Mathematical Geology, Vol. 39, No. 2, February 2007 (© 2007)
DOI: 10.1007/s11004-006-9071-7

A Fixed-Path Markov Chain Algorithm for Conditional Simulation of Discrete Spatial Variables[1]

Weidong Li[2]

*The Markov chain random field (MCRF) theory provided the theoretical foundation for a nonlinear Markov chain geostatistics. In a MCRF, the single Markov chain is also called a "*spatial Markov chain*" (SMC). This paper introduces an efficient fixed-path SMC algorithm for conditional simulation of discrete spatial variables (i.e., multinomial classes) on point samples with incorporation of interclass dependencies. The algorithm considers four nearest known neighbors in orthogonal directions. Transiograms are estimated from samples and are model-fitted to provide parameter input to the simulation algorithm. Results from a simulation example show that this efficient method can effectively capture the spatial patterns of the target variable and fairly generate all classes. Because of the incorporation of interclass dependencies in the simulation algorithm, simulated realizations are relatively imitative of each other in patterns. Large-scale patterns are well produced in realizations. Spatial uncertainty is visualized as occurrence probability maps, and transition zones between classes are demonstrated by maximum occurrence probability maps. Transiogram analysis shows that the algorithm can reproduce the spatial structure of multinomial classes described by transiograms with some ergodic fluctuations. A special characteristic of the method is that when simulation is conditioned on a number of sample points, simulated transiograms have the tendency to follow the experimental ones, which implies that conditioning sample data play a crucial role in determining spatial patterns of multinomial classes. The efficient algorithm may provide a powerful tool for large-scale structure simulation and spatial uncertainty analysis of discrete spatial variables.*

KEY WORDS: Markov chain random field, spatial Markov chain, transiogram, multinomial classes, interclass relationship, nearest known neighbor.

INTRODUCTION

Conditional simulation is widely recognized and increasingly used, not only for prediction of spatial distribution but also for spatial uncertainty analysis through stochastic imaging (i.e., generating alternative conditional realizations) and probability mapping of discrete spatial variables (Deutsch and Journel, 1998, p. 119–191). Conventional geostatistics for conditional simulation of discrete variables usually use auto-covariance models and indicator kriging simulation algorithms

[1]Received 15 January 2005; accepted 10 September 2006; Published online: 12 April 2007.
[2]Department of Geography, Kent State University, 179 Dale Drive, Apt. 204, Kent, OH 44240, USA; e-mail: weidong6616@yahoo.com.

(Goovaerts, 1997, p. 376–378). The Markov chain random field (MCRF) theory recently proposed by Li (2006a) supports a non-linear Markov chain geostatistics. A MCRF refers to a random field of a single Markov chain. In a MCRF, the single Markov chain is also called a "spatial Markov chain" (SMC). The MCRF theory uses transiograms (i.e., transition probability diagrams) (Li, 2006b) to represent spatial self-dependencies (i.e., auto-correlations) and interdependencies (e.g., cross-correlations, juxtapositions, and directional asymmetries) of classes and thus requires using transiogram models to serve continuous transition probability input to Markov chain simulations. SMC models are potentially efficient in computation and capable of dealing with many classes, because they have explicit solutions for conditional probability functions of the unknown locations to be estimated. At the meantime they do not suffer from the constraints of existing multi-chains-based multidimensional (multi-D) Markov chain conditional simulation models (e.g., Elfeki and Dekking, 2001; Li and others, 2004) such as the small-class underestimation problem (relative to large-class overestimation). Particularly, interclass dependencies are naturally incorporated by cross-transiograms in a simulation. Directional asymmetry of class occurrence sequences may be incorporated by using cross-transiograms estimated unidirectionally.

1-D Markov chains (or transition probabilities) have long been used in the geosciences; see Carle and Fogg (1997) and Li, Li, and Shi (1999) for some reviews in geology and soil science, respectively. Multi-D Markov chain models can be traced to Krumbein (1968) and Lin and Harbaugh (1984) in geology for modeling lithological (or sedimentological) structures. However, earlier Markov chain models could not do conditional simulations, and it has long been a difficulty to condition a Markov chain to observed data, even more than one boundary (Koltermann and Gorelick, 1996, p. 2630–2632). Carle and Fogg (1997) developed 3-D continuous-lag Markov chain models by interpolating transition rate matrices established for three principal directions, but conditional simulation has to be performed through indicator kriging approaches and simulated quenching (see Weissmann and Fogg, 1999). Attempts in using multi-D Markov chains for conditional simulation of discrete variables appeared only in recent years; see Elfeki and Dekking (2001) for simulating lithofacies from well logs and Li and others (2004) for simulating soil types from survey line data, with the latter incorporating the idea of the former. Zhang and Li (2005) further applied the method of Li and others (2004) to spatial uncertainty assessment of land cover classes. However, these multi-D Markov chain conditional simulation models have some constraints. For example, they have the tendency of underestimating small classes because of exclusion of unwanted transitions caused by using multiple chains and the full independence assumption (Li, 2006a), and conditioning to point samples is also a difficulty. The MCRF theory theoretically eliminates unwanted transitions because there is only one chain in one random field, no matter how many dimensions are involved. For other related approaches that use transition probabilities (but not Markov chains) for conditional simulations, see Carle and Fogg (1996) and Norberg and

others (2002). The former used transition probability models to replace variograms in indicator kriging. The latter applied a Markov random field (MRF) model to conditional simulation on sparse data through simulated annealing.

It should be noticed that here I do not review the vast literature in MRF simulations in other fields such as image restoration although an investigation was made. Because MRF models consider only adjacent cliques (i.e., groups of pixels) and have to use iterative updating algorithms in simulation (e.g., Geman and Geman, 1984; Besag, 1986), they are unable to condition a simulation directly on sparse samples through distant interactions. Thus, they are not practical in conditional simulation from sparse samples and reproduction of large-scale patterns of categorical variables; see related review and attempt in Tjelmeland and Besag (1998), related attempt and discussions in Norberg others (2002), and related reviews in Qian and Terrington (1991) and Wu and others (2004). That is probably why MRFs are seldom used in practice in geospatial modeling. As a special subclass of MRFs, Markov mesh models (sometimes also called Markov chain models in image processing) were developed for image analysis and efficient reproduction of large-scale image patterns (e.g., Abend, Harley, and Kanal, 1965; Qian and Terrington, 1991; Wu and others, 2004). Markov mesh models are still cliques-based, but use asymmetric neighborhoods and generate realizations by a one-pass way. However, so far it is infeasible to use Markov mesh models for conditional simulation on samples.

While the MCRF theory provides the theoretical foundation for conditional 1-D to multi-D Markov chain simulation, how to use it in practice depends on the development of practical simulation algorithms. Using four nearest known neighbors in orthogonal directions to estimate the unknown point is the central idea of the Markov chain model implemented in Li and others (2004), which show that considering only four nearest known neighbors in orthogonal directions is feasible for generating 2-D fields of categorical variables. The MCRF theory also proved that the conditional independence assumption is correct for nearest known neighbors in cardinal directions in a Pickard random field (Pickard, 1980) in the sparse data situation. Therefore, the simulation algorithm proposed in this paper for working with point data is based on the four-nearest-known-neighbors idea. The practicality of the proposed algorithm is demonstrated by conditional simulations on two soil type datasets – a dense one and a sparse one from the same study area. Note that in this paper the word class refers to category generally.

SIMULATION METHODS

SMC Models

A MCRF is defined as the random field of a SMC that moves or jumps in a space, obeying the different (or same) transition probability rules in different

directions, with its state changes entirely depending on its nearest known neighbors found in different directions. The MCRF theory provides a general solution to the conditional probability distribution of a discrete random variable Z at an unsampled location u as

$$\Pr(Z(u) = k \mid Z(u_1) = l_1, \Lambda, Z(u_m) = l_m) = \frac{\prod_{i=2}^{m} p^i_{kl_i}(h_i) \cdot p^1_{l_1 k}(h_1)}{\sum_{f=1}^{n} \left[\prod_{i=2}^{m} p^i_{fl_i}(h_i) \cdot p^1_{l_1 f}(h_1)\right]} \tag{1}$$

(Li, 2006a), where p^i represents the transition probability in the ith direction; u_1 represents the neighbor from or across which the SMC moves to the current location u; m represents the number of nearest known neighbors; k, l_i, and f all represent classes in the state space $S = (1, \ldots, n)$; h_i is the distance from the current location to the nearest known neighbor u_i. With increasing lag h_i from zero to a certain distance, $p^i_{kl_i}(h_i)$ forms a transiogram (a continuous transition probability diagram) from class k to class l_i in the ith direction.

If we only consider four orthogonal directions – 1, 2, 3, and 4, Eq. (1) becomes

$$\begin{aligned} &\Pr(Z(u) = k \mid Z(u_1) = l, \quad Z(u_2) = q, \quad Z(u_3) = m, \quad X(u_4) = o) \\ &= \frac{p^1_{lk}(h_1) \cdot p^2_{kq}(h_2) \cdot p^3_{km}(h_3) \cdot p^4_{ko}(h_4)}{\sum_{f=1}^{n} \left[p^1_{lf}(h_1) \cdot p^2_{fq}(h_2) \cdot p^3_{fm}(h_3) \cdot p^4_{fo}(h_4)\right]} \end{aligned} \tag{2}$$

where k represents the state of the unsampled location to be estimated, and l, q, m, and o represent the states of the four nearest known neighbors in corresponding directions. Here 1, 2, 3, and 4 may represent the leftward, rightward, upward, and downward directions, respectively in any sequence.

Equation (2) is the principal SMC model to be used in the algorithm. For outer boundary simulation, the number of nearest known neighbors in orthogonal directions is less than four. So Eq. (2) has to be simplified to allow less than four nearest known neighbors. For example, if there are three nearest known neighbors, we have

$$\begin{aligned} &\Pr(Z(u) = k \mid Z(u_1) = l, \quad Z(u_2) = q, \quad Z(u_3) = m) \\ &= \frac{p^1_{lk}(h_1) \cdot p^2_{kq}(h_2) \cdot p^3_{km}(h_3)}{\sum_{f=1}^{n} \left[p^1_{lf}(h_1) \cdot p^2_{fq}(h_2) \cdot p^3_{fm}(h_3)\right]} \end{aligned} \tag{3}$$

If there are two nearest known neighbors, we have

$$\Pr(Z(u) = k \mid Z(u_1) = l, \quad Z(u_2) = q) = \frac{p^1_{lk}(h_1) \cdot p^2_{kq}(h_2)}{\sum_{f=1}^{n} \left[p^1_{lf}(h_1) \cdot p^2_{fq}(h_2)\right]} \tag{4}$$

If there is one nearest known neighbors, we have

$$\Pr(Z(u) = k \mid Z(u_1) = l,) = \frac{p^1_{lk}(h_1)}{\sum_{f=1}^{n} p^1_{lf}(h_1)} = p^1_{lk}(h_1) \tag{5}$$

Normally, a SMC has at least one nearest known neighbor, that is, its previous stay location. The only possible exceptional case is the initial location in a simulation, which occasionally may have no known neighbors in orthogonal directions. If so, we can randomly choose one state for it from the stationary probabilities—proportions of all involving classes.

Simulation Algorithm

Consider we simulate the spatial distribution of n mutually exclusive classes, which is conditional to a sample dataset at N grid nodes. The n classes may be in any random sequence, denoted as any labels $1, 2, \ldots, n$ or A, B, C, ... Monte Carlo sampling is used to draw labels from the cumulative conditional probability distribution for each node. The simulation algorithm proceeds as follows (Fig. 1):

- First connect peripheral observed points in the study area into simulated lines using the aforementioned SMC models – depending on the numbers of nearest known neighbors found in the orthogonal directions, which may be 1, 2, or 3 for different nodes (a corresponding Eqs. (5), (4), or (3) can be chosen as the SMC model). If these simulated lines do not overlay with outer boundaries of the study area, the outer boundaries can be simulated by conditioning on the already simulated lines.
- Connect all close pairs of observed data points that have not been connected in the first step, into simulated lines using the four-nearest-known-neighbors SMC model given in Eq. (2) by following a symmetric (or quasi symmetric) path. These simulated lines form a network composed of many meshes with known (already simulated) boundaries.
- In each mesh, simulate each unknown nodes using the four-nearest-known-neighbors SMC model given in Eq. (2) by following a symmetric (or quasi symmetric) path.

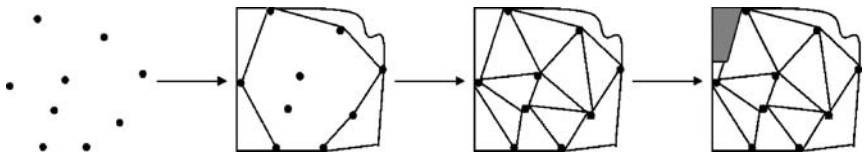

Figure 1. Illustrating the simulation process from netting to mesh-filling.

Such a simulation procedure includes a netting process and a mesh-filling process. In this study the alternate advancing (AA) path suggested in Li and others (2004) was used as the quasi symmetric path. The AA path may be regarded as an adaptation of the herringbone method suggested by Sharp and Aroian (1985) for solving the directional effect (i.e., pattern inclination) problem in simulation of multi-D autoregressive processes. Such a path means that simulation has to be conducted line by line alternately along opposite directions from one side to the other side within the simulation area. To calculate the conditional probability distribution for each unsampled node, transiograms are used to acquire the needed 1-D transition probabilities. Note that order relation problems suffered by indicator kriging do not exist in multi-D Markov chain simulation.

A visual interactive software system will be developed in the future so that simulation can be conducted interactively. For example, after users click on any two neighboring observed points the software will save the point pair (or all the nodes crossed by the line connecting the point pair); after all of these pairs of observed points (or all the nodes crossed by planned lines) are selected and saved, a network can be generated by connecting each pair of observed points into a simulated line during a simulation process. Based on the network, then the software simulates the other unknown locations. The interactive software also can allow users to choose specific simulation paths and directions at different steps and sections of the study area.

TRANSIOGRAM AND TRANSIOGRAM MODELING

In conditional Markov chain simulation, transition probabilities in one to multiple spatial steps are needed. If line data (continuous borehole logs or exhaustive observations of survey lines) are available, both one-step and multi-step transition probabilities can be easily estimated. However, estimation of transition probabilities from sparse point samples has long been a difficulty. Carle and Fogg (1996) used the transition rate matrix method (Krumbein, 1968) to infer continuous transition probability models from one-step transition probabilities and mean boundary spacings (e.g., mean lengths of facies) so that transiton probabilities at any needed lags can be acquired. Elfeki and Dekking (2001) and Li and others (2004) implicitly used the more conventional one-step transition probability matrix (TPM) method to calculate n-step transition probabilities by applying a power of n to a one-step TPM. These two methods generate similar results. They are simple but not widely applicable in practice. One reason is that directly estimating one-step transition probabilities from point samples is not feasible. While line data may be obtained from continuous borehole logs or field line survey for some visually observable variables, point sample data are widely available on the surface and in horizontal dimensions in the subsurface. The second reason is that they are

based on the first-order stationary Markovian assumption and thus have an intrinsic constraint that assumes boundary spacing (e.g., class polygon lengths) to be exponentially distributed. However, it may not be suitable to assume exponential distribution of class polygon sizes in many cases. For example, the thicknesses of lithofacies and alluvial soil textural layers usually tend to be lognormally distributed (Potter and Blakely, 1967; Li and others, 1997). The third reason is that the non-Markovian property (or called the high-order Markovian property) of sample data cannot be reflected on transition probability diagrams derived from one-step transition probabilities based on the first-order stationary Markovian assumption. Considering these reasons, Li (2006b) proposed the transiogram concept and suggested estimating continuous transition probability diagrams directly from sample data. Transition probability diagrams derived from one-step transition probabilities based on the first-order stationary Markovian assumption were thus called idealized transiograms (Li, 2006b), which are not widely available and have limited use.

A transiogram is defined as a two-point transition probability diagram over the distance lag h (Li, 2006b):

$$p_{ij}(h) = \Pr(Z(u+h) = j \mid Z(u) = i) \tag{6}$$

Here the random variable Z is assumed spatially second-order stationary, that is, $p_{ij}(h)$ is only dependent on the lag h and not dependent on the specific location u. $p_{ii}(h)$ represents the auto-transiogram of class i, and $p_{ij}(h)(i \neq j)$ represents the cross-transiogram from class i to class j. When cross-transiograms are estimated unidirectionally (e.g., east-to-west), we normally have $p_{ij}(h) \neq p_{ij}(-h)$. Therefore, they can be used to deal with directional asymmetry of class patterns. When cross-transiograms are estimated omni-directionally, we normally still have $p_{ij}(h) \neq p_{ji}(h)$. This means that juxtaposition relationships, such as class A frequently occurs beside class B but class B has more chances to occur beside class C, still can be captured.

Similar to estimation of variogram models in kriging geostatistics, estimation of transiogram models from sample data involves two steps: first estimate experimental transiograms from sample data and then infer transiogram models from experimental transiograms. Because transiograms are transition probability–based spatial dependence measures, they are different from variograms. For example, compared with indicator variograms, transiograms change (decrease or increase) along opposite directions with increasing lag h and their sills are theoretically equal to the proportions of the related tail classes (i.e., j in $p_{ij}(h)$). Further, transiograms are always positive and for exclusive classes they should have no nuggets theoretically and physically.

The basic mathematical models that are used in variogram modeling (Deutsch and Journel, 1998, p. 25) may be adapted to fit experimental transiograms. For

fitting experimental auto-transiograms, two basic models are adapted as follows (also see Ritzi, 2000):

a) Exponential model

$$p_{ii}(h) = 1 - (1 - p_i)[1 - \exp(-3h/a_i)] \tag{7}$$

b) Spherical model

$$\begin{aligned} p_{ii}(h) &= 1 - (1 - p_i)[1.5(h/a_i) - 0.5(h/a_i)^3]; \quad h < a_i \\ p_{ii}(h) &= p_i; \quad h \geq a_i \end{aligned} \tag{8}$$

For fitting experimental cross-transiograms, these two basic models are adapted as:

a) Exponential model

$$p_{ij}(h) = p_j[1 - \exp(-3h/a_{ij})] \tag{9}$$

b) Spherical model

$$\begin{aligned} p_{ij}(h) &= p_j[1.5(h/a_{ij}) - 0.5(h/a_{ij})^3]; \quad h < a_{ij} \\ p_{ij}(h) &= p_j; \quad h \geq a_{ij} \end{aligned} \tag{10}$$

In above models, a_i represents auto-correlation ranges, a_{ij} represents cross-correlation ranges, and p_i and p_j are proportions of corresponding class i and j, respectively. The sills of these models are explicitly set to the proportions of corresponding tail classes. To effectively fit the complex features of experimental transiograms, more complex models, such as a linear combination of basic models, may be necessary.

Given a fixed lag h, all transition probability values from the head class i to itself and all other classes should sum to 1, that is,

$$\sum_{j=1}^{n} p_{ij}(h) = 1 \tag{11}$$

To meet this condition in model fitting of experimental transiograms, one experimental transiogram $p_{ik}(h)$ among those involving the same head class i should be

left unfitted and its model can be calculated by

$$p_{ik}(h) = 1 - \sum_{\substack{j=1 \\ j \neq k}}^{n} p_{ij}(h) \tag{12}$$

To guarantee the non-negative and well-fitting of the transiogram model calculated by Eq. (12), other fitted transiogram models may need to be repeatedly adjusted.

When available sample data are too few and it is difficult to estimate reliable experimental transiograms from samples, expert empirical knowledge is needed to determine the sill, range, and shape of a transiogram model.

CASE STUDY

Datasets

The case study was conducted in a small area of 4×1.7 km with seven soil classes (here mean soil series) (Li and others, 2004). The physical meaning of the soil classes is not a concern here. The area was discretized into an 80×34 grid with a pixel size of 50×50 m. In this study we used raster data so that all of our data could be processed by GIS software. Two datasets were used for conditional simulations. A dense sample dataset consisted of 136 points (5% of total pixels), and a sparse sample dataset consisted of 45 points (1.6% of total pixels). Samples were distributed regularly.

Estimated Transiogram Models

Experimental transiograms were estimated from the 136 points and fitted by exponential and spherical models (with seven left ones calculated by Eq. (12)). Normally the most complex experimental transiograms were left to simplify the modeling process. The fitted transiogram models were used for simulations with both datasets. The sparse dataset of the 45 points was too small to estimate reliable experimental transiograms of seven classes (it was just used to demonstrate that the algorithm could work with very sparse data). Considering the dense dataset was still small, the transition frequencies in the four orthogonal directions were pooled together to get only one set (i.e., 49) of experimental transiograms. This means that anisotropies and directional asymmetries were not specifically considered in the case study. Because raster data were used, the lag h was represented as the number of pixels (i.e., grid units), not the exact distance.

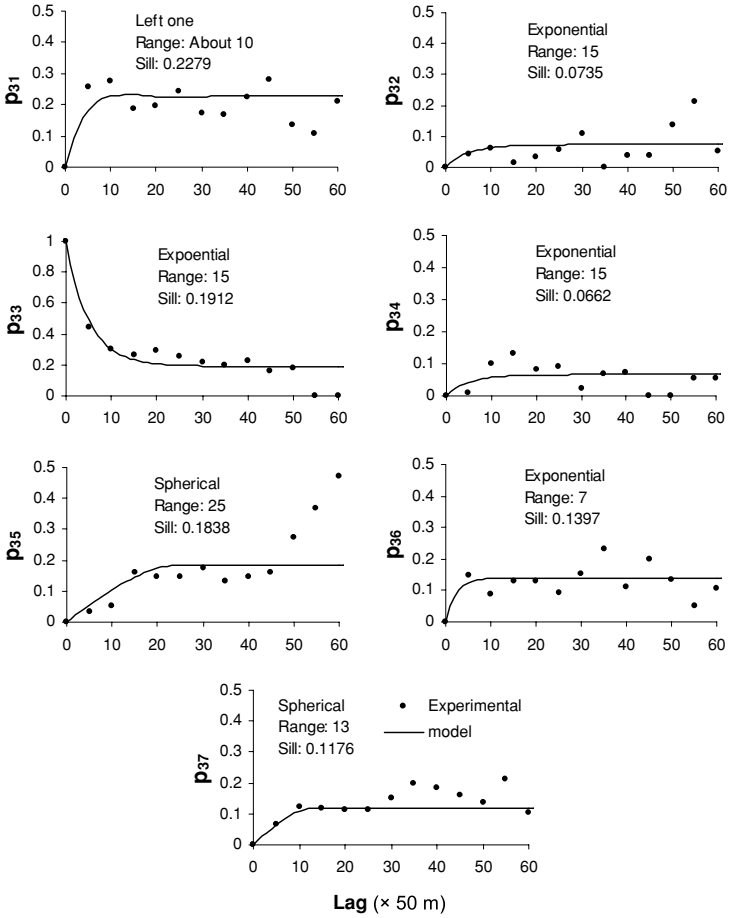

Figure 2. A subset of experimental auto/cross-transiograms and fitted models related with the same head class—class 3. Experimental transiograms (*dots*) are estimated from the dense dataset (136 regular points). Sills of transiogram models (*line*) are set to the proportions of the corresponding tail class.

Figure 2 illustrates a subset of experimental auto/cross transiograms that share the same head class (i.e., soil class 3) and their fitting models. It can be seen that most of these experimental transiograms can be approximately fitted by an exponential or spherical model, with the proportion of the corresponding tail soil class serving as the model sill. In Figure 2, the experimental cross-transiogram $p_{31}(h)$ is the left one, modeled by Eq. (12). Normally if other experimental transiograms are model-fitted appropriately, the left one should also be well-fitted.

Experimental transiograms have apparent fluctuations that are difficult to fit using the two basic models. This may be caused by the insufficiency of observed data and the non-Markovian effect of the real data (note that idealized transiograms based on the first-order stationary Markovican assumption are smooth curves). Fitted transiogram models capture only part of the features of experimental transiogams, depending on the complexity of the mathematical models used. Using composite hole-effect models (see Ma and Jones (2001) for composite hole-effect models for auto-variograms) may capture more details, such as periodicities, of experimental transiogams.

Simulated Results and Analysis

100 realizations were simulated for each dataset. Occurrence probability maps and the maximum occurrence probability map were estimated from each set of simulated realizations. Based on the maximum occurrence probability map for the corresponding dataset, an optimal prediction map was generated. A prediction map also represents the optimally interpolated result of a sparse sample dataset. Simulated results are displayed in Figures 3, 4 and 5.

Figure 3 shows the data used, prediction maps and two realizations for each dataset. It can be seen that all seven soil classes are well captured based on the observed data. Small soil classes such as class 2 and class 4, are well produced, even with the sparse dataset, in response to their occurrences in observed data and their correlation situations. An interesting and probably significant characteristic of simulated realizations is that large-scale soil patterns are generated as polygons, instead of fragmentary pixels as usually seen in simulated realizations of conventional geostatistical methods. Different realizations based on the same conditioning dataset show similar patterns. This should be attributed to the incorporation of interclass relationships (i.e., cross-correlations and juxtaposition relationships here) among different soil classes and the nonlinearity of the SMC estimator.

Occurrence probability maps are very useful in spatial uncertainty analysis of spatial data, which is actually one of major purposes of conditional simulation and has also been a central topic in geographical information science for two decades (Zhang and Li, 2005). Occurrence probability maps of single soil classes are displayed in Figures 4 and 5, respectively for the dense dataset and the sparse dataset. The maps clearly show where and with how much probability a soil class tends to occur in the study area. A dark area in an occurrence probability map means that the soil class has a high probability to occur at that place as a polygon. The gradual transition from a dark color to the background color implies the locational uncertainty of a polygon boundary. Maximum occurrence probability maps (Figs. 4 and 5, top left) indicate the robustness of the corresponding prediction

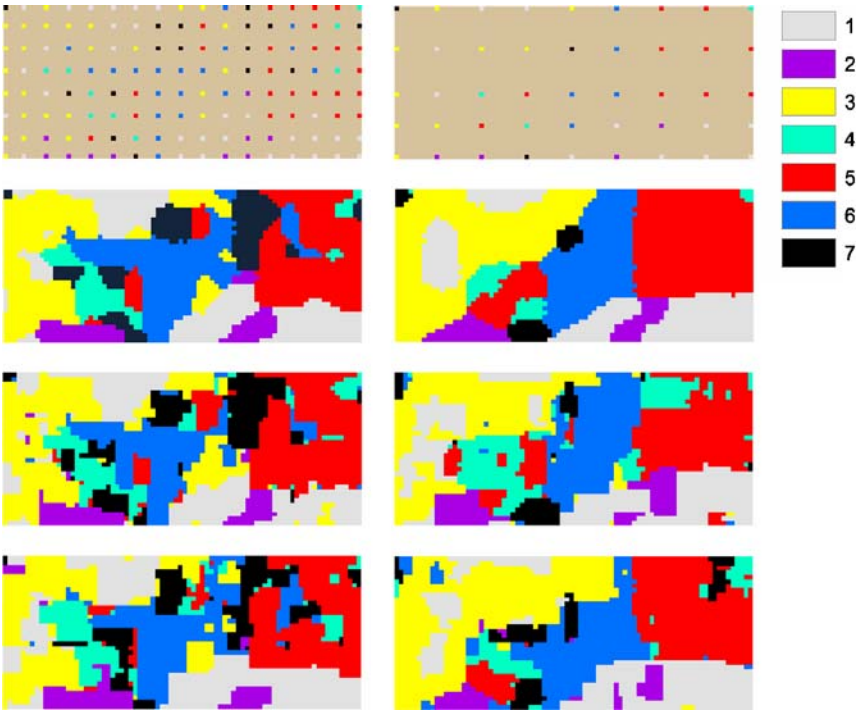

Figure 3. Conditioning datasets and simulated results. The first row shows the two datasets: the dense one has 136 points and the sparse one has 45 points. The second row shows prediction maps based on maximum occurrence probabilities. The third and forth rows show simulated realizations. Simulated results in each column are based on the corresponding dataset at the top of the column.

maps, that is, where the predicted result has lower quality and where it has higher quality. Particularly, the transition zones (or uncertain boundaries) between different soil classes may be revealed in a maximum occurrence probability map. The shallow grey stripes in the two maximum occurrence probability maps indicate the transition zones, which imply approximate class boundary locations and the locational uncertainty of boundaries.

Figure 6 shows a subset of transiograms with the same head class—soil class 1, estimated from the first realizations of the two simulations. Surprisingly, it can be seen that simulated transiograms conditioned on the dense dataset strongly follow the experimental transiograms, rather than the input transiogram models. This means that simulated realizations are more strongly affected by the conditioning data, from which the experimental transiograms are estimated, than by the input transiogram models. When data are relatively abundant, conditional Markov chain simulations seem not very sensitive to the transition probability input, as

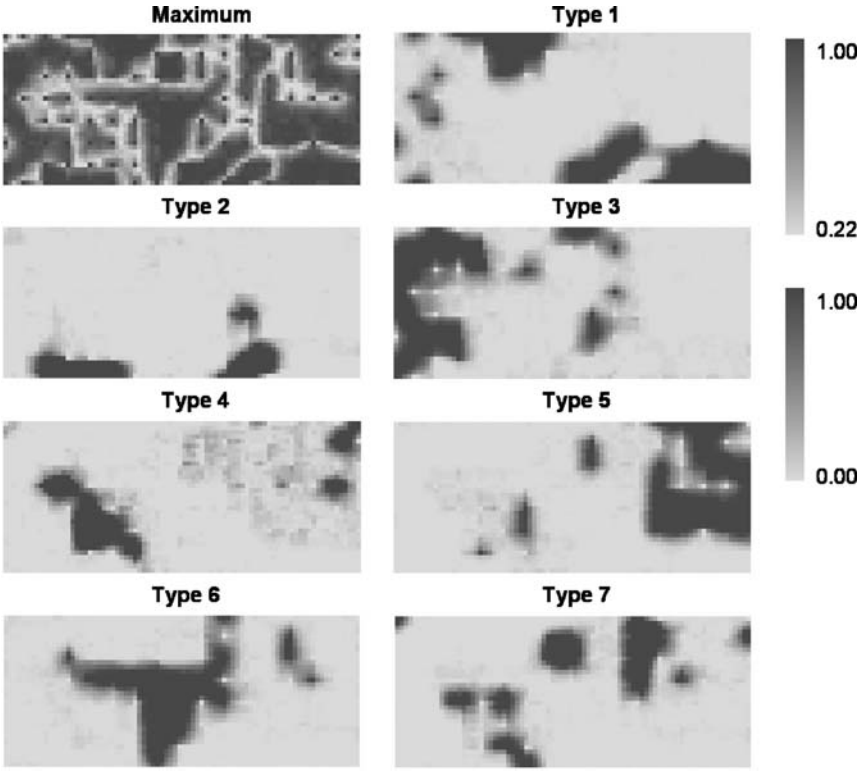

Figure 4. Occurrence probability maps estimated from 100 simulated realizations conditioned on the dense dataset.

also indicated in Li and others (2004). Simulated transiograms conditioned on the sparse dataset is more deviated from corresponding experimental transiograms and transiogram models. This may be because the experimental transiograms estimated from the dense dataset do not effectively reflect the spatial variation structure of the sparse dataset. Figure 7 shows simulated results of three different auto/cross-transiograms ($p_{11}(h)$, $p_{51}(h)$, and $p_{45}(h)$), estimated from the first 10 simulated realization maps conditioned on each of the two datasets. Apparently, the 10 simulated auto/cross-transiograms fluctuate normally around the corresponding experimental transiogram estimated from the dense conditioning dataset. Such fluctuations are called ergodic fuctuations in geostatistics, which is normal because of the limited extent of the spatial domain (Deutsch and Journel, 1998, p. 128–132). The small size of the study area (compared to the average polygon size of the seven soil classes) may be one reason leading to the obvious deviations of some simulated transiograms, especially those conditioned on the

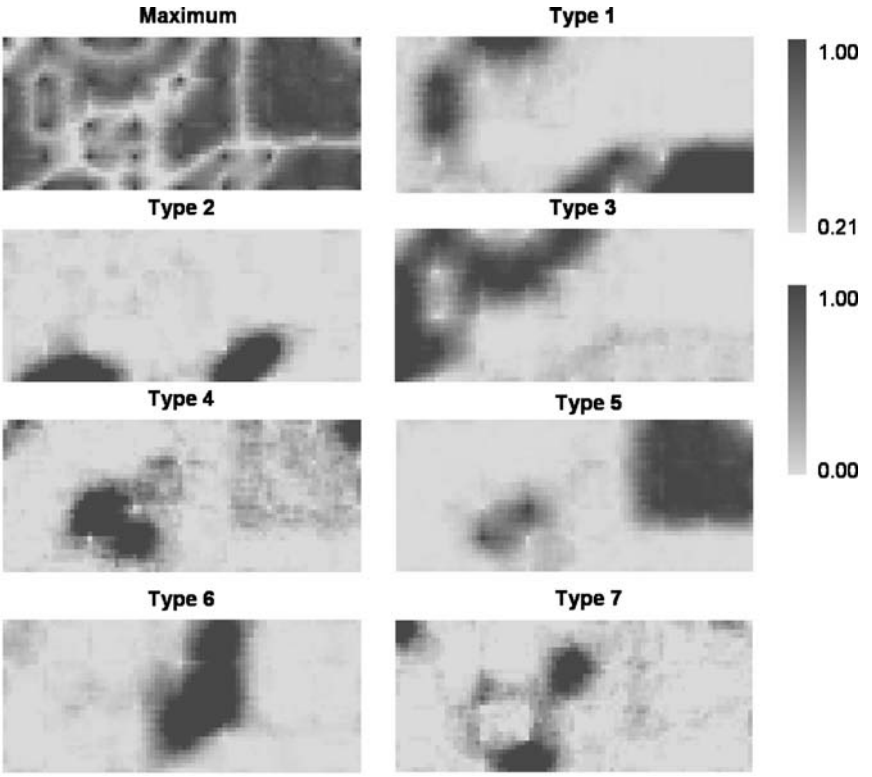

Figure 5. Occurrence probability maps estimated from 100 simulated realizations conditioned on the sparse dataset.

sparse dataset, because the soil distribution in a small area should be strongly non-stationary.

The tendency that simulated transiograms based on the dense dataset follow corresponding experimental transiograms is understandable, because the simulation is directly conditioned on the dataset used to estimate the experimental transiograms. With the incorporation of interclass relationships, close sample data are apparently more decisive in determining the state of an unknown location in a SMC simulation than in a kriging simulation. As a non-linear estimator, the SMC does not necessarily behave similarly as kriging—the linear optimal estimator. This may be why simulated patterns by the SMC algorithm show clear polygons and different realizations are more imitative (but still have locational uncertainty).

Figure 8 demonstrates that $p_{17}(h) \neq p_{71}(h)$, no matter in original data and simulated realizations. This is always true for every class pair. This means that not only cross-correlations, but also juxtaposition tendencies, are also incorporated in

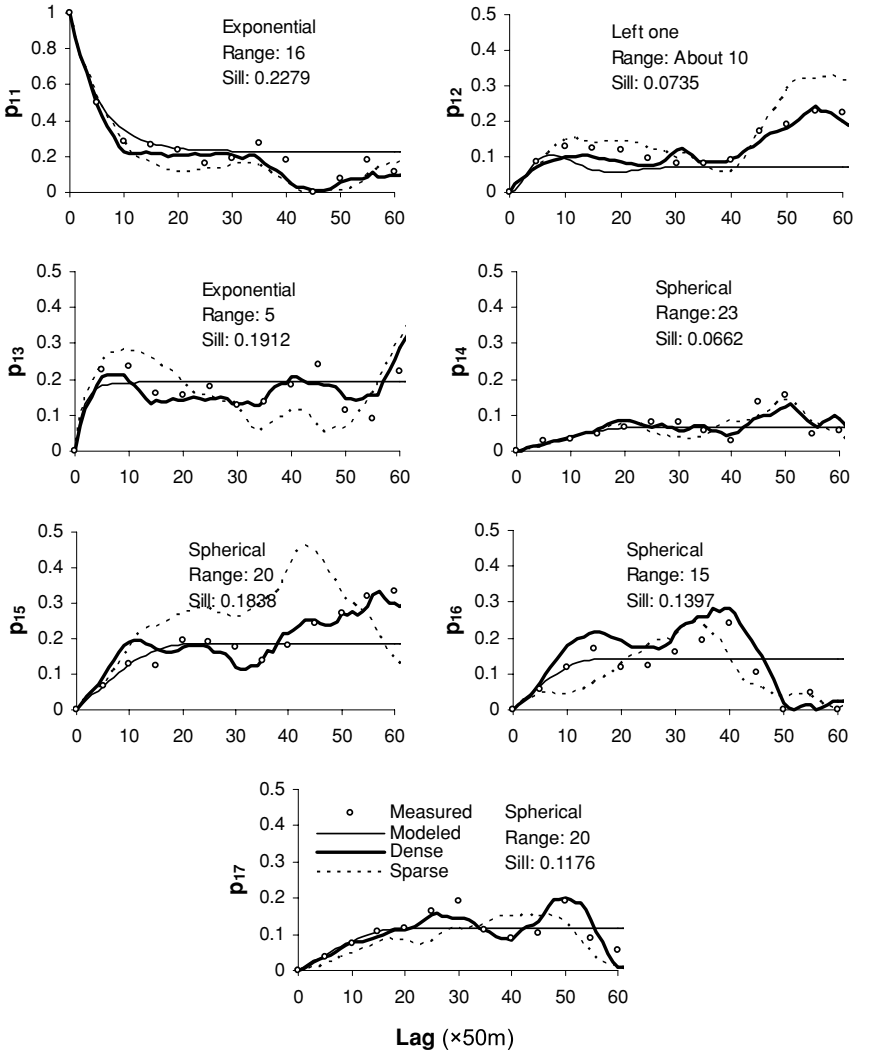

Figure 6. Comparison of simulated transiograms (thick line and dashed line) with experimental transiograms (dots) and input models (thin line) related with the same head class—class 1. Experimental transiograms are estimated from the dense dataset. Simulated transiograms are estimated from the first realizations based on the dense dataset and the sparse dataset, respectively.

the simulations. If transiograms are estimated uni-directionally (e.g., west-to-east, which normally needs a large number of observed data), there will be $p_{17}(h) \neq p_{17}(-h)$ and directional asymmetries of class patterns can be incorporated into a simulation; then simulated results should be better.

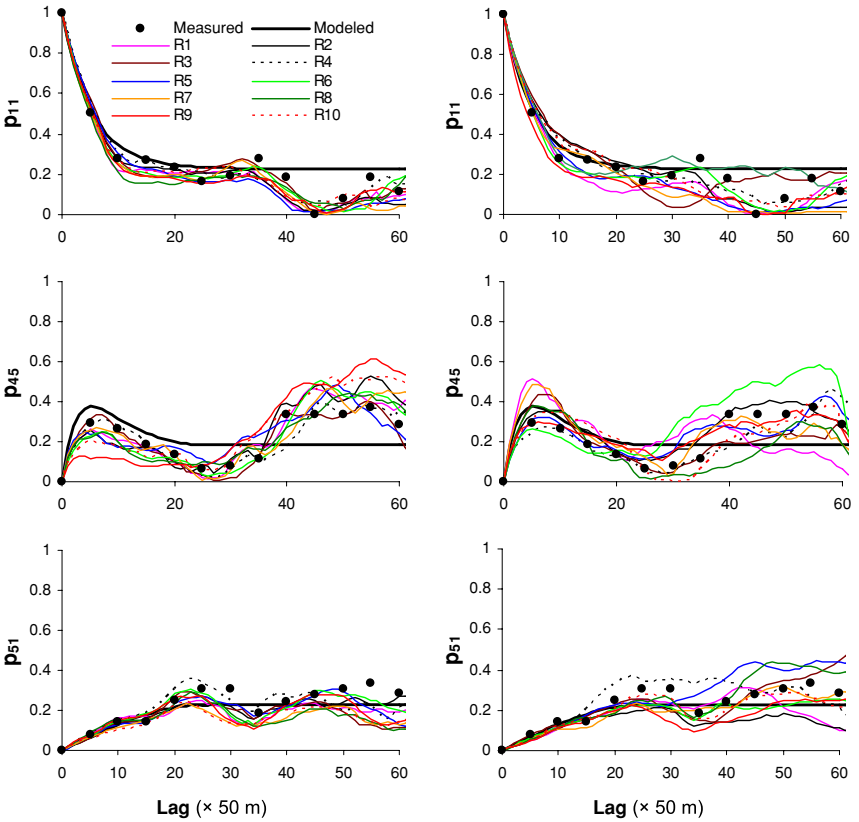

Figure 7. Some auto/cross-transiograms estimated from the first 10 simulated realizations conditioned on the dense dataset (left column) and sparse dataset (right column).

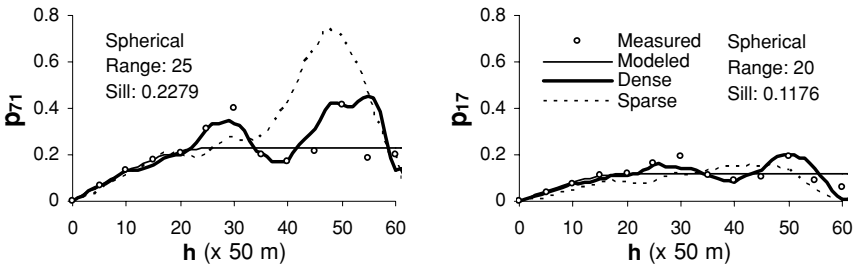

Figure 8. Experimental cross-transiograms $p_{17}(h)$ and $p_{71}(h)$ estimated from the original dense dataset (dots), input transiogram models (thin line), and simulated ones estimated from first simulated realizations based on the two datasets (thick line and dashed line).

CONCLUSIONS

A simple and efficient fixed-path SMC algorithm for stochastic simulation conditioned on point samples has been presented. The MCRF theory provided an explicit solution for the conditional probability function of an unknown location with different nearest-known-neighbor structures. The major power of the algorithm lies with its nonlinear estimator and its capability of incorporating interclass relationships of multinomial classes, which is not well accomplished in conventional geostatistical methods. Therefore, the algorithm is more capable of structure-imitating. Not only can large-scale patterns be produced as polygons, the simulated patterns in different realizations also look similar (but with obvious locational uncertainty reflected in occurrence probability maps). The computational efficiency makes the algorithm well-suited to spatial uncertainty analysis of multinomial classes by generating a large number of realizations and estimating occurrence probability maps.

Using transiogram models to provide transition probabilities at any lags constitutes an advantage of the algorithm in incorporating complex features of spatial variations of discrete variables that a first-order stationary 1-D Markov chain model (e.g., a TPM or a transition rate matrix) cannot capture. But how to effectively model experimental transiograms remains an issue to explore in the future.

The simple algorithm has basically indicated the applicability of the MCRF theory in dealing with discrete variables. More complex and effective algorithms based on this theory may be expected. For example, the data in the off-orthogonal directions may be conditioned in a simulation by using a suitable path. The case study in this paper is just a simple simulation on regular data to demonstrate the practicability and potential of the SMC algorithm and the MCRF theory. The long-term goal is to develop a visual interactive software system to fulfill the mission so that users have a full control on each step of the simulation process.

REFERENCES

Abend, K., Harley, T. J., and Kanal, L. N., 1965, Classification of binary random patterns: IEEE Trans. Info. Theory, v. 11, no. 4, p. 538–544.

Besag, J., 1986, On the statistical analysis of dirty pictures (with discussions): J. Royal Stat. Soc., Ser. B, v. 48, no. 3, p. 259–302.

Carle, S. F., and Fogg, G. E., 1996, Transition probability-based indicator geostatistics: Math. Geol., v. 28, no. 4, p. 453–477.

Carle, S. F., and Fogg, G. E., 1997, Modeling spatial variability with one- and multi-dimensional continuous Markov chains: Math. Geol., v. 29, no. 7, p. 891–918.

Deutsch, C. V., and Journel, A. G., 1998, GSLIB: Geostatistical software library and user's guide 2nd ed.: Oxford University Press, New York, 369 p.

Elfeki, A. M., and Dekking, F. M., 2001, A Markov chain model for subsurface characterization: Theory and applications: Math. Geol., v. 33, no. 5, p. 569–589.

Geman, S., and Geman, D., 1984, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images: IEEE Trans. Pattern Analy. Mach. Intell., v. 6, no. 6, p. 721–741.

Goovaerts, P., 1997, Geostatistics for natural resources evaluation: Oxford University Press, New York, 483 p.

Koltermann, E. C., and Gorelick, S. M., 1996, Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches: Water Resour. Res., v. 32, no. 9, p. 2617–2658.

Krumbein, W. C., 1968, FORTRAN IV computer program for simulation of transgression and regression with continuous time Markov models: Computer Contribution 26, Kansas, Geol. Surv., 38 p.

Li, W., 2006a, Markov chain random fields for estimation of categorical variables: Math. Geol., v. 39, no. 3.

Li, W., 2006b, Transiogram: A spatial relationship measure for categorical data: Inter. J. Geog. Info. Sci., v. 20, no. 6, p. 693–699.

Li, W., Li, B., and Shi, Y., 1999, Markov-chain simulation of soil textural profiles: Geoderma, v. 92, no. 1, p. 37–53.

Li, W., Li, B., Shi, Y., and Tang, D., 1997, Application of the Markov-chain theory to describe spatial distribution of textural layers: Soil Sci., v. 162, no. 9, p. 672–683.

Li, W., Zhang, C., Burt, J. E., Zhu, A. X., and Feyen, J., 2004, Two-dimensional Markov chain simulation of soil type spatial distribution: Soil Sci. Soc. Am. J., v. 68, no. 5, p. 1479–1490.

Lin, C., and Harbaugh, J. W., 1984, Graphic display of two- and three-dimensional Markov computer models in geology: Van Nostrand Reinhold Company, New York, 180 p.

Ma, Y. Z., and Jones, T. A., 2001, Teacher's aide: Modeling hole-effect variograms of lithology-indicator variables: Math. Geol., v. 33, no. 5, p. 631–648.

Norberg, T., Rosen, L., Baran, A., and Baran, S., 2002, On modeling discrete geological structure as Markov random fields: Math. Geol., v. 34, no. 1, p. 63–77.

Pickard, D. K., 1980, Unilateral Markov fields: Adv. Appl. Probab., v. 12, no. 3, p. 655–671.

Potter, P. E., and Blakely, R. F., 1967, Generation of a synthetic vertical profile of a fluvial sandstone body: J. Soc. Petr. Eng. AIME, v. 7, no. 3, p. 243–251.

Qian, W., and Titterington, D. M., 1991, Multidimensional Markov chain models for image textures: J. Royal Stat. Soc, Ser. B, v. 53, no. 3, p. 661–674.

Ritzi, R. W., 2000, Behavior of indicator variograms and transition probabilities in relation to the variance in lengths of hydrofacies: Water Resour. Res., v. 36, no. 11, p. 3375–3381.

Sharp, W. E., and Aroian, L. A., 1985, The generation of multidimensional autoregressive series by the herringbone method: Math. Geol., v. 17, no. 1, p. 67–79.

Tjelmeland, H., and Besag, J., 1998, Markov Random fields with higher-order interactions: Scand. J. Statist. v. 25, no. 3, p. 415–433.

Weissmann, G. S., and Fogg, G. E., 1999, Multi-scale alluvial fan heterogeneity modeled with transition probability geostatistics in a sequence stratigraphic framework: J. Hydrol., v. 226, no. 1, p. 48–65.

Wu, K., Nunan, N., Crawford, J. W., Young, I. M., and Ritz, K., 2004, An efficient Markov chain model for the simulation of heterogeneous soil structure: Soil Sci. Soc. Am. J., v. 68, no. 2, p. 346–351.

Zhang, C., and Li, W., 2005, Markov chain modeling of multinomial land-cover classes: GIScience Remote Sens., v. 42, no. 1, p. 1–18.