Comparing a Fixed-Path Markov Chain Geostatistical Algorithm with Sequential Indicator Simulation in Categorical Variable Simulation from Regular Samples

Chuanrong Zhang1 and Weidong Li
Department of Geography, Kent State University, Kent, Ohio 44242

Abstract: The Markov chain random field (MCRF) theory provided the theoretical support for a new, nonlinear Markov chain geostatistic. This study compares a MCRF-based netting-and-mesh-filling (NMF) algorithm with indicator kriging simulation algorithms and demonstrates their differences in simulating categorical variables from regular samples. The MCRF-NMF algorithm and indicator kriging simulation algorithms—SIS-SIK (sequential indicator simulation with simple indicator kriging) and SIS-OIK (SIS with ordinary indicator kriging)—are all applied to the same datasets. Results show that: (1) MCRF-NMF generates higher simulation accuracy in realizations and lower spatial uncertainties than SIS-OIK does; for a medium dataset, the former achieves a relative increase of 13.2% in simulation accuracy of simulated realizations; (2) MCRF-NMF obeys interclass relationships in simulated maps but SIS-SIK and SIS-OIK do not; and (3) MCRF-NMF generates polygons in simulated realizations, but SIS-SIK and SIS-OIK typically generate dispersed patterns. An imperfection with MCRF-NMF is that simulated polygons in realizations exhibit rough boundaries. However, optimal results based on maximum occurrence probabilities provide an attractive prediction map. Although MCRF-NMF has some limitations as a fixed-path algorithm, its advantage for improving simulation accuracy of simulated realizations makes it valuable.

INTRODUCTION

In determining spatial patterns of complex categorical variables (i.e., multinominal classes) such as soil types and land cover classes, interclass relationships are crucial. Interclass relationships may include cross-correlations, neighboring relationships (i.e., juxtaposition), and directional asymmetries of class sequences (Li and Zhang, 2005). For example, the spatial features of soil types are revealed in their complex spatial interdependencies, specific neighboring relationships, diverse polygon geometries, site-specific sequences, and anisotropies. These complexities pose challenges to conventional geostatistics in effectively capturing the spatial patterns of soil types at watershed scales from limited survey data because of the difficulty in incorporating interclass relationships. For example, if interclass relationships are ignored, non-neighboring classes may occur as neighbors in simulated results. Bogaert (2002) extended the Bayesian Maximum Entropy (BME) approach proposed

1Email: czhang2@kent.edu

251

Copyright © 2007 by V. H. Winston & Son, Inc. All rights reserved.
by Christakos (1990) to deal with categorical variables and pointed out that the linearity of indicator kriging estimators was also an important reason for the unsatisfactory performance of indicator kriging (IK) in modeling multinomial classes. While being relatively time-consuming in computation, the BME approach is nonlinear and can incorporate cross-correlations between classes. Recently, D’Or and Bogaert (2004) applied the BME approach to water table classes and demonstrated advantages of the approach over IK. For example, the BME approach could generate higher maximum occurrence probabilities than IK.

Multi-dimensional Markov chain models have been used in conditional simulation of categorical variables in recent years. For example, Carle and Fogg (1997) combined unconditional three-dimensional Markov chain models (interpolated from one-dimensional Markov chain models in principal directions) with indicator kriging and simulated quenching for conditional simulation of hydrofacies (see Weissmann and Fogg, 1999). The continuous-lag transition probabilities used in the approach were derived from one-step transition probabilities and facies mean lengths. Because they were based on the stationary first-order Markovian assumption, such transition probability curves were also called “idealized” transiograms in Li (2006), which suggested the transiogram to be a spatial relationship measure for categorical data.

To extend Markov chains into a Markov chain–based geostatistic for conditional simulation of categorical variables, many issues had to be solved. Major issues included pattern inclination, transition probability estimation from samples, direct conditioning to sample data, small-class underestimation, theoretical generalization, and algorithm development. Li et al. (2004) suggested a triplex Markov chain (TMC) approach to simulate soil type spatial distribution from survey data. The TMC method avoided the pattern inclination problem in two-dimensional Markov chain simulation by using an alternate advancing path. The proposition of the transiogram concept solved the problem of estimating continuous-lag transition probability models from sparse point samples and thus generalized the TMC method for working with grid point data (Li and Zhang, 2005, 2006). Because transiogram models can provide transition probabilities with any lags, the limitation of sampling intervals for one-step transition probability matrix estimation on resolution of Markov chain simulation was also eliminated. More recently, Li (2007) proposed a single-chain-based Markov chain random field (MCRF) theory. The MCRF theory solved the small-class underestimation problem that usually intrinsically occurred in multi-chain–based Markov chain methods, and also theoretically generalized Markov chains into a new, nonlinear geostatistics, outside conventional clique-based Markov random field (MRF) methods, covariance-based geostatistical methods (e.g., kriging), and the BME approach. As to the problem of direct conditioning to sample data, it was gradually addressed in the developing process of the Markov chain conditional simulation approach, and finally completely solved by the MCRF theory. With the development of practical simulation algorithms, the MCRF-based Markov chain approach will, therefore, become a new practical geostatistical approach for categorical variables and possibly for continuous variables in the future.

The objective of this research was to conduct a comparative study between the MCRF-based netting-and-mesh-filling (NMF) simulation algorithm and the widely used IK simulation algorithms—SIS-OIK (sequential indicator simulation with ordinary indicator kriging) and SIS-SIK (SIS with simple indicator kriging), to
demonstrate the differences between the two approaches in simulated the results of categorical variables from regular samples. It also introduces the MCRF-based Markov chain approach to the field of geographical information science. All of the simulation algorithms are applied to the same datasets. Case simulations are conducted on regular point datasets at a watershed scale.

**METHODS**

Markov Chain Random Field

The Markov chain random field (MCRF) refers to a special random field built on a single Markov chain (Li, 2007). It provided the theoretical foundation for a new, nonlinear geostatistics, called Markov chain geostatistics. Because a MCRF contains only a single Markov chain, it avoids exclusion of unwanted transitions. Thus, MCRF-based models overcome the small-class underestimation problem that usually occurred in the conventional multi-chain-based Markov chain methods, such as the TMC method suggested by Li et al. (2004), which provided an imperfect estimator for Markov chain simulation of categorical variables and works well only when samples are relatively abundant.

MCRF-based models generally differ from previous multi-dimensional conditional and unconditional Markov chain models that were built on multiple chains, because they are single chain—based models. Furthermore, MCRF-based models also differ from conventional MRF models (e.g., Besag, 1986; Norberg et al., 2002), which used cliques in simulations and could not condition directly a simulation on observed data, and Markov mesh (or chain) models for image processing (e.g., Qian and Titterington, 1991; Gray et al., 1994), which also used cliques and could do only unconditional simulations. On the contrary, MCRF-based models are natural conditional models, similar to kriging.

The conditional probability distribution of a MCRF $Z$ at an unknown location $u$ is derived as

$$
\Pr(Z(u) = k | Z(u_1) = l_1, \ldots, Z(u_m) = l_m) = \frac{\prod_{i=2}^{m} p_{kl}^{i}(h_i) \cdot p_{l_1(k)}^{1}(h_1)}{\sum_{f=1}^{n} \left[ \prod_{i=2}^{m} p_{lf}^{i}(h_i) \cdot p_{l_1(f)}^{1}(h_1) \right]} \tag{1}
$$

(Li, 2007), where $p_{kl}^{i}(h_i)$ represents a transition probability in the $i$th direction from state $k$ to state $l$ 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 $S = \{1, \ldots, n\}$; and $h_i$ is the distance from the current location to its nearest known neighbor $u_i$. With increasing lag $h$, any $p_{kl}(h)$ forms a transition probability diagram—a transiogram.
Equation (1) is the general solution of the MCRF. In real applications, considering only the four nearest known neighbors in cardinal directions (Fig. 1B) is sufficient for simulation of multinomial classes in horizontal two dimensions. Thus, Eq. (1) is simplified to

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

where, \(h_1, h_2, h_3, \) 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 simplified MCRF at the five locations \(u, u_1, u_2, u_3, \) and \(u_4,\) respectively. It should be noticed that Eq. (2) is not a multiplication of four one-dimensional Markov chains.

When using regular samples, simulation of outer boundaries only needs to consider two nearest known neighbors along one line (Fig. 1A). Thus, the above Eq. (2) can be simplified as

\[
p_{k|lm} = \Pr(Z(u) = k|Z(u_1) = l, Z(u_2) = m) = \frac{p_{km}^2(h_2) \cdot p_{lk}^1(h_1)}{\sum_{f=1}^{n} [p_{fm}^2(h_2) \cdot p_{lf}^1(h_1)]}
\]
Note that $u_1$ always refers to the start location of the Markov chain to the unknown location to be estimated. Such a location differs from other nearest known neighbors and should always exist.

**Transiogram Model Estimation**

To obtain continuous transiogram models from discontinuous experimental transiograms (which are directly estimated from sample data), one may construct mathematical models to fit experimental transiograms (Li and Zhang, 2006) or just interpolate experimental transiograms into continuous models (Li and Zhang, 2005). The constraint conditions of transiogram modeling for simulation are easy to meet—normally only two constraint conditions need to be met: (1) non-negativeness; and (2) at any lag $h$, values of transiogram models headed by the same class sum to 1. In addition, for exclusive classes, the nugget effect is irrational to transition probabilities and should not be considered in transiogram modeling.

When samples 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, and experimental transiograms may be merely used as a reference. However, when sampled data are sufficient and distributed relatively uniformly, experimental transiograms are usually 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. Thus, linear interpolation of experimental transiograms can be a feasible way for joint transiogram modeling in Markov chain simulation. In this study, we use the interpolation method introduced in Li and Zhang (2005) to acquire transiogram models from experimental transiograms.

**Simulation Algorithm**

Consider we simulate the spatial distribution of $n$ mutually exclusive classes, which is conditional to a sample dataset at $N$ grid nodes. The $n$ classes may be in any random sequence, denoted as any labels 1, 2, …, $n$ or A, B, C, …. Monte Carlo sampling is used to draw labels from the cumulative conditional probability distribution for each node. So the MCRF-NMF simulation algorithm simply consists of the following steps:

1. **Step 1.** Simulate outer boundaries using Eq. (3).
2. **Step 2.** Connect all neighboring sample points into data lines through simulation using Eq. (2), so that simulated lines form a network.
3. **Step 3.** In a mesh formed by simulated lines, simulate unknown nodes row by row from top to bottom still using Eq. (2).
4. **Step 4.** Proceed to next mesh (usually along the ordering from top to bottom and from left to right in the network) until all of meshes in the network are filled and a realization is produced.
5. **Step 5.** Repeat steps 1 to 3 to generate another realization.

In a simulation, transition probabilities needed at any lags are directly drawn from transiogram models. To avoid the pattern inclination (i.e., diagonal trend) effect that was met in extending different unilateral processes into multi-dimensions (Sharp and
Aroian, 1985; Gray et al., 1994; Li et al., 2004), simulations should use the alternate advancing path used in Li et al. (2004), which required a simulation be performed lattice row by lattice row along opposite direction alternately. It should be noted that the alternate advancing path is identical to the path of the herringbone method suggested by Sharp and Aroian (1985) for extending autoregressive processes into two dimensions.

Sequential Indicator Simulation

Sequential Indicator Simulation (SIS) have been widely used and well-documented in the literature (e.g., Goovaerts, 1997; Deutsch and Journel, 1998). SIS is an efficient indicator geostatistical simulation algorithm and it may use either of simple indicator kriging (SIK) and ordinary indicator kriging (OIK) as its estimator. The difference between SIK and OIK is that the former considers the indicator mean known and constant in the whole study area, while the latter allows one to account for local variation of the indicator mean. The following are the steps in SIS for simulation of spatial distribution of $K$ mutually exclusive classes $s_k$ conditional to the data set $\{s(u_\alpha), \alpha = 1, \ldots, n\}$ (Goovaerts, 1997):

1. Transform each class datum $s(u_\alpha)$ into a vector of $K$ hard indicator data, defined as

\[
i(u_\alpha; s_k) = \begin{cases} 
1 & \text{if } s(u_\alpha) = s_k, \\
0 & \text{otherwise}
\end{cases} \quad k = 1, \ldots, K.
\]  

2. Define a random path visiting each node of the lattice discretizing the study area once.
3. At each node $u'$:
   1. Determine the conditional probability of occurrence of each class $s_k$, $[p(u'; s_k(n))]^*$, using simple or ordinary indicator kriging.
   2. Correct these probabilities for order relation deviations.
   3. Define any ordering of $K$ classes and build a conditional distribution function (CDF) by adding the corresponding probabilities of occurrence; e.g.,

\[
[F(u'; s_{k'}(n))]^* = \sum_{k=1}^{K'} [p(u'; s_{k'}(n))]^* \quad k^* = 1, \ldots, K.
\]  

4. Draw a random number $p$ uniformly distributed in $[0, 1]$. The simulated class at location $u'$ is the one that corresponds to the probability interval that includes $p$: $s^{(l)}(u') = s_k$, such as $[F(u'; s_{k-1}(n))]^* < p \leq [F(u'; s_k(n))]^*$.
5. Add that simulated value to the conditioning data set.
6. Proceed to the next node along the random path, and repeat steps 1–5.

• Repeat the entire sequential procedure with a different random path to generate another realization.
Compared with SIS, which deals with single classes one by one, Markov chain algorithms always manage all classes simultaneously with consideration of interclass relationships. Similar to SIK, the MCRF model used in the NMF algorithm considers the mean (i.e., area proportion) of each class constant in the whole study area.

MATERIALS

Data Sets

We chose a study area in a floodplain in northeastern Iowa County, Wisconsin. The 9 km² square area has seven soil types (soil series) (Table 1). A soil series map of the study area is shown in Figure 2A. This map was assumed to be correct and served as a reference map to validate the simulated results. The study area was discretized into a grid of 22,400 (i.e., 175 × 128) pixels with a pixel size of 20 m.

In the study area, three regular datasets—a dense one of 660 points (i.e., 2.9% of total pixels), a medium one of 180 points (i.e., 0.8% of total pixels), and a sparse one of 48 points (i.e., 0.2% of total pixels)—were sampled (see Figs. 2B–2D). These datasets are consistent with the reference soil map. In fact, all the three datasets are sparse samples. Here the words dense, medium, and sparse are used to differentiate the three datasets with different densities. Using these three datasets, we may examine the differences between simulated results conditioned on different densities of samples.

Parameters

Experimental omnidirectional transiograms were estimated from the dense dataset and the medium dataset, respectively. Using omnidirectional transiograms means that we did not consider anisotropies and directional asymmetries of soil type patterns. Thus, only one set of transiogram models was sufficient for conducting simulation. But neighboring relationships between soil types were still considered...
because omnidirectional cross-transiograms are asymmetric. These experimental transiograms were composed of a series of points, and they were interpolated into continuous transiogram models for use in simulations. Figure 3 shows a subset of interpolated transiogram models headed by soil type 7. Because the sparse dataset was too small to generate reliable experimental transiograms, we used the transiogram models estimated from the dense dataset. In real applications, expert knowledge may be used to build transiogram models for the sparse dataset.

We estimated experimental omnidirectional indicator variograms only from the dense dataset. Experimental indicator autovariograms were further fitted with mathematical models and the set of fitted models were used in indicator simulations conditioned on each of the three datasets. Parameters of the seven estimated indicator autovariogram models are given in Table 2. We did not estimate indicator variogram models from the moderate dataset and the sparse dataset because SIS only accepts mathematical autovariogram models and the models estimated from the dense dataset can be regarded as the best. Indicator cross-variogram models were not required in SIS because indicator cokriging of multiple classes is infeasible in applications. In fact, it is
difficult to coregionalize a set of permissible indicator auto/cross-variogram models for multinomial classes because they do not share the same set of basic structures.

RESULTS AND ANALYSES

Simulated Results

Figure 4 shows three simulated realizations generated by MCRF-NMF, SIS-OIK, and SIS-SIK, conditioned on the dense dataset. Clearly the major soil patterns shown
in the reference soil map are well captured by MCRF-NMF. It is normal that some details shown in the reference map are missing because the conditioned samples are limited and the reference soil map may be delineated with detailed field observation and expert interpretation of the mapmakers. Compared with realizations generated by SIS, we find that the SIS-OIK realization (Fig. 4B) is close to the MCRF-NMF realization (Fig. 4A) in the spatial distribution of soil types, but its pattern is obviously dispersed. This means that SIS-OIK cannot generate polygons (i.e., crisp geometries). The SIS-SIK realization (Fig. 4C) is apparently inferior to the SIS-OIK realization because different soil types mix together and the spatial patterns become a mess in the former.

With a decreasing number of point samples, simulated realizations by both approaches become less imitative of the reference map (see Figs. 5 and 6). Situations similar to those in Figure 4 also occur in realizations conditioned on the medium dataset and the sparse dataset. In Figure 6, the SIS-SIK realization conditioned on the sparse dataset (Fig. 6C) has completely lost the general pattern of the spatial distribu-

Table 2. Indicator Autovariogram Models of the Seven Soil types, Estimated from the Dense Dataset\(^a\)

<table>
<thead>
<tr>
<th>Soil type</th>
<th>Model</th>
<th>Nugget</th>
<th>Sill</th>
<th>Range (pixel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Exponential</td>
<td>0.01</td>
<td>0.15</td>
<td>70</td>
</tr>
<tr>
<td>2</td>
<td>Exponential</td>
<td>0.04</td>
<td>0.16</td>
<td>70</td>
</tr>
<tr>
<td>3</td>
<td>Exponential</td>
<td>0.03</td>
<td>0.093</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>Exponential</td>
<td>0.021</td>
<td>0.067</td>
<td>115</td>
</tr>
<tr>
<td>5</td>
<td>Exponential</td>
<td>0.003</td>
<td>0.02</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>Exponential</td>
<td>0.04</td>
<td>0.105</td>
<td>55</td>
</tr>
<tr>
<td>7</td>
<td>Exponential</td>
<td>0.05</td>
<td>0.3</td>
<td>180</td>
</tr>
</tbody>
</table>

\(^a\)660 regular points; pixel size is 20 m.

Fig. 4. Simulated realizations generated by MCRF-NMF (A), SIS-OIK (B), and SIS-SIK (C), respectively, and conditioned on the dense dataset (i.e., 660 points).
tion of soil types; however, the MCRF-NMF realization (Fig. 6A) and the SIS-OIK realization (Fig. 6B) still retain the pattern well.

A special feature demonstrated in transiograms should be noticed: class 7 has no chance to be immediate neighbors of class 4 and class 5 (see cross-transiograms $p_{74}(h)$ and $p_{75}(h)$ in Fig. 3). Checking the realizations generated by MCRF-NMF, we find that these interclass rules are strictly followed no matter how many samples are conditioned. But in realizations generated by SIS, class 7 touches class 4 and class 5 or even mixed with them together in the sparse-dataset case. This means non-neighboring classes may occur as neighbors in simulated realizations of SIS because of the neglect of cross-correlations.

Although alternative realizations are usually used to demonstrate the uncertainty of class spatial distribution, it is desirable that simulated realizations can better capture each class at their approximate locations with approximate polygon shapes. Users may want to use realizations, rather than an optimal interpolation map to be their reference data or input data to their studies because of the uneven smoothing effect existing in an interpolation map (Goovaerts, 1996). This is also the purpose of multiple-point geostatistics in reservoir modeling (Caers and Zhang, 2004). However,
multiple-point geostatistics is inappropriate for complex categorical variables because of the difficulty of obtaining suitable training images for acquisition of multiple-point statistics. Table 3 shows the PCC (percentage of correctly classified locations) values estimated from simulated realizations generated by MCRF-NMF and SIS-OIK. Apparently, the PCC values from MCRF-NMF realizations are much higher than those from SIS-OIK realizations. For the dense dataset and moderate dataset cases, relative differences remarkably reach above 10%. This means that, given the same sample dataset, MCRF-NMF is more capable than SIS in capturing soil types at their correct locations in the simulated realizations—that is, MCRF-NMF can generate realizations with higher simulation accuracy.

An imperfection with simulated realizations of MCRF-NMF is that simulated polygons exhibit rough boundaries. This may be related to the line-by-line simulation algorithm and the grid form of sample data used in the simulations. Based on maximum occurrence probabilities estimated from a number of simulated realizations (here we used 100 realizations) optimal prediction maps were generated by both MCRF-NMF and SIS as shown in Figure 7. The polygon shapes in optimal prediction maps from MCRF-NMF look far better than those in simulated realizations.

### Occurrence Probability Maps

The fact that simulated realizations from MCRF-NMF have higher PCC values than those from SIS also implies that MCRF-NMF generates less spatial uncertainty than SIS does. To clearly display spatial uncertainty, a practical way is to use occurrence probability maps estimated from a number of simulated realizations. The maximum occurrence probability map generated by a simulation demonstrates the spatial uncertainty associated with the optimal prediction map, which may be comparable with a delineated map based on the same sample data. While the optimal prediction maps based on maximum occurrence probabilities generated by MCRF-NMF look

---

### Table 3. PCCs\(^a\) of simulated realizations\(^b\) from MCRF-NMF and SIS-OIK\(^c\)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Dense</th>
<th>Moderate</th>
<th>Sparse</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCRF-NMF</td>
<td>80.8</td>
<td>68.5</td>
<td>54.7</td>
</tr>
<tr>
<td>SIS-OIK</td>
<td>72.2</td>
<td>60.5</td>
<td>52.6</td>
</tr>
<tr>
<td>AD(^d)</td>
<td>8.6</td>
<td>8</td>
<td>2.1</td>
</tr>
<tr>
<td>RD(^e)</td>
<td>11.9</td>
<td>13.2</td>
<td>4</td>
</tr>
</tbody>
</table>

\(^a\)Percentages of correctly classified locations.

\(^b\)Averaged from 100 realizations.

\(^c\)Conditioned on the three datasets (dense = 660 points; moderate = 180 points; sparse = 48 points). PCC values are estimated relative to the reference soil map.

\(^d\)Absolute difference = (MCRF-NMF – SIS-OIK).

\(^e\)Relative difference = (AD/SIS-OIK) \times %.
similar to those generated by SIS-OIK (see Fig. 7) except that the former obeys inter-class relationships and the later may not, the spatial uncertainties associated with these maps are different.

Figure 8 shows the maximum occurrence probability maps generated by MCRF-NMF and SIS-OIK, conditioned on different datasets and estimated from 100 realizations for each simulation. It can be seen that both methods captured similar transition zones (low probability stripes). However, transition zones in maximum occurrence probability maps from MCRF-NMF are apparently clearer than those from SIS-OIK. This means that MCRF-NMF generates higher maximum occurrence probabilities at most unsampled locations and thus less spatial uncertainty associated with related optimal prediction maps. The average maximum probability (AMP) values shown in Figure 9 indicate this point. The AMP values from MCRF-NMF conditioned on the three different datasets are (0.07 to 0.14) higher than those from SIS-OIK. This also means that samples play more important roles in MCRF-NMF than in SIS-OIK in determining the occurrence of a class at an unknown location. This characteristic of MCRF-NMF indicates that its realizations have high credibility, as proved by the PCC values of realizations in Table 3. Besides the nonlinearity of the MCRF estimator, we think that this difference is caused by the incorporation of interclass relationships.
CONCLUSIONS

We compared a Markov chain geostatistics simulation algorithm (MCRF-NMF) with the widely used indicator kriging simulation algorithms (SIS-OIK and SIS-SIK) in conditional simulation of a complex categorical variable—soil types from regular samples. Simulated results indicated that MCRF-NMF was generally superior to SIS.
SIS-SIK was apparently not suitable for simulating categorical variables that are distributed with apparent non-stationarity. With the consideration of local variation of class mean values, SIS-OIK generated better results than SIS-SIK, and simulated patterns by SIS-OIK looked close to those generated by MCRF-NMF, of which the estimator used in this study assumed complete spatial stationarity, similar to SIK.

Comparing MCRF-NMF with SIS-OIK, we found that the former had the following advantages over the latter: (1) given the same dataset, MCRF-NMF generated realizations with apparently higher simulation accuracy; (2) in simulated maps, MCRF-NMF obeyed interclass relationships defined by spatial measure models but SIS could not; (3) given the same dataset, MCRF-NMF produced more certain spatial patterns and thus generated less spatial uncertainty associated with simulated results than did SIS-OIK; and (4) MCRF-NMF generated polygons but patterns generated by SIS were generally dispersed.

Because MCRF-NMF is a fixed-path algorithm and grid point samples were used, simulated polygon boundaries in realizations were not smooth. Despite these limitations, the advantage of the algorithm in generating higher simulation accuracy in realizations makes it valuable in practical applications. Please note that the MCRF model used in this study considered class proportions constant in the whole study area (similar to SIK). However, MCRF-NMF still generated better results than did SIS-OIK. Further study will consider accounting for the spatial non-stationarity of class proportions in MCRF models and thus it is expected that the well-developed MCRF-NMF algorithm will perform better than the current one.

REFERENCES


