Markov Chain Modeling of Multinomial Land-Cover Classes

Chuanrong Zhang
Department of Geography and Geology, University of Wisconsin, Whitewater, Wisconsin 53190

Weidong Li
Department of Geography, University of Wisconsin, Madison, Wisconsin 53706

Abstract: Quantitatively modeling the spatial distribution and uncertainties of land-cover classes is crucial in geographic information science. However, currently suitable methods for modeling multinomial categorical variables for geographical analyses are rare because of the complex spatial interdependence of multinomial classes. In this paper we use a recently presented two-dimensional Markov chain model to simulate multinomial land-cover classes and to estimate occurrence probability vectors for spatial uncertainty representation. Simulated results indicate that the model may provide a potential method for predictive mapping at higher resolutions possibly over large areas and for spatial uncertainty analyses of land-cover classes.

INTRODUCTION

Land-cover classes (or types) are traditionally mapped as area class maps (i.e., choropleth maps) (Mark and Csillag, 1989). Because of the imperfection of knowledge we can acquire about a mapping area in determining class boundaries, it has been recognized that the delineated (or interpreted) area class map inevitably contains spatial (or locational) uncertainties as discussed by Ehlschlaeger (2000), Stine and Hunsaker (2001), McKelvey and Noon (2001), and Atkinson and Foody (2002). In addition, the delineated area class map actually only represents one realization based on the available high quality dataset (e.g., field survey data, or easily discerned areas on remote sensing images). It is suggested that random field models may be used to represent the spatial uncertainty by generating multiple equally probable realizations through conditional stochastic simulation (stochastic imaging) and inferring occurrence probabilities of the studied categories (e.g., Goodchild et al., 1992; Journel, 1997; Chiles and Delfiner, 1999; Zhang and Goodchild, 2002). Such uncertainty data can be introduced into application models (e.g., ecological models) to get response distributions (called error propagation) for decision making and risk assessment (e.g., Heuvelink, 1998; Kyriakidis and Dungan, 2001; Li et al., 2001). Related spatial uncertainty studies associated with land-cover classes can be found in Kyriakidis and
Dungan (2001) and Zhang and Goodchild (2002). The main method used in these studies is the sequential indicator simulation (SIS) approach initially proposed by Journel (1983) for modeling cutoffs of continuous variables.

Unlike continuous variables that have continuous values varying slightly from one location to another, and also unlike cutoffs of continuous variables that are actually ordinal, multinomial classes (denoted as 1, 2, ..., n for n classes in an arbitrary sequence) of categorical variables such as land-cover classes occur interdependently in space as various patterns. Cross-correlations and juxtapositions between land-cover classes and directional asymmetries of their occurrence sequences may be prominent in many situations. However, in conventional geostatistical methods such as SIS, dealing with cross-correlations by indicator cokriging is quite cumbersome (Chiles and Delfiner, 1999) and accounting for directional asymmetries is actually impossible. The spatial measures used, cross variograms, are defined with the implicit assumption that variables are distributed symmetrically (Goovaerts, 1997). Therefore, cross-correlations between classes (i.e., class interdependence) are normally ignored in variogram-based indicator simulation algorithms (i.e., they only use indicator kriging, not cokriging) (Goovaerts, 1996, p. 911-912). There is no doubt that SIS is a well-developed method for modeling thresholds of continuous variables. But for modeling multinomial classes, computation and parameter estimation workload quickly increases with increasing number of classes and considering anisotropies (given that the observed data are sufficient for parameter estimation for different classes and in different directions). The computation load in other random field models proposed for modeling categorical variables, such as the correlated inter-map cell swapping heuristic (Ehlschlaeger, 2000) and the Bayesian Markov random field model (Norberg et al., 2002), is a more prominent restraining factor for modeling multinomial classes in large areas.

Recently, Li et al. (2004) presented a triplex Markov chain (TMC) model for modeling the spatial distribution of soil classes. The TMC model is highly efficient because it performs conditional simulation by direct conditioning with an explicit conditional probability function and generates realizations in a one-pass way. This feature provides the condition for dealing with a large number of classes in high resolution stochastic imaging. The second feature of the TMC model is that it may use survey line data, rather than point data, to do conditional simulation. This is very different from other spatial statistical models, and may provide an efficient way for better representing spatial continuities of land-cover class parcels. Spatial cross-correlations and directional asymmetries of multinomial classes are naturally included in Markov chain transition probability matrices (TPMs).

This research applies the TMC model to modeling land-cover classes. It introduces an efficient alternative for representing the spatial uncertainties of land-cover classes. The study focuses on: (1) estimation of probability vectors from multiple realizations and their visualizations to express the spatial uncertainty information of land-cover classes; and (2) comparison of indicator (cross) variograms estimated from the original land-cover map and simulated realizations to show whether spatial correlations (or patterns) of land-cover classes are reproduced. Section 2 introduces the method used. In section 3, a case study based on the land-cover classes in the Stone Forest National Park is provided to demonstrate the model’s feasibility in
modeling the spatial distribution of land-cover classes and to visualize the estimated probability vectors.

METHOD

Introduction of the TMC Model

The basic idea of the TMC model is that the occurrence of a spatial state depends on its nearest known neighbors in four orthogonal directions (Fig. 1). Two coupled Markov chains (CMCs) (Elfeki and Dekking, 2001), which consist of three single Markov chains, are used to implement the TMC model. The model conducts simulation by following a fixed path, row by row from top to bottom. Although the model is not suitable for unconditional simulation, it works well for conditional simulation if conditioning data is sufficient. Survey line data have been used as conditioning data in a TMC simulation of soil classes (Li et al., 2004).

Consider three independent one-dimensional Markov chains—$X_i$, $X'_i$, and $Y_i$—all defined on the same discrete state space $S$, where each state $l \in S$ and $l = 1, \ldots, n$. The $X_i$ and $Y_i$ represent two one-dimensional chains in the $x$-direction and the $y$-direction, respectively. But the $X'_i$ represents a one-dimensional chain in the $x'$-(i.e., anti-$x$-) direction (i.e., right-to-left if the $X_i$ chain is in the left-to-right direction). $Y_i$ and $X_i$ are coupled to form one CMC $Z_{ij}^L$, with superscript $L$ standing for the left-to-right direction, and $Y_i$ and $X'_i$ are also coupled to form the other CMC $Z_{ij}^R$, with superscript $R$ representing the right-to-left direction. The two CMCs proceed alternately on a two-
dimensional lattice domain in opposite directions (Fig. 1). Thus, a TMC is defined. The conditional joint transition probability of \( Z_{ij}^L \), with conditioning on future states in the \( x \) and \( y \) directions can be given as

\[
p_{lm,k\varnothing}^L = C \cdot p_{lkq}^x \cdot p_{mk\varnothing}^y = \frac{p_{lq}^x \cdot p_{kq}^{(N_x-i)} \cdot p_{mk}^y \cdot p_{\varnothing q}^{(N_y-j)}}{\sum f \left( p_{lq}^x \cdot p_{fq}^{(N_x-i)} \cdot p_{mf}^y \cdot p_{\varnothing f}^{(N_y-j)} \right)}
\]

where \( p_{lk} \) represents a one-step transition probability from state \( l \) to state \( k \). \( p_{kq}^{(N-i)} \) is a \((N-i)\)-step transition probability, which can be calculated by \((N-i)\) times of self multiplication of the one-step TPM. \( p_{lkq} \) is the probability of cell \((i, j)\) to be in state \( k \), given that the previous cell \((i-1, j)\) is in state \( l \) and the ahead cell \((N, j)\) is in state \( q \). Superscript \( x \) and \( y \) stand for directions. All subscripts stand for states belonging to the state space \( S \). \( C \) is a normalizing constant which is given as \((\sum f \cdot p_{lfq}^x \cdot p_{mf\varnothing}^y)^{-1}\). \( C \) arises because only the transitions to the same state \( k \) in the unknown cell are considered.

The right-to-left CMC \( Z_{ij}^R \) obeys a similar rule except for an opposite proceeding direction. Similarly, we can get \( p_{lm,k\varnothing}^R \) for \( Z_{ij}^R \). The conditional joint transition probability pair \((p_{lm,k\varnothing}^L, p_{lm,k\varnothing}^R)\) define a fully conditional TMC field model. Here the purposes of using two CMCs in opposite directions include: (1) to account for the asymmetric neighborhood problem of the CMC model so that simulation artifacts such as parcel inclination and discontinuities can be avoided; and (2) to effectively impose the influences of surrounding observed data to the unknown cell to be estimated (Li et al., 2004). Note that there is no requirement for what constitutes the \( x \) or \( y \) direction. Users may choose the \( x \) and \( y \) directions according to their arrangement of survey lines so that more lines are along the \( x \) and \( y \) directions.

**Conditioning Data Format and Parameters**

The TMC model is designed to use survey line data to do conditional simulation of “observable” multinomial categorical variables, although it has no limitation on other data types. There are three reasons for using survey line data: (1) A transect survey is a traditional survey method used for field survey of observable natural phenomena such as soil types or land-cover classes and for making choropleth maps. Surveyors only need to observe and record the class boundary changes along a line (straight line or curve) and all the information along the whole survey line is acquired (Fig. 2). Thus, using survey line data dramatically increases the number of conditioning data points with a relatively smaller workload, compared to observing many scattered points. Of course, such a data type is not realistic for modeling continuous variables or non-observable categorical variables. (2) Survey line data is beneficial for modeling categorical variables because the spatial continuities of class parcels can be more effectively represented by survey line data. Point data may not have this
effect. (3) Parameters used in this study, i.e., single-step Markov TPMs, can be estimated directly from survey line data if survey lines are regular. Estimating single-step TPMs indirectly from scattered point data is possible but relatively difficult because of the spatial discontinuity of point data. Current research is focused on developing a methodology to estimate continuous transition probability curves with model fitting from point data so that the model is more versatile in the future. Single-step TPMs may also be estimated using expert knowledge, but that will inevitably bring subjectivity into modeling. Such ambiguity is hopefully avoided, especially when evaluating a method.

High quality survey line data may also be discerned partially from remote sensing images, where some places show clear boundaries. These survey lines needn’t cross the entire study area; line segments also can be conditioned. The TMC model, when simulating the current cell on a grid, actually only requires two nearest surveyed cells as conditioning data. If outer boundaries of a study area are known, the interior observed data can be in various formats (points, lines, even small areas or a mixture). The resemblance of simulated realizations to the real situation in terms of both locations and class pattern detail will depend on the conditioning data available.

If we do not consider anisotropies and asymmetries of the distribution of classes, one single-step TPM is sufficient for conducting simulation using the TMC model. But normally the input parameters for the TMC model are three single-step TPMs in the $x$, $x’$, and $y$ directions so that anisotropies and asymmetries in the $x$ and $y$ directions are incorporated. If survey lines are sufficient, TPMs estimated from survey lines can be used directly. To estimate a TPM, for example, in the $x$ direction, the transition frequency matrix (TFM) in the $x$ direction needs be estimated first from the survey lines in the $x$ direction by counting the frequencies of a given state (e.g., $S_i$) followed by itself or the other states (e.g., $S_j$) in the direction on the lattice. A single-step TPM can be calculated from the corresponding TFM by dividing each entry of the TFM with the row total value. Note that the TFM in the $x$-direction is the transpose matrix of the TFM in the $x’$-direction. This means that if we get the TFM in the $x$-direction, we also get the TFM in the $x’$-direction.

If survey lines are insufficient (or most are curves), TPMs estimated from the survey lines in the axis directions may not be reliable. Under this situation, expert knowledge may have to be used to adjust the TPMs. Owing to the intuitiveness and interpretability of single-step Markov chain transition probabilities, adjusting transition probabilities based on expert knowledge is easier than adjusting variograms that are used in conventional indicator geostatistical approaches. In addition, it is also

![Fig. 2. Illustration of a survey line across several land-cover class boundaries.](image-url)
possible to estimate the transition probabilities from the exhaustive soft information interpreted from remote sensing images or previously hand-delineated maps, or even analogous known areas. Although these land-cover class maps contain spatial uncertainty, the soft information contained in the maps, namely the spatial dependence relationships of multinomial classes, should be feasible for parameter estimation if they can reflect the actual spatial correlations. One reason is that the TPMs in the TMC model only serve as spatial measures, i.e., they represent the spatial dependence relationships, and do not contain locational information about any land-cover classes. The locations of classes in simulated realizations are decided only by conditioning data. The second reason is that if an interpreted map could be regarded as equivalent to a realization, it also should bear the similar spatial statistical estimates (e.g., variograms and proportions).

Note that while interpreted land-cover maps contain locational errors, there is no way to state that maps interpreted from remote sensing images or limited survey data are low quality or model-simulated maps are high quality. In fact, a well interpreted map should represent a good realization of the high quality data configuration based on the interpretation that the map creator made (with subjectivity). On the contrary, the quality of a simulated realization not only depends on the quality of conditioning data and parameters, but also depends on the quality and ability of the random field model used.

Probability Vectors

The concept of probability vectors is used to represent the spatial uncertainty of multinomial land-cover classes. Mark and Csillag (1989) and Goodchild et al. (1992) discussed using probability vectors to represent the transition between two classes if there are cartographic (or locational) errors. According to Goodchild et al. (1992), a probability vector is a set of values representing the likelihood that each class is found in a particular grid cell. A probabilistic multinomial class map can be defined as a set of maps, one map for each class, where each class map’s cell represents the probability that the class is in that cell. Edges between classes are represented by “transition zones” appropriate to the multinomial class map’s locational uncertainty. Because the TMC model can easily produce many equally probable realizations, a probability vector for all classes at each location (i.e., grid cell) can be readily calculated. Let a realization be recorded as an indicator map for class \( k \) as

\[
I_k(z) = \begin{cases} 
1, & u_z = k \\
0, & \text{otherwise} 
\end{cases} 
\]

where \( I_k \) is the indicator of class \( k \) and \( u_z \) is the value of the class at location \( z \). From \( N \) realizations, we can get the probability of the \( k \)th class occurring at location \( z \) as

\[
p_k(z) = \frac{1}{N} \sum_{i=1}^{N} I_k^{(i)}(z), \quad i = 1, \ldots, N
\]
where \( i \) is the realization number. Thus, for each location \( z \) and all \( n \) classes, we get a probability vector:

\[
P(z) = (p_1(z), \ldots, p_k(z), \ldots, p_n(z)).
\] (4)

Probability vectors can be visualized by GIS tools as a series of probability maps for each class, where each cell of a probability map represents the possibility that a class will occur in the cell. By finding the maximum probability value at each grid cell and assigning the corresponding class to the cell, probability vectors can be “hardened” to a best-guess area class map, which stands for the optimal prediction map.

**CASE STUDY AND DISCUSSION**

**Data and Conditioning Schemes**

A simulation case study was conducted on a 5.9 × 5.9 km\(^2\) area of the Lunan Stone Forest National Park in Yunnan, China. Figure 3 shows the land-cover class map of the study derived from IKONOS imagery. Here we use it as a reference map for comparison with the simulated results—that is, we assume the remote sensing derived land-cover map is correct. The seven land-cover classes include: 1–built-up areas; 2–farmland; 3–forest; 4–highland; 5–Shilin with vegetation; 6–Shilin without vegetation; 7–water bodies (Zhang et al., 2003; Zhang, 2004).

The study area is composed of a 295 × 295 grid lattice with a cell size of 20 × 20 m. To test whether the TMC model is suitable for modeling land-cover classes,
two sets of conditioning data were used, a sparse set and a dense set. The sparse dataset consisted of 11 rows and 11 columns of survey lines, i.e., having a survey line interval of 600 m. The dense dataset consists of 21 rows and 21 columns of survey lines, i.e., having a survey line interval of 300 m. Three single-step TPMs in the \(x\), \(x'\), and \(y\) directions (i.e., south to north, north to south, and west to east, respectively) were used in each simulation. Single-step TPMs were directly estimated from the survey line data. Table 1 shows the single-step TPMs estimated from the two datasets. Diagonal transition probabilities (from the top left corner to the bottom right corner in each TPM) represent class autocorrelations, and off-diagonal transition probabilities represent class interdependences (i.e., juxtapositions and cross-correlations). Three single-step TPMs in different directions were used to account for directional asymmetries of land cover class sequences. Multi-step transition probabilities used in

### Table 1. Single-Step Markov Transition Probability Matrices (TPMs) Estimated from the Two Land-Cover Datasets

<table>
<thead>
<tr>
<th>Class</th>
<th>From the sparser dataset</th>
<th>From the denser dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>TPM in the x-direction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.878</td>
<td>.033</td>
</tr>
<tr>
<td>2</td>
<td>.013</td>
<td>.944</td>
</tr>
<tr>
<td>3</td>
<td>.004</td>
<td>.016</td>
</tr>
<tr>
<td>4</td>
<td>.004</td>
<td>.029</td>
</tr>
<tr>
<td>5</td>
<td>.000</td>
<td>.005</td>
</tr>
<tr>
<td>6</td>
<td>.005</td>
<td>.011</td>
</tr>
<tr>
<td>7</td>
<td>.020</td>
<td>.000</td>
</tr>
<tr>
<td>TPM in the anti-x-direction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.877</td>
<td>.068</td>
</tr>
<tr>
<td>2</td>
<td>.007</td>
<td>.953</td>
</tr>
<tr>
<td>3</td>
<td>.004</td>
<td>.021</td>
</tr>
<tr>
<td>4</td>
<td>.010</td>
<td>.026</td>
</tr>
<tr>
<td>5</td>
<td>.002</td>
<td>.004</td>
</tr>
<tr>
<td>6</td>
<td>.005</td>
<td>.016</td>
</tr>
<tr>
<td>7</td>
<td>.020</td>
<td>.000</td>
</tr>
<tr>
<td>TPM in the y-direction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.914</td>
<td>.033</td>
</tr>
<tr>
<td>2</td>
<td>.007</td>
<td>.963</td>
</tr>
<tr>
<td>3</td>
<td>.026</td>
<td>.005</td>
</tr>
<tr>
<td>4</td>
<td>.000</td>
<td>.019</td>
</tr>
<tr>
<td>5</td>
<td>.000</td>
<td>.001</td>
</tr>
<tr>
<td>6</td>
<td>.000</td>
<td>.014</td>
</tr>
<tr>
<td>7</td>
<td>.038</td>
<td>.038</td>
</tr>
</tbody>
</table>

\(a\)Number of classes = 7; grid columns = 295; grid rows = 295; grid-cell size = 20 m × 20 m.
Simulation were calculated from single-step transition probabilities by imposing a power of the number of steps to single-step TPMs (i.e., a single-step TPM multiplies itself $n$ times to generate a $n$-step TPM).

As mentioned above, survey line data may be acquired by recording the class boundary changes along a line. Therefore, it is not proper to simply compare the number of data points on these survey lines with the number of scattered survey points used in other random field simulations. The number of data points (i.e., grid cells) of survey lines is also related with the simulation resolution, i.e., the grid cell size.

**Simulated Realizations and Prediction Maps**

Simulations were performed on an a personal computer using FORTRAN 90 computer programs. One hundred realizations were produced for each conditioning dataset. The simulation time for generating each 100 realizations was about 3 hours. Figure 4 shows prediction maps (based on maximum occurrence probabilities) and two realizations for each conditioning dataset. We can see that the general land cover patterns are well displayed in these simulated realizations. Class parcel boundaries are clear and the anisotropies of land cover classes are retained. In particular, the realizations based on the dense conditioning dataset show remarkable resemblance with the reference land-cover map except for some fine detail. It also can be seen that realizations based on the dense dataset have more detail than those based on the sparse dataset. However, some minor land cover classes, such as land cover 7 (water bodies), are relatively underestimated and major land cover classes, such as land cover 2 (farmlands), have the tendency to be overestimated in realizations based on the sparse dataset. The prediction map is quite similar to the realizations, but with smoother parcel boundaries. The areal proportion data shown in Figure 5 provide a clear demonstration of the estimates of class areas. The average values from 100 realizations and the values from the single realizations and the prediction map are similar, which means that the model is very stable and provides similar realizations as long as the conditioning data are the same. They also show a similar tendency, i.e., minor classes are relatively underestimated and major classes are relatively overestimated when conditioning data are sparser. But when conditioning data are relatively dense, such characteristics lessen.

The reason that minor land-cover classes are underestimated is that the CMCs used in the TMC model have the tendency to underestimate minor components (Li et al., 2004). Minor land-cover classes in this simulation case are also some special land-cover classes. For example, land-cover class 1 is built-up area and class 7 is water bodies. People normally have easier accesses to these kinds of land-cover classes; therefore, it is possible to acquire more survey data about these land-cover classes. Thus, the underestimation problem of some minor land-cover classes may be mitigated in real world applications.

Underestimation of minor components is not unusual in other random field methods such as the Markov random field model as discussed by Norberg et al. (2002) or even indicator geostatistics as discussed by Soares (1998), when dealing with very sparse data. Different methods are suggested by Journel and Xu (1994), Goovaerts (1996), and Soares (1998) to overcome this problem. Simulated annealing as a flexible optimization technique has been used by Goovaerts (1997) for post-processing
Fig. 4. Simulated results of land-cover classes in the Stone Forest National Park based on the two datasets: $S =$ sparser dataset; $D =$ denser dataset. Prediction maps are based on the maximum probabilities calculated from 100 realizations. $R_{10} =$ tenth realization; $R_{50} =$ fiftieth realization.
SIS images and by Carle (1996) for further processing the simulated realizations of the transition probability-based geostatistics. If simulated annealing is applied to post-processing of the realizations of our simulation, the underestimation problem of minor components may be overcome or largely mitigated. But that would largely erode the efficiency of the TMC model. Therefore, further study is necessary to find a better way to eliminate this problem.

Additionally, stationary Markov chains are used in this model. The stationary assumption is rarely true in real applications. Undoubtedly, this is also one of the reasons for the imperfect simulation. For large areas, land covers usually have very different patterns in different sub-regions. To effectively reflect this pattern difference, non-stationary Markov chains, e.g., using different sets of TPMs for different sub-regions, may produce better realizations. Of course, this will increase the work of parameter estimation.
Autocorrelations and Cross-correlations

Realization maps only show visual and qualitative results. To know whether or not the spatial dependence relationships (or the spatial patterns) of land-cover classes are effectively generated in simulated realizations, comparison of quantitative spatial measures between the reference map and simulated realizations is necessary. Direct and cross variograms are widely accepted spatial correlation measures used in geostatistics for representing autocorrelations and cross-correlations. We calculated indicator direct and cross variograms from the reference map and simulated realizations based on the two datasets. Part of the results are presented in Figures 6 and 7.

From Figure 6, it can be seen that the omni-directional direct variograms estimated from the reference land-cover map have diverse forms. However, the realization based on the dense dataset (i.e., Realization D-R10 in Fig. 4) reproduces them exactly, except for the land-cover class 3 (forest), which is also a minor land-cover class that is relatively underestimated. The direct variograms estimated from the sparse dataset realization (i.e., Realization S-R10 in Fig. 4) show little departure from the reference results, but still have the same tendencies. In general, autocorrelations are approximately reproduced in simulated realizations.

There are 21 omni-directional cross variograms associated with the seven land-cover classes for each image. Only a few of them are shown in Figure 7. Similarly, with direct variograms, it can be seen that the simulated results approximate the original ones well, with the dense dataset realization doing a relative better job than the sparse dataset realization.

Such matches shown here between a reference map and simulated realizations are actually remarkable, because the forms of some of the variograms estimated from the reference map are not in the usual form of variogram models (e.g., spherical, exponential, power, etc.) recommended in classical geostatistical books such as Deutsch and Journel (1992) and Goovaerts (1997). Therefore, to fit these variograms with variogram models would be difficult, let alone reproduce them using conventional indicator approaches. This means the TMC model does have the ability to reproduce the complex spatial patterns of land-cover classes in conditional simulation when supplied adequate conditioning data. It also shows that sparse data will generate larger deviations in simulated results using the TMC model.

Probability Maps

The spatial uncertainty of land cover classes can be represented by their occurrence probabilities at each cell, displayed as probability maps. Figures 8 and Figure 9 show such probability maps of seven land-cover classes and the maximum probability maps based on the sparse and dense datasets, respectively. These probability maps not only display the coincidence and difference of occurring locations of land-cover classes in multiple realizations; more importantly, they represent the possible occurrence locations, extents, and patterns of land cover classes predicted from field survey data (or high quality data) by the random field model used. In each probability map, black indicates areas where the land-cover class absolutely occurs (probability is 1.0), grey means areas where the land cover class is uncertain (i.e., probability is >0.0 but <1.0), and white (here using light grey to differentiate from the background color) represents areas where the class never occurs (i.e., probability is 0.0). We can see that
Fig. 6. Direct variograms of land-cover classes estimated from the reference land-cover map and simulated realizations based on two datasets. Or = reference land-cover map; SR = tenth realization based on the sparse dataset. DR = tenth realization based on the dense dataset. Numbers in chart legends (e.g., Or-5) represent land-cover classes (corresponds to the class number in Fig. 3).
grey exists between black and white, which shows transition zones from the class to other classes. The extent of transition zones is related to the density of conditioning data. It can be seen that the transition zones in the probability maps based on the denser dataset are relatively less obvious than those based on the sparser dataset. These transition zones are appropriate to represent the positional uncertainty or class parcel boundary uncertainty in multinomial area class maps as suggested by Mark and Csillag (1989). “The degree of fuzziness of the boundary (an object concept) can thus be represented in the steepness of the probability gradient (a field concept)” (Goodchild et al., 1992, p. 91).

The maximum probability maps based on the two datasets (Figs. 8 and 9) provide a clearer display of transition zones between class parcels (see the white to light grey stripes). Comparing the two maximum probability maps based on the sparser dataset

**Fig. 7.** Cross variograms of land-cover classes estimated from the reference land-cover map and simulated realizations based on two datasets. Or = reference land-cover map; SR = tenth realization based on the sparse dataset; DR = tenth realization based on the dense dataset. Numbers in chart legends (e.g., Or-5) represent land-cover classes (corresponds to the class number in Fig. 3).
and the denser dataset, we can see that increasing the density of conditioning data causes the extent of transition zones to decrease. Such uncertainty information is helpful for our understanding of the spatial uncertainty of categorical variables and also for decision making and risk assessment that use area class maps as input data.

CONCLUSIONS

Spatial uncertainty normally exists in land-cover mapping. Characterizing spatial uncertainty of land-cover classes based on limited high quality data is essential in geographical information science for decision making and risk assessment. It is desirable to have a suitable method for assessing and representing the spatial uncertainty in land-cover area class mapping. The TMC model, which was specifically developed for modeling soil classes, is used to simulate the spatial distribution of multinomial land-cover classes. Probability vectors are estimated from multiple simulated realizations.
and probability maps representing the spatial uncertainty of each land cover class are presented. Variogram analyses are conducted between the reference land-cover map and simulated realizations.

The results demonstrate that the TMC model can produce highly imitative realization images, which are in accordance with the reference land-cover map, in a highly efficient way. Land-cover patterns, such as anisotropies, clear boundaries, and spatial juxtaposition relationships, can be effectively mimicked in realizations when a suitable density of survey line data is available. The TMC method also can serve as a spatial uncertainty model for estimating probability vectors at every location. The probability maps and the maximum probability map visualized from probability vectors clearly show the spatial distribution of every land-cover class and the transition zones from one class to others.

Variogram analyses indicate that the land-cover classes in the study area exhibit very complex spatial autocorrelations and cross-correlations. Such complex spatial
patterns are normally difficult to model using conventional methods by fitting variogram models to experimental data. However, the TMC model has the ability to reproduce such complex spatial relationships of land-cover classes. The good match between cross variograms estimated from the reference map and simulated realizations is remarkable.

An interesting and also practical feature of the method is that we need not deal explicitly with the anisotropy problem. Dealing with anisotropies of multinomial classes is quite cumbersome in indicator approaches (Goovaerts, 1997) or even very difficult in Markov random fields (Tjelmeland and Besag, 1998). Because of the high efficiency and simplicity of the TMC model, it is possible to use it in modeling land-cover classes in large areas with anisotropies of many classes.

This study demonstrated that the current TMC method underestimates minor classes when conditioning data are relatively sparse. This may influence the assessment of spatial patterns and spatial uncertainty of land-cover classes when available conditioning data are not adequate. To make the method applicable to sparse datasets, further improvement is necessary to solve this constraint.

ACKNOWLEDGMENTS

We thank Dr. John R. Jensen and anonymous referees’ helpful and insightful comments and suggestions.

REFERENCES


