A BAYESIAN PROCEDURE FOR RESOURCE EVALUATION OF PETROLEUM PROVINCES IN THE EARLY STAGES OF EXPLORATION.

Elena Medova

May 1981
WP-81- 66

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

PREFACE

Worldwide concern about the "oil crisis" has led to various endeavors to enable a better assessment of oil and gas resources. Professor Michel Grenon summarized results of such work carried out at IIASA in 1979 in his paper *World Oil Resources Assessment and Potential for the 21st Century* (IIASA WP-80-6). In the paper he demonstrated that there was no real foundation for the "consensus" for ultimately recoverable world oil resources of around 2000 billion barrels (most studies were non-independent) and that our understanding of world oil and gas resources is remarkably poor. In an attempt to improve this situation, work in the IIASA Resources Group has focussed on various methodologies for resource assessment.

The research described in this paper is an extension of work on the modeling of oil resources presented in *A Model for Resource Assessment and Exploration/Production Processes*. Medova, IIASA WP-80-44.

ACKNOWLEDGEMENTS

I would like to express my thanks to Professor Gordon Kaufman of MIT whose papers stimulated this work.

The research described in this paper and currently in progress has been undertaken with the collaboration of M.A.H. Dempster of System and Decision Sciences, IIASA. I would like to thank him for consultancy and for preparing the Appendix. His comments on an earlier draft of this paper have materially improved the exposition.

I would also like to thank Diana Barrow for preparing the paper for publication.

ABSTRACT

A major open problem in quantitative methods for petroleum resource evaluation concerns the provision of statistical techniques for geological provinces in the early stages of exploration—for example, if all exploratory wells drilled to date have been found to be dry or if only a few fields have been discovered. A considerable body of literature exists concerning statistical methods for mature provinces in which discovery volumes are on a general declining trend; all these methods use exploration history within the province to project total petroleum resources. In the early stages of exploration however, historical data is an insufficient basis for resource estimation and a simple method must be found to transfer relevant information from explored provinces with similar geology.

Following a survey of the existing quantitative methods for petroleum resource evaluation, this paper describes a new method for Bayesian updating of discovery probabilities corresponding to a rough field size classification in the light of current dry hole data. It is based on spatial Poisson dry hole sampling distributions fitted to geologically similar provinces. The method is applied to and the underlying assumptions statistically tested on some typical partly-explored provinces in Brazil. Finally, a Monte Carlo method for the resource assessment of immature provinces, based on revised discovery probabilities and building on earlier work (Medova, 1980) will be suggested for future development.

CONTENTS

A BAYESIAN PROCEDURE FOR

RESOURCE EVALUATION OF PETROLEUM PROVINCES

IN THE EARLY STAGES OF EXPLORATION.

Elena Medova

...it should be recognized that without a rapid inventory of its oil and gas resources the (U.S.) nation risks repeating the mistake of basing policy on illusion rather than on information.

M.W. Menard (1981)

Perhaps the most challenging strategic research problem underlying methods for forecasting petroleum supplies from new discoveries is to design a sequence of basically compatible models of deposition and discovery that span the information spectrum from frontier to mature.

G.M. Kaufman (1980)

INTRODUCTION

The dependency of today's world on oil and gas and the critical importance of accurate assessment of these resources for energy policy are all too familiar facts. In the U.S. for example--where oil currently provides nearly 45% of the nation's energy and oil and gas together supply 73%--Menard (1981) notes that "one of the reasons for uncertainty in national energy policy is that makers of policy lack the information needed to choose national objectives." He goes on to say that national petroleum policy should be based in particular on "knowledge" of both the total amount of recoverable oil and gas and potential field sizes.

Clearly such knowledge can at best be statistical in nature and Kaufman (1980) points out that "a forecast of the amount of petroleum remaining to be discovered in a large region is a forecast of an uncertain quantity several orders of magnitude more uncertain than a forecast of what is ultimately recoverable with current technology from discovered deposits."

Following Kaufman, we note that orderly incremental resolution of uncertainty regarding petroleum resource assessment is costly, and depends critically on current estimates of the resource base and alternative available technologies, and on prevailing politics and economics.

The nature of formal quantitative analysis of future petroleum supply from a particular geological region depends in turn on the quality and quantity of the data available from predrilling exploration activities and from drilling itself--i.e. on the *exploration history* of the region. Although it "is not...obvious...how the characters of the models used to generate supply forecasts should change, if at all, as more and more predrilling and drilling information accumulates," there is a growing awareness that a combination of Bayesian and economic analysis must consistently be employed (*cf.* Meisner, 1981). Such techniques are relatively well developed for ***mature regions*** where a large quantity of drilling data is available and detailed reservoir engineering studies for many fields may be performed. For ***immature*** (or ***frontier) regions***, however, it is impossible to base these techniques on sparse or inconsistent data from the region and new methods based on information from geologically similar regions--but compatible with the more developed methods for mature regions--must be provided. Such a method is the subject of this paper, and it is hoped that it can eventually be used as a basis for exploration policy analysis to influence supply in a way that the traditional approach to frontier regions through geometric-volumetric appraisal (see Section 2) can never do.

Within a single region it is usually the case that some stratigraphic units are intensely drilled and others are unexplored, so that any new method developed for a geological region should be able to take the situation at this lower level of geological aggregation into account.

In order for a model of discovery and supply from a mature region (or stratigraphic unit) to be logically compatible with a model for a region (or unit) in the early stages of exploration, a ***simple method*** must be proposed to transfer relevant information from explored regions to unexplored regions with similar geology. In this paper the transfer is based on dry hole sampling distributions.

Following a survey of the existing quantitative methods for petroleum resource evaluation (Section 2), this paper will describe the new method for Bayesian updating of discovery probabilities corresponding to a rough field size classification in the light of current dry hole data (Section 3). It is based on spatial Poisson dry hole sampling distributions fitted to geologically similar provinces. The method is applied to (Section 4) and the underlying assumptions statistically tested on (Section 5) some typical partly-explored provinces in Brazil. Finally, a Monte Carlo method for the resource assessment of immature provinces, based on revised discovery probabilities, will be suggested for future development (Section 6). The last section (Section 7) contains conclusions and directions for future research.

2. PRINCIPAL APPROACHES TO PROJECTING AMOUNTS OF UNDISCOVERED OIL AND GAS.

It is reasonable to begin a survey of the existing quantitative methods for petroleum resource evaluation by introducing definitions of applicable terminology (taken from literature) since there is some confusion between exploratory and geological terms.

GEOLOGICAL TERMINOLOGY (A.I. Levorsen, 1967)

Sedimentary basin. These basins include all the areas known to contain large volumes of sediments. They contain not only all the petroleum provinces discovered so far, but all those that are likely to be discovered in the future.

Sedimentary basins have the common characteristics of being geological depressed areas, with thick sediments in the interior and thinner sediments at the edges, but otherwise they may be quite different in origin and character.

Commercial petroleum deposits are classified as ***pools***, ***fields***, and ***provinces***. Terms such as "pool," "field," "province," and "subprovince" are useful in describing and locating the various oil and gas accumulations and occurrences. They combine both geographic and geologic factors that are commonly understood by the geologists, geophysicists, and engineers of the petroleum industry. But these terms, like many others in geology, grade into one another, which makes it difficult, at times, to divine their exact meaning. Local usage generally prevails eventually, even though it may not reflect the best or most accurate scientific classification and terminology.

Province. A petroleum province is a region in which a number of oil and gas pools and fields occur in a similar or related geologic environment. Since the term is loosely used to indicate the larger producing regions of the world, the boundaries of a so-called province are often indistinct. The Mid-Continent province of the south-central United States, for example, has definite regional characteristics of stratigraphy, structure, and oil and gas occurrence. Consequently, the term has a specific meaning for geologists and the petroleum industry. Subprovinces may occur within provinces; within the Mid-Continent province, for example, we find the Cherokee sand subprovince of southeastern Kansas and northeastern Texas, the Reef subprovince of west-central Texas, the Panhandle subprovince of northwestern Texas, and many others.

Field. When several pools are related to a single geologic feature, either structural or stratigraphic, the group of pools is termed a ***field***. The individual pools comprising a field may occur at various depths, one above another, or they may be distributed laterally throughout the geologic feature. Geologic features that are likely to form fields are salt plugs, anticlinally folded multiple sands, and complex combinations of faulting, folding and stratigraphic variables. The amount of oil that a pool or a field will produce is not a distinguishing characteristic. In the East Texas pool and in many of the Middle East pools, for example, the oil is obtained from a single reservoir; yet the ultimate production of each of these pools will be greater than that of many fields or even provinces. Since a field may contain several closely related pools, the terms "pool" and "field" are often confused, especially during the early development stages.

Pool. The simplest unit of commercial occurrence is the pool. It is defined as the body of oil or gas or both occurring in a separate reservoir and under a single pressure system. A pool may be small, underlying only a few acres, or it may extend over many square miles. Its content may be entirely gas, or it may be entirely or mainly oil. The size of an oil pool is generally given as the number of barrels of crude oil that may be produced and recovered at the surface of the ground. This is but a fraction of the crude oil in place underground, usually ranging from one-quarter to three-quarters of the total amount and depending on the current technology. The oil left behind is called ***nonrecoverable*** oil; the oil produced, the ***recoverable*** oil. The ***total***, ***original*** amount of oil in the pool underground is called the oil ***in- place***.

Anomoly. A deviation in the geologic structure or stratigraphy of a basin usually used in the sense of a *seismic* anomoly, an apparent structure observed from seismic records.

EXPLORATION TERMS (P.D. Newendorp, 1975)

Play. An area of concentrated exploration activity and/or interest within a sedimentary basin.

Project. An investment opportunity, a drilling prospect.

Prospect. An area under which is thought to exist a geological trap having oil or gas deposits. A seismic anomoly, for example. The area being considered to locate and drill an exploratory well.

A petroleum province is in a *mature* exploration stage when, after drilling a relatively large number of exploration wells, the discoveries are on a general declining trend. If the discoveries are on a general rising trend, the province is *immature*, and if the discoveries show a general constant trend, the area is in transition from immature to mature.

Unfortunately, each of the above terms is used differently by different authors and therefore special attention should be paid in the application of the various methods.

SURVEY OF EXISTING QUANTITATIVE METHODS FOR
PETROLEUM RESOURCE EVALUATION

Industry approaches to forecasting future discoveries were discussed in a report to the U.S. Energy Information Administration by ICF Incorporated (1979) and a comparison of private sector supply forecasting and decision-making methods appears in the Energy Modeling Forum (1979).

An excellent review of resource forecasting methods has been prepared by G. Kaufman (1980) in a forthcoming publication. The following is a condensed form summary of this work, together with some additional models published last year.

The principal approaches to projecting amounts of undiscovered oil and gas may be loosely classified as shown in Figure 1.

- life cycle
- rate of effort
- geologic-volumetric
- subjective probability
- exploration play or province discovery process.

LIFE CYCLE MODELS

This class of model is based on the assumption that there is a relatively simple functional relationship between time and the amount of oil and gas in place and that the proportions of them that are recoverable are parameters to be inferred from observation of what has been discovered and produced per unit time to date. Life cycles, like most statistical time series models, "divorce" themselves from the physics and engineering of discovery and geological description, and do not incorporate economic effects.

RATE OF EFFORT MODELS

Rate of effort models are similar to life cycle models; incremental additions to the total amount of hydrocarbons discovered, to production, or to reserves, are regarded as a function of cumulative exploratory effort to date. Exploratory effort is generally measured by the number of wildcat or exploratory wells drilled.

The hypothesis underlying Hubbert's analysis of discovery rates is that the average rate of discovery per foot of drilling declines monotonically with increasing cumulative footage drilled.

In Bromberg and Hartigan's study, data series for discoveries of additions to reserves from extension well drilling have been treated as statistical series, i.e. explained by models that explicitly characterize the nature of fluctuations about a trend by postulating a probability distribution for them. Their model projects an exponential decline in addition to reserves from extensions, from revisions, and from discoveries per unit of effort as cumulative effort E_t up to time t increases:

$$\frac{dR_t}{dE_t} = \beta e^{-\alpha E_t} \tag{2.1}$$

where

E_t is the cumulative effort to time t, R_t is the cumulative reserve found by time t, and α and β are fixed parameters. If this model (2.1) were to hold exactly reserves remaining to be found at time t are:

$$R_\infty - R_t = (\beta / \alpha)[e^{-\alpha E_t} - e^{-\alpha E_\infty}]. \tag{2.2}$$

Omitting an explanation of the method, the difference from earlier applications of exponential rate of effort models is exemplified in two ways: an explicit characterization of random fluctuations about a trend and the introduction of uncertainty about the parameters α , β in the form of a probability distribution for α and β capturing a priori uncertainty about these parameters. This model generates a probability distribution for projecting the uncertain quantity R_∞, given that at a time period t only the cumulative effort E_t and cumulative reserves R_t are known with (near) certainty.

GEOLOGIC VOLUMETRIC APPRAISAL

A geologic-volumetric appraisal of petroleum basins begins with an analysis of geological, geochemical, and geophysical data the aim of which is to determine:

(a) the yield in barrels per unit area or the volume of unexplored productive sediment in that basin, and

(b) the volume of productive sediment remaining to be explored.

In essence, this approach to forecasting undiscovered oil and gas is an "extrapolation of data on the abundance of mineral deposits from explored to unexplored ground on the basis of either the area or the volume of broadly favorable rocks" (McKelvey, 1972).

Geologic-volumetric methods are well illustrated by Mallory's method ANOGRE (Accelerated National Oil and Gas Evaluation):

Reasoning by geological analogy, it is assumed that the amount of hydrocarbon found in the volume of rock already drilled within a stratigraphic unit is functionally related to the amount of hydrocarbon in the volume of rock within that unit which has not yet been drilled.

Definitions: $V_{drilled}$ is the volume of rock tested by development wells in known pools plus the volume of rock drilled and found barren, $V_{potential}$ is the volume of rock that appears to be capable of producing but has not been drilled. HC_{known} is the volume of hydrocarbon discovered and $HC_{unknown}$ the computed volume of hydrocarbon yet to be found.

The basic functional relation between the amount of $HC_{unknown}$ to be discovered in a stratigraphic unit is of the form:

$$HC_{unknown} = f\,(HC_{known},\ V_{drilled},\ V_{potential}).$$

It is actually assumed that:

$$\frac{V_{drilled}}{HC_{known}} = \frac{V_{potential}}{HC_{unknown}} f \qquad (2.3)$$

The factor f is chosen *subjectively* after much consideration.

SUBJECTIVE PROBABILITY METHODS

USGS Circular 725, entitled "Geological Estimates of Undiscovered Recoverable Oil and Gas Resources in the United States" is the first U.S. government mineral resource appraisal expressed explicitly in subjective probability terms. The estimates of the undiscovered resources were made:

(1) by reviewing and analyzing all available geological and geophysical information compiled on more than 100 geological provinces,

(2) by applying resource appraisal techniques which include extrapolation of known producibility into untested sediments of similar geology for well developed areas and volumetric techniques using geologic analogs with ranges of yield factors,

(3) by using group appraisals (in a modified Delphi procedure) determined by geologic experts applying subjective probability procedures, and

(4) by reporting final results as probability ranges rather than as simple number values.

The EMRC report "Energy, Mines and Resources: A Canadian Resource Appraisal" gives the principal conclusions in the same form as the USGS Circular 725, i.e. right tail probabilities for the amount of resources remaining to be discovered. There are significant differences between the two studies. The concept of petroleum exploration play underlies the assessment procedures employed in the EMRC study. The exploration play model is composed of: oil and gas occurrence attributes which describe geologic conditions that must be obtained for hydrocarbons to be present in an anomaly, potential equation variables that jointly determine the quantity of hydrocarbons in a prospect conditional on some positive amount being present, and an equation describing how the amount of hydrocarbons in a prospect depends on prospect potential equation variables.

DISCOVERY PROCESS MODELS

A discovery process model is a model built from the assumption that directly describes both physical features of the deposition of individual pools and fields and the fashion in which they are discovered. Discovery process models are applied to a target population which consists of geologically similar deposits.

Arps- Roberts; Drew- Schuenemeyer- Root Models

The pioneering work of Arps and Roberts and the extension of this work by Drew, Schuenemeyer and Root portrays exploration as a process in which wildcats are "randomly" placed at coordinate points within the boundary of a well-defined play or basinal area. If the wildcat lies within the perimeter of the projective area of a field, a discovery is made, otherwise the wildcat is a dry hole. This model interrelates a ***specific field size*** of area A, the ***number W of wells drilled***, and the ***number N(A) of fields of size A***, and in this respect must be distinguished from rate of effort models that do not incorporate such features. The number of fields $\Delta N(A,W)$ found by the next increment ΔW of new field wildcats drilled is proportional to the total area $[N(A)-N(A,W)]A$ of fields of size A remaining to be discovered after N(A,W) fields of size A have been found:

$$\frac{\Delta N(A,W)}{\Delta W} \approx [N(A)-N(A,W]A. \qquad (2.4)$$

If the factor of proportionality rendering the left- and right-hand sides of (2.4) equal is assumed to be a constant c_0, and $N(A,W)$ is interpreted as a (deterministic) continuous function of W, then:

$$N(A,W) = N(A)\,[1-\exp(-c_0 AW)] \qquad (2.5)$$

The Arps and Roberts study and that of Drew, Schuenemeyer, and Root assume the existence of a *finite number* $A_1, \cdots, A_m$ *of target sizes (areas of fields)*, and that Nature has deposed N_i, $i = 1, 2, \cdots, m$, fields of areal extent A_i in a play or basin of area B_0. The N_i's and B_0 are fixed parameters, none of which are known with certainty. The methods they used to estimate uncertain parameters are strictly *marginal* –i.e. each size class is considered *separately* from all others.

The aim of the model is to predict the ultimate *production* in the basin, using the projection of the number of fields in each size category. The projections of increments of cumulative ultimate recoverable oil in each size category are a function of cumulative wildcat wells drilled.

LIKELIHOOD MODELS

Models of this type estimate the number and size distribution (in terms of oil in place equivalents) of remaining fields using a probabilistic model of the discovery process to interpret the actual discovery history in the area of a petroleum play. The discovery process is viewed as independent sequential sampling without replacement from an underlying (geophysical depositional) distribution of field sizes with the probability of discovery of a field by the exploration process at each trial related to field size or number of exploratory wells. Conditional on knowledge of the original size distribution of deposits and the parameters of the discovery process, such a model determines the probability of any possible future discovery sequence. Conversely, given the historical discovery sequence, it is possible to derive maximum likelihood point estimates (or Bayesian distributional estimates) of the parameters of the field size distribution and discovery process. These may then be used to derive or simulate estimates of the total resource distribution. The advantage of the Bayesian over the maximum likelihood approach to estimation of the field size and discovery process parameters is that with Bayesian methods (see Section 3) estimates may usually be easily updated in the light of new (or revised estimates of) discoveries.

In the pioneering work of Barouch and Kaufman (1975, 1978), the underlying field size distribution is taken to be lognormal and field discovery probabilities are taken to be directly proportional to their (random) sizes. The latter assumption was successfully tested against an alternative of equipable discovery of fields of all sizes in the first paper. In Barouch and Kaufman (1978), the basic model was used "to compute maximum likelihood estimates of the parameters of the lognormal depositional size distribution and the number of deposits in the play" for the purposes of simulating an actual discovery history. Conditional expectations of the model sequence of discovered field sizes were "computed using maximum likelihood estimates as point estimates of model parameters, i.e. using the imposed estimated lognormal size distribution." The fluctuations of actual field sizes in discovery order about these conditional expectations are large, but the aggregate amount of discoveries differ from the conditional expectations by only about 7%.

In the (Royal Dutch Shell) model of Meisner-Demirman the size distribution is taken to be lognormal but both its mean (which declines linearly) and discovery probabilities (which decline according to the linear logistic model) are assumed to decline with advancing exploration. Bayesian techniques are used to update normal-gamma and diffuse priors for, respectively, the size distributions and discovery probabilities to obtain posterior parameter distributions upon which predictive field size distributions--and hence future discovery distributions for a given exploration effort (using Monte Carlo techniques)--are based. Loglinear regression methods were involved in parameter estimations as in the work of Barouch and Kaufman, and the field size data were found consistent with the lognormal hypothesis.

In the studies of O'Carroll and Smith (1980) and Smith and Ward (1980), the question at issue is whether the imposition of specific functional forms (such as the lognormal distribution) on field sizes and discovery probabilities increases the precision of the estimates or, on the contrary, merely introduces *misspecification* error leading to *biased* forecasts. These authors work with discretized (multinomial) field size distributions and treat the exploration process as independent sampling without replacement from the field size (depositional) distribution with discovery probability proportional to an (estimated) power and field size. They impose distributional forms by computing the appropriate multinomial discrete approximation to the specified size distribution as functions of its unknown parameters, and compute posterior field size and discovery process parameter distributions from the data using diffuse priors. To make forecasts and compare posterior likelihoods of field size specifications, nonlinear optimization techniques are used to find maximum posterior likelihood estimates of the parameters which include the number of fields in each of the discrete field size categories. These studies found distributional constraints imposed by lognormal and Weibull distributions not "significant" in that the corresponding posterior likelihoods differed little from that of the *unconstrained* field size distribution. Further, they found that the discovery process, while not completely random, seemed best described by discovery probabilities proportional to the cube root of field size and that an attempt to incorporate dry hole data did not improve prediction. However, Smith and Ward (1980) produce simulation results to show that the maximum (posterior) likelihood procedures employed give "evenly pessimistic results when the discovery sample is of limited size" and are even more biased downwards when deposition (field size distribution) and discovery process parameters must be estimated simultaneously.

All the discovery process models described above are applicable to areas where discoveries are on a generally declining trend, in other words for petroleum provinces that are in a mature exploration stage. The models described mainly use data where historical records on discovery, reserves and production exist in some abundance:

Drew-Schuenemeyer-Root Denver-Julesberg

Barouch-Kaufman--Alberta North Sea

O'Carroll-Smith Northern North Sea

Smith-Ward Northern North Sea

Meisner-Demirman North Sea

In the early stages of exploration, however, historical data is not a sufficient basis for resource estimation and this is the problem that is addressed by the Bayesian discovery process model set out in the next section.

3. THE BAYESIAN EXPLORATION MODEL

This paper will attempt to discover what possibility there is of making a discovery in an unexplored or partly explored basin, if a certain amount of drilling is carried out. For example, is it still possible to find a giant (or commercial/subcommercial) oil field, if 70 wells have been drilled to date and all are dry? A Bayesian statistical procedure can be used to answer this question.

As mentioned in Section 1, simple methods must be found to incorporate in an exploration model of a frontier petroleum province data derived from exploration of geologically similar regions. In the simple model developed in this section, this transfer of *a priori* geological knowledge is based on two ingredients-- *prior* (subjective) *discovery probabilities* and corresponding *dry hole* statistics in the form of *sampling distributions* for the number of dry holes prior to discovery--for fields of various sizes.

The results of a simplified exploration process to discovery of a *single* field in a basin are represented by the possible field discovery events:

- discovery of a *giant* field S_g
- discovery of a *commercial* field S_c
- discovery of a *subcommercial* field S_s
- basin dry S_0

whose union is a universal set (sample space).

The corresponding set of subjective *discovery probabilities*, describing the *prior* view of results of exploration in a basin are:

$$P(S_g),\ P(S_c),\ P(S_s),\ P(S_o). \tag{3.1}$$

Dry hole statistics will be incorporated using Bayes Theorem to define the posterior probability of discovery of each field size given a specific number of dry holes prior to discovery.

Let N denote the event of a specific number of dry holes prior to discovery of a field of a particular size and define the conditional ***dry hole sampling probabilities:*** $P(N \mid S_g), P(N \mid S_c), P(N \mid S_s), P(N \mid S_o)$.

Bayes Theorem gives the ***posterior discovery probabilities*** of a field of each size (upon drilling the $(N+1)^{st}$ well) as:

$$P(S_g \mid N) = \frac{P(N \mid S_g)\, P(S_g)}{\sum_{i=g,c,s,o} P(N \mid S_i) P(S_i)}$$

$$P(S_c \mid N) = \frac{P(N \mid S_c)\, P(S_c)}{\sum_{i=g,c,s,o} P(N \mid S_i)\, P(S_i)}$$

$$P(S_s \mid N) = \frac{P(N \mid S_s)\, P(S_s)}{\sum_{i=g,c,s,o} P(N \mid S_i)\, P(S_i)} \tag{3.2}$$

$$P(S_o \mid N) = \frac{P(N \mid S_o)\, P(S_o)}{\sum_{i=g,c,s,o} P(N \mid S_i)\, P(S_i)}$$

Given an actual number N = n of dry wells drilled in a basin, the corresponding values of these posterior probabilities (3.2) are the relative discovery probabilities of field sizes if the next well is a discovery. (For this purpose the event of a *dry* basin may formally be taken to be discovery of a field of *zero* size.) The probabilities (3.2) may be used to answer the question raised above, which is of considerable importance for many countries with partly explored petroleum provinces such as Brazil.

Alternatively, together with the *total dry whole sampling* probabilities given by

$$\sum_{i\,=\,g,c,s,o} P(N\,|\,S_i)\,P(S_i) \qquad N = 1,2,\ldots, \tag{3.3}$$

they may be used in a Monte Carlo model for total in place resource estimation for an unexplored basin. Such a model simulates the discovery processes for *individual* fields in terms of a number of dry holes to the discovery well until the dry basin event is observed (see Section 6).

The values of the prior discovery probabilities (3.1) may be assigned using a worldwide analysis of the geology of sedimentary basins, i.e. the classification scheme of Klemme (1975). It remains to find a method for specifying the dry hole sampling probabilities conditional on the field size categories from an explored basin geologically similar to the unexplored basin of interest. One possibility is to use the *empirical* dry hole sampling distributions from the explored basin for the various size categories. As the number of fields (particularly giants) in a fully explored basin is usually relatively small however, this data may be more parsimoniously used to estimate parameters of specific functional form for the dry hole sampling distributions of the explored basin. If, at the same time these parametric distributions may be *rescaled* to apply to the frontier basin, our full set of aims will have been achieved.

For this purpose, let us suppose that the empirical dry hole sampling distributions of the explored basin may be approximated by the *spatial Poisson distribution*, whose (discrete) density is given by

$$P(n \mid S_i) = e^{-\lambda_i B} \, (\lambda_i B)^n / n! \qquad n = 0,1,2,\ldots \tag{3.4}$$

where n denotes the number of dry holes prior to discovery, B is the (explored) basin area, λ_i is the basin *spatial dry hole rate* (say per KM^2) and i denotes the field size category g, c, s or o as before. The corresponding distribution is given by

$$F_N(N \mid S_i) = \sum_{n=0}^{N} e^{-\lambda_i B} (\lambda_i B)^n / n! \qquad N = 0,1,2,\ldots \tag{3.5}$$

If, for a *particular* field size category, $n_1, n_2, \ldots, n_m$ and $A_1, A_2, \ldots, A_m$ are, respectively, the *observed* numbers of dry holes prior to discovery and the areas of the m fields in the explored basin in discovery order, the maximum likelihood estimate of the spatial rate is given by

$$\lambda = \frac{1}{m}\left[\frac{n_1}{B} + \frac{n_2}{B-A_1} + \frac{n_3}{B-A_1-A_2} + \ldots + \frac{n_m}{B-A_1-\ldots A_{m-1}}\right] \approx \frac{\sum_{i=1}^{m} n_i}{mB} \tag{3.6}$$

The mean of the corresponding spatial Poisson distribution is given by

$$\lambda B \approx \frac{1}{m} \sum_{i=1}^{m} n_i \tag{3.7}$$

If it is assumed that the discovery processes for basins of similar geologic type have common characteristics, then the *unexplored* basin of interest will have the *same* spatial dry hole rate (in each field size category) and by multiplying this rate by its area B' the corresponding spatial Poisson dry hole sampling distribution will have mean given by

$$\lambda B' \approx \frac{\sum_{i=1}^{m} n_i}{m} \quad B'/B \tag{3.8}$$

Thus a simple method has been found to transfer the relevant information from explored to unexplored (or partly explored) provinces of similar geological type. The next two sections of this paper discuss the application of this model—and the statistical testing of its underlying assumptions—on Brazilian data.

4. PETROLEUM PROVINCES IN BRAZIL

As noted above, Brazilian sedimentary basins were chosen for the purpose of developing resource assessment modeling for provinces in the early stages of exploration.

Brazil is determined to find out whether or not it can reach or approach petroleum self-sufficiency. Petrobras, the Brazilian oil agency, says that its current "strategic exploratory program" began in 1978. This program requires 505 wildcat wells to be drilled in the first 4 years--325 of these wildcats are to be offshore. Figure 2 shows the sedimentary basins of petroleum resource interest in Brazil.

Petroleum proved reserves in 1979 were claimed to be 1.373 billion barrels of oil, but some of the Brazilian sedimentary basins are not very widely explored. Recently Brazil has greatly stepped up exploratory drilling efforts, accelerated production plans, and permitted foreign companies to attempt to find and produce oil offshore. In spite of all these efforts the question as to how long Brazil can continue to invest billions of dollars into offshore drilling that has as yet yielded only extremely modest results was discussed in "World Oil" (March, 1980).

The Bayesian statistical procedure described in the previous section will be developed and tested on Brazilian data (source: Petroconsultants) for two marginal continental basins. The marginal continental basins in Brazil are presented in Figure 3.

Modeling future discoveries using the Bayesian procedure is applicable to the Sergipe-Alagoas basin (Figure 4), which is a partly-explored basin of some 12,000 KM^2 currently in transition from the immature to the mature stage. The 2,000 KM^2 area of the basin which lies offshore has so far been little explored. The data required for analysis, which are summarized in Table 1, provide an insufficient basis for making assumptions about the discovery process.

At present the Reconcavo basin is the best studied of all the Brazilian marginal basins (Figure 5). Petroleum exploration in the 10,000 KM^2 area of the Reconcavo basin began in 1937. Data characterizing Reconcavo's exploration and production are summarized in Table 2.

Geological analysis has identified the Reconcavo and Sergipe Alagoas basins as basins with similar geology *(cf.* Asmus and Ponte, 1969). Using Klemme's classification (1975) based on world statistics, both of them are ***pull- apart basins*** (Type V of Klemme's classification scheme). For this type Klemme gives the probability of discovery of a giant field $P(S_g) = 0.2$; and the probability of discovery of a commercial field as $P(S_c) = 0.3$. The probability of the sample space is 1, so the residual 0.5 is the probability of a subcommercial field and/or no discovery, i.e. a dry basin.

The definition of a subcommercial field is of course a question of economic analysis (price of oil, cost of installation, etc.). Let us assume that the probability of discovery of a subcommercial field is $P(S_s) = 0.4$, therefore the probability of a dry basin is $P(S_o) = 0.1$

This completes the set of subjective discovery probabilities for the sample space, describing the prior view of the results of exploration.

The field size classification, chosen for modeling is taken to be:

Subcommercial field : Between 0 and 20,000,000 barrels of oil

Commercial field : Between 20,000,000 and 200,000,000 barrels of oil

Giant field : > 200,000,000 barrels of oil.

Although this classification is arbitrary, it corresponds roughly to current general usage (*cf.* R. Nehring, 1978; J. Smith and L. Ward, 1980).

We are assuming that the discovery processes for our two geologically similar basins have similar dry hole sampling statistics, so next we must construct the appropriate distributions. A map of the Reconcavo basin with exploration legend is shown in Figure 6. Figures 7 and 8 illustrate typical basin data, with exploration legends from which dry exploration wells may be counted, used to construct Tables 1 and 2. The figures refer respectively to the Candeias and Guaricema fields in the basins.

Table 3 summarizes the basic dry hole statistics for both test basins. (Figures in brackets for Sergipe-Alagoas give Reconcavo values rescaled for the relative areas of the two basins).

For each size category in both basins, the empirical (cumulative) distribution of the number of dry holes to discovery was constructed, the spatial Poisson dry hole rates estimated according to (the approximate formula) of (3.6), and the corresponding (spatial) Poisson distribution calculated. The results for both empirical and theoretical distributions are plotted in Figures 9 and 10.

5. TESTING MODEL ASSUMPTIONS ON BRAZILIAN PROVINCES.

The assumptions underlying the model and statistical analysis of the previous two sections together with possible evidence and statistical test for them, are presented in Figure 11. These assumptions were made in order to provide a simple transfer of dry hole sampling distributions from an explored basin to a partly explored basin of similar geological type. For the investigated case of Sergipe-Alagoas basin, they should permit the use of data on the Reconcavo basin for predictive purposes. The two principal assumptions, labelled A1 and A2, concern the suitability of the spatial Poisson distribution and the hypothesis that basins of the same geological type have identical dry hole rates per unit area for the discovery of fields in a particular size category. Both these assumptions have been carefully statistically tested on the Reconcavo and Sergipe-Alagoas basin data. A description of the statistical tests utilized is presented in the Appendix. The results of the statistical tests of assumptions A1 and A2 are summarized in Tables 4 and 5 respectively. Inspection of Table 4 shows that the data are consistent with spatial Poisson dry hole sampling distributions.

Assumption A2 permits the transfer of the distributions of dry holes prior to field discovery from an explored to a partly explored basin. Table 5 reveals that this transfer is statistically acceptable for the two test basins in two of the three field size categories. Figure 12 depicts graphically the suitability of the transferred theoretical (spatial Poisson) dry hole sampling distributions for the various field size categories. The transferred theoretical distributions for commercial and subcommercial fields are a marginal improvement to the fit of the spatial Poisson distributions directly estimated from the empirical data for the Sergipe-Alagoas basin (see Figure 10). However, the fit for giant fields is poor. This could be due to the small sample size (2) as is suggested by the acceptance of the null hypothesis A2 by the nonparametric (Smirnov) test. The truth of this hypothesis would be consistent with the future discovery of a giant offshore field in the Sergipe-Alagoas basin following a relatively large number of exploratory dry wells. Notice from Table 5 that the data also support the hypothesis that dry hole sampling distributions in the two basins are *identical* --possible since the areas of the two test basins differ only by a factor of 1.2. This finding could also be explained by the hypothesis that the areal correction to Reconcavo statistics utilized in the transferred spatial Poisson distributions for Sergipe-Alagoas are partially inappropriate due to an overall improvement in exploration efficiency since the earlier exploration of the Reconcavo basin (*cf.* Table 3). Future research should involve testing assumption A2 on geologically similar basins of widely different areas.

Assumption A3 is consistent with the hypothesis that the geophysical processes responsible for the deposition of petroleum fields *correlate* the spatial locations of fields of the various size classes (e.g. commercial and subcommercial fields may lie relatively near giants) but that within *local* areas exploration processes for individual fields are completely random (as specified by the spatial Poisson distribution, A2) and independent of each other (A0.1). Specifically A3 states that exploration processes for fields in *different* size categories are statistically *dependent.*

If this hypothesis were false--i.e. dry hole sampling distributions for different size categories are statistically *independent* --but the underlying distributions are actually spatial Poisson (A1), then the underlying total dry hole sampling distribution would be spatial Poisson with dry hole rate given by a weighted mixture of the individual size category dry hole rates according to the corresponding discovery probabilities. Figure 13 shows the multimodal nature of the empirical total dry hole sampling histograms for the Reconcavo and Sergipe; Alagoas basins (refer Figure 11, A0.2). The corresponding empirical total distributions were tested for their spatial Poisson character by means of the dispersion test (see Appendix). The tests for both basins *rejected* the spatial Poisson null hypothesis at above the 0.1% level of significance. Thus the data are consistent with A3 and dependent exploration processes for fields of different size categories.

In order to answer the question of Section 3 by giving posterior relative discovery probabilities for fields of various sizes in the Sergipe-Alagoas basin after a specific number of wells have been drilled, it remains to specify the dry hole sampling distribution corresponding to a dry basin. The hypothesis of no discovery in a basin may only be proven when the entire area of the basin has been drilled and all holes are dry. In Zapp's study (1962) the well density to test all potentially productive onshore and offshore U.S. regions was defined as being equal to one well for each two square miles. If this density is applied to the Sergipe-Alagoas basin with an area of 12,000 sq. km, then the approximately 2300 wells have to be drilled for testing a no-discovery hypothesis. The distribution of dry holes for the no-discovery event may therefore be represented by a spatial Poisson distribution with mean equal to 2300.

This constructed distribution of dry holes prior to discovery of an exhausted basin permits the calculation of the posterior conditional discovery probability for each size category field given by the number of exploration dry holes. Figure 14 shows the results of the calculation which allows answers to the questions posed in section 3.

6. THE MONTE CARLO SIMULATION METHOD FOR RESOURCE ASSESSMENT OF IMMATURE PROVINCES.

The Monte Carlo simulation concept allows the analysis of options regarding uncertainty in future discoveries by providing the forecasting results in the form of distributions of possible resource values. The general logic of operating a simulation model is simply to define the distribution of undiscovered resources by a series of repetitive runs.

A disadvantage of resource assessment simulation models for a mature petroleum play based on the likelihood methods discussed in Section 2--one of whose underlying assumptions is sampling without replacement from a *finite* population--is that estimates of the number of fields in the play must be *a priori*. The alternative purely statistical approach, involving estimation of the number of fields in each size category from the data using maximum likelihood methods, tends to underestimation and currently appears plagued with numerical stability difficulties (see Section 2).

The simulation model currently under development for an unexplored (or partly explored) petroleum basin arrives at a probability distribution for the number of fields in the basin only *implicitly*. It simulates the actual exploration process in terms of the number of dry holes to discovery of each field, decides the size category of the field by means of the Bayesian posterior relative discovery probabilities as computed in the previous section, and terminates the simulation run only upon drawing the dry basin event.

While research is continuing on the process of designing and programming the model the following points are worth mentioning:-

(1) The original deposition of reservoirs/fields is assumed to follow an arbitrary size distribution. Since a variety of studies support the assertion that the size distribution of oil fields is adequately represented by a lognormal distribution, this hypothesis will be made in a first version of the model for statistical calibration tests on known basins

(2) The calibration tests will be performed on the Reconcavo and Sergipe-Alagoas basins studies in this paper.

(3) The model will then be used to provide resource assessments for Brazilian petroleum provinces in the early stages of exploration.

(4) Further tests of the assumptions and model should be made for petroleum basins of other geological types.

It should be noted that before simulation, a geological analysis of the chosen basins must be made.

The determination of the basin type requires detailed analysis of a number of geological parameters. This analysis should be made in close contact with geologists.

7. CONCLUSIONS

The method of resource evaluation proposed in this work connects subjective opinion about the probability of finding particular petroleum resources with information accumulated from exploratory drilling. The main advantage of this method is its applicability to resource assessment in partly explored or unexplored areas. The results of the analysis show that:-

1. Spatial Poisson distributions provide a good fit to empirical dry hole distributions in the (rough) field size classification chosen for selected Brazilian data.
2. Spatial Poisson dry hole rates for two pull-apart basins in Brazil are not statistically different in the field size classifications chosen.
3. A large number of dry holes must be drilled before the Bayesian posterior probability of an empty basin is significant.
4. A Monte Carlo simulation model for the resource assessment of petroleum basins in the early stages of exploration based on the Bayesian updating of prior discovery probabilities in the light of dry hole information has been proposed and is currently under development.

LIFE CYCLE MODELS:

Hubbert (1962, 1966)
Moore, (1966)
Ryan, (1965, 1966)

RATE OF EFFORT MODELS:

Hubbert (1974)
Hartigan-Bromberg (1967)

GEOLOGIC-VOLUMETRIC APPRAISAL:

Zapp (1962)
Hendricks (1965)
Mallory (1975)
Jones (1975)

SUBJECTIVE PROBABILITY METHODS:

USGS Circular 725 (1975)
Energy, Mines, Resources Canada EP77-1 (1977)
Canada EP77-1 (1977)

DISCOVERY PROCESS MODELS:

Arps-Roberts (1958)
Drew-Schunemeyer-Root (1978)
Kaufman (1980)
Barouch-Kaufman (1975, 1978)
O'Carroll-Smith (1980)
Smith, (1980)
Meisner-Demirman (1981)

FIGURE 1. PRINCIPAL APPROACHES TO PROJECTING AMOUNTS OF UNDISCOVERED OIL AND GAS

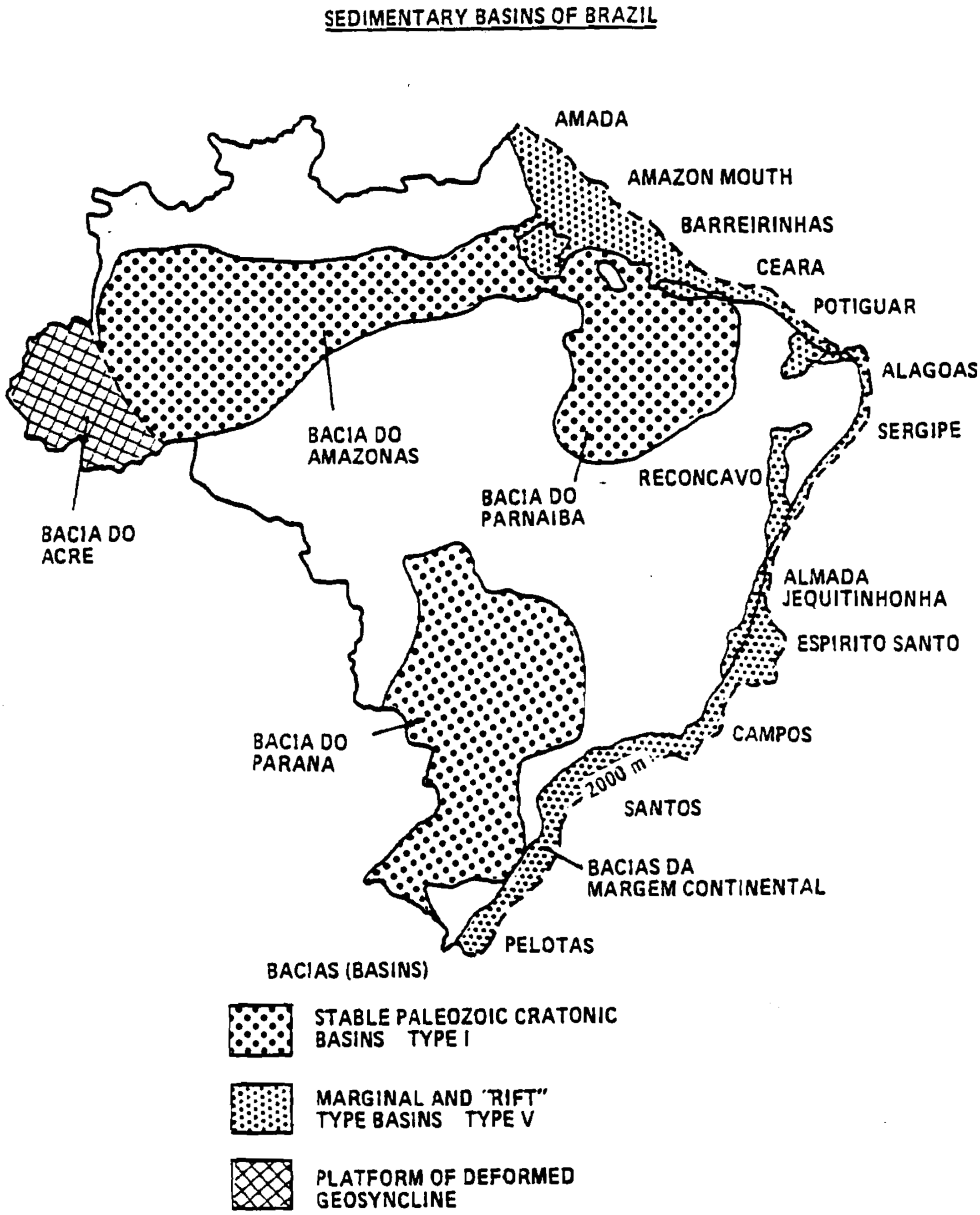
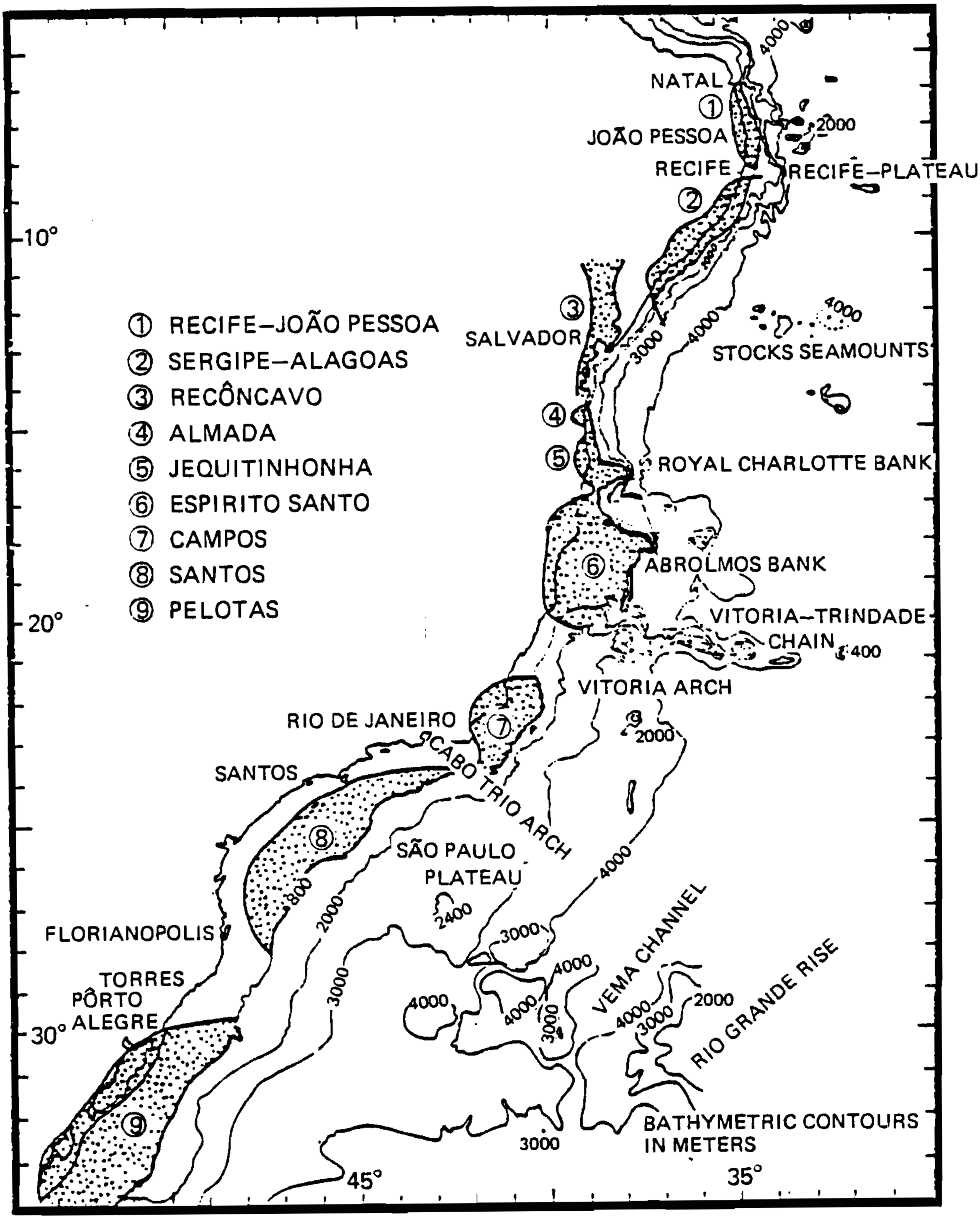


FIGURE 2. SEDIMENTARY BASINS OF BRAZIL.



FROM: R. LEYDEN et al. (1976)

FIGURE 3. EASTERN BRAZIL MARGINAL COASTAL BASINS

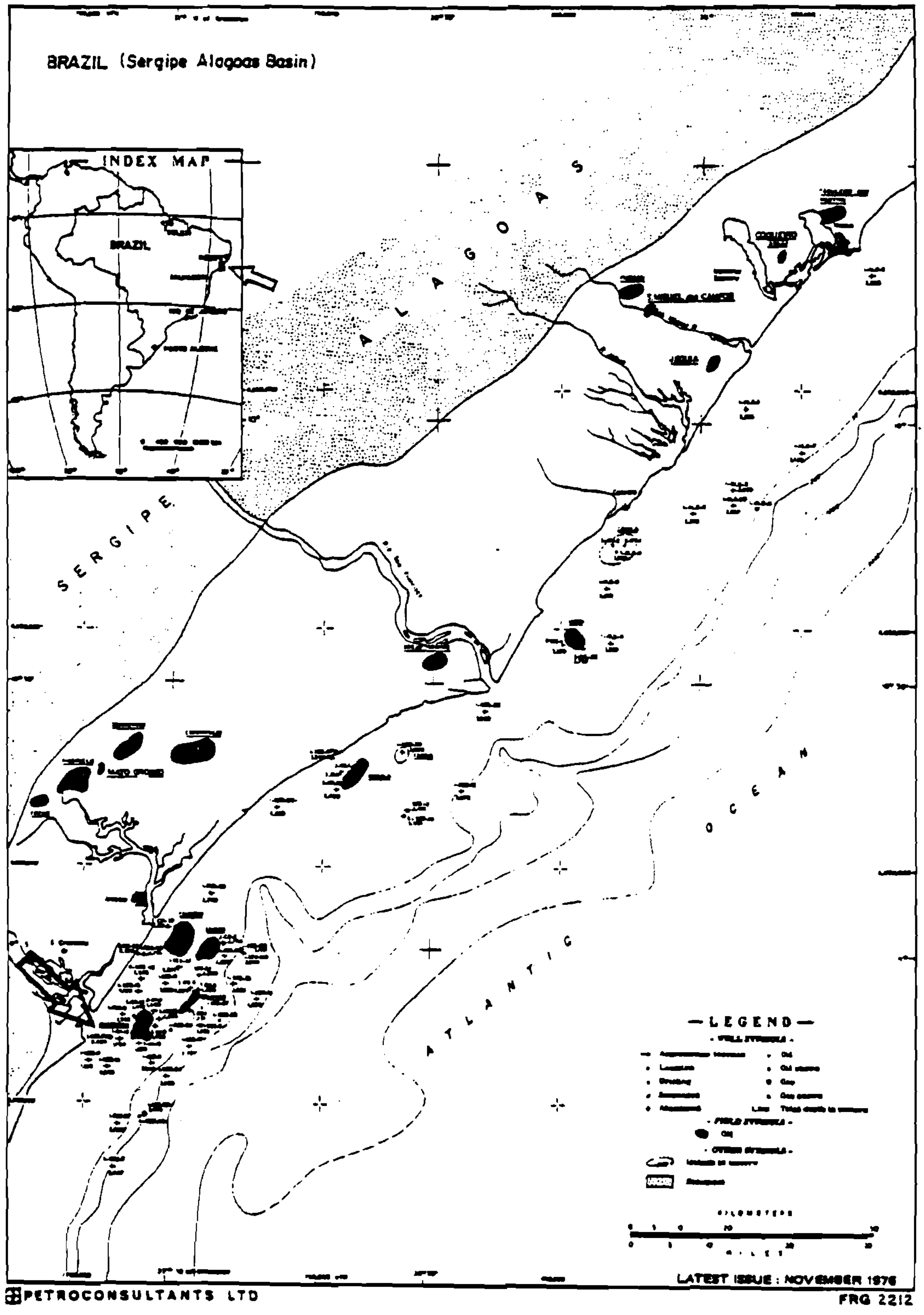


FIGURE 4. SERGIPE ALAGOAS BASIN

TABLE 1. SERGIPE-ALAGOAS BASIN

	NAME OF FIELD	DISCOVERY WELLS	AREA/HA	DRY WELLS	RESERVES/BBLS	CUMULATIVE PRODUCTION/BBLS	ONSHORE/ OFFSHORE
1	Jequia	JA-1-AL Oct. 1957	25	8		56,093/1977	Onshore
2	Riachuelo	1-RO-1-SE Nov. 1961	2300	8	151,940,000	5,347,270/1975	Onshore
3	Tabuleiro dos Martins	TM-2-AL 1962	170	12	22,500,000	1,559,466/1975	Onshore
4	Carmopolis	1-CP-1-SE Aug. 1963	2950	8	1,224,000,000	64,586,720/1975	Onshore
5	Coquero Seco	CS-1-AL Sep. 1963	22	6	5,000,000	283,321/1975	Onshore
6	Treme	TR-1-SE Dec. 1965	25	2		Abandoned	Onshore
7	Aquilhada	Ag-1-SE Sep. 1966	8	4		Abandoned	Onshore
8	Siririzinho	1-SZ-1-SE Aug. 1967	1800	14	210,400,000	14,185,418/1975	Onshore
9	Guaricema	1-SES-1A Nov. 1968	750	7	72,000,000	16,450,000/1977	Offshore
10	Sao Miguel dos Campos	1 CSMC 1 AL May 1969	22	6		6,692/1975	Onshore
11	Ponto dos Mangues	1 PDM 1 SE Jun. 1969	20	4		47,200/1978	Offshore
12	Furado	1 Fu 1 AL Aug. 1969	146	6		4,226,435/1976	Onshore
13	Caíoba	1-SES-6 Jan. 1970	1500	3	20,000,000	9,492,000/1976	Offshore
14	Brejo Grande	1-BRG-1-SE Feb. 1970	800	2		629,974/1977	Onshore
15	Dourado	1-SES-5 Jun. 1970	600	5		570,012/1976 (annual)	Offshore
16	Camorim	1-SES-10 Nov. 1970	2500	5	30,000,000		Offshore
17	Robalo	1-SES-23 May 1973	1400	3	55,350,000		Offshore
18	Mero	1-ALS-10 Aug. 1974	900	6		149,237/1976 (annual)	Offshore
19	Tainha	1-SES-39 Jan. 1975	25	2		Developing	Offshore
20	Cavala	1-ALS-11 Dec. 1975	250	1		Undeveloped	Offshore
21	4 SES 44	4 SES 44 May 1977		0		Shut in	Offshore

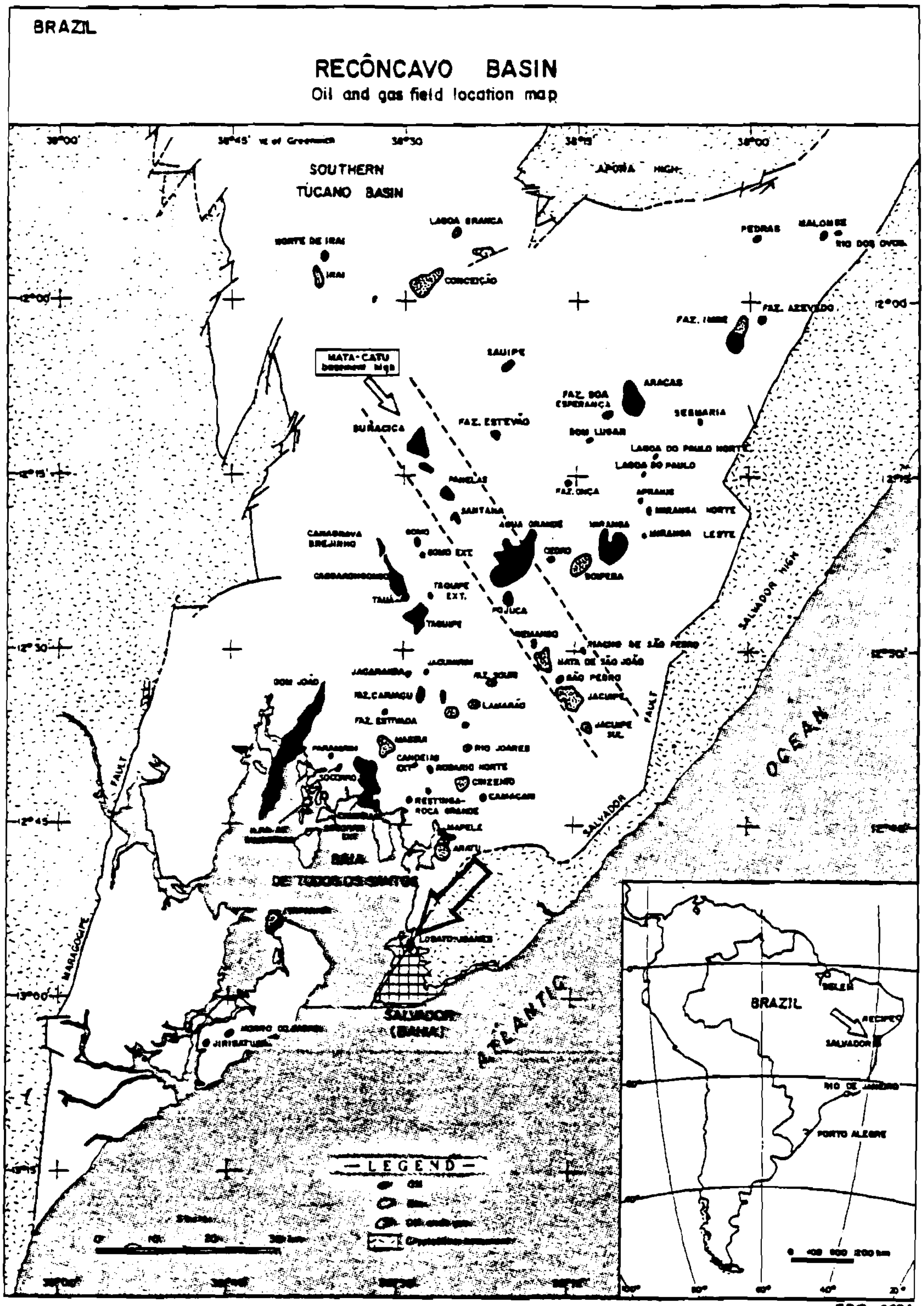


FIGURE 5. RECONCAVO BASIN

TABLE 2. RECONCAVO BASIN

	NAME OF FIELD	DISCOVERY WELLS	AREA/HA	DRY WELLS	RESERVES/BBLS	(BBLS) CUMULATIVE PRODUCTION	ONSHORE/ OFFSHORE
1	Lobato-Joanes	L-3 1939	60	10	Depleted	91,296/1975	Onshore
2	Candeias	C-1-BA 1941	505 218 } 1368 645	19	400,000,000	66,831,000/1975	Onshore
3	Aratu	A-1-BA 1942	600	4		57,200/1960	Onshore
4	Itaparica	1-2-BA 1942	160	9	26,700,000		Off + onshore
5	Dom Joao	Dj-1-BA 1947	400	25	831,000,000	63,894,241/1975	Off + onshore
6	Agua Grande	AG-1-BA 1951	5000	11	558,000,000	230,672,000	Onshore
7	Pedras	RS-5-BA 1951	57	10		431,000/1968	Onshore
8	Paramirim	PV-1 1951	16	6	1,800,000	426,000/1975	Onshore
9	Mata de Sao Joao	Mj-1-BA 1951	1000	11	63,200,000	5,362,500/1970	Onshore
10	Pojuca	PC-1-BA 1953	600	5	63,800,000	321,000/1964	Onshore
11	Jacuipe*	JA-1-BA 1956	1800	4	No reserve or prod. data		Onshore
12	Sao Pedro	SP-1-BA 1957	60	2	15,200,000	45,100	Onshore
13	Socorro	EP-7-BA 1958	16	2	15,300,000	14,970 annual	Onshore
14	Cassarongongo	Cs-1-BA 1959	180	6	71,200,000		Onshore
15	Taua*	Ta-1-BA			No data		Onshore
16	Buracica	BA-2-BA 1959	1475	20	480,000,000	90,228,700/1975	Onshore
17	Taquipe	T-1 1959	800	20	214,500,000	60,611,252	Onshore
18	Brejinho	Bj-1-BA 1960	200	2	6,000,000	2,263,000/1975	Onshore

*Not included in sample

contd.

RECONCAVO BASIN

	NAME OF FIELD	DISCOVERY WELLS	AREA/HA	DRY WELLS	RESERVES/BBLS	(BBLS) CUMULATIVE PRODUCTION	ONSHORE/ OFFSHORE
19	Jacaranda	jo-1-BA 1960	12	2		1,617/1962	Onshore
20	Mapele*	ME-1-BA 1960	110	10	Gas field/no adequate data		Onshore
21	Canabrava		∿100	4	6,000,000		Onshore
22	Fazenda Caruacu	FC-1-BA 1961	50	2		1,258/1976	Onshore
23	Ilha de Bimbarras*	IB-1-BA 1961		1	Gas field/no adequate data		Onshore
24	Gomo	GO-1-BA 1961	40	3		591,400/1976	Onshore
25	Fazenda Estivada	FE-1-BA 1962	30	0			Onshore
26	Morro do Barro	MB-1-BA 1962	25	5	503		Onshore
27	Fazenda Panelas	FP-1-BA 1962	600	1	60,000,000		Onshore
28	Santana	SA-1-BA 1962	50	11	35,000,000		Onshore
29	Fazenda Azevedo	FA-1-BA 1962	729	3	1,320,000		Onshore
30	Massui	MUI-1-BA 1967	160	5		51,524	Onshore
31	Massape	Mp-3-BA 1967	40	6		1,283,318/1976	Onshore
32	Fazenda Imbe	FI-1-BA 1967	160	7	43,500,000	8,838,598/1976	Onshore
33	Jiribatuba	Ji-1-BA 1964	18	6		117,900/1976	Onshore
34	Biriba	Bb-2-BA 1964	600	3		13,397/1977	Onshore
35	Aracas	1-AR-2-BA 1965	700	20	366,000,000	6,967,842	Onshore

*Not included in sample

cont.

RECONCAVO BASIN

	NAME OF FIELD	DISCOVERY WELLS	AREA/HA	DRY WELLS	RESERVES/BBLS	(BBLS) CUMULATIVE PRODUCTION	ONSHORE/ OFFSHORE
36	Miranga	Jul 1965	2500	10	590,000,000	119,824,268	Onshore
37	Malombe	MI-1-BA 1966	287	5	36,733,600	3,841,861/1977	Onshore
38	Fazenda Onca	FO-1-BA 1966	16	3		235,101/1973	Onshore
39	Sesmaria	Si-2-BA 1966	15	4	1,006,400	227,237	Onshore
40	Lagoa do Paulo	LP-1-BA 1966	15	2		545,859/1977	Onshore
41	Fazenda boa Esperanca	FBE-1-BA 1966	813	4	59,126,000	9,236,008/1977	Onshore
42	Cinzento*	CZ-2-BA 1966	800	1	No data		Onshore
43	Camacari	CA-2-BA 1966	16	4		87,707	Onshore
44	Lamarao*	Lm-1-BA 1967			No data		Onshore
45	Fazenda Santo Estevao	FSE-1-BA 1967	25	5	9,000,000	57,340	Onshore
46	Lagoa do Paulo Norte*	LPN-1-BA 1967	15	2			Onshore
47	Norte de Rosario	1-NRR-1-BA 1968	15	0		11,410	Onshore
48	Bom Lugar	1-BL-1-BA 1968	15	4		120,591/1977	Onshore
49	Sauipe	Se-1-BA 1970	30	4	5,000,000	148,000	Onshore
50	Miranga Norte	1-MGN-1-BA 1971	18	5	21,000,000	3,813,615/1977	Onshore
51	Remanso	1-RO-1-BA 1971	18	10		2,776,613/1976+	Onshore
52	Apraius	1 Apr 1 BA 1973	15	2		349,874/1976	Onshore
53	Riacho de Sao Pedro*	1-RSP-1-BA 1973	15	2	No reserve or prod. data		Onshore
54	Rio dos Ovos	1-ROV-1-BA 1974	15	3		130,776/1976	Onshore

*Not included in sample

+Possibly misclassified as subcommercial since production is large.

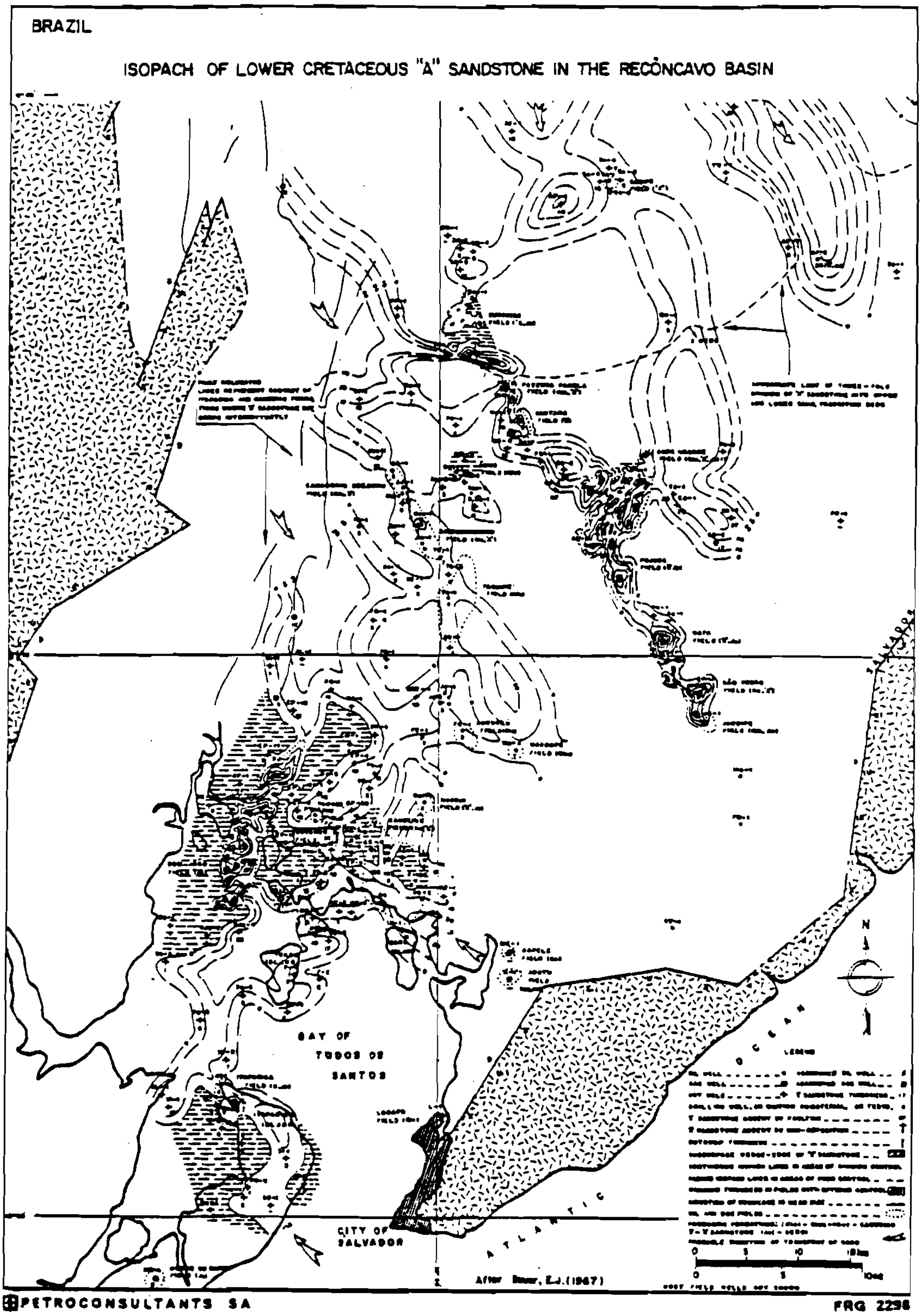
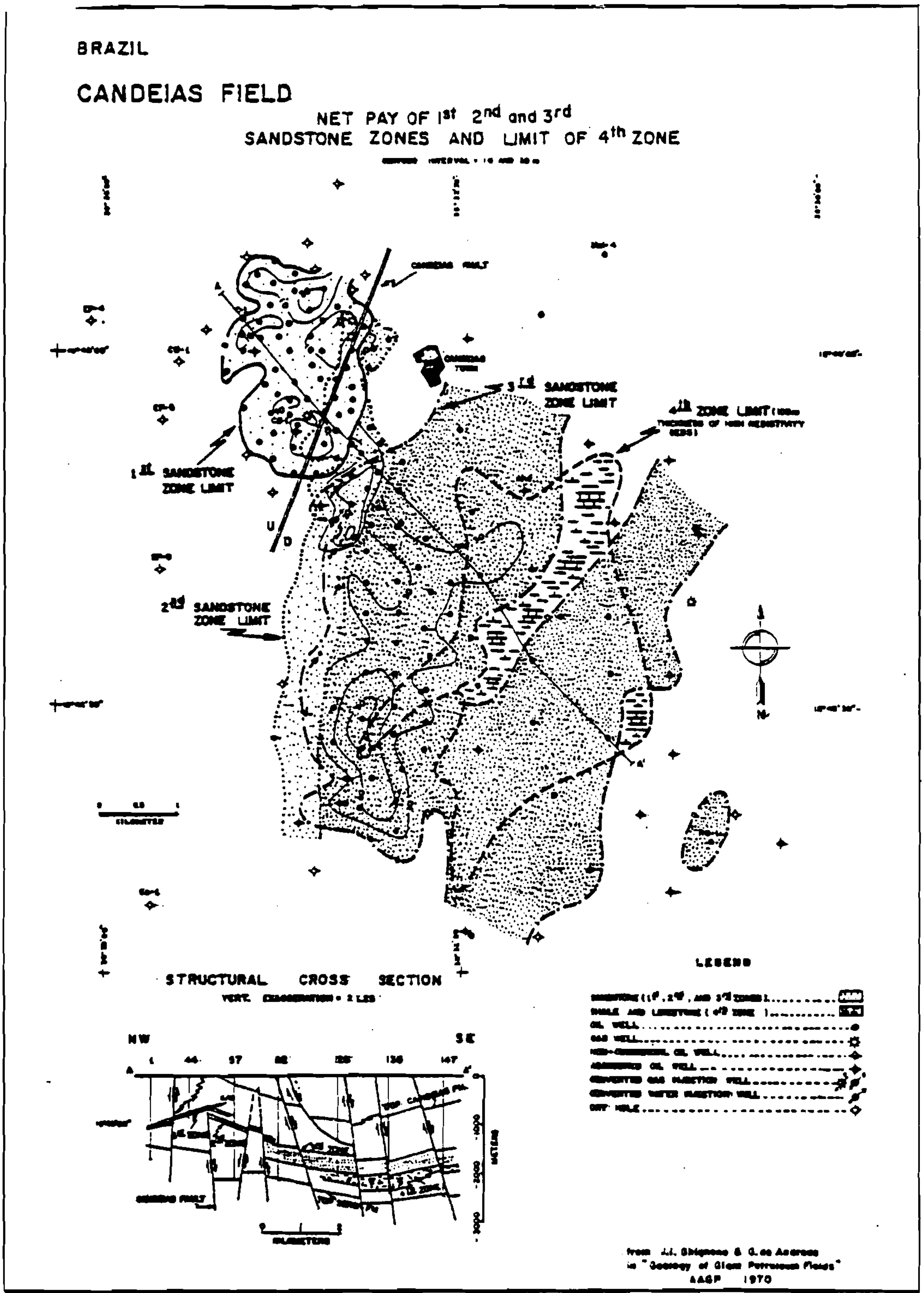


FIGURE 6. ISOPACH OF LOWER CRETACEOUS "A" SANDSTONE IN THE RECONCAVO BASIN.



PETROCONSULTANTS SA

FRG 2051

FIGURE 7. RECONCAVO BASIN: CANDEIAS FIELD

BRAZIL
GUARICEMA FIELD

STRUCTURAL MAP OF TOP OF GUARICEMA SANDSTONE

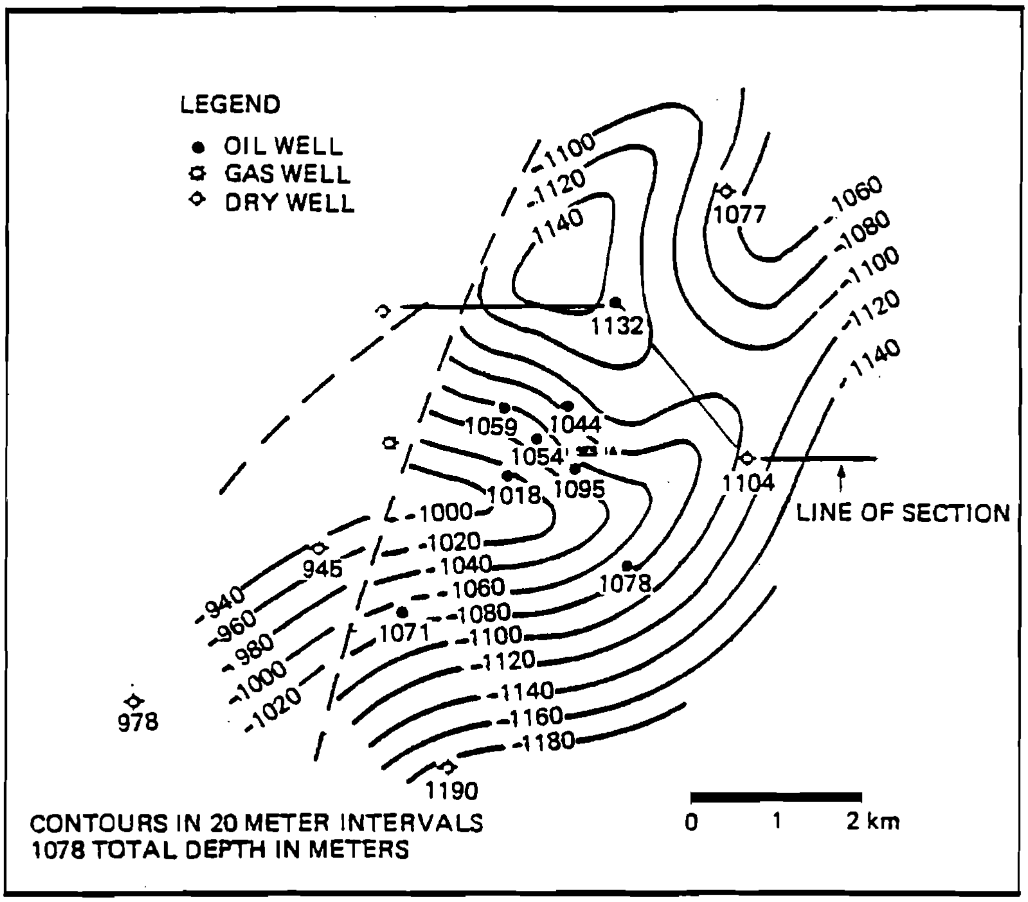


FIGURE 8. SERGIPE-ALAGOAS BASIN: GUARICEMA FIELD

TABLE 3: SUMMARY OF TEST BASIN DRY HOLE STATISTICS FOR ROUGH FIELD SIZE CLASSIFICATION CHOSEN

FIELD SIZE CATEGORIES:

GIANT	200 M	Bbls +
COMMERCIAL	20-200 M	Bbls
SUBCOMMERCIAL	0-20 M	Bbls

TEST BASIN 1: RECONCAVO
AREA: 10,000 KM^2

TEST BASIN 2: SERGIPE-ALAGOAS
AREA: 12,000 KM^2
(2,000 KM^2 OFFSHORE)

	1	2	Predict
GIANT FIELDS			
NUMBER OF FIELDS	7	2	
AV.NO. OF DRY HOLES	17.9	11.0	(21.4)
STANDARD DEVIATION	5.4	4.2	(6.5)
COMMERCIAL FIELDS			
NUMBER OF FIELDS	10	6	
AV.NO. OF DRY HOLES	6.4	6.3	(7.7)
STANDARD DEVIATION	3.2	3.4	(3.8)
SUBCOMMERCIAL FIELDS			
NUMBER OF FIELDS	29	13	
AV.NO. OF DRY HOLES	3.9	4.0	(4.7)
STANDARD DEVIATION	2.4	2.4	(2.9)

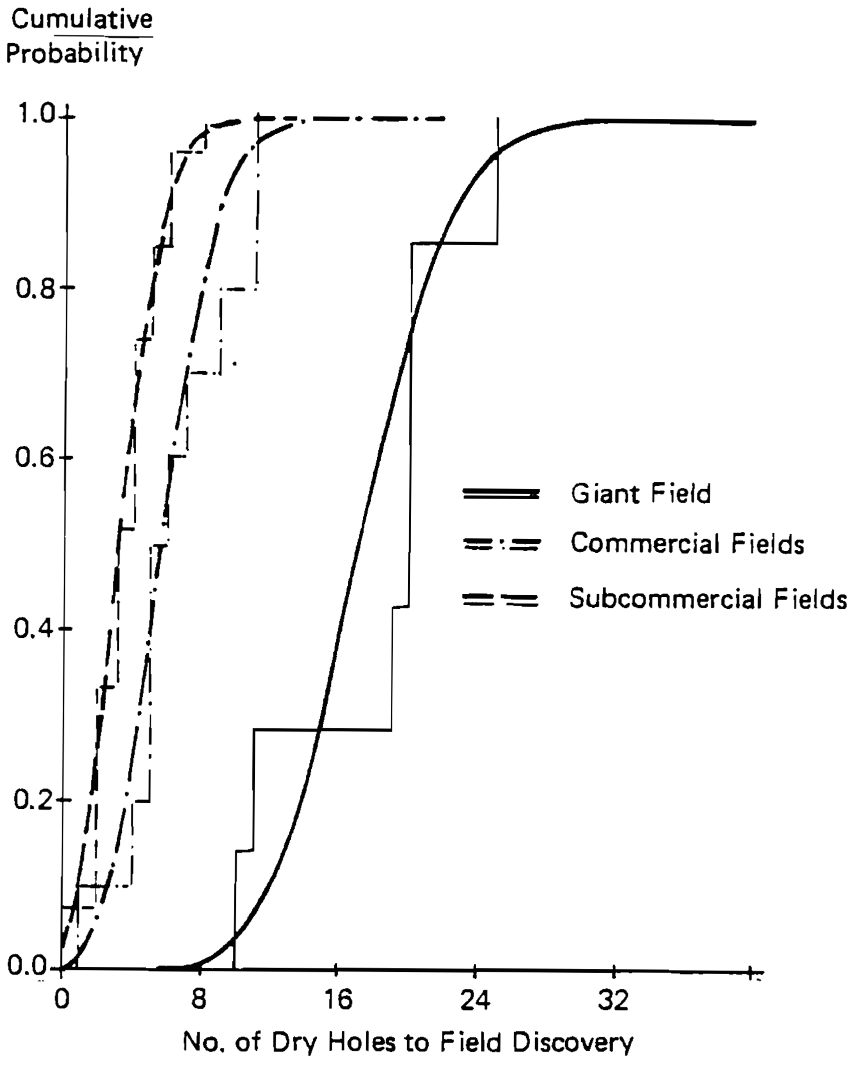


FIGURE 9. RECONCAVO BASIN DRY HOLE SAMPLING DISTRIBUTIONS

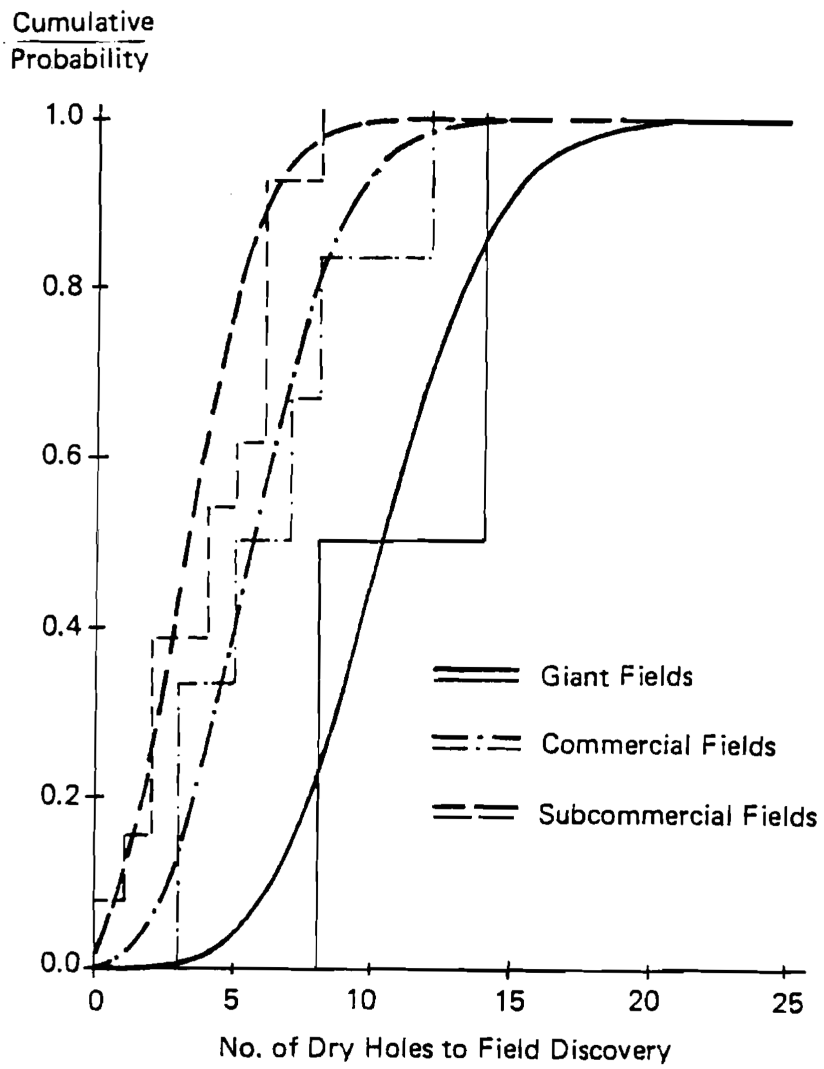


FIGURE 10. SERGIPE ALAGOAS DRY HOLE SAMPLING DISTRIBUTIONS

AO.1: The exploration process for each field in a given basin is *independent* of the exploration processes for all others in the same size category.

- Enhanced by careful data analysis.

  AO.2: All exploration processes have common underlying statistical characteristics depending on the field size category discovered and the basin geology.

- The total empirical dry hole sampling distribution is multimodal.

  A1: Dry hole sampling distributions for each field size category in a given basin are spatial Poisson.

- Poisson dispersion and goodness-of-fit tests.

  A2: Dry hole rates for each field size category in basins of similar geological type are identical.

- Smirnov Goodness-of-fit test of empirical distributions and two sample likelihood ratio test of spatial Poisson rate.

  A3: Exploration processes for fields in *different* size categories are statistically *dependent* due to the nature of geophysical deposition processes.

- Rejection of spatial Poisson goodness-of-fit test for total empirical dry hole sampling distribution.

FIGURE 11. ASSUMPTIONS UNDERLYING THE ANALYSIS

TABLE 4: SUMMARY OF STATISTICAL TESTS OF ASSUMPTION 1: DRY HOLE SAMPLING DISTRIBUTIONS FOR EACH FIELD SIZE CATEGORY IN A GIVEN BASIN ARE SPATIAL POISSON

N.B. EXCEPT AS NOTED ALL TESTS ARE CONDUCTED AT THE 5% SIGNIFICANCE LEVEL

TEST BASIN 1: RECONCAVO   TEST BASIN 2: SERGIPE-ALAGOAS

GIANT FIELDS	1	2
NUMBER OF FIELDS (m)	7	2
χ^2 STATISTIC FOR DISPERSION TEST	9.79	1.63
$\chi^2_{m-1;0.05}$	12.59	3.84
	PASS	PASS
MAXIMUM DEVIATION STATISTIC FOR		0.357 (Emp.)
KOLMOGOROV-SMIRNOV GOODNESS-OF-FIT TEST	0.291	0.842 (Pred.)
$K_{m;0.05}$	0.483	
	PASS	PASS
COMMERCIAL FIELDS		
NUMBER OF FIELDS (m)	10	6
χ^2 STATISTIC FOR DISPERSION TEST	14.3	9.37
$\chi^2_{m-1;0.05}$	16.9	11.07
	PASS	PASS

TABLE 4 (continued)

TEST BASIN 1: RECONCAVO TEST BASIN 2: SERGIPE-ALAGOAS

	1	2
MAXIMUM DEVIATION STATISTIC FOR KOLMOGOROV-SMIRNOV GOODNESS-OF-FIT TEST	0.116	0.281
$K_{m;0.05}$	0.409	0.519
	PASS	PASS

SUBCOMMERCIAL FIELDS

	1	2
NUMBER OF FIELDS (m)	29	13
χ^2 STATISTIC FOR DISPERSION TEST	42.71	17.5
$\chi^2_{m-1;0.05}$	41.3	21.0
$\chi^2_{m-1;0.02}$	45.4	PASS
(AT 2% SIG. LEVEL)	PASS	
MAXIMUM DEVIATION STATISTIC FOR KOLMOGOROV-SMIRNOV GOODNESS-OF-FIT TEST	0.062	0.170
$K_{m;0.05}$	0.246	0.361
	PASS	PASS

TABLE 5: SUMMARY OF STATISTICAL TESTS OF ASSUMPTION 2: DRY HOLE RATES FOR EACH FIELD SIZE CATEGORY IN PULL-APART (KLEMME TYPE 5) BASINS ARE SIMILAR

N.B.: EXCEPT AS NOTED ALL TESTS ARE CONDUCTED AT THE 5% SIGNIFICANCE LEVEL

TEST BASIN 1:	RECONCAVO	TEST BASIN 2:	SERGIPE-ALAGOAS
AREA:	10,000 KM^2	AREA:	12,000 KM^2
GIANT FIELDS	1:7		2:2
MAXIMUM DEVIATION STATISTIC FOR SMIRNOV TWO SAMPLE TEST OF EMPIRICAL DISTRIBUTIONS			0.714
$d_{0.9444;7,2}$			0.857
			PASS
MAXIMUM DEVIATION STATISTIC FOR RESCALED EMPIRICAL DISTRIBUTIONS TO TEST IDENTICAL DRY HOLE RATES (INDEPENDENT OF POISSION ASSUMPTION)			0.714
$d_{0.9444;7,2}$			0.857
			PASS
STANDARD NORMAL STATISTIC FOR LIKELIHOOD RATIO TEST OF IDENTICAL SPATIAL POISSON DRY HOLE RATES			2.84
0.975 (5% SIG. LEVEL)			1.96
0.995 (1% SIG. LEVEL)			2.58
			FAIL*
0.9977 (0.66% SIG. LEVEL)			2.84

*Probably due to small sample size

TABLE 5 (continued)

TEST BASIN 1: RECONCAVO TEST BASIN 2: SERGIPE-ALAGOAS

<u>COMMERCIAL FIELDS</u>	1:10	2:6

MAXIMUM DEVIATION STATISTIC FOR SMIRNOV TWO SAMPLE TEST OF EMPIRICAL DISTRIBUTIONS	0.233
$d_{0.9580;10,6}$	0.633
	<u>PASS</u>
MAXIMUM DEVIATION STATISTIC FOR RESCALED EMPIRICAL DISTRIBUTIONS (TO TEST IDENTICAL <u>RATES</u>)	0.233
$d_{0.9580;10,6}$	0.633
	<u>PASS</u>
STANDARD NORMAL STATISTIC FOR IDENTICAL SPATIAL POISSON DRY HOLE RATES	0.28
$N_{0.975}$	1.96
	<u>PASS</u>

<u>SUBCOMMERCIAL FIELDS</u>	1:29	2:13

<u>N.B.</u>: TABLES FOR MAXIMUM DEVIATION STATISTIC NOT AVAILABLE SO OBSERVATIONS RANDOMLY DELETED TO 1:20 2:12

MAXIMUM DEVIATION STATISTIC FOR SMIRNOV TWO SAMPLE TEST OF EMPIRICAL DISTRIBUTIONS	0.183
$d_{0.9571;20,12}$	0.467
	<u>PASS</u>
MAXIMUM DEVIATION STATISTIC FOR RESCALED EMPIRICAL DISTRIBUTIONS (TO TEST IDENTICAL <u>RATES</u>)	0.133
$d_{0.9579;20,12}$	0.467
	<u>PASS</u>
STANDARD NORMAL STATISTIC FOR IDENTICAL SPATIAL POISSON DRY HOLE RATES (FULL SAMPLE)	1.62
$N_{0.975}$	1.96
	<u>PASS</u>

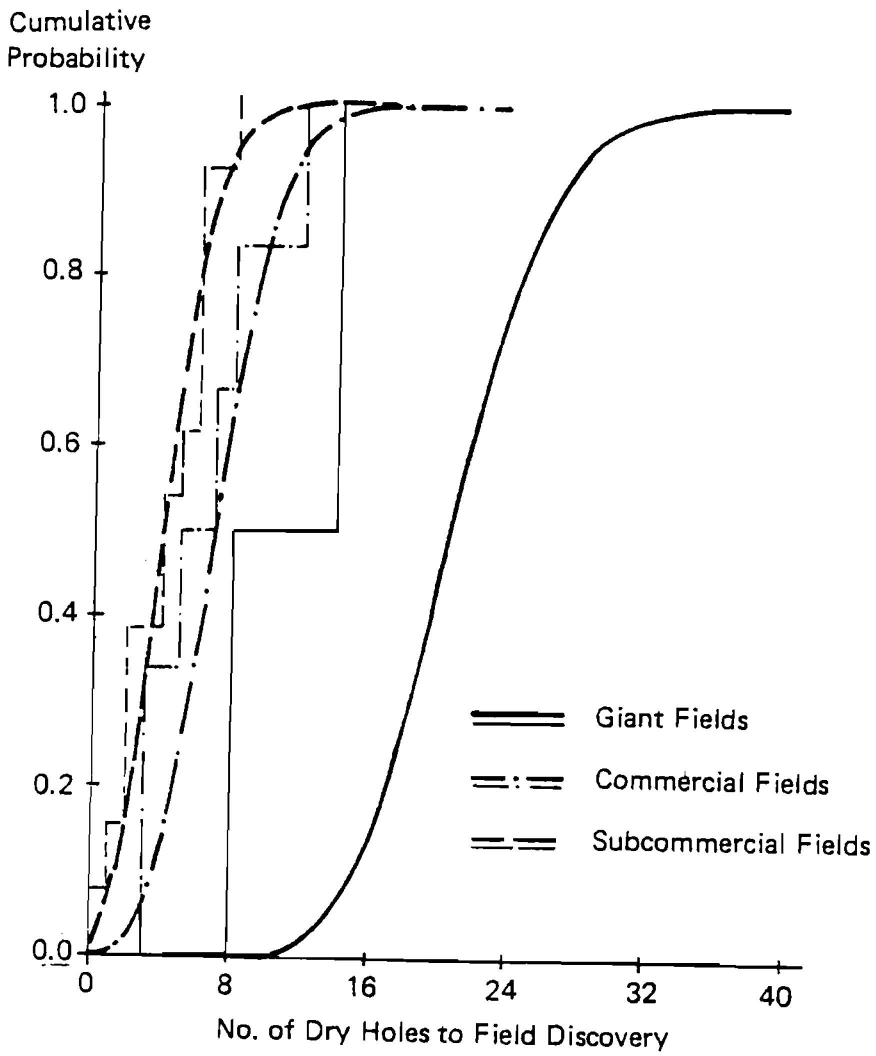
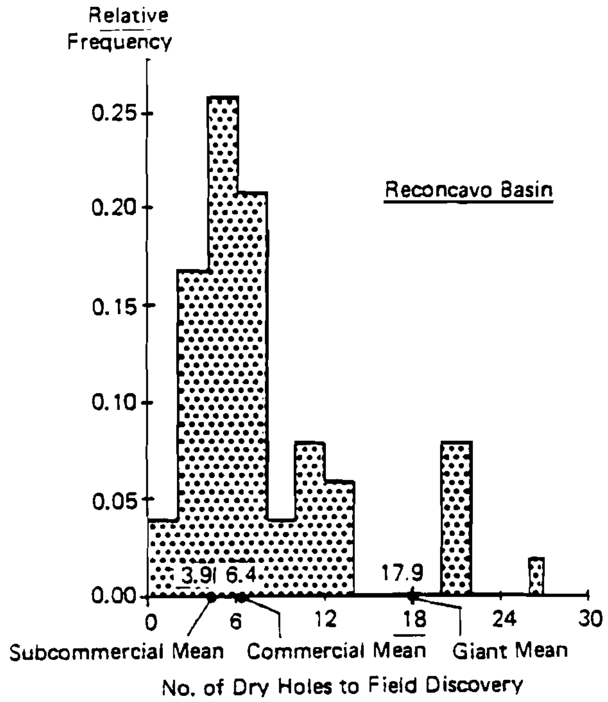


FIGURE 12. TRANSFERRED THEORETICAL VERSUS EMPIRICAL DRY HOLE SAMPLING DISTRIBUTIONS FOR THE SERGIPE ALAGOAS BASIN.



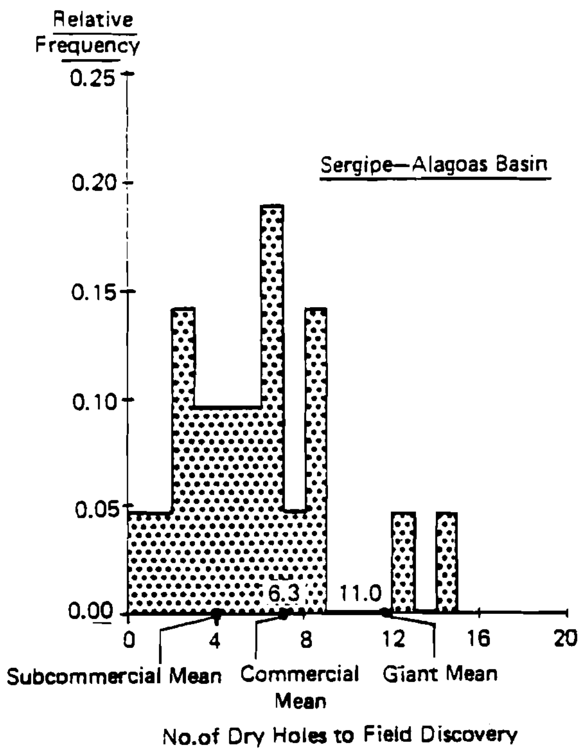


FIGURE 13. DRY HOLE SAMPLING HISTOGRAMS FOR THE TEST BASINS

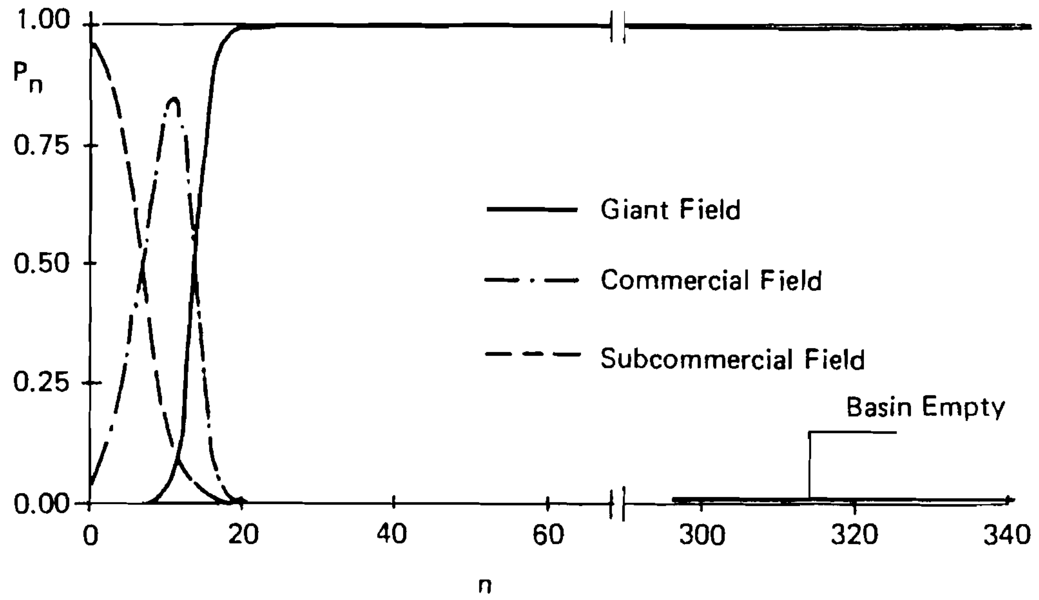


FIGURE 14. POSTERIOR RELATIVE DISCOVERY PROBABILITIES GIVEN THE NUMBER OF DRY HOLES PRIOR TO DISCOVERY.

APPENDIX

We have the following assumptions underlying our analysis which can be statistically tested:

(A1) Dry hole sampling distributions are spatial Poisson in a given basin.

(A2) Spatial Poisson dry hole rates in basins of similar geological type are identical.

(A3) Exploration processes for fields in different size categories are statistically dependent.

TESTS OF A1

1. Intrinsic test--the ***dispersion test*** -- of Poisson distribution by comparison of sample mean and variance which are theoretically identical (i.e. both λ) in the Poisson distribution

$$H_0 : f_{\underset{\sim}{n}} \textit{ Poisson vs. } H_1 : \sim H_0$$

Here $f_{\underset{\sim}{n}}$ denotes the unknown underlying sample density.

Test statistic:

$$t_1 := \sum_{i=1}^{m} (n_i - \bar{n})^2 / \bar{n}$$

where n_i is the number of dry holes for field i,

$$\bar{n} := \frac{1}{m} \sum_{i=1}^{m} n_i$$

is the average number of dry holes for the given field size category in the basin, and m is the number of fields in the given field size category for the basin.

$$\underset{\sim}{t_1} \sim \chi^2_{m-1} \textit{ approximately},$$

Therefore reject H_0 if

$$t_1 > \chi^2_{m-1;0.05}$$

where

$$\chi^2_{m-1;0.05}$$

is the 5% significance point for the χ^2 distribution with m-1 degrees of freedom *cf.* χ^2 tables in Hoel, p. 401).

Perform on each field size category for both test basins.

REF: D.R. Cox & D.V. Hinkley (1974). *Theoretical Statistics*. Methuen, London. pp. 73-74.

2. One-sample goodness-of-fit test to Poisson distribution.

$$H_0 : f_{\underset{\sim}{n}} \textit{ Poisson vs. } H_1 : \sim H_0$$

Kolmogorov-Smirnov ***maximum deviation test*** of the hypothesized spatial Poisson distribution.

Test statistic:

$$t_2 := \max_n |S_m(n) - F_{\underset{\sim}{n}}(n)| (:=K)$$

where $S_m(n) = r / m$ is the proportion r of the m fields with number of dry holes $\leq n$ and $F_{\underset{\sim}{n}}(n)$ is the value of the (cumulative) Poisson distribution at n.

$\underset{\sim}{t_2} \sim$ as in Table XIII of Bradley, pp. 367-9.

Therefore, reject H_0 if $t_2 > K_\alpha$ (table value for $\alpha = 0.025$) to give a 2-tailed test at the 5% level of significance. Note that due to a discrete distribution the true significance level is ***underestimated*** (see p. 303). Perform on each field size category for both test basins.

REF. J.V. Bradley (1968). *Distribution Free Statistical Tests*. Prentice-Hall, Englewood Cliffs, N.J. Section 13.5, pp. 296-304.

TESTS OF A2

3. Two-sample goodness of fit test ***independent*** of spatial Poisson assumption.

$$H_0 : f_{\underset{\sim}{n}_1} = f_{\underset{\sim}{n}_2} \;\; vs. \;\; H_1 : \sim H_0$$

where $\underset{\sim}{n}_j$ represents the random number of dry holes in a field size category in basin $j, j=1,2$. *Smirnov's* maximum deviation ***test*** for identical populations.

Test statistic:

$$t_3 := \max_i |d_i| \; (=D)$$

where

$$d_i := \frac{r_i}{m_1} - \frac{s_i}{m_2}$$

as defined in Bradley, pp. 289.

N.B: However, since we are really interested in spatial rates, sample numbers for the second basin B_2 should be multiplied by $\frac{A_2}{A_1}$, i.e. by the ratio of the basin areas before ranking to compute the d_i.

REF: F.J. Massey (1952). Distribution table for deviation between two sample cumulative *Annals. Math. Stat.* 23 (1952), 435-441.

$\underset{\sim}{t_3} \sim$ *as in Massey's tables* .

Reject H_o if $t_3 > d_\alpha$ (table value *nearest* $\alpha = 0.975$) to give a 2-tailed test nearest the 5% level of significance.

For uncommercial fields the closest we can get is $m_1 = 15$, $m_2 = 20$ *rather than 30* but observations may be randomly deleted to reach this sample size.

4. Two sample test for the spatial Poisson location parameter (i.e. mean): exact *likelihood ratio test.*

First note that if the $\underset{\sim}{n}_j$ are m *independent* Poisson variates with means λ_j, then

$$\sum_{j=1}^{m} \underset{\sim}{n}_j \sim \textit{Poisson with mean } \sum_{j=1}^{m} \lambda_j$$

REF: N.A.J. Hastings & J.B. Pearch (1975) *Statistical Distributions* Butterworth. London. p. 110.

This means that if we *assume:* -

(AO) The exploration process for each field in a given basin is *independent* of the exploration processes for all others and all have *common* underlying statistical characteristics depending on the basin geology.

then we may conduct a two sample test on the *sum* of the observed dry hole numbers in each field size category across the two basins.

$$y^1 := \sum_{j=1}^{m^1} \underset{\sim}{n}_j^1 \sim \textit{Poisson with mean } m^1\lambda^1 := \mu^1$$

and

$$y^2 := \sum_{j=1}^{m^2} n$$

where m^i is the sample size in basin i and λ^i is the mean (rate) in basin i, $i = 1,2$

$$H_o : \lambda^2 = \lambda^1 A^2 / A^1 \Leftrightarrow m^2\lambda^2 = \left[\frac{m^2 A^2}{m^1 A^1}\right] m^1 \lambda^1$$

$$i.e.\ \mu^2 = \left[\frac{m^2A^2}{m^1A^1}\right]\mu^1 \quad or \quad \mu^1 = \left[\frac{m^1A^1}{m^2A^2}\right]\mu^2 := \psi_o\mu^2 \quad vs. \quad H_1 : \sim H_o.$$

Test statistic:

$$s := \sum_{j=1}^{m^1} n_j^1 + \sum_{j=1}^{m^2} n_j^2$$

(a complete sufficient statistic). (The one-sided LR test is uniformly most powerful similar).

For accurate Gaussian approximation with continuity correction, use

$$t_4 := \frac{y^1 - (s\psi o/(1+\psi_o)) - 1/2}{\{s\psi_o/(1+\psi_o)^2\}^{1/2}}$$

$t_4 \sim N(0,1)$, i.e. Gaussian (standard normal), approximately.

For a 2-tailed test of H_o *vs.* H_1 at the 5% level of significance, reject H_o if

$$t_4 < N_{0.025} \ \ or \ \ t_4 > N_{0.975}$$

using standard normal tables from DeGroot, p. 577.

REF: D.R. Cox & D.V. Hinkley (1974). *op. cit.*, pp. 136-7.

TEST OF A3.

5. Dispersion test for *total* dry hole sample (same as 1).

$$H_o : f_{\underset{\sim}{n}}\ Poisson\ \ vs.\ \ H_1 : \sim H_o$$

Accept A3 if test rejects H_o at least at the 1% level of significance. Perform for both test basins.

REFERENCES

Arps, J.J. and J.G. Roberts (1958) *Economics of Drilling for Cretaceous Oil Production on the East Flank of the Denver- Julesberg Basin.* AAPG Bulletin, 42,11.

Asmus, H.E. and F.C. Ponte (1969) *The Brazilian Marginal Basins.* Petrobras unpublished report.

Barouch, E. and G.M. Kaufman (1978) *The Interface between Geostatistical Modeling of Oil and Gas Discovery and Economics.* Mathematical Geology, Vol. 10, No. 5.

Barouch, E. and G.M. Kaufman (1976) *Probabilistic Modeling of Oil and Gas Discovery* in Energy: Mathematics and Models, SIAM Institute for Mathematics and Society.

Bromberg, L. and J.A. Hartigan (1967) *Report to the Federal Energy Administration: United States Resources of Oil and Gas.* Department of Statistics, Yale University.

O'Carroll, M. and J.L. Smith (1980) *Probabilistic Methods for Estimating Undiscovered Petroleum Resources.* Energy Laboratory Working Paper No. MIT-EL 80-008WP.

Drew, L.J., J.M. Schuenemeyer and D.M. Root (1978) *The Use of a Discovery Process Model Based on the Concept of Area of Influence of a Drill Hole to Predict Discovery Rates in the Denver Basin.*, Rocky Mountain Association of Geologists. 1978 Symposium.

Energy, Mines and Resources, Canada EP77-1 (1977).

Hendricks, T.A. (1965) *Resources of Oil, Gas and Natural Gas Liquids in the United States and in the World.*

Holaday, B. and J. Houghton (1979) *Private Sector Supply Forecasting and Decision Making* in Energy Modeling Forum Vol. 5.

Hubbert, M.K. (1962) *Energy Resources: A Report to the Committee on Natural Resources: National Academy of Sciences.* National Research Council Pub. 1000-D.

Hubbert, M.K. (1966) *History of Petroleum Geology and its Bearing Upon Present and Future Exploration.* AAPG Bulletin, 50.

Hubbert, M.K. (1974) *U.S. Energy Resources: A Review as of 1972.* A National Fuel and Energy Policy Study, Part 1.

ICF Incorporated (1979) *A Review of the Methodology of Selected U.S. Oil and Gas Supply Models*

Jones, R.W. (1975) *A Quantitative Geologic Approach to Prediction of Petroleum Resources* in Methods of Estimating the Volume of Undiscovered Oil and Gas Resources. J.D. Hawn (ed.) American Association of Petroleum Geologists, Tulsa, Oklahoma.

Kaufman, G.M. (1980). Unpublished report.

Klemme, H.P. (1975) *Giant Oil Fields related to their Geology Setting: A Possible Guide to Exploration.* Bulletin of Canadian Petroleum Geology. Vol. 23, No. 1.

Levorsen, A.I. (1967) *Geology of Petroleum.* W.M. Freeman and Co. San Francisco, California.

Mallory, P.E. (1975) *Accelerated National Gas Resource Appraisal (ANOGRE).* in J.D. Haun (ed.) Method of Estimating the Volume of Undiscovered Oil and Gas Resources. Tulsa, Oklahoma.

McKelvey (1972) *Mineral Resources Estimates and Public Policy* American Scientist. Co.

Medova, E. (1980) *A Model for Resource Assessment and Exploration/Production Processes.* WP-80-44.

Menard, M.W. (1981) *Toward a Rational Strategy for Oil Exploration.* Scientific American Vol. 244. No. 1.

Meisner, J. and F. Demirman (1981) *The Creaming Method: A Bayesian Procedure to Forecast Oil and Gas Discoveries in Mature Exploration Provinces.* J. Royal Stat. Soc. A 144.1

Moore, C.L. (1966) *Projections of U.S. Petroleum Supply to 1980.* U.S. Dept. of the Interior. Office of Oil and Gas, Washington. D.C.

Newendorp, P.P. (1975) *Decision Analysis for Petroleum Exploration.* Petroleum Publishing Co.

Nehring, R. (1978) *Giant Oil Fields and World Oil Resources.* R-2284-CIA.

Ryan, J.M. (1965) *National Academy of Sciences Report on Energy Resources: Discussion of Limitations of Logistic Projections.* AAPG Bulletin, 49.

Ryan, J.M. (1966) *Models for Inferring Relationships Between Group Size and Potential Group Activity.* Behavioral Science, 11.

Smith, J.L. (1980) *A Probabilistic Model of Oil Discovery.* Review of Economics and Statistics. Vol. 62. No. 4.

Smith, J.L. and G.L. Ward (1980) *Maximum Likelihood Estimates of the*

Size Distribution of North Sea Oil Deposits. Energy Laboratory Working Paper No. MIT-EL 80-027 WP.

Steward-Gordon, T.J. *Brazil: The Target is Oil at any Cost.* World Oil. March 1980.

USGS Circular 725 (1975) *Geological Estimates of Undiscovered and Recoverable Oil and Gas Resources* in the United States.

Zapp, A.D. (1962) *Future Petroleum Producing Capacity of the United States.* in U.S.G.S. Bulletin. 1142-M.