Comments on "Spatial hidden Markov chain models for estimation of petroleum reservoir categorical variables"

Weidong Li and Chuanrong Zhang
Department of Geography, University of Connecticut, Storrs, CT 06268

Huang et al. (2016a) claimed that they proposed a "spatial hidden Markov chain (SHMC) model" based on the Markov chain random field (MCRF) theory proposed by Li (2007). However, we found that their so-called SHMC model is exactly the same as the co-located MCRF cosimulation (coMCRF) model with one auxiliary categorical dataset, which was first presented in Li et al. (2013) and further described and used in Li et al. (2015). Huang et al. (2016a) ignored the coMCRF model in Li et al. (2013, 2015), and also ignored all of our publications in soil science related to MCRF simulation and transiogram modeling. Obviously Huang et al. has committed a plagiarism to our coMCRF model by claiming they proposed a SHMC model in Huang et al. (2016a). In addition, Huang et al. (2016a) made many wrong statements about geostatistics, Markov chain models, and Markov random fields, which are very misleading.

Introduction

In the "Introduction" section, Huang et al. (2016a) stated "Spatial Markov chain (SMC) models have also been widely adopted in petroleum reservoir to characterize the spatial heterogeneity of categorical variables through the conditional probabilities (transition probabilities) from different directions (Carle and Fogg 1997; Weissmann and Fogg 1999)". However, Carle and Fogg (1997) introduced continuous-lag Markov chain models, which were actually a joint modeling method of spatial transition probabilities for providing transition probability parameters to transition probability-based indicator geostatistics (TPROGS) (Carle and Fogg 1996), and Weissmann and Fogg (1999) was focused on hydrofacies modeling using TPROGS. What SMC models were adopted by them? Then Huang et al. (2016a) grouped both the coupled Markov chain (CMC) model proposed by Elfeki and Dekking (2001) and the MCRF model proposed by Li (2007) into SMC models. Thus, they regarded the CMC model, the MCRF model, and the TPROGS model all as SMC models, no matter whether these models are theoretically sound or not and whether they are kriging model or nonlinear Markov chain models. We don't think it is proper to call all of these three models "SMC models", although they all used transition probabilities. Did Carle and Fogg call their method a "spatial Markov chain" model? Did Elfeki and Dekking call their model a "spatial Markov chain" model? If they did not, isn't it proper to simply rename their models? Indeed, Li (2007) used the term "spatial Markov chain" to refer to the Markov chain in a specific MCRF and also used the name of "spatial Markov chain model" for specific MCRF models with different numbers of nearest neighbors and spatial dimensions (i.e., one to six nearest neighbors in cardinal directions, and one to three dimensions). However, considering that this term was not sufficiently unique to represent "single-chain-based multidimensional Markov chain model", Li (2007) and our later papers used the name of "Markov chain random field" to differentiate the new Markov chain model from other Markov chain models for spatial data. Once the name of a model was decided by the model proposer(s), it is not respectful for others to rename it without sufficient reasons and scientific progress.

Huang et al. (2016a) stated "Most traditional geostatistical models, like Markov chain random field (MCRF), use well data only and make prediction based on SMC, which results in a relatively low prediction accuracy (Huang et al. 2016b)". This is not true! First, traditional geostatistical models such as cokriging and indicator cokriging for making use of auxiliary data have existed for decades. The MCRF model is not a traditional geostatistical model. It even has not been well developed as a software system for release. Second, Huang et al. (2016b) did not use a proper MCRF model to make prediction, but just played a trick to mess the MCRF approach. Third, coMCRF model, which may use one or multiple auxiliary datasets, has already been proposed (see Li et al. 2013, 2015). Huang et al. (2016a) stated "Huang et al. (2016c) introduced a beta-transformed Bayesian updating model to boost the classification accuracy of category random field". Really?! We don't think that Huang et al. (2016c) used the proper simulation algorithms and computer programs to perform simulations, except that they just wanted to quickly claim new models even without understanding the differences between spatial models and nonspatial models. They further stated: "Auxiliary information, however, has not been taken into consideration". If this was true, then what had Li et al. (2013) and Li et al. (2015) done? Didn't these two papers present the coMCRF model for incorporating auxiliary information into the MCRF model? Didn't these two papers conduct cosimulations for soil class map updating and land cover post-classification, respectively? After the proposition of the MCRF model as a geostatistical model, its extension for collaborative modeling is a natural thing. In addition, both the MCRF model and the coMCRF model are not limited to two dimensions theoretically.

Model

The second section of the paper is "Review of Markov models". In this section, both Markov mesh models and spatial Markov chain models were reviewed. Let's see how Huang et al. (2016a) made their reviews.

For Markov mesh models, Huang et al. (2016a) wrote their model formula as a conditional probability

$$\Pr\left(f_s | f_{\eta_s}\right) = \Pr\left(f_s | f_{s-1}, f_{s_1}, f_{s_2}, \ldots, f_{s_l}\right) \quad , \tag{1}$$

and stated that "where s_1, s_2, …, s_l is its nearest known locations of current cell s in different directions; s-1 is always the start cell of the Markov chain to the unknown cell s, which is to be estimated". It is widely known that Markov mesh models are unilateral lattice models. Here Huang et al. (2016a) suddenly redefined them on surrounding nearest known locations as a geostatistical model. The purpose of their making such a misleading definition appeared later at the end of this section, where Huang et al. (2016a) stated that "A spatial Markov chain model may be viewed as a special case of Markov mesh models, whereas a Markov mesh model is an extension of spatial Markov chain models, called a generalized spatial Markov chain model". Such a statement is absolutely wrong. Markov mesh models have existed as a special subclass of Markov random fields for half century since 1960s (Abend 1965). How could they generalize the MCRF model proposed in 2007? Here Huang et al. (2016a) turned so-called SMC models, including the CMC model of Elfeki and Dekking (2001), the MCRF model of Li (2007), and the TPROGS model according to their classification, all into special cases of Markov mesh models. Huang et al. (2016a) "generalized" all of these non-unilateral models into unilateral lattice models. Isn't it rational? Couldn't Huang et al. understand the differences between unilateral models and non-unilateral models and the differences between lattice models and geostatistical models?

For spatial Markov chain models, Huang et al. (2016a) first presented a "generalized form" of the CMC model

$$\Pr\left(f_s | f_{s-1}, f_{s_1}, \ldots, f_{s_l}\right) = \frac{p_{f_{s-1}f_s} p^1_{f_{s_1}f_s} \cdots p^l_{f_{s_l}f_s}}{\sum_{f_s} p_{f_{s-1}f_s} p^1_{f_{s_1}f_s} \cdots p^l_{f_{s_l}f_s}} \tag{2}$$

by mimicking the generalized form of the MCRF model simplified using the conditional independence assumption. While the CMC model of Elfeki and Dekking (2001) is composed of two 1-D Markov chains, Huang et al. (2016a) made it composed of l+1 1-D Markov chains. What scientific problem did Huang et al. (2016a) want to solve by writing the CMC model like this? Then Huang et al. (2016a) stated that "Using the conditional independence assumption, Li (2007) gives the general expression of the conditional probability formula in Eq. (1) at any location s as"

$$\Pr(f_s|f_{s-1},f_{s_1},\ldots,f_{s_l}) = \frac{p_{f_{s-1}f_s}p^1_{f_sf_{s1}}\cdots p^l_{f_sf_{s_l}}}{\sum_{f_s} p_{f_{s-1}f_s}p^1_{f_sf_{s1}}\cdots p^l_{f_sf_{s_l}}} . \tag{3}$$

Here our questions are: 1) Could Huang et al. use the conditional independence assumption to get such a model without the derivation process of using Bayes' theorem for factorizing the spatial conditional probability distribution function? 2) How could one use such a model to conduct spatial simulation directly conditional on sample data without first defining the conditional probability function on sparse sample data? 3) Why weren't Markov mesh models used for conditional geostatistical simulations previously, but now suddenly become a generalized spatial Markov chain model in the eyes of Huang et al.?

The third section of the paper is "Spatial hidden Markov chain model". Huang et al. (2016a) stated "The SHMC is an extension of SMC". Which SMC? They stated "A spatial Markov chain … is characterized by its states and conditional probabilities through Eqs. (2) or (3)…". Thus, both the CMC model and the MCRF model were included, but the TPROGS model was not. They defined a local conditional probability distribution for their SHMC model as

$$\Pr(f_s|f_{s-1},f_{s_1},\ldots,f_{s_l},w_s) = \frac{\Pr(w_s|f_s)\Pr(f_s|f_{s-1},f_{s_1},\ldots,f_{s_l})}{\sum_{f_s}\Pr(w_s|f_s)\Pr(f_s|f_{s-1},f_{s_1},\ldots,f_{s_l})} . \tag{4}$$

But when they defined their prior conditional probabilities, they used the formula

$$\Pr(f_s|f_{s-1},f_{s_1},\ldots,f_{s_l}) = \frac{p_{f_{s-1}f_s}p^{(1)}_{f_sf_{s_1}}\cdots p^{(l)}_{f_sf_{s_l}}}{\sum_{f_s} p_{f_{s-1}f_s}p^{(1)}_{f_sf_{s_1}}\cdots p^{(l)}_{f_sf_{s_l}}} , \tag{5}$$

which is exactly the simplified MCRF model based on the conditional independence assumption provided in Li (2007), the same as above Eq. (3). Note that both a generalized MCRF model equation and specific MCRF models with nearest data only in cardinal directions were provided in Li (2007, p. 327-329, see "MCRF-based SMC models"). Also note that the superscripts used in the MCRF model for directions were unnecessary because the subscript labels were already sufficient; we, therefore, didn't use the superscripts anymore in MCRF model equations in our later papers after 2009.

In the above Eq. (4), their $\Pr(w_s|f_s)$, that is, their emission probability, was called cross-field transition probability in Li et al. (2013, 2015). Because the coMCRF model may incorporate multiple auxiliary datasets, it may have multiple cross-field transition probability terms, which represent the cross-field relationships between the primary categorical field and multiple different auxiliary fields. Here by combining the above Eq. (4) and Eq. (5) and using a cross-field transition probability (or their emission probability) expression $b_{f_sw_s}$ for the conditional probability $\Pr(w_s|f_s)$, the complete form of their so-called SHMC model is

$$\Pr(f_s|f_{s-1},f_{s_1},\ldots,f_{s_l},w_s) = \frac{b_{f_sw_s}p_{f_{s-1}f_s}p_{f_sf_{s_1}}\cdots p_{f_sf_{s_l}}}{\sum_{f_s}[b_{f_sw_s}p_{f_{s-1}f_s}p_{f_sf_{s_1}}\cdots p_{f_sf_{s_l}}]} . \tag{6}$$

One can see that their SHMC model (i.e. Eq. (6)) is exactly the same as the co-located coMCRF model with one auxiliary dataset presented in Li et al. (2013) and Li et al. (2015). For the sake of clarity, the co-located coMCRF model with one auxiliary dataset in Li et al. (2013) is copied here:

$$p\left[i_0\left(\mathbf{u}_0\right) \mid i_1\left(\mathbf{u}_1\right), \ldots, i_m\left(\mathbf{u}_m\right); r_0\left(\mathbf{u}_0\right)\right] = \frac{b_{i_0 r_0} p_{i_1 i_0}\left(\mathbf{h}_{10}\right) \prod_{g=2}^{m} p_{i_0 i_g}\left(\mathbf{h}_{0g}\right)}{\sum_{f_0=1}^{n}\left[b_{f_0 r_0} p_{i_1 f_0}\left(\mathbf{h}_{10}\right) \prod_{g=2}^{m} p_{f_0 i_g}\left(\mathbf{h}_{0g}\right)\right]}. \quad (7)$$

The co-located coMCRF model with one auxiliary dataset used in Li et al. (2015) is also copied here:

$$p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1), \ldots, i_m(\mathbf{u}_m); r_0(\mathbf{u}_0)] = \frac{q_{i_0 r_0} p_{i_1 i_0}(\mathbf{h}_{10}) \prod_{g=2}^{m} p_{i_0 i_g}(\mathbf{h}_{0g})}{\sum_{f_0=1}^{n} [q_{f_0 r_0} p_{i_1 f_0}(\mathbf{h}_{10}) \prod_{g=2}^{m} p_{f_0 i_g}(\mathbf{h}_{0g})]}. \quad (8)$$

Note the differences in symbols: the neighborhood size for the MCRF model was denoted as *m* rather than *l*+1, and the cross-field transition probability term was denoted as $b_{i_0 r_0}$ and $q_{i_0 r_0}$ in Li et al. (2013) and Li et al. (2015), respectively. Readers who have interests in the MCRF approach (i.e., Markov chain geostatistics) may read Li (2007), Li and Zhang (2008), and Li et al. (2013, 2015) for details.

Case study

The fourth section of Huang et al. (2016a) is "Case study". Let's see how they did their simulation. They estimated experimental transiograms from 509 data points and fitted all of them with exponential models. They did not use our transiogram joint modeling methods (Li and Zhang 2010), nor use the transition rate method for continuous-lag Markov chain models (also a transition probability joint modeling method) suggested by Carle and Fogg (1997). How did they guarantee their transiogram models (see Huang et al. 206a, figure 7) sum to unity in each row? Without satisfying this condition, how could they implement their simulations using transiogram models?

What are unbelievable are their simulated results provided in Figure 1 (i.e., figure 10 in Huang et al. (2016a)). Huang et al. (2016a) claimed that these results were generated by the "SMC" defined by Eq. (5), which is actually the simplified MCRF model based on the conditional independence assumption. These results, however, have no way to be the results produced by the real MCRF model. The authors also pointed out some problems about the results, as they stated "It is obvious that conditional data have played a role in controlling the distribution of lithofacies near the wells. However, the further counterparts are fragmented and random in the grid". We are afraid that the true problem is not fragmentation or randomness. The true problem with the results should be that small classes were strongly underestimated. These results are very similar to the results generated by the unconditional CMC model (see Elfeki and Dekking 2001). This means that Huang et al. (2016a) used the unconditional CMC model to conduct the simulations. It is not surprising, because they regarded both the CMC model and the MCRF model as SMC models, and their CMC model was always unconditional (to sample data ahead) (see Liang et al. 2014, Huang et al. 2016b).

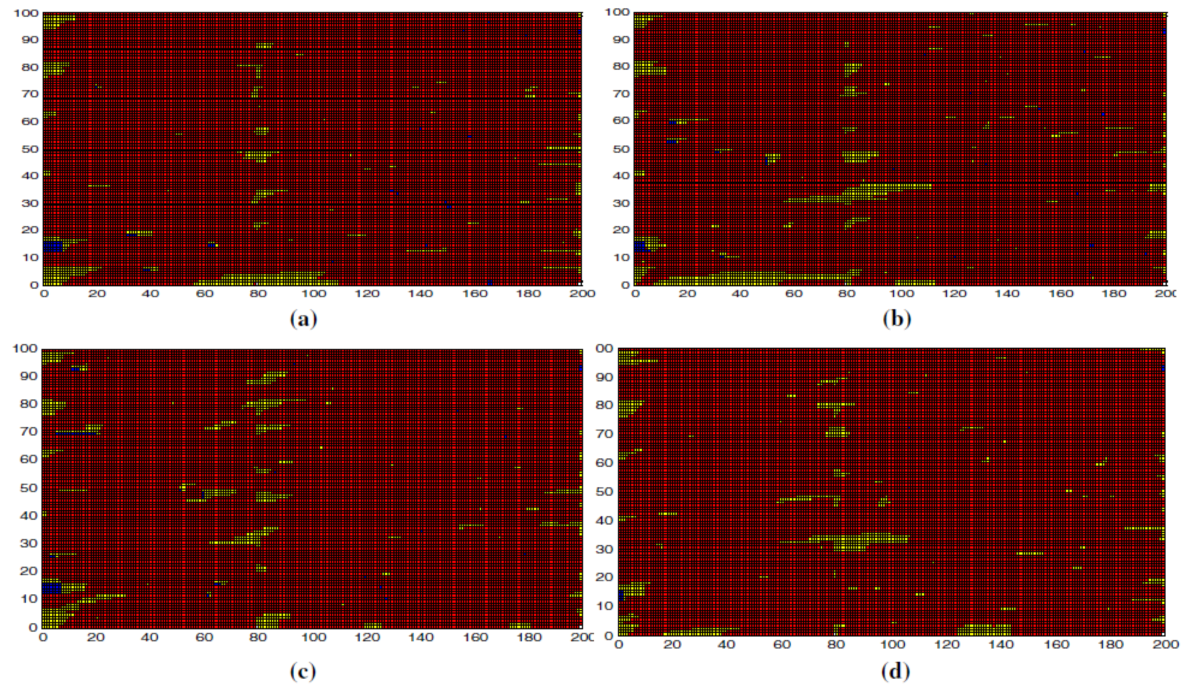

Fig. 10 Four stochastic simulation results based on three wells

Fig. 1 Four stochastic simulation results based on three wells using SMC (from Huang et al. 2016a).

Conclusions

Huang et al. (2016a) claimed "We have presented an SHMC model for geological facies modeling. This combines spatial Markov chain theory and Bayes estimation. We have adopted the specification of earlier published hidden Markov models. SHMC is based on neighborhood and cliques and has a solid theoretical foundation". These claims are wrong. Huang et al. (2016a) essentially proposed their "SHMC" model by plagiarization, because what they presented and claimed as a new model is exactly the co-located coMCRF model with one auxiliary dataset proposed by Li and his coauthors (see Li et al. 2013, 2015). What they did was just ignoring the existence of the co-located coMCRF model and giving it a different interpretation. They did not solve any scientific issues. More surprisingly, in their comparative simulations Huang et al. (2016a) actually implemented the CMC model without conditioning to data ahead, but claimed they implemented the real MCRF model. And this is a research sponsored by the Fundamental Research Funds for Central Universities (No. 2016zzts011) and National Science and Technology Major Project of China (No. 2011ZX05002-005-006).

Apparently by playing various tricks, such as by obscuring geostatistical models defined on sparse sample data and lattice models, the only purpose of Huang and his coauthors was to take the coMCRF model as theirs. According to their logic, cokriging could be called hidden kriging and taken as a new model, and indicator cokriging also could be called hidden indicator kriging and taken as a new model. Is this rational? What is their contribution to the science and what existing scientific issues did they solve, except for claiming a model from others as theirs and messing a new geostatistical approach?

In summary, Huang et al. (2016a) is misleading audiences through tricks. It's our responsiblitily to clarify this fact.

References

Abend K, Harley TJ, Kanal LN (1965) Classification of binary random patterns. IEEE Trans Inf Theory 11:538-544

Carle SF, Fogg GE (1996) Transition probability-based indicator geostatistics. Math Geol 28(4): 453-477

Carle SF, Fogg GE (1997) Modeling spatial variability with one and multidimensional continuous-lag Markov chains. Math Geol 29(7):891-918

Elfeki A, Dekking M (2001) A Markov chain model for subsurface characterization: theory and applications. Math Geol 33(5):569-589

Huang X, Li J, Liang Y, Wang Z, Guo J, Jiao P (2016a) Spatial hidden Markov chain models for estimation of petroleum reservoir categorical variables. J Petrol Explor Prod Technol, doi: 10.1007/s13202-016-0251-9

Huang X, Wang Z, Guo J (2016b) Theoretical generalization of Markov chain random field from potential function perspective. J Cent South Univ 23(1):189-200

Huang X, Wang Z, Guo J (2016c) Prediction of categorical spatial data via Bayesian updating. Int J Geogr Inf Sci 30(7):1426-1449

Li W (2007) Markov chain random fields for estimation of categorical variables. Math Geol 39(3): 321-335

Li W, Zhang C (2008) A single-chain-based multidimensional Markov chain model for subsurface characterization. Environ Ecol Stat 15:157-174

Li W, Zhang C (2010) Linear interpolation and joint model fitting of experimental transiograms for Markov chain simulation of categorical spatial variables. Int J Geogr Inf Sci 24(6): 821-839

Li W, Zhang C, Dey DK, Willig MR (2013) Updating categorical soil maps using limited survey data by Bayesian Markov chain cosimulation. Sci World J, Article ID 587284, doi: 10.1155/2013/587284

Li W, Zhang C, Willig MR, Dey DK, Wang G, You L (2015) Bayesian Markov chain random field cosimulation for improving land cover classification accuracy. Math Geosci 47(2):123-148

Liang YR, Wang ZZ, Guo JH (2014) Reservoir lithology stochastic simulation based on Markov random fields. J Cent South Univ 21(9): 3610-3616

Weissmann GS, Fogg GE (1999) Multi-scale alluvial fan heterogeneity modeled with transition probability geostatistics in a sequence stratigraphic framework. J Hydrol 226(1):48-65