Statistical Issues in the Assessment of Undiscovered Oil and Gas Resources

by

Gordon M. Kaufman

MIT-CEEPR 92-010WP September 1992

STATISTICAL ISSUES IN THE ASSESSMENT OF UNDISCOVERED OIL AND GAS RESOURCES

Gordon M. Kaufman*

ABSTRACT - For Dave Wood

Prior to his untimely death, my friend Dave Wood gave me wise counsel about how best to organize a paper describing uses of statistics in oil and gas exploration. A preliminary reconnaissance of the literature alerted me to the enormous range of topics that might be covered. Geology, geophysics with particular attention to seismology, geochemistry, petroleum engineering and petroleum economics--each of these disciplines plays an important role in petroleum exploration and each weaves statistical thinking into its fabric in a distinctive way. An exhaustive review would be book length. Dave and I agreed that a timely review paper of reasonable length would:

(1) Illustrate the range of statistical thinking of oil and gas exploratists.

(2) Concentrate on topics with statistical novelty, show how statistical thinking can lead to better decision making and let the reader know about important controversies that might be resolved by better use of statistical methods.

(3) Focus on topics that are directly relevant to exploration decision making and resource estimation.

In response to Dave's sensible suggestions, the Department of Interior's 1989 assessment of U.S. undiscovered oil and gas will be a tour map for a short trip through a large territory of statistical methods and applications. Were he here to review this review, I know that it would be better than it is.

INTRODUCTION

Statistical methods in oil and gas exploration serve as aids to exploration decision making at three levels: prospect, play and basin. ["Prospect" and "play" will be defined shortly.] In addition, statistical analyses of the petroleum potential of large geographic regions guides state and federal regulatory, tax and policy analysis. While there is overlap, statistics as used in exploration decision making has features distinct from statistical analysis in support of aggregate appraisal of the petroleum potential of petroleum basins and larger geographic units.

Oil and gas company managers are responsible for finding and producing oil and gas. They wish to harness statistical thinking to a plow that turns up answers to questions like these: **Where** in a petroleum basin are geologic anomalies that might contain oil and gas located? **How many** such anomalies are present and **what** is their size distribution? **What is the likelihood** that a particular **anomaly** contains producible oil and gas? If an anomaly does contain producible **hydrocarbons**, **how much can be profitably produced**?

A geological anomaly that may contain producible hydrocarbons and is a target for drilling is called a "prospect". Prospects are naturally grouped into classes or sets. A set composed of geologically similar prospects within

*Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139-4307. Supported in part by the Center for Energy and Environmental Policy Research.

a petroleum basin is called a "petroleum play". More precisely a play is a family "...of prospective and/or discovered petroleum pools that share a common history of oil or gas generation, migration, reservoir development and trap configuration [Podruski et al. (1988), White (1980)]. Deposits and prospects in a petroleum play possess common measurable characteristics and so constitute a finite population of objects to which finite population sampling theory and statistical analysis may be applied. Geochemical, geological and geophysical methods are deployed in order to measure characteristics of petroleum plays in a fashion that tells exploration managers where to drill and what the return to exploratory effort may be. **In sum, the concept of a petroleum play is central to exploration strategy and the set of deposits and prospects in a play is a natural object for statistical analysis.**

Projections of aggregate amounts of oil and gas remaining to be discovered in basins and larger geographic areas help corporate management make **strategic** exploration decisions such as how to allocate corporate funds available for exploration onshore vs. offshore and within the U.S. vs. outside the U.S. More broadly, assessments at the state, regional and national level are used by policy analysts, government officials and politicians to formulate energy policy. Whether or not to explore in environmentally sensitive areas like Georges Bank, the Arctic National Wildlife Refuge and offshore California depends in part on the magnitudes of producible hydrocarbons that might be found in these areas. State and federal energy tax policies and regulatory rulings also depend on such projections. Because most basins in the lower forty-eight states are mature producing basins, a large fraction of fields remaining to be discovered in them are at the economic margin. As a consequence, exploration activity is very sensitive to wild swings in oil prices like those that have afflicted the U.S. since 1973. All of this has important implications for effective uses of statistical methodology in appraisal of aggregate amounts of petroleum remaining to be found in U.S. basins. Controversy about statistical methods employed to make projections of U.S. undiscovered oil and gas has hummed merrily along since the first national assessment was done in the early 1900's by a single geologist.

About every four years geologists at the U.S. Geological Survey (USGS) and at the Minerals Management Service (MMS) appraise how much undiscovered oil and gas remains in the United States. The USGS is responsible for onshore estimates and the MMS is responsible for offshore estimates. The most recent (1989) national assessment was controversial because estimates of mean amounts of undiscovered oil and undiscovered natural gas were respectively about 33% and 41% lower than estimates of these quantities made in the 1981 national assessment (Page ii of Undiscovered Oil and Gas in the United States, Mast et al. (1989)) Some industry experts and state geologists disagreed with the 1989 USGS and MMS estimates and criticized the methods employed to create these projections. [See Undiscovered Oil and Gas Resources: An Evaluation of the Department of Interior's 1989 Assessment Procedures, National Academy Press, Washington D.C. 1991].

The 1989 U.S. oil and gas assessment is a useful template for examining statistical issues in oil and gas exploration. Many are highlighted in the 1991 NAS/NSF review of the 1989 assessment. Among them are: (1) The importance of periodic audits of data quality;(2) How to design finite population models for encountered and size biased data composed of samples of oil and gas field magnitudes in order of discovery; (3) Comparison of strengths and weaknesses of schemes for estimating both the number of undiscovered deposits in a petroleum play and the size distribution of these undiscovered deposits; (4) Investigation of the shape of the empirical distribution of all deposits in a petroleum play from a sample that is subject to economic truncation as well as being encountered and size biased; (5) How to use spatial point process data describing wildcat well outcome data to structure dependencies between well outcomes--and why an appropriate characterization of such dependencies is of practical importance; (6) How to use probabilistic models of oil and gas deposit discovery as guideposts for calibrating expert judgement about undiscovered oil and gas fields. A statistical tour of these particular topics covers but a very small fraction of active applications of statistics to oil and gas exploration. Inverse problems in seismology and stereology, uses of multivariate techniques in geochemistry and geology, trend surface analysis and kriging are examples of what is left out. The literature is large: for example, Garrett's 1985 review of uses of computers and statistics in petroleum and hardrock geochemical prospecting references over 230 papers.

DATA QUALITY AUDITS

Data bases describing geologic settings at basinal and more disaggregate levels are one indispensable component of oil and gas resource analysis. Dry hole and discovery well data bases are another. Variables that determine reservoir productivity are a third. Examples are zones classified by rock type (sandstone, carbonate, etc.),trapping mechanism, areal extent, thickness, porosity and permeability. Pre-drilling geophysical and geological search costs together with well drilling and production cost data are coupled with geological and reservoir engineering data to make drilling decisions. According to the NAS/NSF review,"Assessors [of undiscovered oil and gas] commonly understate the inadequacy of their data, if they mention it at all." [op. cit. p.28]

As no single publicly accessible, comprehensive data base describing United States petroleum basins, fields and wells in sufficient detail for effective nationwide resource analysis exists, the USGS employed many different data sources of varying completeness and quality in its assessment. To appreciate the magnitude of the task they faced the reader must recognize that there are more than 100,000 documented U.S. oil and gas reservoirs and over 3,000,000 oil and gas wells have been drilled in U.S. oil and gas territories [Brashear et al. (1989)]. Well and reservoir data is generated by oil and gas field operators and collected and processed by petroleum companies, professional societies such as the American Association of Petroleum Geologists, the American Petroleum Institute, the American Gas Association, the Gas Research Institute, state geological surveys and federal agencies as well as by oil and gas data marketing firms. Each agent has his own agenda so that the data collection and processing methods employed differ from agent to agent. As a consequence, reconciliation of oil and gas data bases is often a major undertaking.

Mirroring finding of a review by the American Association of State Geologists of the geological information used by the USGS in the national assessment, the NAS/NSF review panel recommended that the USGS institute formal data audit procedures: "Good statistical practice in resource appraisal begins with an audit of the data used in the assessment...A data audit is a framework for: (1) evaluating the data's accuracy and completeness; (2) identifying areas where the data require improvements; and (3) providing explicit measures of the data's quality to assessment users." [op. cit. p 58] The Energy Information Administration's procedures for insuring the quality of the data it gathers, processes and distributes is an excellent model of how to do this.

ESTIMATING THE EMPIRICAL SIZE DISTRIBUTION OF UNDISCOVERED DEPOSITS IN A PETROLEUM PLAY

In 1989, for the first time, the USGS adopted individual play analysis on a national scale. Two hundred and fifty plays in U.S. oil and gas provinces were defined with the help of a proprietary (NRG Associates) oil and gas field data base. Defining a play is an exercise in the interpretation of complex quantitative and qualitative data. There is broad agreement about how to define a play, but in practice disagreement is frequent. A typical general definition is given in Box 2.1 from the NAS/NSF review:

BOX 2.1 How to Define a Play

In defining a play as "a group of geologically similar prospects having basically the same source-reservoir-trap controls of oil and gas," White stressed the importance of geologic commonality in play definition(White, 1980). Achieving geologic commonality is essential for insuring that each set of reservoirs and prospects being evaluated is as homogeneous as possible. Yet, in practice, achieving commonality is not a straightforward exercise. There is no single operational formula for play definition that fits all geologic settings.

In the great majority of cases, plays are limited to a single formation, because each formation is associated with a distinct set of reservoir characteristics. Yet in some areas, notably California and some Rocky Mountain basins, a single play may encompass several producing formations, the play being essentially defined by a structural trend. Within a single formation, depositional system can be a key parameter in play definition, because differences among types of depositional systems are associated with differences in reservoir size distributions and patterns of reservoir location. Differences in petroleum sources, in thermal maturity within a source, and in migration history as they affect petroleum type and characteristics can also be important factors in determining play boundaries.

While both the NAS/NAS review panel and the Association of State Geologists criticize the imprecision of USGS play definition, they applaud their use of play analysis.

Knowledge of magnitudes of discovered deposits (measured in barrels of oil and/or MCF of gas in place) and discovery dates allow known deposit magnitudes to be arranged in order of discovery. Ordered sequences of deposits magnitudes are building blocks for **discovery process models**, probabilistic models built from assumptions about the fashion in which oil and gas deposits are discovered. Three important features of discovery data are:

* Deposit magnitude data is **encountered** and **size biased.**

* Deposit discovery is akin to **sampling without replacement** elements of a finite population of objects.

* The magnitude of a discovered deposit is **known with precision late** in its production life.

Discovery process models are devices for estimating the empirical distribution of in place deposit magnitudes in a play and projecting magnitudes of future discoveries in order of occurrence. These are their principal raisons d'etre. Once such inferences are made, the empirical distribution of undiscovered deposit magnitudes is then easily calculated. Methods designed to make inferences about the empirical distribution of magnitudes of the complete population from an incomplete sample must take in account the effects of size bias and sampling without replacement. Standard statistical methods for inferring the shape of empirical distributions from samples of independent identically distributed uncertain quantities are not applicable until almost all deposits in a play have been discovered--at which point all the needed data is already in hand!

That larger deposits in a play are, on average, found early in the exploration history of a play has been part of exploration folklore since the early days of oil and gas exploration. Arps and Roberts (1958) were the first to build a model that incorporated this idea. In addition to deposit magnitudes in order of discovery, their model requires wildcat drilling history as input: deposit magnitudes are attached to the sequence of exploratory well successes and failures in a play. In order to apply their model the finite population of deposits deposed by nature in a play must be partitioned into discrete size classes. Each size class is then regarded as functionally independent of all other size classes, so that no a priori assumptions relating the in place numbers of deposits in size classes is necessary. This flexibility is an advantage. However, it is counter-balanced by a disadvantage: interactions between discoveries in different size classes are ignored. Interactions of this type are a principal feature of discovery process models that succeed Arps and Roberts model. Drew, Scheunemeyer and Root have vigorously pursued development of the Arps-Roberts paradigm [Drew, Scheunemeyer and Root (1980a), Root and Scheunemeyer (1980b)] and applied it to discoveries in the Denver-Julesburg Basin, the Permian Basin [Drew, Schuenemeyer and Bawiec (1979) and the Western Gulf of Mexico [Drew, Schuenemeyer and Bawiec (1982)].

In 1975 Kaufman, Balcer and Kruyt proposed a probabilistic discovery process model based on two assumptions: let $U=\{1,2,...,N\}$ be a set of labels of N deposits and $A=\{a_1,a_2,...a_N\}$ be an associated set of deposit magnitudes. Then

I Elements of A are a realization of a sample of N independent identically distributed random quantities each with range $(0,\infty)$ and cumulative distribution function $F(x)$, $x > 0$.

II A realization of elements of A generated according to I is sampled proportional to magnitude and without replacement.

Knowledge of drilling outcomes are not required, so this model is more parsimonious than the Arps-Roberts model. Because the model represented by Assumption II generates discovery magnitudes in order of discovery according to a probabilistic scheme that simultaneously employs all undiscovered deposits, an artificial division of magnitudes into discrete size classes as required by the Arps-Roberts model is not necessary. The interplay between all deposits in A is captured by the model. In accord with early work on oil and gas size distributions, Kaufman et al. assumed that the super-population process generating values of A, the "super-population process", is lognormal, although this is not a necessary assumption.

A graph of marginal expectations of discovery magnitudes in chronological order produced by Assumptions I and II is called the **discovery decline curve.** The discovery decline curve mirrors the empirical observation that as discovery progresses, magnitudes of discoveries tend to decrease. An exact integral representation of these marginal expectations is given by Barouch and Kaufman (1976) for any super-population process with bounded expectation. An interesting feature of this representation is that it can be interpreted as the expectation of an order statistic generated by N iid random variables whose distributions are composed from the LaPlace transform of the super-population distribution. Figure 1 shows an exploration decline curve computed for a lognormal super-population process for N=150,300 and 600 deposits in place.

[Figure 1 here]

The important role played by exploration decline curves in the 1989 assessment procedure will be discussed in section 4.

Arps and Roberts adopted deposit area as a measure of magnitude in their work on the Denver-Julesburg Basin. Most of the deposits in this basin are stratigraphic lens traps not easily identified by seismic search. In this particular basin the area of a deposit is a natural measure of deposit magnitude. The assumption that the probability of discovery of a deposit is proportional to its area may be less justifiable in geological settings where different geological attributes may control discoverability. Then it is appropriate to define magnitude differently. Bloomfield, Deffeyes, Watson, Benjami and Stine (1979) suggested a simple generalization of II that allows inferences to be made about the degree to which a particular definition of magnitude may control discoverability: replace a_i in II with a_i^β where ß is a scalar independent of the labels 1,2,...,N. Although further generalizations are possible (see Wang and Nair (1988) for examples) this particular one gives enough addition modelling flexibility for most applications and expedites testing of the assumption that some version of proportional to size sampling obtains. When $\beta=0$, proportional to size sampling reverts to equal probability sampling of the labels 1,2,...,N without replacement. As $\beta\to\infty$, $a_1, a_2, \ldots, a_N$ are sampled in exact order of magnitude, largest to smallest, with probability one. MLE estimators of ß may not be very robust with respect to play definition. Adelman et al.(1983) showed that if ß is estimated for a mixture of two or more petroleum plays, the resulting estimate may be a drastically biased estimate of discoverability parameters of individual plays in the mixture.

The sampling scheme represented by Assumption II is familiar in the finite population sampling and sample survey literature where it is called "successive sampling of a finite population" [see the books by Hájek (1981) and Cassel, Sarndal and Wretman (1977) for example]. In contrast to the usual assumption made in the sample survey literature that the parameters a_1, $a_2, \ldots a_N$ are known functions of an auxiliary attribute (a ppswor sampling scheme), here the a_i's are unknown until observed. This feature distinguishes the problem of estimation of properties of unobserved elements of A from its sample survey counterpart and led to two distinct lines of work: development of methods of estimation for discovery process models based on assumption II alone and a parallel development of methods for models based on assumptions I and II. Call them the "finite population model" and the "super-population model" respectively.

Super-Population Estimation

The super-population process generating deposit magnitudes was assumed to be lognormal in initial work on super-population parameter estimation. This particular choice of shape was motivated by the first set of studies of empirical size distributions of deposits in petroleum plays. The shape of oil and gas deposit size distributions has been a subject of controversy ever since and we will review the debate after a discussion of methods of estimation of discovery process model parameters. This is a necessary prelude, as methods for inferring the shape of oil and gas deposit size distributions should account for how deposits are discovered.

Maximum likelihood methods for joint estimation of lognormal super-population parameters and the total number N of deposits in a play were studied by Barouch and Kaufman (1977), Barouch, Kaufman and Nelligan (1983) and applied to North Sea and Canadian data. Lee and Wang (1983a, 1983b) continued this line of development with numerous applications to the Canadian Western Sedimentary Basin. Their computational scheme figures prominently in the Geological Survey of Canada's (GSC) undiscovered oil and gas assessment methodology. The GSC's periodic reports of Canadian petroleum endowment present probabilistic assessments of future discoveries in individual plays derived by application of Lee and Wang's treatment of super-population models. The use of deposit size rank plots is an interesting feature of this method. MLE estimates of super-population parameters and the number N of deposits in place are used to compute the marginal distributions of all N deposit magnitudes in order of discovery. (These marginal distributions are conditioned on the observed data by the MLE estimation procedure.) Then each discovered deposit size is assigned a pool rank by pairing that deposit size to a marginal distribution.(For example, match the kth largest discovery to the closest median value of a marginal distribution.) After matching of $n<N$ discoveries to N marginal distributions, the range of possible values of each of the N-n undiscovered deposits is recalculated so as to restrict this range to lie within the range of the two actual discovery sizes closest in rank to it. Lee and Wang call this reduction in predicted range "conditioning on the match" (Podruski et al. (1988) p 12.). Figure 2 from the GSC 1987 assessment is an example of the result for the Slave Point-Golden play.

[Figure 2 here]

Wang and Nair (1986) showed how to do parametric MLE estimation using the Expectation-Maximization algorithm and how to estimate a multivariate version of the discoverability parameter suggested by Bloomfield et al. In a related paper that builds on their earlier work Wang and Nair (1988) do non-parametric MLE estimation of the shape of the super-population distribution. The advantage of this approach is that it does not require any assumption about the shape of the super-population distribution. However, a non-parametric estimate of distributional shape assigns mass only to observed data points, so unless sample size is large, the estimate will not provide convincing evidence for or against lognormality or any other reasonable distribution shape. They illustrate this estimation procedure with an application to the Rimbey-Meadowbrook reef play in Central Alberta at the point of 23 discoveries.

Geologists at the Australian Bureau of Mineral Resources use a statistical method for assessing undiscovered oil and gas that builds on Meisner and Demirmen's model of discovery [Meisner and Demirmen (1981)]. The latter authors assume that the discoverability parameter (the scalar assigned as a power of magnitude in the finite population sampling model represented by Assumption II above) is a decreasing linear function of the number of wildcats drilled. They call a reduction in size of discovery with advancing exploration effort the "creaming phenomenon" and designed the decrease in the value of the discoverability parameter to capture both a decline in discovery size and changes in the probability of success as exploratory drilling increases. The probability distribution of size of discovery at any well is in the form of a size biased distribution for discovery size that would result from letting the number of fields in place go to infinity in the super-population models presented earlier. Consequently the decline of mean discovery size with increasing drilling is,in their model, due solely to decline in the discoverability parameter. The shape of the in-place size distribution is assumed to be lognormal. This assumption coupled with a linear decline in the discoverability parameter allows the expectation of the log of the size of the nth discovery to be expressed as a linear function of the cumulative number of wells drilled. Following Meisner and Demiren's lead, Foreman and Hinde (1985) claim that log field size and discovery number are linearly related for some of Australia's petroleum basins and investigate the quality of fit of such a relation using Australian field size data. This claim may be interpreted as saying that the regression of log deposit size on discovery number is linear. Foreman and Hinde claim to have tested the quality of fit of this relation using monte carloed values from a successively sampled finite population of deposits, but it is hard to follow what they did. Their assertion is testable provided that it is coupled with a precise specification of the covariance structure of residuals.

None of the estimation procedures discussed thus far explicitly incorporate expert judgment. All are "objective" in the sense that estimates are calculable from observed data alone. In a sharp departure from evidentially objective estimation the USGS used discovery process models in its 1989 assessment as an aid to subjective appraisal of petroleum play potential. This aid is based on Houghton's claim that as exploration advances, changes in the shape of discovery magnitude distributions can be reasonably fit with a Pareto distribution equipped with a shift parameter and truncated right tail. He called this three parameter distribution a "truncated shifted Pareto or TSP distribution". Marginal distributions of discovery magnitudes in order of observation centered at their respective expectations--the exploration decline curve values-- are interpretable as distributions of residuals about a non-linear regression function. Houghton's assumption raise two unanswered questions: (1) What super-population shape, if any, leads to a TSP for residuals about the exploration decline curve? (2) Is the TSP a robust approximation to the shape of decline curve residuals? The first question remains unanswered, but most probably there is no super-population process that leads to TSP-type residuals about the exploration decline curve. At best the TSP distribution may be a reasonable approximation. The NAS/NSF review emphasizes the importance of answering the second question: "Houghton studies the quality of the fit of a TSP distribution to the discovery history of a very mature play, the Minnelusa play, in the Powder River Basin and compared this fit with some alternatives [Houghton (1988)]. Houghton's study appears to be the only attempt, prior to the national assessment, to validate the model on which the USGS assessment procedure is based. Given the geological diversity of the approximately 250 plays included in the national assessment, a much more aggressive effort to validate the procedure prior to its adoption was warranted." [op. cit. p.71]

USGS geologists participating in the 1989 assessment were presented with three chronological size distributions--TSP distributions--fit to the first, second and third thirds of discovered deposit magnitudes in a play. Then each was asked to appraise the size distribution of undiscovered deposits by choosing parameters of a TSP that best fit her/his judgements the size distribution of remaining deposits. A "...notable feature of this assessment protocol is that it employs properties of an objective first principles model of discovery as an aid in making subjective assessments **without** exploiting the predictive capabilities of such a model." [op. cit. p.71] Here is an alternative that more fully exploits discovery process modelling: use objective discovery data (deposit magnitudes in order of observation, for example) to project the exploration decline curve; present this projection to the geologist for modification in light of evidence not captured by the model. Another alternative is to ask the geologist to assess the size distribution of in-place deposits, use this distribution to generate an exploration decline curve and then ask for modifications in the assessment of the in-place size distribution based on the geologist's interpretation of the projected exploration decline curve.

Finite Population Estimation

It is possible to set aside the super-population assumption and base inferences about future discoveries on a model composed from Assumption II alone. Three types of estimators of properties of unobserved elements of a set of deposit magnitudes in a play have been studied: maximum likelihood, conditional maximum likelihood and moment type estimators.

In the first of three papers on successive sampling theory and its applications, Gordon (1981) motivates the connection of the finite population sampling scheme represented by Assumption II to deposit discovery by suggesting that this scheme be visualized as a dart board game like that in Figure 3 below:

[Figure 3 here]

In his 1983 paper he showed that a successive sampling scheme can be represented in terms of order statistics of independent but non-identically distributed random variables. This important fact plays a key role in the study of asymptotic (large sample size) properties of successive sampling schemes and in addition suggests the functional form that moment type estimators of successively sampled finite population properties should take. If the probability that a generic population element will be included in a successive sample of size n<N is known a priori, then classical unbiased finite population estimators such as the Horvitz-Thompson estimator [see Hájek (1981) or Cassels et al. (1977)] could be applied to estimate finite population characteristics from an incomplete sample. Unfortunately, these inclusion probabilities depend on population magnitudes which are unknown a priori in the successive sampling scheme represented by Assumption II, so they must be approximated. Gordon shows how to do this by "moment matching"; i.e. matching approximations of expectations of sample characteristics and solving pairs of matches to construct estimators of inclusion probabilities. These approximate inclusion probabilities are then used to do Horvitz-Thompson type unbiased estimation of population characteristics. A full account appears in Gordon (1992). In their examination of finite sample properties of Gordon's estimators, Barouch, Chow, Kaufman and Wright (1985) provide a simple sufficient condition--the average of an "early" fraction of the complete sample of discovery magnitudes must be larger than the complete sample average--that guarantees an odd number of solutions to moment matching pairs, but were unable to prove uniqueness. Fortunately, extensive monte carlo simulation suggests that this condition guarantees uniqueness.

Andreatta and Kaufman (1986) adapt Murthy's (1957) estimator, a close relative of Horvitz and Thompson's estimator, to successive sampling in a different way. If any one population characteristic such as the number N of deposits, the sum of all deposit magnitudes or a fractile of the empirical distribution of in-place deposits is assumed to be known with certainty, then this knowledge be used to compute an estimate of inclusion probabilities from an incomplete sample of the population. They call this "anchored estimation", the known population characteristic being the "anchor". An application to North Sea data partitioned into seven size classes recovers MLE estimates for each of these size classes so closely as to suggest a tight link between conditional (on the anchor) MLE and unbiased estimation via anchoring.

Rabinowitz (1991) extends the successive sampling model to include the effects of drilling dry holes. He defines the "area of influence" of a dry hole to be a fixed area within which no deposit may be located. Drilling locations are supposed to be sufficiently sparse so that "clumping" of areas of influence and/or deposit areas may be ignored. Exploration drilling is assumed to diminish prospective exploration acreage by the sum of areas of influence of dry holes and of discovered deposits. Rabinowitz shows how to do both unbiased and partial maximum likelihood estimation of properties of in place deposits, provides confidence intervals for his estimators and presents sufficient conditions for asymptotic normality of his estimators. (Loosely, the "partial likelihood" is that portion of the joint probability of realizing a sequence of discoveries and dry holes that includes only discovered deposit magnitudes in the numerator, but accounts for reduction of prospective acreage by both dry holes and discoveries in the denominator.) In a follow-up paper [Rabinowitz (1992)] he extends his analysis of partial likelihood estimation to encompass a wider class of functionals and develops an adaptive approach to estimation.

The latest word on MLE for successive sampling from a finite population is provided by Bickel, Nair and Wang (1992). For successive sampling from a finite population divided into a fixed number of discrete size classes they pin down the structural similarity of unbiased Horvitz-Thompson estimators and MLE and show how anchored estimation is related to conditional MLE. Further insight comes from their demonstration that a successive sampling scheme can be embedded in a Poisson sampling scheme, a probabilistic "dual" to Gordon's embedding of successive sampling in an exponential process.

THE SHAPE OF DEPOSIT SIZE DISTRIBUTIONS

Empirical distributions of deposit attributes such as area, rock volume, hydrocarbons in place and recoverable oil and gas are positively skewed with fat right tails. Among analytically tractable distributions, the lognormal distribution has been a popular choice of shape. It first appears in the oil and gas literature with Arps and Roberts (1958) study of the size distribution of oil and gas fields of lower Cretaceous stratigraphic traps in the Denver-Julesburg Basin. (A year earlier, Allais published a massive review of the mineral resources of the Sahara; lognormal size distributions play a central role in this study. Oil and gas size distributions were omitted from publication because of the Algerian question). These studies prompted Kaufman (1963) and Drew and Griffiths (1965) to examine lognormal fits to several U.S. oil and gas data sets. In 1969, McCrossan examined the fit of the lognormal distribution to twenty-eight individual petroleum plays in Western Canada. With the exception of Arps and Roberts work, early attempts to infer the adequacy of the lognormal distribution did not take into account size bias and most sample sizes were moderate--less than 100. Furthermore, a monte carlo study of the shape of **discovered** fields generated by sampling proportional to size and without replacement [Kaufman et al. (1975)] shows that an incomplete sample, even if composed of a substantial fraction of all deposits in a play, is **not lognormal**. Early work did not rigorously confirm or reject the lognormal distribution as a viable candidate for the shape of oil and gas data.

In an excellent short review of the history of the lognormal distribution in the geological sciences, Bloomfield, Hudson, Kim and Watson (1979) emphasize the absence of a plausible theoretical explanation supporting its adoption as an oil and gas size distribution and outline technical problems that arise in testing the quality of fit of alternative shapes. As early as 1964, Prokhorov rigorously demonstrated the difficulty of distinguishing the lognormal distribution from other natural candidates--the gamma distribution and the stable distribution with parameters 1/2 and 1, for example--for samples of less than 100 independent identically distributed sample values.

An ideal data set for examining shape would have three attributes: (1) all possible deposits in a play have been discovered,(2) individual deposit magnitudes are accurately measured,(2) the number of deposits in the set is very large. While no such data set exists, at least one comes close to meeting these criteria, the Lloydminster play in Alberta and Saskatchewan.Over 2500 deposits have been discovered and attributes of each measured according to a well designed protocol by McCallum, Stewart and Associates. Figure 4 displays a lognormal probability plot of the data taken from McCrossan, Procter and Ward (1981).

[Figure 4 here]

The large sample size permits study of a fractile range from .0024 to .99976. Andreatta, Kaufman, McCrossan and Procter (1988) applied a fractile method for fitting distributional shape proposed by Tukey (see Hoaglin and Peters (1983) for details and Hoaglin and Tukey (1992) for complements) and concluded that only the very largest and very smallest (one well fields) do not adhere to lognormal shape.

Most U.S. oil and gas provinces are mature and far down the exploration decline curve. As a result future discoveries in mature provinces are expected to be much smaller than discoveries early in their exploration histories. Returns to exploration effort may be expected to come in the form of "left-tail" discoveries. If only small deposits remain to be discovered, the rate of discovery will be very sensitive to changes in oil prices and discovery costs. This essential economic fact renewed interest in investigation of left tail shape. Schuenemeyer and Drew (1983) suggest that small oil and gas fields may be unreported because they are not profitable to produce. As a consequence, economic imperative truncates reported size distributions. They provide evidence of economic truncation of reported deposit sizes in the Permian Basin, Denver-Julesburg Basin and Gulf of Mexico, apply the Arps-Roberts model to estimate numbers of deposits in place in thirteen log base 2 size classes and use these estimates to impute the shape of an in-place size distribution. In sharp contrast to lognormality, Drew and Scheunemeyer conclude that the in-place size distribution is J-shaped and conjecture that it is log-geometric in form. A corollary of this study is that where lognormal fits appear reasonable, it is because of economic truncation [Attanasi and Drew (1985)].

Davis and Chang (1989) adopt a different approach in an analysis of oil and gas deposit size distributions of two mature regions, the Central Kansas Uplift and the Denver-Julesburg Basin. In order to avoid estimation bias caused by size biased sampling and left tail truncation of small discoveries that go unreported because they are not economical to produce, they adopt the following tactic: a truncation point designed to include all large and moderate size deposits is chosen and the quality of chi-squared fits of truncated lognormal and truncated Pareto distributions are computed. The left tail shape of in place deposits is tested by examining the quality of fit of lognormal and Pareto right tails to samples composed of moderate to large sized deposits only. They conclude that the best fit Pareto distribution is "wildly optimistic and predicts an unreasonably large number of large fields" and claim that it is unlikely that in-place size distributions--for Central Kansas and the Denver-Julesburg Basin at least--are not J-shaped. A vigorous debate ensued. Schuenemeyer, Drew, Root and Attanasi (1990) replied that the approach of Davis and Chang "...ignores wildcat wells and the dynamic nature of petroleum exploration...". As evidence they cite the steady rate of discovery of small fields throughout the exploration history of the Permian Basin and the character of the change in the empirical distribution of magnitudes of discovered fields: "Over time, the mode of the distribution migrates to the left, as many small fields and relatively few fields are discovered. This evidence suggests that the underlying distribution is not lognormal but is J-shaped". They argue in addition that observation of a tapered left tail of the distribution of discovered fields is caused by economic truncation. In a riposte Davis and Chang (1989) assert that the Permian Basin is not a suitable test case because "...exploration is significantly zoned by depth, and progressively deeper parts have opened as technically and economically accessible frontier areas at the same time that shallower parts were entering their maturity. This has not been a significant factor in the Denver-Julesburg Basin." They go on to repeat their commitment to modelling "...field size distributions directly, using only data from their most completely known parts" in place of interposing a discovery process model between the data and the fit. The last word of this debate is not yet written.

Between 1954 and 1963 Ahrens and several other hard rock geologists championed the lognormal distribution, declaring that it arose as a result of fundamental geochemical activity in the formation of mineral deposits. As Bloomfield et al. (op. cit.) point out, elevation of the lognormal distribution to "a fundamental law of geochemistry" was severely criticized by a suite of subsequent researchers, as much for the fashion in which statistical analysis of the data used to support the hypothesis was done as for the hypothesis itself. A convincing theoretical explanation in support of the lognormal or any other particular distributional shape grounded in geologic or geochemical laws is still absent. However, there is a compelling explanation for the ubiquitousness of the lognormal distribution grounded in Aitchison and Brown's quizzical observation: "It is a curious fact that when a large number of items is classified on some homogeneity principle, the variate defined as the number of items in a class is often approximately lognormal."[Aitchison and Brown (1957), p.27].

In an important paper entitled "Lognormal Genesis" that has not yet made its way to the mineral resource literature, Brown and Saunders (1981) rigorously demonstrate how the lognormal distribution can arise as a result of classifying objects into increasingly fine descriptive categories. "**THE FUNDAMENTAL CONCEPT IS ONE OF CLASSIFICATION**", according to Brown and Saunders, irrespective of the particular set of objects being classified. They make some mild assumptions about the nature of the classification scheme to be used and then demonstrate the following procedure leads to lognormality: (1) elements of a population of objects, each possessing a measurable attribute Q, are individually assigned to categories defined by a partitioning of the population according to distinct observable characteristics,(2) the quantity Q assigned to each object classified into a particular element of the partition is normalized by dividing by the total of all quantities Q of objects in that particular element of the partition, (3) the ratio of normalized Q and the relative (to the total number of objects) frequency of objects in that element is computed. This ratio is the normalized quantity Q in each element of the partition averaged over the relative frequency of objects in that particular element of the partition.(Mathematicians call such ratios "Radon-Nikodym derivatives".) Standardize the logarithm of this ratio by subtracting the mean of its logarithm and dividing the difference by the standard deviation of its logarithm. Brown and Saunders principal result is that as the number of characteristics becomes large (as the partition is successively refined) the distribution of this standardized log of the ratio approaches a normal distribution.

A successively refined descriptive partitioning of a set of objects can lead to a lognormal distribution of a measurable quantity possessed by each object in the set. This is a unifying idea superseding "... a multitude of models which incorporate a mechanism peculiar to a specific application in order to use the law of proportionate effect..." [Brown and Saunders, op. cit.] Brown and Saunder's model provides a plausible explanation of why the lognormal distribution appears to be a reasonable fit to oil and gas deposit data. Deposits classified into play types are subjected to a successively refined descriptive partitioning of the type that Brown and Saunders describe. To visualize their model applied to oil and gas deposits imagine a hierarchical tree of descriptive characteristics, each of which has a finite number possibilities. Construction of the tree might begin with identification of geological age--Jurassic, Cretaceous, Triassic, etc. Deposits may next be classified according to source rock characteristics, then migration path attributes followed by classification according to the nature of seal. Deposits may be sorted into trap mechanism types and further sorted according to more refined lithologic characteristics of reservoir rocks. This process sorts deposits into descriptive bins of like objects and if carried far enough may yield approximate lognormality of quantities like deposit area, hydrocarbons in place, rock volume, etc. A classificatory scheme of this types bears some relation to the geochemical and geophysical processes by which hydrocarbons are formed and trapped because of the fashion in which descriptive categories are defined, but there is no need to incorporate the physics and chemistry of the processes themselves into a particular model to achieve approximate lognormality--classification does it alone!

The lognormal distribution figures prominently in the 1975 and 1981 DOI assessments of undiscovered oil and gas, but in a different role than discussed thus far. Expert opinions about the aggregate endowment of hydrocarbons in individual basins were solicited. Each geologist was asked to record fractiles of her/his personal probability distribution for the aggregate amount of undiscovered (recoverable) oil/gas in a basin conditional on the presence of hydrocarbons in the basin. Individual variations in personal probabilities were "averaged out" by application of a Delphi method and the resulting group fractile assessment was then approximated by fitting a lognormal distribution. In order to compute the distribution of undiscovered hydrocarbons in larger regions such as the aggregate for all coterminous U.S. onshore basins for example, basinal assessments were assumed to be probabilistically independent and aggregated by monte carlo simulation

[Crovelli (1986, 1987)]. A detailed study of asymptotic properties of **unnormalized** sums of lognormal random variables by Barouch, Kaufman and Glasser (1986) uncovered complicated behavior: such sums behave drastically differently in different asymptotic regimes. Nevertheless, Crovelli's monte carlo studies show that for parameter values within the range of those that appear in individual basinal assessments, a lognormal fit to the convolution of lognormal random variables is reasonable. Only the extreme tails are out of kilter.

The prominent role of the lognormal distribution in USGS assessment methodology disappeared with the 1989 assessment. The focus of attention became individual petroleum plays and in lieu of appraising in-place or recoverable undiscovered petroleum for an entire basin, experts were asked to assess directly properties of the exploration decline curve for individual plays in the basin. To this end the TSP distribution displaced the lognormal distribution and was employed as described in the previous section.

SPATIAL STATISTICS

The Minerals Management Service (MMS), established in 1982 to administer mineral resources in federal offshore territories, is required by legislative mandate to provide assessments of Outer Continental Shelf (OCS) undiscovered oil and gas every two years. To this end it maintains a large staff of experienced scientists; more than four hundred participated in the 1989 OCS assessment. The MMS has access to a vast amount of data: over one million miles of common depth point seismic lines and detailed information from more than 25,000 offshore wells. As a consequence it is possible for the MMS to project remaining undiscovered oil and gas in partially explored and mature offshore petroleum plays by use of modelling techniques that incorporate more geophysical and geological detail than is typical of USGS play analysis.

The NAS/NSF review suggested that the MMS examine observed sequences of drilling outcomes for possible dependencies and that if significant dependencies are uncovered, they should be incorporated into their models. In addition, the review emphasized the importance of correctly accounting for unidentified prospects in a play. The MMS, in contrast to the USGS, has access to maps identifying locations of prospects uncovered by seismic search: "MMS assessors obtained distributions showing the number of undiscovered fields by probabilistic sampling and economic screening of identified prospects...The MMS made no systematic attempt to calculate the number of prospects that the seismic grid missed, which could be considerable even in basins with apparently extensive exploration. Likewise,the MMS did not estimate the number of prospects that were crossed by the seismic grid, but were not identified. Whenever unidentified prospects were included, their size distribution was such that they could never survive the economic screening." [op. cit. pp 91-92].

Because the location, size and other information about prospects identified by seismic search is available to the MMS, they generate a distribution of the number of undiscovered fields in an offshore play by probabilistic sampling and economic screening of such prospects. Their procedure treats most uncertain quantities as probabilistically independent (although a positive correlation between prospects creeps in through the fashion in which a basinal risk factor is treated). Analytical simplicity and the absence of specific guidelines for incorporation of spatial dependencies are sensible reasons for ignoring such dependencies. However, if dependencies are dictated by the geometry of deposition then ignoring them may distort the assessment. According to the NAS/NSF review, "In the Atlantic offshore region, for example, the MMS used five generic play types for the 1989 assessment... Seismic profiles indicate that structures that are well-defined at depth 'die out' and porosity increases upward from the Triassic strata. The implication is that structure, size and porosity are negatively correlated as functions of depth. In addition, structural relief and extent appear to be correlated with depth. These empirical findings suggest that an assessment procedure in which uncertain quantities assessed are assumed to be probabilistically independent may misrepresent an essential feature of the depositional environment. In particular, if structural quality is depth dependent then prospect risk must also be depth dependent."

In addition to the obvious influence of well location on its outcome, probabilities of future well outcomes may be dependent on past drilling history. For example, are future wildcat successes and failures probabilistically dependent on past wildcat successes and failures and if so how? Detailed data describing when and where each offshore exploratory well was drilled together with attributes such as target depth, thickness, porosity and permeability allows application of statistical point process theory to describe spatial patterns of drilling outcomes. The theory is a useful mechanism for studying the character of spatial dependencies among well attributes. Patterns of wells in a play can be interpreted as a marked spatial point process, the "mark" being one or more descriptive attributes of a well at a given location.

What types of statistical analysis have been done to uncover the structure of such probabilistic dependencies? While spatial point process theory is well developed (see Ripley (1987) for a good short review) not much of it has been applied to oil and gas exploration. Gibbs-Markov point processes are potentially attractive models for well locations interpreted as random points in a region of the plane subject to spatial interaction, but their formidable analytical and computational complexity coupled with lack of a fully developed theory of inference for model parameters block easy application. Stoyan's (1987) application to Karst sinkholes in the North Harz region of the former East Germany is illustrative. Beyond the early work of Drew (1967) and Drew and Griffith (1965), spatial regression in three forms has been used to study oil and gas attributes: trend surface analysis, Kriging (a generalization of trend surface analysis) and spatial versions of logit regression. Harbaugh, Doveton and Davis (1977) give an extensive account of how trend surface analysis and universal Kriging may be used to map subsurface features and to identify areas with high probabilities of drilling success. Most applications are to very mature, intensively drilled provinces such as those in Kansas. Trend surface and Kriging models are considered well specified when residuals from the regression surface appear to be independent; i.e. when spatial dependencies are captured by the explanatory variables. An excellent exposition of this version of geostatistics appears in Cressie (1991).

The spatial pattern of exploratory well successes and failures in partially explored regions is of great interest to explorationists as these patterns provide clues about where drilling is most likely to be successful. Hohn (1988) uses an earlier study of well patterns on the northwest shelf of the Delaware Basin in New Mexico by Kumar (1985) to show how indicator Kriging can be applied to spatial patterns of exploratory well successes and failures. (Here the indicator variable is defined to equal one that if a well is a success and zero otherwise.) Trend surface and Kriging residuals are specified to be continuous variables. As a consequence these models must be fit to proportions such as success rates in areal blocks (or functions of them such as log odds) in order to work well.

Logit regression is an alternative specifically designed to handle discrete valued dependent variables. It has been used for target area identification by Canadian hardrock geologists [Agterberg (1974,1984), Chung (1978), Chung and Agterberg (1980)] and more recently by Wrigley and Dunn (1985) to study drilling successes and failures as a function of local geological structure, zone thickness and shale content. They demonstrate the use of graphical diagnostics to improve model specification. Kaufman and Lee (1992) employ logit regression to examine the influence of the spatial pattern of drilling history on probabilities of future exploratory well successes. Their objective is to estimate the probability of success of the n+1st wildcat at a particular location given locations and outcomes (success or failure) of the preceding n wildcats. They assume that well locations are non-random covariates appearing as part of observed drilling history and in addition that both locations and outcomes of the first n wildcats within a distance d of the n+1st well may influence that well's outcome. An application to the Leduc Reef-Windfall and Swan Hills-South Kaybob plays in Western Canada shows how logit regression can be used to uncover spatial dependencies of well outcomes and to predict future exploratory well outcomes as a function of location. Their analysis suggests that Swan Hills well outcomes are spatially dependent. Figure 5 displays probability of success iso-contours for the 415th Leduc well based on outcome of the first 414 wells. Table 1 presents 95% confidence intervals for point estimates of success probabilities for four wells.

[Figure 5 and Table 1 here]

The precision of these spatial probability estimates may possibly be improved by inclusion of geological variables of the type used by Wrigley and Dunn, but this is yet to be done.

In sum, some reasonably simple modeling techniques are available to explore probabilistic spatial dependencies of well outcomes. Applications to a sample of plays analyzed for the 1989 assessment would show where and how USGS and MMS assessment procedures could be modified to account for such dependencies.

The spatial pattern of exploratory wells in a petroleum basin is typically highly variable; some areas are densely drilled and others sparsely drilled. By definition an exploratory well cannot be drilled within the boundaries of a discovered deposit. Hence, the inventory of prospective acreage remaining to be explored must exclude the combined area of all discovered deposits. Even dry holes may "condemn" acreage within a spatial window about it. Singer and Drew (1976) formalize this observation by defining the area of influence of an exploratory well in relation to a deposit target area of given shape, areal extent and orientation. It is the area surrounding a well within which it is not possible to locate such a target without penetration by that well. Computation of the area of influence determined by a collection of wells drilled close to one another is a tricky exercise in geometric probability. Singer and Drew develop a method for calculating the fraction of prospective exploration acreage in a region that is eliminated or exhausted by areas of influence of drilled wells. Root and Schuenemeyer (1980b) show how the Arps-Roberts model can be modified to incorporate areal exhaustion as a measure of exploration effort. This modification has not been widely adopted in practice because it requires so much computation.

Return to the NAS/NSF panel's suggestion that the MMS employ statistically sound methods for estimating the number and size distribution of unidentified prospect structures in a seismically searched area. How might this be done? Seismic search for drillable targets is a particular type of size biased sampling because large structures of a particular type are usually easier to identify than smaller structures and the larger a structure is, the more likely it is to be crossed by a seismic grid of fixed dimension. If a structure is small enough to fit entirely within a seismic grid rectangle it may not be detected at all! Hard rock geologists have studied a closely related problem in which search effort is composed of bore holes aimed at detecting mineral deposits such as massive sulphide or porphyry copper deposits (McCammon (1977), Singer (1972, 1975)). Early efforts at analysis of continuous line search by Pachman (1966) and others assumed fixed target orientation and size.

Seismic search for geological anomalies is carried out by surveying along linear grid lines which partition the search region into rectangles. The simplest two-dimensional version of the seismic search problem supposes that a geological anomaly or structure is detected if at least one grid line crosses the projective surface area of that anomaly. If search of this type identifies the location and two dimensional projective shape (surface shape) of n such anomalies, how should the number of anomalies **missed** by the survey be estimated? What inferences can be made about the empirical distribution of two-dimensional shapes of anomalies missed by the survey? The early stages of exploration strategy depends in part on information provided by "reconnaissance surveys" of this type.

Barouch and Kaufman (1992) point out that"... this is a problem of **selection bias sampling** in which the probability of inclusion of an element of the finite population of anomalies in the observed sample is solely a function of attributes of that element. The observed data is both **size biased and encountered** [Patil and Ord (1976)]" and the sampling scheme represented by this version of seismic search is called **minus sampling** by spatial statisticians and stereologists. It is **not** a typical stereological problem in which three dimensional properties of an anomaly are to be inferred from two dimensional samples of it. The objective here is to estimate two dimensional characteristics of a population of anomalies from a two-dimensional sample of these characteristics. Detailed seismic surveys may be designed to measure three dimensional attributes of an anomaly from two dimensional slices of it, but only after a reconnaissance survey has identified its location and projective area of closure.

The first step is to compute the probability of detection of an anomaly. Suppose that the center of an elliptically shaped target is uniformly randomly located on the plane and has uniformly random orientation;i.e. this target is equipped with kinematic measure on the plane. The probability that it will be detected by rectangular grid search is a function of grid rectangle dimensions and major and minor axes of the ellipse and is expressible in terms of complete and incomplete elliptic integrals of the second kind. More generally, the probability of detection of an anomaly with irregular boundary curve can be computed as a function of rotations of a rectangular box whose sides are tangent to the target as in Figure 6. [Barouch and Kaufman op. cit.]

[Figure 6 here]

As the probability of detection of a target is the probability that it will be included in a sample, unbiased estimation of properties of the entire finite population of targets can be estimated by employing sample survey estimation techniques. Fan (1992) studied results of seismic search of an offshore U.S. Basin which identified 94 anomalies. As only the major and minor axis of each anomaly was available, he approximated boundary curves with ellipses and used Horvitz-Thompson type estimation to estimate both the number of anomalies missed by seismic search and the empirical distribution of areas of missed anomalies, suggesting that the search failed to discover approximately 16 anomalies in the course of identifying 94. The bulk of the anomalies missed are small relative to what was identified, as Figure 7 illustrates.

[Figure 7 here]

While an elliptical approximation to target shape may allow better approximation to actual shape, estimates obtained by circular approximations of these particular anomalies do not differ appreciably from elliptical approximations, reinforcing McCammon's observation that simple circular approximations may be fairly robust (McCammon, op. cit.).

Several features of reconnaissance seismic search are not incorporated in the problem as formulated above. For example, the probability of detection of a target may be made to depend on the number of grid line crossings. Seismic search does not always discriminate perfectly: what appears to be subsurface structure may be a "ghost" leading to false target identification. A "large" anomaly may be broken up into several smaller ones after a detailed seismic survey of it. These features of seismic search require more complex models and estimation procedures. In practice there is the usual tradeoff between simple robust models and methods and more complex realistic ones.

CONCLUSIONS

A final commentary: only a fraction of the statistical methods reviewed here has made its way into practice in a substantial way. Although most of the current generation of oil and gas geologists and exploration managers have been exposed to more statistical training than their older counterparts, there is a cultural gap that still remains to be bridged. Statisticians must in turn continue to build bridges to the geological community by demonstrating how new statistical methods can be fruitfully applied to problems that the geologists regard as important.

REFERENCES

Adelman, M. A., J. C. Houghton, G. M. Kaufman, M. B. Zimmerman (1983). Energy Resources in an Uncertain Future: Coal, Gas, Oil, and Uranium Supply Forecasting. Cambridge, MA: Ballinger Publishing Company.

Agterberg, F. P. (1974). *Geomathematics*: 596. Elsevier Scientific Publishing Co., Amsterdam.

Agterberg, F. P. (1984). Trend Surface Analysis, in G. L. Gaile and C. J. Willmott (eds), Spatial Statistics and Models. Dordrecht, Holland: Reidel Publishing Co., 147-171.

Aitchison, J. and J. A. C. Brown (1957). The Lognormal Distribution. New York: Cambridge U. P.

Andreatta, G. and G. M. Kaufman (1986). "Estimation of Finite Population Properties When Sampling is Without Replacement and Proportional to Magnitude." *American Statistical Association* 81(395): 657-666.

Andreatta, G., G. M. Kaufman, R. G. McCrossan and R. M. Procter (1988). "The Shape of Lloydminster Oil and Gas Deposit Attribute Data." C. F. Chung et al. (eds), *Quantitative Analysis of Mineral and Energy Resources*: 411-431.

Arps, J. K. and T. G. Roberts (1958). "Economics of Drilling for Cretaceous Oil and Gas on the East Flank of the Denver-Julesberg Basin." *American Association of Petroleum Geologists Bulletin* 42(11): 2549-2566.

Attanasi, E. D. and L. J. Drew (1985). "Lognormal Field Size Distributions as a Consequence of Economic Truncation." *Mathematical Geology* 17(4): 335-351.

Barouch, E., S. Chow, G. M. Kaufman and T. H. Wright (1985). "Properties of Successive Sample Moment Estimators." *Studies in Applied Mathematics* 73: 239-260.

Barouch, E. and G. M. Kaufman (1976). "Oil and Gas Discovery Modelled as Sampling Proportional to Random Size." MIT Working Paper 888-76, 65 pp.

Barouch, E. and G. M. Kaufman (1977). "Estimation of Undiscovered Oil and Gas." in *Proceedings of Symposia in Applied Mathematics* (21): 77-91. Providence, RI: American Mathematical Society.

Barouch, E. and G. M. Kaufman (1992). "Estimation of Properties of Finite Populations of Geologic Anomalies from Grid Line Search Samples." In process.

Barouch, E., G. M. Kaufman and M. L. Glasser (1986). "On Sums of Lognormal Random Variables." *Studies in Applied Mathematics* 75: 37-55.

Barouch, E., G. M. Kaufman and J. Nelligan (1983). "Estimation of Parameters of Oil and Gas Discovery Process Models Using the Expectation-Maximization Algorithm." *Energy Modelling and Simulation* 109-117 pp. A. S. Kydes et al. (eds.) IMACS: North-Holland Publishing Co.

Bickel, P. J., V. N. Nair and P. C. C. Wang (June 1992). "Nonparametric Inference Under Biased Sampling from a Finite Population." *The Annals of Statistics* 20(2): 853-878.

Bloomfield, P., K. S. Deffeyes, G. S. Watson, Y. Benjamini, and R. A. Stine (Feb. 1979). "Volume and Area of Oilfields and Their Impact on Order of Discovery." Resource Estimation and Validation Project Departments of Statistics and Geology Princeton University.

Bloomfield, P., M. Hudson, K. Kim and G. S. Watson (Feb. 1979). "The Use of the Lognormal Distribution in Estimating Crude Oil Resources." Resource Estimation and Validation Project Departments of Statistics and Geology Princeton University. Paper

Brashear, J. P., A. Becker, F. Biglarbigi and R. M. Ray (1989). "Incentives, Technology, and EOR: Potential for Increasing Oil Recovery at Lower Oil Prices." *Journal of Petroleum Technology* 41(2): 164-170.

Brown, G. and J. W. Saunders (1981). "Lognormal Genesis." *J. Appl. Prob.* 18: 542-547.

Cassel, C. M., C. E. Sarndal and J. H. Wretman (1977). Foundations of Inference in Survey Sampling. New York: John Wiley & Sons.

Chung, C. F. (1978). "Computer Program for the Logistic Model to Estimate the Probability of Occurrence of Discrete Events." *Geological Survey of Canada Paper* 78-11, Ottawa, 23 p.

Chung, C. F. and F. P. Agterberg (1980). "Regression Models for Estimating Mineral Resources from Geological Map Data." *Mathematical Geology* 12: 473-488.

Cressie, N. (1991). Statistics for Spatial Data. Somerset, NJ: John Wiley & Sons.

Crovelli, R. A. (1986). "A Comparison of Analytical and Simulation Methods for Petroleum Play analysis and Aggregation." *U. S. Geological Survey Open-File Report* 86-79: 21 p.

Crovelli, R. A. (1987). Probability Theory Versus Simulation of Petroleum Potential in Play Analysis, in Albin, S.L. and Harris, C. M. (eds.), Statistical and Computational Issues in Probability Modeling, pt. 1. *Annals of Operations Research* 8: 363-381.

Davis, J. C. and T. Chang (August 1989). "Estimating Potential for Small Fields in Mature Petroleum Province." *The American Association of Petroleum Geologists Bulletin* 73(8): 967-976.

Drew, L. J. (1967). "Grid-drilling exploration and its application to the search for petroleum." *Economic Geology* 62(5): 698-710.

Drew, L. J. and J. C. Griffiths (1965). "Size, Shape and Arrangement of Some Oilfields in the USA." *1965 Symposium, Computer Applications in the Mineral Industries*. Pennsylvania State University, pp. FF-1-FF-31.

Drew, L. J., J. H. Schuenemeyer and W. J. Bawiec (1979). "Estimating future rates of petroleum discoveries in the Permian basin, in *SPE-AIME Symposium on Economics and Evaluation*, 8th, Dallas, Tex., 1979, Proceedings: Dallas, American Institute of Mining, Metallurgical and Petroleum Engineers, Inc., Society of Petroleum Engineers, 101-106.

Drew, L. J., J. H. Schuenemeyer and W. J. Bawiec (1982). "Estimation of the Future Rates of Oil and Gas Discoveries in the Western Gulf of Mexico." *Geological Survey Professional Paper* 1252. Washington: U. S. Government Printing Office.

Drew, L. J., J. H. Schuenemeyer and D. H. Root (1980a). "Petroleum -Resource Appraisal and Discovery Rate Forecasting in Partially Explored Regions, Part A: An Application to the Denver Basin." *U. S. Geological Survey Professional Paper* 1138-A, B, C: A1-A11.

Fan, M. Y. (1992). "The Probability Assessment of Elliptically Approximated Seismic Anomalies." Thesis submitted to the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Forman, D. J. and A. L. Hinde (Jan. 1985). "Improved Statistical Method for Assessment of Undiscovered Petroleum Resources." *The American Association of Petroleum Geologists Bulletin* 69(1): 106-118.

Garrett, R. G. (1985). "Geochemical Data-Retrospect and Prospect." *The Role of Data in Scientific Progress*: 25-32, P. S. Glaeser (ed.). Elsevier Science Publishers B. V. (North-Holland).

Gordon, L. (1981). "Estimating Reserves and Resources." *Bulletin of the International Statistical Institute* 43: 1303-1320.

Gordon, L. (Feb. 1992). "Estimation for Large Successive Samples with Unknown Inclusion Probabilities." Forthcoming in *Advances In Applied Mathematics*.

Hájek, J. (1981). Sampling From a Finite Population. New York: Marcel Dekker, Inc.

Harbaugh, J. W., J. H. Doveton and J. C. Davis (1977). Probability Methods in Oil Exploration. New York: John Wiley & Sons.

Hoaglin, D. C. and S. Peters (1983). "g- and h-Distributions." Encyclopedia of Statistical Sciences. N. Johnson and S. Kotz (eds). New York: John Wiley & Sons.

Hoaglin, D. C. and J. W. Tukey (1992). "Empirical Bounds for Quantile-Based Estimates of g in the g- and -h Distributions." In preparation.

Hohn, M. E. (1988). Geostatistics and Petroleum Geology. New York: Van Nostrand Reinhold.

Houghton, John C. (1988). "Use of the Truncated Shifted Pareto Distribution in Assessing Size Distribution of Oil and Gas Fields." *Mathematical Geology* 20(8): 907-937.

Kaufman, G. M. (1963). Statistical Decision and Related Techniques in Oil and Gas Exploration. Englewood Cliffs, NJ: Prentice-Hall.

Kaufman, G. M., Y. Balcer and D. Kruyt (1975). "A Probabilistic Model of Oil and Gas Discovery: in Haun, J. D., ed. *Methods of Estimating the Volume of Undiscovered Oil and Gas Resources, American Association of Petroleum Geologists, Studies in Geology* No 1: 113-142.

Kaufman, G. M. and P. J. Lee (1992). "Are Wildcat Well Outcomes Dependent or Independent." *Nonrenewable Resources* 1(3). Published by Oxford University Press for the International Association for Mathematical Geology.

Kumar, N. (1985). "Estimating Exploration Potential in Mature Producing Areas, Northwest Shelf of Delaware Basin New Mexico." *Bull. Amer. Petrol. Geology* 69: 903-1916.

Lee, P. J. and P. C. C. Wang (1983a). "Probabilistic Formulation of a Method for the Evaluation of Petroleum Resources." *Mathematical Geology* 15: 163-181.

Lee, P. J. and P. C. C. Wang (1983b). "Conditional Analysis for Petroleum Resource Evaluations." *Mathematical Geology* 15: 353-363.

Mast, R. F., G. L. Dolton, R. A. Crovelli, D. H. Root, and E. d. Attanasi (1989). "Estimates of Undiscovered Conventional Oil and Gas Resources in the United States--A Part of the Nation's Energy Endowment. *U. S. Geological Survey*, Books and Open-File Reports Section, Federal Center, Box 25425, Denver. CO 80225.

McCammon, R. B. (1977). "Target Intersection Probabilities for Parallel-Line and Continuous-Grid Types of Search." *Mathematical Geology* 9(4): 369-382.

McCrossan, R. G., R. M. Procter and W. J. Ward (1981). "Estimate of Oil Resources, Lloydminster Area, Alberta," Proceedings of the First International Conference on the Future of Heavy Crude and Tar Sands.

Meisner, J. and F. Demirmen (1981). "the Creaming Method: A Bayesian Procedure to Forecast Future Oil and Gas Discoveries in Mature Exploration Provinces." *Journal of the Royal Statistics Society Series A* 144: 1-31.

Murthy, M. N. (1957). "Ordered and unordered estimators n sampling without replacement." *Sankhya* 18: 378-390.

Pachman, J. M. (1966). "Optimization of Seismic Reconnaissance Surveys in Petroleum Exploration." *Management Science* 12(8): B-312-322.

Patil, G. P. and J. K. Ord (1976). "On Size-biased Sampling and Related From-Invariant Weighted Distributions." *Sankhya* 38(Ser B, Pt. 1): 48-61.

Podruski, J.A., J. E. Barclay, A. P. Hamblin, P. J. Lee, K. G. Osadetz, R. M. Procter, and G. C. Taylor (1988). "Conventional Oil Resources of Western Canada." *Geological Survey of Canada Paper 87-26*. Geological Survey of Canada, Ottawa, Canada.

Rabinowitz, D. (1991). "Using Exploration History to Estimate Undiscovered Resources." *Mathematical Geology* 23(2): 257-274.

Rabinowitz, D. (1992). "Adaptive Estimation of Undiscovered Deposits Following Partial Exploration." [In review]

Ripley, B. D., (1987). "Point Processes for the Earth Sciences." Quantitative Analysis of Mineral Energy Resources. C. F. Chung, et al., eds. Reidel Publishing Company.

Root, D. H. and J. H. Schuenemeyer (1980b). "Petroleum-Resource Appraisal and Discovery Rate Forecasting in Partially Explored Regions, Part B: Mathematical Foundations." *U. S. Geological Survey Professional Paper* 1138-A, B, C, pp. B1 - B9.

Schuenemeyer, J. H. and L. J. Drew (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* 15(1): 145-161.

Schuenemeyer, J. H., L. J. Drew, D. H. Root and E. D. Attanasi (Nov. 1990). "Estimating Potential for Small Fields in Mature Petroleum Province: Discussion." *The American Association of Petroleum Geologists Bulletin* 74(11): 1761-1763.

Singer, D. A. (1972). Elipgrid, a FORTRAN IV program for calculating the probability of success in locating elliptical targets with square, rectangular and hexagonal grids: *Geocom Programs*, No. 4, London, 16 pp.

Singer, D. A. (1975). "Relative Efficiencies of Square and Triangular Grids in the Search for Elliptically Shaped Resource Targets." *U. S. Geological Survey, Journal of Research* 3(2): 163-167.

Singer, D. A. and L. J. Drew (1976). "The Area of Influence of an Exploratory Hole." *Economic Geology* 71(3): 642-647.

Stoyan, D. (1987). "Applications of Stochastic Geometry in Geology." Proceedings of the 1st World Congress of the Bernoulli Society, Prohorov, Yo and V. V. Sazonou eds. *Mathematical Statistics Theory and Applications* 2: VNU Science Press Utrecht, The Netherlands.

Wang, P. C. C. and V. N. Nair (Jan. 1986). "Estimation for Super-population Parameters of a Successively Sampled Finite Population."

Wang, P. C. C. and V. N. Nair (1988). "Statistical Analysis of Oil and Gas Discovery Data." C. F. Chung et al. (eds). *Quantitative Analysis of Mineral and Energy Resources* 199-214.

White, D. A. (1980). "Assessing Oil and Gas Plays in Facies-Cycle Wedges." *American Association of Petroleum Geologists* Bulletin 64: 1158-1178.

Wrigley, N. and R.Dunn (1985). "Graphical Diagnostics for Logistic Oil Exploration." *Mathematical Geology* 18(4): 355-374.

Figure 1. Graph of $E(\tilde{Y}_n)$ for Lognormal Superpopulation with Mean = 1808.04, Variance 62.3910×10^6 ($\mu = 6.0$, $\sigma^2 = 3.0$)

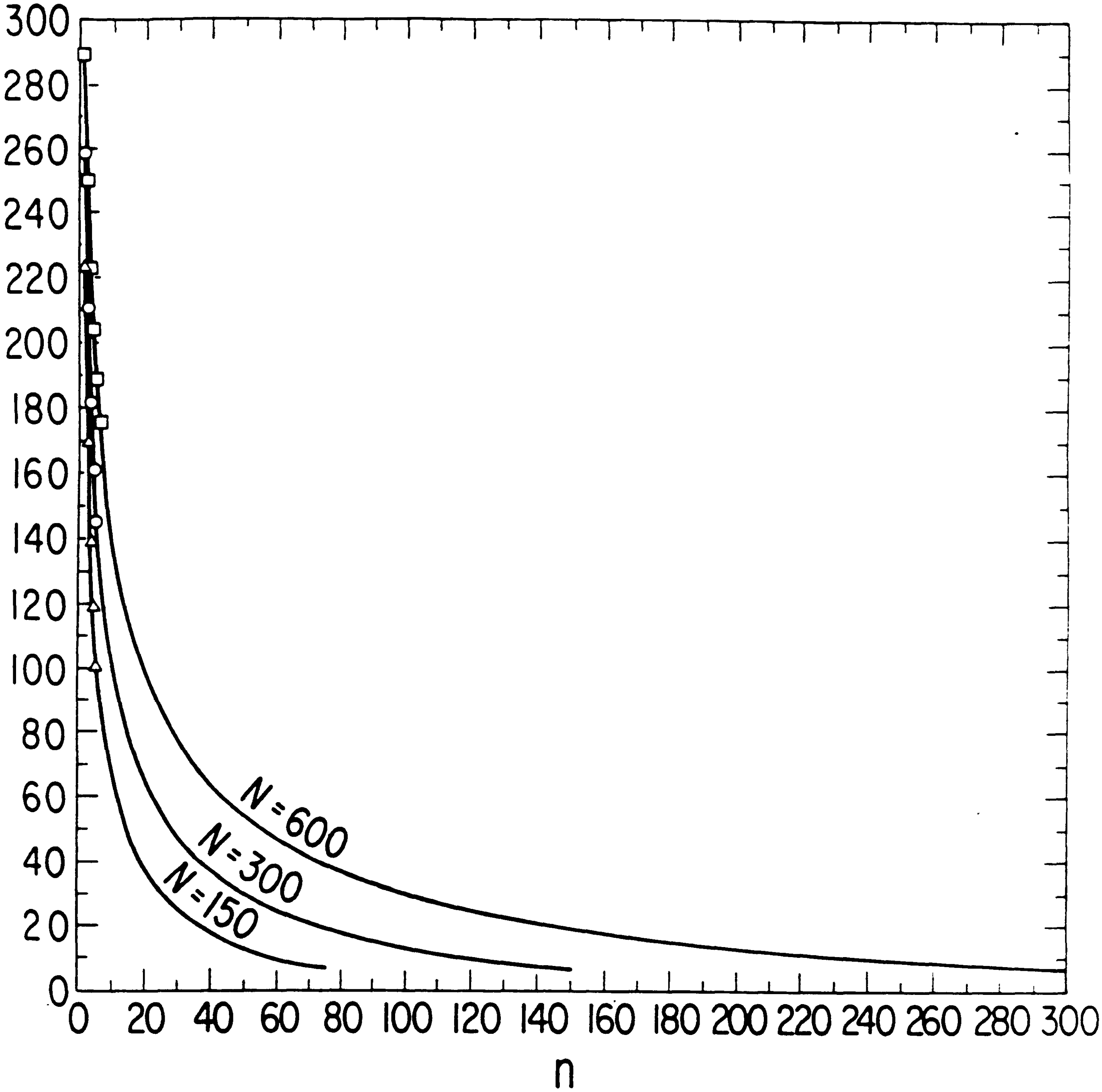
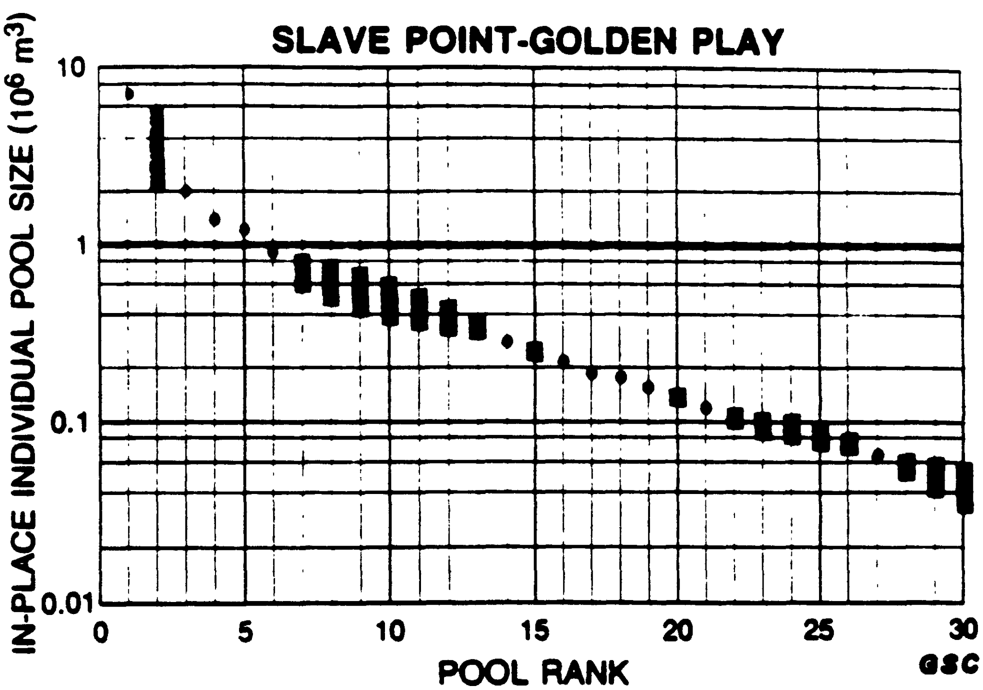


Figure 2. Slave Point-Golden Play Log Sizes vs. Rank Plot*

SLAVE POINT — GOLDEN PLAY

Rank	Pool Name	In-Place Pool Volume (10^6 m^3)	Discovery Year
1	Golden A	6.88	1970
3	Slave G	1.96	1980
4	Seal	1.40	1974
5	Evi B	1.21	1979
6	Evi A	0.88	1979
14	Evi C	0.28	1981
16	Evi D	0.22	1982
17	Evi L	0.19	1982
18	Evi J	0.18	1982
19	Evi I	0.15	1982
21	Evi F	0.12	1982
27	Evi E	0.07	1982

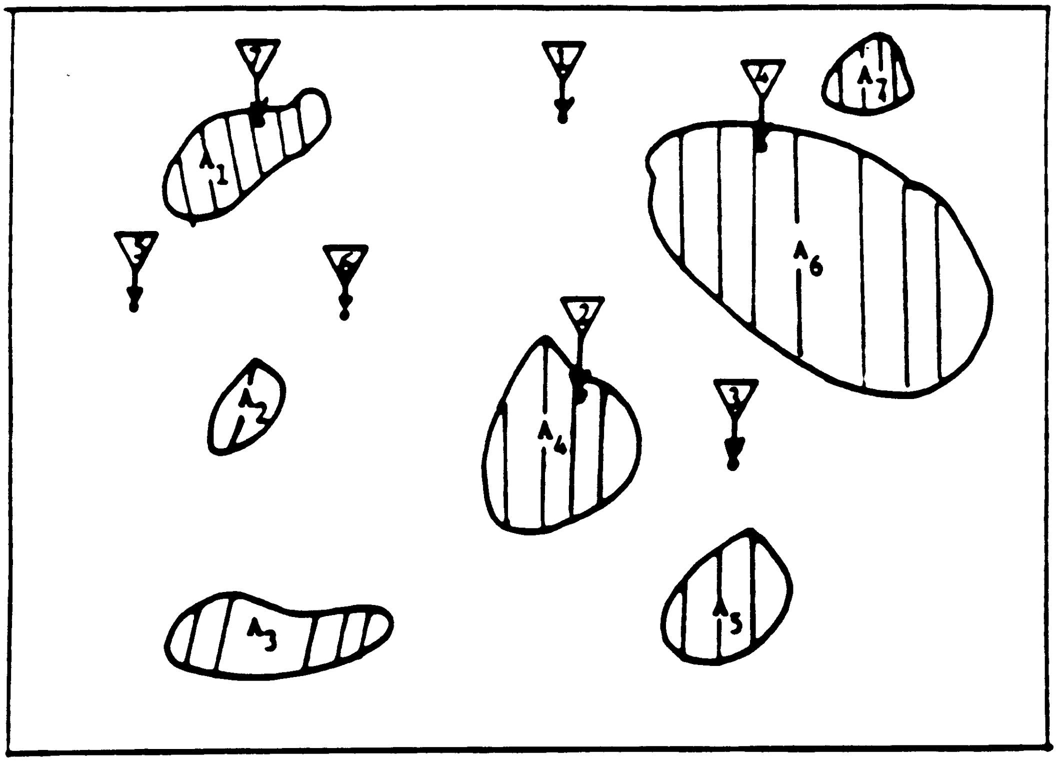
- •Total Discoveries : 12
- •Discoveries in the Top 30 Pools : 12
- •Total Pool Population : 40



$\mu = -2.0$ $\sigma^2 = 3.0$ $\beta = 1$ N=40

*"Conventional Oil Resources of Western Canada", Geological Survey of Canada, Paper 87-26,1988.

Figure 3. Successive Sampling as a Dart Board Game



- location of well number i

$(0, A_4, 0, A_6, 0, 0, A_1)$ - history of first seven wells

Figure 4. Lloydminster Play

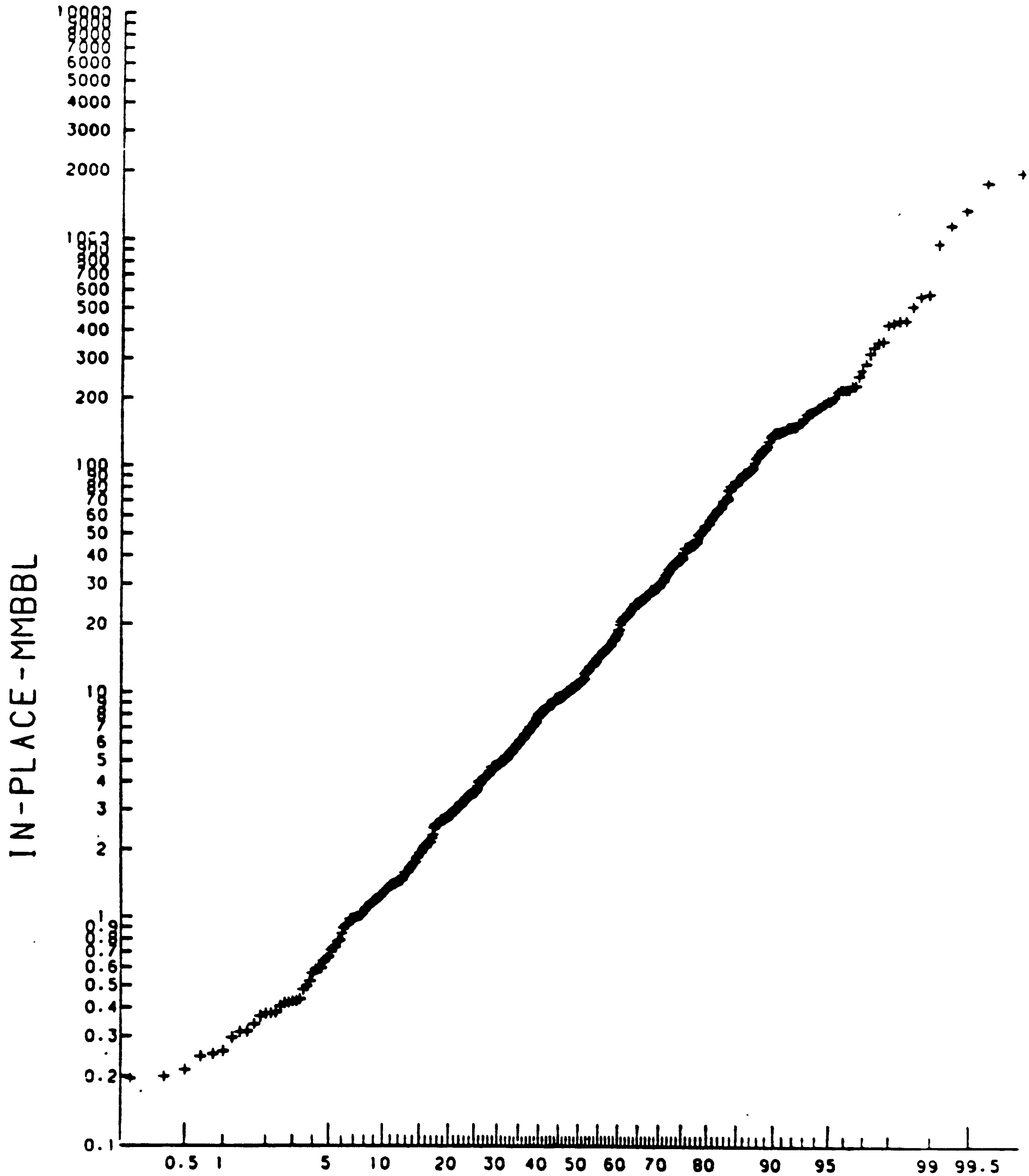
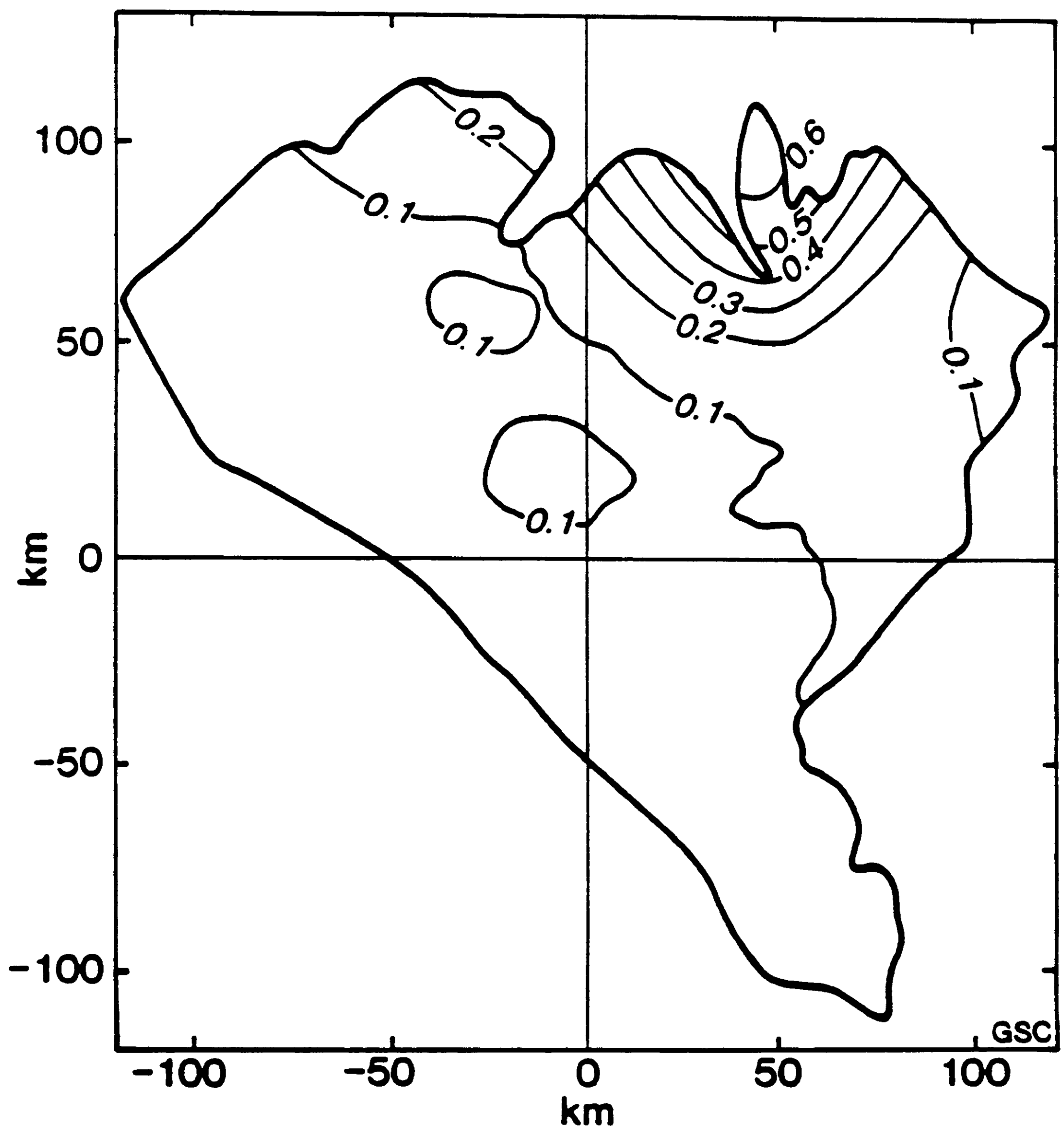


Figure 5. Swan Hills Shelf Margin - Kaybob South Play
d = 5 km, RECIP, RECIPS



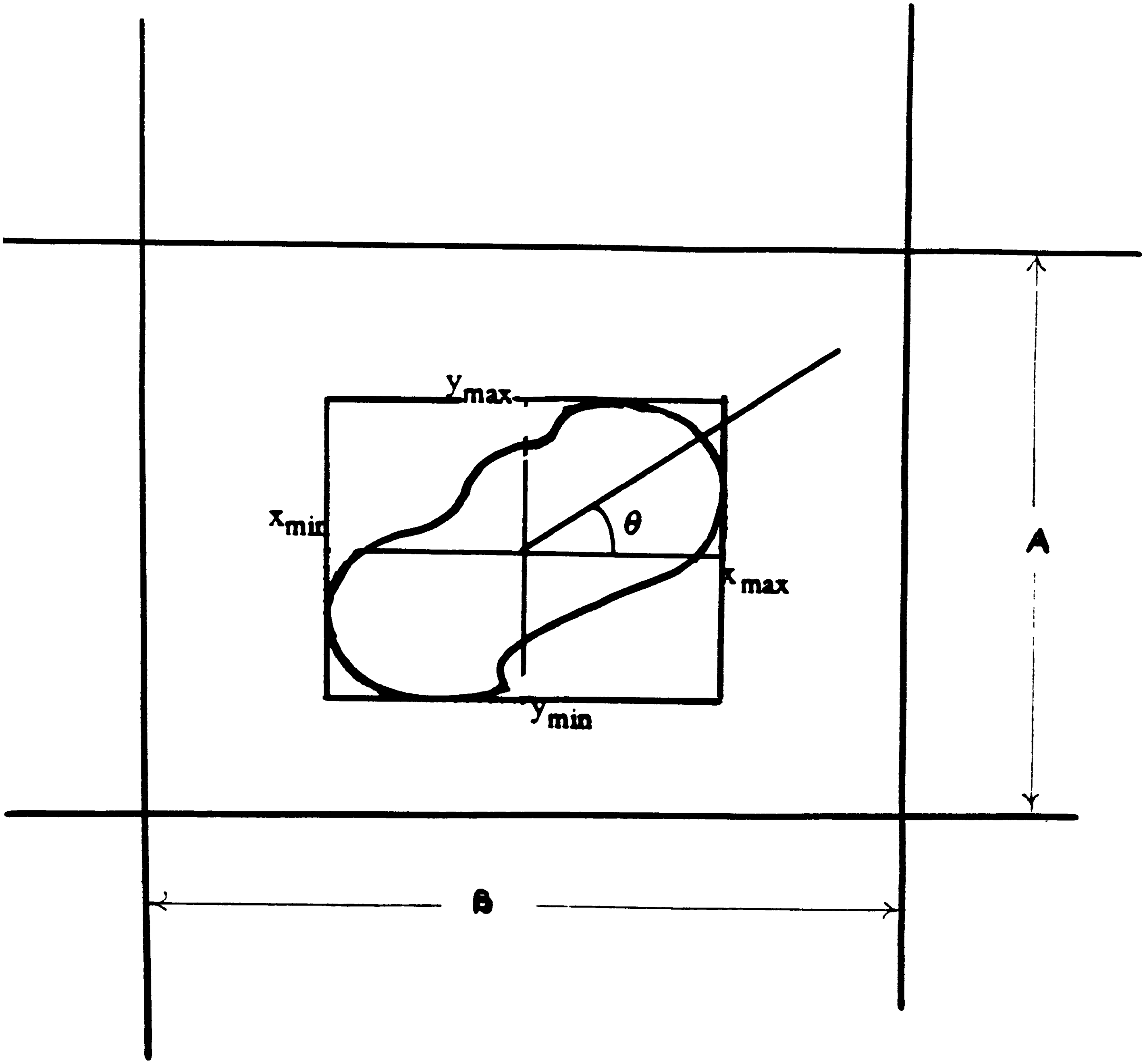


Figure 6. Caliper Lengths of an Anomaly:
Prob Detection at angle θ = $\frac{[y_{max} - y_{min}][x_{max} - x_{min}]}{AB}$

Figure 7. Missed and Observed Target Frequencies vs. Log Area

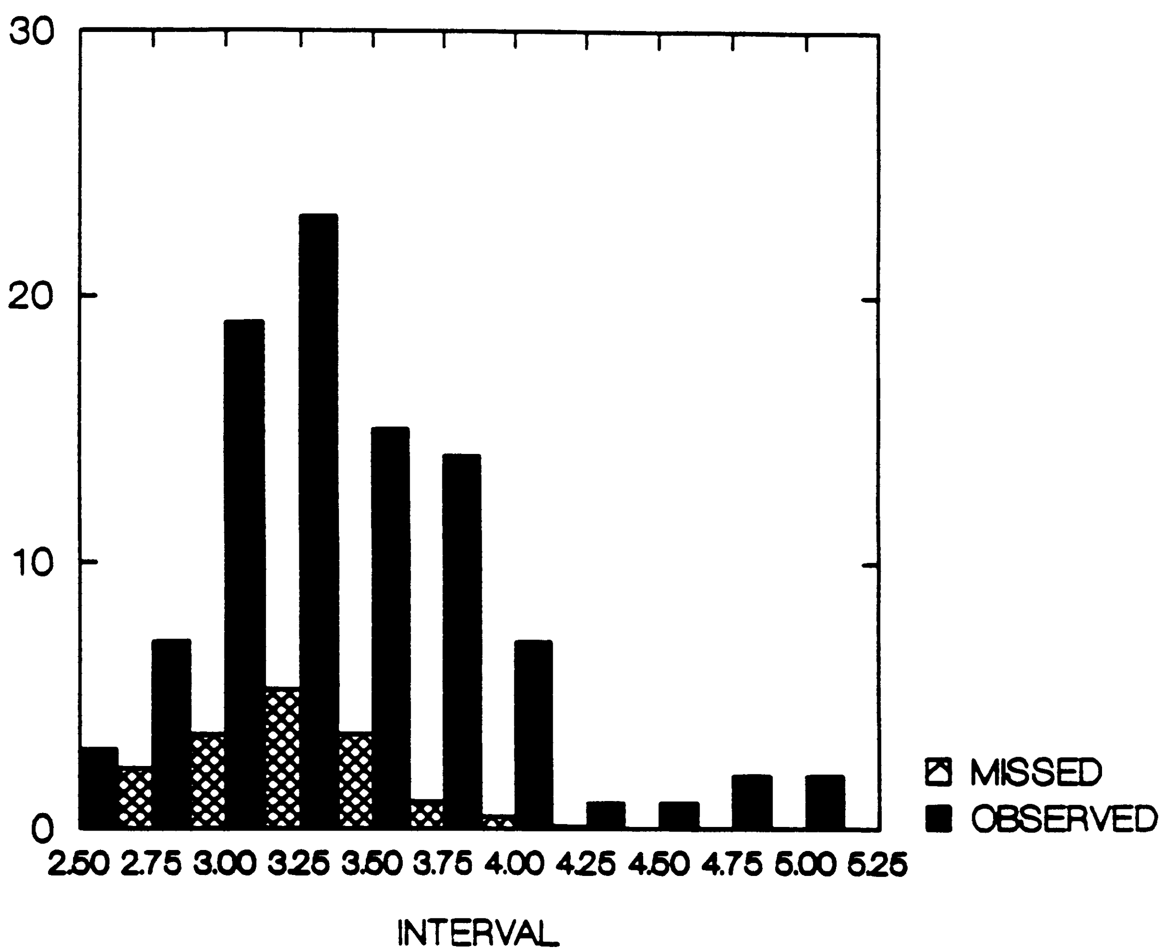


Table 1. 95% Upper and Lower Confidence Bounds for $\hat{p}(\underline{x}_i \mid H_{i-1})$

Well #	Lower	$\hat{p}(\underline{x}_i \mid H_{i-1})$	Upper
411	.106	.323	.375
412	.077	.103	.168
413	.069	.096	.179
414	.090	.118	.156