

Bayesian Inversion in Hidden Markov Models with Varying Marginal Proportions

Selamawit Serka Moja¹ · Zeytu Gashaw Asfaw¹ · Henning Omre²

Received: 17 April 2018 / Accepted: 29 May 2018 / Published online: 6 July 2018
© International Association for Mathematical Geosciences 2018

Abstract Knowledge of the sub-surface characteristics is crucial in many engineering activities. Sub-surface soil classes must, for example, be predicted from indirect measurements in narrow drill holes and geological experience. In this study, the inversion is made in a Bayesian framework by defining a hidden Markov chain. The likelihood model for the observations is assumed to be in factorial form. The new feature is the specification of the prior Markov model as containing vertical class proportion profiles and one reference class transition matrix. A criterion for selection of the associated non-stationary prior Markov model is introduced, and an algorithm for assessing the set of class transition matrices is defined. The methodology is demonstrated on one synthetic example and on one case study for offshore foundation of windmills. It is concluded that important experience from the geologist can be captured by the new prior model and that the associated posterior model is, therefore, improved.

Keywords Bayesian analysis · Non-stationary prior model · Offshore windmills · Sub-surface soil classes

1 Introduction

Collection of exact observations of the sub-surface is usually costly and time-consuming since it involves drilling of vertical wells from which samples can be obtained. Hence, indirect measurements about the sub-surface are frequently col-

✉ Selamawit Serka Moja
selamawitserka.get@gmail.com

¹ School of Mathematical and Statistical Sciences, Hawassa University, P.O. Box 05, Hawassa, Ethiopia

² Department of Mathematical Sciences, NTNU, 7491 Trondheim, Norway

lected, and predictions of the characteristics of interest are based on these indirect observations. The measurements can be collected by log-responses in narrow drill holes or by active geophysical procedures. In order to define a prediction rule for the sub-surface characteristics of interest based on the indirect measurements, both exact observations of the characteristic and indirect measurements are collected in sub-surface profiles in a nearby location. Only a small number of such calibration profiles are usually available due to the cost of obtaining them.

The objective of the study is to define a prediction rule for the facies characteristics along a vertical sub-surface profile in which one has indirect measurement in a narrow drill hole. Calibration data of both facies occurrences and indirect measurements are available in a small number of locations in the neighborhood of the narrow drill hole. The ultimate aim is to predict the facies distribution in a three-dimensional sub-surface volume based on data in a large number of drill holes and a small number of calibration wells. The results from the current study provides an important intermediate result on the path to obtain this three-dimensional sub-surface volume prediction.

Classification of soil classes in the sub-surface of an off-shore windmill park is used as a case study. The log-responses are resistance and friction of a bore head along the vertical profile, and the soil classification is traditionally made by using standardized charts (Robertson 2010). In the current study, a more formal statistical classification approach is used.

The inversion problem is cast in a Bayesian setting. The posterior probability distribution of the categorical facies variable is defined by a likelihood and a prior model. The likelihood model links the indirect measurements and the facies variable, while the prior model represents the experience with the facies variable. The current model, with a conditional independent, single-site response likelihood and Markov chain prior, falls in the class of hidden Markov models. This model class is extensively studied and frequently applied in statistics; see Dymarski (2011) for an overview. In subsurface modelling, facies/lithology characterization is often based on a hidden Markov model. The prior model for the categorical facies variable along a vertical profile is a Markov chain, in a long tradition following Krumbein and Dacey (1969) and Harbaugh and Bonham-Carter (1970). Traditionally, this Markov chain is assumed to be stationary and parametrized by one facies transition matrix, which usually is assessed by a counting estimator in one vertical well (Weissman and Fogg 1999). In Eidsvik et al. (2004) the assessment of the facies transition matrix is cast in a hierarchical Bayesian setting.

The use of a prior stationary Markov chain along the profile causes the prior marginal distributions of facies at all depths to be identical. For sub-surface phenomena this assumption often appears unrealistic since the facies occurrences vary with depth (Avseth et al. 2005). A quantitative assessment of the vertical facies proportions is discussed in Ravenne et al. (2002), but the inclusion of these proportion curves into a Markov model for facies classification remains a challenge, which is addressed in the current study. A prior Markov model that is non-stationary is defined, hence with facies transition matrices that vary with depth. A similar non-stationary Markov chain prior model is used in Ulvmoen et al. (2010), but the assessment of the model parameters are made very ad hoc.

This depth-dependence complicates estimation of the transition matrices. The problem is parametrized by specifying one transition reference-matrix being the transition

matrix in the best stationary model and geologist-specified facies proportion curves along the vertical profile. Based on this parametrization, a criterion for identifying the facies transition matrices for each depth is specified and an algorithm for assessing them is defined.

The likelihood model is defined as conditionally independent with single-site response. Hence, the posterior model appears as a non-stationary Markov chain, which can be exactly assessed by the recursive Forward–Backward and Viterbi algorithms (Viterbi 1967; Baum et al. 1970; Scott 2002).

Bold lower, \mathbf{x} , and upper, \mathbf{X} , case letters denote vectors and matrices, respectively. The vector \mathbf{x}_{-i} represents the vector \mathbf{x} with element i removed. Similarly, the set \mathcal{T} with element $\{i\}$ removed is denoted \mathcal{T}_{-i} . The term $p(\cdot)$ is used for probability density/mass function and the Gaussian n -vector \mathbf{x} has probability density function $\varphi_n(\mathbf{x}; \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ with expectation n -vector $\boldsymbol{\mu}_x$ and covariance $(n \times n)$ -matrix $\boldsymbol{\Sigma}_x$. The identity n -vector and $(n \times n)$ -matrix are denoted \mathbf{i}_n , and \mathbf{I}_n , respectively. Lastly, inequality operators between vectors or matrices entails element-wise inequality.

In Sect. 2, the problem definition and notation are discussed in more detail. In Sect. 3, the likelihood and two prior and posterior models are defined and also the depth-dependent transition matrices and an algorithm for assessing them are specified. In Sect. 4, the categorical Bayesian inverse method is tested on both synthetic and real data sets. Finally, in Sect. 5, conclusion and recommendation for further work are presented.

2 Problem Definition and Notation

The sub-surface characteristic of interest is represented along a vertical profile discretized to $\mathcal{T} : \{1, \dots, T\}$. At each depth $t \in \mathcal{T}$ the categorical facies class $\kappa_t \in \Omega_\kappa : \{1, \dots, K\}$ belong to one out of K classes, which might be un-ordered. The facies profile is represented by the vector, $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_T)'$, and assessing this vector is the objective of the study. Along the profile one real variable associated with the facies class is observed $\mathbf{d} = (d_1, \dots, d_T)'$, $d_t \in \mathcal{R}^1$. Hence, the objective of the study is to predict $\boldsymbol{\kappa}$ given \mathbf{d} , that is predict $[\boldsymbol{\kappa}|\mathbf{d}]$.

Prediction is approached in a probabilistic setting using Bayesian inversion

$$p(\boldsymbol{\kappa}|\mathbf{d}) = \text{const} \times p(\mathbf{d}|\boldsymbol{\kappa}) p(\boldsymbol{\kappa})$$

where the likelihood function $p(\mathbf{d}|\boldsymbol{\kappa})$ defines the link between the observations \mathbf{d} and variable of interest $\boldsymbol{\kappa}$, the prior probability density function (pdf) $p(\boldsymbol{\kappa})$ represents the general experience with the variable of interest $\boldsymbol{\kappa}$, while “const” is a normalizing constant. The posterior pdf $p(\boldsymbol{\kappa}|\mathbf{d})$ is uniquely defined by the likelihood function and the prior pdf, and it constitutes the ultimate solution to a Bayesian inversion problem.

The posterior pdf $p(\boldsymbol{\kappa}|\mathbf{d})$ must usually be explored by simulation, ideally based on a set of realizations

$$[\boldsymbol{\kappa}|\mathbf{d}]_1, \dots, [\boldsymbol{\kappa}|\mathbf{d}]_s \text{ iid } p(\boldsymbol{\kappa}|\mathbf{d}).$$

These realizations are most often generated approximately by Markov chain Monte Carlo (MCMC) simulation, while under very particular model assumptions the generation might be done exactly by repeating a sequential algorithm. In the current study, model assumptions, which make efficient sequential generation possible are defined.

The prediction of $[\kappa|\mathbf{d}]$ can be based on a maximum posterior (MAP) criterion since the sample space Ω_κ is categorical, hence the MAP predictor is

$$\hat{\kappa}_{MAP} = \text{MAP} \{ \kappa | \mathbf{d} \} = \arg \max_{\kappa} \{ p(\kappa | \mathbf{d}) \}.$$

Computing the MAP predictor will generally require complex discrete optimization in high dimension, while particular model assumptions might simplify the calculations. The model assumptions defined later make sequential optimization possible. Moreover, it is interesting to present posterior probability profiles for each class $\kappa \in \Omega_\kappa$, which is defined as

$$\mathbf{p}_\kappa = [p(\kappa_t = \kappa | \mathbf{d})]_{t \in \mathcal{T}}; \quad \kappa \in \Omega_\kappa.$$

From these profiles, the uncertainty associated with the MAP prediction $\hat{\kappa}_{MAP}$ can be assessed. Lastly, an alternative marginal maximum posterior (MMAP) predictor can be defined

$$\hat{\kappa}_{MMAP} = \text{MMAP} \{ \kappa | \mathbf{d} \} = \left[\arg \max_{\kappa} \{ p(\kappa_t = \kappa | \mathbf{d}) \} \right]_{t \in \mathcal{T}}.$$

3 Model Definition

The posterior pdf, which is the focus of this study, is defined by the likelihood function and prior pdf. The posterior also involves the normalizing constant; however, it is usually very complex to calculate. In this section, model assumptions are defined such that it is possible to assess the posterior pdf very efficiently.

3.1 Observation Likelihood

The likelihood function $p(\mathbf{d}|\kappa)$ defines the link between the observations, \mathbf{d} and the characteristic in focus, κ . In the function, \mathbf{d} is a fixed vector of observations while the vector of facies classes κ is the variable. If it is reasonable to assume that the likelihood can be factorized as follows

$$p(\mathbf{d}|\kappa) = \prod_{t \in \mathcal{T}} p(d_t|\kappa) = \prod_{t \in \mathcal{T}} p(d_t|\kappa_t) \quad (1)$$

the calculations of the posterior pdf might be dramatically simplified. The assumption entails conditional independence and single-site response in the observations. Observations with independent errors from one depth t to another cause the former, while responses without spatial convolutions cause the latter. These assumptions are often

justifiable at an initial stage of modeling, and they may be generalized at the expense of higher computational demands. The likelihood function is defined by the partial likelihood functions

$$p(d_t|\kappa_t); \kappa_t \in \Omega_\kappa ; t \in \mathcal{T}.$$

3.2 Prior Experience

The prior pdf $p(\kappa)$ represents the initial knowledge about the variable in focus κ . This sub-surface characteristic will normally be spatially coupled and represented by a Markov chain

$$p(\kappa) = p(\kappa_1) \prod_{t \in \mathcal{T}_{-1}} p(\kappa_t|\kappa_{t-1}, \dots, \kappa_1) = p(\kappa_1) \prod_{t \in \mathcal{T}_{-1}} p(\kappa_t|\kappa_{t-1}). \tag{2}$$

The first-order Markov assumption entails that the knowledge of the previous state screens out the influence of the history, causing the latter factorial expression. Note that this assumption entails that the spatial dependence is $p(\kappa_t|\kappa_{-t}) = p(\kappa_t|\kappa_{t-1}, \kappa_{t+1})$. These screening assumptions introduce geometrically decaying spatial dependence in the variable, and they may be generalized on the expense of higher computational demands.

The prior pdf is defined using the uni- and bi-variate properties, which define the initial pdf and the set of transition matrices

$$\begin{aligned} \mathbf{p}_1 &= [p(\kappa_1)]_{\kappa_1 \in \Omega_\kappa} \\ \mathbf{P}_{t-1, t} &= [p(\kappa_t|\kappa_{t-1})]_{\kappa_{t-1}, \kappa_t \in \Omega_\kappa}; t \in \mathcal{T}_{-1} \end{aligned} \tag{3}$$

which in turn defines the set of marginal pdfs by the recursion

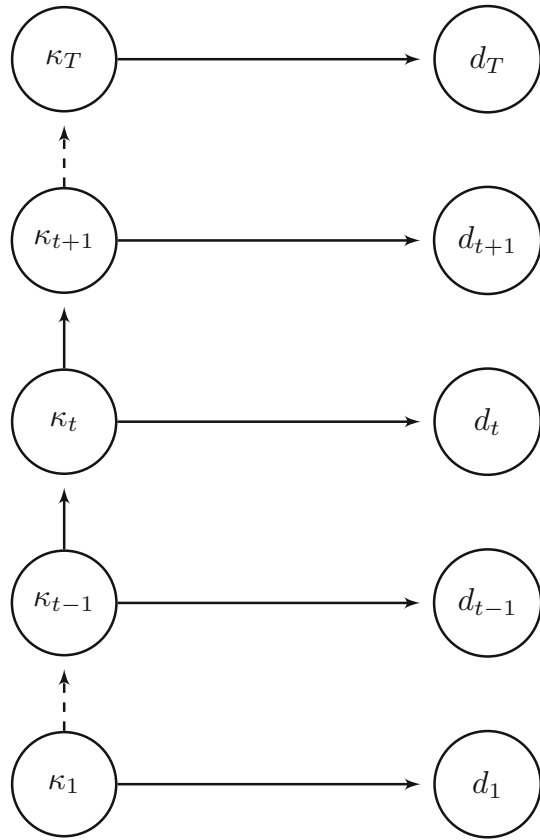
$$\mathbf{p}_t = [p(\kappa_t)]_{\kappa_t \in \Omega_\kappa} = \mathbf{P}'_{t-1, t} \mathbf{p}_{t-1}; t \in \mathcal{T}_{-1}. \tag{4}$$

Note that the set of marginal pdfs $\mathbf{p}_t; t \in \mathcal{T}$ does not uniquely define the prior pdf since many sets of transition matrices will provide identical sets of marginal pdfs.

3.3 Posterior Model

The posterior pdf is fully defined by the likelihood function and the prior pdf. The assumptions enforced in the two previous sub-sections entails that the model can be displayed as a directed graph (Fig. 1). The resulting posterior pdf can be expressed as

Fig. 1 Dependence structure of the hidden Markov model



$$\begin{aligned}
 p(\boldsymbol{\kappa}|\mathbf{d}) &= \text{const} \times \prod_{t \in \mathcal{T}} p(d_t|\kappa_t) \times p(\kappa_1) \prod_{t \in \mathcal{T}_{-1}} p(\kappa_t|\kappa_{t-1}) \\
 &= \text{const} \times p(d_1|\kappa_1) p(\kappa_1) \times \prod_{t \in \mathcal{T}_{-1}} p(d_t|\kappa_t) p(\kappa_t|\kappa_{t-1}) \quad (5) \\
 &= p(\kappa_1|\mathbf{d}) \prod_{t \in \mathcal{T}_{-1}} p(\kappa_t|\kappa_{t-1}, \mathbf{d}).
 \end{aligned}$$

Hence, the posterior pdf will also be a first-order Markov chain with initial pdf and transition matrices dependent on the observation vector \mathbf{d} ; see “Appendix A”. The corresponding spatial expression is $p(\kappa_t|\kappa_{-t}, \mathbf{d}) = p(\kappa_t|\kappa_{t-1}, \kappa_{t+1}, \mathbf{d})$.

The posterior pdf, under the assumptions specified, can be assessed by the highly efficient, recursive Forward–Backward algorithm, see “Appendix B”, and the solution is the initial posterior pdf and the set of posterior transition matrices (Baum et al. 1970; Scott 2002)

$$\begin{aligned} \mathbf{p}_{1|d} &= [p(\kappa_1|\mathbf{d})]_{\kappa_1 \in \Omega_\kappa} \\ \mathbf{P}_{t-1, t|d} &= [p(\kappa_t|\kappa_{t-1}, \mathbf{d})]_{\kappa_{t-1}, \kappa_t \in \Omega_\kappa}; \quad t \in \mathcal{T}_{-1}. \end{aligned} \tag{6}$$

The corresponding set of marginal posterior pdfs can also be recursively assessed

$$\mathbf{p}_{t|d} = [p(\kappa_t|\mathbf{d})]_{\kappa_t \in \Omega_\kappa} = \mathbf{P}'_{t-1, t|d} \mathbf{p}_{t-1|d}; \quad t \in \mathcal{T}_{-1}.$$

Since the posterior pdf is a first-order Markov chain with easily assessable initial pdf and set of transition matrices, generating a set of realizations from the posterior pdf, $[\kappa|\mathbf{d}]_s; s = 1, \dots, S$ is trivial and very fast. Moreover, determining the MAP state, $\hat{\kappa}_{MAP}$, in a first-order Markov chain can be done recursively by the very efficient Viterbi algorithm (Viterbi 1967); see ‘‘Appendix C’’.

Lastly, the posterior probability profile for each class, $\mathbf{p}_\kappa; \kappa \in \Omega_\kappa$, can be calculated from the set of marginal posterior pdfs, $\mathbf{p}_{t|d}; t \in \mathcal{T}$. Consequently, all features of the posterior pdf of interest can be assessed extremely efficiently by these recursive algorithms.

3.4 Alternative Prior Formulations

The prior pdf is assumed to be a first-order Markov chain, possibly a non-stationary one

$$p(\kappa) = p(\kappa_1) \times \prod_{t \in \mathcal{T}_{-1}} p(\kappa_t|\kappa_{t-1}). \tag{7}$$

The model is defined by the initial pdf \mathbf{p}_1 and the set of transition matrices $\mathbf{P}_{t-1, t}; t \in \mathcal{T}_{-1}$. The resulting set of marginal prior pdfs is $\mathbf{p}_t; t \in \mathcal{T}$.

3.4.1 Traditional Prior Model

The prior pdf is usually defined to be spatially stationary (Krumbein and Dacey 1969; Eidsvik et al. 2004). This definition entails that the elements in the set of transition matrices are identical, $\mathbf{P}_{t-1, t} = \mathbf{P}_{t'-1, t'} = \mathbf{P}; t, t' \in \mathcal{T}_{-1}$ and the initial pdf is $\mathbf{p}_1 = \mathbf{p}_s$ where \mathbf{p}_s is the associated stationary pdf, that is $\mathbf{p}_s = \mathbf{P}' \mathbf{p}_s$. Consequently, all marginal pdfs are identical $\mathbf{p}_t = \mathbf{p}_s; t \in \mathcal{T}$. Hence, spatial stationarity entails transition invariant marginal pdfs and transition invariant spatial dependence (Billard and Meshkani 1995).

This prior stationarity assumption is often suitable, since prior models are experience-based and globally anchored, while local characteristics should be a consequence of the site specific observations introduced by the likelihood function.

3.4.2 Trend Prior Model

Spatial stationarity is often a suitable assumption for the horizontal dimension, while the vertical dimension often is known to have trends. The proportions of facies are known to vary with depth in a vertical sub-surface profile, for example.

In the current study of the categorical variable κ along a vertical sub-surface profile one may want to specify varying vertical proportions a priori. A suitable prior model parameterization would be the varying marginal pdfs along the profile $\mathbf{p}_t^0; t \in \mathcal{T}$. The fact that this model specification does not uniquely define the prior Markov chain model $p(\kappa)$, since the set of marginal pdfs does not fully characterize the multivariate pdf, creates problems. One possible solution to this challenge is to define a reference transition matrix \mathbf{P}_r and to define a prior model specification for the multivariate pdf $p(\kappa)$ that has transition matrices that deviate as little as possible from \mathbf{P}_r , which still reproduce the marginal pdfs $\mathbf{p}_t^0; t \in \mathcal{T}$. This approach is used in defining the trend prior model in this study, and the approach is outlined below with a more detailed presentation in “Appendix D”.

Specify the marginal pdf profile $\mathbf{p}_t^0; t \in \mathcal{T}$ and a reference transition matrix $\mathbf{P}_r = [p_r(\kappa|\kappa')]_{\kappa, \kappa' \in \Omega_\kappa}$. The required prior model parameters are the initial pdf \mathbf{p}_1 and the set of transition matrices $\mathbf{P}_{t-1, t} = [p(\kappa_t|\kappa_{t-1})]_{\kappa_{t-1}, \kappa_t \in \Omega_\kappa}; t \in \mathcal{T}_{-1}$, which are defined as follows.

Let

$$\mathbf{p}_1 = \mathbf{p}_1^0$$

For $t = 2, \dots, T$

$$\mathbf{P}_{t-1, t} = \arg \min_{\mathbf{P}} \{ \|\mathbf{P} - \mathbf{P}_r\|_{wL^2} \}$$

constrained by

$$\begin{aligned} \mathbf{p}_t^0 &= \mathbf{P}' \mathbf{p}_{t-1}^0 \\ \mathbf{i}_k &= \mathbf{P} \mathbf{i}_k \\ \mathbf{P} &\geq 0 \mathbf{I}_k \end{aligned} \tag{8}$$

where \mathbf{P} is a $(K \times K)$ -matrix and the deviation measure $\|\cdot\|_{wL^2}$ is an element-wise weighted squared distance metric. The first set of constraints ensures reproduction of the marginal pdfs $\mathbf{p}_t^0; t \in \mathcal{T}_{-1}$, the second set ensures that the rows of \mathbf{P} add to unity, while the third set ensures that all elements in \mathbf{P} are non-negative. Consequently, \mathbf{p}_1 and the set $\mathbf{P}_{t-1, t}; t \in \mathcal{T}_{-1}$, define a first-order Markov chain with marginal pdfs $\mathbf{p}_t; t \in \mathcal{T}$ that reproduces $\mathbf{p}_t^0; t \in \mathcal{T}$.

The minimization system at each t contains a quadratic object function with two linear equality constraints and one linear inequality constraint. An optimization problem of this type cannot generally be solved analytically.

Sequential optimization procedure involving analytical solutions in each step is prescribed.

1. Solve the minimization system with the two linear equality constraints analytically by using Lagrange minimization. If the solution $(K \times K)$ -matrix $\mathbf{P}_{t-1,t}$ contains only non-negative elements, the solution is identified. If, however, some elements in $\mathbf{P}_{t-1,t}$ are negative, proceed to Step 2.
2. Set the negative elements in $\mathbf{P}_{t-1,t}$ to zero, and solve the minimization system with the linear equality constraints for the remaining elements. If the updated solution $\mathbf{P}_{t-1,t}$ has only non-negative elements, the solution is identified. If, however, some elements are negative iterate Step 2.

Note that if the solution is identified in Step 1, the optimal solution is found since the inequality constraints are inactive. Otherwise, neither existence of a solution nor optimality of the obtained solution are demonstrated. Note, however, if a solution of $\mathbf{P}_{t-1,t}$ is identified, it is a valid transition matrix and it reproduces the set of marginal pdfs $\mathbf{p}_t^0, t \in \mathcal{T}$. The algorithm has computational demands of decomposing a $((K + 1)^2 - 2) \times ((K + 1)^2 - 2)$ matrix at each t and is, therefore, linear in T and very efficient.

The current application, involving varying class proportions along a vertical sub-surface profile, will naturally have smoothly varying marginal pdf curves. The reference transition matrix \mathbf{P}_r can be interpreted as an average transition matrix of the Markov chain. By defining \mathbf{P}_r consistently with the marginal pdf profile, and by using natural binomial weighting in the object function, one expects the appearance of non-negative elements in the solution $\mathbf{P}_{t-1,t}$ to be unlikely; see “Appendix C”.

3.5 Assessment of Posterior Model

Both posterior models that are based on the traditional and the trend prior models can be exactly assessed by the recursive Forward–Backward algorithm; see “Appendix B”. The computational demands are proportional to $T \times K^2$ for both models, hence linear in the length of the profile. The solution identifies

$$\begin{aligned} \mathbf{p}_{1|d} &= [p(\kappa_1|\mathbf{d})]_{\kappa_1 \in \Omega_\kappa} \\ \mathbf{P}_{t-1,t|d} &= [p(\kappa_t|\kappa_{t-1}, \mathbf{d})]_{\kappa_{t-1}, \kappa_t \in \Omega_\kappa}; \quad t \in \mathcal{T}_{-1}. \end{aligned} \tag{9}$$

4 Results and Discussion

To compare the traditional and trend prior models, the MAP predictor, the MMAP predictor, the marginal probability profile, and realizations simulated from the posterior model are used. The methods are applied on two different data sets, one synthetic example and one real case study.

4.1 Synthetic Example

Consider an example with a vertical profile discretized to $\mathcal{T} = \{1, \dots, 100\}$ with categorical variable of interest $\kappa_t \in \Omega_\kappa : \{\text{black, gray}\}$, see reference profile κ^r

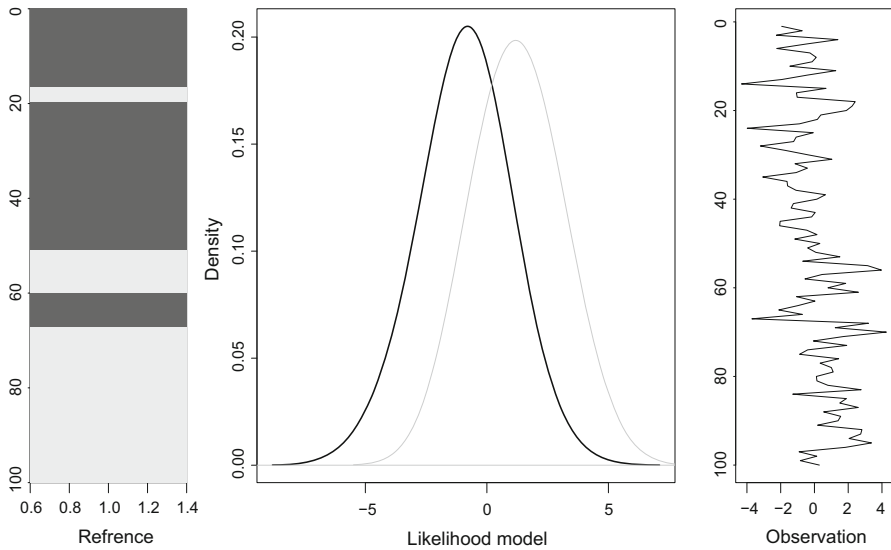


Fig. 2 Reference profile, likelihood model and reference observations

in Fig. 2. Note that there appears to be more black than gray on top of the profile and vice versa. The likelihood model is defined by $p(d_t|\kappa_t) = \varphi_1(d_t; \mu_{\kappa_t}, \sigma_{\kappa_t}^2)$ with $(\mu_{\kappa}, \sigma_{\kappa}^2)$; $\kappa \in \Omega_{\kappa}$ equal $(1, 1.5^2)$ and $(-1, 1.5^2)$ for κ being black and gray, respectively; see Fig. 2.

The synthetic reference observations \mathbf{d}^0 are also displayed in Fig. 2. The objective of the example is to assess κ given the observation \mathbf{d}^0 , hence assess $[\kappa|\mathbf{d}^0]$. Bayesian inversion as described in Sect. 2 is applied, and the correct likelihood function displayed in Fig. 2 is used. Two alternative prior models $p(\kappa)$ are evaluated. One traditional prior model with stationary transition matrix and stationary pdf as

$$\mathbf{P}_{t-1, t} = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix} \quad \mathbf{p}_s = (0.5, 0.5)'$$

One trend prior model with marginal pdf profile as displayed in Fig. 3b top third from left and reference transition matrix \mathbf{P}_r identical to the transition matrix for the traditional prior model. Note that the prior marginal profile assigns higher probability for black at the top and gray at the bottom and that this profile is subjectively assigned by the geoscientists. Based on this pdf profile and the reference matrix unique model parameters for the prior non-stationary Markov chain model are defined by the approach outlined in Sect. 3.2. The elements must be non-negative. In this example, negative values occurred in Step 1 of the algorithm, see “Appendix D”, at two depths. The actual depths are 79 and 80 and the values are -0.008 and -0.002 . Hence, the absolute values are very small compared to the range of probabilities $[0, 1]$, and they occur at break points close to 0 in the prior marginal profile. In this example, with two classes {black,gray} only, Step 2 in the algorithm will assign a binary transition for the corresponding facies at these depths. Experience tells that if the prior marginal

profile had been smoothed, the negative values are less likely to occur. Note that in this example with two classes {black,gray} the transition matrix to be determined is a (2×2) -matrix with four unknown elements. Each line in the matrix must sum to unity, hence only two unknown elements remain. The reproduction of the marginal probabilities defines two linear equality constraints which actually are linearly dependent and hence only enforce one active linear equality constraint. This reduces the number of unknown elements to only one which is identified by the quadratic minimization. In cases with more classes, hence large K , the number of free parameters in the optimization will be larger.

The solutions to the inversion problem $[\kappa|\mathbf{d}^0]$ are presented in Fig. 3. The upper display contains results based on the traditional prior model while the lower display contains results from the trend prior model. In both displays the reference profile κ^r and observations \mathbf{d}^0 are identical, and they are presented upper left. Afterwards, the actual prior marginal pdf profiles for the two prior models are presented in each display. The stationary and trend features are easily observable. The upper right profiles in each display are predictions of $[\kappa|\mathbf{d}]$, that is $\hat{\kappa}_{MAP}$ and $\hat{\kappa}_{MMAP}$ respectively, and the posterior probability profile \mathbf{p}_κ ; $\kappa \in \Omega_\kappa$, for each prior model. The impact of the trend in the latter display is easily observable in the posterior probability profile, there is a much clearer vertical black-gray drift than for the stationary model in the upper display. Note also that the MAP and MMAP predictions deviate at the very bottom of the profile. The predictions based on both prior models have a vertical black-gray drift due to conditioning on \mathbf{d}^0 , but the drift is emphasized by having the trend in the prior model. The lower part in each display contains six realizations from the posterior pdf for each prior model. These realizations expose higher heterogeneity than the corresponding predictions, of course. Note that the trend prior model in the lower display tend to reduce the heterogeneity at top and bottom of the profile. The emphasized vertical black-gray drift can be observed in the realization based on the trend prior model.

Note lastly, that introduction of trends in the prior models should be justified by external geoscience experience with the phenomenon under study, using a stationary prior model would be a natural choice without such experience.

4.2 Case Study

Focus is on predicting the soil classes in the sub-surface off-shore, in order to plan construction of an off-shore windmill park. Consider a vertical sub-surface profile discretized with $T = 1186$. The true profile in one location $\kappa^r = (\kappa_1^r, \dots, \kappa_T^r)'$ with $\kappa_t^r \in \Omega_\kappa : \{\text{black, red, green, blue}\}$ is displayed in Fig. 4a. The classes represent various sand/clay types with different geomechanical characteristics. In the same location, the log-responses, cone resistance and sleeve friction are collected.

These responses are normalized to obtain the actual observations $\mathbf{d}^0 = (\mathbf{d}_1^0, \dots, \mathbf{d}_T^0)$ with $\mathbf{d}_t^0 \in \mathcal{R}^2$ being normalized cone resistance and normalized sleeve friction, respectively (Fig. 4). In this case study the aim is to assess the true profile based on the observed log-response $[\kappa|\mathbf{d}^0]$.

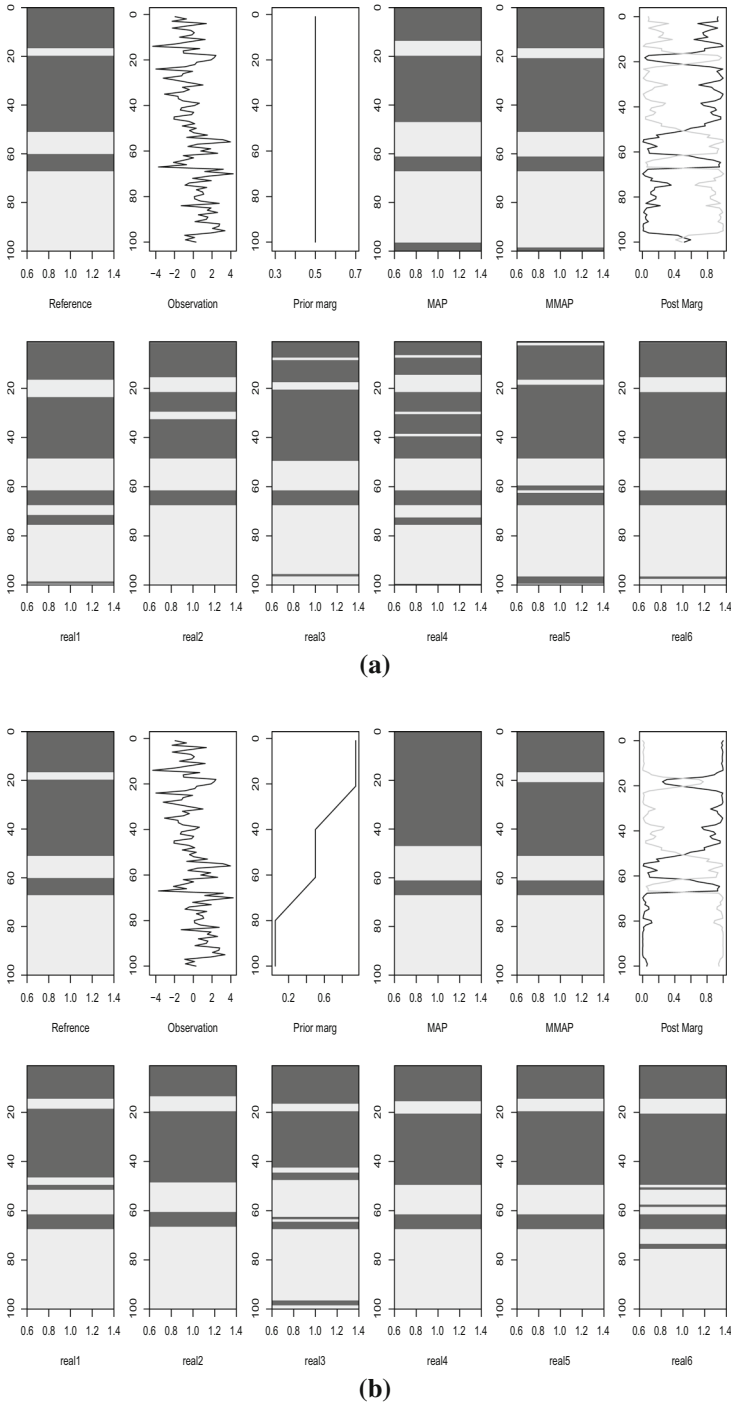


Fig. 3 Results from Bayesian inversion of synthetic example. **a** Results from traditional prior model. **b** Results from trend prior model

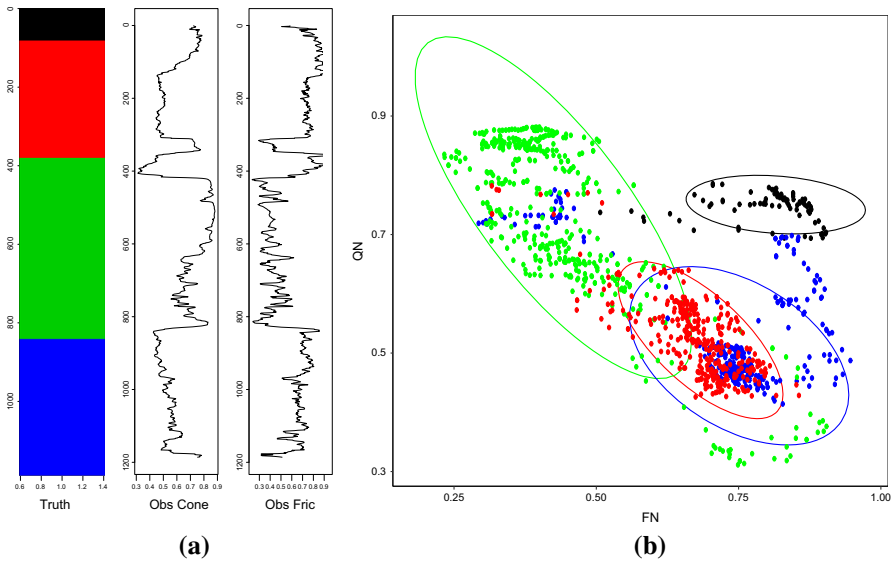


Fig. 4 Case study data and estimated likelihood model

The data in the well make it possible to present a bi-variate plot of likelihood function $p(\mathbf{d}_t^0 | \kappa_t)$, and a likelihood model is fitted,

$$p(\mathbf{d}_t^0 | \kappa_t) = \varphi_2(\mathbf{d}_t^0; \boldsymbol{\mu}_{\kappa_t}, \boldsymbol{\Sigma}_{\kappa_t}) \quad \kappa_t \in \Omega_{\kappa}$$

which also is displayed in Fig. 4b. Note that the model is very rough and does not over-fit the observations that actually appear as multi-modal; see for example the green class.

The Bayesian inversion results based on two alternative prior models are compared. One traditional Markov chain model with stationary class proportions along the profile, having stationary transition matrix and stationary proportions,

$$\mathbf{P} = \begin{bmatrix} 0.85 & 0.05 & 0.05 & 0.05 \\ 0.05 & 0.85 & 0.05 & 0.05 \\ 0.05 & 0.05 & 0.85 & 0.05 \\ 0.05 & 0.05 & 0.05 & 0.85 \end{bmatrix} \quad \mathbf{p}_s = (0.25 \ 0.25 \ 0.25 \ 0.25)'$$

The other model is a trend prior model with depth varying class proportions, \mathbf{p}_t^0 ; $t \in \mathcal{T}$, presented in Fig. 5b. No class has < 0.05 and more than 0.85 prior probability to occur. The reference transition matrix \mathbf{P}_r is identical to the transition matrix above. These two parameters do not uniquely define a prior Markov model, but the approach outlined in Sect. 3.2 to specify the prior pdf $p(\kappa)$ is used. In this study there are nine free parameters to minimize the object function and the object function is centered close to zero, at 0.05, for twelve of the 16 parameters.

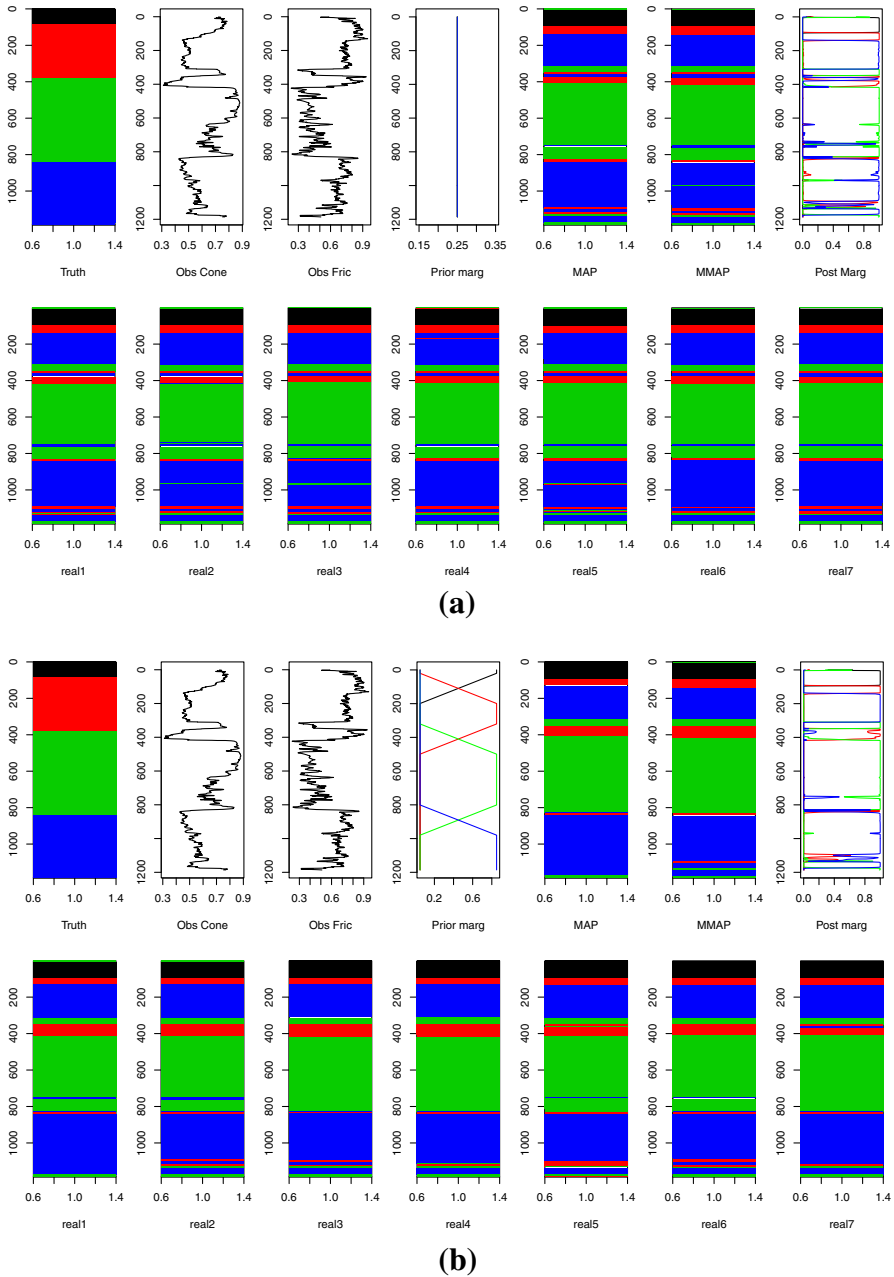
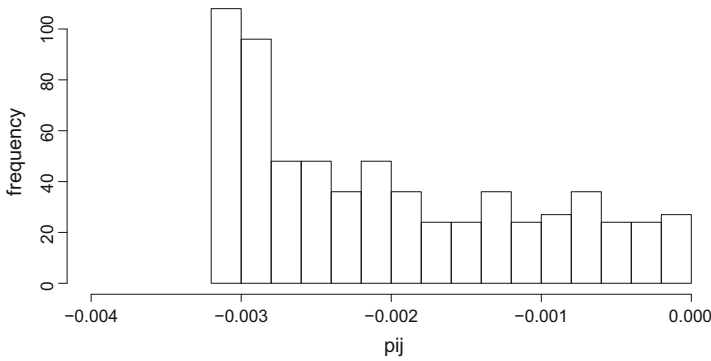
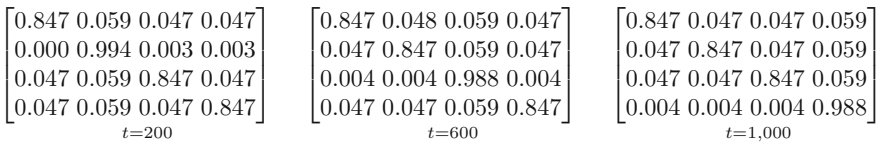


Fig. 5 Results from Bayesian inversion of Case study. **a** Results from traditional prior model. **b** Results from trend prior model



(a)



(b)

Fig. 6 Results from optimization procedure for Case study. **a** Histogram of negative values after Step 1 in optimization procedure for Case study. **b** Prior transition matrix after optimization procedure for Case study

In the algorithm, see “Appendix D”, $T - 1 = 1185$ transition (4×4)-matrices are identified, hence 18,960 transition probabilities. After Step 1 666 negative values appeared and these values are displayed in the histogram in Fig. 6a. The absolute values of all these negative values are very small. In Step 2 of the algorithm these transition probabilities are assigned the value zero, and a reduced optimization is solved. The solution after this step contained no negative values.

These results are encouraging since the final solution is very close to the initial one after Step 1. Figure 6b display the calculated transition matrices $\mathbf{P}_{t-1,t}$ for $t = 200, 600, 1000$. Note that the matrix for $t = 200$ has one zero-valued element which is assigned in Step 2 of the algorithm. The transition matrices appear as expected with relatively large diagonal elements ensuring high spatial dependence of the class. The identical off-diagonal elements are caused by the marginal probability for several classes being identical in sections of the prior specification.

Note that the two prior models are fairly general and based on expert experience about soil distribution in the sub-surface in the area. The results from the Bayesian inversion of $[\kappa|\mathbf{d}^0]$ based on the two alternative prior models are displayed in Fig. 5 in a format similar to Fig. 3 in the synthetic example. The MAP, MMAP and posterior marginal pdfs based on the traditional prior model, in Fig. 5a upper right, do largely have the correct ordering of the classes compared to the true profile at far left in the figure. The predictions appear with many more class-transition than the true profile, however. The associated posterior realizations in Fig. 5a bottom row are even more heterogeneous. The reason for heterogeneity can be seen in the likelihood display,

Fig. 4b. Observe that the likelihood model for the blue class, for example, is very wide due to the appearance of two blue clusters in the data, hence it overlaps the likelihood for the red class. A more adaptive likelihood model could be defined, of course, but this could cause over-fitting. Alternatively, a more informative prior transition matrix could be defined, also at the cost of possibly over-fitting.

The MAP, MMAP and posterior marginal pdfs based on the trend prior model is displayed in Fig. 5b upper right, with the posterior realizations in the bottom row of the figure. These results have many features in common with the results based on the traditional prior model, of course, since both models draw on the same likelihood model. Observe, however, that the predictions and posterior realizations based on the trend prior model appear with significantly less heterogeneity than the ones based on the traditional prior model. The varying proportion profile enforced through the trend prior model reduces the number of class-transitions in the posterior model.

This case study is based on one calibration well only, hence the model inference and posterior model evaluation are based on the same data. This coupling favors over-fitted models which are avoided by use of rough likelihood models. A fair comparison should involve posterior model evaluation on data from a blind calibration well, which will be made in a later study.

5 Conclusions

A categorical inversion method for modelling the sequence of classes along a sub-surface profile based on a observed logging response is presented. The inversion is cast in a Bayesian framework as a hidden Markov model and assessed by the Forward–Backward and Viterbi algorithms. Two different prior model formulations are studied, one traditional stationary model and one trend prior model. The trend prior model is parametrized by a vertical class-proportion profile and a reference transition matrix. By using an efficient sequential analytical procedure depth dependent transition matrices reproducing the class-proportions are identified. The two prior models are tested on both synthetic and real data sets.

User-specified class-proportions along the vertical profile may have large impact on the class predictions and simulations. The number of class transitions is often lower than the corresponding number for a traditional stationary model. The algorithm for estimating the depth-dependent transition matrices from the proportion profiles and the reference matrix appears as efficient and reliable.

The possibility for geologists to constrain the categorical inversion by class-proportion curves is expected to have a large user-potential.

Acknowledgements The research is made as a part of Ph.D.-study at School of Mathematical and Statistical Sciences, Hawassa University, Ethiopia. The funding is provided by the Ethiopian Department of Education and the Norwegian Agency for Development Cooperation. Also thanks to Ivan Depina, Sintef, Norway for providing and supporting the real geotechnical data.

Funding Funding was provided by Ethiopian Department of Education, Ethiopia and Norwegian Agency for Development cooperation, Norway.

Appendix A: Markov Property of Posterior Model

Consider posterior model

$$\begin{aligned}
 p(\boldsymbol{\kappa}|\mathbf{d}) &= \text{const} \times p(d_1|\kappa_1)p(\kappa_1) \prod_{t \in \mathcal{T}_{-1}} p(d_t|\kappa_t)p(\kappa_t|\kappa_{t-1}) \\
 &= p(\kappa_1|\mathbf{d}) \prod_{t \in \mathcal{T}_{-1}} p(\kappa_t|\kappa_{t-1}, \mathbf{d}).
 \end{aligned}$$

Rephrase $p(\boldsymbol{\kappa}|\mathbf{d})$ by a general conditioning decomposition

$$p(\boldsymbol{\kappa}|\mathbf{d}) = p(\kappa_1|\mathbf{d}) \prod_{t \in \mathcal{T}_{-1}} p(\kappa_t|\kappa_{t-1}, \dots, \kappa_1, \mathbf{d}).$$

By demonstrating that for $t \in \mathcal{T}_{-1}$

$$p(\kappa_t|\kappa_{t-1}, \dots, \kappa_1, \mathbf{d}) = \frac{p(\kappa_t, \dots, \kappa_1, \mathbf{d})}{p(\kappa_{t-1}, \dots, \kappa_1, \mathbf{d})} = p(\kappa_t|\kappa_{t-1}, \mathbf{d})$$

the Markov property of the posterior model is proven.

Use notation $\boldsymbol{\kappa}_{1:s} = (\kappa_1, \dots, \kappa_s)$ and $\mathbf{d}_{s:T} = (d_s, \dots, d_T)$ and define

$$\begin{aligned}
 p(\boldsymbol{\kappa}_{1:s}|\mathbf{d}) &= \sum_{\kappa_T \in \Omega_\kappa} \cdots \sum_{\kappa_{s+1} \in \Omega_\kappa} p(\boldsymbol{\kappa}|\mathbf{d}) \\
 &= \text{const} \times p(d_1|\kappa_1)p(\kappa_1) \prod_{i=2}^s p(d_i|\kappa_i)p(\kappa_i|\kappa_{i-1}) \\
 &\quad \times \sum_{\kappa_T \in \Omega_\kappa} \cdots \sum_{\kappa_{s+1} \in \Omega_\kappa} \prod_{j=s+1}^T p(d_j|\kappa_j)p(\kappa_j|\kappa_{j-1}) \\
 &= \text{const} \times p(d_1|\kappa_1) p(\kappa_1) \prod_{i=2}^s p(d_i|\kappa_i) p(\kappa_i|\kappa_{i-1}) \times v_s(\kappa_s, \mathbf{d}_{s+1:T}).
 \end{aligned}$$

The latter factor is only a function of κ_s and $\mathbf{d}_{s+1:T}$ since $(\kappa_T, \dots, \kappa_{s+1})$ is marginalized out. Hence, we may write

$$\begin{aligned}
 p(\kappa_t|\kappa_{t-1}, \dots, \kappa_1, \mathbf{d}) &= \frac{p(\boldsymbol{\kappa}_{1:t}|\mathbf{d})}{p(\boldsymbol{\kappa}_{1:t-1}|\mathbf{d})} \\
 &= p(d_t|\kappa_t)p(\kappa_t|\kappa_{t-1}) \frac{v_t(\kappa_t, \mathbf{d}_{t+1:T})}{v_{t-1}(\kappa_{t-1}, \mathbf{d}_{t:T})} \\
 &= p(\kappa_t|\kappa_{t-1}, \mathbf{d}_{t:T}) \quad \text{QED.}
 \end{aligned}$$

Note that the demonstration also holds when the prior model $p(\boldsymbol{\kappa})$ is a non-stationary Markov chain.

Appendix B: Forward–Backward Algorithm

Consider posterior model

$$\begin{aligned}
 p(\kappa|\mathbf{d}) &= \text{const} \times p(d_1|\kappa_1)p(\kappa_1) \prod_{t \in \mathcal{T}_{-1}} p(d_t|\kappa_t)p(\kappa_t|\kappa_{t-1}) \\
 &= p(\kappa_1|\mathbf{d}) \prod_{t \in \mathcal{T}_{-1}} p(\kappa_t|\kappa_{t-1}, \mathbf{d}).
 \end{aligned}$$

Moreover let $\mathbf{d}_{1:t} = [d_1, \dots, d_t]$ be the subset of $\mathbf{d} = \mathbf{d}_{1:T}$ up to time t .

Algorithm 1: Forward–Backward Algorithm

```

1 Step for Forward Probability
2 for  $\kappa_1 \in \Omega_\kappa$  do
3   |  $p(\kappa_1|d_1) = \text{const} \times p(d_1|\kappa_1) p(\kappa_1)$ 
4 end
5 const such that  $\sum_{\kappa_1} p(\kappa_1|d_1) = 1$ 
6 for  $t = 2$  to  $T$  do
7   | for  $\kappa_t \in \Omega_\kappa$  do
8     | for  $\kappa_{t-1} \in \Omega_\kappa$  do
9       | |  $p(\kappa_{t-1}, \kappa_t|\mathbf{d}_{1:t}) = \text{const} \times p(d_t|\kappa_t) p(\kappa_t|\kappa_{t-1}) p(\kappa_{t-1}|\mathbf{d}_{1:t-1})$ 
10      | end
11     end
12     const such that  $\sum_{\kappa_{t-1}} \sum_{\kappa_t} p(\kappa_{t-1}, \kappa_t|\mathbf{d}_{1:t}) = 1$ 
13      $p(\kappa_t|\mathbf{d}_{1:t}) = \sum_{\kappa_{t-1}} p(\kappa_{t-1}, \kappa_t|\mathbf{d}_{1:t})$ 
14   end

15 Step for Backward Probability
16 for  $t = T$  to  $2$  do
17   | for  $\kappa_{t-1} \in \Omega_\kappa$  do
18     | for  $\kappa_t \in \Omega_\kappa$  do
19       | |  $p(\kappa_{t-1}|\kappa_t, \mathbf{d}) = p(\kappa_{t-1}|\kappa_t, \mathbf{d}_{1:t}) = \frac{p(\kappa_{t-1}, \kappa_t|\mathbf{d}_{1:t})}{p(\kappa_t|\mathbf{d}_{1:t})}$ 
20      | end
21     |  $p(\kappa_{t-1}|\mathbf{d}) = \sum_{\kappa_t} p(\kappa_{t-1}|\kappa_t, \mathbf{d}) p(\kappa_t|\mathbf{d})$ 
22     | for  $\kappa_t \in \Omega_\kappa$  do
23       | |  $p(\kappa_t|\kappa_{t-1}, \mathbf{d}) = \frac{p(\kappa_{t-1}|\kappa_t, \mathbf{d}) p(\kappa_t|\mathbf{d})}{p(\kappa_{t-1}|\mathbf{d})}$ 
24      | end
25     end
26 end

```

Appendix C: Viterbi Algorithm

Consider maximum posterior prediction

$$\begin{aligned} \hat{\kappa}_{MAP} &= \text{MAP} \{ \kappa | \mathbf{d} \} = \arg \max_{\kappa} \{ p(\kappa | \mathbf{d}) \} \\ &= \arg \max_{\kappa} \left\{ p(\kappa_1 | \mathbf{d}) \prod_{t \in \mathcal{T}_{-1}} p(\kappa_t | \kappa_{t-1}, \mathbf{d}) \right\} \\ p_{MAP} &= p(\hat{\kappa}_{MAP} | \mathbf{d}). \end{aligned}$$

Moreover, let $\kappa_t(\kappa) = [\hat{\kappa}'_1, \dots, \hat{\kappa}'_{t-1}, \kappa]$ be MAP-trace up to t given $\kappa_t = \kappa$ with associated MAP-probability $p_t^M(\kappa)$ for $\kappa \in \Omega_{\kappa}$, and $p_t^M(\kappa | \kappa')$ be MAP-probability for $\kappa_t = \kappa$ given that $\hat{\kappa}'_{t-1} = \kappa'$ for $\kappa, \kappa' \in \Omega_{\kappa}$.

Algorithm 2: Viterbi Algorithm

```

1  Start Algorithm
2  for  $\kappa \in \Omega_{\kappa}$  do
3     $p_1^M(\kappa) = p(\kappa_1 = \kappa | \mathbf{d})$ 
4     $\kappa_1(\kappa) = [\kappa]$ 
5  end
6  for  $t = 2$  to  $T$  do
7    for  $\kappa \in \Omega_{\kappa}$  do
8      for  $\kappa' \in \Omega_{\kappa}$  do
9         $p_t^M(\kappa | \kappa') = p_{t-1}^M(\kappa') p(\kappa_t = \kappa | \kappa_{t-1} = \kappa', \mathbf{d})$ 
10      end
11       $\hat{\kappa}'_{t-1} = \arg \max_{\kappa'} \{ p_t^M(\kappa | \kappa') \}$ 
12       $\kappa_t(\kappa) = [\kappa_{t-1}(\hat{\kappa}'_{t-1}), \kappa]$ 
13       $p_t^M(\kappa) = p_t^M(\kappa | \hat{\kappa}'_{t-1})$ 
14    end
15  end
16   $\hat{\kappa}'_T = \arg \max_{\kappa} \{ p_T^M(\kappa) \}$ 
17   $\hat{\kappa}_{MAP} = \kappa_T(\hat{\kappa}'_T) = [\hat{\kappa}'_1, \dots, \hat{\kappa}'_T]$ 
18   $p_{MAP} = p_T^M(\hat{\kappa}'_T)$ 

```

Appendix D: Trend Prior Model

Consider spatial discretization $\mathcal{T} : \{1, \dots, T\}$ and categorical variable $\kappa = [\kappa_1, \dots, \kappa_T]$; $\kappa_t \in \Omega_{\kappa} : \{1, \dots, K\}$. Define prior Markov chain model

$$p(\kappa) = p(\kappa_1) \prod_{t \in \mathcal{T}_{-1}} p(\kappa_t | \kappa_{t-1})$$

parametrized by initial pdf $\mathbf{p}_1 = [p(\kappa_1)]_{\kappa_1 \in \Omega_\kappa}$ and a set of transition matrices $\mathbf{P}_{t-1, t} = [p(\kappa_t | \kappa_{t-1})]_{\kappa_{t-1}, \kappa_t \in \Omega_\kappa; t \in \mathcal{T}_{-1}}$.

Consider a non-complete set of model parameters, where one is the set of marginal pdfs $\mathbf{p}_t^0 = [p^0(\kappa_t)]_{\kappa_t \in \Omega_\kappa; t \in \mathcal{T}}$ and one is the reference transition matrix $\mathbf{P}_r = [p_r(\kappa | \kappa')]_{\kappa, \kappa' \in \Omega_\kappa}$. The challenge is to define a set of model parameters for the prior model $p(\kappa)$ where the marginal pdfs reproduces $\mathbf{p}_t^0; t \in \mathcal{T}$ in the non-complete set of model parameters and where the set of transition matrices do not deviate too much from \mathbf{P}_r . The set of parameters for the prior Markov chain model is defined by the following set of constrained optimization problems.

Initial

$$\mathbf{p}_1 = \mathbf{p}_1^0$$

for $t = 2, \dots, T$ let

$$\mathbf{P}_{t-1, t} = \arg \min_{\mathbf{P}} \{ \|\mathbf{P} - \mathbf{P}_r\|_{wL^2} \}$$

with

$$\begin{aligned} \|\mathbf{P} - \mathbf{P}_r\|_{wL^2} &= \sum_{\kappa \in \Omega_\kappa} \sum_{\kappa' \in \Omega_\kappa} w_{\kappa \kappa'} [p(\kappa | \kappa') - p_r(\kappa | \kappa')]^2 \\ \text{and } w_{\kappa \kappa'} &= [[1 - p_r(\kappa | \kappa')] p_r(\kappa | \kappa')]^{-1} \end{aligned}$$

constrained by

$$\begin{aligned} \mathbf{p}_t^0 &= \mathbf{P}' \mathbf{p}_{t-1}^0 \text{ providing } p_t^0(\kappa) = \sum_{\kappa' \in \Omega_\kappa} p(\kappa | \kappa') p_{t-1}^0(\kappa'); \quad \kappa \in \Omega_\kappa \\ \mathbf{i}_k &= \mathbf{P} \mathbf{i}_k \text{ providing } 1 = \sum_{\kappa \in \Omega_\kappa} p(\kappa | \kappa'); \quad \kappa' \in \Omega_\kappa \\ 0 \mathbf{I}_k &\leq \mathbf{P} \text{ providing } 0 \leq p(\kappa | \kappa'); \quad \kappa, \kappa' \in \Omega_\kappa \end{aligned}$$

end.

The solution to this optimization problem for a given t , $\mathbf{P}_{t-1, t}$, will be a valid transition matrix, which reproduces the marginal pdf \mathbf{p}_t^0 . Moreover $\mathbf{P}_{t-1, t}$ will appear with minimum weighted deviation from the reference transition matrix \mathbf{P}_r .

Since $1 = \mathbf{p}_t^{0'} \mathbf{i}_k; t \in \mathcal{T}$ and $\mathbf{i}_k = \mathbf{P}' \mathbf{i}_k$ the first set of equality constraints is linearly dependent for each t . This linear dependence is canceled by removing the constraint for $\kappa = K$.

Parameterize $\mathbf{P}_{t-1, t}$ and \mathbf{P} by $\alpha : \{\alpha_{\kappa' \kappa} = p(\kappa | \kappa'); \kappa, \kappa' \in \Omega_\kappa\}$ and obtain for each t

$$\mathbf{P}_{t-1, t}(\alpha) = \arg \min_{\alpha} \left\{ \sum_{\kappa \in \Omega_{\kappa}} \sum_{\kappa' \in \Omega_{\kappa}} \frac{[\alpha_{\kappa' \kappa} - p_r(\kappa|\kappa')]^2}{(1 - p_r(\kappa|\kappa')) p_r(\kappa|\kappa')} \right\}$$

constrained by

$$\begin{aligned} p_t^0(\kappa) &= \sum_{\kappa \in \Omega_{\kappa}} \alpha_{\kappa' \kappa} p_{t-1}^0(\kappa') \quad \kappa \in \Omega_{\kappa}/K \\ 1 &= \sum_{\kappa \in \Omega_{\kappa}} \alpha_{\kappa' \kappa} \quad \kappa' \in \Omega_{\kappa} \\ 0 &\leq \alpha_{\kappa' \kappa} \quad \kappa', \kappa \in \Omega_{\kappa}. \end{aligned}$$

This constitutes an optimization problem with quadratic object function with two sets of linear equality constraints and one set of linear inequality constraints. No closed form analytical solution to this optimization problem exists. Note, however, that the reference transition matrix \mathbf{P}_r at which the object function is centered, obey the two latter sets of constraints. Moreover, note that the weights $w_{\kappa \kappa'}$ make deviations from elements of \mathbf{P}_r close to the border of the inequality constraints very costly. If \mathbf{p}_{t-1}^0 and \mathbf{p}_t^0 do not deviate dramatically and \mathbf{P}_r is not chosen in conflict with these two marginal pdfs, most likely the inequality constraints will be non-active.

For each t , the optimization is performed sequentially

1. Minimization of the object function with the two sets of equality constraints is made. This optimization can be done analytically by Lagrange optimization in dimension $(K^2 + (K - 1) + K)$ and a closed form solution is identified. If $\alpha_{\kappa' \kappa} \geq 0 \quad \kappa', \kappa \in \Omega_{\kappa}$, the solution to the optimization is identified. If $\alpha_{\kappa' \kappa} < 0$ for some $\kappa', \kappa \in \Omega_{\kappa}$, for example $\kappa', \kappa \in \Omega^- \subset \Omega_{\kappa}$, go to Step 2
2. Set $\alpha_{\kappa' \kappa} = 0 \quad \kappa', \kappa \in \Omega^-$. Minimization of the object function with the two sets of equality constraints with respect to the remaining elements $\alpha_{\kappa' \kappa}; \kappa', \kappa \in \Omega_{\kappa} \setminus \Omega^-$. The optimization is done by Lagrange optimization in dimension $((K - \ell)^2 + (K - 1) + K)$ with $\ell = |\Omega^-|$, and a closed form solution is identified. If $\alpha_{\kappa' \kappa} \geq 0; \kappa', \kappa \in \Omega_{\kappa} \setminus \Omega^-$, the solution of the optimization is identified, otherwise iterate Step 2.

If the solution, $\mathbf{P}_{t-1, t}$, is identified in Step 1, the exact solution to the optimization problem is found and the inequality constraints are inactive. If, however, the solution $\mathbf{P}_{t-1, t}$, is identified in Step 2, it is ensured to be a valid transition matrix reproducing the marginal pdfs, but it need not be the exact solution to the optimization problem. Lastly, no demonstration of the existence of a solution is currently made.

References

Avseth P, Mukerji T, Mavko G (2005) Quantitative seismic interpretation: applying rock physics tools to reduce interpretation risk. Cambridge University Press, Cambridge

Baum LE, Petrie T, Soules G, Weiss N (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Ann Math Stat 41(1):164–171

Billard L, Meshkani MR (1995) Estimation of stationary Markov chains. J Am Stat Assoc 90(429):307–315

- Dymarski P (ed)(2011) Hidden Markov models, theory and applications. InTechOpen. <http://www.intechopen.com/InTechOpen>
- Eidsvik J, Mukerji T, Switzer P (2004) Estimation of geological attributes from a well log: an application of hidden Markov chains. *Math Geol* 36(3):379–396
- Harbaugh JW, Bonham-Carter G (1970) Computer simulation in geology. Wiley, New York
- Krumbein YC, Dacey MF (1969) Markov chains and embedded Markov chains in geology. *Math Geol* 1(1):79–96
- Ravenne G, Galli A, Doligez B, Beucher H, Eschard R (2002) Quantification of facies relationships in a proportion curves. In: Armstrong M, Bettini C, Champigny N, Galli A, Remacre A (eds) *Geostatistics Rio 2000. Quantitative geology and geostatistics*. Springer, Dordrecht, pp 7–51
- Robertson PK (2010) Soil behavior type from the CPT: an update. In: 2nd international symposium on cone penetration testing, CPT'10. Huntington Beach, CA, USA, pp 9–11
- Scott SL (2002) Bayesian methods for hidden Markov models: recursive computing in the 21st century. *J Am Stat Assoc* 97:337–351
- Ulvmoen M, Omre H, Buland A (2010) Improved resolution in Bayesian lithology/fluid inversion from prestack seismic data and well observations: part 2—real case study. *Geophysics* 75(2):B73–B82
- Viterbi A (1967) Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans Inf Theory* 13(2):260–269
- Weissman GS, Fogg GE (1999) Multi-scale alluvial fan heterogeneity modeled with transition probability geostatistics in a sequence stratigraphic framework. *J Hydrol* 226(1–2):48–65