BUREAU OF MINERAL RESOURCES, GEOLOGY AND GEOPHYSICS

RECORD 1984/4

RECORD

LAMDA: A COMPUTER PROGRAM FOR ESTIMATING HYPOTHETICAL PETROLEUM RESOURCES USING THE 'CREAMING METHOD'

by

A.L. HINDE

RECORD 1984/4

LAMDA: A COMPUTER PROGRAM FOR ESTIMATING HYPOTHETICAL PETROLEUM RESOURCES USING THE 'CREAMING METHOD'

by

A.L. HINDE

CONTENTS

SUMMARY

Computer program, LAMDA, is available on BMR's HP-1000 computer. It provides probablistic forecasts of the amounts of oil and gas that will be discovered, in individual petroleum provinces, as a result of drilling a specified number of new-field wildcat wells. The program fits a straight line to the historical log field size versus new-field wildcat number plot and the forecast is obtained by extrapolation of this line. This particular method of estimating petroleum resource potential is described by Forman and Hinde (in prep.).

Input to the program LAMDA consists of:

(1) Alphanumeric data for identification and documentation purposes.

(2) The exploration history specified as a series of oil (or gas) field sizes with corresponding new-field wildcat numbers and the total number of new-field wildcat wells drilled in the basin or province.

(3) Parameters and options pertaining to the forecast of future discoveries. These include the number of future new-field wildcat wells, a cutoff size for undiscovered fields, and a triangular distribution for the future success rate.

The input data is used to produce a series of graphic plots on the HP printer/plotter. They display discovery rate of oil (or gas), wildcat success rate, cumulative field size, log field size plotted against new-field wildcat number, and a normal plot of the recursive residuals of the fitted model. The recursive residuals are also written onto a file which can be read by another program (program 'W'). This program tests whether or not the residuals are normally distributed. This test, and the normal plot, aid in testing the adequacy of the linear model (program W is documented in Appendix I).

The assessment is given as a distribution and cumulative probability curve of resources together with the mean, standard deviation, and some percentiles. A seriation of average size and standard deviation of the undiscovered fields is printed as well as the distribution of the numbers of new discoveries.

INTRODUCTION

The well known tendency to find the large petroleum fields early during petroleum exploration is called the 'creaming effect' (Meisner and Demirmen, 1981). LAMDA is a FORTRAN program which models the decline in field size that results from the creaming effect. It uses the model described by Forman and Hinde (in prep.). The model is consistent with Australian experience and also the two hypotheses: (1) that fields are discovered from a population with a lognormal distribution and (2) that the probability of discovery of each field is proportional to field size raised to a constant power lambda (λ). The model predicts that a straight line is a reasonable fit to the plot of log field size versus discovery number. It also predicts a dependent relationship among lambda and the average slope and intercept of the fitted line and the average standard deviation of the residuals.

Program LAMDA uses extrapolations of a whole family of possible straight-line fits to the data as a basis for predicting the sizes of future discoveries. The user is required to specify the wildcat success rate, as a triangular distribution, and the number of future new-field wildcat wells.

A random sampling technique, described in the next section, is used to produce the assessment in terms of a cumulative probability curve of petroleum resources. Naturally, the further the extrapolations, the more spread out the probability curve. Similarly, the less data available, or the less it fits the model, the more spread out the probability curve.

In some areas there may be an initial learning period during which the size of field discovered tends to increase. In such a case it is necessary to ignore the data from the learning period and use only the subsequent data.

FORECASTING UNDISCOVERED RESOURCES

The steps carried out to forecast undiscovered petroleum resources are: (1) estimating lambda using the available exploration history, (2) obtaining the distribution of this estimate, and (3) simulating future discoveries (using random sampling). The results are presented as a probability distribution of undiscovered resources.

Estimating the value of lambda

The method of maximum likelihood (Kendall and Stuart, 1960) is used to estimate the value of lambda, given the complete population of field sizes, $(z_1, z_2, \ldots, z_n)$, z_1 having been discovered first, followed by z_2, etc. This method finds the value of lambda that maximises the likelihood function, $L = L(\lambda)$. The likelihood function is the probability of discovering the fields in the observed order given a value of lambda. In practice, the logarithm of the likelihood function is easier to maximise and gives the same result.

The likelihood function for the observed discovery order of all n fields is, by hypothesis (1):

$$L(\lambda) = \prod_{m=1}^{n} \frac{z_m^{\lambda}}{\sum_{i=m}^{n} z_i^{\lambda}} \qquad (1)$$

and its logarithm is:

$$\log L(\lambda) = \sum_{m=1}^{n} (\lambda \log z_m - \log(\sum_{i=m}^{n} z_i^{\lambda})) \qquad (2)$$

the log-likelihood function is maximised when its derivative is zero, i.e. when

$$\frac{\partial \log L(\lambda)}{\partial \lambda} = \sum_{m=1}^{n} \log z_m - \sum_{m=1}^{n} \left(\sum_{i=m}^{n} z_i^{\lambda} \right)^{-1} \left(\sum_{i=m}^{n} z_i^{\lambda} \log z_i \right) = 0$$

i.e. when

$$\sum_{m=1}^{n} \frac{\sum_{i=m}^{n} z_i^{\lambda} \log z_i}{\sum_{i=m}^{n} z_i^{\lambda}} = \sum_{m=1}^{n} \log z_m \qquad (3)$$

This equation can be solved numerically, yielding the maximum likelihood estimate for lambda, $\hat{\lambda}$, for the fields that have already been discovered. Note that the estimate of lambda will change as the sample size is increased with new discoveries. The true value can only be estimated when all fields have been discovered.

Distribution of the estimate of lambda

Maximum likelihood theory allows a probability density function to be assigned to the estimate of lambda (Kendall and Stuart, 1960). This density function assymptotically approaches a normal distribution when the number of observations is large.

The mean of this normal distribution is $\hat{\lambda}$ and its variance is given by:

$$\mathrm{var}(\hat{\lambda}) = -\left\{\frac{\partial^2 \log L(\lambda)}{\partial \lambda^2}\right\}^{-1}$$

$$= \left\{ \frac{\sum_{i<j} \sum_{j=m}^{n} z_i^{\lambda} z_j^{\lambda} (\log z_i - \log z_j)^2}{\left(\sum_{i=m}^{n} z_i^{\lambda}\right)^2} \right\}^{-1} \qquad (4)$$

Assessment of undiscovered resources

The assessment of undiscovered resources is carried out for a particular number of new-field wildcat wells using a triangular distribution of success rates, both of which must be specified by the user. The user also specifies the number of iterations. An iteration consists of one future exploration sequence.

During each iteration, one random value is selected from the specified triangular distribution of the success rate. The success or failure of each of the future new-field wildcats is determined by comparison of the sampled success rate with random numbers. If, for a given future well, the random number is less than the success rate, the well is considered a success. The number of successful wells is tallied throughout the iteration.

Also during each iteration, one random value is selected from the distribution of the estimate of lambda. A normal variate random number generator is used. The corresponding expected values of the slope, B, intercept, A, and standard deviation of the residuals, σ, of a straight line fitted to log field size versus discovery number, is determined via equations (5) to (7) following (Forman and Hinde, in prep.):

$$E(B) = B_{max}(1-e^{-\sigma_y \lambda}) \qquad (5)$$

where $E(B)$ is the expected or average value of the slope, B, of the fitted line

B_{max} is the slope (negative) when λ is infinite

σ_y is the standard deviation of the log field sizes

$$E(A) = \bar{y} - E(B)\bar{x} \qquad (6)$$

$$E(\sigma^2) = n(\sigma_y^2 - E(B)\sigma_x^2)/(n-2) \qquad (7)$$

where $\bar{y}$ = average log field size

$\bar{x}$ = average discovery number = $(n+1)/2$

σ_x = standard deviation of the discovery numbers $= \sqrt{(n^2-1)/12}$

n = number of discoveries

Only positive values of lambda are accepted, i.e., the fitted line must have a negative slope.

Log field sizes are then randomly generated for each discovery using a normal variate random number generator with standard deviation given by (7) and mean given by the straight line specified by (5) and (6).

The log field sizes are converted to actual field sizes. These are summed and when all iterations have been carried out the results are expressed as a probability distribution, or cumulative probability curve, of resources. The distribution of the number of discoveries and a seriation of the average field size, standard deviation, and discovery number of the undiscovered fields is also produced.

As explained by Forman and Hinde (in prep.), the method described here will be in error to some degree because it uses average values of slope, intercept, and standard deviation (equations (5) to (7)) instead of using their joint distribution (which is unknown). It should, however, be more reliable than using the least squares method proposed by Meisner and Demirmen (1981) as the latter method does not take into account the dependence among slope and intercept of the fitted line and the standard deviation of the residuals about the fitted line.

DESCRIPTION OF THE PROGRAM

Program LAMDA consists of the following sections: data input, calculation of model parameters, output of data and model parameters, graph plotting, forecasting section, and output of forecast.

Data input

The first section of the program handles data input. The program steps involved are as follows:

1. Read in and print a title to be used as a unique identification for the run. (This also appears on the graph plot output).
2. Read in and print the basin name.
3. Read in and print the subdivision.
4. Read in and print the date the data was compiled.
5. Read in and print whom the data was compiled by.
6. Generate and print the date of the computer run.
7. Read in and print the characters 'OIL' or 'GAS'.
8. Read in and print the volume units.

9. Read in and print remarks.

10. Read in: the number of wells drilled to date, N; the number of iterations (or runs) for the Monte Carlo simulation, NRUNS; the well number to be considered as the first well, NFIRST; the number of future wells to be drilled, MDRILL; a 5-digit integer to seed the random number generator, ISEED; whether a graph plot is to be produced ('Y' or 'N'); cutoff size for future fields, CUTOFF; the maximum value for the range of the histogram and cumulative probability of resources, HMAX; the triangular distribution for the success ratio.

11. Read in, for each successful well, the new-field wildcat number, the field size, the well date, and the name of the field. If NFIRST>0, subtract (NFIRST-1) from the well sequence number.

Step 11 is repeated until the end of data is reached. The total number of successful wells, NSUC, is then determined.

Calculation of model parameters

Although not used for assessments, a linear least squares model is fitted to log field size versus new-field wildcat number, along the lines of Meisner and Demirmen (1981). This is followed by the procedure to estimate lambda.

12. Calculate the recursive residuals, w_k, according to equations (3.16) and (3.17) of Meisner and Demirmen (1981), for k = 3 to NSUC, and print them onto file 'DATW'.

13. Calculate the estimates of the slope, (B), intercept, (A), and standard deviation of residuals, (σ^2), and compute the correlation coefficient (r), and the corresponding value of the z-statistic.

14. Use the method of maximum likelihood to calculate the estimate of lambda and the standard deviation of this estimate (subroutine GETLAM).

Output of data and of parameter estimates

15. Print the sequence of successful wells read in at step 10 (adjusted if NOFF>0).

16. Print the total volume of all fields, and the total number of wells drilled to date, N.

17. Print the estimates for B, A, σ^2, along with the T-matrix.

18. Print the values for r^2, r, and the z-statistic.

19. Print the estimate of lambda and the standard deviation of this estimate.

20. If this is an assessment run (NRUNS>0) then
    20.1 Print the cutoff size for additional fields.
    20.2 Print the triangular distribution for the success ratio.

Graph plots (uses subroutines PLOTD and RECTG, function CNORM, Calcomp plot routines).

Subroutine PLOTD takes care of producing the graph plots. Subroutine RECTG is a small routine to assist with plotting the plot frames. Function CNORM calculates the value of a standard normal variate such that the area under the curve to the right equals a specified value. For the actual plotting commands, use is made of the Calcomp subroutines: PLOTS, PLOT, SYMB, and NUMB.

The overall sequence in PLOTD is as follows:

21. Plot cumulative volume against number of exploratory wells drilled (showing overall decline).

22. Plot cumulative number of discoveries against number of wells drilled (showing decline in success ratio).

23. Plot cumulative volume against the number of discoveries (showing the decline in field size).

24. Plot log field size versus the new-field wildcat number. Superimpose the fitted straight line.

25. Produce a normal plot of the recursive residuals making use of function CNORM. Fit a straight line to these residuals.

26. Plot the identification.

Assessment section

The assessment section consists of the following subsections: initialization, iterative loop, calculation of statistics, and output of results. The subroutines and functions used are: GETTH, PHIST, MOMNT, PERC, URAN, GRAN, LSFIT, ORDER.

Initialization

27. Calculate the average assessed resource volume and its variance.

28. Zero the histogram arrays, and set the upper range limit of volume to 6 standard deviations above zero, unless this limit has been specified.

29. Calculate constants used in calculation of slope, intercept, etc.

30. Arrange the field sizes in descending order (subroutine ORDER).

31. Fit a straight line to the ordered fields to get the maximum slope (subroutine LSFIT).

32. Initialize variables for Monte Carlo loop.

Iterative loop

The following set of instructions are carried out NRUNS times to accumulate the histogram of recoverable volume.

33. Generate a random value of lambda from a normal distribution, with mean and standard deviation obtained in 14, using function GRAN.
    33.1 If this value is negative, go back to 33.

34. Calculate the slope, B, intercept, A, and standard deviation, σ, of the line to be extrapolated using equations (5) to (7).

35. Generate a random value of the success rate, θ, using the inputted triangular distribution (subroutine GETTH).

36. Set the total undiscovered reserve, RNM, to zero, and the number of new discoveries, NDISC, to 1.

37. For j = 1, 2, ... , MDRILL (the number of additional new-field wildcats).
    37.1 Generate a uniform random variate between 0 and 1 using function URAN.

37.2 If this is greater than θ, jump back to 37 and increment j.

37.3 Generate a random field size, VNJ, from a log-normal distribution, with mean equal to A + B(N+NDISC) and standard deviation, σ.

37.4 RNM = RNM + VNJ.

37.5 Return to 37 and continue the loop.

38. Update the histogram of RNM values. Update the number of discoveries histogram. Update the maximum field so far generated. Update sums for calculation of statistical moments.

39. Return to 33 and continue the loop.

Calculation of statistics and output of results.

40. Set up conversion factor from metric to imperial measure.

41. Print the number of future wells, MDRILL, and the number of Monte Carlo iterations, NRUNS.

42. Call PHIST to print and page plot the histogram of future reserve volume.

43. Convert the histogram to decreasing cumulative form.

44. Call PHIST to print and page plot the cumulative histogram of future reserve volume. Some vertical bars are plotted beneath the curve to give an indication of where there is greatest probability density. The more bars in a region, the greater the density there.

45. Call MOMNT to calculate and print the first four statistical moments of future reserve volume - both in metric and imperial units.

46. Use function PERC to calculate, then print some percentiles of the distribution of future reserve volume. These are expressed as percent probability greater than.

47. Calculate and print maximum field size ever obtained, and average maximum field size for each iteration and its standard deviation.

48. Print the seriation of average field size versus well number for the future fields.

49. Call PHIST to print and page plot the histogram of number of future discoveries.

50. Calculate and print the mean and standard deviation of the number of future discoveries.

51. End of program.

SUBROUTINES AND FUNCTIONS USED BY THE PROGRAM

Various subroutines and functions are used by the program. These are now described in more detail. Unless indicated as double precision, the routines are single precision.

Subroutine PLOTD

Purpose: This subroutine handles the graph plotting.

Method: The subroutine sequence was described in the previous chapter.

Usage: CALL PLOTD (IWELL,FSIZE,N,NSUC,ALPHA1,ALPHA2,BETA1,BETA2,S2,WK, ITITLE,IUNIT)

where

IWELL is an integer array containing the well sequence numbers.

FSIZE is a double precision array containing the corresponding field sizes.

N is the number of wells drilled to date.

NSUC is the number of successful wells.

ALPHA1, ALPHA2 are the values of the estimates $\hat{\alpha}_1$, $\hat{\alpha}_2$ (double precision).

BETA1, BETA2, S2 are the values of the estimates β_1, β_2, $\hat{\sigma}^2$ (double precision).

WK is a double precision array containing the recursive residuals, w_k, in elements 3 to NSUC.

ITITLE is an array containing the identification.

IUNIT is an array containing the units of volume.

Subroutine RECTG

Purpose: To plot a rectangular frame whose bottom left corner is at (x_o, y_o) with specified lengths of sides and a specified number of ticks (dashes) along the lower side. The ticks are numbered at every 10.

Usage: CALL RECTG (XO,YO,XLEN,YLEN,NX)

where

XO, YO are the coordinates of the bottom left corner.

XLEN is the length of the sides parallel to the x-axis.

YLEN is the length of the sides parallel to the y-axis.

NX is the number of ticks to be plotted.

Function CNORM

Purpose: Given a probability, p, CNORM calculates the value of x, under a standard normal curve, such that the areas to the right of x is equal to p.

Method: The algorithm is taken from Abramowitz and Stegun (1972) section 26.2.23. It uses a polynomial approximation:

$x \doteq t - (c_o + c_1 t + c_2 t^2)/(1 + d_1 t + d_2 t^2 + d_3 t^3)$

where $t = (\ln(1/p^2))^{\frac{1}{2}}$

$c_o = 2.515517$, $c_1 = .802853$, $c_2 = .010328$

$d_1 = 1.432788$, $d_2 = .189269$, $d_3 = .001308$

The error in the calculated value of x is less than 4.5×10^{-4}, i.e., x is determined to at least 3 decimal places.

Usage: X = CNORM(P)

where X and P are equivalent to x and p above.

Subroutine PHIST (double precision)

Purpose: To print and page-plot a histogram of 100 cells, either in frequency or cumulative frequency form.

Sequence of steps:

1. Determine the maximum value of frequency.
2. Set the first element of the plot array to '!' symbol, and the rest to '-' symbols.
3. For each cell of the histogram.
   - 3.1 Calculate the upper limit of the range for the cell.
   - 3.2 Calculate the standard deviation of the frequency.
   - 3.3 Determine the element of the plot array that corresponds to the frequency (the maximum frequency corresponds to the 50'th element).
   - 3.4 If histogram is not in cumulative form or a vertical bar is required below the curve then
     - 3.4.1 Fill the plot array with asterisks, up to the element of the plot array determined above.

     else
     - 3.4.2 Place an asterisk '*' in the element of the plot array determined above.
   - 3.5 Print out the upper limit of the cell's range, the frequency, its standard deviation, and the plot array.
   - 3.6 Fill the plot array with blanks, set the first element to the symbol '!'.

3.7 Go to 3 until all cells have been processed.

4. Return

Usage: CALL PHIST (IFREQ,NVERT,NRUNS,CMIN,RANGE,NHIST,ITYPE,LOUT)
where

IFREQ is an integer array, of length NHIST, containing the number of observations falling in each cell.

NVERT is an integer array, of length NHIST, containing 1's at the positions where vertical bars are to be plotted.

NRUNS is the number of observations.

CMIN is the lowest end of the range of the histogram.

RANGE is the length or range spanned by the histogram.

NHIST is the number of cells in the histogram.

ITYPE = 0 if the histogram is in cumulative form,
= 1 if it is in frequency form.

LOUT = the logical unit number of the output print device.

Subroutine MOMNT (double precision)

Purpose: To calculate and print the mean, standard deviation, coefficient of skewness and coefficient of kurtosis of a distribution.

Method: Consider a sample of data, $x_1, x_2, \ldots\ldots x_n$, from an unknown distribution and let

$$X1 = \sum_{i=1}^{n} x_i, \quad X2 = \sum_{i=1}^{n} x_i^2, \quad X3 = \sum_{i=1}^{n} x_i^3, \quad X4 = \sum_{i=1}^{n} x_i^4$$

The first four statistical moments about the origin are given by:

$$\bar{\mu_1} = X1/n, \quad \bar{\mu_2} = X2/n, \quad \bar{\mu_3} = X3/n, \quad \bar{\mu_4} = X4/n$$

and the first four statistical moments about the mean are given by:

$$\mu_1 = \bar{\mu_1}, \quad \mu_2 = \bar{\mu_2} - \bar{\mu_1}^2, \quad \mu_3 = \bar{\mu_3} - 3\bar{\mu_1}\bar{\mu_2} + 2\bar{\mu_1}^3,$$

$$\mu_4 = \bar{\mu_4} - 4\bar{\mu_1}\bar{\mu_3} + 6\bar{\mu_1}^2\bar{\mu_2} - 3\bar{\mu_1}^4$$

In terms of these, the mean is μ_1, the variance is μ_2 (hence the standard deviation is $\sqrt{\mu_2}$), the coefficient of skewness is given by, $\gamma_1 = \mu_3/\sqrt{(\mu_2^3)}$, and the coefficient of kurtosis is given by, $\gamma_2 = \mu_4/\mu_2^3 - 3$ (Kendall and Stuart, 1960, Vol I, page 85). The coefficients of skewness and kurtosis, so defined, have the property of equalling zero for a normal distribution.

Usage: CALL MOMNT (X1,X2,X3,X4,N,LOUT)

where

X1, X2, X3, X4 are the sums defined above.

N is the sample size n.

LOUT is the logical unit number of the output print device.

Function PERC (double precision)

Purpose: To calculate percentiles of a cumulative distribution.

Method: Given a histogram in decreasing cumulative frequency form, PERC uses linear interpolation to calculate the value at which the cumulative frequency equals a specified value, p.

Usage: V = PERC(IHIST,NHIST,NRUNS,HMIN,HMAX,P)

where

IHIST is an integer array, length NHIST, containing the decreasing cumulative number of observations in each cell.

NHIST is the number of cells.

NRUNS is the number of observations.

HMIN, HMAX are the minimum and maximum limits of the range of the histogram.

P is the decreasing cumulative probability at which V is to be calculated.

Subroutine GETTH (double precision)

Purpose: To generate a random value from a triangular distribution (to be used for the success ratio).

Method: Let the minimum, most likely and maximum values of the triangular variate be x_1, x_2, x_3 respectively. The cumulative probability for this distribution is given by:

$$F(x) = (x-x_1)^2/[(x_3-x_1)(x_2-x_1)] \qquad x_1 \leq x \leq x_2$$

$$= 1 - (x_3-x)^2/[(x_3-x_1)(x_3-x_2)] \qquad x_2 \leq x \leq x_3$$

A random uniform variate between 0 and 1 is generated, using the HP function URAN, and this is set equal to F(x). The corresponding value of x, found by inverting the above equation, is given by:

$$x = [F(x)(x_3-x_1)(x_2-x_1)]^{\frac{1}{2}} + x_1 \quad , \quad 0 \leq F(x) < (x_2-x_1)/(x_3-x_1)$$

$$x_3 - [(1-F(x))(x_3-x_1)(x_3-x_2)]^{\frac{1}{2}}, \quad (x_2-x_1)/(x_3-x_1) \leq F(x) \leq 1$$

This value is returned by the subroutine.

Usage: CALL GETTH (TH,X1,X2,X3,LOUT)
where
TH is the returned value of the triangular variate.
X1, X2, X3 are x_1, x_2, x_3 specifying the triangular distribution.
LOUT is the logical unit number of the output print device.

Subroutine ORDER (double precision)

Purpose: To order the first M elements, of the vector array P, in descending order.

Method: The elements are ordered by a bubble sort.

Usage: CALL ORDER (P,M)
where
P is a one-dimensional array.
M is the number of elements to be ordered.

Subroutine LSFIT (double precision)

Purpose: To fit a straight line to the logarithms of a set of y values, and return estimates of the slope, intercept and variance of residuals and the standard deviations of the estimates of slope and intercept. The x-values are the first n integers, where n is the number of y-values.

Method: The method of least squares is used (Kendall and Stuart, 1967, chapter 19).

Usage: CALL LSFIT (Y,N,A1,B1,VAR,SDA1,SDB1)
where
Y is a vector array containing the set of y-values.
N is the number of y-values.
A1 is the estimate of the intercept.
B1 is the estimate of the slope.
VAR is the estimate of the variance of residuals.
SDA1 is the standard deviation of the estimate of A1.
SDB1 is the standard deviation of the estimate of B1.

Subroutine GETLAM (double precision)

Purpose: To get the maximum likelihood estimate of lambda, and the standard deviation of the estimate, given an ordered set of field sizes, y.

Method: The first and second derivatives of the log-likelihood equation are calculated for a trial value of lambda. This value is repeatedly improved, using Newton's method, until the change in lambda is less than 5×10^{-5}.

Usage: CALL GETLAM (X,N,ALAM,SDALAM)

where

X is a vector array containing the field sizes in order of discovery.

N is the number of fields.

ALAM is the estimate of lambda.

SDALAM is the standard deviation of the estimate of lambda.

INPUT DATA

This chapter describes how to prepare the input data file required by program LAMDA. The name of this file is 'DATL'.

Line	Cols	Variable	Description	Format for entire line
1	1-80	ITITLE	Used to identify the computer run	(40A2)
2	1-80	ITITL2	Basin name	(40A2)
3	1-80	ITITL2	Sub-basin or infrabasin	(40A2)
4	1-80	ITITL2	Date the data was compiled	(40A2)
5	1-80	ITITL2	Who compiled the data	(40A2)
6	1-80	ITITL2	The characters 'OIL' or 'GAS'	(40A2)
7	1-10	IUNIT	The units of volume (See Note)	(5A2)
8	1-80	ITITL2	Remarks	(40A2)
9	1-5	N	Number of exploration wells	(5I5,A1,2F10.0)
	6-10	NRUNS	Number of Monte Carlo iterations to be carried out (2000 or 5000 is suggested as reasonable numbers). If no assessment is required, leave blank.	
	11-15	NFIRST	The well number in the data to be considered as the first well (this is intended to be used for eliminating an early incline or learning period). If not required, leave blank.	
	16-20	MDRILL	The number of additional wells to be drilled. If no assessment is required, leave blank.	
	21-25	ISEED	A 5-digit number to seed the random number generators, URAN and GRAN. If omitted, 12345 is used.	
	26	IFPLT	If 'Y' - allows the graph plots to be produced.	
			If 'N' or blank - inhibits graph plot production.	

Line	Cols	Variable	Description	Format for entire line
	27-36	CUTOFF	Cut-off or minimum size for future fields generated; if a field is generated smaller than CUTOFF it is ignored. If not required, leave blank.	
	37-46	HMAX	Maximum value of the range in the histograms of undiscovered reserves. If left blank, the program calculates a value.	
10	1-10	XTH1	Minimum value of the success ratio for the triangular distribution of success ratios.	(3F10.0)
	11-20	XTH2	Most likely value of the success ratio for this distribution.	
	21-30	XTH3	Maximum value of the success ratio for this distribution.	

If no assessment is required, this card may be left blank.

11	1-5	IWELL(I)	Sequence number (new-field wildcat number) of I'th successful well.	(I5,F10.5,A2,1X,10A2)
	6-15	ESIZE(I)	Size of the I'th successful well.	
	16-17	IYEAR(I)	Year I'th successful well was completed.	
	19-38	INAME(I,J), J=1,10	Name of the I'th successful well.	

Line 11 is repeated until all successful wells have been entered.

Note: The units for volume, variable IUNIT, are usually $10^6 m^3$ (abbreviated to MCM) for oil and $10^9 m^3$ (abbreviated to BCM) for gas. The program converts these to 10^6 BBL (MMB) and $10^{12} ft^3$ (TCF) respectively for output.

RUNNING THE PROGRAM

The program resides in the BMR's Hewlett-Packard computer, and is called LAMDA. To run the program, after the data file 'DATL' has been prepared, use the commands:

```
RP,LAMDA    (to restore program LAMDA)
RU,LAMDA    (to run program LAMDA;  output to the terminal)
```

The output will be sent back to the terminal. To send the output to the printer instead, use:

```
RU,LAMDA,6  (to run program LAMDA;  output to LU6)
LP,-N       (to print the output file)
```

Example terminal session

The following example of a complete terminal session, including log-on and log-off, is given, so that persons with little or no experience of the computer system may run the program. Data or information that would be displayed on the terminal is underlined; data or information that is to be typed in by the user is not underlined. The symbol '␣' indicates that a blank space is to be typed. The 'return' key is to be pressed at the end of each line typed in.

```
Press any key
PLEAS LOG-ON:  ALAN.RESAS
PREVIOUS TOTAL SESSION TIME etc....
SECURITY CODE?  915 *
!ED,-                          (Run the editor in input mode)
:AMADEUS OIL                   (Beginning of data: identification)
:AMADEUS BASIN                 (Basin name)
:-                             (Sub-basin - not specified)
:DEC 1982                      (Compilation date)
:D.J.F.                        (Data compiled by)
:OIL                           (Oil or gas)
:MCM                           (Units)
:-                             (Remarks)
```

* example only, security code is confidential

:-^^500^5000^^^^^^^^5012479Y	(500 wells so far drilled; 5000 Monte Carlo iterations; 50 additional wells; seed for random number generators is 12479; produce graph plot)
:.1^^^^^^^^.2^^^^^^^^.25	(Specification of the triangular distribution of success ratio)
:20^^^5.5^^^^^^^FIRST DISCOVERY	(The first discovery occurred with the 20'th well; field size 5.5; field name 'First Discovery')

(continue with the rest of the discoveries)

:742^^^.03^^^^^^^LAST DISCOVERY	(This is the last card of the data specifying the last discovered field).
:*EOD	(Exit from input mode)
:DU,DATL	(Overwrite the file 'DATL', if an error occurs here try SF,DATL to save it as a new file)

DATL CLOSED 26 LINES

!RP,LAMDA	(Restore program LAMDA)
!RU,LAMDA,6	(Run program LAMDA, output to the printer)

/LAMDA: STOP 00000
LAMDA LUN 06 SPOOL FILE - 0444

	(The spool file, -0444, vaies from run to run)
!LP,-N	(Send output to the printer)
!EX	(Log off)

APPENDIX I - A PROGRAM TO TEST FOR NORMALITY

Program 'W' is available on the BMR's HP computer for testing whether a sample comes from a normal distribution or not. The program is for small samples only (50 or less).

Method The program reads up to 50 values to be tested and calculates the W-statistic, as described in Shapiro and Wilk (1965), for the sample. It uses the coefficients in Table 5 of Shapiro and Wilk to calculate the W-statistic.

Usage A file called 'DATW', containing the sample values in F10.0 format (one value per line) must be prepared. This file is automatically produced whenever program LAMDA is run.

The program is then restored with the command:

```
RP,W
```

and run with the command

```
RU,W
```

The program asks whether the data are to be log-transformed or not. Type 'N' in response to this question if testing the model in CREAM. (The option of taking logarithms of the data enables log-normality to be tested).

Output The program lists the data in order of increasing value.

The value of the W-statistic is given, along with 3 percentage points of the distribution of W : 1%, 5%, 50%. Small values of W are significant, i.e. they indicate non-normality.

Example: 10 observations gave a value of W of 0.80.

The percentage points for 10 observations are:

percentage	1%	5%	50%
W-statistic	.781	.842	.938

The value of W is less than the 5% percentage point but greater than the 1% point, hence the observed value of W is significant at the 5% confidence level (indicating non-normality) but not at the 1% level.
(Note: statisticians usually use the 5% confidence level).

ACKNOWLEDGEMENTS

The author wishes to thank Dr M.H. Nederlof, of Shell International, and Dr D.J. Forman for valuable discussions.

REFERENCES

FORMAN, D.J., and HINDE, A.L., in prep. - A model of oil and gas discovery and an improved statistical method for assessment of undiscovered resources.

MEISNER, J., and DEMIRMEN, F., 1981 - The creaming method: A Bayesian procedure to forecast future oil and gas discoveries in mature exploration provinces. Journal of the Royal Statistical Society A, 144.

NEDERLOF, M.H., 1980 - Calibrated computer simulation as a tool for exploration prospect assessment. In: Assessment of undiscovered oil and gas. Proceedings of the seminar. Kuala Lumpur, Malaysia.

ABRAMOWITZ, M., and STEGUN, I.A., eds., 1972 - Handbook of mathematical functions. Dover Publications, Inc., New York, N.Y.

SHAPIRO, S.S., and WILK, M.B., 1965 - An analysis of variance test for normality (complete samples). Biometrika, 52, parts 3 and 4, pp 591-611.

KENDALL, M.G., and STUART, A., 1960 - The advanced theory of statistics, vol. 2. Charles Griffin & Co., London.