On the reconstruction and sampling of random fields based on information from limited-size marginals

Knud Skou Cordua[1], Thomas Mejer Hansen[2], and Klaus Mosegaard[3]

[1]NA
[2]Aarhus University
[3]Københavns Universitet,Klaus Mosegaard

November 24, 2022

Abstract

In an ideal application of sequential simulation, parameters are simulated one at a time conditioned to all previously simulated parameters. This requires that marginal distributions of all dimensions (used to derive the conditional distributions) from the random field can be extracted and used for the simulation. However, in practice, only incomplete information from limited-size marginals is used for sequential simulation due to, e.g., computational unwieldiness or to ensure adequate pattern statistics. In this paper, we start out by addressing the problem of how to reconstruct an unknown random field that is consistent with known limited-size marginals. This problem turns out to be highly underdetermined (i.e., infinitely many solutions exist). Therefore, we describe possible additional constraints to supplement the marginals in order to reconstruct well-defined random fields. Secondly, we investigate which random field (out of infinitely many) that is sampled by sequential simulation algorithms using limited-size marginals. We show that sample distributions of such algorithms may depend on the sampling sequence and, sometimes, are inconsistent with the known marginals. We reviewed a formulation of a Markov random field that provides a well-defined solution to the underdetermined problem. Finally, we investigate the relation between marginal-size and information content of reconstructed random fields.

On the reconstruction and sampling of random fields based on information from limited-size marginals

Knud Skou Cordua[+], Thomas Mejer Hansen[*], Klaus Mosegaard[+]

+ Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, 2100 København Ø, Denmark.

* Department of Geoscience, Aarhus University, Høgh-Guldbergsgade 2-4, 8000 Aarhus C, Denmark.

Corresponding author: Klaus Mosegaard, mosegaard@nbi.ku.dk, phone +4527260794.

Short title: Reconstruction of random fields based on limited-size marginals

Abstract

In an ideal application of sequential simulation, parameters are simulated one at a time conditioned to all previously simulated parameters. This requires that marginal distributions of all dimensions (used to derive the conditional distributions) from the random field can be extracted and used for the simulation. However, in practice, only incomplete information from limited-size marginals is used for sequential simulation due to, e.g., computational unwieldiness or to ensure adequate pattern statistics. In this paper, we start out by addressing the problem of how to reconstruct an unknown random field that is consistent with known limited-size marginals. This problem turns out to be highly underdetermined (i.e., infinitely many solutions exist). Therefore, we describe possible additional constraints to supplement the marginals in order to reconstruct well-defined random fields. Secondly, we investigate which random field (out of infinitely many) that is sampled by sequential simulation algorithms using limited-size marginals. We show that sample distributions of such algorithms may depend on the sampling sequence and, sometimes, are inconsistent with the known marginals. We reviewed a formulation of a Markov random field that provides a well-defined solution to the underdetermined problem. Finally, we investigate the relation between marginal-size and information content of reconstructed random fields.

Keywords: Two- and multiple-point geostatistics, sequential simulation algorithms, Shannon entropy, information theory, prior model, training image.

1 Introduction

In a probabilistic formulation of inverse problems, geostatistical algorithms can express prior information about the Earth which, in combination with a likelihood function measuring the fit to geophysical or production data, allows us to sample the posterior probability distribution (e.g., Hansen et al., 2008; Irving and Singha, 2010; Cordua et al., 2012; Toftaker and Tjelmeland, 2013). Conceptually, prior information provided by geostatistical sampling algorithms can be expressed by a random field, i.e., a joint probability distribution defined over a set of random (model) parameters, which, in a pixel-based formulation, are associated with positions in space (normally arranged in a regular grid) (Toftaker and Tjelmeland, 2013; Cordua et al., 2015).

Sequential simulation is a commonly applied technique in geostatistical sampling algorithms (Journel and Alabert, 1989). This algorithm simulates one model parameter at a time conditioned by all previously simulated model parameters. In a strictly theoretical sense, all marginal probability distributions (of all possible sizes/dimensions and geometries), from the random field to be simulated, have to be known in order to derive all the necessary conditional distributions needed for the simulation.

In practice, however, the only information about the prior random field that is used by a sequential simulation algorithm is a limited set of conditional probability distributions (derived from limited-size marginals distributions), namely those expressing the conditional dependence between the individual pixel values given a set of pixel values (typically) within some local neighborhood (Gómez-Hernández and Journel, 1993; Strebelle, 2002). Such information, in form of conditional dependencies described by limited-size marginals, is obtained from e.g. training images or 'old data sets' that are believed to represent typical Earth structure, and, at the same time, are assumed to be realizations from the ‘underlying’ unknown prior random field that describes the subsurface. We will denote such realizations a sample model.

From a sample model, an (M-dimensional) pattern histogram can be obtained by scanning the sample model with an M-dimensional template. By assuming that the unknown random field is stationary, the (M-point) statistics in the pattern histogram will be an approximation of the marginal distributions over the M model parameters associated with the geometry of the chosen template. In the case of random fields based on covariances (i.e., two-point statistics), one is never confronted with the (two-dimensional) pattern histograms themselves, but only a semivariogram from where a covariance matrix can be constructed. I should be noted, however, that a semivariogram basically describes the relation between (co)variances from different two-dimensional pattern histograms as a function of separation distance of the (two-point) template.

The single normal equation simulation (SNESIM) algorithm is an example of a geostatistical sequential simulation algorithm that samples an unknown random field using a limited-size marginal distribution obtained from a sample model (i.e., a training image) (Strebelle, 2002). This algorithm uses limited-size marginal distributions in order to ensure sufficient pattern statistics from the training image (Guardiano and Srivastava, 1993; Strebelle, 2002). Another example is to assume the unknown random field to be Gaussian and then determine a set of two-dimensional marginal distributions through a semivariogram/covariance analysis from a sample model (typically in form of scattered observations) (e.g., Journel and Huijbregts, 1978). In this way, the unknown random field can typically be completely determined and sampled through a sequential Gaussian simulation (SGSIM) algorithm (e.g., Journel and Alabert, 1989). However, in practice the SGSIM algorithm uses a set of Gaussian conditional distributions, derived from limited-size marginals related to a local neighborhood, in order to reduce the computational workload (e.g., Gómez-Hernándes and Journel, 1993; Journel, 1994).

In this paper, the fundamental problem of how to reconstruct a random field that is consistent with a set of limited-size marginal distributions will be addressed. As we shall see, this mathematical problem turns out to be underdetermined and, hence, infinitely many possible solutions exist. Possible additional assumptions (i.e., constraints) about the unknown random field, necessary to obtain a well-defined random field, will be described. Secondly, we take a closer look at which random field (out of infinitely many possible) that is actually sampled by practical implementations of sequential simulation algorithms using limited-size marginals. In relations to this, we investigate the implicit assumptions used by such algorithms.

Fig. 1 provides a schematic overview of the outline of this paper. This figure illustrates the steps involved from 1) having a known/observed sample model from an 'underlying' unknown random field, 2) extracting pattern statistics from the sample model based on a choice of template configuration (one-, two- or multiple-point statistics of arbitrary geometry) and 3) how this statistics is used to estimate a limited-size marginal distribution from the 'underlying' unknown random field, utilizing the assumption of stationarity. As already mentioned, we shall see 4) that the problem of determining an unknown random field, that is consistent with known marginal distributions, is underdetermined. Consequently, some constraints, in addition to the known marginals, have to be used to produce solution to this problem. 5) It will be investigated which constraints that are (implicitly) used by sequential simulation algorithms (using limited-size marginals) to defeat the under-determination. Moreover, possible inconsistency of random fields sampled by such algorithms (i.e., their sampling distribution) with respect to the known marginals will be investigated. The canonical formulation of a Markov random field provides a marginal-consistent solution of the underdetermined problem. We describe how sequential simulation from this random field results in a sample distribution consistent with the known marginals.

Finally, we discuss and quantify the information content (which is inversely proportional to the entropy (Shannon, 1948)) of random fields based on limited-size marginals and how marginal-size influences the information content.

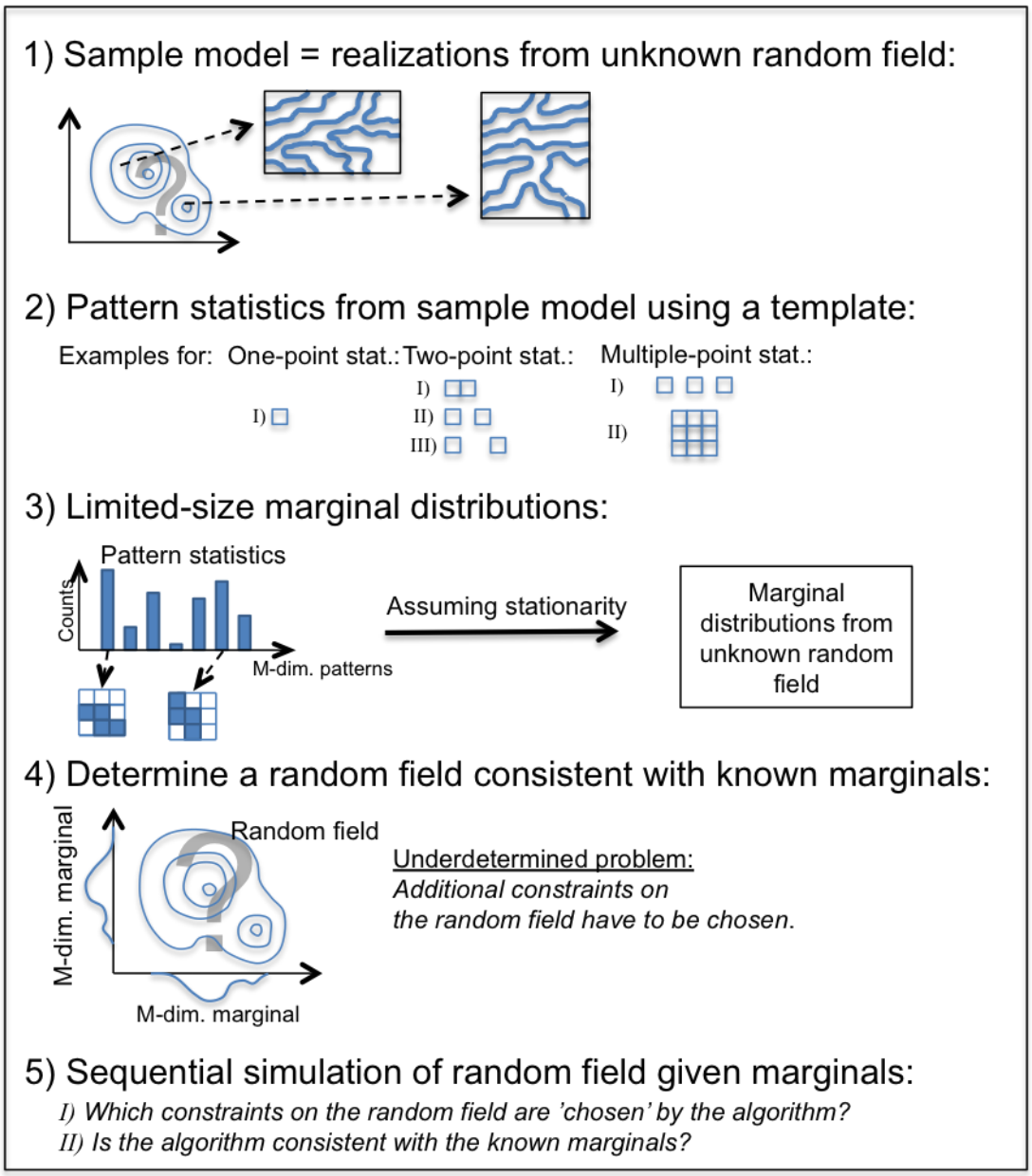

Figure 1. Outline of the fundamental steps of going from 1) one or a few realizations (i.e., sample model) from an 'underlying' unknown random field to 2) chose a template that can be used to obtain one-, two-, or multiple-point-based patterns statistics from the sample model. 3) Under assumption of stationarity, the pattern statistics provides an approximation of marginal distributions from the 'underlying' unknown random field. 4) We show (in appendix A) that the problem of determining a random field that is consistent with the known marginals is underdetermined, and we discuss the additional constraints typically used to uniquely determine a random field. 5) We investigate the actual sampling distributions of sequential simulation algorithms when only limited-size marginal distributions are used/available.

2 Random fields and conditional dependencies

In this study, we consider a set of variables $\mathbf{m} = (m_1, m_2, ..., m_N)^T$ that are associated with nodes in a regular grid. Unless mentioned, these variables are considered continuous. We will denote these variables as model parameters. A random field f (potentially un-normalized) defined over the model parameters is denoted

$$f(\mathbf{m}) = f(m_1, m_2, ..., m_N) \tag{1}$$

An associated marginal probability distribution p (of some dimension lower than f) over some subset $\mathbf{m}^{(k)}$ of the model parameters $\mathbf{m}$ is defined as

$$p\left(\mathbf{m}^{(k)}\right) = \int f(\mathbf{m}) d\mathbf{m}^{(\varnothing k)}, \tag{2}$$

where the integral is taken over the variables $\mathbf{m}^{(\varnothing k)}$ that are not in $\mathbf{m}^{(k)}$.

The relation between the random field and the marginal probability distribution is given by the product rule

$$f(\mathbf{m}) = p\left(\mathbf{m}^{(\neg k)} \mid \mathbf{m}^{(k)}\right) p\left(\mathbf{m}^{(k)}\right), \tag{3}$$

where $p\left(\mathbf{m}^{(\neg k)} \mid \mathbf{m}^{(k)}\right)$ is the conditional probability distribution over $\mathbf{m}^{(\neg k)}$ given the data event $\mathbf{m}^{(k)}$.

If a random field is completely known and all possible marginal distributions from the random field can be obtained, then the field can be rewritten into a product over marginal and conditional probability distributions:

$$\begin{aligned} f(\mathbf{m}) &= p(m_N \mid m_{N-1}, ..., m_1) p(m_{N-1}, ..., m_1) \\ &= p(m_N \mid m_{N-1}, ..., m_1) p(m_{N-1} \mid m_{N-2}, ..., m_1) p(m_{N-2}, ..., m_1) \\ &\quad \vdots \\ &= p(m_2 \mid m_1) p(m_1) \prod_{n=3}^{N} p(m_n \mid m_{n-1}, ..., m_1) \end{aligned} \tag{4}$$

The product rule in eq. 4 is the backbone of sequential simulation because it allows sequential simulating of one model parameter at a time when this parameter can be conditioned to all previously simulated model parameters (see e.g., Gómez-Hernández and Cassiraga (2000) and Hansen et al. (2012)). In this general formulation, a permutation of the model parameters in the conditional probability distributions does not influence the formulation of the random field.

However, in practical implementations of sequential simulation, where only limited-size marginal distributions are used, it is not possible to condition to all previously simulated parameters.

2.1 Markov random fields

A probabilistic graph can be used to express the dependency structure of probability distributions, such as a random field. The nodes in such a graph represent the individual model parameters and edges between two nodes (i.e., model parameters) denote a conditional dependence between these parameters. Two model parameters will be denoted neighbors if they are connected with an edge. The set of all model parameters $\mathbf{m}_{\partial i}$ that are neighbors with a model parameter m_i is denoted the neighborhood of m_i.

Assuming a *Markov property* of the random field in Eq. 1 involves that the individual model parameters have a specified neighborhood (Besag, 1974). For a model parameter m_i associated with the *i*'th node in a grid, a Markov property involves that the conditional probability distribution for this model parameter given all other model parameters satisfies

$$p(m_i \mid m_N, \ldots, m_{i+1}, m_{i-1}, \ldots, m_1) = p(m_i \mid \mathbf{m}_{\partial i}), \tag{5}$$

where $\mathbf{m}_{\partial i}$ is a subset of the model parameters that are in the neighborhood of m_i. When this (Eq. 5) applies to a random field (Eq. 1), the random field is a Markov random field (Cressie and Davidson, 1998).

An example of a random field that satisfies this is

$$f(\mathbf{m}) \propto \prod_{i=1}^{N} p(m_i \mid \mathbf{m}_{\P i}) = \prod_{i=1}^{N} \frac{p(m_i, \mathbf{m}_{\P i})}{p(\mathbf{m}_{\P i})}, \tag{6}$$

where $\mathbf{m}_{\P i}$ is the neighborhood of the *i*'th node.

A clique C_i in a probabilistic graph is a subset of nodes where every pair of nodes is connected, i.e., they are all conditionally dependent on each other (Kindermann and Snell, 1980). A maximum clique (also denoted a maximum clique neighborhood) is a clique that is not a strict subset of a larger clique. A Markov random field can now be defined over a chain of maximum clique neighborhoods C_i. In that case, the random field is defined as (Castillo et al., 1997)

$$f(\mathbf{m}) = p(\mathbf{m}_{R1}) \prod_{i=2}^{L} p(\mathbf{m}_{Ri} \mid \mathbf{m}_{Si}) = p(\mathbf{m}_{R1}) \prod_{i=2}^{L} \frac{p(\mathbf{m}_{Ri}, \mathbf{m}_{Si})}{p(\mathbf{m}_{Si})}, \quad (7)$$

where L is the number of clique neighborhoods and $\mathbf{m}_{Ci} = (\mathbf{m}_{Ri}, \mathbf{m}_{Si})$ is the model parameters associated the i'th clique of the graph. $\mathbf{m}_{Ri}$ and $\mathbf{m}_{Si}$ are model parameters associated with the i'th residual and separator in the chain of cliques that spans the graph. The probability distribution $p(\mathbf{m}_{Ri}, \mathbf{m}_{Si}) = p(\mathbf{m}_{Ci})$ is a marginal probability distribution over the model parameters related to the i'th clique neighborhood in the random field (i.e., Markov random field). More details about Eq. 7 and definitions of cliques, residuals and separators are found in section 4.2.

According to the product rule in Eq. 3, it is seen that both of the Markov random fields defined in Eqs. 6 and 7 above are composed of local marginal probability distributions, namely $p(m_i, \mathbf{m}_{\partial i})$ and $p(\mathbf{m}_{Ri}, \mathbf{m}_{Si})$, respectively. Hence, an algorithm based on sequential simulation that only uses limited-size marginal distributions will sample from some type of Markov random field. More details on this topic will be outlined in section 4.

3 Reconstruction of random fields from limited-size marginals

As previously described, information provided about an Earth model typically exists in form of marginal distributions (sometimes converted into conditional probability distributions) from some 'underlying' unknown random field. According to theorem 1 demonstrated in appendix A1, there are infinitely many random fields over the model parameter space that are consistent with the same known marginal probability distributions. Consequently, additional constraints to the problem have to be applied to deliver a unique random field consistent with the marginals. The conditional dependencies found in a random field, that is a solution to this problem, can only be expected to describe spatial dependencies that are captured by the known limited-size marginals. Hence, unless the additional constrains happens to provide information about spatial dependencies outside of the known marginals, a solution to this underdetermined problem cannot be expected to be the 'underlying' unknown random field itself. Only consistency with respect to the known marginals can be expected to be a common property of the calculated and 'underlying' random fields.

There are several (not necessarily mutually exclusive) ways of reducing the under-determination of this problem such that a unique random field, consistent with known marginals, can be determined. In what follows, we will list some of them.

3.1 Assuming the random field to be parametric

If the random field is assumed to be parametric of some form, the degrees of freedom are reduced to the number of parameters in the probability distribution. For instance, by assuming the random field to be Gaussian, the marginal distributions will also be Gaussian. In this case, the degrees of freedom reduce to $\frac{N^2+3N}{2}$ (N means plus N variances plus $\frac{N(N-1)}{2}$ covariances). These parameters are in practice determined through a semi-variogram analysis and a mean value of a sample model (e.g., Journel and Huijbregts, 1978). The mean and covariance parameters uniquely determine all two- (or larger) dimensional (Gaussian) marginal probability distributions from the Gaussian random field, which, in turn, uniquely determine the full Gaussian random field.

3.2 Assuming the random field to have Markov properties

In general, a (non-parametric) categorical random field has L^{N-1} degrees of freedom, where L is the number of categories and N is the dimension of the random field. Assuming that the random field is a Markov random field implies a reduction in the degrees of freedom. A Markov random field is completely defined by a set of marginal probability distributions as defined in Eqs. 6 and 7. Hence, in case of categorical variables, the degrees of freedom reduce to NL^{M-1}, where M is the dimension of the marginal distributions, L^{M-1} is the degrees of freedom of the individual marginals, and the number of marginals is assumed to be (approximately) equal to the dimension of the random field N. Thus, if all the N marginals are completely known, the number of degrees of freedom is reduced to zero.

3.3 Assuming the random field to be stationary

That a random field is stationary means that the size of its neighborhoods and the marginals over these neighborhoods do not change when shifted in space. Moreover, the marginal distributions defined over these neighborhoods are all equal.

Assuming both stationary and a Markov property means that only a single marginal probability distribution has to be known in order to uniquely determine the random field. I.e., the degrees of freedom reduce to the degrees of freedom associated with a single marginal (L^{M-1} degrees of freedom for the case of a (non-parametric) marginal of categorical variables).

If the random field is assumed to be both Gaussian and stationary, the assumption of stationarity further reduces the degrees of freedom in the Gaussian random field because, in this case, only one mean, one variance, and *N*-1 covariances are need to completely define the random field. Hence, the total number of degrees of freedom reduces to *N+1*.

3.4 Assuming the random field to have maximum entropy/minimum information

Assuming maximum entropy of a random field is often considered the most "neutral" assumption possible. This assumption involves that the random field satisfying the marginal probability distributions, carries least possible additional information (in form of additional constraints on the random field), i.e., it is the most noncommittal with regard to missing information (Jaynes, 1957).

The entropy H of a variable $\mathbf{m}$ with probability distribution $p(\mathbf{m})$ is denoted as

$$H(\mathbf{m}) = -\int_{\mathbf{m}} p(\mathbf{m}) \log_2 \left(p(\mathbf{m}) \right) d\mathbf{m}, \quad (8)$$

for the continuous case and

$$H(\mathbf{m}) = -\sum_i p(\mathbf{m}_i) \log_2 \left(p(\mathbf{m}_i) \right), \quad (9)$$

for the discrete case (Shannon, 1948). Here, $\log_2$ is the logarithm with base two. When using the base-two logarithm, the entropy is measured in bits.

The maximum entropy assumption involves that the additional constraint used for the random field, is a constraint that leads to a random field that maximizes Eq. 8 or 9. E.g., given a set of all two-dimensional marginals from the 'underlying' unknown random field, the Gaussian assumption is the constraint that leads to a random field with maximum entropy (i.e., the least informative random field given the known marginals). If no marginals are known, a uniform distribution is the one that carries least information and, therefore, has maximum entropy.

4. Sequential simulation of random fields using limited-size marginals

The SGSIM algorithm, the Direct Sequential SIMulation (DSSIM) algorithm, the SNESIM algorithm, and the Direct Sampling method (DS) are all examples of a geostatistical sampling algorithms based on implementations of sequential simulation that use limited-size marginal probability distribution to simulate realizations of Earth models (Deutsch and Journel, 1998; Strebelle, 2002; Mariethoz, 2010). Limited-size marginals are typically used due to computational unwieldiness (e.g., in SGSIM) (Journel and Alabert, 1989; Gómez-Hernández and Journel, 1993)) or due to a limited-size template, which is needed to obtain sufficient pattern statistics (e.g., in SNESIM) (Strebelle, 2002)). In the following, we will take a closer look at which random field, out of infinitely many possible, that is sampled by such algorithms.

4.1 Influence of the simulation sequence

As a sequential simulation algorithm proceeds, the model parameters are simulated one at a time from a local conditional probability distribution (based on the known limited-size marginal distribution, see Eq. 6). As the simulation evolves, the algorithm forms a directed graph, where each simulated model parameter is associated with a node in the graph and the directed edges between the nodes represent conditional dependencies between the model parameters. Since each node is only visited once during the sequential simulation, no cycle will ever occur in such a graph. Hence, the graph is said to be a directed acyclic graph (DAG) (see e.g. Bishop (2006) for an introduction to graphical models).

As an example, figure 2(I) shows a set of nodes marked by letters A – I to be sequentially simulated. Figure 2(II) shows a DAG that occurs as a result of sequential simulation with the simulation sequence marked by numbers on the nodes. The conditional dependencies defined by the directed edges result from a square-shaped marginal (related to a choice of neighborhood/template) of 3 by 3 inter-connected nodes. Note that any size and geometry of marginal can be used; the small marginal of 3 by 3 inter-connected nodes used here is only chosen for illustrative purpose.

The random field over a DAG can be factorized as follows (e.g. Whittaker, 1990)

$$f(m_1,\ldots,m_N)=\prod_{i=1}^{N} p(m_i \mid pa(m_i))=\prod_{i=1}^{N}\frac{p(m_i,\, pa(m_i))}{p(pa(m_i))}\,. \qquad (10)$$

$pa(m_i)$ is the set of so-called parents of the model parameter m_i. This expression is a description of a partially ordered Markov model (Cressie and Davidson, 1998; Toftaker and Tjelmeland, 2013). If the simulation sequence used to simulate from this distribution is performed using a unilateral path (i.e., raster scan), this probability distribution is known as a Markov mesh model (Kjønsberg et al., 2012; Daly, 2004).

As an example, the model parameters associated with the nodes D, E, G, and I in Fig 2(II) are parents of the model parameter associated with the node H. From Eq. 10 we now have a general formulation of the random field that results from a sequential simulation algorithm based on limited-size marginals (such as e.g. SGSIM and SNESIM) for a specific simulation sequence and with certain size and geometry of the known marginal related to the individual model parameters. Notice, that in practice, some sequential simulation algorithms (e.g., SNESIM) may reduce the size of the set of parent parameters (i.e., reduce the marginal-size) during the simulation in order to avoid data event with zero probability.

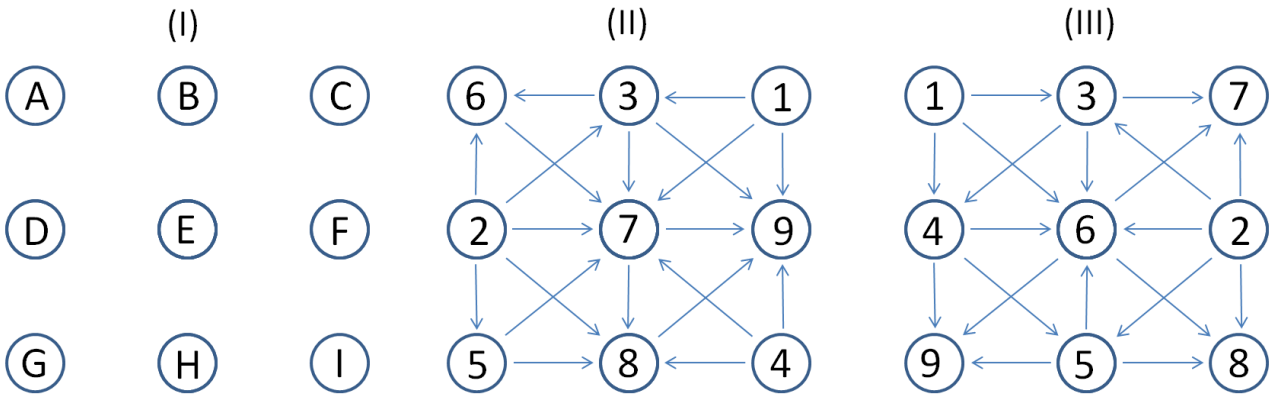

Figure 2. (I) Graph with 9 nodes labeled with letters A – I. (II) First example of sequential simulation sequence leading to a directed acyclic graph. The numbers are associated with the order by which the nodes are sequentially simulated. The arrows show the conditional dependencies between the nodes as a result of a marginal of 3 by 3 inter-connected nodes with the "central node" located in the center of the marginal. Hence, using this marginal, the individual nodes can only be dependent on other nodes that are located next to them. (III) Second example of a sequential simulation path (as a result of the same marginal).

The random field in Eq. 10 that is sampled using the simulation sequences (seq. 1) as seen in Fig 2(II) is given as:

$$f_{seq1}(A,B,\ldots,I) = p(C)p(D)p(B\,|\,C,D)p(I)p(G\,|\,D)p(A\,|\,B,D)\times \\ p(E\,|\,A,B,C,D,G,I)p(H\,|\,D,E,G,I)p(F\,|\,B,C,E,H,I) \quad (11)$$

Using the product rule, this expression can be rewritten as:

$$f_{seq1}(A,B,...,I) = \frac{p(C)p(B,C,D)p(I)p(G,D)p(A,B,D)p(E,A,B,C,D,G,I)}{p(C,D)p(B,D)p(A,B,C,D,G,I)p(D,E,G,I)p(B,C,E,H,I)} \cdot \\ p(H,D,E,G,I)p(F,B,C,E,H,I) \quad (12)$$

Another choice of simulation sequence (seq. 2) is shown in Fig 2(III). The associated random field over this graph is given as:

$$f_{seq2}(A,B,...,I) = p(A)p(F)p(B|A,F)p(D|A,B)p(H|D,F)p(E|A,B,D,F,H) \cdot \\ p(C|B,E,F)p(I|E,F,H)p(G|D,E,H) \\ = \frac{p(A)p(F)p(B,A,F)p(D,A,B)p(H,D,F)p(E,A,B,D,F,H)}{p(A,F)p(A,B)p(D,F)p(A,B,D,F,H)p(B,E,F)p(D,E,H)} \\ p(C,B,E,F)p(I,E,F,H)p(G,D,E,H) \quad (13)$$

From these examples (expressed in Eqs. 12 and 13) it is seen that two different simulation sequences will, in general, lead to two different formulations of the random field because these fields are composed of a product of marginals, $p(m_i, pa(m_i))$, with different dimensions at different locations *i*. Hence, these two random fields have different dependency structure. This is not the case if conditional dependency between all model parameters, and not only a dependency structure provided by limited-size marginals, was used. In that case, the random field is formulated by Eq. 4 and has the same dependency structure independent of the simulation sequence used. In sections 5.1 (using categorical variables) and 6.1 (using continuous variables) we find, through numerical examples, that partially ordered Markov models (i.e., the random field in Eq. 10) for different simulation sequences (i.e., random paths used) has different probability density values when evaluated in the same point in the model parameters space.

The examples show that, when only limited-size marginal distributions are used in sequential simulation, the actual random field sampled by the algorithm depends on (1): the size (i.e., dimension) of the used marginal distributions (i.e., the size of template/neighborhood T used to form the individual conditional distributions) and (2): the simulation sequence. Moreover, since the dimension of the used marginals is changing for different locations across the field and is generally smaller than the known marginal, the random field is not consistent with the known marginal distribution. Moreover, the random field is no longer stationary because the conditional dependency structure does change when shifted in space. The effect of non-stationarity in partially ordered Markov models is also described by Tjelmeland and Toftaker (2012).

The above findings apply to algorithms such as SGSIM, Direct Sequential SIMulation (DSSIM), Sequential Indicator SIMulations (SISIM), SNESIM, direct simulation method, and Extended Normal Equations SIMulation (ENESIM), which are all based on implementations of sequential simulations using limited-size marginals. It should, however, be noted that the above-mentioned effects on the sampling distribution might in some cases be small, and therefore negligible. This is important for practical applications, but it is often unclear how to quantify these effects and how to decide if they are important or not.

4.2 Random fields consistent with marginals defined over chains of cliques

In this section we review an alternative way of formulating a Markov random field that can be sampled through sequential simulation while being consistent with the known marginal distributions at all locations across the random field.

Marginal probability distributions obtained from a sample model have a dependency structure defined by the maximum cliques in an undirected graph (i.e. forms a maximum clique neighborhood). In other words, the marginals based on pattern statistics obtained from a sample model contains information about conditional dependencies between all model parameters within the marginal. Based on marginal distributions with a maximum cliques dependency structure, a Markov random field defined over a chain of maximum cliques based on these marginals can be defined.

For a chain of cliques we have that: (1) every clique C_i occurs exactly once and (2) $C_i \cap (C_1 \cup \ldots \cup C_{i-1})$ is a subset of at least one of the previous cliques $(C_1 \cup \ldots \cup C_{i-1})$. The set $S_i = C_i \cap (C_1 \cup \ldots \cup C_{i-1})$ is said to be the *i*'th separator. Hence, S_i is the overlap between clique C_i and the previous cliques in the chain. The set $R_i = C_i \setminus S_i$ is said to be the *i*'th residual.

The random field defined over a chain of cliques is a Markov random field that can be expressed as (see e.g. Castillo et al., 1997):

$$f(m_1,\ldots,m_N) = p(\mathbf{m}_{R1}) \prod_{i=2}^{L} p(\mathbf{m}_{Ri} \mid \mathbf{m}_{Si}) = p(\mathbf{m}_{R1}) \prod_{i=2}^{L} \frac{p(\mathbf{m}_{Ri}, \mathbf{m}_{Si})}{p(\mathbf{m}_{Si})} = p(\mathbf{m}_{R1}) \prod_{i=2}^{L} \frac{p(\mathbf{m}_{Ci})}{p(\mathbf{m}_{Si})}, \quad (15)$$

where $\mathbf{m}_{Ri}$ and $\mathbf{m}_{Si}$ are the model parameters associated with the nodes of the *i*'th set of residuals R_i and separators S_i, respectively. *L* is the total number of cliques in the chain of

cliques. $\mathbf{m}_{Ci}$ are model parameters associated with nodes contained in the i'th clique C_i, which equals the set of model parameters given by the set $(\mathbf{m}_{Si}, \mathbf{m}_{Ri})$. The distribution in Eq. 15 will be referred to as the canonical Markov model.

The graph depicted in Fig 3(I) is an example of a chain of cliques given by the subsets $(ABDE, BCEF, DEGH, EFHI)$. The associated separators and residuals are given by the subsets $(\varnothing, BE, DE, EFH)$ and $(ABDE, CF, GH, I)$, respectively. The resulting random field obtained by simulating along this chain of cliques is given as:

$$p(A,B,...,I) = \frac{p(A,B,D,E)\,p(B,C,E,F)\,p(D,E,G,H)\,p(E,F,H,I)}{p(B,E)\,p(D,E)\,p(E,F,H)}, \qquad (16)$$

where the simulation sequence is seen in Fig 3(II). Note that the exemplified chain of cliques is not the only one possible.

The canonical Markov model in Eq. 15 is a random field that is consistent with the known marginal distributions. Therefore, it is a possible solution (to the underdetermined problem) of a random field that is consistent with the known marginals. The Markovian dependency structure, defined by a chain of maximum clique neighborhoods, constitutes the additional constraint that leads to this well-defined solution.

It should be noted that, Eq. 15 will only be consistent with the known marginal distribution if the detailed balance condition

$$p(\mathbf{m}_{Si})\,p(\mathbf{m}_{Ri} \mid \mathbf{m}_{Si}) = p(\mathbf{m}_{Ri})\,p(\mathbf{m}_{Si} \mid \mathbf{m}_{Ri}) \qquad (17)$$

is satisfied for all sets of cliques $\mathbf{m}_{Ci} = (\mathbf{m}_{Ri}, \mathbf{m}_{Si})$, which is guaranteed due to the assumption of stationarity.

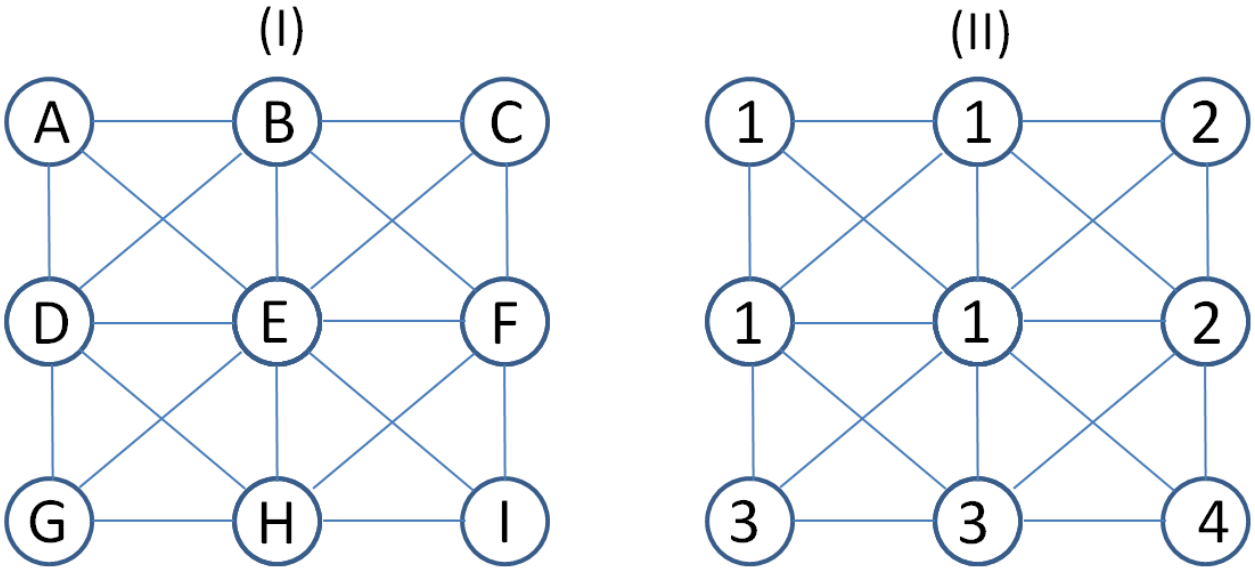

Figure 3. (I) An example of an undirected graph and the conditional dependencies among the nodes, which are labeled A, B,…, I. (II) A possible sequential simulation sequence along a chain of cliques, when sampling from the canonical Markov model, is indicated by numbers.

4.3 Sampling of the canonical Markov model

The formulation of the canonical Markov model in Eq. 15 provides a means of sequentially simulating one residual at a time conditional to the associated separator. The resulting sample distribution from sequential simulation of this random field is given by Eq. 15 and is: 1) consistent with the known marginals (i.e. satisfies Eq. 2A) and 2) is invariant with respect to the simulation sequence, as long as the simulation sequence follows a chain of cliques. Neither was the case for the partially ordered Markov model defined in Eq. 10.

During sequential simulation, lack of detailed balance not only means a lack of consistency with the known marginal distribution, it also means that simulation along a chain of cliques will not satisfy the same marginal distribution as if the simulation is performed in the reverse direction along the chain.

As seen in Fig 3(II), sequential simulation of the canonical Markov model may even allow sequential simulation of more than one model parameter at a time (i.e., each residual may constitute more than a single model parameter), which may reduce the computational cost when sampling this random field (Faucher et al., 2013; Razaee et al., 2013). The canonical Markov model in Eq. 15 can easily be reformulated such that (a correct, sequence independent) sequential simulation of one model parameter at a time (based on Eq. 4) can be obtained using the product

rule within the individual residuals: For the *i*'th residual, the conditional probability can be rewritten as:

$$p(\mathbf{m}_{Ri} \mid \mathbf{m}_{Si}) = \frac{p(\mathbf{m}_{Ri}, \mathbf{m}_{Si})}{p(\mathbf{m}_{Si})} = \frac{\prod_{j=3}^{J} p(m_{Ri,j} \mid m_{Ri,j-1}, \ldots, m_{Ri,1}, \mathbf{m}_{Si}) \ldots p(m_{Ri,2} \mid m_{Ri,1}, \mathbf{m}_{Si}) p(m_{Ri,1} \mid \mathbf{m}_{Si})}{p(\mathbf{m}_{Si})} \quad (18)$$

where $m_{Ri,j}$ is the *j*'th model parameter in the *i*'th residual and J is the number of nodes within the *i*'th set of residuals.

4.4 Relation between sample model and sample distribution

Consider a 16-dimensional marginal probability distribution $p(\mathbf{m}_{Ci})$ based on statistics obtained from the sample model seen in Fig. 4 using an inter-connected template of 4 by 4 pixels. Realizations obtained by sequentially simulating the canonical Markov model based on this marginal probability distribution are seen in Fig. 5. These realizations are consistent with (the information contained in) the known marginals $p(\mathbf{m}_{Ci})$.

Information about conditional dependencies between model parameters that are not inside the known limited-size marginals is not provided. Therefore, the realizations seen in Fig. 5 cannot are be expected to exhibit the same long-range spatial structures (longer that 4 by 4 pixels) as seen in the sample model in Fig. 4. Both the 'underlying' unknown random field, from where the sample model is drawn, and the canonical Markov model, are just two possible random fields (out of infinitely many) that are consistent with the known marginals. Hence, it can only be expected that consistency between the sample model and the sample distribution from the canonical Markov model (Figs. 4 and 5) exist in form of pattern statistics obtained with the same template as used to obtain the known marginals.

It is, in this context, important to note the difference between conditional dependency and spatial correlation. Spatial correlation found in realizations from Markov processes, as e.g. described by the canonical Markov model, may have spatial correlations that extend mush further than the spatial dependencies found in realizations from a Markov random field. This can also be acknowledged in Fig. 5, where the extension of the channel structures are much longer than the marginal distribution of 4 by 4 pixels used to construct the random field.

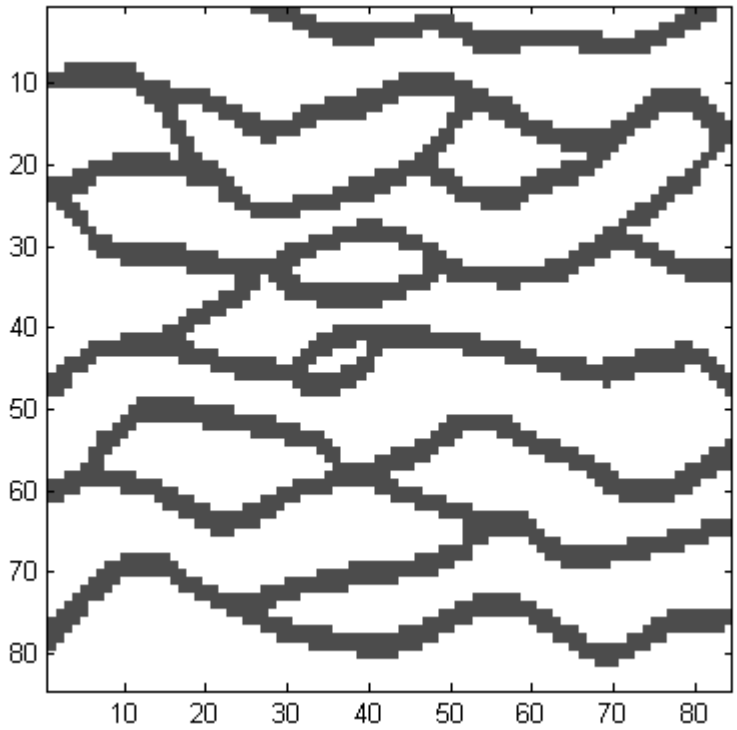

Figure 4. Sample model from where the marginal distribution is estimated based on a 4 by 4 pixel template.

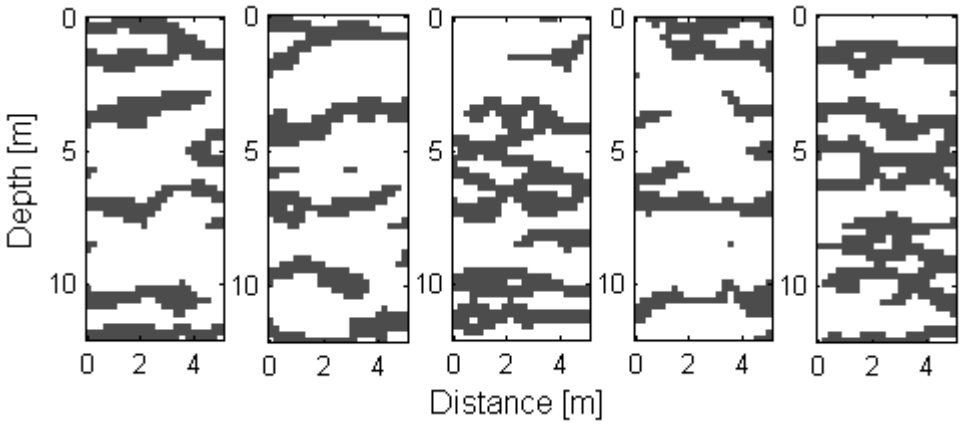

Figure 5. Five outcome realizations obtained from the canonical Markov model in Eq. 15 based on a marginal distribution obtained from the sample model seen in Fig 4 using a 4 by 4 pixel template. Each cell is 0.2 by 0.2 m.

5 Influence of marginal-sizes on information content of random fields

In this section, we discuss the relation between conditional dependence defined by limited-size marginals and Shannon entropy (i.e., information content). Moreover, we calculate and compare

the entropy of two different formulations of a random field, namely partially ordered Markov models based on different simulation sequences and the canonical Markov model.

The entropy of a random field that is decomposed into conditional probability distributions assuming conditional dependence between all model parameters (as expressed by Eq. 4) is given as (Cover and Thomas, 2006)

$$H(m_N,...,m_1) = H(m_1) + H(m_2 \mid m_1) + \sum_{k=3}^{N} H(m_k \mid m_{k-1},...,m_1) \quad . \tag{19}$$

It can be shown (Cover and Thomas, 2006) that conditioning leads to a decrease in entropy (i.e., increase of information):

$$H(m_1 \mid m_2) \leq H(m_1) \quad , \tag{20}$$

with equality if m_1 and m_2 are conditionally independent. Now, consider the case where conditional dependency does not exist between all model parameters (because only limited-size marginals are used to derive the conditional distributions), but only between a subset of the model parameters (for instance when $p(m_5 \mid m_1, m_2, m_3, m_4)$ is required in the sequential simulation using Eq. 4, but only $p(m_5 \mid m_1, m_2)$ is used). In such a case, the entropy of Eq. 4 (as defined by Eq. 19) will increase according to the inequality in Eq. 20 (e.g. $H(m_5 \mid m_1, m_2, m_3, m_4) \leq H(m_5 \mid m_1, m_2)$ under the assumption that m_5 is conditionally dependent on m_3 and m_4). Hence, every time a conditional dependency between model parameters is not used (e.g., as part of a sequential simulation), the entropy of the resulting random field that is actually sampled will increase, and information is lost. This is the case for the partially ordered Markov model, which is the random field sampled by typical sequential simulation algorithms using limited-size marginals, because this random field is not using all conditional dependencies otherwise provided by the known marginals. Additional information is lost, when e.g. SNESIM is pruning the size of the conditioning data event (i.e., the size of the marginal) in order to find a data event with non-zero probability.

A general problem related to marginals obtained from sample models is a trade-off between marginal-size (i.e., the information content described by conditional dependencies in the marginals) and, on the other hand, statistical uncertainty related thinning of the pattern histogram

for increasing marginal-size. A quantification of this type of information loss related to pattern-statistics uncertainty is beyond the scope of this study.

5.1 Quantification of information content for different marginal-sizes

Figure 6A shows a sample model, and Figure 6F shows the non-zero multiple-point probabilities (i.e., marginal distributions) extracted from the sample model using the template seen in Fig 6B. Figures 6D and 6E are two examples of simulation sequences that can be used to obtain the realization of the model parameters (a,b,c,d,e,f,g,h,i) seen in Fig 6C using Eq. 10.

The probability value of partially ordered Markov models (Eq. 10) based on the two different simulation sequences (6E and 6D) when evaluated in the point, as seen in Fig 6C, are

$$f_{seq1}(a=0,b=0,c=0,d=0,e=1,f=1,g=0,h=0,i=1)=\frac{5}{48}\frac{8}{48}\frac{4}{9}\frac{26}{48}\frac{24}{48}$$
$$=20.90\cdot10^{-4} \quad (21)$$

and

$$f_{seq2}(a=0,b=0,c=0,d=0,e=1,f=1,g=0,h=0,i=1)=\frac{18}{48}\frac{18}{48}\frac{21}{24}\frac{6}{48}\frac{3}{5}\frac{4}{24}\frac{26}{48}$$
$$=8.33\cdot10^{-4}, \quad (22)$$

respectively. This numerical example confirms that two different simulation sequences lead to sampling of two different random fields.

The probability value of the same point when evaluated in the canonical Markov model (Eq. 15) along the chain of maximum cliques (a,b,d,e), (b,c,e,f), (d,e,g,h) and (e,f,h,i) is

$$f_{canonical}(a=0,b=0,c=0,d=0,e=1,f=1,g=0,h=0,i=1)=\frac{3}{48}\frac{3}{9}\frac{2}{7}\frac{2}{4}$$
$$=30\times10^{-4} \quad (23)$$

The entropy of the two formulations of the partially ordered Markov models are given as

$$H_{seq1}(a,b,c,d,e,f,g,h,i)=H(e)+H(b)+H(i\,|\,e)+H(a)+H(h\,|\,e)+$$
$$H(f\,|\,b,e)+H(g)+H(c\,|\,b)+H(d\,|\,a)$$
$$=7.58bits \quad (24)$$

and

$$H_{seq2}(a,b,c,d,e,f,g,h,i) = H(g) + H(d) + H(h \mid g,d) + H(c) + H(f \mid c) + H(a) + H(i \mid f,h) + H(e \mid a,d) + H(b \mid a) = 7.31 bits \quad , \tag{25}$$

respectively.

The entropy for the canonical Markov model is in this case given by

$$H_{canonical}(a,b,c,d,e,f,g,h,i) = H(a,b,d,e) + H(c,f \mid b,e) + H(g,h \mid d,e) + H(i \mid e,f,h) = 6.10 bits \tag{26}$$

The entropy for the conditional probabilities is calculated using the expression

$$H(m_N \mid m_1, m_2, ..., m_{N-1}) = \sum_{m_1} \sum_{m_2} ... \sum_{m_N} p(m_1, m_2, ..., m_N) \log_2 \left(\frac{p(m_1, m_2, ..., m_{N-1})}{p(m_1, m_2, ..., m_N)} \right) \tag{27}$$

Sequential simulation sequence no. 2 leads to a partially ordered Markov model that is composed of conditional probability distributions based on marginals, which describes more conditional dependencies than is the case for simulation sequence no. 1. Consequently, the entropy of the partially ordered Markov model related to sequence no. 1 is higher than for sequence no. 2. This means that more information will be propagated from the sample model statistics to the random field being sampled when using simulation sequence no. 2 than sequence no. 1.

The canonical Markov model uses all the information provided by the known marginal distributions, which leads to a random field with lower entropy (i.e., more information) than for both of the partially ordered Markov models.

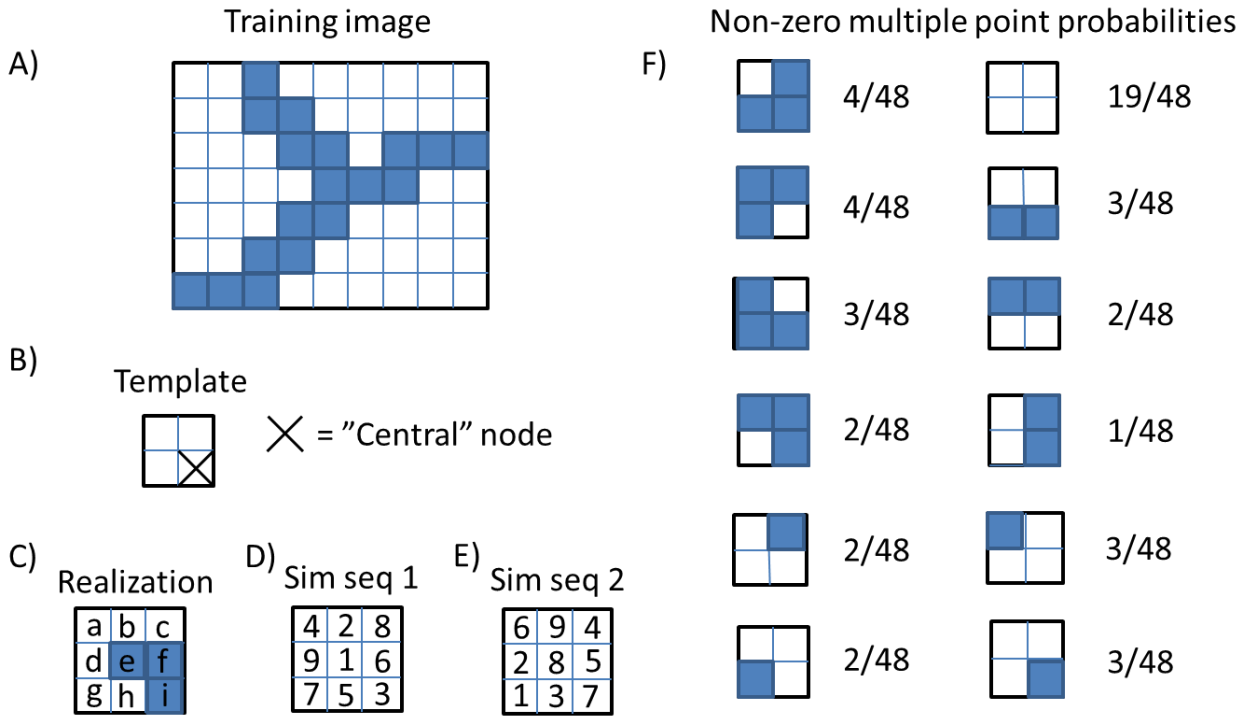

Figure 6. A) Sample model in form of a training image. The white pixels refer to model parameters with value 0 and the blue pixels to the value 1. B) Template used to extract multiple-point statistics. C) One example of a realization. D) Sequential simulation sequence no. 1. E) Sequential simulation sequence no. 2. F) The non-zero multiple-point probabilities (i.e., the estimated marginal) based on sample model statistics using the displayed template.

6 The effect of limited-size marginals in SGSIM

As already discussed, partially ordered Markov models resulting from different simulation sequences may have different probability density values for the same point in the model parameter space. In this section, we will take a look at the probability density values for partially ordered Markov models based on Gaussian marginals. This is the random field sampled by (i.e., sample distribution of) a SGSIM algorithm when the algorithm is using a local neighborhood. The probability density values from such Gaussian partially ordered Markov models evaluated for a certain point in the model parameters space, using different choices of simulation sequence and size of neighborhood, will be compared. Moreover, these results are also compared with the density values of a canonical Markov model based on Gaussian marginals of different sizes.

The ratio between the probability density values of the ‘correct’ Gaussian random field using a full neighborhood and the above mentioned Markov random fields is seen in Fig. 7.

The evaluated points are realizations from the 'correct' Gaussian random field (41 x 41 pixels), which is described by a spherical covariance function with a horizontal range of 20 pixels and vertical range 4 pixels, variance 1, and a mean value of 10.

Different continuous square-shaped marginals (resulting from a choice of neighborhood) with a side length of 9 pixels, 17 pixels, 25 pixels, 33 pixels, and 41 pixels, respectively, have been used to define the different Gaussian partially ordered Markov models (green curves) and the Gaussian canonical Markov model (black curves). For each marginal/neighborhood size, partially ordered Markov models for 10 different simulation sequences have been evaluated (seen as 10 green curves for the individual realizations). The Gaussian marginal distribution used to establish these different Markov models has the same mean, variance, and covariances as the 'correct' Gaussian random field.

A ratio of one in Fig. 7 indicates equality with the 'correct' Gaussian random field. This happens for the canonical Markov model when it uses a neighborhood with a side-length of 41 pixels (i.e., equivalent to known the full Gaussian random field and not only a limited-size marginal). It is seen, that in general the canonical Markov model will converge faster towards a ratio of one than the partially ordered Markov models. Moreover, for a neighborhood size of 41 pixels (double as much as the maximum correlations length of the Spherical covariance function used) the partially ordered Markov model provides probability values that are much different from the 'correct' density value (up to 27 times larger for realization #6 in Fig. 7). This means that the probability of drawing this point in the model parameter space from the 'correct' formulation of the Gaussian random field using a full neighborhood is up to 27 times more probable than if this point had been drawn from this particular formulation of the partially ordered Markov model.

The tendency of higher probability ratios for the partially ordered Markov model (green curves in Fig. 7) as compared to the probability ratios for the canonical Markov model confirms that more information from the 'correct' probability distribution propagates (through the marginals) to the canonical Markov model than to the partially ordered Markov model because the partially ordered Markov model has a sample distribution that is not consistent with the known marginal distributions.

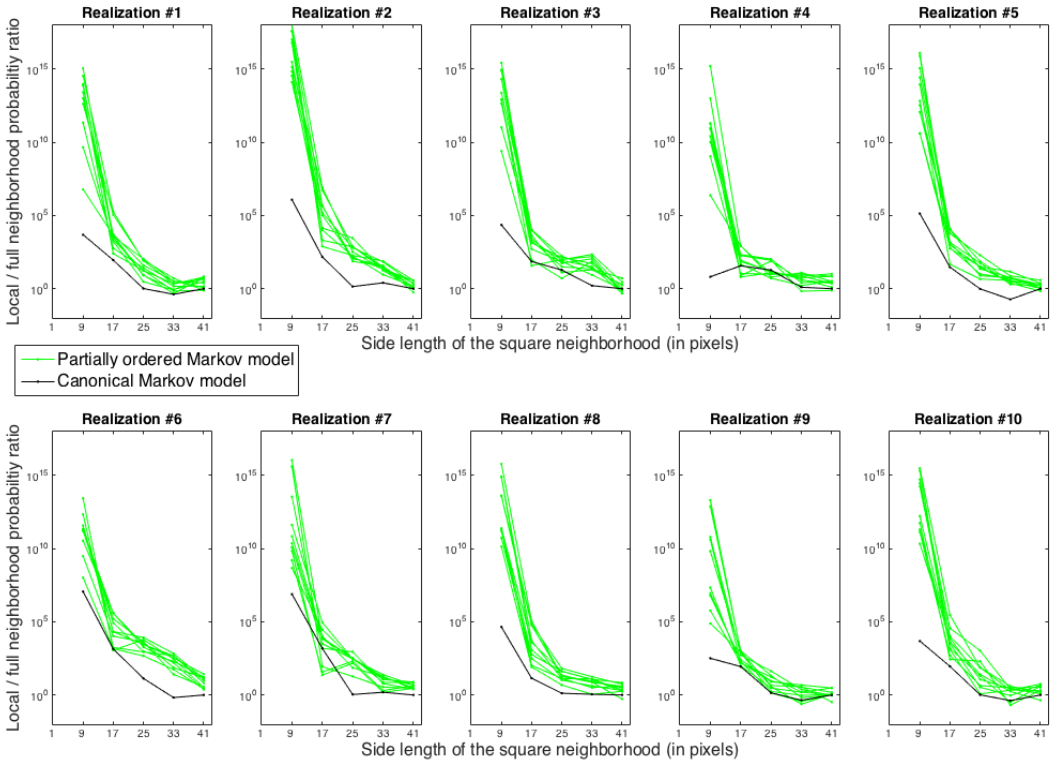

Figure 7. Ratios of probability density values between the 'correct' Gaussian random field using a full neighborhood and different formulations of Markov random fields. Green curves: partially ordered Markov model formulated for different simulation sequences. Black curve: canonical Markov model. The ratios are compared for density values calculated for 10 different points (i.e., realizations) in the model parameters space.

7 Random paths in sequential simulation algorithms

Practical implementations of sequential simulation use a new randomly chosen path for each new simulation sequences. This random path is chosen from a uniform distribution. This gives a probability of $1/N!$ for a specific path to be chosen by the sequential simulation algorithm. Here, N is the number of model parameters to be sampled. Hence, the random field actually sampled by a sequential simulation algorithm will not be a single partially ordered Markov model, but a mixture of partially ordered Markov models, which is given as (Daly, 2004; Toftaker and Tjelmeland, 2013; Cordua et al., 2015):

$$f(m_1,\ldots,m_N)=\sum_{path}\frac{1}{N!}\prod_{i=1}^{N}p(m_i\mid pa(m_i)) \tag{28}$$

However, since the individual partially ordered Markov models are consistent with a set of marginals of different sizes for different simulation paths, Eq. 28 is cumbersome to evaluate. For a more thorough discussion on this topic see Cordua et al. (2015).

8 Conclusion

We outlined the fundamental steps of going from one or a few realizations (i.e., a sample model) from an 'underlying' unknown random field describing the subsurface, to reconstructing a random field that is, at least, consistent with the known (limited-size) marginals distributions (based on pattern statistics from the sample model) from this unknown random field. Firstly, under the assumption of stationarity, marginal probability distributions from the unknown random field can be obtained through (one-, two- or multiple-point-based) pattern statistics from a sample model. Then, a random field consistent with these known marginal distributions can be determined. We showed that this problem of determining an unknown random field consistent with known marginal distributions is, in general, an underdetermined problem. Different assumptions (i.e., additional constrains) about the random field must be chosen in order to produce a unique solution.

Practical implementations of sequential simulation use limited-size marginal distributions (i.e., conditional distributions derived from the marginals), to draw realizations from some random field. The typical implicit assumption (i.e., constraint) used by such algorithms is that the random field is stationary, has Markov properties and/or is a parametric distribution (e.g., Gaussian). In this case, the random field being sampled is a partially ordered Markov model.

The random field actually sampled by sequential simulation algorithms (e.g., SGSIM and SNESIM) can be formulated as a partially ordered Markov model. We find that these random fields depend, not only on the size of the marginal distributions used to derive the necessary conditional distributions (i.e., the size of the chosen template/neighborhood), but also on the sequential simulation sequence (i.e., the random path used). Furthermore, these distributions are generally not consistent with the known marginals and do not, in general, represent a stationary distribution.

We reviewed the canonical Markov model, which is an example of a possible solution that is consistent with known marginals. When this random field is sampled by sequential

simulation, it is exactly consistent with the known marginals, and also invariant with respect to the simulation sequence as long as the simulation follows a chain of cliques.

It is found that the information content (using Shannon entropy) of the canonical Markov model is higher than for the partially order Markov model. This is a result of the fact that, in contrast to the canonical Markov model, the partially order Markov model does not use all the information in form of conditional dependencies from the known marginal distributions.

Acknowledgements

We wish to thank DONG E&P for financial support.

Appendix A

Theorem

Consider an N-dimensional space M with points $\mathbf{m} = (m_1, m_2, ..., m_N)$, and a (joint) probability density (e.g., random field) $f(\mathbf{m})$ with marginal distributions $p_1(\mathbf{m}^{(1)}), p_2(\mathbf{m}^{(2)}), ..., p_k(\mathbf{m}^{(k)})$, each defined over a strict subset $\mathbf{m}^{(k)}$ $(k = 1, 2, ..., \mathrm{K})$ of the N variables $m_1, m_2, ..., m_N$. Assume further that for each m_N there is a finite-length closed interval I_n such that $f(\mathbf{m}) > 0$ in the product interval $I = I_1 \times \ldots \times I_N$. Then, besides $f(\mathbf{m})$, there exists infinitely many (joint) probability densities over M having $p_1, p_2, ..., p_k$ as their marginal distributions.

Proof

Since $f(\mathbf{m}) > 0$ in a closed interval I, there must be a positive real number a such that $f(\mathbf{m}) > a$ everywhere in I. Let us now define nonzero, integrable functions $\varepsilon_n(m_n)$ over each m_n with the following properties:

1. $\varepsilon_n(m_n) = 0$ outside the interval I_N.
2. $|\varepsilon_n(m_n)| < a^{1/N}$ within the interval I_N.
3. $\int_{-\infty}^{\infty} \varepsilon_n(m_n)\, dm_n = 0$.

It is clear that infinitely many such functions ε_n exist. For each choice of functions ε_n, the product function $\varepsilon(\mathbf{m}) = \varepsilon_1(m_1)\varepsilon_2(m_2)\ldots\varepsilon_N(m_N)$ satisfies the following properties:

1. $\varepsilon(\mathbf{m}) = 0$ outside I.
2. $|\varepsilon(\mathbf{m})| < a$ inside I.
3. $\int_M \varepsilon(\mathbf{m})d\mathbf{m} = 0$

Hence, the sum function $f^*(\mathbf{m}) = f(\mathbf{m}) + \varepsilon(\mathbf{m})$ is non-negative and integrable, and satisfies $\int_M f^*(\mathbf{m})\, d\mathbf{m} = 1$. We can therefore conclude that f^* is a (joint) probability density function over M.

Consider now the subset $\mathbf{m}^{(k)}$ of variables. The corresponding marginal distribution of f^* is

$$p^{*(k)}(\mathbf{m}^{(k)}) = \int f^*(\mathbf{m})d\mathbf{m}^{(\neg k)} \tag{A1}$$

where the integral is over variables that are not in $\mathbf{m}^{(k)}$. We now get

$$\begin{aligned} p^{*(k)}(\mathbf{m}^{(k)}) &= \int (f(\mathbf{m}) + \varepsilon(\mathbf{m}))\ d\mathbf{m}^{(\neg k)} \\ &= p^{(k)}(\mathbf{m}^{(k)}) + \int \varepsilon(\mathbf{m})d\mathbf{m}^{(\neg k)}. \end{aligned} \tag{A2}$$

However, per definition of the functions $\varepsilon_n(m_n)$, the latter integral is zero, meaning that the *k*'th marginals ($p^{*(k)}$ and $p^{(k)}$), of the two different (joint) probability densities f^* and f, are identical. Noting that each of the infinitely many choices of $\varepsilon_n(m_n)$ define a new f^*, we have completed the demonstration.

References

Besag, J., 1974. Spatial Interactions and the Statistical Analysis of Lattice Systems. Journal of the Royal Statistical Society. Series B (Methodological), 36 (2): 192-236.

Bishop, C.M., 2006. Pattern recognition and machine learning. Springer Science, New York.

Castillo, E., Gutiérrez, J.M., Hadi, A.S., 1997. Expert systems and probabilistic network models. Springer-Verlag New York, Inc.

Cordua, K.S., Hansen, T.M., Mosegaard, K., 2012. Monte Carlo full-waveform inversion of crosshole GPR data using multiple-point geostatistical a priori information. Geophysics, 77: H19 – H31.

Cordua, K.S., Hansen, T.M., Mosegaard, K., 2015. Improving the pattern reproducibility of multiple-point-based prior models. Mathematical Geosciences, 47, 317 – 343, DOI: 10.1007/s11004-014-9531-4.

Cover, T., Thomas, J., 2006. Elements of Information Theory. John Wiley & Sons, Inc., Hoboken, New Jersey.

Cressie, N., Davidson, J., 1998. Image analysis with partially order Markov models. Comput Stat Data Anal, 29: 1- 26.

Daly, C., 2004. Higher order models using entropy, Markov random fields and sequential simulation. In: O. Leuangthong and C. V. Deutsch (eds.), Geostatistics, Banff 2004, 215 - 224.

Deutsch, C.V., Journel, A.G., 1998. GSLIB: Geostatistical Software Library, Oxford Press. Pp. 369.

Faucher, C., Saucier, A., Marcotte, D., 2013. A new patchwork simulation method with control of the local-mean histogram. Stoch Environ Res Risk Assess, 27: 253 – 273.

Gómez-Hernández, J., Cassiraga, E.F., 2000. Sequential conditional simulations with linear constrains. In: Monestiez P, Allard D, Froideveaux R (Eds.), Geostatistics 2000 Cape Town. Geostatistical Association of Southern Africa.

Gómez-Hernández, J., Journel, A.G., 1993. Joint Sequential Simulation of MultiGaussian Fields. In: Geostatistics Tróia ´92, Soares, A. (Ed.), Kluwer Academic Publishers, volume 1, pp. 85 – 94.

Guardiano, F., Srivastava, M., 1993. Multivariate geostatistics: beyond bivariate moments. In: Geostatistics - Troia, pp. 133 – 144. Kluwer Academic, Dordrecht.

Hansen, T.M., Mosegaard, K., Cordua, K.S., 2008. Using geostatistics to describe complex a priori information for inverse problems, in J. M. Ortiz, and X. Emery, eds., Geostatistics 2008: Chile, vol. 1.

Hansen, T.M., Cordua, K.S., Mosegaard, K., 2012. Inverse problems with non-trivial priors: Efficient solution through sequential Gibbs sampling. Computational Geosciences, 16: 593 – 611.

Irving, J.D., Singha, K., 2010. Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities: Water Resources Research, 46, no. 11, W11514, doi: 10.1029/2009WR008340.

Jaynes, E.T., 1957. Information theory and statistical mechanics. The Physical Review, 106(4): 620 - 630.

Journel, A.G., 1994. Modeling uncertainty: some conceptual thoughts. *in* Dimitrakopoulos, R., ed., Geostatistics for the Next Century: Kluwer Academic Pub., Dordrecht, The Netherlands, p. 30–43.

Journel, A.G., Alabert, F., 1989. Non-Gaussian data expansion in the earth science. Terra Nova, v. 1:123 – 134.

Journel, A.G., Huijbregts, C.J., 1978. Mining geostatistics. Academic Press.

Kindermann, R., Snell, J.L., 1980. Markov Random Fields and their Applications. Contemporary Mathematics, American Mathematical Society, Vol. 1.

Kjønsberg, H., Stien, M., Kolbjørnsen, O., Fjellvoll, B., Abrahamsen, P., 2012. Using Multiple Grids in Markov Mesh Facies Modeling. Ninth International Geostatistics Congress, Oslo, Norway, June 11. – 15., 2012

Mariethoz, G., Renard, P., Straubhaar, J., 2010. The Direct Sampling method to perform multiple-point geostatistical simulations. Water Resources Research, 46: W11536.

Rezaee, H., Mariethoz, G., Koneshloo, M., Asghari, O., 2013. Multiple-point geostatistical simulation using the bunch-pasting direct sampling method. Computers and Geosciences, 54: 293 - 308.

Shannon, C.E., 1948. A Mathematical Theory of Communication. The Bell System Technical Journal. 27: 379 – 423.

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

Tjelmeland, H., Toftaker, H., 2012. On the construction of discrete distributions from training images. Ninth International Geostatistics Congress, Oslo, Norway, June 11. – 15., 2012.

Toftaker, H., Tjelmeland, H., 2013. Construction of Binary Multi-grid Markov Random Field Prior Models from Training Images. Mathematical Geosciences, 45, 383 – 409.

Whittaker, J., 1990. Graphical Models in Applied Multivariate Statistics. Wiley.