THE IMPACT OF DEPENDENCIES AND INTERACTIONS USING SYSTEMS APPROACH ON INVESTMENT DECISION MAKING IN AN ENVIRONMENT OF UNCERTAIN OUTCOMES

By

Mansoor Hamood Al-Harthy

THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF ADELAIDE AUSTRALIA

The University of Adelaide
Australian School of Petroleum

May 2006

To my father: Hamood

Who passed away as I started the PhD Program.
He demanded from me nothing except my best.
Father: This is my best

To my Dearest friend: Turki Al-Busaidy

Who passed away a few months after my father
during his second year of the PhD Program.
Turky: This is your PhD

To my Mother: Sheikha

Who never stops believing in me.
Mother: This is a small gift
&
The best is yet to come.

To my Wife and my Daughter

For their endless support and encouragement especially
for their patience when I was working on Saturdays and Sundays
and even sleeping in the department

ACKNOWLEDGMENTS

In the name of Allah, the most gracious, the most merciful. Looking at the whole PhD thesis, it is no longer merely about me, but it is about those who gave their effort, time and made this work possible. I wish to express my sincere appreciation to my supervisors, Professor Ashok Khurana, Professor Steve Begg and Professor Reidar Bratvold for their help and support throughout this research. Especially Professor Ashok Khurana for his continuous guidance and support and the hours he spent with me questioning and challenging my arguments, asking questions and guiding me through this PhD thesis. Also, I would like to thank the following:

From Australian School of Petroleum, University of Adelaide, Australia

Professor Peter Behrenbruch for his guidance on offshore development cases. Professor Hemanta Sarma for always spending time with me and providing support and guidance. Steve Mackie and Matthew Welsh for providing valuable comments on my papers.

From School of Mechanical Engineering, University of Adelaide, Australia

Karen Adams for reviewing and editing my papers.

From School of Mathematical Sciences, University of Adelaide, Australia

Professor John von der Hock for his valuable comments on the mathematical side of copulas and Professor David Green for his comments on the dependencies and interactions using the systems approach.

From the faculty of Engineering, Computer & Mathematical Sciences, Australia

Executive Dean Professor Peter Dowd for his financial assistance for this research.

From Santos, Adelaide office, Australia

Alan Beeston for his support and guidance concerning the current oil and gas practice in oil and gas evaluations.

From the Sultanate of OMAN

I wish to thank Sultan Qaboos University for funding this research and their continuous support throughout this research.

From Risk analysis Consultant firm; Houston, Texas, USA.

Dr. James Murtha for his valuable comments on the copulas paper, Dr. Murtha is a consultant in risk analysis in the oil and gas industry.

From Decision Decisions Consultant firm; Norman, Oklahoma, USA

Professor John Campbell for his comments in the impact of dependencies at the portfolio levels.

For the anonymous external examiners who reviewed this thesis.

I would like to thank you for your effort and time in bringing this work to life.

Finally, I would like to thank my brothers, sisters and friends who I met during the PhD program for their support and encouragement.

ABSTRACT

Current development decision-making practices in the oil and gas industry focus on detailed modelling of every decision parameter such as reservoir behaviour, oil and gas reserves, production schedule, facilities design, costs and prices individually without paying adequate attention to the dependencies and interactions between these parameters. Modelling of dependencies and interactions promotes the integration of all the relevant parameters of petroleum project evaluation in a holistic manner. Separate and sequential modelling of individual components of investment decisions limits the ability to examine how changes in one component impact on other components of the system. Literature review revealed that current oil and gas practices ignore dependencies at reserves, project and portfolio levels.

This thesis first investigates the impact of dependencies at the reserves level and tests the copulas approach as a new statistical tool for modelling dependencies in reserves calculations. Secondly, it hypothesizes that modelling dependencies and interactions should not be limited to estimating reserves, but should be extended to model the total system representing the whole project. Thirdly, it explores the impact of dependencies and interactions at the portfolio level. Finally, this thesis compares the systems approach for investment decision making, with the current industry practice of using decision tree analysis.

This thesis concludes that at the reserves level dependencies are significant, and the copulas approach proved to be a better method than the envelope and regression methods, but yielded similar results as the Iman-conover method. At the project level, the impact of dependencies and interactions are significant. This is demonstrated by comparing the systems approach which captures dependencies, to the sequential approach which ignores them. At the portfolio level, including dependencies and interactions leads to higher efficient portfolio frontier and yields greater returns at the same level of risk. Finally, the systems approach proposed in this thesis is superior in

capturing dependencies and interactions compared to the decision tree analysis approach currently used in the oil and gas industry. This research concludes that in the context of dependencies and interactions and the problem investigated, the systems approach is the way forward for economic evaluations of oil and gas projects.

PUBLICATIONS

Peer reviewed Journal Papers

Al-Harthy, M. Begg, S. Bratvold, R. Copulas: A New Technique to Model Dependence in Petroleum Decision Making. Special issue: Oil and Gas Exploration and Production Research in Australia 2005. *Journal of Petroleum Science and Engineering, 2006 (in Press). Accepted in 18 March, 2005.*

Al-Harthy, M. Khurana, A. Begg, S. Bratvold, R. Sequential and Systems Approaches for Evaluating Investment Decisions: Influence of functional dependencies and interactions. *The Australian Petroleum Production and Exploration Association (APPEA) Journal.* Volume 46, part 1, *7-11 May, 2006.* (in press). *Accepted in 21 September, 2005.*

Conference Papers

Al-Harthy, M. Khurana, A. Begg, S. Bratvold, R. Sequential and Systems Approaches for Evaluating Investment Decisions: Influence of functional dependencies and interactions. *Presented at the Gold Coast APPEA 2006. Gold Coast, Australia. 7-11 May, 2006. Accepted in 21 September, 2005.*

Al-Harthy, M. Khurana, A. Portfolio Optimization: Systems vs. Sequential. SPE 101152 *Asia Pacific Oil & Gas Conference and Exhibition to be held in Adelaide, Australia*, 11–13 September, 2006. Accepted in 23 February, 2006.

STATEMENT OF ORIGINALITY

This work contains no material which has been accepted for the award of any other degree or diploma at any university or other tertiary institution and, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

Signed:-

Date:---4/7/2006--------

TABLE OF CONTENTS

CHAPTER 2.

DO STOCHASTIC AND INTEGRATED MODELS ADD VALUE?

CHAPTER 3.

TECHNIQUES FOR MODELLING STATISTICAL DEPENDENCE

CHAPTER 4.

FUNCTIONAL DEPENDENCE AND SYSTEMS MODELLING AT PROJECT AND PORTFOLIO LEVELS

CHAPTER 5.

METHODOLOGY: CONSTRUCTION OF COPULAS AND SYSTEMS APPROACH

CHAPTER 6.

TESTING COPULAS: MODELLING DEPENDENCE IN RESERVES CALCULATIONS

CHAPTER 7.

SIGNIFICANCE OF CAPTURING FUNCTIONAL DEPENDENCIES AND INTERACTIONS

LIST OF FIGURES

Chapter 5

Chapter 6

Chapter 7

Chapter 8

Chapter 9

LIST OF TABLES

Chapter 8

Chapter 9

NOMENCLATURE

The following nomenclature is used in this thesis. Generally, only one meaning is assigned to each symbol. In those cases where more than one definition is possible, the correct one will be evident from the context in which is used.

A	Area, acres
a	Half the major axis of the drainage ellipse, ft
α	Long term growth for oil price, fraction
B_{oi}	Oil formation volume factor RB/STB
b	Intercept of a line
C	A copula
$C(u,v)$	Copula function
C_u	The derivative of copula with respect to u
C_u^{-1}	The inverse of the derivative of a copula
C_w	Net capital cost per well, $ per well
D	Discount rate, fraction
D_t	Decline rate at a time t
E	Escalating or de-escalating factor, fraction per year
E_w	Net OPEX per well, $ per well
E_R	Rccovery efficiency, fraction
F_1	First marginal distribution
F_2	Second marginal distribution
$f(x)$	Function of a variable x
$f'(x)$	The derivative of a function
H	Joint distribution
h	Average thickness, ft
IRR	Internal Rate of Return
η	Reversion speed, years
θ	Correlation parameter (theta)
K	Nonparametric estimate of the copulas
Kc	Parametric estimates for the copulas
$K'c$	The derivative of the parametric estimates
K_a	Absolute permeability, md
K_h	Horizontal permeability, md
K_R	Reversion factor, 1/year
K_v	Vertical permeability, md

L_h	Length of horizontal well, ft
MD	Minimum distance
M^*	Correlated matrix
m	Slope of a line
NPV	Net Present Value ($)
N_o	Original Oil in Place (OOIP), STB
$N(0,e)$	Normal distribution of zero mean and standard deviation of error
n	Number of samples
μ_0	Oil viscosity at bubble point, cp
P_a	Abandonment pressure, psi
P_b	Bubble point pressure, psi
P_d	Daily production rate, barrels per day
PI	Productivity index, STB/day/psi
P_n	Net oil price after all taxes and royalties, $/barrel
P_0	First year oil price, $/barrel
P_{pl}	Natural log of the long-term price, $/barrel
p_p	Long term oil price, $/barrel
P_{Td}	Total daily production or sum of all well daily production, barrels per day
P_{tl}	Natural log of the current price, $/barrel
P_{yt}	Yearly production, barrels per year
Q	Production Capacity = Facility limit, barrels per day or barrels per year
q_i	Initial production, barrels per day
σ	Short term volatility of oil price, fraction per year
R_0	Technical Reserves (initial reserves), million barrels
R_t	Remaining Reserves, million barrels at time t
R_x	Rank of x values
$\bar{R}_x$	Mean of the rank of x values
R_y	Rank of y values
$\bar{R}_y$	Mean of the rank of y values
R^*	Rank correlation matrix
R *and* P	Output distributions
r	Pearson correlation
r_e	Drainage radius of horizontal well, ft
r_s	Spearman correlation
r_w	Wellbore radius, ft
S_{wi}	Initial water saturation, fraction
Tab	Abandonment time, years
T_{21}	The movement from point 2 to a reference point 1
T_i	Pseudo-observations for i values

T_p	The movement of other points to a reference point
t	1. A variable between zero and 1 (CDF) 2. Time
τ	Kendall's tau correlation
Φ	Average porosity, fraction
ϕ	Generator for the copulas
$\phi^{/}$	The first derivative of the generator
$\phi^{[-1]}$	The inverse of the generator
U	Uniform distribution
u, v, s and q	Uniform distributions
W	Total number of wells for the field
w_0	Initial number of wells. Number of wells for the first year
w_t	Yearly number of wells
X_i	a point value
$\bar{X}$	Mean of x values
X, Y and N	Input distributions
Y_i	A point value
$\bar{Y}$	Mean of y value
$(Y_{lower})_x$	Lowest possible value of Y, given X
$(Y_{upper})_x$	Highest possible value of Y, given X
$(Y_{norm})_x$	Sampled value from dimensionless normalized distribution
ψ	Normal distribution with N- (0,1)

CHAPTER

1 Why is this research important?

1. Current challenges in the petroleum industry

Current practices of investment decision-making in the oil and gas industry have many challenges that need to be addressed in order to improve performance. Firstly, the current practice in oil and gas industry focuses on detailed, independent modelling of each component of the investment (such as geology, reserves estimates and production) without paying adequate attention to dependencies and interactions that combine all the elements together. The approach proposed in this thesis avoids the need for excessively detailed independent models and substitutes them with simpler models that preserve the logic and physics of the real interactive behaviour. Current research indicates that excessive details do not necessarily yield more accurate results.

Secondly, current industry practice manifests in the modelling of uncertainties in some parts of the petroleum system and not in others. Treating some system parts to be stochastic and others to be deterministic is basically ignoring the fact that uncertainty propagates through the system. For example, uncertainty in reserves

Note: * $ is US dollar unless stated other wise estimates will impact uncertainty in production as well as economics, because the economic value of the field is dependent on the uncertainty in both reserves and production. Such an approach breaks down the system into components for analysis and usually ignores dependencies and interactions between the parts. The breakdown of the system introduces the need for systematic inclusion of dependencies and interactions at the reserves, project and portfolio levels.

In the context of this research, the terms dependence and interactions are closely related. Variable x is said to be dependent on variable y if a change in x affects the value of y and vice- a-versa. The form of the relationship between x and y is termed their mutual dependence. If two variables are dependent on each other, they are also said to interact with each other.

Current industry approach uses decision tree analysis and attempts to model stochastic dependencies and interactions. However, this is not easy and, is therefore, more often than not, is ignored to some extent.

The above challenges lead to the need for creating tools that are able to capture stochastic dependencies and interactions.

1.1. Investment performance of the upstream oil and gas industry

The low investment performance of oil and gas industry was examined by Merrow (2003). He analysed performance of upstream projects for the last several decades and concluded that the 1990's might be called a decade of unprofitable growth for many upstream companies. Merrow analysed 1000 projects in exploration and production; two thirds were offshore projects. These projects ranged in size from $1 million to more than $1 billion. Volatility in project outcomes shows that there is a huge gap between the economic outcome of average industry projects, best practice and the disasters categories. Merrow, as shown in Figure 1-1 defined disaster as those projects, which carried at least two of the following three attributes:

1. Cost growth of 30% over budget
2. A schedule slip up to 35% or more
3. And operability index of less than 50% of the plan in the first year

He defined best practice as those projects, which had cost within 6% of budget estimate, a schedule close to estimated and operability within 80% of the plan in the first year. Merrow found that one in eight of all major offshore developments fell into the "disaster" category.

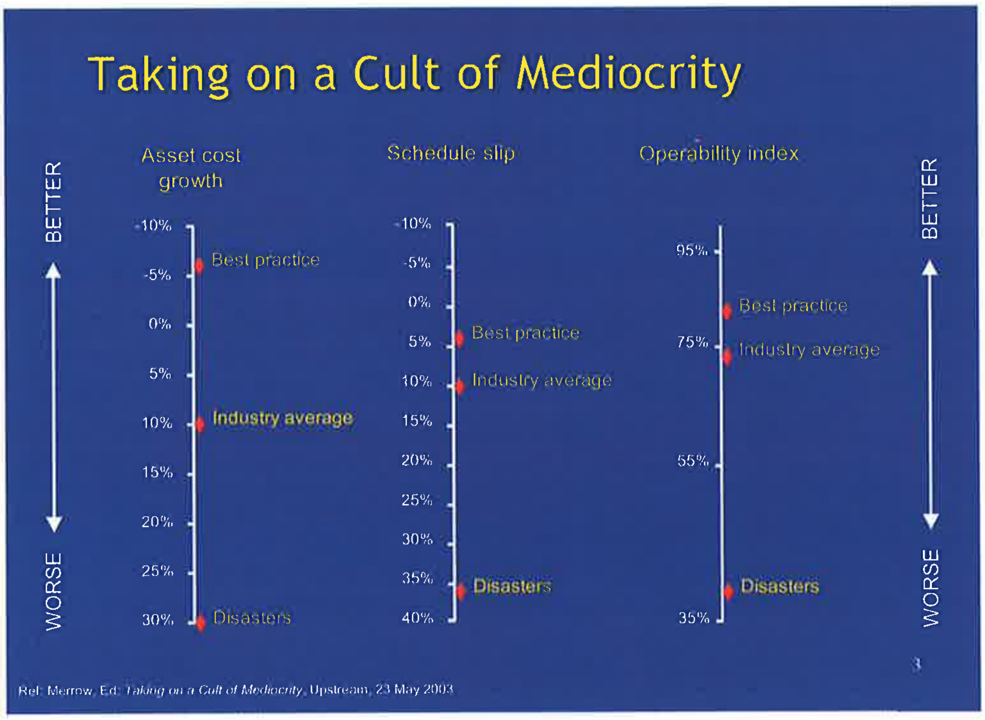


Figure 1-1. Oil and gas performance metric, cost growth, schedule slip and operability index (Merrow, Personal communication)

This result indicates that the current decision tools are not good enough to get the expected results. To uncover the reasons of the low performance of the upstream oil and gas industry a panel discussion was conducted by Jerry Brashear, John Campbell, Truett Enloe, Craig Narum, Richard Rowe and Michael Walls (Poruban, 2000).

The panel focused on the reasons for the low performance of many upstream companies and concluded that the use of decision analysis tools would improve the decision-making and would boost rates of return on capital in exploration and production. Brashear pointed out that with the current practice there is an overstating of expected values, understating of risk and misallocation of capital. Furthermore, Narum indicated that there are many reasons why company returns are so low; the primary reason being the poor link between uncertainty and returns. Campbell however indicated that the tools are not the problem. The problem is the unwillingness of people and senior management to use risk and decision analysis tools. Furthermore, Campbell proposed to overcome this problem by developing a visualization of how portfolio projects look as the key for managers to adopt the process and gain more insights. Narum indicated that the implementation of risk and portfolio techniques should take 2-3 years not 2-3 months. Walls pointed out the lack of defining risk tolerance and not being consistent every time a new project is undertaken is one of the factors in the low performance of oil and gas projects.

The panel discussion pointed out the need to make investment decisions using Decision and Risk Analysis (D&RA) tools in order to improve the performance of oil firms. This is supported by Jonkman et al (2000) who discussed the best practices and their relationship to performance in the petroleum industry. They showed that companies that have been introduced to D&RA have performed better than others, which were not exposed to the same methods.

Further evidence is reported by Simpson et al (2000) who discussed the practice of using D&RA in the petroleum industry in the UK. They indicated, after interviewing different oil companies, that there is a non-parametric relationship between using D&RA and the performance of oil companies. Firms that are using D&RA are better performers than those who are not.

The above discussed issues provide strong evidence that using D&RA will lead to better performance. This is not to suggest that company failure or poor performance is only due to the lack of the implementation of D&RA techniques. It could be argued that there are many reasons why a company fails or has poor performance. However, the fact is that there is a relationship between poor understanding of risk and uncertainty and the poor performance of the oil companies.

Experience and literature review also shows that a competitive advantage in decision quality can be gained by not focusing exclusively on technology and cost reduction, but by also capitalising on the skills and knowledge of people, their ability to process information and to act on it.

Good decision-making practices in complex and difficult situations are crucial to a successful firm. In oil and gas investment projects, the difference between a good decision and a bad one can be the difference between success and failure. Intuition and rules of thumb are no longer sufficiently reliable. They are valuable in stable environments but once changes occur, these rules breakdown. A good investment decision-making process is one that is able to assess risk and uncertainty and is able to manage these in a balanced manner. In order to improve performance through good investment decisions, there is a need to know where the oil and gas industry currently is in terms of its management of risk and uncertainty and what are the possible ways to move forward to achieve quality investment decisions.

1.2. Current practice in the oil and gas industry

Current practice in the petroleum industry is more focused on detailed modelling (precision) and there is little effort in trying to integrate all the components of petroleum projects as one unit, a holistic view as indicated by Begg et al (2001). It lacks the balanced management of risk and uncertainty in all project components to varying degrees. Bos (2005) further emphasized this idea and presented the current practice on a "modelling cube" with three axes: precision, integration and uncertainty (Figure 1-2). Current practice is characterized with high precision, medium integration and low to medium uncertainty modelling.

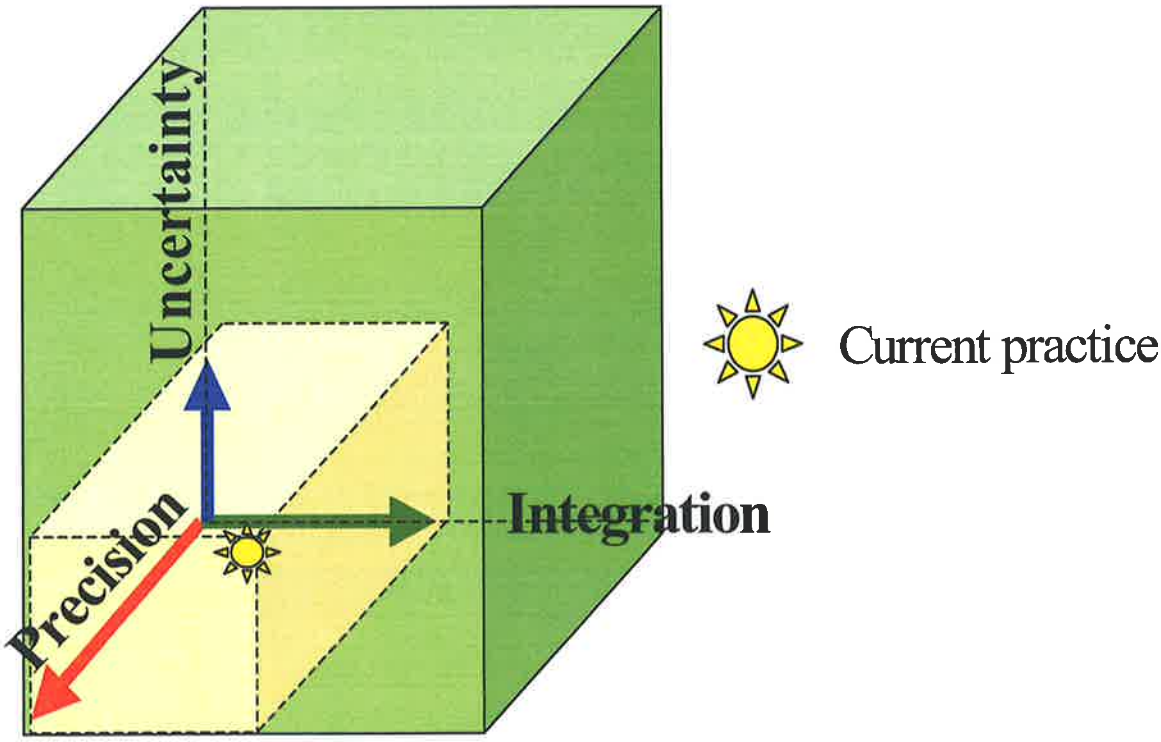


Figure 1-2. "The model cube". Current industry practice: high precision, medium integration and low to medium uncertainty (Bos, 2005)

1.2.1 Precision axis

The current practice puts too much detail into modelling components of the oil industry projects such as geology, reserves, facilities and production without paying little attention to the overall holistic process of combining all the elements together through dependencies and interactions. Consideration of dependencies helps define the input needed to move the project evaluation from one function to another. Excessive details on a single component such as geology is not as important as considering what input does a geologist need to convey to the reservoir engineer or to the economist to achieve better decision quality. This emphasis should be based on business not technological considerations.

It is also essential to point out that models that are too simple are not recommended because the goal is to capture the main drivers of uncertainty. If the model is too simple, then it will be of no use as a petroleum investment decision support tool.

1.2.2. Uncertainty axis

Current practices put more emphasis on deterministic rather than uncertainty models. The modelling of uncertainty is generally restricted to certain areas like reserves. Introducing uncertainties over the deterministic models introduces new challenges and opportunities in the modelling of dependencies and interactions.

1.2.2.1 Below and above ground uncertainties

The current practice is more concerned with modelling below ground uncertainties while ignoring the above ground uncertainties. There is a need to model reserves and incorporate subsurface uncertainties relating to rock and fluid properties. For exploration projects, other factors such as source rock, migration and seal integrity also need to be considered. However, limiting views to below ground uncertainties and ignoring above ground uncertainties such as price, production, development schemes, environmental concerns and fiscal regimes can have a huge impact on the overall value of the project. Brashear et al (1999) indicated that some projects selected to have expected returns of 10 to 20% ended up yielding returns in the region of 5 %. They attributed the poor outcomes to an inadequate appreciation of the above ground uncertainties.

Simpson et al (2000) conducted a survey of UK companies and concluded that most of the companies (82.5% -16 companies) use Monte Carlo Simulation (MCS) for reserves calculation recognizing uncertainties in input variables but only 7.5% (three companies) used MCS in economic evaluations. In general, companies assumed production, cost and economic parameters to be known deterministically (Figure 1-3).

Criteria	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	% usage
Analysis	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	97.5
Holistic view	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	95
Discounted cash flow	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	97.5
Risk and uncertainty definitions	0	0	0	0	1	0	0	1	0	2	1	0	2	0	2	1	2	1	1	1	37.5
Use MC for prospect reserves	0	0	1	2	2	2	2	2	2	2	2	1	2	2	2	2	1	2	2	2	82.5
Take p10,p50,p90	0	0	1	0	0	0	1	0	1	0	0	1	1	2	2	1	1	2	2	2	42.5
EMV, via decision tree	0	0	1	2	1	2	0	2	2	2	2	2	2	2	2	2	1	2	2	2	77.5
Use MC for prospect economics	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	7.5
Use MC for production reserves	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2.5
Use MC for production economics	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Portfolio theory	0	0	0	0	0	0	1	0	1	0	1	2	1	0	1	1	2	2	1	1	35
Option theory	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	0	1	1	15
Preference Theory	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	1	10
Qualitative and Quantitative	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	5
Total	3	5	9	10	10	10	10	11	12	12	12	13	15	15	15	15	15	16	17	17	

Figure 1-3. Use of decision-making techniques among UK companies. (Simpson, 2000).

Figure 1-3 shows the extent and intensity of decision criteria (technique) used by the surveyed companies denoted by the letter A through T. A value of zero indicates no usage, a value of one indicates moderate and value of two indicates frequent usage. The last column shows the usage level of each criteria by all the companies surveyed. Hence 100% would indicate that all the companies are using that technique to its fullest where as 0% would indicate that none are using that technique.

Furthermore, similar results from US companies in using decision-making tools were presented by Schuyler (1997). The results of Schuyler's are shown in Figure 1-4. The first figure in each box is the mean followed by the standard deviation and the sample size. A scale of 1 to 5 was chosen to measure the use of risk and uncertainty methods (Monte Carlo and Decision Tree tools) where 1 represents infrequent use and 5 represents common or frequent use.

Survey response as $\bar{x}$ (Mean), s (Standard deviation), n (Sample size)			
	A Exploration	B Production	C Downstream
1. Decision Tree	3.00,1.62, 20	2.63, 1.46, 19	1.90, 0.99,10
2. Monte Carlo	3.0, 1.60, 19	2.12, 1.22, 17	1.33, 0.71, 9
3. Option Theory	1.11, 0.33, 9	0.89, 0.33, 9	0.80,0.45, 5

Figure 1- 4. Use of decision tools in US (Schuyler, 1997)

From the above Figure it can be seen that there is a low level of application of risk and uncertainty in exploration, production and downstream. Analysing the results and comparing between the UK and the US it seems the US is more acknowledging of the fact that uncertainty should be considered. However like the UK, the result indicated very low application of uncertainty in production, costs and economic parameters. The results of the UK survey indicated that the modelling of uncertainty is largely restricted to reserves while assuming that production, development cost and economics are all certain. This assumption clearly indicates a breakdown of the system, where for example, reserves and production are treated independently by introducing uncertainty in reserves whilst assuming a certain production profile. In reality, variables that impact the uncertainty in reserves are the same variables that impact the production profile. Modelling reserves stochastically and production deterministically is implying that reserves and production are not interacting or not dependent on each other. In other words, the uncertainty in production is not dependent on the uncertainty in reserves.

1.2.2.2. Dependencies in an uncertain environment

Introducing complete modelling of uncertainty below and above ground introduces another challenge, which is modelling stochastic dependencies among variables that are correlated. Stochastic dependency is different from deterministic dependency. Deterministic tools alone are not enough to capture stochastic dependencies; there is a need for other methods that are suitable for a stochastic environment. Stochastic dependencies can be applied in many places in the holistic petroleum model. For example, in calculating the Original Oil in Place (OOIP), porosity and water saturation are negatively related and ignoring this fact can have an impact on the OOIP calculated.

Newendorp (1976) recommended that simulation programs should model dependencies and considered it to be a critical factor in prospect analysis. Furthermore, dependencies and interactions should not only be considered in the reserves but also should be extended to model dependency between other components of the whole petroleum system.

1.2.3. Integration axis

The challenge of investment decisions is to integrate all the components of the petroleum model and avoid the narrow picture where the focus is only integrating geology with reservoir or integrating production with economics. Integration implies a holistic view looking at the petroleum model as one system from geology to economics. In addition, integration creates another challenge, that of capturing interaction among the components of the petroleum systems.

1.2.3.1 The narrow picture

Current petroleum investment decisions practice focuses primarily on stochastically modelling uncertainties in the reservoir and ignoring other uncertainties in cost, production, development and prices or at best dealing with those uncertainties deterministically. Murtha (1993) presented a model that focuses on the calculation of OOIP and how its input affected the calculated OOIP. Murtha's model lacks the complete modelling of uncertainties. It looks at a narrow picture instead of focusing on the overall impact of all components on the whole project. However, years later in a different paper (Murtha, 1997), stated clearly that there is a need for an integrated model that will not only focus on reserves but also on other components such as production forecasts and both capital and operation expenses. This clearly shows the need for an integrated model that deals with uncertainty in all of its components.

1.2.3.2 Missing interactions at the project levels

In order to achieve integration between components such as reserves, production and economic parameters, it is not only important to account for uncertainty in all the components but also to account for dependencies and interactions among those components. The traditional approach lacks the inclusion of dependencies and interactions. For example, the need to see how a change in the geologic success factor (a measure of the risk in the subsurface) impacts the Net Present Value (NPV) is vital. Small changes in the geologic success factor could lead to large variation in NPV. This interaction between geology and economics could be achieved in an integrated stochastic model. Other interactions should be considered as well, such as reserves and production or production and development schemes.

1.2.3.3 Interactions at the portfolio level

Current portfolio approaches take into account the impact of inter-dependence between projects. For example, projects that are from the same play will be correlated with each other from a technical point of view by having similar geology. Furthermore, on the economic level, these projects could be correlated by the same oil price. However, none of the literature so far considered the impact of including intra-dependence of projects at the portfolio level. This research argues that there is a potential value in combining both intra- and inter- dependence at the portfolio level.

The output from the integrated model such as mean and standard deviation of NPV could be used as an input to a portfolio optimisation model, which is a tool to choose optimal projects. There are two important points from the portfolio optimisation that are essential to the valuation of projects. Firstly, an optimal portfolio can only be achieved by recognizing correlation among projects. Secondly, the economic value of a single project should not be judged alone but should be based on how it contributes to the optimal portfolio of projects.

1.3. Current industry approach: decision tree analysis

The use of decision trees to incorporate risk and uncertainty is a common method in oil and gas investment decisions. This method is good for evaluating alternative scenarios and sequential decisions. However, this research argues that in practice this method does not adequately incorporate the impact of dependencies and interactions. The author believes that the value of using a method that captures dependencies and interactions would be better than the current industry approach. A detailed introduction and application of the decision tree approach is covered by Mian (2000) and Newondorp and Schuyler (2000). A short introduction is provided in Chapter 9.

1.4. Vision for the future, where we want to be

In order to overcome these challenges, a vision for future models was proposed. Narum proposed a solution or a direction, indicating that

" Industry should keep moving toward finding a way to integrate technical tools with economic ones on a more dynamic simulation model such as through earth modelling. To look at the physics of technical tools dynamically linked in a stochastic way to the economic model- subsurface to surface, to different markets, different well schemes, platform schemes- and being able to optimise the entire solution with the technical end of the business would be the next big step" (Poruban, 2000).

Narum's vision supported the need for a stochastic integrated model that captures uncertainty and integrates all the components of the petroleum system from geology to economics. This is a key in investigating the impact of dependencies and interactions at the project and portfolio levels.

Both Begg et al (2001) and Bos (2005) propose a vision for the future where models have lower precision; enough to preserve the physics and logic of the detailed model and to fit the purpose of the decision being made. Furthermore, they encourage maximum integration (holistic view) and maximum uncertainty modelling. The level of precision required should be determined by the needs of the decision to be modelled (Figure 1-5).

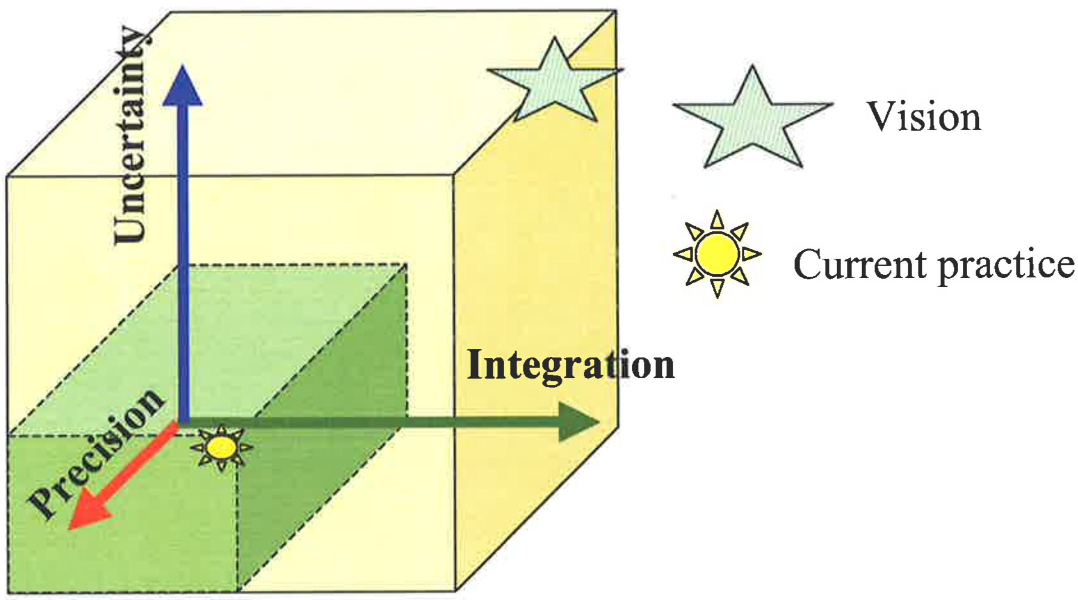


Figure 1-5. "The model cube" the future vision: lower precision, high integration and high uncertainty, modified (Bos, 2005)

1.5. Objectives and scope of the research

The aim of this research is to contribute to the future vision model, specifically to both the uncertainty and integration axes as well as the precision axis by using simpler models that preserve the behaviour and physics of the detailed models. This contribution to the three axes is captured in the value of modelling dependencies and interaction at the reserves, project and portfolio levels.

The research hypothesis is that there is value in modelling dependencies and interactions in the whole petroleum system. Dependencies and interaction should not only be considered in the reserves but should be extended to model dependency between other components of the petroleum system in a single project and at the portfolio levels. This should be done in a stochastic environment where all the components of the petroleum system can capture risk and uncertainty.

Furthermore, this research hypothesizes that an approach that is able to capture stochastic dependencies and interaction at reserves and project level will add value to the quality of investment decisions compared to the current decision tree approach which is commonly used in the oil and gas industry.

This thesis will focus on investigating the impact of dependencies and interactions at three levels:

- Dependencies at the reserves level
- Dependencies at the project level
- Dependencies at the portfolio level
- And finally the systems and the decision tree approaches will be compared with each others.

This research has the following objectives:

1. **Build a stochastic integrated asset model, which is called the systems approach, that displays the following:**
    - Coverage of all components of a typical petroleum project from reserves to economics.
    - Ability to model above and below ground uncertainties;
    - A simple design of a stochastic model that avoids unnecessary details;
2. **At the reserves level: To investigate the impact of statistical dependencies on technical reserves.**

    This objective has two specific sub-objectives:

- It focuses on the impact of the *dependence structure* of the correlation model on the technical reserves.
- It compares and contrasts the current techniques in oil and gas project evaluation with the copulas method as a new statistical tool for modelling dependencies in reserves.

The impact of statistical dependence is investigated only at the reserves level. The decision to study the impact on reserves only was taken to reduce the complexity of the problem and to focus on understanding the impact of dependence structure and the difference between dependence methods. Similarly, as discussed below, a decision was made to investigate functional dependency at the project level.

3. **At the project level: To investigate the impact of functional dependencies and interactions on the NPV of a development decision through a comparison of the systems and sequential approaches.**

The statistical dependence can also be used to investigate the impact at the project level. However, this research focuses on studying functional dependency at the project level, which captures interaction between components of the petroleum system. Impact of the functional dependency is illustrated using the sequential (traditional) as well as the systems approach developed in the first objective.

4. **At the portfolio level: To investigate the impact of dependencies and interactions (intra-dependence) for a mix of development decision projects.**

This objective investigates the impact of intra-dependence for a mix of development projects using the Markwitz mean-variance portfolio model.

5. **And finally, Compare the impact of the systems approach and the current industry approach (Decision tree analysis) on the NPV of an offshore development decision.**

   This objective investigates the impact of evaluating a development decision using the systems approach suggested in this research compared to the decision tree approach which is common in the oil and gas industry.

1.6. How the thesis objectives contribute to the vision model

In terms of the Bos modelling cube, the first objective focuses on building a stochastic integrated asset model that addresses all three axes by having simpler models that represent maximum uncertainty and integration. The second objective focuses on enhancing the uncertainty axis and specifically shows the impact of statistical dependence between uncertain variables at the reserves level. The third objective focuses on both the uncertainty and integration axes, by which it explores the impact of functional dependencies or interaction among stochastic components of the petroleum model. On the uncertainty axis, it builds a complete model that incorporates both above and below ground uncertainties. On the integration axis, it focuses on the integration of all the components of petroleum system such as reserves, production, drilling, facility and economic parameters through building interaction and dependencies that link all the components of the petroleum system. The fourth objective enhances the value of dependencies and interactions from individual project level to a portfolio level. This contributes greatly to the uncertainty and integration axes. The final objective compares this author's vision approach, called the systems approach, with the current industry approach and shows the value added by using the systems approach.

1.7. Outline of this thesis

The outline of this research (Figure 1-6) is as follows:

In Chapter 1, the investment performance of the oil and gas industry for the last decade was characterised as mediocre and it was demonstrated that there is a need to improve. In order to improve performance there is a need to know where we are with regard to the current practice of incorporating uncertainty and integration in the oil and gas industry. Current industry status and future vision of where we want to be were defined.

The current practice is viewed in three axes, precision, uncertainty and integration, Bos's cube model. Problems with the current practice were pointed out as challenges. This research formulates these problems into objectives to be explored by this thesis. And by exploring and achieving these objectives, this thesis will contribute to the future vision of where we want to be. By viewing the current practice through the three axes of precision, uncertainty and integration, Chapter 1 points out two main problems:

- Lack of using stochastic modelling over the entire petroleum project system.
- The lack of integration through not modelling dependencies and interactions.

Chapter 2 will investigate the two problems through literature review and show that stochastic models are better than deterministic models in every discipline of the petroleum process (geology, reserves, production, facilities and economics) and points out that there is value in integration through modelling dependencies and interactions. Furthermore, this Chapter will set the foundation for building the systems approach which is stochastic and integrated.

Chapter 3 will introduce the literature review on the current methods used in oil and gas evaluations to model statistical dependencies. Further, it will

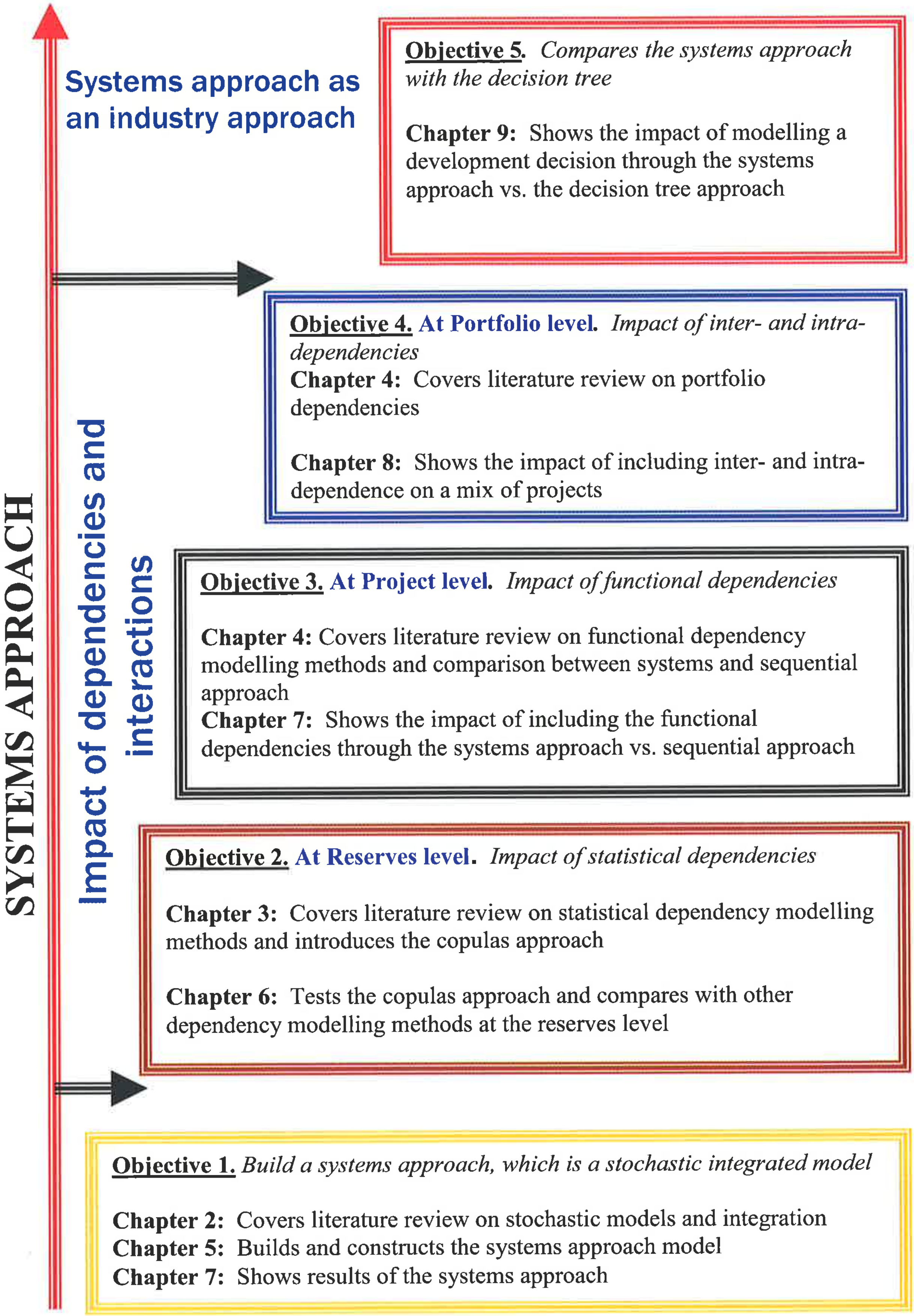


Figure 1-6. Outline of this research and connected Chapters

introduce the copulas method, which is used in financial risk management, but not in the petroleum industry. These methods are used to investigate dependencies at the reserves level which is the second objective of this thesis.

Chapter 4 will review the literature and present simple experiments which illustrate the third and fourth objectives that are concerned with examining the impact of dependencies at the project and portfolio levels.

Chapter 5 will introduce the methodology used to investigate the objectives stated above. By first introducing the methodology used in construction and comparison of the copulas model, and then showing an example of the construction of the simulation algorithm used. Furthermore, Chapter 5 introduces the construction of the stochastic integrated model; the systems approach suggested by this thesis.

Chapter 6 will focus on the experiments, results of the statistical dependencies model and provides a discussion and conclusion of the main findings of the impact of statistical dependence through the importance of capturing dependence structure of the correlation variables at the reserves level.

Chapter 7 will illustrate the experiments and results of the impact of functional dependencies on the development decision at the project level through the sequential and systems approaches.

Chapter 8 will provide the results and conclusions of the impact of inter- and intra-project dependencies at the portfolio level.

In Chapter 9, the use of the systems approach developed in this research, and used to capture dependencies and interactions through the previous objectives, will be compared to the decision tree approach.

And finally, Chapter 10 will provide a summary and conclusion of all the findings focusing on recommendations and future work.

CHAPTER

2

Do stochastic and integrated models add value?

2. Introduction

This Chapter presents the literature review that lays the foundation and explains the reasons why the first objective of this thesis should be attempted. The literature review points out the need to develop a holistic, integrated stochastic model that covers all the components of the petroleum system from reserves to economics as well as the need to model above and below ground uncertainties. This holistic stochastic integrated asset model has been termed the systems approach. The two key words are stochastic (probabilistic) and integrated (holistic). This research argues that using a stochastic approach in every part of the petroleum process is only a part of the solution. The other part is the integration of all the stochastic parts of the petroleum system through dependencies and interactions.

The literature review will focus on two main points:

1. The use of the probabilistic approach versus the deterministic approach in estimating risk and uncertainty in every area or discipline in the petroleum system, such as reserves, production, facilities and economic parameters.

2. The value of integration in the petroleum system, the value of a holistic view, and the need for inclusion of dependencies and interactions among the components of the petroleum system.

The components of the petroleum development project evaluation are primarily focused on answering five main questions;

- What are the chances of finding hydrocarbon?
- What is the original oil (hydrocarbon) in place and how much can be recovered?
- What is the production forecast?
- How to develop?
- And finally, is it economically viable?

The first part of this Chapter will focus on the current practices in the oil and gas industry of using stochastic models to capture risk and uncertainty by answering each one of the above questions. The second part will focus on the value of integration of all the components of petroleum development project.

2.1. What are the chances of finding hydrocarbon?

Several geological elements are essential for oil to accumulate in economical quantities (Figure 2- 1). These elements include a source rock to generate oil or gas, a migration pathway, a reservoir rock to contain and allow its production and a trap to prevent it from migrating. A geologist describes these elements in a qualitative way and then transfers the qualitative description into quantitative probabilities for each element. The risk elements are then multiplied to get the probability of geological success.

Anders (1994) discussed the subjective and objective methods of determining the probability of geological success. He indicated that a subjective method relies on intuition and personal experience. It has the advantage of being quick and simple but it preforms poorly. On the other hand, an objective method that uses frequencies and distributions has the advantage of removing subjectivity and personal biases, but the

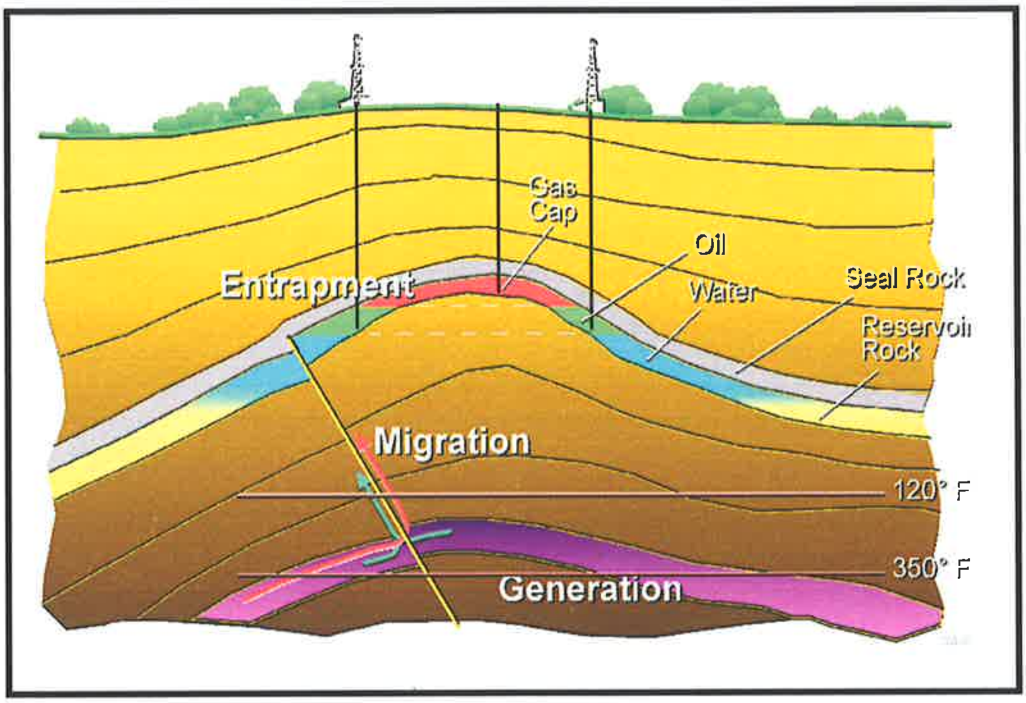


Figure 2- 1. Geological elements for oil accumulation (Houston geological society, 2003)

need to gather substantial amount of data is essential for its success. He concluded that in order to improve exploration economics, uncertainty should be included in the input variables using the objective method.

Otis and Schneidermann (1997) enhanced and encouraged the use of the probabilistic approach by presenting a checklist of the geological factors that need to be considered for determining the probability of existence of petroleum accumulations. The checklist consists of four main areas; source rock, reservoir, trap and migration and its timing. Furthermore, they provided a standard method for transferring qualitative judgements on geologic risk to quantitative probability of geologic success.

2.2 What is the original oil (hydrocarbon) in place and how much can be recovered?

This question focuses on estimating Original Oil (hydrocarbon) in Place (OOIP) and reserves. A model that focuses on the calculation of OOIP and how various inputs affect the calculated OOIP was presented by Murtha (1993). He assigned to each input variable a distribution and then ran a Monte Carlo Simulation (MCS) to produce an outcome distribution. Murtha concluded that by using field data and the probabilistic approach, better results could be achieved and more insight gained into the modelling of dependencies among input variables.

Behrenbruch et al (1985) discussed the estimation of hydrocarbon reserves using probabilistic methods (such as MCS) applied to an offshore field in northwest Australia. They explained in detail the probability distribution used for each input variable in order to calculate the hydrocarbons in place. Furthermore, they indicated that probabilistic methods are important in situations where the system is very complex or if the various elements of the system have to be described in terms of uncertainties due to incomplete information. They concluded that the MCS approach has the advantage over the deterministic method in terms of the following:

- Better handling of partial dependencies
- Allows for realistic gross rock volumes distribution when compared to the deterministic approach and that it can be achieved with less data preparation
- Uses seismic and geological information more directly compared to only mapping the most likely situation.

- And finally, allows for sensitivity evaluation of individual parameters in order to see which ones have a higher impact on the final output.

O'Dell and Lamers (2003) categorize subsurface uncertainties in groups based on their study in Oman. The main categories were: gross rock volume, saturation, reservoir architecture, faults and fractures, reservoir properties such as porosity and permeability and their dependence on lithology and litho facies, relative permeability, PVT data and finally compressibility and compaction. They concluded that this approach helped to identify and rank uncertainties and that quantifying these uncertainties would help to choose the best development options in order to maximize the value of the project.

The use of experimental design and MCS to quantify reservoir uncertainties in developing new fields was discussed by Kabir et al (2002). They indicated that when many uncertain variables exist, it is better to perform an experimental design approach first and produce a regression equation of the form $Y= f (X_1, X_2, X_3, X_4)+ e$, where X_n's are the input variables and e is the random error. Once this equation is produced then it is easy to see which variables have the biggest impact on the dependent variable. Then the second stage is to use MCS on the regression equation. They concluded that the experimental design is a quick, unbiased approach to identify the significant input variables influencing the dependent variable.

Holtz (1993) bridged the gap between geological factors and the calculation of hydrocarbon volumes. He showed how geological factors influence engineering parameters such as area, net pay, porosity and fluid saturation that are used to calculate the hydrocarbons in place and reserves. He divided the geological scale into three; megascopic, macroscopic and microscopic. For example, mega scale geologic features such as depositional system, trap types, and faulting correspond to engineering parameters such as reservoir size, drive mechanism and fluid characteristics. He concluded that consideration of the inter-relationship between the geological factors and reserves would improve the assessment of risk and uncertainty and therefore produce more accurate reserves estimates.

The above literature review results clearly show that in reserves estimates, the probabilistic approach is superior to the deterministic approach and there is a need to model dependencies between geological parameters to estimate reserves.

2.3. What is the production forecast?

Once oil or gas are found and the technical reserves have been estimated, there is a need to estimate how much could be produced economically, what would be the production schedule and what are the uncertainty factors that affect production. The use of probability distributions rather than a single production profile as a way to account for risk and uncertainty in production forecasting was discussed by Spencer and Morgan (1998). They developed a "choke model" where they assigned distributions to different input variables. For example, they represented well capacity by a normal distribution truncated to eliminate non-physical values. Once all the distribution inputs were specified, a simulation was run to forecast the annual production schedule. They concluded that applying uncertainty to production is not an academic exercise, but that it helps to improve the accuracy of production forecast and results in a better understanding of the production model.

Venkataraman (2002) discussed the application of experimental design to quantify uncertainty in production profiles. He used the following steps to implement the experimental design; identify the key variables, conduct the experiment using

MCS, validate and make predictions. He presented a case from the North Sea on a gas project where he found that uncertainty in production arises from reservoir permeability, transmissibility across a fault, the depth of the gas water contact and finally the shale barriers within the formation. He concluded that the experimental design improved the quantification of uncertainty because it allowed for non-linear dependence of the production profiles on the input parameters.

Salomao and Grell (2001) examined uncertainty in the production profile. First they quantified uncertainty in oil in place using probability distributions. The recovery factor was analysed as a function of different uncertainty variables such as vertical and horizontal permeability, relative permeability, viscosity of oil and well productivity. Each one of these variables was modelled as a distribution. Furthermore, they used decline curve analysis to estimate the production forecast through MCS. They concluded that using risk and uncertainty helped to make better decisions and to define a better strategy to develop the field.

Steagall and Schiozer (2001) identified uncertainties that affect volume in place and those that affect fluid flow. They identified uncertainties which have high, moderate, low and no impact on the reservoir production forecasts (Table 2-1). These uncertainties were assigned probability values and numerical simulation software was used for production forecasting. They concluded that the use of numerical simulation in the production forecasts improves the quality of analysis. In addition, they indicated that the reliability of production forecasts depends on the quality of uncertainty analysis and the probability distribution used. Finally, they found that risk analysis in production forecasts requires specific knowledge of the discipline of geology, reservoir engineering and production engineering which implies the need for a multidisciplinary team.

Table 2-1. Reservoir uncertainty attributes, Steagall and Schiozer (2001)

Attributes	**Volumetric Dependence**	**Fluid flow Dependence**
External Geometry, structure	H	M
Reservoir top	M	L
Zones, sub-zones top	L	M
Reservoir base	M	L
Reservoir limits	H	L
Faults	M	H
Porosity	H	M
Lythofacies distribution	H	H
Fluid contacts	H	M
Aquifer geometry	H	M
Absolute Permeability	N	H
Relative Permeability	N	H
Oil and gas Proprieties (PVT)	M	H
Capillary Pressure	M	M
Rock and Fluid compressibility	L	M
Fluid Formation Volume factor	M	L
H = High M = Moderate	L = Low	N = None

Johannessen et al (1994) investigated the use of risk analysis by incorporating probability distributions to well deliverability. Traditionally, a single flow rate is used to calculate well performance. In addition, the flow rate is often restricted by completion configuration not by reservoir deliverability. They computed the well deliverability as a probability distribution reflecting uncertainties relating to input variables. They found that the use of probability distributions resulted in increasing the tubing size which in turn led in increased flow rate and the pre-plateau period being reduced quickly. The overall impact was a significant increase in the Net Present Value (NPV).

The above review shows that the stochastic approach yields better results than the deterministic approach in production forecasting.

2.4. How to develop?

Once technical reserves have been estimated and the production forecast generated then the next step is to determine how to develop the field. The petroleum business is known for being a high capital-intensive industry. The cycle of cash flows shows that the development stage is the most critical in terms of the investment involved. Large amounts of money are spent during this stage; a typical project revenue distribution shows that capital expenditure during this stage can be as much as 40% of the revenue. This stage involves the conceptual facilities design to extract and process hydrocarbons.

Demirmen (2001) indicated that there are three types of development risk:

1. Opportunity loss: this is when a prospect is abandoned as uneconomic while in fact it is economic. This could be the result of poor estimation of the recoverable volume.
2. Uncommercial development: this is when an uneconomic field is developed under the wrong assumption that it is economically viable.
3. Suboptimal development: this is when the development of a field yields less than the maximum economic return that could be achieved with a more appropriate reservoir model.

Suboptimal development results from a mismatch between the capacity of the production facility and the volume of hydrocarbons in the reservoir. The facility could be oversized or undersized, both of which lead to economic loss. Demirmen (2001) showed that this is the most undetected type of risk in the development stage. He showed an example from Asia where a water injection facility was oversized. In another example, in the Gulf of Mexico, the facility was undersized and hence was not able to capture the full production potential of the reservoir.

Demirmen concluded that using the value of information analysis technique and development scenarios helps provide an optimal development design that captures uncertainties in the reserves and the development stage.

Further evidence by Etebar (1995) indicated the use of development scenarios to capture reserves uncertainties. He showed a case from the North Sea where a field discovered in 1977 was unattractive with a conventional plan of 200 conventional development wells. Another look in 1993 at the same field, using the new technology of horizontal wells, and a more detailed look at development scenarios reduced the cost significantly. Better development scenarios consisting of using a Well Head Platform and Floating Production, Storage and Offloading Vessel (FPSO) proved to be significant. Estimated well capacity increased from 740 to 6500 barrels of oil per day (BOPD). The use of development scenarios to capture reserves uncertainty decreased costs estimated and increased estimated production significantly.

Coopersmith and Cunningham (2002) presented a decision analysis approach to appraisal and development evaluation. They discussed the use of decision mapping where each decision is defined and decision trees are used to map development scenarios for an oil field. They presented a case study for developing two offshore areas; area A has proved reserves of 50 Million barrels of Oil (MMBO) and area B has three reserves possibilities (30, 70 and 200 MMBO) with a 70% chance of oil being present in the area.

The question they sought to answer was: what is the best facility structure for both areas and what is the optimum number of wells? The development facility choice ranged from Tie back to Minimum Tension Leg Platform (Mini TLP) to a SPAR.

Using the decision tree analysis helps to produce different development scenarios that will capture uncertainty in reserves and help to determine the facility option yielding the highest economic NPV.

It is clear that the reserves will dictate which facility option to choose. This indicates the importance of capturing uncertainty and how it impacts the choice of development option. For the case study, if the reserves of area A&B are 120 MMBO then the optimal choice is Mini TLP with 8 wells.

2.5. Is it economically viable?

In order to evaluate if an oil field is economically viable, we need to calculate various economic indicators. One of the economic indicators is the NPV for the field. In order to calculate the NPV we need values of oil prices and costs, both capital and operating cost or CAPEX and OPEX, to develop the field. Modelling of oil prices will be explained first and then followed by costs.

The most uncertain variable in the economic model is the oil price. There are no models that can predict prices perfectly. Most of the models of oil prices are discussed in the area of real options and finance (Table 2-2). Dias (1998) discussed different types of stochastic price models that exist in the literature.

These models are:

1. Geometric Brownian Motion,
2. Pure Mean Reversion,
3. Two and Three factors model,
4. Reversion to Uncertain Long Run level
5. And Mean Reversion with Jumps.

Table 2-2. Stochastic models for oil prices, (Dias, 1998)

Type of Stochastic Model	Name of the Model	Main Reference
Unpredictable Model	Geometric Brownian Motion (GBM)	Paddock, Siegel & Smith (80's)
Predictable Model	Pure Mean-Reversion Model (MRM)	Schwartz (1997, model 1)
More Realistic Models	Two and Three Factors Model	Gibson & Schwartz (1990), and Schwartz (models 2 and 3)
	Reversion to Uncertain Long-Run Level	Pindyck (1999) and Baker, Mayfield & Parsons (1998)
	Mean-Reversion with Jumps	Dias & Rocha (1998)

The Geometric Brownian Motion model is the simplest model and Dias (1998) indicated that it is a good approximation, but it is not adequate when spot prices are far from the long run equilibrium. The Mean Reversion model is considered to be a better model because it argues that prices will revert to the long run equilibrium, which tends to make sense with market conditions. For example, if the long run equilibrium is $20 per barrel and if prices increased to $35 per barrel then OPEC will increase production to sell more oil and prices will go down or revert to the long run equilibrium. On the other hand, if price falls below $20 per barrel, then OPEC will restrict production for the price of oil to go up and again the price will revert to the long run equilibrium.

Gibson and Schwartz (1990) developed the Two-Factor model, which is the mean reversion model with convenience yield of oil added. Schwartz (1997) presented a two-factor model but without using convenience yield and he concluded that it performs better than the previous model. The Three-Factor model is the two-factor model with interest rate added as a third stochastic variable following a mean reverting process.

The Reversion to Uncertain Long Run is the Mean Reversion with the long run equilibrium price modelled as Geometric Brownian Motion. The last model is the Mean Reversion with Jumps, Dias looked at oil prices for the last 30 years and indicated that oil prices tend to jump either up or down due to unexpected events such as war, weather, market crashes or the state of the economy; boom or recession.

Edwards and Hewett (1994) looked at the management of oil and gas prices and the quantification of production prediction. They estimated the impact of managing or controlling oil prices by hedging and what impact does production prediction have on hedging. They concluded the following: firstly, the risk and uncertainty in oil prices can be reduced or controlled by the use of financial derivative securities such as future, forward and options trading.

Secondly, the ability to manage oil and gas prices is dependent on the ability to quantify uncertainty in the production prediction. If the production profile is known for sure, then the risk of oil price can be reduced or eliminated by hedging, but if the production profile is uncertain then the hedging of oil prices is less than perfect. The second conclusion is really important to this study where dependency and interaction between models is important in terms of capturing uncertainty rather than looking at a single model alone.

Peterson and Murtha (1993) looked at risk analysis incorporating MCS to estimate the Authorization for Expenditures (AFE) for drilling operations. They used distributions instead of a single number to represent the uncertainties in the number of days of problem free time and problem time in order to calculate the total time for a drilling operation. They concluded that the AFE estimates generated using MCS were a better predictor of actual well days than the conventionally generated AFE estimates. Furthermore, the use of a probabilistic approach offered more insights into the problem free and problem days in drilling operations.

Wood (2001) discussed the use of the probabilistic approach to predict time and cost of a project. He presented an example of a project involving the drilling of a sidetrack well and the installation of separation equipment. Each activity of the project has a cost and duration of how long it will take. All the activities in the project are connected by a critical path network, which determines the optimal path of the project. A distribution is assigned to each activity's cost, duration and dependencies among activities that are related are also modelled. Then a MCS was run to calculate the total cost and time for the whole project. He concluded that producing an average of cost and time helps the project manager to achieve target budgets. In addition, it helps project managers to monitor progress and establish realistic project schedules.

All the previous studies conclude that consideration of uncertainty in each discipline or area rather than using single deterministic values is important for making good decisions. The results of applying the probabilistic approach yield better results than the traditional deterministic approach. However, all the previous studies focus on exploring the impact of uncertainty in separate independent areas or disciplines by themselves. For example, the probabilistic approach was applied to reserves estimating but not to production or economic parameters. None of the previous studies discussed a complete stochastic integrated model from reserves to economics as one unit or one system. Furthermore, the need for integration of all the disciplines from reserves to economics requires the modelling of dependencies and interactions between the disciplines.

The next section will show the value of integration through dependencies and interactions and the need for a probabilistic approach to be applied in all the areas of the petroleum system (Geology, Reservoir Engineering, Production Engineering, Facilities Engineering and Economics).

2.6. The value of integration through dependencies and interactions

The understanding of why projects fail will lead to a better understanding of the need for integration. For example, one could do a good job understanding the description of the reservoir and not focus on wells and facility design, which may lead to a suboptimal solution. Decisions are often made by individuals who do not understand the impact of dependency of one area to another, for example from reservoir to facility design. Furthermore, it is not common to have an individual who has a background in all the required areas (Geology, Reservoir Engineering, Production Engineering, Facilities Engineering and Economics) and in general, teams do better than individuals.

Chow et al (2003) stated the value of integration clearly, "…with a traditional approach to development planning, the reservoir engineer would commonly make estimations of the reserves in place and formation parameters (e.g., pressure and permeability), then hand them off to the completions engineer to size production tubing and determine flow rates. Then, the facilities engineers would be given the assumed production profiles to design the facilities. This approach severely limits any ability to optimize the project interactively between the engineering disciplines or to investigate the effect of changes in any one part of the system with corresponding effects elsewhere in the system (e.g., reserve size, tubing size, completion technique, or various facility options)."

Chow et al (2003) statement clearly shows that the lack of integration can be seen as the lack of modelling dependencies and interaction among the parts of the petroleum system. If dependencies and interactions were included in the model, then it would be possible to see how a change in one part of the system will affect another. Furthermore, the inclusion of dependencies will make it even easier to optimise the entire solution.

Gayton et al (2000), point out that in the traditional approach, each aspect of the project is evaluated individually and optimised and then the results are carried to the next level. For example, reservoir engineers will study the subsurface and develop a reservoir model with production profiles. A single production profile case will be given to the facility engineer to design the facility and the economic value will be calculated based on the production profile and the CAPEX of the facility. The problems with this approach as indicated by Gayton et al (2000) are:

1. The direction of information flow is one way
2. Facility concept is based on one single case of production profile, which might lead to suboptimal design
3. The next area/discipline assumes that the information from previous area is correct, which implies no uncertainties.

The traditional approach implicitly assumes that there is no interaction or dependency among the components of the system. Furthermore, the lack of interaction assumes that a change in one component of the system has no effect on the next one. This lack of interaction implies that a change in reserves will not have an impact on production or the choice of facility, which is not true.

The message from Chow et al (2003) and Gayton et al (2000) is that integration means the ability to see how a change in one part of the system affects others and the ability to optimise the entire solution. This clearly leads to the conclusion that the true value of integration is captured in its ability to model dependencies and interactions.

2.7. Integration of people

One of the questions that could be asked is: How does creating a model that incorporates a complete stochastic approach, systems approach, and integration through the modelling of dependencies and interactions affect petroleum professionals? The answer is that combining the treatment of reserves, production, facilities and economics leads to the needs to create an integrated or multidisciplinary team of geoscientists, reservoir engineers, production technologists, facility engineers and commercial analysts.

Creating integration through modelling of dependencies and interactions requires communication among the petroleum personnel. In this way, any change in reserves estimates will be communicated to all the people in the team. The author believes that having the right integrated model will lead to the creation of the right integrated team. This belief is supported by Schrage (2000) who reports that an innovative model will create an innovative team. If the model does not change or stimulate the way individuals interact in the team then the model is not useful. Effective models do more than change awareness, they change behaviour and the choices people make. Furthermore, Schrage (2000) indicated that managers should not just ask what will this model do? But how will this model affect the behaviour of people? The right integrated model should be able to change behaviour and help the team to innovate. And as indicated by Schrage, it is behaviour, not knowledge, or insight that drives innovation.

2.8. Summary

The literature review in this Chapter addressed the first objective of this research by addressing the need for a stochastic integrated model. This Chapter has shown two important points;

The first part of this Chapter has described that the probabilistic/stochastic approach can be applied to every area in the petroleum system and yields better results than the deterministic approach.

The second part of this Chapter has indicated that the value of integration is captured through the modelling of dependencies and interactions among the parts of the petroleum system.

Combining the first part with the second part showed clearly the value of a holistic stochastic integrated asset model as a means to achieve better results in investment decisions in oil and gas economic evaluations.

Moreover, this Chapter has demonstrated that using the stochastic approach in every part of the petroleum system is only part of the solution. The other part is the integration of all the stochastic parts of the petroleum system through the modelling of dependencies and interaction among the components of the system. An advantage of integration of all the components into a single model is that it will lead to the integration of people from different disciplines, formation of multidisciplinary teams and therefore lead to more innovation.

CHAPTER

3 Techniques for modelling statistical dependence

3. Introduction

The previous Chapter discussed the value of the stochastic approach and the need to model dependencies and interactions. This Chapter continues with the same theme and focuses on introducing methods used to model statistical dependence. This Chapter will review the relevant literature with regard to the second objective of this research, which is investigating the impact of statistical dependencies at the reserves level. The objectives of this Chapter are to:

1. Introduce the current techniques in oil and gas evaluation to model statistical dependencies.
2. Introduce the copulas method, including an overview of its construction and simulation algorithm.

The copulas method is traditionally used in financial and risk management industries, but has not been used in petroleum industry. This technique has the advantage of capturing the dependence structure. The dependence structure is the shape or the pattern of the joint distribution of two correlated variables.

3.1. Correlations and Dependencies

The terms correlation and dependence in the literature are often used interchangeably. Although this is not strictly correct, we will also do so in this thesis.

Although many associate the word "correlation" with a linear correlation between two variables, there are many approaches to model dependencies. In order to analyse dependence there are three measures; Pearson, Spearman and Kendall's tau. These are data analysis techniques used to capture the direction and the magnitude of the correlation.

3.1.1 Pearson correlation

Pearson (linear) correlation is the most common method used to measure dependence between two variables. The Pearson correlation coefficient (r), which correlates two variables - X and Y - is defined by

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2}\sqrt{\sum (Y_i - \bar{Y})^2}} \tag{1}$$

where X_i and Y_i are pair of points, $\bar{X}$ *and* $\bar{Y}$ are their means. The Pearson correlation is appropriate for showing linear relationship between variables. It is generally not invariant under monotonic non-linear transformation. For example the correlation of X and Y is not the same as the correlation of Ln (X) and Ln (Y).

Another disadvantage with the Pearson correlation is that in some situations the perfect negative correlation (-1) and perfect positive correlation (+1) cannot be attained (Iman and Conover, 1982) (Embricts et al, 1999). For example, consider that variable X is lognormally distributed with a mean of zero and a variance of 1 and Y is lognormally distributed with a mean of zero and a variance of sigma squared. For this case the minimum and maximum correlation coefficient are calculated and plotted against sigma as shown in Figure 3-1 which shows that for a value of sigma = 1 the minimum correlation coefficient is approximately -0.38 and for sigma = 2, the minimum correlation coefficient is approximately -0.15 and the maximum correlation coefficient is 0.67. This clearly shows the inability of the Pearson correlation to attain the perfect positive and negative correlation even if variables X and Y are perfectly correlated (either negatively or positively).

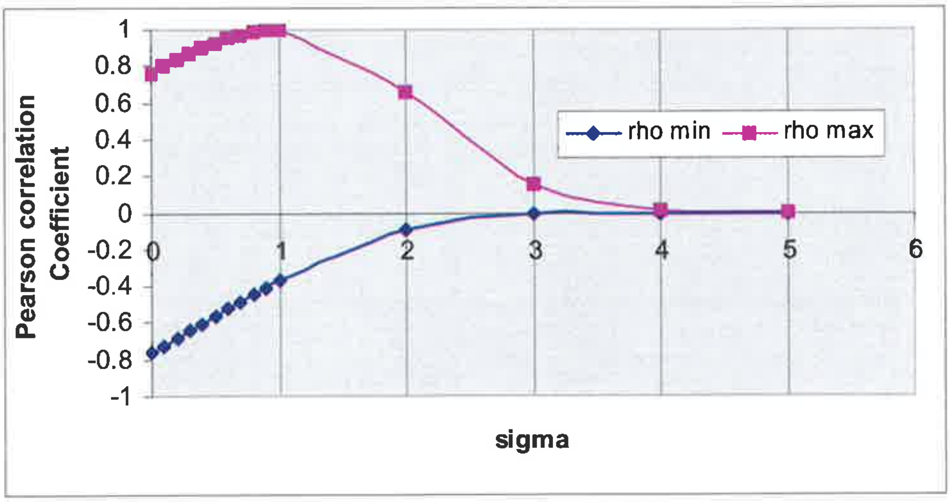


Figure 3 –1. Pearson correlation: minimum and maximum correlation (Embricts et al, 1999)

3.1.2 Spearman rank correlation

Spearman rank correlation coefficient was developed by Spearman in the early 1900's. It is calculated using the rankings of the values, not the actual values as in the Pearson correlation coefficient. Spearman's rank correlation is used as a measure of linear relationship between two sets of ranked data clusters. Spearman's rank correlation coefficient, like other classical correlation coefficients, will take on values between -1 and +1.

The Spearman correlation coefficient is calculated using the following equation

$$r_s = \frac{\sum(R_x - \bar{R_x})(R_y - \bar{R_y})}{\sqrt{\sum(R_x - \bar{R_x})^2}\sqrt{\sum(R_y - \bar{R_y})^2}} \tag{2}$$

where R_x *and* R_y are ranks of X *and* Y respectively; $\bar{R_x}$ *and* $\bar{R_y}$ are the means of the ranking variables.

Because the correlation is computed on the ranks, the method is known as a "distribution free" approach as there are no assumptions regarding the underlying distributions.

3.1.3 Kendall's tau correlation

Kendall's tau (τ) correlation is similar to Spearman's rank order correlation in its assumptions. It does not require any assumptions about the distribution nor does it require the relationship to be linear. As with Pearson and Spearman, the correlation coefficient takes on values between -1 and +1. Kendall's tau can be calculated from the following equation:

$$\tau = \frac{\#(concordant\ pairs) - \#(discordant\ pairs)}{\frac{n\,(n-1)}{2}} \tag{3}$$

where n is the number of samples. Concordant pairs means the number of pairs that are moving in the same direction and discordant pairs are those pairs that are moving in the opposite direction to each other.

A major shortcoming of the Pearson correlation is that it is not invariant under non-linear transformation. Both the Spearman and the Kendall approach overcome this issue. However, none of the three methods discussed so far are able to capture the dependency structure or correlation pattern. As it will be illustrated later, the copulas approach is able to capture and model the dependence structure.

The Pearson, Spearman and Kendall's tau correlations are data analysis techniques and as a stand-alone cannot be used to generate dependence in a stochastic environment. In order to model dependence in a stochastic form, technique that generate the stochastic spread of data as well as the trend are required. These techniques are the Envelope method, the Iman-Conover method (which uses the Spearman rank correlation) and the Copulas method (which uses the Kendall's tau correlation).

3.1.4 The Envelope method

The Envelope method, sometimes called the box method, is a useful technique for modelling partial dependencies. The method is well described by Newondorp and Schuyler (2000), Mian (2002) and Murtha (2000) and uses an "envelope" (see Figure. 3-2) around the data to capture the partial dependency.

In order to capture the dependence structure two lines are estimated Y upper and Y lower (Figure 3-2). For a simulation, a random X value is generated and the corresponding Y upper and Y lower values for that X are calculated. As shown in equation 4, together with a normalized value, they are used to generate the correspondence value of Y to an X.

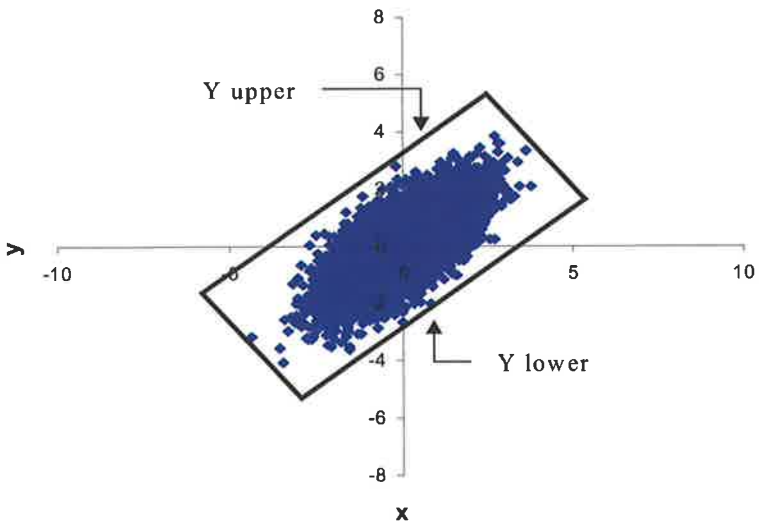


Figure 3-2. Envelope method

$$Y_x = (Y_{lower})_x + [(Y_{upper})_x - (Y_{lower})_x](Y_{norm})_x \tag{4}$$

where

$(Y_{lower})_x$ = lowest possible value of Y, given X

$(Y_{upper})_x$ = highest possible value of Y, given X

$(Y_{norm})_x$ = sampled value from dimensionless normalized distribution

This approach, which may lack mathematical rigour, is very flexible and can capture many structural patterns. Although the envelope method is conceptually simple, its implementation may seem somewhat tricky and involved.

3.1.5 Iman –Conover method

The Iman-Conover method (Iman and Conover, 1982)) is the algorithm used in the Monte Carlo Simulation software such as @Risk™ and Crystal Ball™. To model dependence between variables, the marginal distributions of each variable are defined. Then the Spearman's rank order correlation is calculated. For a simulation of correlated variable X and Y with a rank correlation matrix R^*: firstly, the method generates an independent score matrix for each variable using the van der Waerden scores. The use of van der Waerden scores is based on the inverse of the standard normal cumulative distribution function. Secondly, it will correlate between the two independent score matrices using the Cholesky decomposition to generate a correlated matrix M^* that is close to the rank correlation R^*. Thirdly, it generates independent random samples from X and Y and finally, rearranges X and Y so they will have the same rank correlation matrix as M^*.

3.1.6 Regression fitting

A simple regression approach finds the best-fit line using regression analysis. The regression calculation not only finds the slope and intercept of the line but also the standard error. The standard error is important for simulation because it defines the standard deviations of the errors. The residual is defined as the difference between the input data and the estimated data, based on the equation for the line

$$Y = mX + b + N(0,e) \tag{5}$$

where m is the slope of the regression line, b is the intercept and N (0, e) is the standard normal distribution with standard deviation e.

3.2. Copulas

The primary disadvantage of the dependency models discussed above is their lack of ability to capture and model the dependence structure. In this section an alternative is discussed in which a joint distribution is constructed using a copula.

Even though copulas are well known in the insurance and finance industries, their application in the petroleum industry is still in its early stages. According to extensive research, only two papers have been published about the applications of copulas in the petroleum industry. Armstrong et al (2004) investigated the value of an option to acquire new information in which Bayesian analysis was used to account for new information. The traditional framework of the Bayesian analysis is based on the joint normal distribution where the lower and upper tails are symmetrical. They proposed an alternative Bayesian analysis that is based on Archimedean copulas where the joint distribution does not have to be normal and there is flexibility to have a lower tail or upper tail dependence based on the specific type of copula.

Accioly and Chiyshi (2004) also discussed the suitability of using copulas to model dependence for non-Gaussian data. They presented two estimation procedures: nonparametric and semiparametric, to estimate copulas parameters. They used a sample of 188 exploratory offshore wells drilled in the Gulf of Mexico and explored the relationship between drilling duration and measured depth. They concluded that the Clayton copula reproduced the pattern of the original data and recommended the use of copulas to model the relationship between operating expenditure (OPEX) and production, since cost elements are production dependent.

This Chapter introduces the copulas approach to model dependence in the probabilistic reserves estimates. Chapter 6 will further elaborate on the potential benefit of the copula approach by comparing it with current correlation approaches, highlighting and discussing its relevance and advantages in petroleum applications. The first step in this discussion is to introduce the copulas, their construction and how to run simulations with them.

The essence of the copulas approach is that a joint distribution of random variables can be expressed as a function of the marginal distributions. Sklar first introduced the key theorem on copulas in 1959:

Let H be a joint distribution function with margins F_1 and F_2. Then there exists a copula C such that for all x, y, $F_1, F_2 \in R$:

$$H(x,y) = C\left[F_1(x), F_2(y)\right] \tag{6}$$

If the margins F_1 and F_2 are continuous, then C is unique; otherwise, C is unique in the range of $F_1 \times F_2$. Nelsen (1999)

This theorem and the characteristics of copulas are extensively discussed by Nelsen (1999). Sklar's theorem is completely general for any type of distribution where a joint distribution can be constructed using the copula function. One of the essential properties of copulas is that the copula separates the marginal distribution from the correlation and the copula itself can capture the dependence structure. This is an essential property of copulas. To show how this property makes copulas superior to the traditional dependence measures, Embrechts et al (1999) presented a fallacy, which stated, "Marginal distribution and correlation determine the joint distribution". This is true in the elliptical world in which normal distribution is one of elliptical distribution. By having two marginal distributions and a linear correlation, there is only one joint distribution that might fit. Yet, outside the elliptical distribution, there are many joint distributions that might fit. This can be seen clearly by simulating two variables with the same marginal distribution that is represented by a normal distribution with a mean of zero and a standard deviation of 1. These two marginal distributions have a linear correlation of 0.7 (Figure 3 -3).

The distribution on the left side of Figure 3-3 shows the case where the joint distribution is normal and the dependence structure shape is elliptical. While using the same marginal distribution and the same linear correlation, a different joint distribution can be constructed showing a completely different dependence structure (right side, Figure 3-3). This illustrates the limited ability of linear correlations to capture the dependence pattern. If the linear (Pearson) correlation was able to capture the shape of the dependence structure, we would expect both dependence structures in Figure 3-3 to be the same given that they have the same marginal distribution and the same linear correlation as input.

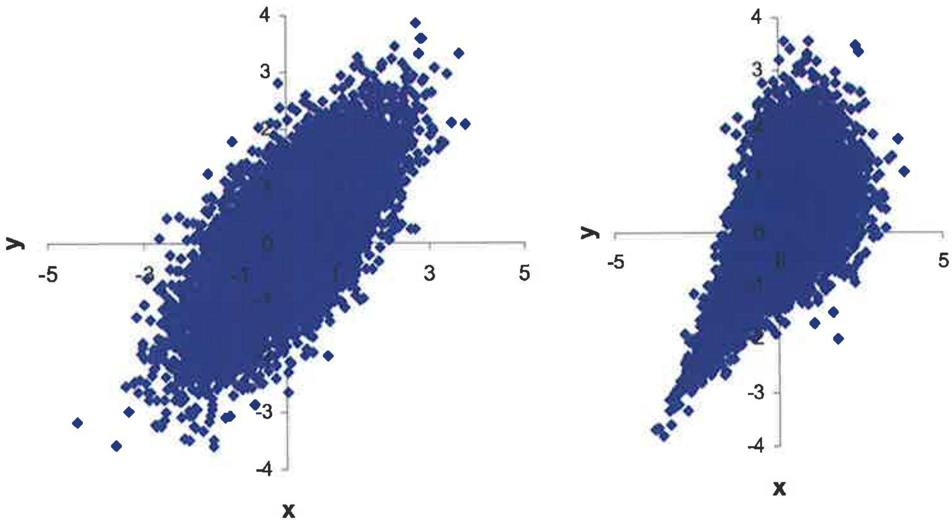


Figure 3- 3. Two identical distribution x~N (0, 1), y~N (0, 1) and same correlation = 0.7, but different dependence structure

There are different classes and families of copulas but, for this study, we will work with Archimedean copulas, because of their simplicity which are defined as:

$$C(u,v) = \phi^{[-1]}(\phi(u) + \phi(v)) \quad for\ u, v \in [0,1] \tag{7}$$

$C(u,v)$ is the copula function with u and v as uniform distributions, ϕ is the generator and $\phi^{[-1]}$ is the inverse generator. Genest and Mackay (1986) presented several properties of this class of copulas. The choice of generator determines the copula family (Table 3-1).

The Archimedean copulas have different families. The popular ones are:

Clayton Family (Clayton, 1978) This family has lower tail dependence for $\theta > 0$ and has the following function:

$$C_{clatyon}(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{\frac{-1}{\theta}} \tag{8}$$

Gumbel Family (Gumbel, 1960) This copula family has upper tail dependence for $\theta \geq 1$ and it has the following function:

$$C_{Gumbel}(u,v) = \exp\left\{-\left[(-\ln u)^{\theta} + (-\ln v)^{\theta}\right]^{\frac{1}{\theta}}\right\} \tag{9}$$

Frank Family (Frank, 1979). This copula family has a flexible θ that ranges $-\infty \le \theta \le \infty$ and has the following function:

$$C_{Frank}(u,v) = -\frac{1}{\theta}\ln\left(1 + \frac{\left(e^{-\theta u} - 1\right)\left(e^{-\theta v} - 1\right)}{e^{-\theta} - 1}\right) \tag{10}$$

Table 3-1. Archimedean copulas with their generators and ranges

Family	Generators ϕ (t)	Ranges of θ
Clayton 1978	$t^{-\theta} - 1$	$(0, \infty)$
Gumbel 1960	$(-\ln t)^{\theta}$	$(1, \infty)$
Frank 1979	$\ln\left(\frac{e^{t\theta} - 1}{e^{\theta} - 1}\right)$	$(-\infty, \infty)$

The relationship between the generator and the copula function is easily identified. If the generator $\phi(t)$ of Clayton copula from Table 3- 1 is substituted back into the general form of the Archimedean copulas (Equation 7), we get the Clayton copula (Equation 8). This means that the choice of the generator determines the copulas family.

Once the types of copulas are identified then the next step is to construct and simulate the copula.

3.3. Construction and simulation of Archimedean copulas

Genest and Rivest (1993) illustrated that Clayton, Gumbel and Frank copulas have the following cumulative distribution function $Kc\,(t) = P\{C(u, v) \le t\}$.

$$Kc\,(t) = t - \frac{\phi\,(t)}{\phi\,'(t^{+})} \qquad \textit{for any } t \textit{ in } [0,1] \tag{11}$$

where $\phi\ (t)$ is the generator and $\phi\,'(t^{+})$ is the derivative of the generator. Equation (11) is known as the parametric estimate. In order to compare and simulate a copula, the value of theta needs to be determined. Genest and Rivest (1993) proposed a relationship between the type of copulas and Kendall's tau and showed that through this relationship the value of theta can be determined. This relationship is expressed in the following theorem:

Let X and Y be random variables with an Archimedean copula C generated by $\phi\ (t)$ *, Kendall's tau of X and Y is given by:*

$$\tau_c = 1 + 4\int_0^1 \frac{\phi\,(t)}{\phi\,'(t)}\,dt \tag{12}$$

For example, for a Gumbel copula the Kendall's tau correlation can be expressed as a function of theta:

$$\tau_c = 1 + \frac{4}{\theta}\int_0^1 t\,\ln t\; dt = 1 - \frac{1}{\theta}$$

Hence by assuming a certain type of copula and calculating the Kendall's tau correlation using equation (3) the value of theta can be determined. Using equation (12) the value of Kendall's tau for each family is shown in Table 3-2.

Table 3-2. Kendall's tau values as a function of theta for each family and their ranges

Family	Ranges	Kendall's Tau
Clayton 1978	$(0, \infty)$	$\frac{\theta}{\theta + 2}$
Gumbel 1960	$(1, \infty)$	$1 - \frac{1}{\theta}$
Frank 1979	$(-\infty, \infty)$	$1 - \frac{4}{\theta}(D_1(-\theta) - 1)$

Note: D_1 is the Debye function of the first order $D_1 = \frac{1}{\theta}\int_0^{\theta} \frac{t}{e^{t} - 1}\,dt$

Once Kendall's tau is calculated and the value of theta is determined for each copula, then the next step is to compare the dependency methods. Frees and Valdez (1998) pointed out that to find the generator that fits the data one needs to compare the parametric estimate $Kc(t)$, by the procedure described earlier for generating the copula, with the nonparametric estimate $K(t)$. The nonparametric estimate is generated as follows according to Frees and Valdez (1998) :

1. Define the pseudo-observations

$$T_i = \left\{ \begin{matrix} number\ of\ (X_{1j}, X_{2j})\ such\ that \\ X_{1j} < X_{1i}\ and\ X_{2j} < X_{2i} \end{matrix} \right\} /(n-1)\ for\ i = 1, \ldots\ldots, n. \tag{13}$$

2. Construct the estimate of K as $K(t) =$ Proportion of $T_i \le t$

To construct the nonparametric estimate, a pseudo-observation (T_i) is estimated. This pseudo-observation captures the movement of each point with the movement of other points. It is possible to determine the correlation between variables in a similar approach as used in the Kendall's tau correlation method. Then construct a cumulative distribution function of the pseudo observations $(T_i's)$ for all the data points. This process generates the nonparametric estimates, which are then compared with the cumulative distribution function of the parametric estimates. For each generator chosen from Table 3-1, the value of theta is substituted into the parametric equation (Equation 11) to be compared with the nonparametric $K(t)$. We then choose the generator that best resembles the nonparametric results. The 'best' generator is determined by calculating the Minimum Distance (MD):

$$MD = \int [Kc(t) - K(t)]^2\ dK(t) \tag{14}$$

Another possible approach to determine the best generator is a Qauntile-Quantile (Q-Q) plot, which plots the cumulative distribution function of the nonparametric estimate against the cumulative distribution function of the parametric estimates.

Once the optimal copula has been determined, it is used to generate the desired correlated marginal distributions U and V in the Monte Carlo Simulation model. Several types of simulation algorithms have been proposed in the literature. A general simulation algorithm was discussed by Mari and Kotz (2001):

1. *Generate two uniform and independent random variables u and q*
2. *Calculate* $v = C_u^{-1}(q)$, *where* $C_u = \dfrac{\partial C(u,v)}{\partial u}$.

The second step is implemented by taking the derivative of the copula function with respect to u. The next step is to find the inverse of the derivative C_u^{-1}. In general this procedure works well where C_u^{-1} has a closed form expression and where the inverse function can be solved analytically. This algorithm will work well for the Clayton but not for the Gumbel copula because it has no closed form expression for the inverse. To overcome this problem, another simulation procedure can be used as suggested by Nelsen (1999):

1. *Simulate two independent random uniform variables s and q on* $[0,1]$.
2. *Set* $t = Kc^{-1}(q)$, *Where Kc is the distribution function of C (u, v)*
3. *Set* $u = \phi^{-1}(s\phi(t))$ *and* $v = \phi^{-1}((1-s)\phi(t))$

The second step in this simulation uses a numerical root finding method to solve for the inverse of the parametric estimates. Once the value of t is found, it is substituted into the equations in the third step to derive the desired correlated variables u and v. This simulation is powerful and can be used to solve for all three type of copulas; Gumbel, Frank and Clayton.

This section has introduced the copulas methods together with their construction and their simulation algorithms. A detailed explanation of the comparisons, construction and simulation of the copulas methods will be given in Chapter 5.

3.4 Summary

In this Chapter, the current techniques to model dependencies used in the oil and gas industry have been introduced. This Chapter shows that the Pearson (linear) correlation cannot capture the dependence structure. Furthermore, this section has discussed Regression fitting, the Envelope method and the Iman-Conover method, which use the Spearman rank correlation as an input. In addition, the copulas approach has also been explained, which has the advantage of capturing the dependence structure.

The questions to be asked next are: How do the Regression Fitting, Envelope and the Iman-Conover methods capture the dependence structure and how do they differ from the copulas approach? What is the impact of this dependence structure and does it really matter? These questions pertain to the second objective of this research and will be investigated by testing these techniques at the reserves level. The answer to these questions will be presented in Chapter 6 followed by conclusions.

CHAPTER 4

Functional dependence and systems modelling at project & portfolio levels

4. Introduction

The second Chapter showed that stochastic models yield better results than deterministic ones and the value of integration can be achieved by the modelling of dependencies and interactions. The third Chapter presented statistical dependence methods for application at the reserves level. This Chapter focuses on explaining the third and fourth objectives of this research and shifts the emphasis from statistical dependency to functional dependency. This section will discuss the value of modelling dependencies at the project and portfolio levels. The objectives of this Chapter are to:

1. Explain the third objective that this research proposes, which is the impact of functional dependence on a development decision at a project level. This is illustrated by:
    a. Showing what functional dependence means
    b. Defining what systems and sequential approaches are

   c. Showing how both systems and sequential approaches treat functional dependence.

2. Explain the fourth objective of this research. That is the assessment of the impact of dependencies and interactions at the portfolio level.

Terms used in this Chapter that require definition are:

Functional dependency is a dependency that relates two variables. For example, y = f (x), then x has a functional dependence with y or x has a functional interaction with y.

Sub-system is defined as more than two variables that have functional dependency on each other.

System is a set of interdependent sub-systems.

4.1. Systems approach

The holistic, stochastic integrated asset model discussed in Chapter 2, termed the systems approach, comprises different sub-systems which are related to and interact with each other. In order to capture the uncertainty in the model, a Monte Carlo Simulation (MCS) is carried out for the whole system. Each iteration in a simulation takes samples from all sub-systems (Figure 4-1). Each variable is treated as stochastic and is represented by a probability distribution function.

The systems approach views a development decision holistically as one integrated system where a sample of reserves is estimated from the reserves model and then used as an input into the production model to yield a production profile, which in turn is used as an input into the economic model to calculate the Net Present Value of the decision. This process is repeated until the NPV distribution is generated based on the number of iterations chosen.

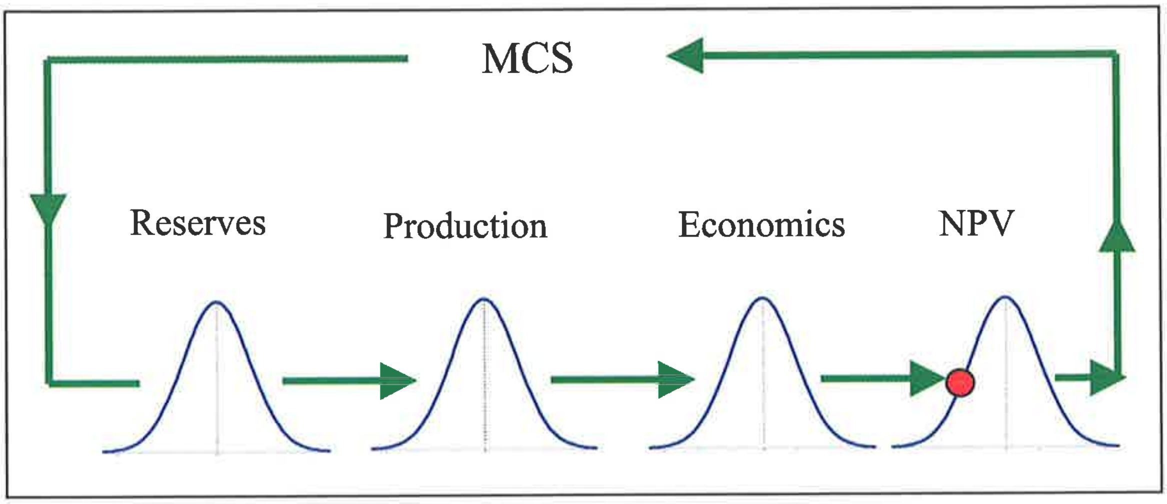


Figure 4-1. Systems approach

4.2. Sequential approach

A sequential approach is defined as a set of sub models that are simulated sequentially and the result of each sub-model is carried to the next sub-model which is again simulated in isolation with its result carried to the next sub-model (Figure 4-2). This is known as the reductionist view, which decomposes a problem into smaller parts, and is the traditional sequential oil and gas industry practice. A key difference between the sequential approach and the common industry practice is that the current oil and gas industry approach assumes some sub-models to be stochastic and others to be deterministic. In this research, the sequential approach is used assuming that all sub-models of petroleum development decisions are stochastic. To analyse a hypothetical development decision using the sequential stochastic approach, a Monte Carlo Simulation was run on the reserves model to determine a distribution of technical reserves, which was then used as an input into the production model to calculate a distribution of the production profile. The production profile generated was used as input into the economic model, which incorporates costs and prices to calculate the NPV. The key difference between the systems approach and sequential approach is that in the systems approach the individual samples are used in the next step, while in the sequential approach the whole distribution is carried from one sub-model to the next.

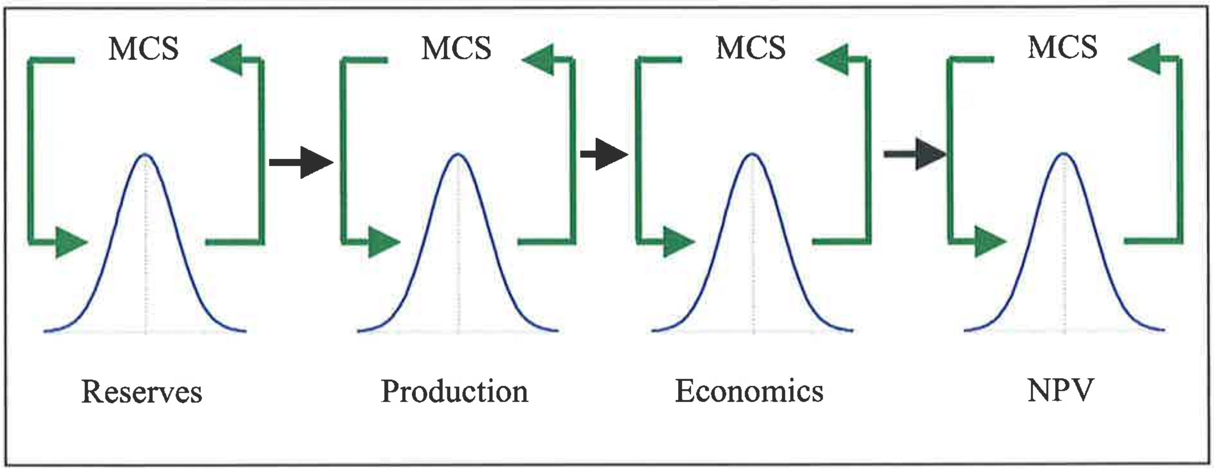


Figure 4- 2. Sequential approach

4.3 Treatment of dependencies and interactions in the two approaches

To show how the systems and sequential approaches treat functional dependencies, two simple examples will be considered, one without and the other with functional dependence between sub-systems.

4.3.1 Example 1: No functional dependence

Let us assume that there are two systems as shown in Figure 4-3: system M and system S. System M has three variables X, Y and R, where R is a function of X and Y. System S has three variables R, N and P where P is a function of R and N. R output in system M is used as an input into system S as in example 1 (Figure 4-3). Note that we cannot say that R has a functional dependence with itself; dependence has to be between one variable and another. Clearly there is no functional interaction between the two systems (M and S).

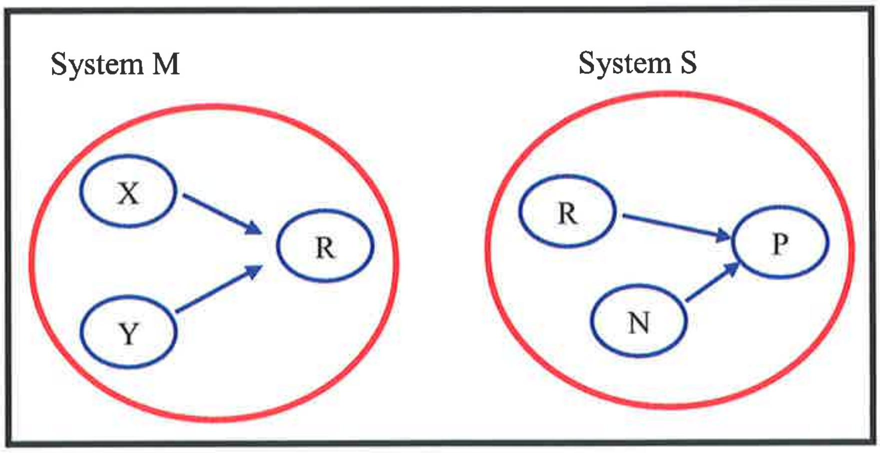


Figure 4-3. Example 1: no functional dependence between systems

The functional forms in the example have arbitrarily been assumed to be

$$R = X * (-Ln(Y))$$

$$P = \frac{R^2}{N}$$

where the variables X, Y and N are assumed to have triangular distributions defined as follows:

Table 4-1. Input values for example 1

Variables	Values		
	Minimum	Most likely	Maximum
X	5	8	10
Y	0.1	0.3	0.5
N	20	40	60

The sequential approach views example 1 as two sub-models with two independent Monte Carlo Simulation runs used to characterise them (Figure 4-4). This approach executes the first simulation to generate an output distribution for R and then uses that output distribution in a second simulation together with input distribution of variable N to generate the output distribution P.

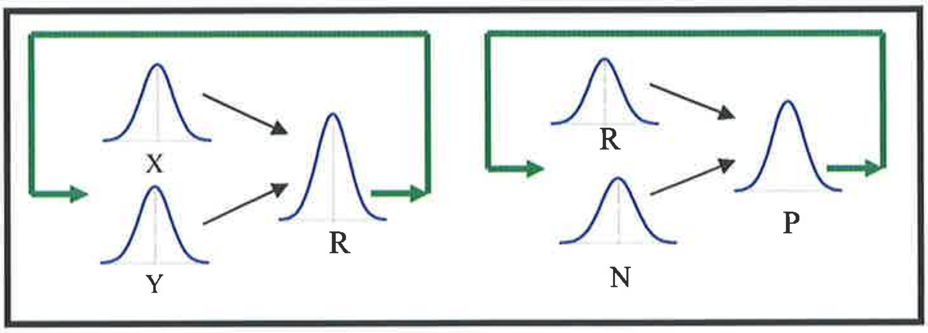


Figure 4-4. Sequential approach for example 1

The systems approach, on the other hand, views both system M and system S as one integrated system and uses a single integrated simulation. The process, shown in Figure 4-5, is as follows: samples of X and Y are taken from their input distributions and used to calculate a value of R. This output value of R (not its distribution) is taken with a sample from the input distribution of N to calculate a possible value of P. This process is repeated for a large number of iterations to generate the output distribution for P.

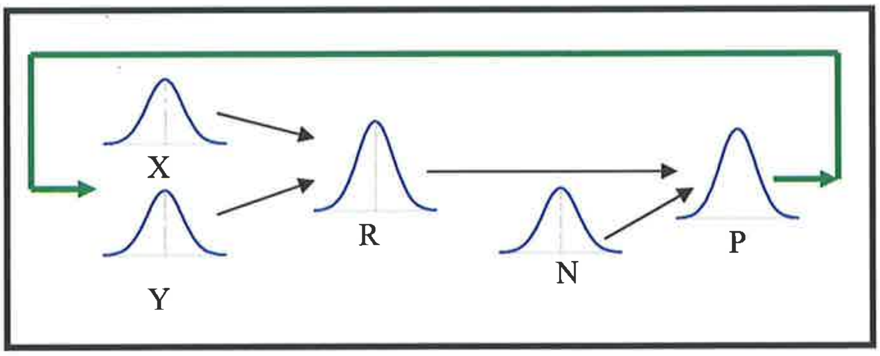


Figure 4-5. Systems approach for example 1

4.3.2 Example 2: With functional dependence

The second example has the same input distributions for X and Y as in example 1. The only difference is that the second example assumes P to be a function of R and Y, instead of R and N leading to the deletion of N (Figure 4-6). The functional form of R is the same as in example 1; the arbitrary functional form of P is $P = \frac{R^2}{Y}$

The difference between the two examples is that in example 1 there is no interaction or dependence between system M and system S, while in example 2 there is an interaction between the two systems, as shown in Figure 4-6.

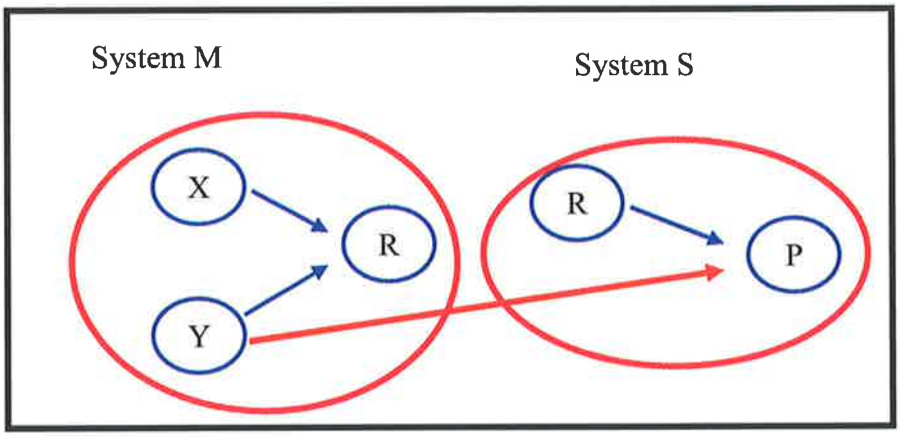


Figure 4-6. Example 2: With functional dependence between systems

The results from examples 1 and 2 are shown in Table 4-2, which shows that for example 1 the systems and sequential approaches yield the same values for the mean and standard deviation of the final distribution for P.

The results from example 2, however, show a significant difference between the two approaches. The mean and the standard deviation for the systems approach are higher than the sequential approach by 17% and 70% respectively. The systems approach yields a higher value of the standard deviation because it captures the functional relationship between variables whereas the sequential approach ignores them. In addition, the systems approach yields a higher value of the mean because of the non-linearity of the functional relationships as will be demonstrated in chapter 7.

Table 4-2. Result of examples 1 and 2

Experiments	**Approaches**	
	Sequential	Systems
Example. 1	Mean P = 2.563 Stand Dev = 1.62	Mean P = 2.563 Stand Dev = 1.62
Example. 2	Mean P = 355 Stand Dev = 241	Mean P = 415 Stand Dev = 410

The difference between the results of the sequential and systems approaches in example 2 is due to incorporation of functional dependencies in the systems approach. This can be observed by plotting the Y variable against P output (Figure 4-7). The systems approach honours the negative functional dependency between Y and P whereas the sequential approach clearly ignores this functional dependency, as shown in Figure 4-8. The conclusions presented in this section are for arbitrarily chosen functional relationships. Logic suggests that they should be valid regardless of the form of the functional relationship and were found to be valid for various other functional forms tested.

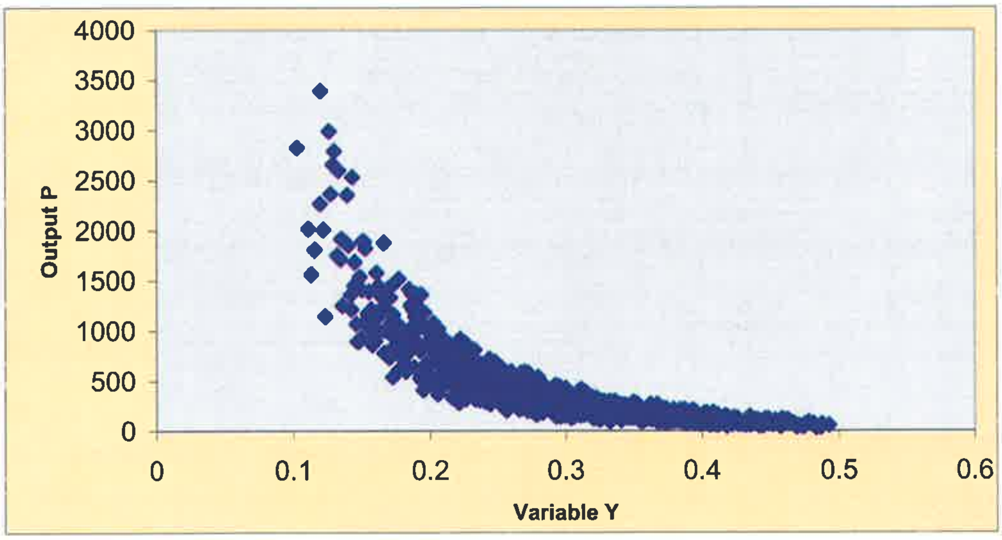


Figure 4-7. Systems stochastic model showing functional dependence

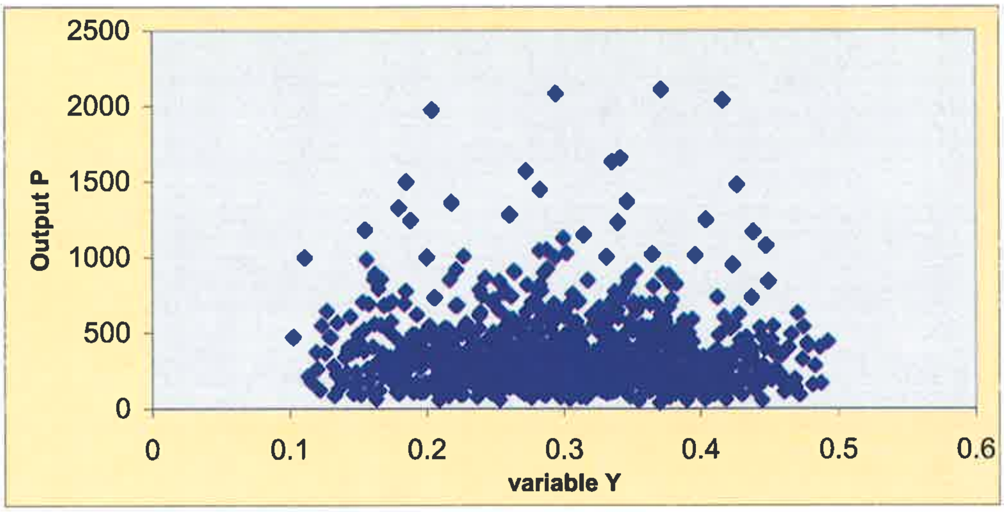


Figure 4-8. Sequential stochastic model showing lack of functional dependence

The impact of the functional dependency can be seen indirectly presenting a statistical dependence between Y and P. This statistical dependence could be measured to show its existence. For the systems approach, the Pearson correlation is -0.7, while for the sequential approach the Pearson correlation is 0, this can clearly be seen from Figure 4-7 and Figure 4-8.

The impact of the holistic stochastic integrated asset model, systems approach, compared to the sequential approach on dependencies and interactions has been investigated on a hypothetical offshore development decision. The results of this investigation are presented in Chapter 7.

4.4 Dependencies and interactions at the portfolio levels.

Harry Markowitz is known as the father of the portfolio theory. His work is built on the concept of diversification among stocks (Markowitz, 1952). The key idea of his work is that the interaction of the stocks is more important at the portfolio level than the value of the stock alone. This is captured in his famous quote "Don't tell me what a certain stock will do …….. tell me what will it do for my portfolio". The same analogy applies to petroleum projects. How projects interact with each other is more important than the evaluation of single project alone. This means that a single project should not be valued alone but the evaluation should be based on how it contributes to the firm's portfolio maximization.

The fourth objective of this research is intended to build on the third objective of investigating the impact of systems and sequential approaches and carrying this difference onto the portfolio level, thus showing how the difference in the systems and sequential approaches impact decision-making at the portfolio level. In other words, this section asks the question: whether the efficient frontier generated using the systems approach to model projects is different from the efficient frontier generated using the sequential approach to model the same set of projects. The objective is to show the impact of the difference between the systems and sequential approaches at the portfolio level.

4.5. Current Portfolio modelling: what is missing?

The Markowtiz model recognizes the importance of interaction among projects in the optimisation of the firm's portfolio rather than the value of the project alone. This model focuses on the interaction among projects as shown in Figure 4-9. Ball and Savage (1999) discussed the main sources of interaction and dependencies among projects as follows:

Places: The NPV's of two projects in a similar area, meaning similar geological setting, will produce positively correlated NPV's, which implies, according to the portfolio theory, less diversification. While two projects in different geological settings may be negatively correlated and hence lead to a more diversified portfolio.

Prices: Oil prices in the world market tend to be the same and therefore positively correlated, which leads to less diversification. However, gas prices are not or at most weakly correlated with each other or with oil prices. So a portfolio with two oil projects will have less diversification compared to a portfolio with two projects when one is oil and the other is gas.

Profiles: The timing of the cash flows and production profiles are also important in the construction of the portfolio. Two positively correlated cash flows from two projects will lead to less diversification compared to two cash flows that are less correlated. A lower correlation between cash flow profiles leads to more stable cash flows and hence more diversification.

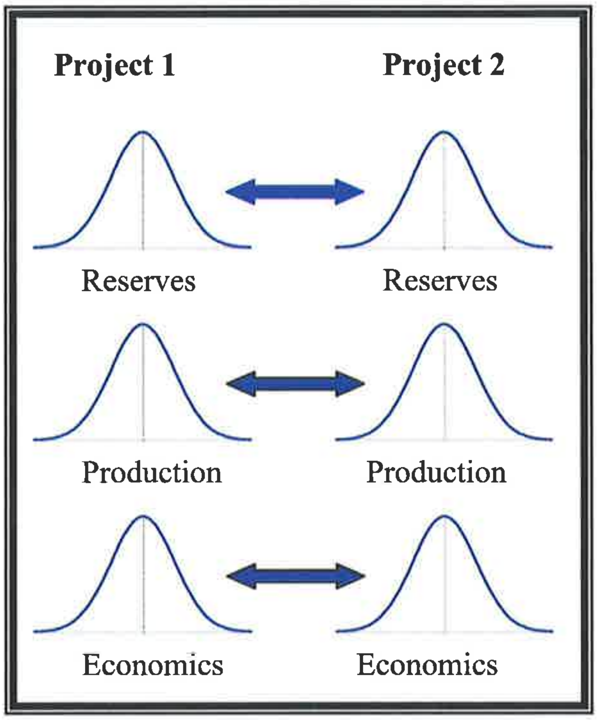


Figure 4-9. Portfolio selection: Inter-dependence among projects

Politics: Political uncertainties are also significant in portfolio characterization. For example, an Iraqi firm with two projects in Iraq will have less diversification compared to one project in Iraq and another in Australia. Different political climates create a more diversified portfolio model then similar ones.

The portfolio model clearly emphasizes the dependencies and interactions among projects. This study argues that portfolio models should not only recognize the impact of inter-dependence, but also the impact of intra-dependence within each project as well. Furthermore, it is author's belief that the combination of both inter- and intra-dependence will lead to better portfolio optimisation models, as shown in Figure 4-10. The idea is to move away from Figure 4-9 that considers only inter-project dependence to Figure 4-10, which considers the impact of both intra- and inter-project dependence. The results of this investigation are discussed in Chapter 8.

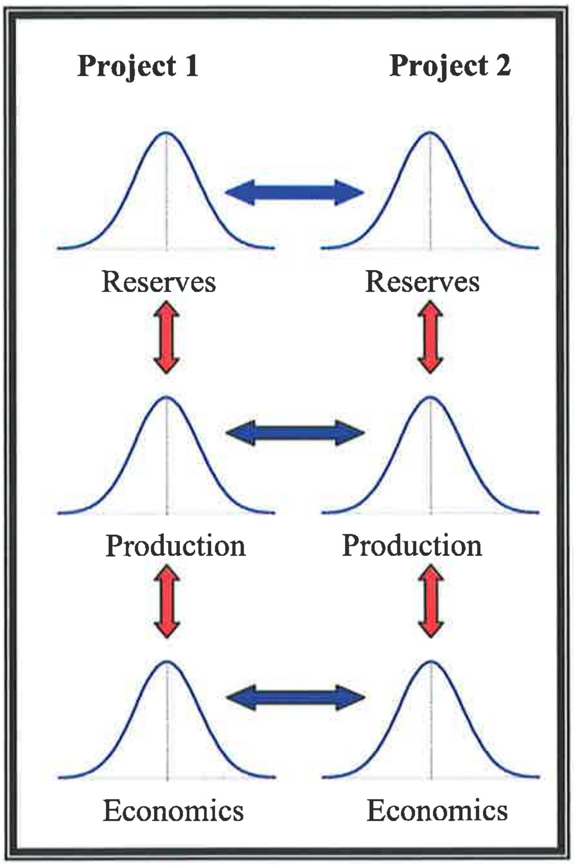


Figure 4-10. Portfolio selection: intra- and inter-dependence within project variables and among projects

4.6. Summary

This Chapter has discussed the third and fourth objectives of this research. Concerning the third objective, it has explained the difference between a sequential and a systems approach. In addition, it illustrated how both approaches can capture the functional dependence when an interaction exists between two systems. The second part of this Chapter investigated the impact of including intra- and inter-project dependence using the systems approach compared to the conventional portfolio approach of including inter-project dependence only. The focus of the two objectives is to show the impact of dependencies and interactions at the project as well as at the portfolio level.

CHAPTER

5 Methodology: Construction of copulas and systems approach

5. Introduction

Chapter 3 introduced the oil and gas industry techniques to model statistical dependence. In addition, the copulas method was introduced. The second objective of this research is to explore the impact of the statistical dependence methods on reserves estimates. Since the copulas method is not traditionally used in the petroleum industry there is a need to outline the detailed steps of how to construct and simulate copulas.

In order to investigate the impact of functional dependence at the project and portfolio level, Chapter 2 and 4 illustrated the need for a holistic stochastic integrated model called the systems approach. This model should therefore be able cover all petroleum components from reserves to economics. In addition, the systems approach model should be able to capture both above and belowground uncertainties with simple designs that avoid unnecessary details. This Chapter is divided in two sections; the first section will introduce the methodology and the equations used to develop, construct and simulate the copulas. The second section will introduce the methodology used for the holistic stochastic integrated model, called the systems approach.

5.1 The copulas methodology

This section is divided in to two main parts. The first part focuses on how to construct copulas that could be used to describe stochastic dependency. The second part focuses on comparison of copulas with one another and with other dependency description methods.

5.1.1 The construction and simulation of copulas

In Chapter 3, two simulation procedures were introduced to construct and simulate copulas. This section shows step by step how to build and simulate these two procedures.

5.1.1.1 General analytical simulation technique

Analytical solution means being able to solve for one variable in terms of others without iteration. A general simulation procedure introduced in Chapter 3 by Mari and Kotz (2001) which uses the distribution function of a copula is as follow:

- *First step: Generate two random variables s and q independent and uniform on (0,1)*
- *Second step: Calculate $v = C_u^{-1}(q)$. The pair (s, v) has the desired distribution. Knowing that $C_u = \dfrac{\partial C(u,v)}{\partial u}$.*

The (s, v) is the desired correlated marginal distribution where, for example, s could be the area and v could be the reservoir thickness. This section will show how this procedure works using the Clayton copula as an example. The same procedures can be used for the Frank copula, but not for the Gumbel copula as will be shown later. The following are the detailed steps for using general analytical simulation technique to simulate the Clayton copula:

1. The distribution function for the Clayton copulas is $C(u,v)=(u^{-\theta}+v^{-\theta}-1)^{\frac{-1}{\theta}}$, this is equation (8) in Chapter 3.
2. Find the derivative of the copulas with respect to u

   $C_u = \dfrac{\partial C(u,v)}{\partial u}$, which is equal to

$$C_u = \frac{\partial C(u,v)}{\partial u} = (u^{-\theta}+v^{-\theta}-1)^{\frac{-1}{\theta}-1} u^{-\theta-1}. \tag{1}$$

3. Set $q=C_u$, since the inverse is expressed in term of q. This is an important step to relate the correlated variable v in term of the uniform variable q as shown in the second step in the general analytical simulation technique where, $v=C_u^{-1}(q)$. Therefore equation (1) is set as:

$$q=(u^{-\theta}+v^{-\theta}-1)^{\frac{-(1+\theta)}{\theta}} u^{-(\theta+1)} \tag{2}$$

4. Solve for the inverse of C_u and therefore solve for $v=C_u^{-1}(q)$. Looking at equation (2), it can be seen that v could be solved analytically without using numerical root finding methods.
5. Simplify equation (2) by raising both sides by the power of $-\dfrac{\theta}{1+\theta}$ and solve for variable v

$$v^{-\theta}=\left(\frac{q^{\frac{-\theta}{1+\theta}}}{u^{\theta}}-u^{-\theta}+1\right) \tag{3}$$

6. Raising both sides of equation (3) by the power of $\dfrac{-1}{\theta}$. The generated equation for v variable with Clayton copula is $v=\left(\dfrac{q^{\frac{-\theta}{1+\theta}}}{u^{\theta}}-u^{-\theta}+1\right)^{\frac{-1}{\theta}}$ (4)

For simulation the value of theta (θ) is determined for Clayton copula as discussed in Chapter 3, Table 3-2, with it's relation with Kendall's tau through the following equation: Kendall's tau $=\dfrac{\theta}{\theta+2}$, by calculating the Kendall's tau a value of theta can be determined.

Once the value of theta is determined and two random variables s and q are generated then Equation (4) generates the desired pair(s, v), which will provide the desired Clayton copula for area and thickness as an example. Unfortunately this procedure only works for copulas with analytical solutions or copulas that do not require a numerical root finding method. It does not work for copulas such as the Gumbel copula. Furthermore this can be seen if we try to derive the derivative $C_u = \dfrac{\partial C(u,v)}{\partial u}$ for the Gumbel copula.

5.1.1.2 A numerical root finding simulation technique

Since an analytical solution cannot be found for the Gumbel copula, another simulation algorithm that is capable of handling numerical root finding is suggested by Nelsen (1999):

Step 1: *Simulate two Uniform random numbers $s \sim U(0, 1)$ and $q \sim U(0, 1)$*

Step 2: *Set $t = K_C^{-1}(t)$, K_C is the copula distribution*

Step 3: *Set $u = \phi^{-1}(s\,\phi(t))$, $v = \phi^{-1}((1-s)\,\phi(t))$*

To illustrate how to construct this algorithm a Gumbel copula will be used as an example. However, this method can also be applied to the Clayton and the Frank copulas even though they can be solved analytically. The reason for using the general analytical simulation for the Clayton and Frank copulas is simplicity. Detailed explanation of the steps required in the numerical simulation follows:

Step 1: *Simulate two Uniform random numbers s ~ U (0, 1) and q ~ U (0, 1)*

The preliminary steps required prior to generating two uniform distributions are:

1. Calculate the Kendall's tau using the following equation:

$$\tau = \frac{\#(concordant\ pairs) - \#(discordant\ pairs)}{\frac{n\,(n-1)}{2}}.$$

   This is the same as equation (3) in Chapter 3.

2. Next calculate the value of theta (θ), from Table 3- 2 in Chapter 3. When using the Gumbel Copula there is a relationship between Kendall's tau and theta, $Kendall's\ tau = \tau = 1 - \frac{1}{\theta}$. Once the value of Kendall's tau is calculated then the value of θ can be estimated.

Step 2: *Set* $t = K_C^{-1}(t)$, K_C *is the copula distribution*

For the second step the following equations are needed.

The parametric estimate distribution is $K_C = t - \frac{\phi\,(t)}{\phi\,'(t)}$ (5)

The generator for the Gumbel copulas is $\phi\,(t) = (-\ln(t))^{\theta}$, Table 3-1. (6)

The first derivative of the generator $\phi\,'(t) = -(\ln(t))^{\theta-1}\frac{1}{t}$. (7)

Substitute (6) and (7) into (5) and we get $K_C = t - \frac{t\ln(t)}{\theta}$. (8)

The next step is to solve for the inverse of K_C which means to solve for t, $t = K_C^{-1}\ (t)$. It is known from algebra that equation (8) cannot be solved analytically to find t. Therefore, numerical root finding must be used to solve for t. Although, there are many methods for numerical root finding, the Newton method was chosen.

Other methods might work as well. The Newton method can be derived from the first order Taylor expansion or from two points and a tangent line.

The Taylor expansion is expressed as $f(x) = f(x_i) + f'(x_i)(x_{i+1} - x_i)$.

In order to solve for the root x, the function is set $f(x) = 0$, and the formula to calculate the value of x_{i+1} using the Newton method is

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \quad (9)$$

Using equation (9), to solve for the numerical root, the following steps and criterion are required:

1. An initial guess of x_i
2. The first derivative of $f(x)$ which is $f'(x)$
3. A convergence criterion, say $f(x) < 0.0000000001$.

The procedures is as follows,

1. Start with the initial guess of $x_i = x_0$ and calculate $f'(x)$.
2. The next step is to find x_1, using is $x_1 = x_0 - \dfrac{f(x_0)}{f'(x_0)}$.
3. Then $x_2 = x_1 - \dfrac{f(x_1)}{f'(x_1)}$ and $x_3 = x_2 - \dfrac{f(x_2)}{f'(x_2)}$ and so on.
4. Continue this process until the value of $f(x) < 0.0000000001$ is reached.

Now using equation (8) we include the q generated by uniform distribution of U (0, 1) in the first step as follows $K_C = t - \dfrac{t\ln(t)}{\theta} - q$. (10)

Based on the root finding method we need the first derivative of K_C with respect to t which is $K'_C = \dfrac{-\ln t}{\theta} - \dfrac{1}{\theta} + 1$. (11)

Now it is possible to solve $t = K_C^{-1}(t)$ for the Gumbel copula, with an initial guess of $x_i = t_0 = 0$, using equation (10) and (11) and substituting them in equation (9). This process can be done in Excel or with Visual Basic Application (VBA), the latter being easier and faster.

Step 3: *Set* $u = \phi^{-1}(s\,\phi(t))$, $v = \phi^{-1}((1-s)\,\phi(t))$

The final step is to get the desired correlated marginal distribution of u and v

Set $u = \phi^{-1}(s\,\phi(t))$ and (12)

$v = \phi^{-1}((1-s)\,\phi(t))$. (13)

For the Gumbel copula, the generator is $\phi(t) = (-\ln(t))^{\theta}$ as given in equation (6). Numerical root finding is used to solve for t and find ϕ^{-1}, which is given by

$$\phi^{-1}(t) = \exp(-t)^{\frac{1}{\theta}} \tag{14}$$

Now equation (14) can be substituted back in equation (12) with s generated from the first step and t estimated from second step, a value of u is estimated, and then repeating the same process will solve for v. This will generate the desired correlated marginal distribution (u, v) using the Gumbel copula. An example, Figure 5-1, shows the correlation between area and net thickness generated by using a Monte Carlo Simulation (MCS) of 1000 iterations for the Gumbel copula.

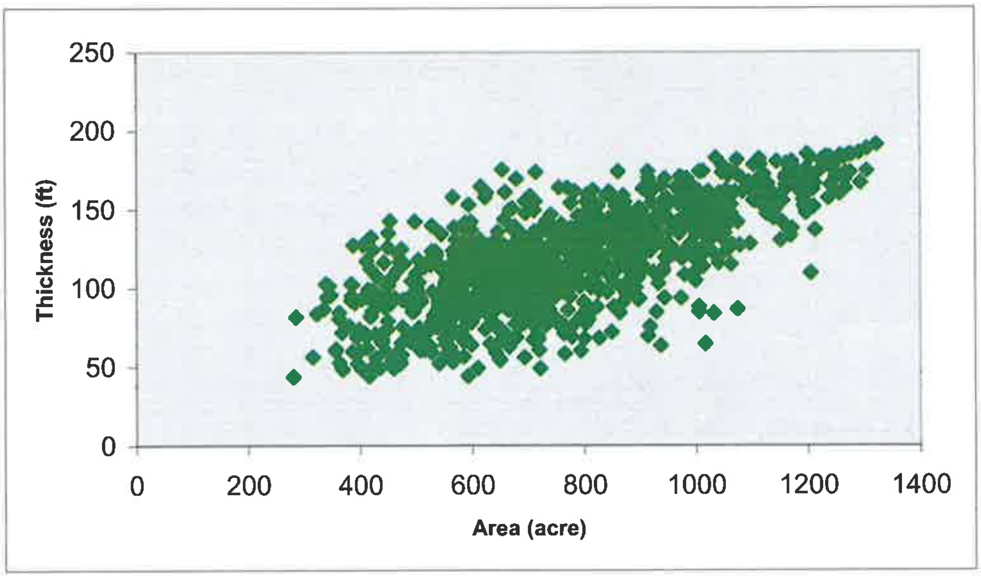


Figure 5-1. Gumbel simulation for area and thickness for 1000 iteration

5.1.2 The comparisons of copulas and other methods

Once the simulation procedure, described above, is set up for the copula, then a MCS can generate correlated distribution resulting in the generation of a plot such as Figure 5-1. The remaining question is how to determine which dependency method can best reproduce the original data? The answer lies in comparing the parametric and non-parametric estimates. Reviewing the parametric and non-parametric estimates is suggested as a way to compare copulas and to compare copulas with other dependency methods, such as Regression fitting, the Envelope method and the Iman-Conover method.

5.1.2.1. Non-Parametric Estimates

According to Frees and Valdez (1998) the nonparametric estimate, $K(t)$, is determined as follows:

1. First, define the pseudo-observations

$$T_i = \begin{Bmatrix} number\ of\ (X_{1j}, X_{2j})\ such\ that \\ X_{1j} < X_{1i}\ and\ X_{2j} < X_{2i} \end{Bmatrix} /(n-1)\ for\ i = 1, \ldots\ldots, n. \quad (15)$$

2. Second, construct the estimate of K as $K(t) = proportion\ of\ T_i \leq t$

Pseudo-observation captures the movement of the other points with respect to a reference point. This reference point is then changed and the movement is captured according to the new reference point. A simple example is provided to demonstrate how the pseudo-observation measure is calculated. Let's assume two variables, X_1 and X_2 are given with three paired points (n = 3) as shown in Table 5-1. This set of data could be the original data or it could be the data generated from the simulation of the dependency method.

To calculate the pseudo-observation measure the following assumptions are made:

- If the movement of a point relative to a reference point is upward (positive direction) then $T_p = 1$.
- If the movement of a point relative to a reference point is downward (negative direction) then $T_p = 0$.

The two assumptions can be mathematically represented as an IF logic function from Excel spreadsheet as follows:

IF(logical_test ,value_if_true ,value_if_false) as

IF (movement direction is upward, upward (positive)= 1, downward (negative)= 0)

Table 5-1. Input data for pseudo observation example

Sample (n)	Variable X_1	Variable X_2
1	2	3
2	4	5
3	7	8

The following steps are then carried out:

First step: first point is the reference point, so point 1, (2, 3), will be the first reference point and T_1 will capture the movement of other points with respect to this point (Figure 5 –2).

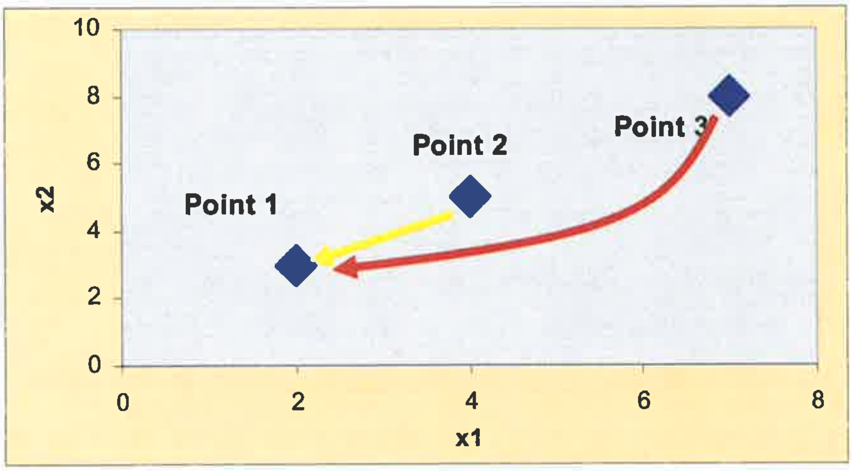


Figure 5- 2. Pseudo-observation for the first point

The movement from point 2 to point 1 (T_{21}) is captured through the logic function as follows: IF (AND (4<2, 5< 3), 1, 0), the result of this is $T_{21} = 0$. The logic test of (4<2, 5< 3) is only used to show how equation (15) works. Since 4 is not less than 2 and 5 is not less than 3, this is a negative direction movement and thus the result of $T_{21} = 0$. Also Figure 5-2 shows that the movement from point 2 to point 1 is downward (negative direction) and therefore $T_{21} = 0$.

Then the movement from point 3 to point 1 (T_{31}) is captured through the logic function as follows: IF (movement upward, 1, 0), the result of this is $T_{31} = 0$.

Therefore $T_{p1} = T_{21} + T_{31} = 0 + 0 = 0$.

And, $T_1 = T_{p1} /$ (n-1) = 0/ (3-1) = 0.

The (n-1) factor is an adjustment for a number of points. The value of $T_1 = 0$ means that the movement from the other points (point 3 and point 2) to the reference point 1 is negative or in the opposite direction. This is how the pseudo-observation estimate captures the movements of points according to each other's position.

Second step: now the second point is set as the reference point, so point 2, (4, 5), will be the reference point and the value for T_2 will capture the movement of the other points with respect to this point as shown in Figure 5-3.

The movement from point 1 to point 2 (T_{12}) is captured through the logic function as follows: IF (movement upward, 1, 0), the result of this is $T_{12} = 1$. This a gain be seen from Figure 5-3 as the movement from point 1 to 2 is upward.

Then the movement from point 3 to point 2 is captured through the logic function as follows: IF (movement upward, 1, 0), the result of this is $T_{32} = 0$ (downward movement).

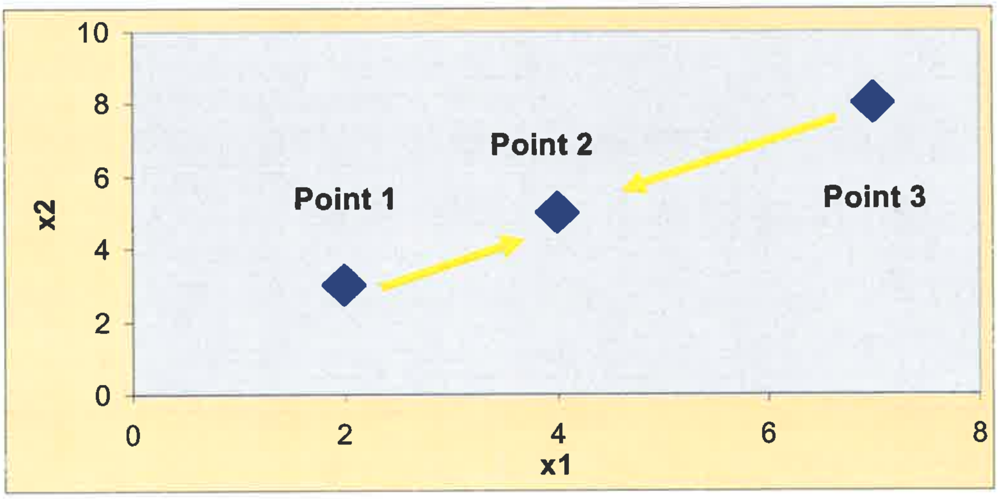


Figure 5-3. Pseudo-observation for the second point

Therefore $T_{p2} = T_{12} + T_{32} = 1 + 0 = 1$,

and $T_2 = T_{p2} / \text{(n-1)} = 1/(3\text{-}1) = 0.5$.

Repeating the same process for the third point we will arrive at $T_{p3} = 2$ and $T_3 = 1$.

The summary of the all the pseudo-observations is given in Table 5-2.

Table 5- 2. Results of the pseudo observation

Sample (I)	Variable X_1	Variable X_2	T_i
1	2	3	0
2	4	5	0.5
3	7	8	1

Once the values of the T_i's are calculated, then a cumulative distribution function (CDF) is developed for the T_i values. The IF.. AND logic function are used here to show how this approach works, but in order to do this a Visual Basic Application (VBA) code was developed which made this process easier to carry out in Excel. This is very useful especially when a simulation is run for 1,000 or 10,000 iteration which means repeating the above procedures for 1,000 or 10,000 points instead of the 3 points shown in the above example. This process will generate the CDF of T_i's for the original data or for the data generated through the simulation of dependency methods. The CDF of T_i's is called the non-parametric estimate of $K(t)$.

5.1.2.2. Parametric Estimates

Once the T_i's are developed for any two variables, then the parametric estimate is compared with the non-parametric estimate for any copula. The parametric equation is: $Kc\left(t\right) = t - \dfrac{\phi\left(t\right)}{\phi^{\prime}\left(t\right)}$ *for any t in the range of* [0, 1] (16)

Once the T_i's are developed we substitute each value of T_i for t in equation (16) and then by substituting for the generator $\phi\left(t\right)$ and its derivative $\phi^{\prime}\left(t\right)$ for each copula. This process will derive the parametric estimates of $Kc\left(t\right)$.

Once the parametric and non-parametric estimates are developed then a comparison between the two approaches can be made. This is achieved by calculating the Minimum Distance (MD) using the following equation:

$$MD= \int [Kc(t) - K(t)]^2 \, dK(t) \tag{17}$$

This is taking the sum of the square difference between the parametric and non-parametric estimates. Graphically, this is achieved by plotting the CDF of the non-parametric estimate and the CDF of the parametric estimate. The original data will be estimated by the non-parametric and the copulas by the parametric method. The graph resulting from plotting the two CDF's is known as the Quantile- Quantile (Q-Q) Plot. The copula that produces the minimum distance is the one that best reproduces the original data. The results and the application of the copula methodology and other dependency methods will be discussed in Chapter 6.

5.2 The stochastic integrated model: A systems approach

This section of the Chapter will describe the development of the stochastic integrated model as a holistic approach that will combine reserves, production, facility and economics data for an asset. This model is called the systems approach. An Excel spreadsheet and VBA codes will be used to build the model and link all the elements from reserves to economics. Then the input variables for the model are defined as distributions in order to capture uncertainty. Once the inputs are verified a simulation package such as @Risk will be used to perform Monte Carlo Simulation. Each section will describe the specific techniques that are used in the model.

Furthermore, this section will show the integration and mutual dependence of the decision parameters such as reserves, production profile, drilling rig selection, production facility selection, price and costs (Figure 5-4). The modelling process will be explained by discussing one complete iteration in a Monte Carlo Simulation of the entire system.

5.2.1 Reserve section

Reserves are calculated as a product of Original Oil in Place (OOIP) and recovery factor using a volumetric calculation of the OOIP and the API recovery factor equation (Arps, 1968) as shown in (Figure 5-5). The OOIP samples stochastic input distributions of the component parameters which are: area, net thickness, porosity, water saturation and the inverse of formation volume factor. Simultaneously, in recognition of the existence of functional dependencies of both OOIP and recovery factor on porosity and water saturation, samples generated for input into the OOIP calculation are also used with other inputs, such as permeability, for calculating the recovery factor. The value calculated for OOIP and the corresponding value of recovery factor are both used to calculate a corresponding value of reserves.

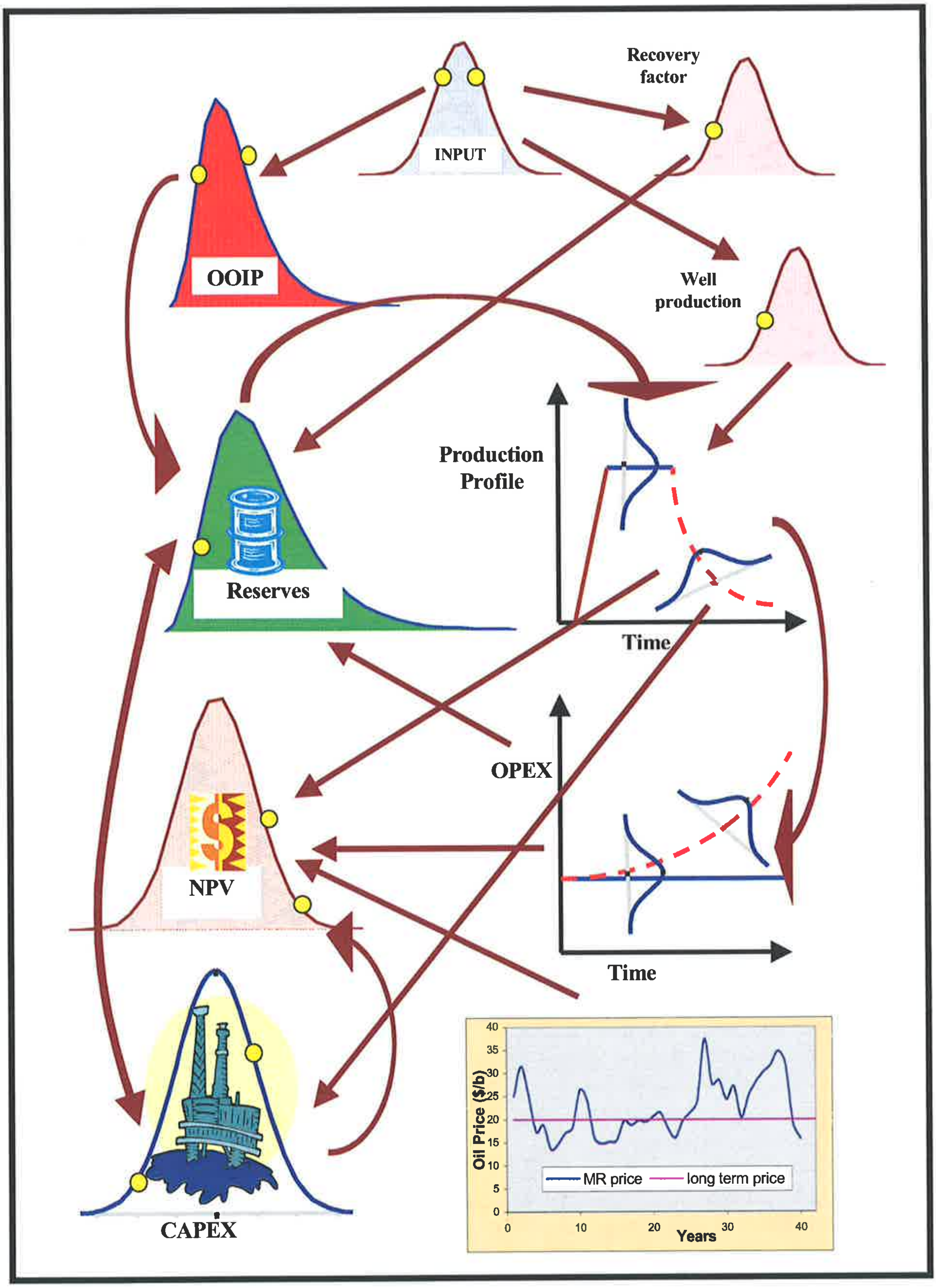


Figure 5-4. Systems approach model with dependencies and interactions

The reserve calculation assumes a tank model with constant thickness where net thickness is equal to gross thickness. The equation to calculate OOIP is as follows:

$$N_o = \frac{7758 * A * h * \Phi * (1 - S_{wi})}{B_{oi}} \tag{18}$$

N_o = OOIP, STB

A = Area, acres

h = Average thickness, ft

Φ = Average porosity, fraction

S_{wi} = Initial water saturation, fraction

B_{oi} = Oil formation volume factor, RB/STB.

This model assumes that the amount of gas produced is small, therefore it will be flared and the model will focus on oil reserves only. The recovery factor for the model is based on the assumption of a solution gas drive and uses the following equation.

$$E_R = 41.815\left(\frac{\Phi(1 - S_{wi})}{B_{0i}}\right)^{0.1611} * \left(\frac{K_a}{\mu_0}\right)^{0.0979} * (S_{wi})^{0.3722} * \left(\frac{P_b}{P_a}\right)^{0.1741} \tag{19}$$

E_R = Recovery factor =Recovery efficiency, fraction

Φ = Average porosity, fraction

S_{wi} = Initial water saturation, fraction

B_{oi} = Oil formation volume factor RB/STB

K_a = Absolute permeability, md

μ_0 = Oil viscosity at bubble point, cp

P_b = Bubble point pressure, psi

P_a = Abandonment pressure, psi

As shown in Figure 5-5, Technical Reserves (R_0) = OOIP * E_R, (20)

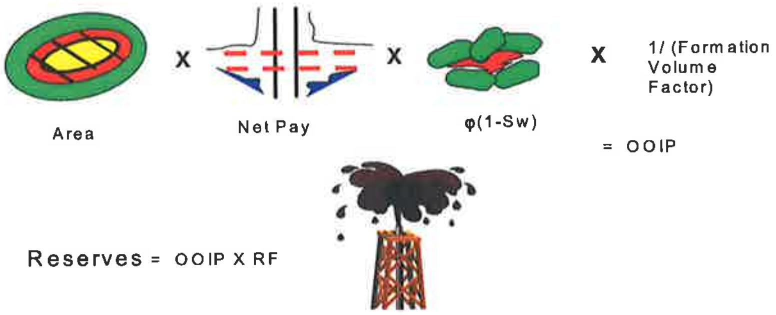


Figure 5- 5. Reserves elements

It is important to point out that this thesis computes technical reserves using the technical parameters of the reservoir. However, once the stochastic integrated model is developed, reserves are defined as commercially recoverable oil and are an output of the economic model, not an input, since they are a function of many economic parameters such as oil price and operating costs

Each one of the variables used to calculate the OOIP is modelled stochastically. The same applies for variables that are used to calculate the recovery efficiency. This results in the calculated technical reserves being a distribution.

5.2.2 Rig selection and drilling program

Rig selection and drilling programs are a function of water depth, costs, rig capability as shown in Table 5-3 as well as rig availability. For example, at water depths above 1000 ft, the decision maker would have three choices: semi sub 3rd generation, semi sub 4th generation or drill ship. Next costs are incorporated and the least expensive is chosen, the semi sub 3rd generation based on the assumed costs will be chosen.

Table 5-3. Drilling selection matrix

Water depth (ft)	Rig type	Costs ($/day)
250	Jack up	40,000
350	Jack up	60,000
1000	Semi sub 3rd generation	70,000
1000	Semi sub 4th generation	115,000
1000	Drill ship (DP)	150,000

In addition to rig selection, the number of wells drilled in any year is a function of rig availability. The model is able to cater for single or multiple rig-drilling schedules, which in turn affects the production schedule.

Rig availability is modelled as two choices:

= 1, is limited to only one rig

= Multiple, 2 or more rigs.

To determine the number of wells drilled by a rig per year the following two equations are used as average estimates:

$$Number\ of\ days\ to\ drill\ a\ well = drilling\ depth\ (ft) / daily\ rate\ (ft / day) \tag{21}$$

$$Number\ of\ wells\ drilled\ by\ a\ rig / year = (30 * 12) / \ Number\ of\ days\ to\ drill\ a\ well \tag{22}$$

This means that if one rig is available then only a certain number of wells can be drilled in a year, but if more rigs are available then more wells can be drilled, as in an acceleration program. For example, if rig availability is limited to only one rig and the number of days to drill a well is 49 days, then this rig can approximately drill only 7 wells in a year. Therefore, the total number of wells in any year is constrained by the number of wells that can be drilled by a rig and the number of rigs available.

5.2.3 Production Profile

This model assumes that all the wells are horizontal and have the same initial production capacity, with the functional dependency between OOIP and initial well rate arising from the fact that both are functions of net reservoir thickness. The same sample of net thickness used to calculate the OOIP is used to calculate a corresponding value of the initial well rate. Furthermore, due to the functional dependency between the recovery factor and the initial rate, both being dependent on permeability, the model uses the same permeability sample generated for calculating the recovery factor as an input into calculating a corresponding value of initial well rate. Initial rate per well was calculated using the Joshi method (Economides et al, 1994) as follows:

<u>Horizontal well flow (Joshi's Method) (barrels per day)</u>

$$PI = \frac{0.00708 * K_h * h}{\mu_0 * B_{0i} * \left\{ Ln\left(\frac{a + \sqrt{a^2 - (L_h/2)^2}}{L_h/2} \right) + \left(\frac{K_h h}{K_v L_h} \right) * Ln\left(\frac{h}{2r_w} \right) \right.} \qquad (23)$$

$$a = (L_h/2)\left[0.5 + \sqrt{0.25 + (2r_e/L_h)^4}\right]^{0.5}$$

$$q_i = PI * (\bar{P} - P_{wf}) \qquad (24)$$

where

h = Average thickness, ft

K_h = Horizontal permeability, md

K_v = Vertical permeability, md

L_h = Length of horizontal well, ft

r_e = Drainage radius of horizontal well, ft

r_w = Wellbore radius, ft

a = Half the major axis of the drainage ellipse, ft

B_{oi} = Oil formation volume factor RB/STB

μ_0 = Oil viscosity, cp

PI = Productivity index, barrel/day/psi

q_i = Initial production, barrel/day

Once a value of reserves and a value of initial well rate are generated for a single iteration they are used to calculate the production profile of the field. This model ignores the build up phase of production. The model also assumes that the well production capacities will start declining immediately as production starts from that well. Maintaining plateau production, therefore, requires combined well capacities to exceed the plateau rate. The rate of production decline for each well is a function of the depletion level of the entire reservoir. Therefore, the initial rates of all future wells are functions of the remaining reserves of the total reservoir. The calculated values of reserves and the initial well rate are used in a feedback mechanism to generate the production profile. At the beginning of production, the initial reserves are the same as the remaining reserves, and the well rate is the same as the initial well rate. In the second year, the remaining reserves are calculated using the production in the previous year and the initial reserves estimate. Then, the decline rate is calculated based on the reserves remaining and the initial reserves estimates. This process yields a decline rate that is similar to the exponential decline rate. This decline rate affects the production in the second year, which in turn affects the total production. This cycle is repeated until the economic limit is reached. This process is mathematically described as follows and is shown in Figure 5-6.

where

Q = Production capacity = Facility limit, barrel per day or barrels per year

R_t = Remaining reserves, million barrels at time t

R_0 = Technical reserves (initial reserves), million barrels.

P_d = Daily production, barrels per day

P_{Td} = Total daily production or sum of all well daily production, barrels per day

P_{yt} = Yearly production, barrels per year

D_t = Decline rate at a time t

w_0 = Number of wells for the first year or at the beginning of plateau

w_t = New additional wells for a given year.

Assume that $t = 0$, is the current time, or the year production will start

1. At $t = 0, R_0 = R_t$, therefore $D_0 = 1$, where $D_t = R_t / R_0$
2. $P_{d0} = q_i$
3. $P_{Td0} = q_i * w_0$, where $w_0 = Q / q_i$
4. $P_{y0} = P_{Tdo} * 365$

At $t = 1$,

5. Calculate remaining reserves at year 1, $R_1 = R_0 - P_{y0}$
6. Calculate decline rate, $D_1 = R_1 / R_0$
7. Calculate daily well rate in second year, $P_{d1} = q_i * D_1$
8. Calculate Total daily Production rate at year 2, $P_{Td1} = P_{d1} * w_0 + P_{d1} * w_0$
9. Calculate Total daily Production rate at year 2 that will reach plateau, $P_{Td1} = Min\,(Q, P_{d1} * w_0 + P_{d1} * w_1)$
10. Total yearly production at year 2, $P_{y1} = P_{Td1} * 365$
11. At $t = 2$, Continue the process and go back to step 5 again

This process in general is as follow:

a. Starting with remaining reserves $R_{t+1} = R_0 - P_{yt}$

b. Calculate Decline rate $D_{t+1} = R_{t+1} / R_0$

c. Calculate daily well rate in t years, $P_{dt+1} = q_i * D_{t+1}$

d. Calculate Total daily Production rate at year t,

$P_{Tdt+1} = P_{dt+1} * w_0 + P_{dt+1} * w_0$, where $w_0 = Q / q_i$

e. Calculate Total daily Production rate at year t that will reach plateau,

$P_{Tdt+1} = Min\,(Q, P_{dt+1} * w_0 + P_{dt+1} * w_{t+1})$

f. Total yearly production at year t, $P_{yt+1} = P_{Tdt+1} * 365$,

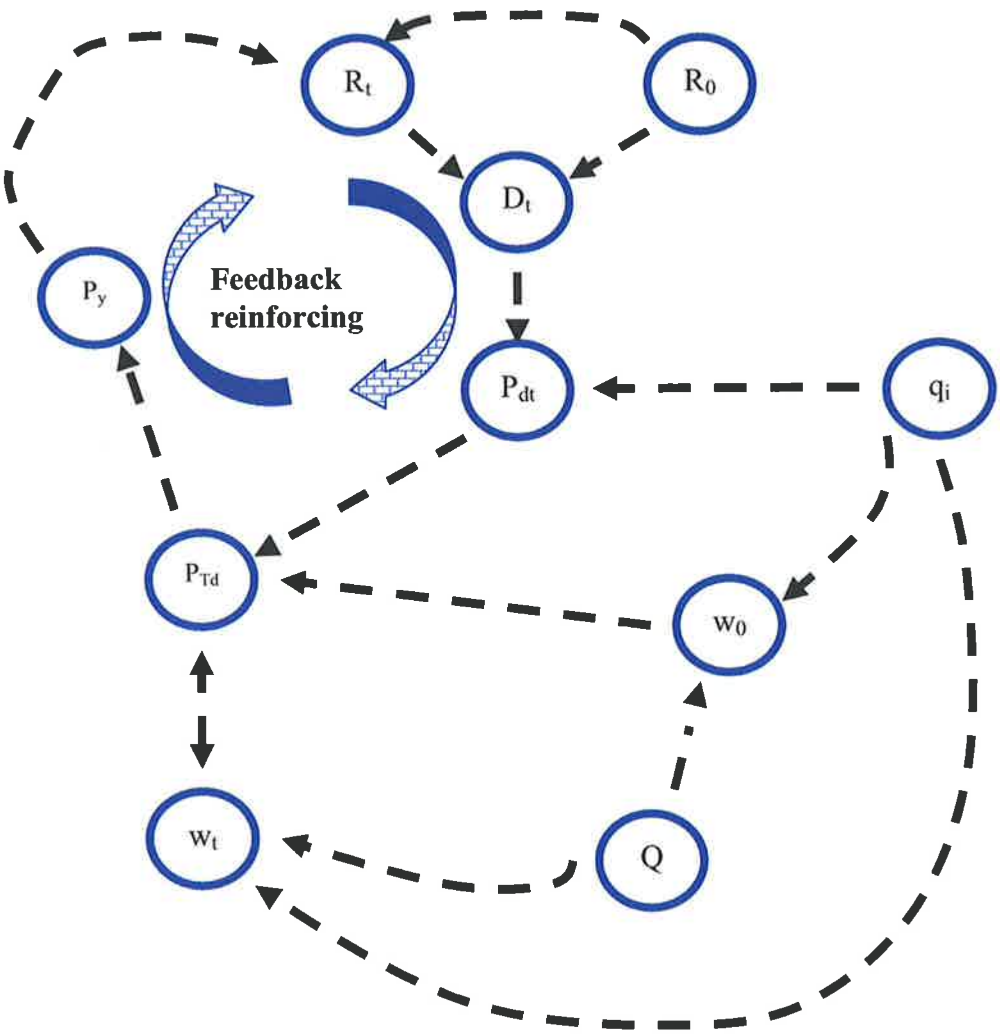


Figure 5-6. Production Model

This model assumes that the decision maker will assign a production capacity or facility limit based on the expected value of reserves. This process of using the calculated individual value of reserves, initial well rate and drilling parameters generates a production profile (Figure 5-4). The production start-up time is also interlinked with economics to show the impact of production start-up timing on NPV.

Each year, the number of additional wells (w_t) to be drilled is determined as a function of the shortfall in production from existing wells relative to facility capacity and the initial rate of the new wells, which declines due to the overall reservoir depletion. If the combined capacity of the current wells exceeds the facility capacity then no additional wells are drilled. Otherwise, more wells are added to maintain production at facility capacity. Number of wells is mathematically calculated as follows:

1. Number of wells per year to maintain plateau (w_t).

In order to calculate this value, the following assumed conditions need to be met, other assumptions can also be made:

- Number of wells per year should be less than, or equal to, the number of wells that can be drilled by Rig availability.
- Number of wells per year should be less than, or equal to, the total number of wells required to produce the whole field.

Number of new wells in a given year $(w_t) = (Q -$ previous yearly production$)/ q_i$.

2. Total number of wells for the field (W).

In the systems approach model, the total number of wells is a function of reserves, initial well rate, capital cost per well, operating cost per well and net oil price after tax (Cunningham et al, 2004). Assuming everything else is constant if reserves increase (decrease), total number of wells would be expected to increase (decrease) accordingly. Also, if well rate capacity is higher (lower), total number of wells would be lower (higher) again assuming everything else is constant. The same analogy is applied with Oil price; if oil prices are higher, an increase in number of wells would be expected to gain more profit, this is simply an acceleration program. Another way to view dependency between total number of wells and reserves, initial well rate, capital cost per well, operating cost per well and net oil price after tax is as an optimisation problem where the objective is to find the optimum number of wells for the field. The following equations and conditions are used to calculate total number of wells:

This model has two options to model total number of wells for the field:

1. Deterministic where the decision maker chooses the total number of wells to produce the whole field based on the expected value. This is done for the sequential approach.
2. Total number of wells (W) can be a function of reserves, production, costs and oil price using the following equation (Cunningham et al, 2004):

$$W = \frac{R * D_r}{q_i} \frac{\left(\sqrt{P_n - \frac{\alpha * E_w}{q_i}} - \sqrt{\frac{C_w * D_r + E - \alpha * E_w}{q_i}}\right)}{\sqrt{\frac{C_w * D_r + E - \alpha * E_w}{q_i}}} \tag{25}$$

$$\alpha = (1 + D_r)^{-Tab}$$

q_i = Oil production, barrels per well per year (production model)

R = Technical reserves, million barrels (reserves model)

P_n = Net oil price after all taxes and royalties, \$/barrel (economic model)

D_r = Discount rate, fraction (economic model)

C_w = Net capital cost per well, $ per well (economic model)

E_w = Net OPEX per well, $ per well (economic model)

Tab = Abandonment time, years (economic model)

Cunningham et al (2004) derived equation (25) using the assumption of exponential decline analysis, which this model is using. This equation fits the model well. If any of the input parameters are changed, the number of wells will change accordingly.

The systems approach model sums all the production from all the wells and provides the production profile for the field

At this stage the systems model has options such as:

- Changing production start time.
- Changing production capacity.

In addition, the systems approach model will calculate the following as an output:

- Daily and yearly production at Plateau
- Number of years at Plateau
- % of Reserves produced at Plateau

5.2.4 Facility and export section

The value of reserves, calculated earlier and used as input into the production model, is also used as input into the facility model which uses simple rules of thumb to determine the appropriate facility's type as a function of reserves, water depth, drilling and storage options. The hypothetical oil field of this study has the expected value reserves of 63 million barrels, is remote and in water depth of 1000 ft. In these circumstances the model chooses an FPSO with shuttle tankers for exporting production. More sophisticated facilities selection processes and stochastic facilities capacity estimating models can also be incorporated. The process of choosing the FPSO and shuttle tanker is as follows:

A production facility choice will depend on the following factors:

- Reserves volumes
- Number of wells
- Water Depth
- Distance to infrastructure
- Storage and drilling facilities

Based on the above factors, then a choice of facility will be used for each development scenario and the optimal solution chosen.

A decision matrix was developed using the following arbitrary guidelines:

Reservoir volumes: If reserves are less than 250 million barrels, then SPAR and Tension leg Platform (TLP) facility will not be chosen because of their high costs.

Water depth: If water depth is greater than 1000 feet, no fixed facility will be used due to the high cost of installing it at this depth based on the current literature.

Number of wells If the number of wells is greater than 5 wells then min-TLP will not be used, this is a technical limitation. Furthermore, if number of wells is greater than 20 then SPAR cannot be used. Based on the current literature most of the SPARs used have a technical limitation of 20 wells. However, this could be easily changed in the model with the increased use of better technologies.

Distance: This impacts the decision of which export mechanism to choose. There are two options: a pipeline or using a shuttle tanker taking account of initial CAPEX and ongoing OPEX. On this basis a shuttle tanker was chosen for this analysis.

Storage and drilling facilities: This is based on whether the facility requires drilling or not and whether the facility requires storage. If the facility requires storage, then an FPSO or SPAR might be a suitable choice. And if a facility requires drilling then FPSO cannot be chosen. A semi-submersible has two types, one with drilling and the other without drilling.

For this study an offshore field development case was used with the following inputs:

Reserves expected value	= 63 million barrels
Water Depth	= 1000 ft
Number of wells	= 7
Requires storage	= Yes
Distance	= 60 miles from nearest location.

According to the decision matrix explained above the FPSO is the suitable choice with a shuttle tanker as an export mechanism.

It is important to indicate that the above arbitrary assumptions about reserves volumes, water depth, number of wells, distance and process selection are based on the current literature and simplified as rules of thumb. The objective of this study is to simplify a detailed model so that it captures the main parameters and shows the process of decision-making.

5.2.5 Economic Modelling

Economic calculations are divided into two groups: cost and revenue. As the calculated realisation of the production profile is generated, it is directly linked to operating expenditure (OPEX), which is divided into fixed and variable costs, with the latter being dependent on the production rate (Figure 5-4). The capital expenditure (CAPEX) associated with production and export facilities together with the number of wells drilled per year are used to calculate the annual capital expenditure (CAPEX). CAPEX dependence on production capacity is input in tabular form. Abandonment costs are incurred when the field operating costs are greater than revenues. Fixed and variable OPEX as well as all components of CAPEX are stochastic. The following are used as an input into the calculation of CAPEX and all are presented as stochastic

- Exploration well costs ($ Million per well)
- Dry hole costs ($ Million per well)
- Production well costs ($ Million per well)
- Floating Production Storage and Offloading (FPSO) costs ($Million)
- Export system: Cost of shuttle tanker or pipeline ($Million)
- Sub-sea equipment &installation ($Million)
- Other costs, administration and management ($Million)
- Production Capacity costs ($Million)

The sum of all the inputs above is the total CAPEX, which is modelled in the year that it is incurred. For this model a CAPEX matrix was assumed, as function of number of years to start production as shown in Table 5-4.

Table 5-4. Capital expenditure profile as a function of production start years

Production Start, years	CAPEX, % of Total				
	Year 1	Year 2	Year 3	Year 4	Year 5
Year 3	40	60	✧		
Year 4	40	40	20	✧	
Year 5	20	40	20	20	✧
Production start up ✧					

The CAPEX for the wells is incurred when the wells are drilled. This is modelled through a direct relationship between production and well capital expenditure.

Revenue calculations use production and oil prices as inputs. Finally the model calculates the NPV using production, price, costs, discount rate and taxes as inputs (Figure 5-4).

The economic section is mainly focused on calculating the Net Present Value (NPV). NPV is defined mathematically by the following formula:

$$\text{Net Present Value (NPV)} = \sum_{t=1}^{n} \frac{Net\ Cash\ Flow_t}{(1 + Discount\ Rate)^n} \tag{26}$$

$$\text{Net Cash flow} = \text{Income} - \text{Taxes} \tag{27}$$

$$\text{Income}_t = \text{Revenue}_t - \text{CAPEX}_t - \text{OPEX}_t \tag{28}$$

Taxes = 0.34 * Income, This model assume 34% tax rate. (29)

$$\text{Revenue}_t = \text{Yearly Production}_t * \text{Oil Price}_t \tag{30}$$

Discount rate is assumed to be 15% per year.

5.2.5.1 Price models

The model provides five options to predict oil price:

1. Deterministic.

2. Increasing or decreasing oil price with time

$$P_t = P_0\ (1 + E)^{(t - t_0)} \tag{31}$$

t : Time (years)

P_0 : First year oil price ($/barrel)

E : Escalating or de-escalating factor (%)

3. Stochastic oil price

Assuming price could have a normal, lognormal or any other type of defined distribution.

4. Geometric Brownian Motion (GBM)

$$P_{t+1} = P_0 \left(\exp(\alpha - 0.5\sigma^2)t + (\sigma\sqrt{t})\psi\right. \quad (32)$$

α : Long term growth, fraction

σ : Short term volatility, fraction per year

t : Time (years)

ψ : Normal distribution with mean of zero and variance of 1, N- (0, 1).

One realization of the geometric brownian motion oil price model is shown in Figure 5-7.

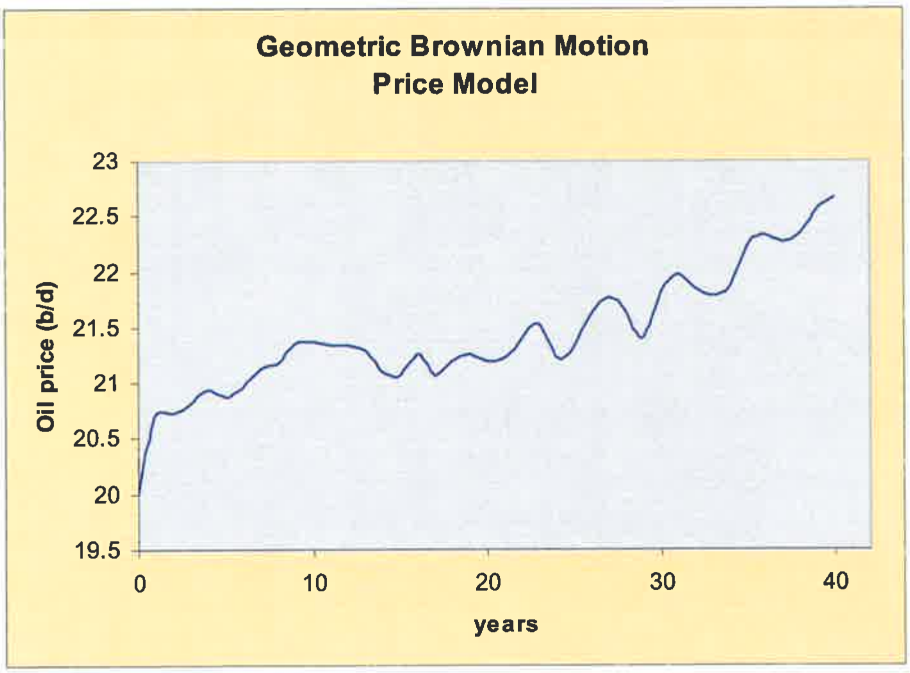


Figure 5-7. GBM with a random price realization

5. Mean Reversion

A sample of one realization of the mean reversion is shown in Figure 5-8 and is derived using the following equations:

$$P_x = P_{tl}\exp(-Kt) + P_{pl}(1-\exp(Kt)) + ((1-\exp(-2Kt))(\frac{\sigma^2}{2K}))^{0.5}\psi \quad (33)$$

$$P_{t+1} = \exp(P_x - 0.5((1-\exp(-2Kt))(\frac{\sigma^2}{2K}))) \quad (34)$$

$P_{tl} = \ln(P_t)$: Natural log of the current price ($/barrel), this is deterministic

$P_{pl} = \ln(p_p)$: Natural log of the long-term price ($/barrel) p_p, this is deterministic

σ : Short term volatility (fraction per year)

t : Time (years)

ψ: Normal distribution with mean of zero and variance of N- (0, 1)

η : Reversion speed (Years)

$K = \frac{\ln 2}{\eta}$: Reversion factor (per year)

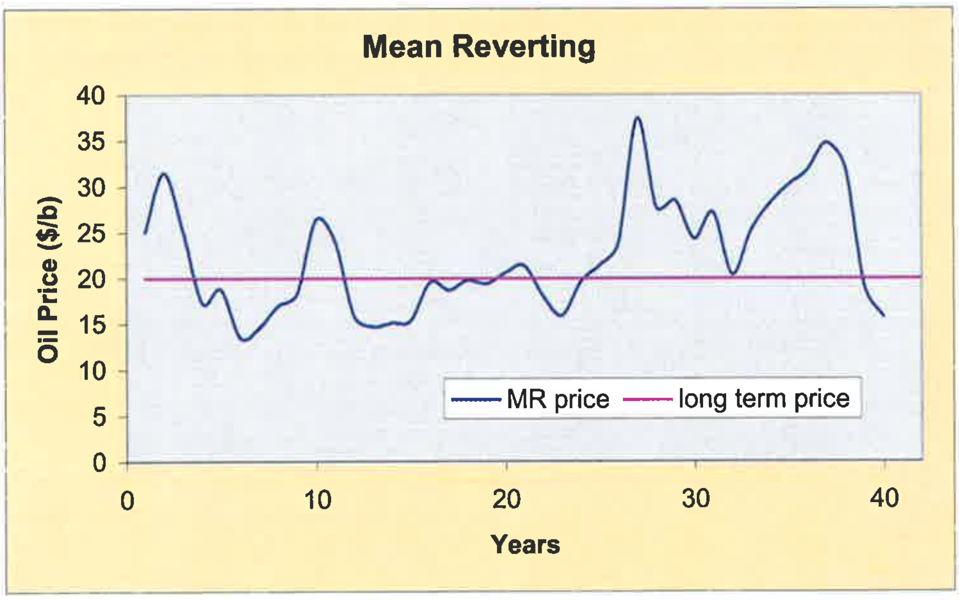


Figure 5-8. A random realization of the mean revering price

The output of the economic section for the systems approach model is the following:

- Net Present Value (NPV)
- Internal rate of Return (IRR)
- Payback period
- Statistical measures such as P10, P50 and P90 values of NPV.

The preceding sections; reserves, rig selection and drilling program, production profile, facility and export and economics describe one iteration of the systems approach model (Figure 5-4) to calculate the value of NPV, taking dependencies into account. The model then goes back and carries out another iteration by picking new samples of inputs required to calculate the NPV and repeats the whole process. The iterations are continued until a stochastic representation of NPV is generated. This process is presented in a simplified version in Chapter 4 and is shown in Figure 5-9 below.

In contrast as shown in Figure 4-2, the sequential approach calculates the distribution of each of the decision parameters independently of each other and then combines these distributions to arrive at the NPV.

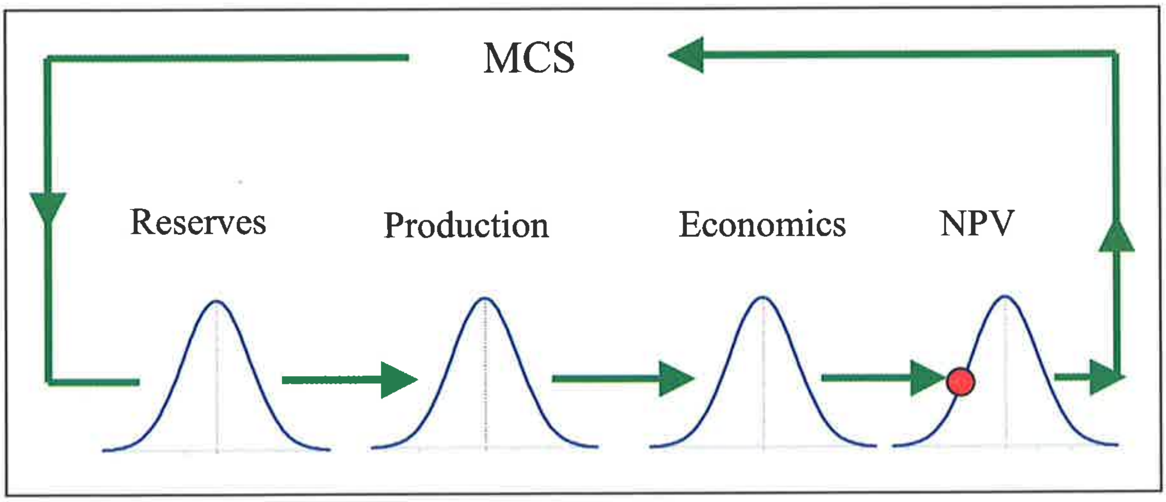


Figure 5-9. Systems approach

CHAPTER 6

Testing copulas: Modelling dependence in reserves calculations

6. Introduction

This Chapter builds on Chapter 3 which introduced the statistical dependency methods including the Copulas approach and Chapter 5 which illustrated the construction and simulation of Copulas. This Chapter illustrates the potential benefits of using copulas to model dependencies in oil and gas applications with a particular focus on estimating reserves. The copulas method has been applied to a simple reserve calculation and the results are compared and contrasted with the results of some of the more commonly used approaches to model dependence. The comparisons show that the traditional methods have problems in accurately reproducing the dependence structure in the tails of the variable distributions. Finally, this Chapter illustrates how the dependence structure can be captured and modelled using the copulas approach.

6.1. Reserves estimation: data and assumptions

The example this study will discuss is an oil reservoir with a focus on technical reserves, which are estimated using the volumetric equation and a recovery factor. The focus is on probabilistic reserves, so area, net thickness, porosity, water saturation, formation volume factor and recovery factor are all treated as uncertain. The reserves estimates are for a new oil discovery with limited data to conduct a probability analysis. Data were collected from nearby oil producing fields that have similar geology to the discovered field. That data for each variable such as area, thickness and porosity were input into a Best Fit™ software (Palisade group, 2004) to determine the best fit distribution using the chi square measure “goodness of fit”.

In addition to finding the best fit distribution, the software was used to estimate the statistical parameters for each variable. For example, area was best fitted with a lognormal distribution with a mean of 700 acres and a variance of 220 acres. This process is carried out for all the variables and summarized in Table 6-1. A plot of net thickness versus recovery factor shows a positive correlation (Figure 6-1). Furthermore, as shown in Figure 6-2, porosity and water saturation are negatively correlated. These inputs are correlated and the objective is to explore the impact of the choice of correlation model.

Table 6-1. Input used for Calculation of reserves

Original Oil in Place (OOIP) and reserves input:	Distribution	Minimum	Most likely	Maximum	Mean	Variance
Area (acres)	Lognormal				700	220
Thickness (ft)	Triangular	40	170	300		
Porosity (%)	Triangular	0.2184	0.35	0.4653		
Water saturation (Sw) (%)	Triangular	0.1915	0.30184	0.4969		
Formation Volume Factor (FVF)	Triangular	1.1	1.15	1.2		
Recovery factor (fraction)	Triangular	.04	.37	.37		
Note: Lognormal distribution for area is truncated between 400 and 1200 acres						

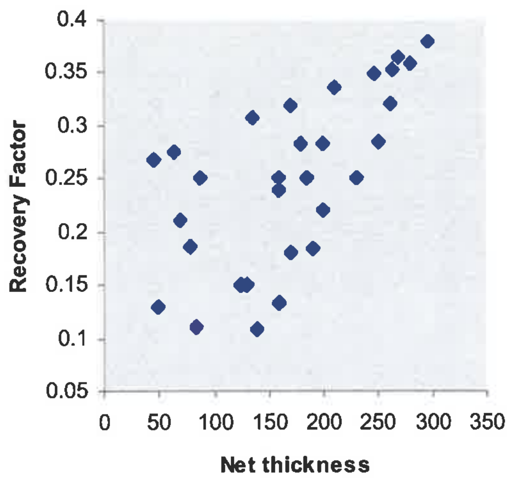


Figure 6-1. Original data with positive dependency between thickness and recovery factor

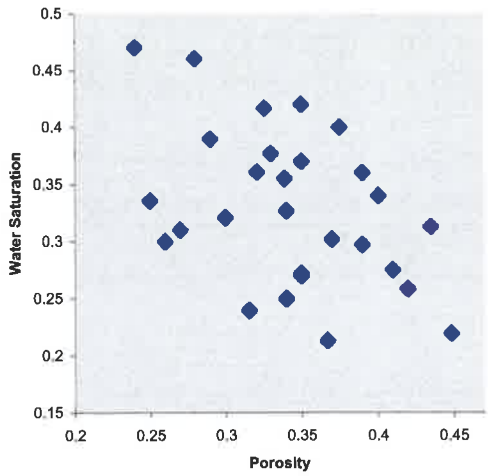


Figure 6-2. Original data with negative dependency between porosity and water saturation

6-2. Results

To investigate the impact of dependency methods, the data in Table 6-1 were used to calculate the Original Oil in Place (OOIP). Furthermore, each dependency method was used to model the positive dependence between net thickness and recovery factor and the negative dependence between porosity and water saturation. The multiplication of original oil in place and recovery factor yielded the technical reserves distribution.

6.2.1 Iman – Conover Method

To model the correlation between net thickness and recovery factor using the Iman-Conover method, triangular distributions were found to be the best fit for both variables. The Spearman correlation for net thickness and recovery factor is 0.67. Triangular distributions were also found to be the best fit for porosity and water saturation with a Spearman correlation of – 0.388.

6.2.2 Envelope Method

Lower and upper bounding lines were estimated for both of the two sets of dependent variables and were used for simulation of the variables as shown in Figure 6-3. For the recovery factor and net thickness a multiple line envelope was used to capture the shape of the dependence structure while for the porosity and water saturation a classic two lines envelope method was used.

6.2.3 Regression fitting method

A Regression analysis was used to find the best fit line for both the net thickness and the recovery factor and for porosity and water saturation dependence structure. The standard error was also estimated for both sets of dependent variables. The regression gave the following equation to correlate net thickness and the recovery factor.

$$\text{Recovery factor } = 0.1267 + 0.000721\,(\textit{Net Thickness}) + \textit{Normal}(0,\, 0.0622)$$

where Normal (0, 0.0622) is the normal distribution with a mean of zero and standard deviation equal to the standard error of the recovery factor and net thickness.

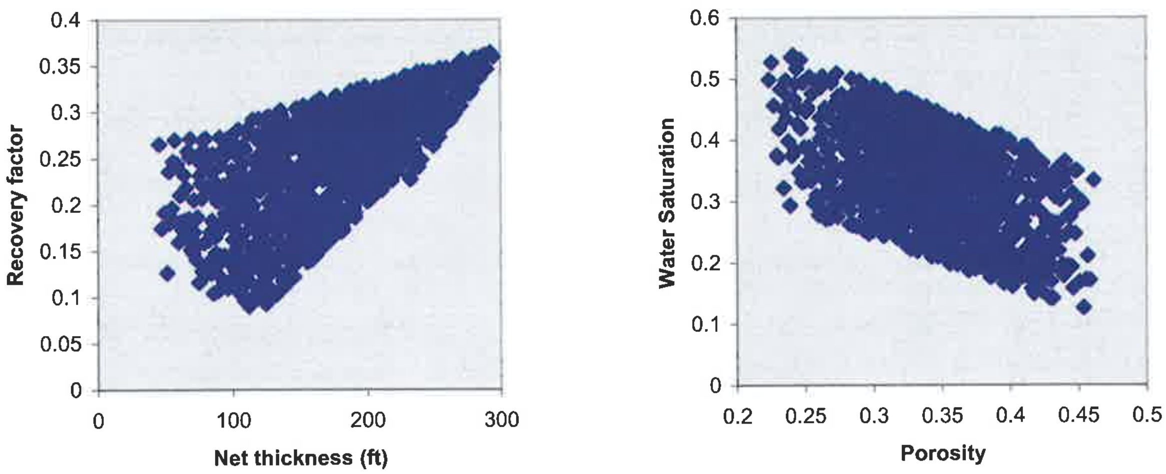


Figure 6-3. Envelope method net thickness vs. recovery factor and porosity vs. water saturation

Note that the above equation was modified to eliminate the possibility of a negative recovery factor. A cut-off for recovery factor of 0.10 was used, which is the minimum value in the original data.

6.2.4 Copula Method

To model dependence using copulas, Kendall's tau was calculated using the following equation:

$$\tau = \frac{\#(concordant\ pairs) - \#(discordant\ pairs)}{\frac{n\,(n-1)}{2}}$$

The Kendall tau is 0.51 for net thickness versus recovery factor and $\tau = -0.283$ for porosity versus water saturation. The Gumbel copula was chosen for the net thickness and the recovery factor as the best fit among the copulas. The reason for the choice will be explained further later. The next step was to calculate the value of theta; $\theta = 2.04$. For porosity and water saturation the Frank copula was chosen to reproduce the original correlation and the value of theta is $\theta = -2.72$. This is because of the ability of the Frank copulas to model positive and negative dependency whereas the Gumbel and Clayton copulas are suitable for positive dependencies only.

6.2.5. Summary and analysis

Table 6-2 shows the mean and standard deviation of reserves distribution using the four correlation models. It also shows the mean and standard deviation assuming no dependency. In the last column of Table 6-2, which includes both dependencies, a significant impact can be seen on both the mean and the standard deviation of the reserves estimates compared with the no dependency case. This confirms that dependencies do matter and should be included in any simulation to ensure capturing the appropriate statistical distribution of computed results to support optimal decision-making.

Table 6-2. Reserves (Million of barrels) mean and standard deviation and dependency methods

Methods	Measure	No Dependency	Porosity and Water Saturation only (1)	Net thickness and RF only (2)	Both (1) and (2)
Iman-Conover	Mean	53.52	53.59	55.70	55.83
	SD	27.67	28.16	31.80	32.17
Envelope	Mean	53.52	53.93	47.15	47.06
	SD	27.67	29.23	28.52	30.05
Regression fitting	Mean	53.52	53.73	48.57	48.58
	SD	27.67	28.31	29.25	30.60
Copulas	Mean	53.52	53.62	55.84	55.92
	SD	27.67	28.11	32.68	32.67

All of the models indicated that dependencies have a larger impact on the standard deviation than on the mean. The difference in the means, comparing the no-dependency results with the correlated results ranged between 4 and 12 %. It is important to point out that conclusion based on the mean apply equally to the P50 since this is a common industry measure. The standard deviation, however, increased by 8 to 18 % in the correlated versus no-dependency cases. This confirmed what other authors have shown: ignoring dependence might lead to a significant underestimation of the uncertainty in the simulated results.

From the cases where only one dependency was included at a time, it is observed that the dependence between net thickness and recovery factor has a larger impact on both the mean and standard deviation than the dependence between porosity and water saturation. This is supported by a tornado plot (Figure 6-4), which showed that net thickness and recovery factor have a larger impact on the reserves estimation than porosity and water saturation. This confirms that it is particularly important to include dependencies between those variables that have the largest impact on the output. Therefore, the analysis will focus on the dependency between the net thickness and the recovery factor since they had the largest impact on technical reserves.

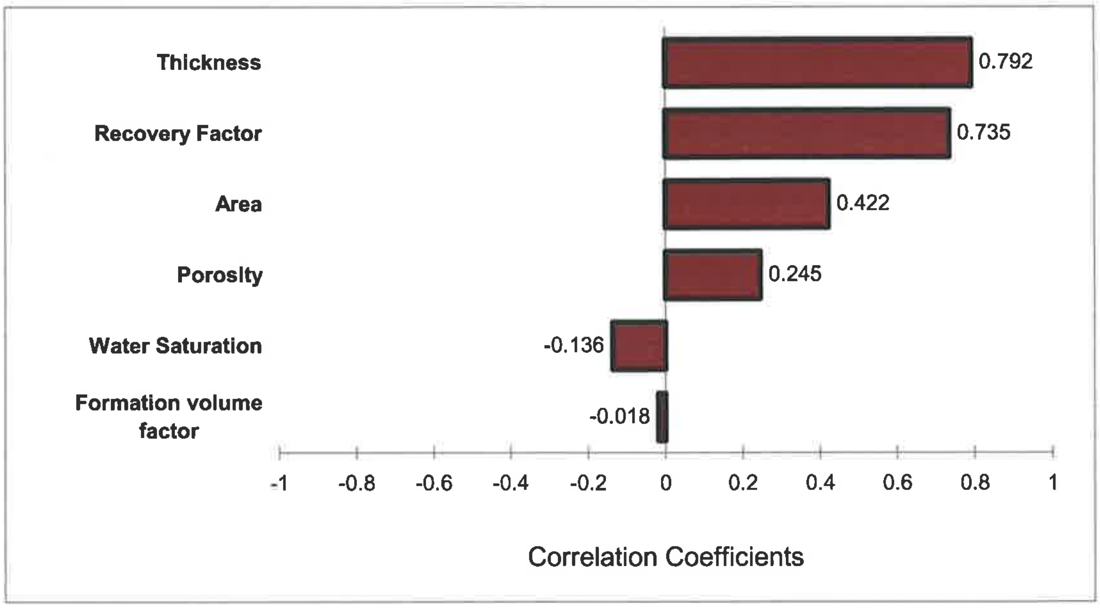


Figure 6-4. Sensitivity analysis for the technical reserves

The next important question this study addresses is: Which one of the correlation models best represents the original data? The answer to this question will show which method clearly reproduces the dependence structure of the original data (Figure 6-1) and will also quantify the impact of capturing the dependence structure.

The Frees approach with both graphical and quantitative analysis to find the best fit was used. Figure 6-1 shows that the original data correlation pattern has strong upper tail dependence. Upper tail dependence means that at higher values of net thickness, an increase in net reservoir thickness will lead to a higher increase in recovery factor compared to the case of lower values of net thickness.

The Q-Q Plots (Quantile-Quantile plots)represent the x-axis with Cumulative Distribution Function (CDF) of the original data and the y-axis with CDF of the dependency methods.

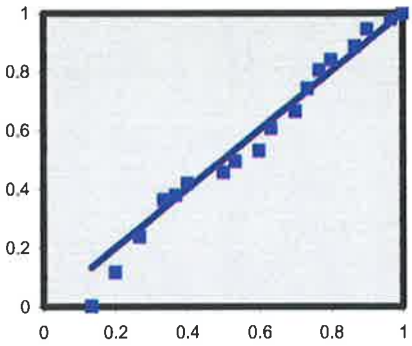


Fig. 6-5. Q-Q Plot Gumbel

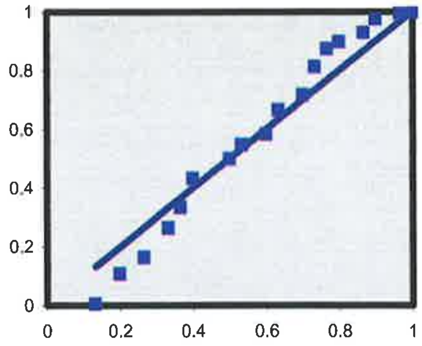


Fig.6-6. Q-Q Plot Regression

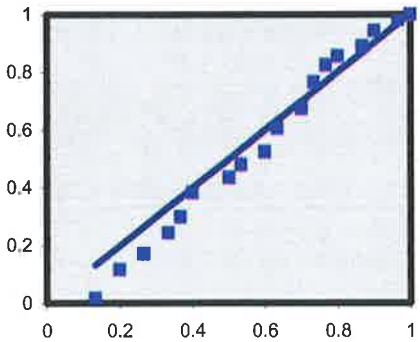


Fig.6-7. Q-Q Plot Envelope

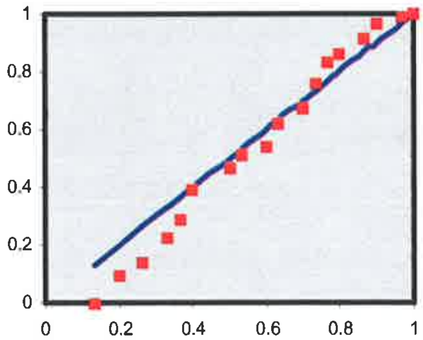


Fig. 6-8. Q-Q Plot Iman-Conover

In Figures 6-5 to 6-8, the 45-degree line represents the original data and the points are the modelled dependency methods. Using a Q-Q plot to analyse the correlation in particular above the 50th percentile (Figures 6-5 to 6-8), it was found that the Gumbel copula gave the best match with the original data.

Reviewing the Q-Q plots for the Iman-Conover, the envelope and the regression fitting methods, it was very difficult to find a significant difference between the three models.

A quantitative approach confirmed the graphical interpretation. The difference between the original data and the data generated using the dependency method was quantified by using the minimum distance, equation 17 in Chapter 5. The Minimum

Distance (MD) is defined as the sum of the square difference between the two sets of data. Quantitative analysis showed that the Gumbel copula, with a Minimum Distance (MD) of 0.0417 as shown in Figure 6-9, was the best dependence method that reproduces the original data among Archimedean copulas where the Frank copula had a MD of 0.098 and the Clayton copula had a MD of 0.17. The second best is the Envelope method with a MD of 0.064. The third is the Iman-Conover method with a MD of 0.08 and the fourth, Regression fitting with a MD of 0.081. This confirmed that the Gumbel copula fitted the original data best. These results confirm the assertion that the Gumbel copula is particularly suited to capturing upper tail correlation patterns. None of the more commonly used methods, the Iman-Conover, Envelope and Regression fitting methods, were able to capture the upper tail structure as well as the Gumbel copula.

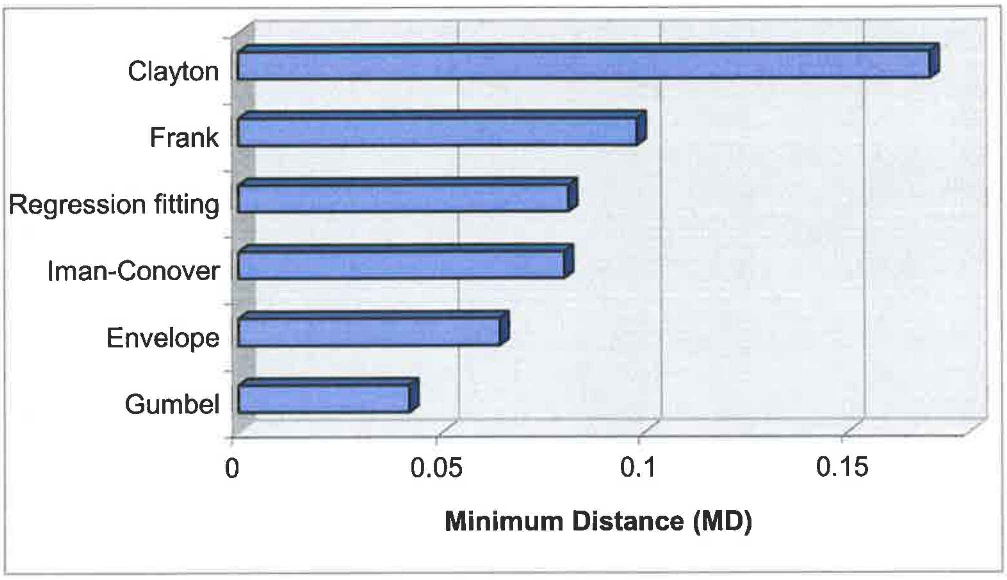


Figure 6-9. Minimum distance for all the dependencies methods

The Envelope method is very flexible and powerful. With the appropriate subjective choice of bounding lines and the distribution to use to sample between these lines, it can be tailored to capture almost any dependence structure. In modelling dependency for the envelope method two lines were initially used. This was not enough to capture the shape of the dependence structure and, consequently, more lines were added. It was found that by adding multiple lines the envelope method yielded better results and the MD could be reduced from 0.117 to 0.064. It was this multiple lines envelope method, which was used for the comparison with the other dependency methods, not the classical two-line model (Figure 6-10).

Furthermore, the process of constructing two lines only makes sense if the joint distribution is normal, or in the general case, if the shape of the dependence structure is an ellipsoid. However, with a dependence structure that has a lower or an upper tail pattern, it was found out that constructing an envelope using more lines yields better results. The challenge, of course, is to subjectively pick the most appropriate bounding lines for the various points in the correlation structure.

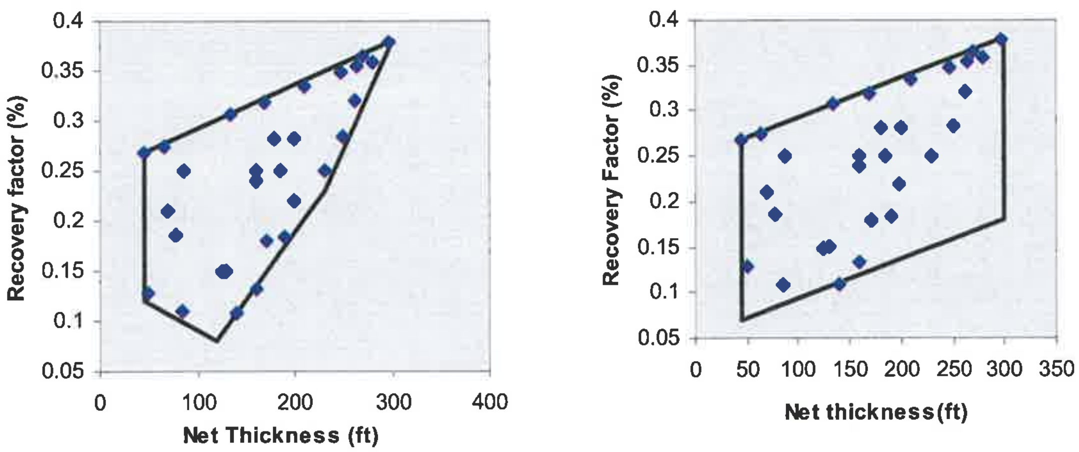


Figure 6-10. Envelope method: on the left with multiple lines and on the right with two lines

In the study example, using the standard approach with two lines to capture the dependence structure led to an underestimation of the mean by 13% and an underestimation of standard deviation by 17% compared with the use of multiple lines (Figure 6-11).

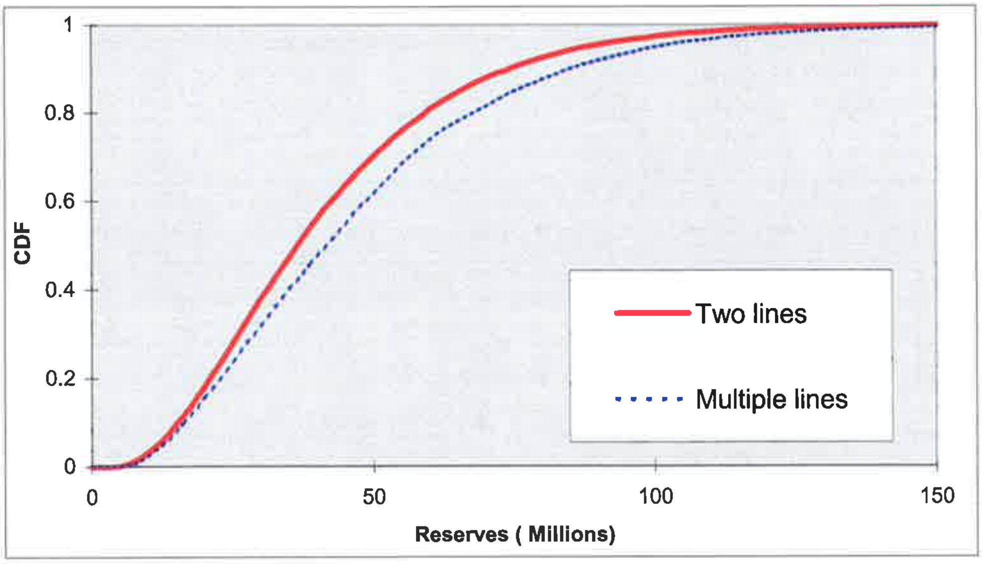


Figure 6-11. Impact of envelope method with multiple lines

6.3. Discussion

This research has investigated how well various correlation models represent the original distributions of the input variables. When the Monte Carlo simulation is initiated, the dependence methods would be expected to reproduce the original dependence pattern as well as to maintain the original distribution of the input variable. The Iman-Conover and copulas retain the same triangular distribution as the original while the regression-fitting model simulates a normal distribution (recovery factor). This illustrates that irrespective of what the input distribution is, using the regression fitting method, the simulation will always result in a normal distribution. This should be expected since the design of the equation has a normal distribution assumption embedded in it as the standard error is normally distributed.

The envelope method tends to retain the original distribution shape only in a weak form. If multiple bounding lines are not used, the envelope method tends to produce other distributions such as a normal distribution.

Compared with the copula approach; the Iman-Conover, envelope and regression fitting models all fail to capture upper tail dependence pattern.

It can be concluded that the Gumbel copula fitted the original data best. Although this, in itself, is interesting, the key question is whether the choice of correlation model will influence the decision metrics (Technical reserves). Let us define the percentile of data from 0 - 0.2 as a *lower tail* and the percentile of data from 0.8 – 1 as an *upper tail*. Note that an ascending cumulative distribution function to define the percentiles is used.

Figure 6-12 shows two significant results. Firstly, the Gumbel and Iman – Conover methods seem to have the same results and the Envelope and Regression fitting methods seem to have similar results. Both the Gumbel and the Iman- Conover methods yield higher reserves for a given probability than the Envelope and Regression fitting methods in the range of 10-15 million barrels. Secondly, examining the upper portion, the Gumbel copula gives the highest reserves in the upper tail (Figure 6-13). This is because it captures the correlation between thickness and the recovery factor in the high probability data. For example, at the 90th percentile, the difference between the Gumbel copula and both the Envelope method and the Regression fitting methods is approximately 15 million barrels. This is a significant difference amounting to 16% of the total estimates, which shows the superiority of the copulas approach in capturing original data upper tail dependence structure based on the minimum distance measure. However when compared with the Iman-Conover the difference is only 1- 2 million barrels (2%). This is a very small difference suggesting that the impact of capturing the dependence structure is really not significant when compared with the Iman- Conover method.

In conclusion, even though the copulas approach captured the upper tail dependence structure and the Iman- Conover did not, the impact of capturing the upper tail dependence structure via the use of the copulas when compared to Iman-Conover is not material. It did not significantly change the results of the decision matrices.

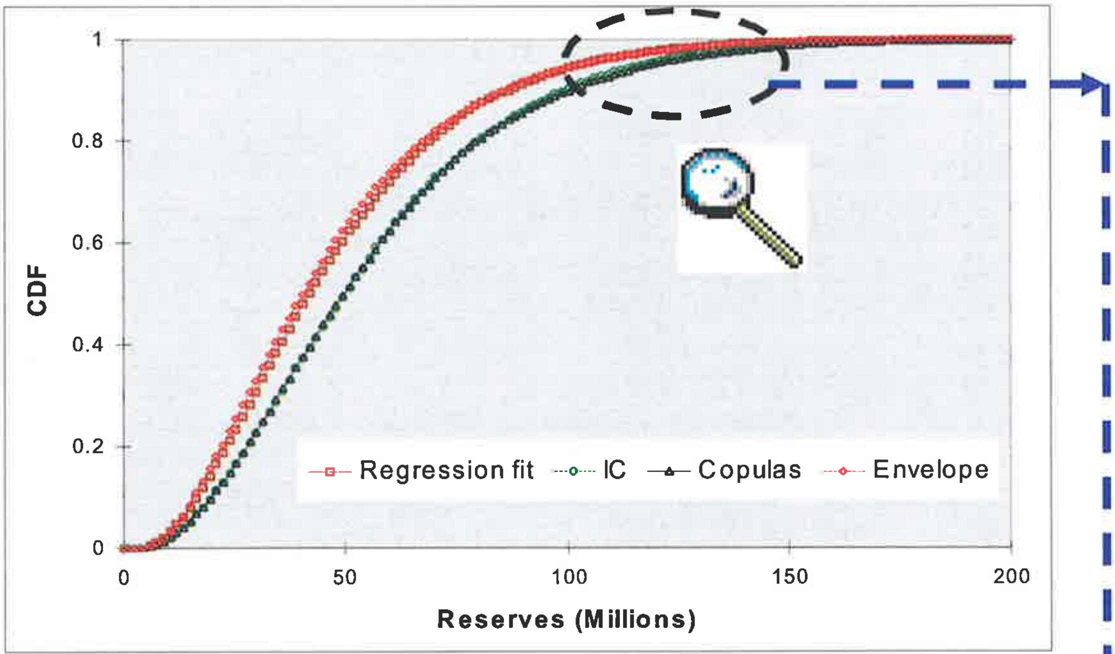


Figure 6-12. CDF's of technical reserves for dependency methods (upper tail)

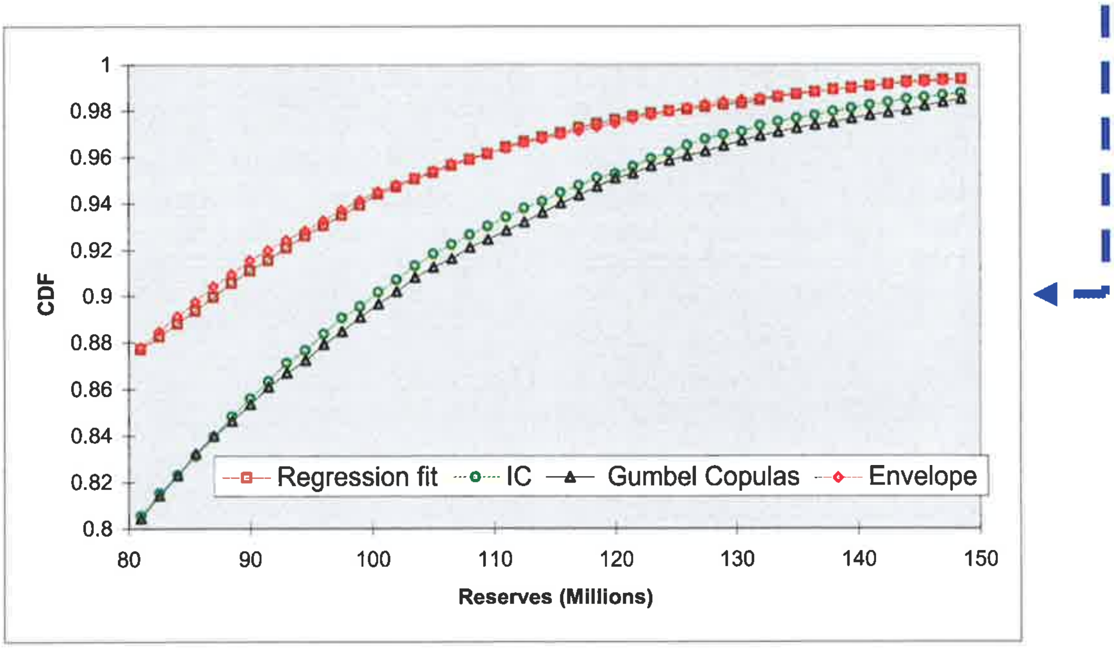


Figure 6-13. CDF's of technical reserves showing the upper tail

The previous analysis explored the upper tail dependence but in order to investigate how the different dependence methods will perform for lower tail dependence a different dependence structure for net thickness and recovery factor is created (Figure 6-14). This dependence structure had the same marginal distributions (net thickness with triangular distribution and recovery factor with triangular distribution) and the same rank correlation (spearman rank correlation = 0.67) as the previously used correlation in the original data (Figure 6-1) and Table 6-1. The only difference is the shape of the dependence structure which is a lower tail dependence structure.

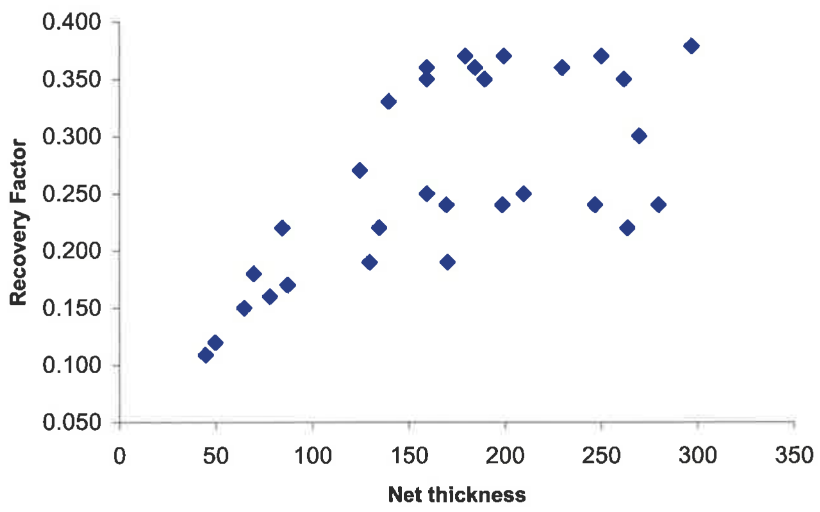


Figure 6- 14. Positive dependency between net thickness and recovery factor

The graphical (Q-Q plots) approach (Figures 6-15 to 6-18) indicates that the Clayton copula fitted the original data best, whilst the other methods showed large discrepancies, in particular below the 40th percentile probability. Among the Archimedean copulas the Clayton copula seems to best fit with a MD of 0.0287 compared to the Gumbel copula with a MD of 0.12 and the Frank copula with a MD of 0.092 (Figure 6-19). The Gumbel copula, as expected in this case was the worst because it is designed to capture upper tail dependence, not lower tail dependence.

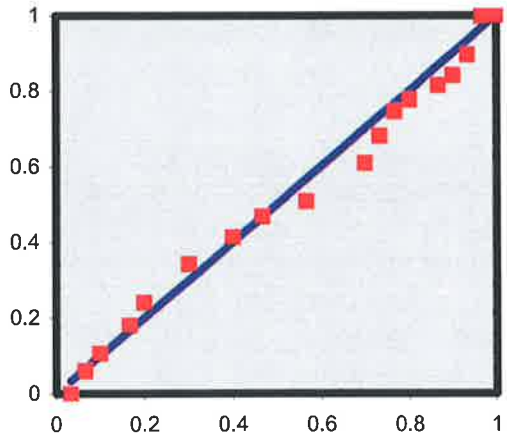


Figure 6-15. Q-Q Plot Clayton copula

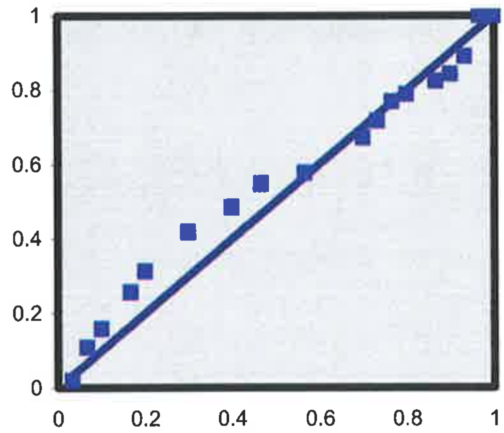


Figure 6-16. Q-Q Regression fit

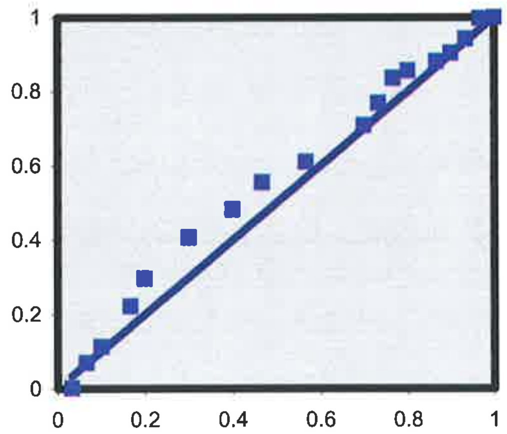


Figure 6-17. Q-Q Plot Envelope

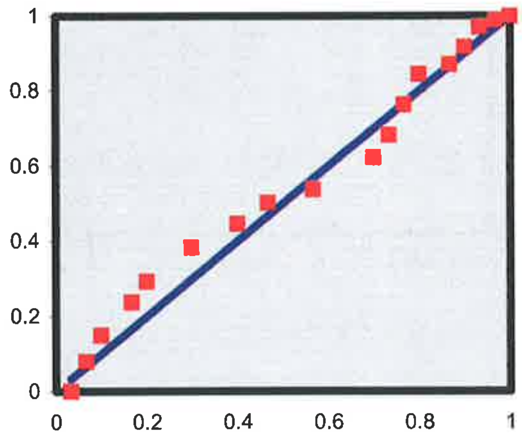


Figure 6-18. Q-Q Plot Iman-Conover

Among the dependency methods, the quantitative analysis showed that the Clayton copula had a MD of 0.0287, the Envelope method a MD of 0.052, the Iman-Conover method a MD of 0.042, and the regression fitting had a MD of 0.063 as shown in Figure 6-19. Again, a copula model captured the dependence in the lower tail better than the other models.

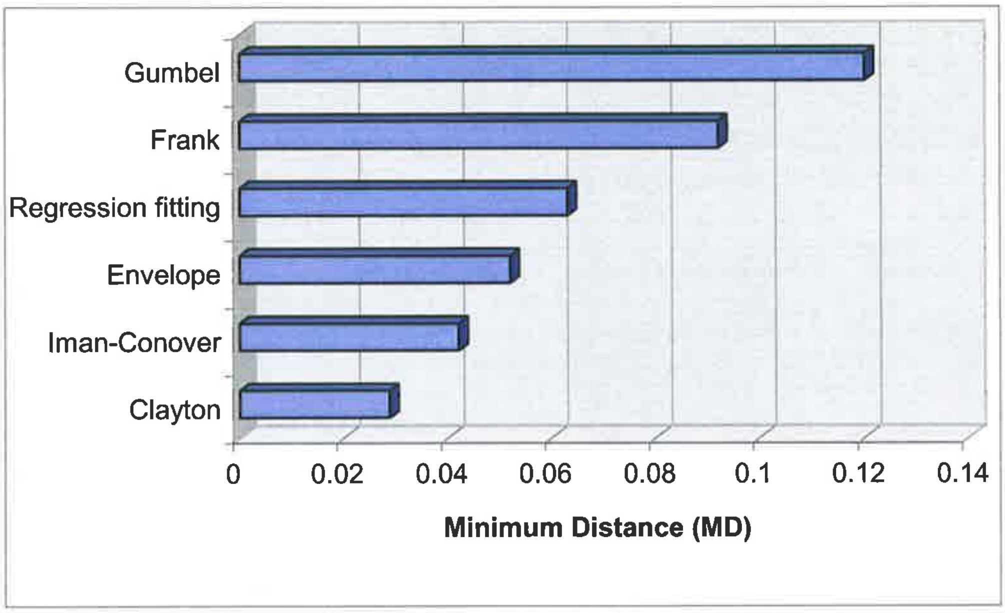


Figure 6-19. Minimum Distance for all the dependencies methods for lower tail

The next step was to investigate the impact of the different techniques on the technical reserves. As shown in Figure 6-20, below the 20th percentile, all of the dependency methods seem to have similar results to the copulas with a range difference of 1 to 3 million barrels; while above the 50th percentile the difference grew to 8 million barrels for the envelope and regression fitting. In this case the Iman-Conover and the copulas again seem to be closer to each other than the other techniques. Investigating the lower tail dependence by zooming in at the 5th percentile, Figure 6-21, the copulas approach responded best to the lower tail dependency but, when compared with the Iman- Conover, the difference is only 2 to 3 million barrels, which is very small. This showed that the copulas approach is superior for capturing lower tail dependence structure compared with the existing methods. Yet the impact of capturing the dependence structure, is low compared to the Iman- Conover method.

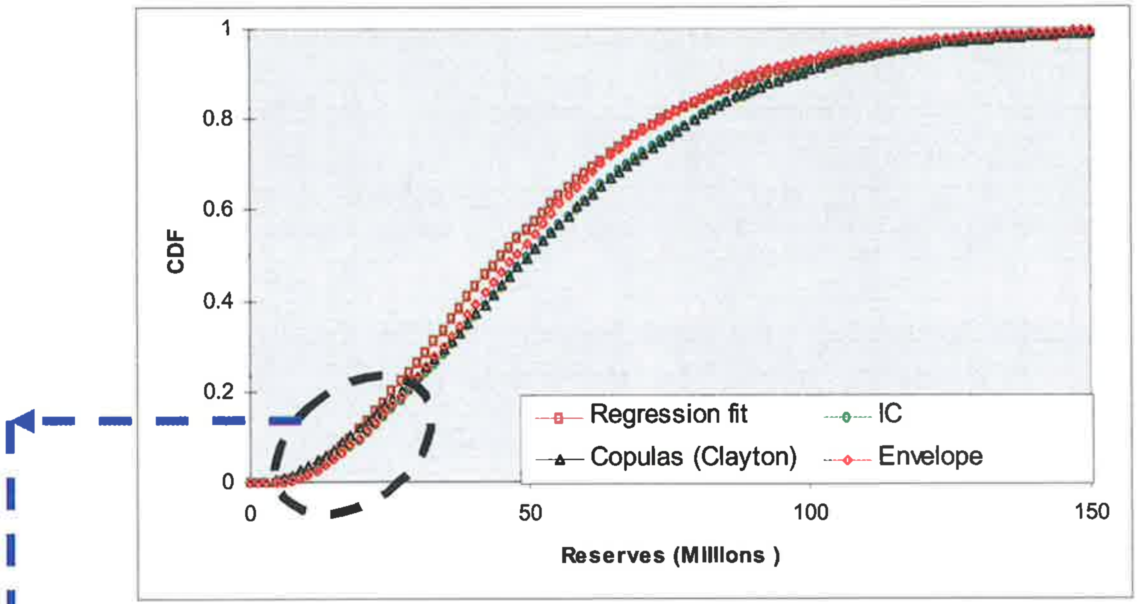


Figure 6-20. CDF of technical reserves and dependency methods (lower tail)

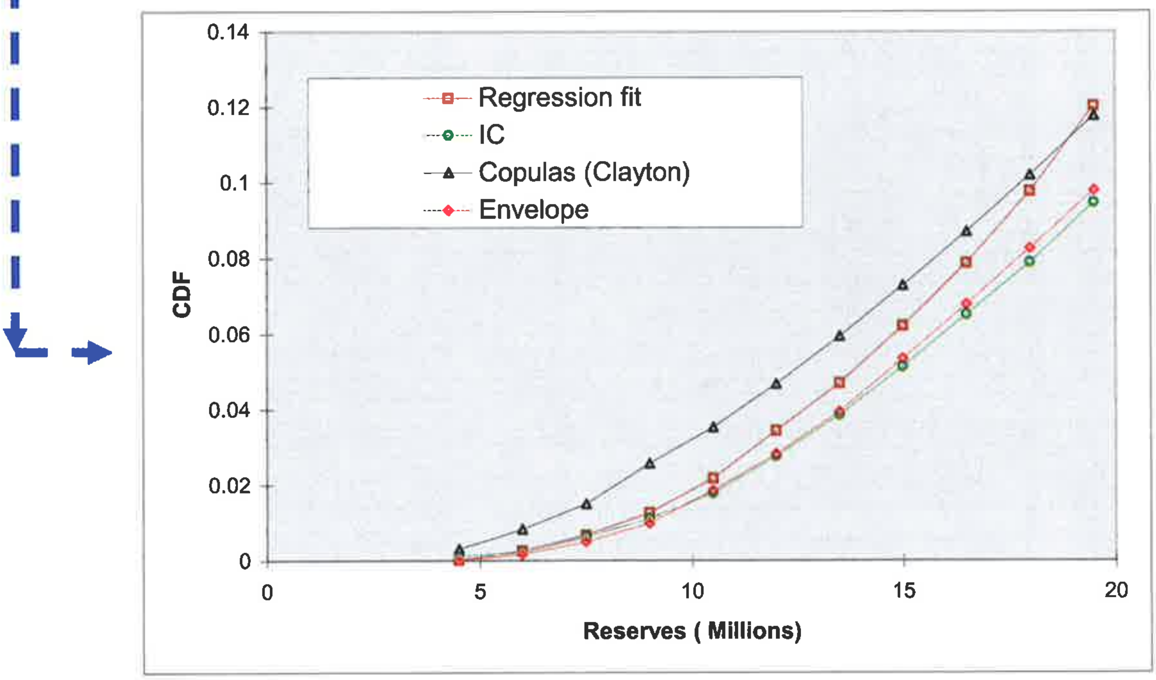


Figure 6-21. CDF of technical reserves showing lower tail dependence

It is important to realize that Figure 6-14 has the same distribution with same mean and standard deviation as Figure 6-1. Furthermore, the rank correlation is the same. Only the dependence structure is different. Indicating that, for the data with important upper tail characteristics (Figure 6-1) and data with important lower tail characteristics (Figure 6-14) the Iman-Conover results are essentially the same. The Iman-Conover method does not capture the correlation pattern as the copulas methods do it and does not distinguish between upper and lower tail dependency structures. It uses a joint distribution that does not represent the original distribution for either lower or upper tail dependence structure. Regression fitting, in both cases, assumes that the dependence structure is joint normal no matter what the dependence pattern is; whereas the Envelope method can be tailored to capture various dependence structures. However, the subjectivity in drawing the bounding envelope makes it more cumbersome, less rigorous and appealing than the copula approach. Copulas should be the choice where upper or lower tail dependency patterns are important. However, in our example, the impact of the dependency structure is small and it is not significant with a difference of only 1 to 3 million barrels (2%) between the Iman-Conover and the copula method.

It is important to point out that this thesis has chosen to investigate the impact of statistical dependence at the reserves level only. The result obtained at the reserves level can be generalized at the project level as well. We believe that at the project level the impact of copulas and Iman-Conover methods will be close to each other for Archimedean copulas similar to that at the reserves level. In addition the copulas approach will outperform the Envelope and Regression fitting methods at the project level similar to that at the reserves level.

6.4 Conclusion

This Chapter has illustrated the potential impact of the choice of dependence models in the probabilistic simulation of reserves and discussed the issue of modelling dependence in the probabilistic reserves estimates. The analysis compared and contrasted the most commonly used approaches for correlation modelling with the newer copulas models.

Firstly, this study has confirmed that dependencies are important and can have a large impact on the metrics used for decision-making, particularly for those variables that impact on the output the most.

Secondly, the copulas approach is technically superior in capturing dependency patterns in the lower and upper tail based on their ability to reproduce stochastically the original data with lower and upper tail dependence.

Thirdly, this Chapter has illustrated that the most commonly used methods in the oil and gas industry (Iman-Conover, Envelope and Regression fitting methods) fail to capture the upper and lower tail dependency patterns. This can lead to significant errors in the simulated results when comparing copulas with the Regression fitting and the Envelope method. However, when compared with the Iman - Conover model, the impact of the dependence structure was not significant, the difference in estimates being less than 3 million barrels (2%).

From a theoretical point of view, the priority of the methods to model dependencies is seen as:

1. The copulas method, because it captures dependence structure, provides technically more accurate results.

2. The Iman-Conover method is the second choice because although it does not capture dependence structure as well as copulas, it still provides quite accurate results.
3. The Envelope method with multiple lines is the third choice because it has the flexibility of producing very good results and captures the dependence structure. However it is subjective
4. Finally, the Regression fitting is ranked fourth because it changes the input distribution. Nevertheless, it produces reasonable results.

From a practical point of view, The Iman- Conover method is easier to run because the software to do this already exists. Using the copulas method needs more understanding of mathematics and modelling skills. If no software were available, the Regression fitting or Envelope methods could be used because they are very easy to construct in an Excel spreadsheet.

Based on the results of the impact of dependence structure and the copulas with oil and gas dependence methods at the reserve levels. These results at the reserves level can be generalized for Archimedean copulas at the project level.

CHAPTER 7

Significance of capturing functional dependencies and interactions

7. Introduction

Chapter 2 introduced the difference between stochastic and deterministic approaches and demonstrated that the stochastic approach is better in evaluating investment decisions. Chapter 4 built on this by introducing two integrated stochastic process approaches: the sequential approach and the systems approach. Furthermore, simple examples were used to show how both the sequential and the systems approaches capture the existence and non-existence of functional dependence. This Chapter uses the systems approach discussed in Chapter 5 and introduces the experiments, results and conclusions of comparing this approach with the sequential approach for computing NPV of a hypothetical development decision. Specifically, this Chapter presents the model developed for the first objective and addresses the third objective of this research.

7.1 Experiments

In order to investigate the third objective, that is evaluating the impact of functional dependencies and interactions, the following two experiments were conducted:

1. No functional dependence case
2. Inclusion of functional dependence.

Both experiments were conducted for both the sequential and the systems approaches.

7.1.1 No functional dependence experiment

This experiment focuses on answering the following question: what is the difference between the systems (Figure 7-1) and the sequential (Figure 7- 2) approaches on the NPV of a development decision?

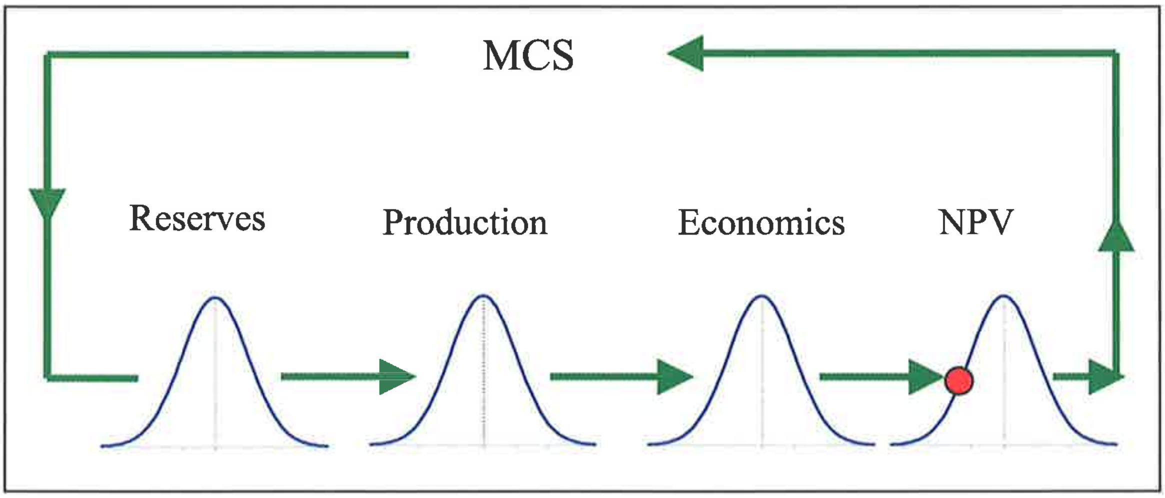


Figure 7-1. Systems approach; reproduction of Fig. 4-1

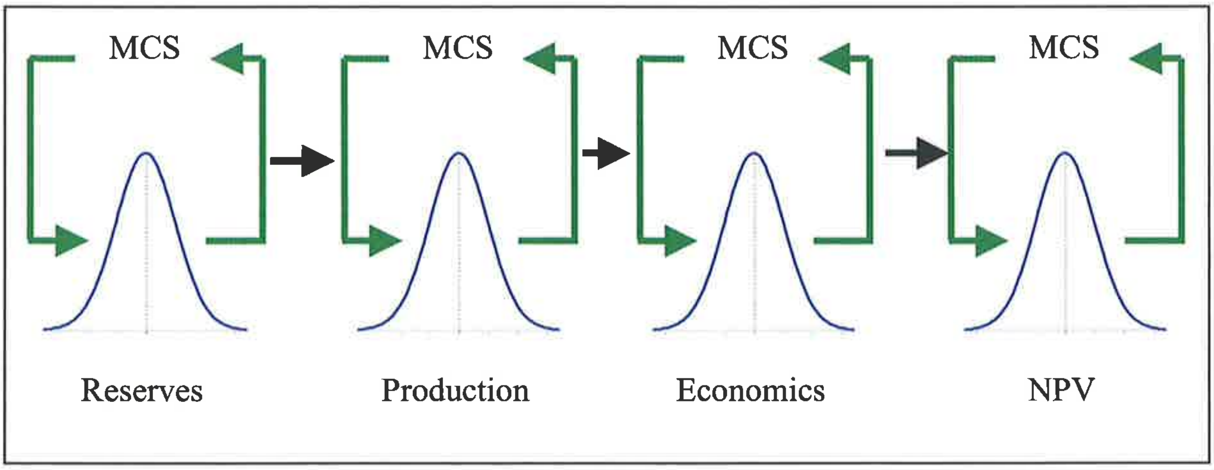


Figure 7- 2. Sequential approach, reproduction of Fig 4-2

7.1.2 Inclusion of functional dependence

This experiment also compares the sequential approach to the systems approach, but this time including functional dependence among the individual parameters of a development decision. Specifically, this experiment investigates the functional dependence within the reserve components as well as the functional dependence between the reserves and production parameters. Finally, functional dependence among parameters that are related to reserves, production and economics were also considered.

For parameters that are related between reserves and production, the following two functional dependencies were used.

a. Original Oil in Place (OOIP) and well productivity both being dependent on net reservoir thickness

b. Recovery factor and well productivity, both being dependent on permeability

For parameters within the reserves calculation, the following functional dependencies were used:

c. Original Oil in Place (OOIP) and recovery factor both being dependent on water saturation

d. Original Oil in Place (OOIP) and recovery factor both being dependent on porosity

For parameters that are related between reserves, production and economics the following functional dependencies were investigated:

e. Total number of wells as a function of production, reserves, well cost and oil price

7.2 Offshore development project: data and assumptions

The example used for the experiment is assumed to be an oil reservoir in an offshore field. The characteristics of this field are similar to those of an oil field in the North West shelf in Australia. The data used are representative of the oil reservoir. This is an oil field with low gas oil ratio. In this study, only oil production is modelled and gas production is ignored since it is small. The model includes four key elements: technical reserves, production, facilities and economic parameters. The input parameters for all the four key elements are the same for both the sequential and the systems approaches. The parameters listed in Table 7-1 were used as an input into the calculation of the Original Oil in Place (OOIP).

Table 7- 1. Original Oil in Place input variables and distributions

Variables	Distribution	Minimum	Most Likely	Maximum	Mean	Standard Deviation
Area (acres)	Lognormal				1000	300
Thickness (ft)	Triangular	50	160	250		
Porosity (fraction)	Triangular	0.3	0.35	0.4		
Water Saturation (fraction)	Triangular	0.3	0.35	0.4		
Formation volume factor (RB/ STB)	Triangular	1.15	1.18	1.2		

In order to calculate reserves, the inputs listed in Table 7 –2 were used as an input into the recovery factor calculations using the API correlation.

Table 7- 2. Recovery factor variable inputs

Variables	Distribution	Minimum	Most Likely	Maximum	Constant
Viscosity (cp)	Triangular	1.65	1.7	1.75	
Permeability (md)	Triangular	40	70	200	
Porosity (fraction)	Triangular	0.3	0.35	0.4	
Bubble point pressure (psi)					4000
Abandonment pressure (psi)					300
Water Saturation (fraction)	Triangular	0.3	0.35	0.4	
Formation volume factor (RB/STB)	Triangular	1.15	1.18	1.2	

This model assumes that all the wells have an initial average production and the inputs listed in Table 7-3 were used in calculating well productivity.

Table 7-3. Well productivity input variables and distribution

Variables	Distribution	Minimum	Most Likely	Maximum	Constant
Thickness (ft)	Triangular	50	160	250	
Permeability (md)	Triangular	40	70	200	
Vertical Permeability (md)	Triangular	20	35	50	
Length of horizontal section (ft)	Uniform	1000		2500	
Drainage radius (in)	Triangular	2000	2980	3960	
Wellbore radius (in)	Triangular	0.3	0.328	0.357	
Formation volume factor (RB/STB0	Triangular	1.15	1.18	1.2	
Average reservoir pressure (psi)	Constant				4000
Wellbore flowing pressure (psi)	Constant				3300

Drilling

Drilling assumptions are: number of days to drill a well is 49 days and there is only one rig available.

Production

Production starts in the third year. This is assumed and deterministic. Optimum number of wells for the field is six. Plateau production capacity was deterministically assumed to be 75,000 barrels per day.

Facility

An FPSO with a shuttle tanker as an export mechanism was assumed to be economically more attractive.

Economics

The following input and assumptions were made for economic variables

Table 7- 4. Economic variable input and distribution

Variables	Distribution	Minimum	Most Likely	Maximum	Constant	Units
Development well	Triangular	10	12	15		$ Million/well
Exploration well	Triangular	7	8	9		$ Million/well
Dry well	Triangular	5	6	7		$ Million/well
Discount rate					15%	Percent
Tax rate					34%	Percent
Shuttle tanker					45	$Million
Current Oil price					25	$/barrel
FPSO	Triangular	50	100	200		$Million
Fixed OPEX	Triangular	10	15	18		$Million/year
Variable OPEX	Triangular	0.1	0.5	0.6		$/barrel
Abandonment	Triangular	7	10	14		$Million
Subsea Equipment	Triangular	40	60	80		$Million
Other Capex	Triangular	3	5	10		$Million

Capital Expenditure (CAPEX) before Production

Table 7-5 indicates what percentage of the capital expenditure will be spent according to the number of years prior to the start of production. This is done for three cases, 3 years, 4 years and 5 years.

Table 7-5. Capital expenditure profile as function of production start up

Production Start, years	CAPEX, % of Total				
	Year 1	Year 2	Year 3	Year 4	Year 5
Year 3	40	60			
Year 4	40	40	20		
Year 5	20	40	20	20	
Production start up					

Production Capacity was assumed to be a function of CAPEX: Table 7- 6 lists each production capacity (barrels per day) with the amount of CAPEX to build the production facility.

Table 7- 6. Production Capacity vs incremental CAPEX

Production Capacity Barrel per day	CAPEX $Million
15,000	15
20,000	30
30,000	45
40,000	60
50,000	95
75,000	115
90,000	160
120,000	180
150,000	200

7.3 Results and discussion

This section will discuss the results from each of the experiments introduced previously for both the sequential approach and the systems approach. Furthermore, it will present step by step how the results were computed.

7.3.1 The no functional dependence (Interaction) case

Monte Carlo Simulations were used to generate the probabilistic reserves estimates from the sequential and the systems approaches. The sequential approach generates a separate distribution for OOIP using stochastic distributions of area, net thickness, porosity, water saturation and formation volume factor and a separate distribution for the recovery factor, then multiplying those two distributions together to calculate a distribution for reserves. The systems approach, however, generates a reserves distribution by sampling all the variables simultaneously. The two approaches produced essentially the same technical reserves distribution as shown in Figure 7-3 and Table 7-7 for all the main statistical parameters; mean, standard deviation, P10, P50 and P90. This clearly shows that there is no major difference in the probabilistic reserves estimates between the systems and sequential approaches, when dependencies are not taken into account.

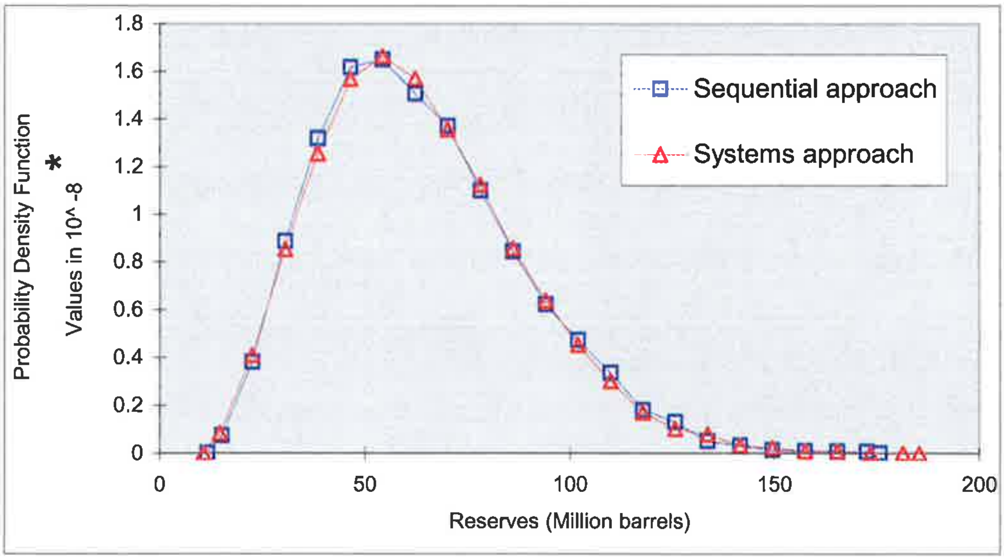


Figure 7-3. Technical Reserves: sequential vs. systems approaches
* Vertical axis: Probability Density Function (PDF) is the relative frequency value divided by the width of the bin (@Risk manual, Palisade group)

Table 7-7. Technical Reserves (Million barrels of oil)

Parameter/ Approach	Sequential	Systems
Mean	63.6	63.7
Std Deviation	24.7	24.9
10% Percentile	34.1	33.9
50% Percentile	60.5	60.5
90% Percentile	97.9	97.3

As the second step in the experiment, production profiles for both the sequential and the systems approaches were generated. In the sequential approach, the output distribution of reserves was used as input to generate the production profiles. For the systems approach all original variables contributing to production were sampled.

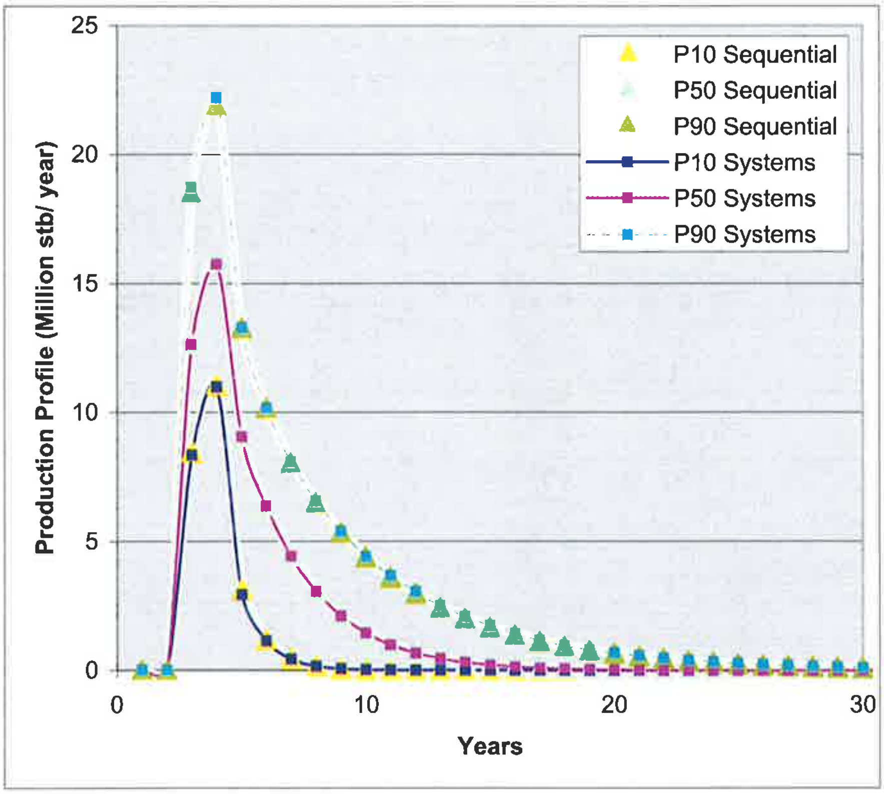


Figure 7-4. Production forecast: sequential vs. systems approach

Figure 7-4 compares the production profiles for the P10, P50 and P90 probability levels while Table 7-8 shows the ultimate recovery and illustrates that both the systems and sequential approaches produce the same production profiles, when no dependencies are involved.

Table 7-8. Production sum (Million barrels of oil)

Parameters/Approach	Sequential	Systems
Mean	63.8	63.9
Std Deviation	24.3	24.5
10% Percentile	35.0	34.9
50% Percentile	60.5	60.5
90% Percentile	97.8	97.3

As the third step in the sequential approach, the output production profile is used as an input into the economic model. Stochastically, this is not an issue for the systems approach and for the sequential approach it is easy to model deterministically. However, it is difficult if done stochastically, because the production distribution in each year is related to the cumulative recovery and therefore the production distributions in the preceding years. This is due to the fact that the production capacity is a function of the depletion level of the reservoir and implies that production in a given year cannot be modelled independently of the production in the preceding years.

To overcome this problem with the sequential approach, the complete range of production profiles (from 1% to 99% probability ranges) was linearly proportioned to generate the production profiles for various probability levels. Each production profile generated is used to calculate NPV. However, another problem is that other parameters in the economic model such as FPSO CAPEX, development well CAPEX and abandonment costs are all stochastic. Again to overcome the problem of using production percentiles with stochastic costs, proportionally distributed percentiles of all the costs were generated and used with the corresponding percentile production profile to generate percentiles of NPV. The final challenge was that even with this, the multiplication of all P10 inputs together does not generate the P10 of the output. So using the Capen approximation (Rose, 2001) to relate between input and output percentiles, as an example, an output with three uncertain variables the multiplication of the three P10's is equal to P1.3 (P10*P10*P10 = P1.3). Using this Capen approximation the corresponding NPV values were generated and then used to construct the NPV distribution to be compared with the NPV distribution calculated from the systems approach.

As stated previously, none of these complications are relevant for the systems approach. The results clearly show that in the absence of interactions there is no significant difference between the NPV distribution produced by the sequential and the systems approaches (Figure 7-5) and Table 7-9 shows the main statistical parameters for both approaches.

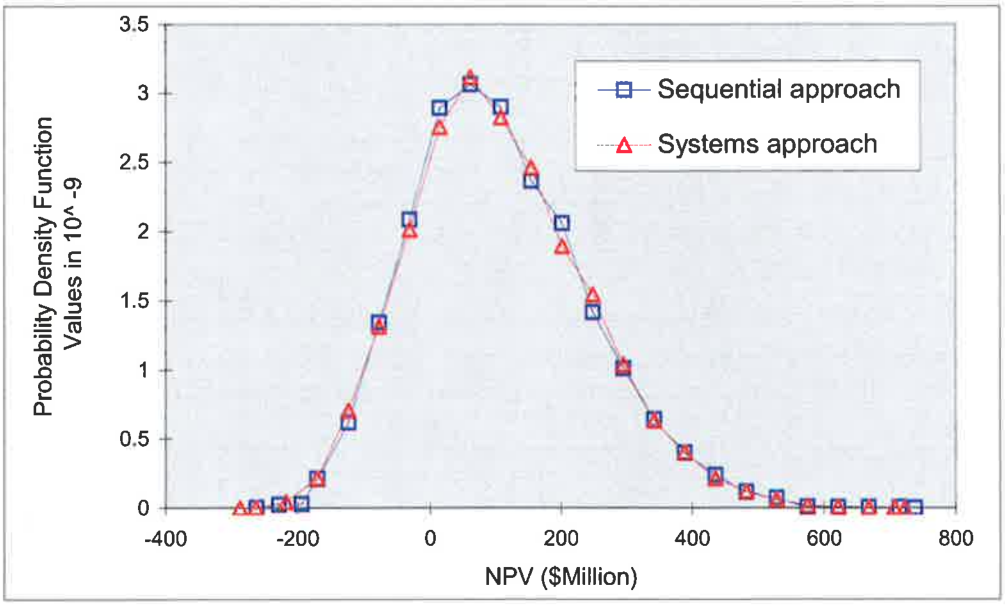


Figure 7-5. NPV: sequential vs. systems approach

Table 7- 9. Net Present Value ($Million)

Parameters/Approach	Sequential	Systems
Mean	107	107
Std Deviation	133	135
10% Percentile	-55.7	-58.3
50% Percentile	93.0	95.2
90% Percentile	287	286

The sequential approach uses a sample of area, thickness, porosity, water saturation and the formation volume factor to generate an output distribution of OOIP through multiple iterations. It then independently produces an output distribution for well productivity by sampling all the inputs such as net thickness, permeability and viscosity. In the sequential approach, there are two simulations independent of each other. In the system approach, there is only one simulation, which concurrently generates distributions for OOIP and well productivity by sampling the common net thickness variable independently for the two calculations. However, because there is no interaction between the net thickness used for OOIP and well production capacity, both the sequential and systems approaches produce the same results. The same logic applies for economic calculations.

This is a very important conclusion because it shows logically and quantitatively that the systems approach without interaction is equivalent to the sequential approach.

7.3.2 Sensitivity analysis

After generating the net present values for both approaches, a sensitivity analysis was conducted for both results to identify the variables, that have the greatest impact on the NPV using the advanced analytical capabilities of @Risk™. Figure 7-6 and 7-7 show the correlation coefficient between input variables and the computed

NPV. The sequential approach generates a sensitivity analysis, which shows that OOIP is the most uncertain variable affecting the NPV, followed by well productivity, FPSO CAPEX and recovery factor. Other parameters have minor impact (Figure 7-6). For the systems approach, thickness, area and permeability are the most important factors affecting the NPV followed by FPSO CAPEX, porosity and the length of the horizontal wells. The rest of the variables have minor impacts (Figure 7-7).

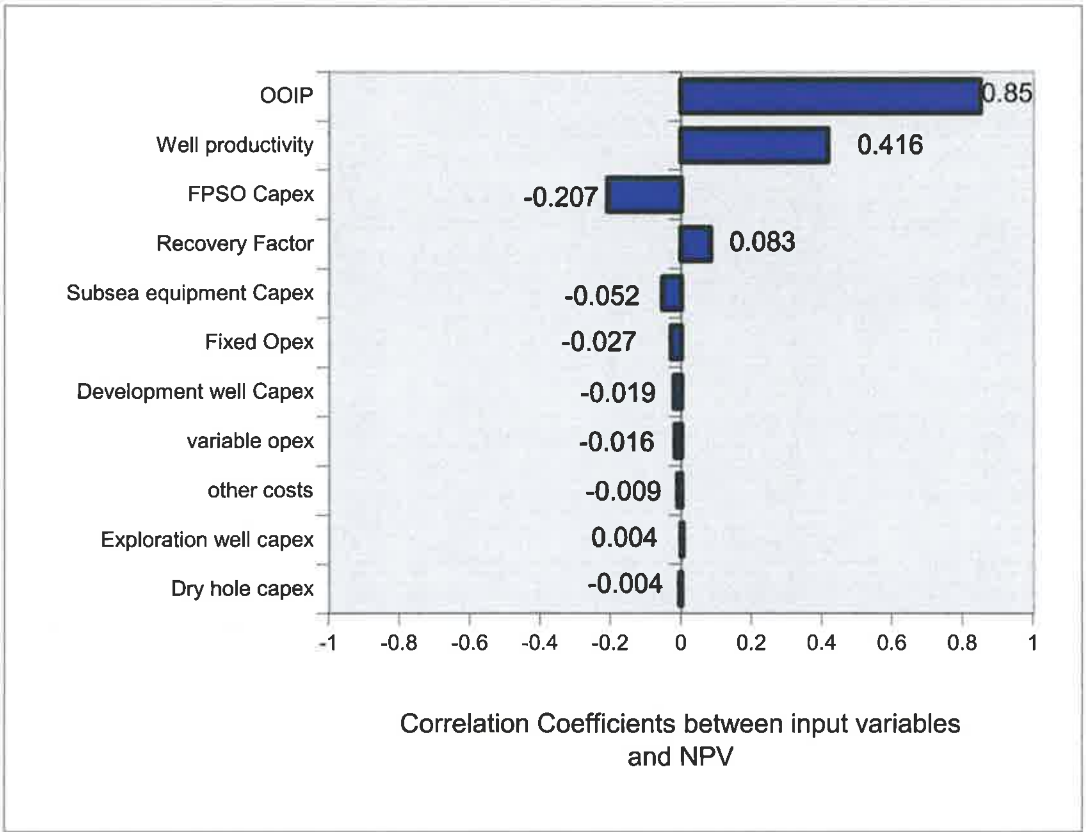


Figure 7-6. NPV sensitivity analysis for sequential approach

The systems approach is better at identifying the basic variables that have the biggest impact on NPV. This is an important aspect in field development planning, because there is a need to know which variables have the most impact on the NPV.

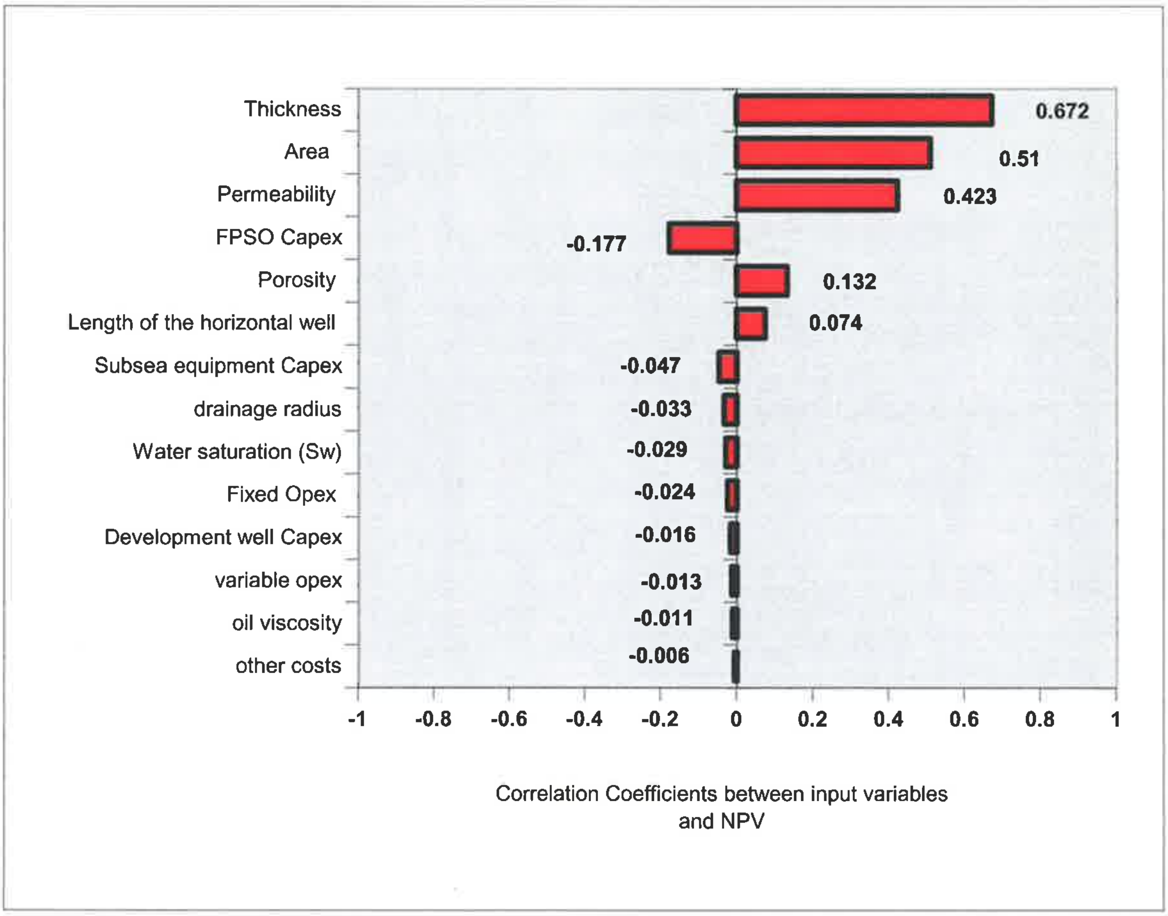


Figure 7-7. NPV sensitivity analysis for systems approach

Whereas the systems approach clearly shows the impact of each individual variable on the NPV, the sequential approach only shows the sensitivity of the sequential steps carried for the analysis. For example, for the sequential approach, it is clear that the OOIP is the most sensitive variable that impacts the NPV, but looking at the tornado chart it is not possible to know which variables in the OOIP calculation contribute most to this impact, it could be porosity, area or thickness (Figure 7-6). It is also important to point out that at a higher sequential step, the decision maker will not be able to observe any parameters that relate to reserves or well performance. At a higher sequential steps only yearly production profile and economic parameters will be observed. The systems approach, however, clearly shows that thickness and area are the most sensitive variables that impact the NPV (Figure 7-7).

7.3.3 The functional dependence case

We now introduces the functional dependencies described on section 7.1.2 and compare the results using the methods described by Figure 7-1 and Figure 7-2. This section will investigate the impact of introducing each functional dependence and interaction and see how that impacts upon the output.

7.3.3.1 The functional dependence case: Thickness

This case investigates the impact of the interaction between OOIP and well productivity (which are both functions of reservoir thickness) using the same process as discussed before, for both the sequential and the systems approaches. There is a marked difference in the resulting NPV probability density function between the two approaches, as shown in Figure 7-8 and Table 7-10.

The interaction between OOIP and well productivity results in an increase in the mean and standard deviation of the NPV distribution. The sequential approach underestimates the mean by 6 % and the standard deviation by 12 % in this example. The higher standard deviation of the results of the systems approach shows that it captures more of the uncertainty range than the sequential approach. The difference in the P10 and P90 range between 10 – 27 % between the two approaches.

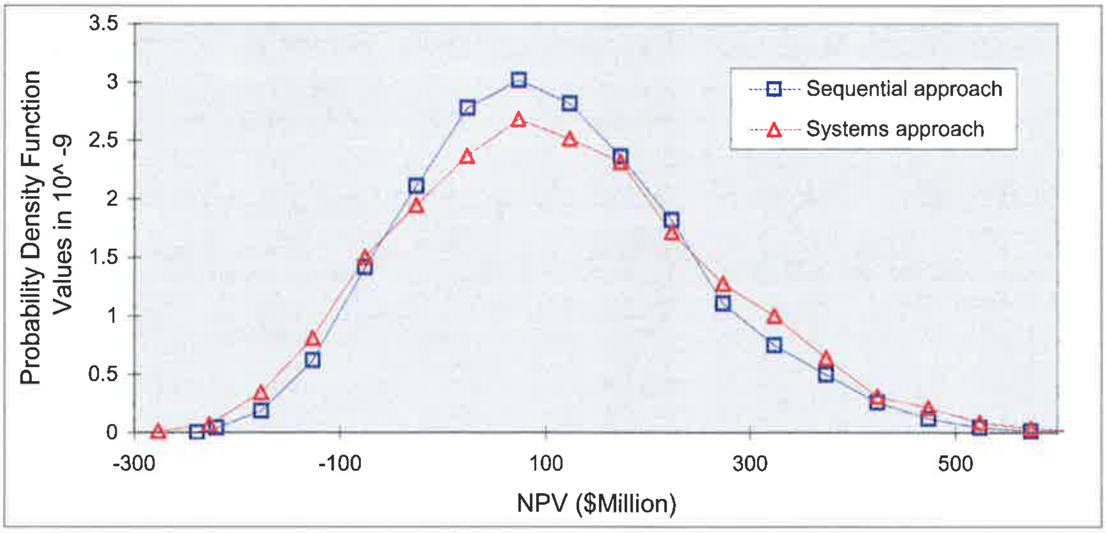


Figure 7-8. Thickness interaction: NPV sequential vs. systems approach

Table 7-10. NPV ($ Million) Thickness interaction

Parameters/ Approach	Sequential	Systems
Mean	106.69	113.13
Std Deviation	133.09	148.97
10% Percentile	-58.67	-74.06
50% Percentile	96.03	102.59
90% Percentile	285.91	315.78

Figure 7-9 (a) shows the calculated values of well productivity when the values of net reservoir thickness are sampled independently of the OOIP calculations. Figure 7-9 (b) on the other hand shows a similar plot of the calculated well productivity versus net reservoir thickness when the net reservoir thickness values used are the same as those used for OOIP calculations. Figure 7-9 (a) clearly shows no trend when the interaction is ignored (the sequential approach). The results of the systems approach however Figure 7-9 (b) show a clear logical trend, which takes the interaction into account.

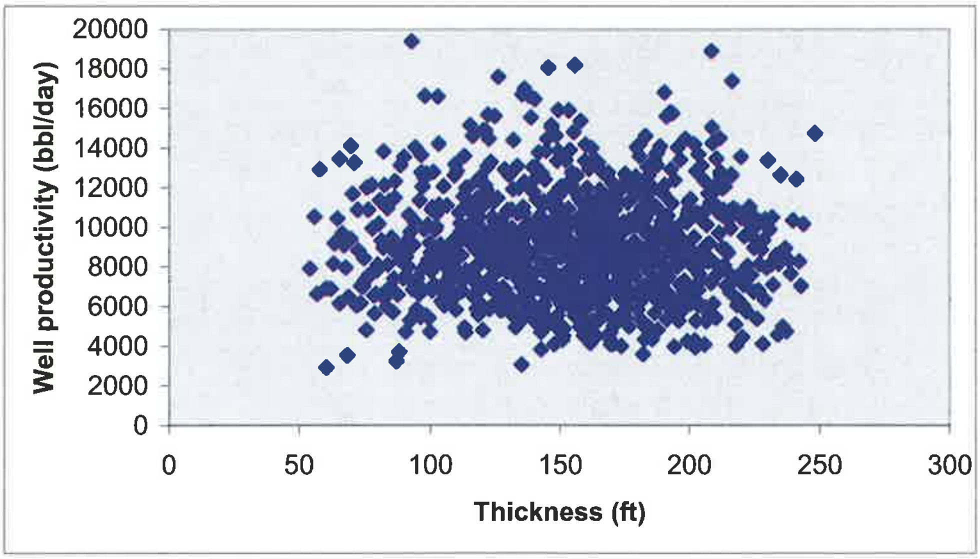


Figure 7-9 (a). Sequential approach: thickness interaction

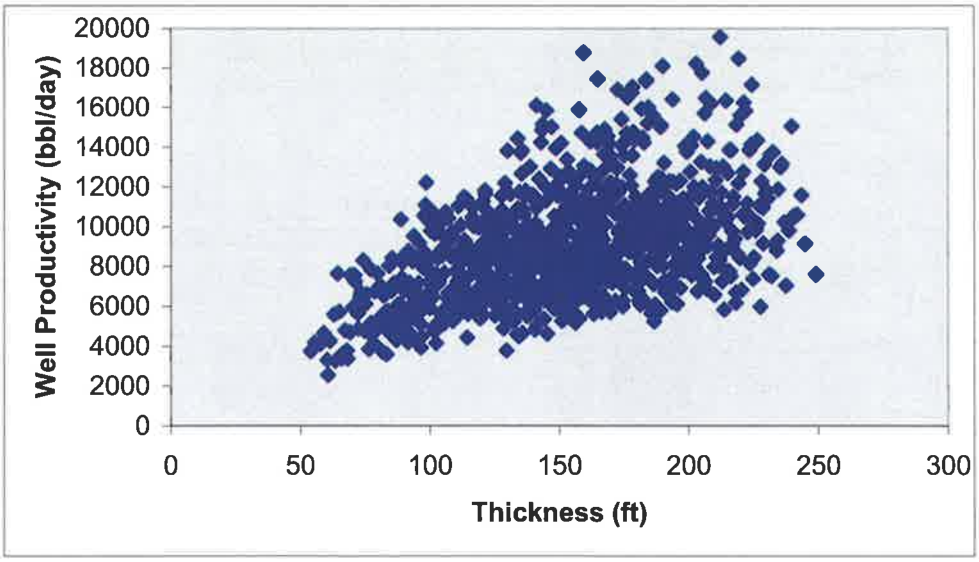


Figure 7-9 (b). Systems approach: thickness interaction showing a clear trend

7.3.3.2 The functional dependence case: Permeability

In this case the interaction between the recovery factor and well production capacity, both being functions of permeability is investigated for both approaches. The results are shown on Figure 7-10 and Table 7-11.

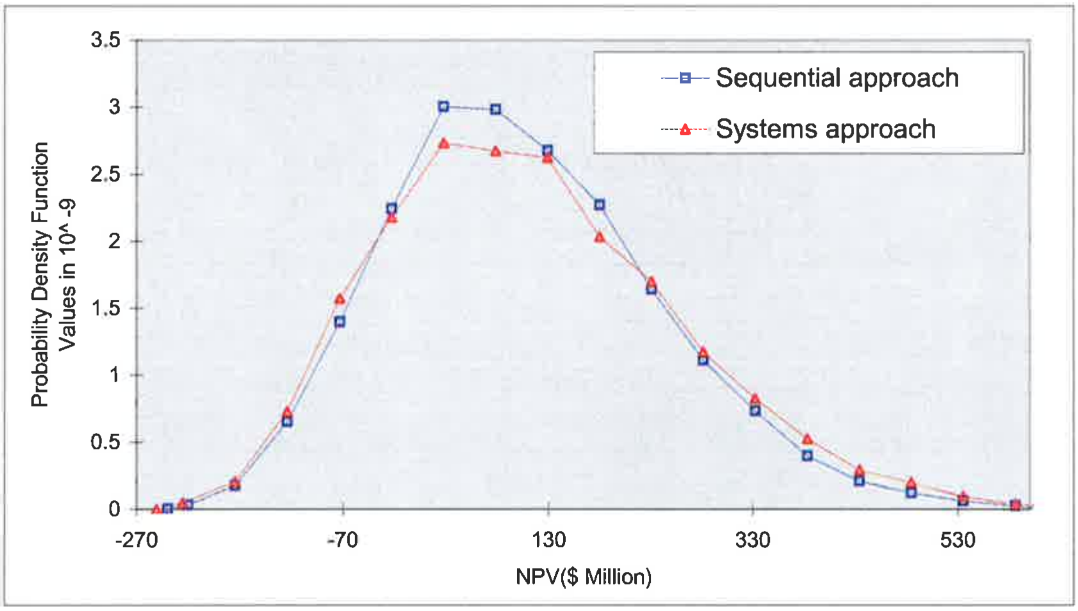


Figure 7-10. Permeability interaction: Net Present Value (NPV) sequential vs. systems approach

The sequential approach underestimates the mean by 6% and the standard deviation by 9%. Because of the higher standard deviation, the extreme values are captured well with the systems approach rather than with the sequential approach. The difference in P10, P50 and P90 values between the two approaches ranged from 5-12%.

Table 7-11. NPV ($ Million) Permeability interaction

Parameter/ Approach	Sequential	Systems
Mean	106.78	113.08
Std Deviation	133.44	145.59
10% Percentile	-56.71	-63.67
50% Percentile	94.33	99.00
90% Percentile	288.04	310.69

7.3.3.3 The functional dependence case: Water saturation

This case investigates the impact of interaction between OOIP and recovery factor, water saturation being an input into both calculations. There is very little difference in the result as shown in Figure 7-11 and Table 7-12. This is due to the fact that the recovery factor is a very weak function of water saturation in the API equation ($RF\, \alpha\, (S_{wi})^{0.3722}$).

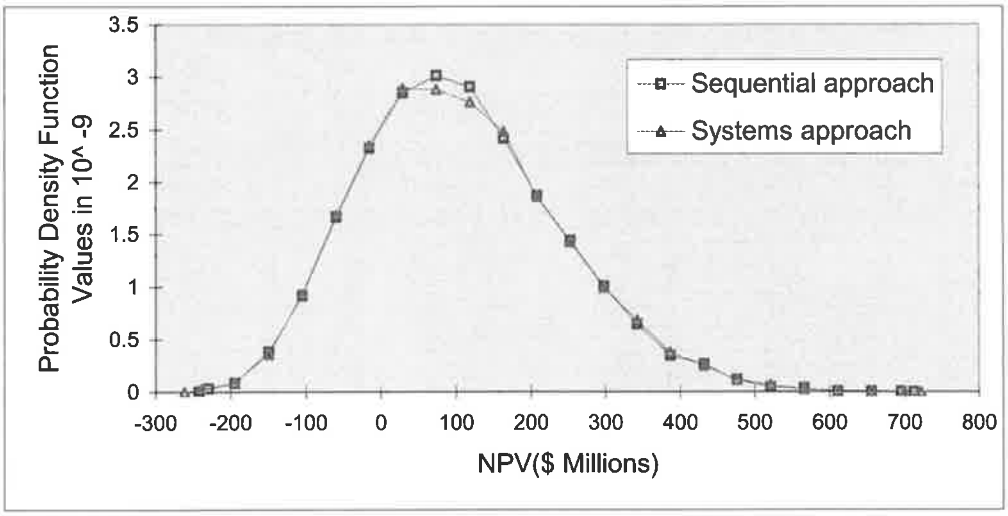


Figure 7-11. Sw interaction: NPV sequential vs. systems approach

Table 7-12. NPV ($ Million) S_w interaction.

Parameter/ Approach	Sequential	Systems
Mean	106.46	107.33
Std Deviation	133.28	135.27
10% Percentile	-56.12	-59.27
50% Percentile	95.52	95.83
90% Percentile	284.80	289.49

7.3.3.4 The functional dependence case: Porosity

This case focuses on the interaction between OOIP and the recovery factor with porosity being the interaction parameter. Once again the two approaches show very little difference as shown in Figure 7-12 and Table 7-13. This small difference is, again, due to the fact that recovery factor is a weak function of porosity ($RF\alpha\ (\Phi)^{0.1611}$) in the API equation.

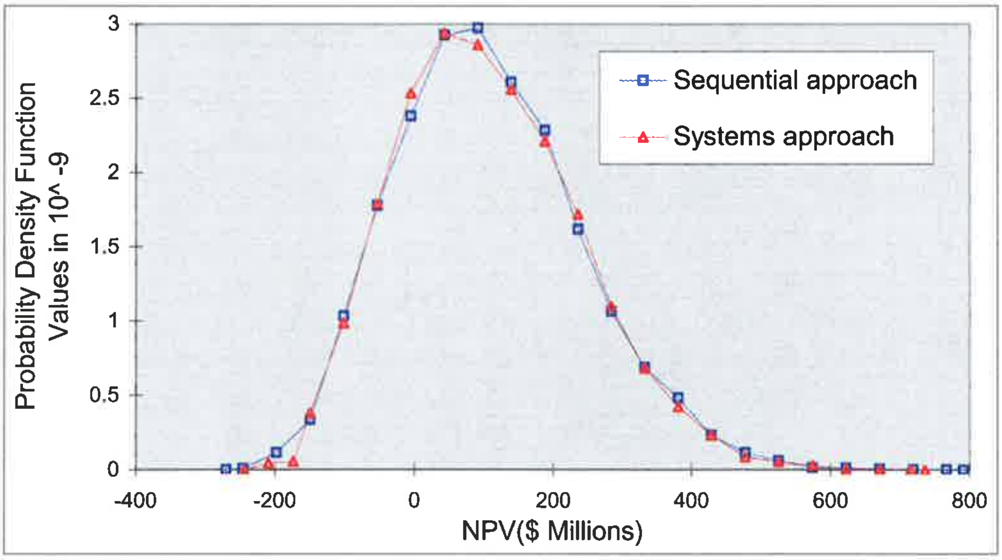


Figure 7-12. Porosity interaction: NPV sequential vs. systems approach

Table 7-13. Porosity interaction NPV ($ million)

Parameter/ Approach	Sequential	Systems
Mean	107.24	106.50
Std Deviation	134.25	133.44
10% Percentile	-58.16	-57.68
50% Percentile	96.97	94.72
90% Percentile	286.25	284.02

7.3.3.5 The functional dependence case: Total number of wells

The sequential approach assumes that the number of wells are determined once and used from that point forward in the simulation. This is due to the nature of the sequential approach where each discipline input is evaluated individually and separately. The systems approach assumes that the number of wells is a function of the reserves, well productivity, well costs (both capital and operating) and oil price.

A mean reversion oil price model uses the current oil price, long-term oil price, volatility and reversion factor which brings the oil price from its high or low values to the mean at the end of the assigned time period. This model attempts to mimic the OPEC price management policy as explained in Chapter 2

The price was modelled using the Mean Reverting technique with an assumption that a current price of $25 per barrel and a long term price of $40 per barrel with a minimum volatility of 15%, most likely volatility of 30% and maximum volatility of 40% and a reversion half life that lies between 1-2 years (Figure 7-13).

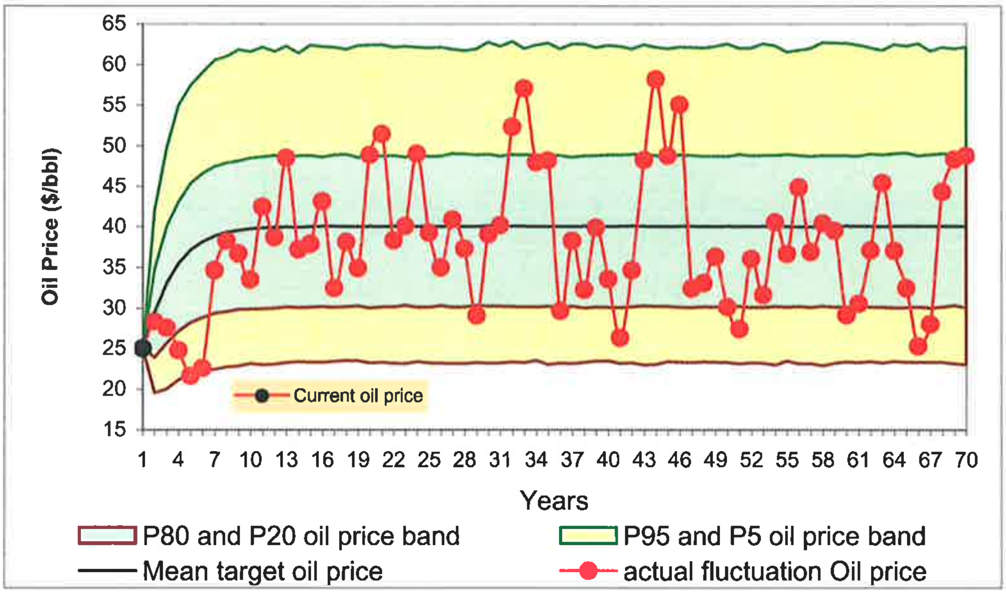


Figure 7-13. Mean reverting price model

Figure 7-13 shows one realization of the oil price with fluctuations both above and below the mean target oil price bandwidth generated through the mean reversion oil price model.

The results of the systems and sequential approaches for the total number of wells interaction are markedly different, with the NPV distribution for the systems approach shifting to the right (Figure 7-14) and Table 7-14. The sequential approach underestimates the mean by 40% and the standard deviation by 29% compared to the systems approach in our example. In addition, the systems approach captures the uncertainty on the high side better than the sequential approach. The difference in the P10 and P90 ranges between 0 - 34% between the two approaches.

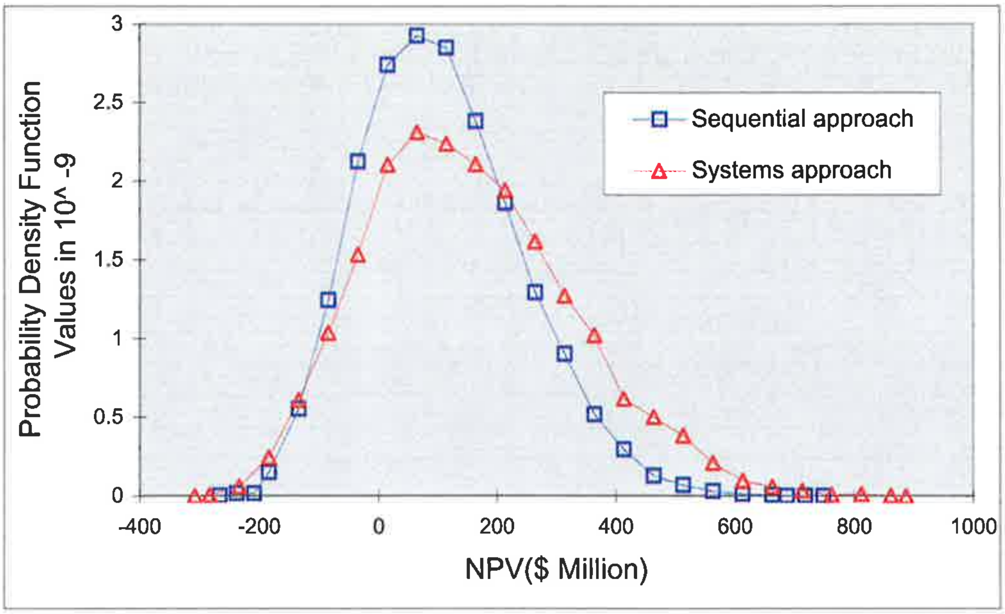


Figure 7-14. Total No. of wells interaction: Net Present Value sequential vs. systems approach

Table 7-14. NPV ($ Million) Total number of wells interaction

Parameter/ Approach	Systems	Sequential
Mean	154.14	106.39
Std Deviation	173.07	133.99
10% Percentile	-56.87	-57.30
50% Percentile	137.64	95.73
90% Percentile	386.33	287.74

7.3.3.6 All functional dependencies combined

All the interactions discussed above (i.e. OOIP with initial well rate, recovery factor with initial well rate, OOIP with recovery factor and total number of wells with reserves, production, CAPEX and oil price) were combined together, and another simulation was run to compare the impact of including or excluding them. The results show that ignoring interactions underestimates the mean NPV by 54 % and underestimates the standard deviation by 44% (Figure 7-15). Furthermore, the P10, P50 and P90 values of NPV are all underestimated by approximately 20%, 50% and 50% respectively (Table 7-15). These results clearly show that the impact of dependencies and interactions can be very significant.

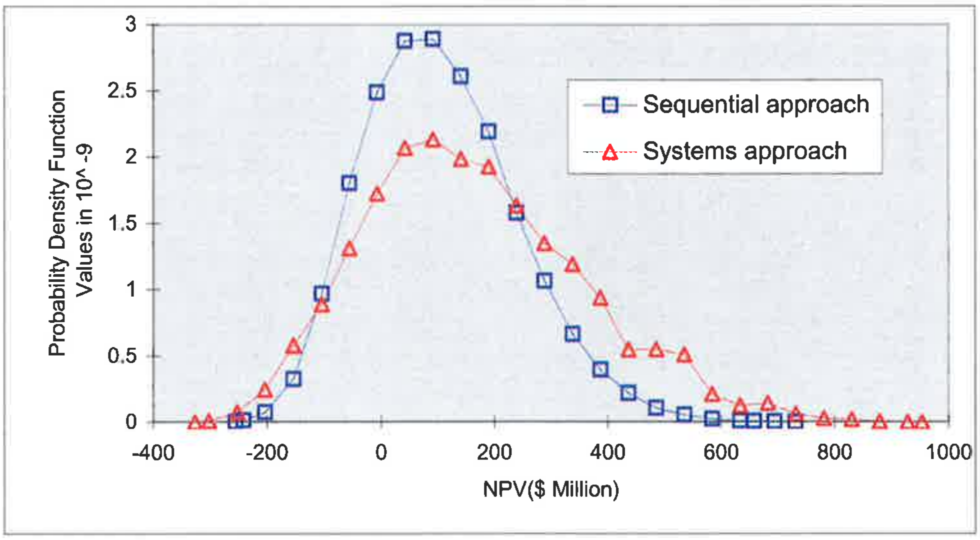


Figure 7-15. Impact of all dependencies and interactions on NPV

Table 7-15. NPV ($ Million) All interactions

Parameter /Approach	Systems	Sequential
Mean	164.63	106.38
Std Deviation	193.29	134.36
10% Percentile	-70.04	-58.28
50% Percentile	144.24	95.21
90% Percentile	428.36	285.33

Generally, it might be expected that the difference between the systems and the sequential approaches with interaction present could be attributed also to the non-linearity of the functional form. This was investigated and found that even with a linear function, in the presence of interaction, the results between the systems and the sequential approaches are different. An example to illustrate this is shown. Given X and Y are triangular distributed (Table 7 –16)

Table 7- 16. Linear function example input variable

Variables/values	Minimum	Most likely	Maximum
X	50	80	100
Y	100	300	500

where the functional forms are linear and given by

$$R = X + Y$$
$$P = R + Y$$

Table 7-17. Summary of Results

Experiment/Approaches	Sequential	Systems
Linear function results	Mean P = 676 SD = 163	Mean P = 676 SD = 118

Clearly, this example shows that the even if the functional form is linear, while the mean is almost the same, the standard deviation is different (Table 7-17). This example, indicates that the functional form does not have to be non-linear for sequential and systems approaches to yield different results.

7.4 Conclusions

This Chapter has discussed the impact of input parameter dependencies and interactions on NPV through the analysis of a hypothetical offshore field development and concludes the following:

- The impact of dependencies and interactions on NPV can be significant and could be material to development decisions. The systems approach captures interactions and dependencies while the sequential approach ignores them. Ignoring interactions underestimates the mean as well as the standard deviation (by 54% and 44% respectively in our example). Furthermore, the P10, P50 and P90 values of NPV are all underestimated by 20%, 50% and 50% respectively in the sequential approach. The wider standard deviation of the systems approach shows that it honours the increased uncertainty when interaction occurs.
- In the systems approach the standard deviation and the mean value of the output are greater than the corresponding values calculated using the sequential approach. The systems approach yields higher values of the standard deviation because it captures the uncertainty and the functional dependency between variables whereas the sequential approach ignores them. The systems approach also yields higher values of the mean due to the non-linearity of the functional relationship as discussed in section 7.3.3.6 and shown in Table 7-17. Where the functional dependencies are linear, the calculated mean values using either the systems approach or the sequential approach are almost the same but the standard deviation for the systems approach is still higher.

- In a stochastic environment, where functional dependence and interaction amongst input parameters are ignored, both the sequential and the systems approaches yield almost identical results. The difference is merely statistical. In the presence of functional dependence and interaction, the systems approach is superior in capturing their impact.
- The impact of the functional dependence and interaction on the mean and standard deviation of the development decision depends upon:
    - **i.** The modelling approach, the systems approach captures interaction where as the sequential approach does not.
    - **ii.** The functional form of the interaction.
    - **iii.** The sensitivity of a variable. Variables that have higher impact on the output will have a higher interaction impact compared to variables that have lower impact on the output variables.
    - **iv.** Even if the functional form of the interaction is linear the systems and the sequential approach yield different results.

CHAPTER 8

Portfolio optimisation: systems vs. sequential approach

8. Introduction

As discussed in previous Chapters, traditional investment decisions in the oil and gas industry often ignore modelling of dependencies and interactions among decision parameters such as reserves, production, facilities and costs. This limits the ability to examine how a change in one parameter may affect the others. It was demonstrated that for a single project, the systems approach better captures the impact of dependencies compared to the sequential approach. In reality, however, companies are not only faced with evaluating single projects but with many projects that compete for funds. The decision maker must choose a portfolio of projects, which balance risk and reward and together contribute most to the organization's profit.

This Chapter presents the results of the fourth objective of this research, which investigates the impact of the systems and sequential approaches at the portfolio level and shows how the difference in the approaches impacts the decision-making at the portfolio level.

This research argues that portfolio models should not only recognize the impact of inter-dependence of properties due to parameters such as geological and political environment but also simultaneously recognize the impact of intra-dependence of calculated parameters within each project. The author believes that the combination of both inter- and intra-dependence will lead to better portfolio optimisation models.

8.1 The mean – variance model

In 1952, Nobel laureate Harry Markowitz established the basis and foundation of the modern portfolio theory (Markowitz, 1952). A portfolio can be characterized by two measures; the central tendency or the expected return and a measure of dispersion or the standard deviation of the expected return. The objective of portfolio optimisation is to maximize returns given a standard deviation or minimize the standard deviation of the portfolio given an expected return. Markowitz showed that the standard deviation of a portfolio could be reduced through an understanding of the relationship between projects. If stocks move in the same direction then they create less diversification compared to stocks moving in the opposite direction, which tend to have more diversification. This is similar to the idea of spreading one's bets in order to reduce risk.

Markowitz's model was initially developed for stocks, but many authors have implemented it in the oil and gas industry (Hightower and David (1991), Orman and Duggan (1998) and Ball and Savage (1999)). The calculation of the mean –variance model involves estimating the expected net present value E(NPV) and the standard deviation of the NPV's (SD). The expected net present value of the portfolio is the weighted average of the expected net present value of the individual projects as shown in the following equation:

$$E(NPV_p) = \sum_{i=1}^{n} x_i E(NPV_i) \qquad (1)$$

Where (x_i) is the participation level of the i^{th} project. The standard deviation of the portfolio σ_p is defined as follows:

$$\sigma_p = \left[\sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \sigma_{ij} \right]^{\frac{1}{2}} \tag{2}$$

$$\text{where } \sigma_{ij} = \sigma_i \, \sigma_j \rho_{ij}, \tag{3}$$

and ρ_{ij} is the correlation coefficient between projects i and j

σ_i is the standard deviation of a project i and σ_j is the standard deviation of project j.

Both the expected net present value for each project $E(NPV_i)$ and the standard deviation σ_i are output from the Monte Carlo Simulation model for both the systems and sequential approaches.

8.2 Portfolio optimisation: constructing the efficient frontier

As mentioned above, the mean-variance model aims to maximize the expected NPV or minimize the standard deviation of the expected NPV. This problem can be solved for a single objective by setting the other objective as constant. For example, let us consider the case where the objective is to minimize standard deviation of the expected NPV while holding expected NPV constant and repeating the process to generate the efficient frontier. Before constructing the efficient frontier it is important to set the constraints defining the minimum risk standard deviation and maximum expected NPV. These two values set the boundaries for the efficient frontier. The optimisation process for the efficient frontier is as follows:

Minimize

$$\sigma_p = \left[\sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \sigma_{ij} \right]^{\frac{1}{2}} \tag{4}$$

Subject to:

$$E(NPV_p) = \sum_{i=1}^{n} x_i E(NPV_i) \text{, this is equation (1)}$$

$$M(R_p) = \sum_{i=1}^{n} x_i M(R_i) \leq B_p \tag{5}$$

$$0 \leq x_i \leq 1 \tag{6}$$

Where $M(R_p)$ is the cost of the portfolio and $M(R_i)$ is the individual cost for each project and B_p is the budget constraint. The total cost of the portfolio has to be less than or equal to the available budget. The second constraint is important because for every optimisation we need to set $E(NPV_p)$ as constant. The third constraint is basically setting the participation levels to be between zero and 1.

8.3 Experiments

In order to investigate the difference between the systems and the sequential approaches at the portfolio level, this thesis considered five hypothetical offshore projects, all modelled using the systems and the sequential approaches. It has been assumed that all of the 5 projects are oil related and are located in the same play or have similar geologic characteristics. As demonstrated in Chapter 7, the systems approach captures intra-dependence within single projects compared to the sequential approach, which ignores them. The intra-dependencies within the five projects have been assumed to be:

- Functional dependency between OOIP and initial well rate, both being dependent on reservoir net thickness.
- Functional dependency between recovery factor and initial well rate, both being dependent on permeability.
- Functional dependency between OOIP and recovery factor, both being dependent on porosity and water saturation.
- Total number of wells being dependent upon reserves, production, well CAPEX and oil price.
- Functional dependency between reserves and facility capital expenditure.

The inter-dependence between the five projects, has been assumed to be due to:

- Porosity representing similar geological setting.
- Operating expenditure (OPEX), which depends upon intrinsic costs and geographical location. Since all the five projects are offshore with similar locations, the operating costs per barrel are similar on average. Conceptually, this is the same as setting a single operating cost per barrel distribution for the northern North Sea or deepwater Gulf of Mexico fields.
- Oil price as one of the cash flow elements that is common to all projects.

It is very important to note that in portfolio analysis, the same variables may affect both the inter-dependence in a portfolio as well as intra-dependence in specific projects. For example, in a given portfolio porosity may be an inter-and intra-dependence parameter concurrently. Conversely, there may be parameters which only relate to either intra-dependence or inter-dependence.

Once all the inter-and intra-dependencies were set the following four experiments were run:

- Experiment 1: The five projects were sequentially evaluated without inter-dependence correlations.
- Experiment 2: The five projects were sequentially evaluated with inter-dependence correlations.
- Experiment 3: The five projects were evaluated using the systems approach without inter-dependence correlations.
- Experiment 4: The five projects were evaluated using the systems approach with inter-dependence correlations.

Comparison of the results of experiments 1 and 2 and comparison of results of experiments 3 and 4 show the impact of project inter-dependence for Markowitz's portfolio concept. In addition, the comparison of the results of experiments 1 and 3 and experiments 2 and 4 show the impact of the systems approach versus the sequential approach at the portfolio level. Comparison of results of experiments 1 and 4 show the impact of combining both intra- and inter-dependence together with a case that ignores them. Each one of the experiments was run in an Excel model, which interfaces with @Risk (Monte Carlo Simulation software).

8.4 Results and discussion

Each experiment consists of five projects that were run in a Monte Carlo Simulation and then the output was recorded for the mean and the standard deviation of the NPV and the expected cost for each project. Furthermore, the NPV's for each of the individual projects were extracted from the simulations and a correlation matrix was developed for each experiment. All of these outputs from simulations were used as an input into the portfolio model to generate the efficient frontier.

The results of the portfolio analysis for experiments 1, 2, 3 and 4 are shown on Table 8- 1 through to Table 8-4.

Table 8-1. Results of sequential portfolio analysis without inter-dependence (Experiment 1)

			Participation levels				
E (NPV)	E (SD NPV)	E (costs)	Project 1	Project 2	Project 3	Project 4	Project 5
($ Million)	($ Million)	($ Million)	Fraction	Fraction	Fraction	Fraction	Fraction
1,248	435	1,500	1.00	1.00	0.90	0.63	0.17
1,308	444	1,500	1.00	1.00	0.83	0.56	0.25
1,368	464	1,500	1.00	0.80	0.85	0.58	0.28
1,428	487	1,500	1.00	0.59	0.89	0.60	0.32
1,488	513	1,500	0.92	0.41	0.93	0.62	0.35
1,548	541	1,500	0.74	0.25	0.97	0.65	0.37
1,608	571	1,500	0.57	0.10	1.00	0.68	0.41
1,668	602	1,500	0.34	0.00	1.00	0.71	0.44
1,728	636	1,500	0.00	0.00	1.00	0.70	0.50
1,788	679	1,500	0.00	0.00	0.95	0.61	0.58
1,848	730	1,500	0.00	0.00	0.88	0.54	0.66
1,908	788	1,500	0.00	0.00	0.80	0.47	0.74
1,968	851	1,500	0.00	0.00	0.73	0.41	0.83
2,028	918	1,500	0.00	0.00	0.65	0.34	0.91
2,088	988	1,500	0.00	0.00	0.58	0.27	0.99
2,148	1,012	1,500	0.00	0.00	0.93	0.00	1.00

Table 8-2. Results of sequential portfolio analysis with inter-dependence (Experiment 2)

			Participation levels				
E (NPV)	E (SD NPV)	E (costs)	Project 1	Project 2	Project 3	Project 4	Project 5
($ Million)	($ Million)	($ Million)	Fraction	Fraction	Fraction	Fraction	Fraction
1,127	695	1,500	1.00	1.00	1.00	0.81	0.00
1,192	702	1,500	1.00	1.00	1.00	0.67	0.09
1,257	716	1,500	0.98	1.00	0.92	0.60	0.18
1,322	736	1,500	0.49	1.00	0.94	0.61	0.22
1,387	757	1,500	0.30	0.82	0.99	0.64	0.25
1,452	780	1,500	0.14	0.66	1.00	0.69	0.29
1,517	803	1,500	0.00	0.49	1.00	0.74	0.32
1,582	828	1,500	0.00	0.27	1.00	0.78	0.36
1,647	855	1,500	0.00	0.06	1.00	0.82	0.40
1,712	883	1,500	0.00	0.00	1.00	0.74	0.47
1,777	917	1,500	0.00	0.00	0.99	0.61	0.56
1,842	955	1,500	0.00	0.00	0.90	0.55	0.65
1,907	998	1,500	0.00	0.00	0.81	0.48	0.74
1,972	1,043	1,500	0.00	0.00	0.71	0.41	0.83
2,037	1,092	1,500	0.00	0.00	0.62	0.35	0.92
2,102	1,146	1,500	0.00	0.00	0.93	0.00	1.00

Table 8-3. Results of systems portfolio analysis without inter-dependence (Experiment 3)

			Participation levels				
E (NPV)	E (SD NPV)	E (costs)	Project 1	Project 2	Project 3	Project 4	Project 5
($ Million)	($ Million)	($ Million)	Fraction	Fraction	Fraction	Fraction	Fraction
1,277	404	1,500	1.00	1.00	0.70	0.48	0.16
1,352	416	1,500	1.00	0.94	0.71	0.51	0.20
1,427	433	1,500	1.00	0.73	0.73	0.53	0.24
1,502	456	1,500	0.98	0.53	0.74	0.56	0.27
1,577	484	1,500	0.83	0.36	0.77	0.59	0.31
1,652	514	1,500	0.69	0.20	0.79	0.62	0.34
1,727	547	1,500	0.54	0.04	0.82	0.64	0.38
1,802	583	1,500	0.21	0.00	0.82	0.66	0.42
1,877	624	1,500	0.00	0.00	0.76	0.66	0.48
1,952	673	1,500	0.00	0.00	0.65	0.62	0.55
2,027	729	1,500	0.00	0.00	0.54	0.59	0.63
2,102	791	1,500	0.00	0.00	0.44	0.56	0.70
2,177	857	1,500	0.00	0.00	0.33	0.52	0.78
2,252	926	1,500	0.00	0.00	0.22	0.49	0.85
2,327	998	1,500	0.00	0.00	0.11	0.46	0.93
2,402	1,071	1,500	0.00	0.00	0.00	0.43	1.00

Table 8-4. Results of systems Portfolio analysis with inter-dependence (Experiment 4)

			Participation levels				
Expected NPV	E SD of NPV	E costs	Project 1	Project 2	Project 3	Project 4	Project 5
($ Million)	($ Million)	($ Million)	Fraction	Fraction	Fraction	Fraction	Fraction
906	647	1,500	1.00	1.00	0.63	0.30	0.00
1,006	651	1,500	1.00	1.00	0.76	0.43	0.00
1,106	665	1,500	1.00	1.00	0.83	0.52	0.03
1,206	683	1,500	0.87	1.00	0.85	0.55	0.07
1,306	705	1,500	0.66	0.89	0.89	0.60	0.12
1,406	729	1,500	0.46	0.67	0.92	0.64	0.16
1,506	757	1,500	0.25	0.46	0.95	0.68	0.21
1,606	786	1,500	0.04	0.25	0.99	0.72	0.26
1,706	818	1,500	0.00	0.00	0.99	0.75	0.31
1,806	855	1,500	0.00	0.00	0.84	0.72	0.41
1,906	897	1,500	0.00	0.00	0.68	0.69	0.50
2,006	945	1,500	0.00	0.00	0.52	0.65	0.60
2,106	997	1,500	0.00	0.00	0.36	0.62	0.70
2,206	1,052	1,500	0.00	0.00	0.20	0.59	0.80
2,306	1,111	1,500	0.00	0.00	0.05	0.56	0.90
2,406	1,170	1,500	0.00	0.00	0.00	0.43	1.00

The resulting efficient frontiers for experiments 1 and 2 are shown on Figure 8-1, where as the efficient frontiers resulting from experiments 3 and 4 are shown in Figure 8-2.

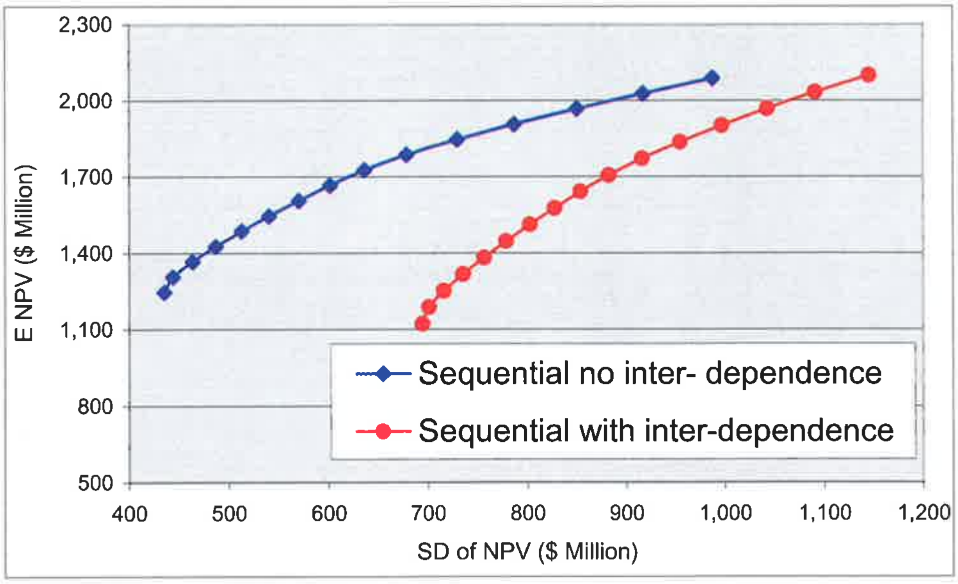


Figure 8-1. Comparison of efficient frontier resulting from the sequential approach with and without inter-dependence

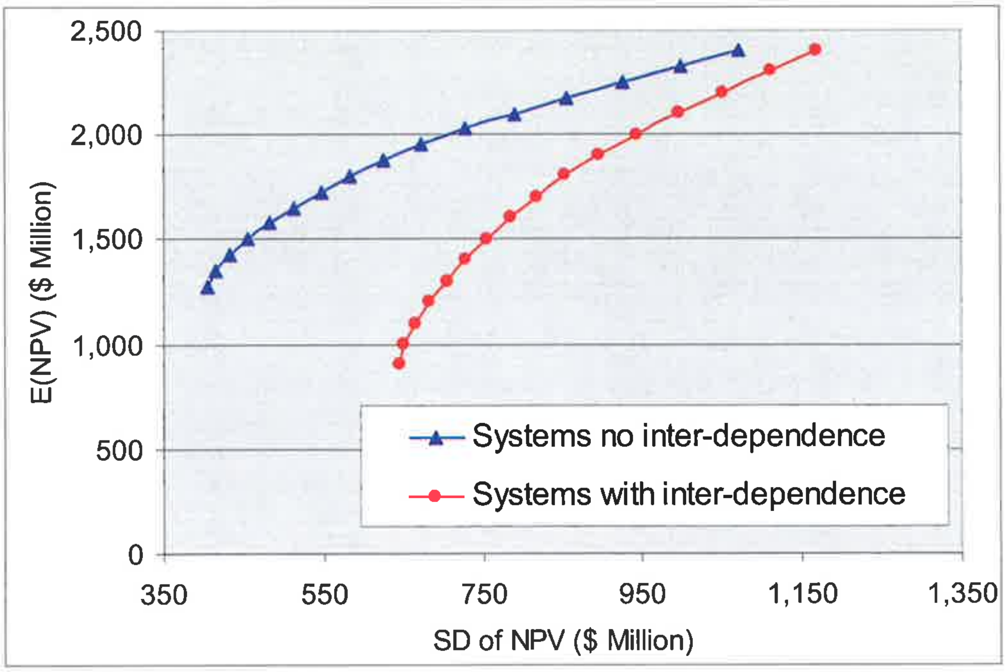


Figure 8-2. Comparison of efficient frontier resulting from the systems approach with and without inter-dependence

A comparison of experiment 1 and 2 shows the impact of portfolio modelling which accounts for inter-dependence. This result shows that the inclusion of inter-dependence causes the efficient frontier curve to shift down and move to the right. This results in a decrease in the expected NPV and an increase in the standard deviation. This is true in our case because all of the inter-dependence variables such as porosity, OPEX and oil price have positive correlations which lead to increases in the standard deviation compared to those cases that ignore them. The same effect is observed when comparing the systems approach with and without inter-dependence. In conclusion these results confirm what we would expect in a portfolio model with only inter-dependence.

Comparing the resulting of experiments 2 and 4 illustrates the impact of systems approach in the context of portfolio analysis, as shown in Figure 8-3. This result is very important because it clearly shows the value of including intra-dependence which the systems approach captures but the sequential approach does not. It shows that for the same level of risk, a higher expected net present value results when using the systems approach compared to the sequential approach. Furthermore, for a smaller level of risk the difference between the sequential and the systems approach is smaller. As the level of risk increases the difference between the two approaches grows from 8 to 15% difference in expected NPV. This can be attributed to the impact of dependencies and interactions as shown in Chapter 7, where a systems approach for a single project has a higher mean NPV and standard deviation compared to the sequential approach. The results shown in Figure 8-3 clearly demonstrate the importance of the systems approach and its ability to capture intra-dependence that a portfolio model using the sequential approach is not able to.

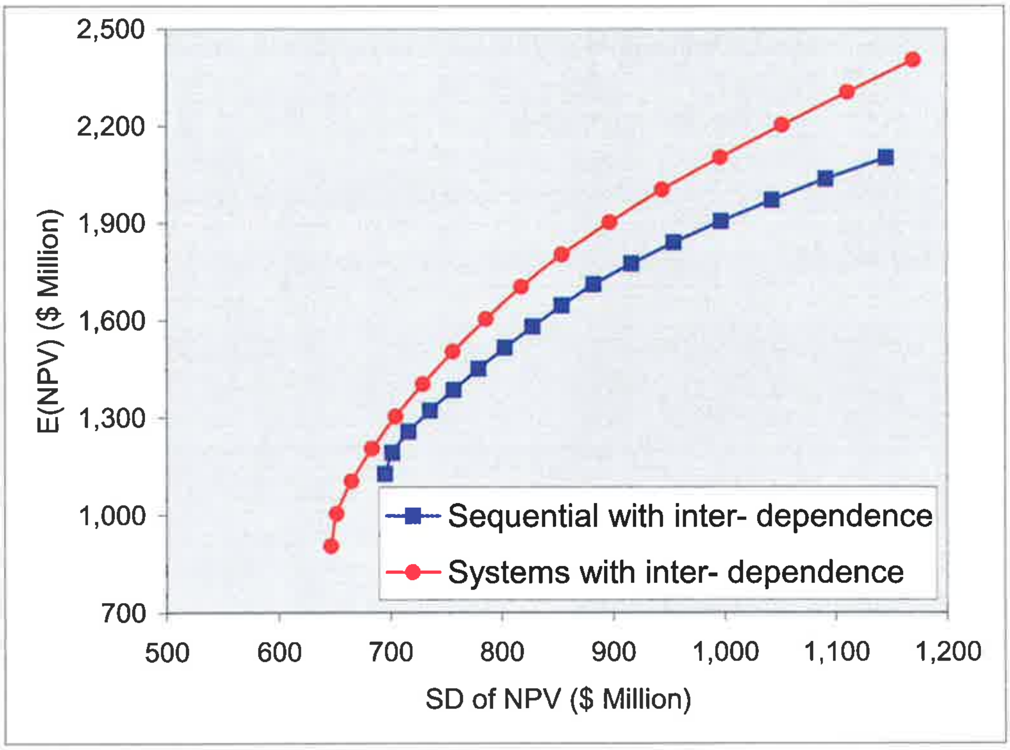


Figure 8-3. The impact of intra-dependence captured by the systems approach at the portfolio level

In order to investigate the impact of including or ignoring both inter and intra-dependence, we compare the sequential approach with no inter-dependence with the systems approach with inter-dependence (Figure 8-4).

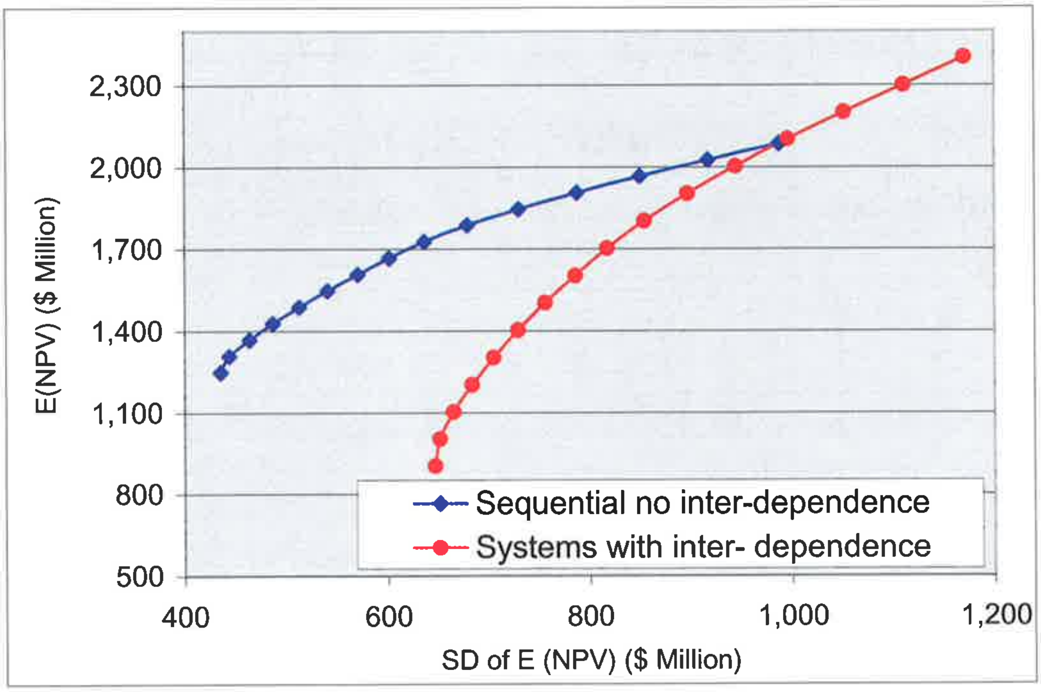


Figure 8-4. Both portfolio and systems approaches capture inter and intra-dependence compared to sequential approach

The inability of the sequential approach to capture both inter- and intra-dependencies at the lower level of risk leads to overestimate of the expected return for the same value of risk up to 30% at the lower levels as shown in Figure 8-4. At higher levels of risk the systems approach outperforms the sequential approach by capturing more expected NPV for the same level of risk that the sequential approach is not able to achieve. This clearly shows that ignoring dependencies at the lower level of risk tends to give an optimistic view and a pessimistic view at higher levels of risk.

The portfolio model has the advantage of capturing inter-dependence among projects, thus interaction among projects is more important than single projects alone. However, the systems approach is also designed to capture intra-dependence in single projects, which sometimes could be the same as inter-dependence. Combining both the portfolio and the systems approaches produce better results than compared to each approach separately. Use of both approaches captures both inter-dependence and intra-dependence among the projects and within the projects themselves.

8.5 Conclusion

This Chapter addressed the question whether the efficient frontier resulting from using the systems approach to model a portfolio is different from the efficient frontier generated using the sequential approach. The conclusions are:

- Markowtiz portfolio model captures interaction and dependencies between projects and leads to lower expected net present value and higher standard deviation when inter-dependencies are included. This result was observed when comparing the results of both the sequential and the systems approach with and without inter-dependence.
- The systems approach captures the intra-dependence within a project and leads to a higher expected net present value for a given level of risk when compared to the sequential approach at the portfolio level.
- Combination of the portfolio model using the systems approach is able to capture both intra- and inter-dependence within a project and between the projects in the portfolio set. The results show that ignoring both inter- and intra-dependence leads to overestimation of expected

net present value at lower levels of risk and underestimation of the expected net present value at the higher levels of risk.

- The portfolio and the systems approaches complement each other by capturing both inter-and intra-dependence and can add significant value to petroleum project investment decision-making at a portfolio level.

CHAPTER

9

Systems approach vs. Decision tree approach: A comparison

9. Introduction

This chapter presents the results and conclusions for the fifth objective of this research, which is to compare the systems approach, introduced in this research, with the industry approach of using decision trees for analysing investment decisions using a hypothetical offshore development decision as an example.

9.1 The hypothetical development decision

Results of the decision tree and the systems approach were compared by investigating their impact on the expected NPV of the offshore development decision example introduced in chapter 5, and the assumptions and data used in chapter 7. The objective is to use the tried and tested decision tree approach to show the validity of the systems approach and to show that with the incorporation of an infinite number of outcomes, decision tree analysis results would approximate the results of the systems approach.

9.2 Decision tree analysis

Decision trees are important in solving complex problems where these problems involve subsequent decisions. A tree consists of decision, chance and terminal nodes connected by branches. These nodes are represented as shown in Figure 10-1.

1. A square symbol, ■, is a decision node
2. A circle symbol, ●, is a chance node which represents uncertain outcome
3. A terminal node, ◄, is the end node which represents the outcome values of that branch.

This thesis uses the term “Decision tree” to include the term “ Event tree”. Where a tree is used without a decision along its branches, it is considered to be an event tree.

Decision trees are often used to calculate the Expected Monetary Value (EMV). A decision tree can be solved using the backward induction (sometimes called rollback) method. In order to calculate the EMV, two inputs are required:

- Probabilities have to be assigned to each uncertain variable knowing that the sum of probabilities at every chance node must equal to one. This is called the normalization process.
- Value outcome for each branch terminal node. In our case it is the NPV of each outcome.

The example shown in Figure 9-1 depicts the decision node to drill or not drill. The chance node represents the uncertainty in the outcome of deciding to drill with input probabilities and three possible outcomes; dry hole, 20 MMSTB or 50 MMSTB. The end nodes represent the input NPV values of -$20 million, $50 million and $100 million for each scenario outcome. The EMV is $51 million and, given that this is positive, the decision is to drill.

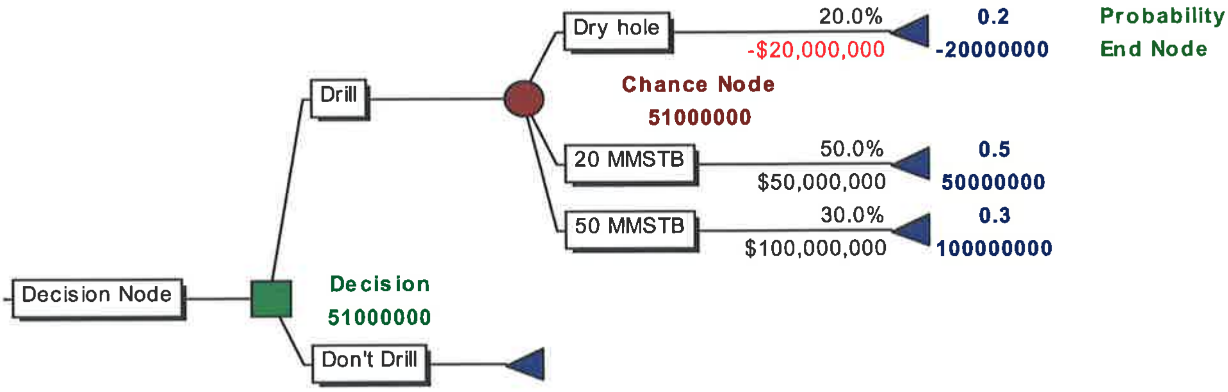


Figure 9-1. Drilling decision

In order to use a decision tree, we need to determine the decision nodes together with their possible outcomes. The second step is to determine the choice or the uncertain variables. For simplicity, this analysis assumes that reserves, initial production, FPSO capital expenditure and price will be treated stochastically and the rest of the variables in the model will be treated deterministically. These stochastic variables have the tendency to impact the net present value the most. All the uncertain variables that are used in the calculation of reserves were represented by the reserves variable. A similar assumption was made for initial production. It is important to point out that the decision maker can choose to have any number of variables, as stochastic. All the 29 variables in the development decision analysis could be treated as uncertain when modelling using the decision tree approach. However, this would become extremely cumbersome to evaluate.

The third step is to assign the input values and the probabilities associated with each scenario. The decision tree approach does not use stochastic continuous distributions but uses discrete values that represent a continuous distribution. To convert from the stochastic continuous distribution to discrete values, the Swanson-Megill approximation was used for the distribution of reserves. This approximation converts the continuous distribution into 3 discrete values by assigning a 30%

weighting to a P10 outcome, 40% weighting to a P50 outcome and 30% weighting to a P90 outcome. The Swanson-Megill approximation is very good for converting a continuous distribution to discrete values. This was done for reserves, initial production, FPSO CAPEX and price. The rest of the variables were treated as deterministic. Reserves, initial production, FPSO CAPEX and price were each represented by three discrete outcomes as shown on Figure 9-2. The same set of equations used for calculating NPV for the systems approach were used for the decision tree model in order to have a fair comparison of the two approaches. The fact that four variables were uncertain and each one of these variables was represented with three nodes requires the whole decision model to have 81 scenarios ($3^4 = 81$).

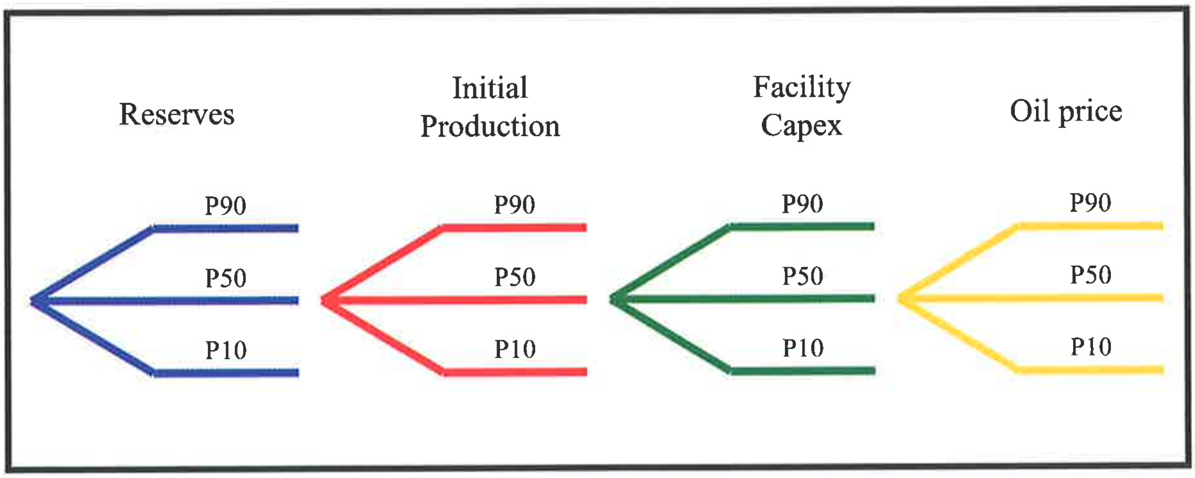


Figure 9-2. Development decision using decision tree

The fourth step is to calculate the NPV value of each outcome scenario. In order to calculate these the reserves tree nodes P10, P50 and P90 were assigned three values 1, 2 and 3. The same was done for initial production, FPSO CAPEX and oil price. Assigning these values helped to calculate the NPV for each outcome scenario. For example, the first scenario will take the following numbers 1, 1, 1, 1. These numbers represent the P10's of reserves, initial production, FPSO CAPEX and oil price. This is the first scenario as shown on Table 9-1.

The calculation of NPV for the first scenario involves using the P10 of reserves as an input and the P10 of initial production. Based on these two variables and other deterministic variables, the decision tree model will calculate the production profile and the number of wells required for this specific P10 reserves and P10 initial production. Once the production profile is generated the model will pick the P10 of FPSO CAPEX and the P10 of oil price to calculate the NPV of the first scenario.

Table 9- 1. List of scenarios required for the calculation of each NPV scenario for the decision tree approach

Scenarios	Reserves	Initial Production	FPSO Capex	Price
Scenario 1	1	1	1	1
Scenario 2	1	1	1	2
Scenario 3	1	1	1	3
Scenario 4	1	1	2	1
Scenario 5	1	1	2	2
Scenario 6	1	1	2	3
Scenario 7	1	1	3	1
Scenario 8	1	1	3	2
Scenario 9	1	1	3	3
Scenario 10	1	2	1	1
Scenario 11	1	2	1	2
Scenario 12	1	2	1	3
Scenario 13	1	2	2	1
Scenario 14	1	2	2	2
Scenario 15	1	2	2	3
Scenario 16	1	2	3	1
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
Scenario 75	3	3	1	3
Scenario 76	3	3	2	1
Scenario 77	3	3	2	2
Scenario 78	3	3	2	3
Scenario 79	3	3	3	1
Scenario 80	3	3	3	2
Scenario 81	3	3	3	3

The same steps need to be repeated to calculate the NPV value for the second outcome scenario. The calculation of EMV of the development decision requires 81 scenarios, which if conducted manually would require a lot of work and time. To overcome this, we have used the power of Monte Carlo Simulation to calculate the NPV's for all the scenarios. To do this a variable was created, which is called a choice value, which is a uniform integer distribution with three values {1, 2, 3}, which represent the P10, P50 and P90 of a given variable. Then for uncertain variables in the decision tree model, such as reserves, a VLOOKUP function was created to pick up the choice value from the uniform distribution and assign the corresponding value to it. For example, if the choice value is 1, then the VLOOKUP function will pick P10, which corresponds to 33 million barrels. As the choice value changes the reserves value will change accordingly. This process was done for the initial production, FPSO capex and oil price as they represent the uncertain variables in the decision tree model. The next step is to create a function that identifies the choice values generated for each scenario and the corresponding NPV for them. Using Table 10-1, for each scenario generated, an IF (AND) logic function was used to match the NPV's and the scenarios. For example, from Table 10-1 the first scenario is 1, 1, 1, 1 , the IF (AND) function will match the choice values with the calculated NPV for that scenario; if they match it will record the NPV otherwise it will find the appropriate corresponding scenario and record it. The NPVs recorded are directly input into the Precision Tree™ software as an input along with their calculated probabilities. For example, the probability of getting the NPV of the first scenario is 0.0081 (0.3*0.3*0.3*0.3), which is the multiplication of the four variables probabilities of P10's. Running a Monte Carlo Simulation for this process will generate all the scenarios in less than 5 minutes. This research has found that using the above-described procedure is really simple, very powerful and it saves a lot of time, assuming a scenario table has been created. This process is strongly recommended for decision tree analysis with multiple scenarios.

After generating all the NPV's with their distributions, the EMV value is calculated by folding the tree back. The results is an EMV = $166 million using the Precision Tree^{TM} software. The decision tree therefore supports the conclusion to develop the field. The complete decision tree for the example development decision model is attached in Appendix I.

The output from the decision tree approach is the EMV, which is the mean of the NPV of the analysis. The systems approach provides the output as a continuous distribution with all the statistical parameters required. This is a very important element in the decision making process. The focus should not only be on the mean, but also on the standard deviation, percentiles, and the probability of achieving a certain NPV. For example, what is the probability that the NPV's will be greater than zero? Or what is the probability of the low or high side NPV's? These types of questions cannot be answered by the decision tree approach. If the decision maker is concerned with statistical parameters other than the EMV value then the systems approach is the method to use.

Relying on the mean only as the judging criteria for the difference between the systems approach and the decision tree approach is not sufficient. To have a complete comparison of the whole NPV distribution output, a cumulative distribution function was generated using the NPV's calculated and their corresponding probabilities in a Monte Carlo Simulation. This process made it easy to compare the two approaches.

9.3 The development decision using the systems approach

Chapter 5 introduced the construction of the systems approach, which was compared with the sequential approach in chapter 7. The systems approach was constructed using 29 uncertain variables. The dependencies and interactions were modelled as indicated and results were presented in chapter 7.

9.4 Decision tree and systems approach: comparison and impact

The decision tree approach has been used, tested and validated in the oil and gas industry. The aim here is to compare the results of the two approaches and show that the systems approach produces similar results. In order to validate the systems approach, four experiments were conducted with the decision tree model:

- Experiment 1 used only one uncertain variable (reserves)
- Experiment 2 used two uncertain variables (reserves and initial production)
- Experiment 3 used three uncertain variables (reserves, initial production and FPSO CAPEX)
- Experiment 4 used four uncertain variables (reserves, initial production, FPSO CAPEX and Oil price)

The results of these experiments are presented graphically and quantitatively. Graphically, Figure 9-3 shows the cumulative distribution function (CDF) for the four experiments. As we move from once uncertain variable to four uncertain variables, the decision tree results are getting closer to the systems approach results. It is also very interesting to note that as we move from one to four uncertain variables, the difference in the mean value stays close to the mean found for the decision tree with four variables. In other words, the mean NPV does not change significantly from one to four uncertain variables. Other statistical parameters such as standard deviation change significantly as the number of uncertain variables increases.

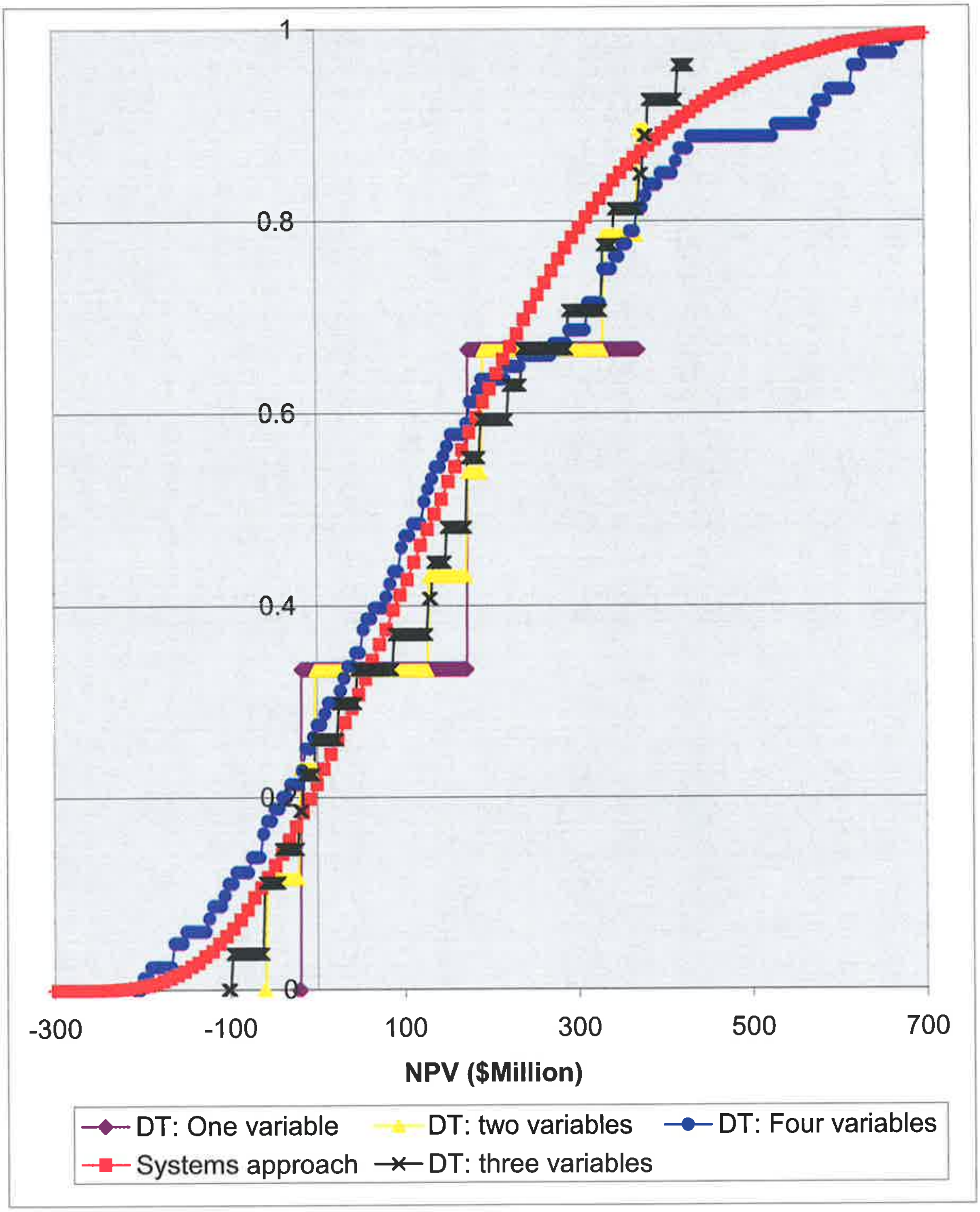


Figure 9-3. Decision tree CDF with variables vs. systems approach

A Q-Q plot was developed by calculating the cumulative distribution values for the x-axis in Figure 9-3 and then plotting the curves again. The results show that once again the decision tree approach with four uncertain variables is similar to the systems approach as shown in Figure 9-4.

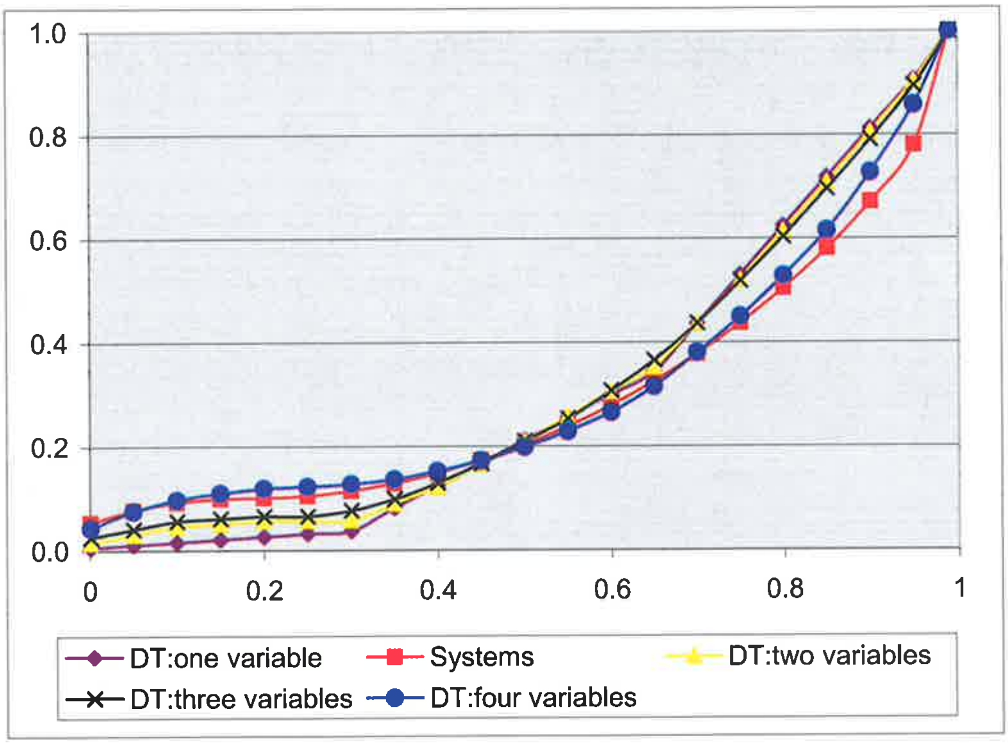


Figure 9-4. Q-Q Plot for Decision Tree CDF's vs. Systems approach

The Q-Q Plot made it easier to show the result quantitatively by summing the square difference between each curve and the systems approach. As we would expect, the curve that produces the minimum value is the one that is closest to the systems approach. This is the same concept as the minimum distance approach presented in the copulas methodology section. The quantitative results are summarized in Table 9-2, which clearly shows that the results of decision tree analysis with four variables are close to the results of the systems approach. These results validate the systems approach compared to the tried and tested decision tree approach. Furthermore, the results show that the decision tree approach used by the industry is just a subset or part of the systems approach. As the number of uncertain variables increases, the results of the decision tree approach converges with the results of the systems approach. Specifically, as the experiments move from one uncertain variable to four uncertain variables, the distance between the decision tree approach and the systems approach gets smaller. This clearly indicates that the decision tree approach is a subset of the systems approach.

Table 9- 2. Minimum values that shows which curve is closer to the systems approach

Decision Tree Curves	Minimum values
Decision tree with one variable	0.12
Decision tree with two variables	0.09
Decision tree with three variables	0.07
Decision tree with four variables	0.012

Now that we have validated the systems approach, the next issue to address is the impact of using the systems approach compared to the decision tree approach on the NPV of a development decision.

The results obtained are very interesting because they clearly show that even though both approaches use the same set of equations, the results are different. The mean NPV (EMV) generated from the decision tree approach compared to the mean value generated from the systems approach shows that these two values are approximately the same. However, the decision tree approach tends to underestimate the low side NPV's (P5 and P10's) by 30 - 47% (Figure 9-5). While on the high side (P90 and P95's) the decision tree tends to overestimate the NPV by more than 30%. It shows that the project has a higher upside potential than the actual potential. The mean values for both approaches are close to each other with a difference of 5%, while the standard deviation has a difference of 20%. The result of this comparison is shown in Figure 9-5 and Table 9-3.

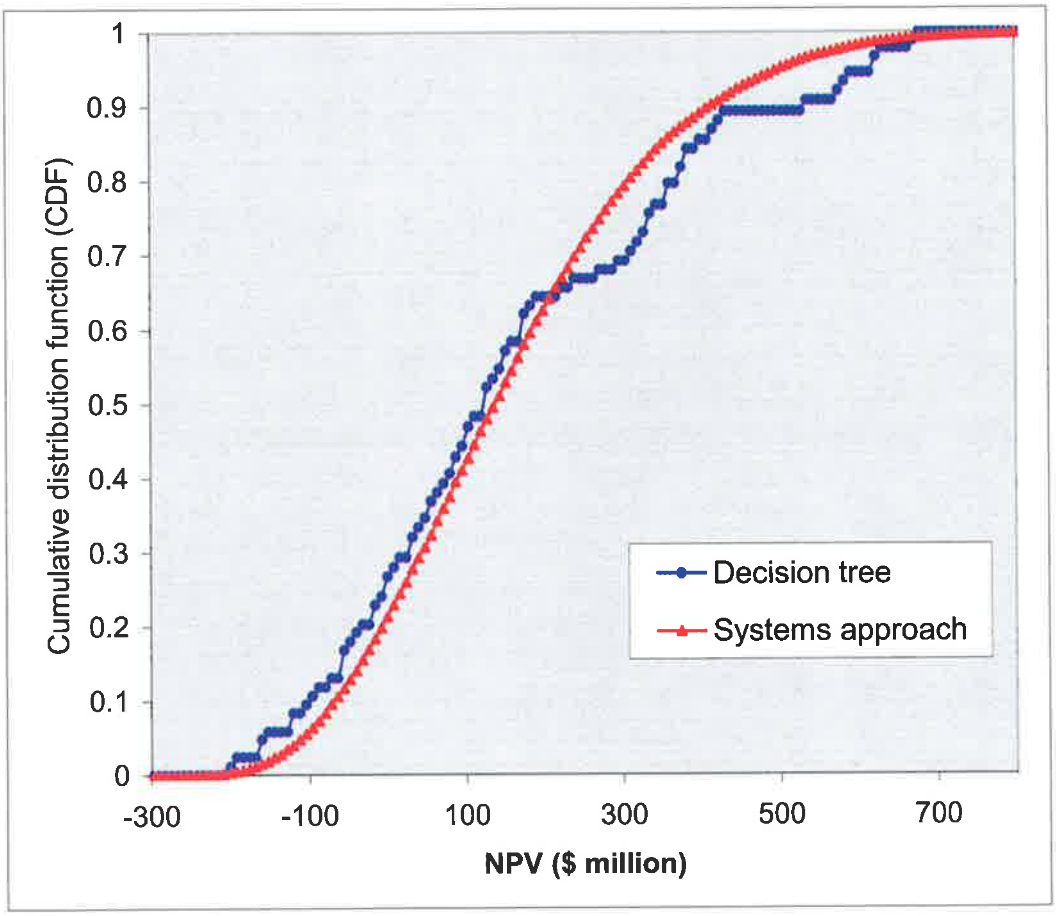


Figure 9-5. CDF of NPV comparison between decision tree and systems approach models

Table 9- 3. Comparison of decision tree and systems approaches model NPV's ($ million)

Parameter/ Approach	Decision tree	Systems approach
Mean	164.77	156.81
Std Dev	223.18	185.18
10% Percentile	-102.75	-69.43
50% Percentile	123.17	138.26
90% Percentile	530.22	410.90

9.5 Results and discussion

The results above clearly indicate that the decision tree approach tends to underestimate the low side by showing it is worse than it actually is. It also overestimates the high side by showing that it looks higher than it actually is. Whilst the difference between the two approaches is large away from the mean, it is very small closer to the mean. The difference in results of the two approaches arises due to the difference in their core characteristics:

9.5.1 Number of uncertain variables (Complexity)

In this example, the decision tree analysis considers only four variables that are stochastic while the systems approach considers all 29 variables to be stochastic. The assumption to limit the number of stochastic variable to four was made because adding more uncertain variables to the decision tree approach will cause the number of branches to grow exponentially as the number of variables increases. In the decision tree approach, just assuming four variables with three nodes leads to 81 scenarios. If the decision tree approach was to use 29 variables as used in the systems approach there would be 69 trillion scenario outcomes of the decision tree as shown on Figure 9-6. From a practical point of view, this is very difficult to accomplish and there is no point in doing it since the systems approach can achieve similar results in less than 5 minutes of computation.

It is important to indicate here that if the 29 uncertain variables modelled by the systems approach could be modelled by a decision tree, we would expect the results of both approaches to be almost identical. It is also important to point out that not all the 29 uncertain variables are important because some have a minor impact compared to others.

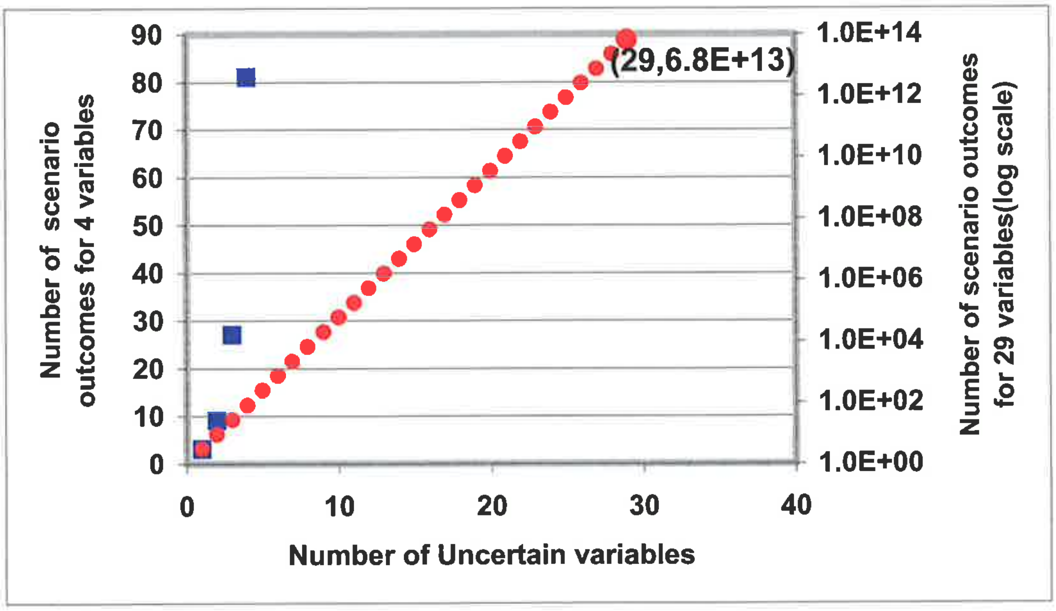


Figure 9-6. Number of scenario outcomes vs. number of uncertain variables for a decision tree approach

9.5.2 Account for uncertainties

In our example, the decision tree approach only accounted for uncertainties in four variables; reserves, initial production, FPSO CAPEX and oil price. The rest of variables were assumed certain. In reality, uncertainty exists in all the parameters. Part of the difference in the result between the systems approach and the decision tree approach is due to the ability of systems approach to account for uncertainty in all the variables compared to the limited number of uncertain variables typically accounted for in decision tree analysis for practical reasons.

Furthermore, the decision tree approach assumes three discrete values for reserves, initial production and FPSO CAPEX. For example, a decision tree scenario could be middle reserves, low production, low FPSO CAPEX and low oil price. These low values represent the P10's of the last three variables and the middle value represents the P50 of reserves. By doing this the decision tree analysis ignores all scenario outcomes that may exist between the three discrete points as shown in Figure 9-7.

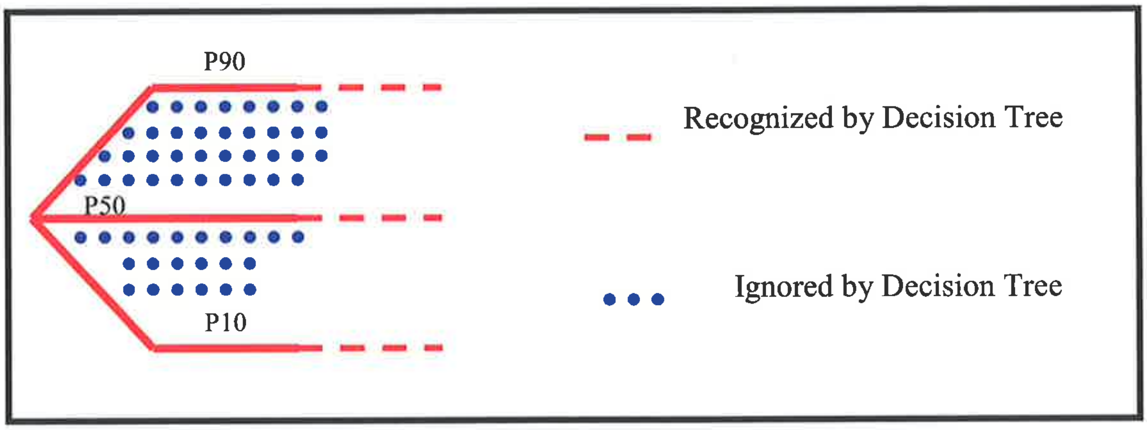


Figure 9-7. Scenarios recognized and ignored by decision tree

9.5.3 Dependencies and interactions

Even with a complete modelling of 29 uncertain variables using decision tree analysis, the final results of the decision tree will not be the same as the systems approach. This is because the decision tree approach ignores dependencies and interactions. For example, the modelling of dependencies between area and net thickness, porosity and water saturation are completely ignored. Decision trees do not easily allow the use of dependency modelling technique such as the Copulas and Iman-Conover methods. They are not designed to capture statistical stochastic dependencies unlike the systems approach. Furthermore, functional dependencies are also difficult to achieve in decision tree. For example, as net thickness increases, reserves increase and initial well production increases because of the dependency between net thickness and production. Also as permeability increases, recovery factor increases and this also aligns with higher initial well production. These types of dependencies are difficult to incorporate in the decision tree approach. However, they are built into the systems approach and are incorporated in every iteration.

9.5.4 Handling stochastic oil price models.

Decision trees are not good in capturing dynamic behaviour. For example, it is very difficult to capture the Mean Reversion oil price model or the Geometric Brownian Motion model using the decision tree approach because the distributions are continuous and each current price is a function of volatility, reversion factor, long-term price and previous year price. These prices have two elements: an uncertainty which is captured in a continuous distribution and another element, which represents the dynamic behaviour. These types of price models clearly cannot be handled by the decision tree approach while the systems approach can handle them easily.

9.6 Conclusion

The objective of this chapter is to compare and contrast the systems approach, introduced by this research, and the decision tree approach. This research shows that in the experiments conducted the decision tree results differs by 5% for the mean and 20% for the standard deviation when compared to the systems approach. The decision tree approach tends to underestimate the low side and also overestimates the high side. This difference is due to the following factors:

- **Complexity** (Number of uncertain variables) as the number of uncertain variables increases, the decision tree approach yields results which are closer to the systems approach. However, as number of uncertain variables increase, the problem becomes complex and very difficult to handle by decision tree approach.

- **Uncertainty** can only be represented by discrete branches, generally three chosen are P10, P50 and P90 and everything in between, below or above is ignored with the decision tree approach, while all uncertainties are accounted for with the systems approach.
- **Dependencies and interactions** are not captured by the decision tree approach. Decision trees are not designed to capture statistical stochastic dependency methods such as the Copula and the Iman-Conover methods.
- **Model output.** The systems approach produces an output distribution with all the statistical parameters such as the standard deviation and percentiles to help understand uncertainty while the decision tree approach produces a single measure, the EMV.
- **Tolerate changes:** The systems approaches can handle changing variables from deterministic to stochastic easily compared to the decision tree approach.
- **Stochastic oil price models.** Clearly the systems approach is better designed to deal with dynamic and continuous behaviour models, such as the Mean Reversion oil price model, compared to the decision tree approach.

In the context of investment decision making in the oil and gas industry accounting for complexity, uncertainty, dependency and interaction, and dynamic behaviour is important. This research concludes that the systems approach is superior to the decision tree approach in addressing these.

CHAPTER 10 In Conclusion…

10. Introduction

This chapter presents a summary of all the conclusions of this study. It examines the objectives of this research stated in Chapter 1 and combines the main findings and conclusions to address these objectives. Furthermore, it presents some recommendations for implementation and future work in the area of dependencies and interactions using the systems approach.

10.1 Conclusion

The research hypothesis is that there is value in modelling dependencies and interactions at each of reserves, project and portfolio levels. Furthermore, it hypothesizes that the systems approach developed in this research is superior to the current decision tree analysis used by the oil and gas industry. This research has theoretically and experimentally confirmed these hypotheses by addressing the objectives proposed of this study.

1. **Objective 1: Build a stochastic integrated asset model (called the systems approach).**

   The fully stochastic integrated model (systems approach) has been developed. The construction of this model is explained in detail in chapter 5 and the application of this model is shown and explained in chapter 7. The model is able to integrate all components of the petroleum system from reserves to economics including above and below ground uncertainties. It is also able to capture dependencies and interactions.

2. **Objective 2: At the reserves level: Investigate the impact of statistical dependencies on the technical reserves.**

   This objective addresses two points:

   a) It focuses on the impact of the *dependence structure* of the correlation model on reserves.

      This research has concluded that the copulas approach is superior in capturing dependency patterns in the lower- and upper-tail dependence structure compared to the existing dependency methods such as the Iman-Conover, Regression fitting and the Envelope method.

   b) It compares and contrasts the current techniques in oil and gas evaluations with the copulas method

      This research has confirmed that dependencies are important and can have a large impact on the metrics used for decision-making. Furthermore, this research has shown that the copulas methodology is superior to regression fitting and the envelope method, which tend to underestimate the mean and the standard deviation. However when compared to the Iman- Conover method the impact of the dependence structure is not significant. Furthermore, the concept of multiple lines to construct the envelope method has been introduced. This performs better in capturing the dependence structure than the traditional approach. However, it is still not as good as the copulas method.

The result obtained at the reserves level can be generalized at the project level as well. At the project level the impact of copulas and Iman-Conover methods will be close to each other for Archimedean copulas similar to that at the reserves level. In addition the copulas approach will outperform the Envelope and Regression fitting methods at the project level similar to that at the reserves level. It is important to point out that the conclusions related to the mean value apply equally to the P50 value since this is a common industry measure.

This research recommends the following from a *theoretical* point of view as methods to be used in modelling statistical dependencies in order of importance;

1. The Copulas method should be preferred, because it captures dependence structure and provides accurate results.
2. Next in line is the Iman-Conover method because its results are almost as the results of the copulas method even though it does not capture dependence structure.
3. The Envelope method with multiple lines is third because captures the dependence structure and generally produces reasonable results. However, its use can be very subjective

4. And finally, Regression fitting is ranked fourth because it has theoretical reasoning behind it and produces reasonable results.

From *practical* point of view, the Iman- Conover method is easier to apply because the software to do this already exists. The Copulas method needs more understanding of mathematics and modelling skills. Where no software is available it is recommended that the Regression fitting or the Envelope methods be used because they are very easy to construct in an Excel spreadsheet.

3. **Objective 3: At the project level: Investigate the impact of functional dependencies and interactions on a development decision through a comparison of the systems and sequential approaches.**

This objective investigates the impact of dependencies and interactions on NPV through the analysis of a hypothetical offshore field development and concludes the following:

- The impact of dependencies and interactions on NPV is significant and could be material to the development decision. The systems approach captures interactions and dependencies while the sequential approach ignores them. Ignoring interactions underestimates the mean as well as the standard deviation (by 54% and 44% respectively in our example). Furthermore, the P10, P50 and P90 values of NPV are underestimated by 20%, 50% and 50% respectively in the sequential approach in our example.
- In a stochastic environment, where functional dependence and interaction amongst input parameters are ignored, both the sequential and the systems

approaches yield almost identical results. The difference is merely statistical. In the presence of functional dependence and interaction, the systems approach is superior in capturing their impact.

- The impact of the functional dependence and interaction on the mean and standard deviation of the development decision depends upon:
    - **i.** The modelling approach, the systems approach captures interaction where as the sequential approach does not.
    - **ii.** The functional form of the interaction.
    - **iii.** The sensitivity of a variable. Variables that have higher impact on the output will have a higher interaction impact compared to variables that have lower impact on the output variables.
    - **iv.** Even if the functional form of the interaction is linear the systems and the sequential approach yield different results.

4. Objective 4: At the portfolio level: Investigate the impact of dependencies and interactions (intra-dependence) for a mix of development decisions projects.

This objective questions whether the results obtained from the efficient frontier using the systems approach produces different results from the efficient frontier generated using the sequential approach when modelling the same set of projects. This research concludes the following:

- Markowtiz portfolio model captures interaction and dependencies between projects and leads to lower expected net present value and higher standard deviation. This result was observed when comparing both the sequential and the systems approach with and without inter-dependence.

- The systems approach captures the intra-dependence within a project and leads to a higher expected net present value for a given level of risk when compared to the sequential approach at the portfolio level.
- Combination of the portfolio model using the systems approach is able to capture both intra- and inter-dependence within a project and between the projects in the portfolio set. The results show that ignoring both inter- and intra-dependence leads to overestimation of expected net present value at lower levels of risk and underestimation of the expected net present value at the higher levels of risk.
- The portfolio and the systems approaches complement each other by capturing both inter-and intra-dependence and can add significant value to petroleum project investment decision-making at a portfolio level.

5. **Objective 5: Compare impact of the Systems approach and the current industry approach, Decision Tree Analysis, on the NPV of an offshore development decision.**

This objective compares and contrasts the systems approach, introduced by this research, and the decision tree approach. This research shows that in the experiments conducted the decision tree differs by 5% for the mean and 20% for the standard deviation when compared to the systems approach. The decision tree approach tends to underestimate the low side and overestimates the high side. This difference is due to the following factors:

- **Complexity** (Number of uncertain variables) as the number of uncertain variables increases, the decision tree approach yields results which progressively get closer to the systems approach. However, as the number of uncertain variables increases, the problem becomes complex and very difficult to handle by the decision tree approach.
- **Uncertainty** can only be represented by discrete branches, generally three chosen are P10, P50 and P90. Values in between, below or above are generally ignored with the decision tree approach, while all uncertainties are accounted for with the systems approach in a continuum.
- **Dependencies and interactions** are not captured by the decision tree approach. Decision trees are not designed to capture statistical stochastic dependency methods such as the Copula and the Iman-Conover methods.

In addition, the systems approach has the following advantages over decision tree analysis:

- **Model output.** The systems approach produces an output distribution with all the statistical parameters such as the standard deviation and a continuous probability distribution of the results to help understand uncertainty while the decision tree approach produces a single measure, the EMV.
- **Tolerate changes:** The systems approach can handle changing variables from deterministic to stochastic easily compared to the decision tree approach.

- **Stochastic oil price models.** Clearly the systems approach is better designed to deal with dynamic and continuous behaviour models, such as the mean reversion oil price model, compared to the decision tree approach.

10.2 Recommendations for future work

This research recommends the following two areas for further research:

1. Even though it was found that the Copulas method provided similar results to the Iman-Conover method; it is recommended that the impact of copulas at the portfolio level be explored further. For example, the inter-dependence among projects could be modelled using a copula instead of using the Pearson correlation. This could be interesting because the impact of diversification is based on the correlation and one could investigate whether the choice of the correlation model matters at the portfolio level.

2. It is recommended that the simple systems approach model built as part of this research be extended to incorporate the appropriate level of sophistication and programmed in a user-friendly windows format via the use of a multidisciplinary team with functional expertise in each of the key input variables. We believe that the use of such a model would enhance the quality of investment decision making in the oil and gas industry.

REFERENCES

Accioly, R. and Chiyshi, R. (2004). *Modeling Dependence with Copulas: A useful tool for field development decision process*. Journal of Petroleum Science and Engineering **44**: 83-91.

Anders, J. (1994). *Incorporating Risk in Exploration Economics*. World Oil: 49-55.

Armstrong, M., Galli, A., Bailey, W. and Couet, B. (2004). *Incorporating technical uncertainty in real option valuation of oil projects*. Journal of Petroleum Science and Engineering **44**: 67-82.

Ball, B. and Savage, S. (1999). *Portfolio Thinking: Beyond Optimization*. Petroleum Engineer International: 54-56.

Ball, B. and Savage, S. (1999). *Holistic vs. Hole-istic E&P Strategies*. Journal of Petroleum Technology: 74-84.

Begg, S., Bratvold, R. and Campbell, J. (2001). *Improving Investment Decisions Using a Stochastic Integrated Asset Model*. Paper SPE 71414 presented at the 2001 SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana. 30 September-3 October.

Begg, S., Bratvold, R. and Campbell, J. (2002). *The Value of Flexibility in Managing Uncertainty in Oil and Gas Investments*. Paper SPE 77586 Presented at the SPE Annual Technical Conference and Exhibition. San Antonio, Texas, USA.

Behrenbruch, P., Turner, G. and A, B. (1985). *Probabilistic Hydrocarbon Reserves Estimation: A Novel Monte Carlo Approach*. Paper SPE 13982, Aberdeen, Scotland. September 10- 13.

Bos, C. F. (2005). *A Framework of Uncertainty Quantification and Technical-to-Business Integration for Improved Investment Decision-Making*. SPE 94109 presented at the 2005 SPE/EAGE Annual Conference, Madrid, Spain. June 13-16.

Brashear, J., Becker, A. and Gabriel, S. (1999). *Interdependencies Among E&P Projects and Portfolio Risk Management*. Paper SPE 56574 presented at the SPE Annual Technical Conference and Exhibition, Houston, Texas. 3-6 October.

Charles, T., Guemene, J., Corre, B. and Dubrule, O. (2001). *Experience with the Quantification of Subsurface Uncertainties*. Paper SPE 68703 presented at the SPE Asia Pacific Oil and Gas Conference and Exhibition, Jakarta, Indonesia. 17- 19 April.

Chow, C., Mothershead, J., DeGolyer, MacNaughton and Cox, D. (2003). *Case History: Field Development Design Using A Risk-Based Integrated Production Model*. JPT **55**(1).

Clayton, D. (1978). *A model for association in bivariate life tables and its applications in epidemiological studies of familial tendency in chronic disease incidence*. Biometrika **65**: 141-151.

Coopersmith, E. and Cunningham, P. (2002). *A Practical Approach to Evaluating the Value of Information and Real Option Decisions in the Upstream Petroleum Industry*. Paper SPE 77586 presented at the 2002 SPE Annual Technical Conference and Exhibition, San Antonio, Texas, USA. 29 September- 2 October.

Cunningham, P. (2004). *NPV Formulas - A Fast Method for Calculating and Optimising NPV in Screening Studies*. www.omgem-ideas.com.

Demirmen, F. (2001). *Subsurface appraisal: the road from Reservoir Uncertainty to Better Economics*. Paper SPE 68603 presented at the 1993 SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, Texas. April 2- 3.

Demirmen, F. (2005). *Reliability and Uncertainty in Reserves: How the industry Fails, and a vision for improvement*. SPE 94680 presented at the SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, Texas. April 3- 5.

Dias, M. (1998). *Real Options in Upstream Petroleum: Overview of Models and Applications*.

http://sphere.rdc.puc-rio.br/marco.ind/overview.html.

Dias, M. and Rocha, K. (1998). *Petroleum Concessions with Extendible Options Using Mean Reversion with Jumps to Model Oil Prices*. Working Paper, first presented at the present at "Workshop on Real Options", Stavanger, Norway.

Economides, M., Hill, A. and Ehlig-Economides, C. (1994). *Petroleum Production Systems*. Prentice Hall. New Jersey.

Edwards, R. and Hewett, T. (1994). *Quantification of Production Prediction Uncertainty and its Impact on the Management of Oil and Gas Price Risk*. Paper SPE 28330 presented at the 1994 SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana. September 25 - 26.

Embrechts, P., McNeil, A. and Straumann, D. (1999). *Correlation: pitfalls and alternatives*. RISK: 69-71.

Etebar, S. (1995). *Captain Innovative Development Approach*. Paper SPE 30369 presented at the SPE Offshore Europe, Aberdeen, United Kingdom. September 5-8.

Frank, M. (1979). *On the Simultaneous Associativity of F (x, y) and x + y-F (x, y)*. Aequationes Mathematicae **19**: 194- 226.

Frees, W. and Valdez, E. (1998). *Understanding Relationships Using Copulas*. North American Actuarial Journal **2**(1): 1-23.

Gayton, P., Miller, D., Baker, J. and R, N. (200). *Innovative Development Engineering Techniques*. Paper SPE 65202 presented at the 2000 SPE European Petroleum conference, Paris, France. 24-25 October.

Genest, C. and MacKay, J. (1986). *The joy of copulas: bivariate distributions with uniform marginals*. The American Statistician **40**.

Genest, C. and Rivest, L. (1993). *Statistical inference procedures for bivariate Archimedean copulas*. Journal of the American Statistical Association **88**(423): 1034-1043.

Gibson, R. and E, S. (1990). *Stochastic Convenience Yield and the Pricing of Oil Contingent Claims*. Journal of Finance **45**(3): 959 -976.

Gumbel, E. (1960). *Bivariate exponential distributions*. Journal of the American Statistical Association **55**: 698 -707.

Hightower, M. and David, A. (1991). *Portfolio Modeling: A Technique for Sophisticated Oil and Gas Investors*. SPE paper 22016 Presented at the SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, TX. April 11-12.

Hotz, M. (1993). *Estimating Oil Reserves Variability by combining Geologic and Engineering Parameters*. Paper SPE 25827 presented at the 1993 SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, Texas. March 29- 30.

Iman, R. and Conover, W. (1982). *A distribution-free approach to inducing rank order correlation among input variables*. Communication in statistics, simulation and computation **11**: 311-334.

Johannessen, L., Bakken, H. and Leach, P. (1994). *Risk Analysis of Well Performance: New Approach To Optimize Well Deliverability*. Paper SPE 27594 presented at the SPE European Production Operations Conference and Exhibition, Aberdeen, United Kingdom. March 15-17.

Jonkman, R., Bos, C., Breunese, J., Morgan, D., Spencer, J. and Sondena, E. (2000). *Best Practices and Methods in Hydrocarbon Resources Estimation, Production and Emissions Forecasting, Uncertainty Evaluation and Decision Making*. Paper SPE 65144 presented at the SPE European Petroleum Conference, Paris, France. Oct 24-25.

Kabir, C., Chawathe, A., Jenkins, S., Olayomi, A., Aigbe, C. and Faparusi, D. (2002). *Developing*

New Fields Using Probabilistic Reservoir Forecasting. Paper SPE 77564 presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas. September - 2 October 8.

Mari, D. and Kotz, S. (2001). *Correlation and Dependence*. Imperial College Press. London, U.K.

Markowitz, H. (1952). *Portfolio Selection*. Journal of Finance **7**(1): 77-91.

Merrow, E. (2003). *Taking on a Cult of Mediocrity*. Journal of Upstream.

Mian, M. (2002). *Project Economics and Decision Analysis, volume two: Probabilistic Models*. Penn Well Publications. Oklahoma, USA.

Murtha, J. (1993). *Incorporating Historical Data in Monte Carlo Simulation*. Paper SPE 26245 presented at the 1993 SPE Petroleum Computer Conference, New Orleans, Louisiana, USA. July 11- 14.

Murtha, J. (1997). *Monte Carlo Simulation: Its Status and Future*. Paper SPE 37932 presented at the 1997 SPE Annual Technical Conference and Exhibition, San Antonio, Texas, USA. October 5 - 8.

Murtha, J. (2000). *Decisions involving Uncertainty, An @risk tutorial for the petroleum industry*. Palisade Corporation.

Nelsen, R. (1999). *An introduction to copulas*. Springer Verlag. New York.

Newendorp, P. (1976). *A method for treating Dependencies Between Variables in Simulation Risk Analysis Models*. SPE 5581, SPE-AIME. October.

Newondorp, P. and Schuyler, J. (2000). *Decision analysis for petroleum exploration*. Penn Well Publishing Co. Tulsa, Oklahoma.

O'Dell, M. and Lamers, E. (2003). *Subsurface Uncertainty Management in the Harweel Cluster, South Oman.* Paper SPE 84189 presented at the SPE Annual Technical Conference and Exhibition, Denver, Colorado. 5-8 October.

Otis, R. and Schneidermann, N. (1997). *A Process for Evaluating Exploration Prospects.* AAPG Bulletin **81**(7): 1087-1109.

Peterson, S. and Murtha, J. (1993). *Risk Analysis and Monte Carlo Simulation Applied to the Generation of Drilling AFE Estimates.* Paper SPE 26339 presented at the SPE Annual Technical Conference and Exhibition, Houston, Texas. 3-6 October.

Poruban, S. (2000). *Panel: risk-based decision analysis evolving into key tool of E& P sector.* The Oil and Gas Journal: 64 -68.

Rose, P. (2001). *Risk Analysis and Management of Petroleum Exploration Ventures.* AAPG. Oklahoma, USA.

Salomao, M. and Grell, A. (2001). *Uncertainty in Production Profiles on the Basis of Geostatistic Characterization and Flow Simulation.* Paper SPE 69477 presented at the SPE Latin American and Caribbean Petroleum Engineering Conference, Buenos Aires, Argentina. 25- 28 March.

Schrage, M. (2000). *Serious Play: How the World's Best Companies Simulate to Innovate.* Harvard Business School Press.

Schuyler, J. (1997). *Best Practices in Project Evaluation and Influence on Company Performance.* Paper SPE 37945 presented at the 1997 SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, Texas. March 16-18.

Schwartz, E. (1997). *The Stochastic Behaviour of Commodity Prices: Implications for Valuation and Hedging.* Journal of Finance **52**(3): 923-973.

Simpson, G., Lamb, F., Finch, J. and Dinnie, N. (2000). *The Application of Probabilistic and Qualitative Methods to Asset Management Decision Making*. Paper SPE 59455 presented at the SPE Asia Pacific Conference on Integrated Modeling for Asset Management, Yokohama, Japan. April 25-26.

Spencer, J. and Morgan, D. (1998). *Application of Forecasting and Uncertainty Methods to Production*. Paper SPE 49092 presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana. September 27 - 30.

Steagall, D. and Schiozer, D. (2001). *Uncertainty Analysis in Reservoir Production Forecasts During Appraisal And Pilot Production Phases*. Paper SPE 66399 presented at the SPE Reservoir Simulation Symposium, Houston, Texas. 11- 14 February.

Venkataraman, R. (2000). *Application of the Method of Experimental Design to Quantify Uncertainty in Production Profiles*. Paper SPE 59422 presented at the SPE Asia Pacific Conference on Integrated Modeling for Asset Management, Yokohama, Japan. April 25-26.

Wood, D. (2001). *Probabilistic Methods with Simulation Help Predict Timing, Costs of Projects*. Oil and Gas Journal: 79 - 82.

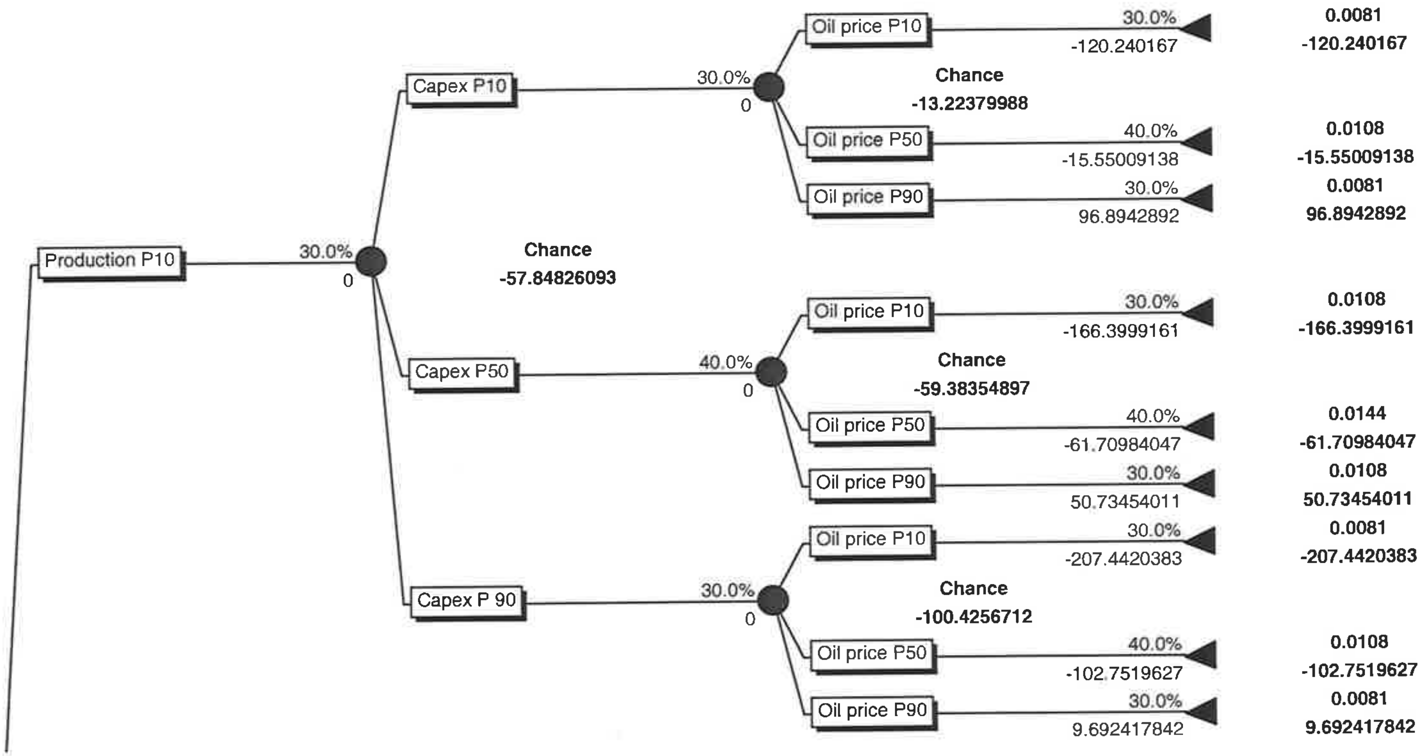
(2004). *Houston Geological Society*. http://www.hgs.org, Key word, oil and gas and reservoir.

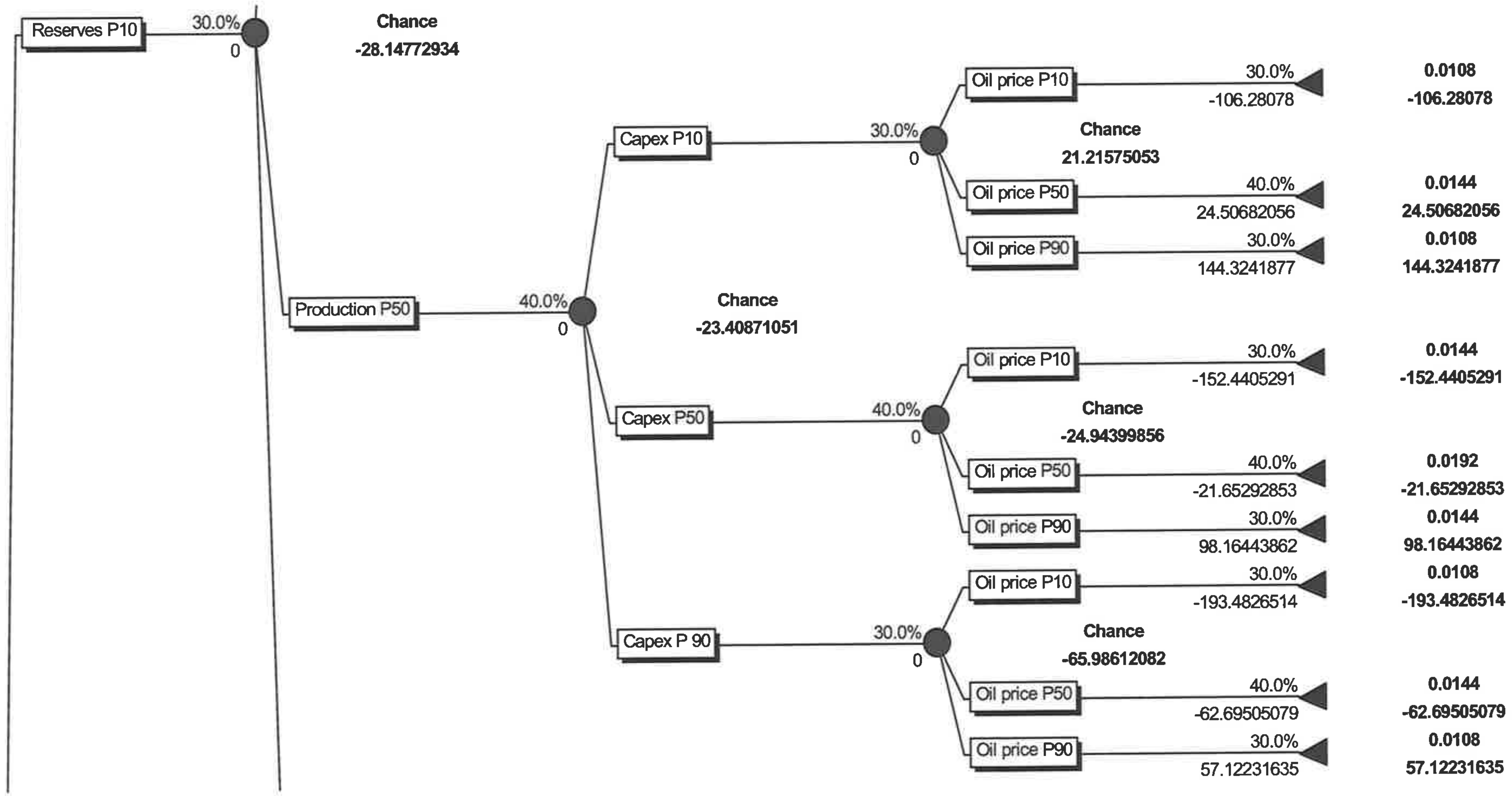
APPENDIX I
Detailed Decision tree analysis

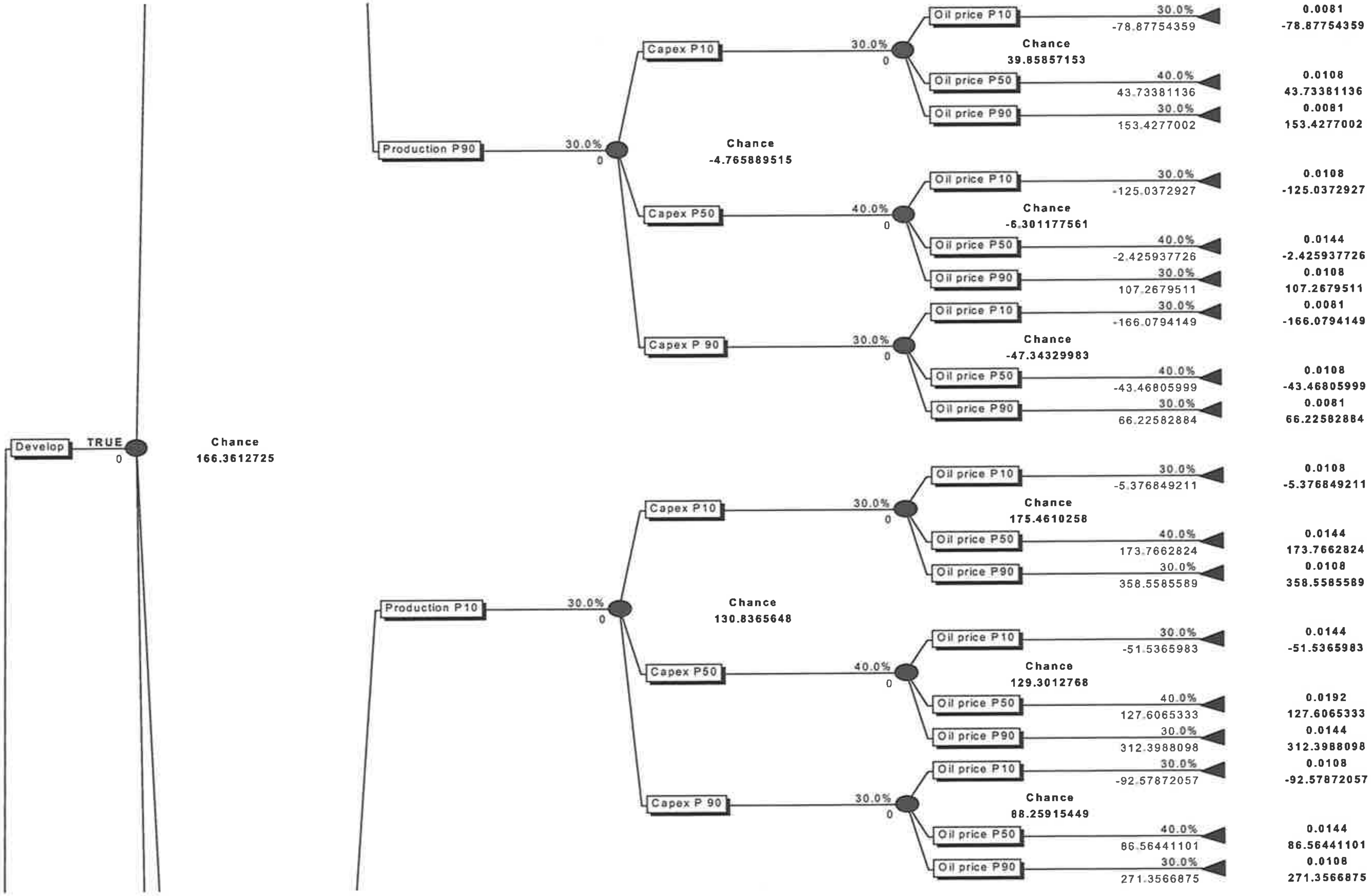
Offshore Development Decision

Decision Tree Approach

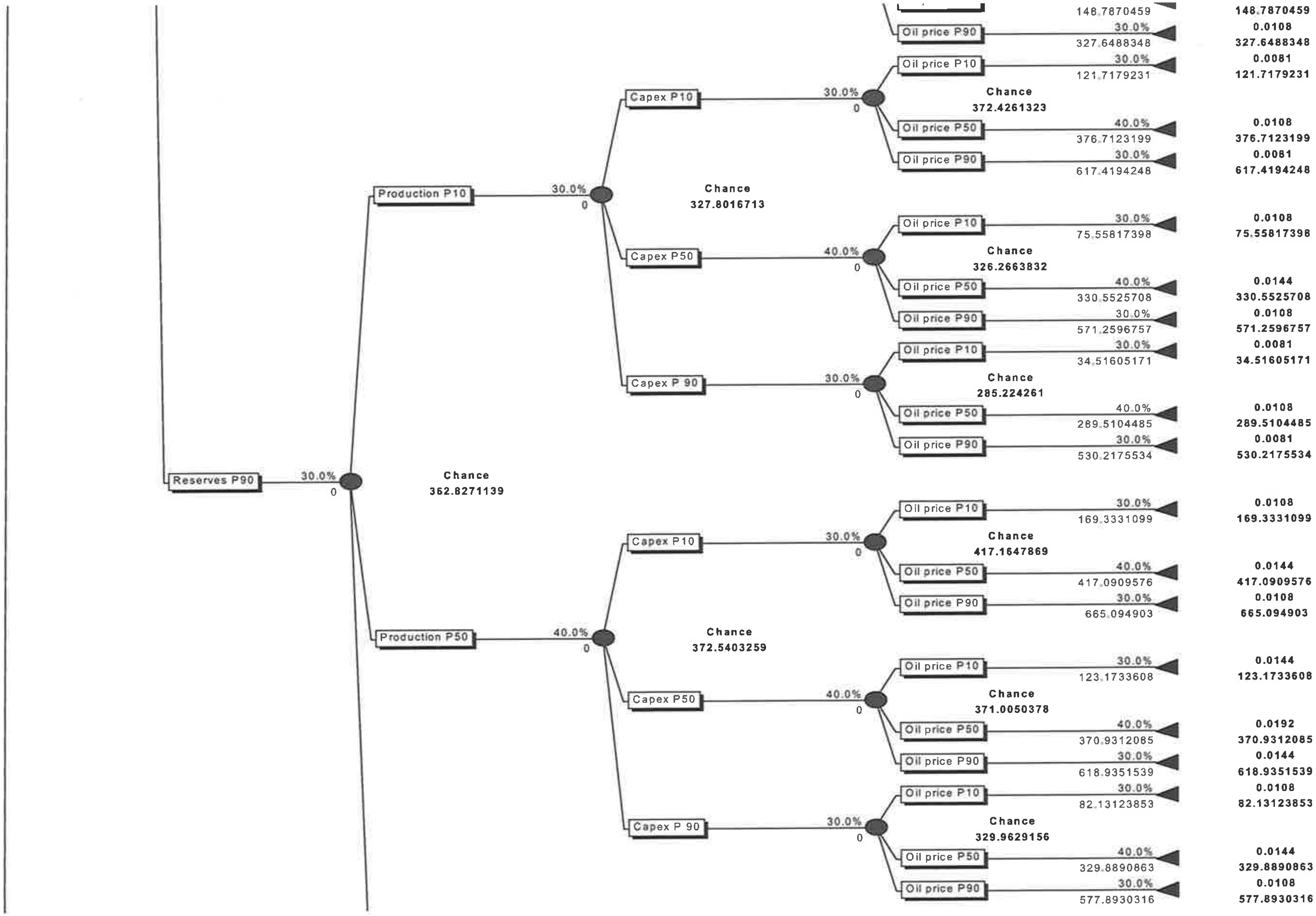
Author	Mansoor Al-Harthy
Decision	Offshore development
Number of uncertain variables	4
Number of path ways	81



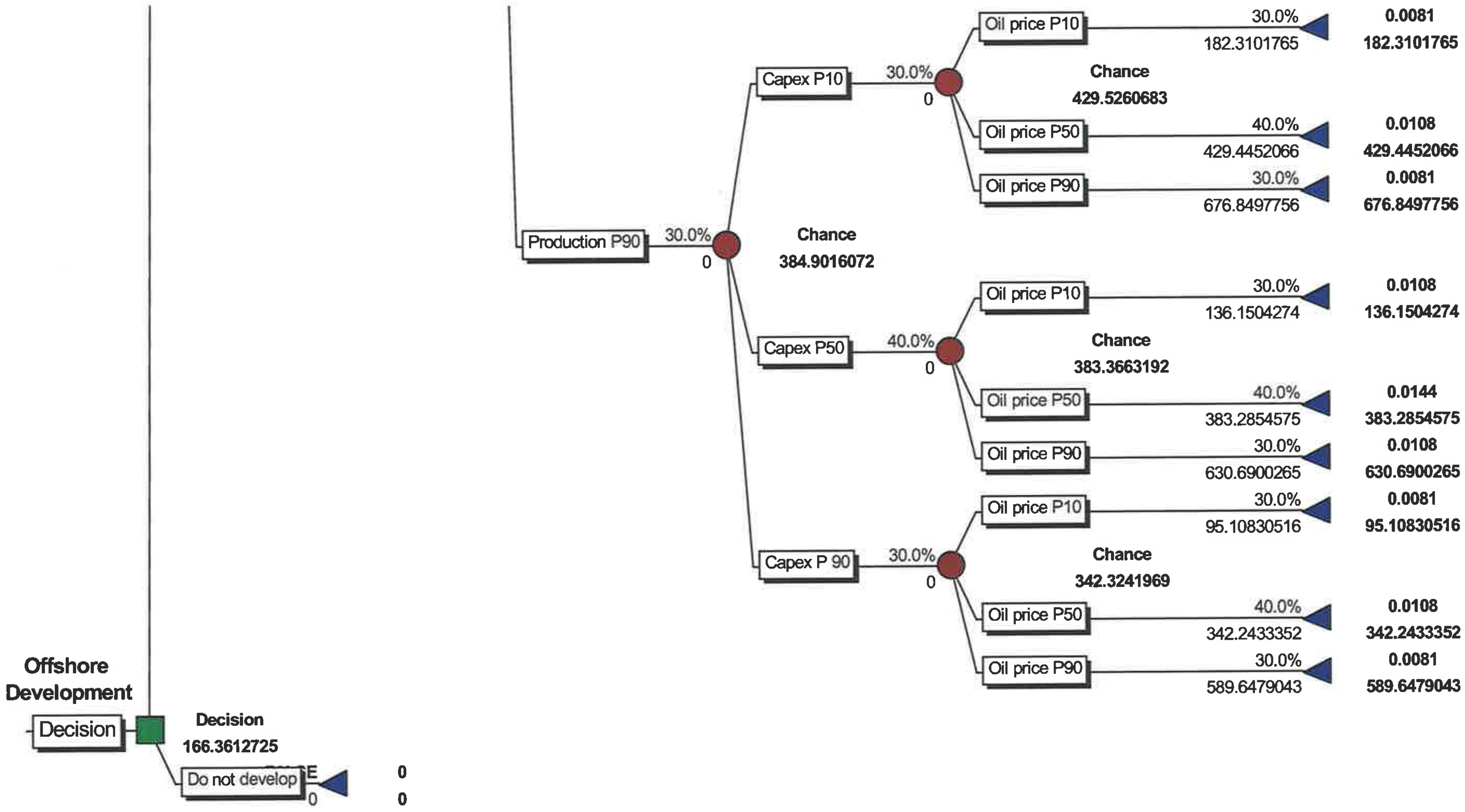

Reserves P10
30.0%
0
Chance
-28.14772934
Production P50
40.0%
0
Chance
-23.40871051
Capex P10
30.0%
0
Chance
21.21575053
Oil price P10
30.0%
-106.28078
0.0108
-106.28078
Oil price P50
40.0%
24.50682056
0.0144
24.50682056
Oil price P90
30.0%
144.3241877
0.0108
144.3241877
Capex P50
40.0%
0
Chance
-24.94399856
Oil price P10
30.0%
-152.4405291
0.0144
-152.4405291
Oil price P50
40.0%
-21.65292853
0.0192
-21.65292853
Oil price P90
30.0%
98.16443862
0.0144
98.16443862
Capex P 90
30.0%
0
Chance
-65.98612082
Oil price P10
30.0%
-193.4826514
0.0108
-193.4826514
Oil price P50
40.0%
-62.69505079
0.0144
-62.69505079
Oil price P90
30.0%
57.12231635
0.0108
57.12231635



Develop
TRUE
0
Chance
166.3612725
Production P90
30.0%
0
Chance
-4.765889515
Capex P10
30.0%
0
Chance
39.85857153
Oil price P10
30.0%
-78.87754359
0.0081
-78.87754359
Oil price P50
40.0%
43.73381136
0.0108
43.73381136
Oil price P90
30.0%
153.4277002
0.0081
153.4277002
Capex P50
40.0%
0
Chance
-6.301177561
Oil price P10
30.0%
-125.0372927
0.0108
-125.0372927
Oil price P50
40.0%
-2.425937726
0.0144
-2.425937726
Oil price P90
30.0%
107.2679511
0.0108
107.2679511
Capex P 90
30.0%
0
Chance
-47.34329983
Oil price P10
30.0%
-166.0794149
0.0081
-166.0794149
Oil price P50
40.0%
-43.46805999
0.0108
-43.46805999
Oil price P90
30.0%
66.22582884
0.0081
66.22582884
Production P10
30.0%
0
Chance
130.8365648
Capex P10
30.0%
0
Chance
175.4610258
Oil price P10
30.0%
-5.376849211
0.0108
-5.376849211
Oil price P50
40.0%
173.7662824
0.0144
173.7662824
Oil price P90
30.0%
358.5585589
0.0108
358.5585589
Capex P50
40.0%
0
Chance
129.3012768
Oil price P10
30.0%
-51.5365983
0.0144
-51.5365983
Oil price P50
40.0%
127.6065333
0.0192
127.6065333
Oil price P90
30.0%
312.3988098
0.0144
312.3988098
Capex P 90
30.0%
0
Chance
88.25915449
Oil price P10
30.0%
-92.57872057
0.0108
-92.57872057
Oil price P50
40.0%
86.56441101
0.0144
86.56441101
Oil price P90
30.0%
271.3566875
0.0108
271.3566875



148.7870459
148.7870459
Oil price P90
30.0%
327.6488348
0.0108
327.6488348
Capex P10
30.0%
0
Oil price P10
30.0%
121.7179231
0.0081
121.7179231
Chance
372.4261323
Oil price P50
40.0%
376.7123199
0.0108
376.7123199
Oil price P90
30.0%
617.4194248
0.0081
617.4194248
Production P10
30.0%
0
Chance
327.8016713
Capex P50
40.0%
0
Oil price P10
30.0%
75.55817398
0.0108
75.55817398
Chance
326.2663832
Oil price P50
40.0%
330.5525708
0.0144
330.5525708
Oil price P90
30.0%
571.2596757
0.0108
571.2596757
Capex P 90
30.0%
0
Oil price P10
30.0%
34.51605171
0.0081
34.51605171
Chance
285.224261
Oil price P50
40.0%
289.5104485
0.0108
289.5104485
Oil price P90
30.0%
530.2175534
0.0081
530.2175534
Reserves P90
30.0%
0
Chance
362.8271139
Capex P10
30.0%
0
Oil price P10
30.0%
169.3331099
0.0108
169.3331099
Chance
417.1647869
Oil price P50
40.0%
417.0909576
0.0144
417.0909576
Oil price P90
30.0%
665.094903
0.0108
665.094903
Production P50
40.0%
0
Chance
372.5403259
Capex P50
40.0%
0
Oil price P10
30.0%
123.1733608
0.0144
123.1733608
Chance
371.0050378
Oil price P50
40.0%
370.9312085
0.0192
370.9312085
Oil price P90
30.0%
618.9351539
0.0144
618.9351539
Capex P 90
30.0%
0
Oil price P10
30.0%
82.13123853
0.0108
82.13123853
Chance
329.9629156
Oil price P50
40.0%
329.8890863
0.0144
329.8890863
Oil price P90
30.0%
577.8930316
0.0108
577.8930316


Oil price P10
30.0%
182.3101765
0.0081
182.3101765
Capex P10
30.0%
0
Chance
429.5260683
Oil price P50
40.0%
429.4452066
0.0108
429.4452066
Oil price P90
30.0%
676.8497756
0.0081
676.8497756
Production P90
30.0%
0
Chance
384.9016072
Oil price P10
30.0%
136.1504274
0.0108
136.1504274
Capex P50
40.0%
0
Chance
383.3663192
Oil price P50
40.0%
383.2854575
0.0144
383.2854575
Oil price P90
30.0%
630.6900265
0.0108
630.6900265
Oil price P10
30.0%
95.10830516
0.0081
95.10830516
Capex P 90
30.0%
0
Chance
342.3241969
Oil price P50
40.0%
342.2433352
0.0108
342.2433352
Oil price P90
30.0%
589.6479043
0.0081
589.6479043
Offshore Development
Decision
Decision
166.3612725
Do not develop
0
0
0

APPENDIX II
Related Publication work

Al-Harthy, M. Khurana, A. Begg, S. Bratvold, R. (2006) Sequential and Systems Approaches for Evaluating Investment Decisions: Influence of functional dependencies and interactions.
The Australian Petroleum Production and Exploration Association (APPEA) Journal, v. 46, (part 1), pp. 7-11, May, 2006.

Copulas: A New Technique to Model Dependence in Petroleum Decision Making

Mansoor Al-Harthy [a*], Steve Begg [b], Reidar B Bratvold [c]

[a] *Australian School of Petroleum, University of Adelaide, Australia*
Email: mansoor.alharthy@adelaide.edu.au

[b] *Australian School of Petroleum, University of Adelaide, Australia*
Email: steve.begg@adelaide.edu.au

[c] *Australian School of Petroleum, University of Adelaide, Australia*
Now with the University of Stavanger
Email: reidar.bratvold@ uis.no

Abstract

A key step in valuing petroleum investment opportunities is to construct a model that portrays the uncertainty inherent in the investment decision. In almost all such cases, several of the random variables that are relevant for the model are correlated. Properly accounting for and modelling these correlations is essential in deriving reliable valuations for decision support.

The Envelope method and the Iman - Conover method are popular approaches to model dependency in the petroleum industry. Although these models work well in many cases, there are situations where they fail to accurately account for important characteristics of the correlations.

In many cases the structure of the dependence between two random variables is important. The approaches typically used to model dependence in oil and gas evaluations often fail to address the dependence structure. The copulas technique, which is well known in financial risk management and insurance applications, has

[*] Corresponding author. Email: mansoor.alharthy@adelaide.edu.au
Tele +618 83038041, Fax + 618 83038030.
Australian School of Petroleum, The University of Adelaide
South Australia 5005

proven to be a superior tool for modelling dependency structures. Yet, to our knowledge, it has rarely been used in petroleum applications.

A copula is a statistical concept that relates random variables. It is a function that links the marginal distributions to the joint distribution. A copula can model the dependence structure given any type of marginal distribution, which is not possible with other correlation measures. This is illustrated by the fact that the copula approach is able to separate the marginal distribution from the correlation.

The objective of this paper is to illustrate the potential benefits of using copulas to model dependencies in oil and gas applications with a particular focus on the reserves problem. This paper introduces the copulas method and then compares and contrasts it with some of the more commonly used approaches to model dependence. We then show that the traditional methods have problems in accurately revealing the dependence structure in the tails of the variable distributions. Finally, we illustrate how the dependence structure can be captured and modelled using the copulas approach.

Key words: Copulas; Correlations; Probabilistic Reserves Estimation; Risk and Decision Making

1. Introduction

The performance of oil and gas upstream projects for the last decade has been less than stellar for many upstream companies. Merrow (2003) concluded that one in eight of all major offshore developments fall into the "disaster" category. It is our belief that inaccurate reserve estimates is one of the major contributors to the poor performance of the industry. Studies on North Sea fields (UK and Norway) show that the official estimates of reserves are changing considerably, in some cases as much as fourfold, during field life. The same trend has been observed in the Gulf of Mexico

(Demirmen, 2005). One of the most common problems in exploration and development projects is the failure to estimate the low end such as proven reserves (Rose, 2001). Rose also points out the problem of overestimating the low sides and underestimating the high sides. For example, knowing the high potential is essential in making decisions about the sizing of the production capacity. This is particularly relevant for development projects where the extreme values are important. Demirmen further indicates that the concept of reduction in reserves uncertainty throughout the life of the field is rarely observed in practice due to many factors that impact the estimates over time. One of these factors is that the rules of statistics are not honoured; specifically, dependencies are not properly accounted for in the calculation of reserves.

This paper deals with modelling dependencies in probabilistic reserves estimates. Ignoring these dependencies may lead to erroneous estimates of proven and probable reserves, which in turn lead to poor estimates of Net Present Values (NPV). The outline of the paper is as follows: Section 2 will discuss dependence measures that are used to analyse statistical dependence and then introduce the data generating methods that are used in oil and gas evaluations to model dependencies. Section 3 will introduce the copulas, their construction and how to run simulations with them. Section 4 introduces a simple reserves estimates problem to investigate the impact of using these dependency methods. Sections 5, 6 and 7 consist of the results, discussion and conclusions resulting from this investigation.

2. Correlations and Dependencies

The terms correlation and dependence are often used interchangeably in the literature. Although this is not strictly correct, for clarity we will also do so in this paper.

Although many associate the word "correlations" with linear correlations, there are many approaches to model dependencies. There are three common measures used to analyse statistical dependence: Pearson, Spearman and Kendall's tau. These are data analysis techniques designed to capture the direction and the magnitude of a correlation.

Pearson's (linear) correlation is the most common method used to measure dependence between two variables. The Pearson correlation coefficient (r), which correlates two variables - X and Y - is defined by

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2} \sqrt{\sum (Y_i - \bar{Y})^2}} \qquad (1)$$

Where X_i and Y_i are pair of points, $\bar{X}$ *and* $\bar{Y}$ are their means. The Pearson regression correlation is appropriate for showing the linear relationship between variables. The Pearson correlation is generally not invariant under monotone non-linear transformation. For example, the correlation between X and Y is not the same as the correlation between Ln (X) and Ln (Y).

Spearman's rank correlation coefficient was developed by Spearman in the early 1900s. It is calculated using rankings of values rather than actual values (as the Pearson correlation coefficient). Spearman's rank correlation is used as a measure of linear relationship between two sets of ranked data clusters around a straight line. Spearman's rank correlation is similar to the Pearson correlation method in that it takes on values between -1 and +1.

The Spearman correlation coefficient is calculated as follows:

$$r_s = \frac{\sum (R_x - \bar{R}_x)(R_y - \bar{R}_y)}{\sqrt{\sum (R_x - \bar{R}_x)^2} \sqrt{\sum (R_y - \bar{R}_y)^2}} \qquad (2)$$

Where R_x and R_y are the ranks of X and Y respectively; $\bar{R}_x$ and $\bar{R}_y$ are the mean rank of the variables.

Because the correlation is computed on the ranks, the method is known as a "distribution free" approach as there are no assumptions regarding the underlying distributions. This also means that the Spearman rank order correlation doesn't require that the relationship between two variables is linear.

Kendall's tau (τ) correlation is similar to the rank order correlation in its assumptions. It does not require any assumption about the distribution nor does it require the relationship to be linear. As with Pearson and Spearman, the correlation coefficient takes on values between -1 and +1. Kendall's tau can be calculated from the following equation:

$$\tau = \frac{\#(concordant\ pairs) - \#(discordant\ pairs)}{\frac{n\,(n-1)}{2}} \tag{3}$$

Where n is the samples size and concordant pairs means the number of pairs that are moving in the same direction and discordant pairs are those pairs that are moving in the opposite direction to each other's.

Spearman's rank correlation is more widely used because it is much easier to compute than Kendall's tau. Another interpretation of Kendall's tau is that it is the difference between probability of concordant and discordant pairs. As a result of this interpretation, it tends to have better statistical properties than the Spearman rank correlation.

A major shortcoming of the Pearson correlation is that it is not invariant under non-linear transformation. Both the Spearman and the Kendall approach overcome this deficiency. However, none of the three methods discussed so far are able to capture the dependency structure. As we will illustrate later, the copulas approach is able to capture and model the structure.

The Pearson, Spearman and Kendall's tau are data analysis techniques and they cannot, as stand-alone approaches, be used to generate dependence in a stochastic environment. In order to model dependence in a stochastic form, techniques for generation data correlations are required. The most common techniques are the Envelope method, the Iman-Conover method, which uses the Spearman rank correlation, the regression fitting method, and the Copulas method, which uses the Kendall's tau.

The Envelope method, sometimes called the box method, is a useful technique for modelling partial dependencies. The method is well described by Newendorp and Schuyler (2000), Mian (2002) and Murtha (2000) and uses an "envelope" around the data to capture the partial dependency.

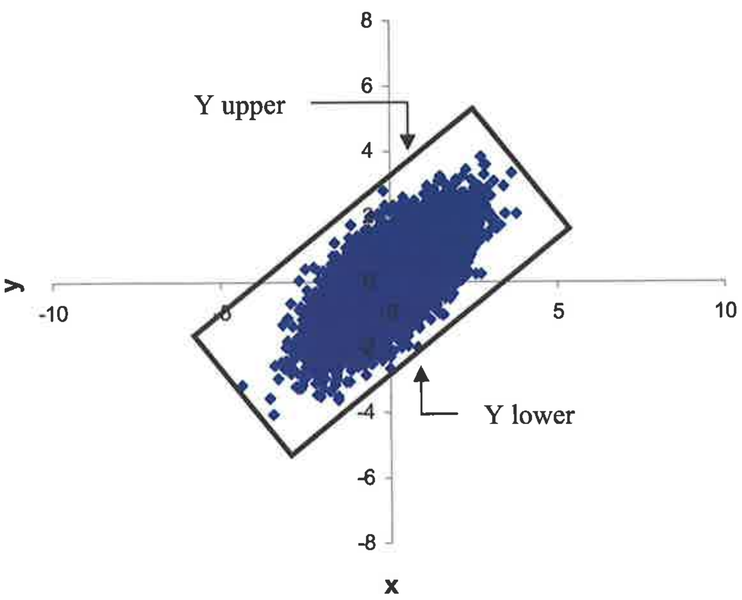


Fig. 1. Envelope method

In order to capture the dependence structure two lines are estimated Y upper and Y lower (Fig.1). For a simulation, a random X value is generated and the corresponding Y upper and Y lower values for that X are calculated. Together with a normalized value are used to generate the correspondence value of Y to an X (Equation 5). This approach, which may be lacking in mathematical rigour, is very flexible and can capture many structural patterns. However, although the envelope method is conceptually simple, its implementation may seem somewhat tricky and involved.

$$Y_x = (Y_{lower})_x + [(Y_{upper})_x - (Y_{lower})_x](Y_{norm})_x \tag{5}$$

where

$(Y_{lower})_x$ = lowest possible value of Y, given X

$(Y_{upper})_x$ = highest possible value of Y, given X

$(Y_{norm})_x$ = sampled value from dimensionless normalized distribution

Iman –Conover method

The Iman-Conover (Iman and Conover, 1982) is the algorithm used in Monte Carlo Simulation software such as @Risk™ and Crystal Ball™. To model dependence between variables the marginal distributions of each variable are defined. Then the Spearman's rank order correlation is calculated. For a simulation of correlated variable X and Y with a rank correlation matrix R^*: first, the Iman-Conover generates an independent score matrix for each variable using the van der Waerden scores. The use of van der Waerden scores is based on the inverse of the standard normal cumulative distribution function. Second, it will correlate between the two independent score matrices using Cholesky decomposition to generate a correlated matrix M^* that is close to the rank correlation R^*. Third, it generates independent random samples from X and Y and finally, rearranges X and Y so they will have the same rank correlation matrix as M^*.

Regression fitting

A simple regression approach finds the best-fit line using regression analysis. The regression calculation not only finds the slope and intercept of the line but also the standard error. The standard error is important for simulation because it defines the standard deviations of the errors. The residual used in estimating the standard error is defined as the difference between the input data and the estimated data. For simulation the regression fitting method will generate the correlated variables based on the following equation.

$$Y = mX + b + N(0, e) \tag{6}$$

Where m is the slope of the regression line, b is the intercept and N (0,e) is the standard normal distribution with standard deviation (e).

3. Copulas

The primary disadvantage of the dependency models discussed above is their lack of ability to capture and model the dependency structure or pattern. In this section an alternative is discussed in which a joint distribution is constructed using a copula. Although copulas are well known in insurance and finance, their application in petroleum is still in its early stages. To the best of our knowledge, there are only two papers discussing petroleum applications. Armstrong et al. (2004) investigated the value of an option to acquire new information in which Bayesian analysis was used to account for new information. The traditional framework of the Bayesian analysis is based on the joint normal distribution where the lower and upper tails are symmetrical. Armstrong et al. (2004) proposed an alternative Bayesian analysis that is based on Archimedean copulas where the joint distribution does not have to be normal and there is flexibility to have a lower tail or upper tail dependence based on the specific type of copula.

Accioly et al. (2004) also discussed the suitability of using copulas to model dependence for non-Gaussian data. They presented two estimation procedures: nonparametric and semiparametric to estimate copulas parameters. They used a sample of 188 exploratory offshore wells drilled in the Gulf of Mexico and explored the relationship between drilling duration and measured depth. They concluded that the Clayton copula reproduced the pattern of the original data and recommended the use of copulas to model the relationship between operating expenditure (Opex) and production, since cost elements are production dependent.

This paper illustrates how copulas can be used to model dependence in the probabilistic reserves estimates. Furthermore, it will show the potential benefit of the copula approach by comparing it with current correlation approaches, highlighting and discussing its relevance and advantages in petroleum applications. The first step in this discussion is to introduce the copulas, their construction and how to run simulations with them.

The essence of the copula approach is that a joint distribution of random variables can be expressed as a function of the marginal distributions. The key theorem on copulas is by Sklar (1959):

Let H be a joint distribution function with margins F_1 and F_2. Then there exists a copula C such that for all x, y $F_1, F_2 \in$ R:

$$H(x,y) = C\left[F_1(x), F_2(y)\right] \tag{7}$$

If the margins F_1 and F_2 are continuous, then C is unique; otherwise, C is unique on the range F_1 x F_2 .

This theorem and the characteristics of copulas are extensively discussed in Nelsen (1999). Sklar's theorem is completely general. For any type of marginal distribution a joint distribution can be constructed using the copula function. The copula separates the marginal distribution from the correlation and the copula itself can capture the dependence structure. This is an essential property of copulas. To show how this property makes copulas superior to the traditional dependence measures. Embrechts et al. (1999) presented a fallacy, which stated, "Marginal distribution and correlation determine the joint distribution". This is only true in the case elliptical distribution, which the normal distribution is one. In these cases, having two marginal distributions and a correlation, there is only one joint distribution that might fit. Yet, outside the elliptical distributions, there are many joint distributions that might fit. This can be seen clearly by simulating two variables with the same marginal distribution, represented by a normal distribution with a mean of zero and standard deviation of 1. These two marginal distributions have a linear correlation of 0.7 (Fig. 2).

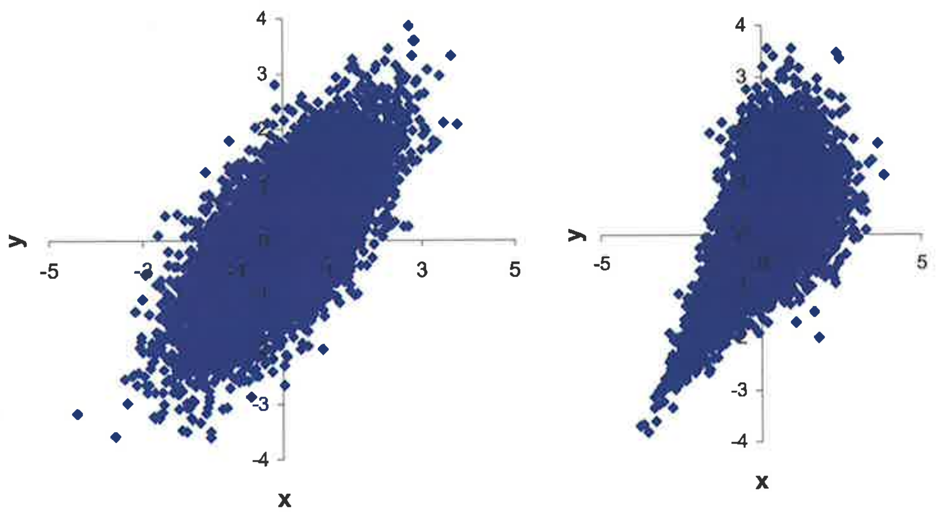


Fig. 2. Two identical distribution x~N (0,1), y~N(0,1) and same correlation = 0.7, but different dependence structure

The distribution on the left side shows the case where the joint distribution is normal and the dependence structure shape is elliptical and constructed using a normal copula. While using the same marginal distribution and the same linear correlation, a different joint distribution can be constructed showing a completely different dependence structure. The distribution on the right side is constructed using a Clayton copula. This illustrates the limited ability of linear correlation to capture the dependence pattern for non-elliptical distributions.

There are different classes and families of copulas but, for this study, we will work with Archimedean copulas, which is defined as:

$$C\left(u,v\right)=\phi^{[-1]}\left(\phi\left(u\right)+\phi\left(v\right)\right) \quad for\ u,\ v\in[0,1] \tag{8}$$

$C\left(u,v\right)$ is the copula function with u and v as uniform distributions, ϕ is the generator and $\phi^{[-1]}$ is the inverse generator. Genest et al. (1986) presented several properties of this class of copulas. The choice of generator determines the copula family (Table 1). The Archimedean copulas have different families and the popular ones are:

Clayton Family (Clayton, 1978) This family has lower tail dependence for $\theta > 0$ and has the following function:

$$C_{clatyon}\left(u,v\right)=\left(u^{-\theta}+v^{-\theta}-1\right)^{\frac{-1}{\theta}} \tag{9}$$

Gumbel Family (Gumbel, 1960) This copula family has upper tail dependence for $\theta \geq 1$ and it has the following function:

$$C_{Gumbel}\left(u,v\right)=\exp\left\{-\left[\left(-\ln u\right)^{\theta}+\left(-\ln v\right)^{\theta}\right]^{\frac{1}{\theta}}\right\} \tag{10}$$

Frank Family (Frank, 1979). This copula family has a flexible theta that ranges $-\infty \leq \theta \leq \infty$ and has the following function:

$$C_{Frank}(u,v) = -\frac{1}{\theta}\ln\left(1+\frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{e^{-\theta}-1}\right) \tag{11}$$

Table 1: Archimedean copulas with their generators and ranges

Family	Generators $\phi\ (t)$	Ranges of theta
Clayton 1978	$t^{-\theta}-1$	$(0,\infty)$
Gumbel 1960	$(-\ln t)^{\theta}$	$(1,\infty)$
Frank 1979	$\ln\left(\frac{e^{t\theta}-1}{e^{\theta}-1}\right)$	$(-\infty,\infty)$

The relationship between the generator and the copula function is easily identified. If the generator $\phi(t)$ of Clayton copula from Table 1 is substituted back into the general form of the Archimedean copulas (Equation. 8), we get the copula function of Clayton (Equation. 9). This means that the choice of the generator determines the copulas family.

Once the types of copulas are identified then the next step is to construct and simulate the copula.

3.1 Construction and simulation of Archimedean copulas

Genest et al. (1993) illustrated that these types of families have the following distribution function $Kc\,(t) = P\{C(u,v) \le t\}$.

$$Kc\,(t) = t - \frac{\phi\,(t)}{\phi^{\prime}\left(t^{+}\right)} \qquad \textit{for any } t \textit{ in } [0,1] \tag{12}$$

Where $\phi\ (t)$ is the generator and $\phi^{\prime}\left(t^{+}\right)$ is the derivative of the generator. Equation (12) is known as the parametric estimate. In order to compare and simulate a copula, the value of theta needs to be determined. Genest proposed a relationship between the type of copulas and Kendall's tau and showed that through this relationship the value of theta can be determined. This relationship is expressed in the following theorem:

Let X and Y be random variables with an Archimedean copula C generated by $\phi(t)$, Kendall's tau of X and Y is given by:

$$\tau_c = 1 + 4\int_0^1 \frac{\phi(t)}{\phi'(t)}\,dt \tag{13}$$

For example, for a Gumbel copula the Kendall's tau correlation can be expressed as a function of theta:

$$\tau_c = 1 + \frac{4}{\theta}\int_0^1 t\ln t\,dt = 1 - \frac{1}{\theta}$$

Hence by assuming a certain type of copula and calculating the Kendall's tau correlation using equation (3) the value of theta can be determined. Using equation (13) the value of Kendall's tau for each family is shown in Table 2.

Table 2: Kendall tau values as a function of theta for each family and their ranges

Family	Ranges	Tau
Clayton 1978	$(0, \infty)$	$\frac{\theta}{\theta+2}$
Gumbel 1960	$(1, \infty)$	$1 - \frac{1}{\theta}$
Frank 1979	$(-\infty, \infty)$	$1-\frac{4}{\theta}(D_1(-\theta)-1)$

Note: D_1 is the Debye function of the first order $D_1 = \frac{1}{\theta}\int_0^{\theta}\frac{t}{e^t-1}\,dt$

Once Kendall's tau is calculated and the value of theta is determined for each copula, then the next step is to compare among dependency methods. Frees and Valdez (1998) points out that to find the generator that fits the data one needs to compare the parametric estimate $Kc(t)$, by the procedure described earlier for generating the copula, with the nonparametric estimate $K(t)$. The nonparametric estimate is generated as follows:

1. Define the pseudo-observations

$$T_i = \left\{ \begin{array}{l} number\ of\ (X_{1j}, X_{2j})\ such\ that \\ X_{1j} < X_{1i}\ and\ X_{2j} < X_{2i} \end{array} \right\} / (n-1)\ for\ i = 1, \ldots\ldots, n. \tag{14}$$

2. Second, construct the estimate of K as $K(t) =$ Proportion of $T_i \le t$

To construct the nonparametric estimate, a pseudo-observation (T_i) is estimated. This pseudo-observation captures the movement of each point with the movement of other points. In this way we determine the correlation between variables in a similar approach as used in the Kendall's tau correlation method. We then construct a cumulative distribution function of the pseudo observations $(T_i{}'s)$ for all the data points. This process generates the nonparametric estimates, which are then compared with the cumulative distribution function of the parametric estimates. For each generator chosen from Table 1, the value of theta is substituted into the parametric equation (Eq. 12) to be compared with the nonparametric $K(t)$. We then choose the generator that best resembles the nonparametric results. The 'best' generator is determined by the Minimum Distance $MD = \int [Kc(t) - K(t)]^2\, dK(t)$.

Another possible approach to determine the best generator is a Q-Q plot, which plots the cumulative distribution function of the nonparametric estimate against the parametric estimates.

Once the optimal copula has been determined, it is used to generate the desired marginal distributions U and V that will be used in the Monte Carlo Simulation model. Several types of simulation algorithms have been proposed in the literature. A general simulation algorithm was discussed by Mari et al, (2001):

1. Generate two uniform and independent random variables u and q
2. Calculate $v = C_u^{-1}(q)$, where $C_u = \dfrac{\partial C(u,v)}{\partial u}$.

The second step is implemented by taking the derivative of the copula function with respect to u. The next step is to find the inverse of the derivative C_u^{-1}. In general this procedure works well where C_u^{-1} has a closed form expression and where the inverse function can be solved analytically. This algorithm will work well for the Clayton but not for the Gumbel copula because it has no closed form expression for the inverse. To overcome this problem, another simulation procedure can be used (Nelsen, 1999):

1. Simulate two independent random uniform variables s and q on $[0,1]$.
2. Set $t = Kc^{-1}(q)$, Where Kc is the distribution function of C (u, v)
3. Set $u = \phi^{-1}(s\phi(t))$ and $v = \phi^{-1}((1-s)\phi(t))$

The second step in this simulation uses a numerical root finding method to solve for the inverse of the parametric estimates. Once the value of t is found, it is substituted into the equations in the third step to derive the desired correlated variables u and v. This simulation is powerful and can be used to solve for all three type of copulas; Gumbel, Frank and Clayton.

The previous sections introduce the dependency methods such as Envelope, Iman-Conover, Regression fitting and how they are used in simulation. This section has introduced the copulas methods, the construction and the simulation algorithm. The next step is to explore which of these dependency methods captures dependence structure and what impact this has on a decision variable. The example used to investigate this is an oil reserves estimation problem.

4. Example - Reserves Estimation

The example we will discuss is an oil reservoir with a focus on technical reserves, which are estimated using the volumetric equation and a recovery factor. The focus is on probabilistic reserves, so area, net thickness, recovery factor, porosity, water saturation and formation volume factor are all treated as uncertain. The stochastic inputs used in calculating the Original Oil in Place (OOIP) are listed in Table 3. These inputs are correlated and the objective is to explore the impact of the choice of correlation model. Data were collected from nearby producing fields, and a plot of net thickness versus recovery factor shows a positive correlation (Fig. 3). Furthermore, as shown in Fig. 4, porosity and water saturation are negatively correlated. The best distribution for the input variables is derived by using the chi square measure.

Table 3. Input used for Calculation of Original Oil in Place (OOIP)

Original Oil in Place (OOIP) input:	Distribution	Minimum	Most likely	Maximum	Mean	Variance
Area (acre)	Lognormal				700	220
Thickness (ft)	Triangular	40	170	300		
Porosity (%)	Triangular	0.2184	0.35	0.4653		
Water saturation (Sw) (%)	Triangular	0.1915	0.30184	0.4969		
Formation Volume Factor (FVF)	Triangular	1.1	1.15	1.2		

Note: Lognormal distribution is truncated between 400 and 1200 acre

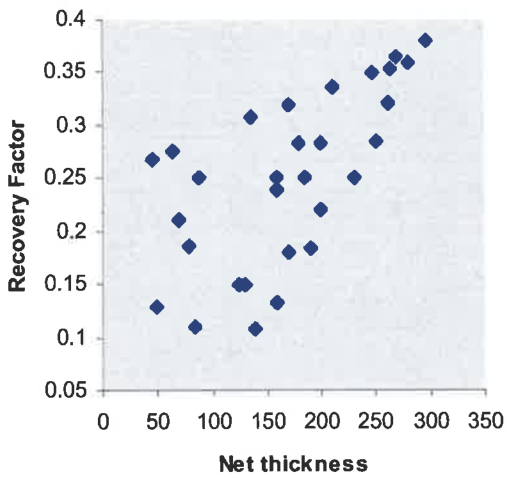


Fig.3. Positive dependency between thickness and recovery factor

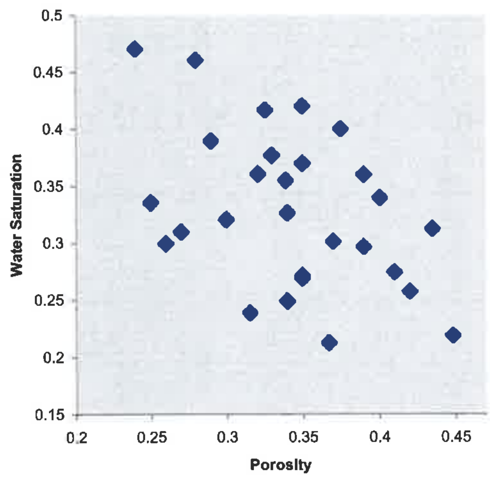


Fig.4. Negative dependency between Porosity and Water saturation

5. Results

To investigate the impact of dependencies methods, the data in Table 3 were used to calculate the Original Oil in Place (OOIP). Furthermore, each dependency method was used to model the positive dependence between net thickness and recovery factor and the negative dependence between porosity and water saturation. The multiplication of original oil in place and recovery factor yielded the technical reserves.

Iman – Conover Method

To model the correlation between net thickness and recovery factor using the Iman-Conover method, triangular distributions were chosen as the best fit for both variables. The Spearman correlation for net thickness and recovery factor is 0.67. Triangular distributions were also found to be the best fit for porosity and water saturation with a Spearman correlation of –0.388.

Envelope Method

Lower and upper lines were estimated for both of the two sets of dependent variables and were used for simulation of the variables as shown in Fig.5. For the Recovery factor and net thickness a multiple line envelope was used to capture the shape of the dependence structure while for the porosity and water saturation a classic two lines envelope method was used.

Regression Fitting

A regression analysis was used to find the best line for both net thickness and recovery factor and for porosity and water saturation dependence structure. The standard error was also estimated for both of the two sets of dependent variables. The regression gave the following equation to correlate net thickness and recovery factor.

$$\mathrm{Recovery}\ Factor = 0.1267 + 0.000721\,(Net\ Thickness) + Normal(0,\ 0.0622)$$

Note that the above equation was modified to eliminate the possibility of a negative recovery factor. A cut-off for the recovery of 0.10 was used, which is the minimum value in the original data.

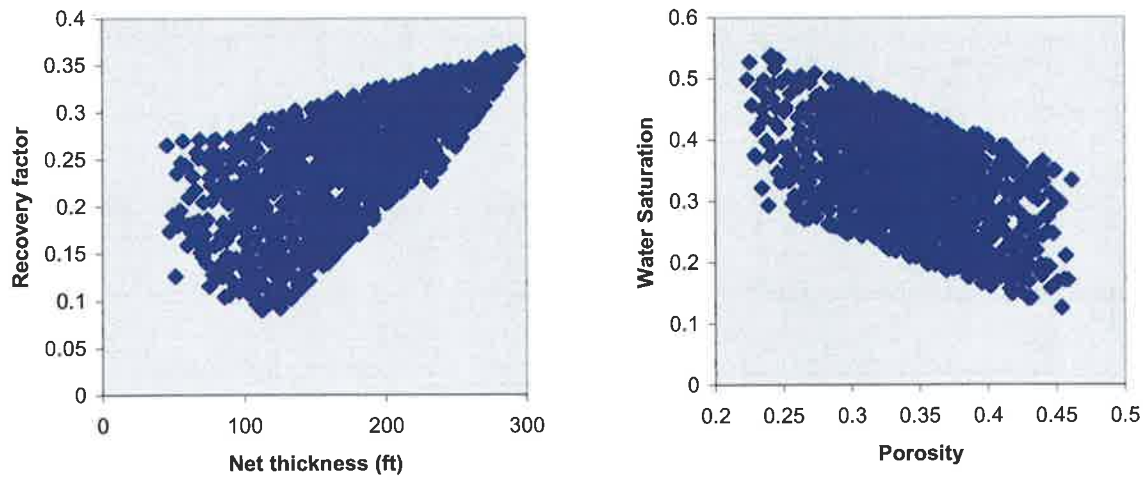


Fig.5. Envelope Method net thickness vs. Recovery factor (left) and Porosity vs. water saturation (right)

Copula Method

To model dependence using copulas, Kendall's tau was calculated using equation (3) giving $\tau = 0.51$ for net thickness versus recovery factor and $\tau = -0.283$ for porosity versus water saturation. The Gumbel copula was chosen for the net thickness and recovery factor as the best fit among copulas as it will be explained later. The next step was to use Table 2 to calculate the value of theta - $\theta = 2.04$. For porosity and water saturation the Frank copula was chosen as the best fit and the value of theta is $\theta = -2.72$. This is explained because of the ability of Frank copula to model positive and negative dependency where as Gumbel and Clayton are suitable for positive dependencies.

Table 4 shows the mean and standard deviation using the four correlation models. We also show the mean and standard deviation assuming no dependency.

In the last column of Table 4, which includes both dependencies, we see a significant impact on both the mean and standard deviation of the reserves estimates compared with the no dependency case. This confirms that dependencies do matter and should be included in any simulation to ensure optimal decision-making.

Table. 4. Reserves (Million of barrels) mean and standard deviation and dependency methods

Methods	Measure	No Dependency	Porosity and Water Saturation only (1)	Net thickness and RF only (2)	Both (1) and (2)
Iman-Conover	Mean	53.52	53.59	55.70	55.83
	SD	27.67	28.16	31.80	32.17
Envelope	Mean	53.52	53.93	47.15	47.06
	SD	27.67	29.23	28.52	30.05
Regression fitting	Mean	53.52	53.73	48.57	48.58
	SD	27.67	28.31	29.25	30.60
Copulas	Mean	53.52	53.62	55.84	55.92
	SD	27.67	28.11	32.68	32.67

All of the models indicated that dependencies have a larger impact on the standard deviation than on the mean. The difference in the means, comparing the no-dependencies with the correlated results ranged between 4 – 12 %. The standard deviation, however, increased by 8 -18 % in the correlated versus no-dependencies cases. This confirmed what other authors have shown: ignoring dependence might lead to a significant underestimation of the uncertainty in the simulated results. From the cases where we only include one dependency at a time, we see that positive dependence has a larger impact on both the mean and standard deviation than negative dependence except for the Envelope and regression fitting approach. This was further supported with a tornado plot (Fig. 6), which showed that thickness and recovery factor have a larger impact on the reserves simulation than porosity and water saturation. This confirms that it is particularly important to include dependencies between those variables that have the largest impact on the output. From this point our analysis focussed on the dependency between thickness and recovery factor since they had the largest impact on technical reserves.

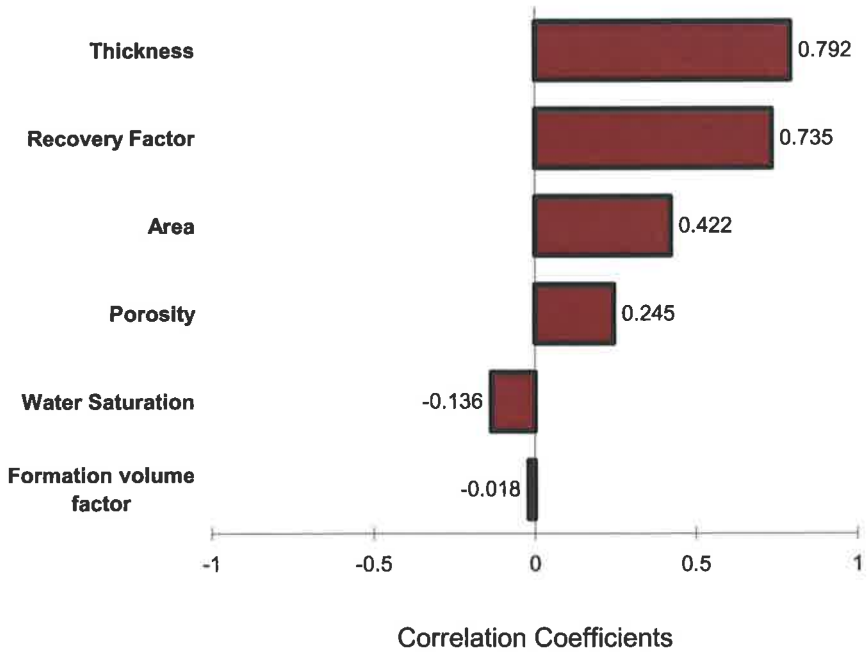


Fig. 6. Sensitivity analysis for the technical reserves

The next question we wanted to address was: Which one of the correlation models best represents the original data?

The Frees approach with both graphical and quantitative analysis to find the best fit was used. From Fig. 3, we see that the original data correlation pattern has strong upper tail dependence. Using a Q-Q plot to analyse the correlation in particular above the 50th percentile (Figs. 7 - 10), we saw that the Gumbel copula gave the best match with the original data.

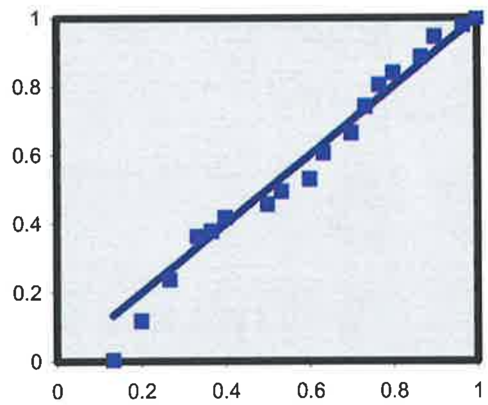


Fig. 7. Q-Q Plot Gumbel

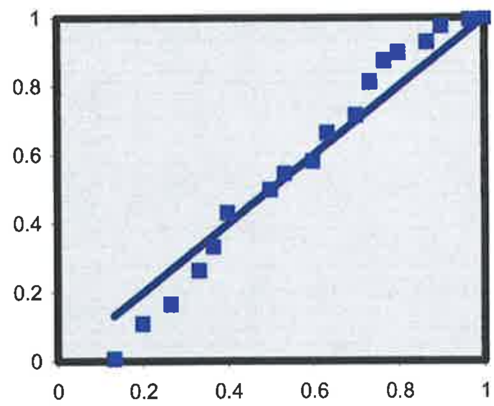


Fig.8. Q-Q Plot Regression

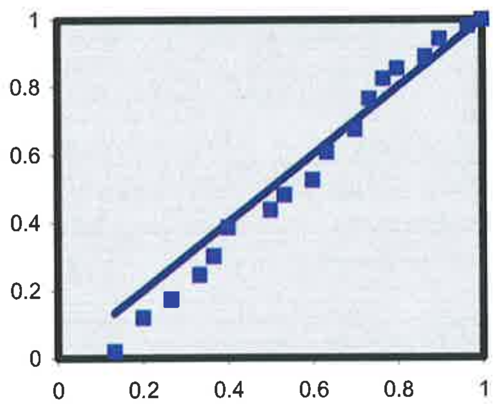


Fig.9. Q-Q Plot Envelope

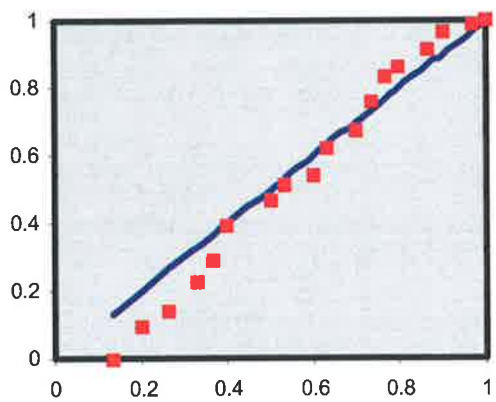


Fig. 10. Q-Q Plot Iman-Conover

From Q-Q plots of the Iman-Conover, envelope and linear regression methods, it was very difficult to see a significant difference between the three models. The quantitative approach confirmed the graphical interpretation. Quantitative analysis showed that the Gumbel copula, with a Minimum Distance (MD) = 0.0417, was the best dependence method that reproduces the original data among Archimedean copulas where the Frank copula had an MD = 0.098 and the Clayton copula with an MD = 0.17. Second is the Envelope MD = 0.064. Third is the Iman-Conover method with MD = 0.08 and finally, Regression fitting with MD = 0.081. This confirmed that the Gumbel copula fitted the original data best. This was not surprising since the Gumbel copula is particularly suited to capture upper tail correlation patterns. None of the more commonly used methods, the Iman-Conover, Envelope and linear regression methods, were able to capture the upper tail structure.

Clearly, the Envelope method is very flexible and powerful. With the appropriate, subjective choice of bounding lines and the distribution to use to sample between these lines, it can be tailored to capture almost any structure. In our approach, two lines in the initial model were used. This was not enough to capture the shape of the dependence structure and, consequently, more lines were added. We found that this process yielded better results and the MD was reduced from 0.117 to 0.064. This multiple lines envelope method was used for the comparison with other dependency methods (Fig. 11). Furthermore, the process of constructing two lines only makes sense if the joint distribution is normal, or in the general case if the shape of the dependence structure is an ellipsoid. However, with a dependence structure that has a lower or an upper tail pattern we found out that constructing more lines yield better results. The challenge, of course, is to subjectively pick the most appropriate bounding lines for the various points in the correlation structure.

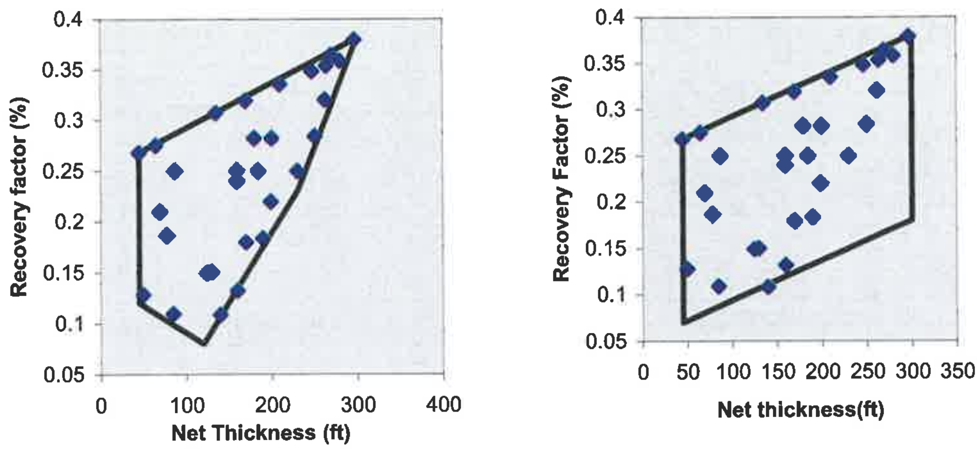


Fig. 11. Envelope method: on the left with multiple lines
and on the right with two lines

Using the standard approach with two lines to capture the dependency structure led to an underestimation of the mean by 13% and an underestimation of standard deviation by 17% compared with the use of multiple lines (Fig. 12).

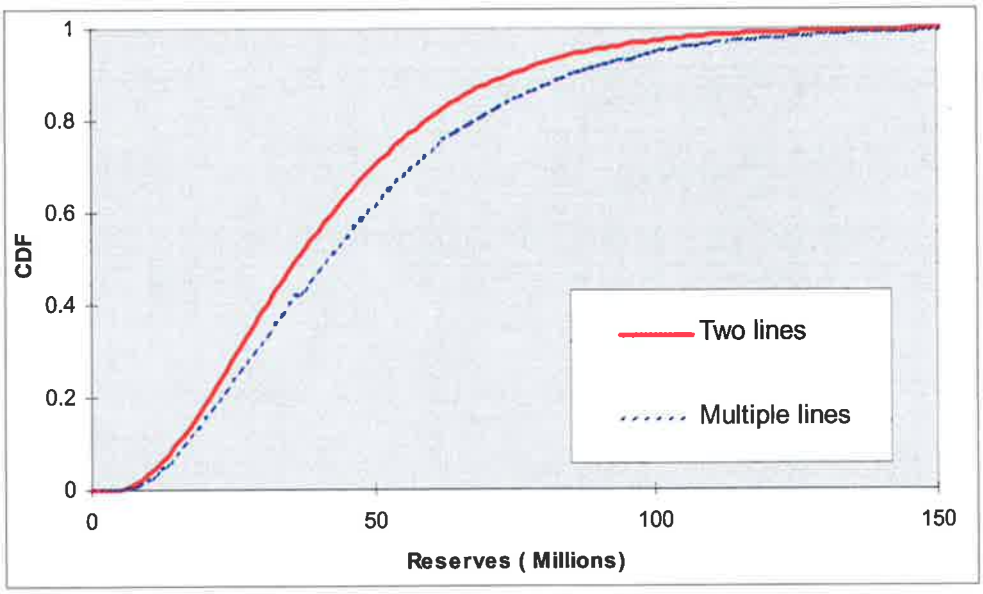


Fig. 12. Impact of envelope method with multiple lines

6. Discussion

We have investigated how well various correlation models represent the original distributions of the input variables. The Iman-Conover and Copulas methods retain the original distribution of the input variable. For example, recovery factor had the best fit with a triangular distribution in the original data. The Iman-Conover and copulas retain the same triangular distribution as the original while the regression fitting model simulates the recovery factor as a normal distribution. This illustrates that irrespective of what the input distribution is, using the linear regression method, the simulation will always result in a normal distribution. This should be expected since the design of the equation has a normal assumption embedded in it as the standard error is normally distributed.

The envelope method tends to retain the original distribution only in a weak form. If multiple bounding lines are not used, the envelope method tends to produce other distributions such as normal distribution.

Compared with the copula approach, the Iman-Conover, envelope and linear regression models all fail to capture upper tail dependence pattern.

We have now concluded that the Gumbel copula fitted our original data best. Although this, in itself, is interesting, the key question is whether the choice of correlation model will influence the decision metrics (Technical reserves). Let us define the percentile of data from 0 - 0.2 as a *lower tail* and the percentile of data from 0.8 – 1 as an *upper tail*. Note that here we use the ascending cumulative distribution function to define the percentiles.

Looking at Fig. 13, we see two main results. First, the Gumbel and Iman –Conover seem to have same results and the Envelope and Regression fitting also seem to have identical results. Both Gumbel and Iman- Conover results yield higher reserves than the Envelope and Regression fitting in the range of 10- 15 million barrels. Second, if we zoom on the upper portion, the Gumbel copula gives the highest reserves in the upper tail (Fig. 14). This is because it captures the correlation between thickness and recovery factor in the high percentile data. For example, at the 90th percentile, the difference between the Gumbel copula and both the envelope method and the linear regression method is approximately 15 million barrels. This is a significant difference amounting to 16% of the total estimates, which shows the superiority of the copulas approach. However when compared with the Iman-Conover the difference is only 1- 2 million barrel. This is a very small difference suggesting that the impact of capturing the dependence structure is really not significant when compared with the Iman-Conover method.

This part concludes that even though the copulas approach captured the upper tail dependence structure and the Iman- Conover did not, the impact of the upper tail dependence structure of copulas when compared to Iman- Conover does not really matter. It did not have any significant results on the decision matrices.

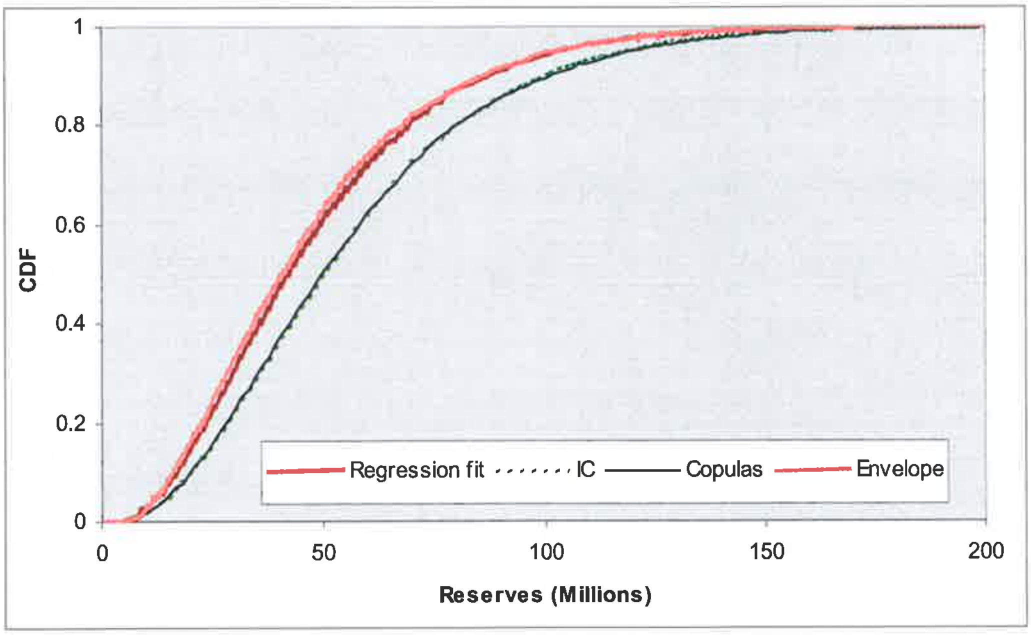


Fig. 13. Technical Reserves for dependency methods

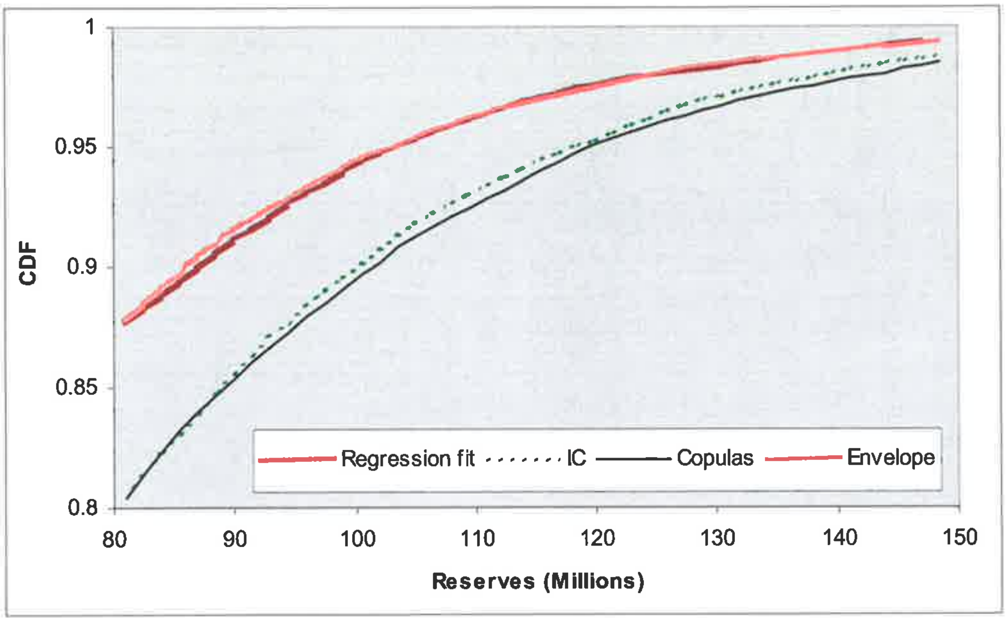


Fig. 14. Technical reserves at the upper tail

The previous analysis explored the upper tail dependence but in order to investigate how the different dependence methods will perform for lower tail dependence we started with a different dependence structure for net thickness and recovery factor (Fig. 15). This dependence structure had the same marginal distributions and the same rank correlation as the previously used correlation (Fig. 3).

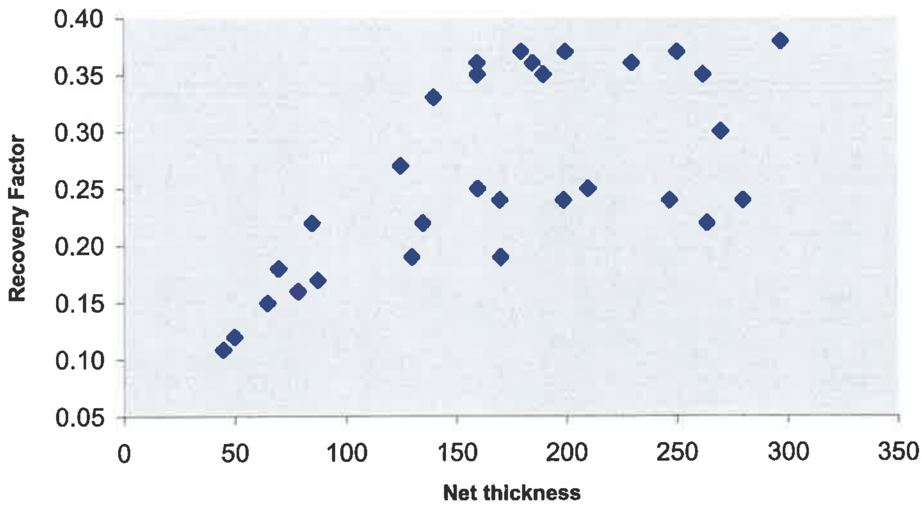


Fig. 15. Positive dependency between net thickness and recovery factor

The graphical (Q-Q plots) approach (Figs. 16-19) indicated that the Clayton copula fitted the original data best, whilst the other methods showed a larger discrepancy, in particular below the 40th percentile. Among the Archimedean copulas the Clayton copulas seems to fitted best with an MD =0.0287 compared to the Gumbel with MD =0.12 and Frank with MD =0.092. The Gumbel copula as expected in this case was the worst because it is designed to capture upper tail dependence not lower tail dependence.

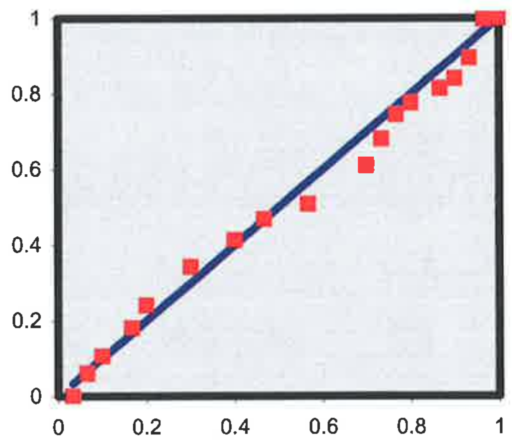


Fig.16. Q-Q Plot Clayton copula

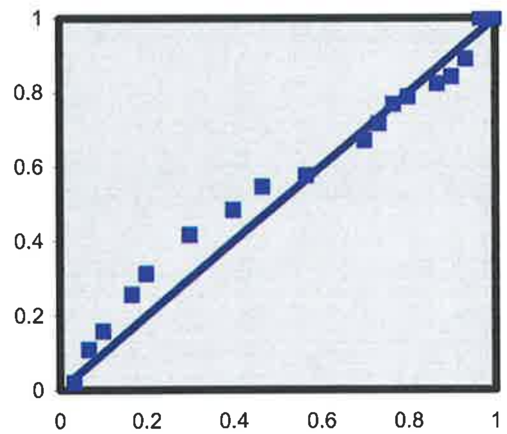


Fig.17. Q-Q Regression fit

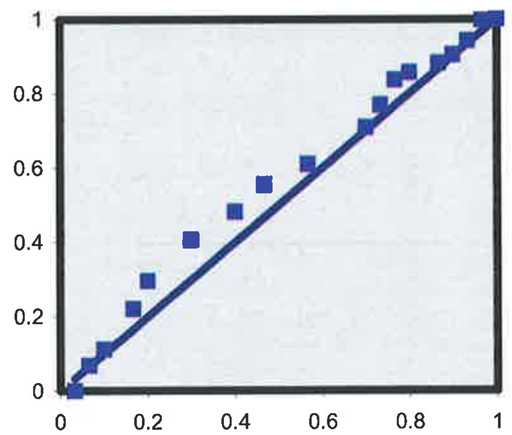


Fig.18. Q-Q Plot Envelope

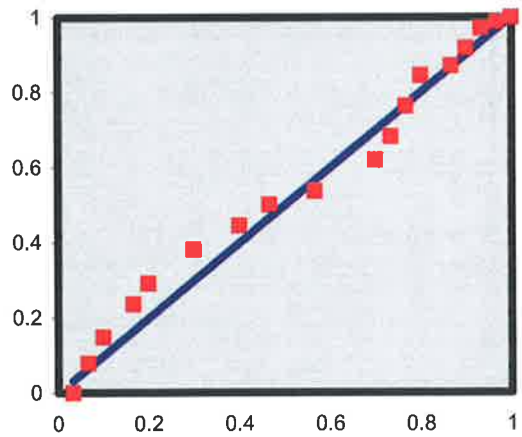


Fig.19. Q-Q Plot Iman-Conover

Among the dependency methods, the quantitative analysis showed that the Clayton copula had an MD = 0.0287, Envelope MD = 0.052, Iman-Conover MD = 0.042, and regression fitting had MD = 0.063. Again, the copula model captured the dependence in the lower tail better than the other models.

The next step was to investigate the impact of the different techniques on the technical reserves. Below the 20th percentile all of the dependency methods seemed to have similar results to the copulas with a range difference of 1-3 million barrel while above the 50th percentile the difference grew to 8 million barrel for the envelope and regression fitting. In this case the Iman- Conover and copulas again seem to be closer to each other than the remaining techniques (Fig. 20). Investigating the lower tail dependence by zooming in at the 5th percentile the copulas approach respond best to the lower tail dependency but, when compared with the Iman- Conover, the difference is only 2 million barrels which is very small (Fig. 21). This showed that the copulas approach is superior for capturing lower tail dependence structure better than the existing methods. Yet the impact that capturing that dependence has is low compared to the Iman- Conover.

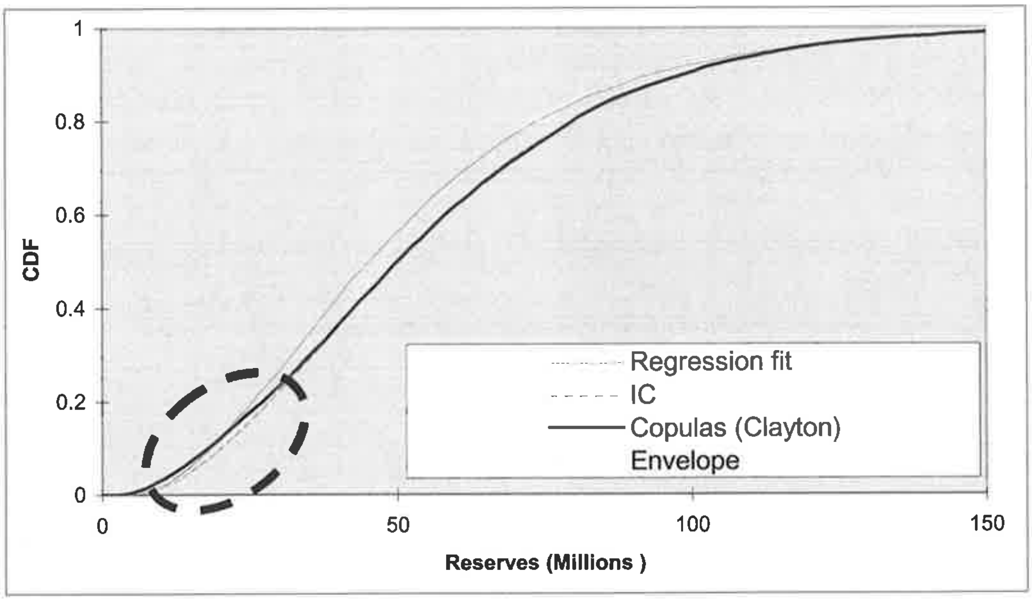


Fig.20. Technical reserves and dependency methods (lower tail)

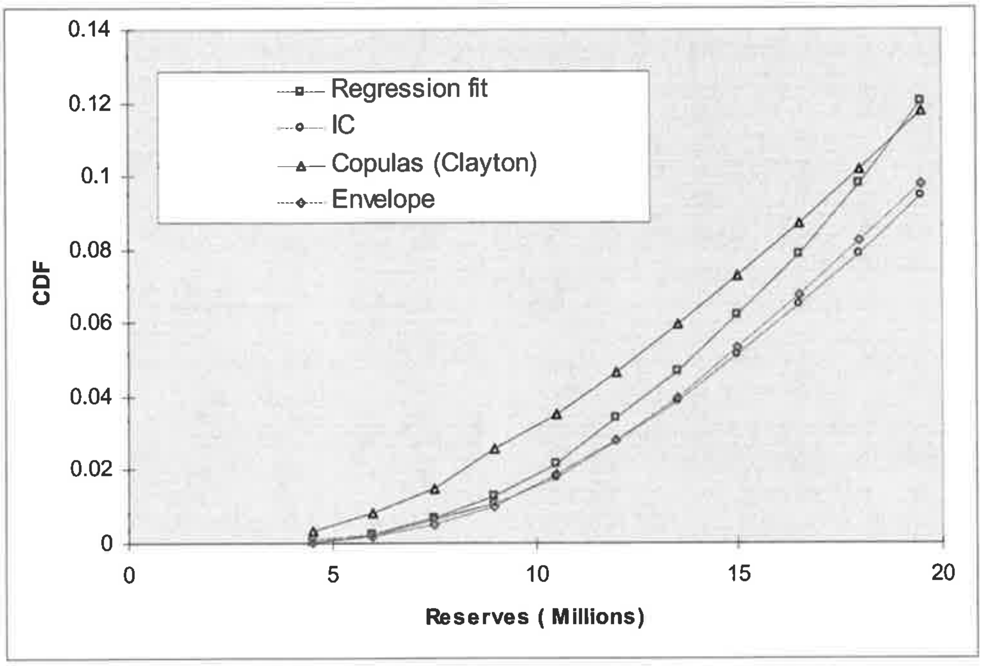


Fig.21. Technical reserves at lower tail dependence

In understanding these results, it is important to realize that Fig. 15 has the same distribution with same mean and standard deviation as Fig. 3. Furthermore, the rank correlation is the same while the dependence structure is different, indicating that, for the data with important upper tail characteristics (Fig. 3) and data with important lower tail characteristics (Fig 15) the Iman-Conover results are the same. The Iman-Conover method does not capture the correlation pattern as the copulas method and does not distinguish between upper and lower tail dependency structures. It uses a joint distribution that does not represent the original distribution for either lower or upper tail dependence structure. The regression fitting, in both cases, assumes the dependence structure is joint normal no matter what the dependence pattern is where as the Envelope can be tailored to capture various dependence structures. However, the subjectivity in drawing the bounding box makes it more cumbersome and less rigorous and appealing than the copula approach. Copulas should be the choice where upper- or lower tail dependency patterns are important. However, in our example, the impact of the dependency structure is small and it is not significant with a difference of 1-3 million barrels between the Iman-Conover and the copulas method.

7. Conclusion

In this paper we have illustrated the potential impact of the choice of dependence models in the probabilistic simulation of reserves and discussed the issue of modelling dependence in the probabilistic reserves estimates. We compared and contrasted the most commonly used approaches for correlation modelling with a more recent approach, the copula model.

Firstly, we have confirmed that dependencies are important and can have a large impact on the metrics used for decision-making, particularly for those variables that impact on the output the most.

Secondly, we have shown that the copulas approach is superior in capturing dependency patterns in the lower- and upper tail.

Thirdly, we have illustrated that the most commonly used correlation models in oil and gas evaluations fail to capture the upper and lower tail dependency patterns. This can lead to significant errors in the simulated results when comparing copulas with regression fitting and the envelope method. However, when compared with the Iman - Conover model, the impact of the dependence structure was not significant, the difference in estimates being less than 3 million barrel. This indicates that, for our example, the Iman – Conover model is doing a good job in capturing the dependency structure.

Based on these results, we believe that the copula approach has several benefits compared with the more commonly used correlation approaches and we encourage further research into and use of these models in oil and gas evaluations. We strongly recommend using copulas as an alternative measure of modelling dependence in economic oil and gas evaluations. The fact that the current software such as @Risk™ and Crystal Ball™ uses Monte Carlo Simulation should further encourage the use of copulas in the oil industry.

Acknowledgement

The Authors would like to thank the following for their help James Murtha, Edson Nakagawa, Ashok Khurana, John Van der Hoek, Steve Mackie, and Matthew Welsh.

References

Accioly, R, Chiyshi F. 2004. Modelling dependence with copulas: a useful tool for field development decision process. Journal of Petroleum Science and Engineering 44, 83 –91.

Armstrong, M, Galli, A Bailey, W, Couet, B. 2004. Incorporating technical uncertainty in real option valuation of oil projects. Journal of Petroleum Science and Engineering 44, 67 –82.

Clayton, DG. 1978. A model for association in bivariate life tables and its applications in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65,141– 151.

Demirmen, F. 2005. Reliability and Uncertainty in Reserves: How the industry Fails, and a vision for improvement. SPE 94680 presented at the SPE Hydrocarbon Economics and Evaluation Symposium held in Dallas, Texas, April 3- 5.

Embrechts, P, McNeil, AJ, Straumann, D. 1999. Correlation: pitfalls and alternatives. *RISK,* May 1999, 69-71.

Frank, MJ. 1979. On the Simultaneous Associativity of F(x, y) and x + y-F(x, y). Aequationes Mathematicae 19, 194– 226.

Frees, W, Valdez, E, 1998. Understanding Relationships Using Copulas. North American Actuarial Journal, 2, 1, 1 –23.

Genest, C, MacKay, J. 1986. The joy of copulas: bivariate distributions with uniform marginals. The American Statistician 40, 280– 283.

Genest, C, Rivest, LP. 1993. Statistical inference procedures for bivariate Archimedean copulas. Journal of the American Statistical Association 88 (423), 1034–1043.

Gumbel, E.J. 1960. Bivariate exponential distributions. Journal of the American Statistical Association, 55, 698 –707.

Iman, R, Conover, W. 1982. A distribution-free approach to inducing rank order correlation among input variables. Communication in statistics, simulation and computation, 11, 311-334.

Mari, D, Kotz, S. 2001. Correlation and Dependence. Imperial College Press, London, U.K., 220

Merrow, E. 2003. Taking on a Cult of Mediocrity. Journal of Upstream, May 23. http://www.upstreamonline.com.

Mian, M. 2002. Project Economics and Decision Analysis, volume two: Probabilistic Models. PennWell Publications , Oklahoma, USA.

Murtha, J. 2000. Decisions involving Uncertainty, An @risk tutorial for the petroleum industry. Palisade Corporation.

Nelsen, RB. 1999. An introduction to copulas. Springer Verlag, New York.

Newendorp, P. 1976. A method for treating Dependencies Between Variables in Simulation Risk Analysis Models. SPE 5581, SPE-AIME, October.

Rose, P. 2001. Risk Analysis and Management of Petroleum Exploration Ventures. AAPG. Tulsa, Oklahoma, USA.

Al-Harthy, M. Khurana, A. (2006) Portfolio Optimization: Systems vs. Sequential Approach/ Society of Petroleum Engineers. SPE 101152 Asia Pacific Oil & Gas Conference and Exhibition to be held in Adelaide, Australia, *11–13 September, 2006*.