UNITED STATES DEPARTMENT OF THE INTERIOR
GEOLOGICAL SURVEY

Use of Field Size Distributions in Analog Basins for Hydrocarbon Resource Assessment

by

John C. Houghton[1]

Open-File Report 86-180

1 Reston, Va

CONTENTS

Section

1 INTRODUCTION
- Definitions
- The Benefits of Including Field Size Distributions in Hydrocarbon Estimation
- Methods Presented in This Report
- Other Methods for Developing Field Size Distributions

2 DATA ON U.S. OIL FIELDS

3 GEOLOGIC DATA

4 THE LOGNORMAL DISTRIBUTION APPLIED TO FIELD SIZES
- Fitting the Lognormal Distribution
- Fitting a Truncated Lognormal Distribution

5 RESULTS
- Choosing the Best Fitting Procedure
  - Two criteria for assessing the fit
  - Method of moments
  - The maximum likelihood method
  - The truncated lognormal
  - The choice: the maximum likelihood
- The Shape of the Field Size Distribution
- Influence of Geology

6 A METHOD TO ESTIMATE THE SHAPE OF FIELD SIZE DISTRIBUTIONS

7 CONCLUSIONS

8 REFERENCES

9 FIGURES

10 TABLES

ILLUSTRATIONS

Figure

1 Fields of the East Texas Basin Fit Using Method of Moments
2 Fields of the East Texas Basin Fit Using Maximum Likelihood
3 Lognormal Distributions with Different Shape Parameters
4 Shape Parameter as a Function of Basin Type

TABLES

Table

1 Basin Classification by Tectonic Type
2 Basic Statistics of the 50 Basins
3 Lognormal Fit to the 50 Basins
4 A Comparison of Quantiles for Selected Basins

INTRODUCTION

This report presents a new method for estimating and describing basin-wide, undiscovered oil and gas resources that attempts to make the estimates more useful for "downstream" analysis, such as engineering analysis, economic analysis, and policy analysis. Although the report contains much information on computerized records of fields and on the results of applying statistics to these files, the analysis and the report are intended to emphasize the concepts of applying field size distribution rather than the results themselves. More work is needed to improve the data base and some of the statistical techniques. A description of the overall statistical methods and further benefits follows a section on definitions.

Definitions

Oil and gas resources occur in many different size packages. Although definitions vary, the smallest individual accumulation is commonly called a pool or a reservoir and is usually defined as an accumulation in which the hydrocarbon phases are physically continuous. In other words, withdrawing some of the hydrocarbon in one point in the reservoir will affect the fluid pressure in the rest of the reservoir. There are about 100,000 hydrocarbon reservoirs in the United States.

Reservoirs are usually aggregated to form fields by including in a field all those reservoirs that overlap one other in a vertical dimension. Fields will often contain oil in reservoirs in which oil is "trapped" by the same "structure", for example, several sandstone layers in an anticline. This definition is a logical one because reservoirs are usually much wider and longer than they are thick, and sometimes they "stack" on top of each other like irregularly-shaped cards in a deck. There are about 30,000 fields in the United States.

Fields are aggregated into "plays", which contain geologically similar accumulations. The hydrocarbons in a single play are usually considered to have the same geologic history, that is, come from the same source, migrate through the same rocks, have been affected by the same thermal history, and are trapped by the same mechanisms. There are several hundred plays in the United States, depending on definitions.

An aggregation unit broader than a play is the geologic basin or province. There are two geographic systems commonly used in the U.S., one presented by the AAPG (Meyer, 1970) and one used primarily by the USGS (Dolton, 1981). These basins or provinces are described in terms of broad geologic features. There are about 50 basins and provinces in the United States that contain more than ten fields.

The Benefits of Including Field Size Distributions in Hydrocarbon Estimation

Undiscovered oil and gas resource estimates are often presented for an entire basin or province. Circular 860 (Dolton et al., 1981) is an example of a set of hydrocarbon resource estimates in which the total amount of oil and gas for a province or basin is assessed, but for which no information is provided on the predicted sizes of the fields or reservoirs. When little field-specific data for a particular region is available, and when assessment is based on expert judgment of total undiscovered resources, field-level estimates are difficult to generate after the resource estimates are made. This lack of information on field sizes makes it difficult for those analysts "downstream", such as economists, engineers, and policy analysts to use this kind of resource estimate in their own analysis. Field size distributions benefit the estimation process because they provide a foundation for the application of, for instance, discovery and production models.

Another benefit is that the resource estimates themselves can be improved if field size distributions are considered by the persons performing the estimation during the assessment process. If these assessors have information on the size of the largest field, they may decide, for instance, that there is not enough space in the basin to fit such a large field and that preliminary estimates are too high. In addition, these techniques can be used as a check on other techniques that generate field sizes. For instance, are field sizes generated from play analysis techniques similar to records from well-explored basins?

Methods Presented in this Report

What is needed to improve aggregate estimates is a way to divide statistically the total amount of estimated oil into probable field sizes. The statistical divisions are modeled by "field size distributions". The fields in a basin are assumed to be statistically distributed, in statistical jargon, as observations in a sample. The most common assumption currently is that the fields are distributed as a two-parameter lognormal distribution (Procter et al., 1982).

This report presents a method for dividing the estimated oil into probable field sizes. Field size distributions are developed by fitting distributions to the record of field discoveries for a set of U.S. basins. Most of these basins are relatively mature, that is, they have been explored enough that the discovery record contains most of the fields that are ever expected to be found.

The essential information that is inferred from the mature basins, which can be used in assessing analog basins, is the "shape" of the field size distribution. A statistical definition of shape will be given in a later section and is defined as the standard deviation of the log-transformed sample. This parameter is a one-to-one function of both the coefficient of variation (c.v.) and the coefficient of skew of the untransformed sample. The coefficient of skew is probably the most intuitive. A high skew indicates that relatively more of a basin's oil is found in the larger fields. A low skew means that the oil is more equally disbursed by field size. In other words, the shape refers to how much larger the largest fields are than the rest of the fields in the basin.

The shape of the fitted distribution of the analog basins can be selected for the basin being assessed on the basis of a rough geologic analogy based on tectonic classification. A set of techniques is presented that develop analog field size distributions from a file of 30,000 fields in 50 basins in the U.S. and develops rules for choosing an analog basin and the associated field size distribution.

Only some of the techniques necessary to estimate undiscovered resources are addressed. The assessment of overall risk (the probability of not finding any commercial oil at all) or of determining the likely number of fields is not addressed directly, neither is the problem of assessing the total amount of oil in a basin or play. Instead, the strategy is to gather information on oil fields in the U.S., group the fields by basin, fit the fields in each basin with a two-parameter lognormal distribution using a variety of fitting procedures, group the basins by regional geology (using their tectonic setting), and then infer the typical shape (rather than the magnitude) of the field size distribution as a function of the geologic type of basin.

Other Methods for Developing Field Size Distributions

There are three other major classes of methods for developing field size distributions for areas with hydrocarbon accumulations (some are presented in Baker et al., 1984). The first, usually applied to plays, is based on prospects for which there is much indirect data, such as geophysics, but no direct drilling information. Area, net feet of pay, and other information are interpreted from geophysical data and stratigraphic analysis and are used to make an estimate of recoverable hydrocarbons for each prospect in a region being assessed. These estimates are collected for the region and together constitute an empirical field size distribution of prospects.

The second class is a stochastic version of the first and is usually applied to plays. Probability distributions are assigned to various volumetric factors, such as porosity, percent fill up, and area, and computer techniques are used to combine the factors and simulate field size distributions. Examples include NPRA (Office of Minerals Policy and Research Analysis, 1979) and PRESTO (Minerals Management Service, 1985), although the field size distributions are embedded implicitly in the PRESTO model.

The third class, applied to both plays and basins, uses discovery process models, such as Arps and Roberts (1958) and newer modifications, such as those presented by Meisner and Demirmen (1981) and Forman and Hinde (1985). These models are applied to relatively mature basins with extensive drilling records. The Arps-Roberts models present estimates of field sizes by broad size category, and the other two present future discoveries as a function of the amount of drilling but do not give explicit estimates of the ultimate number and size of fields expected to be found (Houghton et al., 1985).

DATA ON U.S. OIL FIELDS

The fundamental information on oil fields in the U.S. was taken from the PDS (Petroleum Data System of North America at the University of Oklahoma, 1981). The raw data was purchased in 1983 (most production data is current through 1981) in the form of a fixed-record-length file, containing information on each of the approximately 100,000 domestic hydrocarbon reservoirs in the PDS file they call TOTL. This file represents by far the most complete data set, measured by number of records, of any commercially-available, non-proprietary source.

The data file was adjusted by first deleting the approximately 8000 reservoirs designated "combined pools". Including these reservoirs would result in double counting, because as reservoirs expand and are combined, the cumulative production history from the original records are retained.

Information on oil accumulations at the reservoir level is highly varied across the U.S., in large part because of the varied definitions of a reservoir and because data collection procedures change so much from State to State. Although there is some ambiguity about the definition of a field, field information is more commonly used. Fields are used by the EIA (Energy Information Administration of the Department of Energy), for instance, and they have published a catalog with field codes for every field in the U.S. (EIA, 1982). The reservoirs in the PDS file were combined to fields on the basis of matching EIA field code numbers, of which there are about 30,000. The set of 30,000 field sizes was grouped by AAPG Basin (Meyer, 1970).

The PDS data set has information on annual production and cumulative production but does not have information on estimates of reserves. Reserves can be approximated from production history by multiplying the last year's annual production times a reserve to production ratio. Richard Meyer (Meyer, 1983; Woods, 1985; or for an alternative method see Root and Attanasi, 1985) finds that an R/P ratio of 10.0 is a good approximation. The results in this report are based on that ratio.

One shortcoming in the PDS data set is that some of the entries are missing or incorrect. Although records exist for about 30,000 fields, over 7000 fields lacked either the AAPG Basin number or the EIA field code. Although this represents a large number of errors, it represents only a relatively small amount of oil; only 50 of the over 7000 problem entries have cumulative oil production greater than 1 MM bbls.

GEOLOGIC DATA

Many geologic variables affect the field size distribution of a basin. Structural style, trap types, thermal history, reservoir rock type, source rock type, and several other attributes are probably important. Because we were unable to compile a collection of such geologic descriptors for each of the AAPG basins, a broad-scale tectonic classification was used as a surrogate.

Klemme (1983, see also Coustau (1980)) has classified basins worldwide according to their broad tectonic features and has connected some aspects of their field size distributions to their tectonic structure and other characteristics. This classification has been refined by Klemme, so that the categories are not the same throughout his papers. In this report we use the categories presented by Nehring (1981, after Klemme) as shown in Table [1]. Table [2] assigns each AAPG basin to one of the eight categories presented in Table [1].

Klemme uses BOE (barrels of oil equivalent) in his analysis, but oil is used in this report. Each has advantages. Shapes of the field size distributions of the tectonic setting perhaps should correlate better with BOE than with either oil or gas individually, because the geologic forces that determine the size of the traps are not necessarily correlated with the types of hydrocarbon that fill them. The formula to combine oil and gas should, in that case, be based on volume equivalence rather than energy equivalence, as Mast is now developing (Mast, 1986). On the other hand, once the shape of the distributions of BOE's are determined, then they must be split to derive the distributions for oil and gas separately and their correlation one to the other. Because this report is designed to emphasize techniques rather than results, the simpler case of analyzing just oil field size distributions is presented.

This paper focuses on field size distributions at two extremes, a concentrated habitat and a dispersed habitat (Coustau, 1980). In the concentrated habitat, the largest few fields make up a higher percentage of the total oil in the basin. In contrast, a dispersed habitat displays a more equal distribution of field sizes.

Klemme argues that his graphs show that rifted margin basins (as defined in Klemme, 1983) tend to show a concentrated habitat and the craton margin and craton interior basins tend to be dispersed. He also argues that increasing basin size (measured by total hydrocarbons) correlates with a more dispersed habitat. This may be due, however, to Klemme's measure of shape. His measure is the ratio of the sum of the five largest fields to the total amount of oil and gas in the basin. One might expect that as the total number of fields increases, the percent that the largest five make of the total would decrease.

Klemme finds that exploration maturity is an important factor in the measure of dispersion. It is well known that as a basin is progressively explored, the large fields tend to be discovered early. Immature basins, Klemme claims, will tend to be more concentrated and mature basins will tend to be dispersed. He also argues that field sizes in basins with clastic reservior rocks tend to be more dispersed than carbonate reservoirs.

THE LOGNORMAL DISTRIBUTION APPLIED TO FIELD SIZES

Three methods of fitting a lognormal distribution were tested. Two of the methods require that logarithms of the data be taken before the distribution is fit. Small observations influence greatly the fitting of a distribution when logarithms are taken, so before fitting the samples, the smallest fields were excluded by first deleting zero values and then deleting those fields less than a factor, arbitrarily chosen, of one thousand smaller than the sample mean of the nonzero fields. The record of the small fields is influenced greatly by economics, so it is assumed that virtually no significant information is lost. In addition, the computer file of field sizes was constructed in such a way that fields that were mostly gas with a little oil would be included in the oil file as just small oil fields.

Deleting these extremely small fields is considered different from the truncating included in the third method described below. In the first two methods even though the lognormal distribution is applied to these samples with the very small fields deleted, the sample is considered complete. In the truncated case, much more of the sample is removed and, when fitting, the sample is assumed to be truncated with an unknown number of missing observations, and the fitting procedure estimates the missing number.

This report focuses on the shape of the field size distribution. The data is analyzed to determine whether basins are, for instance, concentrated or dispersed, rather than how much oil each basin contains. The analysis is based on the shape of the field size distribution, which mesures the amount of oil accounted for by the largest few fields, rather than how many barrels of oil is contained in, for instance, the largest field. For this reason, each sample, after deleting the very small fields, was normalized by dividing the observations by the sample mean. This means that, for most of the analysis, it does not matter how large absolutely each of the fields is, because each is expressed as a multiple of the sample mean. In this way the larger observations and the scale parameters from each of the basins are comparable across basins.

Fitting the Lognormal Distribution

The two parameter lognormal distribution was fit using both arithmetic moments and the method of maximum likelihood (moments after taking logs). If Y is distributed as a lognormal, and if

$$X = \ln(Y), \qquad [1]$$

then X is distributed with a normal density function

$$f(X) = \frac{1}{b\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-a}{b}\right)^2}, \qquad [2]$$

where a = mean
and b = standard deviation of the normal deviate.

The method of moments uses the following equations for the estimates of the parameters:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i, \qquad [3]$$

$$s_y^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2, \qquad [4]$$

and

$$\hat{a} = \ln\left(\frac{\bar{y}}{\sqrt{1 + \frac{s_y^2}{\bar{y}^2}}}\right), \qquad [5]$$

$$\hat{b}^2 = \ln\left(1 + \frac{s_y^2}{\bar{y}^2}\right). \qquad [6]$$

The method of maximum likelihood uses:

$$\hat{a} = \frac{1}{n} \sum_{i=1}^{n} x_i, \qquad [7]$$

and

$$\hat{b}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{a})^2 \ . \qquad [8]$$

Fitting a Truncated Lognormal Distribution

For partially explored basins, a greater percentage of the larger fields have been found than of the smaller ones. Even in well-explored basins, some of the smaller fields still remain to be discovered. One way to partially neutralize this problem is to fit the lognormal distribution to only the larger fields in such a way that the number of smaller fields is implied by the distribution. This assures a better fit over the portion of the curve of interest, the larger fields, and in most cases predicts a number of small fields greater than the number already discovered and recorded.

The fitting procedure (Cohen, 1959) operates as follows. A cutoff point is arbitrarily determined *a priori*, in this case a factor of 1000 smaller than the largest field. All fields smaller that the cutoff are ignored. The lognormal distribution is fit through the remaining larger fields, assuming that they represent all the fields above the cutoff. The criteria for the fit is based on maximum liklihood for those largest fields, and the fitting procedure delivers as output an implied number of fields and their distribution below the cutoff point.

This method involves solving for the parameters using an implicit equation in one unknown. First, let

$x_0 \equiv$ truncation point (an undetermined number of observations $< x_0$ are excluded).

and let ζ be the (unknown) standardized truncation point

$$\zeta \equiv \frac{(x_0-\mu)}{\sigma} \qquad [9]$$

Let $\psi(\zeta)$ and $F(\zeta)$ be standardized normal frequency and distribution functions, respectively, and define

$$Z(\zeta) \equiv \frac{\psi(\zeta)}{1-F(\zeta)} . \qquad [10]$$

Then solve the implicit formula for $\hat{\zeta}$:

$$\frac{1 - Z(\hat{\zeta})\left(Z(\hat{\zeta})-\hat{\zeta}\right)}{\left(Z(\hat{\zeta})-\hat{\zeta}\right)^2} = \frac{s^2}{(\bar{x}-x_0)} , \qquad [11]$$

where $\bar{x}$ and s^2 are the mean and variance of the truncated sample. The parameter $\hat{\theta}$ is estimated by

$$\hat{\theta} = \frac{Z(\hat{\zeta})}{Z(\hat{\zeta})-\hat{\zeta}} , \qquad [12]$$

and the lognormal parameters by

$$\hat{a} = \bar{x} - \hat{\theta}(\bar{x}-x_0)^2 \qquad [13]$$

and

$$\hat{b} = s^2 + \hat{\theta}(\bar{x}-x_0)^2 . \qquad [14]$$

RESULTS

Choosing the Best Fitting Procedure

Before discussing the connection between the statistics and the geology, it is important to decide which of the three fitting procedures best represents the basins. Goodness of fit tests were not applied because most samples contained so many observations that most tests would reject the lognormal hypothesis. The coefficient of skew and kurtosis can give a rough measure of the normality of the logarithms, however, and the values given in Table [2] (as well as the estimates of the parameters given in Tables [2] and [3]), show general agreement with the lognormal hypothesis for most of the samples. For 35 of the 50 basins, the skew of the logarithms, which is 0. for a normal distribution, is between -0.3 and 0.4. The kurtosis values on the logs center between 2.5 and 3.0, a little lower than 3.0, the exact value for a normal distribution.

Two criteria for assessing the fit:

Two measures of the quality of fit that provide more interpretive information were used instead of the usual goodness of fit tests. First, plots of both the empirical distribution function and the fitted distribution were made for each of the basins and each of the methods of fitting.

The second measure, a statistic in this report called the quantile measure, was developed that describes how well each observation is fit by the distribution. The quantile measure is defined as follows. A sample is fit by a lognormal distribution using any one of the fitting procedures. The density function of the order statistic for each of the ranked observations is approximated from the fitted lognormal distribution. The actual ranked observation is compared with its predicted density function by calculating its quantile measure, that is, the fraction of the density function (of the order statistic) to the left of the observation. If that quantile measure (fraction) is high, say close to 1., then the observation is higher than one would expect from that fitted lognormal distribution. The quantile measures are expected to be approximately uniformly distributed between 0. and 1.0. The closer the values of the quantile measures are to either 0. or 1.0, the poorer the fit over that portion of the curve. Note, however, that any test for goodness of fit based on this measure would need consider the serial correlation between the ranked observations and the fact that the parameters of the lognormal distribution are estimated from the sample (which tends to move the quantile measures away from the endpoints).

Method of moments:

Analysis of both the graphs and the quantile measures indicates that each of the three fitting procedures has difficulty fitting properly all of the samples. For example, Figures [1] and [2] show the fit with method of moments and maximum likelihood applied to the East Texas basin. Figure [1] shows that the method of moments fits very poorly over most of the distribution when the graph is plotted on a logarithm scale. This lack of fit is common for many of the basins.

The maximum likelihood method:

Figure [2] shows that the maximum likelihood method seems to fit the sample well, and the East Texas basin is typical in that regard. A drawback with the maximum likelihood method, however, is that the distribution does not predict the same amount of oil as the sample being fit. The average size of fields in each of the samples is 1.0, because the samples are normalized. But the average size for the fitted distribution for the East Texas basin is 0.47, only half that of the sample. The maximum likelihood method does not preserve the sample mean.

The East Texas basin is not typical in that the fitted distribution for that basin predicts less oil than the sample. For the group of all 50 basins, the average predicted amount is 2.1 times that of the sample. Forty of the 50 basins predict more oil than was present in the sample.

This overprediction is explained in part by the quantile measures displayed in Table [4]. In general, for those basins with a predicted amount of oil greater than the sample (of which the East Texas basin is the opposite), the top three fields have low quantile measures. This means the fitted distribution overestimates the sizes of those fields. The interpretation is that the maximum likelihood procedure forces the lognormal curve to fit over the majority of the sample, even though the larger fields are overestimated. Thus the total amount of oil is overestimated as a consequence of overestimating the largest fields. Although this overestimation is a shortcoming of the fitting procedure, it is of relatively less importance to the methods in this report, in which the amount of oil in the analog basin is ignored and the shape of its field size distribution is the characteristic of note.

The truncated lognormal:

As expected, the Cohen method of fitting a truncated lognormal fits better in most cases, using the quantile measure as the criterion. It adds one degree of freedom, the number of implied fields smaller than the cutoff, and it only fits a portion of the sample. But the method does not always converge on a solution, for example in the East Texas case. The total amount of oil predicted by the fitted distribution is even higher on average than the maximum likelihood method (it averages 2.8), and the increase is due mostly to the fitting procedure rather than an increased amount of oil in the portion of the curve below the cutoff.

The choice: the maximum likelihood method:

The maximum likelihood method applied to each of the samples was chosen for the analysis presented later in this section. It is not dependent on an *a priori* determination of a cutoff, as in the truncated procedure; the plots showed that it fit most of the basins at least moderately well; there were no basins in which the fitting procedure failed; and it is probably the most common method for fitting a lognormal distribution. The fact that it did not preserve the sample mean is relatively less important than other considerations, because only the shape of the distribution was used in the later analysis.

The Shape of the Field Size Distribution

The parameter b, which is the standard deviation in logarithm space (after taking logarithms), was used as the measure of the shape of the distribution for each of the samples. Using the parameter b as the measure of shape is equivalent to using either the coefficient of variation or the coefficent of skew, both in arithmetic space (before logarithms are taken), because all three statistics are related by a one-to-one function. Examples of lognormal distributions with the same mean and different b's are shown in Figure [3]. High values of b indicate that a greater fraction of oil is contained in the larger fields, that is, the distribution has a thicker right hand tail. A high value of b would indicate a concentrated habitat and vice versa.

Influence of Geology

An important test of this method is whether or not geologic information, such as the plate tectonic classification presented in Table [1], correlates with the estimated shape parameter $\hat{b}$. The relationship between the two parameters is presented in Figure [4] and shows that plate tectonic classification has a marked influence on shape.

The set of 50 basins was disaggregated by tectonic category. Confidence limits were calculated (Ryan et al., 1982) and plotted for each of the collections of $\hat{b}$'s for each category. For instance, for category 1, Craton Center, most of the $\hat{b}$s were between about 2.1 and 2.6. The median value is plotted as a +, in this case at about 2.4. The confidence limits can be interpreted to mean that the $\hat{b}$ values are "significantly different" at a 5% level for any pair if the confidence limits do not overlap, which they do not, for instance, for types 2 and 3.

Figure [4] shows that the plate tectonic classification information is important in influencing the shape. The Figure can be interpreted to mean that basin types 1,2, and 5, (Craton centers, Craton margin shallow, and Craton margin other) have a dispersed habitat and basin types 3,4, and 7 (Craton margin deep, Craton margin northern and W. Rockies, and Subduction) have a concentrated habitat. Type 6 (Downwarp) is in the middle tending to concentrated and type 8 (Delta) has only one example, which shows it as concentrated. Other than the result for the single delta basin, these results are very similar to those presented by Klemme.

From analysis of the 50 basins, other statistical measures also correlate with the plate tectonic classification. The total amount of oil and the number of fields in a basin present similar results. This leads to a problem of interpretation. Is there something about large basins, measured by number of fields or total amount of oil, that influences the shape of the statistical distribution more than this measure of regional geology? More investigation is necessary to answer that question.

A METHOD TO ESTIMATE THE SHAPE OF FIELD SIZE DISTRIBUTIONS

Consider an example of a simplified assessment of a frontier basin. The basin risk, θ, is defined as the probability that no commercial oil will be discovered, the expected number of fields N, and expected total amount of oil T have already been determined. As the first step in assigning a field size distribution, the frontier basin is classified into one of the eight tectonic categories presented in Table [1], based on known geology. The shape parameter is inferred from Figure [4] by using the median value for that category. For instance, if the category is Subduction (Type 7), the shape parameter b of the lognormal distribution would be 2.8.

The total oil T and number of fields N determine the average field size in arithmetic space, so the other lognormal parameter a can be calculated by:

$$a = \ln\left(\frac{T}{N}\right) - \ln\left(e^{b^2/2}\right) \qquad [15]$$

The resource estimate for the frontier basin is presented as having resource potential of zero with probability θ and, with probability $(1 - \theta)$, has an uncertain amount of resources represented by a random sample of N observations drawn from a lognormal distribution with parameters a and b.

This simple approach can be modified to model more complex resource assessment tasks. The traditional assumption is that risk includes two possible outcomes -- zero oil with probability $(1 - \theta)$ and a successful scenario with probability θ. This approach can be refined to include several outcomes with probabilities θ_1, θ_2, etc., as long as the probabilities sum to 1.0. None of the outcomes need be a zero oil option.

In the example above, the total oil T was determined before application of the field size distribution. Using numerical approximations, the parameters a and b could be estimated using information on the amount of the oil contained in the largest field, instead of T. This is close to the method used by Exxon as reported by (White et al., 1983).

The approach could be made more general. The simple example described above can be thought of as based on determination of risk and two more "degrees of freedom", T and N. Other information could substitute for these two estimates. The degree of freedom represented by the amount of total oil T could be replaced by the amount of oil contained in, for instance, the largest ten fields. The expected number of fields N could be replaced by the number above a higher cutoff, for instance a factor 100 smaller than the largest field.

There are other potential modifications of the method. A minor modification would be to decrease the number of analog values of b from eight to two. Each of the eight tectonic types would be reclassified into either dispersed or concentrated habitats, each with a corresponding b.

Alternatively, the choice of the shape parameter, b, could be based on information other than the place tectonic classification. Size in terms of number of fields or total amount of oil could be used to determine the appropriate analog b.

Another potential use of the results is to apply the information in a more disaggregated form. A frontier basin could be compared individually, based on geology and size, to any one of the 49 analog basins. The total amount of oil, the size of the largest field, the expected size of the largest fields as fit by the lognormal, the number of fields greater than a particular cutoff, could all be compared to predictions made by other techniques.

CONCLUSIONS

In this report, a method was developed for resource assessment in a frontier basin that uses oil field size distributions from analogue basins to infer a shape for the field size distribution in the frontier basin. The numerical results show an apparently statistically significant connection between the geology and the shape of the field size distribution for U.S. basins. The method is substantially different from other techniques, such as those described in the Introduction. An approach to perform a simple example of resource assessment was presented, and modifications were suggested that would aid this type of resource assessment technique. For instance, total oil could be inferred by anchoring on an estimate of the largest field and the number of fields.

Another use for the method is to corroborate other methods. When enough information is available that the other methods are feasible, this method can affirm the shape of the field size distributions generated from other methods by comparing those field size distributions with analogue basins.

This method required a choice among fitting techniques and the development of new statistical perspectives, both of which proved useful in understanding the application of field size distributions. Although each of the three methods of fitting a lognormal distribution has some drawbacks, the method of maximum liklihood proved best for the methods presented in this report. It is interesting to note, however, that maximum liklihood methods affect the prediction of total oil as much as was shown in Table [3]. The quantile measure was valuable in analyzing the success of the fitting procedures. Normalizing the samples was helpful in comparing the shapes of the distributions.

Finally, certain intermediate numerical results and the data sets themselves should provide useful information for hydrocarbon resource assessment. The results generated by this analysis are similar to those developed by Klemme even though Klemme's data set was based on a different field file, a different measure of shape, and on barrels of oil equivalent rather than barrels of oil.

REFERENCES

Arps, J. J. and Roberts, T. G., 1958, Economics of drilling for Cretaceous oil and east flank of Denver-Julesburg basin: Amer. Assoc. Pet. Geol. Bull., v. 42, p. 2549-2566.

Attanasi, E. D. and Drew, L. J., Lognormal Field Size Distributions as a Consequence of Economic Truncation, Internal Publication, 12/83, 24 p.

Baker, R. A., and others, Geologic Field Number and Size Assessments of Oil and Gas Plays, AAPG Bulletin v. 68, no 4, April 1984, p. 426-437.

Cohen, A. Clifford, Jr., Simplified Estimators for the Normal Distribution When Samples Are Singly Censored or Truncated, Technometrics, v. 1, no. 3, Aug 1959, p. 217-237.

Coustau, Henri, Habitat of Hydrocarbons and Field Size Distribution: A First Step Towards Ultimate Reserve Assessment, in Assessment of Undiscovered Oil and Gas, Proceedings of the Seminar, Kuala Lumpur, Malaysia, March 1980, Issued by CCOP Project Office, UNDP Technical Support for Regional Offshore Prospecting in East Asia, (RAS/80/003), p. 180-194.

Dolton, G. L., and others, 1981, Estimates of Undiscovered Recoverable Conventional Resources of Oil and Gas in the United States, Geological Survey Circular 860, 87 p.

Energy Information Administration, Dept. of Energy, An Economic Analysis of Natural Gas Resources and Supply, Working Draft, August 1985, 239 p.

Energy Information Administration, Dept. of Energy, Oil and Gas Field Code Master List -- 1982, DOE/EIA-0370 (82), Dist. Cat. UC-92, 539 p.

Forman, David J. and Hinde, Alan L., Improved Statistical Method for Assessment of Undiscovered Petroleum Resources, AAPG Bulletin v. 69, no. 1, Jan 1985, p. 106-118.

Houghton, J. C., and others, 1985, Some Statistical Methods for Estimating Undiscovered Resources in the U.S. Oil and Gas Assessment Project [abs]: Second Annual V.E. McKelvey Forum on Mineral and Energy Resources, U.S. Geological Survey, Denver, Colorado, Feb. 1986.

Klemme, H. D., Field Size Distribution Related to Basin Characteristics, Oil and Gas Journal, Dec 26, 1983, p. 168-176.

Mast, Richard F., personal communication, 1986.

Meisner, J. and Demirmen, F., The Creaming Method: A Bayesian Procedure to Forecast Future Oil and Gas Discoveries in Mature Exploration Provinces, J.R. Statist Soc A(1981), 144, Part 1, p. 1-31.

Meyer, R. G., Geologic Province Code Map for Computer Use, Amer. Assoc. of Petro. Geol. Bulletin, v. 54, No. 7, July 1970, p. 1301-1305.

Meyer, R. G., and others, Chapter 2: The Resource Potential of Small Oil and Gas Fields, in The Future of Small Energy Resources, R. F. Meyer and J. C. Olson, editors, McGraw Hill, N.Y., 1983.

Minerals Management Service, Estimates of undiscovered, economically recoverable oil and gas resources for the outer continental shelf as of July 1984, OCS report # MMS 85-0012, M.M.S., Dept. of the Interior, 1985, 45 p.

Nehring, Richard, The Discovery of significant oil and gas fields in the United States, v. 1, Jan 1981, R-2654/1-USGS/DOE, p. 93.

OCS Lands Act Amendments, 1978, PL95-372; and Outer Continental Shelf Lands Act of 1953, PL 82-212.

Office of Minerals Policy and Research Analysis, Department of the Interior, Final Report of the 105(b) Economic and Policy Analysis, Alternative overall procedures for the exploration, development, production, transportation and distribution of the petroleum resources of the National Petroleum Reserve in Alaska (NPRA), (GPO Stock Number 024-000-00865-7), December 15, 1979, 145 p.

Procter, R. M., McCrossan, R. G., Andreatta, G., and Kaufman, G. M., The Shape of Oil Deposit Attribute Data, MIT Energy Laboratory Working Paper, MIT-EL 82-028WP, Cambridge, Mass., 1982, 48 p.

Root, D. H., and Attanasi, E. D., 1985, Estimation of Ultimate Oil Field Recovery from Production Data using Rules of Thumb, personal communication, 31 p.

Ryan, Thomas A. Jr., Minitab Reference Manual, section on BOXPLOT with NOTCH option, companion to Minitab Student Handbook, Duxbury Press, Boston, Mass., 11/1/82, 154 p.

University of Oklahoma Information Systems Program, 1981, Petroleum Data System for North America users guide: Norman, Oklahoma, University of Oklahoma, 274 p.

Woods, Thomas J., Long-Term Trends in Oil and Gas Discovery Rates in Lower 48 United States, AAPG Bulletin, v. 69, No. 9, Sept 1985, p. 1321-1326.

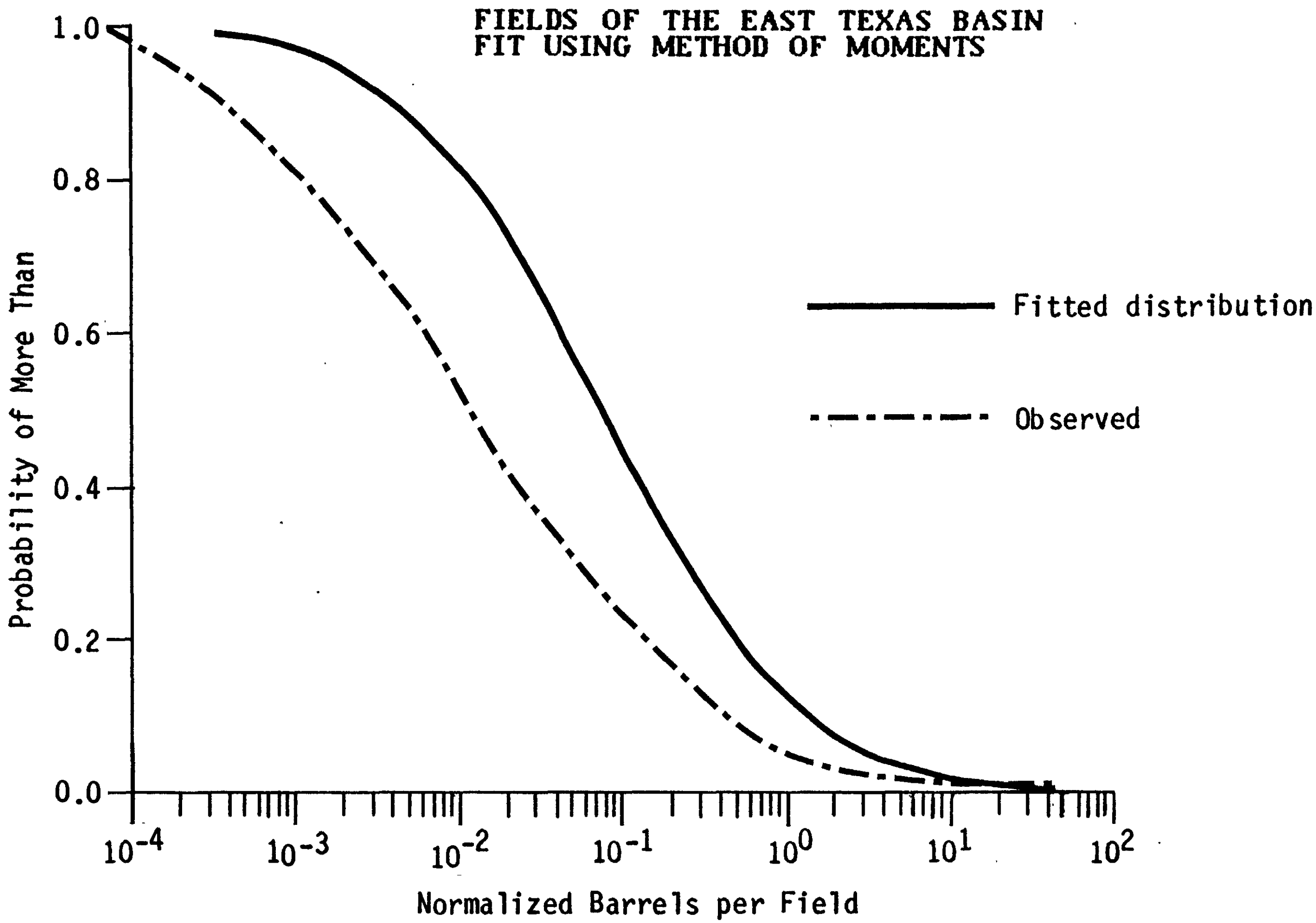
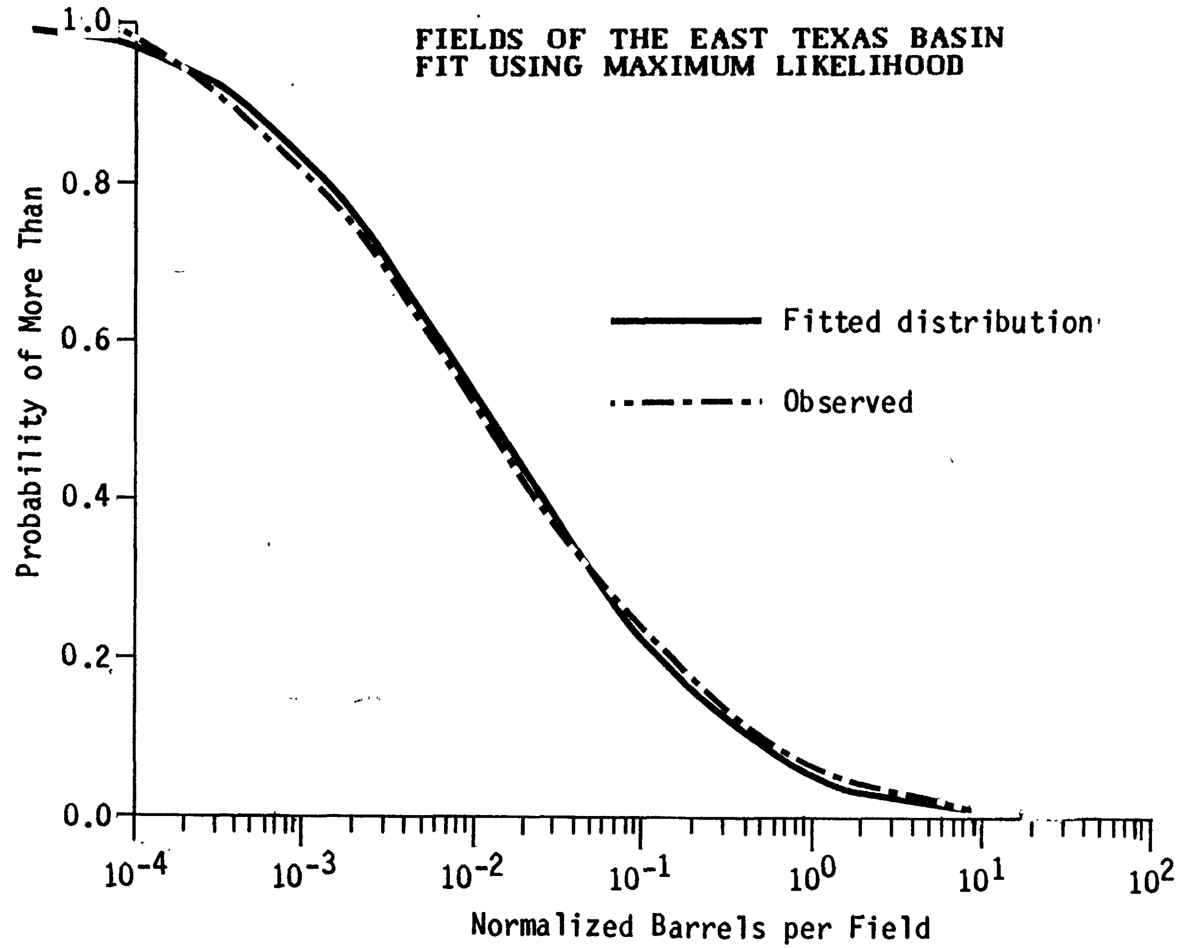

FIELDS OF THE EAST TEXAS BASIN
FIT USING METHOD OF MOMENTS
Fitted distribution
Observed
Probability of More Than
1.0
0.8
0.6
0.4
0.2
0.0
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
Normalized Barrels per Field

Figure 1



FIELDS OF THE EAST TEXAS BASIN
FIT USING MAXIMUM LIKELIHOOD
Fitted distribution
Observed
Probability of More Than
1.0
0.8
0.6
0.4
0.2
0.0
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
Normalized Barrels per Field

Figure 2

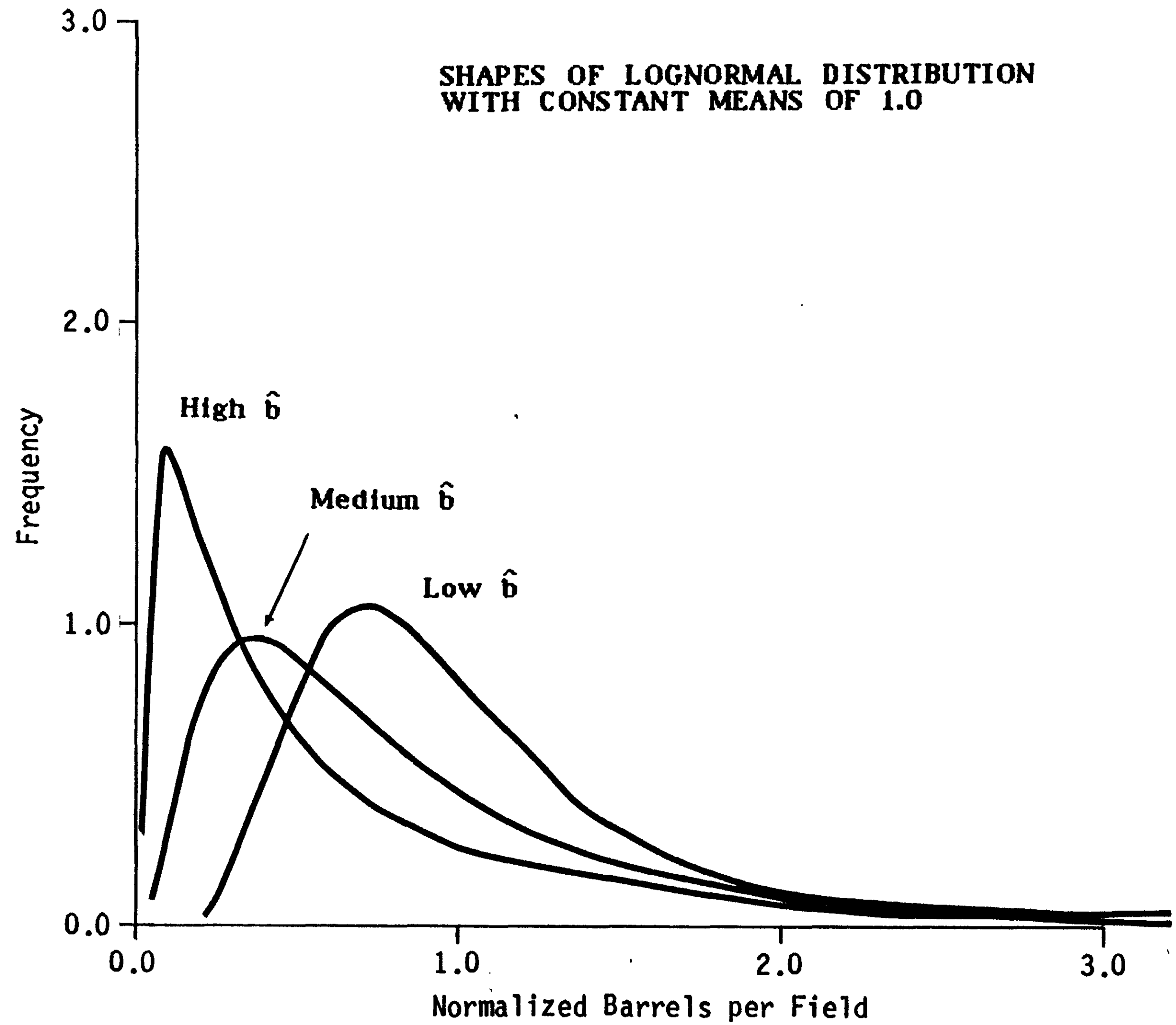

3.0
2.0
1.0
0.0
Frequency
SHAPES OF LOGNORMAL DISTRIBUTION
WITH CONSTANT MEANS OF 1.0
High $\hat{b}$
Medium $\hat{b}$
Low $\hat{b}$
0.0
1.0
2.0
3.0
Normalized Barrels per Field

Figure 3

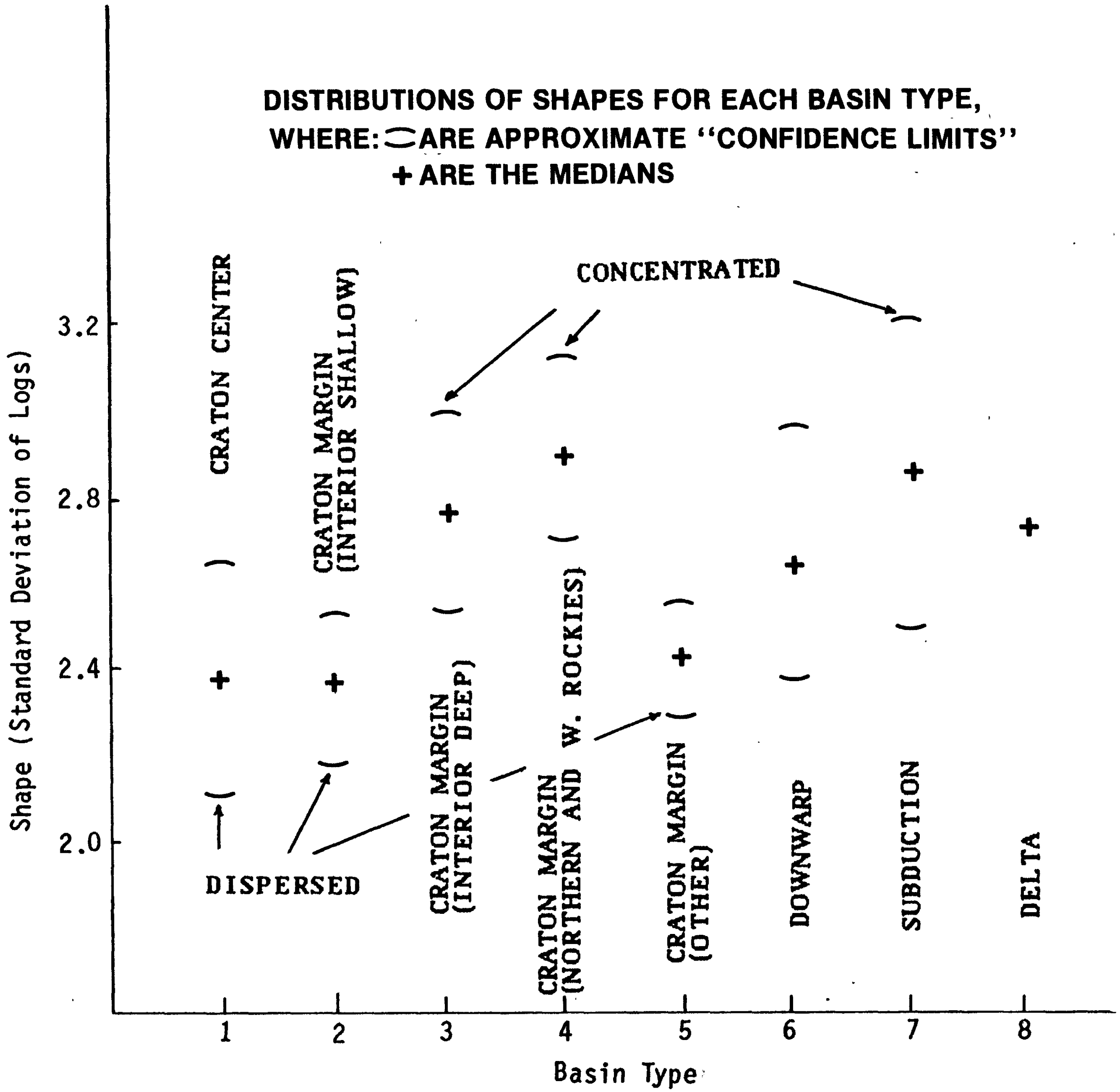

DISTRIBUTIONS OF SHAPES FOR EACH BASIN TYPE,
WHERE: ⊂ ARE APPROXIMATE "CONFIDENCE LIMITS"
+ ARE THE MEDIANS
CONCENTRATED
DISPERSED
CRATON CENTER
CRATON MARGIN (INTERIOR SHALLOW)
CRATON MARGIN (INTERIOR DEEP)
CRATON MARGIN (NORTHERN AND W. ROCKIES)
CRATON MARGIN (OTHER)
DOWNWARP
SUBDUCTION
DELTA
Shape (Standard Deviation of Logs)
3.2
2.8
2.4
2.0
1
2
3
4
5
6
7
8
Basin Type

Figure 4

Table 1

BASIN CLASSIFICATION BY TECTONIC TYPE

	Type	Example	No basins
1.	Craton Center	Williston	3
2.	Craton Margin (Interior Shallow)	Chautauqua Platform	10
3.	Craton Margin (Interior Deep)	Anadarko	6
4.	Craton Margin (Northern and W. Rockies)	Big Horn	8
5.	Craton Margin, other	Warrior Basin	9
6.	Downwarp	Gulf Coast	6
7.	Subduction	L. A. Basin	7
8.	Delta	Louisianna Offshore	1

			50

Table 2

BASIC STATISTICS OF THE 49 BASINS

No.	AAPG Basin	Type	no. fds	no.> cut.	total oil	largest field	(on logs) skew	(on logs) kurt.
160	Appalachian Basin	3	1230	114	1164122.	646750.	-0.96	3.59
200	Warrior Basin	5	61	22	2650.	868.	-0.51	3.41
210	Mid-Gulf Coast Basin	6	582	524	2856856.	551056.	-0.24	3.19
220	Gulf Coast basin	6	5521	3553	21294126.	783115.	0.05	2.56
230	Arkla basin	6	746	568	3502903.	362487.	-0.22	2.54
260	East Texas basin	6	721	456	8711999.	5240452.	.28	*
300	Cincinnati arch	5	283	39	106256.	105295.	1.42	7.88
305	Michigan basin	5	683	418	998582.	55721.	-0.29	2.53
315	Illinois basin	5	1499	1194	4033550.	402659.	-0.07	2.59
335	Forest City basin	1	140	91	103494.	21662.	-0.27	2.60
345	Arkoma basin	3	236	76	437306.	163765.	0.49	3.34
350	S. Oklahoma province	3	519	441	3692084.	1351664.	0.23	2.97
355	Chautauqua platform	2	1645	1420	6212344.	751933.	-0.07	2.97
360	Anadarko basin	3	1914	1557	2546917.	322858.	-0.12	2.94
365	Cherokee basin	2	281	219	269221.	33337.	-0.25	2.43
370	Nemaha anticline	2	331	293	727490.	295345.	-0.15	2.66
375	Sedgwick basin	2	604	504	744367.	76265.	-0.18	2.86
380	Salina basin	1	38	33	46478.	26529.	0.03	2.55
385	C. Kansas uplift	2	1741	1613	2752722.	267689.	-0.14	3.28
390	Chadron arch	5	76	74	93367.	51047.	0.07	2.68
395	Williston basin	1	307	287	1650924.	126249.	-0.34	3.48
400	Ouachita province	6	88	77	55165.	18150.	0.17	1.85
410	Llano uplift	5	105	69	11626.	2969.	0.00	2.23
415	Strawn basin	5	53	12	1775.	1553.	1.03	3.05
420	Fort Worth syncline	2	752	639	640640.	61910.	-0.19	2.69
425	Bend arch	2	2345	2051	2016850.	277548.	-0.09	3.30
430	Permian basin	3	3525	2934	27602381.	1869386.	0.13	2.62
435	Palo Duro basin	2	246	226	1346003.	575375.	0.17	3.63
440	Amarillo arch	3	23	11	1047416.	1043962.	1.56	5.11
450	Las Animas arch	5	93	83	54011.	11642.	-0.26	2.89
500	Sweetgrass arch	4	82	39	323252.	166521.	0.01	2.75
510	C. Montana uplift	5	34	29	55001.	10088.	-0.51	2.61
515	Powder River basin	2	408	384	1966765.	121944.	-0.41	2.64
520	Big Horn basin	4	135	117	3165289.	849428.	0.22	2.67
530	Wind River basin	4	136	92	1859797.	1199625.	0.10	2.64
535	Green River basin	4	291	154	844115.	228593.	-0.20	2.35
540	Denver basin	2	1127	987	919638.	63188.	-0.55	3.24
575	Uinta basin	4	64	44	393940.	83323.	0.28	2.22
580	San Juan basin	4	143	89	236508.	39400.	-0.31	2.44
585	Paradox basin	4	100	65	497155.	378280.	0.32	2.62
595	Piceance basin	4	77	19	965195.	802949.	1.58	4.92
730	Sacramento basin	7	123	5	60430.	38163.	-0.32	1.24
740	Coastal basins	7	15	13	542476.	456644.	0.68	2.76
745	San Joaquin basin	7	133	109	9667843.	1875070.	-0.19	2.13
750	Santa Maria basin	7	38	32	1713382.	218367.	-0.36	1.95
755	Ventura basin	7	50	45	1864448.	920366.	0.39	2.43
760	Los Angeles basin	7	89	75	9368242.	2722006.	-0.33	2.37
820	Cook Inlet basin	7	21	6	1267632.	638153.	-1.18	3.36
952	Louisiana offshore	8	688	302	9121242.	497415.	-0.76	3.06
956	Texas offshore	6	42	9	54448.	24737.	-0.64	3.02

Notes for Table [2]

Type:	tectonic classification from Table [1]
no. fds:	number of fields greater than zero
no. > cut.:	number of fields greater than .001 of average size
skew:	coefficient of skew on logs
kurt:	kurtosis on logs
*:	unable to calculate parameter

Table 3

LOGNORMAL FIT TO THE 49 BASINS

			maximum liklihood				truncated lognormal						
							no.	impli	implied				
No.	AAPG Basin	Type	xmu	xsd	ymu	ysd	trun.	no.	amount	xmu	xsd	ymu	ysd
160	Appalachian Basin	3	-2.57	2.87	4.71	291.5	75	94.	784956.	-1.47	1.59	0.81	2.7
200	Warrior Basin	5	-2.14	2.51	2.74	64.5	19	20.	3673.	-1.70	2.05	1.51	12.4
210	Mid-Gulf Coast Basin	6	-2.44	2.34	1.35	21.1	261	628.	3206453.	-2.79	2.33	0.93	14.2
220	Gulf Coast basin	6	-3.36	2.81	1.83	96.7	1162	3152.	33203470.	-2.92	2.64	1.75	57.2
230	Arkla basin	6	-2.76	2.71	2.53	101.2	311	435.	3796781.	-1.69	2.01	1.41	10.7
260	East Texas basin	6	-4.33	2.67	0.47	16.8	*	*	*	*	*	*	*
300	Cincinnati arch	5	-5.82	2.33	0.04	0.7	*	*	*	*	*	*	*
305	Michigan basin	5	-2.56	2.72	3.16	129.8	290	356.	1464173.	-1.82	2.17	1.71	18.1
315	Illinois basin	5	-3.31	2.71	1.43	56.6	434	2274.	4568917.	-4.63	2.87	0.59	36.5
335	Forest City basin	1	-2.66	2.74	3.02	131.7	62	72.	136296.	-1.72	2.11	1.65	15.2
345	Arkoma basin	3	-4.26	2.83	0.77	42.8	25	241.	18566909.	-9.85	4.99	13.34	a
350	S. Oklahoma province	3	-3.52	2.49	0.66	14.9	112	5388.	7212444.	-10.11	4.07	0.16	634.1
355	Chautauqua platform	2	-3.02	2.49	1.10	25.0	429	2100.	7227519.	-4.03	2.75	0.78	35.1
360	Anadarko basin	3	-2.75	2.43	1.24	24.1	514	1894.	2680053.	-3.10	2.43	0.86	16.6
365	Cherokee basin	2	-2.57	2.72	3.13	128.5	142	167.	366948.	-1.52	2.04	1.77	14.3
370	Nemaha anticline	2	-2.88	2.60	1.64	48.5	110	171.	681340.	-1.42	1.94	1.59	10.5
375	Sedgwick basin	2	-2.72	2.54	1.64	41.4	289	539.	883253.	-2.74	2.38	1.10	19.1
380	Salina basin	1	-2.53	2.41	1.47	27.2	23	28.	58407.	-2.06	2.20	1.45	16.6
385	C. Kansas uplift	2	-2.72	2.28	0.90	12.3	556	5670.	3278861.	-5.84	3.08	0.33	39.4
390	Chadron arch	5	-3.06	2.60	1.37	40.8	38	87.	177296.	-3.67	2.88	1.61	102.3
395	Williston basin	1	-1.88	2.12	1.44	13.7	241	288.	1866443.	-1.81	1.96	1.12	7.6
400	Ouachita province	6	-3.56	3.23	5.23	968.2	37	42.	95973.	-1.16	2.15	3.17	32.2
410	Llano uplift	5	-2.24	2.37	1.76	29.1	55	75.	23194.	-2.49	2.49	1.83	40.6
415	Strawn basin	5	-3.02	2.29	0.66	9.1	9	*	*	*	*	*	*
420	Fort Worth syncline	2	-2.13	2.23	1.44	17.4	406	524.	601348.	-1.45	1.77	1.14	5.4
425	Bend arch	2	-2.60	2.12	0.70	6.7	523	6118.	1634292.	-4.97	2.71	0.27	10.6
430	Permian basin	3	-3.60	2.77	1.28	60.4	706	3035.	39655959.	-3.68	2.83	1.38	76.8
435	Palo Duro basin	2	-3.44	2.29	0.44	6.0	73	*	*	*	*	*	*
440	Amarillo arch	3	-5.94	3.08	0.30	35.6	2	*	*	-17.17	8.39	*	*
450	Las Animas arch	5	-2.35	2.46	2.00	42.0	64	82.	84890.	-2.26	2.33	1.57	23.9
500	Sweetgrass arch	4	-3.32	2.86	2.14	127.7	20	24.	408240.	-1.77	2.23	2.04	24.5
510	C. Montana uplift	5	-1.24	1.92	1.82	11.5	28	28.	86791.	-1.13	1.79	1.61	7.9
515	Powder River basin	2	-2.37	2.64	3.05	100.6	262	299.	2224903.	-1.34	1.85	1.45	7.9
520	Big Horn basin	4	-3.01	2.57	1.34	36.8	66	*	*	*	*	*	*
530	Wind River basin	4	-3.60	2.77	1.28	60.6	38	116.	2687700.	-4.14	2.92	1.14	81.9
535	Green River basin	4	-3.20	3.02	3.98	388.8	91	270.	2120054.	-4.48	3.11	1.43	182.3
540	Denver basin	2	-1.83	2.22	1.89	22.4	679	839.	923549.	-1.19	1.64	1.18	4.4
575	Uinta basin	4	-3.74	3.15	3.40	491.2	24	33.	1477097.	-2.77	2.95	4.94	392.6
580	San Juan basin	4	-2.23	2.58	3.01	84.8	69	78.	411460.	-1.67	2.16	1.97	20.6
585	Paradox basin	4	-4.60	3.05	1.07	114.8	20	92.	1178055.	-5.79	3.55	1.66	909.7
595	Piceance basin	4	-6.76	3.42	0.40	142.7	3	*	*	*	*	*	*
730	Sacramento basin	7	-2.62	3.53	37.82	a	3	*	*	*	*	*	*
740	Coastal basins	7	-3.92	2.90	1.32	89.0	6	11.	7509475.	-4.42	3.78	15.29	*
745	San Joaquin basin	7	-3.02	3.10	5.91	723.3	66	75.	15223678.	-1.46	2.13	2.26	21.9
750	Santa Maria basin	7	-2.37	3.00	8.54	787.6	25	26.	5219265.	-1.38	2.32	3.72	55.1
755	Ventura basin	7	-2.79	2.42	1.15	21.8	27	41.	3881203.	-2.78	2.68	2.25	81.9
760	Los Angeles basin	7	-2.70	2.87	4.22	266.5	50	56.	11444956.	-1.42	1.95	1.63	10.9
820	Cook Inlet basin	7	-0.69	1.54	1.64	5.1	6	6.	2092724.	-0.69	1.54	1.65	5.2
952	Louisiana offshore	8	-2.12	2.77	5.64	265.3	237	248.	11907310.	-1.11	1.77	1.58	7.5
956	Texas offshore	6	-0.99	1.68	1.51	6.0	9	9.	85277.	-1.01	1.70	1.55	6.5

Notes for Table [3]

xmu: mean after taking logs

xsd: standard deviation after taking logs

ymu: mean without taking logarithms, also ratio of predicted total oil to total oil

ysd: standard deviation without taking logarithms

no trun: number of fields greater than the cutoff of 1000 times smaller than the largest field

impli no: number of total fields implied by the truncated lognormal fit

implied amount: total amount of oil implied by the truncated lognormal fit

*: unable to calculate

a: greater than 10,000.

Table 4

A COMPARISON OF QUANTILES FOR SELECTED BASINS

			.....Size.......				..................Quantiles..........									
		Num	amount of oil				maximum liklihood			method of mom.			truncated			ratio
No.	AAPG Basin	fds	tot	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	
160	Appalachian Basin	114	1.2	63.3	4.0	3.1	.3	.0005	.0007	.91	.03	.03	.98	.1	.1	4.7
200	Warrior	22	.003	7.2	7.2	3.2	.3	.7	.7	.5	.7	.96	.5	.8	.8	2.7
210	Mid-Gulf Coast basin	524	2.9	101.1	37.2	28.4	.5	.3	.3	.9	.7	.8	.6	.4	.5	1.4
260	East Texas basin	456	8.7	274	46.2	27.3	.96	.91	.92	.93	.7	.6	*	*	*	0.5
305	Michigan basin	418	1.0	23.3	21.3	20.9	.0005	.002	.01	.4	.7	.9	.02	.06	.2	3.2
335	Forest City basin	91	.1	19.0	12.8	11.7	.2	.3	.5	.7	.9	.96	.4	.5	.8	3.0
365	Arkoma basin	219	.3	27.1	19.4	16.7	.03	.05	.1	.7	.8	.9	.2	.3	.5	0.8
400	Ouachita province	77	.06	25.3	7.8	7.3	.25	.16	.35	.84	.6	.8	.4	.2	.4	5.2

Notes

Num fds: number of fields

tot: total oil in the basin in billion barrels

maximum liklihood: method of fitting 2-parameter lognormal based on moments of logarithms

method of mom.: method of fitting 2-parameter lognormal based on moments without logarithm transform

truncated: method of fitting 2-parameter lognormal based on fitting a truncated sample

ratio: ratio of predicted total oil to total oil, also ymu

Size
- 1st: size of the largest field relative to the mean
- 2nd: size of the second largest field relative to the mean
- 3rd: size of the third largest field relative to the mean

Quantile
- 1st: quantile of the largest field (see text)

*: unable to calculate