International Journal of Geographical Information Science
Vol. 24, No. 6, June 2010, 821–839

Taylor & Francis
Taylor & Francis Group

Linear interpolation and joint model fitting of experimental transiograms for Markov chain simulation of categorical spatial variables

Weidong Li[a,b,c]* and Chuanrong Zhang[c]

[a]*Key Laboratory of Subtropical Agriculture Resource and Environment, Ministry of Agriculture, College of Resource and Environment, Huazhong Agricultural University, Wuhan, Hubei, China;* [b]*Department of Resource and Environmental Information Science and Engineering, Huazhong Agricultural University, Wuhan, Hubei, China;* [c]*Department of Geography and Center for Environmental Sciences and Engineering, University of Connecticut, Storrs, CT, USA*

(*Received 10 April 2009; final version received 15 June 2009*)

In Markov chain random field (MCRF) simulation of categorical spatial variables with multiple classes, joint modeling of a large number of experimental auto and cross-transiograms is needed. This can be tedious when mathematical models are used to fit the complex features of experimental transiograms. Linear interpolation can be used to perform the joint modeling quickly regardless of the number and the complexity of experimental transiograms. In this paper, we demonstrated the mathematical validity of linear interpolation as a joint transiogram-modeling method, explored its applicability and limitations, and tested its effect on simulated results by case studies with comparison to the joint model-fitting method. Simulations of a five-class variable showed little difference in patterns for interpolated and fitted transiogram models when samples were sufficient and experimental transiograms were in regular shapes; however, they neither showed large difference between these two kinds of transiogram models when samples were relatively sparse, which might indicate that MCRFs were not much sensitive to the difference in the detail of the two kinds of transiogram models as long as their change trends were identical. If available, expert knowledge might play an important role in transiogram modeling when experimental transiograms could not reflect the real spatial variation of the categorical variable under study. An extra finding was that class enclosure feature (i.e., a class always appears within another class) was captured by the asymmetrical property of transiograms and further generated in simulated patterns, whereas this might not be achieved in conventional geostatistics. We conclude that (i) when samples are sufficient and experimental transiograms are reliable, linear interpolation is satisfactory and more efficient than model fitting; (ii) when samples are relatively sparse, choosing a suitable lag tolerance is necessary to obtain reliable experimental transiograms for linear interpolation; (iii) when samples are very sparse (or few) and experimental transiograms are erratic, coarse model fitting based on expert knowledge is recommended as a better choice whereas both linear interpolation and precise model fitting do not make sense anymore.

Keywords: categorical spatial variable; expert knowledge; transiogram; Markov chain random field

1. Introduction

A one-step transition probability means a probability of transitioning from one state to another (or itself) in a single step in time (or space). A Markov chain is said to be stationary if

*Corresponding author. Email: weidong6616@yahoo.com

ISSN 1365-8816 print/ISSN 1362-3087 online

DOI: 10.1080/13658810903127991
http://www.informaworld.com

its transition probabilities are independent of specific time (or space) index. The one-step transition probability matrix (TPM), which consists of all of the relevant one-step transition probabilities among states of a discrete variable, has been traditionally used as a correlation measure and parameter input in Markov chain analysis of one-dimensional time (or space) series in many fields, such as geography (Brown 1970, Bell 1974, Collins *et al*. 1974, Tang *et al.* 2007), geology (Potter and Blakely 1967), biology, and ecology (Horn 1975, Balzter 2000). However, for conditional Markov chain simulation of categorical spatial variables (e.g., various landscape types) based on sparse sample data, one-step TPMs are not sufficient anymore because transition probabilities at different lag distances are needed for estimating the states at unsampled locations. Although idealized continuous-lagged transition probability functions (called idealized transiograms) may be inferred from one-step TPMs, on the one hand, they are not sufficient to reflect the complex spatial heterogeneity of categorical spatial variables, and, on the other hand, it is also difficult, if not impossible, to estimate one-step TPMs from sparse sample data. For solving these problems, the transiogram was proposed by Li (2006) as a spatial relationship measure of categorical spatial variables. Except for serving as the accompanying spatial measure of Markov chain geostatistics (MCG), transiograms also can be used to characterize the spatial variability of categorical spatial variables as an independent spatial metric.

The transiogram may generally refer to various transition probability curves over distance lags, of which some had appeared in earlier studies in spatial analysis of geological facies (Schwarzacher 1969, Luo and Thomsen 1994, Luo 1996, Carle and Fogg 1997, Ritzi 2000). Transiograms may be simply classified into three types – experimental transiograms (measured directly from sample data), exhaustive transiograms (measured from images or maps), and idealized transiograms (derived from single-step transition probabilities and/or domain knowledge based on the first-order Markovian assumption); their respective features and properties were demonstrated in Li (2007a). Among these different kinds of transiograms, experimental transiograms are closely attached with MCG because their models are directly used in Markov chain random field (MCRF) models for estimating the conditional probability distribution of a categorical variable at an unsampled location (Li 2007b).

The MCRF-based sequential simulation (MCSS) algorithm has proved to be an effective method for simulating categorical variables with multiple classes (e.g., soil classes) and complex interclass relationships (e.g., neighborships) (Li and Zhang 2007, Zhang and Li 2008). It demonstrated obvious advantages over sequential indicator simulation – the conventionally used method for simulating categorical variables. For example, MCSS generates polygon-like patterns in simulated realizations, which not only have higher accuracy but also obey the complex interclass relationships conveyed by input transiogram models (see Li and Zhang 2007). Thus one does not need to use post-processing methods as suggested by some researchers (Journel and Xu 1994, Goovaerts 1996, Deutsch 1998, 2006) to further improve the patterns of simulated realizations. MCSS also does not have the order relation violation problem in estimating conditional probability distributions of classes. However, to conduct Markov chain simulation of categorical spatial variables, one must first make a joint modeling of experimental auto- and cross-transiograms to obtain a valid set of transiogram models. For example, to simulate 10 classes, 100 auto- and cross-transiogram models are needed. Although some of them may be inferred from others by making use of the properties of transition probabilities, there are still at least 45 transiogram models to construct from scratch based on experimental transiograms and expert knowledge. This usually means a large work load in a tedious process, particularly when experimental transiograms have complex shapes to fit by using complex (or nested) mathematical models. Hence, how to efficiently get a valid set of transiogram models is a crucial issue in MCG.

In indicator kriging (Deutsch and Journel 1998), the indicator variogram served as the spatial correlation measure. Experimental indicator variograms were normally fitted by some recommended mathematical models such as spherical, Gaussian, and exponential models. Some of them have been adapted for fitting experimental transiograms (Li 2007a). Unfortunately, conventional indicator geostatistics does not provide an adequate method for joint modeling of auto- and cross-indicator variograms of categorical variables (Deutsch 1998, 2006, Machuca-Mory *et al*. 2008). Thus it normally ignores cross correlations among classes when modeling categorical variables. Carle and Fogg (1997) suggested the transition rate method to jointly model auto- and cross-experimental transiograms in the transition probability-based indicator geostatistics for implementing full indicator cokriging. A major limitation of the method is that it is based on the first-order Markovian assumption and thus implicitly assumes that class patch sizes (e.g., length and width) conform to an exponential distribution. This makes the method not widely applicable to many categorical spatial variables; for example, it apparently cannot be used to fit complex features such as multiple ranges and periodicity of experimental transiograms, which may appear in many situations and have to be fitted by complex models (e.g., nested or hole-effect models) (Li 2007a). When experimental transiograms are reliably estimated from a large number of data, their fluctuations tend to be smooth because of more data pairs and shorter lag tolerance; such smooth fluctuations normally represent the complex spatial heterogeneity of the categorical variable under study and therefore should not be ignored as noise.

MCG is much simpler in its estimators (i.e., MCRF models) than aforementioned indicator kriging approaches and thus provides the flexibility of using simpler methods to jointly model experimental transiograms merely based on the basic properties of transition probabilities. For a matrix of experimental transiograms $\hat{\mathbf{P}}(\mathbf{h}) = [\hat{p}_{ij}(\mathbf{h})]$, joint modeling needs to be done only row by row, rather than the whole matrix together. Thus, constraint conditions apply only to individual rows. Li (2007a) suggested a mathematical model-fitting procedure for joint modeling of experimental transiograms, which requires that one transiogram model take the remaining part of unity minus other models in a row of a transiogram model matrix. Although it is simple theoretically and methodologically, this method still can be very tedious for precise fitting of a large number of experimental transiograms, and it is also difficult to obtain a set of perfectly fitted models when experimental transiograms have complex shapes. Li and Zhang (2005) also proposed a simpler fast joint modeling method – linear interpolation of experimental transiograms. This method was found to be practical when samples can provide reliable experimental transiograms (e.g., no steep fluctuations at used lags). When reliable experimental transiograms are available, not only can this method largely release the transiogram-modeling burden in Markov chain simulation of categorical variables but it also provides a way to incorporate complex features of reliable experimental transiograms into simulations. However, it was not mathematically proved whether such a simple method could always provide a valid set of transiogram models. Therefore, a mathematical demonstration in its validity and some case analyses in its applicability and limitations are necessary to clarify related confusions and guide its application. In addition, such a fast method may also be applicable to modeling other spatial measures.

The objectives of this study are (1) to prove the validity and applicability of linear interpolation with comparison to model fitting in transiogram modeling, (2) to test the sensitivity of Markov chain simulation to transiogram models derived by the two different methods with increasing numbers of samples, and (3) to find out the practical ways of joint transiogram modeling when experimental transiograms vary from reliable to unreliable. The rest of this paper is arranged as follows: We first introduce the constraint conditions for joint modeling of experimental transiograms and their rationality in Section 2. In Section 3 we prove that linear interpolation mathematically meets all the constraint conditions and is therefore rational. Then in

Section 4 we show the applicability and limitations of linear interpolation by estimating transiograms from three sample sets of a binary variable with different numbers of samples. We further conduct some simulations of a five-class variable to demonstrate and analyze the sensitivity of Markov chain simulation to different kinds of transiogram models obtained by linear interpolation and model fitting in Section 5. Finally we conclude this paper by summarizing the case studies and making recommendations in the usage of the two transiogram joint modeling methods.

2. Constraint conditions

In Markov chain simulation of categorical spatial variables, the modeling of experimental transiograms must meet some constraint conditions to generate a valid set of auto- and cross-transiogram models to ensure a proper simulation and valid results. These constraint conditions are simply some basic properties of transition probabilities and categorical spatial variables. Three constraint conditions were defined for transiogram modeling in Li (2007a): (a) non-negative, (b) transiograms headed by the same class (i.e., a row of elements in a transiogram matrix) summing to one at any used lags, and (c) no-nugget effect.

Assuming $\mathbf{P}(\mathbf{h}) = [p_{ij}(\mathbf{h})]$ is a TPM at the distance lag $\mathbf{h}$ (or a matrix of transiogram models), we have the first constraint condition defined as

$$p_{ij}(h) \geq 0, \tag{1}$$

the second constraint condition defined as

$$\sum_{j=1}^{n} p_{ij}(\mathbf{h}) \equiv 1, \tag{2}$$

and the third constraint condition defined as

$$p_{ij}(0) = 0 \quad \forall i \neq j \quad \text{and } p_{ii}(0) = 1 \tag{3}$$

where $i, j = 1, \ldots, n$, indicate the numbers of n classes, and $p_{ij}(\mathbf{h})$ represents a transiogram, which is defined as a transition probability function from state i to state j across the distance lag $\mathbf{h}$. Any transiogram-modeling method that meets the above conditions can be used in Markov chain modeling of categorical variables.

The non-negative condition is required because negative probabilities are simply irrational. The second condition is also required because it is the basic property of a valid TPM to have its row elements sum to one. As to the third condition, it is not strictly required in Markov chain modeling as long as the summing-to-one condition can be met at all used lag values. However, it will be better to designate this condition as a constraint for joint transiogram modeling for the following reasons: First, classes of categorical spatial variables are mutually exclusive; that means it is physically irrational for their auto-transition probabilities at the same location to be less than one and for their cross-transition probabilities at the same location to be nonzero. Second, in geostatistics, the nugget effect in variogram models is thought to be caused by measurement errors and/or short-range variations missed by sparse sample data, and therefore in variogram modeling it may be retained or simply replaced by a short-range model (or a short-range component in a nested model) (Goovaerts 1997, p. 101–103). So it is proper to discard nugget effects in practical use by honoring origin points in transiogram modeling because measurement errors (e.g., sample misclassification) in categorical variables are actually unquantifiable (or may be ignored). Third, not

using nugget effects also simplifies the transiogram joint modeling procedure. In general, the three constraint conditions for joint transiogram modeling have their rationalities and are also sufficient to ensure proper Markov chain modeling.

3. Joint transiogram modeling

3.1. Mathematical rationalities of linear interpolation

An experimental transiogram $\hat{p}_{ij}(\mathbf{h})$ may be directly estimated from sample data by counting the transition frequency from a class to itself or another with different lags (e.g., numbers of pixels for raster data) by the following equation:

$$\hat{p}_{ij}(\mathbf{h}) = \frac{f_{ij}(\mathbf{h})}{\sum_{k=1}^{n} f_{ik}(\mathbf{h})} \tag{4}$$

where $f_{ij}(\mathbf{h})$ represents the frequency of transitions from class i to class j at the lag $\mathbf{h}$ and n is the total number of classes. To acquire reliable experimental transiograms from sparse samples, one has to consider a lag tolerance $\Delta\mathbf{h}$ around each lag value, which may be decided by users according to the density of samples.

Assume $\mathbf{h}_k$ and $\mathbf{h}_{k+1}$ are two specific neighboring lag values in an experimental transiogram $\hat{p}_{ij}(\mathbf{h})$ with $\mathbf{h}_{k+1} > \mathbf{h}_k$ and $\hat{p}_{ij}(\mathbf{h}_k)$ and $\hat{p}_{ij}(\mathbf{h}_{k+1})$ the transition probability values measured from sample data at these two corresponding lags (Figure 1). The linear interpolation equation for calculating the transition probability value $p_{ij}(\mathbf{h})$ at the lag $\mathbf{h}$ between lags $\mathbf{h}_k$ and $\mathbf{h}_{k+1}$ can be expressed as

$$p_{ij}(\mathbf{h}) = \frac{\hat{p}_{ij}(\mathbf{h}_k) \cdot (\mathbf{h}_{k+1} - \mathbf{h}) + \hat{p}_{ij}(\mathbf{h}_{k+1}) \cdot (\mathbf{h} - \mathbf{h}_k)}{\mathbf{h}_{k+1} - \mathbf{h}_k} \tag{5}$$

(Li and Zhang 2005, Zhang and Li 2008). The above equation can meet all the three constraint conditions listed in the last section. The following proves these points:

(a) *Non-negative*: Measured transition probabilities from samples are always non-negative, thus we have $\hat{p}_{ij}(\mathbf{h}_k) \geq 0$ and $\hat{p}_{ij}(\mathbf{h}_{k+1}) \geq 0$. According to the definition of the linear interpolation equation, we also have $\mathbf{h}_{k+1} > \mathbf{h}_k$, $\mathbf{h}_{k+1} \geq \mathbf{h}$ and $\mathbf{h} \geq \mathbf{h}_k$ (see Figure 1). Thus, it is easy to see that in the right-hand side of Equation (5) the numerator is non-negative and the denominator is positive; then we have $p_{ij}(\mathbf{h}) \geq 0$, that is, linear interpolation meets the non-negative condition.

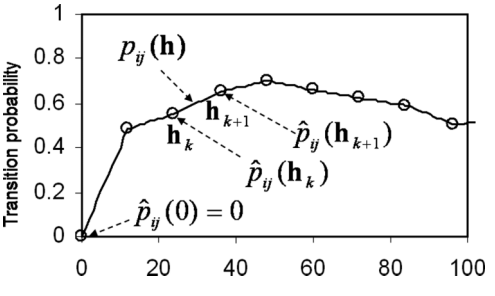

Figure 1. Linear interpolation of an experimental cross-transiogram (dots for measured values, solid line for interpolated model).

(b) *Summing-to-one*: Assume $\hat{\mathbf{P}}(\mathbf{h}_k) = [\hat{p}_{ij}(\mathbf{h}_k)]$ and $\hat{\mathbf{P}}(\mathbf{h}_{k+1}) = [\hat{p}_{ij}(\mathbf{h}_{k+1})]$ are two measured valid transition probability matrices at lags $\mathbf{h}_k$ and $\mathbf{h}_{k+1}$, respectively (here 'valid' means the transition probabilities are correctly calculated from transition frequencies estimated from samples). Then they have the basic property that elements in each row sum to one, that is, $\sum_{j=1}^{n} \hat{p}_{ij}(\mathbf{h}_k) \equiv 1$ and $\sum_{j=1}^{n} \hat{p}_{ij}(\mathbf{h}_{k+1}) \equiv 1$. From Equation (5) we further have

$$\begin{aligned}\sum_{j=1}^{n} p_{ij}(\mathbf{h}) &= \sum_{j=1}^{n} \left[\frac{\hat{p}_{ij}(\mathbf{h}_k) \cdot (\mathbf{h}_{k+1} - \mathbf{h}) + \hat{p}_{ij}(\mathbf{h}_{k+1}) \cdot (\mathbf{h} - \mathbf{h}_k)}{\mathbf{h}_{k+1} - \mathbf{h}_k}\right] \\ &= \frac{1}{\mathbf{h}_{k+1} - \mathbf{h}_k} \cdot \left[(\mathbf{h}_{k+1} - \mathbf{h}) \cdot \sum_{j=1}^{n} \hat{p}_{ij}(\mathbf{h}_k) + (\mathbf{h} - \mathbf{h}_k) \cdot \sum_{j=1}^{n} \hat{p}_{ij}(\mathbf{h}_{k+1})\right] \\ &= \frac{1}{\mathbf{h}_{k+1} - \mathbf{h}_k} \cdot [(\mathbf{h}_{k+1} - \mathbf{h}) + (\mathbf{h} - \mathbf{h}_k)] \equiv 1\end{aligned} \tag{6}$$

This proves that linear interpolation surprisingly meets the summing-to-one condition though it is very simple. This also can be seen from Figure 2, where interpolated experimental auto- and cross-transiograms for a binary variable always meet on the 0.5 probability line, that is, they sum to one.

(c) *No-nugget effect*: By counting the origin points (i.e., (0, 1) for auto-transiograms and (0, 0) for cross-transiograms) into experimental transiograms, linear interpolation meets the no-nugget-effect condition. This is because linear interpolation respects measured values exactly (see Figure 1): From Equation (5), one can easily get $p_{ij}(\mathbf{h}) = \hat{p}_{ij}(\mathbf{h}_k)$ when $\mathbf{h} = \mathbf{h}_k$ and $p_{ij}(\mathbf{h}) = \hat{p}_{ij}(\mathbf{h}_{k+1})$ when $\mathbf{h} = \mathbf{h}_{k+1}$. That means we have $p_{ij}(0) = 0$ and $p_{ii}(0) = 1$ when $\mathbf{h} = 0$.

Therefore, mathematically linear interpolation is suitable for joint transiogram modeling, and it can provide a valid set of transiogram models for Markov chain simulation of categorical spatial variables. Of course, mathematical validity of a method does not mean it is always good to use at any situation in real-world cases. As demonstrated later, when samples are very sparse and cannot provide reliable experimental transiograms, this method is not recommended.

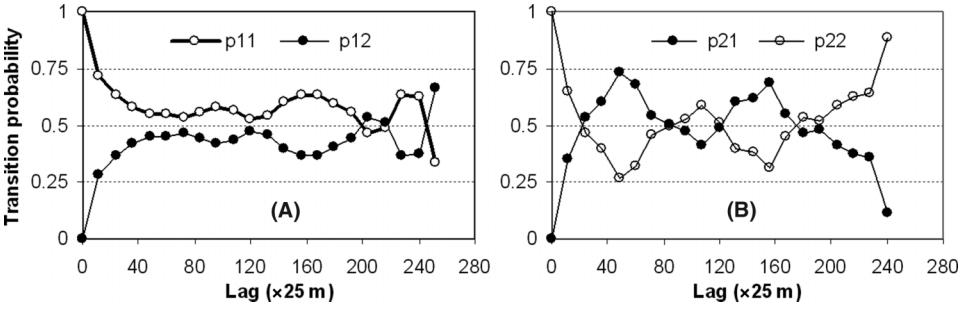

Figure 2. Interpolated experimental auto- and cross-transiograms estimated from a binary variable. Each pair has a common head class and sums to 1 at any lag values. (A) From 487 samples with a lag tolerance of 12-p length; (B) From 97 samples with a lag tolerance of 12-p length.

3.2. Model fitting

Assume $p_{ik}(\mathbf{h})$ is a transiogram model from the transiogram model matrix $\mathbf{P}(\mathbf{h}) = [p_{ij}(\mathbf{h})]$. To get a valid set of transiogram models for Markov chain simulation, one needs to conduct the model fitting of experimental transiograms subset by subset (i.e., row by row in a transiogram matrix, each row has a common head class). For each subset one model must take the left part of unity minus others as

$$p_{ik}(\mathbf{h}) = 1 - \sum_{\substack{j=1 \\ j \neq k}}^{n} p_{ij}(\mathbf{h}) \tag{7}$$

so that the summing-to-one condition is ensured. One can ensure the other two constraint conditions to be met simply by including the origin points into experimental transiograms and tuning fitted models. Mathematical models for fitting experimental variograms may be adapted to fit experimental transiograms. Li (2007a) provided some basic mathematical models for transiogram modeling, which include linear, spherical, exponential, Gaussian, cosine-exponential, and cosine-Gaussian models.

Because experimental transiograms are usually estimated omnidirectionally (or bidirectionally), they and consequently their models meet the property of

$$p_{ij}(\mathbf{h}) = \frac{p_j}{p_i} \cdot p_{ji}(\mathbf{h}) \tag{8}$$

where p_i and p_j are proportions of class i and class j, respectively. Apparently, if $p_{ji}(\mathbf{h})$, p_i, and p_j are all known, one can use Equation (8) to infer the transiogram model $p_{ij}(\mathbf{h})$. Thus, using Equations (7) and (8) and class proportion data, one may fit only 10 experimental transiograms to get all of the 25 (auto and cross) transiogram models for five classes. In this study, we use class proportions of sample data to approximately serve as the real class proportions in the simulation domain in case studies.

4. Applicability and limitations

Data of a binary variable were used to test the applicability and limitations of linear interpolation and model fitting in joint modeling of experimental transiograms. Estimated experimental transiograms from three different sample sets and their models are shown in Figure 3. In the first row (Figure 3 A1, A2, and A3), a dense sample set (487 points) and three different lag tolerances (i.e., 12, 6, and 3 p length) were used to obtain three versions of an experimental cross-transiogram (here we regard them as three experimental transiograms). An exponential model with a sill of 0.43 and a range of 35 p (i.e., 35 pixels) length was used to fit all three of them. It can be seen that these experimental transiograms with different lag tolerances have little difference. This means that when sample data are sufficient, very reliable experimental transiograms can be estimated and their features should be the real reflection of spatial variation of the variable under study. Although the exponential model can basically fit all the three simple experimental transiograms well, it still misses the irregular convex and concave at medium and high lags. This indicates that for even such simple experimental transiograms a perfect fitting using a mathematical model is not feasible. However, linear interpolation perfectly captures all features; this means that linear interpolation should be appreciated under this situation.

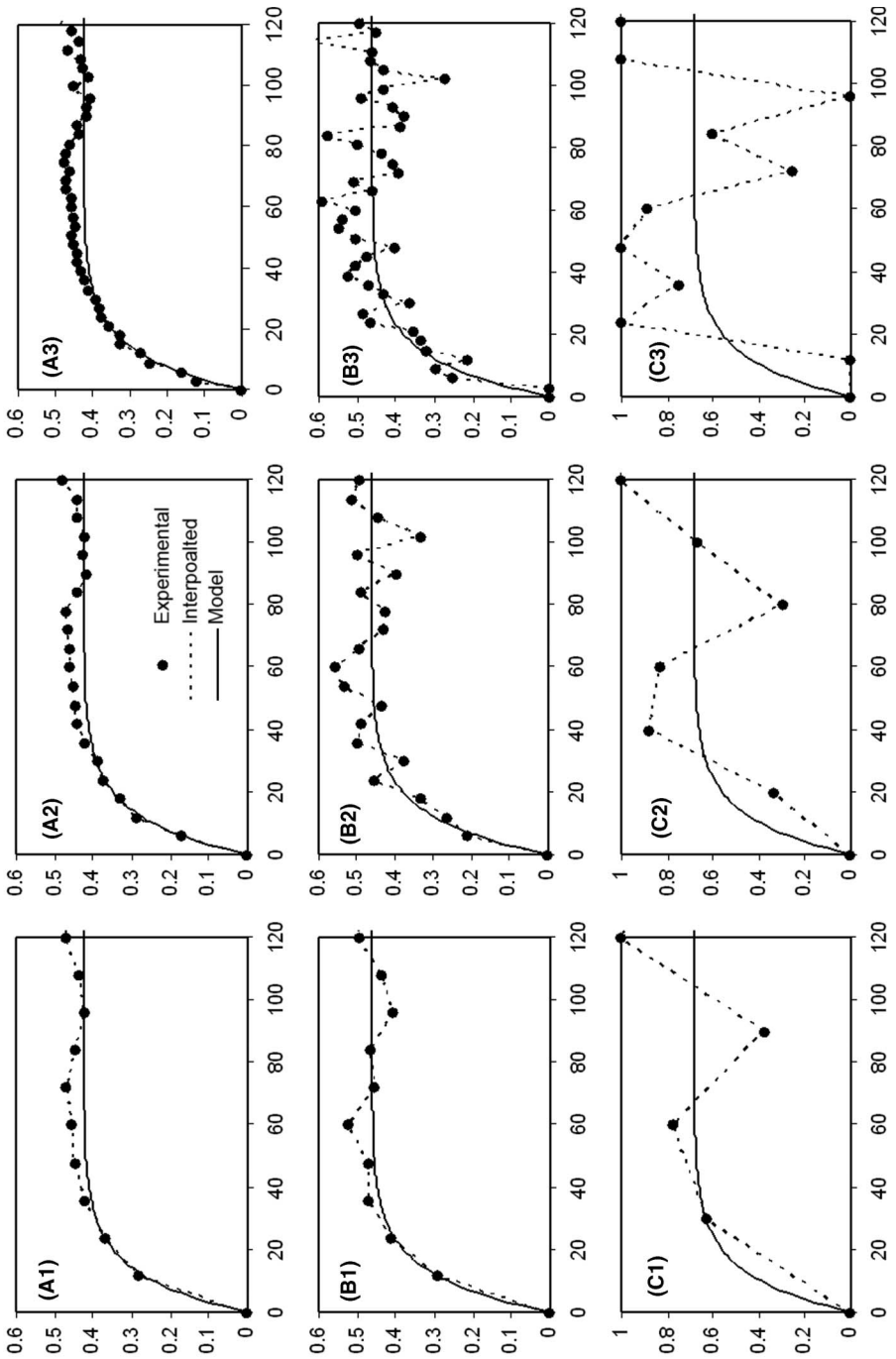

Figure 3. Different versions of an experimental cross-transiogram of a binary variable and their models when it is estimated from different numbers of samples with different lag tolerances. (A1), (A2), and (A3) Estimated from 487 sample data using different lag tolerances of 12-, 6-, and 3-p length, respectively. (B1), (B2), and (B3) Estimated from 97 sample data using different lag tolerances of 12-, 6-, and 3-p length, respectively. (C1), (C2), and (C3) Estimated from 19 sample data using different lag tolerances of 30-, 20-, and 12-p length, respectively. Math models: exponential, with a range of 35-p length, and sills of 0.4251, 0.4639, and 0.6842 for (A), (B), and (C) rows, respectively, based on different sample sets.

When the number of samples decreases to 97, the experimental transiogram estimated with a small lag tolerance (Figure 3 B3) shows steep fluctuations. But when the lag tolerance increases to 12 p length (Figure 3 B1), the experimental transiogram becomes reliable with comparison to those estimated from the dense sample set. This means linear interpolation is still proper to use as long as a suitable lag tolerance is chosen to decrease the steep fluctuations of the experimental transiogram. In fact, when an experimental transiogram steeply fluctuates, not only linear interpolation loses its charm, it is also not easy to decide the fitting model's type (e.g., exponential, spherical, or Gaussian) and parameters.

When sample data are very few (here decrease to 19 points), experimental transiograms become unreliable even if the lag tolerance increases to a very large value (Figure 3 C). The transiogram model provided in row C of Figure 3 is not fitting to the experimental ones because the latter do not provide sufficient information for a confident modeling. On the contrary, it is an assumed model based on the knowledge from the dense or medium sample set. Apparently linear interpolation is not proper to use anymore for this small sample set in the sense of modeling. However, simply using a mathematical model to fit the values of such an unreliable experimental transiogram also does not make sense, even if the fitting is statistically optimal. Under this situation, a proper choice should be to choose a simple mathematical model based on expert knowledge to 'coarsely fit' the experimental transiogram because anyway the exact features of the spatial variation of the variable under study are unknown. Here a 'coarse model fitting' means one may ignore the detail of the experimental transiogram and only take care of its general trend and his/her confident personal knowledge by a mathematical model. There is no doubt that different experts may choose different models based on the experimental transiogram and their different personal knowledge about the study area and the target variable (Figure 4). The more knowledge an expert has about the target variable in the study area, the better model he can choose. Here a better model means it is closer to the truth even though it appears to deviate a lot from the experimental one. For example, in Figure 4, the transiogram model inferred from the dense sample set is used to represent the 'truth' that expert knowledge should approach; however, the transiogram models chosen by experts 1 and 2 may have apparent deviations because of their poor knowledge, though these models may better fit the unreliable experimental transiogram from the sparse sample set.

5. Sensitivity analysis

To test the sensitivity of Markov chain simulation to transiogram models derived in different ways, we chose a relatively dense sample set (139 points) and a relatively sparse one (47 points, thinned from the dense one) of a five-class categorical variable for case studies.

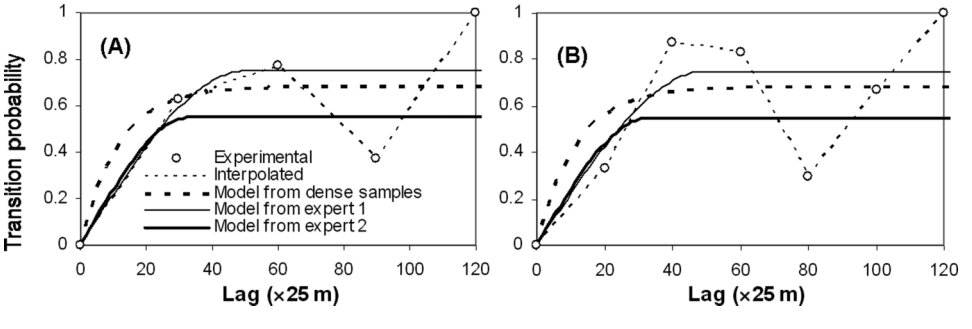

Figure 4. Different models to fit an experimental cross-transiogram estimated from a sparse sample set (19 sample data). (A) Using a lag tolerance of 30-p length; (B) Using a lag tolerance of 20-p length.

Table 1. Class proportions in sample data sets and averaged from simulated realizations (100 per simulation) based on transiogram models obtained by different ways.

Sample set	Data source	Class proportion				
		Class 1	Class 2	Class 3	Class 4	Class 5
Dense sample set (139 points)	Samples	0.3597	0.1439	0.0576	0.0863	0.3525
	Realizations, using interpolated transiogram models, 12-p lag tolerance	0.3599	0.1452	0.0687	0.0963	0.3299
	Realizations, using fitted transiogram models, 12-p lag tolerance	0.3615	0.1465	0.0673	0.0958	0.3289
Sparse sample set (47 points)	Samples	0.4043	0.1064	0.0638	0.0213	0.4043
	Realizations, using interpolated transiogram models, 12-p lag tolerance	0.3995	0.1038	0.0721	0.0218	0.4028
	Realizations, using interpolated transiogram models, 20-p lag tolerance	0.4258	0.1265	0.0774	0.0175	0.3528
	Realizations, using fitted transiogram models with poor expert knowledge (i.e., fitted transiogram models, 20-p lag tolerance)	0.4295	0.1385	0.0720	0.0127	0.3474
	Realizations, using fitted transiogram models with perfect expert knowledge (i.e., fitted transiogram models using the dense sample set, 12-p lag tolerance)	0.3634	0.1252	0.0780	0.0672	0.3662

The study area is about 15 km^2 and discretized into a 239 $\times$ 97 grid. The proportions of different classes are different in each sample set and between the two sample sets (Table 1). We estimated experimental transiograms from the two sample sets using different lag tolerances (12 p length for the dense sample set and 12 and 20 p length for the sparse sample set) and then inferred corresponding transiogram models by linear interpolation (i.e., interpolated transiogram models) and model fitting (i.e., fitted transiogram models). Six simulations based on each set of transiogram models were conducted using the MCSS algorithm. For each simulation, 100 realizations were generated to estimate averaged class proportions (see Table 1) and the optimal prediction map.

5.1. Results based on the dense sample set

Figure 5 shows two subsets of transiogram models obtained through linear interpolation and model fitting of corresponding experimental transiograms estimated from the dense sample set with a lag tolerance of 12 p length, one subset for the largest class (class 1) and the other for the smallest class (class 3). It can be seen that experimental transiograms headed by the largest class have little fluctuation and thus can be well fitted by mathematical models such as the exponential model, the spherical model, and the dampened hole-effect model. Therefore, linear interpolation and mathematical model fitting make little difference in this case, whereas the latter needs much more time and tedious adjustments, which may be unendurable when the number of classes is large. It is worth mentioning here that the dampened hold-effect model (Deutsch and Journel 1998, p. 26) is quite useful for approximately fitting the first peak of experimental transiograms if the peak is not very sharp. For experimental transiograms headed by the smallest class, some fluctuations appear but we still can use mathematical models to fit well major features although having to miss some of them. Clearly under this situation, model fitting has no advantages over linear interpolation, whereas the latter can be done quickly

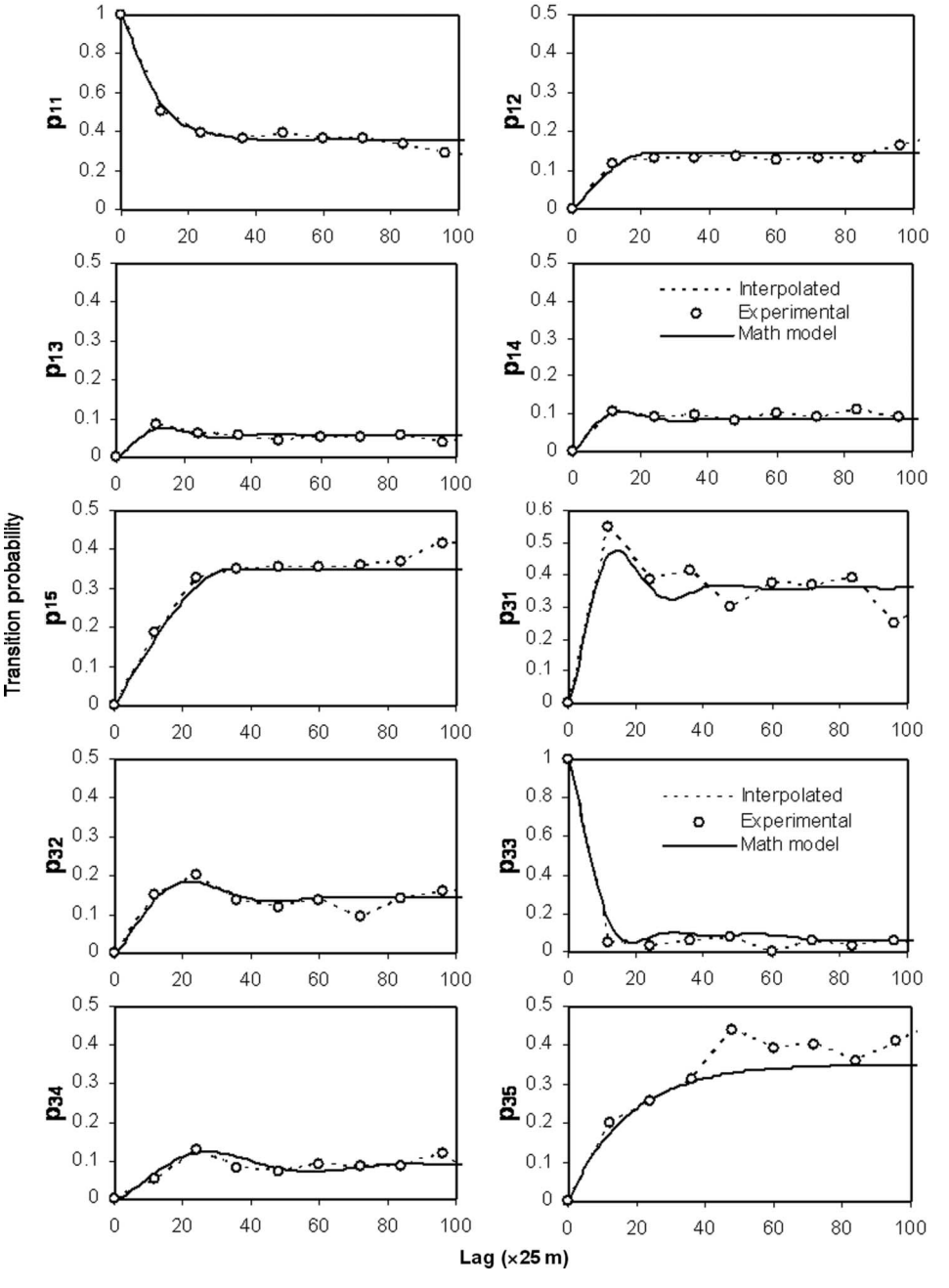

Figure 5. Two subsets of experimental transiograms of a five-class categorical variable and their models, headed by the largest class (class 1, proportion 0.3597) and the smallest class (class 3, proportion 0.0576), estimated from a dense sample set (139 points) using a lag tolerance of 12-p length.

regardless of the number of involved transiograms. Note that sills of the fitted transiogram models were set to the proportions of corresponding tail classes, which are quite identical with the sills of corresponding experimental transiograms, as shown in Figure 5.

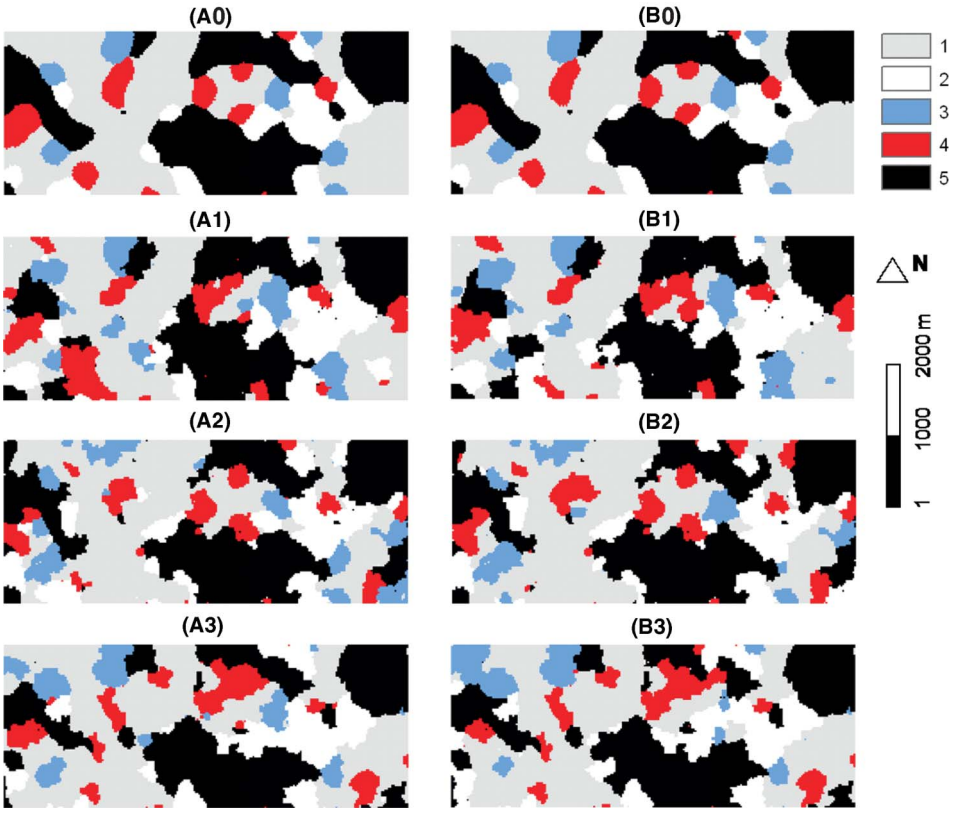

Figure 6. Simulated results of a five-class categorical variable, based on the dense sample set (139 points). Column A: Using interpolated transiogram models. Column B: Using fitted transiogram models. 1, 2, and 3 in map labels (e.g., A1, A2, and A3) refer to three different realizations. A0 and B0 are optimal prediction maps based on maximum occurrence probabilities estimated from 100 simulated realizations.

Because transiogram models obtained by linear interpolation and model fitting do not have much difference in this case study, it is expected that simulated results using these two sets of transiogram models are similar. This is verified by the simulated results, some of which are shown in Figure 6. It can be seen that class patterns in optimal prediction maps of the two simulations are basically the same (Figure 6: A0 vs. B0), and if the same random number sequence is used class patterns in realizations of the two simulations are very similar (Figure 6: A1 vs. B1, A2 vs. B2). Class proportions estimated from simulated realizations based on the two different sets of transiogram models also basically have no difference (Table 1). But they do have some deviation from the original class proportions (i.e., estimated from the sample data). This is easy to understand because class proportions in simulated realizations are not only related to transiogram models (mainly their sills) but also related to the spatial distributions of samples of different classes. The relative underestimation of class 5 in simulations is probably because this class tends to occur more frequently at boundaries (i.e., boundary effect).

5.2. Results based on the sparse sample set

For the sparse sample set, experimental transiograms were estimated using two lag tolerances – 12 and 20 p length, as shown in Figure 7 for the subset headed by the largest class (class 5) and

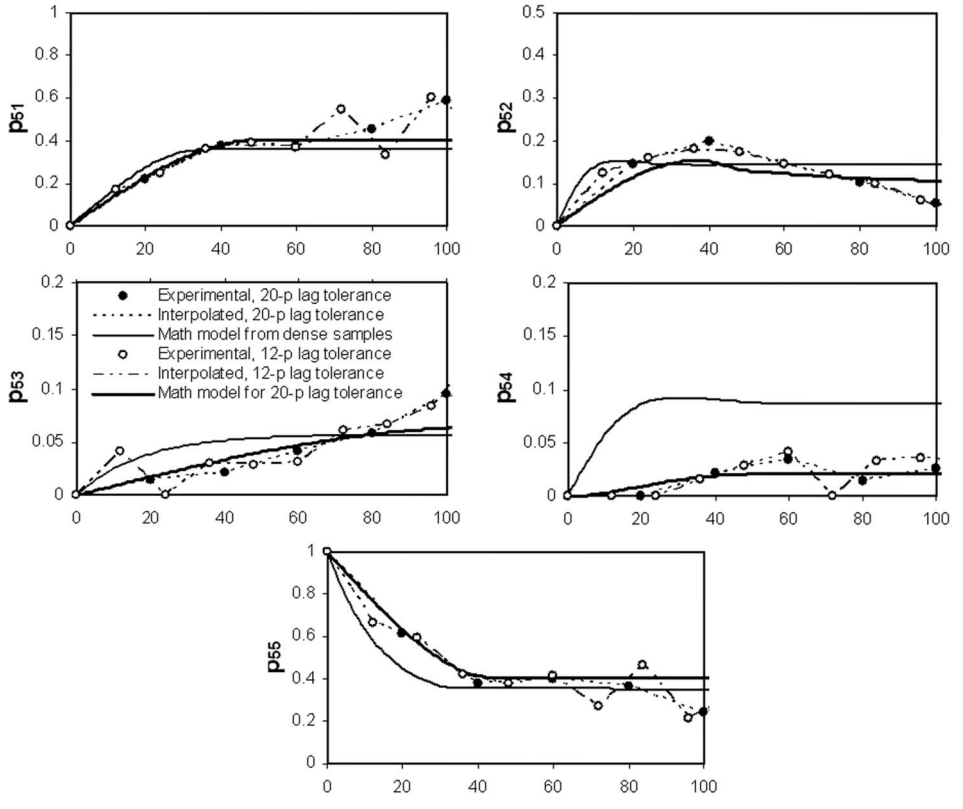

Figure 7. A subset of experimental transiograms headed by a large class (class 5, proportion 0.4043), estimated from a sparse data set (47 sample points) using different lag tolerances (12- and 20-p length), and their models inferred using different methods (interpolation and model fitting) and different levels of experts (poor expert and perfect expert). Here the perfect expert's models (thin solid lines) are the same as those inferred from the dense sample set, and the poor expert's models (thick solid lines) are those simply fitting the experimental transiograms estimated from the sparse sample set with the lag tolerance of 20-p length.

in Figure 8 for the subset headed by the smallest class (class 4), respectively. They were interpolated and model-fitted. Because of the small number of samples, experimental transiograms estimated using the 12-p lag tolerance show strong fluctuations, especially for those headed by small classes, which imply that these transiograms are less reliable. Increasing the lag tolerance to 20 p length can obviously smooth the fluctuations, although the number of measured transition probabilities also becomes smaller. Because of the fluctuations, the small number of measured values, and the occurrence of zero values in experimental transiograms, it is difficult to fit these experimental transiograms precisely using mathematical models. Thus, as we proposed in the last section, a 'coarse fitting' to these experimental transiograms by mathematical models usually can work well for simulation. Because the experimental transiograms with the 20-p lag tolerance are smoother, we use mathematical models to fit this set of experimental transiograms, and the sills of models are set to the proportions of corresponding tail classes as suggested for transiogram modeling (Li 2007a). It should be noted that class proportions in the sparse sample set changed because of the reduced number of samples relative to those in the dense sample set (Table 1). This change implies that the spatial correlation information conveyed by sparse samples in a limited area may largely deviate from the truth.

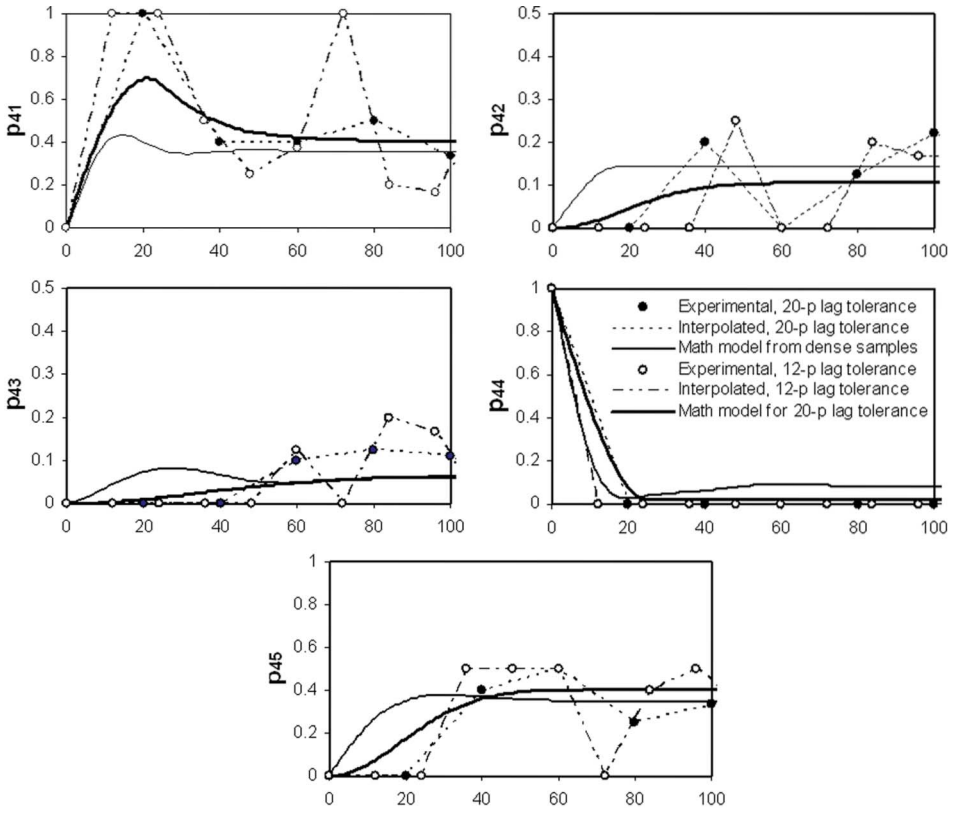

Figure 8. A subset of experimental transiograms headed by a small class (class 4, proportion 0.0213), estimated from a sparse data set (47 sample points) using different lag tolerances (12- and 20-p length), and their models inferred using different methods (interpolation and model fitting) and different levels of experts (poor expert and perfect expert). Here the perfect expert's models (thin solid lines) are the same as those inferred from the dense sample set, and the poor expert's models (thick solid lines) are those simply fitting the experimental transiograms estimated from the sparse sample set with the lag tolerance of 20-p length.

Simulated results using the two sets of interpolated transiogram models are shown in Figure 9. It can be seen that simulated realizations based on the interpolated transiograms measured with 12-p lag tolerance (Figure 9 column A) and those based on the interpolated transiograms measured with 20-p lag tolerance (Figure 9 column B) show apparent difference in their spatial patterns, particularly those parts related to both classes 2 and 5. Clearly class 2 is always enclosed by class 5. This should result from the difference of experimental cross-transiograms between classes 2 and 5. Checking the experimental transiograms, we did not find any problem on $\hat{p}_{52}(\mathbf{h})$, but indeed found a problem on $\hat{p}_{25}(\mathbf{h})$. Figure 10 shows clearly that the first measured transition probability value in $\hat{p}_{25}(\mathbf{h})$ is 1 for the 12-p lag tolerance but not for the 20-p lag tolerance. This situation irrationally causes a large difference at the low lag section between the two interpolated transiograms using 12- and 20-p lag tolerances. The reason should be that class 2 samples are surrounded by class 5 samples in the sparse sample set and thus no transitions from class 2 to other classes can be captured within a short lag distance. The consequence is that class 2 is completely enclosed by class 5 in simulated patterns using that interpolated transiogram. The occurrence of such a problem may mean that the 12-p lag tolerance is too small for estimating appropriate transition probabilities from the sparse sample set. It might

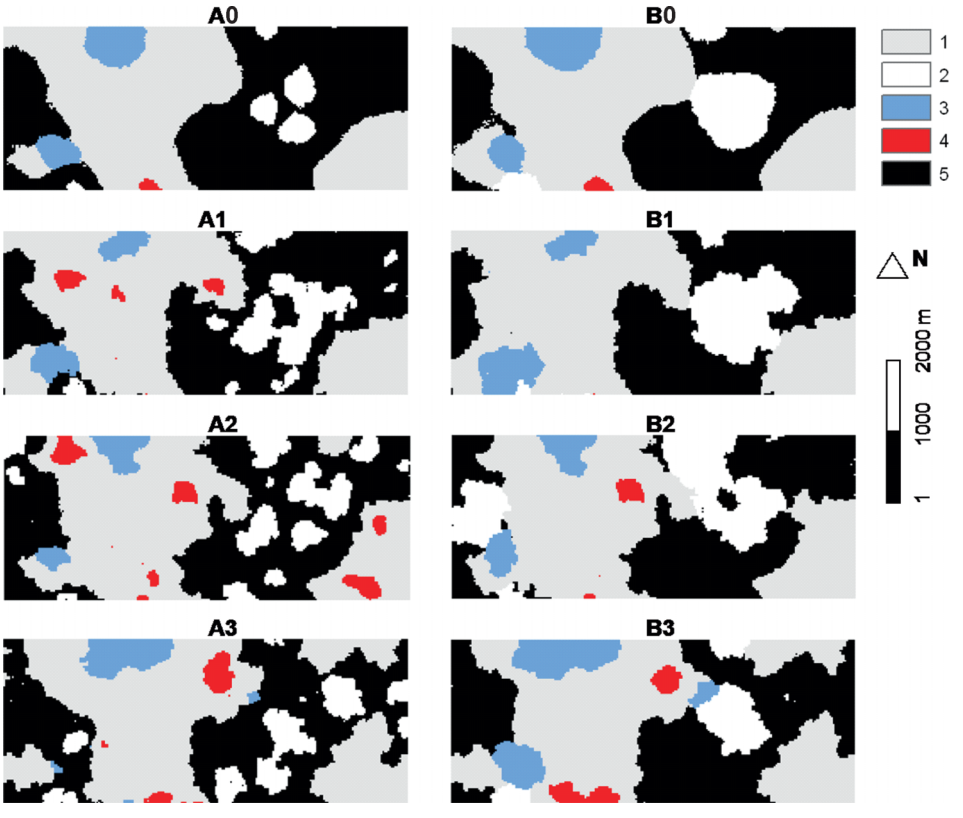

Figure 9. Simulated results of a five-class categorical variable, based on the sparse sample set (47 points). Column A: Using interpolated experimental transiograms measured with the lag tolerance of 12-p length. Column B: Using interpolated experimental transiograms measured with the lag tolerance of 20-p length. 1, 2, and 3 in map labels (e.g., A1, A2, and A3) refer to three different realizations. A0 and B0 are optimal prediction maps based on maximum occurrence probabilities estimated from 100 simulated realizations.

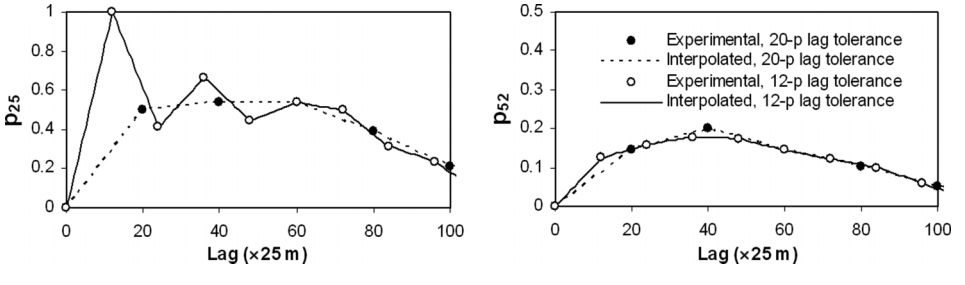

Figure 10. Experimental transiograms $\hat{p}_{25}(\mathbf{h})$ and $\hat{p}_{52}(\mathbf{h})$, estimated from a sparse data set (47 sample points) using different lag tolerances (12- and 20-p length), and their interpolated models.

also mean that the 20-p lag tolerance is too large and misses a special feature of the real pattern (if the feature is real), because the use of the 20-p lag tolerance avoids the problem and consequently destroys the enclosure feature in simulated patterns. Despite obvious difference in pattern details, the general patterns in simulated realizations, especially those in optimal prediction maps, still show some similarity to some extent. This might imply that the MCRF model is not very

sensitive to the fluctuations of transiograms if the fluctuations are not irrational. Note that from both simulations it can be seen that class 4 is always enclosed in class 1. This also corresponds to the characteristics of experimental transiogram $\hat{p}_{41}(\mathbf{h})$, where the first measured transition probability value is 1 for both 12- and 20-p lag tolerances (see Figure 8). The generation of such a feature is typically an advantage of transiograms as an asymmetric metric over symmetric metrics such as indicator variogams. There are some obvious deviations between the class proportions from the sparse sample set and those from the simulated realizations using the interpolated transiograms with 20-p lag tolerance, which may be caused by the excessive smoothing effect of the large lag tolerance.

The simulated results based on transiogram models fitting the 20-p lag tolerance experimental transiograms are shown in Figure 11 column A. Although the model fitting is very coarse, the simulated results using this set of transiogram models and those using interpolated transiograms with 20-p lag tolerance (see Figure 9 column B) are quite similar in patterns and class proportions. This indicates that the MCRF model is not much sensitive to some fluctuations in transiogram models as long as their change trends are the same.

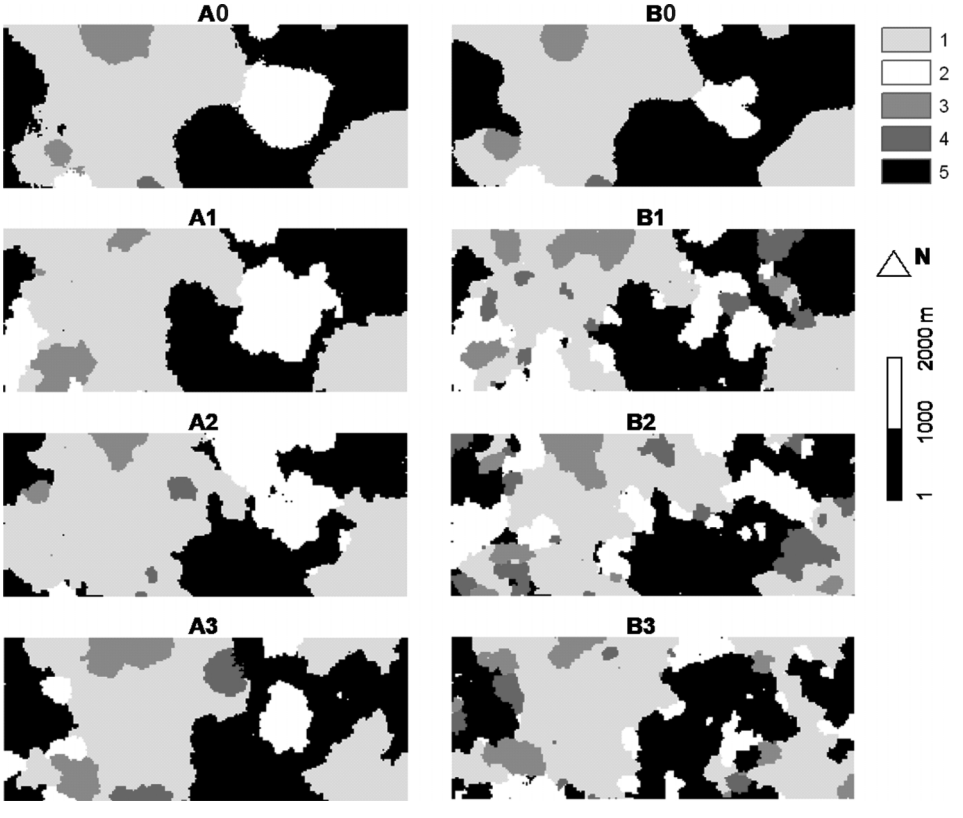

Figure 11. Simulated results of a five-class categorical variable, based on the sparse sample set (47 points). Column A: Using transiogram models fitting the experimental transiograms estimated from the sparse sample set with the lag tolerance of 20-p length, representing the poor expert's knowledge. Column B: Using transiogram models fitting the experimental transiograms estimated from the dense sample set with the lag tolerance of 12-p length, representing the perfect expert's knowledge. 1, 2, and 3 in map labels (e.g., A1, A2, and A3) refer to three different realizations. A0 and B0 are optimal prediction maps based on maximum occurrence probabilities estimated from 100 simulated realizations.

5.3. Results based on assumed expert knowledge

Precise model fitting to the experimental transiograms measured from the sparse sample set is difficult and also unnecessary because of their unreliability. Thus, 'coarse model fitting' to these experimental transiograms based on expert knowledge should do a better job, because expert knowledge can be assumed to be more reliable than the experimental transiograms measured from a small number of samples. Here an issue arises: How much valuable knowledge does an expert (i.e., the modeler) have? Different experts may be at different levels in knowing the study area: some may have very rich knowledge about the study area and thus can construct good transiogram models even without considering much the experimental transiograms; some others may have little knowledge about the study area and have to attempt to fit the unreliable experimental transiograms with mathematical models based on the basic principles of transiogram modeling. For this reason, here we assume we have two experts, a 'perfect' one and a 'poor' one, to conduct the transiogram modeling with mathematical models. The perfect expert knows the correct proportions of all classes (used to set model sills) and their spatial auto- and cross-correlation structures (used to decide model types and ranges) in the study area. He or she provides a set of transiogram models, which are thought to correctly reflect the truth and ignore the unreliable experimental transiograms. Here we use the set of fitted transiogram models inferred from the dense sample set to serve as the models provided by the perfect expert, assuming that the dense sample set can perfectly reflect the spatial variability of the classes in the study area. From Figures 7 and 8, it can be seen that the fitted transiogram models from the dense sample set (i.e., the perfect expert's models) sometimes largely deviate in sills, ranges, and model types from the experimental transiograms measured from the sparse sample set, because the sparse sample set does not reflect the truth in proportions and correlation structures of some classes. The poor expert has no knowledge at all about the study area. All that he or she has is the sparse sample set. Thus he or she calculates the class proportions and the experimental transiograms (each has only several measured values because of the use of a large lag tolerance) from the sample set and then he or she does his or her best to choose some mathematical models to fit the experimental transiograms with the sills of these models being set to the proportions of the corresponding tail classes. From Figures 7 and 8, one can see that the transiogram models fitting the experimental transiograms (i.e., the poor expert's models) may differ largely from the perfect expert's models in sills, ranges, and model types.

In the real world, a modeler's expert knowledge about his study area may not be so poor or so rich as the two extreme cases assumed above. So when a modeler estimates the transiogram models from a small number of samples, the quality of the models should fall between the above two extreme cases. Here we just use these two extreme cases to test the response of MCSS to the difference between transiogram models. Columns A and B in Figure 11 show simulated results using the poor expert's transiogram models and the perfect expert's transiogram models, respectively. It can be seen that simulated realizations using these two different sets of transiogram models have apparent differences in patterns. Although conditioning data are the same, the patterns of realizations in column B are much more complex than those in column A and tend to mimic the simulated realizations based on the dense sample set. One reason should be that the perfect expert's models effectively convey the spatial variation information of the variable into the simulation. The second reason is that the class proportion information of the dense sample set carried by the perfect expert's transiogram models affects simulated realizations. Apparently the smallest class, class 4, was reproduced in a larger proportion in the realizations in column B of Figure 11. Table 1 also shows that the class proportions estimated from the realizations based on the perfect expert's transiogram models (i.e., the transiogram models inferred from the dense sample set) are close to those of the dense sample set. However, the optimal maps based on

maximum probabilities (Figure 11: A0 vs. B0) do not display such differences for the two different sets of transiogram models. This is because the spatially certain information – the conditioning samples and their locations—are the same for the two simulations.

6. Conclusions

The validity of linear interpolation for joint transiogram modeling is mathematically demonstrated in this paper. The method can meet all the three constraint conditions for joint transiogram modeling. So as long as transition probabilities are correctly measured from a set of samples, linear interpolation can provide a valid set of transiogram models for MCRF simulation of the corresponding categorical variable. The major advantage of linear interpolation over model fitting is that the former can finish the transiogram-modeling process quickly regardless of the number and complexity of experimental transiograms, whereas the latter can be very tedious and time-consuming when the number of classes (or transiograms) is large.

When samples are abundant and experimental transiograms are reliable, linear interpolation is satisfactory or even superior to mathematical model fitting because it is efficient and can capture every small feature (i.e., peak and trough) of experimental transiograms. When samples are relatively sparse, we still can obtain reliable experimental transiograms by choosing a suitable lag tolerance so that linear interpolation can be used. When samples are very sparse (or very few), experimental transiograms become erratic with only a few measured values within the useful lag distance. Under this situation, both linear interpolation and precise model fitting do not make sense; therefore, we suggested using 'coarse model fitting' based on expert knowledge to obtain approximate transiogram models. Here 'coarse model fitting' means that model fitting should mainly consider expert knowledge and the general trend of experimental transiograms, and ignore their details because of their unreliability.

Our simulations of a five-class variable show that simulated results are similar for interpolated and fitted transiogram models when samples are dense and experimental transiograms are in regular shapes. But in this case study they also do not show large differences between these two kinds of transiogram models when samples are relatively sparse and a suitable lag tolerance is chosen, which may indicate that Markov chain simulation is not very sensitive to the difference in the detail of transiogram models as long as their change trends are identical. A special finding is that MCRF can capture the class enclosure feature in simulated results, which apparently should be attributed to the asymmetrical property of transiograms. Such a feature may not be captured by conventional indicator geostatistics.

Our simulations also show that expert knowledge can play an important role in transiogram modeling and Markov chain simulation, especially when samples are sparse and experimental transiograms cannot reflect the real spatial variation of the categorical variable under study. Although model fitting is tedious because of the large number of experimental transiograms involved in modeling a complex categorical variable, it actually becomes relatively easier to do when the number of samples is small because expert knowledge has to play a role and precise fitting to unreliable experimental transiograms becomes unnecessary.

References

Balzter, H., 2000. Markov chain models for vegetation dynamics. *Ecological Modelling*, 126, 139–154.

Bell, E.J., 1974. Markov analysis of land use change. An application of stochastic processes to remotely sensed data. *Socio-Economic Planning Science*, 8, 311–316.

Brown, L.A., 1970. On the use of Markov chains in movement research. *Economic Geography*, 46, 393–403.

Carle, S.F. and Fogg, G.E., 1997. Modeling spatial variability with one- and multi-dimensional continuous Markov chains. *Mathematical Geology*, 29, 891–918.

Collins, L., Drewett, R., and Ferguson, R., 1974. Markov models in geography. *The Statistician*, 23, 179–209.

Deutsch, C.V., 1998. Cleaning categorical variable (lithofacies) realizations with maximum a-posteriori selection. *Computers & Geosciences*, 24, 551–562.

Deutsch, C.V., 2006. A sequential indicator simulation program for categorical variables with point and block data: BlockSIS. *Computers & Geosciences*, 32, 1669–1681.

Deutsch, C.V. and Journel, A.G., 1998. *GSLIB: seostatistics software library and user's guide*. New York: Oxford University Press.

Goovaerts, P., 1996. Stochastic simulation of categorical variables using a classification algorithm and simulated annealing. *Mathematical Geology*, 28, 909–921.

Goovaerts, P., 1997. *Geostatistics for natural resources evaluation*. New York: Oxford University Press.

Horn, H.S., 1975. Markovian processes of forest succession. *In*: M.L. Cody and J.M. Diamond, eds. *Ecology and evolution of communities*. Cambridge, MA: Belknap Press, 196–211.

Journel, A.G. and Xu, W., 1994. Posterior identification of histograms conditional to local data. *Mathematical Geology*, 26, 323–359.

Li, W., 2006. Transiogram: a spatial relationship measure for categorical data. *International Journal of Geographical Information Science*, 20, 693–699.

Li, W., 2007a. Transiograms for characterizing spatial variability of soil classes. *Soil Science Society of America Journal*, 71, 881–893.

Li, W., 2007b. Markov chain random fields for estimation of categorical variables. *Mathematical Geology*, 39, 321–335.

Li, W. and Zhang, C., 2005. Application of transiograms to Markov chain simulation and spatial uncertainty assessment of land-cover classes. *GIScience & Remote Sensing*, 42, 297–319.

Li, W. and Zhang, C., 2007. A random-path Markov chain algorithm for simulating categorical soil variables from random point samples. *Soil Science Society of America Journal*, 71, 656–668.

Luo, J., 1996. Transition probability approach to statistical analysis of spatial qualitative variables in geology. *In*: A. Foster and D.F. Marriam, eds. *Geologic modeling and mapping*. New York: Plenum Press, 281–299.

Luo, J. and Thomsen, A., 1994. On direct estimation of the bivariate probability distribution of a regionalized variable from its spatial samples. *In*: A.G. Fabbri and J.J. Royer, eds. 3rd CODATA conference on geomathematics and geostatistics. *Sci. de la Terre, Ser. Inf.*, Nancy, 32, 115–123.

Machuca-Mory, D.F., Ortiz, J.M., and Deutsch, C.V., 2008. On the challenge of using sequential indicator simulation for the estimation of recoverable reserves. *International Journal of Mining, Reclamation and Environment*, 22, 285–299.

Potter, P.E. and Blakely, R.F., 1967. Generation of a synthetic vertical profile of a fluvial sandstone body. *Society of Petroleum Engineers of AIME Journal*, 7, 243–251.

Ritzi, R.W., 2000. Behavior of indicator variograms and transition probabilities in relation to the variance in lengths of hydrofacies. *Water Resources Research*, 36, 3375–3381.

Schwarzacher, W., 1969. The use of Markov chains in the study of sedimentary cycles. *Mathematical Geology*, 1, 17–39.

Tang, J., Wang, L., and Yao, Z., 2007. Spatio-temporal urban landscape change analysis using the Markov chain model and a modified genetic algorithm. *International Journal of Remote Sensing*, 28, 3255–3271.

Zhang, C. and Li, W., 2008. Regional-scale modelling of the spatial distribution of surface and subsurface textural classes in alluvial soils using Markov chain geostatistics. *Soil Use and Management*, 24, 263–272.