Comparison of Simple Indicator Kriging, DMPE, Full MV Approach for Categorical Random Variable Simulation

Yupeng Li and Clayton V. Deutsch

Inference of conditional probabilities at unsampled locations is a critical problem in geostatistics. In the context of categorical variables, the probabilities of each category could be calculated by indicator kriging or multiple point statistics - two mature conditional probability estimation approaches. These techniques are compared with the newly developed Direct Multivariate Probability Estimation technique. The comparison is based on the information provided by the different techniques. The informative strength function is used as a quantitative uncertainty assessment to the estimation results. The informative strength function measures how our uncertainty is reduced and how the unsampled location gains information from related neighbouring locations.

Introduction

Categorical variables such as facies or rock types usually reflect the origin of a rock unit. Although, the true facies for a specific unsampled location is unique, it is usually viewed as a random variable when the true facies on some location is inaccessible. The paradigm of geostatistics is to characterize any facies or rock types for unsampled location as a categorical random variables which are defined as the set of possible values or states that can take over the study area or at any particular location. Usually, the random variables are location dependent and donated as $Z(\mathbf{u})$, the upper-case letter such as $Z(\mathbf{u})$ will refer to a random variable at location $\mathbf{u}$. The set of possible outcomes of the random variable is denoted as $\{k, k=1,2,...,K\}$. Then, the probability of one category k to exist at location $\mathbf{u}$ is expressed as a probability:

$$p(\mathbf{u};k) = prob\{Z(\mathbf{u}) = k\} \tag{1}$$

The aim is to give a more accuracy and precise estimation for $p(\mathbf{u};k)$. For example, given category number with no other prior information, we can use the uniform distribution (least information probability) as the estimation, which will be: $p(\mathbf{u};k) = 1/K$. More practically, we have some global information for the probability of the category k to prevail at location $\mathbf{u}$, will be $p(\mathbf{u};k)$ and denoted as: $p(\mathbf{u};s_k) = p(k)$. If more information are given, such as the outcomes of the related surrounding neighboring locations, denoted as $\{Z(\mathbf{u}_\alpha) = z_\alpha, \alpha = 1,...,n\}$ or $\{(n)\}$, the probability distribution function as expressed in (1) will be updated to a posterior probability, which is called conditional probability and can be written as:

$$p(\mathbf{u}_0 \mid \mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n) = prob\{Z(\mathbf{u}_0) = z_0 \mid Z(\mathbf{u}_\alpha) = z_\alpha, \alpha = 1,...,n\} \tag{2}$$

In equation(2), the lower case $\{z_0, z_1, ..., z_n\}$ are the outcomes of the random variables $\{Z(\mathbf{u}_\alpha) = z_\alpha, \alpha = 0,1,...,n\}$, which are $z_\alpha \in [1,2,...K], \alpha = 1,...,n$.

Based on the conditional probability definition, the probability $P(\mathbf{u}_0 \mid \mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n)$ can be calculated from the multivariate probability as:

$$P(\mathbf{u}_0 \mid \mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n) = \frac{P(\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n)}{P(\mathbf{u}_1, ..., \mathbf{u}_n)} \tag{3}$$

In geostatistics, characterization of uncertainty about a spatially distributed phenomenon on the unsampled location is done through conditional simulation on the univariate cumulative distribution function as in equation(2). More over from the view of information gaining, we are more interested in estimating the conditional probability based on the neighboring data as expressed in equation (2) because it will provide more informative estimation for specific location after obtaining some surrounding locations information. Thus, the inference of this conditional probability distribution is the central problem of geostatistics.

There are several established and new development approaches to infer the conditional probability directly or indirectly, such as indicator kriging or multiple point geostatistics. In indicator kriging approach (Goovaerts, 1994; Journel, 1983), the conditional probability multivariate probability is calculated as a linear combination of indicator data:

$$P(\mathbf{u}_0 = z_0 \mid \mathbf{u}_1 = z_1, ..., \mathbf{u}_n = z_n) = \sum_{i=1}^{n} \lambda_i P(\mathbf{u}_0 = z_0 \mid \mathbf{u}_i = z_i) \quad (4)$$

The weights characterize the dependence between the sampled locations $\mathbf{u}_i$ and the information from sampled locations to unsampled location. The weights λ_i are estimated by indicator kriging (IK) based on the indicator covariance.

Theoretically, the conditional probability can be calculated from equation(3), the numerator is an n+1 variates multivariate probability, while the denominator is the n-variates marginal probability of this n-variates multivariate probability. In the multipoint statistics approach, the required n+1 variate multivariate probability is constructed from a training image. Then, the required conditional probability in expression (3) can be calculated by just counting the relatively states number occurred in the training image:

$$P(\mathbf{u}_0 = z_0 \mid \mathbf{u}_1 = z_1, ..., \mathbf{u}_n = z_n\} = \frac{Count(\mathbf{u}_0 = z_0, \mathbf{u}_1 = z_1, ..., \mathbf{u}_n = z_n)}{Count(\mathbf{u}_1 = z_1, ..., \mathbf{u}_n = z_n)} \quad (5)$$

The new developed DMPE approach estimate the conditional probability by estimate the multivariate probability, which is the numerator in equation(3), from the bivariate marginal. In this approach, in this algorithm, the bivariate marginal probabilities are imposed to an initial multivariate probability as the constraints and satisfied by iteratively modifying the multivariate probability. And also in DMPE, the lower marginal probabilities are used as a tool to characterize the information and the redundancy between the sampled locations and the unsampled location.

The theory of all these three methods is not the main point of this paper. The details of indicator kriging approach were written in details in paper (Deutsch, 2006; Goovaerts, 1994; Journel, 1983). The multiple-point approach can be found in(Liu, 2006; Ortiz and Deutsch, 2004; Strebelle, 2002; Wang, 1996). The details of the new developed DMPE approach was given in the paper in this report. In this paper, the estimation results were compared from the informative point between those two mature approaches and the newly developed approach DMPE (Direct Multivariate Probability Estimation).

Spatial statistics for those three methods

In this research, the main point is given an objective comparison with the aim of algorithm improvement. To make a very good estimation, there are many practical details for anyone of them. In order to exclude the practical issue such as variogram model construction, all the spatial statistics are inferred from the same multivariate probability that is obtained by scanning the training image with certain data configuration.

Multivariate probability from training image

Given N categorical random variables, and assuming each of them may have *K* categories $\{k = 1, 2, ..., K\}$ outcomes, these N random variables will form a multivariate probability $P_{mv}(\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_N)$ which is defined as:

$$P_{mv}(\mathbf{u}_1, ..., \mathbf{u}_N) = prob(Z(\mathbf{u}_1), ..., Z(\mathbf{u}_N)); \quad \alpha = 1, ..., N \quad (6)$$

Later, the notation of $P_{mv}(\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_N)$ will refer to the multivariate probability mass function, while $p_{mv}(z_1, z_2, ..., z_N)$ will refer to one probability state, which represents the probability of a specific configuration of category $z_\alpha = k \ with \ (k = 1, ..., K, \alpha = 1, ..., N)$ existing at locations $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_N$. Totally, the probability states number will be K^N. In this research, the original multivariate probability is obtained from scanning a training image with certain data configuration. The training image can be any kinds of categorical variables distribution map which can represent the heterogeneities characters in the spatial space that the geologist expect to see in the research area. It is used most often in multiple point statistics. From this training image, under the stationary assumption, the N point statistics will be retrieved by scanning the data event from the training image.

$$P(\mathbf{u}_1 = z_1, \mathbf{u}_2 = z_2, ..., \mathbf{u}_N = z_N \} = \frac{Count(\mathbf{u}_1 = z_1, \mathbf{u}_2 = z_2, ..., \mathbf{u}_n = z_N)}{Count(\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_N)} \quad (7)$$

The expression(7) characterizes the joint uncertainty about the N actual values $z_1, ..., z_N$. It should satisfy the constraint of: $p_{mv}(z_1, z_2, ..., z_N) \in [0,1]$ and $\sum_{1}^{K^N} p_{mv}(z_1, z_2, ..., z_N) = 1$.

Transition Probability from multivariate probability

Knowing the multivariate probability, the related any lower order marginal probability can be calculated through the multivariate probability marginalization. For example, denote the *n* variate multivariate probability as $P(\mathbf{u}_\ell = z_\ell, \ell = 1, ..., n)$, the second order marginal or bivariate marginal probability

$$P(\mathbf{u}_i, \mathbf{u}_j; z_i, z_j), i, j = 1, 2, ..., N, i \neq j$$

can be calculated from the multivariate probability as:

$$P(\mathbf{u}_i, \mathbf{u}_j; z_i, z_j) = \sum_{\substack{All\ \ell \\ with\, \mathbf{u}_i = z_i\ \&\ \mathbf{u}_j = z_j;\ i \neq j}} P(\mathbf{u}_\ell = z_\ell; \ell = 1, ..., n, z_\ell = 1, ..., K) \quad (8)$$

The bivariate marginal probability satisfy $0 \leq P(\mathbf{u}_i, \mathbf{u}_j; z_i, z_j) \leq 1$ and $\sum_{z_i=1}^{K} \sum_{z_j=1}^{K} P(\mathbf{u}_i, \mathbf{u}_j; z_i, z_j) = 1$.

For any two random variables $(\mathbf{u}_i, \mathbf{u}_j; i, j = 1, ..., N, i \neq j)$, this bivariate probability will be a $K \times K$ matrix which is also called transition probability matrix. For example, if there are 3 locations with possible 3 categories in all the locations, there will be 27 possible multivariate states 3 transition probability matrices as shown in Figure 1.

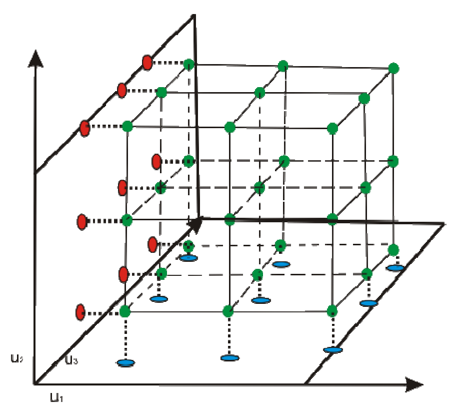

Figure 1 Diagram of multivariate probability and the related bivariate marginal probability (All the multivariate probability states are shown as a green point in the space. The bivariate probabilities between random variable u2 and u3 are shown as a red and those between u1 and u3 are shown as blue circle, the bivariate between u1 and u2 is not shown in this figure).

For the multivariate probability distribution $P_{MV}(\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_N)$, there are totally K^N possible values.

Assuming three facies can be exit at 100 grid locations, the probability spatial space will be 3^{100}, which is a huge number. Thus, it is a big challenge to save and retrieve it in the estimation process. While, the lower order marginal distribution has a lower stationary requirement and can be easily obtained. In DMPE approach, the transition probability is used to reconstruct the original multivariate probability. Thus, the conditional probability can be calculated directly from equation(3).

Covariance from Transition probability

In indicator kriging approach, indicator covariance, which is a bivariate statistics, is used to characterize the spatial relationship. While in DMPE, the transition probability is used. In this part, it will show that the covariance can be inferred from the transition probability with some simplicity assumption. This point will ensure that the differences of conditional probability estimation for unsampled location only come from the algorithm itself.

In indicator kriging approach, a categorical random variable is always transformed to a binary indicator variable. For a categorical variable *Z*(**u**), the indicator variable for category k is defined as:

$$I(\mathbf{u};k)=\begin{cases}1, & if\ k\ exist\ at\ location\ \mathbf{u}\\ 0, & otherwise\end{cases}\qquad k=1,\dots,K \tag{9}$$

The indicator covariance model is usually given by indicator variogram model, which is a measure of the spatial correlation between every two locations. In stationary assumption, the indicator variogram $\gamma_I(\mathbf{h};k)$ for an indicator variable at two locations with departure distance of **h** is defined as:

$$2\gamma_I(\mathbf{h};k)=E\left\{\left[I(\mathbf{u};k)-I(\mathbf{u}+\mathbf{h};k)\right]^2\right\},\quad k=1,\dots,K \tag{10}$$

From indicator variogram calculation as in equation(10), the indicator variogram of category k will only count those categorical transitions that category k changes to other categories from the one location to the other location. From previous session, the bivariate transition probability between any two categories is defined as:

$$p(\mathbf{h};k,k');k,k'=1,\dots,K$$

Thus, for each indicator variogram, it summarizes these 2(K-1) different bivariate transition probabilities (Deutsch, 2005).

$$2\gamma(\mathbf{h};k)=\sum_{\substack{k'=1\\k'\neq k}}^{K}p(\mathbf{h};k,k')+\sum_{\substack{k'=1\\k'\neq k}}^{K}p(\mathbf{h};k',k) \tag{11}$$

For example, for three categories ($k=1,2,3$), the indicator variogram and the transition probability have the relationship as:

$$\begin{cases}2\gamma(\mathbf{h};1)=p(\mathbf{h};1,2)+p(\mathbf{h};1,3)+p(\mathbf{h};2,1)+p(\mathbf{h};3,1)\\ 2\gamma(\mathbf{h};2)=p(\mathbf{h};2,1)+p(\mathbf{h};2,3)+p(\mathbf{h};1,2)+p(\mathbf{h};3,2)\\ 2\gamma(\mathbf{h};3)=p(\mathbf{h};3,1)+p(\mathbf{h};3,2)+p(\mathbf{h};1,3)+p(\mathbf{h};2,3)\end{cases} \tag{12}$$

With the asymmetric assumption ($p(\mathbf{h};1,2)=p(\mathbf{h};2,1)$ and $p(\mathbf{h};1,3)=p(\mathbf{h};3,1)$) and the quality of transition probability matrix ($p(\mathbf{h};1,1)+p(\mathbf{h};1,2)+p(\mathbf{h};1,3)=p_1$):

$$\begin{cases}\gamma(\mathbf{h};1)=p(\mathbf{h};1,2)+p(\mathbf{h};1,3)=p(1)-p(\mathbf{h};1,1)\\ \gamma(\mathbf{h};2)=p(\mathbf{h};2,1)+p(\mathbf{h};2,3)=p(2)-p(\mathbf{h};2,2)\\ \gamma(\mathbf{h};3)=p(\mathbf{h};1,3)+p(\mathbf{h};2,3)=p(3)-p(\mathbf{h};3,3)\end{cases} \tag{13}$$

Where p_1 is the univariate probability for category one. Based on the relationship of $C_1(\mathbf{h})=C_1(0)-\gamma_1(\mathbf{h})$, from equation(12) and(13), the indicator covariance for category 1 can be calculated from the transition probability as**:**

$$\begin{aligned}C_1(\mathbf{h})&=C_1(0)-p_1+p_{11}(\mathbf{h})\\ &=p_1(1-p_1)-p_1+p_{11}(\mathbf{h})\\ &=p_{11}(\mathbf{h})-p_1\bullet p_1\end{aligned} \tag{14}$$

It is the same for category two and three. The indicator covariance can be calculated from equation(14) given a transition probability matrix, and used in the simple indicator kriging directly.

The comparison criteria

The conditional probability $p(\mathbf{u}_0 | \mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n)$ for a given data event can be calculated from full multivariate probability, DMPE or from indicator Kriging. For comparison purposes, it is helpful to have a quantitative measurement to say how informative of an estimation is given the conditioning data. In other words, it is necessary to quantitatively evaluate how much uncertainty has been reduced regarding of this estimation for the unsampled location?

The information content is related to the conditional probability $p(\mathbf{u}_0; k | \mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n)$. Say, if $p(\mathbf{u}_0; k | \mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n) = 1$, it is certain that at location $\mathbf{u}_0$ is category k given the categories at the surrounding locations $(\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n)$, therefore this estimation is very informative on unsampled location. Similarly for the case when $p(\mathbf{u}_0; k | \mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n) = 0$. it is certain that k is not going to happen given the situation at the surrounding locations, hence $p(\mathbf{u}_0; k | \mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n) = 0$ is also very informative estimation on unsampled location. Conversely, if $p(\mathbf{u}_0; k | \mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n) = 1/K$ it is not certain which category is going to happen on unsampled location, hence the information content of this estimation reaches the minimum. In more practical case, we have some prior information about category k, say, its global proportion $p(\mathbf{u}; s_k) = p(k)$ then the lowest informative estimation should be shifted the point where $p(\mathbf{u}_0; k | \mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n) = p(k)$

Based on this understanding, we can make a generalized definition of informative strength function ω_e for the estimation satisfying the following conditions(Liu, 2005):

(1). $\omega_e \in [0,1]$ and ω_e is a function of estimation $\omega_e = f\{p(\mathbf{u}_0; k | \mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n)\}$;

(2). $\omega_e \to 1$ when $p(\mathbf{u}_0; k | \mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n)$ is most informative ($p(\mathbf{u}_0; k | \mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n) \to 1 \; or \; 0$);

(3). $\omega_e \to 0$ when B is not informative ($p(\mathbf{u}_0; k | \mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n) \to 1/K \; or \; p(k)$);

(4). ω_e decrease within $[0, 1/K] \; or \; [0 \; p(k)]$ and increase within $[1/K, 1] \; or \; [p(k), 1]$ moronically;

Linear informative strength definition

Given sampled around locations, the posterior probability (the estimated conditional probability distribution for unsampled location) will be $p(\mathbf{u}_0; k | \mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n)$, which will bring more information regarding to the prior probability $p(k)$ (Global proportion). The informative strength ω_e can be defined as a linear function (Liu, 2005):

$$\omega_e(\mathbf{u}; k | \mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n) = \begin{cases} \dfrac{p(\mathbf{u}_0; k | \mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n) - p(k)}{1 - p(k)} & if \;\; p(\mathbf{u}_0; k | \mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n) \geq p(k) \\ \dfrac{p(k) - p(\mathbf{u}_0; \mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n | d)}{p(k)} & if \;\; p(\mathbf{u}_0; k | \mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n) < p(k) \end{cases} \quad (15)$$

Figure 2 is an example of informative strength function given different prior probabilities.

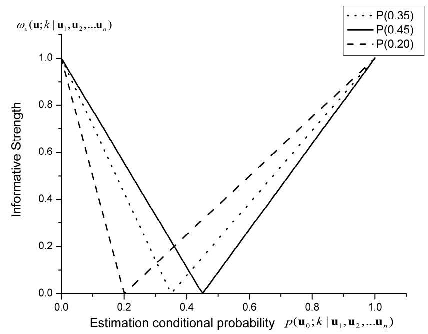

Figure 2 linear informative strength function given three different univariate marginal probabilities

Non-linear informative strength definition

In the information theory(Cover and Thomas, 2006), the uncertainty of knowledge is measured by entropy $H(X)$. The entropy of a random variable *X* with a probability mass function $p(x)$ is defined as:

$$H(X) = -\sum_x p(x)\log(p(x)) \tag{16}$$

The more unpredictable, the higher it's entropy. The entropy reaches its maximum value when X is uniformly distributed, corresponding to minimum informative estimation. It reaches the minimum value when there is no uncertainty about X, i.e. X happens with probability 1 or 0.

Consider the categorical random variable $\mathbf{u}$ is defined on a spatial domain such that all possible outcomes of $(k; k = 1,...,K)$ is the equal probability $p(\mathbf{u}_0;k\,|\,\mathbf{u}_1,\mathbf{u}_2,...\mathbf{u}_n) = 1/K$. The entropy $H_e(\mathbf{u})$ will be:

$$\begin{aligned} H_e(\mathbf{u}) &= -\sum_{k=1}^{K}\frac{1}{K}\ln(\frac{1}{K}) \\ &= -\ln(\frac{1}{K}) \\ &= \ln K \end{aligned} \tag{17}$$

With these uniform probabilities, $H_e(\mathbf{u})$ is the upper bound for the average entropy with lowest informative estimation $p(\mathbf{u}_0;k\,|\,\mathbf{u}_1,\mathbf{u}_2,...\mathbf{u}_n)$ on the unsampled location. Thus, the uniform probability is called least informative probability. Practically, one more information resource is the global proportion of each category $p(k), k = 1,...,K$. Based on this the average entropy would be:

$$H_e(\mathbf{u}) = -\sum_{k=1}^{K} p(k)\bullet\ln p(k) \tag{18}$$

After some conditioning data are obtained, the estimation for unsampled location is updated from the global proportion to a posterior probability distribution denoted as $p(\mathbf{u};k\,|\,\mathbf{u}_1,\mathbf{u}_2,...\mathbf{u}_n)$, for the conditional probability distribution for one unsampled location, the local entropy would be:

$$H_e(\mathbf{u}\,|\,\mathbf{u}_1,\mathbf{u}_2,...\mathbf{u}_n) = -\sum_{k=1}^{K} p(\mathbf{u};k\,|\,\mathbf{u}_1,\mathbf{u}_2,...\mathbf{u}_n)\bullet\ln(p(\mathbf{u};k\,|\,\mathbf{u}_1,\mathbf{u}_2,...\mathbf{u}_n) \tag{19}$$

This new entropy $H_e(\mathbf{u}\,|\,\mathbf{u}_1,\mathbf{u}_2,...\mathbf{u}_n)$ will always be less than $H_e(\mathbf{u})$, says that knowing more informative can only reduce the uncertainty. That is: Information can't hurt (Cover and Thomas, 2006). But it is still need to say how much of uncertainty is reduced after gaining the new information resource. In Bayesian statistics the *KL divergence* is used as a measure of the information gain in moving from a prior distribution to a posterior distribution. *KL divergence* also called relative entropy or Kullback-Leibler distance (KL distance). For a random variable X, with the probability mass functions $p(x)$ and $q(x)$, *KL* distance is defined as:

$$D(p \| q) = \sum_x p(x) \ln \frac{p(x)}{q(x)} \tag{20}$$

In the above definition, $0 \bullet \ln(0/0) = 0$, $0 \bullet \ln(0/q(x)) = 0$ and $p(x) \bullet \ln(p(x)/0) = \infty$.

It is shown that $D(p \| q)$ is always nonnegative and is zero if and only if $p(x) = q(x)$. In information theory, KL distance is used to say how many expected number of bits would have added to the message length by using the original code based on $q(x)$ instead of using a new codes based on the $p(x)$ (Cover and Thomas, 2006).

If the $p(x)$ in equation(20) is the conditional probability $p(\mathbf{u}_0; k \mid \mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_n)$ and $q(x)$ is the prior probability $p(k)$, this therefore represents the amount of useful information, or information gaining, about location $\mathbf{u}_0$ after upgrading the estimation from the global proportion to conditioning probability. The KL distance will be:

$$\begin{aligned} D(p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n) \| (1/K)) &= \sum_{k=1}^{K} p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n) \ln\{\frac{p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n)}{(1/K)}\} \\ &= \sum_k \{p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n) \ln\{p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n)\} - p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n) \ln(1/K)\} \\ &= \sum_k p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n) \ln(K) + \sum_k \{p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n) \ln(p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n))\} \\ &= \ln(K) + \sum_k \{p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n) \ln(p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n))\} \\ &= H_e(\mathbf{u}) - H_e(\mathbf{u} \mid \mathbf{u}_1, \ldots, \mathbf{u}_n) \end{aligned} \tag{21}$$

The larger the distance is, the better the informative, the greater our uncertainty reduced on the unsampled location. So the *KL* distance D can be used as a measure to show how the informative the estimation is after upgrading from a prior probability to posterior probability. The lower bound will be 0, when using the least informative probability as the estimation. The maximum will be 1 when there is no uncertainty for each category.

It is interesting to show that the distance to the least informative probability distributions for binary random variable. At this case, let the logarithms to base 2, the equation(21) will be:

$$\begin{aligned} D(p(\mathbf{u}_0; k \mid \mathbf{u}_1 \ldots \mathbf{u}_n) \| (1/K)) &= \log_2(2) - H(\mathbf{u}_0 \mid \mathbf{u}_1, \ldots, \mathbf{u}_n) \\ &= 1 + p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n) \log_2(p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n)) \\ &\quad + (1 - p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n)) \log_2(1 - p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n)) \end{aligned} \tag{22}$$

The above distance satisfies the previous informative strength function ω_e requirements. When the probability distribution $p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n)$ approaches to the uniform distribution, the informative strength is decreasing, and it increases when $p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n)$ is farther away from the uniform distribution. On both sides, the informative strength ω_e monotonically increases or decreases.

For binary variable, when the prior global probability $p(k), k = 1, \ldots, K$ is given, the lowest point in the informative strength function should be $p(k), k = 1, \ldots, K$, the equation (22) will be modified as:

$$\begin{aligned} D(p'(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n) \| p(k)) &= 1 + p'(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n) \log_2(p'(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n)) \\ &\quad + (1 - p'(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n)) \log_2(1 - p'(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n)) \end{aligned} \tag{23}$$

With

$$p'(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n) = \begin{cases} p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n) / 2p(k) & \text{if } p'(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n) \le p(k) \\ (p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n) - 2p(k) + 1) / 2(1 - p(k)) & \text{if } p(\mathbf{u}_0; k \mid \mathbf{u}_1, \ldots, \mathbf{u}_n) > p(k) \end{cases}$$

Figure 3 are some non-linear informative strength function curves.

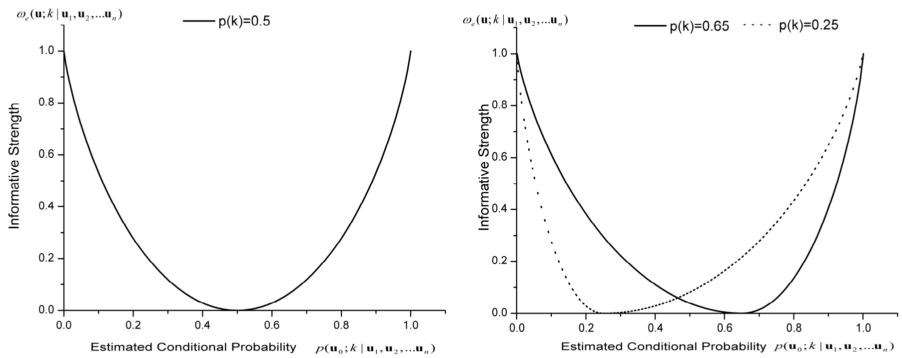

Figure 3 Non-linear informative strength curves given different global marginal probabilities (Left: the binary variable without knowing the univariate probability; Right: after obtaining the univariate probability, the lowest informative strength should shift to the univariate marginal)

There are also some other established approach approaches to compare models, such as cross validation and jackknife. In cross validation, data are left out one at a time and re-estimated from the surrounding data. In the jackknife, a separate set of validation data are held back from the very beginning and only used at the end for checking. More details can be found in (Deutsch, 1999; Goovaerts, 1997).

Comparison methods and comparison results

Because the data configuration will have a sever effect on the estimation results, all efforts should be taken to make sure our comparison is representative. Based on the previous session the following comparison workflow as shown in Figure 4 is adopted.

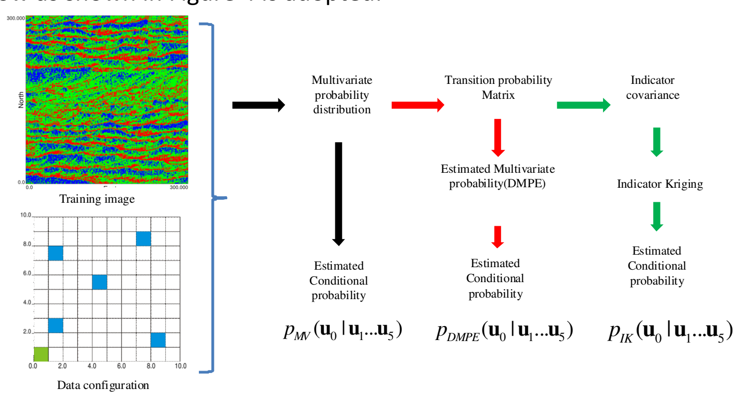

Figure 4 three algorithm comparison workflow

The data configuration as shown in Figure 4 is composed by 6 locations. Sampled data locations are $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4$ and $\mathbf{u}_5$. The location $\mathbf{u}_0$ is unsampled and moving in the data configuration window (10 by 10 grids). Totally, there will be 95 data configurations comparison. The first step is multivariate probability construction which is based on equation(7). Along the red arrow, the transition probability is calculated according equation(8). For indicator Kriging approach, the indicator covariance is coming from transition probability based on(11).

The first comparison could be using the traditional cross validate approach. Using 500 locations in the training image as the crosscheck location, the accuracy plots with those three approaches are plotted in Figure 5. It shows that the DMPE estimation is better than SK approach for depicting the heterogeneity of this training image.

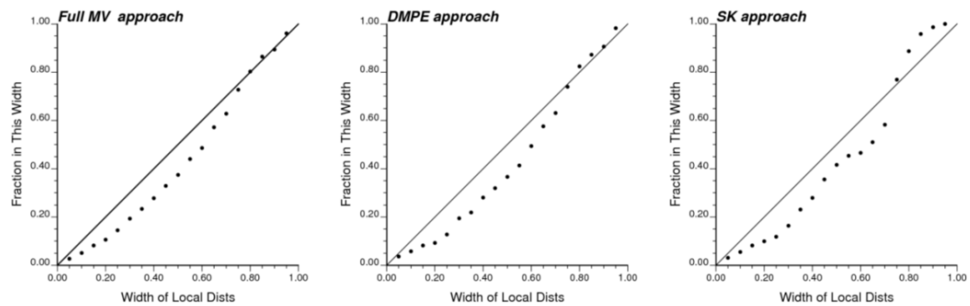

Figure 5 the *accplt* with different estimation approach

From the view of uncertainty, this improvement can be quantitively measured. The linear function of the informative strength ω_e for each category $k, k = 1,2 \text{ and } 3$ is plotted in Figure 6. As shown in Figure 6, the informative strength order for category 1 and 3 is $\omega_e(MV) > \omega_e(DMPE) > \omega_e(IK)$; for category 2 is $\omega_e(MV) > \omega_e(IK) > \omega_e(DMPE)$.

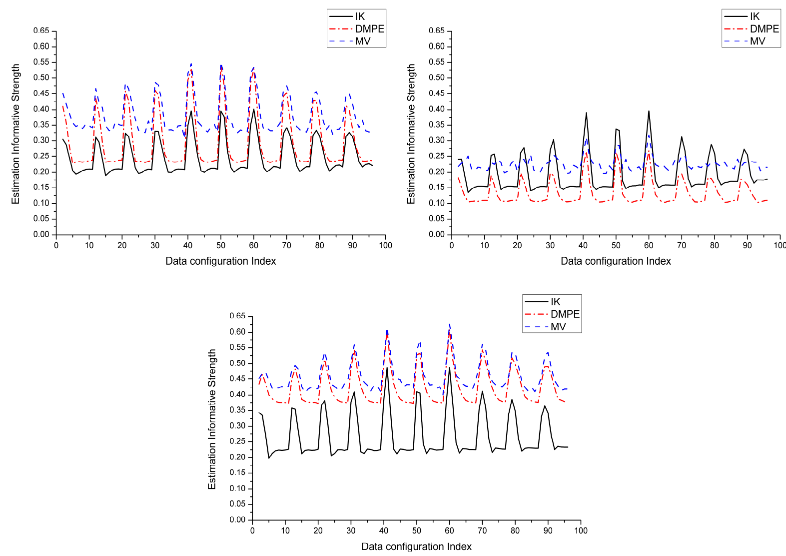

Figure 6 Line Informative Strength comparisons for three categories with different estimation approaches

The informative strength measured by the non-linear approach is shown in Figure 7.

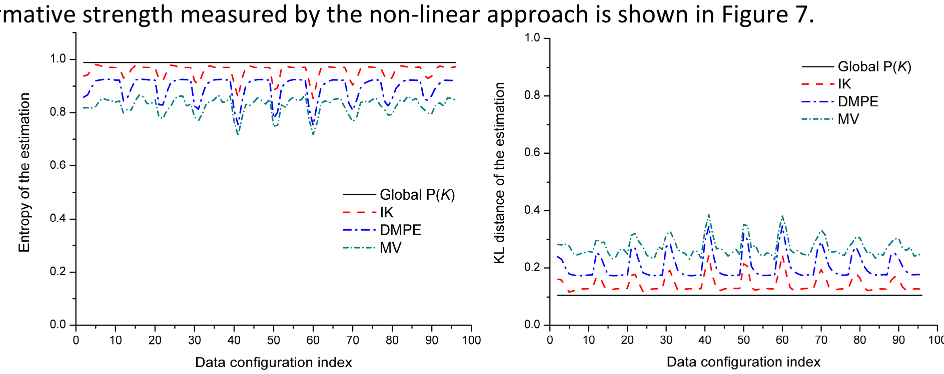

Figure 7 the non-linear uncertainty decreasing using different estimation approaches (Left: entropy; Right: KL distance)

From the KL distance, we can calculate how the estimation is improved from the view of uncertainty reduction. The maximum improvement of informative strength $\omega_e(DMPE) - \omega_e(IK)$ is 10.6%, the minimum of $\omega_e(DMPE)-\omega_e(IK)$ is 4.5%; on average, it is 6.4%. From the informative strength criteria, the estimation improvement on the uncertainty reduction from DMPE is larger than IK, which will encourage a widely application of this new algorithm in the future.

Discussion

Although the multiple point approach has the best informative strength, it is difficult to get enough replications for stable higher order multivariate probabilities because of the huge state space. The stationary assumption in multiple point geostatistics may also cause some difficulty. For IK approach, the linear probability combination approach may not result in models that show the appropriate level of spatial detail as the results shown in this research. It has the minimum informative strength to the unsampled location. The advantage is its computational speed. While for DMPE, the lower marginal probabilities in DMPE will have minimal stationary requirement and the nonlinear combination will be more appropriate in many cases. It will give better informative estimation than IK approach. The spatial statistics tool in DMPE is the transition probability which could be obtained from a training image or vertical profile. Inference of three dimensional transition probabilities is problematic and CPU requirements are large.

Reference

Cover, T.M., and Thomas, J.A., 2006, Element of Information Theory, John Wiley & Sons, Inc.

Deutsch, C.V., 1999, A short Note on Cross VAlidation of Facies Simulation Methods, *in* Deutsch, C.V., ed., Centre of Computational Geostatistics Annual Report, Volume 1: Edmonton, Alberta.

—, 2006, A sequential indicator simulation program for categorical variables with point and block data: BlockSIS: Computers & Geosciences, v. 32, p. 1669-1681.

Goovaerts, P., 1994, Comparative performance of indicator algorithms for modeling conditional probability distribution functions: Mathematical Geology, v. 26, p. 389-411.

—, 1997, Geostatistics for Natural Resources Evaluation, Oxford University Press.

Journel, A., 1983, Nonparametric estimation of spatial distributions: Mathematical Geology, v. 15, p. 445-468.

Liu, Y., 2005, An Information Content Measure Using Multiple-point Statistics, Geostatistics Banff 2004, p. 1047-1056.

—, 2006, Using the Snesim program for multiple-point statistical simulation: Computers & Geosciences, v. 32, p. 1544-1563.

Ortiz, J., and Deutsch, C., 2004, Indicator simulation accounting for multiple-point statistics: Mathematical Geology, v. 36, p. 545-565.

Strebelle, S., 2002, Conditional Simulation of Complex Geological Structures Using Multiple-Point Statistics: Mathematical Geology, v. 34, p. 1-21.

Wang, L., 1996, Modeling complex reservoir geometries with multiple-point statistics: Mathematical Geology, v. 28, p. 895-907.