International Journal of Geographical Information Science
i*First*, 2012, 1–15

Taylor & Francis
Taylor & Francis Group

TECHNICAL COMMUNICATION

Comments on 'Combining spatial transition probabilities for stochastic simulation of categorical fields' with communications on some issues related to Markov chain geostatistics

Weidong Li* and Chuanrong Zhang

Department of Geography and Center for Environmental Sciences and Engineering, University of Connecticut, Storrs Mansfield, CT, USA

(*Received 5 March 2012; final version received 5 March 2012*)

Keywords: Markov chain random field; transiogram; categorical spatial variable; conditional independence; Bayesian maximum entropy; transition probability

Cao *et al.* (2011a) recently presented a method to implement the Tau model suggested by Journel (2002) for conditional simulation of categorical fields. In this article, a spatial Markov chain (SMC) model proposed by Li (2007a), as stated by them, was compared. Li (2007a) presented the Markov chain random field (MCRF) theory (or ideas) with a case study using a specific MCRF model and further suggested the Markov chain geostatistics (MCG) framework based on the primitive MCRF theory. A MCRF model may be called a SMC model and was also called in this way by Li (2007a), because a MCRF means a random field generated or defined by a special SMC. One reason that we always used MCRF instead of SMC (or more simply Markov chain) to name our models or algorithms in our later publications is because SMC is a very general name and may cover all of those existing Markov chain models used for spatial data. Unfortunately, Cao *et al.* (2011a) provided an incorrect SMC model equation with misinterpretations on the MCRF approach. Following Allard *et al.* (2011), Cao *et al.* (2011a, 2011b) also stated that the MCRF approach was a simplified form of the Bayesian maximum entropy (BME) approach proposed by Bogaert (2002); Li and Zhang (2012) commented on Allard *et al.* (2011), which was published in a soil science journal, to clarify some misunderstandings. Although this comment aims to clarify some misinterpretations by Cao *et al.* (2011a, 2011b) on the MCRF approach in geographical information science, our main purpose is to communicate with Cao and other colleagues who may have interests or misunderstandings in related issues, including those new issues raised by Allard *et al*. in their response letter to our comments on their paper. We think such a comment and communication is necessary to boost mutual understanding because it concerns the fate of MCG and the whole field of categorical spatial variable modeling. Therefore, we will only explain and discuss on several major issues rather than pick out specific sentences.

*Corresponding author. Email: weidongwoody@gmail.com

ISSN 1365-8816 print/ISSN 1362-3087 online

http://dx.doi.org/10.1080/13658816.2012.673617
http://www.tandfonline.com

1. On Markov chain random field models

In the work by Cao *et al.* (2011a), the spatial Markov chain (SMC) model proposed by Li (2007a) was written as equation [7] in the following form:

$$\begin{aligned} &P\{C(x_0) = k \mid c(x_1), \ldots, c(x_N)\}_{\text{SMC}} \\ &\quad = P\{I_k(x_0) = 1 \mid I_{k_L}(x_0^L) = 1, I_{k_U}(x_0^U) = 1, I_{k_R}(x_0^R) = 1, I_{k_B}(x_0^B) = 1\}_{\text{SMC}} \\ &\quad = \frac{\pi_{k|k_L}(h_1^x)\pi_{k_U|k}(h_2^x)\pi_{k|k_R}(h_3^y)\pi_{k_B|k}(h_4^y)}{\sum_l [\pi_{l|k_L}(h_1^x)\pi_{k_U|l}(h_2^x)\pi_{l|k_R}(h_3^y)\pi_{k_B|l}(h_4^y)]} \end{aligned} \tag{1}$$

where x and y represent the axis directions; k_L, k_U, k_R, and k_B all represent the states of the Markov chain (class indicators) at neighboring locations, x_0^L, x_0^U, x_0^R, and x_0^B, in four cardinal directions; h_1^x, h_2^x, h_3^y, and h_4^y represent the distances between these four nearest neighbors and the target location x_0, as stated by Cao *et al.* (2011a).

However, the above equation was wrongly written. One cannot find such a SMC model in Li (2007a). Besides the simplified general solution of Markov chain random fields (MCRFs), equation [19] in Li (2007a) that considered only four nearest neighbors in four cardinal directions is also different from this model. Although the style of Equation (1) is rather confusing, from the transition directions of the transiogram terms, we can guess this might be the extended coupled Markov chain (CMC) model by Li and Zhang (2006) for dealing with grid sample data. The CMC model of Elfeki and Dekking (2001) for subsurface characterization used two one-dimensional Markov chains and, therefore, had four such transition probability terms after being extended for horizontal two-dimensional simulation. However, here the directions represented by x and y (axes) and those by letters L, U, R, and B are still not identical. Writing the SMC model proposed by Li (2007a) like this, Cao *et al.* essentially ignored all of the contents of our publications that established the MCRF approach in and after 2007.

The CMC model has drawbacks as pointed out with explanation in Li (2007a) and our other papers (e.g., Li and Zhang 2008), and the drawbacks could also be seen in some figures in Elfeki and Dekking (2001) and their other papers (e.g., Elfeki and Dekking 2005). However in our point of view, the CMC model represented a progress compared with the two-dimensional linear Markov chain model of Lou (1993, 1996), which generated quite fragmentary patterns in simulated realizations; and both the models represented some progresses in terms of conditional simulation compared with the earlier unconditional multidimensional Markov chain models suggested in geosciences (e.g., Lin and Harbaugh 1984). One may argue that Carle and Fogg (1997) also proposed a multidimensional Markov chain model. That has been a misunderstanding or confusion. In Carle and Fogg (1997), the continuous-lag one-dimensional and multidimensional Markov chain models were essentially sets of idealized transition probability models (i.e., idealized transiograms). Because only vertical transition probabilities could be effectively estimated from borehole data due to the sparsity of boreholes, transition probability models in other principal directions had to be inferred based on the vertical models, expert knowledge, and geological interpretation, according to Carle and Fogg (1997). Transition probability models in any direction other than principal directions were obtained by interpolating transition probability models in principal directions. This is similar to the estimation of anisotropic variogram models. However, the interpolated transition probability models (or the transition probability interpolation method) were called continuous-lag

multidimensional Markov chain models in Carle and Fogg (1997) and might not accurately express their actual meanings. Please note that the past traditional uses of transition probabilities and one-dimensional Markov chains in geosciences did not well differentiate them from each other. Differentiating transition probabilities (the spatial measure) from Markov chain models (the estimator or simulator) became necessary after conditional multidimensional Markov chain models emerged in geosciences.

Our initial effort aimed to solve the flaws of the nonlinear CMC model while extending it for use in predictive soil mapping, because this model could generate convergent patterns and worked well under some special conditions (e.g., dense conditioning data and similar class proportions). The real MCRF ideas proposed by Li (2007a) owed these pioneer studies in geosciences a lot. It was also related with those studies in Markov random field models, particularly Markov mesh models, in image analysis and processing. Indeed, extending classical one-dimensional Markov chains or Markov random fields to a practical geostatistical approach is not a simple task. Both obtaining ideas to solve related problems and theoretically rationalizing the novel MCRF ideas were proved to be difficult tasks. To reach the goal, we had to absorb knowledge, ideas, and lessons related to Markov chain modeling from geosciences, statistics, and image engineering and solve a series of issues from transition probability estimation to algorithm design.

Although the term of 'Markov chain random field' came to our mind and used in our manuscripts as early as in 2004, literature search found that we were not the first to use this term. Veijanen (1993) used this term to refer to a Markov mesh random field model for image analysis. This was reasonable because Markov mesh models, which constitute a subclass of Markov random fields, were sometimes also called Markov chain models in image analysis in the past due to their unilateral nature, but this term was rarely used later. The random fields generated or defined by a multidimensional Markov chain model certainly can be called MCRFs. Some even called the simulated realizations by the transition probability geostatistics proposed by Carle and Fogg (1996) MCRFs due to its connection to transition probability. However, the MCRF model proposed by Li (2007a) is different because it is a single-chain random field model for dealing with sparse sample data. It may be regarded as an extension of Markov mesh random fields. There is no such issue called 'we assumed an underlying random field' or 'we claim we were able to build explicitly a random field.' If a Markov random field model could not define or generate a random field, there would have been no reason to call it a 'random field' model.

The MCRF approach came from a single-chain Markov random field idea, which uses one special Markov chain to do multidimensional simulations. As to how such an idea came up, see Li (2007a) and Li and Zhang (2008). The MCRF definition states that *a single spatial Markov chain moves or jumps in a space and decides its state at any uninformed location by interactions with its nearest neighbors in different directions and its last stay location*, as illustrated in figure 1 by Li (2007a). For a MCRF $Z(\mathbf{u})$, assume that $i_1(\mathbf{u}_1)$ is the state of the last stay location of the SMC, $i_2(\mathbf{u}_2)$ to $i_m(\mathbf{u}_m)$ are the states of the nearest neighbors in different directions around an uninformed location $\mathbf{u}_0$, and $\mathbf{u}_1$ may or may not be one nearest neighbor of $\mathbf{u}_0$. The conditional probability distribution of $Z(\mathbf{u})$ at the uninformed location $\mathbf{u}_0$ can be notated as $p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1), \ldots, i_m(\mathbf{u}_m)]$ based on the Markov property, where $i_0(\mathbf{u}_0)$ denotes the state of $z(\mathbf{u}_0)$ to be estimated. Emphasizing the single-chain nature of a MCRF and the last stay location, this conditional probability distribution can be factorized as

$$\begin{aligned} p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1), \ldots, i_m(\mathbf{u}_m)] &= \frac{1}{A} p[i_m(\mathbf{u}_m)|i_0(\mathbf{u}_0), \ldots, i_{m-1}(\mathbf{u}_{m-1})] \\ &\cdots p[i_2(\mathbf{u}_2)|i_0(\mathbf{u}_0), i_1(\mathbf{u}_1)] p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1)] \end{aligned} \tag{2}$$

where $A = p[i_1(\mathbf{u}_1), \ldots, i_m(\mathbf{u}_m)]/p[i_1(\mathbf{u}_1)]$ is a normalizing constant and $\mathbf{u}_1$ indicates the last stay location or the location that the SMC goes through to the current location. Please note that $p[i_m(\mathbf{u}_m), \ldots, i_2(\mathbf{u}_2)|i_0(\mathbf{u}_0), i_1(\mathbf{u}_1)] = p[i_m(\mathbf{u}_m)|i_0(\mathbf{u}_0), \ldots, i_{m-1}(\mathbf{u}_{m-1})] \cdots p[i_2(\mathbf{u}_2)|i_0(\mathbf{u}_0), i_1(\mathbf{u}_1)]$; so as long as the last stay location $\mathbf{u}_1$ and the current uninformed location $\mathbf{u}_0$ are considered together, the factorization result is the same.

Equation (2) is the *explicit full general solution* of MCRFs, which was provided as equation [6] in Li (2007a), although not specifically emphasized. This model is essentially also a multiple-point geostatistical model. The reason for the conditional probability distribution of $Z(\mathbf{u})$ to be factorized like this (i.e., with one conditional probability term and multiple likelihood terms) is due to the single-chain Markov random field idea. Otherwise, it might be regarded as irrational. Li (2007a) may be the first to use such a way to factorize a multiple-point conditional probability; now many know this also works and even works better. Similar factorization can be done if a multiple-point joint probability is used for $Z(\mathbf{u})$. The single-chain Markov random field idea seems simple, but it is indeed a result of a decade-long concentrated exploration in order to extend Markov chains to a geostatistical approach. The initial motivation for this research was to find a suitable method to simulate the spatial distributions of soil types and soil layers.

Assuming the SMC is stationary and its last stay location $\mathbf{u}_1$ is far away from the current uninformed location $\mathbf{u}_0$, one can delete $i_1(\mathbf{u}_1)$ from all of the probability terms in Equation (2) because its influence may be ignored. Then Equation (2) becomes

$$p[i_0(\mathbf{u}_0)|i_2(\mathbf{u}_2), \ldots, i_m(\mathbf{u}_m)] = \frac{1}{A} p[i_m(\mathbf{u}_m)|i_0(\mathbf{u}_0), i_2(\mathbf{u}_2), \ldots, i_{m-1}(\mathbf{u}_{m-1})] \cdots p[i_2(\mathbf{u}_2)|i_0(\mathbf{u}_0)] p[i_0(\mathbf{u}_0)] \tag{3}$$

which can be rewritten as

$$p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1), \ldots, i_m(\mathbf{u}_m)] = \frac{1}{A} p[i_m(\mathbf{u}_m)|i_0(\mathbf{u}_0), i_1(\mathbf{u}_1), \ldots, i_{m-1}(\mathbf{u}_{m-1})] \cdots p[i_1(\mathbf{u}_1)|i_0(\mathbf{u}_0)] p[i_0(\mathbf{u}_0)] \tag{4}$$

where $A = p[i_1(\mathbf{u}_1), \ldots, i_m(\mathbf{u}_m)]$ is a normalizing constant and $\mathbf{u}_1$ is just a nearest neighbor of $\mathbf{u}_0$ and the last stay location of the SMC is not included. Equation (4) is the traditional method used for factorizing a multiple-point conditional probability (similarly for a multiple-point joint probability), and it turns out to be a special case of MCRFs.

Both the Equations (2) and (4) cannot be directly estimated from sparse sample data due to the multiple-point probability terms, which are also the functions of multiple lag distances. Therefore, further simplification is necessary. If we apply the conditional independence assumption, Equation (2) becomes

$$p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1), \ldots, i_m(\mathbf{u}_m)] = \frac{1}{A} p[i_m(\mathbf{u}_m)|i_0(\mathbf{u}_0)] \cdots p[i_2(\mathbf{u}_2)|i_0(\mathbf{u}_0)] p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1)] \tag{5}$$

which, written in the transiogram style, further becomes

$$p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1), \ldots, i_m(\mathbf{u}_m)] = \frac{p_{i_1 i_0}(\mathbf{h}_{10}) \prod_{g=2}^{m} p_{i_0 i_g}(\mathbf{h}_{0g})}{\sum_{f_0=1}^{n} [p_{i_1 f_0}(\mathbf{h}_{10}) \prod_{g=2}^{m} p_{f_0 i_g}(\mathbf{h}_{0g})]} \tag{6}$$

Here $p_{i_0 i_g}(\mathbf{h}_{0g})$ represents a transition probability function, that is, a transiogram, from class i_0 to class i_g over the lag distance $\mathbf{h}_{0g,}$ which is not dependent on specific locations with the intrinsic stationarity assumption.

Similarly, after the conditional independence assumption is applied, Equation (4) finally becomes

$$p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1), \ldots, i_m(\mathbf{u}_m)] = \frac{p_{i_0} \prod_{g=1}^{m} p_{i_0 i_g}(\mathbf{h}_{0g})}{\sum_{f_0=1}^{n} [p_{f_0} \prod_{g=1}^{m} p_{f_0 i_g}(\mathbf{h}_{0g})]} \tag{7}$$

where p_{i_0} is the marginal probability or expectation of class i_0. Note that estimating p_{f_0} from a set of samples within a local window does not eliminate the stationarity assumption on p_{f_0}.

Equation (6) is the simplified *general solution* of MCRF models, provided in Li (2007a) as equation [12]. Equation (7) is a special case of Equation (6) and also can be derived from Equation (6), as explained in Li and Zhang (2012). With Equation (6), Equation (7) may not be much useful because as long as the uninformed location $\mathbf{u}_0$ has nearest neighbors we always can use one of them to serve as the last stay location by assuming that the SMC comes to the current uninformed location through the nearest neighbor. However, Equation (7) was proposed by Allard *et al.* (2011) to be a *simplified BME* (Bayesian maximum entropy) and a generalization of the MCRF approach, and it was also provided by Cao *et al.* (2011a) (i.e., equation [8] in the paper) as a simplified variant of BME and a so-called multinomial regression model. Following Allard *et al.* (2011), Cao *et al.* (2011a, 2011b) also stated that the MCRF (or SMC) model proposed by Li (2007a) was Equation (1), which could be generalized to a simplified BME, by ignoring the existence of Equations (2) and (6) in Li (2007a), and thus further messed the MCRF approach (or Markov chain geostatistics (MCG)).

Cao *et al.* (2011a) stated that the naïve Bayesian network was tantamount to the MCRF (i.e., SMC) model. This is a misunderstanding. Bayesian networks (including naïve Bayesian networks) are a kind of methods proposed for artificial intelligence or machine learning. A Bayesian network is a network of directed, acyclically connected events (or variables), in which variables represented by nodes may be observed (evidence nodes) or not (hidden or latent nodes) and may have very different natures (e.g., chair, back, and ache), and edges represent causal relations. Naïve Bayesian networks used the conditional independence assumption as noted in Li (2007a), but using such an assumption did not make a naïve Bayesian network a MCRF model. Bayesian networks do not consider spatial locations and lag distances and are rarely used as spatial models, let alone as a geostatistical model for modeling a multiclass categorical variable based on sparse sample data. If one had the imagination to adapt a naïve Bayesian network into a geostatistical model to simulate categorical spatial variables conditioned on sample data before the emergence of MCG, he/she should have made a significant progress in geostatistics. The MCRF model is a nonlinear geostatistical model for estimating the local conditional probability distribution of a categorical spatial variable based on nearest data. The relationships between an uninformed location and its nearest neighbors (informed locations) in a MCRF model are not parent–children relationships. The arrow directions illustrating a MCRF in our papers represent transition probability directions. The MCRF approach has its unique characteristics as introduced above, which make it different from other nonlinear geostatistical models such as the BME model proposed by Bogaert (2002). Although the special case of the simplified general solution of MCRFs – Equation (7) –

has some similarity in its expression to naive Bayesian networks due to the common uses of the Bayes' theorem and conditional independence assumption, they are essentially different things for different purposes. If the MCRF approach was initially adapted from or inspired by naive Bayesian networks, we would have had no reason to hide it and turn to connect our model to the CMC model of Elfeki and Dekking (2001). As to Equation (1), the wrong equation provided by Cao *et al.* (2011a), it even has no similarity in expression to naïve Bayesian networks. Here we do not discuss the rationality of the implementation method of the Tau model proposed by Cao *et al.* (2011a) and how it was inspired. However, attempting to classify both the MCRF approach and the Tau model into Bayesian networks seems difficult. There seems no ratio involved in Bayesian networks.

A MCRF is a special Markov chain and also a special causal Markov random field. Single-chain, transition probabilities (including multidimensional multiple-point transition probabilities for the full solution) and directional interactions are the basic features of MCRF models; and for the simplified general solution, the conditional independence assumption is also applied. The single-chain random field idea has proved to be unique. How to implement the MCRF ideas in algorithm design and parameter estimation, how to deal with data clustering, whether to write the model as a joint probability distribution, and whether to use the log-linear form or ratios are all topics to study.

A real exhaustive dataset (i.e., an image) or a simulated realization is usually regarded as a random field. The probability expression of a whole random field (all uninformed pixels in a finite lattice) is a theoretical random field, which can be decomposed as a general sequential algorithm like

$$p[j_1(\mathbf{u}_1),\ldots,j_N(\mathbf{u}_N)|(n)] = p[j_N(\mathbf{u}_N)|(n+N-1)]\cdots p[(j_2(\mathbf{u}_2)|(n+1)]p[j_1(\mathbf{u}_1)|(n)] \quad (8)$$

where $j_1(\mathbf{u}_1),\ldots,j_N(\mathbf{u}_N)$ represent all of the uninformed pixels to be simulated, n informed pixels (i.e., sample data), and $N+n$ the total number of pixels in the whole finite random field. n is zero if no sample data are available. The simulation of these unknowns has to follow a visiting sequence (e.g., a random path) in a simulation process, that is, one needs to first simulate $j_1(\mathbf{u}_1)$ conditioned on the n sample data, then simulate $j_2(\mathbf{u}_2)$ conditioned on the n sample data plus $j_1(\mathbf{u}_1)$, . . . , and finally simulate $j_N(\mathbf{u}_N)$ conditioned on the n sample data plus all previously simulated $N-1$ data. Such a general sequential simulation algorithm does not contain a specific estimation model, that is, it does not provide any information on how to estimate a specific unknown such as $j_k(\mathbf{u}_k)$. One has to find an estimation model (i.e., estimator), which may be linear or nonlinear, to estimate the conditional probability term $p[j_k(\mathbf{u}_k)|(n+k-1)]$, where $k-1$ is the number of previously simulated data. For the conventional sequential Gaussian simulation algorithm, the estimation model is kriging. For the conventional sequential indicator simulation algorithm, the estimation model is indicator kriging. For a MCRF sequential simulation algorithm, the estimation model must be a MCRF model. The MCRF model deals with the conditional probability $p[j_k(\mathbf{u}_k)|(n+k-1)]$, which, after the Markov assumption and the single-chain assumption are applied, becomes $p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1),\ldots,i_m(\mathbf{u}_m)]$. Apparently, the MCRF approach or MCG is not just a sequential simulation algorithm based on the trivial decomposition (8). A MCRF model is an estimation model of a multiple-point conditional probability.

In addition, other than sequential simulation, one may also use a probability estimation model to do optimal interpolation, as kriging has been traditionally used. However, optimal interpolation doses not truly generate a random field due to the smoothing effect.

2. On the transiogram spatial measure

Cao *et al.* (2011a) thought that the MCRF (i.e., SMC) model was proposed by Li (2007a) based on the concept of a transiogram. But the fact is that the concept of a transiogram was proposed by Li (2007b) purely as an accompanying spatial measure of MCG. The CMC model, which was first extended by Li *et al.* (2004) to work with soil survey line data in horizontal two dimensions, also needed this spatial measure when it was further extended for simulations conditioned on regular sample data. If it was not because our Markov chain methods needed such a spatial measure to conveniently work with sparse sample data, there would have been no reason for Li to propose such a concept. On the contrary, merely with such a concept, one may not be able to develop a reasonable Markov chain model for conditional simulation.

The term of 'transiogram' is an abbreviation of transition probability diagram and is also an analog to the 'variogram' term used in classical geostatistics. A transiogram theoretically refers to a transition probability function over the lag distance. The purpose of suggesting such a simple name was merely to provide a convenient, nonconfusing term for the transition probability spatial measures (Li 2007b). As one can see that when a Markov chain or Markov random field model is extended into a geostatistical model and written in the form of Equation (6) – the simplified general solution of MCRFs – the transiogram naturally emerges as a two-point spatial measure and is also an indispensable spatial measure for MCG (Li and Zhang 2008). We were aware that visual transition probability diagrams and transition probability functions over lag distances had been used by earlier researchers in geology with different names, as noted in Li (2007b) – the first full paper to address the related concepts and estimation and modeling issues of transiograms. For our knowledge, a diagram of multistep transition probabilities calculated from a one-step transition probability matrix first appeared in Schwarzacher (1969); Lou (1993, 1996) used the notation of transition probability functions and extensively demonstrated such kind of transition probability diagrams for geological structure analysis, and called them 'Markov diagrams.' Carle and Fogg (1996) used transition probability functions as spatial measures to reformulate indicator kriging/co-kriging systems, and they also proposed using transition rate-derived models (called 'continuous-lag Markov chain models') to jointly fit the Markov diagrams estimated from borehole data (Carle and Fogg 1997). The pioneer studies, particularly Carle and Fogg (1996, 1997), were important in pushing transition probabilities into a spatial structure measure. However, both Markov diagram and transition probability diagram have been often used to refer to the schematic diagrams of Markov chain models. This situation necessitates a new specific term for the lag-based transition probability spatial measures. The transition rate-derived transition probability models were found to be identical with idealized transition probability diagrams calculated from one-step transition probability matrix, and they both were regarded as 'idealized transiograms' by Li (2007b). Li (2007a) and Li and Zhang (2008) extended Markov chains into a theoretically sound practical geostatistical approach for conditional simulation on sample data and consequently established the transiogram as the spatial measure of MCG. Because the transiogram is indispensable to MCG, the estimation and modeling, particularly joint modeling, of transiograms become crucial issues of MCG, and we therefore extensively explored these issues in recent years. We are the first to estimate transiograms from sparse point sample data, and we also proposed two practical transiogram joint modeling methods and a set of transiogram models (Li 2007b, Li and Zhang 2010, Li *et al*. 2012).

Cao *et al.* (2011a) stated that 'an exponential form of transiograms was also discussed in Li (2006), in the absence, however, of a clear underlying theory.' In fact, as a technique

note, Li (2006) only talked about the basic concepts about transiograms and their meanings and demonstrated several typical shapes of idealized and experimental transiograms. The major purpose of this short article was to propose the use of transiograms as an independent spatial measure for characterizing the intra- and inter-class relationships of categorical spatial variables. An extensive introduction of transiograms with examples was provided by Li (2007b). There is no other complex underlying theory related to a transiogram itself, except for the basic properties of transition probabilities and stationary first-order Markov chain transition probability matrix, as presented in Li (2007b). The transiogram spatial measure emerged because it was needed in Markov chain or transition probability-based spatial models for conditional simulations. How to estimate transiogram models from expert or geological knowledge reasonably is another issue, which was first explored by Carle and Fogg (1997).

3. On conditional independence assumption

The conditional independence assumption was used in Li (2007a) to obtain a simplified general solution of MCRFs. This assumption is just a practical assumption and has been often used in statistical models. It apparently does not hold for all situations but its practicality was proved by many application cases (e.g., Besag 1986). A very early application of this assumption can be found in Chow (1962) where adjacent neighbors in cardinal directions were assumed to be conditionally independent in recognition network. However, under some special conditions such as in Pickard random fields (PRFs), this assumption is true for adjacent neighbors in cardinal directions. Li (2007a) found that this assumption also can be regarded as proper for nearest neighbors in cardinal directions in a sparse data space if the unknown real field is assumed to be a PRF. This does not mean that such an assumption is always true for real-world data. The conditional independence of adjacent neighbors in cardinal directions for a rectangular lattice is a property of PRFs (Pickard 1980, Haslett 1985, Besag 1986, Rosholm 1997), rather than a theoretically proved common property of real-world datasets.

Verhagen (1977) proposed a special two-state process model in order to help explain the stochastic behavior of a hard-core lattice gas with on/off states in chemical physics. This work inspired Pickard to give the proposed new process a rigorous theoretical treatment in Pickard (1977, 1980). The associated class of random fields was later named after him as a tribute to his theoretical contributions. The original description of PRFs was mainly concerned with a binary process, but the model is valid for both the continuous and multicategorical phenomena. Haslett (1985) should be credited as the first to use this type of random fields intelligently in a classification context. A systematical introduction of PRFs and its properties can be found in Rosholm (1997). PRFs have found many applications in image and signal processing during the last two decades.

A PRF possesses the property of being simultaneously a Markov random field and a so-called Markov mesh (Besag 1986, Rosholm 1997) on a finite rectangular lattice. Besides this property, a PRF has the following four important characteristic properties: (i) a PRF is a homogeneous (i.e., stationary) second-order Markov random field with respect to a neighborhood of eight nearest neighbors; (ii) the distribution of a PRF is uniquely determined by the marginal distribution of any 2×2 subset in the image; (iii) the four adjacent neighbors (or four adjacent pixel blocks along a cross) of a given pixel are conditionally independent given the value of the surrounded central pixel; and (iv) a PRF contains embedded Markov chains along every monotone path in an image lattice (Haslett 1985, Besag 1986, Rosholm 1997). A MCRF does not necessarily have all of these properties because it is a generalized

single-chain-based random field and may not be stationary. However, for the situation of the four (or less) nearest neighbors found in four cardinal directions by assuming a rectangular lattice, the conditional independence property of a PRF is applicable to the sparse data space (Li 2007a, Li and Zhang 2008). This finding can strongly support the neighborhood choice of using four nearest neighbors in four cardinal directions or quadrants in MCRF algorithm design, as done by Li and Zhang (2007). Such neighborhood design may avoid the difficulty in coping with data clustering effect. Apparently, simply ignoring data clustering effect is not preferable unless it is inevitable. Therefore, regarding the PRF theory as the theoretical foundation of MCRFs is a misunderstanding.

Cao *et al.* (2011a) thought that the Tau model proposed by Journel (2002) based on the permanence of ratios implied conditional independence, and therefore, as a special case of the Tau model (when $\tau_n = 1$), the permanence of ratios is a general form of conditional independence. This viewpoint seems arguable. First, the permanence of ratios assumes that ratios of information increments are typically more stable than the increments themselves. This is not a conditional independence assumption of data events. Using D_1 and D_2 to represent two data events and using A to represent the event to be inferred (Cao *et al*. 2011a), the permanence of ratios amounts to assuming

$$\frac{1-P(A|D_1,D_2)}{P(A|D_1,D_2)}\Big/\frac{1-P(A|D_1)}{P(A|D_1)} \approx \frac{1-P(A|D_2)}{P(A|D_2)}\Big/\frac{1-P(A)}{P(A)} \tag{9}$$

This assumption means that 'the incremental contribution of data event D_2 to knowledge of A is the same after or before knowing D_1,' as stated by Journel (2002). It seems difficult to say that the data independence implied in Equation (9) or in the statement is the conditional independence between D_2 and D_1 given A. This equation seems containing some ingredient of the full independence between A and D_1, but using ratios may largely reduce the negative effect. Journel (2002) suggested to add a power parameter (i.e., the tau parameter) on the right-hand side of Equation (9) to reduce the effect of ignoring D_1 after D_2 appears. Second, Journel (2002) opposed the uses of both the full independence and the conditional independence assumptions in dealing with the dependence of data events. The data independence in Equation (9) is neither full nor conditional independence (Journel 2002). Here, we are not saying the Tau model is not a good model, but it seems difficult to say the permanence of ratios to be a general form of conditional independence as asserted by Cao *et al.* (2011a).

It is true that Journel (2002) demonstrated that the permanence of ratios hypothesis appeared more robust than the full independence and conditional independence hypotheses and that the latter two might not guarantee some limit conditions. However, the article did not obtain a suitable equation for examining conditional independence, because the law of total probability with conditional independence was not applied to the denominator (i.e., the marginal probability) in deriving equation [15] for examining the conditional independence hypothesis. This may not impact the permanence of ratios and the Tau model but may lead to unsafe conclusions. Because of the difficulty in parameter estimation, Journel (2002) did not implement the Tau model.

Both the full independence and conditional independence assumptions can be practical if used in suitable places. Full independence is traditionally used in multivariate linear regression models, where predictor variables are assumed to be fully independent on each other. This gives multivariate linear regression the power to incorporate a number of different variables, which cannot be implemented practically in co-kriging. Conditional independence is often assumed in nonlinear models, such as Bayesian inference models

and Markov random field models, and is often found to be practical. When conditional independence is assumed, the law of total probability has to be applied to the marginal probability (e.g., $P(D_1, D_2)$ at here) simultaneously with the same assumption and there will be no violation of limit properties. However to seek optimality, the data clustering effect (or redundancy) is still an issue to be considered; after all, statistical assumptions and the properties of random processes may deviate from the properties of real-world datasets.

4. On Bayesian inference and the maximum entropy principle

Recently, D'Or *et al.* (2008) and Allard *et al.* (2009, 2011) suggested an efficient method for modeling categorical spatial variables. It was first called 'multinomial regression' although it had nothing to do with regression, then called 'multinomial categorical simulation,' and finally called 'Markovian-type categorical prediction' (MCP). The MCP model was claimed to be a simplified BME and a generalization of the MCRF approach. The MCRF ideas were clearly presented in Li (2007a) and Li and Zhang (2008) with introductions on why the ideas came up. However, Allard *et al.* (2011) not only ignored the contents of these two papers but also the MCRF model was wrongly written and misinterpreted as a two-chain model. The so-called MCP model is actually a special case of the simplified general solution of MCRFs, that is, Equation (7) in this comment. The neighborhood structure used in the MCP algorithm is also an inferior one because it ignored the effect of data clustering in conditional simulation, as pointed out by Li and Zhang (2012). Following D'Or *et al.* and Allard *et al*., Cao *et al.* (2011a, 2011b) also made misinterpretations on the MCRF approach and stated that the so-called simplified BME was a generalization of the MCRF approach. Cao *et al.* (2011a) provided a comparison study between the Tau model and the SMC model. The SMC model implemented by Cao *et al*. (2011a) should be the so-called MCP method of Allard *et al.* (2011) rather than Equation (1) because the extended CMC model still underestimates minor classes when samples are not dense enough. Cao *et al*. (2011b) conducted a comparison study between a so-called multinomial logistic mixed model (a variant or equivalent of the simplified BME) and the simplified BME (i.e., the so-called MCP method, which was claimed to be a generalization of the MCRF approach but is essentially a special case of the simplified general solution of MCRFs). Li and Zhang (2012) commented on Allard *et al*. (2011) and pointed out their misunderstandings. However, they stated they deliberately wrote the wrong equation and then they raised some new issues to question the MCRF approach in their response letter. We have proven above that those new issues raised by them were also misunderstandings, probably caused by their lack of knowledge in Markov random fields and sequential simulations.

To further clarify these confusions and misinterpretations, we need to clarify the meanings of BME. BME is essentially a combination of the Bayesian inference principle and the maximum entropy inference principle (i.e., the principle of maximum entropy). These two are very general statistical principles that one may consider when constructing a probability model, no matter whether the model is spatial or not. For our knowledge, the Bayesian inference principle means that when one builds a probability model the model construction should follow the style of

$$posterior \propto likelihood \times prior$$

where the *posterior* probability represents an update of the *prior* probability based on the evidence (e.g., observed data) through the *likelihood*. If the Bayes' theorem or simply the

quantitative relationship between joint probability and conditional probability is used in the derivation of a probability model, the probability model is possible to be in accordance with this principle. The maximum entropy principle means that when one builds a probability model for inferring a specific variable he/she should choose the one that has the maximum information entropy, subject to a set of constraints that represent our incomplete information about the target distribution (Jaynes 1957). Although arguments still exist (between frequentists and Bayesians objective Bayesians and subjective Bayesians), these two principles have been promoted and widely used in practice in recent years. A probability model constructed adhering to these two principles may be called a BME model.

However, we think at least the following three points regarding the two principles should be kept in mind. First, there is no unique operable and widely applicable BME, maximum entropy, or Bayesian model for everything. On the one hand, many specific probability models for similar or different purposes have been proposed with adherence to these two principles or one of them in recent years. On the other hand, many existing probability models constructed without the knowledge of or concern with these two principles were later found to be in accordance with these two principles or one of them; this should also be the reason why these two principles were established. For example, the Gibbs, the Gaussian, and the uniform distributions were all found to be a maximum entropy distribution. As the Gibbs distribution function is the general probability function characterizing Markov random fields, some Markov random field models may be in accordance with the maximum entropy principle. Second, any practical maximum entropy probability model is subject to a set of specific constraints. These constraints may be related to statistical assumptions, summary statistics, domain knowledge, target variable properties, local environmental conditions, expert experience, and/or new findings. Applying different constraints will generate different maximum entropy models. Third, without an in-depth study on a specific topic, one may not be able to build a practical model simply with the two principles, because these two principles do not provide a specific solution for a specific issue. For example, how to solve a complex likelihood function, how to set a prior probability, and how to measure and estimate the spatial dependence of a set of data or variables are tasks for modelers to deal with.

Now let us check the MCRF models again. Looking at Equation (2) – the explicit full solution of MCRFs – one can find it is in accordance with the Bayesian inference formulation: $p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1), \ldots, i_m(\mathbf{u}_m)]$ is the posterior probability distribution; $p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1)]$ is the prior probability distribution; and the other terms in the right-hand side except for the constant constitute the likelihood part. Looking at Equation (4) – the special stationary case of MCRFs – one can find it is similar: $p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1), \ldots, i_m(\mathbf{u}_m)]$ is still the posterior probability distribution; $p[i_0(\mathbf{u}_0)]$ becomes the prior probability distribution; and the other terms in the right-hand side except for the constant are the likelihood part. Similarly, Equations (6) and (7) – the simplified general solution of MCRFs and its special case – are also in accordance with the Bayesian inference formulation. Therefore, a MCRF model also can be explained from the viewpoint of Bayesian inference: A spatial Markov chain moves or jumps in a space and at any uninformed location its probability distribution is decided by its state at its last stay location and the new evidence (i.e., nearest data) around the current uninformed location. If no nearest data is available other than its last stay location, its probability distribution at the current location is solely decided by the state at its last stay location, that is, a transition probability vector from the last stay location (with informed state) to the current location (with uninformed state), which is the prior probability distribution; but if nearest data around the current location (other than the last

stay location) is available as new evidence, its probability distribution at the current location will become the updated form of the prior probability distribution based on the new evidence through likelihood functions.

It is possible that some MCRF or MCRF-based models are in accordance with the maximum entropy principle, of course, subject to the constraints of the specific models. However, even with this it is still not proper to say MCRF models to be simplified forms of the BME model proposed by Bogaert (2002), because (i) they came from different ideas and (ii) the basic characteristics of MCRFs – the single-chain property, the Markov property, and the directional interactions through transition probabilities – are not the basic characteristics of that BME model. The conditional independence assumption, which was used to simplify the MCRF full solution, was also not used in that BME model.

5. Final remarks

When we began to explore how to use a two-dimensional Markov chain model to simulate the spatial distributions of soil types and layers more than a decade ago, we did not have much knowledge in spatial statistics. In order to solve the problems encountered in two-dimensional Markov chain conditional simulation of categorical spatial variables, we spent many years to learn related knowledge and investigate earlier literature in Markov chain modeling in geosciences and in Markov random field (particularly Markov mesh) simulations in image analysis and processing. The nonlinear multidimensional Markov chain simulation approach for categorical spatial variables has been evolving almost in parallel with the multiple-point geostatistics and the BME method for categorical spatial variables. Although the initial ideas and justifications may be different, these three different non-kriging nonlinear geostatistical approaches actually share some similar purposes and have overlapped or probably will overlap their functions in some respects with the growth and extension of each. Thus, it should be normal that one may find some plausible similarities among them and some connections with other methods for categorical spatial variables such as indicator kriging and Markov random fields. Many statistical or probability models, whether they are spatial or not, often turn out to have some relations with each other even if they were developed separately, due to the explicit or implicit common assumptions and principles followed intentionally or accidentally. MCRF models were found to be accidentally in accordance with the Bayesian inference formulation after the MCRF approach was proposed. Alternative methods should be able to coexist. We should have an objective attitude to treat different approaches and their possible connections.

Christakos (1990, 2000) suggested the BME framework (or paradigm) for geostatistical modeling and provided example models for modeling continuous variables. Under this framework, Bogaert (2002) proposed a spatial model for modeling categorical spatial variables by following the maximum entropy principle and it was also named as BME model. Kriging variants were sometimes also regarded as Bayesian models or special cases of BME. We recognized all of those well-established geostatistical models or frameworks in our publications and also recognized those pioneer studies related to our Markov chain method even if their models were not practical or were rarely used. It was widely known in geosciences that simulating Markov chains in two or three dimensions was very difficult. However, multidimensional Markov chain modeling has its long root in geosciences. The MCRF approach is essentially in alignment with those early two- or three-dimensional Markov chain models in geosciences, which mainly include the models of Krumbein (1968), Lin and Harbaugh (1984), Lou (1993, 1996), and Elfeki and Dekking (2001). Carle and Fogg (1996, 1997) are related to this line but their simulation method is essentially an

extension of indicator kriging. The evolution processes along this line are from multiple-chain models to single-chain models, from linear models to nonlinear models, and from unconditional simulation models to conditional simulation models. This sounds lowborn due to the theoretical flaws of some intermediate models, but it is very reasonable for us when we started this research topic from soil-type simulation more than 10 years ago (i.e., 1998). However, in order to solve problems and seek theoretical supports, we have to turn to Markov random fields. Therefore, the MCRF approach is closely related to Markov random fields (mainly Markov meshes), which were regarded as the natural multi-dimensional extension of one-dimensional Markov chains in statistics. Of course, the MCRF approach is also connected with classical geostatistics. We think the MCRF approach represents a jump both in multidimensional Markov chain models and in practicality because of the unique single-chain Markov random field idea and the support of existing Markov random field theories.

The basic framework of MCG was presented with some implementation methods in recent years, but development was very limited. Extension toward spatiotemporal models with data and knowledge from diverse sources should be on the way. However, considering our difficult situations in these years, we know what we can do is very limited in developing the MCRF approach or MCG. Other researchers are welcome to join in its development or adopt the features of this approach for their purposes, and we also welcome statisticians and mathematicians to provide a rigorous treatment to the MCRF idea if they have the interest. Nowadays in an era of internet, one may easily look across the field of 'geostatistics' to see more similar ideas in other disciplines, including those researches conducted very early. However, finding some plausible connections or similarities in some aspects of an existing geostatistical approach with other statistical approaches or principles should not be a reason to doubt the originality of the existing approach or deny its contributions, let alone with wrong equation citations and misinterpretations in disregard of the contents of the core papers establishing the approach. It is obvious that the MCRF approach has contributed to geostatistics and inspired some related researches in categorical spatial variable modeling. Even if the MCRF geostatistical approach were adapted from an existing non-geostatistical method, that still would have erected a new geostatistical approach and made a contribution to geostatistics. The facts are that the single-chain feature and the general solution forms of MCRFs are unique, and no similar geostatistical methods existed before ours.

References

Allard, D., D'Or, D., and Froidevaux, R., 2009. *Estimating and simulating spatial categorical data using an efficient maximum entropy approach*. Avignon, France: Unit´e Biostatistique et Processus Spatiaux Institut National de la Recherche Agronomique, Technical Report No. 37.

Allard, D., D'Or, D., and Froidevaux, R., 2011. An efficient maximum entropy approach for categorical variable prediction. *European Journal of Soil Science*, 62, 381–393.

Besag, J., 1986. On the statistical analysis of dirty pictures (with discussions). *Journal of the Royal Statistical Society, Series B*, 48, 259–302.

Bogaert, P., 2002. Spatial prediction of categorical variables: the Bayesian maximum entropy approach. *Stochastic Environmental Research and Risk Assessment*, 16, 425–448.

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

Cao, G., Kyriakidis, P.C., and Goodchild, M.F., 2011b. A multinomial logistic mixed model for the prediction of categorical spatial data. *International Journal of Geographical Information Science*, 25, 2071–2086.

Carle, S.F. and Fogg, G.E., 1996. Transition probability-based indicator geostatistics. *Mathematical Geology*, 28, 453–477.

Carle, S.F. and Fogg, G.E., 1997. Modeling spatial variability with one- and multi-dimensional continuous Markov chains. *Mathematical Geology*, 29, 891–918.
Chow, C.K., 1962. A recognition method using neighbor dependence. *IRE Transactions on Electronic Computers*, 11, 683–690.
Christakos, G., 1990. A Bayesian/maximum-entropy view to the spatial estimation problem. *Mathematical Geology*, 22, 763–777.
Christakos, G., 2000. *Modern spatiotemporal geostatistics*. New York: Oxford University Press, 312.
D'Or, D., *et al*., 2008. Simulating categorical random fields using the multinomial regression approach. *In*: J. Ortiz and X. Emery, eds., *Proceedings of the eighth international geostatistics congress*. Santiago de Chile: Gecamin Ltd, 59–68.
Elfeki, A.M. and Dekking, F.M., 2001. A Markov chain model for subsurface characterization: theory and applications. *Mathematical Geology*, 33, 569–589.
Elfeki, A.M. and Dekking, F.M., 2005. Modeling subsurface heterogeneity by coupled Markov chains: directional dependency, Walther's law and entropy. *Geotechnical and Geological Engineering*, 23, 721–756.
Haslett, J., 1985. Maximum likelihood discriminant analysis on the plane using a Markovian model of spatial context. *Pattern Recognition*, 18, 287–296.
Jaynes, E.T., 1957. Information theory and statistical mechanics. *Physical Review Series II*, 106, 620–630.
Journel, A.G., 2002. Combining knowledge from diverse sources: an alternative to traditional data independence hypothesis. *Mathematical Geology*, 34, 573–596.
Krumbein, W.C., 1968. *FORTRAN IV computer program for simulation of transgression and regression with continuous time Markov models*. Computer Contribution 26, Kansas Geology Survey.
Li, W., 2006. Transiogram: a spatial relationship measure for categorical data. *International Journal of Geographical Information Science*, 20, 693–699.
Li, W., 2007a. Markov chain random fields for estimation of categorical variables. *Mathematical Geology*, 39, 321–335.
Li, W., 2007b. Transiograms for characterizing spatial variability of soil classes. *Soil Science Society of American Journal*, 71, 881–893.
Li, W. and Zhang, C., 2006. A generalized Markov chain approach for conditional simulation of categorical variables from grid samples. *Transactions in GIS*, 10, 651–669.
Li, W. and Zhang, C., 2007. A random-path Markov chain algorithm for simulating categorical soil variables from random point samples. *Soil Science Society of American Journal*, 71, 656–668.
Li, W. and Zhang, C., 2008. A single-chain-based multidimensional Markov chain model for subsurface characterization. *Environmental and Ecological Statistics*, 15, 157–174.
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, 821–839.
Li, W. and Zhang, C., 2012. Comments on 'An efficient maximum entropy approach for categorical variable prediction' by D. Allard, D. D'Or & R. Froidevaux. *European Journal of Soil Science*, 63, 120–124.
Li, W., Zhang, C., and Dey, D.K., 2012. Modeling experimental cross transiograms of neighboring landscape categories with the gamma distribution. *International Journal of Geographical Information Science*. doi:10.1080/13658816.2011.603336.
Li, W., *et al*., 2004. Two-dimensional Markov chain simulation of soil type spatial distribution. *Soil Science Society of American Journal*, 68, 1479–1490.
Lin, C. and Harbaugh, J.W., 1984. *Graphic display of two- and three-dimensional Markov computer models in geology*. New York: Van Nostrand Reinhold Company.
Lou, J., 1993. Konditionale Markovsimulation 2-dimensionaler geologischer probleme. *Berliner Geowissenschaftliche Abhandlungen, Reihe D*, Band 4, 103.
Lou, J., 1996. Transition probability approach to statistical analysis of spatial qualitative variables in geology. *In*: A. Forster and D.F. Merriam, eds., *Geologic modeling and mapping*. New York: Plenum Press, 281–299.
Pickard, D.K., 1977. A curious binary lattice process. *Journal of Applied Probability*, 14, 717–731.
Pickard, D.K., 1980. Unilateral Markov fields. *Advances in Applied Probability*, 12, 655–671.
Rosholm, A., 1997. *Statistical methods for segmentation and classification of images*. Dissertation (PhD). Technical University of Denmark, Lyngby, 187.

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

Veijanen, A., 1993. Contextual estimators of mixing probabilities for Markov chain random fields. *Pattern Recognition*, 26, 763–769.

Verhagen, A.M.W., 1977. A three parameter isotropic distribution of atoms and the hard-core lattice gas. *The Journal of Chemical Physics*, 67, 5060–5065.