Environ Ecol Stat (2008) 15:157–174
DOI 10.1007/s10651-007-0045-9

A single-chain-based multidimensional Markov chain model for subsurface characterization

Weidong Li · Chuanrong Zhang

Received: 19 June 2005 / Revised: 16 October 2006 / Published online: 25 October 2007

Abstract Multidimensional Markov chain models in geosciences were often built on multiple chains, one in each direction, and assumed these 1-D chains to be independent of each other. Thus, unwanted transitions (i.e., transitions of multiple chains to the same location with unequal states) inevitably occur and have to be excluded in estimating the states at unobserved locations. This consequently may result in unreliable estimates, such as underestimation of small classes (i.e., classes with smaller than average areas) in simulated realizations. This paper presents a single-chain-based multidimensional Markov chain model for estimation (i.e., prediction and conditional stochastic simulation) of spatial distribution of subsurface formations with borehole data. The model assumes that a single Markov chain moves in a lattice space, interacting with its nearest known neighbors through different transition probability rules in different cardinal directions. The conditional probability distribution of the Markov chain at the location to be estimated is formulated in an explicit form by following the Bayes' Theorem and the conditional independence of sparse data in cardinal directions. Since no unwanted transitions are involved, the model can estimate all classes fairly. Transiogram models (i.e., 1-D continuous Markov transition probability diagrams) are used to provide transition probability input with needed lags to generalize the model. Therefore, conditional simulation can be conducted directly and efficiently. The model provides an alternative for heterogeneity characterization of subsurface formations.

Keywords Conditional independence · Conditional simulation · Pickard random field · Subsurface formation · Transition probability · Transiogram

W. Li (✉) · C. Zhang
Department of Geography, Kent State University, Kent, OH 44242, USA
e-mail: weidong6616@yahoo.com

1 Introduction

Characterization of subsurface heterogeneity is crucial in many fields of earth science such as soil science and hydrogeology, because subsurface facies (or layers) determine preferential paths of water and solute transport (Koltermann and Gorelick 1996; Weigand et al. 2001). Borehole logs are the major information available for human to understand the spatial distribution of subsurface facies. While providing sufficient knowledge along the vertical one-dimensional (1-D) direction, a single borehole gives little information about the lateral extension of subsurface facies. Because of high cost, the densities of boreholes available from field surveys are normally too sparse for surveyors to delineate reliable spatial distribution maps of subsurface facies in multiple dimensions for many applications.

Subsurface facies normally exhibit features such as sharp boundaries, long-range layers, asymmetric sequences, juxtaposition relationships and interdependencies (i.e., cross-correlations) between different facies types (i.e., classes) (Li et al. 1997; Carle and Fogg 1997). While stochastic simulation conditional to observed data using conventional geostatistics has been recognized as a feasible way for spatial pattern prediction and uncertainty analysis of random variables from sparse samples (Deutsch and Journel 1998), conventional geostatistical methods met difficulties in capturing effectively these features, particularly spatial connectivity, of subsurface formations (Bierkens and Weerts 1994), because of the difficulties in incorporating interclass relationships (Li and Zhang 2005) and the linearity of kriging estimators.

Markov chain models can naturally incorporate interclass relationships through cross transition probabilities. But early multidimensional Markov chain models developed in geosciences (e.g., Krumbein 1968; Lin and Harbaugh 1984) did not condition a simulation on samples. Conditional simulation models based on Markov chains emerged only in recent years (Lou 1996; Elfeki and Dekking 2001; Li et al. 2004). In Lou's model, the estimator is linear and no high-order interaction is considered; consequently, simulated patterns are fragmentary. Estimators in Elfeki and Dekking (2001) and Li et al. (2004) are non-linear with consideration of high-order (i.e., distant) interactions of observed data. The non-linear Markov chain conditional simulation approach is efficient in computation, because conditional transition probabilities for estimating unknown locations are explicitly given (Li and Zhang 2005). However, their simulated results underestimate small classes in simulated realizations because of the exclusion of unwanted transitions in the models.

It is a conventional idea in geosciences that multidimensional Markov chain models for subsurface characterization may be built on multiple 1-D chains, one in each direction, and assumed those 1-D chains to be independent of each other (Carle and Fogg 1997). The coupled Markov chain model suggested by Elfeki and Dekking (2001) for subsurface characterization, although being nonlinear, still followed the similar idea: Two Markov chains—one in the vertical direction and one in the horizontal direction—are assumed to be fully independent of each other and they move to the same location with equal states. Thus, the transitions through which the two chains reach the same location with unequal states become unwanted and have to be excluded in estimating the state of the unobserved location. This exclusion consequently results in unreliable estimates at unobserved locations: for example, small classes (classes or layer types

with smaller than average areas) are underestimated in simulated realizations of the coupled Markov chain model. The triplex Markov chain model suggested by Li et al. (2004) used the idea of coupled Markov chains for calculating conditional probability distributions; therefore, it also has the problem of underestimating small classes [see simulated class proportion data in Li et al. (2004)].

Development of 2-D Markov models in recent years for other applications in earth science such as generating and analyzing thematic remote sensing raster maps (or images) can be seen in Patil and Taillie (2001). This paper introduces a single-chain-based multidimensional Markov chain model for subsurface characterization using borehole data on a regular lattice. By following the Bayes' Theorem and the conditional independence of sparse data in cardinal directions in Pickard random fields (Pickard 1980), an explicit solution is derived without the occurrence of unwanted transitions. Transiogram models (i.e., 1-D continuous Markov transition probability diagrams) are used to provide transition probability input with needed lags. The new model is theoretically sound and its workability is demonstrated by simulated results of a soil vertical transect.

2 Problems and progresses

Consider the case that we have two borehole logs and the ground surface information, which make the three boundaries of a vertical section. The task of simulating subsurface facies in vertical 2-Ds becomes a simulation of unobserved locations between the two boreholes conditional on the three boundaries (Fig. 1). The vertical section is normally discretized into a lattice and the estimation of each unknown pixel x may be conducted by considering the influence of observed data in boundaries. To effectively simulate the facies, the model used needs to meet the following conditions: (1) simulation (or estimation) is conditioned on borehole data and the top layer; (2) interclass relationships are incorporated; (3) the large-scale features and their geometrical shapes (i.e., connectivity) of subsurface facies are captured approximately; (4) all classes are produced fairly in their approximate proportions; and (5) input parameters can be estimated from observed data with expert knowledge. Many existing methods have difficulties meeting all of these conditions. For example, if we use indicator kriging, we usually get dispersed patterns in simulated realizations, which are not effective in representing the spatial connectivity of facies with clear boundaries, and realizations usually do not obey the interclass relationships between different facies because only class auto-correlations are considered in indicator kriging simulation algorithms. Although Markov random field (MRF) models can naturally incorporate interclass relationships, they met other problems.

2.1 Directional effect

Conventional MRF methods (e.g., Besag 1986; Geman and Geman 1984; Norberg et al. 2002) have difficulties accounting for anisotropies and large-scale features in simulations because of the use of small, symmetric neighborhood systems (usually four to eight adjacent neighbors) and iterative updating methods (Qian and Titterington

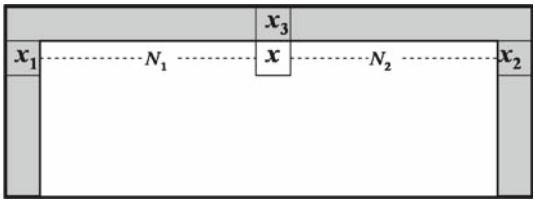

Fig. 1 Illustrating a vertical section, which has three known boundaries—two boreholes (left and right boundaries) and one ground surface (top boundary). A simulation means filling the unknown area with conditioning on the known boundaries. For an unknown pixel x, we may consider only the influence of known pixels x_1, x_2 and x_3. Here N_1 and N_2 represent the distances from x_1 to x and from x_2 to x, respectively

1991; Tjelmeland and Besag 1998). Markov mesh models (e.g., Abend et al. 1965; Qian and Titterington 1991) were proposed for image structure analysis and might generate large-scale features by considering interactions between an unknown pixel and its close predecessors (i.e., upper pixels and left pixels) (see explanation in Besag 1986). However, not only do they consider no distant interactions between an unknown pixel and its surrounding sample data, when producing large-scale features they generate undesirable *directional effect* (i.e., layer inclination along the simulation direction) as specifically demonstrated by Gray et al. (1994). This directional effect problem was thought to be intrinsic because of the inherent asymmetry of neighborhood systems used in Markov mesh models, as noticed and criticized by Besag (1986), Qian and Titterington (1991), and Tjelmeland and Besag (1998). Both MRF and Markov mesh methods use cliques (structures of pixels adjacent of each other) and fixed neighborhood systems, which makes it difficult to use them for direct conditional simulation on sample data (i.e., generate a realization conditioned on observed data by one pass rather than iterative updating).

The coupled Markov chain (CMC) model of Elfeki and Dekking (2001) is similar to a simple Markov mesh model but with an extra *full independence* assumption (i.e., the coupled two 1-D Markov chains in the vertical and horizontal directions are assumed independent of each other). With this assumption, only 1-D transition probabilities are necessary in simulating subsurface layers with borehole data. The CMC model could condition a simulation on borehole data through multiple-step transition probabilities. However, given the inherent asymmetry of the neighborhood, the CMC model cannot avoid the directional effect problem when simulating large-scale features, as mentioned by Li et al. (2004). To deal with the problem, Li et al. (2004) suggested the triplex Markov chain (TMC) model with an "alternate advancing" path (AA) by importing a third 1-D Markov chain. The TMC model was previously called a tripled Markov chain model, which also used the full independence assumption but with three single Markov chains for conditional simulation of subsurface formations through a "middle insertion" (MI) path. Although using these special paths does not effectively solve the underestimation problem of small classes, the undesirable directional effect problem has indeed been resolved.

The MI path means that the current location to be estimated always falls in the middle of two known pixels in a row. The AA path means that the Markov chain

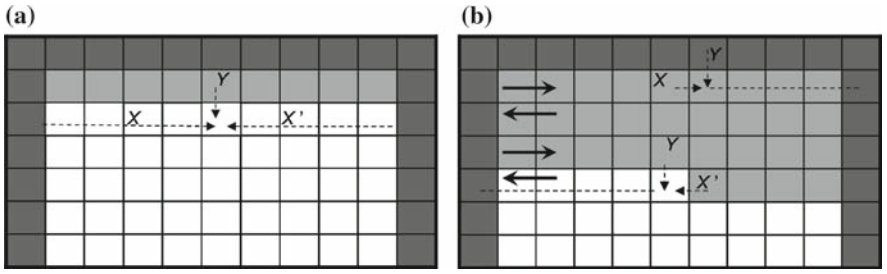

Fig. 2 Illustration of the two paths for the tripled and triplex Markov chain models when used for shallow subsurfacet characterization. (**a**) The middle insertion (MI) path. (**b**) The alternate advancing (AA) path. Thick solid arrows represent simulation proceeding directions. Dash arrows represent the directions of 1-D Markov chains and point to the cell to be determined. Dash lines show the conditioning on a remote known cell. Dark gray cells represent observed data. Light gray cells represent already simulated data

proceeds alternately in opposite directions row by row. The novelty of these two special paths is that they make the simulated predecessors distributed at both (left and right) sides of the unknown pixel to be estimated, making the simulation process symmetric or quasi-symmetric in the x-direction (i.e., the lateral direction) and therefore imposing balanced influences of known (observed or previously simulated) data to the unknown pixel to be estimated (Fig. 2). The development of these two paths implies that the undesirable directional effect problem in Markov chain (or mesh) modeling is solvable. In fact, Wu et al. (2004) implicitly used the AA path in reconstruction of soil microsection images using their Markov mesh model and therefore avoided the directional effect. Both paths are recommended for the model to be proposed in this paper. We later also found that a similar path was used more early by Sharp and Aroian (1985) in the herringbone method for dealing with a similar problem in extending autoregressive processes into multiple dimensions.

2.2 Underestimation of small classes

Underestimation of small classes in simulated realizations results from the full independence assumption of multiple chains in the CMC model and the TMC model. Such an assumption is theoretically unproved in a Markov random field environment. With the full independence assumption, the exclusion of unwanted transitions is inevitable (see Elfeki and Dekking 2001, p. 575). With unwanted transitions being excluded, as long as the current location of the Markov chain departs from observed data, underestimation of small classes (correspondingly overestimation of large classes) begins to occur and strengthens with increasing departure distance. The problem reaches its peak in unconditional simulations, usually causing small classes to disappear. But the problem weakens with increasing density of conditioning data (see Li et al. 2004; Li and Zhang 2005).

This is also why the tripled Markov chain model was changed to the TMC model to accommodate the idea of "*conditioning to future state*" in Elfeki and Dekking (2001), because the requirement of three Markov chains moving to the same pixel with equal states in the tripled Markov chain model implies more exclusion of unwanted transi-

tions. The TMC model uses two CMCs with an AA path; therefore, it does not make the small-class underestimation problem severer. On the contrary, by extending the CMC to condition on data in both x and y directions and by balancing the influences of conditioning data on both left and right sides of the unknown location to be estimated, the problem is largely eased. For the small-class underestimation problem, see simulated results and discussions in Li et al. (2004).

Since exclusion of unwanted transitions is inevitable with the full independence assumption, the only way to eliminate this problem is not to use this assumption. This means we have to find a new way to build a theoretically sounder multidimensional Markov chain model. This will be the main task of this paper.

2.3 Estimating transition probabilities via transiograms

One-step transition probability matrices (TPMs) were used as parameter inputs in Elfeki and Dekking (2001) and Li et al. (2004) for conducting simulation, where multi-step transition probabilities were calculated from one-step TPMs based on the first-order stationary Markovian assumption. While that represents a way to obtain multi-step transition probabilities, in practical use, however, one-step TPMs are difficult to estimate from sparse data. In subsurface lateral directions, there is actually no way to estimate one-step TPMs directly from borehole data. In vertical directions, if the observation interval is larger than pixel size used in a simulation, there is also no way to estimate one-step TPMs directly from discontinuous borehole logs. In addition, multi-step transition probabilities obtained by this way cannot capture effectively the complex spatial heterogeneity of classes because the spatial distributions of natural phenomena are normally not first-order Markovian.

Recently, Li (2006) proposed the transiogram concept as a new two-point spatial relationship measure for categorical data. A transiogram is defined as a 1-D two-point conditional probability (i.e., transition probability) function as

$$p_{ij}(h) = \Pr(Z(x+h) = j | Z(x) = i) \tag{1}$$

where h represents the distance lag, i and j represent states and $p_{ij}(h)$ represents a transition probability from state i to state j over the lag h. Here the random variable Z is assumed spatially stationary, that is, the transiogram $p_{ij}(h)$ is only dependent on the lag h and not dependent on the specific locations x. Therefore, transiograms represent a two-point spatial continuity/discontinuity measure, similar to indicator variograms. Also similar to auto-variograms and cross-variograms, we can call $P_{ii}(h)$ *auto-transiograms* and $p_{ij}(h)(i \neq j)$ *cross-transiograms*. Auto-transiograms represent auto-correlations of single classes and cross-transiograms represent inter-class relationships between different classes.

Transiograms may be directly estimated from data or indirectly calculated from a one-step TPM. Transiograms calculated from one-step TPMs based on the first-order Markovian assumption are called *idealized transiograms*, because they exhibit very smooth curves and have quantitative relations with class proportions and class polygon mean sizes. If available, they are helpful for interpreting observed (or real)

transiograms, which are normally complex. The features of observed transiograms that were not captured by idealized transiograms may be regarded as the high-order Markovian (or non-Markovian) effect of data. Therefore, idealized transiograms are not recommended to be used in simulations because of their oversimplification and the difficulty in acquiring one-step TPMs from sparse samples.

We may directly estimate a limited number of multi-step transition probabilities with different lags (called *experimental transiogram*) from sparse data and then model the experimental transiogram with mathematical models. The method of using mathematical models to fit experimental transiograms was introduced in Li and Zhang (2006). When experimental transiograms are reliable, we even can simply interpolate experimental transiograms into continuous models for use in Markov chain simulation (Li and Zhang 2005). Considering the advantages of transiograms over one-step TPMs, the transiogram is defined as a component of the proposed Markov chain model in this study.

3 Conditional independence of sparse data

The conditional independence assumption states that given a pixel (or an event) its surrounding adjacent pixels (or events) are conditionally independent. The applications of this assumption in lattice data were usually limited to adjacent neighbors in cardinal (or axis) directions.

In fact, if only considering cardinal directions, this assumption holds in Pickard random fields (PRF) (Pickard 1980). Pickard (1980) proved the existence of a curious unilateral MRF, which has a special property that for a generic clique $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$, given any pixel D among A, B, C and D within the clique its two diagonally adjacent pixels B and C are conditionally independent. So in a PRF, we have

$$\Pr(B|D, C) = \Pr(B|D) \tag{2}$$

Based on this property, Pickard presented his model for binary processes.

This model was adapted by Haslett (1985), Idier et al. (2001), Fjortoft et al. (2003), and others for image processing, where they assumed that given a pixel x all its four adjacent pixels y_1, y_2, y_3 and y_4 (i.e., the upper, left, right, and underlying adjacent pixels) in a neighborhood like

$$\begin{pmatrix} & y_1 & \\ y_2 & x & y_3 \\ & y_4 & \end{pmatrix} \tag{3}$$

are conditionally independent. This adaptation is straightforward and it is easy to see this is a PRF.

Pickard's Theorem is usually used for interactions of adjacent pixels or pixel blocks in image processing. Note that the pixels in the generic clique $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ need not be single pixels; they can be pixel blocks, as used in Derin et al. (1984) and Rosholm

(1997). Here we want to apply the conditional independence to sparse data for high-order (or remote) interaction.

Assume we have a general sparse-data structure in a PRF as

$$\begin{pmatrix} & x_1 & \\ & \vdots & \\ x_2 \cdots & x & \cdots x_3 \\ & \vdots & \\ & x_4 & \end{pmatrix} \tag{4}$$

where x_1, x_2, x_3 and x_4 are four nearest *known* neighbors of x along the cardinal directions on a regular lattice. They may be distant from the pixel x with the distance N_1, N_2, N_3 and N_4 (numbers of steps or pixels), respectively. Given the pixel x, are they still conditional independent as the case of adjacent pixels in expression (3)? Indeed, they still can be conditionally independent.

To prove this, let's use E, F, G and H to represent pixel blocks of $\begin{pmatrix} x_1 \\ \vdots \end{pmatrix}$, $\begin{pmatrix} x_2 \cdots \end{pmatrix}$, $\begin{pmatrix} \cdots x_3 \end{pmatrix}$ and $\begin{pmatrix} \vdots \\ x_4 \end{pmatrix}$, respectively, we immediately have that given x the four pixel blocks of E, F, G and H in a PRF are conditionally independent, that is, for any pixel block of E we have

$$\Pr(E|x, F, G, H) = \Pr(E|x) \tag{5}$$

Since only one pixel is known in each of these pixel blocks, from Eq. 5 we further have

$$\Pr(x_1|x, x_2, x_3, x_4) = \Pr(x_1|x) \tag{6}$$

Thus, we have shown that the conditional independence of sparse data with high-order interactions along cardinal directions in a PRF is correct. That is, within a PRF, Eq. 6 holds. Equation 6 can be directly extended to the 3-D situation.

Further relaxing (6) to include data in non-cardinal directions is feasible only in some situations; for example, when nearest known neighbors in different directions are distributed symmetrically, such a relaxation should work. So we can generally write the conditional independence assumption as

$$\Pr(x_i|x, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = \Pr(x_i|x) \tag{7}$$

4 The Markov chain model

To avoid unwanted transitions, we have to assume that one space is only described by one single Markov chain. Here we assume that there is only one Markov chain

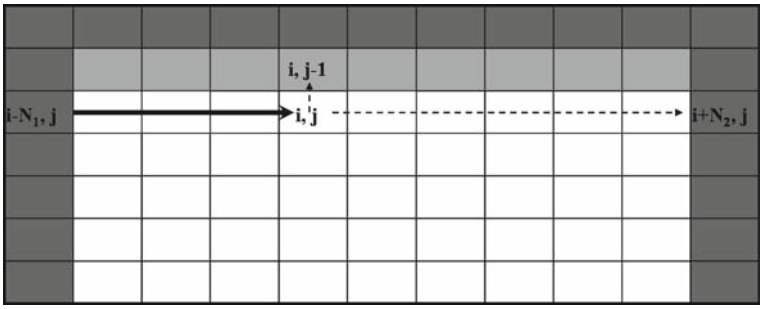

Fig. 3 Illustration of the proposed single-chain-based Markov chain model with conditioning on two boreholes. In simulation, it needs to follow the AA path or MI path. Arrows represent the interactions between known pixels and the unknown pixel to be estimated as well as the directions of transition probabilities. Dark gray cells represent observed data. Light gray cells represent already simulated data

moving or jumping in a 2-D space. The Markov chain interacts with its nearest known neighbors (observed or previously simulated data) in cardinal directions by the same or different transition probability rules. Such a Markov chain random field with the property (6) will be a special PRF.

Let's go back to the subsurface vertical section case as shown in Fig. 1. With three boundaries (left, right, and top) being known, our simulation will be conducted from top to bottom row by row. For each lattice cell (or pixel) x, it has three nearest known neighbors x_1, x_2 and x_3 in cardinal directions at its left, right, and top. These nearest known neighbors may be remote or adjacent. We assume that a Markov chain moves from x_1 to the unknown location x and interacts with x_1, x_2 and x_3. With the Bayes' Theorem and Eq. 6 we have the conditional probability distribution of the Markov chain at x factorized as follows:

$$\begin{aligned}\Pr(x|x_1,x_2,x_3) &= \frac{\Pr(x,x_1,x_2,x_3)}{\Pr(x_1,x_2,x_3)}\\ &= \Pr(x_3|x,x_1,x_2)\cdot\Pr(x_2|x,x_1)\cdot\Pr(x|x_1)\cdot\frac{\Pr(x_1)}{\Pr(x_1,x_2,x_3)}\\ &= C\cdot\Pr(x_3|x)\cdot\Pr(x_2|x)\cdot\Pr(x|x_1)\end{aligned} \tag{8}$$

where C is a constant because it does not involve x. C is given as

$$C = \frac{\Pr(x_1)}{\Pr(x_1,x_2,x_3)} \tag{9}$$

Rewrite (8) more concisely as

$$\Pr(x|x_1,x_2,x_3) = C\cdot\Pr(x_3|x)\cdot\Pr(x_2|x)\cdot\Pr(x|x_1) \tag{10}$$

Now we use $Z_{i,j} = k$, $Z_{i-N_1,j} = l$, $Z_{i+N_2,j} = q$, and $Z_{i,j-1} = m$ (see Fig. 3) to replace location labels x, x_1, x_2 and x_3, and use the transition probability formulism,

we have Eq. 10 rewritten as

$$\begin{aligned}\Pr(Z_{i,j} = k | Z_{i-N_1,j} = l, Z_{i,j-1} = m, Z_{i+N_2,j} = q) \\ = C \cdot p^h_{kq}(N_2) \cdot p^{v'}_{km}(1) \cdot p^h_{lk}(N_1)\end{aligned} \tag{11}$$

where superscripts h and v' represent directions of transition probabilities along axes (i.e., horizontally from left to right and vertically from bottom to top, respectively), not powers; N_1, and N_2 represent distances (i.e., number of spatial steps) between cell $(i - N_1, j)$ and cell (i, j) and between cell $(i - N_2, j)$ and cell (i, j), respectively; $p(N)$ represents a N-step transition probability, and k, l, m and q represent states. Renormalizing the right hand side of Eq. 11 to cancel the constant C, we have

$$\begin{aligned}\Pr(Z_{i,j} = k | Z_{i-N_1,j} = l, Z_{i,j-1} = m, Z_{i+N_2,j} = q) \\ = \frac{p^h_{kq}(N_2) \cdot p^{v'}_{km}(1) \cdot p^h_{lk}(N_1)}{\sum_{f=1}^{n} \left[p^h_{fq}(N_2) \cdot p^{v'}_{fm}(1) \cdot p^h_{lf}(N_1) \right]}\end{aligned} \tag{12}$$

Here it can be seen that no transitions are excluded in deriving Eq. 12 because there is only one single Markov chain and no unwanted transitions need be excluded.

The model of Eq. 12 is the 2-D Markov chain model we can use for subsurface characterization in the vertical two dimensions. If we use the MI path, we fix $N_1 = N_2$ (or $N_1 = N_2 \pm 1$ when the distance between two nearest known data on a row is even). If we use the AA path as the TMC model did in Li et al. (2004), we fix $N_1 = 1$. Anyway, (i) we can eliminate the directional effect by using the novel paths aforementioned and (ii) we have eliminated the root for underestimation of small classes because there are no unwanted transitions excluded.

In case the upper boundary is unknown, it can be simulated using a 1-D conditional Markov chain model. Such a 1-D model can be simplified from Eq. 12 by deleting the terms related with the upper cell. We have it as

$$\Pr(Z_i = k | Z_{i-N_1} = l, Z_{i+N_2} = q) = \frac{p_{kq}(N_2) \cdot p_{lk}(N_1)}{\sum_{f=1}^{n} \left[p_{fq}(N_2) \cdot p_{lf}(N_1) \right]} \tag{13}$$

Equation 12 appears to be similar to Eq. 21 in Elfeki and Dekking (2001), but they are actually quite different. The final equation of the CMC model (Elfeki and Dekking, 2001, Eq. 21) can be written as

$$\begin{aligned}\Pr(Z_{i,j} = k | Z_{i-1,j} = l, Z_{i,j-1} = m, Z_{i+N,j} = q) \\ = \frac{p^h_{kq}(N) \cdot p^v_{mk}(1) \cdot p^h_{lk}(1)}{\sum_{f=1}^{n} \left[p^h_{fq}(N) \cdot p^v_{mf}(1) \cdot p^h_{lf}(1) \right]}\end{aligned} \tag{14}$$

where v represents the vertical direction from top to bottom. The terms p^v_{mk} and p^v_{mf} in (14) are completely different transition probabilities compared with the terms $p^{v'}_{km}$ and $p^{v'}_{fm}$ in (12). They have different transition directions and also are different elements in a TPM. They can be equal unless all classes account for equal proportions and the Markov chain is reversible (e.g., soil layer sequences in soil profiles are vertically symmetric), which are rarely true in geosciences.

Simply following the procedure from (8) to (11) and applying the conditional independence assumption (7), we can get a general solution about the conditional probability distribution of a Markov chain at an unknown location with m nearest known neighbors in different directions:

$$\Pr(X(x) = k | X_1(x_1) = l_1, \ldots, X(x_m) = l_m) = \frac{\prod_{i=2}^{m} p^i_{kl_i}(h_i) \cdot p^1_{l_1 k}(h_1)}{\sum_{f=1}^{n} \left[\prod_{i=2}^{m} p^i_{fl_i}(h_i) \cdot p^1_{l_1 f}(h_1) \right]} \tag{15}$$

where, p^i represents the transition probability in the ith direction; x_1 represents the neighbor from which the Markov chain moves to the current location x; m represents the number of nearest known neighbors; k, l_i, and f all represent states in the state space $S = (1, \ldots, n)$; h_i is the distance from the current location to the nearest known neighbor x_i. h_i can be numbers of pixels or exact distance measures such as meter. Note that for any nearest known neighbor x_i, if h_i is large enough (beyond related correlation ranges), both $p^i_{kl_i}(h_i)$ and $p^i_{fl_i}(h_i)$ will approach p_{l_i} (the stable probability) and they can be canceled from the numerator and denominator.

Equation 15 represents a general solution, which also serves as the general solution of Markov chain random fields (Li 2007). This general solution includes the solutions for cases of only considering nearest known neighbors in cardinal directions in 1-D to 3-D space. In this paper, we will not test other solutions by simulation except for (12).

5 Parameter estimation

Both one-step transition probabilities and multi-step transition probabilities are needed for conducting simulation using the 2-D Markov chain model. These transition probabilities with different numbers of spatial steps can be drawn from transiograms $p_{ij}(h)$ in different directions.

Experimental transiograms can be directly estimated from observed data. But when data are too few, experimental transiograms may not be reliable. A good practice is to apply expert knowledge and experimental transiograms together to infer transiogram models for Markov chain simulation. Thus the practical way includes two steps: (i) First estimating a dozen of transition probabilities with different lags from sample data (i.e., experimental transiograms), and (ii) then fitting the estimated data with mathematical models and expert knowledge. Even if there is no data available in some

directions, by assessing the three determinant parameters of a transiogram model—sill, correlation range, and transiogram shape (e.g., exponential or spherical) from expert knowledge acquired from other ways, a transiogram model may be acquired. This is also the normal practice used in estimating variograms. With model fitting, transition probabilities at any lag can be acquired.

To guarantee that at any lag h all transition probabilities involving the same head class (i.e., class i in $p_{ij}(h)$) sum to 1, we need always leave one transiogram (e.g., $p_{ik}(h)$) not model-fitted and infer it by

$$p_{ik}(h) = 1 - \sum_{\substack{j=1 \\ j \neq k}}^{n} p_{ij}(h), \tag{16}$$

where n is the number of classes (Li and Zhang 2006). Otherwise, the constraint condition of summing to 1 may be easily violated in the processes of model fitting. To guarantee that $p_{ik}(h)$ be non-negative and well-fitted with the experimental transiogram, the model fitting process of other transiograms may need repetitive tuning. Our experiences suggest that this is generally easy to do.

Figure 4 shows a subset of the 16 experimental auto/cross-transiograms in a lateral direction (west-to-east) estimated from 20 boreholes of alluvial soils in a 2-m depth with equal intervals and fitted models using exponential and spherical curves. Figure 5 shows a subset of experimental auto/cross-transiograms and fitted models in a vertical direction (bottom-to-top). It should be noticed that the effective fitting of the first several points of experimental transiograms is important because these low-lag points are more reliable than others and it is the low-lag section of transiogram models that is used in simulations. In vertical directions, the Markov chain model needs only one-step transition probabilities in the bottom-to-top direction, which can be drawn from these transiogram models. One-step transition probabilities in vertical directions may also be directly estimated from borehole logs if the observation interval in borehole logs is the same as the vertical pixel size used in simulation.

There are clear relations between $p_{ij}(h)$ and proportions of individual classes. The sill of $p_{ij}(h)$ is theoretically equal to the proportion p_j of the tail class j. This is because with increasing lag h, $p_{ij}(h)$ will approach the proportion of class j in a stationary and irreducible Markov chain process. The equality between sills of transiograms and proportions of tail classes (i.e., class j in $p_{ij}(h)$) provides a guidance for model fitting of experimental transiograms, because proportions can be approximately estimated from observed borehole data. It can be seen from Figs. 4 and 5 that sills of some experimental transiograms (e.g., cross-transiogram $p_{12}(h)$, represented by p(1, 2) in Fig. 4) apparently deviate from the proportions of the tail classes (e.g., the estimated proportion of layer type 2 is 0.28. Note that sills of all fitted models are set to the proportions of corresponding tail classes). This is because of the boundary effect (Li 2006) and the nonstationary distribution of classes in a small domain (e.g., Layer type 2 is mainly located at the upper part). For such situations, it is proper to set these fitted transiogram models to the expected sills. But if the simulation domain (i.e., the length and width) is sufficiently large, there normally has no such deviation.

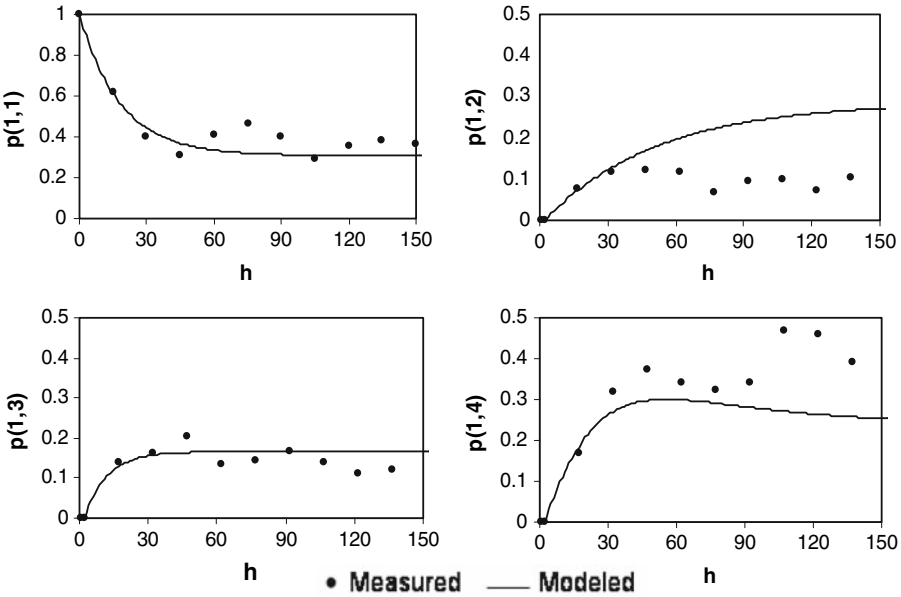

Fig. 4 Experimental auto/cross-transiograms in a lateral direction (west-to-east) estimated from 20 borehole logs and fitting models using exponential and spherical distribution curves (sills are set to correct proportions of tail classes). Boreholes are 2-m deep and located along a transect with an equal interval of 15 pixels (about 254-m)

6 Simulation example

Below a simulation example is given to show the feasibility of the 2-D Markov chain model in Eq. 12. The example is a simulation of a vertical section of alluvial soils with four different layer types — sand, sandy loam, loam, and clay (labeled with 1, 2, 3, and 4, respectively). The vertical section has a 5,250-m length and a 2-m depth. A reference map of the vertical section is hand-delineated based on a large number of densely distributed borehole logs, which are normally not available in soil survey. Such a map does not represent the absolute truth. It is just an expert interpretation based on dense data and may represent an expected result. Exhaustive information of the reference map is not used for simulation. It is provided only for the purpose of a visual comparison with simulated results. Considering the density of boreholes is normally low in soil survey, in the following simulations only a small portion of boreholes are used as conditioned data. These borehole logs are continuous observation data.

The vertical section is randomly discretized into a lattice of 310 columns and 43 rows (so that images can be easily displayed). 20 borehole logs with approximately equal intervals are used for transiogram estimation. Some of transiograms used are already shown in Figs. 4 and 5. The simulation will draw needed transition probabilities from those transiogram models. In this simulation, simulated results using the AA path, which are close to the simulated results using the MI path, will be shown. Simulation

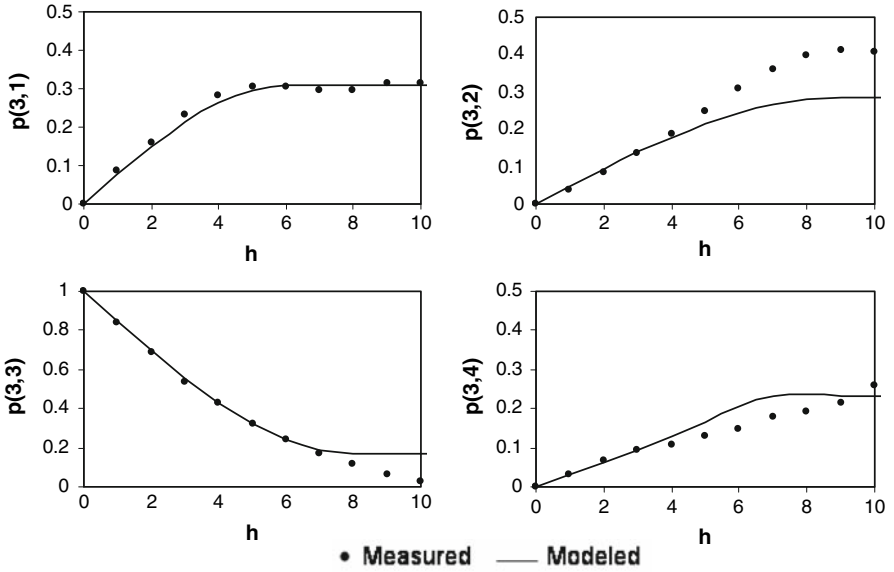

Fig. 5 Experimental auto/cross-transiograms in a vertical direction (bottom-to-top) estimated from 20 borehole logs and fitting models using exponential and spherical distribution curves (sills are set to correct proportions of tail classes). Boreholes are 2-m deep and located along a transect with an equal interval of 15 pixels (about 254-m). Vertical observational interval is 5-cm. Only one-step transition probabilities (close to the origin point (0, 0) for cross-transiograms or (1, 0) for auto-transiograms) in this vertical direction are used in the proposed model

will be conducted by conditioning on different numbers of borehole logs so that we can see whether or not all layer types are effectively represented in simulated realizations.

Hundred realizations are generated and probability maps are estimated from them. Figure 6 shows simulated results conditioned on 11 boreholes. The borehole interval is 525-m (or 31 data columns), which is a quite normal interval used in a soil survey. Results show that both simulated realizations and the prediction map mimic the original map to a large extent. The occurrence probability maps display clearly the distribution of spatial uncertainties of occurrence of different layer types. The maximum occurrence probability map shows where larger uncertainties (lighter color) occur: interestingly, they occur where the boundaries of simulated layers are most probably located.

When increasing the number of boreholes to 22, that is, decreasing the borehole interval to about 254-m (or 15 data column), which is still an acceptable interval in a detailed soil survey, simulated realizations mimic the reference map very closely (Fig. 7). At the same time, the spatial uncertainties in simulated results become small (see the occurrence probability maps). The maximum occurrence probability map displays the uncertainty of layer boundaries.

Important features that should be noticed are that the directional asymmetry of layer occurrence sequence in the vertical direction is well kept and the layer boundaries are clear in prediction maps based on maximum occurrence probabilities and all single realizations no matter how many boreholes are conditioned. These features should

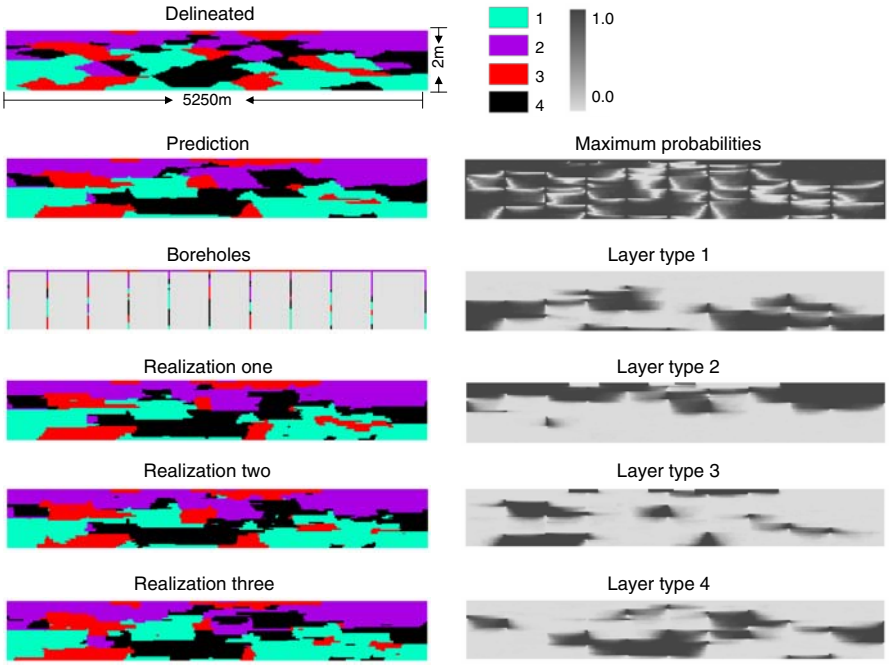

Fig. 6 Simulated results with conditioning on 11 boreholes (about 525-m interval) and the top boundary. The top map is a reference map of the alluvial soil transect (5,250-m long and 2-m deep) with four types of textural layers (it is delineated based on a large number of boreholes record, provided here just for a visual comparison). The prediction map is based on the maximum occurrence probabilities estimated from 100 realizations. Occurrence probability maps of individual soil layer types are estimated from 100 realizations

be attributed to the unidirectional characteristic of transiograms, the incorporation of interclass relationships described by cross-transiograms, and the non-linearity of the model. The simulated patterns are not necessarily exactly the same as that in the reference map, because simulation is based on the limited information conveyed by the conditioned borehole data and the input transiogram models. Uncertainty also normally exists in the transiogram estimation and modeling process (i.e., different modelers may model them with some differences).

Figure 8 provides the proportions of the four layer types in the reference map (serving as expected proportions), borehole data, and simulated realizations. Simulated proportions are averaged from 100 realizations. Though a little deviation between simulated and expected proportion data can be found, apparently all layer types are fairly generated.

7 Conclusions

In this paper a single-chain-based 2-D Markov chain model is proposed for conditional simulation of subsurface formations. The model has only a single Markov chain, which

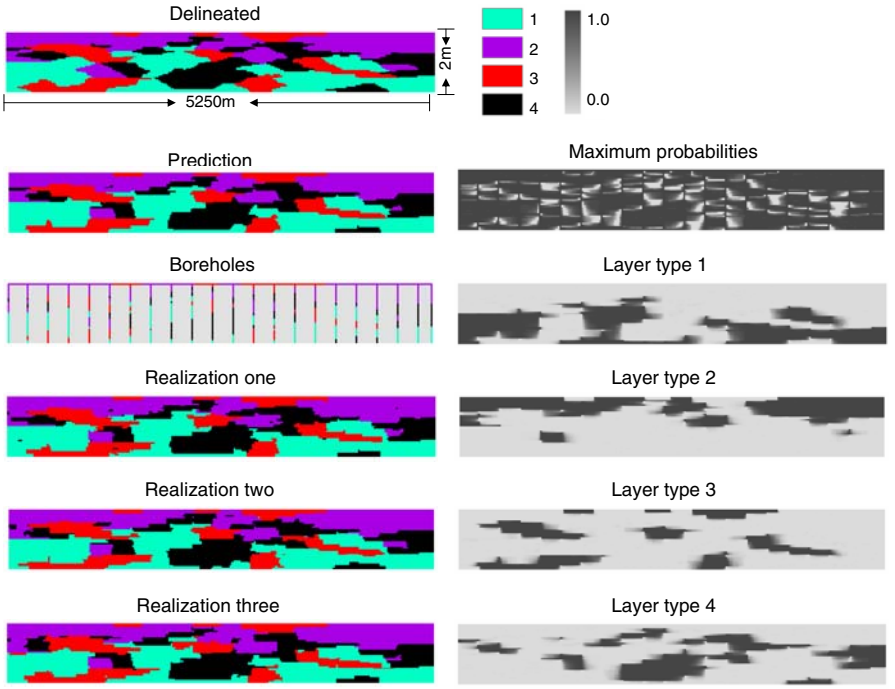

Fig. 7 Simulated results with conditioning on 22 boreholes (about 254-m intervals) and the top boundary

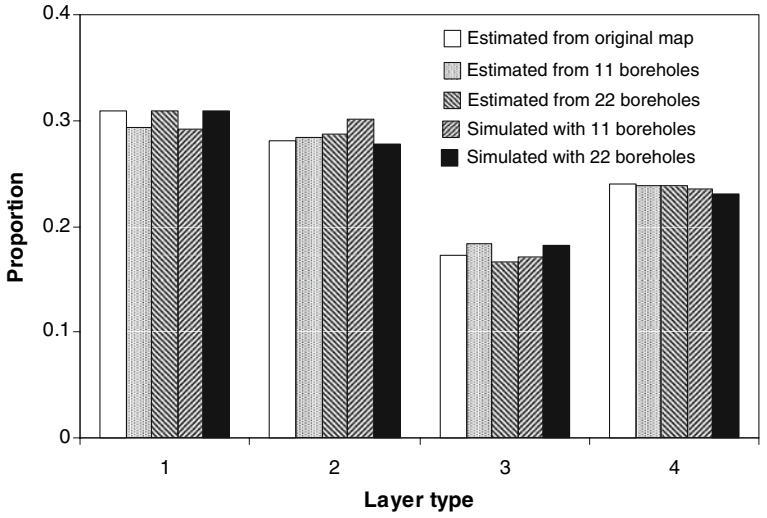

Fig. 8 Proportions of the four layer types estimated from the original map, borehole data, and 100 simulated realizations in each simulation

moves in the space and interacts with its nearest known neighbors in different cardinal directions. With the justification that the conditional independence of sparse spatial data in cardinal directions holds in a PRF, the conditional probability distribution of an unknown location to be estimated is obtained explicitly.

The model avoids unwanted transitions occurred in previously proposed non-linear 2-D Markov chain conditional simulation models (i.e., the CMC and the TMC models). Thus, both small and large classes can be estimated fairly. Experimental transiograms were estimated from limited observed data and then fitted by models, which further provided transition probabilities at any lags needed in simulations.

A simulation example showed that the model worked well with different numbers of borehole logs serving as conditioning data. Because of its efficiency in computation, the model is suitable for performing spatial uncertainty analysis of subsurface formations by generating a large number of realizations and conducting probability mapping. The model may be further extended for 3-D simulation.

Acknowledgments Insightful comments and suggestions from the Editor-in-Chief, Dr. G. P. Patil and three anonymous reviewers, which largely improved this paper, are greatly appreciated. The authors would like to thank Prof. D. M. Titterington at University of Glasgow for examining the Markov chain model presented in this paper.

References

Abend K, Harley TJ, Kanal LN (1965) Classification of binary random patterns. IEEE Trans Inf Theory 11:538–544

Besag J (1986) On the statistical analysis of dirty pictures (with discussions). J Roy Stat Soc Ser B 48:259–302

Bierkens MFP, Weerts HJT (1994) Application of indicator simulation to modelling the lithological properties of a complex confining layer. Geoderma 62:265–284

Carle SF, Fogg GE (1997) Modeling spatial variability with one- and multi-dimensional continuous Markov chains. Math Geol 29:891–918

Derin H, Elliott H, Cristi R, Geman D (1984) Bayes smoothing algorithms for segmentation of binary images modeled by Markov random fields. IEEE Trans Pattern Anal Mach Intell 6:707–720

Deutsch CV, Journel AG (1998) GSLIB: geostatistics software library and user's guide, 2nd edn. Oxford University Press, New York

Elfeki AM, Dekking FM (2001) A Markov chain model for subsurface characterization: theory and applications. Math Geol 33:569–589

Fjortoft R, Delignon Y, Pieczynski W, Sigelle M, Tupin F (2003) Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields. IEEE Trans Geosci Remote Sens 41:675–686

Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell PAMI-6:721–741

Gray AJ, Kay IW, Titterington DM (1994) An empirical study of the simulation of various models used for images. IEEE Trans Pattern Anal Mach Intell 16:507–513

Haslett J (1985) Maximum likelihood discriminant analysis on the plane using a Markovian model of spatial context. Pattern Recognit 18:287–296

Idier J, Goussard Y, Ridolfi A (2001) Unsupervised image segmentation using a telegraph parameterization of Pickard random fields. In: Moore M (ed) Spatial statistics: methodological aspects and some applications, Lecture Notes in Statistics, vol 159. Springer Verlag, New York, pp 115–140

Koltermann EC, Gorelick SM (1996) Heterogeneity in sedimentary deposits: a review of structure-imitating, process-imitating, and descriptive approaches. Water Resour Res 32:2617–2658

Krumbein WC (1968) FORTRAN IV computer program for simulation of transgression and regression with continuous time Markov models, Computer Contribution 26, Kansas Geology Survey

Li W (2006) Transiogram: a spatial relationship measure for categorical data. Int J Geogr Inf Sci 20:693–699
Li W (2007) Markov chain random fields for estimation of categorical variables. Math Geol 39:321–335
Li W, Li B, Shi Y, Tang D (1997) Application of the Markov-chain theory to describe spatial distribution of textural layers. Soil Sci 162:672–683
Li W, Zhang C (2005) Application of transiograms to Markov chain simulation and spatial uncertainty assessment of land-cover classes. GISci Remote Sens 42:297–319
Li W, Zhang C (2006) A generalized Markov chain approach for conditional simulation of categorical variables from grid samples. Trans GIS 10:651–669
Li W, Zhang C, Burt JE, Zhu AX, Feyen J (2004) Two-dimensional Markov chain simulation of soil type spatial distribution. Soil Sci Soc Am J 68:1479–1490
Lin C, Harbaugh JW (1984) Graphic display of two- and three-dimensional Markov computer models in geology. Van Nostrand Reinhold Company, New York
Lou J (1996) Transition probability approach to statistical analysis of spatial qualitative variables in geology. In: Forster A, Merriam DF (eds) Geologic modeling and mapping. Plenum Press, New York, pp 281–299
Norberg T, Rosen L, Baran A, Baran S (2002) On modeling discrete geological structure as Markov random fields. Math Geol 34:63–77
Patil GP, Taillie C (2001) A multiscale hierarchical Markov transition matrix model for generating and analyzing thematic raster maps. Environ Ecol Stat 8:71–84
Pickard DK (1980) Unilateral Markov fields. Adv Appl Probab 12:655–671
Qian W, Titterington DM (1991) Multidimensional Markov chain models for image textures. J Roy Stat Soc Ser B 53:661–674
Rosholm A (1997) Statistical methods for segmentation and classification of images, PhD Dissertation, Technical University of Denmark, Lyngby
Sharp WE, Aroian LA (1985) The generation of multidimensional autoregressive series by the herringbone method. Math Geol 17:67–79
Tjelmeland H, Besag J (1998) Markov random fields with higher-order interactions. Scand J Stat 25:415–433
Weigand H, Totsche KU, Huwe B, Kogel-Knabner I (2001) PAH mobility in contaminated industrial soils: a Markov chain approach to the spatial variability of soil properties and PAH levels. Geoderma 102:371–389
Wu K, Nunan N, Crawford JW, Young IM, Ritz K (2004) An efficient Markov chain model for the simulation of heterogeous soil structure. Soil Sci Soc Am J 68:346–351

Author Biographies

Weidong Li has studied soil and geographical information science for about 20 years and since 1994 he began to develop Markov chain models for description and simulation of categorical soil variables. He got his PhD in soil and water science in 1995 from China Agricultural University. In recent years he worked at several places over the world such as Catholic University of Leuven, Israel Institute of Technology, and University of Wisconsin-Madison for research in soil, hydrology, and geographical information science. He has published more than 30 first-author papers in Chinese and international journals.

Chuanrong Zhang got her PhD in 2004 from UW-Milwaukee. Her research is focused on Internet-GIS and geostatistical modeling for natural resources and environmental management as well as spatial uncertainty analysis of geospatial data.