Spatial heterogeneity Characterization with Bivariate Probability Matrix Diagram

Yupeng Li and Clayton.V. Deutsch

Characterizing the spatial heterogeneity of facies or rock type is a first step before spatial estimation or simulation for unsampled locations. In this paper, the bivariate probability diagram is proposed as a spatial characterization tool in spatial estimation and simulation. Its construction usually needs many data which is not always possible in reservoir modeling, one approach is to transform the vertical direction to any spatial distance. The concept, the construction approach and how to use them in the spatial facies modeling is introduced in this paper.

1 Introduction

Spatial patterns of geological bodies can exhibit a range of different characteristics, varying in a qualitative manner from random to highly organized. A fundamental concept in geology is that different facies or rock type always show some specific contact patterns. The way the spatial correlation changes with distance between two points characterizes the degree of spatial continuity of the variable studied.

There are many different tools to quantify the spatial dependence including two point statistics and multiple point statistics. Multiple point statistics can reproduce more complex geological reality but inference needs extra effort and more abound data[1]. Some other multiple point statistics such as connectivity functions[2, 3], and the distribution of runs[4] can also be obtained from a training image or some high density sampled data. They can serve as descriptive statistics but none of them has successful implementation in conditional simulation.

Traditionally, two point statistics tools such as variogram are used extensively because they are easy to implement with relatively few data. They cannot, however, reproduce higher order statistics[5]. Another two point spatial heterogeneity characterization tool is transition probability matrix which is mostly used in transition probability based geostatistics [6, 7] . Both of them have been used in one, two and three dimensional spatial heterogeneity characterization [8, 9, 10, 11].

In this paper, the proposed spatial characterization tool is the bivariate probability matrix diagram which is an extension of the traditional Markov transition probability matrix in the approach proposed by Carle and Fogg[6]. The construction, the characteristics of a bivariate probability matrix diagram are explained in this paper.

As its construction usually needs densely sampled data which is not always possible in reservoir modeling, one approach that can transform vertical direction to any spatial distance is proposed here. The bivariate probability calculated from this method will be used as the input to the new proposed multivariate probability estimation algorithm.

2 Heterogeneity characterization tools

A first step in most geostatistical analysis consists of computing functions such as variogram to describe the spatial variation in a region of interest. These spatial heterogeneity characterization functions are usually obtained from sampled data, and are usually referred to as experimental or sample variogram spatial functions. After fitted with some continuous mathematical functions, they are used to infer the underlie random function models[5].

2.1 Indicator variogram

For categorical random variable, indicator covariance/variogram is usually used through indicator kriging in traditional geostatistics. In indicator kriging approach, the spatial relationship between binary variable is characterized by the indicator variogram function $\gamma_{e_k}(\mathbf{u}, \mathbf{u}+\mathbf{h})$ which is:

$$\begin{aligned} 2\gamma_k(\mathbf{u}, \mathbf{u}+\mathbf{h}) &= Var[I_k(\mathbf{u}) - I_k(\mathbf{u}+\mathbf{h})] \\ &= E\{[I_k(\mathbf{u}) - I(\mathbf{u}+\mathbf{h})]^2\} - \{E[I_k(\mathbf{u}) - I_k(\mathbf{u}+\mathbf{h})]^2\} \end{aligned} \tag{1}$$

Note, if the indicator random variable is stationary,the variogram in above will be written as:

$$2\gamma_k(\mathbf{h}) = E\{[I_k(\mathbf{u}) - I_k(\mathbf{u}+\mathbf{h})]^2\} \tag{2}$$

Define the bivariate probability $p(\mathbf{u}, \mathbf{u}+\mathbf{h})$ between any two location is:

$$p(\mathbf{u}, \mathbf{u}+\mathbf{h}) = Pr(z(\mathbf{u}) = k, z(\mathbf{u}+\mathbf{h}) = k'), \quad k, k' = 1, \cdots, K \tag{3}$$

Where $(z(\mathbf{u}) = k, z(\mathbf{u}+\mathbf{h}) = k')$ is a bivariate event that is the joint outcomes at these two locations. The bivariate probability will characterize the probability for this event to exist. Later on, the bivariate event will be simplified as $(\mathbf{u} = k, \mathbf{u}+\mathbf{h} = k')$.

There is a good relationship between the indicator variogram and bivariate probability. As shown in Equation 2, the indicator variogram will only count those transition probabilities that satisfy $I_k(\mathbf{u}) - I_k(\mathbf{u}+\mathbf{h}) \neq 0$. For example, define two locations with departure distance $\mathbf{h}$ are $\mathbf{u}_1$ and $\mathbf{u}_1 \pm \mathbf{h}$, the bivariate probability between those two locations will be: $p(\mathbf{u}, \mathbf{u}+\pm\mathbf{h})$. Then, the indicator variogram in Equation (2) can also be calculated from the bivariate probability as:

$$2\gamma_k(\mathbf{u}, \mathbf{u}+\mathbf{h}) = \sum_{\substack{k'=1 \\ k'\neq k}}^{K} p(\mathbf{u} = k, \mathbf{u}+\mathbf{h} = k') + \sum_{\substack{k'=1 \\ k'\neq k}}^{K} p(\mathbf{u}_1 = k, \mathbf{u}_1 - \mathbf{h} = k') \tag{4}$$

As shown from Equation(4), 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. Only those bivariate probabilities that character the cross relationship will contribute to the indicator direct variogram. For each indicator variogram, it will count and average $2 \times (K-1)$ bivariate probability terms [12]. Actually, some information lost here,comparing with the bivariate probability. The cross indicator variogram could be integrate this information and used in indicator system. However, it has been shown that this information does not bring too much improvement in indicator kriging system[13, 14].

2.2 Markov transition probability matrix

In many situations, a sequence of events in either time or space is observed as a succession of states. And these events are taken from a limited set of alternatives. If the natural processes exhibit an effect in which previous events influence, but do not rigidly control subsequent events, these processes are named "Markov chain" after the work of the Russian mathematician Markov. Markov properties have been recognized in many geological phenomena, including stratigraphic sequence of lithologic units, sedimentary processes, succession of mineral occurrencesin igneous rocks, etc[15, 16].

When a Markov chain has a very short "memory" which only extends for a single step at a time and ceasing beyond a single step. Such a chain is termed as a "first-order" Markov chain. The relationship between adjacent events will be summarized by the transition probability matrix in which each entry is the probability of the transition from a particular event (pertaining to the particular row in the matrix) to the next state (pertaining to the particular column)which is denoted as $t_{kk'}(\mathbf{h}) = p(\mathbf{u}+\mathbf{h} = k|\mathbf{u} = k')$. All the transition probability will compose a transition probability matrix $T(\mathbf{h}) = \{t_{kk'}\}; k, k' = 1, \cdots, K$. A Markov transition matrix $T(\mathbf{h} = 1)$ with three events may be written as:

$$T(\mathbf{h} = 1) = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \tag{5}$$

The stochastic process is defined as single dependence chain because only a single preceding state is involved to decide the current state. When a single dependence chain is shown to have a one-step Markov property, then one transition probability matrix is a significant descriptor of the transition properties and enough to character it.

Generally, any multiple steps of single dependence chain can also be defined as $T(\mathbf{nh}), n = 1, 2 \cdots, L$ which means that the current state will be dependent on the state that is located at **nh** distance. In this case, the process will have a long distance memory and named as multiple step single dependence chain. Theoretically, if the stochastic process is a stationary one-step single dependence process, this multiple step transition probability matrix $T(\mathbf{nh})$ can be calculated from the multiply of one-step transition probability matrix. That is the transition probability at any lag distance ($\mathbf{nh}$) can be calculated as a matrix exponential:

$$T(\mathbf{nh}) = T(\mathbf{h} = 1)^n \tag{6}$$

where $T(\mathbf{h} = 1)$is the one step transition probability matrix.

From its definition, each entry in the transition probability matrix will be a conditional probability $t_{kk'}(\mathbf{h}) = p(\mathbf{u}+\mathbf{h} = k|\mathbf{u} = k')$ and it relates to bivariate probability as:

$$t_{kk'}(\mathbf{h}) = p(\mathbf{u}+\mathbf{h} = k|\mathbf{u} = k') = \frac{p(\mathbf{u}+\mathbf{h} = k, \mathbf{u} = k')}{p(\mathbf{u} = k')} \tag{7}$$

The transition probability matrix provides a framework for exploring the underlying physical, chemical, and biological control on sedimentary processes and deposits with superimposed random fluctuations introduced by the "built-in" probabilistic mechanism. Markov chain analysis once gained the most popular and favored tools among sedimentologists as a useful statistical technique to explain and understand geological cyclicity. It has been employed to determine whether the occurrence of a facies at one point in a stratigraphic succession is completely dependent on the facies at the immediately underlying point [17, 18, 19, 20].

It is more easy to interpret the geological meaning of the Markov transition probability than a variogram. This makes Markov transition probability matrix also popular in geostatistics research. Many geostatistical researchers proposed to use Markov transition probability as an alternative approach to describe the spatial structure such as the transition probability-based indicator kriging [6]. In this methodology, the indicator (co)kriging equations are reformulated based on the 3D continuous-lag Markov TPM which allow to integrate some geological information such as sedimentary juxtaposition.

3 The Bivariate Probability Model

As shown from indicator variogram definition in Equation (4), the indicator variogram can be calculated from bivariate probability with some assumptions. Also, each entry of transition probability matrix can be expressed as a bivariate probability as shown in Equation (7). Both of them are related to the bivariate probability. Thus, it would be better to use the bivariate probability directly if possible. The bivariate probability will be used as the spatial heterogeneity characterization tools in this research. Some researchers have begun to use it as the spatial statistic tool such as the Bayesian Maximum Entropy approach [21, 22].

3.1 Definition and construction

Let's define the bivariate marginal probability for two location locations separated by lag distance $\mathbf{h}$ as $p_{kk'}(\mathbf{h}) = p(\mathbf{u} = k, \mathbf{u} + \mathbf{h} = k')$. It will be a bivariate probability matrix with $K \times K$ entries. For example, for a stochastic with three categories, the bivariate probability matrix could be written as:

$$P_{kk'}(\mathbf{h}) = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \tag{8}$$

Typically, data are abundant for characterizing vertical spatial variability of facies. Thus, the bivariate probability matrix is usually constructed from the vertical profile where has a high sampled data density. The first step is to construct a transition tally matrix by observing the facies outcome at successive equal intervals. For example, a vertical profile consisting of four categories as shown in Figure 1 can be summarized in a tally matrix as;

	A	B	C
A	10	12	16
B	15	14	33
C	13	36	23

Where for example, the number of times C succeeds A is 16 which means that finding the number of C and A at these two locations at the same is 16. One property of the tally matrix is that i^{th} row total is equal to the i^{th} column total. The row sum (or column) may be written as the vector

$$[38 \quad 62 \quad 72\,]$$

The second step is the division of the tally matrix by the total sum of the tally matrix. The results would be a bivariate probability matrix:

	A	B	C
A	0.2632	0.3158	0.4211
B	0.2419	0.2258	0.5323
c	0.1806	0.5	0.3194

Division of the tally matrix by each of the row total leads to the traditional transition probability matrix. While dividing the row or column vector by the total sum of the tally matrix will result in the univariate probability vector. In this small example it is:

$$[0.2209 \quad 0.3605 \quad 0.4186\,]$$

The bivariate probability matrix at each step also satisfy:

$$\sum_{k'=1}^{K} p_{kk'} = p(k) \quad k = 1, \cdots, K \tag{9}$$

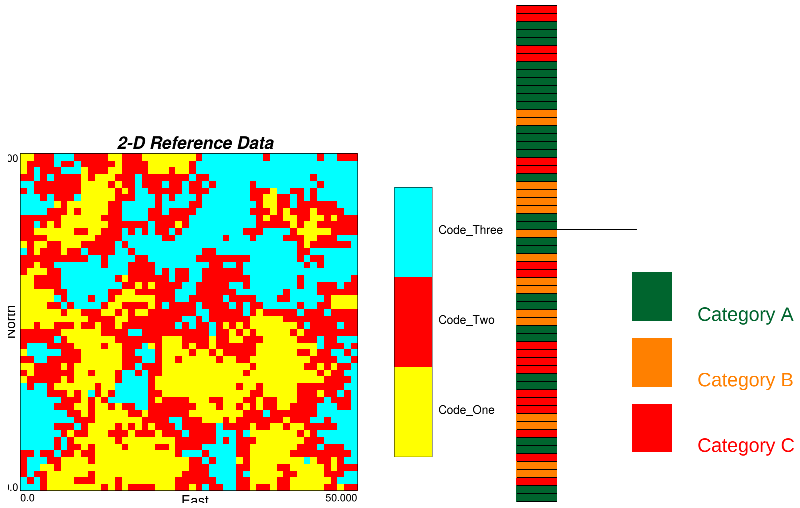

Figure 1: The training image and vertical profile from the outcrop for bivariate probability diagram inference

Which means the row sum or column sum of a bivariate probability would be the univariate probability.

From above two steps, one bivariate probability matrix is obtained given the chosen lag separation. As the lag distance increase, different bivariate probability matrices for each step would be obtained. For example, from the training image along the north direction the bivariate marginal diagram will be plotted as shown in Figure 2. Thus, the experimental bivariate probability model is directly used in this research.

In traditional indicator geostatistical approach, after doing the experimental indicator covariance or indicator variogram calculation, some mathematical models are used through the curve-fitting method. If the same technique is done for bivariate marginal probability model, it will be a real tedious job. Also, in reality, the stationary assumption is very difficult to satisfy. The approach of matrix multiply used in Markov transition probability based model is not appreciate for bivariate marginal probability model. The multiple-steps bivariate probability matrix calculated from the sequence will be used in the multivariate probability estimation directly.

3.2 Characteristic of bivariate probability model

With the lag distance $\mathbf{h}$ increasing from zero to a further distance, bivariate probability $p_{kk'}(\mathbf{h})$ form a diagram. The lag $\mathbf{h}$ may be an exact distance measure, or the number of spatial steps(pixels or grid cells). Under the similar second-order stationary definition, the bivariate probability $p_{kk'}(\mathbf{h})$ is only dependent on the lag $\mathbf{h}$ and not on any specific location $\mathbf{u}$, so that bivariate probability $p_{kk'}(\mathbf{h})$ diagram could be estimated from data pairs in a space as the traditional variogram calculation. Some basic properties are:

1. As shown in Figure 2, there are direct bivariate probability $p(\mathbf{u} = k, \mathbf{u} + \mathbf{h} = k)$ and cross bivariate probability $p(\mathbf{u} = k, \mathbf{u} + \mathbf{h} = k')$. Direct bivariate probability represent auto-correlations of individual categories, and cross-bivariate probability represents cross-relationship between different categories.

2. Although it is named a bivariate probability, it is different with the traditional bivariate probability in pure statistics. In pure statistics, a bivariate probability for two random variable $p(x, y)$ would be a symmetric statistics that is $p(x, y) = p(y, x)$. While here, the bivariate probability has the asymmetric property that is $p_{kk'}(\mathbf{h}) \neq p_{kk'}(-\mathbf{h})$. Thus, the head and tail categories would be defined. For bivariate probability $p_{kk'}(\mathbf{h})$, the number k would be called head category, k' would be called tail category.

3. They are non-negative as they are probabilities; at any specific lag, values of bivariate probability headed by the same tail category will sum to the univariate probability of head category;

4. The initial value of a direct bivariate probability should be the univariate probability p_k. Where the initial value of cross bivariate should be zero. As the lag distance increase to ∞, the bivariate probability should satisfy:

$$p_{kk'}(\mathbf{h}) = p_k \cdot p_{k'} \qquad \mathbf{h} \to +\infty; \quad k, k' = 1, \cdots, K \tag{10}$$

   Which could be defined as the sill of a bivariate probability diagram. The univariate proportion p_k, p_k' are used to calculated the sill, for an area sufficiently large, should be equal to the global proportion. Similar as the variogram, the distance of bivariate probability reaches its sill could also be called as ranges which means after this distance there is no clear relationship between the observed two locations.

One more characteristic of bivariate probabilities is that the mean patch size of each category or the juxtaposition information can be easily integrated with cross transition probabilities. A cross bivariate probability represents the existence between two categories at different locations. An idealized cross bivariate probability $p_{kk'}(\mathbf{h})$ starts from the point $(0, 0)$ and gradually increases to a stable value–the sill as shown in Figure 3.

The univariate probability of the head and tail category should be equal to the global proportion when the area is sufficient large. As shown in Figure 3, before the cross bivariate probability stably approaches its sill, depending on the spatial distribution of the two involved categories(whether they are frequent neighbours or not) the cross bivariate probability shape may be different. If class k frequently occurs adjacent to class k', the bivariate probability will have a peak first and then approach to its sill. If they are seldomly occurs contact to each other, the bivariate probability would have a lower value at the beginning and then approach to its sill later as shown in Figure 3.

4 Three Dimensional Bivariate Probability Inference

Usually, bivariate probability matrices are inferred from the vertical direction or along the specific direction where the data sampling is dense enough for a reliable model inference. Along the vertical direction, data density is sufficient to infer reliable bivariate statistics. But, building the horizontal heterogeneity statistics model is a challenge, especially in clastic reservoir where the facies change rapidly in the horizontal plane. The available well data are rarely sufficient to construct a reliable horizontal variogram especially in the appraisal stage. In this section, an approach to relate the horizontal transition probabilities to vertical transition probabilities is developed based on the sequence stratigraphy.

4.1 Stacking concept model and anisotropy ratios

Before the quantitive tools such as variogram were introduced in geological analysis, Walther's concept of facies stacking was widely used in facies analysis. This idea can be usefully extended to consider not just a vertical sequence, but a whole body of rock. A genetic increment of strata is a mass of sedimentary rock in which the facies or subfacies are genetically related to one another. A typical genetic increment of strata would consist of a single prograding delta sequence containing delta platform, delta front and pro-delta deposits. For example in one fluvial dominated delta building sedimentary process, the sedimentary facies will have an up-coarsing pattern along the vertical direction. As along the lateral chronostratigraphical boundary, from the proximal to distal direction, the same pattern can be found which is shown in Figure **??**

More importantly, Walther's Law is the basis for the theory of Sequence stratigraphy which is a recent revolutionary paradigm in the field of sedimentary geology. Sequence stratigraphy is the study of genetically related facies within a framework of chronostratigraphically significant surfaces. The sequence stratigraphic approach has led to improve understanding of how stratigraphic units, facies tracts, and depositional elements relate to each other in time and space within sedimentary basins[23]. After doing facies and sequence stratigrapy analysis, the facies or facies tracts stacking pattern will always be collected.

Another important information can be obtained from doing some geological analysis based on the sequence stratigraphic analysis is the anisotropy ratio which is a measurement of how much of the vertical sequence stacking pattern will extend to the horizontal direction.

It is not very easy to get a very specific interval because of different scale, different sedimentary models and different data resource et al. But before modeling, this information is clear in the geologist's mind and it should be integrated into the modeling effort such as some efforts has already done on this point[24].

4.2 From vertical to 3D spatial

After building the conceptual sedimentary model in the research area, the next step is define the main anisotropy ratios by doing high-resolution chronostratigraphic correlation. As shown in Figure 5, assuming the vertical facies bivariate probability diagram are already obtained from location a, along the direction from a to f (named as dip direction), it will be more distal, from a to d it will be more proximal, the anisotropy ratio denoted as a_{dip}. Along the direction from a to d (named as strike direction) will be just a sedimentary source shifting, denoted as a_{strike}. Also, there could be some fluturation of vertical anisotropy ratio which is denoted as a_{vert} Location c is an arbitrary spatial location departing from location a. In practical reservoir modeling, these three anisotropy ratios along different directions can always be obtained from the sequence stratigraphic analysis and geological understanding.

The second step is rescaling any 3D spatial distance to an equivalent vertical direction. Any 3D spatial distance vector will be decomposed into three vectors along three main facies transition directions that is vertical, dip and strike direction. After that, based on the anisotropy ratio, an effective distance is obtained by combining them as:

$$\mathbf{h} = \sqrt{(\frac{h_{dip}}{a_{dip}})^2 + (\frac{h_{strike}}{a_{strike}})^2 + (\frac{h_{vert}}{a_{vert}})^2} \tag{11}$$

In Equation (11), it is assumed that the anisotropy ratio along three directions will have a geometric shape. Actually, there could be some other shapes. By this approach, any 3D spatial distance will be related to an effective distance along the vertical direction and obtain the transition probability from there.

5 Conclusion

As both the indicator variogram and the Markov transition probability can calculated from the bivariate probability, it can be used in characterization of heterogeneity of facies or rock types. Using the bivariate probability matrix diagram, the facies pattern can be captured and could be reproduced in the later modeling realization. The Walther's law in sedimentatology provides a way to transform three dimensional distance to vertical direction. In practice, the sequence stratigraphic knowledge could be integrated by this transformation.

References

[1] S. Strebelle. Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, 34:1–21, 2002.

[2] Katerina Michaelides and Adrian Chappell. Connectivity as a concept for characterising hydrological behaviour. *Hydrological Processes*, 23(3):517–522, 2009.

[3] C. F. Harvey and S. Gorelick. Mapping hydraulic conductivity: Sequential conditioning with measurements of solute arrival time, hydraulic head, and local conductivity. *Water Resources Research*, 31(7):1615–1626, 1995.

[4] J.M. Ortiz. *Characteration of high order correlation for enhanced indicator simulation.* PhD thesis, University of Alberta, 2003.

[5] C. V. Deutsch. *Geostatistical Reservoir Modeling.* Oxford University Press, 2002.

[6] Steven Carle and Graham Fogg. Transition probability-based indicator geostatistics. *Mathematical Geology*, 28:453–476, 1996. 10.1007/BF02083656.

[7] Weidong Li, Chuanrong Zhang, James E. Burt, and A-Xing Zhu. A markov chain-based probability vector approach for modeling spatial uncertainties of soil classes. *Soil Sci Soc Am J*, 69(6):1931–1941, 2005.

[8] Weidong Li and Chuanrong Zhang. A single-chain-based multidimensional markov chain model for subsurface characterization. *Environmental and Ecological Statistics*, 15(2):157–174, 2008.

[9] Eungyu Park, Amro M. M. Elfeki, Yungoo Song, and Kangjoo Kim. Generalized Coupled Markov Chain Model for Characterizing Categorical Variables in Soil Mapping. *Soil Sci Soc Am J*, 71(3):909–917, 2007.

[10] G. S. Weissmann and G. E. Fogg. Multi-scale alluvial fan heterogeneity modeled with transition probability geostatistics in a sequence stratigraphic framework. *Journal of Hydrology*, 226(1-2):48 – 65, 1999.

[11] Anne Louise Larsen, Marit Ulvmoen, Henning Omre, and Arild Buland. Bayesian lithology/fluid prediction and simulation on the basis of a markov-chain prior model. *Geophysics*, 71(5):R69–R78, 2006.

[12] Oy Leuangthong, Daniel Khan, and Clayton V. Deutsch. *Solved Problems in Geostatistics.* John Wiley and Sons, 2008.

[13] P. Goovaerts. Comparison of coik, ik and mik performances for modeling conditional probabilities of categorical variables. In R. Dimitrakopoulos, editor, *Geostatistics for the next century*, pages 18–29. Kluwer Academic, 1994.

[14] P. Goovaerts. Comparative performance of indicator algorithms for modeling conditional probability distribution functions. *Math Geology*, 26(3):385–410, 1994.

[15] A. D. Miall. *The Geology of Fluvial Deposits.* Springer-Verlag, New York, 1996.

[16] 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.

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

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

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

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

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

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

[23] O. Catuneanu. *Principles of Sequence Stratigraphaphy.* Elsvier, 2006.

[24] A. C. MacDonald and J. O. Aasen. *A Prototype Procedure for Stochastic Modeling of Facies Tract Distribution in Shoreface Reservoirs*, chapter 9, pages 91–108. AAPG, 1995.

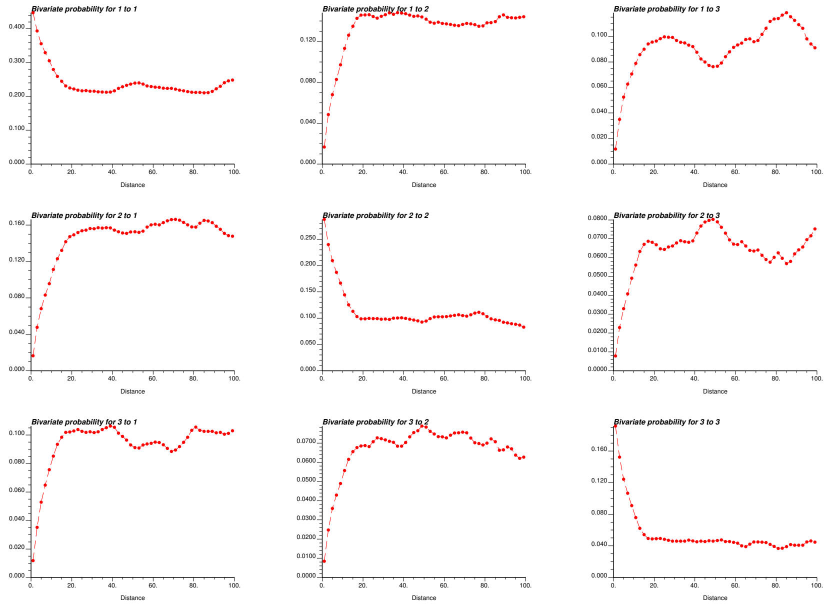

Figure 2: One example of bivariate marginal probability diagram from the training image

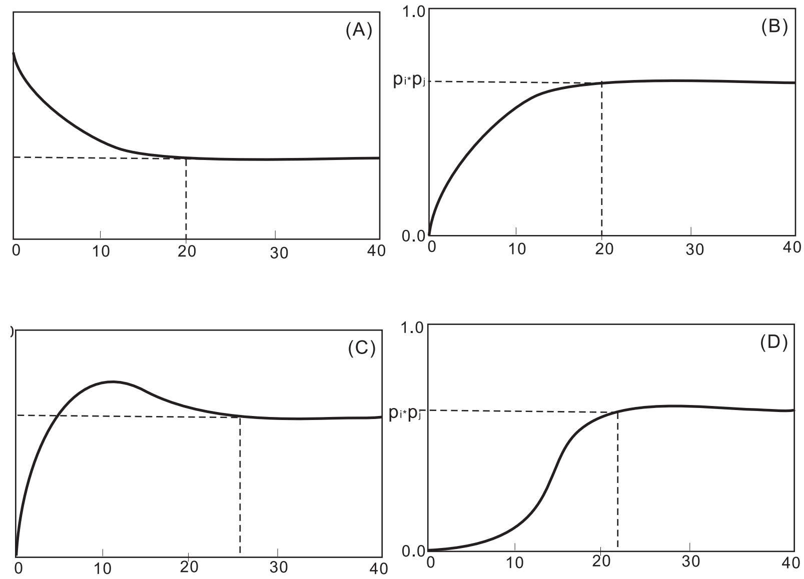

Figure 3: Illustration of typical features of idealized bivariate probability diagram(Modified from W. Li[]). (A): Typical direct bivariate probability diagram;(B):Typical cross bivariate probability diagram;(C):Two categories that are frequent neighbours(D): Two categories that are infrequent neighbours.

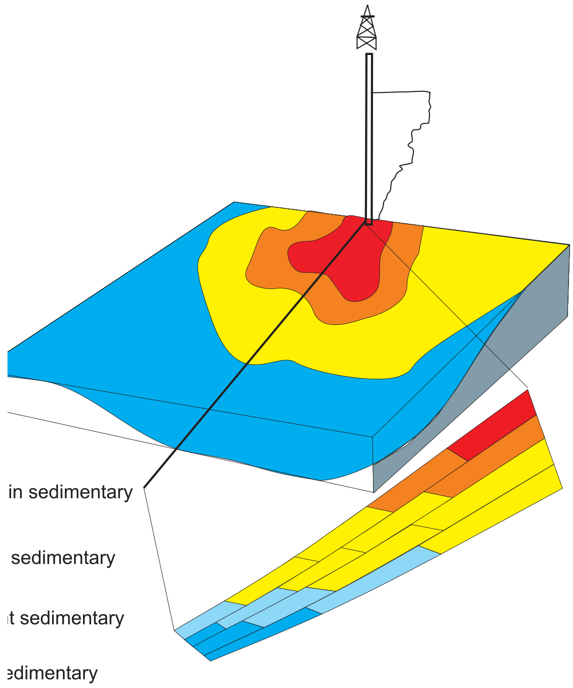

Figure 4: One example of facies stacking pattern in fluvial dominated delta sedimentary system

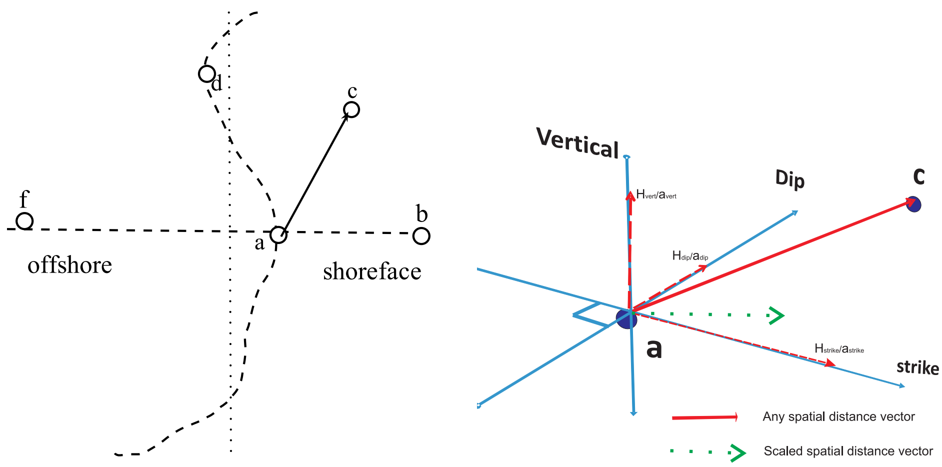

Figure 5: One sedimentary example of shore line and its distance transformation