Doctoral theses at NTNU, 2019:16

Xin Luo

Bayesian inference for Markov mesh models - applied to inversion of seismic data

ISBN 978-82-326-3642-6 (printed ver.)
ISBN 978-82-326-3643-3 (electronic ver.)
ISSN 1503-8181

NTNU
Norwegian University of
Science and Technology

Doctoral theses at NTNU, 2019:16

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Information Technology and
Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

Xin Luo

Bayesian inference for Markov mesh models - applied to inversion of seismic data

Thesis for the Degree of Philosophiae Doctor

Trondheim, January 2019

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

ISBN 978-82-326-3642-6 (printed ver.)
ISBN 978-82-326-3643-3 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2019:16

Printed by NTNU Grafisk senter

Preface

The thesis is submitted in partial fulfillment of the requirements for the degree of Philosophiae Doctor (Ph.D.) at the Norwegian University of Science and Technology (NTNU). The research is funded by the URE (Uncertainty of Reservoir Evaluation) project at the Department of Mathematical Sciences, NTNU.

First of all, I would like to thank my exceptionally great advisor Professor Håkon Tjelmeland. He is one of the nicest professors I have ever met. I am faithfully grateful for his excellent supervision and solid support during these four years. Not only is he resembling a lighthouse in the ocean of my research that leads me to the bright future, but also I could talk with him about the life outside of the academic work. This experience is priceless to me. I also would like to thank Professor Henning Omre and Professor Jo Eidsvik for their help when I was participating in conferences and meetings.

Secondly, I would like to express my gratitude to my colleagues Jacob Skauvold and Torstein Fjeldstad. Jacob was also my office mate I have worked for four years with. He gave me both scientific and emotional help during these years. Torstein was one of the authors I cooperated with to complete paper III. He provided the source of data and pre-processed the data set. I thank them again for their help throughout my Ph.D. life.

Last but not least, I am particularly grateful to my parents for their financial support so that I had an opportunity to start a unique experience in Norway. I also would like to thank my friends for enriching and vitalizing my spare time. They are an unforgettable and invaluable part of my life.

Trondheim, Norway *Xin Luo*
September 2018

Thesis Outline

Background

Paper I **Prior specification for binary Markov mesh models**

Xin Luo and Håkon Tjelmeland

Accepted for publication in *Statistics and Computing*, doi: 10.1007/s11222-018-9813-7.
Presented at Petroleum Geostatistics, Biarritz, France 2015.
Presented at 10th International Geostatistical Congress, Valencia, Spain 2016.

Paper II **A multiple-try Metropolis–Hastings algorithm with tailored proposals**

Xin Luo and Håkon Tjelmeland

Submitted for publication (under revision).

Paper III **A Bayesian model for lithology/fluid class prediction using a Markov mesh prior fitted from a training image**

Håkon Tjelmeland, Xin Luo and Torstein Fjeldstad

Submitted for publication.

Background

1 Introduction

In the development of a petroleum reservoir, seismic data and observations in wells are two types of data giving information about the rock properties and the fluids present in a reservoir. To integrate the information in these observations with general geological knowledge and estimate the reservoir properties, it is natural to adopt a Bayesian formulation. Each type of observed data is modeled by a likelihood function while the general geological knowledge about the characteristics of the reservoir is quantified in a prior distribution. The likelihood function is typically constructed by laws of physics or other natural sciences. Seismic data are noisy and heavily blurred in the vertical direction, and well observations are sparse in an early phase of a reservoir development. This implies that the formulation of a realistic prior distribution is crucial, as the information content in the available data is not sufficient to dominate the properties of the posterior distribution. This is in contrast to the typical situation in image analysis applications, where token priors with totally unrealistic large scale properties often produce satisfactory estimation results.

Ideally, we want to derive the properties of the posterior distribution in an analytical form, but this is usually impossible in the case of reservoir modeling. Therefore, we need to address this by numerical techniques, i.e., estimating properties of the posterior distribution by stochastic simulation. The forms of the prior and likelihood models must therefore be chosen to make sampling from the resulting posterior distribution possible.

For the specification of a prior, it has over the past years become common to estimate a prior model for the spatial distribution of reservoir properties from one or several training images. A training image can be outcrop data from an area of a similar geological origin as the area under study, or it can be hand drawn or computer-generated scenes believed to have similar statistical properties as in the reservoir in focus for the study. The most popular alternative for prior modeling is multiple-point statistics (MPS), see for example Strebelle (2002) and Journel and Zhang (2006). MPS models aim at bridging the gap between physical realism and spatio-temporal stochastic modeling and they are algorithmically defined. The distribution of the estimated prior model is typically only implicitly defined through a simulation process in which the nodes are visited and simulated in a random order. Thus, it is infeasible in most cases to sample correctly from the resulting posterior when using such a prior model. The literature regarding MPS is focused on how to construct simulation algorithms that can reproduce observed data, but without conditioning on the data probabilistically correct. This is not going along with the manner that can be used for the

Bayesian setup with conditioning on observed data. To get a posterior that we are able to sample from, we need to come up with prior models ensuring an explicit expression is available for the posterior distribution, at least up to a normalizing constant. Then Markov random fields (MRFs) and Markov mesh models (MMMs) are two useful tools, see for example Kindermann and Snell (1980) and Hurn et al. (2003) for introductions to MRFs, and Abend et al. (1965) and Cressie (1993) for introductions to MMMs. Discrete MRFs can be used to model available prior information about unobserved data of a discrete variable. This prior is integrated with a likelihood function describing the relation between the unobserved and observed variables into a posterior distribution, and this posterior is fundamental for making inference about the unobserved variables. However, the usefulness of discrete MRFs is restricted by a computationally intractable normalizing constant. This makes any likelihood-based parameter estimation difficult, and a fully Bayesian approach where a prior is specified for the MRF parameters also becomes computationally intensive. In contrast, the normalizing constant for an MMM is computationally available, which makes the likelihood-based parameter estimation and a fully Bayesian approach for this model class much easier than for MRFs. In the following, we discuss MPS, MRFs and MMMs in more detail, and in particular express mathematically why we claim that MMMs are the preferred prior in the situation discussed above.

2 Multiple-point statistics

Multiple-point statistics (MPS) are used to describe the relation between multiple, spatially located points. In the relevant literatures, MPS is firstly introduced by Guardiano and Srivastava (1993) and then developed by Strebelle (2002) for a more detailed simulation process. Caers (2001) and Ortiz and Deutsch (2004) also propose variations of the MPS algorithm. There are a number of different MPS techniques that have been proposed, but they own an algorithmic core in common. In the following we introduce this common essence in general.

Consider an $m \times n$ $(m, n \geq 1)$ lattice $\chi = \{(i, j) : i = 1, \ldots, m; j = 1, \ldots, n\}$ and let $v \in \chi$ denote a node in this lattice. We associate a discrete value x_v to each node $v \in \chi$. For an arbitrary set $A \subseteq \chi$, we define x_A to be the set of values at the nodes corresponding to the elements of A, i.e., $x_A = \{x_v : v \in A\}$, and use $x = x_\chi$ for the values in the whole lattice.To simulate x, MPS first draws a random visiting order of the nodes, $(v_1, \ldots, v_K)$ say, where $K = mn$. Except for the first node v_1, the value of node $v_k, 1 < k \leq K$ can be simulated if the values of all previous nodes $v_1, \ldots, v_{k-1}$ are known. An

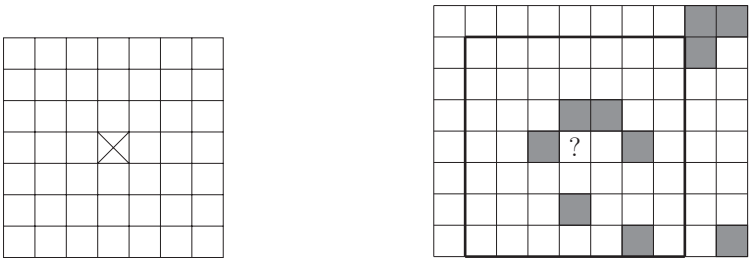

Figure 1: Illustration of MPS simulation based on a template neighborhood τ. The left figure shows one template neighborhood $\tau = \{(i,j)| -3 \leq i,j \leq 3\} \setminus \{(0,0)\}$, where the node $(0,0)$ is represented with $\boxtimes$ and all other nodes are the elements of τ. The right figure shows the resulting MPS simulation after 10 iterations in a 8×10 lattice where $v_{11} = (5,5)$ is the node with ? and $v_1, \ldots, v_{10}$ are represented by the nodes in gray. The black thick lines are the borders of the translation of τ.

MPS neighborhood ν_{v_k} with respect to a node $v_k \in \chi$ is defined to be a subset of the nodes simulated before v, i.e., $\nu_{v_k} \subseteq \{v_1, \ldots, v_{k-1}\}$. It is common to assume that the MPS neighborhood ν_{v_k} is generated by a translation of a template neighborhood τ by

$$\nu_{v_k} = (\tau \oplus v_k) \cap \{v_1, \ldots, v_{k-1}\}, \tag{1}$$

where $\tau \oplus (i,j) = \{(l+i, h+j) : (l,h) \in \tau)\}$. An example of τ and MPS simulation process is illustrated in Figure 1. The figure on the left depicts a template neighborhood $\tau = \{(i,j)| -3 \leq i,j \leq 3\} \setminus \{(0,0)\}$, where the node $(0,0)$ is represented with $\boxtimes$ and all other nodes are the elements of τ. The right figure shows the resulting MPS simulation after 10 iterations in a 8×10 lattice where $v_{11} = (5,5)$ is the node with ? and $v_1, \ldots, v_{10}$ are represented by the nodes in gray. The black thick lines represent the boundaries of the translation of τ, so there are six nodes in $\nu_{v_{11}}$.

The values in x are simulated by firstly simulating x_{v_1} from a $p(x_{v_1}|x_{\nu_{v_1}})$, then simulating x_{v_2} from a $p(x_{v_2}|x_{\nu_{v_2}})$, next simulating x_{v_3} from a $p(x_{v_3}|x_{\nu_{v_3}})$ and so on until x_{v_K} is simulated from a $p(x_{v_K}|x_{\nu_{v_K}})$. Thereby, the joint distribution in the whole lattice χ is given by

$$p(x) = \frac{1}{K!} \sum_{(v_1,\ldots,v_K)} \left[\prod_{i=1}^{K} p(x_{v_i}|x_{\nu_{v_i}}) \right], \tag{2}$$

where the sum is over all possible visiting orders for the nodes and the factor $\frac{1}{K!}$ comes from the random visiting order. The next step is to decide, for an arbitrary node $v \in \chi$, the conditional probability distribution

$$p(x_v|x_{\nu_v}), v \in \chi. \tag{3}$$

In MPS it is common to define (3) by scanning a training image for replicates of x_{ν_v}. This can be done under a prior assumption of stationarity. Let η denote the set of all nodes in the training image and z_u the value associated to node $u \in \eta$. Note that the size of the training image can be the same as that of the lattice χ or not. To estimate (3), the number of replicates of x_{ν_v}, denoted by n_1, in η needs to be counted. A replicate should have the identical geometric configuration and the same values, i.e.,

$$n_1 = \sum_{\substack{u \in \eta \\ (\nu_v \ominus v \oplus u) \subseteq \eta}} I\left(x_{\nu_v} = z_{\nu_v \ominus v \oplus u}\right), \tag{4}$$

where $I(\cdot)$ is the identity function and $\nu \ominus (i,j) = \{(l-i, h-j) : (l,h) \in \nu\}$. From the previous n_1 replicates, the number of replicates n_2 with z_u equal to x_v also needs to be counted, i.e.,

$$n_2 = \sum_{\substack{u \in \eta \\ (\nu_v \ominus v \oplus u) \subseteq \eta}} I\left(x_{\nu_v} = z_{(\nu_v \ominus v \oplus u) \cap \eta}, x_v = z_u\right). \tag{5}$$

Then the probability (3) is defined as

$$p(x_v|x_{\nu_v}) = \frac{n_2}{n_1}. \tag{6}$$

One should notice that the larger the size of ν_v is, the more specific the configuration x_{ν_v} is, and thereby we can find less number of replicates over the training image for estimating (6). In addition, the distribution (3) may be too specific to be estimated from the training image if ν_v is too large. Therefore, some articles define modifications of the procedure above, see for example Strebelle (2002) and Ortiz and Deutsch (2004).

If we use (2) as a prior and adopt the Metropolis–Hastings algorithm to simulate from the resulting posterior distribution, it becomes computationally infeasible to evaluate the acceptance probability. When putting (2) into the expression for the Metropolis–Hastings acceptance probability, there will be a ratio of the prior for the original value x to the prior for the proposed value x', i.e.,

$$\frac{p(x')}{p(x)} = \frac{\sum_{(v_1,\ldots,v_K)} \left[\prod_{i=1}^{K} p(x'_{v_i}|x'_{\nu_{v_i}})\right]}{\sum_{(v_1,\ldots,v_K)} \left[\prod_{i=1}^{K} p(x_{v_i}|x_{\nu_{v_i}})\right]}, \tag{7}$$

where both of the sums are over all possible permutations of the nodes in χ. This ratio does not cancel when computing the acceptance probability. Note that in (7) the number of terms in the summation in the numerator equals $K!$ as well as that in the denominator. This number increases very fast as K increases, so it is computationally intensive or infeasible to evaluate this ratio in general cases. In the next section, we discuss the effect of using a Markov random field prior in the same Bayesian formulation.

3 Markov random fields

In this section we introduce Markov random fields (MRFs) and discuss the use of MRFs as prior in a Bayesian model. For general theory and applications of MRFs, see for example Hammersley and Clifford (1971), Kindermann and Snell (1980) and Hurn et al. (2003).

Markov random fields are a collection of undirected graphical models frequently used in spatial statistics. We assume an undirected graph $\mathcal{G} = \{\chi, \mathcal{E}\}$ where χ is defined as in Section 2 and $\mathcal{E}$ is a set of edges $\mathcal{E} \subseteq \{\{u, v\} | u, v \in \chi, u \neq v\}$. To each node $v \in \chi$ it is associated a discrete value x_v. Following the notation in Section 2, we let x_A denote the vector of values for an arbitrary set $A \subseteq \chi$ and set $x = x_\chi$ for all nodes in χ. The neighborhood with respect to a node $v \in \chi$, denoted by ν_v, in MRF is a set of nodes directly connected to v in $\mathcal{E}$, i.e., $\nu_v = \{u : \{u, v\} \in \mathcal{E}\}$. An element of ν_v is called a neighbor of v. If all pairs of the nodes in a subset $W \subseteq \chi$ are neighbors, then W is called a clique. A maximal clique ω is a clique that is not a subset of another clique, and we denote by Ω the set of all maximal cliques of $\mathcal{G}$. Based on these concepts, x is an MRF with respect to $\mathcal{G}$ if the joint distribution parameterized by θ, $p(x|\theta)$, fulfills the positivity condition $p(x|\theta) > 0$ and has the local Markov property

$$p\left(x_v | x_{\chi \setminus \{v\}}, \theta\right) = p\left(x_v | x_{\nu_v}, \theta\right), \ \forall v \in \chi. \tag{8}$$

The Hammersley-Clifford (1971) theorem states that an x that has a probability distribution $p(x|\theta)$ is an MRF with respect to $\mathcal{G}$ if and only if it can be factorized over the maximal cliques of $\mathcal{G}$, i.e.,

$$p(x|\theta) = \frac{1}{C(\theta)} \cdot \exp\left\{ -\sum_{\omega \in \Omega} \Phi_\omega(x_\omega; \theta) \right\}, \tag{9}$$

where $C(\theta)$ is a normalizing constant and $\Phi_\omega(x_\omega; \theta)$ a clique potential function for the maximal clique ω. In other words, the Hammersley-Clifford theorem is equivalent to a sufficient and necessary condition with which a

positive probability distribution can be represented by a Markov random field.

Letting $U(x;\theta) = \sum_{\omega\in\Omega} \Phi_\omega(x_\omega;\theta)$, the normalizing constant is expressed as

$$C(\theta) = \sum_x \exp\left\{-U(x;\theta)\right\}, \tag{10}$$

where the sum is over all possible scenes x. The number of terms in the sum is thereby exponential in the number of nodes. In most cases $C(\theta)$ is therefore not computationally available. When it comes to an MRF being used as a prior, if we adopt the Metropolis–Hastings algorithm to simulate from the resulting posterior distribution, the ratio of the normalizing constants will cancel. Thereby,

$$\frac{p(x'|\theta)}{p(x|\theta)} = \frac{\exp\left\{-U(x';\theta)\right\}}{\exp\left\{-U(x;\theta)\right\}}, \tag{11}$$

where x and x' represent the current and potential new values, respectively, is simple to evaluate. However, fitting an MRF to a given training image is computationally intractable. The normalizing constant, $C(\theta)$, is a function of the parameter vector θ, so using the maximum likelihood principle to estimate θ is computationally intractable. To resolve this problem, one may estimate $C(\theta)$, as a function of θ, via Markov chain Monte Carlo simulation, see for example Geyer (1991) and Tjelmeland and Besag (1998). Also this process is, however, computationally intensive for models with large and complex interaction structures and many model parameters. To be able to fit also the neighborhood and interaction structure to a given training image, not only the values of the model parameters, a Bayesian formulation may be adopted. A prior for θ, and potentially also for the neighborhood and interaction structure of the MRF, can be formulated. Considering the training image as a sample from the MRF, one may fit the MRF to the training image by simulating from the resulting posterior distribution. However, if adopting a Metropolis–Hastings algorithm for this one runs into trouble. The Metropolis–Hastings acceptance probability will include a ratio of two normalizing constants,

$$\frac{C(\theta')}{C(\theta)}, \tag{12}$$

where $C(\theta)$ and $C(\theta')$ are the normalizing constants of the current and potential new MRFs, respectively. Standard Markov chain Monte Carlo can therefore not be used to simulate from such a posterior distribution. Møller et al. (2006), Murray et al. (2006), Walker (2011) and Lyne et al. (2015) device how to circumvent this problem by including an auxiliary variable and Arnesen and Tjelmeland (2017) propose to obtain an approximate solution

to the problem by replacing the MRF with a Markov mesh approximation. All these procedures are, however, computationally intensive and to the best of our knowledge it has not yet been demonstrated to work within reasonable computation time when the training image requires a large neighborhood and a complex interaction structure.

4 Markov mesh models

The class of Markov mesh models (MMMs) was introduced already in Abend et al. (1965) and was later generalized by Cressie and Davidson (1998). An example about how to fit an MMM is discussed in Stien and Kolbjørnsen (2011). We introduce MMMs in this section and show that this type of models can overcome the problems in the estimation of models and the simulation conditional on data that we have mentioned in Sections 2 and 3.

We assume an $m \times n$ lattice with the same definition as in Section 2. For MMMs, we assume a lexicographical numbering order of the nodes in the lattice χ, from one to mn. To each node $v \in \chi$ it is associated a discrete value x_v. Following the notation in Sections 2 and 3, let x_A denote the vector of values for an arbitrary set $A \subseteq \chi$ and set $x = x_\chi$ for all nodes in χ. For each node $v = (i, j) \in \chi$, the predecessor set ρ_v is defined to be

$$\rho_v = \{(k, l) \in \chi : nk + l < ni + j\}. \tag{13}$$

The MMM is then based on the distribution of x given by

$$p(x) = \prod_{v \in \chi} p(x_v | x_{\rho_v}). \tag{14}$$

In addition, the MMM adopts the Markov assumption that

$$p(x_v | x_{\rho_v}) = p(x_v | x_{\nu_v}), \tag{15}$$

where $\nu_v \subseteq \rho_v$ is called the sequential neighborhood. We assume that the sequential neighborhood ν_v is a translation of a template sequential neighborhood τ given by

$$\nu_v = (\tau \oplus v) \cap \chi, \tag{16}$$

where the operator $\oplus$ is the same as in Section 3. In the following assume that the MMM in (14) is a binary model i.e., $x_v = 0$ or 1 for all $x \in \chi$, and that the model is parameterized by $\theta = \{\theta(x_{\nu_v}) : v \in \chi\}$. Without loss of generality, the right-hand side of (15) can then be expressed as

$$p(x_v | x_{\nu_v}, \theta) = \frac{\exp\{x_v \cdot \theta(x_{\nu_v})\}}{1 + \exp\{\theta(x_{\nu_v})\}}, \tag{17}$$

where $\theta(x_{\nu_v})$ is a parameter value depending on the configuration of x_{ν_v}. Note that we substitute $p(x_v|x_{\nu_v})$ with $p(x_v|x_{\nu_v}, \theta)$ in order to represent that the model is parameterized by θ. Combining (14) with (15) and (17), the expression for the MMM becomes

$$p(x|\theta) = \prod_{v\in\chi} \frac{\exp\{x_v \cdot \theta(x_{\nu_v})\}}{1 + \exp\{\theta(x_{\nu_v})\}}. \tag{18}$$

When using an MMM as a prior, conditioning on data in the Bayesian setting and adopting the Metropolis–Hastings algorithm to simulate from the resulting posterior distribution, the acceptance probability will include the ratio

$$\frac{p(x'|\theta)}{p(x|\theta)} = \prod_{v\in\chi} \left[\exp\{x'_v \cdot \theta(x'_{\nu_v}) - x_v \cdot \theta(x_{\nu_v})\} \cdot \frac{1 + \exp\{\theta(x_{\nu_v})\}}{1 + \exp\{\theta(x'_{\nu_v})\}}\right], \tag{19}$$

where the current and potential values are denoted by x and x', respectively. The product in (19) is over all the nodes in χ, so it is computational available to evaluate it. When it comes to fitting an MMM to a training image, we can first choose a prior distribution for the parameters θ and let (18) be the likelihood function in the Bayesian setup, and then adopt the Metropolis–Hastings algorithm to simulate from the resulting posterior distribution. Assume that we choose a suitable prior $p(\theta)$ for θ so that when evaluating the acceptance probability it is valid to compute the ratio of the prior for the potential new MMM to that for the current MMM. The ratio of likelihoods is given by

$$\frac{p(x|\theta')}{p(x|\theta)} = \prod_{v\in\chi} \left[\exp\{x_v \cdot [\theta'(x_{\nu_v}) - \theta(x_{\nu_v})]\} \cdot \frac{1 + \exp\{\theta(x_{\nu_v})\}}{1 + \exp\{\theta'(x_{\nu_v})\}}\right], \tag{20}$$

where θ and θ' are the parameters for the current and potential new MMMs, respectively. This ratio is feasible to compute since (20) has only one factor for each node $v \in \chi$.

5 Summary of papers

The ultimate goal of my Ph.D. work was to construct a stochastic spatial model as prior when conditioning on real seismic data in the Bayesian setting and to simulate from the resulting posterior distribution. For the reasons discussed above we ended up using a Markov mesh prior. To focus mainly on the methodological aspects and to simplify implementation and simulation, we limited the attention to the binary case. We present our work in three

papers. The papers can be read independently, but we recommend reading paper I before papers II and III. The thesis is organized as follows.

In paper I, "Prior specification for binary Markov mesh models", we use the Bayesian framework to fit a Markov mesh model (MMM) to a given training image. We develop two equivalent parameterizations for the class of MMMs and use these to formulate a prior distribution for all parts of the specification of an MMM, namely the sequential neighborhood, the interaction structure and the parameter values. Assuming the given training image to be a realization generated from an MMM, we fit the MMM to the training image by simulating from the resulting posterior distribution. To sample from the posterior distribution we construct a reversible jump Markov chain Monte Carlo algorithm, where we propose to add or remove one interaction in each step. We test the fitting procedure on two training images, one with a very noisy pattern and the other with a more complex pattern. When simulating realizations from the posterior models, the spatial patterns in the training images are to a large degree reproduced. Not surprisingly, the generated posterior models have larger sequential neighborhoods and more complex interaction structures when using the training image with a more complex pattern than when using the more noisy training image. The computation time to generate one posterior model is acceptable for both training images, but to explore the whole state space is computer intensive. It is therefore of interest to consider how parallel computing can be used to make the process more effective.

In paper II, "A multiple-try Metropolis–Hastings algorithm with tailored proposals", we propose a novel Metropolis–Hastings (MH) algorithm with multiple proposals tailored to the target distribution. The algorithm is based on a directed acyclic graph in which one node, k say, represents the current state and the remaining nodes represent proposed states. In this algorithm, several new states are proposed in parallel and each of them is generated by using the identical proposal distribution tailored to the target distribution. Note that the new states are generated in parallel, so we use computing in parallel implemented on the multiple cores of CPU. The MH algorithm is based on two types of updates. The first type of update is generating new states for all nodes except node k. The second type of update is generating a new value of k. These two updates alternate in the algorithm. Experimental tests are carried out based on the training images also used in paper I. The simulation results involving the burn-in period and mixing for both training images are promising, and the result for one training image is clearly better than the corresponding result in paper I.

In paper III, "A Bayesian model for lithology/fluid class prediction using a Markov mesh prior fitted from a training image", we construct a Bayesian

model for the inversion of observed seismic data to lithology/fluid classes, and study how the choice of prior distribution for the lithology/fluid classes affects the inversion results. We compare two different prior choices. One is a Markov random field prior with a first order neighborhood while the other one is the Markov mesh prior, as discussed in paper I and paper II, with a much larger neighborhood estimated from a training image. We use the same linearized Gaussian distribution for the likelihood in both cases and estimate the resulting posteriors by the Metropolis–Hastings algorithm. The advantage of adopting the Markov mesh prior on the marginal posterior probabilities for the lithology/fluid classes is modest, but observable. The effect of the prior on the connectivity properties in the posterior realizations is however much more larger. The larger neighborhood of the Markov mesh prior enables us to identify connectivity better than the first-order neighborhood Markov random field prior.

References

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

Arnesen, P. and Tjelmeland, H. (2017). "Prior specification of neighbourhood and interaction structure in binary Markov random fields." *Statistics and Computing*, 27, 737–756.

Caers, J. (2001). "Geostatistical reservoir modelling using statistical pattern recognition." *Journal of Petroleum Science and Engineering*, 29, 177 – 188.

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

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. 2nd ed. New York: John Wiley.

Geyer, C. J. (1991). "Markov chain Monte Carlo maximum likelihood." In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, ed. E. M. Keramidas, 156–163. Interface Foundation.

Guardiano, F. B. and Srivastava, R. M. (1993). "Multivariate Geostatistics: Beyond Bivariate Moments." In *Geostatistics Tróia '92: Volume 1*, ed. A. Soares. Netherlands.

Hammersley, J. M. and Clifford, P. (1971). "Markov fields on finite graphs and lattices." Unpublished manuscript.

Hurn, M., Husby, O., and Rue, H. (2003). "A tutorial on image analysis." In *Spatial Statistics and Computational Methods*, ed. J. Møller, Lecture Notes in Statistics, 87–139. Springer.

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

Kindermann, R. and Snell, J. L. (1980). *Markov random fields and their applications*. Providence: American Mathematical Society.

Lyne, A.-M., Girolami, M., Atchadé, Y., Strathmann, H., and Simpson, D. (2015). "On Russian roulette estimates for Bayesian inference with doubly-intractable likelihoods." *Statistical Science*, 30, 443–467.

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.

Murray, I., Ghahramani, Z., and MacKay, D. (2006). "MCMC for doubly-intractable distributions." In *Proceedings of the Twenty-Second Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-06)*, 359–366. AUAI Press.

Ortiz, J. M. and Deutsch, C. V. (2004). "Indicator simulation accounting for multiple-point statistics." *Mathematical Geology*, 36, 545–565.

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.

Walker, S. (2011). "Posterior sampling when the normalising constant is unknown." *Communications in Statistics — Simulation and Computation*, 40, 784–792.

Paper I

Prior specification for binary Markov mesh models

Xin Luo and Håkon Tjelmeland

Accepted for publication in *Statistics and Computing*
doi: 10.1007/s11222-018-9813-7

Paper II

A multiple-try Metropolis–Hastings algorithm with tailored proposals

Xin Luo and Håkon Tjelmeland

Submitted for publication (under revision)

Paper III

A Bayesian model for lithology/fluid class prediction using a Markov mesh prior fitted from a training image

Håkon Tjelmeland, Xin Luo and Torstein Fjeldstad

Submitted for publication