Geophysical Journal International

Geophys. J. Int. (2017) **208,** 1184–1200
Advance Access publication 2016 November 2
GJI Seismology

doi: 10.1093/gji/ggw411

Bayesian inversion of seismic attributes for geological facies using a Hidden Markov Model

Muhammad Atif Nawaz and Andrew Curtis

School of Geosciences, Grant Institute, University of Edinburgh, Edinburgh, United Kingdom. E-mail: Muhammad.AtifNawaz@ed.ac.uk, Andrew.Curtis@ed.ac.uk

Accepted 2016 November 1. Received 2016 October 28; in original form 2016 June 3

SUMMARY

Markov chain Monte-Carlo (McMC) sampling generates correlated random samples such that their distribution would converge to the true distribution only as the number of samples tends to infinity. In practice, McMC is found to be slow to converge, convergence is not guaranteed to be achieved in finite time, and detection of convergence requires the use of subjective criteria. Although McMC has been used for decades as the algorithm of choice for inference in complex probability distributions, there is a need to seek alternative approaches, particularly in high dimensional problems. Walker & Curtis (2014) developed a method for Bayesian inversion of 2-D spatial data using an exact sampling alternative to McMC which always draws independent samples of the target distribution. Their method thus obviates the need for convergence and removes the concomitant bias exhibited by finite sample sets. Their algorithm is nevertheless computationally intensive and requires large memory. We propose a more efficient method for Bayesian inversion of categorical variables, such as geological facies that requires no sampling at all. The method is based on a 2-D Hidden Markov Model (2D-HMM) over a grid of cells where observations represent localized data constraining each cell. The data in our example application are seismic attributes such as *P*- and *S*-wave impedances and rock density; our categorical variables are the hidden states and represent the geological rock types in each cell—facies of distinct subsets of lithology and fluid combinations such as shale, brine-sand and gas-sand. The observations at each location are assumed to be generated from a random function of the hidden state (facies) at that location, and to be distributed according to a certain probability distribution that is independent of hidden states at other locations – an assumption referred to as 'localized likelihoods'. The hidden state (facies) at a location cannot be determined solely by the observation at that location as it also depends on prior information concerning the spatial distribution of other hidden states elsewhere. The prior information is included in the inversion in the form of a training image which represents a conceptual depiction of local geologies that might be expected, but other forms of prior information can be used in the method as desired. The method provides direct estimates of posterior marginal probability distributions over each variable, so these do not need to be estimated from samples such as in McMC. Nevertheless, in the case that samples are desired, these can be generated. On a 2-D test example the method is shown to outperform previous methods significantly, at a fraction of the computational cost. In many foreseeable applications there are therefore no serious impediments to extending the method to 3-D cases.

Key words: Inverse theory; Spatial analysis; Probability distributions.

1 INTRODUCTION

Geological phenomena always exhibit some degree of spatial correlation and continuity, but also apparent randomness in space at various scales. Such spatial variability may be described by model that is ultimately governed by geological processes. Geoscientists are usually able to build plausible models of spatial distributions of facies in a certain area or volume using prior knowledge about the local geological context. The spatial variability in geophysically detectable rock properties (e.g. elastic or electromagnetic properties or density) generally depends on the spatial distribution of rock facies (distinctly classifiable rock types), but is often more complex than the spatial variability of the facies themselves. For this reason, geoscientists can provide better a priori constraints on the spatial

distribution of geological facies than on the variability in such rock properties.

Herein we describe a method to estimate facies in a subsurface geological volume, a process which we refer to as *geological modelling*. The general approach used in geological modelling is to classify various facies from available seismic data based on inferred seismic attributes such as *P*- and *S*-wave impedances, Poisson's ratio or amplitude variations with reflection angle. Although such attributes have been used successfully to discriminate various facies defined in terms of lithology and fluid type, such a classification method does not produce the expected geologically realistic spatial correlations of facies. An alternative approach is therefore required to combine prior knowledge about expected facies distributions in space, with information in the data given in terms of facies probabilities derived from discriminative seismic attributes.

The prior information about realistic spatial distributions of facies is usually represented in terms of joint or conditional probability distributions of various facies over a graphical structure called a *neighbourhood*, which in turn is often defined as a rectangular array of cells in 2-D models. These probabilities are conveniently embodied in the form of *training images*. A training image is a conceptual and pictorial spatial depiction of facies patterns that are likely to exist in the subsurface based on the subjective opinion of a geoscientist, or on other objective data about the local geology. Training images allow statistical information about the spatial correlations of facies to be estimated from graphical representations of their spatial distribution.

There is always uncertainty in the estimation of geological facies from the observed data at a given location. The uncertainty is either due to uncertainty in the estimates of seismic attributes or due to the intrinsic uncertainty in the relationship between facies and the seismic attributes, or both. This implies that the seismic attributes inferred at a given location are related to facies at that location according to some probability distribution. Although the actual observation in a seismic experiment is the raw seismic data, the inferred seismic attributes are referred to as 'observed data' or 'observations' herein. This explicitly distinguishes them from the geological facies which are treated as 'hidden' (not observed) variables. The probability of observing (or inferring) a specific set of seismic attributes at a fixed location given that a particular facies exists at that location, is called the data *likelihood*. Uncertainty in the attributes is accounted for within the likelihood. Since spatial correlation of facies is controlled by the prior probabilities it is commonly assumed that the data likelihood is localized: that is, given the facies at a location, the seismic attributes at that location are conditionally independent of both facies and the seismic attributes at all other locations. This assumption is henceforth referred to as the condition of *localized likelihoods*.

The contextual information expressed as prior probabilities of spatial correlations of facies may be combined with the local information provided by likelihood probabilities in a Bayesian framework. Thus we obtain posterior probabilities of the spatial distribution of geological facies that conform to both the observed data (seismic attributes) and prior constraints. However, a major problem is that the full posterior distribution is usually analytically intractable for standard high-dimensional models and must be explored through simulation and sampling based inference, for example, by using Markov-chain Monte Carlo (McMC) methods. McMC based sampling is computationally demanding as it requires many samples to converge to the true distribution (see Walker & Curtis 2014). The problem escalates in high-dimensional models such that convergence is not guaranteed to be achieved even when a very large but finite number of samples are generated: the chain of samples generated may be confined to a region of locally high probability in the sample space. The samples are then strongly dependent on the initial steps in the McMC random walk, which may lead to false apparent convergence to the posterior probability distribution. As a consequence of this, any fixed finite set of samples may be biased in the sense that they do not represent the true Bayesian posterior distribution. This bias is related to the use of subjective stopping criteria to detect convergence of the chain. To address these challenges, researchers have suggested McMC methods for posterior distributions that can be expressed in factorizable forms, which allow more efficient and reliable sampling than a naïve implementation of McMC in such cases (e.g. Sambridge & Mosegaard 2002; Gallagher *et al*. 2009; Rimstad *et al*. 2012).

Walker & Curtis (2014) developed a method for Bayesian inversion of 2-D spatial data using an exact sampling alternative to McMC. This allows independent samples of the target distribution to be drawn without requiring convergence, thus circumventing convergence related biases. They developed a method for Bayesian inversion of geological facies from seismic data that incorporates prior geological information extracted from a training image and the localized data likelihood computed using a Gaussian Mixture Model (GMM) based on neural networks (Meier *et al*. 2007a,b; Shahraeeni & Curtis 2011; Shahraeeni *et al*. 2012). Their algorithm computes the posterior distribution $\mathcal{P}(\boldsymbol{G} \mid \boldsymbol{D})$ of geological facies given data from the partial conditional posterior distributions $\mathcal{P}(g_i \mid \boldsymbol{D}, \boldsymbol{G}_{<i})$ as

$$\mathcal{P}(\boldsymbol{G} \mid \boldsymbol{D}) = \prod_{i=1}^{M} \mathcal{P}(g_i \mid \boldsymbol{D}, \boldsymbol{G}_{<i}) \tag{1}$$

where $\mathcal{P}$ represents probability, M is the total number of cells in the model, i represents a cell index according to a linear indexing scheme, g_i represents model parameters (geological facies) in cell i, $< i$ represents the set of indices from 1 to $i-1$, $\boldsymbol{G} = [g_1, \ldots, g_i, \ldots, g_M]$ is the set of model parameters, and $\boldsymbol{D}$ is the set of observed data in all cells in the model. The partial conditionals (right hand terms in the above equation) must also usually be calculated since prior information often comes in the form of full conditional representations, $\mathcal{P}(\boldsymbol{G}_i \mid \boldsymbol{D}, \boldsymbol{G}_{\setminus i})$ where $\boldsymbol{G}_{\setminus i} = \boldsymbol{G} \setminus \{g_i\}$ (where $\setminus$ stands for set theoretic difference). Their algorithm computes partial conditional distributions of facies in each cell in the model in turn, given the facies in all previously sampled cells. This process requires large memory and is computationally intensive for real-scale seismic data and geological modelling problems. If distribution functions such as marginals of the posterior distribution are required, these must then be calculated from the set of samples generated.

Here we take a different approach: we compute marginal posterior distributions of facies in each cell in the model directly, without sampling. This incorporates prior geological knowledge and the data likelihood in a similar manner to Walker & Curtis (2014). However, computation of marginal posterior distributions in each cell in the model is computationally more efficient and requires less memory, compared to the computation of partial conditional distributions used in the previous work.

Before we proceed, we present the notation used in this paper. We use boldface capital letters for sets, such as $\boldsymbol{G}$ for the set of geological facies and $\mathbf{D}$ for the set of data in each cell in the model. We use boldface small letters for vectors, for example, $\boldsymbol{d}_{ij}$ represents a vector of data values in cell (i, j); and we use regular small letters for scalar variables, for example, g_{ij} represents geological facies in cell (i, j), where the indexing refers to a 2-D indexing scheme

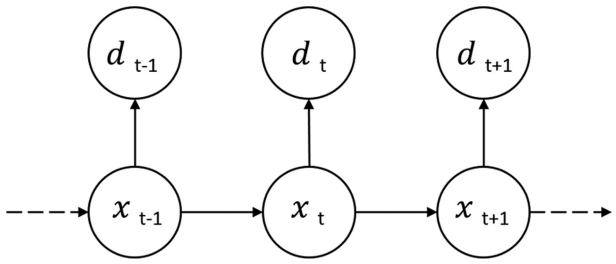


Figure 1. An illustration of a1-dimensional Hidden Markov chain (1D-HMM). Arrow directions represent the directions of probabilistic influence between hidden states x and observations d. Subscripts represent the index (time) of the corresponding state or observation.

followed in this paper. However, at a later stage in this paper, we define the notion of a partition of the set of cells represented by P, and we use a linear indexing of cells within a partition, such that $g_{P,i}$ (note the comma in the subscript) represents the geological facies in the ith cell within the partition P. We also define ordered relationships between cells, such as $<$ and $>$, which are described below.

2 MODEL

2.1 Hidden Markov chain (1D-HMM)

A stochastic process is a non-deterministic method to generate random variables as a function of an independent variable, such as time or space. A Hidden Markov Model (HMM) is a graphical model that represents a dual stochastic process: a stochastic process representing observations with an underlying stochastic process representing hidden states or model parameters. Hidden Markov Chains, or 1D-HMM, are 1-D representations of a more general class of HMM's and are used to represent probability distributions over sequences of observations (Fig. 1)—see Stratonovich (1960), Baum & Petrie (1966), Baum *et al*. (1970) and Baum (1972). The observations are produced by underlying unobserved (hidden) states that represent local (in time or space) instances of the underlying stochastic process. In a 1D-HMM, the observations and the underlying hidden states are indexed with a parameter t, which commonly refers to time but may also refer to some other measurement index such as space. A 1D-HMM assumes that the observation d_t at time t was generated by a stochastic process whose state x_t is hidden from the observer. It also assumes that the hidden states are sequentially distributed according to an underlying stochastic process that satisfies the *Markov property*: given the hidden state x_{t-1} at time $t-1$, the current state x_t at time t is conditionally independent of all of the previous states $x_1, \ldots, x_{t-2}$ at times prior to $t-1$:

$$\mathcal{P}(x_t \,|\, x_1, \ldots, x_{t-1}) = \mathcal{P}(x_t \,|\, x_{t-1})\,. \tag{2}$$

This means that an HMM is a memory-less process: the state x_{t-1} at time $t-1$ is assumed to encapsulate all relevant information from the history of the current state x_t at time t, and knowing the current state x_t is sufficient to generate the future states at times $t+1$ and beyond. Another fundamental assumption of an HMM is that for a given state, the observation from that state is conditionally independent of all other observations and hidden states in the model. This assumption is similar to the assumption of *localized likelihoods* and is called the *first-order Markov Property* with respect to data. HMM has found applications in computational biology (e.g. Krogh *et al*. 1994; Pachter *et al*. 2002; Liang *et al*. 2007; Yoon 2009; Shih *et al*. 2015), natural language processing (e.g. Collins 2002; Nivre 2002; Sun *et al*. 2012), speech recognition (e.g. Rabiner 1989; Dymarski 2011), computer vision (e.g. Li *et al*. 2000; Othman & Aboulnasr 2003; Baumgartner *et al*. 2013), earthquake seismology (e.g. Alasonati *et al*. 2006; Beyreuther & Wassermann 2008; Can *et al*. 2014), petroleum geoscience (e.g. Eidsvik *et al*. 2004; Lindberg & Omre 2014, 2015) and many other fields of research.

Previous work in the field of petroleum geoscience used Markov-chains and HMMs for inversion of seismic data for geological facies (e.g. Larsen *et al*. 2006; Ulvmoen & Omre 2010; Ulvmoen *et al*. 2010; Hammer & Tjelmeland 2011; Rimstad & Omre 2013; Lindberg & Omre 2014, 2015). Larsen *et al*. (2006) inverted pre-stack seismic data using a 1-D Markov-chain prior model of lithology-fluid classes along vertical profiles through a reservoir zone. Ulvmoen & Omre (2010) and Ulvmoen *et al*. (2010) extended the model of Larsen *et al*. (2006) by introducing lateral alignments among neighbouring 1-D vertical Markov-chains to model lateral coupling of lithology-fluid classes as commonly found in geological strata. Such a graphical structure is called a profile Markov random field (see e.g. Eddy 1998). Rimstad & Omre (2013) also used a profile Markov random field based prior but with a different parametrization. Rimstad *et al*. (2012) inverted seismic AVO data for lithology/fluid classes, elastic properties and porosity using an MRF prior model. Lindberg & Omre (2015) used a convolved two-level 1D-HMM for inversion of categorical variables (such as lithology-fluid classes) represented as the bottom hidden-layer of the model, continuous system response variables (such as reflection coefficients) represented as the middle hidden-layer, and the measured convolved data represented in the observation layer. A common feature among all of these approaches for facies inversion is that they are based on inference from full posterior distribution which must be explored through simulation (sampling) based inference, e.g. using McMC methods which suffer from the convergence and bias problems described earlier.

By contrast, in our approach, we derive analytic expressions for marginal posterior distributions of geological facies conditioned on the seismic attribute data using a 2-D Hidden Markov Model (2D-HMM) (Section 3.3 below) which is computationally efficient and allows for exact sampling (without using McMC) using a copula function (Section 4.1).

2.2 2-D Hidden Markov Model (2D-HMM)

Many extensions of hidden Markov-chains to 2-D have been proposed in the literature for applications to 2-D data such as images in computer vision, but these either convert 2-D data into 1-D and then apply a pseudo-2-D approach (e.g. Abend *et al*. 1965; Bevilacqua *et al*. 2007; Ma *et al*. 2008; Daleno *et al*. 2010), or attempt to obtain approximate results by introducing assumptions which limit the spatial dependence among neighbouring cells (locations) in the model (e.g. Li *et al*. 2000; Othman & Aboulnasr 2003; Baumgartner *et al*. 2013). The main contribution of this paper is that it presents analytic, closed-form solutions for approximate marginal posterior distributions of hidden geological states conditioned on the seismic data using a 2D-HMM which incorporates the full 2-D coupling of hidden states.

We build a 2D-HMM over a rectangular 2-D grid where hidden states correspond to the geological facies, and observations correspond to localized seismic attributes such as P- and S-wave impedances. In order to overcome the computational limitations of probabilistic dependence in 2-D, we assume that the geological

facies in a given cell is conditionally independent of the seismic attributes in cells outside of a certain predefined region around that cell, given the seismic attributes and geological facies within the pre-defined region. This assumption does not limit any spatial dependence among neighbouring locations, and is almost always valid for all types of models defined with the localized likelihoods assumption. The hidden states in a 2D-HMM form a special case of a Markov random field (MRF), called a *hidden Markov mesh* or a *causal MRF* (Abend *et al*. 1965; Cressie & Davidson 1998). Causality is induced in the grid by directional conditional dependence among the cells in the model, and allows the analytic derivation of marginal posterior distributions.

In a 2D-HMM, the geological facies are assumed to follow a geological model reflecting the joint spatial distribution of geologically plausible facies patterns over a rectangular 2-D grid. Marginal posterior distributions of facies were computed in each cell while incorporating prior knowledge about the spatial distribution of facies presented in the form of a training image, and the localized data likelihood is computed in each cell using neural networks similar to Shahraeeni & Curtis (2011) and Shahraeeni *et al*. (2012). We use a *conditional copula* function computed from the training image to use the marginal posterior distributions to draw random samples from the joint posterior distribution of facies in a predefined neighbourhood structure.

In this paper, we first introduce some definitions and notation which are used in the rest of the paper. These definitions allow a 1D-like treatment of the 2D-HMM, while fully acknowledging the higher dimensional spatial dependence among cells in the model. We analytically derive expressions for marginal posterior distributions at each location in the model given the data and the neighbouring geological facies, followed by the spatial sampling method using a conditional copula function. Then we present test results of spatial stochastic sampling and of computing marginal probability distributions from an application of this method to a synthetic geological model of siliciclastic-filled river channels in shale, with three geological facies—shale, brine-sand and gas-sand. We finally provide a brief discussion comparing this method with previous research with reference to the test results, before concluding.

3 COMPUTATION OF MARGINAL POSTERIOR DISTRIBUTION IN A 2D-HMM

3.1 Some relevant definitions

Herein a 2D-HMM is represented by a rectangular grid of cells with horizontal, vertical and diagonal neighbours defined for each cell in the grid. Causality is induced on the graph by defining ordered relationships that introduce the notions of past and future with respect to a given cell (e.g. see Fig. 2). The causality also defines the flow of probabilistic influence across all of the cells in the model. A 2D-HMM therefore defines a directed acyclic graph $\mathbb{G}(\boldsymbol{G}, \mathcal{E})$ over the set of vertices (or cells) $\boldsymbol{G}$ defining hidden states and $\mathcal{E}$ is the set of directed edges (relationships) where each edge connects exactly one vertex (or cell) to another in the graph. Being a directed graph, $\mathbb{G}$ defines a relationship $\rightarrow$ on its vertices such that for any two vertices x and y, the relationship $x \rightarrow y \in \mathcal{E}$ holds when there exists a direct path from x to y in $\mathbb{G}$. The graph $\mathbb{G}$ defines an order $<$ on its vertices such that for any two vertices x and y, $x < y$ when there exists an unblocked path from x to y in $\mathbb{G}$; that is, when

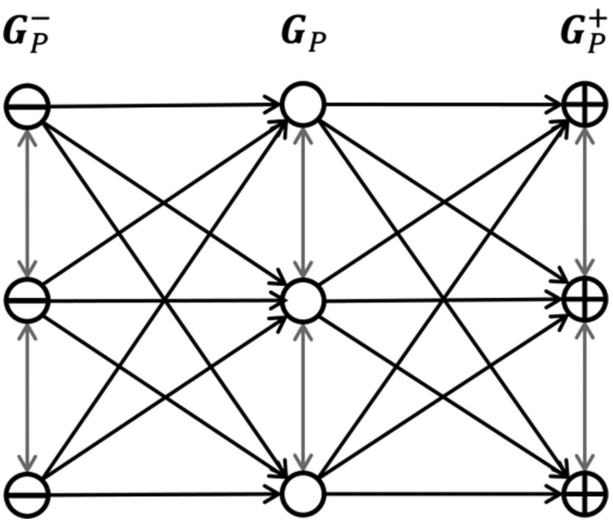


Figure 2. An example of a causal 2D-HMM. The arrow directions represent the directions of probabilistic dependence between various cells (circles) in the model. Circles with − sign, blank and + sign in their interior represent cells in the past partition $\boldsymbol{G}_P^-$, the current partition $\boldsymbol{G}_P$, and the future partition $\boldsymbol{G}_P^+$, respectively.

the probabilistic influence may flow from x to y either directly or indirectly through any other vertices. Similarly, $\mathbb{G}$ also defines a partial order $\leq$ which is similar to the order $<$ except that it also allows that $x = y$. The orders $>$ and $\geq$ are similarly defined such that $x > y$ implies that there exists no direct or indirect path from x to y in $\mathbb{G}$; and $x \geq y$ implies that either there is no direct or indirect path from x to y in $\mathbb{G}$, or that $x = y$. Since there is a one-to-one mapping between each cell $(i, j) \in \boldsymbol{G}$ in the model and the corresponding hidden state g_{ij}, we denote the cells in the model with the corresponding hidden state so we use the same notation $\boldsymbol{G}$ to denote the set of vertices (cells) in the model as well as the set of geological states in the model.

The *neighbourhood* of a cell x, denoted as $Ne(x)$, is the set of cells which are directly connected to cell x according to the underlying graphical model. This implies that the facies in a given cell is either directly influenced by the facies in the neighbouring cells, or it directly influences the facies in the neighbouring cells, or both. A cell x is said to be a *neighbour* of cell y if $x \in Ne(y)$. This also implies that $y \in Ne(x)$—in words, that y is also a neighbour of x. It should be noted that the cell x is not defined to be a neighbour of itself so $x \notin Ne(x)$. The definition of neighbourhood implies that a given cell is conditionally independent of all other cells outside of its neighbourhood

$$\mathcal{P}\left(g_{ij} \mid \boldsymbol{G}_{\setminus ij}\right) = \mathcal{P}\left(g_{ij} \mid \boldsymbol{G}_{Ne(i,j)}\right), \tag{3}$$

where $\mathcal{P}$ is probability, g_{ij} represents geological facies in some cell $(i, j) \in \boldsymbol{G}$, $\boldsymbol{G}_{\setminus ij}$ represents the set of geological facies in all cells in the model except cell (i, j), and $\boldsymbol{G}_{Ne(i,j)}$ represents the set of geological facies in all cells in the neighbourhood of cell (i, j).

Define *neighbourhood cardinality*, denoted as $\mathcal{C}_{Ne}(x)$, as the number of cells in the neighbourhood of a given cell x. The neighbourhood cardinality of a cell at the boundary of a model is usually lower than that of a cell further away from the model boundaries.

Now define a *partition element* or simply a *partition* as a non-empty ordered set of cells (nodes) where each cell is a neighbour of the next cell in the set, and the first and last cells in the set

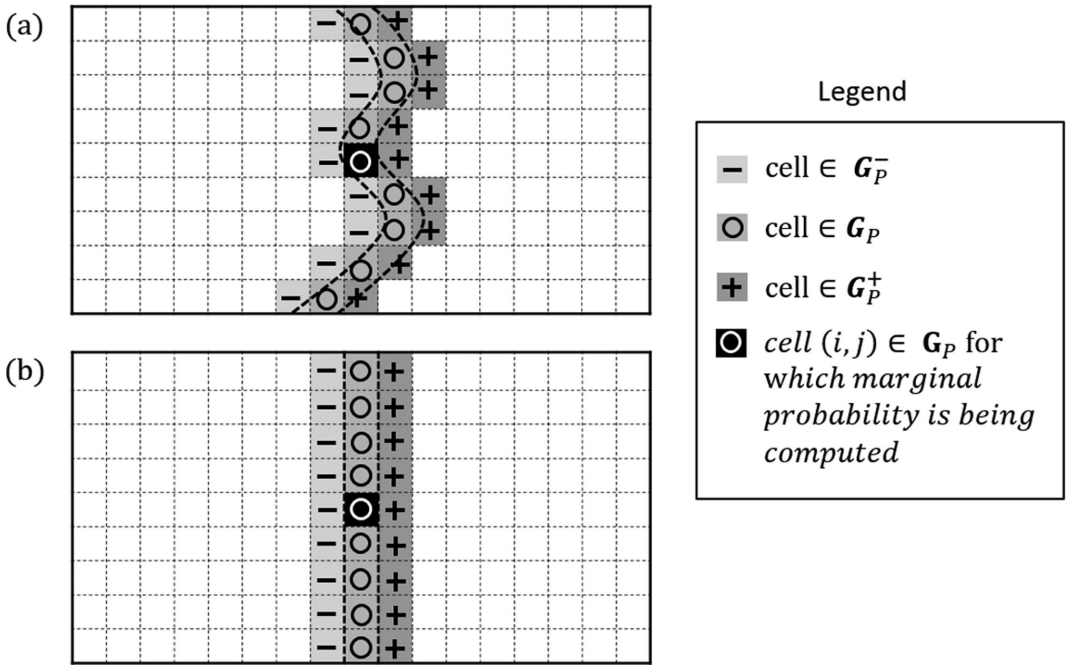


Figure 3. Examples of a partition defined over a graphical model as a set of nodes that divides the model into two non-overlapping parts. An ordered set of such partitions defines a partition family over the graphical model. (a) An arbitrary partition, and (b) a partition (element) defined over a column (or the shorter dimension) in the model. The cells shown with symbol '−' form the previous (past) partition $\boldsymbol{G}_P^-$, the cells shown with symbols 'o' form the current partition $\boldsymbol{G}_P$, and the cells shown with symbol '+' form the next (future) partition $\boldsymbol{G}_P^+$. The black 'o' cell represents the cell $(i, j) \epsilon \boldsymbol{G}_P$ for which the marginal posterior probability is being computed.

lie at the boundary of the model. A non-empty ordered set of disjoint partitions can be defined such that all of the neighbours of any cell in one partition lie either in the same, the previous or the next partition. Such a family of non-empty, ordered, disjoint partition elements defines a *partition family* over the (graphical) model.

In order to create a recursive formulation of the marginal posterior distribution $\mathcal{P}(g_{ij} \mid \boldsymbol{D})$ of facies g_{ij} in a cell $(i, j) \in \boldsymbol{G}$ conditioned to data $\boldsymbol{D}$, we define a partition $\boldsymbol{G}_P$ as a set of cells ordered from 1 to n, such that $g_{ij} \in \boldsymbol{G}_P$; that is, $\exists g_{P,k} \in \boldsymbol{G}_P$ such that $g_{P,k} = g_{ij}$, for some k, and

$$\boldsymbol{G}_P = \left\{ g_{P,1}, g_{P,2}, \ldots, g_{P,k}, \ldots, g_{P,n} : g_{P,k} = g_{ij} \text{ for some } k \right\}. \tag{4}$$

Just as the notation with double letters ij in the subscript represents the location of a cell in the model, the notation $g_{P,\bullet}$ (with a $P, \bullet$ in the subscript) is used in eq. (4) to represent ordering of cells within the partition $\boldsymbol{G}_P$. Since by definition of a partition, the first and the last cell (e.g. $g_{P,1}$ and $g_{P,n}$) must lie on the boundary of the model, we must have $\mathcal{C}_{Ne}(g_{P,1}) < c_{\max}$ and $\mathcal{C}_{Ne}(g_{P,n}) < c_{\max}$, where $c_{\max}$ is the maximum neighbourhood cardinality of a cell inside the model boundaries. So by definition of a partition $g_{P,1} \in Ne(g_{P,2})$, $g_{P,2} \in Ne(g_{P,3})$, $\ldots$, $g_{P,n-1} \in Ne(g_{P,n})$.

A partition may be defined as a row, a column or an arbitrary set of cells that satisfies eq. (4). The shape of the partition is chosen with consideration of computational convenience, the size and shape of the computational model, and the neighbourhood structure. It is preferable to define the partition along the shorter dimension of the model in order to limit the partition size, as the memory required to store the joint distribution of facies within a partition grows exponentially with the partition size.

We define $\boldsymbol{G}_P^-$ as the set of cells which constitute the immediate past of the partition $\boldsymbol{G}_P$ based on the direction induced by causality (Fig. 3)

$$\boldsymbol{G}_P^- = \left\{ g_{kl} : \exists g_{ij} \in \boldsymbol{G}_P, \text{ such that } g_{kl} \rightarrow g_{ij} \in \mathcal{E} \right\}. \tag{5}$$

Similarly, define $\boldsymbol{G}_P^+$ as the set of cells which constitute the immediate future of the partition $\boldsymbol{G}_P$ based on the direction induced by causality (Fig. 3)

$$\boldsymbol{G}_P^+ = \left\{ g_{kl} : \exists g_{ij} \in \boldsymbol{G}_P, \text{ such that } g_{ij} \rightarrow g_{kl} \in \mathcal{E} \right\}. \tag{6}$$

By definition $\boldsymbol{G}_P^- \cap \boldsymbol{G}_P = \boldsymbol{G}_P \cap \boldsymbol{G}_P^+ = \boldsymbol{G}_P^- \cap \boldsymbol{G}_P^+ = \emptyset$ and $\boldsymbol{G} = \cup_P \boldsymbol{G}_P, \forall P$, where $\cap$ represents intersection, $\cup_P$ represents union over all P, and $\emptyset$ is the empty set.

Also, define $\boldsymbol{G}_{\leq P}$ as

$$\boldsymbol{G}_{\leq P} = \left\{ g_{kl} : \exists g_{ij} \in \boldsymbol{G}_P, \text{ such that } g_{kl} \leq g_{ij} \right\}. \tag{7}$$

It follows that $\boldsymbol{G}_P \subset \boldsymbol{G}_{\leq P}$. Similarly, define $\boldsymbol{G}_{>P}$ as

$$\boldsymbol{G}_{>P} = \left\{ g_{kl} : \exists g_{ij} \in \boldsymbol{G}_P, \text{ such that } g_{ij} < g_{kl} \right\} \tag{8}$$

Fig. 3 shows examples of partitions $\boldsymbol{G}_P^-$, $\boldsymbol{G}_P$ and $\boldsymbol{G}_P^+$ defined (a) arbitrarily, and (b) as a column of cells in the model. Fig. 4 shows the regions corresponding to $\boldsymbol{G}_{\leq P}$ and $\boldsymbol{G}_{>P}$ for a partition defined as a column of cells in the model. Similarly define

$$\boldsymbol{D}_P = \left\{ \boldsymbol{d}_{kl} : \exists g_{kl} \in \boldsymbol{G}_P, \text{ such that } \mathcal{P}(\boldsymbol{d}_{kl} \mid \boldsymbol{G}) = \mathcal{P}(\boldsymbol{d}_{kl} | g_{kl}) \right\} \tag{9}$$

$$\boldsymbol{D}_{\leq P} = \left\{ \boldsymbol{d}_{kl} : \exists g_{kl} \in \boldsymbol{G}_{\leq P}, \text{ such that } \mathcal{P}(\boldsymbol{d}_{kl} \mid \boldsymbol{G}) = \mathcal{P}(\boldsymbol{d}_{kl} | g_{kl}) \right\} \tag{10}$$

$$\boldsymbol{D}_{>P} = \left\{ \boldsymbol{d}_{kl} : \exists g_{kl} \in \boldsymbol{G}_{>P}, \text{ such that } \mathcal{P}(\boldsymbol{d}_{kl} \mid \boldsymbol{G}) = \mathcal{P}(\boldsymbol{d}_{kl} | g_{kl}) \right\}. \tag{11}$$

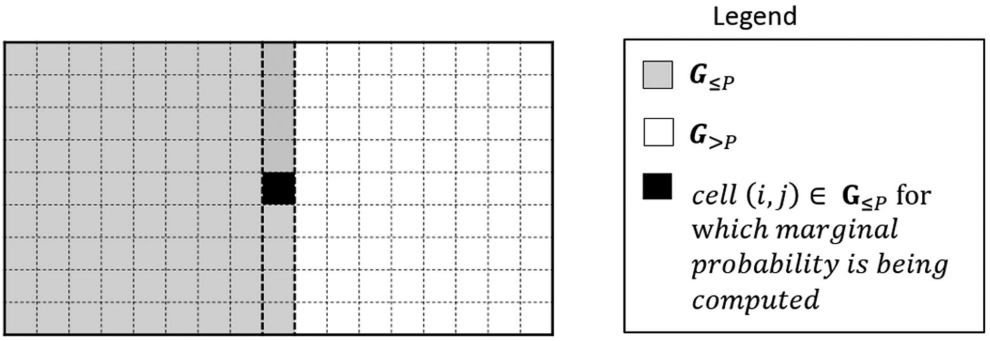


Figure 4. Illustration of partitions $\boldsymbol{G}_{\leq P}$ and $\boldsymbol{G}_{>P}$. The cells with dashed border represent the partition $\boldsymbol{G}_P$, which are also part of partition $\boldsymbol{G}_{\leq P}$.

A key assumption in our algorithm for computing marginal posterior distributions using a 2D-HMM is that $\boldsymbol{D}_{>P}$ and $\boldsymbol{D}_{\leq P}$ are conditionally independent given the geology $\boldsymbol{G}$. Note that the data $\boldsymbol{d}_{kl}$ in a cell (k, l) may be vector, hence the bold-face notation.

3.2 Conditional dependence between partitions

Due to causality, the probability of a facies being present in a given cell depends on the facies in the previous partition $\boldsymbol{G}_P^-$ as well as in the current partition $\boldsymbol{G}_P$. Such a dependence can be computed from the conditional probabilities of facies in the current partition $\boldsymbol{G}_P$ given the facies in the previous partition $\boldsymbol{G}_P^-$. From the above definitions it follows that (allowing for different numbers of cells in $\boldsymbol{G}_P$ and $\boldsymbol{G}_P^-$ in general)

$$\mathcal{P}\left(\boldsymbol{G}_P \mid \boldsymbol{G}_P^-\right) = \mathcal{P}\left(g_{P,1}, g_{P,2}, \ldots, g_{P,n} \mid g_{P,1}^-, g_{P,2}^-, \ldots, g_{P,r}^-\right),$$

where $g_{P,\bullet}$ and $g_{P,\bullet}^-$ represent the ordering of elements within partitions $\boldsymbol{G}_P$ and $\boldsymbol{G}_P^-$ respectively. Then,

$$\begin{aligned}
&\mathcal{P}\left(\boldsymbol{G}_P \mid \boldsymbol{G}_P^-\right) \\
&\quad = \mathcal{P}\left(g_{P,1} \mid g_{P,2}, \ldots, g_{P,n}, g_{P,1}^-, g_{P,2}^-, \ldots, g_{P,r}^-\right) \\
&\qquad \cdot \mathcal{P}\left(g_{P,2}, \ldots, g_{P,n} \mid g_{P,1}^-, g_{P,2}^-, \ldots, g_{P,r}^-\right) \\
&\quad = \mathcal{P}\left(g_{P,1} \mid \boldsymbol{G}_{Ne}^-(P,1)\right) \\
&\qquad \cdot \mathcal{P}\left(g_{P,2}, \ldots, g_{P,n} \mid g_{P,1}^-, g_{P,2}^-, \ldots, g_{P,r}^-\right) \\
&\quad = \mathcal{P}\left(g_{P,1} \mid \boldsymbol{G}_{Ne}^-(P,1)\right) \\
&\qquad \cdot \mathcal{P}\left(g_{P,2} \mid g_{P,3}, \ldots, g_{P,n}, g_{P,1}^-, g_{P,2}^-, \ldots, g_{P,r}^-\right) \\
&\qquad \cdot \mathcal{P}\left(g_{P,3}, \ldots, g_{P,n} \mid g_{P,1}^-, g_{P,2}^-, \ldots, g_{P,r}^-\right) \\
&\quad = \mathcal{P}\left(g_{P,1} \mid \boldsymbol{G}_{Ne}^-(P,1)\right) \cdot \mathcal{P}\left(g_{P,2} \mid \boldsymbol{G}_{Ne}^-(P,2) \setminus \{g_{P,1}\}\right) \\
&\qquad \cdot \mathcal{P}\left(g_{P,3} \mid g_{P,4}, \ldots, g_{P,n}, g_{P,1}^-, g_{P,2}^-, \ldots, g_{P,r}^-\right) \\
&\qquad \cdot \mathcal{P}\left(g_{P,4}, \ldots, g_{P,n} \mid g_{P,1}^-, g_{P,2}^-, \ldots, g_{P,r}^-\right) \\
&\quad = \mathcal{P}\left(g_{P,1} \mid \boldsymbol{G}_{Ne}^-(P,1)\right) \cdot \mathcal{P}\left(g_{P,2} \mid \boldsymbol{G}_{Ne}^-(P,2) \setminus \{g_{P,1}\}\right) \\
&\qquad \cdot \mathcal{P}\left(g_{P,3} \mid \boldsymbol{G}_{Ne}^-(P,3) \setminus \{g_{P,1}, g_{P,2}\}\right) \\
&\qquad\quad \cdots \\
&\qquad \cdot \mathcal{P}\left(g_{P,n} \mid \boldsymbol{G}_{Ne}^-(P,n) \setminus \{g_{P,1}, g_{P,2}, \ldots, g_{P,n-1}\}\right)
\end{aligned}$$

and therefore,

$$\mathcal{P}\left(\boldsymbol{G}_P \mid \boldsymbol{G}_P^-\right) = \prod_{i=1}^{n} \mathcal{P}\left(g_{P,i} \mid \boldsymbol{G}_{Ne}^-(P,i) \setminus \{g_{P,<i}\}\right) \tag{12}$$

where $\{g_{P,<i}\} = \{g_{P,h} : h < i\}$ and $\boldsymbol{G}_{Ne}^-(P,i) = (\boldsymbol{G}_P^- \cup \boldsymbol{G}_P) \cap Ne(P,i)$.

The conditional probabilities on the right-hand side of eq. (12) represent the prior information on the spatial correlation of geological facies. These can be computed directly from the patterns of facies distributions depicted in the training image which correspond to the various shapes and sizes of partitions and the neighbourhood structure. As an example, see Toftaker & Tjelmeland (2013) for a proposed method to build a prior model from a training image using a binary MRF and its partially ordered approximation, and Arnesen & Tjelmeland (2016) for a proposed prior distribution for parameters and structure of a binary MRF. For spatial inversion of geological facies, we stipulate that the spatial correlations read from the training image are assumed to be stationary, that is, they are assumed to be independent of location within the model.

3.3 Derivation of marginal posterior distribution

The idea of a partition $\boldsymbol{G}_P$ thus imposes a natural ordering which (in the following) allows 1-D-like treatment of the underlying 2D-HMM while fully acknowledging the 2-D structure of probabilistic dependence between cells in the model. Using the above definitions, we can derive the recursive formulation for the marginal posterior distribution conditioned to the data $\boldsymbol{D}$ because $\mathcal{P}(g_{ij} \mid \boldsymbol{D}) \propto \mathcal{P}(g_{ij}, \boldsymbol{D})$ since the data $\boldsymbol{D}$ is measured and fixed. Setting $g_{ij} = g_{P,q}$,

$$\begin{aligned}
\mathcal{P}(g_{ij} \mid \boldsymbol{D}) &\propto \mathcal{P}(g_{P,q}, \boldsymbol{D}) \\
&= \sum\nolimits_{G_P \setminus \{g_{P,q}\}} \mathcal{P}(\boldsymbol{G}_P, \boldsymbol{D}) \\
&\qquad \text{[by definition of a marginal distribution over } g_{P,q}\text{]} \\
&= \sum\nolimits_{G_P \setminus \{g_{P,q}\}} \mathcal{P}(\boldsymbol{G}_P, \boldsymbol{D}_{\leq P}) \mathcal{P}(\boldsymbol{D}_{>P} \mid \boldsymbol{G}_P) \\
&\qquad \text{[since } \boldsymbol{D}_{\leq P} \text{ is independent of } D_{>P}\text{]} \\
&= \sum\nolimits_{\boldsymbol{G}_P \setminus \{g_{P,q}\}} \alpha(\boldsymbol{G}_P)\, \beta(\boldsymbol{G}_P)
\end{aligned} \tag{13}$$

where $\alpha(\boldsymbol{G}_P) = \mathcal{P}(\boldsymbol{G}_P, \boldsymbol{D}_{\leq P})$ and $\beta(\boldsymbol{G}_P) = \mathcal{P}(\boldsymbol{D}_{>P} \mid \boldsymbol{G}_P)$ are the equivalent 2-D forward and backward probabilities of those used for 1-D hidden Markov chains in the dynamic programming based algorithms of Baum & Petrie (1966), Viterbi (1967), Baum *et al.* (1970), Baum (1972) and Forney Jr. (1973). Note that the summation in eq. (13) represents summations over all of the cells $g_{P,\bullet}$ in the partition $\boldsymbol{G}_P$ except the cell $g_{ij} = g_{P,q}$.

Since $\boldsymbol{G}_P$'s, by definition, form a partition over the model space, $\alpha(\boldsymbol{G}_P)$ can be expressed using the recursive formulation of Baum's

forward-backward algorithm (Baum 1972) for a 1-D hidden Markov chain as

$$\alpha(\boldsymbol{G}_P) = \mathcal{P}(\boldsymbol{G}_P, \boldsymbol{D}_{\leq P}) = \mathcal{P}(\boldsymbol{D}_P \mid \boldsymbol{G}_P) \sum_{\boldsymbol{G}_P^-} \mathcal{P}\left(\boldsymbol{G}_P \mid \boldsymbol{G}_P^-\right) \alpha\left(\boldsymbol{G}_P^-\right) \quad (14)$$

where summation is over all of the facies in all of the cells in $\boldsymbol{G}_P^-$. On substitution from eq. (12) for $\mathcal{P}(\boldsymbol{G}_P \mid \boldsymbol{G}_P^-)$ and assuming localized likelihoods, eq. (14) takes the form

$$\alpha(\boldsymbol{G}_P) = \prod_{i=1}^{n} \mathcal{P}(\boldsymbol{d}_{P,i} \mid g_{P,i}) \cdot \sum_{\boldsymbol{G}_P^-} \left(\prod_{j=1}^{n} \mathcal{P}\left(g_{P,j} \mid \boldsymbol{G}_{Ne}^-(P, j) \setminus \{g_{P,<j}\}\right)\right) \cdot \alpha\left(\boldsymbol{G}_P^-\right) \quad (15)$$

The factors $\mathcal{P}(\boldsymbol{D}_P \mid \boldsymbol{G}_P) = \prod_{i=1}^{n} \mathcal{P}(\boldsymbol{d}_{P,i} \mid g_{P,i})$ in eq. (15) represent the data likelihood given the geological facies in each cell assuming localized likelihoods. The data likelihood is given by the probabilistic forward model and is explained in the next section.

Similarly, $\beta(\boldsymbol{G}_P)$ can be expressed in a recursive formulation as

$$\beta(\boldsymbol{G}_P) = \mathcal{P}(\boldsymbol{D}_{>P} \mid \boldsymbol{G}_P) = \sum_{\boldsymbol{G}_P^+} \mathcal{P}\left(\boldsymbol{G}_P^+ \mid \boldsymbol{G}_P\right) \mathcal{P}\left(\boldsymbol{D}_P^+ \mid \boldsymbol{G}_P^+\right) \beta\left(\boldsymbol{G}_P^+\right) \quad (16)$$

where the summation is over all possible combinations of facies in all of the cells in $\boldsymbol{G}_P^+$. On substitution from eq. (12) for $\mathcal{P}(\boldsymbol{G}_P^+ \mid \boldsymbol{G}_P)$ and assuming localized likelihoods, eq. (16) takes the form

$$\beta(\boldsymbol{G}_P) = \sum_{\boldsymbol{G}_P^+} \left(\left(\prod_{i=1}^{n} \mathcal{P}(\boldsymbol{d}_{P^+,i} \mid g_{P^+,i}) \right) \cdot \left(\prod_{j=1}^{n} \mathcal{P}\left(g_{P^+,j} \mid \boldsymbol{G}_{Ne}^-(P^+, j) \backslash \{g_{P^+,>j}\}\right)\right) \cdot \beta(\boldsymbol{G}_P^+) \right). \quad (17)$$

Substituting eqs (15) and (17) into eq. (13) gives a recursive formulation for the marginal posterior distribution in a given cell in the model. $\alpha(\boldsymbol{G}_P)$ in eq. (15) is computed in the forward direction (increasing P) while $\beta(\boldsymbol{G}_P)$ in eq. (17) is computed in the backward direction (decreasing P). This process is repeated for each cell of interest (i, j) in the model.

The number of summations in eqs (13)–(17) increases exponentially with the model size. This means that a naïve recursive computation of forward and backward probabilities becomes intractable for models of practical size. In order to limit the computational time and memory, an approximate marginal posterior distribution can be obtained by limiting the grid size considered around each cell. This introduces a practical and fundamental assumption that the facies at a given cell (i, j) is conditionally independent of data observed at locations outside of a certain region of influence $\mathcal{R}(i, j)$ around the cell (i, j), that is

$$\mathcal{P}\left(g_{ij} \mid \boldsymbol{D}\right) = \mathcal{P}\left(g_{ij} \mid \boldsymbol{D}_{\mathcal{R}(i,j)}\right) \quad (18)$$

where $\boldsymbol{D}_{\mathcal{R}(i,j)}$ represents the set of data within $\mathcal{R}(i, j)$. Fig. 5 shows an illustration of the concept of a region of influence. Also, the choice of size of a partition allows us to further limit the number of summations required in eqs (13)–(17) by summing over only the plausible geological facies, for example, by summing over only those facies configurations which are found in the training image. This was achieved by directly scanning the training image for the facies configurations in a manner similar to that used in so-called direct sampling (Mariethoz *et al*. 2010). Tjelmeland & Austad (2012) used a different approach to approximate recursive calculations in a binary MRF by approximating the interaction parameters between neighbouring nodes to zero when they are very small.

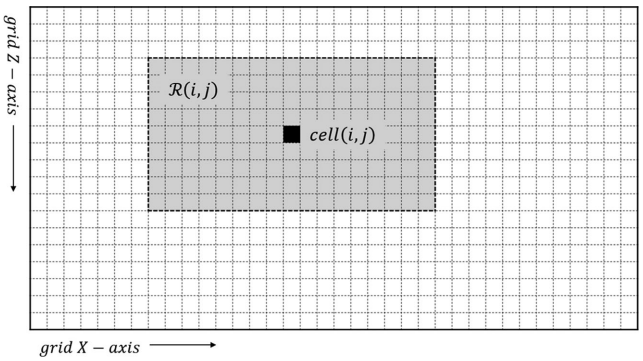


Figure 5. Illustration of the complete model, region of influence $\mathcal{R}(i, j)$ shown as grey shaded cells defined over a subset of the model around cell (i, j), and the cell (i, j) shown in black colour. The marginal posterior distribution of cell (i, j) is computed with the assumption that the facies at cell (i, j) depend on the facies in the neighbouring cells which in turn depend on their neighbours and so on. Thus the facies at cell (i, j) show a spatial correlation with facies across the complete model. The assumption of $\mathcal{R}(i, j)$ in the algorithm, however, removes the conditional dependence of the facies at (i, j) on data observed at locations outside of this region.

4 GENERATING RANDOM SAMPLES—STOCHASTIC REALIZATIONS

The theory above shows how posterior probability of each facies in each cell can be computed directly, rather than being estimated indirectly through sampling algorithms such as in McMC. However, in case samples are desired, we now show how stochastic realizations of the joint distribution can be computed from the marginal distributions using the copula function determined from the prior continuity model (in our case, from the statistical information in the training image). Before we describe the sampling method, we give a brief introduction to copula functions below.

4.1 Copula

A *copula* is a function C of d variables on the unit d-cube $[0, 1]^d$ representing a multivariate distribution of any random vector $\boldsymbol{U} = (u_1, \ldots, u_d)$ that has univariate standard Uniform marginal distributions (on the unit d-cube). A copula therefore represents a multivariate probability distribution with standard Uniform marginal (univariate) distributions, and describes the complete joint dependence among the various random variables in $\boldsymbol{U}$ (Frechet 1951).

Sklar's theorem states that a joint multivariate cumulative distribution function (CDF) equals the copula function of all univariate marginal CDFs (Sklar 1959; Sklar 1973). In other words, any multivariate distribution can be completely described by its copula and its marginal distributions. This implies that for the joint CDF $F(\boldsymbol{Z})$ of a d-variate random vector $\boldsymbol{Z} = (z_1, \ldots, z_d)$ with marginal CDFs $F_1(z_1), \ldots, F_d(z_d)$, there exists a d-copula C (Jaworski *et al*. 2010) such that

$$F(\boldsymbol{Z}) = C(F_1(z_1), \ldots, F_d(z_d)) \quad (19)$$

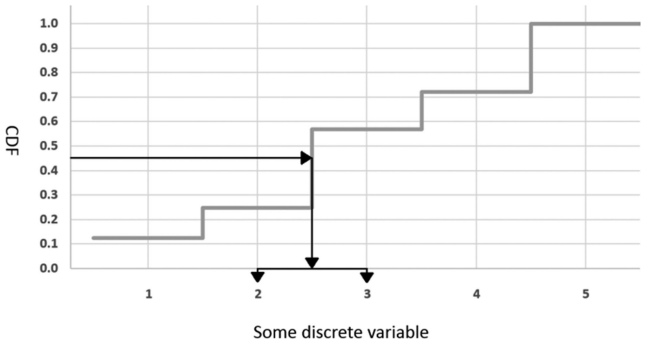


Figure 6. Illustration of possible non-unique values of the inverse-CDF for discrete variables. A cumulative probability of 0.46 cannot be uniquely assigned to a discrete (integer) variable in the interval [1, 5] since the CDF of variable 2 is 0.25 (<0.46) and that of variable 3 is 0.57 (>0.46). Use of the infimum function for this case assigns the value 3 which is the minimum value of the discrete variable for which the cumulative probability is ≥0.46. Another possibility could be to assign the output of the inverse-CDF equal to the variable whose cumulative probability is closest to the input value 0.46. In this example, the result would still be 3. It is therefore important to use any preferred definition for the inverse-CDF for discrete variables consistently.

which can be defined as follows. If $F_1^{-1}, F_2^{-1}, \ldots, F_d^{-1}$ represent inverse marginal distributions defined as

$$F_i^{-1}(x) = \inf\{y \mid F_i(y) \geq x\} \tag{20}$$

where inf is the *infimum* function, which provides the largest possible lower bound for the set on the right side, then for $\boldsymbol{U} = (u_1, u_2, \ldots, u_d)$ in the unit d-cube,

$$C(\boldsymbol{U}) = F\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right) \tag{21}$$

Eq. (21) shows that a copula expresses the conditional dependence of multivariate random variables in terms of conditional dependence of *Uniformly* distributed random variables. A copula can be determined uniquely if all of the marginal distributions are continuous. If any of the marginal distributions is discrete, the copula function becomes non-unique as the inverse distribution of a discrete variable cannot be uniquely determined (see Fig. 6). Nevertheless, the infimum function in eq. (20) defines a consistent way of computing inverse distribution functions (and hence copula functions) for discrete variables.

Since a copula encompasses only the joint dependence of multivariate random variables, it remains unchanged even if the scale or the location of their joint distribution changes. So, given a copula and the marginal distributions of individual random variables, the complete joint dependence structure embodied in the copula may be harnessed to yield multivariate samples from the univariate marginal distributions. Multivariate random numbers $\boldsymbol{Z} = (z_1, z_2, \ldots, z_D)$ can be generated from their marginal distributions by generating multivariate random variables $\boldsymbol{U} = (u_1, u_2, \ldots, u_D)$ with *Uniform* marginal distributions from the copula function (see Algorithm 1), then applying the inverse marginal distribution function to each component of U independently. A copula, therefore, allows the joint multivariate probability distribution to be reconstructed from the known univariate marginal distributions of individual variables.

A *conditional copula* function $C_d(u_d \mid u_1, \ldots, u_{d-1})$ represents a univariate distribution function for a random variable $u_d \in [0, 1]$ given the Uniform random variables $u_i \in [0, 1], i = 1, \ldots, d-1$. The inverse copula function C_d^{-1} can be computed from the copula function C_d using eq. (20). A conditional copula function allows a random variable to be sampled from a combination of its marginal probability distribution and the marginal probability distributions of conditioning variables. We illustrate this with the following example.

Example Consider two random variables z_1 and z_2, a conditional copula function $C_2(u_2 \mid u_1)$ which encodes the probability distribution $\mathcal{P}(z_2 \mid z_1)$, and marginal probabilities $F_1(z_1)$ and $F_1(z_2)$. We can generate samples from $\mathcal{P}(z_2 \mid z_1)$ which honour the marginal distributions as well as the joint distribution of the two variables as follows:

1. Compute $u_1 = F_1(z_1)$
2. Draw a random sample $u \sim U(0, 1)$
3. Compute $u_2 = C_2^{-1}(u \mid u_1)$, where C_2^{-1} is the inverse conditional copula defined by eq. (20)
4. Compute $z_2 = F_2^{-1}(u_2)$

We can extend the above sampling procedure for two variables to a more general case of d variables. Given a conditional copula function $C_d(u_d \mid u_1, \ldots, u_{d-1})$, marginal distributions $F_i(z_i), i = 1, \ldots, d$, and known random variables $z_1, \ldots, z_{d-1}$, conditional samples for the random variable z_d may be generated using the following algorithm:

Algorithm 1 Generation of conditional samples for z_d from the conditional distribution $F_d(z_d \mid z_1, \ldots, z_{d-1})$ using conditional copula function $C_d(u_d \mid u_1, \ldots, u_{d-1})$ and marginal distributions $F_i(z_i), i = 1, \ldots, d$:

1. Compute $u_i = F_i(z_i), u_i \in [0, 1], i = 1, \ldots, d-1$
2. Draw a random sample $u \sim U(0, 1)$
3. Compute $u_d = C_d^{-1}(u \mid u_i), i = 1, \ldots, d-1$, where C_d^{-1} is the inverse conditional copula defined by eq. (20)
4. Compute $z_d = F_d^{-1}(u_d)$

4.2 Sampling from computed marginal distributions using a conditional copula function

A conditional copula function can be derived from the joint spatial distribution of geological facies depicted in the training image within a predefined *neighbourhood* structure around each cell in the model. The geological facies in a given cell may then be simulated from the marginal posterior distribution in that cell conditioned on the previously sampled neighbouring cells, using the algorithm described above. The samples of facies generated in this fashion are conditionally dependent on the facies in the neighbourhood of each of the sampling locations. Stochastic posterior realizations of the geological cross-section may be generated by sequentially drawing random samples for each cell in the model following any linear path.

5 SYNTHETIC TEST

In order to test the algorithm and to benchmark it against pre-existing algorithms we applied it to the same synthetic inverse problem as was used by Walker & Curtis (2014). The synthetic example is based on two 2-D geological cross-sections extracted from a 3-D geological process model of channels with filled and overbank sand deposits emplaced in background shale. Most of the channels are filled with brine. Gas is introduced in some of the channels while obeying gravitational ordering of the two fluids. The sample space of the facies in each cell is therefore given by

$$\mathcal{G} = \{\text{shale, brine-sand, gas-sand}\}.$$

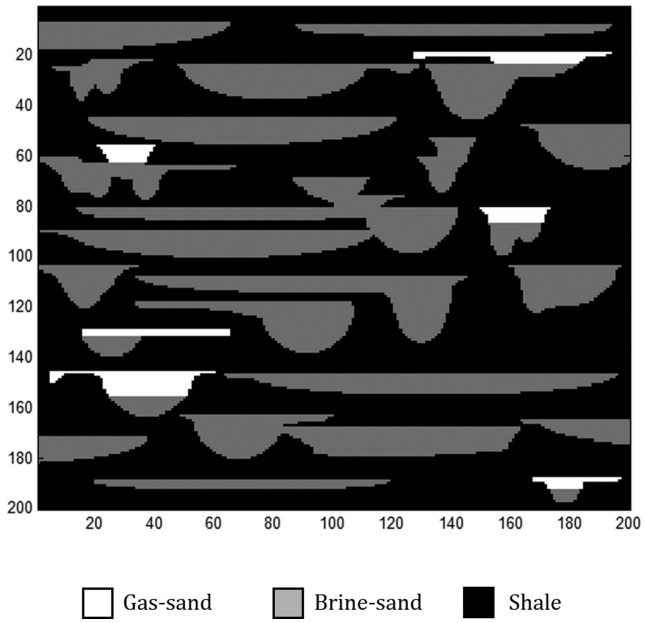


Figure 7. Training image (TI) extracted as a 2-D cross-section from a 3-D geological process model containing channels with filled and overbank sand deposits with shale in the background. The sand is filled with brine or gas, obeying gravitational ordering of the two fluids. The training image represents only a conceptual picture that preserves typical forms of expected types of structures and facies. It lacks any location-specific information about the real geology of the target subsurface in Fig. 8(a). It was used to obtain prior information related to the lateral and vertical continuity and association of various geological facies in the form of the conditional probability distributions $\mathcal{P}(\boldsymbol{G}_P \mid \boldsymbol{G}_P^-) = \prod_{i=1}^{n} \mathcal{P}(g_{P,i} \mid \boldsymbol{G}_{Ne}^-(P,i) \backslash \{g_{P,<i}\})$ in eq. (12).

One of the geological cross-sections (with dimensions of 200 $\times$ 200 model cells) defines the training image (Fig. 7), while the other was used as a target cross-section (with dimensions of 100 $\times$ 100 model cells) representing the true Earth (Fig. 8a). The training image was used to define the prior spatial conditional distributions of facies and the conditional copula function in order to be able to sample from the marginal posterior distributions computed using the main algorithm. The size of the region of influence $\mathcal{R}$(i, j) was arbitrarily taken to be 7 and 9 model cells in each dimension and the partition G_P was defined as a column of seven cells. The size of the partition was chosen arbitrarily whereas its shape was chosen with computational convenience in mind. The conditional copula function was computed for a square 3 $\times$ 3 neighbourhood template for $Ne(\cdot)$. This allowed us to sample the joint posterior distribution within a 3 $\times$ 3 matrix of cells given the marginal posterior distributions in each of these cells. The size of the neighbourhood template could easily be increased without incurring any computational limitations. While a 3 $\times$ 3 neighbourhood template may not be sufficient to reproduce more complex patterns of facies distributions such as a map view of channels in a deltaic environment, we found that the shape of geological features in the target cross-section could easily be recovered with a 3 $\times$ 3 neighbourhood template.

The prior information is extracted from the training image in the form of prior probabilities $\mathcal{P}(g_{ij} \mid \boldsymbol{G}_{Ne(i,j)})$ and $\mathcal{P}(\boldsymbol{G}_P \mid \boldsymbol{G}_P^-)$. The expression $\mathcal{P}(g_{ij} \mid \boldsymbol{G}_{Ne(i,j)})$ represents the probability of existence of a facies g_{ij} in a cell $(i, j) \in \boldsymbol{G}$ given facies configuration $\boldsymbol{G}_{Ne(i,j)}$ in the neighbourhood of cell (i, j), and $\mathcal{P}(\boldsymbol{G}_P \mid \boldsymbol{G}_P^-)$ represents the spatial correlation of facies configurations in consecutive partitions G_P^- and G_P. We assume that the prior information extracted from the training image is stationary over the model space and the probabilities computed therefrom encapsulate the expected spatial correlations of facies. In order to confirm that, we generated realizations from prior probabilities (see Fig. 9). Given that the prior realizations were generated using a partition of size 7 cells along a column, we expect the prior information (and hence these realizations) to preserve small-scale geometrical features and fluid orderings but not the large-scale shapes of the channels. In Fig. 9 we observe that this is the case. Where gas exists it is never beneath oil, flat tops of channels are preserved, but the overall semi-circular valley-style channel shape is not. This means that by incorporating prior information we are only constraining the spatial correlations of various facies, and not the shapes of the channels – which ideally should come from the data likelihoods. If so, the prior probabilities combined with the data-derived likelihoods might produce subsurface structures with geologically plausible spatial correlations of facies.

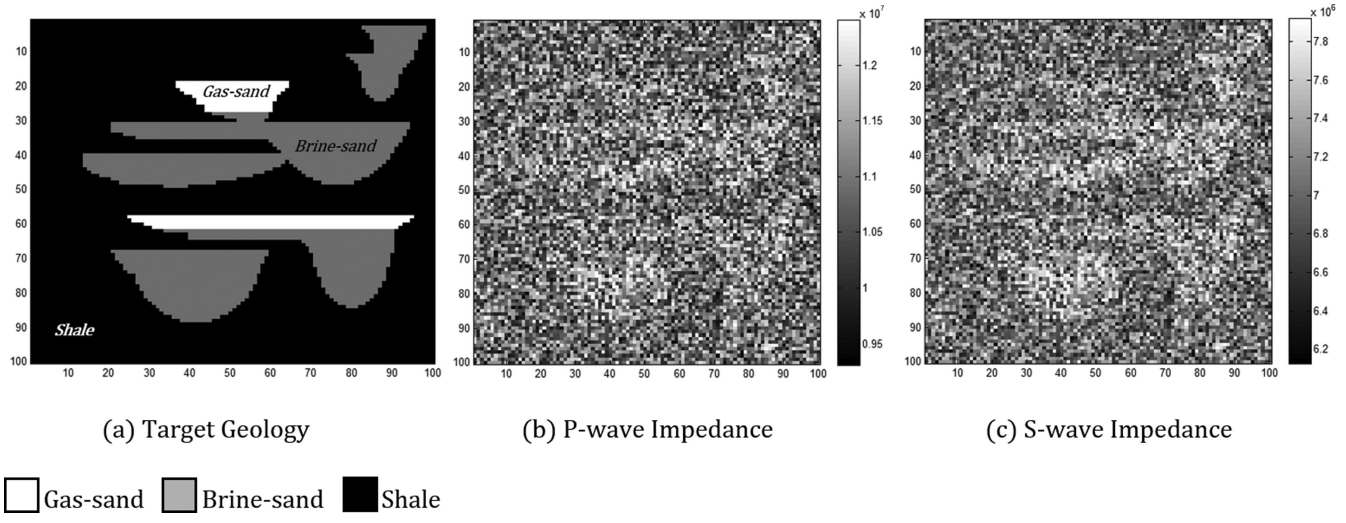


Figure 8. (a) A 2D cross-section that represents the true geological model which is the target for spatial facies inversion. It is extracted from the 3D geological process model using the same process parameters (facies distributions) as the training image in Fig. 7. (b,c) *P*- and *S*-wave impedance attributes generated independently in each cell in the target cross-section using a probabilistic forward model based on the Yin-Marin shaly-sand rock physics model (Marion 1990; Yin *et al*. 1993; Avseth *et al*. 2005) with added Gaussian noise.

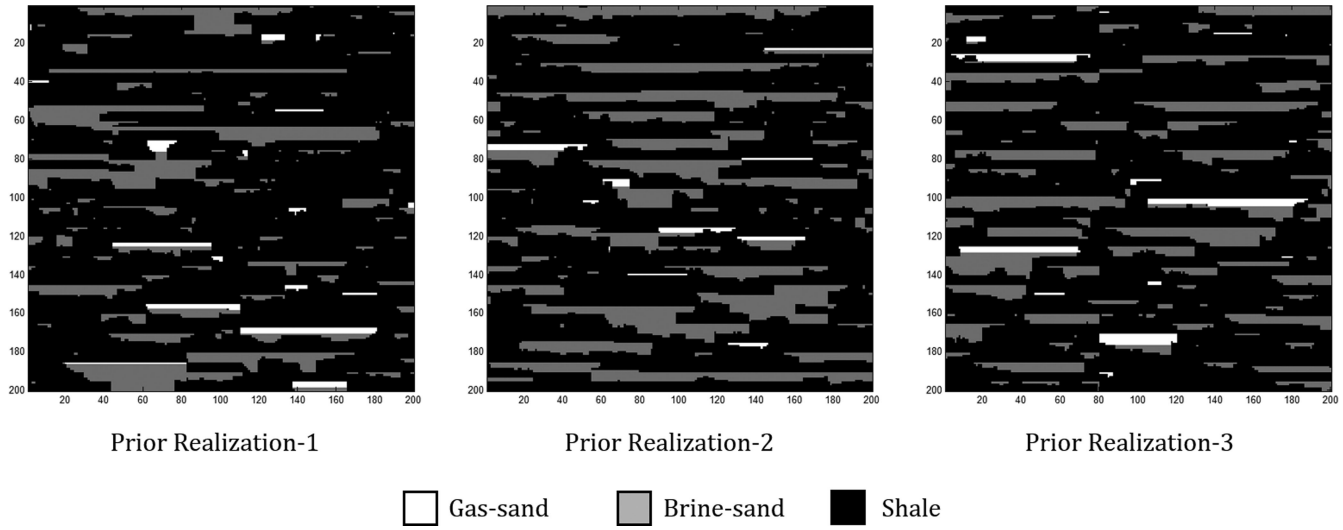


Figure 9. Three realizations generated from prior probabilities computed from the training image shown in Fig. 7 using the partition defined as a column of seven cells.

Table 1. Lower and upper bounds used to define Uniform distributions $\mathcal{P}(\boldsymbol{m}_k \mid g_{ij})$ [same as those used in Walker & Curtis (2014)].

Lithology-fluid class	Clay content by volume (V_{clay})	Sandstone matrix porosity (ϕ_{sand})	Water saturation (S_w)
Shale	[0.20, 0.40]	[0.20, 0.40]	[1.00, 1.00]
Gas-sand	[0.00, 0.20]	[0.20, 0.40]	[0.05, 0.60]
Brine-sand	[0.00, 0.20]	[0.20, 0.40]	[0.60, 1.00]

The target cross-section (Fig. 8a) was extracted from the same 3-D geological process model as the training image, and it therefore contains similar spatial distributions of facies as the training image. The target cross-section was used as a model to generate synthetic seismic attributes which were used to represent real data-derived attributes in our example. These were then inverted for facies using our algorithm with the aim to reproduce the original target cross-section.

Collocated synthetic seismic attributes, *P*- and *S*-wave impedances $\boldsymbol{d}_{ij}$, were generated independently in each cell (i, j) in the target cross-section using the localized likelihood assumption and a probabilistic forward model $\mathcal{P}(\boldsymbol{d}_{ij} \mid g_{ij})$. The Yin-Marion shaly-sand model (Marion 1990; Yin *et al.* 1993; Avseth *et al.* 2005) was used to predict *P*- and *S*-wave impedances from the given geological facies G_i, where

$$G_i \in \mathcal{G} = \{\text{shale, brine-sand, gas-sand}\}$$

The Yin-Marion model is defined by rock-physics parameters $\boldsymbol{m}_k = [V_{\text{clay}}, \phi_{\text{sand}}, S_w]_k$ where V_{clay} is the volume of clay, ϕ_{sand} is the matrix porosity of sand, S_w is the water saturation (with gas saturation given by $S_g = 1 - S_w$), and the subscript k refers to each facies. Gaussian random noise was then added to the predicted model in order to formulate the model probabilistically as $\mathcal{P}(\boldsymbol{d}_{ij} \mid \boldsymbol{m}_k)$. The likelihood $\mathcal{P}(\boldsymbol{d}_{ij} \mid g_{ij})$ is then given in terms of rock-physics parameters $\boldsymbol{m}_k$ by

$$\mathcal{P}\left(\boldsymbol{d}_{ij} \mid g_{ij}\right) = \iiint_{L}^{B} \mathcal{P}\left(\boldsymbol{d}_{ij} \mid \boldsymbol{m}_k\right) \mathcal{P}\left(\boldsymbol{m}_k \mid g_{ij}\right) d\boldsymbol{m}_k \tag{22}$$

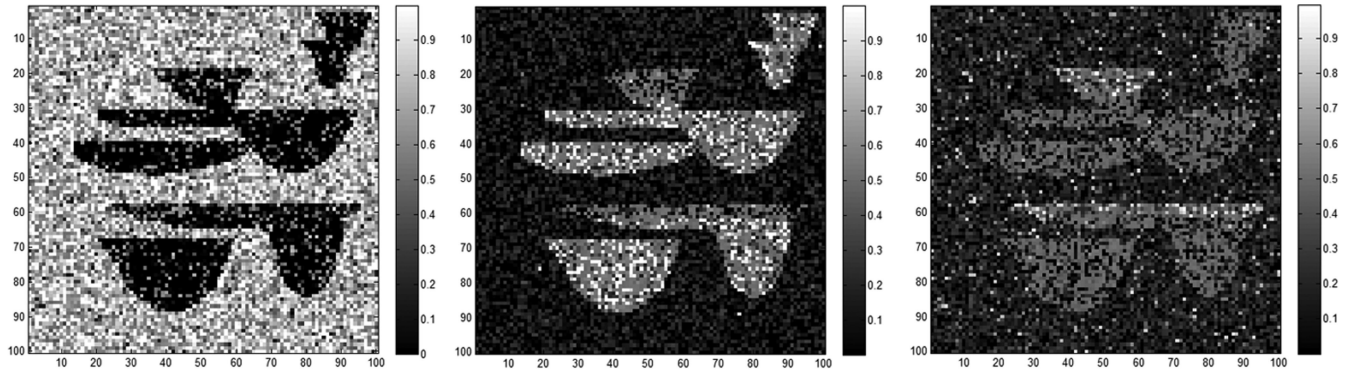
where **L** and **B** (bold-face letters to represent vector bounds) respectively represent the lower and upper bounds on each parameter in $\boldsymbol{m}_k$. The conditional distribution $\mathcal{P}(\boldsymbol{m}_k \mid g_{ij})$ describing the probabilistic relationship between rock-physical parameters $\boldsymbol{m}_k$ and the geological facies g_{ij} in each cell of the target cross-section, was set to Uniform within predefined lower and upper bounds [**L**, **B**] on each parameter in $\boldsymbol{m}_k$ given in Table 1. The distribution $\mathcal{P}(\boldsymbol{d}_{ij} \mid \boldsymbol{m}_k)$ is given by the deterministic Yin-Marion shaly-sand model $\boldsymbol{f}(\boldsymbol{m}_k)$ and a stochastic component in the form of Gaussian random noise $\boldsymbol{e}$ added to the predicted model in order to formulate the model probabilistically

$$\boldsymbol{d}_{ij} = \boldsymbol{f}\left(\boldsymbol{m}_k\right) + \boldsymbol{e}, \text{ where } \boldsymbol{e} \sim \varphi\left(0, \Sigma_d\right), \text{ and}$$
$$\Sigma_d = \begin{bmatrix} \sigma_P^2 & 0 \\ 0 & \sigma_S^2 \end{bmatrix}. \tag{23}$$

Here φ represents the Gaussian function, and Σ_d represents the data covariance matrix with *P*-wave standard deviation $\sigma_{\text{P}} = 1.5 \times 10^4$ s^{-1} m^{-2} kg, *S*-wave standard deviation $\sigma_{\text{S}} = 1.0 \times 10^4$ s^{-1} m^{-2} kg, and off-diagonal elements equal to zero.

The distribution $\mathcal{P}(\boldsymbol{d}_{ij} \mid g_{ij})$ in eq. (22) was sampled sequentially: first for $\boldsymbol{m}_k$ from $\mathcal{P}(\boldsymbol{m}_k \mid g_{ij})$ and then for $\boldsymbol{d}_{ij}$ from $\mathcal{P}(\boldsymbol{d}_{ij} \mid \boldsymbol{m}_k)$, to obtain synthetic data $\boldsymbol{d}_{ij}$ given the facies g_{ij} in each cell in the model (target cross-section). The data thus obtained is shown in Figs 8(b) and (c).

The likelihood $\mathcal{P}(\boldsymbol{d}_{ij} \mid g_{ij})$ was computed for each of the geological facies (Fig. 10) from a Gaussian Mixture-Model (GMM) using neural networks (Meier *et al.* 2007a,b; Shahraeeni & Curtis 2011; Shahraeeni *et al.* 2012). Since the likelihood only uses attributes to discriminate between the facies it allows reasonable discrimination between sand and shale (Fig. 10a), but could hardly discriminate between brine-sand and gas-sand (Figs 10b and c). Also, the likelihood functions are noisy and do not adhere to the statistical spatial distribution of facies as depicted in the training image. This corroborates the need to introduce prior geological knowledge incorporating the spatial correlation of facies.



(a) Likelihood of Shale (b) Likelihood of Brine-sand (c) Likelihood of Gas-sand

Figure 10. Likelihood functions $P(\boldsymbol{d}_{ij} \mid g_{ij})$ for each of the geological facies, (a) shale, (b) brine-sand and (c) gas-sand, given the seismic attributes. Results are computed from a Gaussian Mixture-Model using neural networks (Meier *et al*. 2007a,b; Shahraeeni & Curtis 2011; Shahraeeni *et al*. 2012). In each plot, white is high probability (close to 1) and black is low probability (close to 0). The likelihoods are normalized so that the sum of likelihoods for each of the facies in any cell equals 1.

The marginal posterior distributions for each of the facies in each cell in the model were computed (left column in Figs 11a, c and e) incorporating both the prior geological knowledge elicited from the training image (Fig. 7) and the likelihood functions (Fig. 10).

Stochastic realizations were then generated by sequentially drawing random samples for each cell in the model using Algorithm 1 and a linear path traversing a row at a time. The random samples were drawn in each cell (i, j) using the conditional copula function as discussed in Section 4.1 which was derived from the training image (Section 4.2), and the marginal posterior distributions $\mathcal{P}(g_{ij} \mid \boldsymbol{d})$ computed in that cell (Figs 11a, c and e) from eqs (13), (15) and (17). Fig. 12 shows 4 of a total of 15 stochastic realizations that were generated with the aim to reproduce the target cross-section (Fig. 8a). The comparison of sampled realizations and the target cross-section shows that the quality of realizations is more accurate in our method compared to those generated by Walker & Curtis (2014); we return to this in the discussion below.

In order to assess the quality of samples taken using our method, we also estimated stochastic marginal posterior distributions in each cell in the model by constructing histograms from the 15 simulated stochastic realizations. These are compared with the marginal posterior distributions computed using 2D-HMM in Fig. 11. In summary, the left column in Figs 11(a), (c) and (e) shows the marginal posterior distributions for shale, brine-sand and gas-sand, computed from eq. (13) (with the forward and backward probabilities computed using eqs (15) and (17), respectively). The right column in Figs 11(b), (d) and (f) shows the corresponding marginal posterior distributions for shale, brine-sand and gas-sand estimated from 15 stochastic samples, out of which 4 such randomly selected samples are displayed in Fig. 12. The marginal posterior distributions show an excellent match with the original model as shown in Fig. 8(a).

6 COMPUTATIONAL EFFICIENCY

The computational cost of this algorithm may be expressed mathematically as the maximum number of floating point operations required to compute the posterior marginal distributions in each cell in the model:

$$4 \times r^2 \times (r-1) \times c \times |\mathcal{G}|^r \qquad (24)$$

where r is the number of rows and c is the number of columns in the region of influence $\mathcal{R}(i, j)$ around the cell (i, j) under consideration, and $|\mathcal{G}|$ represents the size of the sample space of geological facies (i.e. the number of geological facies considered). It is assumed in deriving the above expression that the partition $\boldsymbol{G}_P$ is defined as a column of r cells. The variable r has the maximum order 3 in eq. (24) and it also appears in the exponent of $|\mathcal{G}|$. This means that it is desirable to define the partition G_P along the shorter dimension of $\mathcal{R}(i, j)$.

The size of the space of geological facies (i.e. the number of discrete facies classes) $|\mathcal{G}|$ is an important factor in the above expression. Because of its exponentiation it must be chosen to be as small as possible. Typically a small number of facies are estimated in previous studies, for example, Walker & Curtis (2014) inverted for the same 3 classes as above, and Rimstad & Omre (2010) inverted for 4 classes. If the range of relevant rock facies is too large in a given subsurface volume, one can reduce the number of classes considered in the inversion by nesting various facies within one another, for example, limestone and dolomite lithologies may be considered as a single class of carbonates. If the number of facies is large and cannot be reduced, or if the size of the partition (or the shorter dimension of the region of influence) is large, the number of required floating point operations can be significantly reduced by limiting the facies configuration over a partition to those observed in the training image. Journel (2004) and Mariethoz & Caers (2014) provide a detailed account of the incorporation of prior geological knowledge in the form a training image, whereas Toftaker & Tjelmeland (2013) and Arnesen & Tjelmeland (2016) proposed methods to build prior models from a training image using a binary MRF. As the size of the partition increases, a naïve approach to storing the joint distribution of facies would require an exponential amount of computer memory. However, since a training image may only depict a finite number of facies configurations, it limits the shapes and scales of the facies configurations that are considered geologically plausible. As a consequence, we only need to compute probabilities and perform sampling for a limited number of configurations. The partition size can, therefore, be taken as large as any one of the dimensions of the training image, thus allowing this method to be easily extensible to 3-D without becoming computationally intractable.

The number of floating point operations in the above expression shows a computational efficiency that is many times faster

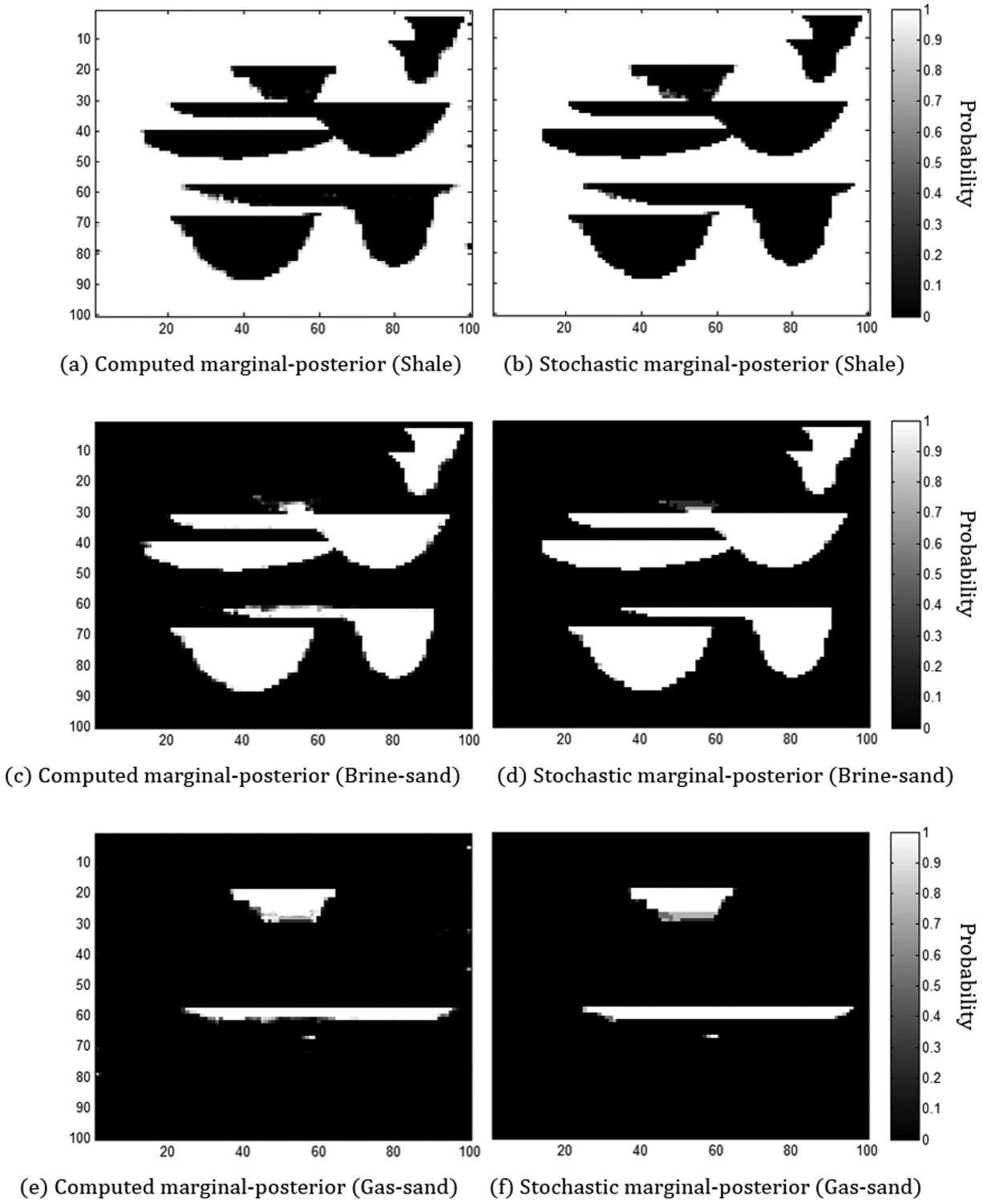


Figure 11. Comparison of marginal posterior distributions computed using 2D-HMM (a,c,e) and the marginal posterior distributions estimated stochastically from 15 realizations (b,d,f) for each of the facies: shale (top row), brine-sand (middle row) and gas-sand (bottom row).

than that of the joint probability computation method of Walker & Curtis (2014) (based on expression 16 in Walker & Curtis 2014). As a comparison, the computation of marginal distributions in each cell in a 5 × 7 grid on a processor with speed 1 GHz takes 0.2 s while the computation of the joint distribution (as in Walker & Curtis 2014) in each cell in the same grid and on the same processor takes 7.5 s whereafter marginal posterior distributions could be computed. Although this comparison is not fair because the two algorithms compute different quantities, it is clear that if our goal is to generate marginal posterior distributions then our method can do so using far lower computation power.

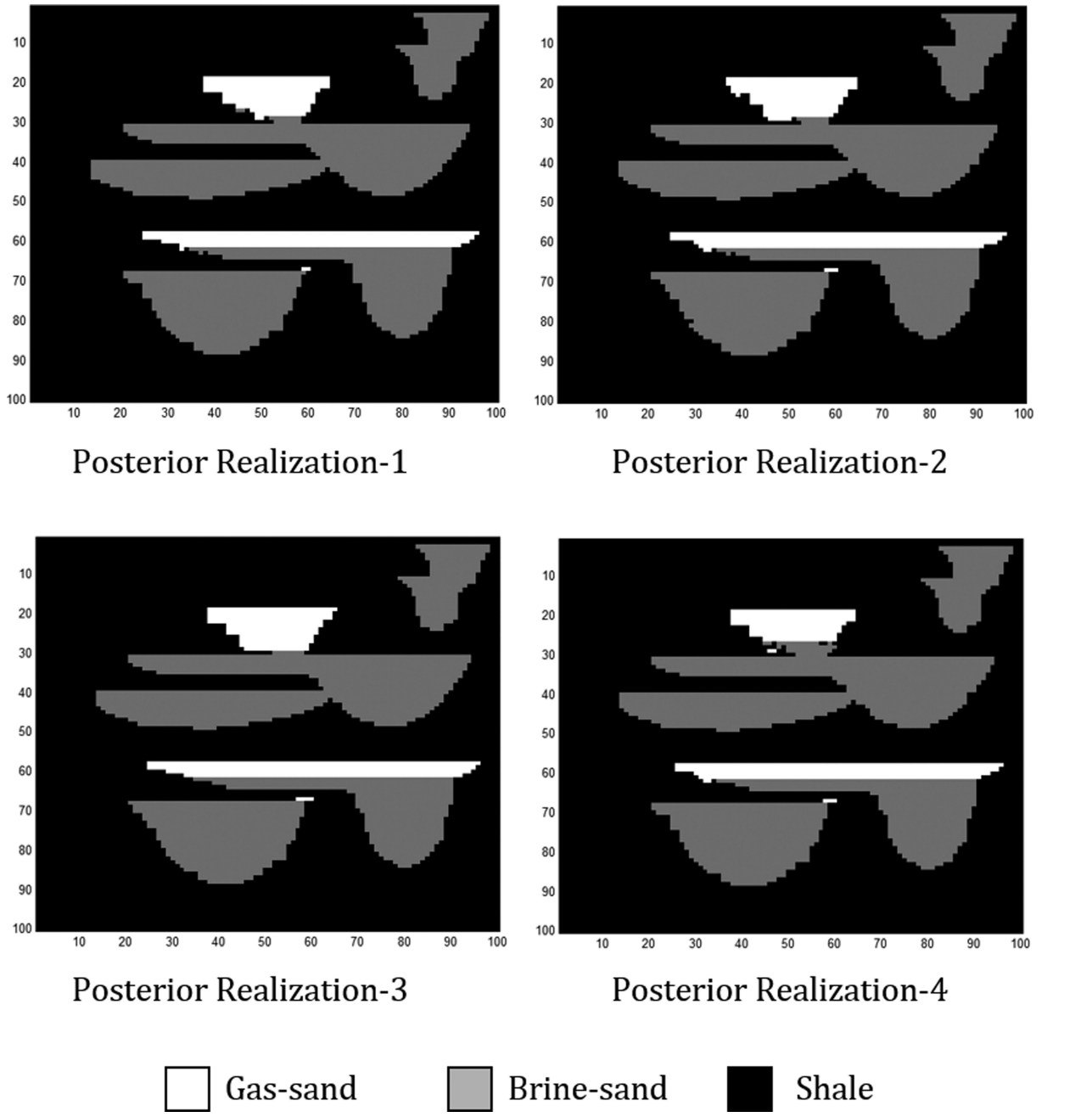


Figure 12. Four realizations of the target cross-section drawn from the conditional copula computed using eqs (13), (15) and (17) from a combination of the training image and the marginal posterior distributions $P(g_{ij} \mid \boldsymbol{D})$ in each cell (i, j) in the grid.

The memory required to store the marginal distributions of facies in each cell is proportional to the number of facies considered. Walker & Curtis (2014) did not present an expression to evaluate the memory required to store the partial conditional distributions in each cell, however, it can be assessed to be exponential in the smallest dimension of the grid (number of columns in their implementation). The memory required by our algorithm to store the marginal distributions of facies in each cell requires far less memory compared to that required to store conditional distributions of facies conditioned upon the previously sampled cells. Also, defining a region of influence around each cell in the model makes the working memory of the algorithm insignificant for modern computers.

7 DISCUSSION

The computation of a full joint distribution $\mathcal{P}(\boldsymbol{G} \mid \boldsymbol{D})$ of geological facies conditioned to seismic and well data is computationally intractable even for small synthetic models. Previous research in probabilistic seismic inversion by Walker & Curtis (2014) relied on the computation of approximate posterior conditional distributions of facies. We took a different approach and computed marginal posterior distributions for each of the facies in each cell in the model, conditioned to the data in that cell and to prior facies distributions. Computation of marginal posteriors is orders of magnitude faster than the previous approach, and requires far lower memory. Also,

because Walker & Curtis (2014) used a 1-D underlying graphical model, sampling of each cell had to be conditioned on a large number of previously sampled cells in one of the model dimensions, well beyond the neighbourhood structure. Since our graphical model is 2-D, we do not have any such limitation and sampling only needs to be conditioned on the previous sampled cells within the neighbourhood structure. The previous work generates samples from the partial conditional distributions whereas our method generates samples from the computed marginal posterior distributions. In terms of quality of prior information incorporated into the inversion process, our method clearly outperforms the method of Walker & Curtis (2014): the realizations from prior distributions (Fig. 9) and the computed marginal posterior distributions (Fig. 12) show flat tops of channels in our example model, whereas the previous method could not produce flat tops in samples from the same example. The main reason for this difference is likely to be that our method computes prior probabilities of spatial distribution of facies over partitions (7 cells in a column in our example) as compared to the neighbourhood structure (3×3 cells as was used in previous work). The size of partition is typically larger than the size of neighbourhood structure in the same dimension (7 cells versus 3 cells).

Other previously existing methods that invert seismic data for geological facies using hidden Markov or similar models (e.g. Larsen *et al*. 2006; Ulvmoen & Omre 2010; Ulvmoen *et al*. 2010; Hammer & Tjelmeland 2011; Rimstad & Omre 2013; Lindberg & Omre 2014, 2015) rely on sampling from full posterior distributions using McMC methods. As described earlier in Section 1, McMC based methods are slow to converge for high-dimensional problems and they suffer from convergence related bias. For this reason, a comparison of our method with McMC based methods in terms of computational efficiency would be essentially meaningless. Nonetheless, a comparison can be made in terms of the amount and quality of prior information incorporated in the inversion process. Since our method is based on full 2-D relationships among cells in neighbouring partitions, it incorporates more prior information as compared to the 1-D Markov-chain based methods (e.g. Larsen *et al*. 2006). However, the amount and quality of prior information incorporated is comparable between our 2D-HMM based priors and the profile Markov random field based priors of Ulvmoen & Omre (2010) and Ulvmoen *et al*. (2010). The advantage of our method over the latter methods remains that we perform inference directly for posterior marginal distributions (in contrast to the full joint distribution) while avoiding the use of McMC.

The localized likelihood assumption is a fundamental assumption in this algorithm as the observations (here seismic attributes) must be conditionally independent given the hidden states (geological facies) in a 2D-HMM. The localized likelihood assumption also allowed us to factorize the likelihood probability $\mathcal{P}(\boldsymbol{D}_P \mid \boldsymbol{G}_P) = \prod_{i=1}^{n} \mathcal{P}(\boldsymbol{d}_{P,i} \mid g_{P,i})$ as a product of factors each involving the local likelihood in each cell in the model, as used in eqs (15) and (17). This means that the seismic attributes are assumed to possess spatial correlations that are only due to the spatial correlations present in the geology (facies and rock properties). However, spatial correlations are also induced in the seismic attributes by the non-localized nature and limited resolution of seismic data, and by correlated noise that was not accounted for during the process that estimated the attributes. Since we used seismic attributes as discriminators of geological facies under a localized likelihoods assumption, it is of paramount importance that the seismic attributes are as localized as possible. This means that they must be derived from seismic data which has been processed carefully and corrected for non-localizing effects of seismic wave propagation such as attenuation, Fresnel zone smearing, etc. This in turn requires that the input seismic attributes are computed after proper de-noising and migration of seismic data in which all wave propagation effects have been accounted for.

Spatial correlations in attributes due to the correlations in geology, on the other hand, are exploited in the inference to improve the spatial correlations in the inverted facies. Therefore, although the assumption of localized likelihoods allows us to compute approximate marginal posterior distributions in a closed form solution, it effectively limits the amount of information present in the seismic attributes that could otherwise be useful in the reconstruction of spatial correlations of facies (specifically, our method ignores correlations in the attributes between cells). As a consequence, our method relies significantly on the prior information, rather than the data, to reconstruct the spatial correlations expected in the geology. The data, therefore, provides the location specific information and the prior knowledge provides information on the spatial correlations to be recovered in the inversion. An advantage of using prior information in this way is that it reduces sensitivity to random noise in the data. However, a more sophisticated approach would exploit the fact that the seismic attributes at neighbouring locations are spatially correlated, depending on the temporal and spatial resolution of the seismic data. We leave such an approach as a direction for future research.

The incorporation of prior information from a training image is dependent on the configuration of pixels that are scanned to compute spatial conditional probabilities. Since we scan the training image with a stencil that is the same as the partition G_P, it is important to define the shape and size of the partition such that the information gathered can reproduce structures present in the training image up to any desired accuracy. In our synthetic case we defined a partition as a column of 7 cells as we found that the prior information thus gathered from the training image was sufficient to reconstruct the actual marginal distributions with reasonable accuracy. This was because the vertical variations in facies in our training image are correlated over smaller length scales than the lateral variations, and hence 7 cells were sufficient. Because we compute the conditional probabilities of facies patterns by sampling from the training image using a direct sampling approach (Mariethoz *et al*. 2010), we can easily extend the size of the neighbourhood structure within the memory limits of modern computers. A reasonably large neighbourhood structure increases the computational time but still remains tractable. The shape of the partition can also be chosen with arbitrary complexity to model complex spatial distributions of facies provided the ordering of partitions can still be defined as required by the algorithm. Since the size of partition defines the size of the region of influence along any one of the dimensions, the partition size should be chosen large enough that the region of influence may contain any large scale recoverable features in the training image.

The assumption of the region of influence $\mathcal{R}(i, j)$ around each cell (i, j) in the model is based on the observation that the facies at any location in the sub-surface have probabilistic dependence only on the data observed in a certain region around it. This region can be made reasonably large but finite. In fact it could be as large as the size of the training image. If the region-of-influence is large enough to capture the large scale facies patterns depicted in the training image, this assumption only limits the data correlations outside of this region and not the correlation of geological facies. As a consequence, the concept of region-of-influence not only makes the algorithm tractable without limiting the size of the overall model,

it also offers a reliable estimation of posterior marginal distributions of facies at the point of interest. Since this assumption is no stricter than the assumption of localized likelihoods, it is therefore valid for all models that are built with the assumption of localized likelihoods. This also applies to models in various other fields of research, such as image and video processing. Other researchers who used 2-D extensions of HMM either limited the spatial interactions of neighbouring cells in the model, or they assumed a 1-D underlying graphical model (pseudo 2D HMM). Both of these approaches prohibited incorporation of full 2-D interactions of cells (facies correlations in our synthetic example). The assumption of the region of influence allowed us to derive the equations to compute marginal posterior distributions with full 2-D spatial interactions among neighbouring cells in the model.

Although we demonstrated the use of a 2D-HMM for spatial inversion of geological facies from seismic data, extension of the method to 3-D or higher dimensions is straightforward. Since marginal posterior distributions can be computed in each cell independently, this approach can be parallelized on heterogeneous computer architectures to exploit the maximum efficiency deliverable from the modern day computational and graphical processors.

Sampling using a conditional copula ensures that each sampled location is conditionally dependent upon previously sampled locations in the neighbourhood of the location being sampled. A copula function can be determined uniquely if all of the marginal distributions are unique. This is not the case for the distribution of facies since these are discrete variables, so a consistent approach is required when defining the copula. Eq. (20) addresses this problem by using the infimum function.

The realizations generated using our method (see Fig. 12) conform to the marginal posteriors better compared to the method of Walker & Curtis (2014). They used an underlying graphical model based on Markov Random Fields, and stipulated that the so-called *positivity condition* is satisfied in their formulation of the algorithm as is required by the Hammersley-Clifford theorem (proved by Besag 1974). However, they could not prove that their graphical model conforms to this requirement. This might be the reason that their sampling results do not conform so well to the spatial correlations of facies depicted in the training image, but we cannot prove this; on the other hand, in our method the Hammersley-Clifford theorem is not applicable, making the point moot.

8 CONCLUSIONS

This paper introduces a method to compute marginal posterior probabilities of geological facies from prior knowledge of spatial facies correlations and data likelihoods by using a 2-D hidden Markov model. The prior knowledge is incorporated in terms of spatial statistics of facies distributions in space, and can be represented in the form of a training image or otherwise. The prior information is assumed to be non-localized and is interpreted as the conceptual depiction of geology. The prior probabilities are therefore independent of data, and only contribute non-localized contextual information. Since the data are observed, they are fixed and the data likelihood is computed as the probability of observing the measured data at any point given that the geological facies at that point is known. The observed data represents any type of data (e.g. *P*-wave and *S*-wave impedances) that can discriminate between geological facies present at any point in the model to some degree of confidence. The likelihood is therefore assumed to be localized; such an assumption is called the condition of *localized likelihoods*. This implies that the data values observed at any point are assumed to depend only on the geological facies present at that point, and are assumed to be independent of the presence of geological facies and data observed at any other point in the model. Our algorithm, in essence, inverts seismic attributes for geological facies by combining the localized likelihoods with the conceptual prior knowledge about spatial correlations of geological facies. The implication of the localized likelihoods assumption is that the seismic data is assumed to be processed and corrected for any non-localized effects of seismic wave propagation.

Previous researchers have used 2-D hidden Markov models to solve problems in various fields such as computational biology and computer vision. They made assumptions that limit the interaction between neighbouring cells. Our method makes no such assumptions and models the full 2-D interactions between neighbouring cells in the model. However, our method does assume that there exists a region of influence around each cell in the model such that any observations (data) outside of this region have no correlation with the observation in the cell under consideration. Such an assumption does not limit the spatial correlations among the hidden states, and is therefore valid for any model that is based on the localized likelihoods assumption.

This paper also describes an approach for stochastic simulation of geological facies where samples are generated from the computed marginal posterior distributions by using a copula function, without requiring partial conditional posterior distributions. The simulation results based on marginal posterior distributions are shown to provide a significant improvement over methods from previous research in this field. As opposed to sampling partial conditional posterior distributions, sampling from marginal posterior distributions is many orders of magnitude faster and does not involve normalization approximations.

ACKNOWLEDGEMENTS

We are thankful to TOTAL UK for their sponsorship of this research. We would also like to show our gratitude to Mohammed Shahraeeni and Benoit Paternoster of TOTAL UK for providing useful suggestions and sharing their expertise during the course of this research. We are also grateful to Prof Henning Omre and Prof Malcolm Sambridge for their comments and constructive criticism on an earlier version of this manuscript.

REFERENCES

Abend, K., Harley, T.J. & Kanal, L.N., 1965. Classification of binary random patterns, *IEEE Trans. Inf. Theory,* **11**(4), 538–544.

Alasonati, P., Wassermann, J. & Ohrnberger, M., 2006. Signal classification by wavelet-based hidden Markov models: application to seismic signals of volcanic origin, in *Statistics in Volcanology,* eds Mader, H., Connor, C. & Coles, S., Geological Society of London.

Arnesen, P. & Tjelmeland, H., 2016. Prior specification of neighbourhood and interaction structure in binary Markov random fields, *Stat. Comput.,* doi:10.1007/s11222-016-9650-5.

Austad, H.M. & Tjelmeland, H., 2016. Approximations and bounds for binary Markov random fields, *Stat. Comput*., doi:10.1007/s11222-016-9685-7.

Avseth, P., Mukerji, T. & Mavko, G., 2005. *Quantitative Seismic Interpretation,* Vol. 1, Cambridge Univ. Press.

Bartolucci, F. & Besag, J., 2002. A recursive algorithm for Markov random fields, *Biometrika,* **89**(3), 724–730.

Baum, L.E., 1972. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes, *Inequalities,* **3,** I–8.

Baum, L.E. & Petrie, T., 1966. Statistical inference for probabilistic functions of finite state Markov chains, *Ann. Math. Stat.,* **37**(6), 1554–1563.

Baum, L.E., 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.

Baumgartner, J., Flesia, A.G., Gimenez, J. & Pucheta, J., 2013. A new approach to image segmentation with two-dimensional hidden Markov models, in *2013 BRICS Congress on Computational Intelligence and 11th Brazilian Congress on Computational Intelligence (BRICS-CCI & CBIC),* Ipojuca, pp. 213–222.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems, *J. R. Stat. Soc. B,* **36**(2), 192–236.

Bevilacqua, V., Daleno, D., Cariello, L. & Mastronardi, G., 2007. Pseudo 2D hidden Markov models for face recognition using neural network coefficients, in *2007 IEEE Workshop on Automatic Identification Advanced Technologies,* Alghero, pp. 107–111.

Beyreuther, M. & Wassermann, J., 2008. Continuous earthquake detection and classification using discrete hidden Markov models, *Geophys. J. Int.,* **175**(3), 1055–1066.

Can, C.E., Ergun, G. & Gokceoglu, C., 2014. Prediction of earthquake hazard by hidden Markov model (around Bilecik, NW Turkey), *Cent. Eur. J. Geosci.,* **6**(3), 403–414.

Collins, M., 2002. Discriminative training methods for hidden Markov models: theory and experiments with perceptron algorithms, in *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing,* Vol. 10, Philadelphia, PA, pp. 1–8.

Cressie, N. & Davidson, J.L., 1998. Image analysis with partially ordered Markov models, *Comput. Stat. Data Anal.,* **29**(1), 1–26.

Daleno, D., Cariello, L., Giannini, M. & Mastronardi, G., 2010. Pseudo 2D hidden Markov model and neural network coefficients in face recognition, in Face Recognition, ed. Oravec M., InTech.

Dymarski, P., 2011. *Hidden Markov Models, Theory and Applications,* InTech, 326 pp.

Eddy, S.R., 1998. Profile hidden Markov models, *Bioinformatics,* **14,** 755–763.

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), 379–397.

Forney, G.D., Jr., 1973. The Viterbi algorithm, *Proc. IEEE,* **61**(3), 268–278.

Frechet, M., 1951. Sur les tableaux de corrélation dont les marges sont données, *Ann. Univ. Lyon. Sect. A.,* **14**(3), 53–77.

Gallagher, K., Charvin, K., Nielsen, S., Sambridge, M. & Stephenson, J., 2009. Markov chain Monte Carlo (McMC) sampling methods to determine optimal models, model resolution and model choice for Earth Science problems, *Mar. Pet. Geol.,* **26**(4), 525–535.

Hammer, H.L. & Tjelmeland, H., 2011. Approximate forward–backward algorithm for a switching linear Gaussian model, *Comput. Stat. Data Anal.,* **55**(1), 154–167.

Jaworski, P., Durante, F., Härdle, W.K. & Rychlik, T., 2010. Copula theory and its applications, *Lecture Notes in Statistics Proceedings*, Proceedings of the Workshop Held in Warsaw, 25–26 September 2009, pp 12.

Journel, A.G., 2004. Beyond covariance: the advent of multiple-point geostatistics, in *Geostatistics Banff 2004,* Vol. 14 of the series Quantitative Geology and Geostatistics, pp. 225–233, eds Leuangthong, O. & Deutsch, C.V., Springer.

Krogh, A., Brown, M., Mian, I.S., Sjolander, K. & Haussler, D., 1994. Hidden Markov models in computational biology: applications to protein modeling, *J. Mol. Biol.,* **235,** 1501–1531.

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

Li, J., Najmi, A. & Gray, R.M., 2000. Image classification by a two-dimensional hidden Markov model, *IEEE Trans. Signal Process.,* **48**(2), 517.

Liang, K.C., Wang, X. & Anastassiou, D., 2007. Bayesian basecalling for DNA sequence analysis using hidden Markov models, *IEEE/ACM Trans. Comput. Biol. Bioinform.,* **4,** 430–440.

Lindberg, D. & Omre, H., 2014. Blind categorical deconvolution in two-level hidden Markov models, *IEEE Trans. Geosci. Remote Sens.,* **52**(11), 7435–7447.

Lindberg, D. & Omre, H., 2015. Inference of the transition matrix in convolved hidden Markov models and the generalized Baum-Welch algorithm, *IEEE Trans. Geosci. Remote Sens.,* **53**(12), 6443–6456.

Ma, X., Schonfeld, D. & Khokhar, A., 2008. Image segmentation and classification based on a 2D distributed hidden Markov model, *Proc. SPIE,* **6822,** 51 pp., doi:10.1117/12.766112.

Mariethoz, G. & Caers, J., 2014. *Multiple-point Geostatistics: Stochastic Modeling with Training Images,* Wiley-Blackwell.

Mariethoz, G., Renard, P. & Caers, J., 2010. The Direct Sampling method to perform multiple-point geostatistical simulations, *Water Resour. Res.,* **46,** doi:10.1029/2008WR007621.

Marion, D.P., 1990. Acoustical, mechanical, and transport properties of sediments and granular materials, *PhD thesis,* Stanford University, Department of Geophysics.

Meier, U., Curtis, A. & Trampert, J., 2007a. A global crustal model constrained by non-linearised inversion of fundamental mode surface waves, *Geophys. Res. Lett.,* **34,** L16304, doi:10.1029/2007GL030989.

Meier, U., Curtis, A. & Trampert, J., 2007b. Global crustal thickness from neural network inversion of surface wave data, *Geophys. J. Int.,* **169,** 706–722.

Nivre, J., 2002. On statistical methods in natural language processing, in *Promote IT Second Conference for the Promotion of Research in IT at New Universities and University Colleges in Sweden,* pp. 684–694, ed. Bubenko, J., University of Skövde.

Othman, H. & Aboulnasr, T., 2003. A separable low complexity 2D HMM with application to face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.,* **25**(10), 1229–1238.

Pachter, L., Alexandersson, M. & Cawley, S., 2002. Applications of generalized pair hidden Markov models to alignment and gene finding problems, *J. Comput. Biol.,* **9,** 389–399.

Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition, *Proc. IEEE.,* **77,** 257–286.

Rimstad, K. & Omre, H., 2010. Impact of rock-physics depth trends and Markov random fields on hierarchical Bayesian lithology/fluid prediction, *Geophysics,* **75**(4), R93–R108.

Rimstad, K. & Omre, H., 2013. Approximate posterior distributions for convolutional two-level hidden Markov models, *Comput. Stat. Data Anal.,* **58,** 187–200.

Rimstad, K., Avseth, P. & Omre, H., 2012. Hierarchical Bayesian lithology/fluid prediction: a North Sea case study, *Geophysics,* **77**(2), B69–B85.

Sambridge, M. & Mosegaard, K., 2002. Monte Carlo Methods in Geophysical Inverse Problems, *Rev. Geophys.,* **40**(3), 3-1–3-29.

Shahraeeni, M.S. & Curtis, A., 2011. Fast probabilistic nonlinear petrophysical inversion, *Geophysics,* **76**(2), E45–E58.

Shahraeeni, M.S., Curtis, A. & Chao, G., 2012. Fast probabilistic petrophysical mapping of reservoirs from 3D seismic data, *Geophysics,* **77**(3), O1–O19.

Shih, M.C., Renuka, S. & Rose, K., 2015. 2D hidden Markov model with spatially adaptive state-space for tracing many cells in image sequence, in *IEEE 12th International Symposium on Biomedical Imaging (ISBI),* New York, pp. 1452–1456.

Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris,* **8,** 229–231.

Sklar, A., 1973. Random variables, joint distributions, and copulas, *Kybernetica,* **9,** 449–460.

Stratonovich, R.L., 1960. Conditional Markov processes, *Theory Probab. Appl.,* **5**(2), 156–178.

Sun, S., Liu, H., Lin, H. & Abraham, A., 2012. Twitter part-of-speech tagging using pre-classification Hidden Markov model, in *IEEE International Conference on Systems, Man, and Cybernetics (SMC),* Seoul, pp. 1118–1123.

Tjelmeland, H. & Austad, H.M., 2012. Exact and approximate recursive calculations for binary Markov random fields defined on graphs, *J. Comput. Graph. Stat.,* **21,** 758–780.

Toftaker, H. & Tjelmeland, H., 2013. Construction of binary multi-grid Markov random field prior models from training images, *Math. Geosci.,* **45,** 383–409.

Ulvmoen, M. & Omre, H., 2010. Improved resolution in Bayesian lithology/fluid inversion from prestack seismic data and well observations, Part 1 — Methodology, *Geophysics,* **75**(2), R21–R35.

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,* **IT-13**(2), 260–269.

Walker, M. & Curtis, A., 2014. Spatial Bayesian inversion with localized likelihoods: an exact sampling alternative to McMC, *J. geophys. Res.,* **119,** 5741–5761.

Yin, H., Nur, A. & Mavko, G., 1993. Critical porosity—a physical boundary in poroelasticity, *Int. J. Rock Mech. Min. Sci. Geomech. Abstr.,* **30**(7), 805–808.

Yoon, B.J., 2009. Hidden Markov models and their applications in biological sequence analysis, *Curr. Genomics.,* **10**(6), 402–415.