Construction of binary multi-grid Markov random field models from training images

Håkon Toftaker and Håkon Tjelmeland

Abstract We define the class of multi-grid discrete Markov random field (MRF) models and discuss how to estimate associated model parameters from a given training image. The intention is to use the resulting model as a prior for the spatial facies distribution in a Bayesian model. The multi-grid MRF model includes normalisation constants which it is computationally infeasible to compute. To cope with this complication we use a partially order Markov model (POMM) approximation to each MRF included in the multi-grid MRF model. We thereby get an explicit expression for the resulting estimated (approximate) multi-grid MRF model. This enables direct unconditional simulation from the model. Moreover, used as a prior model in a Bayesian context, we also have an explicit expression, up to an normalising constant, of the corresponding posterior. The Metropolis–Hastings algorithm can thereby be used to generate samples from the posterior. It is also possible to adopt once more the POMM approximation to MRF idea and generate realisations from a POMM approximation to the posterior distribution without resorting to iterative algorithms.

1 Introduction

When modelling the spatial facies distribution in for example petroleum reservoirs it has become common practice to fit a model to a training image which is believed to represent the spatial phenomenon under study. Multi point statistics (MPS) (Strebelle, 2002; Journel and Zhang, 2006) are often used as a modelling strategy in this

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

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

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

situation. The MPS models are constructed so that it is straight forward to fit the model to a training image. Unconditional simulation from MPS models is also easy. Many of the various MPS models that have been proposed are quite successful in reproducing the characteristics of the training image in unconditional realisations from the fitted model. As such, the MPS modelling strategy is successful. However, there are two important problems related with MPS. First, the number of model parameters that has to be estimated from the training image is typically huge. This clearly makes the model flexible and able to adapt to a wide variety of training images, but it also imply a risk of over fitting. Second, the fitted model is algorithmically defined and comes with no closed form, easy to compute expression for the probability of each possible realisation. This represents no problem as long as the interest is only in the (unconditional) fitted model, but is a major complication if one has also available data that one wants to condition on. In particular, it is not clear how to simulate realisations from the conditional distribution corresponding to the unconditional model and observed data. The standard MPS solution to this problem is to modify somehow the algorithm used to simulate unconditionally, such that the observed data are reproduced. The details of how to modify the algorithm depend on the type of MPS model used and the type of data observed. For some types of data this strategy is quite successful, at least as far as it is possible to see by visual inspection of the generated facies fields.

Alternatives to the MPS modelling strategy includes Markov mesh and partially ordered Markov models (POMM) (Abend et al., 1965; Cressie and Davidson, 1998) and Markov random fields (MRF) (Besag, 1974, 1986; Geman and Geman, 1984; Hurn et al., 2003). However, there are complications also related to these two modelling strategies. Markov mesh models and POMMs are defined through a unilateral path and as a result the realisations from such models typically have a strong directionally induced by this path. See however Stien and Kolbjørnsen (2011) for a Markov mesh formulation which seems almost to have overcome this complication. The main problem with the MRF modelling strategy is that it includes a computationally intractable normalising constant and this is a major complication in the model fitting phase because this normalising constant is a function of the model parameters that we want to estimate. Thereby, it becomes difficult to find for example the maximum likelihood estimator. See however Geyer and Thompson (1995), Descombes et al. (1995) and Tjelmeland and Besag (1998) for a possible procedure based on Markov chain Monte Carlo (MCMC) simulation.

In the present paper we define a multi-grid MRF model. Thus our model consists of a hierarchy of MRF models where the distribution on one level depends on values on previous levels. Clearly the computationally intractable normalising constant complication of MRFs is inherited into our multi-grid MRF formulation. Our strategy to cope with this problem is to approximate each MRF with a POMM by adopting the approximation procedure introduced in Austad and Tjelmeland (2011). This approximation can be used both in the model fitting phase and for (approximate) unconditional and conditional simulation from the fitted model. As the POMM approximation is a (reasonably good) approximation to an MRF the directionality problem of typical POMMs discussed above does not become a severe problem in our case.

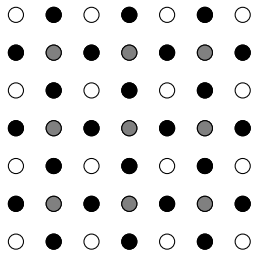

Fig. 1 Illustration of the splitting of S for a toy 7×7 lattice into three sub-lattices S_1, S_2 and S_3. The white, gray and black nodes are in S_1, S_2 and S_3, respectively.

In the following sections we discuss the most important aspects of our multi-grid MRF model and provide some simulation examples. A more detailed discussion can be found in Toftaker and Tjelmeland (2012). Both in the present paper and in Toftaker and Tjelmeland (2012) we limit the attention to the binary case, but the modelling procedure can be generalised to a situation with more than two facies.

2 Multi-grid MRF

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 lattice nodes. To node $(i,j) \in S$ we associate a binary variable $x_{ij} \in \{0,1\}$. In the following we use the notations $x = (x_{ij}, (i,j) \in S)$, $x_A = (x_{ij}, (i,j) \in A)$ and $x_{-A} = x_{S \setminus A}$ when $A \subseteq S$, and $x_{-(i,j)} = x_{-\{(i,j)\}}$ when $(i,j) \in S$.

In the multi-grid approach the nodes in S are split into a series of an odd number, T say, of sub-lattices, which we denote by $S_1,\ldots,S_T$. Figure 1 illustrates this for a small toy example where $n = m = 7$ and $T = 3$. The sub-lattice S_1 is an $n_1 \times m_1$ rectangular lattice, where $n_1 < n$ and $m_1 < m$. The sub-lattice S_2 form an $(n_1 - 1) \times (m_1 - 1)$ rectangular lattice, and the nodes in S_2 are placed between the nodes in S_1 as illustrated in Figure 1. Correspondingly, the nodes in S_3 are placed between the nodes in $S_1 \cup S_2$, and this process is continued until S_T.

To define a distribution for x, $p_\theta(x)$, we define a marginal distribution for x_{S_1}, and for each $t = 2,\ldots,T$ a conditional distribution for x_{S_t} given $x_{S_{1:t-1}}$, where $S_{1:t-1} = S_1 \cup \ldots \cup S_{t-1}$. Thereby we get

$$p_\theta(x) = p_{\theta_1}(x_{S_1}) \prod_{t=2}^{T} p_{\theta_t}(x_{S_t} | x_{S_{1:t-1}}), \tag{1}$$

where we have a separate parameter vector for each of the T distributions, so $\theta = (\theta_1,\ldots,\theta_T)$. We let each of the T distributions on the right hand side of (1) be MRFs. The $p_{\theta_1}(x_{S_1})$ is defined by an energy function $U_{\theta_1}(x_{S_1})$ via the relation $p_{\theta_1}(x_{S_1}) \propto \exp\{-U_{\theta_1}(x_{S_1})\}$ and, for each $t = 2,\ldots,T$, the MRF $p_{\theta_t}(x_{S_t} | x_{S_{1:t-1}})$ is correspondingly defined by an energy function $U_{\theta_t}(x_{S_t}, x_{S_{1:t-1}})$ via the relation $p_{\theta_t}(x_{S_t} | x_{S_{1:t-1}}) \propto \exp\{-U_{\theta_t}(x_{s_t}, x_{S_{1:t-1}})\}$, where the proportionality is as a function

of x_{S_t}. Related to simulation examples, we discuss in Section 6 one possible choice of neighbourhood structures and energy functions for these MRFs.

3 The POMM approximation

In the following we first give a very brief review of the strategy used in Austad and Tjelmeland (2011) for approximating a given MRF with a POMM. Thereafter we discuss how to adapt this procedure to the multi-grid MRF situation.

3.1 POMM approximations of an MRF

The starting point in Austad and Tjelmeland (2011) is that a binary MRF $p_\theta(x) = c(\theta)\exp\{-U_\theta(x)\}$ is assumed given, where $c(\theta)$ is the normalising constant and $U_\theta(x)$ is the energy function. In principle the normalising constant $c(\theta)$ can be computed by summing $p_\theta(x)$ over each component of x in turn and require the sum to be equal to one. Moreover, as a result of this summation process the MRF $p_\theta(x)$ can be reformulated as a POMM. Except for MRFs defined on very small lattices, however, the summation procedure require far too much computation time to be computationally feasible in practice. Austad and Tjelmeland (2011) define a computationally feasible approximate variant of the summation process by including an approximation step before summing out a component whenever it would require too much computation time to do the exact summation. Thereby an approximation $\widetilde{c}(\theta)$ of $c(\theta)$, and a POMM approximation $\widetilde{p}_\theta(x)$ of $p_\theta(x)$ are obtained.

If the interest is in the conditional distribution $p_\theta(x_A|x_{-A})$ corresponding to $p_\theta(x)$, two variants of the summation procedure discussed above can be defined. The first possibility is to set in actual values for x_{-A}, if such are available. The resulting $p_\theta(x_A|x_{-A})$ is then an MRF, just as the original $p_\theta(x)$, and thereby the exact or approximate summations procedures discussed above may be applied. The second alternative is to consider $p_\theta(x_A|x_{-A})$ as a function of both x_A and x_{-A}. In the summation procedure we may then sum out only the components of x that are in A. In the exact, but computationally infeasible, variant of the algorithm we then get a POMM formulation of $p_\theta(x_A|x_{-A})$ and the normalising constant $c(\theta,x_{-A})$ in $p_\theta(x_A|x_{-A}) = c(\theta,x_{-A})\exp\{-U_\theta(x)\}$ as a function of x_{-A}. In the approximate version of the algorithm we correspondingly get a POMM approximation of $p_\theta(x_A|x_{-A})$ and an approximation of the normalising constant $c(\theta,x_{-A})$, also as a function of x_{-A}.

3.2 The POMM approximation of a multi-grid MRF

To obtain a POMM approximation of the multi-grid MRF $p_\theta(x)$ defined in Section 2 we can make an approximation to each of the MRF components. For the first component, $p_{\theta_1}(x_{S_1})$, we get a POMM approximation $\widetilde{p}_\theta(x_{S_1})$ by summing over all components in x_{S_1}.

When it comes to $p_{\theta_t}(x_{S_t}|x_{S_{1:t-1}})$ we can consider this as a conditional variant of the unconditional MRF

$$f_{\theta_t}(x_{S_{1:t}}) \propto \exp\{-U(x_{S_t}, x_{S_{1:t-1}})\} \tag{2}$$

and adopt one of the approximation strategies discussed in Section 3.1. When doing model fitting, actual values for $x_{S_{1:t-1}}$ are available from the training image. Thereby the approximation procedure discussed in Section 3.1 where actual values are inserted for conditional values can be used. Another situation where actual values of the conditional variables are available is when doing unconditional simulation from the fitted model. Then one should start by sampling x_{S_1} from $\widetilde{p}_{\theta_1}(x_{S_1})$. Thereby these simulated values are available for x_{S_1} when one should sample x_{S_2}. Next the simulated values for $x_{S_{1:2}}$ are available when one should sample from $p_{\theta_3}(x_{S_3}|x_{S_{1:2}})$, and so on.

When the interest is in generating realisations from a fitted multi-grid MRF model conditioned to observed data, it is no longer possible to simulate $x_{S_1}, x_{S_2}, \ldots, x_{S_T}$ sequentially as in the unconditional case, and thereby we need an approximation of $p_{\theta_t}(x_{S_t}|x_{S_{1:t-1}})$ as a function of both x_{S_t} and $x_{S_{1:t-1}}$. Starting with the MRF (2) we should then apply the approximation strategy discussed in Section 3.1 where one sums over x_A with $A = S_t$ only. We thereby obtain both an approximation $\widetilde{c}(\theta_t, x_{S_{1:t-1}})$ of the normalising constant $c(\theta_t, x_{S_{1:t-1}})$ in $p_{\theta_t}(x_{S_t}|x_{S_{1:t-1}}) = c(\theta_t, x_{S_{1:t-1}}) \exp\{-U(x_{S_t}, x_{S_{1:t-1}})\}$, and a POMM approximation $\widetilde{p}_{\theta_t}(x_{S_1}|x_{S_{1:t-1}})$, both as a function of $x_{S_{1:t-1}}$. The approximation of the normalising constant can be used to define the following approximation of $p_\theta(x)$,

$$\widetilde{p}_\theta(x) \propto \exp\{-U_{\theta_1}(x_{S_1})\} \prod_{t=2}^{T} \left[\widetilde{c}(\theta_t, x_{S_{1:t-1}}) \exp\{-U_{\theta_t}(x_{S_t}, x_{S_{1:t-1}})\}\right]. \tag{3}$$

This approximation is not a POMM, but except for the associated normalising constant $\widetilde{p}_\theta(x)$ can easily be evaluated for any x and will be our starting point when defining the algorithm for generating conditional realisations in Section 5. From the POMM approximation $\widetilde{p}_{\theta_t}(x_{S_t}|x_{S_{1:t-1}})$ of $p_{\theta_t}(x_{S_t}|x_{S_{1:t-1}})$ we also get a POMM approximation of $p_\theta(x)$,

$$p^\star_\theta(x) = \widetilde{p}_{\theta_1}(x_{S_1}) \prod_{t=2}^{T} \widetilde{p}_{\theta_t}(x_{S_t}|x_{S_{1:t-1}}). \tag{4}$$

4 Parameter estimation

Let now x denote the training image to which we want to fit the multi-grid MRF defined in Section 2. We estimate the parameter vector θ by maximum likelihood. Because we have a separate parameter vector θ_t for each MRF component in the multi-grid MRF, the maximisation can be done with respect to each θ_t separately. Thereby we have to maximise $p_{\theta_1}(x_{S_1}) = c(\theta_1)\exp\{-U_{\theta_1}(x_{S_1})\}$ with respect to θ_1, and for each $t = 2,\ldots,T$ maximise

$$p_{\theta_t}(x_{S_t}|x_{S_{1:t-1}}) = c(\theta_t, x_{S_{1:t-1}})\exp\{-U_{\theta_t}(x_{S_t}, x_{S_{1:t-1}})\} \tag{5}$$

with respect to θ_t. One should remember that the normalising constants of these MRFs are computationally intractable, so direct numerical maximisations are not possible. It may be tempting to find an approximation to the maximum likelihood estimator by optimising numerically the corresponding POMM approximations in stead, but this may become problematic as the POMM approximations are not continuous functions of the parameters. Numerical optimisations may therefore quickly become stuck in a local maxima originating from the approximation. In stead we adopt the POMM approximations as proposal distributions in an importance sampling setting. Thereby we obtain estimates of the computationally intractable normalising constants that are smooth functions of the model parameters and these can be used as basis for numerical optimisation. A more detailed discussion of our maximisation procedure can be found in Section 4.3 of Toftaker and Tjelmeland (2012).

5 Unconditional and conditional simulation

No direct methods are available for unconditional simulation from a fitted multi-grid MRF. However, efficient simulation is possible from approximations to the multi-grid MRF. As also discussed in Section 3.2, we can sample $x_{S_1},\ldots,x_{S_T}$ sequentially. Thus, we first establish a POMM approximation $\widetilde{p}_{\theta_1}(x_{S_1})$ to $p_{\theta_1}(x_{S_1})$ and generate a corresponding sample x_{S_1} from this POMM approximation. Next we insert the actual simulated values for x_{S_1} into the expression for $p_{\theta_2}(x_{S_2}|x_{S_1})$, establish a POMM approximation $\widetilde{p}_{\theta_2}(x_{S_2}|x_{S_1})$ to $p_{\theta_2}(x_{S_2}|x_{S_1})$ and generate x_{S_2} from this POMM approximation. We next insert the actual simulated values for $x_{S_{1:2}}$ into the expression for $p_{\theta_3}(x_{S_3}|x_{S_{1:2}})$ and so on. If only one or a few unconditional realisations are required this is a feasible simulation procedure. One should note, however, that the computationally most expensive part of the procedure is to establish a POMM approximations to a give MRF, and this has to be done $T-1$ times for each new realisation. So if a large number of unconditional realisations should be generated it would be better to start out be establishing the POMM approximation (4) to the whole multi-grid MRF. Having done this, a very large number of unconditional realisations from (4) can be generated very efficiently.

As discussed in Section 3.2, when conditioning to data it is not possible to simulate $x_{S_1},\ldots,x_{S_T}$ sequentially as in the unconditional case. Instead one should start out with the approximation $\widetilde{p}_\theta(x)$ to $p_\theta(x)$ defined in (3), and form the corresponding conditional distribution $\widetilde{p}_\theta(x|z) \propto \widetilde{p}_\theta(x)\psi(z|x)$, where z is the observed data and $\psi(z|x)$ the associated likelihood. Next one can either simulate from $\widetilde{p}_\theta(x|z)$ by a Metropolis–Hastings algorithm, or one may consider $\widetilde{p}_\theta(x|z)$ as just a new MRF to which we can establish a POMM approximation. Having established the POMM approximation to the conditional distribution we can efficiently generate a large number of realisations from it.

6 Simulation examples

We evaluate the performance of our multi-grid MRF scheme by applying it to the three 121×121 training images shown in the upper row of Figure 2. In each case we

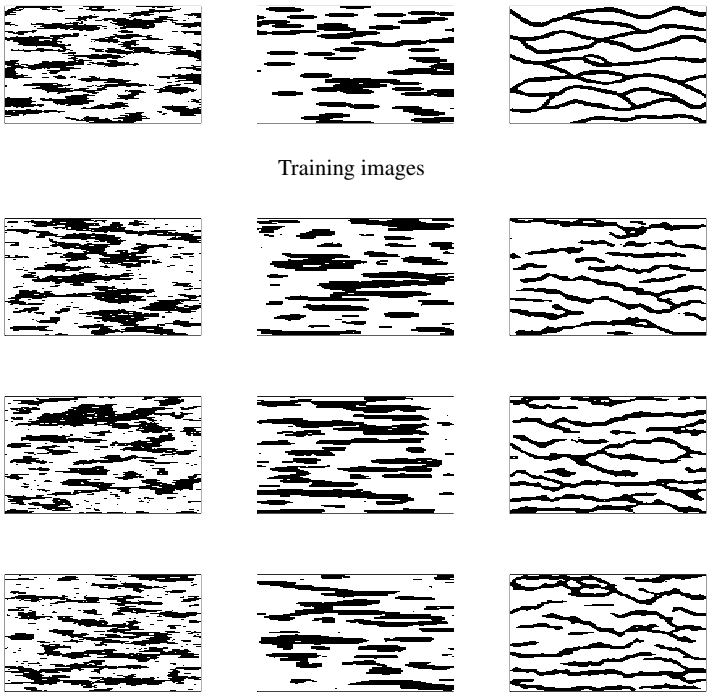

Fig. 2 The three training images (upper row) and for each of the three training images three realisations from the fitted multi-grid MRF.

let S_1 be a 16×16 lattice and use $T = 7$ sub-lattices. In the following we discuss the most important parts of the definition of the energy functions for the various MRF components in our multi-grid MRF model, for a more detailed definition see Toftaker and Tjelmeland (2012). In all the MRFs defining the multi-grid MRF we use a second order neighbourhood system (Besag, 1974), so each interior node $(i, j) \in S_t$ has the eight nearest nodes in S_t as its neighbours. Nodes on the boundary of the lattice have correspondinly fewer neighbours. This gives ten types of cliques and, assuming the potential functions to be translation invariant, a corresponding number of model parameters. Without loss of generality the potential of one of these ten cliques can be set to zero, so we end up with nine model parameters for each level $t = 1, \ldots, T$. The largest cliques contain four nodes, so the model includes in addition to a first order effect and pairwise interactions between neighbour nodes, also triple and quadruple interactions. For $t \geq 2$, the components of $x_{S_{1:t-1}}$ are treated as covariates in the MRF model for x_{S_t}. Here node $(i, j) \in S_t$ is linked to the four nodes in $S_{1:t-1}$ located closest to (i, j), and these four nodes are allowed to influence the first order effect of node (i, j). We only include pairwise and triple interactions between (i, j) and the four closest nodes in $S_{q:t-1}$, which results in ten more parameters for each $t \geq 2$.

For each of the three training images in the upper row of Figure 2 we have fitted the multi-grid MRF described above, and thereafter generated three independent realisations from the fitted model as discussed in Section 5. The unconditional realisations are shown in the three lower rows of Figure 2. We can see that much of the characteristics of the training images are reproduced, but not all. For the training image to the left, the realisations contains too many small black objects relative to the training image. The same tendency can also be observed for the other two training images, but the effect is less clear here. For the training image to the right the fitted model is not able to reproduce the very strong continuity of the channels in the training image. In addition to the visual comparison of the training images with realisations from the fitted model considered here, Toftaker and Tjelmeland (2012) also do a comparison via a number of descriptive statistics.

As discussed in Section 5 we can also generate realisations from POMM approximation to a conditional distribution given some data. In Figure 3 we have assumed to have exact observations in two vertical wells, where the observed values are taken from the corresponding training image. We can observe that the wells do not "stand out" from the simulated values and that the characteristics of the conditional realisations is not much different from the corresponding unconditional realisations. Again Toftaker and Tjelmeland (2012) contain also comparisons via descriptive statistics.

7 Closing remarks

We have defined a multi-grid MRF and discussed how to use the POMM approximation of MRF idea handle the multi-grid MRF model computationally. In particular we have expressions for the probability distribution of the multi-grid MRF when

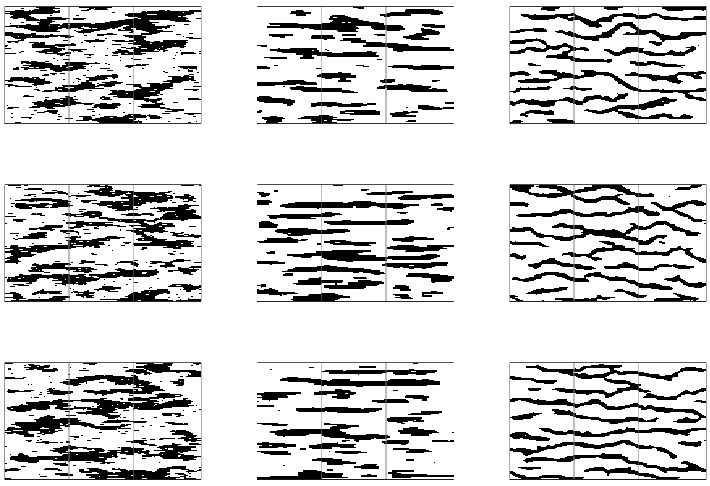

Fig. 3 For the fitted multi-grid MRF to each of the three training images in the upper row of Figure 2, three realisations from the POMM approximation of the corresponding conditional distribution given two vertical "wells".

fitted to a training image, and this allows us to construct general algorithms for generating realisations from the corresponding conditional distribution when data are observed. In simulation examples with three training images we have demonstrated the flexibility, and limitations, of our approach.

In our discussion, and in the examples, we have limited the attention to 2D binary models. Our modelling strategy can easily be extended to include more than two facies, but the computational complexity of the approach grows rapidly with the number of facies and neighbourhood size of the MRFs, so we expect the approach to feasible only for a reasonably low number of facies. A direct generalisation of our approach to a 3D situation is also possible, but for this to become computationally feasible the estimation and simulation procedures must be carefully implemented. A better alternative is perhaps to model the 3D case with a Markov chain of 2D multi-grid MRFs.

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

References

Abend, K., Harley, T. and Kanal, L. (1965). Classification of binary random patterns, *IEEE Transactions on Information Theory* **11**: 538–544.

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/.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society, Series B* **36**: 192–225.

Besag, J. (1986). On the statistical analysis of dirty pictures (with discussion), *Journal of the Royal Statistical Society, Series B* **48**: 259–302.

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

Descombes, X., Mangin, J., Pechersky, E. and Sigelle, M. (1995). Fine structures preserving model for image processing, *Proc. 9th SCIA 95, Uppsala, Sweden*, pp. 349–356.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**: 721–741.

Geyer, C. J. and Thompson, E. A. (1995). Annealing Markov chain Monte Carlo with applications to ancestral inference, *Journal of American Statistical Association* **90**: 909–920.

Hurn, M., Husby, O. and Rue, H. (2003). A tutorial on image analysis, *in* J. Møller (ed.), *Spatial statistics and computational methods*, Vol. 173 of *Lecture Notes in Statistics*, Springer, pp. 87–139.

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

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. and Besag, J. (1998). Markov random fields with higher order interactions, *Scandinavian Journal of Statistics* **25**: 415–433.

Toftaker, H. and Tjelmeland, H. (2012). 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/.