Facies Modeling with Direct Multivariate Probability Estimation

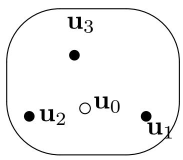
Yupeng Li and Clayton V. Deutsch

In this paper, a new facies modelling procedure using the direct multivariate probability estimation is proposed. In this facies modeling procedure, the bivariate probability diagram is used as the spatial heterogeneity characterization tool. The conditional probability for the unsampled location is calculated from directly estimate the multivariate probability with the constraints of the full set of bivariate probability of the conditioning data set. The mains steps of the facies modeling work flow with the proposed method are presented in this paper.

1 Introduction

Petroleum reservoir management requires numerical models of reservoir heterogeneity. Most geostatistical practitioners agree that the major heterogeneity is caused by facies due to different sedimentary environment or different digenetic process after deposition. The most popular geostatistics techniques can be classified as the cell-based and object-based approaches. Usually, in cell-based reservoir modelling, the model is defined by a group of cells or grids defined in two or three dimensions. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be a group of locations in the model and let E is a set of all the possible categories in the model and is composed as $\{e_1, \cdots, e_K\}$ which are mutually exclusive and exhaustive in the sense that only one category can be find for one specific location $\mathbf{u}_\alpha$.

A location that is not perfectly known and needs to be predicted is usually called "unknown", "unsampled" or "unclassified" and is denoted as $\mathbf{u}_0$. Usually, there is some information about the outcomes for this unsampled location such as the global probability of each category to exist in the domain. Also, some information sources such as the relationsip of the sampled locations $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ related to the unsampled location$\mathbf{u}_0$. These spatial related information are characterized by some spatial characterization tools. $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$.



Traditionally, probability theory is used to give a quantitative characterization of the information coming from the data. For example the information from each sampled location to the unsampled location is usually defined as a conditional probability as $p(\mathbf{u}|\mathbf{u}_i), i = 1, \cdots, n$. Thus, the problem is given n information resource $p(\mathbf{u}|\mathbf{u}_i), i = 1, \cdots, n$, how to infer the conditional probability $p(\mathbf{u}_0|\mathbf{u}_1, \ldots, \mathbf{u}_n)$ to express our uncertainty on unsampled location given the information from the sampled locations. This probability is the core problem in geostatistics. After obtaining the conditional probability distribution $p(\mathbf{u}_0|\mathbf{u}_1, \ldots, \mathbf{u}_n)$, the outcomes probability at location $\mathbf{u}_0$ given the outcomes at location $\mathbf{u}_1, \ldots, \mathbf{u}_n$ has occurred will be charactered by this conditional probability.

There are some well established approaches to estimate $p(\mathbf{u}_0|\mathbf{u}_1, \ldots, \mathbf{u}_n)$ in geostatistics such as Indicator Kriging which is based the indicator variogram and the multipoint geostatistics which is based on training image. Although those two are used widely to infer the conditional probability for categorical variable, there are some constraints in them. In Indicator Kriging approach, after doing an unlinear transforming the original spatial random variable to a binary random variable, the Kriging approach is used to combine the information from sampled locations together. Kriging is a linear approach and may not be appropriate for probability estimation. While for the multiple point geostatistics, the required multivariate probability is obtained through scanning a training image. The strongly stationary assumption of this method brings some challenges to broad application.

Theoretically, the definition of a conditional probability distribution $p(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)$ itself is

$$P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n) = \frac{P(\mathbf{u}_0,\mathbf{u}_1,\ldots,\mathbf{u}_n)}{P(\mathbf{u}_1,\ldots,\mathbf{u}_n)} \tag{1}$$

Where $P(\mathbf{u}_0,\mathbf{u}_1,\ldots,\mathbf{u}_n)$ is the $(n+1)$ multivariate probability, $P(\mathbf{u}_1,\ldots,\mathbf{u}_n)$ is the n^{th} order of the marginal probability of the estimated multivariate probability $P(\mathbf{u}_0,\mathbf{u}_1,\ldots,\mathbf{u}_n)$. The n^{th} marginal probability can be calculated as:

$$P(\mathbf{u}_1,\ldots,\mathbf{u}_n) = \sum_{\mathbf{u}_0} P(\mathbf{u}_0,\mathbf{u}_1,\ldots,\mathbf{u}_n) \tag{2}$$

Merging Equation (2) and Equation (1) permits the conditional probability to be calculated directly after knowing the multivariate probability as:

$$P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n) = \frac{P(\mathbf{u}_0,\mathbf{u}_1,\ldots,\mathbf{u}_n)}{P(\mathbf{u}_1,\ldots,\mathbf{u}_n)} = \frac{P(\mathbf{u}_0,\mathbf{u}_1,\ldots,\mathbf{u}_n)}{\sum_{\mathbf{u}_0} P(\mathbf{u}_0,\mathbf{u}_1,\ldots,\mathbf{u}_n)} \tag{3}$$

Then the best way to estimate the conditional probability in Equation (1) could be to directly estimate the the $(n+1)$ multivariate probability. That is exactly the aim of this research. Explicitly estimating the multivariate probability distribution is proposed in this paper. It is based on the principle of Minimum Relative Entropy (MRE) which is a general form of the maximum entropy principle. The maximum entropy principle is a well established and often used approach in information theory[1, 2]. While its application in geostatistics is limited to the Bayesian Maximum Entropy approach [3, 4]. The details of that point is discussed in paper 122 of this volume. The rest of this paper is constructed as following sections. First, the spatial Markov model is introduced in this paper which is the basis of simulation with the bivariate probability constructed multivariate probability. Finally the simulation procedure with the new proposed multivariate probability estimation is illustrated with a case study.

2 Spatial Markov Model

In traditional Markov stochastic process, under the stationary assumption, the one- step single-dependence transition matrix is enough to character the stochastic process. That is to say, the process is characterized by the one-step single- dependence transition matrix where the current state is dependent only upon one preceding state. In practice, one-step single dependence Markov model is often sufficiently and useful representation for descriptive purposes in a stationary case in one dimensional case[5, 6, 7, 8].

The perfect theory in one dimension and stationary case of Markov model spur many efforts to extend it to two or three dimensions[9]. In spatial geological environment, the outcome for one location usually depends on many surrounding locations. More previous states can be produce influence to the current location especially in two dimension or three dimension. In order to obtain more geological realistic simulation results, more outcomes from previous locations should be counted in when doing simulation.

In Markov chain theory, the multi-step and multi-dependence matrix could used in two or three dimension [10]. The multi-step and multiple-dependence is defined as a stochastic process whose dependency relationships involve more than one preceding state as $P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)$. That is, the outcome for current location would depend on more than one previous locations (multi-dependence), and also these n dependent locations could be any steps from the current location, especially in two or three dimensions (multi-step).

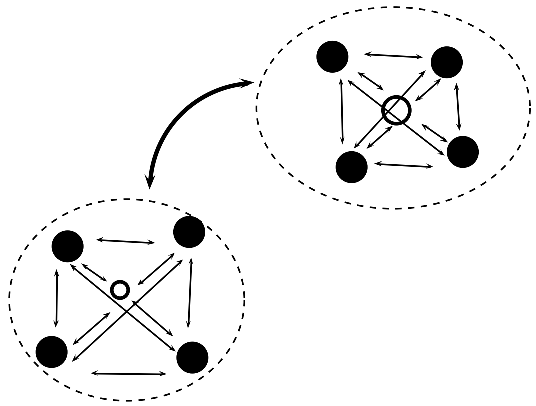


Figure 1: Illustration of a multiple steps and multiple dependence Markov random function field model

It is assumed that one multi-step and multi-dependence Markov chain exist in this space and may move or jump within the space randomly or along a predefined path. It will obey the same or different transition probability rules in different directions and having its states at any unknown location entirely depending on its nearest known neighbours found in different locations. This means that the stochastic structure will keep the same throughout all the locations of the spatial space[11].

All the surrounding sampled locations could bring information to the current unsampled location and are also related to each other. That is to say, the posterior probability will be a function between all the bivariate statistics, written as:

$$P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n) = f(p(\mathbf{u}_i,\mathbf{u}_j), i,j = 0,1,\cdots,n) \tag{4}$$

Comparing the definition of multiple-dependent transition probability and the posterior probability definition, they are exactly the same. The conditional probability estimation in equation (1) can be obtained through direct estimation of the multivariate probability $P(\mathbf{u}_0,\mathbf{u}_1,\ldots,\mathbf{u}_n)$ and using the definition in Equation (3). The multivariate probability can be used to calculate the multiple-dependence from single-dependence along vertical sequence and do vertical stacking pattern simulation. From the lower order bivariate probability, any order of higher order multivariate can be obtained the proposed algorithm DMPE. That is the multiple dependent and multiple-step transition probability matrix can be estimated from all its surrounding locations.

In reality, constrained by the computing power, one more assumption implemented is that the closer neighbours would screen the information in further locations. This assumption could also be found in the traditional sequential simulation technique. Using the same assumption, in Markov multiple-dependence model, it is believed that the current location will be dependent to the nearby locations in certain area which will screen the further data effect[12].

3 Cell-Based Facies Modeling

In geostatistics, characterization of uncertainty about a spatially distributed phenomenon on the unsampled location is done through estimates of conditional probabilities. Using the conditional probability calculation approach as introduced in Equation (1), the conditional probability map for each category can be obtained as one example shown in Figure 2.

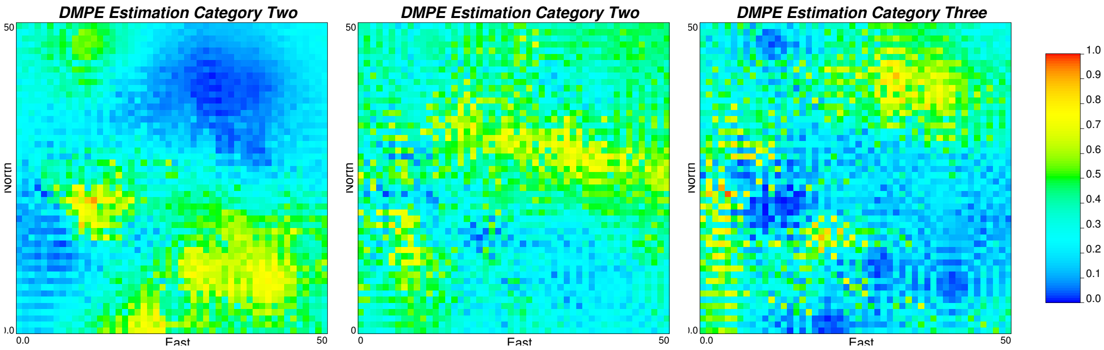


Figure 2: One estimation example results based on the DMPE

In Figure 2, at each time, only one unsampled location is considered and the conditional probability is calculated for this unsampled location, a map of probability of each facies in all the locations are obtained. This process is called interpolating or "estimation".

Another technique is called " conditional simulation". In this case, the simulated facies using the Monte-Carlo draw from the estimated conditional probability are plotted as a maps, where a facies category is presented at every grid cell. One advantage of simulation is that the facies map obtained from simulation would be used as a constraint for next steps in modeling. More importantly, the simulation can characterize the joint uncertainty together to reflect the spatial variability. The "conditional" means that the hard data are reproduced exactly.

The introduced spatial Markov simulation model in Section Two is very similar to the traditional sequential simulation model used in classical geostatistics[12]. In sequential simulation, each realization is one drawing from its multivariate probability function. Ths same as the sequential simulation principle, the spatial Markov simulation model is theoretically valid with no approximation or assumption, but in practice, as the conditioning data increases, the dimension of multivariate probability space will be very huge. In implementation, the Markov assumption is adopted to solve the size problem. It is assumed, the nearest data will screen the far location data's effect. Then for each unsampled location, only limited surrounding conditioning data are used.

The cell-based spatial Markov simulation model would proceed as all simulation grid cells are visited only once along a random path and simulated cell values become conditioning data for cells visited later in the sequence. Building up the reservoir or ore body model usually will include the following steps:

1. Defining the lithotypes or facies. The facies constituting the reservoir have to be defined from the available data: core samples, well logs and seismic data. They have to honour the geological information. In general, the geologists define more facies than could be reasonably simulated. As considering the probability space constraints, the facies type should be grouped into no more than five categories.

2. Geological analysis and bivariate probability modeling. From the vertical core analysis, obtain a pattern and believe this pattern will prevail in certain sedimentary units which may be defined by sequence stratigraphic surface. Inferring the bivariate probability from vertical direction. Define the dip and strike direction and the anisotropy ratio along these three directions which is also the geologic understanding about this modeling domain. Those geological understanding will be integrated into the bivariate probability scaling in the modeling process. The details on bivariate modeling can be found in paper 123 of this volume.

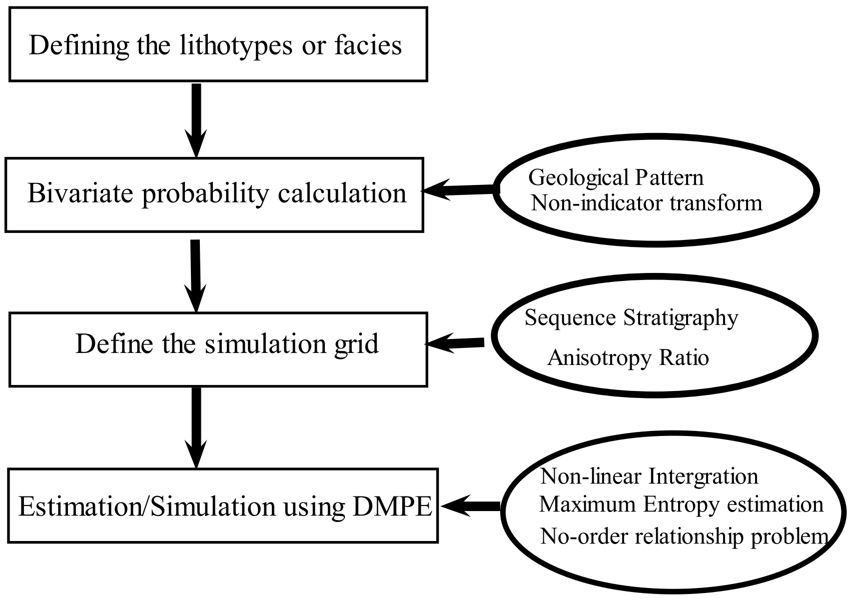


Figure 3: The main work flow estimation/simulation with DMPE and characters of each step

3. Define the simulation grid. Generally, a regular orthogonal simulation grid is adopted in most geological algorithm design. In this research, as the bivariate statistics constrained by the sequence stratigraphic surface, it is expected that the grid of x-y plane is aligned to the interpreted equal time surface. The reference level for the simulation is a specific geological layer which is used to restore the geometry of the reservoir at the time of deposition. The level have been deposited horizontally during sedimentation and should, if possible, correspond to a time line.

4. Doing sequential simulation for each unsampled location. For each unsampled location, the spatial relationship is characterized by the distance and relative direction along the dip and strike direction. The bivariate probability will be obtained from the vertical transition profile.

   (a) Look for the n conditioning data (original well data or previously simulated cell values) closest to current unsampled location $\mathbf{u}_0$;
   (b) Based on the distance between every two locations, retrieve bivariate probability from the experimental bivariate probability diagram;
   (c) Using the proposed DMPE algorithm to estimate the multivariate probability;
   (d) Using the conditional probability definition to build the local cumulative conditional probability at $\mathbf{u}_0$ as in Equation(1);
   (e) Draw a simulated facies value using Monte-Carlo sampling, and assign that value to the grid cell $\mathbf{u}_0$;
   (f) Go to the next unsampled location until all the unsampled location are visited.

The main steps of the work flow with DMPE can be illustrated with the work flow in Figure 3. The spatial Markov simulation model is very similar to the sequential simulation model. One difference compared to the traditional Indicator simulation is that the facies simulated from DMPE are all done in one multivariate probability estimation, there is no order relation problems as in the indicator approach. Another difference is the non-linear character in the multivariate probability estimation process. All the bivariate probability/transition probability are integrated together based on the Maximum Entropy principle.

4 Conditional Simulation

Practically, an experimental bivariate probability is computed from the data set using the same practical constraints as for indicator variogram. The data set could be the training image, well vertical profiles or from outcrops.

For example, from the training image in Figure 3 of paper 123 in this volume, along the north to south direction, as the distance between location pair $(\mathbf{u}_i, \mathbf{u}_j)$ apart from each other to a further distance, the bivariate probability between facies 1, 2 and 3 will compose a diagram as shown in Figure 3 in paper 123 . Traditionally, it is needed to find a function to these experimental bivariate probabilities so that a bivariate probability is available for any distance vector, which will be difficult to keep it as a positive definition function. Because the experimental bivariate probability is consistent with each other, it is used directly in this simulation example. Also because the experimental bivariate probability diagram is a discrete diagram, the distance between every two locations is rounded to the closest lag distance in the simulation process.

Using the calculated bivariate probability matrix as input, the DMPE simulation can proceed as the traditional sequential simulation. The covariance model is replaced with the bivariate probability matrices. Three conditional simulations are shown in Figure 4. Notice the spatial structure from North to South in both of these two simulations is category 1 has different transition probability to category 2 and 3. This asymmetric spatial structure is capture by bivariate probability from the training image and reproduce in the simulation. The used bivariate probability is inferred from the training image.

5 Conclusion

The multivariate probability distribution is estimated directly from the direct and cross bivariate probability. Theoretically, the multivariate probability estimation is in accordance with the Maximum Entropy principle. The conditional probability distribution is calculated from the estimated multivariate probability. There is no simplicity assumption. Theoretically, this conditional probability would reflect the most information from the conditional probability constraints.

There are some constraints in implementation. The first constraint is the conditioning data number because of computing the multivariate probability distribution. One of the results is there are some artifacts in estimation results. As it is can be seen from the estimation results, there are some strips along certain directions. It can be expected that artifacts should be reduced as more conditioning data number can be used in the estimation or simulation.

A second constraint is the univariate proportion reproduction. The simulation results will dependent on the univariate proportion embedded in the bivariate probability diagram. The univariate proportion from the conditioning data set may not be well reproduced. One of the solution to this problem is using some post processing. This point would be the next step research point.

References

[1] H. Ku and S. Kullback. Approximating discrete probability distributions. *Information Theory, IEEE Transactions on*, 15(4):444–447, Jul 1969.

[2] E. T. Jaynes. On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, 70(9):939, 1982.

[3] P. Bogaert. Spatial prediction of categorical variables: the bayesian maximum entropy approach. *Stochastic Environmental Research and Risk Assessment*, 16:425 – 448, 2002.

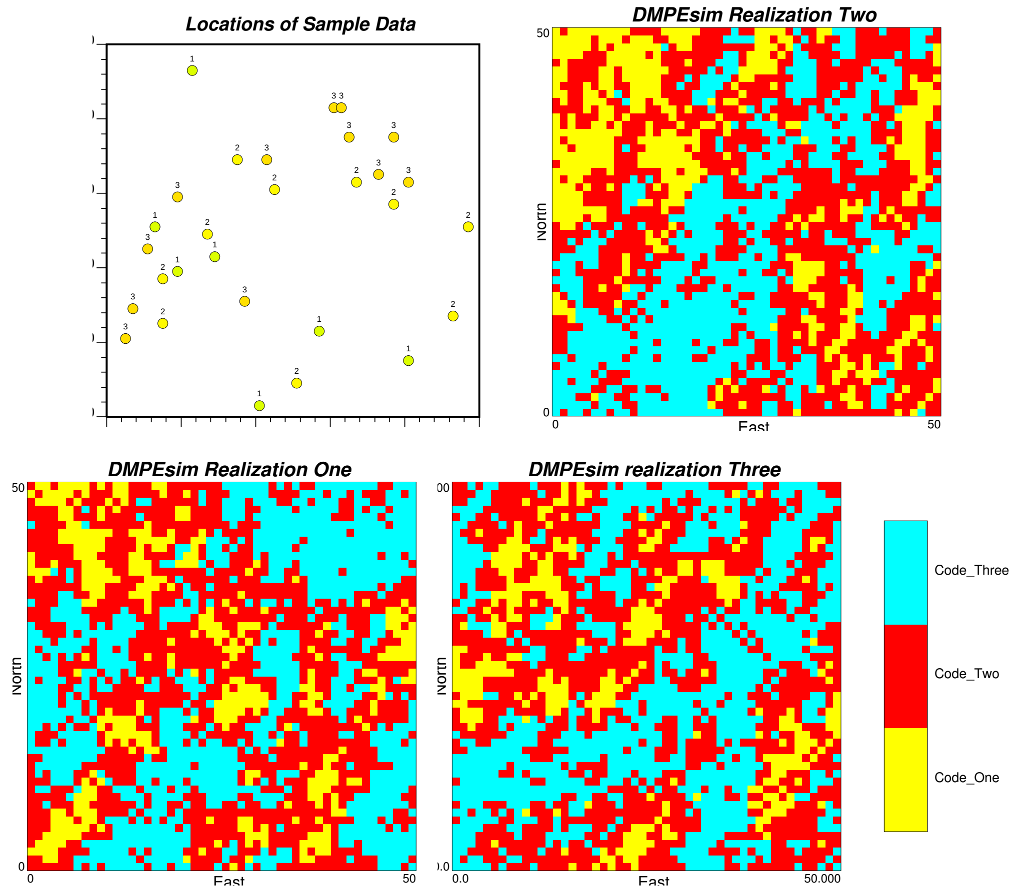


Figure 4: The conditioning data set and three conditional simulation examples based on the DMPE

[4] Dimitri D'Or. *Spatial prediction of soil properties : the Bayesian Maximum Entropy approach.* PhD thesis, Universite catholique de Louvarin, 2003.

[5] W. C. Krumbein and Michael F. Dacey. Markov chains and embedded markov chains in geology. *Mathematical Geology*, 1:79–96, 1969.

[6] W. Krumbein. Markovian models in stratigraphic analysis. *Mathematical Geology*, 2:172–191, 1970.

[7] K. R. Johnson and A. C. Cook. Cyclic characteristics of sediments in the moon island beach subgroup, newcastle coal measures, new south wales. *Mathematical Geology*, 5(1):91–110, 1973.

[8] A.D. Miall. Markov chain analysis applied to an ancient alluvial plain succession. *Sedimentology*, 20:347–364, 1973.

[9] Cunshan Lin and John W. Harbaugh. *Graphic display of two and three dimensional markov computer models in geology.* Van Nostrand Reinhold Company, 1984.

[10] John W. Harbaugh and Graeme Bonham-C Carter. *Computer simulation in geology.* Wiley-Interscience, 1970.

[11] Weidong Li. Markov chain random fields for estimation of categorical variables. *Mathematical Geology*, 39:321–335, 2007. 10.1007/s11004-007-9081-0.

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