UNITED STATES DEPARTMENT OF THE INTERIOR
GEOLOGICAL SURVEY

PROCEDURES FOR PETROLEUM RESOURCE ASSESSMENT USED BY THE U.S. GEOLOGICAL SURVEY--STATISTICAL AND PROBABILISTIC METHODOLOGY

By

Robert A. Crovelli

Open-File Report 83-402

1983

This report is preliminary and has not been reviewed for conformity with U.S. Geological Survey editorial standards. Any use of trade names is for descriptive purposes only and does not imply endorsement by the USGS.

PROCEDURES FOR PETROLEUM RESOURCE ASSESSMENT USED BY THE U.S. GEOLOGICAL SURVEY--STATISTICAL AND PROBABILISTIC METHODOLOGY

By

Robert A. Crovelli

INTRODUCTION

Estimating the quantity of undiscovered crude oil or natural gas in an assessment area or sedimentary basin is a complex process consisting of four stages.

The first stage is the information- or data-gathering stage; types of information and data compiled relate to the following:

1. Petroleum geology.
2. Exploration history.
3. Seismic exploration.
4. Production and reserves statistics.
5. Analog basins or formations.

The second stage involves the analysis of the data by one or more of the following methods:

1. Extrapolation of historical trends (finding rates).
2. Areal or volumetric yield analysis.
3. Geochemical material balance analysis.
4. Play analysis.

The third stage is direct subjective estimation of a petroleum resource in an assessment area. A team of geologists uses a Delphi type of forecasting procedure which involves the following steps:

1. Presentation and discussion of the geological and geophysical data and interpretation.
2. Presentation and discussion of the data analyses made in stage 2.
3. Each team member independently makes his probabilistic assessments.

4. Team reviews all estimates and individuals may revise their estimates.
5. One or more iterations of steps 1-4 are possible.
6. Individual resource estimates are averaged.

In the third stage, Step 6, the team produces the following estimates: (a) an estimate of the likelihood of any recoverable resource being present, called the marginal probability, i.e., a subjective probability of the condition that the resource is actually present in recoverable quantities, and (b) conditional upon recoverable resource being present,

1. A low resource estimate corresponding to a 95-percent probability of more than that amount - this estimate is the 95th fractile (F_{95}).
2. A high resource estimate corresponding to a 5-percent probability of more than that amount - this estimate is the 5th fractile (F_5).
3. A modal ("most likely") estimate of the quantity of resource associated with the greatest likelihood of occurrence.

The fourth stage is the determination of a probability distribution for the quantity of undiscovered recoverable resource from the three conditional estimates and the marginal probability. This report deals with the fourth and last stage.

METHODOLOGY FOR PROCESSING PROBABILISTIC ASSESSMENTS OF UNDISCOVERED HYDROCARBON RESOURCES[1]

The procedures described above for estimating the undiscovered recoverable resources for an area involved subjective probabilities. For each assessment area the resource appraisal team expressed judgments as estimates of a marginal probability, two fractiles (F_{95} and F_5), and a modal value. These procedures were followed for all areas, assuming independence among the areas being assessed. Oil and total gas are separately assessed in foreign assessment.

Conditional Probability Distribution

The conditional low, high, and modal estimates of undiscovered recoverable resource are used to determine a conditional probability distribution of the quantity of undiscovered recoverable resource for an assessment area. The conditional probability distribution represents the judgmental probability distribution of the quantity of undiscovered recoverable resource conditioned on the recoverable resource being present. (This distribution is also referred to as the "unrisked" distribution.) The lognormal distribution is used as a probability model for the conditional probability distribution in an area.

A probability model can be interpreted as:

1. A subjective probability measure of uncertainty, as in risk analysis, where the probability distribution is called a subjective or judgmental probability distribution. This is the subjective probability concept.

---

[1]For a more detailed discussion of the probabilistic methodology for oil and gas resource appraisal, see Crovelli (1981).

2. A statistical probability measure of relative frequency, as in statistical analysis, where the probability distribution is a model for an empirical distribution. This is the statistical probability concept.

In the USGS method, the lognormal distribution is used to describe the uncertainty in the team's estimate of the quantity of undiscovered oil (or gas) in an area. It is a subjective, not statistical, probability distribution.

Because each pair of values among the three conditional estimates determines a lognormal distribution, there are three possible lognormal distributions. The fitted lognormal distribution with the largest standard deviation can be chosen in order to have the greatest spread. A more recent approach is to fit a three parameter lognormal to all three conditional estimates.

The conditional probability distribution can be described in several ways. One standard approach, which is used by the USGS, is the conditional "more-than" (or complementary) cumulative distribution function (cdf) (fig. 1 A) that gives the probability of more than a specific amount. From this function, all the fractiles can be obtained easily. The graph of this function has the value of 1 at the origin. Another approach is the conditional probability density function (fig. 1 B) in which an area under its curve represents probability. The shape of the curve of this function helps to visualize where the probability is located.

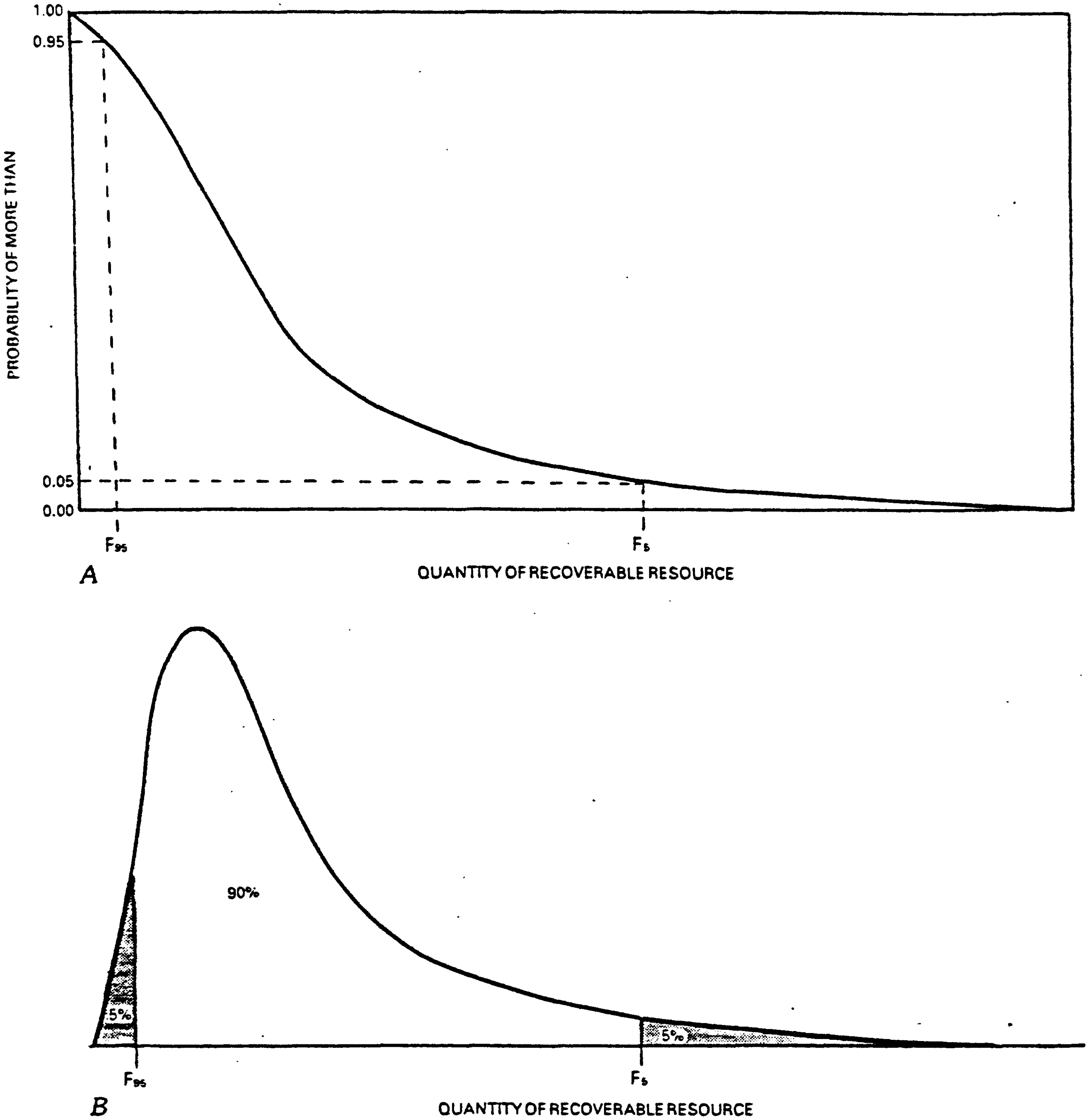

Figure 1.--Typical conditional probability distribution of an undiscovered recoverable resource shown as *A*, conditional *more-than* cumulative distribution function, and *B*, conditional probability density function. F_{95} denotes the 95th fractile; the probability of *more than* the amount is 95 percent. F_5 denotes the 5th fractile; the probability of *more than* the amount is 5 percent.

The Lognormal Distribution as a Resource Probability Model

The lognormal distribution is used by the USGS and many others as a probability model for resource estimation. Some favorable properties of the lognormal distribution as a resource probability model are the following:

A. Subjective properties

1. The lognormal distribution is a reasonable judgmental probability distribution from a subjective probability or decision analysis point of view. The lognormal distribution has the desired shape like many other distributions (e.g., triangular, beta, and gamma) that are used as probability models of uncertainty in risk analysis. These models are flexible and unimodal, and can be positively skewed. However, geologists particularly like the heavy-tailed characteristic of the lognormal for a resource assessment model. Selection of a judgmental probability distribution is relatively arbitrary.

B. Empirical properties

1. Arps and Roberts (1958) found empirically in the Denver-Julesburg basin that a lognormal curve provided an excellent fit to the relative frequency distribution of field size for 338 oil fields.

2. Bloomfield and others (unpublished) showed from data on all oil fields in Kansas that the distributions of the area and of the total oil production of these fields were found to be lognormal to good approximation.

3. Lee and Wang (in press) found empirically that the lognormal distribution is an excellent probability model for the relative frequency distribution of pool sizes in a play from the east coast of Canada; i.e., the pool-size distribution is approximately lognormal.

C. Theoretical properties

1. A well-known result from probability theory states that the lognormal distribution arises from the product of many independent random variables. This result is based upon a general version of the central limit theorem which holds under certain regularity conditions. That is, under fairly general conditions, the distribution of the product of many independent positive random variables is approximately lognormal.

2. A well-known multiplicative reproductive property is that the product of independent lognormal random variables is also lognormal. There are many other reproductive properties of the lognormal which make it a very mathematically tractable distribution--an important characteristic of a good model.

3. The independence condition can be dropped in the previous property when assuming a multivariate lognormal distribution which also is capable of easily incorporating correlations in the model; this is one advantage that the lognormal has when compared with almost any other choice of model other than the normal distribution.

4. Harris (1982) proved the equivalence of a negative exponential finding rate and the lognormality of the remaining resource base.

5. Kaufman (1963) has shown that under certain assumptions reported field size is lognormally distributed.

6. Barouch and Kaufman (1976) have demonstrated theoretically that the sum of independent and identically distributed lognormal random variables can be adequately modeled by a three parameter lognormal.

Probability Distribution

The marginal probability of a resource for an assessed area is applied to its corresponding conditional probability distribution to produce the judgmental probability distribution of the quantity of undiscovered recoverable resource. (This distribution is also referred to as the "unconditional" or "risked" distribution.) From the probability distribution, the final low (F_{95}), high (F_5), and mean (μ) estimates of the quantity of undiscovered recoverable resource are obtained for the assessed area. The mean estimate is the mean of the probability distribution. The curve of the more-than cumulative distribution function (cdf) has the value of the marginal probability at the origin. When the marginal probability of a resource is equal to 1, the probability distribution is the same as the conditional probability distribution.

Aggregate Probability Distribution

An aggregate probability distribution is the judgmental probability distribution of the total quantity of undiscovered recoverable resource in a region consisting of two or more assessed areas. Hence, the aggregate probability distribution is the convolution of two or more probability distributions (the distribution of the sum of two or more independent random variables). The mean value for a region equals the sum of the mean values for the areas making up the region. Similarly, the variance (square of the standard deviation) for a region equals the sum of the variances. The aggregate marginal probability is the subjective probability of the resource being present in commercial quantities within the region (two or more areas) considered. The aggregate marginal probability is calculated from the area marginal probabilities.

A Monte Carlo simulation technique is used to aggregate the probability distributions of a resource for more than one area. The result is the aggregate probability distribution of the resource for that region composed of more than one area. From this distribution, the low (F_{95}), high (F_5), and mean (μ) estimates of the quantity of undiscovered recoverable resource are obtained for the region.

The Monte Carlo computer program develops aggregations as follows: For each area in a region, a random value is selected from the probability distribution of a particular resource. These random values are added together to give a random value for the total resource in the region. This sampling procedure and summation is repeated a large number of times in order to generate an approximate probability distribution of total resource for the region. It has been found that a sampling number of 5,000 is quite adequate.

NUMERICAL EXAMPLE

A product of the World Energy Resources Program of the USGS will be used as an illustrative example. The report by Masters and Peterson (1981) is an assessment of the conventionally recoverable petroleum resources of northeastern Mexico. The region of interest consists of two areas: 1) Sabinas-Parras basins, and 2) the Burgos basin. The resources considered are oil and total gas for the individual areas and for the combined region of northeastern Mexico.

Estimation of Oil

A USGS resource appraisal team made the following assessments of undiscovered recoverable oil in the Sabinas-Parras basins:

1. The chance of oil being present in recoverable quantities was estimated to be 85 percent, i.e., the marginal probability of oil is 0.85.

2. Conditional upon recoverable oil being present, three conditional resource estimates were expressed in billion barrels (BB) as follows:
   a) F_{95} = 0.11, a low conditional resource estimate
   b) F_5 = 3.15, a high conditional resource estimate
   c) ML = 0.72, a modal conditional resource estimate

The conditional probability distribution of the quantity of undiscovered recoverable oil for the Sabinas-Parras basins is displayed in figure 2. The graph of the conditional more-than cdf has the value of 1 at the origin. Some numerical characteristics (in BB) of the lognormal distribution in figure 2 are the following:

$$F_{95} = 0.11;\ F_{75} = 0.30;\ F_{50} = 0.59;\ F_{25} = 1.17;\ F_5 = 3.15$$

$$\text{mean } \mu = 0.99 \text{ and standard deviation } \sigma = 1.34$$

The probability distribution of the undiscovered recoverable oil is also displayed in figure 2. The graph of the more-than cdf has the value of the marginal probability (0.85) at the origin. Some numerical characteristics (in BB) of this distribution are the following:

$$F_{95} = 0.00;\ F_{75} = 0.18;\ F_{50} = 0.47;\ F_{25} = 1.02;\ F_5 = 2.91$$

$$\text{mean } \mu = 0.84 \text{ and standard deviation } \sigma = 1.28$$

The mean estimate of 0.84 BB can be used as a point estimate of the quantity of undiscovered recoverable oil in the Sabinas-Parras basins. The range from low resource estimate (0.00) to high resource estimate (2.91) forms an interval estimate.

The USGS resource appraisal team made the following assessments of undiscovered recoverable oil in the Burgos basin:

1. The chance of oil being present in recoverable quantities was estimated to be 100 percent, i.e., the marginal probability of oil is 1.

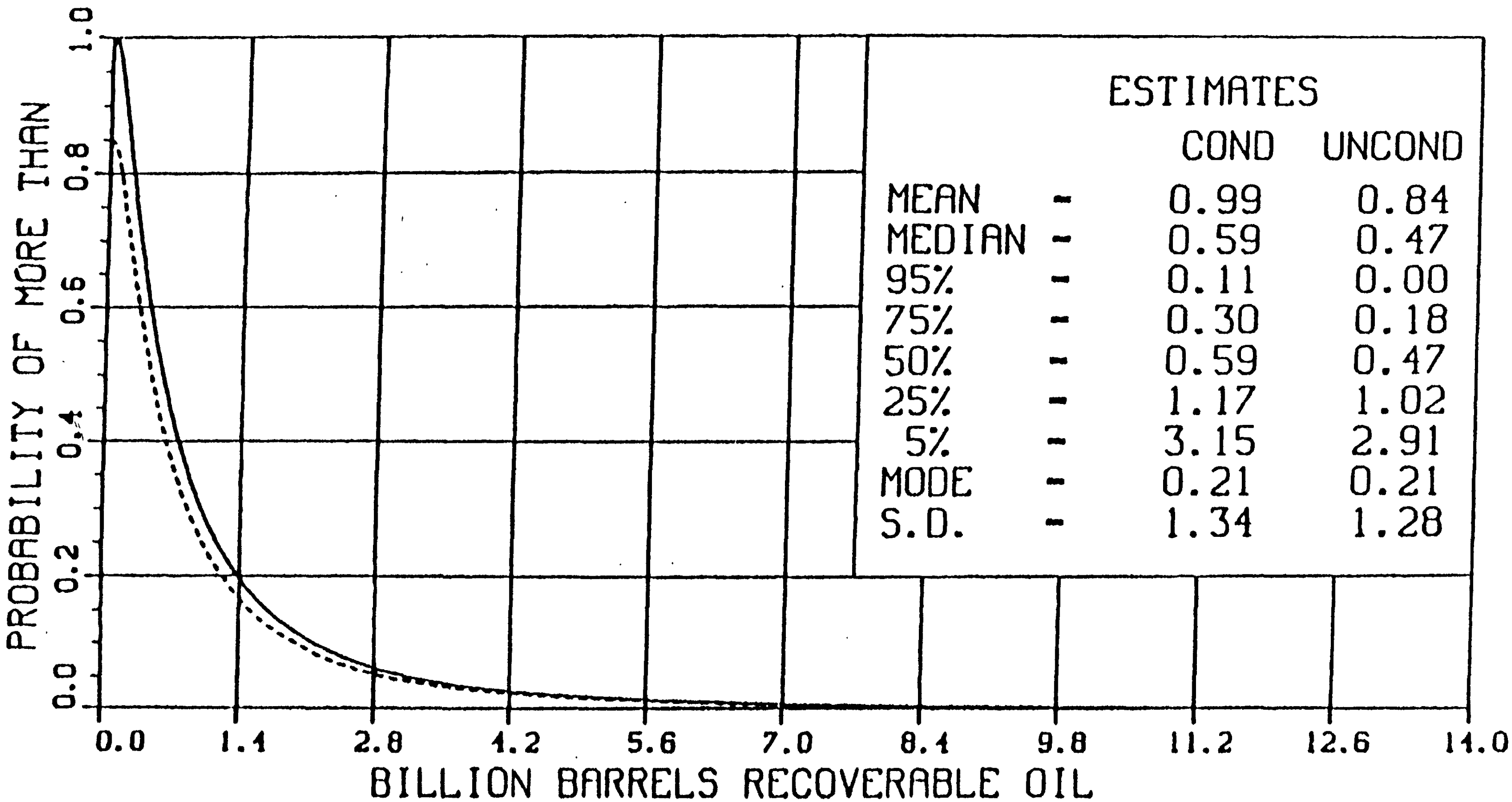

Figure 2.--Probability distribution of the undiscovered recoverable oil for the Sabinas-Parras basins assessment area of northeastern Mexico. Estimates are mean, median, mode, standard deviation (S.D.), and fractiles that correspond to the percentages listed; cond, conditional, solid curve; uncond, unconditional, dashed curve.

2. Conditional upon recoverable oil being present, three conditional resource estimates were expressed in BB as follows:

a) $F_{95} = 0.11$, a low conditional resource estimate

b) $F_5 = 1.18$, a high conditional resource estimate

c) ML = 0.30, a modal conditional resource estimate

The conditional probability distribution of the quantity of undiscovered recoverable oil for the Burgos basin is displayed in figure 3. Since the marginal probability is 1, the probability distribution is the same as the conditional probability distribution. Some numerical characteristics (in BB) of the lognormal distribution in figure 3 are the following:

$$F_{95} = 0.11;\ F_{75} = 0.22;\ F_{50} = 0.36;\ F_{25} = 0.59;\ F_5 = 1.18$$

mean $\mu = 0.47$ and standard deviation $\sigma = 0.39$

The quantity of undiscovered recoverable oil is estimated in the combined region of northeastern Mexico consisting of the two assessment areas: 1) Sabinas-Parras basins, and 2) the Burgos basin. A Monte Carlo aggregation computer program was used to generate an approximate probability distribution; the number of samplings was 5,000. The approximate probability distribution of the quantity of undiscovered recoverable oil for the northeastern Mexico region is displayed in figure 4. Since the aggregate marginal probability is equal to 1, the conditional and unconditional distributions are the same. Some numerical characteristics (in BB) of this distribution are the following:

$$F_{95} = 0.24;\ F_{75} = 0.55;\ F_{50} = 0.93;\ F_{25} = 1.55;\ F_5 = 3.51$$

mean $\mu = 1.31$ and standard deviation $\sigma = 1.34$

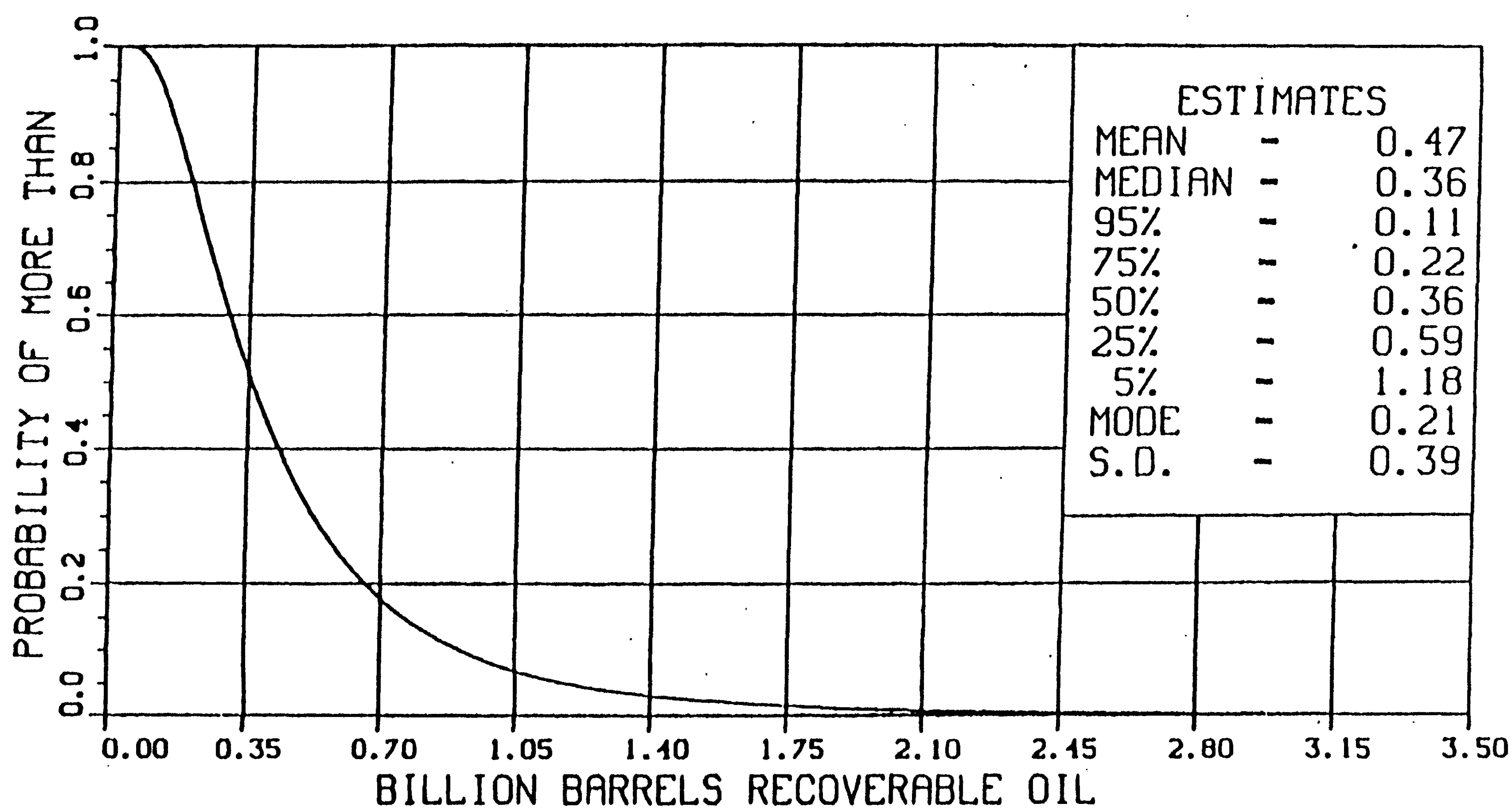

Figure 3.--Probability distribution of the undiscovered recoverable oil for the Burgos basin assessment area of northeastern Mexico. Estimates are mean, median, mode, standard deviation (S.D.), and fractiles that correspond to the percentages listed.

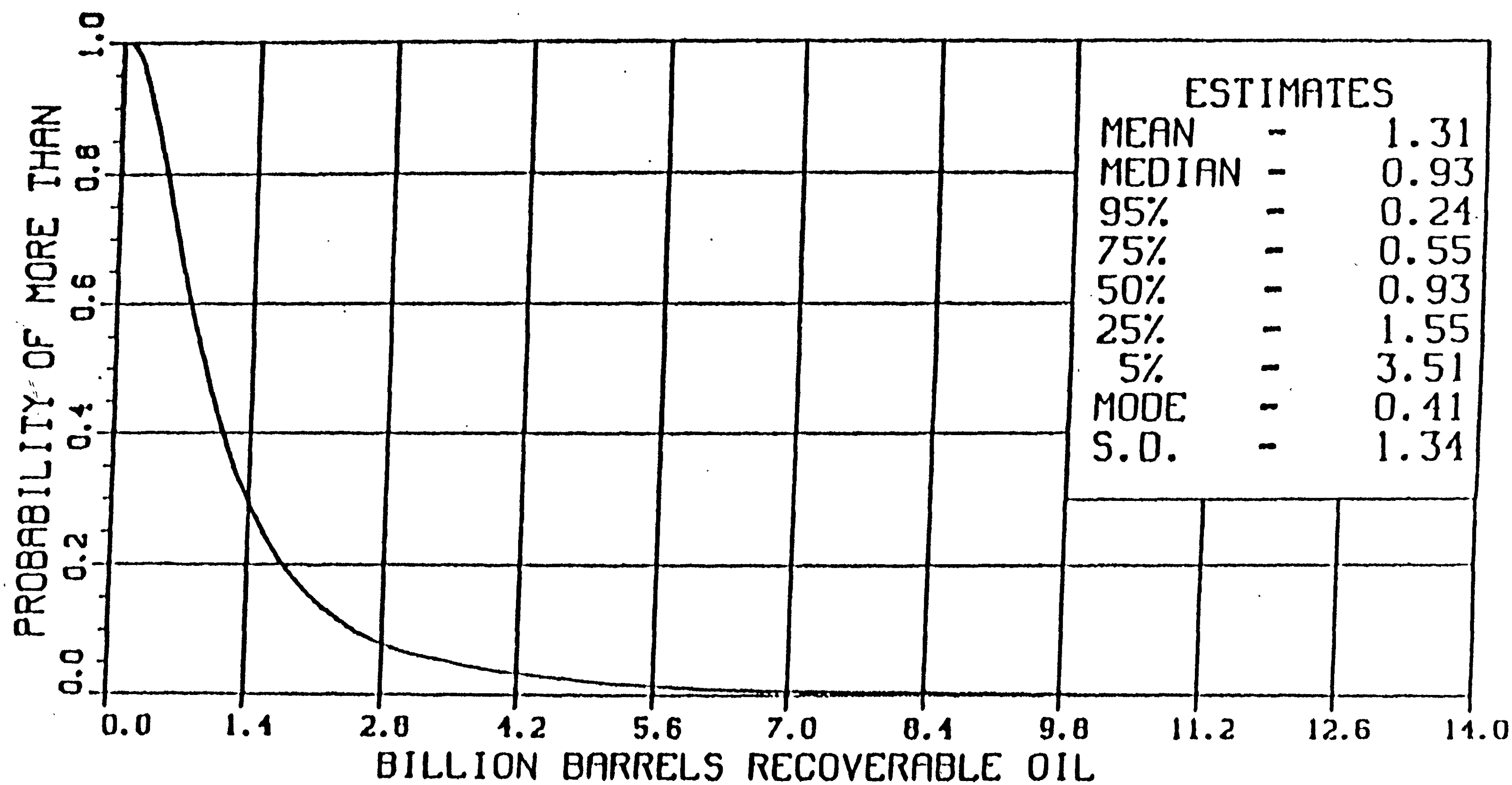

Figure 4.--Aggregate probability distribution of the undiscovered recoverable oil for the combined assessment areas: Sabinas-Parras basins and Burgos basin of northeastern Mexico. Estimates are mean, median, mode, standard deviation (S.D.), and fractiles that correspond to the percentages listed.

The mean estimate of 1.31 BB can be used as a point estimate of the quantity of undiscovered recoverable oil in the northeastern Mexico region. The range from low resource estimate (0.24) to high resource estimate (3.51) forms an interval estimate.

Estimation of Total Gas

The USGS resource appraisal team made the following assessments of undiscovered recoverable total gas in the Sabinas-Parras basins:

1. The chance of gas being present in recoverable quantities was estimated to be 100 percent, i.e., the marginal probability of gas is 1.

2. Conditional upon recoverable gas being present, three conditional resource estimates were expressed in trillion cubic feet (TCF) as follows:
   a) F_{95} = 17.64, a low conditional resource estimate
   b) F_5 = 121.82, a high conditional resource estimate
   c) ML = 48.73, a modal conditional resource estimate

The conditional probability distribution of the quantity of undiscovered recoverable total gas for the Sabinas-Parras basins is displayed in figure 5. Since the marginal probability is 1, the probability distribution is the same as the conditional probability distribution. Some numerical characteristics (in TCF) of the lognormal distribution in figure 5 are the following:

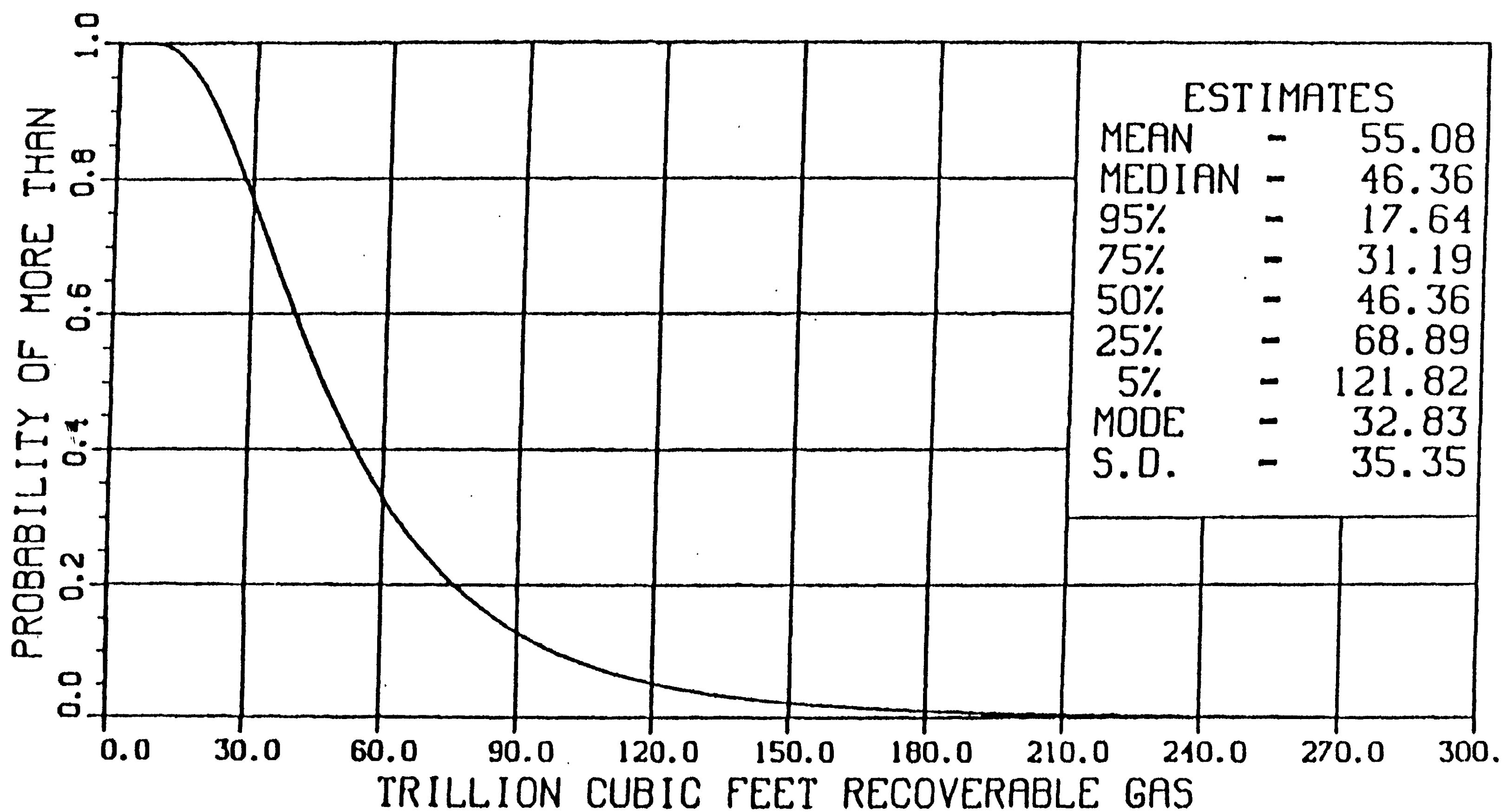

Figure 5.--Probability distribution of the undiscovered recoverable total gas for the Sabinas-Parras basins assessment area of northeastern Mexico. Estimates are mean, median, mode, standard deviation (S.D.), and fractiles that correspond to the percentages listed.

$$F_{95} = 17.64;\ F_{75} = 31.19;\ F_{50} = 46.36;\ F_{25} = 68.89;\ F_{5} = 121.82$$

$$\text{mean } \mu = 55.08 \text{ and standard deviation } \sigma = 35.35$$

The USGS resource appraisal team made the following assessments of undiscovered recoverable total gas in the Burgos basin:

1. The chance of gas being present in recoverable quantities was estimated to be 100 percent, i.e., the marginal probability of gas is 1.

2. Conditional upon recoverable gas being present, three conditional resource estimates were expressed in TCF as follows:

   a) $F_{95} = 1.64$, a low conditional resource estimate

   b) $F_{5} = 12.47$, a high conditional resource estimate

   c) ML = 4.57, a modal conditional resource estimate

The conditional probability distribution of the quantity of undiscovered recoverable total gas for the Burgos basin is displayed in figure 6. Since the marginal probability is 1, the probability distribution is the same as the conditional probability distribution. Some numerical characteristics (in TCF) of the lognormal distribution in figure 6 are the following:

$$F_{95} = 1.64;\ F_{75} = 2.98;\ F_{50} = 4.52;\ F_{25} = 6.85;\ F_{5} = 12.47$$

$$\text{mean } \mu = 5.47 \text{ and standard deviation } \sigma = 3.72$$

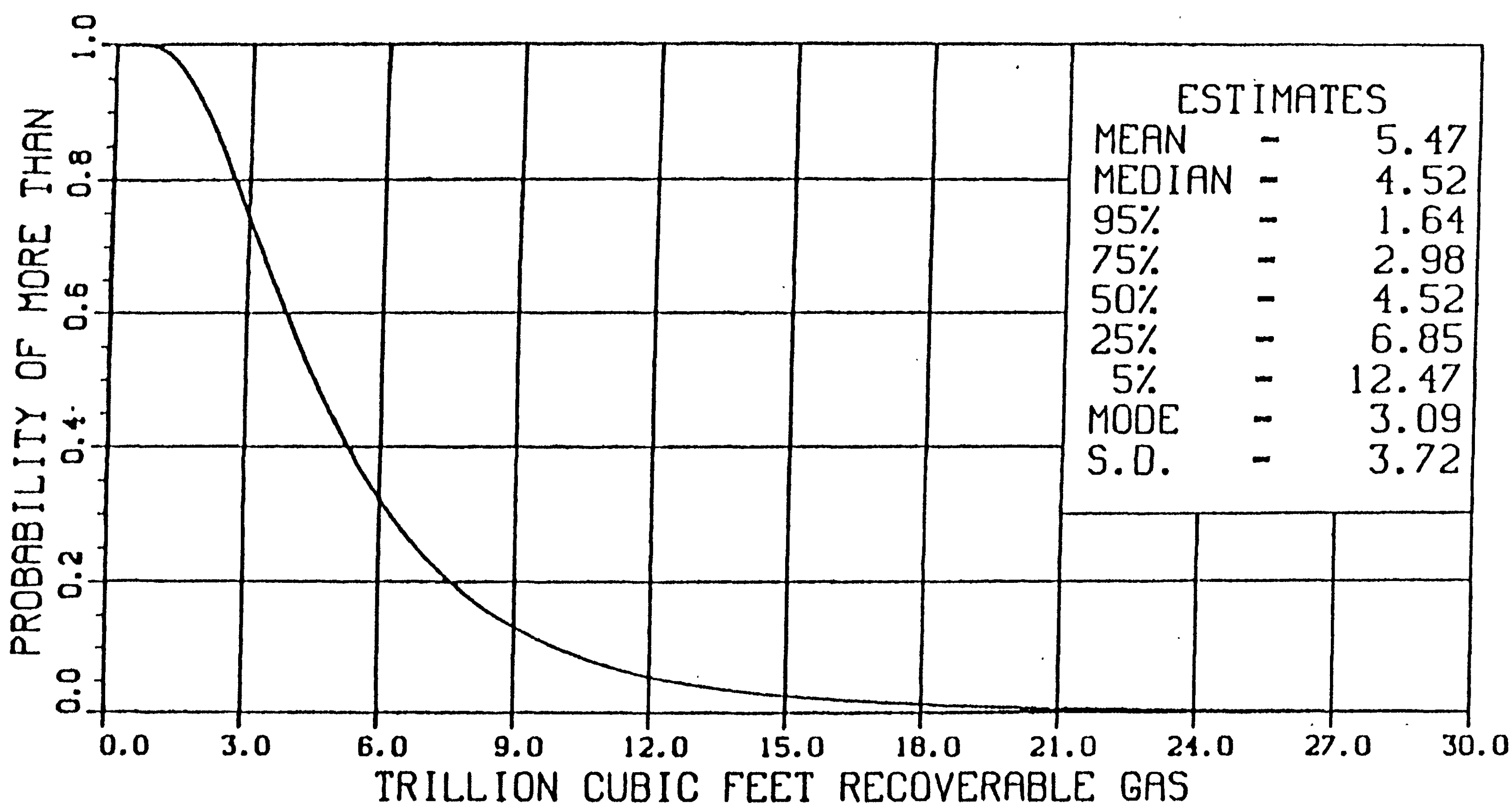

Figure 6.--Probability distribution of the undiscovered recoverable total gas for the Burgos basin assessment area of northeastern Mexico. Estimates are mean, median, mode, standard deviation (S.D.), and fractiles that correspond to the percentages listed.

The quantity of undiscovered recoverable total gas is estimated in the combined region of northeastern Mexico consisting of the two areas: 1) Sabinas-Parras basins, and 2) the Burgos basin. A Monte Carlo aggregation computer program was used to generate an approximate probability distribution; the number of samplings was 5,000. The approximate probability distribution of the quantity of undiscovered recoverable total gas for the northeastern Mexico region is displayed in figure 7. Since the aggregate marginal probability is equal to 1, the conditional and unconditional distributions are the same. Some numerical characteristics (in TCF) of this distribution are the following:

$$F_{95} = 22.77;\ F_{75} = 37.03;\ F_{50} = 52.20;\ F_{25} = 74.74;\ F_{5} = 126.37$$

$$\text{mean } \mu = 60.55 \text{ and standard deviation } \sigma = 35.55$$

The mean estimate of 60.55 TCF can be used as a point estimate of the quantity of undiscovered recoverable total gas in the northeastern Mexico region. The range from low resource estimate (22.77) to high resource estimate (126.37) forms an interval estimate.

SUMMARY OF METHODOLOGY

Individual appraisals are made for each of the assessment areas using petroleum geology, volumetric-yield procedures, exploration histories, and, in some areas, finding-rate studies and structural analyses as a basis for subjective assessments of oil and gas. A lognormal distribution is fitted using subjective low, high, and modal estimates to determine a conditional probability distribution for each area. By applying the marginal probability to the conditional probability distribution, a probability distribution of the quantity of undiscovered resource is established. To obtain total resource

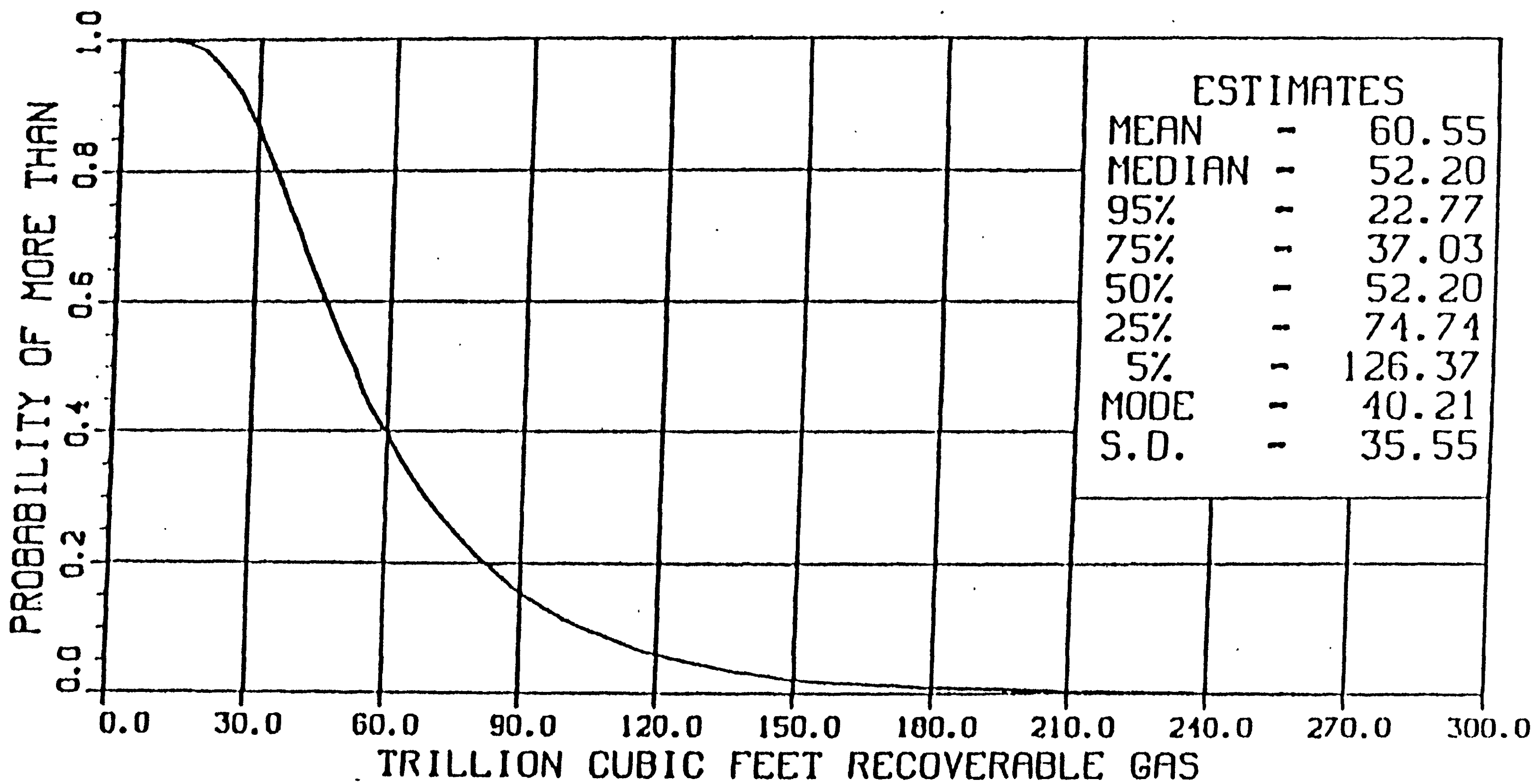

Figure 7.--Aggregate probability distribution of the undiscovered recoverable total gas for the combined assessment areas: Sabinas-Parras basins and Burgos basin of northeastern Mexico. Estimates are mean, median, mode, standard deviation (S.D.), and fractiles that correspond to the percentages listed.

estimates for a region of more than one area, the probability distributions for the individual areas composing the region are aggregated by a Monte Carlo technique. From an aggregate probability distribution, final estimates are obtained for the region.

REFERENCES

Arps, J. J., and Roberts, T. G., 1958, Economics of drilling for Cretaceous oil on east flank of Denver-Julesburg basin: American Association of Petroleum Geologists Bulletin, v. 42, no. 11, p. 2549-2566.

Barouch, E., and Kaufman, G., 1976, Probabilities modelling of oil and gas discovery, in Roberts, F. S., ed., Energy: Mathematics and Models: Philadelphia, Society for Industrial and Applied Mathematics, p. 133-150.

Crovelli, R. A., 1981, Probabilistic methodology for oil and gas resource appraisal: U.S. Geological Survey Open-File Report 81-1151, 77 p.

Dolton, G. L., Carlson, K. H., Charpentier, R. R., Coury, A. B., Crovelli, R. A., Frezon, S. E, Khan, A. S., Lister, J. H., McMullin, R. H., Pike, R. S., Powers, R. B., Scott, E. W., and Varnes, K. L., 1981, Estimates of undiscovered recoverable conventional resources of oil and gas in the United States: U.S. Geological Survey Circular 860, 87 p.

Harris, C. M., 1982, Log-normal resource bases and negative exponential finding rates: U.S. Department of Energy, unpublished memorandum to Data Analysis and Forecasting Branch, June 16, 1982, 5 p.

Kaufman, G. M., 1963, Statistical decision and related techniques in oil and gas exploration: Englewood Cliffs, New Jersey, Prentice-Hall, 307 p.

Lee, P. J., and Wang, P.C.C.. [in press], Probabilistic formulation of a method for the evaluation of petroleum resources: Journal of the International Association for Mathematical Geology (accepted for publication in 1982).

Masters, C. D., and Peterson, J. A., 1981, Assessment of conventionally recoverable petroleum resources of northeastern Mexico: U.S. Geological Survey Open-File Report 81-1143, 7 p.