Further Comments on "Combining spatial transition probabilities for stochastic simulation of categorical fields"

Weidong Li and Chuanrong Zhang

Department of Geography, University of Connecticut, Storrs, CT 06269

In our comment letter (Li and Zhang 2012) on Cao *et al.* (2011), we focused on clarifying their misunderstandings to the Markov chain random field (MCRF) approach (Li 2007). So we did not talk about the Tau model implementation method proposed by them. In their response letter to our comments, Cao *et al.* (2012) first claimed their method and results were sound and their conclusions were valid, then accused the MCRF approach, and finally stated that "*By adopting the general Tau model in a spatial context, Cao et al. (2011) proposed a method to relax the assumption of conditional independence in the spatial prediction and simulation of categorical fields. From this point of view, Cao et al. (2011) actually contributed to the continuing advancement of MCRFs framework and Markov chain geostatistics*". Here we would like to point out that it is not proper to claim their method to be contributive to the continuing advancement of MCRFs framework and Markov chain geostatistics.

The Tau model implementation method

Cao *et al.* (2011) described their Tau model and simulation algorithm as follows:

> The assumption of permanence of ratios is another way to approximate the conditional probability of Equation (5). To condense notation, we use A and $D_1, \ldots, D_N$ to represent the events in sample spaces of $C(\boldsymbol{x}_0)$ and $C(\boldsymbol{x}_1), \ldots, C(\boldsymbol{x}_N)$, respectively. For two events D_1 and D_2, considering the following logistic-type probability ratios, $r_0 = \frac{1-P(A)}{P(A)}$, $r_1 = \frac{1-P(A|D_1)}{P(A|D_1)}$, $r_2 = \frac{1-P(A|D_2)}{P(A|D_2)}$, and $r = \frac{1-P(A|D_1,D_2)}{P(A|D_1,D_2)}$, the permanence of ratios amounts to assuming
>
> $$\frac{r}{r_1} \approx \frac{r_2}{r_0} \tag{9}$$
>
> The idea behind this assumption is that ratios of information increments are typically more stable than increments themselves. Compared to the assumption of conditional independence, this assumption avoids the calculation of the marginal probability in Bayesian expansion (denominator of Equation (5)). Actually, in practice, the summation in the denominator of Equations (7) and (8) does not necessarily equal the marginal probability. It can be easily demonstrated that Equation (9) implies conditional independence (Equation (6)) but the reverse is not necessarily true.

This approximation actually also assumes a certain form of independence between D_1 and D_2. To relax this assumption, Journel (2002) introduced an exponent factor, τ_n, to Equation (9) to account for information redundancy between D_1 and D_2.

$$\frac{r}{r_1} = \left(\frac{r_2}{r_0}\right)^{\tau(D_1,D_2)} \tag{10}$$

Equation (10) can be generalized to N data events (Journel 2002, Krishnan 2008). Denoting $r_n = \frac{1-P(A|D_n)}{P(A|D_n)}, n = 1, \ldots, N$, and reexpressing r as $r = \frac{1-P(A|D_1,\ldots,D_N)}{P(A|D_1,\ldots,D_N)}$, one gets

$$\frac{r}{r_0} = \prod_{n=1}^{N} \left(\frac{r_n}{r_0}\right)^{\tau_n} \tag{11}$$

and thus

$$P(A|D_1, \ldots, D_N) = \frac{1}{1+r} \in [0, 1] \tag{12}$$

The main problem with this model is the determination of the exponent factor τ_n, which actually quantifies the information redundancy between D_n and D_{n-1} (Krishnan 2008). Recently, Chugunova and Hu (2008) showed that the Tau model with constant weights is inapplicable in some cases and suggested the necessity of inference of τ_n in each case and at each simulation point.

In this article, the following procedure is applied to obtain τ_n. First the nearest neighbor $\boldsymbol{x}_1$ of the target location $\boldsymbol{x}_0$ is selected and we let $\tau_1 = 1$. Then we assume the value $c(\boldsymbol{x}_1)$ of this selected location $\boldsymbol{x}_1$ is unknown and perform ordinary kriging (OK) to estimate it using the remaining neighbors as known data taking the OK weights as τ_n, $n > 1$. Equation (11) can be reformulated as

$$r = r_1 \left(\frac{r_2}{r_0}\right)^{\tau_2} \cdots \left(\frac{r_N}{r_0}\right)^{\tau_N} \tag{13}$$

where $\tau_n, n = 2, \ldots, N$, are the OK weights.

This procedure can be interpreted using consensus theory (Benediktsson and Swain 1992) as follows: First, the nearest neighbor $\boldsymbol{x}_1$ of the unknown event location $\boldsymbol{x}_0$ is selected and its 'opinion' on what the unknown event should be is assumed completely credible. Then the degree of agreement between the remaining $N - 1$ neighbors and the first selected nearest neighbor $\boldsymbol{x}_1$ is quantified. The more the class label (or attribute value in general) at $\boldsymbol{x}_n$ agree with that at $\boldsymbol{x}_1$, the larger the OK weights for $\boldsymbol{x}_n$ will be; this implies more redundant information between states at $\boldsymbol{x}_n$ and $\boldsymbol{x}_1$, and thus the 'opinion' of $\boldsymbol{x}_n$ should be suppressed. In kriging, all those weights depend (through the variogram model) on the distances between the sample data locations. For example, if the distance between $\boldsymbol{x}_n$ and $\boldsymbol{x}_0$ is much larger than the variogram range, the OK weight for $\boldsymbol{x}_n$ is 0, that is, $\tau_n = 0$, and its corresponding component in Equation (11) is $\left(\frac{r_n}{r_0}\right)^{\tau_n} = 1$; this means that the observed state at location $\boldsymbol{x}_n$ has no influence on the unknown state at location $\boldsymbol{x}_0$. On the other hand, if the OK weight $\tau_n = 1$, $\left(\frac{r_n}{r_0}\right)^{\tau_n} = \frac{r_n}{r_0}$, which means the 'opinion' of $\boldsymbol{x}_n$ is entirely credible. A nonnegativity constraint is imposed on the OK weights (Deutsch 1996) to ensure each $\tau_n \in [0, 1]$ and the sum of these exponents is 1.

Rationality Analysis

We have no problem with the permanence of ratios and the Tau model suggested by Journel (2002), although this model was not implemented for spatial data in his paper. The Tau model as an empirical model may be a good idea. Here what we want to show is whether the Tau model implementation method of Cao *et al.* (2011) is rational or not.

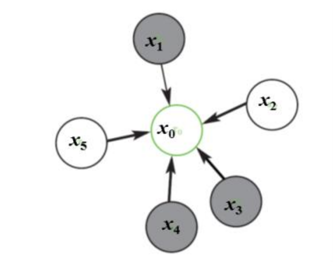

Fig. 1. Example of an unknown state (class label) at location $\boldsymbol{x}_0$ depending on its five nearest neighboring states at locations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_5$.

The left figure is the figure 1 in Cao *et al.* (2011). To estimate the conditional probability distribution of the class label at x_0 using the Tau model, the authors had to obtain the power parameters (i.e., Tau parameters) in above Equation (13). Let's see how they estimated those Tau parameters. We will illustrate their method in the following Figure 2.

They stated that "*In this article, the following procedure is applied to obtain τ_n. First the nearest neighbor x_1 of the target location x_0 is selected and we let $\tau_1 = 1$. Then we assume the value $c(x_1)$ of this selected location x_1 is unknown and perform ordinary kriging (OK) to estimate it using the remaining neighbors as known data taking the OK weights as τ_n, $n > 1$*" (Cao *et al.* p. 1780). According to their statements and explanations made above, they estimated the τ parameters using the following procedure as shown in Figure 2: (1) Assume an unknown state (class label) at location $\boldsymbol{x}_0$ depending on its five nearest neighboring states at locations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_5$. (2) Give x_1 the full credit, that is, allocate $\tau_1 = 1$ to the location x_1. (3) Assume x_1 is unknown, and then estimate the value at x_1 using ordinary kriging from other nearest neighbors, that is, x_2, x_3, x_4, x_5. The kriging weights allocated to these data will serve as their τ parameters. (4) Finally obtain a set of τ parameters for estimating the state at location $\boldsymbol{x}_0$ using the Tau model.

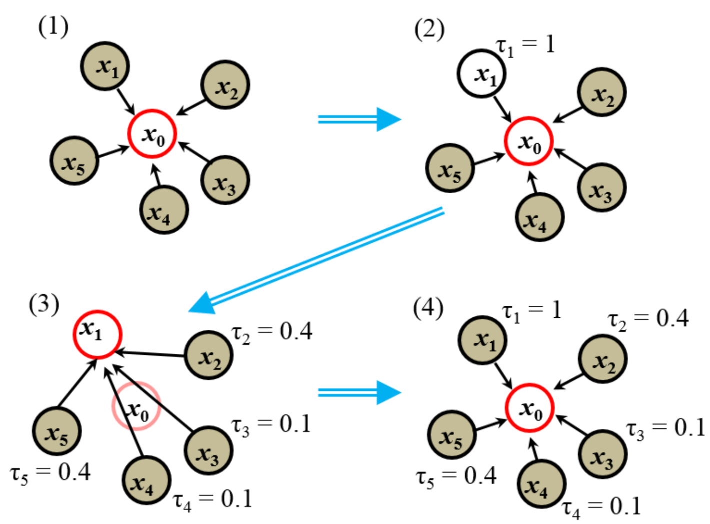

Fig. 2. Illustration of the Tau model implementation method, that is, the τ_n parameter estimation method, proposed by Cao *et al.* (2011): (1) Assume there is a five nearest data neighborhood. (2) Give the datum at x_1 the full credit with $\tau_1 = 1$. (3) Assume x_1 is unknown and estimate the value at x_1 using ordinary kriging from other nearest neighbors (i.e., x_2, x_3, x_4, x_5) to get kriging weights, which are allocated to these data as their τ parameters. (4) Finally obtain a set of τ parameters for estimating the state at location $\boldsymbol{x}_0$.

By this way, apparently the data close to x_1 will get larger weights, that is, x_2 and x_5 will get larger weights, for example, 0.4, no matter what their class labels are. The data far from x_1 will get smaller weights, that is, x_3 and x_4 will get small weights, for example, 0.1. Then the five nearest neighbors x_1, x_2, x_3, x_4, x_5 will have weights (i.e., τ parameters) of 1.0, 0.4, 0.1, 0.1, and 0.4, respectively. Now one can see that the ratio for the local conditional probability distribution at x_0 mainly depends on the nearest neighbors at one side and largely ignores those at the other side. The question here is: how can they know the data with low Tau parameters are redundant and the data with high Tau parameters are not redundant? This is apparently not reasonable. In fact, when they estimate the kriging weights for locations x_2, x_3, x_4, x_5 using x_1 as the uninformed location, they should often get some negative weights because these data are located only at one side of x_1. Constraining negative weights will get some 0 weights, which are still irrational. Even if they used the x_0 as the center to estimate weights for nearest data except for x_1, the method is still irrational, because there is no reason to give x_1 a large weight of 1.0 and give other neighbors small weights. There is no reason to estimate the state at a location mainly based on one of its nearest neighbors and regard others as redundant data.

Cao *et al.* (2011) spent a large volume of their paper to talk about other methods, especially non-spatial methods. For example, the whole section of "Methods" was talking about other things. However, they described the new method (i.e., the Tau model implementation algorithm) they proposed very simply using only a few sentences in a subsection. Apparently from above analysis, one can see that the method suggested by Cao *et al.* (2011) is not rational. Even if it was rational, their method is not a Markov chain spatial model. Therefore, it is not proper to claim it is contributive to the continuing advancement of MCRFs framework and Markov chain geostatistics. Considering that our concern was mainly about their misunderstandings and misinterpretations to our research and that we mainly aimed to communicate with them on Markov chain geostatistics, we did not mention this point in our previous comment letter. This note provides a complement to our previous comment letter and their response letter.

References

Cao, G., Kyriakidis, P.C., and Goodchild, M.F., 2011. Combining spatial transition probabilities for stochastic simulation of categorical fields. *International Journal of Geographical Information Science*, 25 (11), 1773–1791.

Cao, G., Kyriakidis, P.C., and Goodchild, M.F., 2012, Response to 'comments on "combining spatial transition probabilities for stochastic simulation of categorical fields" with communications on some issues related to Markov chain geostatistics'. *International Journal of Geographical Information Science*, 26 (10), 1741–1750.

Journel, A., 2002. Combining knowledge from diverse sources: an alternative to traditional data independence hypotheses. *Mathematical Geology*, 34 (5), 573–596.

Li, W., 2007. Markov chain random fields for estimation of categorical variables. *Mathematical Geology*, 39 (3), 321–335.

Li, W., and Zhang, C., 2012. Comments on 'Combining spatial transition probabilities for stochastic simulation of categorical fields' with communications on some issues related to Markov chain geostatistics. *International Journal of Geographical Information Science*, 26 (10), 1725–1739.