Mathematical Geology, Vol. 34, No. 1, January 2002 (© 2002)

On Modelling Discrete Geological Structures as Markov Random Fields[1]

Tommy Norberg,[2] Lars Rosén,[3] Ágnes Baran,[4] and Sándor Baran[4]

The purpose of this paper is to extend the locally based prediction methodology of BayMar to a global one by modelling discrete spatial structures as Markov random fields. BayMar uses one-dimensional Markov-properties for estimating spatial correlation and Bayesian updating for locally integrating prior and additional information. The methodology of this paper introduces a new estimator of the field parameters based on the maximum likelihood technique for one-dimensional Markov chains. This makes the estimator straightforward to calculate also when there is a large amount of missing observations, which often is the case in geological applications. We make simulations (both unconditional and conditional on the observed data) and maximum a posteriori predictions (restorations) of the non-observed data using Markov chain Monte Carlo methods, in the restoration case by employing simulated annealing. The described method gives satisfactory predictions, while more work is needed in order to simulate, since it appears to have a tendency to overestimate strong spatial dependence. It provides an important development compared to the BayMar-methodology by facilitating global predictions and improved use of sparse data.

KEY WORDS: simulations, predictions, Markov chain Monte Carlo, simulated annealing, incomplete observations.

INTRODUCTION

In many geological applications, there is an interest in predicting properties at locations where no observations have been made previously. The inherent-heterogeneities of geological systems result in spatial variability and make such predictions uncertain. To account for this uncertainty, spatial statistical models

[1]Received 21 January 2000; accepted 16 January 2001.

[2]Department of Mathematical Statistics, Chalmers University of Technology, SE-41296 Göteborg, Sweden.

[3]Department of Geology, Chalmers University of Technology, SE-41296 Göteborg, Sweden; e-mail: lrnmvg@scc.se

[4]Institute of Mathematics and Informatics, Kossuth Lajos University, H-4010 Debrecen, Hungary.

may be applied. Spatial statistical geological models may be used for purposes such as: (1) estimations of probabilities of specific properties, (2) local calculations of the most likely geological property, (3) global predictions of the geological configuration, and (4) simulations of geological property configurations, given the present uncertainties. Many types of earth science data are displayed on a lattice format, which, according to Cressie (1993), is a countable collection of spatial sites that can be either spatially regular or irregular. Examples on earth science lattice data are satellite images and geological maps in raster format.

The most sophisticated spatial lattice models data are based on the Markov random field (MRF) approach, as described by e.g., Besag (1974) and Cressie (1993). Recently, Tjelmeland and Besag (1998) studied a more advanced Markov property with interaction also between directions. Because of computational demands, these models have so far been of limited use in geological applications. However, due to the rapidly increasing CPU capability, MRF methods gain increasing practical interest, especially in image restoration. Rosén and Gustafson (1996) described a computationally simple Bayesian–Markov approach (BayMar) for estimating conditional probabilities for geological properties from data on a regular lattice. The methodology uses Markov chain analysis for modelling the spatial properties and Bayesian updating for estimation of conditional probabilities. The BayMar methodology provides the possibility for local, but not global, predictions of the geological property. Further, the BayMar methodology does not provide possibilities for simulations. Another weakness of the BayMar technique is its inability to properly describe spatial variability in cases of sparse data, which is not unusual in geological applications.

In geological applications the difference between prediction and simulation should be recognized. Simulations are better suited for displaying the inherent variability, whereas predictions give the most likely configurations, given the available data. Typically a prediction is smoother than a simulation.

The present work was performed in order to provide supplementary capabilities to the BayMar methodology. It is the purpose of this paper to describe a procedure for (1) global prediction of the geological configuration, (2) simulating geological configurations, using an MRF approach, and (3) improved estimation of spatial variability in cases of sparse information.

The disposition of the paper is as follows: We commence by discussing the statistical model, how to estimate its parameters and how to do restorations and simulations using a Markov chain Monte Carlo (MCMC) technique (in the former case employing simulated annealing (SA), cf. e.g., Aarts and Korst (1989)). Then we describe our two geological areas of study. The results from the restorations and the simulations are presented next and, finally, in our last section we make conclusions and present some suggestions on how to proceed with this research.

STATISTICAL MODEL

The Probability of Observing a Specific Geological Configuration

We assume below that the set $\mathcal{C}$ of all possible geological configurations $\mathbf{x}$ is of the form S^R, where the *site space* R is a rectangular subset of the d-dimensional integers Z^d and $S = \{1, \ldots, s\}$ is a finite *state space*. In our two particular applications R is a rectangle in Z^2 of the form $\{1, \ldots, n\} \times \{1, \ldots, m\}$. Even for moderate n, m, and s, the total number of possible configurations is extremely large.

The probability of the configuration $\mathbf{x}$ is $p(\mathbf{x}) = Z^{-1} \exp(-\Psi(\mathbf{x}))$, where Z is a normalizing constant. In our applictaions, it is not possible to calculate Z because of the large cardinality of $\mathcal{C}$. Thus, what typically is known about the probability $p(\mathbf{x})$ is that it is proportional to $\exp(-\Psi(\mathbf{x}))$:

$$p(\mathbf{x}) \propto \exp(-\Psi(\mathbf{x})) \tag{1}$$

In statistical physics, $\Psi(\mathbf{x})$ is termed the *energy* of the observation $\mathbf{x}$. Less energy means a likelier observation. We will adopt this terminology below. The constant Z depends on the choice of Ψ, which to some degree is arbitrary.

We have a particular interest in configurations that are only partially known. We thus split $\mathbf{x} = \mathbf{x}_A\mathbf{x}_B$ into a known part $\mathbf{x}_A$ and an unknown part $\mathbf{x}_B$. The sets A and B partition the site space R. The possibility $A = \emptyset$ corresponds to no prior knowledge.

Often one may want to predict $\mathbf{x}_B$. We therefore calculate the conditional probability of $\mathbf{x} = \mathbf{x}_A\mathbf{x}_B$ given $\mathbf{x}_A$, which we write as $p_A(\mathbf{x}_B) = p(\mathbf{x}_A\mathbf{x}_B)/p(\mathbf{x}_A)$, because it is a function of the partial configuration $\mathbf{x}_B$ only. Clearly,

$$p_A(\mathbf{x}_B) = \frac{\exp(-\Psi(\mathbf{x}_A\mathbf{x}_B))}{\sum_{\mathbf{x}_B} \exp(-\Psi(\mathbf{x}_A\mathbf{x}_B))}$$

Typically the sum in the denominator is intractable, so, cf. (1),

$$p_A(\mathbf{x}_B) \propto \exp(-\Psi_A(\mathbf{x}_B)) \tag{2}$$

Two aims of this paper are:

1. to simulate a geological configuration either unconditionally according to (1) or conditionally according to (2)
2. to predict $\mathbf{x}_B$ given $\mathbf{x}_A$:

$$\hat{\mathbf{x}}_B = \arg \max_{\mathbf{x}_B} p_A(\mathbf{x}_B) = \arg \min_{\mathbf{x}_B} \Psi_A(\mathbf{x}_B)$$

Both aims can be achieved using MCMC (in the second case employing simulated annealing). MCMC is a simulation technique in which one just needs to know the energy of the target probability distribution. There is no principal difference between unconditional simulation of (1) and conditional simulation of (2). A difficulty with the second aim is that a quantity such as $\arg\max_{\mathbf{x}} p(\mathbf{x})$ need not be uniquely defined.

Specification of the Energy Ψ

We now proceed to describe the particular class of energy functions that we work with in this paper. We focus on the case of a two-dimensional site space, and assume that there are *mn* sites (or *pixels*) and that the number of possible geological states per site is s:

$$\mathbf{x} = (x_{(i,j)}, i = 1, \ldots, m,\ j = 1, \ldots, n)$$

where $x_{(i,j)} \in \{1, \ldots, s\}$. Specifically we will assume that $\Psi(\mathbf{x})$ can be written as a sum

$$\Psi(\mathbf{x}) = \sum_{i,j} \big(\psi_0(x_{(i,j)}) + \psi_1(x_{(i,j-1)}, x_{(i,j)}) + \psi_2(x_{(i-1,j)}, x_{(i,j)}) + \psi_3(x_{(i-1,j-1)}, x_{(i,j)}) + \psi_4(x_{(i+1,j-1)}, x_{(i,j)})\big) \tag{3}$$

Here, as well as in similar situations, the contribution to the energy is zero if an index pair points to a nonexisting site. This model specifies basic dependence in four directions (1: W–E; 2: N–S; 3: NW–SE; 4: SW–NE). The isotropic Ising model (Guyon (1995)) has $s = 2$ and $\psi_3(x, y) = \psi_4(x, y) = 0$,

$$\psi_1(x, y) = \psi_2(x, y) = \begin{cases} \beta & \text{if } x = y \\ -\beta & \text{if } x \neq y \end{cases}$$

and $\psi_0(x) = \alpha(\alpha \neq 0$ or $\alpha = 0$ depending on whether there is an external field or not).

Any probability distribution (1) with $\Psi(\mathbf{x})$ as above is called a *Gibbs distribution* or a *Markov random field* (*MRF*), see e.g., Guyon (1995). Below, we will think of the function $\psi_0(x)$ as a mean of four functions: $\psi_0(x) = (\psi_{01}(x) + \psi_{02}(x) + \psi_{03}(x) + \psi_{04}(x))/4$. In theory, all four $\psi_{0l}(x)$ coincide with $\psi_0(x)$. Their estimates, however, typically differ.

Statistical Inference

We next discuss how to calculate a nonparametric (N-P), or rather distribution free, estimator $\hat{\Psi}$ of Ψ within a certain class of energy functions which can be

described via the transition matrices of four Markov chains (one chain for each basic dependence direction). We will first describe how to estimate ψ_{01} and ψ_1. How to estimate the other three pairs ψ_{02} and ψ_2, ψ_{03} and ψ_3, and ψ_{04} and ψ_4 will follow by analogy. We first consider the case of a complete observation.

Example 1. Let $\mathbf{x} = x_1, \ldots, x_n$ be n consecutive states of a Markov chain with transition probabilities $p(x, y)$, $1 \leq x, y \leq s$. Assume that the chain is stationary, and denote by $\pi(x)$ its invariant distribution. The probability of the observed configuration $\mathbf{x}$ may be written as

$$p(\mathbf{x}) = \exp\left(\sum_{i=1}^{n} \log \pi(x_i) + \sum_{i=1}^{n-1} \log \frac{p(x_i, x_{i+1})}{\pi(x_{i+1})}\right)$$

Combining (1) and (3),

$$p(\mathbf{x}) \propto \exp\left(-\sum_{i=1}^{n} \psi_0(x_i) - \sum_{i=1}^{n-1} \psi_1(x_i, x_{i+1})\right)$$

By comparison, $\psi_0(x) = -\log\pi(x)$ and $\psi_1(x, y) = -\log(p(x, y)/\pi(y)) = \log \pi(y) - \log p(x, y)$, both modulo some additive constant that is not allowed to depend on $\mathbf{x}$. Replacing $p(x, y)$ and $\pi(x)$ with their maximum likelihood estimators $\hat{p}(x, y)$ and $\hat{\pi}(x)$, respectively, yields the estimators

$$\hat{\psi}_{01}(x) = -\log\hat{\pi}(x) + C_{01}$$

and

$$\hat{\psi}_1(x, y) = \log \hat{\pi}(y) - \log \hat{p}(x, y) + C_1$$

of ψ_{01} and ψ_1, respectively. The arbitrary constants C_{01} and C_1 are not allowed to depend on data. □

In analogy with this example, we define for each basic dependence direction l

$$\hat{\psi}_{0l}(x) = -\log\hat{\pi}_l(x) + C_{0l} \tag{4}$$

and

$$\hat{\psi}_l = \log \hat{\pi}_l(y) - \log \hat{p}_l(x, y) + C_l \tag{5}$$

where

$$\hat{p}_l(x, y) = \frac{f^l_{xy}}{f^l_{x\cdot}}$$

here $f^l_{x\cdot} = \sum_y f^l_{xy}$, where f^l_{xy} is the frequency of transitions from state x into state y in direction l; $\hat{\pi}_l$ is the unique distribution solving $\hat{\pi}_l \hat{P}_l = \hat{\pi}_l$, where $\hat{P}_l$ is the transition matrix formed by the estimates $\hat{p}_l(x, y)$, $1 \le x, y \le s$. Also,

$$\hat{\psi}_0 = \frac{1}{4}(\hat{\psi}_{01} + \hat{\psi}_{02} + \hat{\psi}_{03} + \hat{\psi}_{04}) \tag{6}$$

Any estimator of this kind will later be referred to as an N-P CMEE, based on the complete configuration. The drawback of the CMEE is that it does not take into account the normalizing constant $Z(\Psi)$. Thus it may differ considerably from the MLE of $\boldsymbol{\Psi}$, which we, due to extremely long computing times, have not calculated.

In the case of a partially observed configuration we think in each direction of the data as that of a noncomplete observation of a Markov chain. For each direction l, we calculate a likelihood as in the numerical Example 2 below, which when maximized gives estimates $\hat{\pi}_l(x)$ and $\hat{p}_l(x, y)$ from which the CMEE may be calculated (cf. (4) and (5)).

Example 2. Consider a homogeneous Markov chain with state space $S = \{1, 2\}$ and transition matrix

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$

having invariant distribution $\pi = [\pi_1 \quad \pi_2]$ satisfying $\pi P = \pi$. As in Example 1, our aim is to estimate P and π from a finite realization of the chain. This time, however, we are not able to observe the chain at each moment.

As an example of an incomplete observation, consider

$$\cdots 0\,0\,0\,1\,0\,0\,2\,2\,1\,0\,0\,0\,0\,1\,0\,0\,2\,2\,1\,0\,0\,0\,0\,1\,0\,0\,2\,2\,1\,0\,0\,0\,0\,1\,0\,0\,2\,2\,1\,0\,0\,0 \cdots$$

where 0 denotes an unknown state. Assuming stationarity, its likelihood is

$$L(P) = \pi_1 p^{(3)}_{12} p_{22} p_{21} p^{(5)}_{11} p^{(3)}_{12} p_{22} p_{21} p^{(5)}_{11} p^{(3)}_{12} p_{22} p_{21} p^{(5)}_{11} p^{(3)}_{12} p_{22} p_{21}$$

where we have written $p^{(k)}_{ij}$ for the probability of jumping from state i to state j in k steps ($p^{(1)}_{ij} = p_{ij}$). Notice that we have assumed here that the chain is

in its stationary phase. Writing $p = p_{12}$ and $q = p_{21}$, and noticing $\pi = [q \quad p]/(p+q)$, the expression for $L(P)$ can be simplified to[5]

$$\begin{aligned} L(p,q) = p^4 q^5 (1-q)^4 (3 - 3p + p^2 - 3q + 2pq + q^2)^4 (1 - 5p \\ + 10p^2 - 10p^3 + 5p^4 - p^5 + 10pq - 20p^2 q + 15p^3 q - 4p^4 q \\ - 10pq^2 + 15p^2 q^2 - 6p^3 q^2 + 5pq^3 - 4p^2 q^3 - pq^4)^3 / (p+q) \end{aligned}$$

By maximizing this expression, we get the maximum likelihood estimator of the pair p, q.

There is one slight difficulty, though, since a likelihood like this is not necessarily concave i.e., there may be more than one local maximum. Thus, a careless use of a standard maximization program may produce a wrong answer. □

Instead of routinely employing a standard maximization algorithm, we used a simulated annealing algorithm from Corana and others (1987) (see also Baran and Baran (1997) or Baran and Szabó (1997)) to maximize the (possibly nonconcave) likelihood in the case of noncomplete observations.

OUR MARKOV CHAIN MONTE CARLO ALGORITHM

The idea behind simulating realizations of probability distributions such as (1), in which the normalizing constant is not known, with an MCMC algorithm is to construct a homogeneous Markov chain having the probability distribution itself as its invariant distribution. One of the first algorithms of this kind was described by Metropolis and others (1953).

One difficulty with the MCMC approach to simulation of (1) is that it is difficult to detect when the chain has reached or is close to equilibrium. To handle this problem we have logged and plotted the number of updated sites, the increase or decrease in energy in each new configuration. We have also summed the latter to obtain the energy difference between the current and the initial configuration.

The MCMC algorithm that we chose to realize uses the so called Gibbs sampler introduced by Geman and Geman (1984). Unlike earlier MCMC algorithms, the Gibbs sampler updates one site at a time. After updating all unknown sites, configuration 1 is ready and after updating all unknown sites a second time, configuration 2 is ready, etc.

A difficulty with MCMC algorithms is that they often are very time and computer consuming. Ours is not an exception. Often the Gibbs sampler needed 1.5–2.5 days for a run on a SUN workstation (10–15 min per configuration and maybe 250 configurations to reach or come close to equilibrium).

[5]We employed Mathematica to do the simplification.

By running MCMC with invariant distribution

$$p_k(\mathbf{x}) \propto \exp\left(-\frac{\Psi(\mathbf{x})}{T_k}\right)$$

(we will refer to T_k simply as the temperature of configuration k) and slowly lowering T_k until the sequence of configurations converge there is good hope that the final configuration is close to one with minimal energy (cf. our second aim above). This is often referred to as simulating annealing (SA). The interested reader is referred to Laarhoven and Aarts (1987) or Aarts and Korst (1989).

One often has to do experiments with different so called *cooling schedules* $\{T_k, k = 0, 1, \ldots\}$. A quite popular class of cooling schedules are the *exponential* ones. Any exponential cooling schedule can be written

$$T_k = T_0 c^k$$

The initial temperature is T_0 and we will refer to the constant c, which determines the speed of the convergence of T_k towards 0, in terms of the *half temperature*

$$k_{1/2} = -\log 2/\log c$$

In other words

$$c = 2^{-1/k_{1/2}}$$

so that

$$T_k = T_0 2^{-k/k_{1/2}}$$

The "speed" parameter $k_{1/2}$ is the number of iterations needed to decrease the temperature by a factor of 2. Many of our runs use an exponential cooling schedule with $T_0 = 4$ and $k_{1/2} = 100$.

STUDY AREAS

Two areas were selected for test applications during this work. The first area is situated in Lerum, approximately 25 km east of Göteborg in southwestern Sweden (Fig. 1, upper left). The Lerum area is 5×5 km^2 in size and covered by a geological map of unconsolidated materials in raster format with a pixel size of 50×50 m^2. This area was selected because of two reasons: (1) the distribution of unconsolidated materials is completely known and (2) the area displays a large

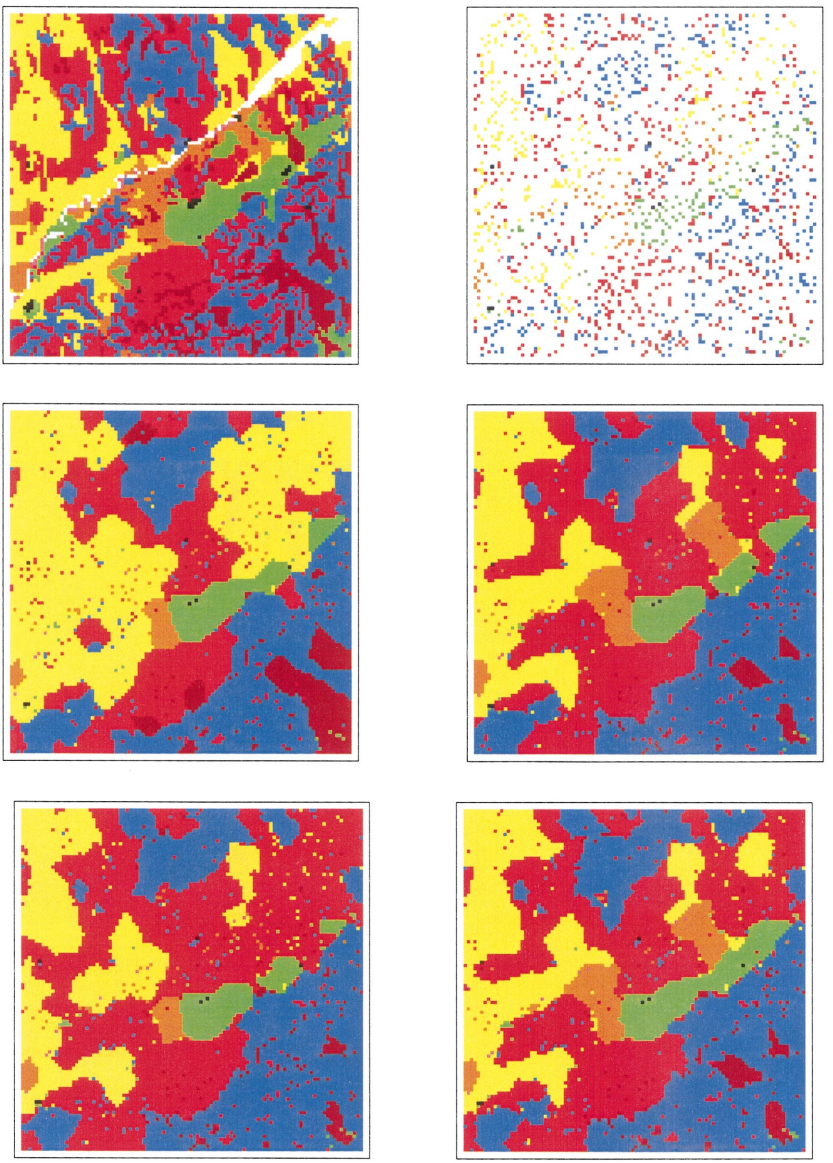

Figure 1. Upper left: Lerum map (SML); upper right: The thinned Lerum map (TSML); middle left: MAP d4 of the thinned Lerum map (N-P CMEE based on SML, $T_0 = 4$, $k_{1/2} = 100$, $k_s = 354$, $p_C = 62.6\%$); middle right: MAP d5 of the thinned Lerum map (N-P CMEE based on TSML, $T_0 = 4$, $k_{1/2} = 100$, $k_s = 400$, $p_C = 65.7\%$); lower left: Conditional simulation d8 of the Lerum area; known is the thinned Lerum map (N-P CMEE based on TSML, $T_0 = 1$, $k_{1/2} = \infty$, $k_s = 250$, $p_C = 63.4\%$); lower right: Conditional simulation a6 of the Lerum area; known is the thinned Lerum map (N-P CMEE based on TSML, $T_0 = 4$, $k_{1/2} = 100$, $\min_k T_k = 1$, $k_s = 457$, $p_C = 66.8\%$).

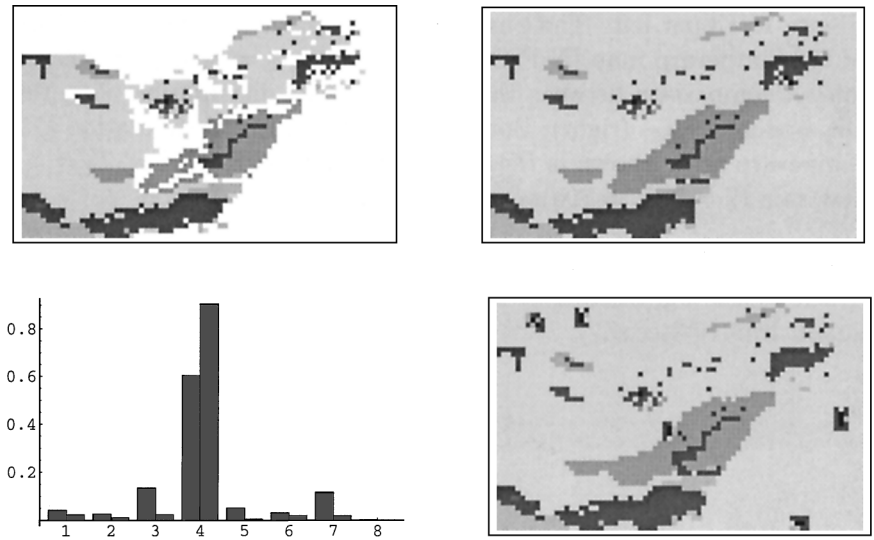

Figure 2. Upper left: The Simpevarp map (BMS); upper right: MAP b14 of the Simpevarp map (N-P CMEE, $T_0 = 4, k_{1/2} = 100, k_s = 400$); lower left: a comparison between the measured state distribution ρ_{BMS} (left) and the estimate $\hat{\pi}_{\text{BMS}}$ (right); lower right: Conditional simulation c11 of the Simpevarp area; known is the Simpevarp map (N-P CMEE, $T_0 = 4, k_{1/2} = 100, \min T_k = 1, k_s = 700$).

variability. The second area is a 7.5×4.6 km^2 area around Simpevarp in the southeastern part of Sweden (Fig. 2, upper left). The area includes a hard rock laboratory for testing of bedrock properties for repository of high level spent nuclear waste and is subject is intensive geoscientific research (see Rosén and Gustafson (1996)). This area is covered by a bedrock map in raster format with a pixel size of 100×100 m^2. There are a total of $102 \times 102 = 10404$ pixels (sites) in the Lerum map of which 265 represent surface water (colored white). The Simpevarp map displays much less variation than the Lerum map, and portions of the bedrock is unknown, since the Baltic Sea covers parts of the area.

The two selected areas exhibit large differences, which were considered to be important for evaluation purposes. The Lerum map (Fig. 1, upper left) displays $s = 8$ different categories of glacial deposits. The area is situated in a northeast–southwest oriented valley in Precambrian bedrock of granitic composition with gneissic structures. The lower elevations of the area are dominated by glaciomarine clays. Glaciofluvial esker deposits appear in the central part of the valley, and are partly covered by the glaciomarine clays. On the valley sides, glacial till is the dominating type of deposit. In the upper elevations, especially in the northern part of the area, bedrock outcrops frequently.

The bedrock geology (Fig. 2, upper left) of the Simpevarp area is dominated by granitoids belonging to the Trans-Scandinavian Igneous Belt (Wikberg, Gustafson,

and Stanfors (1991). The bedrock composition is locally heterogeneous with some older xenoliths, mainly metavolcanics, and some younger intrusions of brittle fine-grained granite. The dominating rock type is a medium-grained granite, which covers approximately 60% of the mapped area. The older, basic, rock types, such as metavolcanics, make up approximately 15% of the area.

RESULTS

The Lerum Area

To check the capacity of our algorithms to predict nonobserved states, we made a Bernoulli thinning of the Lerum map (Fig. 1, upper right). The probability for a site to be kept in the thinned map was 0.15. We estimated the energy function from the data in this map and we made several restorations and conditional simulations.

We report a total of seven maximum posterior predictions (MAPs) of the unknown part of the thinned Lerum map (TSML). For six predictions we used a relatively slow cooling schedule with half temperature, $k_{1/2} = 100$, and initial temperature, $T_0 = 4$, whereas we for one used a much slower schedule with $T_0 = 50$ and $k_{1/2} = 200$. The reason for doing an MAP with this very time consuming schedule was to see whether better results can be obtained with a much slower schedule. The result indicates that this is not the case.

Of the six MAPs based on the fast cooling schedule, three were produced using as energy function the N-P CMEE based on the complete configuration SML, while the other three were produced using the N-P CMEE based on the thinned Lerum configuration TSML. Refer to Figure 1, middle left, as an example of the former three, and to Figure 1, middle right, as an example of the latter three (k_s is the total number of sweeps). For the former three we obtained the proportions 61.7, 62.2, and 63.0% correct pixels, whereas for the latter three we obtained the proportions 65.7, 66.6, and 66.0%. It is remarkable that an MAP that uses the N-P CMEE based on the thinned map seems to perform better than an MAP based on the knowledge in the complete map. A t test of the null hypothesis of no difference versus the alternative that there is a difference between MAPs using the N-P CMEE based on the complete and the thinned map clearly rejects the null hypothesis of no difference, the two-sided p-value being less than 0.002.

As mentioned already, we calculated two N-P CMEEs. One was based on the complete Lerum map SML and the other one was based on the thinned Lerum map TSML. Both CMEEs produced an estimate of the distribution of the $s = 8$ different states denoted $\hat{\pi}_{\mathrm{SML}}$ and $\hat{\pi}_{\mathrm{TSML}}$, respectively, which can be compared to the estimate ρ_{SML} obtained simply by counting frequencies in the Lerum map SML (Fig. 3, left). Notice that the agreement between $\hat{\pi}_{\mathrm{TSML}}$ and ρ_{SML} is remarkably good, and much better than that between $\hat{\pi}_{\mathrm{SML}}$ and ρ_{SML}. Also

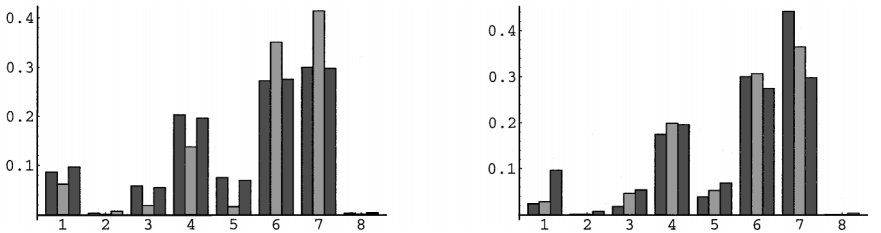

Figure 3. Left: A comparison between the measured state distribution ρ_{SML} (left) and the two estimates $\hat{\pi}_{\mathrm{SML}}$ (middle) and $\hat{\pi}_{\mathrm{TSML}}$ (right); right: A comparison between the measured state distribution in the two maps TSML-d8 (left) and TSML-a6 (middle) with $\hat{\pi}_{\mathrm{TSML}}$, which is obtained from the N-P CMEE based on the thinned Lerum map (right).

notice that $\hat{\pi}_{\mathrm{SML}}$ overestimates the two most frequent states and underestimates the others.

Figure 1, lower left, shows a conditional simulation with constant temperature $T_k = 1$ throughout the whole run. Figure 1, lower right, shows a conditional simulation in which we have used cooling to obtain a reasonable initial configuration. The first 200 iterations were produced using a cooling schedule with $T_0 = 4$ and $k_{1/2} = 100$, whereas the following ones all had $T_k = 1$. Figure 3, right, shows a comparison of the distribution of the different states in the maps in Figure 1, lower left, (denoted $\rho_{\mathrm{TSML-d8}}$) and Figure 1, lower right, (denoted $\rho_{\mathrm{TSML-a6}}$) with $\hat{\pi}_{\mathrm{TSML}}$. Notice that, similar to classical smoothing, both TSML-d8 and TSML-a6 contain too many of the most frequent states and too few of the less frequent states. Notice also that $\rho_{\mathrm{TSML-a6}}$ is closer to the target distribution $\hat{\pi}_{\mathrm{TSML}}$ than $\rho_{\mathrm{TSML-d8}}$. The difference between run a6 and run d8, is that we used cooling to obtain a better initial configuration than for run a6.

The Simpevarp Area

For all restorations and simulations of the Simpevarp map (BMS), the N-P CMEE based on all available information (calculated as outlined in Section) was used. In Figure 2, lower left, we compare the estimate of the state distribution obtained from the CMEE (denoted $\hat{\pi}_{\mathrm{BMS}}$) with the one obtained by just counting states (which as denote ρ_{BMS}). Notice that $\hat{\pi}_{\mathrm{BMS}}$ severely overestimates the most frequent state and underestimates the other. This is in agreement with the results regarding $\hat{\pi}_{\mathrm{SML}}$ for the Lerum map.

We did four MAPs of the unknown part of the Simpevarp map (BMS). For all MAPs a cooling schedule with $T_0 = 4$ and $k_{1/2} = 100$ was used. All predictions show similar results; for an example, see Figure 2, upper right. Notice that in all MAPs the configuration with $k = 200$ turned out very similar to the final configuration. Thus, very little changed after the first 200 iterations.

Figure 2, lower right, shows a conditional simulation, the initial configuration of which is obtained by cooling with $T_0 = 4$ and $k_{1/2} = 100$. There seems to be no principal difference between the four MAPs (compare Fig. 2, upper right, and this conditional simulation).

CONCLUSIONS AND FURTHER WORK

The following conclusions and suggestions for further work were made from this study:

- The methodology seems to be stable, because several different runs, with the same parameter setting, return approximately the same fraction p_C of correct pixels. In addition, the structure of restored configurations having the same parameter setting appear to be very similar for both the Lerum and the Simpevarp areas.
- Our method does not seem to be able to reproduce in a simulated map, the transition frequencies of the original map. This is true also for the state distribution, because the simulations are based indirectly on the transition frequencies (Fig. 3, right). Here the agreement between $\rho_{\mathrm{TSML-a6}}$ and the target distribution $\hat{\pi}_{\mathrm{TSML}}$ is better than the agreement between $\rho_{\mathrm{TSML-d8}}$ and $\hat{\pi}_{\mathrm{TSML}}$. Thus, starting a restoration and running it until T_k becomes 1 in order to obtain an initial configuration seems to improve the result of a conditional simulation. A possible explanation of this phenomena could be that the CMEE may have a tendency to overestimate spatial dependencies. See also the last remark below.
- Because of this tendency to overestimate spatial dependencies, this method is presently not well-suited for uncertainty analyses by means of repeated simulations.
- The state distribution estimate $\hat{\pi}_{\mathrm{TSML}}$ derived from the N-P CMEE based on only 15% of the data of the Lerum map is almost identical to the true state distribution ρ_{SML} of the complete map. Of interest here is how much we have to sample a geological configuration in order to obtain a reasonably good estimate of its true state distribution.
- It is odd that the N-P CMEE based on the thinned Lerum map yields significantly better restorations than the CMEE based on the complete Lerum configuration. Figure 3, left, shows that the N-P CMEE based on the thinned Lerum map better reproduces the known state distribution of the Lerum map than the N-P CMEE based on the complete map. Further work is needed to get a better understanding of this aspect.
- It seems that we derive better estimates of the state distribution ρ in a map if only a part of the configuration is used. This contradicts our intuition.

More work is needed in order to see whether this is truly the case or not, and if so to find an explanation.

- It should be noted that the modelling has so far been performed without consideration of geological processes. The Lerum area consists of two separate geological domains, (1) above and (2) below the highest shoreline after the glaciation. These two domains exhibit different processes for the formation of Quaternary deposits. Estimators of geological properties therefore tend to become a weighted average rather than representing one or both of the geological domains. We believe that a similar effect may be present also in the Simpevarp area. For more realistic predictions and simulations different geological domains should be carefully separated.
- So far we have only studied nonparametric CMEEs. There are parametric CMEEs that may perform better. To obtain a more thorough understanding of this kind of modelling several different kinds of CMEEs should be compared.
- A problem with modelling spatial phenomena with Markov random fields is the possible existence of phase transitions in infinite site spaces (see e.g., Guyon (1995)). Finite models like the ones we study often experience similar phenomena. Our results indicate that this kind of effect is present in the estimated Simpevarp model. It may also be present in the N-P CMEE of the Lerum area based on the complete map. We believe that when this effect is present, the CMEE overestimates the dependence and that the MCMC simulations/restorations further emphasise the dependence so that an unconditionally simulated map does not look realistic.

In summary, our method gives satisfactory predictions, while more work is needed in order to simulate without the mentioned tendency to overestimate spatial dependency. The described method provides an important development compared to the BayMar-methodology by facilitating global predictions and improved use of sparse data.

ACKNOWLEDGMENTS

Research supported by the Swedish Nuclear Fuel and Waste Management Company, the Swedish Natural Science Research Council and the TEMPUS Structural Joint European project No. 9521/95.

REFERENCES

Aarts, E., and Korst, J., 1989, Simulated annealing and Boltzmann machines: Wiley, New York, 242 p.

Baran, Á., and Baran, S., 1997, An application of simulated annealing to ML-estimation of a partially observed Markov chain: Report. Centre for Applied Mathematics and Statistics, Chalmers and Göteborg University, Sweden.

Baran, S., and Szabó Á., 1997, An application of simulated annealing to ML-estimation of a partially observed Markov chain, *in* Proceedings of the 3rd International Conference on Applied Informatics, Eger-Noszvaj, Hungary, August 24–28.

Besag, J., 1974, Spatial interaction and the statistical analysis of lattice systems (with discussion): J. Roy. Statist. Soc. B, v. 36, p. 192–236.

Corana, A., Marchesi, M., Martini, C., and Ridella, S., 1987, Minimising multimodal functions of continuous variables with the "simulated annealing" algorithm: ACM Trans. Math. Softw., v. 13, no. 3, p. 262–280.

Cressie, N., 1993, Statistics for spatial data, Rev. ed.: Wiley, New York, 900 p.

Geman, S., and Geman, D., 1984, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images: IEEE Trans., v. PAMI-6, p. 721–741.

Guyon, X., 1995, Random fields on a network: Springer, New York, 255 p.

Laarhoven, P. J. M., and Aarts, E. H. L., 1987, Simulated annealing: Theory and applications, Kluwer Academic, Dordrecht, 186 p.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E., 1953, Equations of state calculations by fast computing machines: J. Chem. Phys., v. 21, p. 1087–1091.

Rosén, L., and Gustafson, G., 1996, A Bayesian Markov geostatistical model for estimation of hydrogeological properties: Ground Water, v. 34, p. 865–875.

Tjelmeland, H., and Besag, J., 1998, Markov random fields with higher-order interactions: Scand. J. Stat., v. 25, p. 415–433.

Wikberg, P., Gustafson, G., and Stanfors, R., 1991. Äspö Hard Rock Laboratory: Evaluation and conceptual modelling based on the pre-investigations: SKB TR 91–22.