GROUPING SIMILARITIES AMONG GEOSTATISTICAL REALIZATIONS OF RESERVOIR PROPERTIES USING MULTIRESOLUTION WAVELET ANALYSIS

Mr. Worapot Laopom

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering in Petroleum Engineering
Department of Mining and Petroleum Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2004

ISBN 974-17-5895-2

การจัดกลุ่มการกระจายตัวของคุณสมบัติของแหล่งกักเก็บที่ได้จากการทำธรณีสถิติโดยใช้เวฟเลตแบบมัลติเรโซลูชันในการวิเคราะห์

นาย วรพจน์ ลาวป้อม

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมปิโตรเลียม ภาควิชาวิศวกรรมเหมืองแร่และปิโตรเลียม
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2547

ISBN 974-17-5895-2

วรพจน์ ลาวป้อม: การจัดกลุ่มการกระจายตัวของคุณสมบัติของแหล่งกักเก็บที่ได้จากการทำธรณีสถิติโดยใช้เวฟเลตแบบมัลติเรโซลูชันในการวิเคราะห์ (GROUPING SIMILARITIES AMONG GEOSTATISTICAL REALIZATIONS OF RESERVOIR PROPERTIES USING MULTIRESOLUTION WAVELET ANALYSIS) อาจารย์ที่ปรึกษา: อ.ดร. สุวัฒน์ อธิชนากร, อาจารย์ที่ปรึกษาร่วม: อ.ดร. สุนทร พุ่มจันทร์ จำนวนหน้า 88 หน้า,

ISBN 974-17-5895-2

ตัวแปรในการประมาณศักยภาพของแหล่งผลิตทางด้านปิโตรเลียมที่สำคัญเช่น ความพรุนของหิน ซึ่งเป็นตัวแปรที่มีความสัมพันธ์กับตำแหน่งนั้น สามารถประมาณค่าโดยใช้วิธีธรณีสถิติ ดังนั้นในงานวิจัยนี้จึงเลือกตัวแปรนี้มาใช้ในการศึกษาโดยใช้ Stochastic Simulation ในการประมาณค่าความพรุนของหินที่ตำแหน่งต่างๆ ในพื้นที่ศึกษาหนึ่ง แต่วิธีการประมาณค่านี้จะให้ผลการประมาณค่าออกมาหลายรูปแบบ (realizations) ที่มีความเป็นไปได้เท่าๆกัน ดังนั้นถ้านำผลการประมาณค่าทั้งหมด (ทุกแบบที่ประมาณค่าได้) มาประมาณศักยภาพของแหล่งผลิต จะเป็นการเสียเวลาและทรัพยากรอย่างมาก

ในงานวิจัยนี้ผู้ทำการวิจัยประมาณค่าของความพรุนของหินจากข้อมูลความพรุนของหินชุดหนึ่ง โดยใช้ Sequential Gaussian Simulation โดยสร้างรูปแบบที่สามารถเป็นไปได้ของความพรุนของหินเป็นจำนวน 60 รูปแบบ จากนั้นนำรูปแบบที่เป็นไปได้ทั้งหมดมาปรับลดความแตกต่าง (denoising) โดยใช้ multiresolution wavelet analysis (Daubechies 4 wavelet) จากนั้นจะจับคู่รูปแบบที่ถูกปรับลดความแตกต่างแล้วเพื่อหาค่าความเหมือน (correlation) โดยเกณฑ์การเปรียบเทียบและจัดกลุ่มรูปแบบคือ กำหนดค่าความเหมือนขึ้นมาค่าหนึ่ง (cut-off value) โดยคู่รูปแบบใดที่มีค่าความเหมือนน้อยกว่าค่าที่กำหนดจะถูกตัดออก และคู่ใดที่มีค่าความเหมือนสูงจะถูกจัดเข้ากลุ่มเดียวกัน จากการจัดกลุ่ม รูปแบบใดที่มีสมาชิกในกลุ่มมากที่สุดจะเป็นตัวแทนของรูปแบบทั้งหมดที่จะนำไปประมาณศักยภาพของแหล่งผลิตต่อไป

ภาควิชาวิศวกรรมเหมืองแร่และปิโตรเลียม	ลายมือชื่อนิสิต.....
สาขาวิชาวิศวกรรมปิโตรเลียม	ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา 2547	ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

4471610021 : MAJOR PETROLEUM ENGINEERING
KEY WORD : /GEOSTATISTICS/ WAVELET/ POROSITY DISTRIBUTION/ DENOISING/ SEQUENTIAL GAUSSIAN SIMULATION

WORAPOT LAOPOM. THESIS TITLE: GROUPING SIMILARITIES AMONG GEOSTATISTICAL REALIZATIONS OF RESERVOIR PROPERTIES USING MULTIRESOLUTION WAVELET ANALYSIS. THESIS ADVISOR: DR. SUWAT ATHICHANAGORN, THESIS CO-ADVISOR: DR. SUNTHORN PUMJAN, 88 pp.
ISBN 974-17-5895-2

Geostatistics offers a way to describe the spatial continuity of reservoir properties such as permeability and porosity. One of the most important parameters in reserve estimation is porosity; thus, it is chosen as a parameter of interest in this study. Realizations of spatial distribution of porosity can be generated from stochastic simulation at equal probability due to the inherent uncertainty in the data set. Anyway, it is impossible to use all realizations as input for reservoir simulation because it will consume too many resources.

In this study, 60 realizations of porosity data set were generated from Sequential Gaussian Simulation. All realizations were denoised by multiresolution wavelet analysis, using Daubechies 4 wavelet. These denoised realizations were then compared and grouped based on correlation between each pair. The criterion used in the comparison and grouping is a cut-off value that is specified to select highly correlated realizations. The realization that has the largest number of similar realizations based on a certain cut-off value of correlation was chosen as the most representative realization and can be used as a porosity distribution reservoir model.

Department of Mining and Petroleum Engineering Student's signature…….
Field of study: Petroleum Engineering Advisor's signature……
Academic year: 2004 Co-advisor's signature……

ACKNOWLEDGEMENTS

I would like to express my appreciation to my advisor, Dr. Suwat Athichanagorn and my co-advisor, Dr. Sunthorn Pumjan, for their advice, guidance throughout the course of this study. I would like to thank Associate Professor Yingyos Khemayodhin and Dr. Jirawat Chewaroungroaj, who are the thesis committee for their revision of the manuscript of this thesis and suggestions.

I would like to thank my parents for their love and encouragement, which helped me to pass through the hard time of this study.

Finally, I also would like to thank for my wife, who is an unconditional helper.

CONTENTS

Abstract (in Thai)……iv

Abstract (in English)……..…...v

Acknowledgements……...vi

Table of Contents…….....…vii

List of Tables……..….. ix

List of Figures……...x

Nomenclature……xii

1. Introduction……..1

1.1 Problem Statement……... 2

1.2 Outline of Approach……3

1.3 Dissertation Outline……..4

2. Literature Review…….……5

2.1 Previous Work on Geostatistics…...…….….. 5

2.2 Previous Work on Wavelet Denoising…… 7

3. Methodology……..……10

3.1 Geostatistics……..10

3.1.1 Theory of Geostatistics……..10

3.1.1.1 Theory of Regionalized Variables……..... 11

3.1.2 Variogram Analysis……..……..11

3.1.3 Geostatistical Simulation……...18

3.1.3.1 Gaussian Simulation…… 20

3.1.3.2 Sequential Gaussian Simulation Procedure……..… 23

3.2 Denoising using Wavelet Analysis……..…….26

3.2.1 Wavelet Transform……26

3.2.2 Multiresolution Analysis……...… 29

3.2.3 Denoising……...33

CONTENTS (continued)

3.3 Correlation Analysis……38

4. Simulations and Results……..….41

4.1 Generating Raw Data……....……... 41

4.2 Simulation of Realizations……... 45

4.2.1 Constructing the Variogram Model……...45

4.2.2 Checking for Bivariate Normality……50

4.2.3 Sequential Gaussian Simulation……52

4.3 Denoising of Realizations…… 61

4.4 Grouping Similar Denoised Realizations…….....65

5. Conclusions and Recommendations…….....……..83

References…….....…….……...86

Vitae…….. 88

LIST OF TABLES

4.1 The locations and porosity values of raw data……..…42
4.2 Statistics of porosity data……...……..…….44
4.3 The locations, porosity values, and normal score values……....…….46
4.4 The statistical analysis of the normal score data……....….. 48
4.5 Variogram model parameters of normal score data.……...…..50
4.6 Variogram parameters using for checking Bivariate Normality……...51
4.7 Calculation of threshold values at each level of decomposition……...…62
4.8 Correlation of realizations……...…….66
4.9 Statistical analysis of correlation values. ……...……..67
4.10 Correlation values among 4 images: **A**, **B**, **C**, and **D**…...…….....70
4.11 Ranking of highly correlated denoised realizations with
a cut-off value of 0.35…….…….....71
4.12 Grouping of original realizations using different cut-off values…….……. 74
4.13 Grouping of denoised realizations using different cut-off values…….75
4.14 The number of the grouping members of original realizations...…....……..... 81
4.15 The number of the grouping members of denoised realizations……...82

LIST OF FIGURES

3.1 Basic components of a variogram…… 12
3.2 Power model……. 14
3.3 Spherical model…… 15
3.4 Exponential and gaussian model ……..…….. 16
3.5 Variogram parameters (angle tolerance, lag distance, and bandwidth)……....18
3.6 Geostatistical simulation maps and Kriging map……..……... 19
3.7 Graphical display of normal score transform …….. 21
3.8 Sequential Gaussian Simulation process…….. 25
3.9 The Haar function……27
3.10 Haar wavelet examples……28
3.11 Approximations of a signal decomposed by wavelet decomposition…….....30
3.12 Details of a signal decomposed by wavelet decomposition……..... 31
3.13 Schematic representation of the decomposition algorithm……..….32
3.14 Schematic representation of the reconstruction algorithm……... 32
3.15 The hard and soft thresholding functions…….…… 34
3.16 Decomposition of a signal using Daubechies 4 for 2 levels…….35
3.17 Thresholding detail signals...……..……..……....36
3.18 Original and denoised signals...……....37
4.1 Location map of porosity data……..43
4.2 Histogram of porosity data……... 44
4.3 Histogram of the normal score porosity data……48
4.4 The omnidirectional variogram plot and its variogram model for normal score data…….. 49
4.5 Experimental indicator variogram and Gaussian model-derived indicator variogram at median cut-off ……..... 52
4.6 Parameters file of the Sequential Gaussian Simulation…...……53
4.7 Realizations number 1-8……... 54
4.8 Realizations number 9-16…….... 55
4.9 Realizations number 17-24……... 56
4.10 Realizations number 25-32……... 57

LIST OF FIGURES (continued)

4.11 Realizations number 33-40……...58
4.12 Realizations number 41-48……...59
4.13 Realizations number 49-56……...60
4.14 Realizations number 57-60……...61
4.15 Original realization number 1 and its denoised maps reconstructed from 5 different levels of resolution ……63
4.16 Denoising a realization using MATLAB program……...64
4.17 Examples of denoised realizations……65
4.18 Histogram of correlation values between 3,540 pairs of original realizations……....68
4.19 Histogram of correlation values between 3,540 pairs of denoised realizations……68
4.20 The first 8 members of the group number 4 using denoised realizations at cut-off value of 0.28 (realization number 4, 35, 36, 12, 30, 34, 18, and 58) ……76
4.21 The other 8 members of the group number 4 using denoised realizations at cut-off value of 0.28 (realization number 52, 56, 22, 47, 26, 23, 59, and 46)…….....…… 77
4.22 The 8 members of the group number 22 using denoised realizations at cut-off value of 0.35 (realization number 22, 36, 27, 20, 3, 11, 51, and 12)……..78

NOMENCLATURE

a	range distance
$C(h)$	covariance function
C_0	nugget value
C_1	sill value
$c_{j_0,k}$	coefficient of scaling function
$d_{j_0,k}$	coefficient of wavelet function
$d_{j,k}^{H}$	detail coefficient of hard thresholding at level j
$d_{j,k}^{S}$	detail coefficient of soft thresholding at level j
$F_Z(Z)$	cumulative distribution function
r	correlation coefficient
ReV	Regionalized variables
RF	Random function
RV	Random variables
S_{xy}	covariance of variable x and y
S_x^2	variance of variable x
S_y^2	variance of variable y
$Wf(j,k)$	wavelet transform
$\overline{x}$	mean of variable x
$\overline{y}$	mean of variable y
$Z(x_i)$	Realization
$Z_{SK}^{*}(u)$	simple kriging estimator at location u

GREEK LETTERS

$\gamma(\vec{h})$	variogram function
$\lambda_\alpha(u)$	weight for α sample ($\alpha = 1,...,n$) at location u

λ_j	threshold value at level j
$\psi(x)$	wavelet function
$\phi(x)$	scaling function
$\hat{\sigma}$	the median of absolute deviation (MAD)
$\sigma^2_{SK}(u)$	simple kriging variance at location u

CHAPTER I

INTRODUCTION

Porosity is an important reservoir parameter for oilfield development. If the porosity in a reservoir is high, it is likely that the reservoir has a lot of void to contain fluid. On the other hand, if the porosity is low, the reservoir has a little void. In oil and gas reserve calculations, porosity is one of several parameters to be concerned. Porosity data are generally collected and analyzed in order to design the development plan of a field.

In practice, only a few wells are drilled in a reservoir because the cost of drilling is very expensive. As a result, porosity data obtained in the field of interest is limited. When designing a development plan, drilling wells in the areas which lack porosity data poses a problem that challenges petroleum engineers. In the past, invert distance and polygon methods were used to find those missing data. However, these two methods may not give satisfying results because of the assumptions involved.

Later, geostatistics was introduced to estimate data at unsampled locations using a relationship between distance and variogram value, which is calculated from the summation of the squares of the difference between each pair of data, which are far apart from each other at a particular distance, divided by the number of pairs. This technique uses the variogram to find the influential range of data to a location of interest. There are two approaches to generate distribution of reservoir properties, which are kriging and simulation. Kriging provides a single realization of the spatial distribution, while simulation gives many realizations at equal probability due to the uncertainty of data.

To evaluate reservoir performance, reservoir parameters such as permeability or porosity are needed in reservoir simulation. When these parameters are combined with other parameters such as fluid properties, production plan, etc., reservoir simulation study can be carried out to forecast reservoir performance.

1.1 Problem Statement

In general, there may be a numerous number of realizations generated for reservoir property such as porosity when Stochastic Simulation is used. Every realization resulting from Stochastic Simulation can possibly represent the real structure of the reservoir with the same probability. It is difficult to pinpoint which one is more accurate than others until the reservoir has produced for a certain period and the production profile is historically matched. It would be better, if there is a method to choose the realization which best represents the real structure as the reservoir model in reservoir simulation. To find the best representative model, similar realizations have to be grouped and the realization to which most realizations are similar should be the best representative of the reservoir.

Unfortunately, generated realizations have little similarity among them since there are a lot of small-scale heterogeneities. So, the grouping does not give a proper result. In order to solve this problem, realizations of porosity distribution need to be smoothed first, i.e., getting rid of local variations.

Denoising is a technique that can be used to find the smooth structure of reservoir porosity, making it easier to compare similarities among realizations. The technique used in this study is developed from multiresolution wavelet analysis. When the multiresolution analysis is applied to a realization, the realization is decomposed to several layers. At this stage, the denoising algorithm determines which components are small-scale variations and which are the main structure. By suppressing the local variations and reconstructing the realization, a smoother or denoised realization is obtained. By doing this, the denoised realization still possesses the same underlying structure.

The value that is used to determine the degree of similarity among realizations is correlation between each pair of realizations. Based on the computed correlations, these realizations will be grouped and the best representative realization will be determined from the biggest group.

1.2 Outline of Approach

To find the best representative of the realizations estimated from geostatistics using wavelet analysis, the following approach can be taken:

1. A data set of porosity is constructed in the same nature as the real field data such as the spacing of wells. In this study, the raw data, which are 47 porosity sampling data from 35 regular wells and 3 horizontal wells, are artificially created. In addition, the study area covers 10,240 feet in north-south direction and 10,240 feet in east-west direction.
2. In practice, it is hard to find directional variograms from porosity data because the number of data is not enough due to high cost. As a result, an omnidirectional variogram model, a variogram model representing all directions, is used to examine the correlation between the data set and distance among data.
3. 60 realizations of porosity field are generated by Sequential Gaussian Simulation.
4. To make it easier to compare among realizations, each realization needs to be denoised. The soft thresholding method in wavelet analysis is used. The chosen wavelet function is Daubechies 4, which is the most frequently used function. The threshold values are calculated from universal thresholding method presented by Donoho and Johnstone (1994).
5. All denoised realizations are compared based on the correlation between each pair. A pair that has a high correlation value means that they have a high degree of similarity between them. However, a cut-off correlation value is to be specified to eliminate realizations that have a small degree of similarities before they are grouped.
6. The grouping criterion in this study is that members of a group are chosen from the highest to lowest correlation values. In addition, in case that the next realization, which is chosen by the order of correlation value, does not match with all the previous members, it cannot be a member of the group.

All of the procedure can be divided into 3 parts, which are geostatistical modeling, denoising via wavelet analysis, and grouping. GSLIB software is used in the modeling part while MATLAB program is used for other two parts.

1.3 Dissertation Outline

Chapter II presents previous works concerning with this study.

Chapter III introduces the methodology used in this study including geostatistics, wavelet analysis, and correlation analysis. This chapter is divided into 3 sections, which are presented as follows:

- Section 3.1 discusses the geostatistics analysis. The theory of geostatistics is first presented. After that, a procedure to determine the relationship among the set of data and the separating distance is introduced in term of variogram analysis. The last topic of this chapter is Sequential Gaussian Simulation, explaining how to estimate the variable value at each location of interest. All of the topics in this chapter can be applied to any variable of interest that exhibit a certain spatial relationship, including porosity, permeability, water saturation, and etc.

- Section 3.2 introduces the wavelet analysis. It presents a brief theory of wavelet transform and introduces multiresolution analysis. A procedure of data denoising is then discussed. This part includes threshold value calculation, which plays an important role in denoising.

- Section 3.3 presents the correlation analysis used to find similarity between two denoised realizations.

Chapter IV presents the approach taken in this study including generating raw data, simulating realizations from the data using Sequential Gaussian Simulation, denoising generated realizations by applying wavelet analysis, and finally grouping similar denoised realizations using correlation.

Chapter V summarizes the results from the study. The conclusions and recommendations are also presented.

CHAPTER II

LITERATURE REVIEW

Since geostatistics and denoising techniques have been introduced to petroleum industry for many years, applications of these techniques appear in several studies. Most of the studies on geostatistics focused on finding possible structures of interested parameters such as permeability, porosity, and thickness of the reservoir rock. On the other hand, studies on denoising techniques give us several alternatives to smooth raw data and upscale reservoir properties.

In this study, Sequential Gaussian Simulation is used to determine possible distributions of porosity in an artificial field and then these distributions are denoised using multiresolution wavelet analysis to find smooth small-scale variations in these distributions.

2.1 Previous Work on Geostatistics

Aasum, Kelkar, and Gupta (1991) studied the application of geostatistics and fractal geometry to examine 2D reservoir characterizations, porosity distribution and permeability distribution in a dolomitic layered-cake reservoir. Their study accommodated both hard data such as core data and soft data such as geophysical data to perform conditional simulation and Monte Carlo simulation to characterize rock properties. The hard data are 100 to 200 values of permeability and porosity. The authors constructed both ergodic and nonergodic indicator variograms to find the correlation structure of each layer of the study area. They found that the nonergodic variogram gives better results than the ergodic one because the nonergodic variogram is less restrictive. After that the conditional simulation was performed. Hence, some realizations were generated from the simulation because of the uncertainty of the data. In this study, reservoir simulation was not conducted.

Poquioma, Intevep, and Kelkar (1994) applied geostatistics to forecast performance of waterflooding in an oilfield. The field studied was sandstone field.

There were log data from 13 wells, core data from 3 wells, and a single well test. The data set was used to forecast distributions of porosity, permeability, net oil thickness, initial water saturation, and top of the structure, using geostatistical methods. They first carried out both conventional and nonergodic variograms and finally discovered that the nonergodic variogram gave better results because it accounted for biased sampling in the data sets. After that, all distributions of variables were found by applying both conventional kriging and Sequential Indicator Simulation to the data sets. The result from Sequential Indicator Simulation gave good results, while kriging technique gave too smooth results. When the results come from both techniques were put into flow simulation, they found that the Sequential Indicator Simulation results could construct much more realistic reservoir description than the other one.

Hand et. al. (1994) studied the integration of geological, petrophysical, and outcrop data for evaluation of gravity drainage infill drilling at Prudhoe Bay. The field studied, Romeo zonation in Prudhoe Bay was 6,600 X 8,874 ft^2 in size in which 41 conditioning wells were located. The authors obtained zonation pick, facies association picks from 39 conditioning wells, and log and core permeability and porosity data from 37 conditioning wells. Since the authors chose truncated Gaussian simulation to generate 3D facies association maps, data had to be transformed to proper forms. Hence, facies associations were transformed into indicator variables, and permeability and porosity data were transformed into normal score distributions. After that, an experimental variogram of each facies association were performed in the vertical direction. For lateral direction, two experimental variograms were implemented in principle and minor directions, which were determined by trial and error. After the facies association variograms were constructed, the permeability and porosity variograms were then built. To find lateral variograms, the authors assumed that the variogram model and sill of the lateral variogram equaled to vertical variogram for each facies variogram. Then, the authors implemented truncated Gaussian simulation, combining indicator variables and Sequential Gaussian Simulation used to distribute 3D porosity and permeability separately within each facies association, to find the 3D facies associational model. The 3D facies

association description generated using this methodology was judged by geologists to represent a very reasonable geological picture of the studied area.

2.2 Previous Work on Wavelet Denoising

Wavelet analysis has been used in many fields such as signal processing, image analysis, or data compression. The principle of wavelet analysis is that it can break down a signal or function to approximation and detail components. Furthermore, it can recombine approximation and detail components back to the original signal using wavelet multiresolution analysis. By applying these analyses to an interested signal, image, or data, researchers are allowed to diagnose their data. The following previous works are applications of wavelet analysis.

Donoho and Johnstone (1993) studied ideal spatial adaptation by wavelet shrinkage. The authors described the principle for spatially-adaptive estimation, selective wavelet reconstruction, comparing with other spatially-adaptive estimators and developed a practical spatially-adaptive method, RiskShrink, which works by shrinkage of empirical wavelet coefficients. In their experience, the empirical wavelet coefficients at the finest scale were pure noise. Naturally, there was an upward bias estimation due to the presence of some signal at that level. However, this method can control the bias effectively. Finally, the authors introduced optimal thresholds calculated from various types of threshold such as soft and hard, oracle type (projection, shrinkage), and universal. These thresholds are used to cut off noise when performing wavelet denoising.

Chu, Shatzinger, and Tham (1998) presented an application of wavelet analysis to upscale rock properties. Due to the limitation of computational power, reservoir simulation cannot be run at fine scale. However, reservoir property obtained from geostatistical technique is available at fine scale. Thus, the reservoir properties have to be scaled up to coarse scale before they can be used to run reservoir simulation. The researchers noticed that although numerous upscaling techniques are reported in the literature, reasonably accurate equivalent rock properties from the data at finer scale remain challenging and there is no method that can be used consistently for both single and multiphase-flow conditions. Hence, the authors introduced

multiresolution wavelet analysis, to scale up the fine scale properties to coarse scale properties. The decomposition of multiresolution wavelet analysis is the step that transforms the original reservoir property distribution into approximations and details at different resolutions. By applying one level of the decomposition to the original 2-dimensional distribution, the number of grid blocks in the approximation can be reduced by a factor of four, making the approximation coarser than the original one. To upscale reservoir property distributions, the authors performed decomposition using standard Daubechies wavelet function. After that, both upscaled distributions and original distribution were then used in flow simulation to compare the results. The authors found that the original distribution and its approximation gave almost the same result of flow simulation.

Kikani and Meiqing (1998) studied analysis of long-term pressure data using wavelet methods. Since the pressure is measured in a long-term fashion, the data can be used to characterize or improve reservoir models. Unfortunately, the data are huge in volume and prone to uncertainty. Hence, the data have to be preprocessed before they can be used in reservoir characterization process. In this study, a processing procedure is denoising, upscaling, and event detection. The denoising and upscaling stages were implemented using multi-resolution wavelet analysis. The authors denoised the data using the Haarlet in the decomposition and reconstruction. Moreover, they compared the efficiency of data denoising using Butterworth filter and wavelet and found that the wavelet gave a better result. The next stage is upscaling the denoised data to the reservoir simulation scale. In their study, they chose Daubechies 4 wavelet as a wavelet function to scale up the denoised data. They concluded that the upscaling using wavelet technique could preserve the structure of data.

Athichanagorn, Horne, and Kikani (1999) proposed using wavelet analysis to process long-term data obtained from permanent downhole pressure gauges. In their study, there were enormous pressure data, which could be useful in characterizing reservoir. However, the data may include unexpected measurements because they are measured in operational conditions. Hence, the data had to be processed before they can be used to characterize the reservoir. In their study, outliers in the data were first removed by applying wavelet analysis. After outlier data were eliminated, the noise of

the data was the next target. The authors used a hybrid thresholding method, which combines the advantages of the hard and soft thresholding methods to denoise the data. As a result, the denoised data were ready for interpretation.

In summary, stochastic simulation of geostatistical method is often used to determine reservoir property distributions such as permeability, net oil thickness, top of reservoir structure, and porosity because it gives a better result comparing with other methods. Indicator simulation is chosen to find distributions of facies while Gaussian simulation is often chosen to find distributions of porosity and permeability. Some previous works composed 3D models of reservoir property using indicator simulation in the vertical direction and Gaussian simulation in the horizontal direction.

Regarding wavelet analysis, it was used to process data such as upscaling, denoising. Upscaling uses multiresolution wavelet analysis to decompose data to coarser levels. As a result, the output from upscaling using wavelet analysis exhibits the same structure as the original data. Denoising also uses multiresolution wavelet analysis to decompose data into approximation and detail components. However, there is a process, which is wavelet thresholding, to eliminate noise in the data in the detail parts before reconstructing the data back to the original level. In theory, the soft thresholding method gives good results at continuous part of data while the hard thresholding method does at the area of discontinuities in the data. In practice, there is a hybrid thresholding method that combines the advantages of the hard and soft thresholding methods.

CHAPTER III

METHODOLOGY

This chapter presents the theory and technique used in grouping similarities among geostatistical realizations of reservoir properties using multiresolution wavelet analysis. It involves three steps, which are geostatistical simulation, wavelet analysis, and correlation analysis.

3.1 Geostatistics

This section discusses the theory and some applications of geostatistics. The most important characteristic of geostatistics lies in its ability to incorporate the spatial variability structure of variables into its simulation model. It is specially designed for variables of which their values vary with the change of distance and direction, which are called spatial variables. Due to shortcoming of the classical statistics in characterizing the phenomenon of spatial variables, geostatistics technique is widely applied for the modeling of the spatial variables. Most of petrophysical properties such as permeability, water saturation, thickness, and porosity can be categorized into spatial variables.

3.1.1 Theory of Geostatistics

Geostatistics plays an important role in describing the phenomenon of spatial variables. Its application has been widely applied to data in many fields such as soil science data, mineral assays data, forestry inventories, environmental science data and petroleum data, etc. Geostatistics was first introduced in the mining industry in the early fifties by D.G. Krige who was a mining engineer to estimate the ore reserve. The estimator is known as Kriging estimation. In the late fifties, the theory of Regionalized Variables, which was developed from kriging concepts, was formulated by George Matheron. After that, geostatistics has been applied into many industries which are concerned with spatial variables.

3.1.1.1 Theory of Regionalized Variables

A Regionalized Variable is any variable distributed in space or sometime time. Any measurement of Regionalized Variable can be viewed as a realization of random function. The theory introduces four definitions, which are Regionalized Variables (ReV), Realization $Z(x_i)$, Random Variable (RV), and Random Function (RF). Regionalized Variables are measurable quantities which characterize the natural phenomena such as porosity of rock, ore grade, level of ground surface, etc. Realization is defined as a collected value of the Regionalized Variables. Random Variable is defined as a variable that takes a certain number of numerical values according to a certain probalility distribution or in specific a univariate distribution function. And, Random Function is the set of auto-correlated random variables or in specific a multivariate distribution function with n Random Variables ($n \in D$; D = study domain). From the definition of Random Function, the phenomenon of study domain is completely described by RF. In reality, it is impossible to have a complete data to characterize natural phenomena. However, it can be said that the Random Function model is an effective way to characterize uncertainty inherited in the model. The spatial variability structure can be found from the Random Function model.

In practice, the spatial variability structure is quantified by variogram analysis. Hence, variogram tool was introduced to find the spatial variability structure of a Regionalized Variable of interest.

3.1.2 Variogram Analysis

Variogram is a graphical display that helps analysts to find the spatial variability structure of ReV at a specified distance and direction. As it was introduced, the variogram function depends only on the separation vector, $\vec{h}$, not on the locations. The variogram equation is defined as:

$$\gamma(\vec{h}) = \frac{1}{2N(h)} \sum_{i=1}^{N} [Z(x_i) - Z(x_i + \vec{h})]^2 \qquad (3.1)$$

where

$\gamma(\vec{h})$ = variogram value at distance $\vec{h}$

$N(h)$ = number of data pair

$[Z(x_i) - Z(x_i + \vec{h})]$ = the difference in value between two sample points separated by distance $\vec{h}$.

The above equation is used to calculated the variogram value at any distance (h) and direction. The plot of variogram values against distance ($\vec{h}$) along a specific direction presents the spatial variability structure of that variable. In another word, the spatial variability structure of variable is captualized by variogram model. Fig. 3.1 shows the basic components of a variogram model.

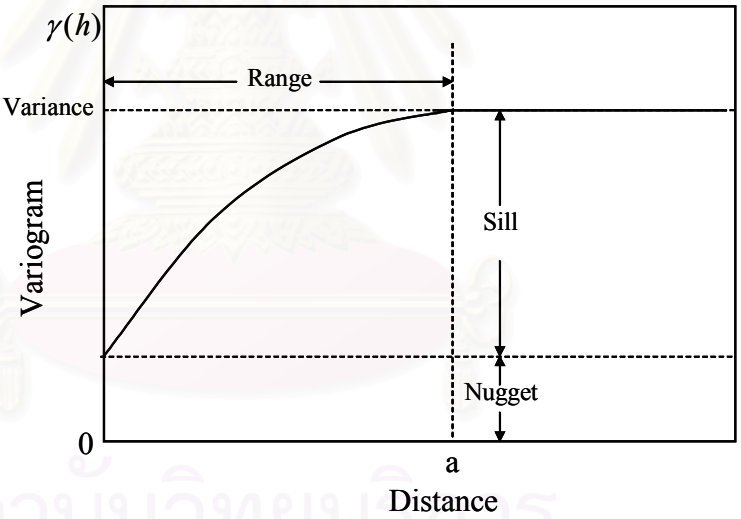

Figure 3.1: Basic components of a variogram.

As seen in Fig. 3.1, three major components of variogram model are sill, range, and nugget. Sill is the maximum variance of variable and is equal to the data variance. Range is the maximum distance in which data still have correlation. Nugget represents the variation at small scale and should be zero at zero distance. But in practice, nugget value comes from two sources, measurement error and small scale variation.

To find the structure of a variable, we start with the calculation of experimental variogram values at given distances using Eq. 3.1. Then, the standard variogram model is fitted to the points of variogram values. The rule of thumb is that the standard models have to fit the points of variogram values as much as possible and is observed by visual inspection. In general, they may be one or more standard models (nested structure) that can be best matched to the variogram plot. Some examples of these standard models are power model, spherical model, exponential model, and gaussian model. The equation and characteristics of these models are described as follows:

(i) Power Model

The power model is called non-transition model with an absence of a sill in the increasing of variogram values, its equation is defined as:

$$\gamma(h) = C_0 + wh^a \tag{3.2}$$

where

C_0 = nugget value

w = slope at origin

a = a real number

h = distance

Fig. 3.2 shows the power model at different values of a. In addition, the power model is called linear model when a equals to one.

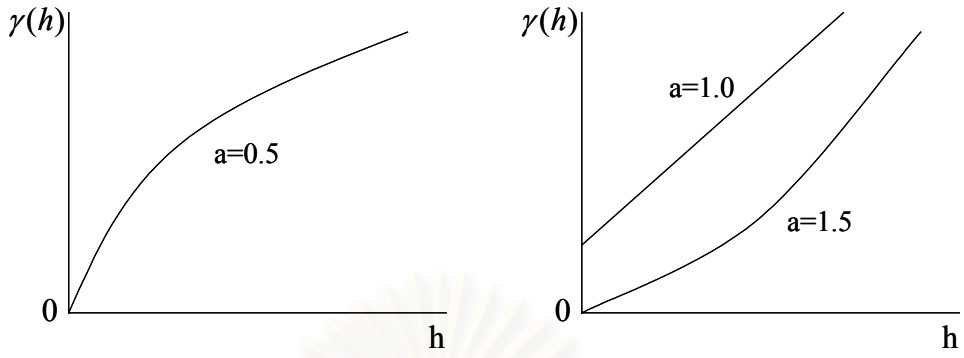

Figure 3.2: Power model.

(ii) Spherical Model

Based on the behavior at the origin and the presence of sill in the increasing of variogram values, the spherical model is called transitional model. Some other models that are defined as transitional model are exponential model and gaussian model. The equation and its definition of spherical model is as follows:

$$\gamma(h) = \begin{cases} C_0 + C_1\left[1.5(\frac{h}{a}) - 0.5(\frac{h}{a})^3\right] & when \quad h \le a \\ C_0 + C_1 & when \quad h > a \end{cases} \tag{3.3}$$

where

C_0 = nugget value

C_1 = sill value

a = range

h = distance

Fig. 3.3 presents the spherical model. The nugget (C_0) is defined as the $\gamma(h)$ at zero distance. The sill is defined at the transition of variogram values when they become stable. The range is defined at the distance where the variogram becomes stable, Fig. 3.3.

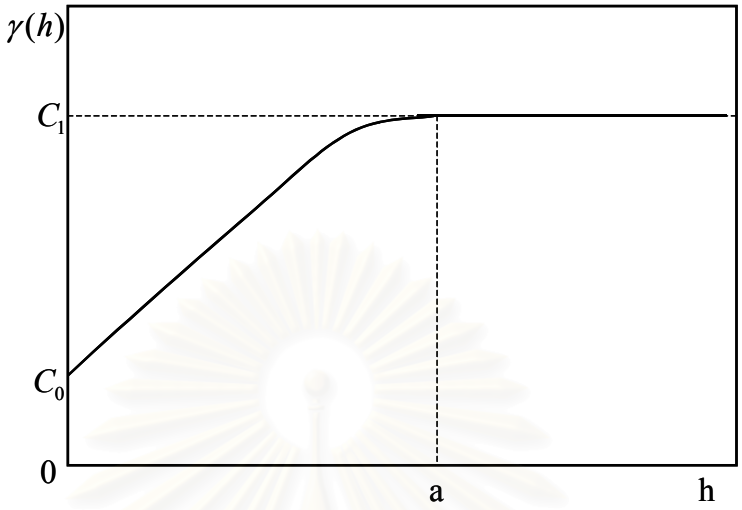

Figure 3.3: Spherical model.

(iii) Exponential Model

Exponential model is a transitional model where the transition of the variogram value takes a longer distance comparing to other models as can be seen in Fig. 3.4. The equation and definitions of the exponential model is as follows:

$$\gamma(h) = \begin{cases} C_0 + C_1\left[1 - \exp(-\frac{h}{a})\right] & when \quad h \le a \\ C_0 + C_1 & when \quad h > a \end{cases} \tag{3.4}$$

where

C_0 = nugget value

C_1 = sill value

a = range

h = distance

(iv) Gaussian Model

Gaussian model is a transitional model with the S-curve behavior at the origin as shown in Fig. 3.4. The equation and definition of gaussian model is:

$$\gamma(h) = \begin{cases} C_0 + C_1\left[1 - \exp(-\frac{h^2}{a^2})\right] & when \quad h \le a \\ C_0 + C_1 & when \quad h > a \end{cases} \tag{3.5}$$

where

C_0 = nugget value

C_1 = sill value

a = range

h = distance

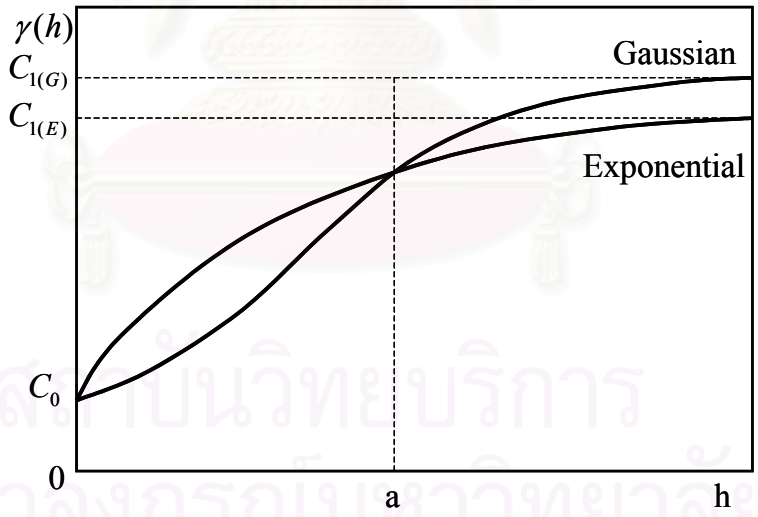

Figure 3.4: Exponential and gaussian model.

Shown in Fig. 3.4, $C_{1(G)}$ is the sill value of gaussian model and $C_{1(E)}$ is the sill value of exponential model.

In practice, a spherical model is frequently used to describe the spatial variability structure of regionalized variables. Nevertheless, a model, which mixes between two or more models, is sometimes used to get the best fit to the variogram plot.

Basically, 2D variograms are calculated in 4 major directions, which are N-S, E-W, NE-SW, and NW-SE. However, an omnidirection variogram, which is the variogram that represents all directions, is calculated for the studied area that has symmetrical structure.

The variogram calculation depends largely on sample distribution. Two parameters which are angle and lag distance are introduced to calculate variogram values in less restrictive way. Angle is the direction of variogram calculation as discussed earlier. Lag distance is the distance (h) for variogram calculation. This process can be applied well to regularly spaced data but not the irregularly spaced data because there are few pairs of data in irregularly spaced data that have the same distance. Hence, tolerance angle, bandwidth, and lag tolerance are introduced to find the more number of calculated pairs. Tolerance angle is an angle which deviates from the observed direction. However, at a large distance, the direction of interest loses its consistency. Therefore, another parameter, which is bandwidth, is introduced. Bandwidth is the offset of the direction of interest. It keeps the pairs of sample that are located between the direction vector and bandwidth lines. To find the variogram value at any distances, lag distance is specified first and the real distance is then calculated from the average distance in each lag. The three parameters are graphically shown in Fig. 3.5.

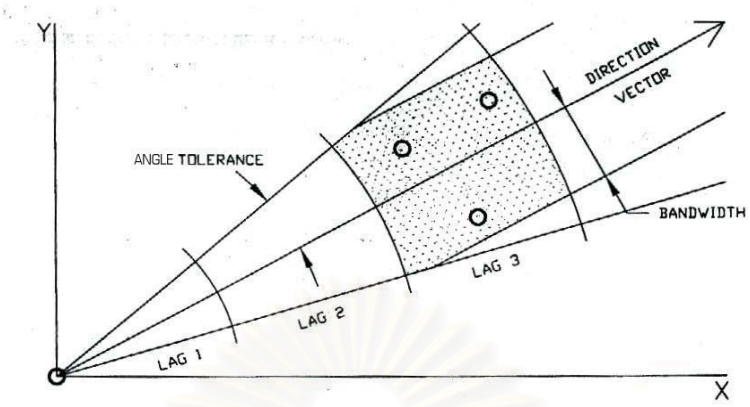

Figure 3.5: Variogram parameters (angle tolerance, lag distance, and bandwidth).

In general, variogram is a significant tool to find the spatial variability structure of Regionalized Variables by calculating variogram values at varying distances in directions of interest and then modeling the variogram function for each direction. The next section describes the estimation process by using the variogram function.

3.1.3 Geostatistical Simulation

There are two forms in geostatistical prediction which are estimation and simulation. In estimation, the best estimated values are obtained by using the method of Kriging estimation. Kriging estimation is based on the knowledge of spatial variability structure of variable and sample distribution surrounding the location to be estimated. It has been well documented that Kriging estimation is a reproduction of variogram model. Therefore, the statistics of original data are preserved to some extent. The major drawback of Kriging estimation is that the estimated values tend to get closer to the mean values, therefore reducing the overall variance. This effect is known as “Smoothing Effect”.

In contrary to geostatistical prediction, conditional simulation aims to simulate the real condition of natural phenomena. It involves the building of alternatives, equally probable and high resolution models of spatial variability structure. Each model represents the reality of natural phenomena in global sense. In geostatistical simulation, the study starts with the finding of spatial variability structure of variable.

And, this structure is used as conditional information together with available samples to construct the conditional probability distribution function (*pdf*) at every location. Then, the simulated values are uniformly drawn from these estimated *pdf*. Finally, many realization maps are then generated. Each realization map is different from the other and conditioned to the available samples and the previously simulated data. The most important aspect in geostatistical simulation is that the statistics of original data are preserved, and hence the spatial distributions of realization maps are the same as sample data. Sample realization maps generated from geostatistical simulation in comparision to Kriging map are shown in Fig. 3.6.

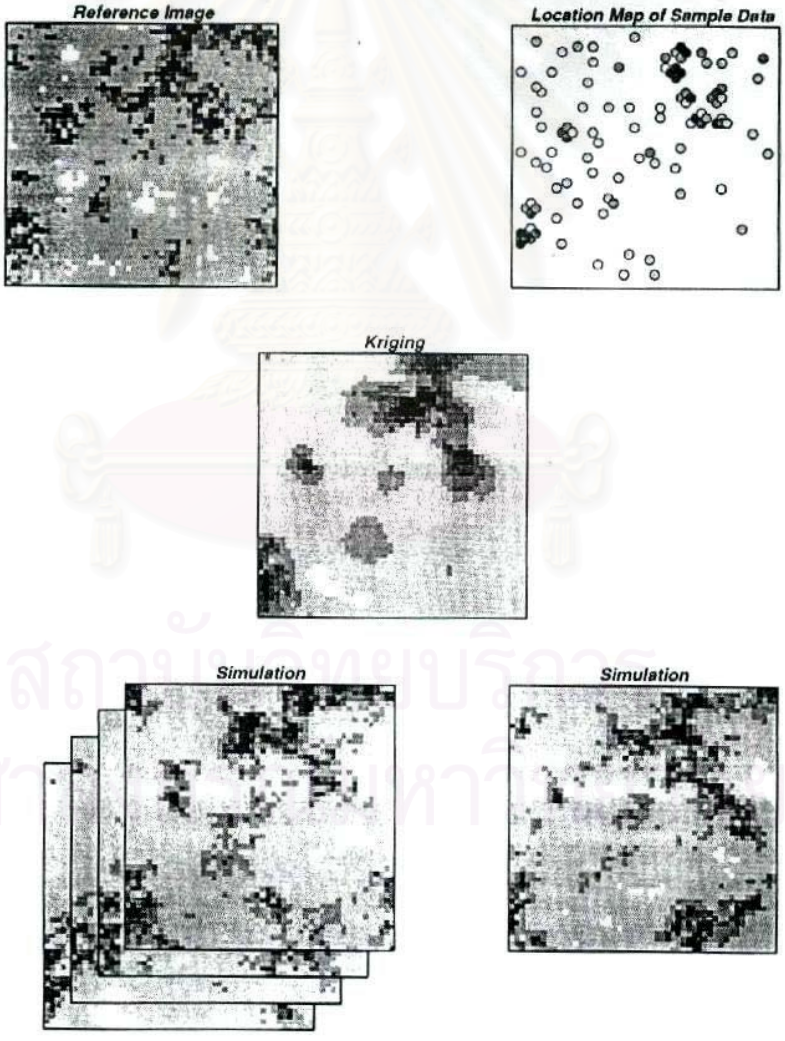

Figure 3.6: Geostatistical simulation maps and Kriging map (Deutsch and Journel, 1992).

3.1.3.1 Gaussian Simulation

Consider the distribution over a field A of one or more attribute(s) $z(u), u \in A$. Geostatistical simulation makes the alternatives, which are equally probable and high-resolution models of the spatial distribution of $z(u)$. To implement Sequential Gaussian Simulation, some related algorithms, which are normal score transform, checking for bivariate normality, and Simple Kriging (Deutsch and Journel, 1992), need to be explained.

(i) Normal Score Transform

The assumption of Gaussian Simulation states that the study variable has to follow standard normal distribution with zero mean and unit variance. The process of transforming original data to standard normal data is carried out using normal score transform function.

Let Z and Y be the two data sets and their *cdf* (Cumulative Distribution Function) are $F_Z(z)$ and $F_Y(y)$, respectively. The transform $Y = \varphi(Z)$ identifies the cumulative probabilities, which correspond to the Z and Y p-quantiles:

$$F_Y(y_P) = F_Z(z_p) = p, \forall p \in [0,1] \tag{3.6}$$

Thus,

$$y = F_Y^{-1}(F_Z(z)) \tag{3.7}$$

with $F_Y^{-1}(\cdot)$ being the inverse *cdf*, or quantile function, of Y data set:

$$y_p = F_Y^{-1}(p), \forall p \in [0,1] \tag{3.8}$$

In case that Y is standard normal with *cdf* $F_Y(y) = G(y)$, the transform $G^{-1}(F_Z(\cdot))$ is the normal score transform. Fig. 3.7 presents the graphical transformation from real data set to normal score data.

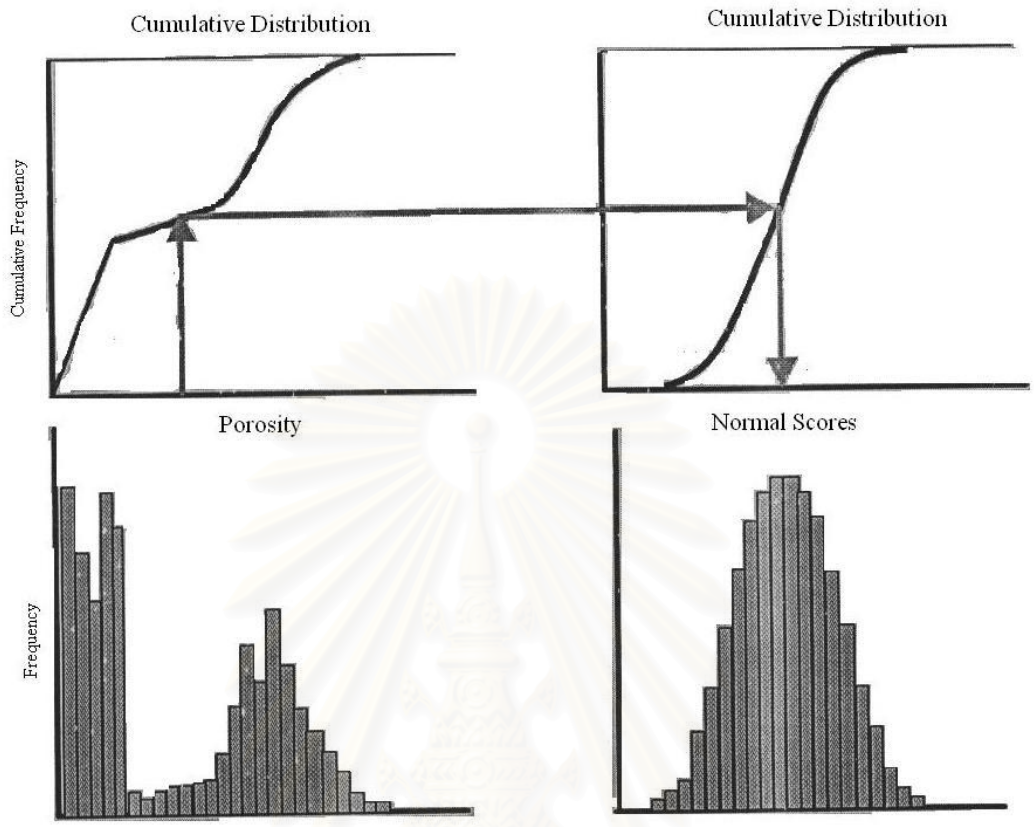

Figure 3.7: Graphical display of normal score transform.

(ii) Checking for Bivariate Normality

To perform the simulation, the bivariate *cdf* of any pair of values $Y(u), Y(u+h)$, $\forall h$, has to be normal. In fact, there are several ways to check the bivariate normality of a normal score data set but the famous method is comparing the experimental bivariate *cdf* of any set of data pairs $\{y(U_\alpha), y(U_\alpha + h), \alpha = 1, ..., N(h)\}$ with covariance function $C_Y(h)$, which is shown as follows:

$$\text{Prob}\{Y(u) \le y_p, Y(u+h) \le y_p\} = p^2 + \frac{1}{2\pi} \int_0^{\arcsin C_Y(h)} \exp(-\frac{y_p^2}{1+\sin\theta}) d\theta \tag{3.9}$$

where $y_p = G^{-1}(p)$ is the standard normal p-quantile and $C_Y(h)$ is the covariance function of the standard normal random function of $Y(u)$.

However, the bivariate probability of the above equation is the non-centered indicator covariance for the threshold y_p:

$$\text{Prob}\left\{Y(u) \le y_p, Y(u+h) \le y_p\right\} = E\left\{I(u;p) \cdot I(u+h;p)\right\} = p - \gamma_I(h;p) \tag{3.10}$$

where

$I(u;p)$ = 1; if $Y(u) \le y_p$,
= 0; otherwise.

$\gamma_I(h;p)$ = the indicator variogram for the p-quantile threshold y_p.

(iii) Simple Kriging

Simple Kriging uses the basic linear regression algorithm and corresponding estimator:

$$\left[Z^*_{SK}(u) - m(u)\right] = \sum_{\alpha=1}^{n} \lambda_\alpha(u)\left[Z(u_\alpha) - m(u_\alpha)\right] \tag{3.11}$$

where

$Z(u)$ = the random variable model at location u.

u_α = the n data locations.

$m(u) = E\left\{Z(u)\right\}$ = the location-dependent expected value of random variable $Z(u)$.

$Z^*_{SK}(u)$ = the linear regression estimator, which is called Simple Kriging.

The Simple Kriging weights $\lambda_\alpha(u)$ are calculated from the following Simple Kriging system:

$$\sum_{\beta=1}^{n} \lambda_\beta(u) C(u_\beta, u_\alpha) = C(u, u_\alpha), \alpha = 1, ..., n \tag{3.12}$$

In the Simple Kriging system, it is required that the means of $m(u)$ and $m(u_\alpha), \alpha = 1, ..., n$ must be known. In addition, the $(n+1)$ by $(n+1)$ covariance matrix $\left[C(u_\alpha, u_\beta), \alpha, \beta = 0, 1, ..., n \right]$ with $u_0 = u$ are required in conducting the Simple Kriging. However, when the random function of $Z(u)$ is stationary with constant mean m, and covariance function $C(h) = C(u, u+h), \forall u$, Eq. 3.11 can be reduced to:

$$Z^*_{SK}(u) = \sum_{\alpha=1}^{n} \lambda_\alpha(u) Z(u_\alpha) + \left[1 - \sum_{\alpha=1}^{n} \lambda_\alpha(u) \right] m \tag{3.13}$$

with the Simple Kriging variance:

$$\sigma^2_{SK}(u) = C(0) - \sum_{\alpha=1}^{n} \lambda_\alpha(u) C(u - u_\alpha) \tag{3.14}$$

where $\sigma^2_{SK}(u)$ is Simple Kriging variance.

In a nutshell, the mean and variance of conditional probability distribution are calculated using Simple Kriging system by which the kriged estimated values represent the means and the Kriging variances represent the variances.

3.1.3.2 Sequential Gaussian Simulation Procedure

Sequential Gaussian Simulation is an estimation model defined under multigaussian assumption. The conditional probability distribution functions are fully characterized by their means and variances given by Simple Kriging System. The estimated means and variances honor both available data and simulated data. The procedure to execute the simulation is presented as follows:

1. Transform the data set into a standard normal distribution.
2. Check for Bivariate Normality of the normal score data. If the data do not meet the condition, other simulation should be considered.
3. Construct variograms analysis and fit a proper model for the data set.

4. Select grid node at random.
5. Perform Simple Kriging to estimate the mean and variance values at the visited node location.
6. Draw a simulated data from that distribution, and add the simulated data to the data set.
7. Select another grid node at random and repeat the procedure for Simple Kriging until all grid nodes are simulated.
8. Back transform the simulated data to the original space, and the realization map is created.
9. Provide different random number sequences for random visited nodes and repeat the same procedure for additional realization maps.

In summary, Sequential Gaussian Simulation is a high performance tool to estimate values at unsampling data locations using the concepts of geostatistical simulation. The flowchart of Sequential Gaussian Simulation procedure is depicted in Fig. 3.8. In practice, this technique involves a lot of calculation so it is impossible to perform it by hand. As a result, computer takes an important role in this part, because the computation time depends on the performance of computer.

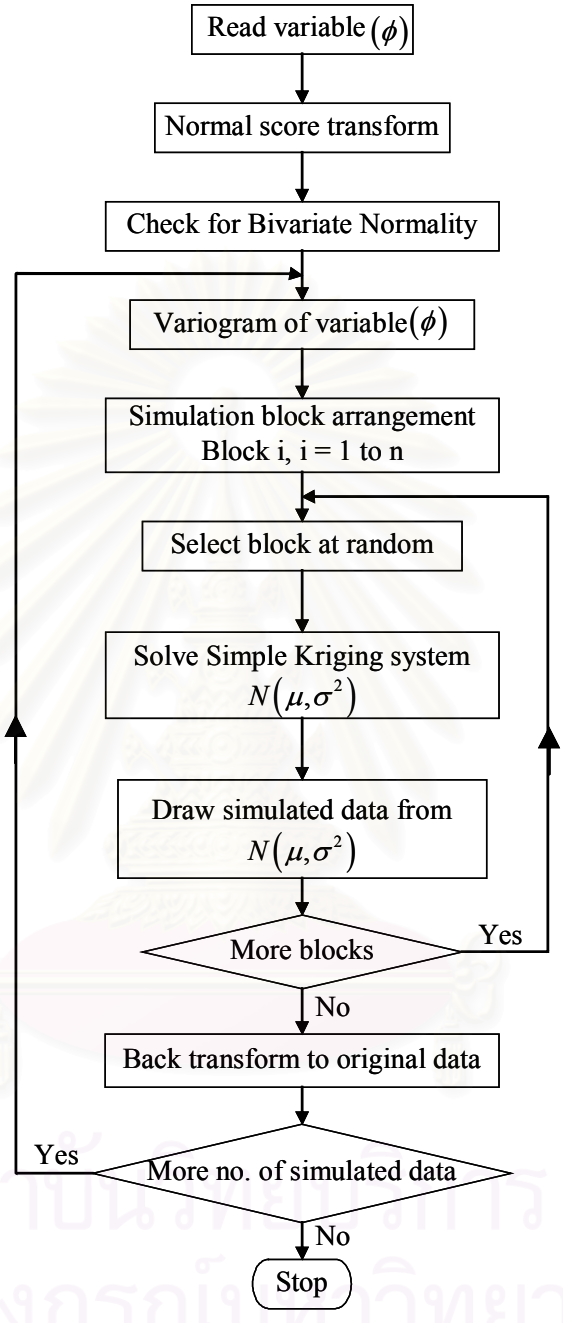

Figure 3.8: Sequential Gaussian Simulation process.

Geostatistical method is an effective way to estimate reservoir parameters such as permeability, net oil thickness, and porosity. In practice, the procedure of geostatistical estimation or simulation involves these steps of work: (1) data preparation and statistical analysis of the variable of interest; (2) structural analysis in

finding variogram model for the variable; (3) kriging estimation or stochastic simulation method selections.

3.2 Denoising using Wavelet Analysis

This section presents the principle of wavelet transform, multiresolution analysis, and denoising. Wavelet transform and multiresolution analysis can be used to break down a complex signal into approximated and detailed components via signal decomposition. The signal at original scale can be rebuilt from the decomposed approximated and detailed components via signal reconstruction. The denoising technique that is based on wavelet multiresolution analysis is applied during signal reconstruction to get rid of local variations in the data.

3.2.1 Wavelet Transform

Both Fourier transform and wavelet transform can be used to analyze signal or data but the advantage of wavelet transform is that the amount of localization in time and frequency is automatically adapted. As a result, the wavelet analysis is widely applied to many data analysis such as signal processing, image analysis, data compression, etc.

The definition of wavelet transform at time $x = k$ and dilation j is shown as follows:

$$Wf(j,k) = \frac{1}{\sqrt{j}} \int_{-\infty}^{\infty} f(x)\psi\left(\frac{x-k}{j}\right)dx \tag{3.15}$$

The function $\psi(x)$ is called a wavelet, which is a function that waves through the x-axis. The definition of wavelet function is presented as follows:

$$\int_{-\infty}^{\infty} \psi(x)dx = 0 \tag{3.16}$$

There are many wavelet functions that can be used in wavelet transform such as Haar wavelet, Biorthogonal wavelet, Meyer wavelet, Daubechies wavelet. To illustrate the principle of wavelet analysis, Haar wavelet may be a good example.

Although Haar function is not often used in practice, it is a simple function that helps us understand wavelet.

Haar wavelet was introduced in 1910 (Haar, 1910). The Haar function can be written as follows:

$$\psi(x) = \begin{cases} 1, & 0 \le x < \frac{1}{2} \\ -1 & \frac{1}{2} \le x < 1 \\ 0, & otherwise \end{cases} \tag{3.17}$$

From the equation, a sketch of the function is shown in Fig. 3.9.

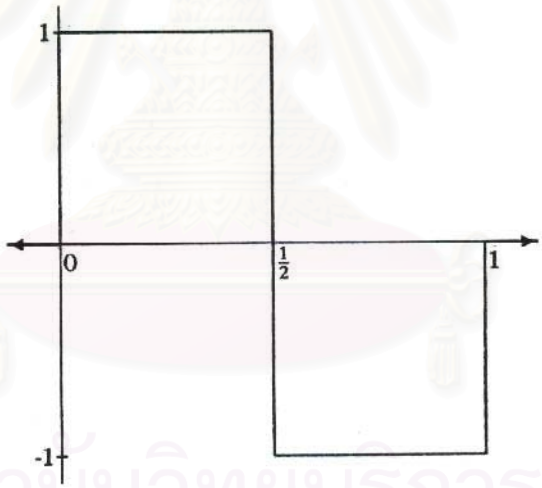

Figure 3.9: The Haar function (Ogden, 1997).

The function, ψ, is called a mother wavelet. The mother wavelet gives a family of wavelets by dyadic dilations and integer translations. Let j be a dilation index and k be translation index. Hence, each wavelet family is indexed by both of these indices. The family wavelets can be written as:

$$\psi_{j,k}(x) = 2^{\frac{j}{2}} \psi(2^j x - k) \tag{3.18}$$

for integer value j and k. Dilation by larger j compresses the function on x-axis and changing k shifts the function along the x-axis. As it was mentioned before that wavelet can describe both time-localized and frequency-localized information, a small value of j is used to analyze high-frequency information while a large value of j is used to capture low-frequency information. Some of these dilated and translated wavelet functions are depicted in Fig. 3.10.

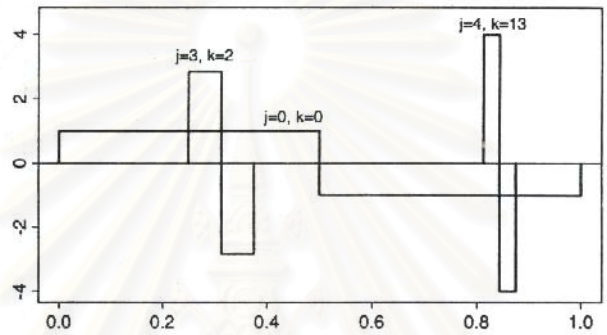

Figure 3.10: Haar wavelet examples (Ogden, 1997).

Another important function is the scaling function, $\phi(x)$, or father wavelet function. In data analysis by wavelet, this function captures low-frequency information while wavelet function captures high-frequency contents. The result of this analysis is described in the next section. Eq. 3.19 presents a simple form of scaling function, the Haar scaling function.

$$\phi(x) = \begin{cases} 1, & 0 \le x < 1 \\ 0, & otherwise \end{cases} \tag{3.19}$$

The dilated and translated scaling function is shown as the following equation:

$$\phi_{j,k}(x) = 2^{\frac{j}{2}} \phi(2^j x - k) \tag{3.20}$$

In summary, there are two important functions, wavelet function and scaling function, which are used in wavelet analysis. These two functions can adjust the frequency of analysis by adjusting parameter j and location of analysis by adjusting

parameter k. The wavelet function captures the high-frequency contents while the scaling function captures the low-frequency contents.

3.2.2 Multiresolution Analysis

The principle of multiresolution analysis, MRA, is to separate an original signal or function into two components, which are approximation part and detail part, and to recombine the approximation component and its detail component to the original signal or function. This analysis can be performed using wavelet function and scaling function. The process of separating is called decomposition. The combining process is called reconstruction. The reconstruction equation at level j_0 is shown as follows:

$$f(x) = \sum_k c_{j_0,k}\phi_{j_0,k}(x) + \sum_{j \geq j_0}\sum_k d_{j,k}\psi_{j,k}(x) \tag{3.21}$$

where

$f(x)$ = a signal or function

$\sum_k c_{j_0,k}\phi_{j_0,k}(x)$ = approximation part at level j_0

$\sum_{j \geq j_0}\sum_k d_{j,k}\psi_{j,k}(x)$ = detail part at level j_0

$c_{j_0,k}$ = coefficients of scaling function at level j_0

$d_{j,k}$ = coefficients of wavelet function at level j

The coefficient of scaling function and wavelet function can be calculated as follows:

$$c_{j,k} = \langle f, \phi_{j,k} \rangle = \int_{-\infty}^{\infty} f(x)\phi_{j,k}(x)dx \tag{3.22}$$

$$d_{j,k} = \langle f, \psi_{j,k} \rangle = \int_{-\infty}^{\infty} f(x)\psi_{j,k}(x)dx \tag{3.23}$$

In Eq. 3.21, when j decreases, the approximation signal is coarser. The coarsest level is j_0. On the other hand, when j increases, the approximation signal is finer. The parameter k in Eq. 3.21 is the location of the data. In decomposition

framework, original signal or function is first decomposed to an approximation and a detail part at level one. The approximation part of level one can then be decomposed again to an approximation, which is coarser than the previous one, of level two and detail of level two. An approximation can be decomposed as many times as desired and the approximation at the next level is coarser than the previous level. The following figure shows approximations of a signal decomposed using decompostion process.

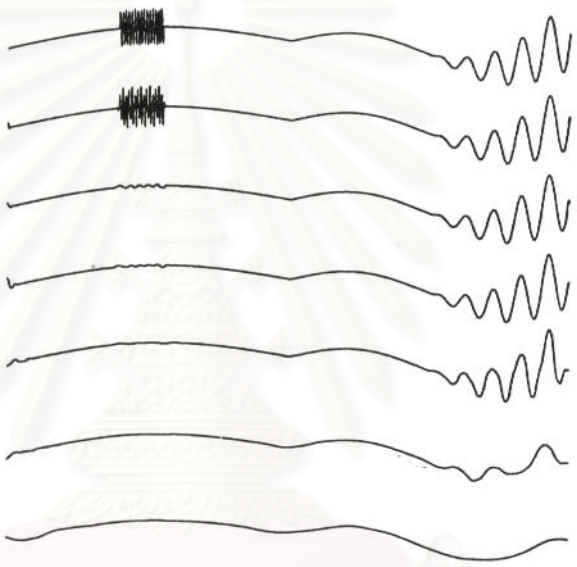

Figure 3.11: Approximations of a signal decomposed by wavelet decomposition (Ogden, 1997).

Depicted in Fig. 3.11, the upper most signal is the original signal and the second signal is its approximation at the next level. Likewise, the next lower signals are approximations of the upper ones. Fig. 3.12 shows the detail part of the signal decomposed in Fig. 3.11.

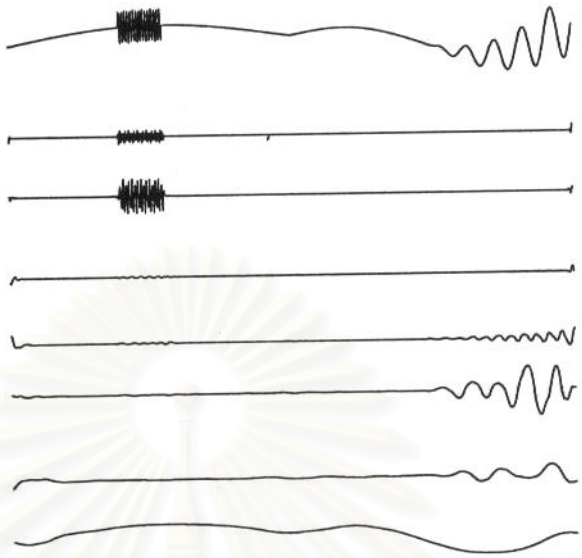

Figure 3.12: Details of a signal decomposed by wavelet decomposition (Ogden, 1997).

The uppermost signal in Fig. 3.12 is the original signal and the bottom signal is the approximation at the last level of decomposition in Fig. 3.11. The signals in between are detail components at different levels of decomposition.

In the reconstruction part, an approximation at any level can be reconstructed using Eq. 3.21 from a coarser approximation and detail components. This process can be implemented until the original signal is obtained.

Fig. 3.13 and 3.14 sketch the mathematical principle of decomposition algorithm and reconstruction algorithm, respectively.

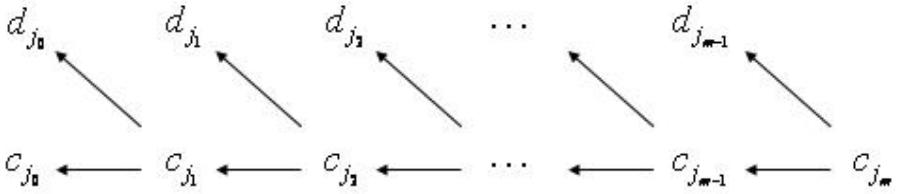

Figure 3.13: Schematic representation of the decomposition algorithm.

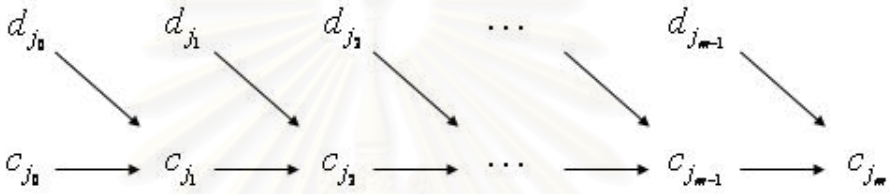

Figure 3.14: Schematic representation of the reconstruction algorithm.

Fig. 3.13 shows the decomposition of a signal from level j_m to j_0, i.e., from the finest resolution to the coarsest one. The detail components represented by the symbol d with a subscript j and the approximation components represented by the symbol c with a subscript j are decomposed from the previous approximation component. As seen in the figure, $c_{j_{m-1}}$ and $d_{j_{m-1}}$ are separated from the approximation component c_{j_m}. In the next decomposition, $c_{j_{m-2}}$ and $d_{j_{m-2}}$ are separated from the approximation component $c_{j_{m-1}}$. This process goes on until the level j_0. Fig. 3.14 shows the reconstruction of a signal. The reconstruction process starts from combining the coarsest approximation component c_{j_0} and detail component d_{j_0} to form the approximation component c_{j_1}. When the approximation component c_{j_1} and detail component d_{j_1} are combined, the result is approximation component c_{j_2}. This reconstruction process is performed until the approximation component c_{j_m} is composed.

Roughly speaking, multiresolution analysis finds the representation of a signal or function at different scales while preserving the underlying features of the original

signal via a decomposition algorithm. On the other hand, it can reconstruct the original signal from an approximation and its detail components via a reconstruction algorithm. Being able to separate detail components from the underlying features of the signal allows us to denoise the signal, which is discussed in the next section.

3.2.3 Denoising

In practice, noise generally exists in data collection process. Hence, it may deviate the results of data analysis. Thus, before doing any analysis, noise should be eliminated. There are many methods to exclude noise but one of the most effective methods is wavelet thresholding method.

The principle of wavelet thresholding is to choose appropriate wavelet coefficients to represent the underlying structures in the signal. In general, large wavelet coefficients should be included in a selective reconstruction while small coefficients should not. Thus, a threshold value must be specified to get rid of the coefficients which are lower than the desired value. The principle of hard thresholding method is that the random noise contains in the small coefficients is eliminated while the large coefficients are kept. However, the principle of the soft thresholding method is that each wavelet coefficient consists of both a signal portion and a noise portion. So, it is desirable to isolate the signal portion by removing the noisy part. The following equations are the keep or kill detail coefficient of hard thresholding and soft thresholding, respectively.

$$d_{j,k}^{H} = \begin{cases} d_{j,k}, & if \left| d_{j,k} \right| > \lambda_j \\ 0, & otherwise \end{cases} \tag{3.24}$$

$$d_{j,k}^{S} = \begin{cases} d_{j,k} - \lambda_j, & if\ d_{j,k} > \lambda_j \\ 0, & if \left| d_{j,k} \right| \le \lambda_j \\ d_{j,k} + \lambda_j, & if\ d_{j,k} < -\lambda_j \end{cases} \tag{3.25}$$

where

$d_{j,k}^{H}$ = detail coefficient of hard thresholding at level j

$d_{j,k}^{S}$ = detail coefficient of soft thresholding at level j

λ_j = threshold value at level j

In the hard thresholding method (Eq. 3.24), detail coefficients larger than the absolute threshold value λ are kept the same while coefficients smaller than the absolute threshold value are set to zero. On the other hand, in the soft thresholding method (Eq. 3.25), detail coefficients larger than the absolute threshold value are shrunken by the absolute threshold value and coefficients smaller than the absolute threshold value are set to zero. Fig. 3.15 graphically shows the hard and soft thresholding functions.

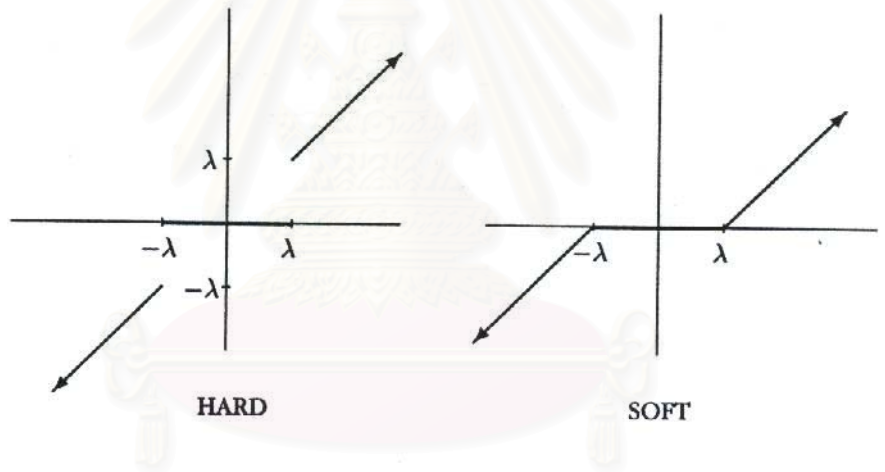

Figure 3.15: The hard and soft thresholding functions (Ogden, 1997).

Fig. 3.15 depicts the functions used in the hard and soft thresholding methods. The x-axis is the detail coefficient before denoising and the y-axis is the value of detail coefficient after denoising. Show in the hard thresholding function in Fig. 3.15, the absolute detail coefficients are set to zero if the coefficients are smaller than the absolute threshold value and are kept the same if the coefficients are larger than the absolute threshold value. Shown in the soft thresholding function in Fig. 3.15, the absolute detail coefficients are set to zero if the coefficients are smaller than the absolute threshold value and are shrunken by the absolute threshold if the coefficients are larger than the absolute threshold value.

The advantage of the hard thresholding method is that it can preserve the signal sharpness in the region of discontinuities in the signal. However, it cannot suppress a few noisy points in the continuous area of the signal while the soft thresholding method can.

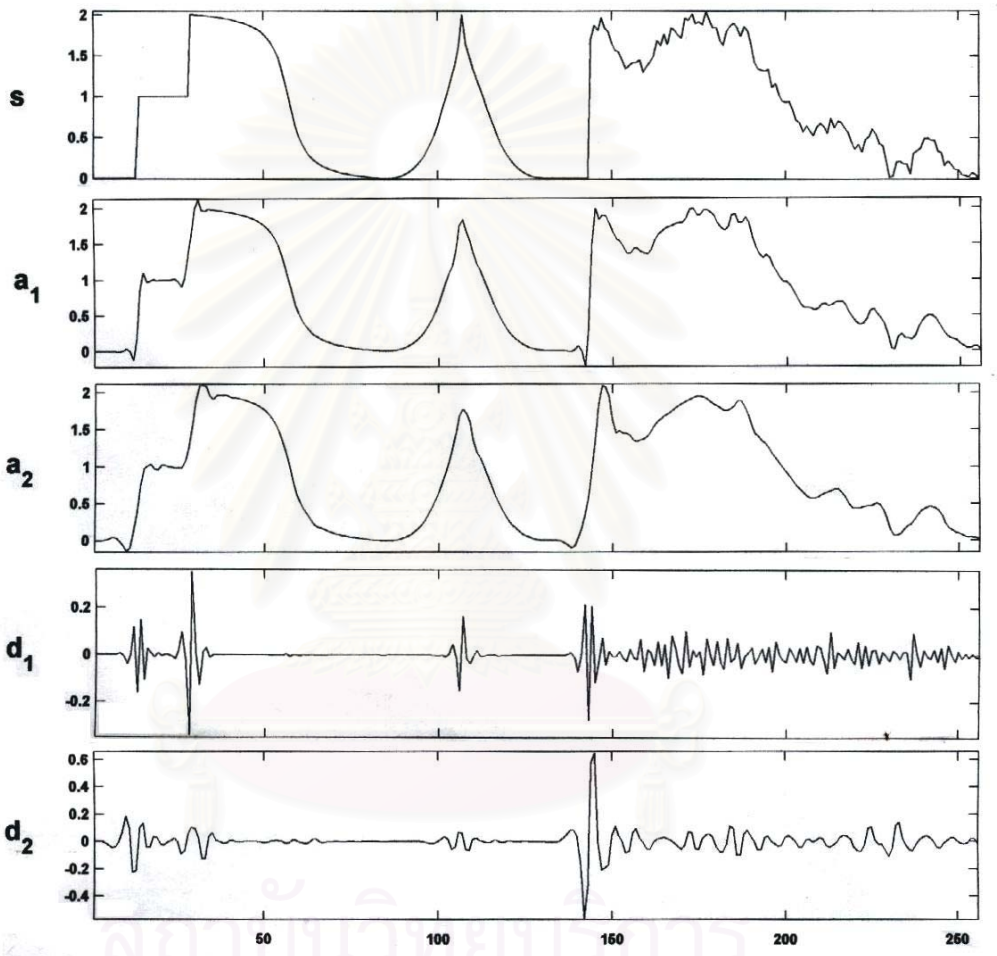

Figure 3.16: Decomposition of a signal using Daubechies 4 for 2 levels.

Fig. 3.16 illustrates the decomposition of a signal using Daubechies 4 wavelet. In Fig. 3.16, s is an original signal and a_2 is an approximation at levels 2 of decomposition. d_1 and d_2 are the detail signals of the original signal at level 1 and 2, respectively. In reconstruction process, a_2 and d_2 can be used to reconstruct an approximation at level 1 (a_1). The approximation at level 1 (a_1) and d_1 then can be combined to reconstruct the original signal. In case of denoising, some coefficients of

which magnitudes are lower than the threshold value, of detail parts at both levels 1 and 2 are set to zero in both soft and hard thresholding methods and the coefficients whose magnitudes are higher than the threshold value are shrunken in the soft thresholding method. The graphical display of the thresholding is shown in Fig. 3.17.

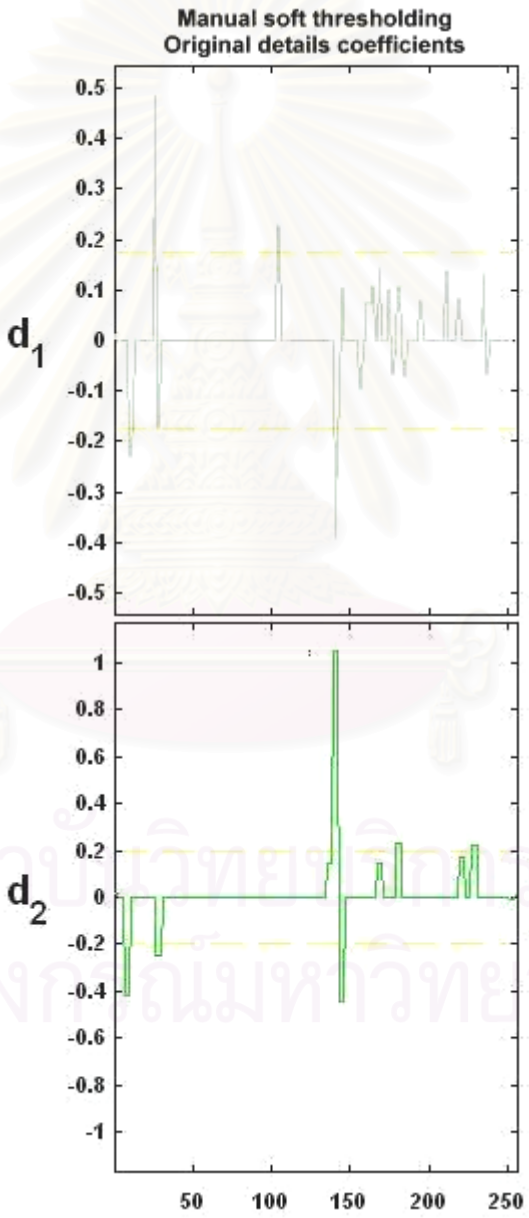

Figure 3.17: Thresholding detail signals.

Fig. 3.17 shows detail components at levels 1 and 2. The denoising of a signal is done by specifying threshold values for detail coefficients seen as dashed lines in the figure. The coefficients having absolute values less than the thresholds, i.e., those lying between the dashed lines, are set to zero. The coefficients having absolute values higher than the threshold, i.e., those lying above the dashed lines, are shrunken in the soft thresholding method or set to zero in the hard thresholding method. After detail coefficients are adjusted, approximation and adjusted details are then combined using signal reconstruction. Fig. 3.18 shows the original signal and its denoised signal.

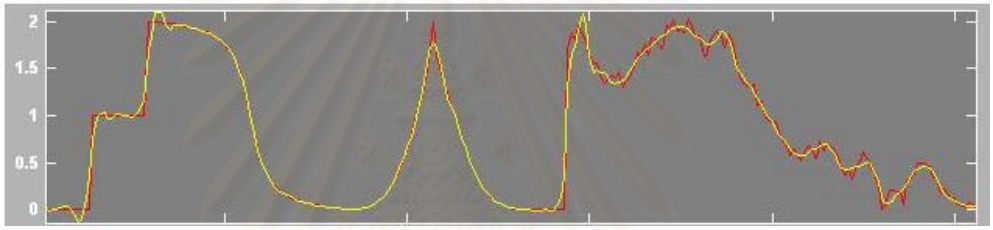

Figure 3.18: Original and denoised signals.

The red signal in Fig. 3.18 is the original signal and the yellow signal is the denoised one. One can see that the yellow signal is smoother than the red one but still preserves the main structure of the original. However, the yellow signal shows a fluctuation in the discontinuity part of the original signal, shown in the left of Fig. 3.18, due to the flaw of the soft thresholding method. The threshold value is an important parameter because the amount of noise eliminated depends on that value. In case that the threshold value is high, much of noise is ruled out but parts of the signal may be smeared out. If the threshold value is low, little of noise is eliminated.

In practice, there is a rule for calculating the threshold values. The equation according to Kikani, J. and He, M. (1998) is expressed as

$$\lambda = \hat{\sigma}\sqrt{2\log n} \tag{3.26}$$

where

$\hat{\sigma}$ = the median of absolute deviation (MAD)

n = the number of data

In Eq. 3.26, the median of absolute deviation (MAD) is the median of absolute deviation of all detail coefficients which can be calculated by Eq. 3.27 (Donoho and Johnstone, 1995). The number of data n and the median of absolute deviation (MAD) are different for different levels of decomposition because of the change in resolution. Therefore, the thresholding value is different for different levels of decomposition.

$$\hat{\sigma} = \frac{median\left(\left|d_{J-1,k}^{(n)} - median\left(d_{J-1,k}^{(n)}\right)\right|\right)}{0.6745} \tag{3.27}$$

where $J = \log_2(n)$.

Briefly speaking, inherent noise in a data set can be eliminated by performing wavelet multiresolution analysis to decompose the data into an approximation component and detail components. Then, threshold values are specified to keep or kill the detail coefficients before reconstructing the new and smooth signal. As a result, the features of the data are preserved while noise is taken out. In practice, there are two techniques for denoising, hard and soft thresholding methods. The advantage of hard thresholding method is that it can preserve the signal sharpness at discontinuities in the signal. However, it cannot suppress a few noisy points in the continuous area of the signal while the soft thresholding method can. The drawback of the soft thresholding method is that the denoised signal may be too smooth.

3.3 Correlation Analysis

Correlation analysis concerns with the relation between X and Y in two-dimensional random variables (X, Y). Let's consider a set of variable consisting of n ordered pairs of values $(x_1, y_1),...,(x_n, y_n)$. The sample means, variances and covariance are shown respectively as follows:

- sample means

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{n} x_j \tag{3.28}$$

$$\bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j \tag{3.29}$$

where

$\bar{x}$ = mean of variable x

$\bar{y}$ = mean of variable y

- sample variances

$$S_x^2 = \frac{1}{n-1}\sum_{j=1}^{n} (x_j - \bar{x})^2 \tag{3.30}$$

$$S_y^2 = \frac{1}{n-1}\sum_{j=1}^{n} (y_j - \bar{y})^2 \tag{3.31}$$

where

S_x^2 = variance of variable x

S_y^2 = variance of variable y

- sample covariance

$$S_{xy} = \frac{1}{n-1}\sum_{j=1}^{n} (x_j - \bar{x})(y_j - \bar{y}) \tag{3.32}$$

where

S_{xy} = covariance of variable x and y

The sample correlation coefficient is

$$r = \frac{S_{xy}}{S_x S_y} \tag{3.33}$$

where

r = correlation coefficient of variable x and y

As shown in the above equations, both S_{xy} and r measure the interrelation between the X and Y values. However, r has the advantage that it does not change under a multiplication of these values by a factor.

Correlation analysis is a way to measure degree of similarity between two variables. The principle of this analysis is to compare between covariance of two variables and multiple of each variable variance. In this study, correlation analysis is used to find the degree of similarity between realizations generated from Sequential Gaussian Simulation.

CHAPTER IV

SIMULATIONS AND RESULTS

This chapter composes of four topics, which are generating raw data, simulation of realizations, denoising of realizations, and grouping of similar denoised realizations. In this study, the raw data, which are 47 porosity sampling data from 35 vertical wells and 3 horizontal wells, were artificially created. The study area covers 10,240 feet in north-south direction and 10,240 feet in east-west direction. After the raw data had been composed, 60 realizations (2D) maps were generated using Sequential Gaussian Simulation (SGS). The next step is denoising the realizations, which are performed by wavelet functions in MATLAB program. Finally, grouping similar denoised realizations are conducted by MATLAB program.

4.1 Generating raw data

To study the phenomenon of realizations in two-dimension from geostatistics simulation, a data set is needed. Since valuable data are confidential, artificial data set was generated for this study. There are two principles in generating the data, which are well spacing and value of variable of interest, which is porosity. In this study, a geological condition of sandstone formation located onshore is chosen to be the case study of this thesis because petroleum is mostly found in this kind of formation. The well spacing for onshore drilling is generally closer than offshore drilling due to the cost of drilling, which facilitates the examination of the porosity spatial structure. In general, well spacing of onshore drilling is about 500 meters or more and porosity value of fine to medium grain size sandstone is between 0.05 and 0.32, which is close to the real data according to Hand et. al. (1994). The study domain has a square area of which each side can be divided into 2^n sections because the area will then be analyzed by multiresolution wavelet analysis in the denoising step, which requires the number of grid cells to be 2^n, where n is an integer. The geology condition of this study is that reservoir formation is fine to medium grain size sandstone having no

major direction of influence because, in reality, the influence ranges are difficult to find due to well spacing limitation. Since variogram values at small distances are crucial in variogram modeling step, data pairs with distance smaller than the well spacing are required. Therefore, horizontal wells having the shorter distance between data than well spacing, are included in this data set. From these principles, 47 porosity sampling data from 35 regular wells and 3 horizontal wells are made up within the study area of 10,240x10,240 ft^2 in north-south and east-west direction respectively, which can be divided into 256x256 blocks with the size of 40x40 ft^2. The following data are the initial data for this study.

Table 4.1: The locations and porosity values of raw data.

X-coordinate (Feet)	Y-coordinate (Feet)	Porosity values	X-coordinate (Feet)	Y-coordinate (Feet)	Porosity values
4,260	10,020	0.31	4,420	1,740	0.13
1,340	9,260	0.26	9,540	1,460	0.23
4,820	9,580	0.22	1,420	7,260	0.32
8,940	9,940	0.14	3,260	6,140	0.18
2,620	8,540	0.31	5,620	4,540	0.20
6,140	8,740	0.16	7,220	5,220	0.17
7,540	8,420	0.20	8,820	8,140	0.16
3,060	7,980	0.07	4,500	8,140	0.12
9,140	7,140	0.29	3,260	2,260	0.17
5,180	6,460	0.27	9,500	4,020	0.24
6,900	6,660	0.05	4,260	3,420	0.05
8,900	6,660	0.20	1,420	4,140	0.10
1,900	5,620	0.05	7,420	8,220	0.26
4,140	5,060	0.16	7,140	8,060	0.26
8,300	4,140	0.12	6,940	8,020	0.19
9,940	4,940	0.16	5,140	6,660	0.29
3,140	3,940	0.21	5,140	7,060	0.17
6,260	3,100	0.16	5,140	7,260	0.21
1,540	2,820	0.17	4,260	3,220	0.07
5,460	2,500	0.26	4,260	3,020	0.09
7,220	2,980	0.26	4,260	2,820	0.11
8,900	2,660	0.22	7,220	1,740	0.21
10,220	2,740	0.30	6,540	780	0.14
			7,220	10,220	0.11

Table 4.1 depicts 47 porosity data at different locations. The $X-Y$ coordinates shown in the table are measured from the origin at the bottom left of the study area. Figure 4.1 shows the location map of the data set.

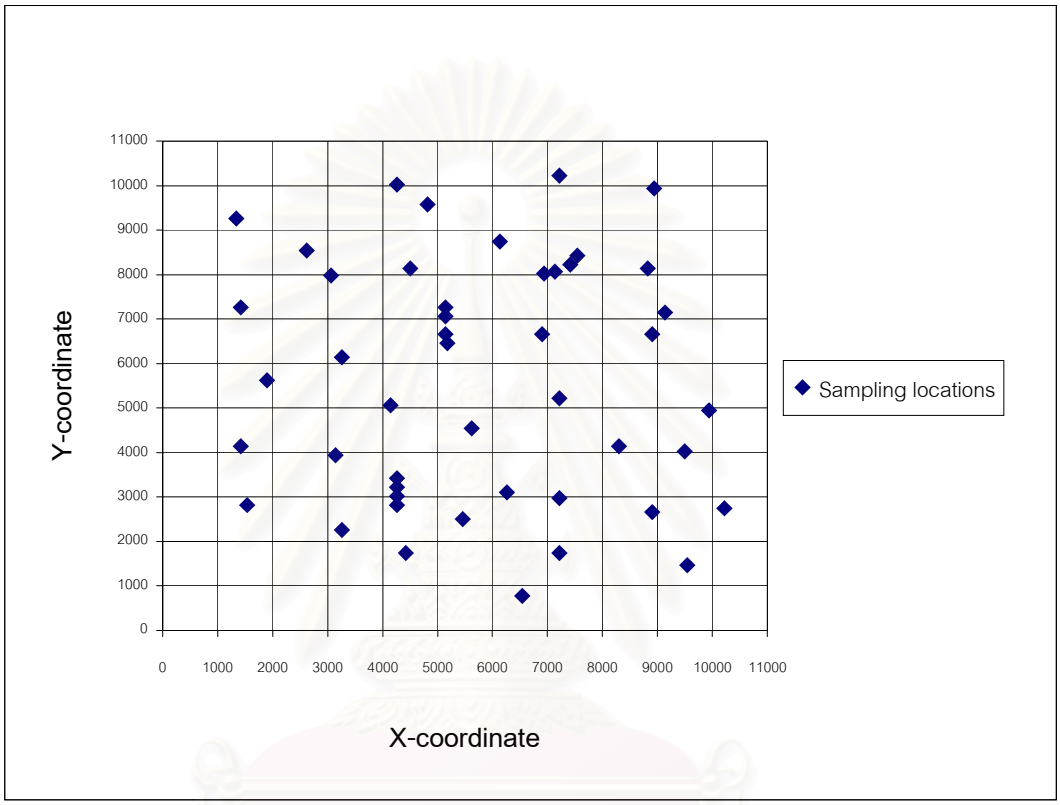

Figure 4.1: Location map of porosity data.

As seen in Fig. 4.1, vertical wells scatter around the study area and three horizontal wells are drilled around the middle part of the area, which can be recognized from close data locations.

Table 4.2 gives a statistical analysis of the porosity data set. The mean of the data is 0.186 and variance and standard deviation are 0.006 and 0.075, respectively. The minimum and maximum value of the data set are 0.050 and 0.320, respectively. The first, second, and third quartiles are 0.127, 0.180, and 0.245, respectively. Coefficient of variation, skewness, and kurtosis are 0.40, -0.051, and 2.160, respectively.

Table 4.2: Statistics of porosity data.

Parameters	values
Mean	0.186
Variance	0.006
Std. Dev	0.075
Minimum	0.050
25^{th} %	0.127
Median	0.180
75^{th} %	0.245
Maximum	0.320
Coefficient of variation	0.40
Skewness	-0.051
Kurtosis	2.160

The following figure is the histogram of porosity data.

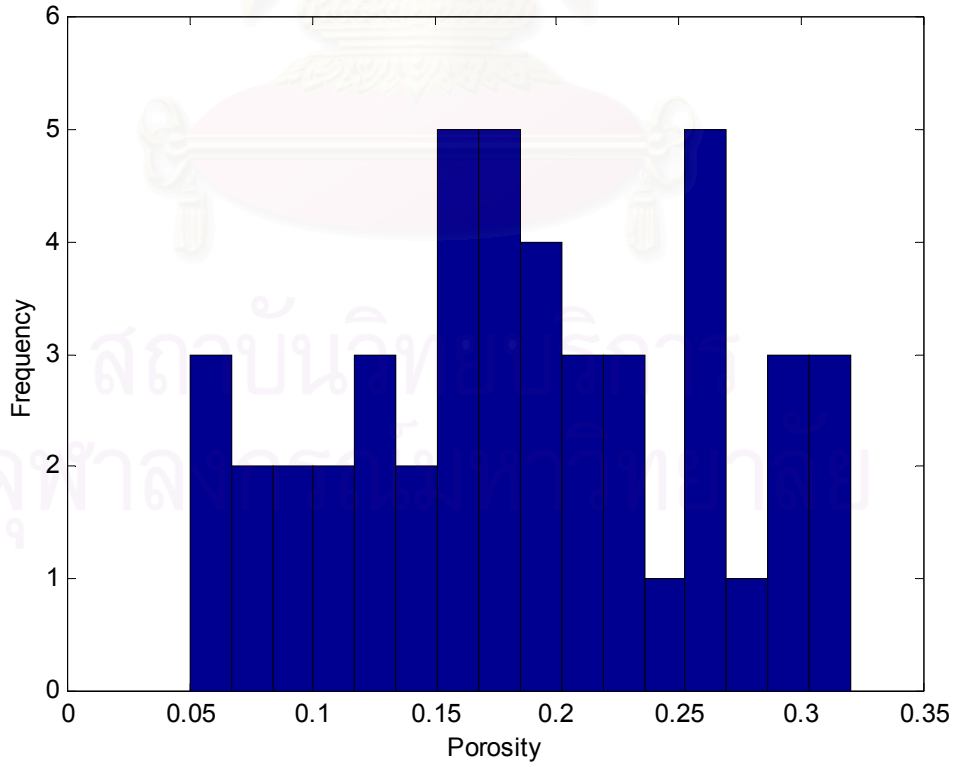

Figure 4.2: Histogram of porosity data.

Fig. 4.2 shows the histogram of porosity data in the study area. As seen in the figure, values of the data spread from 0.05 to 0.32 and have a high frequency at around 0.15 to 0.19. The coefficient of variation records lower than one (0.40) indicates no presence of some erratic high sample values that may have significant impact to the simulation process. The rather small negative skewness describes the sample distribution with a few small values presence at the left of the histogram. It can be said that the histogram is approximately symmetric with the skewness (-0.051) close to zero and the median (0.180) is close to the mean (0.186). The positive kurtosis of 2.160 indicates a peaked distribution.

This 2D porosity data is used as an input in performing Sequential Gaussian Simulation, which will be carried out in the next step.

4.2 Simulation of realizations

This part presents the procedure of conducting the geostatistics simulation of the porosity data generated from the previous step. In this study, Sequential Gaussian Simulation is chosen because it gives good results for homogeneous variable such as porosity, and it is widely used to estimate porosity. This simulation was also presented in Hand et. al. (1994) to find porosity distribution for each facie of Romeo zonation in Prodhoe Bay.

The simulation process can be divided into 2 steps, which are finding the spatial variability structure of porosity variable (known as variogram model), and conducting the simulation. The first step, constructing the variogram model, is executed by Variowin computer program, which is a good graphic display computer program that can show analysts to see how fit of a model in comparison with the calculated variogram values. The second step is performing the Sequential Gaussian Simulation, which is implemented by GSLIB program.

4.2.1 Constructing the variogram model

To perform the Sequential Gaussian Simulation, the sample values have to be transformed to normal score data before further analysis due to its assumption of

multigaussian distribution. In this study, the data are transformed by GSLIB program. The normal score data is tabulated in Table 4.3.

Table 4.3: The locations, porosity values, and normal score values.

X coordinate (Feet)	Y coordinate (Feet)	Porosity values	Normal score values
4,260	10,020	0.31	1.615
1,340	9,260	0.26	0.691
4,820	9,580	0.22	0.441
8,940	9,940	0.14	-0.625
2,620	8,540	0.31	1.853
6,140	8,740	0.16	-0.383
7,540	8,420	0.20	0.161
3,060	7,980	0.07	-1.306
9,140	7,140	0.29	1.306
5,180	6,460	0.27	1.088
6,900	6,660	0.05	-1.853
8,900	6,660	0.20	0.215
1,900	5,620	0.05	-1.615
4,140	5,060	0.16	-0.441
8,300	4,140	0.12	-0.761
9,940	4,940	0.16	-0.270
3,140	3,940	0.21	0.382
6,260	3,100	0.16	-0.326
1,540	2,820	0.17	-0.161
5,460	2,500	0.26	0.834
7,220	2,980	0.26	0.761
8,900	2,660	0.22	0.500
10,220	2,740	0.30	1.443
4,420	1,740	0.13	-0.691
9,540	1,460	0.23	0.562
1,420	7,260	0.32	2.303
3,260	6,140	0.18	0.000
5,620	4,540	0.20	0.107
7,220	5,220	0.17	-0.215

Table 4.3: The locations, porosity values, and normal score values (continued).

X-coordinate	Y-coordinate	Porosity values	Normal score values
4,500	8,140	0.12	-0.834
3,260	2,260	0.17	-0.053
9,500	4,020	0.24	0.625
4,260	3,420	0.05	-2.303
1,420	4,140	0.10	-1.088
7,420	8,220	0.26	0.996
7,140	8,060	0.26	0.912
6,940	8,020	0.19	0.053
5,140	6,660	0.29	1.190
5,140	7,060	0.17	-0.107
5,140	7,260	0.21	0.270
4,260	3,220	0.07	-1.443
4,260	3,020	0.09	-1.190
4,260	2,820	0.11	-0.996
7,220	1,740	0.21	0.326
6,540	780	0.14	-0.562
7,220	10,220	0.11	-0.912

Table 4.3 shows the normal score values in the fourth column of the table of the original data at all locations.

Table 4.4 shows the statistics of the normal score data, which were transformed from the original porosity data. The new data have a mean of zero and variance and standard deviation of one, which are the characteristics of standard normal distribution. The minimum and maximum values are -2.300 and 2.303, respectively. The first, second, and third quartiles are -0.709, 0.000, and 0.642, respectively. Coefficient of variation, skewness, and kurtosis are 0, 0, and 2.718, respectively.

Table 4.4: The statistical analysis of the normal score data.

Parameters	values
Mean	0.000
Variance	0.994
Std. Dev	0.997
Minimum	-2.300
25^{th} %	-0.709
Median	0.000
75^{th} %	0.642
Maximum	2.303
Coefficient of variation	0
Skewness	0
Kurtosis	2.718

The plot of histogram of the normal score porosity data is shown in Fig. 4.3.

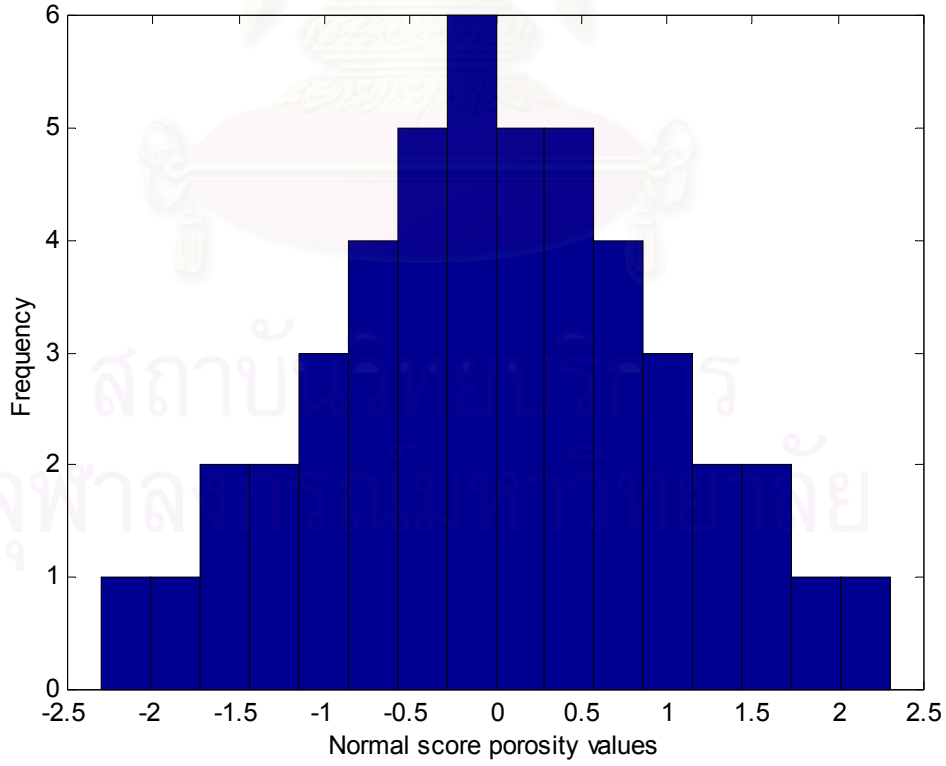

Figure 4.3: Histogram of the normal score porosity data.

Fig. 4.3 shows the histogram of the normal score porosity data. As it can be observed, the normal score porosity follows the standard normal distribution.

After the normal score data were prepared, variogram calculations of these data and a variogram modeling are constructed using the Variowin program. Since there are a few data due to large well spacings, variogram analysis in many directions may give improper results. Therefore, this study used omnidirectional variogram which can represent variogram in all directions. Fig. 4.4 illustrates the plot of the experimental variogram values at difference distances and its model.

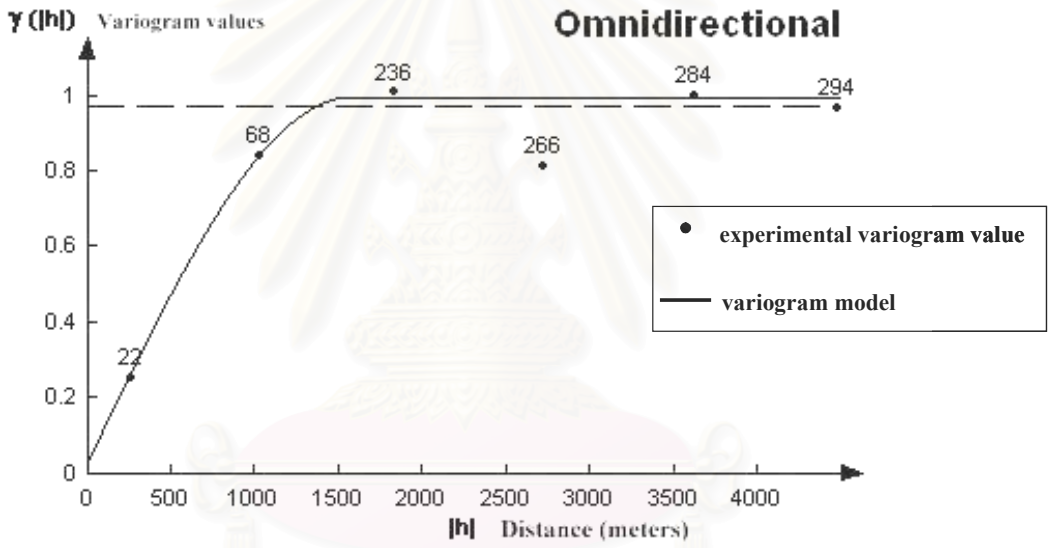

Figure 4.4: The omnidirectional variogram plot and its variogram model for normal score data.

Fig. 4.4 shows the experimental variogram values and its model of normal score data. The parameters used to calculate this variogram plot are: 900 feet of lag spacing, 450 feet of lag tolerance, 5 lags, 0 degree of direction, 90 degree of angular tolerance, and no limits of maximum bandwidth. In Fig. 4.4, the numbers shown near the black dots are the number of pairs that are used in the calculation for each lag distance and the black solid line is the variogram model. The result of fitting the variogram model is shown in the Table 4.5.

Table 4.5: Variogram model parameters of normal score data.

Parameters	Values
Model	Spherical
Nugget	0.03
Range	1,570 feet
Sill	0.97

Table 4.5 shows the model parameters of variogram, which are spherical model with 0.03 nugget, 1,570 feet range, and 0.97 sill.

The variogram model exhibits a small nugget effect, with the nugget value of 0.03 which is approximately 3 percent of the sill value. The normal score porosity data yields a correlation distance of 1,570 feet defined within the range distance and represented all direction. The sill value (0.97) almost equals to the normal score porosity data variance (1.00). In overall, this variogram model represents the spatial variability structure of the transform porosity data, and it will be used as the conditioning information in the simulation process.

In reality, variogram calculation is an important step because the accuracy of the estimated values or realizations mostly depends on the variogram model. Hence, many variogram models were tried when fitting the variogram plot to find the best fit model. After a variogram model is found, there is a condition that the data set has to meet the bivariate normal distribution before conducting Sequential Gaussian Simulation.

4.2.2 Checking for Bivariate Normality

Before applying Sequential Gaussian Simulation to this data set, there is a condition that the data set has to follow bivariate normal distribution. Therefore, the process of checking for Bivariate Normality is necessary. In practice, the checking process is carried out by comparing the theoretical variogram of Bivariate Gaussian model with experimental indicator variogram at several cut-off values, such as second quartile, median, and third quartile. In this study, the median is chosen to be the cut-

off value to examine the Bivariate Normality around the average of data set. In the checking process, the experimental indicator variogram corresponding to a specific cut-off, median cut-off in this case, is compared to the theoretical variogram calculated from Eq. 3.9. The procedure for this check can be elaborated as follows:

1. Calculating an experimental indicator variogram at median cut-off, which is 0.18 for this study.
2. Calculating the theoretical indicator variogram of Bivariate Gaussian model at median cut-off using Eq. 3.9.
3. Comparison of the two indicator variograms, which obtain from procedure number one and two.

In the comparison of these two variograms, some parameters for variogram calculation are specified as shown in Table 4.6.

Table 4.6: Variogram parameters using for checking Bivariate Normality.

Parameters	Values
Number of lag	40
Lag spacing	100
Cut-off	0.18

Table 4.6 shows parameters used in variogram calculation, which are set to be the same for both experimental and gaussian model indicator variograms. The cut-off porosity value is equal to median, which is 0.18.

Fig. 4.5 shows the experimental and gaussian model indicator variograms corresponding to the second quartile, median. As seen in the figure, there is a good correspondence between experimental indicator variogram at median cut-off and theoretical indicator variogram of Bivariate Gaussian model. That means Sequential Gaussian Simulation can be used for this data set.

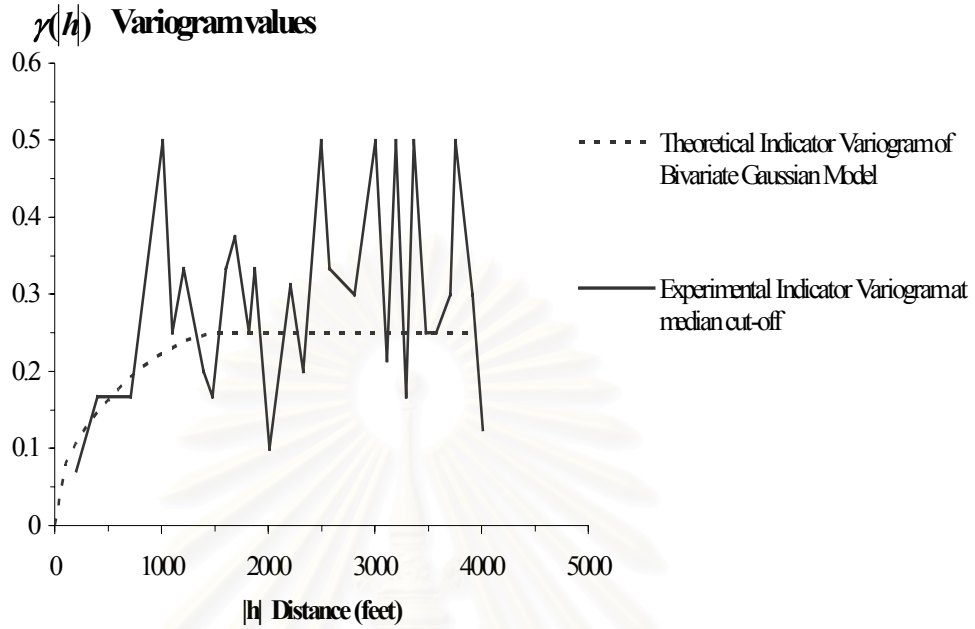

Figure 4.5: Experimental indicator variogram and gaussian model-derived indicator variogram at median cut-off.

4.2.3 Sequential Gaussian Simulation

Sequential Gaussian Simulation can generate many realizations at equal probability from the same data set. The probability distribution (*ccdf*) of the random visited node is constructed by Simple Kriging process, conditioned to the original data and previously simulated data. Then, the realization at the visited node is generated using the random number generator and the constructed *ccdf*. In practice, several realizations are generated to examine the spatial variability structure of a data set. As a result, it takes a lot of time and uses a lot of resources to perform reservoir simulation for multiple realizations. The number of realizations generated in this study are sixty. This is because the comparison process needs many realizations to demonstrate the grouping technique. In this study, 60 realization maps are generated by Sequential Gaussian Simulation of GSLIB program executed in Fortran PowerStation 4.0. The following figure is the parameters file used in simulation.

```
                  Parameters for SGSIM
                  ********************

START OF PARAMETERS:
data.dat                                  \data file
1   2   0   3   0                         \column: x,y,z,vr,wt
-1.0e21    1.0e21                         \data trimming limits
0                                         \0=transform the data, 1=don't
data.trn                                  \  output transformation table
0.05    0.35                               \  zmin,zmax(tail extrapolation)
1       0.0                               \  lower tail option, parameter
1       0.0                               \  upper tail option, parameter
sim2.out                                   \output File for simulation
1                                         \debugging level: 0,1,2,3
sim.dbg                                   \output File for Debugging
222179                                    \random number seed
0                                         \kriging type (0=SK, 1=OK)
60                                          \number of simulations
256   20     40.0                          \nx,xmn,xsiz
256   20     40.0                          \ny,ymn,ysiz
1    0.0     1.0                           \nz,zmn,zsiz
0                                         \0=two part search, 1=data-nodes
2                                         \max per octant(0 -> not used)
1570.0                                      \maximum search radius
0.0  0.0  0.0  1.0  1.0                   \sang1,sang2,sang3,sanis1,2
0    10                                    \min, max data for simulation
10                                         \number simulated nodes to use
1    0.03                                  \nst, nugget effect
1   1570.0   0.97                           \it,  aa,  cc
0.0  0.0   0.0  1.0 1.0                   \ang1,ang2,ang3,anis1,anis2:
```

Figure 4.6: Parameters file of the Sequential Gaussian Simulation.

The parameters shown in Fig. 4.6 provide information that the simulation will simulate 60 realization maps of data set read from "data.dat" file for 256×256 blocks with the size of 40×40 ft^2 in X and Y dimension. The Simple Kriging subroutine is used to calculate the mean and variance of the probability function. The searching subroutine is defined as two parts search, which means the data and previously simulated grid nodes are searched separately with 1,570 feet of maximum search radius. The variogram model used in this simulation is omnidirectional variogram with a nugget effect of 0.03, range of 1,570 feet, and sill of 0.97. The maximum data and simulated nodes used in each node estimation are both 10.

The results of 60 realization maps of porosity data are shown in Fig. 4.7 through 4.14.

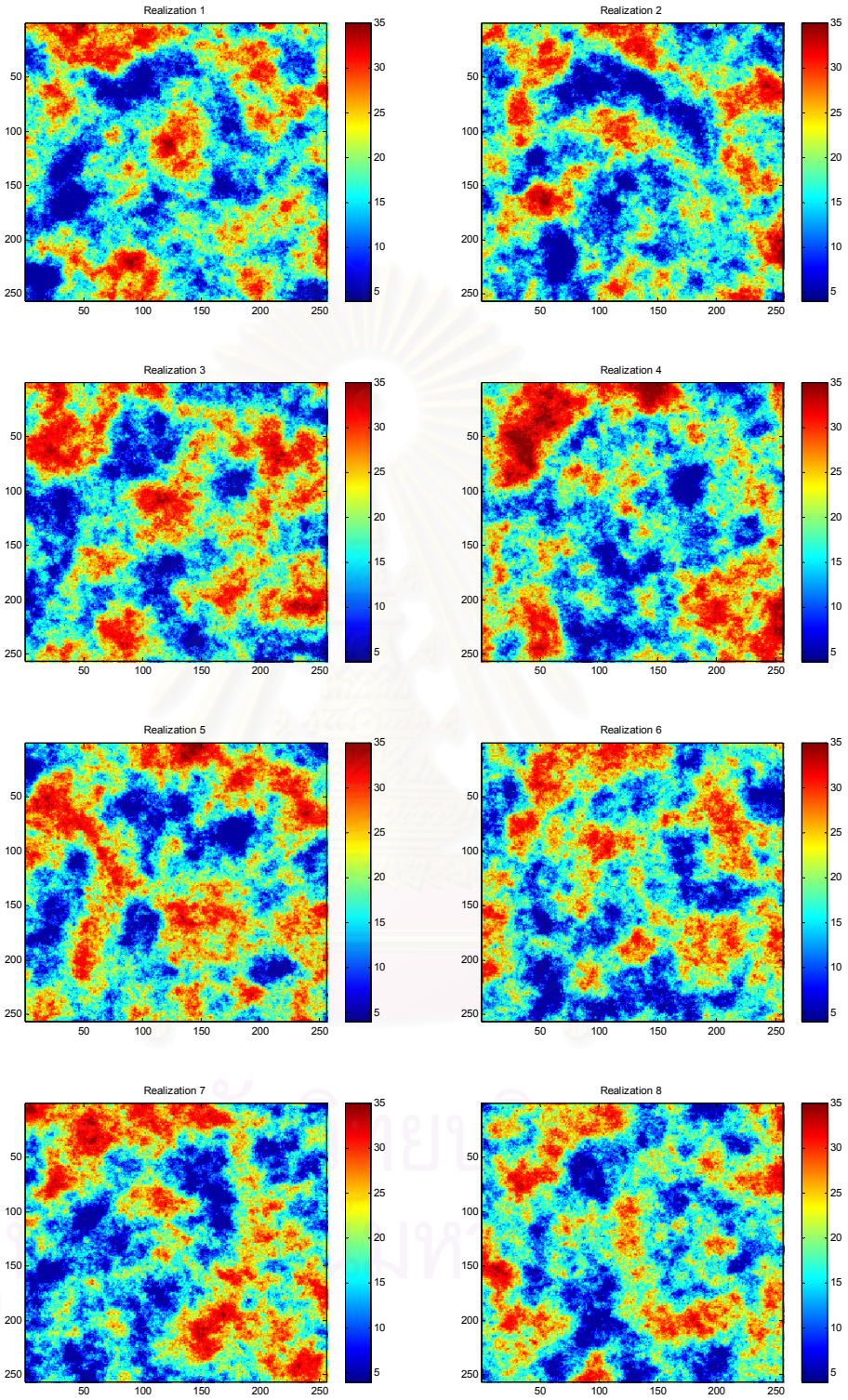

Figure 4.7: Realizations number 1-8.

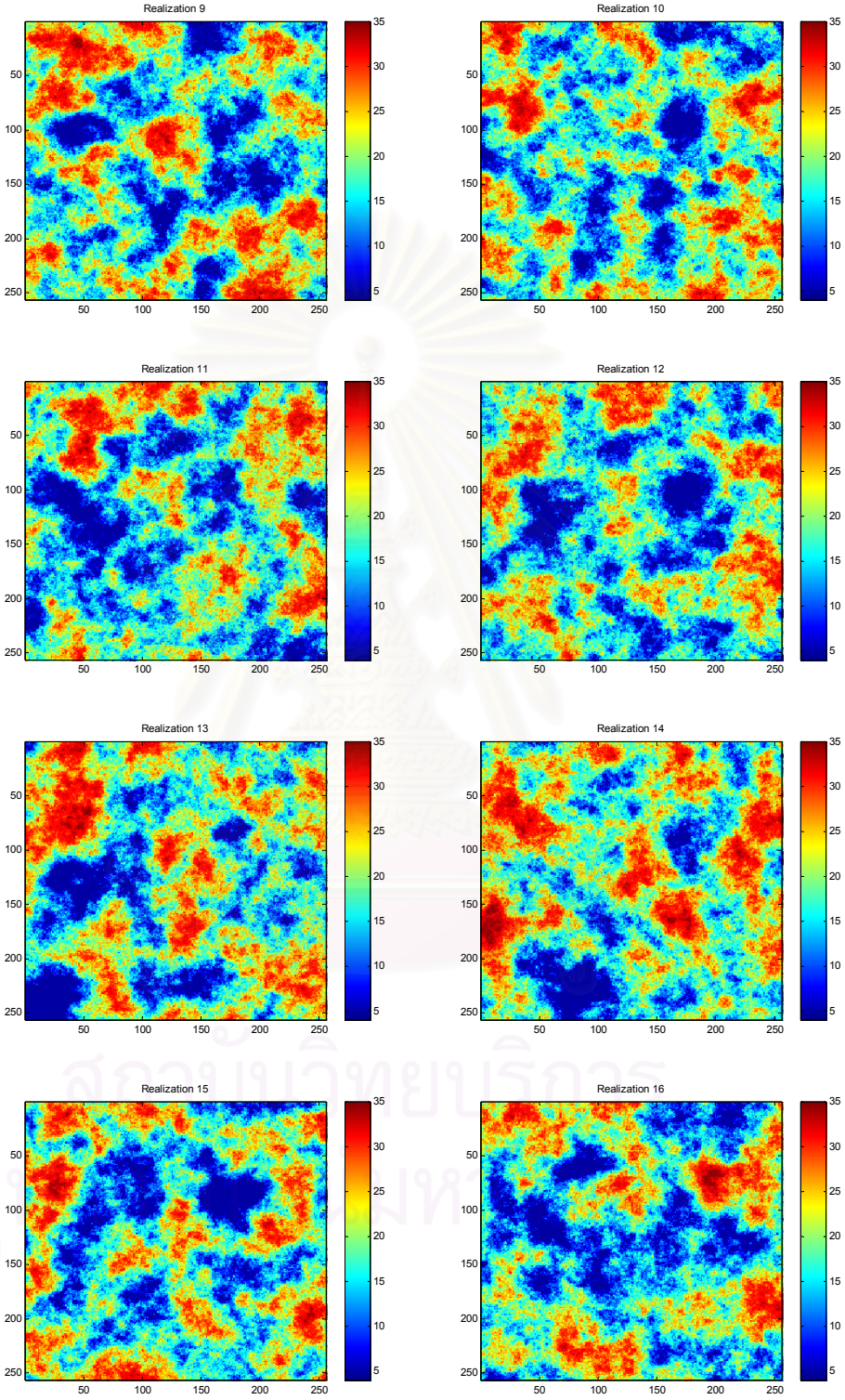

Figure 4.8: Realizations number 9-16.

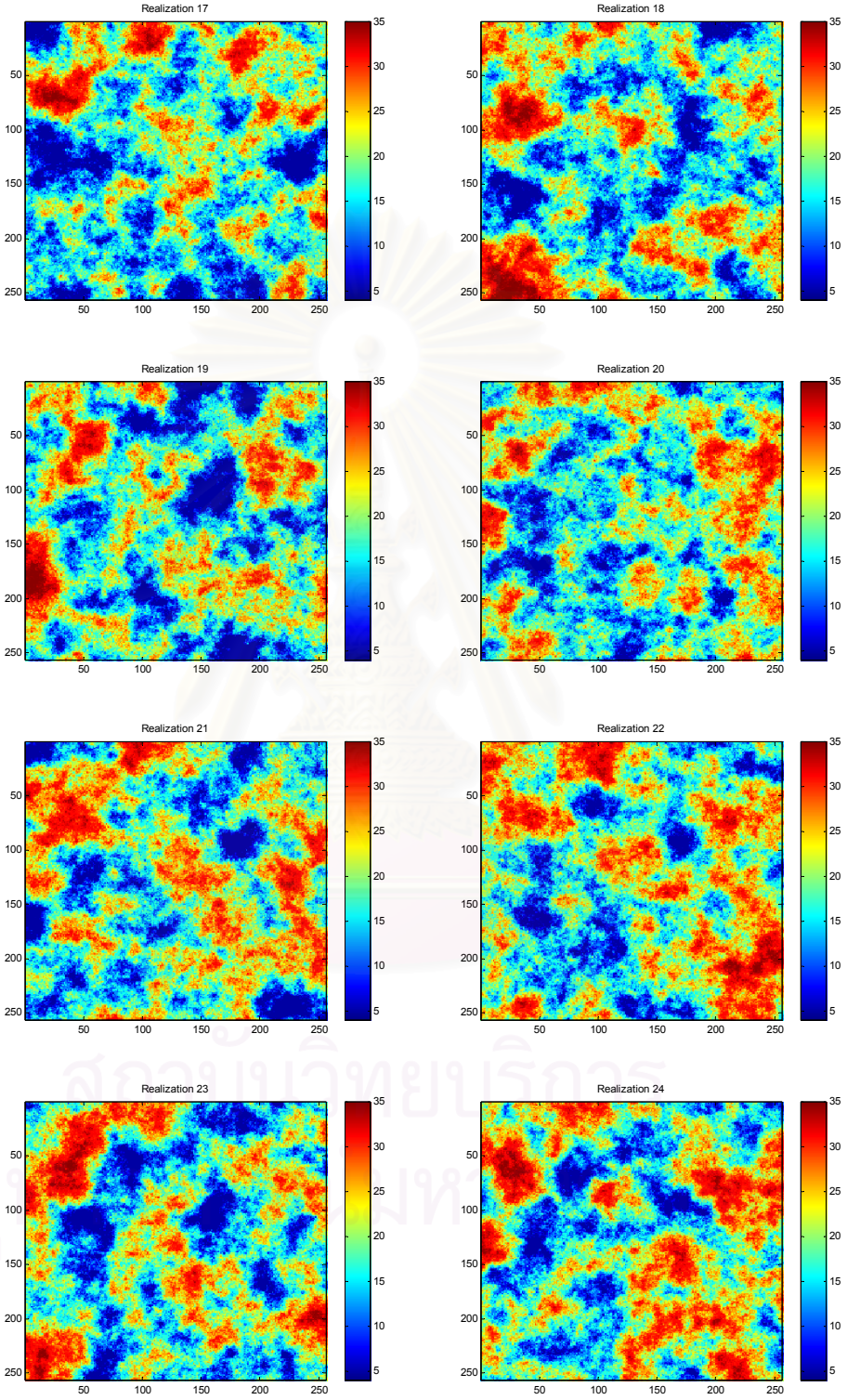

Figure 4.9: Realizations number 17-24.

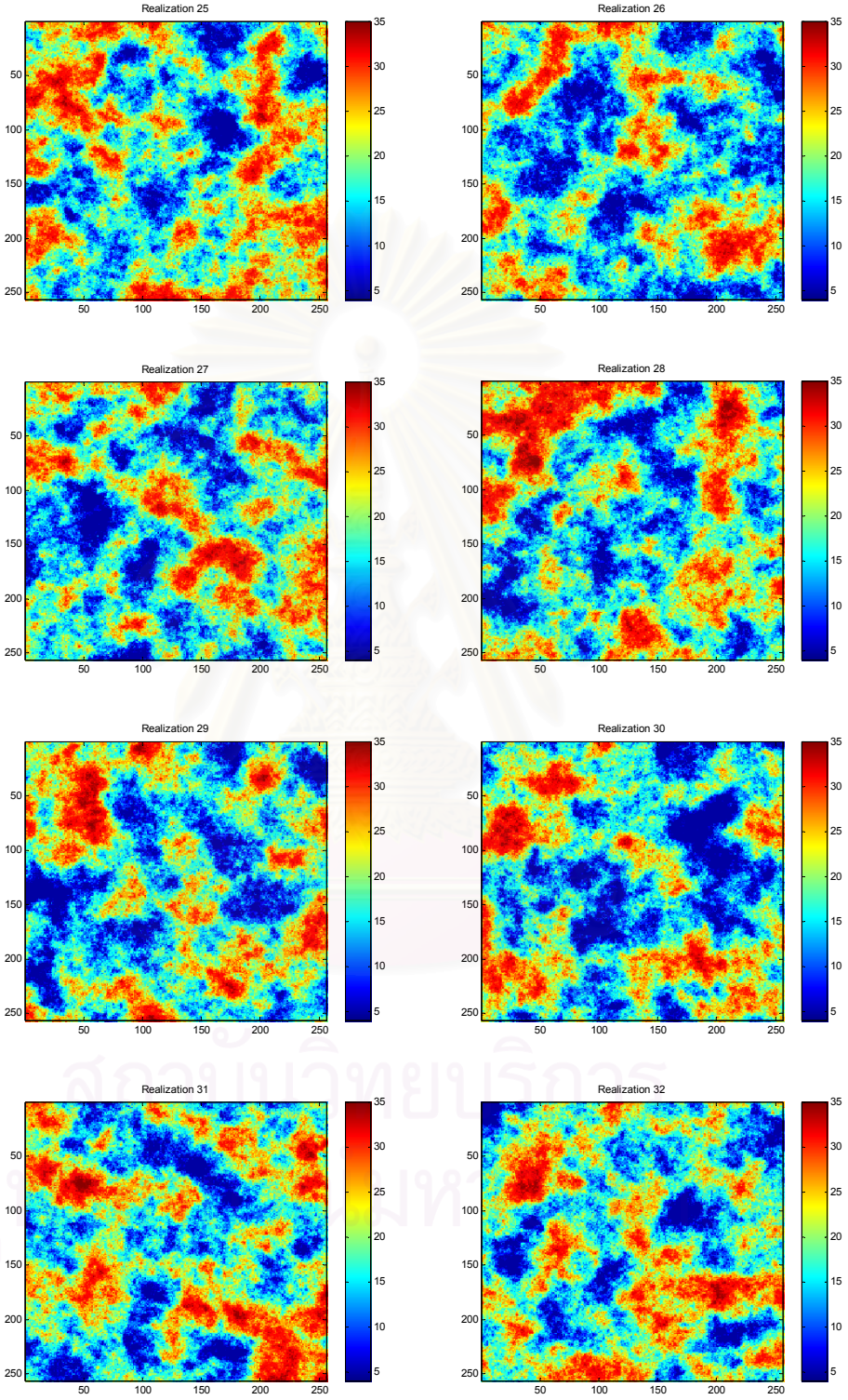

Figure 4.10: Realizations number 25-32.

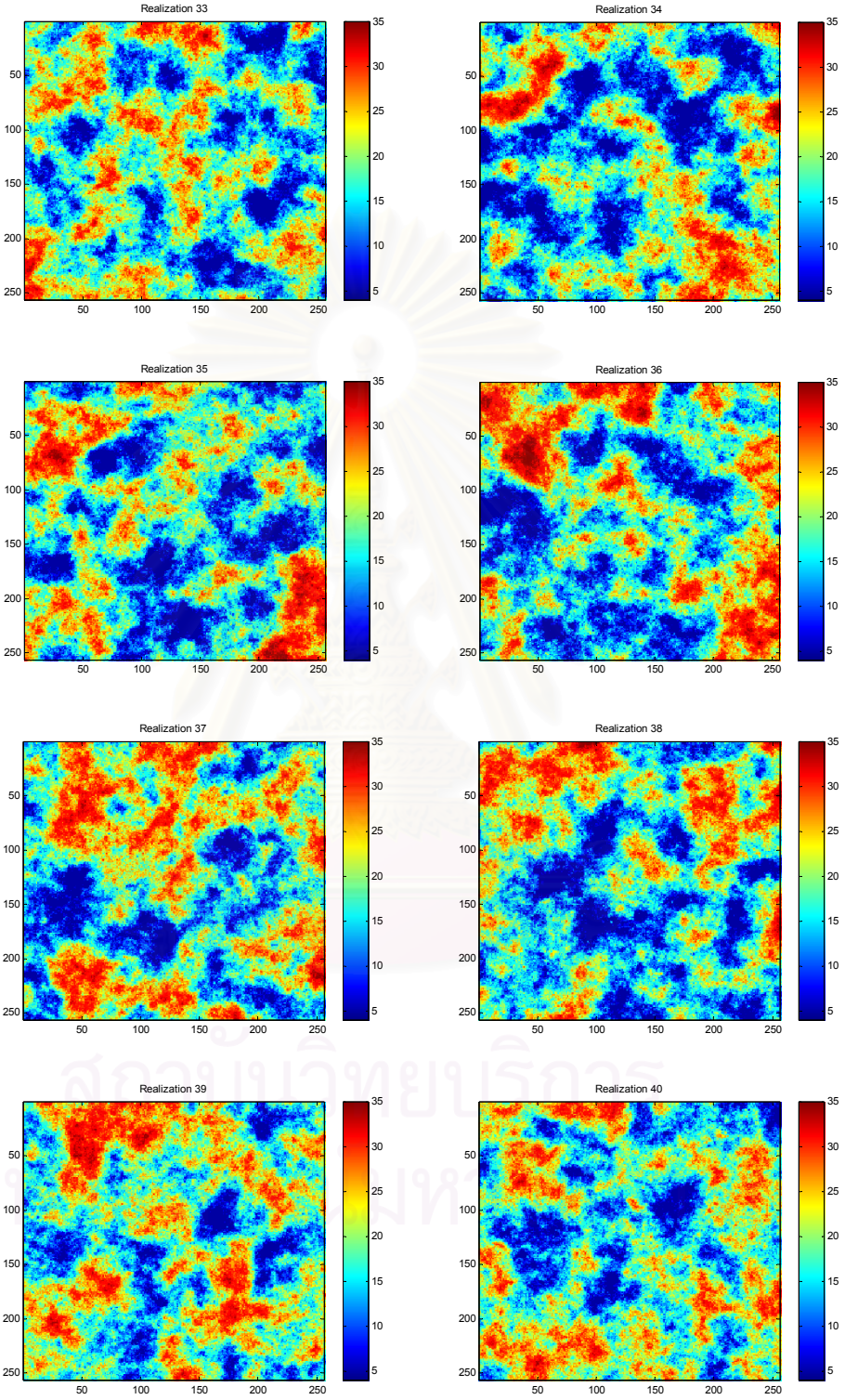

Figure 4.11: Realizations number 33-40.

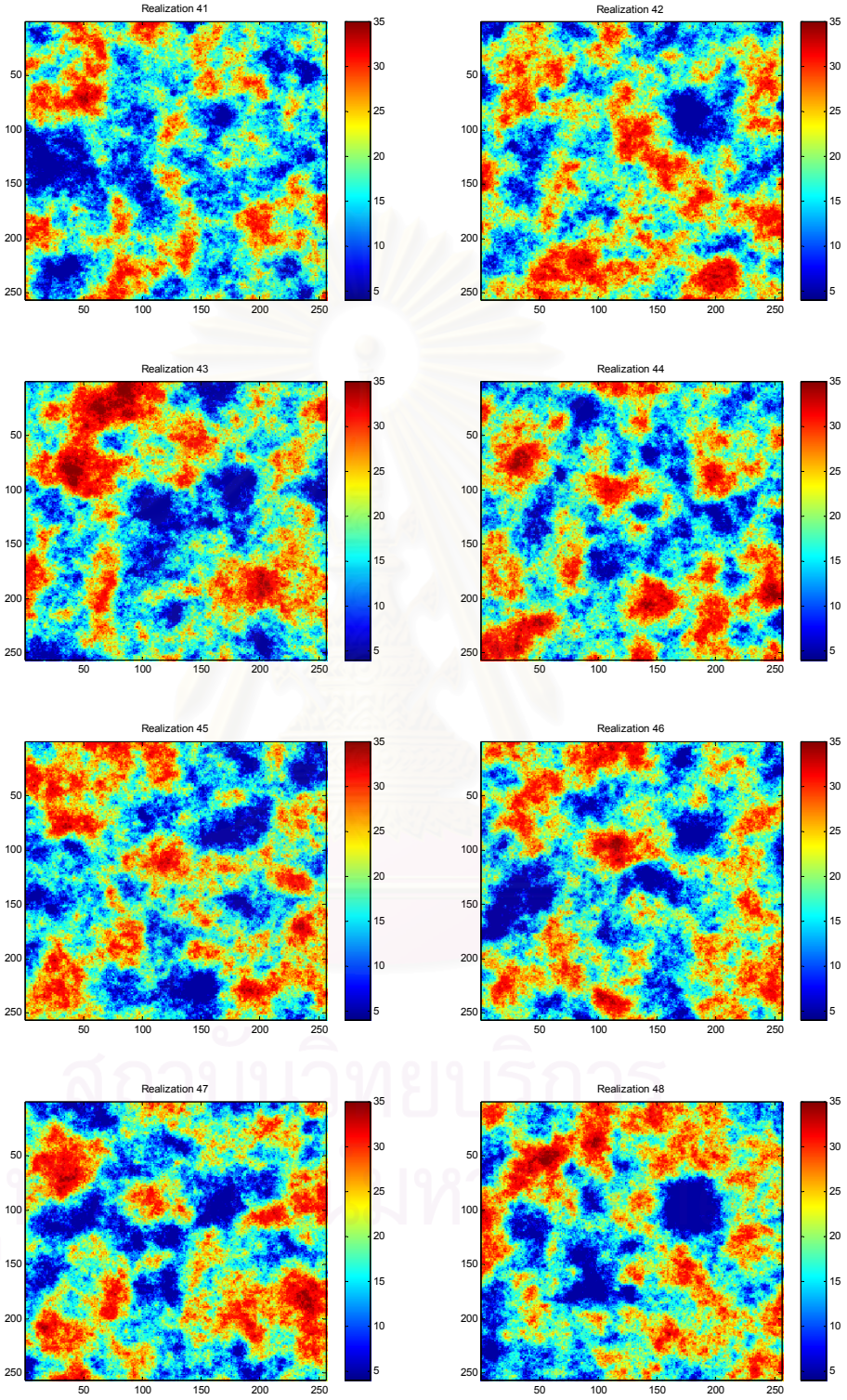

Figure 4.12: Realizations number 41-48.

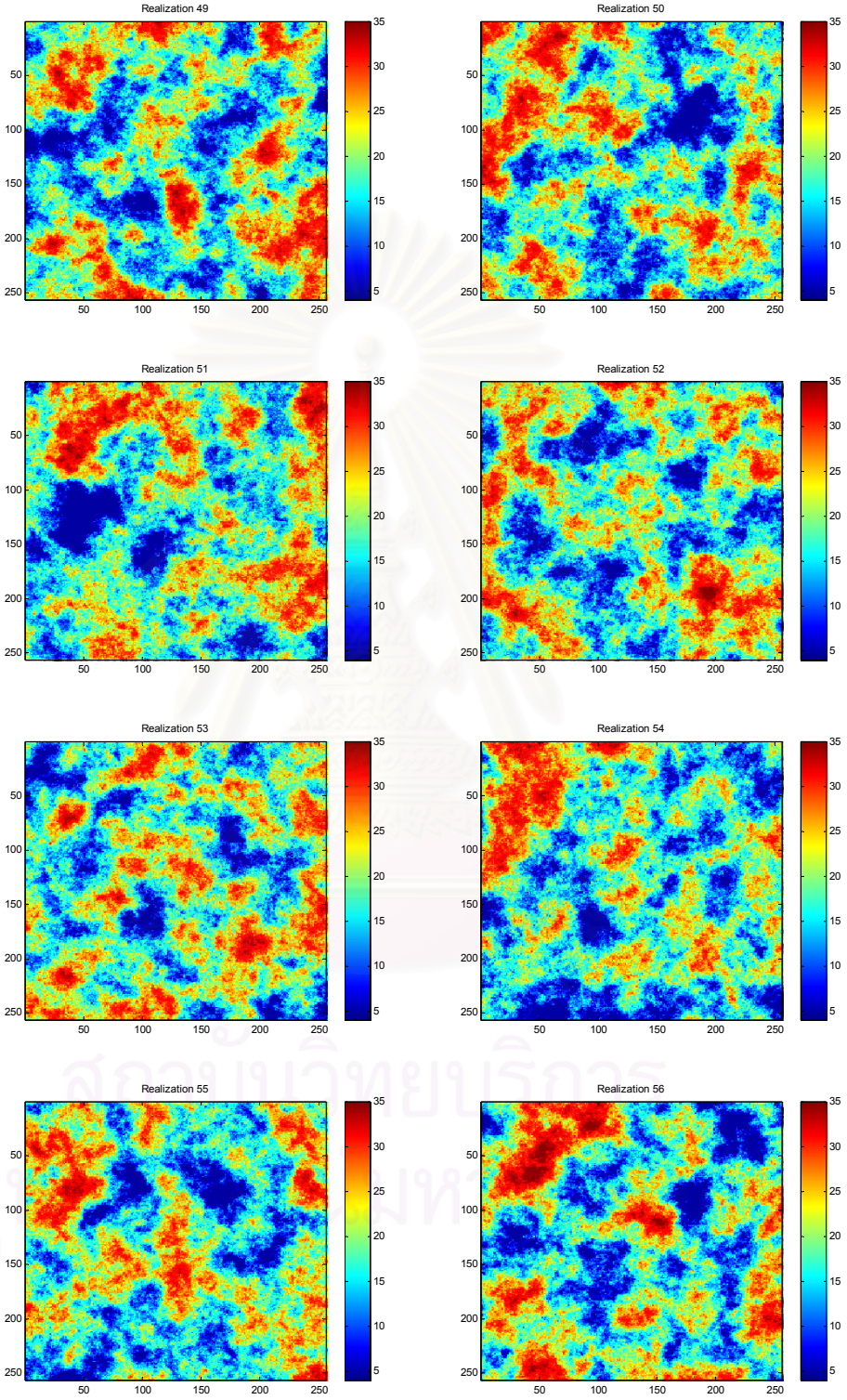

Figure 4.13: Realizations number 49-56.

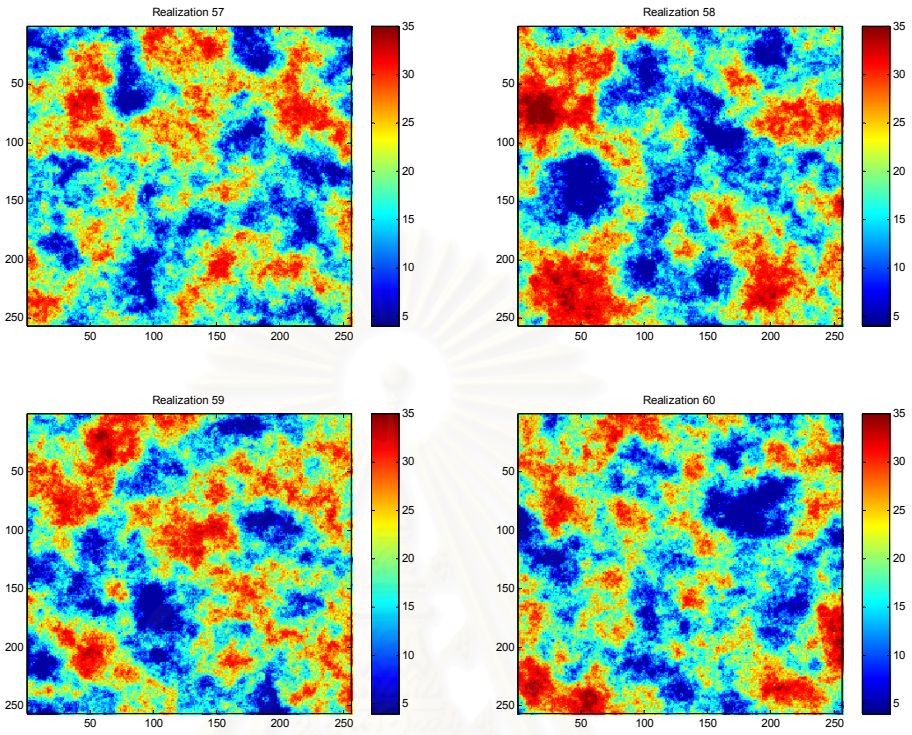

Figure 4.14: Realizations number 57-60.

In global scale, the trend of these 60 realization maps is that there are high porosity values concentrated around the top right and bottom left of the study area and the low porosity values concentrated around the middle of the area. However, there are many differences among these realizations in local scale due to local spatial variation. Therefore, it is very difficult to compare and group similar realizations visually. The next step is to denoise the realizations and then try to group the similar realizations together.

4.3 Denoising of realizations

At this stage, all realizations are denoised using the soft thresholding method. The wavelet function used in this study is Daubechies 4, since it is a prominent function frequently used in data analysis. The threshold values can be calculated from Eq. 3.26. The realizations were denoised by thresholding the three detail components which are horizontal, vertical, and diagonal components. Table 4.7 shows the median

absolute deviation of the horizontal, vertical, and diagonal components and the threshold values at each level of decomposition.

Table 4.7: Calculation of threshold values at each level of decomposition.

Realization	level	Median absolute deviation			Threshold		
		Horizontal	Vertical	Diagonal	Horizontal	Vertical	Diagonal
1	1	1.247	1.209	1.111	3.870	3.752	3.448
	2	2.018	2.046	1.497	5.859	5.940	4.346
	3	4.342	3.854	2.739	11.671	10.359	7.362
	4	7.301	6.236	5.003	17.914	15.301	12.276
2	1	1.221	1.210	1.095	3.790	3.755	3.399
	2	2.116	2.050	1.456	6.143	5.952	4.227
	3	4.602	3.775	2.770	12.370	10.147	7.445
	4	7.064	5.961	5.819	17.333	14.626	14.278
3	1	1.228	1.218	1.070	3.811	3.780	3.321
	2	1.963	2.006	1.427	5.699	5.824	4.143
	3	4.363	4.576	2.677	11.727	12.300	7.195
	4	6.951	6.235	6.142	17.056	15.299	15.071
4	1	1.203	1.191	1.076	3.734	3.697	3.340
	2	2.078	2.018	1.397	6.033	5.859	4.056
	3	4.340	4.130	2.737	11.665	11.101	7.357
	4	6.368	5.640	4.916	15.625	13.839	12.062
5	1	1.242	1.232	1.091	3.855	3.824	3.386
	2	2.155	2.079	1.477	6.256	6.036	4.288
	3	4.694	4.284	2.818	12.617	11.515	7.574
	4	6.152	6.226	5.215	15.095	15.277	12.796
6	1	1.298	1.276	1.155	4.029	3.960	3.585
	2	2.023	2.133	1.570	5.873	6.193	4.558
	3	4.447	4.630	2.863	11.953	12.445	7.695
	4	6.420	8.059	5.236	15.753	19.774	12.848
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
60	1	1.270	1.272	1.096	3.942	3.948	3.402
	2	2.113	2.187	1.529	6.135	6.349	4.439
	3	4.481	3.886	2.891	12.044	10.445	7.771
	4	7.376	5.188	4.745	18.098	12.730	11.643

In this study, many levels of denoising were conducted by trial and error to find the appropriate denoising level of these realizations. Fig. 4.15 shows an example of realization number 1 and its denoised images reconstructed from different levels of decomposition.

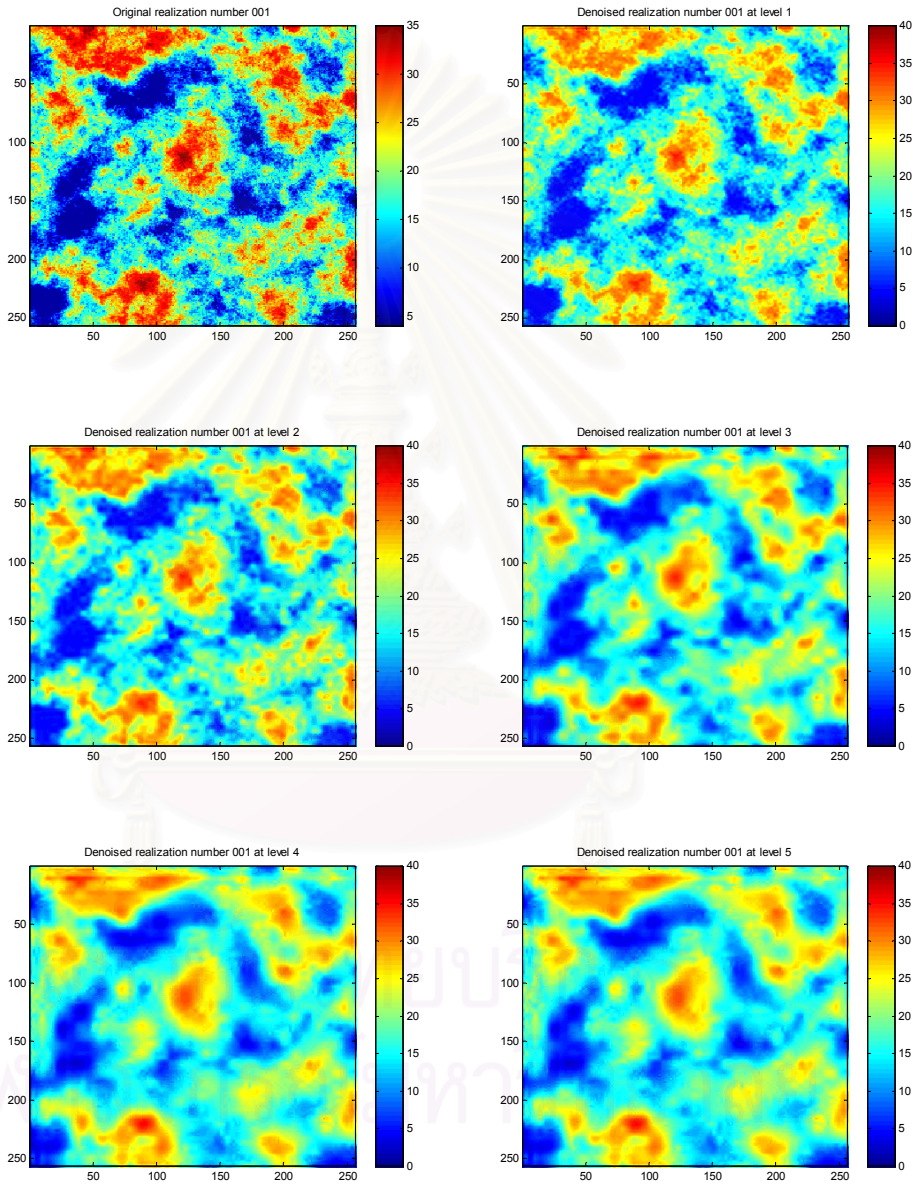

Figure 4.15: Original realization number 1 and its denoised maps reconstructed from 5 different levels of resolution.

As can be seen from Fig. 4.15, the denoised images reconstructed from higher levels are smoother than those reconstructed from lower ones. However, the denoised

realization at level 5 is not much smoother than that at level 4. As a result, level 4 of denoising seems to be an appropriate one to use.

In this study, denoising was done using wavelet function in MATLAB program. Fig. 4.16 shows an example of the denoising process in MATLAB program.

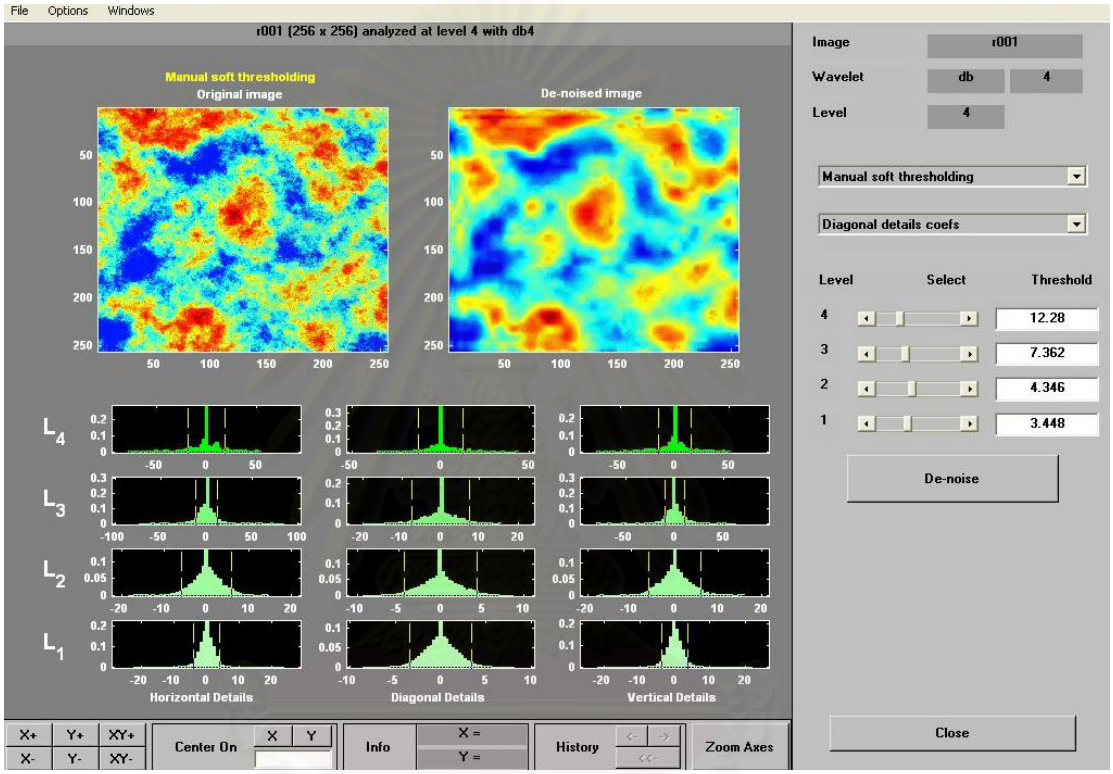

Figure 4.16: Denoising a realization using MATLAB program.

Shown in Fig. 4.16 are the original image of realization number 1 and its denoised image. The letters L with subscript 1 to 4 refer to levels 1 to 4 of the decomposition, respectively. The twelve histograms shown in Fig. 4.16 are the histograms of detail coefficient at each level and direction of decomposition, which are 4 levels in horizontal, vertical, and diagonal direction. The threshold value at each level can be entered into the program by filling the numbers in the boxes located at the middle right or moving the dash lines in the small charts. As seen in the figure, the denoised image is smoother than the original one but still preserves the main structure

of the original image. Some denoised realizations, which are denoised realizations of realizations number 1, 10, 30, and 50, are shown in Fig. 4.17.

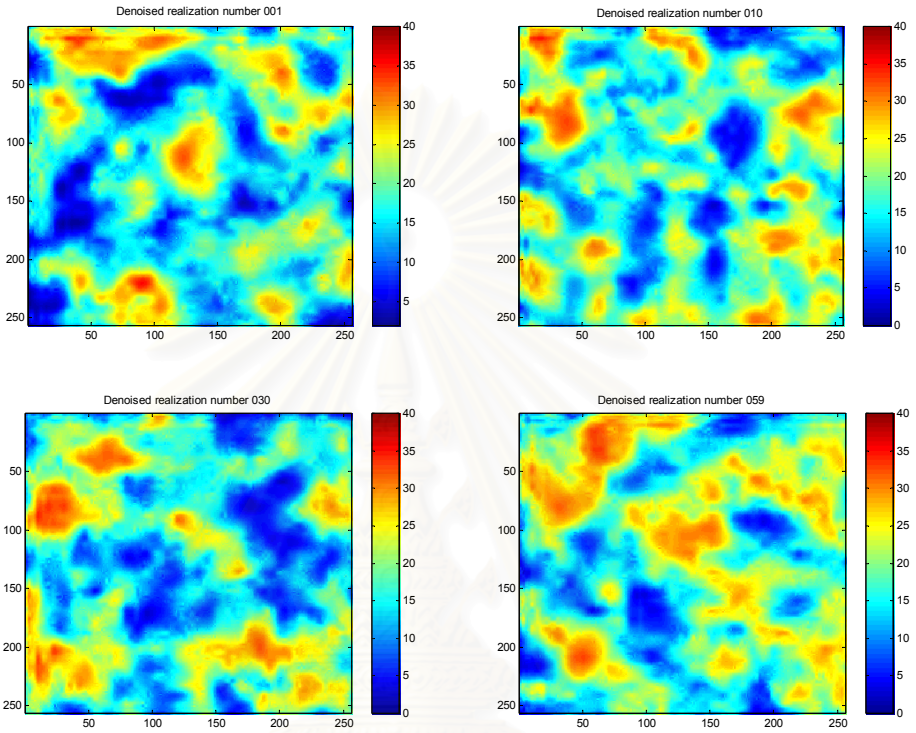

Figure 4.17: Examples of denoised realizations.

After all realizations were denoised, they would then be compared one by one to find similarity among them. The next step is finding the similarity among realizations and grouping these realizations based on correlation between each pair.

4.4 Grouping similar denoised realizations

To compare similarity between two realizations, correlation between properties at the same location of the two realizations is used. The correlation can be calculated from Eq. 3.32. Since 60 realizations are used in this study, the total number of correlations are 3,600 (60 of them are autocorrelation). If the correlation of two denoised realizations is high, it is likely that these two realizations have a high degree of similarity. On the other hand, if the correlation is low, it implies that these two

realizations have a small degree of similarity between them. Tables 4.8 and 4.9 show some examples of the 3,600 correlations and their statistics, respectively.

Table 4.8: Correlation of realizations.

Realization		Correlation	
First	Second	Original realizations	Denoised realizations
1	1	1.000	1.000
1	2	0.216	0.273
1	3	0.412	0.507
⋮	⋮	⋮	⋮
1	59	0.250	0.316
1	60	0.106	0.142
2	1	0.216	0.273
2	2	1.000	1.000
2	3	0.255	0.317
⋮	⋮	⋮	⋮
2	59	0.193	0.244
2	60	0.173	0.217
⋮	⋮	⋮	⋮
59	1	0.250	0.316
59	2	0.193	0.244
59	3	0.353	0.440
⋮	⋮	⋮	⋮
59	59	1.000	1.000
59	60	0.249	0.311
60	1	0.106	0.142
60	2	0.173	0.217
60	3	0.161	0.207
⋮	⋮	⋮	⋮
60	59	0.249	0.311
60	60	1.000	1.000

Table 4.8 shows both correlations of original and denoised realizations. The correlation of original realizations shown in the third column is the correlation between a pair of realizations shown in the first and the second columns. Likewise,

the correlation of denoised realizations shown in the fourth column is the correlation between a pair of denoised realizations shown in the first and the second columns. As seen in the table, the correlations between denoised realizations are higher than correlations between original realizations for the same pair of realizations. This means that the degree of similarity between denoised realizations is higher than that of the original ones.

Table 4.9: Statistical analysis of correlation values.

Parameters	Original realizations	Denoised realizations
Mean	0.226	0.284
Variance	0.05	0.07
Std. Dev	0.069	0.083
Minimum	-0.015	-0.012
25^{th} %	0.178	0.226
Median	0.228	0.286
75^{th} %	0.274	0.343
Maximum	0.474	0.569

In Table 4.9, the parameter values in “Original realizations” column are statistics of the 3,540 correlations paired up from 60 realizations excluding the pairs of the same realization generated by stochastic simulation. The values in “Denoised realizations” column are statistics of the 3,540 correlations paired up from 60 denoised realizations excluding the pairs of the same realization. As seen in the table, the mean of the correlations between pairs of denoised realizations is higher than that of the original realizations because the degree of similarity increases. The variance of the correlations between pairs of denoised realizations is higher than that of the original realizations because the correlations between pairs of the denoised realizations have wider range of values than those of the original realizations.

Fig. 4.18 and 4.19 display the histograms of the correlations for 3,540 pairs of the original realizations and the correlations for 3,540 pairs of denoised realizations, respectively.

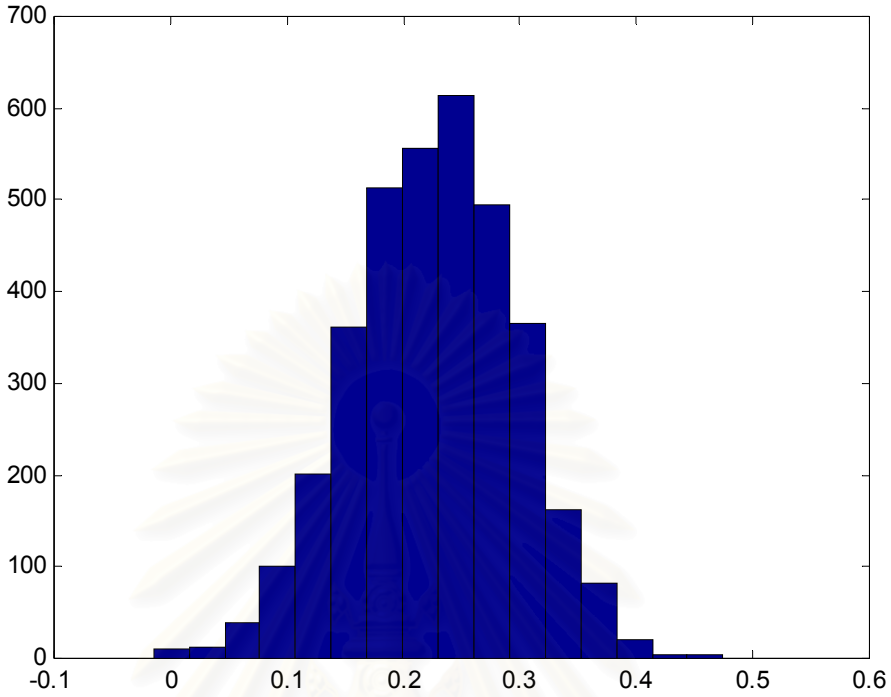

Figure 4.18: Histogram of correlation values between 3,540 pairs of original realizations.

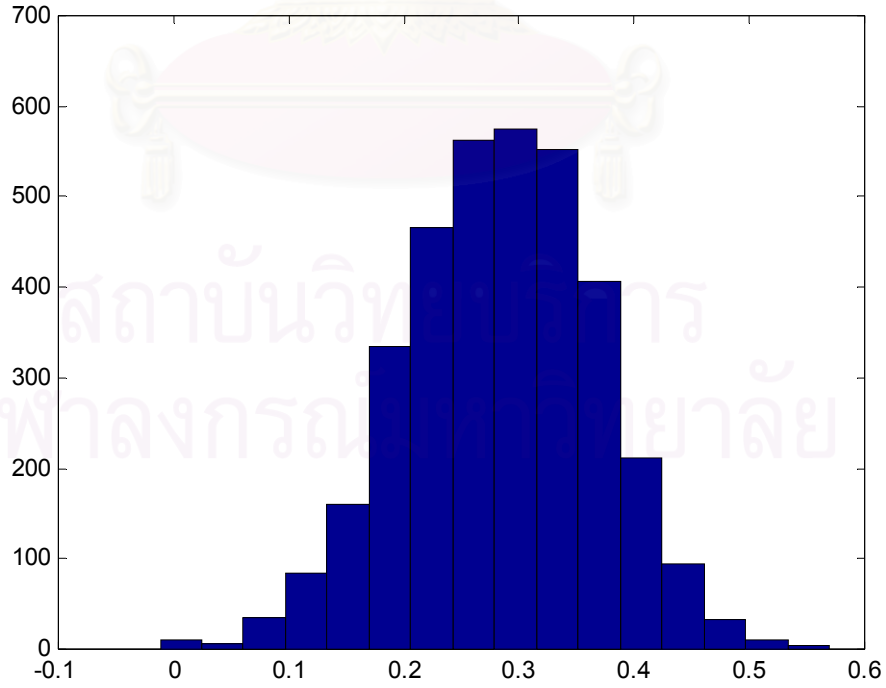

Figure 4.19: Histogram of correlation values between 3,540 pairs of denoised realizations.

As depicted in Fig. 4.18 and 4.19, the distributions of correlations between original and denoised realizations are close to normal distribution but the histogram of the correlations between denoised realizations shifts to the right side comparing with the histogram of the correlations between original realizations. The span of the histogram of the correlations between denoised realizations is wider than that of the original ones. Since the mean of the correlation between denoised realizations is higher, it is better to group denoised images rather than original realizations.

To group the realizations based on the correlation, 60 groups of realizations are composed. There are two criteria used for the grouping process. The first criterion is to specify a cut-off value to eliminate pairs of realizations having a small correlation value. The second principle is that an image with a higher correlation value has a higher priority to be the member of a group than the lower one and the new member of the group has to have correlations with all the previous members of the group not lower than the cut-off value. The procedure of grouping is shown as follows:

1. Calculate correlations of pairs of realizations as shown in Table 4.8.
2. Specify a cut-off value to eliminate pairs that have small correlations.
3. Starting with realization 1, determine realizations that are highly correlated to it (realization that yields a correlation greater than the cut-off value).
4. Rank the realizations that are highly correlated with the base realization based on the correlation values.
5. Check if the correlation between the second most correlated realization and the most correlated realization is higher than the threshold. If it is, include the second most correlated realization as a group member. If not, exclude it. Then, check if the correlation between the third most correlated image and the most correlated image and the correlation between the third most correlated image and the second most correlated image are both higher than the cut-off value. If they are, include the third most correlated realization as a group member. If not, exclude it. Then, perform the same kind of checking for all realizations ranked in step 4.
6. Then continue to find similar realizations based on the same cut-off for realizations 2, 3, 4, …, and 60, respectively.

7. Vary cut-off correlation values and do the grouping.

To clarify the procedure of grouping, grouping of 4 images is presented as an example. The procedure of finding images similar to the base image **A** is presented as follows:

Table 4.10: Correlation values among 4 images: **A**, **B**, **C** and **D**.

Images	Correlation	Images	Correlation	Images	Correlation	Images	Correlation
A-A	1.00	**B-B**	1.00	**C-C**	1.00	**D-D**	1.00
A-B	0.20	**B-A**	0.20	**C-A**	0.40	**D-A**	0.30
A-C	0.40	**B-C**	0.25	**C-B**	0.25	**D-B**	0.10
A-D	0.30	**B-D**	0.10	**C-D**	0.35	**D-C**	0.35

1. The correlation among the four images are calculated and tabulated in Table 4.10.

2. A cut-off value of 0.19 is used.

3. The pairs of images having correlation values lower than cut-off value are eliminated. In this case, images **B**, **C**, **D** all have correlation with **A** higher than 0.19. Thus, none is eliminated.

4. The chosen images are ordered according to their ranks as image **C**, **D**, **B**.

5. Image **C** is a member of the group automatically because it has the highest correlation value with base image **A**.

Then, check whether the correlation value between image **D** (the second most correlated image) and image **C** (the most correlated image) is higher than the cut-off value. In this example, image **D** can be a member of this group because the correlation value between images **D** and **C** is 0.35 which is higher than the specified cut-off value.

Then, check whether the correlation value between image **B** (the third most correlated image) and image **C** (the most correlated image) and the correlation value between image **B** (the third most correlated image) and image **D** (the second most correlated image) are higher than the cut-off value. In this case, correlation values between images **B** and **C** and images **B** and **D** are 0.25 and 0.10, respectively. Since

the correlation value between images **B** and **D** is lower than the cut-off value, image **B** cannot be a member of the group.

Table 4.11: Ranking of highly correlated denoised realizations with a cut-off value of 0.35.

Base Realization	Realization that is highly correlated with the base realization						
	Most correlated	2nd most correlated	3rd most correlated	4th most correlated	5th most correlated	6th most correlated	7th most correlated
1	3	15	9	-	-	-	-
2	36	46	22	-	-	-	-
3	1	38	29	-	-	-	-
4	35	36	12	58	52	-	-
5	17	-	-	-	-	-	-
6	19	23	-	-	-	-	-
7	34	24	22	36	27	-	-
8	59	23	22	3	36	-	-
9	30	35	52	4	-	-	-
10	15	22	12	-	-	-	-
11	28	38	-	-	-	-	-
12	**51**	**36**	**11**	**20**	**22**	**27**	**3**
13	17	51	59	36	27	-	-
14	31	36	-	-	-	-	-
15	23	8	22	3	12	-	-
16	40	51	37	11	-	-	-
17	13	51	59	36	27	-	-
18	58	30	52	4	35	-	-
19	23	6	-	-	-	-	-
20	22	24	36	27	-	-	-
21	54	14	-	-	-	-	-
22	**36**	**27**	**20**	**3**	**11**	**51**	**12**
23	8	15	12	22	3	-	-
24	7	22	34	27	36	-	-
25	10	58	-	-	-	-	-
26	52	22	9	15	-	-	-
27	22	23	3	59	11	36	-
28	11	38	-	-	-	-	-
29	3	36	22	23	-	-	-

Table 4.11: Ranking of highly correlated denoised realizations with a cut-off value of 0.35 (continued).

Base Realization	Realization that is highly correlated with the base realization						
	Most correlated	2nd most correlated	3rd most correlated	4th most correlated	5th most correlated	6th most correlated	7th most correlated
30	18	52	4	58	35	-	-
31	36	13	55	-	-	-	-
32	53	-	-	-	-	-	-
33	23	49	3	-	-	-	-
34	7	4	36	22	-	-	-
35	4	36	58	30	52	-	-
36	22	46	45	-	-	-	-
37	51	59	11	-	-	-	-
38	11	3	22	51	-	-	-
39	57	-	-	-	-	-	-
40	16	37	51	11	-	-	-
41	51	13	56	12	-	-	-
42	60	30	-	-	-	-	-
43	51	59	-	-	-	-	-
44	58	18	-	-	-	-	-
45	46	36	22	-	-	-	-
46	36	45	47	-	-	-	-
47	49	3	36	12	51	-	-
48	7	-	-	-	-	-	-
49	47	46	51	36	-	-	-
50	4	58	12	52	18	-	-
51	12	41	13	56	-	-	-
52	1	18	12	15	-	-	-
53	8	-	-	-	-	-	-
54	21	12	-	-	-	-	-
55	13	36	31	-	-	-	-
56	51	41	13	12	-	-	-
57	39	-	-	-	-	-	-
58	18	52	4	30	35	-	-
59	8	3	22	23	36	-	-
60	30	4	36	58	35	-	-

Table 4.11 shows the ranking of highly correlated denoised realizations when a cut-off value of 0.35 is used. The denoised realization shown in the first column is the base realization for the group. The first member of each group is the denoised realization that has the highest correlation with the base realization. The rest of the

members of the group are selected by the same principle with additional constraint that the new member of the group has to have correlations with all the previous members of the group not lower than the cut-off value. For example, at cut-off value of 0.35, realization number 22 has 7 realizations that are similar to it which are realization number 36, 27, 20, 3, 11, 51, and 12. Realization number 12 also has 7 similar realizations which are realization number 51, 36, 11, 20, 22, 27, and 3. Although both realization 22 and realization 12 have 8 members in its group, realization 22 has a higher average correlation between the base realization and its members. Thus, realization 22 can be considered as the most representative realization among the 60 realizations used in the comparison.

In this study, the cut-off correlation value was varied from zero until the number of members of the biggest group equals to two to examine the results of grouping using original and denoised realizations. Tables 4.12 and 4.13 show the grouping of original and denoised realizations using different cut-off values, respectively. The first column of both tables is the cut-off values and the order of groups is from the biggest group to smaller groups. This study focuses on only the big groups so these tables show only the first four biggest groups. The grouping result of original realizations at a cut-off of 0.0 shows that the biggest group is the group of base realization number 36.

As can be observed from Tables 4.12 and 4.13, the grouping shows static results at a certain range of cut-off values. For original realizations, the biggest groups are the same for cut-off values between 0.06 and 0.09 and between 0.31 and 0.34. For denoised realizations, the biggest groups are the same for cut-off values between 0.02 and 0.06 and between 0.25 and 0.30 and between 0.34 and 0.38. The denoised grouping shows a more static result. For original realizations, the cut-off values between 0.06 and 0.09 are too low, comparing to the rest of correlation values and the cut-off values between 0.31 and 0.34 are very high, which means that only a few realizations are included in the grouping, comparing to the rest of correlation values. For denoised realizations, the static result between cut-off values of 0.02 and 0.06 is not suitable because these values are too low. For the static result at cut-off value of 0.28 (around the median of the correlation values), realization number 4 is the

realization with the biggest member (a total of 16 including the base realization). Members of this group are shown in Fig. 4.20 and 4.21. As seen in the figures, there are still some realizations that have a certain degree of dissimilarity.

Table 4.12: Grouping of original realizations using different cut-off values.

Cut-off	Base realization of the biggest group	Base realization of the second biggest group	Base realization of the third biggest group	Base realization of the fourth biggest group
0.00	36	22	4	12
0.01	36	22	4	12
0.02	4	12	23	56
0.03	4	12	23	58
0.04	4	51	59	56
0.05	22	4	51	56
0.06	36	22	12	4
0.07	36	22	58	52
0.08	36	22	47	5
0.09	36	23	49	7
0.10	42	36	22	12
0.11	26	36	22	12
0.12	12	34	24	22
0.13	22	4	12	34
0.14	15	52	47	22
0.15	15	47	4	18
0.16	52	15	46	1
0.17	52	4	18	35
0.18	36	4	18	35
0.19	36	18	35	9
0.20	47	36	4	30
0.21	4	36	30	18
0.22	4	58	36	30
0.23	4	36	58	18
0.24	21	4	36	58
0.25	27	35	9	4
0.26	45	27	46	60
0.27	22	4	58	18
0.28	22	58	30	18
0.29	11	4	22	58
0.30	22	4	18	58
0.31	18	58	30	34
0.32	18	58	30	36
0.33	18	58	30	24
0.34	18	58	30	60
0.35	36	22	46	23
0.36	36	22	46	45
0.37	36	22	35	4

Table 4.13: Grouping of denoised realizations using different cut-off values.

Cut-off	Base realization of the biggest group	Base realization of the second biggest group	Base realization of the third biggest group	Base realization of the fourth biggest group
0.00	22	36	12	4
0.01	22	36	12	4
0.02	12	4	23	59
0.03	12	4	23	59
0.04	12	4	23	59
0.05	12	23	4	3
0.06	12	23	4	3
0.07	22	30	8	54
0.08	22	4	51	52
0.09	22	36	23	30
0.10	8	25	20	10
0.11	36	13	34	17
0.12	36	56	34	49
0.13	36	34	17	42
0.14	12	27	50	22
0.15	34	12	15	36
0.16	12	47	4	34
0.17	47	12	15	1
0.18	15	47	52	46
0.19	51	15	47	1
0.20	49	52	15	46
0.21	52	15	46	4
0.22	18	52	15	46
0.23	36	18	35	9
0.24	36	4	18	35
0.25	4	36	30	18
0.26	4	36	30	18
0.27	4	36	18	58
0.28	4	58	36	18
0.29	4	36	58	30
0.30	4	36	58	45
0.31	27	22	36	4
0.32	27	29	23	45
0.33	45	46	22	4
0.34	22	30	23	58
0.35	22	12	27	4
0.36	22	4	18	23
0.37	22	18	58	30
0.38	22	18	30	58
0.39	18	30	58	36
0.40	18	30	58	13
0.41	18	30	58	4
0.42	30	58	18	36
0.43	36	22	46	23
0.44	36	22	46	23
0.45	46	45	36	22
0.46	36	22	35	4

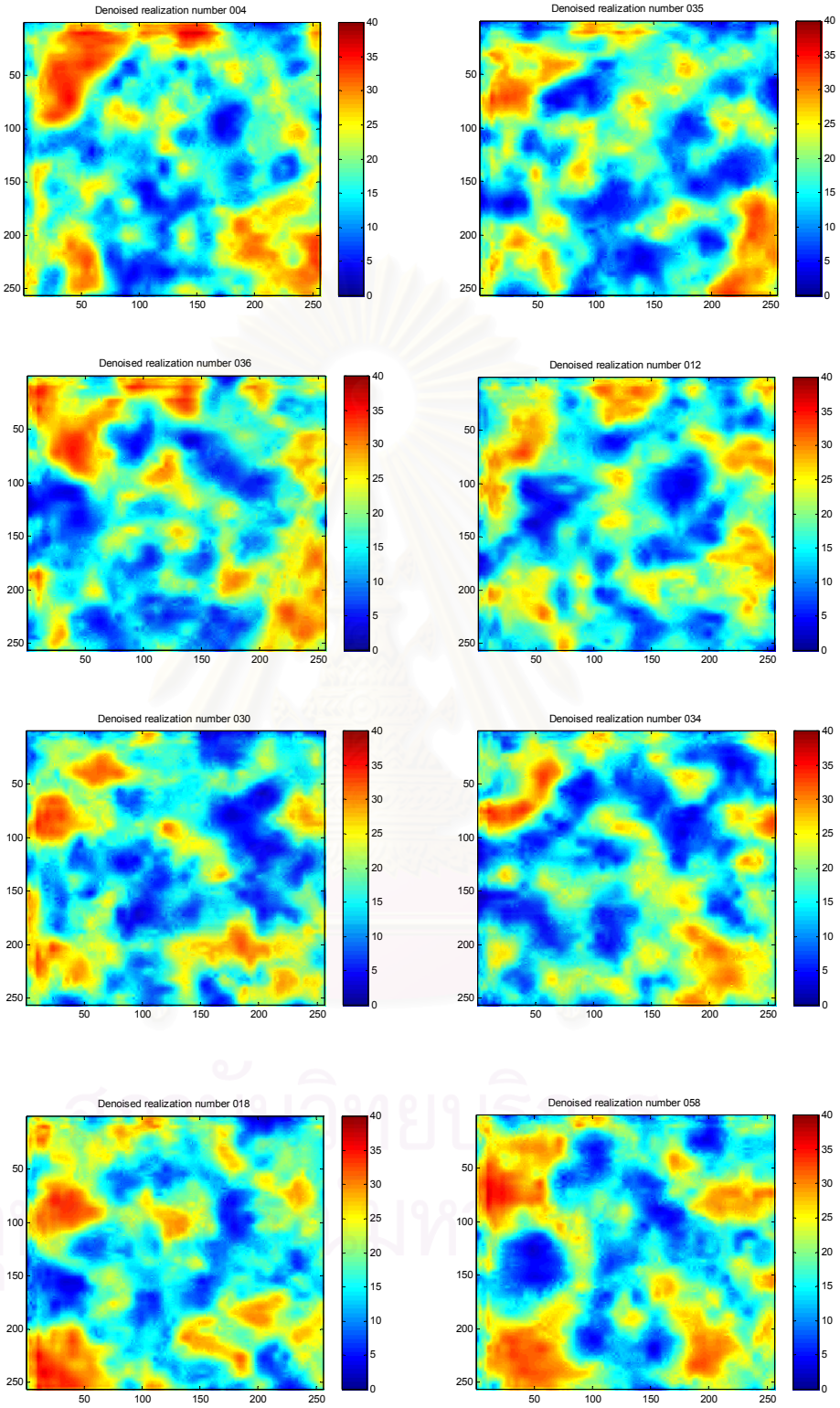

Figure 4.20: The first 8 members of group number 4 using denoised realizations at cut-off value of 0.28 (denoised realization number 4, 35, 36, 12, 30, 34, 18, and 58).

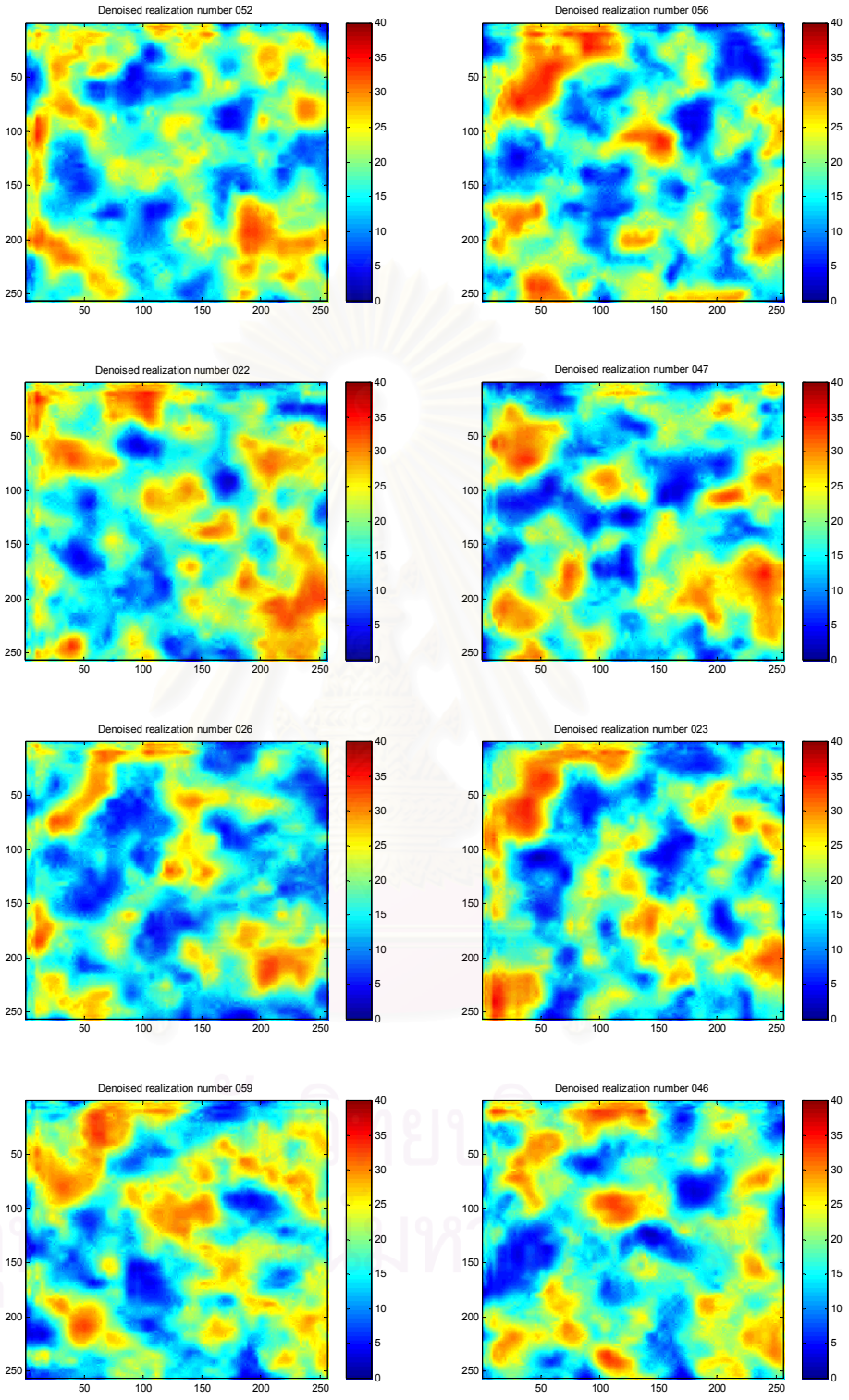

Figure 4.21: The other 8 members of group number 4 using denoised realizations at cut-off value of 0.28 (denoised realization number 52, 56, 22, 47, 26, 23, 59, and 46).

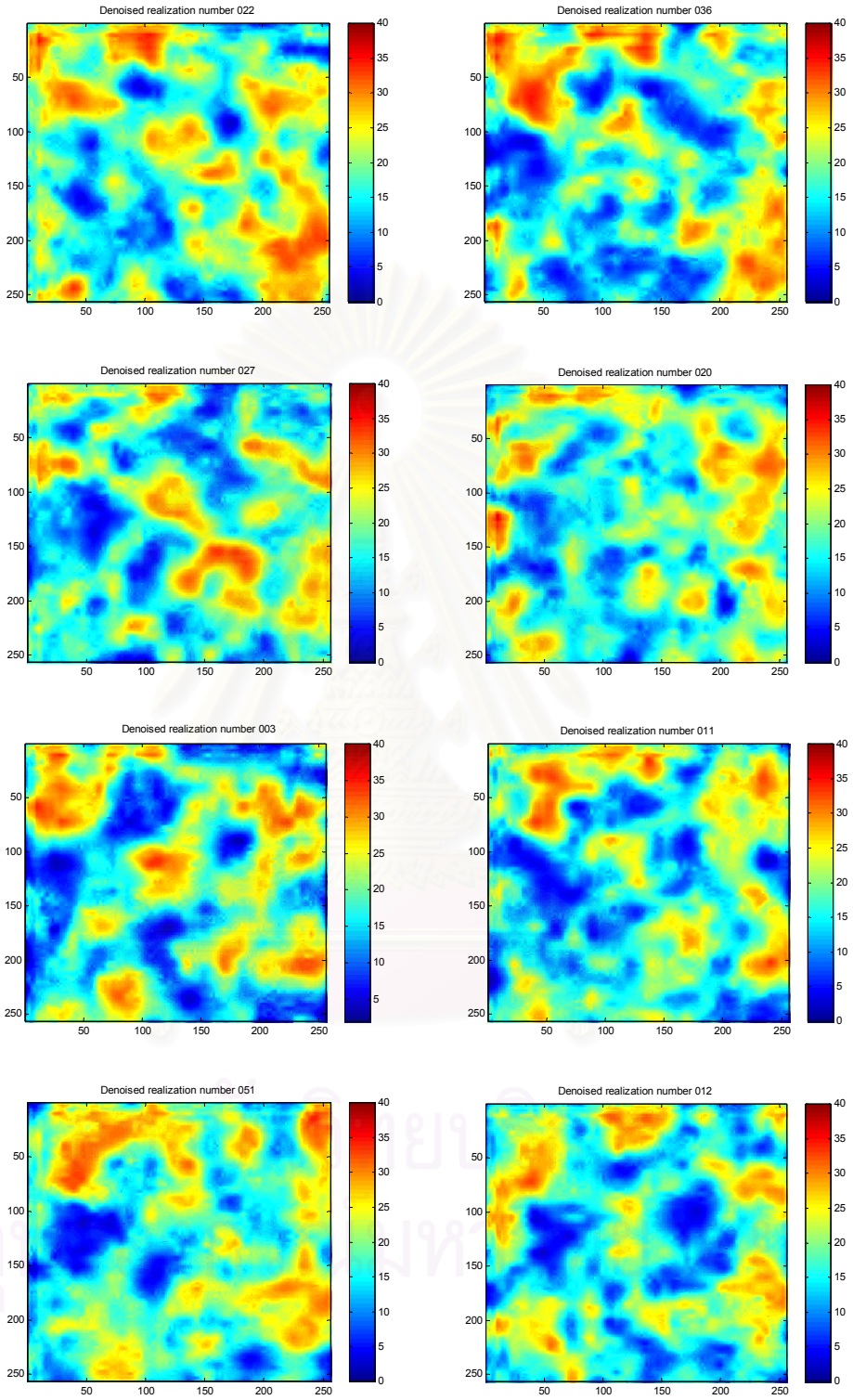

Figure 4.22: The 8 members of group number 22 using denoised realizations at cut-off value of 0.35 (denoised realization number 22, 36, 27, 20, 3, 11, 51, and 12).

For the static result at a cut-off value of 0.35 which is around the third quartile of total correlation values, the members of the group shown in Fig. 4.22 look quite similar. The main structures of these denoised realizations are the same, having high value of porosity at the top and the bottom right of the images and low value of porosity at the bottom left. Therefore, we will use the correlation value at the third quartile as the cut-off value. Having chosen this criteria, realization number 22 is selected as the most representative of all 60 realizations.

However, there is another criterion we can use to find the most representative realization, which is focusing on the number of members of the biggest group. The result of this method depends on the desired number of members of the biggest group. In practice, the proper number of members comes from judgment of analysts who belive that the group, which has this amount of members, is the perfect group having enough degree of similarity. Tables 4.14 and 4.15 show the number of members of the first four biggest groups resulting from original realization and denoised realization grouping at different cut-off values, respectively. The first column of both tables is the cut-off values and the order of groups is from the biggest group to smaller groups. This study focuses on only the big groups, so these tables show only the first four biggest groups. The grouping result of original realizations at a cut-off of 0.0 shows that the number of members of the biggest group is 57, which is the number of members of base realization number 36 as can be observed in Table 4.12.

To find the most representative number of realization in this method, the number of members of the biggest group has to be specified. If the desired number of the members is 9, the most representative realization among original realizations is realization number 45 because it is the base realization of the group that is the biggest group of 9 members at cut-off value of 0.26 (see Tables 4.14 and 4.12). For the same number of members, the most representative realization among denoised realizations is realization number 45 as well because it is the base realization of the group that is the biggest group of 9 members at cut-off value of 0.33 (see Tables 4.15 and 4.13). However, if the desired number of the members is 7, the most representative realization among original realizations is realization number 22 or 11 because they are the base realizations of the groups that are the biggest groups of 7 members at cut-off

value of 0.28 and 0.29, respectively (see Tables 4.14 and 4.12) and the most representative realization among denoised realizations is realization number 22 because it is the base realization of the group that is the biggest group of 7 members at cut-off value of 0.36 (see Tables 4.15 and 4.13).

In summary, 60 porosity realizations were generated from a data set using Sequential Gaussian Simulation. After that, the realizations were denoised using soft thresholding method of wavelet analysis. Then, these denoised realizations were paired to find a correlation between them. After that 60 groups of realizations were compared using a cut-off value to eliminate realizations that are not highly correlated. In this study, cut-off values were varied to observe the less sensitive result of the biggest group. Finally, there are two methods to find the most representative of these 60 realizations, which are focusing on the static results of grouping at different cut-off values and on the proper number of members that is believed to have an appropriate degree of similarity. For the first method, a representative of these 60 realizations, chosen from the biggest group showing less sensitive result, is realization number 22 because it has the highest number of similar realizations at a certain cut-off values. For the second method, the most representative can be any realizations depending on the specified number of members of the biggest group.

Table 4.14: The number of grouping members of original realizations.

Cut-off	No. of members of the biggest group	No. of members of the second biggest group	No. of members of the third biggest group	No. of members of the fourth biggest group
0.00	57	57	57	57
0.01	57	57	57	57
0.02	57	57	57	57
0.03	56	56	56	56
0.04	55	55	55	55
0.05	52	52	52	52
0.06	50	50	50	50
0.07	49	49	49	49
0.08	46	46	46	46
0.09	43	43	43	43
0.10	40	39	39	39
0.11	37	36	36	36
0.12	35	35	35	34
0.13	34	34	34	34
0.14	32	32	32	31
0.15	30	30	29	29
0.16	28	28	28	28
0.17	25	24	24	24
0.18	23	23	23	23
0.19	21	21	21	21
0.20	20	19	19	19
0.21	17	17	17	17
0.22	16	16	15	15
0.23	14	14	14	14
0.24	12	11	11	11
0.25	11	11	11	10
0.26	9	9	9	9
0.27	8	7	7	7
0.28	7	7	7	7
0.29	7	6	6	6
0.30	6	5	5	5
0.31	5	5	5	5
0.32	5	5	5	4
0.33	5	5	5	4
0.34	5	5	5	4
0.35	3	3	3	3
0.36	3	3	3	3
0.37	2	2	2	2

Table 4.15: The number of grouping members of denoised realizations.

Cut-off	No. of members of the biggest group	No. of members of the second biggest group	No. of members of the third biggest group	No. of members of the fourth biggest group
0.00	59	59	59	59
0.01	57	57	57	57
0.02	57	57	57	57
0.03	57	57	57	57
0.04	57	57	57	57
0.05	55	55	55	55
0.06	55	55	55	55
0.07	54	54	54	54
0.08	51	51	51	51
0.09	50	50	50	50
0.10	49	49	49	49
0.11	45	45	45	45
0.12	43	43	43	43
0.13	40	40	40	40
0.14	38	38	38	37
0.15	36	35	35	34
0.16	35	35	34	34
0.17	34	33	33	33
0.18	32	32	31	31
0.19	29	29	29	29
0.20	29	28	28	28
0.21	28	28	28	27
0.22	24	24	24	24
0.23	22	22	22	22
0.24	21	21	21	21
0.25	19	19	19	19
0.26	17	17	17	17
0.27	17	17	17	17
0.28	16	16	15	15
0.29	13	13	13	12
0.30	12	12	12	12
0.31	11	10	10	10
0.32	10	10	9	9
0.33	9	9	8	8
0.34	8	7	7	7
0.35	8	8	7	6
0.36	7	6	6	6
0.37	6	6	6	6
0.38	5	5	5	5
0.39	5	5	5	4
0.40	5	5	5	4
0.41	5	5	5	3
0.42	4	4	4	3
0.43	3	3	3	3
0.44	3	3	3	3
0.45	3	3	2	2
0.46	2	2	2	2

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

A procedure to group geostatistical realizations of reservoir properties using wavelet analysis and correlation analysis has been developed in this study base on artificial data. In practice, the procedure can be divided into three steps: geostatistical simulation, denoising of realizations using wavelet analysis, and grouping realizations that have similar underlying structures. For the stochastic simulation, a variogram model representing the relationship between the data value and the distance between the estimated data and sampled data is created and used as a conditioning information in simulation process. After that, a certain number of realizations which are equally probable are generated by Sequential Gaussian Simulation. In the second step, realizations are denoised in order to find the main feature of the realizations. Finally, these denoised realizations are then compared one by one by correlation analysis. From the analysis, pairs of denoised realizations that have high correlation values are grouped together. The biggest group is the best representative of the realizations simulated from geostatistical simulation.

In this study, 47 porosity sampling data from 35 regular wells and 3 horizontal wells in 10,240x10,240 ft^2 of domain area were made up. Then, the variogram model, which is spherical model with nugget of 0.03, range of 1,570 ft, and sill of 0.97,of normal score data of the data set was determined as the spatial variability structure of the data set. Sequential Gaussian Simulation technique was used to generate 60 realizations at equal probability using variogram model obtained previously. All realizations then were denoised by soft thresholding method applied from multiresolution wavelet analysis using Daubechies 4 wavelet function for 4 levels. These 60 denoised realizations were paired to find the degree of similarity in term of correlation values. Some cut-off values of correlation values were used to find the grouping result.

Important remarks on the geostatistical simulation are presented as follows:

1. The directional variograms cannot be performed due to a small number of sampled data, which is the same as in real situations.
2. There are many variogram models that can fit the variogram plot. However, the spherical model is often used to represent the reservoir property relationship.

In the second part, important points on denoising of realizations using wavelet analysis can be summarized as follows:

1. In this study, the Daubechies 4 wavelet function, which is a famous function, was used to decompose porosity distributions. The number of decompositional levels, which is 4 in this study, comes from trial and error to find the best result for this study.
2. The soft thresholding method is used in the denoising step because it gives a good result for continuous variable. In addition, the threshold values are calculated from the equation, $\lambda = \hat{\sigma}\sqrt{2\log n}$, where n = the number of data, and $\hat{\sigma}$ = the median of absolute deviation (MAD).

The method and results for grouping denoised realizations can be summarized as follows:

1. The correlation values among denoised realizations are higher and have wider range than those among the original realizations. Therefore, it is better to group denoised realizations since they are correlated at a higher degree.
2. All denoised realizations are compared one by one using correlation analysis. A high correlation coefficient value of a pair of realizations means high degree of similarity between the realizations. The criterion used in the comparison and grouping is that a cut-off value is specified to select highly correlated realizations. The realization that has the largest number of similar realizations based on a certain cut-off value of correlation was chosen as the most representative realization.

3. In this study, the cut-off values of the correlation were varied to examine the output of the grouping. However, from the statistical analysis of the correlation values, the proper cut-off value for the particular set of data used in this study is close to the third quartile of the correlation value of all denoised realizations. Therefore, the cut-off value at the third quartile of the correlation value should be a good choice for future study concerning with this type of problem.
4. Another method of finding a representative from the grouping results is specifying the desired number of members of the biggest group. The best representative of realizations is chosen from the base realization of the biggest group that has the number of members equals to the specified number. However, the chosen representative depends on the specified number of the members.

In this study, results of the grouping at different cut-off values come from only a data set and the proper cut-off value is found around the third quartile of the correlation values of all denoised realizations. Hence, performing the same process to different data sets will require different proper cut-off values. Moreover, it is worth to point out that the distributions of realizations may be different for the same data set if the parameters used in Sequential Gaussian Simulation process change. Therefore, changing the simulation parameters may change the grouping results.

REFERENCES

Aasum, Y., Kelkar, M. G., and Gupta, S. P. An application of geostatistics and fractal geometry for reservoir characterization. SPE Formation Evaluation. March 1991: 11-19.

Athichanagorn, S., Horne, R.N., and Kikani, J. Processing and Interpretation of Long-term Data from Permanent Downhole Pressure Guages. presented at the 1999 SPE Annual Technical Conference and Exhibition, Texas, October 1999.

Chu, L. Schatzinger, R.A., and Tham, M.K. Application of wavelet analysis to upscaling of rock properties. SPE Reservoir Evaluation & Engineering, February 1998.

Deutsch, C.V., and Journel, A. G. GSLIB: Geostatistical Software Library and User's Guide. New York: Oxford University Press, 1992.

Donoho, D.L., and Johnstone, I.M. Ideal spatial adaptation by wavelet shrinkage. Technical Report. Department of Statistics, Stanford University, 1993.

Donoho, D.L., and Johnstone, I.M., Adapting to unknown smoothness via wavelet shrinkage, Journal of the American Statistical Association 90, 1995. 1200-1224.

Hand, J.L. and Moritz, A.L.,Jr. Geostatistical integration of geological, petrophysical, and outcrop data for evaluation of gravity drainage infill drilling at Prudhoe Bay. presented at the SPE 69th Annual Technical Conference and Exhibition, Los Angeles, September 1994: 347-358.

Kikani, J. and He, M. Multi-resolution Analysis of Long-term Pressure Transient Data using Wavelet Methods. presented at the 1998 SPE Annual Technical Conference and Exhibition, New Orleans, September 1998: 117-126.

Kreyszig, E. Advanced Engineering Mathematics. 8th ed. New York: John Wiley & Son, 1999.

Ogden, R. T. Essential Wavelets for Statistical Applications and Data Analysis. Boston: Birkhäuser, 1997.

Poquioma, W., Intevep, S. A., and Kelkar, M. Application of geostatistics to forecast performance for waterflooding an oil field. SPE Advanced Technology Series. vol.2, no.1, March 1994: 142-151.

Pumjan, S. 2106530 Geostatistics (Lecture Note). Bangkok: Chulalongkorn University, 2002.

VITAE

Worapot Laopom was born on June 03, 1973 in Pathumtanee, Thailand. He received his B.Eng. in Mining Engineering from the Faculty of Engineering, Chiang Mai University in 1995. After graduating, he has worked for Electricity Generating Authority of Thailand for 6 years and then continued his study in the Master of Petroleum Engineering program at the Department of Mining and Petroleum Engineering, Faculty of Engineering, Chulalongkorn University.