Published online May 16, 2007

Transiograms for Characterizing Spatial Variability of Soil Classes

Weidong Li*
Dep. of Geography
Kent State Univ.
Kent, OH 44242

The characterization of complex autocorrelations and interclass relationships among soil classes call for effective spatial measures. This study developed a transition probability-based spatial measure—the transiogram—for characterizing spatial heterogeneity of discrete soil variables. The study delineated the theoretical foundations and fundamental properties, and explored the major features of transiograms as estimated using different methods and data types, as well as challenges in modeling experimental transiograms. The specific objectives were to: (i) provide a suitable spatial measure for characterizing soil classes; (ii) introduce related knowledge for understanding spatial variability of soil types described by transiograms; and (iii) suggest methods for estimating and modeling transiograms from sparse sample data. Case studies show that (i) cross-transiograms are normally asymmetric and unidirectionally irreversible, which make them more capable of heterogeneity characterization, (ii) idealized transiograms are smooth curves, of which most are exponential and some have a peak in the low-lag section close to the origin, (iii) real-data transiograms are complex and usually have multiple ranges and irregular periodicities, which may be regarded as "non-Markovian properties" of the data that cannot be captured by idealized transiograms, and (iv) experimental transiograms can be approximately fitted using typical mathematical models, but sophisticated models are needed to effectively fit complex features. Transiograms may provide a powerful tool for measuring and analyzing the spatial heterogeneity of soil classes.

Abbreviations: TPM, transition probability matrix.

Many spatial attributes are classified into mutually exclusive multinomial classes (soil type is one example of multinomial categorical variables). For example, in a mesoscale study area such as a watershed that stretches across dozens of square kilometers, there may exist dozens of different soil series, which may be further grouped into several soil associations (Soil Conservation Service, 1962). Such soil series, particularly those within the same association, normally exhibit complex interclass relationships, such as cross-correlations, juxtapositions (i.e., side-by-side), and directional asymmetries in occurrence sequences. To describe the self-dependence (i.e., autocorrelation) of each soil class (e.g., soil series) and the interrelationships between different classes, theoretically sound spatial measures are needed.

Markov transition probability matrices (TPMs) (or one-dimensional first-order stationary Markov chain models) have a long history of use as spatial measures to describe and simulate spatial sequences of lithofacies in geology (e.g., Potter and Blakely, 1967; Krumbein, 1968; Carle and Fogg, 1997) and, more recently, soil layering in soil science (e.g., Li et al., 1999; Weigand et al., 2001). However, one-dimensional transition probability diagrams were rarely used or discussed before Carle and Fogg (1996), who suggested using transition probability diagrams in indicator kriging simulations of geological facies so that directional asymmetry and juxtaposition tendencies may be incorporated. Most of the studies to date have primarily used idealized transition probability diagrams—that is, they calculated transition probability diagrams from one-step transition probabilities based on a first-order stationary Markovian assumption. For example, Schwarzacher (1969) used Markov transition probability diagrams calculated from TPMs to describe the vertical change of lithofacies. Similarly, Weissmann and Fogg (1999) used such diagrams, derived from one-step transition probabilities obtained with the transition rate method in the framework of transition-probability-based indicator geostatistics, for hydrofacies modeling. Some discussions about the properties of idealized transition probability diagrams can be found in Carle and Fogg (1996, 1997) and Ritzi (2000). In contrast to traditional variograms, however, transition probability diagrams have rarely been estimated from sampled point data and also have not been effectively developed as independent spatial measures for heterogeneity characterization. The major reason may be that Markov chains have not been developed into widely applicable multidimensional geostatistical tools for conditional simulations on sparse point samples. Note that although Markov random field methods (including Markov mesh models) have been used in the geosciences for multidimensional simulations (e.g., Norberg et al., 2002; Wu et al., 2004), they did not involve interactions of nonadjacent neighbors (i.e., pixels), which are measured by multistep transition probabilities.

Li (2006) recently proposed the *transiogram* concept and the use of this term for transition probability diagrams, suggesting that the transiogram might be used as an independent spatial relationship measure for categorical data. There are two reasons for proposing such a spatial measure. First, there is a need for a physically meaningful, readily managed, and visual measure of the complex spatial intra-class and interclass relationships among multinomial

Soil Sci. Soc. Am. J. 71:881–893
doi:10.2136/sssaj2005.0132
Received 26 Apr. 2005.
*Corresponding author (weidong6616@yahoo.com).

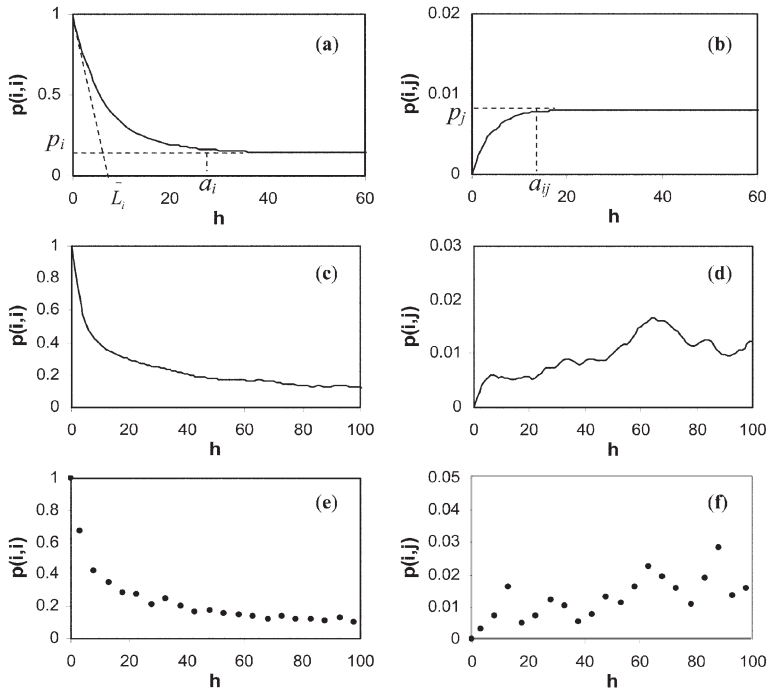

Fig. 1. Illustration of different types of transiograms: (a) idealized auto-transiogram; (b) idealized cross-transiogram; (c) exhaustive auto-transiogram; (d) exhaustive cross-transiogram; (e) experimental auto-transiogram; (f) experimental cross-transiogram. Experimental transiograms are estimated from sparse samples and therefore they are composed of discontinuous points; p(i,j) represents a transition probability from class *i* to class *j* where h **is the spatial lag separation vector.**

classes. Although TPMs may represent interclass relationships, they only measure one-step spatial relationships and fail to quantify the complex spatial relationships among classes across a number of spatial steps. Thus, transiograms may reveal more information than a TPM does. Transiograms directly estimated from data are not limited by the first-order stationary Markovian assumption and manifest the effect of spatial-lag-dependent interactions among observed data, thus they contain a wealth of features that reveal spatial heterogeneity of multinomial classes. Second, Markov chain conditional simulation models (Li, 2007b) need a more powerful spatial measure such that they are more flexible for working with various types of data (e.g., sparse sample point data) and incorporating complex spatial heterogeneity into simulations of multinomial classes.

A comprehensive assessment of transiograms is presented with respect to (i) basic properties and types; (ii) theoretical background of idealized transiograms; (iii) complex features of experimental transiograms estimated from multinomial spatial data (i.e., a mesoscale soil series map); and (iv) modeling of experimental transiograms estimated from sparse point samples. The primary objective was to introduce a suitable spatial measure for characterizing the complex spatial variability of discrete soil variables, particularly the interclass relationships among multinomial soil classes. Note that the word *class* is used as a general term for representing a category and the word *state* refers to the status of a Markov chain at a location.

THEORY

Transiogram

The Transiogram Concept

A brief introduction of the transiogram concept has been provided in Li (2006), and is summarized here. A *transiogram* is defined as a one-dimensional transition probability (i.e., two-point conditional probability) function across the distance lag **h**:

$$p_{ij}(\mathbf{h}) = \Pr\left[z(x+\mathbf{h}) = j \mid z(x) = i\right] \qquad [1]$$

where $p_{ij}(\mathbf{h})$ is the transition probability of random variable z from class i to class j. As lag **h** increases from the origin, $p_{ij}(\mathbf{h})$ generates a diagram, referred to as the transiogram. Therefore, $p_{ij}(\mathbf{h})$ is used to denote a transiogram, with the understanding that $p_{ii}(\mathbf{h})$ denotes the *auto-transiogram* of class i and $p_{ij}(\mathbf{h})$ ($i \neq j$) notates the *cross-transiogram* from class i to class j. Here **h** may be unidirectional (e.g., west to east) or non-unidirectional (e.g., multidirectional or omnidirectional). In a discrete space, **h** can be represented by the number of spatial steps (i.e., pixels). A transition spatial step means that the spatial process (e.g., Markov chain) moves from one pixel to the next adjacent pixel, that is, the distance of a pixel size. In a continuous space, **h** is represented by a continuous distance measure (e.g., meters). Because cross-transition probabilities are normally asymmetric, i and j are not interchangeable in $p_{ij}(\mathbf{h})$; for convenience, class i is called the *head class*, and class j is called the *tail class*.

In Eq. [1], Z is assumed to be a second-order stationary spatial random variable, that is, $p_{ij}(\mathbf{h})$ depends only on the lag **h** and not on the specific location x. This assumption, and the ergodic hypothesis (i.e., that spatial statistics are equivalent to ensemble statistics), allows estimation of transiograms directly from spatial data pairs in a spatial data set. Note that this framework and assumption are similar to that underlying variogram estimation, and as such refers to the model, since there is no physical basis on which the hypothesis can be tested with respect to the data per se (e.g., see Deutsch and Journel, 1998, p. 12–13; Chilès and Delfiner, 1999, p. 19–24). Therefore, a transiogram also represents a two-point spatial continuity/discontinuity measure, similar to an indicator variogram. Auto-transiograms represent self-dependence (i.e., autocorrelations) of single classes; however, cross-transiograms represent interclass relationships (including cross-correlations, juxtapositions, and directional asymmetries).

Types of Transiograms

Transiograms may be divided into two types based on how they are estimated: *idealized transiograms* and *real-data* (or *observed*) *transiograms*. The former refers to transiograms estimated using one-step transition probabilities based on the first-order stationary Markovian assumption. The latter refers to transiograms directly estimated from real data sets. Real-data transiograms may be further split into two subtypes based on the type of data sets used: *exhaustive transiograms* and *experimental transiograms*. Exhaustive transiograms refer to those transiograms directly estimated from maps or images where data are exhaustive. Experimental transiograms refer to those directly estimated from sparsely sampled data.

Figure 1 shows examples of different types of auto- and cross-transiograms. The first row (i.e., Fig. 1a and 1b) is a pair of idealized auto- and cross-transiograms. They are idealized because they are calculated from a one-step TPM, which implicitly imposes the first-order stationary Markovian assumption (i.e., that data are spatially stationary, aperiodic, and dependent on only immediate neighbors). Idealized transiograms are typified by smooth curves with stable sills and apparent correlation ranges. The second row of Fig. 1 shows a pair of exhaustive auto- and cross-transiograms (see Fig. 1c and 1d), which obviously have more features and seldom show stable sills or clear correlation ranges. Here the term *features* refers to the apparent peaks and troughs as well as multiple ranges shown

on a transiogram. These features usually should not be regarded as data noise. The third row of Fig. 1 provides a pair of experimental auto- and cross-transiograms (see Fig. 1e and 1f). Each is composed of a group of discontinuous points. Because real world data are usually strongly non-stationary and non-first-order Markovian, real-data transiograms have far more complex features than idealized transiograms have.

Basic Properties

Both real-data and idealized transiograms, as transition probability diagrams, have the following basic properties:

1. *Non-negative*. As probabilities, transiograms are non-negative, that is,

$$p_{ij}(\mathbf{h}) \geq 0 \qquad [2]$$

2. *Sum to 1*. As transition probabilities, values of transiograms involving the same head class sum to 1 at any estimated lag **h**, that is,

$$\sum_{j=1}^{n} p_{ij}(\mathbf{h}) = 1 \qquad [3]$$

where n is the number of classes, i is the head class, and j is the tail class.

3. *No nugget effect*. For mutually exclusive classes, theoretically and physically, transiograms should not have a nugget effect, that is, we have

$$p_{ij}(0) = 0, \quad i \neq j \qquad [4]$$

for cross-transiograms and

$$p_{ii}(0) = 1 \qquad [5]$$

for auto-transiograms, which must always hold true. Therefore, auto- and cross-transiograms should include the points (0,1) and (0,0), respectively, in their curves (see Fig. 1).

4. *Asymmetry*. Transiograms are typically asymmetric, that is

$$p_{ij}(\mathbf{h}) \neq p_{ji}(\mathbf{h}) \qquad [6]$$

The asymmetry property of transiograms provides more information about spatial juxtapositional relationships between classes. But if transiograms are estimated bidirectionally or omnidirectionally, the following relationship holds:

$$p_{ij}(\mathbf{h}) = \frac{p_j}{p_i} p_{ji}(\mathbf{h}) \qquad [7]$$

where p_i and p_j are the proportions of class i and class j, respectively. This means that if we know $p_{ji}(\mathbf{h})$ and the proportions of the two related classes in a study area (or a data set), we can directly obtain $p_{ij}(\mathbf{h})$.

5. *Irreversibility*. If transiograms are estimated unidirectionally, they are typically irreversible; that is,

$$p_{ij}(\mathbf{h}) \neq p_{ij}(-\mathbf{h}) \qquad [8]$$

The irreversibility property of unidirectional transiograms reflects the directional asymmetry of spatial patterns of classes.

The first three properties also constitute the set of constraint conditions of transiogram modeling for Markov chain simulation. The last two properties are special features of transiograms that other widely used spatial measures (e.g., variograms) may not have.

Idealized Transiograms

Although idealized transiograms are not a true reflection of real-world data or phenomena, they have theoretical and application value. For example, they can capture the basic correlation characteristics of classes and their properties are particularly helpful to interpreting real-data transiograms and modeling experimental transiograms. Idealized transiograms have been used explicitly (Weissmann and Fogg, 1999) or implicitly (through multistep transition probabilities calculated from one-step TPMs; Li et al., 2005) in geostatistical conditional simulations, with one-step transition probabilities estimated from borehole data or survey line data. For the sake of simplicity and efficiency, idealized transiograms may be used as simplified models when one-step transition probabilities are available. Therefore, understanding idealized transiograms is crucial, particularly for understanding real-data transiograms.

There are two simple methods to obtain idealized transiograms from one-step transition probabilities. The first uses one-step TPMs, while the second uses the transition rate matrix method, to derive transiograms (Carle and Fogg, 1997). Both methods are based on the first-order stationary Markovian assumption.

The First-Order Markovian Assumption

The first-order Markovian assumption states that the conditional distribution of a future state, given past states and the present state, is independent of past history and depends only on the present state. The first-order Markovian assumption (or property) can be expressed as

$$\begin{aligned} &\Pr[z(m+1) = k \mid z(m) = l, z(m-1) = r, \ldots, z(0) = s] \\ &= \Pr[z(m+1) = k \mid z(m) = l] = t_{lk} \end{aligned} \qquad [9]$$

where $z(m + 1),\ldots, z(0)$ represent a spatial state sequence of the random variable z; k, l, r, and s represent states of the random variable in a state space (1, 2, …, n); and t_{lk} represents the transition probability from state l to state k.

A first-order stationary Markov chain can be described by a TPM. For example, a three-state first-order Markov chain can be fully described by a TPM **P** with diagonal entries:

$$\mathbf{P} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \qquad [10]$$

where each entry of **P** represents a one-step transition probability between different states or within the same state with a fixed step length. In a discrete space, Eq. [10] is commonly used.

Sometimes, a first-order stationary Markov chain is expressed by a TPM without diagonal entries. Thus, the three-state first-order Markov chain given in Eq. [10] can be rewritten as

$$\mathbf{Q} = \begin{bmatrix} - & \pi_{12} & \pi_{13} \\ \pi_{21} & - & \pi_{23} \\ \pi_{31} & \pi_{32} & - \end{bmatrix} \qquad [11]$$

where each entry of **Q** represents a transition probability between different states. The Markov chain described by Eq. [11] is called an embedded Markov chain, because the spatial continuity of a state has to

be decided by a specific probability distribution and embedded into a Markov chain simulation (Potter and Blakely, 1967; Li et al., 1999).

Idealized Transiograms Derived from Transition Probability Matrices

Under the first-order stationary Markovian assumption, if n and m represent two distances in a spatial sequence and $m < n$, the first-order Markovian property can be formulated as

$$t_{lk}(n) = \sum_j \left[t_{lj}(m) t_{jk}(n-m) \right] \quad [12]$$

where $t_{lk}(n)$ denotes the transition probability from state l to state k across a distance n. This is the Chapman–Kolmogorov equation for homogeneous (i.e., stationary) processes (Agterberg, 1974, p. 420). If transition probabilities are arranged in the matrix format, Eq. [12] reduces to

$$\mathbf{P}(n) = \mathbf{P}(m)\mathbf{P}(n-m) \quad [13]$$

where $P(n)$ represents a TPM with the distance n.

In a discrete space, m and n can be numbers of spatial steps. Letting $m = 1$, we have

$$\mathbf{P}(n) = \mathbf{P}(1)\mathbf{P}(n-1) \quad [14]$$

Successively applying Eq. [14], we get

$$\mathbf{P}(n) = \mathbf{P}(1)^n \quad [15]$$

where $\mathbf{P}(n)$ represents an n-step TPM and $\mathbf{P}(1)$ a one-step TPM. As n increases, the entries in $\mathbf{P}(n)$ gradually become stable and reach the stationary probabilities—that is, different rows in $\mathbf{P}(n)$ will have the same values, with each entry $t_{lk}(n)$ being equal to the proportion of the tail class k.

Equation [15] represents the method to derive multistep transition probabilities from a one-step TPM. As n increases, the calculated multistep transition probabilities $t_{lk}(n)$ form a continuous diagram—an idealized transiogram $p_{lk}(h)$. If the one-step TPM is estimated from sufficient data and the study area is sufficiently large compared with the normal polygon size, the derived transiogram will have a sill equal to the proportion of the tail class k. Therefore, Eq. [15] is one method for deriving idealized transiograms. The prerequisite for this method is that we must have reliable one-step TPMs. The merits of this method are its simplicity and efficiency.

Idealized Transiograms Derived from Transition Rates

A general solution satisfying the Chapman–Kolmogorov equation (Agterberg, 1974, p. 457) is

$$\mathbf{P}(\mathbf{h}) = \exp(\mathbf{R}\mathbf{h}) \quad [16]$$

where $\mathbf{R}$ is a transition rate matrix (Krumbein, 1968), which is independent of the lag $\mathbf{h}$. For an m-state space,

$$\mathbf{R} = \begin{bmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mm} \end{bmatrix} \quad [17]$$

where entry r_{lk} of $\mathbf{R}$ represents the rate of change per unit length (step) from class l to class k. Off-diagonal entries of $\mathbf{R}$ can be calculated from transition probabilities π_{lk} of an embedded Markov chain and mean length $\overline{L_l}$ (for soil classes, a mean length refers to the mean polygon size [i.e., length or width] of a class, and is generally called the *mean boundary spacing*) of a class as

$$r_{lk} = \pi_{lk} / \overline{L_l} \quad \forall l \neq k \quad [18]$$

Analogously, the diagonal entries of $\mathbf{R}$ can be calculated by

$$r_{kk} = -1 / \overline{L_k} \quad [19]$$

For more details on this method, see Carle and Fogg (1997) and Weissmann and Fogg (1999). If $\mathbf{R}$ is known, with increasing lag $\mathbf{h}$, each entry of the TPM $\mathbf{P}(\mathbf{h})$ makes a continuous curve—an idealized transiogram.

Similarly, in a discrete space, $\mathbf{h}$ can be the number of spatial steps. If $\mathbf{h} = 1$, Eq. [16] reduces to

$$\mathbf{P}(1) = \exp(\mathbf{R}) \quad [20]$$

Thus, for n discrete steps, we have

$$\mathbf{P}(n) = \exp(\mathbf{R}n) = \mathbf{P}(1)^n \quad [21]$$

Therefore, in a discrete space, the transition rate method and the one-step TPM method for transiogram estimation are theoretically identical.

Comparing the two methods, it can be seen that the transition rate method does not require estimation of the diagonal transition probabilities, but the mean boundary spacing of each class must be known. When data are few and expert knowledge has to be used, it may be easier to estimate approximate mean boundary spacings than to estimate one-step self-transition probabilities. When detailed line data or exhaustive data (e.g., a training image) are available, both one-step self-transition probabilities and mean boundary spacings can be easily estimated. Major benefits of the transition rate method, according to Carle and Fogg (1997), are continuity in functional representation of the Markov chain model and flexibility afforded in choosing sampling intervals.

Properties of Idealized Transiograms

Besides the basic properties shared by all transiograms, idealized transiograms have the following properties, which may not hold for real-data transiograms.

An idealized auto-transiogram $p_{ii}(\mathbf{h})$ starts from point (0,1) and with increasing $\mathbf{h}$ gradually decreases to a stable value, the sill C_i (Fig. 1a). This sill, in a sufficiently large area, is equal to the proportion p_i of the class i in that area. Therefore, we have

$$\lim_{h \to \infty} p_{ii}(\mathbf{h}) = C_i = p_i \quad [22]$$

The lag distance at which the auto-transiogram approaches its sill is called the *autocorrelation range*, denoted by a_i. The autocorrelation range refers to the range within which observations of the same class are correlated. If the transiogram approaches its sill asymptotically, the *effective range* is set equal to the distance where the transiogram reaches 95% of its sill, similar to the effective range definition for asymptotic variogram models (Deutsch and Journel, 1998, p. 25).

While the autocorrelation range represents the distance of self-dependence of the class i, it does not directly tell the mean size of polygons of the class. If we draw a tangent of the auto-transiogram from

point (0,1) to the x axis, the lag $\mathbf{h}$ where the tangent crosses the x axis is equal to the mean of the polygon size of the class, denoted by $\overline{L_i}$, i.e.,

$$\partial p_{ii}(\mathbf{h})/\partial \mathbf{h}|_{\mathbf{h}=0} = -1/\overline{L_i} \quad [23]$$

(see Fig. 1a) (Carle and Fogg, 1996).

An idealized cross-transiogram $p_{ij}(\mathbf{h})$ starts from point (0,0) and with increasing $\mathbf{h}$ gradually increases to a stable value, the sill C_{ij} (Fig. 1b). The sill, for a sufficiently large area, is equal to the proportion p_j of the tail class j. That is, we have

$$\lim_{h\to\infty} p_{ij}(\mathbf{h}) = C_{ij} = p_j \quad [24]$$

Similarly, we define the *cross-correlation range* a_{ij} as the lag distance at which the sill is reached (or approached asymptotically). The cross-correlation range represents the distance within which the two involved classes are interrelated.

It should be noted that if one-step transition probabilities are estimated from a small area, sills of idealized transiograms, particularly idealized unidirectional cross-transiograms, may not be equal to proportions of corresponding tail classes in the area because boundary effects become substantial in small areas. The term *boundary effects* in this context refers to the fact that a class may have smaller transition frequencies in relation to other classes than those corresponding to the whole field if the class is more frequently adjacent to the boundaries of the study area, because boundary polygons are incomplete and transitions to other classes beyond boundaries are unknown (Li, 2006). In addition, because of the properties of idealized transiograms given in Eq. [22] and [24], apparently if class i and class j have different proportions, idealized cross-transiograms $p_{ij}(\mathbf{h})$ and $p_{ji}(\mathbf{h})$ have different sills. Actually, even if $p_j = p_i$, $p_{ij}(\mathbf{h})$ and $p_{ji}(\mathbf{h})$ will generally not be equal at any arbitrary lag $\mathbf{h}$ less than their respective correlation ranges, because they may not have the same curve shape due to the different juxtaposition characteristics of classes. These explain why idealized cross-transiograms are asymmetric.

Real-Data Transiograms

The Non-Markovian Effect

The term *Markovian property* usually refers to the *first-order* Markovian property and a *Markov chain* typically means a first-order Markov chain. So-called *high-order Markov chains* are actually Markovian representations of discrete non-Markovian processes. We will use the term *non-Markovian*, rather than the term *high-order Markovian*, to represent the non-first-order Markovian property of a discrete process. The non-Markovian property of data has seldom been considered in previous Markov chain models in the geosciences. The first-order Markovian property is an assumption for the convenience of creating first-order Markov chain models, not a property of the data. Real-world data sets typically have many features that are not consistent with a Markovian assumption. These non-Markovian properties imply that the current spatial state not only depends on its adjacent state but also depends on some nonadjacent states in a spatial state sequence. If a real-world data set is strongly non-Markovian, it may not be proper to use a first-order Markov chain model to deal with the data set.

Real-data transiograms have nothing to do with the first-order Markovian assumption. They are a direct reflection of the spatial variation characteristics of the original data. If transiograms from a real data set are basically identical with the corresponding idealized transiograms, we may say the data set is consistent with a first-order Markovian assumption. Otherwise, the extra features manifested by the real-data transiograms are a reflection of the non-Markovian properties of the data. Therefore, we may generally call such features the *non-Markovian effect* because they cannot be captured with a first-order Markov chain model.

Real-data transiograms apparently provide a way to quantitatively and visually represent the non-Markovian property of data, which reveal special features not represented by idealized transiograms. When real-data transiograms or well-fitted transiogram models are used in simulations, the non-Markovian effect of a data set is incorporated into the simulation. Thus, constraints imposed by first-order Markov chain models may be relaxed to some extent.

Exhaustive Transiograms

An exhaustive soil area-class map is essentially a simplified representation of the real soil landscape in a survey area based on field survey data, soil taxonomy, and expert interpretation. It is impossible to obtain a 100% accurate soil map, as an exhaustive survey is not feasible. The information contained in reliable soil survey maps such as those provided by the USDA are the result of practical experience gained through long-term soil survey efforts and represent the collective knowledge of experienced survey teams under support of the government. These soil maps contain considerable knowledge on spatial variability of soil types and their spatial relationships as discerned by experienced soil surveyors. It is thus appropriate to regard a reliable human-delineated soil map as a good representation of soil class spatial distributions (i.e., to assume the soil map bears the spatial statistics [e.g., transiograms] of the real soil distribution in the mapped area). Thus, transiograms estimated from a reliable soil map may reveal many characteristics of soil spatial distributions that are not readily apparent from a visual examination of the soil map. Exhaustive transiograms may be used (i) as a data-mining tool to extract knowledge on spatial correlations and interrelationships of soil classes from the existing digital soil maps, as the extracted knowledge is helpful for users to understand and interpret the mapped soils; and (ii) to provide direct transiogram models or expert knowledge for estimating transiogram models for soils in other areas that bear similarities to the mapped region.

Experimental Transiograms

Samples usually account for only a very small portion of the whole study area and yet often represent the major source of information on which a soil map is delineated. Therefore, the information conveyed by the sample data is most precious. Because of the costs associated with collection, and the subsequent sparsity of data, there is no doubt that experimental transiograms estimated from such sampling will be a series of discontinuous points and may not reveal nearly as abundant information as is provided in exhaustive transiograms. Even so, experimental transiograms can be used to characterize soil spatial variability and develop continuous transiogram models for stochastic simulation, similar to the way traditional experimental variograms have been used.

Modeling of Experimental Transiograms

Accurate modeling of transiograms is crucial to fully exploit the utility of transiograms for both soil heterogeneity characterization and Markov chain simulation. For heterogeneity characterization, following the convention in traditional geostatistical applications, the goal is to choose a suitable type of function model and to specify (statistically fit) model parameters to the sample data, thus providing a characterization of spatial heterogeneity in terms of a few simple parameters (e.g., such as the range, sill, wavelength, and model type, for instance, Gaussian, exponential, or spherical). For Markov chain simulation of soil classes, a crucial step is to parameterize a continuous transiogram model that captures as many major features of

Table 1. Some basic mathematical models for transiograms.

Model type	Function†
	For auto-transiograms‡
Linear	$p_{ii}(\mathbf{h}) = 1 - (1 - p_i)\mathbf{h}/a_i;\ \mathbf{h} < a_i$
	$p_{ii}(\mathbf{h}) = p_i;\ \mathbf{h} \geq a_i$
Spherical	$p_{ii}(\mathbf{h}) = 1 - (1 - p_i)[1.5(\mathbf{h}/a_i) - 0.5(\mathbf{h}/a_i)^3];\ \mathbf{h} < a_i$
	$p_{ii}(\mathbf{h}) = p_i;\ \mathbf{h} \geq a_i$
Exponential	$p_{ii}(\mathbf{h}) = 1 - (1 - p_i)[1 - \exp(-3\mathbf{h}/a_i)]$
Gaussian	$p_{ii}(\mathbf{h}) = 1 - (1 - p_i)\{1 - \exp[-(3\mathbf{h}/a_i)^2]\}$
Cosine–exponential	$p_{ii}(\mathbf{h}) = 1 - (1 - p_i)[1 - \exp(-3\mathbf{h}/a_i)\cos(b\mathbf{h})]$
Cosine–Gaussian	$p_{ii}(\mathbf{h}) = 1 - (1 - p_i)[1 - \exp(-3\mathbf{h}^2/a_i^2)\cos(b\mathbf{h})]$
	For cross-transiograms ($i \neq j$)
Linear	$p_{ij}(\mathbf{h}) = p_j\mathbf{h}/a_{ij};\ \mathbf{h} < a_{ij}$
	$p_{ij}(\mathbf{h}) = p_j;\ \mathbf{h} \geq a_{ij}$
Spherical	$p_{ij}(\mathbf{h}) = p_j[1.5(\mathbf{h}/a_{ij}) - 0.5(\mathbf{h}/a_{ij})^3];\ \mathbf{h} < a_{ij}$
	$p_{ij}(\mathbf{h}) = p_j;\ \mathbf{h} \geq a_{ij}$
Exponential	$p_{ij}(\mathbf{h}) = p_j[1 - \exp(-3\mathbf{h}/a_{ij})]$
Gaussian	$p_{ij}(\mathbf{h}) = p_j\{1 - \exp[-(3\mathbf{h}/a_{ij})^2]\}$
Cosine–exponential	$p_{ij}(\mathbf{h}) = p_j[1 - \exp(-3\mathbf{h}/a_{ij})\cos(b\mathbf{h})]$
Cosine–Gaussian	$p_{ij}(\mathbf{h}) = p_j[1 - \exp(-3\mathbf{h}^2/a_{ij}^2)\cos(b\mathbf{h})]$

† a_i = auto-correlation range; a_{ij} = cross-correlation range; p_i = proportion of class i; $b = 2\pi/\lambda$, where λ is the wavelength of the cosine function.

‡ The linear, spherical, exponential, and Gaussian models for auto-transiograms were also provided in Ritzi (2000).

the experimental transiograms as possible. This allows the soil spatial heterogeneity conveyed by the sample data to be effectively incorporated into the stochastic simulation, thus ensuring that such simulations more accurately reflect the sample data. When sample data are insufficient, experimental transiograms may not convey reliable information (e.g., they may exhibit spurious fluctuations). In this situation, expert knowledge regarding plausible values for correlation ranges, sills, and transiogram model type are required to build a suitable model.

Basic Transiogram Models

One may use any suitable mathematical model to represent an experimental transiogram, such as some of the classic mathematical models that are widely used in modeling auto-variograms (Deutsch and Journel, 1998, p. 25), which may be readily adapted to model auto- and cross-transiograms. Such suitable models include the exponential, spherical, Gaussian, and linear mathematical (theoretical) models. The observed characteristics of experimental auto- and cross-transiograms can be used to develop anisotropic models that satisfy the theoretical constraints implicit in a transiogram.

Recall that sills of idealized transiograms are theoretically equal to the proportions of their respective tail classes. Although experimental transiograms are normally more complex in shape than idealized transiograms and may thus not manifest a stable sill, the magnitudes are still a reflection of the tail class proportions. Therefore, setting the sill of a transiogram model to the proportion of the tail class is appropriate. Ritzi (2000) provided some basic mathematical models for auto-transition probabilities (i.e., auto-transiograms) with the model sills being set equal to the class proportions.

The shape of a cross-transiogram with increasing lag is similar to that of an auto-variogram. Thus, we can simply set sills of mathematical models to the proportions of the corresponding tail classes for modeling cross-transiograms (Li, 2007a). The exponential model for cross-transiograms can thus be given as

$$p_{ij}(h) = p_j\left[1 - \exp(-3h/a_{ij})\right] \quad [25]$$

Since the maximum value of an auto-transiogram occurs at $\mathbf{h} = 0$, the model sills are prescribed differently. The corresponding exponential model for auto-transiograms can be given as

$$p_{ii}(h) = 1 - (1 - p_i)\left[1 - \exp(-3h/a_i)\right] \quad [26]$$

(Ritzi, 2000). Equation [26] ensures that the sill of the auto-transiogram model is equal to the proportion of the modeled class. Sills of other models can be set similarly. Basic mathematical models for auto- and cross-transiograms are given in Table 1. Note that the nugget model used for modeling variograms is discarded here because transiograms for multinomial classes should not have a nugget effect.

In many cases, these basic models are sufficient to fit the general trend of an experimental transiogram. They are not capable of modeling finer details of experimental transiograms, however, particularly a cross-transiogram that exhibits a peak or significant non-Markovian effects.

When a study area is small and the sill of an experimental transiogram obviously deviates from the theoretically expected sill (i.e., the proportion of the tail class) because of the boundary effect, setting the sill of the transiogram model to the proportion of the tail class may result in a fit to the experimental transiogram that appears inaccurate; however, the transiogram model thus derived should be a better reflection of the spatial heterogeneity.

Ma and Jones (2001) suggested using multiplicative-composite models to fit hole-effect auto-variograms, or at least to fit the first peak if a hole-effect auto-variogram is irregular. The cosine–exponential model and the cosine–Gaussian model used by Ma and Jones (2001) are adapted here in the same way for modeling periodicities of transiograms (see Table 1). These two models attenuate the sinusoidal amplitude with increasing lag; however, the cosine–Gaussian model has higher peak heights than the cosine–exponential model. In many situations, one may find these hole-effect models still do not adequately fit experimental transiograms of multinomial classes, because the latter are too complex and irregular. Nested models constructed from these basic models may be useful for some complex transiograms.

Transiogram Model Constraints for Stochastic Simulation

If the purpose in constructing transiogram models is solely to characterize soil spatial variability (e.g., with no attempt to create a stochastic simulation), there is no need to strictly adhere to the theoretical constraints. When creating transiogram models for use in Markov chain simulations, however, the aforementioned three constraint conditions (Eq. [2–4]) must be satisfied to correctly perform simulations.

The requirement that the transiogram model is void of a nugget effect (i.e., Eq. [4]) can be satisfied by including the points at 0 lag [i.e., point (0,1) for auto-transiograms and point (0,0) for cross-transiograms] into experimental transiograms and honoring these values in fitted models. One option for the transiogram models fitting a subset of experimental transiograms headed by the same class to satisfy the summing-to-1 constraint (i.e., Eq. [3]) is to infer one of them from others by the following equation:

$$p_{ik}(\mathbf{h}) = 1 - \sum_{\substack{j=1 \\ j \neq k}}^{n} p_{ij}(\mathbf{h}) \quad [27]$$

where n is the number of classes, i is the head class, and $p_{ik}(\mathbf{h})$ is the inferred model (Li, 2007a).

To guarantee that $p_{ik}(\mathbf{h})$ is non-negative and consistent with the experimental data, the model-fitting process of other transiograms headed by the same class i may need repetitive tuning. Normally, if the sill of every transiogram model is set to the proportion of the corresponding tail class and the section close to the origin is appropriately fitted by choosing a suitable model type and range, $p_{ik}(\mathbf{h})$ will be consistent with the experimental data. The non-negative condition is not a concern for the other experimental transiograms fitted by mathematical models because experimental transiograms are never negative, which means models fitting them will not be negative.

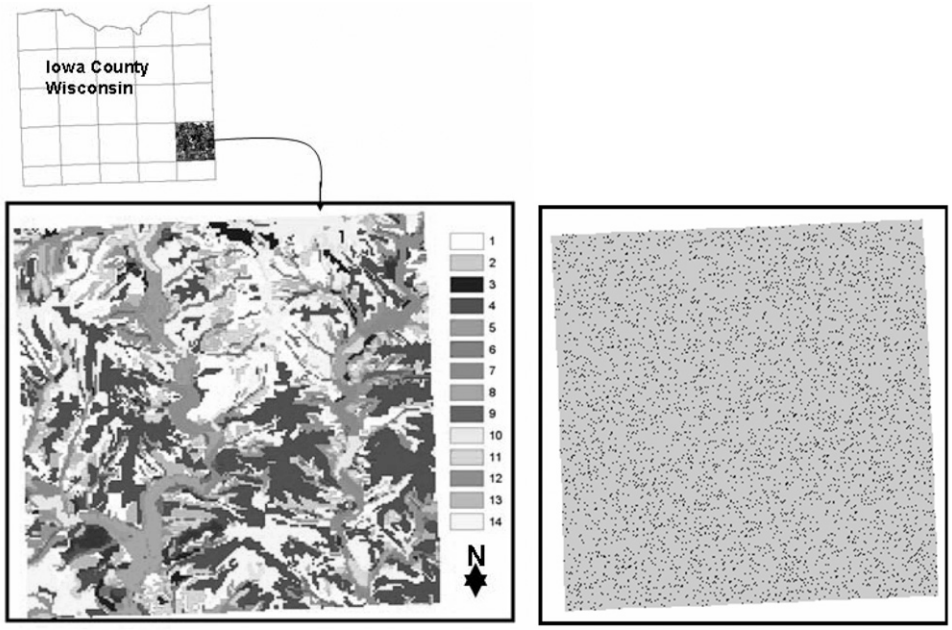

Fig. 2. A large raster soil series map (part of Iowa County, Wisconsin) and a corresponding random sample data set obtained by simulated sampling of the map. The map has a total of 48 soil series; of these, the hues of the 14 most dominant soil series are given (their corresponding names are given in Table 2).

CASE STUDIES

Data Sets

Case studies here aim to demonstrate the features of different types of soil class transiograms and the corresponding modeling of those experimental transiograms.

An exhaustive data set was used for estimating idealized transiograms and exhaustive real-data transiograms. This data set is a large digital soil series map, which covers a part of Iowa County, Wisconsin. The county soil map was downloaded from the NRCS Soil Survey Geographic Database (SoilDataMart.nrcs.usda.gov/). The entire county soil map is too large to display. Therefore, the county was divided into several subareas and only one subarea was chosen for the purposes of this study. The mapping unit in the original map is the soil phase, which is a classification level for the purpose of soil management. The soil phases were merged into corresponding soil series, a higher classification level that is based on pedogenic properties. The clipped soil series map of the chosen subarea is given in the left side of Fig. 2, which has an extent of about 9.6 by 9.6 km^2. For the convenience of data analysis, the map was discretized into a raster image of about 500 by 500 pixels with a pixel size of 20 by 20 m^2. The soil map contains 48 soil series, thus 48 by 48 transiograms may be calculated for the study area in each direction. In this study, only some of transiograms along axis directions were estimated and analyzed.

Among these soil types, the names, proportions, and polygon information for 14 of the 48 are provided in Table 2. These data indicate that some soil types (e.g., 6, 11, and 13) have smaller polygons and some types (e.g., 4 and 5) have larger polygons and that each soil type accounts for a different area proportion. From Fig. 2 it can be seen that Soil Type 4 dominates the study area and the right boundary. The polygon data shown for these soil types may be indicative of the more general complexity of polygon shapes for soils. Note that the soil type numbers listed in Table 2 are identical with those used in the following related figures of transiograms.

To estimate experimental transiograms, two sparse data sets—one relatively large and the other considerably smaller—were used. The larger "sparse" data set was obtained by a random sampling, with n = 12936, of the large raster soil map (Fig. 2, right). The small, sparse data set was obtained using a grid-based sampling with 136 points (Fig. 3, bottom), from a small soil map with seven soil types (Fig. 3, top). The seven soil types of the small soil map are distributed relatively uniformly within a 4.0- by 1.7-km^2 area, which was discretized into an 80 by 34 raster grid with a pixel size of 50 by 50 m^2 (Fig. 3) (also see Li et al. [2005]). The purpose of using this small data set with a small number of soil types is to demonstrate how to jointly model experimental transiograms for Markov chain simulation, because transiogram models for simulation have to be estimated one subset at a time so that they can meet the constraint conditions. Here a subset contains all transiograms headed by the same class. The reason for using regular data is because they are more efficient for estimating reliable experimental transiograms; otherwise the small data set of 136 points would be insufficient to estimate the 49 (i.e., 7 by 7) experimental auto- and cross-transiograms. The specific

Table 2. Names, area proportions, and some polygon characteristics of the first 14 of the 48 soil series (or land types) appearing in the soil map shown in Fig. 2.

No.	Name	Proportion	Polygons	Average area of polygons	Average perimeter of polygons
			no.	m^2	m
1	Northfield, sandy loam	0.1425	168	79076	1790
2	Dubuque, stony silt loam	0.0192	43	41559	1308
3	Fayette, silt loam, valleys	0.0081	30	25049	964
4	Dubuque, deep	0.2389	141	157945	2362
5	Huntsville, silt loam	0.0163	7	217117	2846
6	Chaseburg, silt loam	0.0321	88	33964	1621
7	Orion, silt loam	0.0311	22	131690	2629
8	Ettrick, silt loam	0.0320	12	248491	3321
9	Tell, silt loam	0.0004	1	36844	930
10	Lawson, silt loam	0.0154	15	96009	1871
11	Northfield, loam	0.0661	145	42518	1195
12	Water	0.0013	8	14743	1236
13	Gale, silt loam	0.0438	113	36100	1014
14	Dubuque, ordinary	0.1616	150	100416	1986

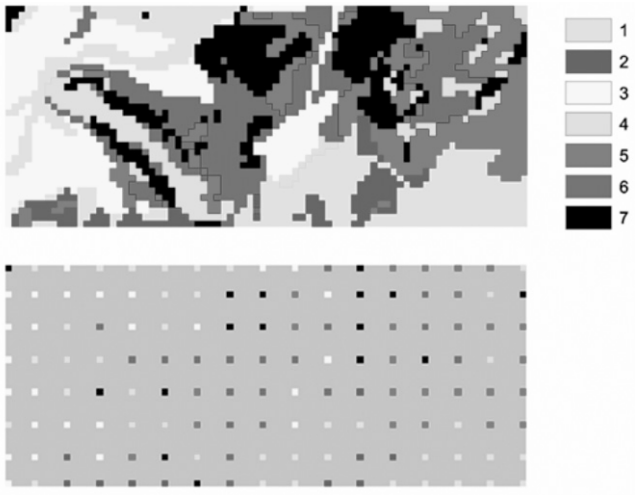

Fig. 3. A small soil map with seven classes and a corresponding regular sample data set (*n* = 136) from this map.

names for these soil types are not relevant here, so the numbers (i.e., 1, 2, 3, …, 7) are used instead to represent the seven different soil types.

Estimation of Transiograms from Data

Transiograms can be estimated by counting transition frequencies of soil types in a data set with different lags. The equation is given as follows:

$$p_{ij}(\mathbf{h}) = F_{ij}(\mathbf{h}) \Big/ \sum_{j=1}^{n} F_{ij}(\mathbf{h}) \qquad [28]$$

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 soil types. If directional asymmetry and anisotropy are not considered, transition frequencies in different directions may be pooled together to get multidirectional or omnidirectional transiograms.

When transition frequencies are counted from exhaustive data sets, the lag $\mathbf{h}$ may be set exactly with a zero tolerance width. Thus, estimated transiograms are continuous functions of lag separation, which may be conveniently expressed as the corresponding number of pixels. As such, they may capture considerable fine-scale resolution of spatial variability structures among and between the several soil types. When data are sparse, however, the number of observed transitions at certain lags will probably not be sufficient to provide an appropriate probability estimate (e.g., estimated transition probabilities may be zero or unrealistically high). Therefore, a tolerance width should be used when estimating experimental transiograms from sparse data. A tolerance width $\Delta\mathbf{h}$ (e.g., 30 m) means that all transitions within a lag interval from ($\mathbf{h} - \Delta\mathbf{h}$) to ($\mathbf{h} + \Delta\mathbf{h}$) (e.g., 400 to 460 m) are counted into the frequency of transitions at one lag $\mathbf{h}$ (e.g., 430 m). Thus, only transition probabilities at a limited number of lags are estimated. Consequently, experimental transiograms are composed of a small number of discrete, typically equally spaced points, similar to the convention followed for variogram estimation with sparse data. In this study, because raster data were used to estimate transiograms, the pixel number was used as a relative distance to represent the lag $\mathbf{h}$.

RESULTS AND DISCUSSION

Idealized Transiograms

To examine the features of idealized transiograms, one-step TPMs were estimated from the large soil map in different directions and transiograms were derived from these TPMs using Eq. [15]. Figure 4 shows some of these idealized transiograms in the west-to-east direction, all of which are smooth curves. Among these, the auto-transiograms are exponential, each with a generally unique correlation range. The sill values for these idealized transiograms are in accordance with the proportions of the corresponding soil types given in Table 2. The shapes of most cross-transiograms are also reasonably consistent with an exponential model, although some of them, such as $p_{1,13}(\mathbf{h})$ and $p_{1,11}(\mathbf{h})$ (see Fig. 4d) are not monotonically increasing, with a peak, or maximum value, occurring before reaching the sill value. This peak is a typical indication of juxtaposition relationships, which results when one class frequently occurs in association with another class. Examining the transiogram sills, it is apparent that these cross-transiograms also have sills that are identical to the proportions of the corresponding tail classes. For example, the sill of $p_{1,14}(\mathbf{h})$ is 0.1616, which is equal to the proportion of Soil Type 14 (see Table 2).

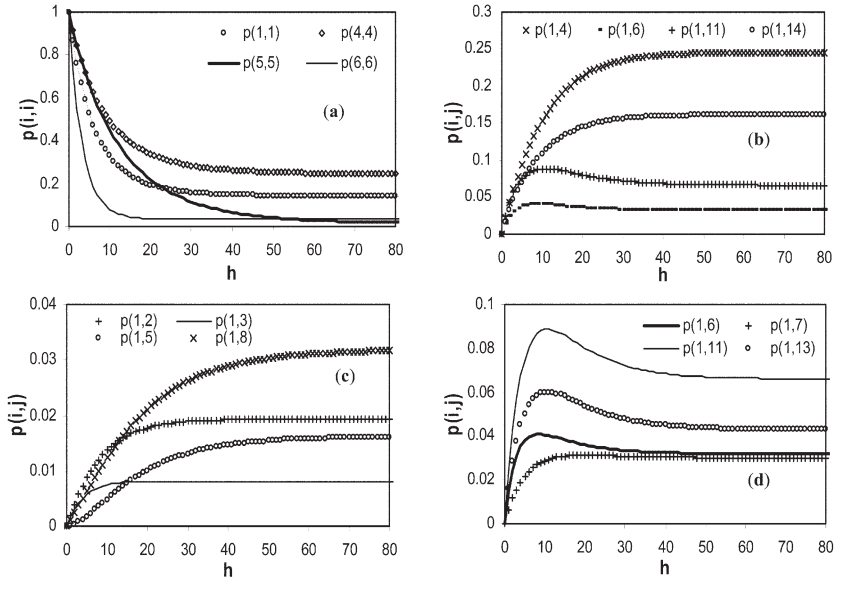

Fig. 4. Idealized transiograms calculated from the one-step transition probability matrix (estimated from the large soil map (shown in Fig. 2) in the west-to-east direction): (a) auto-transiograms; (b), (c), and (d) cross-transiograms. The notation like p(i,j) refers to the transiogram $p_{i,j}$(h), representing a transition probability from class *i* to class *j* where h **is the spatial lag separation vector. The scale along the** h **axis is the number of grid units.**

Idealized cross-transiograms are asymmetric and, if estimated unidirectionally (i.e., on the basis of unidirectional TPMs), are also irreversible. These characteristics can be seen from Fig. 5, which shows idealized cross-transiograms between Class 1 and Class 6 in the west-to-east direction and the east-to-west direction. Asymmetry indicated by the discrepancy between $p_{1,6}(\mathbf{h})$ and $p_{6,1}(\mathbf{h})$ is clear because of their different sills. But the discrepancy between cross-transiograms $p_{1,6}(\mathbf{h})$ and $p_{1,6}(-\mathbf{h})$ is evident only at short lag separations, because these two cross-transiograms have equal sills corresponding to the proportion of the same Tail Class 6.

Notice that the slight difference in the sills of $p_{6,1}(\mathbf{h})$ and $p_{6,1}(-\mathbf{h})$ shown here (see Fig. 5) results from the boundary effect of the limited

study area; theoretically, this effect would not occur if the study area were infinitely large. It also should be noted that the idealized transiograms shown here are similar to the continuous transition probability models derived using the transition rate matrix method as shown in Carle and Fogg (1997) and Weissmann and Fogg (1999), since they are all based on the first-order Markovian assumption and are essentially the equivalent representations in a discrete space.

Although not able to capture non-Markovian effects of spatial data, transiograms derived using one-step TPMs provide a simple and efficient way for transiogram estimation and modeling. This approach may save considerable time and effort if one-step TPMs are available (e.g., from survey line data or training images). A requirement for using this approach is that one-step TPMs estimated from data must be reliable; that is, they should reflect the transition relationships among different classes. If survey line data or large training maps are not available, or the study area is too small compared with the parcel sizes of the soil types, it may be difficult to obtain reliable one-step TPMs.

Exhaustive Transiograms

To demonstrate typical features of exhaustive transiograms, the latter were estimated from the large soil map using a tolerance width of zero. Both unidirectional and "four-directional" transiograms (i.e., computed as the average of the transiograms calculated in the four cardinal directions—east, west, south, and north) were estimated. These transiograms provide an opportunity to gain general understanding and insight into the spatial structure of the variability in soil types, and more importantly, it demonstrates typical features of real transiograms and the spatial features that cannot be captured with a first-order Markovian characterization.

Shape and Periodicity

Figure 6 shows exhaustive real-data auto- and cross-transiograms estimated from the large soil map in an easterly direction. It can be seen that all auto-transiograms shown here can be reasonably described using exponential models. Note that the auto-transiogram $p_{5,5}(\mathbf{h})$ has a longer auto-correlation range than the other soil types, which is a consequence of its large average polygon size (see Table 2). Note also that the fluctuations in auto-transiograms are relatively weak compared with those in cross-transiograms.

Apparent fluctuations can be seen in cross-transiograms, especially with an expanded **h** axis (see Fig. 6d). These cross-transiograms are approximately consistent with exponential or spherical models for lags less than the first peak, although the spatial patterns and periodicities are typically distinct for each cross-transiogram. A periodic pattern in transition frequencies indicates characteristic scales in spatial patterns of soil types, such as occurs in swell–swale landscape sequences (Haws et al., 2004) (i.e., parcels of soil types appear in a space with a spatial regularity or rhythm).

"Four-directional" (not truly omnidirectional) transiograms are displayed in Fig. 7. As a consequence of the pooling across directions, these average curves are considerably smoother than

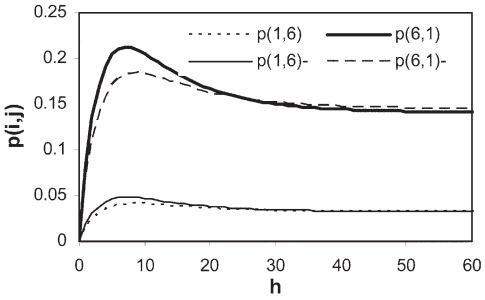

Fig. 5. Illustration of the asymmetry and irreversibility of idealized transiograms. The transiograms p(1,6) and p(6,1) [referring to $p_{1,6}(h)$ and $p_{6,1}(h)$, respectively] are estimated along the west-to-east direction, whereas the transiograms p(1,6)– and p(6,1)– [referring to $p_{1,6}(-h)$ and $p_{6,1}(-h)$, respectively] are estimated along the east-to-west direction.

their unidirectional counterparts. As a consequence of the averaging process, the periodicities displayed in the four-directional cross-transiograms generally differ from the corresponding unidirectional cross-transiograms. Four-directional or omnidirectional transiograms may be used in Markov chain simulations when it is deemed that anisotropies and directional asymmetries are not of interest.

Correlation Range and Non-Markovian Effect

Examination of the correlation ranges associated with the exhaustive transiograms in Fig. 6 and 7 show that some of the auto-transiograms have a principal autocorrelation range of about 15 to 30 pixels (i.e., 300–600 m). That is, they first decrease quickly to a "principal range," and then, rather than remaining stable at a fixed level, gradually decrease and [see $p_{4,4}(\mathbf{h})$ and $p_{1,1}(\mathbf{h})$ in Fig. 6a and 7a] approach their theoretical sills (i.e., the proportions of the corresponding classes). This principal range characteristic and subsequent transition to the theoretical value differs from that of an idealized auto-transiogram, which reaches its theoretical sill quickly and stabilizes at this level with increasing lags. Figure 8 contrasts these two types of transiograms, demonstrating the differences between each.

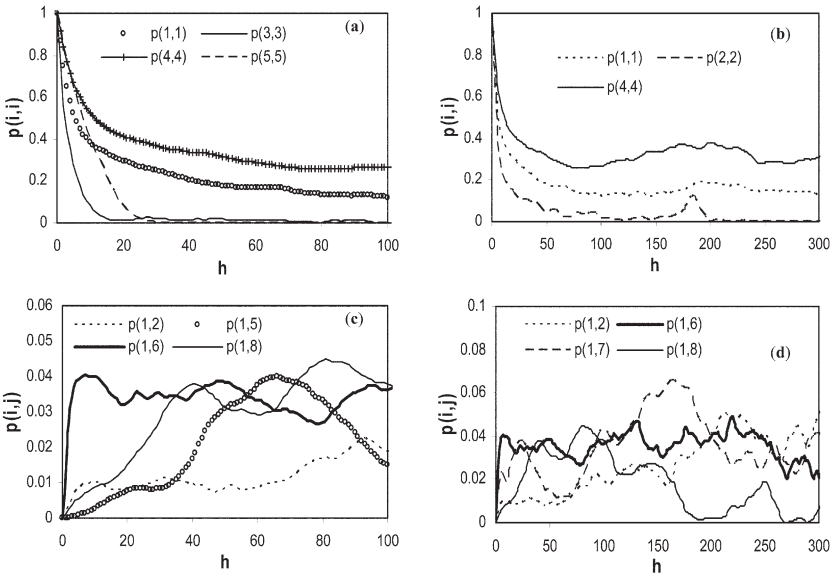

Fig. 6. Exhaustive transiograms estimated from the large soil map (shown in Fig. 2) in the west-to-east direction: (a) and (b) auto-transiograms; (c) and (d) cross-transiograms. The right column [i.e., (b) and (d)] have longer lags so as to demonstrate irregular fluctuations.

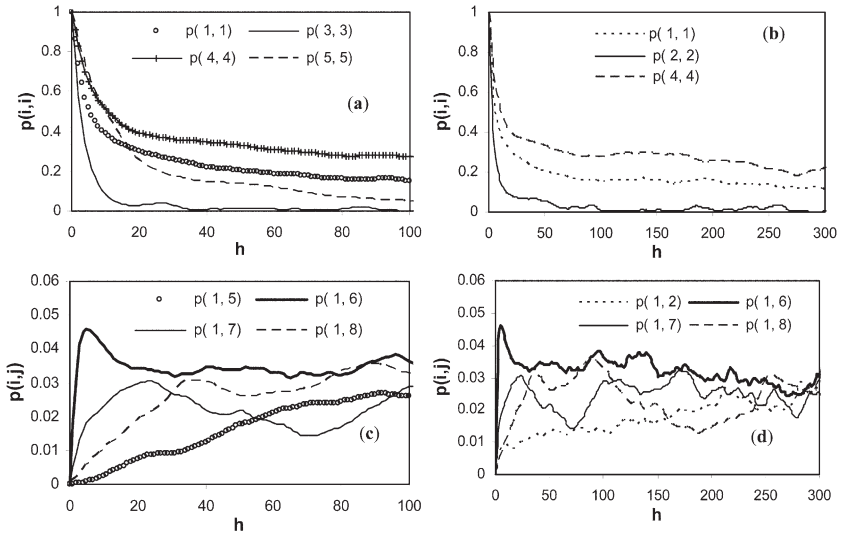

Fig. 7. Exhaustive transiograms estimated from the large soil map (shown in Fig. 2) by pooling transitions along the four cardinal (westward, eastward, southward, and northward) directions: (a) and (b) auto-transiograms; (c) and (d) cross-transiograms. The right column [i.e., (b) and (d)] have longer lags to demonstrate irregular fluctuations.

The principal range of an exhaustive auto-transiogram is a reflection of the first-order dependence (i.e., first-order Markovian property) of the data and the continuing decrease thereafter is a consequence of the high-order dependence (i.e., non-Markovian property) of the data that is not captured in an idealized transiogram.

Cross-correlation ranges of exhaustive cross-transiograms are difficult to discern because of the strong fluctuations in these curves. If the lag immediately subsequent to the first peak is regarded as the principal range of an exhaustive cross-transiogram, then these figures suggest that the principal ranges (see Fig. 6c) lie between about 10 and 65 pixel lengths (i.e., 200–1200 m) in the west-to-east direction. The irregular periodicities of these cross-transiograms (see Fig. 6d and 7d) are a strong indication of the existence of non-Markovian effects among different soil types, which is not captured by idealized cross-transiograms. Some cross-transiograms, such as $p_{1,6}(\mathbf{h})$ in Fig. 7c, have an obvious peak near the origin. This demonstrates that Soil Types 1 and 6 are spatially associated (i.e., are frequently found in close proximity).

In addition, some exhaustive cross-transiograms, such as $p_{1,2}(\mathbf{h})$, reveal two ranges whereas the corresponding idealized cross-transiograms have only one (see Fig. 8), which reflects the fact that exhaustive transiograms may contain long-distance cross-correlations. It is also apparent that near the origin (i.e., for lag distances <10 pixels), idealized auto-transiograms correspond very well to their exhaustive auto-transiogram counterparts (see Fig. 8). This implies that the mean polygon size of a class can be estimated from exhaustive auto-transiograms.

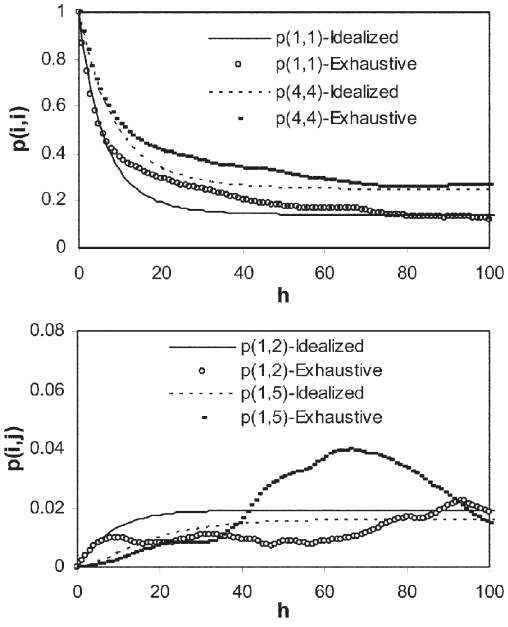

Fig. 8. A comparison of idealized transiograms and exhaustive real-data transiograms as estimated from the large soil map (shown in Fig. 2) in the west-to-east direction.

Sill and Tail Class Proportion

Recall that the sill of an idealized transiogram is equal to the corresponding tail class proportion. Examination of the sills of exhaustive auto-transiograms (Fig. 6 and 7) reveals that they also approach the proportions of the corresponding soil types, although perhaps more gradually. For example, $p_{1,1}(\mathbf{h})$ reaches the proportion of Soil Type 1 (i.e., 0.1425) at a lag of 75 pixels in Fig. 6 and at a lag of 160 pixels in Fig. 7, with slight fluctuations about this level. The strong fluctuations evident in exhaustive cross-transiograms make it difficult to identify sill values, but these values should correspond to the average amplitude in the fluctuating curves after the first peak. Upon examination, it can be seen that the sill values so defined are approximately equal to the proportions of the corresponding tail soil types. In general, as long as the area of the soil map is large enough, the sills of exhaustive transiograms should approach, or fluctuate about, the proportions of the corresponding tail classes. Thus, exhaustive transiograms provide a measure of tail class proportions.

Asymmetry and Irreversibility

Figure 9 shows four exhaustive cross-transiograms between Soil Types 1 and 2, which were estimated from the large soil map in the west-to-east direction [i.e., $p_{2,1}(h)$ and $p_{1,2}(\mathbf{h})$] and the east-to-west direction [i.e., $p_{2,1}(-\mathbf{h})$ and $p_{1,2}(-\mathbf{h})$]. It is evident that $p_{2,1}(\mathbf{h}) \neq p_{1,2}(\mathbf{h}) \neq p_{2,1}(-\mathbf{h}) \neq p_{1,2}(-\mathbf{h})$. The cross-transiograms $p_{2,1}(\mathbf{h})$ and $p_{2,1}(-\mathbf{h})$ have the same theoretical sill (i.e., the proportion of Soil Type 1), which is 0.1425, but their curve patterns (detailed fluctuations) are different; whereas $p_{2,1}(\mathbf{h})$ and $p_{1,2}(-\mathbf{h})$ have similar curve patterns (peaks and troughs), but their theoretical sills, which are actually their curve heights at different lags, are all different. Further, $p_{2,1}(\mathbf{h})$ and $p_{1,2}(\mathbf{h})$ have no common characteristics. Such general features exist for any two soil types, illustrating the properties of asymmetry and irreversibility of unidirectional cross-transiograms.

The asymmetry property of cross-transiograms between two classes [i.e., $p_{ij}(\mathbf{h}) \neq p_{ji}(\mathbf{h})$] holds for both unidirectional and multidirectional cross-transiograms. This follows since their theoretical sills, the proportions of class i and j, are typi-

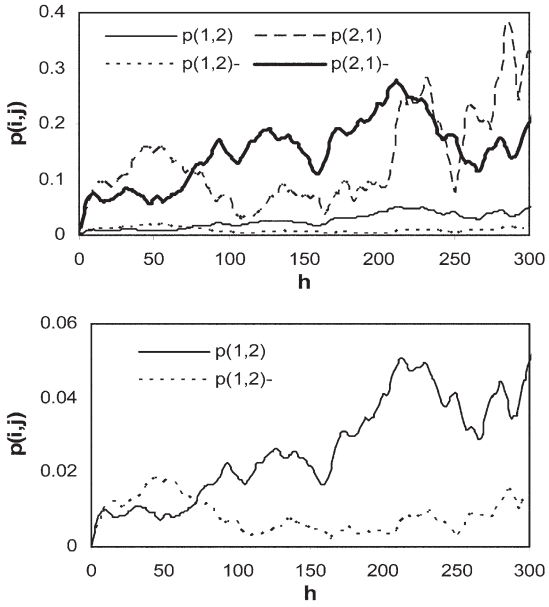

Fig. 9. Exhaustive cross-transiograms between Soil Types 1 and 2 estimated from the large soil map (shown in Fig. 2) along the west-to-east direction and the east-to-west direction. Transiograms p(1,2)– and p(2,1)– [referring to $p_{1,2}(-\mathbf{h})$ and $p_{2,1}(-\mathbf{h})$, respectively] are estimated along the east-to-west direction.

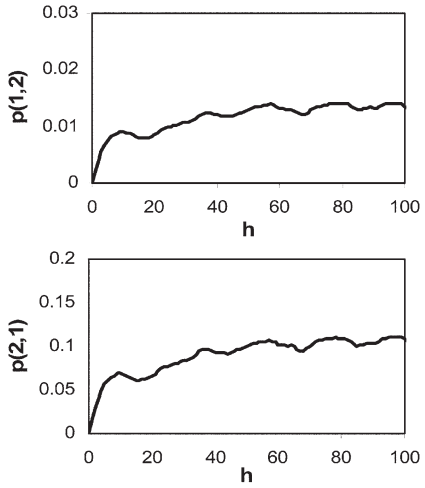

Fig. 10. Exhaustive cross-transiograms between Soil Types 1 and 2 estimated from the soil map by pooling transitions along the four cardinal directions (westward, eastward, southward, and northward) together.

cally different. When transition frequencies in opposite directions are pooled together, however, cross-transiograms $p_{ij}(\mathbf{h})$ and $p_{ji}(\mathbf{h})$ have the same patterns. Figure 10 shows the four-directional transiograms between Soil Types 1 and 2. Clearly, $p_{2,1}(\mathbf{h})$ and $p_{1,2}(\mathbf{h})$ have the same curve shape but different sills, and their quantitative relationship satisfies Eq. [7].

The asymmetric property of cross-transiograms demonstrates the capability of characterizing juxtaposition relationships among classes, in contrast to symmetric spatial measures. For example, consider a hypothetical situation in which Class A has a higher occurrence frequency (e.g., 100), whereas Class B has a lower occurrence frequency (e.g., 10), and further assume that Class B always occurs with Class A as neighbors. In such a situation, it is expected that the transition probability for $p_{A,B}(\mathbf{h})$ at a short lag **h** will be smaller than that for $p_{B,A}(\mathbf{h})$. Such juxtaposition relationships between classes can be detected by cross-transiograms, as can be clearly seen in Fig. 9, which shows that $p_{2,1}(\mathbf{h}) > p_{1,2}(\mathbf{h})$. The irreversible property of unidirectional cross-transiograms (i.e., that the cross-transiogram is not an even function) means that these functions can detect the directional asymmetry of class occurrence sequences. Although cross-transiograms $p_{ij}(\mathbf{h})$ and $p_{ij}(-\mathbf{h})$ have the same theoretical sill, their different curve patterns, particularly differences in the number, lag distance, and magnitude of peaks and troughs, reveal the directional asymmetries.

Modeling Experimental Transiograms

Figure 11 shows experimental transiograms estimated with a tolerance width of two pixels (i.e., $\Delta\mathbf{h} = 2$), which means that the average transition probability per five pixels along the **h** axis is plotted. Several parameters must be estimated to fit experimental transiograms with basic mathematical models. These include the sill (i.e., proportions of tail classes), correlation range, and model type. The tail class proportions can be estimated from the sampled data set. The correlation ranges and model types can be approximately discerned from the experimental transiograms. The resulting estimates of these three parameters are demonstrated in association with the modeled transiograms in Fig. 11, where three of the transiograms are described with exponential models and the remaining with a spherical model. It is apparent that these basic mathematical models have difficulty in capturing the complex features (i.e., peaks, troughs, and sometimes double ranges) of experimental transiograms. This is especially obvious in $p_{1,4}(\mathbf{h})$ and $p_{2,1}(\mathbf{h})$ (see Fig. 12), where the former manifests two ranges and the latter has an irregular periodicity.

With use of a suitable tolerance width and a tolerance angle (the latter is not used here), the data points needed for reliable estimation of experimental transiograms can be significantly

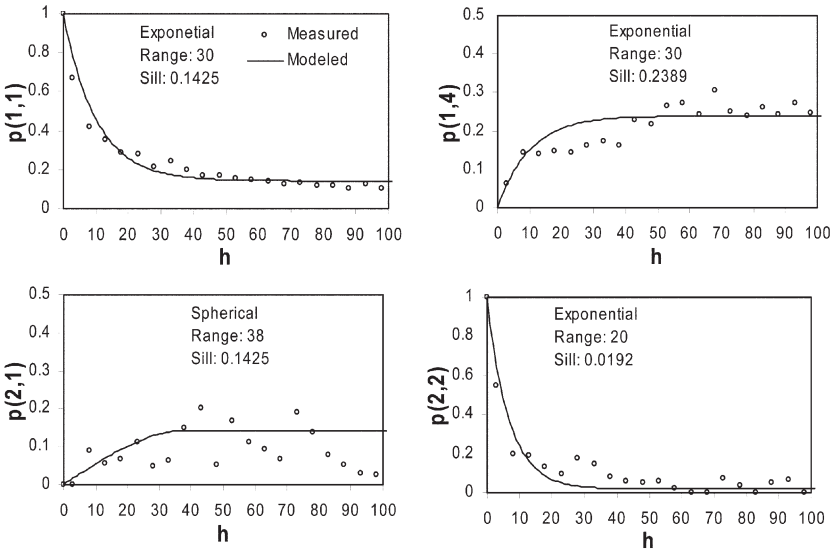

Fig. 11. Four experimental transiograms estimated from the large random data set of soil types along the west-to-east direction with a tolerance width of two pixels and the corresponding mathematical models. Sills were set to the proportions of the corresponding tail classes.

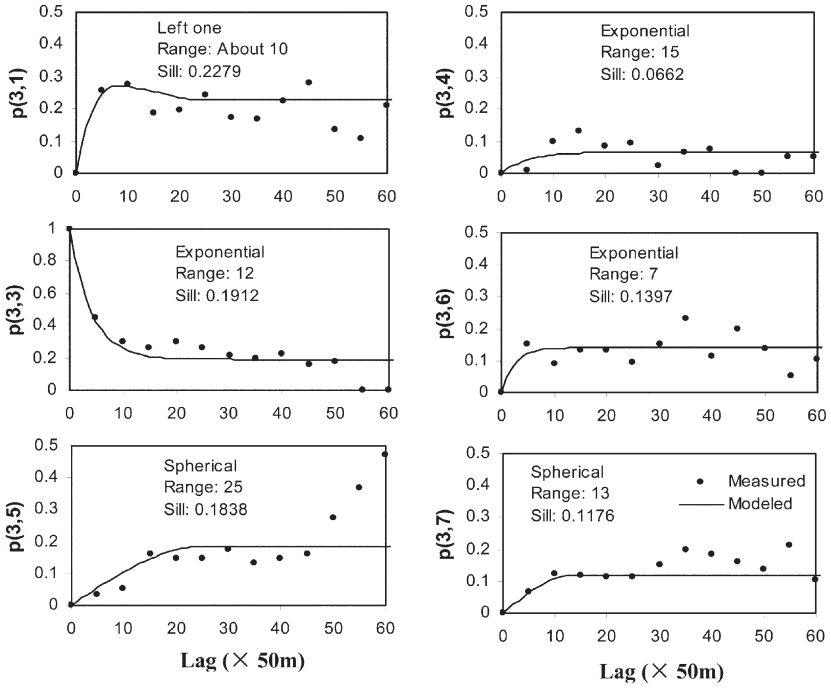

Fig. 12. Experimental transiograms headed by Soil Type 3 and their corresponding models, estimated from the small grid data set (i.e., 136 points). Sills were set to the proportions of the corresponding tail classes in the data set.

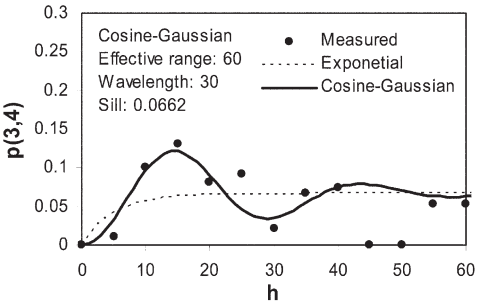

Fig. 13. An example of fitting an experimental cross-transiogram using a cosine–Gaussian

reduced. It should be noticed, however, that with increasing tolerance width, the number of points constituting the experimental transiogram decreases and thus some of the information is lost. Therefore, a trial-and-error approach may be necessary to identify a suitable tolerance width.

Transiogram models used for Markov chain simulation must adhere to the theoretical constraints outlined above. Therefore, each subset of experimental transiograms headed by the same class should be jointly modeled as an entirety so as to meet the summing-to-1 condition. Figure 12 shows some four-directional experimental transiograms headed by Soil Type 3, estimated from the small regular data set (i.e., 136 points) with seven soil types. Exponential models and spherical models were used to fit these experimental transiograms. Sills of the transiogram models were set to the proportions of the corresponding tail classes in the data set, with the ranges and model types discerned from the corresponding experimental transiograms. These basic mathematical models approximately capture the general trends of the experimental data, even for the model calculated as a difference (i.e., $p_{3,1}(\mathbf{h})$, which was calculated using Eq. [27]). This latter transiogram is obviously well characterized. It is recommended that the most complex experimental transiogram corresponding to a specific head class is modeled as the difference indicated in Eq. [27], not only to save time but perhaps to improve the correspondence with experimental cross-transiograms. Because Markov chain simulations generally use only transiogram values near the origin (i.e, small $\mathbf{h}$), the accuracy of the model fit to the first few lag intervals should be maximized. Although a perfect fit is not possible or strictly necessary using mathematical models, such models can provide a general understanding of the soil spatial variation in the study area. For example, in Fig. 12, $p_{3,3}(\mathbf{h})$ shows that the auto-correlation of Soil Type 3 can be described by an exponential model with an auto-correlation range of 12 pixels (i.e., 600 m), and $p_{3,5}(\mathbf{h})$ indicates that Soil Type 3 rarely, if ever, occurs in proximity to Soil Type 5, as evinced by the corresponding long cross-correlation range and spherical curve shape.

It is worth mentioning the differences between modeling transiograms and variograms. Transiograms do not have (and theoretically should not have) a nugget effect for exclusive classes. Conventionally in geostatistics, experimental variograms with the shapes of $p_{3,1}(\mathbf{h})$ and $p_{3,6}(\mathbf{h})$ (see Fig. 12) are fitted with a nugget effect model. This would be improper in transiogram modeling of exclusive classes for Markov chain simulation, however, because it constitutes a violation of the fact that these values represent transition probabilities and as such must satisfy the axiom of summing-to-1. Therefore, the origin (0,0) must be included, which implies that $p_{3,1}(\mathbf{h})$ and $p_{3,6}(\mathbf{h})$ both have short correlation ranges.

Apparently, there are some irregular periodicities in experimental transiograms. For example, $p_{3,4}(\mathbf{h})$ in Fig. 12 shows two peaks and two troughs. A cosine–Gaussian hole-effect model was used to model this pattern, which is shown in Fig. 13. This model provides a reasonable description of the first peak and trough of the experimental transiogram, but the second trough is ignored. Fitting a basic model to such a pattern is quite difficult due to the complexity and irregularity in curve shape, which may be quite diverse for different experimental transiograms; however, accurately fitting the experimental transiograms near the origin would take priority if the model data are to be used for Markov chain simulation or to characterize a complex relationship between classes.

CONCLUSIONS

The transiogram provides a general concept for representing various transition probability diagrams. It not only provides a tool for characterizing spatial heterogeneity of categorical soil variables, but also represents a widely applicable transition probability estimation approach from different data types. For describing interclass relationships, cross-transiograms have special capabilities that symmetric spatial measures do not, i.e, the ability to characterize juxtaposition relationships and directional asymmetry of class distributions.

Although idealized transiograms are relatively simple, they still capture the basic spatial correlation properties of soil classes, including some juxtaposition tendencies. Given the simplicity in deriving idealized transiograms and their availability in some situations (e.g., from borehole data), it is reasonable that these be

used in characterizing subsurface facies and providing transition probability models for geostatistical simulations. Idealized transiograms, however, apparently miss the more complex features of spatial autocorrelations and interclass relationships that are typically present in spatial multinomial classes such as soil types. The more complex features are defined here as the non-Markovian effect. Also, because of the inherent difficulty in estimating one-step transition probabilities from sparse point samples, idealized transiograms have limited uses outside subsurface characterization. Nevertheless, understanding idealized transiograms is important for understanding and interpreting real-data transiograms.

While an idealized transiogram represents only a first-order stationary transition probability model, a real-data transiogram captures considerably more information regarding the spatial structure of variability, as evidenced by the presence of features such as apparent peaks, troughs, and sometimes multiple ranges. These features are most pronounced when computing exhaustive transiograms, as was done here for a large soil type map, which demonstrated the ability to capture the so-called "non-Markovian effect," with some auto-transiograms having two ranges and cross-transiograms generally having irregular peaks and troughs.

Experimental transiograms may be approximately modeled using basic mathematical models, but to fit the complex features of experimental transiograms, more complex models are needed. When transiogram modeling is done for the purpose of Markov chain simulations, experimental transiograms need to be modeled jointly subset by subset so that constraint conditions can be satisfied.

Prospectively, transiograms have the following uses: (i) characterizing spatial variability of categorical soil variables, (ii) providing parameter inputs to Markov chain conditional simulation models, and (iii) data mining of digital soil maps (or databases). Further efforts will focus on developing practical software for estimating transiograms from various data sets and performing automatic model calibration.

ACKNOWLEDGMENTS

The author thanks Dr. C. Zhang for supporting this research, and also thanks Dr. T.R. Ellsworth for helpful comments and revisions, which improved the paper's readability.

REFERENCES

Agterberg, F.P. 1974. Geomathematics. Elsevier, Amsterdam.

Carle, S.F., and G.E. Fogg. 1996. Transition probability-based indicator geostatistics. Math. Geol. 28:453–477.

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

Chilès, J.P., and P. Delfiner. 1999. Geostatistics: Modeling spatial uncertainty. John Wiley & Sons, New York.

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

Haws, N.W., B. Liu, C.W. Boast, P.S.C. Rao, E.J. Kladivko, and D.P. Franzmeier. 2004. Spatial variability and measurement scale of infiltration rate on an agricultural landscape. Soil Sci. Soc. Am. J. 68:1818–1826.

Krumbein, W.C. 1968. FORTRAN IV computer program for simulation of transgression and regression with continuous time Markov models. Comput. Contrib. 26. Kansas Geol. Surv., Lawrence.

Li, W. 2006. Transiogram: A spatial relationship measure for categorical data. Int. J. Geogr. Inf. Sci. 20:693–699.

Li, W. 2007a. A fixed-path Markov chain algorithm for conditional simulation of discrete spatial variables. Math. Geol. (in press), doi:10.1007/s11004-006-9071-7.

Li, W. 2007b. Markov chain random fields for estimation of categorical variables. Math. Geol. doi:10.100S11004-007-9081-0.

Li, W., B. Li, and Y. Shi. 1999. Markov-chain simulation of soil textural profiles. Geoderma 92:37–53.

Li, W., C. Zhang, J.E. Burt, and A.-X. Zhu. 2005. A Markov chain-based probability vector approach for modeling spatial uncertainty of soil classes. Soil Sci. Soc. Am. J. 69:1931–1942.

Ma, Y.Z., and T.A. Jones. 2001. Teacher's aide: Modeling hole-effect variograms of lithology-indicator variables. Math. Geol. 33:631–648.

Norberg, T., L. Rosen, A. Baran, and S. Baran. 2002. On modeling discrete geological structure as Markov random fields. Math. Geol. 34:63–77.

Potter, P.E., and R.F. Blakely. 1967. Generation of a synthetic vertical profile of a fluvial sandstone body. Soc. Pet. Eng. J. 7:243–251.

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

Schwarzacher, W. 1969. The use of Markov chains in the study of sedimentary cycles. Math. Geol. 1:17–39.

Soil Conservation Service. 1962. Soil Survey—Iowa County, Wisconsin. U.S. Gov. Print. Office, Washington, DC.

Weigand, H., K.U. Totsche, B. Huwe, and I. Kogel-Knabner. 2001. PAH mobility in contaminated industrial soils: A Markov chain approach to the spatial variability of soil properties and PAH levels. Geoderma 102:371–389.

Weissmann, G.S., and G.E. Fogg. 1999. Multi-scale alluvial fan heterogeneity modeled with transition probability geostatistics in a sequence stratigraphic framework. J. Hydrol. 226:48–65.

Wu, K., N. Nunan, J.W. Crawford, I.M. Young, and K. Ritz. 2004. An efficient Markov chain model for the simulation of heterogeneous soil structure. Soil Sci. Soc. Am. J. 68:346–351.