PETROLEUM 2015
GEOSTATISTICS

26569

A flexible Markov mesh model for facies modelling

X. Luo* (Norwegian University of Science and Technology) & H. Tjelmeland (Norwegian University of Science and Technology)

SUMMARY

In this presentation we consider the problem of estimating a prior model for spatial discrete variables from a training image. To be able to combine the estimated prior model with a likelihood for observed data, we argue that one needs to use a prior formulation where an explicit formula is available for the prior distribution. Moreover, we argue that to avoid overfitting it is essential to limit the number of parameters in the prior. We propose to formulate a prior within the class of Markov mesh models, for which formulas for the point mass function are available. We define a flexible prior model within the class of Markov mesh models, where we are able to limit the number of model parameters even with a reasonably large sequential neighbourhood by restricting interactions of very high orders to be zero. To fit the Markov mesh prior to a training image we adopt a Bayesian approach, in which we consider the training image as observed data. We fit the model parameters to the training image by simulating from the resulting posterior distribution, and for this we use Gibbs sampler algorithm. We demonstrate the qualities of our approach in simulation examples.

Introduction

In the geostatistical community it has in the last years become common practice to estimate a prior model for the spatial distribution of reservoir properties from one or several training images. Several so called multi-point models have been defined for this purpose, see for example Strebelle (2002) and Journel and Zhang (2006). In Stien and Kolbjørnsen (2011) a Markov mesh model (Abend et al., 1965) is used for the same purpose. An important advantage of Markov mesh models relative to the traditional multi-point statistics models is that an explicit expression for the estimated distribution is available. The distribution of the estimated model in the multi-point statistics case is typically only implicitly defined via a simulation procedure. The importance of having an explicit expression for the estimated model becomes evident when the model is adopted as a prior distribution corporated with a likelihood function related to observed data into a corresponding posterior distribution. If the prior is then only implicitly defined via a simulation algorithm it becomes in most cases impossible to device a simulation algorithm that samples correctly from the corresponding posterior distribution. The focus in the multi-point statistics literature is therefore limited to constructing simulation algorithms that reproduce observed data without conditioning on it in a proper manner. When an explicit expression is available for the estimated model, as for example in the Markov mesh model case, a corresponding explicit expression becomes available also for the posterior distribution, except for an unknown normalising constant. This explicit expression for the posterior can in turn be the basis for constructing simulation algorithms that sample from this distribution. Different algorithms dependent on the mathematical structure of the prior and likelihood can be used for this. The Metropolis–Hastings algorithm is a general approach which can cope with most structures of the posterior distribution, but if the posterior formula has specific structures more efficient alternatives may exist.

In this presentation we discuss the Markov mesh model formulation, and consider for simplicity binary training images only. We device a flexible parameterisation for this model class, and adopt a Bayesian approach to fit the model parameters. The focus of the presentation is on the specification of Markov mesh model and the Bayesian model used to fit the parameter values of this model. We do not consider how to use the fitted prior model in a new Bayesian model in order to condition on observed data from a reservoir under study, but to device a simulation algorithm that is able to simulate from the resulting posterior distribution in such a Bayesian model is clearly a natural next step of the research.

The Markov mesh model

In this presentation we consider a rectangular lattice of size $m \times n$ and use $(i,j), i \in \{1,\ldots,m\}, j \in \{1,\ldots,n\}$ to denote a specific node of this lattice, corresponding to the notation used for the elements in a matrix. We let S denote the set of all the nodes, and to each node in S we have an associated binary variable $x_{ij} \in \{0,1\}$. The set of all these variables is denoted by $x = (x_{ij},(i,j) \in S)$, and we use $x_A = (x_{ij},(i,j) \in A)$ to denote the set of variables in any subset $A \subseteq S$. We define the set of predecessors, B_{ij}, of a node (i,j) to consist of all nodes numbered before (i,j) when the nodes are numbered in the lexicographically order. Thus $B_{ij} = \{(k,l) \in S : k < i\} \cup \{(i,l) \in S : l < j\}$. To each node $(i,j) \in S$ the Markov mesh model associate a sequential neighbourhood $N_{ij} \subset B_{ij}$. A Markov mesh model for x is then assuming the following Markov structure

$$p(x) = \prod_{(ij)\in S} p(x_{ij}|x_{N_{ij}}), \tag{1}$$

and the model is specified by specifying N_{ij} and $p(x_{ij}|x_{N_{ij}})$ for each $(i,j) \in S$. Except for nodes (i,j) close to the boundary of the lattice we assume the sequential neighbourhoods to be translations of a template sequential neighbourhood $T \subset \{(i,j) : i,j \in N, i < 0\} \cup \{(0,j) : j \in N, j < 0\}$, where N is the set of all integers. More specifically, we then define $N_{ij} = (T \oplus (i,j)) \cup S$, where $T \oplus (i,j) = \{(i+k, j+l),(k,l) \in T\}$. A simple example is $T = \{(0,-1),(-1,-1),(-1,0),(-1,1)\}$, in which case N_{ij}'s sufficiently far away from the lattice borders, i.e. $T \oplus (i,j) \subseteq S$, consist of four nodes as illustrated in

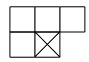


Figure 1 *Illustration of the sequential neighbourhood T when $T = \{(0,-1),(-1,-1),(-1,0),(-1,1)\}$. Node (i,j) is represented as a cell with a $\times$, whereas the elements in the set T are represented with empty cells.*

Figure 1. For all nodes $(i,j) \in S$ where $T \oplus (i,j) \subseteq S$ we assume $p(x_{ij}|X_{N_{ij}})$ to have the form

$$p(x_{ij}|x_{N_{ij}}) = \frac{\exp\{\theta(x_{N_{ij}})x_{ij}\}}{1+\exp\{\theta(x_{N_{ij}})\}}, \tag{2}$$

where $\{\theta(z); z = (z_{ij},(i,j) \in T) \in \Theta = \{0,1\}^{|T|}\}$ is a set of parameters to be specified, and $|T|$ is the number of elements in the set T. We get the most general model by putting no further restrictions on the $\theta(z)$'s, in which case we get one model parameter for each configuration of variables in the template T, and the total number of model parameters becomes $2^{|T|}$. To avoid overfitting, however, we add restrictions on the allowed values for the $\theta(z)$'s. To be able to model the quite complex structure occurring in many training images one would need to let the number of elements in T be quite large, and then the problem of overfitting clearly becomes an important issue to consider. To specify how we put restrictions on the $\theta(z)$'s we first need to introduce an alternative representation of the $\theta(z)$'s.

For two configurations in the sequential neighbourhood, $z = (z_{ij},(i,j) \in T) \in \Theta$ and $z^\star = (z^\star_{ij},(i,j) \in T) \in \Theta$, we write $z^\star \leq z$ if $z^\star_{ij} \leq z_{ij}$ for all $(i,j) \in T$. Thus, $z^\star \leq z$ means that to any zero element in (i,j) the corresponding element in $z^\star$ is also zero. Using this notation we define a one-to-one relation between the $\theta(z)$'s and a corresponding set of $\{\beta(z); z \in \Theta\}$ given by

$$\beta(z) = \sum_{z^\star \leq z} (-1)^{|z|-|z^\star|}\theta(z^\star), \;\; z \in \Theta \tag{3}$$

and

$$\theta(z) = \sum_{z^\star \leq z} \beta(z^\star), \;\; z \in \Theta. \tag{4}$$

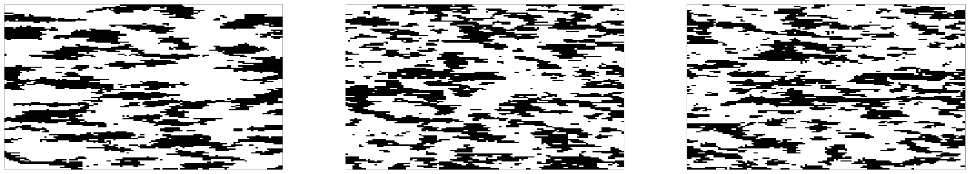
This relation is known as Moebious inversion (Lauritzen, 1906), and for Markov random fields a corresponding reparameterisation is used in Arnesen and Tjelmeland (2015). The $\beta(z)$ parameters have an interpretation as interaction parameters, and we put restrictions on the $\theta(z)$ parameters by fixing some of the $\beta(z)$'s to zero. Typically it is interactions of higher orders that are fixed to zero, for example one may put $\beta(z) = 0$ whenever the number of nodes in z is larger than four (say). We let $\Theta_0 \subset \Theta$ denote the set of z values where we restrict $\beta(z) = 0$, and let $\Theta_1 = \Theta \setminus \Theta_0$. With the restriction $\beta(z) = 0$ for $z \in \Theta_0$ the set of free $\theta(z)$ parameters becomes $\{\theta(z); z \in \Theta_1\}$. Given the values $\theta(z)$ for $z \in \Theta_1$ and that $\beta(z) = 0$ for $z \in \Theta_0$, one can then find the values of $\theta(z), z \in \Theta_0$ from relations (3) and (4). In fact, the $\theta(z)$'s, $z \in \Theta_0$ become linear functions of the elements of $\{\theta(z); z \in \Theta_1\}$.

The Bayesian model used for model fitting

Assuming we have an available training image, x, we now want to fit the vector of model parameters $\theta = \{\Theta(z), z \in \Theta_1\}$ in the Markov mesh model defined above to x . To do this we adopt the Bayesian setup, let $p(x|\theta)$ be given by the above Markov mesh model and assume some prior $p(\theta)$ for θ. The posterior distribution of interest then becomes

$$p(\theta|x) \propto p(\theta)p(x|\theta). \tag{5}$$

The likelihood function $p(x|\theta)$ is thus given by (1) and (2), but now the conditioning in these equations is also on the values of θ and to avoid boundary problems we let the product in (1) only be over nodes



(a) (b) (c)

Figure 2 *(a) The training image used in the simulation exercise. (b) and (c) Two independent realisations from the estimated model.*

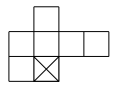


Figure 3 *The sequential neighbourhood used in the simulation example.*

$(i,j) \in S$ for which the sequential neighbourhood is completely inside the training image, i.e. $T \oplus (i,j) \subseteq S$. For the prior we simply assume the elements in the θ vector to be independent and normally distributed with zero mean and some large variance σ^2. In the numerical simulations presented below we use $\sigma = 20$. The posterior distribution $p(\theta|x)$ is not analytically tractable, so to explore the distribution we use a Markov chain Monte Carlo algorithm.

Posterior simulation

To sample from the posterior distribution $p(\theta|x)$ we may adopt a Metropolis–Hastings algorithm. The challenge then is to find a good proposal distribution. What makes this particularly difficult is that a good proposal distribution for one training image may not be a good choice for another training image. To avoid this problem we instead adopt the Gibbs sampler to generate samples from $p(\theta|x)$. Then, for each $z \in \Theta_1$ in turn, we need to sample from the full conditional

$$p(\theta(z)|x,\theta_{-z}) \propto p(\theta,x) = p(\theta)p(x|\theta), \tag{6}$$

where $\theta_{-z} = \{\Theta(z^\star), z^\star \in \Theta_1 \setminus \{z\}\}$. For the Markov mesh model formulated above these full conditionals do not result in well-known distributions from which it is known how to sample. However, it can be shown that the full conditionals are all log-concave, i.e. that

$$\frac{\partial^2}{\partial \theta(z)^2} p(\theta(z)|x,\theta_{-z}) < 0 \tag{7}$$

for all $\theta \in \Theta_1$. Sampling from $p(\theta(z)|x,\theta_{-z})$ can then be done by the adaptive rejection sampling algorithm introduced in Gilks (1992) and we adopt this to run our Gibbs sampler simulation.

Simulation examples

Figure 2(a) shows the training image we use in a simulation exercise. To illustrate the potential of the Markov mesh model formulation we use it for the quite small sequential neighbourhood in Figure 3. In the model we include only interactions up to order four, so $\Theta_0 = \{\theta(z); z \subseteq T \text{ and } |z| > 4\}$ and $\Theta_1 = \{\theta(z); z \subseteq T \text{ and } |z| \leq 4\}$. The resulting number of model parameters is $|\Theta_1| = 57$. Figures 2(b) and (c) show realisations from models simulated by the Gibbs sampler. We see that this rather simple Markov mesh model is able to capture much of the structure in the training image, but at the same time it is easy to see structures in the training image that are not reproduced in the simulations. With a larger sequential neighbourhood and more model parameters this will be improved. However, finding good values for T and Θ_1 by trial and error is clearly a tedious process, so there is inevitably a need to automatise this procedure. We discuss this issue more below.

Closing remarks

In this presentation we have defined a flexible Markov mesh model and discussed how to fit such a model to a given training image by adopting a Bayesian model formulation and applying a Gibbs sampling algorithm. We have demonstrated the properties of the model and fitting procedure by showing results for one training image. Our model formulation differs from procedures previously proposed for such a task. We do not find an optimal set of parameters, but rather simulate from a posterior distribution. We are thereby able to represent the uncertainty in parameter values when making inference based on the training image.

As discussed above, a problematic aspect of our formulation is that we manually have to specify the sequential neighbourhood, T, and the set of model parameter values, Θ_1. The number of possible values for T and Θ_1 is clearly huge, and what values for T and Θ_1 that produce a good model for a specific training image will depend on the spatial structures occurring in that particular training image. So to specify T and Θ_1 manually, as we have done in the example presented above, is clearly a tedious process and will typically result in a suboptimal solution. A better solution would be to extend the prior model part of our model by formulating a prior distribution also for T and Θ_1. In Arnesen and Tjelmeland (2015) a prior distribution is formulated for corresponding quantities for Markov random fields, and as the next step in our research we plan to adapt this prior formulation to the Markov mesh model case, and generalise the sampling algorithm accordingly.

Acknowledgements

We thank the sponsors of the Uncertainty in Reservoir Evaluation (URE) project (https://wiki.math.ntnu.no/ure) at the Norwegian University of Science and Technology (NTNU) for financial support to this work.

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. [2015] Prior specification of neighbourhood and interaction structure in binary Markov random fields. Tech. rep., ArXiv e-prints 1501.06344v1, Available from http://arxiv.org/abs/1501.06344v1.
Gilks, W.R. [1992] Derivative-free adaptive rejection sampling for Gibbs sampling. *Bayesian Statistics 4*, Oxford University Press, Oxford, UK, 641–649.
Journel, J. and Zhang, T. [2006] The necessity of a multiple-point prior model. *Mathematical Geology*, **38**, 591–610.
Lauritzen, S. [1906] *Graphical Models*. Clarenden Press, Oxford.
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.