

A framework of experimental transiogram modelling for Markov chain geostatistical simulation of landscape categories

Jia Yu^{a,b,*}, Weidong Li^{b,c}, Chuanrong Zhang^{b,c}

^a Department of Geography, Shanghai Normal University, Shanghai 200234, China

^b Department of Geography, University of Connecticut, Storrs, CT, USA

^c Center for Environmental Science and Engineering, University of Connecticut, Storrs, CT, USA



ARTICLE INFO

Keywords:

Transiogram
Landscape
Markov chain simulation
Fitting model
Spatial categorical variable

ABSTRACT

Experimental transiogram modelling, especially joint modelling, is crucial for Markov chain random field (MCRF) simulation of categorical spatial variables with multiple classes. Experimental transiogram estimation of landscape categorical variables in a non-visual context and manual model fitting are tedious and time-consuming tasks if the number of classes is large. This study presented a framework which integrated the linear interpolation method and the mathematical model fitting method for transiogram joint modelling to facilitate the whole modelling procedure. The framework was developed as a tool, entitled TGRAM. The tool provides several advanced characteristics and functions which greatly improve the efficiency of transiogram joint modelling. Two case studies were provided to demonstrate the feasibility of the tool. The modelling results can directly provide parameters (i.e., transition probability values at continuous lags in the required pixel size) for further MCRF simulations in related research areas, such as land cover/land use classification, soil and lithofacies mapping, and urban growth detection.

1. Introduction

Spatial analysis and mapping of environmental categorical variables requires spatial modelling of these variables (Truong, Heuvelink, & Gosling, 2013). The spatial variability of categorical spatial variables, which is the typical characteristic of landscape spatial heterogeneity, can represent different kinds of spatial features (e.g. land use/land cover, soil type, lithofacies, sedimentary sequence). To qualify the spatial relationships of different classes of categorical variables, various models have been proposed, such as the transition probability matrix (TPM) (de Almeida et al., 2003; Vistelius, 1949), the indicator variogram (Journel & Huijbregts, 1978; Webster & Oliver, 2007) and the transiogram (Li, 2007a, 2007b). The TPM has been employed to describe the temporal and spatial variations of geographical data for a long time because one-dimensional (1-D) Markov chain models have been earlier used in various fields of research (Burgess & Webster, 1984; Logsdon, Bell, & Westerlund, 1996; Weigand, Totsche, Huwe, & Kögel-Knabner, 2001). TPMs can represent the complex relationships among categories (i.e., classes), but they cannot represent the complex spatial relationships among categories across a number of spatial steps. Currently, the indicator variogram also has been widely used to describe the auto-correlations in single classes of discrete geographic data in the

geosciences (Chiles & Delfiner, 2012). The extensive application of variograms is due to the wide acceptance of kriging-based geostatistics as simulation techniques for spatial variables. But as to the cross-correlations among categories, indicator variograms have the limitation to be utilized because the physical meanings of them are not easy to be interpreted. Meanwhile, because of the symmetric characteristic of cross-variograms, they are not able to interpret the asymmetry in the spatial distribution of categorical variables.

Inspired by the variogram theory and pioneer studies (Carle & Fogg, 1996; Carle & Fogg, 1997; Luo, 1993; Luo, 1996; Ritzi, 2000; Schwarzacher, 1969), the transiogram was proposed to estimate multiple-step transition probabilities and transition probability functions over the lag distance from sample point data for Markov chain random field (MCRF) simulation (Li, 2007a, 2007b). It is a good two-point spatial measure for describing the spatial intra-class autocorrelations and interclass cross-correlations of landscape categories (Li & Zhang, 2013). The advantage of utilizing transiograms lies in two aspects. The first is that the physical meanings of transiograms are easy to be interpreted. They can measure and quantify the spatial relationships among categories (classes) across a number of spatial steps, which means transiograms are not limited by the first-order stationary Markovian assumption on data. Secondly, the transiograms are the

* Corresponding author at: Shanghai Normal University, 100 Guilin Road, Shanghai 200234, China.

E-mail address: yujia@shnu.edu.cn (J. Yu).

requirements for providing essential parameters for multi-dimensional Markov chain spatial simulation. They serve as two-point spatial measures based on sample data to support MCRF models (Li et al., 2015). They are resolution-free and can be used to measure spatial or temporal auto-correlations and cross-correlations over different lag distances.

There are three kinds of transiograms identified in Li (2007a): idealized transiograms, exhaustive transiograms and experimental transiograms. Idealized transiograms refer to the transiograms that are estimated using single-step transition probabilities on the basis of the first-order Markovian assumption and they can be calculated from a one-step or multi-step TPM (Li, Zhang, & Dey, 2012). Exhaustive transiograms are extracted from images or maps of land use/land cover, soil types, and lithofacies, among others. Experimental transiograms are measured directly from sample data in a study area. In these three types of transiograms, idealized transiograms and exhaustive transiograms are useful for understanding experimental transiograms, providing expert knowledge for estimating transiogram models (Li, 2007a). But the experimental transiograms are more frequently utilized because they can be directly used to infer transiogram models for calculating the local probability distributions of categorical variables at unsampled locations in MCRF simulation. The experimental transiograms have been successfully applied in soil mapping (Grunwald, 2009; Zhang & Li, 2008), spatial distribution simulation of clay layer depth (Li & Zhang, 2010a, 2010b), land cover post-classification (Zhang, Li, & Zhang, 2016; Zhang, Li, Zhang, & Li, 2017), and vertical urban growth detection (Zhang, Li, Zhang, & Ouimet, 2017), among others.

The practical application of experimental transiograms benefits from their unique merits as quantitative graphic measures (Li, 2007a). However, as experimental transiograms provide both auto-correlation and cross-correlation information of multiple classes, the transition probability diagrams of them can be very complex in shape. Because of the complexity of diagram shapes of experimental transiograms, it may not be satisfactory to use the basic mathematical models (e.g. the linear model, spherical model, exponential model and Gaussian model) developed for variograms to fit the shapes of experimental transiograms (Ritzi, 2000). Therefore, some more complex mathematical models suggested for variogram modelling (Jones & Ma, 2001) were also recommended for transiograms (Li, 2007a), such as the cosine-exponential and cosine-Gaussian models for cross-transiograms. The shapes of some experimental cross-transiograms may have a peak firstly and then flatten out with increasing lag distance. To fit this type of shapes, gamma-exponential, gamma-spherical and gamma-Gaussian composite cross-transiogram models were also proposed (Li et al., 2012). Furthermore, when sample point data in a study area are sufficient and the experimental transiograms are reliable, linear interpolation, which was proposed as a joint model-fitting method, is satisfactory and more efficient than mathematical model fitting (Li & Zhang, 2010a, 2010b).

From the above discussion, it could be found that, since the shapes of the transiograms are very complex, various fitting models have been proposed in last decade to improve the fitting performance. To support spatial simulation of landscape categories in a study area, the model fitting may be a hard task if the number of categories (classes) is large and all the fitting work is conducted manually. For example, if there are 10 classes, 100 fitting models for transiograms are required. To fit each experimental transiogram, one model must be selected from various model choices and a few parameters also need to be set for each model. Taking gamma distribution-based composite models as an instance, they have five parameters to set, including sill, range, alpha, beta and weight, which must be performed by trial and error or experience (Li et al., 2012). When the experimental transiograms have more complex shapes to fit, modelling by these complex mathematical models means a large and time-consuming task and expert knowledge is crucial to support the modelling process. Although linear interpolation is a simpler joint modelling method which can be conducted without regard to the number and complexity of experimental transiograms, manual

manipulation for fitting many experimental transiograms by the linear interpolation method is still a tedious task for researchers. Under this situation, developing a framework of experimental transiogram modelling and a relatively integrated software tool is essential to facilitating the model-selection and parameter-setting jobs and improving the efficiency of the model fitting process.

In the field of environmental and geographical sciences, a lot of software tools were developed to release the burden in labor-intensive and time-consuming manual data processing, analysis and spatial modelling over recent years (Borouhaki & Malczewski, 2008; Chen, Yu, & Khan, 2010; Chen, Yu, & Khan, 2013; Her et al., 2017; Labiosa et al., 2013; Liu, Guo, & Tian, 2012; Silva, Alçada-Almeida, & Dias, 2014; Tepe & Guldmann, 2017; Yu, Chen, & Wu, 2011). Related to transition probability or variogram estimation, several frameworks or tools have been developed. Carle (1999) released the T-PROGS software to enable implementation of a transition probability-based indicator geostatistical approach for simulation of sedimentary facies, in which a transition rate method was developed for estimating transiogram models (i.e., continuous-lag Markov chain model, see Carle & Fogg, 1997). Faulkner (2002) developed a graphical application for variogram modelling based on Java classes. The application supports a least-squares fitting algorithm for robust parametric variogram model fitting, as well as other traditional variogram plotting and fitting utilities. Truong et al. (2013) developed a web-based tool for applying existing statistical expert elicitation techniques to extract the variogram of a regionalized variable that is assumed to have a multivariate normal or lognormal probability distribution from expert knowledge. Barca, Porcu, Bruno, and Passarella (2017) provided an extensive cross-validation procedure and implemented it in a software application to support practitioners in the variogram model assessment. The software can summarize a large post-processing stage and suggests how to interpret the performed analysis to rate the model to be validated. These frameworks or software tools greatly facilitate later studies in modelling procedure, data processing, and output result analysis. However, for the modelling methodology of transiograms, no visual framework or software tool has been developed till now. It is of great importance to building up a new framework to support the model fitting work of transiograms, especially experimental transiograms, to further enhance the efficiency of parameter preparation for Markov chain simulation.

In this study, a framework of experimental transiogram modelling is developed. The framework integrates the mathematical models and joint model-fitting methods, and it is developed as a tool, entitled TGRAM, to facilitate the model fitting jobs of practitioners and researchers. The modules and processing flows of the framework are described and the usages of the TGRAM tool are introduced. Two case studies are presented to demonstrate the feasibility and practicability of the tool. In the next section, we briefly introduce the main modules of the framework and its workflows. In Section 3, we illustrate a brief user guide for the TGRAM tool. Section 4 and 5 provide case studies and discussions. Concluding remarks are addressed in Section 6.

2. Framework

The framework of modelling experimental transiograms is shown in Fig. 1. The framework includes four modules: (1) spatial data importing module, (2) experimental transiograms estimation module, (3) transiograms modelling module and (4) final transiogram models generation module. These four modules are introduced in detail in the following sections. Before introducing the modules of this framework, we assume that readers/users know or are familiar with the fundamental concepts and theories of the transiogram, transition probability, and one-dimensional Markov chain. The basic theory of transiograms can be found from the papers of Li (2007a), Li and Zhang (2010a, 2010b) and Li et al. (2012).

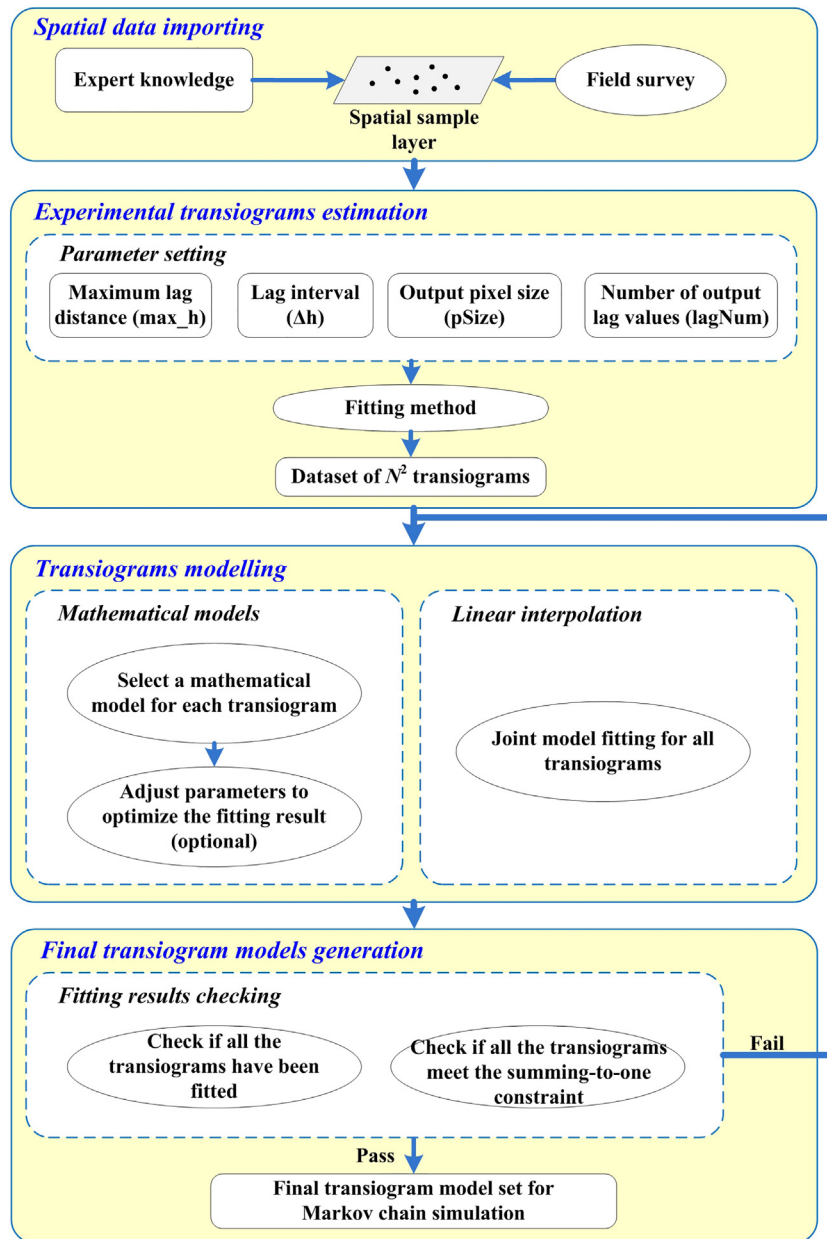


Fig. 1. The framework of modelling experimental transiograms.

2.1. Spatial data importing

The spatial sample data required for estimating experimental transiograms are sample point data. The sample point data in this study is a point layer of the ESRI shapefile format and must have a projected coordinate system. In addition to the spatial location of a point, each sample point needs to have an attribute to save its class value. For example, if the sample data are of land use/land cover, the class values mean different land use/land cover types, denoted as integer numbers. The sample points are usually collected by field surveys and expert interpretation. Investigators can on-site observe real land use/land cover types or interpret the land use types from remotely-sensed images based on expert knowledge. The location and attributes of each point in the sample layer will be read and imported into the framework as the data source. The spatial extent of the sample point layer, which is commonly the minimum bounding rectangle of the layer, also will be recorded for further use in a subsequent module.

2.2. Experimental transiograms estimation

Based on the sample point data, the experimental transiograms are able to be estimated. But four parameters need to be set in advance. The first two parameters are “maximum lag distance” (max_h) and “lag interval” (Δh). Since the sample layer is in a continuous space, max_h and Δh are represented by a continuous distance measure. These two parameters are not easy to be set by users because they may not know what values are appropriate in advance. Therefore, the recommended lag values are provided in the framework on the basis of the spatial extent of the sample point layer (i.e., the minimum bounding rectangle). The recommended Δh and max_h can be calculated as follows:

$$\Delta h = [(\sqrt{H^2 + W^2})/2]/c \tag{1}$$

$$max_h = \Delta h \times c \tag{2}$$

where H and W are the height and width of the minimum bounding rectangle of the sample point layer, and c is a constant, which has the value of 20 in this study. c means the max_h is divided into c intervals,

that is, the maximum number of lag values for estimating experimental transiograms. So we have 20 data values for an experimental transiogram. The users can accept the recommended Δh and \max_h or adjust the values. The values of these two parameters are of great importance since they can directly impact the fluctuations of experimental transiograms.

The other two parameters that need to be preset are “output pixel size” (pSize) and “number of output lag values” (lagNum). These two parameters are related to the Markov chain spatial simulation. Because Markov chain spatial simulation usually uses a discrete space, pSize represents the size of one pixel in the discrete space and lagNum represents the actual number of lag values for final output transiogram models.

The last option which has to be selected is the “fitting method”. There are generally two joint model-fitting methods for experimental transiograms: mathematical model fitting and linear interpolation. The linear interpolation method is different from mathematical model fitting because it does not use any specific mathematical models and it can be utilized regardless of the number and complexity of experimental transiograms, but it only may be used when experimental transiograms are reliably estimated based on sufficient sample data.

After all of the parameters are set, a dataset which contains N^2 experimental transiograms will be generated. Here N is the number of classes in the sample layer. For example, if we have a sample layer with 3 classes ($N = 3$) and $p_{(i, j)}$ is used to denote a transiogram with class i being the tail class and class j being the head class, then 9 experimental transiograms will be generated, including $p_{(1, 1)}$, $p_{(1, 2)}$, $p_{(1, 3)}$, $p_{(2, 1)}$, $p_{(2, 2)}$, $p_{(2, 3)}$, $p_{(3, 1)}$, $p_{(3, 2)}$ and $p_{(3, 3)}$, forming a 3×3 transiogram matrix.

2.3. Transiograms modelling

The “transiograms modelling” module is divided into two parts, which can be selected through the option of “fitting method” in “experimental transiograms estimation” module. If the “Mathematical models” is selected for this option, each experimental transiogram need to be fitted by a mathematical model, except for the one that takes the value of (1.0 - others) in each row of the transiogram matrix. The users may adjust the parameters of different models to optimize the fitting results. If “Linear interpolation” is selected for the “fitting method” option, the model fitting of transiogram will be a batch procedure for all the experimental transiograms. The detailed workflow is given in Fig. 2.

Fig. 2a illustrates the workflow of transiogram joint modelling by mathematical models. Firstly, an experimental transiogram $p_{(i,j)}$ need to be selected. If it is an auto-transiogram, it may have two choices: fitted by (1.0 - others) or fitted by a specific mathematical model. The requirements for fitting using (1.0 - others) is described in Table 1. If we do not fit the transiogram by (1.0 - others), several mathematical models can be selected for this auto-transiogram, including linear, spherical, exponential, Gaussian, Cosine-exponential and Cosine-Gaussian (Table 1). Through adjusting the input parameters of a selected model, the fitting result is optimized to be acceptable. If the selected experimental transiogram is a cross-transiogram, besides the two choices of fitting by (1.0 - others) and fitting by a mathematical model, it may also be fitted by using the strategy of inferring $p_{ij}(h)$ from the fitting model $p_{ij}(h)$ based on the detailed balance relationship, but this latter choice is not applicable to uni-directionally estimated experimental transiograms. The equations of mathematical models for cross-transiograms are different from those for auto-transiograms (Table 1). Except for the mathematical models for both auto and cross-transiograms, there are three additional models for cross-transiograms, including Gamma-exponential, Gamma-Gaussian and Gamma-spherical models (Table 1) for fitting the first or only peak of experimental cross-transiograms of neighboring classes. Input parameters for these three models are more than those for other models. Manually fitting experimental cross-transiograms using these models may be hard and time-

consuming. The framework presented in this paper can greatly reduce the difficulty and improve the efficiency of the fitting tasks. The fitting process is reiterated by manually selecting an unfitted experimental transiogram and visually checking the model fitness, and it ends when each experimental transiogram has a fitted model. If the fitted transiogram model by (1.0 - others) has negative values at some lags, that means other transiograms were not fitted perfectly. Such situation often occurs when the head class of the fitted transiogram model by (1.0 - others) is a minor class (i.e., the experimental transiogram being fitted by (1.0 - others) has a very low sill). To avoid this situation, it is suggested that using the (1.0 - others) model to fit an experimental auto/cross-transiogram with a relatively large head class (i.e., with a relatively high sill). The framework is flexible that if the users are not satisfied by the fitted results, they can adjust them again to get a set of better results.

Fig. 2b presents the workflow of fitting using linear interpolation. The workflow is much simpler than that for the fitting method using mathematical models. The entire procedure is a circulation that the linear interpolation model is applied to each transiogram in the dataset of N^2 experimental transiograms which is obtained in “experimental transiograms estimation” module and it can be done automatically. The equation for all of the experimental transiograms is uniform (Table 1) and the input parameters for linear interpolation are pSize and lagNum, which are also acquired by the previous module and the values of them will not change in the fitting procedure.

2.4. Final transiogram models generation

When the model fitting work has been accomplished in the “transiograms modelling” module, the final result is ready to be generated. Before generating the final outcome, fitting results done in the previous module are required to be checked. Firstly, the framework always checks whether all of the experimental transiograms have been fitted or not. Secondly, if we used mathematical models to fit transiograms, the framework checks whether all transiogram models meet the summing-to-one constraint in each transiogram matrix row. The summing-to-one constraint is ensured by the fitting model of “(1.0 - Others)”. Table 1 can be referred to for the requirements of this model. If either of these two checking processes of the framework fails, the user is asked to go back to the “transiograms modelling” module to check and validate the fitting results. Otherwise, if the checking is passed, the final result will be generated for use in Markov chain spatial simulation. The final result is a data matrix, of which each column is corresponding to a transiogram model $p_{ij}(h)$ in the dataset of N^2 transiogram models. Therefore, the column number of the matrix is N^2 plus one column for lag values. The row number of the matrix is lagNum + 1. The rows are corresponding to the lag values (≥ 0) from 0 to lagNum \times pSize. An example of the final result is given in Section 3.

3. TGRAM tool

A tool called TGRAM was developed to implement the framework of modelling experimental transiograms (Fig. 1). Given the need for cartographic capabilities in spatial visualization and operations, SharpMap was used to read and visualize the sample point layer (shapefile format). C#.Net high-level programming language was used to implement the framework to a software tool. The fitting results and final results are stored in Microsoft Excel files for the convenience of further data processing and diagramming.

Fig. 3 shows the screen shots of the TGRAM tool. It illustrates the operational procedure for experimental transiogram joint modelling as follows:

- (1) Importing sample data: Select *Open Shapefile* in the *File* menu to add a shapefile point layer (Fig. 3a). The sample point layer will be visualized in the view window of the tool (Fig. 3b) after being

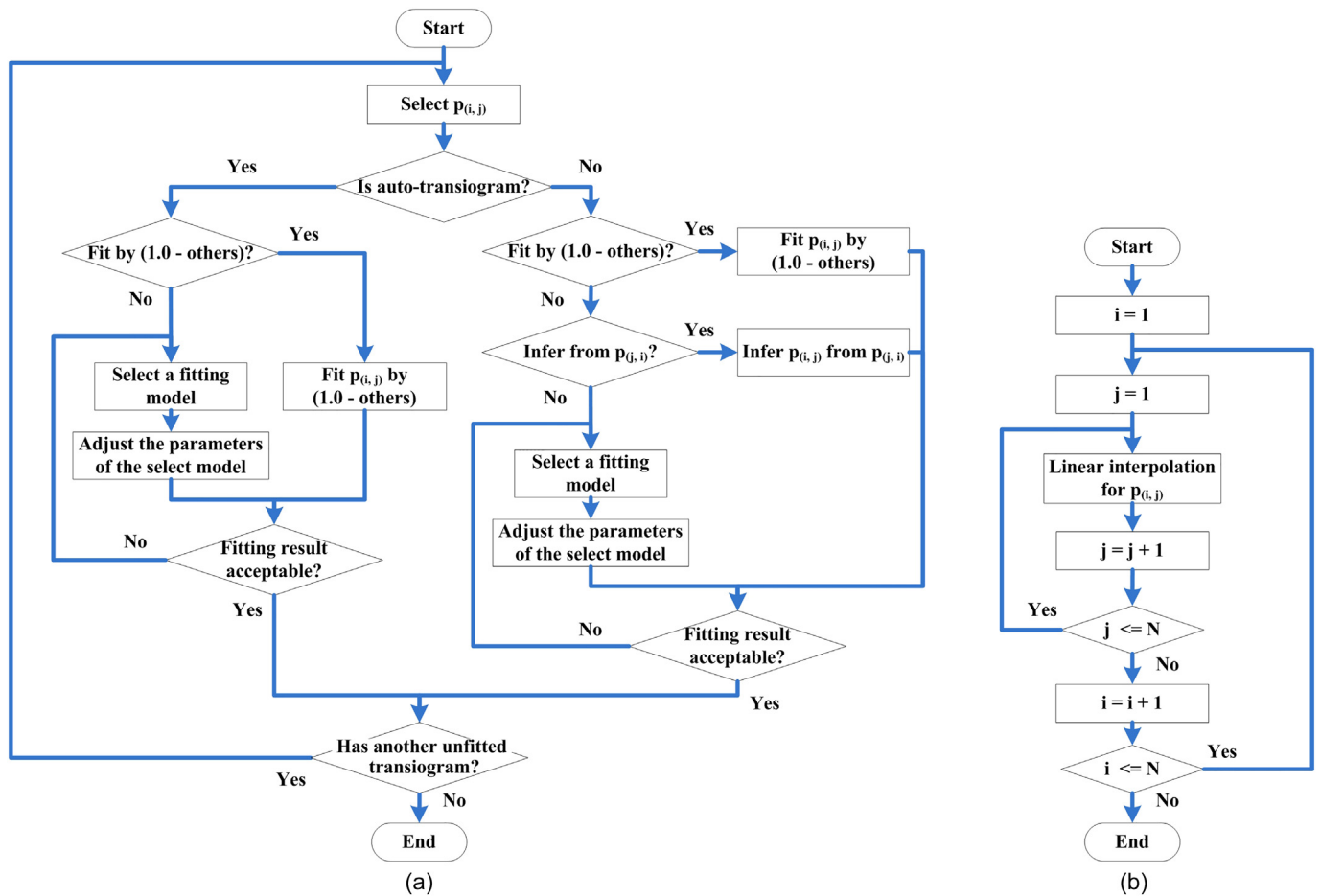


Fig. 2. The workflows of the two model fitting methods of experimental transiograms. (a) joint-fitting by mathematical models. (b) linear interpolation. In the workflow charts, $p_{(i, j)}$ is a transiogram, where class i is the tail class and class j is the head class; N is the number of classes.

- imported. The following modelling processes are all based on this data layer. In the *Estimate Exp. Transiograms* window, the *Layer name* and the *Class field* of the sample data layer need to be selected.
- (2) Recommending lag values: Click on the *Recommended Lag Values* button in the *Estimate Exp. Transiograms* window (Fig. 3b) to provide recommended values of \max_h and Δh (Fig. 3c). If the recommended values are not satisfactory, the user can manually modify the values in the *Estimate Exp. Transiograms* window (Fig. 3d).
 - (3) Estimating experimental transiograms: Other parameters that need to set in the *Estimate Exp. Transiograms* window include the path for saving the dataset of the output transiogram models, the *Output pixel size* ($pSize$), the *Number of output lag values* ($lagNum$) and the *Fitting method* for modelling experimental transiograms (Fig. 3d). The two options for the fitting method are “Linear interpolation” and “Mathematical models”. After setting these parameters, clicking *Generate Transiograms* button can generate the dataset of experimental transiograms.
 - (4) Loading experimental transiograms: In the *View/Fit Exp. Transiograms* window, clicking the *Select the Folder* button can load the dataset of experimental transiograms from its storing path. The dataset is displayed as a list, of which each row corresponds to a transiogram. For the convenience of expression, the name of a transiogram is in the form of $P(i, j)$ (Fig. 3e).
 - (5) Modelling experimental transiograms: Select one row in the list, such as the row of $P(1, 1)$, and then click *View This Transiogram* button (Fig. 3e). If the mathematical model fitting method is selected, a *View Transiogram* window will appear and show the experimental transiogram. Models listed in Table 1 can be selected

- and model parameters can be configured. Clicking the *Fitting by This Model* button will display the fitting result with a green curve. The *Save Fitting Result* button realizes the saving function for the model fitting result (Fig. 3f). If the linear interpolation method is selected, all of the experimental transiograms in the list will be automatically fitted by the linear interpolation model through one iterative run.
- (6) Fitting results checking: All the fitted transiograms are displayed with the green background color. The model names are listed in one column, such as “Exponential Model”, “Gamma-Exponential Model” and “Gaussian Model” (Fig. 3g). If the equation “ $P_{ij}(h) = (P_j/P_i) * P_{ji}(h)$ ” appears in the *Model Name* column for a transiogram (Fig. 3g), that means the transiogram model is inferred using the equation. Similarly, “1.0 - Others” (Fig. 3g) means the corresponding transiogram model is fitted by (1.0 - Others). If linear interpolation is used for transiogram model fitting, all of the model names will be listed as “Linear interpolation” (Fig. 3h). Users can check the list to validate and optimize the fitting results. The *Generate Final Result* command in the *Process* menu (Fig. 3i) can simultaneously check all of the fitting results and signal the problems before generating the whole set of transiogram model data for use in Markov chain simulation (Fig. 1).
 - (7) Generating final result: The final result of modelling experimental transiograms is generated after the fitting results checking process is passed. The final result is saved as a matrix in an Excel file (Fig. 3j). Each column of the matrix represents one fitted result of a transiogram. An element value in the matrix indicates the transition probability from class i to class j at lag value h . The values of the transition probabilities are capable of supporting Markov chain simulation, if it is generated according to the output pixel size.

Table 1
Fitting models of transiograms supported by the proposed framework.

Method	Model name	Equation ^a	Input parameters or requirements ^b
Mathematical model	<i>For auto-transiograms</i>		
	Linear	$p_{ii}(\mathbf{h}) = \begin{cases} 1 - (1 - p_i)\mathbf{h}/a_i, & \mathbf{h} < a_i \\ p_i, & \mathbf{h} \geq a_i \end{cases}$	a_i
	Spherical	$p_{ii}(\mathbf{h}) = \begin{cases} 1 - (1 - p_i)[1.5(\mathbf{h}/a_i) - 0.5(\mathbf{h}/a_i)^3], & \mathbf{h} < a_i \\ p_i, & \mathbf{h} \geq a_i \end{cases}$	a_i
	Exponential	$p_{ii}(\mathbf{h}) = 1 - (1 - p_i)[1 - \exp(-3\mathbf{h}/a_i)]$	a_i
	Gaussian	$p_{ii}(\mathbf{h}) = 1 - (1 - p_i)\{1 - \exp[-3(\mathbf{h}/a_i)^2]\}$	a_i
	Cosine-exponential	$p_{ii}(\mathbf{h}) = 1 - (1 - p_i)[1 - \exp(-3\mathbf{h}/a_i) \cos(2\pi\mathbf{h}/\lambda)]$	a_i, λ
	Cosine-Gaussian	$p_{ii}(\mathbf{h}) = 1 - (1 - p_i)\{1 - \exp[-3(\mathbf{h}/a_i)^2] \cos(2\pi\mathbf{h}/\lambda)\}$	a_i, λ
	<i>For cross-transiograms (i ≠ j)</i>		
	Linear	$p_{ij}(\mathbf{h}) = \begin{cases} p_j\mathbf{h}/a_{ij}, & \mathbf{h} < a_{ij} \\ p_j, & \mathbf{h} \geq a_{ij} \end{cases}$	a_{ij}
	Spherical	$p_{ij}(\mathbf{h}) = \begin{cases} p_j[1.5(\mathbf{h}/a_{ij}) - 0.5(\mathbf{h}/a_{ij})^3], & \mathbf{h} < a_{ij} \\ p_j, & \mathbf{h} \geq a_{ij} \end{cases}$	a_{ij}
	Exponential	$p_{ij}(\mathbf{h}) = p_j[1 - \exp(-3\mathbf{h}/a_{ij})]$	a_{ij}
	Gaussian	$p_{ij}(\mathbf{h}) = p_j\{1 - \exp[-3(\mathbf{h}/a_{ij})^2]\}$	a_{ij}
	Cosine-exponential	$p_{ij}(\mathbf{h}) = p_j\{1 - \exp(-3\mathbf{h}/a_{ij}) \cos(2\pi\mathbf{h}/\lambda)\}$	a_{ij}, λ
	Cosine-Gaussian	$p_{ij}(\mathbf{h}) = p_j\{1 - \exp[-3(\mathbf{h}/a_{ij})^2] \cos(2\pi\mathbf{h}/\lambda)\}$	a_{ij}, λ
	Gamma-exponential ^c	$p_{ij}(\mathbf{h}) = p_j \left[1 - \exp(-3\mathbf{h}/a_{ij}) + \frac{w}{\Gamma(\alpha)\beta^\alpha} \left(\frac{\mathbf{h}}{a_{ij}}\right)^{\alpha-1} \exp\left(\frac{-\mathbf{h}}{\beta a_{ij}}\right) \right], \quad \alpha > 1, \beta > 0$	a_{ij}, α, β, w
	Gamma-Gaussian ^c	$p_{ij}(\mathbf{h}) = p_j \left[1 - \exp(-3\mathbf{h}^2/a_{ij}^2) + \frac{w}{\Gamma(\alpha)\beta^\alpha} \left(\frac{\mathbf{h}}{a_{ij}}\right)^{\alpha-1} \exp\left(\frac{-\mathbf{h}}{\beta a_{ij}}\right) \right], \quad \alpha > 1, \beta > 0$	a_{ij}, α, β, w
	Gamma-spherical ^c	$p_{ij}(\mathbf{h}) = \begin{cases} p_j \left[1.5 \frac{\mathbf{h}}{a_{ij}} - 0.5 \left(\frac{\mathbf{h}}{a_{ij}}\right)^3 + \frac{w}{\Gamma(\alpha)\beta^\alpha} \left(\frac{\mathbf{h}}{a_{ij}}\right)^{\alpha-1} \exp\left(\frac{-\mathbf{h}}{\beta a_{ij}}\right) \right], & 0 < \mathbf{h} < a_{ij} \\ p_j \left[1 + \frac{w}{\Gamma(\alpha)\beta^\alpha} \left(\frac{\mathbf{h}}{a_{ij}}\right)^{\alpha-1} \exp\left(\frac{-\mathbf{h}}{\beta a_{ij}}\right) \right], & \mathbf{h} \geq a_{ij} \end{cases}, \quad \alpha > 1, \beta > 0$	a_{ij}, α, β, w
	Infer p_{ij} from p_{ji}	$p_{ij}(\mathbf{h}) = \frac{p_j}{p_i} p_{ji}(\mathbf{h})$	After $p_{ji}(\mathbf{h})$ has been fitted
<i>For auto and cross-transiograms</i>			
Fit by (1.0 - Others)	$p_{ik}(\mathbf{h}) = 1 - \sum_{j=1, j \neq k}^n p_{ij}(\mathbf{h})$	After all other $p_{ij}(\mathbf{h})$ in a transiogram matrix row have been fitted, the left one is fitted by this equation	
Linear interpolation	Linear interpolation ^d	$p_{ij}(\mathbf{h}) = \frac{\widehat{p}_{ij}(\mathbf{h}_k)(\mathbf{h}_{k+1} - \mathbf{h}) + \widehat{p}_{ij}(\mathbf{h}_{k+1})(\mathbf{h} - \mathbf{h}_k)}{\mathbf{h}_{k+1} - \mathbf{h}_k}$	Output pixel size (pSize), number of output lag values (lagNum)

^a $\mathbf{h} \geq 0$ for all equations.

^b a_i = auto-correlation range; a_{ij} = cross-correlation range; p_i = proportion of class i ; λ is the wavelength of the cosine function; α is the shape parameter of the gamma distribution function; β is the scale parameter of the gamma distribution function; w is a weight parameter for the gamma distribution function component in the composite model of Gamma-exponential, Gamma-Gaussian and Gamma-spherical; In the Gamma distribution function, $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$.

^c Gamma-exponential, Gamma-Gaussian and Gamma-spherical model for cross-transiograms were provided in Li et al. (2012).

^d Linear interpolation model was provided in Li and Zhang (2010a, 2010b). Other fitting models were provided in Li (2007a).

4. Case studies

The following two cases demonstrate the feasibility of the developed framework and TGRAM tool.

4.1. Case study 1

The study area is located in Bristol town, Connecticut, USA, which has an area of 69.45 km² (Fig. 4). It has six different land cover classes, including built-up area (class 1), grass (class 2), agricultural field (class 3), forest (class 4), waterbody (class 5) and barren land (class 6). A dense dataset of 3682 sample points, which were interpreted by experts, were used for estimation of experimental transiograms. Considering the sample points were dense in the study area, the linear interpolation method was used to obtain transiogram models from experimental transiograms. The parameter values for generating the transiogram models are as follows: The max_h was set as 6000 m; $\Delta\mathbf{h}$ was 200 m; pSize was 50 m and lagNum was 120. Because the linear interpolation method is simple and does not need experts' decision and adjustment in model fitting, the joint model fitting process was

completed iteratively for all of the 36 transiograms in several minutes.

Two examples of the generated transiogram models are presented in Fig. 5. In this figure, $P(i, j)$ denotes the transition probability from class i to class j . Fig. 5a shows that there are no strong fluctuations when estimating the transition probabilities of $P(1, j)$ from the dense sample data. The experimental transiograms can be fitted well by linear interpolation method and the fitted results are reliable. Fig. 5b gives the transition probabilities of $P(5, j)$. Because there are only 58 sample points of class 5, which are 1.58% of all the samples, class 5 is a minor class. Although the experimental transiograms display irregular convexes and concaves, there are still no steep irrational fluctuations, which means that the experimental transiograms are reliable, at least in the low lag section. While mathematical model fitting methods may miss such details, linear interpolation can capture all the features, such as $P(5, 1)$ and $P(5, 4)$ in Fig. 5b.

4.2. Case study 2

Another study area is in Yangpu District, Shanghai, China. The study area is a square subarea in the district, which has an area of

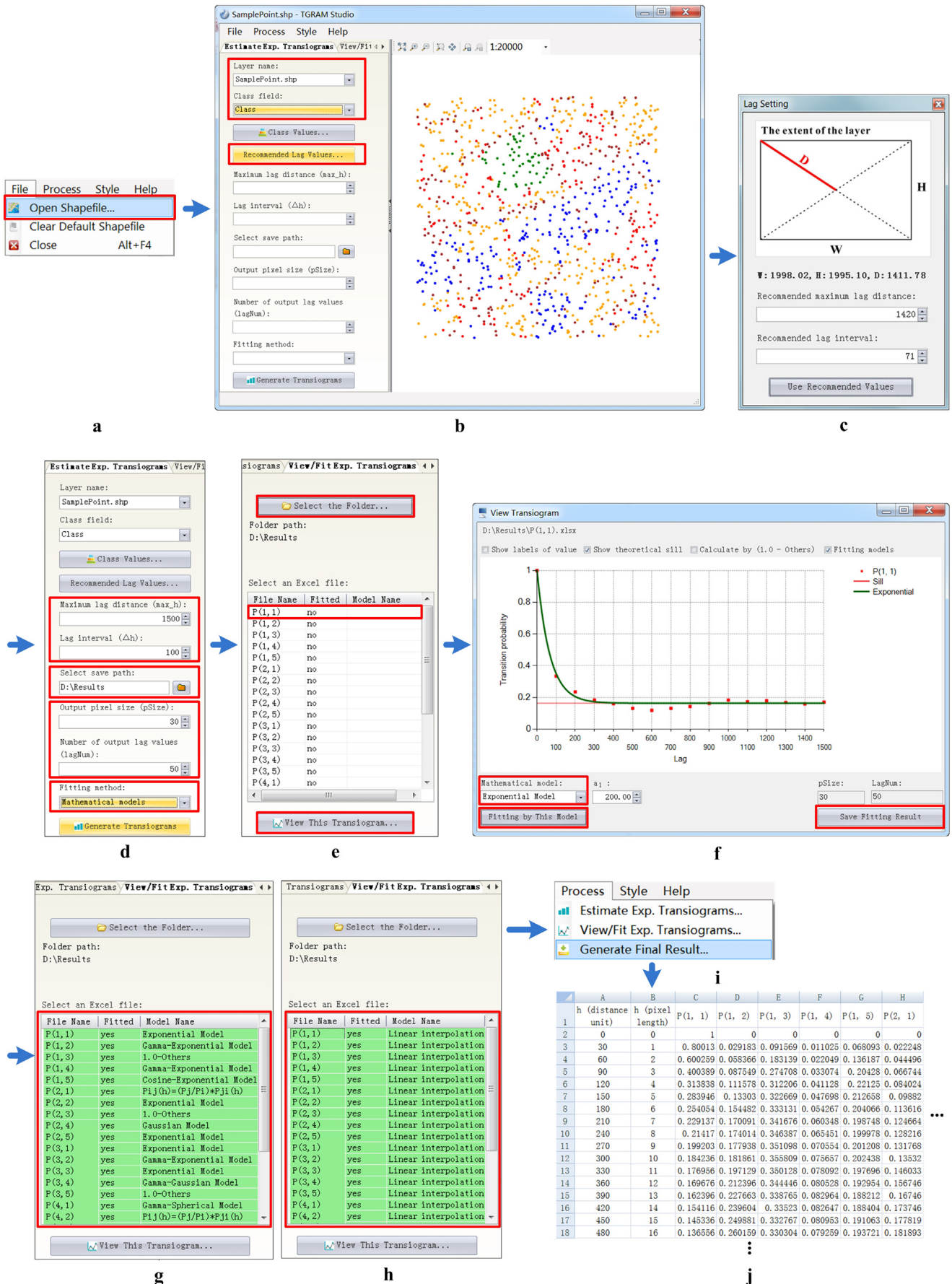


Fig. 3. Steps of modelling experimental transiograms in the TGRAM tool.

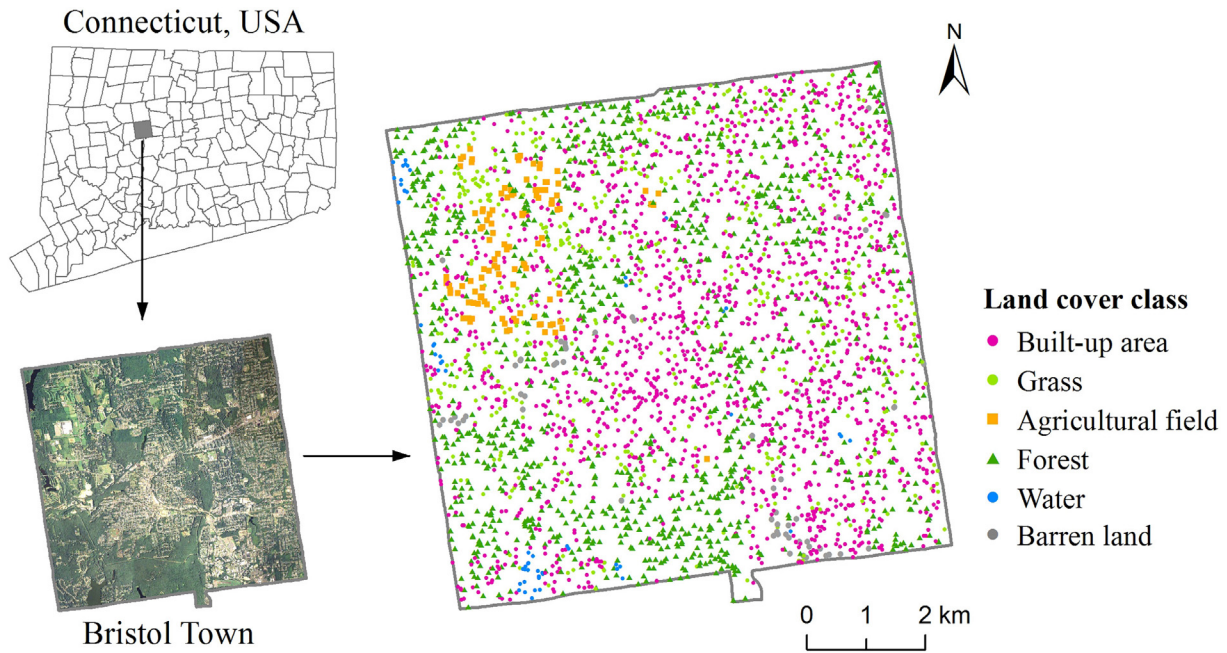


Fig. 4. The study area in Bristol town, Connecticut, USA and the sample point dataset interpreted by experts.

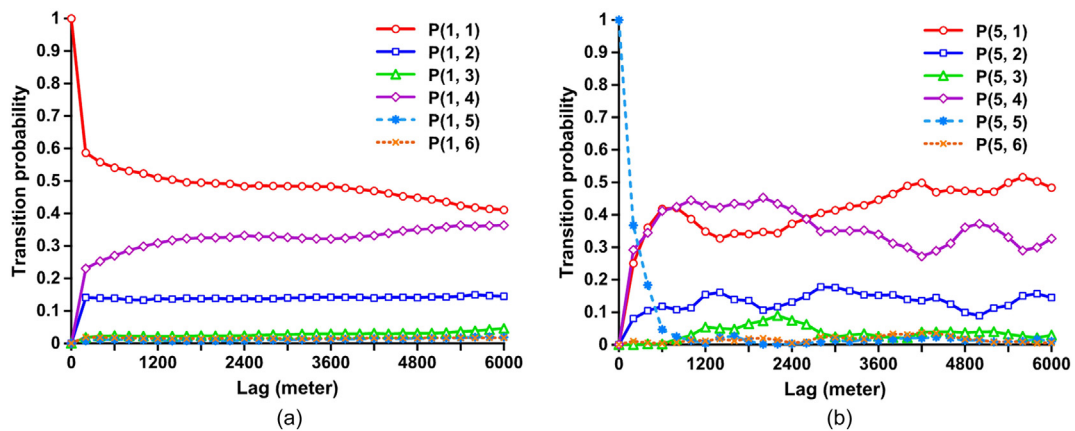


Fig. 5. Two examples of transiogram models fitted by linear interpolation from experimental transiograms for the study area in Bristol town, Connecticut, USA.

4.0 km² (Fig. 6). It is in the urban region of Shanghai and has spatially heterogeneous and complex urban land uses. The main land uses of the study area are commercial and public service land (class 1), industrial land (class 2), residential land (class 3), green infrastructure (class 4) and street (class 5). A relative dense dataset of 998 sample points, which were also interpreted by local experts, were utilized to support the estimation of experimental transiograms. Different from the case study in Bristol town, the fitting method of mathematical model was used in this case study. The parameters were also preset for generating the experimental transiograms. The max_h was set as 1500 m; Δh was 100 m; pSize was 30 m and lagNum was 50. 25 experimental transiograms were generated and each of them was fitted by a mathematical model.

The model fitting results are presented in Fig. 7. Some basic mathematical models were used, such as the exponential model, Gaussian model, and spherical model. The cosine-exponential model was applied to capture the peak in the low lag section of P(1, 5). But this model has its limitation in fitting the peaks of other cross-transiograms. Therefore, the gamma-exponential model, gamma-spherical model, and gamma-Gaussian model were employed to closely fit the single-peak features of some other transiograms, such as P(1, 2), P(4, 1) and P(4, 3). To improve the efficiency of model fitting, we inferred some transiogram

models from other fitted transiogram models. For example, P(1, 2) was inferred from P(2, 1). It can be seen that both of these two experimental transiograms can be fitted very well. In each line of the five transiograms in Fig. 7, there is one and only one transiogram which was fitted by (1.0 - Others). The fitting effect of these transiograms was determined by the fitting results of other transiograms. Judging from Fig. 7, most of these transiograms were fitted well.

5. Discussions

It can be easily found from the above two case studies that the developed framework and TGRAM tool can facilitate the transiogram modelling work. The advantages of the framework are as follows:

- (1) The proposed framework is in a spatial context. This characteristic can greatly enhance the efficiency of experimental transiogram estimation than manual operations. Furthermore, different from traditional sample data formats, such as text file, or raster data, this framework adopts the shapefile data format to store sample points. Because the shapefile format is an open format, various sample data of other formats can be converted into shapefile data. This spatial characteristic improves the compatibility of the TGRAM tool.

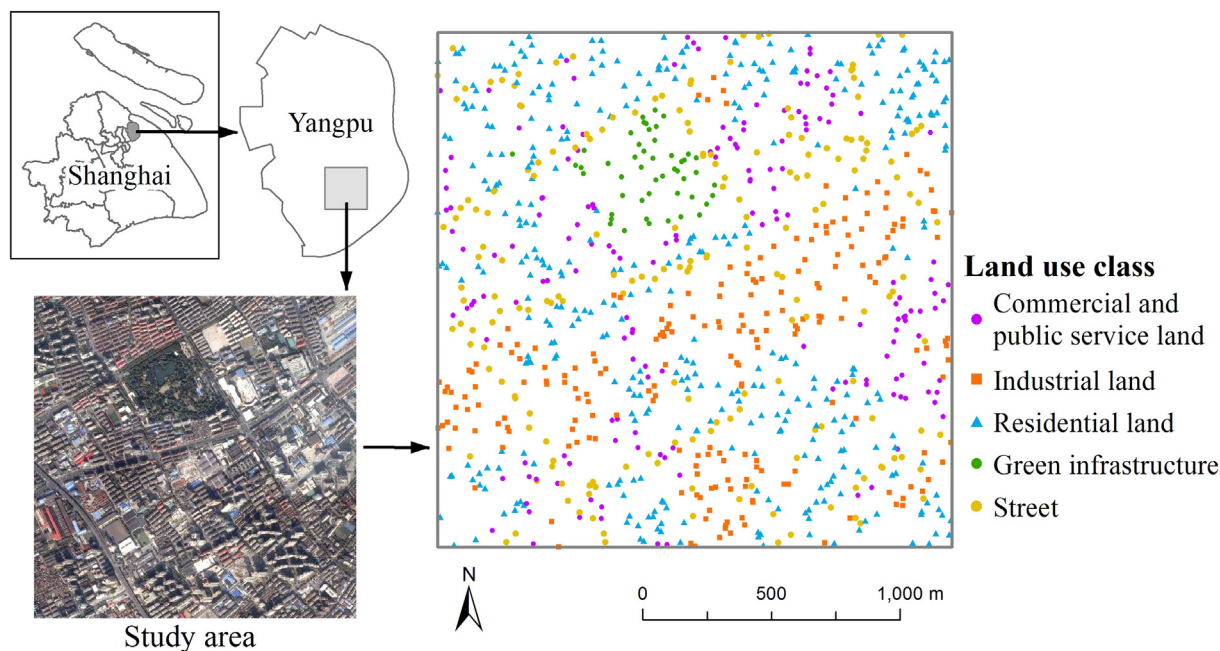


Fig. 6. The study area in Yangpu district, Shanghai, China and the sample point dataset interpreted by experts.

- (2) The option of using mathematical models or using linear interpolation in the tool makes the transiogram modelling work more flexible. Li and Zhang (2010a, 2010b) introduced that mathematical model fitting and linear interpolation have their respective advantages. The main advantage of linear interpolation is that it can complete the transiogram modelling work quickly regardless of the number and complexity of experimental transiograms. When sample points are abundant and experimental transiograms are reliable, linear interpolation is a very good choice. However, when samples are sparse, mathematical model fitting based on expert knowledge is crucial to obtaining reliable transiogram models. If samples are dense, mathematical model fitting is also satisfactory to transiogram modelling. But it takes longer time than linear interpolation. With the TGRAM tool, users can easily make their choices according to the complex situations of different case studies.
- (3) The TGRAM tool provides recommended Δh and \max_h based on Eq. 1 and Eq. 2, respectively. Since different study areas have different scales, it is not convenient for users to determine these two parameters without any reference. The function of *Recommended Lag Values* (Fig. 3b) can provide the parameter values. For example, in the case study in Yangpu District, the recommended Δh is 71 m and the recommended \max_h is 1420 m. Referring to these two values, we set the Δh as 100 m and the \max_h as 1500 m.
- (4) When samples are relatively sparse, choosing a suitable lag interval is critical for obtaining reliable experimental transiograms. However, we have to look for a relatively optimal lag interval by trial and error. Fig. 8 shows the $P(1, 2)$ of the case study in Yangpu District, Shanghai, China. When we set the Δh to 10 m, the experimental transiogram fluctuates steeply. But when we increase the Δh to 100 m, the experimental transiogram becomes much more stable and reliable. It is not easy to estimate which value of Δh is better without trying. Fortunately, the TGRAM tool can conduct the estimation work of a whole set of experimental transiograms with a certain lag interval quickly for a usual study area (only in a few minutes) and then visually display them. Therefore, it is not very time-consuming to find a suitable lag interval using the tool by several tries. The rule of thumb is to avoid strong fluctuations in experimental transiograms while keeping the basic features.
- (5) To fit an experimental transiogram by mathematical models, one needs to select a suitable mathematical model from various model choices. Manual model-selection may also be a hard task for practitioners and researchers because it is not easy to judge whether a mathematical model can fit a transiogram well or not without trial and error. The TGRAM tool has integrated all the modelling choices into an interactive interface (Fig. 3f, Table 1). Users can conveniently select different models by a dropdown list box, which makes possible a quick trial and error process. There are 6 model choices for auto-transiograms and 10 choices for cross-transiograms in TGRAM, if not including the (1.0 - Others) choice.
- (6) After the model-selection step, the selected mathematical model has multiple parameters to set (Table 1). For instance, the gamma-exponential model has four parameters to set, including a_{ij} , α , β and w (still is fixed to p_j in the current version of the tool). The parameter values need to be adjusted repeatedly until getting an optimal fitting result for an experimental transiogram based on the judgement of users (Li et al., 2012). For example, Fig. 9 gives the fitting results of the gamma-exponential model with three sets of different parameter values. Among these three results, it can be found that “Fit 1” has the optimal fitting effect. The TGRAM tool provides a simple one-click drawing function (clicking *Fitting by This Model* button in Fig. 3f), which can real-time display the fitting results after adjusting the parameter values. This function greatly facilitates the trial and error process.
- (7) When the number of classes (or transiograms) is large, the transiogram-modelling process may be time-consuming. To improve the efficiency of model fitting, the framework makes available a function of “Infer p_{ij} from p_{ji} ”. Theoretically, to simulate 10 classes, 45 of 100 transiogram models can be obtained by this function, which can approximately save half the time. When we are trying to fit p_{ij} , the tool can automatically detect if p_{ji} has been fitted. If p_{ji} has been fitted, it will give an option of inferring p_{ij} from p_{ji} . Moreover, if the fitting result is not satisfactory, we can further select other mathematical models to fit the transiogram.
- (8) The tool can automatically check the fitted transiogram models and make them conform to the summing-to-one constraint. Before exporting the final transiogram modelling result, there must be one

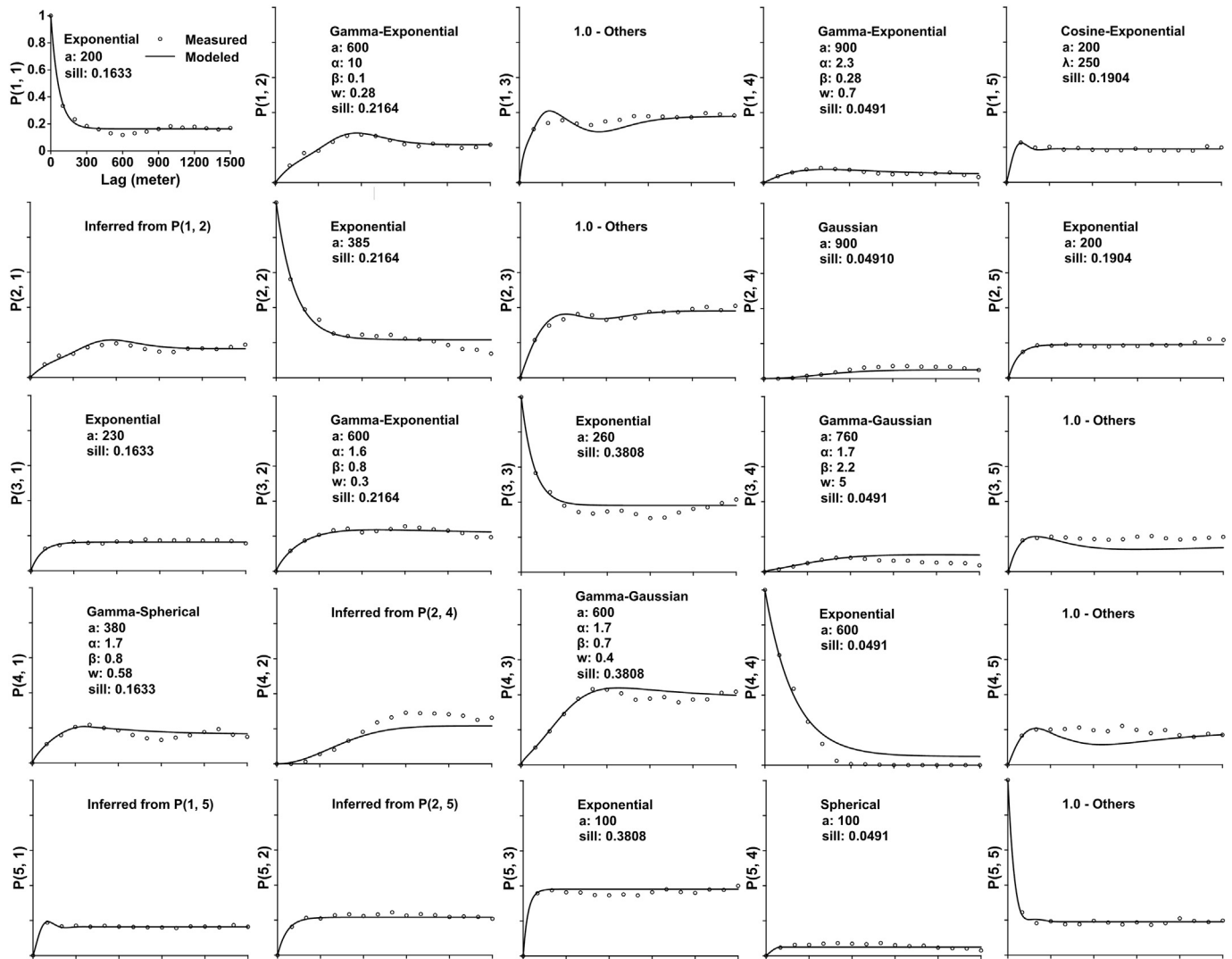


Fig. 7. Joint transiogram modelling by mathematical models for the study area in Yangpu District, Shanghai, China.

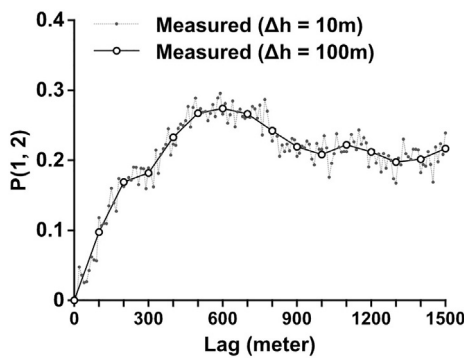


Fig. 8. Experimental transiograms with the lag intervals of 10 and 100 m for the same transition probability function $P(1, 2)$ for the study area in Yangpu District, Shanghai, China.

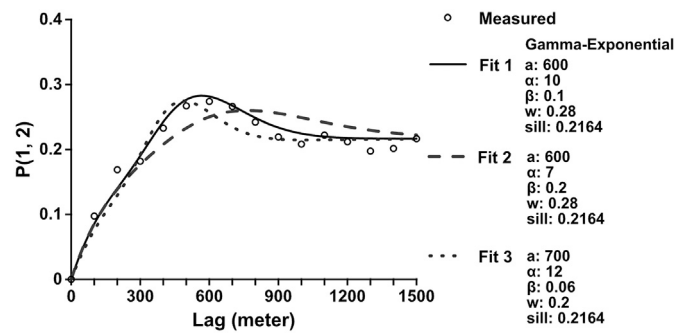


Fig. 9. Fitting an experimental transiogram by the gamma-exponential model with three different sets of parameter values for the Yangpu case study area in Yangpu District, Shanghai, China.

and only one transiogram fitted by (1.0 - Others) in each row of the transiogram matrix. Otherwise, the tool will detect the problem and give prompt feedback to assist users to adjust the model fitting.

- (9) The final result of transiogram modelling is a data matrix, generated from all the transiogram fitting models, and it is saved in an Excel file. The transition probabilities of the fitted auto and cross-transiograms at different lag values can be found from the data

matrix. MCRF simulation can directly obtain necessary parameters from this data matrix in its simulation procedure.

6. Conclusions

This study proposed a framework for modelling experimental transiograms. The framework supports visual model fitting of experimental auto and cross-transiograms using two joint model fitting

methods: the mathematical model method and the linear interpolation method. It has several advanced characteristics, such as recommended lag interval and maximum lag distance, one-click drawing function, and summing-to-one constraint checking. The framework was developed in GIS environment as a software tool, entitled TGRAM, to facilitate the experimental transiogram modelling process. Two case studies, one in Bristol town, Connecticut, USA and the other in Yangpu District, Shanghai, China, were conducted to demonstrate the feasibility of the tool. These two case studies showed that both the linear interpolation method and the mathematical model method in the tool can obtain satisfactory fitting results for the study areas. The final modelling results generated by the tool can directly support Markov chain spatial simulation work for land cover post-classification, urban growth detection, or soil mapping, among others. The framework is flexible and extendable.

However, because the shapes of experimental transiograms can be very complex sometimes and experimental transiograms with extremely minor classes may not be reliable, how to effectively perform transiogram joint modelling is a complex issue to explore. The current TGRAM tool is still an early effort. We will consider extending the functions of the tool in the future to make it more easy-to-use, and further improve the efficiency of the experimental transiogram modelling process.

Availability

The TGRAM tool with documentation is in the public domain. It may be used or distributed freely. The installation package and the source code of the tool can be downloaded for free from the website (<https://archive.org/details/TGRAM1.3>).

Acknowledgements

The TGRAM software was developed using C# computer language by Jia Yu, based on the previous simple Fortran computer programs initially developed by Weidong Li for transiogram modelling with manual operation. Partial supports from the Division of Behavioral and Cognitive Sciences of U.S. National Science Foundation [grant number 1414108], and National Natural Science Foundation of China [grant number 41201548, 5161101688], are appreciated.

References

- de Almeida, C. M., Batty, M., Monteiro, A. M. V., Câmara, G., Soares-Filho, B. S., Cerqueira, G. C., & Pennachin, C. L. (2003). Stochastic cellular automata modeling of urban land use dynamics: Empirical development and estimation. *Computers, Environment and Urban Systems*, 27(5), 481–509.
- Barca, E., Porcu, E., Bruno, D., & Passarella, G. (2017). An automated decision support system for aided assessment of variogram models. *Environmental Modelling and Software*, 87, 72–83.
- Boroushaki, S., & Malczewski, J. (2008). Implementing an extension of the analytical hierarchy process using ordered weighted averaging operators with fuzzy quantifiers in ArcGIS. *Computational Geosciences*, 34(4), 399–410.
- Burgess, T. M., & Webster, R. (1984). Optimal sampling strategies for mapping soil types. I. Distribution of boundary spacings. *Journal of Soil Science*, 35(4), 641–654.
- Carle, S., & Fogg, G. (1997). Modeling spatial variability with one and multidimensional continuous-lag Markov chains. *Mathematical Geology*, 29(7), 891–918.
- Carle, S. F. (1999). *T-PROGS: transition probability geostatistical software version 2.1 user's guide*. Davis, CA: University of California.
- Carle, S. F., & Fogg, G. E. (1996). Transition probability-based indicator geostatistics. *Mathematical Geology*, 28(4), 453–476.
- Chen, Y., Yu, J., & Khan, S. (2010). Spatial sensitivity analysis of multi-criteria weights in GIS-based land suitability evaluation. *Environmental Modelling and Software*, 25(12), 1582–1591.
- Chen, Y., Yu, J., & Khan, S. (2013). The spatial framework for weight sensitivity analysis in AHP-based multi-criteria decision making. *Environmental Modelling and Software*, 48, 129–140.
- Chiles, J. P., & Delfiner, P. (2012). *Geostatistics-Modeling Spatial Uncertainty* (2nd edition). New York: Wiley.
- Faulkner, B. R. (2002). Java classes for nonprocedural variogram modeling. *Computational Geosciences*, 28(3), 387–397.
- Grunwald, S. (2009). Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma*, 152(3–4), 195–207.
- Her, Y., Jeong, J., Arnold, J., Gosselink, L., Glick, R., & Jaber, F. (2017). A new framework for modeling decentralized low impact developments using Soil and Water Assessment Tool. *Environmental Modelling and Software*, 96, 305–322.
- Jones, T. A., & Ma, Y. Z. (2001). Teacher's aide: Geologic characteristics of hole-effect variograms calculated from lithology-indicator variables. *Mathematical Geology*, 33(5), 615–629.
- Journal, A. G., & Huijbregts, C. J. (1978). *Mining Geostatistics*. London: Academic Press.
- Labiosa, W. B., Forney, W. M., Esnard, A. M., Mitsova-Boneva, D., Bernknopf, R., Hearn, P., ... Swain, E. (2013). An integrated multi-criteria scenario evaluation web tool for participatory land-use planning in urbanized areas: The Ecosystem Portfolio Model. *Environmental Modelling and Software*, 41, 210–222.
- Li, W. (2007a). Transiograms for characterizing spatial variability of soil classes. *Soil Science Society of America Journal*, 71(3), 881–893.
- Li, W. (2007b). Markov chain random fields for estimation of categorical variables. *Mathematical Geology*, 39(3), 321–335.
- Li, W., & Zhang, C. (2010a). Simulating the spatial distribution of clay layer occurrence depth in alluvial soils with a Markov chain geostatistical approach. *Environmetrics*, 21(1), 21–32.
- Li, W., & Zhang, C. (2010b). 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., & Zhang, C. (2013). Some further clarification on Markov chain random fields and transiograms. *International Journal of Geographical Information Science*, 27(3), 423–430.
- Li, W., Zhang, C., & Dey, D. K. (2012). Modeling experimental cross-transiograms of neighboring landscape categories with the gamma distribution. *International Journal of Geographical Information Science*, 26(4), 599–620.
- Li, W., Zhang, C., Willig, M. R., Dey, D. K., Wang, G., & You, L. (2015). Bayesian Markov chain random field cosimulation for improving land cover classification accuracy. *Mathematical Geoscience*, 47(2), 123–148.
- Liu, Y., Guo, Q., & Tian, Y. (2012). A software framework for classification models of geographical data. *Computational Geosciences*, 42, 47–56.
- Logsdon, M. G., Bell, E. J., & Westerlund, F. V. (1996). Probability mapping of land use change: A GIS interface for visualizing transition probabilities. *Computers, Environment and Urban Systems*, 20(6), 389–398.
- Luo, J. (1993). *Konditionale Markovsimulation 2-dimensionaler geologischer probleme*. 4. Berlin: Berliner Geowissenschaftliche Abhandlungen, Reihe D, Band.
- Luo, J. (1996). Transition probability approach to statistical analysis of spatial qualitative variables in geology. In A. Forster, & D. F. Merriam (Eds.). *Geologic modeling and mapping* (pp. 281–299). New York, NY: Plenum Press.
- Ritzi, R. W. (2000). Behavior of indicator variograms and transition probabilities in relation to the variance in lengths of hydrofacies. *Water Resources Research*, 36(11), 3375–3381.
- Schwarzacher, W. (1969). The use of Markov chains in the study of sedimentary cycles. *Journal of the International Association for Mathematical Geology*, 1(1), 17–39.
- Silva, S., Alçada-Almeida, L., & Dias, L. C. (2014). Development of a web-based multi-criteria spatial decision support system for the assessment of environmental sustainability of dairy farms. *Computers and Electronics in Agriculture*, 108.
- Tepe, E., & Guldmann, J. M. (2017). Spatial and temporal modeling of parcel-level land dynamics. *Computers, Environment and Urban Systems*, 64, 204–214.
- Truong, P. N., Heuvelink, G. B. M., & Gosling, J. P. (2013). Web-based tool for expert elicitation of the variogram. *Computational Geosciences*, 51, 390–399.
- Vistelius, A. B. (1949). On the question of the mechanism of formation of strata. *Doklady Akademii Nauk SSSR*, 65, 191–194.
- Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists*. Chichester: Wiley.
- Weigand, H., Totsche, K. U., Huwe, B., & Kögel-Knabner, I. (2001). PAH mobility in contaminated industrial soils: A Markov chain approach to the spatial variability of soil properties and PAH levels. *Geoderma*, 102(3–4), 371–389.
- Yu, J., Chen, Y., & Wu, J. (2011). Modeling and implementation of classification rule discovery by ant colony optimisation for spatial land-use suitability assessment. *Computers, Environment and Urban Systems*, 35(4), 308–319.
- Zhang, C., & Li, W. (2008). Regional-scale modelling of the spatial distribution of surface and subsurface textural classes in alluvial soils using Markov chain geostatistics. *Soil Use and Management*, 24(3), 263–272.
- Zhang, W., Li, W., & Zhang, C. (2016). Land cover post-classifications by Markov chain geostatistical cosimulation based on pre-classifications by different conventional classifiers. *International Journal of Remote Sensing*, 37(4), 926–949.
- Zhang, W., Li, W., Zhang, C., & Li, X. (2017). Incorporating Spectral Similarity into Markov Chain Geostatistical Cosimulation for Reducing Smoothing Effect in Land Cover Postclassification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(3), 1082–1095.
- Zhang, W., Li, W., Zhang, C., & Ouimet, W. B. (2017). Detecting horizontal and vertical urban growth from medium resolution imagery and its relationships with major socio-economic factors. *International Journal of Remote Sensing*, 38(12), 3704–3734.