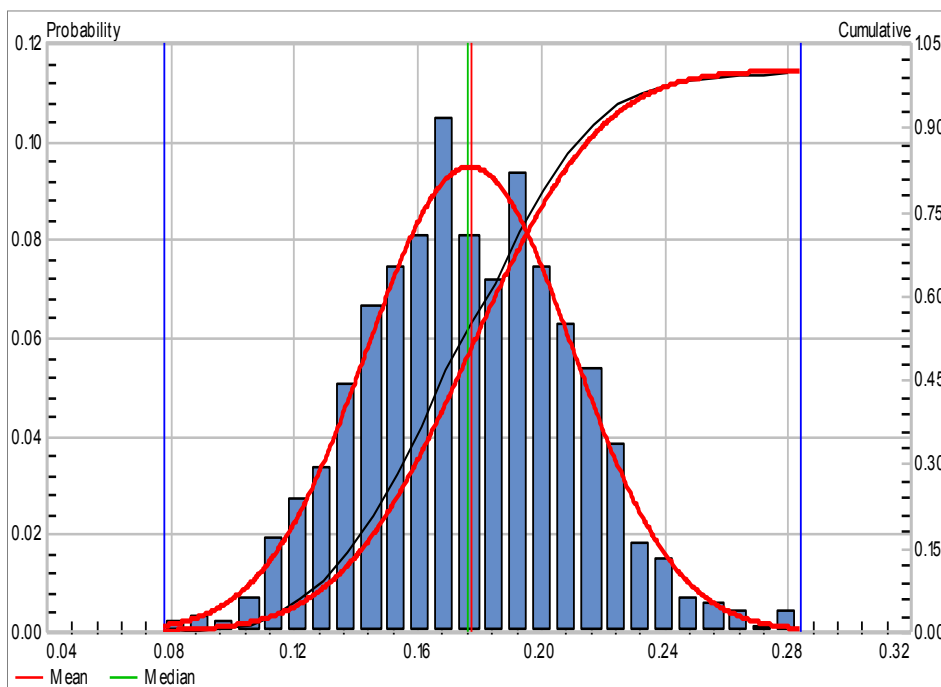


STATISTICAL TOOLS

A Monte Carlo simulation is used to replace a deterministic approach with a stochastic approach, introducing the degree of uncertainty needed for the problem to be solved.

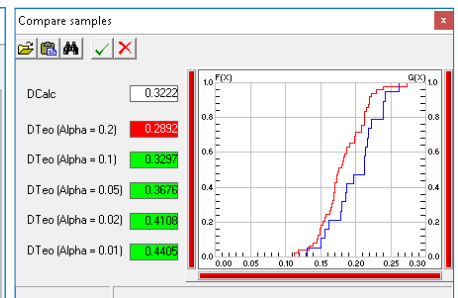
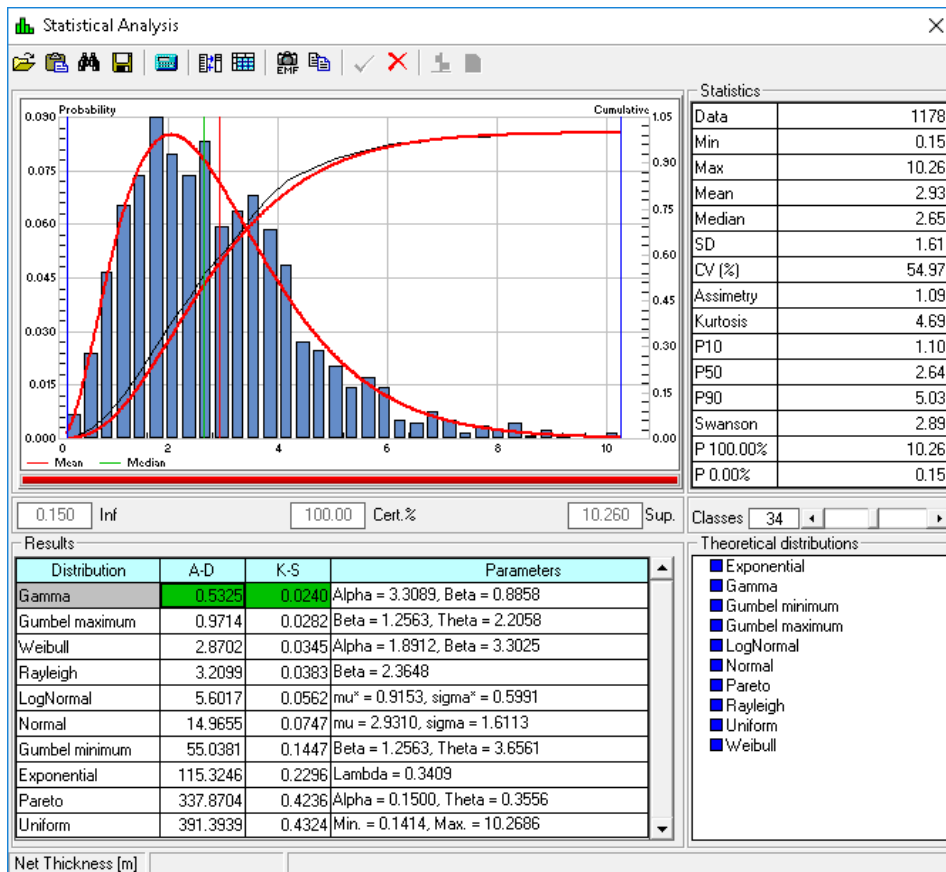
Sahara has an extensive list of known continuous and discrete probability distributions that can be used to define the hypotheses of the variables involved in a Monte Carlo simulation. Knowing the characteristics of probability density functions, the role their parameters play in their changes of form and scale, and their most frequent applications are paramount to ensure the quality of the results. For this, Sahara provides the possibility of performing a statistical analysis on the values of multiple variables.

The Statistical Analysis module allows representing the distribution of the data of a particular variable and comparing it with the different theoretical distributions available. For the best fit distribution, the characteristic parameters can be obtained and saved to be reviewed later or used as hypothesis in a Monte Carlo simulation.



Statistical analysis and Monte Carlo simulations can be performed using data from a Sahara project or with any external data.

Sahara



Comparing Populations. The chart shows the comparison of two samples. In this example, there is not enough evidence to infer that the two samples derive from different populations for the trust levels corresponding to $\alpha = 0.10, 0.05, 0.02$ y 0.01 , but there is for $\alpha = 0.20$.

Statistical Analysis. Example that studies the net thickness of a particular layer for a group of wells. It can be seen that all the theoretical distributions have been selected to carry out the study and among all of them, the one that best represents the data, according to the Goodness of Fit tests, is a Gamma distribution with $\alpha = 3.30$ and $\beta = 0.89$. The representation of the Gamma distribution and its cumulative production is observed on the histogram.

The Statistical Analysis module helps to study distribution of data

The Statistical Analysis window allows analysis for variables in a Sahara project or for external data. Additionally, it can be used to analyze the results obtained from a Monte Carlo simulation.

Statistical data related to the sample can be displayed in this window, such as their maximum, minimum, average, median and standard deviation values and values associated with the 10, 50 and 90 percentiles. The histogram can be modified using the number of classes that are needed for optimal visualization. In addition, sliding the two vertical bars the user can set a range of values and evaluate the percentage of certainty that a value within the sample falls within the established limits.

The window shows a list with the continuous and discrete theoretical distributions available. They can be selected with the purpose of trying to adjust the sample data and find the distribution and parameters that achieve the best fit. To estimate the sample parameters for each of them, Sahara uses the methods of maximum likelihood estimation and the method of moments, as the case may be. Then, to check if the assumptions about a set of data correspond to a known probability distribution, two Goodness of Fit tests are used: Anderson-Darling and Kolmogorov-Smirnov. These tests pose

a null hypothesis which, with a certain degree of significance, is accepted or rejected. If the null hypothesis is rejected, then data is not distributed according to the proposed distribution. If, on the other hand, it is accepted, the data set is considered consistent with the proposed distribution and the degree of agreement between the distribution of the data set and a given theoretical distribution is measured.

After performing the calculation, the selected distributions are ranked according to the degree of adjustment, showing in green the adjustment parameters that represent the best one. These results can be saved to be later studied or used as hypothesis in a Monte Carlo simulation.

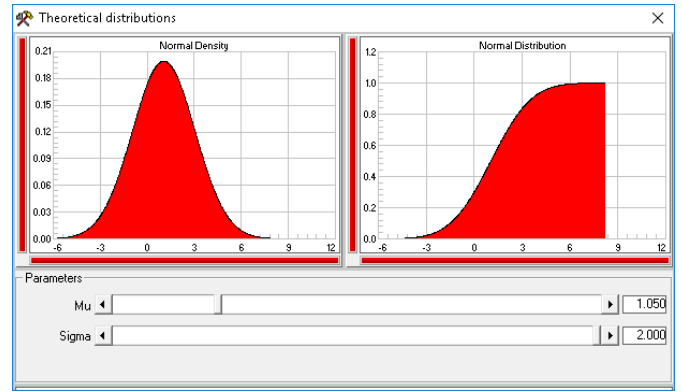
This module also allows for a population comparison. This procedure is used to discern, based on statistical criteria, if two independent samples belong to the same population. This fact is very important to ensure the quality of data used to build the series of historic data of a variable to be used in a Monte Carlo simulation. In this way, more realistic results can be obtained. The method used by Sahara to perform this procedure is the Kolmogorov-Smirnov test for two samples. It is based on the comparison of the sampling cumulative distribution functions, and no assumption is required on the distributions of the compared data.

The purpose of a Monte Carlo simulation is to obtain a stochastic result for any problem set out. To accomplish this, the input data of the model must consist of probability distributions rather than given numbers. The allocation of the distribution to each variable is one of the most complex and important steps of the process, since, if the hypotheses are wrong, they will lead to results that are far from reality.

When the distributions to be used as a hypothesis for each variable must be determined, Sahara has a guide that will help acknowledge the characteristics of each theoretical distribution and the impact of the variation of its parameters. This functionality is of great help in defining the hypotheses. For each of the continuous and discrete distributions available in Sahara, the Theoretical Distributions window provides:

- Chart for the density and cumulative distribution function.
- Sliders to vary the parameters that define the distribution. This option aims to qualitatively visualize the effect of each parameter on the form and scaling of the density and cumulative distribution functions. The range of values for each parameter is arbitrarily delimited by predetermined values.

Monte Carlo technique consists of establishing a model or equation, assigning the corresponding probability distribution to the each input variable, and performing the calculation multiple times. As an additional parameter, a correlation matrix among the variables can be used. In this way, if there is a relationship between some of them, it can be taken



Parameters of Theoretical Distributions. The figure shows a Normal distribution with parameters $\mu = 1.050$ and $\sigma = 2.000$.

into account when generating the sample values. In each run, values for each of the input variables are randomly generated according to the assigned probability distribution. To ensure the results consistency, it is recommended to perform a large number of runs. In a matter of seconds, Sahara will perform the calculations for thousands of runs to produce a result in the Statistical Analysis window. For the results, statistical data can be observed and their distribution compared with the theoretical distributions available in Sahara.

The Latin Hypercube option is another alternative when performing the data sampling. This method of stratified sampling segments the cumulative probability distribution and generates random numbers within each segment, without repeating the sampling in any of them, reducing the number of runs needed to obtain a valid result.

Also, a tornado type sensitivity chart can be visualized, in which the weight of each of the variables involved on the final result can be shown.

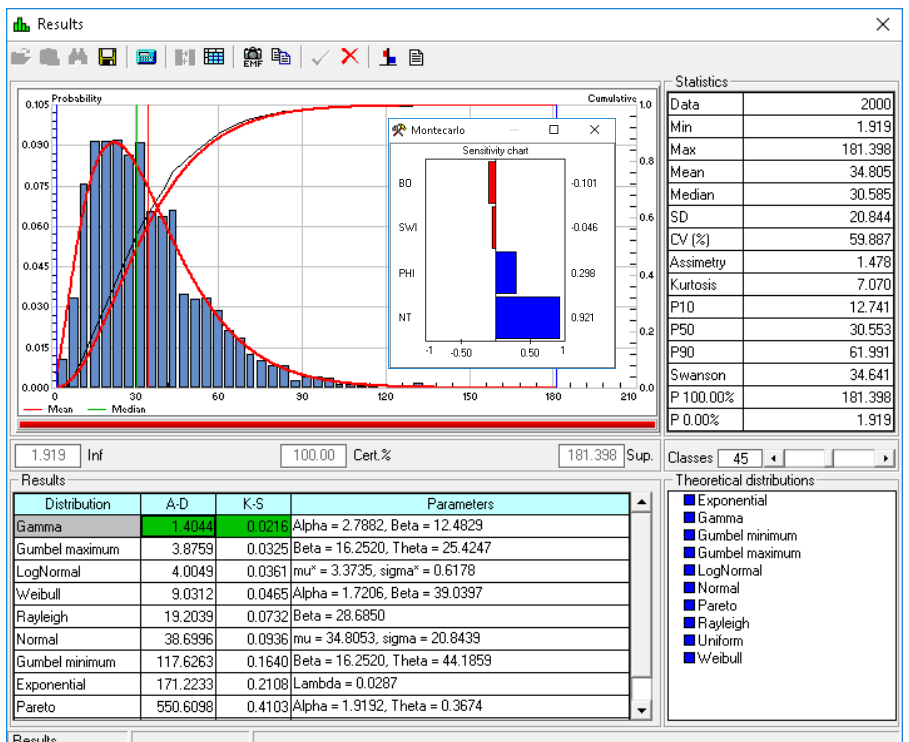
The image shows the "Monte Carlo" setup window. It contains a table for defining variables and their distributions, and a "Correlation matrix" table below it.

Var.	Distribution	Parameters
FI	Normal	Mu = 0.1773, Sigma = 0.0335
HK	Gamma	Alpha = 3.30568, Beta = 0.88658
AREA	LogNormal	P10 = 1540000, P90 = 9900000
Swi	Triangular	P2 = 0.19, Most likely = 0.22, P98 = 0.25
BOI	Uniform	Minimum = 1.05, Maximum = 1.25

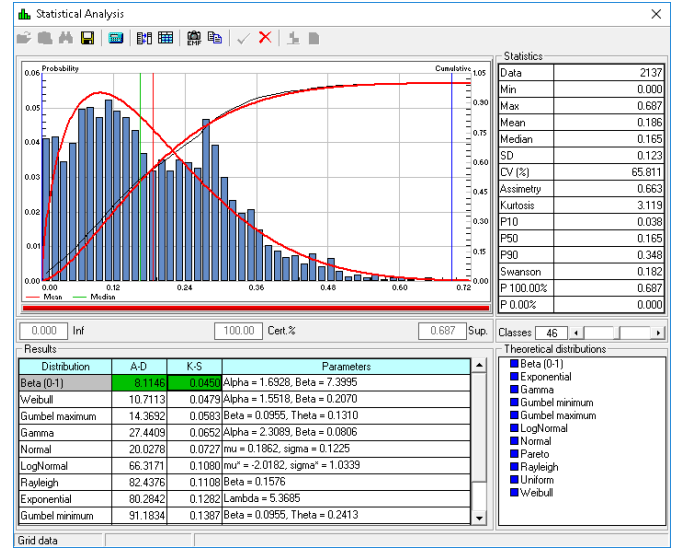
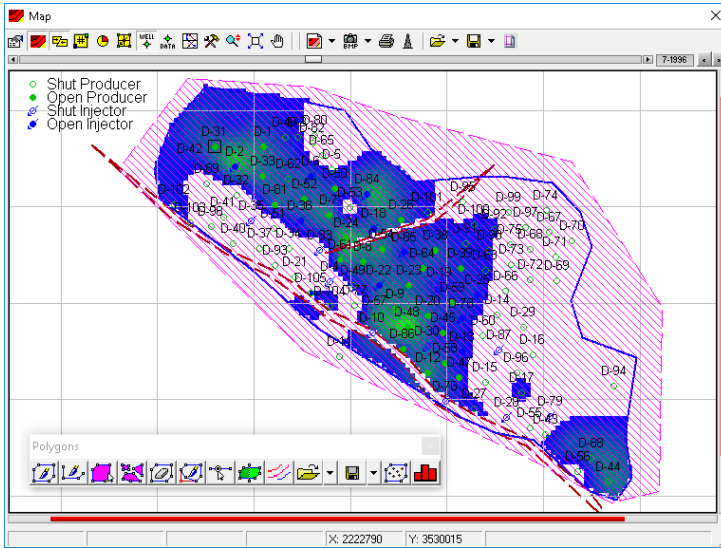
	FI	HK	AREA	Swi	BOI
FI	1	0.5	0.000	0.25	0.000
HK		1	0.000	0.000	0.000
AREA			1	0.000	0.000
Swi				1	0.000
BOI					1

Monte Carlo. Hypothesis definition for the variables of a Monte Carlo simulation with the purpose of analyzing the volume of hydrocarbons in situ. A correlation matrix of the variables is also displayed.

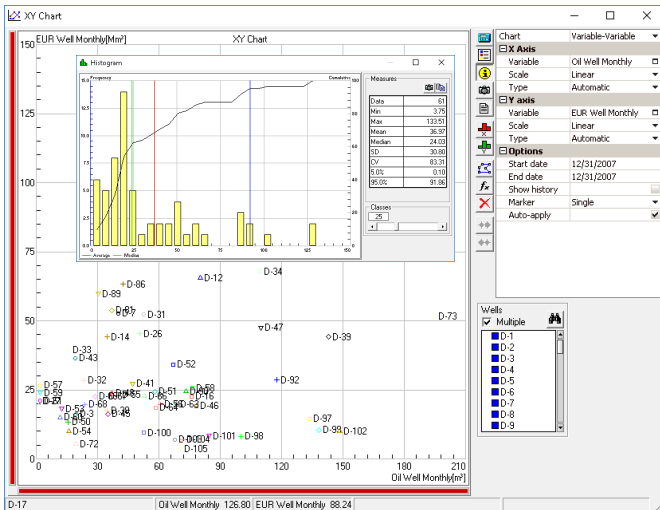
Results: The results of the simulation are shown. Of all available distributions, according to the Goodness of Fit tests, the OOIP volume has a Gamma distribution with $\alpha=2.7882$ and $\beta=12.4829$. The representation of the Gamma distribution and its cumulative production is observed on the histogram. There is also a tornado chart indicating that the variables with greater weight on the result are thickness and porosity.



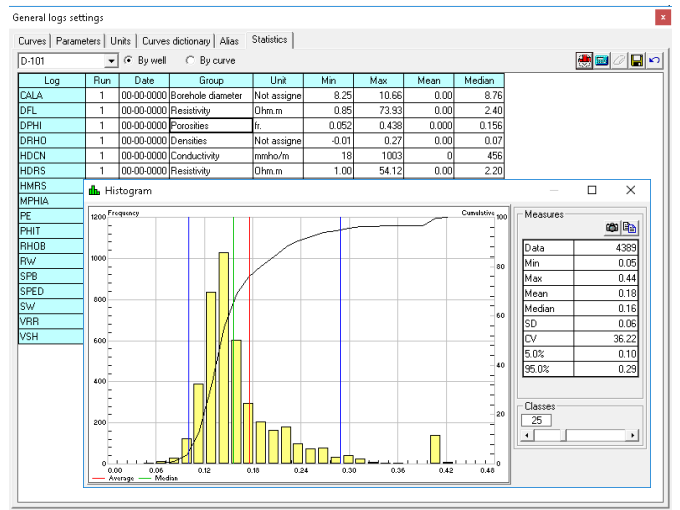
An additional feature of the Statistical Analysis window is that it can be called from many other Sahara tools, allowing analysis for multiple types of data.



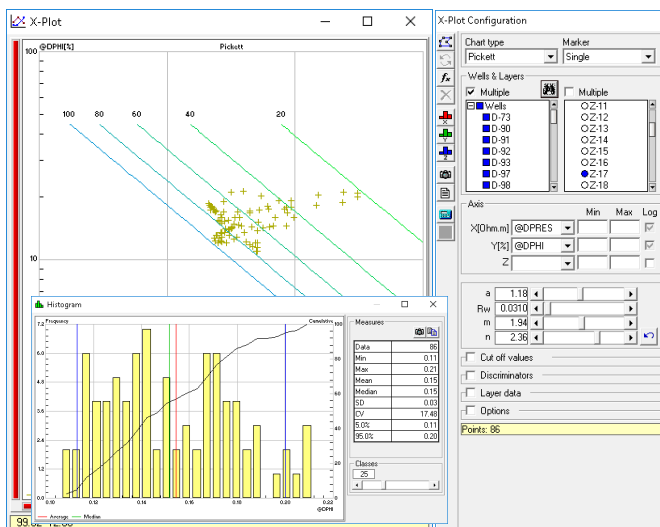
Maps: By making a polygon in the Map window, the histogram of the values of the cells in the map being displayed can be analyzed.



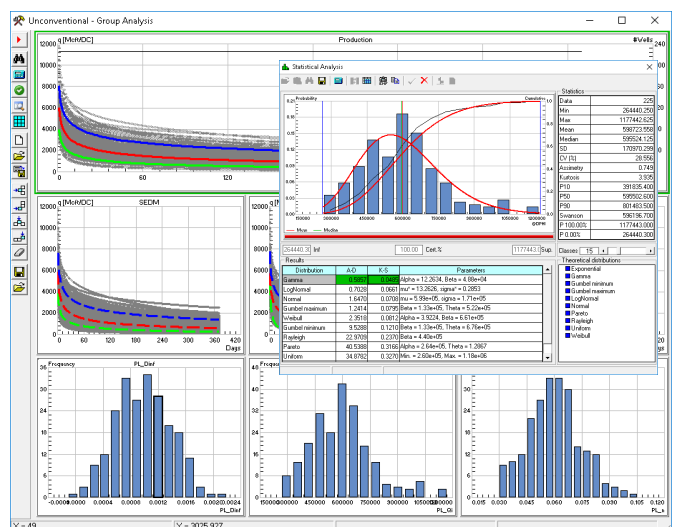
XY Charts: The histogram corresponding to the values plotted in any of the axes of the graphs in the window can be displayed.



Log Settings: For each of the well logs, the corresponding histogram and the statistical parameters of the distribution can be observed. It detects anomalous points.



Cross-Plot: The histogram corresponding to the values of the logs plotted in any of the axes of the graphs in the window can be displayed.



Unconventional - Group Analysis: This module has specific charts to perform statistical analysis, from where the Statistical Analysis window can be accessed.

