Simulation

Simulation

Menu:	QC.Expert	Simulation	Data simulation

This module generates pseudorandom numbers with given statistical properties. Four distributions are available for random sample generation: normal, lognormal, uniform and lambda. Different parameters (mean, variance, skewness) determine the distribution within the four types. Both independent and correlated data (with a given first order autocorrelation) can be simulated. Generated data are useful when simulating real processes or when exploring properties of various statistical procedures before applying them on real data. Uniformly distributed simulated data can be used to construct table of random numbers for acceptance sampling.

Parameters

The procedure does not require any data. Parameters of various distributions are entered in the *Simulation* dialog panel. The following table shows which parameters can be specified for each of the four distributions available:

Table 1 Parameters to be selected

	Normal	Lognormal	Uniform	Lambda
Mean	✓	✓	✓	✓
Standard deviation	✓ positive	✓ positive	✓ positive	✓ positive
Skewness	×	×	✓	✓
Threshold	×	✓	×	×
Autocorrelation	✓-1 to 1	✓-1 to 1	✓-1 to 1	✓-1 to 1
Outliers relative frequency	✓0 to 1	✓0 to 1	✓0 to 1	✓0 to 1

× - parameter cannot be freely specified
✓- parameter can be specified, range of admissible values given

Generated data are (pseudo)random and so their sample characteristics follow the specified parameters only approximately. Nonzero skewness selected for the uniform distribution causes that data from a parallelogram shaped distribution are generated. The standard deviation parameter entered for the uniform distribution is interpreted as the range. Nonzero autocorrelation changes the resulting distributional parameters (parameter specification to the independent random variables generated before transformation to autocorrelated data), e.g. standard deviation is inflated.

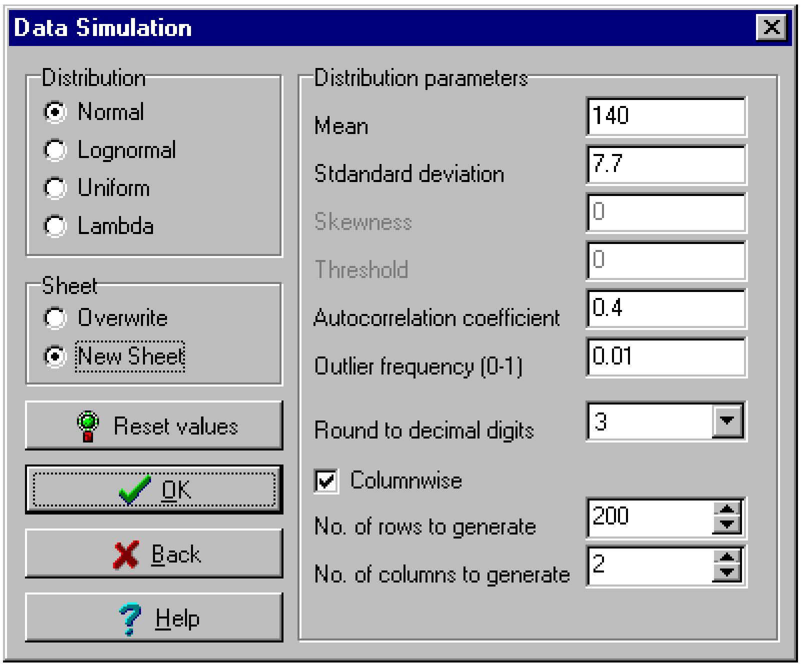


Fig. 1 Simulation dialog panel

Error propagation

Menu:	QC.Expert	Simulation	Error propagation

This module is useful when exploring statistical behavior of a variable which is not measured but computed from measured variables or from variables with a known statistical behavior. Using the module, range of the resulting variable, its mean and confidence interval can be estimated. The individual measured variables contributions to the resulting variable variability can be determined as well. The approach is sometimes called uncertainty evaluation or error propagation. Resulting variable properties are evaluated by the Monte Carlo simulation and by the second order Taylor expansion.

Data and parameters

The input random variables and the function giving the resulting variable must be entered. The input variables can be given by:

a) specifying mean and standard deviation. The program generates normal random variable with the specified parameters. Such a generated random variable is denoted by X or x.

b) entering a sample of real data obtained e.g. by measurement. The program generates random sample according to the empirical distribution function computed from the data. The distribution might not be normal. This is a useful way of specifying the input distribution when a sufficient number of data points (say more than 20) is available and their distribution is not known. The inputted random variables are denoted by Y or y within the procedure. They are entered in columns, as usually.

Both specification types can be combined freely. The maximum number of X and Y variables is 20 (10+10). The Error propagation dialog panel is used to specify means and standard deviations (both parameters have to be specified) of x_i variables. y_i variables are defined by specifying a particular column from the current data sheet. The function of interest (in variables x_1, x_2, ..., y_1, y_2, ...) is entered in the usual syntax in the *Equation* window, for instance `5*x1*(1-x2)/log(3.124*y1)`. The pairwise correlation coefficients (x_i and x_j) or (x_i and y_j) variables can be entered, when they are known. The correlation coefficients values can influence mean, variance, quantile and confidence interval estimates heavily. The *Initialize* button sets all correlation coefficients to zero. When no correlations are specified, their values are set to zero. The correlation coefficients values have to lie within the (-1,1) interval.

For most situations, the number of simulations from 100 to 1000 should be satisfactory. The simulated data can be outputted to a separate sheet in the Protocol window upon request. These data can be used to analyze distribution of the resulting variable using the Basic data analysis module.

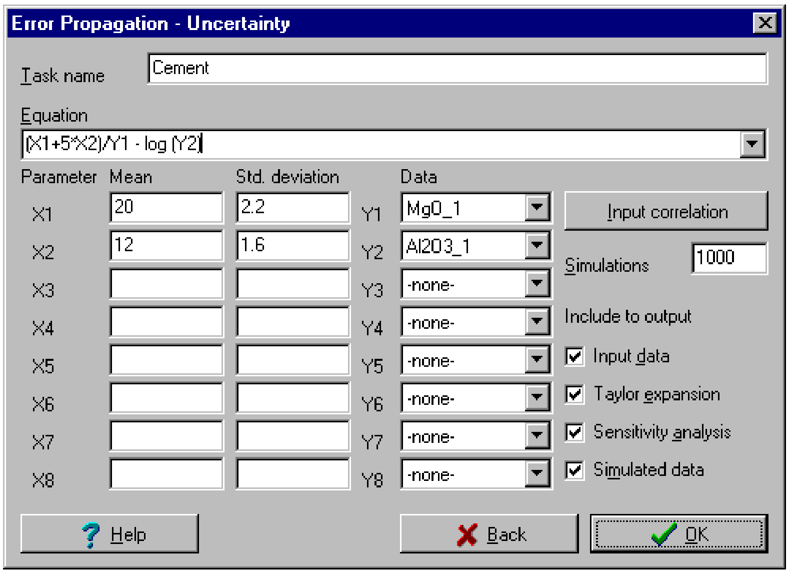


Fig. 2 Error propagation dialog panel

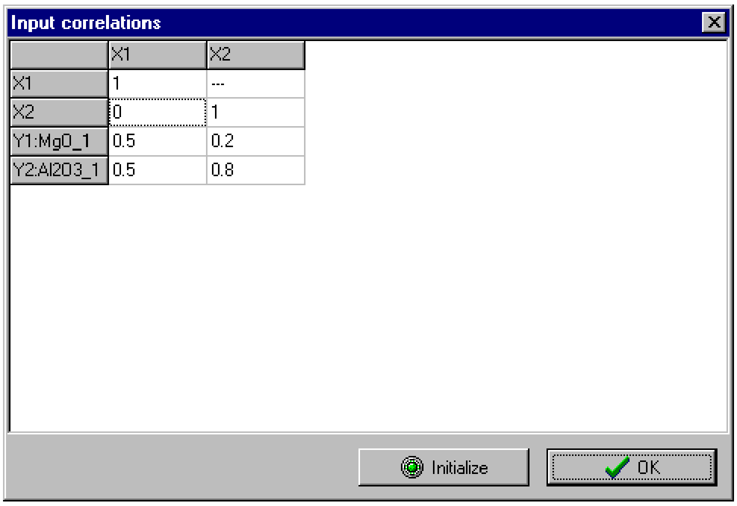


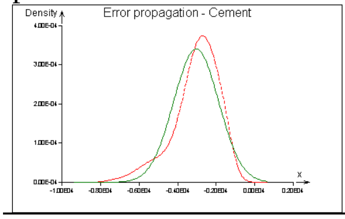
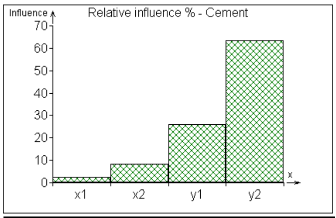
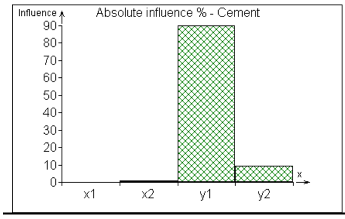
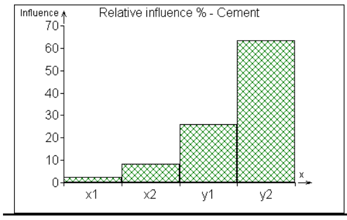
Fig. 3 Entering correlations between variables

Protocol

Function	Function of interest (assumed to possess continuous second order partial derivatives for the Taylor expansion).
Input variables	X variables distributions specified by their means and standard deviations.
Mean	Specified mean.
Standard deviation	Specified standard deviation.
95% interval	Confidence interval, computed via 2.5% and 97.5% percentiles.
+-3sigma	Interval containing 99.73% of data.
Inputted data	Y variables specifying simulation distribution through their empirical distribution function.
Mean	Arithmetic mean of the data as an estimate of the mean.
Standard deviation	Calculated standard deviation.
95% interval	Confidence interval computed via 2.5% and 97.5% data percentile.
+-3sigma	Interval containing 99.73% of data.

Resulting variable	Resulting variable characteristics obtained via Monte Carlo simulations.
Median	Median is often more reliable mean value estimate than arithmetic mean, when the data distribution is asymmetric or it has a high kurtosis.
Mean	Arithmetic mean as an estimate of the mean.
Standard deviation	Standard deviation estimate.
95% interval	2.5% and 97.5% percentiles of simulated data.
+-3sigma	Interval containing 99.73% of data.
Resulting variable range	Minimum and maximum of the resulting variable, generated by Monte Carlo simulation.
Sensitivity analysis	
Absolute sensitivity	Sensitivity of the resulting variable to the individual input variables. It is computed as partial derivative of the specified function with respect to a specified input variable, evaluated at the mean values of independent variables. Therefore, it is the sensitivity to a small change in the input variable. With respect to variability, it can be interpreted as the expected increase in resulting variable standard deviation when standard deviation of the input variable is increased by one unit.
Relative sensitivity	Individual input variable influence upon the resulting variable, which takes the amount of input variability into account. It is computed as the absolute sensitivity multiplied by the standard deviation of the particular input variable. When the relative sensitivity for a particular input variable is small relative to other variables, decreasing its variability or bringing it under statistical control is not very effective. If the relative sensitivity is large, decreasing the input variance or bringing the variable under statistical control can decrease resulting variability substantially.
Taylor expansion approximation	An alternative to the Monte Carlo simulations based on the second order Taylor expansion. Covariances (or correlations) among input variables are taken into account when computing the approximation.
Simple mean	Mean value estimate based on the first order expansion - evaluating the function at the mean of the input variables.
Corrected mean	Mean value estimate based on the second order expansion which does not take the covariances among the input variables into account.
Corrected mean (nonzero covariances)	Mean value estimate based on the second order expansion which does take the covariances among the input variables into account
Corrected standard deviation	Standard deviation estimate based on the second order expansion which does not take the covariances among the input variables into account.
Corrected standard deviation (nonzero covariances)	Standard deviation estimate based on the second order expansion which does take the covariances among the input variables into account.
95% interval	95% interval based on 2.5% and 97.5% percentiles. The computation assumes normality of the resulting variable!
Interval +-3sigma	An interval containing ca. 99.73% of data, based on 0.135% and 99.865% percentiles. The computation assumes normality of the resulting variable!
Simulated data	Simulated values of the resulting variable are saved to a separate sheet.

Graphs

Probability density function plot 	Compares the simulated distribution (red) with the normal model (green). Substantial discrepancies suggest that the resulting variable is not normally distributed.
Relative influence in % 	Relative sensitivity in %, see Protocol.
Absolute influence 	Absolute sensitivity, see Protocol.
Relative influence in % 	Relative sensitivity in %, see Protocol..

Graphical simulation

Menu:	QC.Expert	Simulation	Graphical Simulation

This is a tool for a fast manual data simulation. Points are placed in accordance with a preconceived idea of how the data should look like using mouse. The points are then converted to numerical values and saved in two columns of the current data sheet. WARNING! When the current sheet contains other data, they might be overwritten when output columns are not selected carefully.

Parameters

When the Graphical simulation module is selected using the appropriate menu, the Graphical simulation dialog panel appears, see Fig. 4.

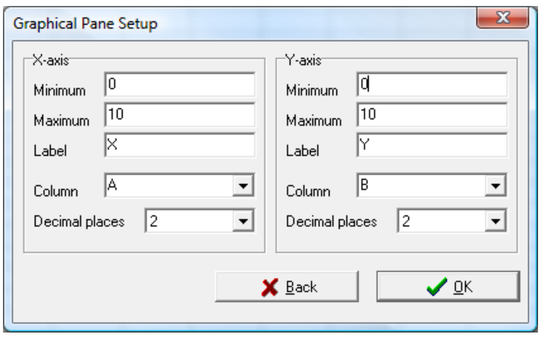


Fig. 4 Graphical simulation dialog panel

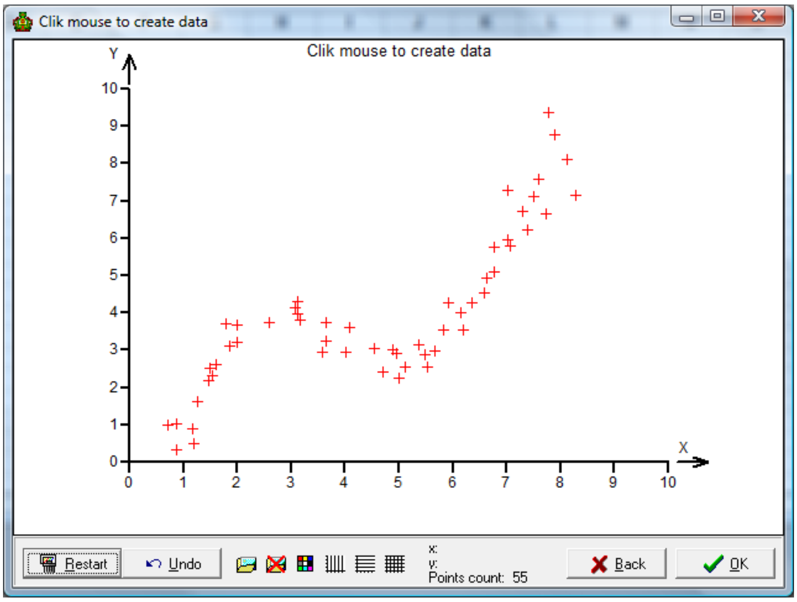


Fig. 5 Graphical simulation

x- and y-axis minimum and maximum values can be selected in the panel. Column names for simulated data and columns to which the data will be outputted can be specified there as well. Upon clicking the *OK* button, an empty plot appears, in which points can be "generated" using mouse, see Fig. 5. The *Restart* button deletes all previously generated points, the *OK* button saves them in the current data sheet, (see Fig. 6) and closes the graphical window. It is recommended to start simulation with an empty data sheet.

DATA

A1 | 72131147541

	X	Y	C
1	1.4672131	4.68	
2	1.9590164	5.88	
3	2.057377	4.92	
4	2.3770492	4	
5	3.0409836	6	
6	3.1393443	4.56	
7	3.704918	4.16	
8	4.3688525	2.72	
9	4.7868852	5.08	
10	5.0327869	4.44	
11	5.7459016	6.04	
12	5.9918033	4.48	

Fig. 6 Simulated data from previous plot, outputted to a data sheet.

The graphical simulation module can also be used to digitalize scanned images from literature, analog plotters etc. Select axes scales to the scanned plot, open image in *jpg*, *gif*, *bmp* or *wmf* format using the *Open file* button at the bottom of the graphical input window and use right mouse button to move and resize the image to fit in the axes, then use mouse to digitalize any point, curve or edge from the image, than press *OK* to transfer the coordinates of the points in the current data sheet. An example is given on Fig. 7 and Fig. 8.

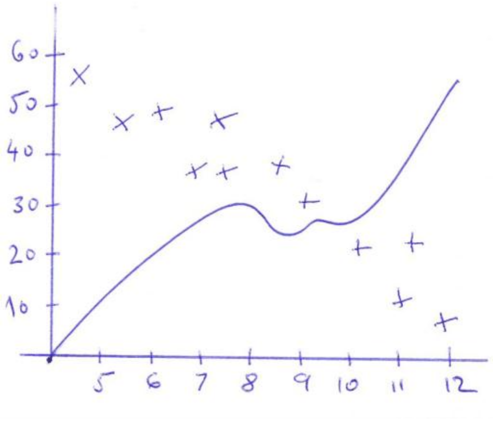


Fig. 7 Scanned jpg image to be digitalized

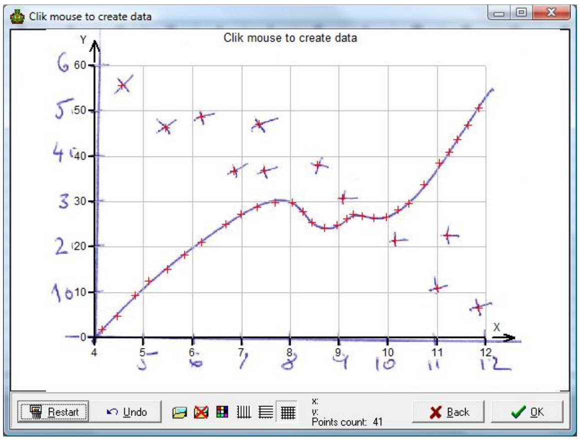


Fig. 8 Imported and digitalized image