Methodology and Implementation Details of DMPE

Yupeng Li and Clayton V. Deutsch

In geostatistics, the simulation of categorical variables such as facies or rock types for an unsampled location requires a conditional distribution based on the information from nearby data locations. The conditional probability can easily be calculated from the multivariate distribution of the unsampled value and all data values. Calculating the conditional probability by estimating the joint multivariate probability is theoretically correct. In this paper, a new multivariate probability estimation algorithm, Direct Multivariate Probability Estimation (DMPE), is proposed. The multivariate probability is estimated from lower order marginal probabilities that are known from the available data. The required conditional probability is then obtained directly by its definition. In this algorithm, the bivariate marginal probabilities are imposed to an initial multivariate probability as the constraints and satisfied by iteratively modifying the initial multivariate probability. The bivariate marginal probabilities are inferred from sampled locations, a training image or profiles along drill holes or wells. The basic concepts and the implementation details about this approach are given in this paper.

Introduction

Assuming that $\mathbf{u}_0$ is the location to be predicted, the outcome for location $\mathbf{u}_0$ will be characterized by its probability distribution $P(\mathbf{u}_0)$ which must be inferred from all available data. One case is that we have no knowledge about how this location relates to measured data locations. In this situation, one possible probability distributions for $P(\mathbf{u}_0)$ is the uniform probability distribution:

$$P(\mathbf{u}_0) = \frac{1}{K} \tag{1}$$

It has been shown that this uniform probability distribution is the maximum entropy estimation [1].

Another situation is that only the global proportions $\{p_k, k = 1, \cdots, K\}$ of all the categories are known. Based on the maximum entropy principle, it will also shown that the best estimation for the event that $\mathbf{u}_0 = e_k$ will be this global proportion written as:

$$P(\mathbf{u}_0) = Pr\{\mathbf{u}_0 = e_k; e_k \in E\} = p_k \tag{2}$$

If there are some sample data and they are related to the unsampled location, as illustrated in Figure 1, the probability distribution for the unsampled location would be something else. The information obtained from the neighbouring locations is denoted as $(\mathbf{u}_1 = e_{k_1}, \cdots, \mathbf{u}_n = e_{k_n})$ or simply as (n). The prior probability distribution $P(\mathbf{u}_0)$ will be updated to a posterior probability distribution $P(\mathbf{u}_0|\mathbf{u}_1, \ldots, \mathbf{u}_n)$ which is a conditional probability distribution and will assign a probability $p(\mathbf{u}_0; e_k|(n))$ to the data event $(\mathbf{u}_0 = e_k)$ given that $(\mathbf{u}_1 = e_{k_1}, \cdots, \mathbf{u}_n = e_{k_n})$ is observed.

That is, given the outcomes at the surrounding locations and the outcome at the unsampled location are dependent on the outcomes of the nearby sampled location, the uncertainty of the outcome of the unsampled location will be characterized by a conditional probability distribution that is written:

$$P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n) = Pr\{\mathbf{u}_0 = e_{k_0}|\mathbf{u}_1 = e_{k_1}, \cdots, \mathbf{u}_n = e_{k_n}\} \tag{3}$$

Updating the prior probability distribution as in Equation (1) or (2) to a posterior probability distribution model as in Equation (3) is the central problem in geostatistics [2].

In probability theory [3], a conditional probability as expressed in Equation (3) is calculated as:

$$P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n) = \frac{P(\mathbf{u}_0,\mathbf{u}_1,\ldots,\mathbf{u}_n)}{P(\mathbf{u}_1,\ldots,\mathbf{u}_n)} \tag{4}$$

If the $n+1$ multivariate probability distribution $P(\mathbf{u}_0,\mathbf{u}_1,\ldots,\mathbf{u}_n)$ in the numerator is known, the n multivariate probability in the denominator can be calculated from marginalization of the $n+1$ multivariate probability $P(\mathbf{u}_0,\mathbf{u}_1,\ldots,\mathbf{u}_n)$. The conditional probability $P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)$ will be calculated directly as:

$$
\begin{aligned}
P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n) &= \frac{P(\mathbf{u}_0,\mathbf{u}_1,\ldots,\mathbf{u}_n)}{P(\mathbf{u}_1,\ldots,\mathbf{u}_n)} \\
&= \frac{P(\mathbf{u}_0,\mathbf{u}_1,\ldots,\mathbf{u}_n)}{\sum_{\mathbf{u}_1}\cdots,\sum_{\mathbf{u}_n}\sum_{\mathbf{u}_0=e_1}^{e_K} P(\mathbf{u}_0,\mathbf{u}_1,\ldots,\mathbf{u}_n)}
\end{aligned} \tag{5}
$$

So, calculating the posterior probability $P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)$ through inferring the multivariate probability $P(\mathbf{u}_0,\mathbf{u}_1,\ldots,\mathbf{u}_n)$ is the theoretically correct approach. This $n+1$ multivariate probability distribution may not be easy to infer. One possible simplifying assumption is to consider that the dependence between the $n+1$ random functions can be modeled by a multi-Gaussian probability distribution, with parameters (essentially the covariance matrix) inferred from the data. With the multi-Gaussian assumption, the conditional probability will also be a Gaussian probability and is obtained through estimating its mean and variance [4, 5, 6].

The multi-Gaussian approach assumes all the random variables are Gaussian distribution which is inappropriate for discrete random variables. Secondly, the categorical data are usually nominal data. There are no simple numerical class boundaries for such data. For discrete multivariate probability inference, some specific approaches are proposed in geostatistics. In the following sections, two main geostatistics techniques, indicator kriging and the training image scanning approach, will be reviewed.

Discrete multivariate random variable

The discrete univariate random variable is a useful tool to characterize the uncertainty at one location. Usually we are more interested in observing several spatial locations together.

If a group of locations are measured together, they define a multivariate data event denoted:

$$\omega : \{z(\mathbf{u}_1) = e_{k_1}, \cdots, z(\mathbf{u}_n) = e_{k_n}\} \tag{6}$$

or simplify denoted as: $\{\mathbf{u}_1 = e_{k_1}, \cdots, \mathbf{u}_n = e_{k_n}\}$.

A multivariate data event of size n is constituted by:

- a location geometry defined by the n locations $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$

- the n outcomes at each location $\{e_{k_1}, \cdots, e_{k_n}\}$

The multivariate data event changes with spatial geometry changing or specific outcomes changing at these n locations.

The set of all possible data events for the defined grouped locations will define the data event event space: $\Omega : \{e_{k_1} \in E, \cdots, e_{k_n} \in E\}$. Each multivariate data event will be in this space:

$$\omega_\ell \in \Omega, \ell = 1, \cdots, N$$

where $N = K^n$ is the space dimension.

One example of all the possible data events composed by three locations and three possible facies outcomes for each location is plotted in Figure 2.

Each data event ω_ℓ will have a probability $p(\omega_\ell)$ to happen according to a multivariate probability mass function $P(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ defined as:

$$P(\mathbf{u}_1, \ldots, \mathbf{u}_n) = Pr(\omega_\ell \in \Omega), \ell = 1, \cdots, N \tag{7}$$

This discrete multivariate probability mass function $P(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ will characterize the probability of the joint outcome for the locations $(\mathbf{u}_1, \ldots, \mathbf{u}_n)$. It will satisfy:

$$\sum P(\mathbf{u}_1, \ldots, \mathbf{u}_n) = 1$$

and

$$P(\mathbf{u}_1, \ldots, \mathbf{u}_n) \geq 0$$

Specifically, if only one location $\{\mathbf{u}\}$ is considered at a time, it will become a univariate random variable $Z(\mathbf{u})$ which will have a univariate probability distribution $P(\mathbf{u})$. If two locations are measured together, it will be a bivariate random variable $Z(\mathbf{u}_\alpha, \mathbf{u}_\beta)$ and will be understood by a bivariate probability distribution distribution $P(\mathbf{u}_\alpha, \mathbf{u}_\beta)$. The same for an m-variate multivariate random variable $Z(\mathbf{u}_1, \ldots, \mathbf{u}_m)$, it will be described by an m-variate probability distribution $P(\mathbf{u}_1, \ldots, \mathbf{u}_m)$.

Index function for multivariate data event

For an n-variate multivariate random variable, one feature is its exponentially large event space dimension. For example, assuming there is a model defined by 100 cells and 3

Table 1: Data event space dimension examples for discrete multivariate random variable

Total data locations	Two facies	Three facies	Four facies	Five facies
3	8	27	64	125
4	16	81	256	625
5	32	243	1,024	3,125
6	64	729	4,096	15,625
7	128	2,187	16,384	78,125
8	256	6,561	65,536	390,625
9	512	19,683	262,144	1,953,125
10	1,024	59,049	1,048,576	9,765,625

categories, then the total data event space dimension would be $3^{100} \approx 5 \times 10^{47}$ which is an almost impossibly large number to deal with computationally.

As shown in Table 1, the multivariate data events number increases very fast as the sampled data locations increase. Thus, indexing and tracking the multivariate probability for each of the multivariate data events is a challenge in handling the multivariate probability distribution.

The solution to the indexing challenge is to order the multivariate data events $\{\omega_\ell, \ell = 1, \cdots, N\}$ into a one dimensional array and calculate its index from the outcomes [7]. The first step is ordering and coding all the categories in the set $\{e_1, e_2, \cdots, e_K\}$ into an integer set $\{1, 2, \cdots, K\}$. The order of the categories is arbitrary. Second, calculate the index from the multivariate outcome. The index ℓ for multivariate each data event ω is calculated as:

$$\ell = f(\mathbf{u}_1 = e_{k_1}, \cdots, \mathbf{u}_n = e_{k_n})$$
$$= 1 + \sum_{\alpha=1}^{n} (k_\alpha - 1) \times K^{\alpha-1}, \quad k_\alpha = 1, \cdots, K \tag{8}$$

where the outcome k_α for each location comes from the integer set $\{1, 2, ..., K\}$.

As an example, assuming the spatial problem defined by three locations and three facies $\{channel, levee, floodplain\}$, the total multivariate events number will be $K^n = 3^3 = 27$ in this simple case. In order to index them numerically, those three location will be ordered as $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ and those three categories will be coded as set $\{1, 2, 3\}$. Then using the proposed equation, each multivariate data event index can be calculated from its outcomes using Equation (8). Some indices of multivariate data events are listed in Table 2. The index of each multivariate event can also be found in Figure 2. Each of the data event will have a probability. There will be a total of 27 probabilities.

Multivariate probability distribution marginalization

Different orders of lower order marginal probability distributions could be calculated from a multivariate probability distribution based on the total probability theorem [8].

Univariate marginalization: The univariate probability distribution $P(\mathbf{u}_\alpha)$ character-

Table 2: Multivariate probability event index calculation example

z_1	z_2	z_3	$1 + \sum_{i=1}^{n}(z_i - 1) \times K^{i-1}$	α
1	1	1	$1 + (1-1) \times 3^{1-1} + (1-1) \times 3^{2-1} + (1-1) \times 3^{3-1}$	1
2	1	1	$1 + (2-1) \times 3^{1-1} + (1-1) \times 3^{2-1} + (1-1) \times 3^{3-1}$	2
3	1	1	$1 + (3-1) \times 3^{1-1} + (1-1) \times 3^{2-1} + (1-1) \times 3^{3-1}$	3
1	2	1	$1 + (1-1) \times 3^{1-1} + (2-1) \times 3^{2-1} + (1-1) \times 3^{3-1}$	4
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
2	3	3	$1 + (2-1) \times 3^{1-1} + (3-1) \times 3^{2-1} + (3-1) \times 3^{3-1}$	26
3	3	3	$1 + (3-1) \times 3^{1-1} + (3-1) \times 3^{2-1} + (3-1) \times 3^{3-1}$	27

izes the distribution of a random variable $Z(\mathbf{u}_\alpha)$ and can be calculated from the multivariate probability distribution $P(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ as:

$$P(\mathbf{u}_\alpha) = \sum_{\mathbf{u}_1 = e_1}^{e_K} \cdots \sum_{\mathbf{u}_\alpha} \cdots \sum_{\mathbf{u}_n = e_1}^{e_K} P(\mathbf{u}_1, \ldots, \mathbf{u}_n) \quad \alpha = 1, \cdots, n \tag{9}$$

The univariate marginal probability distribution $P(\mathbf{u}_\alpha)$ will satisfy $0 \leq P(\mathbf{u}_\alpha) \leq 1$ and $\sum P(\mathbf{u}_\alpha) = 1$. This univariate probability distribution will give the probability $p(\mathbf{u}_\alpha = e_k)$ for each facies e_k to prevail at location $\mathbf{u}_\alpha$ that is:

$$P(\mathbf{u}_\alpha) = Pr\{\mathbf{u}_\alpha = e_k; e_k \in E\} \tag{10}$$

If this univariate probability distribution is the same for all locations, say $P(\mathbf{u}_\alpha) = p_k, \alpha = 1, \cdots, n; k = 1, \cdots, K$, it is called the global proportion.

Bivariate marginalization: A second order marginal probability distribution $P(\mathbf{u}_\alpha, \mathbf{u}_\beta), \alpha \neq \beta$ and $\alpha, \beta = 1, \cdots, n$ is calculated from a multivariate probability distribution as:

$$P(\mathbf{u}_\alpha, \mathbf{u}_\beta) = \sum_{\mathbf{u}_1 = e_1}^{e_K} \cdots \sum_{\mathbf{u}_\alpha} \cdots \sum_{\mathbf{u}_\beta} \cdots \sum_{\mathbf{u}_n = e_1}^{e_K} P(\mathbf{u}_1, \ldots, \mathbf{u}_n) \tag{11}$$

The bivariate marginal probability will satisfy

$$0 \leq P(\mathbf{u}_\alpha, \mathbf{u}_\beta) \leq 1$$

and

$$\sum_{\Omega^2} p(\mathbf{u}_\alpha, \mathbf{u}_\beta) = 1$$

The bivariate probability distribution has a probability to each bivariate data event $(\mathbf{u}_\alpha = e_k; \mathbf{u}_\beta = e_{k'})$ that is:

$$P(\mathbf{u}_\alpha, \mathbf{u}_\beta) = Pr\{\mathbf{u}_\alpha = e_k, \mathbf{u}_\beta = e_{k'}; e_k, e_{k'} \in E\} \tag{12}$$

m-variate marginalization: Following the same logic and notation, any m-order ($m \leq$

n) marginal probability mass function from the n-variate multivariate probability mass function can be calculated as:

$$P(\mathbf{u}_1, \cdots, \mathbf{u}_m) = \sum_{\mathbf{u}_1} \cdots \sum_{\mathbf{u}_m} \sum_{\mathbf{u}_{m+1}=e_1}^{e_K}, \cdots, \sum_{\mathbf{u}_n=e_1}^{e_K} P(\mathbf{u}_1, \ldots, \mathbf{u}_n) \qquad (13)$$

In expression (13), the m random variable $(\mathbf{u}_1, \cdots, \mathbf{u}_m)$ is any arbitrary subgroup from the n random variables $(\mathbf{u}_1, \ldots, \mathbf{u}_n)$. Similarly, all the m-variate marginal probability will satisfy the requirements of

$$0 \leq P(\mathbf{u}_1, \cdots, \mathbf{u}_m) \leq 1$$

and

$$\sum_{\Omega^m} P(\mathbf{u}_1, \cdots, \mathbf{u}_m) = 1$$

Strictly speaking, the word "marginal" in describing probability distribution is unnecessary. The same probability distribution can be either "multivariate" or "marginal" depending on the context. Each is the probability distribution of a set of random variables. When two different sets are under discussion and one is a subset of the other, the "multivariate" will indicate the superset and "marginal" will indicate the subset.

The probability of each marginal data event will sum some probabilities of the multivariate data events in the marginalization computation. Using the data event sets in Figure 2 as an example, the univariate marginalization $p(\mathbf{u}_1 = 1)$ can be calculated from the multivariate probability as:

$$p(\mathbf{u}_1 = 1) = \sum_{k_2=1}^{3} \sum_{k_3=1}^{3} p(k_1 = 1, k_2, k_3)$$

Using the index function of Equation (8), the indices of the multivariate probability states that contribute to the univariate probability $p(\mathbf{u}_1 = 1)$ are: 1, 4, 7, 10, 16, 19, 22 and 25. There are 9 univariate marginal probabilities $(b_1, \cdots, b_9)$. All the univariate probability calculations are shown in Figure 3 where $(p_1, \cdots, p_{27})$ are multivariate probabilities.

The bivariate probabilities for each bivariate data event will also be a sum from a subset of the multivariate probabilities of the multivariate data events. For example, the bivariate probability of data event $(\mathbf{u}_1 = 1, \mathbf{u}_2 = 1)$ will be calculated from three multivariate probabilities as:

$$p(\mathbf{u}_1 = 1, \mathbf{u}_2 = 1) = \sum_{k_3=1}^{3} p(k_1 = 1, k_2 = 1, k_3)$$

The indices for those three multivariate data events are: 3, 6 and 9 calculated using Equation (8). The same could e applied to other bivariate probability data events. For this small example, the bivariate probabilities $(b_1, \cdots, b_{27})$ calculated from the multivariate probabilities $(p_1, \cdots, p_{27})$ are shown in Figure 4.

The above marginalization procedure is also illustrated in Figure 5. In this figure, all the multivariate probability are expressed as round filled circles that are arranged in a cube. The univariate probability, the square shape, will be the sum of nine trivariate probabilities which are shaded together in one slice of this cube. In the image, only one univariate probability is plotted. The bivariate probabilities, plotted with diamond symbols, will be the sum of three of the trivariate probabilities. In this image, only the bivariate probabilities between two variables are plotted.

Illustrating a trivariate probability distribution marginalization using a cube is just for visualization. For a higher order multivariate probability space, it will be a hypercube [9]. Intuitively plotting them in a picture as Figure 5 is a challenge.

Direct multivariate probability estimation

A novel method of multivariate probability estimation in information theory research is implemented in spatial multivariate probability estimation. It is estimated from the known bivariate marginal probability of the desired multivariate probabilities and based on the minimum Kullback-Leibler divergence principle which is an extension of the Maximum Entropy principle [10, 11, 12].

Based on the concept of KL divergence, Kullback proposed the principle of Minimum KL divergence: given new facts, a new distribution $\boldsymbol{p}$ should be chosen that is as hard to discriminate from the original distribution $\boldsymbol{q}$ as possible; so that the new data produces as small an information gain $J(\boldsymbol{p}||\boldsymbol{q})$ as possible, thus no more bias except satisfying the constraints are introduced [13]. It is written as:

$$\text{Minimize:} \quad J(\boldsymbol{p} \parallel \boldsymbol{q}) = \sum_{\ell=1}^{N} p_\ell \log \frac{p_\ell}{q_\ell} \tag{14}$$

subject to:

$$\sum_{\ell=1}^{N} p_\ell = 1 \tag{15}$$

$$\sum_{\ell=1}^{N} a_{m\ell} p_\ell = b_m; m = 1, 2, \cdots, M \tag{16}$$

It is said that if the constraints of Equation (15) and (16) are consistent, there exists a unique probability distribution $\hat{\boldsymbol{p}} : \{\hat{p}_1, \cdots, \hat{p}_N\}$ that satisfies them and is of product form as:

$$\hat{p}_\ell = q_\ell \cdot \mu \prod_{m=1}^{M} \mu_m^{a_{m\ell}} \tag{17}$$

where q_ℓ is the known probabilities; $a_{m\ell} = 0, \text{or}, 1$ are given constant; $\mu, \{\mu_m\}$ are determined from the constraints in Equation (15) and (16). Specifically, this unique probability

distribution $\hat{\boldsymbol{p}}$ is the limit of the iterative sequence $\{\mathbf{p}^{(\delta)}; \delta = 0, 1, 2, \cdots\}$ defined by:

$$\hat{p}_\ell^{(0)} = q_\ell$$

$$\hat{p}_\ell^{(\delta+1)} = \hat{p}_\ell^{(\delta)} \prod_{m=1}^{M} \left[\frac{b_m}{\hat{b}_m^{(\delta)}} \right]^{a_{m\ell}} \quad \delta = 0, 1, 2, \cdots \quad (18)$$

where $\hat{b}_m^{(\delta)} = \sum a_{m\ell} \hat{p}_\ell^{(\delta)}$

From the minimum KL divergence principle, the probability distribution $\boldsymbol{p}$ in the form (17) satisfying (18) will be unique. One proof to this point can be found in Darroch [12]. The proposed approach written in pseudo-code is:

Begin

Input the given lower order marginal probability $b_m, m = 1, \cdots, M$ which is used as the constraints of estimation;

Initialize: *Generate an initial multivariate probability $p_\ell^0, \ell = 1, \cdots, N$;*

Repeat:

1. Calculate a current marginal probability $\hat{b}_m^{(\delta)} = \sum_{m=1}^{M} a_{m\ell} \hat{p}_\ell^{(\delta)}$ with the current estimated multivariate probability $\hat{\mathbf{P}}^\delta$;

2. Calculate a modification factor from the target marginal probability and the estimated marginal probability as: $f_m = \frac{b_m}{\hat{b}_m^\delta}$;

3. Update multivariate probability with the modify factor: $\hat{\mathbf{P}}^{\delta+1} = (\mathbf{F} \times \hat{\mathbf{P}}^\delta)$, where $\mathbf{F} = \{f_m, m = 1, \cdots, M\}$;

4. Set $\hat{\mathbf{P}}^{\delta+1}$ to $\hat{\mathbf{P}}^\delta$ for next iteration;

Until: *the multivariate probability satisfies:* $\Delta = \| \hat{\mathbf{P}}^{\delta+1} - \hat{\mathbf{P}}^\delta \| \approx 0$;

Output: *the final iteration result;*

End

Originally, the above iterative scaling approach is used to infer a matrix with non-negative entries when the row and column sums are known. It was first proposed by Deming and Stephan [14] to infer a two dimensional distribution with known marginal using the empirical distribution of the observed sample "contingency table" as prior guess. Later on, Chow and Liu [15] pointed out that this approach is suitable for binary multivariate random variables estimation based on the second-order distribution with the first-order tree dependency. They pointed out the difficulty to expend to higher order of discrete multivariate random variables.

Program implementation

The proposed stochastic facies modelling methodology is implemented in a program named as `DMPEsim` following the Gslib codes style. Here are several programming implementation notes.

Table 3: CPU time for different model size and number of conditioning data (three categories for all cases)

		Cell numbers in the model			
		50*50	100*100	200*200	300*300
Conditioning data number	4	18.557(s)	73.335(s)	292.039(s)	444.229(s)
	5	95.227(s)	387.967(s)	1522.292(s)	2288.040(s)
	6	486.33(s)	1981.53(s)	8141.99(s)	11585.99(s)
	7	2568.16(s)	10180.37(s)	38204.42(s)	56106.86(s)
	8	23224.71(s)	> 10(h)	> 15(h)	> 15(h)
	9	> 10(h)	> 15(h)	> 15(h)	> 15(h)
	10	> 15(h)	> 20(h)	> 25(h)	> 35(h)
	11	> 20(h)	> 25(h)	> 25(h)	> 35(h)
	12	> 25(h)	> 25(h)	⋯	⋯
	13	⋯	⋯	⋯	⋯

Conditioning data

First is the number of conditioning data to be retained. As already discussed, no more than 11 data is suggested for most cases because of the huge dimension of the full multivariate probability distribution. If the number of facies is more than three, the maximum number of conditioning data should be less than 11 which is already hard coded in the program. The CPU time listed in Table 3 shows how long it takes the program to run on some different size models.

The computer used here has a Xeon 3.19 GHz CPU and 3.00 GB of RAM. As shown, the feasible situations, which are the gray cells in Table 3, will provide a plausible combination for spatial estimation/simulation.

Conditioning data searching

As only a limited number of conditioning data can be used in the simulation, the selection of them will make difference. In the traditional sequential simulation, a local neighbourhood is defined by specifying a search ellipsoid that corresponds to the principle ranges and directions of continuity based on the variogram. Data beyond these ranges will have a limited effect in calculating the posterior conditional probability. Usually, the spiral searching technique is used to search the data on a regular grid until the maximum number of data have been located or the search radius is reached [2]. The same searching strategy is defined in the implementation of the proposed DMPE algorithm. If the data shows some degree of anisotropy, the search area will be elliptical to account for the different ranges of continuity along different directions.

As the maximum number is 11, the basic requirement is that these 11 data should represents the information from all surrounding directions. Thus, the octant searching is adopted as shown in Figure 6. The search will mainly pick the needed conditioning data evenly from the eight octant in 2D and 3D.

Random path and multiple grid

In classical sequential simulation, a random path is followed in the program in order to avoid possible artifacts. In this research, a pseudo-random path is built that considers the original data configuration. During the simulation, all the unsampled locations are ordered according to the number of nearby conditioning data. The one mostly informed will be picked for simulation in each step. For those locations with the same number of conditioning data, one will be picked randomly.

Also, as the simulation proceeds, more and more conditioning data will be obtained within a very close distance. Only using the nearby conditioning data may not reproduce the heterogeneity structure at longer distances. To correct this problem, the multiple grid approach is implemented in order to reproduce longer spatial structures [16]. That is, to simulate the N locations in two or more steps. To start, a coarse grid is simulated to help capture the large-scale features in the model by working with conditioning data over long distances. This grid is then reduced in several steps until the final model grid size is reached. In each simulation grid, a pseudo-random path is used.

Iteration time in DMPE

As the iterative approach is used in the multivariate probability distribution estimation for each unsampled location, the iterative number is also a key parameter to the CPU time. In practice, usually 20 to 25 iterations will permit convergence to the final solution and is enough for most situations.

Conclusion and future works

The proposed DMPE will provide a more accurate and precise posterior probability result than the IK approach under the same conditioning situation. But IK approach can use more conditioning data. Each of them has it own advantages. Although there are some practical computational constraints, it is practical when the grid size of the simulation domain is small and there are relatively few facies types.

At present the conditioning data in the multivariate probability estimation is limited by CPU capacity. The main reason is the huge dimension of the multivariate data event space. Saving the probability of each data event and retrieving them for marginalization brings high computation burdens. One possible future work is finding a more efficient approach that would allow more conditioning locations to be used in the multivariate probability estimation.

References

[1] E. T. Jaynes, "Information theory and statistical mechanics," *Physical Review*, vol. 106, no. 6, pp. 620–630, 1957.

[2] C. V. Deutsch and A. G. Journel, *GSLIB: Geostatistical software library and user's guide.* Oxford University Press, 1998.

[3] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd ed. New York: McGraw-Hill Inc., 1984.

[4] G. Verly, "The multigaussian approach and its applications to the estimation of local reserves," *Mathematical Geology*, vol. 15, no. 2, pp. 259–286, Apr. 1983. [Online]. Available: http://dx.doi.org/10.1007/BF01036070

[5] J. J. Gomez-Hernandez and X.-H. Wen, "To be or not to be multi-gaussian? a reflection on stochastic hydrogeology," *Advances in Water Resources*, vol. 21, no. 1, pp. 47 – 61, 1998. [Online]. Available: http://www.sciencedirect.com/science/article/B6VCF-3SVY576-F/2/78ae9aa0d4f15cf12c14cc37173e2151

[6] P. Goovaerts, "Geostatistical modelling of uncertainty in soil science," *Geoderma*, vol. 103, no. 1-2, pp. 3–26, september 2001. [Online]. Available: http://www.sciencedirect.com/science/article/B6V67-43MJS2V-2/2/f53eebe1835c54a9ad8a67c7c20e08fd

[7] C. V. Deutsch, "Annealing techniques applied to reservoir modeling and the integration of geological and engineering (well test) data," Ph.D. dissertation, Stanford University, Stanford, CA, 1992.

[8] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis.* Englewood Cliffs, NJ: Prentice Hall, 1982.

[9] C. A. Pickover, *Surfing through Hyperspace: Understanding Higher Universes in Six Easy Lessons.* Oxford University Press, 1999.

[10] I. J. Good, "Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables," *The Annals of Mathematical Statistics*, vol. 34, no. 3, pp. 911–934, 1963. [Online]. Available: http://www.jstor.org/stable/2238473

[11] S. Kullback and M. Khairat, "A note on minimum discrimination information," *The Annals of Mathematical Statistics*, vol. 37, no. 1, pp. 279–280, 1966.

[12] J. N. Darroch and D. Ratcliff, "Generalized iterative scaling for log-linear models," *The Annals of Mathematical Statistics*, vol. 43, no. 5, pp. 1470–1480, 1972. [Online]. Available: http://www.jstor.org/stable/2240069

[13] S. Kullback, *Information Theory and Statistics.* Dover Publications, 1968.

[14] W. E. Deming and F. F. Stephan, "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known," *The Annals of Mathematical Statistics*, vol. 11, no. 4, pp. 427–444, December 1940. [Online]. Available: http://www.jstor.org/stable/2235722

[15] C. Chow and C. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 462–467, May 1968.

[16] T. T. Tran, "Improving variogram reproduction on dense simulation grids," *Computers & Geosciences*, vol. 20, no. 7-8, pp. 1161–1168, 1994. [Online]. Available: http://www.sciencedirect.com/science/article/B6V7D-48C7910-4C/2/25ba815d46dc62c051a39c5201d6edad
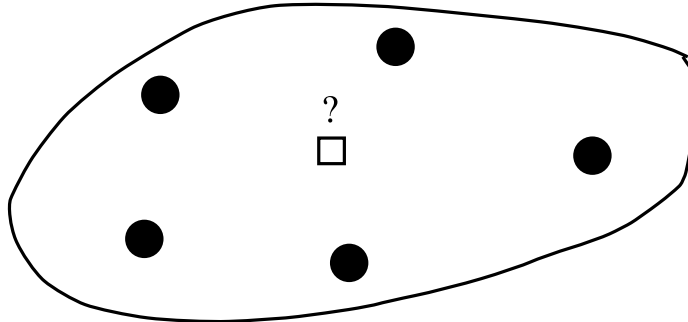
Figure 1: One situation of multivariate probability needed in spatial probability mapping
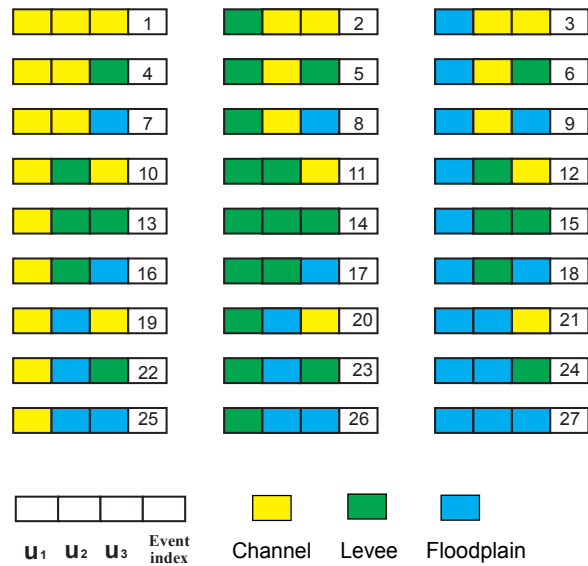
Figure 2: One example of all the possible data events composed by three spatial locations and three possible facies

$$
\begin{bmatrix}
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
\times
\begin{bmatrix}
p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \\ p_7 \\ p_8 \\ p_9 \\ p_{10} \\ p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \\ p_{15} \\ p_{16} \\ p_{17} \\ p_{18} \\ p_{19} \\ p_{20} \\ p_{21} \\ p_{22} \\ p_{23} \\ p_{24} \\ p_{25} \\ p_{26} \\ p_{27}
\end{bmatrix}
=
\begin{bmatrix}
b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \\ b_8 \\ b_9
\end{bmatrix}
$$

Figure 3: A univariate marginalization from a trivariate probability

$$
\begin{bmatrix}
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
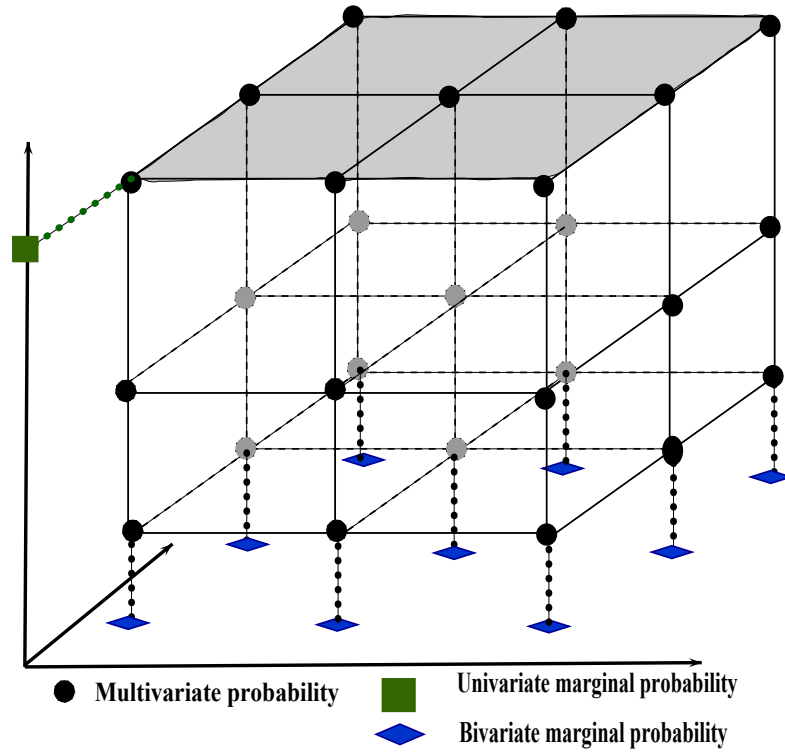0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1
\end{bmatrix}
\times
\begin{bmatrix}
p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \\ p_7 \\ p_8 \\ p_9 \\ p_{10} \\ p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \\ p_{15} \\ p_{16} \\ p_{17} \\ p_{18} \\ p_{19} \\ p_{20} \\ p_{21} \\ p_{22} \\ p_{23} \\ p_{24} \\ p_{25} \\ p_{26} \\ p_{27}
\end{bmatrix}
=
\begin{bmatrix}
b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \\ b_8 \\ b_9 \\ b_{10} \\ b_{11} \\ b_{12} \\ b_{13} \\ b_{14} \\ b_{15} \\ b_{16} \\ b_{17} \\ b_{18} \\ b_{19} \\ b_{20} \\ b_{21} \\ b_{22} \\ b_{23} \\ b_{24} \\ b_{25} \\ b_{26} \\ b_{27}
\end{bmatrix}
$$

Figure 4: The bivariate marginalization from a trivariate probability

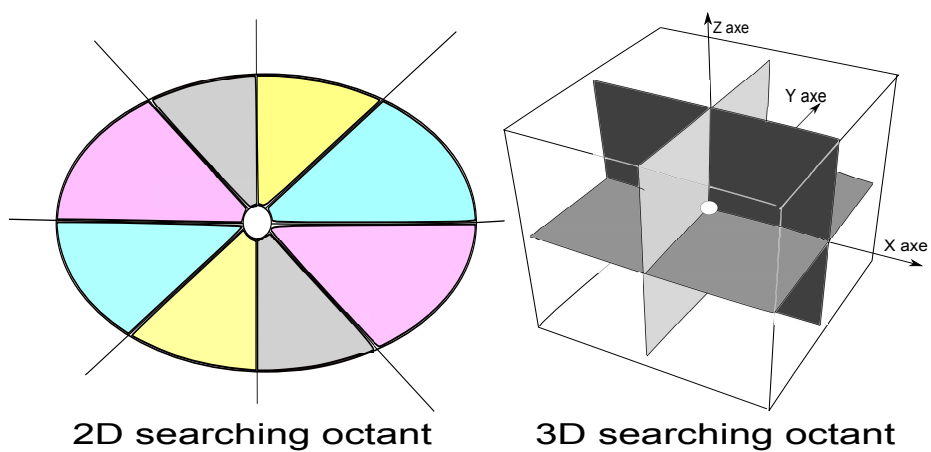**Figure 5:** Univariate and bivariate marginalization from a trivariate probability space

Figure 6: Octant searching in DMPE program implementation