NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET

Estimation of geological attributes from a North Sea well log : an application of hidden Markov chains

by

Jo Eidsvik, Tapan Mukerji and Paul Switzer

PREPRINT
STATISTICS NO. 3/2002

NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL
http://www.math.ntnu.no/preprint/statistics/2002/S3-2002.ps
Jo Eidsvik has homepage: http://www.math.ntnu.no/~joeid
E-mail: joeid@stat.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

Estimation of Geological Attributes from a North Sea Well Log : an application of Hidden Markov Chains

Jo Eidsvik[1], **Tapan Mukerji**[2] **and Paul Switzer**[3]

1) Department of Mathematical Sciences, Norwegian University of Science and Technology, Norway

2) Department of Geophysics, Stanford Rock Physics Laboratory, Stanford University, California

3) Department of Statistics, Stanford University, California

Jo Eidsvik is PhD student, Department of Mathematical Sciences, Norwegian University of Science and Technology, 7491 Trondheim, Norway (email:joeid@math.ntnu.no)

Tapan Mukerji is Research Associate, Department of Geophysics, Stanford Rock Physics Laboratory, Stanford, CA 94305 (email:mukerji@pangea.stanford.edu)

Paul Switzer is Professor, Department of Statistics, Stanford, CA 94305 (email:ps@stat.stanford.edu)

ABSTRACT

Well logs consist of measurements of physical properties such as radioactivity, density and acoustic velocity made in the subsurface within a borehole. The objective is to describe rock type alternations along a well based on the indirect well log observations. Two major strata are recognized from seismic imaging of the reservoir, and rock types and associated rock alternations within the two strata are studied separately and compared using the well log data.

Rock type is modeled by a hidden Markov chain in a Bayesian framework, and we explore the posterior probability distributions for the rock type sequence and the Markov transition matrices conditional on observations in the well. Estimates of hyperparameters for the Markov transition probabilities and parameters of the measurement are obtained by iteratively maximizing the marginal likelihood. Gibbs sampling is an important tool both as part of this estimation scheme and when exploring the posterior probability distributions. We use parametric bootstrap to integrate parameter uncertainty in the posterior distributions for rock types and the Markov transition matrices.

The posterior distributions for the two strata show differences important for reservoir exploitation.

Keywords: hidden Markov models, well logs, marginal likelihood estimation, Gibbs sampling, imputation, sequence stratigraphy

1. INTRODUCTION AND SUMMARY

The alternation, with depth, of sedimentary rock types [facies] plays an important role in the assessment of petroleum reservoirs. Different alternation styles correspond to different ancient depositional, burial and diagenetic (processes related to chemistry, pressure and temperature) environments forming the rock layers. Information on the rock type alternations can be obtained indirectly from measurements [well logs] made along the length of drilled wells. We examine well log data from a productive North Sea reservoir, the Glitne field in Figure 1. These data are a trivariate sequence of radioactivity (gamma ray), density and sonic velocity (M modulus) measurements taken every 15cm at depths between 2150m and 2350m, and displayed in Figure 2a). The 200m zone is divided into upper and lower strata based on external seismic information as shown in Figure 2b). These strata reflect major geological horizons separating the rocks. The well log measurements are informative regarding rock type at the locations where they are taken. For example, high gamma ray counts within the borehole may be indicative of shale, a non-oil bearing rock. The goals are to use the data to assign one of three rock types [sand, shale and sand/shale mix] to each measurement depth, to statistically characterize the rock type alternations separately for each of the two strata, and to compare the alternation statistics for the two strata.

Section 2 describes a hierarchical Markov chain model for the hidden rock type alternations, with a prior distribution model for the transition matrix of the Markov chain. This is followed by a description of the measurement model that is conditional on the underlying rock type. The measurement model is represented by a trivariate Gaussian distribution where the mean values depend on unknown parameters linked to the underlying rock type and known covariates, and where the covariances depend on the underlying rock types. We also examine the possibility of an autocorrelated measurement term that is not related to rock type. The objective is to compute posterior distributions for the underlying rock type sequence and the Markov transition matrix, given the measurement data.

Section 3 describes our approach for approximating these posterior distributions. First, we maximize the marginal likelihood of the data, alternating between estimation of the prior distribution hyperparameters of the Markov transition matrix and estimation of the Gaussian distribution parameters of the measurement model. We then indicate how to sample from the implied posterior distributions for the hidden Markov chain and its transition matrix.

Section 4 describes the application of our modeling procedures to the North Sea data. Posterior distributions for the rock alternations and Markov transition matrix are computed for each of the two strata. A parametric bootstrap scheme studies the uncertainty of the estimated parameters and we attempt to integrate this uncertainty in the posterior distributions. The two strata show interesting differences that are interpreted in the context of the application. This can give valuable information about the way the reservoir was formed through geologic time.

Traditionally geologists have described and understood the different alternation styles of rock types, but only qualitatively. It has been relatively difficult to capture the geological knowledge in a quantitative framework suitable for statistical analysis. The hidden Markov chain formulation can be used to compare rock type transition probabilities in different strata. The hidden Markov model also provides a way to generate what are called 'pseudo-logs', i.e., well logs that are statistically similar to the observed log. These 'pseudo-logs' are important inputs for modeling the expected distribution and uncertainty of seismic signatures between wells where no logs exist. Quantitative analysis of rock distributions in reservoirs has become more important recently, see e.g. Malinverno (1997), Chen and Hiscott (1999) and Avseth et al. (2001). However, few of them study facies transition or facies correlation in time(space). One exception is Deshpande et al. (1997) who estimated the geological correlation along a well using methods from spectral theory on the continuous data directly, without working with categorical rock attributes. Others are Harbaugh and Bonham-Carter (1970) and Weissman and Fogg (1999) who studied Markov transition probabilities along wells using classified rock type values as data, and hence not taking into account the uncertainty of the classification.

2. HIERARCHICAL MARKOV CHAIN MODEL

A hierarchical Markov model is proposed for the rock types in the well, with the two strata modeled independently. Figure 3 illustrates the idea for one of the strata with y being the trivariate well log measurements, X being the rock type, P being the Markov transition matrix, and parameters α and β being hyperparameters for the Markov transition probabilities and parameters of the measurement, respectively.

Let $X = \{X_t; t = 1, \ldots, T\}$ be categorical variables, with X_t being the rock type (facies) at depth t. X represents the facies at every 15cm along the well, and each X_t takes on one of d values, where $d = 3$ in this study with possible rock types being $\{\text{sand}, \text{shale}, \text{mixed}\}$. $X = (X_1, \ldots, X_T)$ is assumed to be a hidden Markov chain with $d \times d$ stationary transition matrix $P = \{P_{ij}; i = 1, \ldots, d; j = 1, \ldots, d\}$, where $P_{ij} = Pr(X_t = j | X_{t-1} = i)$ are the one step facies transition probabilities with $\sum_{j=1}^{d} P_{ij} = 1$ for all i. The initial state of the Markov chain, x_0, is fixed. The probability distribution of X given P is hence;

$$f(x|p) = p_{x_0 x_1} \ldots p_{x_{T-1} x_T} \tag{1}$$

and this is illustrated with an arrow from P to X in Figure 3. The Markov chain models dependency in the alternating rock types along the well, and in this way maybe captures some of the depositional and burial processes. A Bayesian approach is used and the rows of the rock type transition matrix, P, denoted $\underline{P}_i; i = 1, \ldots, d$ are modeled by independent Dirichlet distributions;

$$f(p|\alpha) = f(\underline{p}_1, \ldots, \underline{p}_d | \underline{\alpha}) = \prod_{i=1}^{d} f(\underline{p}_i | \alpha_i) \tag{2}$$

$$\begin{aligned} f(\underline{p}_i|\alpha_i) &= Dirichlet(\alpha_{i1}, \ldots, \alpha_{id}) \\ &= \frac{\Gamma(\alpha_i)}{\prod_{j=1}^{d} \Gamma(\alpha_{ij})} p_{i1}^{\alpha_{i1}-1} \ldots (1 - \sum_{j=1}^{d-1} p_{ij})^{\alpha_{id}-1} \end{aligned} \tag{3}$$

with $\alpha = \{\alpha_{ij}, i, j = 1, \ldots, d$, being the matrix of prior parameters (hyperparameters), and $\alpha_i = \sum_{j=1}^{d} \alpha_{ij}$. In this study $\alpha_i = \kappa$ for all i, and κ measures the expected (prior) variability in the Markov transition probabilities. Larger values of κ indicate less prior variability. The Dirichlet prior relationship is illustrated in Figure 3 with an arrow going from α to P.

Given the hidden Markov sequence X, the well log data $y = \{y_t; t = 1, \ldots, T\}$, with $y_t = (y_t^1, y_t^2, y_t^3)$, are modeled as a sequence of Gaussian random vectors with mean and covariance parameters depending on X. Let $f(y|x, \beta)$ denote the conditional distribution of y given the Markov chain sequence X and the measurement parameters β.

The mean values of the well log data are modeled on a regression form since covariates, depth and pressure, are available. Radioactivity, gamma ray, $\gamma = \{\gamma_t : t = 1, \ldots, T\}$, is assumed to be a constant function of depth and pressure. Density, $\rho = \{\rho_t : t = 1, \ldots, T\}$, is an exponential function of depth. Sonic velocity, M modulus, $m = \{m_t : t = 1, \ldots, T\}$, is linear in the third root of pressure, see Mavko et al. (1995). This gives:

$$[y_t|x_t = k] = \begin{bmatrix} \gamma_t|x_t = k \\ \log(C - \rho_t)|x_t = k \\ m_t|x_t = k \end{bmatrix} = \begin{pmatrix} \eta_k^\gamma \\ h_t^\rho \eta_k^\rho \\ h_t^m \eta_k^m \end{pmatrix} + v_t(x) \tag{4}$$

where $h_t^\rho = [1, s_t]$ with s_t being depth at (time) point t, $h_t^m = q_t^{1/3}$ with q_t being pressure at (time) point t, and with C being a known constant given by the mineralogy, and with η_k^γ, η_k^ρ and η_k^m being unknown parameters depending on the rock type k. The mean vector from equation 4 is then $\mu_{t,k} = (\eta_k^\gamma, h_t^\rho \eta_k^\rho, h_t^m \eta_k^m)$, depending only on the facies k and covariates at time t. The noise term $v_t(x)$ is assumed to be Gaussian, and can in principle depend on more categorical values than just the one at time t. In a situation with uncorrelated measurement noise, the measurements at different depths are uncorrelated, and then $v_t(x) \sim \phi(v_t; 0, \Sigma_k)$ with the variance term depending only on the facies $X_t = k$. Denote by $\beta = (\eta_1^\gamma, \ldots, \eta_d^m, \Sigma_1, \ldots, \Sigma_d)$ the parameters of the measurement process in this case.

Often the case, however, measurements are correlated due to slow fluctuations(lags) in the measurement equipment and because measurements are not made pointwise, but rather integrated over a sliding window. Assume that the $v_t(x)$ terms are correlated according to an autoregressive model of order one, AR(1):

$$v_t(x) = \Psi v_{t-1}(x) + \epsilon_{x_{t-1},x_t} \qquad t = 1, \ldots, T \tag{5}$$

where the $\epsilon_{k,l}$s are uncorrelated and distributed as $\phi(\epsilon_{k,l}; 0, \Sigma_l)$, with k and l being the categorical values at subsequent time points. Denote by $\beta = (\eta_1^\gamma, \ldots, \eta_d^m, \Sigma_1, \ldots, \Sigma_d, \Psi)$ the measurement parameters in this case.

The posterior distribution of the rock type sequence and the Markov transition probabilities conditional on the data y and the parameters $\theta = (\alpha, \beta)$ is given by:

$$f(x, p|y, \theta) = \frac{f(y|x, \beta) f(x|p) f(p|\alpha)}{f(y|\theta)} \tag{6}$$

Figure 3, with equation 1, 2 and 4, fully specifies the model in equation 6, but because of high dimensionality it is not directly available.

3. METHODOLOGY

The objective is to characterize, separately for each stratum, the posterior distributions of the rock types X and Markov transition matrix P conditional on the data y and the parameters $\theta = (\alpha, \beta)$ given by equation 6. An important part of this is to find a good estimate of θ. Note that θ contains two very different sets of parameters; α contains the hyperparameters of the prior distribution for the Markov transition probabilities, β contains parameters of the measurement process. The marginal likelihood of θ equals;

$$l(\theta; y) = \sum_x \int_p f(y|x;\beta) f(x|p) f(p|\alpha)\ dp \tag{7}$$

which requires the sum over d^T possible Markov chains x, and hence seems unavailable, making the maximum likelihood estimate of θ impossible to find directly. Even numerical optimization is unfeasible, since the function to be optimized cannot be written out in an analytical form.

This original likelihood problem is split into two simpler problems below, and the two steps are studied separately. An iterative scheme alternating between the steps is proposed. In the first step the hidden Markov chain $X = x$ is fixed. Doing so, the marginal likelihood with respect to the Dirichlet hyperparameter α can be maximized. In the second step the uncertainty in the Markov transition matrix P is ignored, and $P = p$ is fixed. Then the marginal likelihood with respect to the measurement parameters β can be maximized. The two steps are iterated.

3.1 Estimation of the Dirichlet hyperparameters

Consider first the case with categorical Markov chain values $X = x$ fixed as data instead of the observed values y, see e.g. Maritz (1989) and MacKay and Peto (1995). Categorical values x can be regared as latent variables, missing data, and in this first part they are assumed known. The data y and the measurement parameters β in Figure 3 are hence ignored. The marginal likelihood is then available and equals;

$$l(\alpha; x) = \int_p f(x|p) f(p|\alpha)\ dp = \prod_{i=1}^{d} \frac{\Gamma(\alpha_i)}{\prod_{j=1}^{d} \Gamma(\alpha_{ij})} \frac{\prod_{j=1}^{d} \Gamma(n_{ij} + \alpha_{ij})}{\Gamma(n_i + \alpha_i)} \tag{8}$$

where $n = \{n_{ij}; i = 1, \ldots, d, j = 1, \ldots, d\}$ are transition frequencies of the Markov chain $x = (x_1, \ldots, x_T)$, and $n_i = \sum_{j=1}^{d} n_{ij}$ is the number of transitions from state $i; i = 1, \ldots, d$. One property of the Γ function is that $\Gamma(m+1) = m\Gamma(m)$. Using this the log likelihood equals:

$$\begin{aligned} L(\alpha; x) &= log(l(\alpha; x)) \qquad (9) \\ &= \sum_{i=1}^{d}\left(\sum_{j=1}^{d}\sum_{k=1}^{n_{ij}} log(\alpha_{ij} + n_{ij} - k) - \sum_{k=1}^{n_i} log(\alpha_i + n_i - k)\right) \end{aligned}$$

Setting the derivatives of the log likelihood function in equation 9 equal to 0 gives the maximum marginal likelihood equations. For all $j = 1, \ldots, d$ and $i = 1, \ldots, d$:

$$\sum_{k=1}^{n_{ij}} \frac{1}{\alpha_{ij} + n_{ij} - k} = \sum_{k=1}^{n_i} \frac{1}{\alpha_i + n_i - k} \qquad (10)$$

With the constraints $\alpha_i = \sum_{j=1}^{d} \alpha_{ij} = \kappa$ given after equation 3, equation 10 gives a unique marginal likelihood solution for α.

If $\alpha_{ij} + n_{ij}$ is large an approximation to equation 10 is given by:

$$\sum_{k=a}^{b} \frac{1}{k} \approx \int_{a-\frac{1}{2}}^{b+\frac{1}{2}} \frac{1}{x} dx = log\left(\frac{b+\frac{1}{2}}{a-\frac{1}{2}}\right) \qquad (11)$$

The likelihood equations in equation 10 are hence approximated by, for $i, j = 1, \ldots, d$;

$$\frac{\alpha_{ij} + n_{ij} - \frac{1}{2}}{\alpha_{ij} - \frac{1}{2}} = \frac{\alpha_i + n_i - \frac{1}{2}}{\alpha_i - \frac{1}{2}} \qquad (12)$$

Or, for $i = 1, \ldots, d$:

$$\begin{pmatrix} n_i - n_{i1} & \ldots & -n_{i1} \\ \ldots & \ldots & \ldots \\ 1 & \ldots & 1 \end{pmatrix} \begin{pmatrix} \alpha_{i1} \\ \ldots \\ \alpha_{id} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} n_i - n_{i1} \\ \ldots \\ 2\kappa \end{pmatrix} \qquad (13)$$

where the constraint $\alpha_i = \sum_{j=1}^{d} \alpha_{ij} = \kappa$ has been added. Updating the α estimate by the linear system in equation 13 is much faster than the numerical optimization required in equation 10.

3.2 Estimation of measurement parameters, uncorrelated noise

Consider next the case with Markov transition matrix $P = p$ fixed, see e.g. Baum et al. (1970) and Leroux and Puterman (1992). The hyperparameters

α in Figure 3 is hence ignored. The marginal likelihood of the data y is analytically available in this case and equals;

$$l(\beta; y, p) = \prod_{t=1}^{T} f(y_t|y_{t-1}, \ldots, y_1, p) \tag{14}$$

Assuming that the measurements are conditionally independent given x;

$$f(y|x) = \prod_{t=1}^{T} f(y_t|x_t) \tag{15}$$

the terms in equation 14 are given by;

$$f(y_t|y_{t-1}, \ldots, y_1, p) = \sum_{k=1}^{d} \phi(y_t; \mu_{t,k}, \Sigma_k) Pr(X_t = k|y_{t-1}, \ldots, y_1, \beta, p) \tag{16}$$

where $Pr(X_t = k|y_{t-1}, \ldots, y_1, \beta, p)$ can be calculated recursively through forward(prediction and filtering) equations;

$$Pr(X_t = k|y_{t-1}, \ldots, y_1, \beta, p) = \sum_{j=1}^{d} P_{jk} Pr(X_{t-1} = j|y_{t-1}, \ldots, y_1, \beta, p) \tag{17}$$

$$\begin{aligned} & Pr(X_{t-1} = j|y_{t-1}, \ldots, y_1, \beta, p) \\ = & \frac{\phi(y_{t-1}; \mu_{t,j}, \Sigma_j) Pr(X_{t-1} = j|y_{t-2}, \ldots, y_1, \beta, p)}{\sum_{j=1}^{d} \phi(y_{t-1}; \mu_{t,j}, \Sigma_j) Pr(X_{t-1} = j|y_{t-2}, \ldots, y_1, \beta, p)} \end{aligned} \tag{18}$$

with initial values $Pr(X_1 = k|y_0, \beta, p) = P_{x_0 k}; k = 1, \ldots, d$. The weights in equation 16, $Pr(X_t = k|y_{t-1}, \ldots, y_1, \beta, p)$, are in this way a function of both the fixed Markov transition matrix p and of the measurement parameters β. Equation 14 can be optimized by inserting p and the current estimate of β to obtain the weights, and next maximizing the likelihood function in equation 14 numerically to get an updated estimate of β, keeping the weights fixed.

An alternative approach, see e.g. Leroux and Puterman (1992), is to impute $\Delta_{t,k} = Pr(X_t = k|y_T, \ldots, y_1, \beta, p)$ as an estimate of the indicator $I(X_t = k)$ at each time point, instead of $Pr(X_t = k|y_{t-1}, \ldots, y_1, \beta, p)$ in equation 14 and 16. $\Delta_{t,k}$ is the best estimate of the categorical value based on data y, the fixed p and the current estimate of β. This gives the following marginal likelihood for β:

$$l(\beta; y, p) = \prod_{t=1}^{T} \Delta_{t,k} \phi(y_t; \mu_{t,k}, \Sigma_k) \qquad k = 1, \ldots, d \tag{19}$$

$\Delta_{t,k}$ is calculated through forward and backward(smoothing) equations similar to the ones in equation 17 and 18, see Divijver (1985). Maximization of equation 19 with respect to β gives the following updated marginal likelihood estimates for β:

$$\eta_k^\gamma = \frac{\sum_{t=1}^T \Delta_{t,k} y_t^1}{\sum_{t=1}^T \Delta_{t,k}} \qquad k = 1, \ldots, d \tag{20}$$

$$\eta_k^\rho = (\sum_{t=1}^T \Delta_{t,k} h_t^\rho h_t^{\rho\prime})^{-1} \sum_{t=1}^T \Delta_{t,k} h_t^\rho y_t^2 \qquad k = 1, \ldots, d$$

$$\eta_k^m = (\sum_{t=1}^T \Delta_{t,k} h_t^m h_t^{m\prime})^{-1} \sum_{t=1}^T \Delta_{t,k} h_t^m y_t^3 \qquad k = 1, \ldots, d$$

$$\Sigma_k = \frac{\sum_{t=1}^T \Delta_{t,k}(y_t - \mu_{t,k})(y_t - \mu_{t,k})'}{\sum_{t=1}^T \Delta_{t,k}} \qquad k = 1, \ldots, d \tag{21}$$

The advantage of imputing $\Delta_{t,k}$ as in equation 19 compared to imputing $Pr(X_t|y_{t-1}, \ldots, y_1, \beta, p)$ in equation 14 and 16 is that the estimates given by equation 20 and 21 are analytically available, and hence much faster to calculate.

3.3 Estimation of measurement parameters, correlated noise

Consider again the case with Markov transition matrix $P = p$ fixed, but assume now that the measurements are no longer conditionally independent. A standard way to accomodate AR(1) measurement noise, as given in equation 5, into the model is to work with an augmented state of the underlying Markov chain values $G_t = (X_t, X_{t-1})$, see Schweppe (1973) and Brockwell and Davis (1995). A new transformed measurement can be defined as:

$$[z_t|g_t] = y_t - \Psi y_{t-1} = \mu_{t,x_t} - \mu_{t-1,x_{t-1}} + \epsilon_{g_t} \qquad t = 1, \ldots, T \tag{22}$$

and $z = \{z_t; t = 1, \ldots, T\}$ are conditionally independent given the augmented states $G = (G_1, \ldots, G_T)$. Conditional independence is convenient in the imputation step and the Gibbs sampler. The G_t value is of dimension d^2, and the value $\Delta_{t,k,l} = Pr(G_t = (k, l)|z_T, \ldots, z_1, \beta, p)$ is imputed as an estimate of $I(G_t = (k, l)) = I(X_t = l, X_{t-1} = k)$ in the estimation scheme. An augmented version of the Markov transition matrix, P_{aug}, of size $d^2 \times d^2$ can be defined from P as follows:

$$P^{aug}_{(i,j),(k,l)} = \begin{cases} p_{kl} & \text{if } k = j \\ 0 & \text{else} \end{cases} \tag{23}$$

$\Delta_{t,k,l}$ is hence calculated from P^{aug} and $\phi(z; \mu_{t,k} - \psi\mu_{t-1,l}, \Sigma_k)$ directly through forward and backward equations like in equation 17 and 18, but now in d^2 dimensions. $Pr(X_t = l|z, p, \beta) = \sum_{l=1}^{d} Pr(X_t = l, X_{t-1} = k|z, p, \beta)$ is the marginal probability $\Delta_{t,k}$. Estimates of the $\mu_{t,k}$ related parameters from equation 20 are still unbiased with this extension. Σ_l and ψ values are estimated simultanously:

$$\Sigma_l = \frac{\sum_{t=1}^{T}\sum_{k=1}^{d} \Delta_{t,l,k}(z_t - (\mu_{t,l} - \Psi\mu_{t-1,k}))(z_t - (\mu_{t,l} - \Psi\mu_{t-1,k}))'}{\sum_{t=1}^{T}\sum_{k=1}^{d} \Delta_{t,k,l}} \qquad l = 1, \ldots, d \tag{24}$$

$$\begin{aligned} &\sum_{t=1}^{T}\sum_{k=1}^{d}\sum_{l=1}^{d} \Delta_{t,k,l}\Sigma_l^{-1}\Psi(y_{t-1} - \mu_{t-1,k})(y_{t-1} - \mu_{t-1,k})' \\ = &\sum_{t=1}^{T}\sum_{k=1}^{d}\sum_{l=1}^{d} \Delta_{t,k,l}\Sigma_l^{-1}(y_t - \mu_{t,l})(y_{t-1} - \mu_{t-1,k})' \end{aligned} \tag{25}$$

3.4 Iterative marginal likelihood maximization

Note that equation 10 is independent of the measurement parameters β, and similarly that equation 19 is independent of the Dirichlet hyperparameters α. This separates the original full optimization in equation 7 into two simpler problems that can be solved. An iterative way to maximize the original likelihood in equation 7 is to alternate between updating α by equation 10 and updating β by equation 20 and 21 (uncorrelated measurement noise) or equation 20, 24 and 25 (correlated measurement noise). The maximizations are initiated by imputing values for the rock type sequence x and the Markov transition matrix p. Imputed values of x and p are obtained from Gibbs sampling, see e.g. Robert et al. (1993) and Ryden and Titterington (1998). The Gibbs sampling algorithm keeps parameters α and β fixed at the current estimates, and samples iteratively from the full conditional distributions of P and X, $f(p|x, \alpha)$ and $f(x|y, p, \beta)$, respectively. For uncorrelated noise these full conditionals are given by:

$$f(\underline{p}_i|x, \alpha) = Dirichlet(\alpha_{i1} + n_{i1}, \ldots, \alpha_{id} + n_{id}) \qquad i = 1, \ldots, d \tag{26}$$

$$Pr(X_t = k|x_t^c, y, p, \beta) = C\phi(y_t; \mu_{t,k}, \Sigma_k)P_{kx_{t+1}}P_{x_{t-1}k} \qquad t = 1, \ldots, T \tag{27}$$

where $x_t^c = (x_1, \ldots, x_{t-1}, x_{t+1}, \ldots, x_T)$ and C a normalizing constant. With correlated measurement noise the full conditoinal for P remains the same as in equation 26, whereas X is sampled from;

$$
\begin{aligned}
Pr(X_t = k|z, x_t^c, \beta, p) &= CP_{kx_{t+1}}P_{x_{t-1}k}\phi(z_t; \mu_{t,k} - \psi\mu_{t-1,x_{t-1}}, \Sigma_k) \\
&\quad \phi(z_{t+1}; \mu_{t+1,x_{t+1}} - \psi\mu_{t,k}, \Sigma_{x_{t+1}}) \qquad t = 1, \ldots, T
\end{aligned}
\tag{28}
$$

where C again is a normalizing constant. The burn in time of the Gibbs sampler is short since the parameter values change only slightly from iteration to iteration of the estimation procedure. Algorithmically the methodology goes as follows:

ALGORITHM:

1. Choose starting values, $\theta^1 = (\alpha^1, \beta^1)$, $i = 1$
2. Iterate until convergence:
   - Fix (x, p) from the Gibbs sampler output with parameters β^i and α^i.
   - Calculate α^{i+1} from x using equation 10 or 13.
   - Impute $\Delta_{t,k}; t = 1, \ldots, T, k = 1, \ldots, d$ or $\Delta_{t,k,l}; t = 1, \ldots, T, k = 1, \ldots, d, l = 1, \ldots, d$ from p, y and β^i, and calculate β^{i+1} from y and the $\Delta_{t,k}$'s using equation 20 and 21 (uncorrelated noise) or from y and the $\Delta_{t,k,l}$'s using equation 20, 24 and 25 (correlated noise).
   - Set $i = i + 1$.

The procedure has a flavor of the Expectation Maximization(EM) algorithm to it, see Dempster et al. (1977), and to stochastic EM, see Wei and Tanner (1990) and Nielsen (2000). However, the method presented above works as close to the original marginal likelihood in equation 7 as possible. Imputing categorical rock types x directly to update the measurement parameter β estimates is faster than the above approach, but one level down from the marginal likelihood in equation 14 and 16 with x summed out. Similarly, imputing the Markov transition matrix p directly to update the α estimate is faster than the approach above, but one level down from the marginal likelihood in equation 8 where p is integrated out.

$\theta^1, \theta^2, \ldots$ is a chain of outputs from the algorithm above. With the similarity of EM and stochastic EM noted, the chain should converge to a distribution around the maximum marginal likelihood estimate from equation

7. No proof of this convergence is provided here, but plots of evidence are presented with the example below. A rigorous proof should probably follow the theory and methods on stochastic EM or EM, see e.g. Dempster et al. (1977), Ip and Diebolt (1996) and Nielsen (2000).

The original EM idea is to impute the conditional expectation or MAP values of (x, p). In stochastic EM it is also common to impute a sample of (x, p), see Ip and Diebolt (1996). When a Gibbs sampler is used, several samples are available for a small amount of extra computational time, and hence the MAP estimate or the expectation probably works better.

The algorithm above is iterated until the marginal likelihood in equation 7 converges. Estimates of the marginal likelihood are obtained from the Gibbs output. Let $(x^1, p^1), \ldots, (x^N, p^N)$ be realizations from the Gibbs sampler, and $(x^\star, p^\star)$ a fixed point. The Gibbs stopper, see Yu and Tanner (1999), estimates the posterior density at $(x^\star, p^\star)$ by:

$$\hat{f}(x^\star, p^\star | y, \theta) = \frac{1}{N} \sum_{i=1}^{N} K((x^\star, p^\star)|(x^i, p^i)) \tag{29}$$

where $K((x^\star, p^\star)|(x^i, p^i))$ is the transition density of the Gibbs sampler defined in equation 27 and 26 from state (x^i, p^i) to state $(x^\star, p^\star)$;

$$K((x^\star, p^\star)|(x^i, p^i)) = f(p^\star | x^i, \alpha) f(x^\star | y, p^\star, \beta) \tag{30}$$

The estimate of the posterior at the point $(x^\star, p^\star)$ is then used to estimate the marginal likelihood in equation 7 by the following formula:

$$\hat{l}(\theta; y) = \frac{f(y|x^\star, \beta) f(x^\star|p^\star) f(p^\star|\alpha)}{\hat{f}(x^\star, p^\star|y, \theta)} \tag{31}$$

The chosen point $(x^\star, p^\star)$ does not really matter for the marginal likelihood calculation since it cancels, but preferable a point with large probability mass is chosen, for example the MAP value or results from an initial classification of the facies values. For more on marginal likelihood estimation from the Gibbs or Metropolis-Hastings output, see e.g. Chib (1995), Chib (2001) and Han and Carlin (2001).

Starting values θ^1 are provided from prior knowledge. In some cases an initial classification of the rock types X exists from earlier studies on the same data set. Such initial results are maybe not taking any sort of dependency into account, but nevertheless summarize the experience given in the field and on the particular case study. α^1 is then calculated directly from equation 10 with X equal to the initial classification. β^1 is calculated from equation 20 and 21 with $\Delta_{t,k} = I(X_t = k)$ in this initial phase (uncorrelated measurement noise), or from equation 20, 24 and 25 with $\Delta_{t,k,l} = I(G_t = (k,l))$ in this initial phase (correlated measurement noise) .

3.5 Exploring the posterior distributions of (P, X)

Conditional on the well log data y and fixed parameters from the estimation scheme, the posterior distribution of P and X is explored by Gibbs sampling. (In the example below we propose a parametric bootstrap strategy to integrate parameter uncertainty in the posterior distribution for (P, X), instead of using fixed values of θ.) Markov transition matrix P and facies sequence values X are sampled from the full conditional distributions in equation 26 and 28, respectively. As a fixed, best estimate of θ, we can use the mean of the θ values after convergence (n_0 iterations) of the iterative marginal likelihood estimates;

$$\hat{\theta} = \frac{1}{n} \sum_{i=n_0+1}^{n_0+n} \theta^i \qquad (32)$$

where $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$.

Samples from the Gibbs sampler, $(P^1, X^1), \ldots, (P^N, X^N)$, are approximately from the conditional distribution in equation 6. Associated samples from the posterior distribution of functionals defined from the Markov transition matrix P are easily generated. For each sample P^i, $i = 1, \ldots, N$, an associated sample from the posterior distribution of the stationary distribution or relative proportions, $\pi = (\pi_1, \ldots, \pi_d)$ is given by the relation;

$$\pi = \pi P \qquad (33)$$

Similarly a sample from the posterior distribution of the sojourn times or mean thicknesses, $u = (u_1, \ldots, u_d)$, in the different states is given by the relation;

$$u_i = \frac{\delta}{1 - P_{ii}} \qquad i = 1, \ldots, d \qquad (34)$$

where δ is the spacing in the well log measurements. Samples from the posterior distrbution of standardized jump probabilities, q_{ij} are given by the relation;

$$q_{ij} = \frac{p_{ij}/\pi_j}{\sum_{k \neq i} p_{ik}/\pi_k} \qquad i = 1, \ldots, d \quad j \neq i \qquad (35)$$

Hence, samples from the posterior distribution of P automatically give samples from the posterior distribution of the stationary distribution, sojourn times and normalized jump probabilities.

4. RESULTS

Trivariate observations of gamma ray (radioactivity), density and M modulus (sonic velocity) at every 15cm along the well are plotted in Figure 2a). Two stratigraphic sequences, $s1$ and $s2$, are assumed to represent different depositional environments, and are hence studied independently. The boundary between the two strata is well defined from seismic data and considered fixed in this study. Avseth et al. (2001) studied the data from a rock physics perspective, and calculated relative rock type proportions in the well based on classified facies logs. This study is an extension of Avseth et al. (2001) in which we attempt to explore the posterior distribution for Markov transition matrix P and rock type sequence X conditional on the available data as outlined in section 3. Facies is split into three categories ($d = 3$); (1)sand, (2)mixed; sandy shales and shaly sands, and (3)shale, and focus is on Markov type alternations between the three facies within both strata. Statistical estimates of Markov transitions for rock types in the well, with associated uncertainty could give valuable inputs to geologists.

Autocorrelated measurement noise inputs are modeled as in equation 5. The different data sources, gamma ray, density and M modulus, were found to be uncorrelated, and only the diagonal elements of covariance matrices $\Sigma_1, \ldots, \Sigma_d$ and autocorrelation matrix Ψ were modeled. The well log measurements are of length $T = 590$ in stratum $s1$ and $T = 720$ in stratum $s2$. An initial facies classification exists from Avseth et al. (2001), and this is used to find starting values for the parameters α and β as described in section 3.4, and provides reference variables $(x^{\star}, p^{\star})$ used in the marginal likelihood estimation. κ from equation 13 is fixed at 100 for both strata. The results did not seem sensitive to the choice of κ. In the imputation part of the algorithm, a Gibbs sampler as in equation 26 and 28 was run for 200 iterations. The MAP estimate of (x, p) from the last 100 iterations was picked, and imputed in the estimation algorithm.

Figure 4a) shows the scaled marginal likelihood estimate as a function of iteration number for stratigraphic sequence $s2$. It looks like the marginal likelihood estimate converges after $10-15$ iterations and shows random fluctuations after that. Figure 4b)-f) shows some parameter values for stratum $s2$ as a function of iteration number. These plots show similar convergence and fluctuations as in the likelihood plot in Figure 4a).

Figure 5a)-c) shows the final mean value estimates $\mu_{t,k}$ for $k=1,2,3$ in the Gaussian distribution on a regression form plotted versus depth (dashed) as given in equation 4. Also plotted in Figure 5 are the observed measurements (solid). This is plotted for both strata, and there are jumps at the sequence boundary. The mean parameter values for various facies follow the data fairly well, and the levels are most separated for the gamma ray data. This indicates that gamma ray is the best classifier for facies. Figure 5d) shows the final iteration MAP value for facies in both strata as a function of depth. Shale tends to occur when gamma ray observations are large, but due to the Markov dependency, the facies values show a smoother nature.

Table 1 shows the final iteration MAP solution of the Markov transition matrix for both strata. This is displayed with the sojourn times, u from equation 34, on the diagonal, and the standardized jump probabilities, q_{ij} from equation 35, on the off diagonal. Stratum $s1$ has significantly thicker shale beds and thinner sand beds than stratum $s2$ as seen from the diagonal elements. Further, the jump probabilities in stratum $s1$ appears to be more uniform than in stratum $s2$. In stratum $s2$ the rock type sequence almost always goes through the mixed state, and rarely jumps directly from shale to sand or sand to shale. Stratum $s2$ hence seems more structured than stratum $s1$.

In order to validate the method and to study the parameter uncertainty, we do a parametric bootstrap study. Given the parameter estimates from equation 32, a resampling strategy is carried out based on the hierarchical model in Figure 3. This is done for both strata. Algorithmically;

- For $i=1,\dots,B$:
    1. Draw Markov transition matrix P^{*i} conditional on the Dirichlet distribution specified by parameter $\hat{\theta}$.
    2. Draw a Markov chain of rock types $X^{*i}=(X_1^{*i},\dots,X_T^{*i})$, conditional on the Markov transition matrix P^{*i}.
    3. Draw data y^{*i}, with $y_t^{*i}=(\gamma_t^{*i},\rho_t^{*i},m_t^{*i})$, $t=1,\dots,T$, conditional on the Markov chain X^{*i} and the measurement parameters $\hat{\beta}$.

4. Maximize the marginal likelihood of y^{*i} to obtain $\hat{\theta}^{*i}$.

Bootstrap samples $\hat{\theta}^{*1}, \ldots, \hat{\theta}^{*B}$ represent the distribution of the original marginal likeihood estimates $\hat{\theta}$. In this study we used $B = 25$. Figure 6 displays the bootstrap samples of the α parameter. Samples from both sequences $s1$ (Figure 6a)-c)) and $s2$ (Figure 6d)-f)) show a small variability, and indicate that the diagonal elements of the Markov chain are most likely. Values are clustered close to the original α estimates and this gives evidence to the estimation approach. Table 2 shows the original β estimates obtained by maximizing the marginal likelihood of data y, with the first quartile, the median and the third quartile estimated from the 25 bootstrap samples from the above scheme. The variability seems small for the η^γ parameters, larger for the η^ρ and small for the η^m parameters. For all these parameters, however, the original estimates fall within or just outside the first and third quartiles. For the standard deviation parameters (σ), the bootstrap samples appear to spread out at larger values than the parameters estimated from the well log data. The original estimate from y fall below the first quartile more often than one would expect. Maybe there is significant skewness in the probability distributions for these parameters, or maybe there is a bias in the maximum marginal likelihood estimates that should have been corrected for. It could also be the result of slower convergence of the estimation approach for the variance parameters. The autocorrelation parameter ψ appears to be distributed around the original value, maybe somewhat in the upper tails of the distribution.

In the following we attempt to include the distributional properties of the parameter values when exploring the posterior distribution of Markov transition matrix P and rock type sequence X. For each of the θ^{*i}, $i = 1, \ldots, B$ values, (P, X) are sampled from the conditional distribution given the well log measurements y using Gibbs sampling as described in section 3. Each Gibbs sampler is run for 200 iterations, and every 10th sample from the last 100 iterations are picked out. This gives 10×25 samples ($B = 25$) from the posterior distribution of (P, X). The Gibbs output represents the posterior distribution of P and X conditional on data y, and functionals of P like those described in section 3.5 are particularly interesting in this study. In Figure 7 the stationary distribution $\pi = (\pi_1, \pi_2, \pi_3)$ from equation 33 is plotted for stratum $s1$(left) and stratum $s2$(right) in a ternary plot. Stratigraphic sequence $s1$ contains more shales and less sand and mixed than sequence $s2$. Figure 8 shows estimated probability densities for sojourn times

$u = (u_1, u_2, u_3)$ from equation 34 in states sand(Figure 8a)), mixed(Figure 8b)) and shale(Figure 8c)) for stratum $s1$(solid) and stratum $s2$(dashed). The sojourn times in sand are smaller and the sojourn times in shale larger for stratum $s1$. Figure 9 shows samples of the standardized jump probabilities from mixed to shale given from equation 35 for both strata. The samples are sorted before plotting, and presented as a quantile-quantile plot. Values from sequence $s1$ are on the first(x) axis and values from sequence $s2$ are on the second(y) axis. The jump probabilities are larger for stratum $s2$. Figure 9 also indicates a larger variability in the jump probabilities for sequence $s1$. Samples of standardized jump probabilities in Figure 9 go from approximately $0.2 - 0.7$ for strata $s1$, and only from approximately $0.5 - 0.8$ for strata $s2$.

The main result is that stratum $s1$ has more shales and less sand than stratum $s2$, and that the thicknesses of shale beds are thinner in stratigraphic sequence $s2$. In other words $s1$ has more fine grained sediments (shales) and less coarse grained sediments (sands) than $s2$. Results also indicate that sequence $s1$ is less structured than sequence $s2$. The MAP estimate in Table 1 shows that jumps from shale to sand and sand to shale are extremely rare in stratum 2, the rock type sequence most commonly goes through the mixed state. Also, in Figure 9 the standardized jump probabilities from mixed to shale, q_{23} are larger and show less variability for stratum $s2$. This jump probability indicates a trend of going from coarser to finer material (fining upwards) and is often recognized in well organized deposits.

Sequence stratigraphy is commonly used to limit zones of sea level low or highstands, see Miall (2000). In this dataset $s1$ is believed to represent a relatively high sea level, whereas sequence $s2$ represents a relative lowstand in sea level. Deposits in sequence $s2$ are because of this expected to be richer in coarse grained materials like sand, and also to be more organized in its rock type alternations. The physical explanation for this is based on fluid flow dynamics, and that flow is more structured in shallow waters, since it is closer to the shoreline. This explains the above results physically, but this study further quantifies this by the Markov transition matrix probability distributions. The Markov transition matrix can give geologists a quantitative indicator of burial trends in various depositisional systems.

5. CONCLUDING REMARKS

This paper attempts to quantify geological attributes, with associated uncertainty levels, from indirect well log measurements. Well logs are measurements of radioactivity, density and sonic velocity, and these are informative regaring rock type. Seismic imaging of the earth are used to split the domain of interest into two strata that are believed to represent different depositional environments. The two stratigraphic sequences are studied separately and compared. We model the rock sequence along the well and the rock type alternations in a hierarchical Markov chain setting. The rock types are modeled as a hidden Markov chain, with the indirect measurements distributed as Gaussian variables with parameters depending on the rock type. Autocorrelated measurement noise is possible in the model. We explore the posterior distribution of the underlying rock types and the governing Markov transition matrix conditional on the well log data. Within the hierarchical framework a methodology is outlined for estimating hyperparameters and measurement parameters by iterative marginal likelihood maximization. Gibbs sampling and imputation are important parts of the iterative scheme. Bootstrap methods are used to validate the method and to integrate the uncertainty of the estimated parameters in the posterior distribution for Markov transition probabilities and rock types.

Results show that the two strata are different in rock distributions and rock alternations, and we explain how this could be related to sea level changes in ancient times like indicated by e.g., Miall (2000). Geologists have studied trends in sediment burial and diagenesis (processes of chemical, pressure and temperature reactions) qualitatively. The Markov transition matrix is a quantitative tool to better understand and analyze these different trends. Distributional aspects of the Markov chain goes beyond modeling of just facies proportions. Functionals of the transition matrix, such as sojourn times and standardized jump probabilities, with associated uncertainty levels, give more insight to the phenomenon. The hidden Markov model can also be suitable for generating synthetic well logs, consistent with the geologic environment, which can again be used as an input to study the statistical properties of seismic signatures away from wells.

The Markov model is certainly not the only model for continuous well log measurements, and others may be more valuable. Extensions to the approach include for example more refined measurement models or more complex models for deposition and burial of rocks, possibly taking more geological knowledge into account. Regarding alternative or extended estimation and sampling schemes could also be of interest.

REFERENCES

Avseth,P., Mukerji,T., Mavko,G., and Dvorkin,J., 2001, "Rock physics and seismic interpretation of depositional trends and sequence stratigraphy - a North Sea example," Submitted for publication.
http://srb.stanford.edu/Theses/Avseth_77.pdf

Baum,L.E., Petrie,T., Soules,G., and Weiss,N., 1970, "A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains," The Annals of Mathematical Statistics, 164-171.

Brockwell,B. and Davis,J., 1995, "Time series: Theory and Methods," Wiley.

Chen,C., and Hiscott,R.N., 1999, "Statistical analysis of facies clustering in submarine-fan turbidite successions," Journal of Sedimentary Research, 505-517.

Chib,S., 1995, "Marginal likelihood from the Gibbs output," Journal of the American Statistical Association, 1313-1321.

Chib,S., 2001, "Marginal likelihood from the Metropolis-Hastings output," Journal of the American Statistical Association, 270-281.

Dempster,A., Laird,P.B., and Rubin,J., 1977, "Maximum Likelihood from Incomplete Data via the EM algorithm," Journal of Royal Statistical Society, 1-21.

Deshpande,A., Flemings,P.B., and Huang,J., 1997, "Quantifying lateral heterogenities in fluvio-deltaic sediments using three-dimensional reflection seismic data: offshore Gulf of Mexico," Journal of Geophysical Research, 15385-15401.

Divijver,V., 1985, "On Baum's forward and backward equations," Pattern Recognition, 369-274.

Han,C., and Carlin,B.P., 2001, "Markov Chain Monte Carlo Methods for computing Bayes Factors: A Comparative Review," Journal of American Statistical Association, 1122-1132.

Harbaugh,J.W., and Bonham-Carter,G., 1970, “Computer simulation in geology,” Wiley.

Ip,E.H.S. and Diebolt,J., 1996, In “Markov chain Monte Carlo in practice,” (eds) Gilks, Richardson and Spiegelhalter, Chapman Hall.

Leroux,B.G., and Puterman,M.L., 1992, “Maximum-Penalized-Likelihood Estimation for Independent and Markov-Dependent Mixture Models,” Biometrics, 545-558.

MacKay,N.C.C., and Peto,L., 1995, “Hierarchical Dirichlet modeling in language recognition,” Natural Language Modeling, 450-455.

Malinverno,A, 1997, “On the power law size distribution of turbidite beds,” Basin Research, 263-274.

Maritz,A, 1989, “Empirical Bayes estimation in 2×2 tables,” Communications in Statistics, Theory and Methods, 3215-3233.

Mavko,G., Mukerji,T., and Dvorkin,J., 1995, “The rock physics handbook,” Wiley.

Miall,A.D., 2000, “Principles of sedimentary basin analysis,” Springer Verlag.

Nielsen,S.F., 2000, “The stochastic EM algorithm: Estimation and asymptotic results,” Bernoulli, 457-465.

Robert,C.P., Celeux,G., and Diebolt,J., 1993, “Bayesian estimation of hidden Markov chains: A stochastic implementation,” Statistics and Probability Letters, 77-83.

Ryden,T., and Titterington,D.M., 1998, “Computational Bayesian analysis of hidden Markov models,” Journal of Computational and Graphical Statistics, 194-211.

Schweppe,F.C., 1973, “Uncertain dynamic systems,” Prentice Hall.

Wei,G.C.G., and Tanner,M.A., 1990, "A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms," Journal of American Statistical Association, 699-704.

Weissman,G.S., and Fogg,G.E, 1999, "Multi-scale alluvial fan heterogeneity modeled with transition probability geostatistics in a sequence stratigraphic framework," Journal of Hydrology, 48-65.

Yu,J.Z., and Tanner,M.A., 1999, "An analytical study of several Markov chain Monte Carlo estimators of the marginal likelihood," Journal of Computational and Graphical Statistics, 839-853.

Table 1: MAP estimates of the transition probabilities for stratas $s1$ and $s2$. The diagonal elements are given as sojourn times in meters, the off diagonals are normalized jump probabilities.

	P^{s1}		
	sand	mixed	shale
sand	1.9m	0.94	0.06
mixed	0.39	0.5m	0.61
shale	0.62	0.38	2.8m
	P^{s2}		
	sand	mixed	shale
sand	2.7m	0.99	0.01
mixed	0.29	0.9m	0.71
shale	0.01	0.99	1.1m

Table 2: Parameter estimates from marginal likelihood maximization, $\hat{\theta}$, and bootstrap estimates of the parameter uncertainty given as 1st quartile, median and 3rd quartile estimates for both stratigraphic sequences. Facies 1 is sand, facies 2 is mixed and facies 3 is shale.

	s1:$\hat{\theta}$	s1:$\hat{\theta}_{.25}$	s1:$\hat{\theta}_{.5}$	s1:$\hat{\theta}_{.75}$	s2:$\hat{\theta}$	s2:$\hat{\theta}_{.25}$	s2:$\hat{\theta}_{.5}$	s2:$\hat{\theta}_{.75}$
Facies 1 η^γ	66	64.1	65.0	66.8	62	62.2	63.3	64.6
Facies 2 η^γ	72	67.7	68.4	71.2	70	69.9	70.1	70.2
Facies 3 η^γ	76	75.3	75.8	76.1	81	80.9	81.0	81.4
Facies 1 $\eta^\rho(1)$	-0.4	-137	49	276	0.46	-71	1.5	70
Facies 2 $\eta^\rho(1)$	-0.4	-98	-0.2	69	1.47	-5	6	14
Facies 3 $\eta^\rho(1)$	4.7	-45	21	96	-2.5	-18	5	33
Facies 1 $\eta^\rho(2)$	0.2	-94	34	190	0.8	-45	0	45
Facies 2 $\eta^\rho(2)$	0.3	-66	1.4	48	1.5	-3	4	10
Facies 3 $\eta^\rho(2)$	3.7	-29	15	66	-0.96	-11	4	21
Facies 1 η^m	69	62	68	78	86	79	87	96
Facies 2 η^m	79	70	76	83	83	83	84	87
Facies 3 η^m	62	59	66	71	82	80	82	84
Facies 1 σ^γ	1.7	1.4	1.5	1.9	1.6	2.3	2.5	2.9
Facies 2 σ^γ	3.1	2.5	3.3	3.7	3.0	2.6	3.1	3.3
Facies 3 σ^γ	3.3	3.6	3.9	4.2	4.6	4.7	5.6	5.9
Facies 1 σ^ρ	0.20	0.28	0.30	0.33	0.11	0.14	0.17	0.25
Facies 2 σ^ρ	0.10	0.14	0.20	0.26	0.13	0.24	0.33	0.40
Facies 3 σ^ρ	0.17	0.20	0.28	0.34	0.31	0.40	0.60	0.81
Facies 1 σ^m	0.63	0.89	0.95	1.04	0.36	0.31	0.44	0.54
Facies 2 σ^m	1.44	1.40	1.58	1.67	1.37	1.41	1.62	1.67
Facies 3 σ^m	0.71	0.89	0.95	1.00	0.83	1.14	1.26	1.40
ψ^γ	0.92	0.84	0.87	0.89	0.87	0.70	0.78	0.83
ψ^ρ	0.90	0.85	0.87	0.90	0.88	0.86	0.87	0.88
ψ^m	0.95	0.92	0.94	0.96	0.95	0.93	0.95	0.96

FIGURE CAPTIONS

Figure 1: Location of the Glitne Field in the North Sea.

Figure 2: a) Well observations of γ ray, density and M modulus, b) Seismic section from the Glitne field with a well indicated. Stratigraphic sequences $s1$ and $s2$ are shown with two deeper sequences.

Figure 3: Relations between variables, parameters and data. Observations y are defined from Markov chain X and measurement parameters β, X is defined from the Markov transition matrix P, and P is defined from the hyperparameters α.

Figure 4: Convergence of iterative estimation scheme. All attributes plotted as a function of iteration number. a) Estimate of the marginal likelihood. b) α_{11}. c) α_{12}. d) η^{γ}. e) Σ^{γ}. f) Ψ^{γ}. Convergence appears to take place after $10-20$ iterations. Fluctuations are due to the stochastic nature of the algorithm.

Figure 5: a)-c) Estimated mean values for the measurements in all three facies in both stratas(dashed), with the observed values(solid) as a function of depth. a) γ ray. b) density. c) velocity. d) MAP estimate for facies as a function of depth. Large values of γ ray result in shale, small values in sand.

Figure 6: Estimated variability of the α parameters. a) $\underline{\alpha}_1$ for $s1$. b) $\underline{\alpha}_2$ for $s1$. c) $\underline{\alpha}_3$ for $s1$. d) $\underline{\alpha}_1$ for $s2$. e) $\underline{\alpha}_2$ for $s2$. f) $\underline{\alpha}_3$ for $s2$. The variability appears small, and diagonal elements of α are large.

Figure 7: Ternary plot of stationary probabilities in the different facies classes for $s1$(left) and $s2$(right). Stratum $s1$ has more of shales(class 3) and less of mixed(class 2) and sand(class 1) than stratum $s2$.

Figure 8: Estimated probability densities for the sojourn times of $s1$(solid) and $s2$(dashed). a) sand. b) mixed. c) shale. Sojourn times in shale are larger for stratum $s1$. Sojourn times in sand are smaller for stratum $s1$.

Figure 9: Quantile-quantile plot of standardized jump probabilities from mixed to shale. Values for stratum $s1$ are plotted on the first axis, values for

stratum $s2$ on the second axis. Stratum $s1$ have smaller jump probabilities from mixed to shale, and a larger variability.

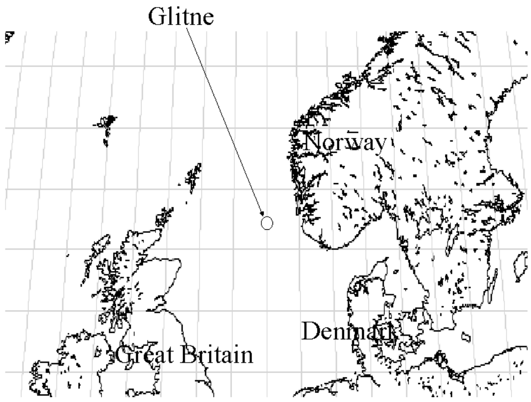

FIG. 1

Figure 1:

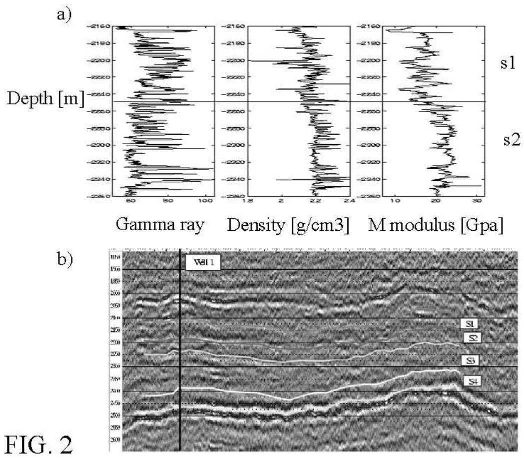

FIG. 2

Figure 2:

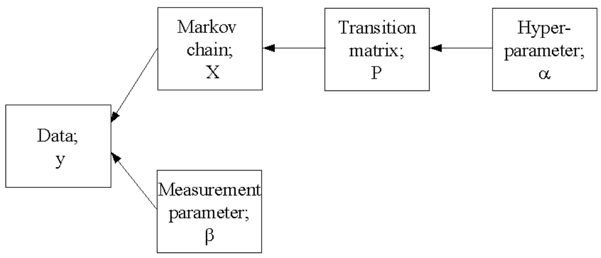

FIG. 3

Figure 3:

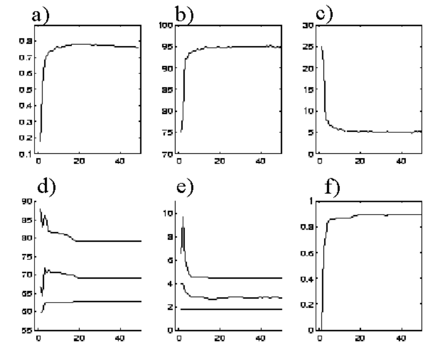

FIG. 4

Figure 4:

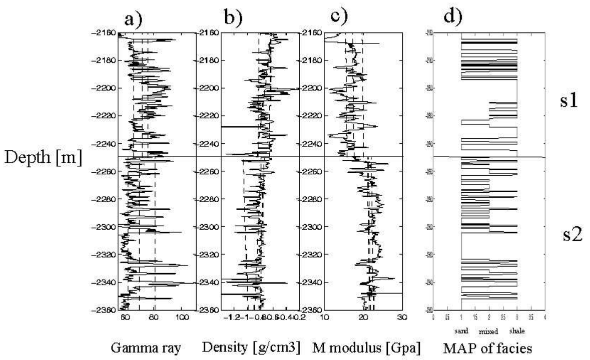

FIG. 5

Figure 5:

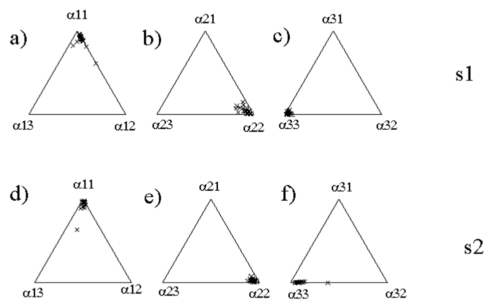

FIG. 6

Figure 6:

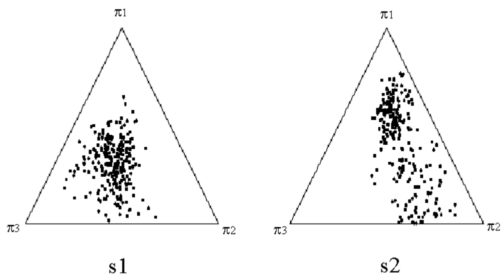

FIG. 7

Figure 7:

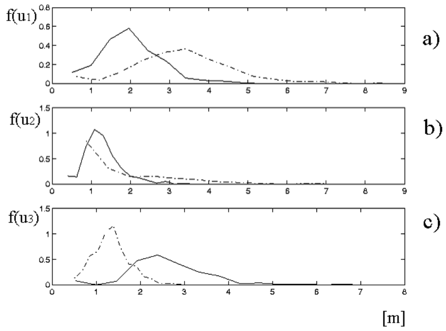

FIG. 8

Figure 8:

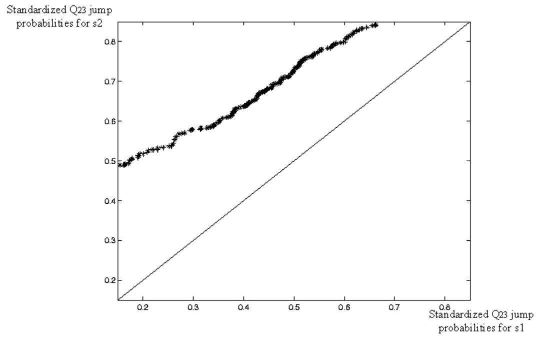

FIG. 9

Figure 9: