Rerum Naturalium Fragmenta

No. 459

PEREC
Prospect Evaluation for Windows User's Guide
Part 2
by
Tamas Jasko

Watford
2002

Rerum Naturalium Fragmenta
Tamas Jasko editor
16 Melrose Place, Watford WD17 4LN, England

Chapter 3
FACTOR MODELS

The automatic validation and distribution fitting that follows the editing of a data file shows little detail of the calculations. The program will only pause if the validation finds errors in the model data. If you want to see the detailed validation messages then select Check Model from the Model menu.

The shape of factor distributions of a validated model can be shown in graphic form by the Factor Display option

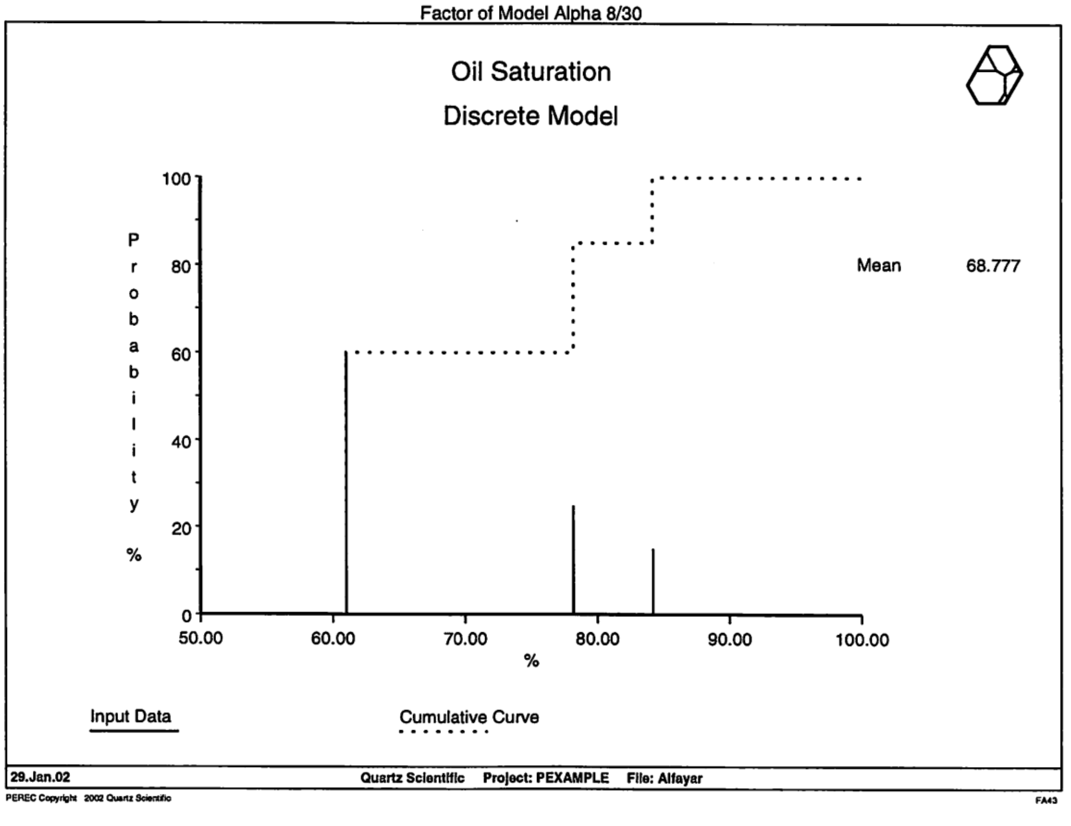


Figure 3.1: Discrete distribution

Constant factors

If the exact value of the factor is known it can be entered as a single value (with 100% probability). The factor is used then as a constant multiplier by the program. On data sheets it will be displayed as 'Constant:'. For a constant it does not matter what you specify as 'distribution model'.

More usually, the exact value of the factor is not known. The value is known only as a random variable within a distribution range. Factor distribution models define the way such random values can be obtained for simulation.

Probability distributions

When you are entering a model, for each of the geological parameters of a reservoir (the factors of a model) you can select a probability model:

Beta (minimum, maximum)
Beta (minimum, most likely, maximum)
Cumulative
Discrete
Expanded beta (minimum, maximum)
Expanded beta (minimum, most likely, maximum)
Frequency
Guided beta fit
Fitted cumulative
Johnson's Su curve
Johnson's Sb curve
Log-normal histogram fit
Log-normal cumulative fit
Normal histogram fit
Normal cumulative fit
Pearson's fitted beta
Smoothed frequency
Triangle
Uniform
Weibull
Explicit normal
Explicit log-normal
Explicit beta

Each of these models defines a probability distribution of the values that the factor can take. The nature of the factor and the available data determine the best choice of model e.g. gross rock volume is generally considered to have log-normal distribution.

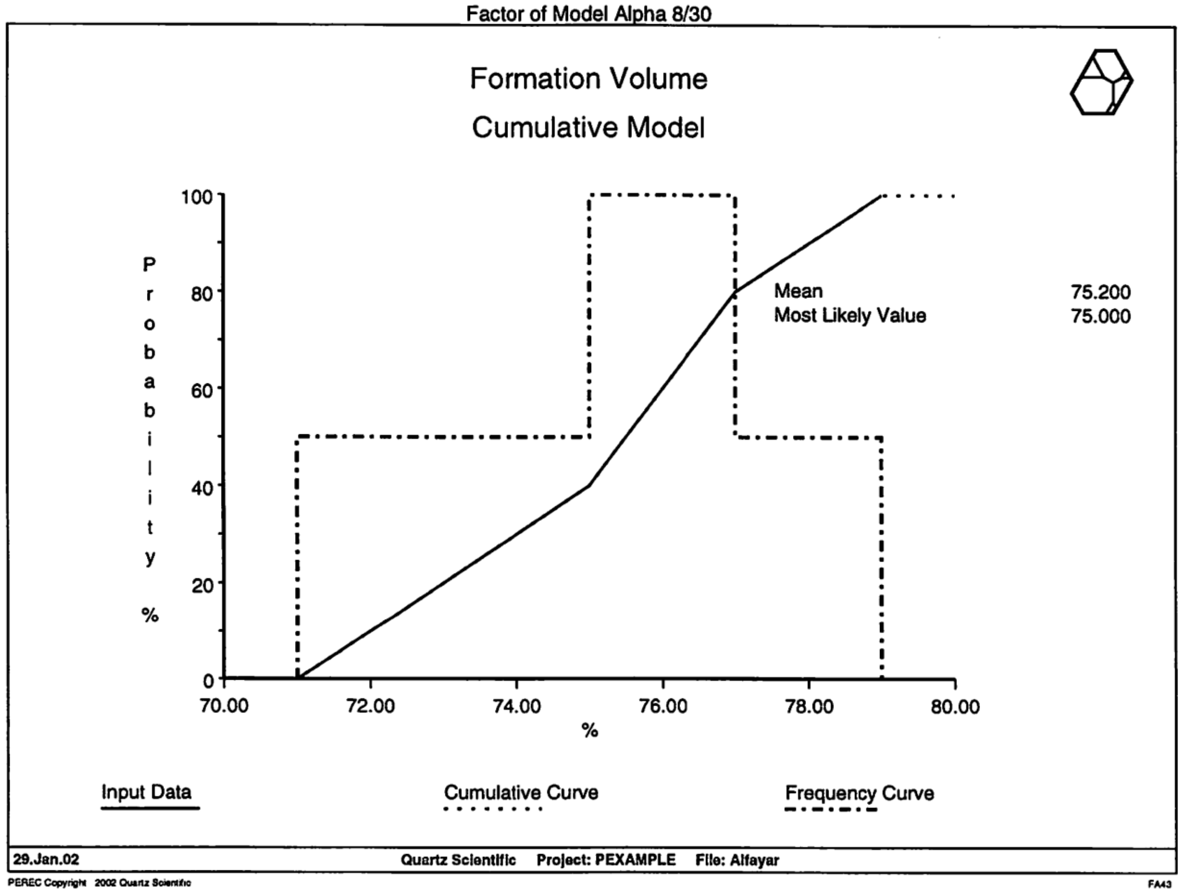


Figure 3.2: Cumulative model

Discrete distribution

Factors that may take a number of discrete values, rather than having a continuous distribution, can be modelled using model D. The model is defined by the actual discrete values and their probabilities. Factor shape plots for discrete distributions display the input data as vertical bars (the height of the bar is the probability of the value) and the corresponding stepped cumulative graph.

Cumulative models

Any shape of continuous distribution can be specified in these models by directly entering values that are read off hand drawn cumulative curves rather than referring to the normal/lognormal fits. This way the model fitting stage is bypassed.

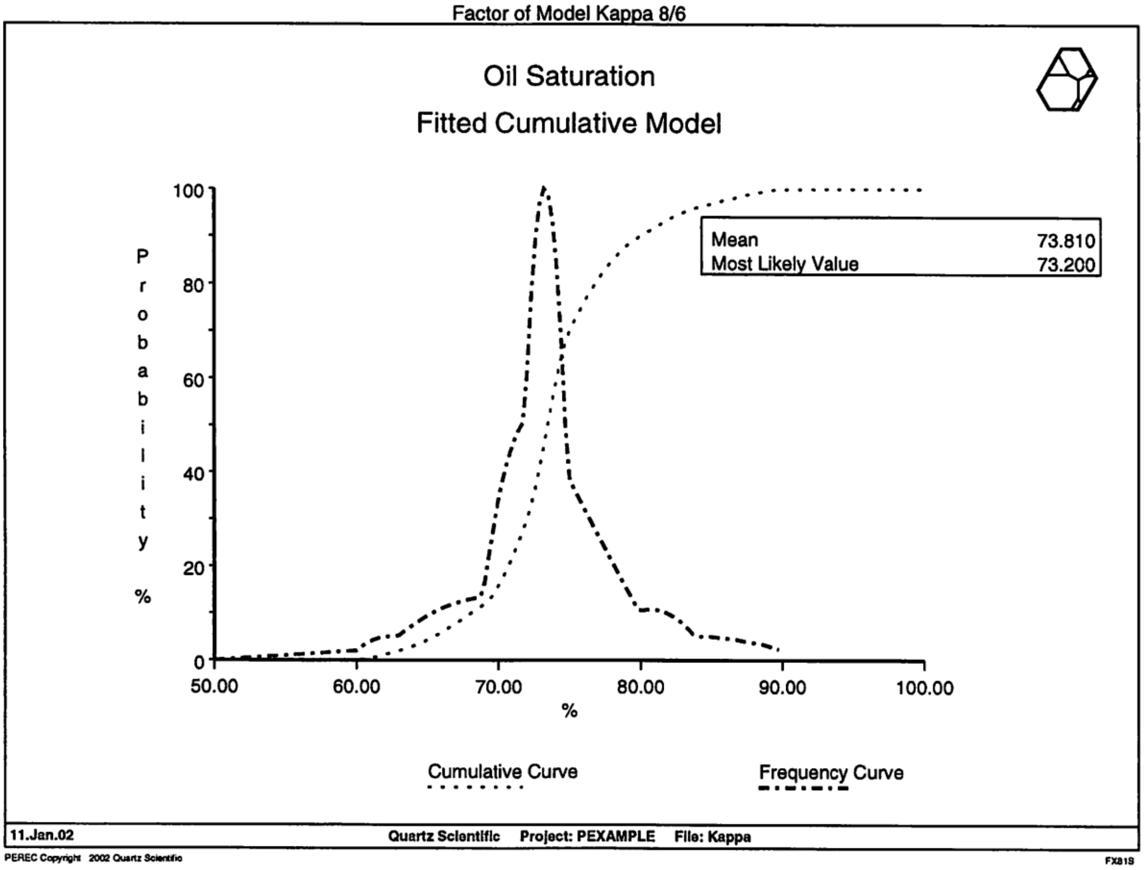


Figure 3.3: Fitted cumulative curve

This is particularly useful if the distribution is known to have an unusual shape that can not be well approximated by normal, lognormal or beta distribution. E.g. if the distribution is bimodal, having more than one peak.

Cumulative distributions (model C) are entered as a series of values with their cumulative probabilities. The first value should be the minimum value and its probability set to 0; the last value should be the maximum value with 100 % probability.

Up to 10 intermediate values can be entered to show the course of the curve between the extremes. These should be ordered by increasing value and probability.

Fitted cumulative

This is similar to the above but an interpolating curve is fitted to the points that define the cumulative model. The shape of the resulting distribution curve, which honours all the data points, appears more natural and pleasing than the straight line segments of the 'ordinary' cumulative option. Data should be entered as above.

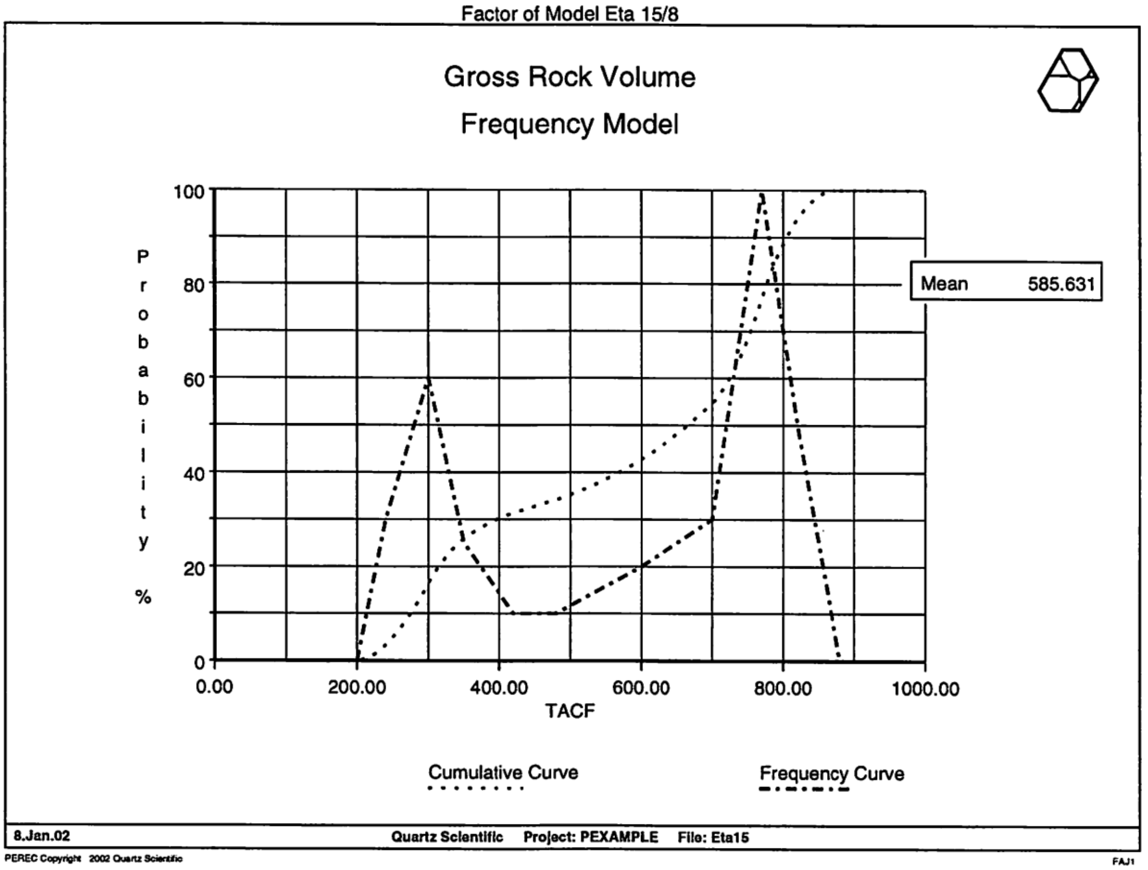


Figure 3.4: Frequency model

Frequency distributions

Frequency distribution models, like cumulative ones, are specified by pairs of values; with the first and last values indicating minimum and maximum, respectively. Both extremes should have 0 probability, the probability of the intermediate values, ordered by value, is given on a relative 0% to 100 % scale.

The example plot shows a bimodal distribution entered as a frequency distribution (model Fr).

Definition by frequency distribution is usually to be preferred to cumulative models as the shape of the curves is less jerky and frequency models are therefore more realistic approximations to actual continuous distributions. Factor validation plots for either model will show both cumulative and frequency curves. One of these is obviously the input itself - the other is computed by the program. Also, the average (mean) of the distribution is computed and displayed.

Smoothed frequency

Model Sf, for Smoothed Frequency, is similar to the above but the defining points are connected by smooth interpolation that gives the distribution curve a natural and pleasing look while honouring the data points.

Data entry and display formats are the same as for the 'ordinary' frequency curve that has straight line segments.

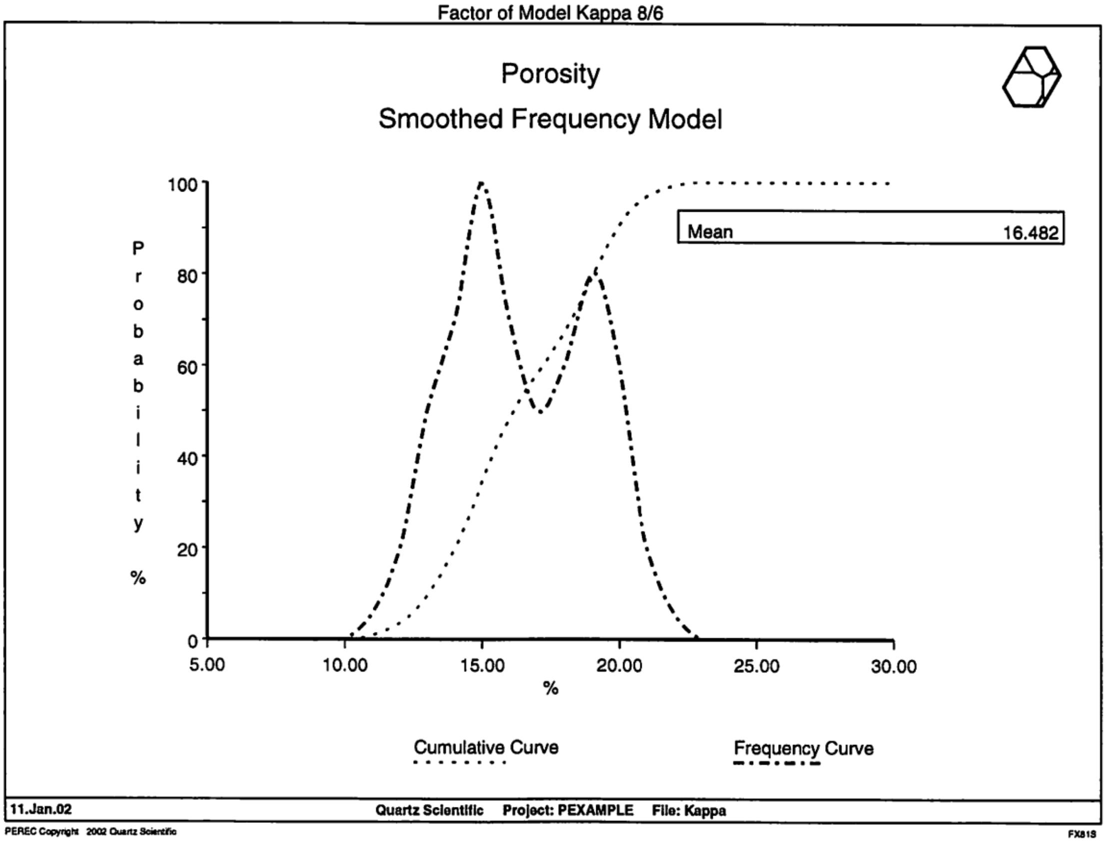


Figure 3.5: Smoothed frequency model

Triangular distribution

Often the only data available about a geological factor are the minimum, maximum and most likely value. For such cases the frequency distribution defined by the three values is known as the triangular distribution (See next figure). The minimum, most likely and maximum values should be entered in the model with 0, 100, 0 % probability, respectively.

Uniform distribution

If only the minimum and maximum values are known these can be entered in the form of a uniform (rectangular) distribution. Enter the minimum value with 0 %, the maximum value with 100 % probability.

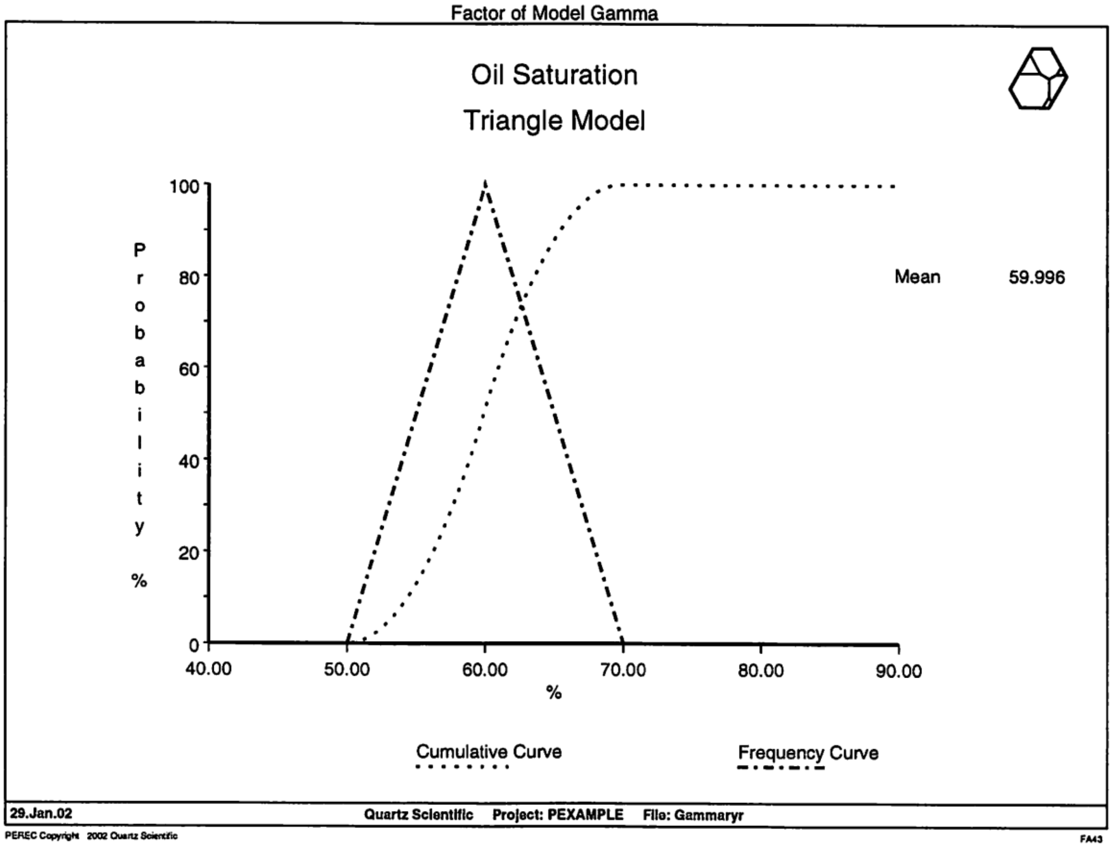


Figure 3.6: Triangular distribution

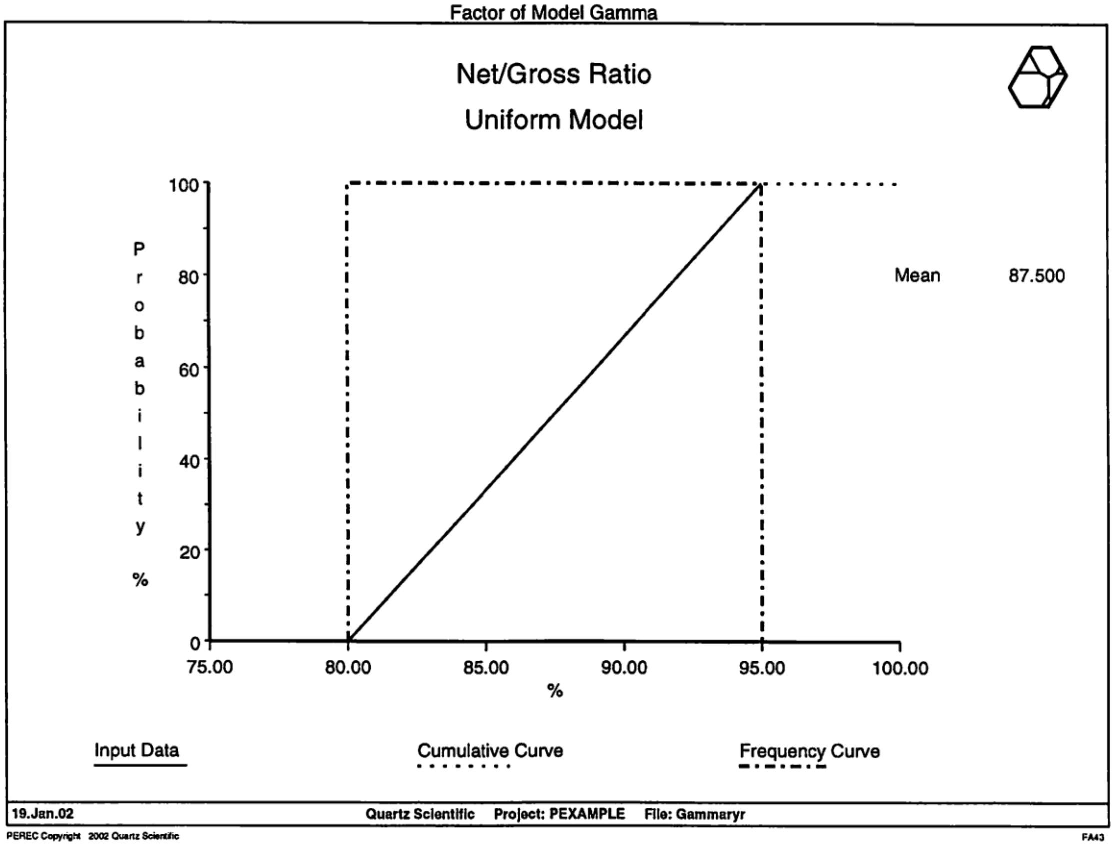


Figure 3.7: Uniform distribution

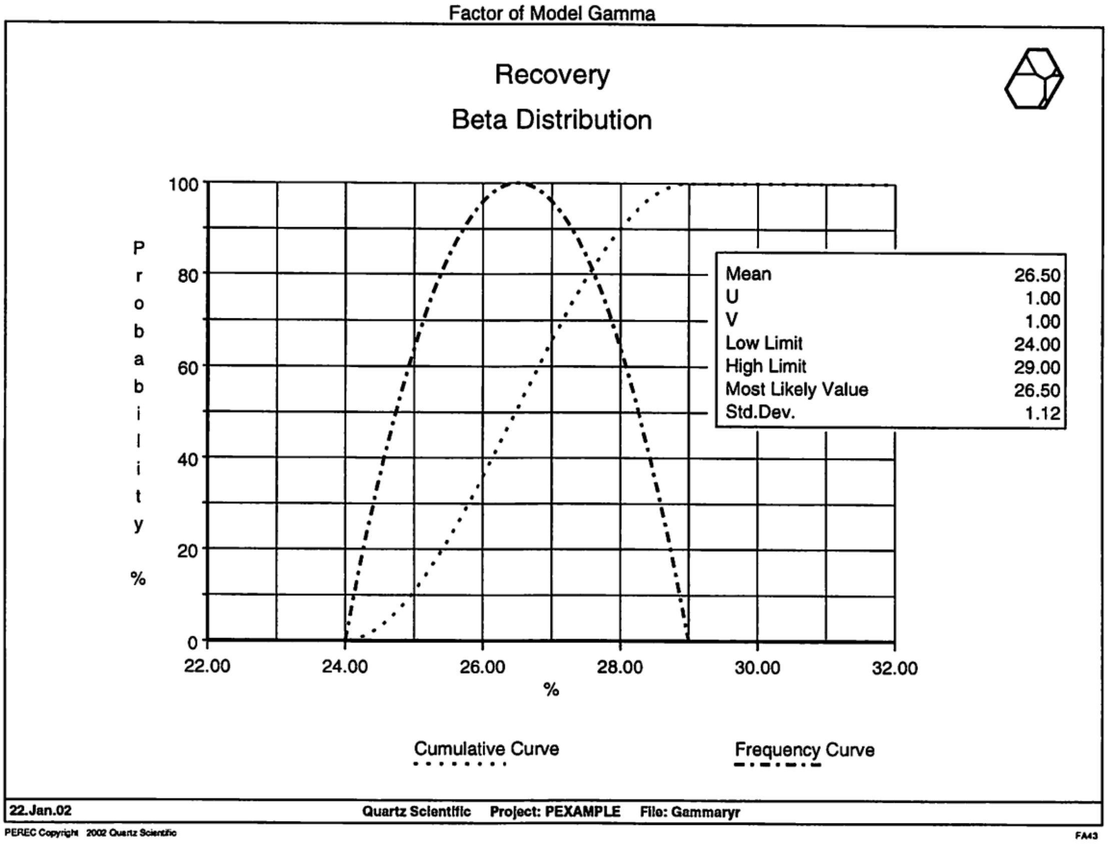


Figure 3.8: Beta distribution (min, max specified)

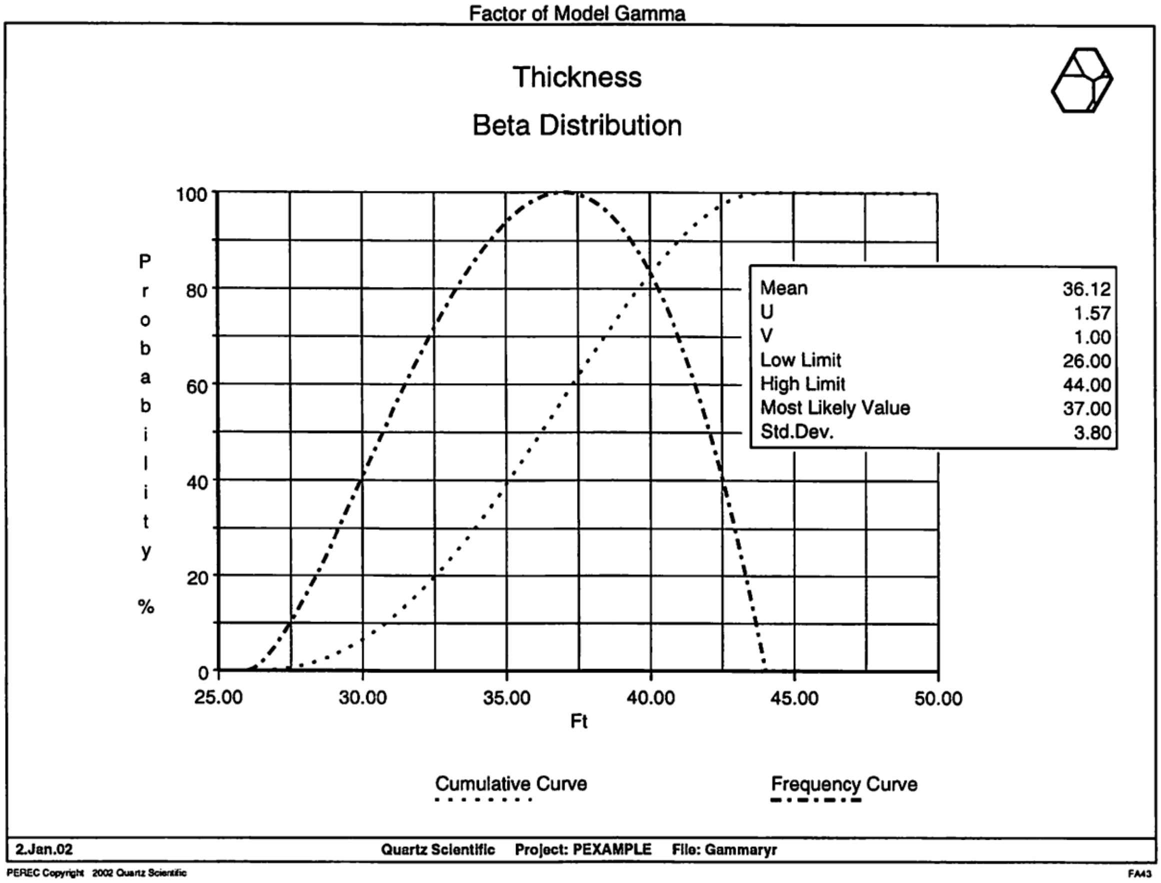


Figure 3.9: Beta distribution (min, most likely, max specified)

Beta distribution

Beta distributions are useful in simulating factors where minimum and maximum values are given and a smooth continuous distribution is required. If you enter 2 values only for minimum and maximum then a symmetrical curve is produced, if you enter minimum, most likely and maximum values (in that order) then the resulting beta distribution will be skewed left or right.

The two (or 3) values are to be entered in the left side in the proper order, the right side can be left empty.

Expanded beta distribution

If you want to specify the distribution by entering minimum, most likely and maximum values but would like a shape that is close to the bell shape of the normal and lognormal distribution then use the expanded beta (Eb) distribution. The minimum and maximum values you specified become the effective range (attained at 1% of maximum frequency) whilst the theoretical range of the curve is stretched beyond the effective range.

Normal & lognormal fits

Models L, M, N, and O are fitting a normal or lognormal distribution to a set of representative values. If either of these model options are used the program will fit the normal or lognormal distributions that approximate best the values (and their probability weightings) supplied by the user.

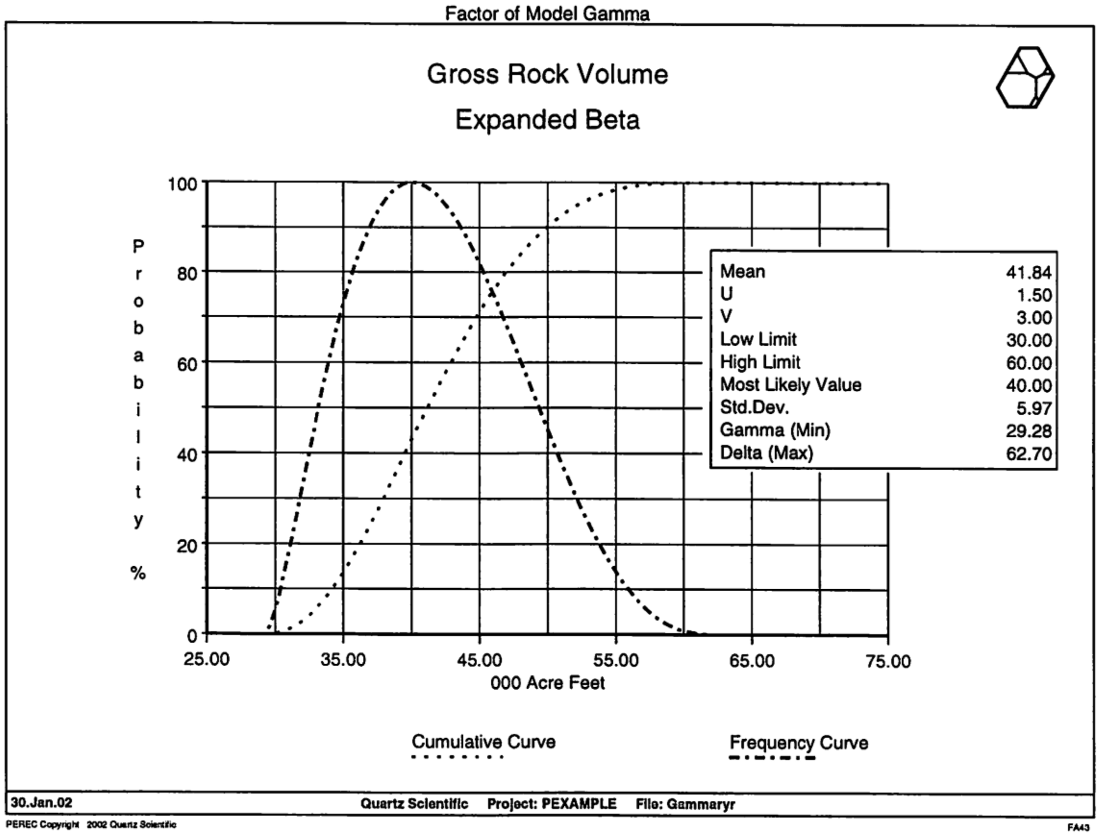


Figure 3.10: Expanded beta distribution

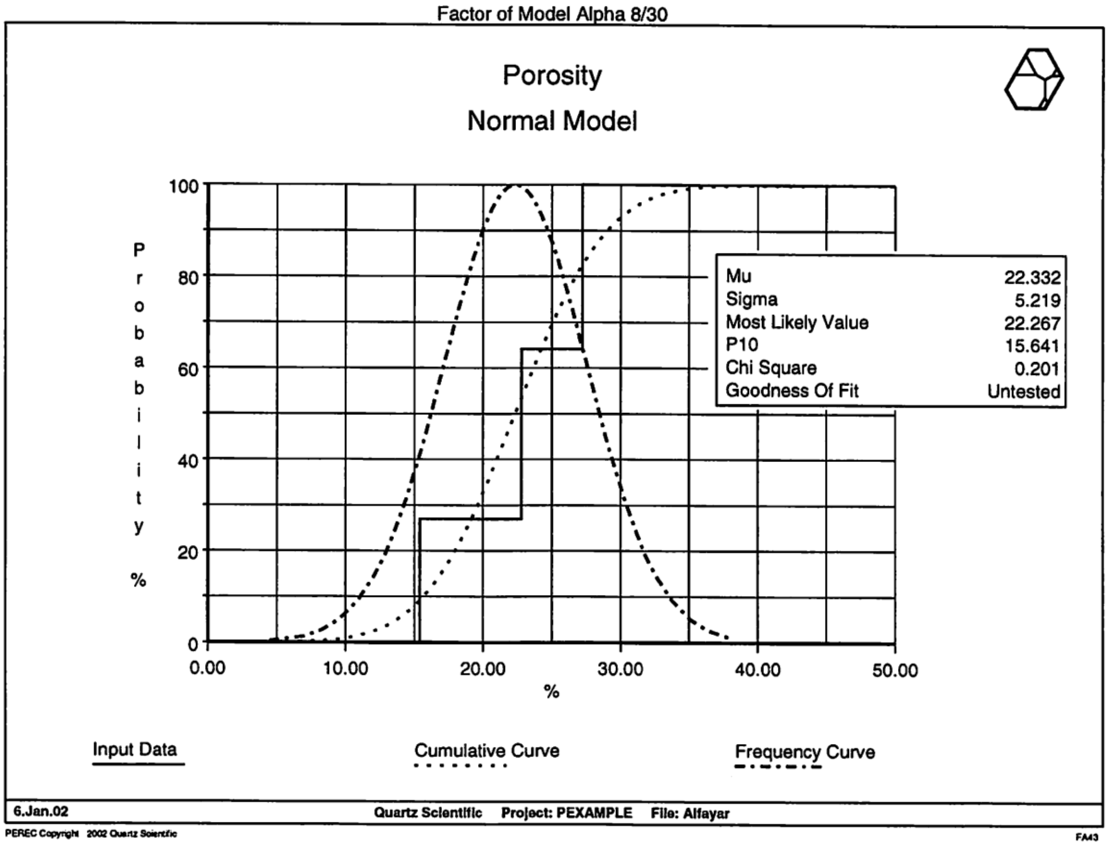
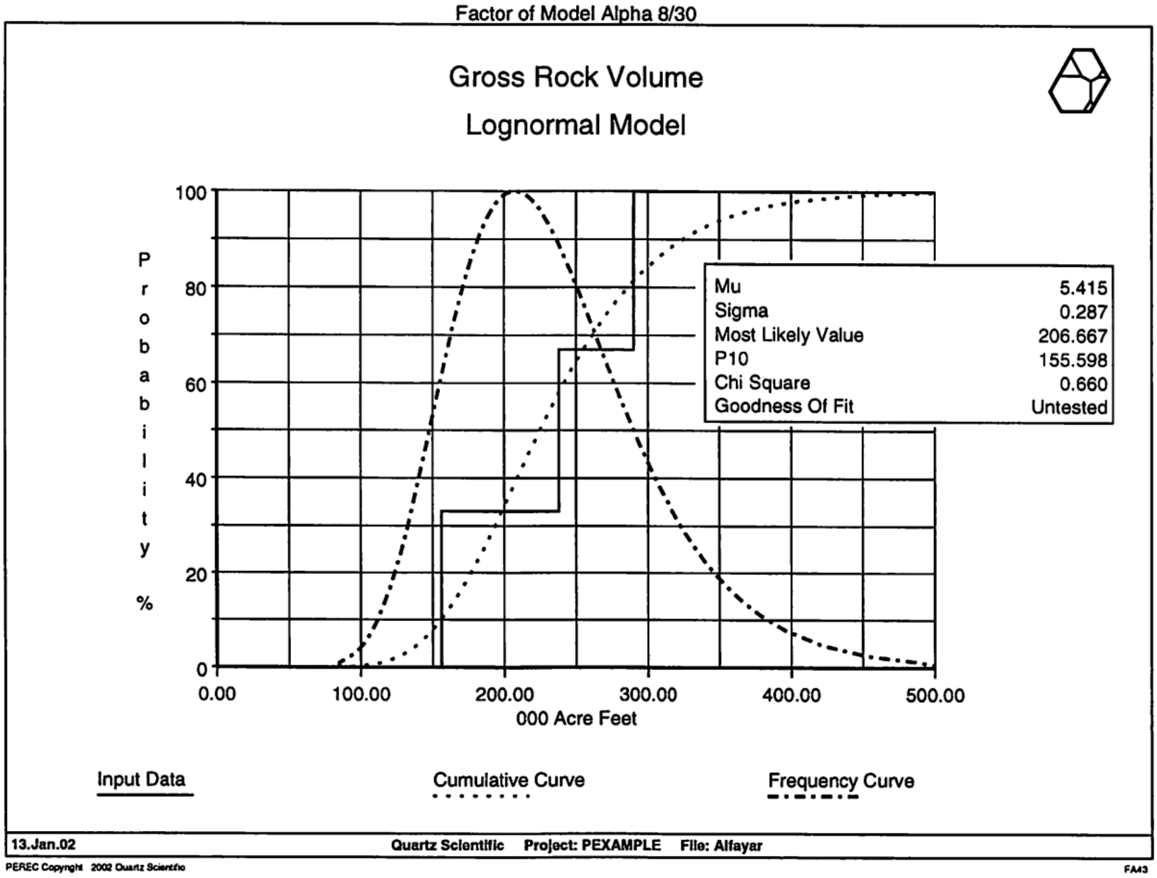


Figure 3.11: Normal histogram fit



Mu	5.415
Sigma	0.287
Most Likely Value	206.667
P10	155.598
Chi Square	0.660
Goodness Of Fit	Untested

Figure 3.12: Lognormal histogram fit

Model N (Normal histogram fit) is based on the principle of the equal area rule; the normal distribution is fitted by the least squares method. Lognormal histogram fit (model L) is obtained by iterative optimisation by the method of steepest descent.

The residual error is calculated as the average difference between the stepped cumulative curve given by the input data and the corresponding values of the fitted distributions. In the plots produced by the factor validation program both the input and the fitted curves are displayed.

The quality of the fit can be judged from the calculated absolute differences, and the chi square measure of fitness. The chi square test value is compared to tables of 95 % confidence level critical values for a given degree of freedom.

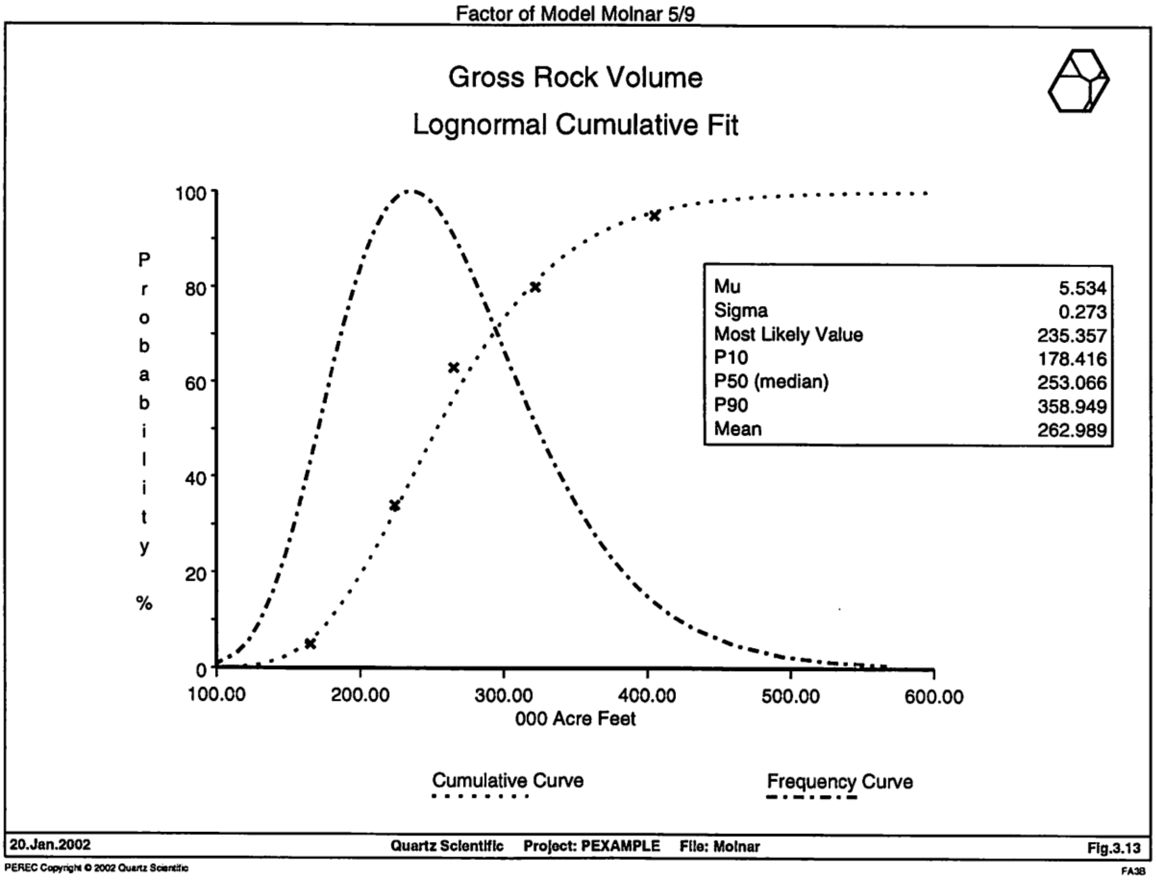


Figure 3.13: Lognormal cumulative fit

The results are expressed as 'adequate' or 'inadequate' goodness of fit. If the number of input values is less than four, the test will not give meaningful results and the program reports 'not tested'.

Lognormal cumulative fits (model M) and Normal cumulative fits (model O) are similarly calculated by the least squares method. Input probability data for these should be given as cumulative percentages (P1 .. P99).

Beta Distribution Fitting

Models Gb, and Pb are fitting a beta distribution to a set of representative values. If either of these model options are used the program will fit the beta distributions that approximate best the values (and their probability weightings) supplied by the user: the two models are differing in the additional constraints applied.

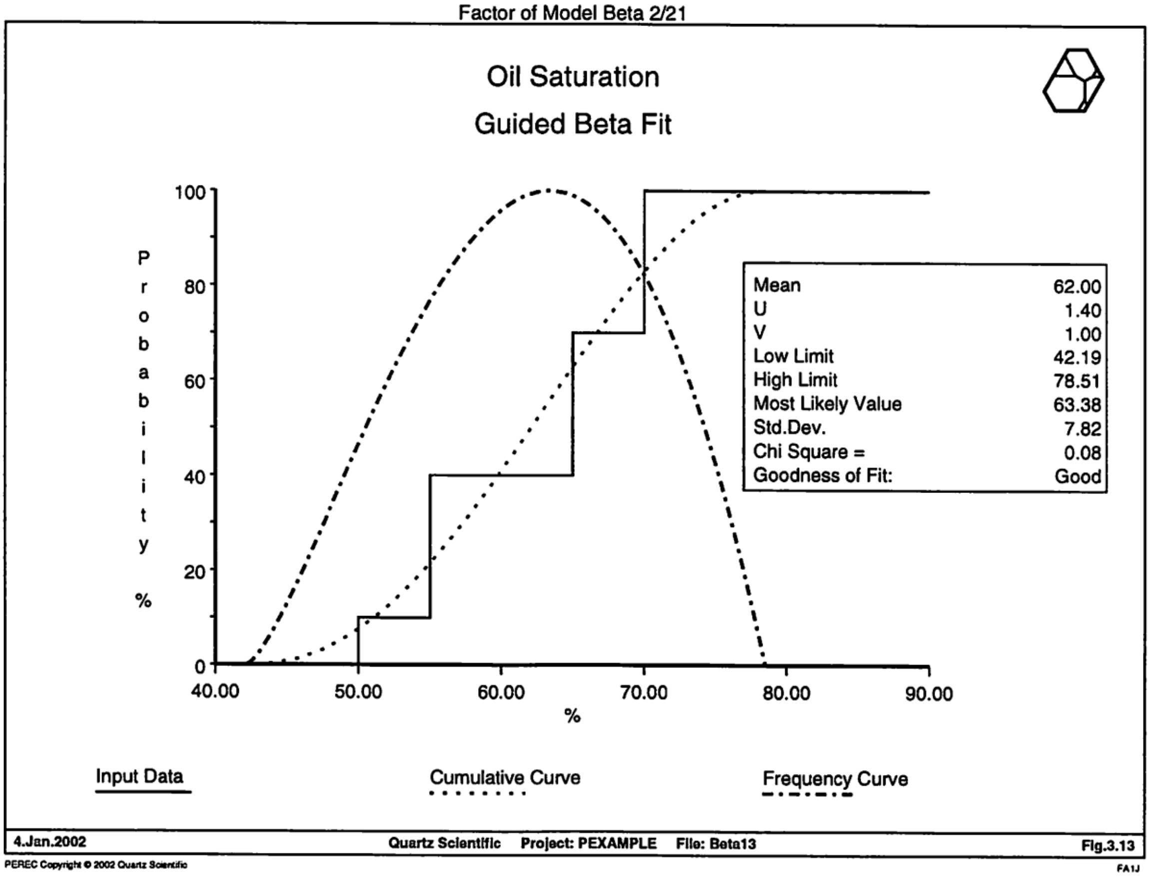
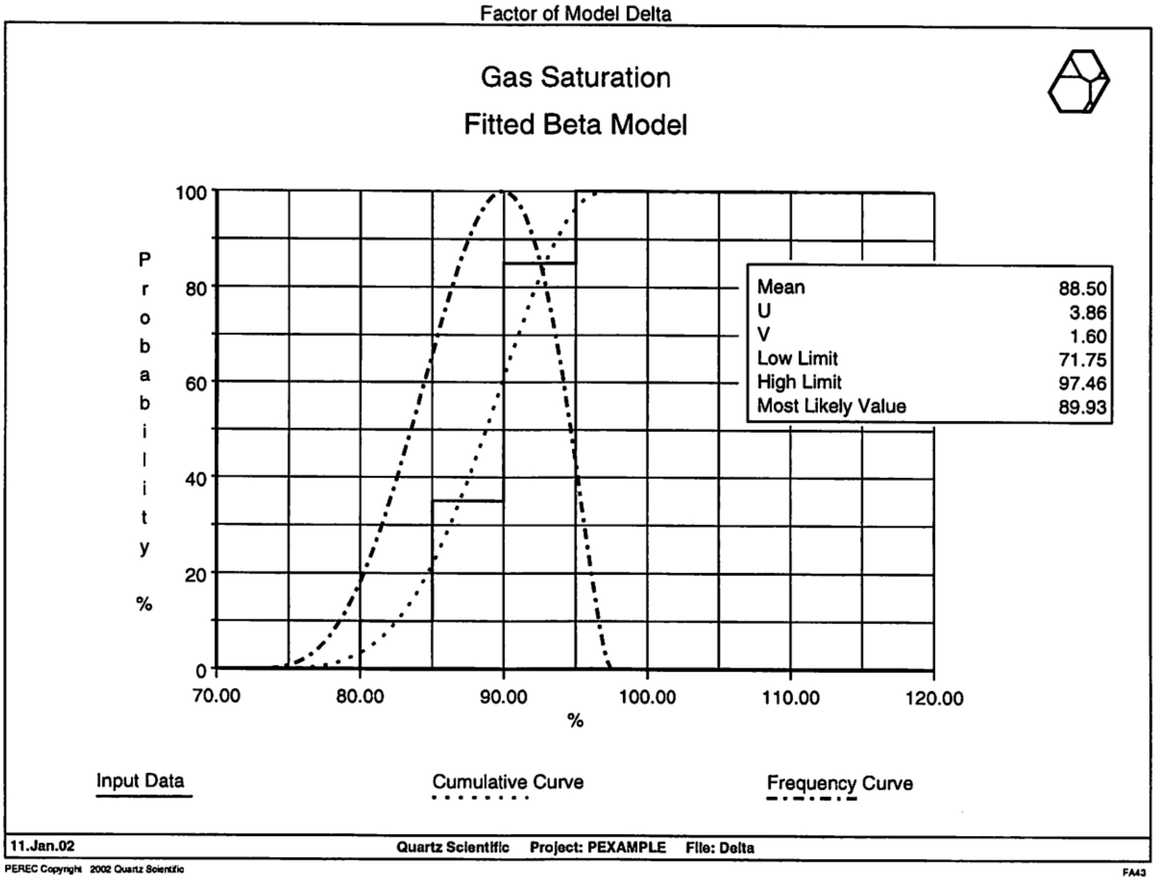


Figure 3.14: Guided beta fit

Guided Beta Fit (Gb) models are fitting a beta distribution that is assumed to be unimodal (single peaked). The distribution parameters are computed from the sample mean, variance and skewness by least squares methods. Restricting the acceptable shapes to unimodal distributions ensures the success of fitting even if the data are ill conditioned. Thus the Gb model procedure is usually the preferred way to fit a beta distribution. The goodness of fit is tested by the usual Chi square test.

If you specify model Pb then the program will try to fit a beta distribution to the data using the Pearson procedure. The program first determines the Pearson distribution class of the data and only proceeds to fit a beta distribution if the Pearson class is appropriate. This means that the Pearson fit (Pb model) does not always succeed to fit a beta model to a given data set.

Another difficulty is that the process may in certain cases lead to contrived curve shapes that do not seem to be natural. It's use is warranted only in special cases, e.g. where the input data have several clearly defined peaks and a multimodal fit is desired.



Mean	88.50
U	3.86
V	1.60
Low Limit	71.75
High Limit	97.46
Most Likely Value	89.93

Figure 3.15: Pearson's fitted beta

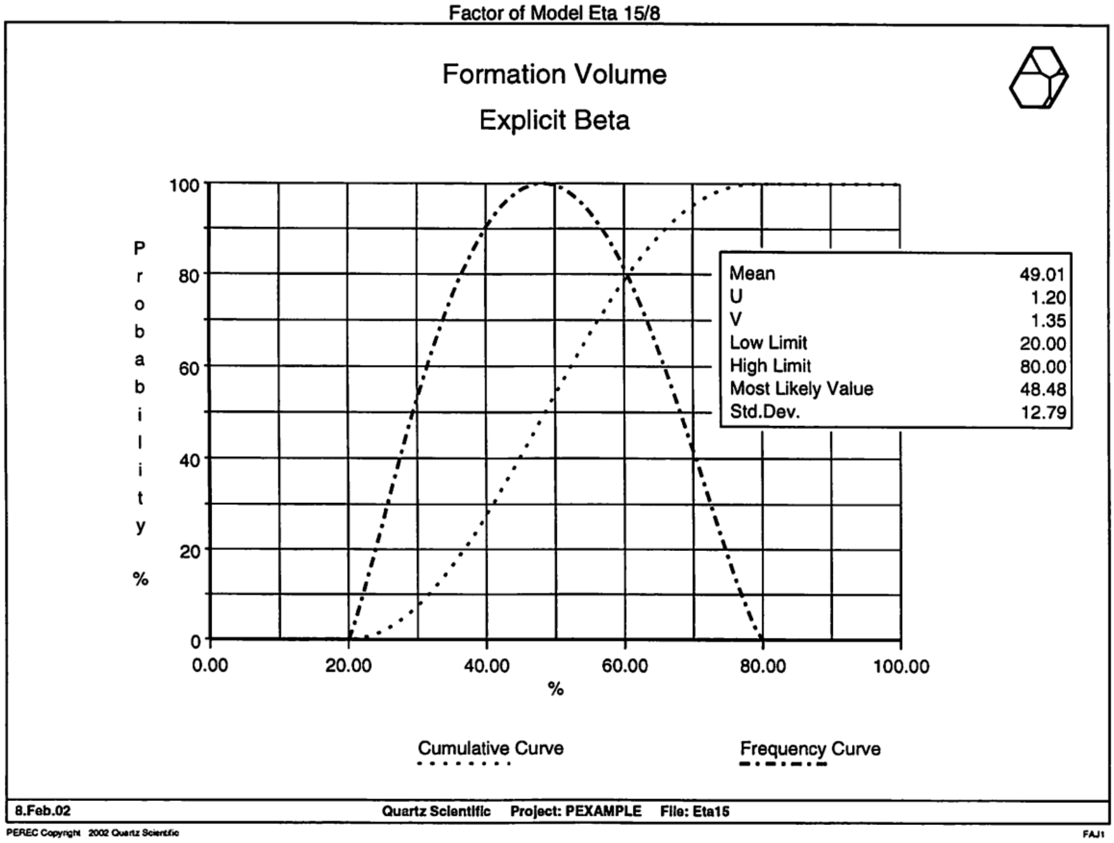


Figure 3.16: Explicit beta model

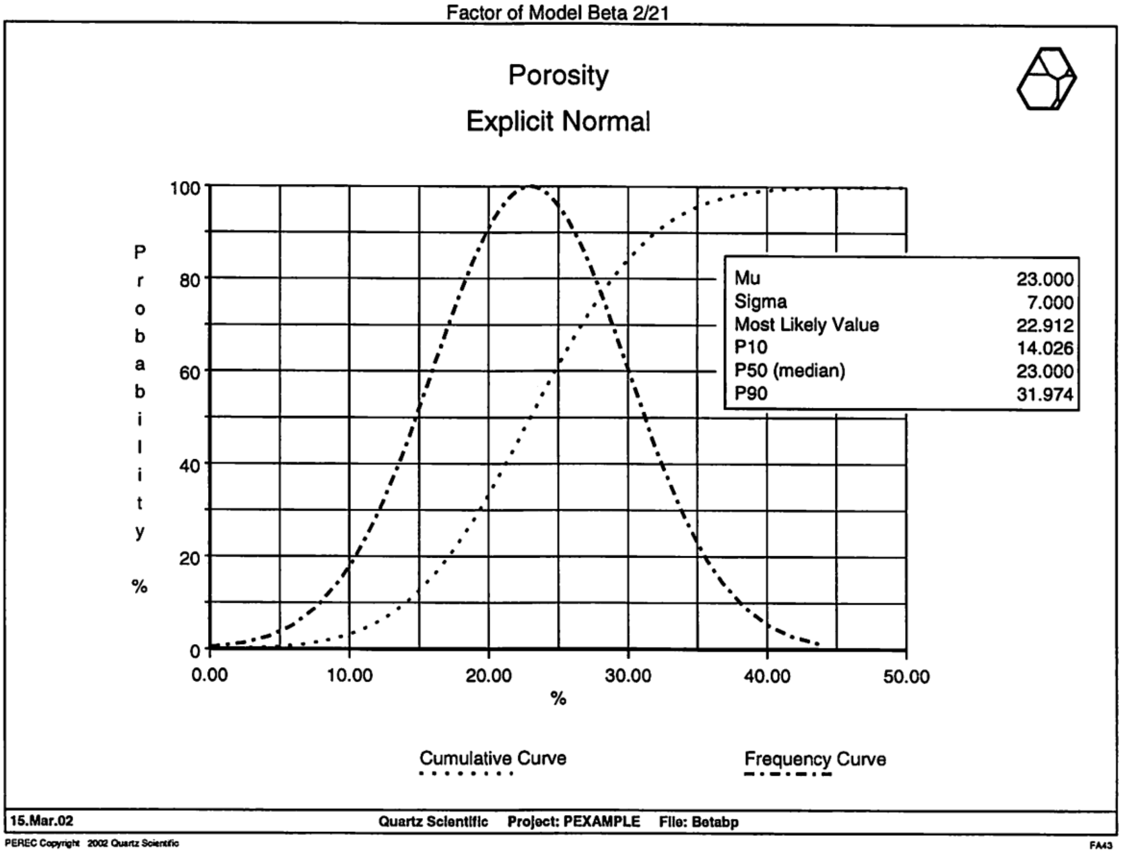


Figure 3.17: Explicit normal model

Explicit models

Factors that have Normal, Lognormal, Weibull or Beta distribution can be modelled directly by entering the distribution parameters, e.g. the mean (Mu) and variance (Sigma). Use model Xn for normal, Y1 for lognormal, Zb for beta, or We for Weibull distribution. Enter Ju or Kb for Johnson's Su and Sb curves, respectively.

For explicit normal models enter the mean and the standard deviation (STD) in the units given.

For explicit lognormal, enter the mean and STD of the (natural) logarithm of the factor values corresponding to the Mu and Sigma parameters shown on summary data sheets.

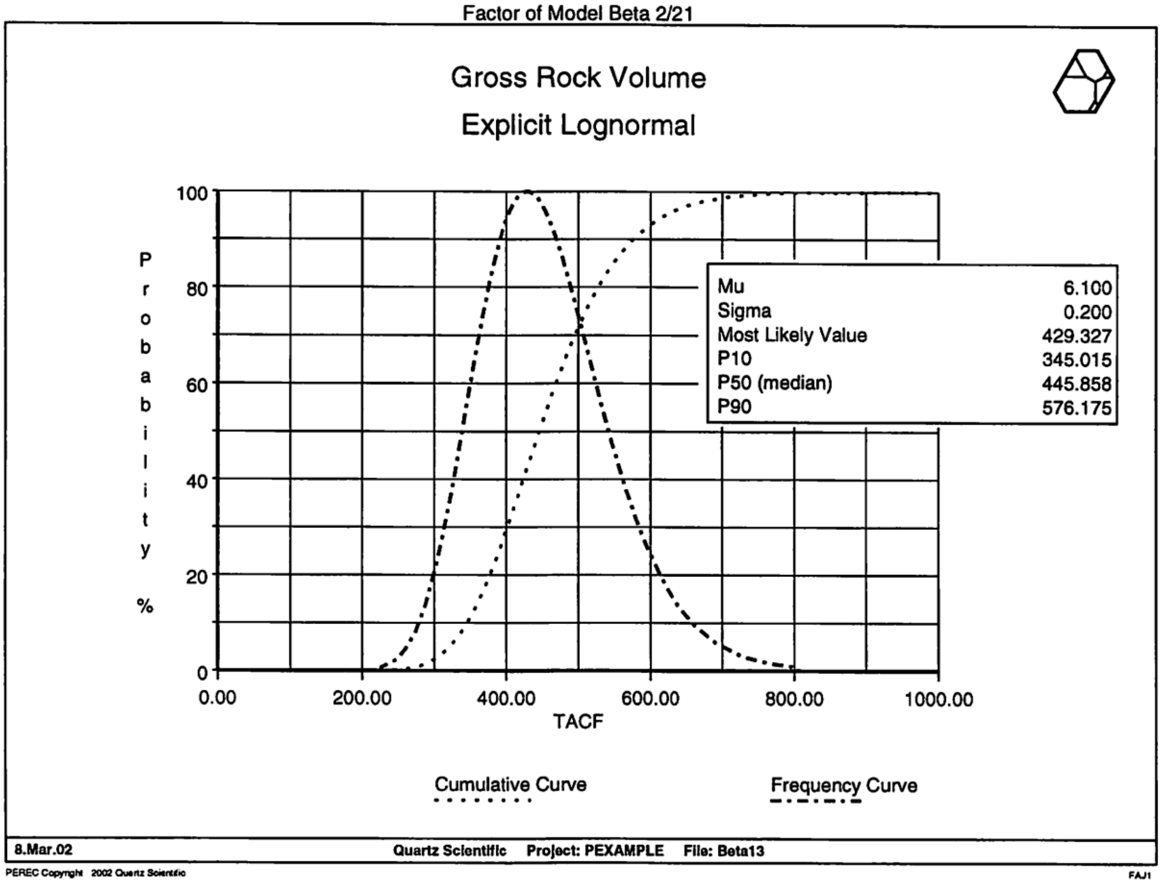


Figure 3.18: Explicit lognormal

For explicit beta, enter the values of the parameters Gamma (minimum value), Delta (maximum value), and the shape parameters (U and V). The U parameter determines the shape of the curve on the left side: small values (0.0 to 1.0) cause the curve to start off from the x axis at a steep angle, whereas high values (3.0 or higher) will make the curve bell shaped - similar to normal or lognormal curves. V determines the shape on the right side. The ratio U/V determines whether the curve is skewed to the left or to the right. If U > V then the most likely value is nearer to the maximum than to the minimum.

For Weibull distribution, enter the values of Rrc and Alpha.

For either kind of Johnson curve, enter the 4 parameters Xi, Lambda, Gamma, Delta, in that order.

Notice, that 'probability' values are not entered: these are not needed for explicit models.

Explicit models can be used to test the effect of varying the parameters of distributions e.g. changing the mean and/or variance obtained by normal or lognormal fits.

Plots of explicitly defined model distributions show no 'input' curve, the factor validation plots display computed cumulative and frequency curves only (See examples).

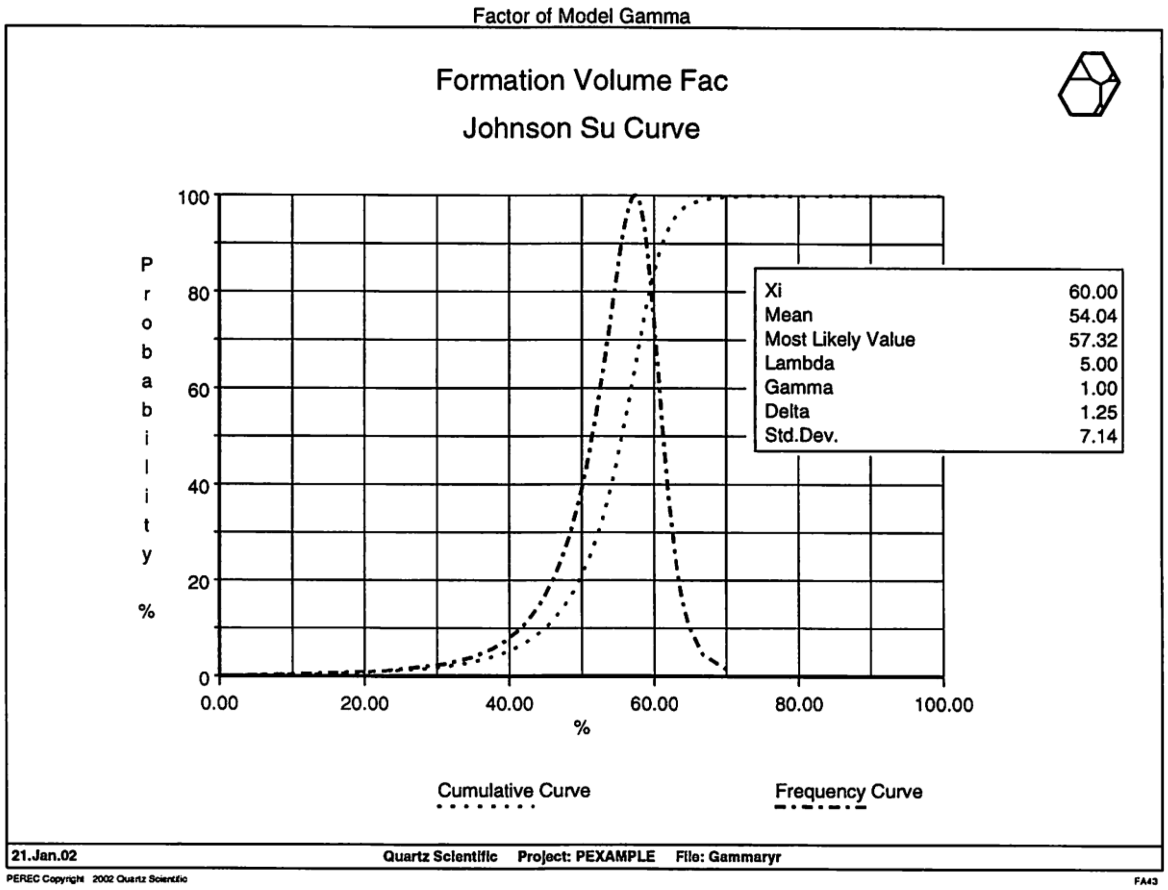


Figure 3.19: Johnson's Su curve

Extreme values

The random values produced for the samples are always within the intended range of acceptable values if the factor distribution is given by discrete, cumulative or frequency models. Most beta models will also produce values within the specified range.

On the other hand, normal and lognormal models, whether fitted or explicitly specified by the user, may pose a problem. The tails of the distribution may yield unacceptable values in certain cases. The factor validation program will check the probability of such excesses and, if necessary, warn the user.

For each factor of normal or lognormal distribution the range of values produced by the model is checked against the limits of minimum and maximum acceptable values.

The limits may be specified by the user in the data file or assumed by the program. The program computes the probability of the factor values falling outside the limits and displays both probabilities i.e. values under minimum, over maximum.

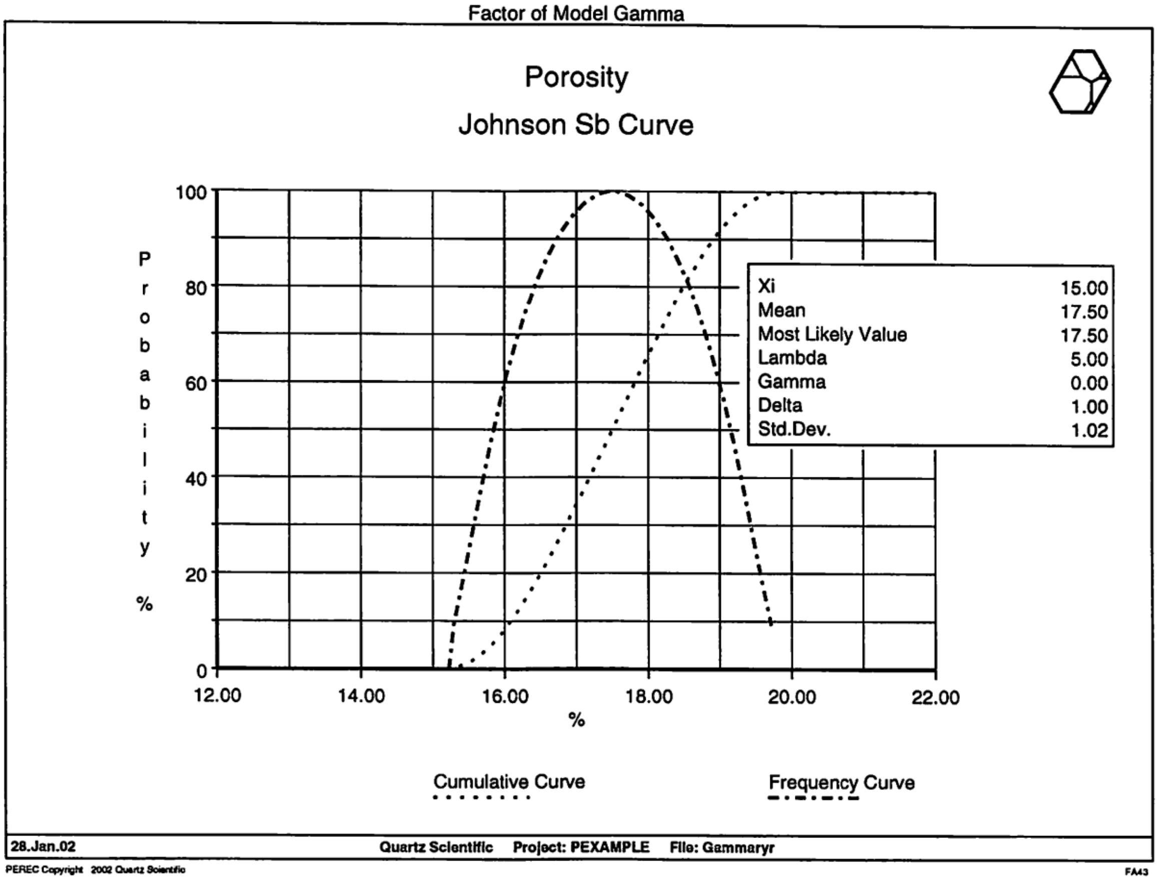


Figure 3.20: Johnson's Sb curve

Levels of severity

A warning is issued if the probability of falling outside of an assumed limit is greater than 1 %.

An error message is sent if the probability of overflow or underflow exceeds 5 % (or 1% if the limit was explicitly set by the user.)

How to specify range limits

If you want to truncate the long tails of a normal and lognormal distribution the range limits can be specified in advance in the data file. For fitted models of types Ln, and No the user can enter the minimum and maximum by entering values with zero probability. Either or both can be entered.

If the factor model is explicit normal/lognormal (types Xn and Yl) the data entry program allows the entry of minimum and maximum limiting values. If explicit limits are required then enter the limiting values in the boxes under the values of Mu and Sigma.

Assumed limits

In the absence of a specified minimum value a lower limit of zero is assumed for all factors as values must be positive at all times. If no upper limit is given, the program is not checking for a maximum value, except if the factor has percentages as values. For percentages, an upper limit of 100 % is assumed.

Action taken

If factor validation indicates a problem, the user may choose to ignore it and let the simulation program to deal with it. Alternatively, the user can change the data file. If you decide to correct the factor shape you can either use the same model type with adjusted parameter values or, change the model type. E.g. take a plot of the factor distribution curve. Redraw the frequency and enter the values read off the plot as a frequency model.

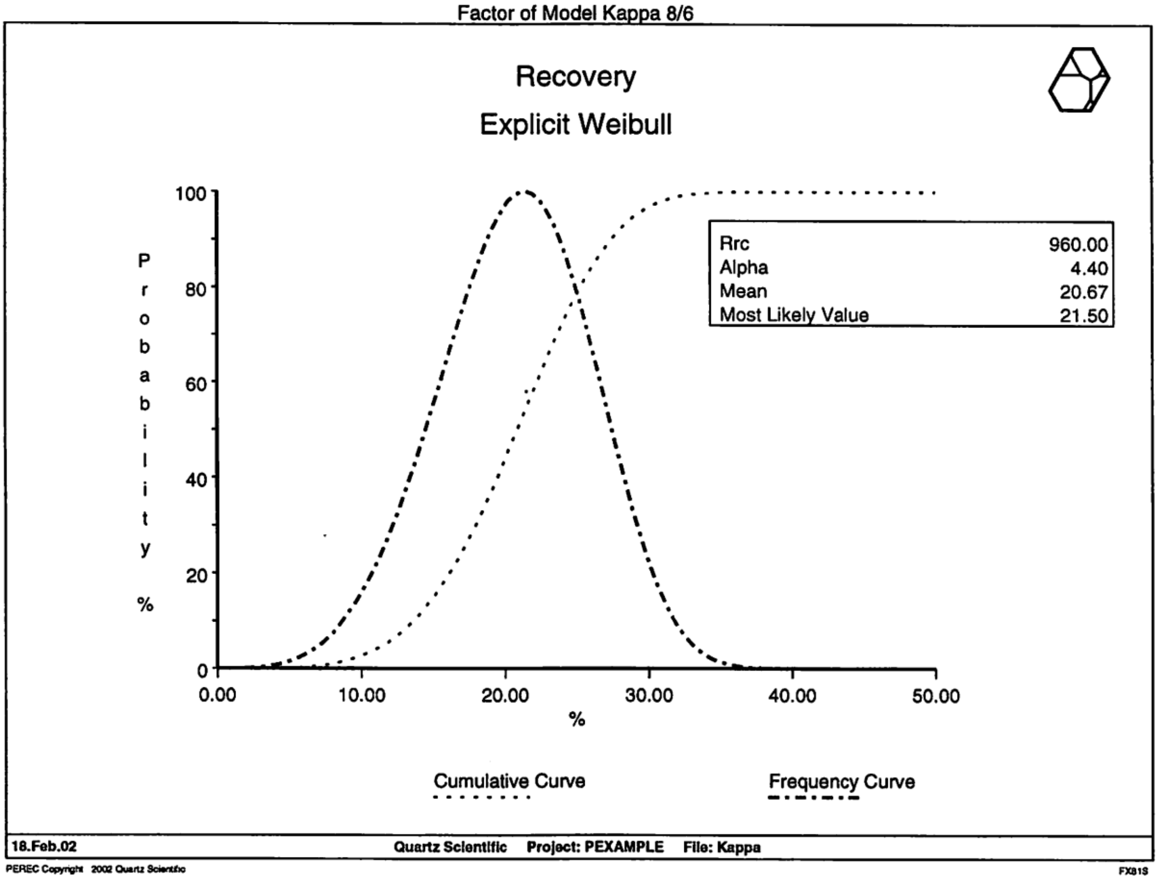


Figure 3.21: Weibull model

Truncated model

If a normal or lognormal fitted model was specified and validation indicates an error condition, then a truncated model is fitted. Truncated normal and lognormal distribution models are normal/lognormal curves cut at the low, high or both tails. Such models have 2 further parameters for low and high limits in addition to mean and variance.

Given the low and high limits the program finds the best fitting values of mean and variance. These are computed in a way similar to those for the unconstrained models, e.g. optimising by the 'equal area' rule. The example plot shows a truncated normal distribution.

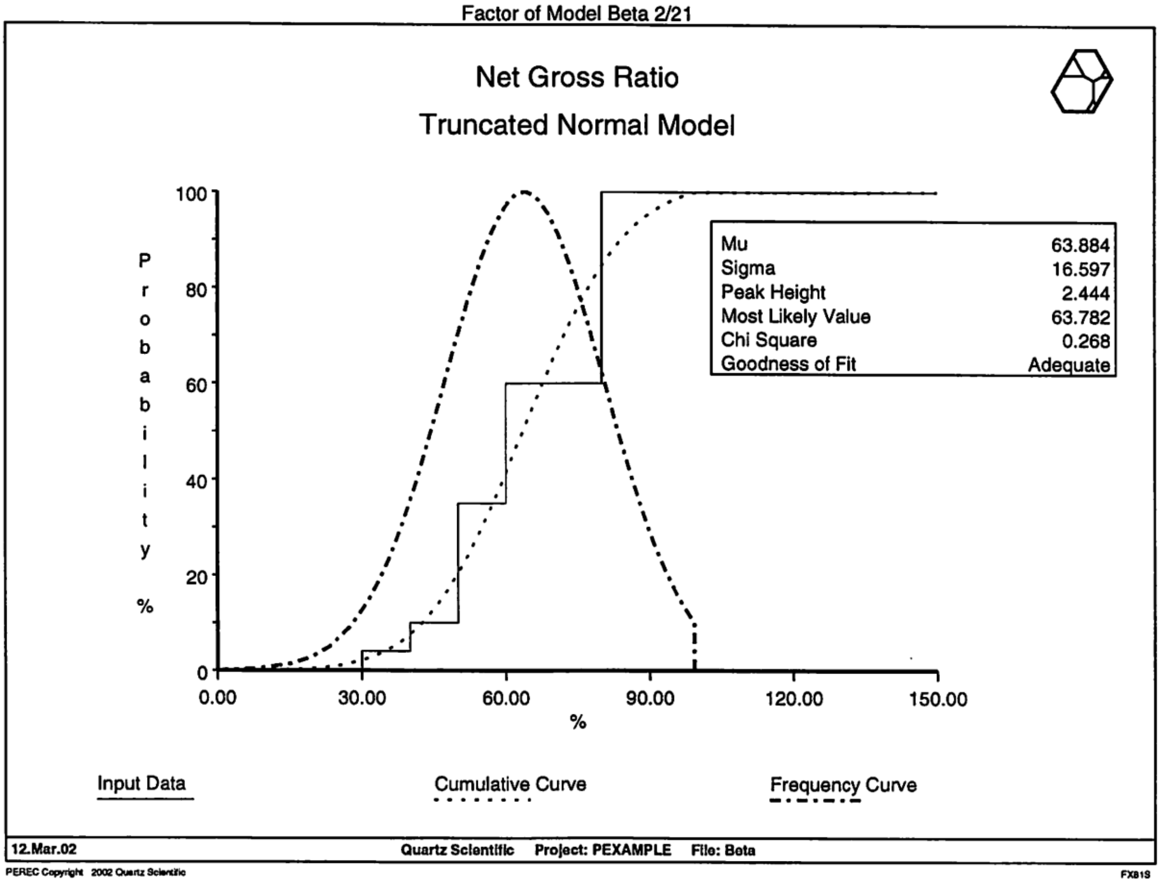


Figure 3.22: Truncated normal model

The effect of dependencies

The prospect evaluation program provides for a degree of dependency between geological factors. The level and nature are described by a single number on a scale between -10 and +10.

The + or - sign indicates positive or negative correspondence with a strength of magnitude 1 to 10, where 10 is the strongest level of correlation. A value of 0, indicating independence, is the usually assumed default.

The treatment of dependency by Monte-Carlo simulation is based on the statistical model of correlated, linearly dependent multivariate data. The joint distribution of two linearly correlated variables is ellipsoid with a sloping main axis that can be determined by regression or factor analysis. The unconditional distribution of the variables can be derived from the joint distribution.

PEREC addresses the inverse problem: given the unconditional distribution of each variable, it generates random samples for the variables with the desired level of correlation.

In the first instance, let us consider two correlated variables. These may be factors to be multiplied in calculating single reservoir expectations or two reservoirs to be added by the Monte-Carlo prospect addition program.

Input to modelling consists of the cumulative distribution curve (any given distribution can be expressed in cumulative form).

Dependency between variables is given as number, representing, in statistical terms, the correlation coefficient of the square transform of the variables.

Correlated pairs of random values are generated as follows: a random number is generated between 0 and 100 % and projected to the cumulative curve of the first variable. This gives the value for the first variable. Then, another random number between 0 and 100 % is generated to obtain a value for the second variable.

Now, if the two variables have no dependency then their correlation coefficient is zero. Random numbers are then generated independently if this is the case and the projected values are not related.

On the other hand, if variable h is dependent on variable *a*, then the choice of the first random number *r1* restricts the choice of *r2*. This is achieved by generating *r1* and *r2* as independent random numbers some of the time and letting *r2* = *r1* at other times.

If negative dependency (correlation) is desired, then the non-random choices of *r*2 will he set to *r*2 = 1 –*r1*.

For example, let the level of dependency between *a* and *b* be minus -30%. Then for each pair of random values *v1* and *v2* three random numbers *r0, r1, r2* are generated in the range of 0 to 100%. The random number *r0* determines whether *r2* is generated independently (70% of the time) or set to equal 1-*r1* (which happens 30% of the time).

The projection of *r1* to curve 1 gives *v1* in the range of variable *a*, *r*2 projected to curve 2 gives *v*2 in the range of variable *b*.

The pairs *(r1, r2)* will he linearly correlated with a correlation coefficient exactly matching the dependency specified. The pairs *(v1, v2)* show a correlation that is a function of the unconditional distribution functions. It will be exact for dependencies of -10, 0, +10 but the intermediate values may produce correlation coefficients differing from the dependency parameter in the same way as the covariance of the bivariate distribution differs from the covariance of its square transform.

Exact correspondence in the above context means that the expected value of the correlation coefficient is equal to the specified dependence.

If the prospect has more than two reservoir properties (geological factors) that are interdependent, data selection is performed using the same principles as for dependency with two variables.

Account is taken of the fact that a particular variable may become indirectly related to other properties as a result of interdependence within the group of variables. E.g. if *B* depends on *A* and *C* is specified as dependent on *B* then this determines a dependency between *A* and *C*. So the program will not allow a further specification for dependency between *C* and *A* to be entered.

Factor display

The validated factor distributions can he displayed in the form of a graph by the Factor display option on the Modelling menu. When you select this option the program asks for the name of a data file and then presents the list of factors in the file as a menu. The factors can he plotted by typing the number on the list.

The default horizontal scale for factor distribution plots is displayed by the program: This can be changed by specifying the desired range when prompted by the program. After displaying the graph the program pauses. Press any key to return to the list of factors. Note that to plot a factor you have to return to the Reservoir Modelling menu and select the Hardcopy option which will produce a paper plot of the last factor that was displayed on screen.

Curves drawn

The plots show the factor distribution as three superimposed curves. The stepped graph of input data is shown by a continuous line, while the fitted distribution is shown by heavier dashed (cumulative curve) or dotted lines (frequency curve). The curves are all scaled to 100 % vertically - this means that the frequency curve is multiplied by a scaling factor.

On colour screens and printers each curve is shown in a different colour.

If you tick the 'Box around values' option (in Options/Preferences) then the program will draw a rectangular box around the tables/lists of numbers.
(Perec for Windows User Guide, Quartz Scientific, Watford, 2002)