On the construction of discrete distributions from training images

Håkon Tjelmeland and Håkon Toftaker

Abstract In reservoir characterisation is has become common practice to define a facies model by fitting a model to one or more training images. Particularly multi-point statistics models are popular. Then node values are simulated sequentially in a random order and conditional distributions for one node value given the previous ones are estimated from the training image(s). This is often combined with a multi-grid approach. In the present paper we focus on binary training images and consider a simplified variant of the multi-point statistics framework. We simulate the node values sequentially in a fixed order and do not include the multi-grid idea. The resulting model is a (higher-order) Markov chain and corresponding theory can be used to understand the properties of the fitted model. In particular the Markov chain is in general not stationary, from which it follows for example that the marginal probabilities vary spatially. This is clearly an unwanted property and we discuss several strategies for coping with the problem. One should note that by restricting us to a fixed simulation order we obtain explicit expressions for the probability distribution of the fitted model, which makes it easy to define a corresponding conditional distribution when observed data are available.

1 Introduction

To model the spatial facies distribution in reservoirs it has become common practice to fit a chosen model class to one or several training images. Different classes of models can be used for this purpose. Various multi-point statistics models (Strebelle,

Håkon Tjelmeland
Department of Mathematical Sciences, Norwegian University of Science and Technology, Gløshaugen, NO-7491 Trondheim, Norway, e-mail: haakont@stat.ntnu.no

Håkon Toftaker
Department of Mathematical Sciences, Norwegian University of Science and Technology, Gløshaugen, NO-7491 Trondheim, Norway, e-mail: toftaker@stat.ntnu.no

Ninth International Geostatistics Congress, Oslo, Norway, June 11. – 15., 2012.

2002; Journel and Zhang, 2006) are frequently considered, but other possibilities that have been used include Markov mesh models (Stien and Kolbjørnsen, 2011) and Markov random fields (Tjelmeland, 1997; Toftaker and Tjelmeland, 2012a). In multi-point statistics (MPS) models the node values are simulated in a random order and the conditional distributions for one node value given the values previously simulated in nearby locations are estimated from the training image(s). With this strategy the number of model parameters to estimate may become huge and thereby over-fitting may become a problem. Another potential problem with the MPS modelling strategy is that there is no way to ensure that the fitted model is stationary. The third, and perhaps most serious, problem is that we have no closed form, easy to evaluate, formula for the probability distribution of a fitted model. Whether or not the latter is a problem depends on what you want to use the MPS model for. MPS is an algorithmically defined model, and as a consequence unconditional simulation from a fitted model is straight forward. If this is all you want to use the fitted model for, the lack of an easy to evaluate formula for the probability distribution represent no problem. If, however, you want to generate also conditional realisations from the fitted model given some observed values, it is not clear how to construct a simulation algorithm that simulates consistent with the induced conditional distribution.

There are serious theoretical and practical problems associated also with the Markov mesh and Markov random field models. The Markov mesh models, and more general partially order Markov models (Cressie and Davidson, 1998), simulates the node values in a fixed order. As a result an easy to evaluate closed form expression is available for the fitted probability distribution, so it is always possible to construct a Metropolis–Hastings algorithm (Gamerman, 1997; Brooks et al., 2011) that simulates from a corresponding conditional distribution. However, just as for MPS models there is no easy way to ensure that the fitted model is stationary. Unconditional realisations often show high correlations in certain directions that are not present in the training image, but is a result of the fixed simulation order. Moreover, the fraction of the various facies in unconditional realisations is often not consistent with corresponding figures in the training image. Markov random fields have very nice theoretical properties. Except for a border effect it is easy to ensure that the fitted model is stationary, and except for a normalising constant an easy to evaluate expression is available for the probability distribution of the fitted model. Just as for Markov mesh models it is easy to construct a Metropolis–Hastings algorithm that simulates from a corresponding conditional distribution. The problem with Markov random fields, however, is that the formulation includes a computationally intractable normalising constant. This represent no problem for conditional simulation, but is a major problem in the model fitting phase. Possible strategies for model fitting have been suggested, but typically to a computationally high price. The possibilities include estimating the normalising constant via Markov chain Monte Carlo (Geyer and Thompson, 1995; Gelman and Meng, 1998; Gu and Zhu, 2001), replacing the need to evaluate the normalising constant with exact sampling (Møller et al., 2006), and replacing the normalising constant with a corresponding approximation (Austad and Tjelmeland, 2011; Tjelmeland and Austad, 2012).

In the present paper we limit the attention to binary fields and study models that simulate the node values in a specific fixed order. As such it is a Markov mesh model and a partially order Markov model. We consider the simulated values as a Markov chain and are thereby able to study the stationarity properties of the model. We first consider a rather naïve modelling strategy, which has clear similarities to the way MPS and Markov mesh models are defined, and we discuss why the resulting model is non-stationary. Thereafter we discuss different strategies for ensuring that the fitted model becomes stationary. Toftaker and Tjelmeland (2012b) give a more detailed treatment of the models discussed here.

2 The naïve approach

Assume we have an $n \times m$ rectangular lattice and let $S = \{(i,j); i = 1,\ldots,n, j = 1,\ldots,m\}$ be the set of nodes. To each node (i,j) we associate a binary variable $x_{ij} \in \{0,1\}$. We use the notations $x = (x_{ij}, (i,j) \in S)$ and $x_A = (x_{ij}, (i,j) \in A)$ for $A \subseteq S$, and $x_{a:b,c:d} = x_{\{(i,j),a\leq i\leq b,c\leq j\leq d\}}$, $x_{a,c:d} = x_{\{(a,j),c\leq j\leq d\}}$ and $x_{a:b,c} = x_{\{(i,c),a\leq i\leq b\}}$ for $a,b \in \{1,\ldots,n\}$, $c,d \in \{1,\ldots,m\}$ when $a < b$ and $c < d$. Our goal is to fit a distribution $p(x)$ to a given training image x_0. We let $p(x)$ be defined from a distribution $\pi(x_T)$ on a very small $q \times r$ lattice $T = \{(i,j), i = 1,\ldots q, r = 1,\ldots,r\}$. Typically q and r will equal two, three or four. Clearly $\pi(x_T)$ is specified by 2^{qr} non-negative probabilities that sum to one. We define $p(x)$ from $\pi(x_T)$ as follows. First we let

$$p(x_T) = \pi(x_T). \tag{1}$$

We assume $x_{1:q,j}, j = r+1,\ldots,m$ to be a $(q-1)$th order Markov where the transition probabilities $p(x_{1:q,j}|x_{1:q,j-r+1:j-1})$ are defined from $\pi(\cdot)$ as the conditional distribution of the last row given the first $r-1$ rows. This defines a distribution $p(x_{1:q,1:m})$ and thereby also a corresponding conditional distribution $p(x_{q,1:m}|x_{1:q-1,q:m})$. We assume also $x_{i,1:m}, i = q+1,\ldots,n$ to be a Markov chain and let the transition probabilities of this Markov chain be given by $p(x_{q,1:m}|x_{1:q-1,q:m})$. This completes the definition of $p(x)$.

One should note that it is straight forward to simulate from $p(x)$ and it is also easy to evaluate $p(x)$ for a given x. The evaluation and simulation of the first q rows includes only low dimensional distributions and the conditional distribution $p(x_{q,1:m}|x_{1:q-1,q:m})$ is a one dimensional Markov random field and can be numerically handled by the so called forward-backward algorithm, see for example Scott (2002).

The distribution $\pi(x_T)$ must be estimated from the training image, and different alternatives exist for this. The simplest alternative (i) is perhaps just to let $\pi(x_T)$ equal the frequency of the block x_T in the training image. A variant of this procedure is (ii) to give a small positive probability also to configurations x_T that does not appear in the training image. An alternative estimation procedure is (iii) to consider the training image as an observed realisation from $p(x)$ and estimate $\pi(x_T)$ by

maximum likelihood. In Figure 1 we show, for three different training images, realisations from the fitted model when $q = r = 2$, when $q = r = 3$ and when $q = 4, r = 3$, and defining $\pi(x_T)$ by alternative (ii). We can observe that this very simple mod-

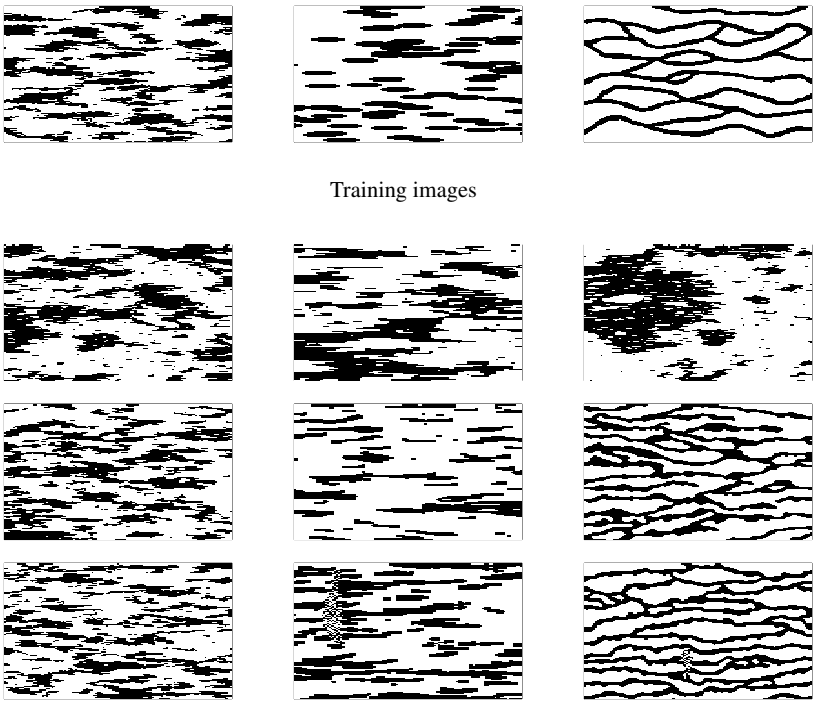

Fig. 1 The naïve approach: The three training images (upper row) and for each of these realisations from a fitted $p(x)$ when $q = r = 2$ (second row), $q = r = 3$ (third row) and $q = 4, r = 3$ (lower row).

elling strategy is capturing a lot of the characteristics of the training images, at least when $q = r = 3$ or $q = 4, r = 3$. However, it is also easy to identify characteristics of the training images that are not reproduced in the fitted models. First, the realisations in the middle and right columns for the $q = 4, r = 3$ case contains artifacts where the fitted model seems to have lost all memory of the training image. This effect occurs frequently in the realisations from the model. After a closer inspection of the training images and realisations we also see that the fraction of white and black in the training image is not reproduced in the fitted model. For example, in the realisations from $q = r = 3$ and $q = 4, r = 3$ in the right column there are much too many channels.

What is the reasons for the shortcomings of the fitted models? The $p(x)$ is essentially defined by two Markov chains where the transition probabilities are given from $\pi(x_T)$. From the way these Markov chains are constructed they will typically not be stationary. The two Markov chains have corresponding limiting distributions, but these will typically not be as we expect from the estimated $\pi(x_T)$. One should note that one should expect the same effect in MPS models. The introduction of a random simulation order and a multi-grid framework perhaps reduce the effect of the non-stationarity, and certainly makes the model less easy to analyse, but there is no reason to believe that the effect is eliminated. It should also be noted that estimating $\pi(x_T)$ in our model formulation by maximum likelihood will not generate stationary Markov chains. Another aspect of our model formulation that should worry us is the number of parameters, which is $2^{qr}-1 = 4\,095$ when $q = 4$, $r = 3$, the minus one coming from the restriction $\sum_{x_T} \pi(x_T) = 1$. The very high number of parameters results in overfitting and thereby in the areas in the realisations where the fitted model appear to have lost all memory of the training image. In the next sections we discuss two strategies for dealing with the non-stationarity and also consider one possibility for how to reduce the number of parameters.

3 A stationary variant of the naïve approach

The distribution $p(x)$ defined in Section 2 is defined via $\pi(x_T)$ and two Markov chains. An "obvious" way to make $p(x)$ stationary is to restrict the distribution $\pi(x_T)$ to be such that the two Markov chains become stationary. It is reasonably easy to see that the first Markov chain, defining $p(x_{1:q,1:m})$, is stationary if and only if $\pi(x_T)$ is such that the corresponding marginal distribution of the first $r-1$ columns equals the marginal distribution of the last $r-1$ columns. Thus, mathematically the requirement is that

$$\sum_{x_{1:q,1}} \pi([x_{1:q,1}, x_{1:q,2:r}]) = \sum_{x_{1:q,r+1}} \pi([x_{1:q,2:r}, x_{1:q,r+1}]) \quad \text{for all} \quad x_{1:q,2:r}. \tag{2}$$

It is less easy to find sufficient conditions for $\pi(x_T)$ that ensures that the second Markov chain is stationary. A necessary condition is clearly the corresponding requirement as (2) for rows, i.e.

$$\sum_{x_{1,1:r}} \pi\left(\begin{bmatrix} x_{1,1:r} \\ x_{2:q,1:r} \end{bmatrix}\right) = \sum_{x_{q+1,1:r}} \pi\left(\begin{bmatrix} x_{2:q,1:r} \\ x_{q+1,1:r} \end{bmatrix}\right) \quad \text{for all} \quad x_{2:q,1r}. \tag{3}$$

In Toftaker and Tjelmeland (2012b) it is shown that sufficient conditions for the second Markov chain to be stationary is (3) and the following two conditional independence assumption for $\pi(x_T)$,

$$x_{1,1:r-1} \perp x_{2:q,r} | x_{2:q,1:r-1} \quad \text{and} \quad x_{q,1:r-1} \perp x_{1:q-1,r} | x_{1:q-1,1:r-1}. \tag{4}$$

These conditional independence assumptions are illustrated in Figure 2. The vari-

Fig. 2 Illustration of the conditional independence assumptions in (4).The variables associated with the nodes in one of the two gray blocks are assumed conditionally independent of the variables associated to the nodes in the other grey block given variables associated to the nodes in the block coloured black.

ables associated with the nodes in one of the two gray blocks are assumed conditionally independent of the variables associated to the nodes in the other grey block given variables associated to the nodes in the block coloured black.

To fit a stationary model $p(x)$ to a training image we follow a modified version of strategy (ii) discussed in Section 2. We consider all $q \times r$ blocks in the training image as independent observations from $\pi(x_T)$ and maximise numerically the corresponding likelihood under the restrictions given by (2), (3) and (4). For each of the three training images in Figure 1, Figure 3 shows one realisation from the

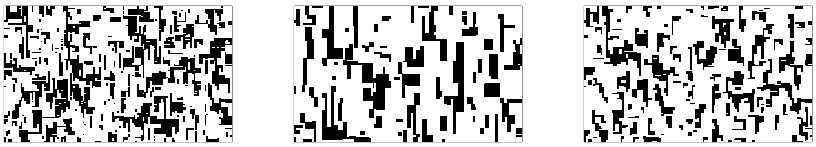

Fig. 3 A stationary model: For each of the three training images in Figure 1, one realisation from a fitted $p(x)$ when $q = r = 3$ and the restrictions in (2), (3) and (4) are enforced.

corresponding fitted $p(x)$ when $q = r = 3$. By construction the fitted models are stationary. From the fitted $\pi(x_T)$ one can easily compute the expected fraction of each facies and it is very close to the corresponding observed figures in the training images. The realisations in Figure 3, however, clearly reveal that restrictions (2), (3) and (4) are far too restrictive to allow reproduction of the characteristics of the three training images.

4 Defining the model component on a cylinder

In this section we consider a model specification procedure that is slightly modified relative to what we discuss in Section 2. In stead of starting with $\pi(x_T)$ for a

small $q \times r$ block we start by specifying $\pi(x_R)$ for a $q \times n$ strip, i.e. $R = \{(i, j), i = 1, \ldots, q, r = 1, \ldots, n\}$, and set

$$p(x_R) = \pi(x_R). \tag{5}$$

From $p(x_R)$ we define $p(x)$ exactly as discussed in Section 2. To ensure that the associated Markov chain is stationary we must have a restriction on $\pi(x_T)$ corresponding to the one in (3). More precisely we must have

$$\sum_{x_{1,1:n}} \pi\left(\begin{bmatrix} x_{1,1:n} \\ x_{2:q,1:n} \end{bmatrix}\right) = \sum_{x_{q+1,1:n}} \pi\left(\begin{bmatrix} x_{2:q,1:n} \\ x_{q+1,1:n} \end{bmatrix}\right) \quad \text{for all} \quad x_{2:q,1:n}. \tag{6}$$

To ensure this property we first define a Markov random field on a cylinder that is n nodes long and have $s \geq q$ nodes around. We define the Markov random field to have rectangular cliques of size $q \times r$ and make the distribution of the Markov random field invariant under rotations around the cylinder by assuming the potential function of all such cliques to be equal. We define $\pi(x_R)$ to be equal to a $q \times n$ section of the cylinder. Thus, we find $\pi(x_R)$ by marginalising over the remaining $(s-q) \times n$ nodes. This marginalisation operation is computationally intensive and limit the values of s and r that can be used to define the model.

To fit a model to a training image we consider all $q \times n$ strips in the training image to be independent observations from $\pi(x_R)$ and maximise the corresponding likelihood. For each of the three training images in Figure 1, Figure 4 shows three realisations from the corresponding fitted $p(x)$ when $q = 3$, $s = 5$ and $r = 3$. To avoid overfitting we include only interactions up the fourth order in the model. This results in a model with 158 parameters. We have found the fitting to be the best if we fit the model to a transposed version of the training images, and this is what is shown in the figure.

From the realisations we can observe that the fitted model captures well the characteristics of the leftmost training image, whereas the black objects in the realisations in the middle column is somewhat larger than in the corresponding training image. The fitted model is not at all able to reproduce the channel structures of the rightmost training image. A more detail analysis of the fitted models, and fitted models for other values of q, r and s can again be found in Toftaker and Tjelmeland (2012b).

5 The naïve approach with a reduced parameter set

To avoid the overfitting observed in Section 2, we define a model where only interactions up to order t are included. We fit the model by maximum likelihood as mentioned in fitting alternative (iii) in Section 2. The resulting $p(x)$ will still be non-stationary, so the best would be to define the model $p(x)$ on a larger lattice and fit the training image to the essentially stationary part in the middle of the lattice. However,

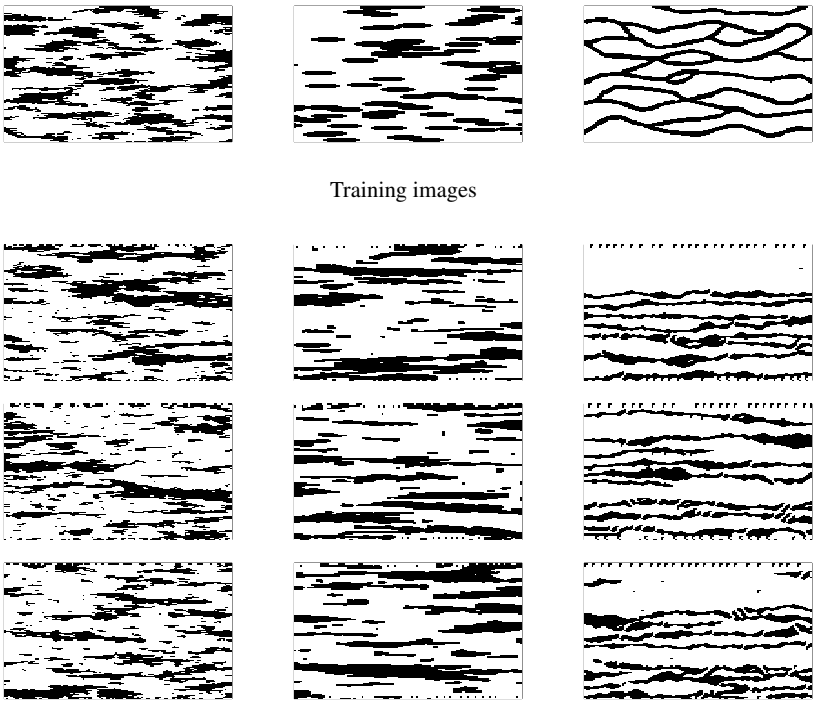

Fig. 4 The cylinder model: The three training images (upper row), and for each of these three realisations from the fitted $p(x)$ when $q = 3$, $s = 5$ and $r = 3$.

this is computationally somewhat more complicated to handle, so we have not implemented this variant yet. Figure 5 shows realisations from the fitted model when $q = 4$, $r = 3$ and $t = 4$, which gave a model with 793 parameters. We can observe that fitted models capture very well the characteristics of the two training images on each side, whereas the black objects in the realisations in the middle column are also here too large compared to corresponding objects in the training image. In particular we can for all the realisations observe that they include no artifacts resulting from overfitting as we observed in the realisations in Figure 1. We have also studied a much larger number of realisations from the fitted models without finding any instances of such artifacts, so to limit the interaction order as we have done seem to be a reasonable strategy to avoid overfitting. To fit successfully a model to the training image in the middle column, however, it seems necessary to include interactions of higher order than four. Again a more detailed study of the fitting procedure can be found in Toftaker and Tjelmeland (2012b).

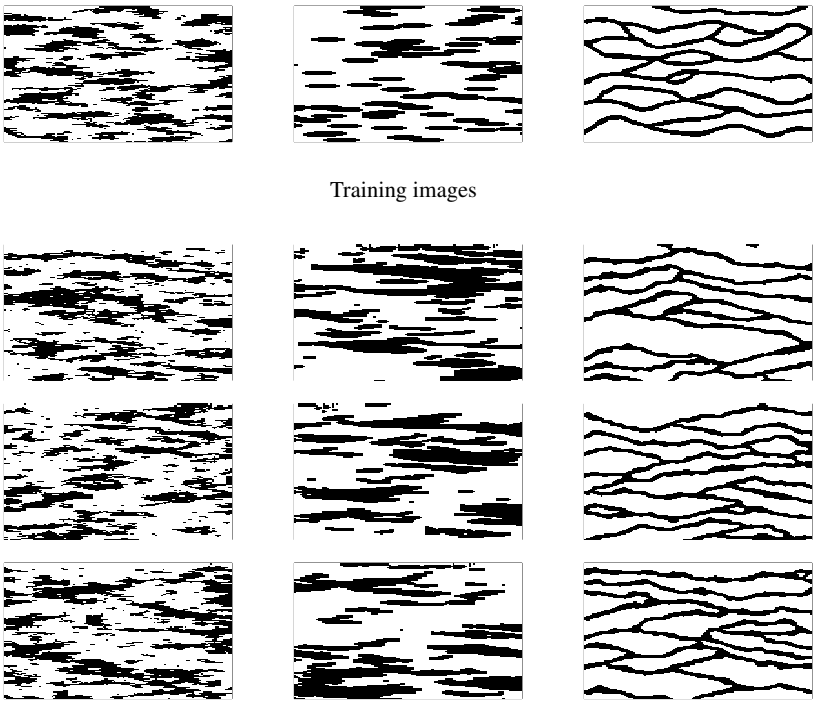

Fig. 5 The naive model with a reduced parameter set: The three training images (upper row), and for each of these three realisations from the fitted model when $q = 4$, $r = 3$ and the interactions up to order $t = 4$ are included in the model.

6 Closing remarks

Inspired by the multi-point statistics formulation we have defined and explored several model formulations that can be fitted to one or more training images. We have limited the attention to models where it is easy to calculate the probability of a realisation as then the corresponding conditional model given some data and a likelihood is immediately defined.

Our simulation results presented above can of course first of all be used to evaluate the model formulations we have proposed. However, the results can also be used to improve our understanding of the properties of the MPS model. Given the simulation order the MPS model is just as our models defined as a Markov chain. Just as our models in Sections 2 and 5 there is nothing in the model formulation ensuring this Markov chain to be stationary. Moreover, as we saw for our model in Section 2, estimating the model parameters from the frequency of configurations in the

training image can perfectly well result in a Markov chain where, for example, the marginal distribution of the limiting distribution is significantly different from what is observed in the training image. We can see no reason why including a random simulation order or a multi-grid should change this situation. Another problematic part of the MPS formulation is the number of parameters that have to be estimated. One should note that the Markov mesh model and partially ordered Markov model formulations also define the model via a Markov chain, so the situation is the same for such models. Our simulations in Section 2 also show another problematic side with the MPS formulation, namely the very high number of parameters in the model which may result in unwanted properties in the fitted model.

Our models defined in Sections 2 and 3 are clearly not useful for model fitting. The formulation in Section 4 may be useful for larger values of q, r and s, but the model fitting process will then be computationally very expensive. We think our formulation in Section 5, preferably defined on a larger lattice than the training image as discussed in Section 5, is useful and a better alternative than both MPS and Markov random fields. However, more consideration should be given to how to parameterise the model, preferably both the number of parameters and which parameters to include in the model should be decided as part of the estimation process and not apriori defined as we have done here. Moreover, the model is only computationally feasible in the 2D situation. A corresponding 3D model can easily be defined but is not computationally feasible.

Acknowledgements We thank the sponsors of the Uncertainty in Reservoir Evaluation (URE) project at the Norwegian University of Science and Technology.

References

Austad, H. and Tjelmeland, H. (2011). An approximate forward-backward algorithm applied to binary Markov random fields, *Technical report 11/2011*, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway. Available from http://www.math.ntnu.no/preprint/statistics/.

Brooks, S., Gelman, A., Jones, G. and Meng, X. L. (2011). *Handbook of Markov chain Monte Carlo*, Chapman & Hall/CRC, London.

Cressie, N. and Davidson, J. (1998). Image analysis with partially ordered Markov models, *Computational Statistics and Data Analysis* **29**: 1–26.

Gamerman, D. (1997). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*, Chapman & Hall, London.

Gelman, A. and Meng, X.-L. (1998). Simulating normalizing constants: from importance sampling to bridge sampling to path sampling, *Statistical Science* **13**: 163–185.

Geyer, C. J. and Thompson, E. A. (1995). Annealing Markov chain Monte Carlo with applications to ancestral inference, *Journal of American Statistical Associa-*

tion **90**: 909–920.

Gu, M. G. and Zhu, H. T. (2001). Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stochastic approximation, *Journal of the Royal Statistical Society, Series B* **63**: 339–355.

Journel, J. and Zhang, T. (2006). The necessity of a multiple-point prior model, *Mathematical Geology* **38**: 591–610.

Møller, J., Pettitt, A., Reeves, R. and Berthelsen, K. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants, *Biometrika* **93**: 451–458.

Scott, A. L. (2002). Bayesian methods for hidden Markov models: Recursive compution in the 21st century, *Journal of the American Statistical Association* **97**: 337–351.

Stien, M. and Kolbjørnsen, O. (2011). Facies modeling using a Markov mesh model specification, *Mathematical Geosciences* **43**: 611–624.

Strebelle, S. (2002). Conditional simulation of complex geolgical structures using multiple-point statistics, *Mathematical Geology* **34**: 1–21.

Tjelmeland, H. (1997). Modeling of the spatial facies distribution by Markov random fields, *in* E. Y. Baafi and N. A. Schofield (eds), *Geostatistics Wollongong '96*, Vol. 1, Kluwer Academic Publishers, London, pp. 512–523. Procedings from the 5th International Geostatistical Congress, Wollongong, Australia, 1996.

Tjelmeland, H. and Austad, H. (2012). Exact and approximate recursive calculations for binary Markov random fields defined on graphs, *Journal of Computational and Graphical Statistics* **21**. To appear.

Toftaker, H. and Tjelmeland, H. (2012a). Construction of binary multi-grid markov random field prior models from training images, *Technical report 1/2012*, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway. Available from http://www.math.ntnu.no/preprint/statistics/.

Toftaker, H. and Tjelmeland, H. (2012b). Construction of discrete prior distributions from training images, *Technical report*, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway. In preparation.