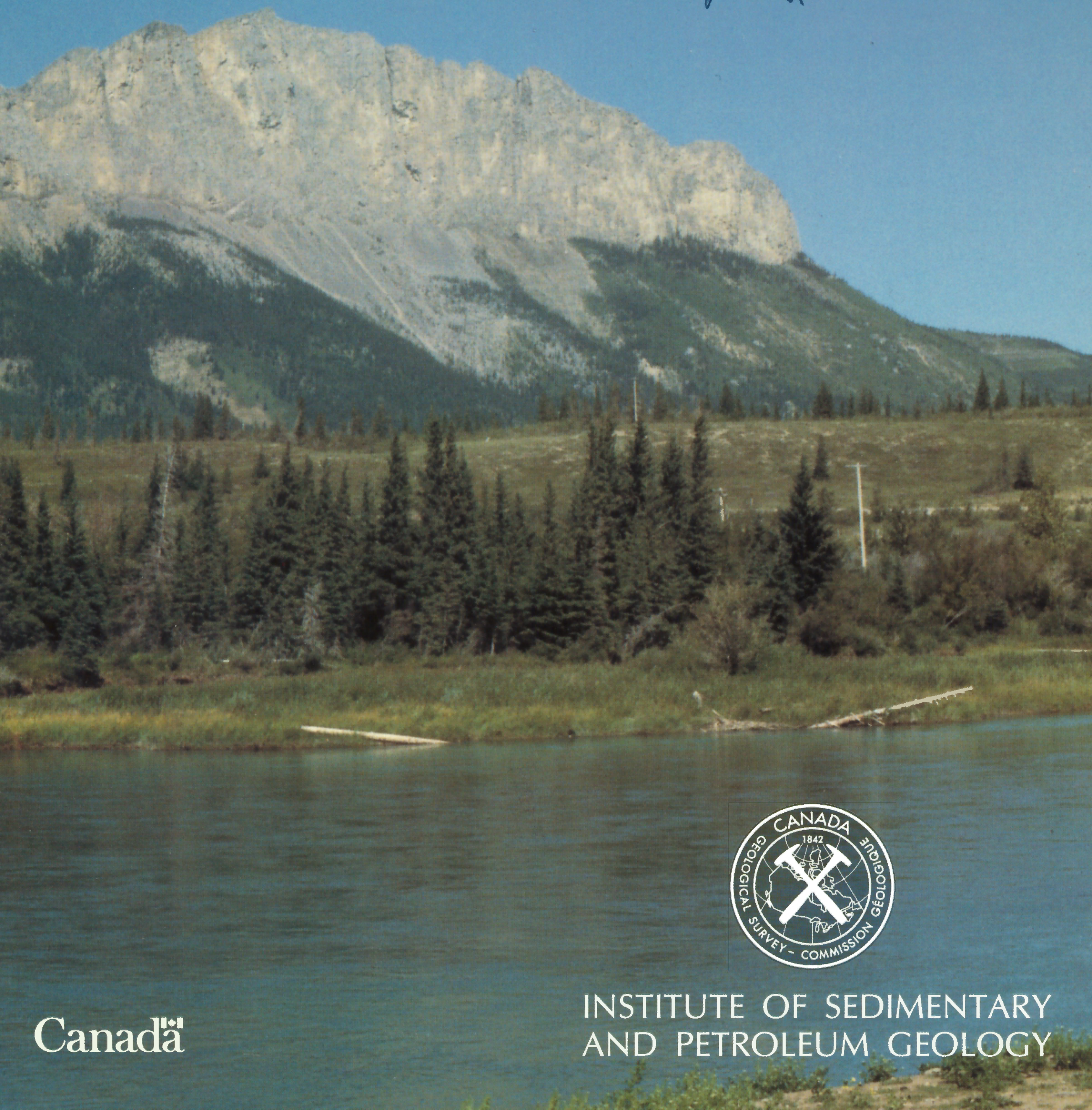
Open File 2703

THE PETROLEUM EXPLORATION AND RESOURCE EVALUATION SYSTEM (PETRIMES) AN INTRODUCTION TO PETROLEUM RESOURCE EVALUATION METHODS

P.J. Lee and P.C.C. Wang

pt 2



CANADA 1842 GEOLOGICAL SURVEY - COMMISSION GÉOLOGIQUE

INSTITUTE OF SEDIMENTARY AND PETROLEUM GEOLOGY

Canada

GEOLOGICAL SURVEY OF CANADA
COMMISSION GEOLOGIQUE DU CANADA

Open File 2703

THE PETROLEUM EXPLORATION AND RESOURCE EVALUATION SYSTEM (PETRIMES)

AN INTRODUCTION TO PETROLEUM RESOURCE EVALUATION METHODS

P.J. Lee[1]
and
P.C.C. Wang[2]

JULY 1993

1 Institute of Sedimentary and Petroleum Geology
Geological Survey of Canada
3303 33 Street N.W., Calgary, Alberta T2L 2A7

2 Keema Corporation
33 Columbia Street West, Waterloo, Ontario N2L 3K4

The authors would like to thank our colleagues Frits Agterberg, Peter Hay, Kirk Osadetz and Lloyd Snowdon for their suggestions and comments.

Geological Survey of Canada contribution number 51789.

First Edition 1984
Second Edition April 1985
Third Edition January 1990

CONTENTS

Chapter 1 Introduction 1

Background 1
Objectives 2
An outline of the evaluation procedure 2
The scope of this book 2

Chapter 2 Evaluation Models 3

Geological models and play definitions 3
Statistical models....................................5
Concepts employed by PETRIMES 6

Chapter 3 Estimating Mature Plays 9

The Beaverhill Lake play 9
Geological setting 9
The discovery sequence 10

The nature of geological populations 10
Outlier proneness 10
Correlation between variables 11
Mixed populations 11
Lognormal distributions....................................11

Lognormal discovery process model 12
Estimation procedure....................................12
Validations of the discovery process model 14
The Beaverhill Lake play 15

Nonparametric discovery process model 16
Estimating distributions 16
Modeling distributions 16
Outliers 17
Long or short tails at both ends 17
Symmetry 18
Plateaus 18
Validations using simulated data sets....................................18
The Beaverhill Lake play 18

Chapter 4 Pool-Size-By-Rank 21

Interpreting pool-size-by-rank 21
Matching process......................................24
An example of matching 26
Pool sizes conditional to pool ranks 27
Play resource distribution 28
Play potential distribution 28

Chapter 5 Evaluating Conceptual Plays 31

Geochemical approaches 31
Pool size equations 32
Construction of probability distributions 34
Geological risk factors 35
Exploration risk 35
Methods for estimating marginal probability 35
Play level risk 36
Prospect level risk 37
More examples of marginal geological risk 38
Dependence of prospect level risks 38

Conceptual plays 39
Pool size equations 39
Exploration risk 40
Number-of-prospects distribution 40
Number-of-pools distribution 40
Play resource distribution 43
Pool-size-by-rank 44
Generation of reservoir parameters 45

Chapter 6 Procedures and Feedback Mechanisms 47

Procedures 47

Feedback mechanisms 49
Can we predict the present situation?...........49
Has the largest pool been discovered ? 50
Pool size conditional to play resource 50

Chapter 7 Concluding Remarks..............................51

References 52

Chapter One

INTRODUCTION

> In order to reach the Truth, it is necessary, once in one's life, to put everything in doubt - so far as possible.
>
> Descartes

BACKGROUND

Petroleum resource evaluations have been performed by geologists, geochemists and statisticians for many decades, in an attempt to estimate resources or potentials that may exist in a given region. Resource evaluations often vary as a result of differences in geological and statistical methods used. Various methods have been compiled by Haun (1975), Grenon (1979), Drew et al (1980, 1982), Schuenemeyer and Drew (1983), Masters (1985), and Rice (1986).

Information required in petroleum resource evaluation generally includes all available reservoir data and data compiled from exploratory and development wells. Other essential geological information comes from regional geological, geophysical, and geochemical studies, as well as from work carried out in analogous basins. To summarize, any comprehensive resource evaluation procedure must combine raw data with information acquired from regional analysis and comparative studies.

In order to achieve accountability, the Hydrocarbon Assessment System Processor, HASP (Energy, Mines, and Resources Canada, 1977; Roy, 1979), blended available exploration data with information previously gathered. Combinations of exploration data and expert judgment were expressed as probability distributions for specific population attributes (pool area, net pay, porosity, etc). Since this procedure was first implemented, demands on the evaluation capability have steadily increased, and the evaluation results have been increasingly applied to economic analysis. Traditional procedures can no longer be adapted to handle the new challenges. Therefore, a probabilistic formulation for HASP was established and new capabilities and features have been added (Lee and Wang, 1983a, 1983b, 1984, 1985, 1986, 1988; Lee, Nair, and Wang, manuscript in preparation). This formulation led to the development of the Petroleum Exploration and Resource Evaluation System, PETRIMES, which is the subject of this book.

OBJECTIVES

The objective of an assessment is to evaluate total resources (the term "resource" is defined as that quantity of hydrocarbons made up of discovered and undiscovered pools or fields), or potentials (the term "potential" is defined as an undiscovered quantity of hydrocarbons). Results of petroleum resource evaluations, however, are usually given as aggregated numbers representing total resources. Aggregated potential values are not specific enough to be used in economic, exploration, or development planning analyses, because all of these processes require a knowledge of the number and size of pools. Consequently, the objectives of a resource assessment are as follows: (1) to estimate the number of yet-to-be discovered pools; (2) to estimate the sizes of the undiscovered pools; (3) to estimate the reservoir characteristics of the undiscovered pools; (4) to provide adequate information for economic analysis; (5) to validate exploration concepts against known information; and (6) to estimate pool size distributions and relate these distributions to geological models.

AN OUTLINE OF THE EVALUATION PROCEDURE

In this book, the procedure for resource evaluation may be outlined as follows:

1. A pool size distribution is estimated using either (1) the exploration-discovery process model for mature plays or (2) multiplication of probability distributions of geological variables according to a pool size equation for conceptual or immature plays.

2. Geological risk factors of a play are identified and their marginal probabilities are estimated.

3. A number-of-pools distribution is derived from the operation of exploration risk on the number-of-prospects distribution.

4. Individual pool sizes of a play can be estimated from the number-of-pools distribution and the pool size distribution.

5. A play resource and/or potential distribution can then be obtained.

THE SCOPE OF THIS BOOK

Chapter 2 of this book explains the meaning and usefulness of geological and statistical models in determining petroleum resource evaluations. In Chapters 3 and 4, the superpopulation model and data from the Beaverhill Lake play are used to illustrate the procedure in an example for which a discovery record is available. In Chapter 5, a frontier play is used to illustrate a conceptual play evaluation. In Chapter 6, the information required to undertake an assessment, the interaction between the assessors and the system, and the mechanisms for the feedback processes during the assessment are given.

Chapter Two

EVALUATION MODELS

How quaint the ways of paradox-
At common sense she gaily mocks.

W.S. Gilbert

GEOLOGICAL MODELS AND PLAY DEFINITIONS

The initial step in the evaluation of any petroleum resource is the identification of an appropriate geological population or model, which can be delineated through subsurface study or basin analysis. A geological model represents a natural population and possesses a group of pools and/or prospects sharing common petroleum habitats. A natural population can be a single sedimentation model, a single structural style, a single type of trapping mechanism or geometry, a statistical population, or any combination of these criteria. Reasons for adopting these criteria in the definition of a model are: (1) geologists can use the data from comparative geological studies as well as that gathered from prior experience; (2) the natural population normally exhibits a probability distribution such as lognormal distribution; (3) statistical concepts, such as the superpopulation concept, can be applied to geological models so that, for specific plays, an estimate of undiscovered pool sizes can be made; and (4) evaluations can be made of mature plays that have hundreds of pools. Data analysis may be required to assist in the differentiation of pools in homogeneous populations for the purpose of generating a statistical population. In this case, the geological meaning of the statistical population may not be clear, but the estimates may be acceptable. Furthermore, it is important to remember that a geological model is merely a working hypothesis that should be revised or redefined as new information becomes available.

Figure 2-1 illustrates a variety of possible sedimentary environments (tidal flats, lagoons, beaches, and patch reefs) that can be used as geological models in resource evaluation. Each of these models has its own distinguishing characteristics of source, reservoir, trapping mechanisms, thermal history of source beds, and migration pathways. To assure the integrity of statistical analysis, each of them should be treated as a separate, natural population in resource evaluation. The logical steps for the description of plays are, therefore: (1) the identification of a single sedimentation model and (2) examination of subsequent geological processes, such as faulting, erosion, folding, diagenesis, biodegradation, thermal history of source rocks, and migration history that may provide a basis for further subdivisions of the model. In some cases, two or more populations may be mistakenly considered as one because of a lack of understanding of the subsurface geology. The resulting mixed population may possess two or more modes in its distribution and this may have a significant impact on the resource evaluation results. An example of a mixed population was described by Lee, Eggen and Vann (1988).

For our first example, let us look at the Leduc reef trend from the Western Canada Basin. Figure 2-2 (left) displays the trend of the Devonian Leduc reefs. The setting includes persistent shelf, slope, and basinal facies belts. Reefs are regions of persistent high energy stromatoporoid rudstone, rooted to the carbonate platform of the underlying regressive hemicycle persistent through the succeeding transgressive hemicycle. Traps on the carbonate shelf are controlled by transgressive-regressive hemicycles of a different order than carbonate build-ups in the persistent basinal facies belt. The traps along the Middle Devonian shelf exhibit a negative correlation between net pay and pool area, whereas the traps along the Rimbey-Meadowbrook chain exhibit a positive correlation. For purposes of petroleum evaluation, the three settings should be separated into three models.

Our second example involves the Slave Point-Elk Point succession. In the northeastern part of British Columbia (Fig. 2-2, right), the Middle Devonian Slave Point and Elk Point succession consists of two predominantly transgressive hemicycles separated by the Watt Mountain regression. A persistent Keg River-Sulphur Point-Slave Point carbonate barrier separates the persistent evaporitic platform to the south and east from the persistent Horn River basin to the north and west (Griffin, 1965a, 1965b Williams, 1984).

The lateral facies transition between persistent carbonate and shale facies belts, referred to as the facies front, generally occupies a zone of several kilometres in width, and extends over a maximum stratigraphic interval of about 430 m. At the front of the shelf, prolific organic growth occurred, in places resulting in the formation of reef structures.

Barrier reefs form reservoirs along the rim of the platform, whereas pinnacle-reefs constitute the reservoirs in the basin adjacent to the shelf. The Slave Point, and probably the Sulphur Point and the Pine Point formations, are dolomitized and diagenetically altered enhancing reservoir development.

In the Slave Point Formation, there are at least two types of reef populations (i.e., reef rim and inner shelf). Thus, the distributions of areal extent and the net pay of these populations may be quite different. The effects of the geology on the accumulation of hydrocarbons may also differ. Consequently, the Slave Point Formation in northeastern British Columbia is divided into two plays with respect to natural gas resource evaluation: i.e., shelf edge and inner shelf.

The important point to emphasize here is that the first step in any resource evaluation is to properly identify geological models that will serve as the framework upon which to formulate a statistical evaluation.

STATISTICAL MODELS

Geological variables include net pay, porosity and others. These variables can be quantified with single values or a range of values. Let us take, for example, the porosity values from a sandstone formation. What is immediately obvious is that some of the values occur more frequently than others. The basic idea, therefore, is to associate each porosity value with a real number, or the likelihood of occurrence of that value (the likelihood that the value may occur: a large number for a likely outcome and a small number for an unlikely one). In other words, all the porosity values of a formation will be associated with a probability that describes their likelihood of occurrence. All these values and their probabilities form a probability distribution (see Fig. 2-4).

We know the probability associated with each value, but we may not be able to explain the process that leads to this distribution. This class of physical phenomenon (so-called random phenomenon), behaves randomly according to a probability distribution. Therefore, if we sample a specimen from a given formation and we wish to predict its value, we need to know the probability distribution of that variable.

One of the steps in resource evaluation is to estimate the probability distributions of geological variables. There are two types of distributions, i.e. discrete and continuous. Let us take, for example, a finite number of pools in a play. Certainly all pools constitute a finite population and will exhibit a discrete distribution (Fig. 2-3, left). On the other hand, pool values can be thought of as coming from an infinite population having a continuous probability distribution. This continuous probability distribution is called a superpopulation or a parent population (Fig. 2-3, right).

In cases where we have a random or a very large sample set, normal statistics can be used to construct a probability distribution. For example, 406 porosity values have been obtained from the Lower Mannville Formation of the Western Canada Basin. This sample set can be used to construct a histogram (Fig. 2-4, top), a cumulative less-than (Fig. 2-4, lower left) or a cumulative greater-than distribution (Fig. 2-4, lower right). These types of continuous distributions are considered to be superpopulations. The greater-than form is generally used to express probability distributions in petroleum resource evaluations. In reality, the samples of certain variables resulting from exploration are neither random, nor large enough to represent the population. Therefore, specifics of the exploration-discovery process are required if we are to estimate the mean and variance of the population.

Petroleum resource estimation procedures make use of the following statistical models:

1. The superpopulation or finite population model. These models are needed to predict individual pool sizes in a population and to measure prediction uncertainties.

2. The exploration-discovery (or discovery) process model. This model can be used to estimate the mean and variance of the population using data resulting

from a discovery process.

3. The lognormal distribution model. If a prior distribution such as a lognormal distribution is specified, then only the mean and variance of a population are required for the distribution to be estimated; the values for each percentile can be generated according to the lognormal distribution. On the other hand, if no prior distribution (nonparametric) is specified, then the values for each percentile must be estimated from the data.

CONCEPTS EMPLOYED BY PETRIMES

Basic concepts employed by PETRIMES are illustrated in Figures 2-5 and 2-6. The upper righthand corner of Figure 2-5 displays the facies distribution of a play containing pools and yet-to-be tested prospects. The discoveries from the play were plotted in terms of the discovery sequence (lower lefthand corner). Some questions that arise from examining the discovery sequence are: "How can these data be used to estimate the sizes of the undiscovered pools in this play? Can conventional statistical methods be used to predict undiscovered resources?" <u>If we adopt the usual method of computing the sample mean and variance for the population, the assumption is either that this is a random sample set from the population, or that it is large enough to represent the population. In fact, neither of these assumptions are valid</u>. In the discovery process, large pools are normally discovered at an early stage. This implies that smaller pools remain to be discovered. Thus, the population mean would be overestimated by the sample mean obtained here, whereas the population variance would be underestimated by the sample variance. Therefore, we believe that the discovery process may be viewed as a sampling process whereby pool discovery probability is proportional to pool size and sampling without replacement.

Consider the patch reef model as an example of how statistical methods may be applied to evaluate a reef play. Firstly, a reef model (Fig. 2-6 top) is defined as a collection of analogous reef pools, and a reef play (upper righthand corner of Fig. 2-6) contains some members of the model. In other words, a reef play possesses a finite number of reef pools, whereas a reef model contains an infinite number of reef pools. Secondly, a reef model can be described in terms of its geological variables such as pool size, pool area, net pay, porosity, and number of pools. The range of all possible values for each variable exhibits a continuous probability distribution because of the infinite number of reef pools, except that the number of pools has a discrete distribution expressed as an integer (upper lefthand corner). Thirdly, for a specific play, the values of a variable are considered to be taken as a random sample from its probability distribution, i.e. they are independently derived from a common (or identical) distribution. The last two statistical assumptions may be verified as follows: (1) a play is defined as a single and natural population; (2) the formation of a trap (before occurrence of hydrocarbon migration) has no influence on the formation of other traps; (3) if excess oil has been generated for all traps, then the formation of a pool (after occurrence of hydrocarbon migration) has no influence on the formation of the other pools in the play. The last statement can be verified using the geochemical material balance method. Fourthly, pool sizes obtained from discoveries of a play (lower righthand corner of Fig. 2-6) can be used as a sample to estimate the population distributions (continuous pool size distribution and the discrete number-of-pools distribution).

A play may contain many, few, or no discoveries at the time of evaluation. A play lacking discoveries (a conceptual play), or one containing few discoveries, is analyzed using the pool size equation (see Chapter 5). If a play has sufficient discoveries (such as those shown in the lower righthand corner of Fig. 2-6), there are two statistical approaches that may be applied to estimate the sizes of the remaining undiscovered pools.

The first approach, called the superpopulation approach or concept (Cochran, 1939; Cassel, Särndal and Wretman, 1977, Baecher, 1979), is used to estimate the continuous pool size distribution and the discrete number-of-pools distribution. The superpopulation approach views a play (the finite population) as one of the possible cases from the geological model (the infinite population or superpopulation), and has been described by Kaufman, Balcer, and Kruyt (1975). The second approach is to estimate the play (upper righthand corner of Fig. 2-6) without using the superpopulation concept. The play has a finite number of pools and possesses a discrete pool size distribution. This approach is called the finite population approach (Kaufman, 1986). In this book, the superpopulation approach is discussed.

Once the superpopulation pool size distribution and the number-of-pools distribution have been estimated, the individual pool sizes of the play can then be estimated, as is shown in the lower lefthand corner of Figure 2-6. The boxes that express the estimation intervals may be matched with the present discoveries (shown in the lower righthand corner). This matching process is one of several kinds of feedback mechanisms provided by PETRIMES that allow geological interpretations to be combined with statistical analysis.

Chapter Three

ESTIMATING MATURE PLAYS

> A discovery process model is one built from assumptions that directly describe both physical features of the deposition of individual pools and fields and the fashion in which they are discovered.
>
> Gordon M. Kaufman

Geological models possess continuous population pool size distributions which are to be estimated from samples. Consequently, it is important to understand the nature of geological populations in order to choose appropriate probability distributions for them. In geological populations, important properties such as outlier proneness and correlation of variables may be observed. In this chapter, the Beaverhill Lake play from the Western Canada Basin is used as an example to illustrate statistical concepts used to evaluate mature plays. The first section provides background information on the Beaverhill Lake play. The second section describes the nature of geological populations. The third and fourth sections explain the lognormal and nonparametric discovery process models.

THE BEAVERHILL LAKE PLAY

Geological Setting

Let us take the Late Devonian Beaverhill Lake play as an example of a mature play. The Late Devonian Beaverhill Lake transgression began with the deposition of the Slave Point carbonate on a broad shelf in northeastern British Columbia, northern Alberta, and the adjacent part of the Northwest Territories. A carbonate reef-front facies, similar to the underlying Elk Point reef carbonate, developed in British Columbia.

Continued transgression terminates the Slave Point carbonate platform which is succeeded by basinal lithofacies of the overlying Waterways Formation in northern Alberta and in the Edmonton area. However, in the Swan Hills region of north-central Alberta, a shallow-water platform, protected to the north by the emergent Peace River Arch and flanked to the southwest by the Western Alberta Ridge, provided a setting conducive to bank development and subsequent reef growth. Emergence of the reefs, followed by rising water levels during Beaverhill Lake deposition, terminated growth of some Swan Hills reefs (Hemphill, Smith and Szabo, 1968).

Subsurface study has revealed a sedimentation model in which the Slave

Point carbonate platform possed laterally into open marine mudstone environment. Most of the pools discovered are situated along the platform margin or are adjacent to the platform (Fig. 3-1). Thus, the play contains traps that are related to organic buildups within the Beaverhill Lake carbonates of the Slave Point platform, and to deeper water equivalent sediments of the platform margin.

The area studied extends over 40,000 square kilometres. There have been over 13,000 wells drilled in this area, but only about 1450 wildcats have penetrated the Beaverhill Lake Group. Since 1956, 32 oil pools as well as several gas pools have been discovered. Over 946 10^6 m^3 (6 billions of barrels) of in-place oil has been discovered.

The Discovery Sequence

In addition to the 32 oil pools, there are 55 exploratory wells which have recovered oil in their drill-stem tests. Some of the 55 wells have been producing since their tests. It is assumed that these 55 wells are capable of producing for about 200 hours at the same rates of the drill-stem tests. Therefore, their reserves were then converted into in-place volume using an average recovery factor of 0.10. These 55 "pools" were combined with the 32 discovered pools to form the discovery sequence in Figure 3-2 used in the resource assessment. The upper half of Figure 3-2 displays the discovery sequence of all defined pools. The gaps on the horizontal axis are failed exploratory wells. The lower half displays results from drill-stem tests.

The reason for combining undefined "pools" with defined pools in an integrated discovery sequence is to obtain representation from the small pools. Additional statistical assumptions are not required in this approach; however, the estimation of reserves from drill-stem test results is time consuming and requires expertise in reservoir engineering.

THE NATURE OF GEOLOGICAL POPULATIONS

Outlier Proneness

An outlier is defined as a member of a population with either a relatively small or a large value in relation to other members of the same population. Outlier characteristics have been described by Neyman and Scott (1971) as follows: if a population distribution has a long tail for the relatively large values (i.e., a large variance), then there is a higher probability of there being one or more outliers in it. Both large and small outliers are observable in many geological populations, but only large outliers are discussed here.

Outliers may be recognized by plotting a variable on a boxplot with a logarithmic scale. Boxplots display where the median of a sample lies, and just how the outliers relate to the median (Velleman and Hoaglin, 1981). For example, Figure 3-3 displays the boxplots for the in-place pool size of several plays in the Western Canada Basin. In the boxplot, the box covers the middle (50%) of the data; the horizontal bar within the box indicates the median of the sample; the vertical short bars beyond the box cover the range occupied by three quarters of the data; and the small squares and crosses outside the box indicate relatively large values. The largest or largest two values in each sample are classed as outliers. The magnitude of an outlier is relative to the values of the sample.

Correlation Between Variables

Correlation between geological variables (such as pool area, net pay, recovery factor, reservoir pressure, and others) is also a common feature of geological populations. For example, the pool area and net pay variables of the Zama reef play of the Western Canada Basin exhibit a negative log-linear association (left of Fig. 3-4). That is, as the values of the pool area increase, the values of net pay decrease. In contrast, the pool area and net pay of the Beaverhill Lake play show a positive log-linear association (right of Fig. 3-4). That is, as pool area value increases, net pay value increases. Correlation between variables is an important element to consider in resource evaluation, because if a correlation is ignored, then the mean of a pool size distribution may be over- or under-estimated (Refer to Chapter 5).

Mixed Populations

Figure 3-5 (left) is a log probability plot of all Keg River reefs known at present from the Black Creek basin of the Western Canada Basin. The lack of linearity in the plot may be indicative of any or all of the following: (1) the data set that was chosen is not from a lognormal population; (2) the data set was not chosen randomly; (3) there is more than one population in the data set; and (4) the sample size is too small. The plot shown in Figure 3-5 (right) displays the reefs from the Rainbow basin, a sub-basin within the Black Creek basin. The majority of the data in Figure 3-5 (right) do follow a straight line, but the plot tends to be slightly convex upward. This convex-upward phenomenon may result from biased sampling during the discovery process, i.e. large pools have higher probabilities of being discovered. Therefore, the nonlinearity in Figure 3-5 (left) is indicative of a mixed population.

Lognormal Distributions

To this point, we have examined the nature of geological populations, possible correlation of variables, and outlier proneness. We shall now choose a probability distribution to represent populations that can handle these properties. In this chapter, a family of lognormal distributions is used to describe the lognormal discovery process model, and the nonparametric discovery model is also applied to illustrate an alternate method. Examples of lognormal distribution shapes are presented in Figure 3-6. In this book, μ is the mean of logarithmic transformed data and σ^2 is the variance of the transformed data, n is the number of discoveries, and N is the total number of pools, including discovered and undiscovered pools in the play. Reasons for adopting a lognormal family have been discussed by Kaufman (1965), Meisner and Demirmen (1981), Lee and Wang (1988) and Lee, Nair and Wang (manuscript in preparation) and are listed as follows:

1. A natural geological population can be adequately approximated from a family of lognormal distributions. Examples from the Western Canada Basin can be used to demonstrate that a lognormal distribution is adequate for approxi-

mations of various large sample sets, e.g. net pay (Fig. 3-7, top, n = 109), porosity (Fig. 3-7, lower left, n = 381), and pool area (Fig. 3-7, lower right, n = 211). Figure 3-7, (lower right) displays a peculiar data set, the big steps between 60 and 100 hectares are due to the assignment of 64 hectares to some of the pools. In these cases, a prior distribution may provide a framework for estimating the population distribution.

2. A lognormal family can handle the outlier-proneness of the variables, and yet it is sufficiently flexible to provide a good approximation of variables that may not have this property. The magnitude of variance measures the outlier proneness; the larger σ^2 is, the higher the probability of there being an outlier in the population.

3. If a pool size distribution is computed from the products and divisions of several dependent or independent lognormal distributions, then the end product is a lognormal distribution. Furthermore, from the central limit theorem, the end product also tends to be a normal or lognormal distribution, irrespective of original probability distribution types.

4. Correlation among variables can be conveniently handled with a lognormal distribution. In lognormal cases, the value of σ^2 is expressed by Equation 5-1 (see Chapter 5, p. 32). It is evident that the value of σ^2 can be conveniently handled by a lognormal distribution.

5. Lognormal distributions adequately approximate the distributions recognized by geologists (Lee and Wang, 1983a).

6. In lognormal distributions, there is independence between μ and σ^2. Consequently, this type of distribution can handle all combinations of μ and σ^2.

7. From a study of the nonparametric discovery process model, it appears that the lognormal distribution may be the best choice among many distributions studied. One disadvantage is that a lognormal distribution is unbounded at the upper limit. However, lognormal distributions have finite means and variances.

LOGNORMAL DISCOVERY PROCESS MODEL

Once a probability distribution is chosen, the next step is to estimate the parameters of the underlying superpopulation distribution from samples obtained from exploration. Taking a lognormal as an example, if the parameters, μ and σ^2 have been estimated, all the upper percentiles of the distribution can then be generated.

Estimation Procedure

We shall now discuss the principle of petroleum resource estimation from a statistical point of view. Figure 3-2 shows that pool size gradually decreases with time; however, variations from that trend or "waves" occur over the course of exploration.

In cases where the discovery data for a play comes from a random sample or, alternatively, if all the discoveries have been made, the sample mean and variance adequately represent the population. In reality, however, discovery is influenced by many factors, such as exploration techniques, drilling technology, acreage availability, and company objectives. Furthermore, geologists tend to test what is perceived as the best or largest prospect, which may not, in fact, turn out to be the largest pool of the play. Testing for the best prospect first tends to characterise the discovery process as a sampling procedure. We are then faced with the question of how to use these types of biased samples to estimate the population. For the superpopulation model, a lognormal pool size distribution is assumed as follows:

$$f_{\theta}(y) = 1 / (y \, \sigma \sqrt{2\pi}) \exp [-1/2 \, (\ln y - \mu)^2 / \sigma^2] \quad \text{for } y > 0 \qquad (3\text{-}1)$$

where Θ = (μ, σ^2) is the population parameter to be estimated.

The estimation is based on the principle that the probability of discovering a pool is proportional to its size, and that a pool will not be discovered twice (Kaufman, 1963; Kaufman, Balcer, and Kruyt, 1975; Barouch and Kaufman, 1977). For the sake of simplicity, the concept of the discovery process model may be expressed as follows. The probability for pool j to be discovered is proportional to its size x_j as:

$$p_j \propto \frac{x_j}{x_1 + \ldots\ldots + x_j + \ldots\ldots + x_N} \qquad (3\text{-}2)$$

where N is a total number of pools in the play.

Take N = 3 and n = 2 (number of discoveries) as an example to illustrate the concept of the discovery model. Let the sizes of the three pools be: $x_1 = 50$, $x_2 = 300$, $x_3 = 100$ MM bbl. The probabilities for all possible discovery sequences are graphed in Figure 3-8, which indicates that the most likely sequence is (x_2, x_3, x_1), even though other sequences are also possible. This is the concept adopted by the discovery process model to characterize the sampling process.

In Equation 3-2, the probability is <u>completely</u> proportional to pool size, but in reality pool size may be merely one of many controlling factors. Thus, Equation 3-2 is generalized by the addition of exponent β into the equation as follows:

$$p_j \propto \frac{x_j^{\beta}}{x_1^{\beta} + \ldots + x_j^{\beta} + \ldots\ldots + x_N^{\beta}} \qquad (3\text{-}3)$$

where the β value ranges from negative to 1 or greater than 1. The larger the β value, the higher the exploration efficiency. When $\beta = 0$, then the discovery process can be considered as a random sampling process.

Therefore, the probability of observing ($x_1, \ldots, x_n$), given Y_i, i = 1, .., N, is expressed as

$$p\{x_1,\ldots,x_n \mid Y_1,\ \ldots,\ Y_N\} = \prod_{j=1}^{n} \frac{x_j^{\beta}}{b_j + Y_{n+1}^{\beta} + \ldots + Y_N^{\beta}} \qquad (3\text{-}4)$$

where $b_j = x_j + \ldots + x_n$ (discovered pool sizes), Y's = undiscovered pool sizes.

In general, the probability of the j-th pool to be discovered is the product of the following probabilities: the probability of pool j in the lognormal pool size distribution $f_\Theta(x_j)$; and the probability of pool j to be discovered in a sequence. Thus, the joint density function of all discovered pools can be shown as follows:

$$L(\Theta) = \frac{N!}{(N-n)!} \prod_{j=1}^{n} f_\Theta(X_j)\ E_\Theta\left[\prod_{j=1}^{n} \frac{x_j^{\beta}}{b_j + Y_{n+1}^{\beta} + \ldots + Y_N^{\beta}}\right] \qquad (3\text{-}5)$$

where Θ represents the distribution parameters (μ,σ^2), $N!/(N-n)!$ is the number of ordered samples of size n without replacement from a population of N pools, and $b_j = x_j^{\beta} + \ldots + x_n^{\beta}$, and $y_{n+1}^{\beta}, \ldots, y_N^{\beta}$ = undiscovered pool sizes.

Quantity $L(\Theta)$ indicates the likelihood of a discovery sequence. The maximum likelihood method is used to obtain solutions for μ, σ^2, ß and N such that $L(\Theta)$ is maximized. The resultant $L(\Theta)$ value is the maximized log-likelihood value. The pool size distribution $f_\Theta(y)$, in fact, can be any probability distribution, but the lognormal family is applied here.

Furthermore, Equation 3-5 consists of two parts, f_Θ and $E_\Theta[\cdot]$. f_Θ represents the pool size distribution which results from tectonism, sedimentation, generation, migration and accumulation of hydrocarbons, whereas $E_\Theta[\cdot]$ represents the manner in which pools are discovered (Fig. 3-9).

Validation of the Discovery Process Model

Simulated data sets may be used to validate the discovery process model. The validation procedure is as follows: A hypothetical population is assigned parameters (e.g. $\mu = -4$ and $\sigma^2 = 20$). A random sample of size 200 is then drawn from the population(e.g. N = 200). A discovery process model is simulated (using ß = 0.3, 0.6, and 1.0). For each ß value, various numbers of "pools" are also "discovered" (given in this example values of n = 20, 40, 80, and 150). The model is then used to analyze each of these discovery sets.

It is important to note that the discovery process model does not require prior distributions for μ, σ^2, and β, or truncation of large values. All available data are used to estimate population mean and variance, because an adequate estimate of population variance cannot be derived from truncation of data. Furthermore, the procedure involves estimating the population rather than fitting a curve to the discovery sequence.

For convenience of calculation, let the values of N range from a minimum to a maximum value. They should not be considered as a prior distribution of N. Results are listed in Tables 3-1, 3-2, and 3-3. Each table gives one value of β. The first column gives the values of n used as sample size, whereas the second column shows the different values of N when n equals a specific value. The third column displays the estimated β value. The 4th to 6th columns display the point (middle value) and 95% interval (first and third values) estimates of μ. The 7th to 9th columns display the point (middle) and the 95% interval (first and third values) estimates of σ^2. The last column, or Log L, of the tables gives the values of the maximized log-likelihood values. Each value can be regarded as an index to the likelihood of the particular combination of N and β, given the discovery sequence. The larger the index, the more plausible the combination for the given sequence of discovery pools. The interpretations are summarized as follows: If the log-likelihood values are given to three decimal places, the discovery process model may underestimate the population value of N. However, if the likelihood values are rounded off, then the model predicts an upper limit of N, or a range of values of N, where the n and/or β values are large enough.

Given that N = 200, the adequacy of estimation for the values of β, μ, and σ^2 from the discovery process model can be examined as follows:

1. Estimated β values approximate the population values, especially when the estimated values are rounded off to one digit.

2. Estimated values of μ and σ^2 fall into all 95% interval estimates, regardless of the values of β and n.

3. As the sample size n increases, the estimation interval is reduced.

4. The point estimates approach the population values as the sample sizes increase.

The principle for determining a value of μ and σ^2 from their intervals is that the values of μ^, σ^2^, and N chosen must be able to predict present and future discoveries, which may be validated by prospect analysis. The procedure for choosing a μ^ and σ^2^ is discussed in the section entitled, "Pool-Size-By-Rank".

The Beaverhill Lake Play

Figure 3-2 shows the graph of the pool size versus discovery year for the Beaverhill Lake play. This figure suggests that pool sizes may have influenced the order of discovery (most of the large pools were discovered early in the play's exploration history). From the discovery model, a range of initial estimates of μ and σ^2 can be obtained, as well as estimates for the number of pools in the play. Using the Beaverhill Lake play example, the steps involved in the estimation of μ and σ^2 are described as follows:

1. Geological and geophysical data indicate the number of undiscovered pools ranges from 3 to 400 (including recoveries from drill-stem tests). Therefore, let the number of pools, N, range from 100 to 400 with an increment of 25 (see column 1 of Table 3-4).

2. Let ß be maximized using the model (see column 2 of Table 3-4).

3. For each N and ß, the discovery model is used to estimate μ and σ^2 such that the likelihood function of the discovery model is maximized. The three estimates under μ^ and σ^2^ are interval estimates and have a 0.9 probability.

The last column of the table gives values of maximized log-likelihoods. The likelihood value jumps from N = 125 to reach a plateau when N = 150. This means that the value of N is at least 150. However, if we examine the group distributions for various values of N, the following results are observed: By increasing the value of N, the number of small pools also increases, but the largest undiscovered pool size, however, may not be yet discovered if N > 152 (Table 3-4). The second prediction is rejected because the drilling density and seismic coverage suggest that the largest pool of the play should have been discovered. For each value of N, there are point and interval estimates for μ and σ^2. These initial interval estimates are used in the pool-size-by-rank estimation.

Figure 3-10 displays pool size distribution derived from the lognormal discovery process model when N = 152 and β = 0.3 (A) and the pool size distribution derived from random sampling assumption (B). It is evident that the mean and variance of the superpopulation pool size distribution are over- and under-estimated, respectively, if the random sampling assumption is made.

NONPARAMETRIC DISCOVERY PROCESS MODEL

Estimating Distributions

In the previous sections, we demonstrated how to use the lognormal superpopulation model (parametric) to estimate pool size distributions. We shall now discuss the use of a nonparametric superpopulation model without benefit of a prior distribution (Wang and Nair, 1988; Lee, Nair and Wang, paper in preparation).

A petroleum play contains N pools within the same underlying cumulative probability distribution F. Furthermore, if n pools are discovered randomly from the play, then the probability density for each pool is simply 1/n. Unfortunately, the n pools are not a random sample, but a biased sample from the play. Therefore, the statistical estimation of F requires use of the lognormal discovery process model as described above. On the other hand, under the discovery process model, the underlying parent distribution F can also be estimated without making any assumptions about its shape. In this case, the probability density for each discovery is estimated and then an empirical pool size distribution, F^(y), is obtained. This empirical distribution is graphically presented using quantile-quantile (Q-Q) plots.

Modeling Distributions

Essentially, probabilistic statistical analysis assumes that a set of data arises as a sample from some classes of probability distributions. This section deals with informal graphic methods used to assess distributional assumptions. The information applied to the graphic methods is based on the results of the nonparametric procedure described in the previous section. The advantage of the procedure is that it is not based on any assumption about the shape of a probability distribution. However, the procedure assigns mass only to the observed data, and assumes that the largest pool in the population is no larger than the largest pool in the sample, and that the smallest undiscovered pool is no smaller than the smallest discovered one. This is a clearly unrealistic situation. In order to overcome this disadvantage, the F estimated nonparametrically is approximated by various probability distributions, and then the best fit among the distributions is judged using the informal graphic procedure.

Suppose F^ is estimated nonparametrically and we test whether $F\hat{} = F_0$ where F_0 is the hypothesized distribution and is completely specified. There are a number of graphic methods which can be applied to test the hypothesis. The percent-percent (P-P) plot is checked to determine whether it falls along a straight line through the origin with a slope of one. However, the P-P plot has several disadvantages. Firstly, it only allows one to check the adequacy of completely specified distributions. In practice, it would be used more to determine the shape of the distribution, such as lognormality, rather than to specify the parameters. Secondly, if the plot is nonlinear, it becomes difficult to determine what alternative shapes one should consider.

The Q-Q plot, on the other hand, is designed to overcome these drawbacks inherited from P-P plots and can be used to assess the adequacy of a hypothesis whether a data set comes from a family $F_0[(y-\mu)/\sigma]$ for unknown location parameter μ and scale parameter σ. Under the hypothesis that the data set is indeed from a distribution with shape F_0, the data will follow a linear configuration. So one just needs to look for linearity without having to estimate values for μ and σ^2. If the linearity does exist, then the intercept of the line is an estimation of μ, and the slope is an estimation of σ. Nair (1984) presents examples about this fitting procedure. Departures from straightness in the theoretical Q-Q plot clearly indicate that the observed and theoretical distributions do not match. When they do not match, the plot may suggest the nature of the mismatches as follows: (1) presence of outliers at either end; (2) curvature at both ends, indicating long or short tails; (3) convex or concave curvature, related to asymmetry; and (4) plateaus. The significance of these mismatches will be discussed below (Chambers, Cleveland, Kleiner and Tukey, 1983).

<u>Outliers</u>. Samples of geological populations often contain outliers. It must be noted that, when outliers are encountered in a set of data, it is prudent to go back to the source of data, if possible, to verify the values. If they are in error, they can be corrected or set aside, but if they really belong to the population, they might be the most important observations in the sample.

<u>Long or short tails at both ends</u>. Another departure from the linearity often observed in Q-Q plots is displayed in Figure 3-11 (top); the ends of the configuration curve upward to the right and downward to the left. A straight line can be fitted to the central portion of the plot. This indicates that these data represent a longer tails than the hypothesized distribution, F_0.

<u>Symmetry</u>. Another possibility is that the theoretical distribution is symmetrical but the data are not. For example, if the plot is an inverted S-shaped (Fig. 3-11, lower left), then the data are more or less symmetrical but have a longer tail than that of F_0. On the other hand, if F_0 is a symmetrical distribution and the Q-Q plot is S-shaped, then the data may be more or less symmetrical, but have a shorter (lighter) tail than that of F_0.

<u>Plateaus</u>. Figure 3-11 (lower right) shows a related phenomenon. There are two rough plateaus. This means that there are two distinct clusters of points which are not accounted for by the theoretical distribution. Indeed, Figure 3-11 (lower right) contains two populations.

Validations Using Simulated Data Sets

The nonparametric discovery model was also validated using the simulated data sets. For purposes of comparison, μ^ and σ^2^ are computed after the estimation is completed. The results are listed in Tables 3-6, 3-7, and 3-8, and may be explained as follows:

1. The nonparametric model does not yield an estimate for N, but the model does provide upper or lower estimates for N.

2. With some exceptions, the model predicts β values reasonably well.

3. The estimated values of μ and σ^2 fall into 95% interval estimates, except in cases where there are high values of β.

4. As the sample size increases, the estimation interval is usually reduced.

5. The point estimates approach the population values as the sample sizes increase.

The Beaverhill Lake Play

The Beaverhill Lake data set was also evaluated using the nonparametric discovery process model. The analysis is summarized as follows: The likelihood value increases to a relatively flat value when N > 152 (Table 3-9). This indicates that the value of N is at least 150. These 150 "pools" include commercial pools and oil recoveries from drill-stem tests. The group distributions for different values of N are listed in Table 3-10 for purposes of comparison. Furthermore, if Table 3-5 and 3-10 are compared, one finds that the lognormal model predicts more both large and small pools than does the nonparametric model. The explanation for this discrepancy is that the small and large pools are predicted by the shape of the lognormal distribution. On the other hand, the nonparametric model cannot predict the lefthand side of distribution because there are no pools in the sample. The nonparametric distributions of the Beaverhill Lake play for N = 150 and β = 0.3 are displayed in Figure 3-12 (discrete distributions).

Various probability distributions (namely: normal, lognormal, power normal, geometric, Pareto, Weibull, one parameter exponential, and two parameter exponential distributions) were hypothesized and were fitted to the nonparametric distribution (Fig. 3-12). The results of the Q-Q plots for these distributions are displayed in Figures 3-13.

The result of the assessment of the distributional assumption is summarized as follows: the Beaverhill Lake data have a longer tail than do normal, power normal (with power = 0.5), uniform, gamma (with shape factor = 5 to 0.01), one parameter and two parameter exponential distributions. The Q-Q plots for the truncated and truncated shifted Pareto distributions display a combination of S- and inverted S-shapes. The lognormal, Weibull, and power normal (with power = 0.001) may have a slightly longer tail. However, the lognormal or power normal with low power is a better choice than other distributions, if one has to use a-prior distribution. The nonparametric discrete distribution of Figure 3-12 is approximated by a continuous lognormal distribution that is used to estimate individual pool sizes.

Chapter Four

POOL-SIZE-BY-RANK

> If you do not expect the unexpected, you will not find it; for it is hard to be sought out, and difficult.
>
> Heraclitus

In resource evaluations, the most useful type of information is the estimation of pool-size-by-rank (for the r-th largest pools), that is, the size of the largest pool, the second largest pool and so on. The minimum data required to obtain this type of information is (1) a pool size distribution and (2) the number-of-pools, N, in the play, or their distribution. The superpopulation concept must be assumed for this estimation. Furthermore, the pool size distribution and the number-of-pools distribution can vary independently, and a lognormal assumption is not required to operate pool-size-by-rank.

INTERPRETING POOL-SIZE-BY-RANK

If N = 1 (i.e. a single pool play), then the distribution of the largest and smallest pool is precisely given by the pool size distribution. More generally, if $X_1, X_2, \ldots, X_N$ are pool sizes generated independently from an identical pool size distribution, denoted by F_Θ where $\Theta = (\mu, \sigma^2)$, then the greater than distribution of the largest pool (assuming there are N pools) is

$$L_{N,r}(x) = 1 - [\,1 - F_\Theta(x)\,]^N, \qquad x > 0 \qquad (4\text{-}1)$$

The greater-than distribution of the r-th largest pool is given by

$$L_{N,r}(x) = \sum_{k=r}^{N} \binom{N}{K} F(x)_\Theta\,[\,1 - F_\Theta(x)\,]^{N-K}, \quad x > 0 \qquad (4\text{-}2)$$

for r = 1, 2, ..., N.

Equations (4-1) and (4-2) are the distributions of the largest and the r-th largest order statistics (Bickel and Doksum, 1977) for a random sample of size N from the superpopulation. When the number of pools is a random variable, then the density of the r-th largest pool can also be derived (Lee and Wang, 1983a) as follows:

$$l_r = \sum_{n=r}^{\infty} \sum_{k=r}^{n} \binom{n}{k} F(x)^k \, [1 - F(x)]^{n-r} \, f(x) \; P(N=n)/P(N \geq r)$$

$$\text{for } x > 0 \text{ and } r = 1, 2, \ldots \qquad (4\text{-}3)$$

where $P(N=n)$ = number-of-pools distribution when $N = n$, $P(N \geq r)$ = number-of-pools distribution when $N \geq r$, for $r = 1, 2, \ldots$ From Equation (4-1) we see that: (1) for a fixed set of parameters μ, σ^2, the probability of depositing a largest pool size of at least x increases to 1 as the total number of pools (N) increases, and (2) for a fixed N and also a given pool size x, the probability of the largest pool being at least x will increase as μ and/or σ^2 increases.

This information is interpreted geologically below:

1. If all pools in a play were deposited as a result of the same geological processes (i.e. they are part of the same population), then as the number of pools deposited rises, the more likely it is that one of them is a relatively large one.

2. The magnitude of the largest pool tends to change with respect to other pools for different values of μ and σ^2, i.e. with respect to different geological models.

For purposes of illustration, let us re-examine the Beaverhill Lake play. Here, as indicated by Figure 4-1 (top A), the Swan Hills A&B pool size (290 10^6 m^3) is located at the upper one percentile on the superpopulation pool size distribution. The interpretation is that the frequency of occurrence of a pool as large or larger than the Swan Hills A&B pool within the superpopulation is about 1%. On the other hand, the probability that the largest pool in the Beaverhill Lake play is as large as the Swan Hills A&B is not 1% but much larger (unless there is only one pool). In the case of more than one pool, the probability can be obtained from the distribution of the largest pool among N pools. The largest pool size distributions for N = 100 and 152, for example, are displayed in Figure 4-1 (top B and C, respectively), together with the superpopulation pool size distribution. The probabilities of having the largest pool size as large as Swan Hills A&B are 0.56% and 0.76% for N = 100 and 152, respectively. Given N = 152, for example, then 152 pools have been deposited with sizes generated from the superpopulation pool size distribution, the chance of having the largest of the 152 pools as large as Swan Hills A&B is 76%. That is, if the processes can be repeated under similar geological conditions, and 152 pools are generated at a time, then roughly 76% of the time the largest pool will have a size at least as large as the size of Swan Hills A&B. This is a frequentist interpretation of probability that employs the superpopulation concept of pool size distribution.

The distributions of the first two largest pool sizes, when N = 152, are displayed in Figure 4-1 (lower). This figure indicates that the probability of the largest pool being bigger than x (say, 290 10^6 m^3) is greater than the probability of the second largest pool being bigger than x. For example, the probability of having the second largest pool size as large as 290 10^6 m^3 is 0.17. This provides an indication about the ranking of the Swan Hills A&B pool.

The difference in size between two adjacent pools can be examined as a function of σ^2, if N and μ remain unchanged. In Figure 4-2 (left), the medians of individual pool size distributions, where $\mu = 0.25$, $\sigma^2 = 7$ and $N = 60$, are displayed by dots, whereas the medians of individual pool size distributions, where $\sigma^2 = 2$ and μ and N remain the same, are displayed by squares. This figure clearly indicates that pool size decreases more rapidly when σ^2 is relatively large than when σ^2 is relatively small. For any skewed pool size distribution, such as a lognormal one, given constant values of μ and N, the larger the value of σ^2, the bigger a single pool tends to be. Hence the magnitude of the first few large pools among the N pools tends to be greater.

Plays from the Western Canada Basin reveal an interesting pattern. They are the Beaverhill Lake, Bashaw, and Zama plays. Values of σ^2^ were estimated from pool size data published by the Energy Resource Conservation Board of Alberta. Figure 4-2 (right) displays the sizes of the largest ten pools for three plays, which have σ^2^ values of 6.6, 3.0, and 1.0, respectively. These ten pools include discovered and undiscovered pools of the plays. The sizes in the Beaverhill Lake play (indicated by the dots) decrease more rapidly than those of the Bashaw reef play (indicated by the crosses) and those of the Zama play (indicated by triangles). The reason for this change is that the pool size distribution for the Beaverhill Lake play has the largest variance of all. The reserves from the first ten pools amount to 91, 68 and 46 percent of their total resources, respectively. This phenomenon demonstrates that the magnitude of σ^2 allocates the resources to individual pools.

A summary of the above follows:

1. Distribution of pool-size-by-rank should be computed from the number of pools, N, and the pool size superpopulation distribution, which does not need to be lognormally distributed.

2. The size of the largest pool increases as the number of pools, N, increases. The amount of increase depends on the magnitude of μ and σ^2. For example, the pool size increases rapidly when σ^2 is large.

3. In resource evaluations, when the constant in Equation (5-1) is not scaled, μ dominates σ^2 in Equation (5-2) for the mean of a pool size. Therefore, parameters μ and N can be thought of as indicators of the richness of the play, whereas σ^2 and N are indicators of the outlier-proneness of the geological model.

4. For each hydrocarbon-bearing play, there is a set of μ, σ^2, and N values associated with the geological model that produced the play. Different geological models may have different values for μ, σ^2 and N, and correspondingly distinct pool sizes.

5. If a play has a pool size distribution with a large σ^2, then the major portion of the play's resources will be made up by the first few largest pools. On the other hand, if σ^2 is relatively small, then the pool sizes of the play will be more or less equal.

The number of pools in a play, which is a finite number, should include all small pools that may not be economically viable at present. One might be concerned that if small pools are included in the assessment, then the mean (of the untransformed data) of the pool size distribution will be reduced substantially. Consequently, the mean would not adequately describe the economic resources. Further explanation is required to clarify this statement: firstly, the mean of a population is not an ideal index for resource measurement, and secondly, a play that includes small pools may not be economically viable at the time of analysis. From the viewpoints of exploration and economic analysis, the remaining largest pool sizes are far more significant than the mean value of the pool size distribution.

As previously mentioned, the superpopulation concept is employed by PETRIMES. Thus, the predictions made by the system are of cases that would occur most frequently. A singular case, for example, is that of the Cardium marine sandstone play, in which the largest pool size is about 10 times larger than the size of the second largest pool. Under such situation, pools of sizes in between the two largest pools may be mistakenly predicted. However, if additional information indicates that no other sizes of pools would exist, then the information can be entered into the system as a condition for predicting the individual pool sizes.

The concept of pool-size-by-rank can also be explained using Monte Carlo simulation. Assume we have a pool size distribution and a number-of-pools distribution. A random number is generated and the number of pools, N_j (say 100), is obtained from the number-of-pools distribution. A total of N_j (= 100 in this case) pool sizes is randomly drawn from the pool size distribution(e.g. x_1, x_2, ..., x_{100}). These pool sizes are sorted in descending order, and the steps are repeated many times. The largest pool sizes from each simulation trial are then used to construct the size distribution of the largest pool. The size of second largest pool, the third largest pool and others are similarly obtained. In practice, the statistical approach for the estimation of individual pool sizes is more effective and can also provide various matching options.

Factors that distort estimations of pool-size-by-rank include the problem of mixed populations, error in the estimation of the number-of-pools and/or pool size distributions, and errors in measurement of pool sizes. The problem of mixed populations is the most severe one, and causes either under- or over-estimation of undiscovered pool sizes when prior distributions are specified. Revisions to the play definitions may solve the mixed population problem.

With respect to the significance of changes in the values of N, μ, and σ^2 and their impact on the estimation of individual pool sizes, the largest pool size is sensitive to the following factors (in decreasing importance): σ^2, N and/or μ, and errors in pool measurements.

MATCHING PROCESS

In the previous section, we described how pool-size-by-rank may be estimated when the total number of pools in the play, N, is known and the pool size distribution is provided. In this section, we shall determine values of $\hat{\mu}$

and σ^2^ from an interval estimate for a given N. The method begins by assigning values to N according to available geological information and values of μ^ and σ^2^ from the intervals. The individual pool sizes are then estimated. The estimated pool sizes should be matched with the discovered pool sizes to determine which values of μ^, σ^2^ and N yield satisfactory discovery fits. A satisfactory fit is one that is reasonable statistically and corresponds to the current geological interpretation of the play. This procedure combines geological interpretation and statistical analysis. This may be a complicated and time-consuming process; nevertheless, it can be quite useful.

Distributions of individual pool sizes can be conveniently characterized as a few selected upper percentiles without much loss of information. Take a pool size distribution as an example. The upper percentiles of 95, 75 50, 25, and 5 of the distribution are 5.5, 8.6, 11.9, 16.4, and 26.4 10^6 m^3 of oil (Fig. 4-3), respectively. For purposes of comparison, the variability of this distribution can be measured by its interquartile range, which measures the variability of the middle 50 percent of the distribution. In this example, the range for 25% to 75% is given as 16.4 - 8.6 = 7.8. The larger the interquartile range, the more variable the distribution; hence, the higher the degree of uncertainty.

There are several reasons for which the upper percentiles, as measurements of the individual pool size distribution, are preferred to the mean and the variance. They are: (1) the mean and standard deviation do not relate directly to probabilities; (2) the mean may tend to over-predict individual pool sizes; and (3) the standard deviation is typically larger than the mean, which makes it less useful for purposes of prediction and comparison.

The interval from the 75th upper percentile (8.6 10^6 m^3) to the 25th upper percentile (16.4 10^6 m^3) is a 50% prediction interval for the pool that contains the median. That is, the probability that the pool will have a value between 8.6 and 16.4 10^6 m^3 is 50%. Similarly, 5.5 to 26.4 10^6 m^3 is a 90% prediction interval for the largest pool. The latter prediction interval has a higher probability of occurrence, but at the expense of having a much wider interval, i.e. more uncertainty. In what follows, we will start with the 75% - 25% prediction interval as a statistical measure of goodness-of-fit, and the median will be used as a point estimator of pool-size-by-rank.

The 75-25% interval, in fact, was derived from pilot studies. In our experience, the most effective method is to divide the sample into two sets. The first sample set is used to establish an interval (such as 75-25% or 95-5%) that can predict present discoveries. The interval derived may then be used to predict future ones. In cases where the 75-25% interval does not match most or all of the discoveries, then the 95-5% interval should be used to match present ones.

Geological gauges for measuring goodness-of-fit compared to statistical fit method are difficult to quantify. Statistical fits may be verified by examining their geological implications. After each statistical fit, we observe whether or not the implications are in accord with the geological model. Examples of the type of questions that one should ask following each fit are:

Have we discovered the largest pool?
What are the sizes of the remaining largest pools?
What is the potential of the remaining undiscovered pools?
Have we predicted enough small pools for the play?
How do recent discoveries, which are not included in the analysis, fit into the prediction picture?

AN EXAMPLE OF MATCHING

If we take, for example, the Beaverhill Lake play, we find that both the lognormal and the nonparametric discovery process models indicate that values of N range from 150 and up. The matching procedure is explained as follows:

1. The intervals for μ and σ^2 are chosen from the output of lognormal and/or nonparametric discovery process models. For example, the intervals from the lognormal model, when N = 150 and β = 0.3 are:

$$-7.93 \leq \mu \leq -5.54, \text{ and}$$
$$19.45 \leq \sigma^2 \leq 41.35.$$

2. The final intervals used in the matching process are:

$$-8 \leq \mu \leq -6 \text{ with an increment of } 1,$$
$$20 \leq \sigma^2 \leq 45 \text{ with an increment of } 10 \text{ or } 5.$$

3. The pool-size-by- rank are computed for each combination of increments, i.e., $\mu\hat{}$ = -8 and $\sigma^2\hat{}$ = 20. The estimated individual pool sizes are then matched with the present discoveries.

4. Repeat steps 1 to 3 for other intervals.

5. When all possible intervals have been analyzed, one acceptable match is chosen among all matches, based on geological criteria.

There are other options to choose pool size distributions for matching purposes. The first option is to take the empirical pool size distribution derived from the nonparametric discovery process model and the second option is to use the values of $\mu\hat{}$ and $\sigma^2\hat{}$ obtained from the lognormal approximation of the empirical pool size distribution. The advantages of the second over the first option has been discussed in the, "Modeling Distribution" section in Chapter Three.

If we take the lognormal approximation in Figure 3-12 having values of μ = -6.80 and σ^2 = 29.55 when N = 152, and β = 0.3 as an example, we find that the pool-size-by-rank listed in Table 4-1 (also see Fig. 4-5) may be interpreted as follows:

1. The best match interval is 95-5% because it matches to all discoveries, except the 8th, 9th, and 10th ranks. The unmatched ranks may result from irregularities of nature, given that the pool sizes estimated are just one of

several possible cases from the superpopulation studied.

2. The largest 17 pools have been discovered.

3. The largest remaining pool size is about 0.7 10^6 m^3.

4. If we examine Table 4-1 carefully, we find that there are a number of options for us to match the discoveries, such as matching the 11th discovered pool rank to the 12th population pool size rank; or matching the 12th discovered pool rank to the 14th rank. In order to verify any of these matches, the following ad-hoc procedure is used: (1) The pool area corresponding to the remaining largest pool size is obtained from the cross-plot of the pool areas vs pool sizes of the Beaverhill Lake play (Fig.4-4), and (2) the pool areas so obtained can be validated against seismic coverage when seismic grids are small enough to reveal prospects having the range of pool areas. For now, the match displayed in Figure 4-5 (left) is assumed to be the final match for purposes of illustration(see below).

POOL SIZES CONDITIONAL TO POOL RANKS

As indicated in Figure 4-5 (left), the predicted pool sizes have a wide range of prediction intervals which overlap with the two adjacent pool sizes. This overlapping phenomenon is a result of the uncertainty in the estimations. In this section, we introduce a method by which the uncertainty in the estimation of pool size can be reduced when pool rank is entered into the analysis (Lee and Wang, 1985).

Once the acceptable match has been obtained, the remaining individual pool sizes and hydrocarbon potential of the play can be estimated by adding conditions to the match. For the Beaverhill Lake play, the remaining pool sizes were estimated by constraining the pool sizes of the 87 discoveries and their ranks. Figure 4-5 (left) displays the result based on N = 152 and is discussed below.

1. The largest remaining pool sizes range (0.9 probability) from 2 10^6 to 6 10^6 m^3 of oil in-place, with a median of 4 10^6 m^3.

2. The range of the prediction intervals following conditional analysis are smaller than those of intervals for which conditional analyses were not performed. For example, the remaining largest pool size ranges from 2 10^6 to 6 10^6 m^3 as compared with the 0.7 10^6 m^3 to 10 10^6 m^3 shown in Figure 4-5 (left).

3. The overlapping range of two consecutive pool sizes is also much smaller than the case shown in Figure 4-5 (left).

4. The degree of uncertainty in the prediction intervals is controlled by four factors: (1) the uncertainty inherited from the superpopulation; (2) errors in measurements of the pools, (3) the ratio of the number of discoveries to total number of pools, and (4) the difference in reserves between the two nearest pools, as illustrated by the 21th to 28nd pools of Figure 4-5 (right).2

5. Individual pool size distributions computed from specified discovery records tend to be less skewed and more concentrated around the medians than those computed without specified conditions.

The estimation of pool sizes constrained to a discovery record serves not only to estimate remaining resources, but can also reduce the uncertainty inherited from its superpopulation.

PLAY RESOURCE DISTRIBUTION

A geological model can generate a variety of play resource values under the superpopulation concept. All these resource values constitute the play resource distribution for the model.

In PETRIMES, the play resource distribution is the sum of all pool sizes of the play. If the pool sizes are approximated using lognormal distributions, then the play resource distributions are the sum of the lognormal distributions. This summation does not have an analytical form. Therefore, the summation is executed numerically by convolution, or by using a Monte Carlo procedure. The mean and the variance of the play resource distribution are as follows:

$$E[\,T\,] = \Theta \times E[\,X\,] \times E\,[\,N\,] \tag{4-4}$$

$$\sigma_T^2 = \sigma^2 \times E[\,N\,] + (E\,[\,X\,])^2 \times \sigma_N^2 \tag{4-5}$$

where $E\,[\,X\,]$ = mean of the pool size distribution,
σ^2 = variance of the pool size distribution,
$E\,[\,N\,]$ = mean of the number-of-pools distribution,
Θ = exploration risk, and
σ_N^2 = variance of the number-of-pools distribution.

Figure 4-6(A) displays the play resource distribution for the Beaverhill Lake play. There is a 50% chance that the play will have a play resource ranging from 887 10^6 to 15,739 10^6 m^3 of oil in-place. An amount of 941.43 10^6 m^3 of oil has been discovered.

This play resource distribution is the superpopulation distribution and contains the uncertainties explained in the previous chapter. If the pool ranks of discoveries are known, then the uncertainties and the variance of the distribution can be reduced by constraining the distribution.

PLAY POTENTIAL DISTRIBUTION

Play potential is defined as undiscovered resources which can be estimated from play resource distribution conditional to matched pool ranks.

Figure 4-6(B) is the play potential distribution (conditional to the output of the match) for the Beaverhill Lake play. Interpretations follows.

1. There is a 50% chance that the remaining play potential ranges from 13 10^6 m^3 to 16 10^6 m^3.

2. There is a 90% chance that the potential ranges from 12 10^6 to 18 10^6 m^3.

3. The mean and standard deviation of the distribution are 15 10^6 m^3 and 1.8, respectively.

4. The median equals 15 10^6 m^3; i.e. there is a 50% chance that the remaining play potential will be greater than 15 10^6 m^3.

Chapter Five

EVALUATING CONCEPTUAL PLAYS

> As time goes on qualitative methods
> are replaced by quantitative methods.
>
> F.Y. Loewinson-Lessing

GEOCHEMICAL APPROACHES

A conceptual play is a play which has not yet been proven by exploration and is postulated from geological information, whereas an immature play is a play in which a number of discoveries have been made but the number of discoveries is not sufficiently large to apply the discovery process model described in Chapter 3. When a conceptual play is evaluated, the amount of data available for assessment may range from no data at one end to some data at the other end. Therefore, the methods for evaluating a conceptual play are based on the amount and types of data available for assessment. Possible methods are outlined as follows:

Types of Data Available	Method
No data	Geologic comparative study
Types of source rocks (conceptual or real)	Types of expected products (oil, gas, or heavey oil)
Conceptual or real stratigraphic columnar section	Burial history, maturation stages, oil and gas windows; timing of hydrocarbon generation
Areal extent and volume of source rock	Amount of oil or gas generated; timing

Detailed descriptions of the above methods are beyond the scope of this book. However, the numerical method presented in this chapter may be applied to conceptual or immature plays. For immature plays, discoveries may be applied to validate estimates obtained.

POOL SIZE EQUATIONS

In reservoir engineering, a pool size can be calculated by using the equation as follows:

Pool size = Constant x Pool Area x Net Pay x Porosity x Hydrocarbon Saturation x Recovery Factor / Gas or oil Formation Volume Factor (5-1)

For resource evaluations, Equation (5-1) is adapted to define a pool size distribution (Roy, 1979). In PETRIMES, the geological variables are jointly approximated by using a multivariate lognormal distribution. Because the result of product and/or division of lognormal random variables is again a lognormal variable (Aitchison and Brown, 1973), it follows from Equation (5-1) that the pool size distribution is lognormal. If we let μ, σ^2 and σ_{ij}, i,j = 1, 2, ..., denote the means, variances and covariances of the natural logarithms of the geological variables, then the mean and variance of a pool size distribution are given by:

$$\text{Mean} = \exp\left(\mu + 1/2\ \sigma^2\right) \tag{5-2}$$

$$\text{Variance} = \exp\left(2\mu + \sigma^2\right)\left[\exp\left(\sigma^2\right) - 1\right] \tag{5-3}$$

$$\mu = \ln(\text{constant}) + \Sigma\, \mu_i \tag{5-4}$$

$$\sigma^2 = \Sigma\, \sigma^2_i + 2 \sum_{i<j} \Sigma\, \sigma_{ij} \tag{5-5}$$

To solve Equation 5-1, the various distributions are multiplied together. This type of multiplication can be accomplished using the Monte Carlo approach or an approximation of a lognormal family to the distributions and the operation of products of lognormal distributions. Usually the result derived from the Monte Carlo approach has a smaller variance compared to that derived using the lognormal approximation method (Lee and Wang, 1983a). Equation (5-1) can be applied to both mature and conceptual plays.

For conceptual plays, we have no discovery record to apply to the discovery process model. The pool size equation may then be used to derive a pool size distribution, as shown in Equation 5-1. Furthermore, distributions of variables such as pool area, net pay, and so on are based on interpretations by geologists and/or comparative studies. These are considered to be superpopulation distributions.

For the Beaverhill Lake play, variabilities in hydrocarbon saturation and the oil shrinkage factor are relatively small in comparison with other variables. Furthermore, there is no significant correlation between hydrocarbon saturation and the oil shrinkage factor with the other variables; hence, not much is lost by ignoring them in the total variance. Thus, we shall only consider pool area, average net pay and average porosity. Equation (5-1) is reduced to:

$$\text{Oil pool size in-place } (10^6\ m^3) = \text{Constant} \times \text{Pool Area} \times \text{Net Pay} \times \text{Porosity} \qquad (5\text{-}6)$$

where the constant is equal to the product of average hydrocarbon saturation, average oil shrinkage factor, and the conversion factor from hectare-metre to million cubic metres, 0.00681.

The reason for computing the oil in-place is that enhanced oil recovery techniques have been applied to some but not all of the pools. Thus, the recovery factor for the play varies greatly from a few percent to as much as 25 percent. Incorporation of the recovery factor here will introduce an inconsistent measurement of pool size. Nevertheless, the system can still handle all relevant factors in Equation (5-1).

Detailed information for each geological variable is given in Table 5-1 (raw data were obtained from the report by Energy Resources Conversation Board, 1989). From this table we can see that the pool area contributes most to the values of μ and σ^2 and therefore, it is most important contributing variable to the pool size equation. Correlations and covariances of the three variables are also given in Table 5-1. The pool area and average net pay variables (Fig. 3-4, right), as well as porosity and pool area are highly correlated, having a correlation coefficient of 0.682. In this example, if the covariances are incorporated, the mean of the pool size is 151 10^6 m^3 of oil. In contrast, if they are all ignored, the mean is reduced to 46 10^6 m^3 of oil. In a similar way, if negative correlations are omitted, then the mean will be overestimated.

The advantage of using the pool size Equation (5-1) is that we can gain a better understanding of the variables, their interdependencies and their influence on pool size distribution. Furthermore, geological variables for an undiscovered pool such as pool area, average net pay, and others can also be regenerated for a given pool size by using Equation (5-1) (see the section on Generation of Reservoir Parameters). Moreover, we usually do not have sufficient data to compute covariances of geological variables for conceptual plays. Therefore, the variance of the pool size distribution may be under- or over-estimated. Furthermore, correlations of variables may change from population to population. For example, log-linear relationships between porosity and water saturation for the Cardium marine sandstone (Fig. 5-1, left) and the Bashaw reefs (Fig. 5-1, right) display very distinct correlation patterns. Examination of possible correlations may lead to justification of the addition or subtraction of variance from the pool size distribution. If a lognormal distribution is adopted, then the variance and covariance can be adjusted. The Beaverhill Lake play was used to describe the roles of Equation 5-4 and 5-5 in the pool size equation. This is a very serious problem, and no adequate solution to it has yet been proposed.

The preceding example was given to demonstrate the impact of correlation on the mean of a pool size distribution. In this example, the sample covariance matrix was computed and used. The population covariance matrix should, in fact, be computed using a multivariate discovery process model (Wang and Nair, 1988; Lee, Nair and Wang, manuscript in preparation).

CONSTRUCTION OF PROBABILITY DISTRIBUTIONS

After pool size equation variables are chosen, probability distributions are estimated either using statistical methods or geological interpretations. In this section, guidelines for constructing probability distributions from geological information are outlined. In constructing probability distributions for frontier plays, all relevant data and information from similar basins should be collected by the assessment team to address the following questions and concerns:

1. In frontier cases, the first question that should be asked is: What is the probability that the play exists? The existence of a play can be analyzed in terms of the presence or absence of factors such as source rocks, maturation, migration, favourable reservoir facies and so on. A marginal probability is applied to each factor to indicate that the factor would exist (See "Geological Risk Factors").

2. If the geological model in question has an extreme range for the values, then the variance of the variable should be relatively large. On the other hand, if values are quite uniform, then variances should be small.

3. Remember that we do not have enough data to compute covariances between variables; however, positive or negative covariances are evident from the geological data. Therefore, if the largest pool size estimated is not what we expect, the following questions should be addressed: Are the mean and variance of the pool size distribution adequate? How much covariance does exist?

4. What is the sensible maximum value that the model can never have? This value will be set at zero in the upper percentile of the probability distribution.

5. What is the largest possible value that the model may have? This value will be set at one or two upper percentiles of the probability distribution.

6. What is the value for each variable above which half of the members of the population will be greater? This value will be placed at the 50 percentile of the distribution.

7. What is the minimum value? This value will be set at 100 in the upper percentile.

8. In determining the geological risk factors that dictate the final accumulation of hydrocarbons, one should ask the question "What are the most unpredictable risk factors in this model?"

9. The number of prospects may be obtained from anomalies showing closure on a structural contour map of time isochrons constructed from seismic data. Some questions, however, remain unanswered: How many anomalies were not detected by the present orientation and density of seismic lines? What is the maximum number of prospects that could exist in this play? How many prospects would be there at 50% chance? The answers to these questions provide us with information needed

to construct probability distributions for the prospects. Other values at various upper percentiles can also be used.

For each probability distribution, values of the four upper percentiles (1.0, 0.5, 0.02 or 0.01, 0.0) are required to construct a probability distribution. The process is as follows: a lognormal distribution is fitted to these four values and generates all other upper percentiles. In general, assessors may either (1) enter the four upper percentiles and let the shape of a lognormal distribution generate other percentiles or (2) enter all percentiles and examine how good is the lognormal approximation.

Table 5-2 presents a format for entry of probability distributions. Formats for entry of exploration risks (Table 5-3) and numbers of prospects and pools are also presented in Table 5-4.

GEOLOGICAL RISK FACTORS

Exploration Risk

A play consists of a number of pools and/or prospects. It is possible that a prospect may not contain hydrocarbons. Thus, associated with each prospect there is an exploration risk that measures the probability of a prospect being a pool. Methods for quantifying exploration risks are described below.

Geological risk factors that determine accumulation of hydrocarbons include, for example, presence of closure and of reservoir facies, as well as adequate seal, porosity, timing, source, migration, preservation, and recovery. For a specific play, only a few of these factors, such as presence of closure and of reservoir facies, and adequate source and seal, are recognized as critical to final accumulation. Consequently, if a prospect located within a sandstone play, for example, is to be tested, it could prove unsuccessful for any of the following reasons: a lack of closure; unfavourable reservoir facies; lack of an adequate source or migration path; and/or absence of cap rock.

Geological risk factors present or absent may be expressed as marginal probabilities. For example, if the marginal probability for the presence of closure factor is 0.9, then this means there is a 90% chance that prospects drilled will have adequate closure. For a prospect to be a pool, simultaneous presence of all the geological factors in the prospect is necessary. This requirement leads us to the analysis of exploration risks.

Methods for Estimating Marginal Probability

When we assess a conceptual play, we begin by formulating a play definition. At this stage, a number of questions emerge:

Does the play in question exist?
Does the play have an adequate source?
Can we recover oil or gas from a play which lies under deep water?

Some of the geological factors such as source, maturation, and migration, for example, would normally exist throughout a play, but at an early stage of exploration, we cannot say whether these factors occur throughout. The system provides ways of handling this type of uncertainty. To this end, we need to explore the concepts of play level and prospect level risk.

Play Level Risk. This risk measures the chance that a geological factor is common to all prospects within the play. Play level risk is a regional phenomenon across an entire play. The occurrence of play level risk is denoted by G (global), and the marginal probability of this event is represented by Θ_g. White (1980) referred to play level risk as play chance, or group risk (White and Gehman, 1979; Gehman, Baker and White, 1981).

If a play contains hydrocarbons, all geological factors are present. Let these factors or events be denoted by $G_1, G_2, \ldots G_j$. The probability of a play having hydrocarbons, then, is

$$\begin{aligned}\Theta_{gi} &= P\,[\ G_i\ ] \\ &= P\,[\text{ the play has factor } G_i\ ] \\ &= P\,[\text{ the geologic factor } G_i \text{ is satisfied for all prospects} \\ &\qquad \text{within the play, } i = 1,\ldots j\ ] \end{aligned} \qquad (5\text{-}7)$$

For example,

G_1 = {adequate source}, G_2 = {adequate preservation}, ...
If all play level geological factors exist, then

$$\begin{aligned}\Theta_g &= P\,[\ G_1 \cap G_2 \cap \ldots\ldots \cap G_j\ ] \\ &= P\,[\text{ play possessing all factors }] \end{aligned} \qquad (5\text{-}8)$$

If any of these values G_i are not satisfied, then the play does not contain hydrocarbons. If Θ_{g1}, Θ_{g2}, ..., Θ_{gj} are statistically independent, then the marginal probability of having all play level risks simultaneously is defined as follows:

$$\Theta_g = \prod_{i}^{j} \Theta_{gi}, \text{ for } i = 1, 2, \ldots\ldots, j \qquad (5\text{-}9)$$

This play level risk can be considered as a parameter to be estimated from data, or may be the expression of an expert's opinion.

White (1980) describes a facies-cycle wedge (Fig. 5-2) as a body of sedimentary rock bounded above and below either by regional unconformities or by the tops of major nonmarine tongues. The ideal wedge represents a transgressive-regressive cycle of deposition including, from base to top, a vertical succession that varies from nonmarine to coarse-textured marine, to fine-textured marine, to coarse-textured marine and back to nonmarine facies. Exploration plays located within a facies-wedge can be allocated into either wedge-base, wedge-middle, wedge-top, wedge-edge, or subunconformity plays. Each play type listed above is associated with a play level risk. White summarizes 1150 plays in 80 productive basins of the free world and presents the relationships between play characteristics and the chances of the play containing hydrocarbons. The results are reproduced in Table 5-5.

Prospect Level Risk. This measures the marginal probability that a geological factor exists for an individual prospect. Prospect level risk is represented by R(local), and its marginal probability is denoted by Θ_r. The risk can also be considered as a superpopulation parameter, and may be estimated from data. For prospect level risk, absence of such factors such as closure, reservoir facies, or porosity will result in a prospect lacking of hydrocarbons. This, however, does not imply that these factors are also absent from other prospects.

Let R_1, R_2, ..., R_k denote the geological factors for an individual prospect at the prospect level. For example,

R_1 = {presence of closure},
R_2 = {adequate seal},... and so on.

Let us define $G = G_1 \cap G_2 \cap \ldots \cap G_l$;
$R = R_1 \cap R_2 \cap \ldots \cap R_k$.

A prospect within a play contains hydrocarbon if and only if (i) the play has all the play level risk factors; and (ii) the prospect has all prospect level risks. In other words, a prospect contains hydrocarbons if and only if $G \cap R$.

Define $\Theta_{ri} = P[R_i \mid G]$ and $\Theta_r = P[R \mid G]$, then the probability of there being hydrocarbons present is defined as:

$$P[\text{ a prospect containing hydrocarbon }] \quad (5\text{-}10)$$
$$= P[G \cap R] = P[R \mid G]\, P[G]$$
$$= \Theta_r \times \Theta_g.$$

If the geological risk factors are independent, then the prospect level risk is defined as

$$\Theta_r = \pi_i^k \Theta_{ri}, \quad \text{for } i = 1, 2, \ldots, k. \quad (5\text{-}11)$$

If the risk factors are not independent, then the rule of multiplication or conditional probability rule must be applied as follows:

$$\Theta_r = P[R_1 \cap R_2 \cap \ldots \cap R_k]. \quad (5\text{-}12)$$

Identification of the presence or absence of a particular prospect level risk factor may be accomplished by integrating information obtained from tested wells together with adjacent wells. For example, presence or absence of closure can be recognized by reviewing stratigraphic correlations after drilling; the existence of reservoir facies is identified from mechanical logs; adequacy of seal may be established by examining (i) the presence or absence of cap rock, (ii) the quality of the seal, and (iii) possible leakage of the closure; adequate source and migration factors means that oil has migrated into the trap. Therefore, if a potential reservoir is shown from drill-stem tests to either contain oil, oil shows, or oil traces, then the factor is considered as present.

More Examples of Marginal Probability Distributions. Figure 5-3 (top) displays a probability distribution for the adequate maturation geologic risk factor. The assumption used here is that either the sample size is large enough to represent the play (population), or the sample is a random sample from the play (population). We also assume that geochemical interpretations are valid.

The distribution suggests that there is a 70% chance that the percent hydrocarbon in extract from the play in question ranges from 40% to 60%, which is considered to be a mature source rock, i.e.:

$$P\ [\ 40\% \leq \text{mature} \leq 60\%\ ] = 0.70.$$

Figure 5-3 (lower) displays the probability distribution for total organic carbon. From this distribution, there is a 70% chance that the play has a total organic carbon content in excess of 0.5. The marginal probability for adequate source is interpreted to be 0.7, i.e:

$$P\ [\ \text{TOC} \geq 0.5\%\ ] = 0.7.$$

DEPENDENCE OF PROSPECT LEVEL RISKS

Traditionally, exploration risk is an expression of the products of marginal probabilities of geological risk factors such as presence of closure, presence of reservoir facies, adequate source, and adequate seal, to name a few of them. The statistical assumption presumed in such a product operation is that risk factors are independent. The assumption of independence of risk factors has been challenged by exploratory well data obtained from the Huang-hua Basin of eastern China.

A total of 242 exploratory wells in a sandstone play from the Huang-hau Basin (Fig. 5-4) were interpreted to determine why a particular well failed. In this case, the presence or absence of closure and of reservoir facies, as well as the adequacy of source and seal were recorded for each well. Table 5-6 presents some of the results obtained by Lee, Qin and Shi (1989). In this table, the number 1 indicates that the factor is present, and the number 0 is used to indicate that the factor is absent.

Firstly, if we assume that these factors are independent of each other, then, the overall prospect level risk is the product of 184/242, 220/242, 185/242, and 228/242, which equals 0.50.

Secondly, the geological factors were analyzed using the following conditional probability formula:

$$\begin{aligned} &P\ [\ \text{Closure} \cap \text{Reservoir Facies} \cap \text{Source} \cap \text{Seal}\ ] \\ &\quad = P\ [\ \text{Closure}\ ]\ \times \\ &\qquad P\ [\ \text{Reservoir Facies} \mid \text{Closure}\ ]\ \times \\ &\qquad P\ [\ \text{Source} \mid \text{Closure} \cap \text{Reservoir Facies}\ ] \\ &\qquad P\ [\ \text{Seal} \mid \text{Closure} \cap \text{Reservoir Facies} \cap \text{Source}\ ] \end{aligned}$$

$$= 184/242 \times 127/184 \times 111/127 \times 109/111 = 0.45. \quad (5\text{-}13)$$

The difference between these two approaches is 0.05. This example demonstrates that geological risk factors may not be independent. The dependency between any two risk factors has been studied further by using chi-square tests which have indicated that three pairs of risk factors (closure and source, closure and seal, and facies and source) are dependent risk factors, whereas other pairs are independent. The data set was also subjected to correlation analysis. In summary, for all dependent pairs of risk factors, significant correlations can be established.

CONCEPTUAL PLAYS

Pool Size Equations

One of the plays from the east coast of Canada was selected as an example to illustrate the application of PETRIMES in a conceptual play. The data used include probability distributions of area of closure, reservoir thickness, porosity, and trap fill. The equation used to calculate a pool size distribution follows:

$$\begin{aligned}\text{Pool size} = c \times & \text{ Area of Closure} \times \text{Reservoir Thickness} \\ & \times \text{Porosity} \times \text{Trap Fill}\end{aligned} \quad (5\text{-}14)$$

where c is the product of hydrocarbon saturation and a conversion factor of cubic feet to millions of barrels, and pool size is oil in-place measured in terms of millions of barrels.

From the superpopulation concept, the probability distributions for reservoir thickness, porosity, and trap fill are superpopulation distributions. The distribution of the area of closure, for example, was derived from structural contour maps based on seismic data. The distribution proposed by geologists was plotted as a solid line on Figure 5-5, and was approximated by a lognormal distribution (indicated by dots in the same figure). This method was used for variables of reservoir thickness, porosity, and trap fill.

If the geological variables are approximated by lognormal distributions with parameters μ, and σ^2, and if they are independent, then

$$\ln X = \ln c + \Sigma \ln Z_i \quad (5\text{-}15)$$

is normally distributed with $\mu\hat{} = 2.882$ and $\sigma^2\hat{} = 2.5$, and its density is given by

$$h(x) = (1 / x\, \sigma \sqrt{2\pi}) \exp [-(\ln x - \mu) / 2\, \sigma^2] \quad (5\text{-}16)$$

where x is pool size in millions of barrels.

Equation (5-16) was plotted as circles on Figure 5-6. At the same time, the pool size distribution was derived using the Monte Carlo approach, based on the original four distributions plotted as solid lines in Figure 5-6.

In this example, the pool size distribution derived from a Monte Carlo simulation agrees remarkably well with the lognormal distribution, except at the 0.5% level, because the Monte Carlo simulation usually tends to yield a less skewed distribution, whereas a lognormal approximation tends to extend the tail of the distribution.

Exploration Risk

Table 5-7 displays the risk factors and their marginal probabilities for the conceptual play. The geological risk factors were interpreted by the assessor as either play or prospect level risks. The first column displays the names of the geological factors, and the second column lists the corresponding marginal probabilities. The last two columns display the interpretations of each risk as prospect level or play level. For Case I, only the adequate timing factor is considered as a play level risk, whereas in Case II, adequate timing, adequate source and adequate preservation factors are considered as play level risks. There is no information to suggest that these risk factors are either dependent or not. Therefore, overall play level risk is calculated from the multiplication of all play level marginal probabilities. Similarly, the overall prospect level risk is the product of all marginal probabilities of prospect level. Exploration risk is the product of overall play and prospect levels. From the table, the two overall risks are very different for these two cases, but the exploration risk is identical. Because of the difference in play level and prospect level risks, the subsequent estimation will be different too.

Number-of-Prospects Distribution

If an identifiable type of trap, such as an anticline, can be mapped on the surface of a play or be detected seismically at depth, then the number of prospects can be counted. Some of the prospects cannot be mapped on the surface because of the presence of vegetation or be detected at depth because the seismic coverage is too sparse to detect small prospects. Three questions should be asked at this point: (1) What is the maximum number of prospects that the play could have? (2) Given a 50% chance, what is the least number of prospects that the play will have? and (3) What is the observed number of prospects? Based on the answers to these questions, one can construct a number-of-prospects distribution. This distribution can be considered as a superpopulation distribution.

Figure 5-7 displays an example of a number-of-prospects distribution for a conceptual play. The mean and variance of the distribution are 103 and 77.09, respectively. Given a 50% chance, the play will have more than 100 prospects.

Number-of-Pools Distribution

The number-of-prospects distribution will be used with exploration risk to derive a number-of-pools distribution. Let M be the random variable denoting the total number of prospects in a play and m be a value of M. Let its probability function be

$$P[m] = P[M = m], \quad m = m_0, \ldots\ldots, m_1$$

This distribution may be obtained from seismic detections and expert knowledge of the play.

Associated with the i-th prospect, we define

$$I = \begin{cases} 1 & \text{if the i-th prospect satisfies the condition R} \\ 0 & \text{otherwise} \end{cases}$$

for i = 1, 2, ... If it is given that event G has occurred, i.e. the play has all the conditions necessary for hydrocarbon occurrence, the total number of pools in the play is given by

$$N = I_1 + I_2 + \ldots\ldots\ldots + I_m$$

Note that N is a sum of random variables. The conditional probability distribution of N, given G (the play exists), is

$$\begin{aligned} &P[N = n \mid G] \\ &= \sum_m P[N = n, M = m \mid G] \\ &= \sum_m P[N = n \mid M = m, G]\, P[M = m] \\ &= \sum_m P[I_1 + I_2 + \ldots + I_m = n \mid M = m, G]\, P[m] \end{aligned} \tag{5-17}$$

where N = random variable for the number of pools; n = a specific value for N. We have assumed { M = m } is statistically independent of G for all m. Moreover, we assume $I_1, I_2, \ldots$ are independent of M and all I_i's are also independent.

Since $P[I_i = 1 \mid G] = \Theta_r$, for all i, then

$$P[N = n \mid G] = \sum_m \binom{m}{n} \Theta_r^n (1 - \Theta_r)^{m-n} P[m], \quad \text{for } n = 0, \ldots, m_1$$

The sum extends from $m = \max(n, m_0)$ to m_1.

The distribution of N is now given by

$$P[N = n] = P[N = n \mid G]\, P[G] + P[N = n \mid G']\, P[G']$$

$$= \begin{cases} (1 - \Theta_g) + \Theta_g \sum_m (1 - \Theta_r)^m P[m], & \text{if } n = 0 \\ \Theta_g \sum_m \binom{m}{n} \Theta_r^n (1 - \Theta_r)^{m-n} P[m], & \text{if } n \geq 1 \end{cases} \tag{5-18}$$

Also

$$P [\text{ play has at least one pool }] = P [N \geq 1]$$
$$= 1 - P [N = 0]$$
$$= \Theta_g [1 - \sum_m (1- \Theta_r)^m P [m]]. \qquad (5-19)$$

For example, Case II gives $\Theta_g = 0.57$ and $\Theta_r = 0.68$. If $M = 6$, then

$P [N \geq 1] = 0.57 \{ 1 - (1 - 0.68)^6\} = 0.57$, and

$P [N = 0] = 1 - 0.57 = 0.43$, or 43%.

The expectation of N is given by

$$E [N] = \Theta_g E [N | G] + (1 - \Theta_g) E [N | G']$$
$$= \Theta_g \sum_n n \Sigma \binom{m}{n} \Theta_r^n (1 - \Theta_r)^{m-n} P [m]$$
$$= \Theta_g \sum_m m \Theta_r \Sigma_n \binom{m-1}{n-1} \Theta_r^{n-1} (1- \Theta_r)^{m-n} P[m] \qquad (5-20)$$

Therefore, $E [N] = \Theta_g \times \Theta_r \times E [M]$, (5-21)

$E [M] = \sum_m m P [m]$.

= expected number of prospects.

Similarly,

$$E [N^2 | G] = \Sigma_m \Sigma_n n^2 \binom{m}{n} \Theta_r^n (1 - \Theta_r)^{m-n} P [m]$$
$$= \Sigma_m [m \Theta_r (1 - \Theta_r) + m^2 \Theta_r^2] P [m]$$
$$= \Theta_r (1 - \Theta_r) E[m] + \Theta_g^2 \times \sigma_M^2 + \Theta_r^2 (E[M])^2$$

Hence,
$$VAR [N] = E [N^2] - [E [N]]^2$$
$$= \Theta_g E [N^2 | G] - \Theta_g^2 \times \Theta_r^2 \times E [M]^2$$
$$= \Theta_g \times \Theta_r^2 E [M]^2 - \Theta_g^2 \times \Theta_r^2 \times E [M]^2 +$$
$$\Theta_g \times \Theta_r (1 - \Theta_r) \times E [M] + \Theta_g \times \Theta_r^2 \times \sigma_M^2. \qquad (5-22)$$

Therefore,
$$\sigma_N^2 = \Theta_g \times \Theta_r [\Theta_r (1 - \Theta_g) \times E [M]^2 +$$
$$(1 - \Theta_r) E [M] + \Theta_r \times \sigma_M^2] \qquad (5-23)$$

It is obvious that σ_N^2 is dominated by $E [M]$, because the contribution from σ_M^2 is diminished by the multiplier Θ_r.

The number-of-prospects distribution (Fig. 5-7) and the risks for Case I and Case II (Table 5-8) were applied to derive the number-of-pools distribution. From the result we can conclude that:

1. Their means are identical, but Case II has a much larger standard deviation.

2. Given a 50% chance, the play will have more than 42 pools for Case I and 62 pools for Case II.

3. For Case I, there is about 5% chance that the play has no pools, whereas the chance for no pools is about 57% for Case II.

4. The play is interpreted as a very risky play in Case II.

The Play Resource Distribution

The operation using a number-of-pools distribution and a pool size distribution will yield a play resource distribution. The play resource distribution is defined as follows:

$$T = X_1 + X_2 + \ldots\ldots + X_N = \sum_{i}^{M} X_i \qquad (5\text{-}24)$$

The play potential distribution is discontinuous at 0 as follows:

$$P\,[\,T = 0\,] = P\,[\,N = 0\,]$$

$$= P\,[\text{ no pools }]$$

$$= (\,1 - \Theta_g\,) + \Theta_g \sum_m (\,1 - \Theta_r)^m\,P\,(m) \qquad (5\text{-}25)$$

Now, for $t > 0$, the greater-than cumulative density function of T is

$$F_T(t) = P\,[\text{ play resource} > t\,]$$

$$= P\,[\,T > t\,]$$

$$= \sum_{n=1}^{m_1} F_n(t)\,P\,[\,N = n\,]$$

where $F_n(t) = P\,[\,X_1 + X_2 + \ldots + X_n > t\,]$.

The probability function of T is given by

$$f_n\,(t) = \begin{cases} P\,[\,N = 0\,], & \text{if } t = 0 \\ \sum_{n=1}^{m_1} f_n(t)\,P\,[\,N = n\,], & \text{if } t > 0. \end{cases}$$

where $f_n(t)$ is the probability density function of the convolution $X_1 + \ldots\ldots + X_n$ of n pool sizes.

The expectation and variance of T are:

$$E[\,T\,] = E[\,X\,] \times E[\,N\,] = \Theta_g \times \Theta_r \times E[\,M\,] \times E[\,X\,] \qquad (5\text{-}26)$$

$$\sigma_T^2 = \sigma_X^2 \times E\,[\,N\,] + (E\,[\,X\,])^2 \times \sigma_N^2 \qquad (5\text{-}27)$$

where $E\,[\,N\,]$ = mean of the number-of-pools distribution,
$E\,[\,X\,]$ = mean of the pool size distribution,
σ^2 = variance of the pool size distribution, and
σ_N^2 = variance of the number-of-pools distribution.

If X is lognormally distributed with μ and σ^2, then

$$E[T] = E[N] \exp(\mu + \sigma^2/2) \quad (5\text{-}28)$$
$$\sigma_T^2 = \exp(2\mu + \sigma^2)\,[\,E(N)\,(\exp(\sigma^2) - 1) + \sigma_N^2] \quad (5\text{-}29)$$

The uncertainty contained in the play resource distribution as measured by its variance is relatively insensitive to the uncertainty inherited from the prospect distribution. This can be examined by substituting the σ_N^2 of Equation (5-23) into Equation (5-29).

The play resource distribution is the superpopulation distribution of the geological model. The uncertainty contained in the distribution can be reduced if we have pool sizes and their ranks as we have discussed for the mature play. In frontier plays, we may not have the information necessary for the reduction of this type of uncertainty.

The play resource distributions for Cases I and II are listed in Table 5-9. The results are interpreted as follows:

1. Their means are identical, but Case II has a much larger standard deviation than Case I.

2. Case I suggests that there is about 10% chance that the play has no potential, while Case II implies there is about 45% chance that the play has no potential (as indicated by: 1 - probability of the first occurrence of play potential, e.g. 1.0 - 0.55).

3. From examination of Table 5-9, Case II has higher potential at the tail of the play resource distribution than does Case I. This is due to different interpretations of the geological factors as play or prospect levels of risk. For Case II, if source and preservation factors do exist in one prospect then they also exist in every prospect. This is the reason why Case II has a higher probability of having more potential (if the potential does exist) than does Case I.

Pool-Size-By-Rank

In frontier cases, pool-size-by-rank is normally obtained from operations of pool size and number-of-pools distributions. Because number-of-pools distribution is used in estimations of individual pool sizes, therefore, a probability of having at least r pools is provided. The results of the two cases are listed in Tables 5-10 and 5-11. These tables can be interpreted as follows:

1. The probabilities for having at least one pool, or two pools, and so on, are very different for the two cases. For example, the probabilities for at least one pool existing are 0.95 and 0.57, respectively, for Case I and II.

2. The sum of the products (of each individual pool size and its probability of existence) equals the mean of the play resource distribution.

3. The estimated pool sizes for Case II are much larger than those of Case I. This variability is inherited from the variances of the play resource distributions.

Generation of Reservoir Parameters

For economic analysis of petroleum resources, it is necessary to find the conditional distribution of the geological variables of Equation (5-1) for a given pool size x. This conditional distribution is also of interest to geologists. For example, the following question may be asked:

Given a pool size equal 714 MM bbl, what are the likely values for its pool area and net pay?

We assume the vector of the geological variables

$$\underline{Z} = (Z_1, Z_2, \ldots, Z_p)$$

associated with the pool size equation

$$X = c \times Z_1 \times Z_2 \times \ldots \times Z_p. \qquad (5\text{-}30)$$

has a multivariate lognormal distribution ($\underline{\mu}$, $\underline{\Sigma}$), where $\underline{\Sigma}$ is positive definite.

The conditional probability distributions for the reservoir parameters were computed for each given pool size in the conceptual play. Examples of the values at the 75th, 50th, and 25th upper percentiles are listed as follows:

Pool size (MM bbl)	Reservoir parameter	Upper percentiles 75	50	25
714	Area ($mile^2$)	35	58	81
	Reservoir thickness (ft)	108	187	331
	Porosity	0.11	0.14	0.19
	Trap fill	0.25	0.39	0.61
409	Area ($mile^2$)	27	46	77
	Reservoir thickness (ft)	82	144	249
	Porosity	0.10	0.14	0.18
	Trap fill	0.21	0.34	0.53

A larger pool size has a larger distribution for the area of closure, the reservoir thickness, the porosity, and the trap fill than for that of a small pool size. This phenomenon is the result of all the geological variables constrained by Equation (5-1). In fact, the conditional distribution of the same variables for a given pool size overlap to a certain extent, reflecting (i) the nature of the irregularities, e.g. small pool size with excellent porosity, and/or (ii) little variations in variables such as porosity. This type of information can then be used for calculating productivity.

Estimated conditional pool area distributions can provide information for calculating the number of wells required for development of an undiscovered pool.

Chapter Six

PROCEDURES AND FEEDBACK MECHANISMS

> Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise.
>
> John W. Tukey

PROCEDURES

A basin or subsurface study is required as the first step in petroleum resource evaluation. Types of data required for petroleum resource evaluations are listed as follows:

Reservoir data - pool area, net pay, porosity, water saturation, oil or gas formation volume factor, in-place volume, recoverable reserves, temperature, pressure, density, recovery factor, gas composition, and so on,

Well data - general information, formation depth, lithology, drill-stem tests, core, gas and fluid analysis, and mechanical logs,

Geochemical data - types of source rocks, burial history, and maturation history, and

Geophysical data - prospect maps.

Well data will assist us to construct structural contour, isopach, lithofacies, porosity and other types of maps. Geophysical data will assist us to construct number-of-prospects distributions and also provide information for risk analysis. The number of dry holes and the reasons why they fail will provide information for estimating the marginal probability of each geological factor. Chronostratigraphic and organic maturation data will assist us in defining the basin's burial and thermal history. All these data will be used to identify a play and its geographic boundary. Furthermore, reservoir and well data retrieved within the geographic play boundary provide information needed to compile an exploration time series for evaluations of mature plays.

In the cases where we do not have enough information to quantify every aspect that we need, we can apply experience gained from other basins, or compile all work completed by previous workers. This type of compilation or comparative study can provide useful information for resource evaluation.

The steps required by the PETRIMES methods are briefly described below:

1. Identification of a play definition and its geographic boundary.

A play has both geographic and stratigraphic limits and is confined to a basin or part of a basin, a structural unit or part of it and is also confined to one or more formations.

2. Identification of geological variables

If it is a conceptual play, what types of variables should be used in Equation 5-1? If it is mature play, to what extent may the discovery process model be applied?

3. Collection of relevant data

For mature plays, an exploration times series consisting of successful and failed exploratory wells, and discoveries and recoveries from drill-stem tests is constructed as the basic input for evaluations. For conceptual or immature plays, all relevant data and information should be compiled. Probability distributions used in Equation 5-1 are then constructed.

4. Data analysis and validations

(1) For mature plays

(i) Plot the pool size, pool area, and other variables by discovery sequence to graphically visualize past exploration history using DSEQ.

(ii) Plot the pool size, pool area, net pay and other variables on log probability paper to graphically examine, using LPLT, the assumption of log-normality and possible mixed populations.

(iii) Cross-plot the geological variables using XPLT to examine the correlation between variables and to detect abnormalities in the data set.

(iv) Use exploratory data analysis to graphically examine whether the play definition is adequate using LPLT, BPLT, HPLT, SPLT, for example.

(2) For conceptual plays

Probability distributions are constructed from expert opinion and the results of comparative studies. The rules for constructing these types of distributions are listed in Chapter 5. These distributions will be approximated by the family of lognormal distributions. In order to verify the approximations, both raw data and approximated distributions should be plotted by CPLT on the same plot.

5. Evaluations

(1) Estimation of pool size distribution

For mature plays, pool size distribution can be estimated using lognormal (LDSCV) or nonparametric (NDSCV) discovery process models or the matching process (MATCH) to estimate pool size distribution. For conceptual plays, pool size distributions may be estimated by using Equation 5-1 (PPSD).

(2) Estimation of number-of-pools distribution

Number-of-pools distribution can be computed by applying the prospect and/or prospect level risks and the number-of-prospects distribution. For conceptual plays, both play and prospect level risks should be presented, while for mature plays only prospect level risks are presented.

(3) Estimation of pool-size-by-rank

(i) Pool-size-by-rank can be computed by the evaluation components, MATCH, or GPSRK for a fixed value of N, or by the evaluation components, or PSRK if N is a random variable.

(ii) The matching process can be accomplished using either an evaluation component (MATCH) or a graphic component (RPLT).

(iii) Individual pool sizes can be further conditioned to the given pool ranks using the PSDR evaluation component.

6. Evaluations of play resource or potential distribution and others

(1) Play resource distribution can be estimated using the PSUM evaluation component. A play potential distribution can be estimated by conditioning the play resource distribution to pool ranks by using the PPDR component.

(2) Ratios between pool sizes can be estimated using the PSRO evaluation component.

(3) Reservoir parameters for each pool size can be estimated from the RVGN component.

(4) Basin, geological province or country resource distributions can be estimated using the PSUM evaluation component.

For each evaluation, there is a report consisting of play map, all input data, final estimates, figures, and other statistics as listed in Table 6-1.

FEEDBACK MECHANISMS

Feedback is essential to any assessment. Figure 6-1 shows different levels of feedback during petroleum assessment. PETRIMES provides various feedback mechanisms.

Can we predict the present situation?

For mature plays, it is highly recommended that the discoveries be divided into two subsets for the following reasons: (1) to examine whether we can predict the second set from the first sample set and (2) to find the adequate prediction interval.

The final estimates must be validated by one or more of the following procedures:

1. Examination of the remaining largest pool size against geological models or exploration concepts.

2. Re-estimation of the undiscovered pool sizes by entering the discovery pool sizes and their ranks.

3. Retrospective study (Lee and Wang, 1986).

Has the largest pool been discovered?

For the Beaverhill Lake play example, where the largest pool appears to have been discovered, geologists may still ask, "What would the largest pool size be if the discovered largest pool is assumed to be the second largest in a play?" Our method allows us to analyze this question.

Take the Beaverhill Lake play as an example. The result of the computation is that given that the largest discovered pool (290 10^6 m^3) is actually the second largest pool, then the prediction interval for the size of the possible largest pool ranges from 320 10^6 to 4129 10^6 m^3. Furthermore, the area of the pool required is as large as the largest pool present today. With this information, we can address the question, "Have we overlooked the largest pool of this play?"

This type of feedback mechanism, allows us to challenge the underlying geological concept or to validate our input data, and is one of the essential features of our evaluation system.

Pool Size conditional to play resource

Individual pool size and number of pools can be estimated for a given play resource. This technique can be used as a feedback mechanism to resolve discrepancies between different estimates, and to validate basic input factors such as exploration risk, number of pools, and pool size distribution.

Having computed the play resource distribution, one measure of the resource is the mean of the distribution. However, geologists might choose a value other than the mean from the distribution as a point estimate of the resource. Figure 6-2 displays the number of pools distributions which are conditional to various play resources such as 646 MM bbl, 2.7 B bbl, and 7.8 B bbl of oil.

Chapter Seven

CONCLUDING REMARKS

....
"Would you tell me please which way I ought to go from here?" "That depends a good deal on where you want to get to," said the Cat. .....

Lewis Carroll

All petroleum resource estimation methods involve a learning process that is characterized by an interactive loop between the geological and statistical models, and their feedback mechanisms. Geological models represent natural populations and are the basic units for petroleum resource evaluation. Statistical models include the distributions for pool size and number-of-pools, and can be estimated from somewhat biased exploration data.

Methods for petroleum resource evaluations have been developed using different geological perspectives. Each of them can be applied to a specific case. When we consider a method to use, the following aspects should be examined:

1. Types of estimates. What types of estimates does the method provide (aggregate estimates vs. pool size estimates)? Do the types of estimates fulfil our economic needs and other requirements?

2. Assumptions required. We must study what specific assumptions must be made and what role they play in the process of estimation.

3. Types of information required. Some methods can only incorporate certain types of information, while other methods can incorporate as much information as is available.

4. Feedback mechanisms. What types of feedback mechanism does the method offer?

PETRIMES is based on a probabilistic framework that uses the superpopulation concept, the discovery process model, and optional use of lognormal distributions. The basic input is an exploration time series. Other types of data are used in different stages during the evaluation process. The system can be applied to both mature and frontier evaluations.

REFERENCES

Aitchison, J. and Brown, J. A. C., 1973, The lognormal distribution with special reference to its uses in economics: Cambridge University Press, England, 176 p.

Baecher, G. B., 1979, Subjective sampling approaches to resource estimation: in M. Grenon, ed., First IIASA Conference on methods and models for assessing energy resources; Pergamon Press, New York, p.186-209.

Barouch, E. and Kaufman, G., 1977, Estimation of undiscovered oil and gas: Proceedings of Symposia in Applied Mathematics, v.21, p.77-91.

Bickel, P. J. and Doksum, K. A., 1977, Mathematical statistics: Basic ideas and selected topics: Holden-Day Inc., 492 p.

Cassel, C. M., Sarndal, C. E. and Wretman, J. H., 1977, Foundations of inference in survey sampling: John Wiley & Sons, New York, 192 p.

Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A., 1983, Graphic methods for data analysis: Wadsworth International Group, Belmont, California, p. 191-242.

Cochran, W. G., 1939, The use of analysis of variance in enumeration by sampling: Journal of American Statistical Association, v. 34, p. 492-510.

Drew, L. J., Schuenemeyer, J. H. and Root, D. H., 1980, Petroleum resource appraisal and discovery rate forecasting in partially explored region-an application to the Denver Basin: U.S. Geological Survey Professional Paper 1138-A, 13 p.

Drew, L. J., Schuenemeyer, J. H. and Bawiec, W. J., 1982, Estimation of the future rates of oil and gas discoveries in the Gulf of Mexico, U. S. Geological Survey Professional Paper 1252, 26 p.

Energy, Mines, and Resources Canada, 1977, Oil and natural gas resource of Canada, 1976: Report EP 77-1, 76 p.

Energy Resources Conservation Board, 1989, Alberta's reserves of crude oil, gas, natural gas liquids, and sulphur: Energy Resources Conservation Board, Alberta, Canada, Table 2-5.

Gehman, H. M., Baker, R. A. and White, D. A., 1981, Assessment methodology - An industry viewpoint: in Assessment of Undiscovered Oil and Gas, Proceedings of the Seminar, United Nations ESCAP, CCOP Technical Publication 10, p. 113-121.

Grenon, M., 1979, First IIASA Conference on methods and models for assessing energy resources: Pergamon Press, New York, 605 p.

Griffin, D. L., 1965a, The Devonian Slave Point, Beaverhill Lake, and Muskwa Formations of northeastern British Columbia and adjacent areas: British Columbia

Department of Mines and Petroleum Resources, Bull. No. 50, 90 p., 20 plates.

Griffin, D. L., 1965b, The facies front of the Devonian Slave Point-Elk Point sequence in northeastern British Columbia and the Northwest Territories: J. Canadian Petroleum Technology, January-March, p. 13-22.

Hemphill, C. R., Smith, R. I. and Szabo, F., 1968, Geology of Beaverhill Lake Reefs, Swan Hills area, Alberta: AAPG Memoir 14, p. 50-90.

Haun, J. D., 1975, Methods of estimating the volume of undiscovered oil and gas resources: AAPG Study in Geology, no. 1, 206 p.

Kaufman, G. M., 1963, Statistical decision and related techniques in oil and gas exploration: Prentice-Hall, Inc., N.J., 307 p.

Kaufman, G. M., 1965, Statistical analysis of the size distribution of oil and gas fields: Symposium on Petroleum Economics and Evaluation, Society of Petroleum Engineers, p. 109-124.

Kaufman, G. M., 1986, Finite Population Sampling Methods for Oil and Gas Resource Estimation: in AAPG Studies in Geology, no. 21, p. 43-53.

Kaufman, G. M., Balcer, Y. and Kruyt, D., 1975, A probabilistic model of oil and gas discovery: in J.D. Haun, ed., Methods of Estimating the Volume of Undiscovered Oil and Gas Resources, AAPG Study in Geology, no. 1, p. 113-142.

Lee, P. J., Qin, Ruo-Zhe and Shi, Yan-Min, 1989, Conditional Probability Analysis of Geological Risk Factors: in F. Agterberg and G. Bonham-Carter ed., Statistical Applications in the Earth Sciences; GSC Paper 89-9, Ottawa Canada.

Lee, P. J. and Wang, P. C. C., 1983a, Probabilistic formulation of a method for the evaluation of petroleum resources: Mathematical Geology, v. 15, no. 1, p. 163-181.

Lee, P. J. and Wang, P. C. C., 1983b, Conditional analysis for petroleum resource evaluations: Mathematical Geology, v. 15, no. 2, p. 353-365.

Lee, P. J. and Wang, P. C. C., 1984, PRIMES: A petroleum resources information management and evaluation system: Oil & Gas Journal, October 1, p. 204-206.

Lee, P. J. and Wang, P. C. C., 1985, Prediction of oil or gas pool sizes when discovery record is available: Mathematical Geology, v. 17, no. 2, p. 95-113.

Lee, P. J. and Wang, P. C. C., 1986, Evaluation of petroleum resources from pool size distribution: in Oil and Gas Assessment Methods and Applications, D.D.Rice (ed), Studies in Geology, no. 21, p. 33-42, AAPG publication.

Lee, P. J., Eggen, S. S. and Vann, I. R., 1990 (in press), Use of a probability approach in assessing the petroleum resources of the northern North Sea: IUGS publication.

Lee, P. J., Nair, V. N. and Wang, P. C. C., 1988, Estimating and modeling the superpopulation distributions in petroleum resource evaluations: Manuscript in preparation.

Lee, P. J. and Wang, P. C. C., 1990 (in press), Petroleum resource evaluation concepts: IUGS Publication.

Masters, C. D., 1985, Petroleum resource assessment: International Union of Geological Sciences, Publication no. 17, 157 p, Ottawa Canada.

Meisner, J. and Demirmen, F., 1981, The creaming method: A Bayesian procedure to forecast future oil and gas discoveries in mature exploration provinces: Journal of Royal Statistical Society, Series A, v. 144, part 1, p. 1-13.

Nair, V. N., 1984, On the behaviour of some estimators from probability plots: Journal American Statistical Association, v. 79, p. 823-836.

Neyman, J. and Scott, E. L., 1971, Outlier proneness of phenomena and of related distributions: Optimizing Methods in Statistics, Academic Press, Inc., New York and London, p. 413-430.

Rice, D. D., 1986, Oil and gas assessment - Methods and applications: AAPG Studies in Geology n. 21, 267 p.

Roy, K. J., 1979, Hydrocarbon assessment using subjective probability and Monte Carlo methods: in M. Grenon, Ed., First IIASA Conference on methods and models for assessing energy resources; Pergamon Press, New York, p. 279-290.

Schuenemeyer, J. H. and Drew, L. J., 1983, A procedure to estimate the parent population of the size of oil and gas fields as revealed by a study of economic truncation: Mathematical Geology, vol. 15, no. 1, p. 145-161.

Velleman, P. F. and Hoaglin, D. C., 1981, Applications, basic, and computing of exploratory data analysis: Duxbury Press, A Division of Wadsworth, Inc. 354p.

Wang, P. C. C. and Nair, V. N., 1988, Statistical analysis of oil and gas discovery data: in Quantitative Analysis of Mineral and Energy Resources, F. Chung, (ed), p. 199-214.

White, D. A. and Gehman, H. M., 1979, Methods of estimating oil and gas resources: AAPG Bull. v. 63, no. 12, p. 2183-2192.

White, D. A., 1980, Assessing oil and gas plays in facies-cycles wedge: AAPG Bull., v. 64, no. 8, p. 1158-1178.

Willams, G. K., 1984, Some musings on the Devonian Elk- Point basin, Western Canada: Bull. Canadian Petroleum Geologists, v. 32, n. 2, p. 216-232.

Wilson, J. L. and Jordan, C., 1983, Middle shelf environment: in Scholle, P. A., Bebout, D. G., and Moore, C. H., ed. Carbonate Depositional Environments: AAPG Memoir 33, Figure 64, p. 335.

Figure 2-1. Examples of geological models (After Wilson and Jordan, 1983). Each model may be defined as a basic unit for evaluation.

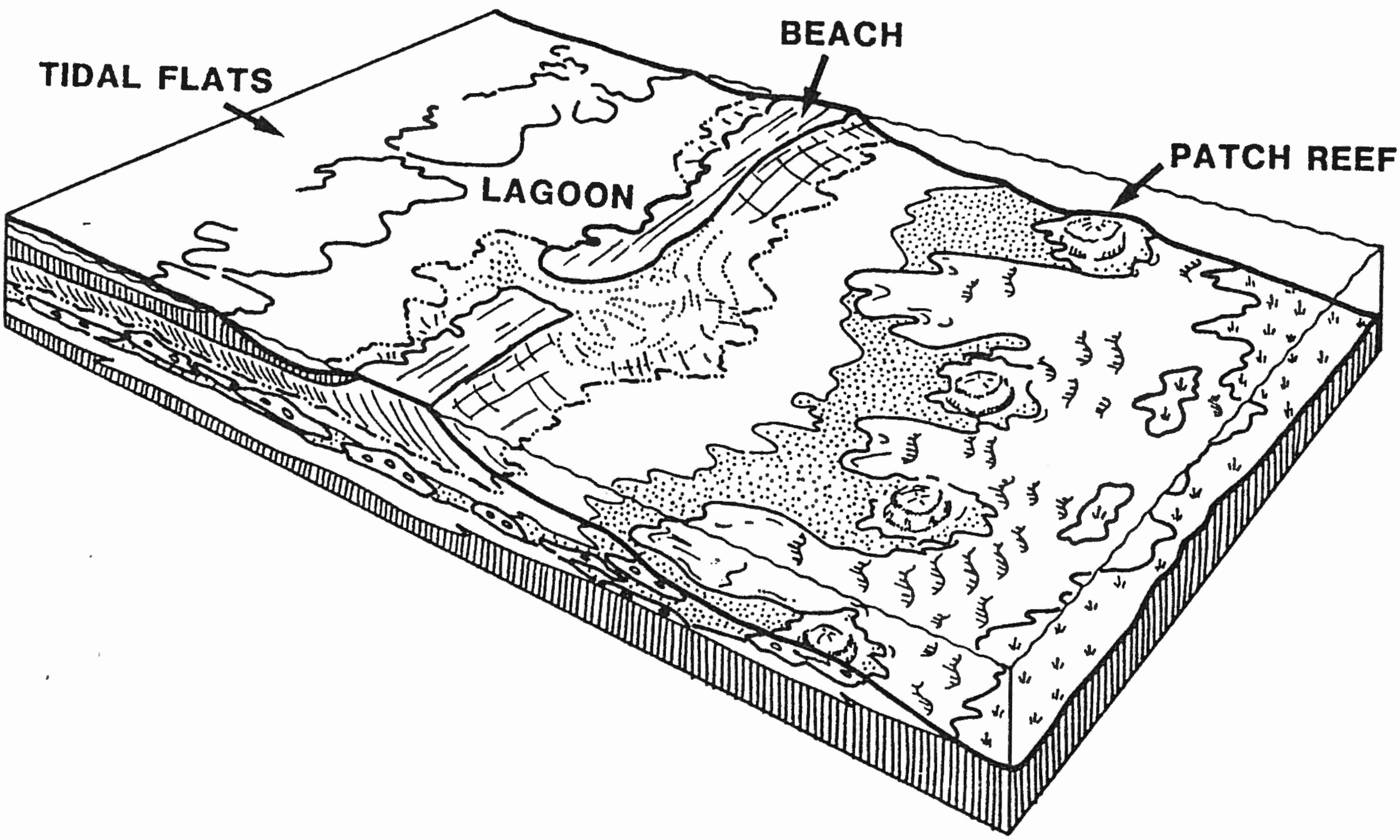


Figure 2-2. Examples of play definitions: Leduc reefs (left), Middle Devonian reefs (right) from the Western Canada Basin.

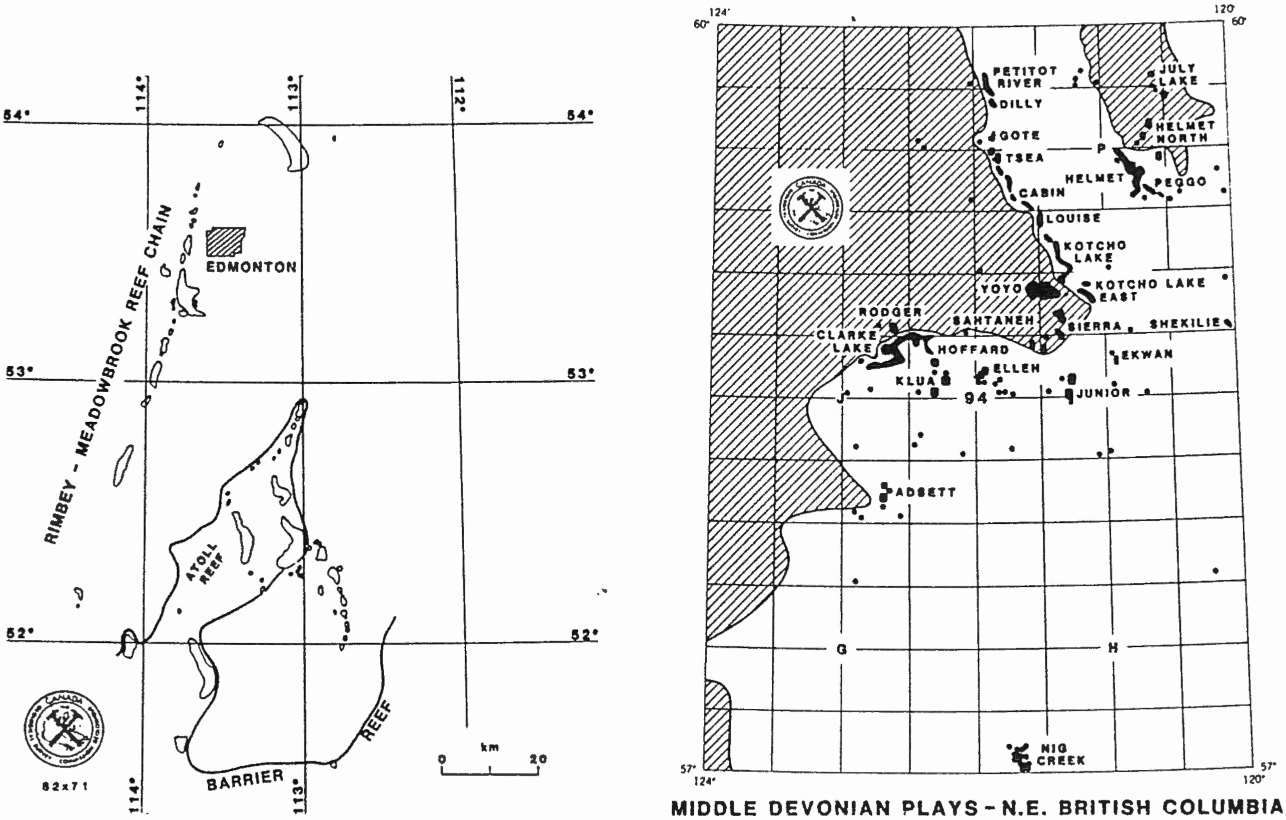
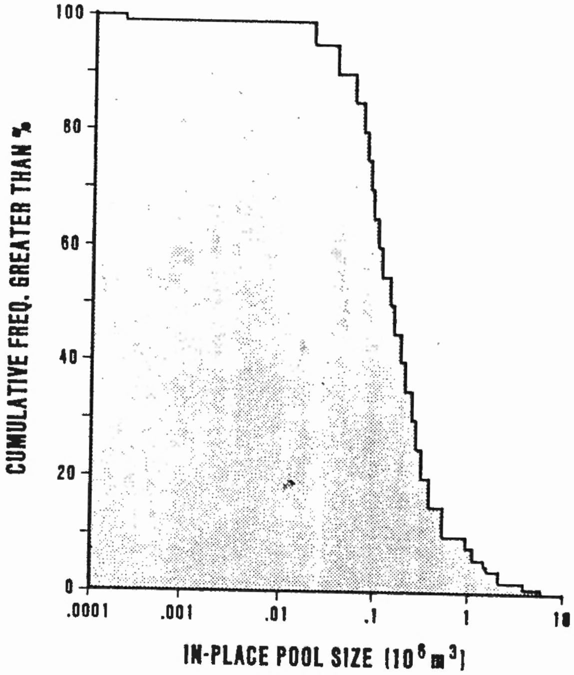


Figure 2-3. Examples of probability distributions: discrete distribution (top), continuous distribution (bottom).



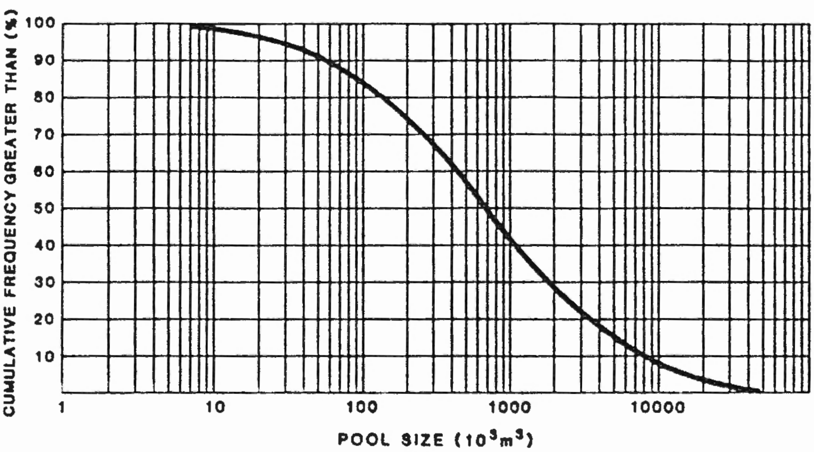


Figure 2-4. An example of histogram for the variable of porosity, the Mannville Formation (top), an example of LESS than cumulative probability distribution (lower left), and an example of GREATER than cumulative probability distribution (lower right).

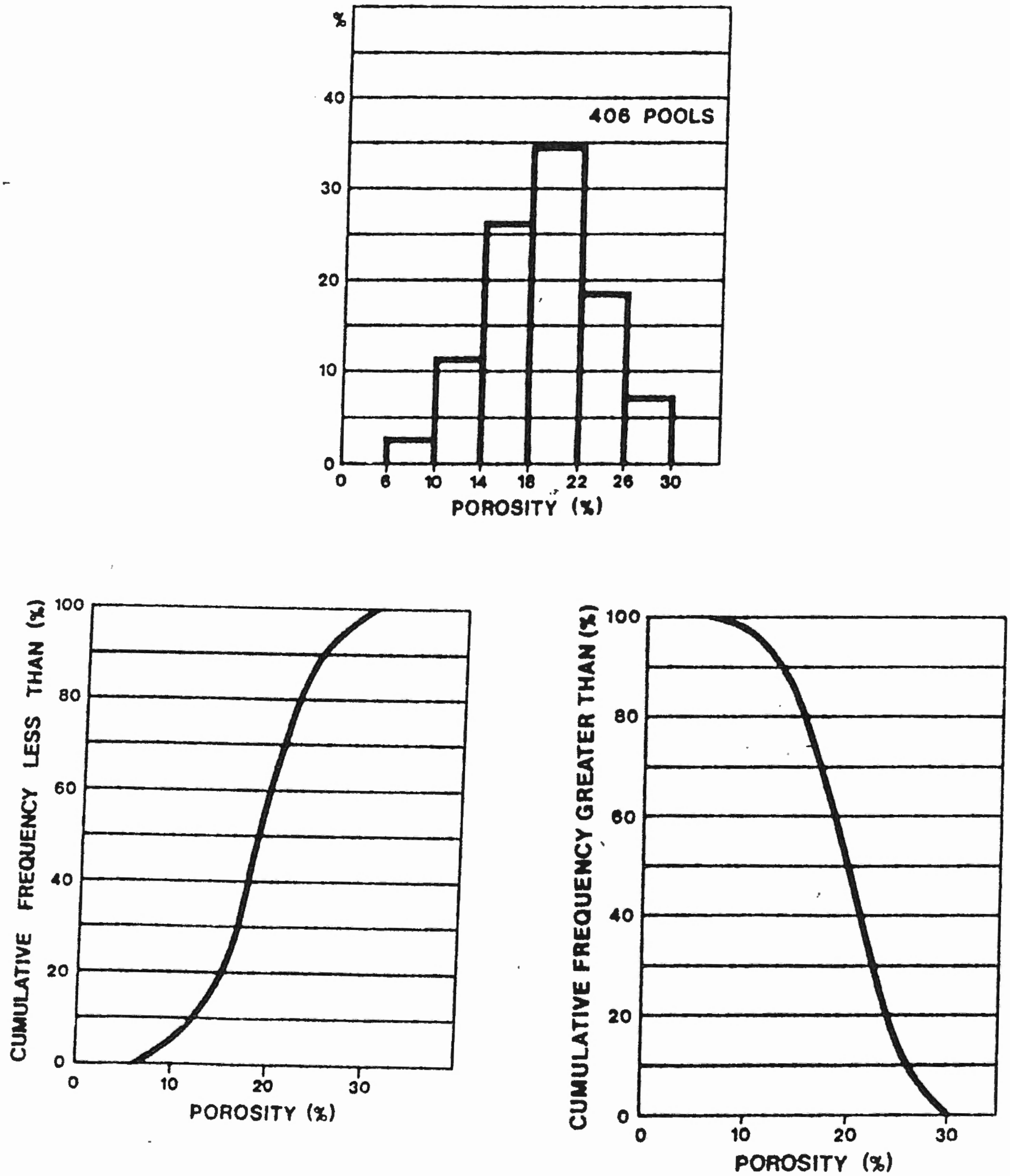
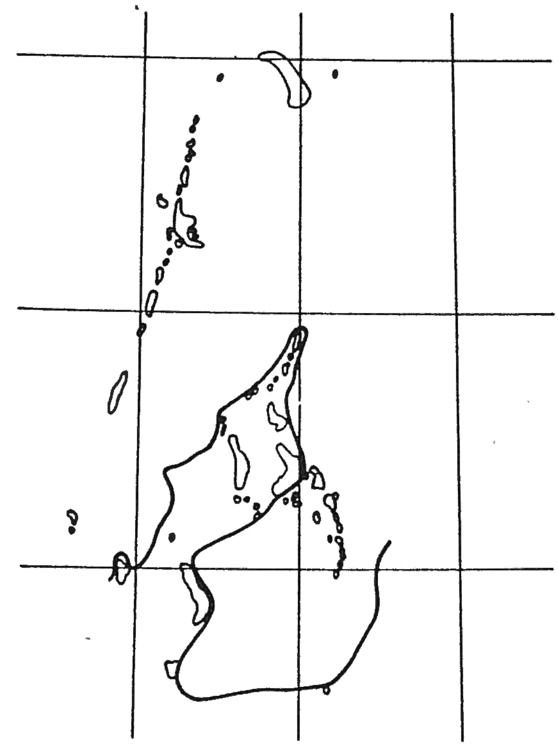


Figure 2-5. Concepts in discovery process as a sampling process.

POPULATION OR PLAY



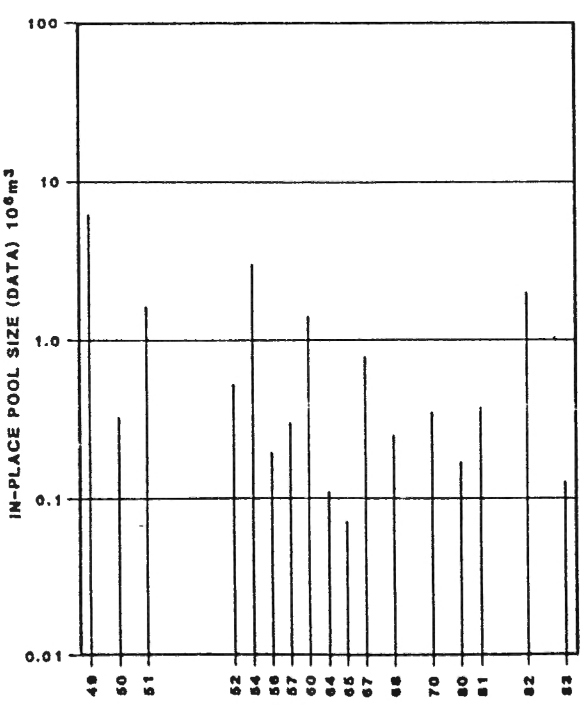
D SCOVERY PROCESS MODEL

CHARACTERIZATIONS:

1. Sampling from the play with N pools proportional to pool size raised to a power β.

2. Sampling without replacement.

$$\frac{X_1^\beta}{X_1^\beta+X_2^\beta+X_3^\beta+\ldots+X_N^\beta} \times \frac{X_2^\beta}{X_2^\beta+X_3^\beta+\ldots+X_N^\beta}$$



ESTIMATION

ASSUMPTIONS:

1. Lognormal distribution, or distribution-free,

2. Superpopulation model, or finite population model,

3. Discovery process model.

Figure 2-6. Statistical concepts employed by PETRIMES.

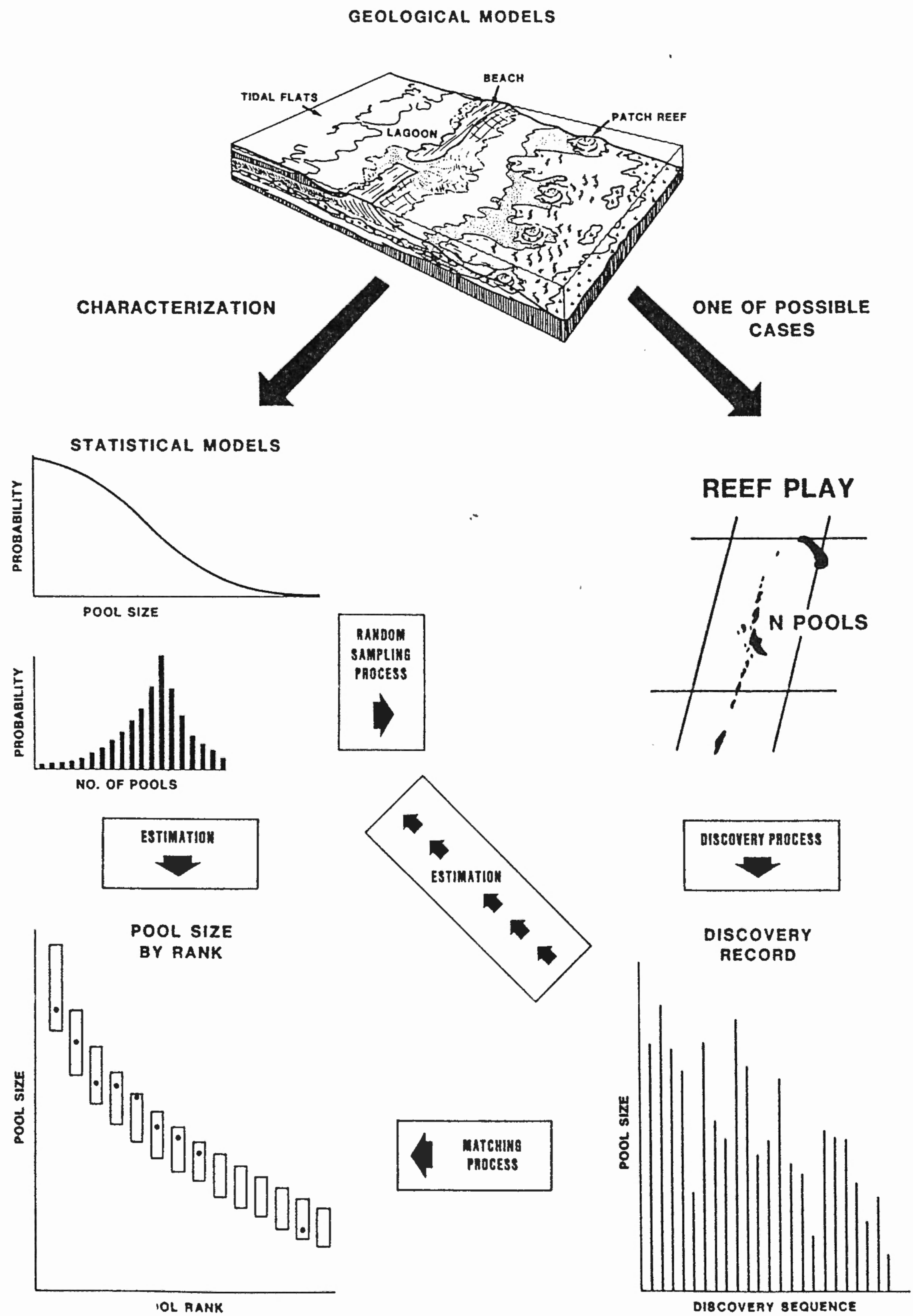


Figure 3-1. Facies map for the Beaverhill Lake play.

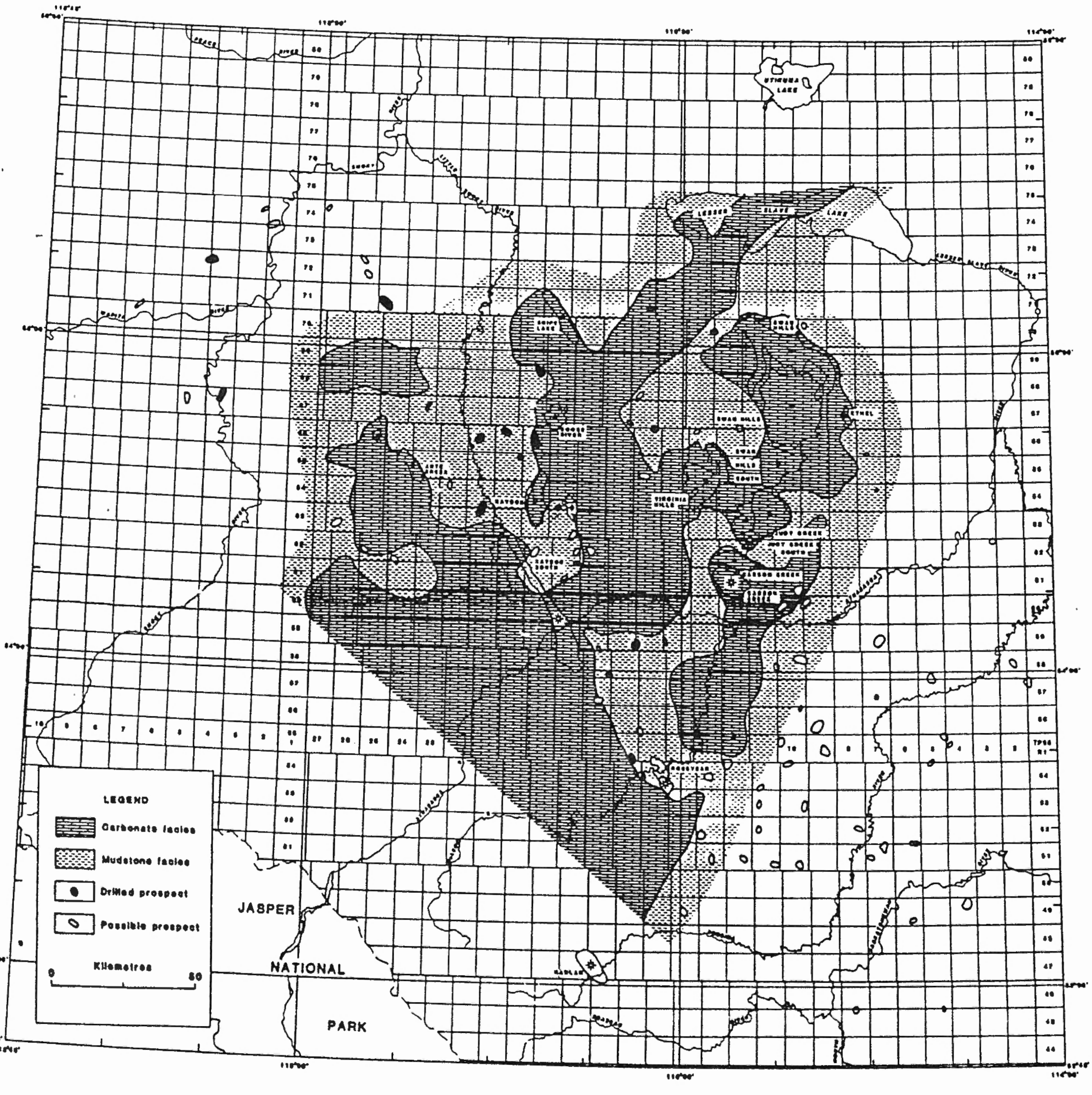


Figure 3-2. Exploration time series for the Beaverhill Lake play.

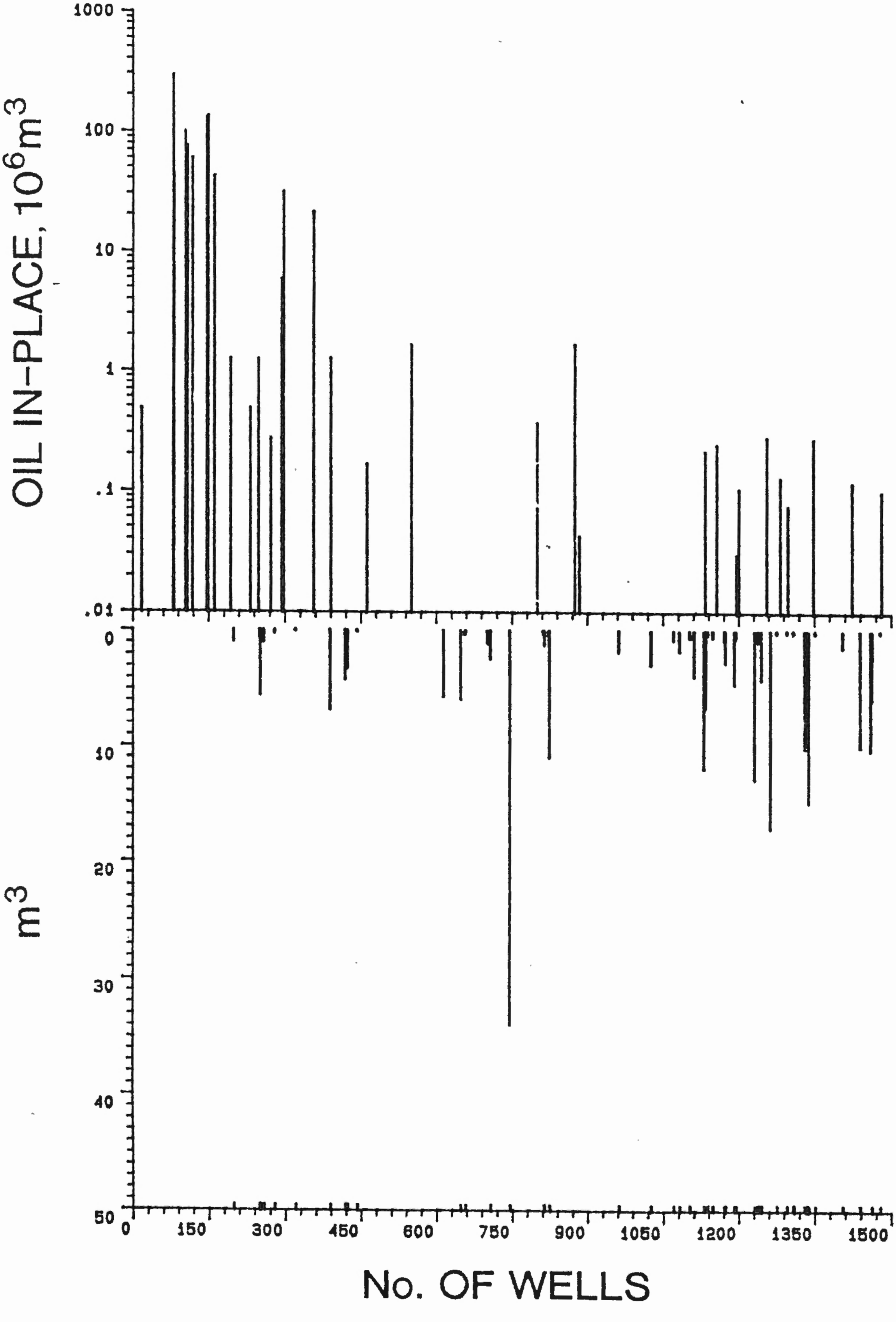


Figure 3-3. The boxplots for the in-place pool size of several plays in the Western Canada Basin.

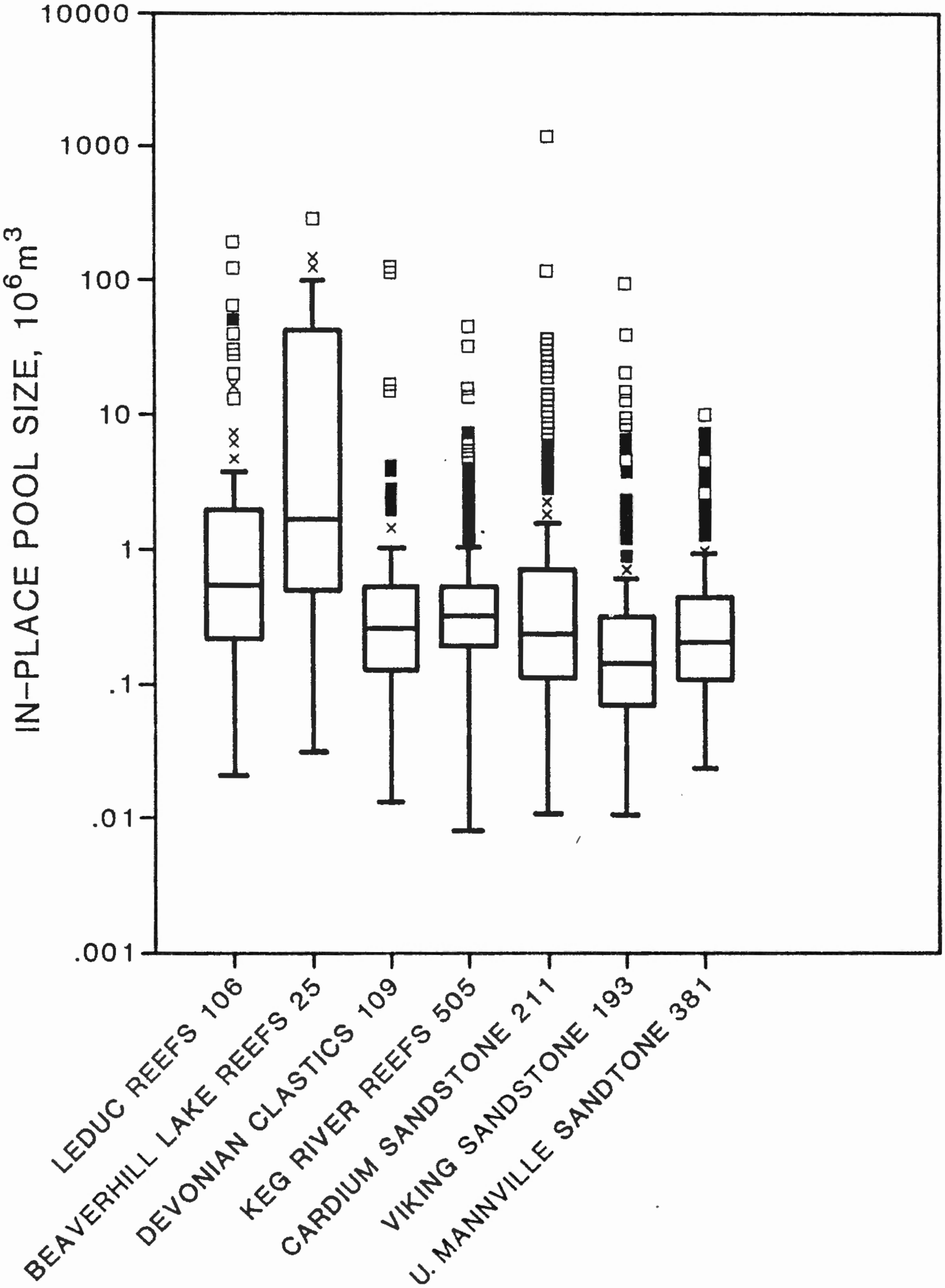


Figure 3-4. Log-linear associations showing: the negative correlation between the pool area and average net pay variables of the Zama reef play (left), and the positive correlation between the pool area and average net pay of the Beaverhill Lake play (right) from the Western Canada Basin.

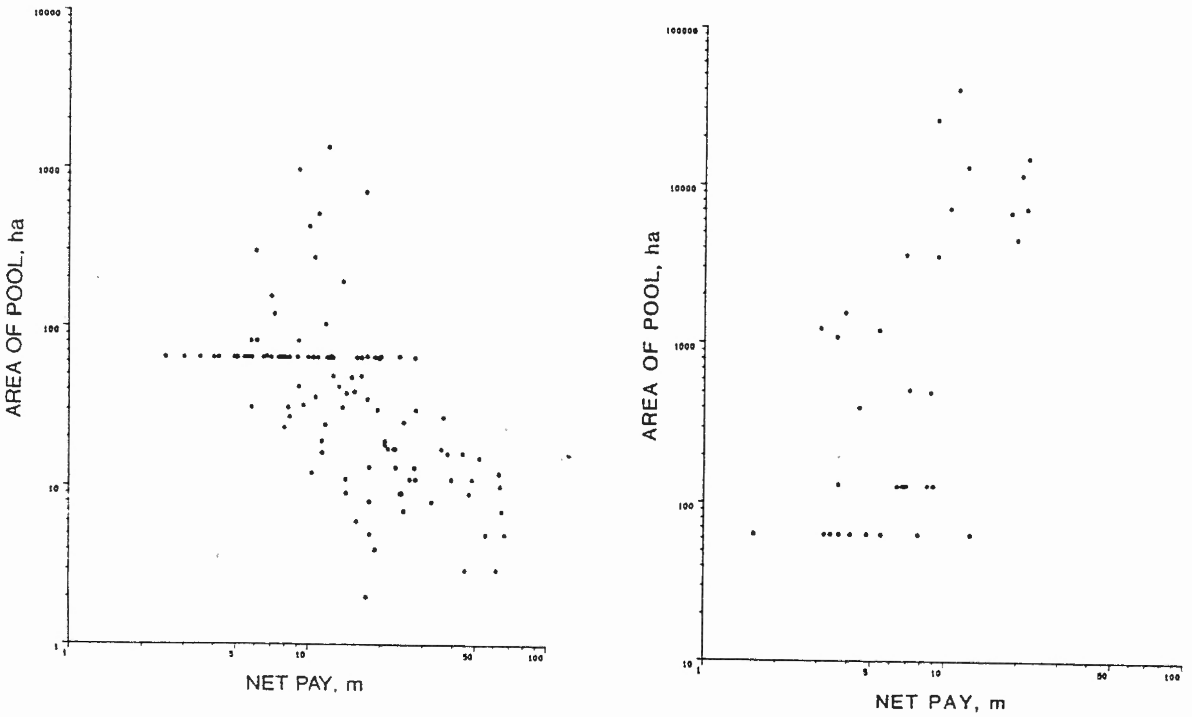
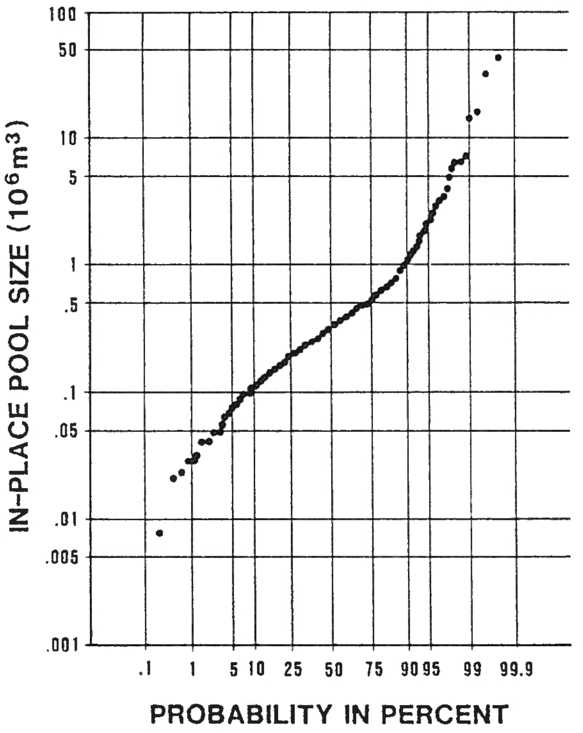


Figure 3-5. Log probability plot for the Keg River reefs from the Black Creek basin (left) and log probability plot for the Keg River reefs of the Rainbow subbasin of the Black Creek basin (right).



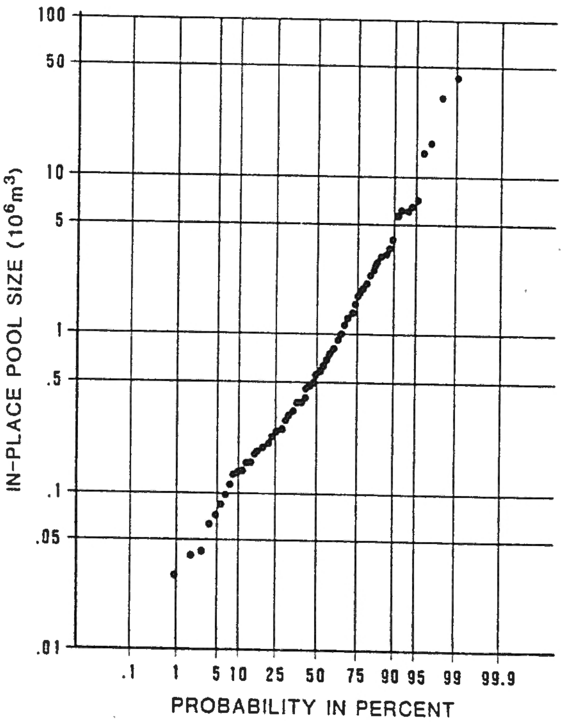
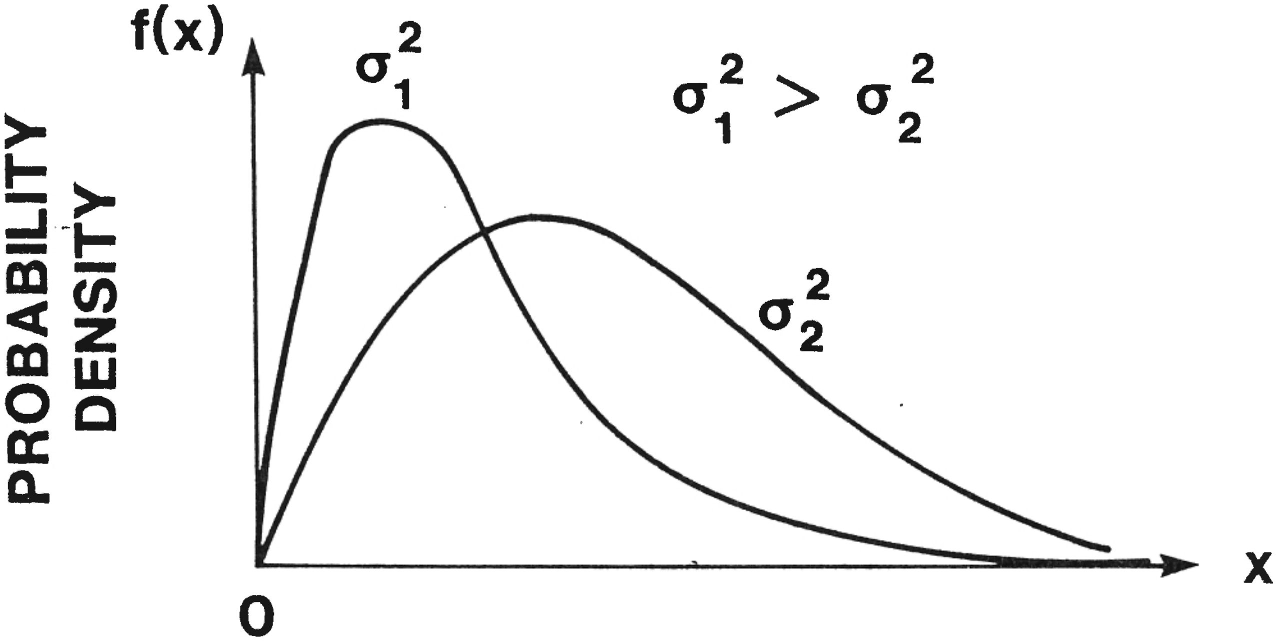


Figure 3-6. Examples of the shapes of lognormal distributions.



$$f(x) = \frac{1}{X\sigma\sqrt{2\pi}} \text{ EXP } [-{}^1/_2 (\ln x - \mu)^2 / \sigma^2]$$

μ = mean of the log transformed pool size

σ^2 = variance of the log transformed pool size

x = pool size

Figure 3-7. Lognormal approximation of: net pay for the Devonian clastics (top), porosity of the Lower Mannville play (lower left), and the Cardium marine sandstone play (lower right) from the Western Canada Basin.

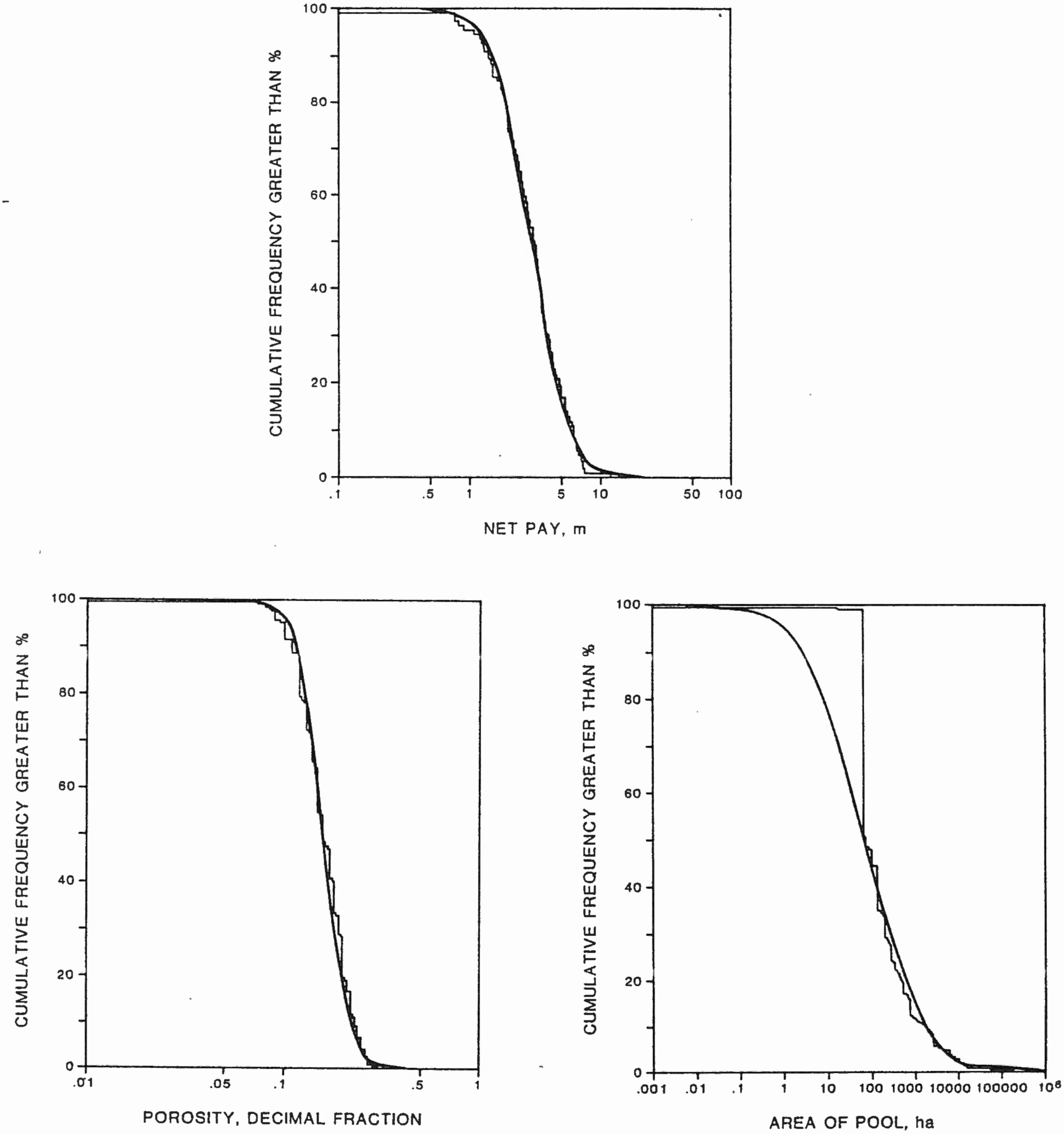
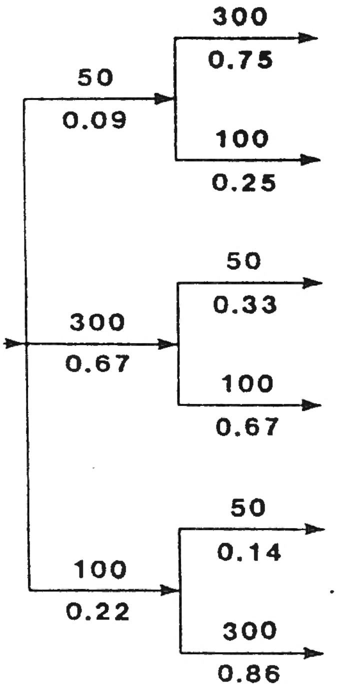


Figure 3-8. A tree showing the probabilities for different discovery sequences.



$$\frac{50}{50+300+100} \times \frac{300}{300+100} = 0.09 \times 0.75 = 0.07$$

$$\frac{50}{50+300+100} \times \frac{100}{300+100} = 0.09 \times 0.25 = 0.02$$

$$\frac{300}{50+300+100} \times \frac{50}{50+100} = 0.67 \times 0.33 = 0.22$$

$$\frac{300}{50+300+100} \times \frac{100}{50+100} = 0.67 \times 0.67 = 0.45$$

$$\frac{100}{50+300+100} \times \frac{50}{50+300} = 0.22 \times 0.14 = 0.03$$

$$\frac{100}{50+300+100} \times \frac{300}{300+50} = 0.22 \times 0.86 = 0.19$$

$W_N = (100, 300, 50)$, $N = 3$, $n = 2$.

Figure 3-9. Geological interpretations based on the discovery process model.

$$\frac{N!}{(N-n)!}\underbrace{\prod_{j=1}^{n} f_\theta(X_j)}\ \underbrace{E_\theta\left[\prod_{j=1}^{n}\frac{X_j^\beta}{b_j+Y_{n+1}^\beta+\cdots\cdots+Y_N^\beta}\right]}$$

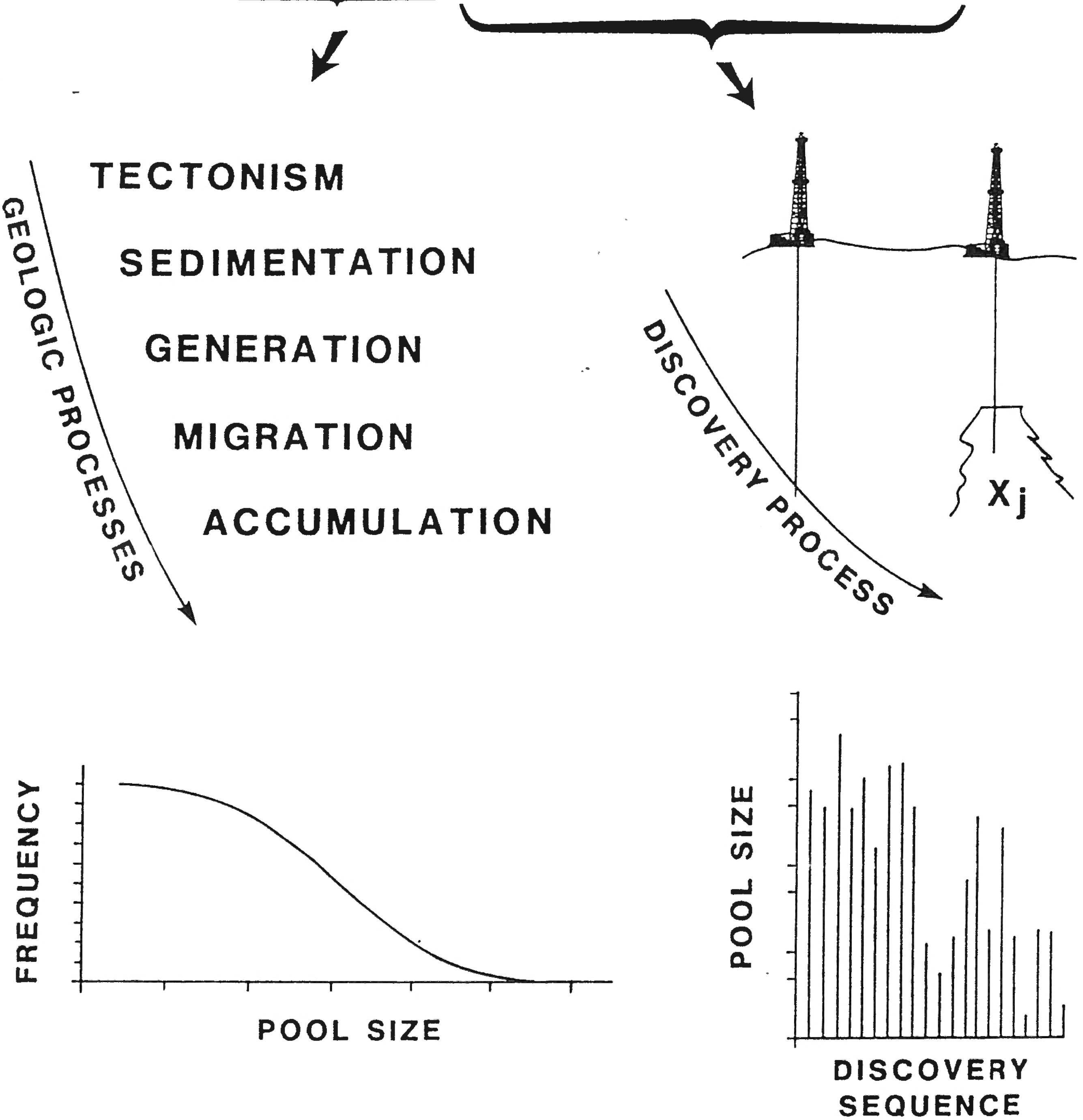


Table 3-1. Estimates derived by the lognormal discovery process model for the simulated population given μ = -4, σ^2 = 20, N = 200 and β = 0.3 for different values of N.

n	N	β^	μ^	σ^2^	Log L
40	50	0.01	-1.47, -0.16, 1.15;	9.95, 17.03, 25.99;	-202.21
	75	0.18	-2.91, -1.39, 0.13;	10.33, 18.91, 30.08;	-201.81
	100	0.23	-4.02, -2.26, -0.49;	10.46, 20.08, 32.80;	-201.60
	125	0.26	-4.89, -2.89, -0.90;	10.50, 20.85, 34.74;	-201.46
	150	0.28	-5.60, -3.40, -1.19	10.49, 21.42, 36.20;	-201.36
	175	0.29	-6.20, -3.81, -1.42;	10.47, 21.86, 37.39;	-201.28
	200	0.31	-6.72, -4.16, -1.61;	10.45, 22.22, 38.38;	-201.22
	225	0.32	-7.17, -4.47, -1.77;	10.43, 22.53, 39.23;	-201.17
80	100	0.22	-2.46, -1.55, -0.64;	11.13, 16.60, 23.15;	-363.38
	125	0.26	-3.39, -2.37, -1.36;	11.81, 18.30, 26.20;	-363.41
	150	0.29	-4.12, -3.00, -1.87;	12.18, 19.42, 28.34;	-363.49
	175	0.30	-4.75, -3.50, -2.25;	12.41, 20.24, 29.99;	-363.58
	200	0.32	-5.30, -3.93, -2.56;	12.57, 20.91, 31.35;	-363.67
	225	0.33	-5.78, -4.29, -2.81;	12.68, 21.45, 32.50;	-363.75
	250	0.34	-6.20, -4.61, -3.02;	12.76, 21.89, 33.49;	-363.82
150	160	0.22	-2.73, -2.12, -2.73;	10.72, 13.92, 17.52;	-578.76
	170	0.24	-3.01, -2.39, -1.78;	11.24, 14.74, 18.71;	-577.88
	180	0.26	-3.27, -2.64, -2.01,	11.65, 15.41, 19.69;	-577.46
	190	0.27	-3.51, -2.86, -2.21;	11.99, 15.99, 20.56;	-577.23
	200	0.28	-3.73, -3.07, -2.41;	12.29, 16.50, 21.33;	-577.12
	210	0.29	-3.94, -3.26, -2.58;	12.56, 16.97, 22.04;	-577.08
	220	0.29	-4.14, -3.44, -2.75;	12.79, 17.39, 22.69;	-577.08
	230	0.30	-4.33, -3.62, -2.90;	13.00, 17.77, 23.29;	-577.11
	240	0.31	-4.51, -3.78, -3.04;	13.20, 18.14, 23.86;	-577.15
	250	0.32	-4.69, -3.93, -3.18;	13.37, 18.47, 24.39;	-577.21
	300	0.34	-5.46, -4.60, -3.75;	14.07, 19.83, 26.63;	-577.57
	350	0.36	-6.11, -5.15, -4.19;	14.57, 20.93, 28.44;	-577.96

Note: The symbols, β^, μ^, and σ^2^ indicate the estimated values of their population values.

Table 3-2. Estimates derived by the lognormal discovery process model for the simulated population given $\mu = -4$, $\sigma^2 = 20$, N = 200 and $\beta = 0.6$ for different values of n.

n	N	β^	μ^	σ^2^	Log L
20	25	0.40	0.05, 1.32, 2.59;	3.61, 8.41, 15.22;	-111.99
	50	0.40	-1.89, -0.12, 1.90;	3.40, 9.86, 19.70;	-110.42
	75	0.46	-3.18, -0.92, 1.33;	3.22, 10.56, 22.14;	-110.32
	100	0.49	-4.12, -1.47, 1.19;	3.06, 11.00, 23.85;	-110.29
	125	0.52	-4.86, -1.87, 1.11;	2.93, 11.31, 25.15;	-110.28
	150	0.53	-5.47, -2.20, 1.08;	2.82, 11.55, 26.20;	-110.27
	175	0.55	-5.98, -2.46, 1.06;	2.71, 11.73, 27.08;	-110.28
	200	0.56	-6.43 -2.69, 1.06;	2.63, 11.89, 27.83;	-110.20
	225	0.57	-6.83, -2.88, 1.06;	2.55, 12.02, 28.50;	-110.29
	250	0.58	-7.19, -3.06, 1.07;	2.48, 12.13, 29.08;	-110.31
40	50	0.40	-1.01 -0.07, 0.87;	5.11, 9.21, 14.50;	-192.17
	75	0.47	-2.46, -1.27, -0.09;	5.61, 11.30, 18.96;	-191.78
	100	0.51	-3.54, -2.09, -0.64;	5.83, 12.59, 21.94;	-191.80
	125	0.54	-4.42, -2.71, -1.00;	5.94, 13.55, 24.25;	-191.86
	150	0.56	-5.15, -3.21, -1.27;	6.01, 14.31, 26.16;	-191.93
	175	0.57	-5.79, -3.64, -1.49;	6.06, 14.95, 27.80;	-192.00
	200	0.58	-6.34, -4.00, -1.66;	6.10, 15.51, 29.24;	-192.06
	225	0.59	-6.84, -4.33, -1.81;	6.13, 16.00, 30.53;	-192.11
	250	0.60	-7.30, -4.62, -1.94;	6.16, 16.44, 31.69;	-192.17
80	100	0.53	-2.42, -1.73, -1.05;	6.74, 10.06, 14.03;	-299.94
	125	0.58	-3.37, -2.58, -1.80;	7.84, 12.17, 17.46;	-299.62
	150	0.61	-4.16, -3.27, -2.39;	8.65, 13.83, 20.22;	-299.77
	175	0.63	-4.85, -3.86, -2.86;	9.32, 15.22, 22.57;	-300.00
	200	0.64	-5.46, -4.37, -3.27;	9.89, 16.45, 24.66;	-300.22
	225	0.65	-6.02, -4.82, -3.62;	10.40, 17.53, 26.52,	-300.43
	250	0.66	-6.52, -5.23, -3.93;	10.86, 18.52, 28.21;	-300.62
150	160	0.54	-4.25, -3.63, -3.00;	11.71, 15.17, 19.08;	-300.24
	170	0.56	-4.66, -4.01, -3.36;	12.96, 16.95, 21.48;	-299.29
	180	0.57	-5.03, -4.36, -3.69;	14.03, 18.51, 23.61;	-299.05
	190	0.58	-5.39, -4.69, -3.99;	15.01, 19.94, 25.57;	-299.09
	200	0.59	-5.72, -5.00, -4.28;	15.90, 21.26, 27.40;	-299.25
	210	0.60	-6.04, -5.30, -4.56;	16.73, 22.50, 29.12;	-299.45
	220	0.61	-6.34, -5.58, -4.82;	17.50, 23.66, 30.76;	-299.69
	230	0.61	-6.63, -5.85, -5.06;	18.22, 24.77, 32.30;	-299.92
	240	0.62	-6.92, -6.11, -5.30;	18.92, 25.81, 33.78;	-300.16
	250	0.62	-7.19, -6.36, -5.52;	19.58, 26.82, 35.20;	-300.39

Table 3-3. Estimates derived by the lognormal discovery process model for the simulated population given $\mu = -4$, $\sigma^2 = 20$, N = 200, and $\beta = 1$ for different values of n.

n	N	β^	μ^	σ^2^	Log L
20	25	1.2	2.09, 3.00, 3.92;	1.87 4.60, 8.55;	-126.46
	50	1.1	-0.32, 1.14, 2.60;	2.42, 7.82, 16.31;	-126.92
	75	1.3	-1.92, 0.02, 1.97;	2.71, 9.72, 21.07;	-127.09
	100	1.3	-3.13, -0.97, 1.54;	2.98, 11.14, 24.50;	-127.20
	125	1.3	-4.13, -1.46, 1.23;	3.11, 12.36, 27.74;	-127.27
	150	1.3	-4.92, -1.99, 0.94;	3.33, 13.26, 29.79;	-127.34
	175	1.3	-5.62, -2.46, 0.70;	3.58, 14.15, 31.60;	-127.38
	200	1.3	-6.32, -2.89, 0.52;	3.67, 15.01, 34.04;	-127.42
	225	1.3	-6.93, -3.28, 0.38;	3.65, 15.79, 36.47;	-128.45
	250	1.3	-7.41, -3.59, 0.22;	3.69, 16.34, 38.00;	-128.48
40	50	1.1	0.31, 1.13, 1.96;	4.36, 7.61, 11.74;	-221.60
	75	1.0	-1.26, -0.26, 0.74;	5.69, 10.65, 17.15;	-219.47
	100	1.0	-2.43, -1.25, -0.07;	6.56, 12.70, 20.84;	-218.96
	125	1.0	-3.37, -2.02, -0.67;	7.25, 14.28, 23.69;	-218.94
	150	1.1	-4.16, -2.65, -1.15;	7.75, 15.61, 26.20;	-218.61
	175	1.1	-4.83, -3.19, -1.55;	8.24, 16.66, 28.01;	-218.54
	200	1.1	-5.44, -3.67, -1.89;	8.70, 17.62, 29.66;	-218.48
	225	1.1	-6.00, -4.09, -2.18;	9.03, 18.59, 31.55;	-218.42
	250	1.1	-6.47, -4.46, -2.45;	9.26, 19.42, 33.31;	-218.40
80	100	1.0	-1.30, -0.65, -0.01;	6.23, 9.19, 12.72;	-351.24
	125	1.1	-2.29, -1.57, -0.84;	7.70, 11.72, 16.59;	-350.60
	150	1.1	-3.13, -2.32, -1.51;	8.84, 13.74, 19.72;	-350.52
	175	1.1	-3.83, -2.95, -2.05;	9.77, 15.38, 22.27;	-350.59
	200	1.1	-4.51, -3.52, -3.52;	10.60, 16.88, 24.62;	-350.63
	225	1.1	-5.06, -4.00, -2.94;	11.29, 18.25, 26.89;	-350.70
	250	1.1	-5.57, -4.43, -3.29;	11.88, 19.26, 28.41;	-350.83
150	160	1.0	-3.26, -2.65, -2.03;	11.53, 14.91, 18.72;	-376.19
	180	1.1	-4.09, -3.43, -2.77;	14.36, 18.81, 23.86;	-375.16
	200	1.1	-4.87, -4.16, -3.44;	16.71, 22.17, 28.41;	-375.44
	225	1.1	-4.88, -5.70, -4.05;	19.20, 25.58, 32.86;	-376.15
	250	1.1	-6.43, -5.67, -4.92;	21.69, 29.30, 38.05;	-376.51

Table 3-4. The lognormal discovery process model estimates of μ, σ^2 and β for the Beaverhill Lake play.

N	β^	μ^	σ^2^	Log L
100	0.2	-5.67, -4.66, -3.65;	16.19; 23.19, 31.45;	-129.65
125	0.3	-6.92, -5.82, -4.72;	18.13, 26.81, 37.18;	-126.06
150	0.3	-7.93, -6.73, -5.54;	19.45, 29.38, 41.35;	-124.39
175	0.3	-8.79, -7.49, -6.19;	20.48, 31.42, 44.70;	-123.38
200	0.4	-9.54, -8.14, -6.73;	21.33, 33.12, 47.50;	-122.70
225	0.4	-10.20, -8.07, -7.20;	22.04, 34.59, 49.94;	-122.21
250	0.4	-10.80, -9.20, -7.60;	22.68, 35.87, 52.06;	-121.84

Table 3-5. Group distributions derived from the lognormal discovery process model for the Beaverhill Lake play.

Class Interval 10^6 m^3	Number of Pools Discovered	Undiscovered N = 152	Undiscovered N = 154
< 0.000000001	1	0	0
0.000000001 - 0.00000001	0	1	1
0.00000001 - 0.0000001	0	4	4
0.0000001 - 0.000001	0	8	8
0.000001 - 0.00001	2	12	12
0.00001 - 0.0001	5	15	16
0.0001 - 0.001	10	14	15
0.001 - 0.01	20	4	5
0.01 - 0.1	19	2	2
0.1 - 1	13	2	2
1 - 10	7	1	1
10 - 100	7	0	0
> 100	3	0	1
	87	65	67

Figure 3-10. Pool size distributions estimated by the discovery process model (A) and the assumption of random sampling (B).

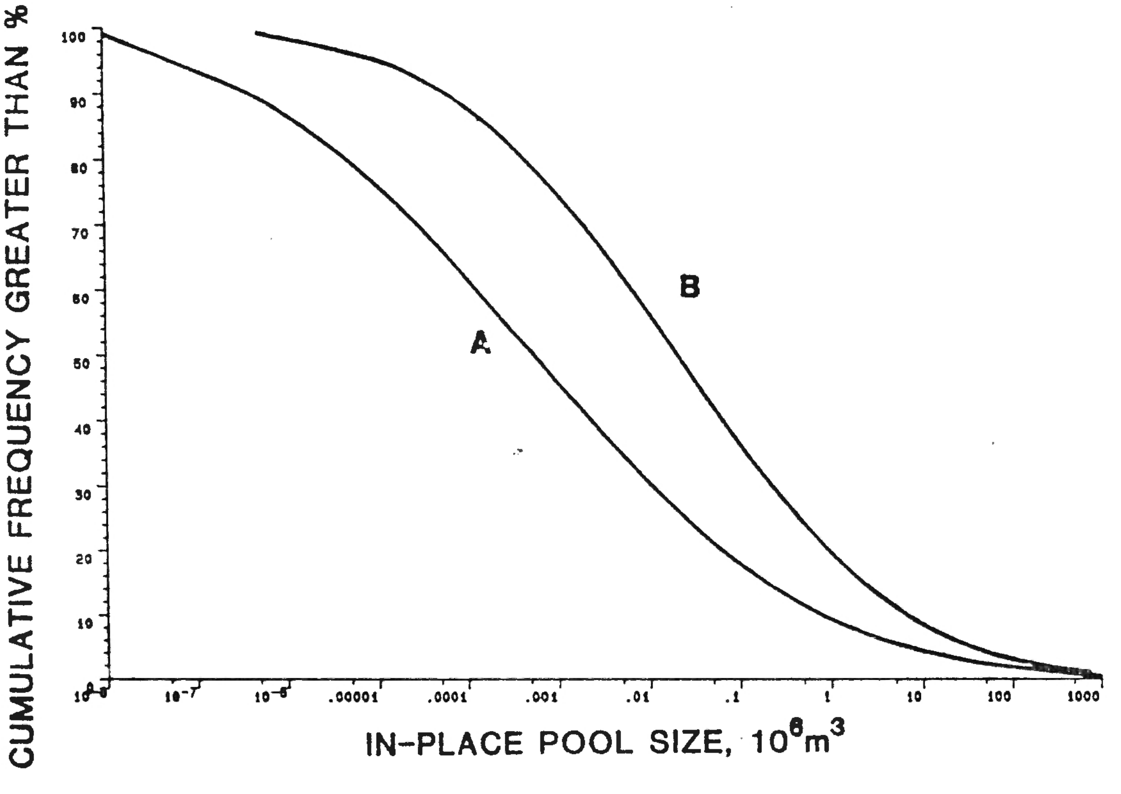


Figure 3-11. The Q-Q plots showing: the ends of the plot curve upward on the right and downward on the left (upper); the data is symmetric but the hypothesized distribution is not (lower left), and the plateaus resulting from mixed populations (lower right).

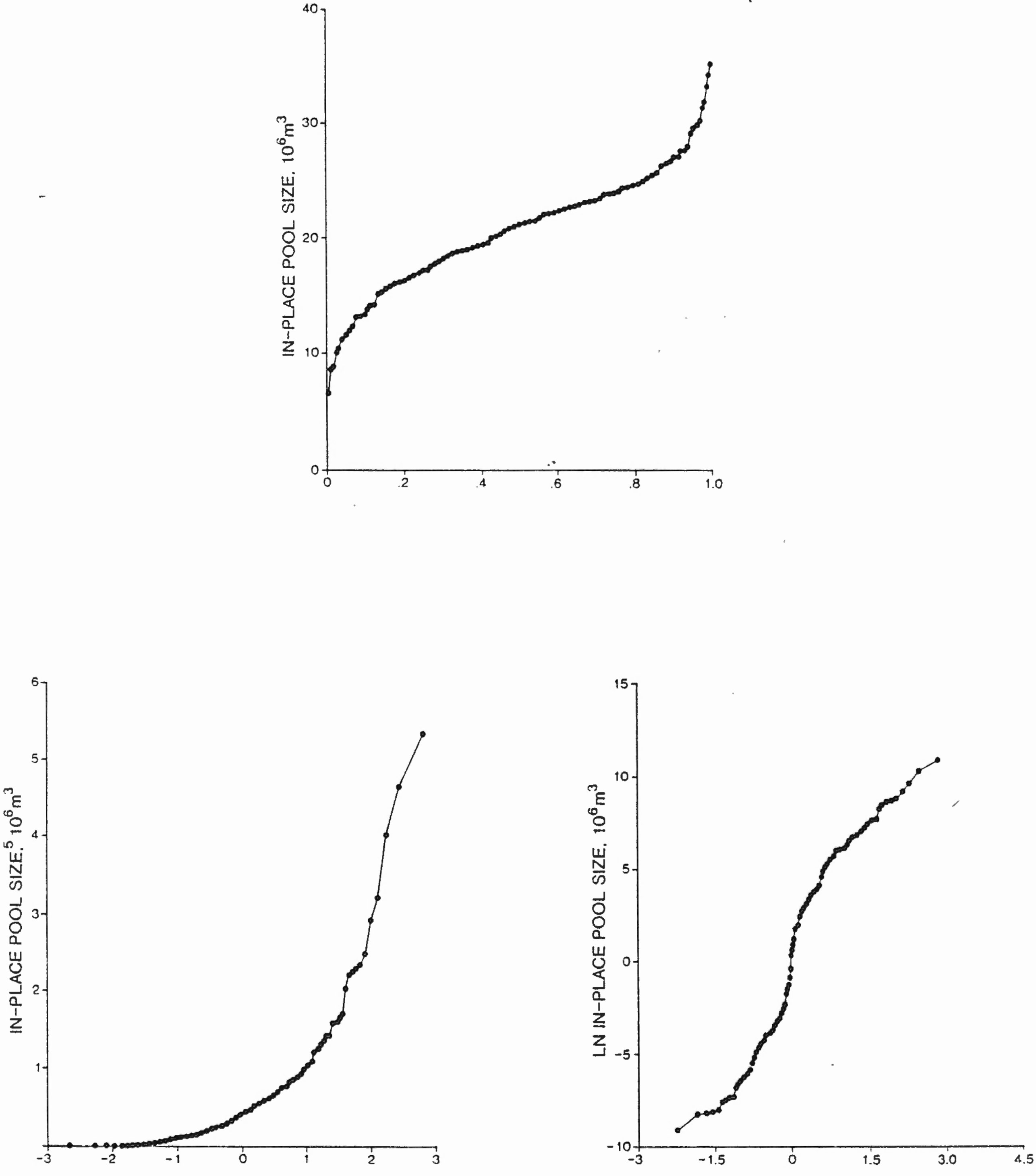


Table 3-6. Estimates derived by the nonparametric discovery process model for the simulated population given $\mu = -4$, $\sigma^2 = 20$, $\beta = 0.3$, and N =200.

n	N	$\hat{\beta}$	$\hat{\mu}$	$\hat{\sigma^2}$	Log L
20	25	0.0	-0.25, 1.58, 3.42;	7.16, 15.83, 24.50;	-102.25
	50	0.0	-0.30, 1.58, 3.47;	7.11, 15.83, 24.54;	-102.25
	100	0.2	-3.94, -1.33, 1.28;	6.09, 16.62, 27.18;	-102.23
	150	0.3	-5.66, -2.69, 0.27;	5.28, 13.89, 22.54;	-102.13
	200	0.3	-5.79, -2.83, 0.14;	5.00, 13.34, 21.68;	-102.11
	250	0.4	-6.94, -3.71, -0.14;	4.37, 10.48, 16.58;	-102.04
40	50	0.1	-1.52, -0.19, 1.13;	9.94, 17.45, 24.96;	-257.22
	100	0.2	-4.09, -2.12, -0.15;	6.52, 19.72, 32.93;	-256.75
	150	0.3	-6.34, -3.56, -0.78;	4.61, 19.90, 35.19;	-256.50
	200	0.3	-6.98, -4.05, -1.13;	5.55, 19.63, 33.70;	-256.39
	250	0.3	-7.39, -4.34, -1.28;	5.89, 19.28, 32.68;	-256.44
80	100	0.2	-2.41, -1.48, -0.54;	10.07, 15.94, 21.81;	-616.02
	150	0.3	-4.08, -2.84, -1.61;	8.82, 17.30, 25.75;	-615.79
	200	0.3	-4.91, -3.48, -2.06;	8.21, 16.90, 25.59;	-615.95
	250	0.3	-5.38, -3.87, -2.36;	8.12, 16.41, 24.70;	-616.50
150	160	0.2	-2.70, -2.09, -1.48;	10.28, 13.58, 16.88;	-1334.75
	180	0.3	-3.22, -2.60, -1.98;	10.97, 14.80, 18.63;	-1334.67
	200	0.3	-3.64, -2.97, -2.29;	10.84, 15.27, 19.70;	-1333.41
	250	0.3	-4.36, -3.61, -2.86;	10.62, 15.51, 20.41;	-1333.18

Table 3-7. Estimates derived by the nonparametric discovery process model for the simulated population given $\mu = -4$, $\sigma^2 = 10$, $\beta = 0.6$, and N = 200.

n	N	β^	μ^	σ^2^	Log L
20	50	0.4	-4.15, -1.84, 0.46;	0.88, 8.22, 16.66;	-101.81
	100	0.6	-6.13, -3.05, 0.03;	1.27, 7.61, 13.94;	-101.71
	150	0.6	-6.92, -3.49, -0.06;	1.65, 7.07, 12.50;	-101.48
	200	0.6	-7.24, -3.73, -0.21;	1.75, 6.68, 11.60;	-101.43
	250	0.6	-7.42, -3.87, -0.31;	1.84, 6.40, 10.96;	-101.45
40	50	0.2	-1.32, -0.55, 0.21;	1.37, 4.81, 8.24;	-257.68
	100	0.4	-3.02, -1.58, -0.14;	1.38, 5.70, 12.13;	-257.13
	150	0.4	-3.49, -1.84, -0.19;	1.45, 5.65, 12.17;	-257.08
	200	0.6	-4.74, -2.56, -0.38;	1.22. 5.78, 11.66;	-256.90
	250	0.6	-5.07, -2.76, -0.46;	0.16, 5.79, 11.41;	-256.71
80	100	0.4	-3.10, -2.22, -1.34;	2.69, 8.74, 14.80;	-610.79
	150	0.6	-5.21, -3.55, -1.89;	1.95, 11.11, 20.27;	-611.48
	200	0.6	-6.45, -4.32, -2.19;	2.12, 11.65, 21.17;	-611.27
	250	0.6	-7.12, -4.85, -2.59;	2.96, 11.62, 20.27;	-611.43
150	160	0.6	-3.60, -3.12, -2.64;	5.74, 8.07, 10.41;	-1296.86
	180	0.6	-4.18, -3.64, -3.11;	6.50, 9.66, 12.81;	-1293.76
	200	0.6	-4.73, -4.09, -3.44;	6.63, 10.82, 15.01;	-1293.19
	250	0.6	-5.83, -4.99, -4.16;	7.31, 12.60, 17.88;	-1311.59

Table 3-8. Estimates derived by the nonparametric discovery process model for the simulated population given $\mu = -4$, $\sigma^2 = 10$, $\beta = 1$, and N = 200.

n	N	β^	μ^	σ^2^	Log L
20	50	0.6	-3.68, -1.70, 0.28;	2.01, 7.14, 12.28;	-99.61
	100	0.8	-5.20, -2.67, -0.14;	1.66, 5.22, 8.79;	-99.67
	150	0.8	-5.50, -3.03, -0.56;	1.68, 4.22, 6.76;	-99.39
	200	0.8	-5.61, -3.21, -0.82;	1.42, 3.62, 5.81;	-99.36
	250	0.8	-5.57, -3.33, -1.09;	1.36, 3.23, 5.09;	-99.42
40	50	0.6	-1.34, -0.54, 0.26;	1.39, 4.65, 7.91;	-250.01
-	100	0.8	-3.17, -1.86, -0.56;	1.26, 4.98, 8.70;	-249.23
	150	1.0	-3.93, -2.38, -0.83;	1.25, 4.28, 7.30;	-249.51
	200	1.0	-4.31, -2.58, -1.01;	1.28, 3.64, 6.34;	-249.27
	250	1.0	-4.50, -2.84, -1.18;	1.28, 3.46, 5.64;	-249.24
80	100	0.8	-2.28, -1.65, -1.03;	1.54, 4.62, 7.69;	-593.09
	150	1.0	-3.60, -2.52, -1.44;	0.89, 5.18, 9.48;	-592.89
	200	1.0	-4.45, -3.06, -1.64;	0.63, 5.20, 9.76;	-593.25
	250	1.0	-4.98, -3.41, -1.85;	0.75, 5.00, 9.25;	-593.85
150	160	1.0	-3.45, -2.69, -2.55;	5.07, 5.92, 8.95;	-1238.86
	180	1.0	-4.03, -3.48, -2.94;	5.12, 8.22, 11.32;	-1240.65
	200	1.0	-4.51, -3.88, -3.25;	5.18, 8.94, 12.70;	-1241.70
	250	1.0	-5.39, -4.63, -3.85;	5.26, 9.64, 13.99;	-1244.10

Table 3-9. The Nonparametric maximum likelihood estimates of μ, σ^2, and β for the Beaverhill Lake play.

N	β^	μ^	σ^2^	Log L
100	0.3	-5.50, -4.60, -3.71;	17.55, 22.07, 27.79;	-682.49
125	0.3	-6.62, -5.70, -4.78;	20.19, 25.92, 31.65;	-677.38
150	0.3	-7.54, -6.58, -5.62;	23.00, 28.97, 34.94;	-675.76
175	0.3	-8.35, -7.35, -6.35;	25.79, 31.61, 37.43;	-675.18
200	0.4	-8.49, -7.45, -6.42;	21.17, 26.88, 32.58;	-674.87
225	0.4	-9.00, -7.93, -6.86;	22.21, 27.80, 33.39;	-673.89
250	0.4	-9.44, -8.37, -7.29;	23.33, 27.68, 34.04;	-673.38

Table 3-10. Group distributions derived from the nonparametric discovery process model for the Beaverhill Lake play.

Class Interval $10^6 m^3$	Number of Pools Discovered	Undiscovered N = 150	Undiscovered N = 175
< 0.0000001	0	13	23
0.0000001- 0.000001	0	0	0
0.000001 - 0.00001	2	9	12
0.00001 - 0.0001	5	13	17
0.0001 - 0.001	10	13	16
0.001 - 0.01	20	13	14
0.01 - 0.1	19	4	5
0.1 - 1	13	0	1
1 - 10	7	0	0
10 - 100	7	0	0
> 100	3	0	0
	87	63	88

Figure 3-12. The empirical pool size distribution estimated using the non-parametric discovery process model (discrete distribution) and lognormal approximation (continuous distribution) for the Beaverhill Lake play.

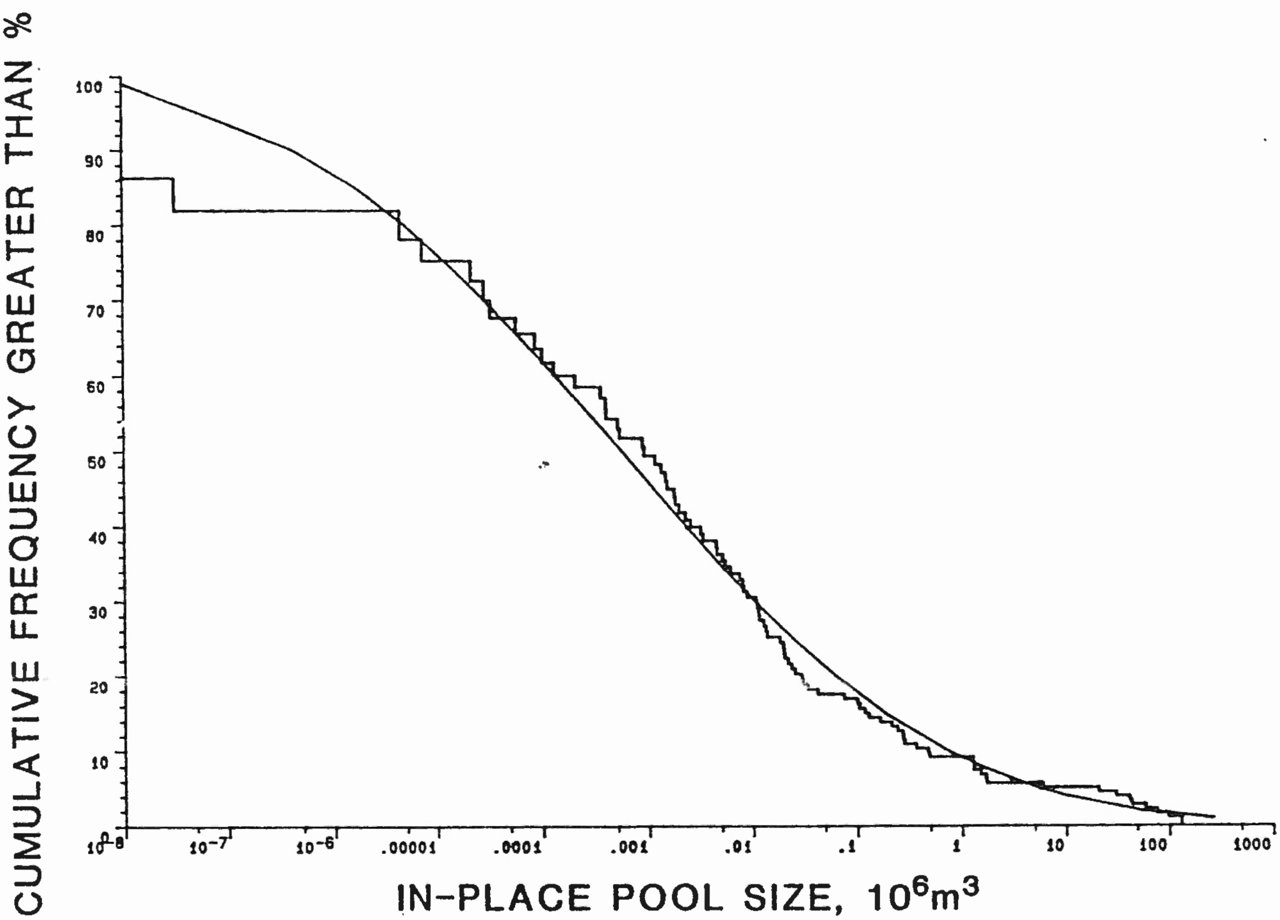
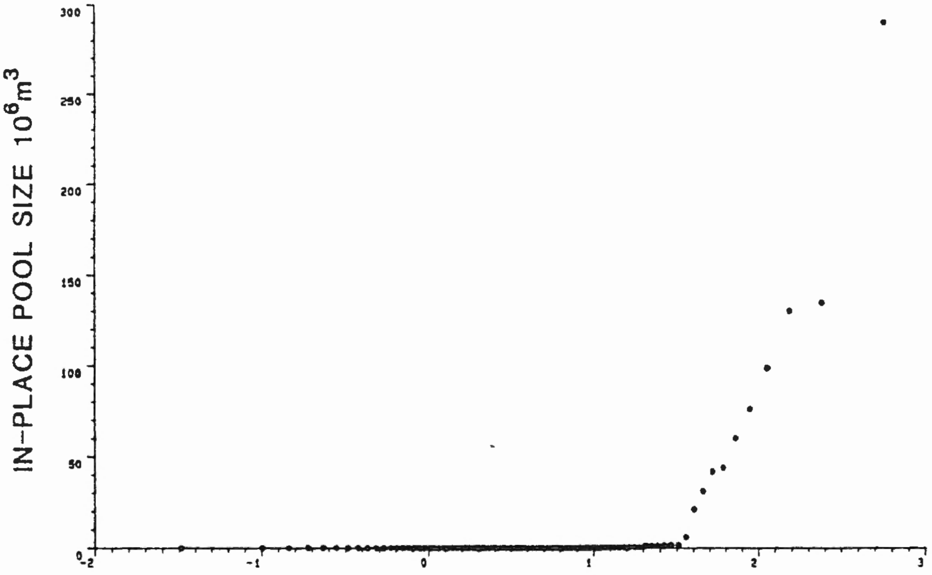
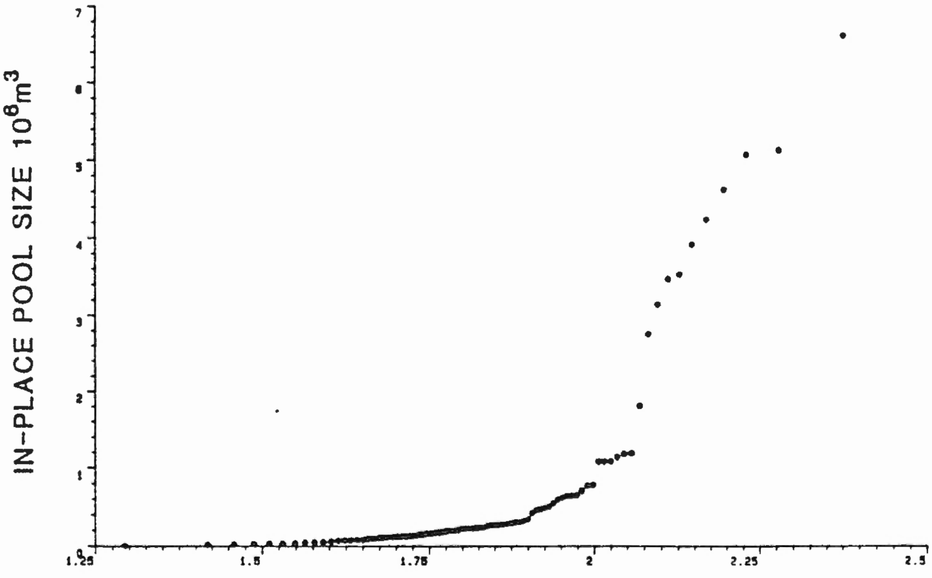


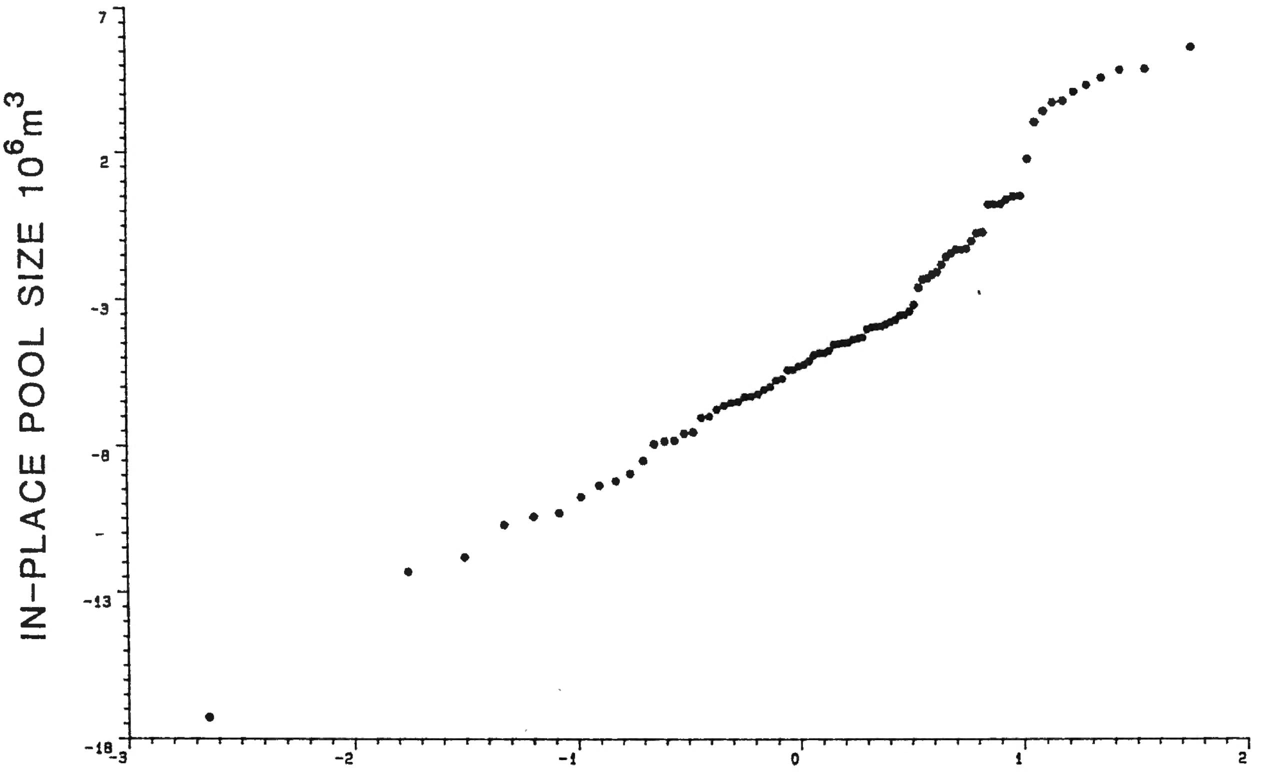
Figure 3-13. The Q-Q plots for the Beaverhill Lake play: (A) normal quantiles; (B) gamma quantiles with shape parameter = 5.0; (C) Weibull quantiles; (D) uniform quantiles; (E) one parameter exponential quantiles; (F) two parameter exponential quantiles; (G) power normal quantiles with power = 0.5; (H) shifted Pareto quantiles; (I) truncated shifted Pareto quantiles with shape factor = 0.1; (J) truncated shifted Pareto quantiles with shape = 0.9; (K) lognormal quantiles; and (L) power normal quantiles with power = 0.001.



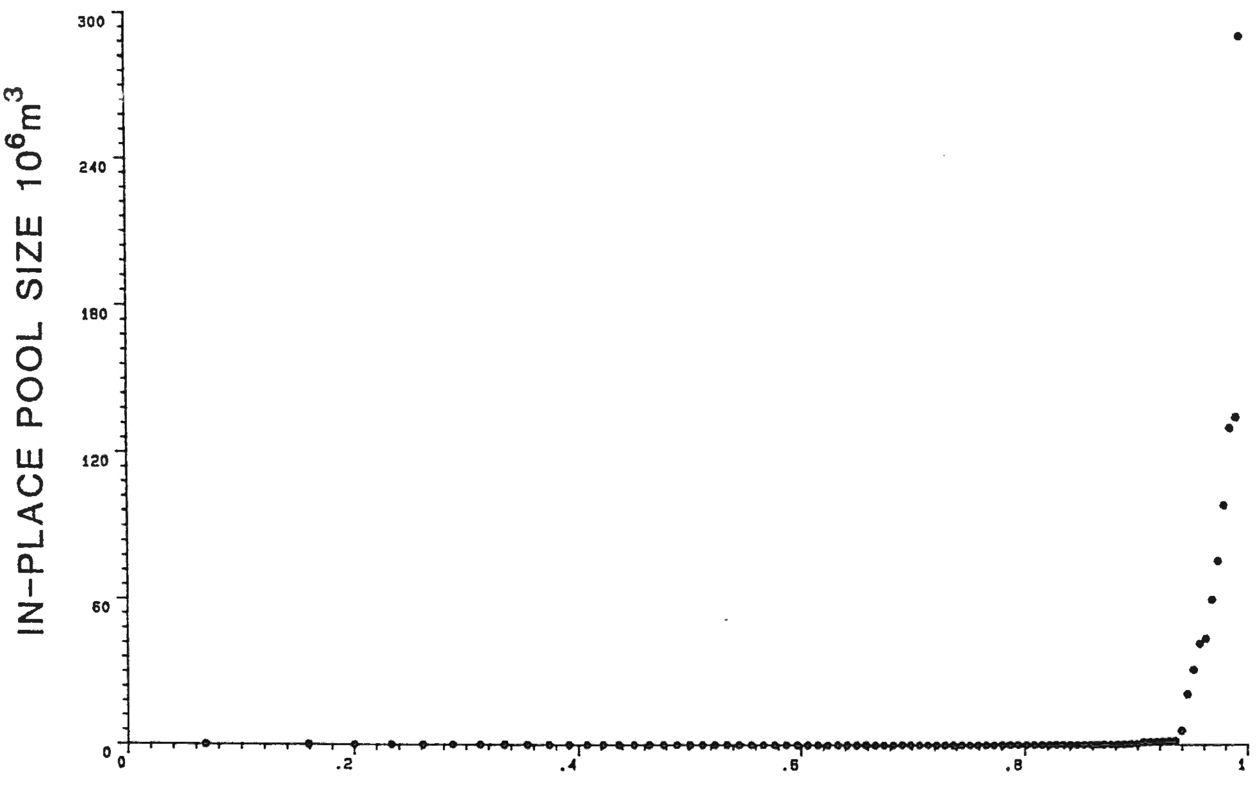
A. NORMAL QUANTILES



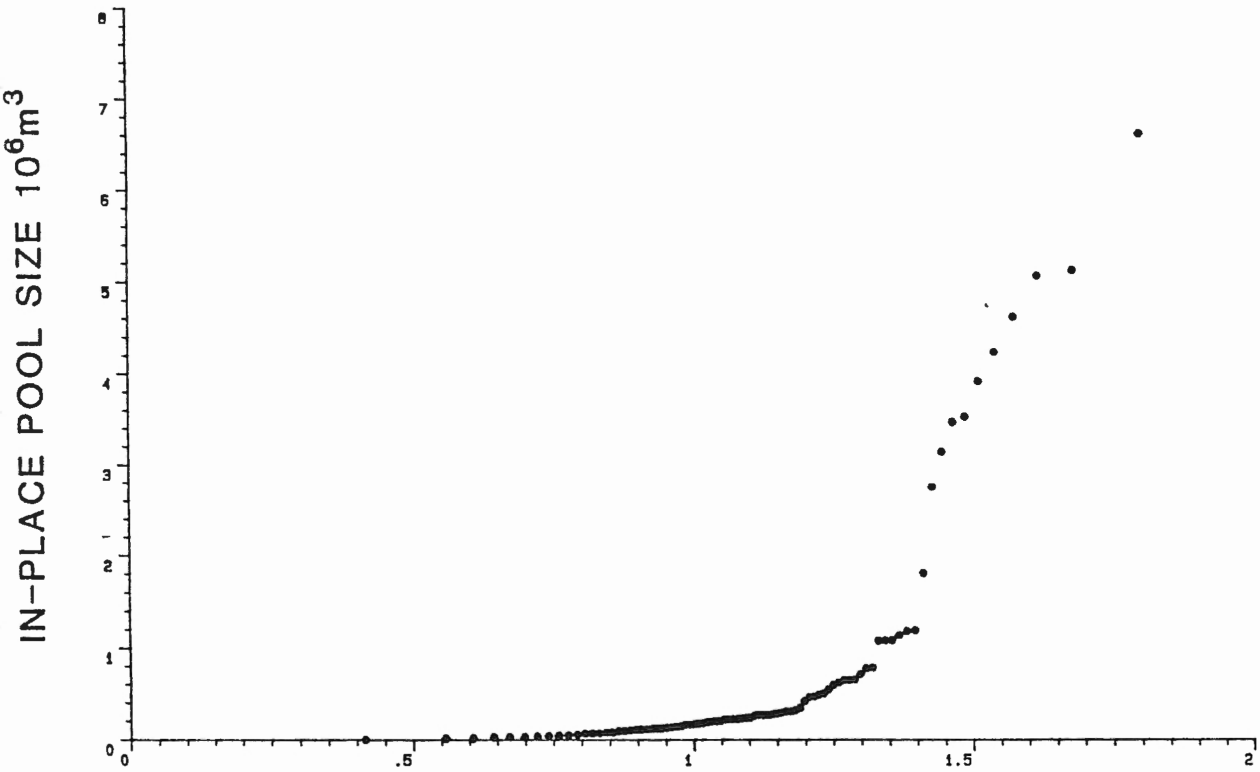
B. GAMMA QUANTILES



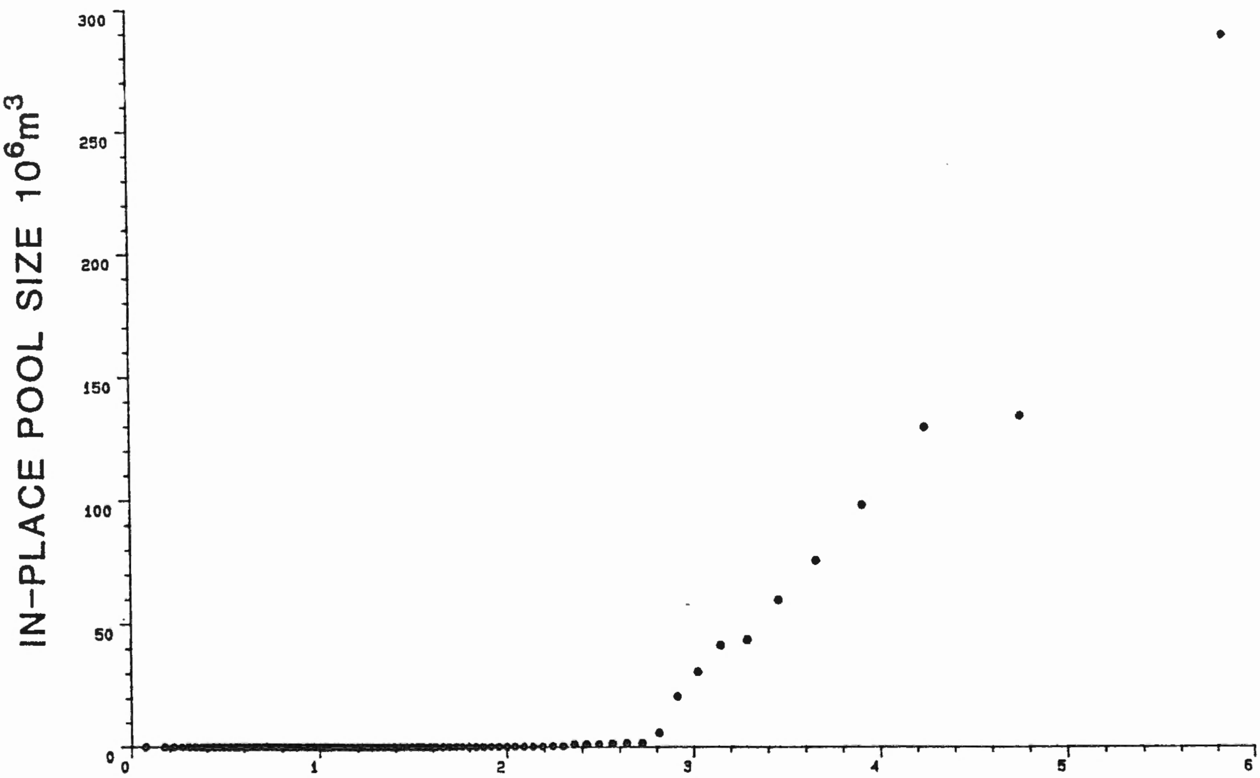
C. WEIBULL QUANTILES



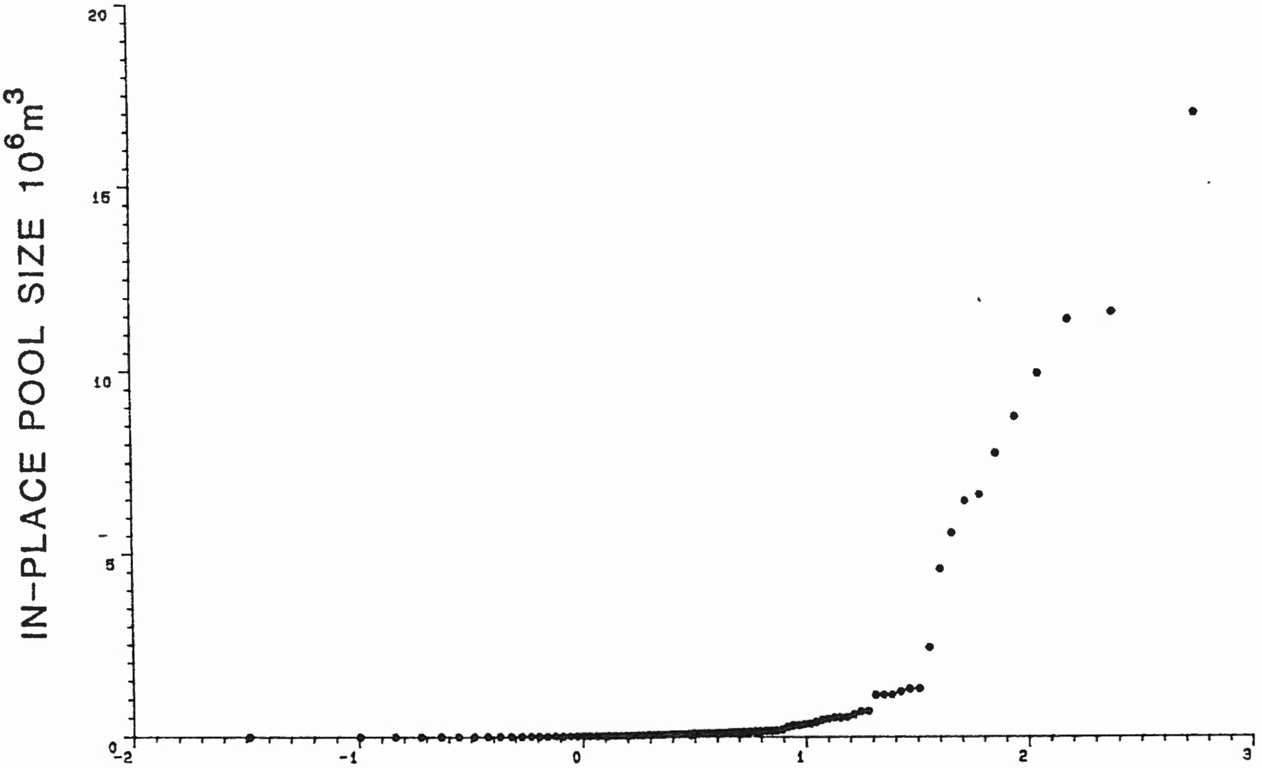
D. UNIFORM QUANTILES



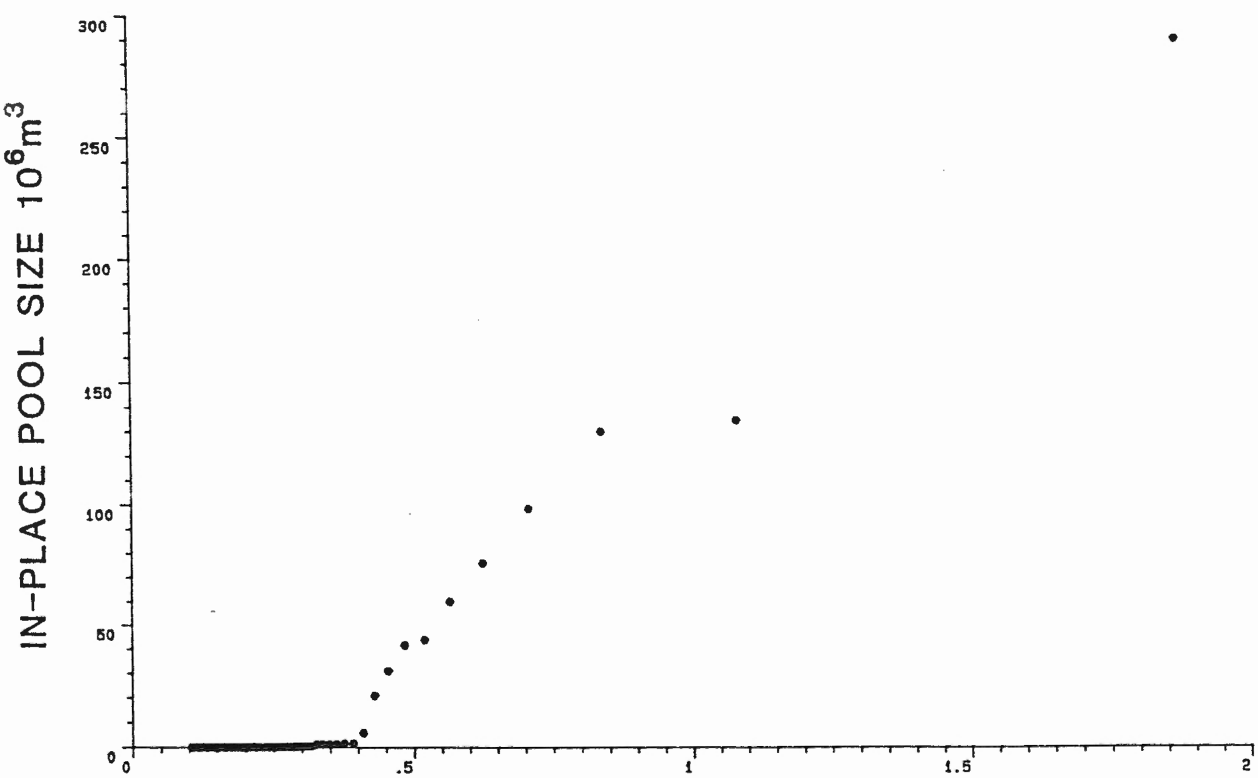
E. ONE PARAMETER EXPONENTIAL QUANTILES



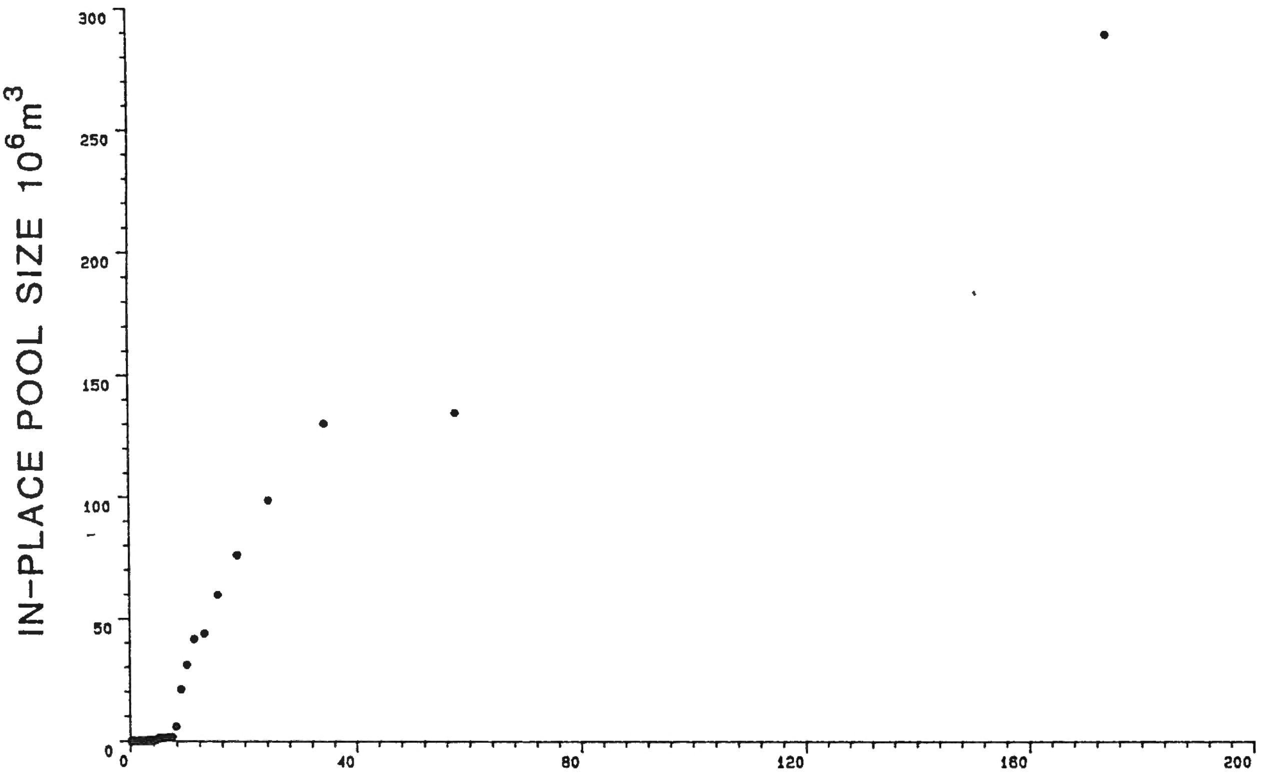
F. TWO PARAMETER EXPONENTIAL QUANTILES



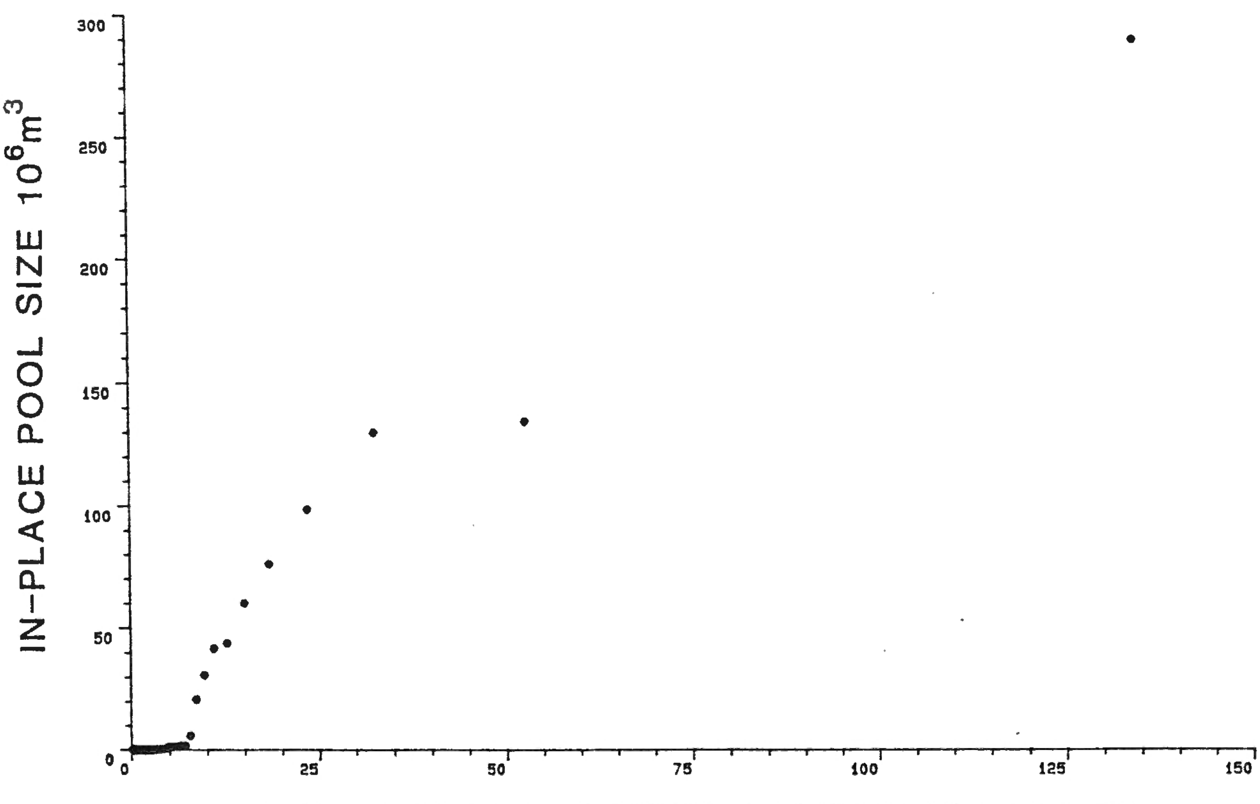
G. POWER NORMAL QUANTILES



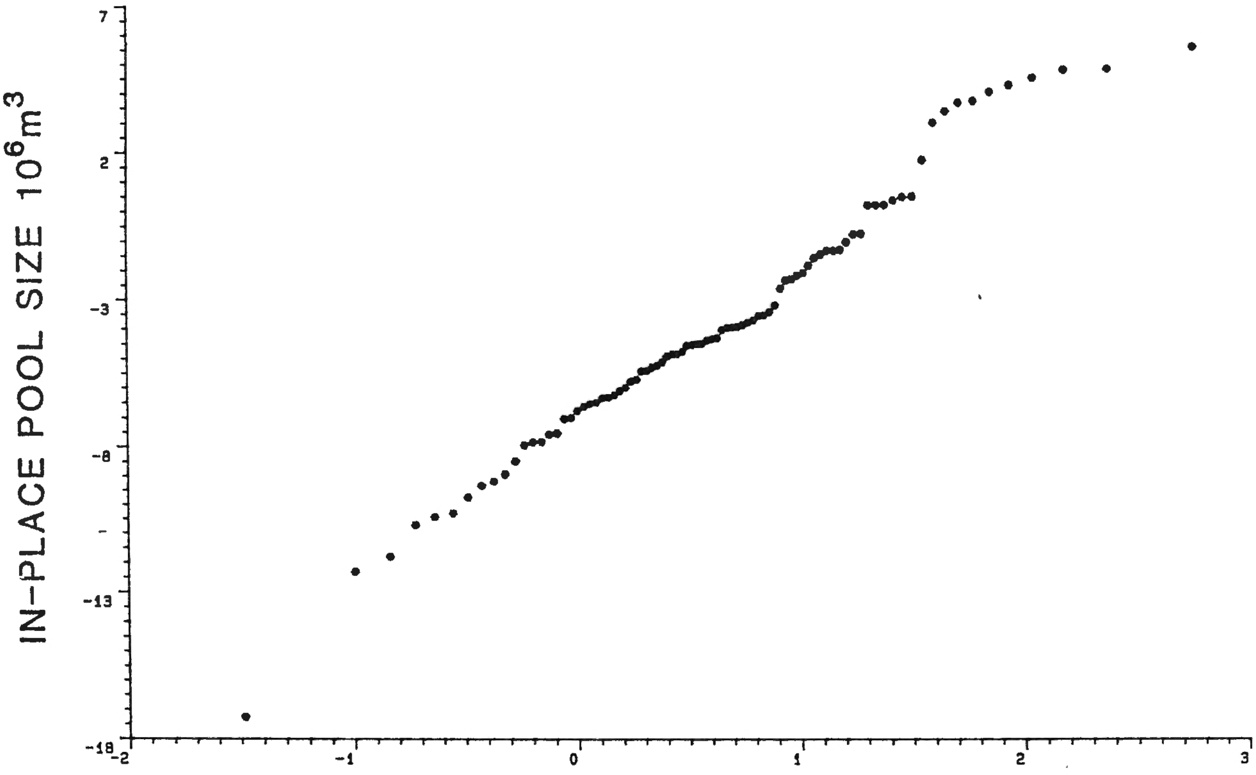
H. SHIFTED PARETO QUANTILES



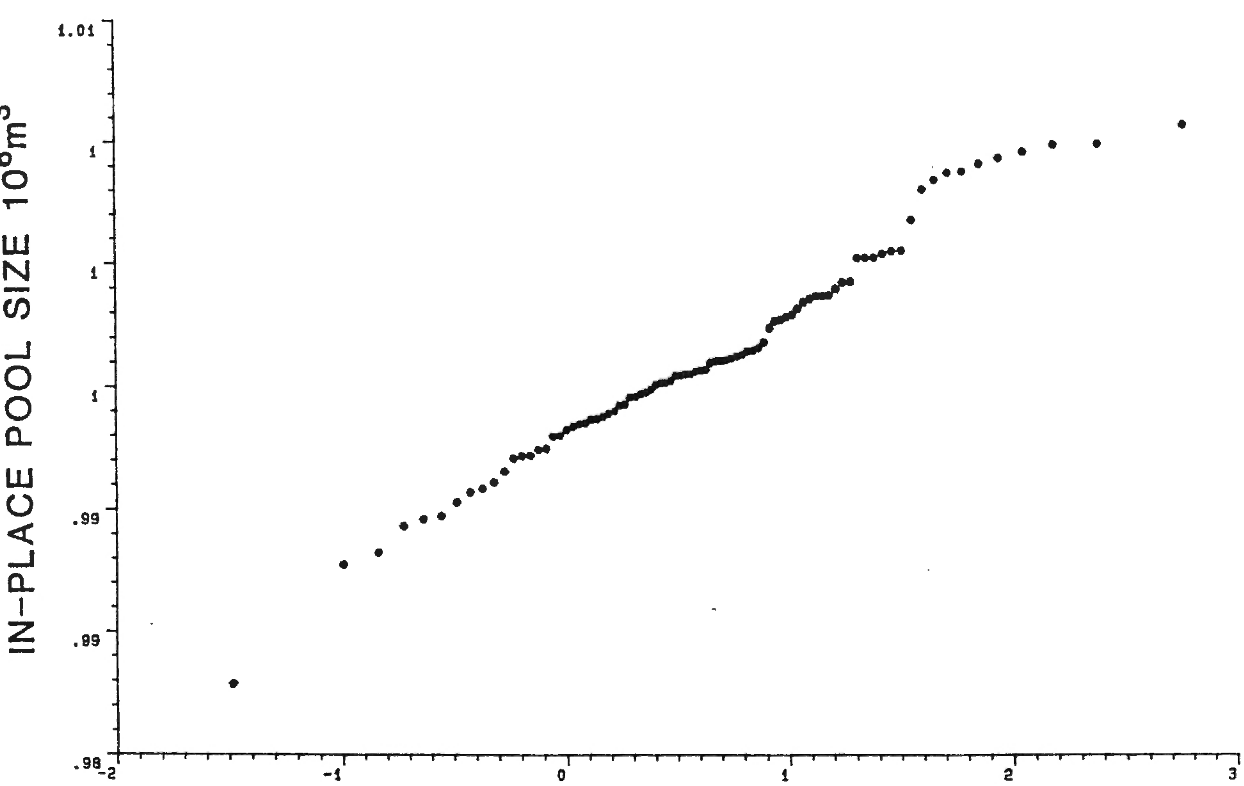
I. TRUNCATED SHIFTED PARETO QUANTILES



J. TRUNCATED SHIFTED PARETO QUANTILES

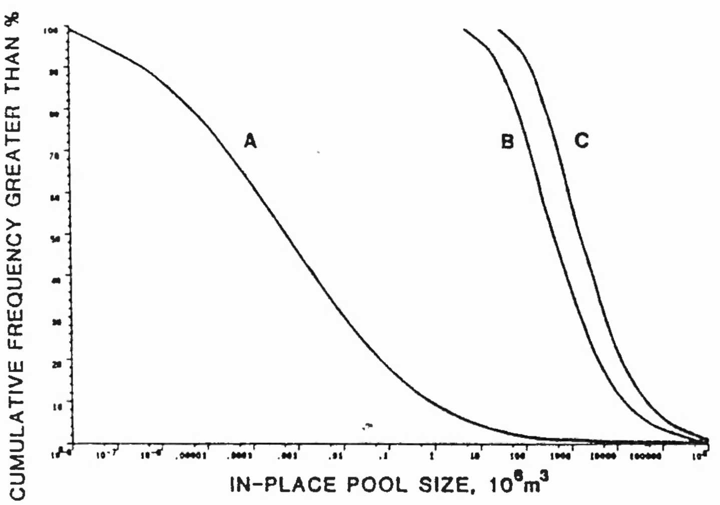


K. LOGNORMAL QUANTILES



L. POWER NORMAL QUANTILES

Figure 4-1. Pool size distribution (top A) and the distributions of the largest pool size distributions for N = 100 (top B) and N = 152 (top C) for the Beaverhill Lake play; and the pool size distribution (lower) and the largest two pool size distributions when N = 152, $\mu\hat{}$ = -6.80, and $\sigma^2\hat{}$ = 29.56 for the Beaverhill Lake play.



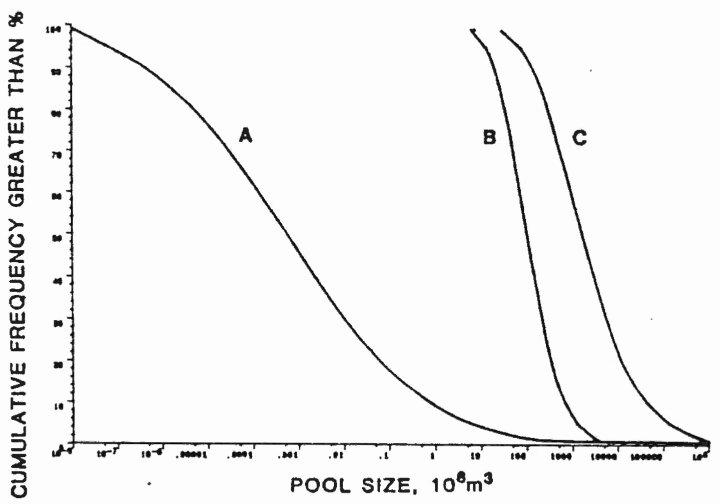


Figure 4-2. Left: individual pool-size-by-rank with σ^2 = 2 and σ^2 = 7 and where μ = 0.25 and N = 60. Right: the pool-size-by-rank for the Beaverhill Lake play (dots) with μ^ = 0.25, σ^2^ = 6.6, and N =60; the Bashaw play (crosses) with μ^ = -0.91, σ^2^ = 3.0, and N = 80; and the Zama play (triangles) with μ^ = -1.5, σ^2^ = 1.0, and N =160.

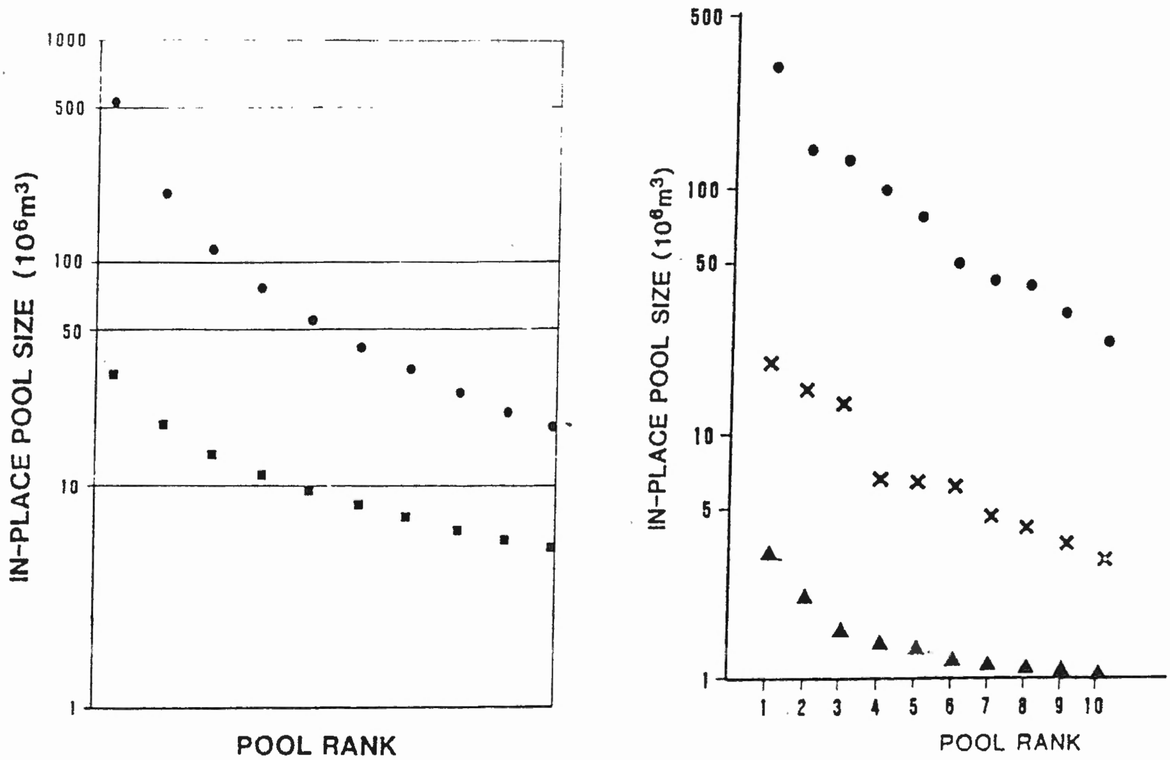


Figure 4-3. Examples of selected upper percentiles.

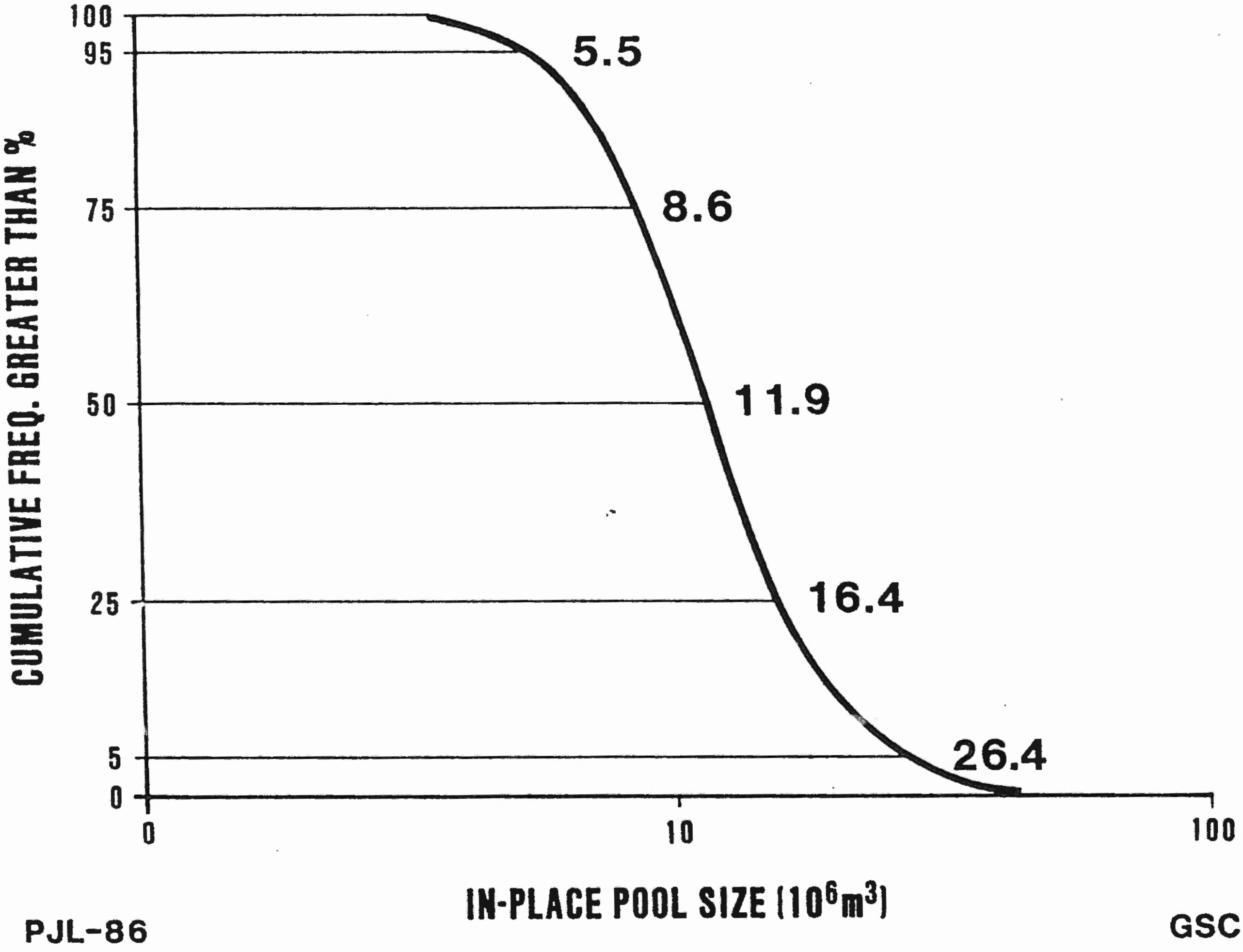


Table 4-1. Pool-size-by-rank for $N = 152$, $\hat{\mu} = -6.80$, $\hat{\sigma}^2 = 29.56$

OBSERVED		ESTIMATED POPULATION					
RANK	POOL SIZE	RANK	99%	95%	50%	5%	1%
1	290.0	1	31.13	83.18	1605.	118600	1180000
2	134.8	2	12.71	28.80	284.1	5297	22780
3	130.0	3	6.90	14.38	105.2	1163	3677
4	98.71	4	4.27	8.43	51.45	425.5	1141
5	75.40	5	2.85	5.42	29.17	199.5	482.2
6	57.72	6	2.01	3.70	18.15	108.3	243.3
7	44.00	7	1.47	2.65	12.04	64.72	137.7
8	41.30	8*	1.11	1.95	8.37	41.43	84.5
9	31.10	9*	0.85	1.48	6.03	27.95	54.96
10	21.04	10*	0.67	1.15	4.47	19.55	37.43
11	5.93	11	0.47	0.90	3.40	14.14	26.42
12	1.70	12	0.38	0.72	2.63	10.49	19.23
13	1.67	13	0.32	0.58	2.07	7.95	14.30
14	1.50	14	0.26	0.46	1.65	6.14	10.86
15	1.29	15	0.22	0.38	1.33	4.81	8.40
16	1.28	16	0.19	0.32	1.08	3.82	6.59
17	1.27	17	0.16	0.27	0.89	3.07	5.24
		18	0.13	0.22	0.74	2.50	4.21
18	0.489	19	0.12	0.19	0.62	2.05	3.42
19	0.477	20	0.10	0.16	0.52	1.69	2.80
20	0.368	21	0.09	0.14	0.45	1.41	2.31
21	0.283	22	0.08	0.12	0.38	1.18	1.93
22	0.275	23	0.07	0.11	0.33	1.00	1.61
23	0.245	24	0.06	0.09	0.28	0.84	1.36
24	0.216	25	0.05	0.08	0.24	0.72	1.15
25	0.167	26	0.04	0.07	0.21	0.62	0.98
26	0.106	27	0.04	0.06	0.18	0.53	0.84
27	0.130	28	0.03	0.05	0.16	0.45	0.72
28	0.119	29	0.03	0.05	0.14	0.39	0.62
29	0.106	30	0.03	0.04	0.12	0.34	0.54
30	0.101	31	0.02	0.04	0.11	0.29	0.44
		32	0.02	0.03	0.09	0.26	0.38
		33	0.02	0.03	0.08	0.23	0.35
		.					
		.					

Figure 4-4. Cross-plot showing the relationship between the pool area and the pool size for the Beaverhill Lake play.

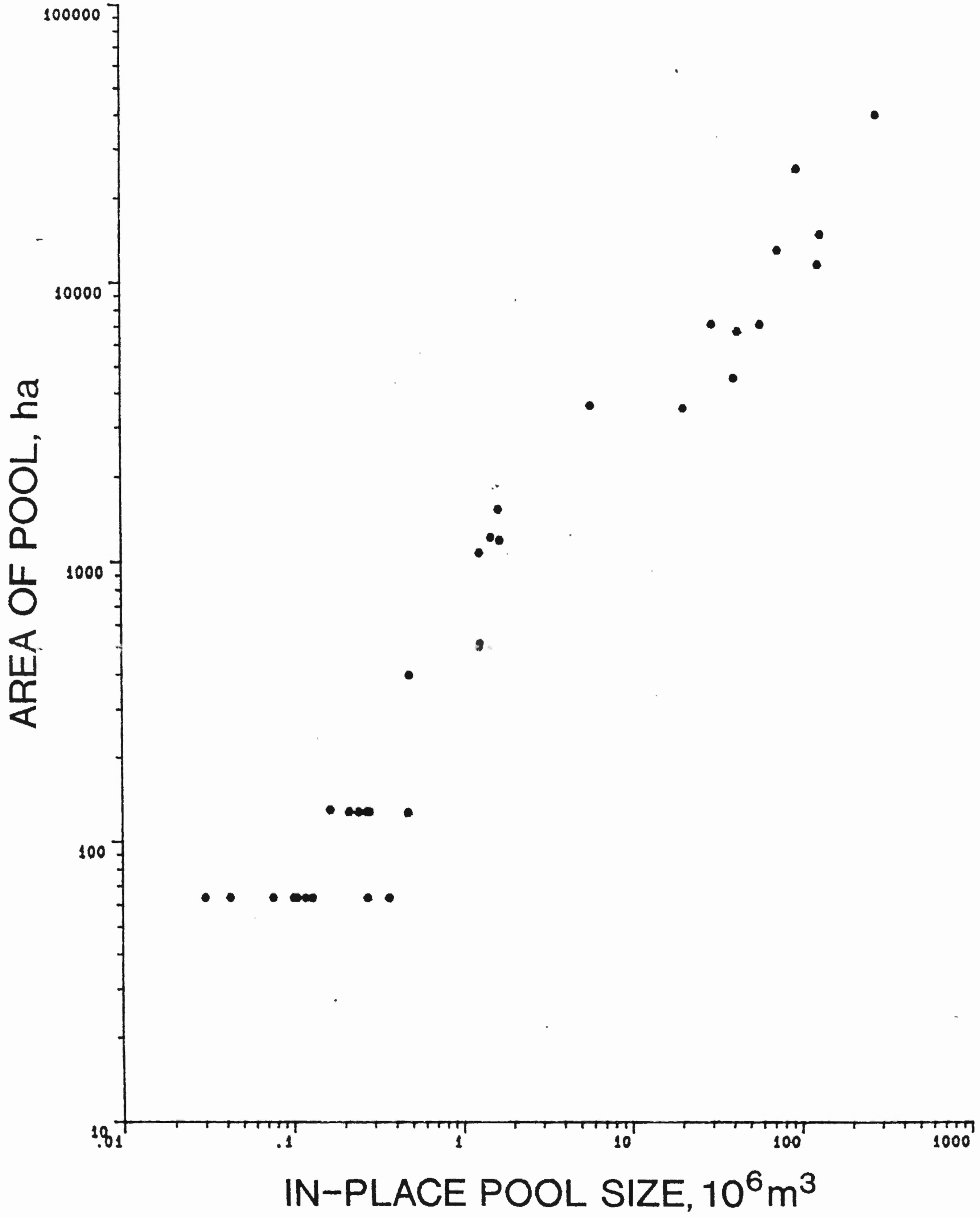
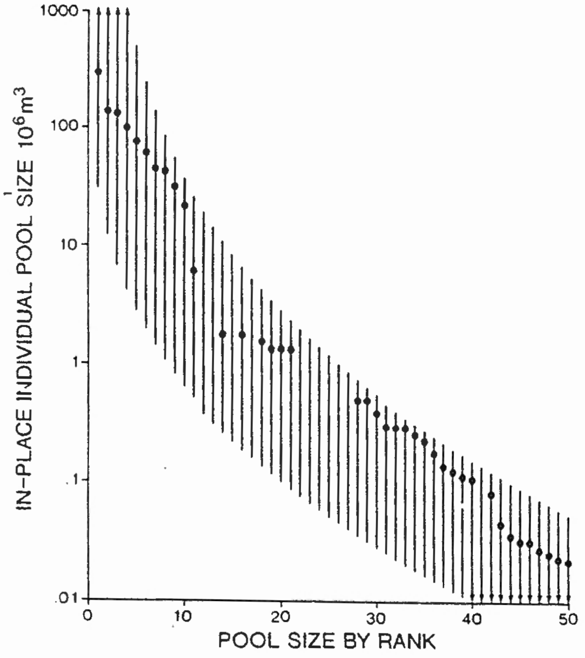


Figure 4-5. Left: Pool-size-by-rank for the Beaverhill Lake play with 95-5% prediction interval. Right: Pool-size-by-rank conditional to the pool ranks for the Beaverhill Lake play.



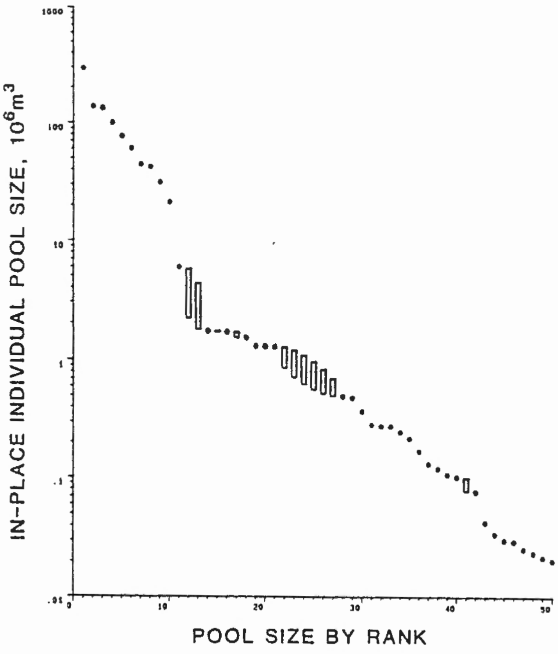


Figure 4-6. Play resource distribution (A) and the play potential distribution (B) for the Beaverhill Lake play.

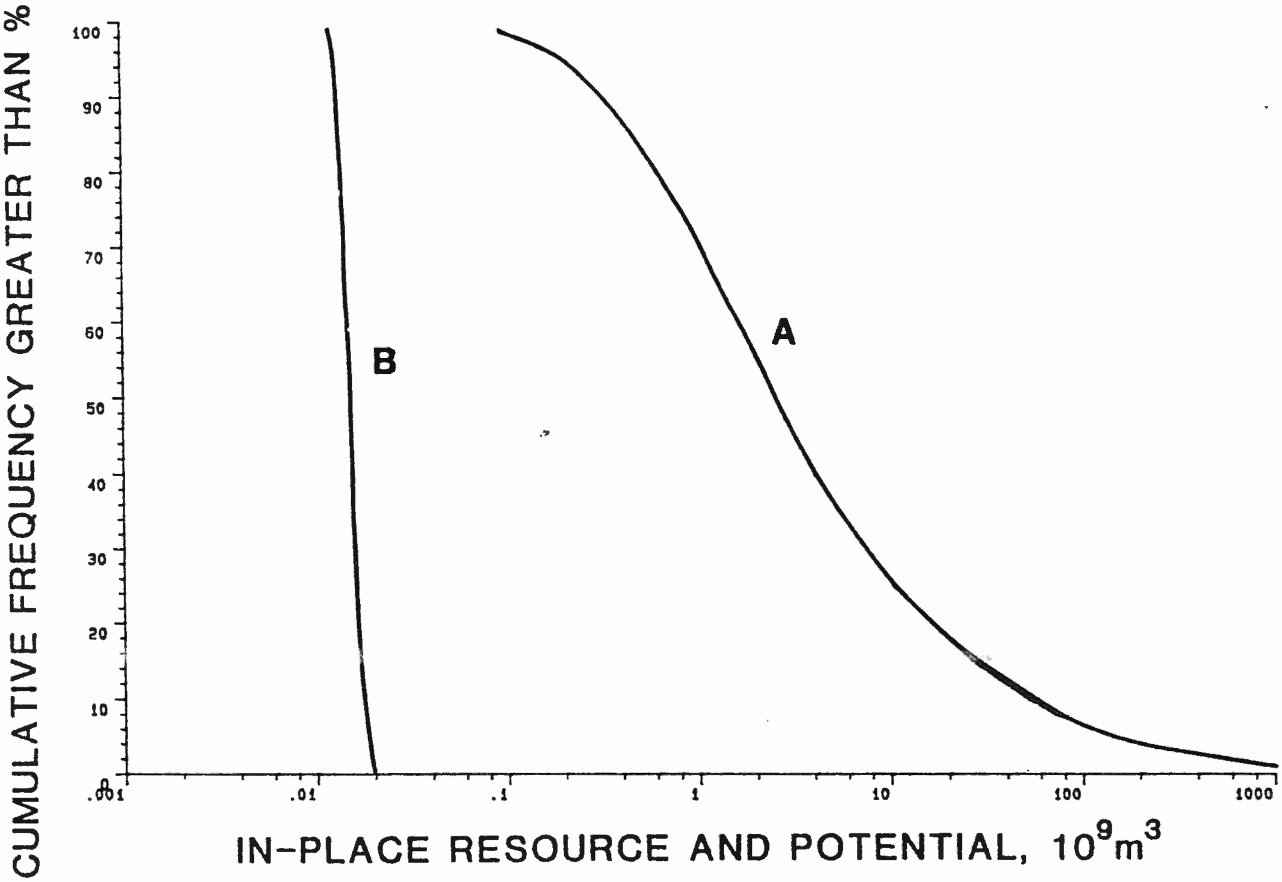


Table 5-1. Lognormal parameters and correlations of geological variables for the Beaverhill Lake play.

Variable	Sample Mean	Variance	Correlations & Covariances		
	$\hat{\mu}$	$\hat{\sigma}^2$	Pool area	Average net pay	Average porosity
Pool area	7.860	2.721	1.000		
Average netpay	2.211	0.422	0.682 (0.731)	1.000	
Average porosity	-2.674	0.068	0.641 (0.275)	0.452 (0.077)	1.000

Constant = 0.681, $2 \sum_{i<j} \sum \sigma_{ij} = 2.164$

Scale factor = 0.001

$\hat{\mu} = 2.408$, $\hat{\sigma}^2 = 3.211 + 2.164 = 5.375$

Figure 5-1. Cross plots showing: the log-linear relationship between the variables of porosity and water saturation for the Cardium marine sandstone play (left) and the Bashaw play (right) from the Western Canada Basin.

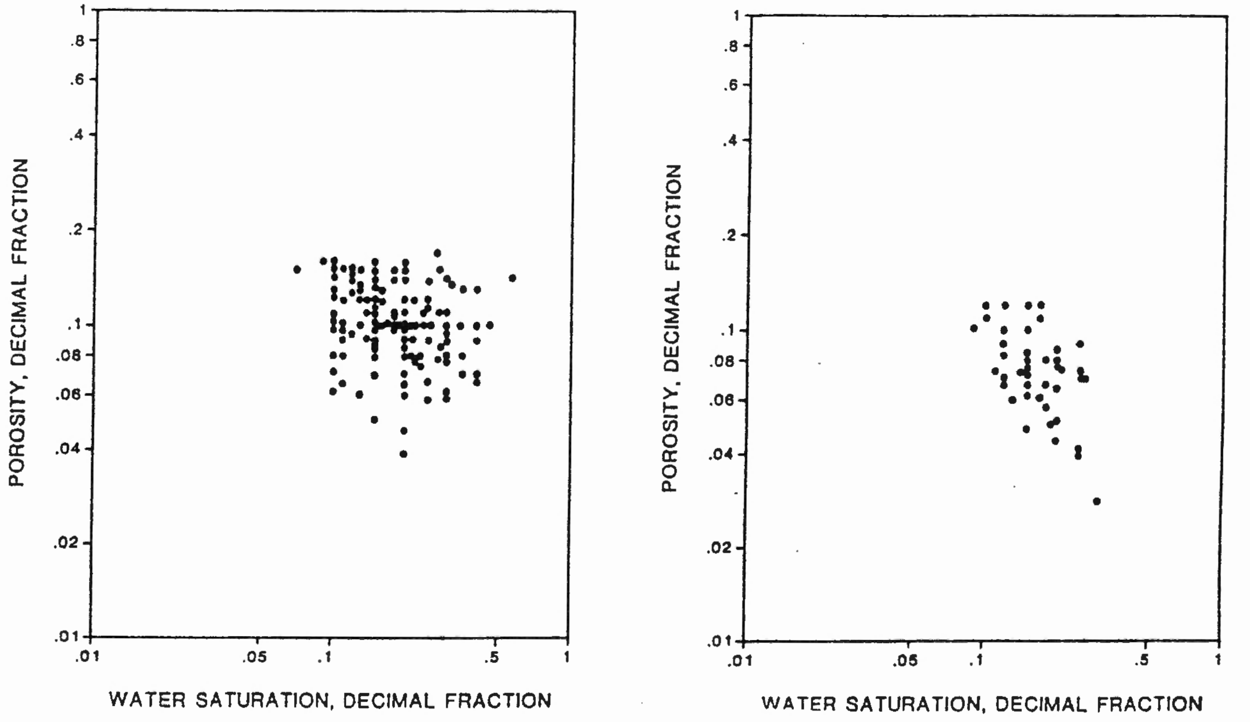


Table 5-2. Format for entry of probability distributions.

Geological variable	Unit of measurement	Probability in upper percentiles 1.0	0.5	0.02/0.1	0.0
Area of closure/pool	$mile^2$ / km^2				
Net pay/no of pay zones	m / ft / no				
Reservoir/ formation thickness	m / ft				
Porosity	decimal fraction				
Trap fill	decimal fraction				
Favourable facies	decimal fraction				
Water saturation	decimal fraction				
Oil/gas saturation	decimal fraction				
Shrinkage factor	decimal fraction				
Formation volume factor	decimal fraction				
Reservoir temperature	Celsius/ Fahrenheit				
Reservoir pressure	kPa/psi				
Recovery factor	decimal fraction				

Table 5-3. Format for entry of geological risk factors and their marginal probability.

Geological factors	Marginal probability	Level: Play	Level: Prospect
Presence of closure			
Presence of reservoir facies			
Presence of porosity			
Adequate seal			
Adequate timing			
Adequate source			
Adequate maturation			
Adequate preservation			
Adequate recovery			
Adequate play conditions			
Adequate prospect conditions			

Table 5-4. Format for entry of number of prospects and pools.

Geological variable	Probability in upper percentiles: 0.99	0.5	0.0
No of prospects			
No of pools			

Figure 5-2. Types of facies-cycle wedge (After White, 1980).

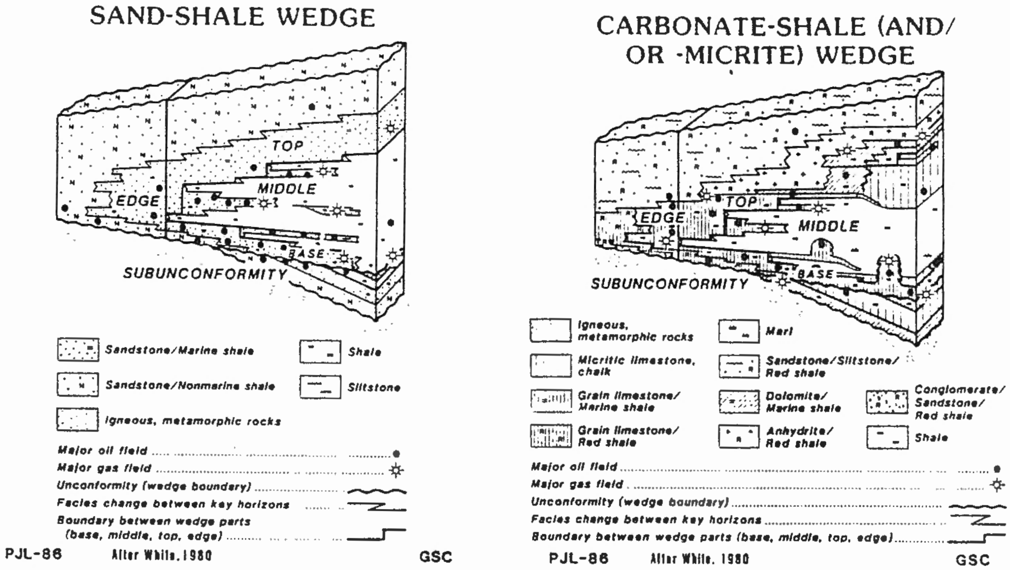
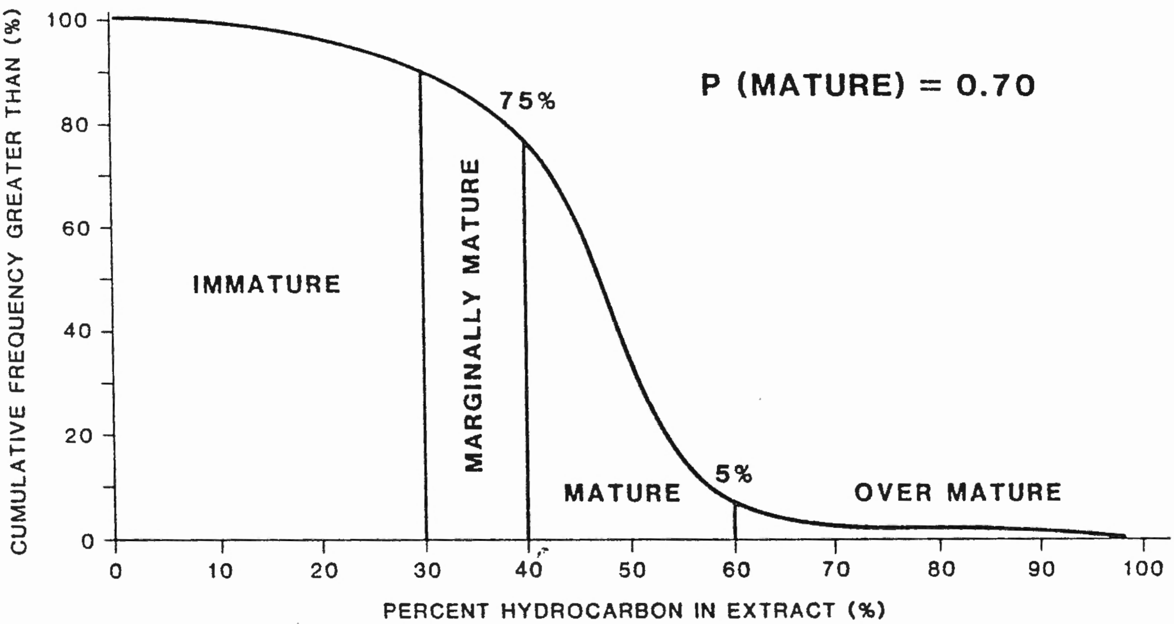


Table 5-5. Examples for play level risks for various geological models

Play type	Example	Exploration risk Sandstone	Exploration risk Carbonate
Edge	Eocene to Miocene		
	Cook Inlet, Alaska	0.15	
	* Jean Marie		0.15
Top	Belly River	0.15	
	Mission Canyon		0.44
Base	Mannville	0.60	
	Beaverhill Lake		0.35
Subunconformity	Jurassic	0.45	
	Mississippian		0.30

Note: * added by the present authors.

Figure 5-3. Top: An example of probability distribution for a variable of source rock maturation. Lower: An example of probability distribution for the amount of total organic matter.



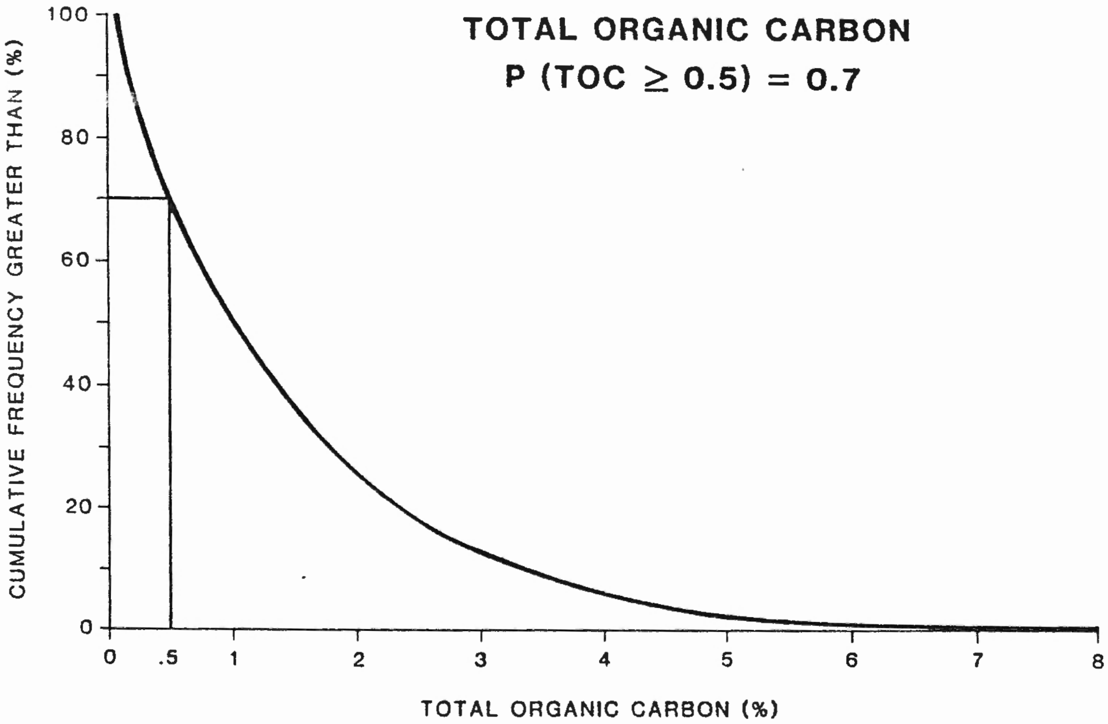


Figure 5-4. Facies map for the Es_1 Formation of the Huang-hua Basin, Eastern China (After Lee, Qin and Shi, 1989).

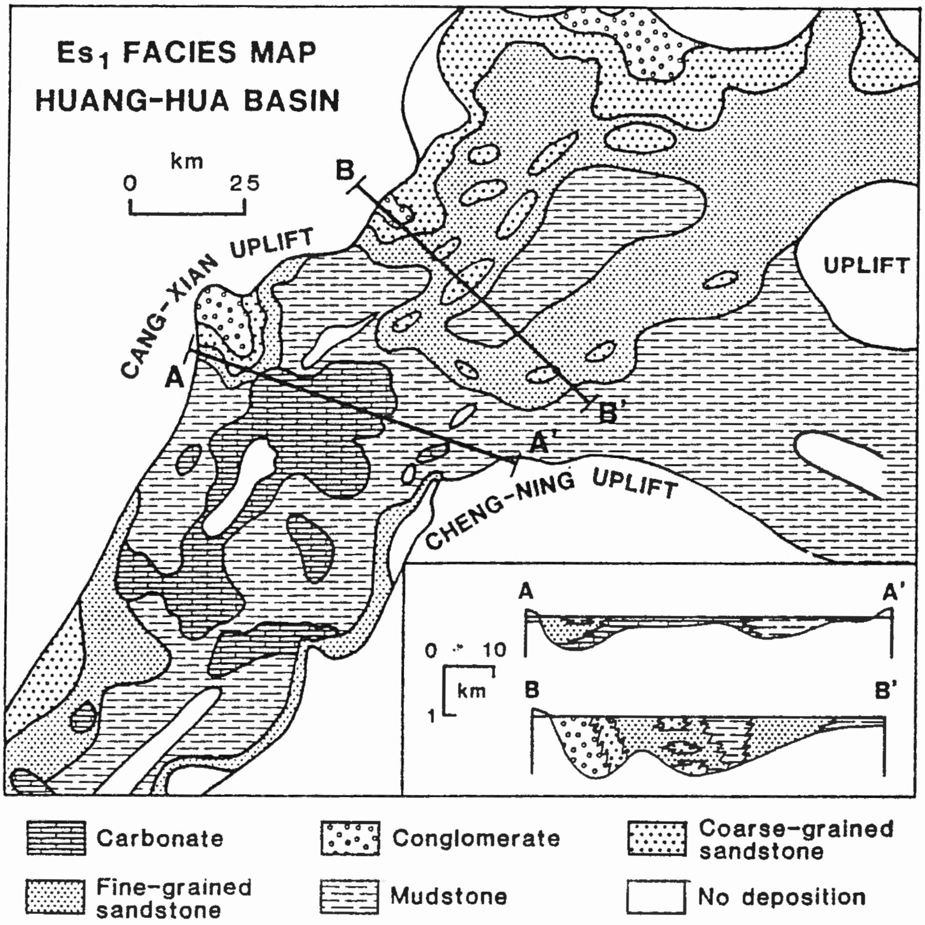


Table 5-6. Example of data set for exploration risk analysis.

Closure	Reservoir facies	Migration	Source
1	0	1	1
0	1	1	1
1	1	0	1
1	1	1	1

Figure 5-5. Distributions of area of closure, reservoir thickness, porosity and trap fill for the conceptual play. The solid line indicates the distributions perceived by geologist and the dots indicate the result of lognormal approximation (After Lee and Wang, 1983a).

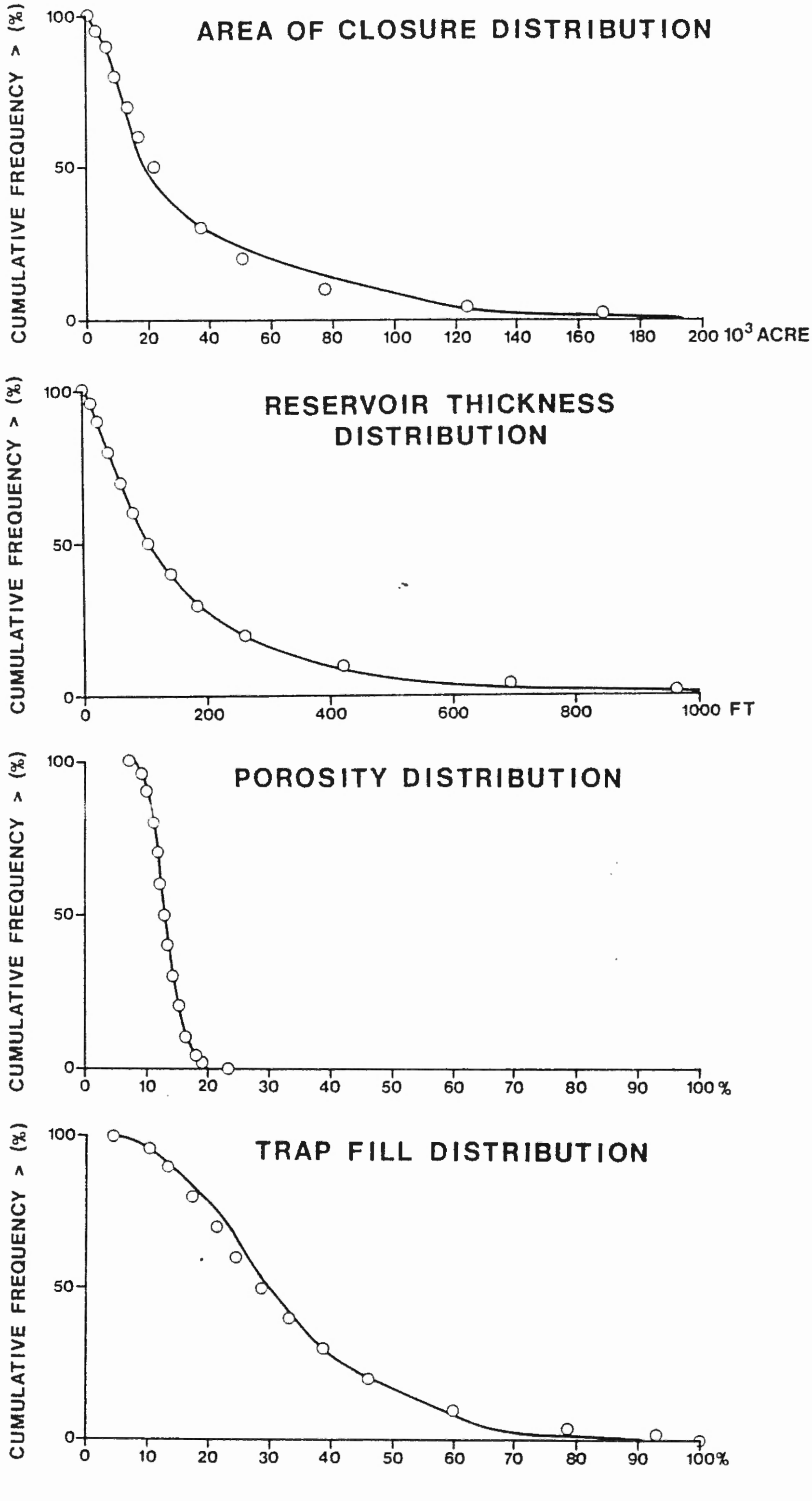


Figure 5-6. Pool size distribution derived by the result of Monte Carlo simulation (solid line) and lognormal approximation (circles) (After Lee and Wang, 1983b).

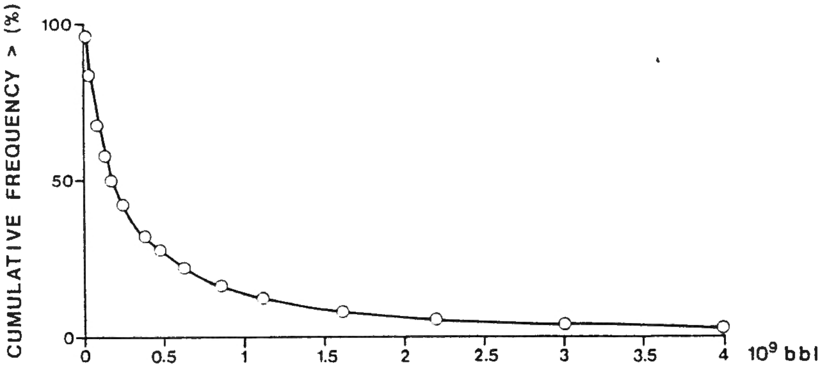


Table 5-7. Exploration risk for the conceptual play.

Geological factor	Marginal probability	Case I	Case II
Presence of closure	0.95	prospect	prospect
Presence of facies	0.90	prospect	prospect
Adequate timing	0.95	play	play
Adequate seal	0.80	prospect	prospect
Adequate source	0.75	prospect	play
Adequate preservation	0.80	prospect	play
Overall play level risk		0.95	0.57
overall prospect level risk		0.41	0.68
Exploration risk		0.39	0.39

Figure 5-7. Number-of-prospects distribution for the conceptual play (After Lee and Wang, 1983a).

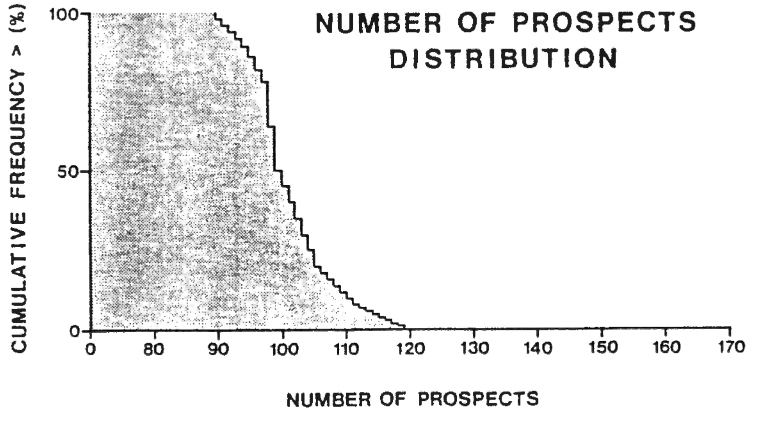


Table 5-8. Number-of-pools distribution for the two cases.

Upper percentile	Number of pools Case I	Case II
0.95	0	0
0.90	33	0
0.75	37	0
0.57	42	0
0.50	42	62
0.25	46	71
0.10	50	78
0.00	80	102

Table 5-9. Play potential distributions for the two cases.

Upper percentiles	Play potential (B bbl) Case I	Case II
0.90	1.22	0.00
0.80	1.57	0.00
0.70	1.82	0.00
0.60	2.05	0.00
0.55	2.29	2.29
0.50	2.28	2.80
0.40	2.55	3.42
0.30	2.86	3.96
0.20	3.30	4.61
0.10	4.05	5.59
Mean	2.50	2.50
Standard deviation	1487	2568

Table 5-10. Pool-size-by-rank for Case I.

Rank	Pr	Mean	S. D.	Upper percentile				
				95	75	50	25	5
1	0.95	782	1008	183	329	522	883	2157
2	0.95	349	244	122	199	286	422	782
3	0.95	231	129	93	145	200	280	469
4	0.95	171	85	75	114	154	209	331
5	0.95	136	62	62	93	124	165	253
6	0.95	112	49	53	78	103	136	203
7	0.95	95	40	45	67	87	114	168
8	0.95	81	33	39	58	75	98	142
9	0.95	71	28	35	51	66	85	123
10	0.95	62	24	30	45	58	74	107
11	0.95	55	21	27	40	51	66	94
12	0.95	49	19	24	35	46	59	83
13	0.95	44	17	21	32	41	53	75
14	0.95	39	15	19	28	37	47	67
15	0.95	35	14	17	26	33	43	61
16	0.95	32	12	15	23	30	39	55
17	0.95	29	11	14	21	27	35	50
18	0.95	26	10	12	19	25	32	46
19	0.95	24	10	11	17	23	29	42
20	0.95	21.7	8.9	9.7	15.4	20.5	26.7	38.1
21	0.95	19.8	8.2	8.6	13.9	18.7	24.5	35.0
22	0.95	18.1	7.7	7.6	12.6	17.0	22.3	32.2
23	0.95	16.5	7.1	6.7	11.3	15.4	20.5	29.6
24	0.95	15.0	6.7	5.8	10.2	14.1	18.8	27.3
25	0.95	13.6	6.2	5.1	9.2	12.7	17.2	25.2
26	0.95	12.4	5.9	4.4	8.2	11.6	15.7	23.2
27	0.95	11.3	5.5	3.8	7.3	10.5	14.4	21.4
28	0.95	10.2	5.2	3.2	6.5	9.5	13.1	19.8
29	0.94	9.3	4.9	2.7	5.8	8.6	12.1	18.3
30	0.94	8.4	4.6	2.3	5.1	7.8	11.0	16.9
31	0.93	7.7	4.5	1.9	4.5	7.0	10.1	15.6
32	0.92	6.9	4.0	1.6	4.0	6.3	9.2	14.4
33	0.90	6.3	3.8	1.3	3.5	5.7	8.4	13.3
34	0.88	5.7	3.5	1.1	3.1	5.1	7.7	12.3
35	0.86	5.2	3.3	1.0	2.7	4.6	7.0	11.4
36	0.82	4.7	3.1	0.8	2.4	4.2	6.4	10.6
37	0.78	4.3	2.9	0.7	2.1	3.8	5.9	9.8
38	0.74	3.9	2.7	0.6	1.9	3.4	5.4	9.1
39	0.68	3.6	2.5	0.6	1.7	3.1	5.0	8.5
40	0.63	3.3	2.4	0.5	1.6	2.8	4.6	7.9
41	0.57	3.1	2.2	0.5	1.4	2.6	4.2	7.3
42	0.51	2.9	2.1	0.4	1.4	2.4	3.9	6.9

Note: Pr : Probability for having r pools.

Table 5-11. Pool-size-by-rank for Case II.

Rank	Pr	Mean	S.D.	Upper percentile				
				95	75	50	25	5
1	0.57	1030	1219	273	464	713	1170	2738
2	0.57	488	311	191	294	409	585	1043
3	0.57	334	168	150	222	296	401	646
4	0.57	257	113	125	179	233	307	468
5	0.57	209	84	107	150	193	249	366
6	0.57	176	66	93	130	164	209	300
7	0.57	152	54	82	113	142	179	253
8	0.57	133	46	74	101	125	157	218
9	0.57	118	39	66	90	112	139	191
10	0.57	106	34	60	81	101	124	169

Note: Pr : Probability for having r pools.

Figure 6-1. Diagram showing different levels of feedback in the process of a petroleum resource evaluation.

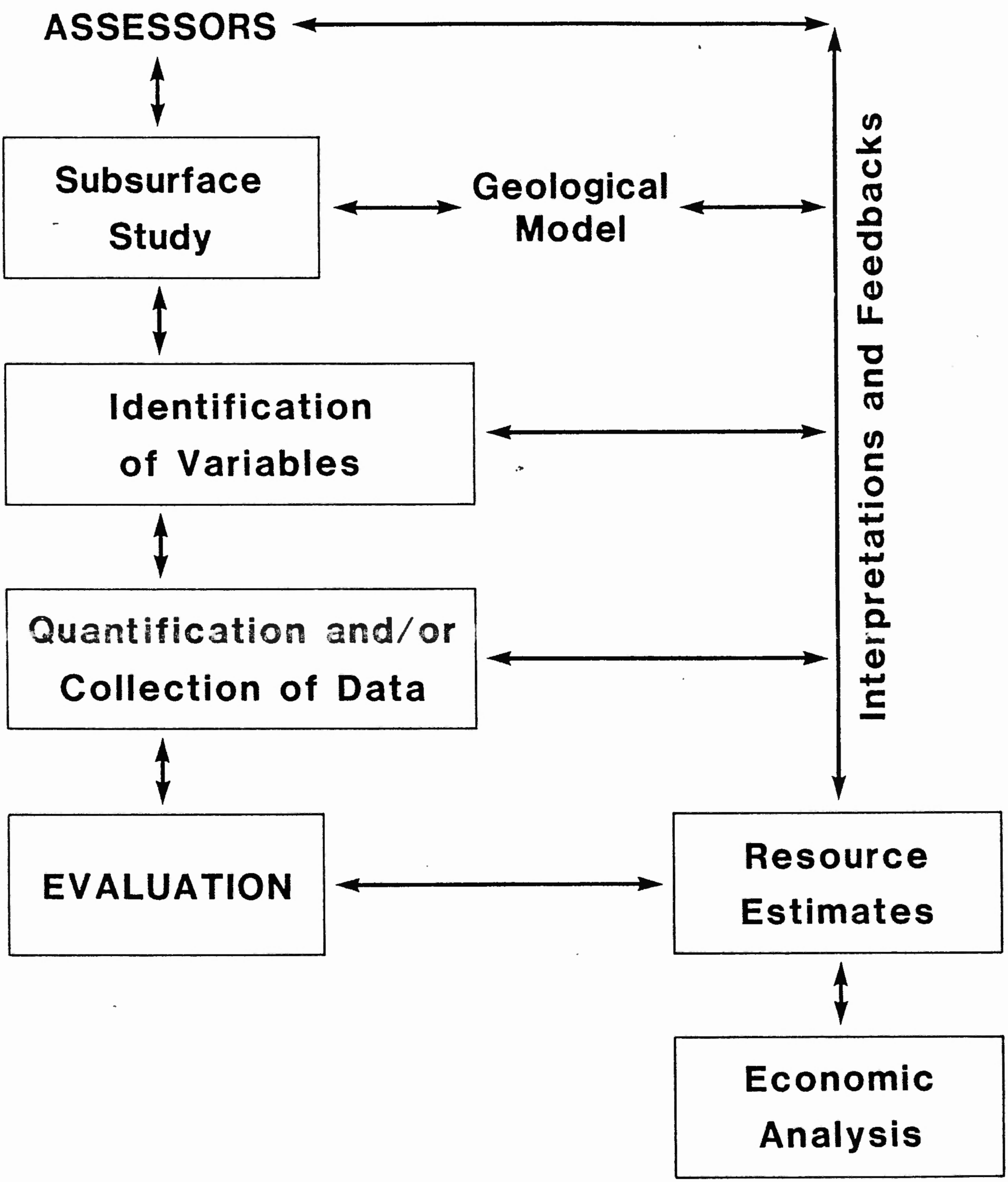


Table 6-1. Petroleum resource assessment record sheet

I. General Information

Country: ________________ Geological province: ________________
Basin: ________________ Play: ________________
Geologists: ________________ Assessors: ________________
Date of assessment: ___/___/____
Date of completion of the sheet: ____/___/____;

II. Basic Information Availability

A. Maps — Source
Structural ________________
Isopach ________________
Facies ________________
Oil & gas pools location ________________
Geophysical ________________
Cross-section ________________
Logs ________________
Others ________________

B. Comments on information availability

C. Level of knowledge concerning this play
No drilling and no seismic _____
No drilling but with seismic _____
Early stage--immaturely explored ______
Intermediate stage--fairly well explored _______
Late stage--maturely explored _________
Completely explored ___________

D. Stratigraphy of each formation or pay

Formation names
Age ________________ Lithology ________________
Thickness ____________ Organic type ________________
Sedimentary environment ________________
Heat flow ____________ Temperature gradient ____________
Oil window ___________ Gas window ________________
Pressure: Normal __________ Abnormal __________
Other comments ________
Indications of oil and/or gas ________

III. Statistics of the Play

A. Play area ____________ $mile^2$ or km^2
Play area explored ____________
Play producing ____________

Play volume ____________ $mile^3$ or km^3
Play volume explored ____________
Play volume producing ____________

B. Reservoir data

	Oil			Gas		
	Min	Average	Max	Min	Average	Max
Pool area ha	_____	_______	_____	_____	_______	_____
Net Pay m	_____	_______	_____	_____	_______	_____
Porosity	_____	_______	_____	_____	_______	_____
Water saturation	_____	_______	_____	_____	_______	_____
Depth m	_____	_______	_____	_____	_______	_____
Recovery factor	_____	_______	_____	_____	_______	_____

C. Hydrocarbon volume

In-place oil volume ________________ 10^6 m^3 or MM bbl
In-place gas volume ________________ 10^6 m^3 or Bcf or Tcf

Primary oil reserve ________________
Primary gas reserve ________________

Enhanced oil reserve ________________
Enhanced gas reserve ________________

Cumulative oil production ____________
Cumulative gas production ____________

D. Drilling history

Number of wells penetrating the play ________________
Number of exploratory wells ________________
Number of exploratory
interpreted as true test wells ________________
Number of development wells ________________

The mean recurrence time for dry well ________________
The mean recurrence time for oil well ________________
The mean recurrence time for gas well ________________
The mean recurrence time for oil & gas well ____________

Exploration risk ________________
Ratio of producing area / play area ________________

Figure 6-2. Unconditional and conditional distributions of the number-of-pools for different play resources (after Lee and Wang, 1983b).