Models of spatial transition probabilities, or equivalently, transiogram models have been recently proposed as spatial continuity measures in categorical fields. In this paper, properties of transiogram models are examined analytically, and three important findings are reported. Firstly, connections between the behaviors of auto-transiogram models near the origin and the spatial distribution of the corresponding category are carefully investigated. Secondly, it is demonstrated that for the indicators of excursion sets of Gaussian random fields, most of the commonly used basic mathematical forms of covariogram models are not eligible for transiograms in most cases; an exception is the exponential distance-decay function and models that are constructed from it. Finally, a kernel regression method is proposed for efficient, non-parametric joint modeling of auto- and cross-transiograms, which is particularly useful for situations where the number of categories is large.

Keywords: categorical data, transition probability, geostatistics, spatial continuity

1. Introduction

Categorical spatial data, such as land use and land cover data in geography and environmental science, rock (lithology) types in earth science and socio-economic survey data in social sciences, etc., are all-important information sources across a wide spectrum of scientific fields. As with continuous spatial data, complex spatial patterns (spatial correlation) exist in such geo-referenced categorical data in accordance with Tobler’s first law of geography ([Tobler](1970)). A successful investigation of the statistical characteristics in these patterns will benefit many of the above mentioned scientific fields, particularly with respect to spatial data classification or clustering, spatial uncertainty modeling and spatial scale effects. In remote sensing imagery classification, for example, spatial pattern information implied in thematic classes (e.g., forest area is more likely adjacent to grass-
land than desert area) can be fully integrated with conventional classifiers to enhance the classification performance (Tso and Mather 2001).

One of the most fundamental concepts in spatial analysis is the choice of spatial continuity measures usually quantifying similarity in attribute values or class labels. In conventional geostatistics, indicator kriging (IK) (Solow 1986) and indicator coKriging (ICK) (Deutsch and Journel 1998) are the most frequently used methods for estimating the posterior (conditional) probability of class occurrence at any unsampled location conditioned on the available observed data. Both IK and ICK rely on two-point spatial continuity measures, indicator (cross)covariance or (cross)variogram models, for characterizing spatial association in categorical spatial data. Although covariances and variograms are suitable for continuous fields, particularly Gaussian random fields, the discrete characteristics of categorical data, along with their sharp boundaries and complex spatial patterns, render the interpretation of such covariances and variograms less intuitive. This, in turn, hinders the applications of the kriging family of methods for deriving probabilities of class occurrence in categorical fields.

Recently, promising alternatives to the indicator (cross-)variogram, namely, spatial transition probabilities (Carle and Fogg 1996), or equivalently, transiograms (Li 2006), have been proposed as alternative spatial continuity measures in categorical fields. The concept of transition probability is not new; but it has only recently been proposed as a continuity measure in categorical fields. Compared to indicator covariances and variogram models, transiograms are more interpretable in categorical fields, and easier to integrate with ancillary information (Carle and Fogg 1996). Based on this concept, Carle and Fogg (1996) reformulated IK and ICK as systems of spatial transition probabilities according to their analytical connections with indicator covarioGrams. More recently, Li (2007b) employed a single Markov Chain moving randomly within a stationary random field (Markov Chain Random Field) for conditional simulation or interpolation of categorical spatial data. Class occurrence probabilities derived by such methods usually satisfy the fundamental probability constraints naturally compared to methods based on variations of IK.

As an extension of transition probabilities in a spatial setting, transiograms naturally inherit basic properties of two-point conditional probabilities, such as asymmetry, non-negativity and unit-sum (Carle and Fogg 1996, 1997). The relationships between the parameters of transiogram models and the information on class proportion, mean length, and class juxtaposition have been investigated by Carle and Fogg (1996), and this information actually provides an interpretation of the behavior of transiogram models and eventually offers a guideline for the construction of such models by incorporating expert knowledge of the spatial distribution of the categories under study. Along these lines, one potential contribution of this paper is to investigate the connections between the behavior of auto-transiogram models near the origin and the spatial distribution of the associated category.

As with variograms, not every function of distance can serve as a valid transiogram. Several basic mathematical models of variograms, such as circular, spherical, exponential, Gaussian, and cosine-Gaussian, have been proposed for transiogram modeling (Li and Zhang 2006). No discussion, however, has yet been made on whether these valid variogram functions can be eligible for transiograms under certain circumstances. In this paper, the validity of transiogram models in the stationary indicator random fields, particularly the excursion sets of Gaussian Random Fields (GRFs), is discussed and it is found that in most cases, only the exponential form and several of its variants are eligible for transiogram modeling.
On another front, even if valid transiograms are assumed to be available, or in cases where the assumption of stationary indicator random fields does not apply, e.g., in a Markov Chain Random Field (MCRF) (Li 2007b), transiogram model fitting from empirical values can become tedious as the number of classes increases, since there are $K^2$ auto- and cross-transiogram curves to be jointly modeled for a set of $K$ classes. In practice, one often finds that basic parametric transiogram models (usually defined by a set of parameters including range, sill, anisotropy and etc.) cannot capture the irregular fluctuations at small scales (e.g., hole effect) often found in empirical transiograms. Incorporating this information in transiogram models could dramatically increase the number of (unknown) parameters to be estimated (Li et al. 2011), and renders the quantitative fitting of such models infeasible. To address these computing and modeling issues, this paper proposes a kernel regression-based non-parametric fitting procedure for efficient and consistent transiogram modeling.

The remainder of this paper is organized as follows: the behavior of auto-transiograms near the origin is examined in Section 3 after briefly reviewing the concepts of transiograms in Section 2. Section 4 is devoted to investigating the validity of basic transiogram models in the excursion sets of GRFs, and the kernel regression based non-parametric transiogram model fitting method is presented in Section 5. Finally, section 6 concludes the paper and provides some discussion.

2. Basic concepts of spatial transition probabilities

Consider a $d$-dimensional geographical region $\Omega$ which is partitioned into $N$ disjoint subregions with a categorical random variable (RV) $C(x) (x \in \mathbb{R}^d)$ which can take one out of $K$ mutually exclusive and collectively exhaustive class labels $c(x) \in \{1, \ldots, K\}$ at any arbitrary location with coordinate vector $x$. Alternatively, one can also define an indicator variable $I_k(x)$ to represent $C(x)$, where $I_k(x) = 1$ if $C(x) = k$ and $I_k(x) = 0$ otherwise.

Given two locations $x$ and $x'$ in $\Omega$, and the associated class labels denoted as $k$ and $k'$, the transiogram $\pi_{k'\mid k}(x', x)$ is typically a parametric model of transition probabilities as a function of the lag vector $h = x' - x$. In words, $\pi_{k'\mid k}(x', x)$ is the probability of, starting from a source location $x$ with class label $k$, arriving at a destination location $x'$ with class label $k'$. Note that since the dimension of $\Omega$ is greater than 1, there are theoretically infinite paths to reach a destination location $x'$ from a source location $x$. This equifinality issue hinders the application of the original definition of 1D transition probability and the rich Markov chain theory based on it, such as the celebrated Chapman-Kolmogrov equation, to high dimensional spaces. To eliminate ambiguity, the definition of spatial transition probabilities is restricted to the path defined by the vector $h = x' - x$. Specifically, spatial transition probabilities could be defined as:

$$\pi_{k'\mid k}(h) = P\{C(x') = k'\mid C(x) = k\}$$
$$= P\{I_{k'}(x') = 1\mid I_k(x) = 1\}$$
$$= \frac{P\{I_{k'}(x') = 1 \text{ and } I_k(x) = 1\}}{P\{I_k(x) = 1\}}$$

Second-order or intrinsic stationarity (Chilès and Delfiner 1999) is implicitly assumed in this definition, since the value of $\pi_{k'\mid k}(h)$ depends only on the lag $h$ and not on the
Figure 1. A sample area-class map (top) with three classes and its associated transiograms curves (bottom) are computed by exhaustive sampling (scanning). All two-point dependence information of the area-class map is encapsulated in these (3 × 3) empirical transiogram curves.

location \( x \) or \( x' \). More specifically, \( \pi_{k|k}(h) \) denotes the auto-transiogram for class \( k \) (when \( k = k' \)), a measure of spatial auto-correlation of class \( k \), and \( \pi_{k'|k}(h) \) denotes the cross-transiogram from class \( k \) to class \( k' \) (when \( k \neq k' \)), a measure of spatial cross-correlation between class \( k \) and class \( k' \). Conventionally, class \( k \) and class \( k' \) in \( \pi_{k'|k}(h) \) are called tail class and head class respectively.

Under the assumption of second-order stationarity, sample transiograms can be obtained by direct computation (exhaustive sampling, no parametric model involved) from sample data on a regular grid whose node spacing coincides with the scale of analysis. Thus for a given \( h \), we have:

\[
\hat{\pi}_{k'|k}(h) \simeq \frac{1}{\pi_k} E\{I_k(x)I_{k'}(x + h)\} \simeq \frac{1}{\pi_k N(h)} \sum_{i=1}^{N(h)} [i_k(x)i_{k'}(x + h)]
\]

(2)

where \( N(h) \) denotes the number of location pairs separated by vector \( h \) and \( \pi_k \) indicates the proportion of class \( k \). Figure 1 provides an area-class map with three categories, and for a certain lag distance \( h \), the associated transiograms (3 × 3) values are computed by exhaustively enumerating the pairs of nodes separated by a template vector \( h \) in the whole sample map.

If a (latent) probabilistic distribution is assumed underpinning the (observable) categorical field, e.g., a truncated multivariate Gaussian field, all auto- and cross-transiograms
can be computed exactly according to the threshold values associated with each class and the analytical form of the latent distribution (Chilès and Delfiner 1999).

The basic models of variograms as well as the classical geostatistical concepts of range, sill, hole effect, and anisotropy, have been discussed in the context of transiograms (Li 2006). Given a lag vector $h$, transiogram values (spatial transition probabilities) have the following basic properties:

- **asymmetry**
  
  $$\pi_{k'|k}(h) \neq \pi_{k'|k}(-h) \quad (3)$$

- **non-negativity**
  
  $$\pi_{k'|k}(h) \geq 0, \forall k, k' \quad (4)$$

- **unit-sum**
  
  $$\sum_{k'=1}^{K} \pi_{k'|k}(h) = 1 \quad (5)$$

- **value at zero distance**
  
  $$\pi_{k'|k}(0) = \begin{cases} 1 & \text{if } k = k' \\ 0 & \text{if } k \neq k' \end{cases} \quad (6)$$

With the basic concepts and properties of the transiogram models reviewed in this section, the remainder of this paper will investigate the additional important properties of these models including the properties near the origin, validity of the transiogram models as well as the fitting procedures of these models.

3. Behaviors of Transiogram Models Near the Origin

As in covariograms, the shape (e.g., regularity) of transiograms reflects the spatial continuity and interaction of categories. In Figure 1 for example, the cross-transiogram curves between class 1 (represented as blue in Figure 1) and class 3 (represented as cyan in Figure 1) stay at 0 for a certain distance before they begin to increase gradually. This transiogram behavior is because class 1 is never adjacent to class 3 in the reference map of Figure 1, and transiogram values should increase after the minimum distance between pixels of class 1 and class 3. This connection between transiogram curves and the shape of objects, particularly the first derivative of an auto-transiogram curve at the origin and the shape of objects of the category associated with that auto-transiogram, is examined in detail in this section.

We start with two particular properties of the indicator representation of categorical spatial variables:

$$P\{I_k(x) = 1\} = E\{I_k(x)\} \quad (7)$$
Figure 2. An illustration of non-center auto-transiogram

\[ P\{I_k(x) = 1 \text{ and } I_{k'}(x + h) = 1\} = E\{I_k(x)I_{k'}(x + h)\} \]

where \( E\{I_k(x)I_{k'}(x + h)\} \) is known as the non-centered indicator cross-covariance or probabilistic version of geometric covariogram (Lantuejoul 2002).

Suppose there is a circular region \( A \) with category \( k \) shifted by \( h \) to region \( A_h \) as illustrated in Figure 2. In this case, the area of intersection \( A \cap A_h \) (blue domain) represents \( E\{I_k(x)I_{k'}(x + h)\} \), and the area of \( A \) (blue domain and red domain) represents \( P\{I_k(x) = 1\} \) if the area of the whole region is assumed to be one. Thus according to the definition of the transiogram (Equation 1), \( \pi_{k|k}(h) \) can be seen as the area of the blue domain divided by the area of the whole circle or the proportion of the blue domain in the white or red domain. If the class proportions \( E\{I_k(x)\} \) are assumed to be constant, one arrives to the analytical links between the transiogram and the indicator (cross-)covariogram and indicator (cross-)variogram by applying the properties of indicators (Equation 7 and Equation 8) to the definition of the (cross-)covariogram (Equation 9) and (cross-)variogram (Equation 10) respectively (Carle and Fogg 1996):

\[
\sigma_{kk'}(h) = \pi_k[\pi_{k'|k}(h) - \pi_{k'}] \tag{9}
\]

\[
\gamma_{kk'}(h) = \pi_k\{\pi_{k'|k}(0) - [\pi_{k'|k}(h) + \pi_{k'|k}(-h)]/2\} \tag{10}
\]

where \( \sigma_{kk'}(h) \) represents indicator (cross-)covariogram and \( \gamma_{kk'}(h) \) represents indicator (cross-)variogram and \( \pi_{k'|k}(-h) \) represents the transition probability in the opposite direction of \( h \). Particularly if we let \( k = k' \) and \( \pi_{k|k}(0) = 1 \), we have a simple linear connection between the auto-transiogram and the indicator auto-variogram:

\[
\pi_{k|k}(h) = 1 - \frac{\gamma_{kk}(h)}{\pi_k} \tag{11}
\]

Because of the linear connection between the transiograms and the indicator (cross-)variogram/covariogram (Equation 9 and Equation 11), transiograms share the
properties of indicator (cross-)variogram/covariograms. As Figure 1 illustrates, auto-
transiograms (diagrams on the diagonal in Figure 1) start from 1 at $h = 0$ and gradu-
ally decrease to the sill value at $h$ equals range, i.e., $\lim_{h \to \infty} \pi_{k|k}(h) = \pi_k$. Cross-
transiograms (diagrams off the diagonal in Figure 1) start from 0 at $h = 0$ and gradually 
increase to the sill value at $h$ equals range, i.e., $\lim_{h \to \infty} \pi'_{k|k}(h) = \pi_{k'}$.

The compactness of a geographic shape is an important property of a polygon in 
GIS/zoning and landscape metrics. One of the simplest compactness measures or indices 
of a shape is the ratio of its perimeter to its area (perimeter-to-area ratio) $\Psi$ (Smith 
et al. 2007). It is well known in the literature that the first derivative of the variogram at 
the origin is related to the derivative or gradient of the surface it represents (Stein 1999,
Chiles and Delfiner 1999). Carle and Fogg (1996) have shown that in a 1D continuous-
space Markov chain, the first derivative of the auto-transiogram at the origin, $\pi'_{k|k}(0; \phi)$, 
termed transition rate, is related to the mean length or mean thickness of the objects of 
category $k$ in direction $\phi$. The empirical transition rate is usually calculated by the total 
length of category $k$ in direction $\phi$ divided by the number of embedded occurrences of 
$k$. In this paper, the relationship between the auto-transiogram for a certain class label 
and the perimeter-to-area ratio of shapes with such a class label in a 2D geographical 
space is given via the following proposition.

**Proposition 3.1:** Under a stationary proportions assumption, the perimeter-to-area 
ratio $\Psi_k$ of the objects of category $k$ in a 2D random sets with unit area can be obtained 
by integrating the derivative $\pi_{k|k}(0)$ in the direction $\phi$ at the origin over all possible 
directions:

$$\Psi_k = -\frac{1}{2} \int_0^{2\pi} \pi'_{k|k}(0; \phi) d\phi$$

(12)

**Proof:** In a 2D space, the perimeter $l_k$ of objects of category $k$ can be obtained by the 
application of Minkowski’s formula (Matheron 1971, Lantuejoul 2002):

$$l_k = -\frac{1}{2} \int_0^{2\pi} K'_\phi(0) d\phi$$

where $K(\cdot)$ is the probabilistic version of geometric covariogram for category $k$ (Lantue-
jour 2002). Replacing the covariogram with the transiogram according to Equation 9, we 
have:

$$l_k = -\frac{\pi_k}{2} \int_0^{2\pi} \pi'_{k|k}(0; \phi) d\phi$$

Under the stationary proportions assumption, $\pi_k$ is proportional to the area of objects 
of category $k$, and without loss of generality, one lets the total area equal one and the 
area of objects of category $k$ is thus $\pi_k$. Equation 12 is obtained per the definition of the 
perimeter-to-area ratio.

For the isotropic case, $\Psi_k$ can then be simply written as:

$$\Psi_k = -\pi \times \pi'_{k|k}(0)$$

(13)
If $\pi'_{k|k}(0) = \infty$ in particular, it is anticipated that the boundaries of category $k$ demonstrate fractal properties. A degenerate case is when $\pi_{k|k}(h)$ is parabolic with $\pi'_{k|k}(0) = 0$.

It is worth noting that, in general, $\Psi_k$ is scale dependent and it is a particular average shape descriptor (of union) of objects of category $k$ instead of a single object; in other words, $\Psi_k$ corresponds to the mean perimeter-to-area ratio in landscape metrics [McGarigal and Marks 1995]. The conclusion of the Proposition 3.1 is for the closed objects. If the union object is open, the boundary of the studied area will be taken into account. This note applies to both convex and concave shapes.

Proposition 3.1 can be verified by the circular region (the radius of this circle $R$ is assumed to be 0.25) in a map with unit area (Figure 2). The perimeter-to-area ratio $\Psi_A = \frac{2}{R} = 8$ and the auto-transiogram for category $k$ can be written as (the area of blue domain divided by the area of the circle):

$$\pi_{k|k}(h) = \frac{2}{\pi} \left\{ \arccos \left( \frac{|h|}{2R} \right) - \frac{|h|}{2R} \sqrt{1 - \left( \frac{|h|}{2R} \right)^2} \right\} \text{ if } |h| \leq R$$

Taking the first derivative of Equation 14 with respect to $h$ and letting $h = 0$, one gets $\pi'_{k|k}(0) = -\frac{2}{\pi R} = -\frac{8}{\pi}$ and Equation 13 is obviously satisfied.

Proposition 3.1 can further be illustrated in Figure 3. The red curve in Figure 3 (a) represents the empirical auto-transiogram of class 3 (represented as yellow) in the sample map Figure 3 (b), and green curve represents that of the sample map Figure 3 (c). The areas of these two sample maps are exactly the same and the proportions of class 3 (yellow regions) are both about 0.40. By comparing Figure 3 (b) and (c), one can easily check that the average perimeter of yellow regions in (c) is much larger than that of (b). Thus according to the proposition 3.1, the auto-transiogram of class 3 (the red curve) in Figure 3 (b) should have a larger first derivative (slope) at the origin than that (the green curve) of (c), which is evident by the two curves in Figure 3 (a).

In this section, the analytical connection between the auto-transiogram of a certain category, particularly its first derivative at the origin, and the shape metrics of the objects with this category in a categorical field is carefully investigated. It provides a quantitative interpretation of the behaviors of transiogram models. More importantly, together with other properties of transiograms [Carle and Fogg 1997, Li 2006], it provides analytical instructions for incorporating domain experts knowledge in transiogram-based applications.

4. Valid Transiogram Models

As in Kriging systems, one often needs to fit empirical transiogram values to certain parametric transiogram models in the transiogram-based methods [Carle and Fogg 1996, Li 2007b]. Several theoretical models of variograms, such as triangular, circular, spherical, exponential and Gaussian, have been proposed as parametric models of transiograms [Li 2006, 2007a] without checking their validity under certain circumstances. In this section, this validity is investigated in the stationary indicator random fields, particularly for indicators of excursion sets of GRFs, which is commonly used in geostatistics and spatial uncertainty modeling.

Let $Z = \{Z(x)\}$ be a stationary GRF with a correlogram $\rho(h)$. Oftentimes, the indicators at location $x$ $i_k(x)$ can be obtained by truncating $Z(x)$ by a specification of a
Figure 3. An illustration of connections between the first derivative of auto-transiograms at the origin and the shape of objects. (a) Empirical auto-transiograms of class 3 in sample area-class maps (b) and (c) (b) a sample area-class map with three classes, and the empirical auto-transiogram of class 3 (represented as yellow) is the red curve in (a) (c) a sample area-class map with three classes, and the empirical auto-transiogram of class 3 (represented as yellow) is the green curve in (a).

For indicator variables in particular, one has:

\[ |I_k(x + h + h') - I_k(x)| \leq |I_k(x + h + h') - I_k(x + h)| + |I_k(x + h) - I_k(x)| \]

For indicator variables in particular, one has:

\[ |I_k(x + h + h') - I_k(x)| \leq |I_k(x + h + h') - I_k(x + h)| + |I_k(x + h) - I_k(x)| \]

Moreover, \( |I_k(x + h) - I_k(x)| = |I_k(x + h) - I_k(x)|^2 \), which leads to a necessary condition for a valid indicator variogram:

\[ \gamma_{kk}(h + h') \leq \gamma_{kk}(h) + \gamma_{kk}(h') \]

(15)
A necessary condition for a valid auto-transiogram can thus be obtained by its analytical connections with an indicator auto-variogram (Equation 11):

$$\pi_{k|k}(h + h') \geq \pi_{k|k}(h) + \pi_{k|k}(h') - 1$$  \hspace{1cm} (16)

According to the triangular inequality of auto-transiogram (Equation 16), the Gaussian form (Equation 19) cannot be a valid auto-transiogram, since if $h = h'$, by the Equation 15 and Taylor expansion, we have $4|h|^2 \leq 2|h|^2$ when $|h| \to 0$, which is obviously a contradiction.

Matheron (1993) provided a more general necessary condition (containing the triangular inequality) for eligible indicator variograms $\gamma_{kk}(x, x')$: for any set of $m$ ($m \geq 2$) points $x_1, \ldots, x_m$ with category $k$, and values $\epsilon_i \in \{-1; 0; 1\}, i = 1, \ldots, m$, such that $\sum_{i=1}^{m} \epsilon_i = 1$, the associated variogram values $\gamma_{kk}(x_i, x_j)$ must satisfy:

$$\sum_{i=1}^{m} \sum_{j=1}^{m} \epsilon_i \epsilon_j \gamma_{kk}(x_i, x_j) \leq 0$$  \hspace{1cm} (17)

Or equivalently, in terms of auto-transiogram values, one has :

$$\sum_{i=1}^{m} \sum_{j=1}^{m} \epsilon_i \epsilon_j (1 - \pi_{k|k}(x_i, x_j)) \leq 0$$  \hspace{1cm} (18)

It is still an open question whether the necessary condition is also sufficient for eligible indicator variograms of general random sets. Emery (2010) recently pursued this question further by suggesting that the properties of triangular, circular, and spherical variograms are rather restrictive in two or three dimensional indicator random fields, and proved that these three variograms are not valid indicator variograms for excursion sets of stationary GRFs. Due to the linear connection between indicator variograms and auto-transiograms, we can hence conclude that the Gaussian (Equation 19), triangular (Equation 20), spherical (Equation 21) and circular (Equation 22) models cannot be valid basic forms of auto-transiograms for indicators of excursion sets of stationary GRFs.

- **Gaussian**

  $$\pi_{k|k}(h) = 1 - (1 - \pi_k) \{1 - \exp(-\frac{|h|}{a})\}$$ \hspace{1cm} (19)

- **Triangular**

  $$\pi_{k|k}(h) = \begin{cases} 
  1 - (1 - \pi_k) \frac{|h|}{a} & \text{if } \frac{|h|}{a} \leq 1 \\
  \pi_k & \text{o.w.} 
  \end{cases}$$ \hspace{1cm} (20)

- **Spherical**

  $$\pi_{k|k}(h) = \begin{cases} 
  1 - (1 - \pi_k) \{1.5 \frac{|h|}{a} - 0.5 (\frac{|h|}{a})^3\} & \text{if } \frac{|h|}{a} \leq 1 \\
  \pi_k & \text{o.w.} 
  \end{cases}$$ \hspace{1cm} (21)
where $a$ represents the range parameter of the model.

To illustrate this conclusion, particularly for spherical cases, we assume that a spherical form (Equation 21) with $a = 1$ is used to model the auto-transiogram of indicator variables $I_k(x)$ truncated by Equation 14 from a stationary GRF with unit variance and a correlogram $\rho(h)$. The class proportion $\pi_k$ can be written as $1 - \Phi(z_k)$ with $\Phi(\cdot)$ indicates the cdf (cumulative distribution function) of Gaussian distribution. This auto-transiogram is equivalent to a spherical auto-variogram with sill $\Phi(z_k)(1 - \Phi(z_k))$ and range $a = 1$. From another perspective, the auto-variogram of the indicator $I(x)$ can be obtained by a function of $\rho(h)$ (Chiles and Delfiner 1999):

$$
\gamma_{kk}(h) = \frac{1}{2\pi} \int_0^1 \exp\left(-\frac{z^2}{1 + u}\right) \frac{du}{\sqrt{1 - u^2}}
$$

(23)

One can thus have the corresponding correlation function $\rho(h)$ of the GRF by inverting Equation 23. Given certain number of locations with coordinates, $\rho(h)$ could lead to a singular covariance matrix (see proof of Proposition 14 in Emery 2010), which shows that the previous assumption of auto-transiogram is not true and thus verifies the conclusion that a spherical form can not be used to model the auto-transiogram of excursion sets of GRFs.

Fortunately, the exponential variogram and its derived models are valid indicator variograms in any Euclidean space (Emery 2010). Specifically, the exponential variogram could be written as $\gamma_{kk}(h) = \pi_k(1 - \pi_k)(1 - \exp(-\frac{|h|}{a}))$, where $\pi_k(1 - \pi_k)$ is the variance of indicator variable $I_k(x)$. According to Equation 11, an eligible auto-transiogram thus can be given in Equation 24. This conclusion is not surprising considering transition probabilities are written as an exponential function of transition rates in continuous-time or in 1D continuous-space Markov chain (Carle and Fogg 1997). The memoryless property of the exponential variogram model, which in a spatial context states that the value of a geo-referenced variable $Z(x)$ depends only on its local neighbors, provides a foundation to simplify computations by reducing the global spatial interactions to local.

**Exponential auto-transiogram**

$$
\pi_k|k(h) = 1 - (1 - \pi_k)\{1 - \exp(-\frac{|h|}{a})\}
$$

(24)

According to the connections between the indicator covariance and spatial transition probabilities (Equation 7), Carle and Fogg 1996 reformulated the indicator Kriging system in terms of transition probabilities to take advantage of the transiogram properties. An immediate consequence of the discussion in this section is that the exponential form (Equation 24) is recommended over the triangular (Equation 20), circular (Equation 22), Gaussian (Equation 19) and spherical (Equation 21) forms for modeling the transiograms in the cases of excursion sets of GRFs and eventually building the transition probability matrix of the reformulated system.
5. Non-parametric Transiogram Modeling

Based on the concepts of transiograms, several approaches have been proposed for categorical spatial data modeling from different perspectives. Li (2007b) proposed a Markov Chain random field (MCRF) by applying a spatial Markov Chain in a 2D geographical space. From a maximum entropy point of view, Allard et al. (2009) introduces a similar concept, namely the bi-probagram, which is basically a bivariate joint probability function of lag distances. Cao et al. (2011), on the other hand, proposed a redundancy model in categorical fields to relax the strict conditional independence assumption commonly imposed in transiogram-based methods. Different from stationary indicator random fields, such as mosaic random fields, the Boolean random sets and the excursion set of GRFs discussed in the previous section, transiograms in these models must only meet basic probability constraints (Equation 4 to Equation 5) and not Matheron’s conditions (Equation 17). Although this results in more options for valid transiograms, the joint fitting of transiograms becomes tedious as the number of classes increases. From another perspective, the shapes of valid transiograms are usually controlled by a set of parameters including range, sill and anisotropy values. These basic shapes, however, tend to be over-smoothed and ignore the small scale effects found in empirical transiogram values.

To account for such effects, new shape parameters and primitives (e.g., trigonometric functions for periodic effects) are usually imposed on the basic transiogram models, and oftentimes, this results in dramatic increases in the complexity of the transiogram models (Li et al. 2011) and makes parameter fitting computationally infeasible. In what follows, a Nadaraya-Watson kernel smoothing regression (Nadaraya 1964) based method is proposed for non-parametric transiogram modeling to address these problems.

5.1. Kernel Regression for Transiogram Modeling

Suppose for $\pi_{k'|k}(h)$, the transiogram from class $k$ to class $k'$ at a certain direction, we have empirical transiogram values $p_{k'|k}(h_1), \ldots, p_{k'|k}(h_N)$ for lag $h_1, \ldots, h_N$ respectively. To compute $\hat{\pi}_{k'|k}(h^*)$, the transiogram from class $k$ to class $k'$ for an arbitrary lag $h^*$ in the same direction, one first finds the range $[h_n, h_{n+1}]$ in which $h^*$ lies, and then $\hat{\pi}_{k'|k}(h^*)$ can be written as a linear interpolation of empirical transiogram values (Li and Zhang 2010):

$$\hat{\pi}_{k'|k}(h^*) = \frac{p_{k'|k}(h_n)|h_{n+1} - h^*| + p_{k'|k}(h_{n+1})|h^* - h_n|}{|h_{n+1} - h_n|}$$

(25)

It has been shown that Equation 25 meets the required probability constraints (Equation 4 to Equation 5), but the formulation is rather limited and only linear effects in transiogram values are taken into account, which is apparently an unrealistic assumption for real cases.

Using principles of kernel density estimation (Parzen window estimation) (Silverman 1986), the Nadaraya-Watson kernel smoothing regression method (Nadaraya 1964) has been proposed for non-linear regression, as non-linear effects can be modeled by carefully selected kernel functions. By applying the Nadaraya-Watson kernel regression, the
resulting transiogram model value $\hat{\pi}_{k'|k}(h^*)$, can be given by:

$$
\pi_{k'|k}(h^*) = \begin{cases} 
\frac{\sum_{n=1}^{N} \kappa(|h_n - h^*|) p_{k'|k}(h_n)}{\sum_{n=1}^{N} \kappa(|h_n - h^*|)} & \text{if } |h^*| \neq 0 \\
1 & \text{if } |h^*| = 0 \text{ and } k = k' \\
0 & \text{if } |h^*| = 0 \text{ and } k \neq k'
\end{cases}
$$

(26)

where $\kappa(\cdot)$ is a kernel function with bandwidth $r$. Note that $\hat{\pi}_{k'|k}(h^*)$ is not continuous at the origin, and this discontinuity can be regarded as nugget effect usually caused by noise and measurement errors.

As a kernel function, $\kappa(\cdot)$ should satisfy the following requirements:

- **non-negative**

$$
\kappa(t) \geq 0, \ \forall t \in \mathbb{R}
$$

- **unit-integral**

$$
\int_{-\infty}^{+\infty} \kappa(t) dt = 1
$$

- **symmetry**

$$
\kappa(t) = \kappa(-t)
$$

The properties of commonly-used kernel functions (Table 1), such as Gaussian, Epanechnikov, Biweight, Triangular, have been studied extensively (Silverman 1986). Usually the Epanechnikov kernel tends to generate the smallest square errors if the smoothing parameter $r$ is chosen correctly. The recently proposed linear interpolation method for transiogram modeling (Li and Zhang 2010) can be regarded as a special case of Equation (26) if a triangular kernel is selected and for each $h^*$, only the two nearest neighbors are chosen. The Gaussian kernel, one of the most commonly-used kernel functions, generates the smoothest curves and thus tends to create smooth category boundaries in the output map. Higher values of $r$ (bandwidth of kernel functions), lead to smoother results, and numerous ways have been proposed to obtain the optimal $r$, including least-squares cross-validation, likelihood cross-validation, reference to a standard distribution and subjective choices (Silverman 1986).

A proof is given to show that Equation (26) always yields valid transiogram values.

**Proof:** Assume that empirical transiogram values are obtained by exhaustive sampling of all observed data, thus given a lab $h$, these empirical transiograms should meet basic transiogram constraints, i.e., for a given $h_n$, we have $p_{k'|k}(h_n) \geq 0$ and $\sum_{k=1}^{K} p_{k'|k}(h_n) = 1$. 

| Kernel      | $\kappa(t)$ where $t = \frac{\Delta h}{r}$ |
|-------------|------------------------------------------|
| Epanechnikov| $\frac{3}{4}(1 - t^2)1_{|t| \leq 1}$     |
| Gaussian    | $\frac{1}{\sqrt{2\pi}} \exp\{-t^2\}$    |
| Biweight    | $\frac{15}{16}(1 - t^2)^21_{|t| \leq 1}$ |
| Triangular  | $(1 - |t|)1_{|t| \leq 1}$                  |

Table 1. Several typical kernel functions for kernel regression
According to the definition of the transiogram (Equation 1) and calculation of empirical transiogram values \( p_{k'|k}(h_n) \) (Equation 2), the values of the transiograms at the origin (Equation 6) is satisfied naturally, i.e., \( p_{k'|k}(0) = 1 \) if \( k = k' \) and \( p_{k'|k}(0) = 0 \) if \( k \neq k' \).

Since we have \( \kappa(\cdot) \geq 0 \) for a valid kernel function, the non-negativity of transiograms is obviously satisfied.

There is only one unknown parameter \( r \), for the head class \( k \) and a given \( h^* \), \( \kappa(|h_n - h^*|) \) will be the same value in all auto- and cross-transiograms. Thus:

\[
\sum_{k=1}^{K} p_{k'|k}(h^*) = \sum_{k=1}^{K} \sum_{n=1}^{N} \kappa(|h_n - h^*|) p_{k'|k}(h_n) = \sum_{n=1}^{N} \kappa(|h_n - h^*|) \frac{\sum_{k=1}^{K} p_{k'|k}(h_n)}{\sum_{n=1}^{N} \kappa(|h_n - h^*|)} = 1
\]

Thus the unit-sum property of transiograms is satisfied. \( \square \)

As an example with two categories, Figure 4 gives a comparison between the proposed kernel regression method and the linear interpolation method (Li and Zhang 2010). Both methods yield valid transiogram values. Not surprisingly, the linear interpolation method (green solid lines) only captures the linear effects in transiogram values, and the proposed method (solid red lines) tends to generate much smoother results by accounting for the non-linear effects in transiogram values via kernel functions, and the shapes of the outcome transiograms of the proposed method can be flexibly adjusted through the parameters of the kernel functions. In contrast to the linear interpolation method that only works within the range of empirical values, the proposed method can be used for extrapolation for distances beyond the empirical range. In addition, the proposed method is not a exact estimator, which means its output estimated transiogram values do not always reproduce the empirical transiogram values (blue circles) as those of the linear interpolation method do. The discrepancy between the estimated and empirical transiogram values is controlled by the kernel bandwidth. Particularly, this discrepancy at the origin (when \( |h^*| = 0 \)) can actually be taken as the nugget effect that the linear interpolation method ignores. To render this proposed non-parametric method operational, the described procedure has been implemented in Matlab (ksr2.m), and integrated with a toolbox for statistical analysis of categorical spatial data, which is available at: [http://www.geog.ucsb.edu/~cao/research.html](http://www.geog.ucsb.edu/~cao/research.html)

6. Conclusions

A collection of statistical methods have been recently proposed for modeling categorical spatial data based on the concept of spatial transition probabilities. Limited discussions, however, have given to the properties of this fairly new spatial continuity measure in the existing literature. In this paper, three findings on basic properties of transiogram models are reported. Specifically, analytical connections between the shape of auto-transiograms near the origin and the spatial distribution of the associated class label was firstly revealed. Similar to variograms, it is not every function that can be used as a valid transiogram model. In the context of stationary indicator random fields, the eligibility of commonly used basic forms of variograms as transiograms was investigated particularly for the excursion sets of Gaussian random fields, one of the most commonly used random sets. It was concluded that the auto-transiogram of indicators in such a random set can not be Gaussian, Spherical, Circular or Triangular forms, which are usually used for variogram modeling. The exponential and its derived forms are recommended for transiogram modeling in the methods based on stationary indicator random fields. Finally, a
non-parametric transiogram fitting procedure was proposed for the cases where the assumption of the stationary indicator random fields does not apply, to capture the small scale effects in transiograms and to address automatic joint fitting issues of transiograms as the number of classes increases. Compared to the recent joint fitting methods based on linear interpolation, the proposed kernel regression-based method is more generic and flexible, and capture the non-linear effects in empirical transiogram values naturally. A Matlab implementation of the proposed joint fitting procedure of empirical transiogram values is also provided. These three findings cover the properties, validity and modeling of transiograms, and provide a better understanding of the behaviors of transiogram models, as well as the transiogram-based methods, and thus to avoid the potential misuse and misunderstanding of this fairly new spatial continuity measure in geographical spaces.

7. Acknowledgments

We gratefully acknowledge the funding provided by the National Geospatial-Intelligence Agency (NGA) to support this research.

References

Allard, D., D’Or, D., and Froidevaux, R., Estimating and Simulating Spatial Categorical Data Using an Efficient Maximum Entropy Approach, 2009. Technical report No.37, Unité Biostatistique et Processus Spatiaux Institut National de la Recherche Agronomique, Avignon, France.

Cao, G., Kyriakidis, P.C., and Goodchild, M.F., 2011. Combining spatial transition probabilities for stochastic simulation of categorical fields. International Journal of Geographical Information Science, 25 (11), 1773–1791.

Carle, S. and Fogg, G., 1996. Transition probability-based indicator geostatistics. Mathematical Geology, 28 (4), 453–476.
Carle, S. and Fogg, G., 1997. Modeling spatial variability with one and multidimensional continuous-lag Markov chains. Mathematical Geology, 29 (7), 891–918.

Chilès, J. and Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty. Wiley-Interscience.

Deutsch, C. and Journel, A., 1998. GSLIB: Geostatistical Software Library and User’s Guide. New York, 369.

Emery, X., 2010. On the existence of mosaic and indicator random fields with spherical, circular, and triangular variograms. Mathematical Geosciences, 42 (8), 969–984.

Lantuèjoul, C., 2002. Geostatistical Simulation: Models and Algorithms. Springer Verlag.

Li, W., 2007a. Transiograms for characterizing spatial variability of soil classes. Soil Science Society of America Journal, 71 (3), 881.

Li, W. and Zhang, C., 2010. Linear interpolation and joint model fitting of experimental transiograms for Markov chain simulation of categorical spatial variables. International Journal of Geographical Information Science, 24 (6), 821–839.

Li, W., 2006. Transiogram: A spatial relationship measure for categorical data. International Journal of Geographical Information Science, 20 (6), 693–699.

Li, W., 2007b. Markov chain random fields for estimation of categorical variables. Mathematical Geology, 39 (3), 321–335.

Li, W. and Zhang, C., 2006. A generalized Markov chain approach for conditional simulation of categorical variables from grid samples. Transactions in GIS, 10 (4), 651–669.

Li, W., Zhang, C., and Dey, D.K., 2011. Modeling experimental cross-transiograms of neighboring landscape categories with the gamma distribution. International Journal of Geographical Information Science.

Matheron, G., 1971. The Theory of Regionalized Variables and Its Applications. Vol. 5. École national supérieure des mines.

Matheron, G., 1993. Une conjecture sur la covariance dun ensemble aléatoire. Cahiers de géostatistique, 107, 107–113.

McGarigal, K. and Marks, B., FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. , 1995., Technical report PNW-GTR-351, U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR.

Nadaraya, E., 1964. On estimating regression. Teoriya Veroyatnostei i ee Primeneniya, 9 (1), 157–159.

Silverman, B., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, CRC.

Smith, M.D., Goodchild, M., and Longley, P., 2007. Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools. 2nd Troubador Publishing.

Solow, A.R., 1986. Mapping by simple indicator kriging. Mathematical Geology, 18 (3), 335–352.

Stein, M., 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer Verlag.

Tobler, W., 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography, 46 (2), 234–240.

Tso, B. and Mather, P., 2001. Classification methods for remotely sensed data. Taylor & Francis.