Blind Categorical Deconvolution in Two-Level Hidden Markov Models

David V. Lindberg and Henning Omre

***Abstract* - A convolved two-level hidden Markov model is defined as an observed top level representing convolutions of an unobserved middle level of responses to an unobserved bottom level containing a Markov chain of categorical classes. The associated model parameters include a Markov chain transition matrix, response levels and variances, a convolutional kernel and an observation error variance. The convolutional kernel and the error variance are defined to be unknown. Focus is on joint assessment of the unknown model parameters and the sequence of categorical classes given the observed top level. This is termed blind categorical deconvolution, and is cast in a Bayesian inversion setting. An approximate posterior model based on an approximate likelihood model on factorisable form is defined. The approximate model, including the likelihoods for the unknown model parameters, can be exactly assessed by a recursive algorithm. A sequence of approximations is defined such that trade-offs between accuracy and computational demands can be made. The model parameters are assessed by approximate maximum likelihood estimation whereas the inversion is represented by the approximate posterior model. A limited empirical study demonstrates that reliable model parameter assessments and inversions can be made from the approximate model. An example of blind seismic deconvolution is also presented and discussed.**

***Index terms* - Blind deconvolution, hidden Markov models, Bayesian inversion, maximum approximate likelihood estimation, forward-backward algorithm.**

I. Introduction

Many recorded signals appear as a convolution of responses from an underlying profile with observation error. The objective is to assess the underlying profile given the convolved signal. If the convolution kernel and error models are known, this assessment is termed signal deconvolution. If the assessment must be made without knowledge of the kernel and error models, it is termed blind deconvolution. We study blind deconvolution into a categorical

The work is part of the Uncertainty in reservoir Evaluation (URE) activity - consortium at Department of Mathematical Sciences, NTNU, Trondheim, Norway.

The authors are with the Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim 7491, Norway (e-mail: davidlin@math.ntnu.no; omre@math.ntnu.no).

underlying profile.

The motivation for our study is seismic inversion in exploration and appraisal of hydrocarbon reservoirs. The recorded seismic signal $\mathbf{d}$ in a discretized vertical profile is gathered as a linear convolution of a wavelet, $\mathbf{w}$, and the earths reflection coefficients $\mathbf{r}$, with an additive error component $\mathbf{e}$. Seismic deconvolution usually refers to removing the effects of the wavelet and the error in order to restore the original reflectivity response sequence $\mathbf{r}$. When the response, wavelet and error parameters are all unknown, we have a blind deconvolution problem. If $\mathbf{r}$ is considered to have some large deviations caused by transitions between two different layers, and frequent smaller variations caused by heterogeneities inside each layer, the conventional model for the response is the Bernoulli-Gaussian (BG) model, see [1], [2]. The standard discrete representation when doing blind seismic deconvolution for this model is given as

$$d_t = \sum_{i=0}^{2a} w_i r_{t-i} + e_t \; ; \quad t = 1, \ldots, T \tag{1}$$

in which $2a + 1$ and T is the discretized length of the wavelet and the seismic trace respectively. Notice that this representation assumes a one-sided wavelet. In Kormylo and Mendel [1], the convolution is modeled by an auto-regressive moving average model and the blind deconvolution problem is solved by maximum likelihood (ML) estimation in which the maximization is made by a numerical block-component method. The ML estimation problem is solved by a stochastic EM (SEM) algorithm [3] in Lavielle [4] that requires Markov chain Monte Carlo (McMC) Gibbs sampling of the latent response field. A Bayesian estimation model assessed by McMC Gibbs sampling is proposed in Cheng et al. [2] for which the wavelet is assigned a Gaussian prior model. The ML estimation model assessed by SEM is compared to the Bayesian estimation method by McMC Gibbs sampling in Rosec et al. [5] and the Bayesian choice is found to be slightly preferable. In Baziw and Ulrych [6], a principle phase decomposition technique is introduced for which overlapping wavelets are estimated through

a Rao-Blackwellized particle filter [7].

We focus here on inversion directly into the underlying lithology/fluid sequence, hence categorical deconvolution. A standard hidden Markov model (HMM) contains two levels; known continuous variables as the top level and hidden unknown categorical variables according to a Markov chain as bottom level (see MacDonald and Zucchini [8] and the references therein). The forward-backward (FB) [9] algorithm is a recursive algorithm that enable us to assess the hidden categorical variables given the observations. We consider the blind deconvolution problem for a convolutional two-level HMM, which is a standard HMM with an additional convolved top layer. For the seismic deconvolution problem in particular, the convolved top level $\mathbf{d}$ represents the observed seismic signal, the middle level $\mathbf{r}$ contains the unknown continuous elastic material properties corresponding to the response whereas the categorical bottom level $\boldsymbol{\pi}$ represents an unobserved sequence of lithology/fluid layers in the subsurface, see Larsen at al. [10]. The discrete representation is now given as

$$d_t = \sum_{i=-a}^{a} w_i r_{t+i} + e_t \quad , \quad r_t = \mu_{\pi_t} + e_{\pi_t} \; ; \quad t = 1, \ldots, T \tag{2}$$

in which $\mathbf{w}$ is the convolution kernel, e_t is additive observation error, μ_{π_t} is the response value given the class of state π_t and e_{π_t} is a class dependent additive response error component. Focus is hence on assessment of the full categorical profile $\boldsymbol{\pi}$. We consider the case where the categorical profile, response profile, convolution kernel and observation error parameters are all unknown, hence the blind categorical deconvolution problem. Notice that in our model the convolution kernel is two-sided and hence according to the general convolution definition, see [11].

Consider the fictive reservoir in Fig.1 for which one exploration well is located in lateral position A and assume we are interested in exploring the subsurface at position B for which only seismic data have been collected. The lithology/fluid characteristics like proportions, ordering rules and material properties are assumed to be stationary throughout the reservoir

and can be assessed from core and log data in well A. The seismic data acquisition parameters may vary significantly in the lateral direction however [5] , due to lateral variations in the overburden. The seismic wavelet and error model parameters should therefore be reestimated for the seismic data in well B. Hence seismic inversion into the lithology/fluid layering in well B constitutes a blind categorical deconvolution problem.

There are several fundamental differences between the traditional BG model and the convolutional two-level HMM model, see the graphical comparison in Fig.2. The BG model can only represent a sequence of two layer types and no thickness information on the layers can be enforced a priori. The convolutional two-level HMM model can represent a sequence of many different layer types and information about transitions between these types and their thicknesses can be enforced a priori. Hence vertical dependencies between the layers is omitted by the BG model, whereas it is assessed by the convolutional two-level HMM model through the Markov chain assumption, see Fig.2. Moreover, when comparing (1) and (2) notice that the BG model only captures one-sided convolutions, whereas the current convolutional two-level HMM model can capture two-sided convolutions, see also Fig.2. Actually, the BG models appear as a small subset of the convolutional two-level HMM models studied here.

The convolutional two-level HMM is first introduced in Larsen at al. [10] who developed an approximate posterior model for the hidden categorical profile given the convolved signal which can be exactly and efficiently assessed by the FB algorithm. The approximate model is generalized in Rimstad and Omre [12]. The blind categorical deconvolution problem is more complicated however, since model parameter estimation usually involves a computer demanding and possibly unstable optimization in addition to the inverse problem. We solve the blind categorical deconvolution problem by assessing the parameters of interest through approximate maximum marginal likelihood estimation applying the approximate posterior model. The model parameter estimates and corresponding prediction of the categorical se-

quence are compared for varying approximation orders to provide a solution with a suitable trade-off between approximation error and computer efficiency.

II. Model Description

Assume that we have a stochastic field on $\mathcal{D} \in \mathbb{R}^1$ discretized on the lattice $\mathcal{L}_\mathcal{D} = \{1, \ldots, T\}$ for which we have a continuous observed field $\mathbf{d} = \{d_t; t \in \mathcal{L}_\mathcal{D}\}$, $d_t \in \mathbb{R}$, see Fig.2. These observations are recorded as a convolution of a continuous latent response field $\mathbf{r} = \{r_t; t \in \mathcal{L}_\mathcal{D}\}$, $r_t \in \mathbb{R}$ which in turn depend on a latent categorical field $\boldsymbol{\pi} = \{\pi_t; t \in \mathcal{L}_\mathcal{D}\}$ that has values from a finite discrete state space $\pi_t \in \Omega_\pi : \{1, \ldots, L\}$, see (2) and Fig.2. Focus of the study is on assessing $\boldsymbol{\pi}$ given $\mathbf{d}$.

In a probabilistic Bayesian framework, the data acquisition model is specified as a likelihood model $p(\mathbf{d}|\boldsymbol{\pi}; \boldsymbol{\theta}_l)$ and we assign a prior probability density function (pdf) $p(\boldsymbol{\pi}; \boldsymbol{\theta}_p)$ to the categorical field under study, depending on some model parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_l, \boldsymbol{\theta}_p)$. The full likelihood model is actually $p(\mathbf{d}, \mathbf{r}|\boldsymbol{\pi}; \boldsymbol{\theta}_l)$ from which we obtain $p(\mathbf{d}|\boldsymbol{\pi}; \boldsymbol{\theta}_l)$ by integrating out $\mathbf{r}$. The objective is to explore the posterior model which is defined by Bayes' rule as

$$p(\boldsymbol{\pi}|\mathbf{d}; \boldsymbol{\theta}) = \frac{p(\mathbf{d}|\boldsymbol{\pi}; \boldsymbol{\theta}_l) \times p(\boldsymbol{\pi}; \boldsymbol{\theta}_p)}{p(\mathbf{d}; \boldsymbol{\theta})} = \frac{1}{p(\mathbf{d}; \boldsymbol{\theta})} \times \int p(\mathbf{d}, \mathbf{r}|\boldsymbol{\pi}; \boldsymbol{\theta}_l) d\mathbf{r} \times p(\boldsymbol{\pi}; \boldsymbol{\theta}_p) \tag{3}$$

with $p(\mathbf{d}; \boldsymbol{\theta})$ being a normalizing constant to ensure that the posterior model becomes a valid probability distribution. For convenience we omit the parameter dependency in the notation throughout the rest of this section.

We assume that the latent underlying categorical states follow a Markov chain as proposed in Krumbein and Dacey [13]. The latent response field appears as conditional independent and single-site dependent on the categorical field whereas the observations appear as a convolution of the response field. We denote our model a convolutional two-level HMM according to the notation in Rimstad and Omre [12]. A directed acyclic graph of the particular convolutional two-level HMM under consideration is shown in Fig.2(a). We also present the

corresponding conventional BG model in Fig.2(b); notice that the vertical dependency in the categorical layer is lacking in the BG model, and that the convolution is one-sided. Moreover, its categorical variables are required to be binary. In the rest of this section, the prior and likelihood model under the convolutional two-level HMM and the corresponding posterior model are described in more detail.

A. Prior Model

We assume that the spatial coupling in the categorical field $\boldsymbol{\pi}$ fulfills a first-order Markov property, $p(\pi_t|\pi_{t-1},\ldots,\pi_1) = p(\pi_t|\pi_{t-1})$. Hence the conditional probability of the state π_t at step t, given all the previous states, only depends on the single previous state π_{t-1}. The transition probabilities are independent on t and are represented by a transition probability $(L \times L)$ matrix $\mathbf{P} : \{p(\pi_t|\pi_{t-1}) : \ \pi_t, \pi_{t-1} \in \Omega_\pi \times \Omega_\pi\}$ which we assume has a stationary limiting distribution. Hence each element $\mathbf{P}_{ij}, \ i,j \in \Omega_\pi$ represents the probability of a transition from class i to class j. The prior model for the latent categorical field is then

$$p(\boldsymbol{\pi}) = \prod_{t=1}^{T} p(\pi_t|\pi_{t-1}) \tag{4}$$

where we set $p(\pi_1|\pi_0) = p(\pi_1)$ for notational ease. Here $p(\pi_1)$ is the initial marginal prior pdf which we define to be the stationary distribution of $\mathbf{P}$. Hence, the prior model is fully defined by the transition probability matrix, $\boldsymbol{\theta}_p = \{\mathbf{P}\}$. The first-order Markov property is graphically represented by directed arrows between the π_t-variables in the graph in Fig.2(a).

B. Likelihood Model

The likelihood model is given by

$$p(\mathbf{d}|\boldsymbol{\pi}) = \int p(\mathbf{d},\mathbf{r}|\boldsymbol{\pi})d\mathbf{r} = \int p(\mathbf{d}|\mathbf{r}) \times p(\mathbf{r}|\boldsymbol{\pi})d\mathbf{r}. \tag{5}$$

Here $p(\mathbf{r}|\boldsymbol{\pi})$ is termed the response likelihood model and $p(\mathbf{d}|\mathbf{r})$ is termed the observation likelihood model. Notice that we define our observation likelihood model such that $\mathbf{d}$ is conditionally independent of $\boldsymbol{\pi}$ given $\mathbf{r}$ as displayed in the graph in Fig.2(a).

The response likelihood model, $p(\mathbf{r}|\boldsymbol{\pi})$, represents physical response variables related to the states of $\boldsymbol{\pi}$. We assume that the marginal response likelihood models are independent Gaussian pdfs, i.e. that the elements in $\mathbf{r}$ are conditionally independent given $\boldsymbol{\pi}$. In particular, this is a single-site dependency as indicated in the graph in Fig.2(a). With these assumptions, $(\boldsymbol{\pi}, \mathbf{r})$ alone constitute a standard HMM [8]. The response likelihood model is then

$$p(\mathbf{r}|\boldsymbol{\pi}) = \prod_{t=1}^{T} p(r_t|\pi_t) \tag{6}$$

with $p(r_t|\pi_t)$ being Gaussian $N(\mu_{r|\pi_t}, \sigma^2_{r|\pi_t})$ for which the marginal mean and variance depend on the class of π_t. Notice that the marginal pdfs of the response appear as Gaussian mixture distributions, $p(r_t) = \sum_{\pi_t=1}^{L} p(r_t|\pi_t)p(\pi_t)$. The full response likelihood model is thus multivariate Gaussian, $p(\mathbf{r}|\boldsymbol{\pi}) = N(\boldsymbol{\mu}_{r|\pi}, \boldsymbol{\Sigma}_{r|\pi})$, where $\boldsymbol{\mu}_{r|\pi} = (\mu_{r|\pi_1}, \ldots, \mu_{r|\pi_T})'$ and $\boldsymbol{\Sigma}_{r|\pi}$ is a diagonal matrix with diagonal terms $(\sigma^2_{r|\pi_1}, \ldots, \sigma^2_{r|\pi_T})$. The response likelihood model is hence fully defined by the model parameters $\boldsymbol{\theta}_{rl} = \{(\mu_{r|\pi}, \sigma_{r|\pi});\ \pi \in \Omega_\pi\}$.

The convolution effect is captured in the observation likelihood model, for which the observation in an arbitrary interior lattice node has the discrete representation in (2). Here $\mathbf{w} = (w_{-a}, \ldots, w_a)$ are the weights of the convolution kernel which we assume to be stationary throughout the observed field and $e_t \sim N(0, \sigma_d^2)$ is additive independent errors which we assume to be Gaussian. Due to the convolution, the observations depend on both deeper and shallower values of the responses as can be seen in the graph in Fig.2(a). The observation likelihood model on matrix form is then

$$p(\mathbf{d}|\mathbf{r}) = N_T\left(\mathbf{W}\mathbf{r}, \sigma_d^2\mathbf{I}\right). \tag{7}$$

Here $\mathbf{W}$ is a band diagonal convolution matrix which rows correspond to the convolution

kernel $\mathbf{w}$ centered along its diagonal. Hence the observation likelihood is fully defined by the model parameters $\boldsymbol{\theta}_{ol} = \{\mathbf{w}, \sigma_d\}$. The observations at the edges of the field are subject to edge effects due to the convolution, see (7), hence some edge corrections must be enforced. We choose to account for these effects by adding pseudo response elements at the edges of $\mathbf{r}$.

With a Gaussian response and observation likelihood model, the full likelihood model in (5) can be shown to be Gaussian

$$p(\mathbf{d}|\boldsymbol{\pi}) = N_T\left(\mathbf{W}\boldsymbol{\mu}_{r|\pi}\ ,\ \mathbf{W}\boldsymbol{\Sigma}_{r|\pi}\mathbf{W}' + \sigma_d^2\mathbf{I}\right). \tag{8}$$

Notice that the variance of the Gaussian likelihood model can be regarded a sum of colored and white errors. The full likelihood model is hence uniquely defined by the model parameters $\boldsymbol{\theta}_l = \{\boldsymbol{\theta}_{rl}, \boldsymbol{\theta}_{ol}\}$. The dependencies defined by this likelihood model are represented by arrows in the graph in Fig.2(a).

C. Posterior Model

The full posterior model of the convolutional two-level HMM in (3) has a normalizing constant which is unfeasible to calculate in practice as it requires summation over all L^T combinations of the state space of $\boldsymbol{\pi}$. Neither are we able to simulate realizations from this posterior distribution, or to compute the maximum a posterior (MAP) prediction directly. The full posterior model for a standard HMM without convolution is however assessable through the FB algorithm. If the posterior model in (3) is on factorisable form it can be assessed by a generalization of the FB algorithm [14], which has higher cost than for the standard HMM. To make our posterior model factorisable, we approximate the likelihood model according to Rimstad and Omre [12]. Define the kth order states $\pi_t^{(k)} = (\pi_{t-k+1}, \ldots, \pi_t);\ t = k, \ldots, T$, which now has a state space of L^k elements. Similarly for the response variables $r_t^{(k)} = (r_{t-k+1}, \ldots, r_t);\ t = k, \ldots, T$ and the observations $d_t^{(k)} = (d_{t-k+1}, \ldots, d_t);\ t = k, \ldots, T$. We

then define kth order approximate observation likelihood marginals, see (7), by

$$\hat{p}^{(k)}\left(d_t^{(k)}\left|r_t^{(k)}\right.\right) = const \times \frac{p_*\left(r_t^{(k)}\middle|\mathbf{d}\right)}{p_*\left(r_t^{(k)}\right)}\;; \quad t = k,\ldots,T \tag{9}$$

in which we choose to define the pdfs $p_*(\mathbf{r}|\mathbf{d})$ and $p_*(\mathbf{r})$ as Gaussian approximations of $p(\mathbf{r}|\mathbf{d})$ and $p(\mathbf{r})$ since the state space of $(\mathbf{r},\mathbf{d})$ is continuous. This approximation is introduced in Larsen et al. [10] to account for the multimodality of the marginal pdf of the response variable $p(r_t)$. Both Gaussian approximations have analytically tractable parameters, see Buland and Omre [15], and the blocks that occur in (9) are their kth dimensional marginals. Further, in Rimstad and Omre [12] it is shown that this approximation performs better than the more intuitive choice of truncated marginals. The full kth order approximate likelihood marginals, see (8), are then

$$\hat{p}^{(k)}\left(d_t^{(k)}\middle|\pi_t^{(k)}\right) = \int \hat{p}^{(k)}\left(d_t^{(k)}\left|r_t^{(k)}\right.\right) \times p\left(r_t^{(k)}\middle|\pi_t^{(k)}\right) dr_t^{(k)}\;; \quad t = k,\ldots,T \tag{10}$$

in which $p\left(r_t^{(k)}\middle|\pi_t^{(k)}\right) = \prod_{i=t-k+1}^{t} p(r_i|\pi_i)$. Higher order approximations capture more of the convolution effect in the observation likelihood model. A kth order approximate posterior model on factorisable form can then be shown, see (3), to be

$$\hat{p}^{(k)}\left(\boldsymbol{\pi}|\mathbf{d}\right) = C_d^{(k)} \times q_{k-1}\left(\pi_{k-1}^{(k-1)}\middle|\mathbf{d}\right) \times \left[\prod_{t=k}^{T} \hat{p}^{(k)}\left(d_t^{(k)}\middle|\pi_t^{(k)}\right)^{1/k} \times p\left(\pi_t^{(k)}\middle|\pi_{t-1}^{(k)}\right)\right] \times q_T\left(\pi_T^{(k-1)}\middle|\mathbf{d}\right) \tag{11}$$

for $k = 1,\ldots,T$ with the edge factors defined as:

$$q_{k-1}\left(\pi_{k-1}^{(k-1)}\middle|\mathbf{d}\right) = \prod_{i=1}^{k-1} \hat{p}^{(i)}\left(d_i^{(i)}\middle|\pi_i^{(i)}\right)^{1/k} \times p\left(\pi_i^{(i)}\middle|\pi_{i-1}^{(i)}\right) \tag{12}$$

$$q_T\left(\pi_T^{(k-1)}\middle|\mathbf{d}\right) = \prod_{i=1}^{k-1} \hat{p}^{(i)}\left(d_T^{(i)}\middle|\pi_T^{(i)}\right)^{1/k} \tag{13}$$

for $k = 2,\ldots,T$ and finally $q_0\left(\pi_0^{(0)}\middle|\mathbf{d}\right) = 1$ and $q_T\left(\pi_T^{(0)}\middle|\mathbf{d}\right) = 1$ for the first order approximation with $k = 1$. An interior kth order categorical state thus follows a kth order Markov chain, with additional edge factors due to the convolution edge effects. The kth

order prior transition probabilities can be shown to depend on the last transition only as the prior Markov chain is of first order, i.e. $p\left(\pi_t^{(k)}\middle|\,\pi_{t-1}^{(k)}\right) = p(\pi_t|\pi_{t-1})$. Notice in particular that due to the multiplicative form of (11) any zero-probability events in the prior model will be inherited by the posterior model.

The performance of the factorisable form is expected to improve for higher order likelihood approximations, which capture more of the spatial structure caused by convolution. The approximate posterior model is however computable in $O((T-k+2)L^k)$ [14] and there is thus a trade-off between algorithm accuracy and computational demands. Rimstad and Omre [12] concluded that an approximation of order $k=2$ is appropriate when all parameters are known. The normalizing constant $C_d^{(k)}$ is implicitly computed in the FB algorithm and it corresponds to a kth order approximation of the marginal likelihood that appears in the posterior model in (3), i.e. $\hat{p}^{(k)}(\mathbf{d}) = 1/C_d^{(k)}$.

III. Model Parameter Estimation

For the blind categorical deconvolution problem we have observed only $\mathbf{d}$, with the variables $\boldsymbol{\pi}$ and $\mathbf{r}$ and the observation likelihood parameters $\boldsymbol{\theta}_{ol} = \{\mathbf{w}, \sigma_d\}$, containing the convolution parameters, unknown. The parameters in the prior model $\boldsymbol{\theta}_p$ and response likelihood model $\boldsymbol{\theta}_{rl}$ are assumed to be known, for the seismic deconvolution problem in particular this corresponds to estimates obtained from observations in well A in Fig.1. Assessment of $\boldsymbol{\theta}_{ol}$ is complicated due to the unknown high-dimensional convolution kernel vector $\mathbf{w}$. In the following, we define a procedure for estimating $\boldsymbol{\theta}_{ol}$, first by a maximum marginal likelihood method and next by Bayesian methods. For notational simplicity we will set $\boldsymbol{\theta}_{ol} = \boldsymbol{\theta}$ throughout the rest of this paper, since we assume $\boldsymbol{\theta}_p$ and $\boldsymbol{\theta}_{rl}$ to be known.

A. Maximum Marginal Likelihood Estimation

In a maximum marginal likelihood (MML) setting we search the parameter values of $\boldsymbol{\theta}$ maximizing the likelihood of the observations $\mathbf{d}$ randomized over $(\boldsymbol{\pi}, \mathbf{r})$. We consider the marginal likelihood of $\mathbf{d}$ for which the unknown categorical field is summed out

$$p(\mathbf{d};\boldsymbol{\theta}) = \sum_{\boldsymbol{\pi}} \left[p(\mathbf{d}|\boldsymbol{\pi};\boldsymbol{\theta}) \times p(\boldsymbol{\pi})\right]. \tag{14}$$

This is identical to the inverse of the normalizing constant in (3) which exact computation is infeasible in practice, as it requires $O(L^T)$ operations when summing over all possible combinations of $\boldsymbol{\pi}$. With the factorisable posterior model approximation in (11) we can however compute a kth order approximate marginal likelihood by the FB algorithm i.e. $\hat{p}^{(k)}(\mathbf{d};\boldsymbol{\theta}) = 1/C_d^{(k)}(\boldsymbol{\theta})$. Hence, a kth order approximate MML estimate is

$$\widehat{\boldsymbol{\theta}}_{MML}^{(k)} = \arg\max_{\boldsymbol{\theta}} \left\{\hat{p}^{(k)}(\mathbf{d};\boldsymbol{\theta})\right\} = \arg\max_{\boldsymbol{\theta}} \left\{-\log\left[C_d^{(k)}(\boldsymbol{\theta})\right]\right\}. \tag{15}$$

Notice that the approximate marginal likelihood estimate is exact with respect to the approximate likelihood model. A workflow diagram of the full procedure including parameter estimation by MML is presented in Fig.3.

B. Estimation by Bayesian Inference

Bayesian inference requires us to assign a prior model to the unknown parameters, $p(\boldsymbol{\theta})$. The parameter posterior model is then given by $p(\boldsymbol{\theta}|\mathbf{d}) \propto p(\mathbf{d}|\boldsymbol{\theta}) \times p(\boldsymbol{\theta})$ whose mode is the parameter maximum a posteriori prediction (PMAP) estimate. The likelihood term $p(\mathbf{d}|\boldsymbol{\theta})$ refers to the marginal likelihood in (14), but in which $\boldsymbol{\theta}$ is now a random variable. A kth order approximate PMAP estimate is thus given by

$$\widehat{\boldsymbol{\theta}}_{PMAP}^{(k)} = \arg\max_{\boldsymbol{\theta}} \left\{\hat{p}^{(k)}(\boldsymbol{\theta}|\mathbf{d})\right\} = \arg\max_{\boldsymbol{\theta}} \left\{\log\left[p(\boldsymbol{\theta})\right] - \log\left[C_d^{(k)}(\boldsymbol{\theta})\right]\right\}. \tag{16}$$

Estimation by PMAP therefor constitutes the same optimization problem as described for MML regularized by the parameter prior model.

Alternatively we can try to assess the entire posterior pdf $p(\boldsymbol{\theta}|\mathbf{d})$ through simulation based Bayesian inference by McMC. We then assess $\boldsymbol{\theta}$ by a large number of simulations from its posterior distribution $p(\boldsymbol{\theta}|\mathbf{d})$ which require evaluations of the complicated likelihood term $p(\mathbf{d}|\boldsymbol{\theta})$. By expanding our algorithm to account for the response field $\mathbf{r}$ we can however eliminate the need for this evaluation. For iteration s, one full update would thus constitute to first simulate $\mathbf{r}^s \sim p(\mathbf{r}|\boldsymbol{\theta}^s, \mathbf{d})$ and then simulate $\boldsymbol{\theta}^s \sim p(\boldsymbol{\theta}|\mathbf{r}^s, \mathbf{d})$. After some burn-in s_b the realizations $\{(\mathbf{r}^s, \boldsymbol{\theta}^s)\}_{s=s_b}^{S}$ will then be realizations from the posterior model $p(\mathbf{r}, \boldsymbol{\theta}|\mathbf{d})$ from which we can make inference of the unknown $\boldsymbol{\theta}$. Contrary to the simpler BG model [2], [5], we are however not able to simulate $\mathbf{r}$ directly from its posterior distribution $p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{d})$. We must thus simulate $\mathbf{r}$ in an McMC step, which can be complicated due to large coupling caused by the convolution. Results from preliminary simulation studies showed either very low acceptance rates or very poor mixing in the McMC step. Further investigation on proper simulation schemes is in progress, but this subject is not discussed further here. Hence we do not consider parameter estimation by McMC throughout the rest of this paper.

C. Parametric Convolution Kernel

In our experience, optimization by MML over the full high dimensional parameter space is a hard problem and the likelihood function in (14) has multiple local maxima. To reduce the dimension of the optimization problem we define parametric convolution kernels. In the seismic case study we use a Ricker wavelet parameterization of the convolution kernel, which is further discussed in Section V. In the empirical simulation study in Section IV we use discretized and normalized Beta functions. These convolution kernels are termed Beta convolution kernels and their first order derivative are termed Beta derivative convolution

kernels:

$$b(t;\alpha,\beta) = c(\alpha,\beta)[t(1-t)]^{\beta-1}, \qquad -\alpha \le t \le \alpha, \quad \beta \ge 1, \quad \alpha \ge 0 \tag{17}$$

$$b'(t;\alpha,\beta) = c(\alpha,\beta)(\beta-1)[t(1-t)]^{\beta-2}(1-2t), \qquad -\alpha \le t \le \alpha, \quad \beta \ge 2, \quad \alpha \ge 0. \tag{18}$$

The parameter α is a discrete width parameter that defines the kernel support $[-\alpha, \alpha]$, β is a continuous shape parameter and $c(\cdot)$ is a normalizing constant which ensures that the weights in the Beta convolution kernel sum to one. Some continuous Beta and Beta derivative models for different shape parameters β and with equal support $[-\alpha, \alpha]$ are displayed in Fig.4. Observe that the Beta convolution kernel parameterization captures both uniform shapes and hemispherical shapes with finite support. The model parameters of interest is then $\boldsymbol{\theta} = \{\alpha, \beta, \sigma_d\}$ with discrete $\alpha \ge 0$ and continuous $\beta \ge 1$ or 2 and $\sigma_d \ge 0$.

The parameters of the Beta model are for some ranges highly coupled, however. For high values of both α and β the resulting kernel can be almost exactly reproduced by a considerably lower α and medium size β. The example with support $[-\alpha, \alpha]$ and $\beta = 20$ in Fig.4(a) can be almost reproduced with a narrower support and somewhat smaller β. In many cases it is favorable to have a narrow support, and this parameterization can be encouraged by Bayesian inference with a suitable prior model for α.

IV. Empirical Study

First we conduct a simulation study for a given model with four different convolution kernel parameterizations. Thereafter we use one convolution kernel parameterization and evaluate the sensitivity to one model parameter; namely the transition matrix in the hidden categorical layer.

A. Simulation Study

The simulation study is based on a convolutional two-level HMM of length $T = 100$ with $L = 3$ possible classes $\Omega_\pi = \{\text{white,grey,black}\}$. The reference prior model transition matrix is defined as

$$\mathbf{P} = \begin{pmatrix} 0.50 & 0.50 & 0 \\ 0.33 & 0.34 & 0.33 \\ 0 & 0.50 & 0.50 \end{pmatrix} \tag{19}$$

with stationary distribution $(0.28, 0.44, 0.28)$. Observe from the transition matrix that the classes white and black can never be neighbors. The response likelihood model parameters are $\mu_{r|\pi} \in \{-2, 0, 3\}$, $\sigma_{r|\pi} = \{0.7, 0.7, 0.7\}$, corresponding to the three classes. A reference categorical profile $\boldsymbol{\pi}$ is generated, see Fig.5. Four observation profiles are simulated from the reference categorical profile $\boldsymbol{\pi}$ using Beta convolution kernels with reference parameters $(\alpha, \beta) = (4, 12.75)$, $(6, 1)$, $(5, 3)$ and a Beta derivative convolution kernel with reference parameters $(\alpha, \beta) = (7, 6)$ resulting in the profiles $\mathbf{d_1}$ through $\mathbf{d}_4$, see Fig.5. All the profiles contain an error term with standard deviation $\sigma_d = 0.3$. The convolution kernels are displayed in Fig.6 in solid thick black. The convolution kernel for case 1 is identical to a discretized truncated standardized Gaussian $N(0, 1)$ convolution kernel and case 2 is identical to a discrete uniform $U[-6, 6]$ convolution kernel. The signal-to-noise (S/N) ratio is defined from the full likelihood model in (8) as $S/N = \text{tr}[\text{Var}(\mathbf{W}\boldsymbol{\mu}_{r|\pi})]/\text{tr}[\mathbf{W}\boldsymbol{\Sigma}_{r|\pi}\mathbf{W}' + \sigma_d^2\mathbf{I}]$ and the S/N-values for the four cases are $[8.77, 5.62, 7.48, 14.41]$.

The observation likelihood parameters $\boldsymbol{\theta} = \{\alpha, \beta, \sigma_d\}$ are estimated by MML assuming a parametric convolution kernel on Beta or Beta derivative form. The MML estimation is performed for approximation orders $k = 1, \ldots, 6$, for which the approximations are expected to perform better for increasing order. As optimizer we use the EGO algorithm [16] which is a statistical derivative-free optimization scheme that has proven to be robust for identifying

the global maxima. We also compute the global MAPs of the categorical field $\boldsymbol{\pi}$ for the same approximation orders by the Viterbi algorithm [17], with plug-in estimated parameter values. These MAPs are compared to the MAP from inversion under the approximate posterior model with $k = 7$ with plug-in reference parameters values which represent the best classification we can obtain.

The convolution kernel estimates $\hat{\mathbf{w}}$ for varying orders k are displayed in Fig.6. The corresponding normalized root mean squared errors (NRMSE) are given in Fig.7, with NRMSE defined by

$$NRMSE = \frac{1}{\mathbf{w}_{max} - \mathbf{w}_{min}} \sqrt{\frac{1}{2a_m + 1} \sum_{i=-a_m}^{a_m} (\mathbf{w}_i - \hat{\mathbf{w}}_i)^2} \tag{20}$$

with $a_m = \max_a\{a_{\mathbf{w}}, a_{\hat{\mathbf{w}}}\}$ being the maximum width for the reference and estimated convolution kernel. The convolution kernel estimates seem to approach the reference convolution kernel for higher order approximations, with only marginal improvements after $k = 4$. All parameter estimates are given in Fig.8, with approximate 90% confidence intervals included for the continuous parameters β and σ_d, computed numerically by Hessians. For cases 1 through 3, the estimated Beta convolution kernels have too wide support for low order approximations, hence too large estimates of α. When the α estimates are close to or equals the reference kernel support for higher order approximations, the β estimates appear to be very reliable. Observe also that all error standard deviations σ_d are consistently underestimated to about 0.18 for $k = 6$ due to slightly overfitting of the model. For case 4, the estimated Beta derivative kernel supports are too narrow, hence too small estimates of α, while the β estimates are very reliable. The underestimation of the support causes a strong overestimation of the error standard deviation σ_d to account for the variance in the data not captured by the reduced convolution effect. To summarize, we obtain increasingly more reliable estimates of the convolution kernel with increasing order in the approximation. Hence there appears to be a trade-off between precision and computer demands. The variance of the error term

tends to be underestimated for Beta models and overestimated for Beta derivative models.

There is one surprising result in Fig.8(a), however. In case 1 the estimated kernel parameters for $k = 4$ provide a far too wide support $[-\alpha, \alpha]$ and severe overestimation of β. The estimated α and β deviates far more from the reference parameter values than the estimates for $k = 2$. Notice that the NRMSE values for the estimated kernel for $k = 4$ is lower than for $k = 2$ and hence closer to the reference Beta kernel in spite of the poor reproduction of the model parameters. Recall that the model parameters for some ranges of values are highly coupled, as previously discussed. If we reparameterize the estimated kernels within a NRMSE-range of $\pm 10^{-3}$ to a narrower support with a suitable β estimate, we obtain the parameter estimates displayed in Fig.9. The corresponding changes in the displays in Fig.6(a) and 7 would not be identifiable. The random error terms in the model are probably making the estimates in Fig.8(a) slightly better than the ones in Fig.9, although the estimated kernels are almost identical.

In Fig.10, the approximate global categorical MAPs for $\boldsymbol{\pi}$ computed by the kth order posterior approximation with corresponding plug-in parameter estimates are compared to the corresponding 7th order MAPs using the reference parameters. The MAPs for increasing approximation orders tend to converge to the reference MAP in cases 1 through 3. This is not surprising since the convolution kernel estimates approach the reference convolution kernel, and hence higher order approximations capture more of the convolution effect. For case 4 with a Beta derivative convolution kernel, the MAPs with plug-in parameter estimates are unable to capture the rapid class transitions due to the severe overestimation of the error variance. To summarize, the blind categorical deconvolution problem with convolution kernels according to the Beta model can be solved very reliably by low-order approximations of the posterior model. For Beta derivative convolution kernel models, the heterogeneity appears to be underestimated.

One single evaluation of the approximate marginal likelihood model for approximation order $k = 4$ (with $L = 3$ classes and profile length $T = 100$) requires about 0.3s on a 2.80GHz CPU, which we consider to be very reasonable. We choose the approximation of order $k = 4$ throughout the rest of this study as it seems to provide a good trade-off between accuracy and efficiency.

B. Sensitivity Study

In this section we only consider parameter estimation for profile $\mathbf{d}_1$ and we are interested in exploring the sensitivity to misspecification of the prior model for $\boldsymbol{\pi}$. We will normally not know the true transition probability matrix $\mathbf{P}$ in (19). In the blind deconvolution we assign prior models with a parameterized transition matrix

$$\mathbf{P}^*(\rho) = \begin{pmatrix} \rho & 1-\rho & 0 \\ (1-\rho)/2 & \rho & (1-\rho)/2 \\ 0 & 1-\rho & \rho \end{pmatrix} \tag{21}$$

with $0 \leq \rho \leq 1$, which can be shown to have identical stationary distribution $(0.25, 0.50, 0.25)$ for all ρ. Class transitions in the categorical field under $\mathbf{P}^*(\rho)$ will for large values of ρ have high probability of staying in the same class, hence favor thick layers, and the opposite for small values of ρ. The transition matrix $\mathbf{P}$ in (19) which is used to generate the profile $\boldsymbol{\pi}$ has equal probability for transitions into any valid class, including transitions back to the current class. We perform approximate MML estimation of order $k = 4$, but now using a prior model with transition matrix $\mathbf{P}^*(\rho)$ with $\rho = 0.2, 0.4, 0.6, 0.8$, denoted case 1 through 4. Notice that case 2 is closest to the transition matrix in (19).

The resulting convolution kernel estimates, their corresponding NRMSE values and the parameter estimates with approximate 90% confidence intervals are displayed in Fig.11. We observe that the convolution kernel estimates appear as fairly reliable for prior models with

small ρ, which entails models favoring frequent transitions. The estimates are much less reliable for models with too large ρ. Observe again that the error standard deviation is consistently underestimated when the convolution kernel is precisely estimated. The corresponding global categorical MAPs with plug-in estimates are displayed in Fig.12. Notice that the wrongly specified prior model has double effect when computing these MAP predictions, both in estimating the model parameters and in performing the inversion. The prediction for case 1, which favors too frequent transitions, resemble the reference profile quite well, but it is unable to capture some of the thicker layers as expected. The prediction for case 2, with a prior model closest to the correct model, provides the best prediction. The predictions for cases 3 and 4, with prior models favoring thick layers, appear with far too little heterogeneity. To summarize, model parameter estimation, i.e. convolution kernel and error variance estimations, appears as robust towards a wrongly specified prior model given that it favors frequent transitions. The MAP predictions, however, appear as sensitive to the choice of prior model in order to reproduce the correct heterogeneity.

V. Case Study - Seismic Inversion

We consider inversion of seismic amplitude-versus-offset (AVO) data into lithology/fluid (LF) classes, with parameters similar to the ones used in Rimstad and Omre [12]. A LF profile of length $T = 61$ units is used with $L = 4$ LF-classes $\Omega_\pi = \{$sand gas, sand oil, sand brine, shale$\}$ represented by colors white, light-gray, dark-gray and black respectively. The real LF profile $\boldsymbol{\pi}$ is displayed in Fig.13. The prior Markov chain transition matrix is defined as

$$\mathbf{P} = \begin{pmatrix} 0.50 & 0 & 0 & 0.50 \\ 0.05 & 0.35 & 0 & 0.60 \\ 0.05 & 0.05 & 0.60 & 0.30 \\ 0.05 & 0.05 & 0.20 & 0.70 \end{pmatrix} \tag{22}$$

with stationary distribution $(0.09, 0.07, 0.28, 0.56)$ defining the prior proportions of the classes. The transition matrix in (22) is assessed by a combination of the empirical transition matrix from the full profile and prior information on the geometry of the layering. We assign slightly higher probabilities to transitions since this appears to give better convolution kernel estimates, see Section IV. Observe the gravitational ordering of the fluids reflected in the transition matrix; sand-brine is never on top of sand-oil which in turn is never on top of sand-gas. The response layer $\mathbf{r}$ is represented by the logarithm of the three elastic material properties P-wave velocity; $\log(v_p)$, S-wave velocity; $\log(v_s)$ and density; $\log(\rho)$, which we assume to be independent given the LF profile. The respective response likelihood parameters are

$$\mu_{r|\pi} \in \begin{cases} \{8.11, 8.12, 8.16, 7.88\} \\ \{7.61, 7.60, 7.58, 7.08\} \\ \{7.66, 7.69, 7.73, 7.77\} \end{cases}, \quad \sigma_{r|\pi} = \begin{cases} 0.0860 \\ 0.1483 \\ 0.0608 \end{cases} \tag{23}$$

where the standard deviation $\sigma_{r|\pi}$ is identical for all LF-classes. The observations $\mathbf{d}$ represent angle-dependent seismic AVO data for the angles $(12°, 22°, 31°)$, see Fig.13, and these observations are synthetically generated following Rimstad and Omre [12]. The observation likelihood in (7) is now defined by $\mathbf{W} = \mathbf{CAD}$ in which $\mathbf{C}$ is the convolution matrix, $\mathbf{A}$ is a matrix of angle-dependent weak Aki-Richards coefficients [18] and $\mathbf{D}$ is a (central) differential matrix that calculates contrasts (see Buland and Omre [15] for more details). The convolution kernel model is chosen as a parametric Ricker wavelet, defined by

$$R(t; \lambda, \gamma) = \gamma \left(1 - \frac{t^2}{\lambda^2}\right) \exp\left\{-\frac{t^2}{2\lambda^2}\right\}, \qquad \lambda > 0, \quad \gamma \geq 0 \, . \tag{24}$$

Here λ is a continuous shape parameter and γ a continuous scaling parameter, related to the wavelet frequency and amplitude respectively. The wavelet is discretized to coincide with the seismic AVO data and truncated at 5λ. The reference wavelets for the three angles provided by the data set owner and some discretized parametric Ricker wavelets for different choices

of (λ, γ) are displayed in Fig.14. Observe that the reference wavelets for different angles appear as almost identical, with very similar amplitude and frequency. The corresponding Ricker wavelet representation (minimum NRMSE) has parameters $(\lambda, \gamma) = (2.43, 25.87)$, see Fig.14(b). We consider these parameter values as the references. The reference error term has standard deviation $\sigma_d = 0.207$, assessed by non-blind MML estimation. The corresponding S/N-ratio is 1.85, which is considerably poorer than in our empirical study in Section IV.

We perform parameter estimation by MML for approximation order $k = 4$ for the observation likelihood parameters $\boldsymbol{\theta} = (\lambda, \gamma, \sigma_d)$ for each angle, hence nine parameters. In the empirical study in Section IV we found this approximation order to be suitable. The Ricker wavelets estimated by MML are displayed in Fig.15, with corresponding NRMSE values $(12.22, 10.14, 10.04)$. The MML parameter estimates with 90% confidence intervals are displayed in Fig.16(a). Observe that for all angles the MML estimates of the Ricker wavelet frequency λ appear as very reliable, whereas the amplitudes γ are clearly overestimated. The estimates of the error standard deviation σ_d appear as reliable although slightly upward biased. Note also that the confidence intervals for the amplitude γ estimates are very wide and cover the reference value for two out of three angles. Our interpretation of the large uncertainty in the amplitude γ estimates is that the Ricker wavelet is almost centered around the zero, hence a scaling of the wavelet has little impact on the scaling of the observations $\mathbf{d}$. Consequently $\mathbf{d}$ carries little information about γ. We therefor also perform Bayesian inference by PMAP for approximation order $k = 4$, assigning a gamma prior model to γ; $p(\gamma) = G(6, 3.8)$ and almost non-informative gamma and inverse gamma prior models to λ and σ_d; $p(\lambda) = G(1.1, 100)$ and $p(\sigma_d^2) = IG(1/1000, 1)$. Hence we assign a prior model with 0.90 range $[10, 40]$ on the wavelet amplitude γ, enforcing a weak restriction on the amplitude being in a range we consider most likely. Our prior belief can be based on observations of both the seismic AVO data and the LF-profile in a representative well in the region, corresponding

to Fig.1, when doing blind categorical deconvolution along trace B, the weak prior restriction on the wavelet amplitude should be assessed from well A. The Ricker wavelet PMAP estimates are displayed in Fig.15, with corresponding NRMSE values $(2.11, 2.26, 4.18)$. The PMAP parameter estimates with 90% confidence intervals are presented in Fig.16(b). Observe that the amplitude γ is estimated more reliably, with small improvements also on the estimated frequencies λ, and with some reduction of the confidence intervals. The error standard deviations σ_d are consistently slightly overestimated, but the differences in the estimates for MML and PMAP are small.

In Fig.17, the approximate global MAPs for the LF-profile $\boldsymbol{\pi}$ with plug-in parameter MML and PMAP estimates are compared to the global MAP with plug-in reference parameters. The MAP of $\boldsymbol{\pi}$ with plug-in MML estimates predicts too rapid class transitions and fails to identify the thick sand-gas layer. This is caused by overestimation of the amplitude γ. The MAP of $\boldsymbol{\pi}$ with plug-in PMAP parameter estimates predicts the correct transitions between shale and sand-brine much better and identifies the sand-gas layer. All MAPs of $\boldsymbol{\pi}$ fail to identify the thin sand-oil layer below the sand-gas layer which obviously is very difficult to capture due to the poor S/N-ratio and the convolution effect. To summarize, blind deconvolution of seismic AVO data is feasible although prior models need to be assigned to the wavelet amplitudes due to lack of amplitude information in the data. Both wavelet frequencies and error variances can be reliably assessed.

VI. Conclusion

We solve the blind categorical deconvolution problem; the inversion of an observed convolved profile with errors into an unobserved categorical profile based on a convolved two-level hidden Markov model. The inversion is cast in a Bayesian inversion framework. The unknown model parameters are the convolutional kernel and the error variance. We perform the model

parameter inference and inversion under an approximate model which can be assessed exactly by a computer efficient recursion. A sequence of approximations are defined such that trade-offs between accuracy and computational demands can be made.

The approach is evaluated in a small empirical test on a categorical profile with three classes and several convolutional kernels. Four parameterized kernel-types are evaluated. It is demonstrated that even a low-order maximum likelihood approximation, actually fourth order, is providing reliable estimates of the convolutional kernel and error variance. Higher-order approximations can provide more precise estimates with larger computational costs. The inversion into a categorical profile can also be made reliably based on the model with plug-in of estimated model parameters.

The approach requires the transition matrix of the underlying categorical Markov model to be specified. We also evaluate the sensitivity of the results with respect to misspecification of this transition matrix. We find that the results are less sensitive towards misspecifications in the direction of more frequent transitions than the opposite.

An inversion example on seismic amplitude-versus-offset data for three angles is presented. The convolutional kernel is parameterized by a Ricker wavelet. The shape of the wavelet and the error variance can be reliably assessed whereas the amplitude of the wavelet is poorly determined. Seismic data provide limited information about the wavelet amplitude. By assigning a prior model on the wavelet amplitude, a reasonable approximate maximum posterior estimate of the amplitude can be made. The inversion into the categorical profile is relatively reliable based on plug-in values of the latter model parameter estimates.

Appendix

Forward-Backward Algorithm

The forward-backward algorithm presented follows Reeves and Pettitt [14].

Forward-backward algorithm

Forward:

- Initiate for $t = k$:
$$z_k\left(\pi_k^{(k-1)}\right) = \sum_{\pi_1} \prod_{i=1}^{k} p(\pi_i|\pi_{i-1}) \times \hat{p}^{(i)}\left(d_i^{(i)}\middle|\,\pi_i^{(i)}\right)^{1/k}$$

- Iterate for $t = k+1, \ldots, T-1$:
$$z_t\left(\pi_t^{(k-1)}\right) = p(\pi_t|\pi_{t-1}) \times \sum_{\pi_{t-k+1}} \hat{p}^{(k)}\left(d_t^{(k)}\middle|\,\pi_t^{(k)}\right)^{1/k} \times z_{t-1}\left(\pi_{t-1}^{(k-1)}\right)$$

- Last element, $t = T$:
$$z_T\left(\pi_T^{(k-1)}\right) = p(\pi_T|\pi_{T-1}) \times \sum_{\pi_{T-k+1}} \prod_{i=1}^{k} \hat{p}^{(i)}\left(d_T^{(i)}\middle|\,\pi_T^{(i)}\right)^{1/k} \times z_{T-1}\left(\pi_{T-1}^{(k-1)}\right)$$

- Normalizing constant:
$$C_d^{(k)} = \left[\sum_{\pi_{T-k+2}} \cdots \sum_{\pi_T} z_T\left(\pi_T^{(k-1)}\right)\right]^{-1}$$

Backward:

- Initiate:
$$p_b\left(\pi_T^{(k-1)}\right) = C_d^{(k)} \times z_T\left(\pi_T^{(k-1)}\right)$$
$$p_b\left(\pi_{T-k+1}\middle|\pi_T^{(k-1)}\right) = \frac{p(\pi_T|\pi_{T-1}) \times \prod_{i=1}^{k} \hat{p}^{(i)}\left(d_T^{(i)}\middle|\,\pi_T^{(i)}\right)^{1/k} \times z_{T-1}\left(\pi_{T-1}^{(k-1)}\right)}{z_T\left(\pi_T^{(k-1)}\right)}$$

- Iterate for $t = T-k, \ldots, 2$:
$$p_b\left(\pi_t\middle|\pi_{t+k-1}^{(k-1)}\right) = \frac{p(\pi_{t+k-1}|\pi_{t+k-2}) \times \hat{p}^{(k)}\left(d_{t+k-1}^{(k)}\middle|\,\pi_{t+k-1}^{(k)}\right)^{1/k} \times z_{t+k-2}\left(\pi_{t+k-2}^{(k-1)}\right)}{z_{t+k-1}\left(\pi_{t+k-1}^{(k-1)}\right)}$$

- Last element $t = 1$:
$$p_b\left(\pi_1\middle|\pi_k^{(k-1)}\right) = \frac{\prod_{i=1}^{k} p(\pi_i|\pi_{i-1}) \times \hat{p}^{(i)}\left(d_i^{(i)}\middle|\,\pi_i^{(i)}\right)^{1/k}}{z_k\left(\pi_k^{(k-1)}\right)}$$

The full kth order approximate posterior model is computed by

$$\hat{p}^{(k)}(\boldsymbol{\pi}|\mathbf{d}) = \left[\prod_{t=1}^{T-k} p_b\left(\pi_t\middle|\pi_{t+k-1}^{(k-1)}\right)\right] \times p_b\left(\pi_T^{(k-1)}\right).$$

References

[1] J. J. Kormylo and J. M. Mendel, "Maximum-likelihood seismic deconvolution," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-21, no. 1, pp. 72–82, 1983.

[2] Q. Cheng, R. Chen, and T. H. Li, "Simultaneous wavelet estimation and deconvolution of reflection seismic signals," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 2, pp. 377–384, 1996.

[3] G. Celeux and J. Diebolt, "A stochastic approximation type EM algorithm for the mixture problem," *Stochastics and Stochastic Rep.*, vol. 41, no. 1-2, pp. 119–134, 1992.

[4] M. Lavielle, "A stochastic algorithm for parametric and non-parametric estimation in the case of incomplete data," *Signal Process.*, vol. 42, pp. 3–17, 1995.

[5] O. Rosec, J. M. Boucher, B. Nsiri, and T. Chonavel, "Blind marine seismic deconvolution using statistical MCMC methods," *IEEE J. Ocean. Eng.*, vol. 28, no. 3, pp. 502–512, 2003.

[6] E. Baziw and J. U. Tadeusz, "Principle phase decomposition: A new concept in blind seismic deconvolution," *IEEE Trans. Geosc. Remote Sens.*, vol. 44, no. 8, pp. 2271–2281, 2006.

[7] N. de Freitas, "Rao-Blackwellized particle filtering for fault diagnosis," *Proc. IEEE Aerosp. Conf.*, vol. 4, pp. 1767–1772, 2002.

[8] I. L. MacDonald and W. Zucchini, *Hidden Markov and Other Models for Discrete-valued Time Series*. Taylor & Francis, 1997.

[9] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *The Ann. of Math. Stat.*, vol. 41, no. 3, pp. 164–171, 1970.

[10] A. L. Larsen, M. Ulvmoen, H. Omre, and A. Buland, "Bayesian lithology/fluid prediction and simulation on the basis of a Markov-chain prior model," *Geophysics*, vol. 71, no. 5, pp. R69–R78, 2006.

[11] R. E. Sheriff and L. P. Geldart, *Exploration Seismology*, 2nd ed. Cambridge University Press, 1995.

[12] K. Rimstad and H. Omre, "Approximate posterior distributions for convolutional two-level hidden Markov models," *Computational Stat. & Data Anal.*, vol. 58, pp. 187–200, 2013.

[13] W. C. Krumbein and M. F. Dacey, "Markov chains and embedded Markov chains in geology," *Math. Geology*, vol. 1, no. 1, pp. 79–96, 1969.

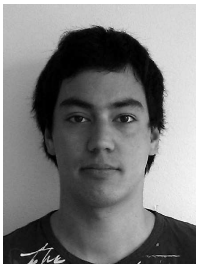
[14] R. Reeves and A. N. Pettitt, "Efficient recursions for general factorisable models," *Biometrika*, vol. 91, no. 3, pp. 751–757, 2004.

[15] A. Buland and H. Omre, "Bayesian linearized AVO inversion," *Geophysics*, vol. 68, no. 1, pp. 185–198, 2003.

[16] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black box functions," *J. of Global Optimization*, vol. 13, pp. 455–492, 1998.

[17] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 260–269, 1967.

[18] K. Aki and P. G. Richards, *Quantitative seismology: Theory and methods.* New York: W.H. Freeman and Co., 1980.



David Volent Lindberg received the M.Sc. degree in industrial mathematics from the Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 2010.

He is currently working toward the Ph.D. degree in statistics at NTNU. His research interests include Bayesian inverse problems, hidden Markov models, blind seismic deconvolution, state-space models, parameter estimation and stochastic optimization.

Henning Omre received the M.Sc. degree in statistics from the Norwegian University of Science and Technology (NTNU), Trondheim, Norway in 1975 and Ph.D. in geostatistics from Stanford University, California, in 1985.

He was employed at the Norwegian Computing Center in the period 1976-1992, and joined the Department of Mathematical Sciences (DMS) of NTNU in 1992 as Professor in Statistics. He is the head of the 'Uncertainty in Reservoir Evaluation' (URE) consortium at DMS/NTNU. His major topics of research are spatio-temporal statistics and Bayesian inversion.

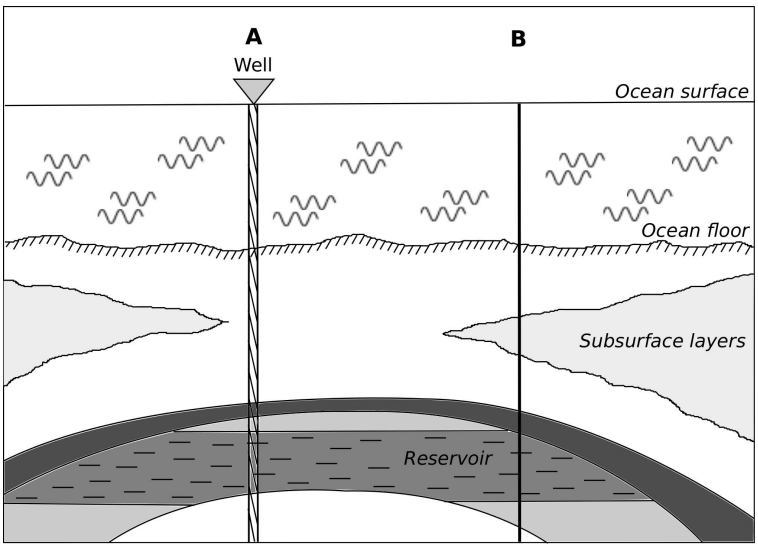
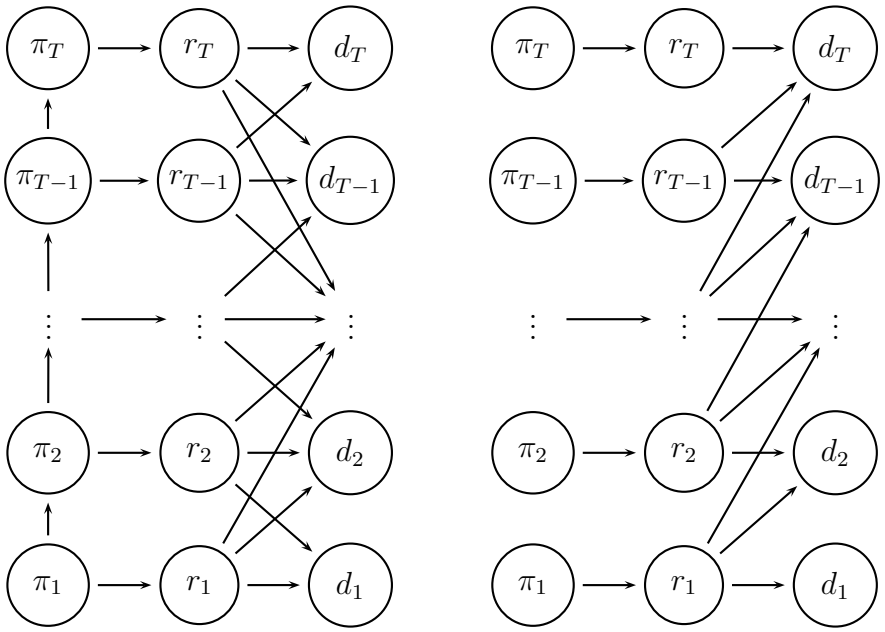


Fig. 1: Cross section of a schematic offshore reservoir, with an exploration well in lateral position A and with seismic data available in position B.



(a) Two-level HMM

(b) BG model

Fig. 2: Directed acyclic graph of (a) the convolutional two-level HMM and (b) the BG model for observed data $\mathbf{d} = (d_1, \ldots, d_T)$ with a latent continuous field $\mathbf{r} = (r_1, \ldots, r_T)$ and a latent categorical field $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_T)$. The directed arrows represent dependencies between the variables. The state space of the categorical variables π_t is binary for the BG model, but may have multiple values for the convolutional two-level HMM.

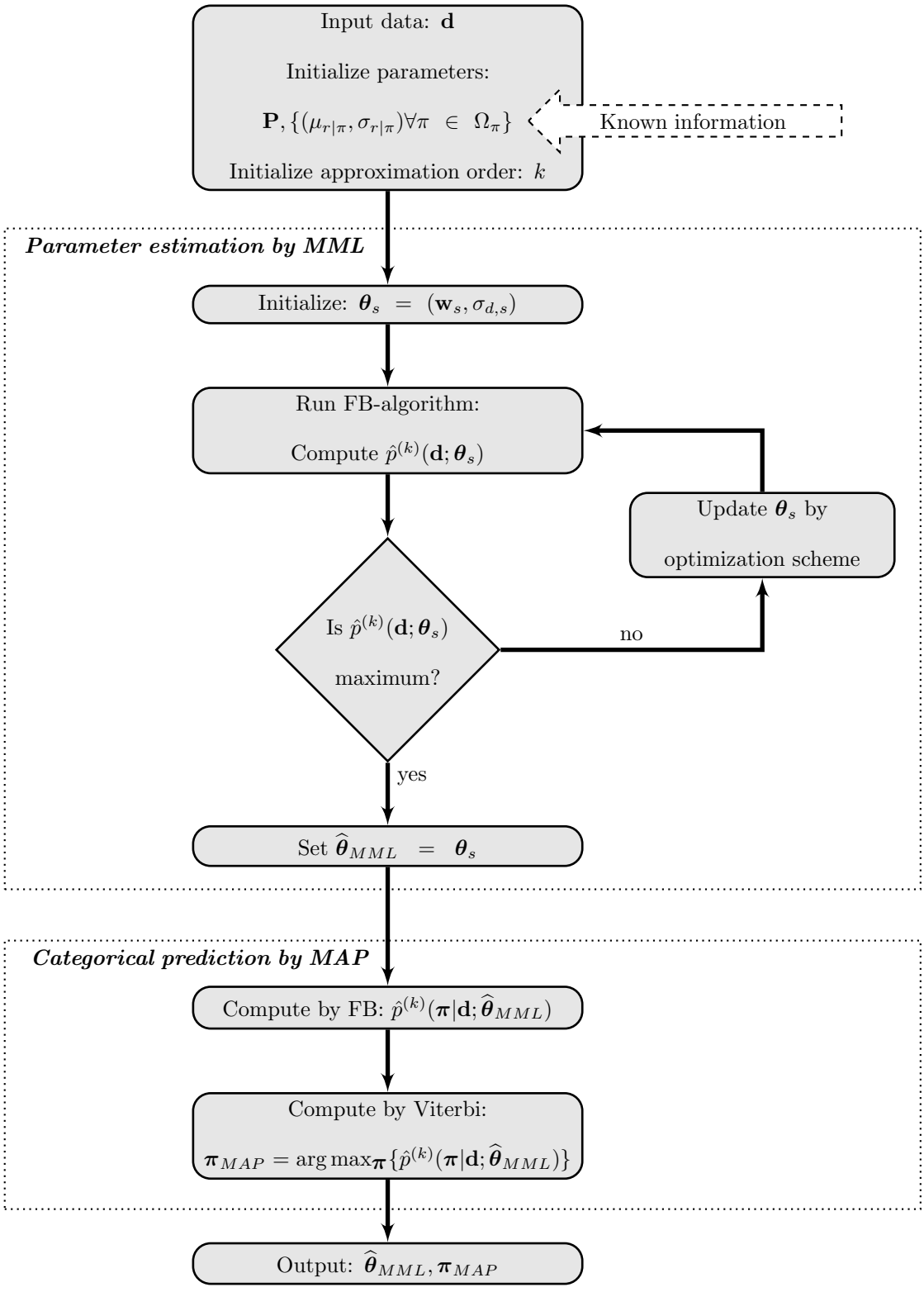
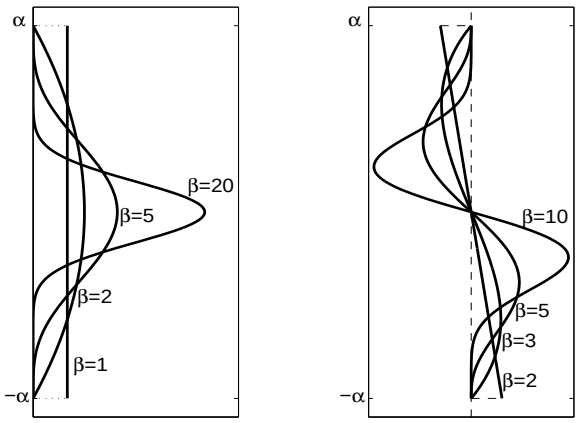


Fig. 3: Workflow diagram of the full procedure for parameter estimation by MML and categorical MAP prediction.



(a) Beta (b) Beta derivative

Fig. 4: Parametric convolution kernels for different parameters (α, β). (a) Continuous symmetric Beta model, $b(t; \alpha, \beta)$, (b) Continuous Beta derivative model, $b'(t; \alpha, \beta)$.

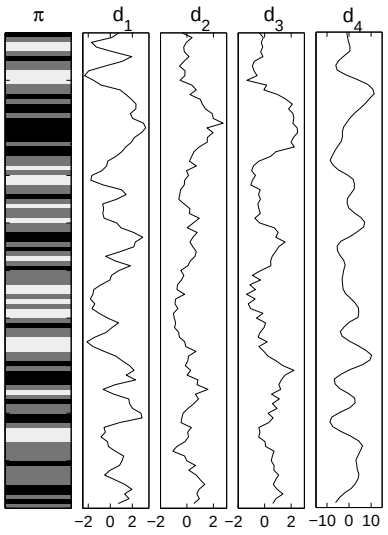
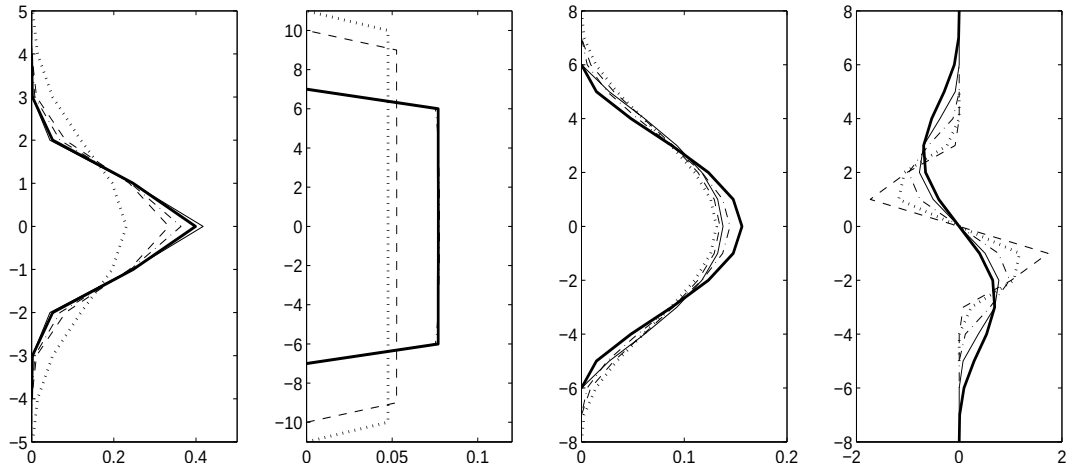


Fig. 5: Reference categorical profile with convolved observations for the four cases in the empirical study.



(a) Case 1 (b) Case 2 (c) Case 3 (d) Case 4

Fig. 6: Estimated parametric convolution kernels compared to the reference convolution kernels. Reference convolution kernels (thick solid), approximate ML estimates for orders $k = 1$ (dotted), $k = 2$ (dashed), $k = 4$ (mixed dotted and dashed) and $k = 6$ (thin solid).

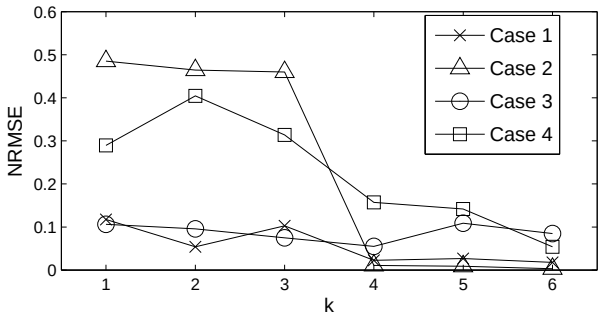
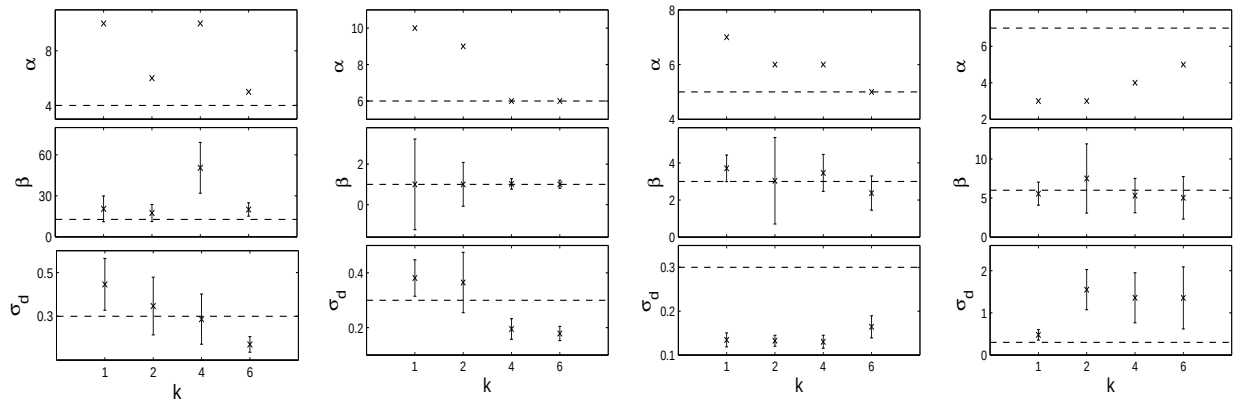


Fig. 7: NRMSEs of the estimated convolution kernels versus approximation order, for $k = 1$ through 6.



(a) Case 1 (b) Case 2 (c) Case 3 (d) Case 4

Fig. 8: Estimated convolution kernel parameters and error standard deviation parameters for orders $k = 1, 2, 4, 6$ with approximate 90% confidence intervals for the continuous parameters. Dashed lines indicate the reference parameter values.

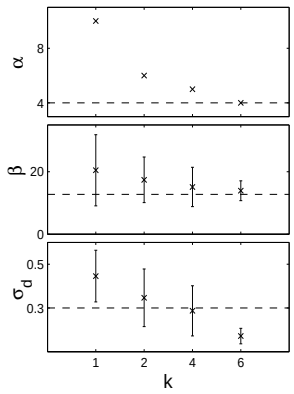
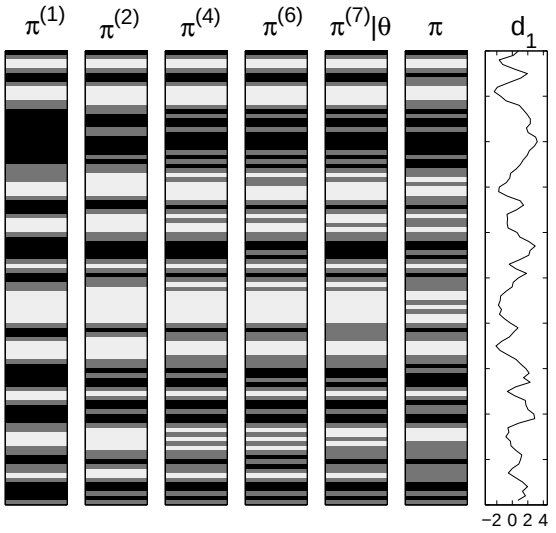
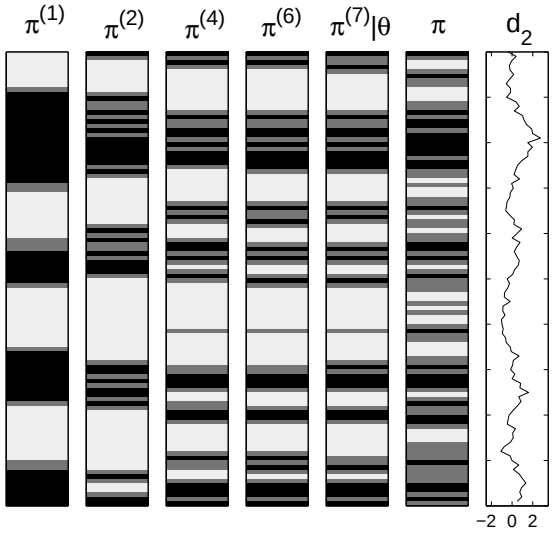


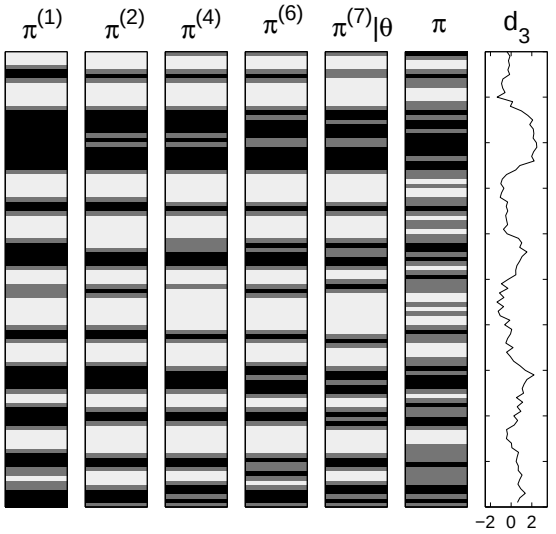
Fig. 9: Reparameterized convolution kernel parameter estimates and estimated error standard deviation parameters for case 1.



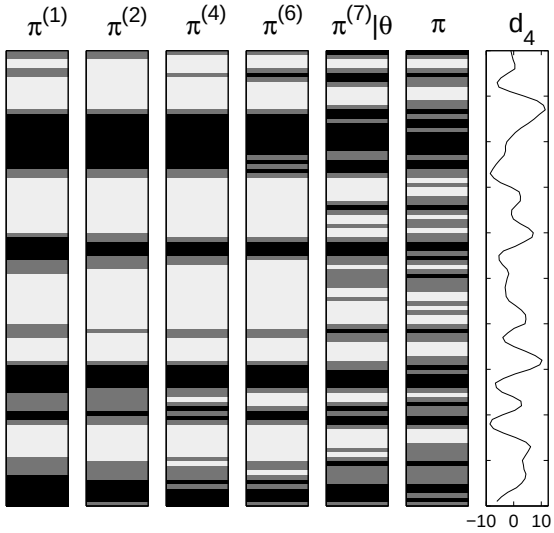
(a) Case 1



(b) Case 2

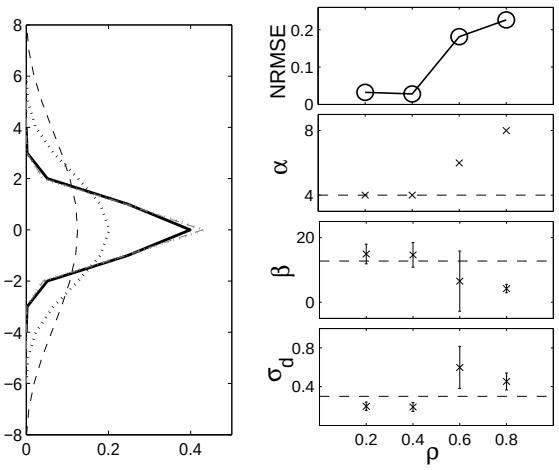


(c) Case 3



(d) Case 4

Fig. 10: Categorical global MAPs with plug-in parameter estimates for $k = 1, 2, 4, 6$, $\boldsymbol{\pi}^{(k)} = \widetilde{\boldsymbol{\pi}}^{(k)}|\mathbf{d}; \widehat{\boldsymbol{\theta}}^{(k)}_{MML}$, compared to the global MAP with plug-in reference parameters for order $k = 7$, $\boldsymbol{\pi}^{(7)}|\boldsymbol{\theta} = \widetilde{\boldsymbol{\pi}}^{(7)}|\mathbf{d}; \boldsymbol{\theta}$, and the reference categorical profile $\boldsymbol{\pi}$.



(a) (b)

Fig. 11: (a) Estimated convolution kernels under sensitivity study for case 1, reference convolution kernel (solid thick black) and approximate estimates for the substitutes $\rho = 0.2$ (dotted grey), $\rho = 0.4$ (dashed grey), $\rho = 0.6$ (dotted black) and $\rho = 0.8$ (dashed black). (b) The convolution kernel estimates NRMSEs and the estimated parameters with 90% confidence intervals for the continuous parameters.

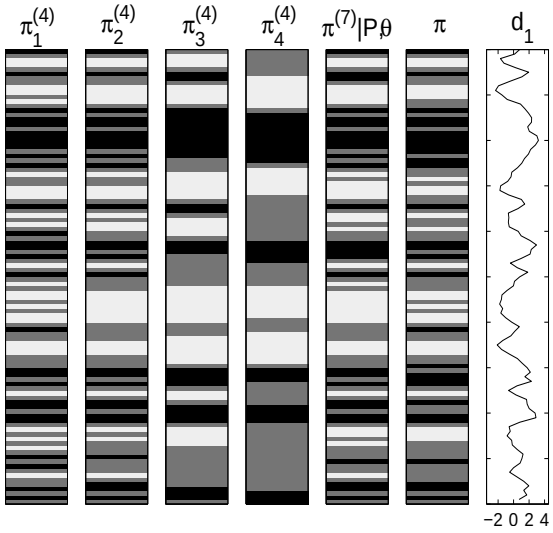


Fig. 12: Categorical global MAPs with plug-in parameter estimates under the transition matrices $\mathbf{P}_i^* = \mathbf{P}^*(\rho_i)$ for $\rho_i \in \{0.2, 0.4, 0.6, 0.8\}$, $\boldsymbol{\pi}_i^{(4)} = \widetilde{\boldsymbol{\pi}}^{(4)}|\mathbf{d}_1; \mathbf{P}_i^*, \left[\widehat{\boldsymbol{\theta}}_{MML}^{(4)}; \mathbf{P}_i^*\right]$, compared to the global MAP with plug-in reference parameters for order $k = 7$, $\boldsymbol{\pi}^{(7)}|\mathbf{P}, \boldsymbol{\theta} = \widetilde{\boldsymbol{\pi}}^{(7)}|\mathbf{d}_1; \mathbf{P}, \boldsymbol{\theta}$, and the reference categorical profile $\boldsymbol{\pi}$.

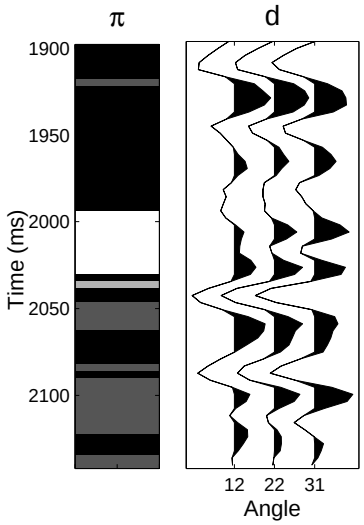
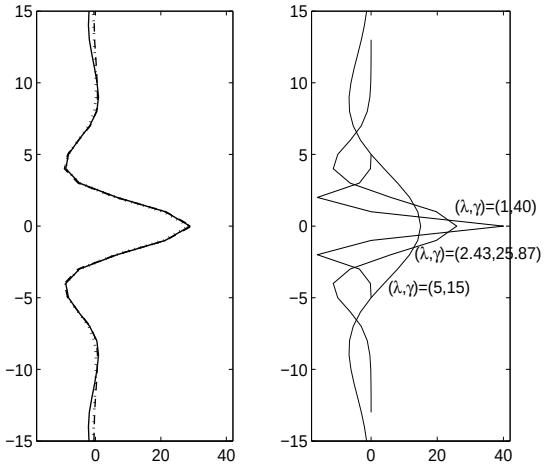
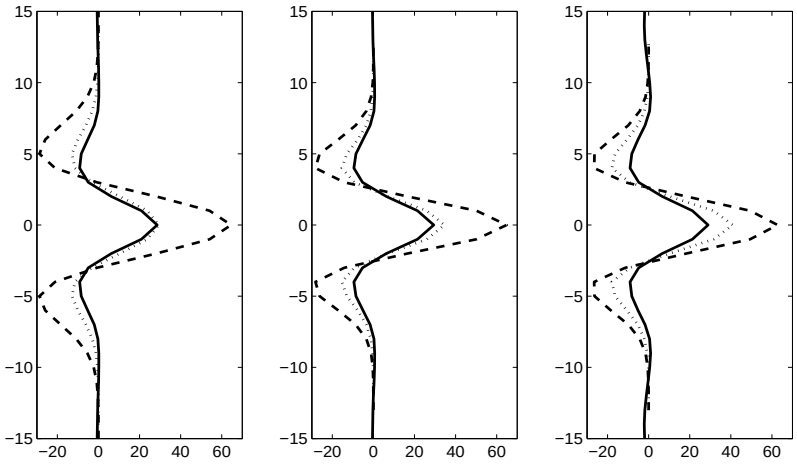


Fig. 13: Real lithology/fluid profile $\boldsymbol{\pi}$ and generated seismic AVO data $\mathbf{d}$ for the angles $(12°, 22°, 31°)$ plotted against time in milliseconds.



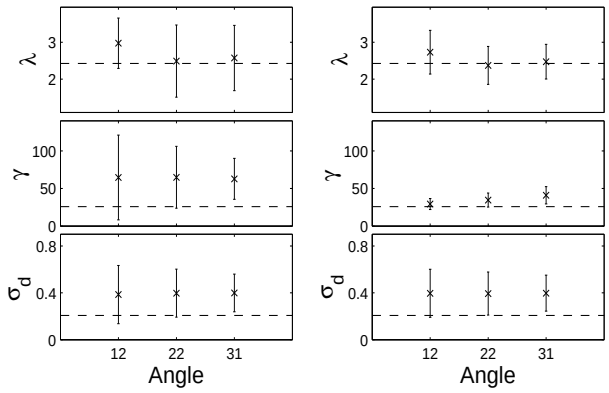
(a) (b)

Fig. 14: (a) Reference wavelets for the angles 12° (dotted), 22° (dashed) and 32° (solid) supplied from the data set owner. (b) Discretized Ricker wavelets $R(t; \lambda, \gamma)$ for some parameter sets including the assumed reference set.



(a) 12° (b) 22° (c) 32°

Fig. 15: Estimated Ricker wavelets compared to the reference wavelets for the AVO data. Reference wavelets (solid black), wavelet MML estimates (dashed black) and wavelet PMAP estimates (dotted black) for the three angles.



(a) MML (b) PMAP

Fig. 16: Estimated Ricker wavelet parameters and error standard deviation parameters with 90% confidence intervals for the AVO data with three angles. Dashed lines indicate the assumed reference parameter values.

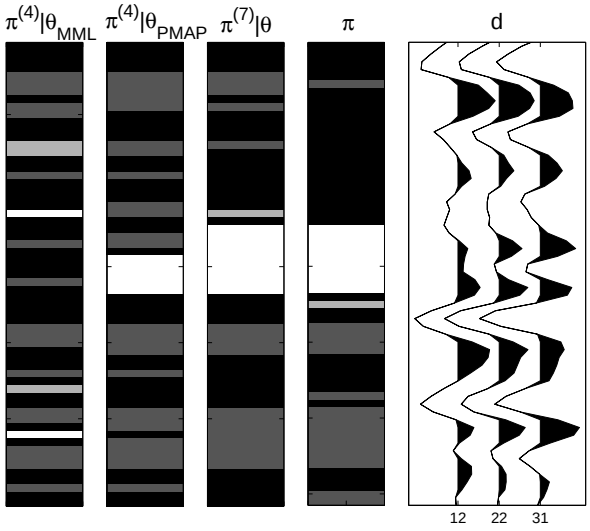


Fig. 17: Categorical global MAPs with plug-in parameter MML estimate, $\boldsymbol{\pi}^{(4)}|\boldsymbol{\theta}_{MML} = \widetilde{\boldsymbol{\pi}}^{(4)}|\mathbf{d}; \widehat{\boldsymbol{\theta}}^{(4)}_{MML}$, and plug-in PMAP estimate $\boldsymbol{\pi}^{(4)}|\boldsymbol{\theta}_{PMAP} = \widetilde{\boldsymbol{\pi}}^{(4)}|\mathbf{d}; \widehat{\boldsymbol{\theta}}^{(4)}_{PMAP}$ compared to the global MAP with plug-in reference parameters for order $k = 7$, $\boldsymbol{\pi}^{(7)}|\boldsymbol{\theta} = \widetilde{\boldsymbol{\pi}}^{(7)}|\mathbf{d}; \boldsymbol{\theta}$, and the reference categorical profile $\boldsymbol{\pi}$.