STOCHASTIC APPROACH IN RESERVE ESTIMATION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

EMİNE BUKET ÜLKER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

THE DEPARTMENT OF PETROLEUM AND NATURAL GAS ENGINEERING

JANUARY 2004

Approval of the Graduate School of Natural and Applied Sciences.

Prof. Dr. Canan Özgen
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Birol Demiral
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Prof. Dr. Mustafa V. Kök
Supervisor

Examining Committee Members

Prof. Dr. Birol Demiral

Prof. Dr. Nurkan Karahanoğlu

Prof. Dr. Suat Bağcı

Prof. Dr. Mustafa Verşan Kök

Assist. Prof. Dr. Evren Özbayoğlu

ABSTRACT

STOCHASTIC APPROACH IN RESERVE ESTIMATION

Ülker, Emine Buket

M.S., Department of Petroleum and Natural Gas Engineering

Supervisor: Prof. Dr. Mustafa V. Kök

January 2004, 75 pages

Geostatistics and more specifically stochastic modeling of reservoir heterogeneities are being increasingly considered by reservoir analysts and engineers for their potential in generating more accurate reservoir models together with usable measures of spatial uncertainty. Geostatistics provides a probabilistic framework and a toolbox for data analysis with early integration of information. The uncertainty about the spatial distribution of critical reservoir parameters is modeled and transferred all the way to a risk conscious reservoir management. The stochastic imaging(modeling) algorithms allow the generation of multiple, equiprobable, unsmoothed reservoir models yet all honoring the data available. This thesis presents stochastic reserve estimation methods as related to the var-

ious stages of development of an oil field. Advances in technology are leading to better deterministic estimates as well as stochastic estimates with narrower ranges. Practices in the industry vary from complete dedication to deterministic or stochastic to a choice of the method depending on the stage of the development.

In this study, reserves are calculated from the data available in Southeastern Turkey by using stochastic methods. Probability density function, number of iterations are important statistical concepts. Increasing number of iterations leads to a normal distribution of histogram.

Keywords: Reserve Estimation, Stochastic Approach, Deterministic Approach, Statistics, ISATIS

ÖZ

REZERV TAHMİNİNDE STOKASTİK YAKLAŞIMLAR

Ülker, Emine Buket

Yüksek Lisans, Petrol ve Doğalgaz Mühendisliği Bölümü

Tez Yöneticisi: Prof. Dr. Mustafa V. Kök

Ocak 2004, 75 sayfa

Jeoistatistik, özellikle de stokastik modelleme, olası rezervin doğruya yakın bir şekilde tahmininin hesaplanması için rezervuar analistleri ve mühendisleri tarafından yoğun olarak kullanılmaktadır. Geoistatistik, olası bir yapı ve veri analizi için bilginin erken ve güncel olarak entegrasyonunu sağlamaktadır. Ayrıca, kritik rezervuar parametrelerinin uzaysal dağılımındaki belirsizlik de risk rezervuar yönetimiyle de modellenmektedir. Bu tez stokastik rezerv hesaplama metodlarının, petrol sahasının gelişme basamaklarına bağlı olduğunu göstermektedir. Teknolojideki gelişmeler hem, analitik metodla yapılan rezerv hesaplamalarında, hem de stokastik metodla yapılan rezerv hesaplamalarında daha küçük aralıklarla tahmin yapılmasını sağlayacaktır. Endüstri alanındaki uygulamalar gelişme süre-

cine baęlı olarak analitik veya stokastik metodun seęilmesine gre deęiřmektedir.

Bu alıřmada, rezerv rakamları, Trkiye'nin gney doęusundaki bir sahadan elde edilen verilerle, olasılık metodları kullanılarak hesaplanmıřtır. Olasılık yoęunluk fonksiyonu, iterasyon sayısı nemli stokastik kavramlardır. Bir model oluřturulurken bu kavramlar dikkatli bir řekilde seilir ve deęerlendirilir. alıřmanın sonuları arasında, iterasyon sayısı arttıka histogramları normal daęılıma yaklařtıęı gzlenmiřtir.

Anahtar Kelimeler: Rezerv Tahmini, Stokastik Yaklařım, İstatistik, Analitik Yaklařım, ISATIS

to my family

ACKNOWLEDGMENTS

I wish to acknowledge the sound advice and guidance of my dissertation director, Prof. Mustafa Versan Kk.

I gratefully appreciate Dr. H. Alkan for their helpful suggestions on the draft of the dissertation.

I also wish to express my gratitude Yıldız Karakee at TPAO for supplying the core data.

Special thanks to my family for their inexhaustible encouragement and motivation during the study. I would like to thank my sister, Demet, for showing great patience and love during my life. I would also like to thank Serhat, his friendship make me happy.

I also wish to express my gratitude TUBITAK, Graduate School of Natural and Applied Sciences, Pet Holding and Kksal Education Council for their supports during this study.

TABLE OF CONTENTS

ABSTRACT	iii
ÖZ	v
DEDICATON	vii
ACKNOWLEDGMENTS	viii
TABLE OF CONTENTS	ix
LIST OF TABLES	xii
LIST OF FIGURES	xiii
CHAPTER	
1 INTRODUCTION	1
2 LITERATURE SURVEY	4
2.1 Definition of Petroleum Reserves	5
2.2 Dan Field Study	10
2.2.1 Derivation of Point-Scale Variogram	13
3 GEOSTATISTICS AND ISATIS SOFTWARE PROGRAM	18
3.1 Introduction	18
3.1.1 Data Integration Method:	19
3.1.2 Calibration of the Seismic Data:	20
3.2 CASE STUDY	20
3.2.1 55 Well Data Case	20
3.2.2 Learning From the Data: Variogram Analysis	22
3.2.3 Finding Relationships	23
3.2.4 Using What Has been Learned and Found	23
3.2.5 Assessing Map Accuracy :Conditional Simulation	24

4	STATEMENT OF THE PROBLEM	26
5	RESULTS AND DISCUSSION	27
5.1	Results of Analyzed Data	27
5.1.1	Estimation of Hydrocarbon Volumes	38
5.2	DISCUSSION	42
6	CONCLUSION	43
	REFERENCES	45
	APPENDICES	48
A	BASIC CONCEPTS OF STATISTICS	48
A.1	INTRODUCTION	48
A.2	BASIC CONCEPTS	48
A.2.1	COMBINATION OF EVENTS	49
A.2.2	PROBABILITY	50
A.2.3	PROBABILITY LAWS	50
A.2.4	UNIVARIATE DISTRIBUTION	52
A.3	THE CUMULATIVE DISTRIBUTION FUNCTION	53
A.3.1	Discrete CDF's	55
A.3.2	Continuous CDF's	55
A.3.3	Empirical CDF's	56
A.4	Order Statistics	57
A.5	PROBABILITY DISTRIBUTION FUNCTIONS	57
A.5.1	Discrete PDF's	57
A.5.2	Continuous PDF's	58
A.5.3	Empirical PDF's	58
A.6	SAMPLING AND STATISTICS	60
B	FEATURES OF SINGLE-VARIABLE DISTRIBUTIONS	61
B.1	NON-CENTERED MOMENTS	61
B.1.1	The Expectation	61
B.1.2	Properties of the Expectation	62
B.2	CENTERED MOMENTS	64
B.2.1	Variance	64
B.2.2	Properties of the Variance	65
B.2.3	Standard Deviation	66

B.3	COMMON PDF'S	66
B.3.1	The Binomial Distribution	66
B.3.2	The Central Limit Theorem	67
B.3.3	The Normal (Gaussian) Distribution	68
B.3.4	The Log-Normal Distribution	69
B.3.5	p-Normal Distribution	69
B.3.6	Uniform Distribution	70
B.3.7	Triangular Distribution	70
B.3.8	Exponential (Boltzmann) Distribution	71
B.4	ESTIMATORS AND THEIR ASSESSMENT	72
B.5	The Elements of Estimation	72
B.6	Desirable Properties of Estimators	73
B.6.1	Estimator Bias	73
B.6.2	Estimator Efficiency	74
B.6.3	Estimator Robustness	74
B.6.4	The Median	74

LIST OF TABLES

2.1	Resource Uncertainty Categories	7
5.1	Core Data of Different Three Wells in Field A	27
5.1	Core Data of Different Three Wells in Field A (cont.)	28
5.2	Descriptive Statistics of Porosity Values in Field A	29
5.3	Descriptive Statistics of Permeability Values in Field A	29
5.4	Regression Analysis	31
5.5	Variance Analysis	31
5.6	Probability Density Function values of Porosity	37
5.7	PDF of Field A	39
5.8	Statistical results of OOIP and ROIP	40
5.9	Statistical results of OOIP and ROIP	40

LIST OF FIGURES

2.1	Reserves and Resources	8
2.2	Resource Status Categories	9
2.3	Volume Terminology	10
2.4	The measurement of volume in different data scales	13
2.5	Variograms for fine-scale and block averaged porosity values	14
2.6	Comparison of Theoretical and Experimental Variograms	15
2.7	Decrease in Sill Value	16
2.8	Gaussian shape model and experimental variograms of the original fine scale simulation	17
5.1	Descriptive Statistics of the Porosity values in Field A	30
5.2	Descriptive Statistics of the Permeability values in Field A	30
5.3	Porosity Permeability Plot	32
5.4	Contour Plot of Porosity	32
5.5	Contour Plot of Permeability	33
5.6	Schematic of auto correlogram with lag distance 0 <i>ft</i>	33
5.7	Schematic of auto correlogram with lag distance 1 <i>ft</i>	33
5.8	Schematic of auto correlogram with lag distance 2 <i>ft</i>	34
5.9	Schematic of auto correlogram with lag distance 5 <i>ft</i>	34
5.10	Porosity Probability Plot	35
5.11	Auto Correlation Function of Porosity	35
5.12	Permeability Probability Plot	36
5.13	Auto Correlation Function of Permeability	36
5.14	Probability Distribution Function	38
5.15	Number of Iteration versus OOIP	41
5.16	Number of Iteration versus ROIP	41

CHAPTER 1

INTRODUCTION

Geostatistics has entered a new age. Until few years ago, geostatistics could still be considered a leading-edge technology, developed and applied by a handful of specialists. Today, geostatistics is routinely applied to most of the reservoir characterization projects worldwide, providing the quantitative geology support needed in the reservoir simulation phase. This introduction chapter describes how a geostatistical model can be built by integrating all the information generated in the individual disciplines, namely Geophysics, Petrophysics, Sedimentology, Stratigraphy and Basic Reservoir Engineering. Stochastic approach guarantees the internal consistency of the reservoir study and provides a robust model to be upscaled in the forward simulation phase. The information contained in the 3D geological model can also be averaged and exported to a conventional mapping algorithm, thus providing a consistent set of traditional 2D maps of reservoir properties (Net to Gross, net sand, porosity, permeability) [17].

This initial stage of the data analysis is very important, as it affects individual

disciplines named previously and thus the final results of the study.

Early stage of the reservoir characterization is geophysics which represents an invaluable source of potential information, both from a structural and a stratigraphic point of view. From a stratigraphic point of view, geophysics is the only static discipline that offers the possibility to directly investigate the inter-well area. Both lithological and petrophysical attributes can be modeled using the seismic volume, and these results can be integrated in the stochastic process to improve the geological modeling. In particular, attribute maps can be generated (e.g., amplitude or seismic impedance), which can be correlated to reservoir properties like average porosity, net to gross or facies distribution. The quantitative log interpretation phase, established for individual wells on the basis of core and log data, can be distributed throughout the reservoir by means of the stochastic model.

In order to build a geostatistical model the data from the Petrophysics must be integrated into other disciplines. The quantitative log interpretation phase, established for individual wells on the basis of core and log data, can be distributed throughout the reservoir by means of the stochastic model. A conventional quantitative well log analysis or a neural network-based approach, allows to characterize the litho-facies from a petrophysical point of view (ϕ , K , S_w). As a consequence, the petrophysical distribution is closely linked to the sedimentological model.

Sedimentological model is the core description which allows to define a facies classification scheme and a depositional environment for each litho-unit. A con-

ceptual sedimentological model based on present day analogue can be useful in the definition of the orientation, shape and size of the main sedimentary bodies. The lithological and the depositional model of the reservoir are represented in the stochastic model by means of facies and relevant distribution functions (variograms, proportion curves,), that can be established by means of well data as well as conceptual model.

The integration of dynamic data into the stochastic model has been a topic of intense research in the last decade, since this type of information is invaluable in the definition of a static reservoir model that will be used in a forward simulation phase. In the stochastic modeling phase, litho-facies are usually simulated first, while petrophysical properties are simulated in a later stage, as continuous or discrete variables, within the litho-facies framework [17].

The scope of this thesis is to estimate reserve with core data from the South-eastern part of Turkey in order to check the validity and the accuracy of stochastic approach.

CHAPTER 2

LITERATURE SURVEY

The increasing importance of worlds energy situation requires more precise studies of hydrocarbon reservoirs. The evaluation of reserves are generally obtained by using various deterministic and stochastic modeling techniques. The input data for the model is acquired from the field tests and experiments. Reserves estimation may be based primarily on initial volumetric estimates or at a later stage in field development on performance analysis involving numerical simulation and material balance methods. It may be necessary in some situations to handle initial volumetric more appropriately by applying suitable probabilistic methods which make use of all available data, including seismic, and which furthermore given an appreciation for uncertainties and accuracy associated with volume estimates.

2.1 Definition of Petroleum Reserves

In February 2000, three major international organizations, the SPE, WPC and AAPG jointly approved and published a significant document entitled Petroleum Resources Classification and Definitions. This achievement represented a major step forward in the industry attempts to improve the level of consistency in the documentation and description of resources that do not satisfy the criteria for reserves. The existing (1997) SPE/WPC [28] reserves definitions were unaffected; the new definitions simply supplemented the 1997 definitions in order to cover the entire resource base. In addition to providing a single set of terminology for industry use, the document also provided a new classification system that was designed to facilitate the development of portfolio management systems by individual countries or companies. The potential benefits of the SPE/WPC/AAPG classification system are twofold. Firstly, it provides the facility to expand the sub-division of project maturity, in order to classify projects on the basis of the work required (reflecting business decisions and budgeting) to move them towards commercial production and revenue. This ensures a continuous focus on the objectives of asset management, as well as providing the potential for a detailed classification of the resource base which illustrates clearly the balance of the portfolio and hence facilitates rational decisions on future acquisitions or divestitures. The second benefit of the new system lies in the consistent approach to establishing a range of uncertainty, regardless of project maturity, which allows the very powerful tool of uncertainty tracking to be applied. Despite being published by APRs as long ago as 1956, this method is surprisingly little used. In

practice, it is the only way to track estimates of recoverable volumes in a meaningful way, and to establish on an on-going basis the validity of the methods used to estimate those volumes. Introduction The establishment of an internationally accepted petroleum resource classification system is long overdue. Despite the obvious benefits of shared definitions of resource categories and related terminology, the industry has developed in such a way that individual countries and companies have been left to first design their own systems which preclude easy comparisons between them. This has been a particular problem in the large corporate mergers seen in recent years, but is also a significant limitation when considering and comparing the resource base of individual countries.

The development of resource definitions has, until recently, largely focussed only on reserves, i.e. those resources that are considered to be discovered and commercially recoverable. Indeed, from 1965 until 1987, the Society of Petroleum Engineers (SPE) published only a definition of proved reserves. In 1987, the definitions were expanded to include probable and possible reserves, but no other resource categories. These definitions were then updated with limited, but significant, changes in 1997, following extensive collaboration with, and approval of, the World Petroleum Congresses (WPC). In 1987, the WPC had published a classification and nomenclature system for all categories of resources as a further development of their co-operation in approving the 1997 reserve definitions, and recognizing the potential benefits to the industry, the SPE and WPC initiated work on developing definitions to encompass the entire resource base. Subsequently, the SPE and WPC were joined by the American Association of Petroleum Geologists (AAPG) and, in February 2000, the three organizations

jointly approved a new document covering Petroleum Resources Classification and Definitions [24].

Table 2.1: Resource Uncertainty Categories

Resource Uncertainty Category	Reserves		Other Resources	
	Deterministic	Probabilistic	Deterministic	
Low estimate	Proved (1P)	P_{90}	Low	
Best estimate	Proved plus Probable (2P)	P_{50}	Most likely	
High estimate	Proved plus Probable plus Possible (3P)	P_{10}	High	

P =probability

It could be said that this process was entirely the reverse of an optimum approach, in which an overall classification system would be the starting point, with the definitions of the various categories (including reserves) following as a refinement of the classification system. Instead, the system has developed from a definition of proved reserves alone (initially in 1965 as far as SPE is concerned, though there were definitions for proved reserves prior to that) and only finally encompassed all resources some 35 years later. In contrast, the Former Soviet Union (FSU) established an integrated resource classification system which has been in use for over 50 years. Resource Classification Systems Much of industry practice has been based, at least implicitly, on the resource classification system published by McKelvey. The so-called McKelvey Box [24] was first published in 1972, with a revised version following in 1975. Figure 2.1 indicates McKelvey Box reserves and resources.

Figure 2.1: Reserves and Resources

The 1972 [23] version is shown in Figure 2.1, which clearly highlights the requirement for reserves to be discovered and economically recoverable. All other volumes are classified as resources. While its simplicity is attractive, it has some limitations when trying to deal with project maturity (likelihood of achieving commercial production) and volumetric uncertainty (range of potentially recoverable volumes, should production occur), which are discussed further below.

Although the SPE/WPC [28] reserves definitions they explicitly require all reserves categories (proved, probable and possible) to be commercially recoverable, industry practice has, in some cases, adopted a rather loose interpretation of commerciality. This problem has perhaps been exacerbated by the lack of a box in which to place discovered volumes which are not currently commercially viable to develop. For example, some companies use the possible reserves category as a sort of catch all for any volumes which are discovered but for which there are no plans for development, e.g. a known gas field in a remote region with no obvious

market. This application of the definitions can be summarized by a modified version of the McKelvey Box [24] in Figure 2.2.

Figure 2.2: Resource Status Categories

Consequently, possible reserves could include upside potential in a producing field or volumes in an accumulation which are unlikely to be developed in the foreseeable future. These two types of reserves are, of potential for establishing a meaningful portfolio management system. In particular, as only a single number (the possible reserve volume) is captured for the undeveloped discovery, no measure of uncertainty in that estimate is documented. Recognizing the need for an integrated internal classification system which specifically includes non-commercial discoveries (as well as undiscovered volumes), some companies have used terms such as contingent resources to categorize such volumes while maintaining the ability to capture a range of uncertainty in the estimated recoverable volumes. Such an approach simply provides terminology for the other parts of the McKelvey Box [24], as shown on Figure 2.3, but represents a significant

improvement in the application of the reserves definitions.

Figure 2.3: Volume Terminology

The term contingent resources is considered particularly appropriate, since the commerciality of the accumulation is contingent on some action or change in circumstances, such as further appraisal and data acquisition, development of a market, lower development costs and/or higher sales prices.

2.2 Dan Field Study

A conventional reservoir geology work, in terms of stratigraphic correlations, log interpretation and facies definitions, represents the basis of geological workflow. The initial stage of the work is very important, as it conditions all the following phases of the work and hence the final results of the integrated study. A case study about Dan Field in Danish North Sea is given in [1].

Reconciling data from different scales is a longstanding problem in reservoir characterization. Data from core plugs, well logs of different types, and seismic

data must all be accounted for in the construction of a geostatistical reservoir model. These data are at vastly different scales and it is inappropriate to ignore the scale difference when constructing a geostatistical model. Geostatistical scaling laws were devised in the 1960s and 1970s primarily in the mining industry where the concern was mineral grades in selective mining unit (SMU) blocks of different sizes. These principles can be extended to address problems of core, log and seismic data. The adoption of these classic volume-variance or scaling relationships presents some challenges. Some specific concerns are (1) the ill-defined volume of measurement, (2) uncertainty in the small-scale variogram structure, and (3) non-linear averaging of many responses including acoustic properties and permeability. The application of volume-variance relations for up scaling and downscaling techniques to integrate data of different scales is demonstrated. Practical concerns are addressed with data from a chalk reservoir in the Danish North Sea. A direct sequential simulation algorithm accounting for data at all scale is documented. Within the petroleum industry and many other fields where geostatistical models are constructed, the treatment of data of different scale is often ignored. The core and log data may be averaged in the vertical direction to the scale of the modelling cells; however, this only partially addresses scale difference. In other examples, a fine-scale model is constructed and then numerically averaged to larger scale. This may be applied in a nested fashion due to computational limitations. The direct simulation of grid block values conditioned to fine-scale data carried out by co simulation using cross-covariance between fine- and coarse-scale values has been described. Notwithstanding the importance of

accounting for data at different scales, the use of geostatistical scaling laws has not seen wide application in petroleum geostatistics. This is due mainly to unfamiliarity with the techniques and scaling laws. Recalling and demonstrating such techniques will address this unfamiliarity. How the variogram changes with volumetric scale is explained by scaling laws. As scale increases, the range of correlation increases, the variance and variogram sill decrease, and the nugget effect also decreases. After a recall of theory, the application of scaling laws is illustrated with a synthetic example and with real data from a Danish chalk reservoir. Core and well log data are used. These data measure significantly different volumes. The volume of the core measurement is well understood; however, the volume of the interpreted well log derived porosity is less well understood. The statistics of each data type together with analytical volume-variance relationships can be used to quantify the volume of investigation of the well log data. Figure 2.4 shows the measurement of volume (in cubic meters) in the different data scales. Note that the scale distance between core and wire line-log is nearly as large as between log- and modelling cell volumes.

Figure 2.4: The measurement of volume in different data scales

An illustration of the different scales Figure 2.4 shows that the change of scale from core to log measurement volumes is nearly as large as the jump from log volume to that of a geological modelling cell. The ultimate goal of this work is to illustrate how data of different scales may be used simultaneously in the construction of high-resolution geostatistical models. When the different types of data are all *hard* in the sense that they do not contain significant errors or uncertainties relative to the property being modelled, it is possible to use block kriging. Certain data types such as seismic contain uncertainties related to the great distance of measurement and calibration of the measured acoustic properties to the petrophysical properties of interest. In this case, it is necessary to use block cokriging.

2.2.1 Derivation of Point-Scale Variogram

The scaling laws developed above may be applied on each nested structure in the variogram. As the first step the point scale variogram must be deduced. The

scaling laws are concerned with the changes to the nugget effect, the variogram range and the sill. For the downscaling to point-scale from our data, we need to consider all three elements. The fine-scale data has no nugget effect in the correlation structure. Therefore the point-scale variogram is assumed to have no nugget. For the spherical structure the correction to point variogram range a_v is based on the fine-scale range a_v is: $a_v = (|v| - |p|)$, where p = point scale, which is zero and v = fine-scale.

The theoretically derived variogram for coarse-scale may be calculated and compared to the experimental variogram from the block-averaged coarse-scale data. The closeness of the match is a measure of the efficacy of the scaling relations. The comparison of the theoretically predicted coarse-scale variogram and the experimental variogram obtained from the block averaged data is shown in Figure 2.5 , having a very good agreement.

Figure 2.5: Variograms for fine-scale and block averaged porosity values

Figure 2.5 indicates the variograms for fine-scale and block averaged porosity values. Upper set is model variogram (line) and experimental variogram (dots)

of simulated fine-scale showing a very good reproduction by the simulation. The lower set is the theoretically predicted variogram (line) and the experimental variogram for the block average porosity values (dots). In order to further outline the performance of the scaling laws, the procedure has been applied at other coarse scale resolutions of 1.0, 2.0 and 4.0 m block averages. The comparison between the theoretical predictions and the experimental variograms gives a near perfect match as is shown in Figure 2.6

Figure 2.6: Comparison of Theoretical and Experimental Variograms

Figure 2.6 indicates the comparison of theoretical and experimental variograms for *a* - fine-scale simulation; *b* - 0.50 m block average; *c* - 1 m block average; *d* - 2 m block average; *e* - 4 m block average. They all show very good match between prediction and experimental variograms.

Figure 2.7: Decrease in Sill Value

In order to illustrate the evolution of the variogram, the decrease in sill (variance) value (normalized to the sill for the 0.02 m data = core-scale) for the theoretically predicted variograms as the length scale of averaging is increased in steps; is shown as a function of the averaging volume in Figure 2.7. However, this function is only valid for this particular model construction, and heavily depends on the variogram structure at the fine scale level. The synthetic fine-scale model has also been used to investigate how non-overlapping volume averaging influences the up scaling of variograms. For this purpose a moving 0.5 m window filter has been applied with a simple square filter function on the fine scale data.

The resulting experimental variogram shows that the variogram structure is changing from the original fine scale spherical model into a more Gaussian shaped

model for the moving average data as seen in Figure 2.8. This figure indicates the Gaussian shaped model and experimental variograms of the original fine scale simulation (upper line and open dots) and the experimental variogram for the up scaled data using a moving window average (lower filled dots) indicating a more Gaussian type variogram model. The open dots are the experimental variogram for the block averaged data. The lower line is the predicted variogram model assuming spherical structure in the scaling calculation.

Figure 2.8: Gaussian shape model and experimental variograms of the original fine scale simulation

CHAPTER 3

GEOSTATISTICS AND ISATIS SOFTWARE PROGRAM

3.1 Introduction

The important question is what we want from a stochastic method. Before discussing various stochastic modeling methods, it will be useful to consider some of the possible goals of a reservoir modeling exercise because the appropriateness of any particular method depends, in large part, on the goal of the study. Stochastic methods vary considerably in their requirements for time, money, human resource and computer hardware and software. In this case ISATIS is used as a software program which is well-adapted to core and seismic data. To run the ISATIS in reservoir modeling the most important challenge is data integration. Relevant information comes from various sources at various scales with varying degrees of reliability. The geological data originates at many scales, from the microscopic image of a pore volume to the basin-scale interpretation of a depositional system;

geophysical (seismic) data, often the most prevalent, span from the macro scale of well logs to the scale of the entire field. The objective is not for each reservoir discipline to maximize its contribution to the final reservoir model, rather the challenge is for each discipline to understand the specific forecasting goal and provide only the relevant information in a usable format and with some assessment of uncertainty. The major contribution of geostatistics to reservoir modeling lies in data integration; providing formalism to encode vital, possibly non-numerical information, combine different data accounting for uncertainty into final forecast. The ISATIS manual [12] encompasses the case study of integration of well and seismic (petrophysical) data using geostatistics.

3.1.1 Data Integration Method:

Geostatistical data integration methods are analogous to a regression problem. The objective problem is to make a prediction at an unsampled location for a variable A based on its relationship (linear combination) with one or more secondary variables. Variable A is sparsely sampled compared to the secondary data. Variable A in this manual case study is average porosity at 55 well locations and with 16900 acoustic impedance AI values comprising the secondary data set.

If a correlation can be established between the primary variable porosity and the secondary variable, AI then it would be desirable to use AI to predict porosity at inter well locations. Such an integration takes the advantage of the locally more precise but sparse well data and the densely sampled, but locally soft data. Several geostatistical data integration methods are available. Generally, the choice depends on the type of data available and the purpose of the study.

3.1.2 Calibration of the Seismic Data:

Acoustic impedance was estimated by point kriging at each well location using a linear variogram model and eight CDP neighborhoods from the 42000 impedances not from the resampled data. The AI values were simply transferred to the grid nodes using a linear variogram. There was no interpolation in the last step because the grid node dimensions are congruent with the 16900 CDP locations in the resampled seismic data.

3.2 CASE STUDY

Because the purpose of the study is to illustrate the effect of integrating seismic data as secondary or soft constraining data into a reservoir description a case study was designed to mimic the early stage in a reservoir's life; late exploration or appraisal.

3.2.1 55 Well Data Case

This case which is given by ISATIS manual, represents a typical Permian Basin well spacing with a drainage area of 40 – *ac*. First the data loaded and the information constituted of two separate ASCII files [12]:

"The file `petroleum_wells.hd`" contains the data collected at 55 wells. In addition to the coordinates the file contains the target variable porosity and the selection which concerns the initial appraisal wells

"The file `petroleum_seismic.hd`" contains a regular grid where one seismic attribute has been measured: the normalized acoustic impedance. The grid is

composed of $260 * 130$ nodes at $40\text{ ft} * 80\text{ ft}$.

Then the study preferences are set: Default unit; X, Y and Z units are in foot unit.

Now the geostatistical method can be applied to quantitatively integrate well and seismic data and quantitatively assess map accuracy. The geostatistical method is a new way to solve these mapping problems. The advantage of this method is that it makes the geoscientists and scientists think about the data before making the map. Not only does the method yield a map with certain statistically desirable properties, but also gives the user a quantitative measure of the reliability of the map.

Geostatistics is the application of several statistical tools that are used to determine the spatial distribution of geologic variables. Variograms, kriging, kriging with external drift, co-kriging and conditional simulation are some of the geostatistical tools commonly used. Geostatistical method elevates the solutions of mapping problems to a new plane. At present engineers collect the data, plot the data on a map and contour the data. The geostatistical method gives the user a methodology for quantitatively determining the spatial characteristics of geologic variables prior to contouring. The spatial information is then used to make the maps and then assess, in a reasonably objective way, the accuracy of the resulting maps.

Briefly the geostatistical method is a four-step procedure that calls on several geostatistical tools. These steps are:

1. Learn from the data through simple statistical analysis (such as calculating

means, variances minimum and maximum values and histogram plots) and variogram analysis.

2. Find relationships between the data sets (if possible) through cross plots usually done in geophysics by trying to find relationships between sparse well data and relatively dense seismic data.
3. Use what has been learned and found in the data to determine the spatial distribution of control points (i.e., grid for subsequent contouring). The tools most often used in this step are kriging with external drift and co-kriging.
4. Assess the accuracy/error/risk of the map made in step 3 using conditional simulation. The assessment of risk is perhaps the greatest leap forward that geostatistics provides in solving mapping problems.

Finding relationship between well and seismic data is not absolutely necessary; however if a relationship is found, it greatly reduces uncertainty in the results.

3.2.2 Learning From the Data: Variogram Analysis

A variogram is a mathematical technique to quantify the spatial correlation or continuity of a geological variable. A variogram is a plot of the average squared difference in the value between data points. Variogram modeling is a mathematical necessity that ensures that kriging will work.

The key parameters that describe the variogram model are:

1. Nugget effect, or the value of the model at zero distance

2. Sill or the variance of the data
3. Range or the break over point from the correlated to uncorrelated zone of the variogram

The nugget is a gauge of measurement uncertainty. If it were zero, then the data would be honored exactly, otherwise the grid values would not honor the well data. Variogram analysis can be used to identify and the quantify the fact that spatial continuity can be longer in one direction than another (anisotropic) in the control points via directional variograms. The direction and magnitude of the anisotropy can be used in subsequent steps of the geostatistical method.

3.2.3 Finding Relationships

The second step in geostatistical method is to find relationships between the seismic and well data. It is the most difficult and least developed (in terms of tools available) of the four steps. Cross plotting of the well variable and the seismic variable at the well locations will sometimes lead to finding usable relations for cokriging.

3.2.4 Using What Has been Learned and Found

Kriging is gridding algorithm that estimates a grid node value such that for the parameter of interest, the squared difference between the grid node value and the surrounding control points is the consistent with the variogram model. Kriging answers the following question :”What value should the grid point be assigned so that it best fits the variogram model?”. Spatial relationships that have been

learned by the variogram from the well control are applied to the areas between the control points in kriging.

Kriging is mathematically achieved by calculating the grid value as a weighted average of the surrounding control points and the grid node to be calculated and how close the control points are to each other and maintains the spatial relationship given by the variogram model. The weights are assigned in such a way to minimize the variance in the least-square sense, thus eliminating systematic overestimation or underestimation error,

Co-kriging looks at not only spatial relationships in the data to be gridded (usually a well parameter such as porosity or depth) but also spatial relationships in a second denser data set is referred as to the guide data. In addition to using variogram models for both the control data and the guide data, cokriging also uses a cross variogram that quantifies the spatial cross correlation between the two data sets.

Co-kriging calculates a grid value as a weighted average of control and guide points. Co-kriging takes into account how far the control and guide points are from the grid point to be computed and how close control and guide the data are, co-kriging also honors spatial relationships found in the variogram for the control data and guide data and in the cross variogram between these data.

3.2.5 Assessing Map Accuracy :Conditional Simulation

Conditional simulation is a mapping process that quantifies contouring uncertainty. When contouring data, an engineer always has the ability to move a contour line.

Conditional simulation is a mathematical way to contour a set of data(with or without a guide data) several hundred different ways. The conditional simulation maps, sometimes referred to as images, honor what is known about the true surface.

Each conditional simulation image

1. has approximately the same mean and variance as control points
2. has approximately the same variogram of the control data approximately honors the control point values

A guide data set(such as seismic data) can be used if so desired. Each image can be considered as a slightly different and reasonable way to contour the data. Honoring the control data and the mean, variance and variogram model of the control data ensures that each map is geologically reasonable. The average of the several hundred images is (by mathematical meaning) the kriged solution. If a guide data set is used then the average of the several hundred conditional simulation images is the KED (kriging with external drift) solution.

Another way to explain conditional simulation is to say that it is a set of pseudorandom numbers in the form of a grid that is organized to "have the same spatial continuity or variogram model as the control points", "have the same mean and variance as the input data" honor the control data.

In general, conditional simulation maps contain much more detail than the map produced by kriging or KED. Conditional simulation maintains detail in a map, but kriging or KED demands that the map be unbiased and minimum variance(or be smooth) at the expense detail [29].

CHAPTER 4

STATEMENT OF THE PROBLEM

The scope of this study is to use some basic concepts of probability and statistics to deal with uncertainty encountered at every stage of the exploration. This thesis aims at using stochastic approach for reserve estimation of the field in the absence of data.

The main objective in this stochastic model application is to find the effects of different probability distribution functions, number of iterations, correlation coefficients and the nature of data. After assuming probability density function for the input variables, the model is constructed and a software program ISATIS is used to make the necessary calculations. In order to check the validity of the process, three different wells from the same reservoir in Southeastern Turkey are chosen.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 Results of Analyzed Data

Knowledge of the reserves and rates of recoveries are important in the sale and exchange of oil properties. Thus, the original oil in place (OOIP) and recovery oil in place (ROIP) is one of the important parameters of reservoir. The core data obtained from these three different wells given in Table 5.1 aim at calculating the OOIP and ROIP using stochastic approach.

Table 5.1: Core Data of Different Three Wells in Field A

<i>well</i>	well head,m	deep interval,m	depth m	porosity %	perm. md	grain density,g/cc	saving %
<i>1</i>	490.9	1926 – 1935	0.9	18.2	4.98	2.7	33
<i>1</i>	490.9	1926 – 1935	0.09	19.09	8.18	2.7	33
<i>1</i>	490.9	1926 – 1935	0.84	14.58	5.07	2.69	33
<i>1</i>	490.9	1926 – 1935	0.5	17.91	3.24	2.69	33
<i>1</i>	490.9	1935 – 1944	0.15	15.91	1.54	2.7	55
<i>1</i>	490.9	1935 – 1944	0.79	17.93	3.2	2.68	55
<i>1</i>	490.9	1935 – 1944	0.34	12.09	0.58	2.68	55
<i>1</i>	490.9	1935 – 1944	0.65	17.65	1.96	2.7	55

Table 5.1: Core Data of Different Three Wells in Field A (cont.)

<i>well</i>	well head,m	deep interval,m	deep m	porosity %	perm. md	grain density,g/cc	saving %
2	475.8	1903 – 1908.8	0.2	18.01	2.74	2.68	41
2	475.8	1903 – 1908.8	0.4	20.37	2.99	2.69	41
2	475.8	1908.8 – 1918	0.4	17.1	2.13	2.69	93
2	475.8	1908.8 – 1918	0	17.07	1.69	2.69	93
2	475.8	1908.8 – 1918	0.29	17.87	2.26	2.69	93
2	475.8	1908.8 – 1918	0.7	15.85	2.17	2.69	93
2	475.8	1908.8 – 1918	0.2	18.92	3.25	2.68	93
2	475.8	1908.8 – 1918	0.75	17.83	3.74	2.68	93
2	475.8	1908.8 – 1918	0.7	19.59	3.07	2.68	93
2	475.8	1908.8 – 1918	0.2	17.63	3.08	2.68	93
2	475.8	1908.8 – 1918	0.65	19.56	3.27	2.69	93
2	475.8	1908.8 – 1918	0.2	16.91	1.45	2.69	93
2	475.8	1908.8 – 1918	0.65	15.67	1.6	2.69	93
2	475.8	1908.8 – 1918	0.29	18.35	1.47	2.69	93
2	475.8	1908.8 – 1918	0.7	13.9	0.73	2.69	93
2	475.8	1908.8 – 1918	0.29	17.48	6.09	2.69	93
2	475.8	1908.8 – 1918	0.75	16.93	1.33	2.68	93
2	475.8	1908.8 – 1918	0.95	16.61	1.67	2.68	93
2	475.8	1908.8 – 1918	0.4	20.58	2.37	2.71	93
2	475.8	1908.8 – 1918	0.65	18.68	2.45	2.68	93
2	475.8	1918 – 1927	0.09	17.6	2.53	2.69	63
2	475.8	1918 – 1927	0.5	20.13	6.63	2.68	63
2	475.8	1918 – 1927	0.9	21.55	6.52	2.68	63
2	475.8	1918 – 1927	0.2	19.68	5.31	2.72	63
2	475.8	1918 – 1927	0.79	18.79	3.78	2.70	63
2	475.8	1918 – 1927	0.04	16.96	2.97	2.68	63
2	475.8	1918 – 1927	0.9	18.15	5.05	2.68	63
2	475.8	1918 – 1927	0.4	20.25	5.03	2.68	63
2	475.8	1918 – 1927	0.9	15.83	1.52	2.68	63
2	475.8	1918 – 1927	0.2	14.95	1.61	2.67	63
2	475.8	1918 – 1927	0.5	15.79	1.5	2.67	63
2	475.8	1918 – 1927	0.9	19.04	2.41	2.69	63
2	475.8	1918 – 1927	0.59	2.65	0.07	2.66	63
2	475.8	1918 – 1927	0.9	7.61	0.09	2.7	63
3	466	1898.5 – 1907.5	0.55	3.89	2.68	2.71	100
3	466	1898.5 – 1907.5	0.4	4.35	0.03	2.71	100
3	466	1898.5 – 1907.5	0.7	5.38	0.1	2.7	100
3	466	1898.5 – 1907.5	0.	16.67	8.21	2.68	100
3	466	1898.5 – 1907.5	0.15	20.19	4.02	2.69	100
3	466	1898.5 – 1907.5	0.45	20.16	4.84	2.69	100
3	466	1898.5 – 1907.5	0.45	16.23	1.9	2.69	100

Table 5.1 indicates the core data of the three different wells in Field A. The third column indicates the deep interval in which core sample is taken. The last column indicates the saving percentage of the core sample when it was taken to the surface. The 5th, 6th and 7th column indicates the physical properties, porosity, permeability and the grain density of the core sample respectively.

Table 5.2: Descriptive Statistics of Porosity Values in Field A

<i>Variable</i>	<i>N</i>	<i>Mean</i>	<i>Median</i>	<i>TRMean</i>	<i>StDev</i>
<i>Porosity</i>	53	16.153	17.48	16.653	4.337
<i>Variable</i>	<i>SE Mean</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Q₁</i>	<i>Q₃</i>
<i>Porosity</i>	0.596	2.650	21.550	15.310	18.855

Table 5.2 indicates the descriptive statistics of the porosity values where the minimum porosity value is 2.65 and maximum value is 21.55 and the mean of 53 porosity value is 16.153.

Table 5.3: Descriptive Statistics of Permeability Values in Field A

<i>Variable</i>	<i>N</i>	<i>Mean</i>	<i>Median</i>	<i>TRMean</i>	<i>StDev</i>
<i>Permeability</i>	53	2.817	2.410	2.683	1.978
<i>Variable</i>	<i>SE Mean</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Q₁</i>	<i>Q₃</i>
<i>Permeability</i>	0.272	0.030	8.210	1.510	3.760

Table 5.3 indicates the descriptive statistics of the porosity values. The mean of 53 permeability values are obtained 2.817 where the minimum value is 0.030 and the maximum value is 8.210.

Figure 5.1: Descriptive Statistics of the Porosity values in Field A

Figure 5.1 shows the descriptive statistics of the porosity values in Field A with the confidence interval 0.95. The red line indicates the normal distribution and the

Figure 5.2: Descriptive Statistics of the Permeability values in Field A

Figure 5.2 shows the descriptive statistics of the permeability values in Field A

with the confidence interval 0.95. The red line indicates the normal distribution and the blue blocks are the histogram of the permeability values.

Regression Analysis

The regression equation is

$$porosity = 12.6 + 1.26 * permeability \quad (5.1)$$

Table 5.4: Regression Analysis

<i>predictor</i>	Coefficient	St.Dev	True Mean	Probability
<i>constant</i>	12.6102	0.8629	14.61	0.000
<i>permeability</i>	1.2575	0.2515	5.0	0.000

Table 5.4 gives the linear relation with the porosity and the permeability value. The equation of a linear is $y = mx + n$ where n is the constant and m is the slope. From Table 5.4 the constant value is 12.6 and the slope value is approximately 1.26. The regression line gives an unbiased estimate for the arithmetic average because this average is a linear combination of the estimates. This is a result of the Gauss-Markov Theorem which assures that all linear combinations of the estimates are unbiased [14].

Table 5.5: Variance Analysis

<i>Source</i>	DF	SS	MS	F-test	Probability
<i>Regression</i>	1	321.75	321.75	25.01	0.0
<i>Residual Error</i>	51	656.23	12.87		
<i>Total</i>	52	977.98			

Table 5.5 indicates the analysis of variance. In Table 5.5 DF refers to degrees of freedom, SS refers to sum of squares and MS refers to mean square values. The

model sum of square is regression sum of squares and the error sum of squares. The mean square values are sums of squares divided by their respective degrees of freedom.

Figure 5.3: Porosity Permeability Plot

Dependency between porosity and permeability often exhibits exponential trend. In practice porosity-permeability data plotted in log-log co-ordinates never lie on a straight line but rather form a cloud of points showing a certain tendency. In such cases as shown in Figure 5.3 deterministic evaluation of permeability based on levels of confidence, can represent minimum, most likely and maximum values of the initial well production rate.

Figure 5.4: Contour Plot of Porosity

Figure 5.4 indicates the distribution of porosity values with respect to permeability (md) and depth (m). It is seen from the Fig 5.4 the porosity distribution

over the Field A.

Figure 5.5: Contour Plot of Permeability

Figure 5.5 indicates the distribution of permeability values with respect to porosity and depth (m).

Figure 5.6: Schematic of auto correlogram with lag distance 0 ft

Figure 5.7: Schematic of auto correlogram with lag distance 1 ft

Figure 5.8: Schematic of auto correlogram with lag distance 2 *ft*

Figure 5.9: Schematic of auto correlogram with lag distance 5 *ft*

Figure 5.6 denotes the porosity of the rock that has been measured on a regular grid along the core. Lag 0 is the property at the location x_i . The location x is a non-negative and always increasing with i and x takes values 1, 2, ...5, and so on. Fig 5.7, 5.8, 5.9 are the schematic indication of the porosity along a line $x_{i+1}-x_i$

Figure 5.10: Porosity Probability Plot

Figure 5.10 represents one-to-one relationship between the value of the porosity and the probability values less than or equal to $F(\text{porosity})$. In Fig 5.10 the vertical axis in terms of F , expressed as percent.

Figure 5.11: Auto Correlation Function of Porosity

The definition of autocorrelation is the correlation between variables of the same sample space taken at different locations. (In Figure 5.11 lag distance is taken from 1 to 13 ft). Auto correlation reduces the information content of data

[4]. Figure 5.11 represents the auto correlation function of porosity. A sample of 53 porosity values is considered to consists of 53 uncorrelated measurements. 53 data are uncorrelated when they come from a population with $Mean = 16.153$ and $Variance = 18.807$ Table 5.2 and Figure 5.1. However if the data have a first lag correlation $\rho = \rho(1) \neq 0$ [14]. Consequently, the variance of the estimate may increase.

Figure 5.12: Permeability Probability Plot

Figure 5.12 represents one-to-one relationship between the value of the permeability and the probability values less than or equal to $F(permeability)$.

Figure 5.13: Auto Correlation Function of Permeability

Figure 5.13 represents the auto correlation function of permeability. A sam-

ple of 53 permeability values is considered to consists of 53 uncorrelated measurements. 53 data are uncorrelated when they come from a population with $Mean = 2.81$ and $Variance = 3.91$ Table 5.3 and Figure 5.2.

Table 5.6: Probability Density Function values of Porosity

x	P(X=x)	x	P(X=x)
4.98	0.00	1.47	0.1354
8.18	0.00	0.73	0.3056
5.074	0.00	6.09	0.00
3.24	0.0021	1.33	0.1647
1.54	0.1219	1.67	0.0989
3.20	0.0024	2.37	0.0241
0.58	0.3372	2.45	0.0198
1.96	0.0584	2.53	0.0163
0.44	0.3621	6.63	0.0000
1.63	0.1057	6.52	0.0000
0.94	0.2565	5.31	0.0000
1.19	0.1965	3.78	0.0003
2.74	0.0093	2.97	0.0048
2.99	0.0046	5.05	0.0000
2.13	0.0413	5.03	0.0000
1.69	0.0957	1.52	0.1257
2.26	0.0310	1.61	0.1092
2.17	0.0379	1.50	0.1295
3.25	0.0020	2.41	0.0219
3.74	0.0004	0.07	0.3980
3.07	0.0036	0.09	0.3973
3.08	0.0035	2.68	0.0110
3.27	0.0019	0.03	0.3988
1.45	0.1394	0.10	0.3970
1.60	0.1109		

Table 5.6 indicates the PDF of porosity values.

Figure 5.14: Probability Distribution Function

5.1.1 Estimation of Hydrocarbon Volumes

Basic Data Requirements: The minimum data requirement for probabilistic reserves calculations involves the basic quantities: 1-Gross rock volume, 2-Net to gross rock thickness, 3-Porosity, 4-Hydrocarbon saturation, 5-Recovery factor.

Area of Field A: In Field A, PDF for area is assumed to be triangular. Area is assumed as $1.092.656 \text{ m}^2$, which is considered as most likely area. As an assumption, minimum area is 25% less value of most likely area and maximum area is 25% more than the most likely area. Minimum area, most likely area and maximum area values are given in Table 5.7.

Net Thickness: As an assumption, net thickness is distributed triangularly. Net thickness values are assumed referring to well data which is given in Southeastern Region. Net thickness values are given in Table 5.7.

Saturation: Average oil saturation and standard deviation of oil saturation values are supplied by TPAO. Average oil saturation and standard deviation of oil values are given in Table 5.7.

Porosity: It is seen in Figure 5.14, PDF for porosity is normal. Average porosity and standard deviation of porosity values are given in Table 5.7.

Recovery Factor(RF): As an assumption, PDF for RF is triangular. RF is assumed to range between 15 – 35% which is accepted an average value for limestone reservoirs. The Field A is not strictly limestone reservoir, but due to the high porosity and low permeability distribution in the Field A the reservoir is considered as limestone reservoir. Thus, minimum value is 0.15, most likely value is 0.25 and maximum RF value is 0.35. RF values are given in Table 5.7.

Table 5.7: PDF of Field A

<i>Minimum area, m²</i>	819.492
<i>Most likely area, m²</i>	1.092.656
<i>Maximum area, m²</i>	1.365.820
<i>Minimum thickness, m</i>	1.8
<i>Most likely thickness, m</i>	26.8
<i>Maximum thickness, m</i>	45.8
<i>Average porosity, %</i>	16.15
<i>Standard deviation of porosity</i>	4.33
<i>Average oil saturation</i>	0.735
<i>Standard deviation of oil saturation</i>	0.119
<i>Minimum RF</i>	0.15
<i>Most likely RF</i>	0.25
<i>Maximum Rf</i>	0.35

The volumetric method for estimating oil in place is based upon core and log data. The following equation is used to calculate OOIP values and when this value multiplied by RF, ROIP values are calculated.

$$OOIP = A * h_p * \phi * (1 - S_{wc}) \quad (5.2)$$

Table 5.8: Statistical results of OOIP and ROIP

<i>number of iterations</i>	100	100	1000	1000
	OOIP	ROIP	OOIP	ROIP
<i>%10</i>	1.969.775	502.532	2.023.274	523.491
<i>%50</i>	1.186.397	277.255	1.174.230	286.559
<i>%90</i>	543.342	128.631	562.343	137.811
<i>standard deviation</i>	541.661	144.200	557.683	153.367
<i>variance</i>	$2.9 * 10^{11}$	$2.08 * 10^{10}$	$3.11 * 10^{11}$	$2.35 * 10^{10}$
<i>skewness</i>	0.368	0.569	0.543	0.829
<i>kurtosis</i>	2.497	2.658	2.958	3.725

Table 5.8 indicates the statistical results of OOIP and ROIP after 100, 1000 iterations. Table 5.8 indicates minimum, most likely and maximum values of OOIP and ROIP at different iterations. Standard deviations, skewness and kurtosis are also given. Skewness tests if the shape of a sample distribution is similar to that of a normal distribution. When the skewness value is within -0.5 and $+0.5$, then the shape distribution is similar to normal distribution. But when it is greater than $+0.5$, indicates a long tail to the right. Kurtosis also examines the shape of the distribution. A distribution is normal when its kurtosis is 3 [4].

Table 5.9: Statistical results of OOIP and ROIP

<i>number of iterations</i>	10000	1000	20000	20000
	OOIP	ROIP	OOIP	ROIP
<i>%10</i>	2.024.266	521.398	2.022.051	516.872
<i>%50</i>	1.176.434	289.350	1.178.850	288.628
<i>%90</i>	557.938	133.529	562.450	133.867
<i>standard deviation</i>	563.047	152.142	560.273	150.283
<i>variance</i>	$3.17 * 10^{11}$	$2.31 * 10^{15}$	$3.13 * 10^{11}$	$2.25 * 10^{10}$
<i>skewness</i>	0.567	0.769	0.543	0.750
<i>kurtosis</i>	2.957	3.444	2.904	3.45

Table 5.9 indicates the statistical results of OOIP and ROIP after 10000, 20000 iterations.

Figure 5.15: Number of Iteration versus OOIP

Figure 5.15 indicates the Original Oil in Place vs iteration number. It is obviously seen from the Figure 5.15 when the iteration number increases the OOIP values leads to normal distribution.

Figure 5.16: Number of Iteration versus ROIP

Figure 5.16 indicates the Recovery Oil in Place versus iteration number.

5.2 DISCUSSION

It is seen that the basics of statistics, how geology and statistics might be used to complement each other in analysis. It is widely recognized that predictions of hydrocarbon recovery are uncertain. In the area of reservoir engineering, the sources of uncertainty have three major causes:

1. The model, because it is an imperfect representation of reality,
2. Geologic parameters, because of a limited sampling in space and/or time,
and
3. Measurement errors in the experiments performed to determine inputs.

Thus, a statistical treatment that recognizes both the lack of knowledge and the uncertainty of the parameters involved in the forecast of reservoir performance is desirable.

CHAPTER 6

CONCLUSION

The aim of this research is to prove that in the absence of adequate data, probabilistic methods lead to reasonable and convenient results in reserve estimation. The field in Southeastern is a marginal field with three wells and oil reserves have limited economic feasibility. The data available were used to define the probability distribution function for each input variable. Due to inadequate data, area, productive thickness and recovery factor were assumed to have triangular distribution. However, porosity was normally distributed.

This study discusses and finally, concludes the following;

1. Reserves calculated with probabilistic methods was not compared directly with those calculated with deterministic methods.
2. As expected, with small number of iterations the normal distribution was not obtained.
3. The skewness and kurtosis give the shape of the OOIP and ROIP are both

normal distributed. This is an expected result, in Field A the most effective parameter is porosity and it is normally distributed.

4. Dependency factor between two variables (porosity-saturation) scatters the output distribution, destroys the symmetrical distribution of the output.

The present study could only reveal the relations in physical properties of rock. This can be taken one step further by inspecting the any potential relation between well data and seismic data.

REFERENCES

- [1] Almeida, A.S. and Frykman, P. *Geostatistical modelling of chalk reservoir properties in the Dan Field, Danish North Sea,* *Stochastic Modeling and Geostatistics; Principles, Methods, and Case Studies*, J.M. Yarus and R.L. Chambers (eds.), American Association of Petroleum Geologists, Tulsa (1995) AAPG Computer Applications in Geology, 3, 273-286.
- [2] Arnold, L., *Stochastic Differential Equations: Theory and Application*, John Wiley and Sons, New York City (1974).
- [3] Berteig, V. et al., *Predictions of Hydrocarbon PV with Uncertainties*, paper SPE 18325 Presented at the 1988 SPE Annual Technical Conference and Exhibition, Houston, October 2-5.
- [4] Box, G.E.P., Hunter, W.G., and Hunter, J.S., *Statistics for Experimenters*, John Wiley and Sons, New York, (1978).
- [5] Chapman, R. E., *Petroleum Geology*, Elsevier Publisher, Amsterdam, New York, (1983).
- [6] Chewaroungroaj, J., Varela, J.O. and Lake, W.L., *An Evaluation of Procedures to Estimate Uncertainty in Hydrocarbon Recovery Predictions* paper SPE 59449 presented at 2000 Asia Pacific Conference in Yokohama, Japan, 25-26 April 2000.
- [7] Corbett, P.W.M., Jensen, J.L., and Sorbie, K.S., *A review of up scaling and cross-scaling issues in core and log data interpretation and prediction*, Core-Log Integration, P.K. Harvey and M.A. Lovell (eds.), Geological Society, London (1998) Geological Society Special Publications, 136, 9-16.
- [8] Da Costa e Silva, A.J., *A New Approach to the Characterization of Reservoir Heterogeneity Based on the Geo mathematical Model and Kriging Technique*, paper SPE 14275 presented at the 1988 SPE Annual Technical Conference and Exhibition, Las Vegas, Sept. 22-25.
- [9] *Determination of Oil and Gas Reserves*, Monograph, Petroleum Society of the Canadian Inst. of Mining, Metallurgy and Petroleum, Calgary (1994).
- [10] Deutsch, C.V. and Frykman, P., *Geostatistical Scaling Laws Applied to Core and Log Data*, paper SPE 56822 presented at 1999 SPE Annual Technical Conference and Exhibition held in Houston, Texas, 3-6 October 1999.

- [11] Grace, J.D., Caldwell, R.H. and Heather, D.I., *Comparative Reserves Definitions: USA, Europe and the Former Soviet Union*, SPE 25828, proceeding of the HEES Symposium, March 1993, 97-108.
- [12] Geffroy, F., Bleines, C., Rambert, F., Touffait, Y. *ISATIS Software Manual* 3rd Edition, Geovariances and Ecole des Mines de Paris, 2001
- [13] Journal, A. G., *Geostatistics for Reservoir Characterization*, paper SPE 20750 presented at the 65th Annual Technical Conference and Exhibition of the SPE, New Orleans, LA, September 23-26, 1990
- [14] Jensen, J.L., Lake, W.L., Patrick, W.M., Corbett and Goggin, J.M., *Statistics for Petroleum Engineers and Geoscientists*, 2nd Edition, Elsevier, 2000
- [15] Jonkman, R.M., Bos, C.F.M., Breunese, J.N., Morgan, D.T.K., Spencer, J.A., and Sandena, E., *Best Practices and Methods in Hydrocarbon Resource Estimation, Production and Emissions Forecasting, Uncertainty Evaluation and Decision Making*, paper SPE 65144 presented at the 2000 SPE European Petroleum Conference held in Paris, FR October 24-25.
- [16] Kristensen, L., et al., *A multidisciplinary approach to reservoir subdivision of the Maastrichtian chalk in the Dan Field, Danish North Sea*, American Association of Petroleum Geologists Bulletin (1995), 1650-1660.
- [17] Lestang A.P., Cosentino L., Cabrera J., Jimenez T., Bellorin O., *Geologically Oriented Geostatistics: an Integrated Tool for Reservoir Studies*, paper SPE 74371 presented at the 2002 SPE International Petroleum Conference held in Villahermosa, Mexico February 10-12
- [18] Martinez, R. D., Cornish, B. E., and Seymour, R. H., *An Integrated Approach for Reservoir Description Using Seismic, Borehole, and Geologic Data*, paper SPE 19581 presented at the 1989 SPJ3 Annual Technical Conference and Exhibition, San Antonio, October 8-11.
- [19] Mata, T., Rojas, L. and Banerjee, S., *Probabilistic Reserves Estimation of Mara West Field, Maracaibo Basin, Venezuela: Case Study*, paper SPE 38805 presented at the 1997 ATCE, San Antonio, TX, October 5-8.
- [20] McKelvey, V.E., *Mineral Resource Estimates and Public Policy*, American Scientist (January-February 1972) 60, 32. Halim D.,
- [21] Patricelli, J.A. and McMichael, C.L., *An Integrated Deterministic or Probabilistic Approach to Reserve Estimations*, JPT (January 1995), 49.
- [22] Ripley, B. D., *Spatial Statistics*, John Wiley and Sons, New York City (1981)
- [23] Ross, J.G., *The Philosophy of Reserve Estimation*, paper SPE 37960 presented at the 1997 SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, March 16-18 .
- [24] Ross, J.G., *Booking Reserves*, JPT (November 1998), 75.

- [25] Sandsdalen, C. et al, *Applied uncertainty analysis using stochastic modelling*, SPE 35533 (1996).
- [26] Simlote, V.N. and Nikolayenko, S.A., *Dealing with Uncertainty: Geostatistics and the New Era in Reservoir Studies*, paper SPE 39546 presented at the SPE India Oil and Gas Conference and Exhibition, New Delhi, India (1998) 383-397.
- [27] Smith, P.J. and Buckee, J.W., *Calculating in Place and Recoverable Hydrocarbons: A Comparison of Alternative Methods*, paper SPE 13776 presented at the 1985 SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, March 14-15.
- [28] *SPE/WPC Reserves Definitions Approved* , JPT (May 1997) 527.
- [29] Yarus, J. and Chambers, R., *Stochastic Modelling and Geostatistics: Principles, Methods, and Case Studies*, American Association of Petroleum Geologists Computer Applications in Geology

APPENDIX A

BASIC CONCEPTS OF STATISTICS

A.1 INTRODUCTION

The modern study of reservoir characterization began in the recent years, driven by the realization that deficiencies in advanced oil recovery techniques frequently had their origin in an adequate reservoir description. The litany of problems became commonplace; wells drilled between existing wells did not have the interpolated characteristics, chemical injected in wells appeared in unforeseen locations displacing agents broke through to producing wells too early, and above all oil recovery was disappointing. Each of these problems can be traced to a lack of understanding of the nature of the distribution of properties of between wells.

A.2 BASIC CONCEPTS

Statistics is a powerful tool that can supply estimates and their uncertainties for a large number of reservoir properties.

Experiment, $\mathfrak{S}$: The operation of establishing certain conditions that may produce one of several possible outcomes or results.

Sample space, Ω : The collection of all possible outcomes of $\mathfrak{S}$.

Event E : A collection of some of the possible outcomes of $\mathfrak{S}$. E is a subset of Ω .

These are the general definitions. An experiment $\mathfrak{S}$ could be the act of taking a rock sample from a reservoir and measuring its porosity (ϕ). The sample space Ω is the collection of all properties of such specimens in the reservoir. That the event E may be the value of specific porosity (ϕ) that is actually observed in a core plug.

Occurrence of E : An event E has occurred if the outcome of an experiment belongs to E .

An experiment always conducted to determine if the a particular result is obtained. An individual outcome (result) is an elementary or simple event; no outcome is also an event and called null set.

A.2.1 COMBINATION OF EVENTS

E_1 and E_2 are outcomes of the experiment $\mathfrak{S}$.

Union of Two Events: $E_1 \cup E_2$ is the set of outcomes that belong to either E_1 or E_2 .

Intersection of Two Events: $E_1 \cap E_2$ is the set of outcomes that belong to both E_1 and E_2 .

Both union and intersection are themselves events. Intersection and union lead to one highly useful concept, mutual exclusivity and exhaustive events.

Mutually Exclusive Events: Events E_1 and E_2 are mutually exclusive if both E_1 and E_2 can not occur simultaneously; that is $E_1 \cap E_2 = \emptyset$.

Exhaustive Sequence of Events: The sequence E_1, E_2, E_n accounts for all possible outcomes such that $E_1 \cup E_2 \cup E_n = \Omega$.

Complementary Event (E^c): The subset of Ω that contains all the outcomes that do not belong to E .

A.2.2 PROBABILITY

Let E be an event associated with an experiment $\mathfrak{S}$. Perform $\mathfrak{S}$ n times and let m denote the number of times E occurs in these n trials. The ratio m/n is called the relative frequency of occurrence E . If, as n becomes large, m/n approaches a limiting value, then

$$p = \lim_{n \rightarrow \infty} \frac{m}{n} \quad (\text{A.1})$$

The quantity p is called the probability of event E , or $\text{Prob}(E) = p$. The relative frequency of occurrence of the event E is a statistic of the experiment $\mathfrak{S}$. The ratio m/n is obtained by conducting n experiments.

A.2.3 PROBABILITY LAWS

The definition of probability lead its axioms which provide a bridge between occurrences and mathematics.

Axiom 1: For every event E , $0 \leq \text{Prob}(E) \leq 1$

Axiom 2: For the special case $E = \Omega$, $\text{Prob}(E) = 1$

Axiom 1 says that the probability of any event must be between 0 (no chance

of happening) and 1 (certainty of happening). Axiom 2 says that some event must occur from the experiment.

Axiom 3 (Additional Law): Let E_1 and E_2 be two events then ;

$$Prob(E_1 \cup E_2) = Prob(E_1) + Prob(E_2) - Prob(E_1 \cap E_2)$$

Axiom 3 is sometimes known as the fundamental law of addition. It simply says that the probability of E_1 and E_2 is the sum of the individual probabilities less the probability of E_1 and E_2 .

Axiom 3' : If E_1 and E_2 are two mutually exclusive events (*i.e.*, $E_1 \cap E_2 = \emptyset$) then

$$Prob(E_1 \cup E_2) = Prob(E_1) + Prob(E_2)$$

Axiom 3'' : If $E_1, E_2, \dots, E_i, \dots$ is a mutually exclusive sequence of events, then

$$Prob(E_1 \cup E_2 \cup \dots E_\infty) = Prob(E_1) + Prob(E_2) \dots Prob(E_\infty) = \sum_{i=1}^{\infty} Prob(E_i)$$

Combining Axiom 1 and 3 gives *Boole's inequality*;

$$Prob(E_1 \cup E_2) \leq Prob(E_1) + Prob(E_2)$$

With a little algebra, this expression becomes a *Bonferroni inequality*

$1 - Prob(E_1 \cup E_2) \geq 1 - Prob(E_1) - Prob(E_2)$ where the left hand side of the equation is the probability of neither event occurring and is also useful for calculating the confidence level of joint hypotheses. The null set may be written as the complement set of the sample space, since $Prob(\emptyset) = 1 - Prob(\Omega) = 0$

Conditional Probabilities: For two events E_1 and E_2 , the conditional probability of E_2 given that E_1 has occurred is defined to be

$$Prob(E_2 | E_1) = \frac{Prob(E_2 \cap E_1)}{Prob(E_1)} \text{ provided that } E_1 > 0$$

Usually the conditional probability definition is used to estimate

$Prob(E_1 \cap E_2)$ so that the definition is sometimes referred to as the fundamental law of multiplication.

Independence: If $Prob(E_1 \cap E_2) = Prob(E_1)Prob(E_2)$ then two events E_1 and E_2 are called *independent events*. This result known as the multiplication law for independent events. When this equation is true, the conditional probability of E_1 is exactly the same as the unconditional probability of E_1 . That is E_2 provides no information about E_1 .

$$Prob(E_2 | E_1) = \frac{Prob(E_2 \cap E_1)}{Prob(E_1)} = Prob(E_2)$$

A.2.4 UNIVARIATE DISTRIBUTION

A random variable $X(\bullet)$ is a mapping from the sample space Ω onto the real line $\mathbb{R}^1$ such that to each element $\omega \in \Omega$ there corresponds a unique real number $X(\omega)$ with all of the associated probabilistic features of ω in Ω .

The idea this should convey is one of uncertainty, not about the rule X but about the outcome of the experiment. The random variable incorporates the notion that

1. certain values will occur more frequently than others
2. the values may be ordered from smallest to largest
3. although it may take any value in a given range, each value is associated with its frequency of occurrence through a distribution function.

The value $X(\omega)$ associated with each element (ω) may not necessarily have any relationship to ω 's intrinsic value. To create a relationship between each element, the data types should be defined. The measurement scales classified into

Nominal Scale: The measurement consists of a classification observations into mutually exclusive categories of equal rank. The classification of fossils as to type is nominal scale.

Ordinal Scale: Observations can be ranked in a hierarchy of steps. The steps between successive states are not equal. As with the nominal scale, a quantitative analysis of ordinal measurements is restricted primarily to counting observations in the various states.

Interval Scale: The interval of a successive intervals is a constant. The interval scale has an arbitrary zero point or point where the magnitude is non-existing.

Ratio Scale: The ordering on a ratio scale has a well-defined zero point where there is non of the property and the variable can always be meaningfully added or multiplied. Ratio scales are the highest form of measurement. All types of mathematical and statistical operations may be performed with them.

The definitions that have been introduced do not rule out dependencies between variables in different sample spaces or between variables within the same sample space.

A.3 THE CUMULATIVE DISTRIBUTION FUNCTION

The random variable X and a real number x , which can be assumed any value between $-\infty$ and ∞ and the probability for the event $X < x$. This probability is a function of x and is called the *cumulative distribution function* (CDF) of X .

$$F(x) = P(X \leq x) \quad (\text{A.2})$$

$F(x)$ is the probability of finding a value of a random variable X that is less than or equal to x . The argument of F is x , the deterministic bounding value, not X the random variable. The CDF uniquely defines all of the probabilistic properties of a random variable.

CDF's have the following properties;

1. $0 \leq F(x) \leq 1$ for all x since $F(x)$ is a probability.
- 2.

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad (\text{A.3})$$

and

$$\lim_{x \rightarrow \infty} F(x) = 1 \quad (\text{A.4})$$

3. $F(x)$ is a non-decreasing function: $F(x+h) \geq F(x)$ for any $h \geq 0$ and all x .
4. $F(x)$ is a continuous function from the right for all x :

$$\lim_{h \rightarrow 0^+} F(x+h) = F(x) \quad (\text{A.5})$$

where $h \rightarrow 0^+$ means h approaches 0 through positive values.

An important use of CDF is that it can be used to find out how often events within a given range of values will occur.

A.3.1 Discrete CDF's

Discrete CDF's are defined by a function $F(x)$ with a set of jumps of magnitude p_i at points x_i for $i = 1, 2, 3 \dots I$ such that

$$Prob(X = x_i) = F(x_i) - \lim_{x \rightarrow x_i^-} F(x) \quad (\text{A.6})$$

and

$$\sum_{i=1}^I p_i = 1 \quad (\text{A.7})$$

Thus a random variable cannot fall within the flat segments of the CDF. This means that X can take on values only at the jump points. In this case, X is a purely discrete or categorical random variable. Nominal or ordinal data have this sort of CDF. If the difference between adjacent x_i 's (*i.e.*, $x_{i+1} - x_i$) is the same for all i , this difference is called the *class size*.

A.3.2 Continuous CDF's

Continuous CDF $F(x)$ will be defined

$$F(x) = \int_{-\infty}^x f(u) du \quad (\text{A.8})$$

with

$$F(+\infty) = \int_{-\infty}^{+\infty} f(u) du = 1 \quad (\text{A.9})$$

The random variable X pertaining to this CDF is a purely continuous random variable. The function $f(u)$ in the definition of $F(x)$ is the *probability distribution function (PDF)*.

A.3.3 Empirical CDF's

If the CDF of a variable can not be known then sample values of the variable can be obtained and an approximate CDF can be produced from those data. Such a CDF is called an *empirical CDF*.

Empirical CDF's are usually produced for non-nominal variables where there is some natural ordering to the data. If there is no natural size to the variable, the empirical CDF might change shape, depending upon the number assigned (i.e., the random variable) to each category.

For variables with ordering, an empirical CDF is produced as follows;

1. Order the data so that $X_1 \leq X_2 \leq X_3 \leq \dots \leq X_I$ where I is the number of data.
2. Assign a probability p_i , to the event $ProbX \leq X_i$. Here some assumptions about the assignment should be made. If X is a continuous variable or is discrete and the uppermost categories may not have been sampled, $p_i = \frac{(i-1/2)}{I}$ is usually adequate. The assumption is made here, without any evidence to the contrary that each sample value has an equiprobable chance of occurring. If X is discrete and all categories have been sampled, $p_i = i/I$.
3. Plot X_i versus p_i and, depending on whether the random variable is continuous or discrete, either connect the lines with stair-steps or continuous curve.

A.4 Order Statistics

Some values of x for the CDF $F(x)$ have special names. $x_{0.50}$ is the value where $x = F^{-1}(0.50) \equiv x_{0.50}$ and is called *median*. $x_{0.25}$ and $x_{0.75}$ are the first and third quartiles, respectively of the CDF $F(x)$. The difference between $x_{0.75}$ and $x_{0.25}$ is sometimes used as a measure of the variability of X and is called the interquartile range.

A.5 PROBABILITY DISTRIBUTION FUNCTIONS

Probability of distribution functions takes a local view with respect to cumulative distribution functions, and describes how the probability of occurrence of X changes with x . As with CDF's there is a distinction between discrete and continuous properties.

A.5.1 Discrete PDF's

$$Prob(x_i - h < X \leq x_i + h) = F(x_i + h) - F(x_i - h) = p_i \quad (\text{A.10})$$

If $h \rightarrow 0$ then $Prob(X = x_i) = p_i$. Thus the jump at the end of each interval represents the probability that X has the value x . The set of numbers $(p_1, p_2, p_3, \dots)$ plotted against $(x_1, x_2, x_3, \dots)$ is called the discrete PDF of the random variable X .

A.5.2 Continuous PDF's

A small interval $(x, x + \delta x)$ was considered from the x axis of a purely continuous random number X with CDF $F(x)$. Then

$$Prob(x < X \leq x + \delta x) = F(x + \delta x) - F(x) = \int_x^{x+\delta x} f(u)du \quad (\text{A.11})$$

This forms the basis for interpreting $f(x)$ as a probability. $f(x)$ represents the frequency of occurrence of a value of x in the neighborhood of x .

The basic properties of continuous PDF's are

1. $f(x) > 0$ for all x since $F(x)$ is a non-decreasing.
- 2.

$$F(+\infty) = \int_{-\infty}^{+\infty} f(u)du = 1 \quad (\text{A.12})$$

3. For any a, b ($a \leq b$),

$$Prob(a < X \leq b) = \int_a^b f(u)du \quad (\text{A.13})$$

A rather curious consequence of the last property is that the probability that X takes a particular value x is zero. If $b \rightarrow a$ then

$$Prob(x = X) = \int_a^a f(u)du = 0$$

Both continuous and discrete PDF's have peaks and troughs. Those values of x where a peak occurs are called *modes*.

A.5.3 Empirical PDF's

Empirical PDF's can be produced for any type of variable. For variables with ordering, an empirical PDF is produced as follows;

1. If the variable is not nominal, order the data so that $X_1 \leq X_2 \leq X_3 \leq \dots \leq X_I$ where I is the number of data.
2. For continuous variables, divide the interval $X_I - X_1$ into convenient intervals. These intervals are called bins or classes. The bins are all of usually equal size, Δx so the height of each interval is proportional to the probability. A rule of thumb is $\Delta x \approx 5(X_I - X_1)/I$
3. The number of data in the i^{th} bin is counted as I_i

then $Prob(x_i \leq X < x_i + \Delta x) = p_i = I_i/I$ for $i = 1$ to $I - 1$
4. p_i versus $x_i \leq X < x_i + \Delta x$ for $I = 1$ to $I - 1$ is plotted.

There are several reasons why we want to know the distribution of a reservoir property therefore CDF's and PDF's are important. They are;

1. *Modeling:* Knowing the PDF/CDF, the models how the property varies within the reservoirs can be produced.
2. *Estimation:* Estimating certain parameters based on their CDF/PDF. Better use of available data can be made if the population PDF is known.
3. *Diagnosis:* The test data can be determined whether they are good or bad by comparing the sample PDF with other sample PDF's or some theoretical reference PDF. PDF's have been used to assess the accuracy of core measurements [14]

A.6 SAMPLING AND STATISTICS

The statistical view of the reservoir is entirely from probability distributions of the random variable X representing some property. This property can be permeability, porosity, or grain size, exists as a random variable in a physical region. Each point in the region has associated with it several PDF's: one for each random variable of interest.

By measuring samples of non-zero volume, characterization of the reservoir is based on averages of the random variable, not the variable itself. Estimating a parameter at a point means estimating the most likely value of the parameter's average. In the range of scales of the geological variability present in the reservoir, some scales may be contained within the measurement, and thus not be recognized in the result.

This point of view has two implications. The first is that all physical measurements are now actually averages and it is possible to have measurements over different volumes of material. Trying to reconcile measurements made on two different volumes can be very difficult, especially since the character of the PDF's and CDF's will change for different volume scales. This is especially true when comparing the statistics of porosity averages derived from core/log data with porosity estimates derived from seismic attributes, even when they are samples over common depth intervals. The second implication lies in regarding the original variable as being random. Many geological process -sedimentation or diagenesis- are very well known and quite deterministic; thus, it is counter-intuitive to think that the result of these processes- the variable under consideration-is random.

APPENDIX B

FEATURES OF SINGLE-VARIABLE DISTRIBUTIONS

B.1 NON-CENTERED MOMENTS

With the probability distribution function $f(x)$ known for a continuous random variable X , the r^{th} moment can be defined as

$$\mu_r = \int_{-\infty}^{+\infty} x^r f(x) dx \quad (\text{B.1})$$

where r is a non-negative integer.

B.1.1 The Expectation

The expected value or the expectation of X is referred to first moment where $r=1$.

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx \quad (\text{B.2})$$

$E(X)$ is the *mean* for a continuous PDF.

A discrete random variable X can take one of a series of values: $(X_1, X_2, X_3, \dots)$ with corresponding probabilities $(p_1, p_2, p_3, \dots)$. That is $p_i = \text{Prob}(X = X_i)$. The expected value of X is defined to be

$$E(X) = \sum_{i=1}^I X_i p_i \quad (\text{B.3})$$

So $E(X)$ is the sum of the possible values of X weighted by their respective probabilities of occurrence. It is related to, but not necessarily the same as the arithmetic average. The arithmetic average is, $\bar{X}$, is computed from M samples $(X_1, X_2, \dots, X_m, \dots, X_M)$ taken from the parent population, which has values $(X_1, X_2, \dots, X_I)$

$$\bar{X} = \frac{1}{M} \sum_{m=1}^M X_m \quad (\text{B.4})$$

Comparing the equations for $\bar{X}$ and $E(X)$, it can be seen that the arithmetic averages appears to equate p_i with $\frac{1}{M}$. For discrete variables, $E(X)$ need not be one of the values X_i . Since there are two definitions for the expectation, corresponding to continuous and discrete PDF's.

B.1.2 Properties of the Expectation

The definition of expectation lead its properties of a sum or the integral

Expectation of a constant c : $E(c)=c$

Expectation of a constant c plus(minus) a random variable: $E(X \pm c)=E(X) \pm c$

An immediate consequence of this property is that the expectation of a random variable with zero mean is zero.

Expectation of a constant c times a random variable: $E(cX) = cE(X)$

Expectation of the sum(difference)of two random variables: $E(X \pm Y) = E(X) \pm E(Y)$ When the expectation deals with more than one random variable, the definition involves the joint probability distribution function for X and Y , $f(X, Y)$.

$$\begin{aligned}
 E(X + Y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + y) f(x, y) dx dy \\
 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x, y) dx dy + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f(x, y) dx dy \\
 &= \int_{-\infty}^{+\infty} x f(x) dx + \int_{-\infty}^{+\infty} y f(y) dy = E(X) + E(Y)
 \end{aligned} \tag{B.5}$$

where $\int_{-\infty}^{+\infty} f(x, y) dy = f(x)$ and $\int_{-\infty}^{+\infty} f(x, y) dx = f(y)$ are the marginal PDF's for Y and X respectively.

Expectation of a the product of two independent random variables:

$$E(XY) = E(X)E(Y)$$

To satisfy this axiom the joint PDF $f(x, y)$ must be written as $f(x, y) = f(x)g(y)$.

$$\begin{aligned}
 E(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f(x) g(y) dx dy \\
 &= \int_{-\infty}^{+\infty} x f(x) dx \int_{-\infty}^{+\infty} y g(y) dy = E(X)E(Y)
 \end{aligned} \tag{B.6}$$

B.2 CENTERED MOMENTS

The r^{th} moment of a random variable is

$$\mu_r = E[(X - E(X))^r] = \int_{-\infty}^{+\infty} (X - E(X))^r f(x) dx \quad (\text{B.7})$$

B.2.1 Variance

One of the central concepts in all statistics is the second ($r=2$) centered moment, known as *variance* and denoted by $Var(X)$.

$$Var(X) = \int_{-\infty}^{+\infty} (X - E(X))^2 f(x) dx \quad (\text{B.8})$$

$Var(X)$ can be written as the difference between the expectation of X^2 and the square of the expectation of X , $E(X)^2$:

$$Var(X) = E[(X - E(X))^2] = E(X^2) - E(X)^2 \quad (\text{B.9})$$

The discrete version of $Var(X)$ for I values is

$$Var(X) = \sum_{i=1}^I (X_i - E(X))^2 p_i \quad (\text{B.10})$$

The estimation of variance from I samples $X_1, X_2, \dots, X_I$

$$s^2 = \frac{1}{I} \sum_{i=1}^I (X_i - E(X))^2 \quad (\text{B.11})$$

which requires that the mean of X be known.

The variance is a measure of the dispersion or the spread of a random variable about its mean.

B.2.2 Properties of the Variance

The most important property of variance is that it is non-negative.

Intrinsic property: $Var(X) \geq 0$

The reason this is so follows directly from the basic definition wherein only squares of differences are used inside an integral and $f(x) \geq 0$.

Except intrinsic property, many of the variance properties follow directly from those of the expectation operator.

Variance of a constant c: $Var(c)=0$

Variance of a random variable plus(minus) a constant c:

$$Var(X \pm c) = Var(X)$$

These properties are consistent with the idea of the variance being a measure of spread, independent of translations. There is of course, no spread in a constant.

Variance of a random variable times a constant c:

$$\begin{aligned} Var(cX) &= E[(cX)^2] - E(cX)^2 \\ &= c^2 E(X^2) - c^2 E(X)^2 \\ &= c^2 Var(X) \end{aligned}$$

This property is analogous for the expectation, this is not intuitive. However, since the constant can be negative, it is entirely consistent with the idea of a non-negative variance, since $Var(cX) \geq 0$.

Variance of the sum(difference)of two independent random variables:

$$\begin{aligned} Var(X \pm Y) &= E[(X \pm Y)^2] - E(X \pm Y)^2 \\ &= E[X^2 \pm 2XY + Y^2] - E(X)^2 \pm 2E(X)E(Y) - E(Y)^2 \end{aligned}$$

$$= E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2$$

$$= Var(X) + Var(Y)$$

$$\text{since } E(XY) = E(X)E(Y)$$

The last two properties can be combined into

$$Var(\sum_i a_i X_i) = \sum_i a_i^2 Var(X_i) \quad (\text{B.12})$$

But this property is valid for independent X_i only.

In addition to these there are a few other related functions of these moments.

B.2.3 Standard Deviation

The standard deviation is the positive square root of the variance

$$SD = +\sqrt{Var(X)}$$

While $Var(X)$ has the units of X^2 , the SD has units of X , making it more comparable to the mean.

B.3 COMMON PDF'S

Probability of distribution functions takes a local view with respect to cumulative distribution functions, and describes how the probability of occurrence of X changes with x . In this section a few common PDF's and their properties will be investigated.

B.3.1 The Binomial Distribution

An experiment with only two possible outcomes: E_1 =a "success" with $Prob(E_1) = p$ and E_2 =a "failure" with $Prob(E_2) = 1 - p$. The sample space can therefore be

expressed as $E = E_1 + E_2$. E_1 and E_2 are obviously mutually exclusive in a single experiment, but several repetitions of the experiment would produce combinations of success and failures. The probability that the first k trials result in E_1 and all of the rest in E_2 is $p^k q^{n-k}$. Using the rules of combinatorics, the event "outcome E_1 s times in I trials" with each combination multiplied by its appropriate probability

$$Prob(R = s) = \frac{I!}{s!(I-s)!} p^s (1-p)^{I-s} \quad (\text{B.13})$$

The mean value of R for I trials is

$$E(R) = \sum_{i=0}^I i({}_I C_i) p^i (1-p)^{I-i} = Ip \quad (\text{B.14})$$

Mixtures of random variables cause the resulting PDF to be a combination of the components PDF's. Such PDF's are called heterogenous or compound distributions.

B.3.2 The Central Limit Theorem

According to a fundamental result of applied probability known as *Central Limit Theorem* (CLT), a random variable X , generated by adding together a large number of independent random variables, will usually have an approximately normal distribution irrespective of the PDF's of the component variables. This remarkable theorem says that a normal distribution will result from the binomial distribution as I approaches infinity regardless of the value of p .

B.3.3 The Normal (Gaussian) Distribution

By far the most important distribution for data analysis is the normal distribution. If a random variable X has a normal distribution, its PDF is uniquely determined by two parameters (μ and σ) according to

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right], -\infty < X < +\infty \quad (\text{B.15})$$

It is obviously seen that $E(X) = \mu$ and $Var(X) = \sigma^2$. The normal distribution is represented as $N(\text{mean}, \text{variance})$.

The normal CDF is

$$F(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x \exp\left[-\frac{(\eta - \mu)^2}{2\sigma^2}\right] d\eta \quad (\text{B.16})$$

The integral function above is called the probability integral. Both the normal distribution and the probability integral can be made parameter free by redefining the variable as $Z = (X - \mu)/\sigma$, known as the *standardized normal variate*.

$$F(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^z \exp\left[-\frac{(\eta)^2}{2}\right] d\eta \quad (\text{B.17})$$

The normal distribution occurs anywhere there are observations of processes that depend on additive random errors. The relevant function is the error function defined as

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp[-(\eta)^2] d\eta \quad (\text{B.18})$$

B.3.4 The Log-Normal Distribution

The log-normal PDF can be derived from the basic definition of the PDF and the standard log-normal variate:

$$z = \frac{\ln x - \mu_{\ln x}}{\sigma_{\ln x}}$$

Substituting this into Equation B.16 gives the following for the CDF:

$$F(x, \mu_{\ln x}, \sigma_{\ln x}) = \frac{1}{x\sigma_{\ln x}\sqrt{2\pi}} \int_{\eta=0}^{\eta=x} \exp\left[-\frac{(\eta - \mu_{\ln x})^2}{2\sigma_{\ln x}^2}\right] d\eta \quad (\text{B.19})$$

and using $y = \ln x$ in the transformation, we have for the PDF;

$$F(x, \mu_{\ln x}, \sigma_{\ln x}) = \frac{1}{x\sigma_{\ln x}\sqrt{2\pi}} \exp\left[-\frac{(x - \mu_{\ln x})^2}{2\sigma_{\ln x}^2}\right] \quad (\text{B.20})$$

The mean and the variance of the log-normal distribution are given by;

$$E(x) = \exp(\mu_{\ln x} + \sigma_{\ln x}^2/2)$$

and

$$Var(X) = \exp(2\mu_{\ln x} + \sigma_{\ln x}^2) \exp(\sigma_{\ln x}^2 - 1)$$

The coefficient of the variation is independent of $\mu_{\ln x}$:

$$C_v^2 = \exp(\sigma_{\ln x}^2) - 1$$

B.3.5 p-Normal Distribution

A more general power transformation of the form is defined [4]

$$y = \begin{cases} [(x+c)^p - 1]/p & p \neq 0 \\ \ln(x+c) & p = 0 \end{cases} \quad \text{which will render a wide range of sample spaces}$$

normal with the appropriate selection of p and c . It is obvious that $p=1$ is the normal distribution and $p=0$ is the log-normal distribution.

B.3.6 Uniform Distribution

This distribution is the simplest of all the commonly used CDF's and PDF's.

It consists of three straight lines. The equation of the CDF is

$$F(x; a, b) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq X < b \\ 1 & b \leq X \end{cases} \quad (\text{B.21})$$

and consequently, for the PDF is

$$F(x; a, b) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leq X < b \\ 0 & b \leq X \end{cases} \quad (\text{B.22})$$

The uniform distribution is a two parameter ($a \leq b$) distribution. The mean of this distribution is $E(X) = (a+b)/2$ and the variance is $Var(X) = (b-a)^2/12$. The popularity of the uniform distribution arises from its simplicity and also from its finite bounds, $X=a$ and $X=b$.

B.3.7 Triangular Distribution

The triangular distribution is the next step up in complexity from the uniform distribution since it is a three-parameter distribution($a < b < c$). The equation for the PDF is;

$$F(x; a, b, c) = \left(\frac{2}{c-a} \right) \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq X < b \\ \frac{c-x}{c-b} & b \leq X < c \\ 0 & c \leq X \end{cases} \quad (\text{B.23})$$

The equation for the CDF is;

$$F(x; a, b, c) = \begin{cases} 0 & x < a \\ \frac{(x-a)^2}{(c-a)(b-a)} & a \leq X < b \\ 1 - \frac{(c-x)^2}{(c-a)(c-b)} & b \leq X < c \\ 1 & c \leq X \end{cases} \quad (\text{B.24})$$

As for the uniform distribution, the triangular distribution is useful when there is a reason to confine the variable to a finite range. However, it frequently occurs that there is a most likely value within this range, but still no priori knowledge of the distribution. In these cases the triangular distribution is appropriate.

B.3.8 Exponential (Boltzmann) Distribution

This distribution is a one-parameter function with the following form for the CDF:

$$F(x; \lambda) = 1 - \exp(-x/\lambda) \quad (\text{B.25})$$

and for the PDF:

$$F(x; \lambda) = \lambda^{-1} \exp(-x/\lambda) \quad (\text{B.26})$$

both for $X > 0$ and $\lambda > 0$. The mean and variance are λ and λ^2 , respectively. Both fracture aperture and spacing may have exponential distributions [14].

B.4 ESTIMATORS AND THEIR ASSESSMENT

The estimation of oil and gas reserves is necessary for assessing the value of oil and gas properties, and for meeting the reporting and disclosure requirements imposed on oil and gas companies by various governmental and regulatory bodies. Geostatistics is a sophisticated technique of forecasting future recoverable volumes and production rates, and is an increasingly popular tool for reservoir management and optimization. Reservoir geostatistics is also progressively being used in the process of forecasting and estimating reserves. However, as with any reservoir engineering technique, certain precautions must be taken when relying on reservoir simulation as the means for estimating reserves. This thesis identifies some of the unique issues encountered in the reserves evaluation process when reservoir simulation is utilized, and offers strategies for addressing these issues.

B.5 The Elements of Estimation

The population represents the entire body from which we could draw examples and make measurements. The properties of the sample is assessed and this assessment procedure is called estimation.

Estimation involves four important elements:

1. The population quantity (the statistics) to be estimated, θ , from a set of data. θ can represent one or more parameters. If it represents more than one

parameter, it is considered to be a vector instead of scalar.

2. The estimator, W . This is the method by which the data values will be combined to produce the estimated or sample quantity.

3. The estimate, $\hat{\theta}$.

4. The confidence interval or standard error is $[Var(\hat{\theta})]^{1/2}$.

The number of parameters required depends on the form of the PDF and the subsequent application of the statistics. Similarly, the normal distribution has two parameters, the mean and variance, which may be estimated to characterize a random variable suspected of being normally distributed.

B.6 Desirable Properties of Estimators

Given an estimator W and a set of data $(X_1, X_2, ..X_I)$ taken from a population with parameter value Θ , W produces an $\hat{\Theta}$ based on the data: $\hat{\Theta} = W(X_1, X_2, ..X_I)$. Knowing this, there are several features that a good estimator will have: small bias, good efficiency, robustness, and consistency.

B.6.1 Estimator Bias

The PDF of an estimate $\hat{\theta}$ should be centered about the population parameter Θ . If this is not true, the estimator W will produce estimates that tend to over- or underestimate Θ , and the estimator is said to be biased. Bias is given by the expression;

$$b = E(\Theta) - \Theta \tag{B.27}$$

Bias is generally undesirable, but sometimes an estimator can be corrected for it. Bias has two principal sources, measurement resolution and sampling.

B.6.2 Estimator Efficiency

The second property of a good estimator is that the variability of its estimates should be as small as possible.

There are several methods to develop estimators. The methods of moments and of maximum likelihood are two examples. Maximum likelihood requires that the population PDF be known (or assumed). Combining the property of estimator bias and variability introduces an important property known as *consistency*. An estimator W is consistent if, as the number of data I , becomes large, $\hat{\Theta}$ approaches Θ (asymptotically unbiased) and the variance of $\hat{\Theta}$ approaches zero.

B.6.3 Estimator Robustness

Estimator robustness is the ability of the estimator to be relatively uninfluenced by errors in a small portion of data set. That errors can occur in the measurement and the recording of the data; a decimal point could get moved or a specimen could be improperly tested. If the unrepresentative values have a large influence on the estimate produced, then the estimator is not robust.

B.6.4 The Median

Substantial work has been done concerning robust estimators. The sample median ($\hat{X}_{0.50}$ is the middle value in a set of I data:

$$(\hat{X}_{0.50} = \begin{cases} X_{\frac{I+1}{2}} & \text{for } I \text{ odd} \\ 1/2[X_{\frac{I}{2}} + X_{\frac{I+2}{2}}] & \text{for } I \text{ even} \end{cases} \quad (\text{B.28})$$

where $X_1 \leq X_2 \leq X_3 \leq \dots \leq X_I$ is an ordered data set. If the data come from a normally distributed population with mean μ , $\hat{X}_{0.50}$ is an unbiased estimate of μ .