University of Alberta

Improved Facies Modelling with Multivariate Spatial Statistics

by

Yupeng Li

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Mining Engineering

Department of Civil and Environmental Engineering

Fall 2011
Edmonton, Alberta

Abstract

This dissertation proposes an improved facies modelling methodology that involves a new geological spatial characterization tool, a geological based spatial distance calculation, and a theoretically sound conditional probability calculation. The full set of bivariate probabilities are proposed as a spatial characterization tool that integrates facies stacking information into the final facies model construction. After inference in the vertical direction from well data, they can be transformed to any spatial distance vector based on a heterogeneity prototype and the calculation approach proposed in this research. The data information carried by the bivariate probabilities will be integrated together into a multivariate probability based on the minimum Kullback–Leibler distance. From this estimated multivariate probability, the conditional probability for each unsampled location is calculated directly. The research developed in this dissertation adds a new geostatistical facies modelling approach to currently available tools. It provides a new approach to integrate more geological understanding in the final model. It could be used in practice and as a seed for further development.

Acknowledgements

I thank Professor Clayton V. Deutsch for his great mentorship in this four years exploration. He gives me confidence and encourages me to continue. Every meeting would always bring excitements. His enthusiasm has influenced me in these four years and in the future. The meeting sheets he used to illustrate his ideas to me will be a treasure and good memories.

Let me express great appreciation to my parents. They always give me full supports in all aspects of my life. It is because of their love that I can go so far. Also deep thanks to my wife, Lihui. She always gives me support and spurs on me when I need it. Albert, it is the time when you were expected that this thesis was being written. Many times, it is because of you, that I can write some sections. I love you.

I would also like to thank the members of my defense committee, for taking their previous time to review my thesis and giving valuable suggestions for my research.

Also, in these four years great friendships have brought more colour to my life. CCG colleagues including Tong, Behrang, Brandon, John, Martha, Mehran, Miguel and Talal and many others were always good candidates to discuss some non-geostatistic topics. Thanks also to my best friends which include Linglin, Xianmin, Hongwen, Qiang, Lucheng, Jianzhao, Andi, Nick, and many others that we shared many wonderful times together. Thank you all.

Finally, I gratefully acknowledge the financial support from China Scholarship Council and Centre for Computational Geostatistics.

Contents

List of Figures

List of Tables

List of Symbols

$\mathbf{u}$	Spatial location defined by its Cartesian coordinates (x,y,z)
$\mathbf{h}$	Spatial lag distance between two locations
$Z(\mathbf{u})$	Spatial discrete random variable
$z(\mathbf{u})$	Realization from the random variable
e_k	Facies type from one location
K	Total number of facies
$P(\mathbf{u})$	Probability distribution for current spatial location
$P(\mathbf{u}_1,\ldots,\mathbf{u}_n)$	Multivariate probability distribution for n locations
ω	A multivariate data event
Ω	Multivariate data event space
ℓ	Index of multivariate data event
N	Multivariate data event space dimension
$I(\mathbf{u};e_k)$	Indicator random variable for facies type e_k
$i(\mathbf{u};e_k)$	Realization from the indicator random variable
p_k	Global proportion for facies type e_k
$Cov(\cdot)$	Covariance function
$\gamma(\cdot)$	Variogram function
$T(\mathbf{h};K,K)$	Markov transition matrix
$P(\mathbf{h};K,K)$	Bivariate probability matrix
$H(\cdot)$	Entropy
$J(\cdot)$	Kullback-Leibler divergence
b	Marginal probabilities

Chapter 1

Introduction

In this chapter, some basic concepts are introduced. Also, the motivation of this thesis is highlighted after stating two major challenges in numerical reservoir modelling.

1.1 Reservoir Simulation

Reservoir flow simulation refers to the dynamic simulation of fluids through a reservoir model over time (Fanchi, 2001; Ezekwe, 2010). It is used extensively by oil and gas companies in the development of new fields as well as in developed fields where production forecasts are needed to help make new investment decisions.

The numerical reservoir model for flow simulation usually refers to the 3D models of gross thickness, net to gross, porosity, and permeability and, where appropriate, descriptions of faults, fractures and aquifers. The field under study will be described by a grid system defined by cells or grid-blocks. Each cell must be assigned reservoir properties to describe the reservoir as sketched in Figure 1.1.

A flow simulator applied to the numerical model will help understanding well production performance and different recovery mechanisms. If possible, the simulator is calibrated using historic pressure and production data. When the simulation results closely match the historical reservoir performance, a higher degree of confidence is placed in its ability to predict the future fluid behaviour under a series of potential scenarios such as drilling new wells and injecting various fluids.

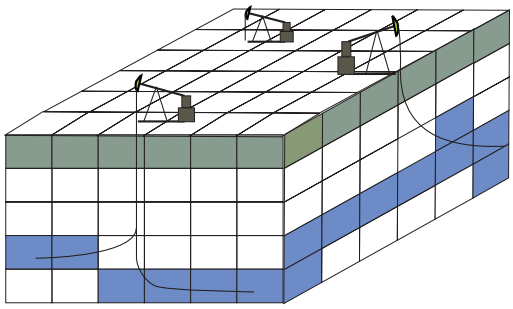


Figure 1.1: A schematic illustration of a numerical model of a reservoir

Reservoir formations show some internal variation in rock properties such as reservoir thickness, porosity and permeability. In flow simulation, the fluid flow is mostly controlled by the connectivity of extreme permeability. For example, the high permeability facies will create preferential flow paths and low permeability facies will create barriers. Building a 3D numerical model that is close to the true geological reality to reflect those reservoir property variation is the main goal of reservoir modelling.

1.2 Reservoir Modelling

Reservoir modelling represents an important step in the work flow of a typical reservoir study (Weber and Geuns, 1990). A reservoir static model from the reservoir modelling will provide a sound geological description of the reservoir, both in geometry and petrophysical properties and it is one of the main input to reservoir flow simulation to correctly evaluate the reservoir production (Haldorsen and Damsleth, 1990).

The reservoir static model includes the structural model, stratigraphic model, facies and petrophysical model from several integrated modelling steps as shown in Figure 1.2. Those models will be built using different data and different techniques, but they all should be integrated and consistent to each other and to the data and information available.

The structural model defines the structural top map and the fault pattern in the

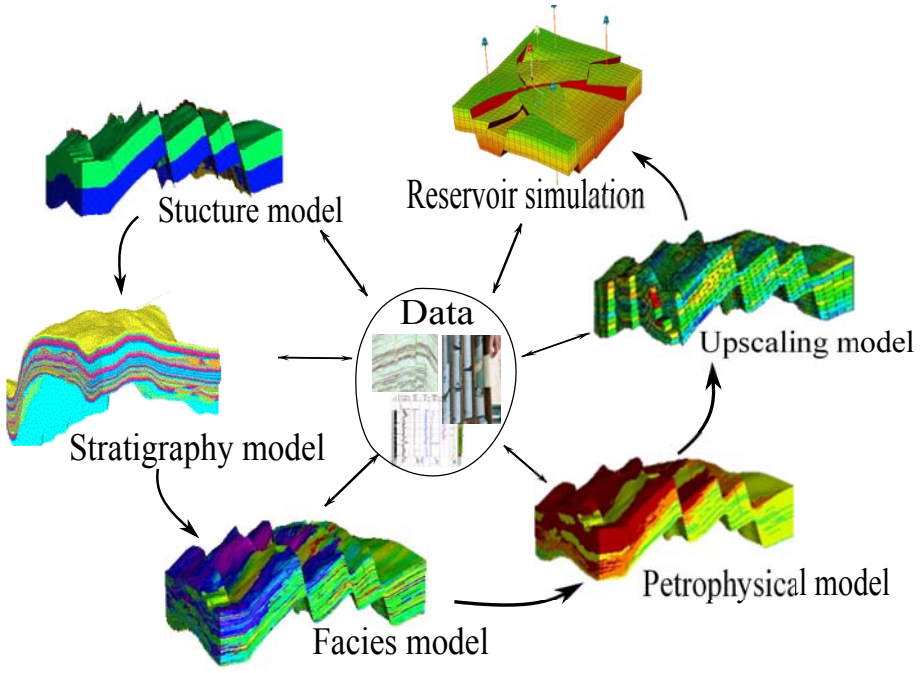


Figure 1.2: Reservoir modelling workflow (modified from ROXAR, 2009)

reservoir. Traditionally, it is the research domain of geophysics and based mostly on seismic survey data.

Stratigraphic modelling mainly refers to the geological analysis such as sequence stratigraphic study, well correlation and division into stratigraphic layers. The stratigraphic model provides the main reservoir flow units. Usually, the data and information for this step will be from seismic data, sedimentology, well logs, minerology, and field/outcrop studies.

The structural and stratigraphic models of a reservoir provide a reference geometric framework for the next stage: facies modelling. By building the facies model, the reservoir structural framework will be filled with a facies distribution to depict the spatial variability of the reservoir rock.

Based on the facies model, the porosity and permeability model characterizes porous network of the reservoir. These models will determine the volume of hydrocarbons, their spatial variability and how easily the fluids will flow towards the producing wells.

The number of cells for a typical geostatistical model is on the order of $10^7 - 10^8$. By contrast, the number of cells in a typical reservoir simulator is on the order of $10^5 - 10^6$. Thus, fine scale geostatistical models are usually upscaled for later reservoir flow simulation as shown in the workflow in Figure 1.2.

1.3 Facies Modelling

The facies are considered to be the foundation of the petrophysical model. The spatial distribution of facies is of major concern due to their influence on fluid flow, and thereby the production of oil and gas (Hatløy, 1995). A detailed facies model will guide the petrophysical distribution modelling since they are highly related together. The steps in facies modelling include: a classification/definition of facies from geological analysis, and a spatial distribution modelling using deterministic or probabilistic/stochastic simulation methods.

1.3.1 Facies identification and classification

A facies is a distinctive rock unit that forms under certain conditions of sedimentation, reflecting a particular process or environment, such as river channels, delta systems, submarine fans, reefs and so on (Reading, 1996). Facies analysis helps geologists reconstruct the paleogeography when the facies were deposited in geological history.

The geologist may define many high resolution facies types from cores, well logs and/or outcrop studies (Glaister and Nelson, 1974; Coudert et al., 1994). Geomodellers will group these facies into a reduced number that would be referred to as lithotypes. The advantage of this is that it provides a simpler description of the reservoir and is more easily handled in stochastic modelling. For example, in a very simple study, the facies could be grouped into pay and non-pay based on a reservoir quality threshold (Deutsch, 2002).

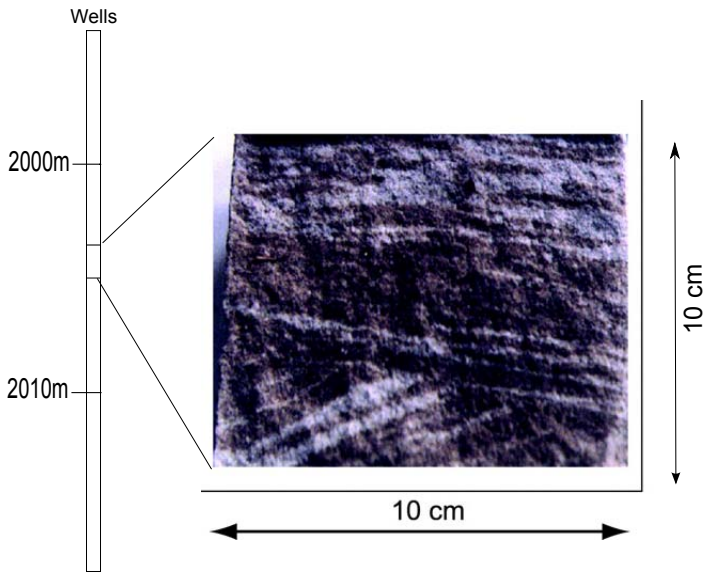


Figure 1.3: A schematic illustration of core data for facies modelling

1.3.2 Data for facies modelling

There are several sources of facies data. Among them, core, well log, seismic and outcrops are the most widely used.

The core is obtained mechanically in the drill hole or along the side of the well at certain depth intervals see Figure 1.3. The porosity, permeability, water saturation, volume of clay, and many other properties can be measured in the lab and are called core data. Core data provide a direct measurement of reservoir properties that may be used to calibrate well logs to improve the well log analysis. As the coring is expensive and slow, there are few data for a modelling project.

Well log data are obtained by lowering logging instruments into the drill hole with wire-lines to measure the rock and fluid properties of the formation from the side walls. Measurements of gamma ray, neutron density, and spontaneous potential of the rock are taken. These logging measurements usually have a vertical resolutions from centimetres to about a meter with a similar lateral penetration around the boreholes. From these logs, a petrophsicist will calculate the permeability and porosity for reservoir formation evaluation. The well log data is a primary data source and usually called hard data for later on reservoir modelling.

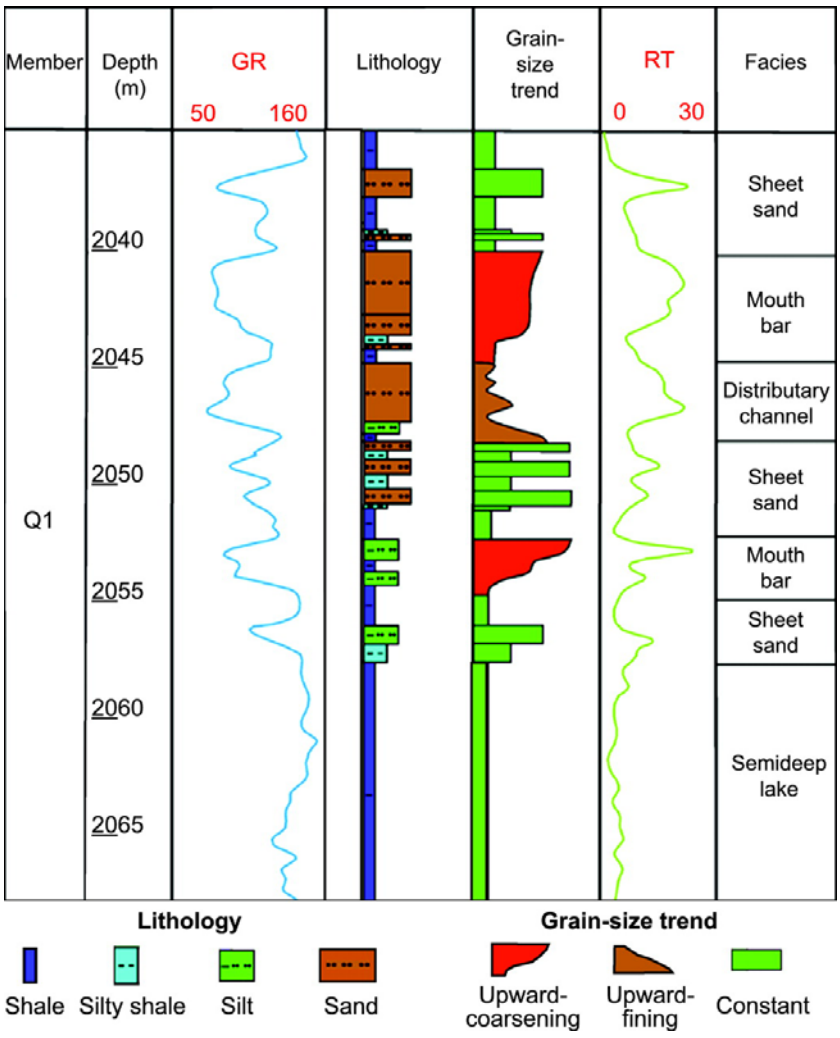


Figure 1.4: One example of well log analysis results

The well log curve shape also helps the basis for geological facies interpretation and other sequence stratigraphy analysis as shown in Figure 1.4. The geologist will recognize different facies through the well log curve stacking pattern study (Ellis and Singer, 2008).

Seismic properties of the subsurface geobodies are measured in terms of acoustic impedance, amplitude and sonic velocity (Veeken, 2007). The scale and resolution of seismic data is usually extensive over the reservoir. The primary limitation of seismic data is its vertical resolution (about 10 meters) and multi-response to the facies variations (Roksandic, 1978). It is usually classified as soft data and is used

as large scale information to aid in the inference of facies after calibrating to the hard data from well logs.

Another data resource is outcrops (Willis and White, 2000; Enge et al., 2007). Usually, the outcrops that will be used in reservoir modelling should have a similar geological background with the reservoir under study. In practice, not every reservoir will have an analogue outcrop for more detailed study.

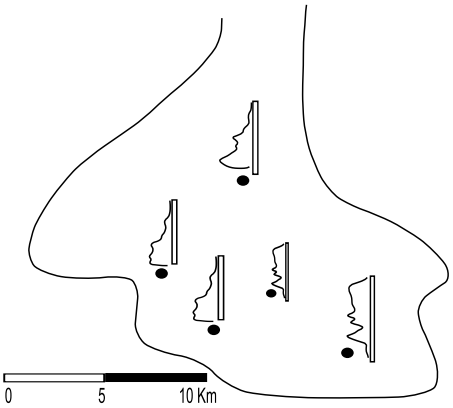
1.3.3 Facies spatial interpolation

After the facies are defined and the relevant data are collected, the numerical models required for flow simulation can be built. As already discussed, the numerical simulation is done on a model where each cell should have a petrophsical value. While in practice, there are relatively few locations that are sampled as shown in Figure 1.5a.

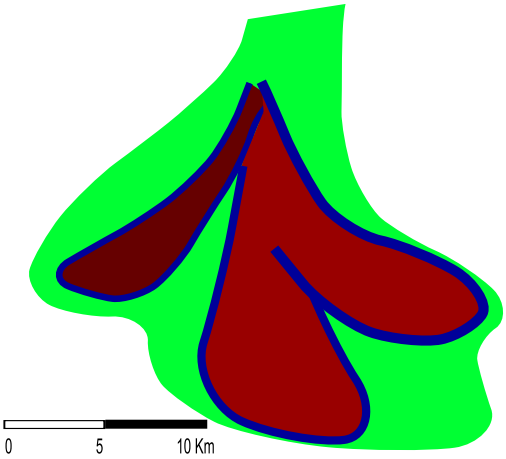
One possible solution comes from geologists. Based on a geological understanding, geologists can draw a facies map from several sampled locations. As shown in Figure 1.5b, the facies distribution in the whole area can be sketched out based on the well data and the geology background of the study area. Then, this facies map can be digitized as shown in Figure 1.5c and used as an input for petrophysical modelling.

Because of heterogeneity and sparse data, our knowledge of the spatial distribution of the physical properties of geologic formations is uncertain. Geostatistics was developed originally to predict probability distributions of ore grades for mining operations (Journel and Huijbregts, 1978). It has become an important tool for estimating reservoir properties at points where data are not available, as well as quantifying the corresponding uncertainty (Deutsch and Journel, 1998).

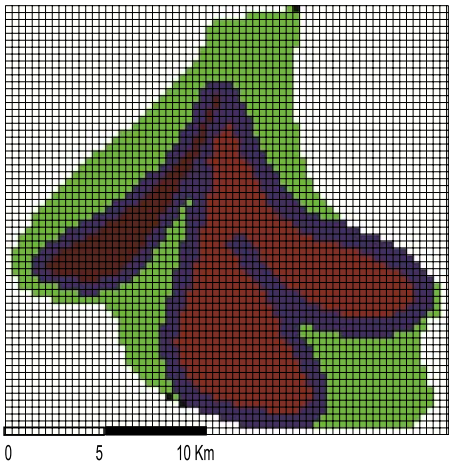
There are many geostatistics algorithms to do spatial interpolation for all the grid cells based on some sampled locations. Very often, they are grouped as cell-based and object-based approaches (Koltermann and Gorelick, 1996). The cell-based approach assign a facies type to the grid cells according to a facies occurrence probability, which is defined as a conditional probability calculated for each cell



(a) Sparse distributed well data



(b) Determined model from geologist



(c) Rasterized determined model

Figure 1.5: The deterministic facies model from the geologists based on well data and the geological understanding

from the known data in the model by using different algorithms. In object-based approaches, the cells are grouped into objects that are simulated together in the domain (Holden et al., 1998). Although the object-based approach can produce clear geometric patterns, it requires a detailed understanding of object properties such as size, orientation, and erosion or intersection relationships that are difficult to obtain for the subsurface geobodies due to the complex sedimentary history. For the cell-based approach, the local hard data are reproduced in the final model and the observed statistical character is reproduced.

1.4 Motivation

This thesis is mainly concerned with methods for facies model construction. The ultimate aim is the facies realization will have a more geologically realistic heterogeneity characters. To reach that aim, the focus is on two major aspects of facies modelling: inference of spatial variability statistic and the calculation of conditional probabilities used to infer facies at unsampled locations.

1.4.1 Facies spatial variability statistics

There are some properties of the data used for facies modelling. The data used for facies modelling are mostly indirect measurements. Although, core data are direct measurements, they are seldom available. The most widely used are well logs that are usually obtained along the vertical direction.

Data for facies modelling are based on small support volume. Usually, a core sample would only be several metres long from a well if they are sampled during the drilling. The diameter of a core is only about 10 centimeters. The facies data from this entire core would provide valuable information for this well, but not every well will have such core data. A well log usually available along the entire well, but the lateral detected diameter of a well log is also limited to several meters from the bore hole. The seismic data is extensive along lateral direction but has limited vertical resolution. The different volume sizes of the available data is shown in

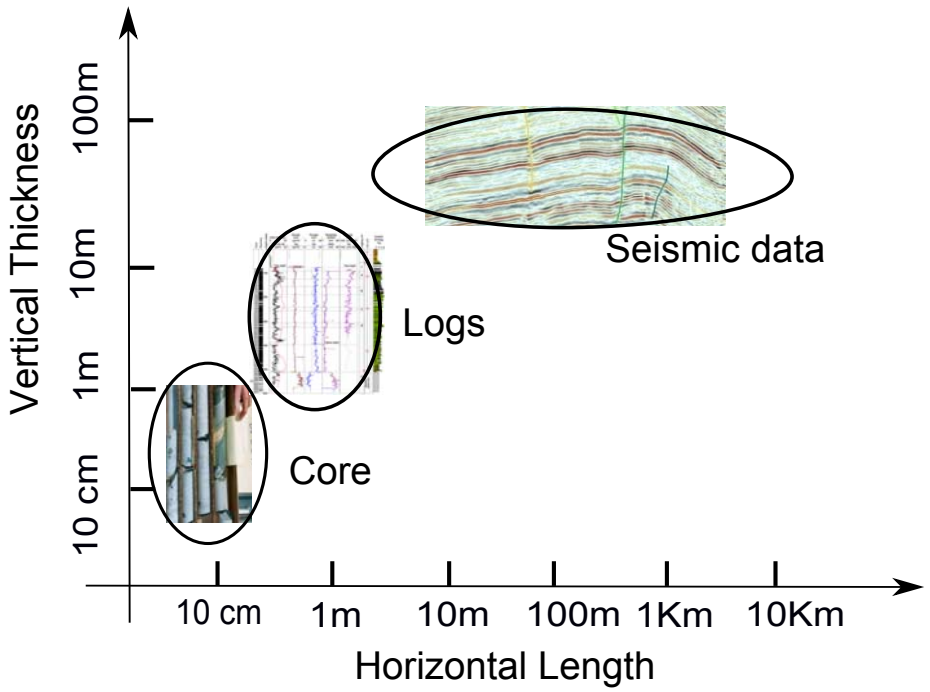


Figure 1.6: Different data resolution along vertical and horizontal direction

Figure 1.6.

Another characteristic of facies data is the different scales along different spatial directions. The hard data from well logs and core are usually only have a high sample rate along the vertical direction while in a vary sparse style horizontally. On one side, it is relative easy to obtain the spatial statistic along the vertical direction where enough data are sampled. On the other side, it is usually difficult to obtain reliable statistics along the lateral direction where data are spaced several hundred meters or even a kilometer or more as shown in Figure 1.7. Furthermore, the spatial statistics such as variogram inferred from spares data sample may not be very reliable.

Interpreting the lateral heterogeneity is always a challenge in geostatistics. In this thesis, a new method to infer more representative spatial statistics from limited data is proposed. Instead of the indicator variogram, the bivariate probability matrix is used as a spatial characterization tool. There are some information such as the facies stacking patterns that is not captured by the traditional variogram. Using the

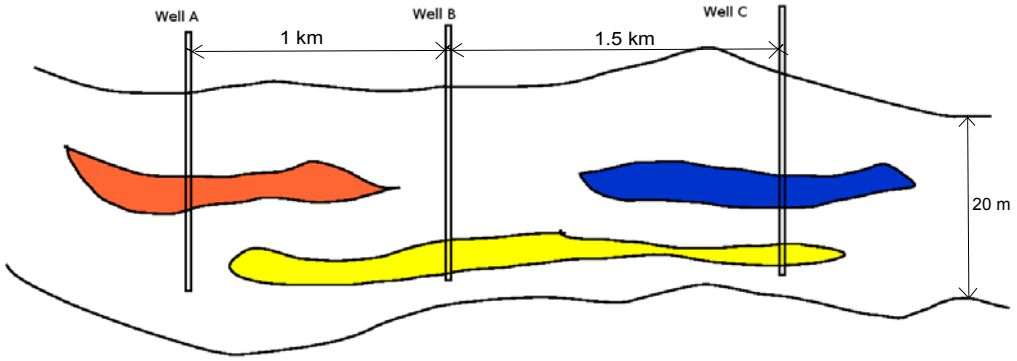


Figure 1.7: One common situation of the difficulty of inferring horizontal statistics from limited sampled wells. Horizontally, the distance between the wells are usually more than 0.5 Km; Vertically, the well may only penetrate no more than 50 meters of the target formation.

bivariate probability matrix, the facies stacking pattern information will be characterized and used for more reliable horizontal statistic inference based on limited data. Thus, it will integrate more geological constraints into the final facies model and provide a more realistic facies model.

1.4.2 Conditional probability calculation

In the cell-based modelling approach, the facies is assigned to each grid according to a calculated conditional probability that can be written as:

$$P(\text{faices at current location} \mid \text{information from nearby locations})$$

There are many mathematical algorithms to estimate such posterior probabilities including: Indicator Kriging (IK) and multi-point statistics (MPS). IK is one of the most popular techniques used in cell-based facies modelling; however, there are limitations such as the linear nature of the kriging predictor and the occurrence of probabilities outside of the $[0,1]$ interval. Also, for MPS, there are some constraints such as the heavy reliance on a training image. There is a need to find a better approach to combine the information from the data locations. In this thesis, a theoretically correct approach is proposed and implemented in facies modelling. The

conditional probability in this approach is calculated directly from the estimated multivariate probability.

1.5 Dissertation Outline

Some preliminary concepts and literature review related to this research are described in Chapter 2. The main concepts and also the related previous research on discrete multivariate probability estimation is reviewed in this Chapter.

Chapter 3 will address the first challenge in facies modelling, that is, how to integrate more geological constraints into the spatial characterization. The traditional variogram is replaced with a new bivariate probability matrix diagram. The characteristics and merits of this new approach and its implementation are discussed in this Chapter.

Another concepts in Chapter 3 is the "heterogeneity prototype" that is used in the spatial distance vector transformation. The geological basis of this concept is presented. Based on this prototype, a new spatial distance calculation approach is proposed instead of using the traditional geometric approach. This approach permits a geological understanding about how the spatial variations will extend along horizontal direction integrated into the final model.

After inferring a bivariate probability model, its use in spatial multivariate probability inference is illustrated in Chapter 4. In this chapter, a new multivariate probability estimation technique, named as Direct Multivariate Probability Estimation (DMPE), is proposed. This technique is based on the Maximum Entropy principle. The theory and numerical implementation of the new proposed approach are given in detail in this fourth Chapter.

The implementation of the new spatial characterization tools and the discrete multivariate probability estimation method in the facies model is given in Chapter 5. A comparison with other approaches is illustrated with the simulation results and other quantitative measurements. Also, the program implementation details regarding the new methodology are presented in this Chapter.

Chapter 6 shows a case study with a real data set. In this case study, it is shown that heterogeneity prototype can be recognized from the data set through geological well correlation. The estimation and simulation results are presented using the proposed spatial distance calculation and DMPE approach.

Final thoughts and future work are given in Chapter 7.

Chapter 2

Preliminary Concepts and Literature Review

Considering a random variable and probability model for the outcome at each unsampled location, a discrete multivariate random variable is defined. Multivariate probability distribution concepts are reviewed. The multivariate probability marginalization is reviewed after introducing the multivariate data event index calculation. Inferring probability distributions from the sample data is an important research topic in geostatistics. For discrete random variables, the Indicator Kriging and the training image scanning approach are common and are reviewed.

2.1 Spatial Random Variable

Geology is governed by the laws of physics. The facies structure in reservoirs are formed and modified by physical sedimentary processes. One approach to facies modelling is simulating the sedimentary processes in a forward style. These techniques are called process-based methods. Some process-based examples are the SEDSIM project (Tetzlaff and Harbaugh, 1989; Wendebourg and Dominik Ulmer, 1992; Hutton and Syvitski, 2008) and the work of Mackey and Bridge (Mackey and Bridge, 1995; Karssenberg et al., 2001). In the process-based model, the diffusion equation is used to describe the process of sediment transport. With properly chosen initial values and boundaries conditions (Tsynkov and Vatsa, 1998), the numerical

solution of the diffusion equation can produce realistic geological models (Tetzlaff and Harbaugh, 1989; Marr et al., 2000).

Advantages of process-based models are that they can help provide genetic interpretations of deposits and can obtain realistic sedimentary architecture. However, it is difficult or impossible to make the simulated deposits from process-based model fit the observed data in sufficient detail in three dimensions (Tetzlaff, 1990). Therefore, process-based models have had limited application in quantitative simulation of the facies architecture of hydrocarbon reservoirs (Koltermann and Gorelick, 1996).

One could focus on the sedimentary facies architecture without considering their actual depositional processes. Geostatistics facies modelling approach belongs to this category (Journel, 1986; Philip and Watson, 1987). This approach considers the facies as outcomes of a spatial stochastic process with a certain probability distribution. A spatial random variable is used to express the uncertainty in the outcomes at each location.

2.1.1 Discrete univariate random variable

Let all the possible outcomes for one observation in the research area be denoted by a set $E : \{e_k, k = 1, 2, \cdots, K\}$. The main characteristic of this set are:

(a) it has limited items (usually K is between 2 and 7);

(b) it is not a numerical set.

In this research, the main observation for each spatial location will be the facies type. As an example, the possible outcome for one location could be one type from the set $\{channel, levee, floodplain\}$ in a fluvial depositional environment (Miall, 1996). When dealing with such observations, a discrete random variable is used.

The outcome at any spatial location $\mathbf{u}$, which could be defined by its Cartesian coordinates (x, y, z), will be considered as a spatial random variable $Z(\mathbf{u})$. Mathe-

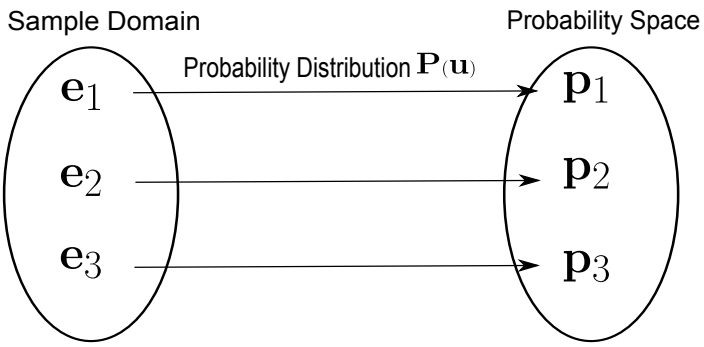


Figure 2.1: A univariate discrete probability distribution example

matically, a spatial discrete random variable $Z(\mathbf{u})$ is defined in words as:

$$Z(\mathbf{u}) = \texttt{the facies at current location}\,\mathbf{u}$$

Common notation will be used in this research such as: upper case letters like Z or Y denote a random variable and lower case letters like z or y denote the outcomes of a random variable. The set E is called the domain of this discrete random variable and all the elements are mutually exclusive and exhaustive in the sense that only one facies type can be found at one specific location $\mathbf{u}$. One possible observation or outcome at a location is written as $z(\mathbf{u}) = e_k$ or simplified as $\mathbf{u} = e_k$ and it will be named as a univariate data event.

After a random variable $Z(\mathbf{u})$ is defined, a probability mass function $P(\mathbf{u})$ is also defined with it. The function that gives the probability that a random variable is exactly equal to some value is called a probability mass function. The probability mass function $P(\mathbf{u})$ for $Z(\mathbf{u})$ is defined as:

$$P(\mathbf{u}) = Pr(z(\mathbf{u}) = e_k; e_k \in E) = Pr(\mathbf{u} = e_k; e_k \in E) \tag{2.1}$$

It is a function that satisfies the properties: $P(\mathbf{u}) \geq 0$ and $\sum P(\mathbf{u}) = 1$.

The outcomes of the random variable $Z(\mathbf{u})$ will obey its probability distribution $P(\mathbf{u})$. A set of probability values $\{p_1, \cdots, p_K\}$ for each facies $e_k \in E$ to be found at location $\mathbf{u}$ will be given from the probability mass function as shown in Figure 2.1.

From the probability mass function of the random variable $Z(\mathbf{u})$, one can understand how likely each facies is at the location. As an example, assuming the proba-

bility mass function of finding one facies from the set $\{channel, levee, floodplain\}$ at one location is $\{0.8, 0.15, 0.05\}$, this means there is a higher probability to find channel at this location than the other two facies.

2.1.2 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:

$$\omega : \{z(\mathbf{u}_1) = e_{k_1}, \cdots, z(\mathbf{u}_n) = e_{k_n}\} \tag{2.2}$$

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 from all locations` $\{e_{k_1}, \cdots, e_{k_n}\}$

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:

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

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 2.2.

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:

$$P(\mathbf{u}_1, \ldots, \mathbf{u}_n) = Pr(\omega_\ell \in \Omega), \ell = 1, \cdots, N \tag{2.3}$$

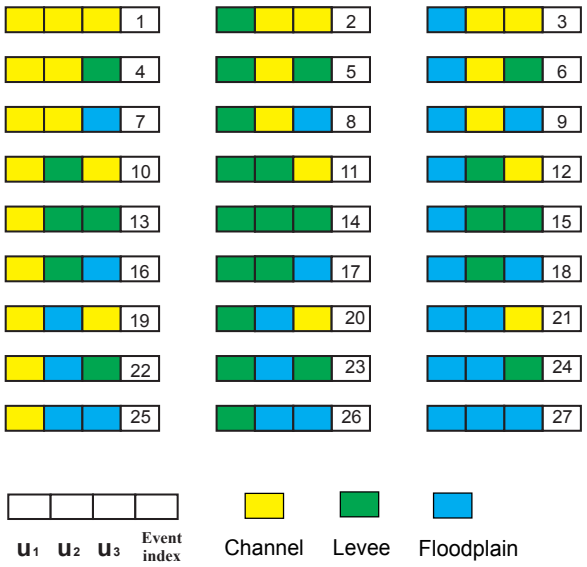


Figure 2.2: One example of all the possible data events composed by three spatial locations and three possible facies

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 $P(\mathbf{u}_1,\ldots,\mathbf{u}_m)$.

The term"probability" will be reserved for a specific probability value which will state for a specified data event such as univariate data event $\mathbf{u} = e_k$ or multivariate data event ω. The probability mass function $P(\cdot)$ will be called a "probability

Table 2.1: 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

distribution".

2.1.3 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 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 2.1, 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 (Deutsch, 1992). 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:

$$\begin{aligned} \ell &= f(\mathbf{u}_1 = e_{k_1}, \cdots, \mathbf{u}_n = e_{k_n}) \\ &= 1 + \sum_{\alpha=1}^{n} (k_\alpha - 1) \times K^{\alpha-1}, \quad k_\alpha = 1, \cdots, K \end{aligned} \tag{2.4}$$

Table 2.2: 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

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 (2.4). Some indices of multivariate data events are listed in Table 2.2. The index of each multivariate event can also be found in Figure 2.2. Each of the data event will have a probability. There will be a total of 27 probabilities.

2.1.4 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 (Johnson and Wichern, 1982).

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

$$P(\mathbf{u}_\alpha) = \sum_{\mathbf{u}_1=e_1}^{e_K} \cdots \sum_{\mathbf{u}_\alpha} \cdots \sum_{\mathbf{u}_n=e_1}^{e_K} P(\mathbf{u}_1,\ldots,\mathbf{u}_n) \quad \alpha = 1,\cdots,n \tag{2.5}$$

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:

$$P(\mathbf{u}_\alpha) = Pr\{\mathbf{u}_\alpha = e_k; e_k \in E\} \tag{2.6}$$

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

$$P(\mathbf{u}_\alpha, \mathbf{u}_\beta), \alpha \neq \beta \quad \alpha, \beta = 1, \cdots, n$$

is calculated from a multivariate probability distribution as:

$$P(\mathbf{u}_\alpha, \mathbf{u}_\beta) = \sum_{\mathbf{u}_1 = e_1}^{e_K} \cdots \sum_{\mathbf{u}_\alpha} \cdots \sum_{\mathbf{u}_\beta} \cdots \sum_{\mathbf{u}_n = e_1}^{e_K} P(\mathbf{u}_1, \ldots, \mathbf{u}_n) \tag{2.7}$$

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:

$$P(\mathbf{u}_\alpha, \mathbf{u}_\beta) = Pr\{\mathbf{u}_\alpha = e_k, \mathbf{u}_\beta = e_{k'}; e_k, e_{k'} \in E\} \tag{2.8}$$

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:

$$P(\mathbf{u}_1, \cdots, \mathbf{u}_m) = \sum_{\mathbf{u}_1} \cdots \sum_{\mathbf{u}_m} \sum_{\mathbf{u}_{m+1} = e_1}^{e_K}, \cdots, \sum_{\mathbf{u}_n = e_1}^{e_K} P(\mathbf{u}_1, \ldots, \mathbf{u}_n) \tag{2.9}$$

In expression (2.9), 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 2.2 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 (2.4), 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 2.3 where $(p_1, \cdots, p_{27})$ are multivariate probabilities.

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

Figure 2.3: A univariate marginalization from a trivariate probability

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 (2.4). 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 2.4.

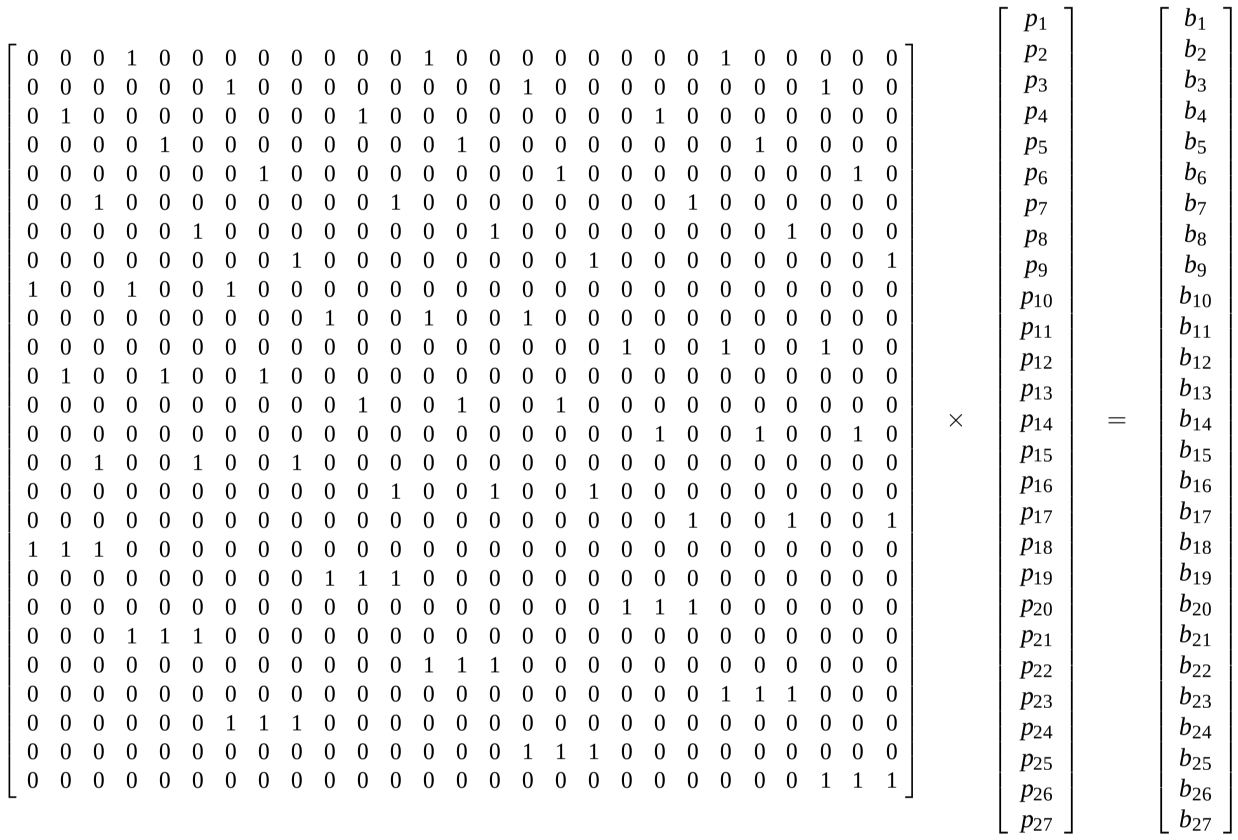


Figure 2.4: The bivariate marginalization from a trivariate probability

The above marginalization procedure is also illustrated in Figure 2.5. 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 (Pickover, 1999). Intuitively plotting them in a picture as Figure 2.5 is a challenge.

2.1.5 Prior and posterior probability

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:

$$P(\mathbf{u}_0) = \frac{1}{K} \tag{2.10}$$

It will be shown that this uniform probability distribution is the maximum entropy estimation.

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:

$$P(\mathbf{u}_0) = Pr\{\mathbf{u}_0 = e_k; e_k \in E\} = p_k \tag{2.11}$$

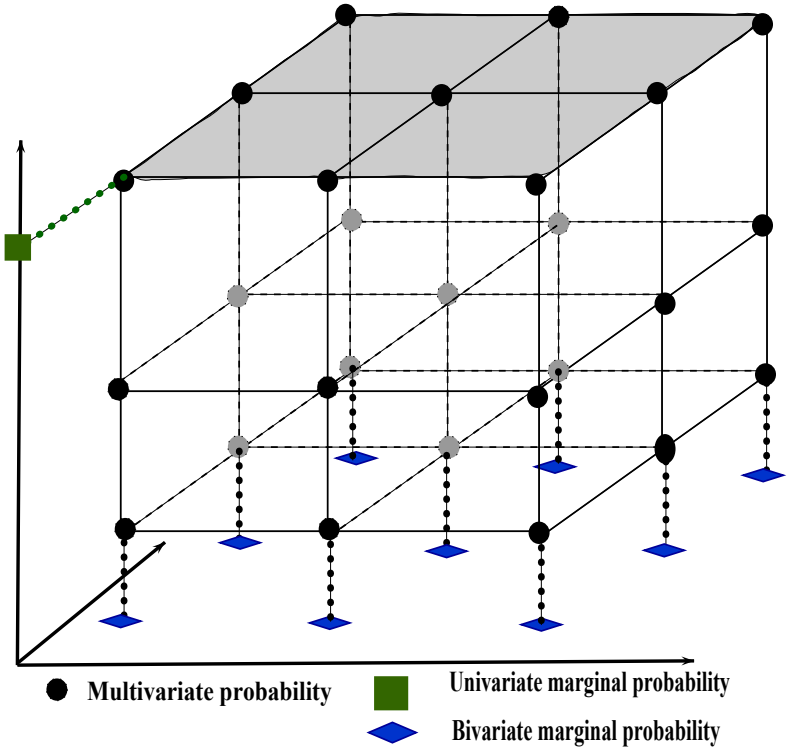


Figure 2.5: Univariate and bivariate marginalization from a trivariate probability space

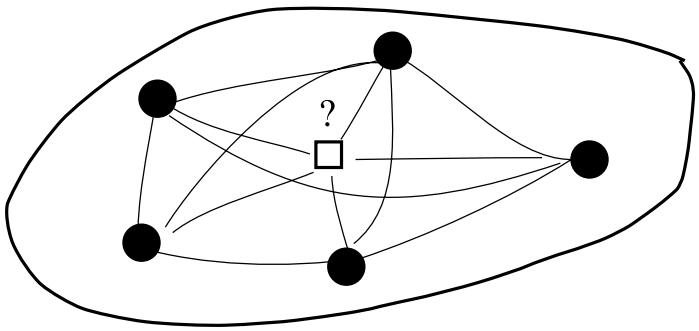


Figure 2.6: One situation of multivariate probability needed in spatial probability mapping

If there are some sample data and they are related to the unsampled location, as illustrated in Figure 2.6, 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:

$$P(\mathbf{u}_0|\mathbf{u}_1, \ldots, \mathbf{u}_n) = Pr\{\mathbf{u}_0 = e_{k_0}|\mathbf{u}_1 = e_{k_1}, \cdots, \mathbf{u}_n = e_{k_n}\} \tag{2.12}$$

Updating the prior probability distribution as in Equation (2.10) or (2.11) to a posterior probability distribution model as in Equation (2.12) is the central problem in geostatistics (Deutsch and Journel, 1998).

In probability theory (Papoulis, 1984), a conditional probability as expressed in Equation (2.12) is calculated as:

$$P(\mathbf{u}_0|\mathbf{u}_1, \ldots, \mathbf{u}_n) = \frac{P(\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_n)}{P(\mathbf{u}_1, \ldots, \mathbf{u}_n)} \tag{2.13}$$

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 $P(\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_n)$ using Equation (2.9). The conditional probability $P(\mathbf{u}_0|\mathbf{u}_1, \ldots, \mathbf{u}_n)$ will be calculated directly as:

$$\begin{aligned} P(\mathbf{u}_0|\mathbf{u}_1, \ldots, \mathbf{u}_n) &= \frac{P(\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_n)}{P(\mathbf{u}_1, \ldots, \mathbf{u}_n)} \\ &= \frac{P(\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_n)}{\sum_{\mathbf{u}_1} \cdots, \sum_{\mathbf{u}_n} \sum_{\mathbf{u}_0=e_1}^{e_K} P(\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_n)} \end{aligned} \tag{2.14}$$

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 multivariate Gaussian probability distribution, with parameters (essentially the covariance matrix) inferred from the data. With the multivariate Gaussian assumption, the conditional probability will also be a Gaussian probability and is obtained through estimating its mean and variance (Verly, 1983; Gomez-Hernandez and Wen, 1998; Goovaerts, 2001).

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 (IK) and the training image scanning approach, will be reviewed.

2.2 Posterior Probability Estimation-IK

The IK approach was first proposed by Journel for continuous variable which is difficult to transform into Gaussian variables (Journel, 1983; Journel and Isaaks,

1984). Later, it was also used to estimate the posterior probability for categorical variables (Bierkens and Burrough, 1993). With the proposed and publicly available program for indicator kriging based simulation programs such as `ISIM3D` (Gomez-Hernandez and Srivastava, 1990) or `SISIM` (Deutsch and Journel, 1998), it is widely used for categorical variable simulation in petroleum and other areas (Marinoni, 2003; Bastante et al., 2005).

2.2.1 Indicator transformation

In order to handle categorical variables numerically, the first step in the IK approach is to consider an indicator transformation of the original discrete random variable. An indicator random variable $I(\mathbf{u}, e_k)$ is computed from a traditional discrete random variable $Z(\mathbf{u})$ as:

$$I(\mathbf{u}_\alpha; e_k) = \begin{cases} 1 & \text{If } e_k \text{ exists at location } \mathbf{u}_\alpha \quad \alpha = 1, \cdots, n \\ 0 & \text{Otherwise} \end{cases} \tag{2.15}$$

It is done for each facies separately, as noted by the facies type e_k.

After the indicator transformation of the random variable, a probability distribution is defined for each sampled location where the rock type is obtained. That is, for each facies e_k, the probability distribution of the indicator random function is given as:

$$P(\mathbf{u}_\alpha; e_k) = \begin{cases} 1 & i(\mathbf{u}_\alpha; e_k) = 1 \quad \alpha = 1, \cdots, n \\ 0 & i(\mathbf{u}_\alpha; e_k) = 0 \end{cases} \tag{2.16}$$

Comparing Equation (2.15) and Equation (2.16), it is seen that the indicator variable can be considered as the probability of facies e_k to be found at the location $\mathbf{u}_\alpha$. When the location is sampled there is no uncertainty and the probability for the facies that prevails at current location is 1, and 0 for others.

The outcome for this indicator random variable will be denoted as $i(\mathbf{u}, e_k)$ and its domain will be $\{0, 1\}$ which is a numerical value and can be treated with traditional statistics. Thus, the mean of the random variable $I(\mathbf{u}; e_k)$ in the research area can

be calculated as:

$$\begin{aligned} E[I(\mathbf{u};e_k)] &= 1\cdot p(i(\mathbf{u};e_k)=1)+0\cdot p(i(\mathbf{u};e_k)=0) \\ &= p(i(\mathbf{u};e_k)=1) \\ &= Pr(\mathbf{u}=e_k)=p_k \end{aligned} \tag{2.17}$$

Practically, the estimated mean of a indicator random variable $I(\mathbf{u},e_k)$ can be estimated from its n outcomes $i(\mathbf{u}_\alpha;e_k)$ as:

$$p_k = \frac{1}{n}\sum_{\alpha=1}^{n} i(\mathbf{u}_\alpha;e_k) \tag{2.18}$$

where p_k is the global mean which will reflect the average frequency of facies e_k that can be found in the study area. A further practical refinement of the global proportion inference will be to consider declustering weights applied to each of the n data (Pyrcz et al., 2006).

Also, the indicator covariance from two locations $\mathbf{u},\mathbf{u}+\mathbf{h}$ will be calculated as:

$$Cov(\mathbf{u},\mathbf{u}+\mathbf{h};e_k,e'_k) = E[I(\mathbf{u};e_k)I(\mathbf{u}+\mathbf{h};e'_k)] - E[I(\mathbf{u};e_k)]E[I(\mathbf{u}+\mathbf{h};e'_k)] \tag{2.19}$$

Under the stationary assumption, the above equation can be written as:

$$\begin{aligned} Cov(\mathbf{h};e_k,e'_k) &= E[I(\mathbf{u};e_k)I(\mathbf{u}+\mathbf{h};e'_k)] - E[I(\mathbf{u};e_k)]E[I(\mathbf{u}+\mathbf{h};e'_k)] \\ &= P(\mathbf{h},e_k,e'_k) - p_k\cdot p_{k'} \end{aligned} \tag{2.20}$$

In Equation (2.19) and (2.20), if the facies from the two locations are the same, it will be called a direct covariance. If not, it is called a cross covariance.

2.2.2 Linear combination model

In the IK approach, the conditional probability distribution $P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)$ is estimated directly by estimating the mean of the corresponding conditional indicator variable $E[I(\mathbf{u}_0;e_k|\mathbf{u}_1,\ldots,\mathbf{u}_n)]$ as shown in Equation (2.21):

$$\begin{aligned} E[I(\mathbf{u}_0;e_k|\mathbf{u}_1,\ldots,\mathbf{u}_n)] &= 1\cdot P(\mathbf{u}_0=e_k|\mathbf{u}_1,\ldots,\mathbf{u}_n)+0\cdot P(\mathbf{u}_0\neq e_k|\mathbf{u}_1,\ldots,\mathbf{u}_n) \\ &= P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n) \end{aligned} \tag{2.21}$$

The final estimated for $p(\mathbf{u}_0; e_k|\mathbf{u}_1, \ldots, \mathbf{u}_n)$ is obtained by a linear combination of the hard data at each location as

$$p^*(\mathbf{u}_0; e_k|\mathbf{u}_1, \ldots, \mathbf{u}_n) - p_k = \sum_{\alpha=1}^{n} \lambda_{\alpha,e_k} \cdot [p(\mathbf{u}_\alpha; e_k) - p_k] \tag{2.22}$$

In Equation (2.22), the hard data at each location is expressed as probability interpretation as expressed in Equation (2.16). The weights λ_{α,e_k} are calculated from the simple kriging system or the system of normal equations as:

$$\sum_{\alpha=1}^{n} \lambda_{\alpha,e_k} Cov(\mathbf{h}_{\alpha,\beta}; e_k) = Cov(\mathbf{h}_{\alpha,0}; e_k), \beta = 1, \cdots, n \tag{2.23}$$

where $Cov(\mathbf{h}_{\alpha,\beta}; e_k)$ is the covariance model obtained from the indicator random variables, $\mathbf{h}_{\alpha,\beta}$ is the distance matrix between all the conditioning data, $\mathbf{h}_{\alpha,0}$ is the distance matrix between the unsampled location and conditioning data location (Cressie, 1990).

Although the indicator transformation is a nonlinear transform, the estimation results for each facies at an unsampled locations are linear interpolations of probability values from the hard data. The results of such probability interpolation for each facies separately has no guarantee that the result will be inside $[0, 1]$. The sum of the probability distribution for each location will not necessarily be one $\sum_{k=1}^{K} p(\mathbf{u}, e_k) \neq 1$ which is called order relationship problem in indicator kriging approach (Journel and Posa, 1990).

Practically, the recommended solution is resetting non admissible probability values to the closest bound 0 or 1, followed by standardization so that they will sum up to one, but the potential consequences of these order relation problems are poorly understood.

A more efficient way of posterior probability updating would be to work directly on all the categories together. Theoretically, from the conditional probability definition, it can be calculated provided that the multivariate probability distribution $P(\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_n)$ is obtained. This will require an estimate of the multivariate probability directly. Training image scanning is such an approach.

2.3 Posterior Probability Estimation-Training Image Scanning

In the traditional IK approach, the $(n+1)$ multivariate probability values are estimated by interpolation from data. Inferring the conditional probabilities from scanning a training image suggested by Guardiano and Srivastava in 1993 is a kind of direct approach (Guardiano and Srivastava, 1993). The required $(n+1)$ multivariate probabilities in Equation (2.14) is constructed by scanning the frequency of data event from a training image. The algorithm of Guardiano and Srivastava was further developed by Wang named growing algorithm trying to capture the structural patterns (Wang, 1996). Strebelle proposed an algorithm to estimate the conditional probability by using the search tree to store the training data events prior to the image simulation (Strebelle, 2002).

2.3.1 Training image

The training image is an important tool used in this multivariate probability estimation method. The needed multivariate probabilities will be obtained from scanning a training image. The training image should represent the heterogeneities characteristics of study area. Although the training image itself is not the underlying phenomenon, it should bear the same spatial structures of the geological sedimentary structure as the reservoir. For example, for a fluvial sand environment, a training image of facies distribution could be like the map shown in Figure 2.7

The training image can come from an unconditional realization generated using the object-based algorithm (Holden et al., 1998; Maharaja, 2008). The geologists hand-drawn map can be digitalized and used as training image. For example the map in Figure 1.5b could be used as a training image. It also can come from the realization of processed-based modelling approach (Cojan et al., 2005).

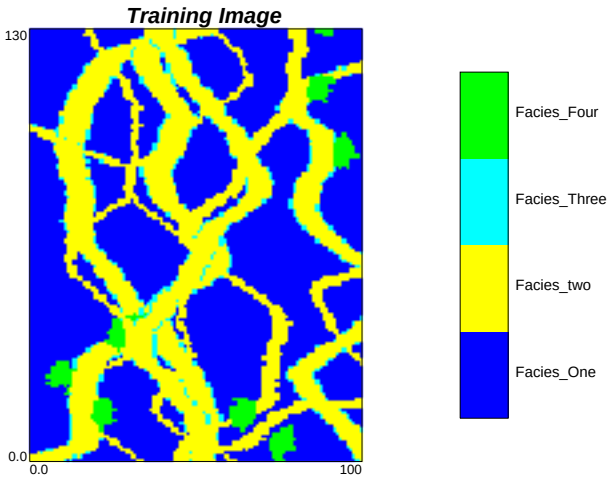


Figure 2.7: One training image for facies modelling

2.3.2 Training image scanning

From the chosen training image, the posterior probability $P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)$ will be retrieved by scanning the data configuration over the training image and counting the relative frequency of events occurred in the training image as:

$$P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n) = \frac{P(\mathbf{u}_0,\mathbf{u}_1,\ldots,\mathbf{u}_n)}{P(\mathbf{u}_1,\ldots,\mathbf{u}_n)} \simeq \frac{Count(\mathbf{u}_0 = e_{k_0},\mathbf{u}_1 = e_{k_1},\cdots,\mathbf{u}_n = e_{k_n})}{Count(\mathbf{u}_1 = e_{k_1},\cdots,\mathbf{u}_n = e_{k_n})} \tag{2.24}$$

In Equation (2.24),the denominator $Count(\mathbf{u}_1 = e_{k_1},\cdots,\mathbf{u}_n = e_{k_n})$ is the number of the conditioning multivariate data event in the training image.While the numerator $Count(\mathbf{u}_0 = e_{k_0},\mathbf{u}_1 = e_{k_1},\cdots,\mathbf{u}_n = e_{k_n})$ is the number of the conditioning locations plus the unsampled location.

The training image approach has several advantages to the conditional probability estimation. First, it brings the information from more than just two point statistics. Thus, it has the ability to reproduce some curvilinear geological shape (Strebelle, 2002). Second, the conditional probability calculated from Equation (2.24) does not suffer any order deviations as may happen in the kriging approach.

One major limitation of the training image approach is that the multivariate data event cannot include too many grid nodes. It depends on the repetition of

the data event from the training image and there should be enough for a reliable probability inference in Equation (2.24). If too many nodes are included in the data event, the training image would have be very large. For example, given a data event composed by three categories and 15 locations, the number of configurations would be 3^{15}. If each of them require 10 repetitions, the dimension of the training image would be $10 \times 3^{15} = 143,489,070$. Handing this high dimension training image is a challenge.

2.4 Remarks

The traditional indicator kriging approach simplifies the random variable to a binary variable. Resulted from linear interpolation to the probabilities, the order relation deviation problem is a concern with this approach.

The posterior probability estimated from the training image is constrained by the size of the data event. The scanned multivariate probability depends on the repetition of each data event and requires a large training image. The training image must be representative of the subsurface heterogeneity.

Chapter 3

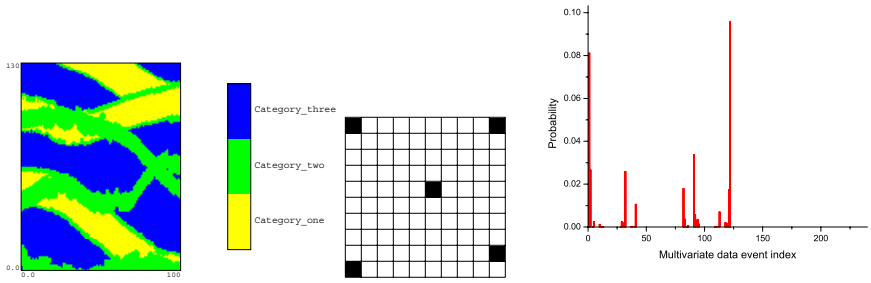
Characterizing Spatial Heterogeneity

Characterization of spatial relationships is the first step in probability estimation at unsampled locations. Using the bivariate probability diagram instead of the variogram will make it possible to integrate more geological constraints into the model. Practically, there is a need to estimate the bivariate probability in all directions to reflect the spatial anisotropy. Traditionally, anisotropy is assumed geometric, which is not appropriate for some geological situations. Based on the proposed sedimentary prototype, a new spatial distance scaling approach is proposed. The bivariate probability for any spatial distance vector will be obtained by transforming this distance vector to an effective vertical distance. The bivariate probability is then obtained from the estimated vertical direction.

3.1 Multiple-Point Statistics

The main challenge to reproduce geological structures in geostatistical models is describing the spatial variability. The natural way would be to describe the entire features at several locations to delineate the full 3D spatial extension. For a geologist, the information from several wells or from different kinds of information are grouped together to build the final facies model which will fit the facies contact pattern of a conceptual model and the current specific situation.

This kind of pattern reproduction is mimicked in geostatistics with multiple-point statistics that will reproduce the spatial relationship from more than two point



(a) Training image **(b)** Configuration **(c)** Multivariate probability

Figure 3.1: One multivariate probability from using one data configure after scanning the training image. In this case, there are totally 5 locations and 3 facies, the multivariate data event total number is 256.

informations (Strebelle, 2002; Arpat and Caers, 2007). Such inference will account for correlations between three or more locations at a time.

The multivariate probability distribution $P(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ will characterize the probability of each multivariate data event $(\mathbf{u}_1 = e_{k_1}, \cdots, \mathbf{u}_n = e_{k_n})$. The information from more than one location are carried with this statistics. Although inference from limited data source is difficult, it can reproduce more features from a properly chosen training image (Mirowski et al., 2009). Also, in facies simulation, the challenge of quickly inferring the conditional probability for one multivariate data event from the whole data event space is relieved by using a search tree to save the frequency of the multivariate data events in the training image. This is the `snesim` algorithm (Strebelle, 2000).

In visualizing multivariate probabilities, one can plot them as the traditional histogram as shown in Figure 3.1. The probability or frequency of each data event is plotted together with its index. Practically, however, without looking at the underlying training image that was used for the multivariate probability inference, it is impossible to know the features from the histogram as plotted in Figure 3.1.

As mentioned, the inference of multiple point statistics requires a vast amount of data on a regular grid which is never available in the subsurface. If we reduce the

focus to two locations, reliable inference can be obtained from well data in many situations.

3.2 Traditional Two-point Spatial statistics

One common observation is that the closer together two locations, the more similar they are. This kind of spatial distance dependent correlation is characterized by the variogram function. The variogram accounts for bivariate statistics and is easier to infer from limited sampled data.

For a discrete spatial random variable, the indicator covariance/variogram usually is used in facies modelling (Lark and Beckett, 1998; Deutsch and Journel, 1998). A variant of the indicator covariance is Markov transition probability matrix (Carle and Fogg, 1996, 1997). These statistics have been used in one, two and three dimensional spatial heterogeneity characterization (Weissmann et al., 1999; Carle et al., 2006). These two will be reviewed first before introducing the bivariate probability matrix model that will be used later in this thesis.

3.2.1 Indicator covariance/variogram

Theoretically, the variogram $\gamma(\mathbf{h})$ is a function that describes the spatial dependence degree of a spatial random variable $Z(\mathbf{u})$. It is the variance of the increment of two random variables at locations $\mathbf{u}$ and $\mathbf{u}+\mathbf{h}$ and with the assumption that $\gamma(\mathbf{h})$ is dependent upon the value of $\mathbf{h}$, and independent of location $\mathbf{u}$, the variogram is defined as:

$$2\gamma(\mathbf{h}) = Var[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})] \tag{3.1}$$

For categorical spatial random variable, the spatial relationship is characterized by the indicator variogram function $\gamma_k(\mathbf{h})$ which is calculated from the indicator random variable as:

$$2\gamma_k(\mathbf{h}) = Var[I(\mathbf{u};e_k) - I(\mathbf{u}+\mathbf{h};e_k)] \tag{3.2}$$

where $I(\mathbf{u};e_k)$ is a binary variable that is obtained with the indicator transformation to the categorical variable see Equation (2.15).

The indicator covariance is needed to solve the kriging equations. The relationship between the indicator variogram and indicator covariance can be written as:

$$\gamma_k(\mathbf{h}) = Cov(0;e_k) - Cov(\mathbf{h};e_k) \tag{3.3}$$

with $Cov(0;e_k) = Var[I(\mathbf{u};e_k)]$ being the stationary variance and $Cov(\mathbf{h};e_k)$ being the stationary covariance defined in Equation (2.20).

From Equation (3.2), it can be seen that the indicator variogram will only count those transition probabilities that satisfy $I(\mathbf{u};e_k) - I(\mathbf{u}+\mathbf{h};e_k) \neq 0$. Thus, only those data pairs that transition from the current facies e_k to another facies will contribute to this indicator variogram. For example, there is no difference for the transition of $e_1 \rightarrow e_2$ and $e_1 \rightarrow e_3$, as both of them are calculated as $1 \rightarrow 0$ when calculating the indicator variogram for facies e_1. From the view of bivariate probabilities, the indicator variogram can be expressed as :

$$2\gamma_k(\mathbf{h}) = \sum_{\substack{k'=1 \\ k'\neq k}}^{K} p(\mathbf{u}_1 = e_k, \mathbf{u}_1 + \mathbf{h} = e_{k'}) + \sum_{\substack{k'=1 \\ k'\neq k}}^{K} p(\mathbf{u}_1 = e_k, \mathbf{u}_1 - \mathbf{h} = e_{k'}) \tag{3.4}$$

If bivariate probabilities could be used directly in the estimation model, more facies transition information would be integrated.

In practice, the experimental indicator variogram is calculated from the sample data as:

$$2\hat{\gamma}_k(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{j=1}^{N(\mathbf{h})} [i(\mathbf{u}_j;e_k) - i(\mathbf{u}_j+\mathbf{h};e_k)]^2 \tag{3.5}$$

Where $N(\mathbf{h})$ is the total number of pairs falling in the same lag.

The experimental variogram is not used directly in geostatistical analysis. One reason is that the experimental variogram is only calculated at specific directions and lag distances $\mathbf{h}$ such as the direction in vertical and horizontal. In most of geostatistics modelling algorithms, the variogram value for all lag vectors is needed;

even, along directions with too few data. The experimental variogram may not satisfy the mathematic requirement of positive definiteness (Wackernagel, 2003). In applied geostatistics, the experimental variograms are modelled by certain mathematical functions to ensure their validity (Cressie and Hawkins, 1980; Chiles and Delfiner, 1999).

3.2.2 Markov transition probability matrix

In many situations, a sequence in either time or space is observed as a succession of states that are taken from a limited set of alternatives. If the natural processes exhibit an effect of previous events, but do not rigidly control subsequent events, these processes are named "Markov chains" after the work of the Russian mathematician Markov (Anderson and Goodman, 1957; Meyn and Tweedie, 2008). Markov properties have been recognized in many geological phenomena, including the stratigraphic sequence of lithologic units, sedimentary processes, succession of mineral occurrencesin igneous rocks, and so on (Miall, 1973; Hattori, 1976; Hiscott, 1981; Carr, 1982; Doveton, 1995; Oliver et al., 1997; Xu and Maccarthy, 1998).

When a Markov chain has a very short memory that only extends for a single step at a time and ceases for larger distances. Such a chain is termed as a first-order Markov chain. The relationship between adjacent events will be summarized by the transition probability denoted as $t_{kk'}(\mathbf{h}) = p(\mathbf{u}+\mathbf{h} = e_k|\mathbf{u} = e_{k'})$. All the K^2 transition probabilities will compose a Markov transition probability matrix $T(\mathbf{h}) = \{t_{kk'}\}; k,k' = 1,\cdots,K$ 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). For example, a Markov transition matrix $T(\mathbf{h})$ with three events may be written as:

$$T(\mathbf{h}) = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \tag{3.6}$$

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

Table 3.1: One step Markov transition probability matrix example

	Facies 1	Facies 2	Facies 3
Facies 1	0.2632	0.3158	0.4210
Facies 2	0.2419	0.2258	0.5323
Facies 3	0.1606	0.5000	0.3194

ability as:

$$t_{kk'}(\mathbf{h}) = p(\mathbf{u}+\mathbf{h} = e_k|\mathbf{u} = e_{k'}) = \frac{p(\mathbf{u}+\mathbf{h} = e_k, \mathbf{u} = e_{k'})}{p(\mathbf{u} = e_{k'})} \tag{3.7}$$

This is only a one-step transition probability, because the current state is only dependent upon the immediately preceding state. Thus, the stochastic process is defined as a 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 enough to character it (Haggstrom, 2002) .

Generally, any multiple steps of the dependence chain can also be defined as $T(\mathbf{nh}), n = 1, 2\cdots, L$ which means that the current state will depend on the state that is located at the **nh** distance. In this case, the process will have a long distance memory and is named a multiple step 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 product of one-step transition probability matrices (Harbaugh and Bonham-Carter, 1970). That is, the transition probability at any lag distance $(\mathbf{nh})$ can be calculated as:

$$T(\mathbf{nh}) = T(\mathbf{h} = 1)^n \tag{3.8}$$

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

As an example, assume a one step transition probability matrix as listed in Table (3.1), using the relationship in Equation (3.8) all the transition probability can be calculated and plotted as shown in Figure 3.2.

Markov chain analysis was popular with sedimentologists as a statistical technique to explain and understand the geological cyclicity. The transition probabil-

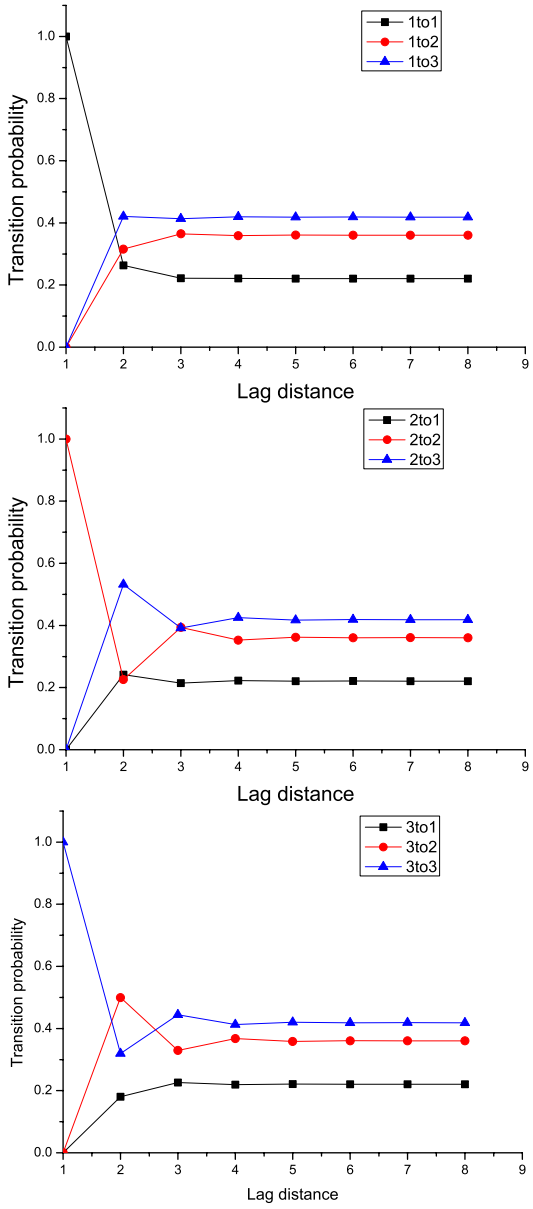


Figure 3.2: One calculated multiple steps transition probability matrix diagram that is calculated using one step Markov transition probability matrix listed in Table (3.1) using Equation (3.8)

ity matrix provides a framework for exploring the underlying physical, chemical, and biological control on sedimentary processes and deposits with superimposed random fluctuations (Allen, 1965; Krumbein and Dacey, 1969; Dacey and Krumbein, 1970; Driese and Dott, 1984). It is also used as an alternative approach to describe the spatial structure, such as in the transition probability-based indicator approach (Carle and Fogg, 1996, 1997; Carle et al., 2006).

3.2.3 Three dimensional anisotropy modelling

Although both indicator variograms and Markov transition probability matrices are used for 3D spatial heterogeneity characterization, little research is available on 3D anisotropy modelling. In this section, the current 3D anisotropy variogram modelling is briefly reviewed.

In the variogram function $\gamma(\mathbf{h})$, the lag distance $\mathbf{h}$ represents a vector with modulus of $|h|$ in a particular direction. The variogram will in general increase with the modulus $|h|$. The spatial variability is characterized by the variogram function in different directions.

Sedimentary systems are anisotropic. There are good reasons to expect the facies distribution to be different in different directions. The data distribution makes it impossible to find the variograms for all the directions. Instead, a valid 3D variogram function will be used (Chiles and Delfiner, 1999). A spatial distance vector $\mathbf{h}$ will be converted to a dimensionless scalar variogram distance. The spatial variability is commonly grouped into geometric and zonal anisotropy by studying the variability of the range with the direction (Gringarten and Deutsch, 2001; Wackernagel, 2003).

The spatial lag vector $\mathbf{h}$ is decomposed into three components h_x, h_y, h_z. A variogram $\gamma(\mathbf{h})$ or a covariance $Cov(\mathbf{h})$ has geometric anisotropy if the anisotropy can be reduced to isotropy by a linear transformation of the coordinates as:

$$\gamma(h_x, h_y, h_z) = \gamma'(\sqrt{(h'_x)^2 + (h'_y)^2 + (h'_z)^2}) \tag{3.9}$$

with

$$h'_x = a_{11}h_x + a_{12}h_y + a_{13}h_z$$
$$h'_y = a_{21}h_x + a_{22}h_y + a_{23}h_z$$
$$h'_z = a_{31}h_x + a_{32}h_y + a_{33}h_z$$

where a_{ij} represents the transformation matrix of the coordinates. Correction of geometric anisotropy is done through such linear transformation of an ellipsoid into a sphere and thus physical anisotropy is expressed as isotropy in the variogram.

This anisotropy modelling procedure is a required step in geostatistical studies. The variogram quantifies the direction-dependent spatial variability of the variable under consideration.

3.3 Bivariate Probability Matrix Model

Although two point statistics do not completely characterize complex shapes, they are easy to infer and implement in modern geostatistics. The indicator variogram $\gamma_k(\mathbf{h})$ quantifies part of the bivariate information from facies transitions. The bivariate probability matrix denoted as $P(\mathbf{h};k,k')$ will be used directly as it contains all of the bivariate information.

3.3.1 Experimental bivariate probability matrix

A bivariate probability for two locations separated by a lag distance $\mathbf{h}$ is defined as

$$p(\mathbf{h};k,k') = Pr(\mathbf{u} = e_k, \mathbf{u}+\mathbf{h} = e_{k'}) \tag{3.10}$$

where k and k' are the integer transformation of the original set $\{e_1, e_2, \cdots, e_K\}$ as done in the multivariate probability index calculation in Equation (2.4). A bivariate probability matrix $P(\mathbf{h};K,K)$ with $K \times K$ entries all possible facies transitions. For example, for a spatial domain with three categories, the bivariate probability matrix

could be written as:

$$P(\mathbf{h};K,K) = \begin{bmatrix} p(\mathbf{h};1,1) & p(\mathbf{h};1,2) & p(\mathbf{h};1,3) \\ p(\mathbf{h};2,1) & p(\mathbf{h};2,2) & p(\mathbf{h};2,3) \\ p(\mathbf{h};3,1) & p(\mathbf{h};3,2) & p(\mathbf{h};3,3) \end{bmatrix} \tag{3.11}$$

Typically, data are abundant along the vertical direction from well log data analysis. Thus, multiple bivariate probability matrices are usually constructed from the vertical profile for different lag vectors.

The lag $\mathbf{h}$ may be an exact distance measure or an offset of spatial steps (pixels or grid cells). Under the assumption that the bivariate probability matrix $P(\mathbf{h};K,K)$ is only dependent on the lag $\mathbf{h}$ and not on any specific location $\mathbf{u}$, the bivariate probability $P(\mathbf{h};K,K)$ could be estimated from data pairs in a similar way as the traditional variogram calculation (Cressie, 1991).

The first step is to construct a facies transition tally matrix by observing the facies outcomes at different equally spaced intervals. For example, a tally matrix calculated from a vertical profile consisting of three facies types is illustrated in Figure 3.3. In this tally matrix, the number of facies C succeeds A is 16 which means that founding facies C and A at these two locations at the same is 16 times totally.

One property of the tally matrix is that the i^{th} row total is equal to the i^{th} column total. For example, the row sum (or column sum) of the calculated tally matrix is:

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

The second step is to divide the tally matrix by the total sum of the tally matrix. The results would be a bivariate probability matrix $P(\mathbf{h}=1;3,3)$ as shown in Figure 3.3c. Division of the tally matrix by each of the row totals leads to the traditional transition probability matrix $T(\mathbf{h}=1;3,3)$ as shown in Table 3.1. 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:

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

(a) Facies vertical profile

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

(b) Tally matrix

	Facies A	Facies B	Facies C
Facies A	0.05813	0.06976	0.09302
Facies B	0.08720	0.08139	0.19186
Facies C	0.07558	0.20930	0.13372

(c) Bivariate probability matrix

Figure 3.3: Example of vertical facies profile and the calculated tally matrix and bivariate probability matrix

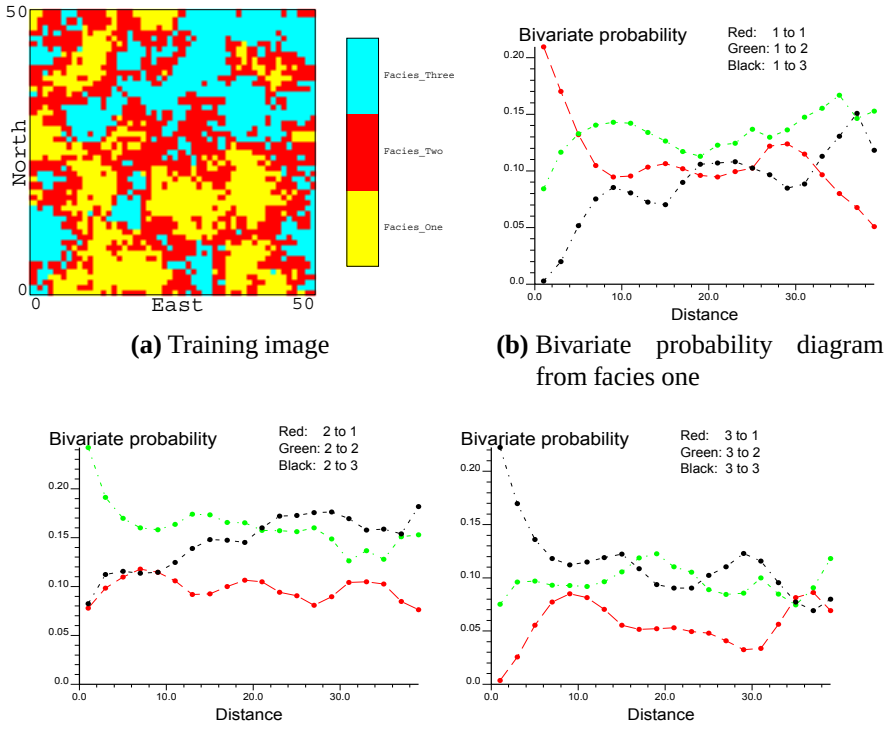
Using the same procedure, a bivariate probability matrix can be obtained from a facies profile given a lag separation.

3.3.2 Properties of bivariate probability matrix diagram

As discussed above, different bivariate probability matrices would be obtained given different lag distances. If the lag distance $\mathbf{h}$ increases from zero to a further distance, the bivariate probability matrices $P(\mathbf{h};K,K)$ will form a diagram. As an example, a bivariate probability diagram calculated from a training image along North-to-South direction is shown in Figure 3.4.

Some basic properties of an experimental bivariate probability matrix diagram are:

1. There are direct bivariate probability $p(\mathbf{h};k,k)$ and cross bivariate probability $p(\mathbf{h};k,k')$. Direct bivariate probability represent auto-correlations of individual categories, and cross-bivariate probability represents cross-relationship between different categories. Under stationarity, the bivariate probability will only be related to the lag distance between two locations without considering



(a) Training image

(b) Bivariate probability diagram from facies one

(c) Bivariate probability diagram from facies two

(d) Bivariate probability diagram from facies three

Figure 3.4: One bivariate probability diagram calculated from the training image along north-to-south

the specific location.

2. Although it is called a bivariate probability, it is different with the traditional bivariate probability in pure statistics. In statistics, a bivariate probability for two random variable X and Y would be a symmetric statistics that is $p(X,Y) = p(Y,X)$. While here, the bivariate probabilities may be asymmetric, that is, $p(\mathbf{h},k,k') \neq p(-\mathbf{h};k,k')$. Thus, the head and tail categories should be defined for each spatial bivariate probability. For bivariate probability $p(\mathbf{h};k,k')$, k is called the head category, k' is called the tail category.

3. They are non-negative and the bivariate probability with the same head and tail category will sum to the univariate probability of the category:

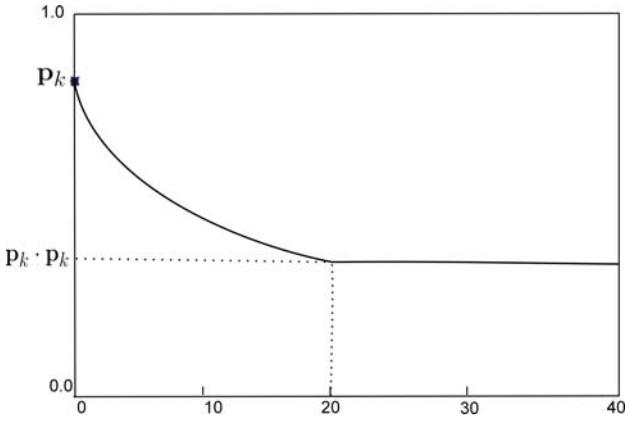
$$\sum_{k'=1}^{K} p(\mathbf{h};k,k') = p_k \quad k = 1,\cdots,K \tag{3.12}$$

4. The initial value of a direct bivariate probability (when $\mathbf{h} = 0$) should be the univariate probability p_k. Where the initial value of cross bivariate (when $\mathbf{h} = 0$) should be zero. As the lag distance increases to ∞ and there is no spatial dependence, the bivariate probability should satisfy:

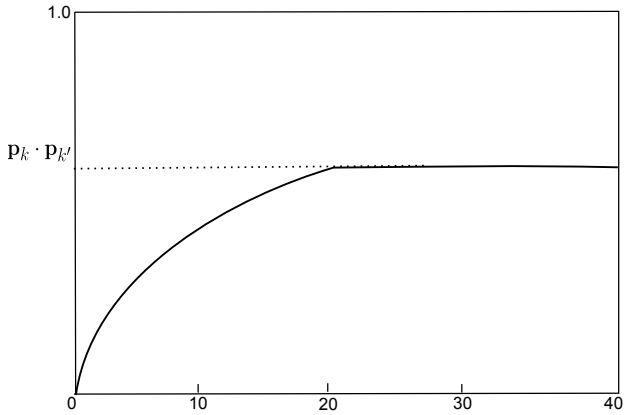
$$p(\mathbf{h};k,k') = p_k \cdot p_{k'} \qquad \mathbf{h} \to +\infty; \quad k,k' = 1,\cdots,K \tag{3.13}$$

   Which is defined as the sill of the bivariate probability diagram. The univariate proportions $p_k, p_{k'}$ are used to calculate the sill. Similar to the variogram, the distance at which the bivariate probability reaches its sill could also be called the range. This is the distance when there is no dependence between the two facies types.

The bivariate probability matrix diagram will be used as the spatial statistic tool in this research. The bivariate probability matrix provides some advantages when quantifying the spatial variability. As list above, the bivariate probability matrix allows a consistent relationship between the model sill and proportions of different



(a) Typical direct bivariate probability diagram



(b) Typical cross bivariate probability diagram

Figure 3.5: Typical features of idealized bivariate probability diagram

facies. For they are always expressed together as a matrix, the global proportion of each facies is embedded in the bivariate probability matrix at each step.

Another important characteristic of the bivariate probability matrix diagram is that the facies dependence relation can be revealed from cross bivariate probability diagrams. Theoretically, a direct bivariate probability diagram $p(\mathbf{h}; k, k)$ will decrease from its global mean p_k to sill value $p_k \cdot p_k$ as shown in Figure 3.5a. An idealized cross bivariate probability diagram $p(\mathbf{h}; k, k')$ starts from the original zero and gradually increases to a stable value–the sill $p_k \cdot p_{k'}$ as shown in Figure 3.5b.

In real data, there could be different shapes. The stacking pattern between different facies is reflected from the shape of the calculated bivariate probability di-

agram (Li, 2006). Before the cross bivariate probability stably approaches its sill, depending on the spatial distribution of the two involved facies (whether they are frequent neighbours or not) the cross bivariate probability shape may be different. If facies k frequently occurs adjacent to facies k', the bivariate probability diagram $p(\mathbf{h};k,k')$ will have a peak first and then approach its sill as shown in Figure 3.6a. If they are not in contact with each other, the bivariate probability would have a lower value at the beginning and then approach its sill later as shown in Figure 3.6b.

Also, the juxtaposition between different facies can revealed from their changing speed. As shown in Figure 3.6c, during the decrease of facies one to one, the bivariate probability of facies one to two will increase faster than the bivariate probability of facies one to three. Thus, facies one will have a high probability of transition to two and less transitions to facies three.

3.3.3 Bivariate probability matrix interpolating

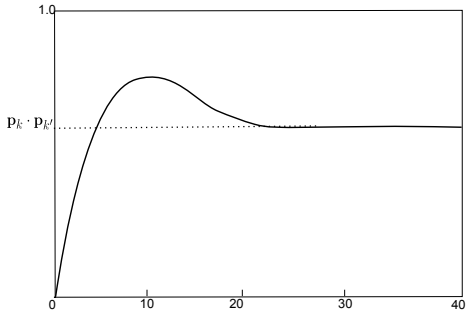
The experimental bivariate probability diagram will only be available for some specific lag distances. In the final construction of a 3D geostatistical model, the bivariate probability could be required for any distance between two neighbour counting lag distance.

The first solution would be to fit the experimental bivariate probability diagrams with some continuous functions. However, there is no procedure to fit all of them with a valid probability model. In this research, a simple interpolation approach is used to estimate all required bivariate probability diagrams.

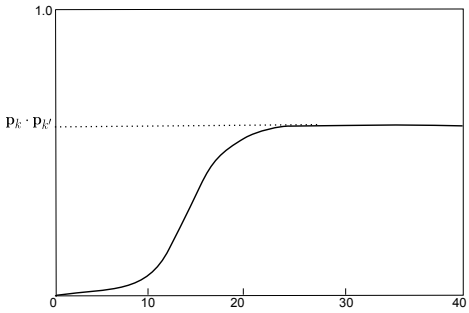
In numerical analysis, polynomial interpolation is a well established interpolation method. Given some points x, it will find a polynomial which goes exactly through these points. Usually the form would be:

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0 \tag{3.14}$$

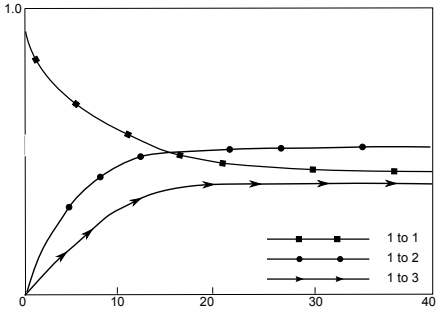
When constructing interpolating polynomials, there is a tradeoff between having a better fit and having a smooth well-behaved fitting function. The more data



(a) Two categories that are frequent neighbours



(b) Two categories that are infrequent neighbours



(c) The stacking pattern information between three categories

Figure 3.6: The juxtaposition information reflected in the bivariate probability diagram shape

points that are used in the interpolation, the higher the degree of the resulting polynomial, and therefore the greater oscillation it will exhibit between the data points. Therefore, a high-degree interpolation may be a poor predictor of the function between points, although the data points would be exactly matched.

In this research, four nearby bivariate probabilities are used to construct the function listed in Equation (3.14). During the interpolation, the cross bivariate probability diagram will be interpolated first. Then, the univariate proportion will be used to calculate the direct bivariate probability. Thus, the univariate probability is enforced during the interpolation. Figure 3.7 shows one bivariate probability diagram interpolation result.

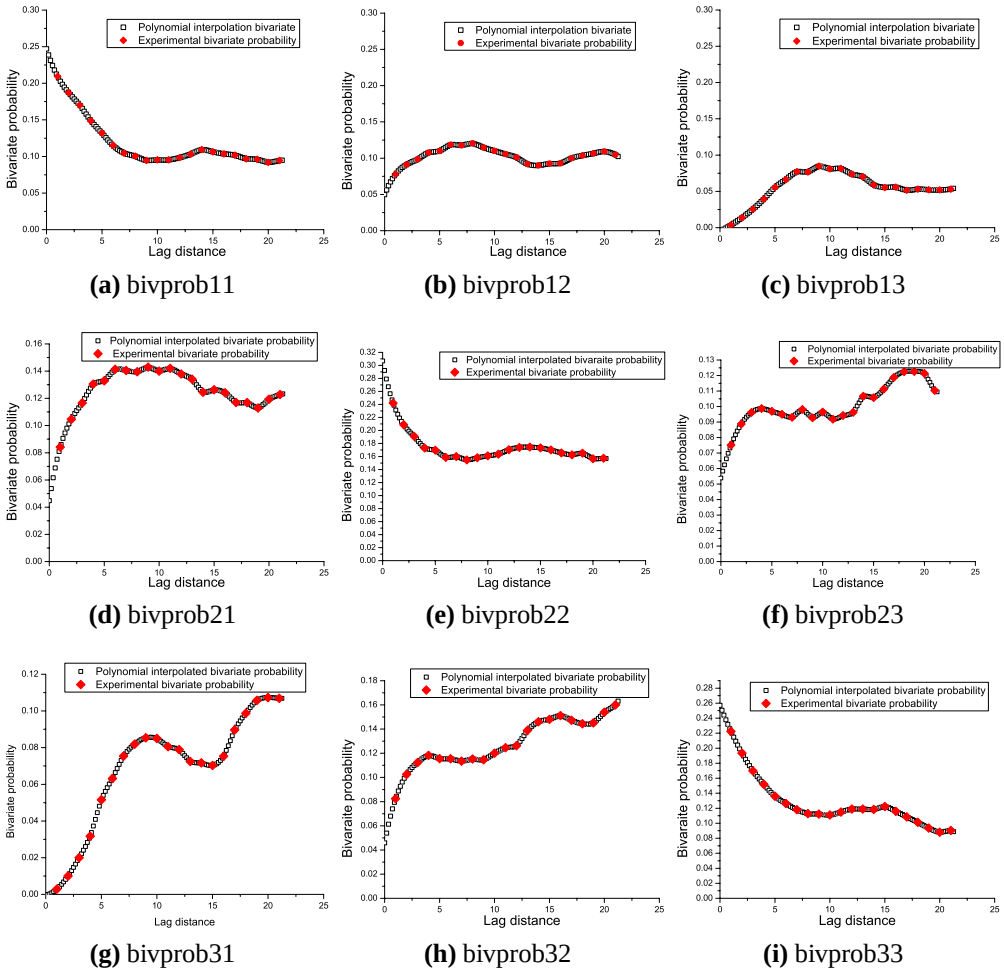
3.4 3D Spatial Heterogeneity Model

As discussed above, there is always a challenge to infer the horizontal spatial variability. In this section, the proposed methodology to infer the horizontal heterogeneity statistics is based on the sedimentary facies model and the theory of sequence stratigraphy. Geologists have a conceptual model of the subsurface. In any geostatistics algorithm, this conceptual model should be exploited and reproduced as much as possible. In the proposed method, this information is integrated through a heterogeneity prototype construction.

3.4.1 Walther's law

The earth's surface can be classified into different sedimentary realms or environments that are physically, chemically and biologically distinct from adjacent areas such as mountain ranges, sand deserts and deltas.

Each environment shows both abrupt and gradational lateral and vertical transitions. Sediments from those transitions will repeat vertically through a sedimentary sequence but may vary in character as a result of environmental and/or evolutionary changes through time. The relationship between depositional environments in space, and the resulting stratigraphic sequences developed through time was sum-



(a) bivprob11 **(b)** bivprob12 **(c)** bivprob13

(d) bivprob21 **(e)** bivprob22 **(f)** bivprob23

(g) bivprob31 **(h)** bivprob32 **(i)** bivprob33

Figure 3.7: One example of bivariate probability interpolation

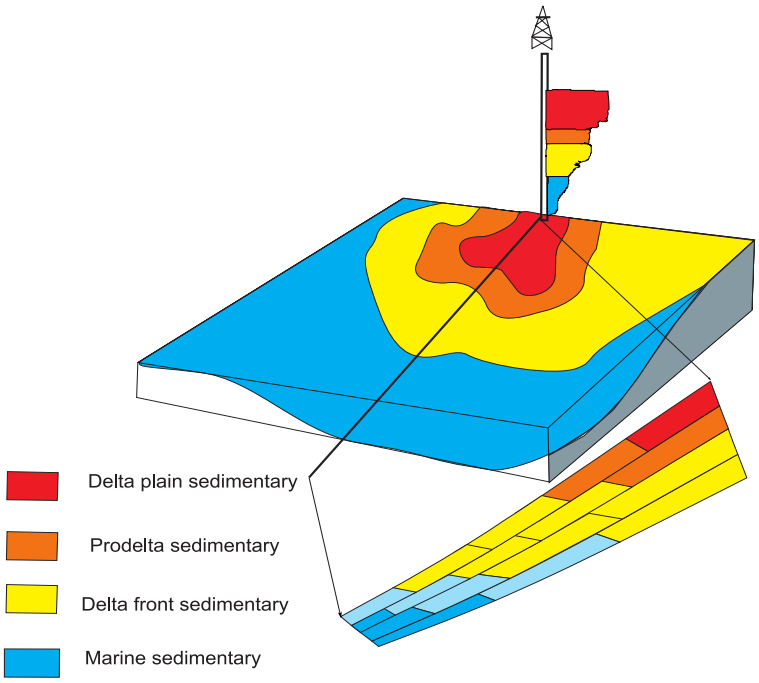


Figure 3.8: The facies stacking pattern in fluvial dominated delta sedimentary system

marized by Walther's Law which states that sedimentary environments that started out side-by-side will end up above one another over time due to transgressions and regressions (Middleton, 1973; Soreghan, 1997). In other words, a vertical sequence of facies is similar to the original lateral distribution of sedimentary environments.

This is a fundamental principle of stratigraphy which allows geologists to predict the facies lateral changes from the vertical changes observed in outcrops, core or well logs. For example, in a fluvial dominated delta building sedimentary process, the sediments will have an upward-coarsing pattern along the vertical direction which is from prodelta, delta front to delta plain. This is the same as the lateral chronostratigraphical boundary, from the proximal to distal direction, the same contacting pattern can be found both from the lateral and vertical direction as shown in Figure 3.8.

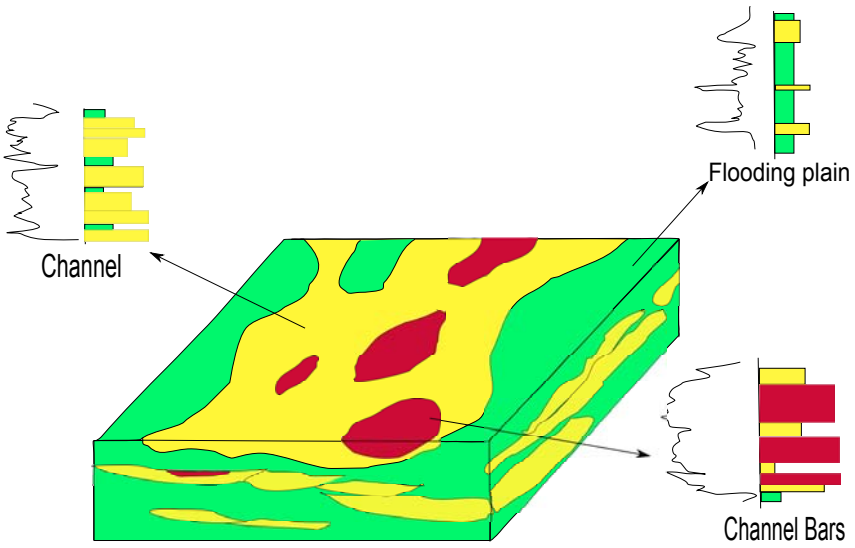


Figure 3.9: One example of braided river facies model shown in three-dimensional block diagrams

3.4.2 Sedimentary facies model

In general, between 2 to 7 facies types are defined and their spatial relations are described according to a conceptual geologic model. This is usually done by using local criteria or by reference to an existing universal facies model that characterizes the facies spatial assemblage (Plint, 1995).

Facies models will show a pattern of environments that prograde side ways to deposit a series of facies arranged in a predictable vertical sequence. Facies models summarize the essential aspects of facies sequences and relate them to the inferred depositional environments.

Facies models have been proposed for most major depositional environments, and there is a large measure of consensus about their general aspects, particularly for fluvial, aeolian, deltaic, and other shallow-water settings (Reading, 1996). They are usually most effectively conveyed graphically as three-dimensional block diagrams that relate the environment and its behaviour to the facies pattern as shown in Figure 3.9. The paleogeographic sketches, and vertical profile logs are also typical components of a published facies model as shown in Figure 3.10.

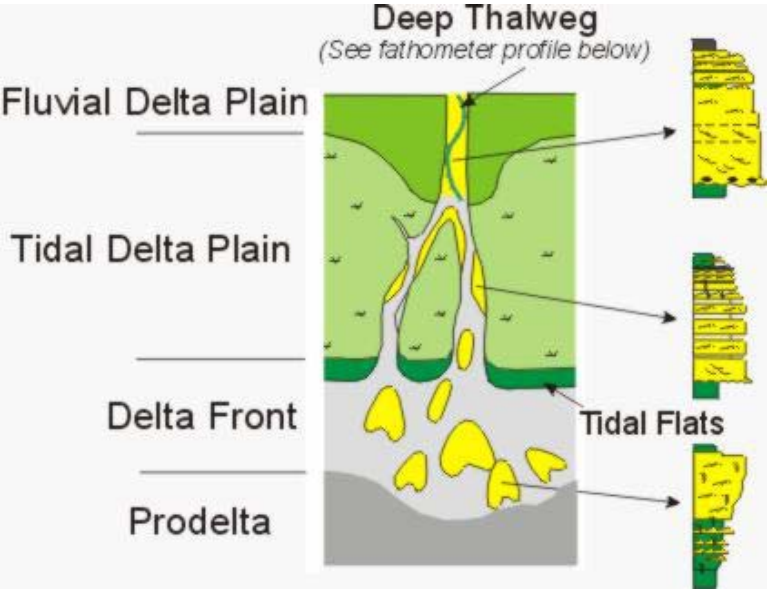


Figure 3.10: Example of facies model diagram in paleogeographic sketches and vertical profile logs (http://fosi.iagi.or.id/mahakam/mah-facies-des.htm)

The universal facies model presented in classical sedimentary textbooks provides valuable facies stacking pattern information for geostatistical facies modelling.

3.4.3 Sequence stratigraphy

The theory of sequence stratigraphy is a relatively recent paradigm in the field of sedimentary geology. It is the study of genetically related facies within a framework of chronostratigraphically significant surfaces. The sequence stratigraphic approach has led to improved understanding of how stratigraphic units, facies tracts, and depositional elements relate to each other in time and space within sedimentary basins (McLaughlin, 2005; Catuneanu, 2006). It was first utilized by the petroleum industry to interpret depositional surfaces on seismic sections (Payton, 1977; Wagoner and Tenney, 1991). Now sequence stratigraphy is used extensively by geologists to explain vertical and lateral changes in sediment rock distribution (Cross and Baker, 1992; Aitken and Flint, 1995; Catuneanu, 2002; Catuneanu et al., 2009).

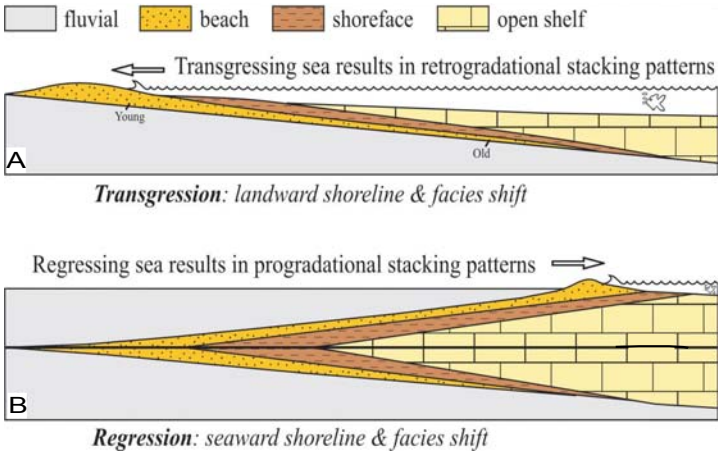


Figure 3.11: Transgression and regressions model in sequence stratigraphy theory (Modified from Catuneanu, 2006)

In sequence stratigraphy theory, the change in accommodation in conjunction with the rates of sedimentation represent a key control on depositional trends which is reflected by specific shoreline shifts. Two major depositional trends are usually recognized from the deposits: transgression and regression. A transgression is defined as the landward migration of the shoreline. This migration triggers a corresponding landward shift of facies. Transgression results in retragradational staking patterns. A regression is defined as the seaward migration of the shoreline. This migration triggers a corresponding seaward shift of facies. Regressions result in progradational staking patterns, for example, nonmarine facies shifting toward and overlying marine facies as shown in Figure 3.11.

3.4.4 Spatial heterogeneity prototype

In facies modelling, the facies will be identified from wells, cores and perhaps from the analogue outcrops. The geologist will select a specific facies model and provide the sequence stratigraphic analysis for the reservoir from correlating a group of wells. The facies model and sequence stratigraphy analysis results show the facies associations and how each of them will be interpreted in context with others.

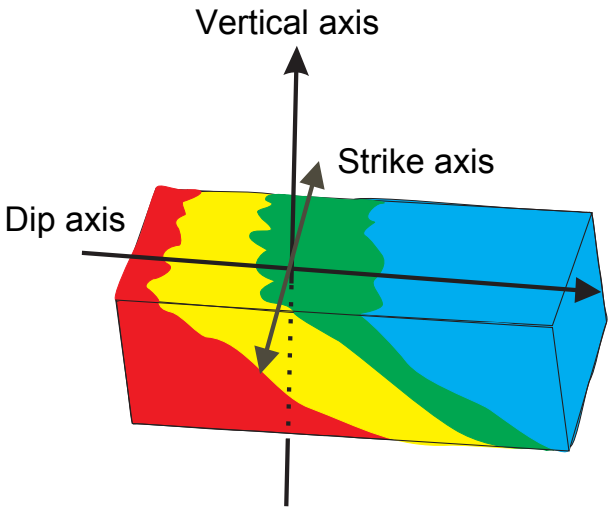


Figure 3.12: Sketch showing the heterogeneity prototype

Although each reservoir is unique, different reservoirs will share some basic characters. Based on the previous discussion, one conceptual heterogeneity prototype is proposed as plotted in Figure 3.12 to integrate the geological facies model and the sequence stratigraphy information in the geostatistical algorithms. In this prototype, three major heterogeneity axes are defined as: sedimentary dip, sedimentary strike and the vertical direction.

The vertical direction is the direction that the main sediments are observed and documented. The vertical profiles are the main form of data obtained from drilling. They are usually documented from logs. Their analysis is the starting point for Walther's law and sequence stratigraphy to interpret depositional processes and sedimentary environments.

The dip axis is associated with the direction of the major facies transitions horizontally. Usually, in channelized sedimentary environments, it will be parallel to the direction from sedimentary source to sedimentary deposition. For example, it will be from proximal to distal in a coastal sedimentary environment. Along the dip and vertical direction, the facies stacking pattern is expected to be similar. For a fluvial dominated delta building sedimentary process, the sedimentary deposits will have an upward-coarsing pattern along the vertical direction as well as a similar

pattern from the flood plain to the delta front direction as shown in Figure 3.8.

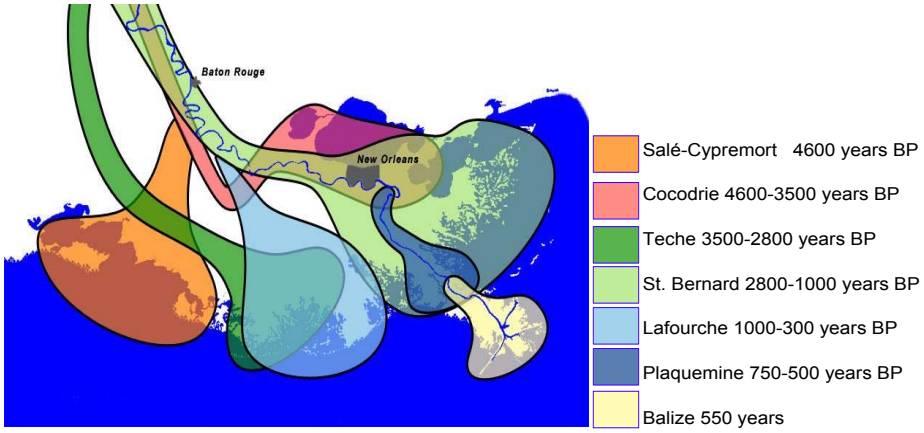
The strike axis will extend perpendicular to the the major sedimentary source direction. In most sedimentary environments, the shifting of sedimentary source is common. This sediments source switching phenomenon can be observed in the current Mississippi delta shown in Figure 3.13a. The Mississippi delta area is the modern area of land (the river delta) built up by alluvium deposited by the Mississippi River as it slows down and enters the Gulf of Mexico. The switching of the Mississippi River delta during the last 4,000 years is well documented (Roberts, 1997; Coleman et al., 1998). This kind of switching phenomenon is characterized by the strike direction in the conceptual model also shown in Figure 3.13b.

3.4.5 Properties of prototype

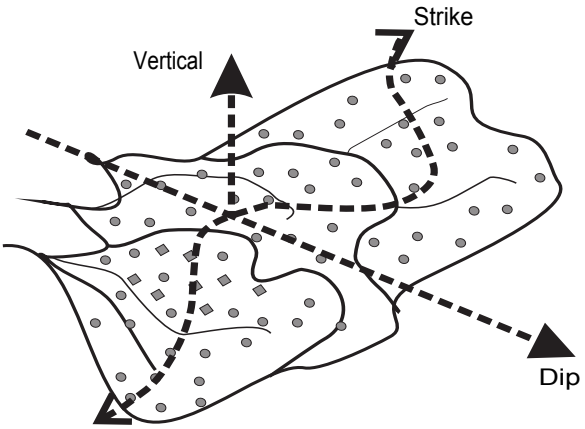
The dip and strike axis can be predicted from the data if the pathway of deposition is directional such as in a delta environment, or in gravity sedimentary environments (Ghibaudo, 1992; Drake and Calantoni, 2001). The dip direction will be toward the source direction and the strike will be perpendicular to the dip direction. The vertical axis will be the main sedimentary stacking direction.

The sedimentary trends or stacking patterns are asymmetric along different directions. Along the vertical axis, the fining upward facies transition pattern would become a coarsening upward pattern in the reverse direction. The dip axis will have the same asymmetric property. Thus, it is necessary to define a positive direction along those two axes for computational implementation.

The direction from the earlier deposits to the later deposits will define a positive direction and usually it will be from bottom to top. For the dip axis, the positive direction will be from the depositional source to the margin, that is, from proximal to distal. The positive direction definition along dip and in the vertical axis should have the same staking trend, that is, their positive direction should be consistent with each other. For example, in a delta building depositional process, the positive axes will be different from delta sediments in a transgressing process as shown in



(a) Switching of Mississippi delta (Coleman et al., 1998)



(b) Delta switching in prototype representation

Figure 3.13: The lateral switching phenomenon in Mississippi delta sedimentary environment and its prototype representation

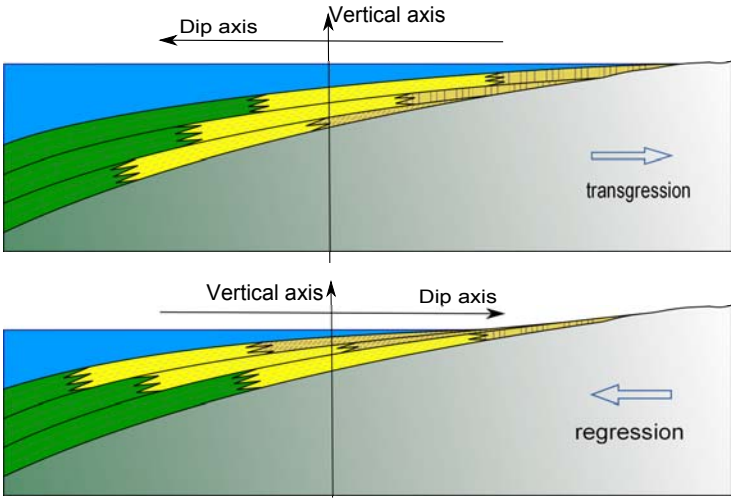


Figure 3.14: The different positive definition for dip direction in case of transgression (top) and regression of the sea (bottom)

Figure 3.14.

The strike direction will have no positive or negative direction. The sediments along this axis will be characterized by a stochastic function defined below.

One more characteristic of these three axes is the anisotropy ratio along the strike and dip axes. With high-resolution chronostratigraphic correlation analysis, the sedimentologists could estimate how much of the heterogeneity stacking pattern from the vertical well profile will be valid along the lateral direction, especially along the sedimentary dip direction. In this research, this geological understanding is quantified by heterogeneity ratio along dip axis. This ratio is denoted as a_x which is equal to:

$$a_x = \frac{H_x}{H_z} \tag{3.15}$$

where the H_x and H_z would be the full pattern thickness along the dip and vertical direction.

While along the strike direction, the relative shifting area will also estimated from the sequence stratigraphy and facies analysis. Along the strike direction, the anisotropy ratio a_y will control the prototype model. It will calculated as:

$$a_y = \frac{H_y}{H_z} \tag{3.16}$$

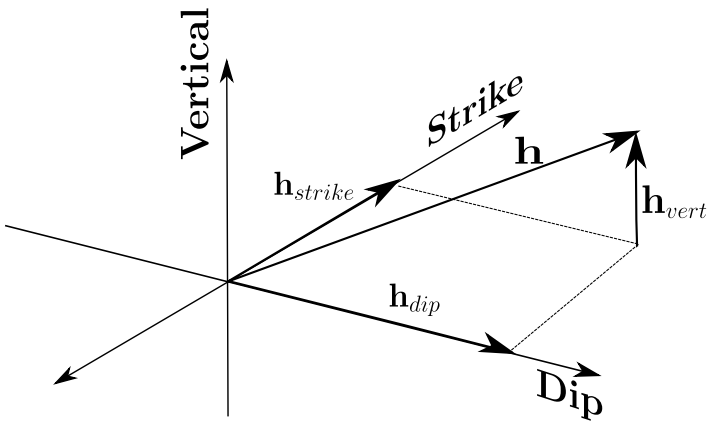


Figure 3.15: Decomposition along three major anisotropy axes

where H_y is the distance that the geologists believe the switching will happen for the particular reservoir.

3.5 3D Spatial Distance Transformation

After setting up the anisotropy axes in the reservoir, during the modelling, any arbitrary spatial distance vector **h** will be decomposed along those three axes as shown in Figure 3.15. Three distance vectors components $\mathbf{h}_{vert}$, $\mathbf{h}_{dip}$ and $\mathbf{h}_{strike}$ will describe the spatial distance between two locations $(\mathbf{u}, \mathbf{u}+\mathbf{h})$.

The orientation of the anisotropy is controlled by the geological features of the reservoir and the orientation of the coordinate system. The anisotropy axes may not coincide with the axes of the model coordinate system. In this case, the components $\mathbf{h}_x, \mathbf{h}_y, \mathbf{h}_z$ of the distance vector **h** in the data coordinate system will have different values when referenced in the coordinate system of the anisotropy axes. Thus, it is necessary to transform the vector from the data coordinate system to the coordinate system of the anisotropy axes. In the following discussions it is assuming that the *x*, *y* and *z* direction of the model coincide with the dip, strike and vertical axes of the prototype model. The decomposed distance vector along the three axes $\mathbf{h}_{vert}, \mathbf{h}_{dip}, \mathbf{h}_{strike}$ will correspond to the $\mathbf{h}_z, \mathbf{h}_x, \mathbf{h}_y$ distance vectors.

Any spatial distance vector **h** will be transformed to an effective distance vector

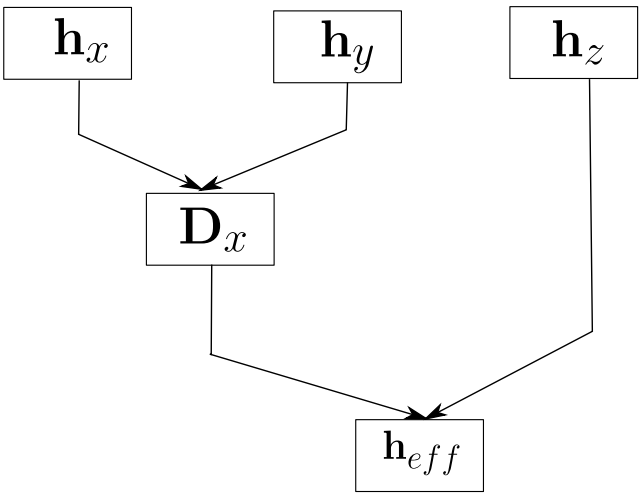


Figure 3.16: Two steps in three dimensional distance combination

$\mathbf{h}_{eff}$ along the vertical direction. It is done in two steps shown in Figure 3.16. In the first step, the dip distance vector component $\mathbf{h}_x$ is modified to a vector D_x reflecting the sedimentary shifting along the strike direction controlled by strike component vector $\mathbf{h}_y$. Then, the modified distance vector D_x from the first step is combined with the vertical component vector $\mathbf{h}_z$ to obtain a final effective vertical vector $\mathbf{h}_{eff}$ based on the sedimentary environment.

3.5.1 Integration of dip and strike direction

The anisotropy along the sedimentary dip and strike direction will be combined together first. The basic idea is that the shifting process along the strike direction follows some random process. The lag distance $\mathbf{h}_x$ along the dip direction will be modified by a factor from the strike directions which will mimic sediment source shifting. It will be the main effect on the final effective distance calculation D_x from the strike direction vector $\mathbf{h}_y$.

As the sediments source shifting along the strike direction, the distance increase along strike direction will not always increase the probability of transition to another facies as it does along dip direction. This geological variability along the strike axis direction is quantified by F_y. For example, it could be modelled as sine

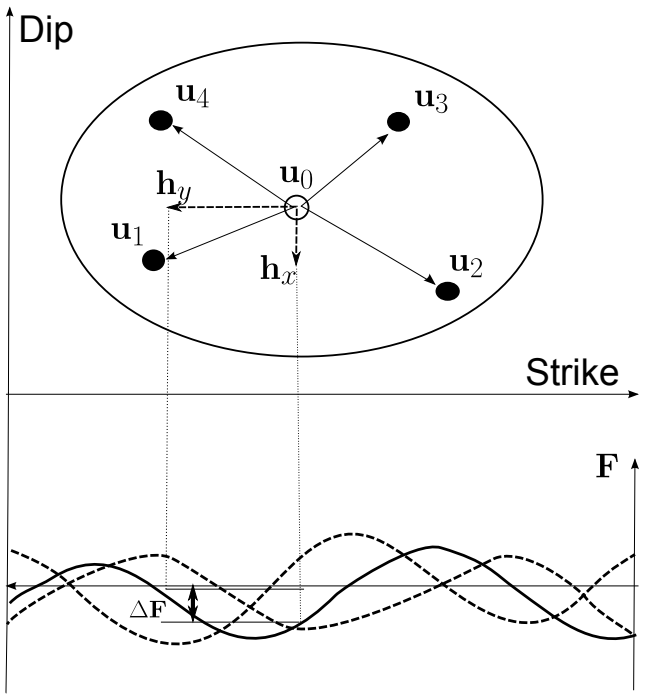


Figure 3.17: Sketch showing the random factor from the switching function along the strike direction

wave function $F_y = f(\mathbf{h}_y) = A \cdot sin(B \cdot \mathbf{h}_y + C)$. For the sine function, the value from valley to peak will be the maximum deviation to the dip distance caused by strike. The period of this sine wave would depends on the research domain

$$\frac{2\pi}{B} = H_y = nper \cdot ny \cdot ysiz$$

where H_y is the full switching distance; and the parameter *nper* will control how many full sine periods will be found along the strike direction. The larger *nper* , the higher frequency of randomness modifications to the distance along the dip direction. In the simulation, each realization would use a different random parameter $\frac{C}{B} = rand(\cdot)$ for the sine wave as shown in Figure 3.17.

The variety caused by the distance vector along strike direction $\mathbf{h}_y$ is denoted as ΔF_y and will cause a modification to the distance vector $\mathbf{h}_x$ along the dip distance. The effective distance along dip direction D_x is combined from the factor ΔF_y and

the actual distance along dip direction h_x as:

$$D_x = h_x - \Delta F_y \tag{3.17}$$

There are different cases for the combination of strike and sedimentary dip together. Some possible cases are shown in Figure 3.18.

Case A : $\mathbf{h}_x = \Delta F_y$. Although there is a dip distance difference ($\mathbf{h}_x \neq 0$), the effect from the strike will cancel that dip difference. Thus, for those two locations, there will be no increment along the dip direction.

Case B : $\Delta F_y < 0$ the increment along the dip direction will be increased from the sedimentary effect according Equation (3.17). Thus, the combined distance will be larger than the original increment $\mathbf{h}_x$.

Case C : $\Delta F_y = 0$ the increment along the dip direction will be kept as the same. The variability will only depend on the dip vector $\mathbf{h}_x$.

Case D : $\Delta F_y > 0$, the increment along the dip direction will be decreased considering the effect from the strike direction;

Case E : $\mathbf{h}_x = 0$ and $\Delta F_y < 0$, although there is no dip distance difference. But from the strike effect, there would be an increment along the dip direction as $D_x > 0$ from Equation (3.17).

By this approach, the distance along strike and dip will be combined together. As shown in the combination example listed above, the geological meaning is integrated into the " effective" sign-dependent distance calculation.

A small example that can be calculated manually will be used to show the combination of dip from strike direction. There are four locations in Cartesian coordinates shown in Figure 3.19a. The distance vectors between all the locations will follow the Euclidean geometry property. In Figure 3.19b, the coordinate system is the geological based as along strike direction and there is a sine wave of

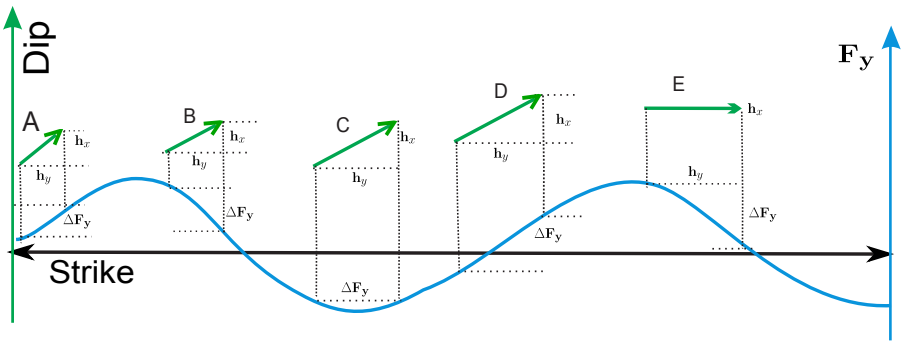


Figure 3.18: Illustration of random modification of the dip vector from the strike vector based on the random switching function

Table 3.2: Distance matrix from coming dip and strike direction

	o	a	b	c
o	0	5.4797	14.4046	-0.5441
a	-5.4797	0	8.9249	-6.0238
b	-14.4046	-8.9249	0	-14.9487
c	0.5441	6.0238	14.9487	0

$(2 \cdot sin(\pi/10 \cdot h_y))$ attached to. The X component of each distance vector is modified by the factor which is calculated from the random function along the strike direction using the Y component of the distance vector as a variable. As can be seen from this example, after accounting the fluctuation caused by Y components, the X component along the dip direction is embedded into a new space. Although this transformation is not Euclidean, the transformed distances along the dip direction satisfy the triangle qualities.

After taking the positive direction into consideration, the distance matrix between the four locations is shown in Table (3.2). In this matrix, the negative sign will affect the facies stacking pattern. The row index is the head of a distance vector. The column index will be the tail of a distance vector.

3.5.2 Integration of dip and vertical direction

Combining the $\mathbf{h}_z$ and D_x into an effective distance h_{eff} based on the anisotropy ratio a_x along the dip direction is the next step for a final effective distance calculation.

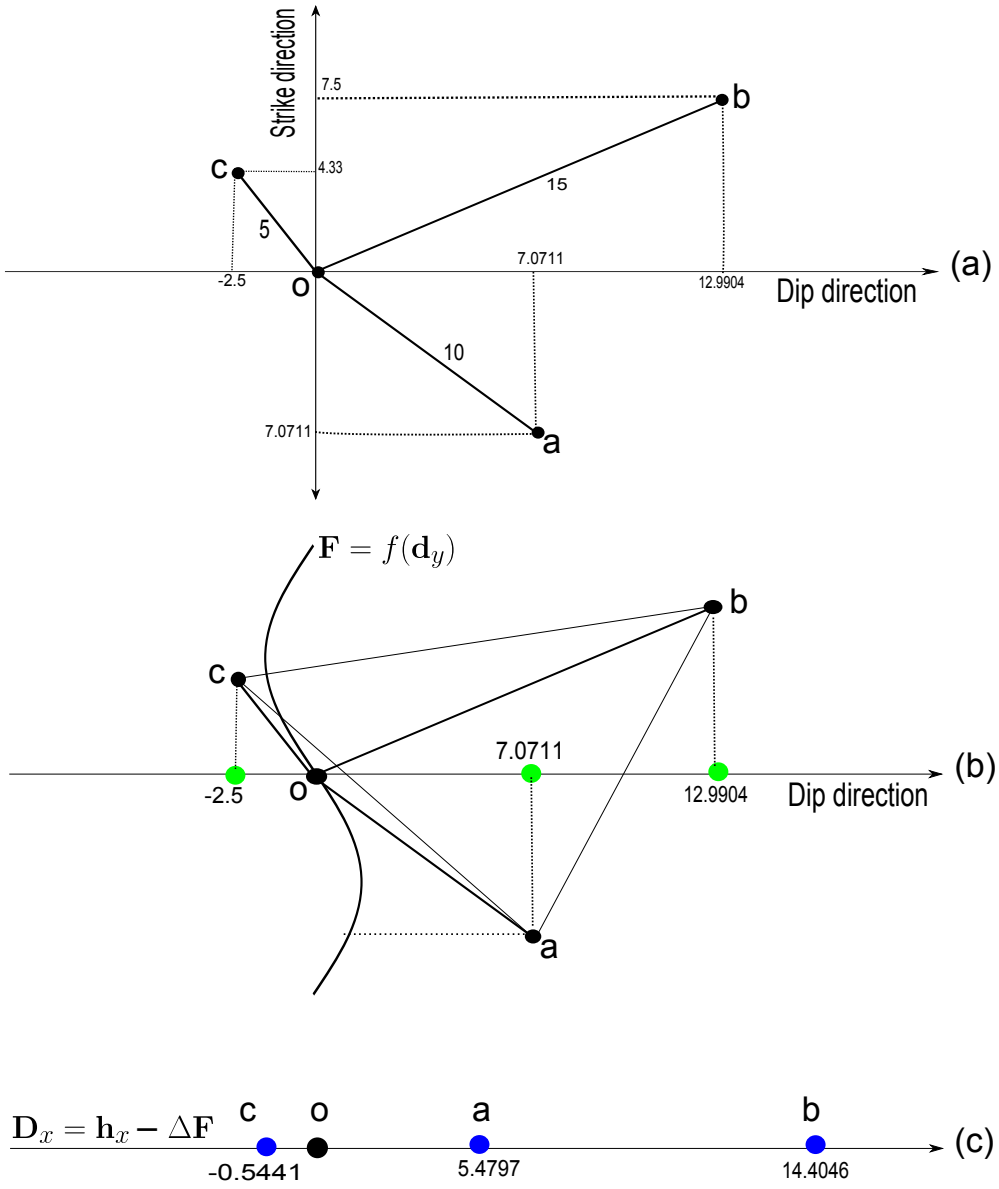


Figure 3.19: One example of the combination between dip and strike direction. In (a), all the distances are in their original space and will be decomposed along dip, and strike direction; in (b), the green locations are the relative dip locations for those four points; After modification by the factor from strike vector, the relative dip locations are shown in (c)

The final combined distance vector is calculated as:

$$h_{eff} = \mathbf{h}_z + \frac{D_x}{a_x} \tag{3.18}$$

The positive direction will be different for a regression and transgresion stacking pattern. There will be four combination results for an arbitrary distance vector as shown in Figure 3.20.

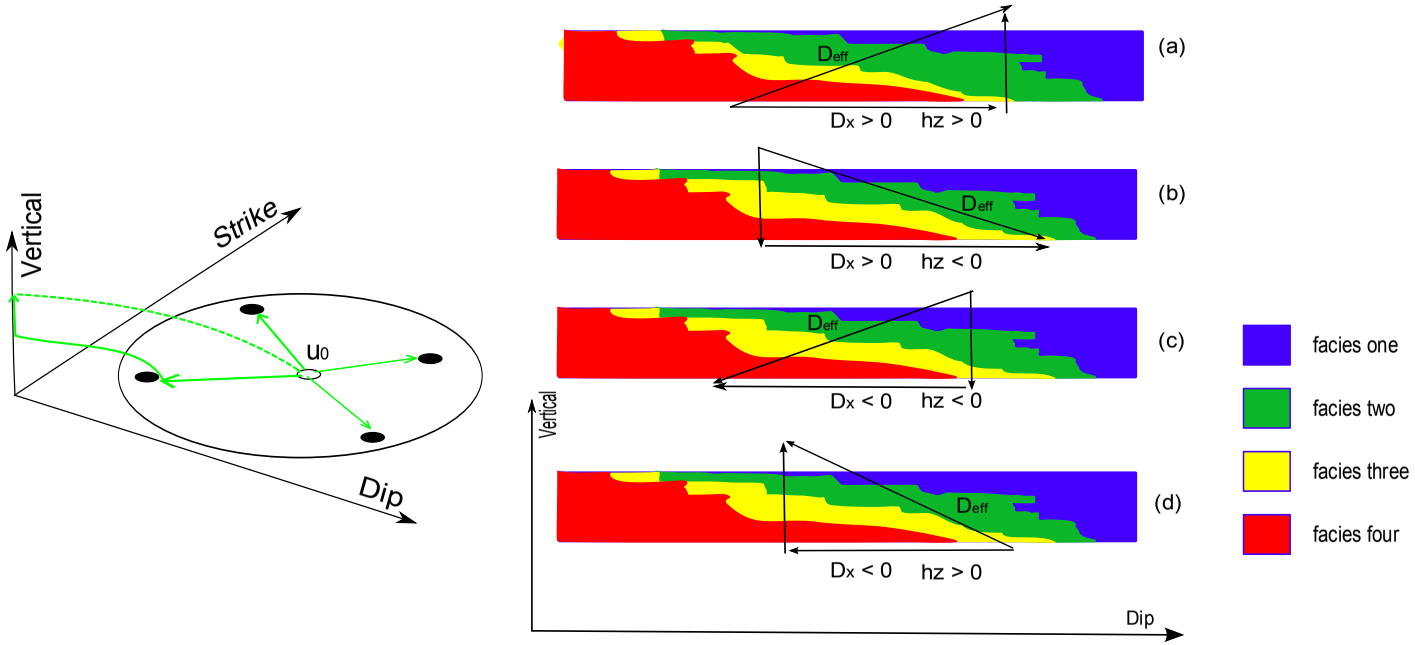


Figure 3.20: Distance combination from vertical direction and dip direction

Case (a) in Figure 3.20 is the situation where $\mathbf{h}_z > 0$ and $D_x > 0$. The vector of $\mathbf{h}_z > 0$ means that along the vertical direction, the facies will have a transition tendency from four to one. While the vector of $D_x > 0$ means that along the dip direction, the facies will have a same trend as the vertical trend. Thus, the effective vector h_{eff} combined from Equation (3.18) will also keep the transition trend that the facies will have a high tendency from four to one. The distance interval $|h_{eff}|$ will be larger than any single one $|\mathbf{h}_z|$ or $|\frac{D_x}{a_x}|$.

In Case (b), the vector of D_x is still positive, while the vector $\mathbf{h}_z$ is negative which means that along the vertical direction and the lateral direction the facies transition trends are different. The direction of the final combination result vector h_{eff} from those two vectors will depend on the stronger one. If $|\frac{D_x}{a_x}| > |\mathbf{h}_z|$, then the direction of vector h_{eff} will be positive. Otherwise, it will be negative. In both situations, the distance interval $|h_{eff}|$ will be shorter than $|\mathbf{h}_z|$ or $|\frac{D_x}{a_x}|$.

In case (c), both the transition trend from vertical and dip direction show that the facies has a higher probability from four to one. They are consistent. From the combination, the value $|\frac{D_x}{a_x}|$ is added to the vertical distance interval $|\mathbf{h}_z|$. The direction vector will have a negative value to indicate the geological meaning.

In case (d), the trend from vertical and lateral contradict each other again. In this situation, if the $|\frac{D_x}{a_x}| > |\mathbf{h}_z|$, the direction of vector h_{eff} will follow the direction of vector D_x. Otherwise, if $|\frac{D_x}{a_x}| < |\mathbf{h}_z|$, the direction of vector D_x will follow the vector $\mathbf{h}_z$, which is positive. The distance from the combination will be shorter than each one.

A small manually calculated example is given in Figure 3.21. This time, the data configuration is the same as the example given in Figure 3.19, but the coordinate axes changed to dip and vertical axes. To the left, all the distance vectors are decomposed along the two axes. To the right, all the distance vectors are combined from vertical component h_z and effective dip vector component D_x using Equation (3.18). Note that the geometrical relations are still valid in this new coordinate system.

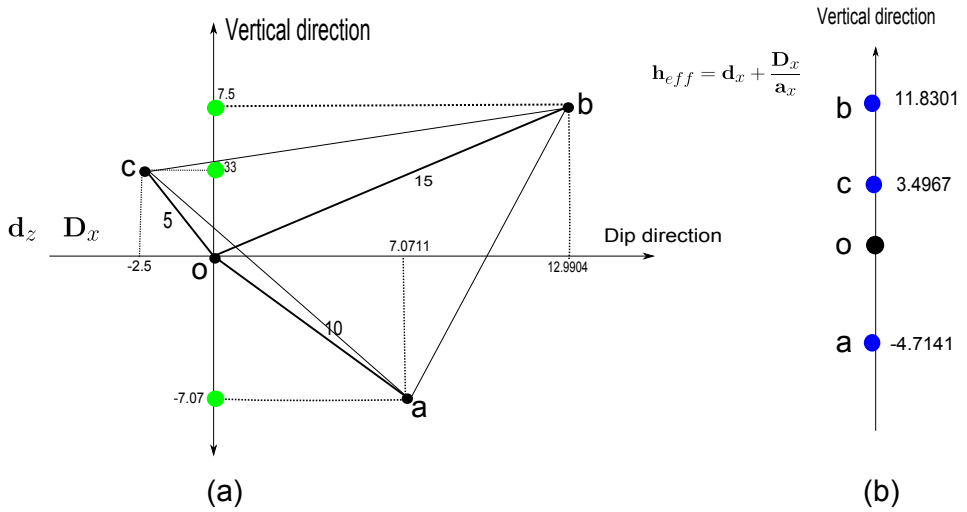


Figure 3.21: One example of vertical and dip combination results. The green relative locations for those four point in (a) indicate their relationship before combination; in (b) is the relative location along the vertical direction. The distance interval between them will be the final effective distance.

3.6 Remarks

How the facies contact each other and what kind of anisotropy variability is expected in the model along different directions are expressed through the facies geological model. Such quantitative information should be integrated in the final geostatistical facies model.

The bivariate probability diagram proposed in this research can integrate the facies stacking patterns information from the facies geological model which provides a way to reproduce them from the final geostatistical model.

The 3D spatial distance transformation scheme proposed in this chapter will integrate geological anisotropy understandings on how the heterogeneity extends spatially. Those will make the final geostatistical models closer to our geological understanding.

Chapter 4

Discrete Multivariate Probability Estimation

In this chapter, 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.

4.1 Entropy of Probability Distributions

Predicting facies outcomes for spatial locations will always come with uncertainty because of lack of perfect knowledge of the study area. The well known probability theory and the related statistics are the most traditional tools to handle uncertainties (Bardossy and Fodor, 2001). For example, the uncertainty of the outcomes for a group of locations is modelled with a discrete probability distribution $P(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ in this research. The probability distribution allows one to assign a numerical assessment to each of the possible events.

4.1.1 Entropy

The term entropy was first defined in thermodynamics as a measure of the change in randomness or disorder in a closed chemical system such as the result of a reac-

tion (Gibbs, 1873; Tribus, 1961). It was originally devised by Claude Shannon in 1948 to study the amount of information in a transmitted message and is expressed in terms of a discrete set of probabilities (Shannon, 1948). Since then, entropy has been widely accepted as a criteria to measure the uncertainty with a probability distribution (Gokhale, 1973; Kullback, 1968; Cover and Thomas, 2006).

Quantitatively, the measured entropy carried by the probability distribution p : $\{p_\ell, \ell = 1, \cdots, N\}$ is defined as:

$$H(p) = -\sum_{\ell=1}^{N} p_\ell \log p_\ell \tag{4.1}$$

Note: in the case of $p_\ell = 0$, it is defined that $H(p_\ell) = 0$. This measurement will indicate the information/uncertainty that goes with this probability distribution.

4.1.2 Measurement of uncertainty

Although the concept of entropy was first proposed for uncertainty evaluation of a discrete probability distribution, it is also used for continuous variables:

$$H(X) = -\int_a^b \log(f(x))\, f(x) dx \tag{4.2}$$

Where random variable X with a range (a,b) and a probability density function $f(x)$ (Cover and Thomas, 2006).

It can also be shown that the entropy is related to the variance for a normal distribution. For a random variable X with a normal distribution $N(\mu, \sigma^2)$, the entropy will be:

$$\begin{aligned} H(X) &= -\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} exp(-\frac{1}{2}[\frac{x-\mu}{\sigma}]^2) \log\left\{ \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{1}{2}[\frac{x-\mu}{\sigma}]^2) \right\} dx \quad (4.3) \\ &= \log(\sigma\sqrt{2\pi}) + \frac{\log e}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-y^2} y^2 dy \\ &= \log(\sigma\sqrt{2\pi}) + \frac{\log(e)}{\sqrt{\pi}} \frac{1}{2} \sqrt{\pi} \\ &= \log(\sigma\sqrt{2e\pi}) \end{aligned}$$

where the substitution $y = \frac{x-\mu}{\sqrt{2}\sigma}$ is used. It is interesting to observe that the entropy for a normal distribution is a function of σ but not of μ. This is showing that the entropy and variance both are a measure of uncertainty for continuous variable.

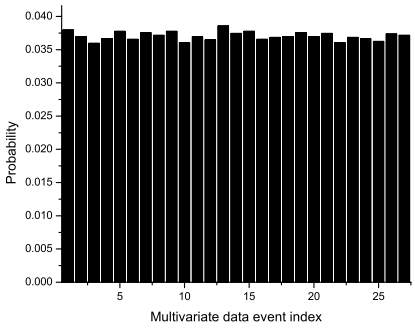
A probability distribution with higher entropy will correspond to less information (more uncertainty or more of a lack of information). After obtaining some informations from the sampled locations, the uncertainty at an unsampled location will likely decrease and it will be characterized by the estimated probability distribution for this location.

For a discrete random variable $Z(\mathbf{u})$, the entropy reaches its maximum value when discrete probability distribution is uniformly distributed as $1/K$. The entropy of this uniform probability distribution would be calculated as:

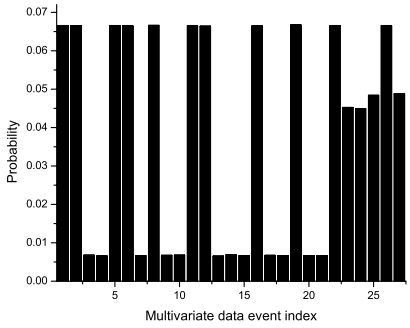
$$H(P) = -\sum \frac{1}{K} \log(1/K) = \frac{\log(K)}{K} \tag{4.4}$$

The entropy will reach the minimum value when there is no uncertainty about the outcomes, that is, when the probability distribution is $P(\mathbf{u}) = 1$ or $P(\mathbf{u}) = 0$. The uncertainty will be zero. Knowing $P(\mathbf{u}) = 1$ of one facies means that we are sure that this facies will be found.

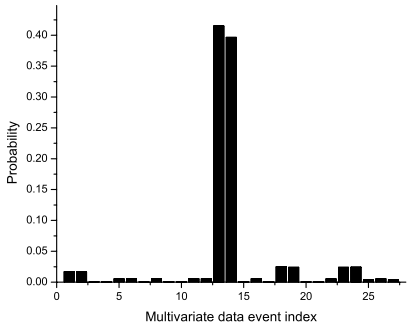
In an example of three location with three categories, three possible multivariate probability distribution would be assigned to this situation as shown in Figure 4.1. The entropy calculated from them are 1.4313, 1.2855 and 0.6699 respectively. Case (a) is the totally random distribution. The probability distribution in (a) has no information. The uncertainty is the highest. While for case (b), the outcomes probability for some data events are larger than others. In probability distribution (c), data events 13 and 14 have a higher probability. There would be less uncertainty about the outcomes based on this probability distribution. It has smallest entropy value of these three probability distributions.



(a) Uniform distribution case



(b) The probability after getting more information



(c) The probability distribution with less uncertainty

Figure 4.1: One example of three possible probability distributions for a multivariate data events space

4.2 Maximum Entropy Principle

When there is no reason to do otherwise, all outcomes will have equal probability. As in the example in Figure 4.1, the probability distribution of case (a) would be the choice when no information is given. This is called the principle of insufficient reason, or principle of indifference (Jaynes, 1957). It corresponds to a decision to use a uniform probability distribution. The challenge is to select a probability distribution to use if some constraints are known from data. The maximum entropy principle is a theory that defines how to construct an appropriate probability distribution.

4.2.1 Constraints in probability estimation

There could be many different constraints on the probability distribution. Denoting the probability distribution as $\mathbf{p} : \{p_\ell, \ell = 1, \cdots, N\}$, the first constraints would be the closure constraint:

$$\sum_{\ell=1}^{N} p_\ell = 1 \tag{4.5}$$

There are lower order marginal constraints on the desired multivariate probability. As shown in Chapter 2, all the lower order marginal probability distribution such as univariate and bivariate marginal probability distributions are linear combinations of the multivariate probability distributions. They should be reproduced in the final estimated multivariate probability distribution.

The different order of marginal probabilities can be expressed as linear operations of the multivariate probability distribution:

$$\sum_{\ell=1}^{N} a_{m\ell} p_\ell = b_m; m = 1, 2, \cdots, M \tag{4.6}$$

where $a_{m\ell}$ is called the marginal construction matrix, b_m are the marginal probabilities.

If all the lower order marginal constraints are fully integrated there will be a large number of constraints. For example, if the multivariate function is composed of 20 binary random variables, then $2^{20-1} = 524,288$ joint lower order of probability constraints are needed to fully specify all the multivariate probabilities.

In practice, many of these constraints are difficult to obtain. In spatial statistics, only the univariate and bivariate constraints are easily and reliably inferred. Thus, without using the full set of lower order marginal probability as constraints for the desired multivariate probability estimation, the inference will be under-constrained. When a system is under-constrained, the desired probability distribution usually has a range of possible outcomes. The estimation results from the maximum entropy principle principle introduced next will be a "best" in the sense of entropy (Cheeseman, 1983).

4.2.2 Maximum entropy principle

For probability estimation under constraints, the most extensively used technique is the Maximum Entropy (ME) principle which is stated as:

> *It is based on the premise that when estimating the probability distribution, the best estimation results will keep the largest remaining uncertainty (the maximum entropy) consistent with all the known constraints. In that way, no more additional assumptions or biases are introduced in the estimation* (Jaynes, 1978).

Formally, assuming any desired multivariate probability is $\mathbf{p} : \{p_\ell, \ell = 1, \cdots, N\}$, the objective function will be:

$$\texttt{Maximum:} \qquad H(\mathbf{p}) = -\sum_{\ell=1}^{N} p_\ell Log(p_\ell) \tag{4.7}$$

and the constraints will be:

$$\sum_{\ell=1}^{N} p_\ell = 1 \tag{4.8}$$

$$\sum_{\ell=1}^{N} a_{m\ell} p_\ell = b_m; m = 1, 2, \cdots, M \tag{4.9}$$

Where the first constraint in Equation (4.8) will make sure the probability distribution will sum to one to satisfy the normal or closure constraint. The constraints

$b_m; m = 1, 2, \cdots, M$ in Equation (4.9) are of any order of marginal for the desired discrete multivariate probability.

The maximization of $H(\mathbf{p})$ subject to the above constraints results in a optimization problem. The well-known solution to the problem of optimizing a function subject to constraints is the method of Lagrange multipliers (Bertsekas, 1999). The Lagrange multipliers procedure can be illustrated in the situation that only the normal constraint is enforced. The first step is to form a new objective function $L(p_\ell, \lambda)$ as:

$$L(p_\ell, \lambda) = -\sum_\ell p_\ell \log(p_\ell) + \lambda\big(\sum_\ell p_\ell - 1\big) \tag{4.10}$$

The second step is equating the derivative of (4.10) to zero with respect to each of the variables $p_\ell, \ell = 1, \cdots, N$ and λ. This results in an equation set:

$$\frac{\partial L}{\partial p_\ell} = -1 - \log(p_\ell) + \lambda = 0 \tag{4.11}$$

$$\frac{\partial L}{\partial \lambda} = \sum_\ell (p_\ell) - 1 = 0 \tag{4.12}$$

From Equation (4.11), $p_\ell = exp(\lambda - 1)$. This is independent of ℓ. Thus, all the probabilities p_ℓ should be equal and sum to 1. Then, the uniform distribution $p_\ell = 1/N$ is the ME estimation.

The Lagrange multiplier principle can be extended to any order of marginal probabilities. In the following illustration, the full constraints from Equation (4.8) and (4.9) are enforced in the objective function. The new objective function $L(p, \lambda, \lambda_m)$ will be defined as:

$$L(p_\ell, \lambda, \lambda_m) = -\sum_\ell p_\ell \log(p_\ell) + \lambda\big(1 - \sum_\ell p_\ell\big) + \lambda_m\big(b_m - \sum_\ell a_{m\ell} p_\ell\big) \tag{4.13}$$

Then, the derivatives of Equation (4.13) are set to zero with respect to each of the variables $p_\ell, \ell = 1, \cdots, N$ and $\lambda, \lambda_m; m = 1, \cdots, M$, that is:

$$\frac{\partial L}{\partial p_\ell} = -\log(p_\ell) - 1 - \lambda - \sum_{m=1}^{M} \lambda_m a_{m\ell} = 0 \tag{4.14}$$

$$\frac{\partial L}{\partial \lambda} = 1 - \sum_{\ell} p_\ell = 0 \tag{4.15}$$

$$\frac{\partial L}{\partial \lambda_m} = b_m - \sum_{\ell} a_{m\ell} p_\ell = 0 \tag{4.16}$$

From Equation (4.14):

$$p_\ell = exp\Big(-\lambda_0 - \sum_{m=1}^{M} \lambda_m a_{m\ell} \Big) \quad (\lambda_0 = \lambda + 1) \tag{4.17}$$

Replacing Equation (4.17) back into Equation (4.15) and (4.16), results in a set of simultaneous equations that can be written as:

$$1 - \sum_{\ell} exp(-\lambda_0 - \sum_{m=1}^{M} \lambda_m a_{m\ell}) = 0 \tag{4.18}$$

$$b_m - \sum_{\ell} a_{m\ell} \Big[exp(-\lambda_0 - \sum_{m=1}^{M} \lambda_m a_{m\ell}) \Big] = 0 \tag{4.19}$$

In principle, it is shown that solving the Equation set (4.18) and (4.19) to get the $\lambda_m; m = 0, 1, \cdots, M$ and substituting back into Equation (4.17), will get the maximum entropy solution from the constraints. Practically, the solution of the $m+1$ set of coupled implicit nonlinear equations cannot be found analytically. Even finding the numerical solution is not an easy task in practice (Newman, 1979; Mead and Papanicolaou, 1984; Ulrych et al., 1990).

4.3 Minimum KL Divergence Principle

The entropy of a probability distribution is a measure of the uncertainty of the random variable. Based on the entropy concept, the Kullback-Leibler divergence (KL divergence or KL distance) is used as a measurement of the distance between two probability distribution in probability theory and information theory.

4.3.1 KL divergence

The KL divergence is an asymmetric measure of the difference between two probability distributions (Kullback and Khairat, 1966; Ireland and Kullback, 1968). For ease of presentation, consider two discrete probability distribution $p : (p_1, \cdots, p_N)$ and $q = (q_1, \cdots, q_N)$. The KL divergence between them is a measure of the difference of the information contained in them and is defined as:

$$J[p \parallel q] = \sum_{\ell=1}^{N} p_\ell \log \frac{p_\ell}{q_\ell} \tag{4.20}$$

One property of KL divergence is that the asymmetric property as $J[p \parallel q] \neq J[q \parallel p]$. Thus, it is not a true distance metric. Some other properties are:

$$J[p \parallel q] \geq 0 \tag{4.21}$$

$$J[p \parallel q] = 0 \quad \text{if and only if} \quad p = q \tag{4.22}$$

4.3.2 Minimum KL divergence principle

Based on the concept of KL divergence, Kullback proposed the principle of Minimum KL divergence: given new facts, a new distribution p should be chosen that is as hard to discriminate from the original distribution q as possible; so that the new data produces as small an information gain $J(p||q)$ as possible, thus no more bias except satisfying the constraints are introduced (Kullback, 1968). It is written as:

$$\text{Minimize:} \quad J(p \parallel q) = \sum_{\ell=1}^{N} p_\ell \log \frac{p_\ell}{q_\ell} \tag{4.23}$$

subject to:

$$\sum_{\ell=1}^{N} p_\ell = 1 \tag{4.24}$$

$$\sum_{\ell=1}^{N} a_{m\ell} p_\ell = b_m; m = 1, 2, \cdots, M \tag{4.25}$$

The minimum KL divergence principle can be seen as an extension of the principle of ME. As given $q = 1/N$, the minimum KL divergence results will be exactly

the ME results which is:

$$\begin{aligned} J(p \parallel q) &= \sum_{\ell=1}^{N} p_\ell \log(p_\ell N) \\ &= \sum_{\ell=1}^{n} p_\ell \log p_\ell + N \log N \\ &= -H(p) + (\text{constant}) \end{aligned} \tag{4.26}$$

As shown in Equation (4.26), minimizing $J(p \parallel q)$ will be the same as maximizing entropy $H(p)$. Therefore, both $J(p \parallel q)$ and $H(p)$ are measures of the deviation of p from a discrete uniform distribution. From this view, it is seen that the ME solution is a special case of the minimum KL divergence. Moreover, the minimum KL divergence formulation is more general and offers greater flexibility, because with the minimum KL divergence formulation the null-hypothesis function q can represent any probability function (Ireland and Kullback, 1968).

4.3.3 Iterative scaling

From section 4.2.2, it was shown that the ME solution will be a coupled nonlinear equation set. It could be solved but it will always be a challenge. In this section, one multivariate probability estimation approach based on the minimum KL divergence principle is introduced based on some previous works (Good, 1963; Kullback and Khairat, 1966; Darroch and Ratcliff, 1972).

It is said that given the constraints in the form of Equation (4.24) and (4.25), there exists an unique probability distribution $\hat{p} : \{\hat{p}_1, \cdots, \hat{p}_N\}$ that satisfies them and is of product form as:

$$\hat{p}_\ell = q_\ell \cdot \mu \prod_{m=1}^{M} \mu_m^{a_{m\ell}} \tag{4.27}$$

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 (4.24) and (4.25). Specifically, this unique probability distribution $\hat{p}$ is the limit of the iterative sequence $\{\mathbf{p}^{(\delta)}; \delta =$

$0, 1, 2, \cdots\}$ defined by:

$$\begin{aligned}
\hat{p}_\ell^{(0)} &= q_\ell \\
\hat{p}_\ell^{(\delta+1)} &= \hat{p}_\ell^{(\delta)} \prod_{m=1}^{M} \left[\frac{b_m}{\hat{b}_m^{(\delta)}} \right]^{a_{m\ell}} \quad \delta = 0, 1, 2, \cdots
\end{aligned} \tag{4.28}$$

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

From the minimum KL divergence principle, the probability distribution p in the form (4.27) satisfying (4.28) will be unique. One proof to this point is given below (Darroch and Ratcliff, 1972).

Assuming π as another probability distribution that is also satisfy the constraints in Equation (4.25) and in the form of (4.27), then the KL divergence between π and q would be:

$$\begin{aligned}
J(\pi \parallel q) &= \sum_{\ell=1}^{N} \pi_\ell \log \frac{\pi_\ell}{q_\ell} \\
&= \sum_{\ell=1}^{N} \pi_\ell \left[\log \frac{q_\ell \cdot \mu \prod_{m=1}^{M} \mu_m^{a_{m\ell}}}{q_\ell} \right] \quad \text{Equation (4.27) into here} \\
&= \sum_{\ell=1}^{N} \pi_\ell \Big[\log \mu + \sum_{m=1}^{M} a_{m\ell} \log \mu_m \Big] \\
&= \log \mu \Big[\sum_{\ell=1}^{N} \pi_\ell \Big] + \sum_{m=1}^{M} \log \mu_m \Big[\sum_{\ell=1}^{N} a_{m\ell} \pi_\ell \Big]
\end{aligned} \tag{4.29}$$

The same for probability of p and q, the KL divergence can be expressed as:

$$\begin{aligned}
J(p \parallel q) &= \sum_{\ell=1}^{N} p_\ell \log \frac{p_\ell}{q_\ell} \\
&= \log \mu \Big[\sum_{\ell}^{N} p_\ell \Big] + \sum_{m=1}^{M} \log \mu_m \Big[\sum_{\ell=1}^{N} a_{m\ell} p_\ell \Big]
\end{aligned} \tag{4.30}$$

From Equation (4.29) and (4.30):

$$J(\pi \parallel q) - J(p \parallel q) = J(p \parallel \pi)$$

where $J(p \parallel \pi)$ is nonnegative and equal to zero if and only if $p = \pi$.

Furthermore, the sequence $\{\mathbf{p}^{(\delta)}; \delta = 0, 1, 2, \cdots\}$ will be a convergence process to the solution in the form of Equation (4.27) (Darroch and Ratcliff, 1972). Providing the result from δ's iteration is:

$$\hat{p}_\ell^{(\delta)} = \hat{p}_\ell^{(\delta-1)} \prod_{m=1}^{M} \left[\frac{b_m}{\hat{b}_m^{(\delta-1)}} \right]^{a_{m\ell}} \tag{4.31}$$

From the convexity of the logarithmic function:

$$\prod_{m=1}^{M} \left[\frac{b_m}{\hat{b}_m^{(\delta-1)}} \right]^{a_{m\ell}} \leq \sum_{m=1}^{M} a_{m\ell} \left[\frac{b_m}{\hat{b}_m^{(\delta-1)}} \right] \tag{4.32}$$

hence:

$$\hat{p}_\ell^{(\delta)} \leq \sum_{m=1}^{M} a_{m\ell} \hat{p}_\ell^{\delta-1} \left[\frac{b_m}{\hat{b}_m^{(\delta-1)}} \right]$$

exists and it minimize $J(p \parallel q)$ subject to constraints in Equation (4.25).

More generally, consider R sets of constraints each of them in the form of (4.25), let the r^{th} set constraint be written as:

$$\sum_{\ell=1}^{N} a_{m\ell}^r p_\ell = b_m^r, \quad r = 1, 2, \cdots, R; \quad m = 1, \cdots, M \tag{4.33}$$

where $\sum_{m=1}^{M} b_m^r = 1$. In the spatial multivariate probability estimation, this r^{th} set constraints could be any order lower marginal probability.

Provided that the constraints in Equation (4.33) are consistent with each other, there exists a unique positive probability distribution p that satisfies them and is of the form:

$$p_\ell = q_\ell \mu \prod_{r=1}^{R} \prod_{m=1}^{M} (\mu_m^r)^{a_{m\ell}^r} \tag{4.34}$$

which means that $\mathbf{p}$ can be obtained as the limit of an iterative scaling process (Bishop, 1969; Ku and Kullback, 1974).

The starting probability of the iteration q can take the uniform distribution which is the simplest and most natural choice (Darroch and Ratcliff, 1972). A more reasonable and practical choice of q assumes that all variables are independent, that is, the initial estimation will be:

$$q = \prod_{\alpha=1}^{n} \prod_{k=1}^{K} P(\mathbf{u}_\alpha; e_k)$$

Thus, the estimated multivariate probability represents a generalized independent distribution subject to the linear constraints coming from lower-marginal distributions. This method is named as **Iterative Scaling** (IS) which has been studied extensively in mathematics and statistics researches (Bishop, 1969; Ku and Kullback, 1969; Darroch and Ratcliff, 1972; Gokhale, 1972; Ku and Kullback, 1974).

The above proposed IS 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: $\hat{\mathbf{P}}^{\delta+\mathbf{1}} = (\mathbf{F} \times \hat{\mathbf{P}}^\delta)$, where $\mathbf{F} = \{ f_m, m = 1, \cdots, M \}$;
4. Set $\hat{\mathbf{P}}^{\delta+\mathbf{1}}$ to $\hat{\mathbf{P}}^\delta$ for next iteration;

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

Output: *the final iteration result*;

End

Originally, the 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 (Deming and Stephan, 1940) 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 pointed out that this approach is suitable for binary random variables estimation based on the second-order distribution with the first-order tree dependency (Chow and Liu, 1968). They

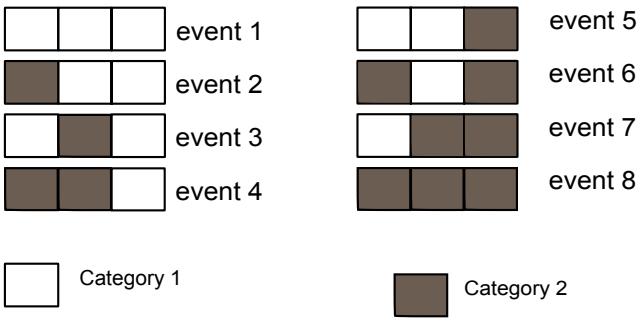


Figure 4.2: One simple multivariate data events space which is composed by three locations and two categories

pointed out the difficulty to expend to higher order of discrete multivariate random variables.

The Bayesian Maximum Entropy (BME) approach has proved its ability to predict categorical variables efficiently and in a flexible way (D'Or et al., 2001; Bogaert, 2002). In the BME approach, still the Lagrangian formulation is used (D'Or and Bogaert, 2003; Wibrin et al., 2006). The final solution to the probabilities estimation is fitting of a non-saturated log-linear model (D'Or and Bogaert, 2004). And the iterative scaling approach discussed in this section is recommended. Although the inference procedure is different with the minimizing of KL divergence used in this chapter, the same solution is used to obtain the final solution.

4.4 Implementation Examples

As already discussed in section 4.2.2, getting the solution of the $m+1$ set of coupled linear equations is a challenge with the Lagrange multiplier approach. In this section, a small example is used to illustrate this point.

4.4.1 Example of Lagrange multiplier challenge

In this case, the number of categories is 2, the random variable number N is 3 and the multivariate data event space will be $2^3 = 8$. All the possible data events are shown in Figure 4.2.

The probability of each data event is denoted as $p : \{p_1, p_2, \cdots, p_8\}$, which must be estimated. The bivariate probability constraints from the three random spatial variables $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ are $b : \{b_1, b_2, \cdots, b_{12}\}$.

The entropy of probability distribution p would be expressed as:

$$H(p) = -\sum_{\ell=1}^{8} p_\ell \log p_\ell \tag{4.35}$$

constraints are:

$$\begin{aligned}
p_1 + p_5 &= b_1 \\
p_3 + p_7 &= b_2 \\
p_2 + p_6 &= b_3 \\
p_4 + p_8 &= b_4 \\
p_1 + p_3 &= b_5 \\
p_5 + p_7 &= b_6 \\
p_2 + p_4 &= b_7 \\
p_6 + p_8 &= b_8 \\
p_1 + p_2 &= b_9 \\
p_5 + p_6 &= b_{10} \\
p_3 + p_4 &= b_{11} \\
p_7 + p_8 &= b_{12}
\end{aligned}$$

and also should satisfy:

$$\sum_{i=1}^{8} p_i = 1 \tag{4.36}$$

Using the Lagrange multiplier $\lambda_1, \lambda_2, \cdots, \lambda_{12}$ for all the constraints of bivariate probability $\{b_1, \cdots, b_{12}\}$ and λ_0 for constraints from (4.36).

Now, using the Lagrange multiplier formalism, the new objective function will

be:

$$L(p,\lambda_0,\lambda_1,\cdots,\lambda_{12}) = -\sum_{\ell=1}^{8} p_\ell \log p_\ell + \lambda_1(p_1+p_5-b_1) + \cdots + \lambda_{12}(p_7+p_8-b_{12}) + \lambda_0(\sum_{\ell=1}^{8} p_\ell - 1) \tag{4.37}$$

From this new objective function, the 21 derivatives are taken and set equal to zero:

$$\frac{\partial L}{\partial p_1} = -(1+\log p_1) + \lambda_1 + \lambda_5 + \lambda_9 = 0 \tag{4.38}$$

$$\vdots$$

$$\frac{\partial L}{\partial p_8} = -(1+\log p_8) + \lambda_4 + \lambda_8 + \lambda_{12} = 0 \tag{4.39}$$

$$\frac{\partial L}{\partial \lambda_1} = p_1 + p_5 - \pi_1 = 0 \tag{4.40}$$

$$\vdots$$

$$\frac{\partial L}{\partial \lambda_{12}} = p_7 + p_8 - \pi_{12} = 0 \tag{4.41}$$

$$\frac{\partial L}{\partial \lambda_0} = \sum_{i=1}^{8} p_i - 1 = 0 \tag{4.42}$$

From Equation (4.38) to (4.39):

$$p_\ell = exp(-\sum_{m=1}^{12} a_{m\ell}\lambda_m); \quad \ell = 1,\cdots,8 \tag{4.43}$$

Substituting into Equation (4.40),(4.41) and (4.42) leads to:

$$\sum_{\ell=1}^{8} exp(-\sum_{m=1}^{12} a_{m\ell}\lambda_m) - 1 = 0 \tag{4.44}$$

$$\sum_{\ell=1}^{8} a_{m\ell} exp(-\sum_{m=1}^{12} a_{m\ell}\lambda_m) - b_m = 0 \tag{4.45}$$

To calculate the probability distribution p_ℓ, one needs to solve this 13 coupled, implicit nonlinear equations to get the Lagrange parameters $\lambda_0,\lambda_1,\cdots,\lambda_{12}$. Although

in principle, the techniques to solve this problem are known, in practice it is not so simple. An efficient algorithm for solving the set of nonlinear equations in (4.44) and (4.45) would be required.

4.4.2 Example of iterative scaling

Instead of looking for the solution of $m+1$ nonlinear equation, the iterative scaling approach greatly simplifies the estimation process. It will be used in later spatial multivariate probability $P(\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_n)$ estimation.

Following is an example using a simple discrete probability $P = \{p_1, p_2, p_3, p_4, p_5\}$ to illustrate the iterative scaling process.

Let the linear constraints be written:

$$\sum_{\ell=1}^{5} a_{m\ell} p_\ell = b_m, \quad m = 1,2,3,4; \quad \ell = 1,2,3,4,5 \tag{4.46}$$

Given a marginal construction matrix $a_{m\ell}$, the linear constraints in Equation (4.46) could be written in traditional matrix form:

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \tag{4.47}$$

Assuming the initial probability distribution is $\{p_1^{(0)}, p_2^{(0)}, p_3^{(0)}, p_4^{(0)}, p_5^{(0)}\}$, the first

iterative scaling process will be proceed as:

$$p_1^{(1)} = p_1^{(0)}(\frac{b_1}{b_1^{(0)}})^1(\frac{b_2}{b_2^{(0)}})^0(\frac{b_3}{b_3^{(0)}})^0(\frac{b_4}{b_4^{(0)}})^1 = p_1^{(0)}\prod_{r=1}^{4}(\frac{b_r}{b_r^{(0)}})^{a_{r1}}$$

$$p_2^{(1)} = p_2^{(0)}(\frac{b_1}{b_1^{(0)}})^1(\frac{b_2}{b_2^{(0)}})^1(\frac{b_3}{b_3^{(0)}})^0(\frac{b_4}{b_4^{(0)}})^1 = p_2^{(0)}\prod_{r=1}^{4}(\frac{b_r}{b_r^{(0)}})^{a_{r2}}$$

$$p_3^{(1)} = p_3^{(0)}(\frac{b_1}{b_1^{(0)}})^0(\frac{b_2}{b_2^{(0)}})^1(\frac{b_3}{b_3^{(0)}})^0(\frac{b_4}{b_4^{(0)}})^0 = p_3^{(0)}\prod_{r=1}^{4}(\frac{b_r}{b_r^{(0)}})^{a_{r3}}$$

$$p_4^{(1)} = p_4^{(0)}(\frac{b_1}{b_1^{(0)}})^0(\frac{b_2}{b_2^{(0)}})^0(\frac{b_3}{b_3^{(0)}})^1(\frac{b_4}{b_4^{(0)}})^1 = p_4^{(0)}\prod_{r=1}^{4}(\frac{b_r}{b_r^{(0)}})^{a_{r4}}$$

$$p_5^{(1)} = p_5^{(0)}(\frac{b_1}{b_1^{(0)}})^1(\frac{b_2}{b_2^{(0)}})^1(\frac{b_3}{b_3^{(0)}})^1(\frac{b_4}{b_4^{(0)}})^0 = p_5^{(0)}\prod_{r=1}^{4}(\frac{b_r}{b_r^{(0)}})^{a_{r5}}$$

so after the first iteration the probability will be:

$$p_\ell^{(1)} = p_\ell^{(0)}\mu\prod_{r=1}^{4}[\frac{b_r}{b_r^{(0)}}]^{a_{r\ell}} \qquad \ell = 1,\cdots,5 \tag{4.48}$$

where μ is used for normalization to make the results a probability. Using the same iterative scaling process, the iterated sequence $\{\mathbf{p}^{(\delta)};\delta = 0,1,2,\cdots\}$ can be obtained. The limit of this sequence would be the solution satisfying the linear constraints.

4.5 Practical implementation

Although the iterative scaling approach has a sound mathematic background, it is not widely used due mainly to computational pressure coming from the marginalization and iterative process. The marginalization from the huge multivariate probability space requires tracing a large number of multivariate probabilities. For each bivariate marginal probability calculation from a multivariate probability with n variables and K discrete categories, the number of multivariate probabilities that need to be traced is $K^n/K^2 = K^{n-2}$. For example, when the random variable number n increases to 20 and the number of categories equal to 3, the number of the

multivariate probabilities that needs to be traced are $3^{18} = 387,420,489$. This operation will be very CPU intensive for large n. Thus, few practical applications have been done with this IS approach in multivariate probability distribution estimation. In order to implement this technique in facies modelling, some numerical developments are proposed below.

4.5.1 Marginalization construction matrix

When the multivariate probability space increases, one intensive computing process is the marginalization operation. In the iteration process, each updating step in IS procedure can be written as:

$$\begin{aligned} \hat{p}_\ell^{(\delta+1)} &= \hat{p}_\ell^{(\delta)} \prod_{m=1}^{M} \left[\frac{b_m}{\hat{b}_m^{(\delta)}} \right]^{a_{m\ell}} \\ \frac{\hat{p}_\ell^{(\delta+1)}}{\hat{p}_\ell^{(\delta)}} &= \prod_{m=1}^{M} \left[\frac{b_m}{\hat{b}_m^{(\delta)}} \right]^{a_{m\ell}} \\ \log \frac{\hat{p}_\ell^{(\delta+1)}}{\hat{p}_\ell^{(\delta)}} &= \sum_{m=1}^{M} a_{m\ell} \log \frac{b_m}{\hat{b}_m^{(\delta)}} \\ \log \hat{p}_\ell^{(\delta+1)} - \log \hat{p}_\ell^{(\delta)} &= \sum_{m=1}^{M} a_{m\ell} \log \frac{b_m}{\hat{b}_m^{(\delta)}} \end{aligned} \tag{4.49}$$

where $\hat{b}_m^{(\delta)} = \sum_{m=1}^{M} a_{m\ell} \hat{p}_\ell^{(\delta)}$.

As it is shown, the marginal construction matrix $a_{m\ell}, m = 1, \cdots, M; \ell = 1, \cdots, N$ is used not only in the marginalization operation but also in the multivariate probability updating step. As the data event space increases, so does the marginalization matrix dimension. Generally, the dimension will be defined from the marginal order m, the total location number n and the facies number K as:

$$K^m \cdot \binom{n}{m} \times K^n \tag{4.50}$$

For example, if there are 20 random variables and 3 categories, the dimension of the marginal construction matrix will be $1710 \times 3,484,784,401$. Some of the marginal construction matrix are as listed in Table 4.1.

Table 4.1: The dimension of marginal probability construction matrix

K	n	m	$a_{m\ell}$ dimension
3	3	2	$\binom{3}{2} \cdot 3^2 \times 3^3 = 27 \times 27$
3	5	2	$\binom{5}{2} \cdot 3^2 \times 3^5 = 90 \times 243$
3	10	3	$\binom{10}{3} \cdot 3^3 \times 3^{10} = 3240 \times 59049$

Handling such a high dimension matrix efficiently is a challenge in the numerical implementation of this method. Building and saving the marginalization construction matrix $a_{m\ell}$ efficiently in the iteration process will release some computational burden. The naive data structure for a matrix is to use an array. Each entry $a_{m\ell}$ can be accessed by the two indices m and ℓ. Huge memory is needed to store all the entries to represent the matrix.

Although the dimension of $a_{m\ell}$ increases dramatically as the number of random variables increases, there are many zero values in the matrix and non-zero values are always equal to one because in the marginalization only some of the multivariate probability values will be summed up as illustrated in Figure 2.3 and Figure 2.4. The marginalization computation can take advantage of this sparse matrix and can proceed with a more efficient computation.

There are many ways to represent a sparse matrix (Stoer et al., 2002). The approach used in this research is List of Lists (LIL). Other expressions could be used. In the LIL approach, only the non-zero column indices are stored. In this research, all the non-zero column indices are calculated from the multivariate event index function as in Equation (2.4). The sparse matrix is saved by a one dimensional array (only the nonzero elements column number) and two parameters: the number of total rows and the number of non-zero elements in each row of the naive matrix $a_{m\ell}$. Substantial memory reduction is obtained and yields huge savings in memory when compared to a naive approach.

The marginalization computation can be done in a fast linear operation style with relatively small storage requirement by taking advantage of the sparse matrix operation. More importantly, the sparse matrix is constant according to the order

of multivariate probabilities and can be built only once. It saves a lot of CPU time when the marginalization is needed in every iteration.

4.5.2 Numerical examples

The die problem

The die problem serves as an excellent illustration of different solution efforts including the iterative scaling and the Lagrange multiplier formalism. This problem was originally proposed by Jaynes as an example to show the ME principle for undetermined problem (Jaynes, 1978). In this case, there is no difference in the final results from using Lagrange multiplier and the iterative scaling approach. However, the IS procedure is more straightforward and easier to implement than the Lagrange multiplier approach.

Consider a die of six faces that is tossed for $T \quad (T \to +\infty)$ times. One is told that the average number of spots up was not 3.5 as we might expected from an "honest" die but 4.5. Given this information, and nothing else, what probability should one assign to i spots on the next toss?

From the ME approach, the solution could be proceed as following procedures. The constraints to the entropy equation would be:

$$\sum_{i=1}^{6} i \cdot p_i = 4.5 \tag{4.51}$$

$$\sum_{i=1}^{6} p_i = 1 \tag{4.52}$$

the new objective function with Lagrange multipliers would be:

$$L = -\sum_{i=1}^{6} p_i \log p_i + \lambda_0 (\sum_{i=1}^{6} p_i - 1) + \lambda_1 (\sum_{i=1}^{6} i \cdot p_i - 4.5) \tag{4.53}$$

The probabilities would be:

$$p_i = exp(\lambda_0 + i\lambda_1) \tag{4.54}$$

$$\Lambda = exp(\lambda_0) = \sum_{i=1}^{6} exp(-i\lambda_1) \tag{4.55}$$

$$\Lambda = x(1-x)^{-1}(1-x^6) \quad \text{where:} \quad x = exp(-\lambda_1) \tag{4.56}$$

$$\frac{\partial \Lambda}{\partial \lambda_1} = -4.5 \tag{4.57}$$

$$3x^7 - 5x^6 + 9x - 7 = 0 \tag{4.58}$$

After obtaining the desired root for the Equation (4.58), the maximum entropy probabilities are calculated to be:

$$\{0.05435, 0.07877, 0.11416, 0.16545, 0.23977, 0.34749\} \tag{4.59}$$

While for the iterative scaling process, the marginal construction function would be: $a_{1i} = \{1,2,3,4,5,6\}$. According to the iterative scaling process, the first estimation to the desired probability would be: $p_i^0 : \{1/6, 1/6, 1/6, 1/6, 1/6\}$. The first time of scaling would be:

$$p_1^1 = \mu p_1^0 (\frac{4.5}{3.5})^1$$
$$p_2^1 = \mu p_2^0 (\frac{4.5}{3.5})^2$$
$$p_3^1 = \mu p_3^0 (\frac{4.5}{3.5})^3$$
$$p_4^1 = \mu p_4^0 (\frac{4.5}{3.5})^4$$
$$p_5^1 = \mu p_5^0 (\frac{4.5}{3.5})^5$$
$$p_6^1 = \mu p_6^0 (\frac{4.5}{3.5})^6$$

All the ten times iteration results are listed in Table 4.2 The Lagrange result and the final IS result are almost exactly the same. The IS process is more straightforward and easier to implement for large problems.

One multivariate probability estimation

The iterative scaling approach is illustrated numerically in this example with the

Table 4.2: Ten times iteration results for the die problem

iteration time	p_1	p_2	p_3	p_4	p_5	p_6
1	0.08123	0.10444	0.13428	0.17265	0.22198	0.28540
2	0.06503	0.08945	0.12306	0.16928	0.23286	0.32032
3	0.05914	0.08365	0.11832	0.16736	0.23672	0.33482
4	0.05660	0.08108	0.11615	0.16639	0.23835	0.34143
5	0.05543	0.07989	0.11512	0.16591	0.23909	0.34456
6	0.05488	0.07931	0.11463	0.16567	0.23944	0.34606
7	0.05461	0.07904	0.11439	0.16556	0.23961	0.34680
8	0.05448	0.07890	0.11427	0.16550	0.23970	0.34715
9	0.05441	0.07883	0.11421	0.16547	0.23974	0.34733
10	0.05438	0.07880	0.11419	0.16546	0.23976	0.34741

simple case $n = 6, K = 3$. The true multivariate probability is scanned from one training image. It is shown that the multivariate probabilities are reproduced from the bivariate marginals.

Given a training image and one data configuration, the multivariate probability concerning the joint outcome for this grouped location will be scanned from the training image as discussed in the traditional multiple point geostatistics such as the work of Guardiano and Srivastava (Guardiano and Srivastava, 1993). Each bivariate probability b_m is calculated from the scanned multivariate probability using Equation (4.25). The estimated multivariate probability $\hat{p}$ is obtained from the full set of bivariate probabilities using the iterative scaling approach introduced above. Then, the estimated multivariate probability distribution $\hat{p}$ is compared with the original scanned true one p as shown in Figure 4.3. The sequence calculated from different iterations is plotted together in Figure 4.3. The process converges to a feasible solution which is very close to the true multivariate probability.

Comparing the difference between the target bivariate probability $b(\mathbf{u}_i, \mathbf{u}_j)$ and the iterated bivariate probability $\hat{b}(\mathbf{u}_i, \mathbf{u}_j)$ shown in Figure 4.4, the difference between them is close to zero after 30 iterations.

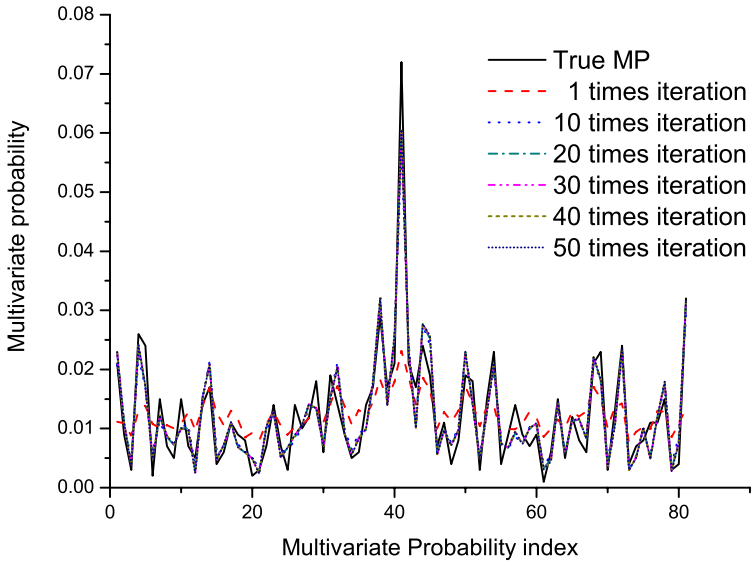


Figure 4.3: The convergence of the iteration results to the true multivariate probability

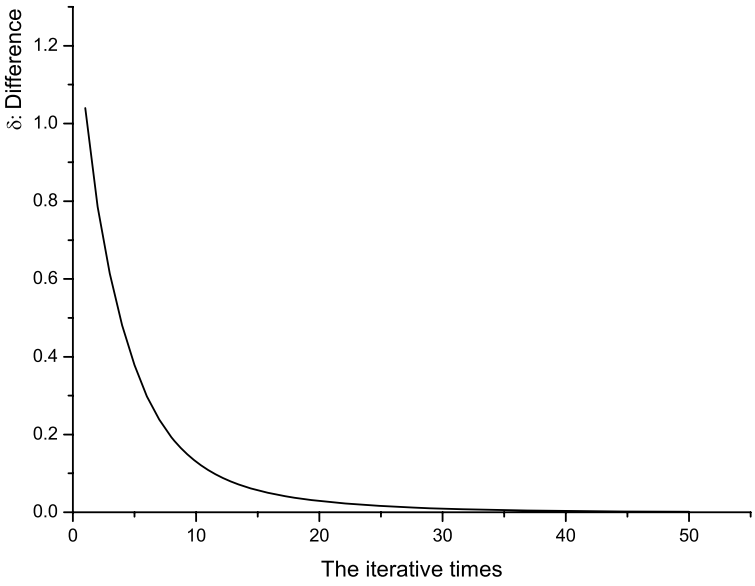


Figure 4.4: The difference to the bivariate probability constraints

4.6 Remarks

The direct multivariate probability estimation from the bivariate marginal probability using the minimum KL divergence principle is theoretically correct. The numerical implementation of the iterative scalling approach proposed in this chapter makes the approach more practical in geostatistical research.

Although the computation of the IS approach is complex, it is not a barrier for implementation in geostatistics research and would be a flexible and powerful approach if the CPU considerations are not a major concern.

Chapter 5

Facies Modelling Using DMPE

Updating the prior probability distribution to a posterior probability distribution at each unsampled location using the new proposed spatial distance calculation in spatial probability mapping is illustrated with examples. These examples show that more geological information can be reproduced. The resulting probability maps obtained from different interpolation algorithms are compared by measures of accuracy and precision. The assessment of the cross validation results and the KL measurements of the maps show that updating the prior probability to posterior probability map using the new proposed DMPE will provide more accurate and precise results than the traditional IK approach. Sequential simulation is implemented to characterize the joint uncertainty over the whole research area. Some programming details for the DMPE implementation, such as the Markov model at the conditioning data choosing, are explained in this chapter.

5.1 Estimation

The main aim of geostatistics is to update the prior probability $P(\mathbf{u}_0)$ to a posterior probability $P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)$ based on the information from the nearby sampled locations $(\mathbf{u}_1,\ldots,\mathbf{u}_n)$. In Chapter 3, the information from the sample locations are expressed as bivariate probabilities between each pair of them. Any arbitrary spatial distance will be scaled to an effective distance along the vertical direction and the bivariate probabilities will be obtained from the well data. In Chapter 4, it is

shown that all these bivariate probabilities can be used to construct a multivariate probability distribution that is based on the minimum KL distance principle. The estimated multivariate probability will be used directly to calculate the posterior probability $P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)$. All those steps are the basis of the proposed spatial probability interpolation methodology called Direct Multivariate Probability Estimation (DMPE).

5.1.1 Spatial probability interpolation

The true probability for each facies at the sample locations is known. Based on the dependence relationship between the locations, the unsampled locations will be assigned a probability value for each facies. These probability distributions will be visually presented as a map for each facies. There will be K maps for a domain with facies space $e_k, k = 1,\cdots,K$. This process is called spatial probability interpolation or "estimation" in geostatistics.

The global proportion p_k will reflect the average frequency of facies e_k that can be found in the study area. The global proportion would be used when one is interpolating probability at an unsampled locations with no further information available. This will satisfy the maximum entropy/minimum KL distance principle as explained above. As an example shown in Figure 5.1, the probability map for each facies would be the global proportion. Of course, the probability would be 1 or 0 at data locations in this situation.

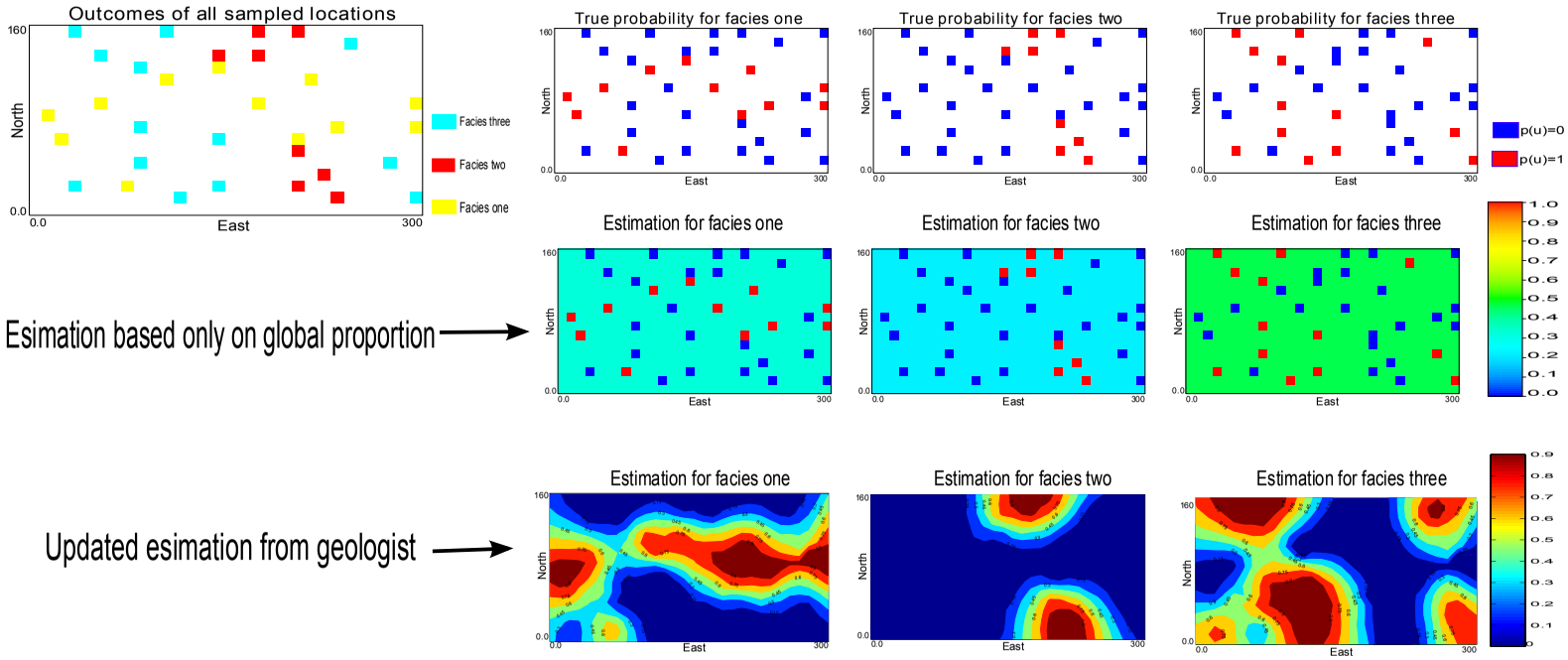


Figure 5.1: Facies spatial interpolation based on different informations

In practice, the facies will always show some spatial correlation in different directions. After more information is gained, the posterior probability maps will be updated. For example, after doing some geological analysis, it may be believed that the geological background of the reservoir is fluvial. It may also be clear that one facies will be the paleochannel and is extended from east to west, a second facies could be alluvial fan, and a third facies is flood plain. An updated probability map for each facies can be obtained through manually sketch by geologists as shown in Figure 5.1. Some other methods, such as IK, can be used if the correlation between the sampled data can be obtained.

In this research, the correlation information between each location will be expressed as bivariate probability distributions between them and will be used as a constraints to the posterior probability map drawing. The DMPE will be used to get the posterior probability map. The updating process will be:

1. Set up the anisotropy model of the study area and set up the model coordinates;

2. Build the vertical bivariate probability diagram;

3. Pick an unsampled location $\mathbf{u}_0$ and search for surrounding conditional data $(\mathbf{u}_1, \ldots, \mathbf{u}_n)$;

4. If there are no conditioning data, use the global proportion;

5. If there is only one conditioning data, then use the input bivariate probability and the global proportion to calculate the conditioning probability as:

$$P(\mathbf{u}_0|\mathbf{u}_\alpha) = \frac{P(\mathbf{u}_0, \mathbf{u}_\alpha)}{P(\mathbf{u}_\alpha)}$$

6. If the number of conditioning data is more than one then:

   (a) Retrieve the bivariate probability distribution $P(\mathbf{u}_\alpha, \mathbf{u}_\beta); \alpha \neq \beta$ from the locations $(\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_n)$ and use the new proposed anisotropy based

distance calculation in order to integrate the geological understanding into the final model;

(b) apply the proposed DMPE algorithm to estimate the $n+1$ multivariate probability distribution $P(\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_n)$;

(c) Calculate the conditional probability for the unsampled location $\mathbf{u}_0$ as:

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

7. Go to next unsampled location, until all the unsampled locations are visited.

Only one unsampled location is considered at a time. The above spatial probability interpolation is implemented in a Gslib style program named `DMPEest`. The program details and the parameter file are given in the Appendix.

5.1.2 2D Spatial Estimation Example

A small example is used to illustrate the DMPE approach for facies spatial estimation. The main aim is to illustrate that the geological understanding can be integrated into the geostatistical model through the proposed spatial distance calculation.

The calculation is done along the horizontal plane defined by strike and dip direction. The hard data for this direction is shown in Figure 5.2. There is a sine wave fluctuation along the strike direction (North-to-South). While along the dip direction (West-to-East), the facies pattern would be like the one calculated from a training image as shown in Figure 3.4. The estimation results from using the anisotropy based spatial distance calculation approach are shown in Figure 5.3. The estimation results from a traditional geometric distance calculation approach are shown in Figure 5.4.

Comparing the results, there are clear geological constraints in the model with the anisotropy based distance calculation. The probability map is not only conditioned to the hard data constraints, but also on our geological understanding. The random trend along strike could be modified with more geological analysis.

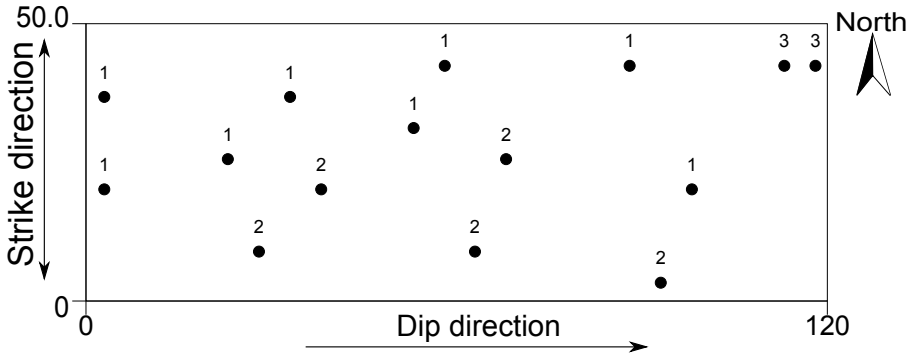


Figure 5.2: The conditioning data along horizontal plane and the strike and dip directions

5.2 Estimation results assessment

The probability distribution $P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)$ at each unsampled location $\mathbf{u}_0$ must be checked. For a probability distribution, the basic check is that the sum of probabilities should equal to 1. This requirement is already satisfied from the DMPE algorithm as the conditional probabilities are from a fully validate multivariate probability distribution. The assessment below is an evaluation of the accuracy and precision information that the probability distribution can provide.

5.2.1 Cross validation assessment

Cross validation is a technique for assessing how accurately a predictive model will perform in practice. It is usually done through comparison between the true value and the estimated probability results and is mainly used in settings where the goal is prediction (Picard and Cook, 1984). In geostatistical cross validation (Davis, 1987), actual data values are dropped one at a time and re-estimated from the remaining neighbouring data. Each datum is replaced in the data set once it has been validated.

From the indicator transformation, it is known that the indicator value of each facies is its probability to exist at the current location. As shown in Figure 5.1, the hard data for each facies spatial probability mapping is only 1 or 0. If it is 1, the current facies is found at this location. If it is 0, the current location is not the facies under consideration. While using any spatial probability interpolation algorithm or

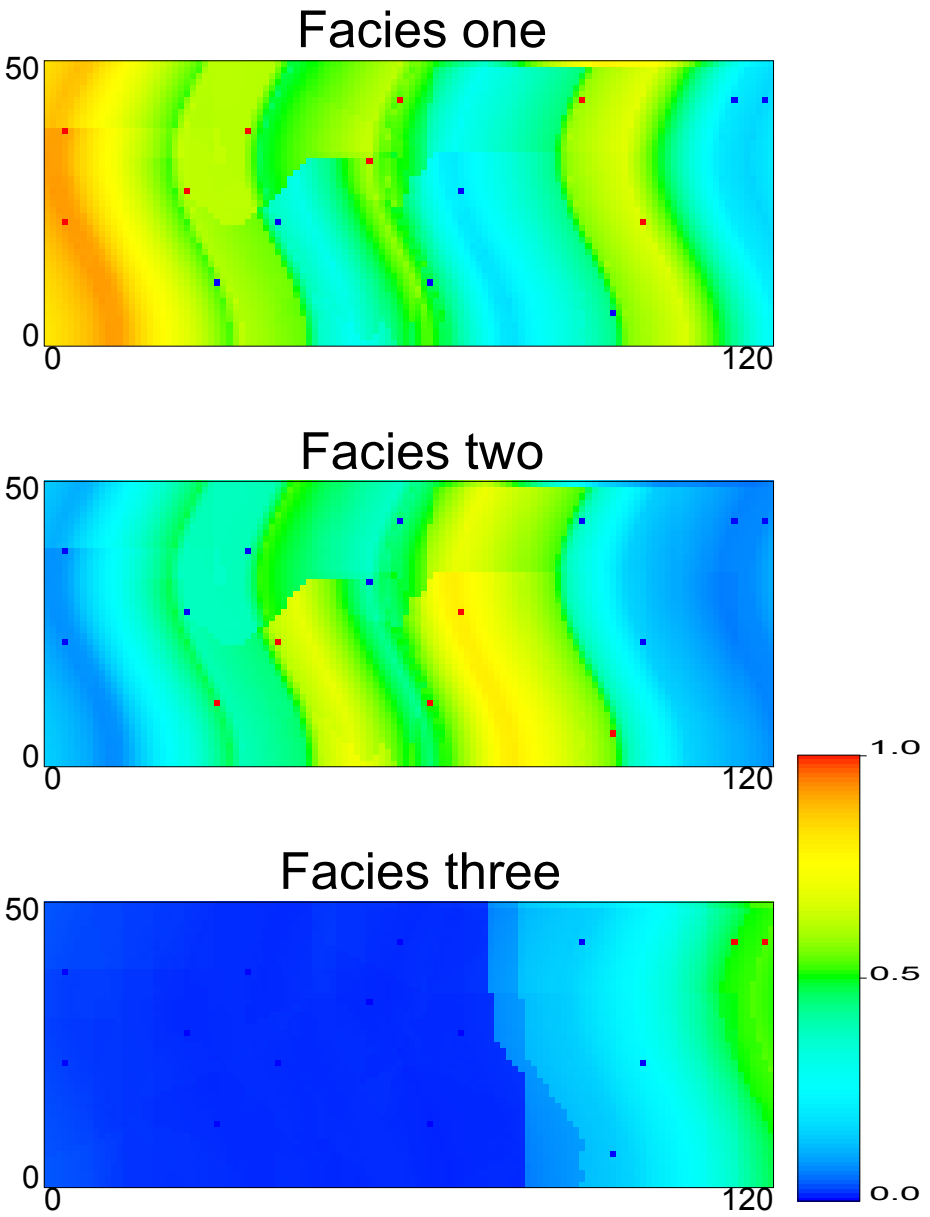


Figure 5.3: The estimation along dip and strike direction using the new proposed spatial distance calculation approach

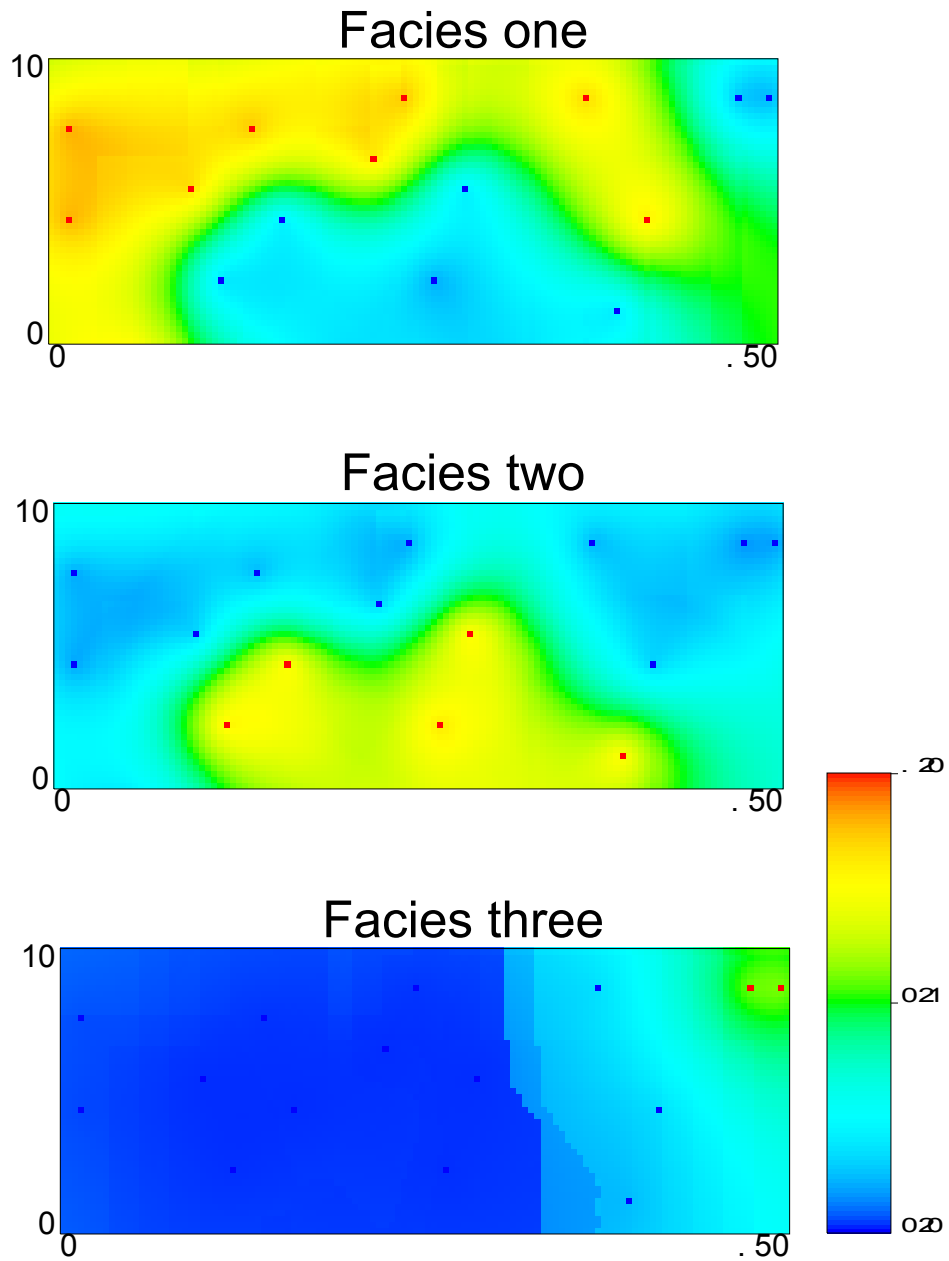


Figure 5.4: The estimation along dip and strike direction using the traditional geometric distance calculation approach

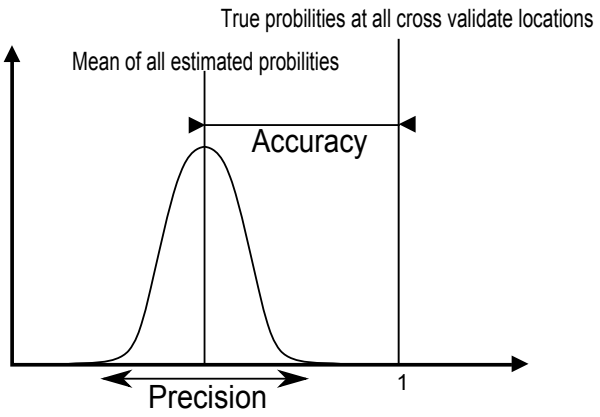


Figure 5.5: The accuracy and precision criteria of cross validation results

estimation approach, the estimated posterior probability for facies $e_k, k = 1, \cdots, K$ would always be between 0 and 1 except at the sampled locations.

One way to evaluate the cross validation results is to check the closeness to true value and the dispersion of the estimated probabilities from this grouped locations. It is done by grouping the estimated probabilities $p(\mathbf{u}; e_k|(n))$ for data event $\mathbf{u} = e_k$ at several locations together. The mean of these grouped estimated probabilities $\mu_k = E\{p[\mathbf{u}; e_k|(n)]\}$ should be as close to one as possible. The closeness to the true value will be calculated as $\{1 - \mu_k\}$ and will be used as one criteria to the estimate accuracy.

The dispersion of the grouped probability values will reflect the precision of the estimation model. The dispersion is quantified using the standard deviation of the grouped probabilities, which is calculated as: $\sigma_k = \sqrt{E\{p[\mathbf{u}; e_k|(n)] - \mu_k\}^2}$. These two quantitative measurements are shown in Figure 5.5. These two checks will be done for each facies separately.

A small comparison is done between the traditional IK approach and the proposed DMPE approach. The training image shown in Figure 5.6 is used as a reference distribution. A total of 1000 locations are picked for checking the estimation of each facies. At every cross validation location, eight surrounding data will be randomly selected as conditioning data. The DMPE algorithm and the IK are used to construct the conditional probability at each cross validation location. In DMPE,

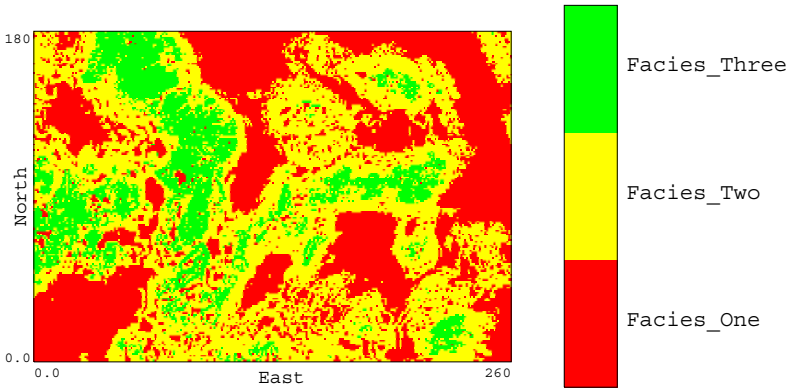


Figure 5.6: The training image used for doing cross validation from some randomly picked locations

Table 5.1: The comparison between the clossness to the true probability from cross validation

statistics	Facies one	Facies two	Facies three
$\{1-\mu_k\}$ from DMPE	0.66	0.60	0.52
$\{1-\mu_k\}$ from IK	0.68	0.64	0.70
σ_k from DMPE	0.32	0.24	0.30
σ_k from IK	0.33	0.24	0.28

the bivariate probabilities for the conditioning data are taken from the bivariate probability diagram along X and Y directions. In the IK approach, order relations problems are corrected immediately following estimation.

As shown in Figure 5.7 and Table 5.1, the estimated results from DMPE are closer to the "true probability distribution". In other words, DMPE is more accurate than IK which is an expected result since more geological information is being used. In this comparison, only bivariate marginal probabilities are used as constraints in the multivariate probability inference. Higher order marginal probability results would help in the process.

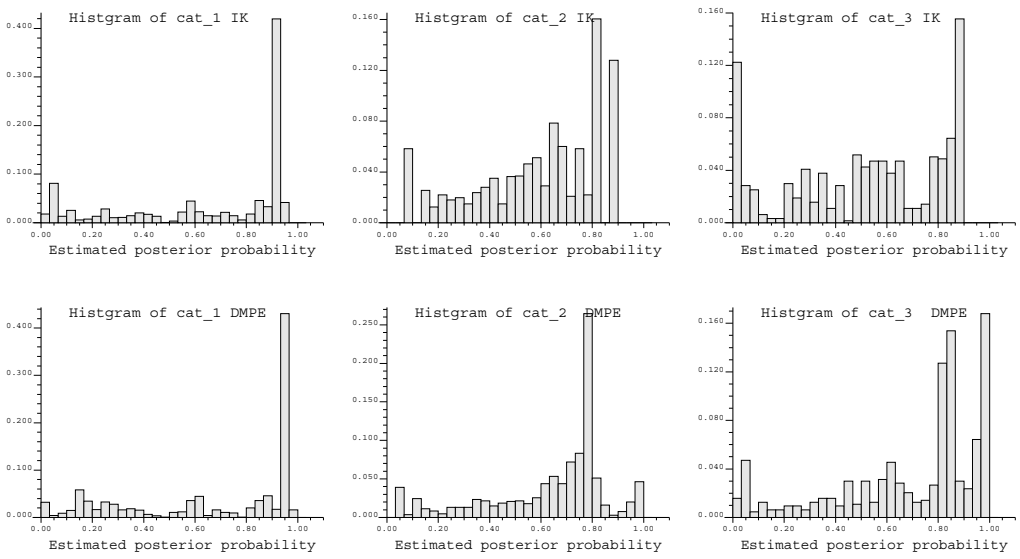


Figure 5.7: Histogram of the estimated probability at the hard data locations

5.2.2 KL distance assessment

The above assessment can be done facies by facies at each cross validation location. The starting point is a prior probability distribution such as the global proportion $\{p_k, k = 1, \cdots, K\}$. A quantitative measurement is needed to evaluate how much of the uncertainty is changed using the posterior probability instead of the global proportions.

The previous introduced KL divergence will be used as a quantitative measurement of information gain in moving from a prior distribution to a posterior distribution. The amount of useful information gain at location $\mathbf{u}_0$ after upgrading the estimation from the prior probability $P(\mathbf{u}_0)$ to posterior probability $P(\mathbf{u}_0|\mathbf{u}_1, \ldots, \mathbf{u}_n)$ will be:

$$J\Big(P(\mathbf{u}_0|\mathbf{u}_1, \ldots, \mathbf{u}_n) \parallel P(\mathbf{u}_0)\Big) = \sum P(\mathbf{u}_0|\mathbf{u}_1, \ldots, \mathbf{u}_n) \log \left[\frac{P(\mathbf{u}_0|\mathbf{u}_1, \ldots, \mathbf{u}_n)}{P(\mathbf{u}_0)}\right] \tag{5.1}$$

For example, upgrading from the uniform probability distribution ($\frac{1}{K}$) to a pos-

terior probability, the KL divergence is:

$$
\begin{aligned}
& J\big(P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n) \parallel (\frac{1}{K})\big) \qquad (5.2)\\
=& \sum P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)\log\left[\frac{P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)}{1/K}\right]\\
=& \sum\left[P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)\log(P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)) - P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)\log(1/K)\right]\\
=& \sum P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)\log(K) + \sum\left[P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)\log P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)\right]\\
=& \log(K) + \sum\left[P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)\log P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)\right]\\
=& H(\mathbf{u}_0) - H(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)
\end{aligned}
$$

Because of $0 \leq J \leq 1$, the new updated entropy $H(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)$ will always be less than $H(\mathbf{u}_0)$, that is, knowing more information can only reduce the uncertainty (Cover and Thomas, 2006). The larger the distance the more informative the estimation is, the greater our uncertainty is reduced at the unsampled location.

It is interesting to understand what the distance defined in Equation (5.1) for binary category variable $K = 2$ will be. Let the logarithms have base 2, the distance in Equation (5.1) will be:

$$
\begin{aligned}
J\big(P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n) \parallel (\frac{1}{K})\big) &= \log_2(2) - H(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n) \qquad (5.3)\\
&= 1 + p(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)\log_2 p(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)\\
&\quad + [1 - p(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)]\log_2[1 - p(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)]
\end{aligned}
$$

The above distance decreases from 1 to 0 when the posterior probability $p(\mathbf{u}|\mathbf{u}_1,\ldots,\mathbf{u}_n)$ increases from 0 to 0.5 which is the uniform distribution for a binary variable. While as the posterior probability continues to increase from 0.5 to 1, the distance increases from 0 to 1, as shown in Figure 5.8.

The prior probability could also be the global proportion of each facies $p(k), k = 1,\cdots,K$. The KL distance would reach zero in the case of the posterior probability equal to the global proportion. In this situation, the distance will be zero.

For the general case, the number of facies would be more than just two. At each location, the KL distance could be calculated from the posterior probability

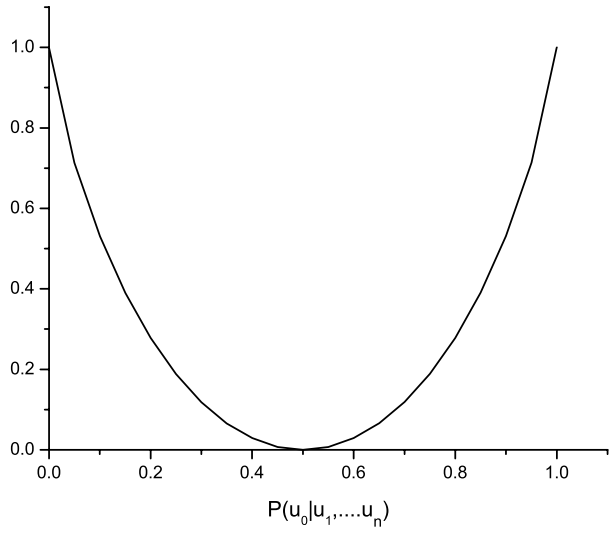


Figure 5.8: The calculated KL distance for different posterior probability distributions for binary variable

$P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)$ and its prior probability $P(\mathbf{u}_0)$. The uncertainty reduction from the prior probability will be characterized by this KL distance. It will give a global assessment to the probability map.

Consider a small example to show the global uncertainty assessment from the probability mapping. The hard data for this example is taken from one training image in Figure 5.9.

The bivariate probability diagram is scanned from the training image along the X direction. During the estimation, the bivariate probability for arbitrary location pairs is used from this direction. The multivariate probability is estimated from the conditioning data for each unsampled location. As a comparison, the IK estimation is also calculated from those bivariate probability based on the relationship between bivariate probability and indicator covariance in Equation (2.20).

The conditional probability maps from both methods are shown in Figure 5.10. One reason of the noise in the maps is that the DMPE estimation is based on the ME principle. The final probability map tends to have a maximum entropy charac-

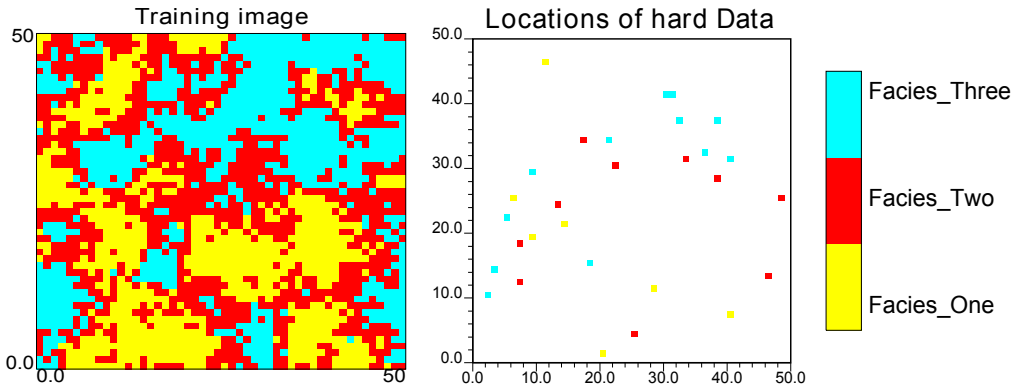


Figure 5.9: The training image (left) and hard data set (right) used in uncertainty assessment with KL distance criteria

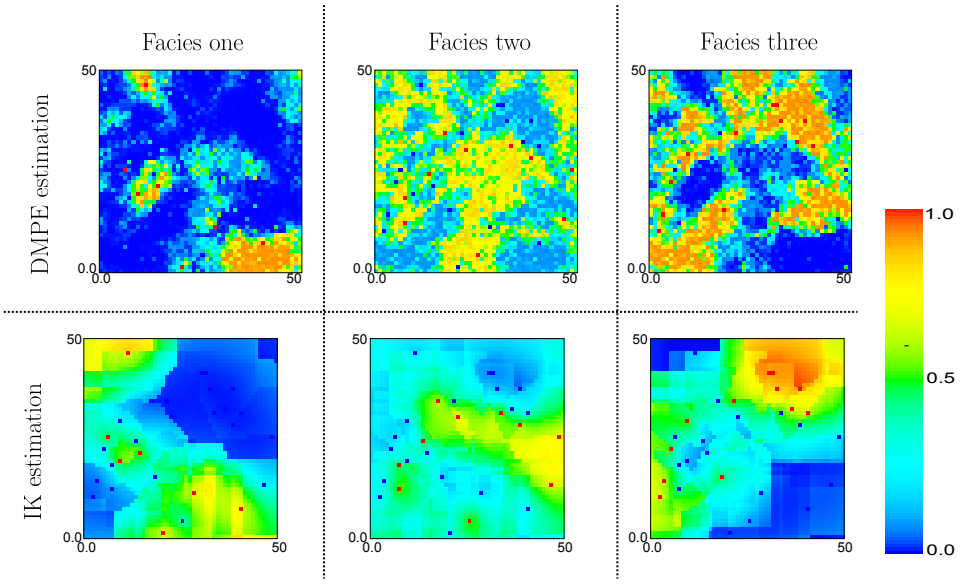
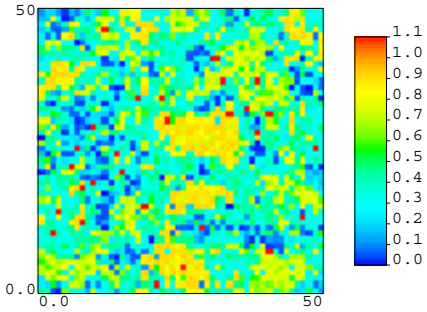
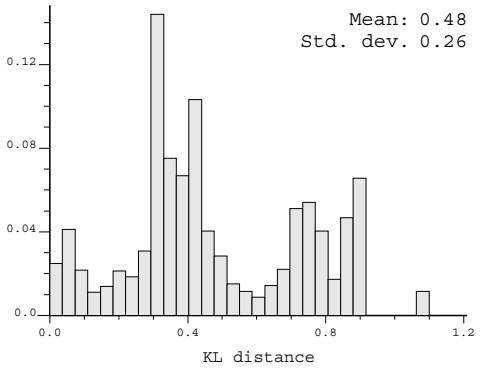


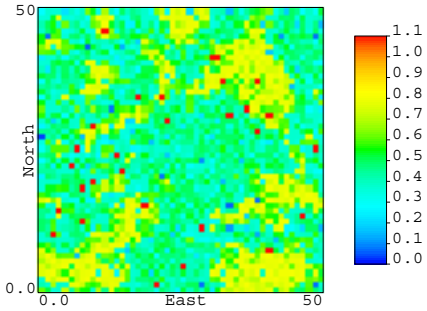
Figure 5.10: Estimation results from DMPE and IK using the same hard data and spatial bivariate probability matrix



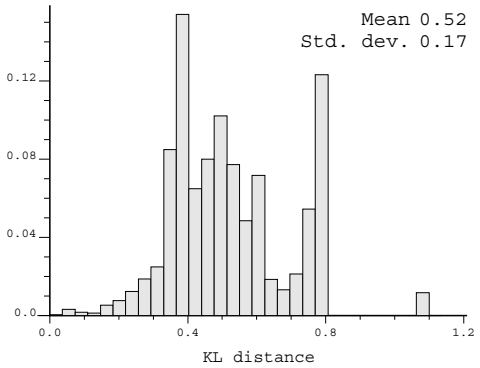
(a) KL distance map calculated from IK



(b) Histogram of KL distance calculated from IK



(c) KL distance map calculated from DMPE



(d) Histogram of KL distance calculated from DMPE

Figure 5.11: The KL distance calculated from two different probability maps

teristic. More conditioning data could be used for IK approach to obtain a smooth map. But the main aim is a comparison between those two methods, thus, the same conditioning situation is used in the estimation.

Using the KL distance defined in Equation (5.1), the information gain from both methods will be plotted as one map shown in Figure 5.11. Comparing the mean value of the KL divergence from those two estimations, the mean of the KL distance is increased from 0.48 to 0.52 from the IK to DMPE method. The standard deviation decreases from 0.48 in IK to 0.17 in DMPE. Globally, the estimation result from DMPE gives a more precise estimation result under same information resource.

5.3 Sequential Simulation

The probability map shown in Figure 5.10 shows the probabilities to find each facies at the unsampled locations. Usually, the final facies models are plotted as a map where a specific facies,not a probability , is assigned at every grid cell. Bierkens and Burrough proposed to use the category with the largest probability of occurrence which is defined as the map purity (Bierkens and Burrough, 1993; Lark and Beckett, 1998). This approach is inadequate if one aims at reproducing the joint outcomes from the entire area.

From a mathematics view, a facies map for N locations together will represent one realization from the data event space defined from these N locations and K categories together which is K^N. The multivariate probability distribution $P(\mathbf{u}_1, \cdots, \mathbf{u}_N)$ will characterize the outcomes of all the drawings. Each drawing is called as a realization which will have a specified probability to exist. Multiple realizations are possible from a multivariate probability distribution that meet criteria derived from measurement data such as given global proportion and bivariate marginal probabilities.

Generally, an extensive exploration of the data event space should be performed to characterize uncertainty. As the data event space is so huge ,well-designed random explorations scheme should be adopted to do the exploration. The sequential simulation is the one used mostly in geostatistical simulation to do this exploration (Gomez-Hernandez and Journel, 1993; Lantuejoul, 2001; Soares, 2001).

5.3.1 Sequential simulation

Exploring the uncertainty described by the multivariate probability distribution involves drawing realizations. The basic idea is that a large number of synthetic geological structures are generated based on a stochastic description of the system. The simulated realizations have the same values at the sample locations as the measurements, that is, the hard data are reproduced exactly. Simulation algorithms yield multiple equally likely outcomes from the algorithm.

In this method, one uses the property that a general multivariate probability defined in n locations can always be decomposed as the product of a marginal probability for one location and a series of conditional probabilities for other locations:

$$\begin{aligned} p(\mathbf{u}_1,\cdots,\mathbf{u}_n) &= p(\mathbf{u}_n|\mathbf{u}_{n-1},\cdots,\mathbf{u}_1)p(\mathbf{u}_{n-1},\cdots,\mathbf{u}_1) \\ &= p(\mathbf{u}_n|\mathbf{u}_{n-1},\cdots,\mathbf{u}_1)p(\mathbf{u}_{n-1}|\mathbf{u}_{n-2},\cdots,\mathbf{u}_1)p(\mathbf{u}_{n-2},\cdots,\mathbf{u}_1) \\ &= \cdots \quad \cdots \quad \cdots \\ &= p(\mathbf{u}_n|\mathbf{u}_{n-1},\cdots,\mathbf{u}_1)p(\mathbf{u}_{n-1}|\mathbf{u}_{n-2},\cdots,\mathbf{u}_1)\cdots p(\mathbf{u}_3|\mathbf{u}_2,\mathbf{u}_1)p(\mathbf{u}_2|\mathbf{u}_1)p(\mathbf{u}_1) \end{aligned} \tag{5.4}$$

In each step, Monte-Carlo simulation is used to draw a facies type from each conditional probability distribution. Monte-Carlo method is a powerful tool in simulating spatial phenomena, while few assumptions are required (Mosegaard and Tarantola, 1995). As shown in Figure 5.12, it goes as:

- Building a quantile function $Q(t)$ from the estimated conditional probability for example using the conditional probability $P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)$. The probability $P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)$ may need be ordered according to $k = 1,\cdots,K$;
- Draw a random number t which follows a uniform distribution;
- Read quantile function $Q(t)$ to find which k that is specified by the random number;
- Assign the outcome k to the current location as a realization from the probability distribution $P(\mathbf{u}_0|\mathbf{u}_1,\ldots,\mathbf{u}_n)$;

In geostatistics, the principle illustrated in Equation (5.4) is used in the classical sequential simulation for simulating a set of random variable jointly conditioned on sampled locations. The implementation of sequential simulation consists of reproducing the desired spatial properties through the sequence of conditional distributions. Each drawing from such a multivariate probability distribution $P(\mathbf{u}_0,\cdots,\mathbf{u}_n)$ represents a simulated realization of the reservoir.

The simulated facies using the Monte-Carlo drawing from the estimated conditional probability are plotted where a facies type is present at every grid cell. As an

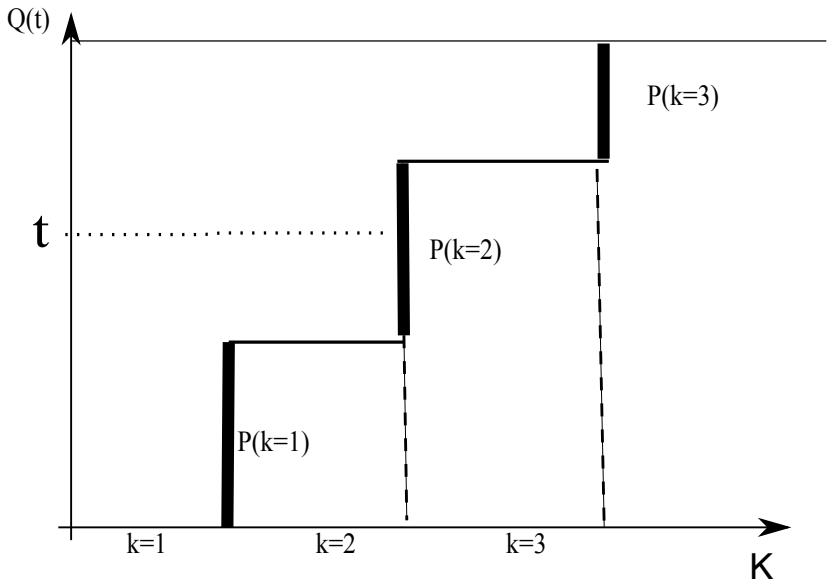


Figure 5.12: Monte-Carlo simulation with discrete probability distribution

example in Figure 5.13, four realizations are coming from the same conditioning data and under the same bivariate probability constraints. Although they share the same spatial structures, there are some differences at small scale.

5.3.2 Markov model in sequential simulation

The sequential simulation principle is theoretically valid with no approximation or assumption. But in practice, as the number of conditioning data increase, the dimension of the multivariate probability space will be difficult to manage. In implementation, the Markov assumption is adopted to solve the size problem.

Markov models based on various data screening hypotheses are often used because they reduce the statistical inference burden. For example, an extensively sampled secondary data are integrated into the geostatistical model in the practice of cokriging with the introduction of the Markov coregionalization model (Shmaryan and Journel, 1999; Journel, 1999). Based on the Markov model, it is assumed that the current location will be dependent on the nearby locations in a certain area which will screen all data that are further away.

With this Markov-type assumption, only some locations *n* from all the surround-

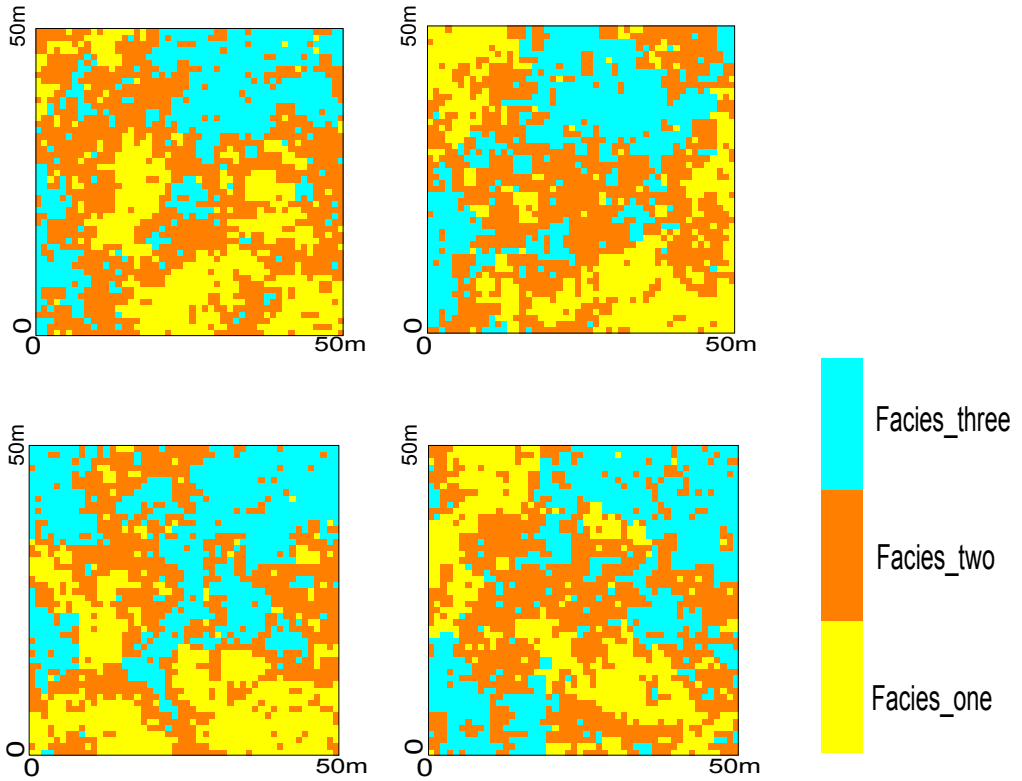


Figure 5.13: Four different realizations from sequential simulation with the same hard data set and bivariate probability constraints

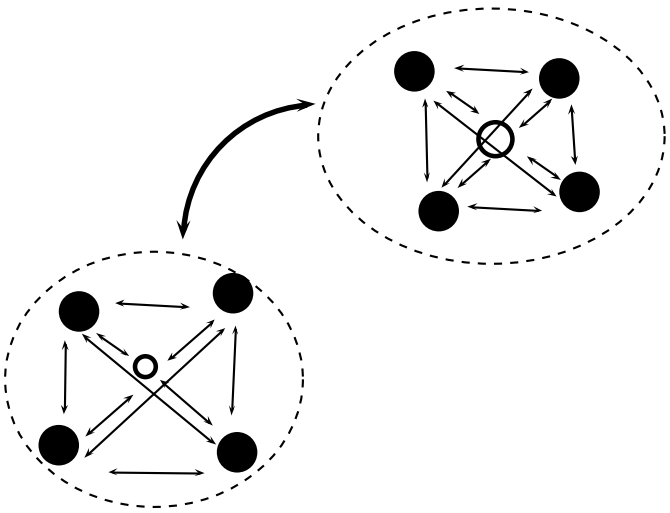


Figure 5.14: Illustration of a Markov random function field model

ing locations will bring information to the current unsampled location. As shown in Figure 5.14, the posterior probability will be a function of the bivariate statistics between these locations $(\mathbf{u}_1, \ldots, \mathbf{u}_n)$, written as:

$$P(\mathbf{u}_0|\mathbf{u}_1, \ldots, \mathbf{u}_n) = f(P(\mathbf{u}_i, \mathbf{u}_j), i, j = 0, 1, \cdots, n) \tag{5.5}$$

Normally around 40 conditioning data are used in traditional IK approach. For DMPE, around 10 conditioning data are used for most of cases. If permitted, more conditioning data should be used in DMPE approach as there are no theoretical constraints.

5.3.3 Stochastic reservoir models assessment

Simulation aims at drawing realizations that reflect the statistics models from the data. Thus, the first assessment to different realizations is how well the input statistics are reproduced. As one example in Figure 5.15, the input bivariate probability diagram is compared with the output from several realizations.

Given that limited conditioning data can be used in the DMPE program, it is impossible to reproduce the input bivariate probability exactly. As shown in Figure 5.15, the bivariate probability diagrams are reasonably reproduced by all re-

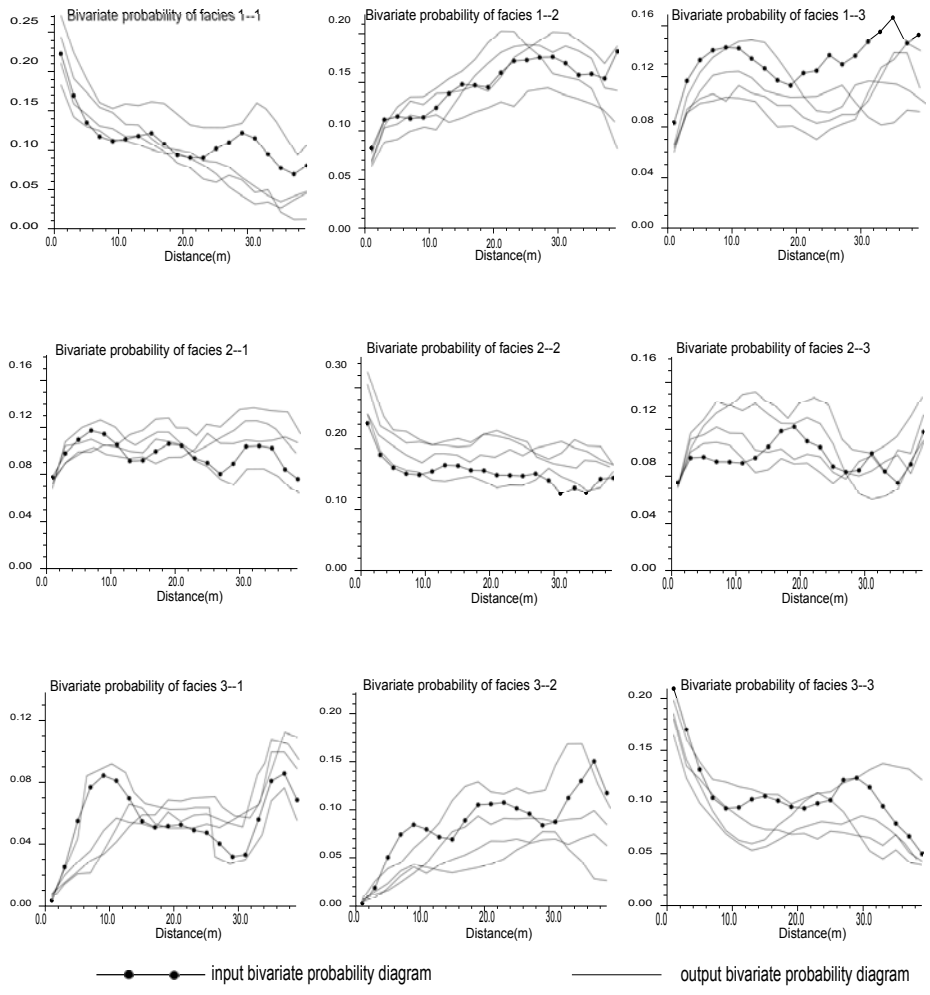


Figure 5.15: One example of bivariate probability diagram reproduction from different realizations

alizations. The calculated bivariate probability from the simulation output reflects the stacking pattern along certain directions and fluctuates around the input bivariate probability constraints. Theoretically, these fluctuations are the character of the multivariate probability for those grid together.

5.4 DMPE Facies Stochastic Modelling

5.4.1 Work flow of the DMPE

The cell-based spatial stochastic simulation will proceed along a random path through the grid cells in the model. At each grid cell, the DMPE is used to estimate the conditional probability, then Monte-Carlo simulation is used to assign a facies codes to the current location which will be used as conditioning data for the cells visited later in the path.

The procedure will include the following major works:

1. **Defining the facies:** The facies constituting the reservoir have to be defined from the available data: core samples, well logs and seismic data. They have to be consistent with the available geological information. In general, the geologists define more facies than can reasonably be simulated. Considering the probability space constraints, the facies types could be grouped into no more than four categories in order to use the DMPE within a reasonable computing time.

2. **Geological analysis and bivariate probability modelling:** A heterogeneity prototype will be obtained from the geological works such as sequence stratigraphic and sedimentological analysis. The dip, strike direction and the anisotropy ratios along these directions will be defined in this prototype. Based on the well sampled direction, usually will be the vertical direction, the bivariate probability diagram will be calculated.

3. **Define the simulation grid:** Generally, a regular orthogonal simulation grid

is adopted. 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 that is used to restore the geometry of the reservoir at the time of deposition.

4. **Perform simulation at each unsampled location:** The spatial relationships are characterized by the effective distance calculated from the dip and strike separation distance constrained from the geological conceptual model.

   (a) Look for the *n* closest conditioning data (original well data or previously simulated cell values) to the current unsampled location $\mathbf{u}_0$;

   (b) Based on the distance between every pair of locations, retrieve the bivariate probabilities from the experimental bivariate probability along the vertical direction;

   (c) Use the DMPE algorithm to estimate the multivariate probability;

   (d) Use the conditional probability definition to build the local cumulative conditional probability at $\mathbf{u}_0$;

   (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 locations are visited.

The main steps of using DMPE in facies modelling are shown in Figure 5.16. There are some differences compared with other traditional approach such as IK. First, the bivariate probability diagram is used as the spatial variability characterization tool. Inferring the bivariate probability diagram will require highly sampled data in the vertical direction. Also, in the spatial distance scaling process, the sequence stratigraphy and the geological understanding is integrated into the model.

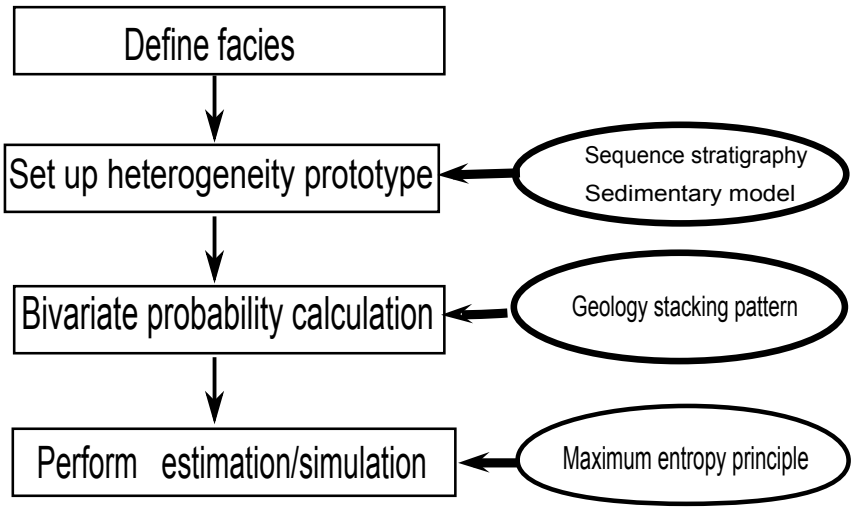


Figure 5.16: Main workflow of estimation/simulation with DMPE

The proposed DMPE follows a maximum entropy solution of the constraints from the surrounding hard data.

5.4.2 Program implementation

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

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 5.2 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 5.2, will provide a

Table 5.2: 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	...	...	...	...

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 (Deutsch and Journel, 1998). 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 5.17. The search will mainly pick the needed conditioning data evenly from the eight octant in 2D and 3D.

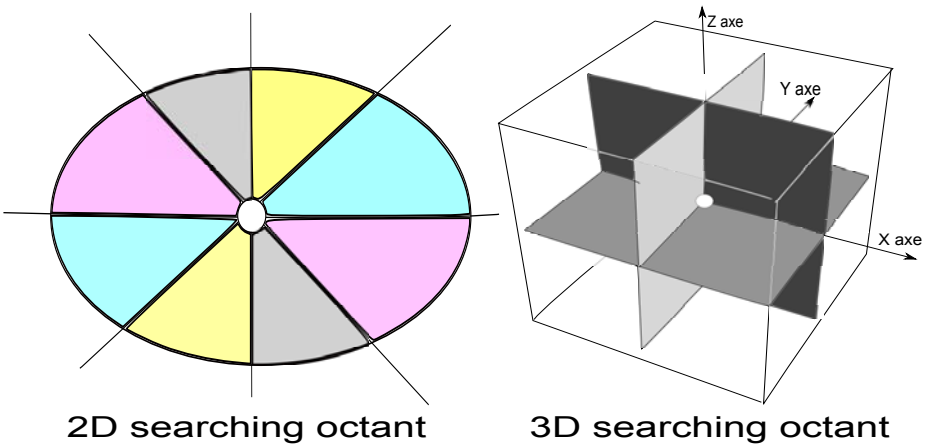


Figure 5.17: Octant searching in DMPE program implementation

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 (Tran, 1994). 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. As shown in Chapter 4, the estimated multivariate probability distribution is very close to the true distribution after 30 iterations. In practice, usually 20 to 25 iterations will permit convergence to the final solution and is enough for most situations.

5.5 Remarks

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.

In the multivariate probability estimation process, the estimated probability distribution should obey the minimum KL distance based on all the information. When it comes to the assessment to the final conditional probability distribution, which is calculated from the multivariate probability, it should have larger KL distance from its prior probability distribution. That is the main idea for the assessment of the posterior probability distribution using the KL distance.

Stochastic simulation will draw realizations from the multivariate probability distribution to characterize its randomness. The fluctuations of the bivariate probability constraints reproduction from the output realization and the differences of the spatial heterogeneity structure between different realizations actually are expected as an expression of our uncertainty in the geology reality.

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.

Chapter 6

Case study

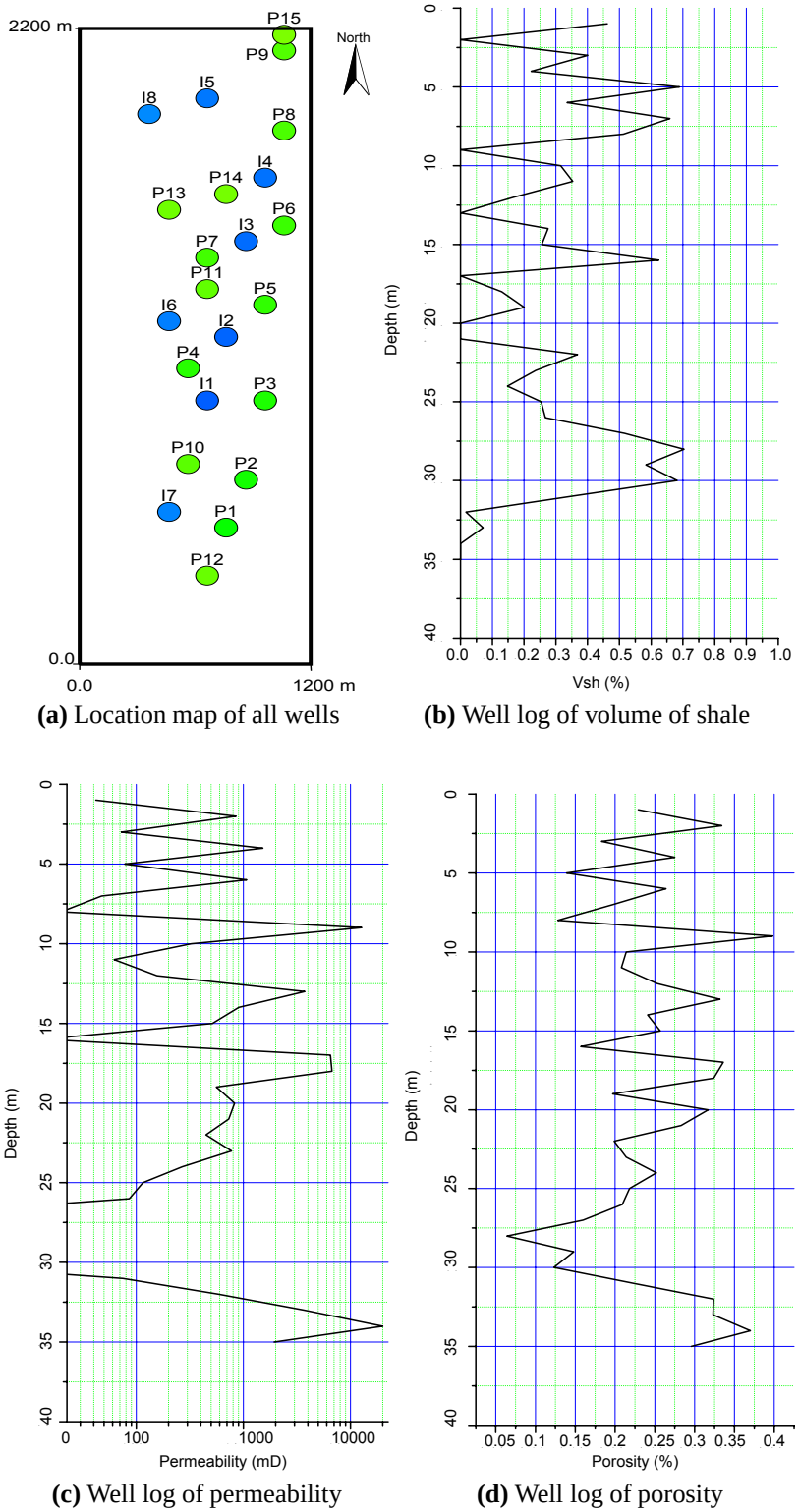
A case study is presented in this Chapter. Detecting the dip and strike directions from real data is illustrated. It is shown that the geological backgound of the data set will provide a conceptual heterogeneity model for the spatial anisotropy based distance calculation. In estimation and simulation, the new proposed anisotropy based spatial distance calculation and the direct multivariate probability estimation is implemented.

6.1 Data Set

The data set in this case study is from the Production forecasting with Uncertainty Quantification project (punq2). The well data have values of permeability, porosity and shale proportion. A total of 23 wells are available, see Figure 6.1.

The wells represent the Brent group from the North Sea basin. The upper part of the well data represents the Tarbert formation which is a prograding near shore sedimentary environment. The bottom part represents the upper Ness formation which has fluvial sediments. A short review of the Brent group is given below.

The Brent group comprises five lithostratigraphic units: the Broom, Rannoch, Etive, Ness and Tarbert formations (Richards, 1992). It is generally interpreted to record the progradation and subsequent transgression of a wavedominated delta (Brown, 1991; Helland-Hansen et al., 1992). The Rannoch and Etive formations record progradation of the wave-dominated delta front and coeval coastal barrier, while



(a) Location map of all wells

(b) Well log of volume of shale

(c) Well log of permeability

(d) Well log of porosity

Figure 6.1: All available wells location map and the well logs from the well P9

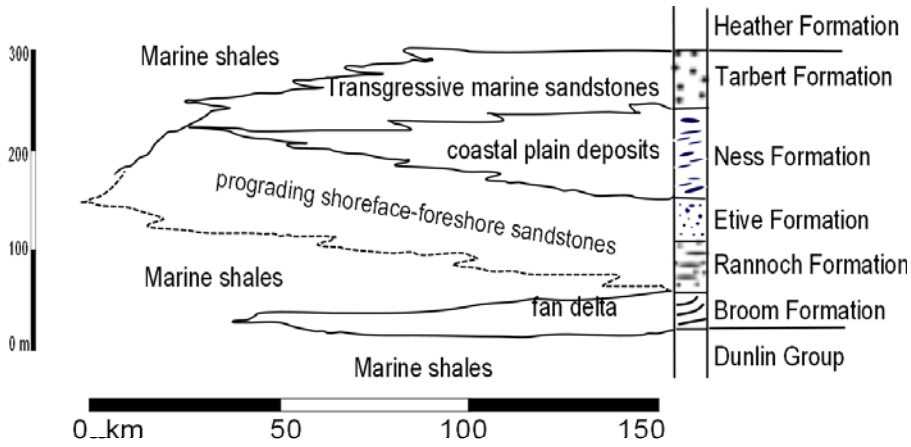


Figure 6.2: Schematic sequence stacking pattern of the Brent group in the North Sea basin (Richards, 1992)

the Ness formation comprises delta plain deposits. The Tarbert formation comprises transgressive shallow marine sandstones, see Figure 6.2.

The Brent group has formed a major exploration target in the North Sea since the discovery of the giant Brent and Ninian fields in the early seventies. As such, its stratigraphy and sedimentology have been the focus of continuous interest and analysis by a large number of geoscientists (Livera and Caline, 1990; Morris et al., 2003; Bullimore and Helland-Hansen, 2009).

6.2 Prototype Definition

In the proposed spatial anisotropy based distance calculation, building the geological prototype is a crucial step to ensure that the final facies model is reasonable and geologically realistic. It is built through geological exploration works based on the available data. In this case study, it will include the conceptual sedimentary model analysis, the facies definition and the modelling prototype coordinate definition.

6.2.1 Conceptual geological model

Based on the available data set, the Tarbert formation in the upper part of the Brent group will be modelled in this case study. The Tarbert formation is recognized by

the first appearance of shoreline sediments (delta front or shoreface foreshore) in the upper part of the Brent group, above the continental deposits of the Ness formation.

Generally, the Tarbert formation has an average thickness of 30 to 50 meters and comprises several upward-shallowing, weakly wave-influenced shoreface sandstone successions that are stacked vertically and contain evidence for tidal current activity. The base of the formation therefore represents a sequence boundary that has been transgressively reworked with little preservation of intervening lowstand deposits. Moreover, the formation underlies, sometimes unconformably, the marine shales which belongs to the Heather formation (Bullimore and Helland-Hansen, 2009).

Thus, the final geological model for the Tarbert formation is illustrated in Figure 6.3. In this conceptual model, each sand body of Tarbert formation would have an upcoarsing trend in the transgressive process. The bottom could be the sand from the Ness formation or the marine shale.

6.2.2 Facies definition

Before facies modelling, the facies types should be defined based on the geological background, well log data and other available data sources. If the DMPE method is used, the cell number in the model and the available CPU time should also be taking into consideration. Usually, three facies type is enough (Deutsch, 2002).

In this data set, there are three properties for each well: porosity, permeability and volume of shale. From the cross-plot between the volume of shale with the porosity and permeability shown in Figure 6.4, it can be seen they have a very high correlation coefficient. This is expected in such clastic reservoirs where the sand with a small percent of shale will have a higher porosity and permeability.

The volume of shale characterizes the sediments. It could be modelled directly as a continuous variable. While in this research, the facies will be constructed from the volume of shale first and used to build the facies model. The final permeability and porosity model will be constrained by the facies model.

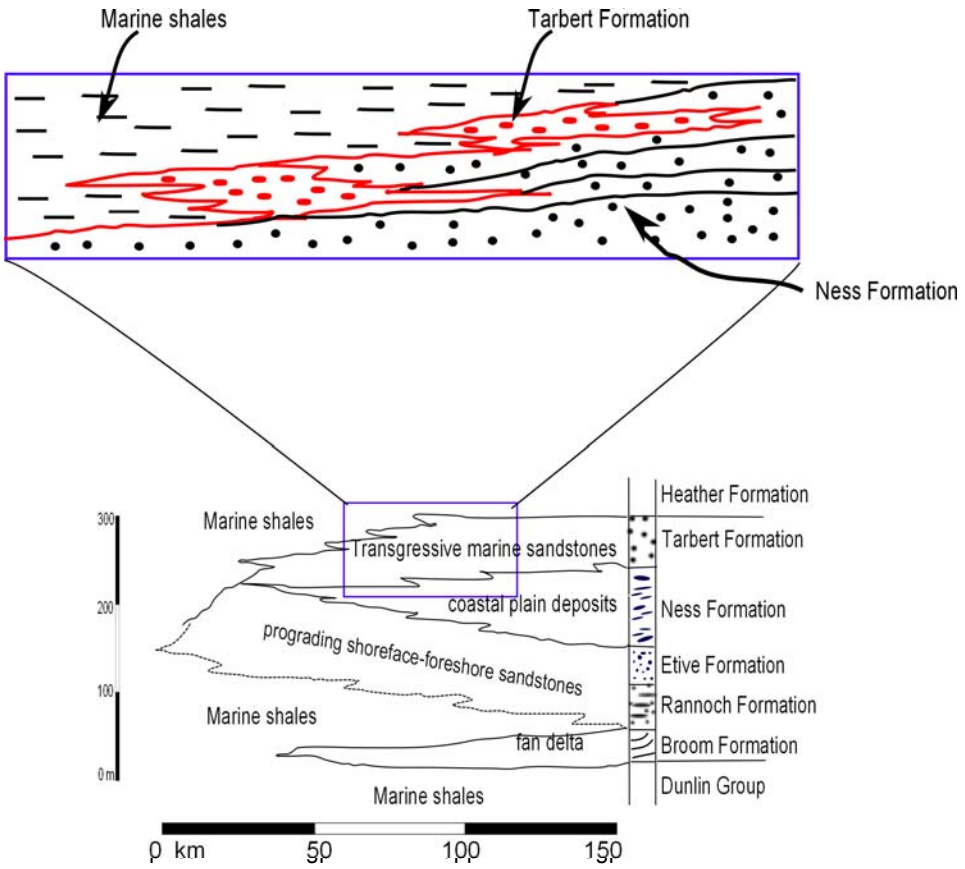


Figure 6.3: Conceptual geological model of Tarbert formation (Modified from Richards, 1992)

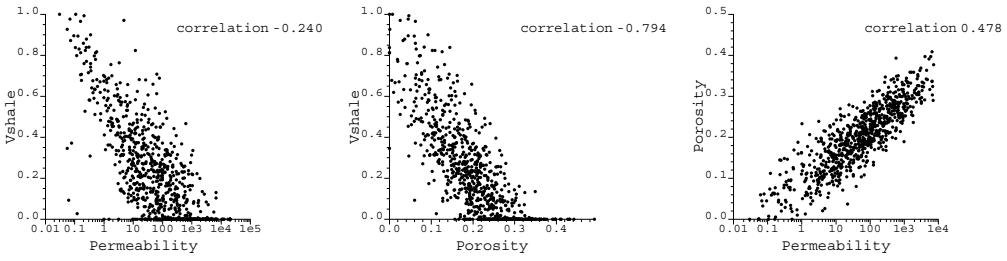


Figure 6.4: The correlation between of the volume of shale with the permeability and porosity

Generally, for a shoreface sedimentary environment, it is divided into upper shoreface, lower shoreface (MacDonald and Aasen, 1995; Reading, 1996). Upper Shoreface refers to the portion of the seafloor that is shallow enough to be agitated by everyday wave action (wave base). The continuous agitation of the sea floor in the upper shoreface environment filters the smallest grains leaving those grains heavy enough that the water cannot keep them suspended.

Lower Shoreface refers to the portion of the seafloor or sedimentary depositional environment that lies below everyday wave base. In this portion of the coastal environment, only the larger waves produced during storms have the power to agitate the sea bottom. Between storms, finer grained sediments accumulate on the seafloor.

Well logs such as spontaneous potential or gamma ray are usually used to define the facies. If core is available, it is important to understand the relationships between core, log facies and the nature of the depositional environments. Based on the above shoreface sedimentary characteristics, three facies types are defined from shale volume log using two arbitrary thresholds. As the Tarbert formation is a part of vertically stacked shoreface sandstone, the litho-facies type one which has a small proportion of shale (15%) can be interpreted as upper shoreface. While the one with volume of shale more than 45%, will be classified as shale. The value of between these two threshold will be lower shoreface as shown in Figure 6.6. Those two thresholds are used just for this study and are chosen based on the limited log data. One example of the facies vertical profile from shale volume log is shown in Figure 6.6.

6.2.3 Model grid definition

In the proposed spatial distance calculation approach, the dip and strike direction in the simulation domain will have a large impact on the final facies distribution. Thus, detecting and defining the dip and strike direction from the data set for subsequent geological facies simulation is an important step. The best way to do this is to

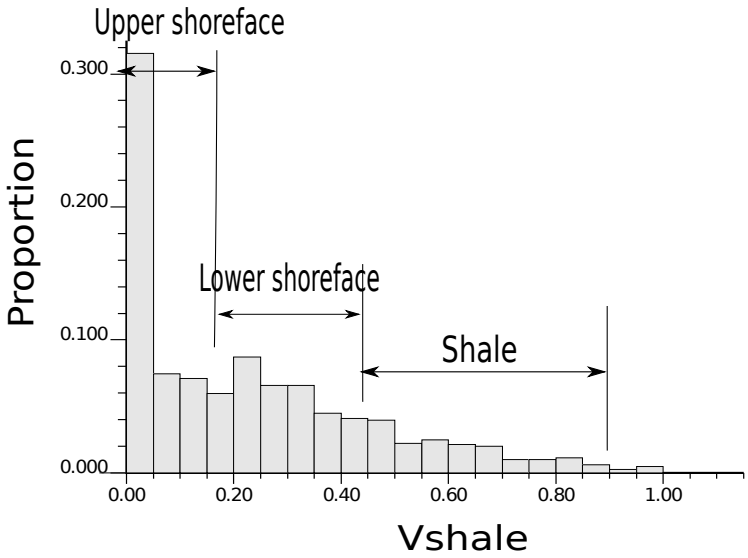


Figure 6.5: Shale volume histogram and the thresholds used to construct the facies

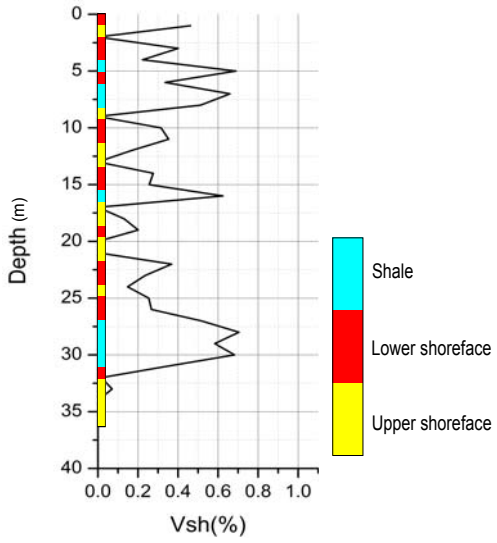


Figure 6.6: One example of vertical facies profile constructed from shale volume log

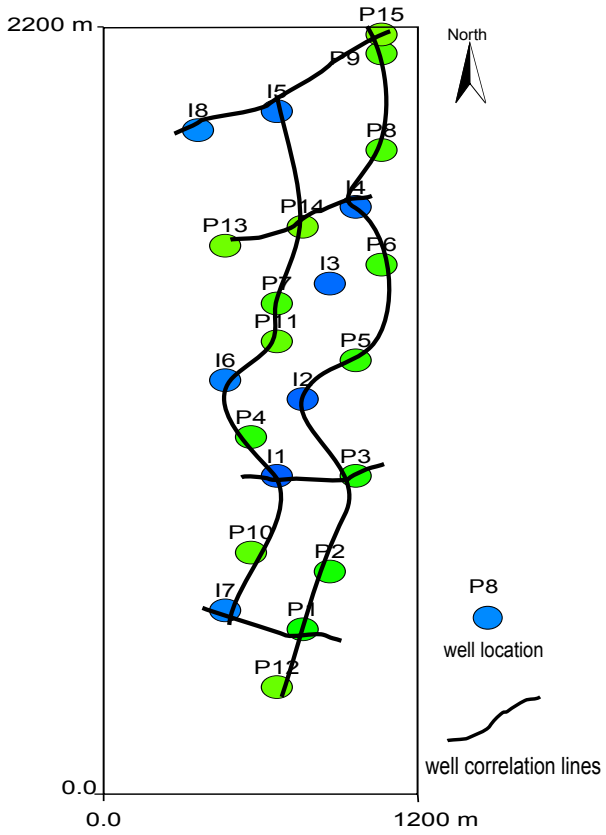


Figure 6.7: Well correlation sections

construct well correlations across the study area. In this research, a total of 6 well correlation sections are built from the well log data as shown in Figure 6.7.

As can be seen from two well correlation lines along the North-to-South direction in Figure 6.8, and from bottom to top, the facies stacking is in a pattern of Shale $\rightarrow$ Lower shoreface $\rightarrow$ Uppor shoreface. Although the proportion of each facies changes, the upwarding stacking pattern doesn't change. Based on the shoreline sedimentary model and the conceptual geological model in Figure 6.3, the direction along North-to-South will be the direction from proximal to distal axis in heterogeneity prototype.

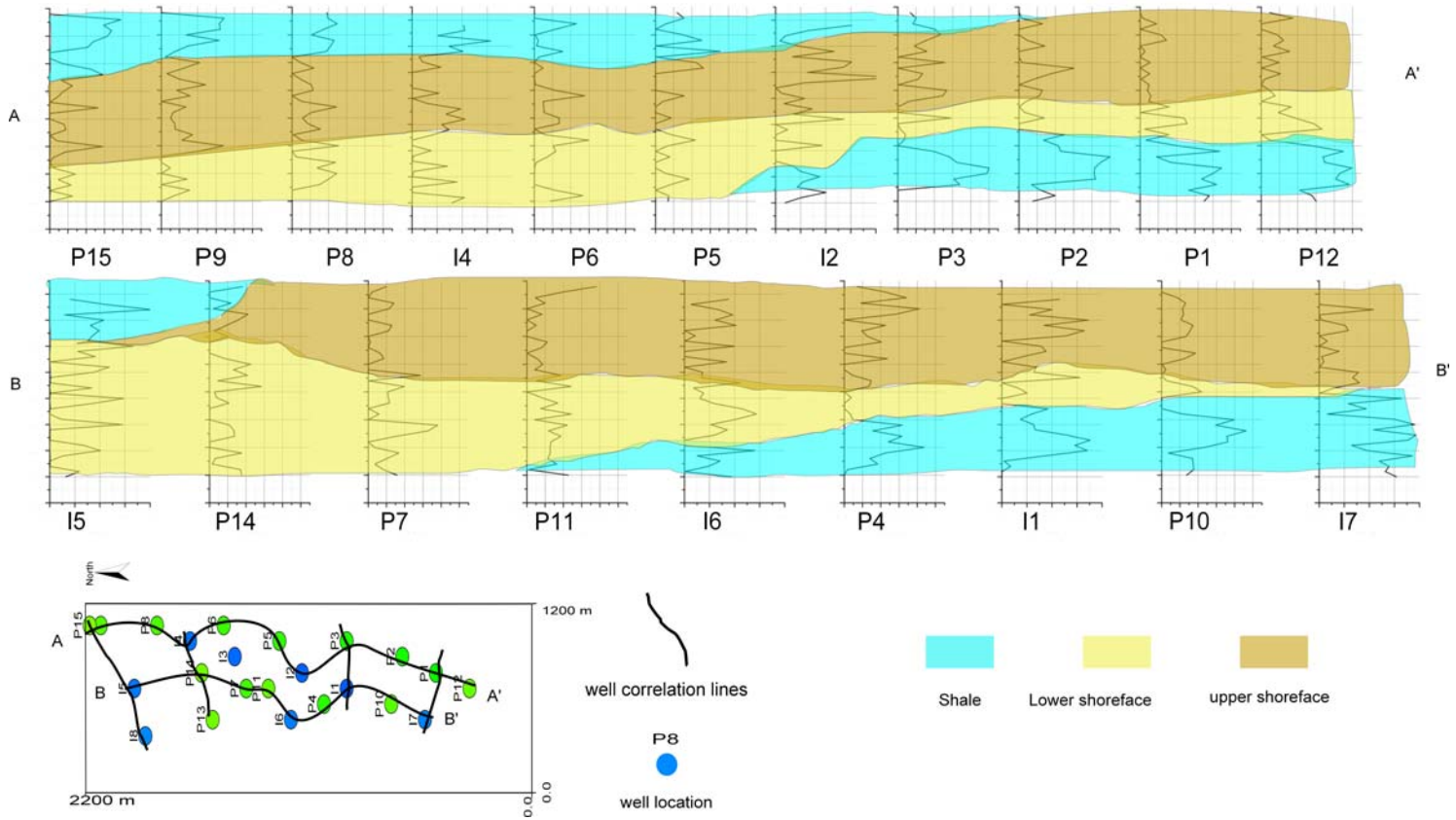


Figure 6.8: The well correlation along dip direction from the well data set

Four vertical well correlation sections perpendicular to the dip direction are shown in Figure 6.9. Although, the well stacking pattern along East-to-West direction will not change much along each line, their stacking patterns are different. Thus, the strike direction will be from East-to-West direction for this data set.

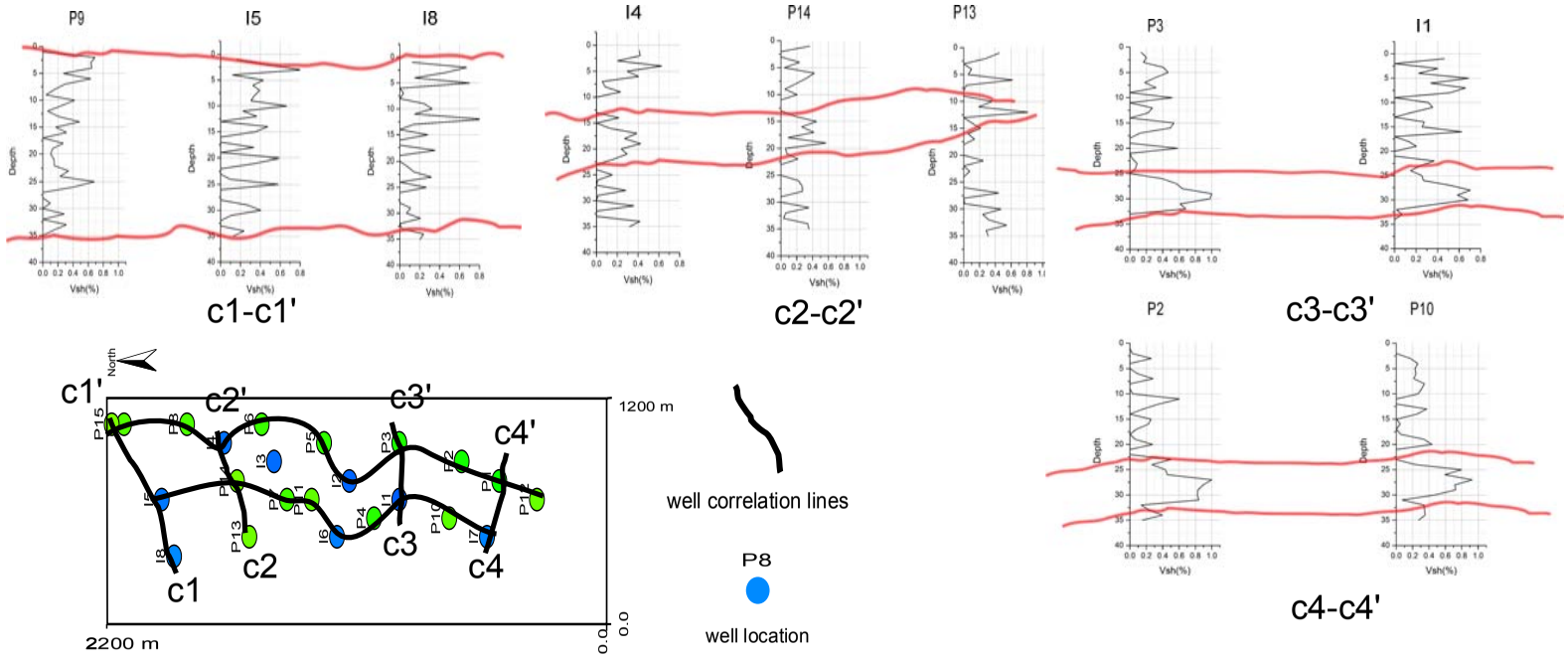


Figure 6.9: Three cross line along strike direction

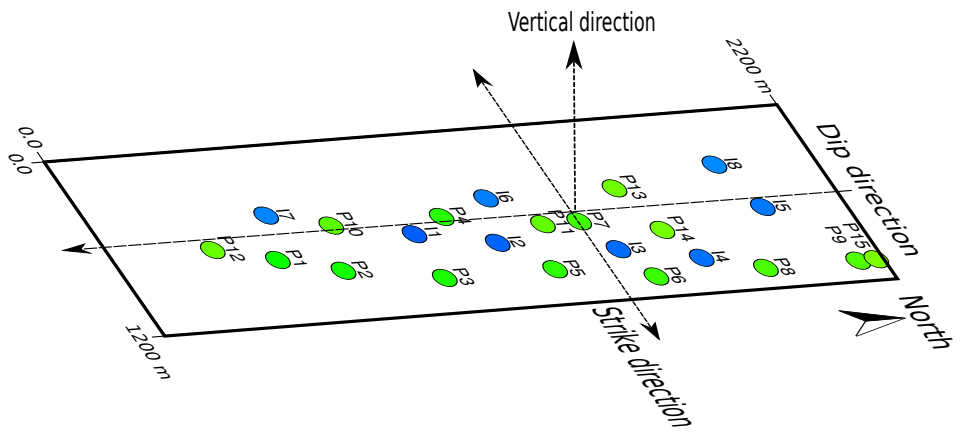


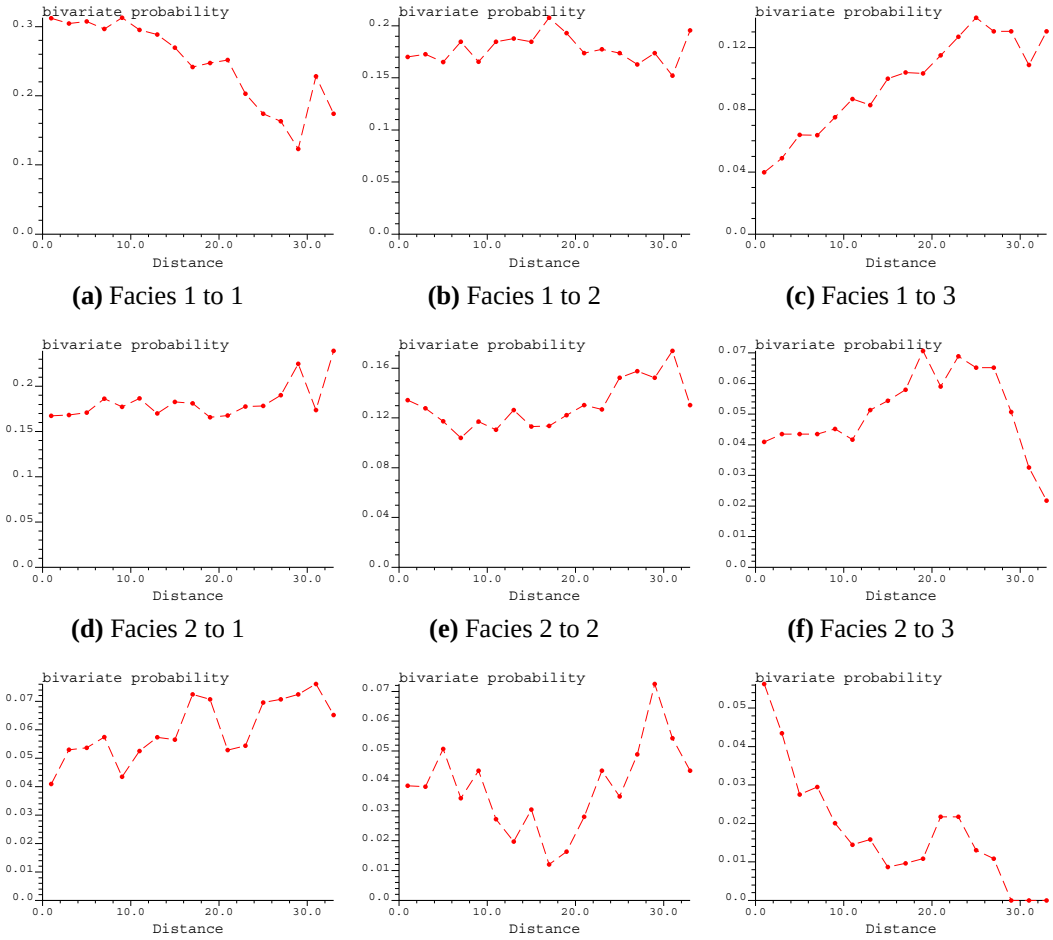
Figure 6.10: The heterogeneity prototype definition for the case study

Based on these well correlations and the geological background of the Tarbert formation shown in Figure 6.3, the dip and the strike directions of the prototype of the model will be from North-to-South and East-to-West respectively, as shown in Figure 6.10.

Another aspect of the conceptual model is the anisotropy ratio between the vertical and dip direction. For the Brent formation, the vertical-to-horizontal ratio will be close to 1:600 along dip direction and 1:5000 along the strike direction based on some studies on the North Sea basin (Brown, 1991). For the final geological model, the model dimensions are $1200 \times 2200 \times 35$ meter. The fine scale cell size is $20 \times 10 \times 1$ meters. Thus, the number of cells in each direction will be: $60 \times 220 \times 35$. The total number of cells for this model would be 462,000.

6.3 Facies Modelling

It has become a standard approach to split reservoir modelling into two steps: First generate the geometry of the facies and second, populate each facies with petrophysical properties such as porosity and permeability (Damsleth et al., 1992; Deutsch, 2002).



(a) Facies 1 to 1 **(b)** Facies 1 to 2 **(c)** Facies 1 to 3

(d) Facies 2 to 1 **(e)** Facies 2 to 2 **(f)** Facies 2 to 3

(g) Facies 3 to 1 **(h)** Facies 3 to 2 **(i)** Facies 3 to 3

Figure 6.11: Vertical bivariate probability diagram in the case study

6.3.1 Vertical bivariate probability diagram inference

For the proposed DMPE approach, the spatial heterogeneity variability is characterized by the bivariate probability diagram. Using the well log data from these 23 wells, the vertical bivariate probability diagram is calculated as shown in Figure 6.11.

These bivariate probability diagrams in Figure 6.11 reveal some geological information. For example, the mean lengths can be read from the direct bivariate probability for each facies. Upper shoreface facies have a mean length of 20 meters which is the longest length along vertical direction. While for shale, the mean

length is around 16. The lower shoreface has the shortest length of 8 meters.

By looking at the cross bivariate probability diagram, it also can be found that the upper shoreface to shale transitions are less frequent than the transition from upper shoreface to lower shoreface. This facies transition pattern is also supported from the well correlation sections as shown in Figure 6.8.

6.3.2 Spatial probability mapping

Although the results from the spatial probability mapping (estimation) will not be used in the final reservoir modelling, it is usful to check the computing environment for later stochastic simulation.

In this model, all 23 wells will be used as hard data for estimation. For each unsampled location, 8 conditioning data will be used as conditioning data.

The Figure 6.12 shows one slice along the dip and vertical direction of the estimation result. While along the strike and dip lateral direction, the estimation results are shown in Figure 6.13. Instead of using the traditional geometric distance calculation approach, the spatial distance is calculated from the proposed anisotropy distance calculation approach introduced in Chapter 3. The random switching function along the strike direction is assumed follow a sine wave function that is clearly reproduced in the estimation results as shown in Figure 6.13.

Of course, any other kind of strike switching function can be used. For example, a simple changing to its amplitude and angular frequency of the sine function along the strike direction will produce different estimation probability maps as shown in Figure 6.14.

6.3.3 Stochastic simulation

The sequential simulation algorithm is used to address the joint uncertainty of facies outcomes in the study area.

As shown in Figure 6.15, there are some small scale noise in the simulation results. Part of the reason is the small number of conditioning data. Only 8 con-

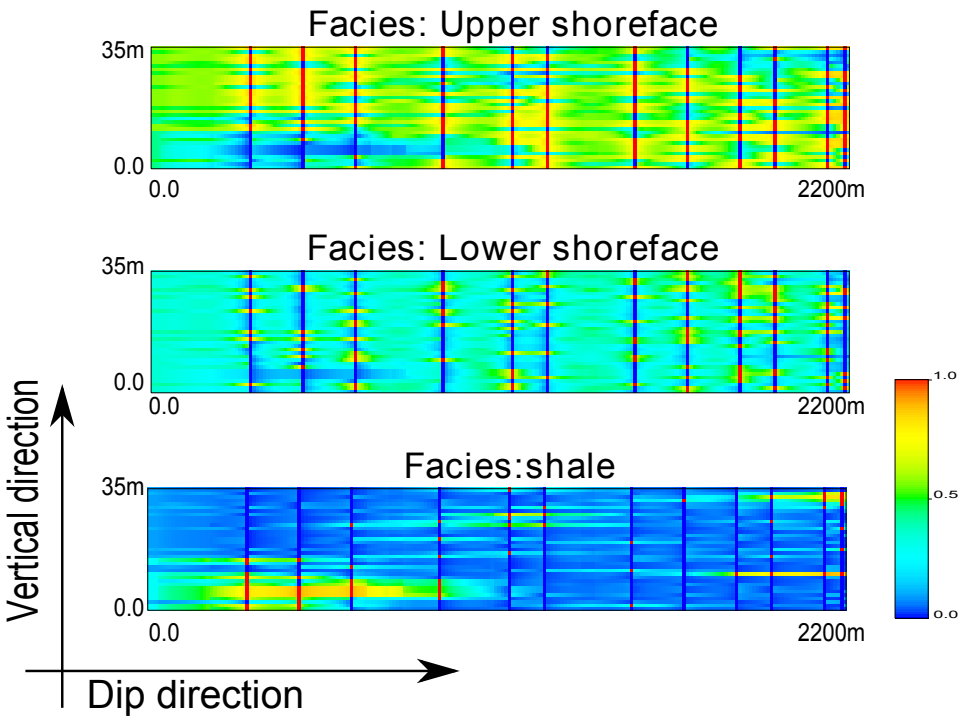


Figure 6.12: One slice along the vertical and dip direction of the estimation model for three facies using DMPE

ditioning data are used. In this case, the actual conditioning data used may change greatly for two very close locations.

As can be seen from some 2D slices along the XY direction in the model, shown in Figure 6.16, the facies distribution shows a kind of wave along the strike direction which is integrated into the model through the spatial distance calculation.

6.4 Remarks

The strike and dip directions can be discerned from real data sets. The stacking pattern for the vertical direction is dependent on the geological analysis to the study area. A detailed well correlation and the relative sequence stratigraphy research will aid a correct direction detecting for the facies modelling work.

The geological understanding will be integrated into the model through the proposed spatial distance calculation approach. Different geological understanding can

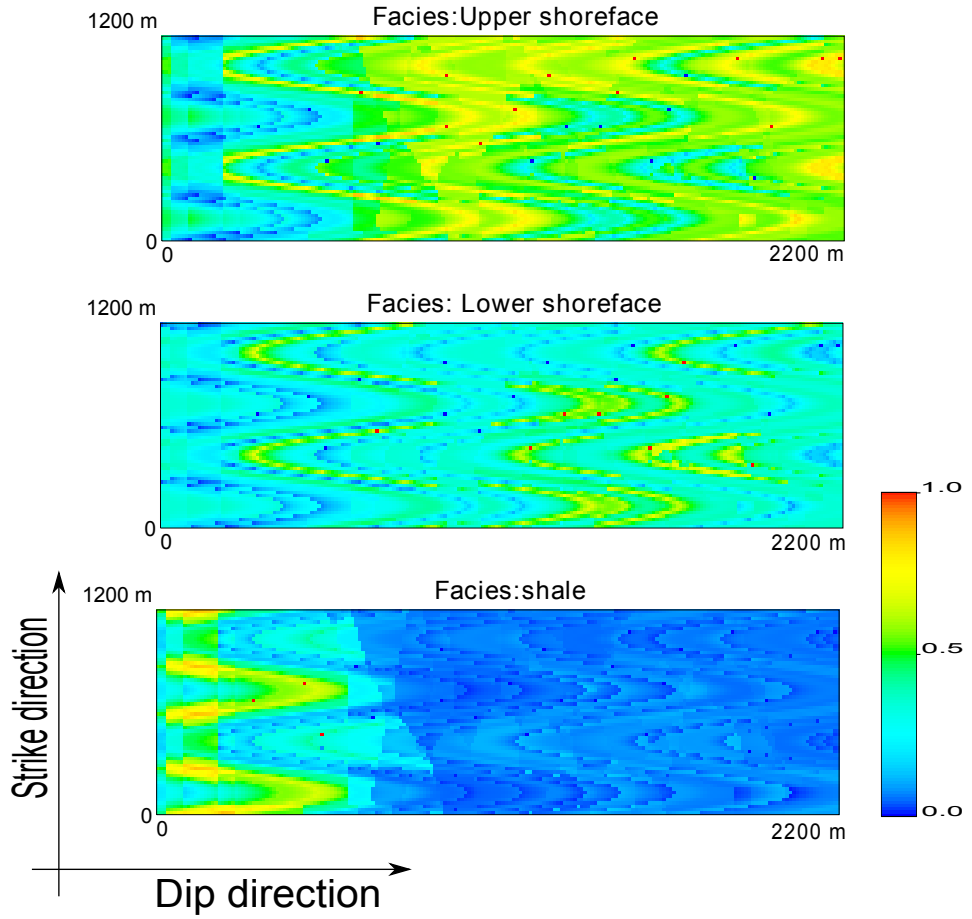


Figure 6.13: One slice along the strike and dip direction of the estimation model for three facies using DMPE

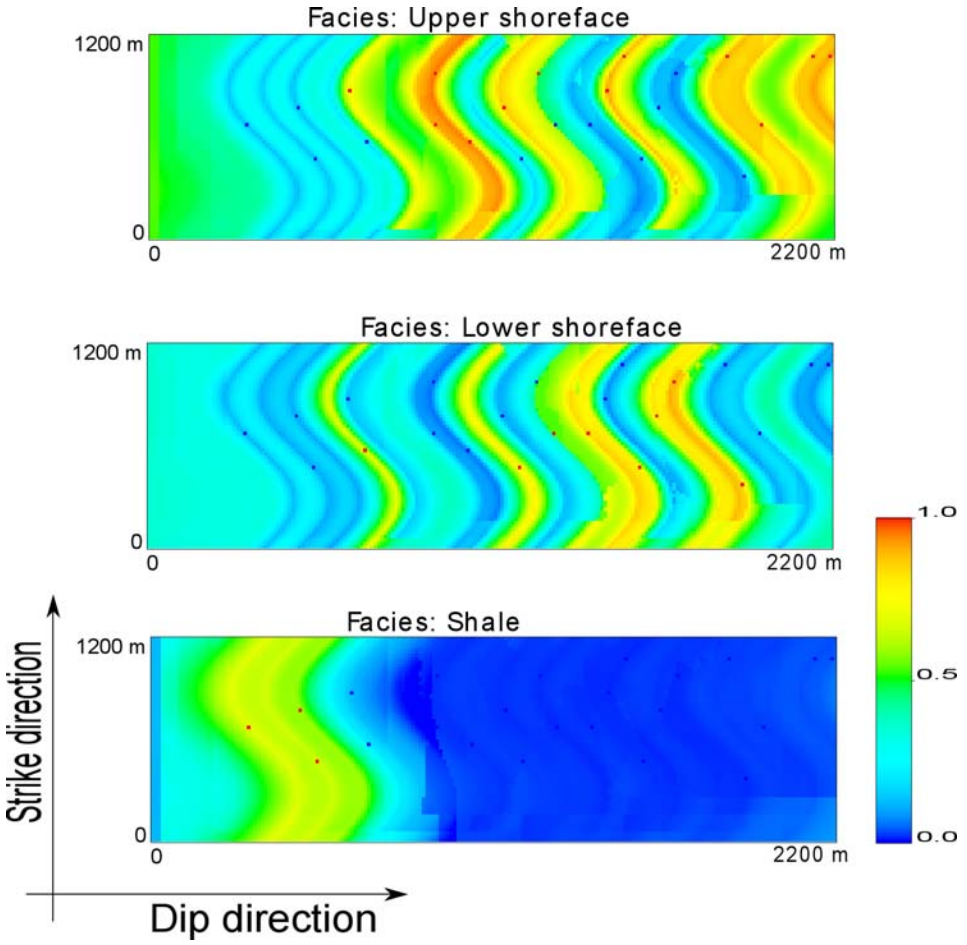


Figure 6.14: One slice of the estimation model with different random setting along strike direction

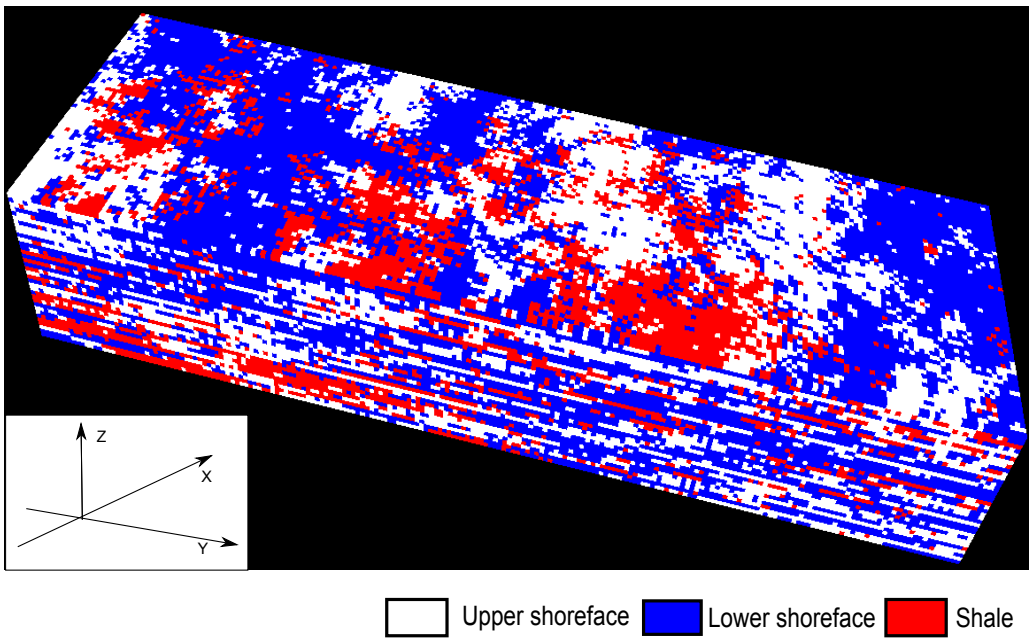


Figure 6.15: One 3D simulation output using the DMPE

be imposed into the final model through modifying the random function along the strike direction.

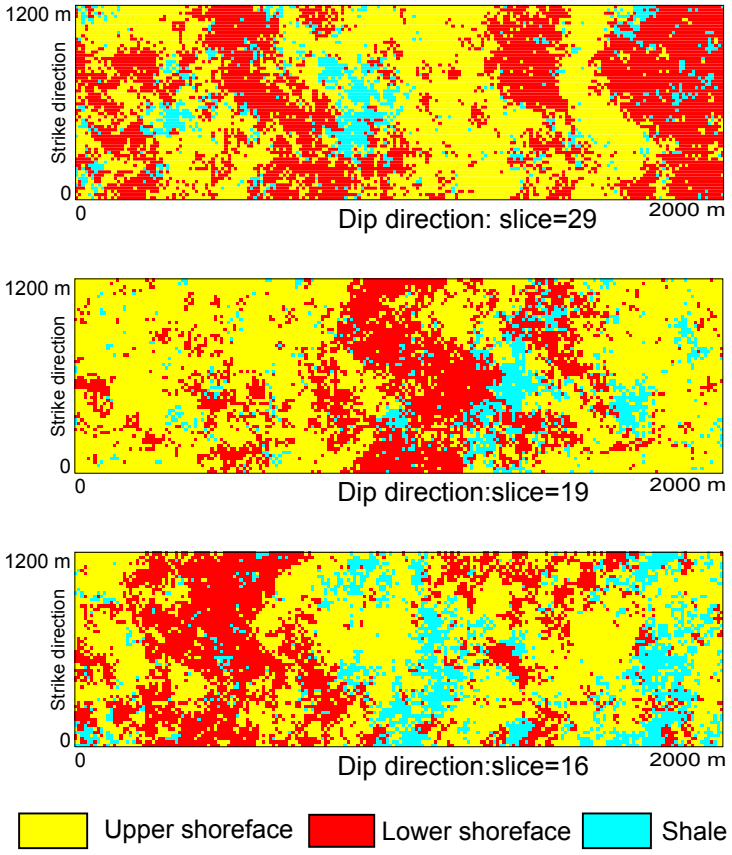


Figure 6.16: Three slices from the 3D simulation in the case study

Chapter 7

Final Remarks and Future Work

Valid reproduction of subsurface heterogeneity features is a fundamental requirement for quantitative geological models used in reservoir management. To meet this requirement, this dissertation proposes a stochastic facies modelling methodology that involves: a new geological characterization tool, a geological based spatial distance calculation approach, and a more theoretically correct conditional probability calculation.

7.1 Contributions

There are several contributions of this thesis. One is the heterogeneity spatial variability characterization. The bivariate probability diagram proposed in this research characterizes the facies spatial variations. Using the bivariate probability diagram instead of the variogram will make it possible to integrate more geological constraints such as the facies stacking pattern into the model.

Practically, there is a need to estimate the bivariate probability along any spatial directions to reflect the spatial anisotropy. Interpreting the lateral heterogeneity is always a challenge in geostatistics. In this thesis, a heterogeneity prototype is used to instruct the transformation of an effective spatial distance in the vertical direction. It provides a new approach to infer the lateral spatial statistics from available data. In this prototype, the vertical, strike and dip direction, are defined to reflect the different character of variation. The vertical direction will be the main

facies stacking direction and usually reflects the sediments deposit history. Along the strike direction, the sediments will show a kind of source shifting which is normal in clastic sediments environments. The dip direction will be the direction from sediment source to the deposition locations. Recognizing these three axes is possible in most sedimentary deposits as shown in the case study chapter of this thesis.

When the spatial variation information is characterized by the bivariate probability, the maximum entropy principle is used to combine all of them together to construct a multivariate probability distribution. In this approach, the bivariate probability between each data pair is considered as a marginal probability of the target multivariate probability which will characterize the facies outcomes probability at these locations taken all together.

After the multivariate probability is constructed, conditional probability can be calculated directly from its definition. This approach is theoretically correct without any other assumptions.

These form the bases of the proposed spatial probability interpolation methodology named Direct Multivariate Probability Estimation(DMPE). The whole procedure of sequential simulation using the DMPE will include the following steps:

1. Defining the 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. Considering the multivariate data event space constraints, there should be no more than four categories in order to use the DMPE within reasonable computing time.

2. Geological analysis and bivariate probability modelling. From the vertical log analysis, obtain a pattern for the sedimentary units that may be defined by a sequence stratigraphic surface. Geological analysis of the data set will correctly define the dip and strike direction of the heterogeneity prototype. The anisotropy ratios for the prototype will also aid in geologically realistic

modelling. The axes directions and anisotropy ratios will reflect the geologic understanding about the study area.

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. Implement traditional sequential simulation for each unsampled location. For each unsampled location, the spatial distance is expressed as an effective distance calculated from the dip and strike separation distance constrained from the heterogeneity prototype. The bivariate probability of each data pair is obtained from the bivariate probability calculated from the vertical direction. The conditional probability is calculated using the DMPE. After each cell is simulated, it will be used as a hard data for later cells simulation.

Finding the solution of the maximum entropy equation using the traditional Lagrange multiplier approach is a challenge to this multivariate probability estimation. Instead, an iterative scaling approach is used. It is based on the minimum Kullback-Leibler distance principle which is a more general maximum entropy approach. Although there are already some research, the contribution of this thesis is proposing an innovative sparse matrix construction approach for doing the lower order marginalization from the multivariate probability. Thus, in the iterative process, the marginalization proceeds quickly. The numerical implementation of the iterative scalling approach proposed in this thesis makes the iterative scaling technique more practical for geostatistical application. The above proposed DMPE facies modelling approach can be used in the situation that the geological patterns are

clear, such as the cyclicity along certain direction.

7.2 Future Work

The bivariate probability diagrams obtained at different lag distance are used directly in this research. The distance between them are interpolated with polynomials. One future research topic could be fitting it with a theoretical function.

The spatial distance calculation approach proposed in this research does not ensure that the covariance matrix is positive definite. Correcting the covariance matrix and ensuring the final solution is valid would be another area of future work. This kind of Non-Euclidean distance calculation approach could be used in some more general situation to reflect the geological constraints.

For the heterogeneity prototype proposed in this dissertation, the random function along strike direction is assumed to follow a sine wave function. Another research topic would be to use more geologically straight forward strike random function into the heterogeneity strike axis, such as adding some trend along strike direction.

The bivariate probabilities are used as constraints to the multivariate probability construction. Another subject for future work would be to consider higher order marginal probability constraints, such as the trivariate marginal probabilities in the multivariate probability construction. Also, there exist an opportunity to integrate a more informative prior probability for the starting point of iterative scaling.

The proposed stochastic simulation methodology is possible to be used with other spatial categorical variables such as the sequence stratigraphic bodies. Considering the sedimentary facies tract as a categorical variable, the stacking pattern can be characterized by the bivariate probability diagram. The multiple simulation results would capture the uncertainty of sequence stratigraphy.

Another area for future work would be computational improvement of DMPE. 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.

Bibliography

Aitken, J. F. and Flint, S. S. (1995). The application of high-resolution sequence stratigraphy to fluvial systems: a case study from the Upper Carboniferous Breathitt Group, Eastern Kentucky, USA. *Sedimentology*, 42(1):3–30.

Allen, J. R. L. (1965). Fining-upwards cycles in alluvial successions. *Geological Journal*, 4(2):229–246.

Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about Markov chains. *The Annals of Mathematical Statistics*, 28(1):89–110.

Arpat, G. B. and Caers, J. (2007). Conditional simulation with patterns. *Mathematical Geology*, 39(2):177–203.

Bardossy, G. and Fodor, J. (2001). Traditional and new ways to handle uncertainty in geology. *Natural Resources Research*, 10(3):179–187.

Bastante, F., Taboada, J., Alejano, L., and Ordez, C. (2005). Evaluation of the resources of a slate deposit using indicator kriging. *Engineering Geology*, 81(4):407 – 418.

Bertsekas, D. P. (1999). *Nonlinear Programming*. Athena Scientific.

Bierkens, M. F. P. and Burrough, P. A. (1993). The indicator approach to categorical soil data. *Journal of Soil Science*, 44(2):369–381.

Bishop, Y. M. M. (1969). Full contingency tables, logits, and split contingency tables. *Biometrics*, 25(2):383–399.

Bogaert, P. (2002). Spatial prediction of categorical variables: the bayesian maximum entropy approach. *Stochastic Environmental Research and Risk Assessment*, 16(6):425–448.

Brown, S. (1991). Stratigraphy of the oil and gas reservoirs: UK Continental Shelf. *Geological Society, London, Memoirs*, 14(1):9–18.

Bullimore, S. A. and Helland-Hansen, W. (2009). Trajectory analysis of the lower Brent Group (Jurassic), northern North Sea: contrasting depositional patterns during the advance of a major deltaic system. *Basin Research*, 21(5):559–572.

Carle, S. and Fogg, G. (1996). Transition probability-based indicator geostatistics. *Mathematical Geology*, 28(4):453–476.

Carle, S. and Fogg, G. (1997). Modeling spatial variability with one and multidimensional continuous-lag Markov chains. *Mathematical Geology*, 29(7):891–918.

Carle, S. F., Esser, B. K., and Moran, J. E. (2006). High-resolution simulation of basin-scale nitrate transport considering aquifer system heterogeneity. *Geosphere*, 2(4):195–209.

Carr, T. R. (1982). Log-linear models, markov-chains and cyclic sedimentation. *Journal of Sedimentary Petrology*, 52(3):905–912.

Catuneanu, O. (2002). Sequence stratigraphy of clastic systems: concepts, merits, and pitfalls. *Journal of African Earth Sciences*, 35(1):1–43.

Catuneanu, O. (2006). *Principles of Sequence Stratigraphy*. Elsevier, Amsterdam.

Catuneanu, O., Abreu, V., Bhattacharya, J., Blum, M., Dalrymple, R., Eriksson, P., ..., and Winker, C. (2009). Towards the standardization of sequence stratigraphy. *Earth-Science Reviews*, 92(1-2):1–33.

Cheeseman, P. (1983). A method of computing generalized bayesian probability values for expert systems. In *Proceedings of the Eighth international joint conference on Artificial intelligence*, volume 1, pages 198–202, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Chiles, J. P. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, New York.

Chow, C. and Liu, C. (1968). Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, 14(3):462–467.

Cojan, I., Fouche, O., Lopez, S., and Rivoirard, J. (2005). Process-based reservoir modelling in the example of meandering channel. In Leuangthong, O. and Deutsch, C. V., editors, *Geostatistics Banff 2004*, volume 14 of *Quantitative Geology and Geostatistics*, pages 611–619. Springer Netherlands.

Coleman, J. M., Roberts, H. H., and Stone, G. W. (1998). Mississippi river delta: An overview. *Journal of Coastal Research*, 14(3):699–716.

Coudert, L., Frappa, M., and Arias, R. (1994). A statistical method for litho-facies identification. *Journal of Applied Geophysics*, 32(2-3):257 – 267.

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. John Wiley & Sons, New York.

Cressie, N. (1990). The origins of kriging. *Mathematical Geology*, 22(3):239–252.

Cressie, N. (1991). *Statistics for Spatial Data*. John Wiley & Sons, New York.

Cressie, N. and Hawkins, D. (1980). Robust estimation of the variogram. *Mathematical Geology*, 12(2):115–126.

Cross, T. and Baker, M. (1992). Application of high-resolution sequence stratigraphy to reservoir analysis. In Eschard, R. and Doligez, B., editors, *Subsurface Reservoir Characterization From Outcrop Observations*.

Dacey, M. and Krumbein, W. (1970). Markovian models in stratigraphic analysis. *Mathematical Geology*, 2(2):175–191.

Damsleth, E., Tjolsen, C. B., Omre, H., and Haldorsen, H. H. (1992). A two-stage stochastic model applied to a North Sea reservoir. *Journal of Petroleum Technology*, 44(4):402–408,486.

Darroch, J. N. and Ratcliff, D. (1972). Generalized iterative scaling for log-linear models. *The Annals of Mathematical Statistics*, 43(5):1470–1480.

Davis, B. (1987). Uses and abuses of cross validation in geostatistics. *Mathematical Geology*, 19(3):241–248.

Deming, W. E. and Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, 11(4):427–444.

Deutsch, C. V. (1992). *Annealing Techniques Applied to Reservoir Modeling and the Integration of Geological and Engineering (Well Test) Data*. PhD thesis, Stanford University, Stanford, CA.

Deutsch, C. V. (2002). *Geostatistical Reservoir modeling*. Oxford University Press.

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

D'Or, D. and Bogaert, P. (2003). Continuous-valued map reconstruction with the bayesian maximum entropy. *Geoderma*, 112(3-4):169 – 178. ¡ce:title¿Pedometrics 2001¡/ce:title¿.

D'Or, D. and Bogaert, P. (2004). Spatial prediction of categorical variables with the bayesian maximum entropy approach: the ooypolder case study. *European Journal of Soil Science*, 55(4):763–775.

D'Or, D., Bogaert, P., and Christakos, G. (2001). Application of the bme approach to soil texture mapping. *Stochastic Environmental Research and Risk Assessment*, 15(1):87–100.

Doveton, J. H. (1995). Theroy and application of vertical variability measuresfrom markov chain analysis. In Yarus, J. M. and Chambers, R. L., editors, *Stochastic Modeling and Geostatistics: Principles, Methods, and Case Studies*, pages 55–64. AAPG Computer Applications in Geology, No. 3.

Drake, T. G. and Calantoni, J. (2001). Discrete particle model for sheet flow sediment transport in the nearshore. *Journal of Geophysical Research*, 106(C9):19859–19868.

Driese, S. G. and Dott, R. H. (1984). Model for sandstone-carbonate "cyclothems" based on upper member of morgan formation (middle pennsylvanian) of northern Utah and Colorado. *Bulletin of the American Association of Petroleum Geologists*, 68(5):574–597.

Ellis, D. V. and Singer, J. M. (2008). *Well Logging for Earth Scientists*. Springer-Verlag, 2nd edition.

Enge, H. D., Buckley, S. J., Rotevatn, A., and Howell, J. A. (2007). From outcrop to reservoir simulation model: workflow and procedures. *Geosphere*, 3(6):469–490.

Ezekwe, N. (2010). *Petroleum Reservoir Engineering Practice*. Prentice Hall.

Fanchi, J. R. (2001). *Principle of Applied Reservoir Simulation*. Gulf Professional Publishing.

Ghibaudo, G. (1992). Subaqueous sediment gravity flow deposits: practical criteria for their field description and classification. *Sedimentology*, 39(3):423–454.

Gibbs, J. (1873). A method of geometrical representation of thermodynamic properties of substances by means of surfaces. *Transactions of the Connecticut Academy*, 2:382–404.

Glaister, R. P. and Nelson, H. W. (1974). Grain-size distributions, an aid in facies identification. *Bulletin of Canadian Petroleum Geology*, 22(3):203–240.

Gokhale, D. V. (1972). Analysis of log-linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(3):371–376.

Gokhale, D. V. (1973). Approximating disrete distributions, with applications. *Journal of the American Statistical Association*, 68(344):1009–1012.

Gomez-Hernandez, J. J. and Journel, A. (1993). Joint sequential simulation of multigaussian fields. In Soares, A., editor, *Geostatistics Troia 1992*, volume 1, pages 85–94. Kluwer Academic Publishers.

Gomez-Hernandez, J. J. and Srivastava, R. M. (1990). ISIM3D: An ANSI-C three dimensional multiple indicator conditional simulation program. *Computers & Geosciences*, 16(4):395–410.

Gomez-Hernandez, J. J. and Wen, X.-H. (1998). To be or not to be multi-gaussian? a reflection on stochastic hydrogeology. *Advances in Water Resources*, 21(1):47 – 61.

Good, I. J. (1963). Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *The Annals of Mathematical Statistics*, 34(3):911–934.

Goovaerts, P. (2001). Geostatistical modelling of uncertainty in soil science. *Geoderma*, 103(1-2):3–26.

Gringarten, E. and Deutsch, C. V. (2001). Teacher's aide variogram interpretation and modeling. *Mathematical Geology*, 33(4):507–534.

Guardiano, F. and Srivastava, R. M. (1993). Multivariate geostatistics: Beyond bivariate moments. In Soares, A., editor, *Geostatistics Troia 1992*, volume 1, pages 133–144. Kluwer.

Haggstrom, O. (2002). *Finite Markov Chains and Algorithmic Applications*. Cambridge University Press.

Haldorsen, H. and Damsleth, E. (1990). Stochastic modeling. *Journal of Petroleum Technology*, 42(4):404–417.

Harbaugh, J. W. and Bonham-Carter, G. (1970). *Computer Simulation in Geology*. John Wiley & Sons.

Hatløy, A. S. (1995). Numerical facies modeling combining deterministic and stochastic method. In Yarus, J. M. and Chambers, R. L., editors, *Stochastic Modeling and Geostatistics: Principles, Methods, and Case Studies*, pages 109–120. AAPG Computer Applications in Geology, No. 3.

Hattori, I. (1976). Entropy in markov chains and discrimination of cyclic patterns in lithologic successions. *Mathematical Geology*, 8(4):477–497.

Helland-Hansen, W., Ashton, M., Lmo, L., and Steel, R. (1992). Advance and retreat of the brent delta: recent contributions to the depositional model. *Geological Society, London, Special Publications*, 61(1):109–127.

Hiscott, R. N. (1981). Chi-square tests for markov chain analysis. *Mathematical Geology*, 13(1):69–80.

Holden, L., Hauge, R., Skare, ø., and Skorstad, A. (1998). Modeling of fluvial reservoirs with object models. *Mathematical Geology*, 30(5):473–496.

Hutton, E. W. H. and Syvitski, J. P. M. (2008). Sedflux 2.0: An advanced process-response model that generates three-dimensional stratigraphy. *Computers & Geosciences*, 34(10):1319–1337.

Ireland, C. T. and Kullback, S. (1968). Minimum discrimination information estimation. *Biometrics*, 24(3):707–713.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, 106(6):620–630.

Jaynes, E. T. (1978). Where do we stand on maximum entropy? In Levine, R. D. and Tribus, M., editors, *The maximum entropy formalism*. The MIT Press.

Johnson, R. A. and Wichern, D. W. (1982). *Applied Multivariate Statistical Analysis*. Prentice Hall, Englewood Cliffs, NJ.

Journel, A. G. (1983). Nonparametric estimation of spatial distributions. *Mathematical Geology*, 15(3):445 – 468.

Journel, A. G. (1986). Geostatistics: Models and tools for the earth sciences. *Mathematical Geology*, 18(1):119–140.

Journel, A. G. (1999). Markov models for cross-covariances. *Mathematical Geology*, 31(8):955–964.

Journel, A. G. and Huijbregts, C. J. (1978). *Mining Geostatistics*. Academic Press, New York.

Journel, A. G. and Isaaks, E. H. (1984). Conditional indicator simulation: Application to a Saskatchewan uranium deposit. *Mathematical Geology*, 16(7):685–718.

Journel, A. G. and Posa, D. (1990). Characteristic behavior and order relations forindicator variograms. *Mathematical Geology*, 22(8):1011–1025.

Karssenberg, D., Tornqvist, T. E., and Bridge, J. S. (2001). Conditioning a process-based model of sedimentary architecture to well data. *Journal of Sedimentary Research*, 71(6):868–879.

Koltermann, C. E. and Gorelick, S. M. (1996). Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches. *Water Resources Research*, 32(9):2617–2658.

Krumbein, W. C. and Dacey, M. F. (1969). Markov chains and embedded Markov chains in geology. *Mathematical Geology*, 1(1):79–96.

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

Ku, H. H. and Kullback, S. (1974). Loglinear models in contingency table analysis. *The American Statistician*, 28(4):115–122.

Kullback, S. (1968). *Information Theory and Statistics*. Dover Publications.

Kullback, S. and Khairat, M. (1966). A note on minimum discrimination information. *The Annals of Mathematical Statistics*, 37(1):279–280.

Lantuejoul, C. (2001). *Geostatistical Simulation: Models and Algorithms*. Springer.

Lark, R. M. and Beckett, P. H. T. (1998). A geostatistical descriptor of the spatial distribution of soil classes, and its use in predicting the purity of possible soil map units. *Geoderma*, 83(3-4):243 – 267.

Li, W. (2006). Transiogram: A spatial relationship measure for categorical data. *International Journal of Geographical Information Science*, 20(6):693–699.

Livera, S. E. and Caline, B. (1990). The sedimentology of the Brent Group in the Cormorant Block IV oilfield. *Journal of Petroleum Geology*, 13(4):367–396.

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

Mackey, S. D. and Bridge, J. S. (1995). Three-dimensional model of alluvial stratigraphy; theory and applications. *Journal of Sedimentary Research*, 65(1b):7–31.

Maharaja, A. (2008). Tigenerator: Object-based training image generator. *Computers & Geosciences*, 34(12):1753–1761.

Marinoni, O. (2003). Improving geological models using a combined ordinary-indicator kriging approach. *Engineering Geology*, 69(1-2):37 – 45.

Marr, J. G., Swenson, J. B., Paola, C., and Voller, V. R. (2000). A two-diffusion model of fluvial stratigraphy in closed depositional basins. *Basin Research*, 12(3-4):381–398.

McLaughlin, P. P. (2005). Sequence stratigraphy. In Selley, R. C., Cocks, L. R. M., , and Plimer, I. R., editors, *Encyclopedia of Geology*, pages 159 – 173. Elsevier, Oxford.

Mead, L. R. and Papanicolaou, N. (1984). Maximum entropy in the problem of moments. *Journal of Mathematical Physics*, 25(8):2404–2417.

Meyn, S. P. and Tweedie, R. (2008). *Markov Chains and Stochastic Stability*. Cambridge University Press.

Miall, A. D. (1973). Markov chain analysis applied to an ancient alluvial plain succession. *Sedimentology*, 20(3):347–364.

Miall, A. D. (1996). *The Geology of Fluvial Deposits*. Springer-Verlag, New York.

Middleton, G. V. (1973). Johannes Walther's law of the correlation of facies. *Geological Society of America Bulletin*, 84(3):979–988.

Mirowski, P., Tetzlaff, D., Davies, R., McCormick, D., Williams, N., and Signer, C. (2009). Stationarity scores on training images for multipoint geostatistics. *Mathematical Geosciences*, 41(4):447–474.

Morris, J., Hampson, G. J., and Maxwell, G. (2003). Controls on facies architecture in the Brent Group, Strathspey Field, UK North Sea: implications for reservoir characterization. *Petroleum Geoscience*, 9(3):209–220.

Mosegaard, K. and Tarantola, A. (1995). Monte carlo sampling of solutions to inverse problems. *Journal of Geophysical Research*, 100(B7):12431–12447.

Newman, W. (1979). Extension to the maximum entropy method II. *IEEE Transactions on Information Theory*, 25(6):705–708.

Oliver, D., Cunha, L., and Reynolds, A. (1997). Markov chain Monte Carlo methods for conditioning a permeability field to pressure data. *Mathematical Geology*, 29(1):61–91.

Papoulis, A. (1984). *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill Inc., New York, 2nd edition.

Payton, C. E. (1977). *Seismic stratigraphy : applications to hydrocarbon exploration*. American Association of Petroleum Geologists Memoir.

Philip, G. M. and Watson, D. F. (1987). Probabilism in geological data analysis. *Geological Magazine*, 124(6):577–583.

Picard, R. R. and Cook, R. D. (1984). Cross-validation of regression models. *Journal of the American Statistical Association*, 79(387):575–583.

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

Plint, A. G. (1995). *Sedimentary Facies Analysis:a tribute to the research and teaching of Harold G. Reading*. Wiley-Blackwell.

punq2. http://www2.nr.no/research/sand/punq/.

Pyrcz, M. J., Gringarten, E., Frykman, P., and Deutsch, C. V. (2006). Representative input parameters for geostatistical simulation. In Coburn, C., Yarus, J. M., and Chambers, R. L., editors, *Stochastic modeling and geostatistics: Principles, methods, and case studies*, volume 2 of *AAPG Computer Applications in Geology*, pages 123–137.

Reading, H. G. (1996). *Sedimentary Environments: Processes, Facies and Stratigraphy*. Blackwell Publishing.

Richards, P. C. (1992). An introduction to the brent group: a literature review. *Geological Society, London, Special Publications*, 61(1):15–26.

Roberts, H. H. (1997). Dynamic changes of the holocene mississippi river delta plain: The delta cycle. *Journal of Coastal Research*, 13(3):605–627.

Roksandic, M. M. (1978). Seismic facies analysis concepts. *Geophysical Prospecting*, 26(2):383–398.

Shannon, C. (1948). A mathematical theory of communication. *Mobile Computing and Commmunications Review*, 5(1):379–423.

Shmaryan, L. and Journel, A. (1999). Two markov models and their application. *Mathematical Geology*, 31(8):965–988.

Soares, A. (2001). Direct sequential simulation and cosimulation. *Mathematical Geology*, 33(8):911–926.

Soreghan, G. S. (1997). Walther's law, climate change, and upper paleozoic cyclostratigraphy in the ancestral rocky mountains. *Journal of Sedimentary Research*, 67(6):1001–1004.

Stoer, J., Bulirsch, and Roland (2002). *Introduction to Numerical Analysis(3rd ed.)*. Springer-Verlag.

Strebelle, S. (2000). *Sequential simulation drawing structures from training images*. PhD thesis, Stanford University.

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

Tetzlaff, D. (1990). Limits to the predictive ability of dynamic models that simulate clastic sedimentation. In Cross, T. A., editor, *Quantitative Dynamic Stratigraphy*, pages 55–65. Prentice Hall, Englewood Cliffs.

Tetzlaff, D. and Harbaugh, J. (1989). *Simulating clastic sedimentation*. Van Nostrand Reinhold Co. Inc.

Tran, T. T. (1994). Improving variogram reproduction on dense simulation grids. *Computers & Geosciences*, 20(7-8):1161–1168.

Tribus, M. (1961). *Thermodynamics and thermostatics*. D. Van Nostrand, New York.

Tsynkov, S. V. and Vatsa, V. N. (1998). Improved treatment of external boundary conditions for three-dimensional flow computations. *AIAA journal*, 36:1998–2004.

Ulrych, T., Bassrei, A., and Lane, M. (1990). Minimum relative entropy inversion of 1d data with applications. *Geophysical Prospecting*, 38(5):465–487.

Veeken, P. C. (2007). *Seismic Stratigraphy, Basin Analysis and Reservoir Characterisation*. Elsevier.

Verly, G. (1983). The multigaussian approach and its applications to the estimation of local reserves. *Mathematical Geology*, 15(2):259–286.

Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Application*. Springer-Verlag.

Wagoner, J. V. and Tenney, C. (1991). Nonmarine sequence-stratigraphic concepts and application to reservoir description in the Statfjord Formation, Statfjord Field, northern North Sea. In Hanslien, S., editor, *Petroleum Exploration and Exploitation in Norway - Proceedings of the Norwegian Petroleum Society Conference*, volume 4 of *Norwegian Petroleum Society Special Publications*, pages 381–411. Elsevier, Stavanger, Norway.

Wang, L. (1996). Modeling complex reservoir geometries with multiple-point statistics. *Mathematical Geology*, 28(7):895–907.

Weber, K. J. and Geuns, V. (1990). Framework for constructing classtic reservoir simulation models. *Journal of Petroleum Technology*, 42(10):1248–1253, 1296–1297.

Weissmann, G. S., Carle, S. F., and Fogg, G. E. (1999). Three-dimensional hydrofacies modeling based on soil surveys and transition probability geostatistics. *Water Resource Research*, 35(6):1761–1770.

Wendebourg, J. and Dominik Ulmer, J. (1992). Modeling compaction and isostatic compensation in SEDSIM for basin analysis and subsurface fluid flow. In Pflug, R. and Harbaugh, J., editors, *Computer Graphics in Geology*, volume 41 of *Lecture Notes in Earth Sciences*, pages 143–153. Springer Berlin/Heidelberg.

Wibrin, M., Bogaert, P., and Fasbender, D. (2006). Combining categorical and continuous spatial information within the bayesian maximum entropy paradigm. *Stochastic Environmental Research and Risk Assessment*, 20(6):423–433.

Willis, B. J. and White, C. D. (2000). Quantitative outcrop data for flow simulation. *Journal of Sedimentary Research*, 70(4):788–802.

Xu, H. Q. and Maccarthy, I. A. J. (1998). Markov chain analysis of vertical facies sequences using a computer software package (SAVFS): Courtmacsherry Formation (Tournaisian), southern Ireland. *Computers & Geosciences*, 24(2):131–139.

Appendix A

Bivariate Probability Diagram

A.1 Bivariate Probability Matrix Calculation

A Gslib style program `TPcalc` is developed to calculate the bivariate probability matrix and the related Markov transition probability matrix. The parameter file is shown in Figure A.1. The program will calculate the bivariate probability or Markov transition probability from a log profile or a chosen direction of a training image.

In line 1 and 2 are the category number and category types that exist in the domain.

In line 3 is an indicator to define the bivariate probability or Markov transition probability calculation form well data or from training image.

In line 4 and 5 will be the file name and the specific column number (well ID, depth and category) needed in the program if the calculation is from well data.

If the calculation is from a training image, in line 6 , 7 and 8 will be the training dimension definition and its file name.

Line 9 will be the scan direction. It could be along x or along y or both of them. The test shows that scan from X and Y directions would improve the reproduction of training image pattern.

Line 10 is length and number of the count interval for well data and training image.

The left lines are different output files from the program. Line 11 is the output of the bivariate probability $p(\mathbf{u}_\alpha, \mathbf{u}_\beta)$ and transition probability $p(\mathbf{u}_\alpha|\mathbf{u}_\beta)$ for DMPE

```
                    parameter file for TPcalc
                   ************************
Line number        START OF PARAMETERS:
1                  3                    -number of catgory
2                  1   2  3             -category types
3                  0                    -0: from well data;1: from training image
4                  wells.out            -input markov chain data file
5                  1   2   3            -well ID,depth, category codes column
6                  50   50   1          -the dimension of training image you would scan
7                  0.5  0.5             -the origin coordinated of training image
8                  true_cat.dat         -input training image you want to scan
9                  0                    -which direction do you want to scan? 0:x; 1:y.
10                 0.1      40          -transition count interval and  interval number
11                 mde3.out             -output  bivariate for DMPE
12                 biv3.out             -output bivariate probability for diagram plotting
13                 tp3.out              -output transition probability for diagram plotting
14                 dbg3.out             -debug file
```

Figure A.1: The parameter file used in bivariate diagram calculation program

program.

Line 12 is the out put of bivariate probability $p(\mathbf{u}_\alpha, \mathbf{u}_\beta)$ for plotting in program *TP-diagram*.

Line 13 is the Markov transition probability $p(\mathbf{u}_\alpha|\mathbf{u}_\beta)$ output for plotting. The output for estimation/simulation from line 12 is in different format as those used for plotting from line 13.

Line 14 is the debug information output file name.

A.2 Bivariate Probability Diagram Plotting

Using the DMPE in estimation and simulation, the experimental bivariate probability or transition probability matrix would be used directly in the programs without modelling. So the plotting and doing some visual checking is one important step. The plotting for visual checking or geological pattern recognition can be done using the program `TPdiagram`. The plotting file is modified from the codes `Vargplt`. The parameter file is shown as in Figure A.2.

The output of the Bivariate Probability and Transition Probability from program `TPcalc` is ordered as the head category first then the tail category. For example it

there are three categories $1, 2, 3$. The bivariate probability/Markov transition probability would output in an order of $1 \to 1, 1 \to 2, 1 \to 3, 2 \to 1, 2 \to 2, \ldots, 3 \to 2, 3 \to 3$. An output file of bivariate probabilities for plotting is shown in Figure A.3.

```
                              Parameters for TPdiagram
                             *************************
Line #        START OF PARAMETERS:
1             bivariate.ps          -file for PostScript output
2             1                     -number of variograms to plot
3             0.0   -20.0           -distance  limits (from data if max<min)
4             0.0    -1.2           -bivariate limits (from data if max<min)
5             0      1.0             -plot sill (0=no,1=yes), sill value)
6             bivariate probability  -Title for bivariate probability
7             biv3.out               -1 file with variogram data
8             1    1   1   1    1    -TP #, dash #, pts?, line?, color
9             biv3.out               -2 file with variogram data
10            3    3   0   1   10    -TP #, dash #, pts?, line?, color
```

Figure A.2: The bivariate probability/transition probability diagram plotting program

```
Bivariate Probability for category   1   1
   1     1.000  0.1236734693877551
    ......................
   40    40.00  0.0747826086956522
Bivariate Probability for category   1   2
   1     1.000  0.0681632653061224
    ......................
   40     40.0  0.0943478260869565
Bivariate Probability for category   1   3
   1     1.000  0.0028571428571429
    ......................
   4     40.00  0.0221276595744681
Bivariate Probability for category   2   1
   1     1.000  0.0612244897959184
    ......................
   40     40.0  0.0943478260869565
Bivariate Probability for category   2   2
   1     1.000  0.2151020408163265
    ......................
   40     40.0  0.0943478260869565
Bivariate Probability for category   2   3
   1     1.000  0.0906122448979592
    ......................
   40     40.0  0.0943478260869565
Bivariate Probability for category   3   1
   1     1.000  0.0053061224489796
    ......................
   40     40.0  0.0943478260869565
Bivariate Probability for category   3   2
   1     1.000  0.0816326530612245
    ......................
   40     40.0  0.0943478260869565
Bivariate Probability for category   3   3
   1     1.000  0.3514285714285714
    ......................
   40     40.0  0.0943478260869565
```

Figure A.3: Example of the bivariate probability output file for plotting

Appendix B

DMPE Programs

B.1 DMPE Estimation Program

After the bivariate probability/Markov transition probability is calculated, the output file from `TPcalc` is ready for spatial estimation and simulation. DMPE is implemented in two programs: `DMPEest` and `DMPEsim`.

The program `DMPEest` is used to do estimation or cross validation with the input hard data. The parameter file of `DMPEest` is shown in Figure B.1. Here are some explanations of the parameter files.

Line 1 is the estimation option. 0 is for doing estimation on grid, 1 is doing cross validation for hard data locations.

Line 2 , 3 and 4 are the number of category type, and its proportion. One note here is that the order of category in line 3 should be the same as the order of category in line 2 of parameter file for program `TPcalc` shown in Figure A.1.

Line 5 and 6 are the information of hard data file.

Line 7 is the debug level and the debug information will output to the file name listed in line 8.

Line 9 is the estimation/cross validate results file name.

Line 10, 11 and 12 are the grid definition.

Line 13 is the conditioning data number for DMPE. This number should be small than 8 because of huge multivariate probability space.

```
                          Parameters for DMPEest
                         **********************
Line #   START OF PARAMETERS:
1          1                           -option: 0=grid, 1=cross
2          3                           -number thresholds/categories
3          1   2    3                  -categories
4          0.21  0.29  0.50            -global pdf
5          cluster.out                 -file with data
6          1  2    3    4   5          -columns for DH,X,Y,Z and variable
7          0                           -debugging level: 0,1,2,3
8          dmpe_est.dbg                -file for debugging output
9          dmpe_est.out                -file for DMPE output
10         10   0.5    1.0             -nx,xmn,xsiz
11         10   0.5    1.0             -ny,ymn,ysiz
12         1    0.0    1.0             -nz,zmn,zsiz
13         6                           -maximum conditioning data for DMPE
14         10.0  10.0  2.0             -maximum search radii
15         0.0   0.0   0.0             -angles for search ellipsoid
16         0                           -max per octant (0=not used)
17         truexy_mde3.out             -the bivariate marginal file name for DMPE
18         40  1                       -the total lag in TP calculation & interval length
19         3.0   3.0   1               -the anisotropy transform ratio
20         20.0                        -minimum iteration time for DMPE
```

Figure B.1: The DMPEest program parameter file

Line 14 and 15 are the searching radius and searching angle.

Line 16 would be used when octant searching is used.

Line 17 is the bivariate probability matrix file name.

Line 18 is the lag interval number and its length used in transition probability calculation.

Line 19 is the anisotropy ratio used to transform the spatial distance to the distance used in transition probability calculation.

Line 20 is the minimum iteration time used to do the multivariate probability estimation.

B.2 DMPE Simulation Program

Classical sequential simulation approach is used. During the simulation, previously simulated nodes are used as hard data for later unsampled location conditional probability calculation and Monte-Carlo simulation. Because of the huge multivariate probability space requirement, the number of conditioning data should be no more than 10. Also, in the parameter file the category list in Line 3 as in Figure B.2

```
                                Parameters for DMPEsim
                               **********************
Line   #     START OF PARAMETERS:
1              3                              -simulation number
2              3                              -number of categories
3              1   2   3                      -category types
4              0.20   0.3   0.5               -global proportions  of each category
5              sample_true.dat                -file with local data
6               1  2  3   4                  -columns for X,Y,Z, and category
7              truexy_mde3.out                -the bivariate marginal file name for DMPE
8              40  1                          -total lag number in  TP calculation & its length
9              30.0   30.0   1                -the anisotropy transform ratio
10             3.0                            -minimum iteration time for DMPE
11             0                              -debugging level: 0,1,2,3,4
12             DMPE_rs.dbg                    -file for debugging output
13             rs0813-est-1.out               -file for estimation/simulation output
14             50    0.5   1.0                -nx,xmn,xsiz
15             50    0.5   1.0                -ny,ymn,ysiz
16             1.0    1    1                  -nz,zmn,zsiz
17             31199                          -random number seed
18             5                              -maximum conditioning nodes for DMPE
19             0                              -maximum data number per octant    (0=not used)
20             100.0   100.0   1.0            -maximum search radius
21             0.0    0.0    0.              -angles for search ellipsoid
22             21     21     1               -size of searching table
```

Figure B.2: The DMPE simulation program program parameter file

should be in the same order as that listed in parameter file for bivariate probability calculation program. In this program `DMPEsim`, the hard data are assigned to grid nodes center in simulation. The parameter file is shown in Figure B.2.

Line 1 is given the simulation realization number.

Line 2 and 3 are the category number and its value of the discrete random variable.

Line 4 are the global proportion of the categories in the domain.

Line 5 and 6 are the input of hard data file name and the column number.

Line 7 is the bivariate probability file name and line 8 is the lag *xlag* in building the bivariate probability matrix.

Line 9 is the anisotropy ratio for three directions which is used to relate the 3D spatial distance to the length of TP calculation interval.

Line 10 is the iterative time used in iterative scaling approach when doing the full multivariate probability estimation. Usually, it will converge to a stable solution after 30 times iterations.

Line 11 and 12 are the debug level and output file for the program.

Line 13 is the simulation output file of the DMPE.

Line 14, 15, and 16 is the simulation grid definition.

Line 17 is the random number seed for random path build when doing simulation.

Line 18 is the maximum conditioning nodes. As the constraint of huge probability data event space, it should be no more than 10 given the category is less than three.

Line 19 is the set for octant searching setting.

Line 20 is the minimum search radius.

Line 21 is the search angles for search ellipsoid.

Line 22 is the searching table built for searching.

Appendix C

Core Subroutines in DMPE Implementation

In the implementation of DMPE for spatial estimation and simulation, there are some subroutines that could be useful for other purposes. Here is a short introduction to some of them.

C.1 Single Multivariate Probability Subroutine

The first one is a single program to do a discrete multivariate probability estimation from specified bivariate probability constraints. In the `DMPEest` and `DMPEsim`, the multivariate probability is explicitly estimated and used to calculate the conditional probability. It could be used outside the program.

In the program `DMPEsingle`, a single discrete multivariate probability is estimated from the input bivariate probabilities. The pair-wise bivariate probabilities are input for all the variables. The output is the estimated multivariate probability. This program can be used to test the iterative scaling method in the multivariate probability estimation. The parameter of program `DMPEsingle` is shown in Figure C.1.

```
                        Parameter file for DMPEsingle
                        ************************
Line   #                START OF PARAMETERS:
1                       3                       -number of catgory
2                       6                       -category types
3                       bivprob.out             -input known mv at here
4                       mvprob.out              -univariate probability marginal file
5                       dmpe.dbg                -output the debuging information
```

Figure C.1: One single multivariate probability estimation program parameter file

```
                             Parameter file for MVmarg
                             ************************
Line   #                     START OF PARAMETERS:
1                            3                         -number of catgory
2                            6                         -category types
3                            mvprob.out                -input known mv
4                            marg.dbg                  -output the debuging information
5                            bivprob.out               -bivariate probability marginal file
6                            uivprob.out               -univariate probability marginal file
```

Figure C.2: The parameter file of marginal program

C.2 Multivariate Probability Marginalization Subroutine

The second one is to multivariate probability marginalization. The univariate probabilities and bivariate probabilitis of an input multivariate probabilities will be calculated from the program `MVmarg`. The parameter file is shown in Figure C.2. The category number and the variable number are input in line 1 and line 2. In line 3 is the known multivariate probability file name. The univariate and bivariate probability will be written into the specified file given in line 5 and 6.