



**Erick Costa e Silva Talarico**

**Seismic to Facies Inversion using Convolved  
Hidden Markov Model**

**Dissertação de Mestrado**

Thesis presented to the Programa de Pós-graduação em Matemática da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Matemática.

Advisor: Prof. Sinésio Pesco

Rio de Janeiro  
April 2018



**Erick Costa e Silva Talarico**

**Seismic to Facies Inversion using Convolved  
Hidden Markov Model**

Thesis presented to the Programa de Pós-graduação em Matemática da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Matemática. Approved by the undersigned Examination Committee.

**Prof. Sinésio Pesco**

Advisor

Departamento de Matemática – PUC-Rio

**Prof. Hélio Côrtes Vieira Lopes**

Departamento de Informática – PUC-Rio

**Prof. José Eduardo Mendonça Lira**

Petrobras

**Prof. Paulo Marcos de Carvalho**

Petrobras

**Prof. Márcio da Silveira Carvalho**

Vice Dean of Graduate Studies

Centro Técnico Científico – PUC-Rio

Rio de Janeiro, April 19th, 2018

All rights reserved.

### **Erick Costa e Silva Talarico**

Majored in Physics by the Pontifical Catholic University (PUC, Rio de Janeiro, Brazil) in 2010, and completed a Masters degree in Physics at the same University. Employed at Petrobras for 7 years as a Geophysicist in the Exploration & Production Executive Department. His work centers on seismic data attributes, primarily AVO (Angle versus Offset) analysis, as tools for exploration risk quantification and mitigation.

#### Bibliographic data

Silva Talarico, Erick Costa

Seismic to Facies Inversion using Convolved Hidden Markov Model / Erick Costa e Silva Talarico; advisor: Sinésio Pesco. – Rio de Janeiro: PUC-Rio, Departamento de Matemática, 2018.

v., 120 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Matemática.

Inclui bibliografia

1. Matemática – Teses. 2. Matemática Aplicada – Teses. 3. Sísmica;. 4. Amplitude versus ângulo;. 5. Inversão Sísmica;. 6. Avaliação de incerteza;. 7. Modelo de Markov Oculto;. 8. Modelo Convolutacional.. I. Pesco, Sinésio. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Matemática. III. Título.

CDD: 620.11

To my family, for their support  
and encouragement.

## Acknowledgments

I would like to first thank my advisor Sinesio for his support, friendship, and guidance.

Then I acknowledge Petrobras, for allowing me to do Masters. Specially my current manager Álvaro Arouca for his patience, understanding and guidance, and my previous manager Raul Damasceno for believing in this project.

I would like to thank my colleagues in Petrobras, for their continuous support, and enlightening discussions and suggestions. Specially, Paulo Carvalho, and José Lira for their feedbacks on the manuscript.

I would like to express my gratitude to all professors, staff and fellow students at PUC-Rio, for all the knowledge they shared, and all the help.

Finally, I would like to thank my wife for her extreme patience during this two years of intense studying, but mainly for helping me in the hard times.

## Abstract

Silva Talarico, Erick Costa; Pesco, Sinésio (Advisor). **Seismic to Facies Inversion using Convolved Hidden Markov Model**. Rio de Janeiro, 2018. 120p. Dissertação de Mestrado – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

Oil and Gas Industry uses seismic data in order to unravel the distribution of rock types (*facies*) in the subsurface. But, despite its widespread use, seismic data is noisy and the inversion from seismic data to the underlying rock distribution is an ill-posed problem. For this reason, many authors have studied the topic in a probabilistic formulation, in order to provide uncertainty estimations about the solution of the inversion problem. The objective of the present thesis is to develop a quantitative method to estimate the probability of hydrocarbon bearing reservoir, given a seismic reflection profile, and, to integrate geological prior knowledge with geophysical forward modelling. One of the newest methods for *facies* inversion is used: Convolved Hidden Markov Model (more specifically the Projection Approximation from (1)). It is demonstrated how Convolved HMM can be reformulated as an ordinary Hidden Markov Model problem (which models geological prior knowledge). Seismic AVA theory is introduced, and used with Convolved HMM theory to solve the seismic to *facies* problem. The performance of the inversion technique is measured with common machine learning scores, in a broad set of realistic experiments. The technique capability of estimating reliable probabilities is quantified, and it is shown to present distortions smaller than 5%. As a conclusion, the studied Projection Approximation is applicable for risk management in Oil and Gas applications, which integrates geological and geophysical knowledge.

## Keywords

Seismic; Amplitude versus Angle; Seismic Inversion; Uncertainty assessment; Hidden Markov Model; Convolutional model.

## Resumo

Silva Talarico, Erick Costa; Pesco, Sinésio. **Inversão Sísmica para fácies usando modelo de Markov oculto com efeito convolutivo**. Rio de Janeiro, 2018. 120p. Dissertação de Mestrado – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

A indústria de óleo e gás utiliza a sísmica para investigar a distribuição de tipos de rocha (*facies*) em subsuperfície. Por outro lado, apesar de seu corriqueiro uso em geociências, medidas sísmicas costumam ser ruidosas, e a inversão do dado sísmico para a distribuição de *facies* é um problema mal posto. Por esta razão, diversos autores estudam esta inversão sob o ponto de vista probabilístico, para ao menos estimar as incertezas da solução do problema inverso. O objetivo da presente dissertação é desenvolver método quantitativo para estimar a probabilidade de reservatório com hidrocarboneto, dado um traço sísmico de reflexão, integrando modelagem sísmica direta, e conhecimento geológico a priori. Utiliza-se, um dos métodos mais recentes para resolver o problema inverso: Modelo de Markov Oculto com Efeito Convolutivo (mais especificamente, a Aproximação por Projeção de (1)). É demonstrado que o método pode ser reformulado em termos do Modelo de Markov Oculto (MMO) ordinário. A teoria de sísmica de AVA é apresentada, e usada conjuntamente com MMO com Efeito Convolutivo para resolver a inversão de sísmica para *facies*. A técnica de inversão é avaliada usando-se medidas difundidas em Aprendizado de Máquina, em um conjunto de experimentos variados e realistas. Apresenta-se uma técnica para medir a capacidade do algoritmo em estimar valores confiáveis de probabilidade. Pelos testes realizados a aproximação por projeção apresenta distorções de probabilidade inferiores a 5%, tornando-a uma técnica útil para a indústria de óleo e gás.

## Palavras-chave

Sísmica; Amplitude versus ângulo; Inversão Sísmica; Avaliação de incerteza; Modelo de Markov Oculto; Modelo Convolutivo.

## Table of contents

1	Introduction	16
1.1	Motivation: Seismic Data	16
1.2	Objectives of this Thesis	17
1.3	Chapters	18
2	Technical Background	19
2.1	Introduction	19
2.2	Rock Types	19
2.3	Quantitative Interpretation	21
2.4	Related Works	22
3	Seismic Forward Modelling	25
3.1	Introduction	25
3.2	Reflection from an Interface	25
3.3	Seismic Modelling from an Elastic Profile	27
3.4	Elastic Modelling from a <i>Facies</i> Profile	29
3.5	A Priori Information about <i>Facies</i>	30
3.6	Seismic Inversion to <i>Facies</i>	31
4	HMM Theory	33
4.1	Introduction	33
4.2	Hidden Markov Model	33
4.3	Closing Remarks	37
5	Convolved HMM	38
5.1	Introduction	38
5.2	Relation between <i>Facies</i> Inversion and HMM	38
5.3	Projection Approximation	40
5.4	Complete Likelihood	46
5.5	Exact Inversion Assessment	48
5.6	Elastic Inversion Conditioned on <i>Facies</i>	49
6	Application on Synthetic Examples	50
6.1	Introduction	50
6.2	Performance Measures	50
6.3	Dataset Generation	53
6.4	Base Case	55
6.5	Transition Matrix	57
6.6	Impedances Variability	59
6.7	Acquisition Error	62
6.8	Wavelet	64
7	Conclusions	68
7.1	Future Works	69

Bibliography	<b>70</b>
A Hidden Markov Model	<b>76</b>
A.1 Introduction	76
A.2 Easy Questions on HMM	76
A.3 Posterior Normalization Constant	77
A.4 Posterior Marginal Probabilities	79
A.5 Sampling from the Posterior Distribution	80
A.6 Most Probable Sequence a Posteriori	81
B Gaussian Distribution	<b>85</b>
B.1 Definitions	85
B.2 Affine Transformations	85
B.3 Bayesian Inversion	87
B.4 Gaussian Marginalization	90
B.5 Affine Transformations and Marginalization	91
C Gaussian Approximation to a Gaussian Mixture	<b>92</b>
C.1 Introduction	92
C.2 Optimization	92
D AVA Matrices	<b>95</b>
D.1 Definitions	95
D.2 Linearisation of AVA Equations	96
D.3 AVA Reflectivity Profile	98
D.4 AVA Seismic	100
D.5 AVA Linear Inversion	102
E Markov Chain Monte Carlo	<b>106</b>
E.1 Introduction	106
E.2 MCMC Outlined	107
F Inversion Figures	<b>111</b>

## List of figures

- Figure 1.1 Seismic experiment schematic<sup>1</sup>(top left). The mechanical perturbation propagates away from the source, reflects at subsurface discontinuities and is recorded by the sensors. The seismic processing produces an image of reflected amplitudes (top right). From the seismic reflection image, geophysicists try to infer the distribution of rocks in the subsurface (illustrated by reservoir rock in yellow and non reservoir in green) and support well drilling (bottom). 17
- Figure 2.1 Synthetic *facies* sequence, and corresponding property profile. The yellow samples represent oil-bearing sandstone, and the green samples represent shale. 20
- Figure 2.2 Sketch, adapted from (2), comparing direct seismic to *facies* inversion, and sequential inversion, from seismic to elastic properties, and then to *facies*. 22
- Figure 3.1 The figure<sup>2</sup> illustrates the incidence of a plane wave on an interface, and the associated reflected, and transmitted waves. In this thesis, only the P-wave reflection will be studied. 26
- Figure 3.2 Example of a matrix  $\mathbf{G}$  plotted in grayscale. The inputs are described in the columns, and the outputs in the rows. The matrix has a banded structure, because of the convolution effect in equation 3-6. The interval  $1+l'$  to  $N-l'$  represents the output sample range of the seismic trace, as discussed in D.4. 28
- Figure 3.3 Typical seismic wavelet in time (top) and frequency domain (bottom). 29
- Figure 3.4 Convolved HMM forward model. The  $\pi$  profile is the *facies* sequence, the  $m$  profile is the corresponding elastic properties, and  $s$  is the recorded seismic AVA. In the illustration,  $k' = 2l' + 1$  is the length of the seismic effective kernel. See appendix D for more details on the convolution formulation. 31
- Figure 5.1 The seismic mean predictions, and 80% confidence intervals for all *facies* patterns are stacked in this figure. The red dashed lines represent a conservative measure of the influence range for patterns of size 11, and seismic kernel of size 31. 44
- Figure 5.2 The figures illustrate the approximate seismic forward model for two possible patterns. The output of pattern in the right shows higher similarity with the seismic profile, so it will tend to be associated with higher posterior probability. Not coincidentally the pattern on the right has higher similarity with the true *facies* sequence, although this is not a rule that higher likelihoods come from solutions closer to the true geology. 45

- Figure 5.3 The figures shows this thesis implementation of the inversion, which uses additional *facies* samples in the inversion. CI stands for confidence interval. 47
- Figure 5.4 The figures illustrates the strategy used to compute the likelihood with missing data, for a pattern size of  $k = 3$ , and seismic kernel size of  $k' = 5$ . In the seismic central part 5.4(a), there is no missing samples, so the full conditional covariance and mean are used. At the borders 5.4(b), some seismic samples will be missing for comparison with the pattern synthetic, so just the corresponding indices in the conditional covariance matrix and mean, will be used. 47
- Figure 6.1 Illustration of precision and recall scores<sup>3</sup>. 51
- Figure 6.2 Probability calibration graph for a well calibrated classification algorithm. The distortion score is the RMSE between the fraction of positives and mean probability estimates 52
- Figure 6.3 The graphs show, as a function of the lag parameter  $k$ , classification scores 6.3(a) accuracy, 6.3(b) precision, and 6.3(c) recall for the base case. There is no significant difference in accuracy, precision, and recall with increasing  $k$ . 55
- Figure 6.4 The graphs show the time spent 6.4(a) in the overhead processing, 6.4(b) in the forward recursion, and 6.4(c) per sampling, as a function of the lag parameter  $k$ . An exponential increase in overhead and forward recursion time can be noticed, and the elapsed time per sampling is approximately constant, as expected. 56
- Figure 6.5 Mean acceptance rate as a function of lag parameter  $k$ . 57
- Figure 6.6 The graphs illustrate *facies* profiles sampled from the prior probability distribution, for the three transition matrix scenarios: 6.6(a), equiprobable, 6.6(b) mildly unbalanced, and 6.6(c) highly unbalanced. 58
- Figure 6.7 The graphs compare the classification scores 6.7(a) accuracy, 6.7(b) precision, and 6.7(c) recall between the three transition matrix scenarios. 58
- Figure 6.8 Mean acceptance rate as a function of transition matrix parametrization. 59
- Figure 6.9 The graph show the experimental marginal posterior probability calibration for each of the three transition matrix scenarios. The size of the points represent the amount of samples in each bin. 59
- Figure 6.10 The graphs illustrate log-impedance profiles sampled from the prior probability distribution, and the corresponding theoretical histograms for the three impedance variability scenarios: 6.10(a) and 6.10(d) small variability, 6.10(b) and 6.10(e) mild variability, and 6.10(c) and 6.10(f) big variability. 60
- Figure 6.11 The graphs compare the classification scores 6.11(a) accuracy, 6.11(b) precision, and 6.11(c) recall between the three impedance variability scenarios. 61

Figure 6.12 Mean acceptance rate as a function of impedance variability parametrization.	61
Figure 6.13 The graph show the experimental marginal posterior probability calibration for each of the three impedance variability scenarios. The size of the points represent the amount of samples in each bin.	62
Figure 6.14 The graphs illustrate the <i>facies</i> profiles sampled from the prior probability distribution and the modelled seismic, for the three signal-to-noise scenarios: 6.14(a) low noise level, 6.14(b) mild noise level, and 6.14(c) high noise level.	63
Figure 6.15 The graphs compare the classification scores 6.15(a) accuracy, 6.15(b) precision, and 6.15(c) recall between the three acquisition error scenarios.	63
Figure 6.16 Mean acceptance rate as a function of acquisition error parametrization.	63
Figure 6.17 The graph shows the experimental marginal posterior probability calibration for each of the three transition matrix scenarios. The size of the points represent the amount of samples in each bin.	64
Figure 6.18 The graphs illustrate the seismic effective kernel (wavelet and derivative effects), and corresponding seismic for the three wavelet scenarios: 6.18(a)-6.18(d) short wavelet, 6.18(b)-6.18(e) medium wavelet, and 6.18(c)-6.18(f) long wavelet.	65
Figure 6.19 The graphs compare the classification scores 6.19(a) accuracy, 6.19(b) precision, and 6.19(c) recall between the three wavelet scenarios.	66
Figure 6.20 Mean acceptance rate as a function of the seismic wavelet.	66
Figure 6.21 The graph shows the experimental marginal posterior probability calibration for each of the three transition matrix scenarios. The size of the points represent the amount of samples in each bin.	66
Figure B.1 Illustration on the posterior distribution dependency on the signal-to-noise ratio (SNR). B.1(c) shows, for SNR= 1, a comparison between prior and posterior distributions. B.1(a) shows the conditional mean as a function of the signal-to-noise ratio. B.1(b) shows the conditional standard deviation as a function of the signal-to-noise ratio.	89
Figure D.1 log approximation (in red) to the elastic contrast terms $\frac{\Delta x}{x}$ (in black)	97
Figure D.2 Error of the log approximation	97
Figure D.3 Convolution calculation for each output sample $s_n$ .	102
Figure D.4 Valid convolution illustration, a limited number of output samples are computed	102

- Figure D.5 The graphs compare the Bayesian linear inversion result, for different prior models. D.5(a) shows the inversion results with the correct prior model, D.5(b) shows the result with a constant deviation in the correct prior, and D.5(c) shows the result with a random oscillatory perturbation (within the wavelet spectrum) to the prior. The recovered impedances are different, but the modelled seismic is undistinguishable. 104
- Figure F.1 The graphs show, the inversion results for the base case. F.1(a), shows the true *facies* and the marginal posterior probabilities, per *facies*. F.1(b) shows 500 *facies* profiles sampled from the posterior distribution, stacked side-by-side. F.1(c) shows the joint impedance inversion, and the forward seismic modelling, as a sanity check. 112
- Figure F.2 The graphs show, the inversion results for the equiprobable case. F.2(b), shows the true *facies* profile and the posterior marginal probabilities, per *facies*. F.2(a) shows the 500 *facies* profiles sampled from the posterior distribution, stacked side-by-side. F.2(c) shows the joint impedance inversion, and the forward seismic modelling, as a sanity check. 113
- Figure F.3 The graphs show, the inversion results for the highly unbalanced case. F.2(b), shows the true *facies* profile and the posterior marginal probabilities, per *facies*. F.2(a) shows the 500 *facies* profiles sampled from the posterior distribution, stacked side-by-side. F.2(c) shows the joint impedance inversion, and the forward seismic modelling, as a sanity check. 114
- Figure F.4 The graphs show, the inversion results for the small impedance variability case. F.4(a) shows the 5000 *facies* profiles sampled from the posterior distribution, stacked side-by-side. F.4(b), shows the corresponding marginal probabilities, per *facies*. And F.4(c) shows the joint impedance inversion, and the forward seismic modelling, as a sanity check. 115
- Figure F.5 The graphs show, the inversion results for the big impedance variability case. F.5(a) shows the 5000 *facies* profiles sampled from the posterior distribution, stacked side-by-side. F.5(b), shows the corresponding marginal probabilities, per *facies*. And F.5(c) shows the joint impedance inversion, and the forward seismic modelling, as a sanity check. 116
- Figure F.6 The graphs show, the inversion results for the small acquisition error case. F.6(a) shows the 5000 *facies* profiles sampled from the posterior distribution, stacked side-by-side. F.6(b), shows the corresponding marginal probabilities, per *facies*. And F.6(c) shows the joint impedance inversion, and the forward seismic modelling, as a sanity check. 117

- Figure F.7 The graphs show, the inversion results for the high acquisition error case. F.7(a) shows the 5000 *facies* profiles sampled from the posterior distribution, stacked side-by-side. F.7(b), shows the corresponding marginal probabilities, per *facies*. And F.7(c) shows the joint impedance inversion, and the forward seismic modelling, as a sanity check. 118
- Figure F.8 The graphs show, the inversion results for the high resolution wavelet case. F.8(a) shows the 5000 *facies* profiles sampled from the posterior distribution, stacked side-by-side. F.8(b), shows the corresponding marginal probabilities, per *facies*. And F.8(c) shows the joint impedance inversion, and the forward seismic modelling, as a sanity check. 119
- Figure F.9 The graphs show, the inversion results for the low resolution wavelet case. F.9(a) shows the 5000 *facies* profiles sampled from the posterior distribution, stacked side-by-side. F.9(b), shows the corresponding marginal probabilities, per *facies*. And F.9(c) shows the joint impedance inversion, and the forward seismic modelling, as a sanity check. 120

## List of Abbreviations

$s$  – seismic data. It is a vector of size  $NM$ , where  $N$  is the number of samples, and  $M$  is the number of angles of observation.

$m$  – log of the elastic properties data. It is a vector of size  $3N$ , where  $N$  is the number of samples, and 3 stands for the 3 isotropic elastic properties ( $\rho$ ,  $v_P$ , and  $v_S$ ).

$\rho$  – density

$v_P$  – compressional wave velocity

$v_S$  – shear wave velocity

$\pi$  – lithology sequence. It is a vector of size  $N$ , where  $N$  is the number of samples. It is a categorical vector, meaning that each entry belongs to a predefined discrete set of values.

$\mathbb{R}$  – set of real numbers

$p(X)$  – probability distribution of stochastic variable  $X$ , in functional form

$p(X = a)$  – measure of the stochastic event  $X = a$

$p(X|Y)$  – conditional probability distribution of stochastic variable  $X$  given stochastic variable  $Y$

$p(X, Y)$  – joint probability distribution of stochastic variables  $X$  and  $Y$

$\mathcal{N}(x|\mu, \Sigma)$  – Gaussian distribution (multivariate or not depends on the context), of variable  $x$ , with parameters  $\mu$  for mean, and  $\Sigma$  for variance (or covariance matrix)

$\mathcal{U}[a, b]$  – uniform distribution between  $a$  and  $b$

$X_{\sim n}$  – the sub vector of vector  $X$  with element  $n$  omitted

$\partial_k(n)$  – index array  $(n - l, \dots, n + l)$ , where  $k = 2l + 1$

$X_{\partial_k(n)}$  – sub vector with elements  $X_i$ , and  $i \in \partial_k(n)$

$[A]_{\partial_k(n), \partial_k(n)}$  – sub matrix with all elements  $A_{i,j}$ , where  $i, j \in \partial_k(n)$

$[A]_{:, \partial_k(n)}$  – sub matrix with all elements  $A_{i,j}$ , where  $j \in \partial_k(n)$ ,  $\forall i$

$[A]_{\partial_k(n), :}$  – sub matrix with all elements  $A_{i,j}$ , where  $i \in \partial_k(n)$ ,  $\forall j$

$\mathbb{1}$  – identity matrix. Its size is explained in the context

$\mathbf{0}$  – matrix of zeros. Its size is explained in the context

$\mathbb{1}_{\{x,y\}}$  – boolean operator, which yields 1 if  $x$  is equal to  $y$ , and 0 otherwise

# 1 Introduction

## 1.1 Motivation: Seismic Data

Seismic is an indirect method that uses mechanical waves to probe the subsurface. The basic elements for a seismic experiment are seismic sources and sensors.

In the seismic experiment, a mechanical excitation is produced by the source. This excitation produces a wave that travels through the subsurface media. Whenever it hits a discontinuity of elastic property, it produces two waves: a reflected wave, and a transmitted wave. The sensors record the reflected waves arriving at the surface.

Figure 1.1 shows a seismic experiment. The arrows represent the mechanical wave propagation, and reflection on the subsurface discontinuities.

A conventional seismic survey comprises the repetition of the latter experiment for many source-sensor positions, in order to cover a large area. The product of seismic imaging is a field of reflection intensities as a function of time ( $t$ ) and space ( $x,y$ ), as illustrated in figure 1.1 (top right). This 3D reflection image is a function of the subsurface rock properties, and the industry rely on this image to determine where to drill in order to find new hydrocarbon reservoirs.

The problem of defining the subsurface properties from a seismic image, is an inversion problem (3). For many reasons (some discussed in section D.5), this inversion is ill-posed, and the measurements are noisy. For this reason, the Oil and Gas companies rely on risk estimates to rank and manage its portfolio of drilling opportunities (4).

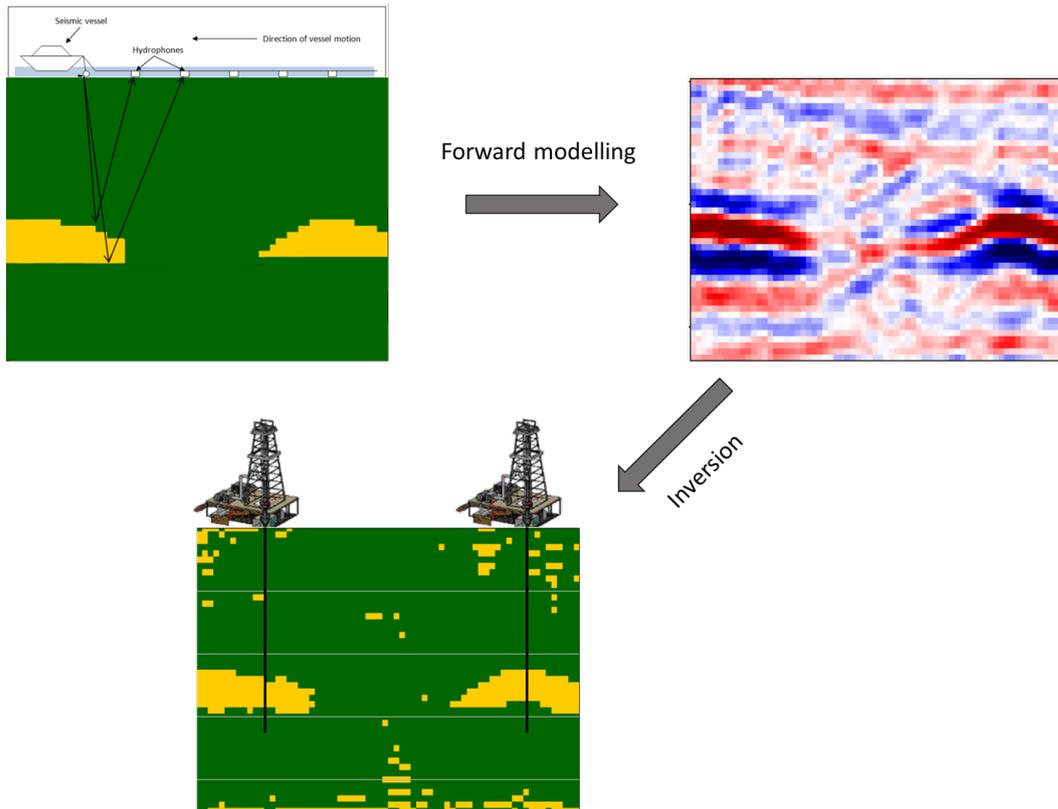


Figure 1.1: Seismic experiment schematic<sup>1</sup>(top left). The mechanical perturbation propagates away from the source, reflects at subsurface discontinuities and is recorded by the sensors. The seismic processing produces an image of reflected amplitudes (top right). From the seismic reflection image, geophysicists try to infer the distribution of rocks in the subsurface (illustrated by reservoir rock in yellow and non reservoir in green) and support well drilling (bottom).

## 1.2 Objectives of this Thesis

The goal of the present thesis is to present a method to quantify the drilling chance of success using seismic reflection data.

In order to accomplish this objective, a probabilistic inversion from seismic reflection image to rock types (*facies*) will be studied. This topic dates back to (5), and is still a topic of research (6, 7, 8, 9).

The inversion formulation adopted in this thesis is known as Convolved Hidden Markov Model (1, 10). This tool will allow to quantify the probability of hydrocarbon reservoir in different points of the subsurface.

More specifically, the projection approximation proposed in (1) will be studied, with focus on the following questions:

- What is the most probable rock type sequence, given a seismic reflection

<sup>1</sup>Figure adapted from Nwhit - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=18527767>, in January 2018

profile?

- What are the marginal probabilities profile for each rock type?
- How to sample possible rock type sequences from the posterior probability distribution?

### 1.3

#### Chapters

Chapter 2 will introduce basic concepts to help understand the general framework in which the present work is situated.

In chapter 3, 1D seismic forward modelling theory will be introduced in order to clarify how the measured seismic dataset is generated. The forward modelling will allow us to estimate the seismic reflection profile, given a *facies* profile.

Chapter 4 will introduce Hidden Markov Model theory, and chapter 5 will show how to generalize this theory as an approximate solution to the Convolved Hidden Markov Model problem.

In order to validate the inversion technique, chapter 6 will analyse its performance with synthetic examples, under different modelling conditions. Among other results, it will be shown that the inversion method estimates reliable probabilities, and thus can be used for risk management by the Oil and Gas industry.

The appendices give detailed mathematical demonstration of the results and tools used throughout this thesis.

## 2 Technical Background

### 2.1 Introduction

The present chapter will discuss two basic concepts for the work. The first one is rock type, or *facies*. And the second one is Quantitative Interpretation. It will also contextualize the present thesis in the seismic to *facies* inversion literature.

### 2.2 Rock Types

Geology studies the types of rocks (lithologies), their properties, occurrence, and genetic relations. *Facies* are the set of textural characteristics that define a rock, such as: mineralogy, pore sizes and shapes, grain sizes and shapes, color, etc. But in this thesis, the terms *facies*, lithology and rock type will be used interchangeably.

Common lithologies in passive margin basins (such as offshore Brazil) are:

- Shales/Silt: fine-grained (micro to nanometer-scale) clay-mineral rich sediments. These are impermeable rocks, and usually do not serve as hydrocarbon reservoirs. Shales represent predominant fraction of typical geological sequences.
- Sandstones: fine to coarse grained (mili to micrometer scale) quartz rich sediments. These are important hydrocarbon reservoirs, but represent small fraction of geological sequences.

As a function of the depositional dynamics, the *facies* sequencing will have some geometrical characteristics, such as: typical *facies* proportion, and thickness. Also, some groups of *facies* will tend to intercalate with others at a certain frequency, while some groups of *facies* will never be in direct contact.

As an example, due to gravitational sorting, a gas bearing sandstone layer can not be directly under an oil or water bearing sandstone layer, as gas

is lighter than oil and water. For the same reason, an oil bearing sand can not be directly under a water bearing sand.

Figure 2.1 shows a synthetic *facies* sequence, together with the corresponding elastic property profile (property which will be introduced in section D.5).

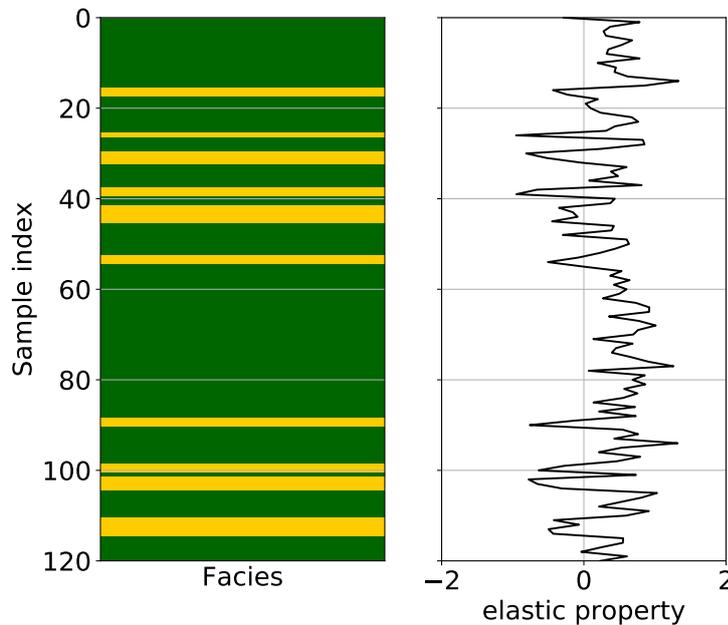


Figure 2.1: Synthetic *facies* sequence, and corresponding property profile. The yellow samples represent oil-bearing sandstone, and the green samples represent shale.

Summarizing, each *facies* occurs in the subsurface, respecting some stacking pattern, which depends on the local geology. First order Markov Chains have been used in geophysics (11, 12) to model *facies* sequences as a stochastic process, respecting prior knowledge about the stacking pattern.

Markov Chains helps to regularize the inversion by reducing the set of possible solutions. Furthermore, Markov Chains also provides a natural way to integrate geological knowledge in the geophysical analysis, while this integration is harder in other approaches, for example in seismic inversion directly to rock properties (as discussed in section D.5).

For the above mentioned reasons, Markov Chain is used in the present thesis.

Another important aspect to be noticed, is that each *facies* presents a range of possible properties (as the dispersion in figure 2.1 illustrates).

The algorithms studied in this thesis will vastly use probability theory, instead of deterministic formalism, to allow modelling the variabilities discussed above.

## 2.3

### Quantitative Interpretation

In chapter 1, it has been introduced the concept of a seismic measurement, and the idea that a seismic image has to be inverted in order to result in useful subsurface information. Historically, this inversion problem has been solved by dividing it in smaller inverse problems (3).

This thesis will study reflection images, which are already the product of some inversion steps, given the acquired seismic data. The present work lies on the last inversion problem, that is to unravel the subsurface geology given seismic reflection images, which is also known as Quantitative Interpretation (6).

Quantitative Interpretation is the field that links *facies*, elastic properties (to be further explained in chapter 3), and seismic reflection data. Figure 2.2 illustrates the relationship between these variables.

The seismic to *facies* inversion studied in this thesis is illustrated in figure 2.2 with arrow 1. It will also be developed the theory to estimate the elastic properties joint conditioned on the inverted *facies* and on the seismic measurement (arrows 2 in figure 2.2).

Other inversion formulations are possible, as illustrated in figure 2.2. One can perform inversion from seismic images to elastic properties (13) (briefly discussed in section D.5). And, given the estimated elastic properties, invert for *facies* (5, 14, 15), which might be referred to as rock physics inversion or *facies* classification.

Even though the *facies* inversion approach (larger arrow in figure 2.2) has to implicitly deal with the elastic parameters, its advantage over the sequential approach (elastic inversion, then rock physics inversion) is that it correctly accounts for geological prior knowledge (presented in chapter 3), and avoids elastic inversion problems (modelling elastic prior knowledge, as discussed in section D.5).

As pointed out in section D.5, seismic to *facies* inversion has an additional advantage. It promotes the interaction between the geoscientists, since it highly depends on the interaction between geophysical forward modelling and geological prior knowledge.

Chapter 3 will describe rock physics and seismic modelling, for these tools represent the forward model of our inversion problem.

However, rock physics topic will be analysed from a statistical point of view. For deeper understanding of rock physics modelling theories, one should read (16).

Although not illustrated in figure 2.2, other variables of interest exist in the Oil and Gas industry. For instance, some authors (17, 18, 19, 20) invert elastic property volumes to petrophysical properties: porosity, oil saturation, and shale content. Another approach would be to invert seismic reflections data directly to petrophysical properties, as in (21, 22).

Finally, some authors even joint invert seismic reflection data to *facies*, elastic properties, and petrophysical properties (23, 24, 25).

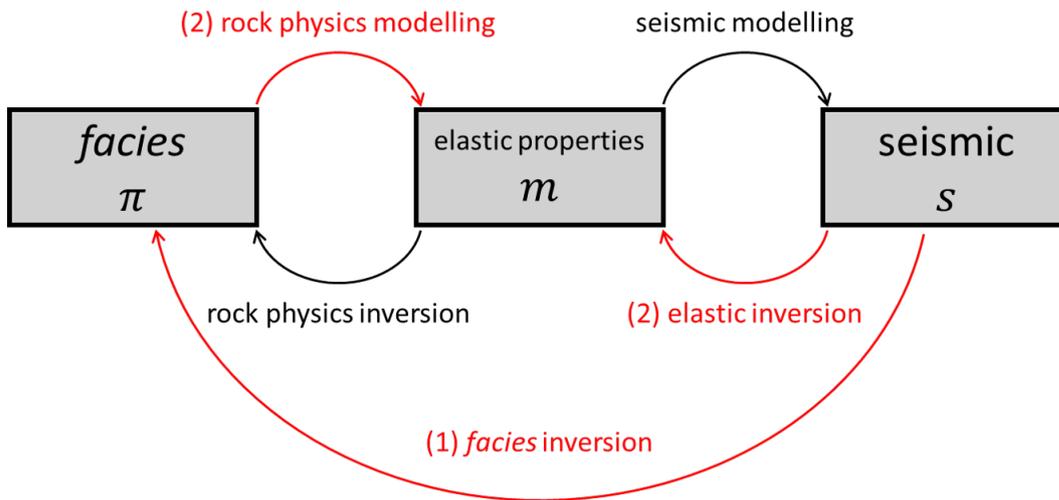


Figure 2.2: Sketch, adapted from (2), comparing direct seismic to *facies* inversion, and sequential inversion, from seismic to elastic properties, and then to *facies*.

From another point of view, an inversion can be deterministic, if it is formulated as a minimization problem, or probabilistic, if it is written in terms of probabilities (using Bayesian formalism (26, 27)).

The formalism studied in this work is probabilistic, because the inverse problem is ill-posed (to be discussed in D.5) and seismic measurement is noisy. So, the best one can do is to take into account all the admissible solutions and measure the probability of finding hydrocarbon bearing reservoir, which cannot be handled by deterministic methods.

## 2.4 Related Works

The present thesis will study the seismic to *facies* inversion technique developed by (1). In this section, we will contextualize the techniques used in this thesis within the literature of seismic to *facies* inversion.

The most simple way to interpret *facies* from seismic is to sequentially invert seismic to elastic properties, and then from elastic properties to *facies*

(5). The problem with this approach is the possible bias introduced by the prior model of the elastic inversion (as will be discussed in section D.5).

(28, 29) are two of the pioneering works that made probabilistic inference directly from reflection images, and it already used the concept of Markov Random Fields, in order to impose a lateral smoothness constrain in the *facies* simulations. These works already used Monte Carlo Markov Chain (MCMC) technique to sample *facies* scenarios from the posterior distribution (this MCMC will also be used in the present work). Both papers assumed the approximation that the reflection was generated by an interface separating two semi infinite spaces, so they inverted a reflection horizon (map) to the possible *facies* in the lower layer, considering a known upper layer.

(8) was the first to elaborate an inversion from a seismic profile directly to *facies* profile. The likelihood approximation used in (8) is maximally factorized, meaning that the likelihood term is approximated as a product of point-wise terms:  $p(s | \pi) \sim \prod_n p(s | \pi_n)$ . Even so, the inversion results were encouraging.

(2) also uses a point-wise approximation for the posterior probabilities as in (8), with a faster algorithm, adequate for inverting large seismic datasets. But, the main reason for the algorithm's efficiency was ignoring *facies* stacking patterns (considering point-wise prior distribution for *facies*) in the *facies* prior distribution. So it did not use Markov Chain.

Both (8) and (2) did not use MCMC technique to assess the exact posterior distribution, but rather presented an approximated posterior distribution as solution to their inversion problem.

Later publications (30, 31, 32) use MCMC sampling in order to assess the exact posterior distribution, but with a Gibbs sampling strategy while the present thesis uses independent proposal Metropolis-Hastings sampling.

The advantage of Gibbs strategy is that one does not need to approximate the likelihood term, since in this technique one always have the full *facies* profile, and perturbs it in order to propose a new *facies* configuration. On the other hand, these publications' algorithms take longer time per *facies* sampling than the present thesis. In (32), 20000 samplings take 30 hours, while section 6.4 shows that with the algorithms used in this thesis 20000 samplings take 50 min, plus a small overhead time for the first forward recursion, which, in the worst case scenario, is approximately 15 min (for  $k = 15$ ).

(33, 34, 35, 36) presents and analyses an extension of (8) model to invert 2D seismic sections. The idea was to include the influence of neighbouring *facies* profiles, in order to sample the current *facies* profile taking into account lateral continuity. Thus, each *facies* profile sampling is done using (8) approximation, while an MCMC Gibbs strategy is used to update the 2D *facies* section

profile-by-profile. The experiments in (33, 36) produced over confident results, thus, underestimating uncertainties.

(37) improves the model in (33), by considering the model parameters (such as transition matrix, wavelet, etc.) uncertainties in the *facies* inversion. As a consequence, they achieve a higher variability in the *facies* posterior distribution, thus, a more robust uncertainty estimation.

(38, 39) improve the point-wise likelihood approximation in (8), by reformulating the problem in terms of *facies* patterns with length  $k$ , just like in this thesis, but with a different likelihood approximation (for a discussion see (1)).

(39, 10) study the joint inference of model parameters (transition matrix, and wavelet) with the posterior *facies* probability estimation.

Some other authors (40, 9) assess the posterior distribution for *facies* by simulating a large amount of *facies* profiles, from the prior probability distributions, and then ranking the simulated profiles by comparing their corresponding modelled seismic profile with the measured seismic profile. This method has the advantage of not using an approximate seismic likelihood, but, on the other hand, it demands a lot of processing time in order to sample enough profiles with good match (high likelihood) to the measured seismic.

Another advantage of this methodology is that it allows for modelling complex prior models, while the convolved HMM developed by (1), demands a Markov Chain type of prior for *facies* profile, and can not appropriately model vertical correlation of rock properties.

In (9), the petrophysical properties associated with the *facies* profile are inferred together with the *facies*.

(41, 7) addressed the idea of joint impedances-facies inversion, using an EM-like algorithm (Expectation-Maximization (26)). The technique is nowadays known as Ji-Fi (joint impedance-facies inversion) as in (24, 42).

In Ji-Fi, one alternates between estimating the best *facies* sequence, given the elastic profile, and then the most probable elastic profile given the *facies* and the seismic reflection profiles. An advantage of this technique is that it uses graphical models to take into account spatial *facies* correlations, other than just vertical (as in Markov Chain). On the other hand, this method is an optimization technique, and does not assess the inversion uncertainty.

## 3 Seismic Forward Modelling

### 3.1 Introduction

Chapter 1 introduces the idea that seismic acquisition provides reflection images (figure 1.1), and that the Oil and Gas industry is interested in determining the subsurface distribution of *facies* given a seismic reflection image.

So, the unknown variable is the *facies* sequence  $\pi$ . The measurements are the seismic profiles  $s$ . And the link between *facies* and seismic are the elastic properties  $m$ .

This chapter will explain seismic forward model using AVA (amplitude versus angle) theory.

First, the equations for reflection intensity on an interface will be discussed. Second, it will be shown how to forward model a seismic reflection profile, given an elastic properties profile. Then, the elastic profile will be modelled from the *facies* profile. Finally, it will be introduced how to use Markov Chain to model the prior knowledge of the *facies* profile.

In the end, Bayes Rule (26) will be used to formulate the inversion problem, and its complexity will be discussed.

### 3.2 Reflection from an Interface

Consider an incident plane compressional wave (P-wave) hitting an interface between two homogeneous isotropic elastic media. From this incident P-wave, four waves are generated: reflected and transmitted compressional and shear waves, as in Figure 3.1.

Nowadays, the most commonly recorded seismic data is compressional wave, so only the  $R_P$  coefficient will be modelled.

Equations 3-1 through 3-5 estimate P-wave reflected amplitude  $R_P$  using Aki Richards 3-term approximation (43). The reflection intensity depends on the elastic contrast between the upper ( $x_1$  in equation 3-5) and lower media ( $x_2$  in equation 3-5).

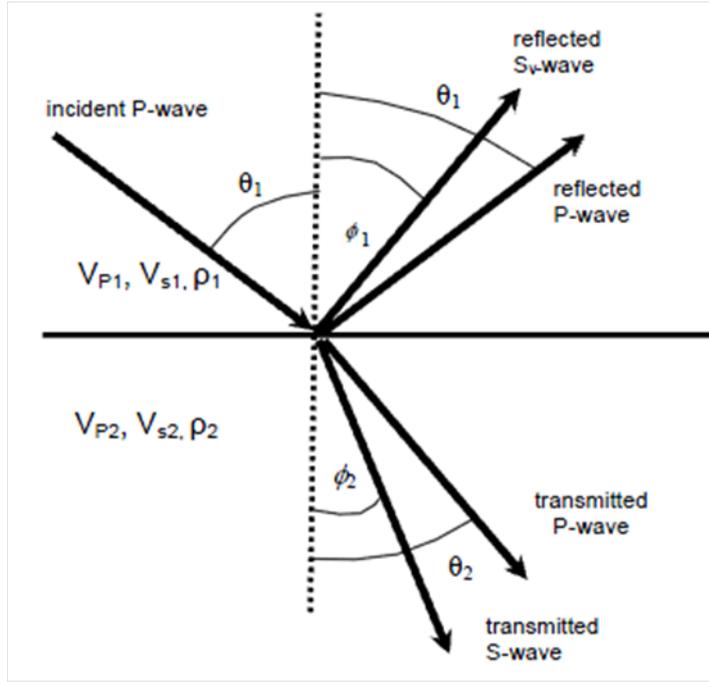


Figure 3.1: The figure<sup>1</sup> illustrates the incidence of a plane wave on an interface, and the associated reflected, and transmitted waves. In this thesis, only the P-wave reflection will be studied.

$$R(\theta) = A + B \sin^2(\theta) + C \tan^2(\theta) \quad (3-1)$$

$$A = \frac{1}{2} \left( \frac{\Delta v_P}{\bar{v}_P} + \frac{\Delta \rho}{\bar{\rho}} \right) \quad (3-2)$$

$$B = -2 \left( \frac{\bar{v}_S}{\bar{v}_P} \right)^2 \left( 2 \frac{\Delta v_S}{\bar{v}_S} + \frac{\Delta \rho}{\bar{\rho}} \right) \quad (3-3)$$

$$C = \frac{1}{2} \frac{\Delta v_P}{\bar{v}_P} \quad (3-4)$$

$$\frac{\Delta x}{\bar{x}} = 2 \frac{x_2 - x_1}{x_2 + x_1} \quad (3-5)$$

In the equations 3-1 through 3-4,  $v_P$ ,  $v_S$ , and  $\rho$  are rock elastic properties.  $v_P$  is the compressional (or P) wave velocity,  $v_S$  is the shear (or S) wave velocity,  $\rho$  is the bulk density.  $\theta$  is an acquisition parameter, and refers to the angle of incidence ( $\theta_1$  in figure 3.1).

From the Appendix D, it has been shown that equation 3-1 can be approximately re-written in matrix form, if parametrized in terms of  $(\log(\rho), \log(v_P), \log(v_S))$  as in (44, 45, 13), and considering  $k = 4 \left( \frac{\bar{v}_S}{\bar{v}_P} \right)^2$  as a known constant.

In fact,  $k = 1$  is a good approximation, and  $k$  is known to vary in

<sup>1</sup>Figure adapted from Evan - Own work, [http://www.subsurfwiki.org/wiki/Zoeppritz\\_equation](http://www.subsurfwiki.org/wiki/Zoeppritz_equation), in January 2018

the range from 0.6 ( $\frac{\bar{v}_P}{\bar{v}_S} = 2.5$ ) to 1.7 ( $\frac{\bar{v}_P}{\bar{v}_S} = 1.5$ ), typically for consolidated sandstones and shales.

### 3.3 Seismic Modelling from an Elastic Profile

This section will show how the seismic trace is generated from an elastic parameters profile.

Suppose that the elastic properties profile is known:  $m_n = (\log(\rho_i), \log(v_{Pi}), \log(v_{Si}))$ ,  $i = n, \dots, N$ . The elastic sequence is distributed on a regular lattice in time domain, where  $n$  is the index along this lattice. Between each sample in this lattice, there is an interface, and, consequently, a reflection.

Earth acts as a bandpass filter to the wave propagation (3). Peg-leg multiples, and seismic attenuation are two physical processes that filter the high and low bands of the wave spectra. For this reason, seismic record will be a blurred version of the reflectivity profile.

The one dimensional convolution model has been used in the industry for decades as a good approximation to the seismic reflection data (for detailed introduction to seismic forward and inverse modelling, refer to (3)). In this context, the reflection profile is modelled as the vertical convolution of the subsurface reflectivity profile and a blurring kernel (commonly known as wavelet), as described in equation 3-6.

$$s(x, y, t, \theta) = w(t, \theta) * R(x, y, t, \theta) := \int_{-\infty}^{+\infty} w(\tau, \theta) R(x, y, t - \tau, \theta) d\tau \quad (3-6)$$

In equation 3-6,  $w(t, \theta)$  is known as the seismic wavelet. It is a function of the seismic apparatus (source spectra for example), as well as a result of the wave propagation dynamics (dispersion, for example).

Appendix D derives the output seismic profile  $\mathbf{s}$  given the elastic profile  $\mathbf{m}$ , in matrix form:

$$\mathbf{s} = \mathbf{WADm} := \mathbf{Gm} \quad (3-7)$$

In equation 3-7, the linear transform  $\mathbf{W}$  computes the convolution with the seismic wavelet from equation 3-6. The matrix  $\mathbf{A}$  represents the reflectivity coefficients (presented in equation D-15). And the linear transform  $\mathbf{D}$  is the differentiation operator.

Also, in equation 3-7,  $\mathbf{s} = \text{vec}(s(\theta_1), \dots, s(\theta_M))$ , and  $\mathbf{m} = \text{vec}(\log(\rho), \log(v_P), \log(v_S))$ , and the operator  $\text{vec}$  is defined in Appendix D.

Figure 3.2 illustrates the general structure of matrix  $\mathbf{G}$ . There are  $M$  output profiles, for this is the number of incidence angles, and there are 3 input profiles:  $v_P$ ,  $v_S$  and  $\rho$ .

It can be seen in figure 3.2, that  $\mathbf{G}$  is a banded matrix, where each band refers to the contribution of an elastic property profile on a seismic profile. The bands are proportional to  $w(\theta) * \frac{d}{dt}$ , the convolution of the seismic wavelet at angle  $\theta$  and the derivative operator. The result of the convolution  $w(\theta) * \frac{d}{dt}$  will be referred to as effective seismic kernel.

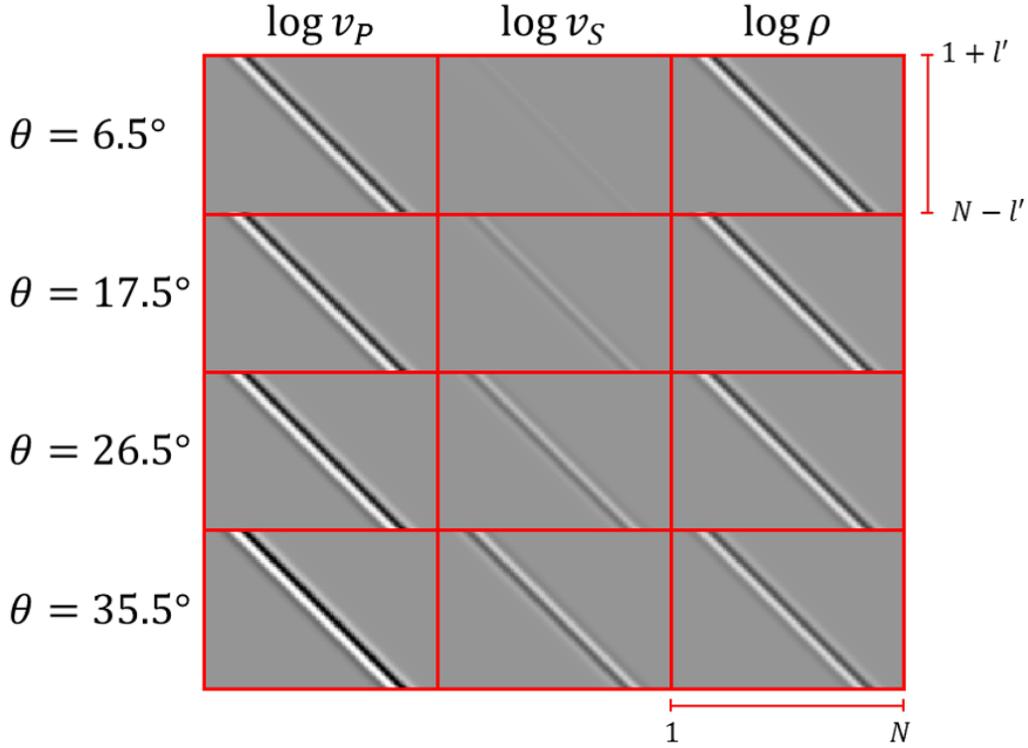


Figure 3.2: Example of a matrix  $\mathbf{G}$  plotted in grayscale. The inputs are described in the columns, and the outputs in the rows. The matrix has a banded structure, because of the convolution effect in equation 3-6. The interval  $1 + l'$  to  $N - l'$  represents the output sample range of the seismic trace, as discussed in D.4.

The seismic effective kernel is responsible for mixing the elastic properties at different positions. A typical seismic wavelet used in seismic modelling is the ricker ("mexican hat"). Figure 3.3 shows the seismic wavelet, the corresponding effective kernel, and its Fourier transform.

From equation 3-7, one can model a noiseless seismic profile, given the elastic profile. Suppose the seismic measurement has some additive Gaussian noise, with covariance matrix  $\Sigma_{ee}$ . Then, the seismic measurement is given by the equation 3-8.

$$p(s | m) = \mathcal{N}(s | \mathbf{G}m, \Sigma_{ee}) \quad (3-8)$$

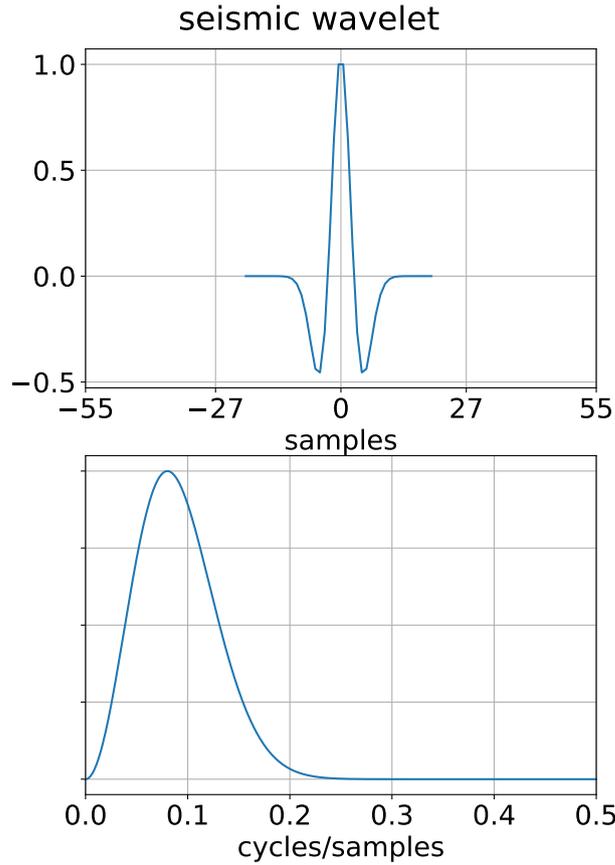


Figure 3.3: Typical seismic wavelet in time (top) and frequency domain (bottom).

### 3.4 Elastic Modelling from a *Facies* Profile

Now, let us assume the *facies* sequence to be known:  $\pi = (\pi_1, \dots, \pi_N)$ , where  $\pi_n \in \Omega = \{0, \dots, K - 1\}$ . Each *facies* have a range of elastic properties, which can be described by a probability distribution:

$$p(m_n | \pi_n = i) = \mathcal{N}(m_n | \mu_i, \Sigma_i) \quad (3-9)$$

In the above equation,  $\mu_i$  is the mean elastic vector ( $\log \rho, \log v_P, \log v_S$ ) for *facies*  $i$ , and  $\Sigma_i$  is the corresponding covariance (of dimensions  $3 \times 3$ ). These parameters might be estimated from well logs.

The present work assumes a probabilistic independence of the elastic parameters in position  $n$  to all the other samples, given the *facies* in position  $n$ . (1) considers spatial dependence among neighbouring elastic parameters, and models this dependency using Gaussian distribution with spatial correlation.

Although elastic properties do have spatial dependency, in real datasets this dependency occurs only within contiguous samples of the same *facies*. As this situation is not easily approximated in the Gaussian approximation step

(equation 5-13), we have decided to avoid modelling spatial dependency.

So, given a *facies* profile, the corresponding elastic profile is described by the following Gaussian distribution:

$$p(m | \pi) = \prod_{n=1}^N \mathcal{N}(m_n | \mu_{\pi_n}, \Sigma_{\pi_n}) = \mathcal{N}(m | \mu_{\pi}, \Sigma_{\pi}) \quad (3-10)$$

$$\mu_{\pi} := \mathbb{E}[m|\pi] \quad (3-11)$$

$$= ([\mu_{\pi_1}]_1, \dots, [\mu_{\pi_N}]_1, [\mu_{\pi_1}]_2, \dots, [\mu_{\pi_N}]_2, [\mu_{\pi_1}]_3, \dots, [\mu_{\pi_N}]_3)$$

$$\Sigma_{\pi} := \text{COV}[m|\pi] \quad (3-12)$$

$$= \begin{bmatrix} [\Sigma_{\pi_1}]_{1,1} & \dots & 0 & [\Sigma_{\pi_1}]_{1,2} & \dots & 0 & [\Sigma_{\pi_1}]_{1,3} & \dots & 0 \\ \dots & \dots \\ 0 & \dots & [\Sigma_{\pi_N}]_{1,1} & 0 & \dots & [\Sigma_{\pi_N}]_{1,2} & 0 & \dots & [\Sigma_{\pi_N}]_{1,3} \\ [\Sigma_{\pi_1}]_{2,1} & \dots & 0 & [\Sigma_{\pi_1}]_{2,2} & \dots & 0 & [\Sigma_{\pi_1}]_{2,3} & \dots & 0 \\ \dots & \dots \\ 0 & \dots & [\Sigma_{\pi_N}]_{2,1} & 0 & \dots & [\Sigma_{\pi_N}]_{2,2} & 0 & \dots & [\Sigma_{\pi_N}]_{2,3} \\ [\Sigma_{\pi_1}]_{3,1} & \dots & 0 & [\Sigma_{\pi_1}]_{3,2} & \dots & 0 & [\Sigma_{\pi_1}]_{3,3} & \dots & 0 \\ \dots & \dots \\ 0 & \dots & [\Sigma_{\pi_N}]_{3,1} & 0 & \dots & [\Sigma_{\pi_N}]_{3,2} & 0 & \dots & [\Sigma_{\pi_N}]_{3,3} \end{bmatrix}$$

### 3.5

#### A Priori Information about *Facies*

Section 2.2 have introduced the idea that vertical *facies* sequences, although not deterministic, follow some stacking pattern. It has also been pointed out that *facies* vertical patterns are usually modelled with first order Markov Chains.

(11) discusses probabilistic formulations to describe the possible stacking patterns: first (and higher) order Markov chains, and discrete time Markov chains. The present work adopts the first order Markov chain formalism, due to its easier parametrization.

Considering the *facies* sequence  $\pi$  as a random array, distributed as a first order Markov Chain, the following probability relation holds:

$$p(\pi_n = j | \pi_{n-1} = i) = T_{ij} \quad (3-13)$$

The probability of occurring a given *facies* sequence  $\pi$  is:

$$p(\pi) = p_s(\pi_1) \prod_{n=2}^N p(\pi_n | \pi_{n-1}) \quad (3-14)$$

Where  $p_s$  is the stationary probability distribution associated with transition matrix  $T$ , as defined in equation A-2.

### 3.6 Seismic Inversion to *Facies*

Thus, given a *facies* sequence, one can model the seismic output, using the results from the previous sections. Figure 3.4 illustrates the probabilistic dependency between each variable in the problem. In figure 3.4, each elastic property depends only on the *facies* at the same position, and each seismic sample depends on the neighbouring elastic properties, where the neighbourhood size is equal to the effective kernel size.

As explained in section D.4, this thesis uses the concept of "valid" convolution. As a consequence the output seismic profile has smaller length than the input elastic and *facies* profiles, as illustrated in figure 3.4.

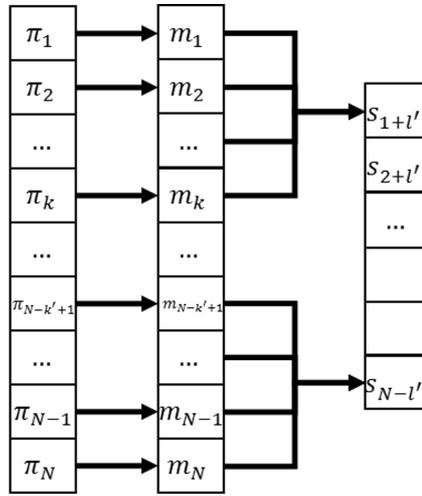


Figure 3.4: Convolved HMM forward model. The  $\pi$  profile is the *facies* sequence, the  $m$  profile is the corresponding elastic properties, and  $s$  is the recorded seismic AVA. In the illustration,  $k' = 2l' + 1$  is the length of the seismic effective kernel. See appendix D for more details on the convolution formulation.

As demonstrated in section B.5, the seismic conditional probability given the *facies* profile is computed by:

$$\begin{aligned} p(s | \pi) &= \int_m p(s | m) p(m | \pi) d^{3N} m \\ &= \mathcal{N}(s | \mathbf{G}\mu_\pi, \mathbf{G}\Sigma_\pi\mathbf{G}^T + \Sigma_{ee}) \end{aligned} \quad (3-15)$$

In order to estimate the inverse conditional probability  $p(\pi|s)$  (also known as posterior probability), the Bayes theorem was used:

$$p(\pi | s) = \frac{p(s | \pi) p(\pi)}{p(s)} \quad (3-16)$$

In equation 3-16, the denominator  $p(s)$ , is a normalization constant, given by the sum of the joint probability distribution  $p(s, \pi) = p(s | \pi) p(\pi)$ , over all possible *facies* sequences:

$$\begin{aligned} p(s) &= \sum_{\pi} p(s | \pi) p(\pi) \\ &= \sum_{\pi_1} \cdots \sum_{\pi_N} p(s | \pi_1, \dots, \pi_N) p(\pi_1, \dots, \pi_N) \end{aligned} \quad (3-17)$$

The great challenge, given a measured seismic profile, is to imply some knowledge about the hidden *facies* sequence, for there are  $K^N$  possible sequences, where typically  $N \approx 100$ , and  $K \approx 3$ .

For example, to find the most probable *facies* sequence, one has to evaluate  $p(s | \pi) p(\pi)$  for all  $K^N$  possible *facies* sequences and find the maximal.

If one is interested in sampling possible *facies* sequences conditioned on the seismic measurement, it would have to, first, compute the normalizing constant (summation over  $K^N$  terms), in order to compute each individual *facies* sequence probability by normalizing the term  $p(s | \pi) p(\pi)$  (normalize each of the  $K^N$  possible sequences).

There is no easy way to evaluate the conditional probability distribution  $p(\pi | s)$ , because each seismic sample depends on the interaction of many neighbouring *facies* samples, making seismic to *facies* inversion a complex combinatorial problem.

Chapter 5 shows the approximate approach used in (1), to efficiently solve the *facies* inversion problem. Since this approach is a generalization of the Hidden Markov Model (HMM) formulation, chapter 4 is devoted to describe HMM probabilistic model and to present its main results.

## 4 HMM Theory

### 4.1 Introduction

In Hidden Markov Models (HMM), there us a sequence of measurements, and each measurement is a property of a hidden categorical variable. One is interested in assessing the categorical variables from the measurements, given predefined knowledge about the stacking patterns of these categories.

Thus, HMM is intimately related to the seismic to *facies* problem, and will be studied further in this chapter.

### 4.2 Hidden Markov Model

Hidden Markov Model (HMM) is a probabilistic model that deals with a sequence of observed measurements, when each observed measurement is associated with a hidden state (categorical variable), also known as latent variable. A good overview of the model is given in (46).

The latent, and the measurement sequences are denoted by  $z$  and  $X$ , as described in equations 4-1 and 4-2.

$$z = (z_1, \dots, z_N), \quad z_n \in \{0, \dots, K - 1\} \quad (4-1)$$

$$X = (x_1, \dots, x_N), \quad x_n \in \mathbb{R}^D \quad (4-2)$$

Where,  $N$  is the size of the hidden states vector, and  $K$  is the number of possible hidden state variable values.  $D$  is the dimension of each measurement.

The name Hidden Markov Model, comes from the fact that the latent sequence is a Markov Chain. The markovian property means, in probability terms, that the conditional probability of a hidden variable, given all previous hidden sequence is only a function of the immediate previous state:

$$p(z_n | z_1, \dots, z_{n-1}) = p(z_n | z_{n-1}) \quad (4-3)$$

Since  $z$  is a categorical variable, equation 4-3 can be stored as a matrix, known as transition matrix:

$$p(z_n = j \mid z_{n-1} = i) = T_{ij} \quad (4-4)$$

Even though, there are many variations of HMM, the canonical definition implies a conditional independence of the measurement in position  $n$  to all other variables, given the hidden state in the same position  $n$ . This relation is described equation 4-5.

$$p(x_n \mid X_{\sim n}, z_1, \dots, z_N) = p(x_n \mid z_n) \quad (4-5)$$

The term  $p(x_n \mid z_n)$  is known as emission probability.

For there are  $K$  possible states for the hidden variables, there are  $K$  different emission probability distributions:  $p(x \mid z_n = i) = f_i(x)$ ,  $\forall n$ . Although not used in this thesis, it would be possible to model an emission probability that depended on  $n$ . It would be useful, for example, to model depth dependent rock-physics.

For an HMM problem, usually, one is interested in accessing the posterior probability distribution  $p(z \mid X)$  for the categorical sequences, given a sequence of measurements. Bayes theorem calculates the posterior probability as:

$$p(z \mid X) = \frac{p(X \mid z) p(z)}{p(X)} \quad (4-6)$$

The major problem with equation 4-6 is the normalization constant  $p(X)$ , which is expensive to be calculated, for it is the sum of  $p(z, X)$ , over all possible hidden sequences  $z$ .

While a naive computation of  $p(X)$  would take  $O(K^N)$  operations, the forward recursion algorithm, derived in section A.3, takes only  $O(NK^2)$  operations:

**Data:** Transition matrix  $T_{ij}$ , and emission probability densities  $f_i$

**Result:** Normalizing constant  $p(X) = \sum_z p(z, X)$ , and  $\alpha$  array

```

1 Initialize  $\alpha_1(i) = f_i(x_1)p_s(z_1 = i)$ ;
2 for  $n = 2$  to  $N$  do
3   | for  $i = 0$  to  $K - 1$  do
4   |   |  $\alpha_n(i) = f_i(x_n) \sum_{j=0}^{K-1} T_{ji} \alpha_{n-1}(j)$ ;
5   | end
6 end
7 Finally:  $p(X) = \sum_{i=0}^{K-1} \alpha_N(i)$ 

```

**Algorithm 1:** Forward Recursion

A direct application of the posterior probability is the computation of the posterior marginal probabilities  $p(z_n = i \mid X)$ . This probability would be

computed with cost  $O(K^N)$  by a naive approach, because it would have to sum  $p(z | X)$  over all possible sequences  $z_{\sim n} = (z_1, \dots, z_{n-1}, z_{n+1}, \dots, z_N)$

The following algorithm, derived in section A.4, computes  $p(z_n = i | X)$  with cost  $O(NK^2)$ :

**Data:**  $\alpha$  array, normalizing constant  $p(X)$ , Transition matrix  $T_{ij}$ ,  
and emission probability densities  $f_i$

**Result:** Posterior marginal probabilities  $p(z_n | X)$  for every sample  
 $n$

```

1 Initialize  $\beta_N(i) = 1, \quad \forall i;$ 
2 Compute: for  $i = 0$  to  $K - 1$  do
3    $p(z_N = i | X) = \frac{\alpha_N(i)\beta_N(i)}{p(X)};$ 
4 end
5 for  $n = N - 1$  to  $1$  do
6   for  $i = 0$  to  $K - 1$  do
7      $\beta_n(i) = \sum_{j=0}^{K-1} \beta_{n+1}(j)f_j(x_{n+1})T_{ij};$ 
8      $p(z_n = i | X) = \frac{\alpha_n(i)\beta_n(i)}{p(X)};$ 
9   end
10 end

```

**Algorithm 2:** Backward Recursion: marginal probability

Another use of the posterior probability is sampling possible hidden sequences  $z$  from it. The sampling algorithm comes from the fact that the posterior distribution can be rewritten as:

$$p(z|X) = p(z_N | X) \times p(z_{N-1} | z_N, X) \times \dots \times p(z_1 | z_2, \dots, z_N, X) \quad (4-7)$$

So, one can sample the hidden sequence, according to the following schedule: first, sample the last hidden variable  $z_N$  from the distribution  $p(z_N | X)$ ; and then, in descending order, sample each hidden variable  $z_n$ , given the previously sampled hidden variables  $z_{n+1}, \dots, z_N$  and the measured sequence  $X$  (according to the distribution  $p(z_n | z_{n+1}, \dots, z_N, X)$ ).

Section A.5 derives the following algorithm, with complexity  $O(NK)$ , for sampling a hidden sequence  $z$  given an observed sequence  $X$ :

**Data:**  $\alpha$  array, posterior marginal probability  $p(z_N | X)$  for the last sample  $N$ , and Transition matrix  $T_{ij}$

**Result:** sequence  $z = (i_1, \dots, i_N)$ , sampled from the posterior distribution  $p(z | X)$

- 1 Sample  $z_N = i_N$  from the probability distribution  $p(z_N | X)$ ;
- 2 **for**  $n = N - 1$  **to** 1 **do**
- 3     Compute  $c = \sum_{j=0}^{K-1} T_{j,i_{n+1}} \alpha_n(j)$ ;
- 4     Sample  $z_n = i_n$  from the probability distribution:  

$$p(z_n = i | z_{n+1} = i_{n+1}, X) = \frac{T_{i,i_{n+1}} \alpha_n(i)}{c}$$
;
- 5 **end**

**Algorithm 3:** Backward Recursion: posterior sampling

The last result in this chapter gives the most probable latent sequence, given an observed sequence. In other words, it finds the hidden sequence that maximizes the posterior probability distribution  $p(z | X)$ .

As for the previous algorithms, a naive implementation would require evaluating  $p(z | X)$  for all possible hidden sequences  $z$ , which is computationally intractable. Section A.6 derives, yet, another flavour of the Forward-Backward Algorithm, commonly known as Viterbi Algorithm:

**Data:** Transition matrix  $T_{ij}$ , and emission probability densities  $f_i$

**Result:** Most probable hidden sequence  $\hat{z}$ :

$$(\hat{z}_1, \dots, \hat{z}_{N-1}, \hat{z}_N) = \arg \max_{z_1, \dots, z_N} p(z_1, \dots, z_N | x_1, \dots, x_N),$$

and its posterior probability:  $p(\hat{z} | X)$

1 **Forward Recursion:**

2 Initialize  $\omega_1(i) = f_i(x_1) p_s(z_1 = i)$ ,  $\forall i \in \{0, \dots, K - 1\}$ ;

3 **for**  $n = 1$  **to**  $N - 1$  **do**

4      $\psi_n(i) = \arg \max_j \{T_{ji} \omega_n(j)\}$ ;

5      $\omega_{n+1}(i) = f_i(x_{n+1}) \max_j \{T_{ji} \omega_n(j)\}$ ;

6 **end**

7 **Backward Recursion:**

8 Initialize:  $\hat{z}_N = \arg \max_i \omega_N(i)$ ;

9 Compute maximal probability:  $p(\hat{z} | X) = \frac{\omega_N(\hat{z}_N)}{p(X)}$ ;

10 **for**  $n = N - 1$  **to** 1 **do**

11      $\hat{z}_n = \psi_n(\hat{z}_{n+1})$ ;

12 **end**

**Algorithm 4:** Viterbi Algorithm

### 4.3

#### Closing Remarks

It is important to notice that the number of operations in the algorithms is  $O(NK^2)$  (except for 3, whose complexity is  $O(NK)$ ), so it is linear in the sequence size, and quadratic on the number of classes (which is usually small, in geologic applications). This chapter has demonstrated that, the posterior probability distribution of an HMM, although defined on state space with high cardinality ( $K^N$ ), is computationally tractable.

Our seismic to *Facies* inversion problem is not a HMM model, because a seismic sample is not independent of the other variables, given the *facies* sample on the same time index  $n$ :  $p(s_n | s_{\sim n}, \pi) \neq p(s_n | \pi_n)$ . Actually, the seismic sample  $n$  is not a function of the *facies* sample  $n$ , but instead it depends on the sum of contributions from neighbouring *facies* samples, due to the convolution effect.

Next chapter presents a reformulation of the seismic to *facies* problem, making it equivalent to the HMM problem. This reformulation makes it possible to apply the algorithms described in the current chapter.

## 5 Convolved HMM

### 5.1 Introduction

Chapter 4 have derived algorithms to assess the posterior probability distribution of a Hidden Markov Model.

The fact that an HMM problem can be expressed as a product of factors involving a maximum of 2 variables makes it possible to develop efficient algorithms to solve for the latent sequence given a sequence of measurements (for a more general solution of factorizable probabilistic models see (47)).

Seismic to *facies* inversion is similar to HMM, for it also involves an observable sequence of measurements (seismic), and a hidden sequence of categorical variables (*facies*). But it does not have the factorizability property of an HMM, because each seismic sample results from the contribution from many neighbouring *facies* samples (making it a more difficult problem to solve).

This chapter introduces approximate methods, developed in (1), to solve seismic to *facies* inversion problem. These methods transform the original problem into a factorizable equivalent problem, making it possible to use the efficient algorithms from the previous chapter.

### 5.2 Relation between *Facies* Inversion and HMM

Assuming a *facies* sequence in a regular lattice in time domain with  $N$  samples. Each sample  $n$  has a rock type denoted by  $\pi_n$ , which belongs to a finite set of  $K$  possibilities  $\Omega = \{0, \dots, K - 1\}$ .

The approximation which will be discussed deal with a strategy to factorize the conditional seismic probability (seismic likelihood) as a product of factors, each one containing  $k = 2l + 1$  *facies* samples (where  $k$  is chosen by the user):

$$p(s | \pi) \approx \prod_{n=l+1}^{N-l} q_n(\pi_{\partial_k(n)}) := \prod_{n=l+1}^{N-l} q_n(\pi_{n-l}, \dots, \pi_{n+l}) \quad (5-1)$$

$\pi_{\partial_k(n)}$  is defined as the *facies* pattern of size  $k$  at position  $n$ .

The prior probability distribution  $p(\pi)$ , although already factorizable in pairs of neighbouring *facies*, can be written in terms of *facies* patterns:

$$\begin{aligned} p(\pi) &= p_s(\pi_1) \prod_{n=2}^N p(\pi_n | \pi_{n-1}) \\ &= p(\pi_{\partial_k(l+1)}) \prod_{n=l+2}^{N-l} p(\pi_{\partial_k(n)} | \pi_{\partial_k(n-1)}) \end{aligned} \quad (5-2)$$

Where equation 5-2 holds, since the following relations hold:

$$p(\pi_{\partial_k(l+1)}) = p_s(\pi_1) \prod_{n=2}^k p(\pi_n | \pi_{n-1}) \quad (5-3)$$

$$p(\pi_{\partial_k(n)} | \pi_{\partial_k(n-1)}) = p(\pi_{n+l} | \pi_{n+l-1}) \quad (5-4)$$

Equations 5-3 and 5-4 are true for first order Markov Chain. They can be adapted for the case of higher order Markov Chains.

By using the approximate likelihood and the formulation of the prior distribution, the posterior probability distribution can also be factorized as a product of terms in  $\pi_{\partial_k(n)}$ :

$$\begin{aligned} p(\pi | s) \propto p(\pi)(s | \pi) &= p(\pi) \prod_{n=l+1}^{N-l} q_n(\pi_{\partial_k(n)}) \\ &= p(\pi_{\partial_k(l+1)}) q_{l+1}(\pi_{\partial_k(l+1)}) \prod_{n=l+2}^{N-l} p(\pi_{\partial_k(n)} | \pi_{\partial_k(n-1)}) q_n(\pi_{\partial_k(n)}) \end{aligned} \quad (5-5)$$

Now, the categorical vector  $(\pi_1, \dots, \pi_k)$ , can be seen as equivalent to a categorical variable  $z$  in the set  $\tilde{\Omega} = \{0, \dots, K^k - 1\}$ , because  $K^k$  is the number of possible sequences of size  $k$  and  $K$  states. A bijective function  $\psi$  can be built from  $\Omega^k$  to  $\tilde{\Omega}$ :

$$\psi(\pi_1, \dots, \pi_k) = \sum_{i=1}^k \pi_i K^{i-1} = z \quad (5-6)$$

By making a correspondence between  $z_1$  and  $\pi_{\partial_k(l+1)}$ ,  $z_2$  and  $\pi_{\partial_k(l+2)}$ , and so on, one can re-write equation 5-5:

$$p(z_1, \dots, z_{N-l} | s) \propto p(z_1) q_{l+1}(\psi^{-1}(z_1)) \prod_{n=2}^{N-2l} p(z_n | z_{n-1}) q_{n+l}(\psi^{-1}(z_n)) \quad (5-7)$$

In other words, convolutional model becomes an HMM problem on the new variable. The terms  $q_{n+l}(\psi^{-1}(z_n))$  work as likelihood terms. The transition

matrix and the stationary probability are given by:

$$\begin{aligned} p(z_n = j \mid z_{n-1} = i) &= p(\pi_{\partial_k(l+n)} = \psi^{-1}(j) \mid \pi_{\partial_k(l+n-1)} = \psi^{-1}(i)) \\ &= T_{\psi^{-1}(i)_k, \psi^{-1}(j)_k} \prod_{n'=1}^{k-1} \mathbb{1}_{\{\psi^{-1}(j)_{n'}, \psi^{-1}(i)_{n'+1}\}} \end{aligned} \quad (5-8)$$

$$\begin{aligned} p(z_1 = i) &= p(\pi_{\partial_k(l+1)} = \psi^{-1}(i)) \\ &= p_s(\psi^{-1}(i)_1) \prod_{n=2}^k T_{\psi^{-1}(i)_{n-1}, \psi^{-1}(i)_n} \end{aligned} \quad (5-9)$$

In the above equations,  $T$  is the transition matrix of the Markov Chain in terms of  $\pi$ , and  $\mathbb{1}_{\{x,y\}}$  equals 1 if  $x$  is equal to  $y$ , and 0 otherwise.

All the efficient algorithms developed in chapter 4 for HMM can be used here to assess the posterior probability distribution over  $z$ , and consequently over  $\pi$ .

The complexity of algorithms 1 and 2 for HMM are  $O(NK^2)$ , for a problem with cardinality  $K$ , and  $N$  samples. Since the number of elements in  $\tilde{\Omega}$  is  $K^k$ , and the number of samples is reduced to  $N - 2l = N - k + 1$ , it yields  $O((N - k + 1)K^{2k})$  for the *facies* inversion case.

Actually, the complexity is  $O((N - k + 1)K^{k+1})$ , since equation 5-8 restricts each state transition to only  $K$  possible states, instead of  $K^k$ . For the same reason, the sampling algorithm 3 has the complexity  $O((N - k + 1)K)$ , in the *facies* inversion case.

The next section will explain the likelihood approximation that factorizes in the required form of equation 5-1.

### 5.3 Projection Approximation

This approach was developed in (1). The strategy is to marginalize the effect of *facies* farther than  $l$  lags from the central sample  $n$  (where the chosen pattern size is  $k = 2l + 1$ ).

$$\begin{aligned} p(s \mid \pi_{\partial_k(n)}) &= \sum_{\pi_1} \cdots \sum_{\pi_{n-l-1}} \sum_{\pi_{n+l+1}} \cdots \sum_{\pi_N} p(s, \pi_{\sim \partial_k(n)} \mid \pi_{\partial_k(n)}) \\ &= \sum_{\pi_1} \cdots \sum_{\pi_{n-l-1}} \sum_{\pi_{n+l+1}} \cdots \sum_{\pi_N} p(s \mid \pi) p(\pi_{\sim \partial_k(n)} \mid \pi_{\partial_k(n)}) \end{aligned} \quad (5-10)$$

This marginalization is not feasible because one would have to sum  $K^{N-k}$  possible *facies* sequences.

Instead, the likelihood approximation is developed by focusing on the elastic properties. If one rewrite the conditional probability distribution  $p(s \mid \pi_{\partial_k(n)})$  as the marginalization in relation to the elastic properties trace:

$$\begin{aligned}
p(s \mid \pi_{\partial_k(n)}) &= \int_m p(s, m \mid \pi_{\partial_k(n)}) \\
&= \int_m p(s \mid m) p(m \mid \pi_{\partial_k(n)}) \\
&= \int_{m_{\partial_k(n)}} \int_{m_{\sim \partial_k(n)}} p(s \mid m) p(m_{\partial_k(n)}, m_{\sim \partial_k(n)} \mid \pi_{\partial_k(n)}) \\
&= \int_{m_{\partial_k(n)}} \int_{m_{\sim \partial_k(n)}} p(s \mid m) p(m_{\sim \partial_k(n)} \mid m_{\partial_k(n)}, \pi_{\partial_k(n)}) p(m_{\partial_k(n)} \mid \pi_{\partial_k(n)})
\end{aligned} \tag{5-11}$$

In equation 5-11, the inner integral is the marginalization over the variables  $m_{\sim \partial_k(n)}$ , for which there is no knowledge about the corresponding *facies* samples. The outer integration is over the variables  $m_{\partial_k(n)}$ , which correspond to the *facies* samples of interest  $\pi_{\partial_k(n)}$ .

Since, the probabilistic model for  $m$  is a Hidden Markov Model ( $\pi$  being the hidden states), the term  $p(m_{\sim \partial_k(n)} \mid m_{\partial_k(n)}, \pi_{\partial_k(n)})$  is equal to  $p(m_{\sim \partial_k(n)} \mid \pi_{\partial_k(n)})$ , due to conditional independence. Now, two approximations comes in place.

The first one is that  $p(m_{\sim \partial_k(n)} \mid \pi_{\partial_k(n)}) \approx p(m_{\sim \partial_k(n)})$ , which is approximately true if the mixing time of the *facies* markov chain is small enough.

The second approximation consists in substituting the probability distribution  $p(m)$ , which is a Gaussian mixture, with a single Gaussian distribution.

$$p(m) = \sum_{\pi \in \Omega^N} p(m \mid \pi) p(\pi) \approx p_*(m) \tag{5-12}$$

The best approximation is achieved by minimizing the Kullback-Leibler divergence from the the approximate distribution  $p_*(m)$  to the true one  $p(m)$ :

$$\begin{aligned}
\arg \min_{\mu_*, \Sigma_*} D_{\mathbf{KL}}(p \parallel p_*) &= - \int_m p(m) \log \frac{p_*(m)}{p(m)} \\
s.t. \quad p_*(m) &= \mathcal{N}(m \mid \mu_*, \Sigma_*)
\end{aligned} \tag{5-13}$$

It is shown in appendix C that the solution to the optimization problem 5-13 is:

$$\mu_* = (\mu_S, \dots, \mu_S), \quad \mu_S = \sum_{i=0}^K p_s(i) \mu_i \quad (5-14)$$

$$[\Sigma_*]_{n,n+l} = \sum_{i=0}^K \sum_{j=0}^K p(\pi_1 = i, \pi_{n+l} = j) (\mu_i - \mu_S)(\mu_j - \mu_S)^T, \quad l \geq 1 \quad (5-15)$$

$$[\Sigma_*]_{n,n} = \sum_{i=0}^K p_s(\pi = i) [\Sigma_i + (\mu_i - \mu_S)(\mu_i - \mu_S)^T] \quad (5-16)$$

The above equations are different from the ones obtained in (1), because of the assumption of elastic conditional independence  $p(m | \pi) = \prod_n p(m_n | \pi_n)$ . In (1), it is assumed that elastic properties are spatially correlated, given the *facies* sequence. But, by the way it is formulated, it creates correlation between properties across *facies* boundaries, which is not reasonable. So, in order to use the cleanest approach, this spatial correlation is not considered in the present work.

Adapting the notation used in (1), we will denote the approximate likelihood as  $p_*^{(k)}(s | \pi_{\partial_k(n)})$ . Using the above mentioned approximations, one can rewrite equation 5-11 as:

$$\begin{aligned} p_*^{(k)}(s | \pi_{\partial_k(n)}) &= \int_{m_{\partial_k(n)}} \int_{m_{\sim \partial_k(n)}} p(s | m) p_*(m_{\sim \partial_k(n)}) p(m_{\partial_k(n)} | \pi_{\partial_k(n)}) \\ &= \int_{m_{\partial_k(n)}} \int_{m_{\sim \partial_k(n)}} p_*(s, m_{\sim \partial_k(n)} | m_{\partial_k(n)}) p(m_{\partial_k(n)} | \pi_{\partial_k(n)}) \\ &= \int_{m_{\partial_k(n)}} p_*(s | m_{\partial_k(n)}) p(m_{\partial_k(n)} | \pi_{\partial_k(n)}) \end{aligned} \quad (5-17)$$

The term  $p_*(s | m_{\partial_k(n)})$  in equation 5-17 might be computed by marginalizing  $p_*(m, s)$  with respect to  $m_{\sim \partial_k(n)}$  and, then, conditioning over  $m_{\partial_k(n)}$ .

First,  $p_*(m, s)$  is computed using the results from section B.2:

$$\begin{aligned} p_*(m, s) &= p(s | m) p_*(m) \\ &= \mathcal{N} \left( \begin{bmatrix} m \\ s \end{bmatrix} \middle| \begin{bmatrix} \mu_* \\ \mathbf{G}\mu_* \end{bmatrix}; \begin{bmatrix} \Sigma_* & \Sigma_* \mathbf{G}^T \\ \mathbf{G}\Sigma_* & \mathbf{G}\Sigma_* \mathbf{G}^T + \Sigma_{ee} \end{bmatrix} \right) \end{aligned} \quad (5-18)$$

The mean background  $\mu_*$  will be constant as shown in equation 5-14, unless the *facies* markov chain is inhomogeneous. Thus the term  $\mathbf{G}\mu_*$  is usually zero, because  $\mathbf{G}$  is a composition of a derivative operator and a wavelet. For now on, it will be assumed that  $\mathbf{G}\mu_* = \mathbf{0}$ . This simplification is not necessary

for the approximation, but will make the equations uncluttered.

Second, the marginalization over  $m_{\sim\partial_k(n)}$  (section B.4) yields:

$$p_*(m_{\partial_k(n)}, s) = \mathcal{N} \left( \begin{bmatrix} m_{\partial_k(n)} \\ s \end{bmatrix} \middle| \begin{bmatrix} [\mu_*]_{\partial_k(n)} \\ \mathbf{0} \end{bmatrix}; \begin{bmatrix} [\Sigma_*]_{\partial_k(n), \partial_k(n)} & [\Sigma_*]_{\partial_k(n), :} \mathbf{G}^T \\ \mathbf{G} [\Sigma_*]_{:, \partial_k(n)} & \mathbf{G} \Sigma_* \mathbf{G}^T + \Sigma_{ee} \end{bmatrix} \right) \quad (5-19)$$

In the above equations:  $[\mu_*]_{\partial_k(n)} = \mathbb{E}_{p_*} [m_{\partial_k(n)}]$ ,  $[\Sigma_*]_{\partial_k(n), \partial_k(n)} := \text{COV}_{p_*} [m_{\partial_k(n)}, m_{\partial_k(n)}]$ , and  $[\Sigma_*]_{\partial_k(n), :} := \text{COV}_{p_*} [m_{\partial_k(n)}, m] = [\Sigma_*]_{:, \partial_k(n)}^T$

Third, using equations B-13 and B-16, one derives the conditional probability:

$$p_*(s \mid m_{\partial_k(n)}) = \mathcal{N} \left( s \mid \mu_{s|m_{\partial_k(n)}}, \Sigma_{s|m_{\partial_k(n)}} \right) \quad (5-20)$$

$$\mu_{s|m_{\partial_k(n)}} = \mathbf{G} [\Sigma_*]_{:, \partial_k(n)} [\Sigma_*]_{\partial_k(n), \partial_k(n)}^{-1} (m_{\partial_k(n)} - [\mu_*]_{\partial_k(n)}) \quad (5-21)$$

$$\Sigma_{s|m_{\partial_k(n)}} = \mathbf{G} \Sigma_* \mathbf{G}^T + \Sigma_{ee} - \mathbf{G} [\Sigma_*]_{:, \partial_k(n)} [\Sigma_*]_{\partial_k(n), \partial_k(n)}^{-1} [\Sigma_*]_{\partial_k(n), :} \mathbf{G}^T \quad (5-22)$$

To simplify the notation, we will denote  $\mathcal{G} = \mathbf{G} [\Sigma_*]_{:, \partial_k(n)} [\Sigma_*]_{\partial_k(n), \partial_k(n)}^{-1}$ .

Finally, in order to compute the integral in equation 5-17, one needs the conditional distribution of elastic properties given the *facies* pattern:

$$p(m_{\partial_k(n)} \mid \pi_{\partial_k(n)}) = \mathcal{N} \left( m_{\partial_k(n)} \mid \mu_{\pi_{\partial_k(n)}}, \Sigma_{\pi_{\partial_k(n)}} \right) \quad (5-23)$$

Where,  $\mu_{\pi_{\partial_k(n)}}$ , and  $\Sigma_{\pi_{\partial_k(n)}}$  are defined the same way as in equations 3-11 and 3-12, respectively.

Conditional distribution  $p_*(s \mid m_{\partial_k(n)})$  works as  $p(Y|X)$  from section B.5 and the distribution  $p(m_{\partial_k(n)} \mid \pi_{\partial_k(n)})$  plays the role of  $p(X)$  in section B.5. The result of the integral 5-17 can be computed using equation B-29:

$$p_*^{(k)}(s \mid \pi_{\partial_k(n)}) = \mathcal{N} \left( s \mid \mu_{s|\pi_{\partial_k(n)}}, \Sigma_{s|\pi_{\partial_k(n)}} \right) \quad (5-24)$$

$$\mu_{s|\pi_{\partial_k(n)}} = \mathcal{G} (\mu_{\pi_{\partial_k(n)}} - [\mu_*]_{\partial_k(n)}) \quad (5-25)$$

$$\Sigma_{s|\pi_{\partial_k(n)}} = \mathcal{G} \Sigma_{\pi_{\partial_k(n)}} \mathcal{G}^T + \Sigma_{s|m_{\partial_k(n)}} \quad (5-26)$$

Equation 5-25 works as an approximate forward model that relates the mean impedances of the *facies* pattern  $\mu_{\pi_{\partial_k(n)}}$  with the corresponding synthetic seismic, taking into account the effect of the mean background  $[\mu_*]_{\partial_k(n)}$ . If  $\mu_{\pi_{\partial_k(n)}} = [\mu_*]_{\partial_k(n)}$ , then,  $\mu_{s|m_{\partial_k(n)}} = 0$ , meaning that if the mean impedances of the *facies* pattern is equal to the background impedances  $[\mu_*]_{\partial_k(n)}$ , then no seismic reflection is expected.

A difference between the present work and the original work (1) is to consider only a vicinity  $s_{\partial_{k'}(n)}$  in the likelihood, instead of the whole seismic trace  $s$ . In fact, given  $\pi_{\partial_k(n)}$ , not much can be said about the whole seismic trace. So, we can consider a parameter  $k'$ , such that the seismic samples  $s_{\sim\partial_{k'}(n)}$  are conditionally independent of  $\pi_{\partial_k(n)}$  given  $s_{\partial_{k'}(n)}$ :

$$\begin{aligned} p\left(s \mid \pi_{\partial_k(n)}\right) &= p\left(s_{\sim\partial_{k'}(n)} \mid s_{\partial_{k'}(n)}, \pi_{\partial_k(n)}\right) p\left(s_{\partial_{k'}(n)} \mid \pi_{\partial_k(n)}\right) \\ &\propto p\left(s_{\sim\partial_{k'}(n)} \mid s_{\partial_{k'}(n)}\right) p\left(s_{\partial_{k'}(n)} \mid \pi_{\partial_k(n)}\right) \\ &\propto p\left(s_{\partial_{k'}(n)} \mid \pi_{\partial_k(n)}\right) \end{aligned} \quad (5-27)$$

Figure 5.1 illustrates this point for patterns of size 11, and seismic kernel of size 31. The mean seismic predictions (equation 5-25) are the black lines, stacked for all possible patterns. The grey interval is the stacked 80% confidence interval for each prediction.

One can notice, that beyond the dashed lines, all predictions are zero. Thus, whatever measurement is found beyond this point, it would not improve our knowledge about the possible *facies* pattern.

The red dashed lines correspond to the seismic kernel size plus the pattern size, and would represent a more conservative range of influence  $k'$ . But in practice, the range of influence can be smaller.

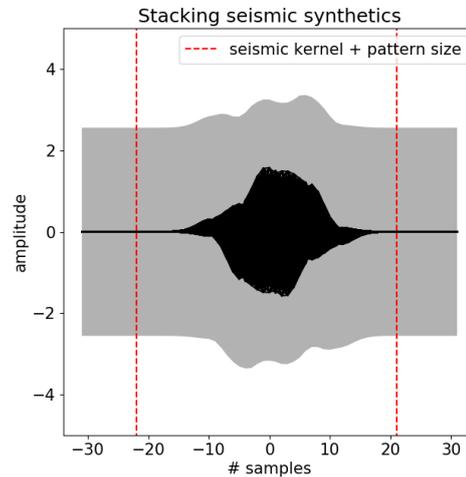


Figure 5.1: The seismic mean predictions, and 80% confidence intervals for all *facies* patterns are stacked in this figure. The red dashed lines represent a conservative measure of the influence range for patterns of size 11, and seismic kernel of size 31.

The conditional mean and covariance for  $s_{\partial_{k'}(n)}$  given  $\pi_{\partial_k(n)}$  are computed by marginalizing out  $s_{\sim\partial_{k'}(n)}$  in equation 5-24:

$$\mathbb{E} [s_{\partial_{k'}(n)} | \pi_{\partial_k(n)}] = [\mu_{s|\pi_{\partial_k(n)}}]_{\partial_{k'}(n)} \quad (5-28)$$

$$\text{COV} [s_{\partial_{k'}(n)} | \pi_{\partial_k(n)}] = [\Sigma_{s|\pi_{\partial_k(n)}}]_{\partial_{k'}(n), \partial_{k'}(n)} \quad (5-29)$$

The above approximation 5-27, allows one to compute and store each  $\mu_{s_{\partial_{k'}(n)} | \pi_{\partial_k(n)}}$ , and  $\Sigma_{s_{\partial_{k'}(n)} | \pi_{\partial_k(n)}}$ , for each  $\pi_{\partial_k(n)} \in \Omega^k$ , since, storing  $\mu_{s|\pi_{\partial_k(n)}}$  and  $\Sigma_{s|\pi_{\partial_k(n)}}$  for each possible  $\pi_{\partial_k(n)}$  would require much more memory. So it saves computation time during inversion, at the expense of some pre-processing time, and memory storage.

Figures 5.2(a) and 5.2(b) illustrate the approximate forward model. They show two proposed *facies* patterns, and their resulting mean seismic prediction  $\mu_{s_{\partial_{k'}(n)} | \pi_{\partial_k(n)}}$ , plus an 80% confidence interval, computed from the conditional covariance matrix  $\Sigma_{s_{\partial_{k'}(n)} | \pi_{\partial_k(n)}}$ .

The pattern in figure 5.2(b) shows a better match to the measured seismic, thus higher likelihood. The pattern in this figure is more similar to the true *facies* sequence than the other pattern, but higher likelihood does not in general translate in proximity to the true *facies* because the forward model is not injective.

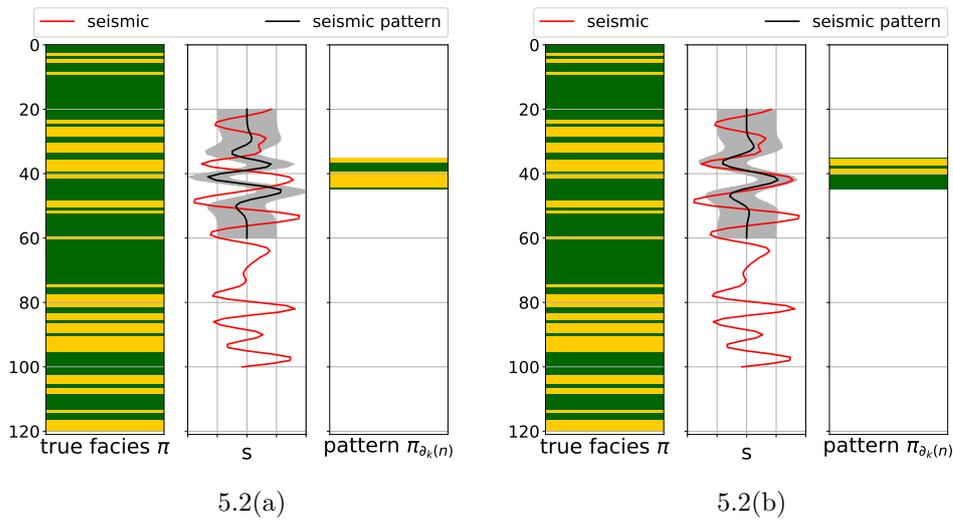


Figure 5.2: The figures illustrate the approximate seismic forward model for two possible patterns. The output of pattern in the right shows higher similarity with the seismic profile, so it will tend to be associated with higher posterior probability. Not coincidentally the pattern on the right has higher similarity with the true *facies* sequence, although this is not a rule that higher likelihoods come from solutions closer to the true geology.

## 5.4

### Complete Likelihood

Last section estimated an approximated conditional distribution  $p_*^{(k)}(s_{\partial_{k'}(n)} \mid \pi_{\partial_k(n)})$  for the seismic given a *facies* pattern centred at position  $n$ . Now, it will be shown how to relate it to the full approximated likelihood  $p_*^{(k)}(s \mid \pi)$ .

The complete likelihood is computed in (1) in a Naïve Bayesian assumption, i.e., the full likelihood  $p_*^{(k)}(s \mid \pi)$  is computed as the product of independent factors  $p_*^{(k)}(s \mid \pi_{\partial_k(n)})$ . It works as if each *facies* pattern  $\pi_{\partial_k(n)}$  carries a bit of independent information about the total seismic trace.

(1) uses an exponent correction, to compensate the number of times each variable  $\pi_n$  is used (each one appears in  $k$  factors in the following equation).

Finally, in the original work (1) (equation 3.30), there are border factors to account for the fact that the *facies* at the border are used less than  $k$  times in equation 5-30. In the present thesis, no boundary terms are used. On the other hand, the *facies* trace is bigger than the seismic trace ( $l_s$  samples in the beginning and  $l_s$  samples in the end, where  $k_s = 2l_s + 1$  is the seismic effective kernel size). With this strategy, the *facies* at the border of the seismic profile are accounted for  $k$  times, and the border distortion is pushed away from the region of interest. The concept is illustrated in 5.3.

From the discussion above and approximation 5-27, the complete likelihood in the present thesis can be estimated by the following expression:

$$p_*^{(k)}(s \mid \pi) = \prod_{n=1}^N p_*^{(k)}(s_{\partial_{k'}(n)} \mid \pi_{\partial_k(n)})^{\frac{1}{k}} \quad (5-30)$$

One important topic to be stressed is that in equation 5-30, the length of the measurement  $s$  is  $N - 2l_s$ , where  $k_s = 2l_s + 1$  is the size of the seismic effective kernel, while the length of the unknown *facies* sequence  $\pi$  is  $N + 2k$  (where the additional  $2k$  variables are marginalized in the result).

The challenge with this formulation is how to compute the likelihood for *facies* patterns away from the seismic top and bottom. The likelihood at these positions is computed taking into account the available seismic samples, and marginalizing out the missing seismic values (as explained in section B.4), like illustrated in the figure 5.4.

This causes *facies* patterns further away from the seismic borders to be less conditioned by seismic (less informative likelihoods). This can be noted in figure 5.3, as the outer *facies* probabilities converge to the prior distribution  $p_s = [0.75, 0.25]$ .

In chapter, only the central  $N - 2l_s$  part of the inversion will be analysed, for the outer *facies* samples are less conditioned by seismic information.

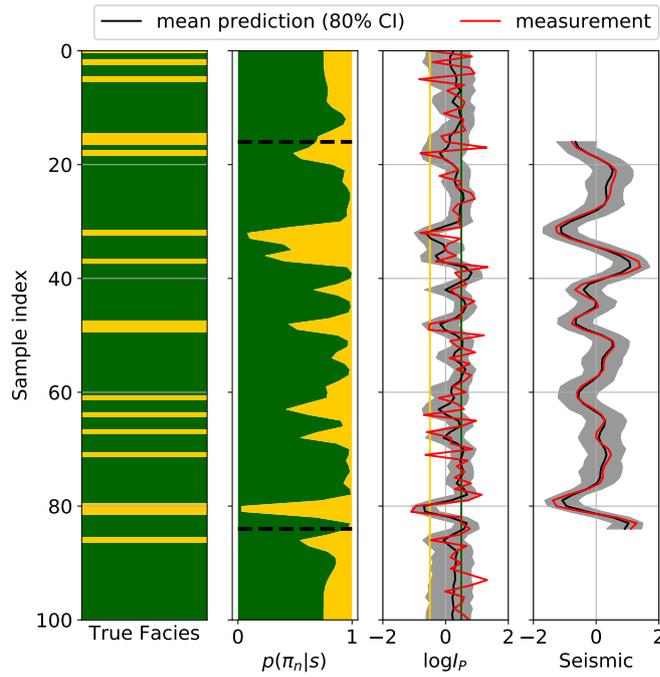


Figure 5.3: The figures shows this thesis implementation of the inversion, which uses additional *facies* samples in the inversion. CI stands for confidence interval.

PUC-Rio - Certificação Digital N° 1613077/CA

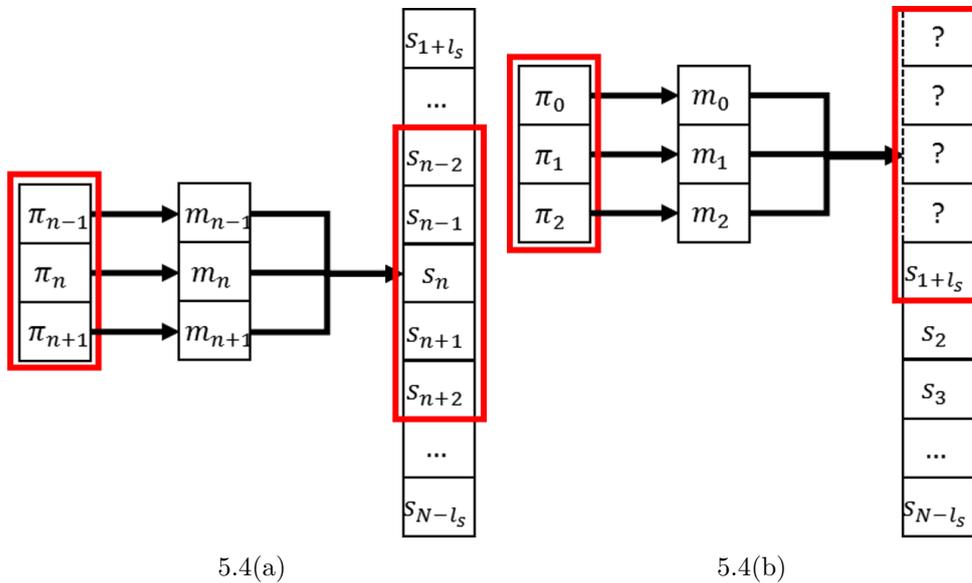


Figure 5.4: The figures illustrates the strategy used to compute the likelihood with missing data, for a pattern size of  $k = 3$ , and seismic kernel size of  $k' = 5$ . In the seismic central part 5.4(a), there is no missing samples, so the full conditional covariance and mean are used. At the borders 5.4(b), some seismic samples will be missing for comparison with the pattern synthetic, so just the corresponding indices in the conditional covariance matrix and mean, will be used.

## 5.5 Exact Inversion Assessment

There is a sampling technique, called Markov Chain Monte Carlo (MCMC) that will allow us to sample from the correct posterior probability distribution  $p(\pi | s)$ , given the approximation  $p_*^{(k)}(\pi | s)$ .

This thesis will use an MCMC strategy known as independent sampler Metropolis-Hastings.

A demonstration for MCMC method and related results is outlined in appendix E).

The MCMC technique is an iterative approach, that allows one to sample a population of sequences  $\pi$ , which will converge to a population drawn from the exact posterior distribution, as the number of sampled sequences grows. Theorem 5.5.1 below shows how the MCMC technique is applied to seismic to *facies* inversion.

**Theorem 5.5.1** (MCMC for *facies* inversion).  *$\pi$  is a random vector, corresponding to the hidden *facies* sequence. Given some measured seismic trace, the *facies* sequence has posterior probability distribution  $p(\pi | s) \propto p(\pi)p(s | \pi)$ . Given an approximate posterior distribution  $p_*^{(k)}(\pi | s)$ , the following algorithm asymptotically samples *facies* sequences from the desired distribution  $p(\pi | s)$ :*

**Data:** Approximate posterior  $p_*^{(k)}(\pi | s)$ , and exact posterior distribution  $p(\pi | s) \propto p(\pi)p(s | \pi)$

**Result:** Sampled population of sequences  $\pi^1, \pi^2, \dots, \pi^M$

- 1 Sample  $\pi^1$ , from  $p_*^{(k)}(\pi | s)$ ;
- 2 **for**  $t = 1$  **to**  $M - 1$  **do**
- 3     sample  $\pi'$  from  $p_*^{(k)}(\pi | s)$ , and sample  $u \sim \mathcal{U}[0, 1]$ ;
- 4     compute  $\alpha_{\pi^t, \pi'} = \min \left\{ \frac{p(\pi')p(s | \pi')p_*^{(k)}(\pi^t | s)}{p(\pi^t)p(s | \pi^t)p_*^{(k)}(\pi' | s)}, 1 \right\}$ ;
- 5     **if**  $u < \alpha_{\pi^t, \pi'}$  **then**
- 6          $\pi^{t+1} = \pi'$ ;
- 7         **else**  $\pi^{t+1} = \pi^t$ ;
- 8     **end**
- 9 **end**

One can use the sampled *facies* sequences (according to the theorem E.2.3) to approximate functionals on  $\pi$ . For example, one might be interested in estimating the marginal probabilities (which is an integral over the indicator function) for a *facies* type  $i$ , at a position  $n$ .

$$p(\pi_n = i | s) = \sum_{\pi \in \Omega^N} \mathbb{1}_{\{\pi_n, i\}} p(\pi | s) = \frac{1}{M} \sum_{t=1}^M \mathbb{1}_{\{\pi_n^t, i\}} \quad (5-31)$$

Sadly, there is no algorithm to estimate the most probable sequence a posteriori, when using MCMC. One ought to use the marginal maximum a posteriori (MMAP) definition as an approximation:

$$\pi_n^{MMAP} = \arg \max_i p(\pi_n = i | s) \quad (5-32)$$

## 5.6

### Elastic Inversion Conditioned on *Facies*

In the present chapter, it has been discussed how to sample *facies* sequences given a seismic measurement. Now, we can use this knowledge to sample the elastic properties conditioned on both the underlying *facies* sequence and on the seismic measurement (as illustrated in figure 2.2:

$$\begin{aligned} p(m | \pi, s) &\propto p(m, s | \pi) \\ &= p(s | m) p(m | \pi) \end{aligned} \quad (5-33)$$

Equation 5-33 is a Bayesian Inversion expression (as in sections B.3, and D.5), with likelihood term  $p(s | m)$ , and prior probability term  $p(m | \pi)$ .

From equation 3-10,  $p(m | \pi) = \mathcal{N}(m | \mu_\pi, \Sigma_\pi)$ , where  $\mu_\pi$  and  $\Sigma_\pi$  are defined in equations 3-11, and 3-12, respectively.

From equation 3-8,  $p(s | m) = \mathcal{N}(s | \mathbf{G}m, \Sigma_{ee})$ . Thus, one can derive the joint distribution  $p(m, s | \pi)$ , analogously to equation 5-18:

$$p(m, s | \pi) = \mathcal{N} \left( \begin{bmatrix} m \\ s \end{bmatrix} \middle| \begin{bmatrix} \mu_\pi \\ \mathbf{G}\mu_\pi \end{bmatrix}; \begin{bmatrix} \Sigma_\pi & \Sigma_\pi \mathbf{G}^T \\ \mathbf{G}\Sigma_\pi & \mathbf{G}\Sigma_\pi \mathbf{G}^T + \Sigma_{ee} \end{bmatrix} \right) \quad (5-34)$$

The conditional probability  $p(m | \pi, s)$  may be computed by conditioning equation 5-34, on  $s$ :

$$p(m | \pi, s) = \mathcal{N} \left( m \mid \mu_{m|s,\pi}, \Sigma_{m|s,\pi} \right) \quad (5-35)$$

$$\mu_{m|s,\pi} = \mu_\pi + \Sigma_\pi \mathbf{G}^T \Sigma_{ss}^{-1} (s - \mathbf{G}\mu_\pi) \quad (5-36)$$

$$\Sigma_{m|s,\pi} = \Sigma_\pi - \Sigma_\pi \mathbf{G}^T \Sigma_{ss} \mathbf{G} \Sigma_\pi \quad (5-37)$$

$$\Sigma_{ss} = \mathbf{G}\Sigma_\pi \mathbf{G}^T + \Sigma_{ee} \quad (5-38)$$

Similar joint *facies*-elastic inversion is discussed in (41, 48, 7, 49).

## 6 Application on Synthetic Examples

### 6.1 Introduction

Chapter 5 has shown a class of approximations parametrized by pattern size  $k$ , used to make the seismic to *facies* inversion problem computationally feasible. A MCMC algorithm was presented in order to iteratively assess the exact posterior distribution .

In this chapter the inversion is tested for a variety of synthetic examples, with different parametrizations: high and low *facies*' elastic variability, high and low signal-to-noise ratio, wavelets with different frequency contents, and different transition matrices.

Different scores are used to evaluate the inversion performance, focusing on the classification error, the processing time, and the reliability of the probability estimates.

The pattern size  $k$  is a user defined parameter, and will be probed in order to understand its impact on the inversion results.

This chapter assumes that the seismic reflection profile comes from a single angle of incidence  $\theta$ . As explained in section D.5, this reduces the number of necessary elastic parameters to one (denoted by log-impedance), simplifying the analysis of the results.

### 6.2 Performance Measures

Measuring performance is a proper way to evaluate the characteristics of an algorithm and to allow reproducibility. The performance scores for classification problems available in (50) are used and described below.

Accuracy is the ratio of right predictions to the total number of samples. It measures how assertive is the inversion. Sometimes a classification algorithm may favour a predominant type of rock, at the expense of making big mistakes in other less common rocks. But, as the Oil and Gas Industry is interested in a rare *facies* (hydrocarbon bearing sand), the above bias should be detected and avoided. Hence, the following two scores will be used: precision and recall.

Precision is the ratio of right predictions of certain *facies* to the total of predictions of that *facies*. It measures how frequently one is right in its prediction regarding a specific *facies*. Recall is the ratio of right predictions of certain *facies* to the total amount of samples of that *facies*. It measures how probable is for a sample of a certain *facies* to be detected by the algorithm.

The ideal scenario would be a high precision and high recall for every *facies* of interest. Figure 6.1 summarizes the explained scores.

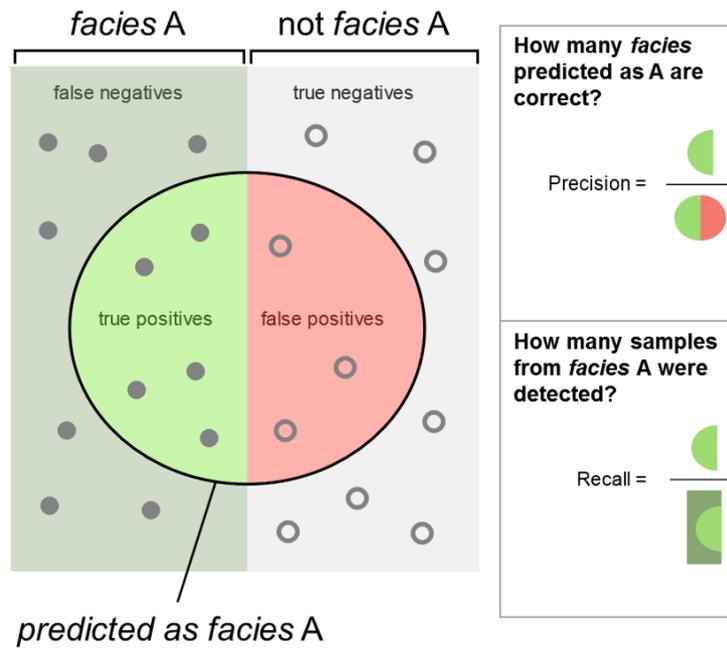


Figure 6.1: Illustration of precision and recall scores<sup>1</sup>.

As discussed in section D.5, seismic inversion is an ill-posed problem, and this is the reason of using probabilistic formalism in this thesis. So, there should be a score to measure the quality of the estimated probabilities.

For example, the algorithm will estimate for some samples 80% of chance to belong to *facies i*. This should truly mean that 80% of those samples are really from that *facies*.

Related to this concept, one can define the probability distortion score in equation 6-1.

$$distortion(i) = \sqrt{\frac{1}{N} \sum_{b=1}^B n_b (\bar{p}_b - f_{i,b})^2} \quad (6-1)$$

In equation 6-1,  $N$  is the number of samples in the dataset, the distortion is being computed for *facies i*. The  $[0, 1]$  segment has been uniformly divided in  $B$  bins.  $n_b$  is the number of samples with probability estimate in that bin,  $\bar{p}_b$

<sup>1</sup>Figure adapted from Walber - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=36926283>, in January 2018

is the mean probability estimate of the samples in bin  $b$ , and  $f_{i,b}$  is the fraction of samples in that bin which are from *facies*  $i$ .

Figure 6.2 exemplifies a graph of  $\bar{p}_b$  (mean estimated probabilities) versus  $f_{i,b}$  (fraction of true samples for *facies*  $i$ ). The graph was generated by a HMM classifier, on a synthetic example with 1001 samples. Although the theoretical distortion should be zero, because there are no approximations involved, there is a non-zero empirical distortion, due to statistical fluctuations.

In figure 6.2, the calibration graphs for *facies* 0 and *facies* 1 are redundant (they are antisymmetric). It only happens for binary problems: high probabilities of one *facies* implies low probabilities for the other, and correctly classifying one *facies* is equivalent to correctly not classifying the other.

For this reason, in the following sections, only the calibration curve for *facies* 1 will be shown.

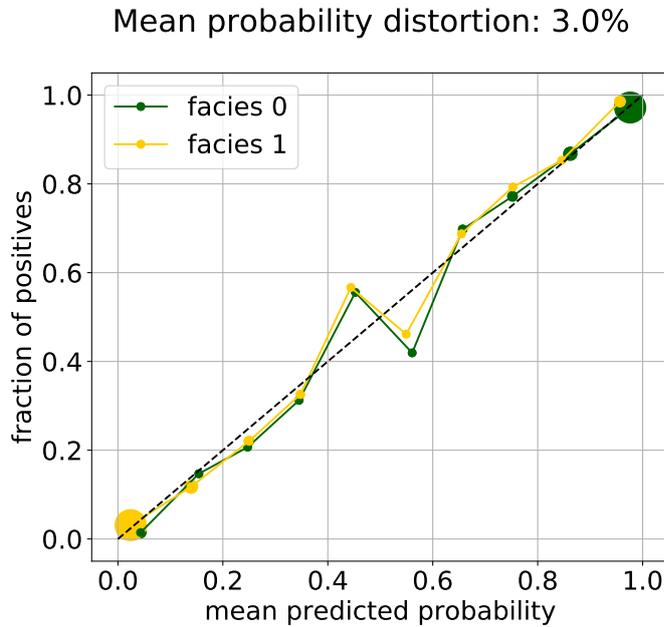


Figure 6.2: Probability calibration graph for a well calibrated classification algorithm. The distortion score is the RMSE between the fraction of positives and mean probability estimates

It is important to notice that an adequate choice for the number,  $B$ , of bins depends on the amount of available samples  $N$ , and on the dataset variability. Small amount of data will cause some bins to have too small number of examples,  $n_b$ , to estimate the probability distortion. In figure 6.2 the size of the circles is proportional to the amount of samples in that bin.

Finally, an algorithm is useful only if it can be computed in a reasonable amount of time. So the seismic to *facies* inversion algorithm will also be evaluated on this premise.

First, the algorithm needs a preprocessing step to store all the *facies* patterns of size  $k$  and the associated approximate responses. This preprocessing has the advantage to speed up the Forward-Backward algorithm, as it inquires the *facies* responses many times. On the other hand, it introduces an overhead time, which can be very long, depending on the number of *facies* types, and the pattern size  $k$ . Hence, this overhead time is used as another performance indicator.

Another time-consuming step in the inversion is the likelihood computation for the forward recursion, because one has to compare  $K^k$  seismic synthetic patterns with  $N$  measured seismic patterns, and this comparison involves computing the inverse of synthetics' covariance matrix 5-26. The time spent in the likelihood computation will be measured and discussed.

The time spent in the Forward recursion (algorithm 1) was insignificant when compared to the overhead and the likelihood computation time, so it was not taken into account.

The last performance measure is the average amount of time elapsed to sample each *facies* sequence from the posterior distribution (algorithm 3). The complexity of this step is  $O((N-k+1)K)$ , but as many posterior samplings are performed in the MCMC algorithm, this step is crucial to the algorithm overall processing time. This measure will be referred to as the time per sampling.

### 6.3 Dataset Generation

This chapter assumes that one has measured seismic from a single angle of incidence. In this situation, section D.5 shows that equation 3-7 can be simplified to:

$$s(\theta) = W(\theta)D \log(I_E) = W(\theta)Dm \quad (6-2)$$

In equation 6-2,  $I_E$  is known as elastic impedance (51). The elastic impedance is already enough, in the single angle case, to fully describe the elastic behaviour of the rock.

In the following the inversion will be studied for different parametrizations.

For each parametrization, 20 *facies* profiles were randomly sampled from the prior distribution and used to generate elastic properties and synthetic seismic. The 20 datasets were inverted and the scores were used to evaluate the algorithm performance. This evaluation routine is similar to the one used in (34).

The model parameters of interest are in the following list. The first two parameters refer to the geological a priori information and the others refer to seismic measurement conditions.

- transition matrix
- log-impedance standard deviation
- signal-to noise ratio (defined in B-23)
- seismic wavelet frequency content

The only fixed parameters, through all the examples, are the size of the chain (121 samples, unless stated otherwise), the number of classes (2), and their log-impedance mean values (0.5, for *facies* 0, and  $-0.5$  for *facies* 1).

The *facies* 0 represent shale (green in the figures), and *facies* 1 represents hydrocarbon bearing sand (yellow in the figures). So, computing  $p(\pi_n = 1 | s)$  actually is the same as computing the probability of success in finding hydrocarbon at position  $n$ .

The noise covariance matrix is defined by equation 6-3.

$$\Sigma_{ee} = c\mathbf{G}\mathbf{G}^T + 10^{-2}\mathbf{1} \quad (6-3)$$

In equation 6-3,  $\mathbf{G}$  refers to the seismic forward model in matrix form. The term  $c\mathbf{G}\mathbf{G}^T$  introduces a coloured noise in the same bandwidth as the seismic effective kernel (as illustrated in figures 6.18(d) - 6.18(f)). The term  $10^{-2}\mathbf{1}$  introduces a negligible level of white noise (uncorrelated noise), just to guarantee that  $\Sigma_{ee}$  is positive definite (since  $\mathbf{G}\mathbf{G}^T$  is positive semi-definite).

Equation B-23 defines the signal-to-noise ratio. In order to compute this ratio, one needs an estimate of  $\Sigma_{XX}$ , which comes from the approximate Gaussian covariance matrix  $\Sigma_*$  (equation 5-13).

So the signal-to-noise ratio will be computed in the following sections as:

$$\gamma = \frac{\text{tr}(\mathbf{G}\Sigma_*\mathbf{G}^T)}{\text{tr}(c\mathbf{G}\mathbf{G}^T + 10^{-2}\mathbf{1})} \approx \frac{\text{tr}(\mathbf{G}\Sigma_*\mathbf{G}^T)}{c \text{tr}(\mathbf{G}\mathbf{G}^T)} \quad (6-4)$$

From equation 6-4, one can estimate the coloured noise level  $c$  from the signal-to-noise ratio  $\gamma$  and vice-versa.

As will be explained further in the next sections, three scenarios were tested for each parameter: a base and two extreme values, corresponding to lowest and highest expected in practical situations (from the author's experience). The three scenarios will be compared based on the performance scores across the 20 sample inversions.

For each inversion, 5000 *facies* sequences will be sampled from the posterior probability distribution.

## 6.4

### Base Case

The base case consists of the following parameters

- transition matrix:

$$T = \begin{bmatrix} 0.83 & 0.17 \\ 0.5 & 0.5 \end{bmatrix} \quad p_s = \begin{bmatrix} 0.75 \\ 0.25 \end{bmatrix} \quad (6-5)$$

- log-impedance standard deviation:  $\sigma = 0.3$
- signal-to-noise ratio:  $\gamma = 4$
- wavelet: ricker with central frequency `0.08cycles/sample` (period `12.5samples/cycle`), and size 31 samples

First, the performance results will be shown as a function of the pattern size  $k$ . Parameter  $k'$  is always defined as  $k$  plus the length of the seismic effective kernel, as discussed in section 5.3.

Figure 6.3 shows the classification scores. There is no significant difference in accuracy, precision, and recall with increasing  $k$ . The similar results may be indicative that the number of MCMC iterations is enough for convergence to the right posterior distributions, for all  $k$  values.

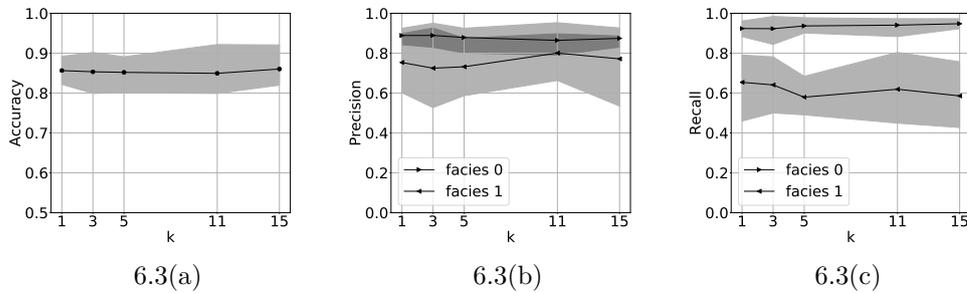


Figure 6.3: The graphs show, as a function of the lag parameter  $k$ , classification scores 6.3(a) accuracy, 6.3(b) precision, and 6.3(c) recall for the base case. There is no significant difference in accuracy, precision, and recall with increasing  $k$ .

From a computational time perspective, figure 6.4 shows an exponential increase in the overhead time and in the forward recursion time for increasing  $k$ , as expected. The time spent per sampling is approximately constant as expected, as algorithm 3 has complexity  $O(K)$ .

For  $k = 15$ , the inversion starts to become not feasible, as the forward recursion itself takes  $1000s \approx 16\text{min}$ . This performance means that the present inversion implementation is not applicable to process a seismic volume, as it

has around  $10^3$  up to  $10^6$  profiles and each profile would spend this amount of time just in the forward recursion.

Although not plotted, another issue with the projection approximation is the amount of computer memory needed to store all the possible  $k$ -lag *facies* vectors, and their associated seismic signatures, which increases exponentially as  $O(K^k)$ , thus limiting the pattern size  $k$ , and the number of possible *facies* types  $K$ .

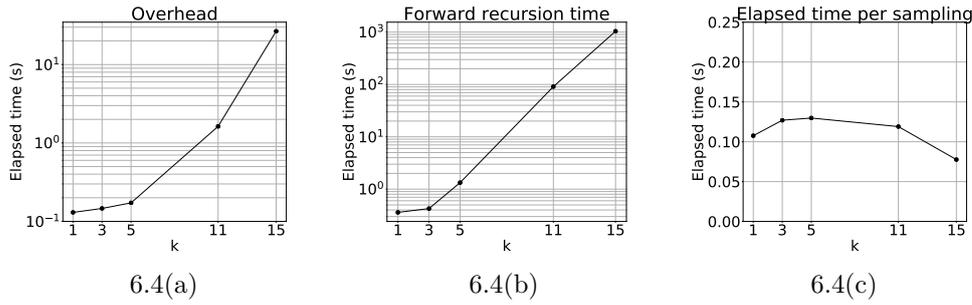


Figure 6.4: The graphs show the time spent 6.4(a) in the overhead processing, 6.4(b) in the forward recursion, and 6.4(c) per sampling, as a function of the lag parameter  $k$ . An exponential increase in overhead and forward recursion time can be noticed, and the elapsed time per sampling is approximately constant, as expected.

Figure 6.5 shows the acceptance rate as a function of pattern size  $k$ . The increase in the acceptance rate with  $k$ , is a weak symptom that the approximate posterior distribution is becoming more similar to the exact posterior distribution. In the limiting case, if  $k$  equals to the whole *facies* profile, there would be no approximation at all.

For higher acceptance rates, it is expected a faster convergence to the exact posterior probability distribution with less iterations. So, there is a trade-off between faster convergence and bigger computational demands (figure 6.4), as  $k$  increases.

In the next sections, the pattern size  $k$  will be fixed at 11, as a compromise between acceptance rate and computation time.

Figure F.1 illustrates the inversion of a seismic trace, for  $k = 11$ .

In the next sections, the base case scenario will be compared to other parametrizations.

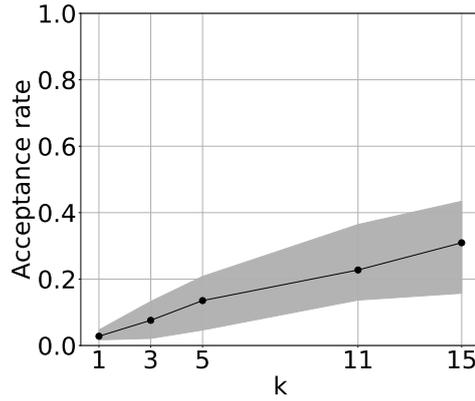


Figure 6.5: Mean acceptance rate as a function of lag parameter  $k$ .

## 6.5

### Transition Matrix

This section compares the inversion results for three different transition matrices.

- Extreme case 1 (equiprobable):

$$T = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \quad p_s = \begin{bmatrix} 0.50 \\ 0.50 \end{bmatrix} \quad (6-6)$$

- Base case (mildly unbalanced):

$$T = \begin{bmatrix} 0.83 & 0.17 \\ 0.5 & 0.5 \end{bmatrix} \quad p_s = \begin{bmatrix} 0.75 \\ 0.25 \end{bmatrix} \quad (6-7)$$

- Extreme case 2 (highly unbalanced):

$$T = \begin{bmatrix} 0.95 & 0.05 \\ 0.67 & 0.33 \end{bmatrix} \quad p_s = \begin{bmatrix} 0.93 \\ 0.07 \end{bmatrix} \quad (6-8)$$

Ten sequences sampled from the above transition matrices are illustrated in figure 6.6 for comparison. It can be seen a decrease in fraction of *facies* 1 from 6.6(a) to 6.6(c)

For each transition matrix, 20 datasets were sampled, and their associated modelled seismic were inverted. The inversion results are shown in figures 6.7 and 6.8.

Figure 6.7 shows that, as the transition matrix becomes more biased towards *facies* 0, the accuracy increases. But this happens at the expense of decreasing the precision and the recall of *facies* 1. Intuitively, the more dominant *facies* 0 becomes, the lower the cost of mistaking *facies* 1 with 0.

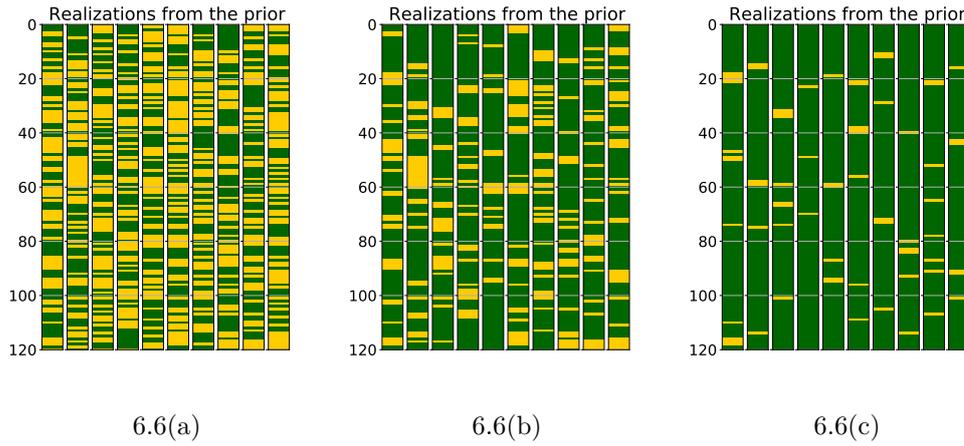


Figure 6.6: The graphs illustrate *facies* profiles sampled from the prior probability distribution, for the three transition matrix scenarios: 6.6(a), equiprobable, 6.6(b) mildly unbalanced, and 6.6(c) highly unbalanced.

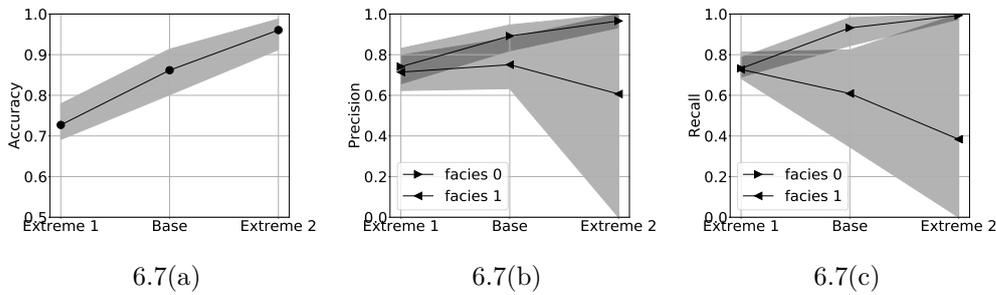


Figure 6.7: The graphs compare the classification scores 6.7(a) accuracy, 6.7(b) precision, and 6.7(c) recall between the three transition matrix scenarios.

The high accuracy values do not reflect a good classification result, but rather just expresses the percentage of *facies* 0 in the profile (a constant prediction will already yield a high accuracy).

In order to validate the marginal probability estimates, probability calibration graphs (explained in section 6.2) are used. Figure 6.9 shows that all the parametrization are well calibrated.

Finally, figures F.2 and F.3 exemplify the inversion for the extreme cases.

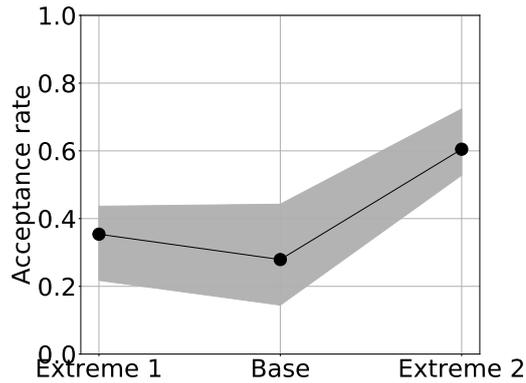


Figure 6.8: Mean acceptance rate as a function of transition matrix parametrization.

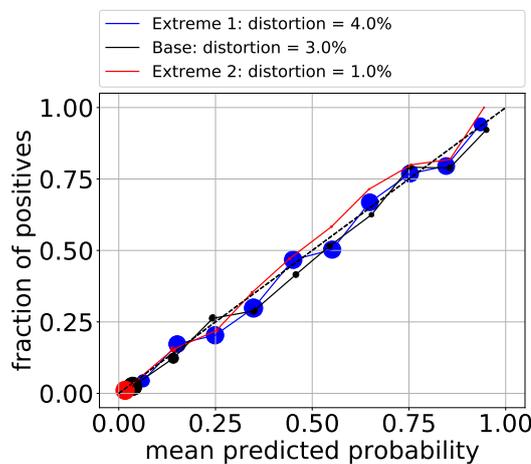


Figure 6.9: The graph show the experimental marginal posterior probability calibration for each of the three transition matrix scenarios. The size of the points represent the amount of samples in each bin.

## 6.6 Impedances Variability

The closer the *facies* are in terms of elastic properties, the harder it is for the inversion to distinguish them. Having the difference between the mean log-impedance of the two *facies* fixed at 1, the impedance standard deviation cases are:

- Extreme case 1:  $\sigma = 0.05$
- Base case:  $\sigma = 0.3$
- Extreme case 2:  $\sigma = 1$

In linear AVA theory, what matters in the seismic forward modelling is the contrast between the log-impedance of neighbouring layers as stated in equation D-7.

On the other hand, given the *facies* difference in mean elastic property, one needs the *facies* standard deviation in order to determine how distinguishable they are. The statistical quantity named the effect size, defined in equation 6-9 (borrowed from (52)) quantifies this concept.

$$\delta_{ij} = \frac{|\mu_i - \mu_j|}{\sqrt{\sigma_i^2 + \sigma_j^2}} \quad (6-9)$$

In equation 6-9,  $\mu_i$  and  $\sigma_i$  are the mean log-impedance and standard deviation for *facies*  $i$ . The synthetic examples studied in the present chapter can be compared to real datasets with similar effect sizes.

- Extreme case 1 (completely separable *facies*):  $\delta = 20.0$
- Base case (reasonable *facies* separation):  $\delta = 3.3$
- Extreme case 2 (almost indistinguishable *facies*):  $\delta = 1.0$

Figure 6.10 shows log-impedance profiles sampled from the prior distribution, for the three cases.

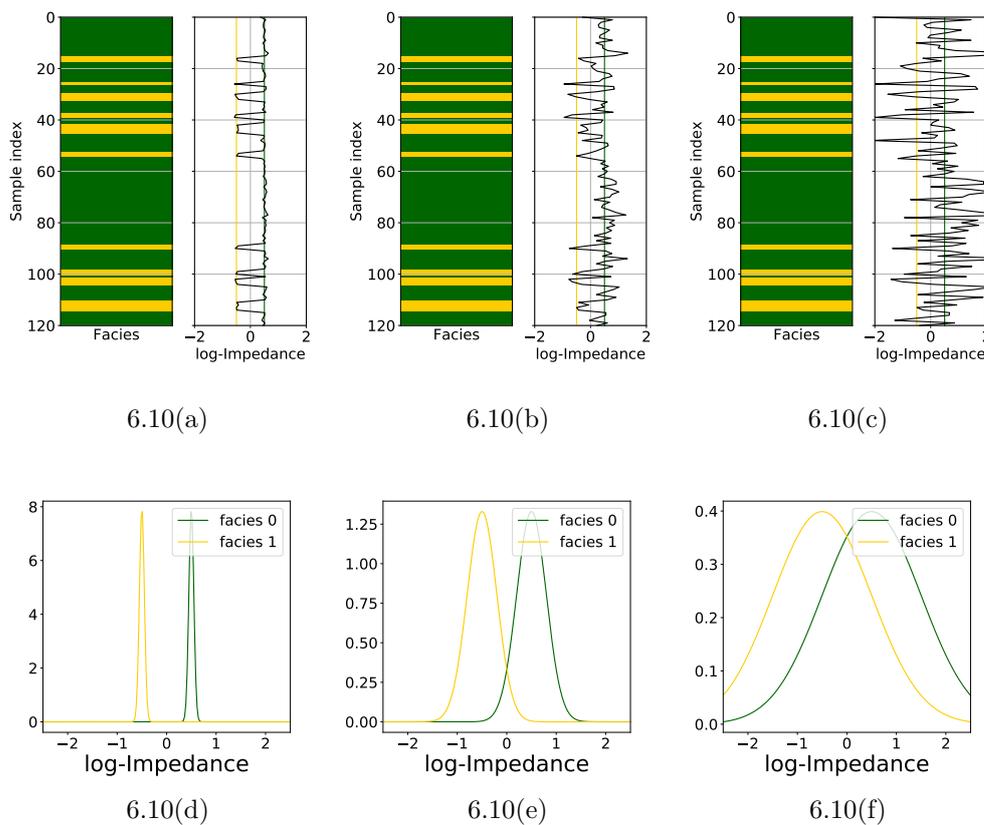


Figure 6.10: The graphs illustrate log-impedance profiles sampled from the prior probability distribution, and the corresponding theoretical histograms for the three impedance variability scenarios: 6.10(a) and 6.10(d) small variability, 6.10(b) and 6.10(e) mild variability, and 6.10(c) and 6.10(f) big variability.

Figure 6.11 compares the inversion results for the three cases in 20 datasets.

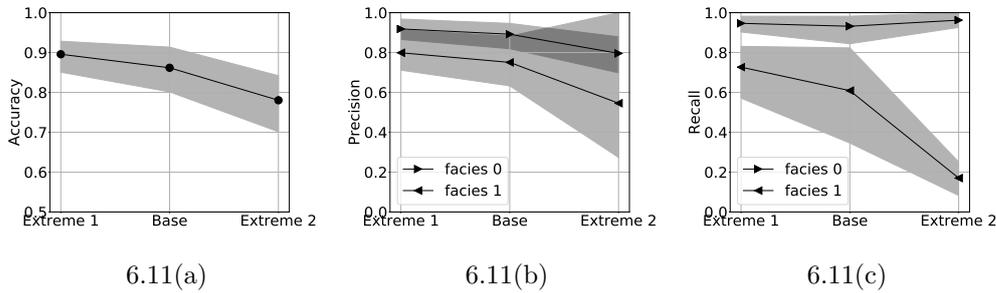


Figure 6.11: The graphs compare the classification scores 6.11(a) accuracy, 6.11(b) precision, and 6.11(c) recall between the three impedance variability scenarios.

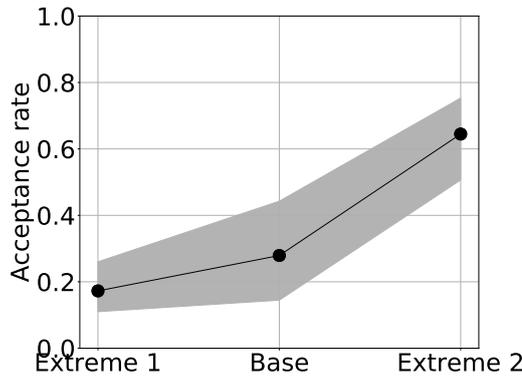


Figure 6.12: Mean acceptance rate as a function of impedance variability parametrization.

The first thing to notice is the degradation of classification scores as the impedance variability increases. As the effect size decreases, the algorithm starts giving more weight to the prior information  $p(\pi)$ , than to the seismic likelihood  $p(s|\pi)$ . As a consequence, the mean precision and recall for *facies* 1 decreases from extreme case 1 to extreme case 2.

Figure 6.12 illustrates the acceptance rate, with higher rates for extreme case 2. This can be explained by the fact that the Gaussian approximation, presented in appendix C becomes accurate, when *facies* elastic distributions' are indistinguishable.

Figure 6.13 shows the probability calibration plots, with distortions of 3% for the three cases, which is low.

Finally, figures F.4 and F.5(c) illustrate the inversion of a single seismic trace using the extreme scenarios.

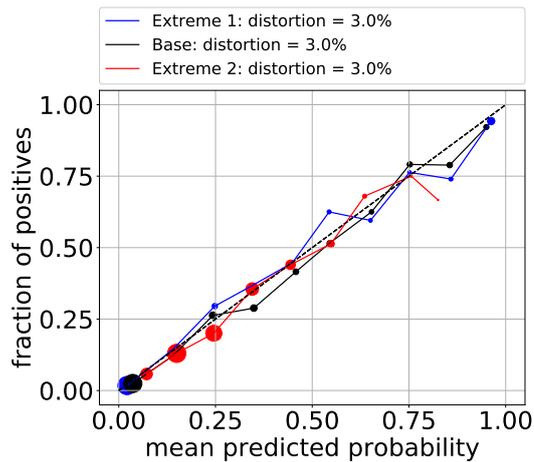


Figure 6.13: The graph show the experimental marginal posterior probability calibration for each of the three impedance variability scenarios. The size of the points represent the amount of samples in each bin.

## 6.7 Acquisition Error

The more noise in the seismic profile, the harder it is for the inversion to distinguish between the *facies*. This problem is even worse if the noise spectrum is similar to the seismic kernel spectrum, which is the present case, because the noise will be indistinguishable from signal. The signal-to-noise ratios studied in this section are:

- Extreme case 1 (almost noiseless seismic):  $\gamma = 20$
- Base case (typical noise level):  $\gamma = 4$
- Extreme case 2 (very noisy seismic):  $\gamma = 1$

Figure 6.14 shows seismic profiles sampled from the forward model 3-8, for the three parametrizations.

Figure 6.15 compares the classification scores between the study cases. There is a decrease in the classification scores as noise goes higher, as would be expected.

The acceptance rate in figure 6.20 increases with decreasing signal-to-noise ratio. It happens because, with increasing noise level, the seismic likelihood becomes less informative, so the difference between the approximate and exact distributions becomes less important.

Figure 6.17 show the probability calibration plots, without meaningful distortions, as in previous sections.

Finally, figures F.6 and F.7(c) show the inversion results for both extreme parametrizations.

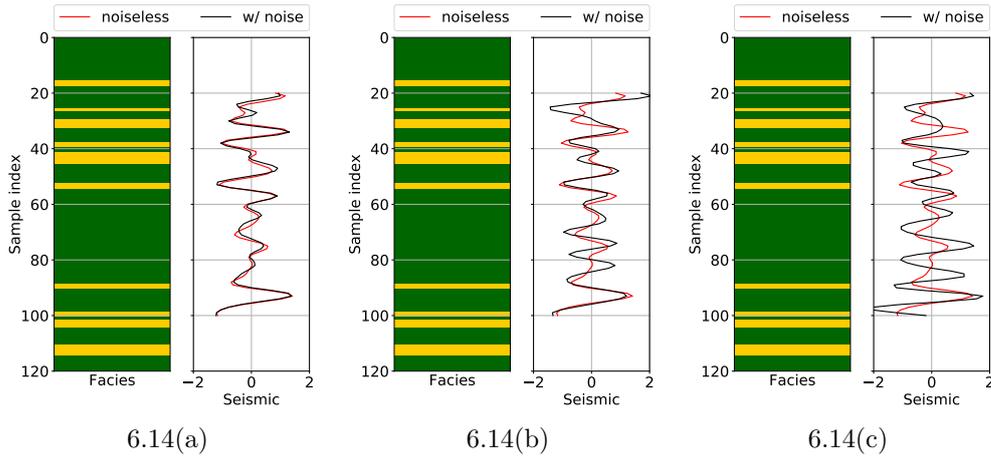


Figure 6.14: The graphs illustrate the *facies* profiles sampled from the prior probability distribution and the modelled seismic, for the three signal-to-noise scenarios: 6.14(a) low noise level, 6.14(b) mild noise level, and 6.14(c) high noise level.

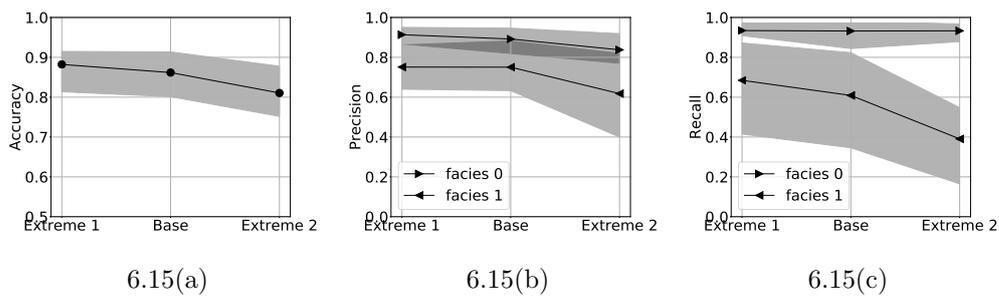


Figure 6.15: The graphs compare the classification scores 6.15(a) accuracy, 6.15(b) precision, and 6.15(c) recall between the three acquisition error scenarios.

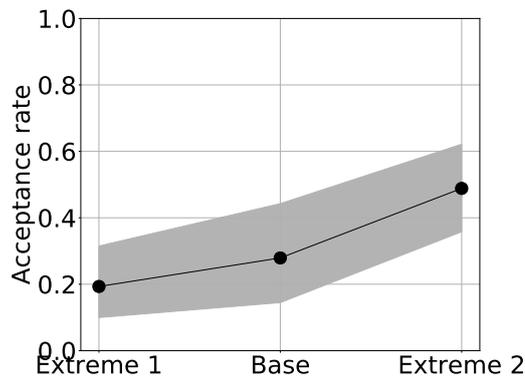


Figure 6.16: Mean acceptance rate as a function of acquisition error parametrization.

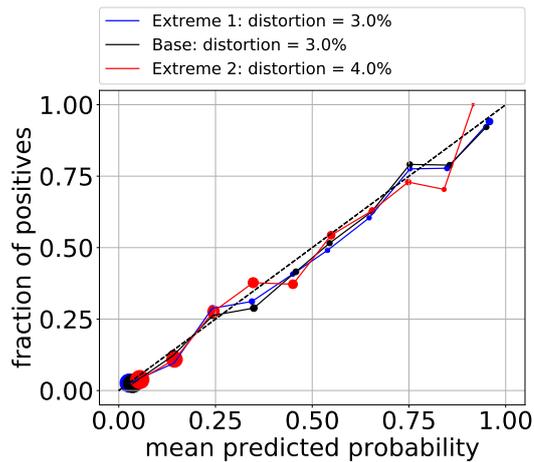


Figure 6.17: The graph shows the experimental marginal posterior probability calibration for each of the three transition matrix scenarios. The size of the points represent the amount of samples in each bin.

## 6.8 Wavelet

From section D.5 it is known that the wavelet high-frequency content will limit the resolution of the inversion. Thus it is expected that wavelets with less resolution will not detect high frequency intercalation between *facies*.

In this section, we will investigate how the wavelet's high frequency content influence *facies* inversion. The possible wavelets are:

- Extreme case 1 (high resolution): Ricker with peak frequency 0.32cycle/samples (period 3.125samples/cycle), and size 21 samples
- Base case (medium resolution): Ricker with peak frequency 0.08cycle/samples (period 12.5samples/cycle), and size 43 samples
- Extreme case 2 (low resolution): Ricker with peak frequency 0.02cycle/samples (period 50.0samples/cycle), and size 113 samples

The above values of frequency are parametrized in terms of cycle/samples, since different seismic data may be measured at different sampling rates.

The following figure 6.18 illustrates the three wavelets (6.18(a) - 6.18(c)). It also illustrates (6.18(d) - 6.18(f)) the modelled synthetic seismic.

Figures 6.18(d) - 6.18(f) show the seismic with and without noise, in order to demonstrate how the modelled noise has the same frequency spectrum as the seismic effective kernel.

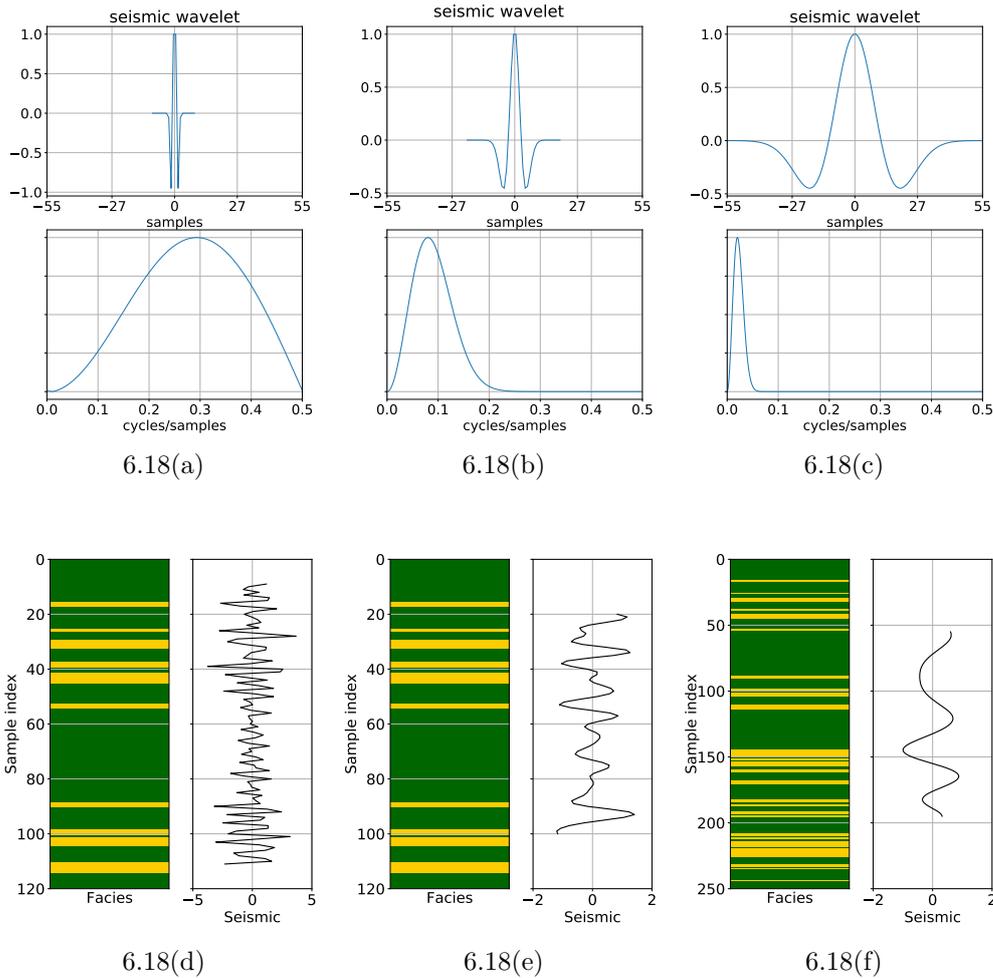


Figure 6.18: The graphs illustrate the seismic effective kernel (wavelet and derivative effects), and corresponding seismic for the three wavelet scenarios: 6.18(a)-6.18(d) short wavelet, 6.18(b)-6.18(e) medium wavelet, and 6.18(c)-6.18(f) long wavelet.

Figure 6.19 compares the scores of the study cases. As the wavelet peak frequency increases, precision and recall for *facies* 1 decreases.

Recall decreases dramatically, because layers of *facies* 1 has typically small thickness, which are not resolved by the wavelet, and thus are not detected. On the other hand, the inversion will predict *facies* 1 only when there is a strong negative reflection, which is typically related to a region with thick *facies* 1 layer (or set of layers), which explains that precision not decreasing so much as recall.

Extreme wavelet 1 has high frequency content, allowing it to discriminate even thin layers of *facies* 1, which explains the higher classification scores.

Figure 6.20 shows the acceptance rate. Figure 6.5 illustrates that in general as the pattern size  $k$  increases in relation to the seismic kernel length, the acceptance rate will increase. This explains the decrease in acceptance rate

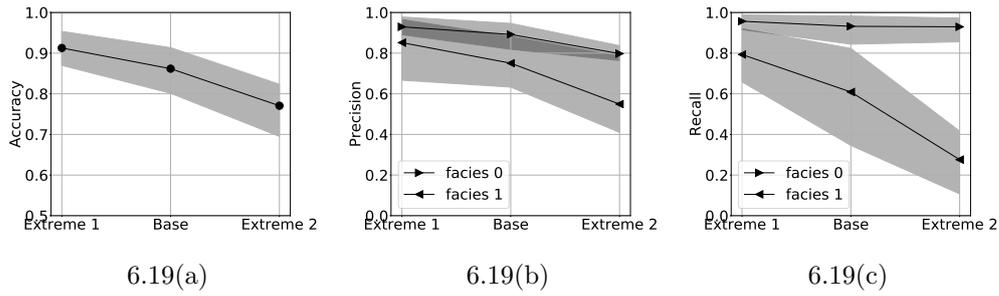


Figure 6.19: The graphs compare the classification scores 6.19(a) accuracy, 6.19(b) precision, and 6.19(c) recall between the three wavelet scenarios.

with the increase in seismic kernel in figure 6.20.

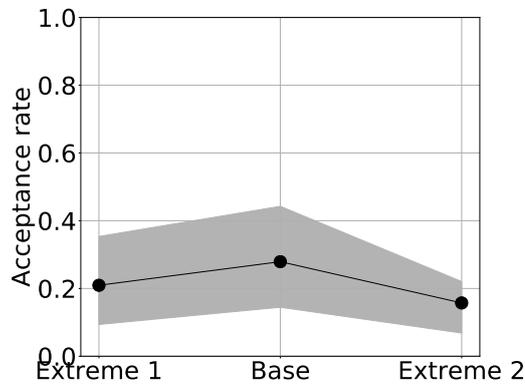


Figure 6.20: Mean acceptance rate as a function of the seismic wavelet.

Figure 6.21 shows the probability calibration plots, with small values of distortion.

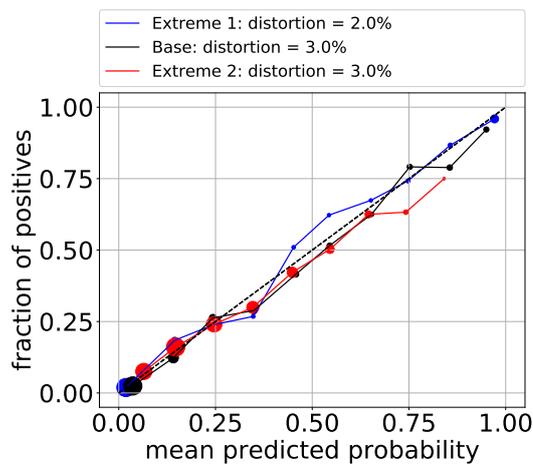


Figure 6.21: The graph shows the experimental marginal posterior probability calibration for each of the three transition matrix scenarios. The size of the points represent the amount of samples in each bin.

Finally, a concrete example for each of the extreme cases is given in figures F.8, and F.9.

## 7

### Conclusions

It has been seen that the Oil and Gas industry uses seismic reflection images in order to infer the spatial distribution of hydrocarbon bearing reservoirs in the subsurface. Owing to the fact that this problem is ill-posed, and the seismic data is noisy, probabilistic methodologies must be used in order to assess the chance of success before drilling.

The present work had the objective to quantify the chance of drilling success given a seismic image, taking into account the physical model that relates *facies* to seismic.

The problem of predicting the spatial distribution of *facies* in the subsurface, from the seismic reflection data, known as Convolved Hidden Markov Model, was demonstrated to be a complex combinatorial problem, due to the long range interaction between the categorical variables that represent *facies*. Other works in the literature have been analysed, all of them using either approximate or iterative methods to solve the problem.

In this thesis, a technique called projection approximation has been used to tackle the problem. It has been shown that this method is based on approximating the high dimensional complex probability distribution with a product of factors, each involving a reduced number of unknowns.

Synthetic experiments were set in order to analyse the technique performance in a wide range of geological and measurement conditions. In order to test the robustness of the approximation in quantifying the posterior probability of *facies* given seismic, some performance scores commonly used in the field of machine learning were used. A not so common score called probability distortion was introduced, which is useful to quality control the posterior probability estimates.

Additionally, the synthetic examples and their parametrization was designed in such a way that the results in this thesis can be compared to real datasets, regardless of absolute values of seismic amplitude, noise level, sampling rate, and absolute values of elastic properties.

Thus it can be stated that the projection approximation to the Convolved Hidden Markov Model yields reliable probability estimates, under a wide range of geological and acquisition conditions. It can be used by the Oil and

Gas industry to invert seismic traces at different positions and compare their probability of success.

## 7.1

### Future Works

The experiments in chapter 6 assess the inversion performance under a wide range of conditions, but under the hypothesis of known model parameters. For this reason, the inversion technique might not perform as well in real datasets as it performed in the synthetic examples in the present thesis. So, an interesting future topic of research is to understand the inversion sensibility to uncertainties in the model parameters.

The application of the developed technique in the inversion of real data is of utmost importance, with the same goal of further validating the robustness of the method.

As stated above, even though the projection approximation presented promising results in synthetic 1D seismic inversion, the present implementation has great limitations in terms of computation time. The time it took for a single seismic profile inversion makes it prohibitive to invert 3D seismic volumes of typical size (up to millions of profiles). This problem is worse in real datasets, where the number of *facies* of interest increases the computation effort in relation to the binary case studied in this work.

In order to deal with this limitation, a future research topic of interest is to apply high performance computing techniques to optimize the algorithm. Once, with a more efficient implementation, another step will be to generalize the method for 3D seismic application.

Another future work will be to generalize the projection approximation to deal with a more complex spatial correlation of elastic properties, since the Gaussian approximation in chapter C was demonstrated only for spatially uncorrelated properties.

A final future direction would be taking into account the non-linearities of AVA model in relation to the pattern properties  $m_{\partial_k(n)}$ , while linearising in relation to the background properties  $m_{\sim\partial_k(n)}$ .

## Bibliography

- [1] FJELDSTAD, T. Bayesian Inversion and Inference of Categorical Markov Models with Likelihood Functions Including Dependence and Convolution, June 2015.
- [2] BULAND, A.; KOLBJORNSEN, O.; HAUGE, R.; SKJAEVELAND, O.; DUFFAUT, K. Bayesian lithology and fluid prediction from seismic prestack data. *Geophysics*, v. 73, n. 3, p. C13–C21, 2008.
- [3] ROSA, A. L. R. *Análise do sinal sísmico*. Sociedade Brasileira de Geofísica, 2010.
- [4] SIMM, R. A ‘sense check’ method for incorporating seismic amplitude information into prospect risk (a tribute to mike bacon). *First Break*, v. 35, p. 45–49, 2017.
- [5] MUKERJI, T.; JORSTADZ, A.; AVSETH, P.; MAVKO, G.; GRANLI, J. R. Mapping Lithofacies and Pore-Fluid Probabilities in a North Sea Reservoir: Seismic Inversions and Statistical Rock Physics. *Geophysics*, v. 66, p. 988–1001, July 2001.
- [6] AVSETH, P.; MUKERJI, T.; MAVKO, G. *Quantitative Seismic Interpretation*. Paperback. ed. Cambridge University Press, 2010.
- [7] KEMPER, M.; GUNNING, J. Rock physics driven joint inversion to facies and reservoir properties, 2012.
- [8] LARSEN, A. L.; ULVMOEN, M.; OMRE, H.; BULAND, A. Bayesian lithology/fluid prediction and simulation on the basis of a markov-chain prior model. *Geophysics*, v. 71, n. 5, p. R69–R78, 2006.
- [9] CONNOLLY, P. A.; HUGHES, M. J. Stochastic inversion by matching to large numbers of pseudo-wells. *Geophysics*, v. 81, n. 2, p. M7–M22, 2016.
- [10] LINDBERG, D.; OMRE, H. Inference of the transition matrix in convolved hidden markov models and the generalized baum–welch algorithm. *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, v. 53, n. 12, p. 6443–6456, 2015.

- [11] KRUMBEIN, W. C.; DACEY, M. F. Markov chains and embedded markov chains in geology. *Mathematical Geology*, 1969.
- [12] DE GROOT, P. F. M.; BRIL, A. H.; FLORIST, F. J. T.; CAMPBELLT, A. E. Monte Carlo Simulation of Wells. *Geophysics*, v. 61, p. 631–638, May 1996.
- [13] BULAND, A.; OMRE, H. Bayesian linearized avo inversion. *Geophysics*, v. 68, n. 1, p. 185–198, 2003.
- [14] YENWONGFAI, H.; MONDOL, N.; LECOMTE, I.; FALEIDE, J. Prestack simultaneous inversion to predict lithology in the realgrunnen subgroup of the goliat field, sw barents sea, 2016.
- [15] ZHAO, L.; GENG, J.; HUA HAN, D.; CHENG, J.; GUO, T. Rock physics based probabilistic lithology and fluid prediction in a heterogeneous carbonate reservoir, 2013.
- [16] MAVKO, G.; MUKERJI, T.; DVORKIN, J. *The Rock Physics Handbook*. Second. ed. Cambridge University Press, 2009.
- [17] JOHANSEN, T.; JENSEN, E.; MAVKO, G.; DVORKIN, J. Inverse rock physics modeling for reservoir quality prediction. *Geophysics*, v. 78, n. 2, p. M1–M18, 2013.
- [18] GRANA, D. Bayesian Linearized Rock-Physics Inversion. *Geophysics*, v. 81, Nov. 2016.
- [19] SALTZER, R.; FINN, C. Predicting vshale and porosity using cascaded seismic and rock physics inversion, 2005.
- [20] BACHRACH, R. Joint estimation of porosity and saturation using stochastic rock-physics modeling. *Geophysics*, v. 71, Sept. 2006.
- [21] CHIAPPA, F.; MAZZOTTI, A. Petrophysical linearised inversion of ava data, 2004.
- [22] BOSCH, M.; CARA, L.; RODRIGUES, J.; NAVARRO, A.; DÍAZ, M. A monte carlo approach to the joint estimation of reservoir and elastic parameters from seismic amplitudes. *Geophysics*, v. 72, n. 6, p. O29–O39, 2007.
- [23] GRANA, D. *Bayesian Inversion Methods for Seismic Reservoir Characterization and Time-Lapse Studies*. Aug. 2013. Tese (Doutorado em Física) - Stanford University, Aug. 2013.
- [24] KEMPER, M.; GUNNING, J. Joint impedance and facies inversion – seismic inversion redefined. *First Break*, 2014.

- [25] FIGUEIREDO, L. P.; GRANA, D.; SANTOS, M.; FIGUEIREDO, W.; ROISENBERG, M.; NETO, G. S. Bayesian seismic inversion based on rock-physics prior modeling for the joint estimation of acoustic impedance, porosity and lithofacies. *Journal of Computational Physics*, 2017.
- [26] BISHOP, C. M. *Pattern recognition and machine learning*. Springer, 2006.
- [27] TARANTOLA, A. *Inverse problem theory and methods for parameter estimation*. SIAM, 2005.
- [28] EIDSVIK, J.; OMRE, H.; MUKERJI, T.; AVSETH, P. Seismic reservoir prediction using bayesian integration of rock physics and markov random field: a north sea example. *The Leading Edge*, p. 290–294, 2002.
- [29] HOUCK, R. Quantifying the uncertainty in an avo interpretation. *Geophysics*, v. 67, n. 1, p. 117–125, 2002.
- [30] KJONSBORG, H.; HAUGE, R.; KOLBJORNSEN, O.; BULAND, A. Integrating stochastic rock physics in seismic pre-drill prospect risk and reservoir quality assessment, 2009.
- [31] KJONSBORG, H.; HAUGE, R.; KOLBJORNSEN, O.; BULAND, A. Bayesian Monte Carlo Method for Seismic Predrill Prospect Assessment. *Geophysics*, v. 75, n. 2, 2010.
- [32] HAMMER, H.; KOLBJORNSEN, O.; TJELMELAND, H.; BULAND, A. Lithology and Fluid Prediction from Prestack Seismic Data using a Bayesian Model with Markov Process Prior. *Geophysical Prospecting*, 2011.
- [33] ULVMOEN, M.; OMRE, H. Improved resolution in bayesian lithology/fluid inversion from prestack seismic data and well observations: Part 1 – methodology. *Geophysics*, v. 75, n. 2, p. R21–R35, 2010.
- [34] ULVMOEN, M.; HAMMER, H. Bayesian lithology/fluid inversion—comparison of two algorithms. *Computational Geosciences*, v. 14, n. 2, p. 357–367, 2010.
- [35] ULVMOEN, M.; OMRE, H.; BULAND, A. Spatially coupled lithology/fluid inversion from real seismic data and well observations, 2009.
- [36] ULVMOEN, M.; OMRE, H.; BULAND, A. Improved resolution in bayesian lithology/fluid inversion from prestack seismic data and well observations: Part 2 — real case study. *Geophysics*, v. 75, n. 2, p. B73–B82, 2010.

- [37] RIMSTAD, K.; AVSETH, P.; OMRE, H. Bayesian lithology/fluid prediction constrained by spatial couplings and rock physics depth trends. *The Leading Edge*, 2010.
- [38] RIMSTAD, K.; OMRE, H. Approximate Posterior Distributions for Convolutional Two-Level Hidden Markov Models. *Computational Statistics and Data Analysis*, v. 53, p. 187–200, 2013.
- [39] LINDBERG, D.; OMRE, H. Blind Categorical Deconvolution in Two-Level Hidden Markov Models. *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, v. 52, n. 11, p. 7435–7447, 2014.
- [40] AYENI, G.; HUCK, A.; DE GROOT, P. Extending reservoir property prediction with pseudo-wells. *First Break*, 2008.
- [41] GUNNING, J.; KEMPER, M. Some newer algorithms in joint categorical and continuous inversion problems around seismic data, 2011.
- [42] NAEINI, E.; EXLEY, R. Quantitative interpretation using facies-based seismic inversion. *Interpretation*, v. 5, n. 3, p. SL1–SL8, 2017.
- [43] AKI, K.; RICHARDS, P. *Quantitative seismology: theory and methods*. 1st ed. W. H. Freeman, 1980.
- [44] BORTFELD, R. Approximations to the reflection and transmission coefficients of plane longitudinal and transverse waves. *Geophysical Prospecting*, 1961.
- [45] STOLT, R.; WEGLEIN, A. Migration and inversion of seismic data. *Geophysics*, 1985.
- [46] RABINER, L. R. tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of The IEEE*, v. 77, n. 2, p. 257–286, 1989.
- [47] REEVES, R.; PETTITT, A. Efficient recursions for general factorisable models. *Biometrika*, v. 91, n. 3, p. 751–757, 2004.
- [48] RIMSTAD, K.; AVSETH, P.; OMRE, H. Hierarchical bayesian lithology/fluid prediction: A north sea case study. *Geophysics*, v. 77, n. 2, p. B69–B85, 2012.
- [49] GRANA, D.; FJELDSTAD, T.; OMRE, H. Bayesian Gaussian Mixture Linear Inversion for Geophysical Inverse Problems. *Mathematical Geosciences*, v. 49, p. 493–515, 2017.

- [50] Scikit-learn: Machine learning in python. Accessed in October, 2017.
- [51] CONNOLLY, P. Elastic impedance. *The Leading Edge*, p. 438–452, 1999.
- [52] PATERNOSTER, R.; BRAME, R.; MAZEROLLE, P.; PIQUERO, A. Using the Correct Statistical Test for the Equality of Regression Coefficients. *Criminology*, v. 36, p. 859–866, 1998.
- [53] RHODES, C. Sampling from the Posterior Distribution for HMM Hidden States. Oct. 2008.
- [54] LU, T.-T.; SHIOU, S. H. Inverses of 2x2 block matrices. *Computers and Mathematics with Applications*, v. 43, p. 119–129, 2002.
- [55] ULRYCH, T. J.; SACCHI, M. D.; WOODBURY, A. A bayes tour of inversion: A tutorial. *Geophysics*, v. 66, n. 1, p. P55–P69, 2001.
- [56] ZOEPPRITZ, K. On the reflection and propagation of seismic waves at discontinuities. *Erdbebenwellen VII B*, p. 66–84, 1919.
- [57] ROSA, R. *Extraction of elastic parameters using seismic reflection amplitude with offset variation*. 1976. Dissertação (Mestrado em Física) - Houston University, 1976.
- [58] Scipy.org. Accessed in January, 2018.
- [59] HAMPSON, D.; RUSSELL, B.; BANKHEAD, B. Simultaneous inversion of pre-stack seismic data, 2005.
- [60] VELIS, D. R. Parametric sparse-spike deconvolution and the recovery of the acoustic impedance, 2006.
- [61] BALL, V.; TENORIO, L.; SCHIOTT, C.; BLANGY, J.; THOMAS, M. Uncertainty in inverted elastic properties resulting from uncertainty in the low-frequency model. *The Leading Edge*, p. 1028–1035, 2015.
- [62] ROSS, S. *Simulation*. Fourth. ed. Elsevier, 2006.
- [63] XIAOYU, X.; YUN, L.; ZHEN, Z.; DESHENG, S.; JIXIANG, L.; JIE, W.; HUIFENG, W. The application of a low-frequency model constrained by seismic velocity to acoustic impedance inversion, 2013.
- [64] SAMS, M.; CARTER, D. Stuck between a rock and a reflection: A tutorial on low-frequency models for seismic inversion. *Interpretation*, v. 5, n. 2, p. B17–B27, 2017.

- [65] PENDREL, J. Low frequency models for seismic inversions: strategies for success, 2015.
- [66] SAMS, M.; SAUSSUS, D. Practical implications of low frequency model selection on quantitative interpretation results, 2013.
- [67] JAMES, B. R. *Probabilidade: um curso em nível intermediário*. 3. ed. Instituto de Matemática Pura e Aplicada, 2004.
- [68] GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, D. J. *Markov Chain Monte Carlo in Practice*. Springer-Science +Business Media B. V., 1996.
- [69] SHONKWILER, R.; MENDIVIL, F. *Explorations in monte carlo methods*. Springer, 2009.

# A

## Hidden Markov Model

### A.1 Introduction

This chapter focuses on deriving the main results concerning Hidden Markov Models. One can find an overview about HMM in the seminal work (46). But the notations, and demonstrations from this appendix were adapted from (26).

### A.2 Easy Questions on HMM

From the definition of a HMM, given in chapter 4, it is easy to derive the following relations.

The first one is the probability distribution of happening a sequence of hidden variables, before taking any measurement. This is known as the prior probability:

$$\begin{aligned} p(z) &= p(z_N | z_1, \dots, z_{N-1}) \times p(z_1, \dots, z_{N-1}) \\ &= p(z_N | z_{N-1}) \times p(z_1, \dots, z_{N-1}) \\ &= \dots = p(z_N | z_{N-1}) \times \dots \times p(z_2 | z_1) \times p(z_1) \end{aligned} \tag{A-1}$$

To compute  $p(z_1)$ , we use the hypothesis that the Chain has translational invariance. Thus,  $p(z_1 = i) = p(z_n = i)$  regardless the position  $n$  in the sequence. We define  $p_s(i)$  the stationary distribution, so that  $p(z_1 = i) = p_s(i)$ .

To compute the stationary probability, one has to marginalize the joint probability  $p(z_n, z_{n-1})$  :

$$\begin{aligned} p_s(j) &= p(z_n = j) = \sum_{i=0}^{K-1} p(z_n = j, z_{n-1} = i) \\ &= \sum_{i=0}^{K-1} p(z_n = j | z_{n-1} = i) p(z_{n-1} = i) \\ &= \sum_{i=0}^{K-1} T_{ij} p_s(i) \end{aligned} \tag{A-2}$$

The above equation is an eigenvector equation (with eigenvalue equal to 1). Its solution exists and is unique from Pierre-Frobenius theorem E.2.1.

The second easy relation is the conditional probability of an observed sequence of measurements, given a sequence of latent variables. This conditional probability is also known as likelihood.

$$\begin{aligned} p(X|z) &= p(x_N|x_1, \dots, x_{N-1}, z) \times p(x_1, \dots, x_{N-1}, z) \\ &= p(x_N|z_N) \times p(x_{N-1}|x_1, \dots, x_{N-2}, z) \times p(x_1, \dots, x_{N-2}, z) \\ &= \dots = p(x_N|z_N) \times \dots \times p(x_1|z_1) \end{aligned} \quad (\text{A-3})$$

### A.3

#### Posterior Normalization Constant

The posterior probability distribution for the hidden states sequence, given a sequence of measurements is given by, from Bayes theorem (a good reference is (26)).

$$p(z|X) = \frac{p(X, z)}{p(X)} = \frac{p(X|z) \times p(z)}{p(X)} \quad (\text{A-4})$$

$$= \frac{p(z_1)p(x_1|z_1) \prod_{n \geq 2}^N p(x_n|z_n)p(z_n|z_{n-1})}{p(X)} \quad (\text{A-5})$$

Given an observable sequence  $X$ ,  $p(X)$  is a constant. It can be calculated by summing  $p(X, z)$  for all possible latent sequences  $z$ :

$$\begin{aligned} p(X) &= \sum_{z \in \Omega^N} p(X, z) = \sum_{z \in \Omega^N} p(X|z)p(z) \\ &= \sum_{z_1=0}^{K-1} \sum_{z_2=0}^{K-1} \dots \sum_{z_N=0}^{K-1} p(z_1)p(x_1|z_1) \prod_{n \geq 2}^N p(x_n|z_n)p(z_n|z_{n-1}) \end{aligned} \quad (\text{A-6})$$

The above summation involves  $O(K^N)$  operations, thus is unfeasible in normal applications. In order to efficiently calculate  $p(X)$ , one uses the fact that the joint probability  $p(X, z) = p(X|z)p(z)$  can be expressed as a product of factors, each involving only 2 variables:

$$p(X, z) = q_1(z_1) \prod_{n \geq 2} q_n(z_{n-1}, z_n) \quad (\text{A-7})$$

$$q_1(z_1) = p(x_1|z_1)p(z_1) \quad (\text{A-8})$$

$$q_n(z_{n-1}, z_n) = p(x_n|z_n)p(z_n|z_{n-1}), n \geq 2 \quad (\text{A-9})$$

From this fact, the summation in equation A-6, can be rearranged as:

$$p(X) = \sum_{z_N=0}^{K-1} \sum_{z_{N-1}=0}^{K-1} q_N(z_{N-1}, z_N) \times \dots \times \sum_{z_2=0}^{K-1} q_3(z_2, z_3) \times \sum_{z_1=0}^{K-1} q_2(z_1, z_2) q_1(z_1) \quad (\text{A-10})$$

Or in recursive form:

$$\alpha_1(z_1) = q_1(z_1) \quad (\text{A-11})$$

$$\alpha_n(z_n) = \sum_{z_{n-1}=0}^{K-1} q_n(z_{n-1}, z_n) \alpha_{n-1}(z_{n-1}) \quad (\text{A-12})$$

$$p(X) = \sum_{z_N=0}^{K-1} \alpha_N(z_N) \quad (\text{A-13})$$

The recursive equations take  $O(NK^2)$  operations, thus is feasible, since it is linear in the sequence size. By substituting the  $q$  factors, one obtain what is known as Forward Recursion:

$$\alpha_1(z_1) = p(x_1|z_1)p(z_1) \quad (\text{A-14})$$

$$\alpha_n(z_n) = p(x_n|z_n) \sum_{z_{n-1}=0}^{K-1} p(z_n|z_{n-1}) \alpha_{n-1}(z_{n-1}) \quad (\text{A-15})$$

$$p(X) = \sum_{z_N=0}^{K-1} \alpha_N(z_N) \quad (\text{A-16})$$

Conjecture A.3.1 gives the interpretation of the  $\alpha$  array.

**Conjecture A.3.1.**  $\alpha_n(z_n) = p(x_1, \dots, x_n, z_n)$

*Proof.* By finite induction. For  $n=1$

$$\alpha_1(z_1) = p(x_1|z_1)p(z_1) = p(x_1, z_1) \quad (\text{A-17})$$

Let the conjecture be true for index  $n - 1$ . Then, for  $n$ :

$$\begin{aligned}
\alpha_n(z_n) &= p(x_n|z_n) \sum_{z_{n-1}=0}^{K-1} p(z_n|z_{n-1})\alpha_{n-1}(z_{n-1}) \\
&= p(x_n|z_n) \sum_{z_{n-1}=0}^{K-1} p(z_n|z_{n-1})p(x_1, \dots, x_{n-1}, z_{n-1}) \\
&= p(x_n|z_n) \sum_{z_{n-1}=0}^{K-1} p(z_n|z_{n-1})p(z_{n-1}|x_1, \dots, x_{n-1})p(x_1, \dots, x_{n-1}) \\
&= p(x_n|z_n) \sum_{z_{n-1}=0}^{K-1} p(z_n, z_{n-1}|x_1, \dots, x_{n-1})p(x_1, \dots, x_{n-1}) \tag{A-18} \\
&= p(x_n|z_n)p(z_n|x_1, \dots, x_{n-1})p(x_1, \dots, x_{n-1}) \\
&= p(x_n|z_n)p(x_1, \dots, x_{n-1}, z_n) \\
&= p(x_n|z_n, x_1, \dots, x_{n-1})p(x_1, \dots, x_{n-1}, z_n) \\
&= p(x_n, z_n, x_1, \dots, x_{n-1}) \\
&= p(x_1, \dots, x_n, z_n)
\end{aligned}$$

■

#### A.4 Posterior Marginal Probabilities

The marginal probability for a given state variable  $z_n$  given the sequence of all measurements is given by  $p(z_n|X)$ . The equation below expresses the marginal probabilities in terms of  $\alpha_n(z_n)$ , and an additional factor:

$$\begin{aligned}
p(z_n|X) &= \frac{p(X, z_n)}{p(X)} = \frac{p(x_{n+1}, \dots, x_N|x_1, \dots, x_n, z_n)p(x_1, \dots, x_n, z_n)}{p(X)} \\
&= \frac{p(x_{n+1}, \dots, x_N|z_n)\alpha_n(z_n)}{p(X)} \tag{A-19}
\end{aligned}$$

If one defines:  $\beta_n(z_n) := p(x_{n+1}, \dots, x_N|z_n)$  for  $n \leq N - 1$ , and  $\beta_N := 1$ , the above equation can be written as:

$$p(z_n|X) = \frac{\alpha_n(z_n)\beta_n(z_n)}{p(X)} \tag{A-20}$$

Just like the Forward Recursion, one can derive the Backward Recursion, which computes all the  $\beta_n$  terms from  $N - 1$  down to 1.

$$\begin{aligned}
\beta_n(z_n) &= p(x_{n+1}, \dots, x_N | z_n) \\
&= \sum_{z_{n+1}=0}^{K-1} p(x_{n+1}, \dots, x_N, z_{n+1} | z_n) \\
&= \sum_{z_{n+1}=0}^{K-1} \frac{p(x_{n+1}, \dots, x_N, z_{n+1}, z_n)}{p(z_n)} \\
&= \sum_{z_{n+1}=0}^{K-1} \frac{p(x_{n+2}, \dots, x_N | x_{n+1}, z_{n+1}, z_n) p(x_{n+1}, z_{n+1}, z_n)}{p(z_n)} \\
&= \sum_{z_{n+1}=0}^{K-1} \frac{p(x_{n+2}, \dots, x_N | z_{n+1}) p(x_{n+1} | z_{n+1}, z_n) p(z_{n+1}, z_n)}{p(z_n)} \quad (\text{A-21}) \\
&= \sum_{z_{n+1}=0}^{K-1} \frac{p(x_{n+2}, \dots, x_N | z_{n+1}) p(x_{n+1} | z_{n+1}) p(z_{n+1}, z_n)}{p(z_n)} \\
&= \sum_{z_{n+1}=0}^{K-1} p(x_{n+2}, \dots, x_N | z_{n+1}) p(x_{n+1} | z_{n+1}) p(z_{n+1} | z_n) \\
&= \sum_{z_{n+1}=0}^{K-1} \beta_{n+1}(z_{n+1}) p(x_{n+1} | z_{n+1}) p(z_{n+1} | z_n)
\end{aligned}$$

## A.5 Sampling from the Posterior Distribution

The topic from this section was adapted from the unpublished work (53), and it can also be found in (1).

From the previous section, one can choose any position  $n$  in the sequence and sample the hidden state given all the measurements. But, if one wants to sample the entire hidden state sequence, it is needed a sampling schedule, where the next chosen position is sampled conditioned on all the measurements, and all the already sampled hidden values, in order to respect the stacking pattern of the hidden variables.

We will use a backward schedule, starting from the last sample  $N$ , and at each step sample the  $z_n$  given all the measurements and all  $z_m$  for  $m > n$ . This schedule can be written as the following factorization A-22.

$$p(z | X) = p(z_N | X) \times p(z_{N-1} | z_N, X) \times \dots \times p(z_1 | z_2, \dots, z_N, X) \quad (\text{A-22})$$

First, one needs the marginal probability for the last index  $z_N$ :

$$p(z_N | X) = \frac{\alpha_N(z_N) \beta_N(z_N)}{p(X)} \quad (\text{A-23})$$

From equation A-13, and the definition of  $\beta_N$ :

$$p(z_N|X) = \frac{\alpha_N(z_N)}{\sum_{z'_N=0}^{K-1} \alpha_N(z'_N)} \quad (\text{A-24})$$

Now, assume one has already sampled all hidden variables from  $n + 1$  to  $N$ . Then:

$$\begin{aligned} p(z_n|z_{n+1}, \dots, z_N, x_1, \dots, x_N) &= p(z_n|z_{n+1}, x_1, \dots, x_n) \\ &= \frac{p(z_n, z_{n+1}, x_1, \dots, x_n)}{p(z_{n+1}, x_1, \dots, x_n)} \\ &= \frac{p(z_n, z_{n+1}, x_1, \dots, x_n)}{\sum_{z'_n=0}^{K-1} p(z'_n, z_{n+1}, x_1, \dots, x_n)} \end{aligned} \quad (\text{A-25})$$

The above equation depends on terms like  $p(z_n, z_{n+1}, x_1, \dots, x_n)$ .

$$\begin{aligned} p(z_n, z_{n+1}, x_1, \dots, x_n) &= p(z_{n+1}|z_n, x_1, \dots, x_n)p(z_n, x_1, \dots, x_n) \\ &= p(z_{n+1}|z_n)p(z_n, x_1, \dots, x_n) \\ &= p(z_{n+1}|z_n)\alpha_n(z_n) \end{aligned} \quad (\text{A-26})$$

Thus,

$$p(z_n|z_{n+1}, \dots, z_N, X) = \frac{p(z_{n+1}|z_n)\alpha_n(z_n)}{\sum_{z'_n=0}^{K-1} p(z_{n+1}|z'_n)\alpha_n(z'_n)} \quad (\text{A-27})$$

## A.6 Most Probable Sequence a Posteriori

Given a measured sequence, one might be interested to know the most probable latent state sequence. This is known as maximum a posteriori sequence  $z_{MAP}$ .

The most probable sequence  $z_{MAP}$  is the solution of the following optimization problem:

$$z_{MAP} = \arg \max_z p(z|X) = \arg \max_z p(z, X) \quad (\text{A-28})$$

Since the cardinality of the set of possible solutions is  $K^N$ , it is infeasible to try all of them. Instead, we will once more use the factorization property of the HMM problem. The first step is to observe the following general property of the max operator:

**Fact A.6.1.** *Let  $f : A \times B \rightarrow \mathbb{R}$  be a real function from set  $A \times B$ . Then:*

$$\max_{a,b} f(a, b) = \max_b \max_a f(a, b) = \max_a \max_b f(a, b)$$

Also, let  $g(b) := \max_a f(a, b)$ , and let  $\hat{b} := \arg \max_b g(b)$ . Then,

$$\arg \max_{a,b} f(a, b) = (\arg \max_a f(a, \hat{b}), \hat{b})$$

To the above fact, we will add the following useful definitions:

**Definition A.6.1.**

$$\Gamma_n(z_n) := \arg \max_{z_1, \dots, z_{n-1}} p(x_1, \dots, x_n, z_1, \dots, z_n), \quad n \geq 2$$

**Definition A.6.2.**

$$\begin{aligned} \omega_n(z_n) &:= \max_{z_1, \dots, z_{n-1}} p(x_1, \dots, x_n, z_1, \dots, z_n) = p(x_1, \dots, x_n, \Gamma_n(z_n), z_n) \\ \omega_1(z_1) &:= p(x_1, z_1) = p(x_1|z_1)p(z_1) \end{aligned}$$

Now, analogously to the previous sections, we will derive a forward-backward type of algorithm.

Let's say that we know at step  $n$  the maximal a posteriori trajectory  $(\Gamma_n(z_n), z_n)$ , for every choice of  $z_n$ . By keeping track of the cost function  $\omega_n(z_n)$  for each maximal trajectory (one for each  $z_n$ ), we can update the maximal trajectories for the next step  $n+1$ : for each possible final state  $z_{n+1}$ , the maximal trajectory  $\Gamma_{n+1}(z_{n+1})$  will be one of the previous  $K$  maximal trajectories  $(\Gamma_n(z_n), z_n)$  plus the final transition from  $z_n$  to  $z_{n+1}$ . In order to decide which of the previous maximal trajectories, one must compute the cost for each one of them, with:

$$\begin{aligned} \omega_{n+1}(z_{n+1}) &= \max_{z_1, \dots, z_n} p(x_1, \dots, x_{n+1}, z_1, \dots, z_{n+1}) \\ &= \max_{z_1, \dots, z_n} \{p(x_{n+1}, z_{n+1}|x_1, \dots, x_n, z_1, \dots, z_n)p(x_1, \dots, x_n, z_1, \dots, z_n)\} \\ &= \max_{z_1, \dots, z_n} \{p(x_{n+1}, z_{n+1}|z_n)p(x_1, \dots, x_n, z_1, \dots, z_n)\} \\ &= \max_{z_1, \dots, z_n} \{p(x_{n+1}|z_{n+1}, z_n)p(z_{n+1}|z_n)p(x_1, \dots, x_n, z_1, \dots, z_n)\} \\ &= \max_{z_1, \dots, z_n} \{p(x_{n+1}|z_{n+1})p(z_{n+1}|z_n)p(x_1, \dots, x_n, z_1, \dots, z_n)\} \\ &= p(x_{n+1}|z_{n+1}) \max_{z_1, \dots, z_n} \{p(z_{n+1}|z_n)p(x_1, \dots, x_n, z_1, \dots, z_n)\} \\ &= p(x_{n+1}|z_{n+1}) \max_{z_n} \max_{z_1, \dots, z_{n-1}} \{p(z_{n+1}|z_n)p(x_1, \dots, x_n, z_1, \dots, z_n)\} \\ &= p(x_{n+1}|z_{n+1}) \max_{z_n} \left\{ p(z_{n+1}|z_n) \max_{z_1, \dots, z_{n-1}} \{p(x_1, \dots, x_n, z_1, \dots, z_n)\} \right\} \\ &= p(x_{n+1}|z_{n+1}) \max_{z_n} \{p(z_{n+1}|z_n)\omega_n(z_n)\} \end{aligned} \tag{A-29}$$

From the fact A.6.1, it is natural to define a maximal  $z_n$  in equation A-29 as a function of  $z_{n+1}$ :

**Definition A.6.3.**

$$\psi_n(z_{n+1}) = \arg \max_{z_n} \{p(z_{n+1}|z_n)\omega_n(z_n)\}, \quad 1 \leq n \leq N - 1$$

And, we can calculate the  $\Gamma_{n+1}(z_{n+1})$  vector, again using fact A.6.1:

$$\begin{aligned} \Gamma_{n+1}(z_{n+1}) &= \left( \arg \max_{z_1, \dots, z_{n-1}} p(x_1, \dots, x_n, z_1, \dots, z_{n-1}, \psi_n(z_{n+1})), \psi_n(z_{n+1}) \right) \\ &= (\Gamma_n(\psi_n(z_{n+1})), \psi_n(z_{n+1})) \end{aligned} \quad (\text{A-30})$$

For  $n = 2$ , we can compute the initial value for the  $\Gamma$  sequence:

$$\begin{aligned} \Gamma_2(z_2) &= \arg \max_{z_1} p(x_1, x_2, z_1, z_2) \\ &= \arg \max_{z_1} \{p(x_2|z_2)p(z_2|z_1)p(x_1, z_1)\} \\ &= \arg \max_{z_1} \{p(z_2|z_1)p(x_1, z_1)\} \\ &= \arg \max_{z_1} \{p(z_2|z_1)\omega_1(z_1)\} \\ &= \psi_1(z_2) \end{aligned} \quad (\text{A-31})$$

Finally, the maximum sequence probability is given by:

$$\begin{aligned} \max_{z_1, \dots, z_N} p(z_1, \dots, z_N | x_1, \dots, x_N) &= \frac{1}{p(X)} \max_{z_1, \dots, z_N} p(x_1, \dots, x_N, z_1, \dots, z_N) \\ &= \frac{1}{p(X)} \max_{z_N} \max_{z_1, \dots, z_{N-1}} p(x_1, \dots, x_N, z_1, \dots, z_N) \\ &= \frac{1}{p(X)} \max_{z_N} \omega_N(z_N) \end{aligned} \quad (\text{A-32})$$

And the last state corresponding to this maximum value is  $\hat{z}_N = \arg \max_{z_N} \omega_N(z_N)$ . Now, one just has to back substitute the maximum states, just like stated in fact A.6.1:

$$\begin{aligned}
z_{MAP} &= \arg \max_{z_1, \dots, z_N} p(z_1, \dots, z_N | x_1, \dots, x_N) \\
&= \arg \max_{z_1, \dots, z_N} p(x_1, \dots, x_N, z_1, \dots, z_N) \\
&= \left( \arg \max_{z_1, \dots, z_{N-1}} p(x_1, \dots, x_N, z_1, \dots, \hat{z}_N), \hat{z}_N \right) \\
&= (\Gamma_N(\hat{z}_N), \hat{z}_N)
\end{aligned} \tag{A-33}$$

Defining  $\hat{z}_n = \psi_n(\hat{z}_{n+1})$ , and using equation A-30, we write down the maximal sequence:

$$\begin{aligned}
z_{MAP} &= \arg \max_{z_1, \dots, z_N} p(z_1, \dots, z_N | x_1, \dots, x_N) = (\Gamma_N(\hat{z}_N), \hat{z}_N) \\
&= (\Gamma_{N-1}(\hat{z}_{N-1}), \hat{z}_{N-1}, \hat{z}_N) = \dots \\
&= (\Gamma_2(\hat{z}_2), \hat{z}_2, \dots, \hat{z}_{N-1}, \hat{z}_N) \\
&= (\hat{z}_1, \dots, \hat{z}_{N-1}, \hat{z}_N)
\end{aligned} \tag{A-34}$$

It's important to notice that each  $\hat{z}_n$  from the maximal sequence depends on  $\hat{z}_{n+1}$ , from the definition A.6.3.

Another important topic is that the maximal sequence is not given by the marginal maximum for each position  $n$  in the sequence. In other words:  $\hat{z}_n \neq \arg \max_{z_n} p(z_n | X)$  (in general). It happens because the sequence obtained this way does not take into account the inter-relations between neighbouring latent states.

Nonetheless, a useful definition is the marginal maximum a posteriori sequence:

$$z_{MMAP} := \left( \arg \max_{z_1} p(z_1 | X), \dots, \arg \max_{z_N} p(z_N | X) \right) \tag{A-35}$$

## B Gaussian Distribution

### B.1 Definitions

A multivariate Gaussian distribution is given by the following PDF:

**Definition B.1.1** (Multivariate Gaussian). *A random vector  $X \in \mathbb{R}^N$  is said to be a multivariate Gaussian, if its probability density function is given by:*

$$p(X = x) = \mathcal{N}(x | \mu; \Sigma) := \frac{1}{\sqrt{2\pi\Sigma}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\}$$

Where:

$$\begin{aligned} \mu &= \mathbb{E}[X] \\ \Sigma &= \mathbb{E}[(X - \mu)(X - \mu)^T] \end{aligned}$$

From the above equation,  $\Sigma$  is a positive-definite symmetric matrix (called covariance matrix), and its elements are given by:

$$\Sigma_{i,j} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]$$

### B.2 Affine Transformations

Say  $X$  is a multivariate Gaussian vector and  $Y$  is another random vector, defined by an affine transformation of  $X$  plus a Gaussian error (independent of  $X$ ).

$$Y = AX + b + E \tag{B-1}$$

$$p(E = \epsilon) = \mathcal{N}(\epsilon | 0; \Sigma_{ee}) \tag{B-2}$$

$$p(X = x) = \mathcal{N}(x | \mu_X; \Sigma_{XX}) \tag{B-3}$$

Another way of expressing the linear system, in terms of the variable  $Y$  is:

$$p(Y = y | X = x) = p(E = y - Ax - b) = \mathcal{N}(y - Ax - b | 0; \Sigma_{ee}) = \mathcal{N}(y | Ax + b; \Sigma_{ee}) \quad (\text{B-4})$$

Then,  $Y$  is also a multivariate Gaussian, its mean and covariance matrix given by:

$$\mu_Y = A\mu_X + b \quad (\text{B-5})$$

$$\begin{aligned} \Sigma_{YY} &= \mathbb{E} \left[ (Y - \mu_Y)(Y - \mu_Y)^T \right] \\ &= \mathbb{E} \left[ (AX + b + E - A\mu_X - b)(AX + b + E - A\mu_X - b)^T \right] \\ &= \mathbb{E} \left[ (AX + E - A\mu_X)(AX + E - A\mu_X)^T \right] \\ &= A\mathbb{E} \left[ (X - \mu_X)(X - \mu_X)^T \right] A^T + A\mathbb{E} \left[ (X - \mu_X)E^T \right] \\ &\quad + \mathbb{E} \left[ E(X - \mu_X)^T \right] A^T + \mathbb{E} \left[ EE^T \right] \\ &= A\Sigma_{XX}A^T + A\mathbb{E}[X - \mu_X]\mathbb{E}[E]^T + \mathbb{E}[E]\mathbb{E}[X - \mu_X]^T A^T + \Sigma_{ee} \\ &= A\Sigma_{XX}A^T + \Sigma_{ee} \end{aligned} \quad (\text{B-6})$$

The crosscovariance between variables  $X$  and  $Y$  is:

$$\begin{aligned} \Sigma_{XY} &= \mathbb{E} \left[ (X - \mu_X)(Y - \mu_Y)^T \right] \\ &= \mathbb{E} \left[ (X - \mu_X)(AX + b + E - A\mu_X - b)^T \right] \\ &= \mathbb{E} \left[ (X - \mu_X)(AX + E - A\mu_X)^T \right] \\ &= \mathbb{E} \left[ (X - \mu_X)(X - \mu_X)^T \right] A^T + \mathbb{E} \left[ (X - \mu_X)E^T \right] \\ &= \Sigma_{XX}A^T \end{aligned} \quad (\text{B-7})$$

By definition, the joint probability of the augmented vector  $(X, Y)$  is:

$$\begin{aligned} p(X = x, Y = y) &= \mathcal{N} \left( \begin{bmatrix} x \\ y \end{bmatrix} \middle| \mu; \Sigma \right) \\ &= \mathcal{N} \left( \begin{bmatrix} x \\ y \end{bmatrix} \middle| \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}; \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \right) \end{aligned} \quad (\text{B-8})$$

By substituting the block matrices in the above equation, with their values from equations B-7 and B-6, one obtains:

$$\begin{aligned}
p(X = x, Y = y) &= \mathcal{N} \left( \begin{bmatrix} x \\ y \end{bmatrix} \middle| \begin{bmatrix} \mu_X \\ A\mu_X + b \end{bmatrix}; \begin{bmatrix} \Sigma_{XX} & \Sigma_{XX}A^T \\ A\Sigma_{XX} & A\Sigma_{XX}A^T + \Sigma_{ee} \end{bmatrix} \right) \quad (\text{B-9})
\end{aligned}$$

### B.3 Bayesian Inversion

Now, consider a situation where one has a realization of  $Y$ , and want to use this information to learn something about  $X$ . We have to estimate  $p(X | Y)$ , also known as posterior distribution:

$$p(X = x | Y = y) = \frac{p(Y = y, X = x)}{p(Y = y)} = \frac{p(Y = y, X = x)}{\int_{x'} p(Y = y, X = x')} \quad (\text{B-10})$$

Thus, to find the posterior distribution, ones has to fix  $Y = y$  in the joint distribution  $p(X, Y)$ , and renormalize in relation to  $X$ .

By defining  $\Lambda = \Sigma^{-1}$  in equation B-8, one can write  $p(X, Y)$  as:

$$\begin{aligned}
p(X = x, Y = y) &\propto \exp \left\{ -\frac{1}{2} \begin{bmatrix} x - \mu_X & y - \mu_Y \end{bmatrix} \Lambda \begin{bmatrix} x - \mu_X \\ y - \mu_Y \end{bmatrix} \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \begin{bmatrix} \Delta x & \Delta y \end{bmatrix} \begin{bmatrix} \Lambda_{XX} & \Lambda_{XY} \\ \Lambda_{YX} & \Lambda_{YY} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[ \Delta x^T \Lambda_{XX} \Delta x - 2\Delta x^T \Lambda_{XY} \Delta y + \Delta y^T \Lambda_{YY} \Delta y \right] \right\} \quad (\text{B-11})
\end{aligned}$$

Where  $\Delta x = x - \mu_X$  and  $\Delta y = y - \mu_Y$ . One can "complete the squares" in relation to  $\Delta x$ , and treat  $\Delta y$  as a constant, in order to arrive at a quadratic form:

$$\begin{aligned}
p(X = x, Y = y) &\propto \exp \left\{ -\frac{1}{2} \left[ \Delta x^T \Lambda_{XX} \Delta x - 2\Delta x^T \Lambda_{XY} \Delta y + \Delta y^T \Lambda_{YY} \Delta y \right] \right\} \\
&\propto \exp \left\{ -\frac{1}{2} (\Delta x - \Lambda_{XX}^{-1} \Lambda_{XY} \Delta y)^T \Lambda_{XX} (\Delta x - \Lambda_{XX}^{-1} \Lambda_{XY} \Delta y) \right\} \quad (\text{B-12})
\end{aligned}$$

By inspection, one checks that the desired probability distribution is Gaussian.

From (54), and since  $\Lambda = \Sigma^{-1}$ , one can simplify equation B-12:

$$\Lambda_{XX}^{-1} = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} \quad (\text{B-13})$$

$$\Lambda_{XY} = \Lambda_{XX}\Sigma_{XY}\Sigma_{YY}^{-1} \quad (\text{B-14})$$

From the above result, one can rewrite the posterior probability in equation B-10:

$$p(X = x | Y = y) = \mathcal{N}(x | \mu_{X|Y}; \Sigma_{X|Y}) \quad (\text{B-15})$$

$$\mu_{X|Y} = \mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y - \mu_Y) \quad (\text{B-16})$$

$$\Sigma_{X|Y} = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} \quad (\text{B-17})$$

In order to gain some insight over the above equations, we will analyse them in the scalar case  $Y = aX + b + E$ :

$$\mu_{X|Y} = \mu_X + a \frac{\sigma_{XX}^2}{\sigma_{YY}^2} (y - a\mu_X - b) \quad (\text{B-18})$$

$$\sigma_{X|Y}^2 = \sigma_{XX}^2 \left( 1 - \frac{a^2 \sigma_{XX}^2}{\sigma_{YY}^2} \right) \quad (\text{B-19})$$

$$\sigma_{YY}^2 = a^2 \sigma_{XX}^2 + \sigma_{ee}^2 \quad (\text{B-20})$$

The term  $a^2 \sigma_{XX}^2$  is the expected variance in the data due to input variance, and  $\sigma_{ee}^2$  is the noise variance. One can define the signal-to-noise ratio as  $\gamma = \frac{a^2 \sigma_{XX}^2}{\sigma_{ee}^2}$ . The equations simplify to:

$$\mu_{X|Y} = \mu_X + \frac{\gamma}{\gamma + 1} \frac{y - a\mu_X - b}{a} \quad (\text{B-21})$$

$$\sigma_{X|Y} = \sigma_{XX} \sqrt{1 - \frac{\gamma}{\gamma + 1}} \quad (\text{B-22})$$

Figure B.1 illustrate the effect of the signal-to-noise on the conditional probability distribution, in the linear equation  $Y = X + E$ , given  $Y = 2$ . The prior mean is  $\mu_X = 1$ , and the prior standard deviation is  $\sigma_{XX} = 1$ .

Figure B.1 illustrates that in the limit of small signal-to-noise ratio, the conditional mean and standard deviation tend to their prior values:  $\mu_X = 1$ , and  $\sigma_{XX} = 1$ . In the opposite limit of high signal-to-noise, the conditional mean tends to the measurement value, and the conditional standard deviation tends to zero.

From the scalar case, one can understand that the prior knowledge about

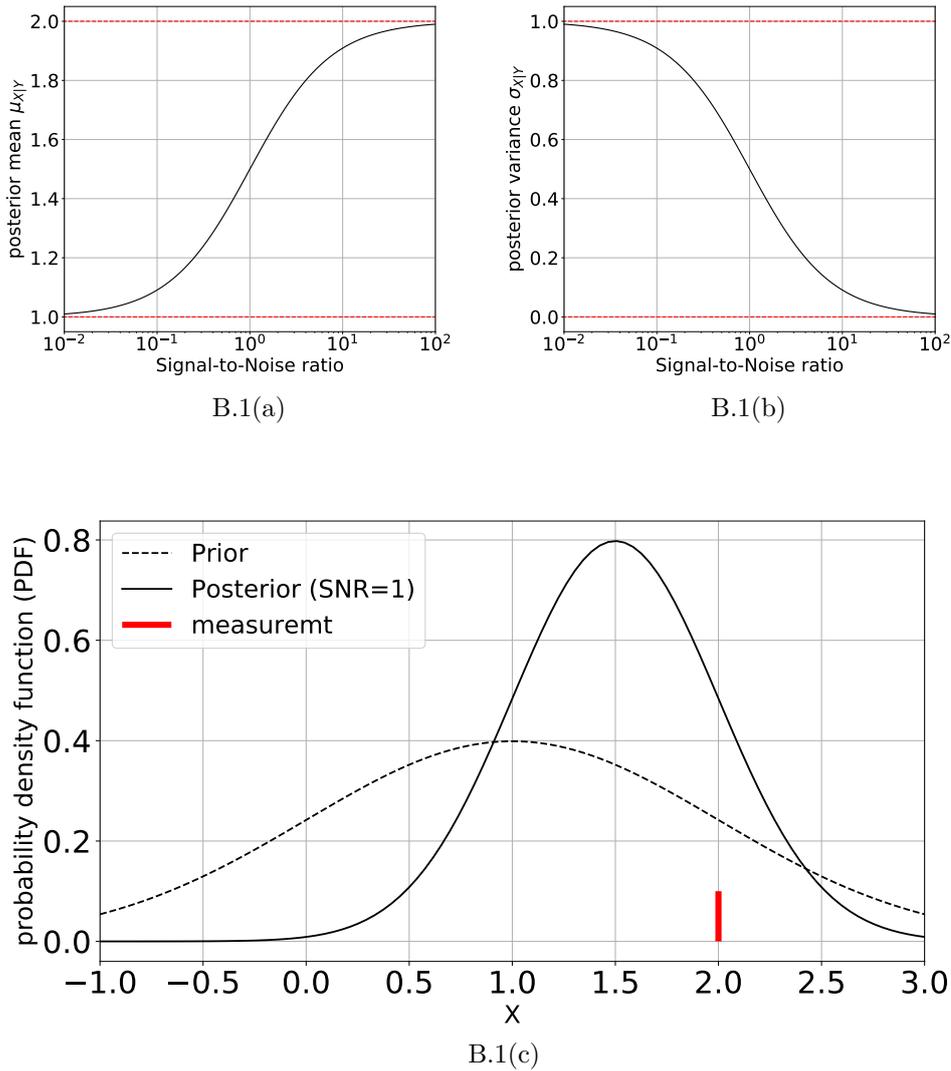


Figure B.1: Illustration on the posterior distribution dependency on the signal-to-noise ratio (SNR). B.1(c) shows, for SNR= 1, a comparison between prior and posterior distributions. B.1(a) shows the conditional mean as a function of the signal-to-noise ratio. B.1(b) shows the conditional standard deviation as a function of the signal-to-noise ratio.

the input variable  $X$  will be gradually replaced by the knowledge given by the linear system, as the signal-to-noise ratio increases. And the conditional variance, will be smaller than the prior variance, and will decrease gradually, also, as the signal-to-noise ratio increases.

Now, we go back to the analysis of the multidimensional case. To treat this case, the definition of signal-to-noise ratio must be extend:

$$\gamma = \frac{\text{tr}(A\Sigma_{XX}A^T)}{\text{tr}(\Sigma_{ee})} \quad (\text{B-23})$$

In this case, if the system of equations is not injective, the posterior variances will not asymptotically reach zero in the limit of  $\gamma \rightarrow +\infty$  (like

in figure B.1(b)). This is the challenge in equation D-26 (and in geophysical applications in general).

For an overview of Bayesian methods see (26). For an introduction to Bayesian inversion, and applications to geophysics, see (27, 55).

#### B.4 Gaussian Marginalization

Suppose, one has a multivariate Gaussian variable  $Z$ , partitioned in two groups of variables:  $Z = (X, Y)$ . Suppose, that one can measure the variables  $X$ , but not the other variables  $Y$ .

Suppose, one needs to describe the probability distribution of this subset of measurable variables  $X$ . It will be necessary to marginalize the remaining variables  $Y$ :  $p(X = x) = \int_y p(X = x, Y = y)$ .

We assume that the joint distribution  $p(X, Y)$  assumes the same functional form as equation B-11. We also assume that  $\Lambda = \Sigma^{-1}$ , and  $\Sigma$  is the joint covariance matrix.

By "completing the squares" in equation B-11 in relation to  $\Delta y$ , one arrives at two quadratic forms:

$$\begin{aligned}
 p(X = x, Y = y) &\propto \exp \left\{ -\frac{1}{2} \left[ \Delta x^T \Lambda_{XX} \Delta x \right. \right. \\
 &\quad \left. \left. + \left( \Delta y - \Lambda_{YY}^{-1} \Lambda_{YX} \Delta x \right)^T \Lambda_{YY} \left( \Delta y - \Lambda_{YY}^{-1} \Lambda_{YX} \Delta x \right) \right. \right. \\
 &\quad \left. \left. - \Delta x^T \Lambda_{XY} \Lambda_{YY}^{-1} \Lambda_{YX} \Delta x \right] \right\} \\
 &= \exp \left\{ -\frac{1}{2} \Delta x^T \left( \Lambda_{XX} - \Lambda_{XY} \Lambda_{YY}^{-1} \Lambda_{YX} \right) \Delta x \right\} \\
 &\quad \exp \left\{ -\frac{1}{2} \left( \Delta y - \Lambda_{YY}^{-1} \Lambda_{YX} \Delta x \right)^T \Lambda_{YY} \left( \Delta y - \Lambda_{YY}^{-1} \Lambda_{YX} \Delta x \right) \right\} \\
 &= Q_1(x) Q_2(x, y)
 \end{aligned} \tag{B-24}$$

Term  $Q_2(x, y)$  is a Gaussian, thus its integral in relation to  $y$  is invariant under translation, thus independent of  $x$ . Thus  $\int_y Q_2(x, y) = C$ , where  $C$  is a constant.

Thus:

$$p(X = x) \propto Q_1(x) = \exp \left\{ -\frac{1}{2} \Delta x^T \left( \Lambda_{XX} - \Lambda_{XY} \Lambda_{YY}^{-1} \Lambda_{YX} \right) \Delta x \right\} \tag{B-25}$$

Inspection of equation B-25, shows that  $p(X)$  is a Gaussian distribution.

Since  $\Sigma = \Lambda^{-1}$ , and using the same reasoning as in equation B-13, it holds that:  $\Sigma_{XX}^{-1} = \Lambda_{XX} - \Lambda_{XY} \Lambda_{YY}^{-1} \Lambda_{YX}$ . Substituting this result in equation

B-25, one obtains:

$$p(X = x) = \mathcal{N}(x \mid \mu_X; \Sigma_{XX}) \quad (\text{B-26})$$

## B.5

### Affine Transformations and Marginalization

Let us assume once again that  $X$  and  $Y$  are multivariate Gaussian random variables with the holding following relationship:

$$p(X = x) = \mathcal{N}(x \mid \mu_X, \Sigma_{XX}) \quad (\text{B-27})$$

$$p(Y = y \mid X = x) = \mathcal{N}(y \mid AX + b; \Sigma_{ee}) \quad (\text{B-28})$$

One might be interested in knowing the marginal distribution  $p(Y)$ . First, one must, compute the joint distribution  $p(X, Y)$ , which corresponds to equation B-9. Then, one use the results in section B.4, to arrive at the result:

$$p(Y = y) = \mathcal{N}(y \mid A\mu_X + b; A\Sigma_{XX}A^T + \Sigma_{ee}) \quad (\text{B-29})$$

# C

## Gaussian Approximation to a Gaussian Mixture

### C.1

#### Introduction

In this appendix, we will derive the formulas for the Gaussian approximation of a Gaussian mixture distribution. This problem was posed in section 5.3.

The first thing to notice, is that the maximization of the  $D_{\mathbf{KL}}(p||q)$  operator is equivalent of finding the probability distribution  $q$  with is the best approximation to the probability distribution  $p$ . In this apendix, it will be demonstrated the results of the optimization problem 5-13.

The tools we will use are the following:

**Fact C.1.1.**

$$\text{tr} \{BA\} = \text{tr} \{AB\} \quad (\text{C-1})$$

**Fact C.1.2.**

$$\partial_A \text{tr} \{BA\} = B \quad (\text{C-2})$$

**Fact C.1.3.**

$$\partial_A \log \det(A) = [A^{-1}]^T \quad (\text{C-3})$$

### C.2

#### Optimization

In section 5.3, we found the following optimization problem:

$$\begin{aligned} \arg \min_{\mu_*, \Sigma_*} D_{\mathbf{KL}}(p||p_*) &= - \int_m p(m) \log \frac{p_*(m)}{p(m)} \\ \text{s.t. } p_*(m) &= \mathcal{N}(m | \mu_*, \Sigma_*) \end{aligned} \quad (\text{C-4})$$

Where,  $p(m)$  is a Gaussian mixture, given by:

$$\begin{aligned} p(m) &= \sum_{\pi \in \Omega^N} p(m | \pi) p(\pi) \\ &= \sum_{\pi \in \Omega^N} \mathcal{N}(m | \mu_\pi, \Sigma_\pi) p(\pi) \end{aligned} \quad (\text{C-5})$$

So, discarding the terms without  $p_*$ , one wants to maximize:

$$\begin{aligned}
& - \int_m p(m) \log(p_*(m)) \\
& = \frac{1}{2} \sum_{\pi \in \Omega^N} p(\pi) \int_m \mathcal{N}(m \mid \mu_\pi, \Sigma_\pi) \left\{ \log \det(\Sigma_*) + (m - \mu_*)^T \Sigma_*^{-1} (m - \mu_*) \right\} \\
& = \frac{1}{2} \sum_{\pi \in \Omega^N} p(\pi) \int_m \mathcal{N}(m \mid \mu_\pi, \Sigma_\pi) \left\{ \log \det(\Sigma_*) + \text{tr} \left[ \Sigma_*^{-1} (m - \mu_*) (m - \mu_*)^T \right] \right\}
\end{aligned} \tag{C-6}$$

The term  $(m - \mu_*)(m - \mu_*)^T$  can be rewritten:

$$\begin{aligned}
(m - \mu_*)(m - \mu_*)^T & = (m - \mu_\pi)(m - \mu_\pi)^T - \Delta\mu(m - \mu_\pi)^T \\
& \quad - (m - \mu_\pi)\Delta\mu^T + \Delta\mu\Delta\mu^T
\end{aligned} \tag{C-7}$$

$$\Delta\mu = \mu_* - \mu_\pi \tag{C-8}$$

So, we can compute the following integrals:

$$\int_m \mathcal{N}(m \mid \mu_\pi, \Sigma_\pi) (m - \mu_\pi)(m - \mu_\pi)^T = \Sigma_\pi \tag{C-9}$$

$$\int_m \mathcal{N}(m \mid \mu_\pi, \Sigma_\pi) \Delta\mu(m - \mu_\pi)^T = 0 \tag{C-10}$$

$$\int_m \mathcal{N}(m \mid \mu_\pi, \Sigma_\pi) (m - \mu_\pi)\Delta\mu^T = 0 \tag{C-11}$$

$$\int_m \mathcal{N}(m \mid \mu_\pi, \Sigma_\pi) = 1 \tag{C-12}$$

So, the optimization problem, can be further rewritten as the optimization of:

$$\frac{1}{2} \sum_{\pi \in \Omega^N} p(\pi) \left\{ -\log \det(\Sigma_*^{-1}) + \text{tr} \left[ \Sigma_*^{-1} (\Sigma_\pi + \Delta\mu\Delta\mu^T) \right] \right\} \tag{C-13}$$

Deriving in relation to  $\mu_*$ , and equalling to zero:

$$\sum_{\pi \in \Omega^N} p(\pi) \Sigma_*^{-1} \Delta\mu = 0 \tag{C-14}$$

$$\mu_* = \sum_{\pi \in \Omega^N} p(\pi) \mu_\pi = (\mu_S, \dots, \mu_S)^T \tag{C-15}$$

$$\mu_S = \sum_{i=0}^K p_s(\pi_1 = i) \mu_i \tag{C-16}$$

Doing the same, in relation to  $\Sigma_*^{-1}$ , one gets:

$$\sum_{\pi \in \Omega^N} p(\pi) \left\{ -\Sigma_* + \Sigma_\pi + \Delta\mu\Delta\mu^T \right\} = 0 \quad (\text{C-17})$$

$$\Sigma_* = \sum_{\pi \in \Omega^N} p(\pi) \left\{ \Sigma_\pi + \Delta\mu\Delta\mu^T \right\} \quad (\text{C-18})$$

$$[\Sigma_*]_{n,n+l} = \sum_{i=0}^K \sum_{j=0}^K p(\pi_1 = i, \pi_{n+l} = j) (\mu_i - \mu_S)(\mu_j - \mu_S)^T, \quad l \geq 1 \quad (\text{C-19})$$

$$[\Sigma_*]_{n,n} = \sum_{i=0}^K p_s(\pi = i) \left[ \Sigma_i + (\mu_i - \mu_S)(\mu_i - \mu_S)^T \right] \quad (\text{C-20})$$

## D AVA Matrices

### D.1 Definitions

A basic tool in this Appendix is the kronecker product:

**Definition D.1.1** (Kronecker Product). *Given two matrices  $A$  and  $B$ , of size  $(n_A, m_A)$ , and  $(n_B, m_B)$ , respectively. We define the kronecker product of the matrices  $A \otimes B$  by the block matrix defined by:*

$$A \otimes B = \begin{bmatrix} a_{1,1}B & \cdots & a_{1,m_A}B \\ \cdots & \ddots & \cdots \\ a_{n_A,1}B & \cdots & a_{n_A,m_A}B \end{bmatrix}$$

The  $A \otimes B$  matrix has size  $(n_A n_B, m_A m_B)$

Another definition of interest to make the notations lighter, is the following:

**Definition D.1.2** (vectorize). *Given a set of vectors  $M$  ( $v_1, \dots, v_M$ ), all with the same number of elements  $N$  (or, given a matrix of size  $(N, M)$ ), we define the vectorize operations over them, as the vertical stacking of their elements:*

$$\text{vec}(v_1, \dots, v_M) = \begin{bmatrix} v_{11} \\ \cdots \\ v_{1N} \\ v_{21} \\ \cdots \\ v_{2N} \\ \cdots \\ v_{M1} \\ \cdots \\ v_{MN} \end{bmatrix}$$

## D.2 Linearisation of AVA Equations

Consider an incident plane compressional wave (P-wave) hitting an interface between two homogeneous isotropic elastic media, as in figure 3.1. From this incident P-wave, four waves are generated: reflected and transmitted compressional and shear waves, as illustrated in figure 3.1.

The system of equations that estimates the amplitudes of each generated wave, given a unit amplitude incident P-wave, are known as Zoeppritz equations (56). The books (3), and (43) are good references to the topic.

$$\begin{bmatrix} R_P \\ R_S \\ T_P \\ T_S \end{bmatrix} = \begin{bmatrix} -\sin(\theta_1) & -\cos(\phi_1) & \sin(\theta_2) & \cos(\phi_2) \\ \cos(\theta_1) & -\sin(\phi_1) & \cos(\theta_2) & -\sin(\phi_2) \\ \sin(2\theta_1) & \frac{v_{P1}}{v_{S1}} \cos(2\phi_1) & \frac{\rho_2 v_{S2}^2 v_{P1}}{\rho_1 v_{S1}^2 v_{P2}} \cos(2\phi_1) & \frac{\rho_2 v_{S2}^2 v_{P1}}{\rho_1 v_{S1}^2} \cos(2\phi_2) \\ -\cos(2\phi_1) & \frac{v_{S1}}{v_{P1}} \sin(2\phi_1) & \frac{\rho_2^2 v_{P2}}{\rho_1 v_{P1}} \cos(2\phi_2) & \frac{\rho_2^2 v_{S2}}{\rho_1 v_{P1}} \sin(2\phi_2) \end{bmatrix}^{-1} \begin{bmatrix} \sin(\theta_1) \\ \cos(\theta_1) \\ \sin(2\theta_1) \\ \cos(2\phi_1) \end{bmatrix} \quad (\text{D-1})$$

In the system of equations D-1,  $R_P$ ,  $R_S$ ,  $T_P$  and  $T_S$  are the reflected compressional wave, compressional shear wave, transmitted compressional wave, and transmitted shear wave, respectively. The variables  $\theta_1$ ,  $\phi_1$ ,  $\theta_2$  and  $\phi_2$  are the corresponding angles with the interface normal vector (in accordance with figure 3.1).

The above system of equations D-1 is non linear on the elastic parameters, so there is no analytic way to formulate its inverse. Some authors (for example, (44), (57) and (43)) have developed approximations to the above equations, in order to make it easier to understand the contribution of each elastic parameter in the reflection behaviour.

As we will show, the approximations to Zoeppritz equation makes it easier to invert.

Usually, one is interested in the reflected compressional wave. A well known approximation for  $R_P$  is the Aki-Richards 3 (43) terms equation:

$$R(\theta) = A + B \sin^2(\theta) + C \tan^2(\theta) \quad (\text{D-2})$$

$$A = \frac{1}{2} \left( \frac{\Delta v_P}{\bar{v}_P} + \frac{\Delta \rho}{\bar{\rho}} \right) \quad (\text{D-3})$$

$$B = -2 \left( \frac{\bar{v}_S}{\bar{v}_P} \right)^2 \left( 2 \frac{\Delta v_S}{\bar{v}_S} + \frac{\Delta \rho}{\bar{\rho}} \right) \quad (\text{D-4})$$

$$C = \frac{1}{2} \frac{\Delta v_P}{\bar{v}_P} \quad (\text{D-5})$$

In equation D-2,  $R$  is the same as  $R_P$ , and  $\theta$  is the same as  $\theta_1$  in equation D-1. The Aki-Richards equation D-2 depends on terms like:

$$\frac{\Delta x}{\bar{x}} = 2 \frac{x_2 - x_1}{x_2 + x_1} \quad (\text{D-6})$$

Where  $x$  denotes the elastic property (density  $\rho$ , shear velocity  $v_S$ , or acoustic velocity  $v_P$ ), and 2 and 1 denote the lower, and upper media, respectively. And  $\bar{x}$  equals to the mean property between both media.

Thus, the dependency of the reflectivity on the elastic parameters is a non linear one. But, for weak contrasts, equation D-6 can be approximated, as in (44) and (45), by:

$$\frac{\Delta x}{\bar{x}} \approx \log(x_2) - \log(x_1) = \Delta \log(x) \quad (\text{D-7})$$

Figures D.1 and D.2 show the quality of the *log* approximation to the elastic contrast. In figure D.2 the elastic contrast error is always below 4% of the total exact elastic contrast, but for practical situations (weak elastic contrast), this error will be lower.

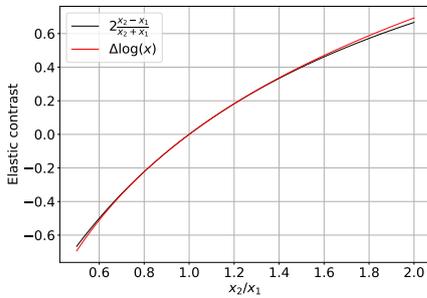


Figure D.1: log approximation (in red) to the elastic contrast terms  $\frac{\Delta x}{\bar{x}}$  (in black)

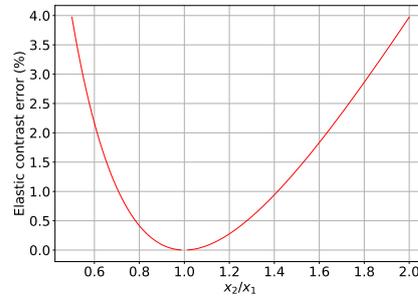


Figure D.2: Error of the log approximation

So, should we re-parametrize our problem in terms of the log of the elastic parameters, instead of in terms of the elastic parameters themselves, we have a linear equation, apart from the term  $k = 4(\frac{v_S}{v_P})^2$ , which is generally, handled by assuming a fixed (or smooth background model), which is not inverted for.

Equation D-2 can be grouped in terms of each elastic parameter (as in (13)), to arrive at equation:

$$R(\theta) = A_{\theta,\rho} \Delta \log(\rho) + A_{\theta,v_P} \Delta \log(v_P) + A_{\theta,v_S} \Delta \log(v_S) \quad (\text{D-8})$$

$$A_{\theta,\rho} = \frac{1}{2} (1 - k \sin^2(\theta)) \quad (\text{D-9})$$

$$A_{\theta,v_P} = \frac{1}{2} (1 + \tan^2(\theta)) \quad (\text{D-10})$$

$$A_{\theta,v_S} = -k \sin^2(\theta) \quad (\text{D-11})$$

Equation D-8 can be written in matrix form as:

$$R(\theta) = \begin{bmatrix} A_{\theta,\rho} & A_{\theta,v_P} & A_{\theta,v_S} \end{bmatrix} \begin{bmatrix} \Delta \log(\rho) \\ \Delta \log(v_P) \\ \Delta \log(v_S) \end{bmatrix} \quad (\text{D-12})$$

We can denote  $A(\theta) := [A_{\theta,\rho} \quad A_{\theta,v_P} \quad A_{\theta,v_S}]$ .

The above equation D-12 is linear, and each elastic parameter's contribution to the output is given by a  $A_{\theta,x}$  coefficient. Each observation angle will yield an additional equation. With three observation angles, the system of equations is invertible. With more than three observation angles, it may be solved using least squares method. For any number of observation angles the tools developed in appendix B are applicable (although they will yield little help with less than 3 angles).

One important point is that one can only hope to invert the AVA equation to the elastic contrasts between the upper and lower media:  $\Delta \log(x) = \log(x_2) - \log(x_1)$ . The absolute elastic values of the lower (or upper) layer could only be determined if one knew beforehand the upper (or the lower, respectively) elastic parameters.

### D.3

#### AVA Reflectivity Profile

The last section has shown how to linearize and express in matrix form the AVA relations in terms of elastic contrasts in an interface. Now, say one has a sequence of elastic properties (from well logs, for example), and wants to simulate the corresponding sequence of reflectivity coefficients. In this section, we will develop the matrix formulation for this problem.

Taking equation D-12, and stacking it for different angles, one obtains the following matrix equation:

$$\begin{bmatrix} R_1(\theta) \\ \dots \\ R_N(\theta) \end{bmatrix} = \begin{bmatrix} A_{\theta,\rho} \mathbb{1} & A_{\theta,v_P} \mathbb{1} & A_{\theta,v_S} \mathbb{1} \end{bmatrix} \begin{bmatrix} \Delta \log(\rho)_1 \\ \dots \\ \Delta \log(\rho)_N \\ \Delta \log(v_P)_1 \\ \dots \\ \Delta \log(v_P)_N \\ \Delta \log(v_S)_1 \\ \dots \\ \Delta \log(v_S)_N \end{bmatrix} \quad (\text{D-13})$$

Where  $\mathbb{1}$  is the identity matrix with size  $(N, N)$ . We assume for equation D-13 that  $(\frac{\bar{v}_S}{\bar{v}_P})^2$  is constant throughout the profile, to make the matrix formulation uncluttered.

Now, say we have measured the reflection coefficients for  $M$  different angles of incidence  $(\theta_1, \dots, \theta_M)$ . We can group the different reflection coefficients equations as follows:

$$\begin{bmatrix} R_1(\theta_1) \\ \dots \\ R_N(\theta_1) \\ R_1(\theta_2) \\ \dots \\ R_N(\theta_2) \\ \dots \\ R_1(\theta_M) \\ \dots \\ R_N(\theta_M) \end{bmatrix} = \begin{bmatrix} A_{\theta_1, \rho} \mathbb{1} & A_{\theta_1, v_P} \mathbb{1} & A_{\theta_1, v_S} \mathbb{1} \\ A_{\theta_2, \rho} \mathbb{1} & A_{\theta_2, v_P} \mathbb{1} & A_{\theta_2, v_S} \mathbb{1} \\ \dots & \dots & \dots \\ A_{\theta_M, \rho} \mathbb{1} & A_{\theta_M, v_P} \mathbb{1} & A_{\theta_M, v_S} \mathbb{1} \end{bmatrix} \begin{bmatrix} \Delta \log(\rho)_1 \\ \dots \\ \Delta \log(\rho)_N \\ \Delta \log(v_P)_1 \\ \dots \\ \Delta \log(v_P)_N \\ \Delta \log(v_S)_1 \\ \dots \\ \Delta \log(v_S)_N \end{bmatrix} \quad (\text{D-14})$$

The above equation can be re-written in terms of kronecker product and the  $vec$  operator:

$$vec(R(\theta_1), \dots, R(\theta_M)) = \mathbf{A} vec(\Delta \log(\rho), \Delta \log(v_P), \Delta \log(v_S)) \quad (\text{D-15})$$

$$\mathbf{A} = \begin{bmatrix} A(\theta_1) \\ \dots \\ A(\theta_M) \end{bmatrix} \otimes \mathbb{1} \quad (\text{D-16})$$

Finally, the vector  $vec(\Delta \log(\rho), \Delta \log(v_P), \Delta \log(v_S))$  can be written as a function of the elastic profiles  $vec(\log(\rho), \log(v_P), \log(v_S))$ . First, we note that:

$$\begin{bmatrix} \Delta \log(x)_1 \\ \Delta \log(x)_2 \\ \dots \\ \Delta \log(x)_N \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix} \begin{bmatrix} \log(x)_1 \\ \log(x)_2 \\ \log(x)_3 \\ \dots \\ \log(x)_N \\ \log(x)_{N+1} \end{bmatrix} \quad (\text{D-17})$$

$$:= D \log(\mathbf{x})$$

Where  $D$  is the matrix that takes the finite differences. Based on equation D-17, one can write equation D-18 for the vector  $vec(\Delta \log(\rho), \Delta \log(v_P), \Delta \log(v_S))$  of all elastic derivatives:

$$\begin{bmatrix} \Delta \log(\rho)_1 \\ \dots \\ \Delta \log(\rho)_N \\ \Delta \log(v_P)_1 \\ \dots \\ \Delta \log(v_P)_N \\ \Delta \log(v_S)_1 \\ \dots \\ \Delta \log(v_S)_N \end{bmatrix} = \begin{bmatrix} D & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & D & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & D \end{bmatrix} \begin{bmatrix} \log(\rho)_1 \\ \dots \\ \log(\rho)_{N+1} \\ \log(v_P)_1 \\ \dots \\ \log(v_P)_{N+1} \\ \log(v_S)_1 \\ \dots \\ \log(v_S)_{N+1} \end{bmatrix} \quad (\text{D-18})$$

$$= \mathbf{D} \text{vec}(\log(\rho), \log(v_P), \log(v_S))$$

$$\mathbf{D} = \mathbf{1} \otimes D \quad (\text{D-19})$$

In the above equation, the identity matrix  $\mathbf{1}$  has size  $(3, 3)$ .

#### D.4 AVA Seismic

In reality, the signal which travels through earth is not a Dirac's delta, but a finite length signal, called wavelet. As a consequence, what one measures at the surface is a convolved version of the reflectivity:

$$s(\theta, t) = \mathbf{w}(\theta) * \mathbf{R}(\theta) := \int_{-\infty}^{+\infty} w(\theta, \tau) R(\theta, t - \tau) d\tau \quad (\text{D-20})$$

In the case of digital signal, the above continuous convolution is substituted by a discrete convolution:

$$s(\theta, i) = \mathbf{w}(\theta) * \mathbf{R}(\theta) := \sum_{j=-\infty}^{+\infty} w_j(\theta) R_{i-j}(\theta) \quad (\text{D-21})$$

Figure D.3 represents the above equation as a calculation work-flow. Since, both the reflectivity profile (size  $N$ ), and the wavelet  $w$  (size  $k = 2l + 1$ ) have finite duration, the above equation can be expressed in matrix form, by using the corresponding Toeplitz matrix:

$$\begin{aligned}
 \begin{bmatrix} s_{1+l}(\theta) \\ \dots \\ s_{N-l}(\theta) \end{bmatrix} &= \begin{bmatrix} w_k & w_{k-1} & \dots & w_1 & 0 & \dots & 0 & \dots & 0 & 0 & 0 \\ 0 & w_k & \dots & w_2 & w_1 & \dots & 0 & \dots & 0 & 0 & 0 \\ & & \dots & & & \dots & & & & & \\ 0 & 0 & \dots & 0 & 0 & \dots & w_k & w_{k-1} & \dots & w_1 & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & w_k & \dots & w_2 & w_1 \end{bmatrix} \begin{bmatrix} R_1(\theta) \\ \dots \\ R_N(\theta) \end{bmatrix} \\
 &:= W(\theta) \begin{bmatrix} R_1(\theta) \\ \dots \\ R_N(\theta) \end{bmatrix}
 \end{aligned} \tag{D-22}$$

In the above equation, only  $N - 2l$  output samples are computed, because only these samples can be computed given the finite duration inputs without padding, as depicted in the sketch D.4. This definition of the convolution is usually referred to as valid convolution in programming packages (for example, in Python's Numpy (58)).

The advantage of using this definition, instead of the usually used full convolution formulation (as in (13)), is that it makes no assumption on the input vector: the full convolution convention assumes that the unknown input samples ( $R_1$  through  $R_l$ , and from  $R_{N-l+1}$  to  $R_N$ , in equation D-22) are zero.

In other words, in the valid convolution formulation, the unknown samples are dealt with correctly, for they are considered additional variables in the input vector. No border effect is expected when inverting this formulation of convolution.

So, for all the observation angles, the equation becomes:

$$\begin{aligned}
 \begin{bmatrix} s_{1+l}(\theta_1) \\ \dots \\ s_{N-l}(\theta_1) \\ s_{1+l}(\theta_2) \\ \dots \\ s_{N-l}(\theta_2) \\ \dots \\ s_{1+l}(\theta_M) \\ \dots \\ s_{N-l}(\theta_M) \end{bmatrix} &= \begin{bmatrix} W(\theta_1) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & W(\theta_2) & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & W(\theta_M) \end{bmatrix} \begin{bmatrix} R_1(\theta_1) \\ \dots \\ R_N(\theta_1) \\ R_1(\theta_2) \\ \dots \\ R_N(\theta_2) \\ \dots \\ R_1(\theta_M) \\ \dots \\ R_N(\theta_M) \end{bmatrix} \\
 &:= \mathbf{W} \text{vec}(R(\theta_1), \dots, R(\theta_M))
 \end{aligned} \tag{D-23}$$

Finally, we can join equations D-12, D-18, and D-23 to write a linear

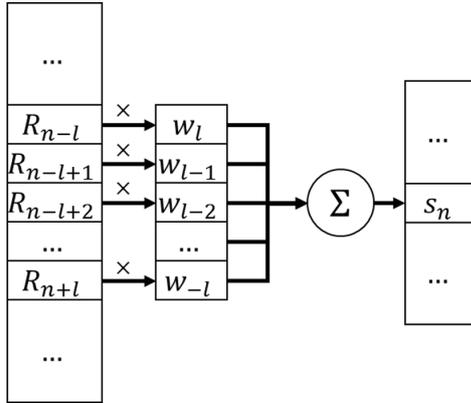


Figure D.3: Convolution calculation for each output sample  $s_n$ .

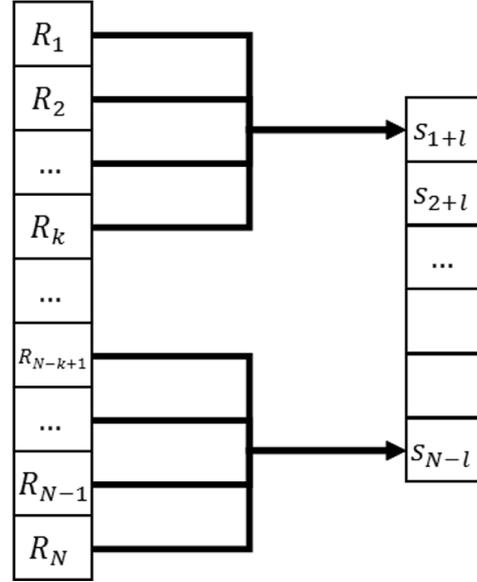


Figure D.4: Valid convolution illustration, a limited number of output samples are computed

system relating the elastic property profile to the reflectivity measurement profile.

$$vec(s(\theta_1), \dots, s(\theta_M)) = \mathbf{WAD}vec(\log(\rho), \log(v_P), \log(v_S)) \quad (\text{D-24})$$

Or, succinctly:

$$s = \mathbf{WAD}m := \mathbf{G}m \quad (\text{D-25})$$

In practice, we have the seismic measurement  $s$ , and wish to invert for the elastic properties  $m$ . The linear inversion problem  $s = \mathbf{G}m$ , can be dealt with using the tools developed in Appendix B.

## D.5

### AVA Linear Inversion

As discussed in section B.3, Bayesian inversion is the methodology used to describe the knowledge about a variable in a probabilistic framework, when this variable is only known from an indirect measurement. It is suitable when the forward model (relation between the variable and the measurement) is ill-posed, and/or the measurement is noisy.

Section 2.3 introduces the idea of inverting seismic data to the elastic properties. For a further discussion on AVA inversion, see (13, 59, 60).

In order to exemplify this technique and discuss its limitations, let's assume that we have seismic data measured at only one angle of incidence  $\theta$ .

In this case, equation D-25 can be simplified to:

$$s(\theta) = W(\theta)D \log \left( \rho^{A_{\theta,\rho}} v_P^{A_{\theta,v_P}} v_S^{A_{\theta,v_S}} \right) := W D m \quad (\text{D-26})$$

Where, in the above equation, the elastic parameter triplet is reduced to only  $m = \log \left( \rho^{A_{\theta,\rho}} v_P^{A_{\theta,v_P}} v_S^{A_{\theta,v_S}} \right) := \log(I_E)$ . The quantity  $I_E$  is known as elastic impedance (51), and we will refer to it simply as impedance.

The matrix  $WD$  expresses the convolution with the seismic effective kernel:  $w * \frac{d}{dt}$ .

Since this equation have only convolution effects, we can write it in the frequency domain, using Fourier transform:

$$\mathcal{F} \left\{ w * \frac{d}{dt} \right\} (\omega) \times \mathcal{F} \{ m \} (\omega) = \mathcal{F} \{ s \} (\omega) \quad (\text{D-27})$$

But, from figure 3.3, one sees that equation D-27 is undefined, or at least highly unstable because the low, and high frequencies of the Fourier spectrum of the seismic kernel are very low. In other words, since the transfer function of the linear system annihilates these frequency components, there is no way of recovering them, by using linear theory.

In terms of low frequencies, one can sum any constant (or low frequency trend) to a solution  $m$  of equation D-26, and it will yield another solution  $m + c$ , as good as the original one. This is a severe problem, since we need the absolute values of the impedances, in order to understand the types of rocks and their petrophysical properties.

The high frequencies determine the details of the solution. So, one can sum any high frequency function  $c(t)$  (with frequency content above  $\sim 0.25$ cycles/sample in figure 3.3) to a solution  $m$ , and the result will be another solution as good as the first one. It means, that we cannot unravel the detailed geology from the seismic data, and it makes sense, for seismic images are a blurred version of the sub-surface properties.

The only hope is that one have some prior knowledge about the solution at low frequencies, and high frequencies. If this is the case, one can use Bayesian framework for inversion.

Even though, the Bayesian Inversion, will not solve the above mentioned ambiguities in the inversion of equation D-26, it will estimate confidence bands, and provide tools to stochastically sample possible solutions conditioned on the seismic measurement and on the prior information.

(61) shows that actual uncertainty in the inverted impedances is bigger than the estimated by Bayesian theory. The reason is, what they call uncertainty in the low-frequency model (prior mean  $\mu_X$  in equation B-3).

Figure D.5 compares Bayesian inversion results, given the correct prior

knowledge, versus the inversion result if one perturb the prior in different frequency bands.

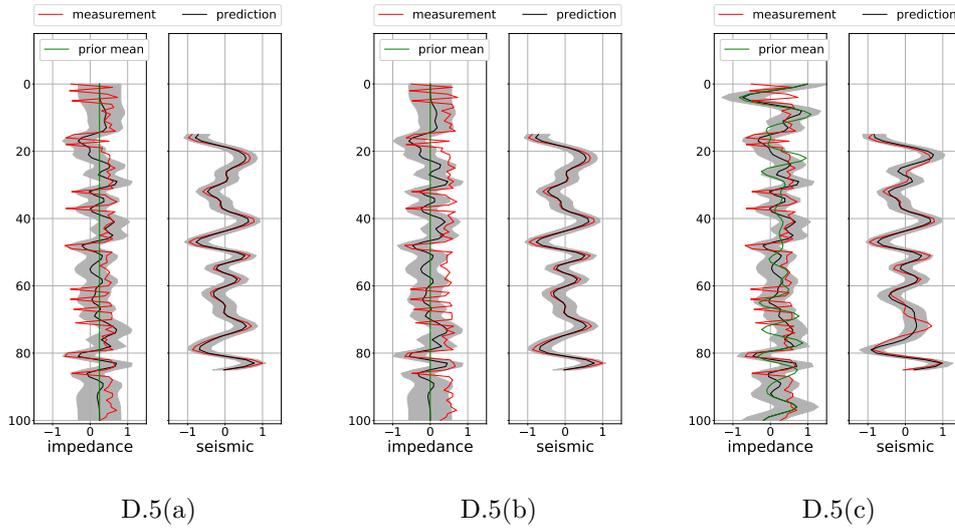


Figure D.5: The graphs compare the Bayesian linear inversion result, for different prior models. D.5(a) shows the inversion results with the correct prior model, D.5(b) shows the result with a constant deviation in the correct prior, and D.5(c) shows the result with a random oscillatory perturbation (within the wavelet spectrum) to the prior. The recovered impedances are different, but the modelled seismic is undistinguishable.

It can be noticed that the constant perturbation produces a bias in the posterior mean solution  $\mu_{X|Y}$ . As explained in D.5, this is due to the lack of low frequencies in the seismic effective kernel (figure 3.3).

In figure D.5(c), the oscillatory perturbation in the prior mean  $\mu_X$  is within the seismic spectrum (figure 3.3), so the inversion was able to correct the prior knowledge properly, given the seismic measurements.

Notwithstanding the biases generated on the inverted impedances  $\mu_{X|Y}$ , the modelled seismic synthetics fit the measurement equally well, for the three cases. So, there is no way of determining the true low-frequency content of the impedance from the seismic measurement. (61) proposes a hierarchical Bayesian approach to include the low-frequencies uncertainty in the inversion formalism.

Similarly, the high frequency components of the impedance profile are filtered out by the seismic effective kernel. The posterior mean  $\mu_{X|Y}$  is a smooth solution to the inversion problem. Although, the high frequency information can be obtained by sampling  $m$  from the posterior probability distribution using equations B-15 to B-17 (see (62) for random variables sampling techniques), this high frequency content is subject to the correct

knowledge of the prior covariance matrix (which is estimated from nearby wells).

Summarizing the above discussion, a prior mean of the impedances profile (also known as the prior model, low frequency model, or background model) is crucial for the seismic inversion. But it depends on a prior knowledge of geology, which is minimal in regions between drilled wells.

Usually, inversion softwares create a prior model based on nearby drilled wells, by laterally interpolating their impedances. (63, 64, 65) show advanced methodologies for prior model building. Still, the available methods are a lot limited, and impossible to apply when there are no nearby wells. See (61, 66, 64) for a discussion on the difficulties and implications of prior model creation.

The above mentioned uncertainties on the elastic properties inverted from seismic impacts *facies* probabilities estimation, if one is to use these properties as input for rock physics inversion (illustrated in figure 2.2).

This thesis' inversion method, on the other hand, does not need an elastic background model, for it works on *facies* sequencing. *Facies* inversion uses nearby wells to determine the elastic properties ranges that each *facies* might have, and by working on the possible *facies* sequences, it can be used to improve elastic inversion (as explained in section 5.6).

Although seismic to *facies* inversion does not depend on an elastic prior model, it depends on the *facies* prior knowledge, which is also hard to be determined in regions away from well control. This problem can be treated more easily for it can be modelled from geological experience and analogues. So, seismic to *facies* inversion highly depends on the interaction between geophysicists and geologists.

# E

## Markov Chain Monte Carlo

### E.1 Introduction

Monte Carlo is the technique of numerically calculating the expectation of some random variable. Usually, it is achieved through use of the Large Numbers Theorem (67). But, the use of the Large Numbers Theorem, requires one to be able to take independent samples from the desired probability distribution. A good reference on random numbers generation is (62).

But for multidimensional problems, or complex combinatorial problems, the sampling itself might be hard to be implemented, or even, the probability distribution is known up to an unknown normalizing constant (such as in equation 3-17). In these cases, one uses the technique known as Markov Chain Monte Carlo (MCMC), which instead of sampling independent identically distributed samples of the proposed distribution, uses a Markov Chain, whose stationary distribution is the required distribution.

Since, this thesis main focus is not on this technique, it will not be demonstrated it in depth, but instead it will be shown its reasoning, and only for the discrete random variable case. For introductory reading on the topic, we suggest (68), (26), and (69).

**Definition E.1.1** (Markov Chain: discrete case). *A random sequence  $X_1, X_2, \dots$ , where  $X \in \Omega = \{0, \dots, K-1\}$ , is said to be a Markov Chain, with starting probability  $\lambda$ , and transition probability matrix  $T$ , if the following relations hold:*

$$\begin{aligned} p(X_n = i_n \mid X_1 = i_1, \dots, X_{n-1} = i_{n-1}) &= p(X_n = i_n \mid X_{n-1} = i_{n-1}) \\ &= T_{i_n, i_{n-1}} \end{aligned} \quad (\text{E-1})$$

$$p(X_1 = i) = \lambda_i \quad (\text{E-2})$$

## E.2 MCMC Outlined

The Markov Chains of interest in this thesis are so called irreducible (equivalently, they are said to be defined by irreducible transition matrices). They are defined by the two equivalent conditions:

$$\forall i, j \in \Omega, \forall n > 0, \exists m > 0, \text{s.t. } p(X_{n+m} = j \mid X_n = i) > 0 \quad (\text{E-3})$$

$$\iff \forall i, j \in \Omega, \forall n > 0, \exists m > 0, \text{s.t. } [T]_{i,j}^m > 0 \quad (\text{E-4})$$

In other words, in an irreducible Markov Chain, every state is achievable from every state along the evolution of the Chain. As a consequence, every state is visited infinitely many times, along the evolution of the Chain. For irreducible chains, the following theorem holds:

**Theorem E.2.1** (Perron Frobenius Theorem). *Given, an irreducible Markov Chain on a finite set  $\Omega$ , with transition matrix  $T$ , there exists a unique vector  $p$ , such that:*

$$\sum_{i=0}^{K-1} p_i = 1 \quad (\text{E-5})$$

$$p_i \geq 0, \quad \forall i \in \{0, \dots, K-1\} \quad (\text{E-6})$$

$$p_j = \sum_{i=0}^{K-1} p_i T_{i,j} \quad (\text{E-7})$$

And let  $\lambda$ , any probability distribution on  $\Omega$ . Let the vector sequence  $\lambda^n$ , defined by  $\lambda^0 = \lambda$ , and  $\lambda_j^n = \sum_{i=0}^{K-1} \lambda_i^{n-1} T_{i,j}$ . The sequence converges:

$$\lim_{n \rightarrow \infty} \lambda^n = p \quad (\text{E-8})$$

The vector  $p$  is known as the stationary distribution of the Markov Chain.

The Perron Frobenius Theorem means that independently of the initial state of the Markov Chain, after enough steps, it will converge to the same probability distribution, given a transition matrix. There is a nice demonstration of this theorem in (69).

**Theorem E.2.2** (MCMC). *Say  $X$  is a random variable with probability distribution  $p$ , on the finite set  $\Omega$ . Say,  $q$  is any irreducible transition matrix. Then the following algorithm gener-*

ates an irreducible Markov Chain with stationary distribution  $p$ :

**Data:** Proposed transition matrix  $q$ , and required probability distribution  $p$

**Result:** Random sequence  $X_1, X_2, \dots, X_N$ , sampled from  $p$

```

1 Initialize  $X_1$ , with whatever heuristic;
2 for  $n = 1$  to  $N$  do
3   sample  $X'$  from  $q_{X_n, X'}$ , and sample  $u \sim \mathcal{U}[0, 1]$ ;
4   compute  $\alpha_{X_n, X'} = \min \left\{ \frac{p_{X'} q_{X', X_n}}{p_{X_n} q_{X_n, X'}}, 1 \right\}$ ;
5   if  $u < \alpha_{X_n, X'}$  then
6      $X_{n+1} = X'$ ;
7   else  $X_{n+1} = X_n$ ;
8   end
9 end

```

**Algorithm 5:** MCMC sampling

*Proof.* The resulting transition matrix, from the above algorithm is:

$$T_{i,j} = q_{i,j} \alpha_{i,j}, \quad j \neq i \quad (\text{E-9})$$

$$T_{i,i} = q_{i,i} + \sum_{j \neq i} q_{i,j} (1 - \alpha_{i,j}) \quad (\text{E-10})$$

Because, in order to transition from  $i$  to  $j$ , first the transition must be proposed, and then it must be accepted. On the other hand, for a transition from  $i$  to itself, either, this transition was proposed and accepted, or any other transition was proposed but rejected.

Then, one must check that  $p$  will be the stationary distribution of this transition matrix, by simply checking that equation E-7 holds. From theorem E.2.1,  $p$  is the unique stationary distribution of the transition matrix  $T$ .

$$\begin{aligned}
\sum_{i=0}^{K-1} p_i T_{i,j} &= \sum_{i \neq j} p_i T_{i,j} + p_j T_{j,j} \\
&= \sum_{i \neq j} \{p_i q_{i,j} \alpha_{i,j}\} + p_j q_{j,j} + p_j \sum_{i' \neq j} \{q_{j,i'} (1 - \alpha_{j,i'})\} \\
&= p_j q_{j,j} + \sum_{i \neq j} \{p_i q_{i,j} \alpha_{i,j} + p_j q_{j,i} (1 - \alpha_{j,i})\} \\
&= p_j q_{j,j} + \sum_{i \neq j} A(i, j)
\end{aligned} \quad (\text{E-11})$$

In the summation of the  $A(i, j)$  terms, for those terms  $i$ , for which  $\alpha_{i,j} = \frac{p_j q_{j,i}}{p_i q_{i,j}} < 1$ , and consequently  $\alpha_{j,i} = 1$ :

$$\begin{aligned}
A(i, j) &= p_i q_{i,j} \alpha_{i,j} + p_j q_{j,i} (1 - \alpha_{j,i}) \\
&= p_i q_{i,j} \frac{p_j q_{j,i}}{p_i q_{i,j}} + p_j q_{j,i} (1 - 1) \\
&= p_j q_{j,i}
\end{aligned} \tag{E-12}$$

For those  $i$ , for which  $\alpha_{j,i} = \frac{p_i q_{i,j}}{p_j q_{j,i}} < 1$ , and consequently  $\alpha_{i,j} = 1$ :

$$\begin{aligned}
A(i, j) &= p_i q_{i,j} \alpha_{i,j} + p_j q_{j,i} (1 - \alpha_{j,i}) \\
&= p_i q_{i,j} + p_j q_{j,i} \left(1 - \frac{p_i q_{i,j}}{p_j q_{j,i}}\right) \\
&= p_i q_{i,j} + p_j q_{j,i} - p_i q_{i,j} = p_j q_{j,i}
\end{aligned} \tag{E-13}$$

Substituting back the above terms, the summation becomes:

$$\begin{aligned}
\sum_{i=0}^{K-1} p_i T_{i,j} &= p_j q_{j,j} + \sum_{i \neq j} A(i, j) \\
&= p_j q_{j,j} + \sum_{i \neq j} p_j q_{j,i} = \sum_{i=0}^{K-1} p_j q_{j,i} \\
&= p_j \sum_{i=0}^{K-1} q_{j,i} = p_j
\end{aligned} \tag{E-14}$$

So, the algorithm really samples from the required probability distribution  $p$ . ■

From the above theorem, given, a proposal transition matrix  $q$ , one can sample from a complex probability distribution  $p$ , the only requirement is that  $q$  is an irreducible transition matrix.

On the other hand,  $p$  is the stationary probability distribution, meaning, that the population sampled from algorithm 5 converges, as the iterations grow, to a population sampled directly from  $p$ . The amount of samples needed for convergence, depends on the quality of the transition matrix  $q$  in aliasing the desired probability distribution  $p$ .

An indication that the approximation is converging fast is the acceptance ratio  $\alpha$ : high acceptance ratios throughout the simulation, means fast convergence.

Theorem E.2.2, together with the next theorem, makes it possible to numerically compute expectation values for complex probability distributions.

**Theorem E.2.3** (Birkoff Theorem: discrete case). *Let  $X$  be a random variable, with values in  $\Omega = \{0, \dots, K - 1\}$ , and probability distribution  $p$ . Let a real function  $f$  over  $\Omega$ . And, say that  $X_1, X_2, \dots$  is an irreducible Markov Chain with stationary distribution  $p$ . Then, the expected value of the random variable  $f(X)$  is given by the formula:*

$$\frac{1}{N} \sum_{n=1}^N f(X_n) = \sum_{i=0}^{K-1} f(i)p(i) \quad (\text{E-15})$$

On the above equation, the left-hand side denotes an average of  $f$  over the population sampled by the Markov Chain, while the right-hand side is the expectation of the function  $f$  over the random variable  $X$ .

So, given a complex probability distribution  $p$ , one devises a Markov Chain using theorem E.2.2. The expectation of any integrable function can be computed using population mean over the chain sampled sequence.

There are many flavours of MCMC, each one using different of type of transition matrix  $q$  ((68) gives a broad review). The one used in this thesis is known as independent sampler Metropolis-Hastings:

**Definition E.2.1** (Independent sampler Metropolis Hastings). *Say  $X$  is a random variable with probability distribution  $p$ . Say,  $q_{i,j} = q_j$  is a transition matrix, which is independent on the initial state. Given that the following restriction holds  $q_j = 0 \implies p_j = 0$  (irreducible transition matrix over the exact posterior probability space), then, the MCMC acceptance ratio will be given by:*

$$\alpha_{i,j} = \min \left\{ \frac{p_j q_i}{p_i q_j}, 1 \right\} \quad (\text{E-16})$$

In this case, if  $q = p$ , then the acceptance ratio will be 1. Thus, the acceptance ratio, in the case of independent sampler MH algorithm, is a measure of similarity of the proposed distribution  $q$  and the true distribution  $p$ .

## **F**

### **Inversion Figures**

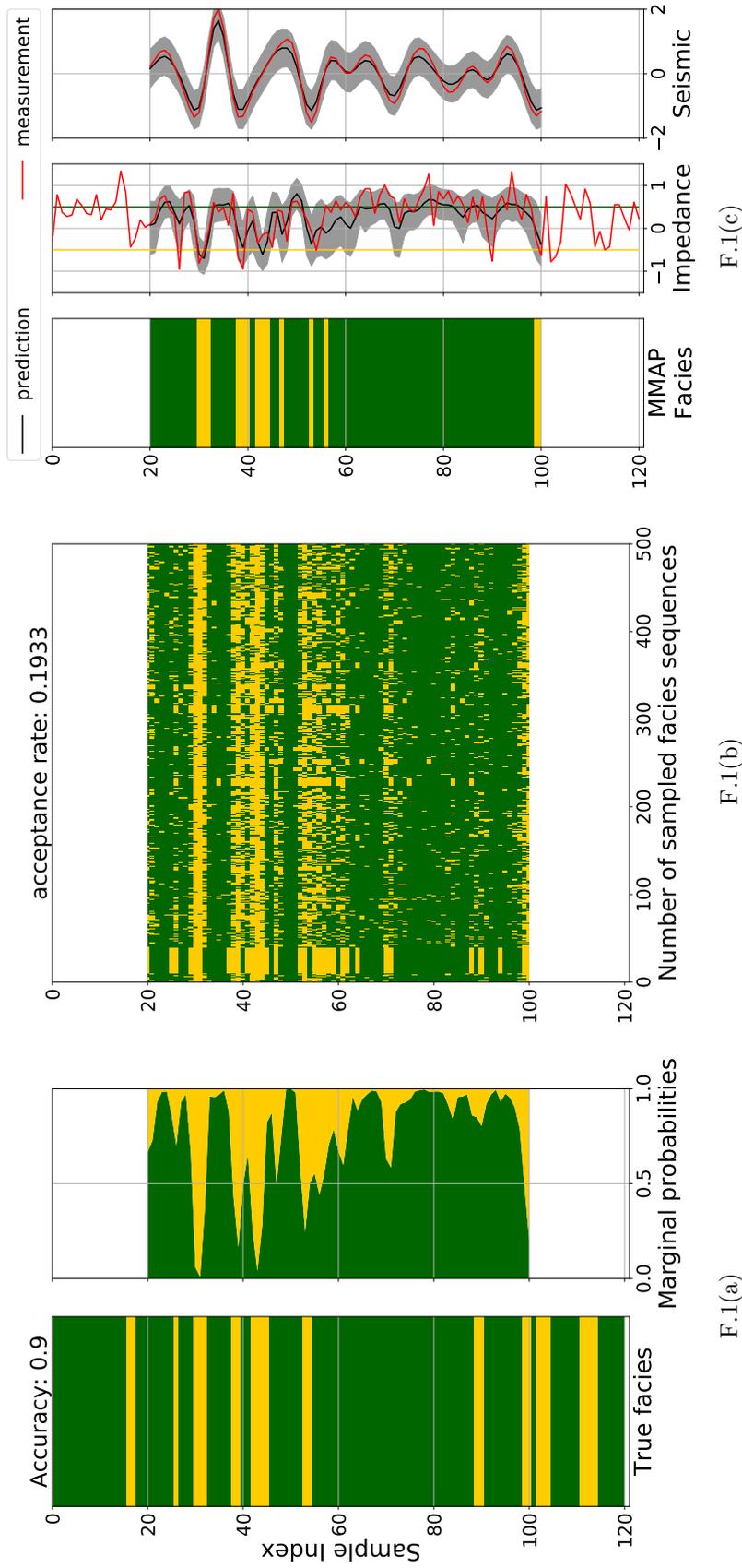


Figure F.1: The graphs show, the inversion results for the base case. F.1(a), shows the true *facies* and the marginal posterior probabilities, per *facies*. F.1(b) shows 500 *facies* profiles sampled from the posterior distribution, stacked side-by-side. F.1(c) shows the joint impedance inversion, and the forward seismic modelling, as a sanity check.

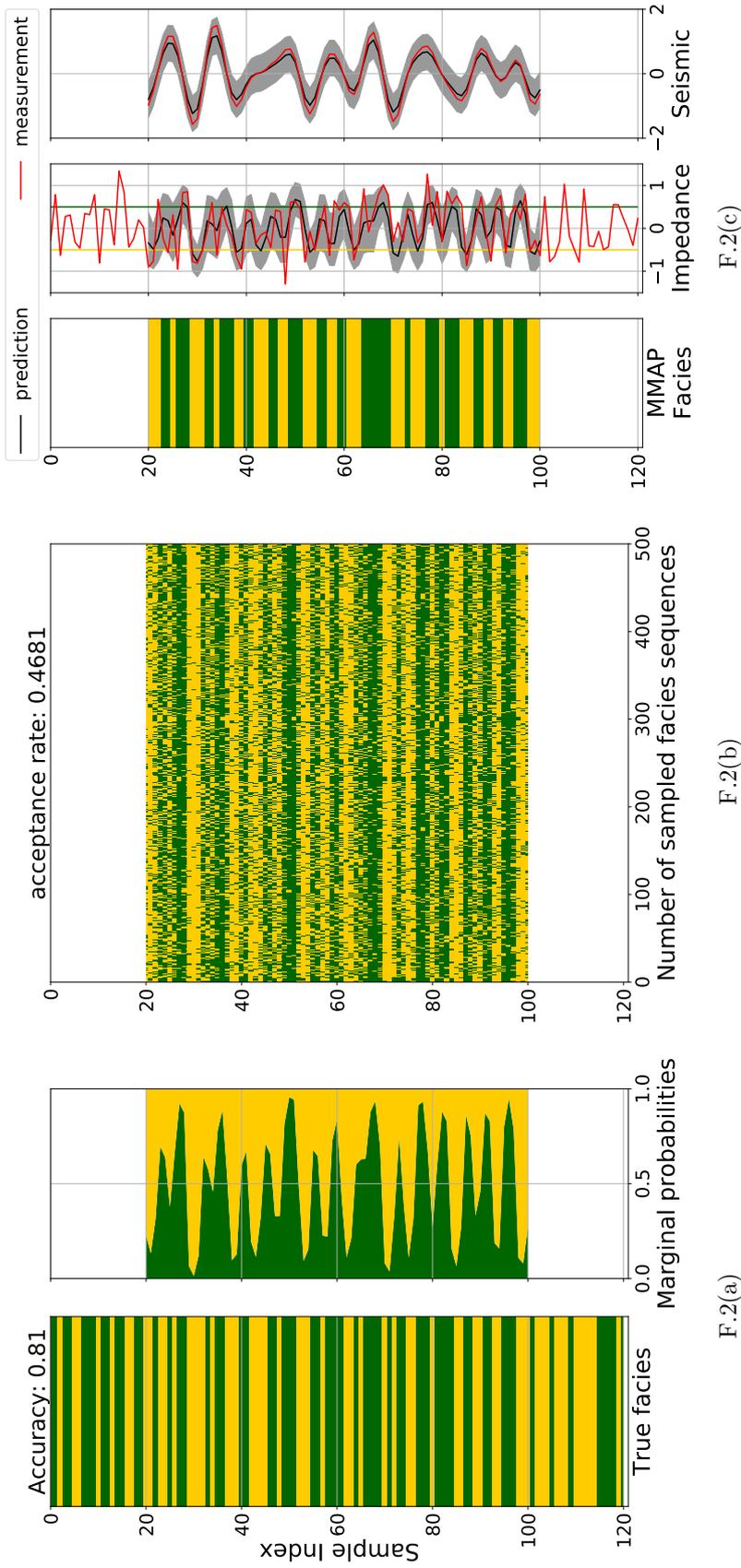


Figure F.2: The graphs show, the inversion results for the equiprobable case. F.2(b), shows the true *facies* profile and the posterior marginal probabilities, per *facies*. F.2(a) shows the 500 *facies* profiles sampled from the posterior distribution, stacked side-by-side. F.2(c) shows the joint impedance inversion, and the forward seismic modelling, as a sanity check.

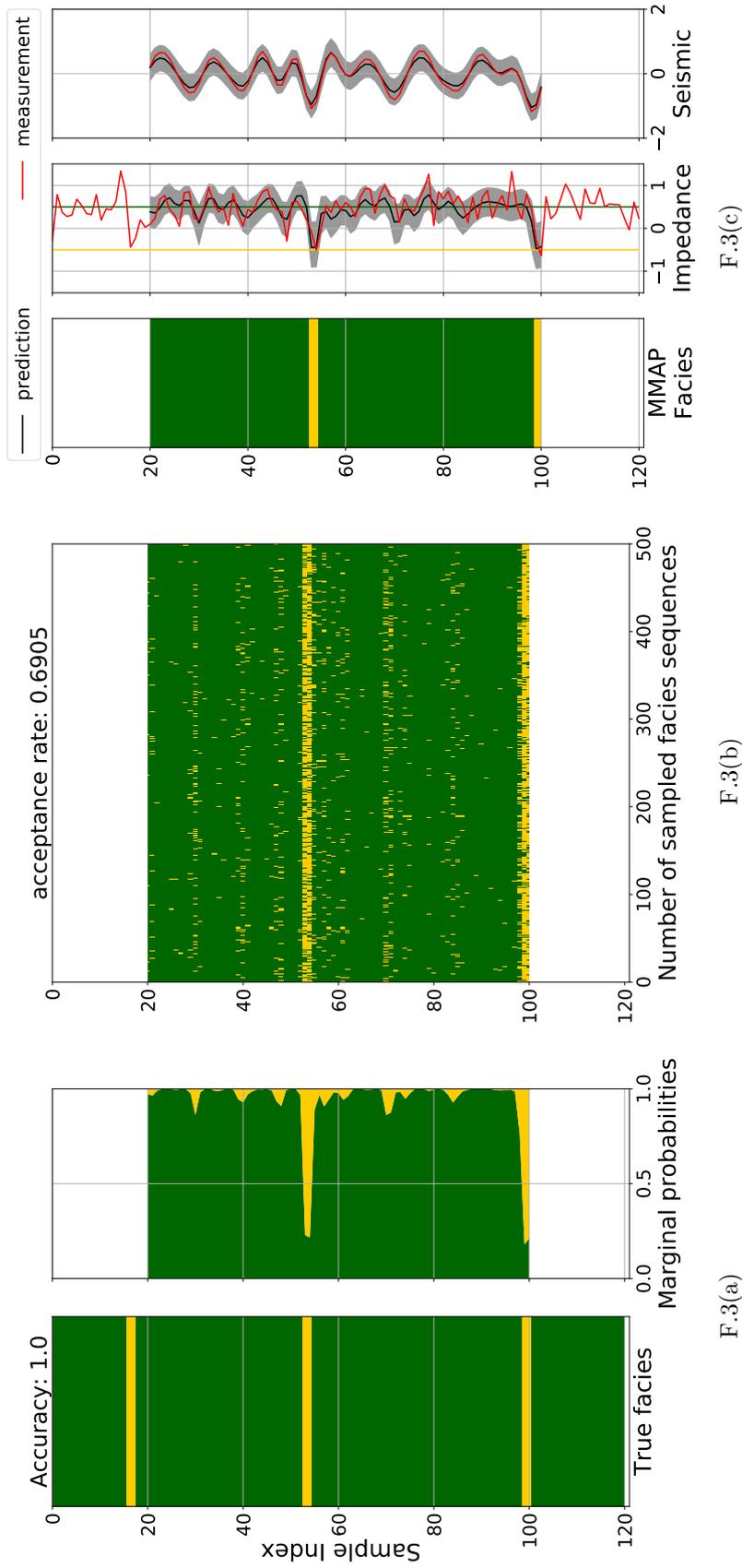


Figure F.3: The graphs show, the inversion results for the highly unbalanced case. F.2(b), shows the true *facies* profile and the posterior marginal probabilities, per *facies*. F.2(a) shows the 500 *facies* profiles sampled from the posterior distribution, stacked side-by-side. F.2(c) shows the joint impedance inversion, and the forward seismic modelling, as a sanity check.

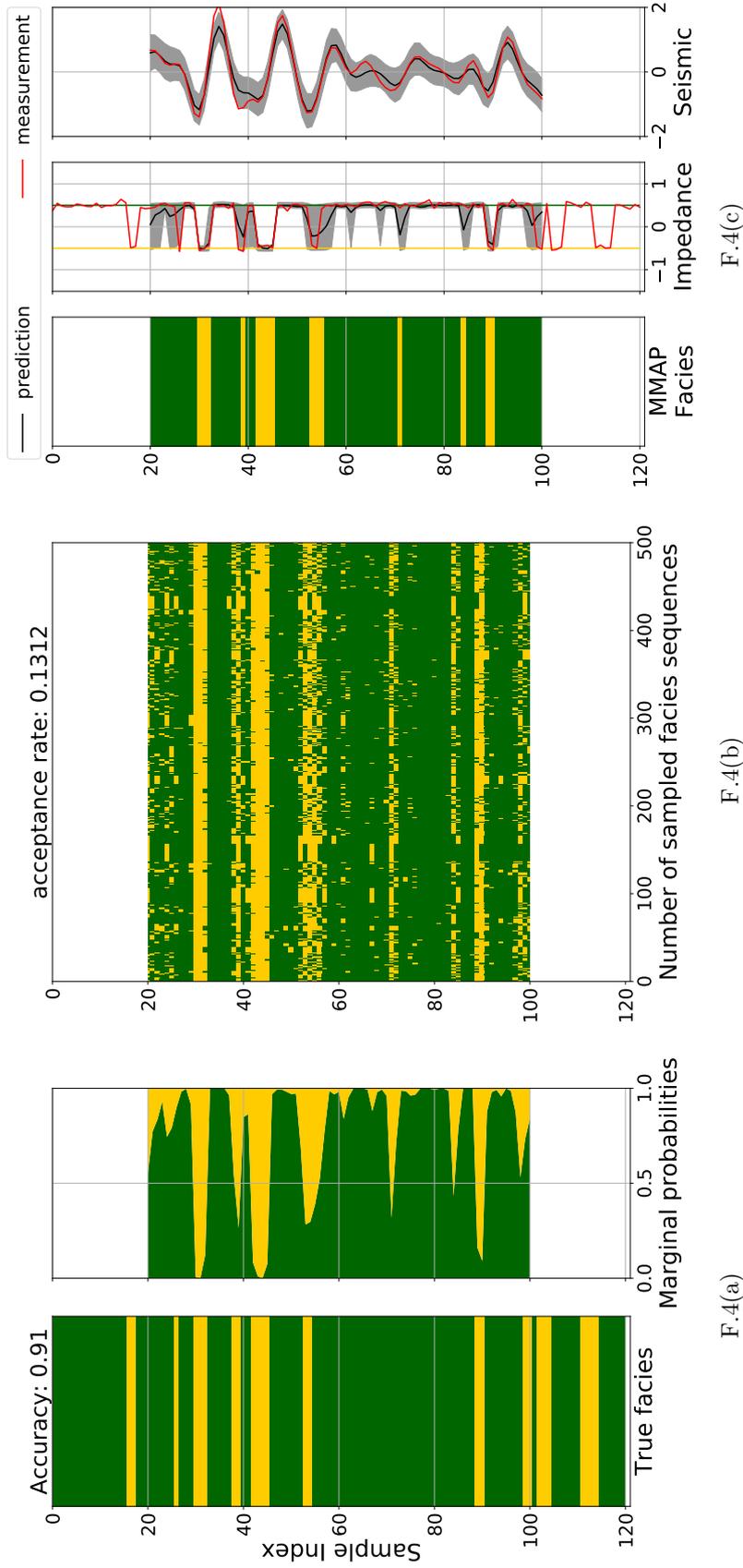


Figure F.4: The graphs show, the inversion results for the small impedance variability case. F.4(a) shows the 5000 *facies* profiles sampled from the posterior distribution, stacked side-by-side. F.4(b), shows the corresponding marginal probabilities, per *facies*. And F.4(c) shows the joint impedance inversion, and the forward seismic modelling, as a sanity check.

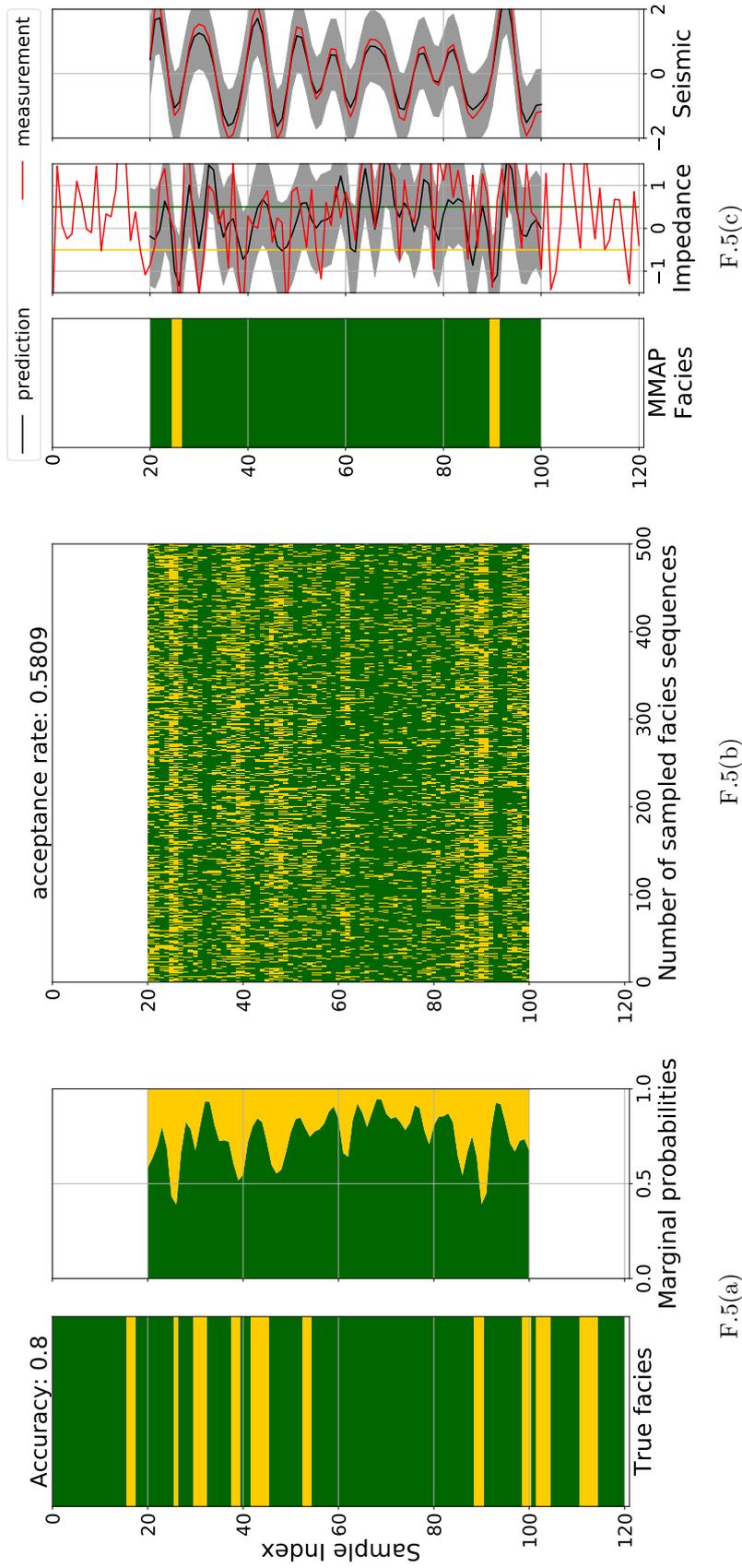


Figure F.5: The graphs show, the inversion results for the big impedance variability case. F.5(a) shows the 5000 *facies* profiles sampled from the posterior distribution, stacked side-by-side. F.5(b), shows the corresponding marginal probabilities, per *facies*. And F.5(c) shows the joint impedance inversion, and the forward seismic modelling, as a sanity check.

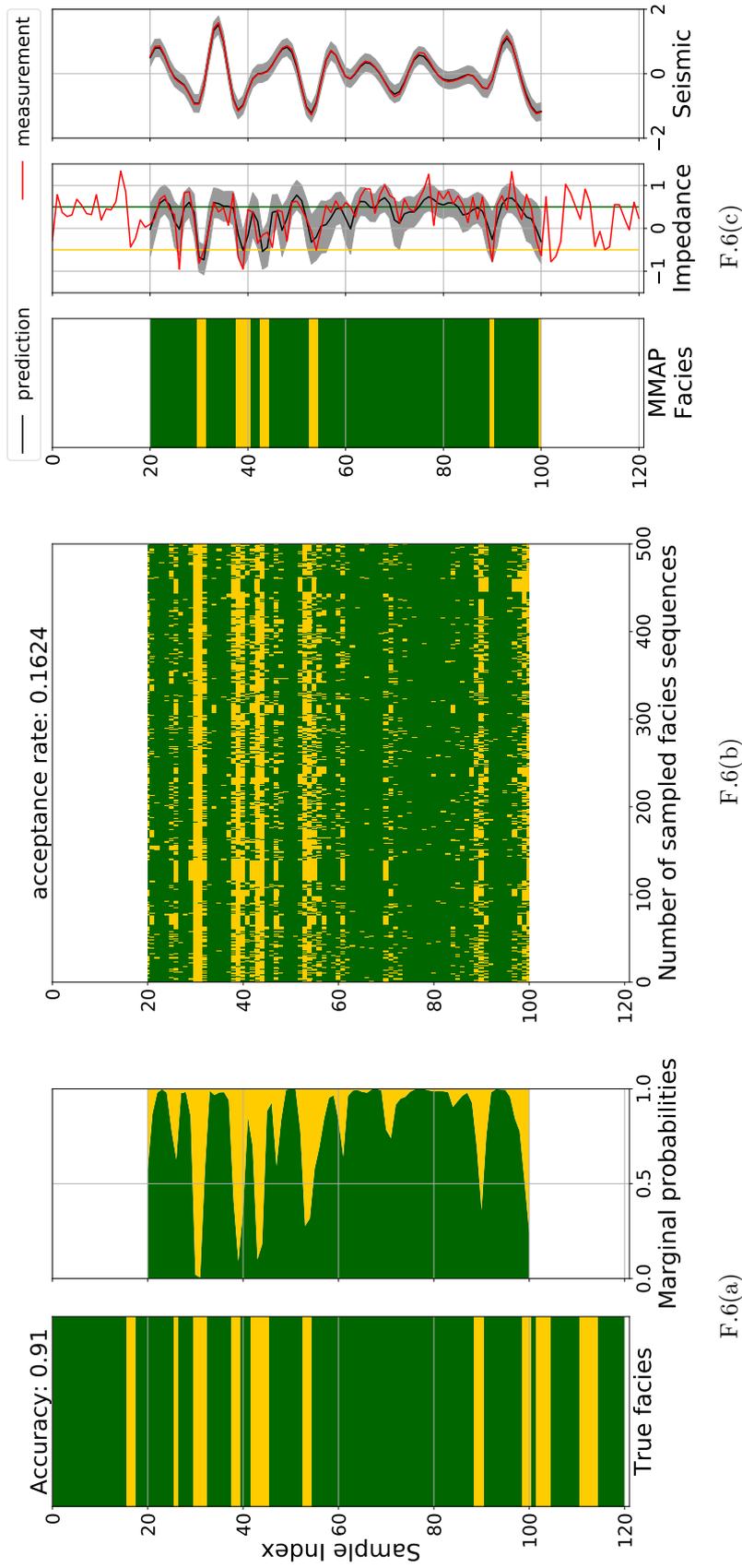


Figure F.6: The graphs show, the inversion results for the small acquisition error case. F.6(a) shows the 5000 *facies* profiles sampled from the posterior distribution, stacked side-by-side. F.6(b), shows the corresponding marginal probabilities, per *facies*. And F.6(c) shows the joint impedance inversion, and the forward seismic modelling, as a sanity check.

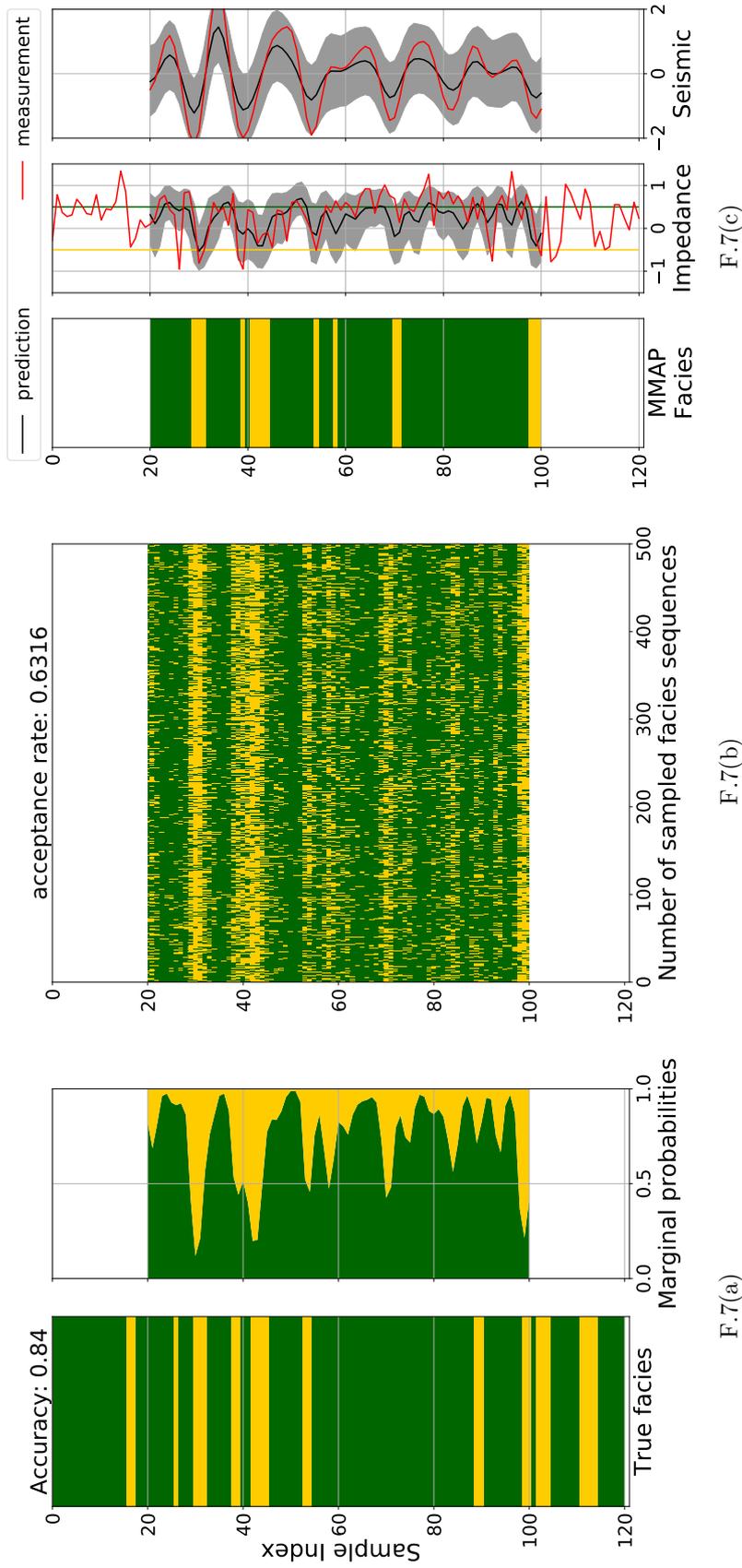


Figure F.7: The graphs show, the inversion results for the high acquisition error case. F.7(a) shows the 5000 *facies* profiles sampled from the posterior distribution, stacked side-by-side. F.7(b), shows the corresponding marginal probabilities, per *facies*. And F.7(c) shows the joint impedance inversion, and the forward seismic modelling, as a sanity check.

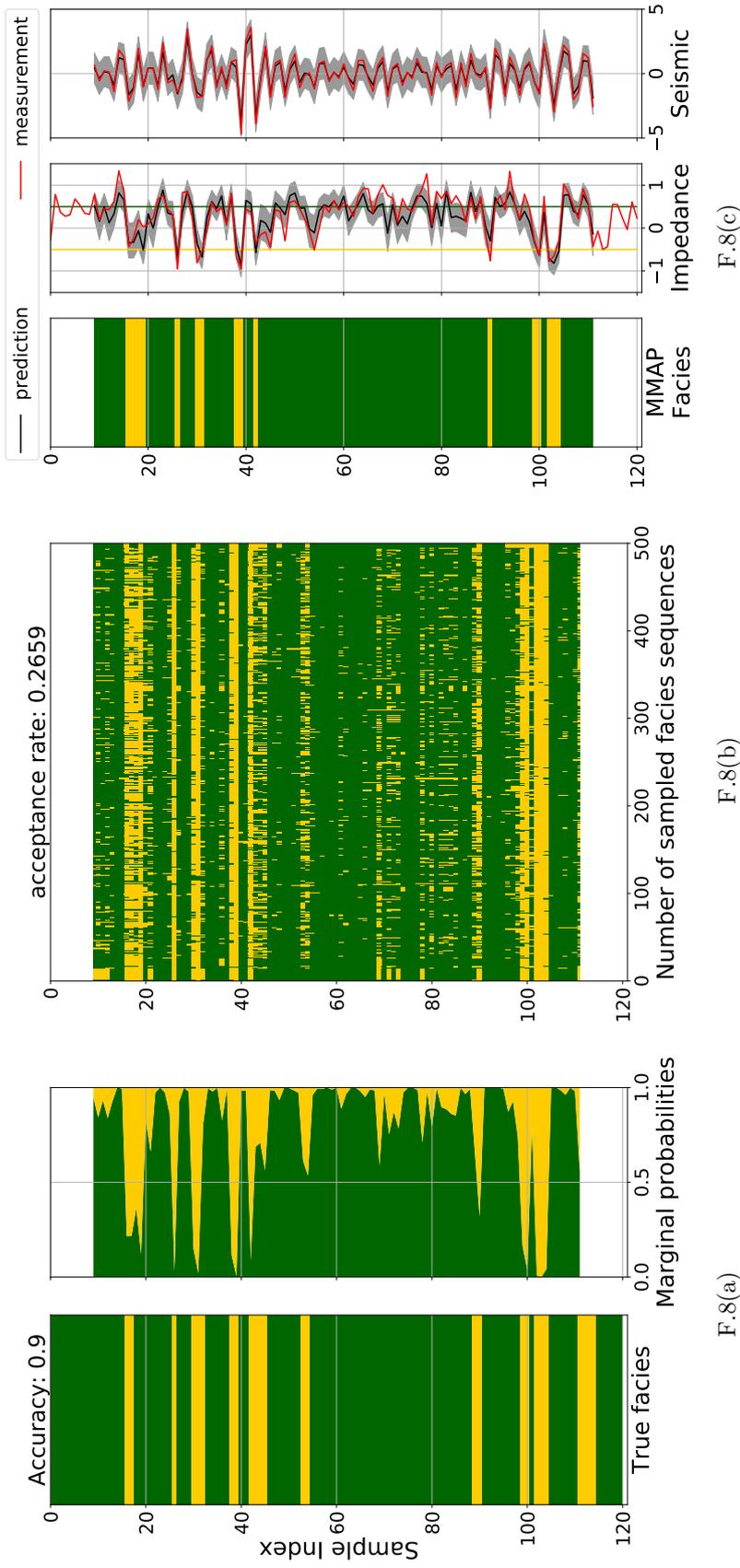


Figure F.8: The graphs show, the inversion results for the high resolution wavelet case. F.8(a) shows the 5000 *facies* profiles sampled from the posterior distribution, stacked side-by-side. F.8(b), shows the corresponding marginal probabilities, per *facies*. And F.8(c) shows the joint impedance inversion, and the forward seismic modelling, as a sanity check.

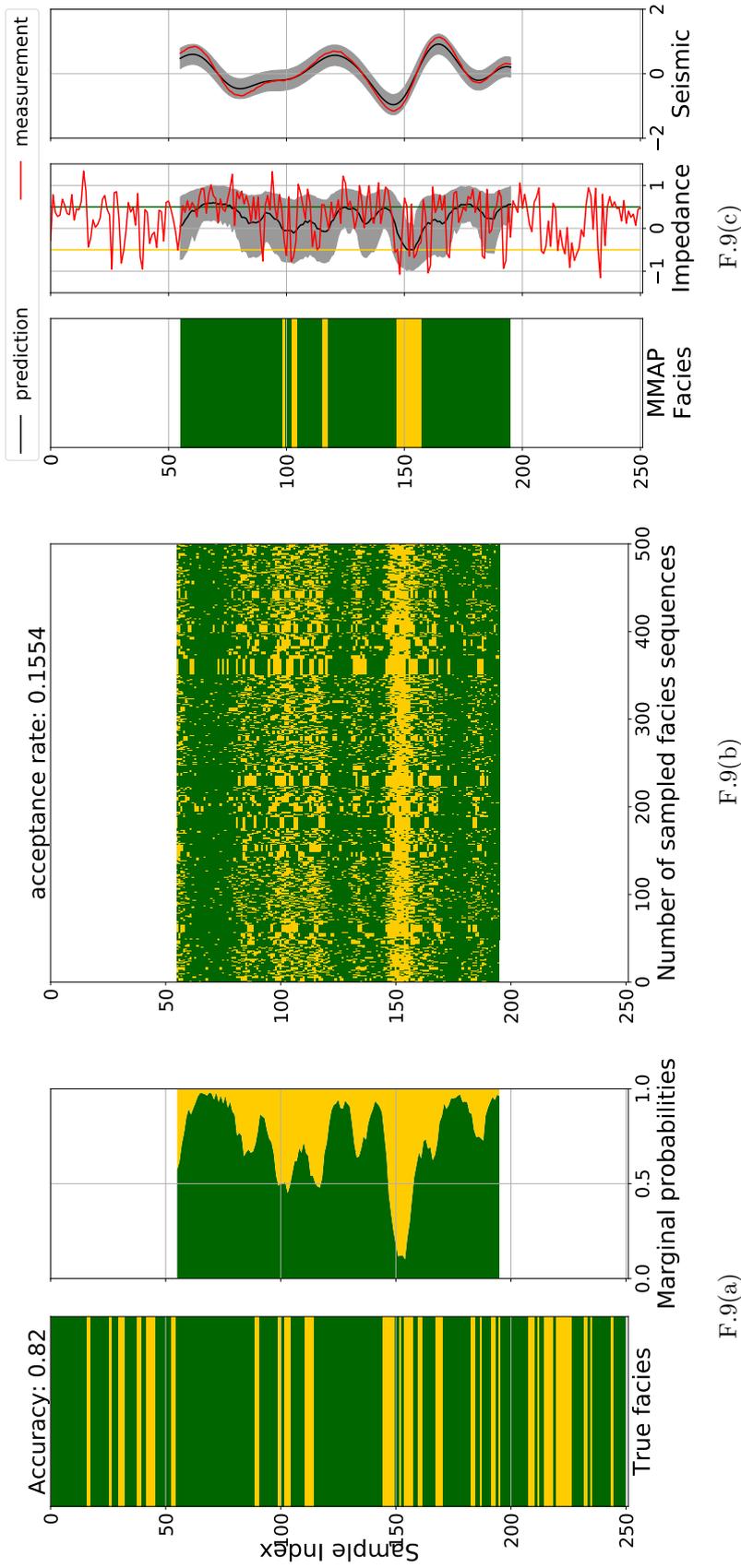


Figure F.9: The graphs show, the inversion results for the low resolution wavelet case. F.9(a) shows the 5000 *facies* profiles sampled from the posterior distribution, stacked side-by-side. F.9(b), shows the corresponding marginal probabilities, per *facies*. And F.9(c) shows the joint impedance inversion, and the forward seismic modelling, as a sanity check.