Markov Chain Random Fields for Estimation of Categorical Variables

Weidong Li

Abstract Multi-dimensional Markov chain conditional simulation (or interpolation) models have potential for predicting and simulating categorical variables more accurately from sample data because they can incorporate interclass relationships. This paper introduces a Markov chain random field (MCRF) theory for building one to multi-dimensional Markov chain models for conditional simulation (or interpolation). A MCRF is defined as a single spatial Markov chain that moves (or jumps) in a space, with its conditional probability distribution at each location entirely depending on its nearest known neighbors in different directions. A general solution for conditional probability distribution of a random variable in a MCRF is derived explicitly based on the Bayes' theorem and conditional independence assumption. One to multi-dimensional Markov chain models for prediction and conditional simulation of categorical variables can be drawn from the general solution and MCRF-based multi-dimensional Markov chain models are nonlinear.

Keywords Multi-dimensional Markov chain · Markov random field · Conditional simulation · Interclass relationship · Nonlinear · Conditional independence

Introduction

Multi-dimensional (multi-D) Markov chain conditional simulation models have great potential for more accurately predicting and simulating categorical variables such as soil types, land cover classes and lithofacies from sample data. This is because they can easily incorporate interclass relationships through cross transition probabilities. Interclass relationships make up a large portion of spatial heterogeneity information conveyed by sampled categorical data and incorporating them into simulations is

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

crucial for effectively capturing spatial patterns of complex categorical variables (so-called multinomial classes) from limited samples.

Earlier multi-D Markov chain models such as Lin and Harbaugh (1984) in geology and Qian and Titterington (1991) in image texture analysis did not condition a simulation on sample data. Recently, a nonlinear 2-D Markov chain conditional simulation approach was developed for simulating categorical variables from survey lines data (Li et al. 2004; Zhang and Li 2005). Further extension of this approach to interpolation and conditional simulation of categorical variables from grid point samples can be found in Li et al. (2005) and Li and Zhang (2005), respectively.

The triplex Markov chain (TMC) model proposed by Li et al. (2004) considered four nearest known neighbors in four cardinal directions and it was built on the coupled Markov chain (CMC) idea of Elfeki and Dekking (2001) in calculating the conditional probability distribution (CPD) of a random variable at an unknown location. The CMC model coupled two 1-D Markov chains together with an independence assumption. Major merits of the CMC model are its simplicity and conditioning idea (i.e., conditioning on a future state), which were retained by the TMC model. Major deficiencies of the CMC model include directional effect (i.e., pattern inclination or diagonal trend) and underestimation of small classes (Li et al. 2004). The directional effect problem was actually a long-standing problem in unconditional multi-D Markov chain simulations in image texture analyses, which was caused by asymmetric neighborhoods used in simulations (Gray et al. 1994). The TMC model effectively overcame the directional effect problem with an alternate advancing (AA) path, and weakened the small-class underestimation problem. Thus, when sufficient data were conditioned, it not only could capture complex spatial patterns of multinomial classes but also approximately reproduce auto and cross variograms or transiograms. Here transiograms refer to transition probability diagrams, which provide a way to estimate transition probabilities with different lags from sparse samples (Li 2006; Li and Zhang 2005). With generalization of the TMC model through transiograms, a middle insertion path was found also effective in overcoming the directional effect problem (Li and Zhang 2005). It should be noted that the directional effect problem was also encountered in extending 1-D autoregressive processes into multiple dimensions, and the herringbone method was proposed to effectively solve the problem (Turner and Sharp 1994; Sharp and Turner 1999). The herringbone method suggested alternating the direction of propagation of the process from lattice row to lattice row in the form of a herringbone pattern so that overall isotropy could be induced. Therefore, the AA path of the TMC model can be regarded as an adaptation of the herringbone idea for conditional Markov chain simulation.

Although the directional effect problem was overcome in the Markov chain conditional simulation approach, the small-class underestimation problem cannot be effectively solved within the framework of the CMC theory. The problem is found to be related with the fundamental assumption of the CMC theory—the full independence of the two 1-D Markov chains in a CMC, which causes unwanted transitions (i.e., transitions of the two 1-D Markov chains to the same location with unequal states). This problem is strong when sample data are relatively sparse, as shown and discussed in Li et al. (2004) and Li and Zhang (2005). Because sparsity of samples is the normal case in real world applications, this problem largely impacts the usefulness of the approach. In addition, the use of multiple chains is also inconvenient

to further expansion of CMC-based models for dealing with random point samples through a random path.

This paper proposes a single-chain based Markov chain random field (MCRF) approach for building one to multi-D Markov chain models for conditional simulation or interpolation. A general solution for the CPD of a random variable in a MCRF is derived explicitly based on the Bayes' theorem and the conditional independence assumption. Multi-D Markov chain models built on the MCRF theory will not have the small-class underestimation problem. With transiograms serving as spatial measures, multi-D MCRF models may work with any kinds of sample data in conditional simulations. Thus, the MCRF and the transiogram will constitute the theoretical backbone of nonlinear Markov chain geostatistics.

It is important to note that the objective of this paper is to define the theoretical basis of the MCRF, not to provide practical implementation techniques. Specific algorithms and implementations of MCRF-based Markov chain models will be provided in future papers or depend on future development of new algorithms.

Small-Class Underestimation

Both the CMC model and the TMC model used couplings of multiple 1-D Markov chains. That means they are multiple-chain models. The occurrence of the small-class underestimation problem is exactly related to the use of multiple chains. In the CMC theory of Elfeki and Dekking (2001), two 1-D Markov chains were coupled together to conduct a 2-D simulation. To couple two 1-D Markov chains, a full independence assumption of two chains in axial directions was employed and these two chains were forced to move to the same location with equal states (Elfeki and Dekking 2001, p. 573). Thus, on a lattice, they had

$$
\begin{aligned}
&\Pr(Z_{i,j} = k \mid Z_{i-1,j} = l, Z_{i,j-1} = m) \\
&= C \cdot \Pr(X_i = k \mid X_{i-1} = l) \Pr(Y_j = k \mid Y_{j-1} = m)
\end{aligned}
\tag{1}
$$

where X_i and Y_j were two 1-D Markov chains in two axial directions and $Z_{i,j}$ was the coupled 2-D Markov chain, all defined in a state space $\{1, 2, \ldots, n\}$ including k, l, and m. In (1), C was a normalizing constant and its necessity was attributed to the exclusion of unwanted transitions (Elfeki and Dekking 2001, p. 575, for exclusion of unwanted transitions).

By normalizing the right-hand side of (1), the CPD of the CMC was obtained as

$$
p_{lm,k} = \Pr(Z_{i,j} = k \mid Z_{i-1,j} = l, Z_{i,j-1} = m) = \frac{p_{lk} \cdot p_{mk}}{\sum_{f=1}^{n}(p_{lf} \cdot p_{mf})} \quad k = 1, \ldots, n
\tag{2}
$$

where p_{lk} represented a transition probability from state l to state k (Elfeki and Dekking 2001, p. 574).

It has been found that the exclusion of unwanted transitions in (2) causes $p_{lm,k}$ to have lower values when k is a small class or higher values when k is a large class. For

example, if we have two 1-D Markov chains of two classes with the following same transition probability matrix

$$
\begin{array}{cc}
\text{Class} & 1 \quad 2 \\
\begin{array}{c} 1 \\ 2 \end{array} &
\begin{bmatrix} 0.1 & 0.9 \\ 0.1 & 0.9 \end{bmatrix}
\end{array}
\tag{3}
$$

given l and m (i.e., they are known), occurrence probabilities of class 1 and class 2 should be 0.1 and 0.9, respectively. However, using (2) we always get the occurrence probability of class 1 (i.e., when $k = 1$)—the small class—at about 0.012 and that of class 2 (i.e., when $k = 2$)—the large class—at about 0.988. Thus, the small class will be underrepresented in simulated realizations (correspondingly, the large class will be overestimated).

In a conditional simulation, the small-class underestimation problem is weakened by conditioning data and may not reach the extent shown in the above example. However, the small-class underestimation problem becomes severe with decreasing density of conditioning data (Li et al. 2004).

Conditional Independence Assumption

The conditional independence assumption used in this study assumes that given a location x its nearest known neighbors in different directions $x_1, \ldots, x_m$ are conditionally independent. This assumption can be generally expressed as

$$
\Pr(x_i \mid x, x_1, \ldots, x_m) = \Pr(x_i \mid x)
\tag{4}
$$

Thus, only pairwise interactions between the unknown location to be estimated and surrounding known data are considered, but the two points of each pair may be remotely located and their interaction is directional through a transition probability with a distance lag (Fig. 1).

The conditional independence assumption has been used widely in many fields such as in Bayesian network (also called Naïve Bayes because this assumption is mathematically difficult to prove) (e.g., Friedman et al. 1997; Ramoni and Sebastiani 2001) and more recently in indicator geostatistics. As suggested by Journel (2002), the assumption of conditional independence is a way around the problem of knowing joint probabilities of multi-points (Ortiz and Deutsch 2004).

Thus far there is no Markov random field (MRF) theory to support the assumption that pixels in any directions (including cardinal directions and non-cardinal directions) are conditionally independent. But in practical uses, it seems nothing can block the applications of this general assumption because of its practicality in simplifying complex issues and in generating acceptable (or even satisfactory) results under some conditions. If the neighborhood is symmetric, we can find a successful application example from Besag (1986) for image processing, where eight adjacent neighbors of x were all assumed conditionally independent. Ripley (1990) suggested that in a pairwise interaction process site interactions between a point and its distant nearest neighbors (in any directions) might be treated independently.
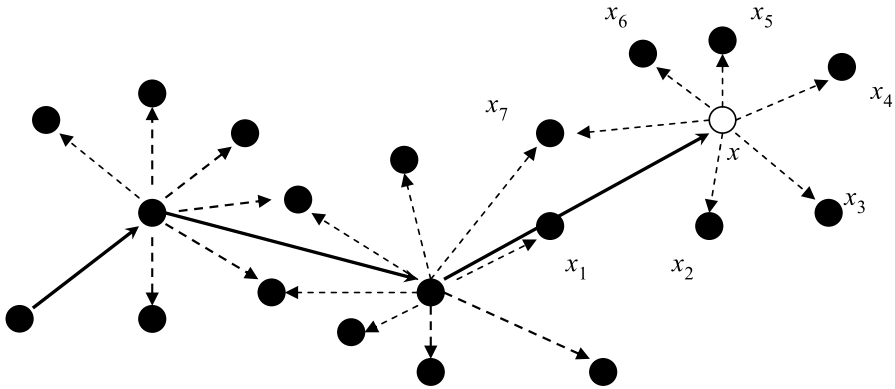
Fig. 1 Illustration of a Markov chain random field. A spatial Markov chain jumps in a multi-dimensional space. The *white cell* refers to the current location with an unknown state, and surrounding *black cells* represent its nearest known neighbors in different directions in the simulation domain (or in a search radius). *Arrows* represent directional interactions. The *solid arrows* also represent the moving directions of the spatial Markov chain. After the spatial Markov chain moves to next location, the current location becomes known

As Journel (2002) suggested, the conditional independence assumption should not be taken lightly and should be checked whenever possible. We found the conditional independence assumption to be theoretically correct in a Pickard random field (Pickard 1980) for sparse samples, only if the nearest known neighbors in cardinal directions are considered (see Appendix). With a conservative attitude, it is suggested herein that if the nearest known neighbors are generally located uniformly throughout the study space and a random path is used in a simulation, inclusion of nearest known neighbors in non-cardinal directions may be considered. But if a fixed path such as the AA path is used, including nearest known neighbors in non-cardinal directions may not be suitable because of the strong asymmetry of structures of nearest known neighbors.

Markov Chain Random Fields

Definition

To avoid the small-class underestimation problem, the suggested MCRF contains only one single Markov chain. Thus, the full independence assumption of multiple chains and exclusion of unwanted transitions will not be needed in the MCRF theory.

A MCRF should be a special MRF because it obeys the transition probability law. The difference is that the former is built on a "directional chain" and the latter does not contain chains. In a MCRF, it is assumed that there is only one single Markov chain in a space which has its CPD at any location entirely dependent on its nearest "known" neighbors in different directions. The distances from such nearest known neighbors to the current location to be estimated may be various. The Markov chain may move or jump within the (any-D) space randomly or along prescribed paths. For

convenience, the Markov chain in a specific MCRF is called a spatial Markov chain (SMC).

A MCRF is generally defined to be a field of a single SMC that moves or jumps in a space, obeying the same or different transition probability rules in different directions and having its states at any unknown location entirely depending on its nearest known neighbors found in different directions. These nearest known neighbors around x form an unfixed neighborhood system. The same labels are used below to represent both locations and states of a random variable at the locations. Assume X is a random variable that obeys the rules of a SMC defined on a state space $S = (1, \ldots, n)$, then according to the above definition of the MCRF, its CPD at any location x is given as

$$\Pr(x \mid (n)) = \Pr(x \mid x_1, \ldots, x_m) \quad m \geq 1 \tag{5}$$

where x represents current location; n represents the n known locations in the whole space; and m indicates the number of nearest known neighbors possibly found in all directions considered in the space.

In (5), m is always greater than or equal to one, because the SMC at least may have its last stay location (already become known) serve as its nearest known neighbor. If the SMC leaps across a known location, the location crossed will replace the last stay location to be the nearest known neighbor in that direction. Figure 1 illustrates a SMC, which jumps from one location to another location in a space (i.e., in a MCRF) and its state at each location depends on its nearest known neighbors, e.g., x_1 to x_7, that could be found in different directions.

General Solution

Using the Bayes' theorem (or the definition of conditional probability), we can decompose the right side of (5) as follows

$$
\begin{aligned}
\Pr(x \mid &x_1, \ldots, x_m) \\
&= \frac{\Pr(x_m, \ldots, x_2, x_1, x)}{\Pr(x_1, \ldots, x_m)} \\
&= \Pr(x_m \mid x, x_{m-1}, \ldots, x_1) \cdots \Pr(x_2 \mid x, x_1) \cdot \frac{\Pr(x_1, x)}{\Pr(x_1, \ldots, x_m)} \\
&= \Pr(x_m \mid x, x_{m-1}, \ldots, x_1) \cdots \Pr(x_2 \mid x, x_1) \cdot \Pr(x \mid x_1) \cdot \frac{\Pr(x_1)}{\Pr(x_1, \ldots, x_m)} \\
&= C \cdot \Pr(x_m \mid x, x_{m-1}, \ldots, x_1) \cdots \Pr(x_2 \mid x, x_1) \cdot \Pr(x \mid x_1)
\end{aligned}
\tag{6}
$$

where C is given as

$$C = \frac{\Pr(x_1)}{\Pr(x_1, \ldots, x_m)} \tag{7}$$

C is a constant because it doesn't involve the unknown location x.

Please note that here, the joint probability $\Pr(x_1, x)$ is decomposed into $\Pr(x \mid x_1) \cdot \Pr(x_1)$, not $\Pr(x_1 \mid x) \cdot \Pr(x)$ so that $\Pr(x)$ will not occur in C. This factorization still obeys the Bayes' theorem.

Applying the conditional independence assumption of (4), the following equations hold

$$\Pr(x_m \mid x, x_{m-1}, \ldots, x_1) = \Pr(x_m \mid x) \tag{8}$$

$$\vdots$$

$$\Pr(x_2 \mid x, x_1) = \Pr(x_2 \mid x) \tag{9}$$

Thus, (6) can be simplified to

$$\Pr(x \mid x_1, \ldots, x_m) = C \cdot \Pr(x_m \mid x) \cdots \Pr(x_2 \mid x) \cdot \Pr(x \mid x_1) \tag{10}$$

If we use transition probabilities with distance lags to replace two-point conditional probabilities in (10), it is

$$\Pr(X(x) = k \mid X(x_1) = l_1, \ldots, X(x_m) = l_m) = C \cdot p_{kl_m}^m(h_m) \cdots p_{kl_2}^2(h_2) \cdot p_{l_1 k}^1(h_1) \tag{11}$$

where $p_{kl_i}^i(h_i)$ represents a transition probability in the ith direction from state k to state l_i with a lag h_i; x_1 represents the nearest known neighbor from or across which the SMC moves to 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 current location to the nearest known neighbor x_i. With increasing lag h, any $p_{kl}(h)$ forms a transition probability diagram, here simply called "transiogram" and suggested as an accompanying spatial measure of MCRF.

Renormalizing the right hand side of (11) to cancel the constant C, it is

$$\Pr(X(x) = k \mid X(x_1) = l_1, \ldots, X(x_m) = l_m) = \frac{\prod_{i=2}^m p_{kl_i}^i(h_i) \cdot p_{l_1 k}^1(h_1)}{\sum_{f=1}^n [\prod_{i=2}^m p_{fl_i}^i(h_i) \cdot p_{l_1 f}^1(h_1)]} \tag{12}$$

Equation (12) gives the general expression of the CPD of the SMC X at any unknown location in a MCRF. The exact form of (12) at a specific location depends on the number of nearest known neighbors found in different directions, distances from these nearest known neighbors to the location to be estimated, and directions along which these nearest known neighbors are located. Using (12), we can calculate the CPD values of X at any unobserved location.

Note that in deriving (12), there is no independence assumption of multiple chains and also no unwanted transitions involved because there is only one single Markov chain in a MCRF. Therefore, the MCRF theory is thoroughly different from the CMC theory.

MCRF-Based SMC Models

As previously mentioned, the conditional independence assumption may not be applicable anywhere. Similar to MRFs which were systematically presented several decades ago (Besag 1974), the general solution of (12) for MCRF may also not practically work for any neighborhood situation. To effectively utilize the MCRF theory in practice, development of practical simulation algorithms is the key.

However, from (12), specific simple SMC models may be drawn and these simple SMC models may be useful in prediction and simulation of categorical variables because they do not underestimate small classes. Because the conditional independence assumption is correct regarding cardinal directions, specific SMC models that consider nearest known neighbors only in cardinal directions are theoretically sound Markov chain models (i.e., not Naïve models. Note that Naïve models may still be practical in some situations). Some of such SMC models in 1-D to 3-D spaces are provided as follows.

In a 1-D space

1. If $m = 1$, that is, if there is only one nearest known neighbor, and $h_1 = 1$ (step length), (12) is reduced to

$$\Pr(X(x) = k \mid X(x_1) = l) = \frac{p_{lk}}{\sum_{f=1}^{n} p_{lf}} = p_{lk} \tag{13}$$

thus, the SMC goes back to the most basic Markov chain—a 1-D unidirectional continuous Markov chain.

2. If m is 2, that is, if an unknown location has two known neighbors, one at each side, we have a 1-D SMC model for conditional simulations drawn from the general solution as

$$\Pr(X(x) = k \mid X(x_1) = l, X(x_2) = q) = \frac{p_{lk}^{1}(h_1) \cdot p_{kq}^{2}(h_2)}{\sum_{f=1}^{n}[p_{lf}^{1}(h_1) \cdot p_{fq}^{2}(h_2)]} \tag{14}$$

Equation (14) has been used in Li and Zhang (2005) for simulating outer boundaries of a study area. Note that if the SMC moves forward continuously (i.e., $h_1 = 1$), the 1-D SMC model can be further simplified as

$$\Pr(X(x) = k \mid X(x_1) = l, X(x_2) = q) = \frac{p_{lk} \cdot p_{kq}(h)}{\sum_{f=1}^{n}[p_{lf} \cdot p_{fq}(h)]} \tag{15}$$

where h stands for the distance from current unknown location to a known neighbor ahead. Equation (14) is a generalized 1-D Markov chain model for conditional simulation, and (15) is a special case of (14). Equation (15) is identical with the 1-D Markov chain model conditioned on a future state on a rectangular lattice introduced by Elfeki and Dekking (2001, p. 572), which can be rewritten as

$$\Pr(X_i = k \mid X_{i-1} = l, X_{i+h} = q) = \frac{p_{lk} \cdot p_{kq}(h)}{p_{lq}(h+1)} \tag{16}$$

in the notation custom of this paper. Equation (15) is a renormalized form of (16). Equation (16) was also the core idea of the CMC theory for conditional simulation.

In a 2-D space

1. First, considering the neighborhood of the CMC model of (2), that is, considering that the SMC have two adjacent known neighbors in orthogonal directions (i.e., $h_1 = 1$ and $h_2 = 1$), we can get the SMC model from the general solution as

$$\Pr(X(x) = k \mid X(x_1) = l, X(x_2) = m) = \frac{p_{lk} \cdot p_{km}}{\sum_{f=1}^{n}(p_{lf} \cdot p_{fm})} \tag{17}$$

This model looks like the CMC model of (2). But carefully checking one can find that they are different: In (2) we have terms p_{mk} and p_{mf}, but here we have p_{km} and p_{fm}. Because of the normal asymmetric property of transition probabilities, given the same transition probability matrix, (17) and (2) generate different results. In addition, (17) is a single-chain 2-D model. It can be used to conduct unconditional simulation without the small-class underestimation problem. If we use the transition probability matrix provided in Expression (3) to (17), we can get the occurrence probability of class 1 (i.e., when $k = 1$)—the small class—exactly at 0.1 and that of class 2 (i.e., when $k = 2$)—the large class—exactly at 0.9 as expected.

2. If we consider subsurface characterization with conditioning on borehole logs, we may use three nearest known neighbors in cardinal directions (left, right, and top). Thus, the general SMC model for subsurface characterization, drawn from the general solution, is

$$
\Pr(X(x) = k \mid X(x_1) = l, X(x_2) = m, X(x_3) = q)
$$
$$
= \frac{p_{lk}^1(h_1) \cdot p_{km}^2(h_2) \cdot p_{kq}^3(h_3)}{\sum_{f=1}^n [p_{lf}^1(h_1) \cdot p_{fm}^2(h_2) \cdot p_{fq}^3(h_3)]} \tag{18}
$$

Similarly, this model is different from the CMC model conditioned on borehole logs given in Elfeki and Dekking (2001, p. 576). Equation (18) will correct the small-class underestimation problem of the CMC model.

3. If we consider four cardinal directions in the horizontal two dimensions, the general SMC model drawn from the general solution can be given as

$$
\Pr(X(x) = k \mid X(x_1) = l, X(x_2) = m, X(x_3) = q, X(x_4) = o)
$$
$$
= \frac{p_{lk}^1(h_1) \cdot p_{km}^2(h_2) \cdot p_{kq}^3(h_3) \cdot p_{ko}^4(h_4)}{\sum_{f=1}^n [p_{lf}^1(h_1) \cdot p_{fm}^2(h_2) \cdot p_{fq}^3(h_3) \cdot p_{fo}^4(h_4)]} \tag{19}
$$

This SMC model is illustrated in Fig. 2. It looks similar to but essentially differs from the TMC model, which was composed of two further extended CMCs (Li et al. 2004; Li and Zhang 2005). Equation (19) will correct the small-class underestimation problem of the TMC model.

In a 3-D space, if only considering cardinal directions, we can get a general 3-D SMC model from the general solution as

$$
\Pr(X(x) = k \mid X(x_1) = l_1, \ldots, X(x_6) = l_6) = \frac{\prod_{i=2}^6 p_{kl_i}^i(h_i) \cdot p_{l_1 k}^1(h_1)}{\sum_{f=1}^n [\prod_{i=2}^6 p_{fl_i}^i(h_i) \cdot p_{l_1 f}^1(h_1)]} \tag{20}
$$

The 3-D herringbone method presented by Sharp and Turner (1999) may be adopted as a practical fixed path in a 3-D Markov chain conditional simulation for subsurface characterization using the above equation.

Above SMC models in (14), (18), (19) and (20) will be useful in 1-D to 3-D interpolation and conditional simulation of categorical variables.

Fig. 2 Illustrating a simplified spatial Markov chain in a two-dimensional space, which considers only nearest known neighbors in cardinal directions. The current cell (i.e., the *white cell*) is conditioned on four nearest known neighbors (*black cells*) at the *top*, *bottom*, and *two sides*. *Black cells* represent known locations. *All arrows* (*dash* and *solid*) represent data interactions and transition probability directions. The *solid arrow* also represents the moving direction of the spatial Markov chain
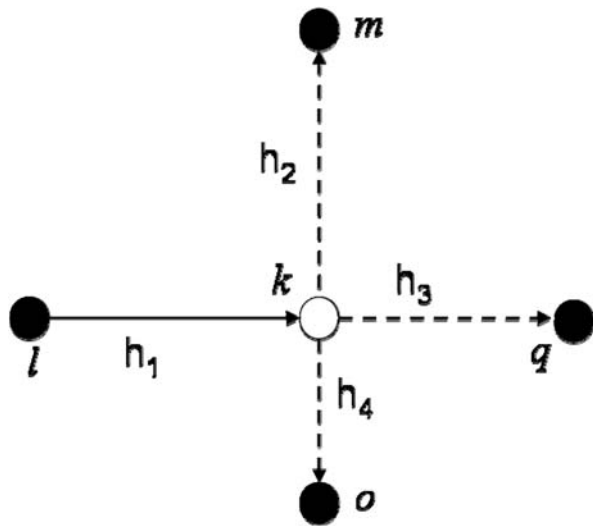
Example

The usefulness of the MCRF theory will depend on the development of practical simulation algorithms. Although multi-D Markov chain models based on the CMC idea underestimate small classes and have other imperfections (Li et al. 2004), they have shown their usefulness in conditional simulations of soil types and land cover classes with relatively dense samples. MCRF-based SMC models can directly replace those models for the same application purposes using similar simulation algorithms. For example, we can directly apply the SMC model of (19) (Fig. 2) to the algorithms developed for the TMC model to acquire better results. However, the MCRF theory can support many different models and algorithms and it is more flexible than the CMC theory.

To show the usefulness and also to demonstrate the settlement of the small-class underestimation problem by the MCRF theory, Fig. 3 shows some simulated realizations of seven land cover classes generated by the TMC model and the SMC model of (19). Details in simulation algorithm and parameter estimation can be seen in Li and Zhang (2005). The simulation domain is a 295×295 grid (i.e., a 5.9×5.9 km^2 area with a pixel size of 20×20 m^2). These realizations are conditioned on three different regular point datasets (1849, 441 and 121 regular points) in simulations. Experimental transiograms were estimated from the most dense sample dataset and interpolated into continuous models (Li and Zhang 2005). The reference land cover map was provided to verify simulated results. It can be seen that small classes such as class 7, class 1 and class 3 are obviously underrepresented or even disappear in realizations generated by the TMC model with decreasing density of conditioning data; however, they are always well-represented in realizations generated by the SMC model, irrespective of the density of conditioning data.

Fig. 3 Simulated realizations of a land cover map with seven land cover classes. **A**, the original map. **B**, **D** and **F**, realizations from the SMC model (i.e., (19) in the text), conditioned on 1849, 441 and 121 regular points, respectively. **C**, **E** and **G**, realizations from the TMC model, conditioned on 1849, 441 and 121 regular points, respectively. The SMC model apparently overcomes the small-class underestimation problem (see class 7, class 1 and class 3)
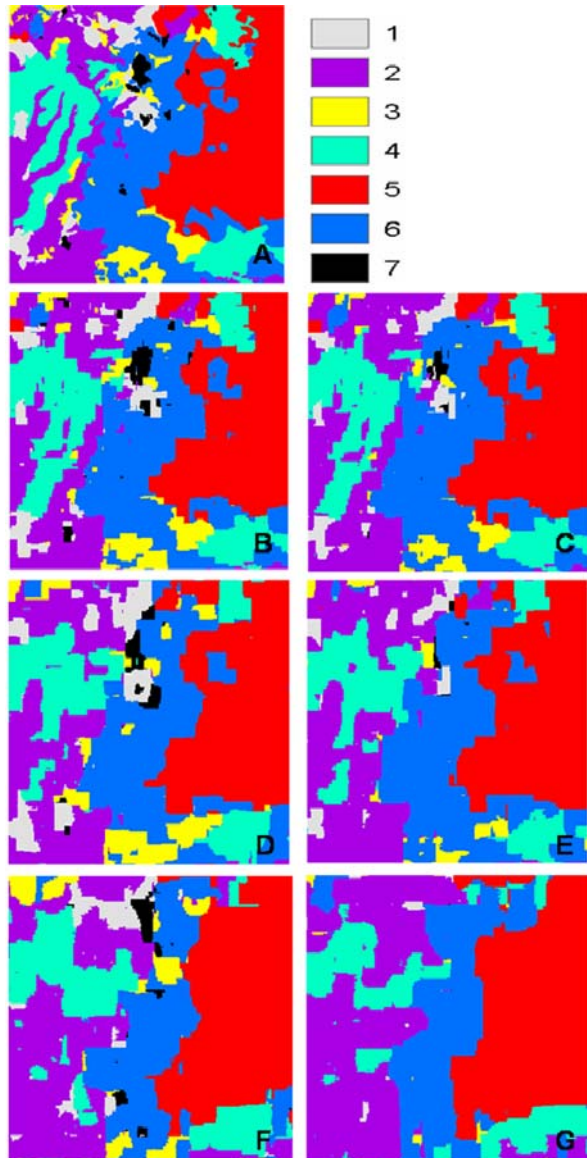
Conclusions and Discussion

A single-chain based MCRF theory for Markov chain estimation of categorical variables in one to multiple dimensions is presented in this paper. A MCRF is a special MRF for dealing with high-order interactions of sparse data. It consists of a single SMC that can move in the whole space. The general solution of the CPD of a random variable at an unknown location in a space of any dimensions is derived explicitly based on the conditional independence assumption and the Bayes' theorem. The general solution only involves 1-D two-point transition probabilities with different lags,

which can be obtained from transiograms. Thus, the MCRF theory apparently simplifies complex problems, facilitates efficient conditional simulation, and meets our needs in dealing with multinomial classes such as soil types. Based on the general solution, different simplified SMC models may be built for different situations and purposes such as simulation and prediction of subsurface formations or surface categorical geographical variables.

Multi-D SMC models overcome the small-class underestimation problem of multi-D Markov chain models based on the CMC theory and are more flexible for algorithm design. Thus, the MCRF theory actually expands Markov chains into a nonlinear Markov chain geostatistics. Simulation algorithms developed for the TMC model can be directly applied to the SMC models that consider only nearest known neighbors in cardinal directions. Development of practical simulation algorithms will be the key for making full use of the MCRF theory.

One issue that should be noted is simulation paths. It is conventionally thought that unilateral processes are difficult to use in practice in multi-dimensions (Martin 1996, 1997) because they cause directional trends in simulated patterns (Gray et al. 1994). That is also why the CMC model (Elfeki and Dekking 2001), as a 2-D unilateral Markov process, generates inclined patterns when conditioning data are sparse or performing unconditional simulations. The herringbone method (Turner and Sharp 1994; Sharp and Turner 1999) solved this problem in simulation of multi-D unilateral autoregressive processes. Similar idea was used in Li et al. (2004) in 2-D Markov chain conditional simulation. As a fixed path, this method should be widely applicable to multi-D unilateral processes including multi-D MCRF-based SMC models. However, to deal effectively with irregularly distributed point samples in conditional simulation, a random path will be more appreciated. The general solution of MCRF provides the flexibility and potential for designing random-path simulation algorithms.

The second issue that should be noted is the conditional independence assumption. This study only indicates that it is theoretically correct to assume conditional independence of nearest known neighbors in cardinal directions. Although the conditional independence assumption has been widely used in recent years in different fields and it is usually applicable, it is not proved (or unprovable) that this assumption can be theoretically correct for nearest (known) neighbors in non-cardinal directions. Thus, the general solution of MCRFs is derived actually based on an assumption that is not fully proved. It should be noted, therefore, that when nearest known neighbors in non-cardinal directions are considered in a simulation, the general solution may not work well for some situations, e.g., asymmetric neighborhoods with fixed paths.

In addition, note that the so-called cardinal directions are not limited to the exact four axial directions. For example, a triangle lattice has only three cardinal directions (Pickard 1980). A suitable tolerance angle for cardinal directions should be feasible for dealing with irregular point samples with a random path, and in a non-lattice space it may be feasible to consider more than four cardinal directions.

References

Besag, J., 1974, Spatial interaction and the statistical analysis of lattice systems (with discussion): J. Roy. Stat. Soc. Ser. B, v. 36, no. 2, p. 192–236.

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

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

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., v. 6, no. 6, p. 707–720.

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., v. 41, no. 3, p. 675–686.

Friedman, N., Geiger, D., Goldszmidt, M., 1997, Bayesian network classifiers: Mach. Learn., v. 29, no. 2-3, p. 131–163.

Gray, A.J., Kay, I.W., Titterington, D.M., 1994, An empirical study of the simulation of various models used for images: IEEE Trans. Pattern Anal. Mach. Intell., v. 16, no. 5, p. 507–513.

Haslett, J., 1985, Maximum likelihood discriminant analysis on the plane using a Markovian model of spatial context: Pattern Recognit., v. 18, no. 3-4, p. 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, New York, p. 115–140.

Journel, A.G., 2002, Combining knowledge from diverse sources: an alternative to traditional data independence hypothesis: Math. Geol., v. 34, no. 5, p. 573–596.

Li, W., 2006, Transiogram: a spatial relationship measure for categorical data: Int. J. Geogr. Inf. Sci., v. 20, no. 6, p. 693–699.

Li, W., Zhang, C., 2005, Application of transiograms to Markov chain simulation and spatial uncertainty assessment of land-cover classes: GISci. Remote Sens., v. 42, no. 4, p. 297–319.

Li, W., Zhang, C., Burt, J.E., Zhu, A.X., 2005, A Markov chain-based probability vector approach for modeling spatial uncertainty of soil classes: Soil Sci. Soc. Am. J., v. 69, no. 6, p. 1931–1942.

Li, W., Zhang, C., Burt, J.E., Zhu, A.X., 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., Harbaugh, J.W., 1984, Graphic display of two- and three-dimensional Markov computer models in geology: Van Nostrand Reinhold Company, New York, p. 180.

Martin, R.J., 1996, Some results on unilateral ARMA lattice processes: J. Stat. Plan. Infer., v. 50, no. 3, p. 395–411.

Martin, R.J., 1997, A three dimensional unilateral autoregressive lattice process: J. Stat. Plan. Infer., v. 59, no. 1, p. 1–18.

Ortiz, J.M., Deutsch, C.V., 2004, Indicator simulation accounting for multiple-point statistics: Math. Geol., v. 36, no. 5, p. 545–565.

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

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

Ramoni, M., Sebastiani, P., 2001, Robust Bayes classifiers: Artif. Intell., v. 125, no. 1-2, p. 209–226.

Rosholm, A., 1997, Statistical methods for segmentation and classification of images: PhD dissertation, Technical University of Denmark, Lyngby, p. 187.

Ripley, B.D., 1990, Gibbsian interaction models, *in* Griffith, D.A., ed., Spatial statistics: Past, present, and future: Institute of Mathematical Geography, Syracuse University, p. 3–25.

Sharp, W.E., Turner, B.F., 1999, The extension of a unilateral first-order autoregressive process from a square net to an isometric lattice using the herringbone method, *in* Lippard, S.J., Nass, A., Sinding-Larsen, R., eds., Proceedings of IMAG'99, 6–11 August 1999, Trondheim, p. 255–259.

Turner, B.F., Sharp, W.E., 1994, Unilateral ARMA processes on a square net by the herringbone method: Math. Geol., v. 26, no. 5, p. 557–564.

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

Appendix: Conditional Independence of Nearest Known Neighbors in Cardinal Directions

If only considering cardinal directions, the conditional independence assumption actually holds in a Pickard random field (PRF). Pickard (1980) proved the existence of a curious unilateral Markov random field, which has a special property that for a generic clique $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix} \right)$, 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

$$\Pr(B \mid D, C) = \Pr(B \mid D) \tag{21}$$

Based on this property, Pickard presented his model, which is still complex and involves three-pixel cliques, for binary processes.

Elfeki and Dekking (2001) adopted a full independence assumption of two 1-D Markov chains to construct their model. The CMC model (Elfeki and Dekking 2001, p. 573) can be simply written as

$$\Pr(D \mid B, C) = \Pr(D \mid B) \cdot \Pr(D \mid C) \tag{22}$$

which requires the two conditional probabilities $\Pr(D \mid B)$ and $\Pr(D \mid C)$ (i.e., two 1-D Markov chains) be fully independent of each other. The consequence of such an assumption has been explained in the text of this paper.

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

$$\begin{pmatrix} & y_1 & \\ y_2 & x & y_3 \\ & y_4 & \end{pmatrix} \tag{23}$$

are conditionally independent. This adaptation is straightforward and it still makes a PRF. To explain this, let's first check the pixel y_1. In a PRF, given x, we have y_1 and y_2 are conditionally independent, and we also have y_1 and y_3 are conditionally independent according to (21). Looking at the direction from y_4 to y_1, we can find y_4 is a past state of y_1 beyond x. So given x, y_1 is independent of y_4. Finally, we have

$$\Pr(y_1 \mid x, y_2, y_3, y_4) = \Pr(y_1 \mid x) \tag{24}$$

and other simpler expressions for conditional independence in a PRF.

The Pickard's theorem is usually used for interactions of adjacent pixels or pixel blocks in image processing. Note that pixels in the generic clique $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix} \right)$ 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) interactions.

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

$$
\begin{pmatrix}
 & & x_1 & & \\
 & & \vdots & & \\
x_2 & \cdots & x & \cdots & x_3 \\
 & & \vdots & & \\
 & & x_4 & &
\end{pmatrix}
\tag{25}
$$

where x_1, x_2, x_3 and x_4 are four nearest known neighbors of x along cardinal directions on a regular lattice. They may be distant from the pixel x with distances h_1, h_2, h_3 and h_4, respectively. Given the pixel x, they can still be regarded as being conditionally independent.

To prove this, let's use E, F, G and H to represent pixel blocks of $\begin{pmatrix} x_1 \\ \vdots \end{pmatrix}$, $(x_2 \cdots)$, $(\cdots x_3)$ 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

$$
\Pr(E \mid x, F, G, H) = \Pr(E \mid x)
\tag{26}
$$

Since only one pixel is known in each of these pixel blocks, from (26) we further have

$$
\Pr(x_1 \mid x, x_2, x_3, x_4) = \Pr(x_1 \mid x)
\tag{27}
$$

Thus, it can be seen that sparse data with high-order interactions along cardinal directions can be conditionally independent in a PRF; that is, within a PRF, (27) holds. Both (24) and (27) can be directly extended to three dimensions.

Further relaxing (27) to include data in non-cardinal directions is practically feasible in some situations. So we may generally write the conditional independence assumption in a multi-D space as

$$
\Pr(x_i \mid x, x_1, \ldots, x_n) = \Pr(x_i \mid x)
\tag{28}
$$

Compared to the "forced" independence of two chains in the CMC theory, this conditional independent assumption of nearest known neighbors is reasonable.