European Journal of Soil Science

European Journal of Soil Science, February 2012, **63**, 120–124 doi: 10.1111/j.1365-2389.2011.01413.x

Letter to the Editor

Comments on 'An efficient maximum entropy approach for categorical variable prediction' by D. Allard, D. D'Or & R. Froidevaux

Recently, Allard *et al*. (2011) published an interesting paper in which they proposed a Markov type Categorical Prediction (MCP) approach for simulating categorical variables. They found that the MCP solution might be regarded as an approximation to the full BME (Bayesian Maximum Entropy) solution derived by Bogaert (2002) and D'Or & Bogaert (2004). Then they suggested that the MCP approach was a generalization of the Markov chain random field (MCRF) approach, in other words, the Markov chain geostatistics (MCG) proposed by Li (2007). However, contrary to their understanding, we knew that the solution was just a special case of the MCRF general solution when the MCRF general solution was first obtained. In Allard *et al*. (2011) there are several incorrect statements and citations referring to the MCRF theory by Li (2007) and the MCRF sequential simulation algorithm by Li & Zhang (2007), which can be misleading. The following comments aim to clarify the misunderstandings.

Incorrect statements

Allard *et al*. (2011) state 'More recently a Markov chain random field approach was proposed by Li (2007) with applications to the simulation of soil type spatial distribution (Li *et al*., 2004; Li & Zhang, 2007) but the conditioning is limited to only four neighbours organized in orthogonal directions. This algorithm relies on a conditional independence assumption between two Markov chains, but its optimality properties are not known'. Later, they also state 'For the estimation and simulation of spatial categorical variables, Li (2007), Li *et al*. (2004) and Li & Zhang (2007) proposed the combination of two orthogonal one-dimensional Markov chains, P^1 and P^2, with a conditional independence assumption between the two chains ....'.

These are misunderstandings of the MCRF theory and MCG. To clarify these, we raise and answer some questions.

Is the MCRF approach composed of two Markov chains or built on the coupled Markov chain model?

Li (2007) states 'This paper proposes a single-chain based Markov chain random field (MCRF) approach for building one to multi-D Markov chain models for conditional simulation or interpolation' and that 'the suggested MCRF contains only one single Markov chain'. This is repeated many times in this paper and Figure 1 in Li (2007) illustrated what a MCRF looks like in a space. In addition, Li usually called the MCRF approach 'Markov chain geostatistics' (MCG), which include different MCRF-based algorithms, transiograms and the MCRF theory. However, Allard *et al*. state that the MCRF algorithm 'relies on a conditional independence assumption between two Markov chains' and that Li and his co-authors 'proposed the combination of two orthogonal one-dimensional Markov chain P^1 and P^2, with a conditional independence assumption between the two chains'.

Although the development of the MCRF theory and approach started earlier than 2007, the paper of Li *et al*. (2004) was definitely not a part of the MCRF approach. Their method was built on the Coupled Markov Chain (CMC) model of Elfeki (1996) and aimed to correct some of its deficiencies. Li (2007) and Li & Zhang (2008) described the evolution of their research from coupled Markov chains to a single chain. However, Allard *et al*. considered Li (2007), Li *et al*. (2004) and Li & Zhang (2007) together and came to a wrong conclusion.

The proposition for the use of the MCRF theory and MCG resulted from research on Markov chain modelling of soil categories by Li and his colleagues using the CMC model proposed by Elfeki (1996) as a starting point. That is because earlier postdoctoral research by Li applied the CMC model to the simulation of categorical soil variables and found that the model had some apparent deficiencies. The MCRF idea resulted from the process of first attempting to correct the deficiencies of the CMC model and then seeking a new way to build theoretically sound multi-dimensional Markov chain models; however, it was thoroughly different from the CMC model.

The CMC model of Elfeki (1996) comprised a pair of independent Markov chains, but there was no conditional independent assumption within it. To couple the two Markov chains, Elfeki (1996) assumed them to be (absolutely) independent and excluded unwanted transitions, which were explained and demonstrated in Elfeki & Dekking (2001) as noted by Li (2007). Li (2007) pointed out the deficiencies of the CMC model, and also demonstrated the problem of minor class under-estimation in simulated realizations (actually in local conditional probability distribution estimation), which was difficult to solve within the CMC framework. He then proposed an innovative single-chain MCRF model, which, as a single Markov chain, does not need an absolute independent assumption and also contains no unwanted transitions.

The conditional independent assumption was used both in Li (2007) and in Allard *et al*. (2011), but the purpose was not to deal with two Markov chains. On the contrary, it was used to simplify the solutions so that they were computable. In the publications of Li and his colleagues, two Markov chains and the conditional independence assumption did not co-exist in one model.

Is the MCRF approach equal to a simulation algorithm?

The MCRF theory essentially provides the foundation of MCG and can support a variety of simulation algorithms. However, Allard *et al*. (2011) note that the conditioning of the MCRF approach 'is limited to only four neighbours organized in orthogonal directions'.

The MCRF approach did not impose such a limitation on simulation algorithm design. Li (2007) not only provided a general solution of MCRFs, but also pointed out that 'Development of practical simulation algorithms will be the key for making full use of the MCRF theory', that (i) 'the so-called cardinal directions are not limited to the exact four axial directions'; (ii) 'It is important to note that the objective of this paper is to define the theoretical basis of the MCRF, not to provide practical implementation techniques. Specific algorithms and implementations of MCRF-based Markov chain models will be provided in future papers or depend on future development of new algorithms' and (iii) 'With a conservative attitude, it is suggested herein that if the nearest known neighbours are generally located uniformly throughout the study space and a random path is used in a simulation, inclusion of nearest known neighbours in non-cardinal directions may be considered'. Li & Zhang (2007) also pointed out that 'In addition, the MCRF theory is also flexible in cardinal directions; three or more than four cardinal directions may be considered in algorithm design'.

In fact, multiple simulation algorithms have been suggested and used for Markov chain modelling by Li *et al*., including two fixed-path algorithms of the alternate advancing path and the middle insertion path for dealing with regular or borehole sample data, and one random-path algorithm, which used quadrant search for dealing with sparse point sample data. Although these simulation algorithms considered four cardinal directions or sectors, it does not mean the MCRF approach cannot have other algorithms.

Why was a quadrant search algorithm suggested for MCRF modelling?

It seems that Allard *et al*. were much influenced by using 'four neighbours organized in orthogonal directions' in the MCRF quadrant search algorithm and note that 'its optimality properties are not known'. Indeed, so far all simulation algorithms that we have published consider only neighbours in four cardinal directions or quadrants. However, these propositions were related to the evolving process of the multi-dimensional Markov chain simulation in geosciences and have rational backgrounds.

First, there is a historical reason. The CMC model, Lou's (1996) model and some Markov mesh models (Gray *et al*., 1994) all considered only neighbours in orthogonal directions on a lattice and used fixed paths. It is understandable that every scientific technology has a step-by-step evolutionary process from initial simple or even incorrect ideas, as demonstrated in our publications on Markov chain modelling since 2004 and by others before 2004 in multi-dimensional Markov chain modelling in geology. Because our research in two-dimensional Markov chain modelling started from the attempt to correct the deficiencies of the CMC model, we had to solve related problems one by one. That is why we first used fixed path algorithms and then developed a random path algorithm.

A second reason is because Li (2007) found that nearest neighbours in cardinal directions could be regarded as conditionally independent in a sparse data space on the basis of the theorem of Pickard (1980). With such a theoretical support, considering neighbours in four cardinal directions in an algorithm (by assuming a rectangle lattice) may be a good choice in algorithm design. However, Li (2007) did not constrain the algorithm design to four neighbours in four orthogonal directions, as explained above.

Third, it is a practical algorithm, although we have not shown any test on the optimality of the quadrant search random-path MCRF sequential simulation algorithm in published papers so far. However, except for the above second reason, quadrant search also represents a de-clustering method. Quadrant search in the random path MCRF simulation algorithm also had a data de-clustering purpose. Choice of a number of neighbours in a neighbourhood without the consideration of directions was tested in unpublished MCRF simulations years ago, but was discarded because it was inferior. Allard *et al*. (2011) also note that it was very easy to code such an algorithm by using their MCP solution.

It seems that Allard *et al*. have ignored the facts that the tests in D'Or & Bogaert (2004) (Figure 5 in their paper) and in Brus *et al*. (2008) (Figure 3 in their paper) in the BME modelling of categorical soil variables both showed that considering four data in the neighbourhood could be a good choice. Brus *et al*. (2008) actually chose the neighbourhood size of four in their simulation algorithm.

Incorrect equation citation

Allard *et al*. (2011) also note 'the prediction equation obtained by Li (2007, equation 12)' as

$$P^{LZ}_{i_0|i_1,\ldots,i_4} = \frac{P^{d_1}_{i_0|i_1}(h_1)\prod_{k=2}^{4} P^{d_k}_{i_k|i_0}(h_k)}{\sum_{i_0=1}^{I} P^{d_1}_{i_0|i_1}(h_1)\prod_{k=2}^{4} P^{d_k}_{i_k|i_0}(h_k)}, \qquad (1)$$

and state that 'where $P^{d_i}(.)$ is obtained from transition probabilities of one of the two Markov chains, depending upon their directions'.

There is no such equation in Li (2007) (or in any other paper). Equation (12) in Li (2007) which is the simplified general solution of MCRFs, and copied here as:

$$\begin{aligned}\Pr(X(x) = k|X(x_1) = l_1, \cdots, X(x_m) = l_m) \\ = \frac{\prod_{i=2}^{m} p^{i}_{kl_i}(h_i)\cdot p^{1}_{l_1 k}(h_1)}{\sum_{f=1}^{n}\left[\prod_{i=2}^{m} p^{i}_{fl_i}(h_i)\cdot p^{1}_{l_1 f}(h_1)\right]}.\end{aligned} \qquad (2)$$

This general solution is based on a single spatial Markov chain, irrelevant to two or coupled Markov chains. It does not determine the number of nearest neighbours, nor provide a direction on the spatial Markov chain. Because the MCRF theory assumes that a spatial Markov chain may move or jump to any place in a space randomly or along prescribed paths, it is impossible that the general solution (that is, the general form of the local conditional

probability distribution) has a direction of *LZ*. Note that this general solution is a directly computable form, simplified using the conditional independence assumption. The full solution of MCRFs involves complex multiple-point statistics and is not directly computable based on sample data.

The notations used in Equation (2) may not be perfect, because a specific superscript notation for directions of transition probabilities was not necessary. Because of this, the superscript notation for directions was deleted in MCRF models (Li *et al*., 2010) and in Li & Zhang (2011) the general solution was written as

$$\Pr(z(\mathbf{u}) = k|z(\mathbf{u}_1) = l_1, \cdots, z(\mathbf{u}_m) = l_m) = \frac{p_{l_1 k}(\mathbf{h}_1) \prod_{g=2}^{m} p_{kl_g}(\mathbf{h}_g)}{\sum_{f=1}^{n} \left[p_{l_1 f}(\mathbf{h}_1) \prod_{g=2}^{m} p_{fl_g}(\mathbf{h}_g) \right]}. \quad (3)$$

However, the meaning is the same, because $\mathbf{h}g$ is a vector with a direction denoted by g.

Incorrect judgement

Allard *et al*. (2011) state 'We also show that the Markov Chain Random Field (Li, 2007) is a special case of our general result . . .'. They note that 'Replacing $P^d_{i0|i_1}(h_1)$ by $P^d_{i_1|i_0}(h_1)p_{i_0}/p_{i_1}$ yields, after simplification by p_{i_1}, to an equation formally similar to the prediction equation (7)'. However, things are not so simple. If the MCRF general solution was really composed of two Markov chains, such a replacement could not do this. If the MCRF general solution was correctly written as the present Equation (2) or (3), replacing $p_{i0|i_1}(\mathbf{h}_{10})$ with $p_{i_1|i_0}(\mathbf{h}_{01})p_{i_0}/p_{i_1}$ would result in an equation that is the same as, rather than formally similar to, the prediction equation (7) in Allard *et al*. (2011). However, to make such a replacement one needs to first make some assumptions.

Allard *et al*. (2011) further state that 'Our method is thus a generalization of the Markov chain random field approach, allowing us to consider as many neighbours in as many directions as desired'. Such a judgement, plus the wrong equation citation, denied the generality of the MCRF theory and played down the MCRF approach. Because Allard *et al*. had claimed that the MCP solution was an approximation to the full BME solution, the MCRF approach thus became a special case of an approximation of the BME approach for categorical variables. We think that this is not true.

First we check the solutions, the general models, to see whether or not the MCP method is a generalization of the MCRF approach. For a clear comparison, here we use the notation style of Allard *et al*. (2011) and use 'two-point conditional probability' (called 'directional bivariate probability function' by Allard *et al*.) to refer to transition probability or transiograms because they did not use the term of transition probability.

The MCP solution (Allard *et al*., 2011, equation (7)) is copied here as

$$p^*_{i0|i_1,\ldots,i_n} = \frac{p_{i_0} \prod_{k=1}^{n} p_{i_k|i_0}(\mathbf{h}_{0k})}{\sum_{i_0=1}^{I} p_{i_0} \prod_{k=1}^{n} p_{i_k|i_0}(\mathbf{h}_{0k})}. \quad (4)$$

We show only the conditional probability form because it is directly related to the MCRF general solution and it was also the form used in simulation by Allard *et al*. (2011). We can easily derive such a solution by first applying the definition of conditional probability, and then using the law of total probability to the denominator and applying the conditional independence assumption.

The MCRF general solution provided in Li (2007) has been given here earlier (Equation (2) or (3)). Using the notation style of Allard *et al*. (2011), the MCRF general solution can be rewritten as

$$p^*_{i0|i_1,\ldots,i_n} = \frac{p_{i0|i_1}(\mathbf{h}_{10}) \prod_{k=2}^{n} p_{i_k|i_0}(\mathbf{h}_{0k})}{\sum_{i_0=1}^{I} p_{i0|i_1}(\mathbf{h}_{10}) \prod_{k=2}^{n} p_{i_k|i_0}(\mathbf{h}_{0k})}. \quad (5)$$

The key to obtaining this solution is to factorize the last joint probability term as $p_{i_0,i_1} = p_{i0|i_1}p_{i_1}$ rather than $p_{i_0,i_1} = p_{i_1|i_0}p_{i_0}$, as explained in Li (2007).

Then what is the difference between the MCP solution and the MCRF solution? Equation (5) does not have the class mean value term p_{i_0} (the mean proportions of classes or marginal probabilities), but one of its two-point conditional probabilities has a different head-tail direction. However, Equation (4) has the class mean value term. This means that Equation (5) can have varying class mean values, but Equation (4) needs the mean values of all classes for estimating the local conditional probability distribution at any uninformed location. It is true that Equation (5) can be converted into Equation (4) by replacing $p_{i0|i_1}(\mathbf{h}_{10})$ with $p_{i_1|i_0}(\mathbf{h}_{01})p_{i_0}/p_{i_1}$. However, to do so one has to first assume the existence of stationary class mean values and the symmetry of two-point conditional probabilities. The latter does not hold for uni-directionally estimated two-point conditional probabilities. In fact, $p_{i0|i_1}(\mathbf{h}_{10}) = p_{i_1|i_0}(\mathbf{h}_{01})p_{i_0}/p_{i_1}$ only approximately holds in real data for omni-directionally estimated two-point conditional probabilities. This means that Equation (4) (the MCP solution) is a special case of Equation (5) (the MCRF general solution).

The general solution of MCRFs (Equation (5)) is completely explained by the MCRF theory. In the MCRF quadrant search simulation algorithm, it is always assumed that one neighbour within the neighbourhood serves as the last stay location of the spatial Markov chain. Equation (4) also can be simply explained by the MCRF theory. By assuming the existence of stationary class mean values and that the last stay location of the spatial Markov chain in a MCRF is always outside the neighbourhood (or correlation range), we have $p_{i0|i_1}(\mathbf{h}_{10}) \approx p_{i_0}$ because of $\lim_{\mathbf{h}_{10}\to\infty} p_{i0|i_1}(\mathbf{h}_{10}) = p_{i_0}$, a basic property of first-order stationary Markov chains. That is, the transition probability $p_{i0|i_1}(\mathbf{h}_{10})$ from the last stay location to the current location to be estimated will decay to its limit (the class mean value p_{i_0}) when the lag is sufficiently large. Equation (5) will directly become Equation (4) (note that i_1 is thus excluded from the neighbourhood because it is outside, and the neighbourhood size becomes $n-1$, which may still be denoted by n). In general, the last stay location of the spatial Markov chain in a MCRF may have three situations:

(i) always within the neighbourhood, (ii) always outside the neighbourhood and (iii) unsure (that is, sometimes inside, sometimes outside). The general solution of MCRFs (Equation (5)) covers all the three situations but Equation (4) covers only the second situation approximately. Again, Equation (4) is a special case of Equation (5). However, Allard *et al*. (2011) mistakenly thought the MCP approach was a generalization of the MCRF approach.

While it is easy to derive, Li (2007) did not provide the solution of Equation (4) because it is an inferior special case. If we compare MCRFs to kriging (or, more precisely, indicator kriging), Equation (5) might be regarded as the counterpart of ordinary kriging, but Equation (4) at best might be regarded as the counterpart of simple kriging. What Li (2007) did provide is the general computable form of Markov chain geostatistics, an ordinary MCRF solution, rather than a special case.

Thus, both Equation (4) and Equation (5) were obtained as solutions of MCRFs. The major idea of the MCRF theory that a single spatial Markov chain moves or jumps in a space and decides its state at any location by interactions with its nearest neighbours in different directions is perfectly explained by Equation (5). However, Equation (4) is just a special case of Equation (5) by assuming the existence of stationary class mean values, the last stay location of the spatial Markov chain always to be outside the range and the transition probability decays to the tail class mean value, or outside the neighbourhood where its influence is screened and thus decays to the tail class mean value. Therefore, Equation (4) was not regarded as a general solution. While simple kriging is useful in geostatistics to deal with stationary continuous variables such as residuals, there are no residuals to deal with in categorical variables. Because Equation (4) was much less useful than Equation (5) and was also included in Equation (5), it was therefore discarded.

In addition to the implied assumptions by Allard *et al*. (2011), the MCP method did not consider data configuration (or data clustering effect) in its algorithm for conditional simulation. We found such an algorithm was inferior and discarded it, as mentioned above. Allard *et al*. (2011) showed some unconditionally simulated realizations as follows: 'The simulation followed a classical sequential process, except that the random path was organized following a multi-grid with four levels. At the first level, the grid was subsampled using an internode distance equal to the basic internode distance multiplied by 2^4. The nodes in this subsample were first simulated following a random path. Then, the operation was repeated for lower levels using powers from 3 to 1 to sample the grid. In this way, the random path was partially organized by regularly spreading the first simulated values. At each node of the grid, the neighbourhood consisted of the eight closest previously simulated nodes'. At the first level, they could not have sufficient simulated nodes or even might have no simulated node in a neighbourhood to meet the required number of eight neighbours. For other levels, the 'eight closest previously simulated nodes' were actually located in eight directions segmenting the search circle equally (except for boundary nodes). Li (2007) stated that 'If the neighbourhood is symmetric, we can find a successful application example from Besag (1986) for image processing, where eight adjacent neighbours of x were all assumed conditionally independent. Ripley (1990) suggested that in a pairwise interaction process site interactions between a point and its distant nearest neighbours (in any directions) might be treated independently' and 'With a conservative attitude, it is suggested herein that if the nearest known neighbours are generally located uniformly throughout the study space and a random path is used in a simulation, inclusion of nearest known neighbours in non-cardinal directions may be considered'.

Li (2007) did not limit the conditioning of the MCRF approach to four neighbours organized in orthogonal directions. The neighbourhood size for the MCRF approach is flexible. Depending on different lattices (or grids), the number of cardinal directions can be different and can be three for a triangular lattice, four for a rectangle lattice or six for a hexagonal lattice. Eight has also been used in Markov random fields (Li, 2007). Li (2007) finally notes 'In addition, note that the so-called cardinal directions are not limited to the exact four axial directions. For example, a triangle lattice has only three cardinal directions (Pickard, 1980). A suitable tolerance angle for cardinal directions should be feasible for dealing with irregular point samples with a random path, and in a non-lattice space it may be feasible to consider more than four cardinal directions'.

Final remarks

The earlier multi-dimensional Markov chain models proposed in geosciences were not practical for real world applications for conditional simulation of categorical spatial variables, and some were primitive, for example the Markov chain model by Lou (1996) and the CMC model of Elfeki (1996). Although they did not effectively extend Markov chains into practical geostatistics, they were initial stages and inspired later research.

The papers of Li and his colleagues published before 2007 in two-dimensional Markov chain modelling attempted to build on the CMC idea and correct its deficiencies. The CMC model by Elfeki (1996) and its extension (Elfeki & Dekking, 2001), although having some deficiencies, were recognized as making valuable progress in multi-dimensional Markov chain modelling by Li and his colleagues. The basic idea of the CMC model of coupling two independent chains was difficult to expand to support a practical Markov chain geostatistical approach and Li (2007) began to explore whether or not a single chain idea could work and that finally generated the MCRF theory. The MCRF theory generally supports Markov chain geostatistics, but the implementation methods, such as model expansion, algorithm design and transiogram estimation, still need to be developed. Thus the development of the MCRF approach or MCG is still at the early stage but the framework is present.

The MCRF theory (Li, 2007) contains the following: (i) the basic idea that a single spatial Markov chain moves or jumps in a space and decides its state at any location by interactions with its nearest neighbours in different directions; (ii) a

directly computable general solution, derived with the conditional independence assumption (plus the conditional probability definition, the Bayes' theorem and the law of total probability); (iii) an accompanying metric to fit the general solution, the transiogram for representing spatial correlations of categories; and (iv) a finding based on the Pickard theorem and its applications that the conditional independence could be regarded as proper for nearest neighbours in cardinal directions in a sparse data space. Although the CMC model of Elfeki (1996) is inevitably related, the core ideas of the MCRF approach are not relevant to two- or coupled-Markov chains. MCRFs are a special kind of Markov random fields, as noted in Li (2007). The general solution of MCRFs can be written in the form of a Gibbs distribution function, which characterizes Markov random fields. Whether it is related to the maximum entropy principle was never a concern. The maximum entropy principle was first expounded by Jaynes (1957a,b), and has been widely used in many applications. Examples of modelling categorical variables are given by Phillips *et al*. (2006, 2008) in species distribution modelling. Therefore, whether or not the MCRF approach can be related to the maximum entropy principle, there seems to be no reason to claim that it belongs to the BME approach proposed by Bogaert (2002) and D'Or & Bogaert (2004). The single chain idea has not been proposed previously in geostatistics or multi-dimensional Markov chain modelling and it does explain MCRF models in any form. The MCRF quadrant search simulation algorithm is just a specific application of the MCRF general solution.

References

Allard, D., D'Or, D. & 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 & Risk Assessment*, **16**, 425–448.

Brus, D.J., Bogaert, P. & Heuvelink, G.B.M. 2008. Bayesian maximum entropy prediction of soil categories using a traditional soil map as soft information. *European Journal of Soil Science*, **59**, 166–177.

D'Or, D. & Bogaert, P. 2004. Spatial prediction of categorical variables with the Bayesian maximum entropy approach: the Ooypolder case study. *European Journal of Soil Science*, **55**, 763–775.

Elfeki, A.M. 1996. *Stochastic characterization of geological heterogeneity and its impact on groundwater contaminant transport*. PhD dissertation, Delft University of Technology, Balkema Publisher, Rotterdam, The Netherlands.

Elfeki, A.M. & Dekking, F.M. 2001. A Markov chain model for subsurface characterization: theory and applications. *Mathematical Geology*, **33**, 569–589.

Gray, A.J., Kay, I.W. & Titterington, D.M. 1994. An empirical study of the simulation of various models used for images. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **16**, 507–513.

Jaynes, E.T. 1957a. Information theory and statistical mechanics. *Physical Review Series II*, **106**, 620–630.

Jaynes, E.T. 1957b. Information theory and statistical mechanics II. *Physical Review Series II*, **108**, 171–190.

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

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

Li, W. & Zhang, C. 2008. A single-chain-based multidimensional Markov chain model for subsurface characterization. *Environmental & Ecological Statistics*, **15**, 157–174.

Li, W. & Zhang, C. 2011. A Markov chain geostatistical framework for land-cover classification with uncertainty assessment based on expert-interpreted pixels from remotely sensed imagery. *IEEE Transactions on Geoscience & Remote Sensing*, **49**, 2983–2992.

Li, W., Zhang, C., Burt, J.E., Zhu, A.X. & Feyen, J. 2004. Two-dimensional Markov chain simulation of soil type spatial distribution. *Soil Science Society of America Journal*, **68**, 1479–1490.

Li, W., Zhang, C., Dey, D.K. & Wang, S. 2010. Estimating threshold-exceeding probability maps of continuous environmental variables with Markov chain random fields. *Stochastic Environmental Research & Risk Assessment*, **24**, 1113–1126.

Lou, J. 1996. Transition probability approach to statistical analysis of spatial qualitative variables in geology. In: *Geologic Modeling and Mapping* (eds A. Forster & D.F. Merriam), pp. 281–299. Plenum Press, New York.

Phillips, S.J. & Dudik, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Pickard, D.K. 1980. Unilateral Markov fields. *Advances in Applied Probability*, **12**, 655–671.

Ripley, B.D. 1990. Gibbsian interaction models. In: *Spatial Statistics: Past, Present, and Future* (ed. D.A. Griffith), pp. 3–25. Institute of Mathematical Geography, Syracuse University, Ann Arbor, Michigan.

Weidong Li & Chuanrong Zhang

Department of Geography and Center for Environmental Sciences and Engineering, University of Connecticut, Storrs, Connecticut 06269, USA.