Comments on "Response to '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 06269, USA

After reading the response of Huang et al. to our comments on their article, and carefully thinking about their attitude and arguments, we conclude that their response is dishonest, wrong, and tricky. We believe that not only is Huang et al. (2016a) a plagiarism but also all of the related publications by this research group (i.e., Liang et al. 2014; Huang et al. 2016a, 2016b and 2016c) are essentially scientific frauds, although their arguments showed their ignorance in scientific research ethics. Below we would like to point out their main dishonest, wrong and tricky arguments in their response:

(1). Huang et al. stated "*As a response, firstly, we must acknowledge our negligence of the investigation and citation on their publications in soil science related to (co)MCRF simulation and transiogram modeling. Such ignorance is embarrassing and we apologize.*" However, we believe they knew what they were doing. W. Li reviewed their submission (the same title and content as Huang et al. 2016a) to *Journal of Data Science* in March 2015 and pointed out the major problems of their manuscript and also provided the references of Li et al. (2013, 2015). In that manuscript, the authors were Wang, Huang and Guo. However, they ignored the review comments, changed author sequence, added three coauthors, and then published the same manuscript in *Journal of Petroleum Exploration and Production Technology.*

(2). Huang et al. stated "*we want to clarify that the SHMC is different from the coMCRF model even though their expressions are nearly identical*". The truth is that their so-called spatial hidden Markov chain (SHMC) model is completely the same as the co-located coMCRF model with one auxiliary dataset, rather than being nearly identical. The Markov chain random field (MCRF) model includes a generalized form and specific MCRF models with nearest data only in cardinal directions (see "MCRF-based SMC models" in Li 2007). So no matter how many and which nearest data they used in their hidden spatial Markov chain (SMC) to construct the SHMC model, The hidden SMC is still a MCRF model, and the SHMC model is still a co-located coMCRF model with one auxiliary dataset.

(3). Huang et al. stated "*These two methods have different theoretical backgrounds, different simulation algorithms, and have been applied to different disciplines. Combing hidden Markov model with multidimensional Markov chain was, and is, original.*" This is obviously a wrong logic. If providing a different theoretical background or different simulation algorithm or applying it to different disciplines can turn an existing spatial statistical model into another new model, why didn't Huang et al. claim kriging and Markov random fields to be their new models by doing so? The MCRF model is original, because IT DID NOT EXIST IN THIS WORLD before it was mathematically derived using Bayes' theorem and the conditional independence assumption of nearest data and proposed as a new spatial statistical model by Li (2007), not just because it has a different theoretical background and simulation algorithm. After it was proposed, it is not a problem for anybody to explore its statistical relationships with other statistical theories, principles or existing models. But claiming it as another new spatial statistical model by giving it a different explanation is wrong, no matter whether the different explanation is rational, correct or not. Similarly, the coMCRF model as a natural extension of the MCRF model also

DID NOT EXIST IN THIS WORLD before it was openly proposed. In addition, what is the different theoretical background of the so-called SHMC model, while Huang et al. directly used a MCRF model to replace the traditional one-dimensional Markov chain in hidden Markov chain model? All what Huang et al. did was reinterpreting the co-located coMCRF model with one auxiliary dataset as a SHMC model. If one can claim a new model using such a behavior, can one reinterpret cokriging as a hidden kriging and claim it as a new model?

(4). Huang et al. stated "*Last but not least, the statements with respect to geostatistics, Markov chain models, and Markov random fields are based on our interpretations. We do not think these analyses are "lies", even though some concepts have been confused. We fully apologize for our misleading and incorrect judgments*." Maybe those statements are just their misunderstandings. However, if Huang et al. even have no knowledge about geostatistics, Markov chain models, Markov random fields, and their differences, what are the reasons for them to quickly publish several papers to reinterpret the MCRF approach and claimed to propose multiple ***new*** spatial statistical models based on MCRF within a very short term?

(5). Huang et al. claimed they proposed a multidimensional Markov chain based on the fully independence assumption by the equation (2) in their response. How could they get such a "generalized" coupled Markov chain (CMC) model simply by a fully independence assumption without any reasons? To obtain the CMC model, Elfeki and Dekking (2001) first assumed there are two 1D Markov chains in a 2D space and they move to the same grid cell from neighboring cells, then assumed the two 1D Markov chains to be fully independent and excluded conflict transitions, at the meantime assumed one 1D Markov chain is conditional to a future state, and finally obtained their conditional CMC model. Although the CMC model has defects and is not much practical when boreholes are not sufficiently dense, it has its contributions as an initial idea. The small class underestimation flaw of the CMC model was proved and corrected by (Li 2007) with the MCRF model. Huang et al. also mentioned this point in their paper. Then what were the reason and purpose for Huang et al. to propose such a model here? What scientific issues did they want to solve?

(6). In the equation (3) in their response, they provided the multidimensional Markov chain based on the conditional independence assumption. How simple! In their eyes, the MCRF model is just the result of a conditional independence assumption. However, the truth is far more complex. Without spending many years in solving the deficiencies of the CMC model, without the single-Markov-chain random field idea, without the sparse data neighborhood definition, without the Bayesian factorization based on the local sequential Bayesian updating idea, and without establishing the conditional independence assumption for nearest data within a neighborhood, there is no way to derive the MCRF model. And without developing the fixed-path and random-path sequential simulation algorithms and the tramsiogram joint modeling methods, there is also no way to implement the MCRF model in any form and publish it in a journal article. However, by writing equations without understanding what they are, Huang et al. dared to take anything as their new models.

(7). Huang et al. stated "*We have missed the pioneering work by Li and his colleagues, we apologize for this. The latest publication, i.e., Li et al. (2015), however, has been cited in Huang et al. (2016a).*" That is exactly their problem – if they knew the existence of the coMCRF model, why did they still claim they proposed a so-called SHMC model based on the MCRF theory, while the so-called SHMC model is exactly the same as the co-located coMCRF model with one auxiliary dataset? The Equation of the co-located coMCRF model with one auxiliary dataset was clearly presented in Li et al. (2015), and Li et al. (2015) cited Li et al. (2013).

(8). Huang et al. stated "*Didn't Bayesian updating and Beta transformation are proper simulation algorithms? Didn't the transiogram models used in Huang et al. (2016c) are valid continuity measurement for spatial models?*" Bayesian updating and Beta transformation are surely not proper simulation algorithms here, because Huang et al. (2016c) was claiming new spatial statistical models. New spatial statistical models cannot be so simply proposed by writing some equations without solving any existing scientific issues and by using other computer algorithms to generate some results. For the transiogram modeling part, the problem is not whether the exponential model used by Huang et al. (2016c) is valid or not. The real problem is that Huang et al. (2016c) did not use any transiogram joint modeling method to obtain their transiogram models, similar as what they did in Huang et al. (2016a). Without a coregionalization model for multiple classes, they cannot make each row of transiogram models sum to unity at any lag value for implementing their spatial models. Obviously they did not understand how to use transiograms in spatial statistical models, and they also did not develop any functional computer algorithms to implement those models.

(9). Huang et al. stated "*This methodology indicates that MRF can also be served as a geostatistical model for handling sparse data set. One can find this application in Fig. 1.*" They provided the study from Norberg et al. (2002) as evidence to mislead readers. The MRF model provided in Norberg et al. (2002) is still a lattice model, implemented by an iterative algorithm from an initial image as the start point. It does not directly deal with distant interactions as typical geostatistical models do.

(10). Huang et al. stated "*The main purpose of introducing such a method is to incorporate auxiliary information into the existing models*." What are the existing models they used? They were incorporating auxiliary information into the MCRF model. Since Li et al. (2013, 2015) have incorporated auxiliary information into the MCRF model and proposed the coMCRF model, what is the reason for Huang et al. further claimed a SHMC model as a special case of the coMCRF model to incorporate auxiliary information?

(11). Huang et al. stated "*Please note that the first attempt for integrating secondary data with Markov chain is by no means introduced by Li et al. (2013), but appears three years ago in Li et al. (2010) who integrate Markov chain model with multi-scale data for lithology stochastic simulation*." We read the Chinese paper of Li et al. (2010) recently, and found that Li et al. (2010) stated they incorporated multi-scale auxiliary data into the CMC model of Elfeki and Dekking (2001) using Bayes' theorem. It is surprising that three years later after the CMC model was theoretically proved defective (e.g., the small class underestimation problem) and corrected by the MCRF model, some people were still attempting to expand it. Unfortunately, Li et al. (2010) did not use Bayes' theorem correctly; otherwise, the so-called SHMC model proposed by Huang et al. (2016a) also can be regarded as a plagiarism to Li et al. (2010) because Huang et al. (2016a) included the CMC model into SMC models.

(12). Huang et al. stated "*Without citation, isn't* $b_{i_0 r_0}$ *in Eq. (7) a "plagiarism" to* $\Pr\left(Z_{i,j}^1 = S_q \mid Z_{i,j}^2 = D\right)$ *in Eq. (6) from the perspective of Li and Zhang?*" Of course not! First, incorporating auxiliary data into a spatial model is not a new idea. Many relevant studies have been published since 1980s. The coMCRF model is just a natural extension of the MCRF model as a basic geostatistical model and its novelty lies with the MCRF model. Second, we suggest Huang et al. look clearly at the two probability terms. They are different things. The cross-field transition probability term $b_{i_0 r_0}$ in the co-located coMCRF model is essentially a likelihood function, which can be written like $b[Z'(\mathbf{u}') = r_0 | Z(\mathbf{u}) = i_0]$ with *Z'* being the auxiliary dataset.

But $\Pr\left(Z_{i,j}^1 = S_q \mid Z_{i,j}^2 = D\right)$ in Li et al. (2010) for incorporating auxiliary data into the CMC model is just an ordinary conditional probability with Z^2 being the auxiliary dataset (see Li et al. 2010, p. 74).

(13). Huang et al. argued that "*We always consider the neighborhood of the unknown cell s in cardinal directions. Therefore, we have at most 4 and 6 neighbors for the 2-D plane and 3-D space respectively, just as shown in Fig. 2. The MCRF model proposed by Li (2007) considers arbitrary directions, 3-D simulation, to the best of our knowledge, however, has not been implemented in their publications.*" The MCRF model includes a generalized form and specific MCRF models with nearest data only in cardinal directions (see "MCRF-based SMC models" in Li 2007). The MCRF model with nearest neighbors only in cardinal directions was our initial idea for correcting the CMC model and our simulation algorithms were also mainly focused on such a neighborhood structure. So no matter how many and which nearest data they considered, their so-called SHMC model is still a co-located coMCRF model with one auxiliary dataset. Implementing an existing model cannot be claimed as proposing a new model.

(14) Huang et al. further argued that "*In addition, the most significant difference between the coMCRF and the SHMC model lies in the simulation algorithm. The former uses a series of expert-interpreted data sets and an image data set pre-classified by the supervised maximum likelihood (SML) algorithm (Li et al. 2015). The latter uses Viterbi algorithm to integrate well data with geological conceptual data (sonic impedance). How can Li and Zhang state that the SHMC model is a plagiarism to their coMCRF model simply due to the similar expression and totally ignore the algorithm innovation? Is this declaration rational? Has the hidden Markov model been combined with multi-dimensional Markov chain in previous publications? Has the 3-D simulation been implemented with MCRF before?*" These arguments are irrational and cannot deny the plagiarism nature of Huang et al. (2016a). The plagiarism we talked here is about the model, not about the case study or simulation algorithm. Huang et al. (2016a) actually implemented the CMC model.

(15) Huang et al. stated "*By using Eq. (10), we can find in Fig. 3 that sills in each row sum approximately to unity*". No matter what model they used for fitting a specific experimental transiogram, without using a correct joint modeling method they cannot normally implement the MCRF model or the coMCRF model in any form. Summing to unity at any specific lag value for each row of transiogram models is strictly required. Probably they even did not use transiogram models in their simulations but pretended they used them.

(16) Huang et al. stated "*As for the simulation results, we agree with Li and Zhang that small classes are strongly underestimated. This undesirable result is caused by our false conception that both the CMC model and the MCRF model are SMC models. Apparently, we should apologize for the misleading results*". False conception cannot cause such a problem. The truth should be that they implemented the CMC model but pretended they implemented the MCRF model. This is also why we believe they did not use transiogram models in their simulations, but just used a one-step transition probability matrix to implement the CMC model.

(17) Huang et al. stated "*We do believe, however, that the Viterbi algorithm-based SHMC model is an advance in lithology stochastic simulation with multi-scale data. We highly respect Li and his colleagues' impressive work in Markov chain geostatistics. The interesting proposition and algorithm of our contribution, however, should also be acknowledged*". Using the Viterbi algorithm to the CMC model, or to the CMC model with auxiliary data suggested by Li et al. (2010), cannot support their claim for a MCRF-based SHMC model, even if the coMCRF model had not been proposed. After the coMCRF model was proposed, reinterpreting a

special case of the coMCRF model (i.e., the co-located coMCRF model with one auxiliary dataset) as a new model (i.e., the so-called SHMC model) is undoubtedly a plagiarism, no matter whether a 3D case study is conducted or a 2D case study is conducted.

In geostatistics/spatial statistics, proposers of new approaches usually spent many years or their whole life to gradually accumulate knowledge, develop new ideas, solve existing scientific issues, and finally propose a new approach. And developing a new fundamental model into a full-fledged approach is also a large and long-term task. Typical examples are kriging and Markov random fields. However, without fundamental knowledge in geostatistics/spatial statistics, without solving any existing scientific issues, even without doing some reasonable application studies using existing spatial statistical methods, and starting from zero (note that the first author Huang is just a graduate student), the research group of Huang et al. suddenly claimed a series of ***new*** spatial statistical methods for categorical spatial variables, all based on the MCRF model, by playing tricks to mislead readers. They took the MCRF model from various angles, while ignoring the Markov chain geostatistics framework we proposed. In order to claim quickly, they even did not develop any correct computer programs for their case studies (e.g., they always implemented the CMC model but pretended they implemented the MCRF model). Their irrationality was also reflected by their arguments in their response to our comments. We can conclude that not only is Huang et al. (2016a) a plagiarism, but all of the related publications by this research group (i.e., Liang et al. 2014; Huang et al. 2016a, 2016b and 2016c) are essentially frauds. Ignorance in scientific research ethics cannot deny the plagiarism and trouble-making nature of their publications.

References

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 J, Xiong L, Fang S, Tang L, Huo H (2010) Lithology stochastic simulation based on Markov chain models integrated with multiscale data. Acta Pet Sin 31(1):73–77 (in Chinese)

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

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

Weidong Li Email: weidongwoody@gmail.com Tel: 1-860-486-2728
Chuanrong Zhang Email: zhangchuanrong@gmail.com Tel: 1-860-486-2610

August 2, 2016