Stoch Environ Res Risk Assess (2008) 22:217–230
DOI 10.1007/s00477-007-0109-2

ORIGINAL PAPER

A comparative study of nonlinear Markov chain models for conditional simulation of multinomial classes from regular samples

Chuanrong Zhang · Weidong Li

Published online: 27 March 2007

Abstract Simulating fields of categorical geospatial variables from samples is crucial for many purposes, such as spatial uncertainty assessment of natural resources distributions. However, effectively simulating complex categorical variables (i.e., multinomial classes) is difficult because of their nonlinearity and complex interclass relationships. The existing pure Markov chain approach for simulating multinomial classes has an apparent deficiency—underestimation of small classes, which largely impacts the usefulness of the approach. The Markov chain random field (MCRF) theory recently proposed supports theoretically sound multi-dimensional Markov chain models. This paper conducts a comparative study between a MCRF model and the previous Markov chain model for simulating multinomial classes to demonstrate that the MCRF model effectively solves the small-class underestimation problem. Simulated results show that the MCRF model fairly produces all classes, generates simulated patterns imitative of the original, and effectively reproduces input transiograms in realizations. Occurrence probability maps are estimated to visualize the spatial uncertainty associated with each class and the optimal prediction map. It is concluded that the MCRF model provides a practically efficient estimator for simulating multinomial classes from grid samples.

Keywords Markov chain random field · Transiogram · Interclass relationship · Conditional simulation · Categorical variable

C. Zhang (✉) · W. Li
Department of Geography, Kent State University,
Kent, OH 44242, USA
e-mail: czhang2@kent.edu

1 Introduction

Complex categorical geospatial variables, such as land cover, land use, soil type, soil quality grade, and lithofacies, are usually categorized into multinomial classes for conveniences of representation and human understanding. Knowing spatial distribution of these multinomial classes is crucial for human management of natural resources and environment. However, it is difficult (actually impossible) to acquire the accurate spatial distribution map of a complex categorical geospatial variable from samples because samples normally account for only a very small portion of the whole study area. In delineating an area-class map from limited samples, human interpretation has to be used in interpolating classes at unsampled (or unobserved) locations. Therefore, it is generally accepted that human-delineated area-class maps based on limited samples are subject to spatial (or locational) uncertainty (e.g., Goodchild et al. 1992; Li and Zhang 2005; Zhang and Goodchild 2002).

In the last two decades, geostatistical conditional simulation methods (mainly various indicator kriging algorithms) have been widely used to simulate discretized continuous variables (i.e., cutoffs or thresholds of continuous variables) and simple categorical variables (e.g., binary variables), and assess their spatial uncertainty by generating alternative realizations and probability maps (see Chiles and Delfiner 1999; Goovaerts 1997). However, for simulating fields of cross-correlated multinomial classes from samples, effective methods lacked, mainly because conventional geostatistics have difficulties in incorporating interclass relationships among multinomial classes (Deutsch 2006) and dealing with their nonlinear spatial distributions with linear estimators (Bogaert 2002). Early studies in simulating categorical variables from

samples (e.g., Bierkens and Burrough 1993a, b; Goovaerts 1996; Kyriakidis and Dungan 2001; Miller and Franklin 2002) commonly used the indicator kriging formalism introduced by Journel (1983). However, they had to ignore interclass relationships or appeal to post-processing of single realizations for interclass correlations because of the difficulty in cokriging multinomial classes, as discussed in Goovaerts (1996) and D'Or and Bogaert (2004).

Major methods incorporating interclass relationships in conditional simulation on samples began to emerge in recent years, when conventional geostatistics met difficulties in expanding its application scope to complex categorical variables with cross-correlated multinomial classes. One approach is the transition probability-based indicator geostatistics introduced by Carle and Fog (1996), which suggests using transition rate-based transition probability models to replace indicator variogram models in indicator kriging so that interclass relationships may be incorporated by avoiding some troubles facing indicator cokriging such as the order relation problem and the parameter permissibility problem. This approach has been applied to three-dimensional (3-D) hydrofacies modeling in recent years due to its advantages in incorporating geological interpretation while sample data are difficult to obtain in subsurface lateral directions (e.g., Weissmann and Fogg 1999; de Marsily et al. 2005). The second is the Bayesian maximum entropy (BME) approach proposed by Christakos (1990, 2000). Bogaert (2002) extended the BME approach for modeling categorical variables. Compared with indicator kriging, the BME approach uses a nonlinear estimator and also incorporates cross-correlations. A recent case study (see D'Or and Bogaert 2004) on ground water table classes using the same datasets used by Bierkens and Burrough (1993b) showed that the BME approach could objectively decrease the spatial uncertainty in predicted results (i.e., generate obviously higher average maximum probabilities) compared with that using indicator kriging. This approach is more general than kriging and may be used to deal with various variables (Christakos 2000). The third is the nonlinear pure Markov chain approach. This approach uses a triplex Markov chain (TMC) model (Li et al. 2004) as its estimator for simulating multinomial classes with incorporation of interclass relationships through transiograms (i.e., transition probability diagrams) (Li and Zhang 2005; Li 2006). Simulation case studies of soil types and land cover classes using this approach (e.g., Li and Zhang 2006) showed that the model could capture complex spatial patterns of multinomial classes when conditioned on a number of samples and it could generate large-scale patterns (i.e., polygons). A major problem with the TMC model is the apparent underestimation of small classes in simulated realizations when conditioning data are relatively sparse. With the proposition of the Markov chain random field (MCRF) theory (Li 2007), which provides theoretically sound multi-D Markov chain estimators, this approach is evolving toward a general and widely applicable nonlinear Markov chain geostatistics.

The TMC model is built on the coupled Markov chain (CMC) idea of Elfeki and Dekking (2001). The CMC theory assumes that two single Markov chains are independent of each other and they move in a 2-D space to the same pixel with equal states. That is, two single chains can be coupled (i.e., multiplied) together. Thus, transitions of the two chains moving to the same pixel with different states become unwanted and have to be excluded in simulation (see Elfeki and Dekking 2001, p. 575). The exclusion of unwanted transitions consequently brings a deficiency—small classes are underestimated in simulated realizations, and the underestimation problem becomes severer with decreasing numbers of conditioning data. This is why the TMC model underestimates small classes. Therefore, the CMC idea is theoretically defective. Because data scarcity is the normal case and small classes are usually important in mapping, this deficiency largely impacts the potential usefulness of the Markov chain approach. The MCRF theory theoretically overcomes the small-class underestimation problem and effectively extends Markov chains into multi-dimensions. A MCRF refers to a single-chain-based random field. Because there is only one single Markov chain in a MCRF, unwanted transitions do not exist. Thus, MCRF-based Markov chain models do not underestimate small classes.

In this paper, we will apply a 2-D MCRF model to conditional simulation of multinomial classes from grid samples. A case study is conducted for testing the new model and comparing it with the TMC model. The objective of this study is to introduce the MCRF model for conditional simulation and spatial uncertainty assessment of multinomial classes and demonstrate that the small-class underestimation problem occurred with the TMC model is effectively solved. Section 2 presents the MCRF model and other related methodologies including the previous TMC model and the estimation of transiogram models. Section 3 introduces the case study dataset and shows estimated transiograms. Simulated results are demonstrated and analyzed in Sect. 4. In the last section we will conclude that the underestimation problem of small classes in the nonlinear pure Markov chain approach is solved.

2 Methods

2.1 Markov chain random field

Differing from conventional multi-D Markov chain models developed in the geosciences, which are built on multiple

1-D chains, a MCRF-based model contains only one single Markov chain no matter how many dimensions are involved. Therefore, MCRF-based models are all single-chain models. In a MCRF, the Markov chain interacts with its nearest known neighbors in different directions through transition probabilities with different lags, which are provided by transiograms. The conditional probability distribution of a MCRF Z at an unknown location u was derived (Li 2007) as

$$\Pr(Z(u) = k|Z_1(u_1) = l_1, \cdots, Z_m(u_m) = l_m) = \frac{\prod_{i=2}^{m} p^i_{kl_i}(h_i) \cdot p^1_{l_1 k}(h_1)}{\sum_{f=1}^{n} \left[\prod_{i=2}^{m} p^i_{fl_i}(h_i) \cdot p^1_{l_1 f}(h_1)\right]} \quad (1)$$

where $p^i_{kl_i}(h_i)$ represents a transition probability in the ith direction from state k to state l_i with a lag h_i; u_1 represents the neighbor from or across which the Markov chain moves to the current location u; m represents the number of nearest known neighbors; k, l_i, and f all represent states in the state space $S = (1, ..., n)$; h_i is the distance from the current location to the nearest known neighbor u_i. With increasing lag h, any $p_{kl}(h)$ forms a transition probability diagram—a transiogram. In directions i, transitions are from the current unknown location u to its nearest known neighbors, but in direction 1 (i.e., the coming direction of the Markov chain), the transition is from the nearest known neighbor u_1 to the current location u.

Equation 1 is the general solution of MCRFs. In real applications, it is not necessary to consider many nearest known neighbors in different directions. Usually considering only the nearest known neighbors in the four cardinal directions is sufficient for simulating multinomial classes in the horizontal two dimensions. Thus, Eq. 1 can be simplified to

$$p_{k|lmqo} = \Pr(Z(u) = k|Z(u_1) = l, Z(u_2) = m, Z(u_3) = q, Z(u_4) = o) = \frac{p^4_{ko}(h_4) \cdot p^3_{kq}(h_3) \cdot p^2_{km}(h_2) \cdot p^1_{lk}(h_1)}{\sum_{f=1}^{n} [p^4_{fo}(h_4) \cdot p^3_{fq}(h_3) \cdot p^2_{fm}(h_2) \cdot p^1_{lf}(h_1)]} \quad (2)$$

where, 1, 2, 3, and 4 represent the four cardinal directions considered, h_1, h_2, h_3, and h_4 represent the distances from the unknown location u to its nearest known neighbors u_1, u_2, u_3, and u_4 in the four cardinal directions, respectively; and k, l, m, p, and o represent the states of the Markov chain at the five locations u, u_1, u_2, u_3, and u_4, respectively.

Equation 2 is the MCRF model for 2-D simulation, as illustrated in Fig. 1. On outer boundaries of a study area, nearest known neighbors in cardinal directions are normally less than four. Equation 2 can always be adapted to those situations by deleting transition probabilities involving those missing neighbors.

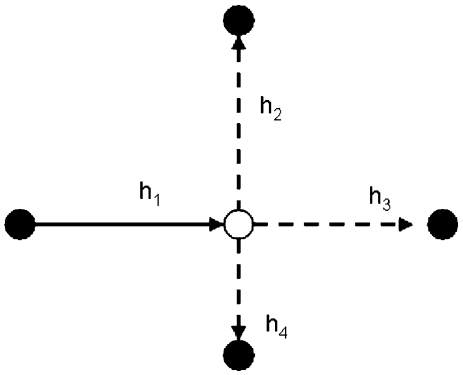

Fig. 1 Illustration of the Markov chain random field model: a single Markov chain is used to conduct two-dimensional simulation. h_1, h_2, h_3, and h_4 all represent distances. *Black cells* represent known locations. The *white cell* stands for the unknown location to be estimated. The *solid arrow* shows the moving of the Markov chain. *Dash arrows* indicate interactions between the Markov chain and its other nearest known neighbors in cardinal directions. All *arrow* directions also represent transition probability directions

2.2 Comparing to the triplex Markov chain model

The TMC model estimates a value at an unknown location by conditioning on its four nearest known neighbors in four cardinal directions. In calculating the conditional probability distribution for each unknown location, it uses the CMC idea—that is, it uses three single Markov chains to make two couplings. Therefore, the TMC model is a multiple-chain model. Figure 2 demonstrates one coupling of two chains used in the TMC model.

To work with point data through using transiograms, the TMC model was extended as follows (Li and Zhang 2005):

$$p_{k|lmqo} = \frac{p^4_{ko}(h_4) \cdot p^3_{kq}(h_3) \cdot p^2_{mk}(h_2) \cdot p^1_{lk}(h_1)}{\sum_{f=1}^{n} [p^4_{fo}(h_4) \cdot p^3_{fq}(h_3) \cdot p^2_{mf}(h_2) \cdot p^1_{lf}(h_1)]} \quad (3)$$

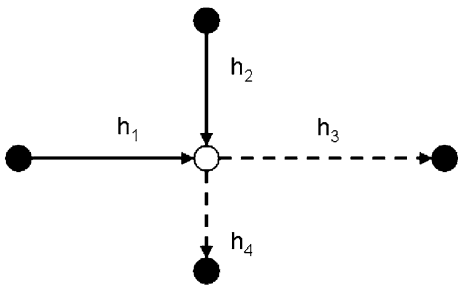

Fig. 2 Illustration of the generalized coupling-based Markov chain model: two chains move to the same location from the lateral and the vertical directions. The meanings of *symbols* are similar to those in Fig. 1

Equation 3 is a generalized multiple-chain model as illustrated in Fig. 2. Equations 2 and 3 look similar in their expressions of the conditional probability distribution. Carefully checking them, however, one can find that terms $p^2_{mk}(h)$ and $p^2_{mf}(h)$ in Eq. 3 differ from terms $p^2_{km}(h)$ and $p^2_{fm}(h)$ in Eq. 2. It is this difference that decides whether there are unwanted transitions in 2-D Markov chain models and consequently whether or not small classes are underestimated in simulated realizations. The reason for this difference is because they are built on the two completely different Markov chain ideas—the multiple-chain idea and the single-chain idea, with different assumptions.

In this paper, Eq. 3 is also used to conduct simulations on the same datasets so that simulated results from the two models can be compared.

2.3 Simulation algorithm

To avoid the pattern inclination problem in simulated realizations of multi-D unilateral processes (see Gray et al. 1994 for the directional effect occurred in 2-D Markov mesh models, and Sharp and Aroian 1985 for the same occurred in 2-D autoregressive models), we need to use the AA (alternate advancing) path (Li et al. 2004) in our simulations. The AA path is similar to the herringbone path suggested by Sharp and Aroian (1985) for overcoming the pattern inclination problem occurred in extending 1-D autoregressive processes in two dimensions. The herringbone method suggested alternating the direction of propagation of the autoregressive process from lattice row to lattice row in the form of a herringbone pattern so that overall isotropy could be induced.

The AA path is used in both the MCRF model and the TMC model for 2-D simulations. In a simulation, transition probabilities needed at any lags are directly drawn from transiogram models. Monte Carlo simulation is used to generate realizations. The simulation procedure consists of the following steps: (1) first simulate outer boundaries, where nearest known neighbors in cardinal directions are less than four; (2) connect (by simulation) all neighboring observed data points that are not connected by the simulation in step 1 so that the simulated lines form a network; and (3) within each polygon formed by simulated lines, perform simulation row by row from top to bottom.

The above procedure ensures that when estimating an unsampled location within a mesh formed by simulated lines, effects of all close sample points are incorporated directly or indirectly. Note that the so-called cardinal directions in the MCRF model and the TMC model refer to just four orthogonal directions and that they can be rotated. So the above algorithm has no limitation on randomly distributed point data.

2.4 Estimation of transiograms

An auto-transiogram $p_{ii}(h)$ represents the auto-correlation of single class i and a cross-transiogram $p_{ij}(h)$ $(i \neq j)$ represents the interclass relationship from class i to class j. Three constraint conditions need be considered in modeling transiorgams for simulation: (1) no nuggets for exclusive classes; (2) nonnegative; and (3) at any lag h, values of transiograms headed by the same class sum to 1 (Li 2006). A typical feature of transiograms is that the height of a transiogram is a reflection of the proportion of the tail class. Under the ideal conditions (e.g., data are stationary and first-order Markovian and the study area is sufficiently large), the sill of a transiogram is equal to the proportion of the tail class.

When sampled data are too sparse, experimental transiograms may not be reliable. Under this situation, it is better to use expert knowledge to assess parameters (such as sills, ranges, and model types) of transiogram models (Li and Zhang 2006), and not to completely depend on experimental data. But when sampled data are sufficient, experimental transiograms are reliable and may effectively reflect the real spatial variation characteristics of studied classes. Under this situation, their features should be respected as much as possible. For the purpose of testing the MCRF model, the case study in this paper will use grid samples so that experimental transiograms are reliable and can be simply interpolated into continuous models.

Data analysis has found that experimental transiograms of multinomial classes in non-stationary areas (as is usually the case) have complex shapes. Classical mathematical models such as the exponential model and the spherical model may not effectively capture the erratic features of experimental transiograms. Li and Zhang (2005) suggested a linear interpolation method for fitting experimental transiograms into continuous models. The linear interpolation method is given in the following equation:

$$X = \frac{A(D_B - D_X) + B(D_X - D_A)}{D_B - D_A} \tag{4}$$

where A and B are the values of two neighboring points in an experimental transiogram, D_A and D_B are the corresponding lags of the two neighboring points with $D_B > D_A$, and X is the value to interpolate at a lag D_X between D_A and D_B.

Interpolated transiograms using Eq. 4 normally can meet the requirements of a simulation for transiogram modeling (i.e., the aforementioned three constraint conditions). But occasionally, the whole subset of transiograms headed by a minor class may be equal to zero at the high-lag section, and thus violate the summing-to-one constraint condition. Therefore, if the high-lag section of transio-

grams is used in a simulation, a good practice is to set the unreliable high-lag section of interpolated transiogram models to the proportions of corresponding tail classes.

3 Datasets and cases study

3.1 Study area and sample datasets

The case study is conducted using land cover classes in an area as an example. The case study area is located in the Lunan Stone Forest National Park in the Yunnan province of China. The study area has a 5.9 km length and a 5.9 km width. Seven land cover classes are classified from the area (Li and Zhang 2005). Among these classes, classes 1, 3, and 7 are small classes, together accounting for about 13% of the area. For simulation, the study area is discretized into a 295 × 295 grid lattice with a cell size of 20 × 20 m.

For estimation of experimental transiograms and conditional simulation, a point dataset is sampled in the study area, which is composed of 1,849 regularly distributed points, accounting for 2% of the total pixels. This dataset is used as a *dense* dataset. From the dense dataset, 441 regularly distributed points are extracted as a *moderate* dataset, accounting for 0.5% of the total pixels. From the moderate dataset, 121 regularly distributed points are further extracted from the moderate dataset as a *sparse* dataset, accounting for 0.139% of the total pixels. In fact, all three datasets may be thought to be sparse compared with the total number of the pixels and the study area. In this paper, the terms *dense*, *moderate*, and *sparse* are used to differentiate the three different densities of sampled data.

Note that the Markov chain models introduced in the last section have no limitation on data formats. Potentially, they may work with any data types (regular/irregular lines, points, small areas or mixtures) with an intensively developed software system. In this paper, regular point data are used to test the new model and demonstrate its feasibility in simulating multinomial classes. In fact, without auxiliary information and previous sampling, regular sampling is generally regarded as the most efficient sampling scheme for mapping purposes (De Gruijter et al. 2006).

Figure 3 shows a land cover map delineated based on the observed dense dataset, field visual observation, and a high-resolution satellite image of the study area. Such a map may serve as a reference map, representing the real land cover patterns of the study area. Accurate land cover distribution maps are normally unavailable for many applications because of the difficulty in acquiring detailed and accurate observations at every location in a large area.

3.2 Transiograms for simulation

In this study, omnidirectional transiograms are used. Thus, one set of transiogram models (i.e., 49 auto/cross-transio-

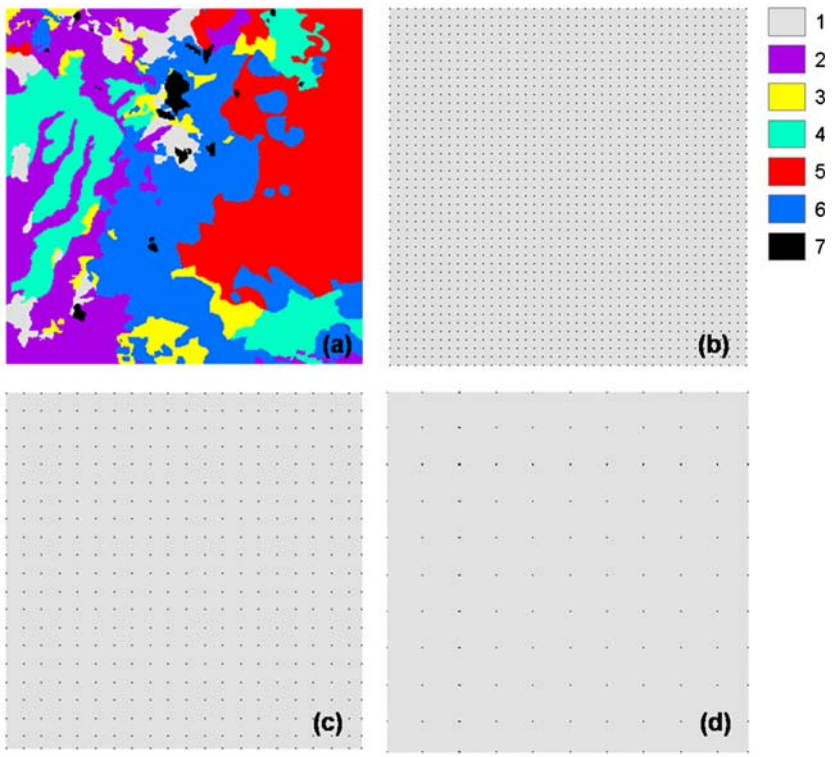

Fig. 3 The reference map with seven land cover classes and the three regular sample datasets: **a** the reference map, **b** the dense dataset (1,849 points), **c** the moderate dataset (441 points), and **d** the sparse dataset (121 points)

grams for the seven classes) is sufficient for conducting simulations. Otherwise, we need several sets of transiorgam models, one set for each direction.

Experimental transiograms are estimated from samples by counting all data pairs with different lags. Figure 4 shows a subset of interpolated experimental transiograms headed by class 2, estimated from the moderate dataset; and Fig. 5 provides that headed also by class 2, but estimated from the dense dataset. The end section of these transiogram models is set to the proportions of corresponding classes. It can be seen that these experimental transiograms are complex but reliable because both datasets (i.e., dense and moderate) provide similar transiograms. Therefore, simply interpolating experimental transiograms into continuous models is suitable and may capture more spatial heterogeneity conveyed by corresponding datasets.

The sparse dataset apparently has too few data points to estimate reliable experimental transiograms for seven classes, particularly for small classes. Therefore, the transiogram models estimated from the dense dataset is used to conduct simulation on the sparse dataset. In real applications, when data are too sparse to acquire reliable experimental transiograms, transiogram models may be estimated completely based on expert knowledge. Use of expert knowledge, though important in practice, undoubtedly brings parameter uncertainty, which may not be helpful here for the model testing purpose.

3.3 Simulations and result processing

One hundred realizations are generated for each of the three datasets using both the MCRF model and the TMC model. Occurrence probability vectors are estimated from

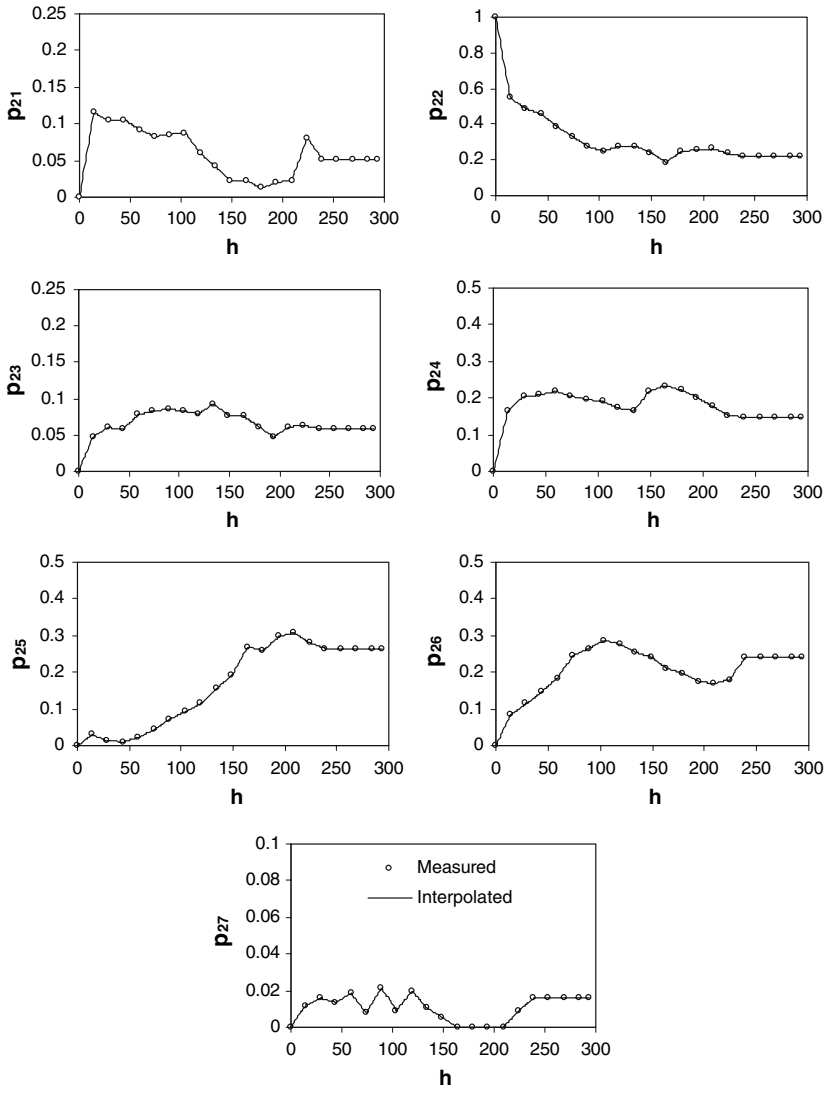

Fig. 4 A subset of transiogram models headed by class 2, estimated from the moderate dataset. The high-lag section is set to the proportion of each class in the sampled dataset. The scales along the *h* axis are numbers of pixels

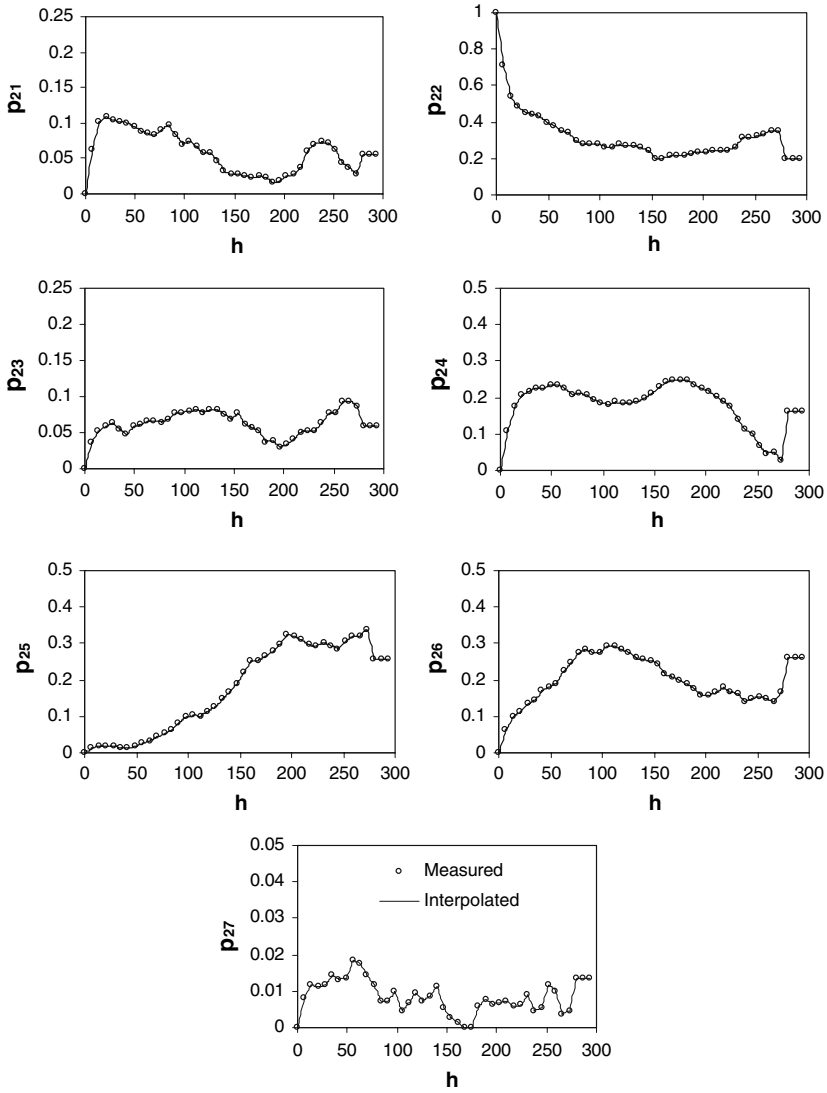

Fig. 5 A subset of transiogram models headed by class 2, estimated from the dense dataset. The high-lag section is set to the proportion of each class in the sampled dataset

100 realizations for each simulation, and occurrence probability maps are visualized from corresponding occurrence probability vectors. Prediction maps are generated based on maximum occurrence probabilities. To see whether the input transiograms are effectively reproduced, transiograms are estimated from the exhaustive data of ten realizations for each simulation and compared with the experimental ones.

4 Results and discussions

4.1 Prediction maps

The early purpose of geostatistical modeling is for spatial prediction through interpolation. Thus far, spatial prediction is still the aim of many studies and applications. Prediction is normally based on the optimal estimates using a regression model such as kriging, not simulation. Most geostatistical simulation algorithms appear after the 1980s and they can be used to generate alternative realizations, and further obtain occurrence probabilities of a cutoff estimated from a large number of realizations (Goovaerts 1997). For Markov chain simulation of categorical variables, an optimal prediction map can be estimated based on the maximum occurrence probabilities estimated from a large number of realizations. An optimal prediction map provides us a good estimate at the spatial distribution of classes and the quality of the prediction depends on the capability of the model used.

Figure 6 shows prediction maps conditioned on the three (dense, moderate, and sparse) datasets using the MCRF model and the TMC model. It can be seen that when sampled data are relatively dense, the prediction maps are very imitative of the reference map in details. With decreasing density of samples, prediction maps

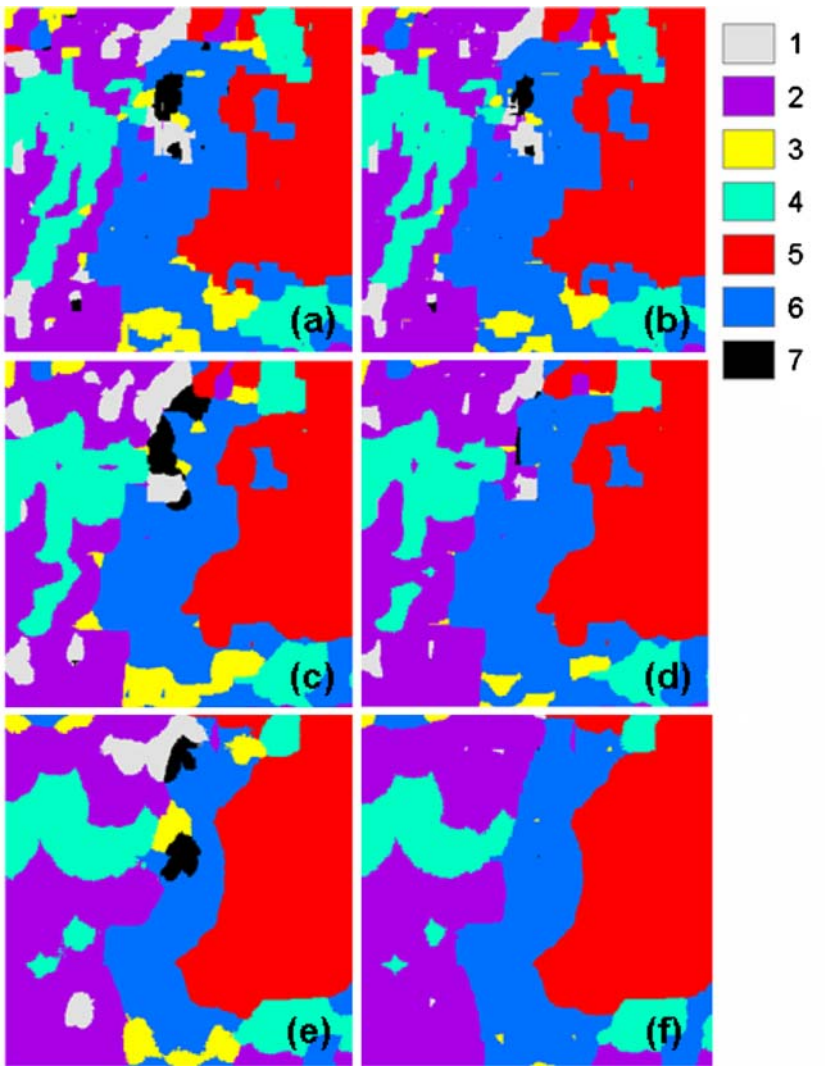

Fig. 6 Prediction maps of land cover classes using the MCRF model (*left column*) and the TMC model (*right column*), conditioned on three sampled datasets with different sampling densities: **a**, **b** conditioned on the dense dataset; **c**, **d** conditioned on the moderate dataset; and **e**, **f** conditioned on the sparse dataset

gradually lose pattern details, but the major patterns (i.e., large class areas) are still captured. The difference between prediction maps generated by the MCRF model and those generated by the TMC model is that with decreasing density of samples the MCRF-predicted maps do not lose small classes (see classes 7, 3 and 1 in Fig. 6a, c, e), but small classes gradually disappear in the TMC-predicted maps (see classes 7, 3 and 1 in Fig. 6b, d, f).

Table 1 Predicted proportions of land cover classes based on maximum occurrence probabilities, conditioned on different sampled datasets using the MCRF model and the TMC model

Class	Reference map	Dense dataset			Moderate dataset			Sparse dataset		
		Data	MCRF	TMC	Data	MCRF	TMC	Data	MCRF	TMC
1	0.0546	0.0525	0.0508	0.0376	0.0522	0.0528	0.0232	0.0331	0.0316	0.0009
2	0.1941	0.2044	0.1989	0.2148	0.2200	0.2040	0.2526	0.2810	0.2807	0.3175
3	0.0589	0.0568	0.0535	0.0336	0.0590	0.0512	0.0184	0.0579	0.0452	0.0037
4	0.1608	0.1601	0.1683	0.1664	0.1474	0.1610	0.1465	0.1488	0.1449	0.1212
5	0.2553	0.2542	0.2558	0.2575	0.2653	0.2720	0.2776	0.2975	0.3008	0.3043
6	0.2628	0.2591	0.2604	0.2836	0.2404	0.2384	0.2803	0.1653	0.1794	0.2524
7	0.0134	0.0130	0.0123	0.0065	0.0157	0.0206	0.0014	0.0165	0.0174	0.0001

Table 1 provides the proportion data of all classes in the sampled datasets and prediction maps. Apparently, the MCRF-predicted proportions are identical with those in the sample datasets; however, small classes are underestimated in the TMC-predicted results (and consequently some large classes are overestimated). The small-class underestimation problem of the TMC model occurs with all three datasets, and the situation becomes worse with decreasing density of samples.

It is normal in geostatistical simulations that proportions of small classes decrease in prediction maps with decreasing conditioning data because of the uneven smoothing effect. However, it is surprising to find that in this case study the MCRF model can keep the proportions of small classes in prediction maps with fewer but bigger polygons when sampled data become sparser. These predicted results means that the MCRF model can perform well no matter how many data are conditioned, but the TMC model is apparently inappropriate to use when conditioning data are relatively sparse. It is desirable that all classes, regardless of their proportions, are well predicted and the predicted proportions are identical with those in the sample dataset. The capability of the MCRF model to keep the suitable proportions of classes in prediction maps in our simulation cases is a little surprising. It may be related with the fact that the small classes are well cross-correlated with large classes. This may be important for the MCRF model to be a good estimator.

4.2 Simulated realizations

Simulated realizations can represent spatial uncertainty by displaying the differences between different realizations, and they are also used for uncertainty propagation analysis by introducing them into application models (Mowrer 1997). It is expected that each realization may represent one possible spatial configuration of the classes. Therefore, effectively capturing all classes and their spatial heterogeneity in simulated realizations is desirable. Good random field models should keep the suitable proportions of all classes in simulated realizations, though different realizations may look more or less different. It is also desirable that simulated realizations can imitate the real patterns of all classes (assuming they are known in model testing).

Figures 7 and 8 show simulated realizations generated by the MCRF model and the TMC model, respectively. It can be seen that simulated realizations by the two models all show patterns similar to those in the reference map, and the similarity with the reference map decreases when conditioning data become sparser. Comparing simulated realizations generated by the MCRF model and those generated by the TMC model, it is clear that the latter gradually lose small classes—that is, small classes are underestimated and the underestimation become more severe when data become more sparse; however, the former do not have this tendency. The underestimation problem of small classes in realizations simulated by the TMC model also can be seen from the simulated class proportions provided in Table 2. For example, the underestimation degrees of the smallest class—class 7 in TMC-simulated realizations are 50.8, 85.4, and 98.2% for the dense, moderate and sparse datasets, respectively. On the contrary, class 7 does not have this tendency in data from the MCRF model. Therefore, the TMC model is obviously inferior to the MCRF model in capturing small classes, and it may be inappropriate to use for simulation when observed data are relatively sparse.

There are obvious advantages with simulated realizations of the Markov chain approach: (1) classes are usually captured at their approximate locations; (2) large polygons can be generated. These characteristics are desirable for simulation of multinomial classes, because such kinds of realizations are probably more useful in applications. For example, in ecological or hydrological modeling, the correct locations and patch sizes of classes are helpful for accurately determining the preferential paths of (energy, solute or water) flows.

Appropriate generation of all classes in realizations, no matter small or large, is crucially important. Sometimes, the smallest class may be the most important one in a study. For example, a small land cover class may represent a special species or disease-infected areas that we want to know the most. While the TMC model may still be used as a workable model in some situations (e.g., sampled data are dense, or all classes have similar proportions), considering the deficiency and the theoretical flaw it has and that the MCRF model is theoretically sound and also does not add any overhead in simulation, the MCRF model should be a better choice for conditional Markov chain simulation of multinomial classes in the future.

4.3 Transiogram analysis

Approximately reproducing input statistics in simulated realizations is the basic requirement for a random field model to be trusted. Li and Zhang (2005) showed that features of original transiograms could be followed by simulated realizations using the TMC model, but deviation between simulated transiograms and the original ones (i.e., experimental transiograms) increased when conditioning data became more sparse, mainly because the underestimation problem of small classes was reflected in the heights of simulated transiograms. Since transiograms are used in this study, it is also necessary to check whether the MCRF model can reproduce the input transiograms.

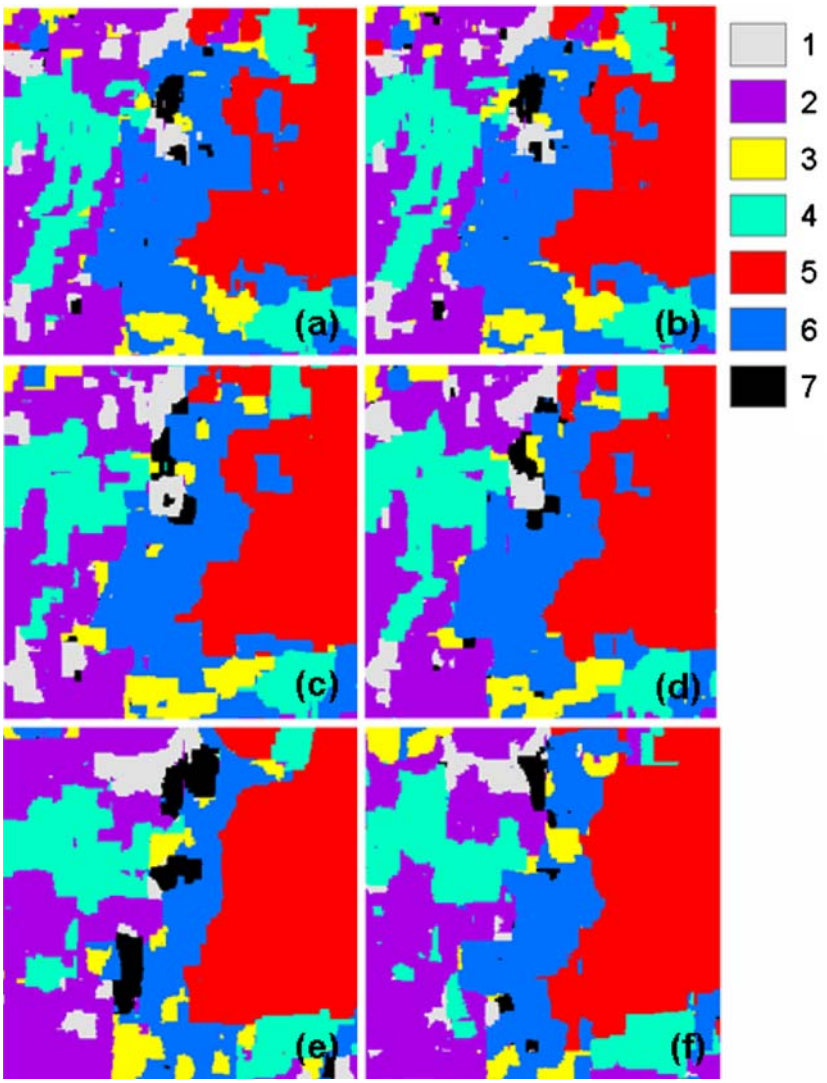

Fig. 7 Simulated realizations generated by the MCRF model conditioned on different datasets: **a**, **b** conditioned on the dense dataset; **c**, **d** conditioned on the moderate dataset; and **e**, **f** conditioned on the sparse dataset

Figure 9 provides a subset of simulated transiograms (i.e., transiograms estimated from simulated realizations) headed by class 2 using the MCRF model, conditioned on the moderate dataset. The first ten realizations are used for estimating simulated transiograms. It can be seen that experimental transiograms estimated from the original dataset are well reproduced in simulated realizations with reasonable ergodic fluctutations (see Deutsch and Journel 1998, pp. 128–132, for explanation of ergodic fluctuation). Particularly interesting is that experimental transiograms are reproduced both at the low-lag section and the high-lag section. This means spatial heterogeneity conveyed by the sample dataset is completely captured by the model. Because small classes are well produced with their due proportions, the heights of simulated transiograms do not obviously deviate from the original ones. Other simulated transiograms conditioned on the same dataset fit corresponding original ones similarly. Simulated transiograms conditioned on the dense dataset (not shown) are basically identical with the original and those conditioned on the sparse dataset (also not shown) show larger ergodic fluctuations. In general, with the small-class underestimation problem being solved, the MCRF model has the capability of effectively reproducing both auto and cross transiograms.

4.4 Occurrence probability maps

To visually show the spatial uncertainty associated with the occurrence of each class and all classes together, the better representation is the use of occurrence probability maps rather than single realizations.

Figure 10 displays occurrence probability maps generated by the MCRF model, conditioned on the moderate dataset. These probability maps of single classes clearly reveal where and with how much certainty (or uncertainty)

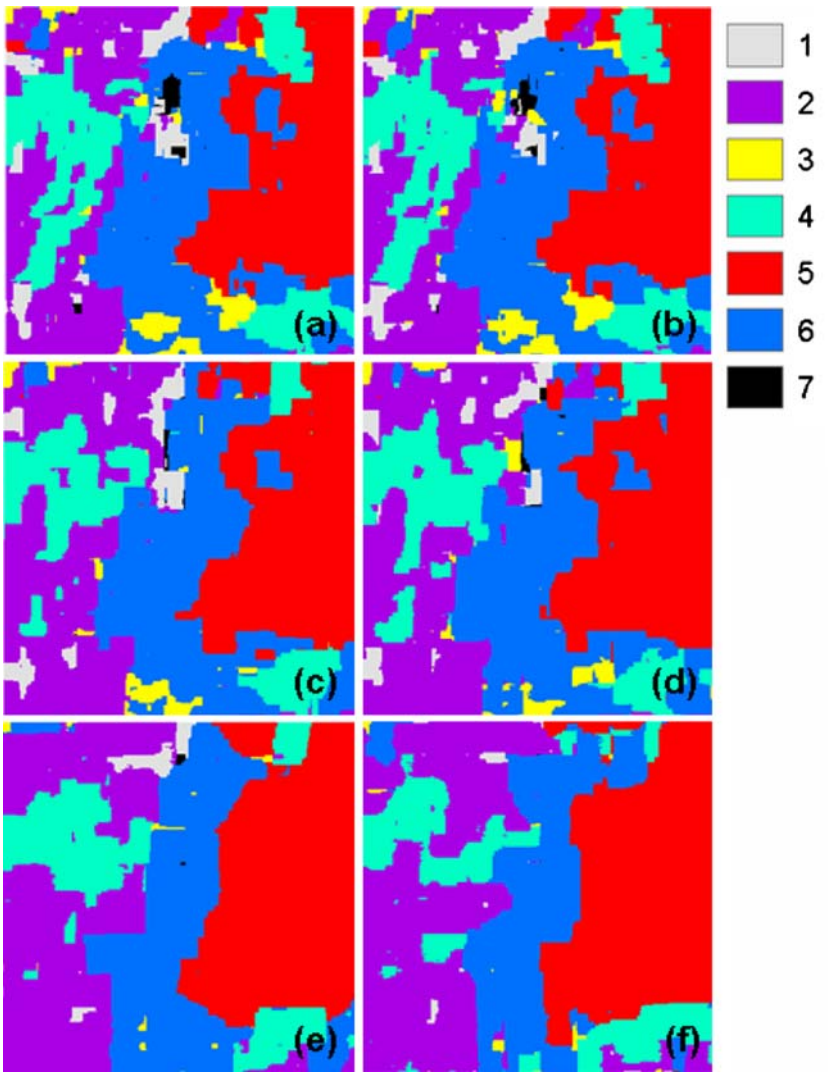

Fig. 8 Simulated realizations generated by the TMC model conditioned on different datasets: **a**, **b** conditioned on the dense dataset; **c**, **d** conditioned on the moderate dataset; and **e**, **f** conditioned on the sparse dataset

Table 2 Simulated proportions of land cover classes, averaged from 100 simulated realizations, conditioned on different sampled datasets using the MCRF model and the TMC model

Class	Dense dataset					Moderate dataset					Sparse dataset				
	Data	MCRF	Dev	TMC	Dev[a]	Data	MCRF	Dev	TMC	Dev	Data	MCRF	Dev	TMC	Dev
1	0.0525	0.0527	0.4	0.0384	26.9	0.0522	0.0564	8.1	0.0250	52.1	0.0331	0.0397	19.9	0.0051	84.6
2	0.2044	0.1963	4.0	0.2143	4.8	0.2200	0.2154	2.1	0.2495	13.4	0.2810	0.2713	3.4	0.3079	9.6
3	0.0568	0.0556	2.1	0.0352	38.0	0.0590	0.0558	5.4	0.0209	64.6	0.0579	0.0560	3.3	0.0074	87.2
4	0.1601	0.1678	4.8	0.1656	3.4	0.1474	0.1500	1.8	0.1464	0.7	0.1488	0.1520	2.2	0.1347	9.5
5	0.2542	0.2551	0.4	0.2572	1.2	0.2653	0.2676	0.9	0.2740	3.3	0.2975	0.2963	0.4	0.3035	2.0
6	0.2591	0.2582	0.3	0.2829	9.2	0.2404	0.2379	1.0	0.2818	17.2	0.1653	0.1663	0.6	0.2412	45.9
7	0.0130	0.0143	10.0	0.0064	50.8	0.0157	0.0169	7.6	0.0023	85.4	0.0165	0.0184	11.5	0.0003	98.2

[a] Deviation (%)

a class may occur. It can be seen that with decreasing density of sampled data, spatial uncertainty associated with each class increases. The most interesting are those maximum occurrence probability maps, which clearly show the transition zones (see the white to shallow-gray stripes) between class polygons (i.e., areas that are more homogeneous). Transition zones become wider with decreasing density of sample data, indicating more spatial uncertainty

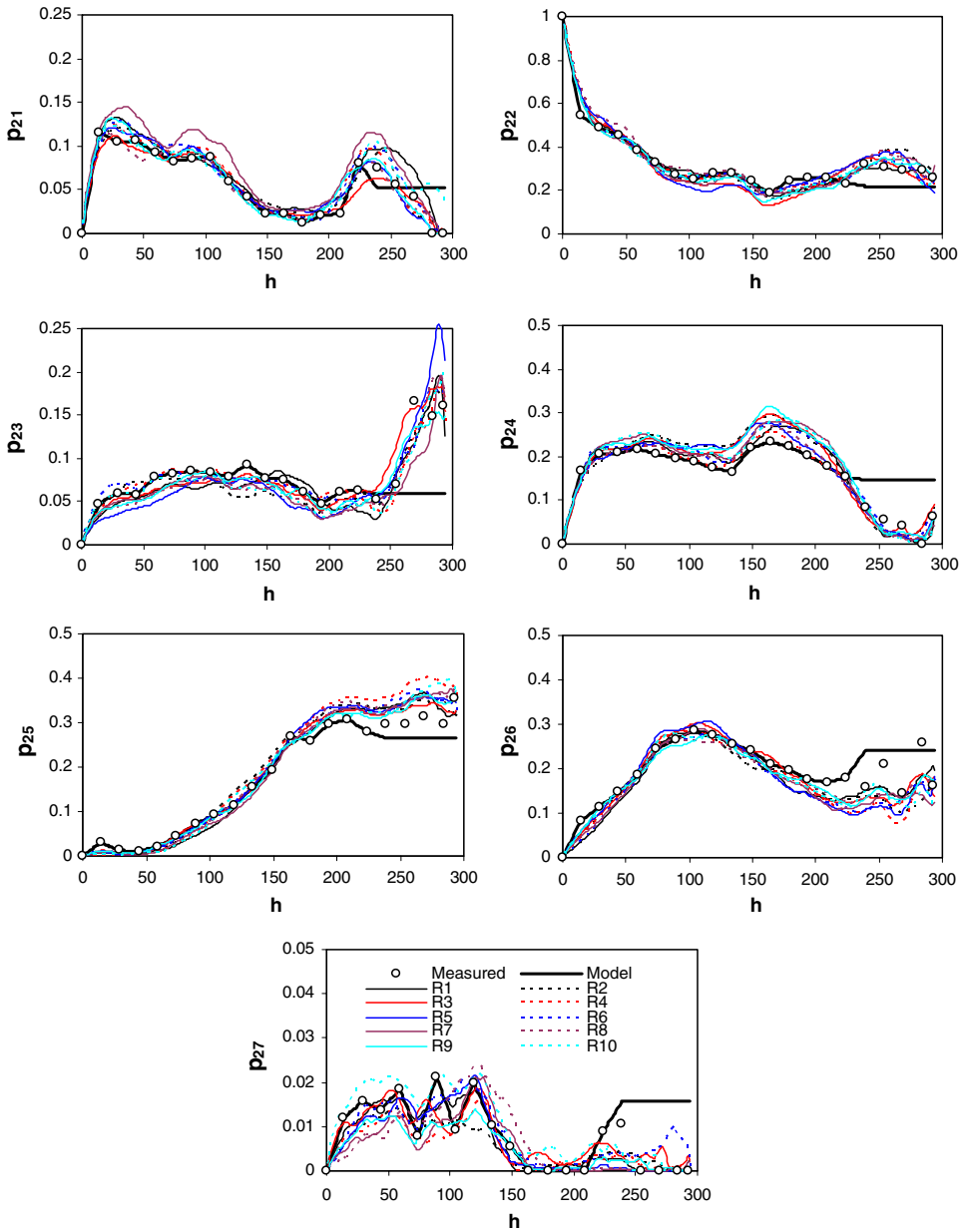

Fig. 9 Simulated transiograms headed by class 2 using the MCRF model, conditioned on the moderate dataset, estimated from the first ten realizations

is contained in related prediction maps (or human-delineated area-class maps).

Compared with realizations, probability maps can demonstrate spatial uncertainty more vividly. The effective capture of transition zones in nonlinear Markov chain simulation of categorical variables is interesting and also meets the expectations of pioneer studies in spatial uncertainty of area-class maps (e.g., Mark and Csillag 1989; Goodchild et al. 1992). As Goodchild et al. (1992, p. 91) pointed out, ''the degree of fuzziness of the boundary (an object concept) can thus be represented in the steepness of the probability gradient (a field concept)''.

5 Conclusions

Although the coupling-based 2-D TMC model is capable of generating imitative patterns of multinomial classes through the incorporation of interclass relationships when conditioned on dense samples, it has an apparent deficiency—underestimation of small classes, which largely impacts the usefulness of the Markov chain geostatistical approach in sparse-data modeling. The MCRF theory (Li 2007) provides the theoretical foundation for a theoretically sound nonlinear Markov chain geostatistics. The MCRF-based model used in this paper has all the merits of the previously used TMC model, but effectively overcomes

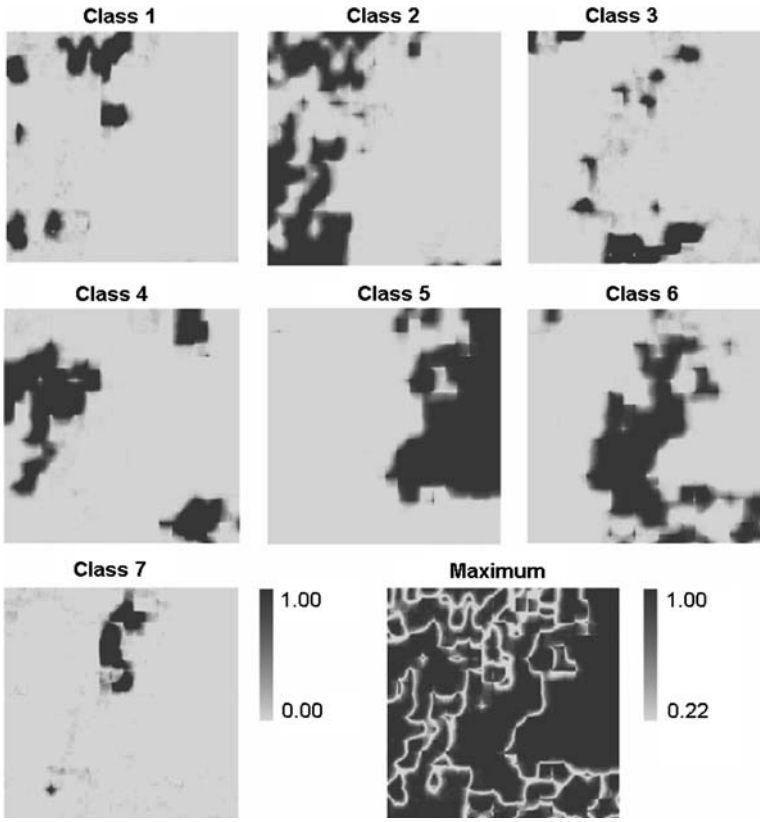

Fig. 10 Occurrence probability maps of land-cover classes generated by the MCRF model, conditioned on the moderate dataset

the remaining major deficiency of the previous model in small-class underestimation and that is proved by simulated results. The MCRF model uses one single Markov chain for a 2-D space; thus it differs from existing multi-D Markov chain models which normally assume multiple chains, one per direction (Carle and Fogg 1997). Because the MCRF model does not have the problem of excluding unwanted transitions, it does not underestimate small classes in conditional multi-D Markov chain simulations.

Simulations are conducted on three datasets with different sampling densities. The simulated results indicate that all classes, particularly small classes, are fairly reproduced in simulated realizations. One interesting phenomenon is that small classes are also well reproduced in prediction maps based on maximum occurrence probabilities. Simulated realizations show patterns imitative of the expected one—here the reference map. Transiogram analysis indicates that simulated transiograms effectively follow the original transiograms (experimental transiograms estimated from sampled data) both in curve shapes and curve heights. Occurrence probability maps demonstrated the spatial uncertainty associated with each classes and the prediction map.

Potentially, the time dimension may also be added in the future for studying spatio-temporal change of multinomial classes. Future effort will focus on: (1) developing a practical software package for further evaluation of MCRF-based models and their applications, and (2) expanding functions of the current simulation approach, for example, dealing with various sample data types and correlating with secondary information.

References

Bierkens MFP, Burrough PA (1993a) The indicator approach to categorical soil data: I. Theory. J Soil Sci 44:361–368

Bierkens MFP, Burrough PA (1993b) The indicator approach to categorical soil data: II. Application to mapping and land use suitability analysis. J Soil Sci 44:369–381

Bogaert P (2002) Spatial prediction of categorical variables: the Bayesian maximum entropy approach. Stoch Environ Res Risk Assess 16:425–448

Carle SF, Fog GE (1996) Transition probability-based indicator geostatistics. Math Geol 28:453–477

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

Chiles J-P, Delfiner P (1999) Geostatistics—modeling spatial uncertainty. Wiley, New York, 695 p

Christakos G (1990) A Bayesian/maximum-entropy view to the spatial estimation problem. Math Geol 22:763–777

Christakos G (2000) Modern spatiotemporal geostatistics. Oxford University Press, New York, 312 p

De Gruijter JJ, Brus DJ, Bierkens MFP, Knotters M (2006) Sampling for natural resources monitoring. Spinger, Berlin, 332 p

D'Or D, Bogaert P (2004) Spatial prediction of categorical variables with the Bayesian maximum entropy approach: the Ooypolder case study. Eur J Soil Sci 55:763–775

Deutsch CV (2006) A sequential indicator simulation program for categorical variables with point and block data: BlockSIS. Comput Geosci 32:1669–1681

Deutsch CV, Journel AG (1998) GSLIB: Geostatistics Software Library and User's Guide. Oxford University Press, New York, 369 p

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

Goodchild MF, Sun G, Yang S (1992) Development and test of an error model for categorical data. Int J Geogr Inf Syst 6:87–104

Goovaerts P (1996) Stochastic simulation of categorical variables using a classification algorithm and simulated annealing. Math Geol 28:909–921

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

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

Journel AG (1983) Nonparametric estimation of spatial distributions. Math Geol 15:445–468

Kyriakidis PC, Dungan JL (2001) A geostatistical approach for mapping thematic classification accuracy and evaluating the impact of inaccurate spatial data on ecological model prediction. Environ Ecol Stat 8:311–330

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 (in press)

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

Mark DM, Csillag F (1989) The nature of boundaries on the 'Area-class' maps. Cartographica 26:65–78

de Marsily Gh, Delay F, Goncalves J, Renard Ph, Teles V, Violette S (2005) Dealing with spatial heterogeneity. Hydrogeol J 13:161–183

Miller J, Franklin J (2002) Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. Ecol Modell 157:227–247

Mowrer HT (1997) Propagating uncertainty through spatial estimation processes for old-growth subalpine forests using sequential Gaussian simulation in GIS. Ecol Modell 98:73–86

Sharp WE, Aroian LA (1985) The generation of multidimensional autoregressive series by the herringbone method. Math Geol 17:67–79

Weissmann GS, Fogg GE (1999) Multi-scale alluvial fan heterogeneity modeled with transition probability geostatistics in a sequence stratigraphic framework. J Hydrol 226:48–65

Zhang J, Goodchild MF (2002) Uncertainty in geographical information. Taylor & Francis, New York, p 266