A Comment on Sartore's "spMC: Modelling Spatial Random Fields with Continuous Lag Markov Chains"

Weidong Li and Chuanrong Zhang

This comment aims to reduce the misunderstandings caused by Sartore (2013), which misused two Markov chain random field (MCRF) simulation algorithms. Recently, Sartore (2013) developed a spMC package using *R* computer language for modeling categorical random fields with transition probabilities. Sartore tested four simulation algorithms, and showed that the fixed-path and random-path MCRF algorithms had the lowest accuracies.

We would like to point out herein that Sartore did not understand well the MCRF model/theory proposed in Li (2007a), and also incorrectly used the fixed-path and random path MCRF algorithms described respectively in Li (2007b) and Li and Zhang (2007). First, Sartore improperly named the fixed-path and random-path MCRF algorithms as "fixed and random path methods". This could easily cause misunderstandings. As stochastic spatial simulation algorithms, they must have a model for estimating local probability distributions.

Second, the equation provided in Sartore (2013) for the simplified MCRF model is improper. Sartore did not write the denominator of the simplified MCRF model equation, probably not intending to show the similarity between the simplified MCRF model proposed in Li (2007a) and the "Multinomial categorical simulation procedure" (a name provided by Sartore (2013) for the "Markovian-type categorical prediction" method suggested by Allard et al. (2011)). In the process that we endeavored to solve the small class underestimation problem of the coupled Markov chain (CMC) model proposed by Elfeki (1996) and extended by Elfeki and Dekking (2001), the single-chain-based MCRF idea (including Bayesian decomposition and the conditional independence assumption) simply led us to the simplified MCRF model (Li 2007a, Li and Zhang 2008), which theoretically corrected the problem of the CMC model. Allard et al. (2011) made a large effort to connect the simplified MCRF model with the maximum entropy model suggested by Bogaert (2002). The work by Allard et al. (2011), if sufficiently rational, would well prove the correctness of the MCRF theory and model from another angle. Unfortunately, Allard et al. (2011) misunderstood the MCRF theory and model - Allard et al. (2011) claimed a new model, and surprisingly wrote and interpreted the simplified MCRF model as the CMC model. This problem was pointed out by our comments on the paper (see Li and Zhang 2012).

Third, Sartore apparently applied the simplified MCRF model to sample data through improper algorithms by only "considering those nearest points along the axial directions within a radius of length one" (see Sartore 2013 p. 24). The fixed-path MCRF algorithm we suggested requires first connecting neighboring sample data in different directions into a mesh and then filling mesh holes (i.e., segmented unobserved subareas), all by simulation. Because of the complexity of the procedure, only a simple algorithm for regular sample data was developed and tested in Li (2007a, 2007b). The random-path MCRF algorithm we suggested divides each neighborhood search area into four quadrants and takes one nearest neighbor from each quadrant. In case there are no data available in some quadrants (e.g., on boundaries or in some subareas where data are extremely sparse at the beginning of simulation), the neighborhood size will be less than four (Li and Zhang 2007). Since the random-path MCRF algorithm can work with both regular and irregular sample data and can generate simulated patterns with smoother boundaries, this random-path algorithm is preferable to the fixed-path algorithm. Because Sartore ignored those neighboring data that are spatially closer but lie in non-axial directions, Sartore's results may not be accurate.

The main ideas for the MCRF approach were misunderstood by some readers in geostatistics and geosciences. A reason might be that some readers were not familiar with the specific ideas and knowledge used in the MCRF approach (i.e., Markov chain geostatistics), such as "a single Markov chain in a multi-dimensional space with local updating", Bayesian decomposition, and conditional independence. In fact, both the Bayes' theorem and the conditional independence assumption are widely known in statistics;

they just were not used in geostatistics previously to deal with the sparse spatial data in a neighborhood for estimating the local conditional probability distribution at an unobserved location. The idea of using a single Markov chain for multidimensional simulation is a new idea, and it is this idea that led us to the MCRF model. This idea should be understandable if readers read the paper of Li (2007a), which explains why such an idea was suggested.

Bibliography

D. Allard, D. D'Or, and R. Froidevaux. An efficient maximum entropy approach for categorical variable prediction. *European Journal of Soil Science*, 62(3):381–393, 2011. doi: 10.1111/j.1365-2389.2011.01362.x.

P. Bogaert. Spatial prediction of categorical variables: The Bayesian maximum entropy approach. Stochastic *Environmental Research and Risk Assessment*, 16:425–448, 2002. http://www.springerlink.com/content/c0d0cgfmrmflqpp4.

A. Elfeki. *Stochastic characterization of geological heterogeneity and its impact on groundwater contaminant transport*. Ph.D. diss. Delft University of Technology, The Netherlands, 1996. Balkema Publisher, ISBN 90-5410-666-2. http://repository.tudelft.nl/view/ir/uuid%3Acffa6701-c543-41e3-b36f-440abf980b5f/.

A. Elfeki and F. Dekking. A Markov chain model for subsurface characterization: Theory and applications. *Mathematical Geology*, 33: 569–589, 2001. doi: 10.1023/A:1011044812133.

W. Li. Markov chain random fields for estimation of categorical variables. *Mathematical Geology*, 39: 321–335, 2007a. http://www.springerlink.com/content/d0j5870n46k80351.

W. Li. A fixed-path Markov chain algorithm for conditional simulation of discrete spatial variables. *Mathematical Geology*, 39(2): 159–176, 2007b. http://www.springerlink.com/content/f120u5h45u147t1k/.

W. Li and C. Zhang. A random-path Markov chain algorithm for simulating categorical soil variables from random point samples. *Soil Science Society of America Journal*, 71(3):656–668, 2007. https://dl.sciencesocieties.org/publications/sssaj/pdfs/71/3/656.

W. Li and C. Zhang, A single-chain-based multidimensional Markov chain model for subsurface characterization. *Environmental and Ecological Statistics*, 15(2):157-174, 2008. http://link.springer.com/article/10.1007/s10651-007-0045-9.

W. Li and C. Zhang. 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(1): 120-124, 2012. doi: 10.1111/j.1365-2389.2011.01413.x.

L. Sartore. spMC: modelling spatial random fields with continuous lag spatial Markov Chains. *The R Journal* 5, 16–28, 2013. https://journal.r-project.org/archive/2013-2/sartore.pdf.

Weidong Li & Chuanrong Zhang
Department of Geography
University of Connecticut,
Storrs, CT 06268.
weidong.li@uconn.edu & chuanrong.zhang@uconn.edu

11/09/2015, Storrs, CT