NTNU
Norwegian University of
Science and Technology

Parameter estimation in convolved categorical models

David Lindberg

Master of Science in Physics and Mathematics
Submission date: July 2010
Supervisor: Karl Henning Omre, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Problem Description

To study discrete convolved models in 1D

- In particular likelihood parameter estimation by maximum marginal likelihood and maximum posterior probability.
- Robustness with respect to model assumptions shall also be evaluated.

This thesis is to be carried out at the Department of Mathematical Sciences at the Norwegian University of Science and Technology under guidance of Professor Karl Henning Omre.

Assignment given: 19. February 2010
Supervisor: Karl Henning Omre, MATH

Abstract

In this thesis, we solve the seismic inverse problem in a Bayesian setting and perform the associated model parameter estimation. The subsurface rock layers are represented by categorical variables, which depends on some response variables. The observations recorded appear as a convolution of these response variables. We thus assess the categorical variables' posterior distribution based on a prior distribution and a convolved likelihood distribution. Assuming that the prior model follows a Markov chain, the full model becomes a hidden Markov model. In the associated Posterior-Prior deconvolution algorithm, we approximate the convolved likelihood in order to use the recursive forward-backward algorithm. The prior and likelihood distributions are parameter dependent, and two parameter estimation approaches are discussed. Both estimation methods make use of the marginal likelihood distribution, which can be computed during the forward-backward algorithm.

In two thorough test studies, we perform parameter estimation in the likelihood. Approximate posterior models, based on the respective parameter estimates, are computed by Posterior-Prior deconvolution algorithms for different orders. The signal-to-noise ratio, a ratio between the observation mean and variance, is found to be of importance. The results are generally more reliable for large values of this ratio. A more realistic seismic example is also introduced, with a more complex model description. The posterior model approximations are here more poor, due to under-estimation of the noise parameter.

Acknowledgments

I would first of all like to thank my supervisor, Henning Omre, for all help and guidance during the work of this master's thesis. When I have lacked motivation he has been understanding and supportive. He has also been available whenever I have needed guidance.

I would also like to thank my family for all their support during my studies. A special thanks to my friends and my fellow class mates for all good times during these five years. Thanks for the many supporting and interesting discussions, during the work of this thesis in particular.

Last I would like to thank my beautiful Lise for always believing in me.

Preface

This master's thesis is written as an ending work of the Master of Technology studies at the Norwegian University of Science and Technology in Trondheim. The thesis is a continuation of my project work from the fall 2009. The work in this thesis corresponds to a workload of 30 study points, which is equivalent to a full semester's work.

Necessary theory is derived, in particular on the marginal likelihood, in order to deduce methods on parameter estimation. Parameter estimation is performed for two thorough test studies and a more realistic seismic empirical example. All algorithms, and corresponding computations, in this thesis have been implemented in MATLAB, version 2009b.

David Lindberg
July 2010, Trondheim

Contents

Chapter 1

Introduction

In this thesis, we want to perform lithology/fluid (LF) prediction in subsurface layers. This is of importance in exploration and development of petroleum reservoirs. In seismic exploration, seismic waves are initiated by a seismic source. The waves are then reflected by the subsurface and recorded by a set of receivers. The propagation of the waves through the different subsurface rock layers is described physically by the wave equation, see References [1] and [2]. Figure 1.1 shows an example of a marine seismic survey, where we see that the waves are initiated and registered from different angles. We will choose to approach the problem in a Bayesian setting.

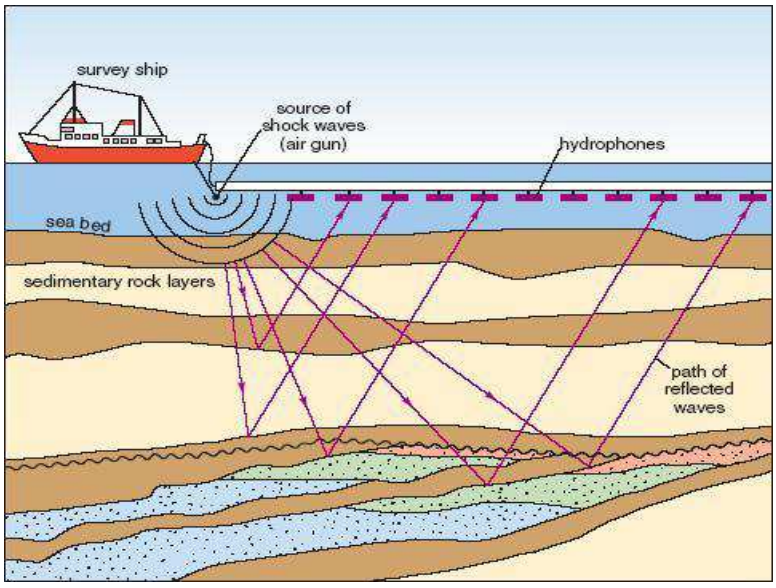


Figure 1.1: *Marine seismic survey with a survey ship towing a wave source, an air gun, and a set of hydrophones, which are wave receivers. Figure from Reference [16].*

1.1 Variables of interest

In this thesis, we study the LF-characteristics along a vertical 1D-profile through the subsurface layers. We discretize the LF-characteristics into a set of LF-classes. The categorical variable, x_t, is associated with one of these classes, and corresponds to the scaled reflection time, t, at which the seismic waves are recorded by the receivers, i.e. the hydrophones in Figure 1.1. The total LF-profile is represented by the full categorical variable vector $\mathbf{x} : \{x_t; t = 1, \ldots, T\}$. Each LF-state, x_t, corresponds to a response, r_t, represented by the full response vector $\mathbf{r}$. The actual seismic data from the LF-states recorded by the hydrophones along the profile are registered as a convolution of these responses, or locally weighted averages. These convolved observations are represented by the discretized vector $\mathbf{d}$.

From a geologic viewpoint, the response variable r_t can represent three elastic medium properties at time t, and thus actually be a three-dimensional vector, see e.g. References [3], [4] and [12]. These properties are P- and S-wave velocity and density, represented by their logarithms in $\mathbf{r}_t$. We only consider r_t at time t as a one-dimensional response variable in this thesis, making $\mathbf{r}$ a $(T \times 1)$ response vector.

Work done in e.g. References [3], [4] and [12] treats observation data, $\mathbf{d}$, from different angles as shown in Figure 1.1. In this thesis, we will consider the simplest case only, assuming the data to be 1D, i.e. from one angle only.

1.2 Seismic inverse problem

For a seismic inverse problem, we want to make inference on the LF-profile, $\mathbf{x}$, based on the observations, $\mathbf{d}$. The general seismic forward model can be written as

$$\mathbf{d} = \mathbf{A}(\mathbf{x}) + \boldsymbol{\epsilon} \ . \tag{1.1}$$

Here $\mathbf{d}$ is an observation vector with seismic data on discrete form related to the full LF-state profile, $\mathbf{x}$, by a transfer function, $\mathbf{A}(\cdot)$, from the state space into the observation space, and $\boldsymbol{\epsilon}$ is observation error. The convolution effect divides the inversion problem in two, giving the divided forward model

$$\mathbf{d} = \mathbf{A}_d(\mathbf{r}) + \boldsymbol{\epsilon}_d \tag{1.2}$$

$$\mathbf{r} = \mathbf{A}_r(\mathbf{x}) + \boldsymbol{\epsilon}_r \ . \tag{1.3}$$

Here $\mathbf{A}_d(\cdot)$ and $\mathbf{A}_r(\cdot)$ are transfer functions from the observation space to the response space and from the response space to the state space respectively. The two forward models descried are ideal cases when we know the true transfer functions. In this thesis, we will set these as matrices, i.e. $\mathbf{A}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ in Expression (1.1), where $\mathbf{A}$ is a $(T \times T)$-matrix, and likewise in Expressions (1.2) and (1.3). This approach leads to model uncertainty in addition to the observation error. Methods on how to set up these seismic forward models are discussed in Reference [2] .

Many LF-states may return the same observation vector, $\mathbf{d}$. The solution is thus non-unique, which makes the inverse problem ill posed and thereby challenging. With our approach, a classical inverse solution based on a square error loss would be

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left(\|\mathbf{d} - \mathbf{A}\mathbf{x}\|^2 \right) \tag{1.4}$$

for the general model in Expression (1.1). This solution may be unstable for an ill-posed problem, i.e. small perturbations in $\mathbf{d}$ by ϵ may cause large change in $\hat{\mathbf{x}}$. The matrix $\mathbf{A}$

may in addition not have full rank. Likewise the inverse solutions for the divided inversion problem in Expressions (1.2) and (1.3) are

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left(\|\hat{\mathbf{r}} - \mathbf{A}_r \mathbf{x}\|^2 \right) \tag{1.5}$$

$$\hat{\mathbf{r}} = \arg\min_{\mathbf{r}} \left(\|\hat{\mathbf{d}} - \mathbf{A}_d \mathbf{r}\|^2 \right) \ . \tag{1.6}$$

In this case, we might have to obtain two unstable inversion solutions, and we might have two matrices, $\mathbf{A}_r$ and $\mathbf{A}_d$, without full rank.

In this thesis, we approach the inverse problem in a Bayesian setting. We then make our inference on the LF-profile, $\mathbf{x}$, based on a combination of the observations, $\mathbf{d}$, and prior knowledge. The solution is represented by the full posterior distribution, $p(\mathbf{x}|\mathbf{d})$. As the objective is to predict the categorical variables in $\mathbf{x}$ based on $\mathbf{d}$, which are registered as a convolution, we name this inversion approach Bayesian categorical deconvolution. The major difference between a Bayesian approach and the classical inversion approach in Expression (1.4) is that the posterior distribution associate a probability to every possible solution. This way we can simulate realizations of the LF-layer based on the seismic observed data and prior knowledge, and in addition obtain information on how likely these predicted LF-layers are.

Throughout this thesis we will assume that the observation error, $\boldsymbol{\epsilon}$ in Expression (1.1), is Gaussian distributed,

$$\boldsymbol{\epsilon} \sim N_T(0, \boldsymbol{\Sigma}_\epsilon) \ . \tag{1.7}$$

The Gaussian distribution is presented in Appendix A.1. This error takes the measurement uncertainty into account. If the error is uncorrelated with equal variance, we may write $\boldsymbol{\epsilon} \sim N_T(0, \sigma_\epsilon^2 \mathbf{I})$. Here $\mathbf{I}$ is the identity matrix of proper dimensions, in this case a $(T \times T)$-matrix.

With our Bayesian approach, the models will depend on some parameters, $\boldsymbol{\theta}$. In the general forward model in Expression (1.1), these parameters may originate from the restrictions on the model matrix, $\mathbf{A}$, including parameters defining the model uncertainty. Also, the Gaussian error would include the variance parameter $\boldsymbol{\Sigma}_\epsilon$ by Expression (1.7). This variance parameter could again depend on other parameters. Likewise for the divided forward model in Expressions (1.2) and (1.3), we would have parameters restricting $\mathbf{A}_r$ and $\mathbf{A}_d$, and variance parameters $\boldsymbol{\Sigma}_{\epsilon_r}$ and $\boldsymbol{\Sigma}_{\epsilon_r}$. All parameters can be set if we know their true values, but this is seldom the case. They should thus be estimated based on the seismic observations, $\mathbf{d}$, and this estimation work will be a main objective in this work, thereby the name of this thesis: Parameter estimation in convolved categorical models.

In Chapter 2, we deduce the Bayesian model in more detail. We define the prior distribution $p(\mathbf{x})$ and a convolved likelihood distribution consisting of a response likelihood $p(\mathbf{r}|\mathbf{x})$ and an acquisition likelihood $p(\mathbf{d}|\mathbf{r})$. Together, these distributions define the posterior distribution $p(\mathbf{x}|\mathbf{d})$. In Chapter 3, we then derive a method on how to compute an approximate posterior distribution $\hat{p}(\mathbf{x}|\mathbf{d})$. The approximation can be computed during a kth order Posterior-Prior deconvolution algorithm, which again makes use of the important forward-backward algorithm. Both of these algorithms are derived in more detail. Study on the efficiency of the first order algorithm, for $k = 1$, is done in Reference [21]. In Chapter 4, we derive methods on how to estimate the model parameters, $\boldsymbol{\theta}$. These are parameters that defines the prior distribution and the two likelihood distributions. The main tool in the parameter estimation is the marginal likelihood, $p(\mathbf{d})$, which we make use of in two different estimation procedures. In Chapters 5 and 6, we apply the models and corresponding methods, derived in the previous chapters, for two thorough test studies. In Chapter 5 we estimate the parameters in the acquisition likelihood for different parameter sets, based on the respective observation

profiles only. The same acquisition likelihood parameter estimation is performed in Chapter 6, which explores the estimations sensitivity to the prior distribution model. In both test studies we then assess the approximated posterior distribution for different orders k, based on the corresponding kth order estimated parameters. In Chapter 7 we estimate the acquisition likelihood parameters for a more realistic empirical example, with reference models related to actual seismic models, see e.g References [12], [14] and [21]. The estimation, and corresponding posterior model assessment, is discussed due to the results from the foregoing test studies. Finally, in Chapter 8, we summarize the findings and look at further work to be done.

Chapter 2

Model formulation

Assume that we have a stochastic process on $\mathcal{D} \in \mathbb{R}^1$ discretised by the lattice ${\mathcal{L}_\mathcal{D}}^t = \{1, \ldots, T\}$. Every x_t has a value from the finite discrete state space $\Omega_x : \{1, 2, \ldots, L\}$. Based on the forward model in Expression (1.1), the objective in a Bayesian setting is to explore the posterior distribution, $p(\mathbf{x}|\mathbf{d})$, where $\mathbf{x} : \{x_t \in \Omega_x; t \in {\mathcal{L}_\mathcal{D}}^t\}$ are the LF-states and $\mathbf{d} : \{d_t \in \mathbb{R}^1; t \in {\mathcal{L}_\mathcal{D}}^t\}$ are observed seismic data. The posterior pdf is

$$p(\mathbf{x}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{x})}{p(\mathbf{d})} p(\mathbf{x}) \tag{2.1}$$

by Bayes' rule. Here $p(\mathbf{d}|\mathbf{x})$ is the likelihood pdf and $p(\mathbf{x})$ is the prior pdf, both assumed to be known, and $p(\mathbf{d})$ is a constant which ensures normality.

In our model, we assume that the distribution of the data points, d_t, depends on the unknown states, x_t. We also assume that these hidden underlying states, x_t, follow a Markov chain with a stationary transition matrix. With these assumptions, the model becomes a hidden Markov model (HMM), see References [17] and [19]. In the rest of this chapter we present the distributions in Expression (2.1) in more detail.

2.1 Prior model

We now assume that the prior chain fulfills the first-order Markov property,

$$p(x_t|x_1, \ldots, x_{t-1}) = p(x_t|x_{t-1}) \quad \forall\, t \quad , \tag{2.2}$$

i.e. that the conditional probability of the state x_t at time t, given all the previous states, only depends on the closest previous state, x_{t-1}. The transition probabilities can be represented by a one-step $(L \times L)$ transition probability matrix, $\mathbf{P}_{x_t}$. If the Markov chain is stationary, the transitions for all times t will have the same transition probability matrix, $\mathbf{P}_x$.

The prior model, $p(\mathbf{x})$, is assumed to be a first order stationary Markov model, by Expression (2.2) we thus have the prior distribution

$$\begin{aligned} p(\mathbf{x}) &= p(x_1, \ldots, x_T) = p(x_T|x_1, \ldots, x_{T-1}) p(x_{T-1}|x_{T-2}, \ldots, x_1) \cdots p(x_2|x_1) p(x_1) \\ &= \prod_{t=1}^{T} p(x_t|x_{t-1}) \quad , \end{aligned} \tag{2.3}$$

where $p(x_1|x_0) = p(x_1)$ for notational ease. Here $p(x_1)$ is the initial marginal prior and $p(x_t|x_{t-1})$ defines the stationary $(L \times L)$ transition probability matrix $\mathbf{P}_x$. In this thesis, we will let $p(x_1)$ be the stationary pdf of $\mathbf{P}_x$.

2.2 Likelihood model

For the most simple HMM we have the conditional independent likelihood model

$$p(\mathbf{d}|\mathbf{x}) = \prod_{t=1}^{T} p(d_t|\mathbf{x}) = \prod_{t=1}^{T} p(d_t|x_t), \tag{2.4}$$

i.e. every observation d_t only depends on the underlying state x_t. A graphical model of the HMM with likelihood by Expression (2.4) and prior by Expression (2.3) is shown in Figure 2.1.

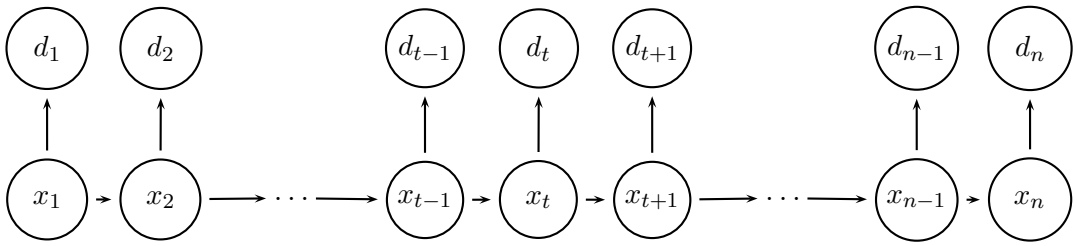


Figure 2.1: *Graphical model of a simple HMM.*

Given all states in Figure 2.1, each node will only depend on the nodes connected to itself by an edge. We observe the dependency relationships between the observations, $\mathbf{d}$, and the states, $\mathbf{x}$, defined by Expression (2.4). We also observe the prior Markov property in Expression (2.3) by the directed edges between elements in $\mathbf{x}$.

In the convolved seismic inversion case, with forward models by Expressions (1.2) and (1.3), we assume a likelihood model by

$$p(\mathbf{d}|\mathbf{x}) = \int p(\mathbf{d}|\mathbf{r})p(\mathbf{r}|\mathbf{x})d\mathbf{r} \ . \tag{2.5}$$

Here $p(\mathbf{d}|\mathbf{r})$ is termed the acquisition likelihood model and $p(\mathbf{r}|\mathbf{x})$ the response likelihood model. We choose the likelihood such that $[\mathbf{d}|\mathbf{r}]$ is independent of $\mathbf{x}$. A graphical model of the HMM with this likelihood model is shown in Figure 2.2. The Markov property in the prior distribution is, by Expression (2.3), the same as for the simple HMM in Figure 2.1, but the likelihood for the observation variables $\mathbf{d}$ and $\mathbf{r}$ makes it more complex. We see that every observation element, d_t, depends on the complete response vector $\mathbf{r}$.

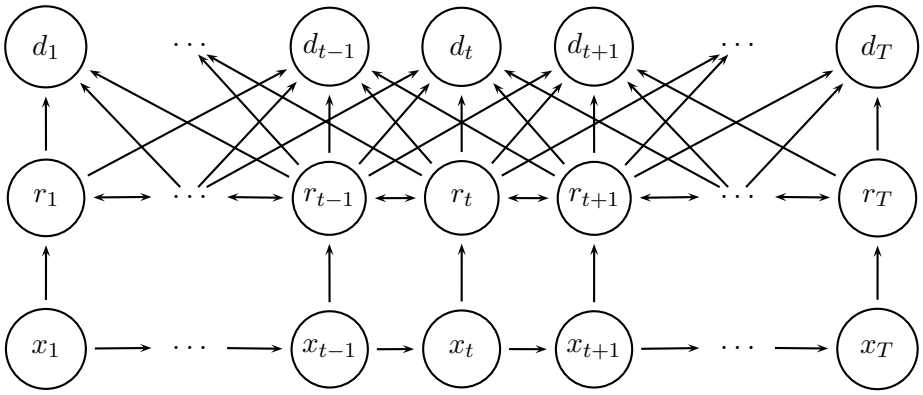


Figure 2.2: *Graphical model of the HMM for our convolved Bayesian model.*

2.2.1 Response likelihood model

The response likelihood model, $p(\mathbf{r}|\mathbf{x})$, represents response variables related to the $\mathbf{x}$-classes that are actually captured by the observations in $\mathbf{d}$,

$$[\mathbf{r}|\mathbf{x}] = \boldsymbol{\mu}_x + \boldsymbol{\sigma}_x e \quad \sim \quad p(\mathbf{r}|\mathbf{x}) = N_T(\boldsymbol{\mu}_x, \boldsymbol{\sigma}_x \boldsymbol{\sigma}_x') \ . \tag{2.6}$$

Here $\mathbf{e}$ is a $(T \times 1)$-vector of independent, standard Gaussian random variables, $\mathbf{e} \sim N_T(0, \mathbf{I})$. In this pdf, $\boldsymbol{\mu}_x = (\mu_{x_1}, \ldots, \mu_{x_T})'$ is the $(T \times 1)$ response expectation vector and $\boldsymbol{\sigma}_x = (\sigma_{x_1}, \ldots, \sigma_{x_T})'$ is the $(T \times 1)$ response variance vector related to the classes. Hence the marginal response likelihoods, $[r_t|x_t]$, will have independent univariate Gaussian pdfs by

$$[r_t|x_t] \quad \sim \quad p(r_t|x_t) = N_1(\mu_{x_t}, \sigma_{x_t}^2) \ \ \forall \ t \ . \tag{2.7}$$

Here μ_{x_t} is the t'th element in $\boldsymbol{\mu}_x$ and σ_{x_t} is the t'th element in $\boldsymbol{\sigma}_x$. The response likelihood distribution in Expression (2.6) can then be factorized by its marginals,

$$p(\mathbf{r}|\mathbf{x}) = \prod_{t=1}^{T} p(r_t|x_t) \ . \tag{2.8}$$

2.2.2 Acquisition likelihood model

The convolution effect is captured by the acquisition likelihood model,

$$[\mathbf{d}|\mathbf{r}] = \mathbf{W}\mathbf{r} + \mathbf{D}\mathbf{e} \quad \sim \quad p(\mathbf{d}|\mathbf{r}) = N_T(\mathbf{W}\mathbf{r}, \mathbf{D}\mathbf{D}') \ . \tag{2.9}$$

Here $\mathbf{W}$ is a $(T \times T)$-convolution matrix, which is fairly broad-banded as we assume the convolution to be relatively wide. The error term, $\mathbf{e}$, is a $(T \times 1)$-vector of independent, standard Gaussian random variables, which makes the acquisition likelihood, given by Expression (2.9), Gauss-linear. The $(T \times T)$-matrix $\mathbf{D}$ captures observation error, and may be colored such that $\mathbf{D}\mathbf{e} = \mathbf{W}\mathbf{e}_1 + \mathbf{e}_2$. Here $\mathbf{e}_1$ and $\mathbf{e}_2$ are both independent Gaussian distributed error terms, i.e. $\mathbf{e}_1 \sim N_T(0, \sigma_{e_1}^2 \mathbf{I})$ and $\mathbf{e}_2 \sim N_T(0, \sigma_{e_2}^2 \mathbf{I})$, and represents colored and white noise respectively. Thus, with colored noise the acquisition likelihood pdf is

$$[\mathbf{d}|\mathbf{r}] \quad \sim \quad p(\mathbf{d}|\mathbf{r}) = N_T(\mathbf{W}\mathbf{r}, \sigma_{e_1}^2 \mathbf{W}\mathbf{W}' + \sigma_{e_2}^2 \mathbf{I}) \ . \tag{2.10}$$

With white noise only, i.e. $\mathbf{D}\mathbf{e} = \mathbf{e} \sim N_T(0, \sigma_e^2 \mathbf{I})$, the acquisition likelihood would be $p(\mathbf{d}|\mathbf{r}) = N_T(\mathbf{W}\mathbf{r}, \sigma_e^2 \mathbf{I})$.

Throughout this thesis, we will assume that the convolution is stationary. The rows in $\mathbf{W}$ will then contain the same elements. These rows corresponds to wavelets, $\mathbf{w}$, which are discussed and estimated in Reference [4]. In the two test studies in Chapters 5 and 6, we assume that the wavelets are equal, normalized second order exponential functions. We discretize this function for a variable interval of length $2a + 1$ centered around zero,

$$\mathbf{w} = [w_i]_{i=\{-a,\ldots,a\}} = \left[\frac{C_w}{\sqrt{2\pi}\sigma_w} e^{\frac{x_w^2}{2\sigma_w^2}} \right]_{x_w=\{-a,\ldots,a\}} \ . \tag{2.11}$$

Hence $\mathbf{W}$ will be a $2a + 1$-diagonal convolution matrix. Here $C_w = \left[\sum_{i=-a}^{a} w_i\right]^{-1}$ is an added constant that ensures normality of the discretization. For a large value of the interval parameter a, more of the functions tails is captured in $\mathbf{w}$ by Expression (2.11).

2.3 Posterior model

For the simple HMM with prior distribution by Expression (2.3) and conditional independent likelihood distribution by Expression (2.4), we have the posterior distribution by Expression (2.1),

$$p(\mathbf{x}|\mathbf{d}) = C_d \cdot p(\mathbf{d}|\mathbf{x})p(\mathbf{x}) = C_d \cdot \prod_{t=1}^{T} p(d_t|x_t)p(x_t|x_{t-1}) \quad . \tag{2.12}$$

Here we find the constant, C_d, by a summation over the states, x_t, for all times t,

$$\sum_{\mathbf{x}} p(\mathbf{x}|\mathbf{d}) = 1 \quad \Rightarrow \quad C_d = \left[\sum_{x_1} \dots \sum_{x_T} \prod_{t=1}^{T} p(d_t|x_t)p(x_t|x_{t-1}) \right]^{-1} \quad .$$

The constant, C_d, corresponds to $1/p(\mathbf{d})$ in Expression (2.1), and ensures normality. In practice, the computation of C_d is very computer demanding, or even impossible as there are L^T possible outcomes of $\mathbf{x}$. The posterior distribution can be rewritten as $p(\mathbf{x}|\mathbf{d}) = p(x_1|\mathbf{d}) \prod_{t=2}^{T} p(x_t|x_1, \dots, x_{t-1}, \mathbf{d})$. It can be shown that $p(x_t|x_1, \dots, x_{t-1}, \mathbf{d}) = p(x_t|x_{t-1}, \mathbf{d})$, hence the posterior distribution is also a Markov chain,

$$p(\mathbf{x}|\mathbf{d}) = \prod_{t=1}^{T} p(x_t|x_{t-1}, \mathbf{d}) \quad , \tag{2.13}$$

where $p(x_1|x_0, \mathbf{d}) = p(x_1|\mathbf{d})$. The posterior transition probability at time t, i.e. $p(x_t|x_{t-1}, \mathbf{d})$, defines the non-stationary $(L \times L)$ posterior transition probability matrix, $\mathbf{P}_{x_t|d}$. This Markov model is non-homogeneous as the conditioning on different data, $\mathbf{d}$, may give different transition matrices, $\mathbf{P}_{x_t|d}$. In the next chapter, we will introduce the forward-backward algorithm in which we can compute the posterior distribution in Expression (2.13) recursively.

For the more complex HMM with the same prior distribution by Expression (2.3), but with the likelihood distribution by Expression (2.5), we have the posterior distribution

$$p(\mathbf{x}|\mathbf{d}) = C_d \cdot \int p(\mathbf{d}|\mathbf{r})p(\mathbf{r}|\mathbf{x})d\mathbf{r} \cdot \prod_{t=1}^{T} p(x_t|x_{t-1}) \quad , \tag{2.14}$$

with likelihood models defined by Expressions (2.8) and (2.9). We have reasons to believe that this posterior distribution is not a first order Markov chain, because of the coupling in the likelihood model. The normalizing constant, C_d, can be found by the same procedure as in Expression (2.13), and is hard to compute in practice as for the simple HMM. In the next chapter, we will approximate the acquisition likelihood in Expression (2.9) in order to be able to use the Posterior-Prior ratio deconvolution algorithm. The posterior distribution in Expression (2.14) is then approximated by a Markov chain as for the simple HMM in Expression (2.13). This way, we can compute the approximated posterior distribution by the forward-backward algorithm.

Chapter 3

Assessing the posterior model

In this chapter, we first introduce the forward-backward algorithm for the simple HMM model in Figure 2.1. We then approximate the posterior distribution in the HMM in Figure 2.2, using the Posterior-Prior ratio deconvolution algorithm, to gain a posterior distribution with Markov properties. This way we can use the forward-backward algorithm in the approximated, more complex HMM also. Most of the procedures presented originate from work done in Reference [14].

3.1 Forward-backward algorithm

We now define two probabilities, the forward probability,

$$p_f(x_1, \ldots, x_t) = p(x_1, \ldots, x_t | d_1, \ldots, d_t) \ , \tag{3.1}$$

and the backward probability,

$$p_b(x_1, \ldots, x_t) = p(x_1, \ldots, x_t | d_1, \ldots, d_T) \ . \tag{3.2}$$

The difference between the forward and the backward probabilities is that p_f conditions on the observations in $\mathbf{d}$ up to the highest state index t, while p_b conditions on all the observations. We also notice that the full backward probability is

$$p_b(\mathbf{x}) = p_b(x_1, \ldots, x_T) = p(\mathbf{x}|\mathbf{d}),$$

i.e. it is equal to the full posteriori distribution which we want to find. In the rest of this section, we deduce the forward-backward algorithm for the simple HMM in Figure 2.1.

The initial marginal forward probability is

$$p_f(x_1) = p(x_1|d_1) = C_{x_1} \cdot p(d_1|x_1)p(x_1) \ , \tag{3.3}$$

where the constant, $C_{x_1} = 1/p(d_1)$, is found by

$$\sum_{x_1} p_f(x_1) = 1 \ \ \Rightarrow \ \ C_{x_1} = \frac{1}{p(d_1)} = \frac{1}{\sum_{x_1} p(d_1|x_1)p(x_1)} \ .$$

The joint forward probability distribution of two neighbor states is

$$\begin{aligned} p_f(x_{t-1}, x_t) &= p(x_{t-1}, x_t | d_1, \ldots, d_t) = \frac{p(x_{t-1}, x_t, d_1, \ldots, d_t)}{p(d_1, \ldots, d_t)} \\ &= \frac{1}{p(d_1, \ldots, d_t)} \cdot p(d_t|x_t, x_{t-1}, d_1, \ldots, d_{t-1})p(x_t|x_{t-1}, d_1, \ldots, d_{t-1}) \\ &\quad \cdot p(x_{t-1}|d_1, \ldots, d_{t-1})p(d_1, \ldots, d_{t-1}) \\ &= C_{x_{t-1,t}} \cdot p(d_t|x_t)p(x_t|x_{t-1})p_f(x_{t-1}) \ , \end{aligned} \tag{3.4}$$

where we can find $p(d_t|x_t)$ by the likelihood pdf in Expression (2.4) and $p(x_t|x_{t-1})$ by the prior pdf in Expression (2.3). In the last equation, we have used the Markov property by Expression (2.2) and the conditional independence properties by the likelihood model in Expression (2.4). The constant, $C_{x_{t-1,t}} = p(d_1, \ldots, d_{t-1})/p(d_1, \ldots, d_t)$, can be found by

$$\sum_{x_{t-1}} \sum_{x_t} p_f(x_{t-1}, x_t) = 1 \quad \Rightarrow \quad C_{x_{t-1,t}} = \frac{1}{\sum_{x_{t-1}} \sum_{x_t} p(d_t|x_t) p(x_t|x_{t-1}) p_f(x_{t-1})} \, ,$$

with the summations being over the state space, $x_t, x_{t-1} \in \Omega_x : \{1, 2, \ldots, L\}$. Now the marginal forward probabilities are simply found by

$$p_f(x_t) = \sum_{x_{t-1}} p_f(x_{t-1}, x_t) = C_{x_{t-1,t}} \cdot \sum_{x_{t-1}} p(d_t|x_t) p(x_t|x_{t-1}) p_f(x_{t-1}) \quad . \tag{3.5}$$

We can thus compute the marginal and joint forward probabilities, by Expressions (3.5) and (3.4) respectively, for all t recursively.

We notice that $p_f(x_T) = p_b(x_T)$, thus we have an initial upper marginal backward probability. The conditional backward probability distribution of two neighbor states is

$$\begin{aligned}
p_b(x_{t-1}|x_t) &= p(x_{t-1}|x_t, d_1, \ldots, d_T) && (3.6)\\
&= p(x_{t-1}|x_t, d_1, \ldots, d_{t-1}) && (3.7)\\
&= p(x_{t-1}|x_t, d_1, \ldots, d_{t-1}, d_t) && (3.8)\\
&= \frac{p(x_{t-1}, x_t|d_1, \ldots, d_t) \cdot p(d_1, \ldots, d_t)}{p(x_t|d_1, \ldots, d_t) \cdot p(d_1, \ldots, d_t)} \\
&= \frac{p_f(x_{t-1}, x_t)}{p_f(x_t)} \quad . && (3.9)
\end{aligned}$$

From Expression (3.6) to Expression (3.7) we have used the Markov property, i.e. x_{t-1} will be independent of $d_t, \ldots, d_T$ when x_t is given. This is also shown graphically in Figure 2.1. Since we can condition on anything that is independent, we condition x_{t-1} on d_t from Expression (3.7) to Expression (3.8) to gain desirable properties. The result in Expression (3.9) only depends on the joint and marginal forward probabilities found by Expressions (3.4) and (3.5) respectively. As in the forward procedure, we find the marginal backward probabilities by

$$p_b(x_{t-1}) = \sum_{x_t} p_b(x_{t-1}, x_t) = \sum_{x_t} p_b(x_{t-1}|x_t) p_b(x_t) \quad , \tag{3.10}$$

with the summation being over the state space, Ω_x.

We can now make the full forward-backward algorithm by Expressions (3.3), (3.4), (3.5), (3.9) and (3.10). The full algorithm is presented in Algorithm 1.

Algorithm 1: Forward-backward algorithm

Forward:

- Initiate:
  $p_f(x_1) = C_{x_1} \cdot p(d_1|x_1) p(x_1)$
  $C_{x_1} = \frac{1}{\sum_{x_1} p(d_1|x_1) p(x_1)}$
- Iterate for $t = 2, \ldots, T$:
  $p_f(x_{t-1}, x_t) = C_{x_{t-1,t}} \cdot p(d_t|x_t) p(x_t|x_{t-1}) p_f(x_{t-1})$
  $C_{x_{t-1,t}} = \frac{1}{\sum_{x_{t-1}} \sum_{x_t} p(d_t|x_t) p(x_t|x_{t-1}) p_f(x_{t-1})}$
  $p_f(x_t) = \sum_{x_{t-1}} C_{x_{t-1,t}} \cdot p(d_t|x_t) p(x_t|x_{t-1}) p_f(x_{t-1})$

Backward:

- Initiate:
  $p_b(x_T) = p_f(x_T)$
- Iterate for $t = T, \ldots, 2$:
  $p_b(x_{t-1}|x_t) = \frac{p_f(x_{t-1}, x_t)}{p_f(x_t)}$
  $p_b(x_{t-1}) = \sum_{x_t} p_b(x_{t-1}|x_t) p_b(x_t)$
  $p_b(x_t|x_{t-1}) = \frac{p_b(x_{t-1}|x_t) p_b(x_t)}{p_b(x_{t-1})}$

After running Algorithm 1, we can compute the full posteriori distribution by Expression (2.13),

$$p(\mathbf{x}|\mathbf{d}) = \prod_{t=1}^{T} p(x_t|x_{t-1}, \mathbf{d}) = p_b(x_1) \prod_{t=2}^{T} p_b(x_t|x_{t-1}) \quad . \tag{3.11}$$

From Expression (3.11), we can simulate realizations $[\mathbf{x}|\mathbf{d}]^s = (x_1^s, \ldots, x_T^s)$ from the posterior distribution by the following algorithm.

Algorithm 2: Simulate full state realizations

- Draw $x_1^s \sim p_b(x_1)$.
- For $t = 2, \ldots, T$: Draw $x_t^s \sim p_b(x_t|x_{t-1}^s)$.

Then, we have the full simulated realization $(x_1^s, \ldots, x_T^s) = \mathbf{x}^s$. The forward-backward algorithm can be used to compute the locationwise maximum a posteriori predictions (MAP),

$$[\hat{x}_t|\mathbf{d}]_{MAP} = \arg\max_{x_t}\{p(x_t|\mathbf{d})\} = \arg\max_{x_t}\{p_b(x_t)\} \ \ \forall\, t \quad . \tag{3.12}$$

The global MAP can be computed by the Viterbi algorithm, which is a recursive optimal solution to the problem of estimating the full state sequence of a discrete-time finite-state Markov process, see Reference [10]. This is a dynamic programming algorithm for finding the shortest route, see Reference [6], where the transition probabilities are weights between the states. We then find the path through the states in time that maximizes the total weight, i.e. maximizes the probability of the full state sequence. A small example on how the VA algorithm works is displayed in Figure 3.1.

Algorithm 3: Viterbi algorithm

- Initiate:
  $\gamma_1(l) = p_b(x_1 = l) \ , \qquad \forall\, l \in \Omega_x = \{1, 2, \ldots, L\}$
  $\hat{\mathbf{x}}_1(l) = \{l\}$
- Iterate for $t = 2, \ldots, T$:
  $\gamma_{t,t-1}(l, l^*) = \gamma_{t-1}(l^*) \cdot p_b(x_t = l|x_{t-1} = l^*) \ , \qquad \forall\, l, l^* \in \Omega_x$
  $\gamma_t(l) = \max_{l^*}\{\gamma_{t,t-1}(l, l^*)\}$
  $\hat{\mathbf{x}}_t(l) = \{\hat{\mathbf{x}}_{t-1}(l^*), l\}$
- Global MAP:
  $l_T = \arg\max_l\{\gamma_T(l)\}$
  $[\hat{\mathbf{x}}|\mathbf{d}]_{MAP} = \hat{\mathbf{x}}_T(l_T)$

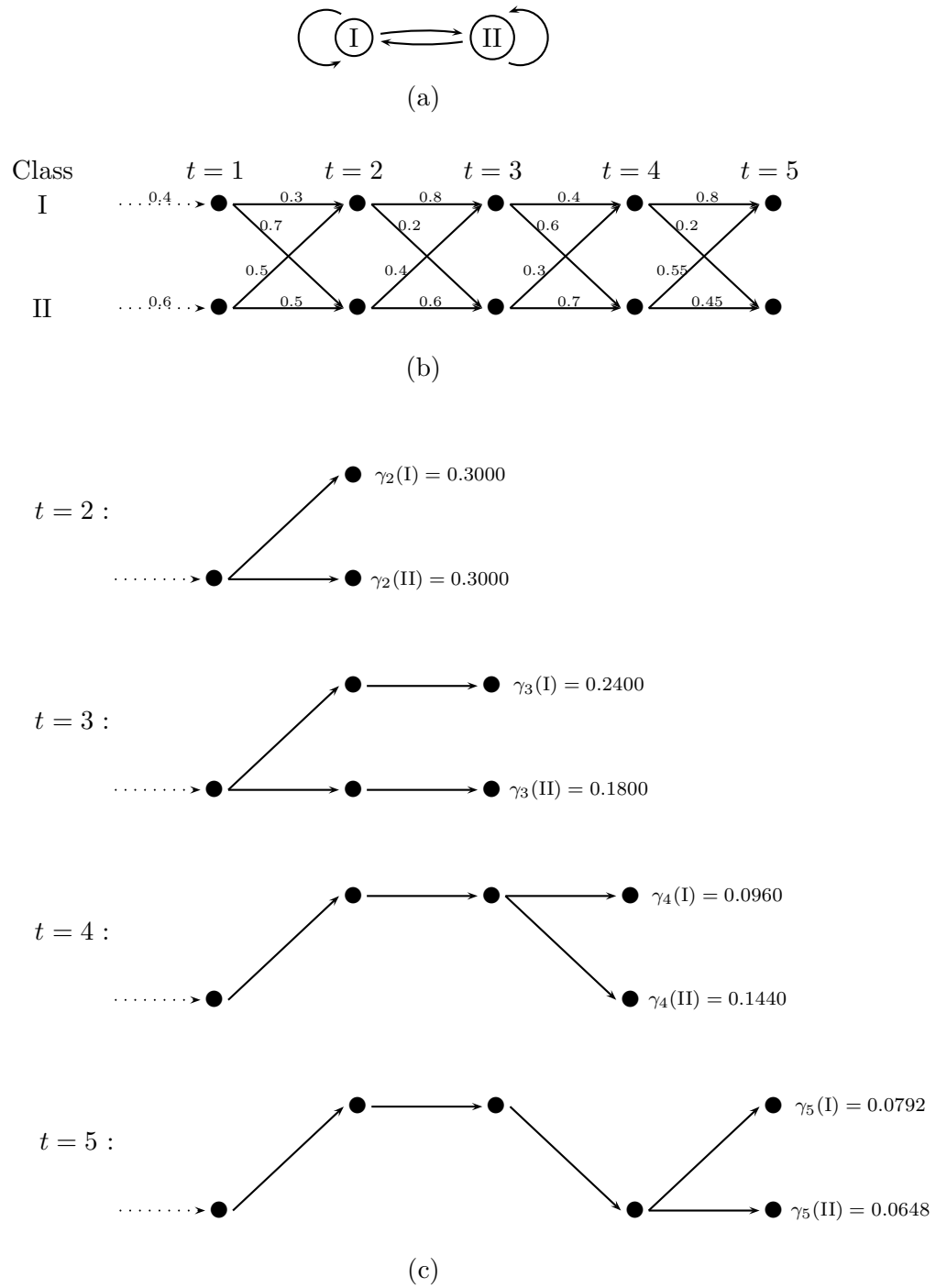


Figure 3.1: *An example of the VA for a Markov chain of length $T = 5$ with $L = 2$ classes, i.e. $\Omega_x = \{I, II\}$, see the state diagram in (a). In (b) we see the backward initial probabilities, $p_b(x_1)$, as the dotted lines weights and the backward transition probabilities, $p_b(x_t|x_{t-1})$, as the solid lines weights. These are computed during the forward-backward algorithm. The recursive shortest path evaluations are shown in (c). As $\gamma_5(I) > \gamma_5(II)$, the global MAP after the VA is $[\hat{\mathbf{x}}|\mathbf{d}]_{MAP} = \hat{\mathbf{x}}_5(I) = \{II, I, I, II, I\}$.*

3.2 Posterior-Prior (P-P) ratio deconvolution algorithm

We want to find an approximation to the full posterior distribution in Expression (2.14) by

$$\hat{p}(\mathbf{x}|\mathbf{d}) = C_{d*} \cdot \prod_{t=1}^{T} l_d(x_t) p(x_t|x_{t-1}) \quad . \tag{3.13}$$

Here $l_d(x_t)$ is a likelihood which only depends on x_t when $\mathbf{d}$ is given. This way the approximate posterior distribution becomes a non-stationary first-order Markov chain by the same reasoning as for Expression (2.13),

$$\hat{p}(\mathbf{x}|\mathbf{d}) = \prod_{t=1}^{T} \hat{p}(x_t|x_{t-1}, \mathbf{d}) \quad , \tag{3.14}$$

with $\hat{p}(x_1|x_0, \mathbf{d}) = \hat{p}(x_1|\mathbf{d})$ for notational ease. We only approximate the posterior distribution by approximations in the likelihood distribution, as the prior distribution often defines constraints on the posterior. These constraints may be hard to ensure in prior model approximations.

We now rewrite the acquisition likelihood model in Expression (2.9) by Bayes' rule,

$$p(\mathbf{d}|\mathbf{r}) = C_{d|r} \cdot \frac{p(\mathbf{r}|\mathbf{d})}{p(\mathbf{r})} \quad .$$

This is valid for all $p(\mathbf{r})$ as long as $p(\mathbf{d}|\mathbf{r})$ is used when calculating $p(\mathbf{r}|\mathbf{d})$. As $p(\mathbf{d}|\mathbf{r})$ is Gauss-linear by Expression (2.9), we choose $p(\mathbf{r})$ to be Gaussian. Then $p(\mathbf{r}|\mathbf{d})$ will also be Gaussian,

$$p(\mathbf{d}|\mathbf{r}) = C_{d|r} \cdot \frac{p_*(\mathbf{r}|\mathbf{d})}{p_*(\mathbf{r})} \quad , \tag{3.15}$$

where

$$p_*(\mathbf{r}|\mathbf{d}) = N_T(\boldsymbol{\mu}_{\mathbf{r}|\mathbf{d}}, \boldsymbol{\Sigma}_{\mathbf{r}|\mathbf{d}}) \tag{3.16}$$

$$p_*(\mathbf{r}) = N_T(\boldsymbol{\mu}_{\mathbf{r}}, \boldsymbol{\Sigma}_{\mathbf{r}}) \quad . \tag{3.17}$$

The prior distribution in Expression (2.3) and the response likelihood distribution in Expression (2.6) indicate that $p_*(\mathbf{r})$ is non-Gaussian, but we can still calculate $\boldsymbol{\mu}_{\mathbf{r}}$ and $\boldsymbol{\Sigma}_{\mathbf{r}}$ by finding the two first moments analytically. The parameters $\boldsymbol{\mu}_{\mathbf{r}|\mathbf{d}}$ and $\boldsymbol{\Sigma}_{\mathbf{r}|\mathbf{d}}$ are also available analytically. The approximation of these four parameters is performed in Appendix B.1, see Expressions (B.9), (B.13) and (B.14). For a first order algorithm we then approximate Expression (3.15) by

$$\hat{p}(\mathbf{d}|\mathbf{r}) = C_{d|r} \cdot \prod_{t=1}^{T} \frac{p_*(r_t|\mathbf{d})}{p_*(r_t)} \quad , \tag{3.18}$$

where $p_*(r_t|\mathbf{d})$ and $p_*(r_t)$ are marginal distributions of Expressions (3.16) and (3.17) respectively. With this approximation, the dependencies in $\mathbf{r}$ and $[\mathbf{r}|\mathbf{d}]$ are ignored since we only use their first order marginals.

We can now approximate the full posterior distribution in Expression (2.14) by the response likelihood distribution in Expression (2.8) and the approximated acquisition likelihood

distribution in (3.18). The approximated posterior distribution is

$$\begin{aligned}\hat{p}(\mathbf{x}|\mathbf{d}) &= C_d \cdot \int \hat{p}(\mathbf{d}|\mathbf{r}) p(\mathbf{r}|\mathbf{x}) d\mathbf{r} \cdot \prod_{t=1}^{T} p(x_t|x_{t-1}) \\ &= C_d \cdot C_{d|r} \cdot \int \left(\prod_{t=1}^{T} \frac{p_*(r_t|\mathbf{d})}{p_*(r_t)} \prod_{t=1}^{T} p(r_t|x_t) \right) d\mathbf{r} \prod_{t=1}^{T} p(x_t|x_{t-1}) \\ &= C_{d*} \cdot \prod_{t=1}^{T} l_d(x_t) p(x_t|x_{t-1}) \ ,\end{aligned}$$

according to Expression (3.13). Here

$$l_d(x_t) = \int \frac{p_*(r_t|\mathbf{d})}{p_*(r_t)} p(r_t|x_t) dr_t \tag{3.19}$$

and $C_{d*} = C_d \cdot C_{d|r}$ is a normalizing constant. We notice that the full observation vector $\mathbf{d}$ is used in $l_d(x_t)$ for all times t when we apply the approximated acquisition likelihood distribution in Expression (3.18). The integral in Expression (3.19) can be approximated numerically by Algorithm 5 in Appendix B.2.

The posterior model has been approximated by a first-order Markov chain according to Expression (3.14), and we can compute the approximated posterior distribution by the forward-backward algorithm in Section 3.1. Instead of using the marginal likelihoods $p(d_t|x_t)$ in the forward recursions in Expression (3.4), we use the approximated likelihood $l_d(x_t)$.

3.3 Higher order P-P ratio deconvolution algorithm

We now generalize the P-P ratio deconvolution algorithm from the last section. First we define a new representation of the LF-states $\mathbf{x} = (x_1, \ldots, x_T)$ by the kth order states $x_t^{(k)} = (x_{t-k+1}, \ldots, x_t)$,

$$\mathbf{x}^{(k)} = \left(x_k^{(k)}, x_{k+1}^{(k)}, \ldots, x_T^{(k)} \right) = [(x_1, \ldots, x_k), (x_2, \ldots, x_{k+1}), \ldots, (x_{T-k+1}, \ldots, x_T)] \ .$$

The kth order full LF-state $\mathbf{x}^{(k)}$ contains the whole first order full LF-state $\mathbf{x}$, but many elements x_t are repeated. The state space for $x_t^{(k)}$ has L^k elements. One can show that the prior model is still Markovian,

$$\begin{aligned}p(\mathbf{x}^{(k)}) &= \prod_{t=k}^{T} p\left(x_t^{(k)} \middle| x_{t-1}^{(k)*} \right) = \prod_{t=k}^{T} p[(x_{t-k+1}, \ldots, x_t)|(x_{t-k}^*, \ldots, x_{t-1}^*)] \\ &= \prod_{t=k}^{T} \left(\prod_{i=1}^{k-1} I\left(x_{t-k+i} = x_{t-k+i}^* \right) \right) \cdot p\left(x_t | x_{t-1}^* \right) \ . \end{aligned} \tag{3.20}$$

Here $I(A)$ is an indicator function, which is equal to 1 if A is true and 0 if A is false, and $p(x_t|x_{t-1}^*)$ is found by the first order transition matrix $\mathbf{P}_x$. The transition probabilities $p\left(x_t^{(k)} \middle| x_{t-1}^{(k)*} \right)$ will thus make a $(L^k \times L^k)$ transition matrix $\mathbf{P}_{x^{(k)}}$ defined by $\mathbf{P}_x$. As in Expression (3.13), for the first order approximation, we want to approximate the full posterior

distribution by an approximation in the acquisition likelihood distribution, but now for a higher order k,

$$\hat{p}^{(k)}\left(\mathbf{x}^{(k)}\middle|\,\mathbf{d}\right) = C_{d*}\cdot\prod_{t=k}^{T} l_d^{(k)}\left(x_t^{(k)}\right) p\left(x_t^{(k)}\middle|\,x_{t-1}^{(k)*}\right) \quad . \tag{3.21}$$

Here

$$l_d^{(k)}\left(x_t^{(k)}\right) = \int \frac{p_*\left(r_t^{(k)}\middle|\,\mathbf{d}\right)}{p_*\left(r_t^{(k)}\right)} p\left(r_t^{(k)}\middle|\,x_t^{(k)}\right) dr_t^{(k)} \quad , \tag{3.22}$$

according to Expression (3.19) for the first order, i.e. for $k = 1$.

The kth order response variable in Expression (3.22) is $r_t^{(k)} = (r_{t-k+1}, \ldots, r_t)$, and the kth order marginal response likelihood $p\left(r_t^{(k)}\middle|\,x_t^{(k)}\right)$ is simply

$$\left[r_t^{(k)}\middle|\,x_t^{(k)}\right] \;\sim\; p\left(r_t^{(k)}\middle|\,x_t^{(k)}\right) = \prod_{i=t-k+1}^{k} p(r_i|x_i) \quad , \tag{3.23}$$

by the assumption of independent marginals in Expression (2.8). The Gaussian distributions $p_*\left(r_t^{(k)}\middle|\,\mathbf{d}\right)$ and $p_*\left(r_t^{(k)}\right)$, which makes the kth order marginal acquisition likelihood distribution, are the kth marginals from Expressions (3.16) and (3.17) respectively. For larger k, more of the smoothness in the acquisition likelihood model is captured, as these kth order marginals utilize more rows in the covariance matrices $\mathbf{\Sigma_{r|d}}$ and $\mathbf{\Sigma_r}$ in Expressions (3.16) and (3.17) respectively. As for the first order approximation in Expression (3.19), we use the full observation vector $\mathbf{d}$ in every element $l_d^{(k)}\left(x_t^{(k)}\right)$, and the approximations are only made in the likelihood distribution.

By Expression (3.21), our approximated posterior distribution is a non-stationary kth order Markov chain,

$$\begin{aligned}\hat{p}^{(k)}\left(\mathbf{x}^{(k)}|\mathbf{d}\right) &= \prod_{t=k}^{T} \hat{p}^{(k)}\left(x_t^{(k)}\middle|x_{t-1}^{(k)*}, \mathbf{d}\right) \\ &= \prod_{t=k}^{T}\left(\prod_{i=1}^{k-1} I(x_{t-k+i} = x_{t-k+i}^*)\right)\hat{p}\left(x_t|x_{t-1}^*, \mathbf{d}\right) \quad .\end{aligned} \tag{3.24}$$

The initial posterior transition probabilities are

$$\hat{p}^{(k)}\left(x_k^{(k)}\middle|x_{k-1}^{(k)*}, \mathbf{d}\right) = \hat{p}\left(x_k^{(k)}\middle|\,\mathbf{d}\right) = C_{x_k^{(k)}}\cdot l_d^{(k)}\left(x_k^{(k)}\right)\prod_{j=1}^{k} p(x_j|x_{j-1}) \quad , \tag{3.25}$$

where $p(x_j|x_{j-1})$ are simply transition probabilities in the first order transition matrix, $\mathbf{P}_x$.

For $k = 1$, we obtain the first order algorithm as discussed in the previous section. For $k = T$ we have, by Expressions (3.21) and (3.22),

$$\begin{aligned}\hat{p}^{(T)}\left(\mathbf{x}^{(T)}\middle|\,\mathbf{d}\right) &= C_{d*}\cdot\prod_{t=T}^{T} l_d^{(T)}\left(x_t^{(T)}\right) p\left(x_t^{(T)}\middle|\,x_{t-1}^{(T)}\right) = C_{d*}\cdot l_d^{(T)}(\mathbf{x})p(\mathbf{x}) \\ &= C_{d*}\cdot\int \frac{p_*(\mathbf{r}|\mathbf{d})}{p_*(\mathbf{r})} p(\mathbf{r}|\mathbf{x})d\mathbf{r} = p(\mathbf{x}|\mathbf{d}) \quad ,\end{aligned}$$

i.e. the full posterior. Here we have used the properties in Expressions (3.20) and (3.23). The algorithm for $k = T$ would use to much cpu-time as it would require too many operations in the forward-backward algorithm. The computations would also be memory intensive, as we gain an $(L^T \times L^T)$ transition matrix, even though this is sparse.

We can now compute the approximated posterior distribution by Algorithm 1 in Section 3.1. By Expressions (3.24) and (3.25) and the definition of the backward probability in Expression (3.2), we have the approximate kth order posterior distribution

$$\hat{p}^{(k)}\left(\mathbf{x}^{(k)}\middle|\,\mathbf{d}\right) = \prod_{t=k}^{T} \hat{p}^{(k)}\left(x_t^{(k)}\middle|\, x_{t-1}^{k*}, \mathbf{d}\right) = p_b^{(k)}\left(x_k^{(k)}\right) \cdot \prod_{t=k+1}^{T} p_b^{(k)}\left(x_t^{(k)}\middle|\, x_{t-1}^{(k)*}\right) \quad , \tag{3.26}$$

similar to Expression (3.11) for the first order simple HMM. We can simulate approximate realizations, $[\mathbf{x}|\mathbf{d}]^s = (x_1^s, \ldots, x_T^s)$, by the procedure in Algorithm 2 in Section 3.1. The locationwise MAPs are by Expression (3.12) approximated by

$$[\hat{x}_t|\mathbf{d}]_{MAP} \quad = \quad \arg\max_{x_t}\{p_b(x_t)\} = \arg\max_{x_t}\left\{\sum_{x_i^{(k)}\setminus x_t} p_b\left(x_i^{(k)}\right)\right\} \quad , \tag{3.27}$$

with the total MAP solution $[\hat{\mathbf{x}}|\mathbf{d}]_{MAP} = \{[\hat{x}_t|\mathbf{d}]_{MAP}; t = 1, \ldots, T\}$. The global MAPs can likewise be computed by Alorithm 3 in Section 3.1.

Chapter 4

Parameter estimation

In this chapter, we will assume that the prior and the likelihood distributions depend on the parameters $\boldsymbol{\theta} = (\boldsymbol{\theta_x}, \boldsymbol{\theta_d}, \boldsymbol{\theta_r})$. We will thus have a prior model by $p(\mathbf{x}; \boldsymbol{\theta_x})$ and a likelihood model by $p(\mathbf{d}|\mathbf{x}; \boldsymbol{\theta_d}, \boldsymbol{\theta_r}) = \int p(\mathbf{d}|\mathbf{r}; \boldsymbol{\theta_d}) p(\mathbf{r}|\mathbf{x}; \boldsymbol{\theta_r}) d\mathbf{r}$. Further, we will discuss two ways on how to estimate these parameters and how to evaluate the estimates. We begin by deriving the marginal distribution $p(\mathbf{d})$ of the observed data, as it plays an important part in the estimation procedure.

4.1 Marginal likelihood distribution

In the simple HMM in Figure 2.1, the full variable space is $[\mathbf{d}, \mathbf{x}, \boldsymbol{\theta}]$, while it is $[\mathbf{d}, \mathbf{r}, \mathbf{x}, \boldsymbol{\theta}]$ for the more complex HMM in Figure 2.2. Here the random variables $\mathbf{d}$, $\mathbf{r}$ and $\mathbf{x}$ depend on the parameter $\boldsymbol{\theta}$ in the prior and the likelihood models. The parameters in $\boldsymbol{\theta}$ can be unknown constants or random variables themselves. By computing the marginal likelihood, $p(\mathbf{d}; \boldsymbol{\theta})$, the state variable, $\mathbf{x}$, and the response variable, $\mathbf{r}$, are integrated out. This way we can make inference on the parameter $\boldsymbol{\theta}$ based on the observations in $\mathbf{d}$ only. For notational ease, we omit $\boldsymbol{\theta}$ in the following developments, hence $p(\mathbf{d}) = p(\mathbf{d}; \boldsymbol{\theta})$, i.e. the marginal likelihood distribution is parameter dependent and the parameters are considered as constants.

The marginal likelihood distribution, $p(\mathbf{d})$, can be written as

$$p(\mathbf{d}) = p(d_T|d_{T-1}, \ldots, d_1) p(d_{T-1}|d_{T-2}, \ldots, d_1) \cdots p(d_3|d_2, d_1) p(d_2|d_1) p(d_1) \ . \tag{4.1}$$

For the simple HMM in Figure 2.1 with likelihood model by Expression (2.4), we have an initial marginal distribution $p(d_1) = \sum_{x_1} p(d_1|x_1) p(x_1)$ which appears in the initial marginal posterior distribution

$$p(x_1|d_1) = \frac{p(d_1|x_1) p(x_1)}{p(d_1)} \ . \tag{4.2}$$

For $t = T, \ldots, 3$ we have

$$\begin{aligned} p(x_t|d_t, \ldots, d_1) &= \frac{p(d_t|x_t) p(x_t|d_{t-1}, \ldots, d_1) p(d_{t-1}, \ldots, d_1)}{p(d_t|d_{t-1}, \ldots, d_1) p(d_{t-1}, \ldots, d_1)} \\ &= \frac{p(d_t|x_t) \sum_{x_{t-1}} p(x_t|x_{t-1}, d_{t-1}, \ldots, d_1) p(x_{t-1}|d_{t-1}, \ldots, d_1)}{p(d_t|d_{t-1}, \ldots, d_1)} \\ &= \frac{p(d_t|x_t) \sum_{x_{t-1}} p(x_t|x_{t-1}) p(x_{t-1}|d_{t-1}, \ldots, d_1)}{p(d_t|d_{t-1}, \ldots, d_1)} \ , \end{aligned} \tag{4.3}$$

which are the marginal forward probabilities in Expression (3.5) with a defined constant, $C_{x_{t-1,t}} = 1/p(d_t|d_{t-1},\ldots,d_1)$. With initial posterior probability by Expression (4.2) and marginal forward probability by Expression (4.3), we thus have by recursion

$$\begin{aligned}
1 &= \frac{1}{p(d_T|d_{T-1},\ldots,d_1)} \sum_{x_T} \sum_{x_{T-1}} p(d_T, x_T, x_{T-1}|d_{T-1},\ldots,d_1) \\
&= \frac{1}{p(d_T|d_{T-1},\ldots,d_1)} \sum_{x_T} \sum_{x_{T-1}} p(d_T|x_T)p(x_T|x_{T-1})p(x_{T-1}|d_{T-1},\ldots,d_1) \\
&= \frac{1}{p(d_T|d_{T-1},\ldots,d_1)} \sum_{x_T} \sum_{x_{T-1}} p(d_T|x_T)p(x_T|x_{T-1}) \frac{p(d_{T-1}|x_{T-1}) \sum_{x_{T-2}} p(x_{T-2}|d_{T-2},\ldots,d_1)}{p(d_{T-1}|d_{T-2},\ldots,d_1)} \\
&= \ldots = \frac{1}{p(d_T|d_{T-1},\ldots,d_1)\cdots p(d_2|d_1)p(d_1)} \sum_{x_1} \ldots \sum_{x_T} \prod_{t=1}^{T} p(d_t|x_t)p(x_t|x_{t-1}) \quad .
\end{aligned}$$

By Expression (4.1), we can thus write the marginal likelihood distribution as

$$p(\mathbf{d}) = \sum_{x_1} \ldots \sum_{x_T} \prod_{t=1}^{T} p(d_t|x_t)p(x_t|x_{t-1}) \quad . \tag{4.4}$$

During the forward-backward algorithm described in Section 3.1, we compute all the constants defined by $C_{x_{t-1,t}} = 1/p(d_t|d_{t-1},\ldots,d_1)$ in Expression (4.3). We can thus compute the marginal likelihood distribution in Expression (4.1) as

$$p(\mathbf{d}) = \frac{1}{\left[\prod_{t=1}^{T} C_{x_{t-1,t}}\right]} \quad , \tag{4.5}$$

where $C_{x_{0,1}} = C_{x_1}$. The marginal log-likelihood is thus

$$l(\mathbf{d}) = \log[p(\mathbf{d})] = -\sum_{t=1}^{T} \log\left(C_{x_{t-1,t}}\right) \quad , \tag{4.6}$$

and this computation can be implemented in the forward-backward algorithm. As the marginal likelihood, the log-likelihood will be parameter dependent, i.e. $l(\mathbf{d};\boldsymbol{\theta})$.

We now look at the more complex HMM in Figure 2.2 approximated by the kth order P-P ratio deconvolution algorithm described in Section 3.3. The model has approximated likelihoods by Expression (3.22). With approximated likelihoods, the marginal likelihood distribution will be an approximation as well. One can by the same reasoning as above show that the kth order approximated marginal likelihood distribution now becomes

$$\hat{p}^{(k)}(\mathbf{d}) = \frac{1}{\left[\prod_{t=k}^{T} C_{x^{(k)}_{t-1*,t}}\right]} \quad , \tag{4.7}$$

where $C_{x^{(k)}_{k-1*,k}} = C_{x^{(k)}_{k}}$. Thus, the kth order approximated marginal log-likelihood is

$$\hat{l}^{(k)}(\mathbf{d}) = \log\left[\hat{p}^{(k)}(\mathbf{d})\right] = -\sum_{t=k}^{T} \log\left(C_{x^{(k)}_{t-1*,t}}\right) \quad , \tag{4.8}$$

and will also depend on $\boldsymbol{\theta}$, i.e. $\hat{l}^{(k)}(\mathbf{d};\boldsymbol{\theta})$. The constants $C_{x^{(k)}_{t-1*,t}}$ are computed in the forward recursions during the forward-backward algorithm described in Algorithm 1 in Section 3.1.

4.2 Parameter estimation methods

In this section, we introduce two similar approaches for estimating the most likely value of $\boldsymbol{\theta}$. These are maximum marginal likelihood estimation and maximum a posteriori prediction. The major difference between the approaches is that we consider the parameters as unknown constants in the first method, and as unknown random variables in the second. An algorithm for computing the parameter estimate is also derived, which can be applied for both these methods. We then define the Fisher information, a quantity for evaluation of the parameter estimates by their corresponding log-likelihood values.

4.2.1 Parameter maximum marginal likelihood estimation (MMLE)

In this subsection, we assume that the parameters in $\boldsymbol{\theta}$ are unknown constants. The approximated marginal likelihood distribution in Expression (4.7), for the complex HMM in Figure 2.2, is parameter dependent, i.e. $\hat{p}^{(k)}(\mathbf{d};\boldsymbol{\theta})$. Hence the constants computed by the forward iterations in the forward-backward algorithm will by Expression (4.7) also depend on the parameters, i.e. $C_{x^{(k)}_{t-1*,t}} = \left[C_{x^{(k)}_{t-1*,t}};\boldsymbol{\theta}\right]$. We can then find the kth order maximum marginal likelihood estimate (MMLE) of the parameters,

$$\widehat{\boldsymbol{\theta}}^{(k)}_{MMLE} = \arg\max_{\boldsymbol{\theta}} \left\{\hat{l}^{(k)}(\mathbf{d};\boldsymbol{\theta})\right\} = \arg\max_{\boldsymbol{\theta}} \left\{-\sum_{t=k}^{T} \log\left(\left[C_{x^{(k)}_{t-1*,t}};\boldsymbol{\theta}\right]\right)\right\} \quad . \tag{4.9}$$

Here $\hat{l}^{(k)}(\mathbf{d};\boldsymbol{\theta})$ is the kth order approximate marginal log-likelihood in Expression (4.8).

4.2.2 Parameter estimation by maximum a posteriori prediction (MAP)

We now assume that the parameters in $\boldsymbol{\theta}$ are random variables with a prior distribution $\pi(\boldsymbol{\theta})$. The approximated marginal likelihood in Expression (4.7) is then a conditional probability, $\hat{p}^{(k)}(\mathbf{d}|\boldsymbol{\theta})$, with $C_{x^{(k)}_{t-1*,t}} = \left[C_{x^{(k)}_{t-1*,t}}\middle|\boldsymbol{\theta}\right]$. As $\hat{p}^{(k)}(\boldsymbol{\theta}|\mathbf{d}) \propto \hat{p}^{(k)}(\mathbf{d}|\boldsymbol{\theta}) \cdot \pi(\boldsymbol{\theta})$, we first define the kth order parameter posterior logarithm value

$$\hat{l}^{(k)*}(\boldsymbol{\theta}|\mathbf{d}) = \hat{l}^{(k)}(\mathbf{d}|\boldsymbol{\theta}) + \log\left[\pi(\boldsymbol{\theta})\right] \;\propto\; \log\left[\hat{p}^{(k)}(\boldsymbol{\theta}|\mathbf{d})\right] \quad . \tag{4.10}$$

The kth order maximum a posteriori prediction (MAP) of $\boldsymbol{\theta}$ is then

$$\begin{aligned} \widehat{\boldsymbol{\theta}}^{(k)}_{MAP} &= \arg\max_{\boldsymbol{\theta}} \left\{\hat{p}^{(k)}(\boldsymbol{\theta}|\mathbf{d})\right\} = \arg\max_{\boldsymbol{\theta}} \left\{\hat{l}^{(k)*}(\boldsymbol{\theta}|\mathbf{d})\right\} \\ &= \arg\max_{\boldsymbol{\theta}} \left\{\log[\pi(\boldsymbol{\theta})] - \sum_{t=k}^{T} \log\left(\left[C_{x^{(k)}_{t-1*,t}}\middle|\boldsymbol{\theta}\right]\right)\right\} \quad . \end{aligned} \tag{4.11}$$

This is similar to the MMLE in Expression (4.9), but we have included the parameter prior distribution in the estimation. This parameter prior distribution will depend on some hyperparameters $\boldsymbol{\alpha}$, which should be fixed constants. If they were not fixed, $\boldsymbol{\alpha}$ could depend on the hyperhyperparameters $\boldsymbol{\alpha}^*$ which again could depend on the hyperhyperhyperparameters $\boldsymbol{\alpha}^{**}$ and so on. This could lead to an infinite dimensional hyperparameter space which makes parameter estimation impossible.

4.2.3 Estimation optimization algorithm

We assume that the parameter $\boldsymbol{\theta}$ is a $(M \times 1)$ vector, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)$. Every parameter $\{\theta_m;\ m = 1, \ldots, M\}$ is now discretized on $\mathcal{D} \in \mathbb{R}^1$ by an independent lattice with length n_m, ${\mathcal{L}_\mathcal{D}}^{\theta_m} = \{h_1^{\theta_m}, \ldots, h_{n_m}^{\theta_m}\}$. These intervals should be restricted by the definition space of the respective parameters, e.g. the variance parameters can not be negative. To ease the notations, we choose $n_m = n\ \forall\ m = 1, \ldots, M$ in the rest of this thesis. Together, the lattices make the M-dimensional parameter grid $\{\mathcal{L}_\mathcal{G} = {\mathcal{L}_\mathcal{D}}^{\theta_1} \times \ldots \times {\mathcal{L}_\mathcal{D}}^{\theta_M}; \mathcal{G} \in \mathbb{R}^M\}$. With this discretization, we should run the forward-backward algorithm for every parameter set in $\mathcal{L}_\mathcal{G}$, i.e. n^M runs, obtaining n^M log-likelihood values by Expression (4.8). The MMLE and MAP estimate will then be the parameter, $\boldsymbol{\theta}_g \in \mathcal{L}_\mathcal{G}$, that maximizes Expressions (4.9) and (4.11) respectively. The two estimates need not be the same parameter $\boldsymbol{\theta}_g$. The procedure is presented in the following algorithm.

Algorithm 4: Parameter estimation on grid

- Initiate:
  - For every $\{\theta_m;\ m = 1, \ldots, M\}$ initiate lattices ${\mathcal{L}_\mathcal{D}}^{\theta_m} = \{h_1^{\theta_m}, \ldots, h_n^{\theta_m}\}$ with $|h_n^{\theta_m} - h_1^{\theta_m}|$ sufficiently large.
  - $\mathcal{L}_\mathcal{G} = {\mathcal{L}_\mathcal{D}}^{\theta_1} \times \ldots \times {\mathcal{L}_\mathcal{D}}^{\theta_M}$
- For every $\boldsymbol{\theta}_g \in \mathcal{L}_\mathcal{G}$:
  - Run the forward-backward algorithm with parameters $\boldsymbol{\theta}_g$.
  - Compute the marginal log-likelihood $\hat{l}_g^{(k)} = \hat{l}^{(k)}(\mathbf{d}; \boldsymbol{\theta}_g)$ by Expression (4.8).
- Parameter estimates:
  - $\widehat{\boldsymbol{\theta}}_{MMLE} = \arg\max_{\boldsymbol{\theta}_g \in \mathcal{L}_\mathcal{G}} \left\{\hat{l}_g^{(k)}\right\}$
  - $\widehat{\boldsymbol{\theta}}_{MAP} = \arg\max_{\boldsymbol{\theta}_g \in \mathcal{L}_\mathcal{G}} \left\{\hat{l}_g^{(k)} + \log\left[\pi(\boldsymbol{\theta}_g)\right]\right\}$

After running this algorithm on an initial grid, ${\mathcal{L}_\mathcal{G}}^1$, we can run it again on a more narrow grid, ${\mathcal{L}_\mathcal{G}}^2$, centered around $\widehat{\boldsymbol{\theta}}_{MMLE}^{(k)} \in {\mathcal{L}_\mathcal{G}}^1$ or $\widehat{\boldsymbol{\theta}}_{MAP}^{(k)} \in {\mathcal{L}_\mathcal{G}}^1$. This could be repeated until we have a grid, ${\mathcal{L}_\mathcal{G}}^j$, where all $\boldsymbol{\theta}_g \in {\mathcal{L}_\mathcal{G}}^j$ are satisfactory close. This procedure should be done separately for MMLE and MAP estimation as they not necessarily give the same estimate.

For the MAP estimation, we can include approximation of the marginal parameter posterior distributions, $p(\theta_m|\mathbf{d})$, $\theta_m \in \boldsymbol{\theta}$, in Algorithm 4. The parameter intervals chosen should thus be relatively wide, with a sufficient number of elements, n. To simplify the computations, the grid's step length should be chosen to be equal, i.e. $|h_j^{\theta_m} - h_{j-1}^{\theta_m}| = \delta_{\theta_m}\ \forall\ j$. The approximate marginal discrete parameter posterior pdf is then

$$\hat{p}^{(k)}(\theta_m|\mathbf{d}) = C_{\theta_m} \cdot \sum_{\boldsymbol{\theta}\setminus\theta_m} \exp\left\{\hat{l}^{(k)}(\mathbf{d}|\boldsymbol{\theta}) + \log\left[\pi(\boldsymbol{\theta})\right]\right\} \quad , \tag{4.12}$$

where $C_{\theta_m} = 1/\left[\delta_{\theta_m} \cdot \sum_{j=1}^n \hat{p}^{(k)}(\theta_m = h_j^{\theta_m}|\mathbf{d})\right]$ is an added constant which assures that the distribution sums to unity. The marginal parameter MAP estimate is then simply found by $\hat{\theta}_{m,MAP}^{(k)} = \arg\max_{\theta_m}\left\{\hat{p}^{(k)}(\theta_m|\mathbf{d})\right\}$.

One challenge in Algorithm 4 is to set the initial parameter intervals, $\{h_1^{\theta_m}, \ldots, h_n^{\theta_m}\}\ \forall\ m$, as the optimal parameter values might fall outside of these intervals. We should thus choose $|h_n^{\theta_m} - h_1^{\theta_m}|$ relatively large. With many parameters, the algorithm gets computer demanding as we run the forward-backward algorithm n^M times. Another challenge is thus to find a reasonable interval length n.

4.2.4 Parameter estimate evaluation

For a parameter set, $\boldsymbol{\theta}= (\theta_1, \ldots, \theta_M)$, with marginal log-likelihood value, $l(\mathbf{d}; \boldsymbol{\theta})$, by Expression (4.6), the Fisher information, or information number, is given as,

$$I(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}} \left[\left(\frac{\partial}{\partial \boldsymbol{\theta}} l(\mathbf{d}; \boldsymbol{\theta}) \right)^2 \middle| \boldsymbol{\theta} \right] = -E_{\boldsymbol{\theta}} \left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2} l(\mathbf{d}; \boldsymbol{\theta}) \middle| \boldsymbol{\theta} \right] \ , \tag{4.13}$$

see Reference [5]. This is the second moment of the partial derivative of the log-likelihood function with respect to the parameter. The last equality in Expression (4.13) holds when $p(\mathbf{d}; \boldsymbol{\theta})$ belongs to the family of exponential distributions, i.e. the Fisher information is then the negative expectation of the second derivative of the log-likelihood function with respect to the parameter, see Reference [5]. The Fisher information gives a bound on the variance of any unbiased parameter estimator $\hat{\boldsymbol{\theta}}_u$ of $\boldsymbol{\theta}$,

$$[I(\boldsymbol{\theta})]^{-1} \leq \mathrm{Var}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_u) \ , \tag{4.14}$$

see Reference [5]. Large information numbers thus makes a smaller bound on the estimators variance, and the best unbiased estimator of $\boldsymbol{\theta}$ will have variance $[I(\boldsymbol{\theta})]^{-1}$.

The complex HMM in Figure 2.2 is approximated by the kth order P-P ratio deconvolution algorithm in Section 3.3. The corresponding marginal likelihood distribution, by Expression (4.7), is an approximation as well, but is exact with respect to the approximated model. As we compute the approximate marginal log-likelihood values numerically, likewise is their approximate second derivatives computed numerically. Assuming that $p(\mathbf{d}; \boldsymbol{\theta})$ belongs to the family of exponential distributions, the numerical approximation of the kth order Fisher information is then by Expression (4.13)

$$\hat{I}^{(k)}(\boldsymbol{\theta}) = -\widehat{\frac{\partial^2}{\partial \boldsymbol{\theta}^2}} \hat{l}^{(k)}(\mathbf{d}; \boldsymbol{\theta}) \ . \tag{4.15}$$

Here $\hat{l}^{(k)}(\mathbf{d}|\boldsymbol{\theta})$ is the kth order approximate marginal log-likelihood by Expression (4.8).

We assume that the parameter MMLE in Expression (4.9) is an unbiased estimator. For univariate parameter estimation, an approximate parameter $100(1-\alpha)\%$ minimum Gaussian-based confidence interval can by Expression (4.14) be computed by

$$\theta \ \in \ \left[\widehat{\theta}^{(k)}_{MMLE} - z_{\alpha/2} \cdot \left[\hat{I}^{(k)}(\theta) \right]^{-1} \ , \ \widehat{\theta}^{(k)}_{MMLE} + z_{\alpha/2} \cdot \left[\hat{I}^{(k)}(\theta) \right]^{-1} \right] \ , \tag{4.16}$$

see Reference [22]. Here $z_{\alpha/2}$ is the $\alpha/2$-quantile in the standard Gaussian distribution. Large values of the Fisher information indicate small variance and thus result in small confidence intervals. Similarly, when estimating M parameters, approximate parameter $100(1-\alpha)\%$ minimum Gaussian-based confidence regions are defined by M-dimensional ellipsoids satisfying

$$(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{(k)}_{MMLE})' \left[\hat{I}^{(k)}(\boldsymbol{\theta}) \right] (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{(k)}_{MMLE}) \leq \chi^2_M(\alpha) \ , \tag{4.17}$$

see Reference [11]. Here $\chi^2_M(\alpha)$ is the upper (100α)th percentile of a chi-square distribution with M degrees of freedom, and $\hat{I}^{(k)}(\boldsymbol{\theta})$ corresponds to a $(M \times M)$ Gaussian covariance matrix. The ellipsoid in Expression (4.17) defining the confidence region boundary is centered in $\widehat{\boldsymbol{\theta}}^{(k)}_{MMLE}$ and have semi axes $\pm\sqrt{\chi^2_M(\alpha)\lambda^{(k)}_i}\mathbf{e}^{(k)}_i$, where $\left\{ (\lambda^{(k)}_1, \mathbf{e}^{(k)}_1), \ldots, (\lambda^{(k)}_M, \mathbf{e}^{(k)}_M) \right\}$ are the eigenvalue-eigenvector pairs for $\left[\hat{I}^{(k)}(\boldsymbol{\theta}) \right]^{-1}$, see Reference [11]. As we expect better model

approximations for higher orders k, the approximate Fisher information in Expression (4.15) should become smaller for increasing orders. The resulting parameter confidence intervals and regions, by Expressions (4.16) and (4.17) respectively, should thus become smaller for increasing orders as well.

4.3 Parameters in the model

We now introduce the parameters $\boldsymbol{\theta}_x$, $\boldsymbol{\theta}_r$ and $\boldsymbol{\theta}_d$ in the prior, the response and the acquisition likelihood model respectively. We thus have a prior distribution by $p(\mathbf{x};\boldsymbol{\theta}_x)$, a response likelihood distribution by $p(\mathbf{r}|\mathbf{x};\boldsymbol{\theta}_r)$ and an acquisition likelihood distribution by $p(\mathbf{d}|\mathbf{r};\boldsymbol{\theta}_d)$. We deduce the MMLE and MAP estimate for these distributions, suggesting conjugate parameter prior distributions in the MAP approach. The dependencies between the parameters in the models and the model distributions by Expressions (2.3), (2.8) and (2.10) are presented in Figure 4.1. The parameters shown will be discussed in more detail.

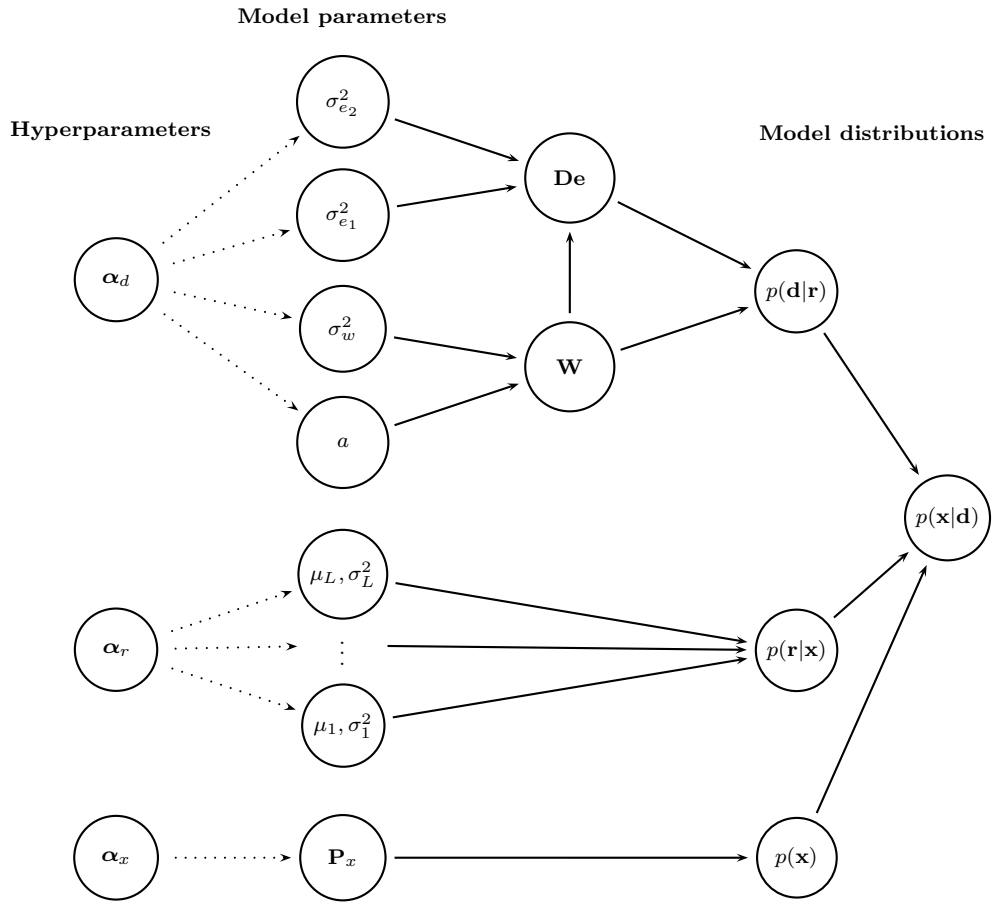


Figure 4.1: *Relations between the parameters in the full model and the model's distributions. The lines between the hyperparameters and the model parameters are dotted as these relations are valid only for the parameter MAP estimation.*

4.3.1 Parameters in the prior model

We assume that the prior model still holds the Markov property as in Expression (2.3), but instead of fixing the transition probabilities in $\mathbf{P}_x$ they are now unknown parameters,

$$\boldsymbol{\theta}_x = \mathbf{P}_x = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,L-1} & p_{1,L} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,L-1} & p_{2,L} \\ \vdots & \vdots & & \vdots & \vdots \\ p_{L,1} & p_{L,2} & \cdots & p_{L,L-1} & p_{L,L} \end{pmatrix} \quad . \tag{4.18}$$

The initial marginal prior distribution, $p(x_1)$, will also depend on $\boldsymbol{\theta}_x$ as we still assume it to be the stationary distribution of $\mathbf{P}_x$. We also still assume that the transition matrix, $\mathbf{P}_x$, is stationary with state space $\Omega_x : \{1, 2, \ldots, L\}$. Every row in $\mathbf{P}_x$ must sum to unity. We will thus have $M = L \cdot (L-1)$ unknown parameters,

$$\boldsymbol{\theta}_x = \{p_{i,j} < 1 \; \forall \; (i,j);\; i = 1, \ldots, L,\; j = 1, \ldots, L-1\} \quad ,$$

with $p_{i,L} = 1 - \sum_{j=1}^{L-1} p_{i,j}$. The kth order MMLE of $\boldsymbol{\theta}_x$ is then, by Expression (4.9),

$$\widehat{\boldsymbol{\theta}}_{x,MMLE}^{(k)} = \arg\max_{\boldsymbol{\theta}_x} \left\{ -\sum_{t=k}^{T} \log\left(\left[C_{x_{t-1*,t}^{(k)}}; \boldsymbol{\theta}_x \right] \right) \right\} \quad ,$$

and can be computed by Algorithm 4 in Section 4.2.3. When we run the algorithm in this case, we assume that all other parameters in the model besides from $\boldsymbol{\theta}_x$ are fixed.

In a Bayesian setting, we can choose to assign an independent Dirichlet distribution to every row in $\mathbf{P}_x$. The Dirichlet distribution is described in Appendix A.2, see also Reference [13]. Each row in $\mathbf{P}_x$, as we have defined it, will be probability parameters in a multinomial distribution, see Reference [20]. The Dirichlet distribution is chosen as it is conjugate prior to the multinomial distribution, see Appendix A.2. The parameter prior pdf for $\boldsymbol{\theta}_x = \mathbf{P}_x$ is then

$$\begin{aligned} \pi(\boldsymbol{\theta}_x) &= p(p_{1,1}, \ldots, p_{1,L}, p_{2,1}, \ldots, p_{L,L}) = \prod_{i=1}^{L} p(p_{i,1}, \ldots, p_{i,L}; \alpha_{i,1}, \ldots, \alpha_{i,L}) \\ &= \prod_{i=1}^{L} \frac{1}{B(\boldsymbol{\alpha}_i)} \prod_{j=1}^{L} p_{i,j}^{\alpha_{i,j}-1} \quad . \end{aligned} \tag{4.19}$$

Here $B(\boldsymbol{\alpha}_i) = \frac{\prod_{i=1}^{L} \Gamma(\alpha_{i,j})}{\Gamma(\sum_{j=1}^{L} \alpha_{i,j})}$ is the multinomial beta function, where $\Gamma(\cdot)$ is the gamma function, and the $\boldsymbol{\alpha}_i$'s are fixed hyperparameter vectors. All hyperparameters in the prior will then make an $(L^2 \times 1)$ vector, $\boldsymbol{\alpha}_x = (\alpha_{1,1}, \ldots, \alpha_{L,L})$. Reference [8] proposes a model where the rows in $\mathbf{P}_x$ are given symmetric Dirichlet distributions, i.e. $\alpha_{i,j} = \alpha \; \forall \; i, j$. The parameter prior pdf is then

$$\pi(\boldsymbol{\theta}_x) = \frac{1}{[B(\alpha)]^L} \prod_{i=1}^{L} \prod_{j=1}^{L} p_{i,j}^{\alpha-1} \quad , \tag{4.20}$$

where $B(\alpha) = \frac{[\Gamma(\alpha)]^L}{\Gamma(L\alpha)}$. Then there is only one hyperparameter, $\boldsymbol{\alpha}_x = \alpha$, to fix. Assuming that all other parameters are fixed, the kth order MAP estimate with parameter prior distribution from either Expressions (4.19) or (4.20) is, by Expression (4.11),

$$\widehat{\boldsymbol{\theta}}_{x,MAP}^{(k)} = \arg\max_{\boldsymbol{\theta}_x} \left\{ \log[\pi(\boldsymbol{\theta}_x)] - \sum_{t=k}^{T} \log\left(\left[C_{x_{t-1*,t}^{(k)}} \middle| \boldsymbol{\theta}_x \right] \right) \right\} \quad .$$

The MAP estimate can be computed by Algorithm 4 in Section 4.2.3.

4.3.2 Parameters in the response likelihood model

Every element in the response likelihood model are independent univariate Gaussian distributed by Expression (2.7), $[r_t|x_t] \sim N_1(\mu_{x_t}, \sigma^2_{x_t})$. Here $(\mu_{x_t}, \sigma_{x_t})$ can take L possible values, $(\mu_{x_t}, \sigma_{x_t}) \in \{(\mu_1, \sigma_1), \ldots, (\mu_L, \sigma_L)\}$. Every parameter set, (μ_i, σ_i^2), is the response mean and variance corresponding to the ith class in the categorical LF-state space $\Omega_x : \{1, \ldots, i, \ldots, L\}$. There are thus $2L$ unknown parameters to determine in the response likelihood, see also Figure 4.1,

$$\boldsymbol{\theta}_r = (\mu_1, \ldots, \mu_L, \sigma_1, \ldots, \sigma_L) \ . \tag{4.21}$$

The kth order MMLE of $\boldsymbol{\theta}_r$ is, by Expression (4.9),

$$\widehat{\boldsymbol{\theta}}^{(k)}_{r,MMLE} = \arg\max_{\boldsymbol{\theta}_r} \left\{ -\sum_{t=k}^{T} \log\left(\left[C_{x^{(k)}_{t-1*,t}}; \boldsymbol{\theta}_r \right] \right) \right\} \ ,$$

and can be computed by Algorithm 4 in Section 4.2.3 assuming that all other parameters are fixed. If we assume that the variance parameters are equal, i.e. $\sigma_i = \sigma_x \ \forall \ i$, there will be $L + 1$ unknown parameters, i.e. $\boldsymbol{\theta}_r = (\mu_1, \ldots, \mu_L, \sigma_x)$. This would simplify Algorithm 4, gaining a parameter grid of lower dimension.

As for the prior model, we can use a Bayesian setting for the response likelihood model parameters in Expression (4.21). We choose to assign independent prior Gaussian distributions to every conditional mean parameter, $[\mu_i|\sigma_i^2] \sim N_1(\gamma_i, \sigma_i^2/\lambda_i)$ and independent prior Inverse Gamma distribution to every variance parameter, $\sigma_i^2 \sim IG(\alpha_i, \beta_i)$. The Inverse Gamma distribution is described in Appendix A.3. These particular prior distributions are chosen as they are conjugate prior distributions for the Gaussian distribution, see Appendix A.1. The full response parameter prior is then normal-scaled inverse gamma distributed,

$$\pi(\boldsymbol{\theta}_r) = \prod_{i=1}^{L} \left[\frac{\sqrt{\lambda_i}}{\sqrt{2\pi}\sigma_i} \frac{1}{\beta_i^{\alpha_i}\Gamma(\alpha_i)} \left(\frac{1}{\sigma_i^2} \right)^{\alpha_i+1} \exp\left(-\frac{2 + \beta_i\lambda_i(\mu_i - \gamma_i)^2}{2\beta_i\sigma_i^2} \right) \right] \ , \tag{4.22}$$

with the $(4L \times 1)$ hyperparameter vector $\boldsymbol{\alpha}_r = (\gamma_1, \ldots, \gamma_L, \lambda_1, \ldots, \lambda_L, \alpha_1, \ldots, \alpha_L, \beta_1, \ldots, \beta_L)$. If we assume equal variance, i.e. $\sigma_i = \sigma_x \ \forall \ i = 1, \ldots, L$ with $\sigma_x^2 \sim IG(\alpha, \beta)$, the parameter prior pdf becomes

$$\pi(\boldsymbol{\theta}_r) = \prod_{i=1}^{L} \left[\frac{\sqrt{\lambda_i}}{\sqrt{2\pi}\sigma_x} \frac{1}{\beta^{\alpha}\Gamma(\alpha)} \left(\frac{1}{\sigma_x^2} \right)^{\alpha+1} \exp\left(-\frac{2 + \beta\lambda_i(\mu_i - \gamma_i)^2}{2\beta\sigma_x^2} \right) \right] \ , \tag{4.23}$$

with the $((2L + 2) \times 1)$ hyperparameter vector $\boldsymbol{\alpha}_r = (\gamma_1, \ldots, \gamma_L, \lambda_1, \ldots, \lambda_L, \alpha, \beta)$. The kth order MAP estimate with parameter prior pdf from either Expressions (4.22) or (4.23), is, by Expression (4.11),

$$\widehat{\boldsymbol{\theta}}^{(k)}_{r,MAP} = \arg\max_{\boldsymbol{\theta}_r} \left\{ \log[\pi(\boldsymbol{\theta}_r)] - \sum_{t=k}^{T} \log\left(\left[C_{x^{(k)}_{t-1*,t}} \middle| \boldsymbol{\theta}_r \right] \right) \right\} \ .$$

Here we have assumed that all other parameters in our HMM besides from $\boldsymbol{\theta}_r$ are fixed. We can then find this estimate by Algorithm 4 in Section 4.2.3.

4.3.3 Parameters in the acquisition likelihood model

The wavelets, $\mathbf{w}$, are the rows in the convolution matrix, $\mathbf{W}$, in the acquisition likelihood model in Expression (2.9), and depend on the parameters a and σ_w by Expression (2.11). We first assume that the error term in the acquisition likelihood model contains colored noise, i.e. the acquisition likelihood is given by Expression (2.10). With the two unknown noise parameters, σ_{e_1} and σ_{e_2}, there are four unknown parameters in total, see also Figure 4.1,

$$\boldsymbol{\theta}_d = (\sigma_w, a, \sigma_{e_1}, \sigma_{e_2}) \ .$$

The acquisition likelihood parameters kth order MMLE is, by Expression (4.9),

$$\widehat{\boldsymbol{\theta}}_{d,MMLE}^{(k)} = \arg\max_{\boldsymbol{\theta}_d} \left\{ -\sum_{t=k}^{T} \log\left(\left[C_{x_{t-1*,t}^{(k)}}; \boldsymbol{\theta}_d \right] \right) \right\} \ . \tag{4.24}$$

These parameters can be estimated by Algorithm 4 in Section 4.2.3, when assuming that all other parameters are fixed. If we only assume white noise, i.e. $\mathbf{De} = \mathbf{e} \sim N_T(0, \sigma_e^2 \mathbf{I})$, and fix a, then $\boldsymbol{\theta}_d = (\sigma_e, \sigma_w)$ only. We can justify fixing a as the wavelet variance parameter, σ_w, constrains the significant wavelet length by Expression (2.11).

In a Bayesian setting, we could assign independent Inverse Gamma distributions to the noise variance parameters, $\sigma_{e_1}^2$ and $\sigma_{e_2}^2$, i.e. $\sigma_j^2 \sim IG(\alpha_j, \beta_j); j \in \{e_1, e_2\}$, presuming that the Gaussian error means are zero by Expression (1.7). We choose this prior distribution family as it is the conjugate prior to the Gaussian distribution when the mean is given, see Appendix A.1. The discrete wavelet function in Expression (2.11) is not actually a probability distribution, but is identical to a discrete zero-meaned Gaussian distribution function. We thus choose to assign an Inverse Gamma prior distribution to the wavelet variance parameter as well, i.e. $\sigma_w^2 \sim IG(\alpha_w, \beta_w)$. Here we assume a to be fixed. The acquisition likelihood parameter prior pdf is then

$$\pi(\boldsymbol{\theta}_d) = \prod_{j=w,e_1,e_2} \left[\frac{1}{\beta_j^{\alpha_j} \Gamma(\alpha_j)} \frac{e^{-1/(\sigma_j^2 \beta_j)}}{\sigma_j^{2(\alpha_j+1)}} \right] \ , \tag{4.25}$$

with the (6×1) hyperparameter vector $\boldsymbol{\alpha}_d = (\alpha_w, \alpha_{e_1}, \alpha_{e_2}, \beta_w, \beta_{e_1}, \beta_{e_2})$. With the white noise assumption the parameter prior pdf is

$$\pi(\boldsymbol{\theta}_d) = \prod_{j=w,e_2} \left[\frac{1}{\beta_j^{\alpha_j} \Gamma(\alpha_j)} \frac{e^{-1/(\sigma_j^2 \beta_j)}}{\sigma_j^{2(\alpha_j+1)}} \right] \ , \tag{4.26}$$

with the (4×1) hyperparameter vector $\boldsymbol{\alpha}_d = (\alpha_w, \alpha_e, \beta_w, \beta_e)$. Assuming that all other parameters beside from $\boldsymbol{\theta}_d$ is fixed, the parameter MAP estimate with prior pdf from either Expressions (4.25) or (4.26) is

$$\widehat{\boldsymbol{\theta}}_{d,MAP}^{(k)} = \arg\max_{\boldsymbol{\theta}_d} \left\{ \log[\pi(\boldsymbol{\theta}_d)] - \sum_{t=k}^{T} \log\left(\left[C_{x_{t-1*,t}^{(k)}} \middle| \boldsymbol{\theta}_d \right] \right) \right\} \ , \tag{4.27}$$

by Expression (4.11). We can find this estimate by Algorithm 4 in Section 4.2.3.

Chapter 5

Test study: Parameter estimation in the acquisition likelihood

In the test study in this chapter, we will use a model with three possible classes, with the state space $\Omega_x = \{\text{white,grey,black}\}$. A representation of the model is displayed in Figure 5.1 along a 1D class profile of length $T = 200$, i.e. $\mathbf{x} : \{x_t; t = 1, \ldots, 200\}$. This particular representation will be the reference class profile, $\mathbf{x}^R$, in our test study. In Figure 5.1, we also have a response profile, $\mathbf{r}^R$, which will be our reference response profile. Observation profiles will be simulated based on this response profile, according to Expression (2.10), for different acquisition likelihood parameter sets containing the convolution variance, σ_w^2, and the noise variance, σ_e^2. The main objective will be to estimate these two acquisition likelihood parameters based on the respective observation profiles. The next objective will then be to predict class profiles, $\hat{\mathbf{x}}$, using the observation profiles and the estimated parameters, and compare the reproductions to the reference profile, $\mathbf{x}^R$.

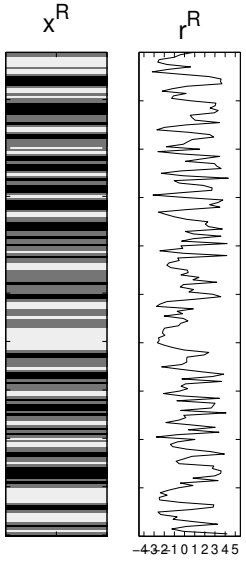


Figure 5.1: *Reference profile,* $\mathbf{x}^R$*, and an associated response profile,* $\mathbf{r}^R$*.*

5.1 Model specifications

In our model, there are $L = 3$ possible states. The reference class profile in Figure 5.1 is simulated from the prior distribution defined in Section 2.1, with the (3×3) transition probability matrix

$$\mathbf{P}_x = \begin{pmatrix} 0.50 & 0.50 & 0 \\ 0.33 & 0.34 & 0.33 \\ 0 & 0.50 & 0.50 \end{pmatrix} \quad . \tag{5.1}$$

The initial marginal prior distribution is defined as the stationary pdf of $\mathbf{P}_x$ in Expression (5.1), i.e. $p(x_1) = (0.2845, 0.4310, 0.2845)$. We note that the classes white and black are defined symmetrically relative to the background class, grey, and that they cannot be neighbors by Expression (5.1). Except from this, there is no informative prior knowledge in $\mathbf{P}_x$, i.e. the transition probabilities are uniform.

The response levels are associated to the states by independent Gaussian univariate response likelihood pdfs as defined in Section 2.2.1, i.e.

$$p(\mathbf{r}|\mathbf{x}) = \prod_{t=1}^{T} p(r_t|x_t) = \prod_{t=1}^{T} \phi_1(r_t; \mu_{x_t}, \sigma_x^2) \tag{5.2}$$

by Expressions (2.7) and (2.8). There is thus no spatial dependency between the conditional response levels, $[r_t|x_t]$, for all times t. In the reference response profile in Figure 5.1, we have assumed equal variance for the three classes, with $\sigma_x = 0.7$. The reference response class means are $\mu_{x_t} \in \{-2, 0, 3\}$, corresponding to white, grey and black respectively. The resulting marginal response pdfs are displayed in Figure 5.2. We see that these Gaussian response pdfs slightly overlap, most for the white and grey class, as their means are closer than between the black and grey mean. The black class should thus be easier to identify and separate from the grey background class than the white class.

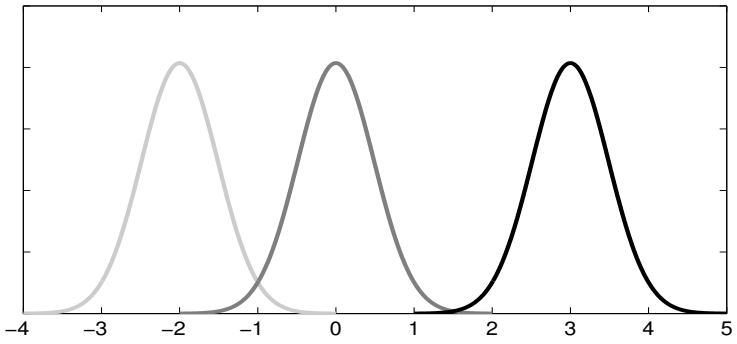


Figure 5.2: *Class response pdfs, $\phi_1(r_t; \mu_{x_t}, 0.7^2)$, with $\mu_{x_t} \in \{-2, 0, 3\}$ corresponding to the classes white, grey and black respectively.*

We assume an acquisition likelihood model, as defined by Expression (2.10), with white noise only, i.e. $p(\mathbf{d}|\mathbf{r}) = N_T(\mathbf{Wr}, \sigma_e^2\mathbf{I})$. The wavelets, $\mathbf{w}$, which are the rows in the convolution matrix, $\mathbf{W}$, are assumed to be the discrete second-order exponential function given by Expression (2.11). There are thus two acquisition likelihood parameters, σ_w and σ_e, and the parameter values used in our test study is presented in Table 5.1. The wavelets studied are presented in Figure 5.3, where the resulting wavelet length parameters used are SC:$a = 2$, MC:$a = 4$ and LC:$a = 10$ with corresponding wavelet lengths $2a + 1$. The marked function values in Figure 5.3 define the discretized wavelets.

		Convolution		
		SC Short kernel $\sigma_w = 0.5$	MC Medium kernel $\sigma_w = 1$	LC Long kernel $\sigma_w = 3$
Error	SN Small noise $\sigma_e = 0.1$	-	MC/SN	-
	MN Medium noise $\sigma_e = 0.3$	SC/MN	MC/MN Base case	LC/MN
	LN Large noise $\sigma_e = 0.8$	-	MC/LN	-

Table 5.1: *Acquisition likelihood parameters in the test study.*

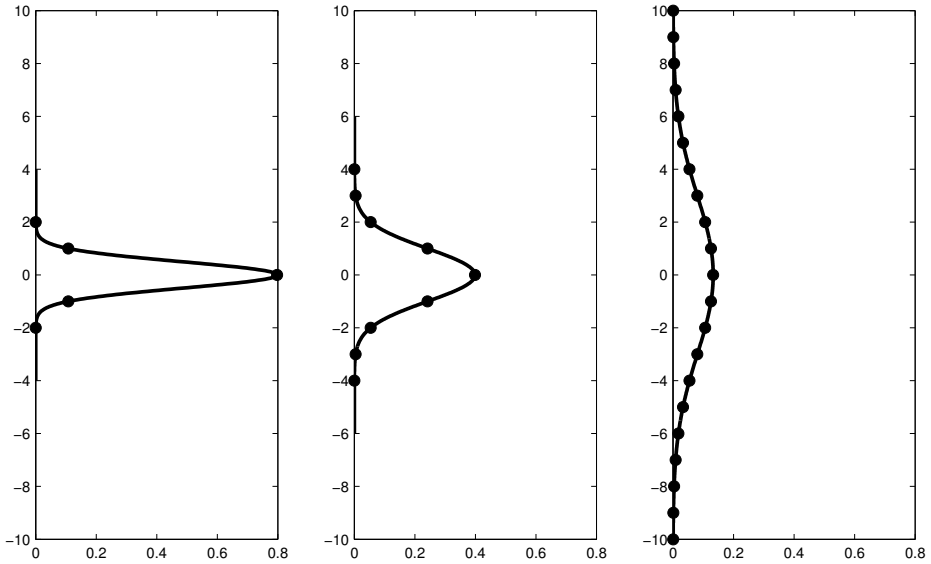


Figure 5.3: *The second-order exponential function, with marked values defining the discrete wavelet vector,* $\mathbf{w}$*, given by Expression (2.11) for (a) short kernel, (b) medium kernel and (c) long kernel.*

The resulting simulated reference observation profiles for the different parameter sets in Table 5.1 are shown in Figure 5.4. Here we have also simulated observation profiles for the four cases in Table 5.1 not present in the study, just to see how the observations vary for the different parameters. We see that the observation profiles are more similar to the response profile for smaller kernels, with smoother profiles for larger kernels. We also observe the influence of the increasing noise, especially clearly for the long kernel observations.

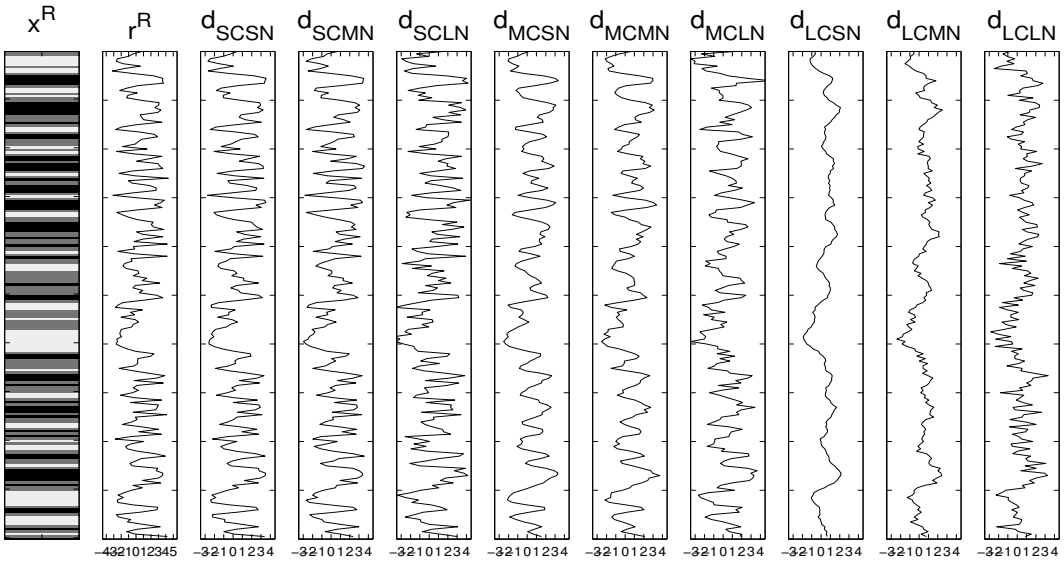


Figure 5.4: *Reference profile,* $\mathbf{x}^R$*, with the associated response profile,* $\mathbf{r}^R$*, and all observation profiles,* $\mathbf{d}$*, in the test study.*

The ratio between the wavelet variance and the noise variance is the signal-to-noise ratio, here approximated by

$$\begin{aligned} S/N &= [\mathrm{Tr}(\mathrm{Var}\{\mathbf{e_2}\})]^{-1} \times [\mathrm{Tr}(\mathrm{Var}\{\mathbf{Wr}\})] = [T\sigma_e^2]^{-1} \times [\mathrm{Tr}(\mathbf{W}\mathrm{Var}\{\mathbf{r}\}\mathbf{W}')] \\ &\approx \frac{\mathrm{Tr}(\mathbf{W}\boldsymbol{\Sigma}_\mathbf{r}\mathbf{W}')}{T\sigma_e^2} \ , \end{aligned} \tag{5.3}$$

i.e. a ratio between the observation mean and variance. Here $\mathrm{Tr}(\mathbf{A})$ is the trace of the matrix $\mathbf{A}$ and $\boldsymbol{\Sigma}_{\boldsymbol{r}}$ is the variance in the Gaussian approximation in Expression (3.17). High values of the signal-to-noise ratio indicate a good read of the response signal, while it is low when the variance, i.e. the noise, is large. The approximated S/N-ratios in our test study, by Expression (5.3), are shown in Table 5.2. We see that the S/N-ratio is larger for smaller noise, and also larger for smaller kernels with observations more similar to the response, see Figure 5.4.

	MC/LN	LC/MN	MC/MN	SC/MN	MC/SN
S/N	3.3	10.6	23.8	36.1	214.0

Table 5.2: *Approximated signal-to-noise ratios in the test study.*

From a typical statistical viewpoint, the parameters should be estimated by the following method. For every test case, we should simulate many observation profiles, then the parameter estimate would be the mean of the estimates from the respective observations. As our model is in space along a profile of length T, approximated by kth order Markov models, we obtain parameter randomization along the profile by the noise. We thus choose not to simulate many observation profiles as the variations are captured along the profile when we choose T to be large. Here we have chosen $T = 200$, and assume that the class transitions and corresponding observation changes are repeated sufficiently.

5.1.1 Test statistics for the approximated posterior model

The quality of the class profiles maximum a posteriori predictions (MAPs) are evaluated by the class profile ratio statistic

$$\Delta_{MAP} = \frac{1}{T}\sum_{t=1}^{T} I\left([\hat{x}_t|\mathbf{d}]_{MAP} = x_t^R\right) \cdot 100\% \ , \tag{5.4}$$

i.e. the ratio in percentage of the lattice nodes, $\hat{x}_t$, in the predicted class profile, $[\hat{\mathbf{x}}|\mathbf{d}]_{MAP}$, which are correctly predicted compared to the true reference class profile, $\mathbf{x}^R$. This can be evaluated for both locationwise and global MAP prediction, as described in Section 3.1. Perfect prediction would naturally yield $\Delta_{MAP} = 100\%$.

The misclassification rate of each class, $l \in \Omega_x$, is defined by the two coverage rate statistics $[a_l, b_l]$, as done in Reference [14]. When using the kth order P-P ratio deconvolution algorithm, these are defined as

$$a_l = \frac{\sum_t I(x_t^R = l)\hat{p}^{(k)}(x_t = l|\mathbf{d})}{\sum_t I(x_t^R = l)} \tag{5.5}$$

$$b_l = \frac{\sum_t \hat{p}^{(k)}(x_t = l|\mathbf{d})}{\sum_t I(x_t^R = l)} \ . \tag{5.6}$$

Here a_l represents the ability to predict the l-class in $\mathbf{x}^R$ correctly, i.e. how well the posterior model predicts the true reference class profile. The value of a_l should be larger than the prior probability for the respective class, i.e. the marginal distribution $p(x_1 = l)$, and close to 1 when using an informative likelihood model. The variable b_l represents over/under-prediction of class l, i.e. the ratio between how much the posterior model favors class l and the representation of class l in the reference profile. The value of b_l should be close to 1 for a good prediction. The difference $(b_l - a_l)$ represents the misclassification rate of class l, thus perfect prediction would yield $a_l = b_l = 1$.

5.2 Test study results

The test study discussed is based on the five test cases defined in Table 5.1. Parameter estimation of the acquisition likelihood parameters is performed as described in Section 4.2, based on the observation profiles displayed in Figure 5.4. The objective will be to predict the reference class profile, $\mathbf{x}^R$, based on these estimated parameters and the observations. All other parameters are kept fixed as the real values. The prediction will be performed by the P-P ratio deconvolution algorithm as defined in Sections 3.2 and 3.3, for orders $k = 1, 2, 3, 4, 5$, as higher order approximations turned out to be too computer demanding.

For each of the five test study cases, estimation of the two acquisition likelihood parameters, $\boldsymbol{\theta}_d = (\sigma_w, \sigma_e)$, separately and estimation of both parameters simultaneously by Algorithm 4 in Section 4.2.3 are discussed. In the estimation, we keep all other parameters fixed. Univariate estimation results are displayed for $k = 1, 2, 3, 4$, see i.e. Figure 5.5, with corresponding 95% parameter minimum Gaussian-based confidence intervals approximated by Expression (4.16). Simultaneous estimation results are displayed for $k = 1, 3$, see i.e. Figure 5.6, with corresponding elliptic 95% parameter minimum Gaussian-based confidence regions approximated by Expression (4.17). The marginal log-likelihood values, by Expression (4.8), which are displayed, are computed during the forward-backward algorithm described

in Section 3.1. Only parameter maximum marginal likelihood estimation (MMLE), as described in Section 4.2.1, is considered here, with estimate by Expression (4.24). Prediction of order k of the reference class profile, $\mathbf{x}^R$, based on the corresponding observation profile and the kth order simultaneous MMLE estimate, $\widehat{\boldsymbol{\theta}}_{d,MMLE}^{(k)}$, is displayed, see i.e. Figure 5.7. This is performed by locationwise class MAP prediction by Expression (3.12), and global class MAP prediction as defined in Algorithm 3 in Section 3.1. The same prediction is also performed based on the true acquisition parameters, $\boldsymbol{\theta}_d$, to be compared with the results from the MMLE estimate. The class MAPs are evaluated by the class profile ratio in Expression (5.4) for the different orders k, see i.e. Table 5.3, where Δ_{MAP} should be close to one for a good prediction. The class coverage rates defined by Expressions (5.5) and (5.6) are also displayed, see i.e. Figure 5.9. Both values should be close to one for a good prediction, and the resulting misclassification rate, $(b_l - a_l)$, should be small.

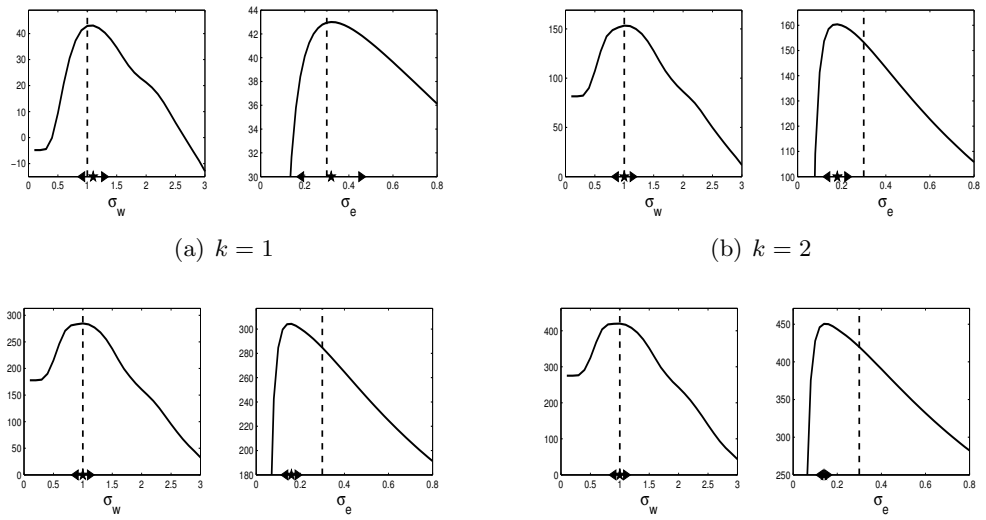
We begin by performing the parameter estimation and corresponding class profile prediction for the base case MC/MN. Then we look at the model's sensitivity in the convolution effect by performing the same estimation for the test cases SC/MN and LC/MN, with smaller and longer wavelet kernel respectively than for the base case. Next we look at the models noise sensitivity by performing the same procedure for the test cases MC/SN and MC/LN, with smaller and larger noise respectively. We want to see how the results in these test cases differ when compared to the base case. For all test cases, we should expect the results to improve for increasing orders k, as the models should be approximated better for higher orders. The kth order approximated models are simplifications of the true model, and the parameters thus have different interpretations for different orders k. We should thus not necessarily expect the parameter confidence intervals and regions to capture the true parameter value, but we should expect them to become smaller for increasing orders k.

5.2.1 Test case: MC/MN, base case

For the base case MC/MN with true parameter values $(\sigma_w, \sigma_e) = (1, 0.3)$, the univariate parameter estimation plots are displayed in Figure 5.5. The dotted lines represents the true parameter values, and the univariate parameter MMLEs correspond to the log-likelihood functions maximums, and are marked with a star. The resulting univariate 95% parameter minimum Gaussian-based confidence intervals centered in the MMLEs are marked by triangular brackets. The univariate wavelet parameter MMLE is very close to the true parameter value for all k, with slightly closer estimates for $k > 1$. The univariate noise parameter MMLE is a bit over-estimated for $k = 1$, and under-estimated for $k > 1$ when compared to the true parameter value. For both parameters, the confidence intervals become smaller for increasing k, and even overlaps the MMLE star. The parameter estimates and corresponding confidence intervals seem to stabilize for $k > 1$.

The simultaneous parameter estimation plots for the base case, MC/MN, are displayed in Figure 5.6. Here darker color indicate lower log-likelihood values, with independent color settings for the two plots. The round marker and the star indicate the true parameter value and the parameter MMLE value respectively. The boundary of the 95% parameter minimum Gaussian-based confidence region is displayed by an elliptic border. The simultaneous parameter MMLEs are here close to the true values for both orders, and for $k = 1$ the MMLE is actually equal to the true parameters. The wavelet parameter's simultaneous MMLE is equal to the true parameter value also for $k = 3$. The noise parameter's simultaneous MMLE is a bit under-estimated for $k = 3$ when compared to the true value, similar to the univariate noise parameter MMLE result in Figure 5.5. The confidence region is much smaller for the

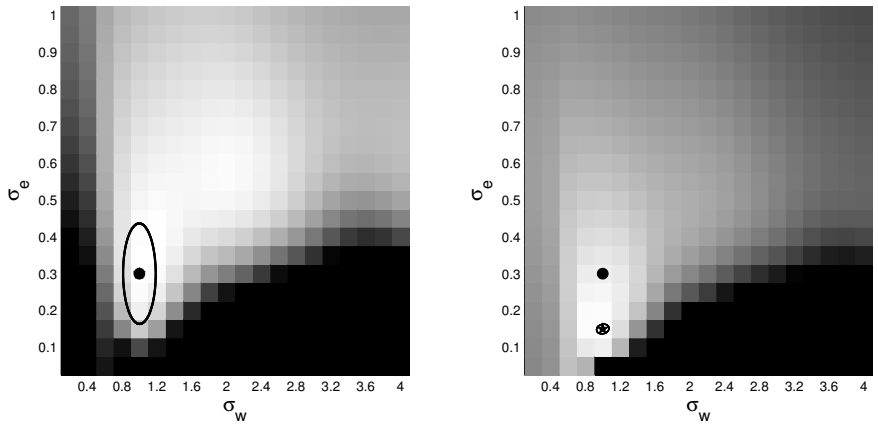
higher order, $k = 3$, where it overlaps the MMLE star.



(a) $k = 1$ (b) $k = 2$

(c) $k = 3$ (d) $k = 4$

Figure 5.5: *MC/MN: Marginal log-likelihood plots for univariate parameter estimation.*



(a) $k = 1$ (b) $k = 3$

Figure 5.6: *MC/MN: Marginal log-likelihood plots for simultaneous parameter estimation.*

The locationwise and global class MAPs based on the simultaneous parameter MMLEs, $\widehat{\boldsymbol{\theta}}^{(k)}_{d,MMLE}$, for the base case MC/MN are displayed in Figure 5.7 together with the reference class and observation profile. The class MAPs based on the true parameters, $\boldsymbol{\theta}_d$, are likewise displayed in Figure 5.8. The resulting class profile ratio values are presented in Table 5.3. We notice that for both cases, the two MAP solutions are mutually almost the same for every k, with stable predictions for $k > 1$. For $k = 1$, the MAPs seem to be quite poor when compared to the reference class profile, $\mathbf{x}^R$, as too large layers of white and black are

predicted. For $k > 1$ the MAPs seem to be better, as the transitions between the layers are better recognized. The class profile ratio values are however almost the same for the MAPs by the MMLE, and only slightly better for increasing k for the MAPs by the true parameter. This could be because too many false transitions are predicted for $k > 1$. We notice that the differences between the MAPs based on the parameter MMLE and the MAPs based on the true parameters are small.

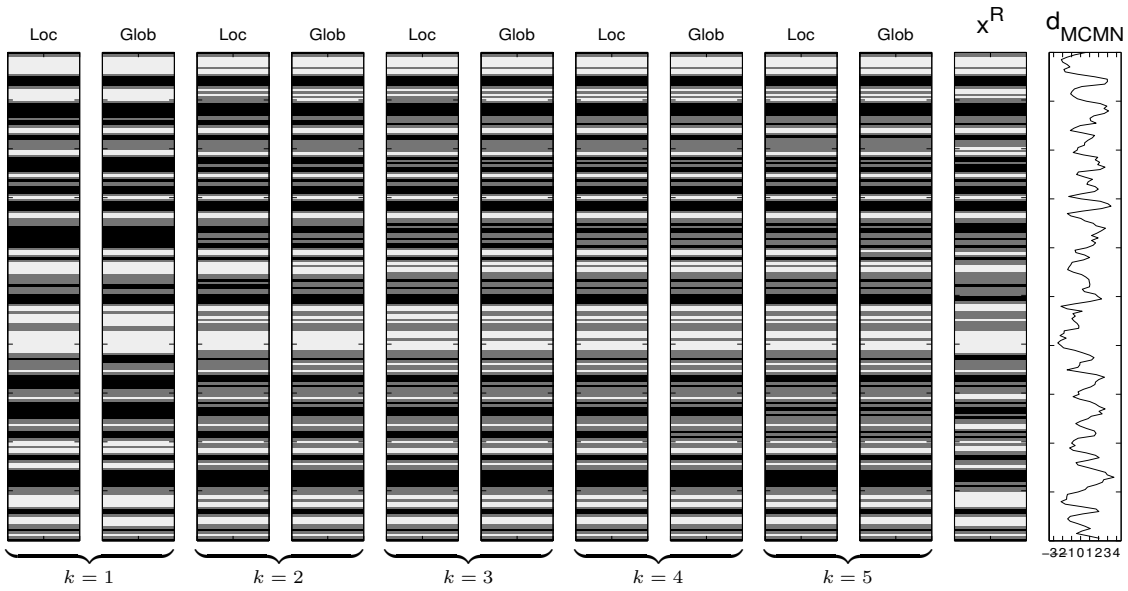


Figure 5.7: *MC/MN: Locationwise and global MAPs based on the parameter MMLE estimate,* $[\hat{\mathbf{x}}|\mathbf{d}_{MCMN}; \widehat{\boldsymbol{\theta}}^{(k)}_{d,MMLE}]_{MAP}$.

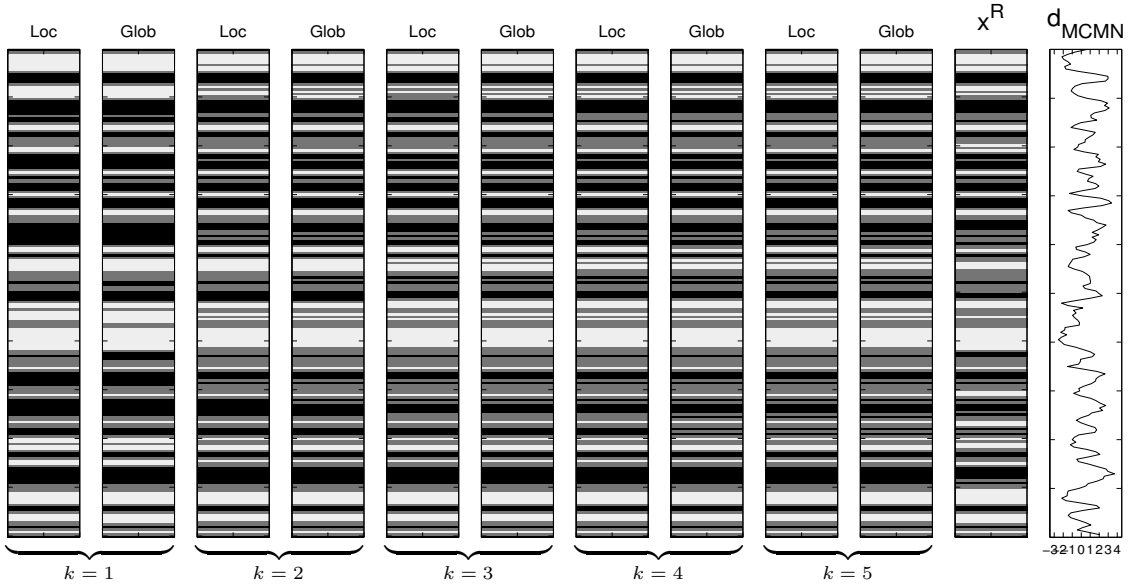


Figure 5.8: *MC/MN: Locationwise and global MAPs based on the true parameter,* $[\hat{\mathbf{x}}|\mathbf{d}_{MCMN}; \boldsymbol{\theta}_d]_{MAP}$.

$(\Delta_{locMAP}, \Delta_{globMAP})$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$\widehat{\boldsymbol{\theta}}^{(k)}_{d,MMLE}$:	$(81.0, 80.5)$	$(82.0, 80.0)$	$(80.0, 80.0)$	$(80.5, 81.0)$	$(80.0, 79.5)$
$\boldsymbol{\theta}_d$:	$(81.0, 80.5)$	$(85.0, 83.5)$	$(84.0, 83.5)$	$(85.0, 85.0)$	$(85.0, 85.0)$

Table 5.3: *MC/MN: Class profile ratios in percentage for the locationwise and global MAPs.*

The classes' coverage rate values and corresponding misclassification rates for the base case, MC/MN, are displayed in Figure 5.9 for the different orders k. In the plots, for every order k, the lower marker indicates the value of the coverage rate statistic a_l and the upper marker the value of the coverage rate statistic b_l. The misclassification rate is then indicated by the interval length $[a_l, b_l]$. The results for the posterior kth order approximations based on the kth order parameter MMLE correspond to the values in black, while the values in grey are based on the true parameters. The lower horizontal line in these plots is the marginal prior probability for the each class, given as $p(x_1) = (0.2845, 0.4310, 0.2845)$, which should be below the coverage rate values. The upper horizontal line is the balanced ratio of 1, which the coverage rates ideally should lie on. We notice that the coverage rate values, and hence the misclassification rates as well, seem to stabilize for all classes for orders $k > 1$. The posterior model's probability rates of the correct classes' depths in $\mathbf{x}^R$, indicated by the value of a_l, are stable and quite good for the white and black class, being relatively close to 1. For the grey class, the value of a_l is quite low for $k = 1$, increasing for $k = 2$ and then stabilizing. For $k = 1$, the white and black class is, by the value of b_l, over-predicted while the grey class is under-predicted. This is in accordance with the large white and black layers predicted in the MAPs in Figures 5.7 and 5.8. For $k > 1$, the white and grey class are slightly under-predicted, and the black class a bit over-predicted. The resulting misclassification rates, $(b_l - a_l)$, are quite good for all classes. For the white and black class, it is increasing for $k > 1$ and stabilizing, while it is stable for all k for the grey class. As for the class MAP results, there is no big difference between the rates based on the parameter MMLE and the rates based on the true parameters, with slightly better results for the true parameters.

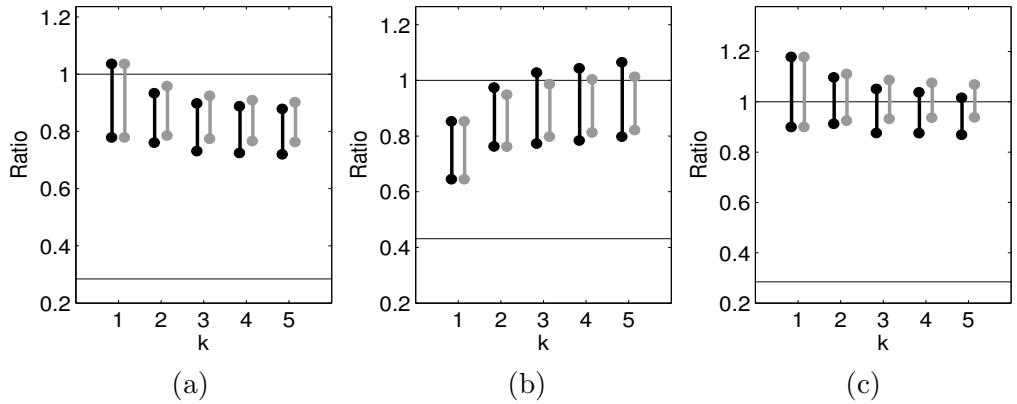
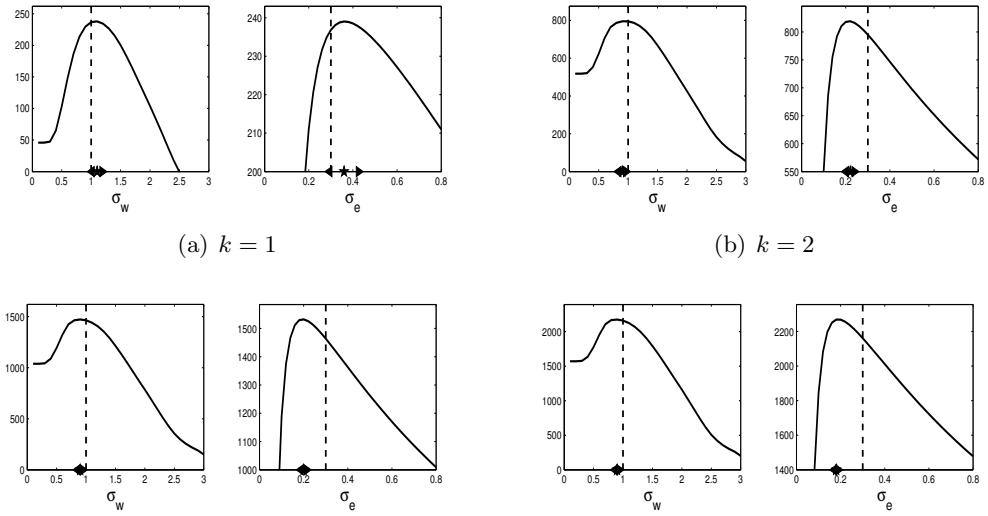


Figure 5.9: *MC/MN: Misclassification rates, $[a_l, b_l]$, with the prior ratios marked for (a) the white class, (b) the grey class and (c) the black class.*

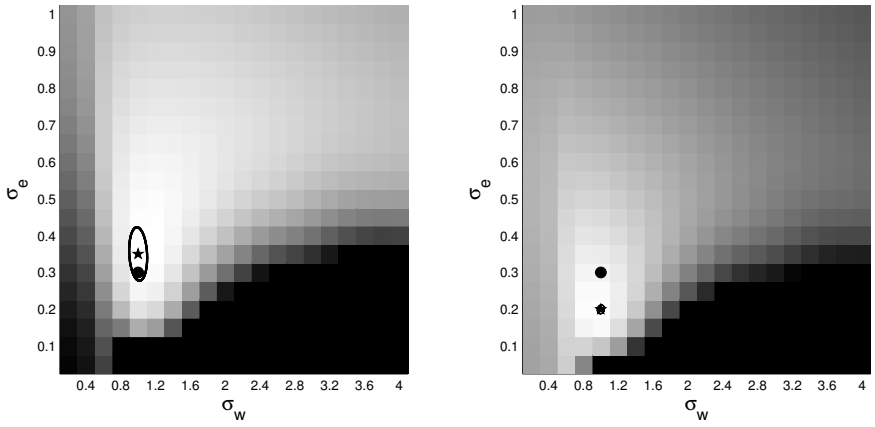
For the base case MC/MN, the class profile, response profile and observation profile, i.e. $\mathbf{x}^R$, $\mathbf{r}^R$ and $\mathbf{d}_{MCMN}$ respectively in Figure 5.4, are actually the 200 first elements of longer profiles with length $T = 1000$. These complete base case profiles are presented in Figure C.1 in Appendix C.1. With longer profiles, we should expect that more of the variations are captured. The univariate and simultaneous parameter estimation plots based on these longer profiles are displayed in Figures 5.10 and 5.11 respectively. We notice how the plots resemble the corresponding plots in Figures 5.5 and 5.6 for the profile of length $T = 200$, with almost the same MMLEs. The confidence intervals and regions are here slightly smaller, due to more data points used in the computations. As the difference is small, we have chosen to simulate profiles of length $T = 200$ only for the other test studies.



(a) $k = 1$ (b) $k = 2$

(c) $k = 3$ (d) $k = 4$

Figure 5.10: *MC/MN: Marginal log-likelihood plots for univariate parameter estimation based on the complete profiles.*

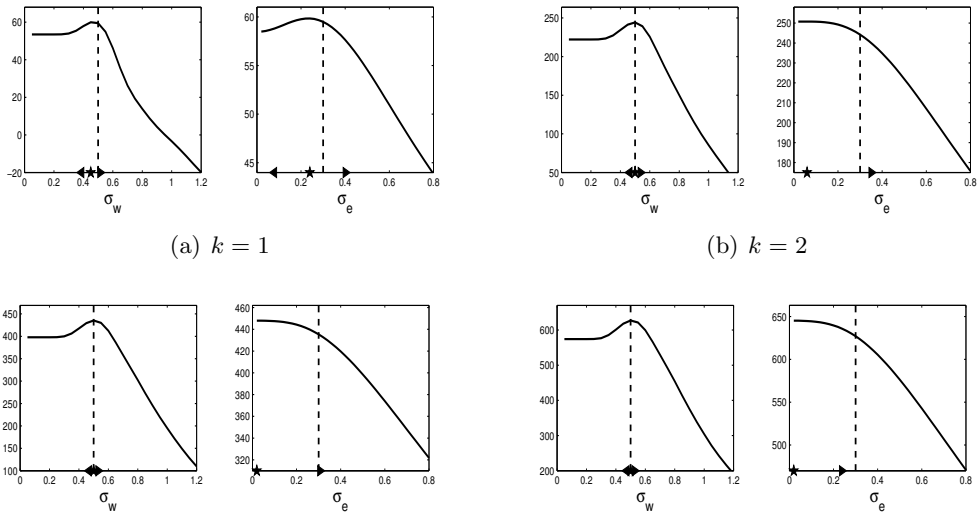


(a) $k = 1$ (b) $k = 3$

Figure 5.11: *MC/MN: Marginal log-likelihood plots for simultaneous parameter estimation based on the complete profiles.*

5.2.2 Test case: SC/MN

For the test case SC/MN with true parameter values $(\sigma_w, \sigma_e) = (0.5, 0.3)$, the univariate parameter estimation plots are displayed in Figure 5.12. The wavelet parameters univariate MMLE is very close to the true value for all orders k, with decreasing confidence intervals. The noise parameter is a bit under-estimated for $k = 1$ in the univariate case, and estimated very close to 0 for $k > 1$, i.e. higher order estimations favors no noise. The simultaneous parameter estimation plots are displayed in Figure 5.13. The simultaneous wavelet parameter MMLE is slightly under-estimated for $k = 1$ compared to the true value, and equal for $k = 3$. The simultaneous noise parameter MMLE is, for both orders, very under-estimated compared to the true value, similar to the univariate estimation result. As in the base case, the confidence region decrease for higher orders k. Due to the strong under-estimation of the noise, some of the confidence intervals and regions in Figures 5.12 and 5.13 fall outside of the possible noise definition space, i.e. it captures negative noise variance parameters. Thus, the assumption of the confidence intervals and regions being Gaussian-based does not hold in these cases.

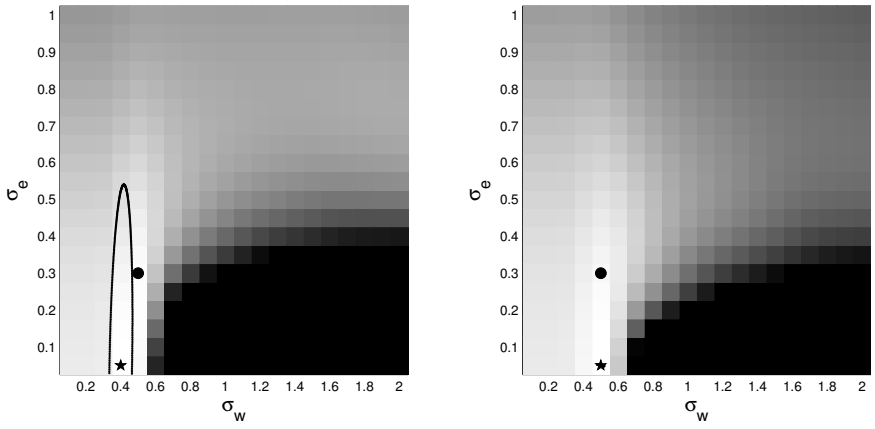


(a) $k = 1$ (b) $k = 2$

(c) $k = 3$ (d) $k = 4$

Figure 5.12: *SC/MN: Marginal log-likelihood plots for univariate parameter estimation.*

The locationwise and global class MAPs based on the simultaneous parameter MMLEs, $\widehat{\boldsymbol{\theta}}^{(k)}_{d,MMLE}$, for the test case SC/MN are displayed in Figure 5.14, and the resulting class profile ratio values are presented in Table 5.4. The MAPs are almost equal for all orders k for both MAP approaches, and they copy the reference profile, $\mathbf{x}^R$, very accurate. The resulting class profile ratio values are close to 100% for all orders k. In this test case, and in the ones left, we will not display the MAPs based on the true parameters as we did for the base case. These MAPs have however been computed, and they will be evaluated compared to the results by the MMLE. For the test case SC/MN, the predictions are almost equally reliable for both parameter sets as seen in Table 5.4.



(a) $k = 1$ (b) $k = 3$

Figure 5.13: *SC/MN: Marginal log-likelihood plots for simultaneous parameter estimation.*

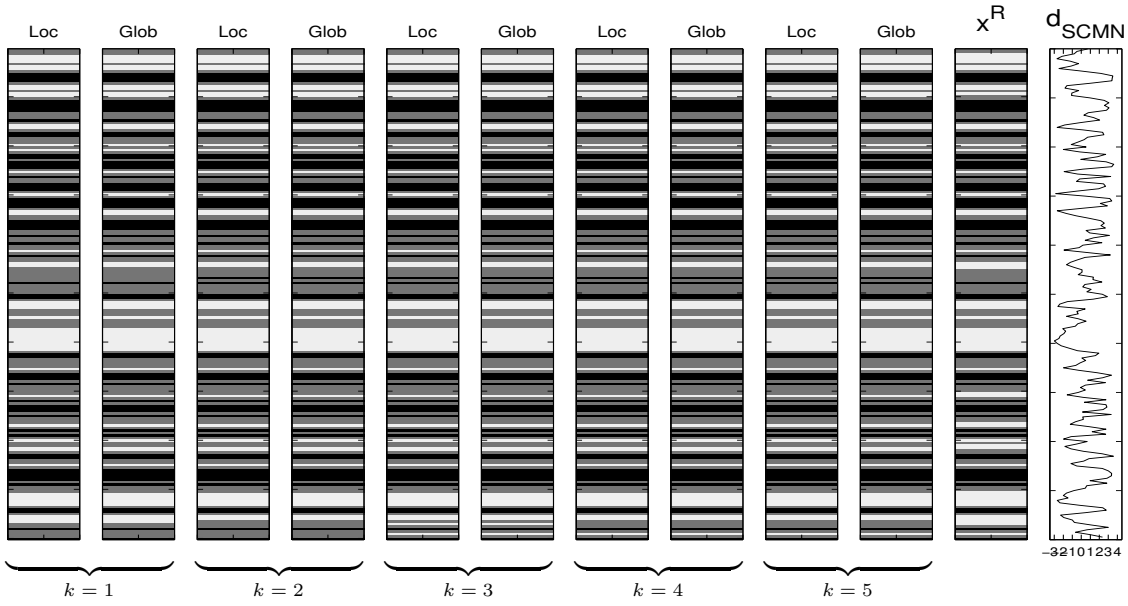


Figure 5.14: *SC/MN: Locationwise and global MAPs based on the parameter MMLE estimate,* $[\hat{\mathbf{x}}|\mathbf{d}_{SCMN}; \widehat{\boldsymbol{\theta}}^{(k)}_{d,MMLE}]_{MAP}$.

$(\Delta_{locMAP}, \Delta_{globMAP})$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$\widehat{\boldsymbol{\theta}}^{(k)}_{d,MMLE}$:	$(95.0, 95.0)$	$(94.0, 94.0)$	$(95.0, 95.0)$	$(94.5, 94.5)$	$(94.5, 94.5)$
$\boldsymbol{\theta}_d$:	$(95.5, 95.5)$	$(95.0, 95.0)$	$(94.0, 94.5)$	$(94.0, 94.5)$	$(94.0, 94.5)$

Table 5.4: *SC/MN: Class profile ratios in percentage for the locationwise and global MAPs.*

The classes' coverage rate values and corresponding misclassification rates for the test case SC/MN are displayed in Figure 5.15. The coverage rate values, and resulting misclassification rates, are very stable, with small improvement for the white and black class for increasing k. The coverage rate values are especially reliable for the black and grey class, while the white class is a bit under-predicted. We can see this in Figure 5.14 where some white layers in $\mathbf{x}^R$ are reduced to thinner layers in the class MAPs, for an example the second last white layer.

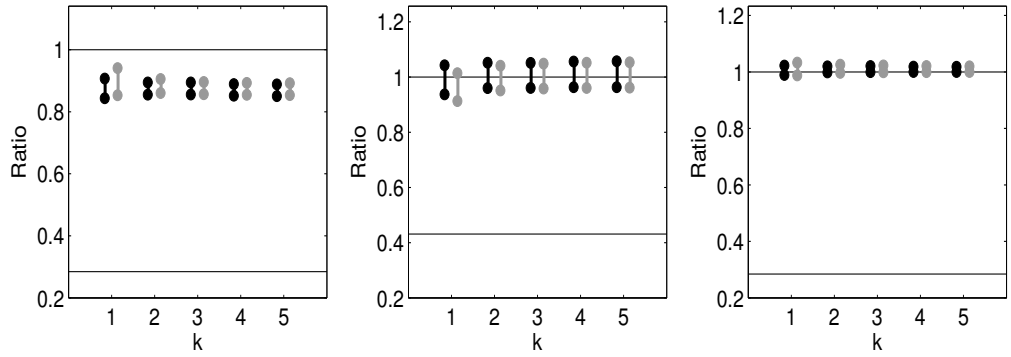


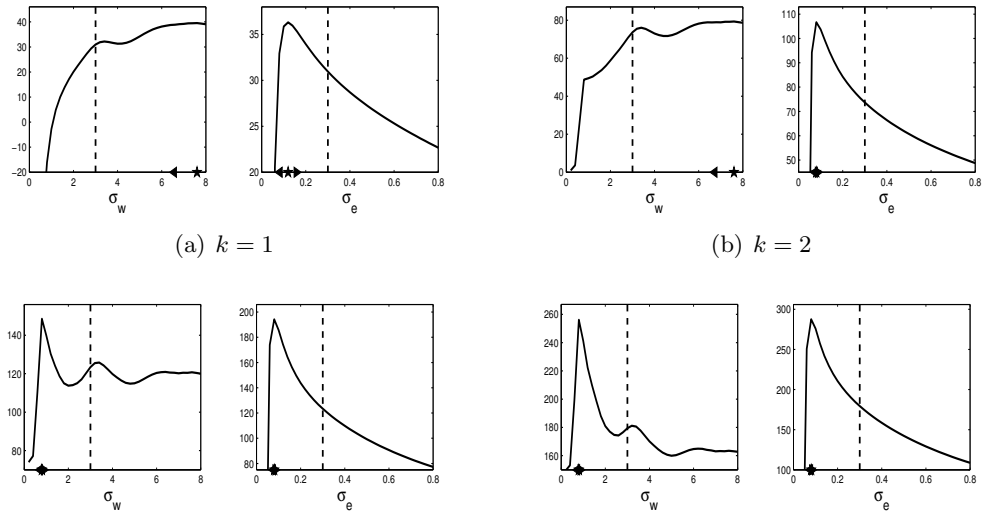
Figure 5.15: *SC/MN: Misclassification rates, $[a_l, b_l]$, with the prior ratios marked for (a) the white class, (b) the grey class and (c) the black class.*

5.2.3 Test case: LC/MN

For the test case LC/MN with true parameter values $(\sigma_w, \sigma_e) = (3, 0.3)$, the univariate parameter estimation plots are displayed in Figure 5.16. The univariate MMLE of the wavelet parameter is, when compared to the true value, strongly over-estimated for $k = 1, 2$, while it is quite under-estimated for $k = 3, 4$. We notice a local log-likelihood maximum close to the true parameter value for all k. The univariate MMLE of the noise parameter is quite under-estimated and stable for all k when compared to the true value. The simultaneous parameter estimation plots are displayed in Figure 5.17. As in the univariate case, when compared to the true value, the wavelet parameter is here strongly over-estimated for $k = 1$ and quite under-estimated for $k = 3$. The noise parameter's simultaneous MMLE is close to the true value for $k = 1$ and under-estimated for $k = 3$. In both simultaneous estimation plots and in the wavelet parameter univariate plots the approximate log-likelihood function is multimodal, i.e. there are multiple maximas. Thus, the assumption of the marginal likelihood belonging to the exponential family of distributions does not apply to this test case, and the confidence intervals and regions may not be representative.

The locationwise and global class MAPs based on the simultaneous parameter MMLEs, $\widehat{\boldsymbol{\theta}}^{(k)}_{d,MMLE}$, for the test case LC/MN are displayed in Figure 5.18, and the resulting class profile ratios are presented in Table 5.5. The MAPs are poor for all k with almost equal predictions for $k > 2$. For $k = 1$, very large black and white layers are predicted in accordance to the strongly over-predicted wavelet parameter. The MAPs are actually best for $k = 2$, while too many and too thin black and white layers are predicted for $k > 2$ with an under-predicted wavelet parameter compared to the true parameter. The class profile ratio values are very

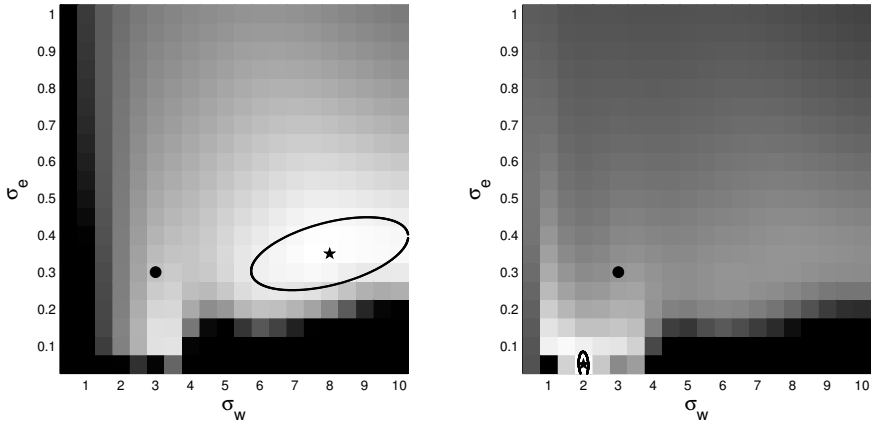
poor for all orders k. We notice that in this test case, the class MAPs based on the true parameters resemble the reference class profile significantly more reliable then the MMLE class MAPs. They are however also quite poor.



(a) $k = 1$ (b) $k = 2$

(c) $k = 3$ (d) $k = 4$

Figure 5.16: *LC/MN: Marginal log-likelihood plots for univariate parameter estimation.*



(a) $k = 1$ (b) $k = 3$

Figure 5.17: *LC/MN: Marginal log-likelihood plots for simultaneous parameter estimation.*

The classes' coverage rate values and corresponding misclassification rates for the test case LC/MN are displayed in Figure 5.19. The coverage rate values, and resulting misclassification rates, are poor for all classes for all orders k, stabilizing for $k > 2$. The black and white class is over-predicted for $k = 1, 2$ corresponding to under-prediction of the grey class, in accordance to the MAPs in Figure 5.18.

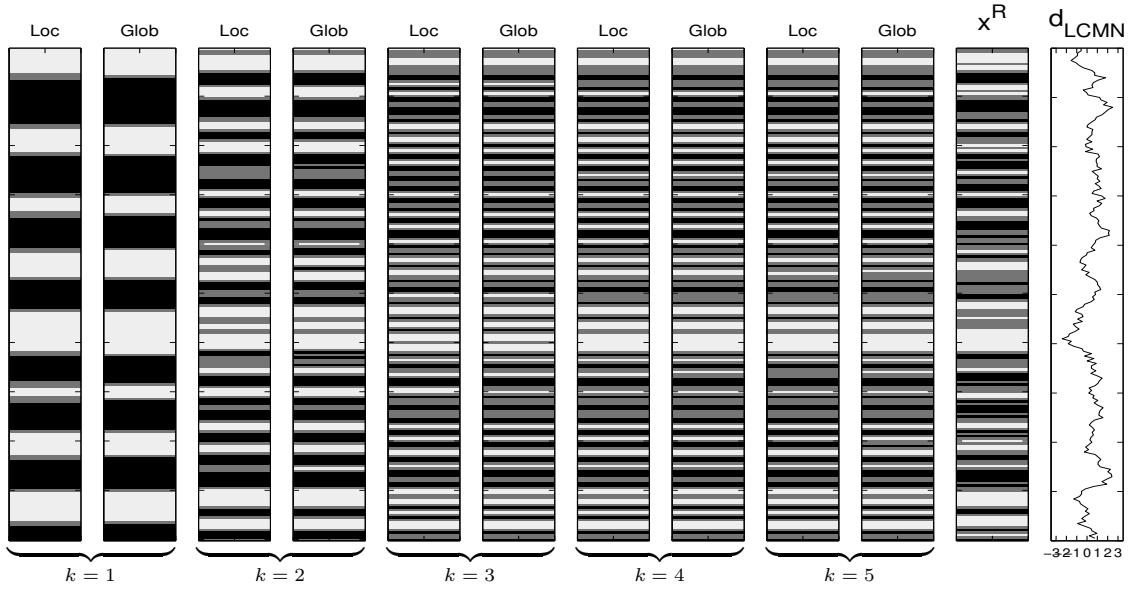


Figure 5.18: *LC/MN: Locationwise and global MAPs based on the parameter MMLE estimate,* $[\hat{\mathbf{x}}|\mathbf{d}_{LCMN}; \widehat{\boldsymbol{\theta}}^{(k)}_{d,MMLE}]_{MAP}$.

$(\Delta_{locMAP}, \Delta_{globMAP})$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$\widehat{\boldsymbol{\theta}}^{(k)}_{d,MMLE}$:	$(42.0, 42.5)$	$(55.5, 53.5)$	$(45.5, 46.0)$	$(45.5, 45.0)$	$(46.0, 46.0)$
$\boldsymbol{\theta}_d$:	$(62.0, 59.0)$	$(63.0, 60.5)$	$(64.0, 64.5)$	$(64.5, 67.0)$	$(63.0, 58.5)$

Table 5.5: *LC/MN: Class profile ratios in percentage for the locationwise and global MAPs.*

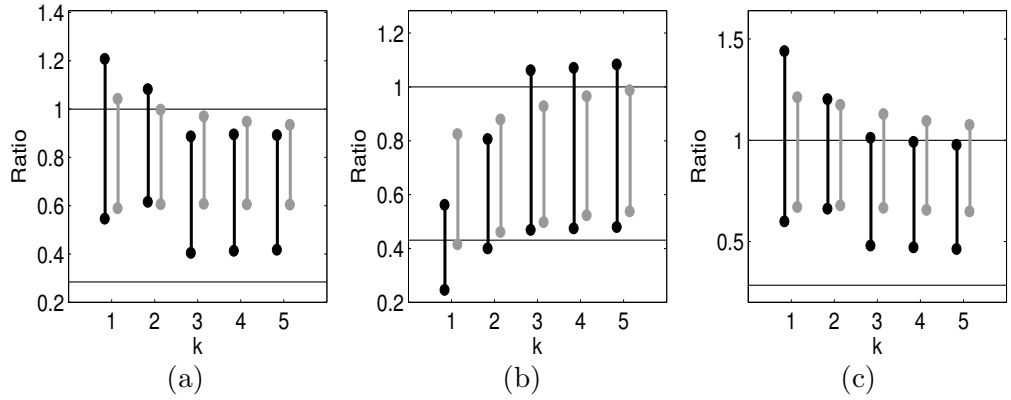
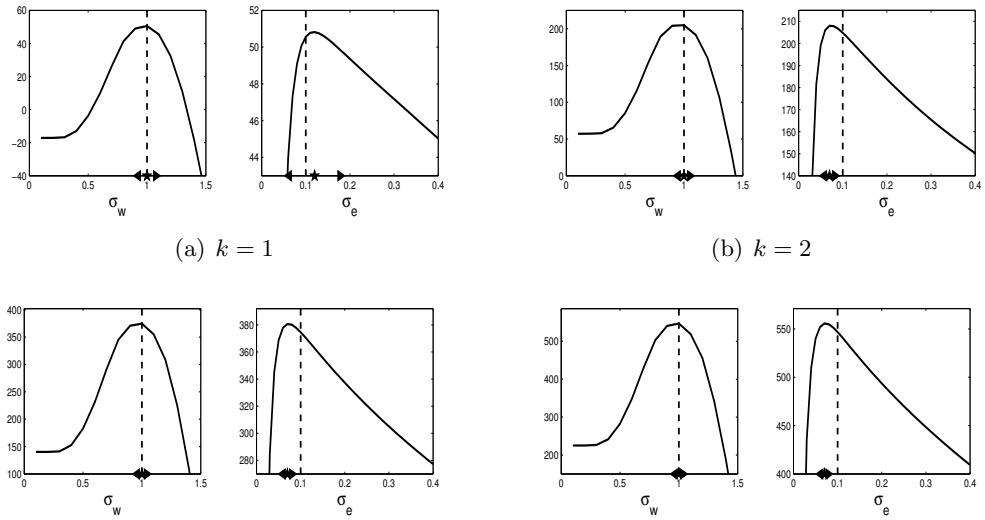


Figure 5.19: *LC/MN: Misclassification rates,* $[a_l, b_l]$*, with the prior ratios marked for (a) the white class, (b) the grey class and (c) the black class.*

5.2.4 Test case: MC/SN

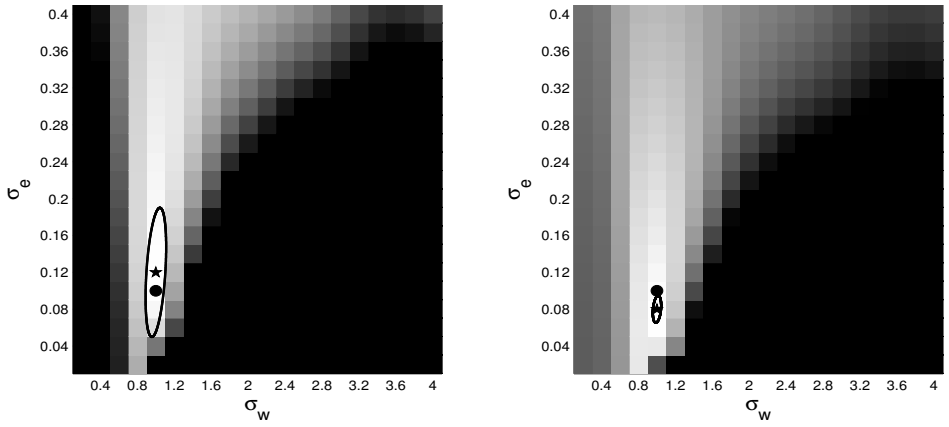
For the test case MC/SN with true parameter values $(\sigma_w, \sigma_e) = (1, 0.1)$, the univariate parameter estimation plots are displayed in Figure 5.20. The univariate MMLEs are stable for $k > 1$ and close to the true values for all orders k, especially for the wavelet parameter. The simultaneous parameter estimation plots are displayed in Figure 5.21. The resulting simultaneous MMLEs are very close to the true values for both orders k, again especially for the wavelet parameter.



(a) $k = 1$ (b) $k = 2$

(c) $k = 3$ (d) $k = 4$

Figure 5.20: *MC/SN: Marginal log-likelihood plots for univariate parameter estimation.*



(a) $k = 1$ (b) $k = 3$

Figure 5.21: *MC/SN: Marginal log-likelihood plots for simultaneous parameter estimation.*

The locationwise and global class MAPs based on the simultaneous parameter MMLEs, $\widehat{\boldsymbol{\theta}}^{(k)}_{d,MMLE}$, for the test case MC/SN are displayed in Figure 5.22, and the resulting class profile ratios are presented in Table 5.6. Both of the MAPs resemble the reference class profile very reliably for all orders k, with no significant improvement for higher orders.

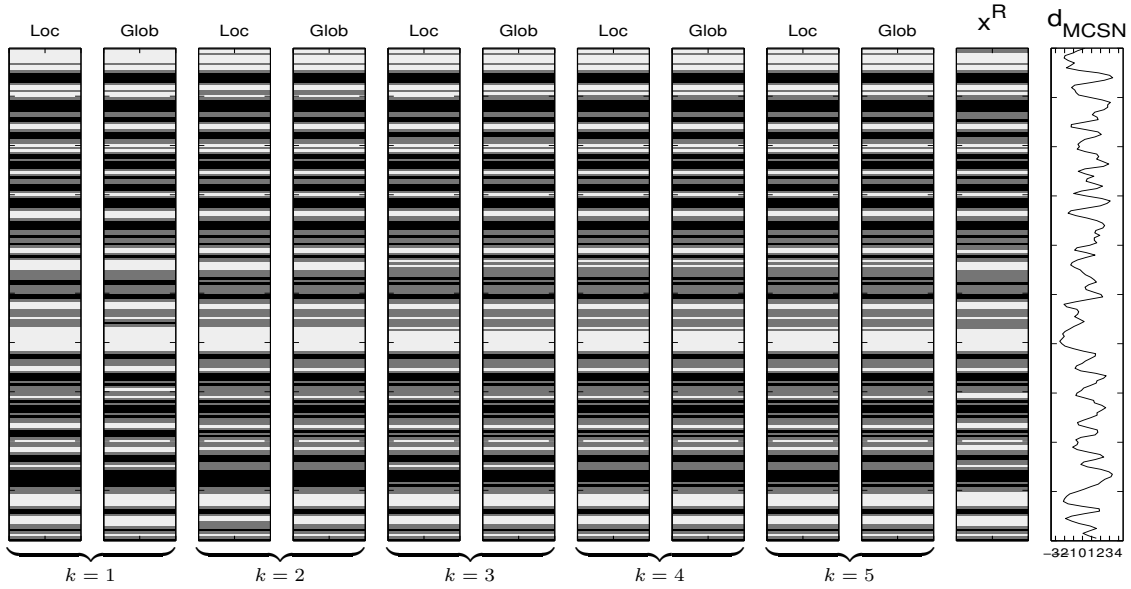


Figure 5.22: *MC/SN: Locationwise and global MAPs based on the parameter MMLE estimate,* $[\hat{\mathbf{x}}|\mathbf{d}_{MCSN}; \widehat{\boldsymbol{\theta}}^{(k)}_{d,MMLE}]_{MAP}$.

$(\Delta_{locMAP}, \Delta_{globMAP})$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$\widehat{\boldsymbol{\theta}}^{(k)}_{d,MMLE}$:	$(91.0, 90.5)$	$(90.5, 91.0)$	$(91.0, 91.5)$	$(90.5, 90.5)$	$(90.5, 90.5)$
$\boldsymbol{\theta}_d$:	$(91.0, 90.5)$	$(90.0, 91.0)$	$(91.0, 90.5)$	$(91.5, 91.5)$	$(91.5, 91.0)$

Table 5.6: *MC/SN: Class profile ratios in percentage for the locationwise and global MAPs.*

The classes' coverage rate values and corresponding misclassification rates for the test case MC/SN are displayed in Figure 5.23, and are in general quite reliable. The coverage rate values seem to stabilize for $k > 1$ for all classes, and are especially satisfying for the black class. The white class is a bit under-predicted for $k > 1$, corresponding to the small over-prediction of the other classes. As for the class MAP comparisons, the rates from the approximated model based on the parameter MMLEs are as reliable as the rates from the approximated model based on the true parameters.

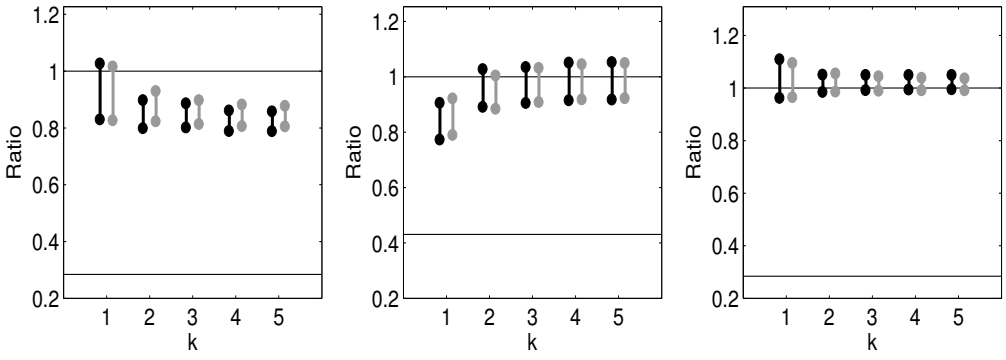
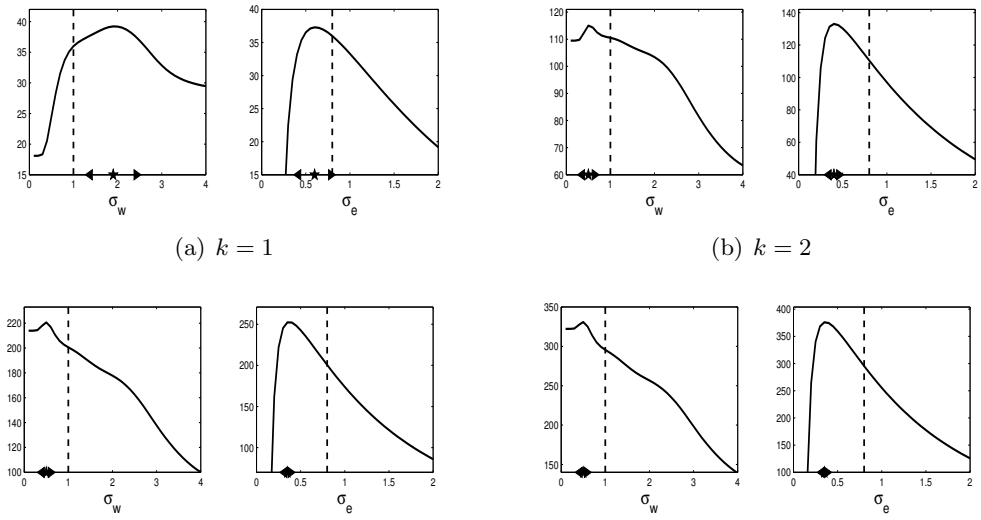


Figure 5.23: *MC/SN: Misclassification rates,* $[a_l, b_l]$*, with the prior ratios marked for (a) the white class, (b) the grey class and (c) the black class.*

5.2.5 Test case: MC/LN

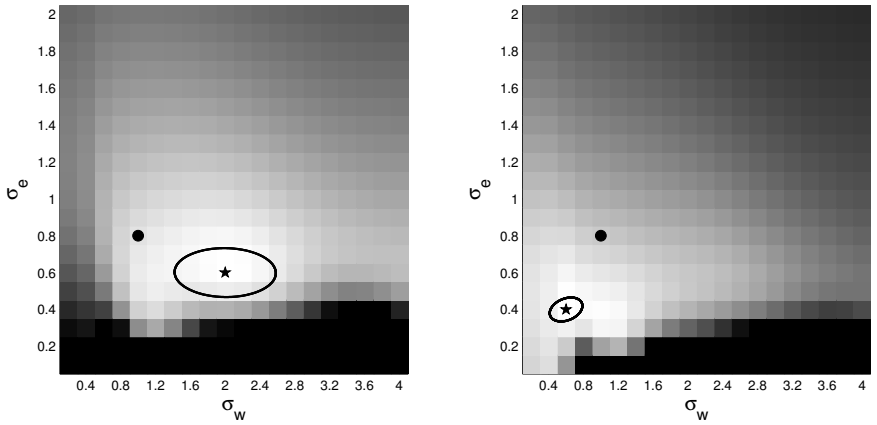
For the test case MC/LN with true parameter values $(\sigma_w, \sigma_e) = (1, 0.8)$, the univariate parameter estimation plots are displayed in Figure 5.24. The univariate MMLE of the wavelet parameter is, when compared to the true value, quite over-estimated for $k = 1$, and under-estimated, but stable, for $k > 1$. The univariate MMLE of the noise parameter is under-estimated for all k. The simultaneous parameter estimation plots are displayed in Figure 5.25. The wavelet parameter is here, as in the univariate case, strongly over-estimated for $k = 1$, and a bit under-estimated for $k = 3$. As in the testcase LC/MN there is a tendency of multi modality in the simultaneous MMLE plots.



(a) $k = 1$ (b) $k = 2$

(c) $k = 3$ (d) $k = 4$

Figure 5.24: *MC/LN: Marginal log-likelihood plots for univariate parameter estimation.*



(a) $k = 1$ (b) $k = 3$

Figure 5.25: *MC/LN: Marginal log-likelihood plots for simultaneous parameter estimation.*

The locationwise and global class MAPs based on the simultaneous parameter MMLEs, $\widehat{\boldsymbol{\theta}}^{(k)}_{d,MMLE}$, for the test case MC/LN are displayed in Figure 5.26, and the resulting class profile ratios are presented in Table 5.7. For $k = 1$, with a strongly over-estimated wavelet parameter, too thick black and white layers are predicted. The MAPs for $k > 1$ recognize the class transitions more correctly, but many false thin layers are also predicted. The values of Δ_{MAP} are a bit poor for all k.

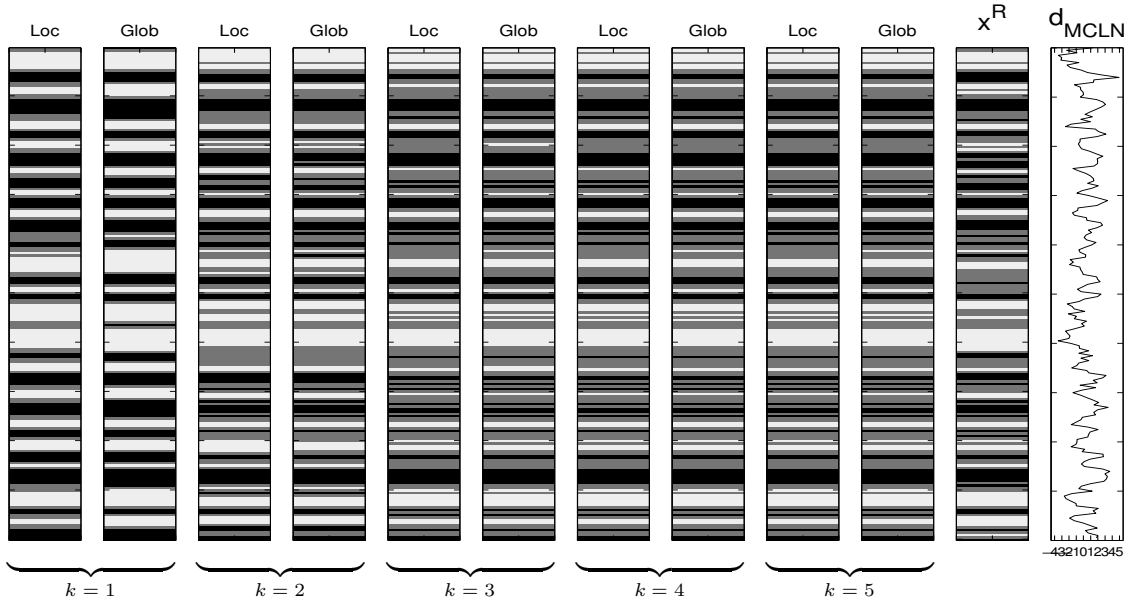


Figure 5.26: *MC/LN: Locationwise and global MAPs based on the parameter MMLE estimate,* $[\hat{\mathbf{x}}|\mathbf{d}_{MCLN}; \widehat{\boldsymbol{\theta}}^{(k)}_{d,MMLE}]_{MAP}$.

$(\Delta_{locMAP}, \Delta_{globMAP})$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
$\widehat{\boldsymbol{\theta}}^{(k)}_{d,MMLE}$:	(71.0, 69.5)	(76.0, 75.5)	(71.5, 72.0)	(71.5, 72.0)	(71.5, 72.0)
$\boldsymbol{\theta}_d$:	(75.5, 72.0)	(75.0, 73.0)	(76.0, 74.5)	(76.0, 75.0)	(76.0, 75.5)

Table 5.7: *MC/LN: Class profile ratios in percentage for the locationwise and global MAPs.*

The classes' coverage rate values and corresponding misclassification rates for the test case MC/LN are displayed in Figure 5.27. The rates seem to stabilize for $k > 2$. The white and the black class are over-predicted for $k = 1$ in accordance to the class MAPs in Figure 5.26, and under-predicted for $k > 2$. The misclassification rates are in general quite poor, especially for the grey class. The coverage rate values from the approximated model based on the true parameters are more reliable then the rates from the approximated model based on the parameter MMLEs, but the misclassifcation rates are almost equal.

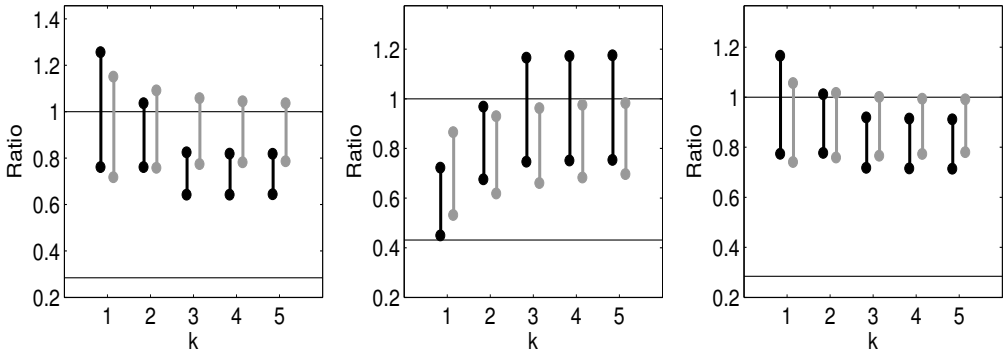


Figure 5.27: *MC/LN: Misclassification rates, $[a_l, b_l]$, with the prior ratios marked for (a) the white class, (b) the grey class and (c) the black class.*

5.3 Discussion

The two parameters in the acquisition likelihood have been estimated by maximum marginal likelihood estimation (MMLE), by Expression (4.24), based on the five test cases in Table 5.1. The shape of the log-likelihood function generally stabilizes for $k = 2$. For both univariate and simultaneous MMLE, the wavelet parameter estimate, $\hat{\sigma}_{w,MMLE}$, is generally close to the true parameter value. The noise parameter estimate, $\hat{\sigma}_{e,MMLE}$, is generally under-estimated when compared to the true value. One could believe that this under-estimation occurs as we have chosen the profile length T too short. The noise parameter could thus be overfitted due to few data points, as our kth order approximated model is a simplification of the exact model. The computed MMLEs may also be biased, but estimation by MMLE is consistent and it should thus approach the true value when T grows large. However, the noise parameter is also under-estimated when based on the complete profiles of length $T = 1000$ in the base case, MC/MN. These longer profiles are thus perhaps too short as well. Another explanation to the under-estimation could be because we use every data point d_t, in the observation profile, k times when approximating the posterior model by the kth order P-P ratio deconvolution algorithm.

For all test cases, the 95% confidence intervals and regions become smaller for increasing orders k when the log-likelihood values are stable. As we expect more exact approximations for increasing orders k, we should also expect smaller confidence intervals and regions. We notice that the confidence intervals and regions often do not cover the true parameter values, which might seem like unreliable results. The kth order MMLEs, however, have a different interpretation than the true parameters, as we use kth order approximated models. When approximating the posterior distribution by a kth order model by the P-P deconvolution algorithm, we use the kth order marginal acquisition likelihood, see Expression (3.22). We thus make different use of the wavelet matrix $\mathbf{W}$ for the different orders k, i.e. different use of the wavelet parameter. In the test case SC/MN, some of the confidence intervals and regions captures negative parameter values, i.e. it falls outside the parameters definition space. Here, the assumption of the confidence intervals/regions being Gaussian-based clearly fails. The intervals/regions are still informative in the sense that they become smaller for increasing orders k as expected.

The approximated posterior models are generally more reliable for early increasing orders k, and stabilizing for $k = 2, 3$. There seems to be a strong correspondence between the S/N-ratios and the results from the posterior model approximations. In Figure 5.28, we have plotted the logarithm of the test study S/N-ratios in Table 5.2, i.e. $\log(S/N)$, against the logit of the class MAP ratio, $\text{logit}(\Delta_{globMAP}/100)$, for $k = 3$. Here $\text{logit}(p) = \log(\frac{p}{1-p})$, where $0 \leq p \leq 1$. The results for the posterior approximations based on the parameter MMLE correspond to the values in black, while the values in grey are based on the true parameters. The plot indicates that the approximated posterior models are more reliable for increasing S/N-ratios, but also more reliable for shorter wavelet kernels. In the test case LC/MN with long wavelet kernel, too many data points are thus utilized in the wavelet. For the test cases with medium kernel (MC) there seem to be a linear trend, with respect to the log-logit transform, when tuning the noise parameter, as indicated by the dotted lines. The class MAPs based on the true parameters are in general only slightly more reliable, except for the test case LC/MN.

As discussed, the results are in general more reliable for early increasing orders k, with no great improvements after $k = 2$. Thus, $k = 3$ should be sufficient when running the estimation algorithms for the two current acquisition parameters, σ_w and σ_e, being relatively less computer demanding. The posterior model approximations for all orders k could then be computed based on this 3rd order simultaneous estimate. We should here expect more reliable results for models with larger S/N-ratio.

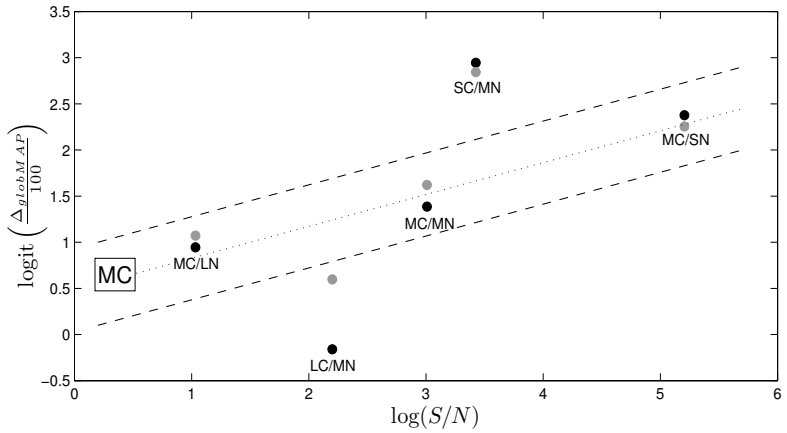


Figure 5.28: *The logarithm of the test cases' S/N-ratios,* $\log(S/N)$ *plotted against the logit of the respective global class MAP profile ratios, logit*$(\Delta_{globMAP}/100)$*, for* $k = 3$.

Chapter 6

Test study: Parameter estimation in the acquisition likelihood - sensitivity to prior model

In the test study in this chapter, we evaluate the prior model's influence on the parameter estimation in the acquisition likelihood model. The same class model as in Chapter 5 will be used, with the state space $\Omega_x = \{\text{white,grey,black}\}$. The study is performed by simulating reference class profiles using different prior models, with associated reference response and observations profiles. We then perform the parameter estimation and class prediction, as we did in Chapter 5, but now using an equal, less informative prior model for all the test cases. We thus assume that the simulated class profiles originate from the same prior model. The estimation of the associated acquisition likelihood parameters, i.e. the convolution variance, σ_w^2, and the noise variance, σ_e^2, will be performed.

6.1 Observation model

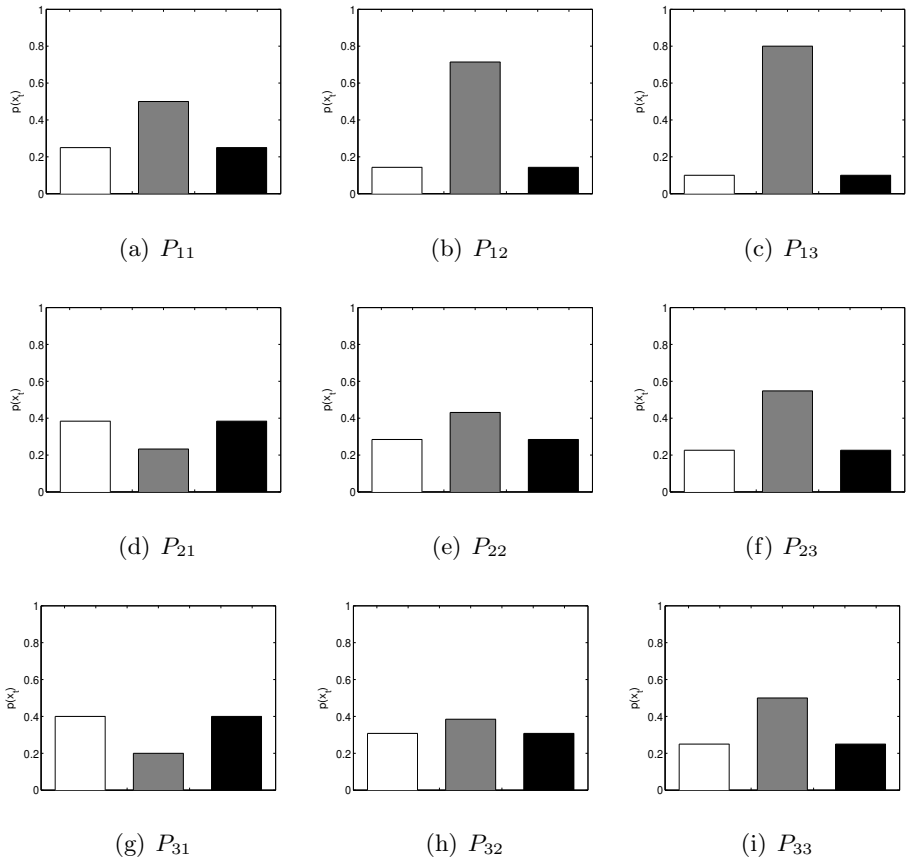
There are $L = 3$ possible states in our model, and we define the test case prior models by the (3×3) transition matrix

$$\mathbf{P}_x = \begin{pmatrix} 1-p_1 & p_1 & 0 \\ p_2 & 1-2p_2 & p_2 \\ 0 & p_1 & 1-p_1 \end{pmatrix} \quad . \tag{6.1}$$

The initial marginal prior distribution is defined as the stationary pdf of $\mathbf{P}_x$, and thus depends on the variables p_1 and p_2 as well. The white and black class are defined symmetrically relative to the grey class, and are defined such that they cannot be neighbors. The prior models used in our test study are presented in Table 6.1, and the test cases will from here on out be referred to by their prior model. We notice that the prior model in the base case, $\mathbf{P}_{22}$, has the same transition matrix used in Chapter 5, defined by Expression (5.1). The corresponding marginal prior distributions are displayed in Figure 6.1, and can be interpreted as the prior ratios between the classes. We notice that for all test cases in Figure 6.1, the ratios for the classes white and black are equal by the defined prior model in Expression (6.1). The approximated signal-to-noise ratios for the test cases, computed by Expression (5.3), are shown in Table 6.2. We notice that the S/N-ratio is especially low for the test case $\mathbf{P}_{13}$.

		Transition variable from white/black		
		$p_1 = 0.2$	$p_1 = 0.5$	$p_1 = 0.8$
Transition variable from grey	$p_2 = 0.1$	$\mathbf{P}_{11}$	$\mathbf{P}_{12}$	$\mathbf{P}_{13}$
	$p_2 = 0.33$	$\mathbf{P}_{21}$	$\mathbf{P}_{22}$ Base case	$\mathbf{P}_{23}$
	$p_2 = 0.4$	$\mathbf{P}_{31}$	$\mathbf{P}_{32}$	$\mathbf{P}_{33}$

Table 6.1: *Prior transition matrix parameters in the test study according to Expression (6.1).*



(a) P_{11} (b) P_{12} (c) P_{13}

(d) P_{21} (e) P_{22} (f) P_{23}

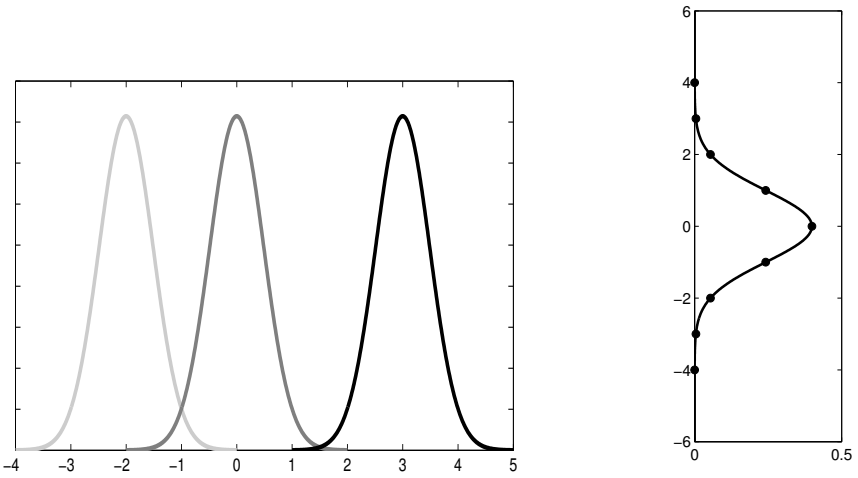
(g) P_{31} (h) P_{32} (i) P_{33}

Figure 6.1: *Marginal prior ratios in the test cases between the white, grey and black class.*

	$\mathbf{P}_{11}$	$\mathbf{P}_{12}$	$\mathbf{P}_{13}$	$\mathbf{P}_{21}$	$\mathbf{P}_{22}$	$\mathbf{P}_{23}$	$\mathbf{P}_{31}$	$\mathbf{P}_{32}$	$\mathbf{P}_{33}$
S/N	29.6	12.9	6.9	44.2	23.8	13.5	45.9	25.5	14.7

Table 6.2: *Approximated signal-to-noise ratios in the test study.*

We next define the likelihood models in the same way as we did in in Chapter 5. The response likelihood model is defined by independent Gaussian univariate response likelihood pdfs, i.e. $p(\mathbf{r}|\mathbf{x}) = \prod_{t=1}^{T} \phi_1(r_t; \mu_{x_t}, \sigma_x^2)$. The reference response class means are $\mu_{x_t} \in \{-2, 0, 3\}$, and the response class variance is equal for all classes, $\sigma_x^2 = 0.7^2$. The black class should thus be easier to identify from grey than the white class. The resulting class response pdfs are displayed in Figure 6.2(a). The acquisition likelihood model is defined by Expression (2.9), where we here assume white noise only, i.e. $p(\mathbf{d}|\mathbf{r}) = N_T(\mathbf{Wr}, \sigma_e^2\mathbf{I})$, with reference noise parameter $\sigma_e = 0.3$. The wavelets, $\mathbf{w}$, in the convolution matrix, $\mathbf{W}$, are given by Expression (2.11). The reference wavelet is displayed in Figure 6.2(b), with parameter value $\sigma_w = 1$ and resulting wavelet length $a = 4$. The acquisition likelihood model defined thus corresponds to the base case in Chapter 5, i.e. MC/MN in Table 5.1 with S/N-ratio 23.8 by Table 5.2.



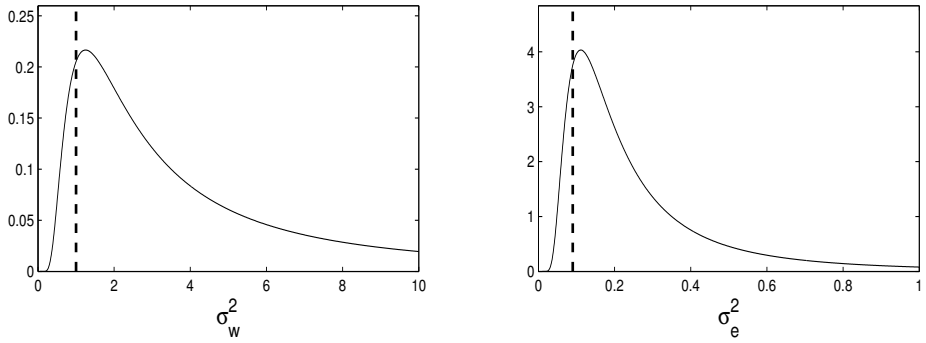
(a) (b)

Figure 6.2: *(a) Class response pdfs, $\phi_1(r_t; \mu_{x_t}, 0.7^2)$, with $\mu_{x_t} \in \{-2, 0, 3\}$ corresponding to the classes white, grey and black respectively. (b) The second-order exponential function given by Expression (2.11) with $\sigma_w = 1$, where the marked values define the discrete wavelet vector,* $\mathbf{w}$.

The acquisition likelihood parameters, σ_w and σ_e, will be estimated by maximum a posteriori prediction (MAP) as defined in Section 4.2.2, with estimates by Expression (4.27). Both parameters are assigned independent univariate prior Inverse Gamma distributions, $IG(\alpha, \beta)$, as described in Section 4.3.3, i.e. $\sigma_e^2 \sim IG(2, 4)$ and $\sigma_w^2 \sim IG(1, 0.4)$. The two prior distributions are displayed in Figure 6.3 where the dotted lines represents the reference values, i.e. $1^2 = 1$ and $0.3^2 = 0.09$ respectively. The parameter posterior logarithm values, as defined in Expression (4.10), are then

$$\hat{l}^{(k)*}(\boldsymbol{\theta}_d|\mathbf{d}) = \hat{l}^{(k)}(\mathbf{d}|\boldsymbol{\theta}_d) + \log\{\pi(\boldsymbol{\theta}_d)\} = \hat{l}^{(k)}(\mathbf{d}|\sigma_e^2, \sigma_w^2) + \log\{\pi(\sigma_w^2)\} + \log\{\pi(\sigma_e^2)\} \quad , \quad (6.2)$$

where $\pi(\sigma_w^2)$ and $\pi(\sigma_e^2)$ are the defined parameter prior distributions. The chosen prior distributions are quite centered around the true values. To justify the noise parameter prior distribution chosen, we refer to Chapter 5, where the noise parameter was often underestimated compared to the reference value. We notice that the wavelet parameter prior distribution is much wider than for the noise parameter.



(a) $\sigma_w^2 \sim IG(1, 0.4)$ (b) $\sigma_e^2 \sim IG(2, 3)$

Figure 6.3: *Acquisition likelihood parameter inverse gamma prior distributions.*

6.2 Test study results

The test study discussed is based on the nine test cases with true prior models defined by Table 6.1. In the study, we will simulate 1D class profiles, and associated response and observation profiles, using these prior models. All the class profiles simulated are of equal length, $T = 200$. The less informative base case prior model, $\mathbf{P}_{22}$, is then set as the prior model for all the test cases when estimating the parameters. The only distinct prior knowledge in $\mathbf{P}_{22}$ is that white and black can not be neighbors. The response likelihood parameters are kept fixed as the reference values. Estimation of the acquisition likelihood parameters is performed, by algorithms of order $k = 1, 3$, based on the observation profiles and the base case prior model. The objective is to reproduce the reference class profiles based on the estimated parameters. The class prediction will be performed by the P-P ratio deconvolution algorithm, as defined in Sections 3.2 and 3.3, for orders $k = 1, 2, 3, 4, 5$. In these predictions, we will for all orders k set the acquisition parameters as the parameter MAP estimate for $k = 3$, i.e. $\widehat{\boldsymbol{\theta}}_{d,MAP}^{(3)}$, as proposed by the results in Chapter 5.

For each of the nine test study cases, simultaneous MAP estimation of the acquisition likelihood parameters, $\boldsymbol{\theta}_d = (\sigma_w, \sigma_e)$, is discussed. The displayed values in the kth order parameter MAP plots are the kth order parameter posterior logarithm values defined in Expression (6.2). The parameter prior distribution values are also displayed in the parameter MAP plots, together with the corresponding approximated marginal parameter posterior distribution values, computed by Expression (4.12). Predictions of the reference profile, $\mathbf{x}^R$, based on the parameter MAP estimate for $k = 3$ are displayed by global class MAPs up to order $k = 5$, computed by Algorithm 3. The coverage rate values, a_l and b_l defined by Expressions (5.5) and (5.6) respectively, are also displayed. The misclassification rate, $(b_l - a_l)$, should be small, and both values should be close to one for a good prediction. As in Chapter 5, these rates are compared to the rates from the posterior distribution based on the true acquisition likelihood parameters, $\boldsymbol{\theta}_d$, which also is computed.

We begin by performing the parameter estimation and corresponding class profile prediction for the test case $\mathbf{P}_{22}$. Then we perform the same estimation and prediction procedures for the remaining test cases in Table 6.1 in turn, i.e. $\mathbf{P}_{11}$, $\mathbf{P}_{12}$, etc. For all test cases, we should expect the results to improve for increasing orders k, as the models should be approximated better for higher orders. We should also expect more reliable results for the test cases with larger S/N-ratio, see Table 6.2, due to the results from Chapter 5.

6.2.1 Test case: Observation model $\mathbf{P}_{22}$, base case

The reference class profile, response profile and observation profile for the test case $\mathbf{P}_{22}$ are displayed in Figure 6.4. These reference profiles are equal to the base case profiles from Chapter 5, which are the first parts of the longer profiles of length $T = 1000$ displayed in Appendix C.1. The prior model, from which the reference class profile, $\mathbf{x}^R$, is simulated from, is here defined by the transition matrix

$$\mathbf{P}_{22} = \begin{pmatrix} 0.50 & 0.50 & 0 \\ 0.33 & 0.34 & 0.33 \\ 0 & 0.50 & 0.50 \end{pmatrix} , \tag{6.3}$$

by Table 6.1. The initial marginal prior distribution is $p(x_1) = (0.2845, 0.4310, 0.2845)$, i.e. it slightly favors the grey class, and is displayed in Figure 6.1(e). The class profile is characterized by many class transitions and medium-sized layers, according to the prior model in Expression (6.3).

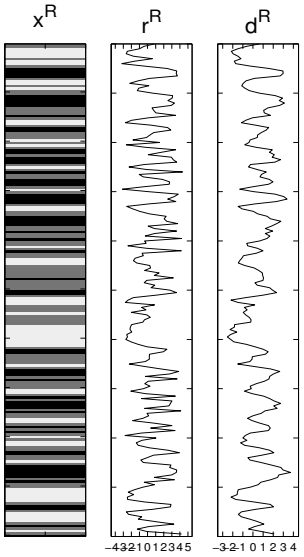
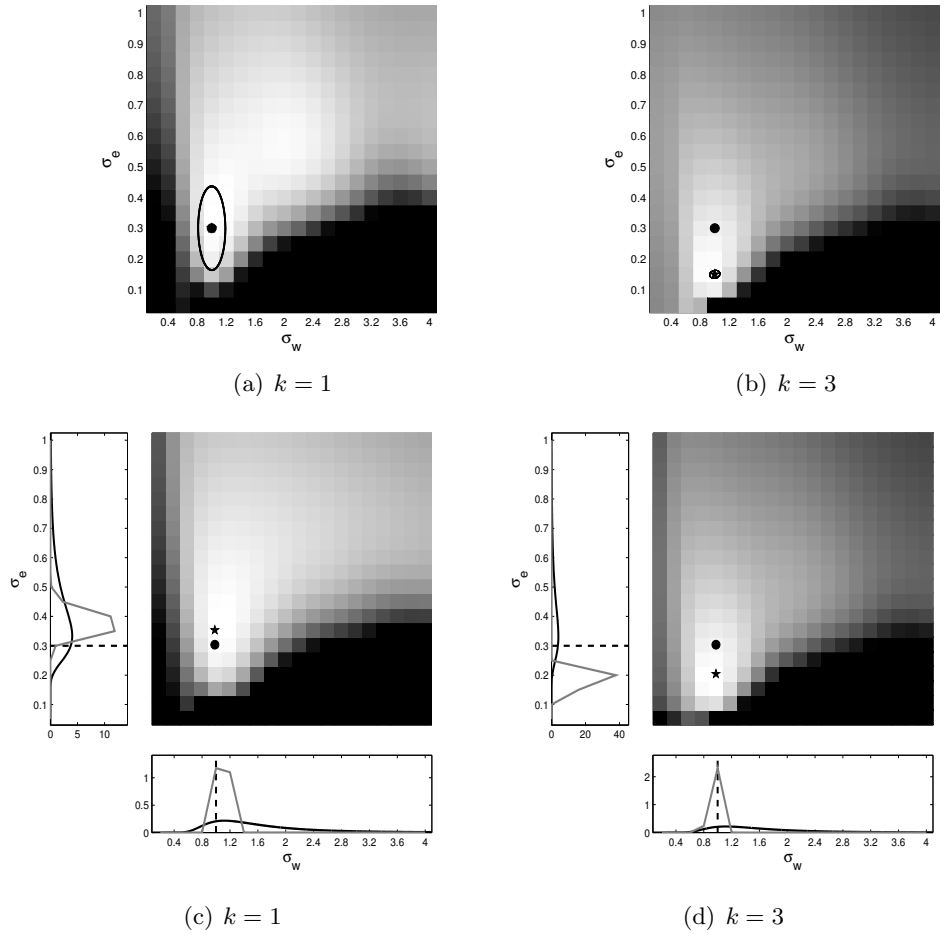


Figure 6.4: $\mathbf{P}_{22}$*: Reference class profile,* $\mathbf{x}^R$*, response profile,* $\mathbf{r}^R$*, and observation profile,* $\mathbf{d}^R$*.*

The parameter estimation plots for the base case, $\mathbf{P}_{22}$, are displayed in Figure 6.5. We have here included the marginal log-likelihood plots, which are displayed in Figures 6.5(a) and 6.5(b). Darker color in these two plots indicate lower log-likelihood values, with independent color settings for the two plots. The round marker and the star indicate the true parameter value, $\boldsymbol{\theta}_d$, and the parameter MMLE value, $\widehat{\boldsymbol{\theta}}_{d,MMLE}^{(k)}$, respectively. These two plots are equal to the simultaneous MMLE plots for the base case MC/MN in Chapter 5, see Figure 5.6. The resulting parameter MAP estimation plots, with computed values by Expression (6.2), are displayed in Figures 6.5(c) and 6.5(d). Here darker color indicate lower function values. The round marker and the star now indicate the true parameter value, $\boldsymbol{\theta}_d$, and and the parameter MAP value, $\widehat{\boldsymbol{\theta}}_{d,MAP}^{(k)}$, respectively. We notice how the marginal log-likelihood plots resemble the MAP plots, as the computed marginal log-likelihood values are large compared to the logarithm of the parameter prior distribution values, see Expression (6.2). A major difference however occurs for small parameter values, where the parameter posterior logarithm values

are considerably smaller due to the small parameter prior probabilities, see especially the plots for $k = 1$. In this case, the noise parameter MAP estimate is slightly larger then the MMLE estimate, while the wavelet estimates are equal. Due to the quite similar estimates, parameter estimation by MMLE would probably be sufficient, as it does not require prior knowledge on the parameters. We have however chosen to perform MAP estimation only for the other test cases, as estimation by MMLE was considered in the test study in Chapter 5.



(a) $k = 1$ (b) $k = 3$

(c) $k = 1$ (d) $k = 3$

Figure 6.5: $\mathbf{P}_{22}$*: Computed marginal log-likelihood values for the two acquisition likelihood parameters for orders (a)* $k = 1$ *and (b)* $k = 3$*, and their corresponding parameter posterior logarithm values in (c) and (d) respectively.*

The acquisition parameters marginal prior and posterior distribution values are also displayed in the MAP plots, in black and grey respectively. The dotted lines here represent the reference parameter values. In these plots, the distribution values are plotted against the unsquared parameter variables, i.e. θ_d against $\pi(\theta_d^2)$, $\theta_d \in \{\sigma_w, \sigma_e\}$. The functions plotted are thus not real probability distributions, but the marginal parameter MAP estimates are still the maximas of these posteriors. We will from here on out refer to them as the marginal parameter prior and posterior functions. In this case, the bivariate parameter MAP estimate is equal to the marginal MAP estimates for both orders. For $k = 1$, the marginal posterior function is wider distributed due to smaller parameter posterior logarithm values for decreasing orders.

The global class MAPs based on the parameter MAP estimate for $k = 3$, $\widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}$, for the base case $\mathbf{P}_{22}$ are displayed in Figure 6.6 together with the reference class and observation profile. The global class MAPs based on the reference parameters, $\boldsymbol{\theta}_d$, are likewise displayed in Figure 6.7, and are equal to the global MAPs in Figure 5.8 in the base case in Chapter 5. The resulting class profile ratio values, as defined by Expression (5.4), are presented in Table 6.3. We notice that the class MAPs reproduce the reference class profile quite well for all orders, and slightly more reliably for the true parameters except for $k = 1$.

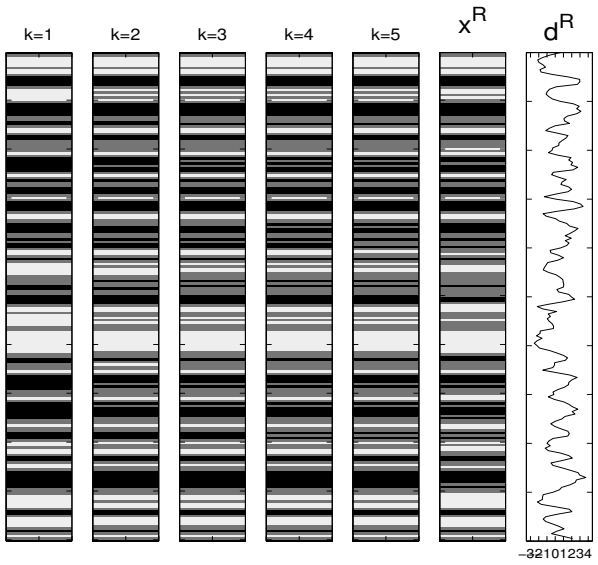


Figure 6.6: $\mathbf{P}_{22}$*: Global class MAPs based on the parameter MAP estimate for* $k = 3$*,* $[\hat{\mathbf{x}}|\mathbf{d}; \widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}]_{MAP}$*.*

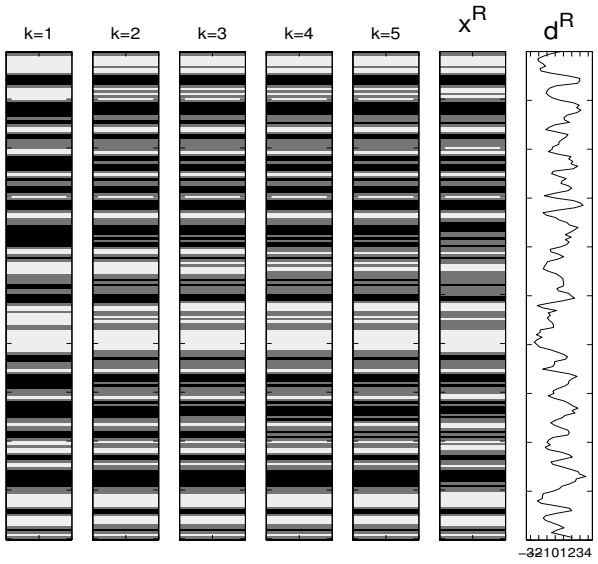
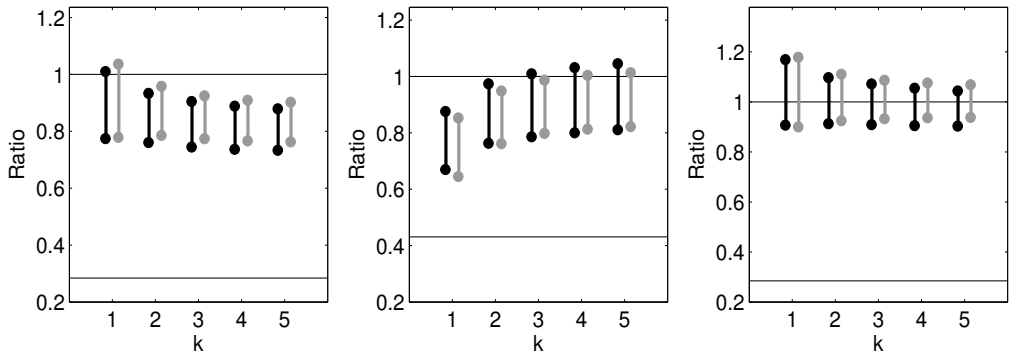


Figure 6.7: $\mathbf{P}_{22}$*: Global class MAPs based on the reference parameters,* $[\hat{\mathbf{x}}|\mathbf{d}; \boldsymbol{\theta}_d]_{MAP}$*.*

$\Delta_{globMAP}$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$\widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}$:	83.5	80.0	82.0	82.5	82.5
$\boldsymbol{\theta}_d$:	80.5	83.5	83.5	85.0	85.0

Table 6.3: $\mathbf{P}_{22}$*: Class profile ratios in percentage for the locationwise and global MAPs.*

The classes' coverage rate values and corresponding misclassification rates for the base case, $\mathbf{P}_{22}$, are displayed in Figure 6.8 for the different orders k. These plots resemble the misclassification plots in Figure 5.9 for the base case in Chapter 5. The results for the posterior kth order approximations based on the 3rd order parameter MAP estimate correspond to the values in black, while the values in grey are based on the true parameters. The rates seem to stabilize for $k > 1$ for all classes, and are almost equal between the two posterior distribution approximations. For orders $k > 1$ the misclassification rates for the white and black class are quite reliable.



(a) (b) (c)

Figure 6.8: $\mathbf{P}_{22}$*: Misclassification rates,* $[a_l, b_l]$*, with the prior ratios marked for (a) the white class, (b) the grey class and (c) the black class.*

6.2.2 Test case: Observation model $\mathbf{P}_{11}$

In Figure 6.9, we see the reference class profile simulated from $\mathbf{P}_{11}$, and the corresponding response and observation profile. The prior model is here defined by the transition matrix

$$\mathbf{P}_{11} = \begin{pmatrix} 0.80 & 0.20 & 0 \\ 0.10 & 0.80 & 0.10 \\ 0 & 0.20 & 0.80 \end{pmatrix} \tag{6.4}$$

with initial marginal prior $p(x_1) = (0.25, 0.50, 0.25)$. This prior model favors the grey class, which is reflected by the reference class profile. When entering a class, the probability of staying there is large by Expression (6.4). This results in quite few transitions, and thick layers as seen in the reference class profile.

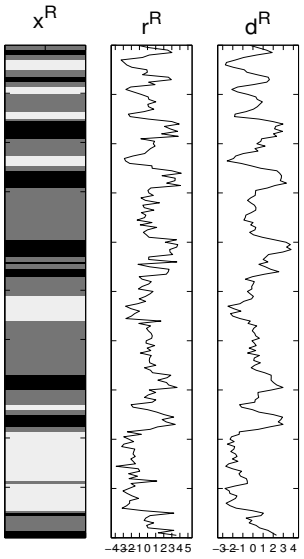
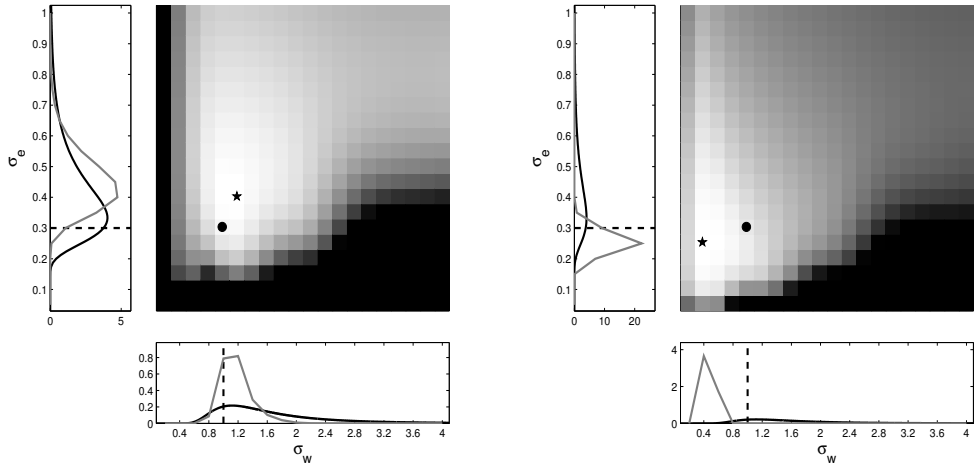


Figure 6.9: $\mathbf{P}_{11}$*: Reference class profile,* $\mathbf{x}^R$*, response profile,* $\mathbf{r}^R$*, and observation profile,* $\mathbf{d}^R$*.*

The parameter MAP estimation plots for the test case $\mathbf{P}_{11}$ are displayed in Figure 6.10. For $k = 1$, both parameters are a bit over-estimated when compared to the reference parameter value. For $k = 3$, the noise parameter is slightly under-estimated when compared to the reference parameter value, and the wavelet parameter is quite under-estimated. As for the base case, the marginal posterior functions are wider for the lower order, and the bivariate MAP estimate is equal to the marginal estimate for both orders.



(a) $k = 1$ (b) $k = 3$

Figure 6.10: $\mathbf{P}_{11}$*: Computed bivariate parameter posterior logarithm values for (a)* $k = 1$ *and (b)* $k = 3$*.*

The global class MAPs based on the parameter MAP estimate for $k = 3$, $\widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}$, for the test case $\mathbf{P}_{11}$ are displayed in Figure 6.11 together with the reference class and observation profile. For all orders k, the predictions reproduce the reference class profile very reliably. Except for a predicted thin white layer above the middle white layer for $k = 1, 2$, there is no significant difference between the kth order class MAPs.

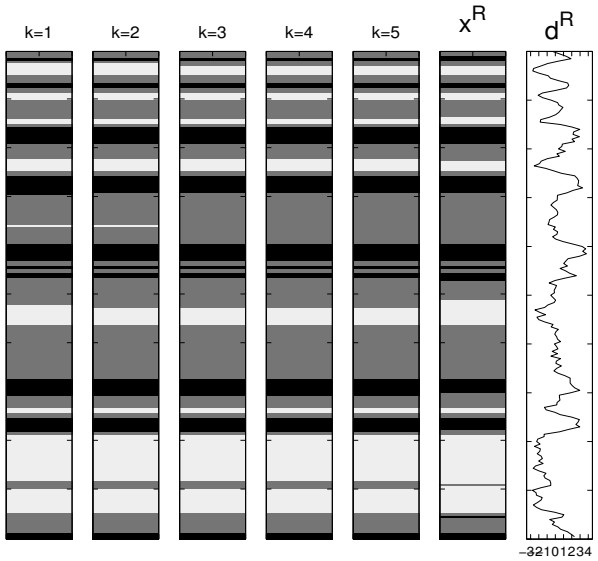


Figure 6.11: $\mathbf{P}_{11}$*: Global class MAPs based on the parameter MAP estimate for* $k = 3$, $[\hat{\mathbf{x}}|\mathbf{d}; \widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}]_{MAP}$.

The classes' coverage rate values and corresponding misclassification rates for the test case $\mathbf{P}_{11}$ are displayed in Figure 6.12 for the different orders k. We notice that the rates are significantly more reliable for the approximations based on the estimated parameters. Especially for the white and black class, the misclassification rates are very small for $k > 1$.

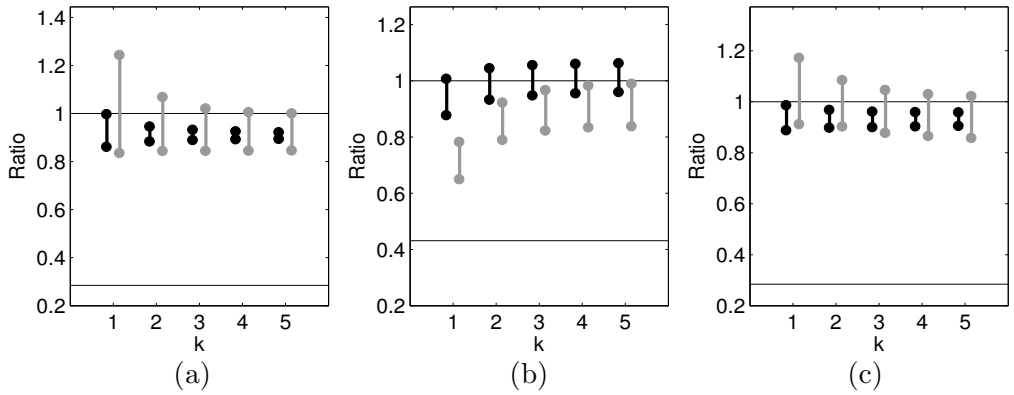


Figure 6.12: $\mathbf{P}_{11}$*: Misclassification rates,* $[a_l, b_l]$*, with the prior ratios marked for (a) the white class, (b) the grey class and (c) the black class.*

6.2.3 Test case: Observation model $\mathbf{P}_{12}$

In Figure 6.13, we see the reference class profile simulated from $\mathbf{P}_{12}$, and the corresponding response and observation profile. The prior model is here defined by the transition matrix

$$\mathbf{P}_{12} = \begin{pmatrix} 0.50 & 0.50 & 0 \\ 0.10 & 0.80 & 0.10 \\ 0 & 0.50 & 0.50 \end{pmatrix} , \tag{6.5}$$

with initial marginal prior $p(x_1) = (0.1429, 0.7143, 0.1429)$. This prior model clearly favors grey, as seen in the class profile, with small transition probabilities from the grey class to the other classes.

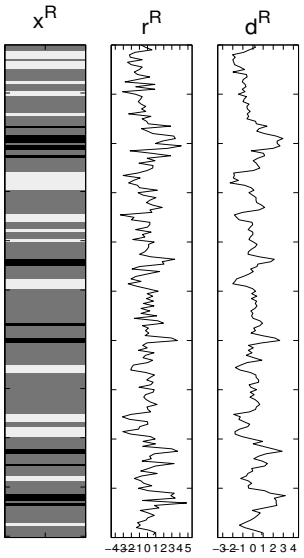
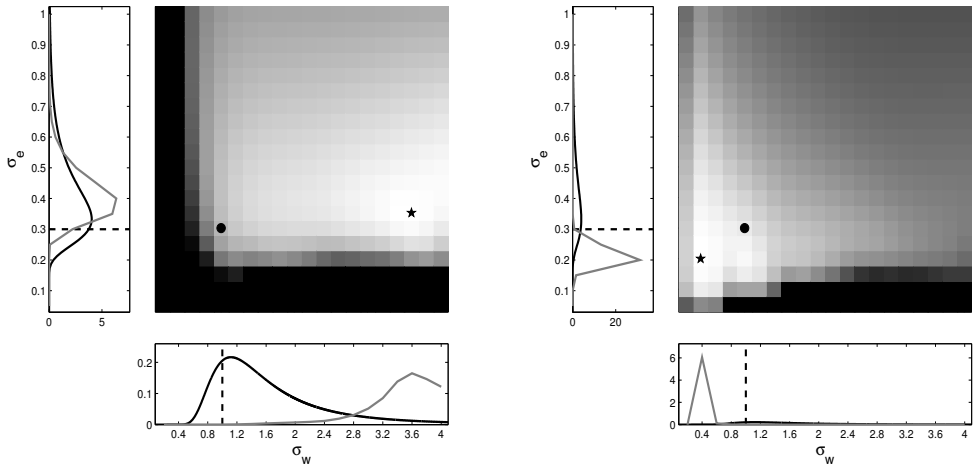


Figure 6.13: $\mathbf{P}_{12}$*: Reference class profile,* $\mathbf{x}^R$*, response profile,* $\mathbf{r}^R$*, and observation profile,* $\mathbf{d}^R$*.*

The parameter MAP estimation plots for the test case $\mathbf{P}_{12}$ are displayed in Figure 6.14. For $k = 1$, the wavelet parameter is strongly over-estimated when compared to the reference value. For $k = 3$, the plot is multimodal. Both parameters are a bit under-estimated, but a local maximum occurs for the posterior logarithm values of $(\sigma_w, \sigma_e) = (1, 0.2)$ which is closer to the reference value. Parts of the marginal wavelet parameter posterior function for $k = 1$ actually falls outside the chosen parameter interval. We also notice that for $k = 1$ the marginal noise parameter posterior MAP estimate is slightly larger then the bivariate noise MAP estimate.

The global class MAPs based on the parameter MAP estimate for $k = 3$, $\widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}$, for the test case $\mathbf{P}_{12}$ are displayed in Figure 6.15 together with the reference class and observation profile. The predictions are quite reliable for all orders k, with no great improvements for increasing orders. Some thin layers in $\mathbf{x}^R$ are not recognized, including transitions in the black layer neighborhoods which are predicted by thicker black layers.



(a) $k = 1$ (b) $k = 3$

Figure 6.14: $\mathbf{P}_{12}$*: Computed bivariate parameter posterior logarithm values for (a)* $k = 1$ *and (b)* $k = 3$*.*

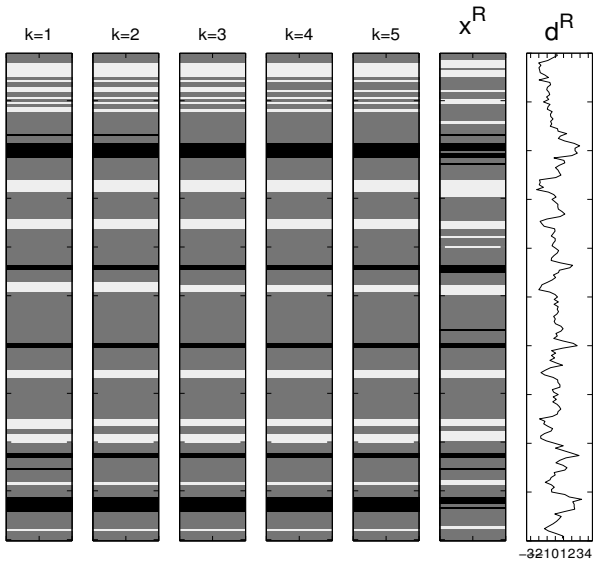


Figure 6.15: $\mathbf{P}_{12}$*: Global class MAPs based on the parameter MAP estimate for* $k = 3$*,* $[\hat{\mathbf{x}}|\mathbf{d}; \widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}]_{MAP}$*.*

The classes' coverage rate values and corresponding misclassification rates for the test case $\mathbf{P}_{12}$ are displayed in Figure 6.16 for the different orders k. The coverage rates seem to stabilize for orders $k > 2$, and the misclassification rates are more reliable for increasing k for the white and black class. As in the test case $\mathbf{P}_{11}$, see Figure 6.12, the rates are much more reliable for the posterior approximations based on the estimated parameters. Both the white and the black class is quite over-predicted for all orders k for the posterior approximations based on the reference parameters.

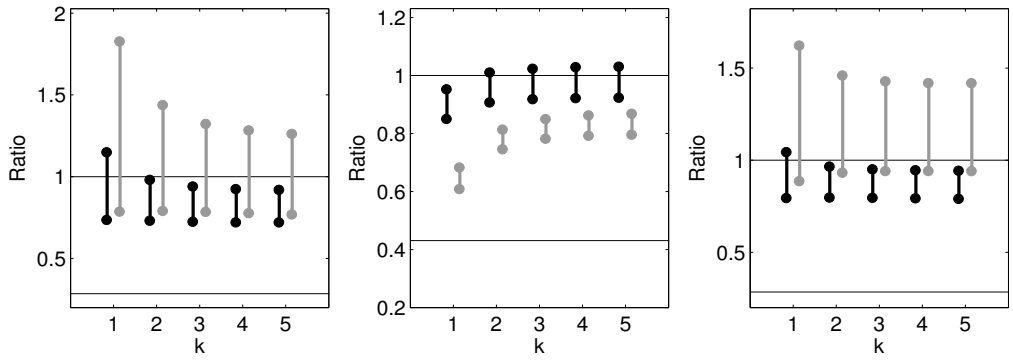


Figure 6.16: $\mathbf{P}_{12}$*: Misclassification rates,* $[a_l, b_l]$*, with the prior ratios marked for (a) the white class, (b) the grey class and (c) the black class.*

6.2.4 Test case: Observation model $\mathbf{P}_{13}$

In Figure 6.17, we see the reference class profile simulated from $\mathbf{P}_{13}$, and the corresponding response and observation profile. The prior model is here defined by the transition matrix

$$\mathbf{P}_{13} = \begin{pmatrix} 0.20 & 0.80 & 0 \\ 0.10 & 0.80 & 0.10 \\ 0 & 0.80 & 0.20 \end{pmatrix} \tag{6.6}$$

with initial marginal prior $p(x_1) = (0.10, 0.80, 0.10)$. Due to large transition probabilities from all classes into the grey class, the class profile is mainly grey with thin layers of white and black.

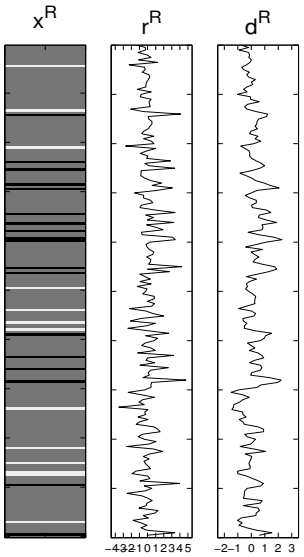
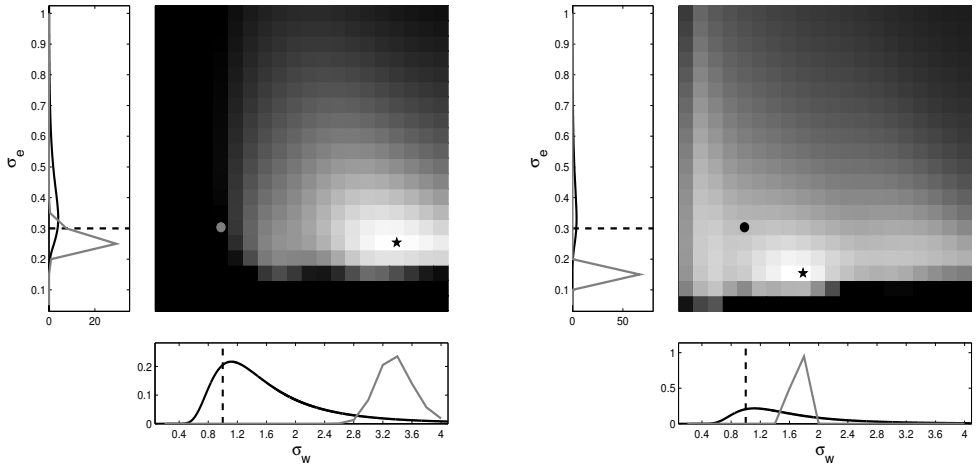


Figure 6.17: $\mathbf{P}_{13}$*: Reference class profile,* $\mathbf{x}^R$*, response profile,* $\mathbf{r}^R$*, and observation profile,* $\mathbf{d}^R$*.*

The parameter MAP estimation plots for the test case $\mathbf{P}_{13}$ are displayed in Figure 6.18. For $k = 1$, the wavelet parameter is strongly over-estimated when compared to the reference value. For $k = 3$, the wavelet parameter is still quite over-estimated while the noise parameter is under-estimated, and the plot seems to be multimodal.



(a) $k = 1$ (b) $k = 3$

Figure 6.18: $\mathbf{P}_{13}$*: Computed bivariate parameter posterior logarithm values for (a)* $k = 1$ *and (b)* $k = 3$*.*

The global class MAPs based on the parameter MAP estimate for $k = 3$, $\widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}$, for the test case $\mathbf{P}_{13}$ are displayed in Figure 6.19 together with the reference class and observation profile. The predictions are very poor for all orders k. Too many and too thick white and black layers are predicted due to the over-estimated wavelet parameter, especially for $k = 1$.

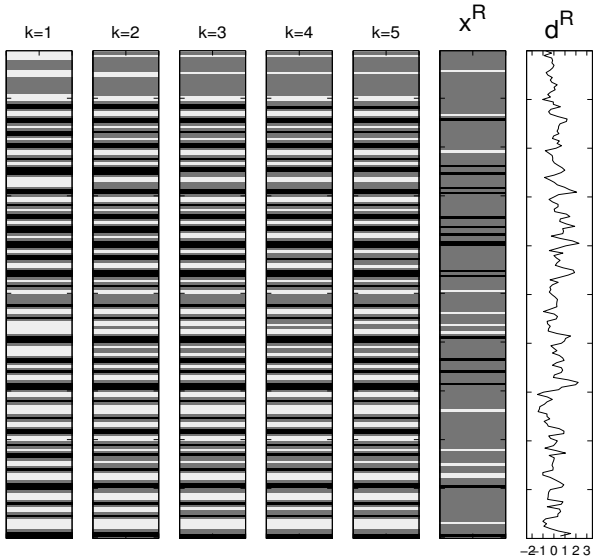
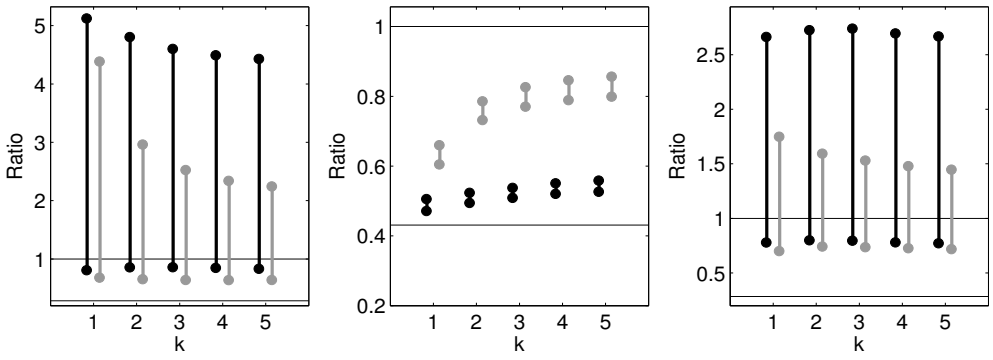


Figure 6.19: $\mathbf{P}_{13}$*: Global class MAPs based on the parameter MAP estimate for* $k = 3$*,* $[\hat{\mathbf{x}}|\mathbf{d}; \widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}]_{MAP}$*.*

The classes' coverage rate values and corresponding misclassification rates for the test case $\mathbf{P}_{13}$ are displayed in Figure 6.20 for the different orders k. The rates reflect the poor result in the class MAPs, i.e. the black and white classes are strongly over-predicted, and the grey class is accordingly under-predicted. The misclassification rates are a bit more reliable for the posterior approximations based on the reference parameters. These rates decrease for increasing orders k, but are also very poor.



(a) (b) (c)

Figure 6.20: $\mathbf{P}_{13}$*: Misclassification rates,* $[a_l, b_l]$*, with the prior ratios marked for (a) the white class, (b) the grey class and (c) the black class.*

6.2.5 Test case: Observation model $\mathbf{P}_{21}$

In Figure 6.21, we see the reference class profile simulated from $\mathbf{P}_{21}$, and the corresponding response and observation profile. The prior model is here defined by the transition matrix

$$\mathbf{P}_{21} = \begin{pmatrix} 0.80 & 0.20 & 0 \\ 0.33 & 0.34 & 0.33 \\ 0 & 0.20 & 0.80 \end{pmatrix} , \tag{6.7}$$

with initial marginal prior $p(x_1) = (0.3837, 0.2326, 0.3837)$. This prior model favors the white and the black classes, due to small transition probability from these classes into the grey class. The reference class thus consists mainly of thick white and black layers.

The parameter MAP estimation plots for the test case $\mathbf{P}_{21}$ are displayed in Figure 6.22. For both orders k, the MAP estimate is close to the reference value. The wavelet parameter is a bit under-estimated for both orders when compared to the reference value.

The global class MAPs based on the parameter MAP estimate for $k = 3$, $\widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}$, for the test case $\mathbf{P}_{21}$ are displayed in Figure 6.23 together with the reference class and observation profile. The predictions are quite reliable for all orders k. The thick black and white layers are quite well predicted, even if some of the internal transitions are not recognized.

The classes' coverage rate values and corresponding misclassification rates for the test case $\mathbf{P}_{21}$ are displayed in Figure 6.24 for the different orders k. The coverage rate values seem to stabilize for orders $k > 1$. For the white and black classes, the misclassification rates are quite reliable, decreasing for increasing orders k. For the grey class the misclassification rates are a bit more poor.

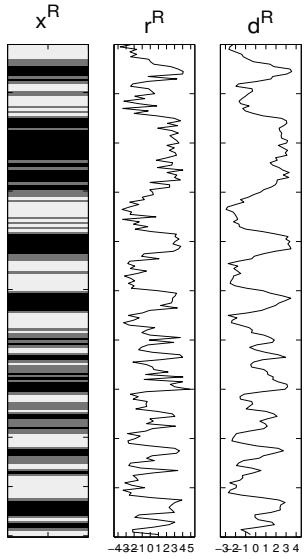
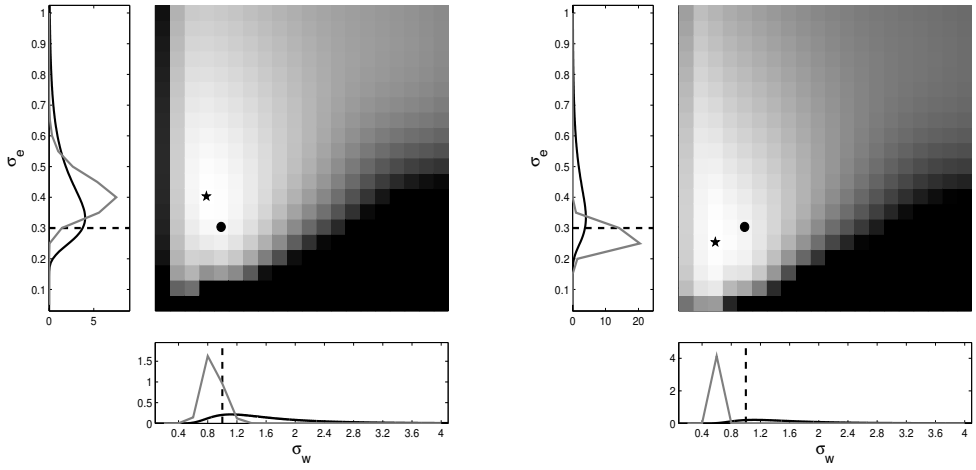


Figure 6.21: $\mathbf{P}_{21}$*: Reference class profile,* $\mathbf{x}^R$*, response profile,* $\mathbf{r}^R$*, and observation profile,* $\mathbf{d}^R$*.*



(a) $k = 1$ (b) $k = 3$

Figure 6.22: $\mathbf{P}_{21}$*: Computed bivariate parameter posterior logarithm values for (a)* $k = 1$ *and (b)* $k = 3$*.*

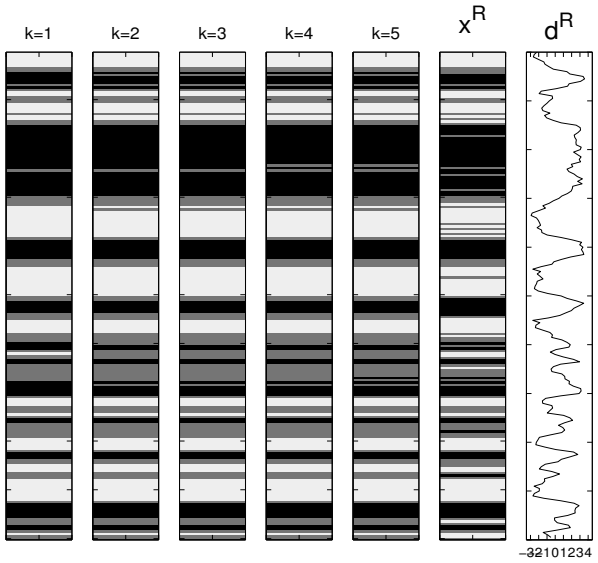


Figure 6.23: $\mathbf{P}_{21}$*: Global class MAPs based on the parameter MAP estimate for* $k = 3$*,* $[\hat{\mathbf{x}}|\mathbf{d}; \widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}]_{MAP}$*.*

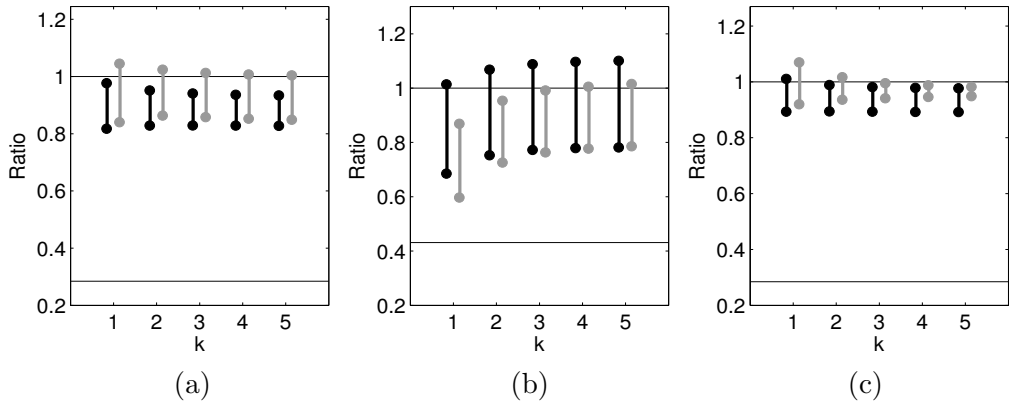


Figure 6.24: $\mathbf{P}_{21}$*: Misclassification rates,* $[a_l, b_l]$*, with the prior ratios marked for (a) the white class, (b) the grey class and (c) the black class.*

6.2.6 Test case: Observation model $\mathbf{P}_{23}$

In Figure 6.25, we see the reference class profile simulated from $\mathbf{P}_{23}$, and the corresponding response and observation profile. The prior model is here defined by the transition matrix

$$\mathbf{P}_{23} = \begin{pmatrix} 0.20 & 0.80 & 0 \\ 0.33 & 0.34 & 0.33 \\ 0 & 0.80 & 0.20 \end{pmatrix} \quad , \tag{6.8}$$

with initial marginal prior $p(x_1) = (0.2260, 0.5479, 0.2260)$. The reference class profile consists of frequent thin white and black layers between medium-sized grey layers, according to the prior model in Expression (6.8). The frequent class changes result in frequent fluctuations in the response and observation profile.

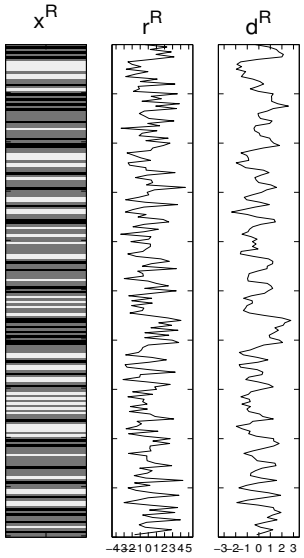
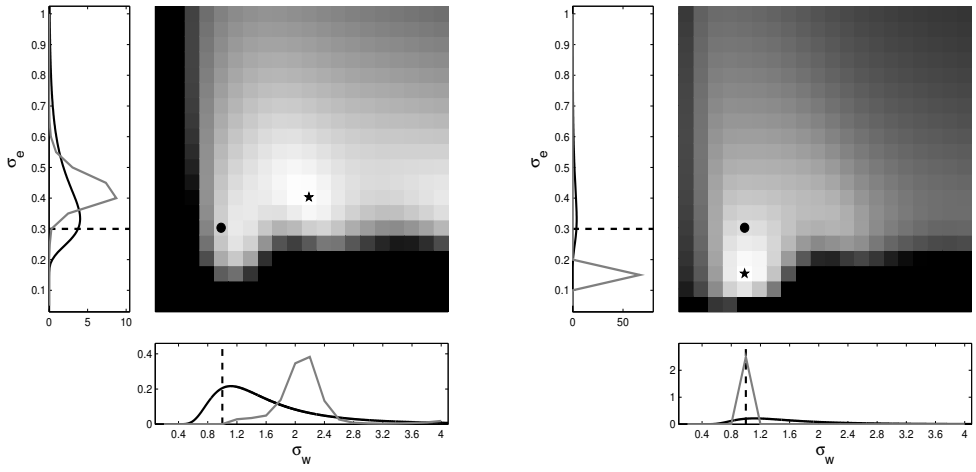


Figure 6.25: $\mathbf{P}_{23}$*: Reference class profile,* $\mathbf{x}^R$*, response profile,* $\mathbf{r}^R$*, and observation profile,* $\mathbf{d}^R$*.*

The parameter MAP estimation plots for the test case $\mathbf{P}_{23}$ are displayed in Figure 6.26. For $k = 1$, the wavelet parameter is over-estimated when compared to the reference value, and the plot is multimodal. For $k = 3$ the wavelet parameter MAP estimate is equal to the reference value, while the noise parameter is under-estimated.



(a) $k = 1$ (b) $k = 3$

Figure 6.26: $\mathbf{P}_{23}$*: Computed bivariate parameter posterior logarithm values for (a)* $k = 1$ *and (b)* $k = 3$*.*

The global class MAPs based on the parameter MAP estimate for $k = 3$, $\widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}$, for the test case $\mathbf{P}_{23}$ are displayed in Figure 6.27 together with the reference class and observation profile. The predictions resemble the reference class profile quite well for all orders k.

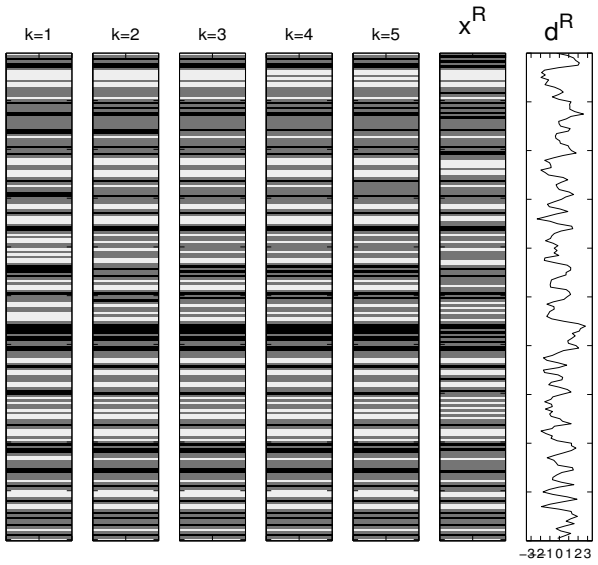


Figure 6.27: $\mathbf{P}_{23}$*: Global class MAPs based on the parameter MAP estimate for* $k = 3$*,* $[\hat{\mathbf{x}}|\mathbf{d}; \widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}]_{MAP}$.

The classes' coverage rate values and corresponding misclassification rates for the test case $\mathbf{P}_{23}$ are displayed in Figure 6.28 for the different orders k. The coverage rate values seem to stabilize for orders $k > 1$, and the misclassification rates are quite reliable. We notice that the rates are slightly more reliable for the posterior approximations based on the parameter MAP estimate.

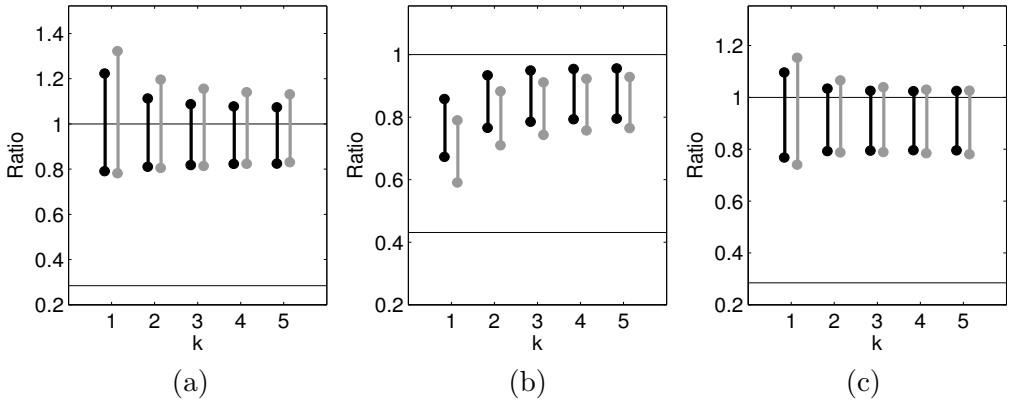


Figure 6.28: $\mathbf{P}_{23}$*: Misclassification rates,* $[a_l, b_l]$*, with the prior ratios marked for (a) the white class, (b) the grey class and (c) the black class.*

6.2.7 Test case: Observation model $\mathbf{P}_{31}$

In Figure 6.29, we see the reference class profile simulated from $\mathbf{P}_{31}$, and the corresponding response and observation profile. The prior model is here defined by the transition matrix

$$\mathbf{P}_{31} = \begin{pmatrix} 0.80 & 0.20 & 0 \\ 0.40 & 0.20 & 0.40 \\ 0 & 0.20 & 0.80 \end{pmatrix} \tag{6.9}$$

with initial marginal prior $p(x_1) = (0.40, 0.20, 0.40)$. This prior model clearly favors the white and black classes, due to the large transition probabilities into these classes. The reference class profile reflects this property, consisting of thick white and black layers, with thin grey layers in between them.

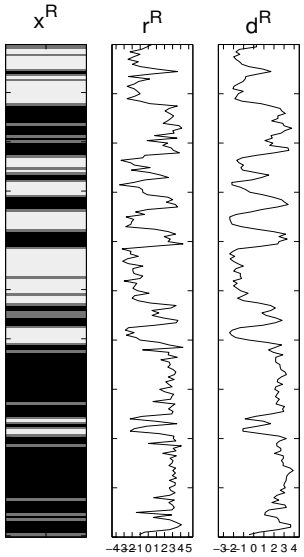
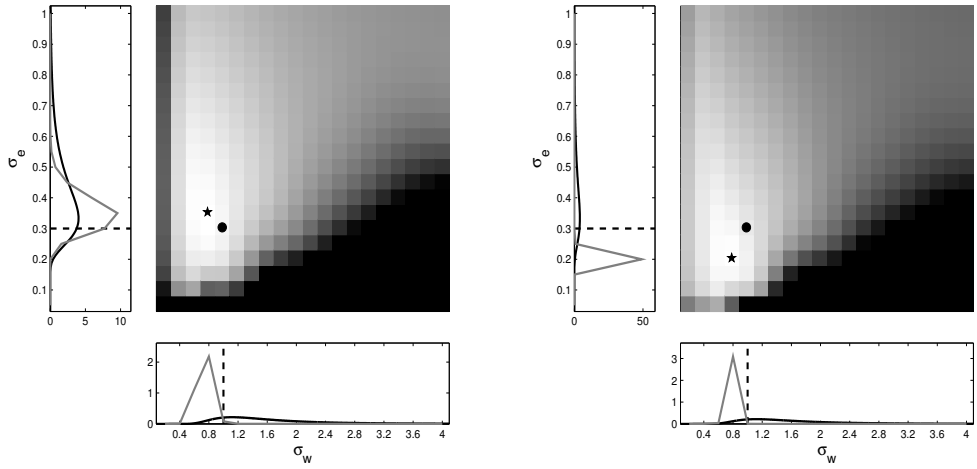


Figure 6.29: $\mathbf{P}_{31}$*: Reference class profile,* $\mathbf{x}^R$*, response profile,* $\mathbf{r}^R$*, and observation profile,* $\mathbf{d}^R$*.*

The parameter MAP estimation plots for the test case $\mathbf{P}_{31}$ are displayed in Figure 6.30. The MAP estimates are for both orders k close to the reference values. For both orders, the wavelet parameter is slightly under-estimated when compared to the reference value. The noise parameter is slightly over-estimated for $k = 1$ and a bit under-estimated for $k = 3$.

The global class MAPs based on the parameter MAP estimate for $k = 3$, $\widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}$, for the test case $\mathbf{P}_{31}$ are displayed in Figure 6.31 together with the reference class and observation profile. The predictions are reliable for all orders k, recognizing the large white and black layers well.

The classes' coverage rate values and corresponding misclassification rates for the test case $\mathbf{P}_{31}$ are displayed in Figure 6.32 for the different orders k. The misclassification rates are very reliable for the white and black class, and poor for the grey class due to few grey layers in the reference class profile.



(a) $k = 1$ (b) $k = 3$

Figure 6.30: $\mathbf{P}_{31}$*: Computed bivariate parameter posterior logarithm values for (a)* $k = 1$ *and (b)* $k = 3$*.*

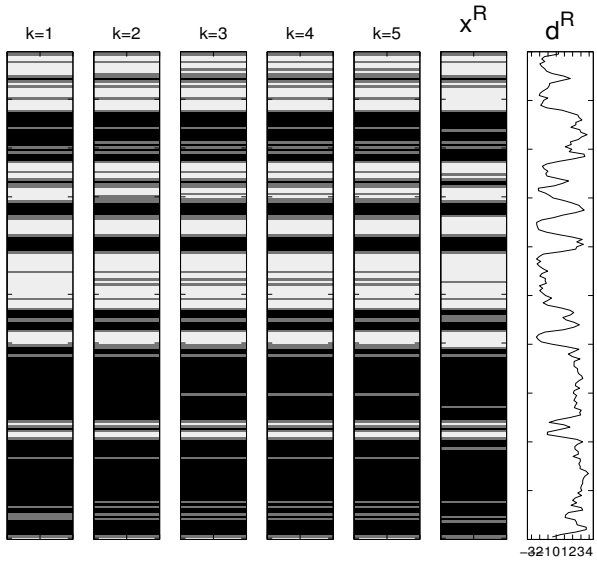


Figure 6.31: $\mathbf{P}_{31}$*: Global class MAPs based on the parameter MAP estimate for* $k = 3$*,* $[\hat{\mathbf{x}}|\mathbf{d}; \widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}]_{MAP}$*.*

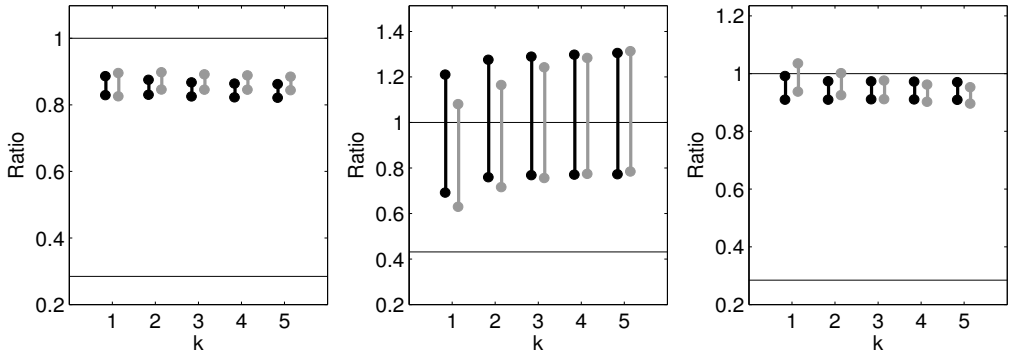


Figure 6.32: $\mathbf{P}_{31}$*: Misclassification rates,* $[a_l, b_l]$*, with the prior ratios marked for (a) the white class, (b) the grey class and (c) the black class.*

6.2.8 Test case: Observation model $\mathbf{P}_{32}$

In Figure 6.33, we see the reference class profile simulated from $\mathbf{P}_{32}$, and the corresponding response and observation profile. The prior model is here defined by the transition matrix

$$\mathbf{P}_{32} = \begin{pmatrix} 0.50 & 0.50 & 0 \\ 0.40 & 0.20 & 0.40 \\ 0 & 0.50 & 0.50 \end{pmatrix} \quad , \tag{6.10}$$

with initial marginal prior $p(x_1) = (0.3077, 0.3846, 0.3077)$. We should thus expect class profiles with frequent transitions between medium-sized white and black layers and thin grey layers in between them, as reflected by the reference class profile.

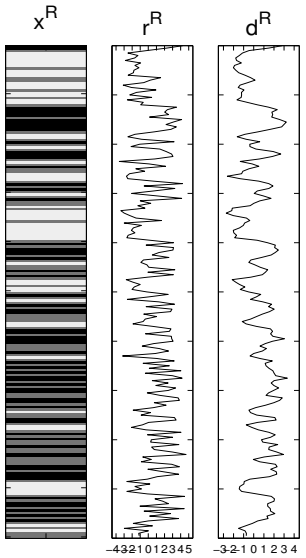
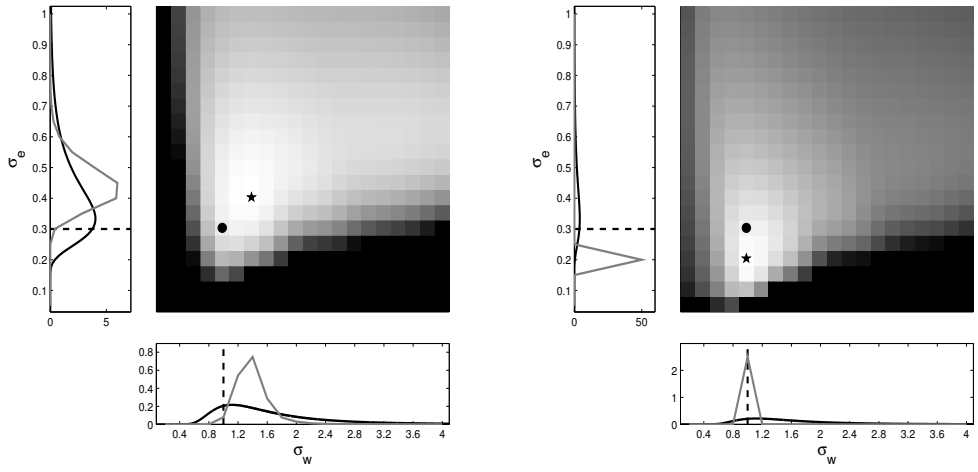


Figure 6.33: $\mathbf{P}_{32}$*: Reference class profile,* $\mathbf{x}^R$*, response profile,* $\mathbf{r}^R$*, and observation profile,* $\mathbf{d}^R$*.*

The parameter MAP estimation plots for the test case $\mathbf{P}_{32}$ are displayed in Figure 6.34. The MAP estimates are for both orders k close to the reference values. For $k = 1$, both parameters are a bit over-estimated when compared to the reference values, and the marginal noise parameter MAP estimate is slightly larger then the bivariate estimate. For $k = 3$ the wavelet parameter MAP estimate is equal to the reference value, while the noise parameter is a bit under-estimated.



(a) $k = 1$ (b) $k = 3$

Figure 6.34: $\mathbf{P}_{32}$*: Computed bivariate parameter posterior logarithm values for (a)* $k = 1$ *and (b)* $k = 3$*.*

The global class MAPs based on the parameter MAP estimate for $k = 3$, $\widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}$, for the test case $\mathbf{P}_{32}$ are displayed in Figure 6.35 together with the reference class and observation profile. For $k = 1$ a bit to thick black layers are predicted, while the predictions for $k > 1$ are slightly more reliable and stable.

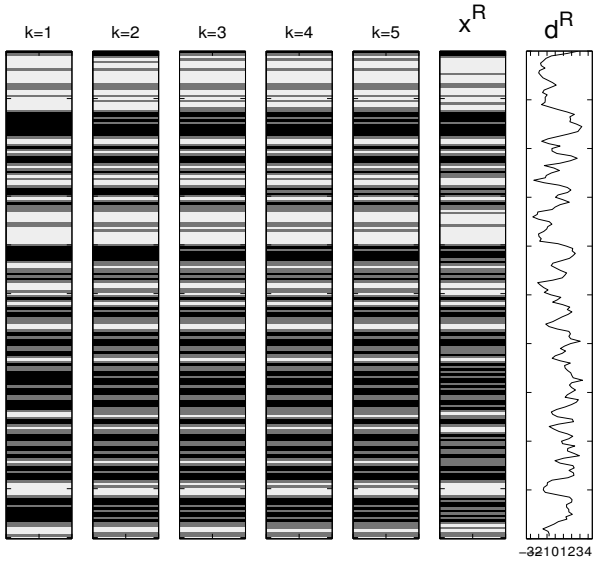


Figure 6.35: $\mathbf{P}_{32}$*: Global class MAPs based on the parameter MAP estimate for* $k = 3$*,* $[\hat{\mathbf{x}}|\mathbf{d}; \widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}]_{MAP}$*.*

The classes' coverage rate values and corresponding misclassification rates for the test case $\mathbf{P}_{32}$ are displayed in Figure 6.36 for the different orders k. The misclassification rates are quite reliable for the white and black class, decreasing for increasing orders k.

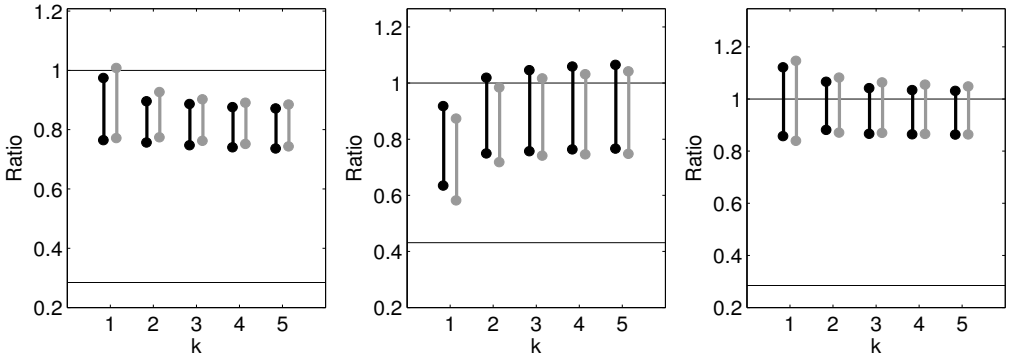


Figure 6.36: $\mathbf{P}_{32}$*: Misclassification rates,* $[a_l, b_l]$*, with the prior ratios marked for (a) the white class, (b) the grey class and (c) the black class.*

6.2.9 Test case: Observation model $\mathbf{P}_{33}$

In Figure 6.37, we see the reference class profile simulated from $\mathbf{P}_{33}$, and the corresponding response and observation profile. The prior model is here defined by the transition matrix

$$\mathbf{P}_{33} = \begin{pmatrix} 0.20 & 0.80 & 0 \\ 0.40 & 0.20 & 0.40 \\ 0 & 0.80 & 0.20 \end{pmatrix} \tag{6.11}$$

with initial marginal prior $p(x_1) = (0.25, 0.50, 0.25)$. The reference class profile now consists of thin layers of all classes with very frequent transitions, due to large transition probabilities out of the classes by Expression (6.11).

The parameter MAP estimation plots for the test case $\mathbf{P}_{33}$ are displayed in Figure 6.38. For $k = 1$, the wavelet parameter is strongly over-estimated when compared to the reference value. As for the MAP plot for the test case $\mathbf{P}_{12}$ for $k = 1$, see Figure 6.14, parts of the marginal wavelet parameter posterior function falls outside the chosen parameter interval. We notice that both the MAP plot and the marginal wavelet parameter posterior function for $k = 1$ are multimodal. For $k = 3$ the joint parameter MAP estimate is close to the reference value.

The global class MAPs based on the parameter MAP estimate for $k = 3$, $\widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}$, for the test case $\mathbf{P}_{33}$ are displayed in Figure 6.39 together with the reference class and observation profile. The predictions are quite reliable and early improving for increasing orders k. For $k = 1$, a bit to thick white and black layers are predicted.

The classes' coverage rate values and corresponding misclassification rates for the test case $\mathbf{P}_{33}$ are displayed in Figure 6.40 for the different orders k. The misclassification rates for the grey and black class are a bit poor, and stable for all orders k. We notice that the rates for the posterior approximations based on the parameter MAP estimate are a bit more reliable, except for the black class.

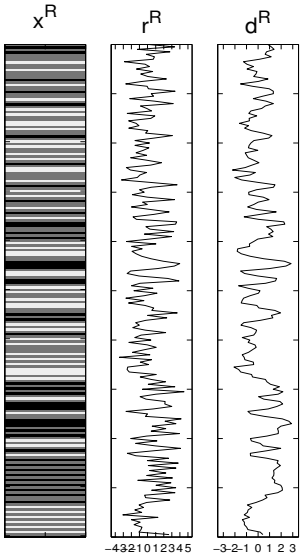
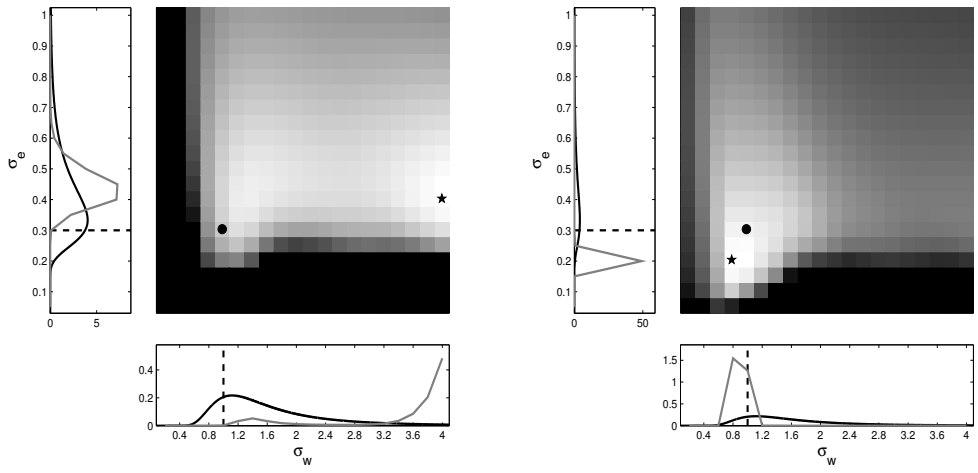


Figure 6.37: $\mathbf{P}_{33}$*: Reference class profile,* $\mathbf{x}^R$*, response profile,* $\mathbf{r}^R$*, and observation profile,* $\mathbf{d}^R$*.*



(a) $k = 1$ (b) $k = 3$

Figure 6.38: $\mathbf{P}_{33}$*: Computed bivariate parameter posterior logarithm values for (a)* $k = 1$ *and (b)* $k = 3$*.*

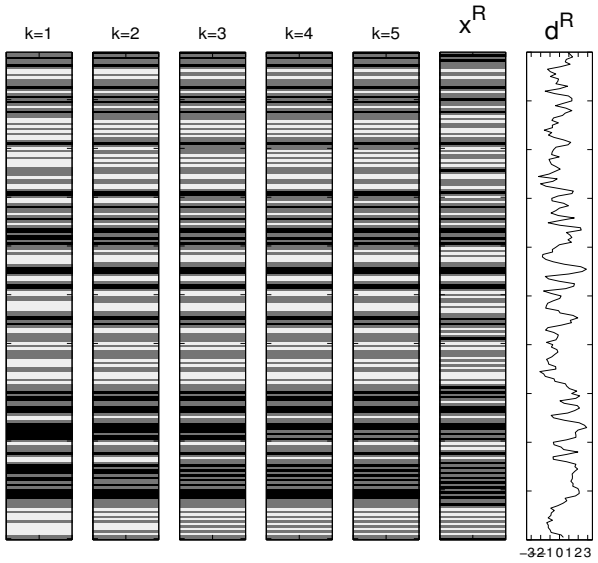


Figure 6.39: $\mathbf{P}_{33}$*: Global class MAPs based on the parameter MAP estimate for* $k = 3$*,* $[\hat{\mathbf{x}}|\mathbf{d}; \widehat{\boldsymbol{\theta}}^{(3)}_{d,MAP}]_{MAP}$*.*

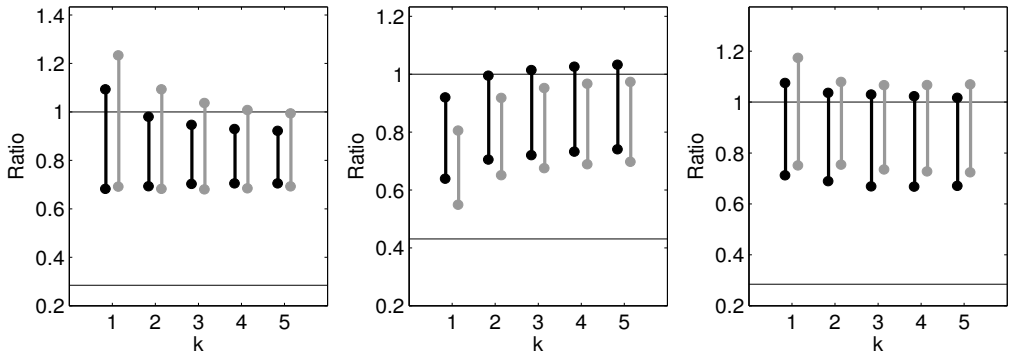


Figure 6.40: $\mathbf{P}_{33}$*: Misclassification rates,* $[a_l, b_l]$*, with the prior ratios marked for (a) the white class, (b) the grey class and (c) the black class.*

6.3 Discussion

The two acquisition likelihood parameters have been estimated by simultaneous maximum a posteriori prediction (MAP) estimation, by Expression (4.11), as described in Section 4.2.2. This estimation is based on the nine test cases in Table 6.1, with nine different reference class profiles simulated from prior models with varying transition matrices. The behavior of the reference class profiles clearly follows from these respective prior transition matrices. The results from the parameter estimation and the posterior model approximation are in general quite reliable, and the reference class profile is well recognized in the predictions.

As for the test study in Chapter 5, the noise parameter is generally under-estimated for $k = 3$ when compared to the reference value. As mentioned in the discussion in Section 5.3 this might be because the profiles are too short, the MMLEs might be biased or due to the fact that we utilize every observation data point k times in the posterior model approximations. The wavelet parameter MAP estimate is in general closer to the reference value for the larger order $k = 3$. For the base case, we showed that the kth order MMLE and MAP plots are not that different because the computed marginal log-likelihood values are very large. Estimation of the parameters by MMLE should thus be sufficient, in which we do not need to assign prior distributions to the parameters.

The approximated posterior models are, as for the test study in Chapter 5, generally more reliable for early increasing orders k, and stabilizing for $k = 2, 3$. We have plotted the logarithm of the test study S/N-ratios in Table 6.2, against the logit of the class MAP ratio for $k = 3$, see Figure 6.41. As for the test study in Chapter 5, see Figure 5.28, the results are more reliable for increasing S/N-ratios. A suggested linear trend, with respect to the log-logit transform, is included in the plot. We notice how the plotted values fall outside of the MC-trend in Figure 5.28 as we now tune the prior model parameters and not the noise parameter. We also notice that the results from the class predictions based on the estimated parameters are as reliable, and often significantly better, than the predictions based on the reference parameters. The outlying test case $\mathbf{P}_{13}$, with the smallest S/N-ratio, is where the estimation and corresponding prediction is most poor. The wavelet parameter is over-estimated only in this test case, probably because the observation profile is too smooth, see Figure 6.17, due to few class transitions.

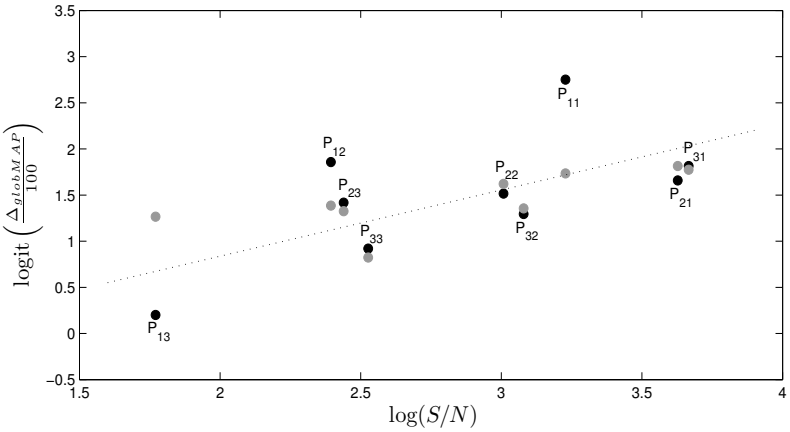


Figure 6.41: *The logarithm of the test cases' S/N-ratios,* $\log(S/N)$ *plotted against the logit of the respective global class MAP profile ratios, logit*$(\Delta_{globMAP}/100)$*, for* $k = 3$*.*

Chapter 7

Example: Seismic Inversion Model

As our empirical example, we will use a model with four possible classes, with the state space $\Omega_x = \{\text{white,light-grey,dark-grey,black}\}$. A representation of the model is displayed in Figure 7.1 along a 1D class profile of length $T = 200$, i.e. $\mathbf{x} : \{x_t; t = 1, \ldots, 200\}$, which will be our reference class profile, $\mathbf{x}^R$. The associated profiles in Figure 7.1, $\mathbf{r}^R$ and $\mathbf{d}^R$, will be our reference response and observation profile respectively. A physical interpretation is that the class profile constitutes a geological sequence of lithology/fluid classes. The four classes in the state space corresponds to the four lithology-fluid classes gas-saturated sandstone, oil-saturated sandstone, brine-saturated sandstone and shale respectively, see e.g. References [14] and [21]. The main objective will be to estimate the parameters in the acquisition likelihood model, i.e. the convolution variance, σ_w^2, and the noise variance, σ_e^2, based on the observation profile. The next objective will then be to predict class profiles, $\hat{\mathbf{x}}$, using the observation profile and the estimated parameters, and compare the reproductions to the reference profile, $\mathbf{x}^R$. The estimation and prediction procedures are based on the results from the two forgoing chapters.

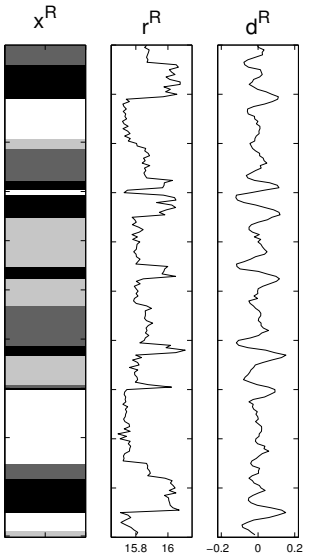


Figure 7.1: *Reference profile,* $\mathbf{x}^R$*, and an associated response profile,* $\mathbf{r}^R$*, and observation profile,* $\mathbf{d}^R$*.*

7.1 Model specifications

In our seismic model, there are $L = 4$ possible states. The reference class profile in Figure 7.1 is simulated from the prior distribution defined in Section 2.1, with the (4×4) transition probability matrix

$$\mathbf{P}_x = \begin{pmatrix} 0.9441 & 0 & 0 & 0.0559 \\ 0.0430 & 0.9146 & 0 & 0.0424 \\ 0.0063 & 0.0230 & 0.9423 & 0.0284 \\ 0.0201 & 0.0202 & 0.1006 & 0.8591 \end{pmatrix} . \tag{7.1}$$

The initial marginal prior distribution is defined as the stationary pdf of $\mathbf{P}_x$ in Expression (5.1), i.e. $p(x_1) = (0.2416, 0.1552, 0.3833, 0.2198)$. Considering the geophysical interpretation, gas-saturated sandstone can naturally only occur beneath shale stone by Expression (7.1), and brine-saturated sandstone can not occur above oil-saturated sandstone. In the parameter estimation algorithms, we will use a less informative prior model, defined by the prior transition matrix

$$\mathbf{P}_x = \begin{pmatrix} 0.95 & 0 & 0 & 0.05 \\ 0.05 & 0.90 & 0 & 0.05 \\ 0.03 & 0.03 & 0.91 & 0.03 \\ 0.03 & 0.03 & 0.10 & 0.84 \end{pmatrix} , \tag{7.2}$$

with initial marginal prior distribution $p(x_1) = (0.4091, 0.1364, 0.2392, 0.2153)$. We notice how the relative small change in $\mathbf{P}_x$ results in a significant change in the marginal prior. The prior transition matrix in Expression (7.1) resembles the prior transition matrix in the test case $\mathbf{P}_{11}$ in Chapter 6, see Expression (6.4), with large transition probabilities from each class into themselves. Based on results from this test case, when using the simpler prior model in Expression (7.2) in the computations, we could expect as good results as when using the true prior model in Expression (7.1).

We assume a response likelihood model with no spatial dependency, i.e. $p(\mathbf{r}|\mathbf{x}) = \prod_{t=1}^{T} \phi_1(r_t; \mu_{x_t}, \sigma_x^2)$, where $\mu_{x_t} \in \{15.7402, 15.8001, 15.8671, 16.0120\}$ and $\sigma_{x_t}^2 \in \{0.0004, 0.0003, 0.0002, 0.0011\}$ corresponding to the respective classes in Ω_x. The resulting marginal response pdfs are displayed in Figure 7.2. We observe how the white and light-grey classes' response pdfs clearly overlap, and the black class pdf is wider corresponding to the larger variance. The response likelihood model corresponds to a rock physics likelihood model, i.e. it represents the seismic impedance for each geological class, see References [3], [4], [12], [14], [21].

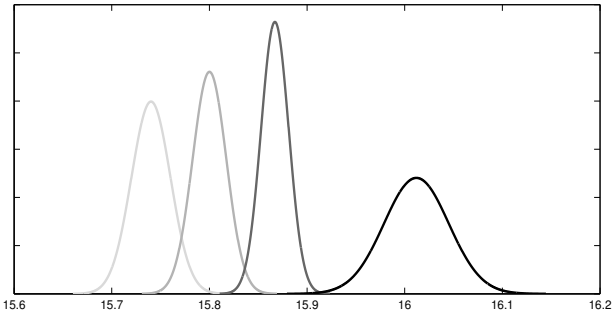


Figure 7.2: *Class response Gaussian pdfs, $\phi_1(r_t; \mu_{x_t}, \sigma_{x_t}^2)$, corresponding to the classes white, light-grey, dark-grey and black respectively.*

We assume an acquisition likelihood model, as defined by Expression (2.10), with white noise only, i.e. $p(\mathbf{d}|\mathbf{r}) = N_T(\mathbf{Wr}, \sigma_e^2\mathbf{I})$ with $\sigma_e = 0.01$. The convolution matrix is now defined as

$$\mathbf{W} = \mathbf{R}\boldsymbol{\Delta} \ ,$$

where Δ is a 3-band wide contrast matrix with band elements $(-0.5, 0, 0.5)$ and $\mathbf{R}$ is a Ricker convolution matrix. The Ricker wavelets, $\mathbf{w}_R$, which are the rows in the Ricker convolution matrix, $\mathbf{R}$, are assumed to be discrete values of the normalized Ricker wavelet function

$$\mathbf{w}_R(t) = \frac{2}{\sqrt{3\sigma_w}\pi^{1/4}} \left(1 - \frac{t^2}{\sigma_w^2}\right) e^{\frac{-t^2}{2\sigma_w^2}} \ . \tag{7.3}$$

This is the negative normalized second derivative of a Gaussian function, i.e., up to scale and normalization, the second Hermite function, see Reference [18]. The standard definition of the Ricker wavelet is

$$\mathbf{w}_R(t) = \left(1 - 2\pi^2 f^2 t^2\right) e^{-\pi^2 f^2 t^2} \ , \tag{7.4}$$

see Reference [15], where f is frequency in Hz. The relationship between the convolution variance parameter and the frequency is thus by Expression (7.3)

$$\sigma_w = \frac{1}{\sqrt{2}\pi f} \ . \tag{7.5}$$

The wavelet parameter used in our empirical example is $\sigma_w = 3$, which by Expression (7.5) corresponds to a 0.075Hz wavelet. The resulting Ricker wavelet and its contrast wavelet is displayed in Figure 7.3, where the marked function values define the discretized wavelets. The wavelet length is 31. The signal-to-noise ratio is $S/N = 23.9$ by Expression (5.3).

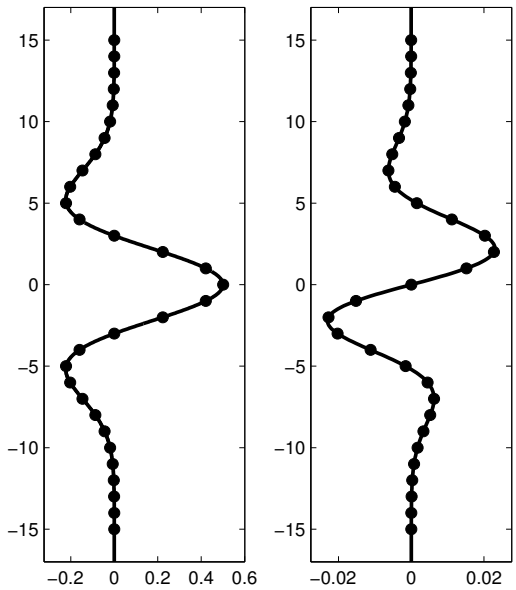
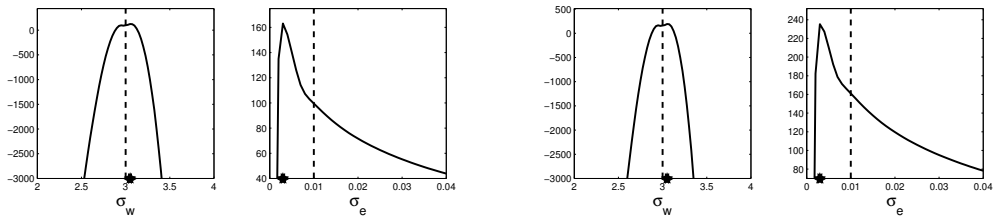


Figure 7.3: *(a) The Ricker wavelet, $\mathbf{w}_R$, given by Expression (7.3) for $\sigma_w = 3$ and (b) its contrast wavelet by $\mathbf{w}_c = \mathbf{w}_R\Delta$.*

7.2 Results with discussion

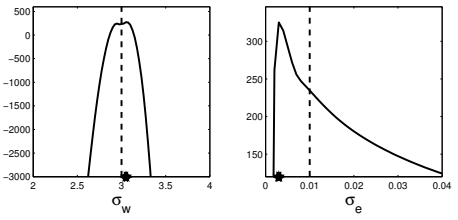
In this seismic example, we perform univariate and simultaneous MMLE estimation of the current two acquisition parameters, σ_w and σ_e, as done in the test study in Chapter 5. The wavelet parameter now has a different interpretation, as we now utilize the Ricker wavelet function in Expression (7.3) instead of the Gaussian wavelet function in Expression (2.11). The estimation is based on the observation profile in Figure 7.1, and all other parameters are kept fixed by their reference values given in Section 7.1. We then assess the approximate posterior model by the P-P ratio deconvolution algorithm for orders $k = 1, 2, 3, 4$, as described in Sections 3.2 and 3.3, based on the MMLE estimate for $k = 3$, i.e. $\widehat{\boldsymbol{\theta}}^{(3)}_{d,MMLE}$. We evaluate the approximate posterior model by the class MAP, and by the coverage rates given by Expressions (5.5) and (5.6) as done in the test studies in Chapters 5 and 6.

The univariate estimation plots for $k = 1, 2, 3, 4$ are displayed in Figure 7.4, where both parameters have very stable estimates for all orders k. The univariate wavelet parameter MMLE is very close to the reference value, while the univariate noise parameter MMLE is quite under-estimated. The simultaneous parameter plots for $k = 1, 3$ are displayed in Figure 7.5, with very stable estimates as well. As for the univariate case, the wavelet parameter estimate is almost equal to the reference value, while the noise parameter is quite under-estimated.

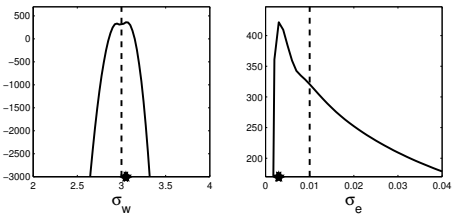


(a) $k = 1$

(b) $k = 2$



(c) $k = 3$

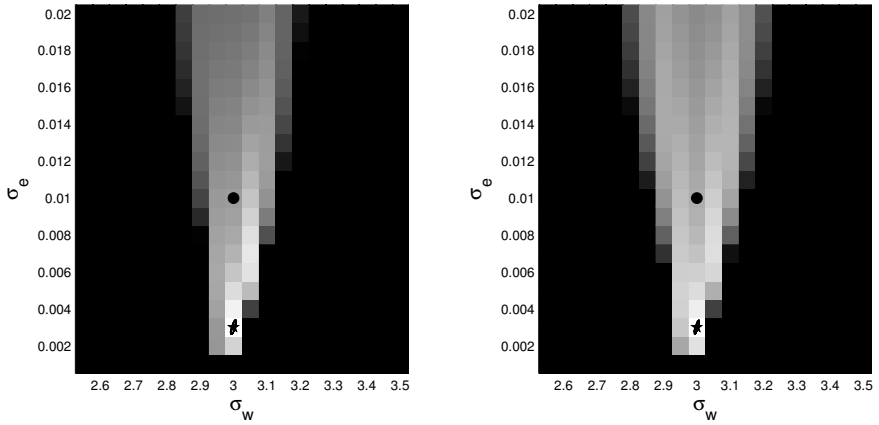


(d) $k = 4$

Figure 7.4: *Marginal log-likelihood plots for univariate parameter estimation.*

The locationwise and global class MAPs for $k = 1, 2, 3, 4$ based on the simultaneous parameter MMLE for $k = 3$, $\widehat{\boldsymbol{\theta}}^{(3)}_{d,MMLE}$, are displayed in Figure 7.6 together with the reference class and observation profile. The predictions are quite poor for all orders k, and are mutually significantly different. The class MAPs based on the reference parameters, $\boldsymbol{\theta}_d$, are likewise displayed in Figure 7.7. We notice how these predictions recognize the transitions in the reference profile significantly more reliable, even though the wavelet parameter is the same in the two approximations. In the class MAPs in both figures, the black class seem to be

recognized most reliably, in accordance to the more distinct response distribution, see Figure 7.2. For the class MAPs based on the reference parameters, the main misclassification is between the white and the light-grey classes, whose response distributions overlap the most, see Figure 7.2. We notice how some of the locationwise MAPs based on the parameter MMLE estimate contain class transitions prohibited by the prior model, see Expression (7.2). These are class transitions from white to light-grey and light-grey to dark-grey. As these MAPs consider the approximate posterior model locationwise independently, thereby the name, these illegal transitions might occur. The global MAPs however do not adopt illegal transitions.



(a) $k = 1$ (b) $k = 3$

Figure 7.5: *Computed marginal log-likelihood values for the two acquisition likelihood parameters for (a)* $k = 1$ *and (b)* $k = 3$*.*

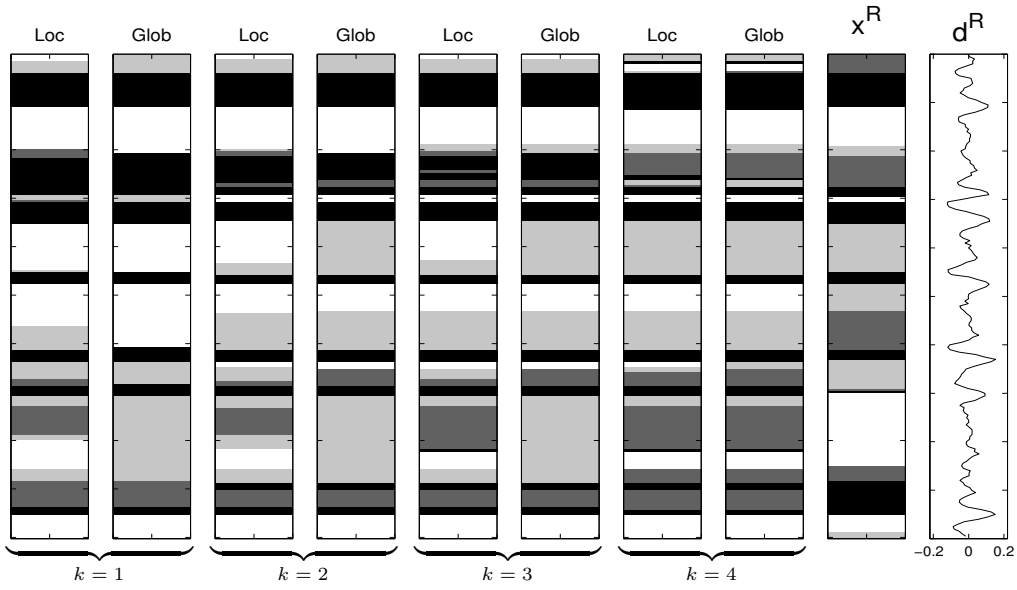


Figure 7.6: *Locationwise and global MAPs based on the 3rd order parameter MMLE estimate,* $[\hat{\mathbf{x}}|\mathbf{d}; \widehat{\boldsymbol{\theta}}^{(3)}_{d,MMLE}]_{MAP}$*.*

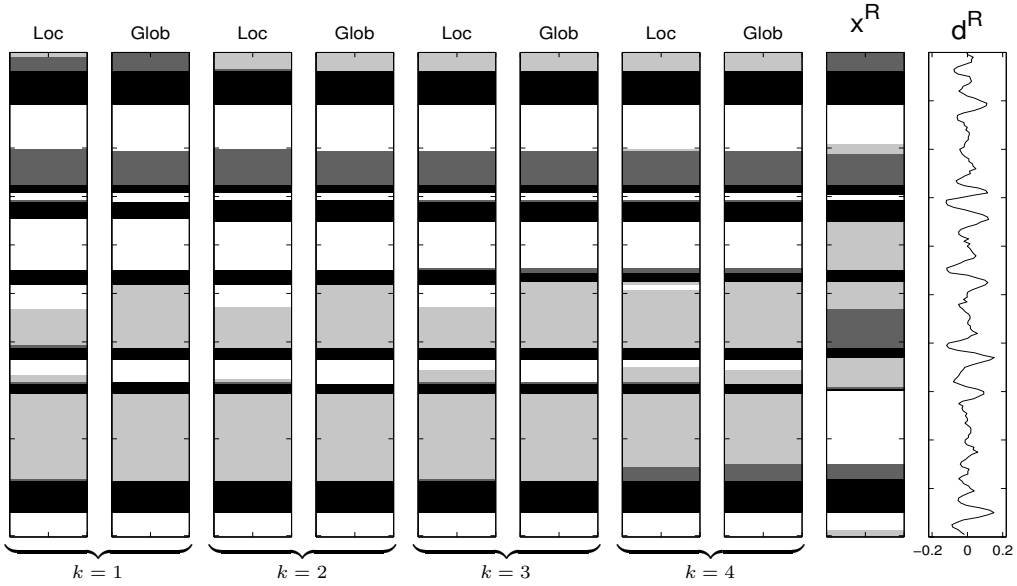


Figure 7.7: *Locationwise and global MAPs based on the reference parameter,* $[\hat{\mathbf{x}}|\mathbf{d};\boldsymbol{\theta}_d]_{MAP}$.

The class profile ratios from the global class MAPs in Figures 7.6 and 7.7 are presented in Table 7.1. The ratios are poor for both parameter sets, and increases for increasing orders k. We also notice that the predictions based on the true parameters are only slightly more reliable. In Table 7.2, we have computed similar class profile ratios, but here we consider white and light-grey as the same class. The physical interpretation is that we consider gas-saturated sandstone and oil-saturated sandstone as one class, i.e. as a hydrocarbon class, which would be the class of most interest. We now observe how the rates increase significantly, especially for the MAPs based on the true parameters. This is in accordance to the MAPs in Figure 7.7 where the main misclassification was between the white and light-grey class. The rates are however still quite poor for the MAPs based on the estimated parameters.

$(\Delta_{locMAP}, \Delta_{globMAP})$	$k=1$	$k=2$	$k=3$	$k=4$
$\widehat{\boldsymbol{\theta}}^{(3)}_{d,MMLE}$:	$(41.0, 35.5)$	$(42.5, 44.0)$	$(42.5, 45.0)$	$(53.5, 54.5)$
$\boldsymbol{\theta}_d$:	$(47.5, 51.5)$	$(45.0, 48.5)$	$(46.0, 48.0)$	$(53.0, 54.0)$

Table 7.1: *Class profile ratios in percentage for the locationwise and global MAPs.*

$(\Delta_{locMAP}, \Delta_{globMAP})$	$k=1$	$k=2$	$k=3$	$k=4$
$\widehat{\boldsymbol{\theta}}^{(3)}_{d,MMLE}$:	$(61.5, 69.5)$	$(64.5, 68.5)$	$(61.0, 68.5)$	$(64.0, 64.0)$
$\boldsymbol{\theta}_d$:	$(82.5, 83.5)$	$(80.5, 81.0)$	$(79.5, 79.5)$	$(82.0, 83.0)$

Table 7.2: *Class profile ratios in percentage for the locationwise and global MAPs considering white and light-grey as the same class.*

The classes' coverage rate values and corresponding misclassification rates are displayed in Figure 7.8 for the different orders k, and are in general quite poor. We notice that the rates seem to be stable only for the black class, whose misclassification rates are quite reliable according to the distinct marginal response distribution. For both the dark-grey and the black class, the misclassification rates are significantly more reliable for the approximate posterior models based on the true parameters. Except for the black class, the coverage rate value of a_l, i.e. the ability to predict the l-class in $\mathbf{x}^R$ correctly, is poor.

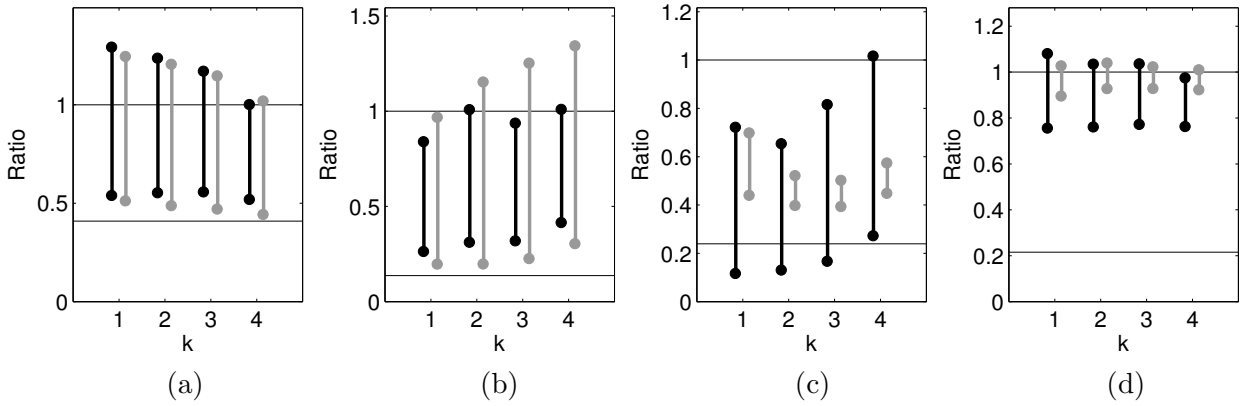


Figure 7.8: *Misclassification rates, $[a_l, b_l]$, with the prior ratios marked for (a) the white class, (b) the light-grey class, (c) the dark-grey class and (d) the black class.*

The S/N-ratio in our seismic example is relatively large, $S/N = 23.9$, and based on the results from Chapters 5 and 6, we should expect more reliable posterior model approximations. The seismic model is, however, quite more complex than the test study models, e.g. there are now four possible classes, and the prior model is not as simple. The convolution matrix is also different, utilizing the Ricker wavelet. This wavelet could have been chosen by Expression (7.4), i.e. not normalized, and based on the results from Chapter 5, the discrete reference wavelet chosen is perhaps to wide, consisting of 31 elements. The MMLE of the wavelet parameter is however very close to the reference value, and the major misclassification factor originate from under-estimation of the noise parameter. In the test studies, the noise parameter was consistently under-estimated when compared to the true value, but this did not affect the posterior model approximation significantly. In our seismic example however, the noise is large when compared to the mean differences in the response pdfs, which makes class recognition more difficult.

Chapter 8

Conclusion and further work

In this master's thesis, we study lithology/fluid (LF) prediction along a 1D-profile based on recorded prestack seismic data. The subsurface LF-classes are denoted by $\mathbf{x}$, which corresponds to a response, denoted by $\mathbf{r}$. The actual observed data, which is recorded as a convolution of $\mathbf{r}$, is denoted by $\mathbf{d}$. The inverse problem is ill-posed, which in addition to model and observation error makes the prediction challenging. We have approached the problem in a Bayesian setting. One objective has been to find the posterior distribution, $p(\mathbf{x}|\mathbf{d})$, given by a prior distribution, $p(\mathbf{x})$, and a convolved likelihood distribution, $p(\mathbf{d}|\mathbf{x})$, consisting of a response likelihood distribution $p(\mathbf{r}|\mathbf{x})$ and an acquisition likelihood distribution $p(\mathbf{d}|\mathbf{r})$. The main objective has been to estimate the model's parameters, in the acquisition likelihood model in particular . Necessary theory and methods, in order to perform these tasks, have been introduced.

In two thorough test studies and an empirical seismic example, we have performed parameter estimation in the acquisition likelihood model and approximated posterior models of order k based on the respective estimates. The results generally stabilize for orders $k = 2, 3$. Estimation for different acquisition likelihood parameter sets, and the estimations sensitivity to the prior model has been studied. The ratio between the observation mean and variance is termed the signal-to-noise ratio (S/N). The results from the test studies indicate that this ratio could be a reliable indication of the quality of the parameter estimates and the corresponding posterior model approximations. For test cases with increasing S/N-ratios, the results were in general more reliable, with posterior model approximations recognizing the true LF-class well. The estimated parameters need however not be very close to the true parameters, even for a model with large S/N-ratio, as the simplified kth order posterior approximations utilize the parameters in different ways for different orders.

In the second test study and in the seismic example, we chose to approximate the posterior models based on the parameter estimate for order $k = 3$. This procedure is less computer demanding and was proposed by the first test study as the estimates were stable. The resulting posterior model approximation were generally quite reliable. Due to the fact that the approximations utilize the parameters differently, this solution might however not be optimal, and should perhaps have been studied in more detail.

The noise parameter is consistently under-estimated for higher order computations when compared to the true value. This property clearly affected the following posterior model approximation in the seismic example. By the results in the test studies, this under-estimation seems not to occur due to use of few data points, i.e. by over-fitting the data, nor due to MMLE bias. Most likely it originates from the fact that every data point is used multiple times in the approximations. We do not explore this property in this thesis, but it should be

studied in more detail.

What should be done in further work is thus mainly a larger parameter estimation study for the model in the seismic example, with respect to the noise parameter in particular. Study on why the noise is under-estimated, and possible solutions by e.g. noise correction or similarly, should be performed. Models including colored noise should also be explored, as the studies in this thesis consider white noise only. Both the observation data and the response variables might be extended into higher dimensions for this seismic model. Thus, data from multiple angles and three-dimensional response variables by the three elastic material properties P-wave velocity, S-wave velocity and density should be considered, as described in the thesis' introduction. Parameter estimation in the prior model and the response likelihood model should be studied. Other parameter optimization methods could be examined, as the optimization method used in this thesis might require too much cpu-time when estimating many parameters.

Bibliography

[1] Aki, K. and Richards, P.G. (1980). *Quantitative seismology: Theory and methods*, W.H. Freeman and Co.

[2] Buland, A. (2002). *Bayesian seismic AVO inversion*, Ph.d. thesis, Chap.1

[3] Buland, A. and Omre, H. (2003). Bayesian linearized AVO inversion, *Geophysics*, **68**: No.1

[4] Buland, A. and Omre, H. (2003). Joint AVO inversion, wavelet estimation and noise-level estimation using a spatially coupled hierarchical Bayesian model, *Geophysical Prospecting* **51**: 531-550

[5] Casella, G. and Berger, R.L. (2002). *Statistical Inference*, 2nd edt., Duxbury, Chapter 7

[6] Cormen, T.H., Leiserson, C.E., Rivest, R.L. and Stein, C. (2005). *Introduction to Algorithms*, 2nd edt., McGraw-Hill Book Company

[7] Dominici, F., Parmigiani, G. and Clyde, M. (2000). Conjugate analysis of multivariate normal data with incomplete observations, *The Canadian Journal of Statistics* **28**: No.3, pages:531-550

[8] Eidsvik, J., Mukerji, T. and Switzer, P. (2003) Estimation of Geological Attributes From a Well Log: An Application of Hidden Markov Chains, *Mathematical Geology*, **36**: No.3

[9] Fink, D. (1997). A compendium of conjugate priors. Available from `http://www.stat.columbia.edu/~cook/movabletype/mlm/CONJINTRnew%2BTEX.pdf`, last checked 26.06.2010

[10] Forney Jr., G.D. (1973). The Viterbi Algorithm, *Proceedings of the IEEE*, **61**: No.3

[11] Johnson, R.A and Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis*, 6th edt., Pearson

[12] Larsen, A.L., Ulvmoen, M., Omre, H. and Buland, A. (2006). Bayesian lithology/fluid prediction and simulation on the basis of a Markov-chain prior model, *Geophysics*, **71**: No.5

[13] Minka, T.P. (2003). Estimating a Dirichlet distribution. Available from `http://research.microsoft.com/en-us/um/people/minka/papers/dirichlet/minka-dirichlet.pdf`, last checked 26.06.2010

[14] Omre, H. and Ulvmoen, M. (2009). Bayesian Categorical Deconvolution, Ph.d. thesis

[15] Ryan, H. (1994). Ricker, Ormsby, Klauder, Butterworth - A Choice of Wavelets, CSEG Recorder

[16] Rigzone homepages, How Does Marine Seismic Work?, `http://www.rigzone.com/training/insight.asp?i_id=303`, last checked 26.06.2010

[17] Ross, S.M. (2007). *Introduction to Probability Models*, 9th edt., Elsevier Inc.

[18] Rottmann, K. (2004). *Matematisk Formelsamling*, 8th edt., Spektrum forlag

[19] Scott, S.L. (2002). Bayesian Methods for Hidden Markov Models: Recursive computing in the 21st Century, *Journal of the American Statistical Association* **97**: 337-351

[20] Tjelmeland, H. and Kvaløy, J.T. (2000). *Tabeller og formler i statistikk*, Tapir trykk

[21] Ulvmoen, M. and Hammer, H. (2009). Bayesian lithology-fluid inversion - algorithm efficiency, Ph.d. thesis, to be published in *Computational Geosciences*

[22] Walpole, R.E., Myers, R.H., Myers, S.L. and Ye, K. (2007). *Probability & Statistics for Engineers & Scientists*, 8th edt., Pearson

Appendix A

Relevant probability distributions

In this appendix, we introduce the probability distributions utilized in the thesis. The given distributions and their properties are in accordance with References [7], [9], [11] and [20]. Concerning the term conjugate distribution, a parameter prior distribution $p(\theta)$ is conjugate to the likelihood $p(x|\theta)$ if the posterior $p(\theta|x)$ is in the same distribution family as the prior.

A.1 Gaussian distribution

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a Gaussian distributed random vector defined on $\mathbf{x} \in \mathbb{R}^n$. The Gaussian probability density function with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is defined as

$$\phi_n(\mathbf{x}) = \phi_n(x_1, \ldots, x_n; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \quad .$$

We say that $\mathbf{X}$ is Gaussian distributed by $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Here $\boldsymbol{\mu}$ is the $(n \times 1)$ mean vector and $\boldsymbol{\Sigma}$ is the $(n \times n)$ covariance matrix of $\mathbf{x}$,

$$\mathrm{E}[\mathbf{X}] = \boldsymbol{\mu} \quad , \quad \mathrm{Var}[\mathbf{X}] = \boldsymbol{\Sigma} \quad .$$

For the one-dimensional Gaussian distribution, $N_1(\mu, \sigma^2)$, the moment generating function is defined as

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx}\phi_1(x)dx = e^{\mu t + \sigma^2 t^2/2} \quad .$$

A.1.1 Properties

Let $\mathbf{X}$ be a Gaussian distributed random vector with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ where

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \quad , \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad , \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \quad .$$

Here $\mathbf{X}_1 = (X_1, \ldots, X_{n_1})$ and $\mathbf{X}_2 = (X_{n_1+1}, \ldots, X_{n_1+n_2})$ with $n = n_1 + n_2$. Then the conditional distribution for $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$ is also Gaussian, $N_{n_1}(\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2})$, with expectation and variance

$$\begin{aligned} \boldsymbol{\mu}_{1|2} &= \mathrm{E}[\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2] = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ \boldsymbol{\Sigma}_{1|2} &= \mathrm{Var}[\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2] = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \quad . \end{aligned}$$

Linear combinations of Gaussian variables are also Gaussian distributed, i.e.

$$\mathbf{Ax} + \mathbf{b} \sim N_n(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}') \ .$$

Here $\mathbf{A}$ is a $(n \times n)$ matrix of constants and $\mathbf{b}$ is a $(n \times 1)$ vector of constants.

For the one-dimensional Gaussian distribution, $N_1(\mu, \sigma^2)$, the conditional conjugate parameter priors are $[\mu|\sigma^2] \sim N(\gamma_*, \lambda_*)$ and $[\sigma^2|\mu] \sim IG(\alpha_*, \beta_*)$. Here $IG(\cdot)$ indicates the Inverse Gamma distribution defined in Appendix A.3. We can assign the conditional conjugate prior distribution $[\mu|\sigma^2] \sim N(\gamma, \sigma^2/\lambda)$ to the mean parameter and the conjugate prior distribution $\sigma^2 \sim IG(\alpha, \beta)$ to the variance parameter . The joint conjugate prior is the normal-scaled inverse gamma distribution, $(\mu, \sigma^2) \sim p(\mu|\sigma^2)p(\sigma^2) = NsIG(\gamma, \lambda, \alpha, \beta)$, with pdf defined as

$$f(\mu, \sigma^2; \gamma, \lambda, \alpha, \beta) = \frac{\sqrt{\lambda}}{\sqrt{2\pi}\sigma} \frac{1}{\beta^\alpha \Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp\left(-\frac{2 + \beta\lambda(\mu - \gamma)^2}{2\beta\sigma^2}\right) \ .$$

A.2 Dirichlet distribution

Let $\mathbf{P} = (P_1, \ldots, P_n)$ be a Dirichlet distributed random vector defined for $\{0 < p_i < 1;\ i = 1, \ldots, n\}$ and $\sum_{i=1}^n p_i = 1$. The Dirichlet probability density function with scale parameter vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$, where $\alpha_i > 0 \ \forall \ i$, is defined as

$$f(\mathbf{p}) = f(p_1, \ldots, p_{n-1}; \alpha_1, \ldots, \alpha_n) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^n p_i^{\alpha_i - 1} \ . \tag{A.1}$$

Here $B(\boldsymbol{\alpha})$ ensures normality and is the multinomial beta function expressed by the gamma function $\Gamma(\cdot)$

$$B(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^n \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^n \alpha_i)} \ .$$

We say that $\mathbf{X}$ is Dirichlet distributed by $D_n(\boldsymbol{\alpha})$. We notice that only $(n-1)$ of the p_i's are found by the probability distribution in Expression (A.1) as $\sum_{i=1}^n p_i = 1$, then $p_n = 1 - \sum_{i=1}^{n-1} p_i$. The mean and variance of each element in $\mathbf{P}$ can be found as

$$\mathrm{E}[P_i] = \frac{\alpha_i}{\sum_{j=1}^n \alpha_j} \quad , \quad \mathrm{Var}[P_i] = \frac{\mathrm{E}[P_i](1 - \mathrm{E}[P_i])}{1 + \sum_{j=1}^n \alpha_j} \ .$$

The Dirichlet distribution is the parameter conjugate prior distribution to the multinomial distribution.

A.3 Inverse Gamma distribution

Let $X > 0$ be an inverse Gamma distributed random variable defined for $x > 0$. The Inverse Gamma probability density function with parameters $\alpha > 0$ and $\beta > 0$ is defined as

$$f(x) = f(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \frac{e^{-1/(\beta x)}}{x^{\alpha+1}} \ .$$

The mean and variance can be found as

$$\mathrm{E}[X] = \frac{1}{\beta(\alpha - 1)} \text{ for } \alpha > 1 \quad , \quad \mathrm{Var}[X] = \frac{1}{\beta^2(\alpha-1)^2(\alpha-2)} \text{ for } \alpha > 2 \ .$$

We say that X is Inverse Gamma distributed by $IG(\alpha, \beta)$. If X is Gamma distributed, then $1/X$ will be Inverse Gamma distributed, thereby the name.

Appendix B

Explicit computations

In this appendix, we first derive the computation of the approximate Gaussian distributions in Expressions (3.16) and (3.17) for the models defined in Chapters 5 and 6. These two approximate distributions are thus computed with the assumption of equal marginal response likelihood distribution variance, i.e. $\sigma_{x_t}^2 = \sigma_x^2$ in Expression (2.7). The computations are similar without this assumption. Next, the likelihood integral $l_d^{(k)}(x_t^{(k)})$ in Expression (3.22) is computed.

B.1 Computing $p_*(\mathbf{r})$ and $p_*(\mathbf{r}|\mathbf{d})$

We want to compute the mean and variance for the Gaussian distributions, $p_*(\mathbf{r}|\mathbf{d})$ and $p_*(\mathbf{r})$ in Expressions (3.16) and (3.17) respectively, which defines the alternative acquisition likelihood distribution in Expression (3.15). This is done in order to use the P-P ratio deconvolution algorithm defined in Sections 3.2 and 3.3.

First we assume that every marginal prior pdf is the stationary distribution of $\mathbf{P}_x$ denoted $\mathbf{p}_{stas} = (p_1, \ldots, p_L)$, i.e. $p(x_t) = \mathbf{p}_{stas}$ for all times t. Then the marginal response pdf will be Gauss-linear by Expression (2.7),

$$p(r_t) = \sum_{x_t} p(r_t|x_t)p(x_t) = \sum_{i=1}^{L} p_i \phi_1(r_t; \mu_i, \sigma_x^2) \ .$$

We will thus have the Gaussian marginal response moment generating function

$$M_{R_t}(s) = \int_{-\infty}^{\infty} e^{sr} p(r_t) dr_t = e^{\frac{\sigma_x^2 s^2}{2}} \sum_{i=1}^{L} p_i e^{\mu_i s} \ .$$

The two first moments are

$$\mathrm{E}(R_t) = \frac{d}{ds} M_{R_t}(s)\Big|_{s=0} = \sum_{i=1}^{L} \mu_i p_i = \mathrm{E}(R) \tag{B.1}$$

$$\mathrm{E}(R_t^2) = \frac{d^2}{ds^2} M_{R_t}(s)\Big|_{s=0} = \sigma_x^2 + \sum_{i=1}^{L} \mu_i^2 p_i = \mathrm{E}(R^2) \ . \tag{B.2}$$

We can then compute the marginal mean and variance by these moments,

$$\mu_r = \mathrm{E}(R) \tag{B.3}$$

$$\sigma_r^2 = \mathrm{E}(R^2) - \mathrm{E}(R)^2 \ . \tag{B.4}$$

The moments in Expressions (B.1) and (B.2) are independent of r_t, and so are then the mean and variance, thus $p(r_t) = N_1(\mu_r, \sigma_r^2)$ for all times t.

The spatial covariance between two marginal response variables over a positive natural numbered timestep, Δ_t, is

$$\mathrm{Cov}(R_t, R_{t+\Delta_t}) = \mathrm{E}(R_t \cdot R_{t+\Delta_t}) - \mathrm{E}(R_t)\mathrm{E}(R_{t+\Delta_t}) = \mathrm{E}(R_t \cdot R_{t+\Delta_t}) - \mu_r^2 \ . \tag{B.5}$$

Here

$$\mathrm{E}(R_t \cdot R_{t+\Delta_t}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} r_t r_{t+\Delta_t} p(r_t, r_{t+\Delta_t}) dr_t dr_{t+\Delta_t} \tag{B.6}$$

by the definition of the expectation function. The joint response marginal is

$$\begin{aligned} p(r_t, r_{t+\Delta_t}) &= \sum_{x_t}\sum_{x_{t+\Delta_t}} p(x_t, x_{t+\Delta_t}) p(r_t, r_{t+\Delta_t}|x_t, x_{t+\Delta_t}) \\ &= \sum_{x_t}\sum_{x_{t+\Delta_t}} p^{\Delta_t}(x_{t+\Delta_t}|x_t) p(x_t) p(r_t|x_t) p(r_{t+\Delta_t}|x_{t+\Delta_t}) \ , \end{aligned} \tag{B.7}$$

where we have used the conditional response independence property in Expression (2.8). In Expression (B.7) $p^{\Delta_t}(x_{t+\Delta_t}|x_t)$ defines the Δ_t-transition probability matrix, $\mathbf{P}^{\Delta_t} = (\mathbf{P}_x)^{\Delta_t}$. The expectation in Expression (B.6) is then

$$\begin{aligned} \mathrm{E}(R_t \cdot R_{t+\Delta_t}) &= \sum_{x_t}\sum_{x_{t+\Delta_t}} p^{\Delta_t}(x_{t+\Delta_t}|x_t) p(x_t) \int_{-\infty}^{\infty} r_t p(r_t|x_t) dr_t \\ &\quad \cdot \int_{-\infty}^{\infty} r_{t+\Delta_t} p(r_{t+\Delta_t}|x_{t+\Delta_t}) dr_{t+\Delta_t} \\ &= \sum_{x_t}\sum_{x_{t+\Delta_t}} p^{\Delta_t}(x_{t+\Delta_t}|x_t) p(x_t) \mu_{x_t} \mu_{x_{t+\Delta_t}} \ . \end{aligned}$$

Here we have used the definition of the expectation function, for $p(r_t|x_t) = N_1(\mu_{x_t}, \sigma_x^2)$ we have $\mu_{x_t} = \mathrm{E}(r_t|x_t) = \int_{-\infty}^{\infty} r_t p(r_t|x_t) dr_t$. The covariance in Expression (B.5) thus becomes

$$\rho_{\Delta_t} = \mathrm{Cov}(R_t, R_{t+\Delta_t}) = \sum_{x_t}\sum_{x_{t+\Delta_t}} p^{\Delta_t}(x_{t+\Delta_t}|x_t) p(x_t) \mu_{x_t} \mu_{x_{t+\Delta_t}} - \mu_r^2 \ . \tag{B.8}$$

With μ_r given by Expression (B.3) we can compute the covariance in Expression (B.8) for all timesteps Δ_t. By Expressions (B.3), (B.4) and (B.8) we can now compute $p_*(\mathbf{r}) = N_T(\boldsymbol{\mu}_\mathbf{r}, \boldsymbol{\Sigma}_\mathbf{r})$ with

$$\boldsymbol{\mu}_\mathbf{r} = \begin{pmatrix} \mu_r \\ \vdots \\ \mu_r \end{pmatrix} \quad , \quad \boldsymbol{\Sigma}_\mathbf{r} = \begin{pmatrix} \sigma_r^2 & \rho_1 & \cdots & \rho_{199} \\ \rho_1 & \sigma_r^2 & & \vdots \\ \vdots & & \ddots & \rho_1 \\ \rho_{199} & \cdots & \rho_1 & \sigma_r^2 \end{pmatrix} \ . \tag{B.9}$$

Next we assume that the marginal likelihood has a Gaussian pdf, $p_*(\mathbf{d}) = N_T(\boldsymbol{\mu}_\mathbf{d}, \boldsymbol{\Sigma}_\mathbf{d})$ in order to find $p_*(\mathbf{r}|\mathbf{d})$. With this assumption we can write

$$\begin{pmatrix} \mathbf{r} \\ \mathbf{d} \end{pmatrix} \sim N_{2T}\left(\begin{bmatrix} \boldsymbol{\mu}_\mathbf{r} \\ \boldsymbol{\mu}_\mathbf{d} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_\mathbf{r} & \boldsymbol{\Sigma}_\mathbf{rd} \\ \boldsymbol{\Sigma}_\mathbf{dr} & \boldsymbol{\Sigma}_\mathbf{d} \end{bmatrix} \right) \ ,$$

where $\boldsymbol{\Sigma}_\mathbf{rd} = \boldsymbol{\Sigma}'_\mathbf{dr}$. If we can find $\boldsymbol{\mu}_\mathbf{d}$, $\boldsymbol{\Sigma}_\mathbf{d}$ and $\boldsymbol{\Sigma}_\mathbf{rd}$ we can compute the mean and variance in the distribution in Expression (3.16) by

$$p_*(\mathbf{r}|\mathbf{d}) = N_T(\boldsymbol{\mu}_{\mathbf{r}|\mathbf{d}}, \boldsymbol{\Sigma}_{\mathbf{r}|\mathbf{d}}) = N_T\left(\boldsymbol{\mu}_\mathbf{r} + \boldsymbol{\Sigma}_\mathbf{rd}\boldsymbol{\Sigma}_\mathbf{d}^{-1}(\mathbf{d} - \boldsymbol{\mu}_\mathbf{d}), \boldsymbol{\Sigma}_\mathbf{r} - \boldsymbol{\Sigma}_\mathbf{rd}\boldsymbol{\Sigma}_\mathbf{d}^{-1}\boldsymbol{\Sigma}_\mathbf{dr}\right) \ . \tag{B.10}$$

With the acquisition likelihood pdf in Expression (2.10), assuming an error term with colored noise, the approximate marginal likelihood distribution mean and variance is

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{d}} &= \mathrm{E}(\mathbf{d}) = \mathrm{E}[\mathrm{E}(\mathbf{d}|\mathbf{r})] = \mathrm{E}(\mathbf{W}\mathbf{r}) = \mathbf{W}\boldsymbol{\mu}_{\mathbf{r}} && \text{(B.11)} \\ \boldsymbol{\Sigma}_{\mathbf{d}} &= \mathrm{Var}(\mathbf{d}) = \mathrm{E}[\mathrm{Var}(\mathbf{d}|\mathbf{r})] + \mathrm{Var}[\mathrm{E}((\mathbf{d}|\mathbf{r}))] \\ &= \mathrm{E}(\sigma_{e_1}^2 \mathbf{W}\mathbf{W}' + \sigma_{e_2}^2 \mathbf{I}) + \mathrm{Var}(\mathbf{W}\mathbf{r}) = \mathbf{W}\boldsymbol{\Sigma}_{\mathbf{r}}\mathbf{W}' + \sigma_{e_1}^2 \mathbf{W}\mathbf{W}' + \sigma_{e_2}^2 \mathbf{I} \ . && \text{(B.12)} \end{aligned}$$

By the properties of the Gaussian distribution and the acquisition likelihood distribution variance in Expression (2.10) we have

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{d}|\mathbf{r}} &= \boldsymbol{\Sigma}_{\mathbf{d}} - \boldsymbol{\Sigma}_{\mathbf{dr}}\boldsymbol{\Sigma}_{\mathbf{r}}^{-1}\boldsymbol{\Sigma}_{\mathbf{rd}} \\ &= \mathbf{W}\boldsymbol{\Sigma}_{\mathbf{r}}\mathbf{W}' + \sigma_{e_1}^2 \mathbf{W}\mathbf{W}' + \sigma_{e_2}^2 \mathbf{I} - \boldsymbol{\Sigma}_{\mathbf{dr}}\boldsymbol{\Sigma}_{\mathbf{r}}^{-1}\boldsymbol{\Sigma}_{\mathbf{rd}} \\ &= \sigma_{e_1}^2 \mathbf{W}\mathbf{W}' + \sigma_{e_2}^2 \mathbf{I} \ , \end{aligned}$$

thus $\boldsymbol{\Sigma}_{\mathbf{dr}}\boldsymbol{\Sigma}_{\mathbf{r}}^{-1}\boldsymbol{\Sigma}_{\mathbf{rd}} = \mathbf{W}\boldsymbol{\Sigma}_{\mathbf{r}}\mathbf{W}'$. We see that $\boldsymbol{\Sigma}_{\mathbf{dr}} = \mathbf{W}\boldsymbol{\Sigma}_{\mathbf{r}}$ is a possible solution, as $\boldsymbol{\Sigma}_{\mathbf{r}}$ is symmetric. With this solution, and with the marginal likelihood distribution parameters by Expressions (B.11) and (B.12), we are able to compute the pdf in Expression (B.10). Its mean and variance is

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{r}|\mathbf{d}} &= \boldsymbol{\mu}_{\mathbf{r}} + (\mathbf{W}\boldsymbol{\Sigma}_{\mathbf{r}})' \left(\mathbf{W}\boldsymbol{\Sigma}_{\mathbf{r}}\mathbf{W}' + \sigma_{e_1}^2 \mathbf{W}\mathbf{W}' + \sigma_{e_2}^2 \mathbf{I}\right)^{-1} (\mathbf{d} - \mathbf{W}\boldsymbol{\mu}_{\mathbf{r}}) && \text{(B.13)} \\ \boldsymbol{\Sigma}_{\mathbf{r}|\mathbf{d}} &= \boldsymbol{\Sigma}_{\mathbf{r}} - (\mathbf{W}\boldsymbol{\Sigma}_{\mathbf{r}})' \left(\mathbf{W}\boldsymbol{\Sigma}_{\mathbf{r}}\mathbf{W}' + \sigma_{e_1}^2 \mathbf{W}\mathbf{W}' + \sigma_{e_2}^2 \mathbf{I}\right)^{-1} (\mathbf{W}\boldsymbol{\Sigma}_{\mathbf{r}}) \ . && \text{(B.14)} \end{aligned}$$

With the Gaussian pdfs, $p_*(\mathbf{r})$ defined by Expression (B.9) and $p_*(\mathbf{r}|\mathbf{d})$ defined by Expressions (B.13) and (B.14), we have an acquisition likelihood model according to Expression (3.15). We can then use the P-P ratio algorithm to compute the approximate posterior distribution.

B.2 Approximating $l_d^{(k)}(x_t^{(k)})$

We want to approximate the integral in Expression (3.22),

$$l_d^{(k)}\left(x_t^{(k)}\right) = \int \frac{p_*\left(r_t^{(k)}\middle|\,\mathbf{d}\right)}{p_*\left(r_t^{(k)}\right)} p\left(r_t^{(k)}\middle|\,x_t^{(k)}\right) dr_t^{(k)} \ , \qquad \text{(B.15)}$$

for every $r_t^{(k)} = (r_{t-k+1}, \ldots, r_t); t = k, \ldots, T$. Here $p_*(\mathbf{r}|\mathbf{d}) = N_T(\boldsymbol{\mu}_{\mathbf{r}|\mathbf{d}}, \boldsymbol{\Sigma}_{\mathbf{r}|\mathbf{d}})$, $p_*(\mathbf{r}) = N_T(\boldsymbol{\mu}_{\mathbf{r}}, \boldsymbol{\Sigma}_{\mathbf{r}})$ and $p(\mathbf{r}|\mathbf{x}) = N_T(\boldsymbol{\mu}_{\mathbf{r}|\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{r}|\mathbf{x}})$ with

$$[a_t^{(k)}] \sim N_k\left([\boldsymbol{\mu}_{\mathbf{a}}]_{t-k+1}^{t}, [\boldsymbol{\Sigma}_{\mathbf{a}}]_{t-k+1}^{t}\right) \qquad \text{(B.16)}$$

for $a \in \{r, [r|\mathbf{d}], [r|x]\}$. The notation $[\cdot]_i^j$ in Expression (B.16) indicates a matrix of rows and columns i to j from the matrix in the brackets. We can approximate the integral in Expression (B.15) by the following algorithm.

Algorithm 5: Estimating the $l_d^{(k)}(x_t^{(k)})$ integral

- For $i = 1, \ldots, n_s$:
  - Draw $[r_t^{(k)}]^{(s)} \sim p(r_t^{(k)}|x_t^{(k)})$
  - $m_i = \frac{\phi_*([r_t^{(k)}]^{(s)}|\mathbf{d})}{\phi_*([r_t^{(k)}]^{(s)})}$

- $\hat{l}_d^{(k)}(x_t^{(k)}) = \frac{1}{n_s}\sum_{i=1}^{n_s} m_i$

We have to estimate this integral for all $\{x_t^{(k)} = (x_{t-k+1}, \ldots, x_t);\ t = k, \ldots, T\}$ for every possible state, $x_t \in \Omega_x : \{1, \ldots, L\}$, to be able to implement the forward-backward algorithm in Algorithm 1. We thus have to run Algorithm 5 $(T - k + 1)L^{(k)}$ times before running Algorithm 1, which becomes computer demanding for large k. A challenge is thus to find a reasonable number of simulations, n_s.

Appendix C

Additional figures

In this appendix we display the complete reference profiles of length $T = 1000$ for the base case in Chapters 5 and 6. The base case profiles displayed in these two chapters are the 200 first elements of the complete profiles.

C.1 Full profiles from Chapters 5 and 6

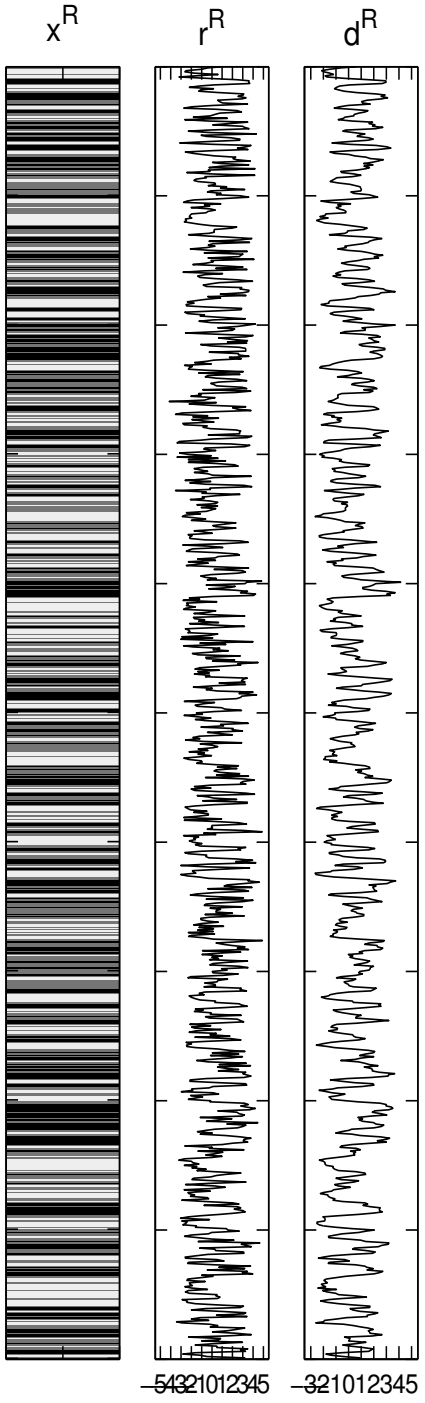


Figure C.1: *The full reference profile,* $\mathbf{x}^R$*, response profile,* $\mathbf{r}^R$*, and observation profile,* $\mathbf{d}_{MCMN}$*, in the base case in Chapter 5 with length* $T = 1000$*.*