NTNU – Trondheim
Norwegian University of
Science and Technology

Prediction of Lithology/Fluid Classes from Petrophysical and Elastic Observations

Elisabeth Straume

Master of Science in Physics and Mathematics
Submission date: June 2012
Supervisor: Jo Eidsvik, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Preface

This thesis is a result of the course TMA 4905 Master thesis in Statistics, at the Norwegian University of Science and Technology(NTNU) in spring 2012. This is the last and final subject at the study program Master of Science in Industrial Mathematics.

This work is a combination of two disciplines, statistics and geophysics. This task was first presented to me by my supervisor, Associated Professor Jo Eidsvik. I have no earlier experience from geophysics, so writing this master's thesis has given me a lot of new insight to this field.

I would like to thank my supervisor Jo Eidsvik for his guidance and assistance through this last semester. His ideas and opinions have been very helpful. I would also like to thank my classmates for all good times during the last five years.

Trondheim, June 2012

Elisabeth Straume

Abstract

The objective of this study is to classify lithology/fluid(LF) variables along depth profiles. The classification is done by a Bayesian inversion method to obtain the posterior probability density functions(PDFs) for the LF classes at every depth, given data in form of petrophysical variables or elastic properties. In this way we determine the most probable lithology/fluid profile. A stationary Markov chain prior model will be used to model the continuity of the LF classes a priori. The likelihood relates the LF classes to data. A statistical rock-physics forward model is used to relate the petrophysical variables to elastic attributes.

This will be done for synthetic test data inspired by a North Sea sandstone reservoir and for real test data in form of a well log from the North Sea. Data for the synthetic case is either the petrophysical variables or the elastic properties. For the real data is only the elastic properties considered.

Samandrag

Formålet med denne oppgåva er å klassifisera litologi/fluid(LF)-variablar langs profiler i djupna. Klassifiseringa er gjort ved bruk av ein bayesiansk inversjonsmetode for å oppnå posterior sannsynstettleiksfunksjonane(PDF) for LF-klassane for kvart djup, gitt data i form av petrofysiske variabler eller elastiske eigenskapar. På denne måten kan vi bestemma den mest sannsynlege LF-profilen. Ei stasjonær markovkjede vil bli brukt som prior model til å modellera kontinuiteten av LF-klassane a priori. Likelihooden relaterer LF-klassane til data. For å relatera dei petrofysiske variablane til dei elastiske attributane vil vi bruka ein bergartsfysisk framover modell.

Dette vil bli gjort for syntetiske test data inspirert av eit sandsteinreservoar frå Nordsjøen og for reelt test data i form av ein brønnlogg frå Nordsjøen. Data for det syntetiske tilfellet er enten dei petrofysiske variablane eller dei elastiske eigenskapane. For det reelle tilfellet er kun dei elastiske eigenskapane betrakta.

Contents

Notation

Table 1: Notation used in the thesis

Variable	Description
x_t	Lithology/fluid(LF) class at depth t
t	Depth index
$\boldsymbol{x}$	LF profile
$\boldsymbol{m}$	The petrophysical variables
$\boldsymbol{m_t}$	The petrophysical variables at depth t
ϕ	Porosity, a petrophysical variable
C	Clay content, a petrophysical variable
$\boldsymbol{y}$	The elastic properties
$\boldsymbol{y_t}$	The elastic properties at depth t
V_P	Pressure velocity, an elastic property
V_S	Shear velocity, an elastic property
ρ	Density, an elastic property
K_{HM}	Bulk modulus based on Hertz-Mindlin grain contact theory
G_{HM}	Shear modulus based on Hertz-Mindlin grain contact theory
n	Coordination number
ϕ_0	Critical porosity
P	Effective pressure
ν	Grain Poisson's ratio
g	The gravity constant
Z	A given depth
ρ_b	Bulk density
ρ_{fl}	Fluid density
K_{mat}	Matrix elastic bulk modulus
G_{mat}	Matrix elastic shear modulus
K_C	The bulk modulus of wet clay
G_C	The shear modulus of wet clay
K_q	The bulk modulus of quartz
G_q	The shear modulus of quartz
K_{dry}	Dry bulk modulus
G_{dry}	Dry shear modulus
K_{sat}	Saturated rock bulk modulus
G_{sat}	Saturated rock shear modulus
ρ_{cl}	The density to the respectively LF class
ρ_c	The density for wet clay
ρ_q	The density for quartz
$\mathbf{P}$	Markov chain
$\boldsymbol{d}$	Generic notation for a data variable
N	Number of observations
p($\cdot$)	Probability density function(pdf)
p($\cdot \mid \cdot$)	Conditional pdf

Chapter 1

Introduction

Predicting lithology/fluid(LF) characteristics is important both in the exploration and the development of petroleum reservoirs. LF characteristics are predicted based on general geological experience and seismic data. That a variety of LF characteristics may result in identical seismic data makes this classification problem a challenging task and actually an illposed inverse problem. This thesis infers LF classes along a vertical earth profile. Vertical coupling of the LF classes along the profile is modeled using a Markov chain prior model, which entails that the probability of a LF class at a certain depth, given all LF classes below it, is dependent only on the LF class immediately below.

The inversion is based on a Bayesian framework, where prior knowledge about the LF characteristics is combined with information contained in the observed data. The Bayesian framework is commonly used to invert LF characteristics from seismic data. Larsen et al. (2006) propose an integrated LF inversion method based on a Markov chain model to infer LF classes from seismic prestack data. The objective of Bayesian inversion is to assess the posterior model. The prediction of LF characteristics is done along 1D profiles. Rimstad et al. (2010) predict LF characteristics in a spatial setting with a Markov random field prior model in 2D and rock physics depth trends.

Grana and Della Rossa (2010) apply a sampling algorithm to produce a priori joint PDFs of the petrophysical variables, and pressure- and shear-wave velocities using petrophysical forward relations and a priori information about LF classes. The use of statistics in rock-physics has lately become a more frequent area. In Avseth et al. (2005) deterministic models are used in combination with statistical models to build relations between elastic properties and reservoir attributes and to quantify uncertainty.

The objective of this study is on predictions of LF variables based on data by including prior information. We demonstrate this approach based on two different synthetic test datasets and a real dataset in form of a well log from the North Sea. The synthetic test data is either the petrophysical variables, which in this thesis are considered to be clay content C and porosity ϕ, or the elastic properties, which are considered to be Pressure(P)- and Shear(S) velocity, V_P and V_S, and density ρ. An

isotropic, elastic medium is completely described by these three elastic parameters. We will use $\boldsymbol{m}$ to indicate the petrophysical variables, and $\boldsymbol{y}$ to indicate the elastic properties for all depth t=$1, \ldots, N$. The LF characteristics are discretized into a LF class x_t at each t, that corresponds to the seismic sampling. The complete LF profile $\boldsymbol{x}$ is represented by a vector of discrete LF classes. To link the LF classes of interest and the available set of data, the elastic properties are defined for each t along the profile. We will use $\boldsymbol{m_t}$ and $\boldsymbol{y_t}$ to indicate respectively the petrophysical variables and elastic properties at depth t.

The posterior model cannot be computed analytical, but it can be assessed by an algorithm extracting probabilities for states in a hidden Markov model(HMM) given a dataset $\boldsymbol{d}$. A HMM is a Markov model where the system is being modeled to be a Markov process with hidden states and the data are indirect observations of these hidden variables. This algorithm is called the forward-backward algorithm and the performance of this algorithm will be studied with use of different scoring rules. When we consider the elastic properties $\boldsymbol{y_t}$ we use a statistical rock-physics forward model, which relates the petrophysical variables to the elastic attributes and link the LF classes to the elastic attributes.

This thesis is organized as follows. In Chapter 2 we present the statistical rock-physics model linking LF classes to petrophysical variables, and petrophysical variables to elastic attributes. In Chapter 3 we start by describing a hidden Markov model, before we present the forward-backward algorithm and an example of this algorithm. In Chapter 4 is the forward-backward algorithm used on synthetic test data and in Chapter 5 is this algorithm used on a well log to predict the posterior PDFs of the LF classes for every depth and determine the most probable LF profile.

Chapter 2

Statistical rock-physics model

In this chapter we will first discuss the statistical model in Chapter 2.1. The petrophysical forward model linking rock class to the response variables is explained in Chapter 2.2.

2.1 Stochastic model

The main objective of this study is to classify a lithology/fluid profile for each set of data. The lithologies consist of mixtures of clay and quartz which are classified as sand or shale, the fluids are gas, oil, and brine. The LF profile $\boldsymbol{x}$ and data $\boldsymbol{d}$ are connected by use of Bayesian inversion.

In a Bayesian inversion setting is the complete setting represented by the posterior probability density function(PDF) $p(\boldsymbol{x}|\boldsymbol{d})$. The posterior model can according to Bayes's theorem be expressed as

$$p(\boldsymbol{x}|\boldsymbol{d}) = \text{const} \cdot p(\boldsymbol{d}|\boldsymbol{x})p(\boldsymbol{x}), \tag{2.1}$$

where 'const' is a normalizing constant which usually is very difficult to compute, $p(\boldsymbol{d}|\boldsymbol{x})$ is the likelihood function and p($\boldsymbol{x}$) is the prior PDF. All PDFs are multivariate since they are defined on vectors of random variables. The probability of any LF combination along the profile, given a dataset $\boldsymbol{d}$, can be determined from the posterior in equation 2.1.

The likelihood function $p(\boldsymbol{d}|\boldsymbol{x})$ ties the LF profile $\boldsymbol{x}$ to the dataset $\boldsymbol{d}$. It contains forward models that represents the physical relation between $\boldsymbol{x}$ and $\boldsymbol{d}$. Details about the likelihood function $p(\boldsymbol{d}|\boldsymbol{x})$ are provided in Chapter 4. In the case where the elastic properties are used as the dataset $\boldsymbol{d}$ we use rock-physics relations in the likelihood function, this relations will be explained in the next subchapter. The prior p($\boldsymbol{x}$) represents knowledge about the LF profile $\boldsymbol{x}$. The prior model in this thesis is represented by a stationary Markov chain prior model for the LF profile $\boldsymbol{x}$, details are provided in Chapter 3. This choice of prior model makes it possible to define an efficient algorithm to approximate the posterior PDF, details about this algorithm are provided in Chapter 3.

We need a link between the LF profile $\boldsymbol{x}$ and the data $\boldsymbol{d}$, which is either the elastic properties or the petrophysical variables, to compute the likelihood function $p(\boldsymbol{d}|\boldsymbol{x})$. This link is given in Figure 2.1. The arrows represent dependence. The LF profile $\boldsymbol{x}$ is dependent of the elastic properties $\boldsymbol{y_t}$ either direct or indirect over the petrophysical variables $\boldsymbol{m_t}$, where t=1,...,N is a depth index.

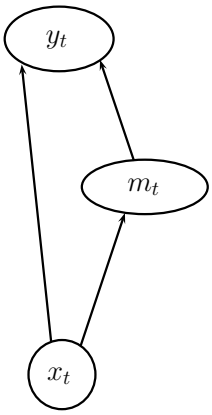


Figure 2.1: The link between the elastic properties $\boldsymbol{y}_t$, the petrophysical variables $\boldsymbol{m}_t$ and the LF class x_t at depth t.

2.2 Petrophysical model

In this subchapter is important rock-physics formulas mention for completeness, but main focus is on establishing a forward model to use in the likelihood $p(\boldsymbol{d}|x)$.

The rock-physics model is a set of equations that transforms petrophysical variables in elastic attributes. This rock-physics model is the stiff-sand model based on Hertz-Mindlin grain-contact theory. The stiff-sand model was selected because it is appropriate to describe a well-consolidated sand, as is typical in the North Sea. In shale, the effective porosity is near zero, so the rock-physics model reduces to the computation of velocities and density of a matrix made of wet clay, by means of a Voigt-Reuss Hill average, and we obtain a good approximation of the velocities in shale (Grana and Della Rossa, 2010).

The bulk modulus K_{HM} and shear modulus G_{HM} of a dry rock, when we assume that the sand frame is a dense random pack of identical spherical grains, are defined as

$$K_{HM} = \left[\frac{n^2(1-\phi_0)^2 G_{\text{mat}}^2 P}{18\pi^2(1-\nu)^2}\right]^{1/3}$$

and

$$G_{HM} = \frac{5-\nu}{10-5\nu}\left[\frac{3n^2(1-\phi_0)^2 G_{\text{mat}}^2 P}{2\pi^2(1-\nu)^2}\right]^{1/3},$$

where G_{mat} is the grain shear modulus. These modulus are subject to an effective pressure P with a critical porosity ϕ_0, and a coordination number n, an average number of contacts per grain, and the grain Poisson's ratio ν. Each of the involved variables are explained next.

The critical porosity is the point where a porous material, such as a sandstone, changes from being the load-bearing phase to becoming suspended in a fluid(Mavko et al., 2003). The critical porosity ϕ_0 used is 0.4.

The average number of contacts per grain depends on the porosity. This relationship between coordination number and porosity has been approximated by the following empirical equation (Avseth et al., 2005)

$$n = 20 - 34\phi + 14\phi^2.$$

From this formula we see that the coordination number n decrease with increasing porosity ϕ. The Poisson's ratio can be expressed in terms of the grain bulk modulus, K_{mat}, and the grain shear modulus, G_{mat}

$$\nu = \frac{3K_{mat} - 2G_{mat}}{6K_{mat} + 2G_{mat}}.$$

Effective pressure P versus depth is obtained with the following formula

$$P = g \int_0^Z (\rho_b - \rho_{fl}) dz,$$

where g is the gravity constant, and ρ_b and ρ_{fl} are the bulk density, and the fluid density, respectively at a given depth, Z. For the bulk density we use the quartz density ρ_q and for the fluid density we use the brine density ρ_{brine}. These values are given in Table 2.1. From this formula we naturally see that the pressure P will increase with the depth Z. In a small depth section this is not very prominent, and we will consider Z as a constant. We have set Z equal to 2200m.

Matrix elastic moduli are obtained by Voigt-Reuss-Hill averages for a matrix made of two component, wet clay and quartz (Mavko et al., 2003) given as

$$K_{\text{mat}} = \frac{1}{2}\left(\frac{CK_c + (1-C)K_q}{1-\phi} + \frac{1-\phi}{\frac{C}{K_c} + \frac{1-C}{K_q}}\right)$$

and

$$G_{\text{mat}} = \frac{1}{2}\left(\frac{CG_c + (1-C)G_q}{1-\phi} + \frac{1-\phi}{\frac{C}{G_c} + \frac{1-C}{G_q}}\right),$$

where C is the clay content, ϕ is the effective porosity, and K_c, G_c, K_q and G_q are the bulk and shear moduli of wet clay and quartz, given in Table 2.1. For effective porosity values between zero and the critical porosity, ϕ_0, this model connect the solid-phase elastic moduli K_{HM} and G_{HM} of the dry rock at porosity ϕ_0, by interpolating these two end members at the intermediate effective-porosity values by means of the modified Hashin-Shtrikman upper bound (Hossain et al., 2011):

$$K_{\text{dry}} = \left[\frac{\phi/\phi_0}{K_{HM} + \frac{4}{3}G_{\text{mat}}} + \frac{1-\phi/\phi_0}{K_{\text{mat}} + \frac{4}{3}G_{\text{mat}}}\right]^{-1} - \frac{4}{3}G_{\text{mat}},$$

$$G_{\text{dry}} = \left[\frac{\phi/\phi_0}{G_{HM} + \frac{1}{6}\xi G_{\text{mat}}} + \frac{1-\phi/\phi_0}{G_{\text{mat}} + \frac{1}{6}\xi G_{\text{mat}}} \right]^{-1} - \frac{1}{6}\xi G_{\text{mat}}, \quad \xi = \frac{9K_{\text{mat}} + 8G_{\text{mat}}}{K_{\text{mat}} + 2G_{\text{mat}}}.$$

Gassman's equation are used for calculating the effect of fluid on velocities using matrix and fluid properties:

$$K_{\text{sat}} = K_{\text{dry}} + \frac{\left(1 - \frac{K_{\text{dry}}}{K_{\text{mat}}}\right)^2}{\frac{\phi}{K_{fl}} + \frac{1-\phi}{K_{\text{mat}}} - \frac{K_{\text{dry}}}{K_{\text{mat}}^2}},$$

$$G_{\text{sat}} = G_{\text{dry}},$$

where K_{fl} is the fluid bulk modulus, which is either the bulk modulus of gas, oil or brine depending of the LF class. For the LF class shale we are using the bulk modulus of brine. The bulk moduli parameters are given in Table 2.1.

From the saturated-rock elastic moduli, we obtain the Pressure-wave and Shear-wave velocities using the following equations:

$$V_P = \sqrt{\frac{K_{\text{sat}} + \frac{4}{3}G_{\text{sat}}}{\rho}} \tag{2.2}$$

and

$$V_S = \sqrt{\frac{G_{\text{sat}}}{\rho}}. \tag{2.3}$$

The density ρ is estimated as a weighted linear average:

$$\rho = \phi\rho_{cl} + (1-\phi)C\rho_c + (1-\phi)(1-C)\rho_q,$$

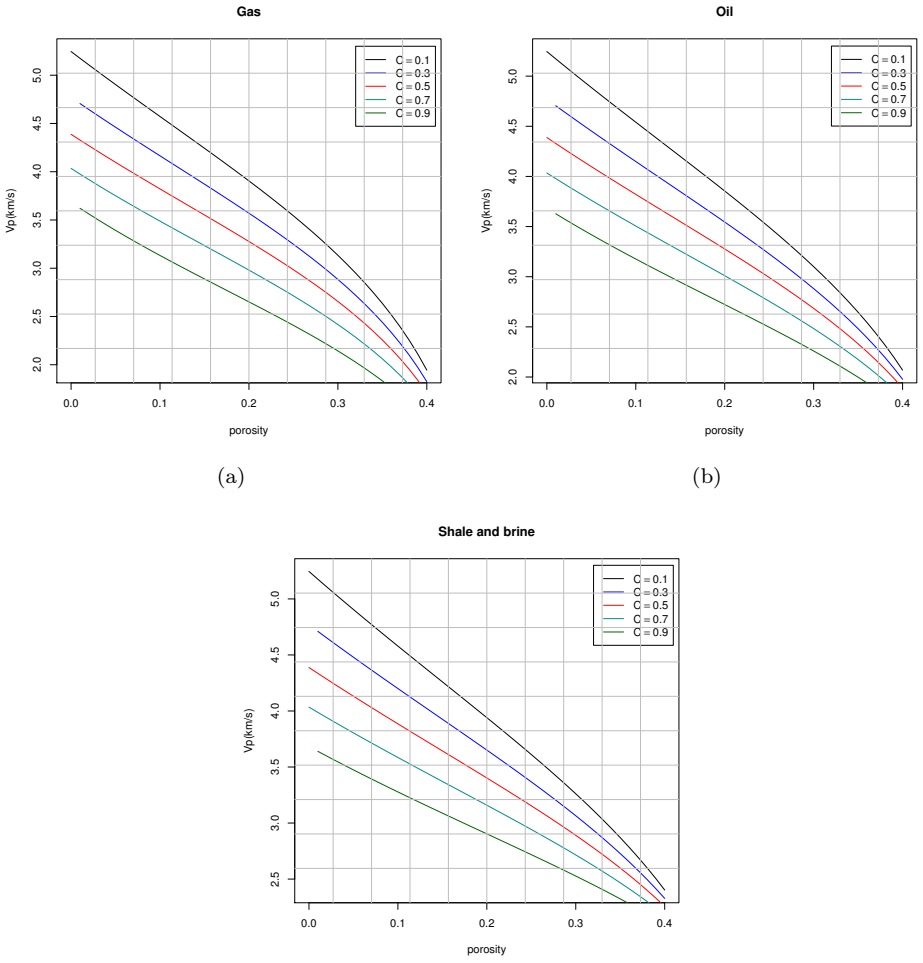
where ρ_{cl} is either ρ_{gas}, ρ_{oil} or ρ_{brine} depending on the LF class. We are using the ρ_{brine} for both the LF classes shale and brine, and ρ_c and ρ_q is respectively the density for wet clay and quartz, all given in Table 2.1. The rock-physics parameters given in Table 2.1 are used in this chapter and in Chapter 4. The rock-physics model is denoted as $f_{RPM}(C, \phi, x_t)$ for depth t.

Table 2.1: Rock-physics model parameters

Matrix components		
K_c = 20 GPa	G_c= 8 GPa	$\rho_c = 2.5 g/cm^3$
K_q= 33 GPa	G_q= 36 GPa	$\rho_q = 2.6 g/cm^3$
Fluid bulk modulus	**for different LF classes**	
K_{gas} = 0.0001 GPa	K_{oil}= 1.0 GPa	K_{brine}= 2.8 GPa
Density for	**different LF classes**	
$\rho_{gas} = 0.26 \cdot 10^{-3} g/cm^3$	$\rho_{oil} = 0.8 g/cm^3$	$\rho_{brine} = 1.1 g/cm^3$

In Figure 2.2-2.5 the rock-physics model are shown as different cross-plots. Since shale and brine are separate only through different a priori models for the petro-physical variables, they are indistinguishable in the cross-plots. For this reason we

have 3 plots. In Figure 2.2 the P-velocity is shown as a function of porosity for all LF classes, gas is given in Figure 2.2(a), oil in Figure 2.2(b), brine and shale in Figure 2.2(c), for different values of the clay content C. We can observe that the P-velocity decrease in both the case when the porosity ϕ increase and in the case when the clay content C increase as a function of porosity ϕ. The three figures seems quite similar, but we can observe that the range between the smallest and largest clay content C is smallest for the two LF classes shale and brine.



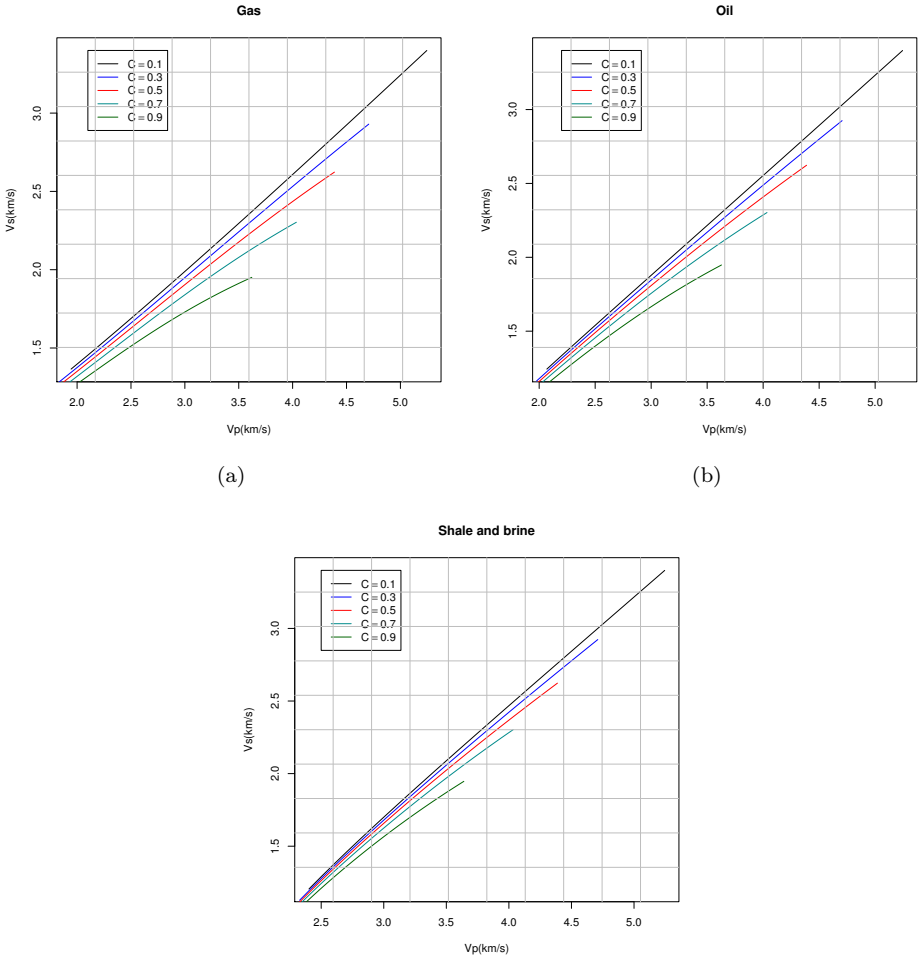
(a) (b)

(c)

Figure 2.2: P-velocity as a function of the porosity for the LF classes (a) gas, (b) oil and (c) shale and brine for different values of the clay content.

In Figure 2.3 the S-velocity is shown as a function of P-velocity for all LF classes

for different values of the clay content C. This is done for value of porosity between zero and the critical porosity equal to 0.4 in the rock-physics model. We can observe that the S-velocity increase with the P-velocity for all LF classes, and that it increase fastest for gas and slowest for shale and brine. We can also observe that the S-velocity decrease as a function of P-velocity when the clay content increase.



(a)

(b)

(c)

Figure 2.3: S-velocity as a function of P-velocity for the LF classes (a) gas, (b) oil and (c) shale and brine for different values of the clay content and porosity between zero and the critical porosity.

The P-velocity and S-velocity as functions of the density are respectively shown in Figure 2.4 and 2.5, for all LF classes with value of porosity between 0 and the

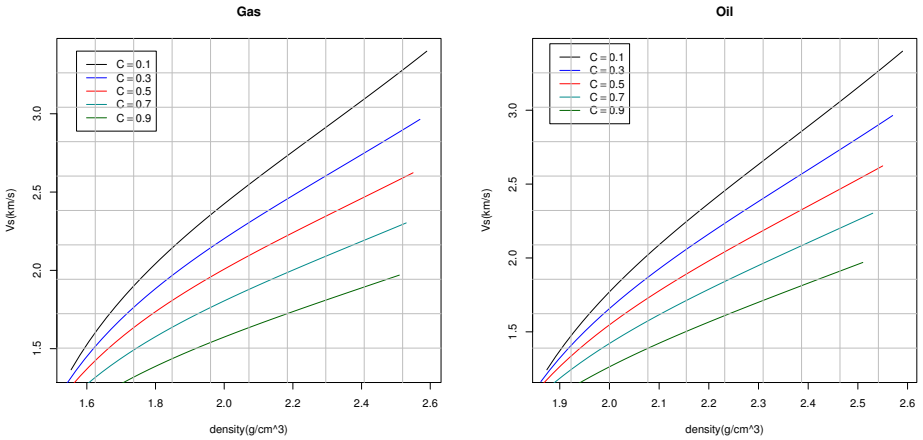
critical porosity, this is shown for different values of the clay content C. We can from Figure 2.4 observe that the S-velocity increase with the density ρ, and that it decrease as a function of density when the clay content increase for all the LF classes. We can also observe that the curve shape seems to be quite similar, but the values of density ρ varies with the LF classes. From Figure 2.5 we can observe that the P-velocity increase with the density ρ, and that it decrease as a function of density when the clay content increase for all the LF classes. We can in the same way as in Figure 2.4 observe that the curve shape seems to be quite similar, but that the values of density ρ varies with the LF classes. The density values are highest for the LF classes shale and brine, which seems correct from Table 2.1. We can observe that the P-velocity, V_P, is higher than the S-velocity, V_S, as a function of the density ρ for all LF classes. A reason for this can be that V_S only depends of G_{sat}, while V_P depends of K_{sat} in addition, see equation 2.2 and 2.3.

In Chapter 4 we will consider two different synthetic set of data

- $\boldsymbol{d_t} = \boldsymbol{m_t} = \begin{pmatrix} C \\ \phi \end{pmatrix}_t$
- $\boldsymbol{d_t} = \boldsymbol{y_t} = \begin{pmatrix} V_P \\ V_S \\ \rho \end{pmatrix}_t$

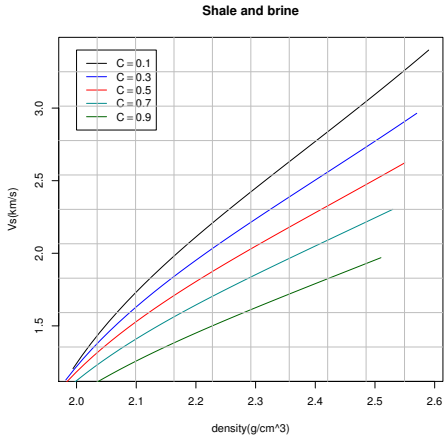
for every depth t=1, ..., N. The likelihood $d_t|x_t$ for the second set of data is Gaussian distributed with expected value from the rock-physics forward model.

In this chapter the rock-physics model has been used for constant values of the petrophysical variables porosity and clay content, but in Chapter 4 we will consider a distribution for the petrophysical variables. In that way we can separate shale and brine. Both the petrophysical variables $\boldsymbol{m}$ and the elastic properties $\boldsymbol{y}$ will be conditional on a LF reference profile $\boldsymbol{x}$ which will be simulated from a Markov chain prior model.



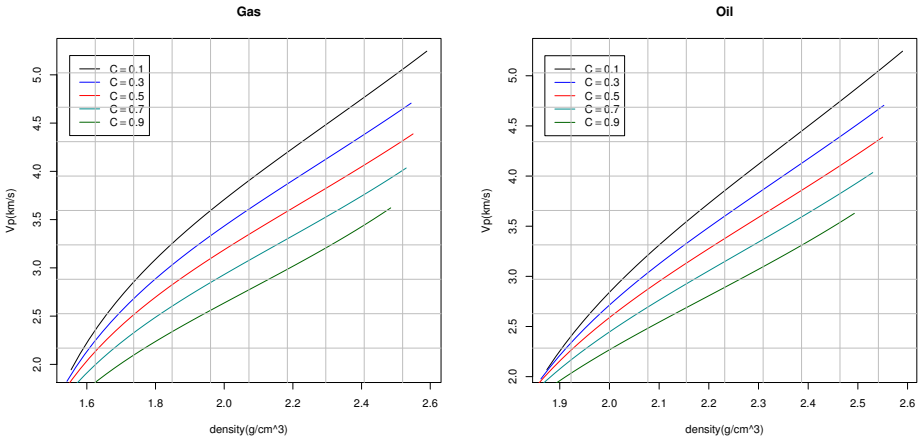
(a)

(b)



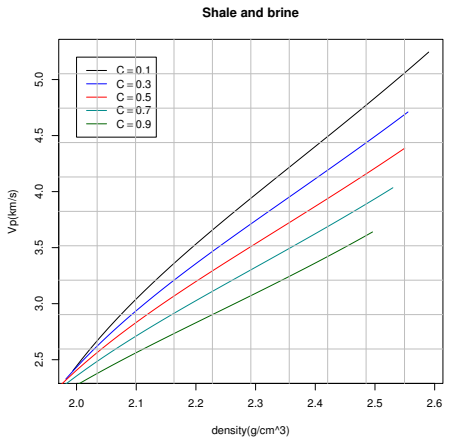
(c)

Figure 2.4: S-velocity as a function of the density ρ for the LF classes (a) gas, (b) oil and (c) shale and brine for different values of the clay content and porosity between zero and the critical porosity.



(a)

(b)



(c)

Figure 2.5: P-velocity as a function of the density ρ for the LF classes (a) gas, (b) oil and (c) shale and brine for different values of the clay content and porosity between zero and the critical porosity.

Chapter 3

Hidden Markov models

The posterior model is not analytically tractable, but it can be assessed by an algorithm computing the posterior probabilities for all hidden states variables, given a sequence of observation. This algorithm is called the forward-backward algorithm. This chapter will be started by introducing a Markov chain, before we introduce a hidden Markov model(HMM), which are needed to assess the posterior model. The forward-backward algorithm will be introduced in Chapter 3.1 and an example of this algorithm will be given in Chapter 3.2.

We consider a stochastic process $\{x_t, t = 1, \ldots, N\}$ that takes on a finite number of possible values, $x_t \in \{1, \ldots, s\}$, where s is the number of states. If $x_t = i$, then the process is said to be in state i at depth t. We suppose that whenever the process is in state i, there is a fixed probability p_{ij} that it will next be in state j. We suppose that

$$p(x_{t+1} = j | x_t = i, x_{t-1} = i_{t-1}, \ldots, x_1 = i_1) = p_{ij}, \quad p_{ij} > 0, \quad \sum_j p_{ij} = 1 \ \forall i$$

for all states $i_1, \ldots, i_{t-1}, i, j$ and for all t ≥ 1. Such a stochastic process is known as a Markov chain (Ross, 2007). For a Markov chain, the conditional distribution of any future state x_{t+1} given the past states $x_1, \ldots, x_{t-1}$ and the present state x_t is independent of the past states and depends only on the present state. The element of $x = (x_1, \ldots, x_N)$ follow a Markov chain with stationary transition matrix $\mathbf{P}$ and initial distribution π_0, which is often taken to be the stationary distribution of $\mathbf{P}$ (Scott, 2002).

For instance, in a petrophysical application there could be four classes. Then $x_t \in \{1, 2, 3, 4\}$, where x_t=1 means that the LF class is equal to gas, x_t=2 means equal to oil, x_t=3 means equal to brine and x_t=4 means equal to shale.

A Markov model in which the system being modeled to be a Markov process is unobserved (hidden). Hidden Markov models assume that the distribution of an observed data point d_t, $t = 1, \ldots, N$, depends on an unobserved(hidden) state x_t. In this thesis the data $\boldsymbol{d_t}$ is continuous. It can consist of the elastic parameters, $\boldsymbol{d_t} = \boldsymbol{y_t}$, or it can consist of the petrophysical variables, $\boldsymbol{d_t} = \boldsymbol{m_t}$, as introduced at the end of Chapter 2.2.

The stochastic process x_t is conditionally independent of all other variables than the previous one and all other observed data except the datapoint at depth t

$$p(x_t|x_1^{t-1}, d_1^t) = p(x_t|x_{t-1}, d_t),$$

where $x_1^{t-1} = (x_1, x_2, \ldots, x_{t-1})$ and $d_1^t = (d_1, \ldots, d_t)$. This is illustrated in Figure 3.1. The full conditional distribution of d_t is

$$p(d_t|d_{-t}, \boldsymbol{x}, P) = p(d_t|x_t),$$

where $d_{-t} = \{d_{t'} : t' \neq t\}$. This means that d_t is conditionally independent of all other missing and observed data given x_t, as illustrated in Figure 3.1. In Figure 3.1 the hidden Markov model is $\boldsymbol{x}$ shown at the bottom with observed data points $\boldsymbol{d}$. The arrows indicate dependence, d_t depends on x_t, and x_{t+1} depends on x_t.

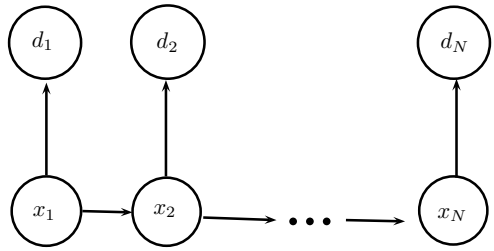


Figure 3.1: Hidden Markov model with unobserved states $\{x_1, \ldots, x_N\}$ and observed data points $\{d_1, \ldots, d_N\}$. The conditional distribution of the value at any node, given values at all others nodes, depends only on nodes which it is connected by an arrow.

3.1 The forward-backward algorithm

The forward-backward algorithm is a method to do inference for a HMM. It computes the posterior PDFs of all hidden states variables given a sequence of observations $d_1, \ldots, d_N$. For all hidden state variables $x_t, t = 1, \ldots, N$, the algorithm computes the posterior marginal distribution $p(x_t|d_1^N)$, where $d_1^N = d_1, \ldots, d_N$. The algorithm computes the values that are required to obtain the posterior marginal distribution in two passes. The first pass goes forward in time, while the second goes backward in time.

In the forward step, the algorithm computes a set of forward probabilities which for all $t = 1, \ldots, N$ provide the probability of ending up in any particular state given the first t observations, i.e. $p(x_t|d_1^t)$. This is done by recursively predicting $p(x_{t+1}|d_1^t)$ and the conditional observations $p(d_{t+1}|d_1^t)$, and updating to get $p(x_t|d_1^t)$. In the last step in the forward step we get $p(x_N|d_1^N)$.

In the backward step, the algorithm computes a set of backward probabilities $p(x_{t-1}, x_t|d_1^N)$, by starting with the last one $p(x_{N-1}, x_N|d_1^N)$ and end with $p(x_1, x_2|d_1^N)$.

From this backward step we can sum out one variable to get the posterior marginal distribution $p(x_t|d_1^N)$:

$$p(x_t = k|d_1^N) = \sum_{x_{t+1}=1}^{s} p(x_t = k, x_{t+1}|d_1^N). \tag{3.1}$$

For each step in the backward algorithm we need to calculate the posterior marginal distribution before we can calculate the next step. The forward-backward algorithm will be shown through an illustrative example in Chapter 3.2.

A HMM classifier or a estimated class at depth t can be computed by

$$\hat{x}_t = \text{argmax}_k p(x_t = k|\boldsymbol{d}). \tag{3.2}$$

Another application of the forward-backward algorithm is to maximize the marginal likelihood (MacDonald and Zucchini, 1997). The marginal likelihood can be evaluated by using the forward step:

$$p(d_1, ..., d_N) = p(d_N|d_{N-1}, .., d_1) \cdots p(d_2|d_1)p(d_1),$$

The forward-backward algorithm can be translated into upward/downward recursion when applied to a seismic profile through a target zone. This is useful since the Markov transitions naturally mimic the geological deposition over time, and the petrophysical properties of the subsurface. For instance, a gas-zone is terminated by an upper shale cap-rock, while oil being lighter than brine is never below a brine-filled sandstone.

To check the reliability to the forward-backward algorithm we can calculate the value of different scoring rules (Gneiting and Raftery, 2007). The scoring rules gives us a measure of the probability to match the reference profile, which is used as the truth. In this thesis we will consider score and logscore, where logscore of a HMM classifier or a estimated profile is defined as

$$\text{logscore} = \sum_t \log p(x_t = x_{\text{profile},t}|\boldsymbol{d}) \tag{3.3}$$

and score of a HMM classifier or a estimated profile is defined as

$$\text{score} = \frac{1}{N} \sum_t I(x_{\text{profile},t} = \hat{x}_t), \tag{3.4}$$

where I$(\cdot)$ is a indicator function defined as

$$I(\cdot) = \begin{cases} 1, & x_{\text{profile},i} = \hat{x}_t \\ 0, & \text{else} \end{cases}. \tag{3.5}$$

In words logscore is the sum of the logarithm to the posterior marginal distribution of the actual profile and score is a classification rule, which tells us if the estimated profile is equal to the actual profile. The higher score, the better is the classification. A score value equal to 1 means perfect classification. Logscore is always non-positive, a value equal to zero means perfect classification.

3.2 Illustrative example of the forward-backward algorithm

We will show the forward-backward algorithm through a simple example, which is not related to LF prediction. We consider a situation with a latent(hidden) Markov chain, only indirectly observed. Data d_t are modeled by

$$d_t = x_t + \epsilon_t, \qquad \epsilon_t \sim N(0,1), \qquad t = 1, \ldots, N, \tag{3.6}$$

where x_t is the hidden variable assumed to have two possible states, $x_t \in \{1, 3\}$ and N is length of the dataset which is 100.

We use a Markov chain as the prior model, given as

$$\mathbf{P} = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}.$$

We simulate a reference profile $\boldsymbol{x}$ from the prior model $\mathbf{P}$. By use of the reference profile $\boldsymbol{x}$ and equation 3.6 a set of data $\boldsymbol{d}$ is simulated. This dataset $\boldsymbol{d}$ is shown in Figure 3.2.

From the data $\boldsymbol{d}$ we can find the naive profile in this example defined as

$$\hat{x}_t^{naive} = \begin{cases} 1, & \text{if } d_t < 2 \\ 3, & \text{if } d_t > 2 \end{cases},$$

where d_t is one element of the dataset $\boldsymbol{d}$ and $t = 1, \ldots, N$. The meaning of the naive profile is illustrated in Figure 3.2, where a line separate $\hat{x}_t^{naive} = 1$ from $\hat{x}_t^{naive} = 3$.

We know from the hidden Markov model that the datapoint d_t at depth t is only dependent of the hidden variable x_t and that x_t is only dependent of the previous hidden variable x_{t-1} and the datapoint d_t. The likelihood $p(d_t|x_t)$ is known from equation 3.6 and from the Markov chain $\mathbf{P}$ we know $p(x_t|x_{t-1})$. In the forward step we start by computing the conditional distribution

$$p(x_1|d_1) = \frac{p(x_1)p(d_1|x_1)}{p(d_1)} = \frac{p(x_1)p(d_1|x_1)}{\sum_{x_1 \in \{-1,1\}} p(x_1)p(d_1|x_1)}$$

and end by computing

$$p(x_N|d_1^N) = \frac{p(x_N|d_1^{N-1})p(d_N|x_N)}{p(d_N|d_1^{N-1})} = \frac{\sum_{x_{N-1} \in \{-1,1\}} p(x_N|x_{N-1})p(x_{N-1}|d_1^{N-1})p(d_N|x_N)}{\sum_{x_N \in \{-1,1\}} p(x_N|d_1^{N-1})p(d_N|x_N)}$$

In the backward step we start by computing

$$p(x_{N-1}, x_N|d_1^N) = p(x_N|d_1^N)p(x_{N-1}|x_N, d_1^N) = p(x_N|d_1^N)\frac{p(x_N|x_{N-1})p(x_{N-1}|d_1^{N-1})}{p(x_N|d_1^{N-1})}$$

and end by computing

$$p(x_1, x_2|d_1^N) = p(x_2|d_1^N)p(x_1|x_2, d_1^N) = p(x_2|d_1^N)\frac{p(x_2|x_1)p(x_1|d_1)}{p(x_2|d_1)},$$

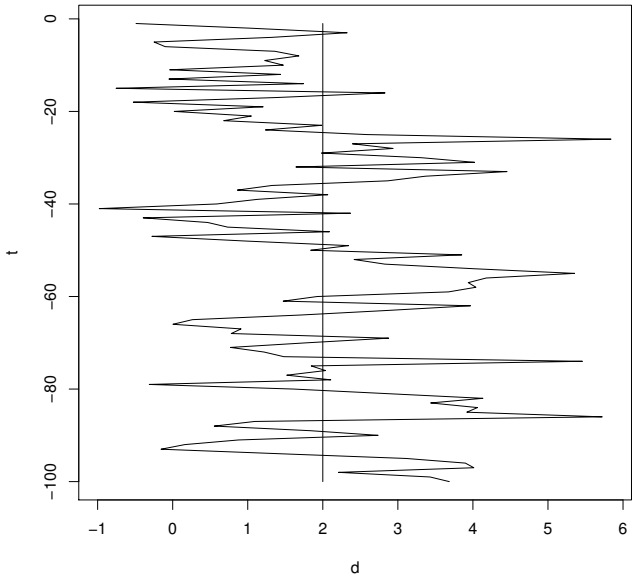


Figure 3.2: Data $\boldsymbol{d}$ used in this example, where the line separate $\hat{x}_t^{naive} = 1$ from $\hat{x}_t^{naive} = 3$.

where $p(x_2|d_1^N)$ is the posterior marginal distribution

$$p(x_2|d_1^N) = \sum_{x_3} p(x_2, x_3|d_1^N),$$

where we sum out x_3 by summing over the two classes in this example.

From the posterior marginal distribution we can calculate the estimated profile $\hat{\boldsymbol{x}}$ by using equation 3.2, where $\hat{x}_t$ is the most likely class at t, given all data. The estimated profile is a HMM classifier. Figure 3.3 shows the reference profile $\boldsymbol{x}$, the naive profile $\hat{\boldsymbol{x}}^{naive}$, the estimated profile $\hat{\boldsymbol{x}}$, and the posterior marginal distributions p($x_t = k|\boldsymbol{d}$) for $k = \{1, 3\}$ given.

We observe that the estimated profile has the same trend as the reference profile, but is not similar for all $t = 1, \ldots, N$, where N=100. The naive profile is similar to the reference profile for some t, but the naive profile jumps more. The naive profile jumps 35 times, the reference profile jumps 11 times and the estimated profile only jumps 9 times. We can also observe that for p($x_t = k|\boldsymbol{d}$) for $k = \{1, 3\}$ is the prediction nearly perfect when the reference profile and the estimated profile is similar. Perfect prediction means that the posterior marginal PDFs would be binary zero-one functions, with the value of one in the actual reference class. The posterior marginal PDFs appear smoothed across the borders where the estimated profile vary from the naive profile. Score and logscore in this example are given

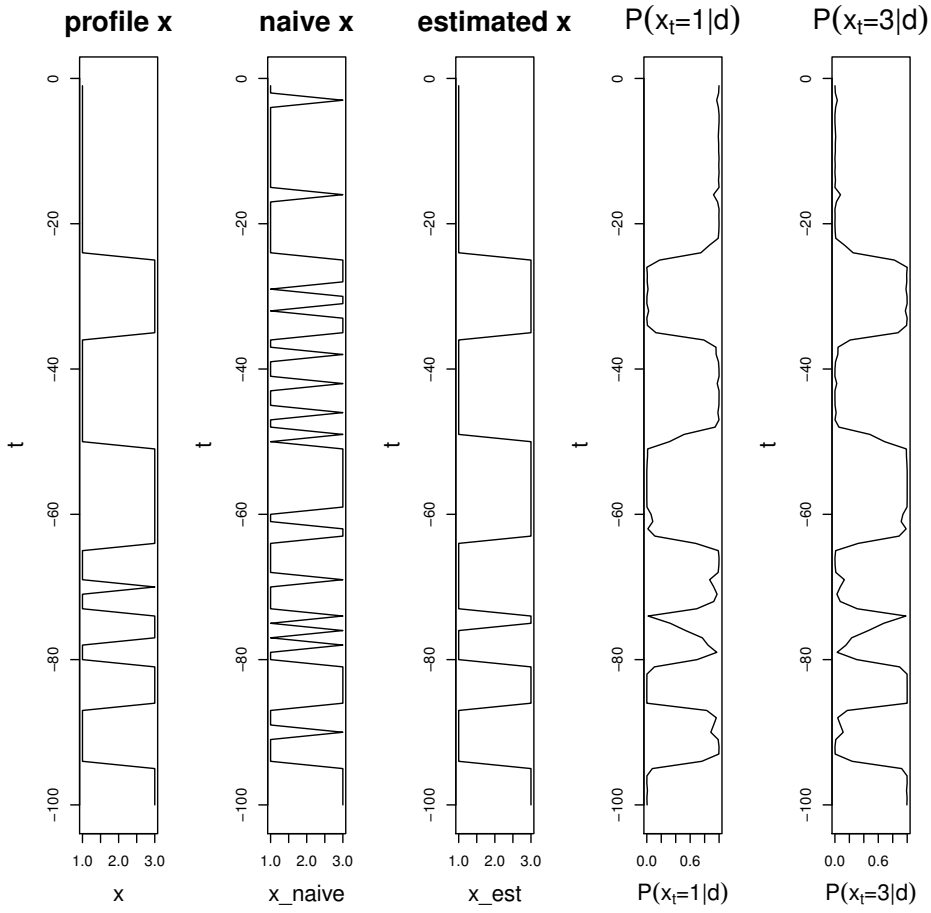


Figure 3.3: The reference profile $\boldsymbol{x}$, the naive profile $\hat{x}^{naive}$, the estimated profile $\hat{\boldsymbol{x}}$, and the posterior marginal distributions $\mathrm{p}(x_t = k|\boldsymbol{d})$ for $k = \{1, 3\}$

in Table 3.1 for the naive profile and the HMM classifier. Logscore for the HMM classifer is defined as in equation 3.3, and logscore for the naive profile is defined as

$$\text{logscore}^{\text{naive}} = \sum_t \log p(x_t = x_{\text{profile},t}|\boldsymbol{d_t}),$$

and consider only data from depth t. Score is defined for the HMM classifer as in equation 3.4, and for the naive profile is $\hat{x}_t$ exchanged with $\hat{x}_t^{\text{naive}}$.

The values for the HMM classifier, $\hat{x}$, is higher than for naive for both score and logscore. This is naturally since the HMM classifier is a better classification method than the naive classifier. The score value for the HMM classifier means that 95 % of the estimated profile is equal to the reference profile and indicate that

the prediction is good. Logscore equal to -12.13 means that the prediction is pretty good.

Table 3.1: Score and logscore for the naive and HMM classifier

	Naive	HMM classifier
Score	0.83	0.95
Logscore	-216.18	-12.13

Chapter 4

Synthetic test data

We will now study the performance of the forward-backward algorithm for classifying LF classes, given various data. Two test studies are defined that use synthetic test data, inspired by a North Sea sandstone petroleum reservoir. The first one consists of the two petrophysical variables, clay content C and porosity ϕ, and the second one consists of the three elastic properties P-velocity V_P, S-velocity V_S and density ρ. The reference reservoir contains four LF classes: gas-, oil-, and brine-saturated sandstone and shale.

The prior model for the LF classes is defined as a stationary Markov chain through the reference reservoir, with transition matrix (Larsen et al., 2006)

$$\mathbf{P} = \begin{pmatrix} 0.980 & 0 & 0 & 0.020 \\ 0.015 & 0.970 & 0 & 0.015 \\ 0.002 & 0.008 & 0.980 & 0.010 \\ 0.007 & 0.007 & 0.036 & 0.950 \end{pmatrix},$$

with rows and columns that corresponds to gas-, oil-, and brine-saturated sandstone and to shale. The stationary prior PDF $\pi_0(\boldsymbol{x})$ which represents the proportions of LF classes, can be computed from $\mathbf{P}$, $\pi_0(\boldsymbol{x})$=(0.233,0.156,0.393,0.218). The transition matrix contains several zero elements that represents impossible upward transitions, brine can never be directly above gas or oil, and only shale can be directly above gas. In Figure 4.2 and 4.5, the leftmost display presents one realization of a LF sequence that was generated from $\mathbf{P}$. This sequence is used as the LF reference profile. The proportions of LF classes in the reference profile are (0.236,0.195,0.355,0.214).

4.1 Petrophysical variables

We consider a Gaussian distribution for the petrophysical variables $\boldsymbol{m} = (\boldsymbol{m}_1, \ldots, \boldsymbol{m}_N)$ with $\boldsymbol{m}_t = (C, \phi)_t^T$ for $t = 1, \ldots, N$ and where the superscript T means transpose. The distribution is given as $\boldsymbol{m}_t = N(\boldsymbol{m}_t; \boldsymbol{\mu}(x_t), \boldsymbol{\Sigma})$ for all t where the expectation depends of the LF class x_t. We have used the expectation

$$\boldsymbol{\mu}(x_t) = \begin{cases} \begin{bmatrix} 0.3 \\ 0.35 \end{bmatrix}, & x_t = 1, 2, 3 \\ \\ \begin{bmatrix} 0.7 \\ 0.3 \end{bmatrix}, & x_t = 4 \end{cases} \tag{4.1}$$

and the covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.1^2 & 0 \\ 0 & 0.05^2 \end{bmatrix}. \tag{4.2}$$

For comparing results, we let the expectation $\boldsymbol{\mu}(x_t)$ to $\boldsymbol{m_t}$ for x_t=1,2,3 and x_t=4 go against each other, and define α to be the interval between the two expectations, as illustrated in Figure 4.1. The expectation $\boldsymbol{\mu}(x_t)$ as a function of α at depth t is defined as

$$\boldsymbol{\mu}(x_t, \alpha) = \begin{cases} \alpha \cdot \boldsymbol{\mu}_{123}^{orig}(x_t) + (1-\alpha) \cdot \boldsymbol{\mu}_4^{orig}(x_t), & x_t = 1, 2, 3 \\ \alpha \cdot \boldsymbol{\mu}_4^{orig}(x_t) + (1-\alpha) \cdot \boldsymbol{\mu}_{123}^{orig}(x_t), & x_t = 4 \end{cases}, \tag{4.3}$$

where $\boldsymbol{\mu}_{123}^{orig}(x_t)$ and $\boldsymbol{\mu}_4^{orig}(x_t)$ are the original expectations for respectively $x_t = 1, 2, 3$ and $x_t = 4$. For example if $\alpha = 1/2$ then the two expectations are similar, and if $\alpha = 1$ then the two new expectations are the same as the original expectations, and for $\alpha > 1$ will the distance between the two expectations become bigger.

We want to find the posterior marginal distribution $p(x_t = k|\boldsymbol{d})$ for all states k=1,2,3,4 and for all $t = 1, \ldots, N$, when data $\boldsymbol{d}$ is the petrophysical variables $\boldsymbol{m}$. To do this, we first simulate the LF reference profile $\boldsymbol{x}$ from $\mathbf{P}$, and simulate the data $\boldsymbol{d_t} = \boldsymbol{m_t}$ from the petrophysical distribution to $\boldsymbol{m_t}$ for all depth t. The forward-backward algorithm runs the LF prediction for this data. The entire procedure is replicated B=50 times for every α value of the expectations in equation 4.3. The code for simulation of data and for the forward-backward algorithm are written in **R** (R Development Core Team, 2011). We compute the mean value of the B posterior marginal distributions and use this to compute a HMM classifier, a estimated profile $\hat{\boldsymbol{x}}$, as in equation 3.2. Figure 4.2 shows the LF reference profile $\boldsymbol{x}$, the estimated LF profile $\hat{\boldsymbol{x}}$ and the posterior marginal distribution $p(x_t = k|\boldsymbol{d})$ for all states k.

We can observe that the estimated profile has the same trend as the reference profile $\boldsymbol{x}$, but this $\hat{\boldsymbol{x}}$ is not similar to the reference profile $\boldsymbol{x}$ for all t=$1, \ldots, N$, where N=1000. We can see that p($x_t = k|\boldsymbol{d}$) for k=4, the LF class shale, is a nearly perfect prediction. Perfect prediction means that the posterior marginal PDFs would be binary zero-one functions, with the value of one in the actual LF

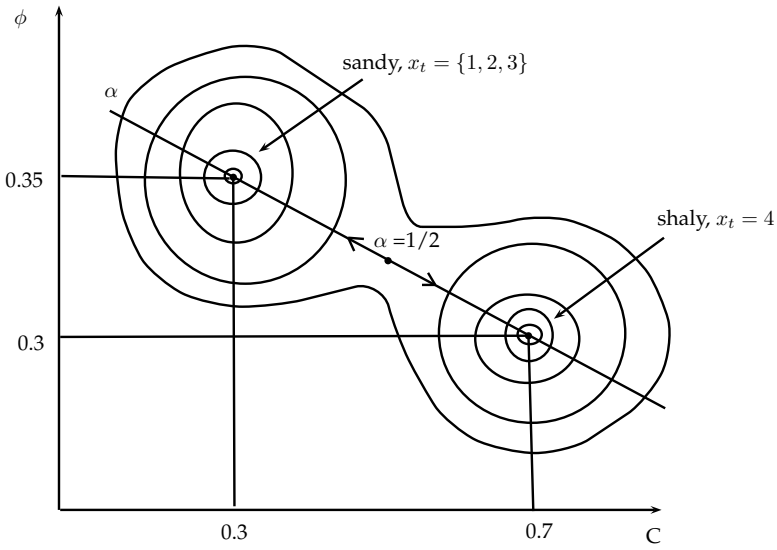


Figure 4.1: The expectation to $\boldsymbol{m_t}$, where the meaning of α is given. For α=1/2 the two expectations are similar and we are in center, α=1 represents the original expectations and for $\alpha > 1$ is the distance between the two expectations bigger.

reference class. For k=1,2,3 we can see that the posterior marginal PDFs appear smoothed across the borders where the LF reference profile is not equal to 4, which tell us that gas-, oil- and brine-saturated sandstone are more difficult to predict than shale. This is reasonable from the expectation to $\boldsymbol{m_t}$ in equation 4.1, where we cannot separate gas-, oil- and brine-saturated sandstone.

At depth t=-200 m are both the reference class x_t and the estimated class $\hat{x}_t$ equal to 4, the LF class shale. At this depth we can say for sure that the rock class is shale and the posterior marginal distribution for shale is equal to 1 and all the others posterior marginal distribution are equal to zero. This is naturally since it is easy to separate shale from sandstone from the petrophysical distribution. At depth t=-600 m is the reference class x_t equal to 2, the LF class oil, and the estimated class $\hat{x}_t$ is equal to 3, the LF class brine. The posterior marginal distribution for shale is zero, which is natural since x_t is different from 4 and it is easy to separate shale and sandstone. The highest posterior marginal distribution is the one for brine which is why $\hat{x}_t = 3$, but the posterior marginal distribution for gas and oil are also different to zero. This is because it is not easy to separate the different types of sandstone. In the next subchapter we will consider the elastic properties as the dataset and we will use the rock-physics model introduced in Chapter 2.2, which will do it easier to separate the different type of sandstone.

We want to see how good the estimated LF profile $\hat{\boldsymbol{x}}$ fits to the LF reference profile $\boldsymbol{x}$ for the different α values of the expectation. We have used values of α between 1/2 and 2. This have been tested by use of score and logscore, defined

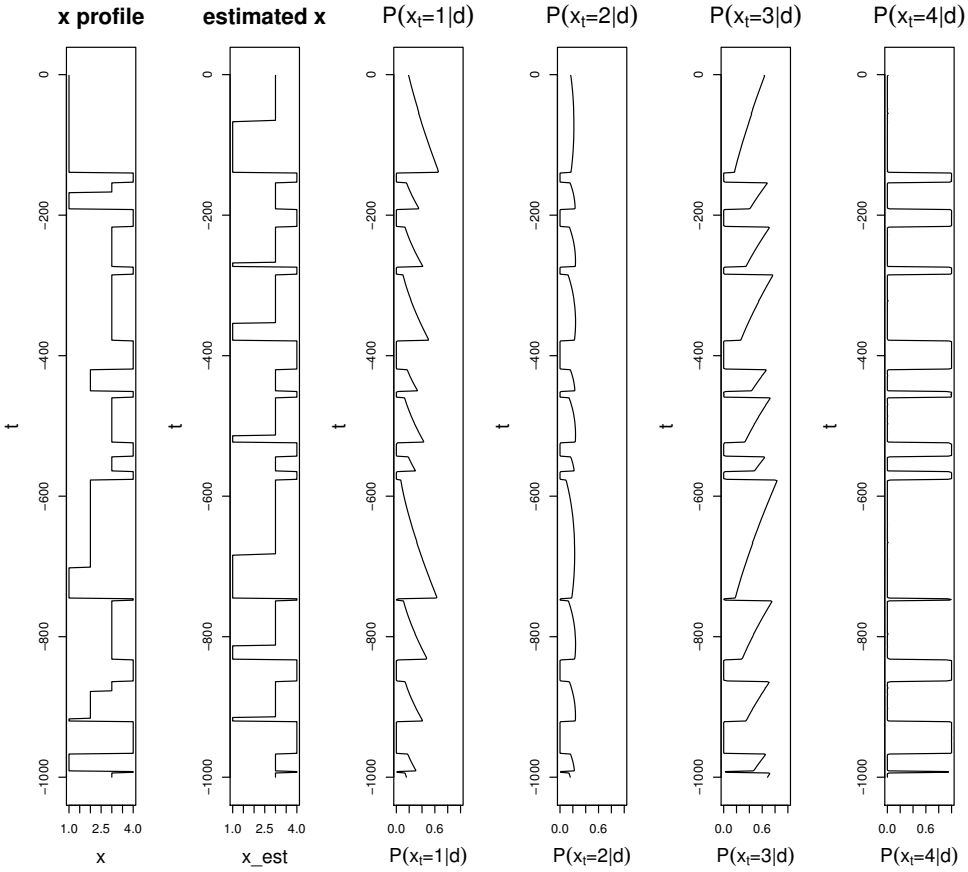


Figure 4.2: The reference profile $\boldsymbol{x}$, the estimated profile $\hat{\boldsymbol{x}}$, and the posterior marginals $\mathrm{p}(x_t = k|\boldsymbol{d})$ for $k = \{1, 2, 3, 4\}$ and all t when the petrophysical variables are used as the dataset $\boldsymbol{d}$ and α=1 so we have the original petrophysical expectations.

in formula 3.3 and 3.4. Figure 4.3 displays logscore and score as functions of α. We can observe that both logscore and score start with values equal to -1380 for logscore and 0.35 for score and seems to converge towards a value, -780 for logscore and 0.63 for score. This means that when the two expectations are similar is the score only 0.35, and when we have the original expectations have score converged to the constant value 0.63. From this we know that 63 % of the estimated profile $\hat{\boldsymbol{x}}$ is equal to the reference profile $\boldsymbol{x}$ when we have the original expectations, given in equation 4.1, or a larger distance between the two expectations. Both score and logscore indicate that the LF classification could have been better.

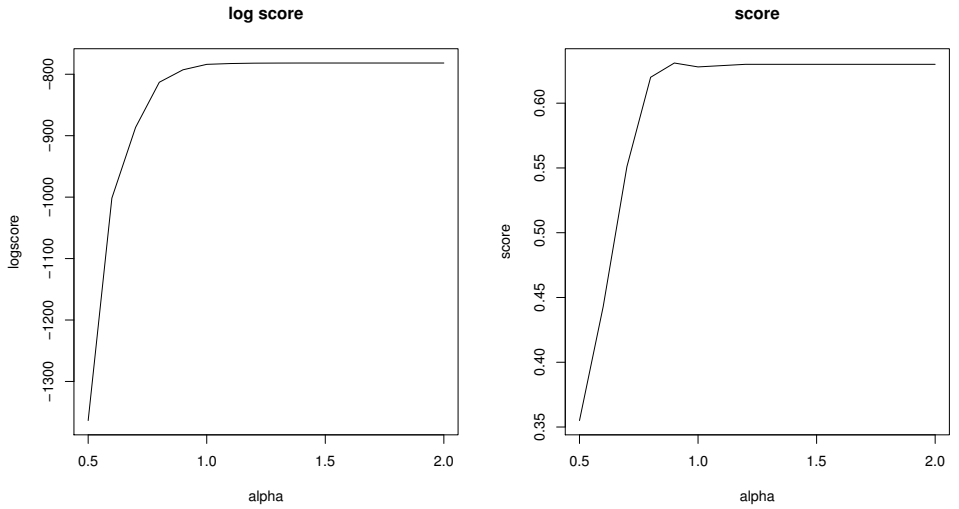


Figure 4.3: Logscore and score as functions of the distance α.

4.2 Elastic properties

We next consider a Gaussian distribution for the elastic properties $\boldsymbol{y} = (\boldsymbol{y_1}, \ldots, \boldsymbol{y_N})$, with $\boldsymbol{y_t} = (V_P, V_S, \rho)_t^T$ for t=1, ..., N and where the superscript T means transpose. The distribution is given as $\boldsymbol{y_t} = N(\boldsymbol{y_t}; f_{RPM}(c, \phi, x_t), T)$, where f_{RPM} is the rock-physics model described in Chapter 2.2 and T is random error, describing the degree of accuracy of the model. We set the random error to be $T = \tau \cdot I_3$, where I_3 is a identity matrix with size 3 and τ is a given value.

We marginalize out $\boldsymbol{m_t}$ from the model to get an approximate likelihood p($\boldsymbol{y}|\boldsymbol{x}$). This is done by linearizing the rock-physics model f_{RPM} around the expectation of the petrophysical variables

$$f_{RPM}(C_t, \phi_t, x_t) \approx f_{RPM}(\mu_C, \mu_\phi, x_t) + F_{x_t}\left(\begin{pmatrix} C_t \\ \phi_t \end{pmatrix} - \begin{pmatrix} \mu_C \\ \mu_\phi \end{pmatrix}\right), \qquad (4.4)$$

where C_t and ϕ_t are the petrophysical variables at depth t, F_{x_t} is a 3×2 matrix consisting of the derivative of f_{RPM} with respect to C_t and ϕ_t when x_t is a given class and μ_C and μ_ϕ are dependent of x_t like equation 4.1.

The expectation to $\boldsymbol{y_t}$ given x_t is

$$\begin{aligned} E(\boldsymbol{y_t}|x_t) &= E(E(\boldsymbol{y_t}|C_t, \phi_t, x_t)) \\ &= E(f_{RPM}(C_t, \phi_t, x_t)) = E\left(f_{RPM}(\mu_C, \mu_\phi, x_t) + F_{x_t}\left(\begin{pmatrix} C_t \\ \phi_t \end{pmatrix} - \begin{pmatrix} \mu_C \\ \mu_\phi \end{pmatrix}\right)\right) \\ &= f_{RPM}(\mu_C, \mu_\phi, x_t), \end{aligned}$$

and the variance to $\boldsymbol{y_t}$ given x_t is

$$\begin{aligned} Var(\boldsymbol{y_t}|x_t) &= E(Var(\boldsymbol{y_t}|C_t, \phi_t, x_t)) + Var(E(\boldsymbol{y_t}|C_t, \phi_t, x_t)) \\ &= E(T) + Var\left(f_{RPM}(\mu_C, \mu_\phi, x_t) + F_{x_t}\left(\begin{pmatrix} C_t \\ \phi_t \end{pmatrix} - \begin{pmatrix} \mu_C \\ \mu_\phi \end{pmatrix}\right)\right) \\ &= T + F_{x_t} Var\left(\begin{pmatrix} C_t \\ \phi_t \end{pmatrix} | x_t\right) F_{x_t}^T. \end{aligned}$$

Then the distribution for the elastic properties $\boldsymbol{y_t}$ given a class $\boldsymbol{x_t}$ is given as

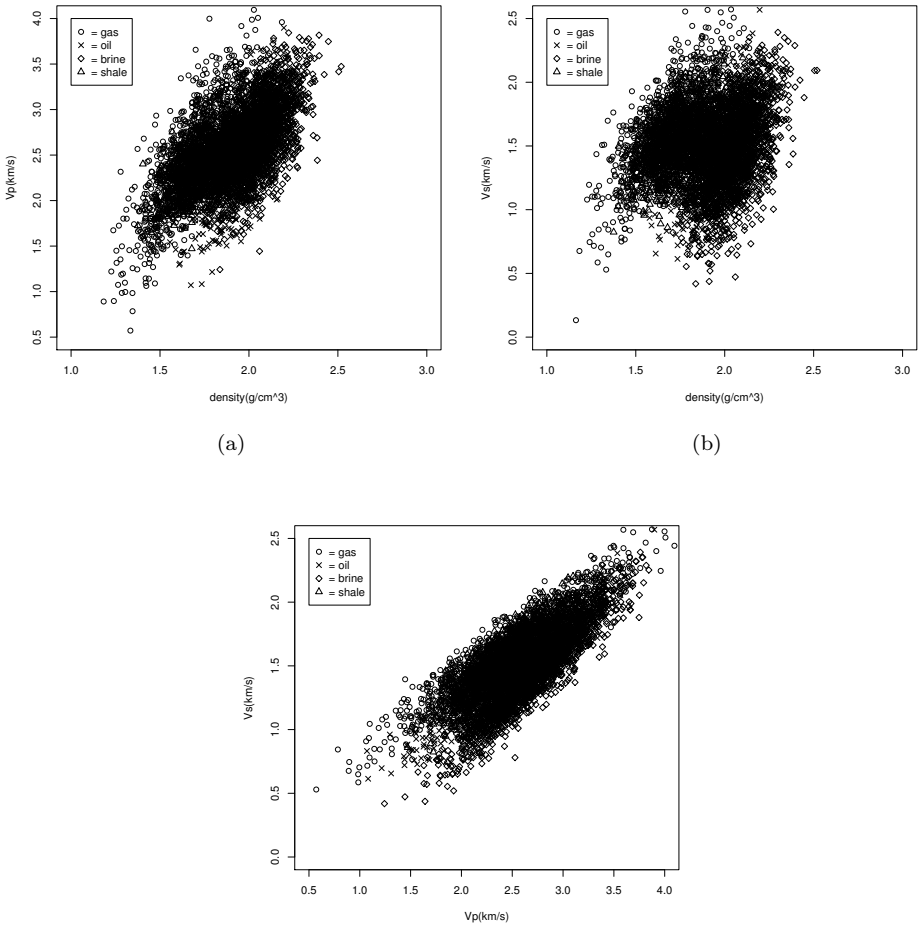
$$\boldsymbol{y_t}|x_t \sim N\left(f_{RPM}(\mu_C, \mu_\phi, x_t), T + F_{x_t} Var\left(\begin{pmatrix} C_t \\ \phi_t \end{pmatrix} | x_t\right) F_{x_t}^T\right), \tag{4.5}$$

for all depth t, and $\boldsymbol{y_t}$ is no longer dependent of the petrophysical variables, only of their distribution.

We have simulated values of the elastic properties V_P, V_S and ρ from this distribution, by first simulating the LF reference profile $\boldsymbol{x}$ from $\mathbf{P}$. Cross-plots of this values are displayed in Figure 4.4. Figure 4.4(a) and 4.4(b) shows the simulated P-velocity and S-velocity as functions of the simulated density ρ. Figure 4.4(c) shows the simulated S-velocity given as a function of the simulated P-velocity. We can observe that these figures seems to be similar to Figure 2.3-2.5, which is expected from the distribution to the elastic attributes.

To find the posterior marginal distribution $p(x_t = k|\boldsymbol{d})$ for all states k=1,2,3,4 and for all $t = 1, \ldots, N$, when data $\boldsymbol{d}$ is the elastic properties, we run the forward-backward algorithm explained in Chapter 3.1. The entire procedure is replicated B=100 times for every τ value of the likelihood variance. The code for simulation of data, for the forward-backward algorithm and for the rock-physics model are written in **R** (R Development Core Team, 2011). We compute the mean value of the B posterior marginal distributions and use this to compute a HMM classifier, a estimated profile $\hat{\boldsymbol{x}}$, as in equation 3.2. Figure 4.5 displays the LF reference profile $\boldsymbol{x}$, the estimated LF profile $\hat{\boldsymbol{x}}$ and the posterior marginal distribution $p(x_t = k|\boldsymbol{d})$ for all states k when the likelihood variance τ=0.1. We can also in this case observe that the estimated profile has the same trend as the reference profile, but is not similar for all $t = 1, \ldots, N$ where N=1000. For the values of t where the estimated LF class are similar to the reference class for a large size of t, e.g from t=-580 m to t=-750 m, is the posterior marginal PDFs nearly perfect predicted. We can see that for these points are the posterior marginal PDFs almost a binary zero-one function for all states k. When the estimated LF class are similar to the reference class for a smaller size of t, e.g from t=-420 m to t=-450 m, is the prediction of the posterior marginal PDFs worse. The posterior marginal PDFs is no longer a binary zero-one function. When the estimated LF profile and the reference LF profile differs are the posterior marginal PDFs smooth across these borders.

The posterior marginal PDF for k=1 is easy to determine, and it is almost a binary zero-one function for all depth t, except the few depths where the estimated profile differs from the reference profile. Gas saturated sandstone is easy to separate from the other LF classes in the petrophysical model since the modulus and



(a) (b)

(c)

Figure 4.4: Plot of the simulated elastic properties: (a) The simulated P-velocity as a function of the simulated density ρ. (b) The simulated S-velocity as a function of the simulated density ρ. (c) The simulated S-velocity as a function of the simulated P-velocity.

the density for gas are quite different the other moduli and densities, see Table 2.1. For k=2 is the prediction of the posterior marginal PDF a little bit worse, we have no longer a binary zero-one function for every t, only for the depths where the estimated LF class are similar to the reference class for a large size of t. Oil saturated sandstone is possible to separate through the petrophysical model, because the fluid modulus and density for oil are different from the other fluid moduli

and densities. The difference are not so great as for gas, which result in a worse prediction for oil than for gas. For k=3 and k=4 are the prediction hardest. The posterior marginal PDFs are only zero when the reference profile is equal to 1 or 2 for a large size of depth t, and it is never equal to one. It is hard to separate brine saturated sandstone from shale since we get the same value from the petrophysical model whether the LF class is shale or brine. We use the petrophysical distribution to separate shale and brine, but this is not enough and we get a bad prediction.

We again want to see how good the estimated LF profile $\hat{\boldsymbol{x}}$ fits to the reference profile $\boldsymbol{x}$. We let τ in the likelihood variance vary between 0.001 and 1. Figure 4.6 shows logscore and score as functions of τ when N=1000. We can observe that both logscore and score decrease as a function of the likelihood variance τ. Logscore starts with a value of 0 and seem to go slowly against a constant value of -790 and score starts with a value of 1 and seem to go slowly against a constant value of 0.64. This means that when the likelihood variance τ is zero is the estimated LF profile the same as the reference profile, and when the likelihood variance become greater is the estimated LF profile more different to the reference profile, which is naturally. In the case where Figure 4.5 is given, for τ=0.1, is score=0.86 and logscore= -308, which indicate a pretty good prediction.

For large τ we have the same situation as for α=1/2 in Figure 4.3 in the previous subchapter, in both cases it is difficult to separate sandstone from shale. When τ is large it is too much uncertainty that we can separate sandstone and shale and when α=1/2 are the expectation for sandstone and shale similar and we cannot separate them.

Table 4.1 compares the predicted LF profile $\hat{\boldsymbol{x}}$ and the reference LF profile $\boldsymbol{x}$, with use of a classification matrix for $\boldsymbol{x}$ versus $\hat{\boldsymbol{x}}$. In a classification matrix we can determine whether the predicted value matches the actual value. Perfect prediction would make $\hat{\boldsymbol{x}}$ identical to $\boldsymbol{x}$ and would make all non-diagonal terms in the classification matrix zero. We can observe that the prediction is good for τ=0.1, but not perfect since not all of the non-diagonal terms in the classification matrix is zero. The prediction for gas saturated sandstone is perfect and the prediction for oil saturated sandstone is nearly perfect, as seen in Figure 4.5. The prediction of brine saturated sandstone and shale are not so good, which agree with the discussion of Figure 4.5. We can from the classification matrix easy observe that 100 of the LF class in the reference class given as shale has been estimated as brine saturated sandstone.

Table 4.1: Classification matrix of $\boldsymbol{x}$ versus $\hat{\boldsymbol{x}}$ for the likelihood variance τ=0.1.

$\boldsymbol{x} \setminus \hat{\boldsymbol{x}}$	Gas sand	Oil sand	Brine sand	Shale	Sum
Gas sand	227	9	0	0	236
Oil sand	0	189	6	0	195
Brine sand	0	0	337	18	355
Shale	0	0	100	114	214
Sum	227	198	443	132	1000

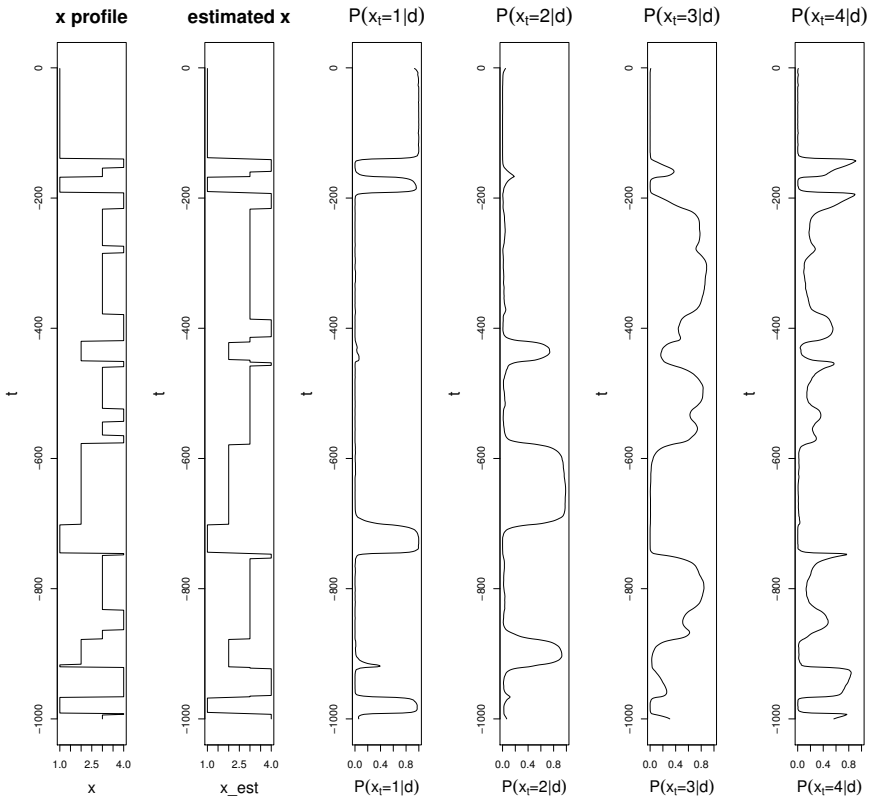


Figure 4.5: The LF reference profile $\boldsymbol{x}$, the estimated profile $\hat{\boldsymbol{x}}$, and the posterior marginals $\mathrm{p}(x_t = k|\boldsymbol{d})$ for $k = \{1, 2, 3, 4\}$ when the likelihood variance τ=0.1.

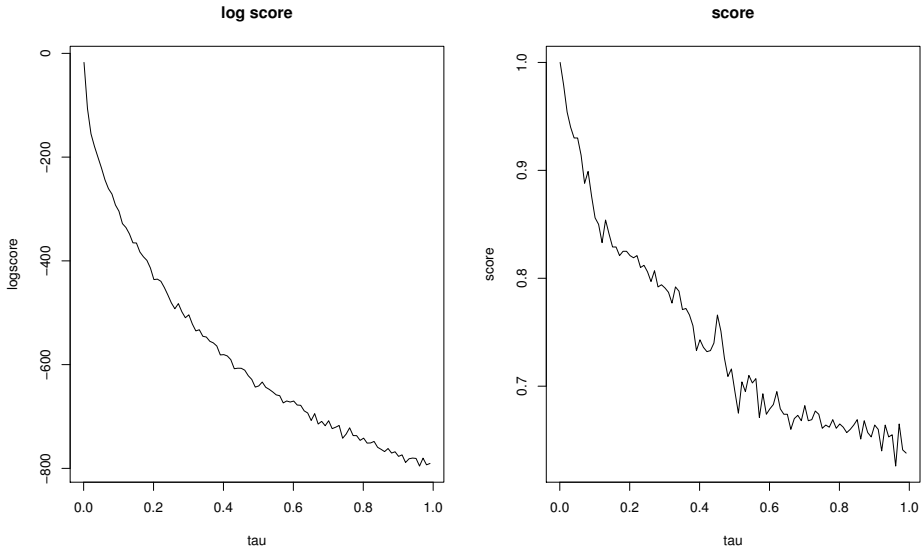


Figure 4.6: Logscore and score as functions of τ in the likelihood variance.

Chapter 5

North Sea well log

We will in this chapter use real test data in the form of a well log from a reservoir in the North Sea. Indirectly, from well log measurements made along the length of drilled wells, we can obtain information on the rock type alternations. This well log consists of gamma-ray radioactivity, porosity, density, pressure- and shear-velocities. These data are a sequence of measurements taken every 15 cm at depths between 2013 m and 2640 m. We consider a subset of this dataset, depths from 2160m to 2250m. This subset of the well log has been studied in Eidsvik et al. (2004). We model the facies with 3 possible states, $x_t \in \{1, 2, 3\}$, which correspond to oil- and brine-saturated sandstone and shale, respectively. The well log measurements give information of the rock type at the locations where they are taken. For example, high clay content may be indicative of shale. The goal for this chapter is to use the dataset to assign one of these three rock types.

The petrophysical variable clay content can be computed from gamma-ray(Gr) by using the following equation:

$$C = \frac{Gr - \min(Gr)}{\max(Gr) - \min(Gr)}.$$

The dataset consisting of the petrophysical variables and elastic properties is shown in Figure 5.1, where the subset of the well log considered is separated by lines.

The prior model for the LF classes is also in this situation defined as a stationary Markov chain through the reference reservoir, with a transition matrix defined as

$$\mathbf{P} = \begin{pmatrix} 0.984 & 0 & 0.016 \\ 0.008 & 0.982 & 0.010 \\ 0.007 & 0.037 & 0.956 \end{pmatrix},$$

with rows and columns corresponding to oil- and brine-saturated sandstone and to shale. This transition matrix is the same as used for the synthetic test data in Chapter 4, from Larsen et al. (2006), where the row and the column corresponding to gas-saturated sandstone are removed and the remaining transition matrix is normalized. The stationary prior PDF for $\mathbf{P}$ is $\pi_0(\boldsymbol{x})$=(0.324,0.455,0.221)

When we cross-plot the clay content C and the porosity ϕ from the well log, see Figure 5.2(a), we can observe that we need to change the petrophysical distribution given in Chapter 4.1, equation 4.1 and 4.2, to match the actual characteristics of the well log. The petrophysical distribution for the real case is assumed to be Gaussian, with expectation

$$\boldsymbol{\mu}(x_t) = \begin{cases} \begin{bmatrix} 0.35 \\ 0.32 \end{bmatrix}, & x_t = 1, 2 \\ \\ \begin{bmatrix} 0.75 \\ 0.39 \end{bmatrix}, & x_t = 3 \end{cases}$$

and covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.15^2 & 0 \\ 0 & 0.02^2 \end{bmatrix}.$$

To study the expected levels and the variability induced by the forward model, we simulate the petrophysical variables $\boldsymbol{m}_t$ given x_t from this distribution for all depth, after first simulating a profile $\boldsymbol{x}$ from the Markov chain prior model $\mathbf{P}$. A cross-plot of the simulated porosity as a function of the simulated clay content is shown in Figure 5.2(b). We can observe that the petrophysical distribution might have been better if the variance $\boldsymbol{\Sigma}$ was dependent of the rock type. We will not consider this further here. We need to tune the rock-physics parameters, defined in Chapter 2.2, from the local geology, to match the well log reasonably accurately. This is done through trial and error, careful tuning would require more geological information. We have tested how the model changes with increasing and decreasing values of every parameters to find a good fit. The parameter values for the best fit are given in Table 5.1. The bulk and shear modulus for clay content are increased, and the bulk and shear modulus for quartz are decreased. The bulk modulus for brine and oil have been increased. The densities for brine, oil and clay content has been increased and the density for quartz has been decreased. The critical porosity ϕ_0 in the rock-physics model has been increased to 0.45 and the given depth Z used is the mean of the depth values in the well log, equal to 2205 m.

Table 5.1: The rock-physics model parameters for the real case

Matrix components		
$K_c = 40$ GPa	$G_c = 12$ GPa	$\rho_c = 2.8 g/cm^3$
$K_q = 20$ GPa	$G_q = 11$ GPa	$\rho_q = 2.5 g/cm^3$
Fluid bulk modulus		
$K_{oil} = 1.5$ GPa	$K_{brine} = 3.6$ GPa	
Fluid density		
$\rho_{oil} = 1.1 g/cm^3$	$\rho_{brine} = 1.2 g/cm^3$	

We have simulated the elastic properties $\boldsymbol{y_t}$ given x_t for all depth t, from the distribution given in Chapter 4.2, equation 4.5. T in the likelihood variance is in

this real case assumed to be

$$T = \tau^2 \cdot T_m,$$

where τ is a given constant and T_m is

$$T_m = \begin{pmatrix} 0.18^2 & 0 & 0 \\ 0 & 0.15^2 & 0 \\ 0 & 0 & 0.1^2 \end{pmatrix}.$$

Different cross-plots are given in Figure 5.3 - 5.5. In each plot, the (a) cross-plot refers to the well log given and the (b) cross-plot to the simulated values of the elastic properties. For the simulated values of the elastic properties in these plots are τ set equal to 1. We can observe that the simulated values fit reasonably well to the the well log. The simulated values could be better tuned from rock-physics relations or the likelihood variance T, but this would require more expert opinion from geophysicists.

To assign the rock types to each measurement depth and to assess the posterior marginal distribution, we use the forward-backward algorithm introduced in Chapter 3.1, in the same way as for the synthetic test data in the previous chapter. Also in this case the code for the forward-backward algorithm and for the rock-physics model are written in **R** (R Development Core Team, 2011). As the data $\boldsymbol{d}$ we use the elastic properties V_P, V_S and ρ from the well log. We will define a naive profile from the well log, where the naive class is defined as

$$\hat{x}_t^{\text{naive}} = \text{argmax}_k[p(d_t|x_t = k)\pi_0(x_t = k)], \tag{5.1}$$

where $p(d_t|x_t)$ is the likelihood function with the same distribution as given in equation 4.5 and $\pi_0(x_t)$ is the stationary distribution for class x_t. This is done for two different τ in $\mathbf{T}$ in the likelihood variance, τ equal to 0.5 and to 2. The naive profile $\hat{\boldsymbol{x}}^{\text{naive}}$, the estimated profile $\hat{\boldsymbol{x}}$, given in Chapter 3.1, equation 3.2, and the posterior marginal distribution $p(x_t = k|\boldsymbol{d})$ for k=1,2,3 as a function of depth t are given in Figure 5.6 for τ equal to 0.5 and in Figure 5.7 for τ equal to 2.

For both values of the τ the estimated profile is equal to the naive profile for some values of the depth t, but not all. The naive estimator is different for the two values of τ. For the smallest τ we get, as expected, a lot more details than for the highest τ in the likelihood variance. For $\tau = 0.5$ the naive profile jumps 42 times, and the estimated profile jumps 24 times. For $\tau = 2$ the naive profile jumps 24 times and the estimated profile jumps 10 times. We can observe that it seems to be oil-saturated sandstone in the top of actual subset of the well log for both cases of the likelihood variance. From the posterior marginal distribution we observe that this is more sure for the smallest value of τ. From these plots it can seem like brine-saturated sandstone and shale are in layers in the rest of the studied subset of the well log. The posterior marginal distribution seems to be a binary zero-one function for mostly all t, at least for the smallest τ. This means that the prediction is good for $\tau = 0.5$. For $\tau = 2$ the prediction is worse. The posterior marginal distribution is not a binary zero-one function for all t, only the depth t where the estimated profile is equal to the naive profile.

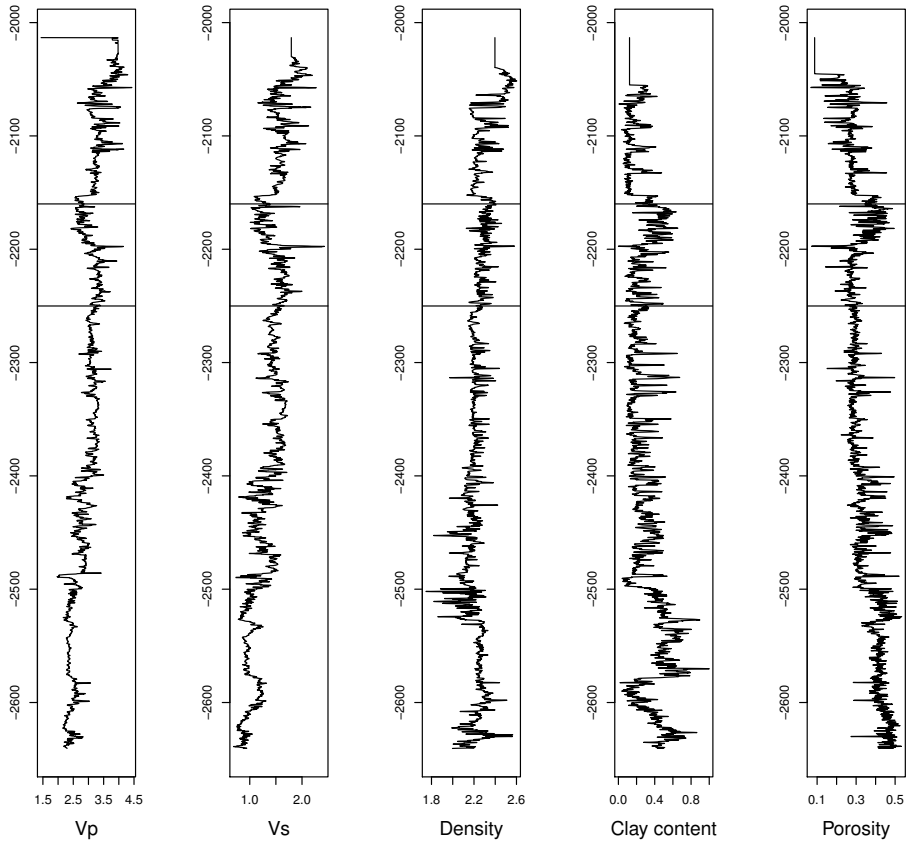
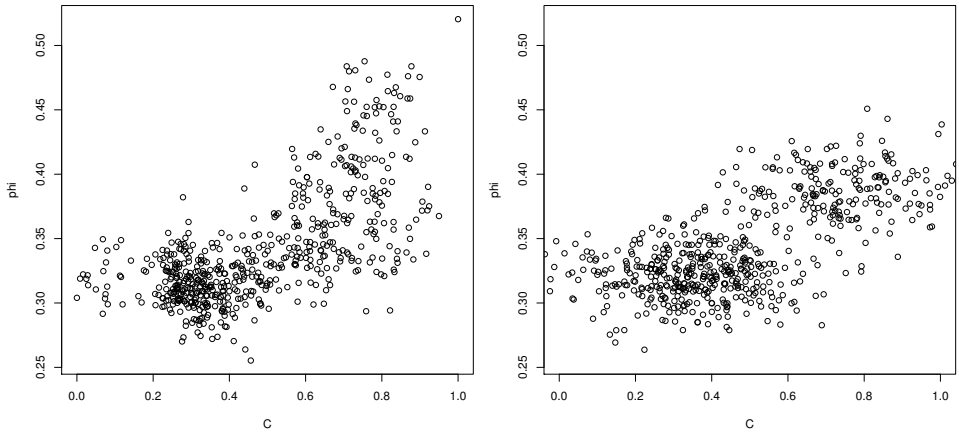


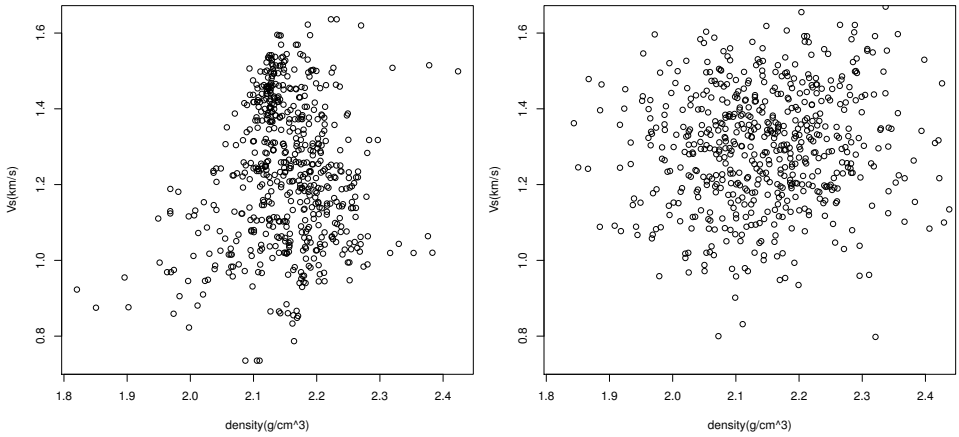
Figure 5.1: Pressure- and shear velocity, V_P and V_S, density, clay content and porosity from the well log, where the studied area between -2160 m and -2250 m is shown.

For the first case, τ equal to 0.5, the posterior PDFs is almost only a binary zero-one functions. Again, a careful tuning of the geophysical parameters, and the likelihood covariance model, could give more reliable results here, and might provide more interpretation of the depositional environment.



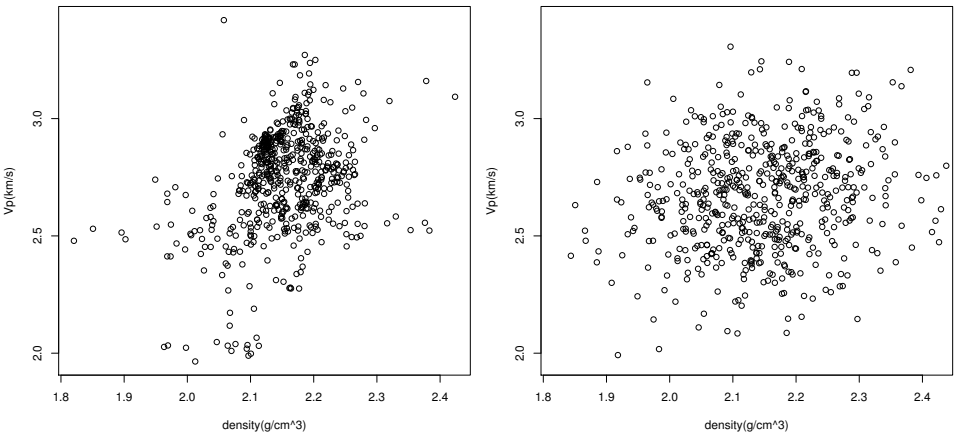
(a) (b)

Figure 5.2: Cross-plot of the petrophysical variables, clay content C and porosity ϕ, from the well log in (a) and simulated in (b).



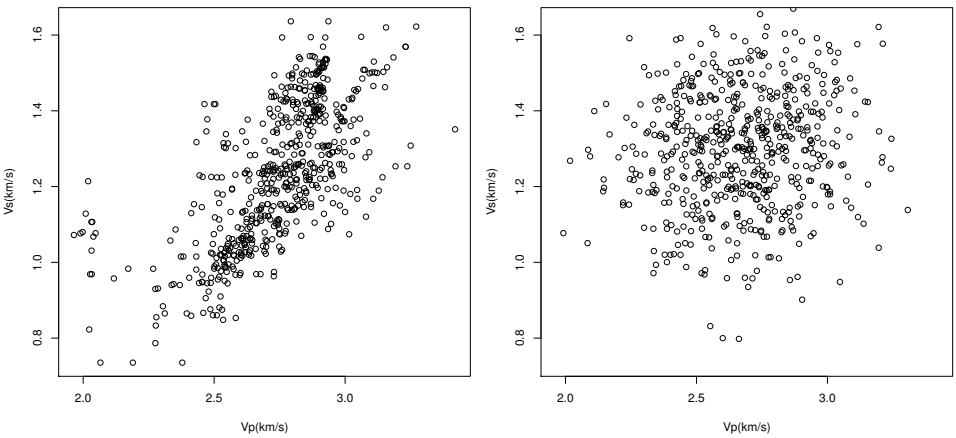
(a) (b)

Figure 5.3: The shear velocity V_S as a function of the density ρ from the well log in (a) and simulated in (b).



(a) (b)

Figure 5.4: The pressure velocity V_P as a function of the density ρ from the well log in (a) and simulated in (b).



(a) (b)

Figure 5.5: The shear velocity V_S as a function of the pressure velocity V_P from the well log in (a) and simulated in (b).

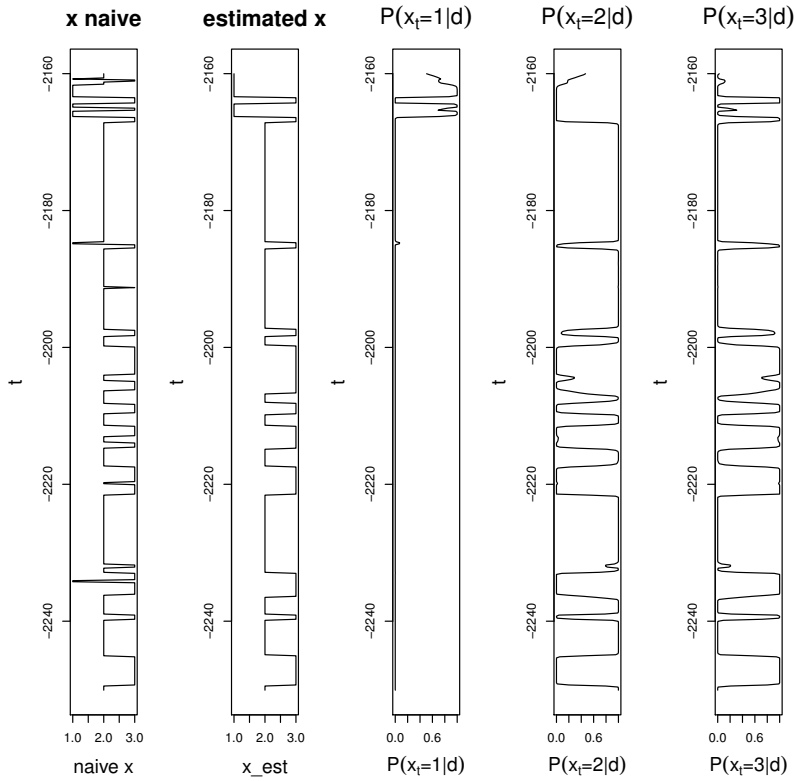


Figure 5.6: The naive profile $\hat{\boldsymbol{x}}^{\text{naive}}$, the estimated profile $\hat{\boldsymbol{x}}$, given in equation 3.2, and the posterior marginal distribution $p(x_t = k|\boldsymbol{d})$ for k=1,2,3 as a function of depth t with τ equal to 0.5.

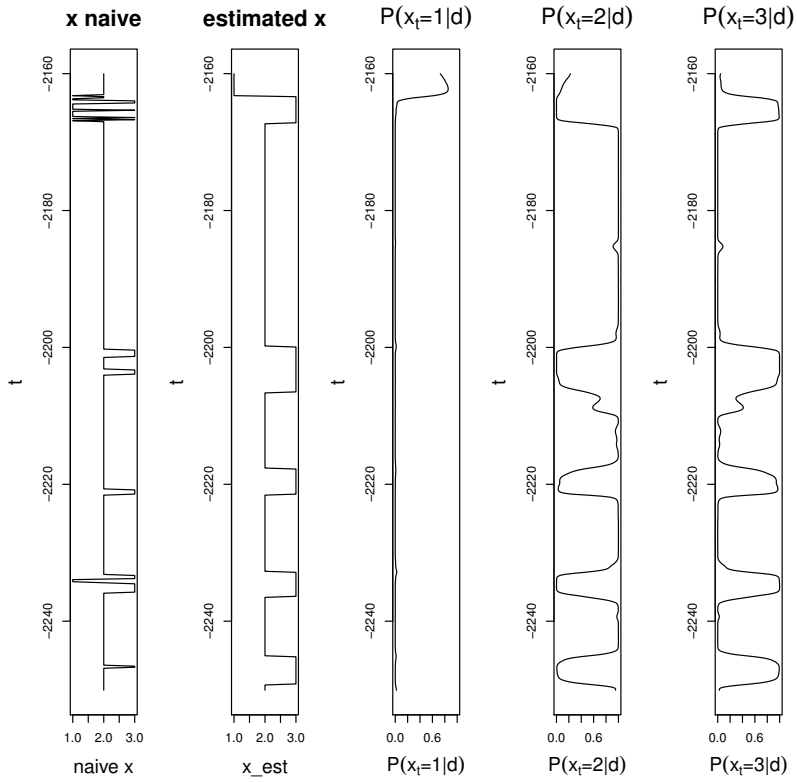


Figure 5.7: The naive profile $\hat{\boldsymbol{x}}^{\text{naive}}$, the estimated profile $\hat{\boldsymbol{x}}$, given in equation 3.2, and the posterior marginal distribution $p(x_t = k|\boldsymbol{d})$ for k=1,2,3 as a function of depth t with τ equal to 2.

Chapter 6

Conclusion

In this thesis a method to predict LF characteristics along a profile is presented. The prediction was done by use of Bayesian inversion to obtain the posterior PDFs of the LF classes and in that way we determined the most probably LF profile. The posterior PDF is not analytically tractable, but has been assessed through a forward-backward algorithm which compute the posterior distribution for all hidden states variables. The prior model considered to reach the posterior PDF is a stationary Markov chain prior model, which has been used to model the continuity of the LF classes along the profile. A likelihood model was used to link LF classes to a set of data. To relate the petrophysical variables to the elastic properties we used forward rock-physics relations.

The most probably LF variables profile has been computed for different set of data. We have first considered synthetic test data inspired by a North Sea sandstone reservoir, where we used either the petrophysical variables or the elastic properties as the data. At the end of this thesis is real data in form of a well log from the North Sea considered as the set of data.

For the first case, the petrophysical variables in the synthetic test data, it was hard to get a good prediction of the lithology/fluid profile. The reason for this was that the petrophysical variables did not discriminate between gas-, oil- and brine-saturated sandstone, which made it impossible to separate them from each other. Through the prior model for LF transitions, we still made a reasonable configuration of the LF profile.

For the second case, the elastic properties in the synthetic test data, we used the forward rock-physics model in the likelihood model to relate the elastic properties to the LF classes. In this case we marginalize over the petrophysical variables. For this model it was easier to separate the different type of sandstone. The rock-physics model did not separate brine-saturated sandstone and shale, this made it hard to separate these two LF classes. In this case the prediction of gas- and oil-saturated sandstone was good, but not the prediction of brine-saturated sandstone and shale. We determined the most probably LF profile, where the configuration for gas- and oil-saturated sandstone was the most sure configuration.

For both cases of the synthetic test data, we started by making a reference

profile, which was used to simulate the synthetic test data. This reference profile was uses as the truth profile and we used different scoring rules to check how good the estimated profile fit to the reference profile. This scoring rules tells us that the estimated profile did not fit too well to the reference profile for the petrophysical variables, and that the prediction of the estimation profile was a lot better for the elastic properties.

For the last case we used the elastic properties from the well log as data. In this case we needed to tune the rock-physics parameters and the petrophysical variables to match the well log. We determined the most probably LF profile also for this case, which shown that it was oil-saturated sandstone in the top of the studied area of the well log and brine-saturated sandstone and shale in layers in the rest of the well log.

In a further research we can estimate model parameters, and use rock-physics relations more carefully. This will result in a more reliable LF profile.

Bibliography

Avseth, P., Mukerji, T., and Mavko, G. *Quantitative seismic interpretation: Applying rock physics tools to reduce interpretation risk*. Cambridge Univ Pr, 2005.

Eidsvik, J., Mukerji, T., and Switzer, P. Estimation of geological attributes from a well log: an application of hidden Markov chains. *Math. Geol.*, 36(3):379–397, 2004.

Gneiting, T. and Raftery, A. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

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

Hossain, Z., Mukerji, T., Dvorkin, J., and Fabricius, I. Rock physics model of glauconitic greensand from the north sea. *Geophysics*, 76:E199, 2011.

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

MacDonald, I. and Zucchini, W. *Hidden Markov and other models for discrete-valued time series*, volume 70. Chapman & Hall/CRC, 1997.

Mavko, G., Mukerji, T., and Dvorkin, J. *The rock physics handbook: Tools for seismic analysis of porous media*. Cambridge Univ Pr, 2003.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. URL `http://www.R-project.org/`. ISBN 3-900051-07-0.

Rimstad, K., Avseth, P., and Omre, H. Bayesian lithology/fluid prediction constrained by spatial couplings and rock physics depth trends, 2010.

Ross, S. *Introduction to probability models*. Academic press, 2007.

Scott, S. Bayesian methods for hidden markov models. *Journal of the American Statistical Association*, 97(457):337–351, 2002.