

SimLab 2.2

REFERENCE MANUAL

INDEX

REFERENCE MANUAL

INTRODUCTION	7
1 SIMLAB MAIN PANEL	8
1.1 STARTING SIMLAB USING A SAVED CONFIGURATION	9
2 SAMPLE GENERATION	13
2.1. NEW SAMPLE GENERATION	16
2.1.1. Input Factors	18
2.1.2. Selection of sampling Method	28
<i>FAST method</i>	29
<i>Fixed samples</i>	31
<i>Latin hypercube</i>	32
<i>Morris</i>	34
<i>Quasi-Random LpTau</i>	35
<i>Random</i>	36
<i>Replicated Latin Hypercube</i>	37
<i>Sobol's method</i>	39
2.1.3. Output Format Selection	45
2.1.4. Sample Generation.....	47
2.2. SAMPLE VISUALISATION	48
2.2.1. Sample visualisation and generation with Cobweb plots	50
2.2.2. Samples visualisation and generation with Scatter Plot	53
2.2.3 Sample visualisation with Histograms	55
3 MODEL EXECUTION	57
3.1 HOW TO SET AND EXECUTE A MONTE CARLO ANALYSIS.....	59
3.1.1 External model execution	60
3.1.2 Internal model execution.....	63
4. MODEL' RESULTS ANALYSIS (Statistical Post Processor)	68
4.1 UNCERTAINTY ANALYSIS	76
4.2. SENSITIVITY ANALYSIS (SRC, SRRC, PCC, PRCC, SPEAR, PEAR INDEXES).....	83
4.3. SENSITIVITY ANALYSIS MORRIS METHOD.....	98
4.4 SENSITIVITY ANALYSIS: FAST	100
4.5 SENSITIVITY ANALYSIS: IMPORTANCE MEASURES	102
4.6 SENSITIVITY ANALYSIS: SOBOL	104
APPENDIX A	107
APPENDIX B: THREE TEST CASES FOR SIMLAB	123

TEST CASE 1: GIS APPLICATION	123
DESCRIPTION OF THE TEST CASE	123
OBJECTIVES	124
INTERFACE BETWEEN PREPSPOP AND ARCVIEW	124
REFERENCES	126
TEST CASE2: ECONOMIC APPLICATION	127
1. INTRODUCTION	127
CASE STUDY 1: SENSITIVITY ANALYSIS OF SINGLE REGRESSION MODEL, SOLOW CLASSIC.....	128
Model selection	128
Monte-Carlo forecast and variance decomposition	129
CASE STUDY 2: SENSITIVITY ANALYSIS OF MULTIPLE REGRESSION MODEL, LINEAR VERSUS NON LINEAR MODELLING; KLEIN1 CONSUMPTION EQUATION.....	131
Model selection	132
Montecarlo forecast and variance decomposition.....	134
REFERENCE.....	138
TEST CASE 3: FINANCIAL APPLICATION (BEI)	138
INTRODUCTION	138
THE USE OF UA AND SA IN THE WORLD OF FINANCE	139
An example of use of UA and SA to manage credit risk.....	139
An example of use of UA and SA to manage market risk.....	139
AN APPLICATION OF UA AND SA.....	140
Introduction.....	140
The HJM model	140
Results and conclusions	142
APPENDIX C	146
REFERENCES.....	155
TABLE OF FIGURES.....	157

We are grateful to Professor R. Cooke for his contribution to the development of the dependence tree correlation method, implemented in this software.

INTRODUCTION

SimLab 2.2 is a software designed for Monte Carlo – based uncertainty and sensitivity analysis.

Monte Carlo (MC) methods are used here for pseudorandom number generation with emphasis on sampling set of points from joint probability distributions; the designation “sample distribution” is often used.

MC-based uncertainty and sensitivity analyses are based on performing multiple model evaluations with probabilistically selected model input, and then using the results of these evaluations to determine 1) the uncertainty in model predictions and 2) the input variables that gave rise to this uncertainty. In general, a Monte Carlo analysis involves five steps.

In the first step, a range and distribution are selected for each input variable (input factor). These selections will be used in the next step in the generation of a sample from the input factors. If the analysis is primarily of an exploratory nature, then quite rough distribution assumptions may be adequate.

In the second step, a sample of points is generated from the distribution of the inputs specified in the first step. The result of this step is a sequence of sample elements.

In the third step, the model is fed with the sample elements and a set of model outputs is produced. In essence, these model evaluations create a mapping from the space of the inputs to the space of the results. This mapping is the basis for subsequent uncertainty and sensitivity analysis.

In the fourth step, the results of model evaluations are used as the basis for uncertainty analysis. One way to characterise the uncertainty is with a mean value and a variance. Other model output statistics are provided.

In the fifth step, the results of model evaluations are used as the basis for sensitivity analysis.

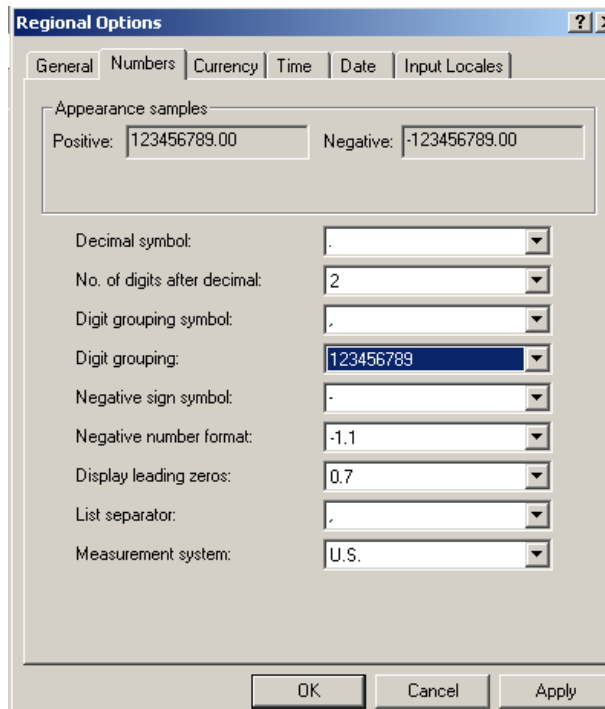
SimLab is composed of three modules. These modules cover all the steps summarised above.

1. The Statistical Pre Processor module executes the first and second steps.
2. The Model Execution module accomplishes the third step.
3. The Statistical Post Processor module carries out the fourth and fifth steps.

1 SIMLAB MAIN PANEL

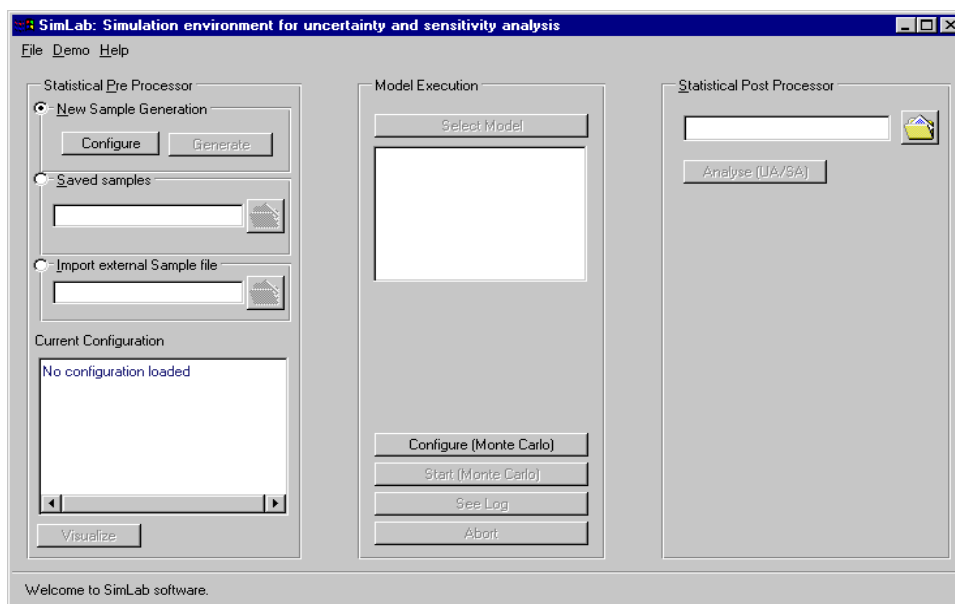
IMPORTANT NOTICE for WINDOWS 2000/XP users: Before using SimLab, please set the following options on your PC:

1. Start the *Control Panel*, click on *Regional Options* and then on *Numbers*.
2. Set the Options as follows:



3. Now you can start SimLab.

At start SimLab displays the main panel; this panel is logically divided in three frames:



**Statistical
Preprocessor module**

**Model execution
module**

**Statistical
Postprocessor module**

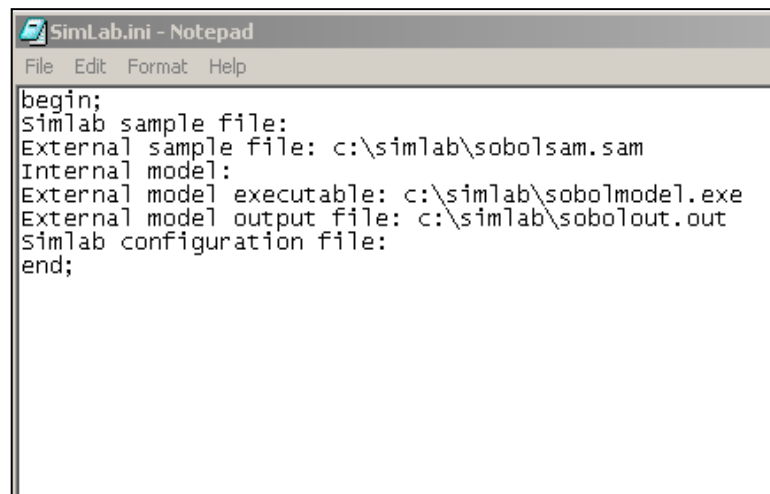
Figure 1: SimLab main panel

1. The **Statistical Pre Processor module**: generates a sample in the space of the input factors.
2. The **Model Execution module**: executes the model for each point in the sample of input factors
3. The **Statistical Post Processor module**: performs the uncertainty and sensitivity analysis.

1.1 STARTING SIMLAB USING A SAVED CONFIGURATION

At start up the application reads the SimLab configuration file (SimLab.ini) placed in the installation directory.

“SimLab.ini” contains the list of all the configuration items that can be loaded from saved computations during the normal execution of the program; to start the application with a specific configuration the user has to insert the complete path of previous saved computations after the corresponding labels.



```
SimLab.ini - Notepad
File Edit Format Help
begin;
simlab sample file:
External sample file: c:\simlab\sobolsam.sam
Internal model:
External model executable: c:\simlab\sobolmodel.exe
External model output file: c:\simlab\sobolout.out
Simlab configuration file:
end;
```

Figure 2 Example of SimLab.ini

**SimLab Main panel:
File Menu
description**

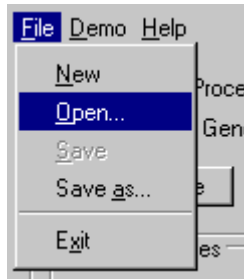


Figure 3: Main panel menu

In the top left part of the main window the file menu is located. Through this menu the user can save the current job or load the work stored from a SimLab project file. The SimLab project file contains a copy of a working session. The file used for the SimLab configuration has as the default extension **psc** (project SimLab configuration).

Menu:

- | | |
|----------------|--|
| New | Start a new session. |
| Open | Load a saved project. |
| Save | Save the active project. |
| Save as | Save the current work in a new project file. |
| Exit | Quit SimLab. |

**SimLab Main panel:
Demo Menu
description**

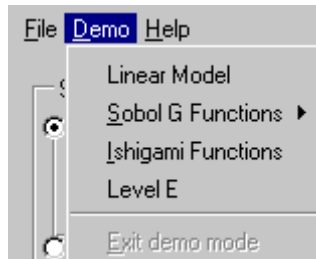


Figure 4 Demo menu

SimLab provides a special configuration called *demo mode*, in which the user can choose among a set of *built-in* examples and analyse results using predefined settings. In the demo mode some options available in SimLab are disabled to force the user following a predefined sequence of operations. In particular, the user cannot:

- Add or delete factors;
- Modify the distribution type of a single factor;
- Correlate factors;
- Load existing factors files;
- Add or delete intervals in the case of uniform distributions;
- Change the order of the input factors;
- Change the sample file;
- Change the model or the model output file;
- Add or delete variables for UA/SA.

Given the constraints above, the user can select among different random sampling techniques and set samples with different number of points for the same built-in example. S/he cannot overwrite demo files. When the FAST sampling technique is used, the user can also set up groups of factors.

The selection of one of the menu items in the Demo menu enables the demo mode. When the user enters the Demo mode, s/he is warned if the working session has not been saved. This release of SimLab provides four different models for the demo mode:

- 1) **Linear model** – Three input factors, X_1, X_2, X_3 , are defined as uniform distributions varying over a specific range:

$$X_1 \sim U(0.5, 1.5)$$

$$X_2 \sim U(1.5, 4.5)$$

$$X_3 \sim U(4.5, 13.5)$$

The output variable Y is simply the sum of the three input factors.

- 2) **G-function of Sobol'** – This is a non-monotonic function, see [17], whose analytical expression takes the form:

$$Y = \prod_{j=1}^k g_j(X_j)$$

$$\text{where } g_j(X_j) = \frac{|4X_j - 2| + a_j}{1 + a_j} \text{ with } a_j \geq 0 \text{ and } k = 8$$

$a_j = 0$ means that the associated factor is very important. For $a_j = 1$ the input factor is slightly less important. If $a_j = 99$, the corresponding factor is absolutely un-important. The eight input factors X_j are uniformly distributed in the range $(0, 1)$. The user can choose among four different sets of a_j parameters:

$$a_j = \{0, 0, 0, 0, 0, 0, 0, 0\} \quad [\text{all the factors very important}]$$

$$a_j = \{99, 99, 99, 99, 99, 99, 99, 99\} [\text{factors equally non-important}]$$

$$a_j = \{0, 1, 4.5, 9, 99, 99, 99, 99\} [\text{in decreasing order of importance}]$$

$$a_j = \{99, 0, 9, 0, 99, 4.5, 1, 99\} [\text{in random order of importance}]$$

- 3) **Ishigami function** – Another non-monotonic function (see [8]) with three input factors X_1, X_2, X_3 . The factors are uniformly distributed in $(-\pi, \pi)$. The model is:

$$Y = \sin X_1 + A \sin^2 X_2 + B X_3^4 \sin X_1$$

The parameters A and B have values $A = 7$ and $B = 0.1$. The main peculiarity of this model is the dependence on X_3 , which has no additive effect on Y but interacts with X_1 .

- 4) **Level E** – The Level E test case was selected for inclusion in the demo mode because of its interesting non-monotonic features that are suitable for testing the techniques implemented in the software. The Level E model is a computer code used in safety assessment for nuclear waste disposal. This code has been the subject of two international benchmarks of risk assessment codes and Monte Carlo methods, and is well known in the literature (see [12], [13] and [18]). The Level E model predicts the radiological dose to humans over geological time scales due to the underground migration of radionuclides from a nuclear waste disposal site through a system of natural and engineered barriers. The core of the model is a set of partial differential equations, which describes the migration of four nuclides through two geosphere layers characterized by different hydro-geological properties. The processes being considered in the model are radioactive decay, dispersion, advection and chemical retention between the migrating nuclides and the porous medium. The model has a total of 33 factors, 12 of which are taken as independent uncertain factors. A description of the input factors and of their probability distributions can be found in [14].

The above models are executable files and are located in the directory Models under the SimLab base directory.

To exit the Demo mode, the user can choose the menu item *Exit Demo mode* in the *Demo menu* or press the button *Exit demo mode* located on the SimLab main panel. The demo configuration will be discarded after exiting the demo mode, so the user has to save it under a different name if s/he has made any change.

2 SAMPLE GENERATION

The first step in the sample generation phase is to select ranges and distributions (probability distribution functions, pdf) for the input factors. This selection makes use of the best information available on the statistical properties of the input factors. In some instances, it is possible to get empirical estimates of pdf's from available underlying data for the input factors. The effort put at this stage is related to the scope of the analysis: if the analysis is at an exploratory stage, then rather crude pdf's could be adequate: the cost of getting better pdf's may be relatively high. The second step is to select a sampling method among the available ones. Note the choice of the sampling method has implications on what type of sensitivity analysis the user is able to perform later on (for example, if we use FAST sampling, we cannot perform the analysis using the method of Morris).

The third step is the actual generation of the sample from the input pdf's. The Statistical Pre Processor guides the user through the choice of the distributions and the visualisation of the sample.

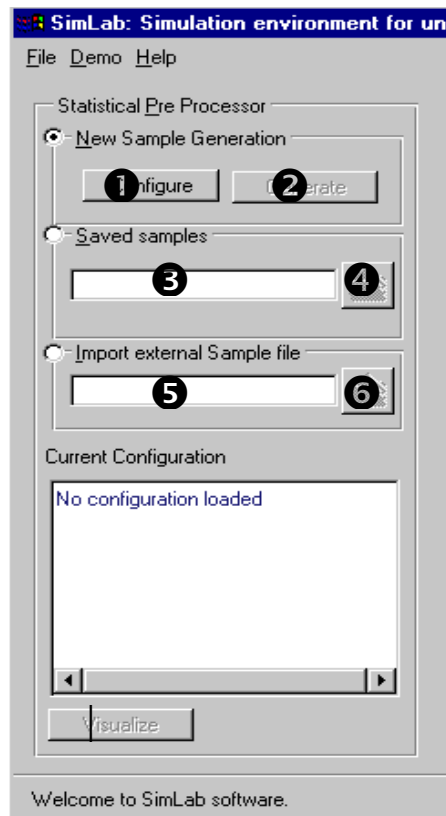


Figure 5: Pre Processor frame

Figure 5: Pre Processor is the left frame of the SimLab main panel. This frame contains the commands to load or create a configuration for the sample generation. Three different options can be alternatively chosen:

- a. create a new sample file (**‘New sample generation’**),
- b. load an existing sample file previously created by the user with SimLab (**‘Load sample file’**), and
- c. import a sample file created with a different application in ASCII format (**‘Import external sample file’**).

The toggle in the left part of each sub-frame allows the user to select among the three options.

**Statistical Pre
Processor module:
Main functionalities**

- ❶ **Configure:** This toggle loads the *Statistical Pre Processor main panel* using which the user can create a new configuration of input factors for subsequent sample generation.
- ❷ **Generate:** This toggle allows the user to *de-facto* create the sample with the desired configuration. This toggle is enabled when the configuration phase is completed.
- ❸ **Existing sample file textbox:** This textbox visualises the file name that is selected using the toggle ≠.
- ❹ **Load existing sample file:** Load a sample file previously generated by SimLab. As default SimLab saves sample files with the extension **.sam**. More details on the format of sample files are in APPENDIX C.
- ❺ **External sample file textbox.** This textbox visualises the file name that is selected using the toggle ≈.
- ❻ **Import external sample file:** Import a sample file created with any application but SimLab in ASCII format. The format specification of the external sample file can also be found in APPENDIX C.

7 Visualise current configuration: This frame visualises the configuration actually loaded in the Pre Processor module. In *Figure 6* an example is given.

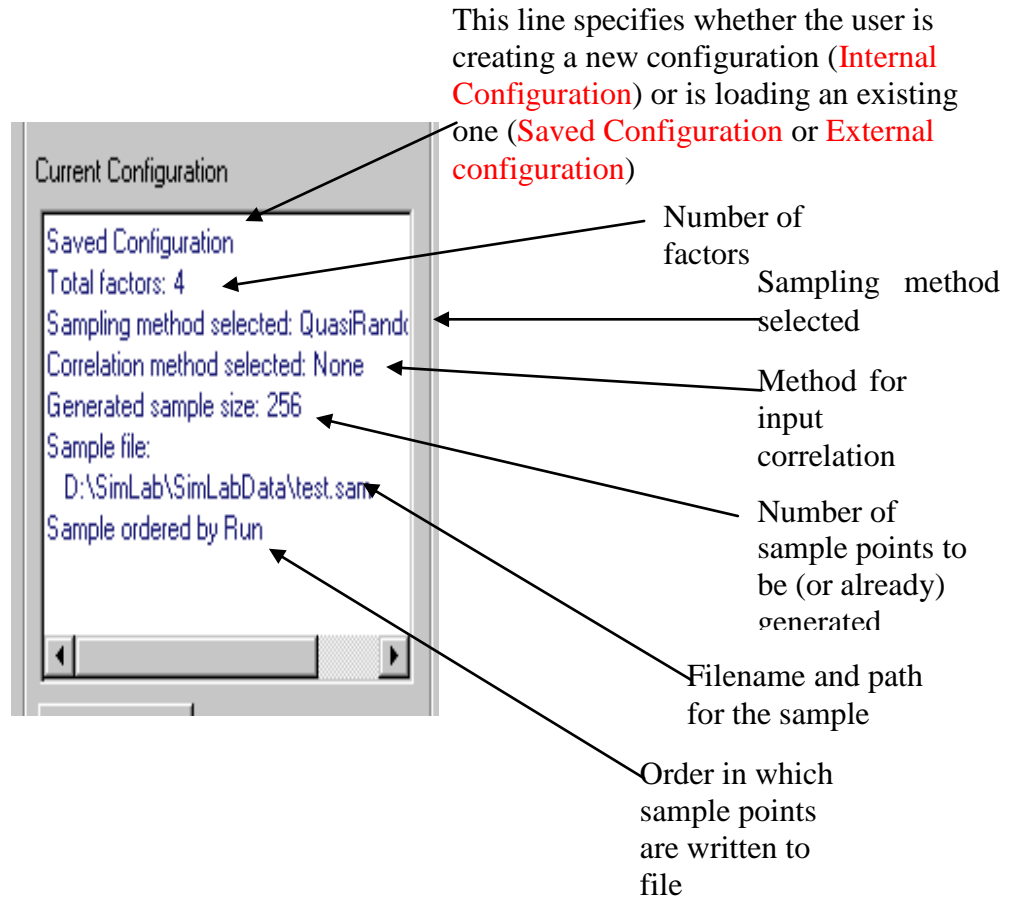


Figure 6: Current configuration area

8 Visualise: This toggle introduces the user to the graphics of the sample set (loaded or generated). A number of graphs (scatter-plots, cobwebs and histograms) and tables are available. The user can visualise the sample correlation structure and the values of all the sample points generated. Using scatter-plot or cobwebs the user can select a subset of the sample and create a new sample (See 2.2.1. Sample visualisation and generation with Cobweb plots, 2.2.2. Samples visualisation and generation with Scatter Plot and 2.2.3 Sample visualisation with Histograms).

SAMPLES GENERATION (Statistical Pre Processor Module)	2.1. New Samples Generation	2
--	------------------------------------	----------

2.1. NEW SAMPLE GENERATION

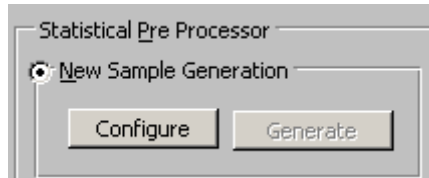


Figure 7: New Sample Generation

The upper part of the Pre Processor area is dedicated to the commands for a new sample generation. A new sample is generated using the information on the input factors distributions and selecting the desired sampling method.

Pre Processor Configuration:

By clicking on the toggle **Configure** (see Figure 7) the user can set the configuration for the sample. The panel *Statistical Pre Processor - Main Panel* (Figure 8) is opened allowing the user:

1. to set the statistical properties of the input factors (i.e. the distributions, the ranges and the truncation values);
2. to select the sampling method;
3. to select the name of the file where the sample will be saved.

Once this has been done, the toggle **Generate** is enabled. The generation of the sample starts when the user clicks on this toggle.

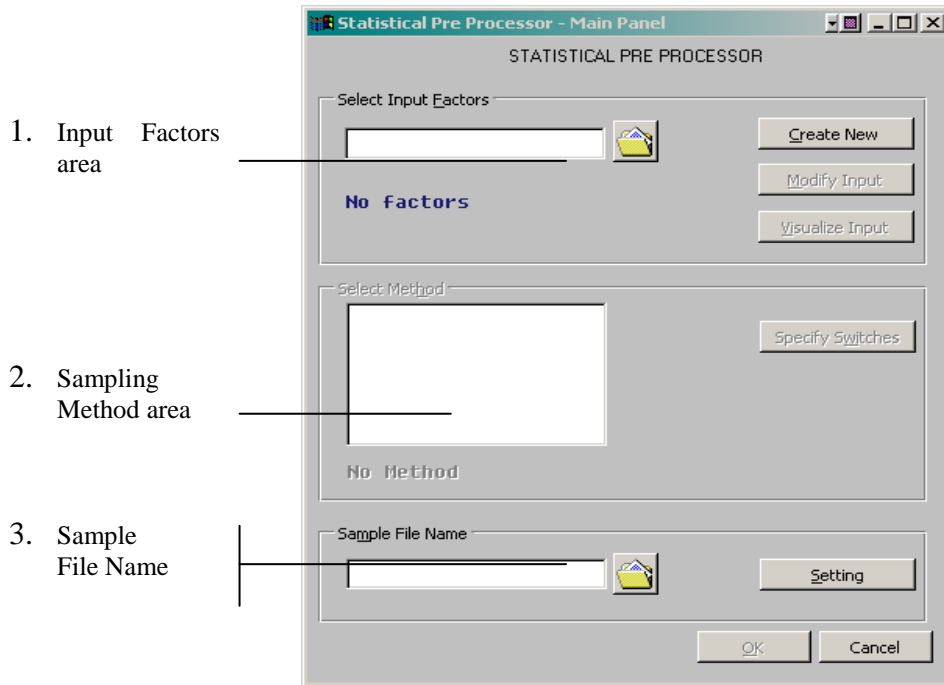


Figure 8: Statistical Pre Processor - Main Panel

The *Statistical Pre Processor - Main Panel* is composed of three frames:

1. **Select Input Factors:** here the user defines the input factors, by specifying the number of input factors, their pdf's, their ranges and their correlations (see 2.1.1. Input Factors).
2. **Select Sampling Method:** here the user can select among a number of available sample methods (see 2.1.2. Selection)
3. **Select Sample File Name:** the sample design generated by SimLab is saved on a text file. The name of such file is selected by the user (see 2.1.3. Output Format Selection).

2.1.1. INPUT FACTORS

In the upper part of the *Statistical Pre Processor Panel* the user defines the input factors, by specifying their number, their pdf's, their ranges and their correlations. Two choices are available:

1. create a new set of input factors from scratch or
2. load a configuration created previously.

As default, the files containing a list of factors have extension **.fac** (see APPENDIX C: “.fac” format file).

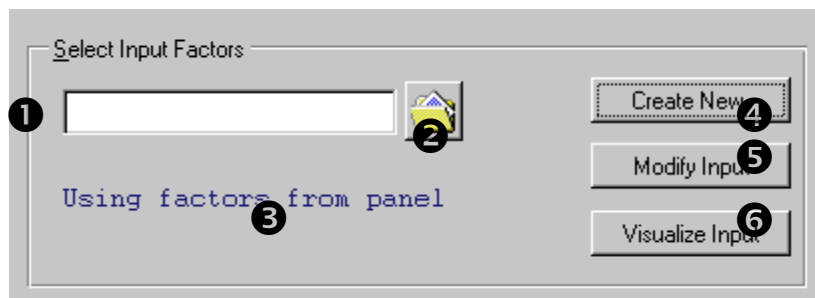


Figure 9 Select Input Factors frame

Select Input Factors area

- ❶ **Loaded factors file textbox.** This textbox shows the name of the file of input factors currently loaded in the system.
- ❷ **Open factors file toggle.** Click to select and load a file of input factors.
- ❸ **Message area.** This area visualises the current status of the input factors. The message *no factors* appears if no input factors are loaded. The message *Using factors from panel* appears if the user has just created a new set. The message *Using factors from file* appears if the user is using a set of input factors previously saved in a file.
- ❹ **Create New toggle.** Clicking on this toggle the user creates a new set of factors. The panel *input factors selection* is opened (see Figure 10).
- ❺ **Modify Input toggle.** Clicking on this toggle the user modifies the parameters of the current input factors configuration.
- ❻ **Visualise Input toggle.** Clicking on this toggle the user inspects the current factor configuration and all the related parameters.

Clicking **Create new** (see *Figure 9*) the following panel related to the visualisation and modification of inputs factors is shown:

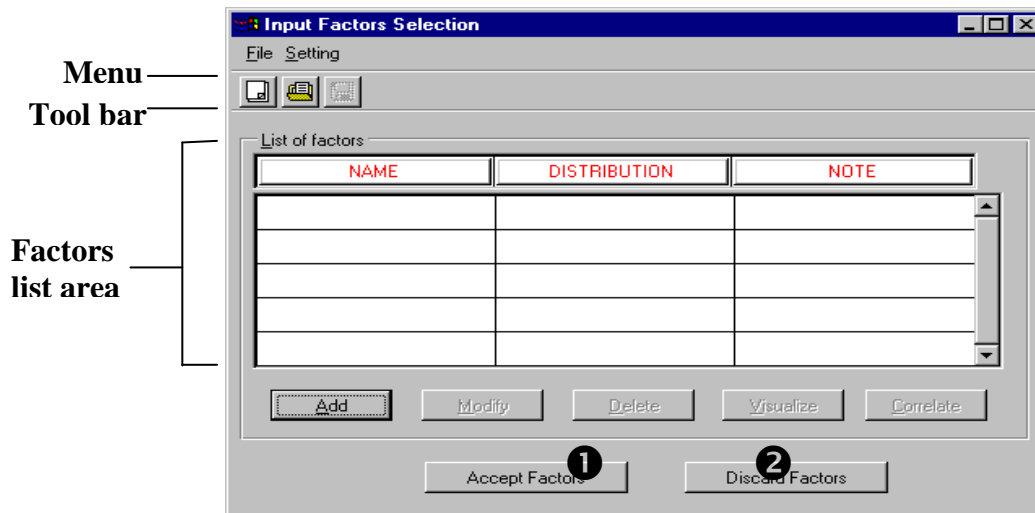


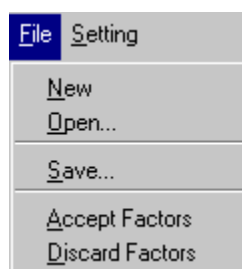
Figure 10 Input Factors Selection panel

New command. A new list of input factors can be created. If some factors were previously loaded a warning appears before removing the old list.

Open command. Opens an existing file of input factors.

Save command. Save the current list of factors into a file with extension **.fac**.

**Input Factors Selection:
File Menu**



Input Factors Selection: Setting Menu



The **Model's Parameters** item allows the user to set, or modify, the distribution default truncation through the following panel:

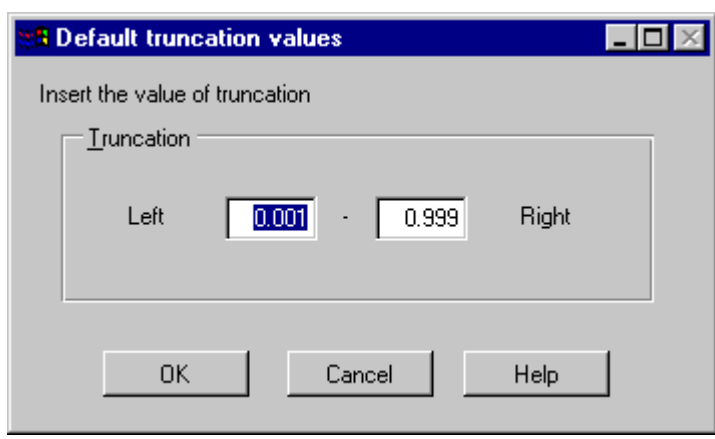


Figure 11 Truncation values

These are the default truncation quantiles. Such truncations affect all the factors having distributions defined over intervals of infinite width. The (left and/or right) truncation values can also be modified for single factors (see distribution panels in APPENDIX A).

Input Factors Selection: Menu bar



- ① **Create new factors toggle.** Create a new factors set (see Input factors selection: File Menu → New command)
- ② **Open list of factors toggle.** Open a file of input factors (see Input factors selection: File Menu → Open)
- ③ **Save list of factors button.** Save the factors list in a file (see Input factors selection: File Menu → Save)

**Input Factors Selection:
Factors list area**

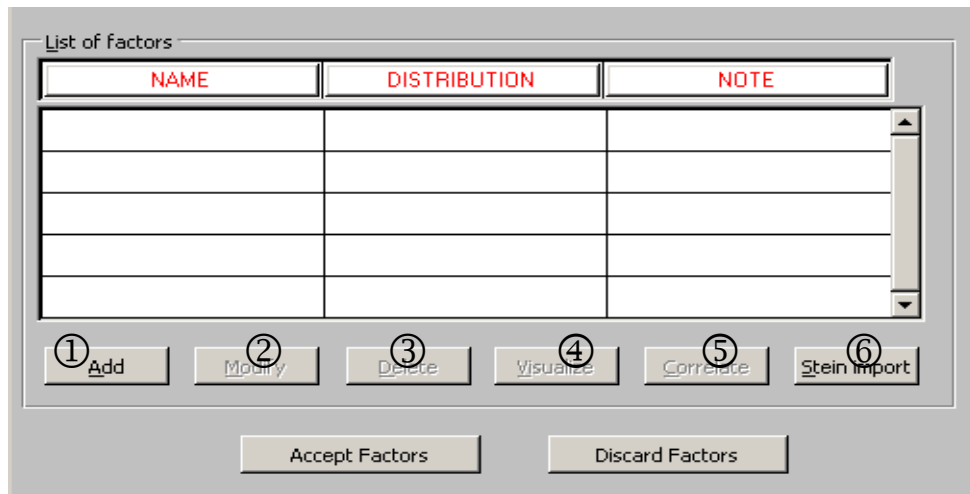


Figure 12 Factors list area

Once the user has loaded an existing factor file, or has created new factors, the area in *Figure 12* contains a list of the statistical characteristics for each factor. With the commands given in *Figure 12* the user can insert a new input factor (a new row is added in the list), modify or remove a selected input factor, build a correlation structure between the input factors.

The line in the factor list contains the following information (see *Figure 13*):

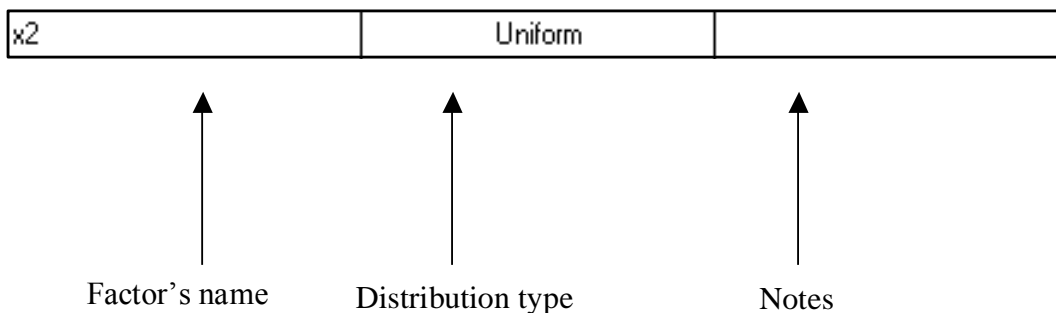


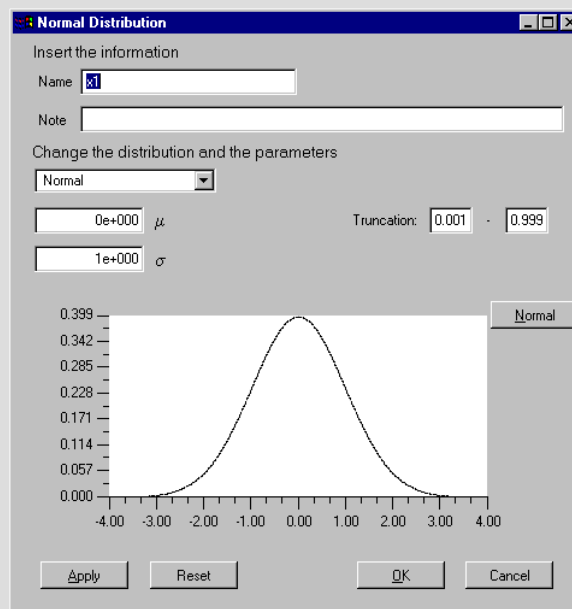
Figure 13 Input factor summary

Factors list area: toggles functionality-

- ① **Add toggle.** The user can add a new input factor to the current list of factors. The same action is obtained by double clicking on an empty row.
- ② **Modify toggle.** Select an input factor and then click on this toggle: the user can modify the statistical parameters of the selected input factor. The same action is obtained by double clicking the factor's row.
- ③ **Delete toggle.** Select an input factor and then click on this toggle: the user deletes the selected input factor.
- ④ **Visualise toggle.** Select an input factor and then click on this toggle: the user examines the statistical parameters of the selected input factor.
- ⑤ **Correlate toggle.** The user can impose a correlation structure to the input factors. Two different correlation methods are available (the Iman&Conover rank correlation (see [5]) and the dependence tree method [9]).
- ⑥ **Stein Import toggle.** The application imports an external ASCII sample file (generally correlated), and uses the method proposed by Stein to create an LHS (or replicated-LHS) sample with the same correlation structure. Format specifications for the external ASCII sample file are given in APPENDIX C.

Example: add an input factor

Click the **Add** toggle: the distribution panel for the normal pdf (default) is visualised.



A set of 12 pdf's is available. A complete list of the distributions and a description of the corresponding panels is given in APPENDIX A.

**Factors list area:
Correlate button**

The user can specify a correlation structure in the sample design.

SimLab implements three methods to induce correlation:

1. **the Iman and Conover method,**
2. **the dependence-tree / copula method and**
3. **the Stein method.**

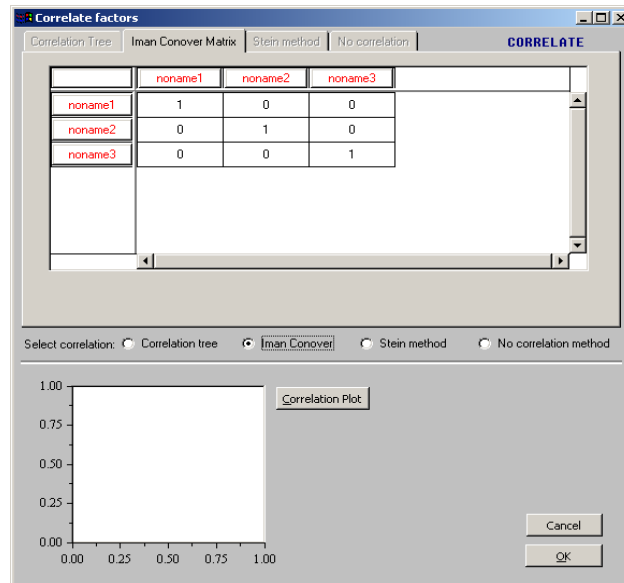


Figure 14: Correlation panel

**Correlation method:
Dependence tree**

The dependence-tree method for modelling the correlation between factors (see [9] and [10]) is related to influence diagrams, but makes use of undirected acyclic graphs instead of directed graphs used in influence diagrams.

The user can specify correlations among input factors that form a tree structure. Whatever correlation values are imposed by the user in this way, it is guaranteed that a joint pdf exists. The joint pdf has minimum information amongst all those joint distributions which satisfy the criteria given by the user. By selecting the **Correlation tree** button in *Figure 14* the user can induce the desired correlation between input factors as follows:

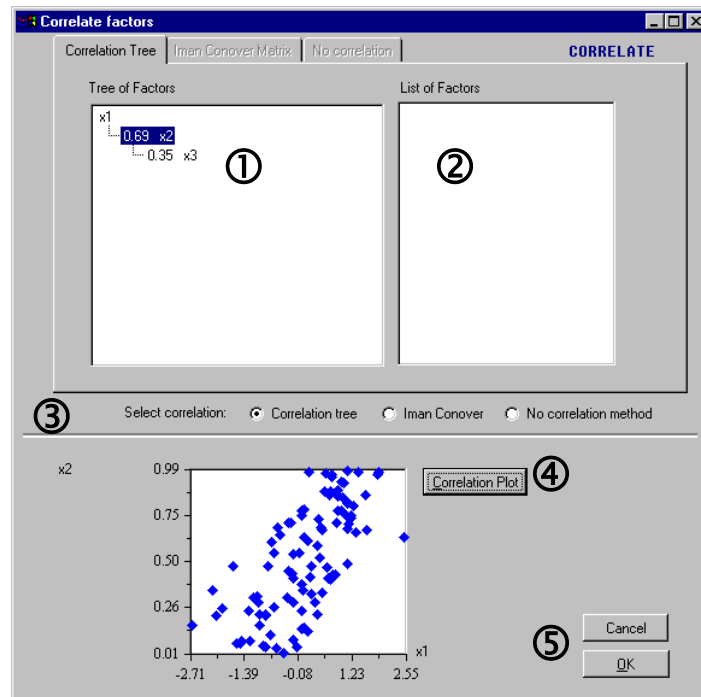
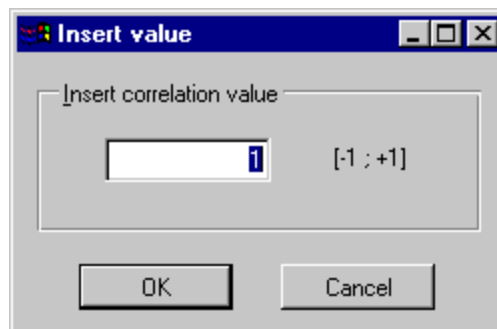


Figure 15 Correlation tree panel

- ① **Factors tree area.** This area visualises the tree correlation structure currently used; in *Figure 15* an example is displayed. The inputs factors X1 and X2 are correlated with value 0.69, X2 and X3 with value 0.35.
- ② **List of Factors area.** This area shows the list of available input factors not yet correlated. To define a new correlation the user can drag an item from this list and overlay it over a factor tree element. The correlation value between the two entities is asked to the user as follows.



- ③ **Select Correlation.** Choose one of the four items here to select the desired correlation method.
- ④ **Correlation Plot button.** Click this button after selecting a correlation structure. SimLab will display a scatter-plot of 100 correlated points for the pair of chosen factors.
- ⑤ **Cancel and OK buttons.** The **Accept** button saves the correlation configuration and will close the panel, while the **Cancel** discards and quits.

The second correlation method available in SimLab is the Iman-Conover method (see [5]). This is used to induce a desired rank correlation on pairs of input factors. Its characteristics are:

**Correlation method:
Iman-Conover**

- Rank correlations can be set independently on marginal distributions,
- the original form of the marginal distributions is preserved,
- May be used with many sample schemes,
- if the correlations imposed are too strong, then the correlation matrix is not positive definite, and a message is displayed.

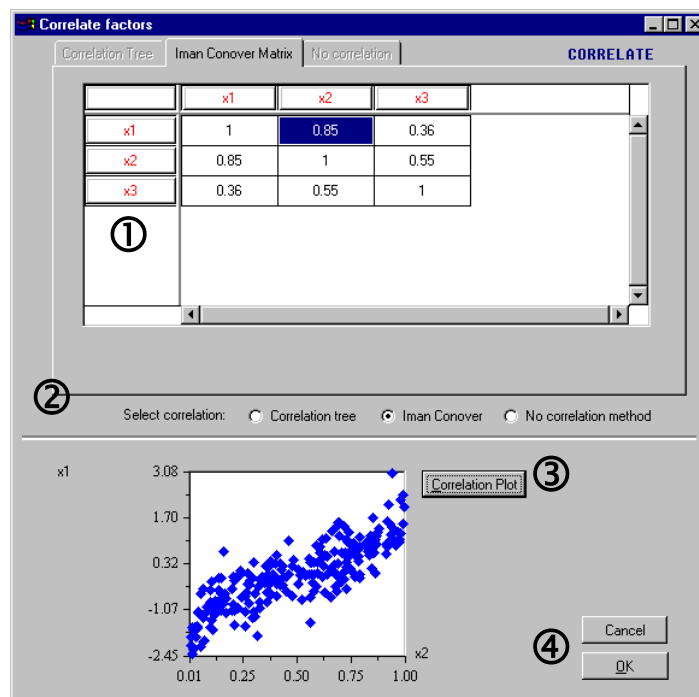


Figure 16 Iman and Conover Correlation Panel

- ① **Correlation matrix area.** This part of the panel visualises the correlation matrix, a square matrix for the elements of the current configuration. To impose a correlation, choose the cell related to the two items to correlate, double click on it and type the value.
- ② **Correlation switch.** Choose the Iman and Conover toggle.
- ③ **Correlation Plot button.** Click this button after selecting a correlation structure. SimLab will display a scatter-plot of 100 correlated points for the pair of chosen factors.
- ④ **Cancel and OK buttons.** The **Accept** button saves the correlation configuration and will close the panel, while the **Cancel** discards and quits.

The third correlation method available is the **Stein method** (see [23]).

Correlation method:
Stein

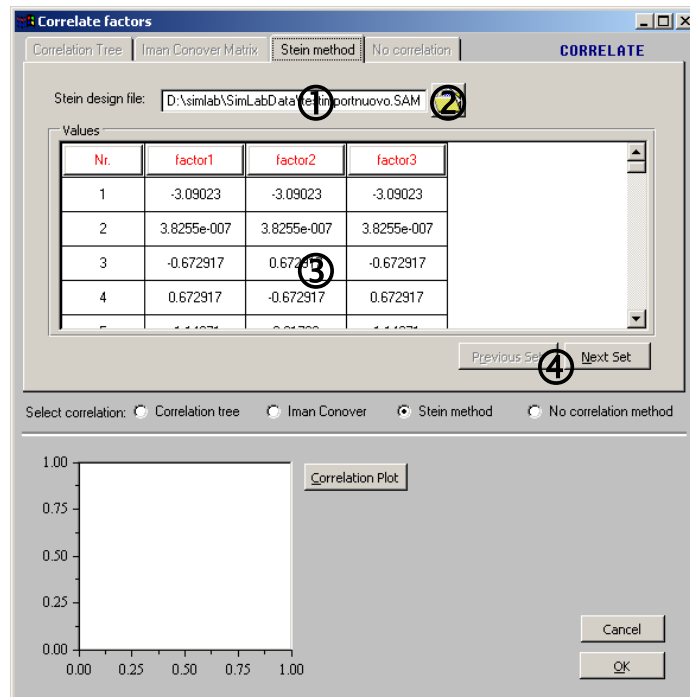


Figure 17 Stein Correlation Panel

The method of Stein allows the user to generate a correlated LHS sample. The user must provide an ASCII file (see format in APPENDIX C) that contains a correlated sample (eg, a random sample, or even an empirical sample generated by an experiment). The method generates an LHS sample with the same correlation of the sample provided by the user.

- ① **Correlated ASCII Sample File.** The user selects the ASCII input file. The format of this file is identical to the format of an external import file (see APPENDIX C).
- ② **Correlated ASCII Sample File Selector.** Browser to help the user selecting the ASCII input file
- ③ **Stein Correlation matrix area.** This area visualises all the elements in the correlated ASCII sample file.
- ④ **Navigation buttons.** Buttons used to inspect groups of sample values of the correlated ASCII sample file.

Input Factors Selection panel:
Accept and Discard buttons

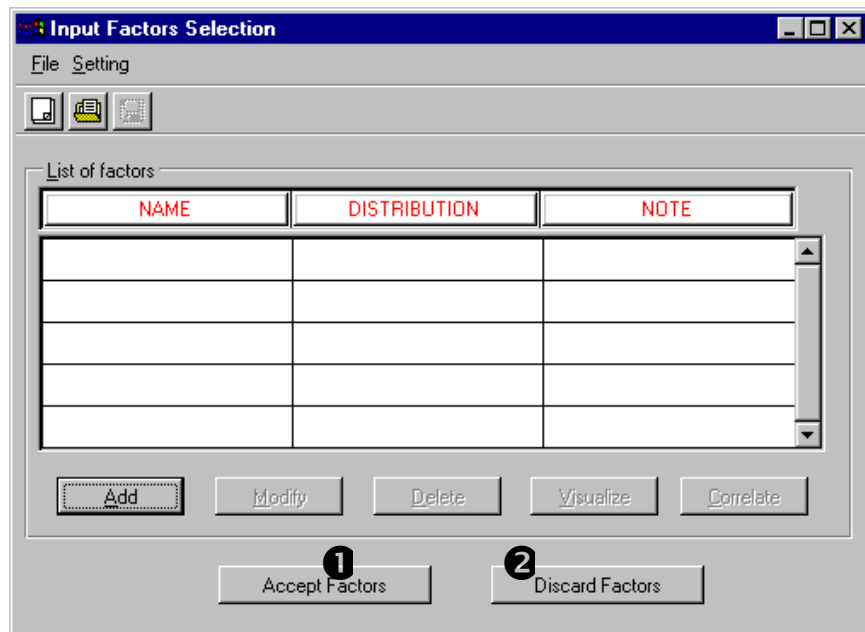


Figure 18 Input factors selection panel

Once the user has created the input factor configuration for the subsequent sample generation, such configuration can be accepted by clicking on “accept factors” and becomes the working configuration.

- ❶ **Accept Factors button.** Save the current input factor configuration as the working configuration and return to the Pre-processor main panel.
- ❷ **Discard Factors button.** Discard the configuration and go back to the Pre-processor main panel.

2.1.2. SELECTION OF SAMPLING METHOD

The ‘Select method area’ of Pre Processor Main panel allows the user to choose among the many sampling methods available.

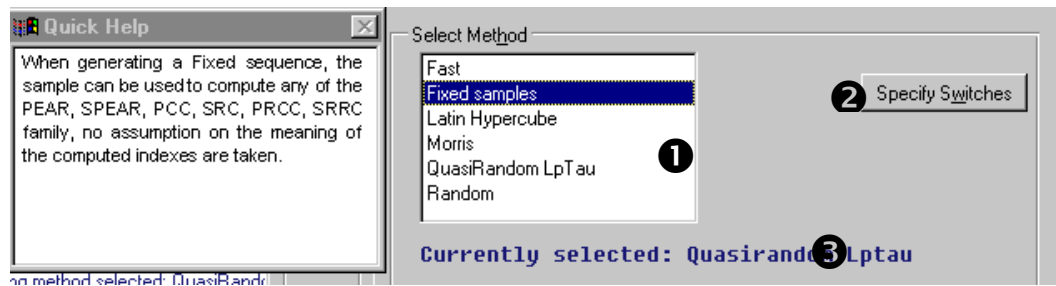


Figure 19 Select Method area

Select Method area

- ❶ **Methods area.** A list of methods available for the current sample generation is displayed. Selecting an item a Quick Help summarising the method characteristics is shown.
- ❷ **Specify Switches button.** Clicking this button the user visualises the panel for the input parameters (switches) to the selected method. The switch panel can also be visualised double clicking the method name in the Methods area.
- ❸ **Message area.** This message visualises which method has been selected in the current configuration.

The sampling techniques and SA method available in the software are FAST, Fixed samples, Latin Hypercube, Morris, QuasiRandom LpTau, Random. The choice of the technique is related to the factor set created. The method area of the Prep main panel lists visualises only those techniques suitable to the current loaded factors’ configuration.

FAST method

The Fourier Amplitude Sensitivity Test (FAST) is a method of sensitivity analysis, and works for monotonic and non-monotonic models alike, see references [14] and [15]. This method is based on performing numerical calculations to obtain the expected value and variance of a model prediction. The basis of this calculation is a transformation that converts a multidimensional integral over all the uncertain model inputs to a one-dimensional integral. Specifically, a search curve, which scans the whole parameter space, is constructed in order to avoid the multidimensional integration. A decomposition of the Fourier series representation is used to obtain the fractional contribution of the individual input variables to the variance of the model prediction.

The analysis is divided into four steps:

- construction of ranges and distributions for the x_j and development of the expected value and variance of y formally in terms of integrals defined on the input parameter space Ω ;
- transformation of the multidimensional integral defined on Ω to a one-dimensional integral;
- estimation of expected value and variance for y ;
- estimation of the sensitivity indices.

The FAST sensitivity indices of the first order are calculated using the terms in the Fourier decomposition of the model output.

An extension of FAST [14] can give, for the same set of model evaluations (i.e. at no extra computational cost) both the first order sensitivity indices and the total ones.

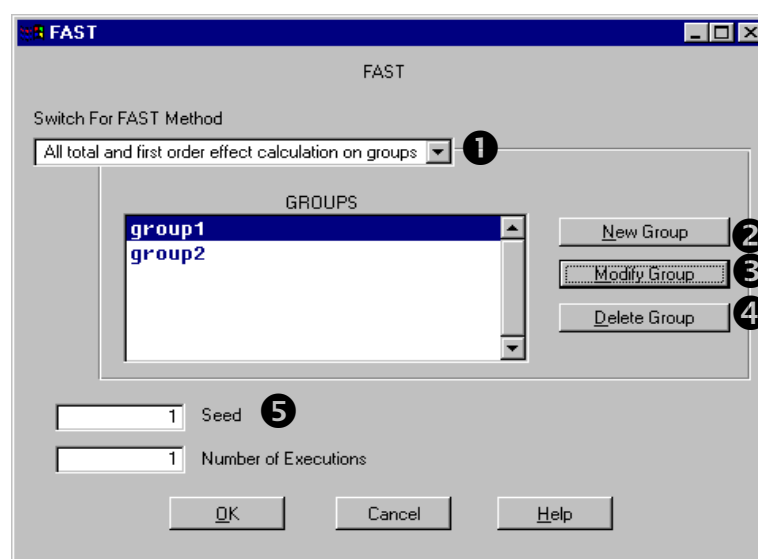
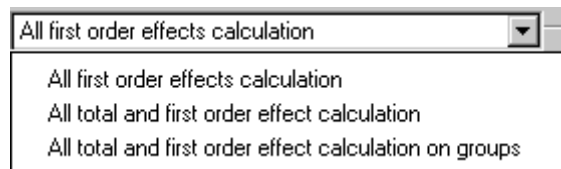
Fast method panel

Figure 20 Fast method panel

- 1 **Select the indices to estimate.** the user can select which FAST sensitivity effects to estimate. The classical FAST method estimates the first order effects. The extension of FAST computes first order effects and total effects. This technique can also be used by grouping sub-sets of factors together.



**Fast method panel:
Groups area**

This area is related to the partition of input factors into distinct groups to compute the partial variances. Enabled only after the selection of “All total and first order effect calculation on groups” in the Effects selection combo.

- 2 **New Group button.** Button for adding a new group. A window, in which the user must insert the name of the new group, will be visualised.
- 3 **Modify Group button.** This button allows the user to insert or delete an item to and from the selected group.
- 4 **Delete Group button.** This button deletes the selected group.
- 5 **Number of executions and seed text boxes.** In this fields the user has to put the cardinality of the sample size that will be generated and the input seed to the random number generator algorithm. The random numbers generation in SimLab is implemented using an iterative function based on a user defined starting point. The seed is the primer value for this procedure (a large value (≥ 7 digits) is advised).
- 6 **OK button.** Load the inserted data in the current configuration, closes the window and returns to the previous Prep Main panel.
- 7 **Cancel button.** Closes the window, losing all the changes.
- 8 **Help button.** This button visualises some practical hints that guides the user in a correct application of the method.

IMPORTANT: The Fast method can be used with a set of uncorrelated factors whose cardinality is greater than one.

Fixed samples

Fixed sample generation is a sampling generation completely controlled by the user. This generation could be useful for testing the behaviour of the model in specific areas of the sample space.

In order to make easier the input of the samples by the user, SimLab generates as default the 50th quantiles for each factor in every new sample unit added. Each sample unit created is visualised as a line in the sample matrix of the “fixed” panel. The user can modify every element of this matrix.

Fixed samples method panel

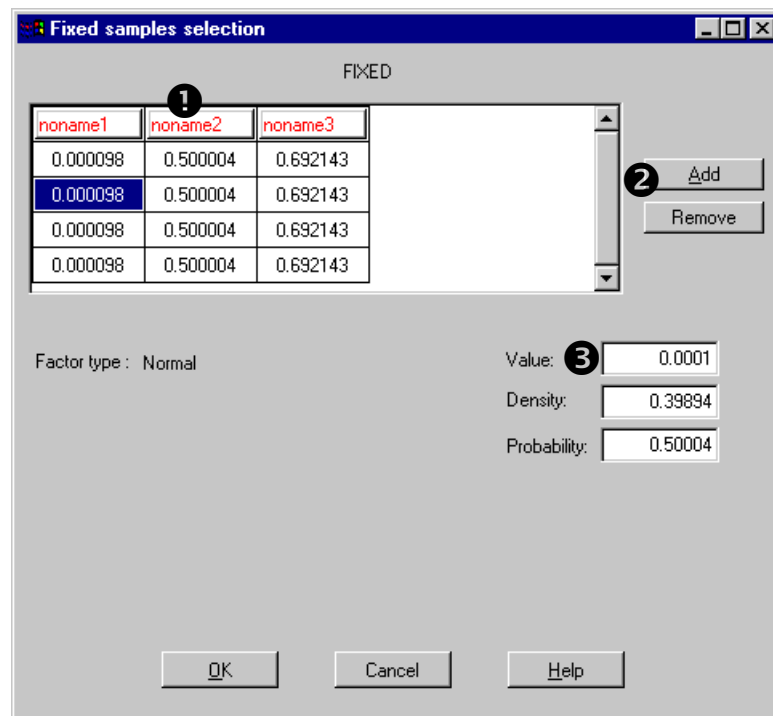


Figure 21 Fixed samples method panel

- ❶ **Sample Matrix.** Sample set matrix.
- ❷ **Add button.** This button adds a new sample unit in the sample matrix. The user can modify every element of the matrix double clicking it.
- Remove button.** This button removes an entire sample unit in the sample matrix.
- ❸ **Distribution area parameters.** This area shows some information that is related to the matrix element selected. The three text boxes show the x value the distribution value and the cumulative value of the element.

IMPORTANT: Fixed samples can be used independently by the factors set selected.

Latin hypercube

Latin hypercube sampling may be considered a particular case of stratified sampling, see [4], [6], [10], [11] and [23].

The purpose of stratified sampling is to achieve a better coverage of the sample space of the input factors. Let the sample space S of the input vector \mathbf{X} be partitioned into L disjoint strata $S_1 \dots S_L$. Represent the size of each S_i , i.e. as $p_i = P(\mathbf{X} \in S_i)$. Obtain a random sample \mathbf{x}_h , $h = 1,$

\dots, n_i from S_i , where $\sum_{i=1}^l n_i = N$. In particular, when $l = 1$, the result is

a random sample over the entire sample space.

In the latin hypercube the range of each input factor, $X_j, j = 1, 2, \dots, k$, is divided into N intervals of equal marginal probability, $1/N$, and one observation of each input factor is made in each interval using random sampling within that interval. Thus, there are N non-overlapping realisations for each of the k input factors. One of the realisations on X_1 is randomly selected (each observation is equally likely to be selected), matched with a randomly selected realisation of X_2 , and so on up till X_k . These collectively constitute a first sample, \mathbf{x}_1 . One of the remaining realisations on X_1 is then matched at random with one of the remaining observations on X_2 , and so on, to get \mathbf{x}_2 . A similar procedure is followed for $\mathbf{x}_3, \dots, \mathbf{x}_N$, which exhausts the observations and results in a latin hypercube sample. The method has the advantage of ensuring input factor has all portions of its distribution represented by input values.

LHS performs better than random sampling when the output is dominated by a few components of the input factors. The method ensures that each of these components is represented in a fully stratified manner, no matter which components might turn out to be important.

LHS is better than random sampling for estimating the mean and the population distribution function. LHS asymptotically is better than random sampling in that it provides an estimator (of the expectation of the output function) with lower variance. In particular the closer the output function is to being additive in its input variables, the more reduction in variance.

Latin Hypercube panel

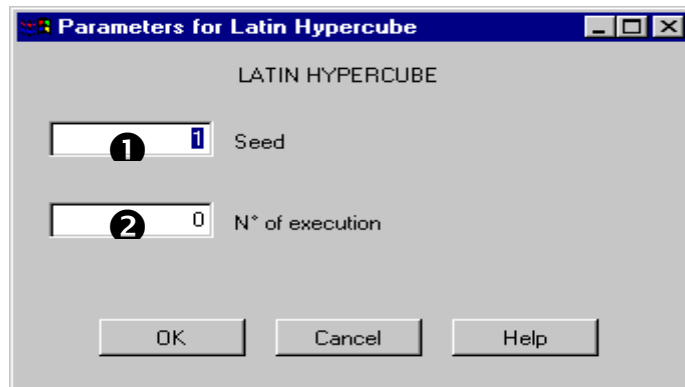


Figure 22 Latin Hypercube panel

- ❶ **Seed text box.** The random numbers generation in SimLab is implemented using an iterative function based on a user defined starting point. The seed is the primer value for this procedure (a large value (≥ 7 digits) is advised).
- ❷ **Number of executions text box.** In this text field the user must insert a number of model executions. This number should not be smaller than $3/2$ the number of input factors; possibly a larger value (e.g. 10 times the number of factors) should be used.

IMPORTANT: Latin hypercube can be used independently by the factors set selected.

Morris

The guiding philosophy of Morris method is to determine which factors may be considered to have effects, which are negligible, linear and additive, or non-linear or involved in interactions with other parameters, see reference [13].

The experimental plan is composed of individually randomised ‘one-factor-at-a-time’ experiments, in which the impact of changing the value of each of the chosen factor is evaluated in turn.

The number of model executions is computed as $r \times (k + 1)$, where r is the number of trajectories (successions of points starting from a random base vector in which two consecutive elements differ only for one component) and k the number of model input factors.

For each factor, the Morris method operates on selected levels. These levels correspond to the quantiles of the factor distribution. In particular:

For 4 levels, the 12.50th, 37.50th, 62.50th and 87.50th quantiles are taken.

For 6 levels, the 8.33th, 25.00th, 41.66th, 58.33th, 75.00th, 91.66th quantiles are taken.

For 8 levels, the 6.25th, 18.75th, 31.25th, 43.75th, 56.25th, 68.75th, 81.25th and 93.75th quantiles are taken.

Morris method panel

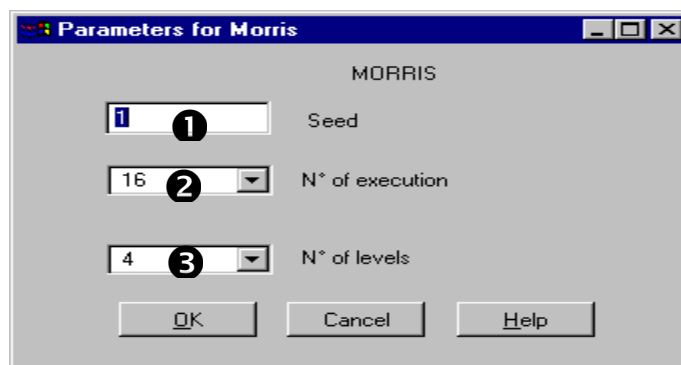


Figure 23 Morris method

- ❶ **Seed text box.** The random numbers generation in SimLab is implemented using an iterative function based on a user defined starting point. The seed is the primer value for this procedure (a large value (≥ 7 digits) is advised).
- ❷ **Number of executions text box.** In this text field the user must select the cardinality of the sample set created. A list of suitable values is available, these values are calculated as (Trajectories) x (Number of model input factors + 1); where (Trajectories) value goes from 4 to 10.
- ❸ **Number of levels text box.** In this field the user can select the number of levels between 4, 6 and 8.

IMPORTANT: The input factor list has to contain uncorrelated items

Quasi-Random LpTau

Quasi-Random LpTau generates quasi-random sequences, see references [2][20] and [22]. The starting point of the generation is the creation of Sobol's sequences. These are uniformly distributed sets of $L=M^N$ points in the N-dimensional unit cube: $I^N = [0,1] \times \dots \times [0,1]$. Each point generated is then used as input to the transformation that calculates the inverse cumulative function of each element of the factor set.

QuasiRandom LpTau panel

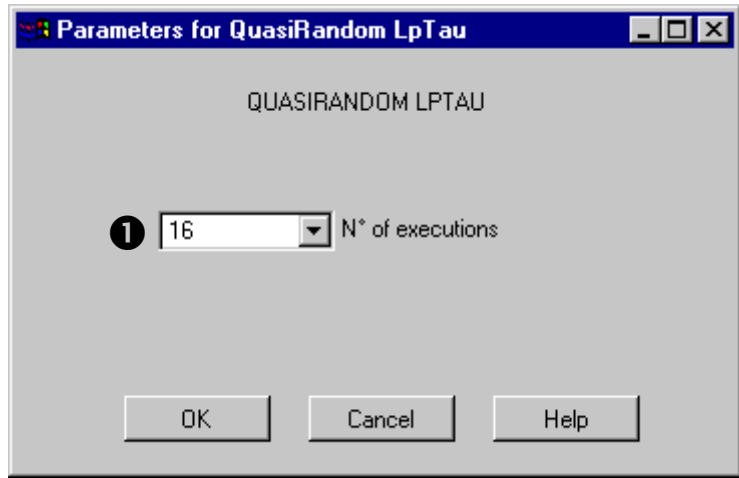


Figure 24 Quasi-Random LpTau method panel

- ❶ **Num of execution text box.** In this text field the user must select the number of model executions. A list of suitable values is available: this list covers all the powers of two from 4 to 20. In Quasi-random LpTau method the number of model executions should be not smaller than the number of variables; possibly a much larger value (e.g. 10 times the number of variables) should be used.

IMPORTANT: Quasi-random LpTau is available only if the inputs factors set is uncorrelated and if the cardinality of the set is less than 52.

Random

In this method a sample (x_1, x_2, \dots, x_m) , of the desired dimension, m , is generated from the joint distribution of the input variables or, when these are independent, from their marginal distributions. Random sampling is also referred to as pseudo random, as the random numbers are machine-generated with deterministic process and not random stricto sensu. From the statistical point of view random sampling has advantages, as it produces unbiased estimates of the mean and the variance of the output variables. In this model the number of model executions should not be smaller than 1.5 the number of variables; possibly a much larger value (e.g. 10 times the number of variables) should be used.

Random method panel

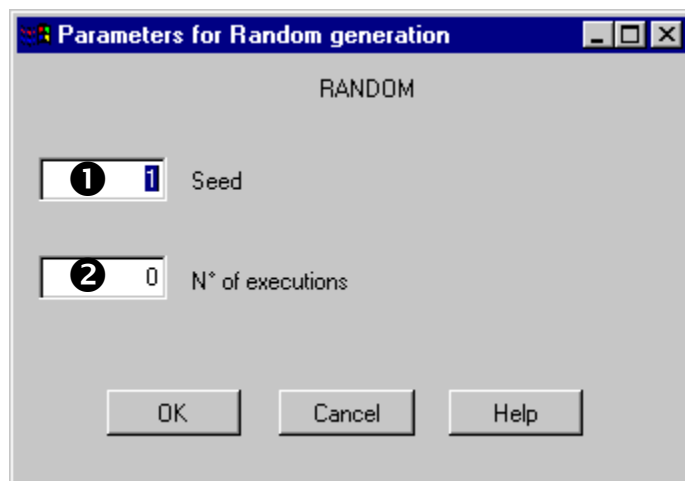


Figure 25 Random method

- ❶ **Seed text box.** The random numbers generation in SimLab is implemented using an iterative function based on a user defined starting point. The seed is the primer value for this procedure (a large value (≥ 7 digits) is advised).
- ❷ **Number of model executions text box.** In this text field the user must insert the cardinality of sample set he wants to generate.

IMPORTANT: Random method can be used independently by the factors set selected

Replicated Latin Hypercube

The replicated Latin Hypercube in SimLab is a sampling method related to the importance measures calculation. The sample design is obtained replicating r times a base sample set created using the Latin Hypercube algorithm. Each replication is created by independent permutations of the columns of the base design (see [24]).

Replicated Latin Hypercube panel

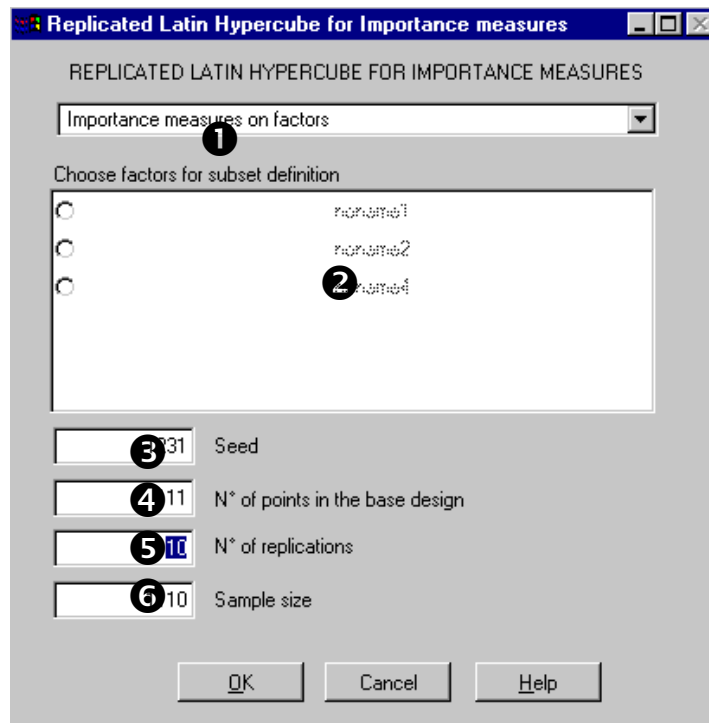


Figure 26: Replicated Latin Hypercube panel

- ❶ **Importance Measures combo box.** This combo box switches the calculation of the importance measures between the evaluation of these indicators on single factors or on two complementary groups.
- ❷ **Factor subset definition area.** This area is activated only with the “Importance measures on a subset” item selected in the **Importance Measures** combo box. The list of all the factors in the configuration is displayed; all the checked factors in this area compose the main subset, all the unchecked factors compose the complementary group.

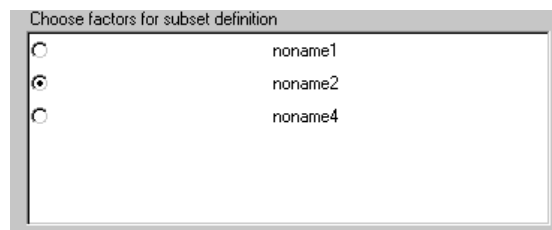


Figure 27: Replicated Latin Hypercube subset selection Area

- ③ **Seed text box.** The random numbers generation in SimLab is implemented using an iterative function based on a user defined starting point. The seed is the primer value for this procedure (a large value (≥ 7 digits) is advised).
- ④ **Number of points in the base design text box.** In this text box the user must insert the cardinality of the base design. This number should not be smaller than $\frac{3}{2}$ the number of input factors; possibly a larger value (e.g. 10 times the number of factors) should be used.
- ⑤ **Number of replications text box.** The number of replications for the method has to be greater than 2.
- ⑥ **Sample size text box.** Inactive text box displaying the sample set size; the value shown is calculated multiplying the base design cardinality by the number of replications.

IMPORTANT: Replicated Latin Hypercube method can be used only with a Iman-Conover correlated factors set and a Stein correlated factors set

Sobol's method

The method of Sobol' [21] is a variance-based technique. The model is $y = f(x_1, x_2, \dots, x_k)$, where x_1, x_2, \dots, x_k are independent input factors and y is the model output. The joint pdf of the input factors is:

$$P(x_1, x_2, \dots, x_k) = \prod_{i=1}^k p_i(x_i) \tag{1}$$

Mean and variance of y are:

$$E(y) = \int \int \dots \int f(x_1, x_2, \dots, x_k) \prod_{i=1}^k p_i(x_i) dx_i \tag{2}$$

$$V(y) = \int \int \dots \int (f(x_1, x_2, \dots, x_k) - E(y))^2 \prod_{i=1}^k p_i(x_i) dx_i = \int \int \dots \int f^2(x_1, x_2, \dots, x_k) \prod_{i=1}^k p_i(x_i) dx_i - E^2(y) \tag{3}$$

If one of the input factors x_j is fixed to a generic value \tilde{x}_j , the resulting (conditional) variance of y is:

$$V(y|x_j = \tilde{x}_j) = \int \int \dots \int (f(x_1, x_2, \dots, \tilde{x}_j, \dots, x_k) - E(y|x_j = \tilde{x}_j))^2 \prod_{\substack{i=1 \\ i \neq j}}^k p_i(x_i) dx_i = \int \int \dots \int (f^2(x_1, x_2, \dots, \tilde{x}_j, \dots, x_k) \prod_{\substack{i=1 \\ i \neq j}}^k p_i(x_i) dx_i - E^2(y|x_j = \tilde{x}_j)) \tag{4}$$

For the purpose of sensitivity analysis one is interested in eliminating the dependence upon the value \tilde{x}_j by integrating $V(y|x_j = \tilde{x}_j)$ over the probability density function of \tilde{x}_j , obtaining

$$E(V(y|x_j)) = \int \int \dots \int f^2(x_1, x_2, \dots, x_j, \dots, x_k) \prod_{i=1}^k p_i(x_i) dx_i - \int E^2(y|x_j = \tilde{x}_j) p_j(\tilde{x}_j) d\tilde{x}_j \tag{5}$$

The dependence on \tilde{x}_j is dropped from the left-hand side of eq. 5, due to the integration. Subtracting Eq. 5 from Eq. 3 one obtains:

$$V(y) - E(V(y|x_j)) = \int E^2(y|x_j = \tilde{x}_j) p_j(\tilde{x}_j) d\tilde{x}_j - E^2(y) \tag{6}$$

Thanks to a statistical identity, the left hand side of Eq. 6 is also equal to $V(E(y|x_j))$, and represents a measure of the sensitivity of y with respect

$$S_j = \frac{V(E(y|x_j))}{V(y)}$$

to factor x_j . If one divides it by the unconditional variance, one obtains the so-called first order sensitivity index. The S_i 's are nicely scaled in $[0,1]$. Eq. 6 is computationally impractical. In a Monte Carlo framework, it implies a double loop: the inner one to compute $E^2(y|x_j = \tilde{x}_j)$, and the

outer to compute the integral over $d\tilde{x}_j$. For this reason the integral in 6 has been rewritten by Ishigami and Homma [8] as:

$$\int E^2(y|x_j = \tilde{x}_j) p_j(\tilde{x}_j) d\tilde{x}_j = \int \left\{ \int \dots \int f(x_1, x_2, \dots, \tilde{x}_j, \dots, x_k) \prod_{\substack{i=1 \\ i \neq j}}^k p_i(x_i) dx_i \right\}^2 p_j(\tilde{x}_j) d\tilde{x}_j = \int \int \dots \int f(x_1, x_2, \dots, \tilde{x}_j, \dots, x_k) f(x'_1, x'_2, \dots, \tilde{x}_j, \dots, x'_k) \prod_{\substack{i=1 \\ i \neq j}}^k (p_i(x_i) dx_i) \prod_{\substack{i=1 \\ i \neq j}}^k (p_i(x'_i) dx'_i) p_j(\tilde{x}_j) d\tilde{x}_j = \int \int \dots \int f(x_1, x_2, \dots, x_j, \dots, x_k) f(x'_1, x'_2, \dots, x_j, \dots, x'_k) \prod_{i=1}^k (p_i(x_i) dx_i) \prod_{\substack{i=1 \\ i \neq j}}^k (p_i(x'_i) dx'_i)$$
 (7)

The expedient of using the additional integration variable primed, allows us to realise that the integral in Eq. 7 is the expectation value of the function F of a set of $(2k - 1)$ factors:

$$F(x_1, x_2, \dots, x_j, \dots, x_k, x'_1, x'_2, \dots, x'_{j-1}, x'_{j+1}, \dots, x'_k) = f(x_1, x_2, \dots, x_k) f(x'_1, x'_2, \dots, x'_{j-1}, x'_j, x'_{j+1}, \dots, x'_k)$$
 (8)

The integral (7) can hence be computed using a single Monte Carlo loop. The following procedure was proposed by Sobol' 1993 and reviewed in Saltelli (2002).

Two input sample matrices \mathbf{M}_1 and \mathbf{M}_2 are generated:

$$\mathbf{M}_1 = \begin{matrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{matrix}, \quad \mathbf{M}_2 = \begin{matrix} x'_{11} & x'_{12} & \dots & x'_{1k} \\ x'_{21} & x'_{22} & \dots & x'_{2k} \\ \dots & \dots & \dots & \dots \\ x'_{n1} & x'_{n2} & \dots & x'_{nk} \end{matrix}$$
 (9)

where n is the sample size used for the Monte Carlo estimate. In order to estimate the sensitivity measure for a generic factor x_j , i.e.

$$S_j = \frac{V(E(y|x_j))}{V(y)} = \frac{(U_j - E^2(y))}{V(y)}$$
 (10)

$$U_j = \int E^2(y|x_j = \tilde{x}_j) p_j(\tilde{x}_j) d\tilde{x}_j$$

we need an estimate for both $E(y)$ and U_j . The former can be either obtained from values of y computed on the sample in \mathbf{M}_1 or \mathbf{M}_2 . U_j can be obtained from values of y computed on the following matrix \mathbf{N}_j :

$$\mathbf{N}_j = \begin{matrix} x'_{11} & x'_{12} & \dots & x_{1j} & \dots & x'_{1k} \\ x'_{21} & x'_{22} & \dots & x_{2j} & \dots & x'_{2k} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x'_{n1} & x'_{n2} & \dots & x_{nj} & \dots & x'_{nk} \end{matrix}$$
 (11)

i.e. by:

$$\hat{U}_j = \frac{1}{n-1} \sum_{r=1}^n f(x_{r1}, x_{r2}, \dots, x_{rk}) f(x'_{r1}, x'_{r2}, \dots, x'_{r(j-1)}, x_{rj}, x'_{r(j+1)}, \dots, x'_{rk})$$
 (12)

If one thinks of matrix \mathbf{M}_1 as the “sample“ matrix, and of \mathbf{M}_2 as the “re-sample” matrix, then \hat{U}_j is obtained from products of values of f computed from the sample matrix times values of f computed from \mathbf{N}_j , i.e. a matrix where all factors except x_j are re-sampled. In this way the computational cost associated with a full set of first order indices S_i is $n(k+1)$. One set of n evaluations of f is needed to compute $E(y)$, and k sets of n evaluations of f are needed for the second term in the product (12).

The problem setting of Sobol’ was that of identifying a subset of the k factors that could account for most of the variance of y . Imagine that the factors have been partitioned into a trial set $\mathbf{u} = (x_{i_1}, x_{i_2}, \dots, x_{i_m})$ and the remaining set $\mathbf{v} = (x_{l_1}, x_{l_2}, \dots, x_{l_{k-m}})$. Then according to Sobol’, the overall effect of the subset \mathbf{u} on the variance of the output can be estimated from:

$$U_v = \int \int \dots \int f(\mathbf{u}, \mathbf{v}) f(\mathbf{u}', \mathbf{v}) d\mathbf{u} d\mathbf{u}' d\mathbf{v} \quad (13)$$

$$V(E(y|\mathbf{v})) = U_v - E^2(y) \quad (14)$$

$$V(E(y|\mathbf{u})) + V(E(y|\mathbf{u}, \mathbf{v})) = V(y) - V(E(y|\mathbf{v})) \quad (15)$$

In Equation (15), $V(E(y|\mathbf{u}))$ is the first order effect of the set \mathbf{u} , while $V(E(y|\mathbf{u}, \mathbf{v}))$ is the interaction term between the sets \mathbf{u} and \mathbf{v} . If $V(y) \cong V(E(y|\mathbf{v}))$, then \mathbf{u} is non-influent, and all factors in \mathbf{u} can be fixed in a subsequent analysis of the model. Formula (13) shows the same expedient of the additional integration variables already described. The Monte Carlo estimate of U_v is:

$$\hat{U}_v = \frac{1}{n-1} \sum_{r=1}^n f(x_{r_{i_1}}, x_{r_{i_2}}, \dots, x_{r_{i_m}}, x_{r_{l_1}}, x_{r_{l_2}}, \dots, x_{r_{l_{k-m}}}) f(x'_{r_{i_1}}, x'_{r_{i_2}}, \dots, x'_{r_{i_m}}, x_{r_{l_1}}, x_{r_{l_2}}, \dots, x_{r_{l_{k-m}}}) \quad (16)$$

i.e. to estimate the total effect of set \mathbf{u} one must now re-sample the variables in the set \mathbf{u} . One can easily see that (12) is a particular case of (16). Error estimates for \hat{U}_j ’s are discussed in the original reference of Sobol’.

Equation (15) is a particular case of a general variance decomposition scheme proposed by Sobol’, whereby the total unconditional variance can be decomposed as:

$$V(y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \dots + V_{12..k} \quad (17)$$

where

$$V_i = V(E(Y|x_i))$$

$$V_{ij} = V(E(Y|x_i, x_j)) - V_i - V_j$$

and so on. The development in (17) contains k terms of the first order V_i , $\frac{k(k-1)}{2}$ terms of the second order V_{ij} and so on, till the last term of

order k , for a total of $2^k - 1$ terms. The V_{ij} terms are the second order (or two-way) terms, analogous to the second order effects described in experimental design textbooks. The V_{ij} terms capture that part of the effect of x_i and x_j that is not described by the first order terms. Sobol's version of formula (17) is based on a decomposition of the function f itself into terms of increasing dimensionality, i.e.:

$$f(x_1, x_2, \dots, x_k) = f_0 + \sum_i f_i + \sum_i \sum_{j>i} f_{ij} + \dots + f_{12\dots k} \quad (18)$$

where each term is function only of the factors in its index, i.e. $f_i = f_i(x_i)$, $f_{ij} = f_{ij}(x_i, x_j)$ and so on. Decompositions (17, 18) are unique provided that the input factors are independent and that the individual terms $f_{i_1 i_2 \dots i_s}$ in (18) are square integrable and have zero mean over the domain of existence.

Homma and Saltelli (1996) introduced the new estimate U_{-j} :

$$\hat{U}_{-j} = \frac{1}{n-1} \sum_{r=1}^n f(x_{r1}, x_{r2}, \dots, x_{rj}, \dots, x_{rk}) f(x_{r1}, x_{r2}, \dots, x_{r(j-1)}, x'_{rj}, x_{r(j+1)}, \dots, x_{rk})$$

As before:

$$V(E(y|\mathbf{x}_{-j})) = \hat{U}_{-j} - \hat{E}^2(y) \quad (19)$$

where $V(E(y|\mathbf{x}_{-j}))$ is the total contribution to the variance of y due to non- x_j . This implies that the difference $V(y) - V(E(y|\mathbf{x}_{-j}))$ is equal to the sum of all terms in the variance decomposition (15) that include x_j .

For a case with $k = 3$:

$$S_1^T = \frac{V(y) - V(E(y|\mathbf{x}_{-1}))}{V(y)} = \frac{E(V(y|\mathbf{x}_{-1}))}{V(y)} = S_1 + S_{12} + S_{13} + S_{123} \quad (20)$$

where e.g. $S_1 = \frac{V(E(y|x_1))}{V(y)}$, and analogous expressions can be written for S_2^T, S_3^T . The S_j^T 's are called *Total effect* terms. The total effects are useful to identify input factors that are non-influential. If the total effect for a given input factor is negligible, then that factor can be fixed to any value within its range of uncertainty (and the dimensionality of the space of input can be reduced accordingly).

Observe here that when the model is purely additive, $\sum_{i=1}^k S_i = 1$, while for a given factor x_j an important difference between S_j^T and S_j flags an important role of interactions for that factor on y .

A complete sensitivity analysis can be obtained by estimating all terms in (17), but these are as many as $2^k - 1$. This problem is called *the curse of dimensionality*. The computational cost of estimating all effects in (17) is

in fact as high as $n2^k$, where again n is the sample size used to estimate one individual effect¹.

For this reason we customarily prefer to compute the set of all S_i plus the set of all S_i^T , that gives a fairly good description of the model sensitivities. With the method of Sobol' this entails a computational cost of $n(2k + 1)$ model evaluations, i.e. nk for the first order terms, nk for the total effect terms, plus n for $\hat{E}(y)$.

Technical note: Better estimates for the first order indices are obtained if the $E^2(y)$ term in (10) is estimated as

$$\hat{E}^2 = \frac{1}{n} \sum_{r=1}^n f(x_{r1}, x_{r2}, \dots, x_{rk}) f(x'_{r1}, x'_{r2}, \dots, x'_{rk}) \quad (21)$$

rather than

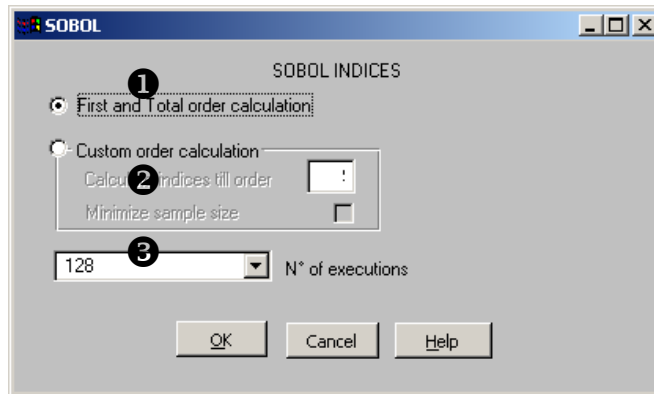
$$\hat{E}^2 = \left(\frac{1}{n} \sum_{r=1}^n f(x_{r1}, x_{r2}, \dots, x_{rk}) \right)^2 \quad (22)$$

Equation (21) is a legitimate estimate of the squared sample mean given the independence of the two sample vectors. It is clear from (10) and (12) that the estimate of S_j goes more naturally to zero for a non-influential factor x_j when (21) is used, as can be seen from:

$$\hat{U}_j - \hat{E}^2(y) = \frac{1}{n-1} \sum_{r=1}^n f(x_{r1}, x_{r2}, \dots, x_{rk}) f(x'_{r1}, x'_{r2}, \dots, x'_{r(j-1)}, x_{rj}, x'_{r(j+1)}, \dots, x'_{rk}) - \frac{1}{n} \sum_{r=1}^n f(x_{r1}, x_{r2}, \dots, x_{rk}) f(x'_{r1}, x'_{r2}, \dots, x'_{rk}) \quad (23)$$

On the other hand the computation of the total effect indices is better achieved using (22). $V(y)$ is computed from \mathbf{M}_1 for all indices. In conclusion, the standard computational strategy so far implemented to compute the full set of total and first order indices entails a total of $n(2k + 2)$ model evaluations, two samples being used to estimate \hat{E}^2 .

¹ $n(2^k - 1)$ would be needed to compute all effects, and n more to compute $\hat{E}(y)$.



- ❶ **Sobol' switch.** To select which of the analyses has to be performed. Estimation of **first order and total effect indices** or estimation **up to a given order (custom)**.
- ❷ **Maximum order of the indices calculated** when the user has selected the **Custom option**.
- ❸ **Number of model executions.** The user selects the sample size n

IMPORTANT: The method of Sobol' as implemented in SimLab can only be used with non-correlated sets of input factors.

2.1.3. OUTPUT FORMAT SELECTION

The last operation for the configuration of a sample generation is the output format selection (see *Figure 28*). Through the ‘Output File area’ of Pre Processor Main panel it is possible to choose the output file for the sample set created and to specify its format. (As default this file has extension *.sam*)

Setting output file panel

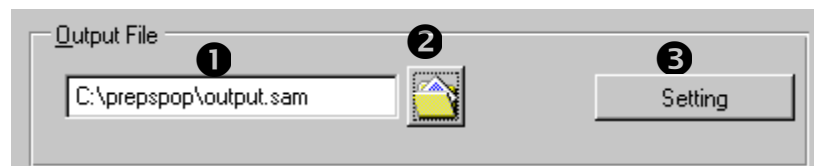


Figure 28 Setting output file panel

- ❶ **Visualisation field.** This field visualises the current file selected for storing the sample generation.
- ❷ **Output file selector button.** This button calls a file selector panel in which the user can navigate the file system and select the output file.
- ❸ **Setting button.** The output format of the sample file can be modified in order to fit better the user needs. The “Output format selection” panel that will be displayed after clicking the setting button allows these modifications.

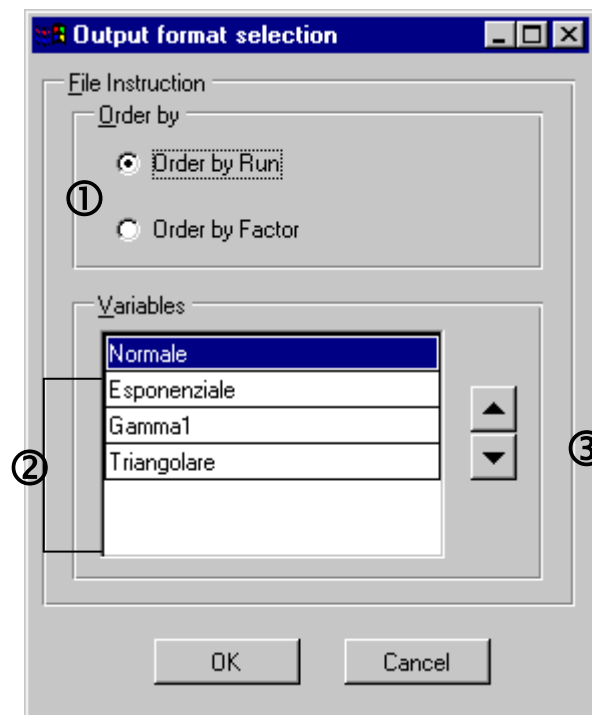


Figure 29 Output format selection

- ① **Order by frame.** The output file can store the sample set generated ordered by run or by factor:
If the user selects the “**Order by run**” check box, the file generated contains NumGen lines (NumGen: cardinality of the sample set), each line with a complete sample unit.
Otherwise if selects ‘**Order by Factor**’ the file generated contains NumFac lines (NumFac cardinality of factor set), each line with all the values generated for a single factor.
- ② **Variables list.** This ordered list visualises the arrangement in which the factor values are written into the output file. This arrangement is used either by the “Order by run” either by the ‘Order by Factor’ policy. The user can modify it using the Arrows buttons.
- ③ **Arrows buttons.** This buttons allow the user to change the arrangement of Variables list. After selecting an input factor the user can move it up or down clicking the relative button.



For more details for the sample file see APPENDIX C

2.1.4. SAMPLE GENERATION

The samples generation is the last step of Statistical Pre Processor module. This action can be accomplished by loading an existing Prep configuration or by creating a new one.

Click on the *Generate* button in the SimLab main panel to start the generation of the sample; the output file selected in the configuration will be filled with the generated sample.

The following status bar will be visualised during the generation. This panel indicates the current status of the job and allows the user to interrupt the generation process by clicking the stop button.

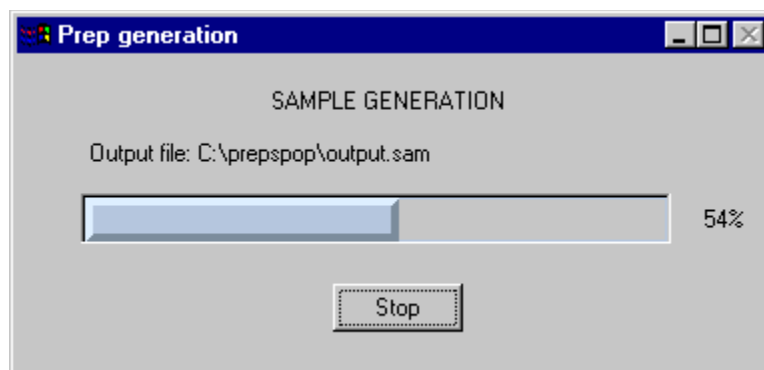



Figure 30 Samples generation

 For more details about the sample format file (as default **.sam** file) see APPENDIX C

Note: For different sensitivity analysis methods, the sample file is generated in a different way. For instance, in the Sobol' method both $f(X_{r1}, X_{r2}, \dots, X_{rk})$ and $f(X'_{r1}, X'_{r2}, \dots, X'_{r(j-1)}, X_{rj}, X'_{r(j+1)}, \dots, X'_{rk})$ are included in the sample file.

2.2. SAMPLE VISUALISATION

Once a sample set has been created or loaded the user can inspect it through three different kinds of graphs: with cobwebs graph, with scatter plot and through histograms. He can also examine all the items generated looking at the tables where the values are listed and inspect the raw correlation imposed through the generation.

A visual analysis of the generated samples could be an easy and powerful way for checking the structure of the outcomes; but this visualisation could also be used in a more powerful way. Using cobwebs or scatterplot the user can select a subset and generate from it a new starting set for the Monte Carlo execution.

The following figure (*Figure 31*) depicts a standard situation after a sample creation. In the “current configuration” visualisation field is summarised the configuration used for the sample generation and in the bottom of the panel the “visualise” button is active.

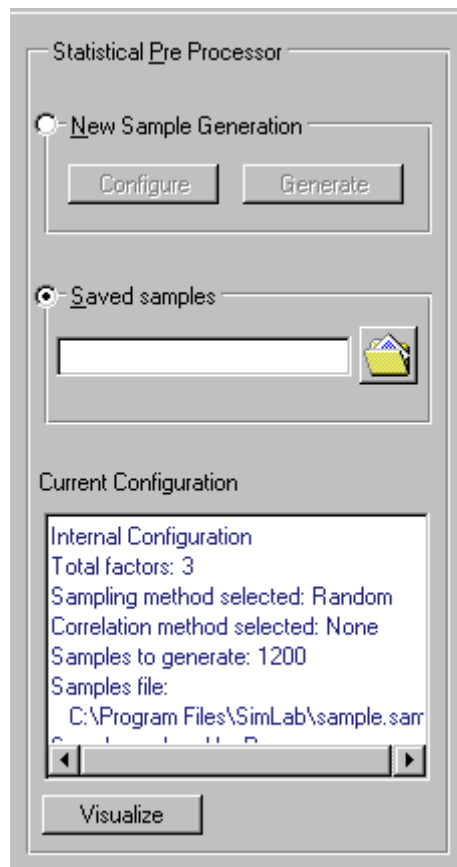


Figure 31 Pre Processor configuration loaded

The “Visualise” button displays the “plots” panel; the user can select the tabs at the top of it to see the graphs selector window, to inspect the tabulated values or to analyse the raw correlation. In the next paragraphs we will analyse the graph selector, in the following the other notebook of the plot window are considered.

Selecting “Tabulated values” every sample unit is listed in the following matrix:

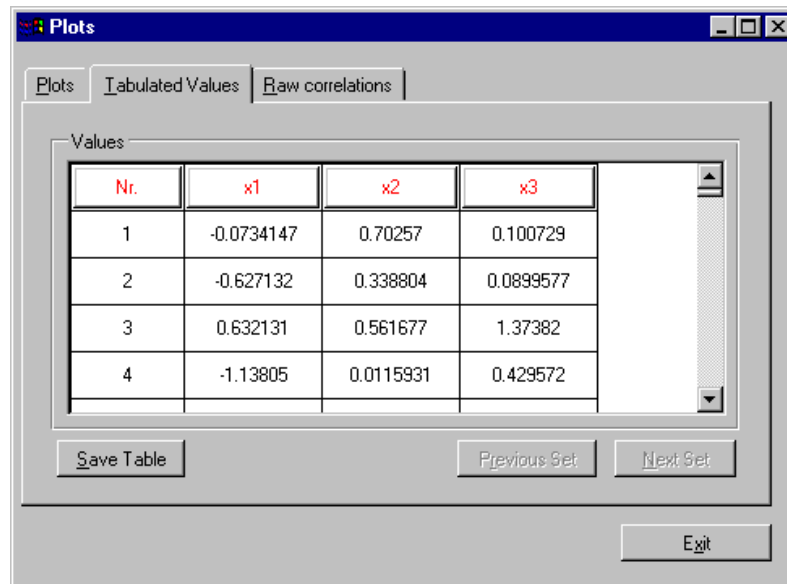


Figure 32 Tabulated values panel

The matrix shows a set of 100 samples at a time, through the “Previous Set” button and the “Next Set” button the user can run through the entire sample generated.

Selecting “Raw correlations” the effective correlation of the sample set is displayed

Raw correlation panel

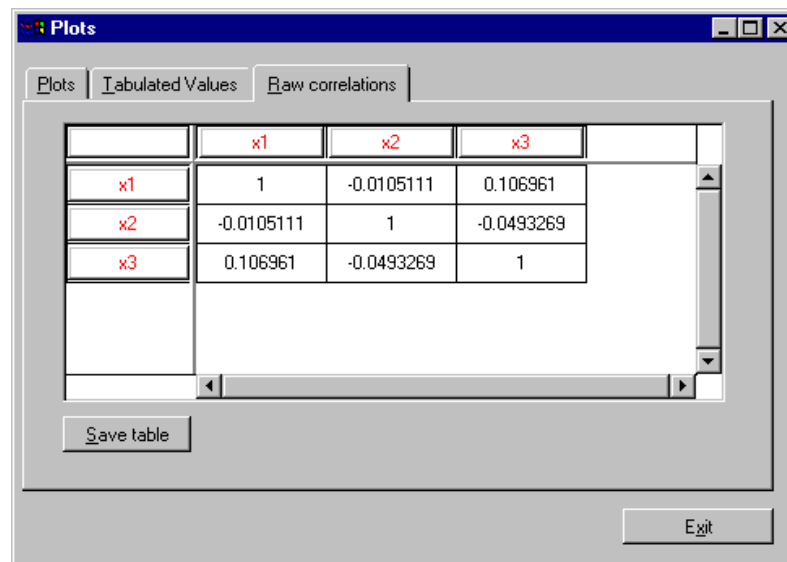


Figure 33 Correlation on sample data

The “Save table” buttons in the previous panels save in a text file the contents of the displayed matrixes.

2.2.1. SAMPLE VISUALISATION AND GENERATION WITH COBWEB PLOTS

The cobwebs plots peculiarity is the ability of displaying in two dimensions higher vectorial spaces.

Plots panel

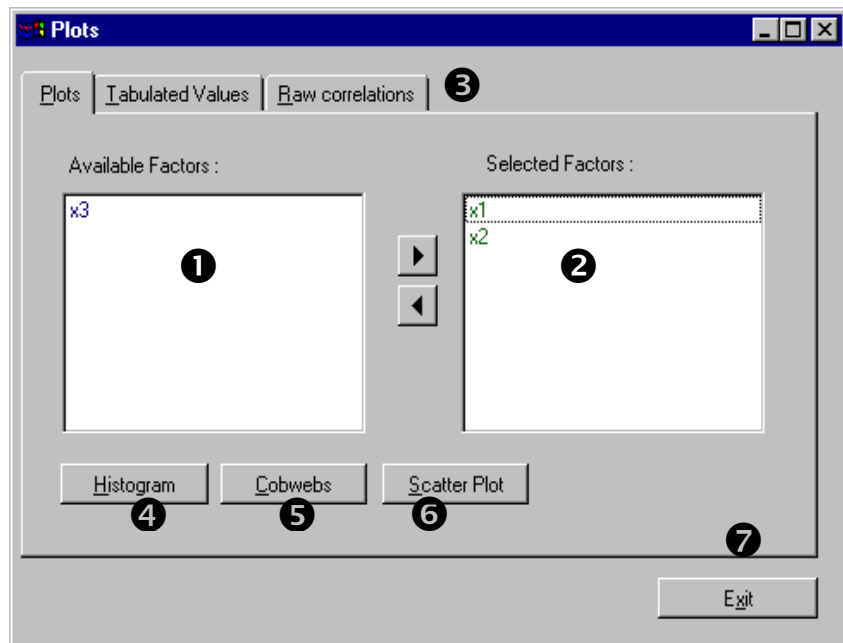


Figure 34 Plots panel

- 1 Available Factors list.** This list prints out the available factors that can be selected for the plots.
- 2 Selected Factors list.** List of factors already chosen by user for the visualisation.
- 3 Move buttons.** Through these buttons the user can add or remove elements from the “Available Factors list” to the “Selected Factors list”. The user can also move one factor from one list to the other by double clicking it.
- 4 Histogram button.** This button recalls the histogram panel, loading the selected factors data (activated if at least one element belongs to the “Selected Factors list”).
- 5 Cobwebs button.** This button shows the factors of the “selected factor” list through cobwebs graph. This button is active only if two or more factors are in the ‘Selected Factors list’.
- 6 Scatter Plot button.** Button related to the visualisation of data through the scatter diagram. This button is active only if two factors belong to the ‘Selected Factors list’.
- 7 Exit button** closes the panel.

Figure 35 shows a cobweb graph. In this figure three factors can be identified (the three vertical bars), while the sample units are depicted as polylines touching all the factors.

SimLab doesn't visualise the entire sample set, it selects randomly 100 sample units of what generated.

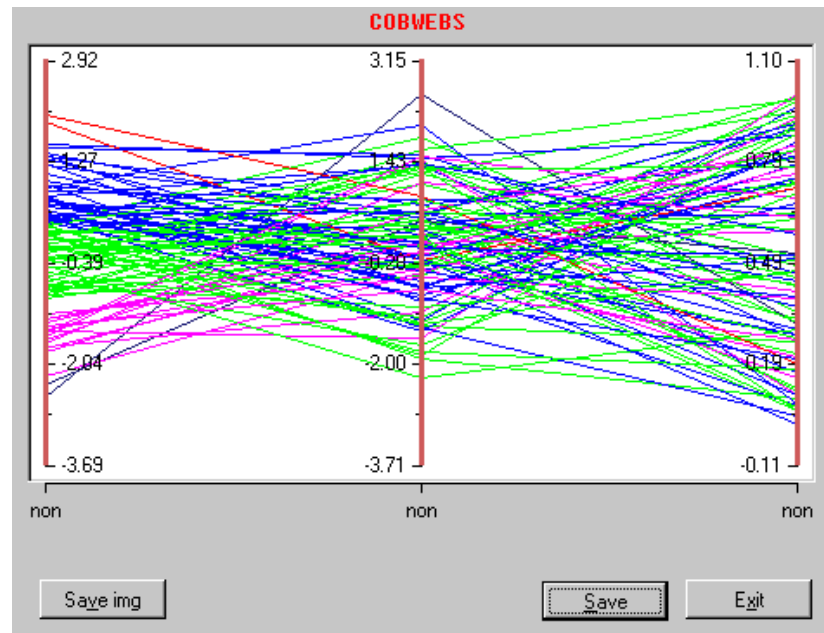


Figure 35 Cobwebs graph panel

The user can select on each vertical bar a sub-interval in order to refine the visualisation. This operation can be accomplished by clicking the mouse left button on the factor bar and drag it till selecting the wished interval. For a more precise interval definition the user can click the right button and the following panel for analytical selection will be displayed:

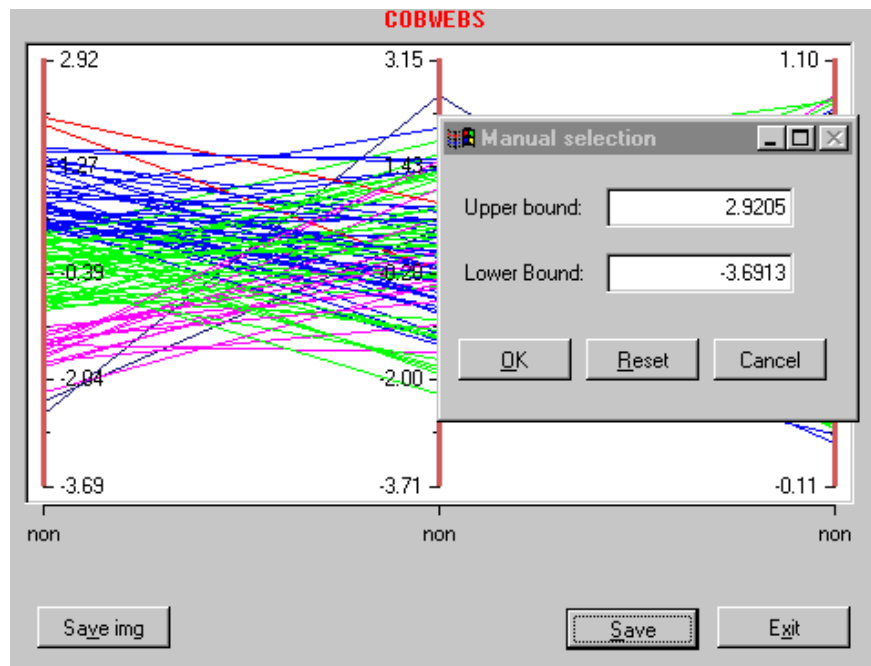


Figure 36 Cobwebs manual selection

The refinement can be useful for checking the behaviour of the sample in a specific area of the sampling space. Furthermore the user can generate a new sample set using the refinement procedure; the save button at the bottom of panel can be clicked anytime for saving a new sample set. This new creation is lead by the selections made on the factor axes: every sample unit in the entire sample set that satisfies the selected ranges in the graph will be added to the new generation.

2.2.2. SAMPLES VISUALISATION AND GENERATION WITH SCATTER PLOT

A different way of displaying the sample is through scatter plots. The user can select this modality in the “plots panel” after the selection of two factors. A subset of 1000 points of the entire set is displayed.

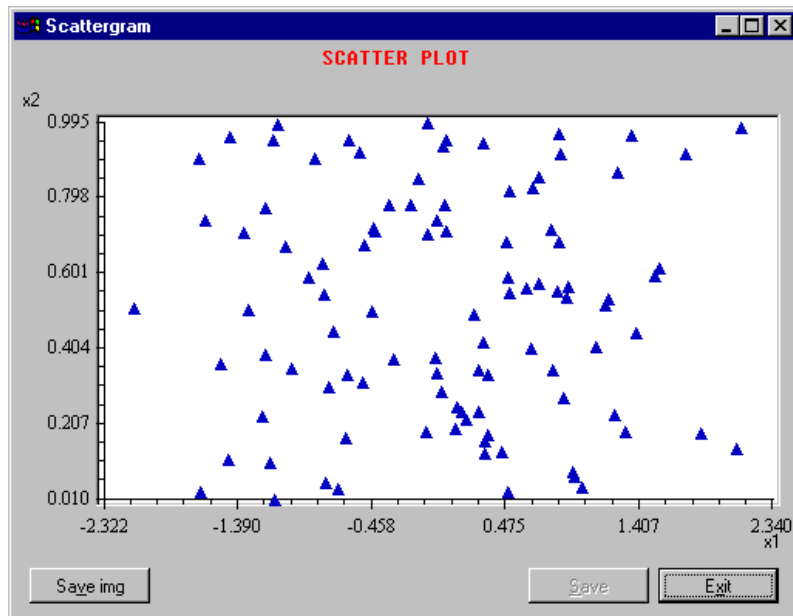


Figure 37 Scatter Plot diagram

As in the cobwebs panel the user can save a subset of the scatter plot graph as a new sample set. He can select a rectangular region using the mouse buttons (clicking the left button in the upper left position and leaving it in the lower right) and save all the sample units belonging to that region in a new sample set.

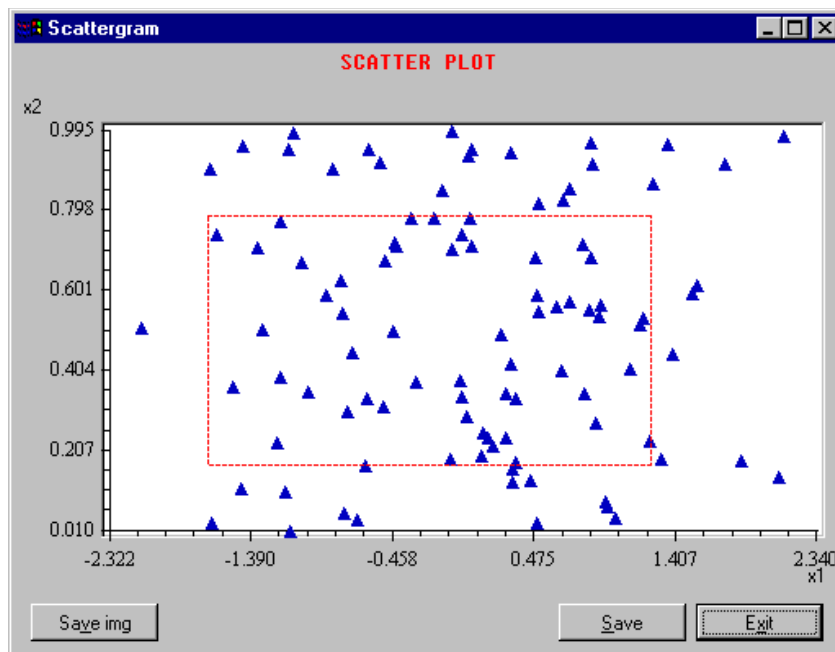


Figure 38 Selection of a region

Figure 38 depicts the selection of a region inside the scatter plot; every region created enables the “save” button in the bottom of the panel. Click “save” and the new sample will be created.

2.2.3 SAMPLE VISUALISATION WITH HISTOGRAMS

Histograms are the third and the last way of displaying the generated sample. As with the other visualisation procedures the user has to add one or more factor in the “Selected Factors list” of the “plot” panels (*Figure 34*) and then click the “Histogram” button. *Figure 39* shows the plots related to three factors.

SimLab divides the dynamic range (maximum value generated minus minimum) of each factor in 100 subintervals before calculating the number of relative occurrences.

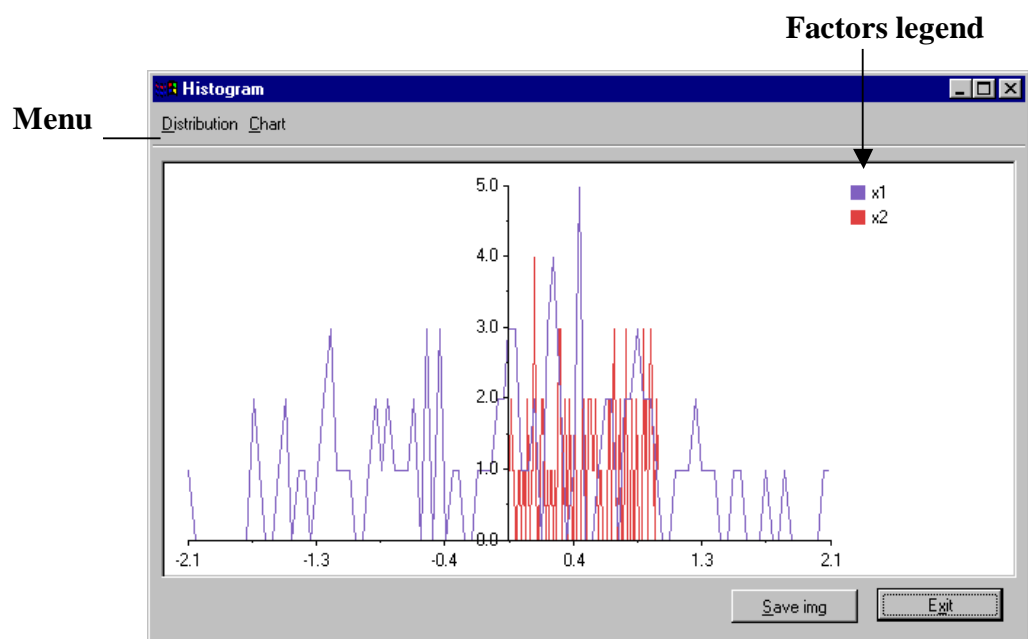
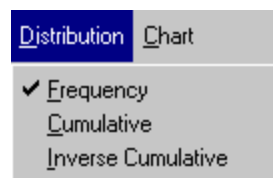


Figure 39 Histogram panel

**Histogram panel:
Menu Distribution**



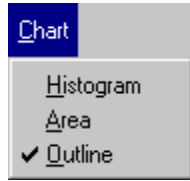
Through this menu the user can select different visualisation for each factor.

Frequency item visualises the frequency distribution of the factors selected.

Cumulative item displays the cumulative distribution.

Inverse Cumulative item displays the inverse cumulative distribution.

**Histogram panel:
Menu Chart**



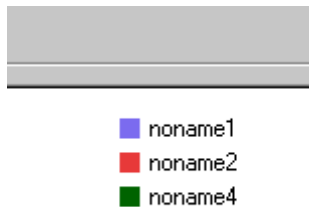
With this menu the user can choose the shape of the graphs in the histogram panel

Histogram item displays the selected factors using histogram layout.

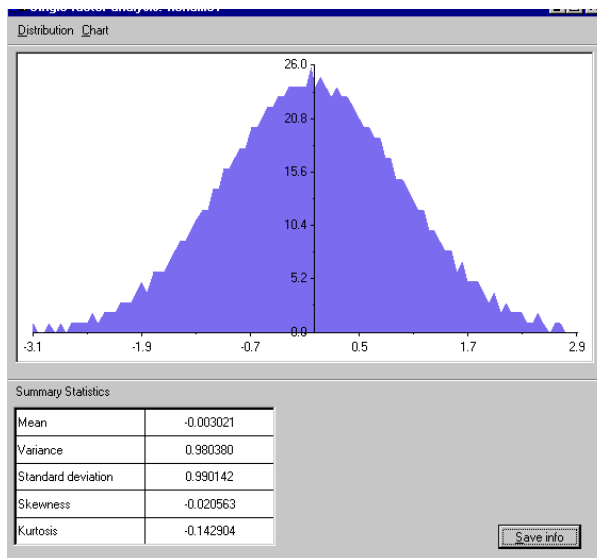
Area item uses filled Areas to represent the selected objects.

Outline item visualises the charts using coloured lines.

**Histogram panel:
Factors legend**



Clicking one of the factors showed in the legend the inspector panel for the item is displayed.



3 MODEL EXECUTION

In this section it is explained how to evaluate the model using the sample set created in the previous phases.

This model execution propagates the sample through the analysis. Conceptually, this task is quite simple. Each element of the sample in the formula:

$$x_i = [x_{i1}, \dots, x_{in}], i = 1, \dots, m,$$

where n is the number of independent (i.e. sampled) factors and m is the sample size, is supplied to the model as input, and the corresponding model predictions are saved for use in later uncertainty and sensitivity studies.

This creates a sequence of results of the form

$$y_i = f(x_{i1}, x_{i2}, \dots, x_{in}) = f(x_i), i = 1, 2, \dots, m, \quad (3.1)$$

Typically, there are many model predictions of interest, in which case y_i would be a vector rather than a single number.

In its simplest form, this step involves little more than putting a “DO loop” around the model within which (1) each sample element is read and supplied to the model as input, (2) the model is evaluated, and (3) the results of each model evaluation are written to a file that is saved after all model evaluations have been completed. In practice, this step can be somewhat more complicated than this. For example, a sampled variable may not be in exactly the form the model takes as input or model predictions may not be in the form desired for subsequent uncertainty and sensitivity analysis. In such cases, a pre-processor and a post-processor can be added to the loop immediately before and immediately after model evaluation to perform the necessary transformations. A more complex situation sometimes arises when the model under consideration is actually a sequence of individual models, each of which supplies input to the next model in the sequence. When each model produces many distinct cases for analysis by the next model, it is sometimes necessary to use a clustering procedure at the interfaces to control the total number of individual cases that are propagated through the entire analysis.

The configuration of the central part of the SimLab Main panel is the way to set the model in the current execution.

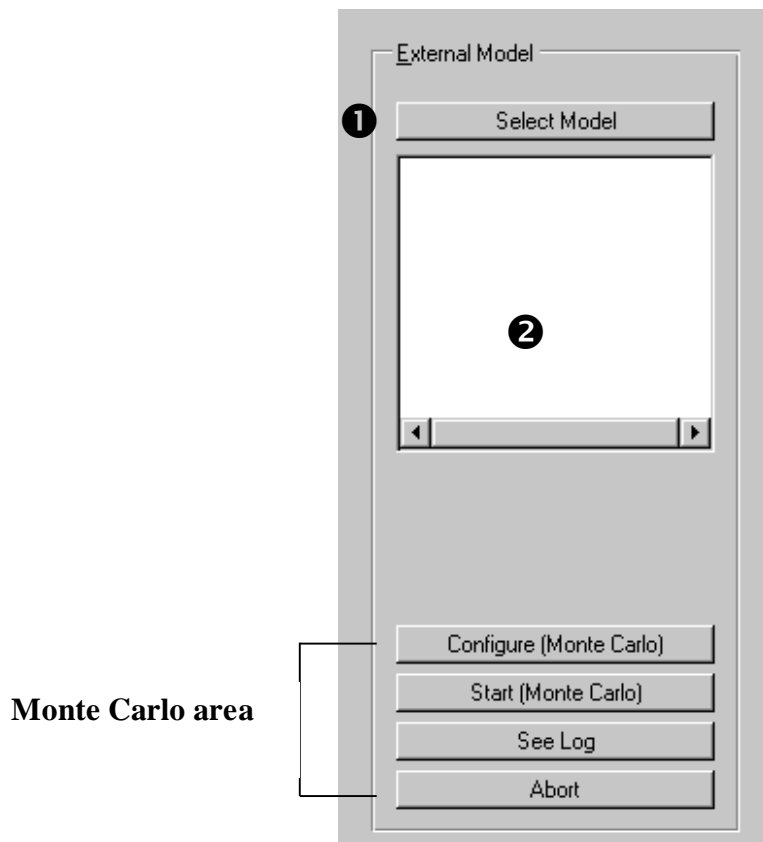


Figure 40 External Model module panel

- ❶ **Select Model button.** The most important setting here is the configuration of the model. The “select model” button opens a new panel in which the user can specify which model to execute.
- ❷ **Visualise field.** This field shows the current model configuration loaded in SimLab.

Monte Carlo area: buttons

Configure (Monte Carlo) button. This button allows the user to activate the configuration of the Monte Carlo execution.

Start (Monte Carlo) button starts a Monte Carlo execution.

See Log button. This button shows a log panel in which the user can open a text file. In this log window the user can load the log file related to the last external model execution.

Abort button stops the current Monte Carlo execution.

3.1 HOW TO SET AND EXECUTE A MONTE CARLO ANALYSIS

When the user clicks the ‘Configure (Monte Carlo) button’ the SimLab Main panel changes status, two yellow arrows are displayed and some buttons are insensitive:

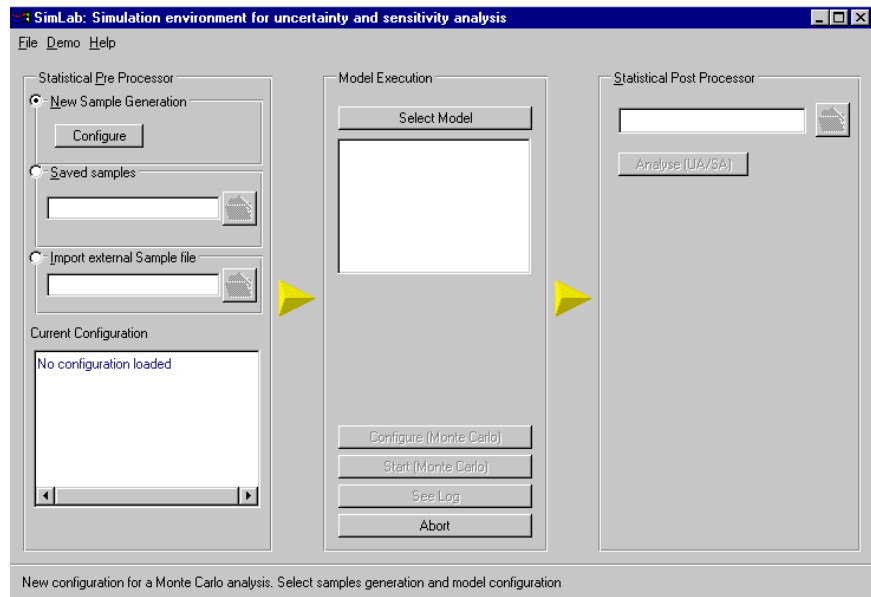


Figure 41 SimLab Main panel for Monte Carlo

The user can now set and execute a complete Monte Carlo analysis; he can select in the Pre Processor panel the input factors or load an old generation. In the former case a new sample set will be generated otherwise the old file is used as input to the model chosen. The setting of the executable file for the model is accomplished clicking the “Select Model” button in the central part of the panel.

By selecting the “Abort” button the user aborts the execution and comes back to the starting SimLab panel.

Clicking the “Select Model” button two types of model could be chosen: an external or an internal model.

In the external modality the user uses a stand-alone application in which he has coded the model to evaluate. This application has to read the sample file and produce an output text file using the format specifications of SimLab.

In the internal model the user can edit a set of equations inside the application and directly evaluate the sample set through them.

3.1.1 EXTERNAL MODEL EXECUTION

The external model execution was studied to allow the link of complex models to the SimLab software. A lot of models are made of complicated equations that can hardly be coded using simple mathematical functions, these models can be easily manage building ad hoc applications.

In the external model execution SimLab launches as a stand alone process the application visualised in the “External Model” textbox passing as argument the sam file generated in the Monte Carlo execution. It then waits for the conclusion of the process and tries to read the file inserted in the “Output File” textbox. The user has to create carefully this file in the format specified by SimLab. If the external executable returns value is different from zero SimLab interprets it as a failure and stops the Monte Carlo execution (see the following picture).

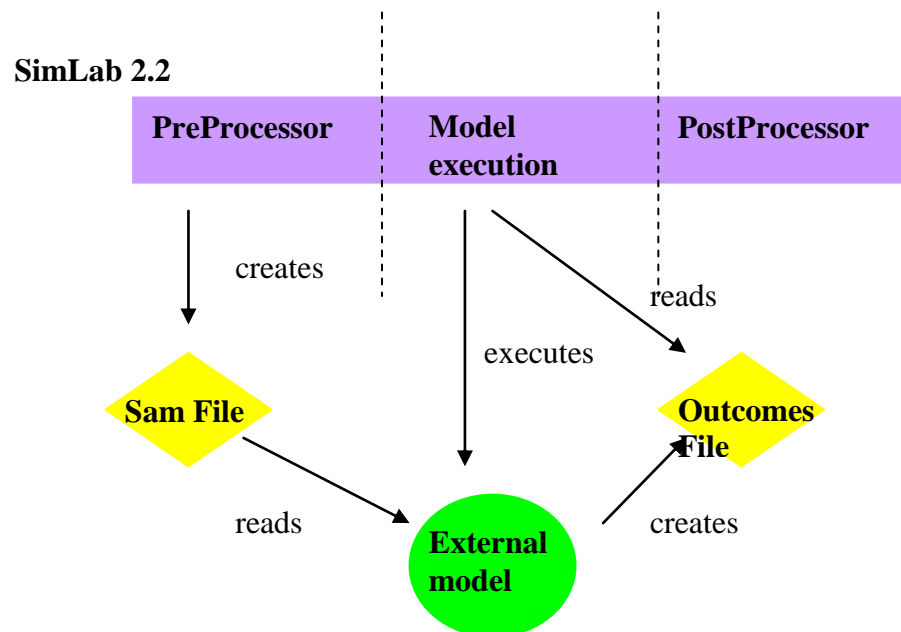


Figure 42: External model execution schema

Model Selection panel: external model

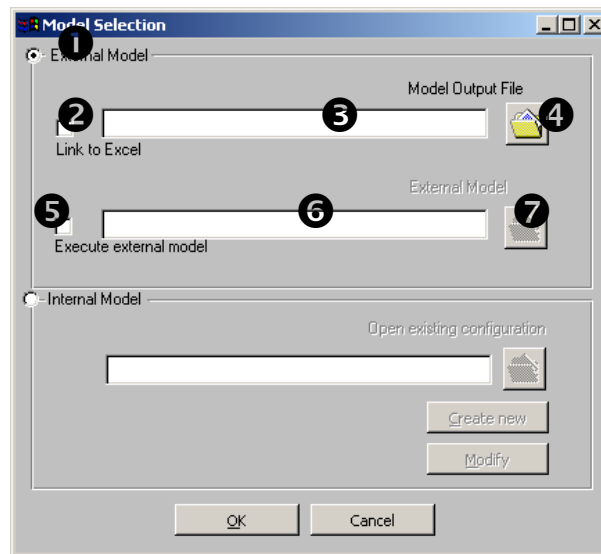


Figure 43: Model selection panel

- 1 External Model.** This switch, if selected, allows specifying an external application that will be launched automatically by Simlab. If you do not want to execute any model but you have an output file available, you can tell SimLab to read directly the output file, yet the checkbox **5** has to be disabled.
- 2 Link to Excel.** This switch links SimLab to Microsoft Excel. First, generate the sample using the statistical pre processor; then prepare an Excel file composed by two Worksheets. Name the first Worksheet as 'Inputs' and leave it blank; name the second Worksheet as 'Outputs'. Write all the output variable names on the first row, and enter the formula of the model in all the outputs cells.

Save this Excel file and specify it in the **output file textbox**. Remember to enable the checkbox **5**. Start the Excel application without opening any Excel file. Click on "Start Monte Carlo" and SimLab will open the Excel file, load the entire sample in the Worksheet "Inputs". SimLab waits till you close the Excel application and then reads the model outputs from the worksheet "Outputs". At the end of this process, the button "Analyse(UA/SA)" is enabled.
- 3 Output file text box.** In this text box the name of the output file created by the external model is specified. See paragraph External Model Output Format File in APPENDIX C for more details about format file description (not for Excel execution).
- 4 Output file selector button.** This button calls a file selector panel where the user can navigate the file system and select the model output file.

- 5 **External execution switch.** To Enable or disable the external model execution.
- 6 **External Model text box.** In this text box the external model executable file is visualised.
- 7 **External Model file selector button.** This button calls a file selector panel where the user can navigate the file system and select the model executable file.

During the execution of the external model in the Monte Carlo simulation the following panel is displayed (not for external execution linked to Excel).

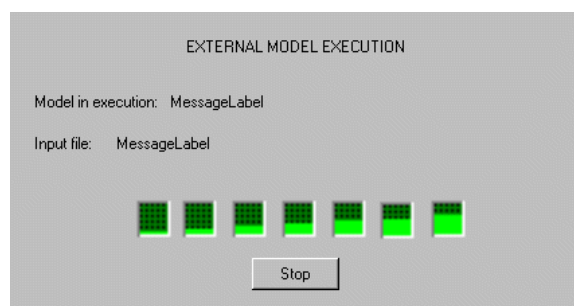


Figure 44: External execution panel

3.1.2 INTERNAL MODEL EXECUTION

Instead of calling an external application for the model execution, the user can create and edit a set of equations inside SimLab and collect them together in a model. The Internal model area is related to this task.

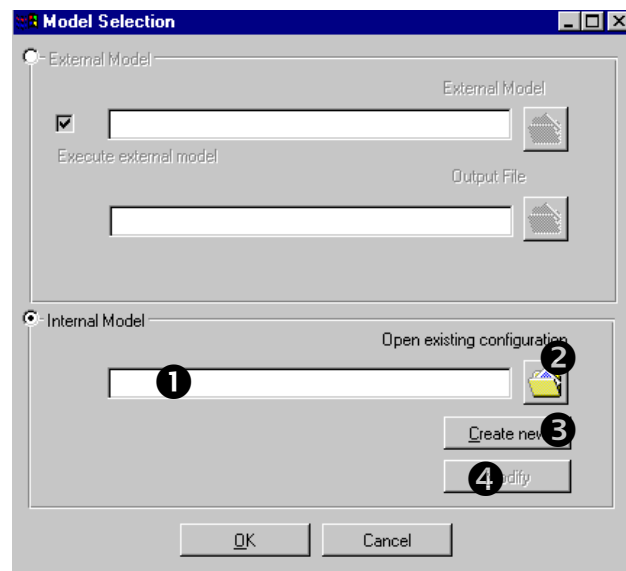


Figure 45: Internal model selection

- ❶ **Internal model file text box.** If the user has loaded an internal model from a previous saved job, this text box contains the file name of the internal model configuration selected.
- ❷ **File selector button.** With this button the user can select a file where an internal model configuration was previously saved. As default the internal model file configuration extension is **.imd**.
- ❸ **Create new button.** This button calls the module for editing a new internal model; the previous configuration is discarded.
- ❹ **Modify button.** Click this button if you want to modify the internal model loaded.

The process of building an internal model is independent from the status of the Pre-Processor configuration, this means that the user is not bound to any constraints regarding the number and the names of the input factors that he wants to use in the model creation. All the compatibility tests for checking the cardinality of the sample factors and the number of the factors presents in the model are evaluated during a Monte Carlo execution.

The main panel for the creation of a new internal model is depicted in the following figure:

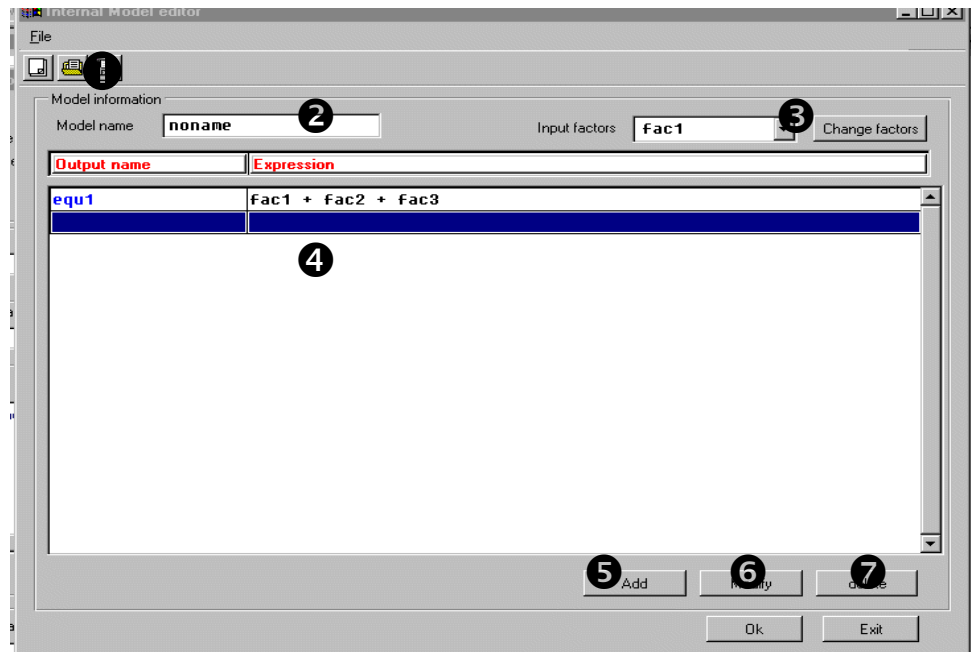


Figure 46: Internal model editor

- ❶ **Save and edit Menu and buttons.** With the menu items and buttons located in this area the user can save or load an internal model configuration.
- ❷ **Model name text box.** This editable text box contains the logical name of the internal model.
- ❸ **List of factors.** The list of factors represents the available factors to be used in the editing of the equations of the model. This list is loaded with the name of the pre-processor factors if a pre-processor configuration is active during the creation of the model. Click the button in the right to modify the list.
- ❹ **Equations area.** This area contains all the equations currently inserted in the internal model.
- ❺ **Add button.** Click this button or double click an empty line in the equations area to add a new equation to the model; the equation panel will be displayed.
- ❻ **Modify button.** This button allows modifying an item in the equations list; click it after the selection of a line in the equations area.
- ❼ **Delete button.** Button for deleting a model equation.

The creation of a model equation is carried on in the expression editor, a specific panel where all the items for building a correct equation are available.

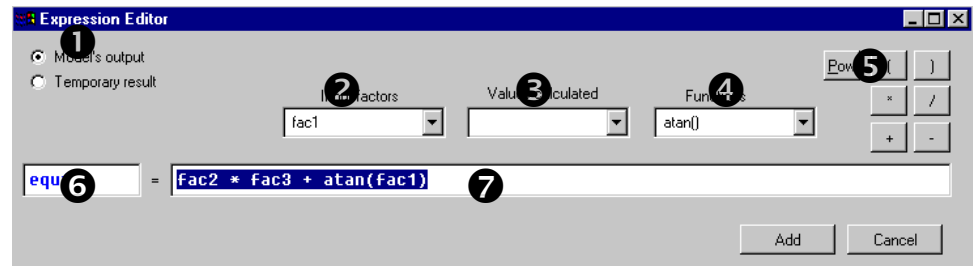


Figure 47: Expression editor

- ① **Equation type switch.** Switching from “**Model’s output**” and “**Temporary result**” the user can select the type of the output calculated by the expression. **Temporary results** marks momentary calculation used for storing intermediate data while **Model’s output** marks models outcomes.
- ② **Input factors list.** Input factors to the model. During the Monte Carlo execution each of these elements will be linked to the pre processor factors and will assume the values of the sample design.
- ③ **Model results list.** List of the expression names already inserted in the current model.
- ④ **Functions list.** Available functions list
- ⑤ **Operators buttons.** Available operations buttons.
- ⑥ **Expression name.** Name of the function calculated by the expression.
- ⑦ **Expression.** Analytical form of the expression.

The factors list, available for the construction of the model equations, can be modified clicking the button “**Change factors**” in the internal model editor. This operation can be useful when an internal model is created when no pre processor configuration is loaded. Changes in the factor list can be accomplished through the “Factor list editor” panel.

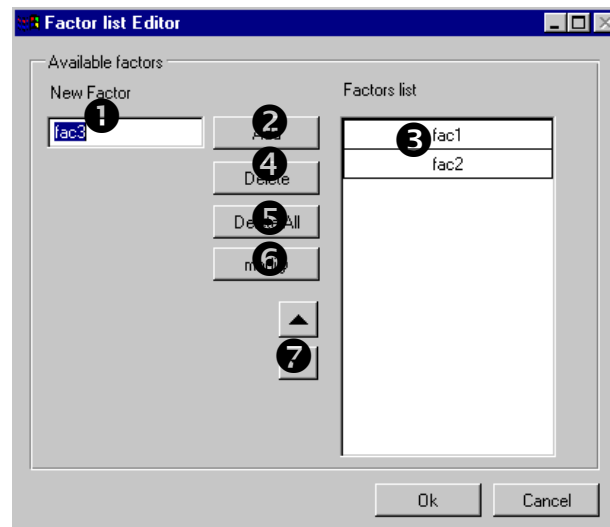


Figure 48: Factor list editor

- ❶ **New factor text box.** Text box for the name of the new factor to add to the factors list
- ❷ **Add button.** Click this button and the name in the text box will be added to the factors list.
- ❸ **Delete button.** Select the item to delete in the factor list and click his button, the item selected will be removed by the list. No factor already inserted in a model expression can be deleted.
- ❹ **Delete all button.** Delete all the items in the factor list. No factor already inserted in a model expression can be deleted.
- ❺ **Modify button.** Select the item to modify in the factor list and click his button, a new panel where the user can modify the factor name will be displayed.

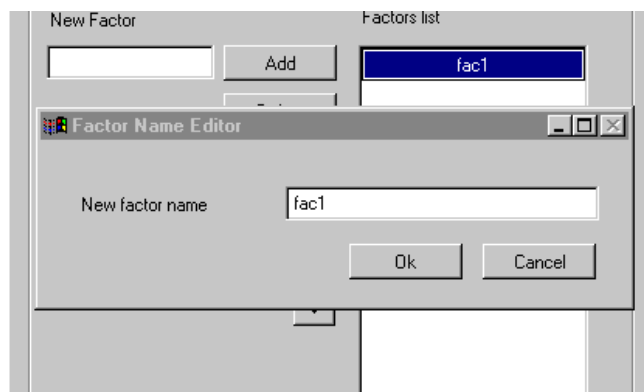


Figure 49: Factor modifier panel

- ❻ **Change factor order buttons.** These buttons allows the user to change the order of the factors available in the factors list.

7 Factor list. List of all the factors in the current model configuration.

At the Start of a Monte Carlo execution with an internal model configuration loaded the SimLab software asks the user to synchronise the factors available in the pre processor configuration to the factors created in the internal configuration. The synchronisation connects the internal model factor to the values generated for the linked factor in the sample design. This operation is necessary because no relation has been supposed during the construction of the internal model between the model and the pre processor configuration loaded.

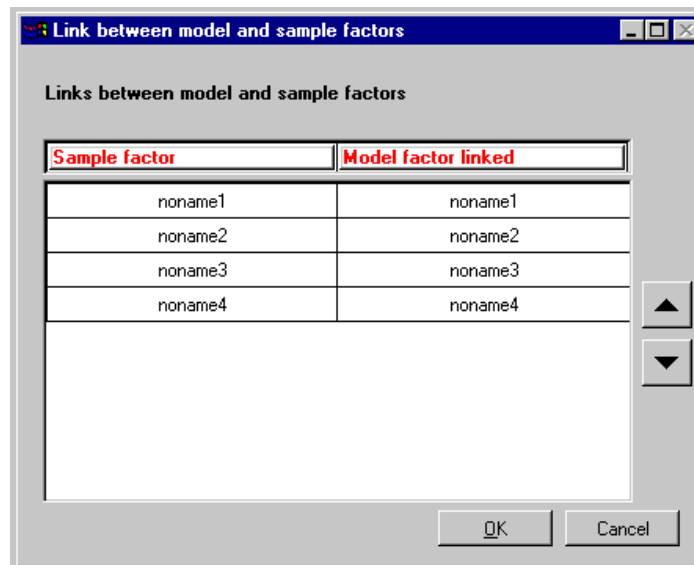


Figure 50: Link factors panel

The buttons on the right part of the panel can be used to change the relation between the sample factors and the model factors. Select one item in the right column and change its position clicking the arrow buttons.

4. MODEL' RESULTS ANALYSIS (Statistical Post Processor)

Statistical Post Processor performs the uncertainty and sensitivity analysis of the model outcomes.

The purpose of uncertainty analysis is to determine the uncertainty in estimates for the dependent variables of interest. The purpose of sensitivity analysis is to determine the relationships between the uncertainty in the resultant dependent variables. Uncertainty analysis typically precedes sensitivity analysis since, if the uncertainty in a dependent variable is under an acceptable bound or within an acceptable range, then there is little reason to perform a sensitivity analysis. Further, when a Monte Carlo analysis is being performed, the generation of summary measures such as means, variances and distribution functions used to represent uncertainty requires little effort once the necessary model evaluations have been performed, see references [14].

When a Monte Carlo study is being performed, propagation of the sample through the model creates a mapping from analysis inputs to analysis results of the form

$$[y_i, x_{i1}, x_{i2}, \dots, x_{in}], \quad i = 1, \dots, m,$$

where n is the number of independent factors and m is the sample size.

Once this mapping is generated and stored, it can be explored in many ways to determine the sensitivity of model predictions to individual input variables.

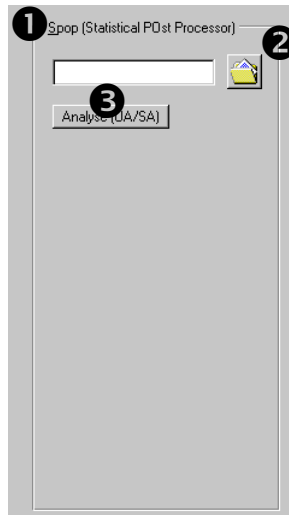


Figure 51 Statistical Post Processor panel

The right part of the SimLab main panel is related to the Uncertainty and Sensitivity analysis. Through the Analyse button illustrated below the user can open an existing UA/SA file stored in the system or execute the last step of a Monte Carlo execution.

SPOP module panel

- ❶ **Visualisation text box** This text box visualises the file selected with the “Open old UA/S” button. This file must contain a saved UA/SA configuration.
- ❷ **Open old UA/SA button.** Through this button the user can navigate the file system and select a saved UA/SA file (**spc** file).
- ❸ **Analyse (UA/SA) button.** This button introduces to the UA/SA analysis; it is available only after the execution of the external model in a Monte Carlo analysis or after loading a saved UA/SA from the file system. This button visualises the Post Processor panel depicted in *Figure 52*.

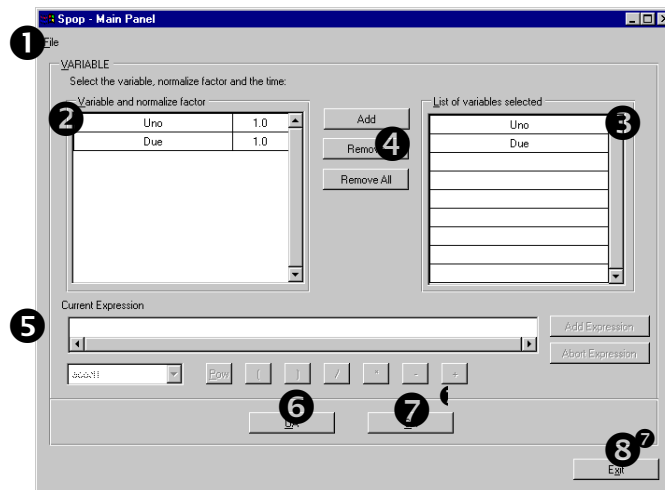


Figure 52 Post Processor Main panel

Post Processor Main panel: File Menu

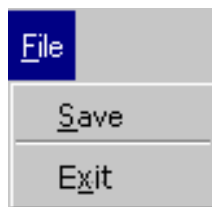


Figure 53 Post Processor File menu

SPOP Main panel

- 1 Save command** saves the current Post Processor configuration in a spc file.

Exit command Closes the window and returns to SimLab Main panel.
- 2 Variable and normalise factor area.** This area lists the output variables computed by the model. Through the add button the user can select one or more of them for the uncertainty and sensitivity analysis. Besides every variable a normalised factor is displayed, this factor will be multiplied by each values of the variable in the analysis phase. This factor can be modified by double click it (only positive values are allowed).
- 3 Selected variables area.** This area contains the variables selected by the user for the uncertainty and sensitivity analysis. The user can add a variable listed in the “Variable and normalise factor area” through the “add” button in the middle of the panel.
- 4 Add button.** The “add” button allows the user to add a new variable to the “selected variables” list (see **3**). By pushing “Add” after selecting an item in “Variable and normalise factor area” the “Select variable for analysis” panel is visualised. Through this panel the user

can analyse the variable as it is or combine it in expression; moreover if the external model is time dependent two time selections are available: “single time” and “multi time”.

Remove button. This button removes the items selected from the “selected variables” area.

Remove All button. This button removes all the items from the “selected variables” area.

Select variable for analysis panel

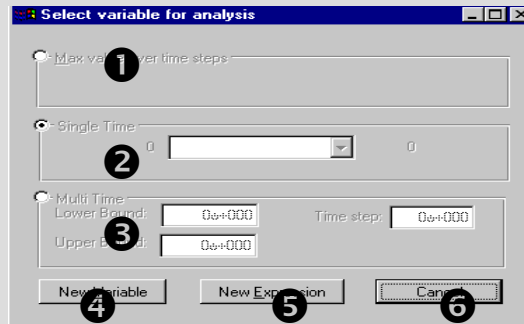


Figure 54 Select variable for analysis panel

- ① **Max value over time steps switch.** (Only available for time dependent models). Selecting the Max value checkbox the user can evaluate the chosen variable at the maximum value for each run independently by the time point of the measurement.
- ② **Single Time switch.** (Only available for time dependent models). Selecting the single time checkbox the user can evaluate the chosen variable at a specific time point. The list of the time points evaluated by the model are available in the combo box in the centre of the frame; otherwise if the user puts a different time value inside the textbox a linear interpolation between model output values of the nearest time points is calculated.
- ③ **Multi Time switch.** (Only available for time dependent models). In this frame the user can select several time points in which the variable will be evaluated. Modifying the lower and upper bound of time evaluation interval and the time step a list of variables at the calculated time points is added to the “Selected variables” area.
- ④ **New Variable button.** This button adds the selected variable into the ‘List of variables selected’ area.
- ⑤ **New Expression button.** This button adds the selected variable into the “Expression” area that will become available for editing.
- ⑥ **Cancel button.** Closes the window and returns to Post Processor panel.

- ⑤ **Expression area.** Area related to the compilation of expressions. SimLab allows the user to create expression using the output model variables and evaluate for them the UA and SA.

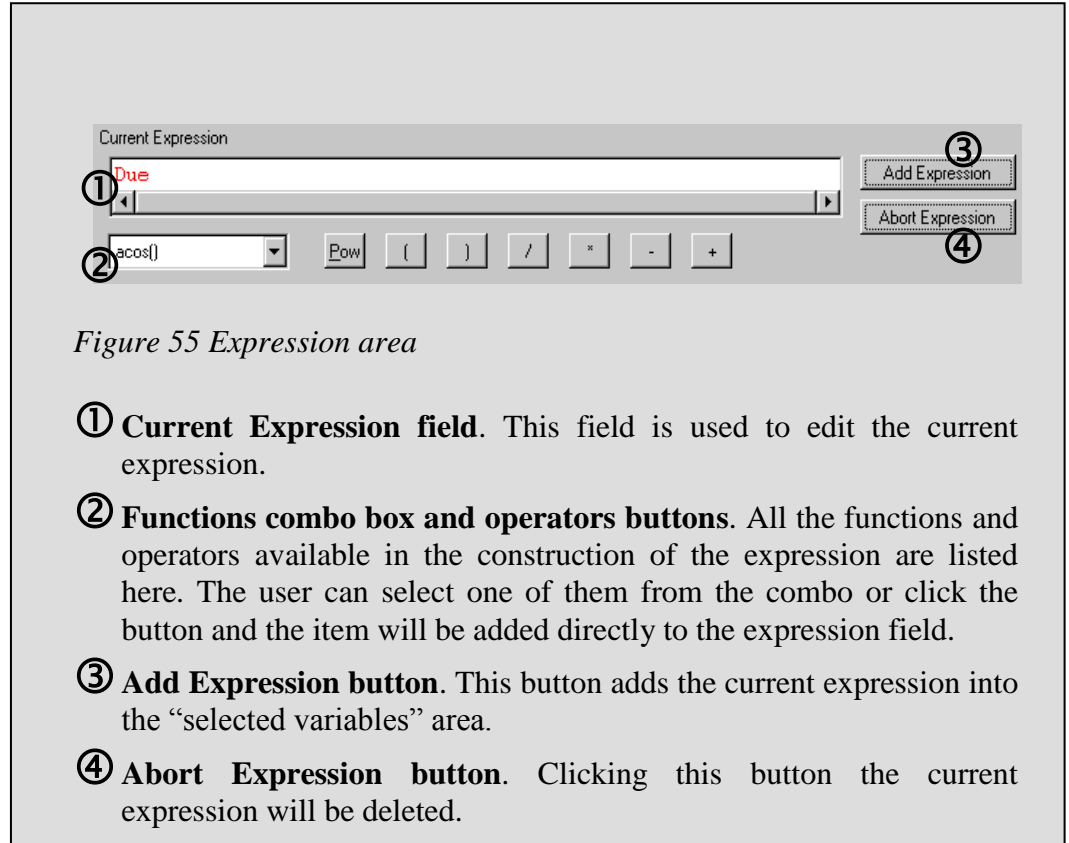


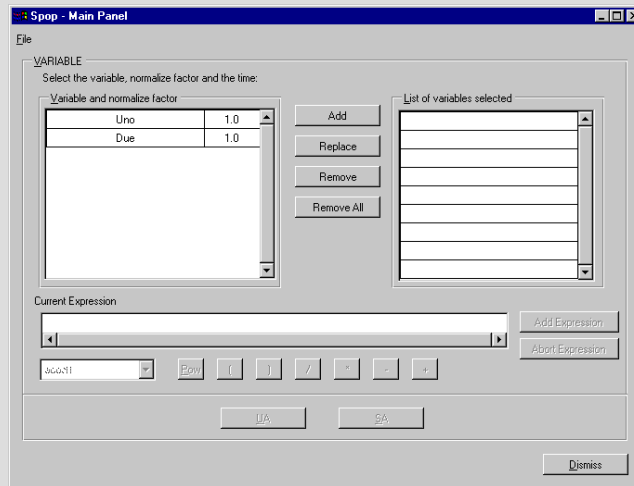
Figure 55 Expression area

- ① **Current Expression field.** This field is used to edit the current expression.
- ② **Functions combo box and operators buttons.** All the functions and operators available in the construction of the expression are listed here. The user can select one of them from the combo or click the button and the item will be added directly to the expression field.
- ③ **Add Expression button.** This button adds the current expression into the “selected variables” area.
- ④ **Abort Expression button.** Clicking this button the current expression will be deleted.

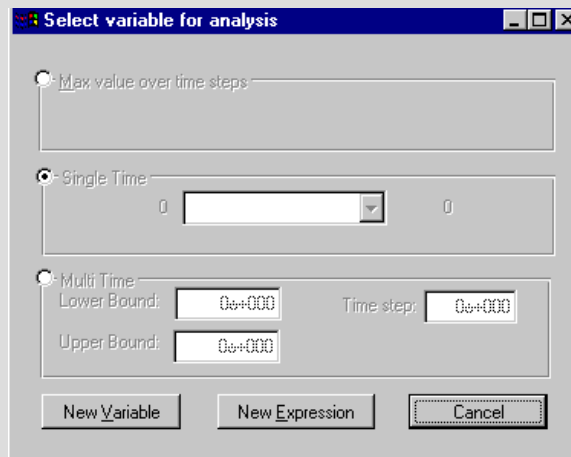
- ⑥ **UA button.** This button calls the procedures and panels related to the uncertainty analysis.
- ⑦ **SA button.** This button calls the procedures and panels related to the sensitivity analysis.
- ⑧ **Exit button** closes the window and returns to SimLab main panel.

Add variables and expression: example

Lets start from the situation depicted in the following panel. The model executed has generated two not time dependent variables (Uno and Due).



Let's suppose that the user wishes to analyse the variables Uno, Due and the following expression: $Uno * \cos(\cos(Due))$. After the selection of the Uno variable and the click on the "Add" button this panel is visualised:



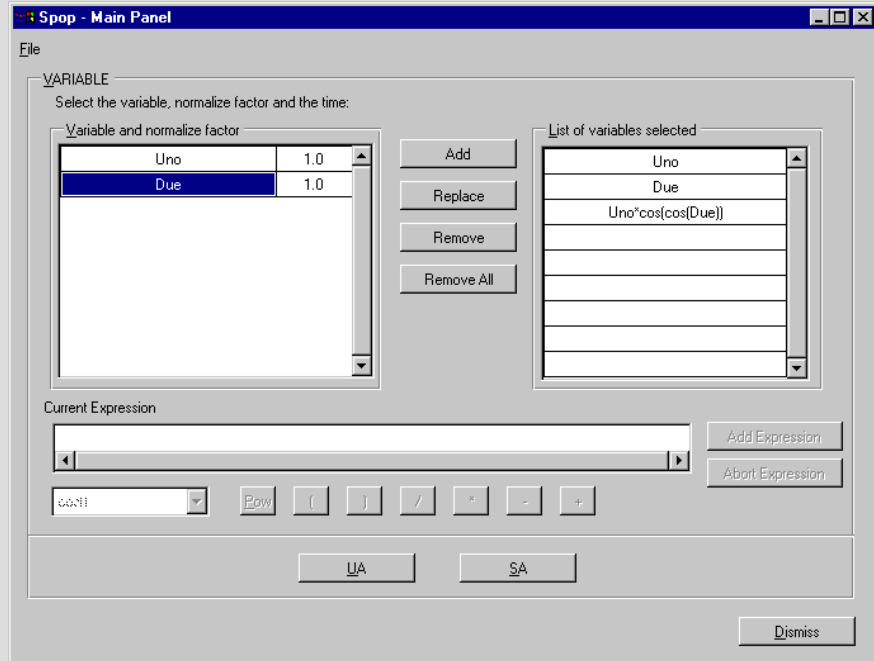
In this example the model's output variables are not time dependent so the "Single Time" frame and "Multi Time" frame are not selectable. Clicking 'New Variable' button the Uno item will be added to the "selected variables" list, the same procedure could be applied to Due. The following panel illustrates the situation:

The screenshot shows the 'Spop - Main Panel' window. At the top, there is a menu bar with 'File'. Below it, the 'VARIABLE' section is active, with the instruction 'Select the variable, normalize factor and the time:'. This section contains two tables: 'Variable and normalize factor' and 'List of variables selected'. The first table has two rows: 'Uno' with a factor of '1.0' and 'Due' with a factor of '1.0'. The second table lists 'Uno' and 'Due'. Between these tables are 'Add', 'Remove', and 'Remove All' buttons. Below the tables is the 'Current Expression' area, which is currently empty. To the right of this area are 'Add Expression' and 'Abort Expression' buttons. Below the 'Current Expression' area is a dropdown menu showing 'cos()' and a set of mathematical symbols: 'Pow', '(', ')', '/', '*', '-', '+'. At the bottom of the window are 'UA', 'SA', and 'Exit' buttons.

To insert the expression $Uno * \cos(\cos(Due))$, select Uno variable into “Variable and normalise factor” area; click the “Add” button and then in the “Select variable for the analysis” panel the “New Expression” button. In the Expression area appears the selected variable (Uno); then insert the maths symbols and trigonometric functions into the Expression area:

The close-up screenshot shows the 'Current Expression' field with the text `Uno*cos(cos())` entered. The 'cos()' dropdown menu is visible below the field, and the mathematical symbols 'Pow', '(', ')', '/', '*', '-', '+' are also visible.

For inserting the Due variable move the cursor inside the brackets, select the Due variable in “Variable and normalise factor” area, click the “Add” button and then in the “Select variable for the analysis” panel the “Add to Expression” button. The Due variable will be inserted into the expression. Now click “Add Expression” button of expression area and the current expression will be inserted into the “selected variables” list.



4.1 UNCERTAINTY ANALYSIS

Once a sample has been generated and propagated through the model, uncertainty analysis is straightforward. Means, variances and distribution functions can be estimated directly from the model predictions.

After the selection of models output variables (at least one variable must be selected), click the UA button. The range on the x-axis of each variable is divided in 100 intervals the frequencies function calculated and then displayed. The result of uncertainty analysis will be visualised in the following panels:

UA main panel

Menu
Legend

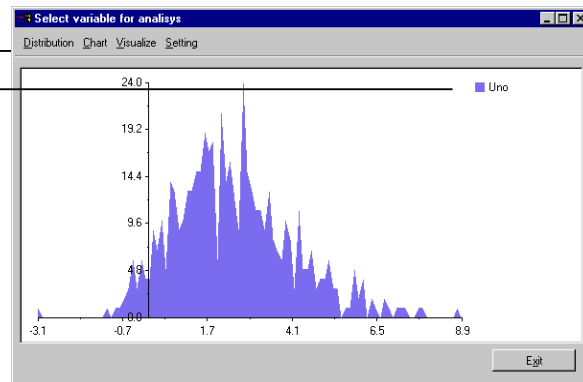


Figure 56 UA Main panel

UA Main panel:
Distribution Menu

Menu:

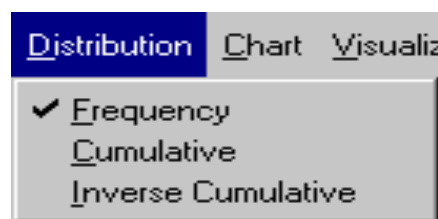


Figure 57 Distribution Menu

Frequency item Selecting this item the user visualises the frequency distribution of the model’s output variables selected. *Figure 56* shows an example.

Cumulative item. This item switches to the cumulative function of the selected variables.

Inverse cumulative item. This item switches to the inverse cumulative function of the selected variables.

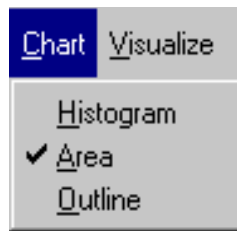


Figure 58 Chart Menu

Histogram item. Setting this item the distribution function selected is displayed using an histogram.

Area item. Through this item the function is shown using a filled area

Outline item shows only the outline of the distribution function of the selected variables.



Figure 59 Visualise Menu

Visualise command this menu allows the selection of the output items to visualise; several variables can be displayed simultaneously, each can be selected by the list of available variables of the “variable selection” panel. (See *Figure 60*)

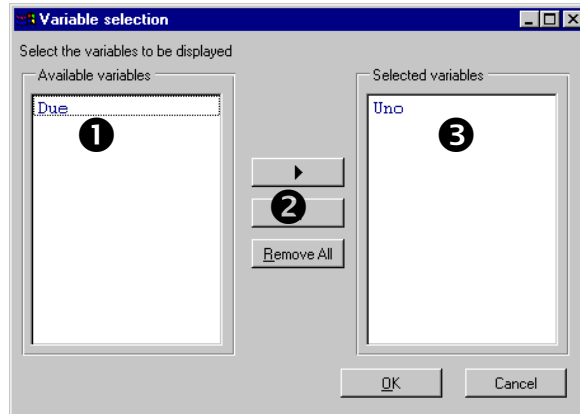


Figure 60: Variable selection panel

Variable selection panel

- ① **Available variables list.** This list contains the available variables for the uncertainty analysis. At first this area contains all the items of “variables selected” area of the Statistical Post Processor panel (see *Figure 52*) except for the first variable which is the visualised as default. For moving the variables from this list to ‘Selected variables’ list and vice versa, the user must select the item to move and use the appropriate arrows buttons or double click it.
- ② **Move buttons.** Through this buttons the user moves the selected variable from ‘Available variables’ area to ‘Selected variables’ area and vice versa, or he removes all variables from the ‘Selected variables’ area.
- ③ **Selected variable area.** This area contains the selected variables. As default it contains only the first variable of ‘List of variables selected’ area of the Statistical Post Processor Main panel.

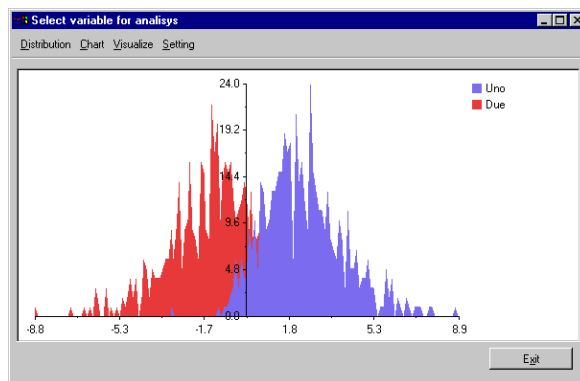


Figure 61: Post Processor panel with two variables

UA Main panel:
Setting Menu



Figure 62: Setting Menu

Choose item allows the user to configure some parameters related to the displaying of the variables analysed and the calculation of some tests.

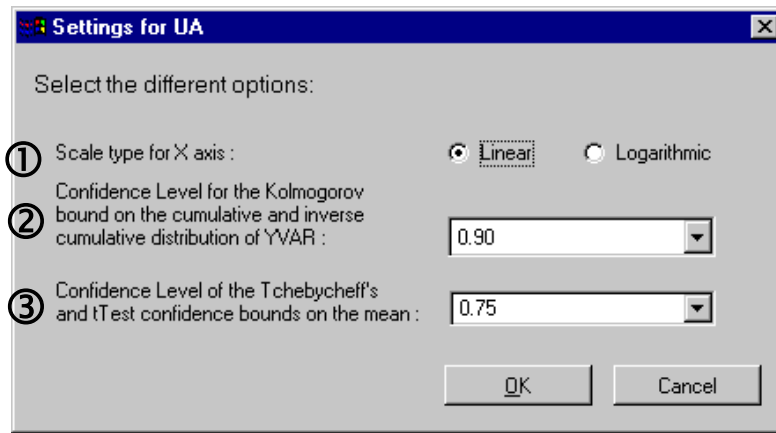


Figure 63 UA setting panel

UA setting panel

- ① **Scale.** This checkbox allows the selection between Linear and Logarithmic scale for the selected variables displayed.
- ② **Kolmogorov-Smirnov test.** Used to test the hypothesis that a sample comes from a particular distribution. The value of the Kolmogorov-Smirnov test is based on the largest absolute difference between the observed and the theoretical cumulative distributions. This goodness-of-fit test, tests whether the observations could reasonably have come from the specified distribution. It is sensitive to any type of difference in the two distributions—shape, location, etc. If $S_N(x)$ is the cumulative distribution of N sample units and $F(x)$ is the theoretical cumulative distribution under hypothesis, then the statistic D_N is defined as:

$$D_N = \max |S_N(x) - F(x)| \quad \text{over all } x.$$

The confidence levels allowed are 0.90, 0.95, 0.98 and 0.99 for the cumulative and inverse cumulative distribution of the selected variables.

- ② **Tchebycheff's and T test confidence bound.** If the Y variable is normal distributed it would be possible to use standard parametric techniques, such as T test, to compute confidence bounds on μ , the population mean of Y , depending on the sample size. When the

distribution of a function is unknown (or simply non-normal) a non-parametric method based on the Tchebycheff's theorem can be used to estimate confidence bounds on the mean. The hypothesis of the tests are:

$$H_0: \mu = \mu_0 \quad \text{against} \quad H_1: \mu \neq \mu_0$$

A confidence interval for a mean specifies a range of values within which the unknown population parameter, in this case the mean, may lie.

The width of the confidence interval gives us some idea about the precision of the estimate. The wider is the interval and the less is the precision of the estimate. We calculated these intervals for different confidence levels, depending on how precise we want to be. We interpret an interval calculated at a 95% level as, we are 95% confident that the interval contains the true population mean. We could also say that 95% of all confidence intervals formed in this manner (from different samples of the population) will include the true population mean.

By selecting the confidence bounds in the previous panel, cumulative and inverse cumulative distribution will be visualised into the selected band.

**UA Main panel:
Legend area**

For each variable displayed in the UA main panel a specific panel could be invoked. By double clicking one item of the legend in the upper left part of the graph area (see *Figure 56*) a new window with only the chosen element is displayed; here all the graphical and tabulated information regarding the specific object are summarised. (See *Figure 64*)

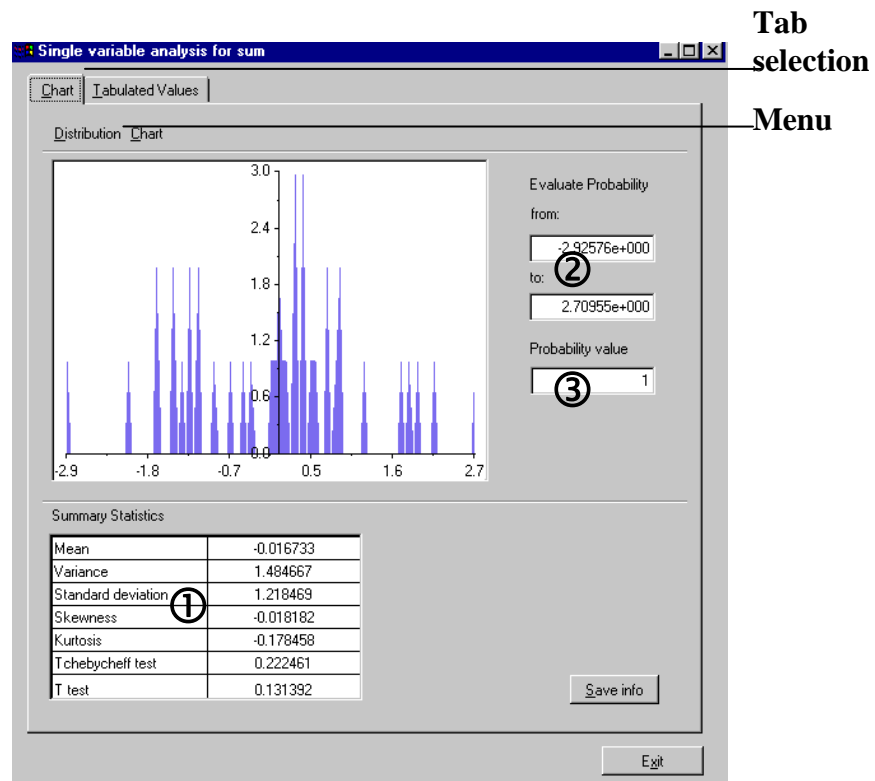


Figure 64: Single variable panel

Menu:

The menu contains the same voices ‘Distribution’ ‘Chart’ and functionality of the UA Main panel menu (see Figure 56 for more details about the Menu functionality).

**Single variable panel:
Tabs selection**

Chart and Tabulated Values Tabs Switching from one tab to the other visual or tabulated information regarding the variable selected are displayed.

**Single variable panel:
Chart tab**

- ① **Statistic area.** This area contains some statistic about the selected variable: the mean value, Tchebycheff bound, Ttest bound variance and standard deviation values.
- ② **Range.** At first these fields visualise the min and the max values assumed by the plot depicted. Changing the values the probability of the variable in the range specified appears in the probability textbox.
- ③ **Prob.** This field visualises the probability value for the variable in the range specified by the text boxes in the left.

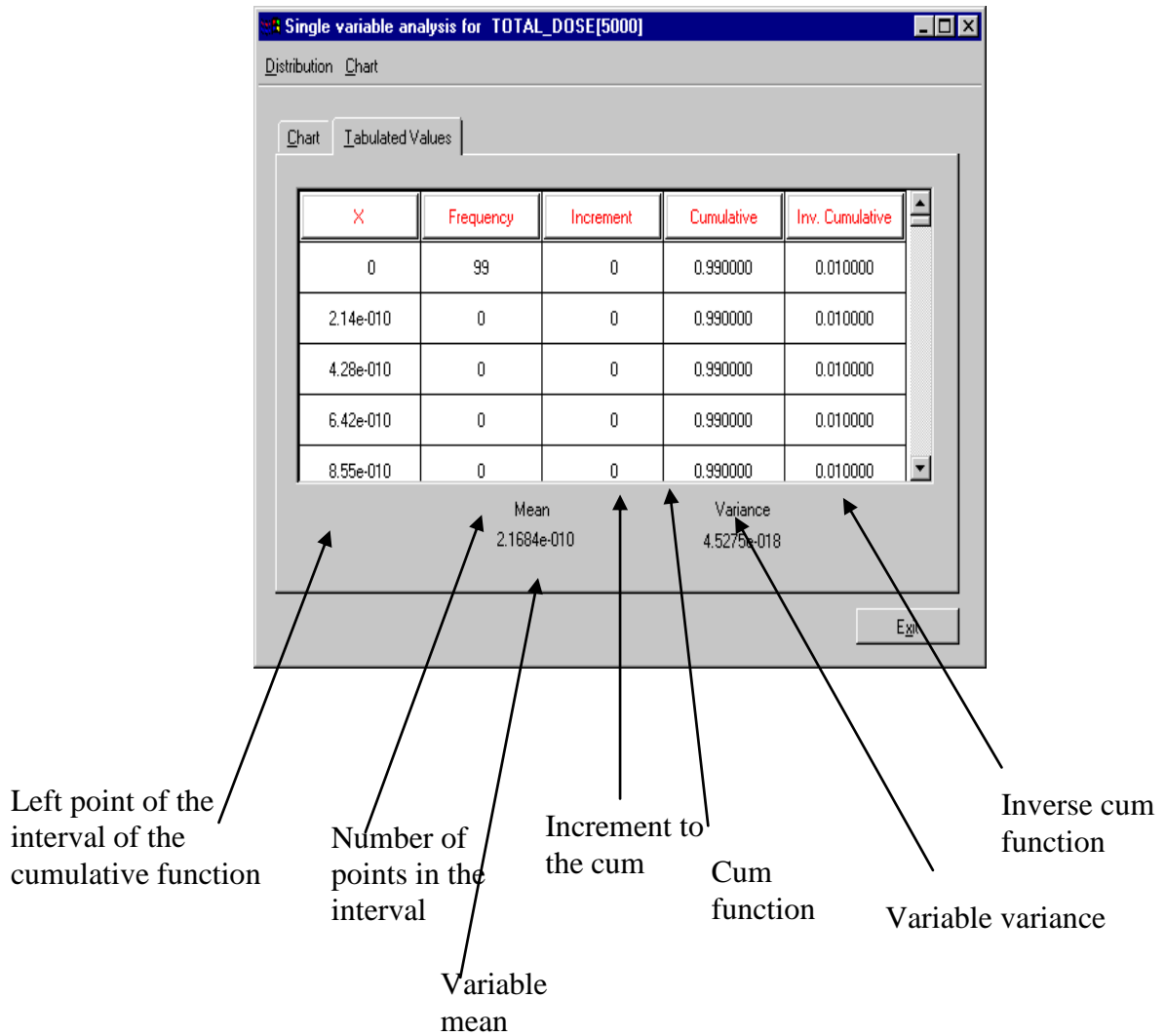


Figure 65 Tabulated Value Panel

MODEL RESULTS ANALYSIS (Statistical Processor)	4.2. Sensitivity Analysis (SRC, SRRC, PCC, PRCC, SPEAR, PEAR, SMIRNOV indices)	4
--	--	---

4.2. SENSITIVITY ANALYSIS (SRC, SRRC, PCC, PRCC, SPEAR, PEAR, SMIRNOV INDICES)

The purpose of uncertainty analysis is to determine the uncertainty in estimates for dependent variables of interest, see references[2], [7], [13], [18] and [19]. The purpose of sensitivity analysis is to determine the relationships between the uncertainty in the independent variables used in an analysis and the uncertainty in the resultant dependent variables. Uncertainty analysis typically precedes sensitivity analysis since, if the uncertainty in a dependent variable is under an acceptable bound or within an acceptable range, then there is little reason to perform a sensitivity analysis. Further, when Monte Carlo analysis is being performed, the generation of summary measures such as means, variances and distribution functions used to represent uncertainty requires little effort once the necessary model evaluations have been performed.

When a Monte Carlo study is being performed, propagation of the sample through the model creates a mapping from analysis inputs to analysis results of the form

$$[y_i, x_{i1}, x_{i2}, \dots, x_{in}] \quad i = 1, \dots, m$$

where n is the number of independent variables and m is the sample size. Once this mapping is generated and stored, it can be explored in many ways to determine the sensitivity of model predictions to individual input variables. This section considers sensitivity analysis techniques based on regression analysis: PEAR and PCC, SRC, and their rank transformation (SPEA, PRCC and SRRC).

Scatterplots, PEAR and SPEA

The generation of scatterplots is undoubtedly the simplest sensitivity analysis technique. This approach consists of generating plots of the points (x_{ij}, y_j) , $i = 1, \dots, m$, for each independent variable x_i .

Scatterplots may sometimes completely reveal the relationship between model input and model predictions; this is often the case when only one or two inputs dominate the outcome of the analysis. Further, they often reveal non-linear relationships, thresholds and variable interactions, so facilitating the understanding of the model behaviour. They can be considered as a global measure of importance, and are model independent, as the plots can reveal even strongly non-linear or non-monotonic features.

MODEL RESULTS ANALYSIS (Statistical Processor)	4.2. Sensitivity Analysis (SRC, SRR, PCC, PRCC, SPEAR, PEAR, SMIRNOV indices)	4
---	--	----------

When there is no relationship between the independent and the dependent variable, the individual points will be randomly spread over the plot. In contrast, the existence of a well-defined relationship between the

MODEL RESULTS ANALYSIS (Statistical Post Processor)	4.2. Sensitivity Analysis (SRC, SRRC, PCC, PRCC, SPEAR, PEAR, SMIRNOV indices)	4
---	--	---

independent and the dependent variable is often revealed by the distribution of the individual points.

One disadvantages of the method is that it needs generating and inspecting a large amount of plots: at least one per input factor, possibly multiplied by the number of time points if the output is time dependent. Further, scatterplots offer a qualitative measure of sensitivity: the relative importance of variables cannot be estimated but not quantified.

Another simple measure of sensitivity is given by the *Pearson product moment correlation coefficient* (PEAR) which is the usual linear correlation coefficient computed on the x_{ij}, y_i ($i = 1, \dots, m$).

For non-linear models the *Spearman coefficient* (SPEA) is preferred as a measure of correlation, which is essentially the same as PEAR, but using the ranks of both Y and X_j instead of the raw values i.e.,

$$\text{SPEA}(Y, X_j) = \text{PEAR}(R(Y), R(X_j)),$$

where R (.) indicates the transformation which substitutes the variable value with its rank.

The basic assumptions underlying the Spearman coefficient are:

- (a) Both the x_{ij} and y_i are random samples from their respective populations;
- (b) the measurement scale of both variables is at least ordinal.

The numerical value of SPEA, commonly known as the Spearman ‘rho’, can also be used for hypothesis testing, to quantify the confidence in the correlation itself. The following hypothesis is first formulated:

H_0 : “no correlation exists between Y and X_j ”;

then the value of SPEA (Y, X_j) is computed from a given number N of simulations and its value is compared with the quantiles of the Spearman test distribution. A level of significance α is selected which indicates the probability of erroneously rejecting the hypothesis, i.e. in this example, the probability that the test indicates correlation when Y and X_j are actually uncorrelated. Then, the hypothesis H_0 is rejected if SPEA falls out of the interval of extremes $[W(\alpha/2), W(1-\alpha/2)]$, where W’s are the quantiles of the test distribution. The values of W can be read from the table of the distribution according to the values of α and N.

Regression analysis

More quantitative measures of sensitivity are based on regression analysis. A multivariate sample of the input \mathbf{x} is generated by some sampling strategy (dimension $m \times k$), and the corresponding sequence of m output values is computed using the model under analysis. If a linear regression model is being sought, it takes the form

MODEL RESULTS ANALYSIS (Statistical Processor)	4.2. Sensitivity Analysis (SRC, SRRC, PCC, PRCC, SPEAR, PEAR, SMIRNOV indices)	4
--	--	---

$$y_i = b_0 + \sum_j b_j x_{ij} + \varepsilon_i$$

where $y_i = 1, \dots$, are the output values of the model, $b_j, j = 1, \dots, k$ (k being the number of input variables) are coefficients that must be determined and ε_i is the error (residual) due to the approximation. One common way of determining the coefficients b_j is using the least square method. In this least square approach, the b_j 's are determined so that the function

$$F(b) = \sum_i \varepsilon_i^2$$

is a minimum. Once the b_j 's are computed, they can be used to indicate the importance of individual input variables x_j with respect to the uncertainty in the output y . In fact, assuming that \mathbf{b} has been computed, the regression model can be rewritten as

$$(y - \bar{y}) / \hat{s} = \sum_j (b_j \hat{s}_j / \hat{s})(x_j - \bar{x}_j) / \hat{s}_j$$

where

$$\bar{y} = \sum_i y_i / m, \quad \bar{x}_j = \sum_i x_{ij} / m$$

$$\hat{s} = \left[\sum_i (y_i - \bar{y})^2 / (m-1) \right]^{1/2}$$

$$\hat{s}_j = \left[\sum_i (x_{ij} - \bar{x}_j)^2 / (m-1) \right]^{1/2}$$

The coefficients $b_j \hat{s}_j / \hat{s}$ are called **Standardised Regression Coefficients (SRC)**. These can be used for sensitivity analysis (when the x_j are independent) as they can quantify the effect of varying each input variable away from its mean by a fixed fraction of its variance while maintaining all other variables at their expected values.

When using the SRC's it is also important to consider the model coefficient of determination,

MODEL RESULTS ANALYSIS (Statistical Processor)	4.2. Sensitivity Analysis (SRC, SRRC, PCC, PRCC, SPEAR, PEAR, SMIRNOV indices)	4
--	--	---

$$R_y^2 = \frac{\sum_{i=1}^m (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^m (y_i - \bar{y})^2}$$

where \hat{y}_i denotes the estimate of y_i obtained from the regression model.

R_y^2 provides a measure of how well the linear regression model based on SRC's can reproduce the actual output y . R_y^2 represents the fraction of the variance of the output explained by the regression. The closer R_y^2 is to unit, the better is the model performance. The validity

of the SRC's as a measure of sensitivity is conditional on the degree to which the regression model fits the data, i.e. to R_y^2 .

Correlation measures

Another interesting measure of variable importance is given by **Partial Correlation Coefficients (PCC)**. These coefficients are based on the concepts of correlation and partial correlation. For a sequence of observations (x_{ij}, y_i) , the correlation $r_{x_j y}$ between the input variable X_j and the output Y is defined by

$$r_{x_j y} = \frac{\sum_{i=1}^m (x_{ij} - \bar{x}_j)(y_i - \bar{y})}{\left[\sum_{i=1}^m (x_{ij} - \bar{x}_j)^2 \right]^{1/2} \left[\sum_{i=1}^m (y_i - \bar{y})^2 \right]^{1/2}}$$

where $\bar{y} = \sum_i y_i / m$ and $\bar{x}_j = \sum_i x_{ij} / m$.

The correlation coefficient $r_{x_j y}$ provides a measure of the linear relationship between X_j and Y . The partial correlation coefficient between the output variable Y and the input variable X_j is obtained from the use of a sequence of regression models. First the following two models are constructed

MODEL RESULTS ANALYSIS (Statistical Processor)	4.2. Sensitivity Analysis (SRC, SRRC, PCC, PRCC, SPEAR, PEAR, SMIRNOV indices)	4
--	--	---

$$\hat{Y} = b_0 + \sum_{h \neq j} b_h x_h \quad \text{and} \quad \hat{X}_j = c_0 + \sum_{h \neq j} c_h x_h$$

Then, the results of these two regressions are used to define the new variables $Y - \hat{Y}$ and $X_j - \hat{X}_j$. The partial correlation coefficient between Y and X_j is defined as the correlation coefficient between $Y - \hat{Y}$ and $X_j - \hat{X}_j$. Thus, the partial correlation coefficients provide a measure of the strength of the linear relationship between two variables after a correction has been made for the linear effects of other variables in the analysis. In other words, PCC gives the strength of the correlation between Y and a given input X_j cleaned of any effect due to any correlation between X_j and any of the $X_i, i \neq j$.

In particular PCC's provide a measure of variable importance that tends to exclude the effects of other variables.

However, in the particular case in which the input variables are uncorrelated, the order of variable importance based either on SRC's or PCC's (in their absolute values) is exactly the same.

Rank transformation

Regression analysis often performs poorly when the relationships between the input variables are non-linear. In this case, the value of the

R_y^2 coefficient computed on the raw values may be low. The problem

associated with poor linear fits to nonlinear data can often be avoided with the use of the rank transformations.

The rank transform is a simple procedure which involves replacing the data with their corresponding ranks, i.e. assign rank 1 to the smallest observation and continue to rank N for the largest observation. The usual least square regression analysis is then performed entirely on these ranks $R(x_k)$. The final regression equation expresses $R(y_i)$ in terms of $R(x_k)$.

The new value for R_y^2 (on ranks) is then computed: if this new value is

higher, then the new coefficients **SRRC, Standardised Rank Regression Coefficients**, can be used for sensitivity analysis instead of

SRC's. The ranked variables are more often used because the R_y^2

MODEL RESULTS ANALYSIS (Statistical Processor)	Post	4.2. Sensitivity Analysis (SRC, SRRC, PCC, PRCC, SPEAR, PEAR, SMIRNOV indices)	4
--	------	--	---

associated with the SRC's are generally lower than that associated with the SRRC's, especially for non-linear models. The difference between the R_y^2 , computed on the raw values and on the ranks, is a useful indicator of the non-linearity of the model.

The performance of the SRRC is shown to be extremely satisfactory when the model output varies linearly or at least monotonically with each independent variable. However, in the presence of strong monotonicity, the accuracy of results may become dubious or completely misleading. Another limitation in the use of ranks is that the transformation alters the model being studied, so that the resulting sensitivity measures (e.g. SRRC) give us information on a different model. The new model is not only more linear but also more additive than the original one, i.e. more variation can be explained in terms of summing elementary effect than was the case with the original model. When this becomes, the SRRC become somehow a qualitative measure, as the relative importance of input cannot be assessed.

The PCC can be computed on the ranks (*Partial Rank Correlation Coefficients*). The performance of the PRCC shows the same features of performance as the SRCC: good for monotonic models, and not fully satisfactory in the presence of non-monotonicity. PCC's provides related but not identical measures of variable importance.

Kolmogorov-Smirnov Test

Model simulations are classified as either behavioural (B) or non-behavioural (\bar{B}). A set of binary elements are defined distinguishing between two sub-sets of each input factor $X_i: (X_i|B)$ of m elements and $(X_i|\bar{B})$ of n elements ($n+m=N$, the total number of Monte Carlo runs performed). The Smirnov two-sample test is performed for each factor X_i independently. Under the null hypothesis that the two distributions $f_m(X_i|B)$ and $f_n(X_i|\bar{B})$ are identical:

$$H_0 : f_m(X_i|B) = f_n(X_i|\bar{B})$$

$$H_1 : f_m(X_i|B) \neq f_n(X_i|\bar{B})$$

The test statistic is defined by:

$$d_{m,n}(X_i) = \sup_y \|F_m(X_i|B) - F_n(X_i|\bar{B})\|$$

where F are marginal cumulative probability functions and f are probability density functions. At what significance level α does the computed value of $d_{m,n}$ determines the rejection of H_0 ? A low level

MODEL RESULTS ANALYSIS (Statistical Processor)	4.2. Sensitivity Analysis (SRC, SRRC, PCC, PRCC, SPEAR, PEAR, SMIRNOV indices)	4
---	---	----------

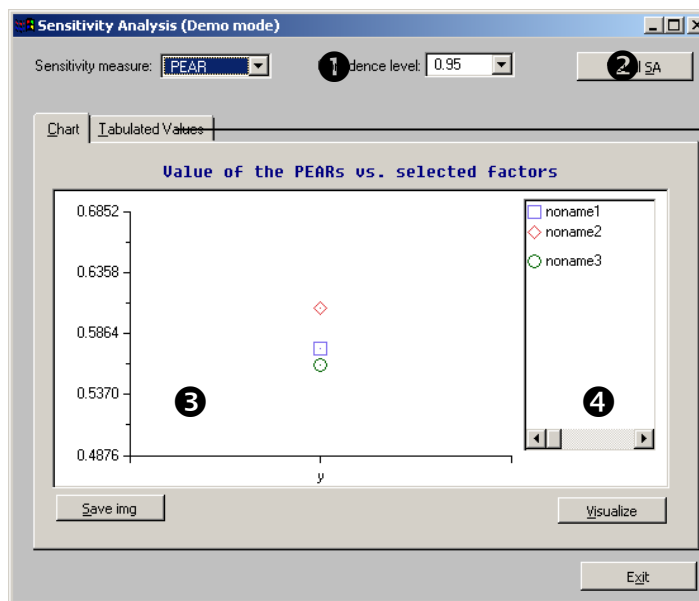
of α implies a significant difference between $f_m(X_i|B)$ and $f_n(X_i|\bar{B})$, suggesting that X_i is a key factor in producing the splitting between B and \bar{B} .

To perform the Smirnov test, we choose the significance level α , which is the probability of rejecting H_0 when it is true (i.e. to recognise a factor as important when it is not). From α we derive the critical level D_α at which the computed value $d_{m,n}$ determines the rejection of H_0 . If $d_{m,n} > D_\alpha$ then H_0 is rejected at significance level α and the factor X_i is considered as important. Simlab uses $\alpha = 5\%$.

MODEL RESULTS ANALYSIS (Statistical Post Processor)	4.2. Sensitivity Analysis (SRC, SRRC, PCC, PRCC, SPEAR, PEAR, SMIRNOV indices)	4
---	--	---

After the selection of the models output variables to analyse (at least one variable must be selected), click the SA button. The Sensitivity Analysis differs on the base of the sample method chosen. For random, Quasi random Lptau, Latin Hypercube, fixed method the following panel is visualised:

SA Main panel:



Tab selection

Tab selection

Figure 66 SA main Panel

Chart and Tabulated Values Tabs allow user to select a different visualisation of the SA indexes.

Chart tab visualises graphically the value of the selected indexes of sensitivity (see Figure 66 SA main Panel SimLab main panel).

Tabulated Values tab visualises the values of the selected indexes as elements in a matrix.

- ① **Sensitivity measure and confidence bound combos.** The first combo contains the list of all the sensitivity indices calculated. Switching from one to the other the SA window loads the related values. The second changes the confidence level when executing test on indices.
- ② **Visual SA button.** Clicking this button the user opens all the panels for the visual SA.
- ③ **Save img button** saves the panel in a postscript file.

MODEL RESULTS ANALYSIS (Statistical Post Processor)	4.2. Sensitivity Analysis (SRC, SRRC, PCC, PRCC, SPEAR, PEAR, SMIRNOV indices)	4
---	--	---

④ **Visualise button** allows the user to choose which variables and input factors displaying on the screen. This button visualised the ‘Select components’ panel:

In the “chart” window the user has to select which items wishes to put in the graphical area. Clicking the button ④ the following panel is displayed:

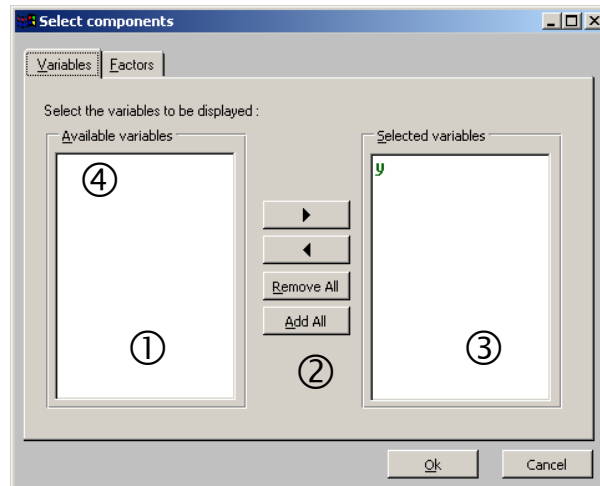
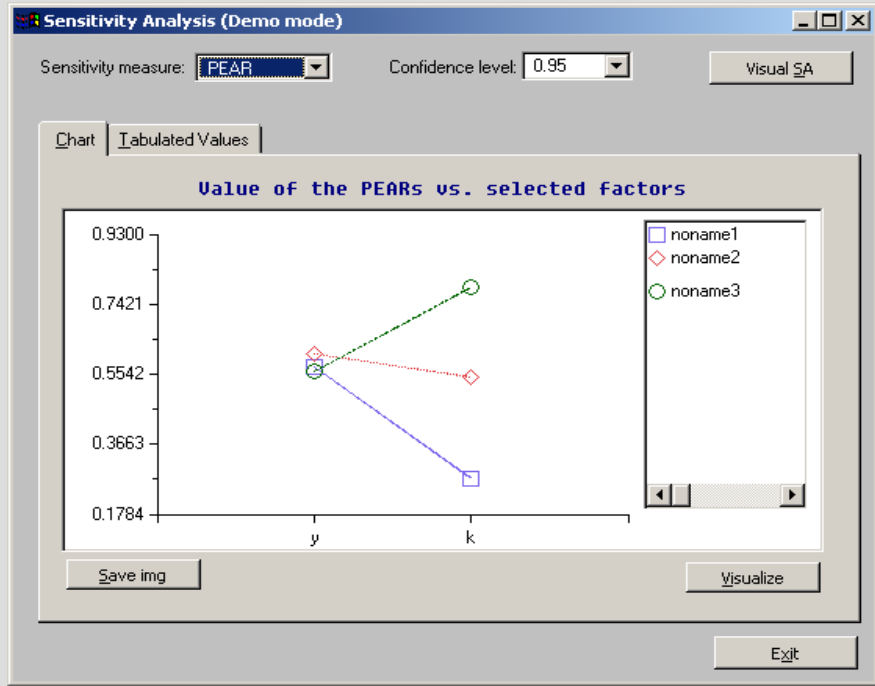


Figure 67 Select Variables and factors Panel

- ① **Available variables area.** This area contains the variables available to be displayed. At first this area contains all the variables of the ‘List of variables selected’ area of the Post Processor Main panel with exception of the first variable which is the only visualised as default. For moving the variables from this area into ‘Selected variables’ area and vice versa, the user must select the variable to move and use the appropriate “arrow” buttons or double click it.
- ② **Move buttons.** Through this buttons the user has the possibility to move the selected variable from ‘Available variables’ area to ‘Selected variables’ area and vice versa, or to remove all variables from the ‘Selected variables’ area.
- ③ **Selected variable area.** This area contains the selected variables. At first this area contains as default only the first variable of ‘List of variables selected’ area of the Post Processor Main
- ④ **Factors tab selection.** This tab visualises a similar panel in which the users can select the input factors to display (max 10 factors).

Below there is an example of the SA panel with two output variables and three input factors selected. The output variable is on the x-axis while the factors on the vertical axis.



Tabs selection:
Tabulated values

Tabulated Values visualises the values of the selected index of sensitivity as elements in a matrix.

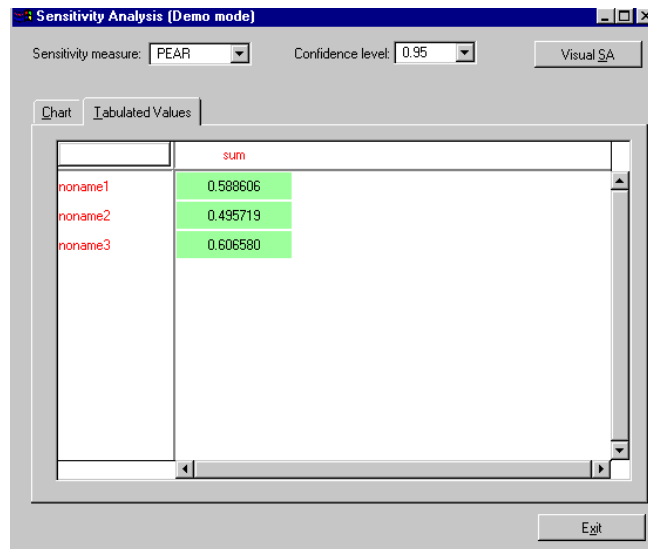
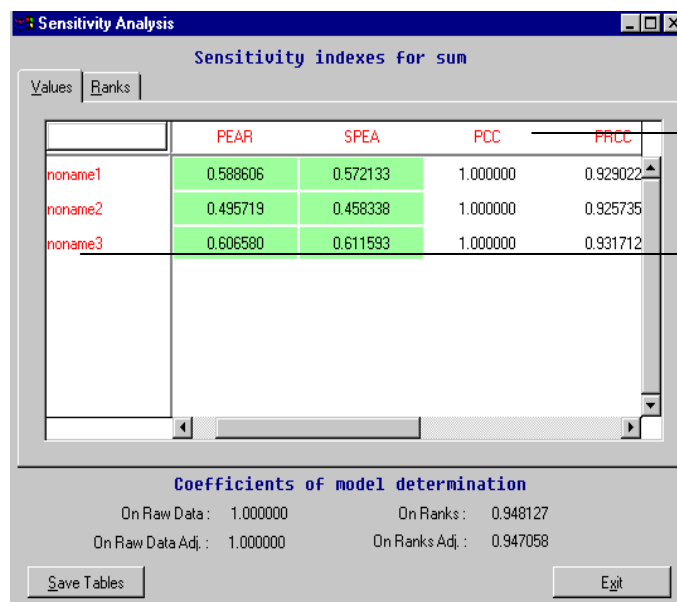


Figure 68 Sensitivity Analysis tabulated value

Double clicking a variable name in the first line of the previous matrix a panel regarding the item selected will be visualised. Here all the sensitivity indices for the specific variable and information about the coefficients of model determination are displayed.



Index of sensitivity

Input factors

Figure 69 Panel generated by clicking the header of output variable

For some sensitivity indices, the Tchebycheff significance test is implemented (see pages 79 and 80 of this manual where $H_0: \mu = 0$). A green background in the sensitivity indices matrix means that the sensitivity estimate is significant, while a red background means that the sensitivity estimate is non significant.

A third way to investigate the values of the SA indexes for a single variable is through the forecast panels. Click the column of the index to visualise and a graphical representation in descending order of the magnitude of the values will appear.

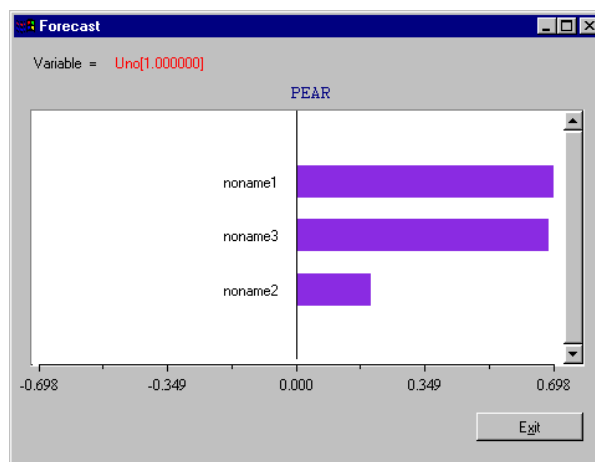


Figure 70 Forecast panel

MODEL RESULTS ANALYSIS (Statistical Post Processor)	4.2. Sensitivity Analysis (SRC, SRRC, PCC, PRCC, SPEAR, PEAR, SMIRNOV indices)	4
---	--	---

Visual SA:

To inspect graphically how the input factors influences the output variable the “visual SA” could be invoked. The visual SA display through scatterplots and cobwebs the factors and the variables selected.

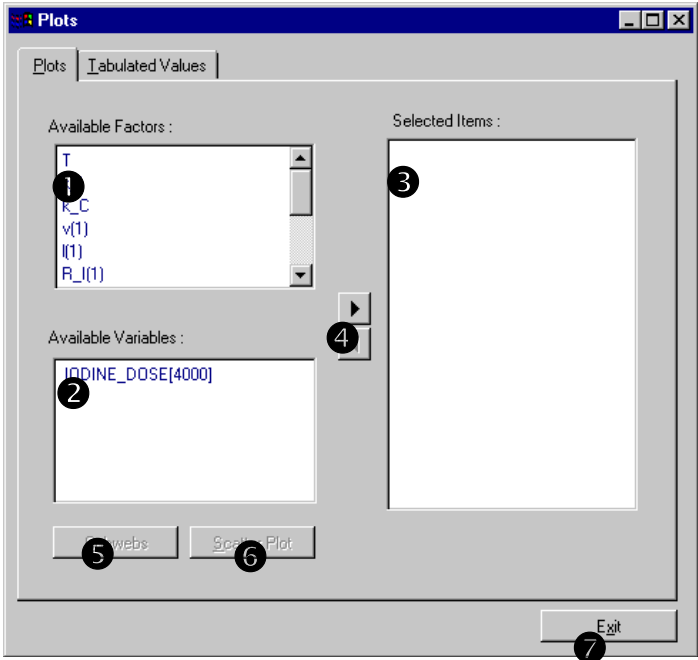


Figure 71 Plots panel

- ① **Available Factors area.** This area visualises the current list of factors.
- ② **Available Variables area.** This area visualises the current list of variable.
- ③ **Selected Items area.** This area visualises the factors and the variables chosen by user. The factor names start with an ‘F’, while the variable names with a ‘V’.
- ④ **Move buttons.** Through this buttons the user can move a selected factor or variable from the ‘Available area’ into ‘Selected Items area’ or vice versa. The same functionality is achieved double clicking an item.

⑤ **Cobwebs button.** This button, available only if the selected items are more than 2, shows the factors and the output variables with a cobweb graph.

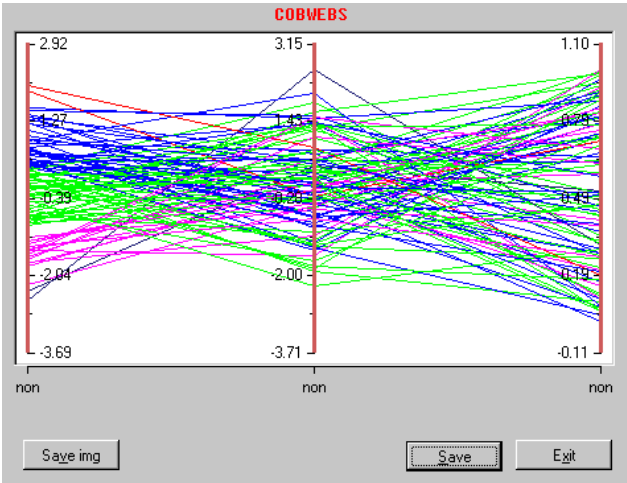


Figure 72 Cobwebs graph

⑥ **Scatter Plot button** shows two item with the scatter diagram. This button is enabled when two items are in the 'Selected Items area'.

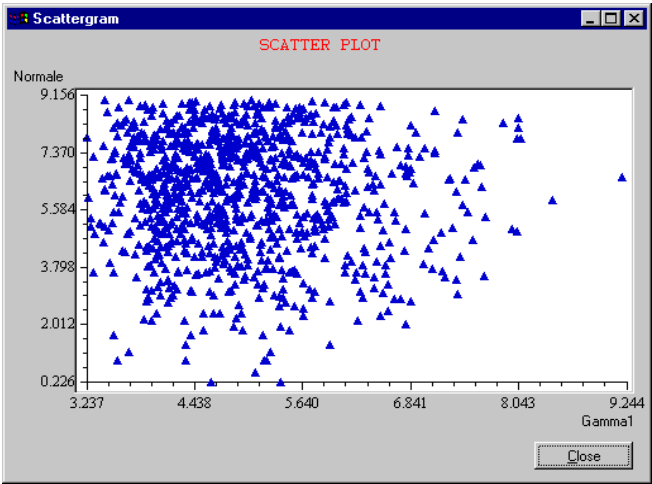
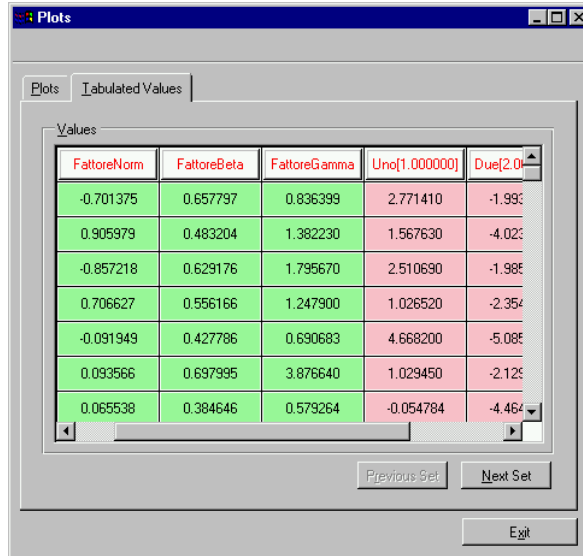


Figure 73 Scatter plot panel

The user can choose to see tabulated values through the following panel:

Tabulated value tab



In this panel all the external model input factors (green columns) and all model's outcomes (pink columns) are shown. The panel visualises a set of 100 samples each time. Through the 'Previous Set button' and the 'Next Set button' the user can navigate through all the values.

- ⑦ **Exit button.** Close the window

4.3. SENSITIVITY ANALYSIS MORRIS METHOD

Morris estimates the main effect of a factor by computing a number r of local measures, at different points x_1, x_2, \dots, x_r in the input space, and taking their average (this reduces the dependence on the specific point that a local experiment has). These r values are selected such that each factor is varied over its interval of experimentation. Morris wishes to determine which factors have negligible effects, linear and additive effects, non-linear or interaction effects, see reference [13].

The Morris method doesn't allow the UA analysis.

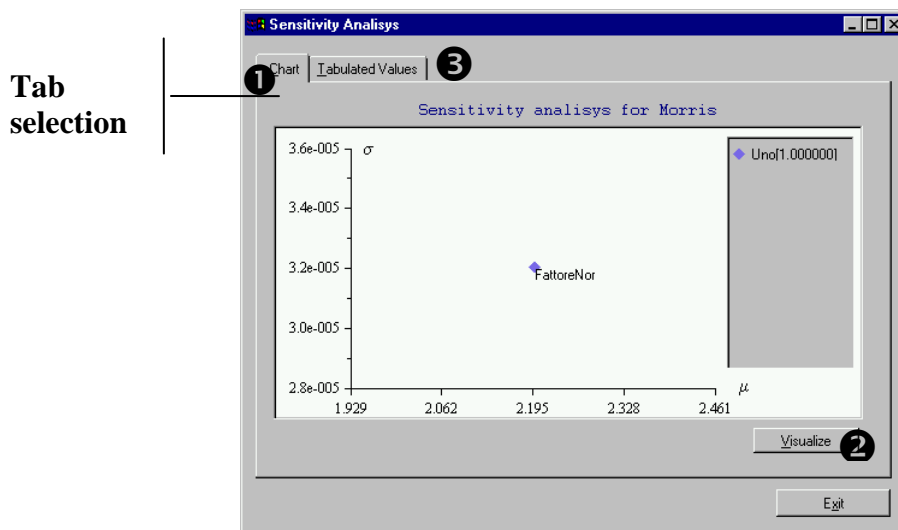


Figure 74 Sensitivity analysis for Morris

Chart and Tabulated Values Tabs allow user to select a different visualisation of the same indexes for the selected factors.

- 1 **Chart.** Visualises on X-axis the μ values and on the Y-axis the σ values of the Morris indicators.



NOTE: The software visualises as default all the values related to the variables and to the input factors.

- 2 **Visualise** button allows the selection of model's output variables and input factors for the visualisation; more items simultaneously can be displayed.

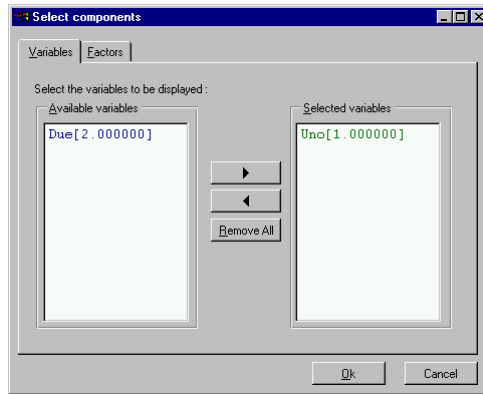


Figure 75 select Variables Panel

- 3 **Tabulated Values.** Visualises the μ values and σ values generated by sensitivity analysis of all output variable and of all input factors

	Uno[1.000000]	Uno[1.000000]	Due[2.000000]	Due[2.000000]
SA - MORRIS	μ	σ	μ	σ
FattoreNorm	2.196993	0.000032	2.196987	0.000029
FattoreBeta	0.544592	0.000005	0.544591	0.000005
FattoreGamma	2.850306	0.622412	2.850305	0.622407

Figure 76 Tabulated values

4.4 SENSITIVITY ANALYSIS: FAST

FAST is a method used to solve non-linear, non-monotonic problems that are also called non-linear sensitivity analysis. The FAST approach is based on performing numerical calculations to obtain the expected value and variance of a model prediction. The basis of this calculation is a transformation that converts a multidimensional integral over all the uncertain model inputs to a one-dimensional integral, see references [14] and [15].

There are also some variations of the basic FAST.

The analysis is divided into four steps: construction of ranges and distributions of the input factors and development of the expected value and variance of Y, transformation of the multidimensional integral, estimation of the expected value and variance of Y and sensitivity analysis.

Fast panel:

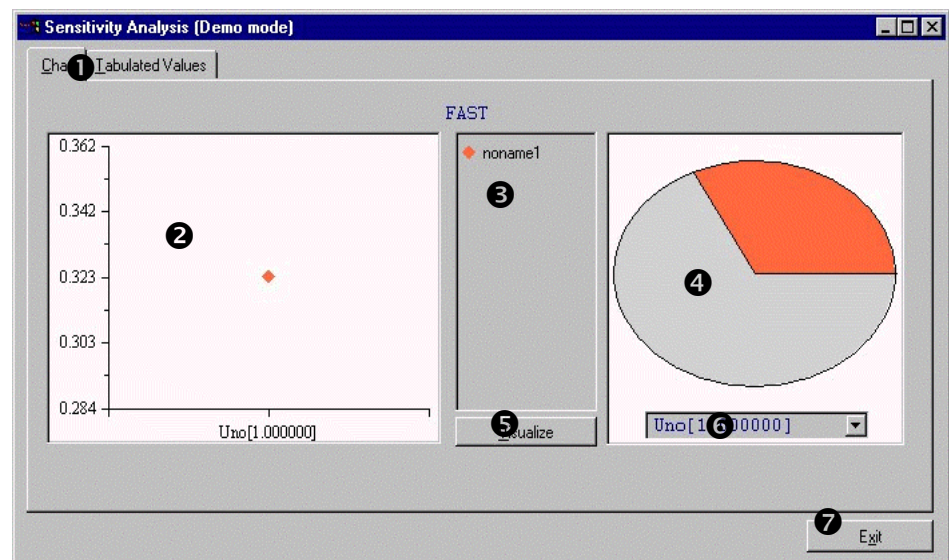


Figure 77 Fast SA panel first orders effects

- ① **Chart and Tabulated Values Tabs** allow user to select a different visualisation of the same indexes for the selected factors (for the tabulated Values panel see *Figure 78*).
- ② **Graphic area.** Area where the indexes are plotted in a system of co-ordinated axis.
- ③ **Legend area.** This area displays the list of factors currently visualised in the window
- ④ **Pie chart.** Pie chart related to the indexes loaded.

- 5 **Visualise button.** Button that allows the selection of variables and factors to be displayed.
- 6 **Pie chart selector.** Combo box in which every variable analysed is listed. This selection affects only the pie chart diagram.
- 7 **Exit button** closes the window and returns to Post Processor main window.

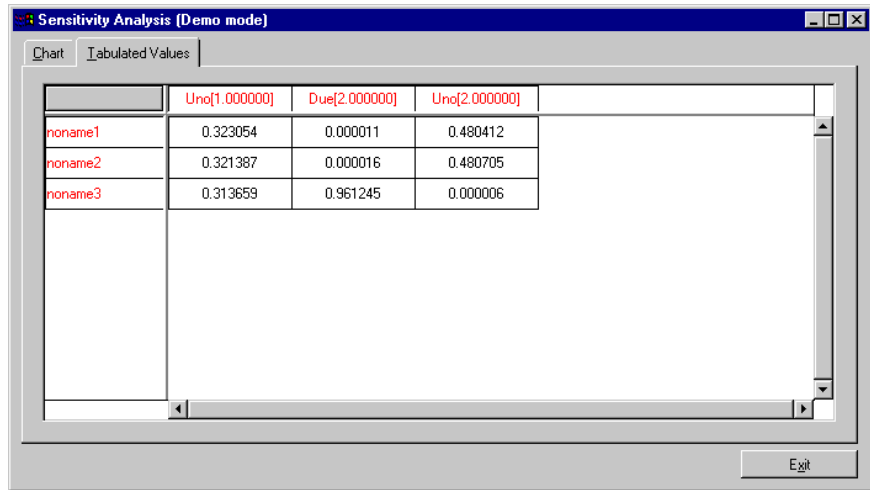


Figure 78 Fast panel tabulated values

In the first and total effects analysis the Fast panels visualises at the same time the two kind of indexes, see figure below:

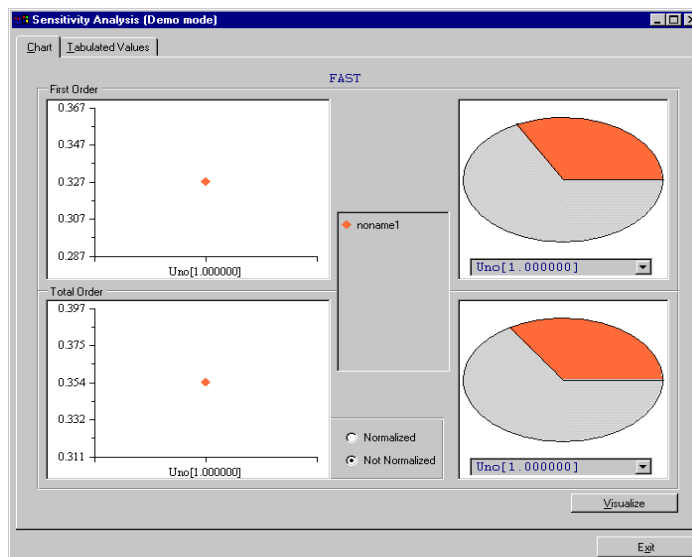


Figure 79 Fast first and total order effects panel

4.5 SENSITIVITY ANALYSIS: IMPORTANCE MEASURES

Importance indices are model free sensitivity measures, which supply the main effect of individual factors. These measures are useful in environments where the inputs to be examined have a correlated structure.

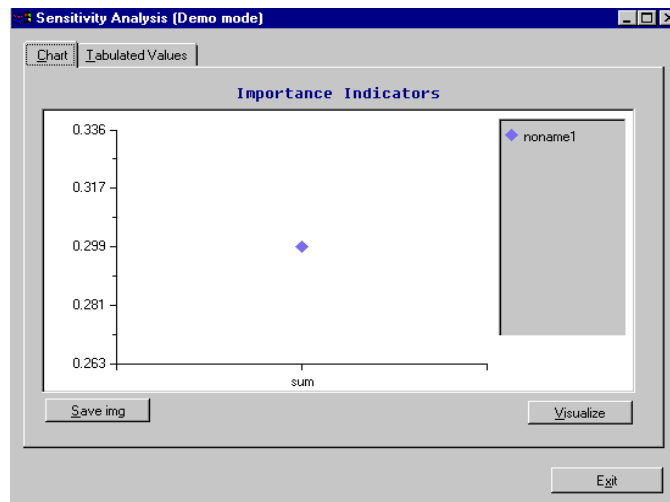


Figure 80: Importance measures panel

Chart and Tabulated Values Tabs allow user to select a different visualisation of the importance indicators for the selected factors.

Chart tab Visualises on a graph the values of the importance measures indicators.



NOTE: The software visualises as default all the values related to the variables and to the input factors.

Visualise button allows the selection of model's output variables and input factors for the visualisation; more items simultaneously can be displayed.

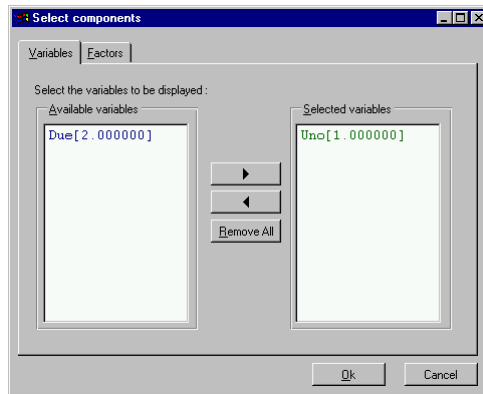


Figure 81 select components Panel

Tabulated Values. Visualises the tabulated values for the importance indicators for all output variables and for all input factors

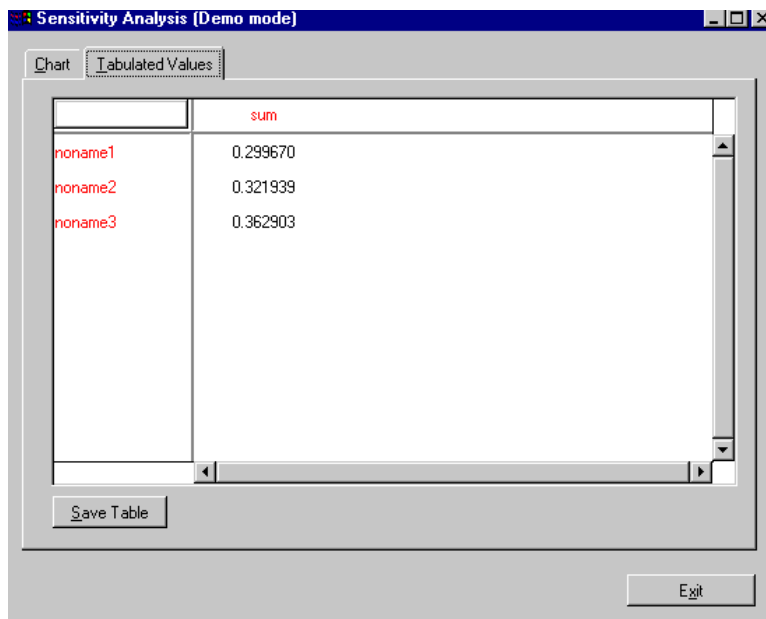


Figure 82: Importance measures tabulated values panel

4.6 SENSITIVITY ANALYSIS: SOBOL

Fast panel:

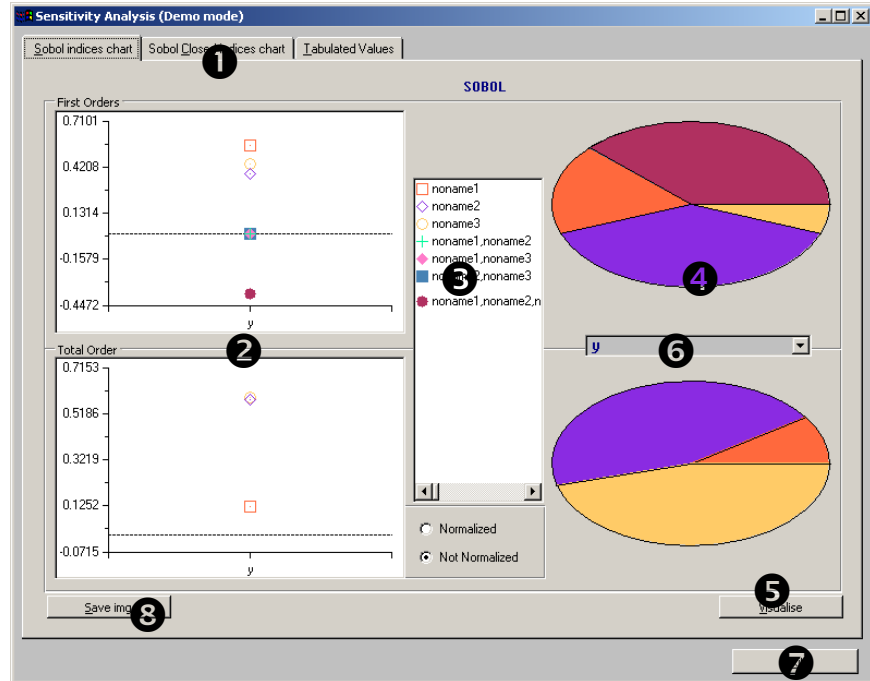


Figure 83 Sobol SA panel: Custom order calculation

- 1 **Sobol indices Chart, Sobol Closed indices and Tabulated Values Tabs** allow user to select different representations of the same indices for the selected factors (for the tabulated Values panel see Figure 85 Sobol SA panel: tabulated values tab).
- 2 **Graphic area.** Area where the indices are plotted in a system of co-ordinated axis.
- 3 **Legend area.** This area displays the list of factors currently visualised in the window
- 4 **Pie chart.** Pie chart related to the indexes loaded.
- 5 **Visualise** button allows the selection of model's output variables and input factors for the visualisation; more items simultaneously can be displayed.

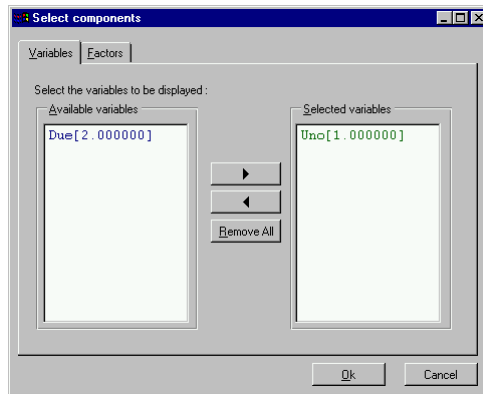


Figure 84 Select variables Panel

- ⑥ **Pie chart selector.** Combo box in which every variable analysed is listed. This selection affects only the pie chart diagram.
- ⑦ **Exit button** closes the window and returns to Post Processor main window.
- ⑧ **Save image button** saves the panel in a postscript file.

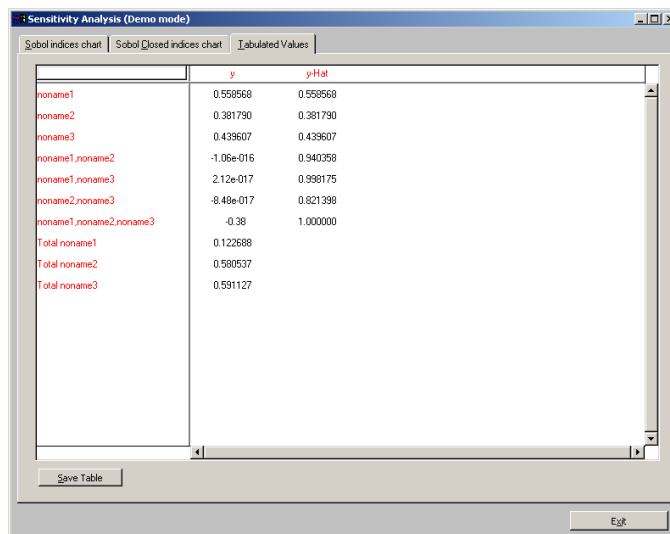


Figure 85 Sobol SA panel: tabulated values tab

MODEL'S RESULTS ANALYSIS (Statistical Processor)	Post	4.6 Sensitivity analysis: Sobol	4
---	-------------	--	----------

figure

will

be

APPENDIX A

Each distribution panel can be divided in four parts (the only exceptions are constants and relations panels):

Information area

Parameters area

Distribution area

Buttons area

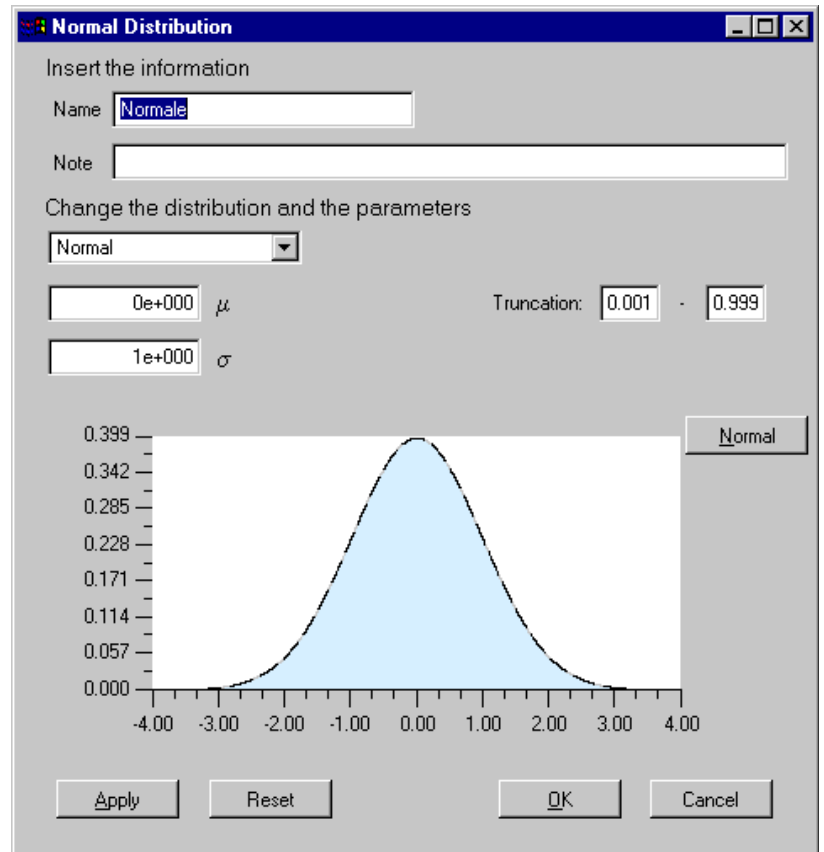


Figure 86 Default panel: Normal distribution

Information area: This area is identical in all distribution panels; in this area the user must specify the name of the current input factor. He may also insert some notes about it.

Parameters area: This area changes in each distribution panel; all the distribution parameters inputs are located here. The only common field is the distribution type selector:

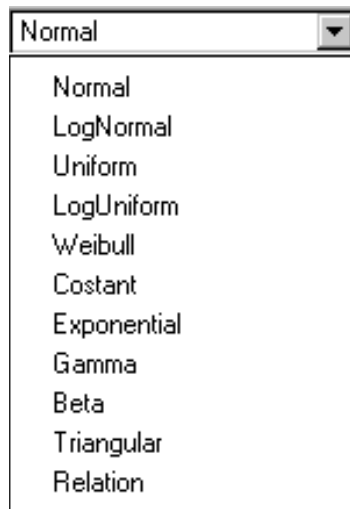


Figure 87 Distribution type selector

Distribution area: Shows a plot of the selected distribution with the current parameters; clicking the button in this area a Quick Help will be visualised; this help shows the analytical expression of the distribution density.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right]$$

Figure 88 Quick Help example

Buttons area: This area is common in every distribution panel.

Apply button: Using this button, after changing the parameters, the new parameters will be validated and the distribution plot will be updated.

Reset button: You can use this button to reset the distribution parameters to their default values.

Normal distribution

The density function of the Normal distribution has the following expression:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right]$$

for x in $(-\infty, \infty)$.

**Distribution
parameters**

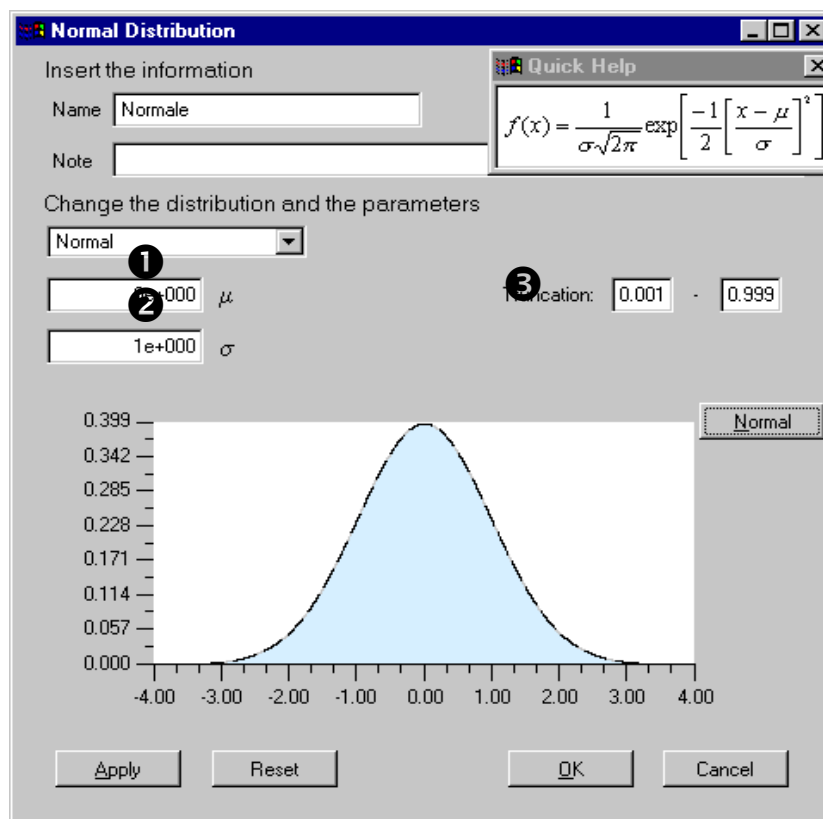


Figure 89 Normal distribution panel

- ❶ Mean value (μ), with μ in $(-\infty, \infty)$; the default value is $\mu = 0e+000$.
- ❷ Standard deviation value (σ), with $\sigma > 0$; the default value is $\sigma = 1e+000$.
- ❸ Truncation value; the default value is in $[0.001, 0.999]$.

There is the possibility to see the density formula through Quick Help by pressing the Normal button and close it by pushing again this button. The Quick Help functionality is the same for all distribution panels.

LogNormal distribution

The density function of the LogNormal distribution has the following expression:

$$f(x) = \frac{1}{x \cdot \sigma \cdot \sqrt{2\pi}} \exp\left[-\frac{1}{2} \left[\frac{\ln(x) - \mu}{\sigma}\right]^2\right] \cdot \mathbf{1}_{(0, \infty)}(x)$$

for x in $(0, \infty)$.

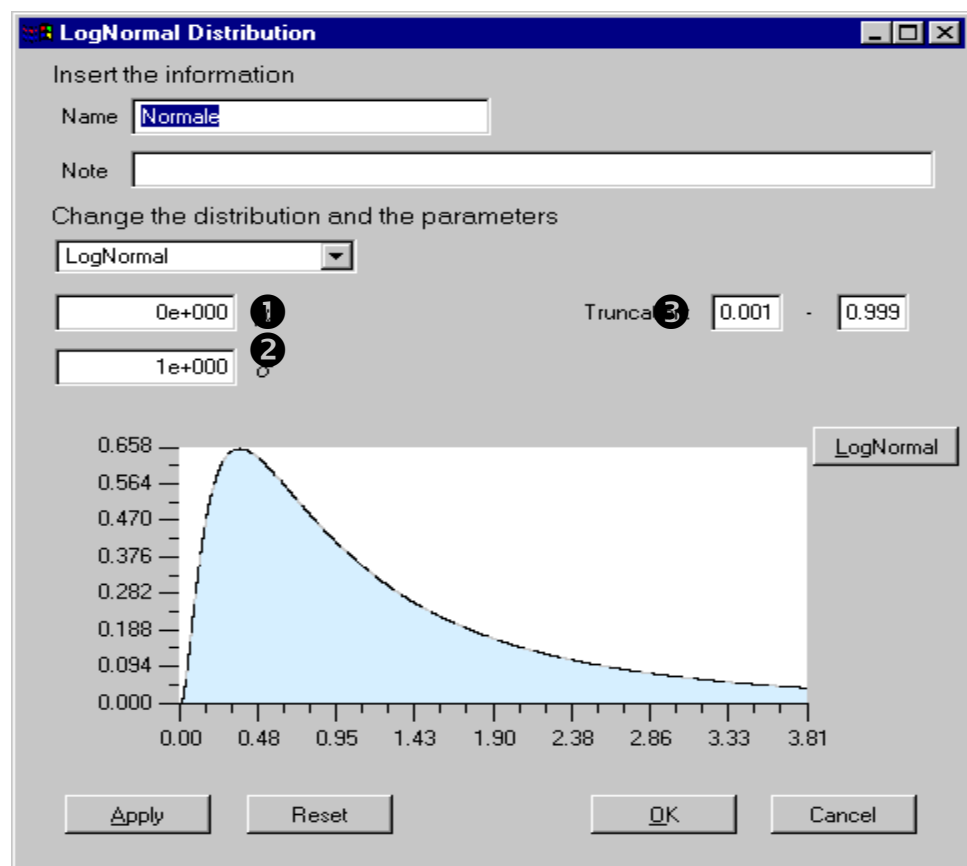


Figure 90 Log-Normal distribution panel

Distribution parameters

- ① μ parameter; the default value for μ is 0.
- ② σ parameter; the default value for σ is 1.
- ③ Truncation value; the default value is in $[0.001, 0.999]$.

Mean value: $\exp\left(\mu + \frac{1}{2}\sigma^2\right)$

Standard deviation: $\exp(2\mu + 2\sigma^2) - \exp(2\mu - \sigma^2)$

Uniform distribution

The density function of the Uniform distribution has the following expression:

$$f(x) = \sum_{i=1}^n \frac{Weight_i}{RightVal_i - LeftVal_i}$$

with n number of intervals.

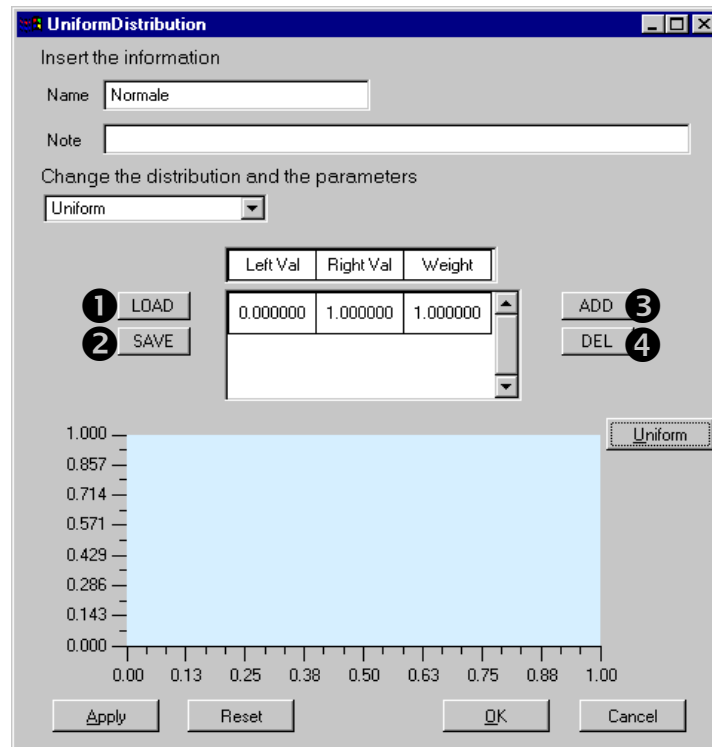


Figure 91 Uniform distribution panel

- Buttons description**
- ➊ **LOAD button.** The user can select this button to load a set of intervals for the current uniform distribution. This list of intervals will overwrite any previous list of intervals present in the matrix.
 - ➋ **SAVE button.** This button allows the user to save in a file the current intervals for the uniform distribution.
 - ➌ **ADD button.** This button is used to insert a new interval for the current uniform distribution. For each interval, the user has to insert the left bound, the right bound and the relative weight of the interval. The intervals can't overlap and the sum of the weights must be equal to 1.
 - ➍ **DEL button.** With this button the user may delete any interval present in the matrix. The user must select one interval before pressing the DEL button. After deleting an interval, the weight of the remaining elements must be corrected.

Parameters area

This area contains the parameters that characterise the uniform distribution: $RightVal_i$ and $LeftVal_i$ are the extreme values of i -th subinterval so that

$(RightVal_i - LeftVal_i) > 0$ for each i ;

$Weight_i$ is the weight of the i -th subinterval, the sum of all the weights must be equal to 1.

The user can modify each value double clicking the matrix item.

$$\text{Mean Value: } \sum_{i=1}^n \left(\frac{LeftVal_i + RightVal_i}{2} \right) * w_i$$

$$\text{Standard deviation: } \sqrt{\sum_{i=1}^n \left(\frac{(LeftVal_i^3 - RightVal_i^3)}{LeftVal_i - RightVal_i} * \frac{1}{3} * w_i \right)}$$

LogUniform distribution

The density function of the LogUniform distribution has the following expression:

$$f(x) = \sum_{i=1}^n \frac{Weight_i}{x \cdot (\ln(RightVal_i) - \ln(LeftVal_i))}$$

with n number of intervals.

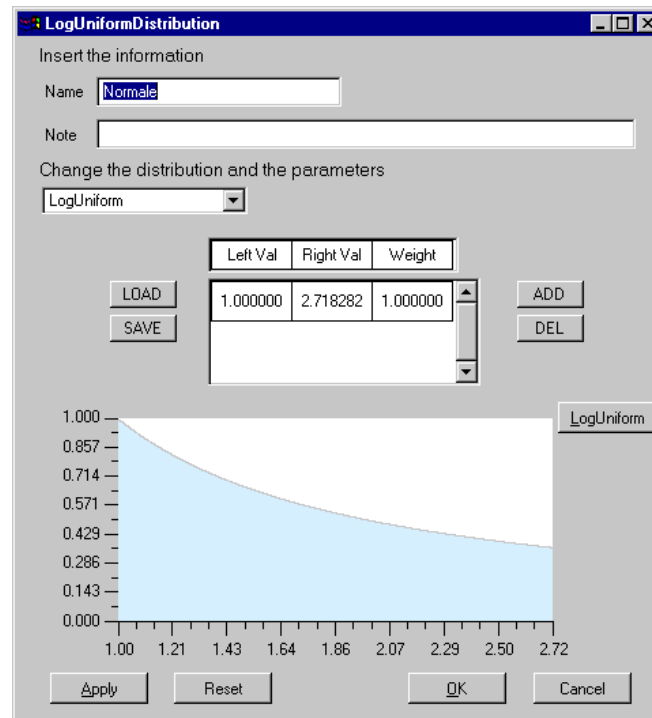


Figure 92 LogUniform distribution panel

$$\text{Mean Value } (\mu): \sum_{i=1}^n \frac{RightVal_i - LeftVal_i}{\log(RightVal_i) - \log(LeftVal_i)}$$

Standard deviation:

$$\sqrt{\sum_{i=1}^n \left(\frac{RightVal_i^2 - LeftVal_i^2}{2 \cdot (\log(RightVal_i) - \log(LeftVal_i))} \right) + \mu^2 - 2\mu \cdot \frac{RightVal_i - LeftVal_i}{\log(RightVal_i) - \log(LeftVal_i)}}$$

Weibull distribution

The density function of the Weibull distribution has the following expression:

$$f(x) = \frac{\beta}{n} * \left[\frac{x - b}{n} \right]^{\beta-1} \cdot \exp \left[- \left(\frac{x - b}{n} \right)^{\beta} \right]$$

Distribution parameters

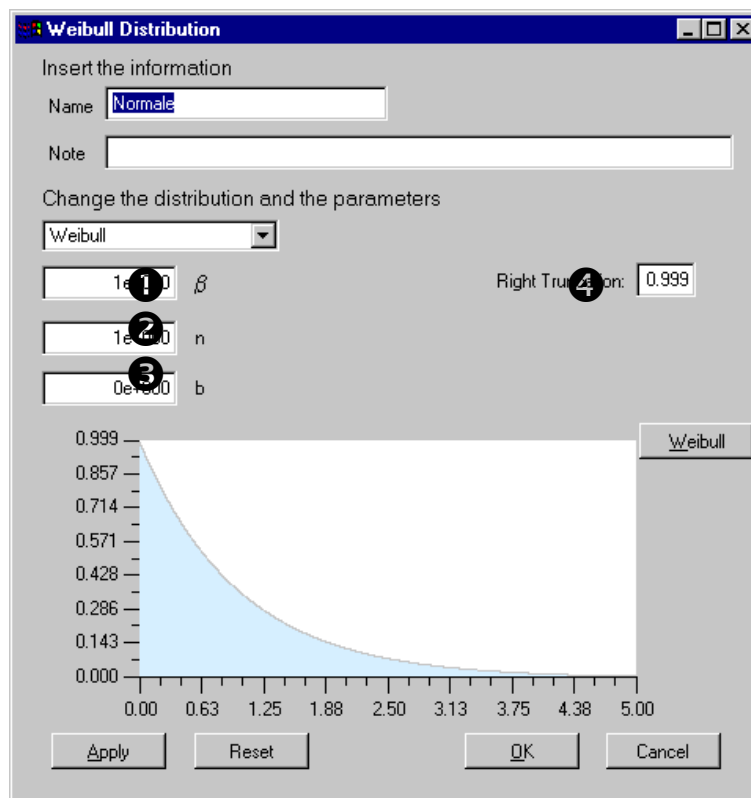


Figure 93 Weibull distribution panel

- ❶ β value with $\beta > 0$; the default value for β is 1.
- ❷ n value with $n > 0$; the default value for n is 1.
- ❸ b value with $b \geq 0$; the default value for b is 0.
- ❹ Right truncation value; the default value is 0.999

Mean value: $n \cdot \Gamma\left(1 + \frac{1}{\beta}\right) + b$

Standard deviation: $\sqrt{n^2 \cdot \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right) \right]}$

Constant factor

The user may insert constant factors using this panel. Constant factors are excluded by the sample generation process.

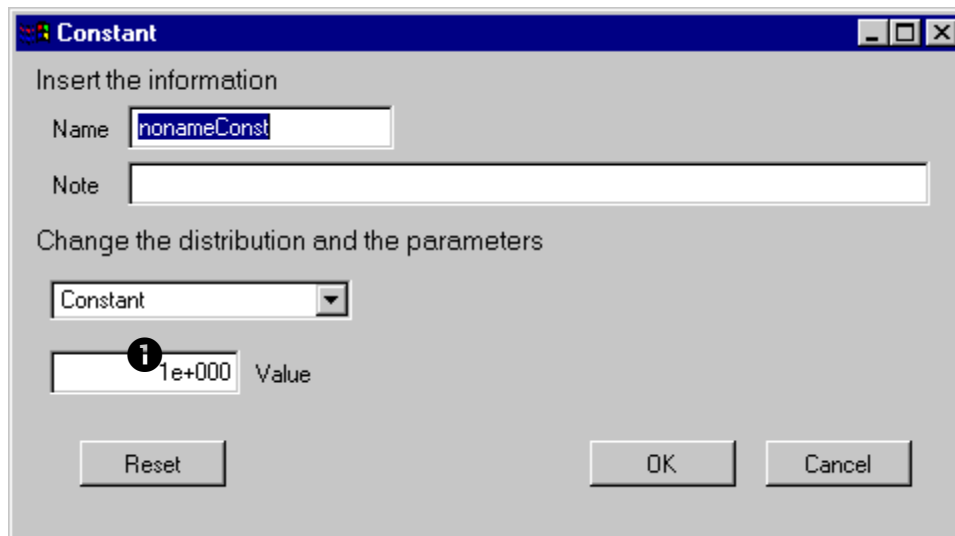


Figure 94 Constant distribution panel

- 1 This field specifies the value of the constant factor; the default value for a constant is 1.

Exponential distribution

The density function of the Exponential distribution has the following expression:

$$f(x) = \lambda \cdot \exp[-\lambda \cdot (x - b)]$$

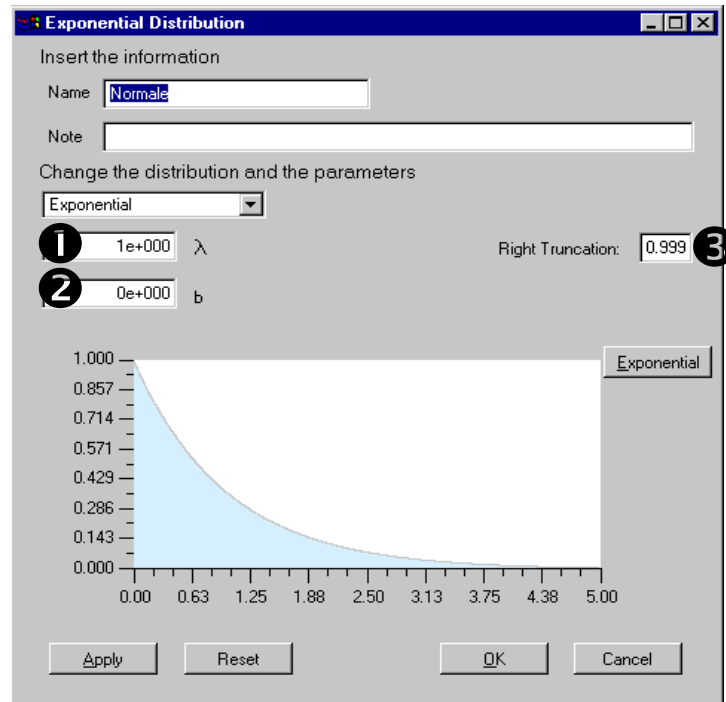


Figure 95 Exponential distribution panel

Distribution parameters

- ❶ λ value with $\lambda > 0$; the default value for λ is 1.
- ❷ b value with $b > 0$ and $x \geq b$; the default value for b is 0.
- ❸ Right truncation value; the default value is 0.999.

Mean value: $\frac{1}{\lambda} - b$

Standard deviation: $\frac{1}{\lambda^2}$

Gamma distribution

The density function of the Gamma distribution has the following expression:

$$f(x) = \frac{\lambda^r}{\Gamma(r)} * (x - b)^{r-1} * \exp(-\lambda * (x - b))$$

Distribution
parameters

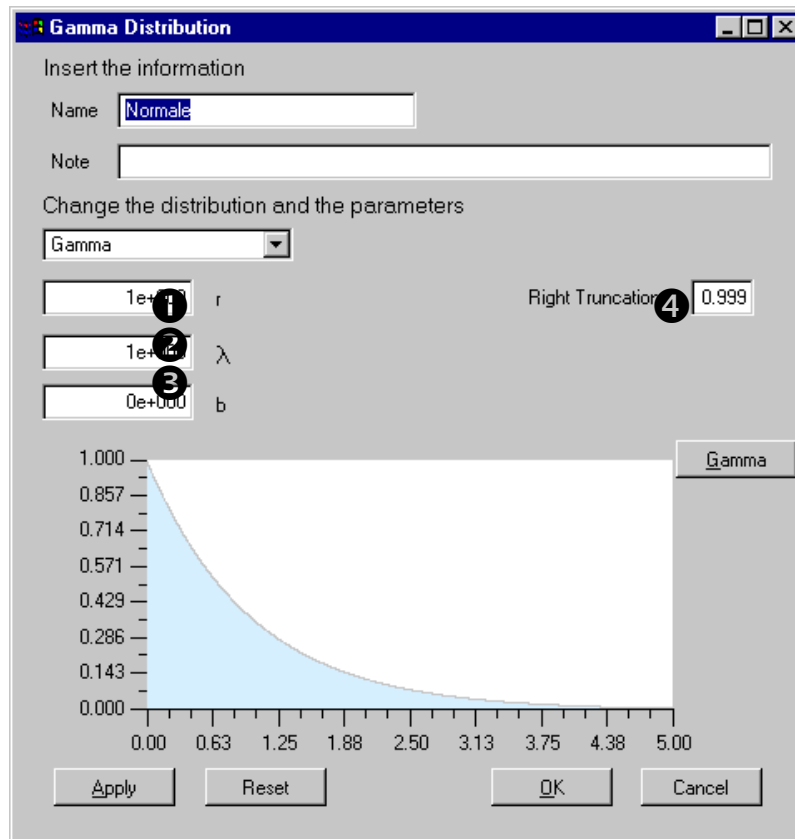


Figure 96 Gamma distribution panel

- ❶ r value with $r > 0$; the default value for r is 1.
- ❷ λ value with $\lambda > 0$; the default value for λ is 1.
- ❸ b value with $b \geq 0$ & $x \geq b$; the default value for b is 0.
- ❹ Right truncation value; the default value is 0.999.

Mean value: $\frac{r}{\lambda} + b$

Standard deviation: $\sqrt{\frac{r}{\lambda^2}}$

Beta distribution

The density function of the Beta distribution density function has the following expression:

$$f(x) = \frac{1}{B(\alpha, \beta)} * \frac{1}{b-a} * \left[\frac{x-a}{b-a} \right]^{\alpha-1} * \left[\frac{b-x}{b-a} \right]^{\beta-1}$$

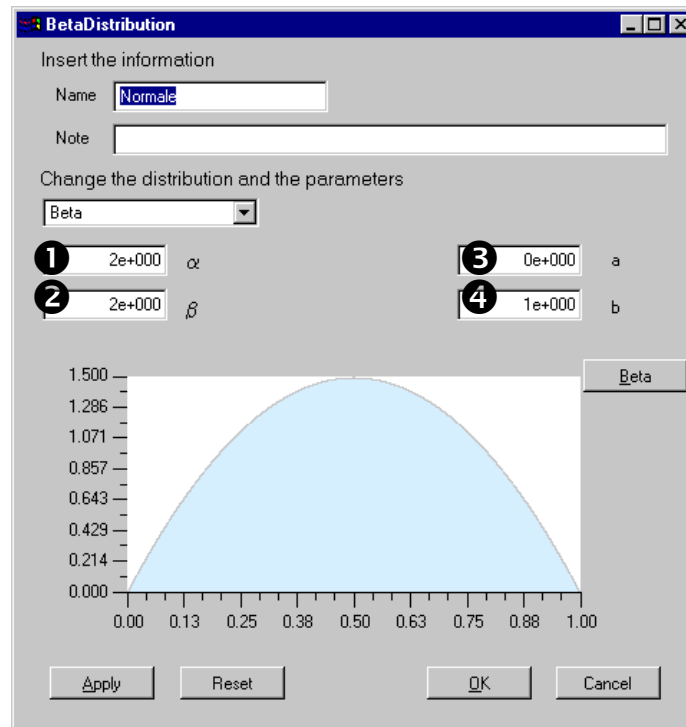


Figure 97 Beta distribution panel

Distribution parameters

- ❶ α value with $\alpha > 0$; the default value for α is 2.
- ❷ β value with $\beta > 0$; the default value for β is 2.
- ❸ a value with $a < b$; the default value for a is 0.
- ❹ b value; the default value for b is 1.

Mean value: $\frac{\alpha}{\alpha + \beta} * (b - a)$

Standard deviation: $\sqrt{\frac{\alpha * \beta}{(\alpha + \beta + 1) * (\alpha + \beta)^2} * (b - a)^2}$

Triangular distribution

The density function of the Triangular distribution has the following expression:

$$f(x) = \begin{cases} \frac{2 * (x - \alpha)}{(\beta - \alpha) * (\gamma - \alpha)} & \alpha \leq x \leq \beta \\ \frac{2 * (\gamma - x)}{(\gamma - \beta) * (\gamma - \alpha)} & \beta \leq x \leq \gamma \end{cases}$$

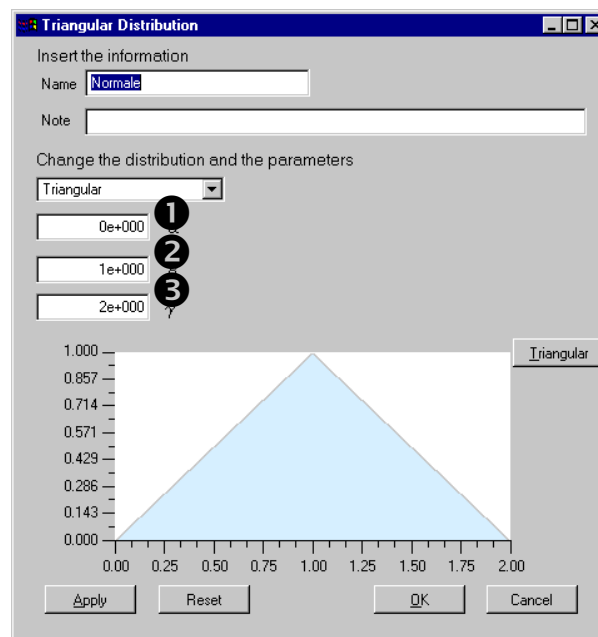


Figure 98 Triangular distribution panel

Distribution parameters

- ❶ α value; the default value for α is 0.
- ❷ β value; the default value for β is 1.
- ❸ γ value with $\alpha \leq \gamma < \beta$ or $\alpha < \gamma \leq \beta$; the default value for γ is 0.5.

Height of β point (h):
$$\frac{2}{(\gamma - \alpha)}$$

Mean value (μ):
$$\frac{h}{3(\beta - \alpha)} \cdot [\beta^3 - \alpha^3] - \frac{\alpha \cdot h}{2(\beta - \alpha)} [\beta^2 - \alpha^2] + \frac{-h}{3(\gamma - \beta)} [\gamma^3 - \beta^3] + \frac{1}{2} \left(\frac{\beta \cdot h}{(\gamma - \beta)} + h \right) \cdot [\gamma^2 - \beta^2]$$

Standard deviation:

$$\left(\frac{h}{4(\beta - \alpha)} [\beta^4 - \alpha^4] - \frac{\alpha \cdot h}{3(\beta - \alpha)} [\beta^3 - \alpha^3] + \frac{-h}{4(\gamma - \beta)} [\gamma^4 - \beta^4] + \frac{1}{3} \left(\frac{\beta \cdot h}{\gamma - \beta} + h \right) \cdot [\gamma^3 - \beta^3] - \mu^2 \right)^{1/2}$$

Relation

This panel represents a special case among the input factors selection panels; the user can create a relation between inputs factors using the expression editor.

Panel description

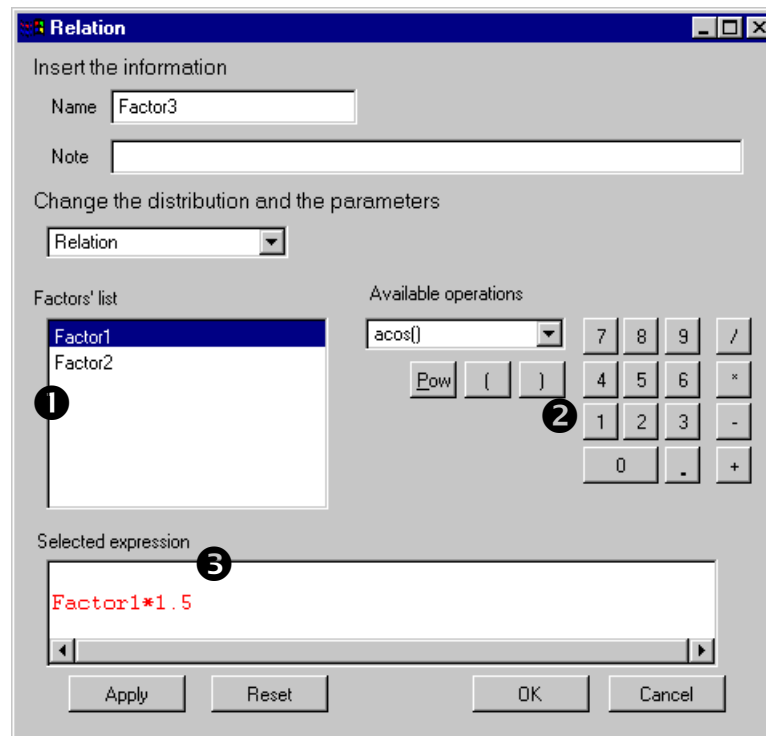


Figure 99 Relation panel

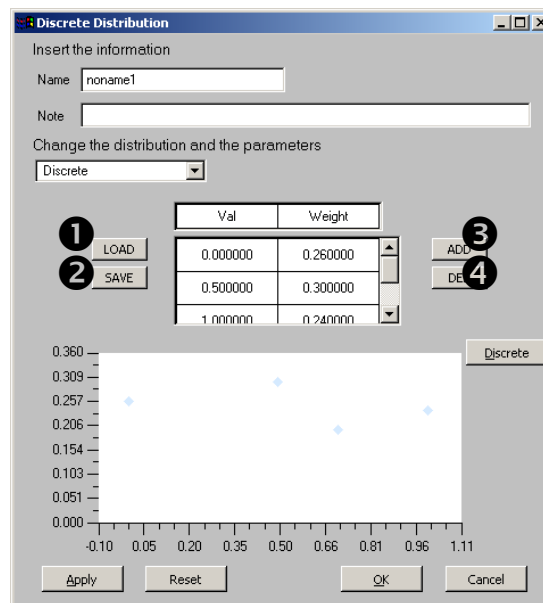
- ❶ **Factors list.** This field shows the available inputs factors; by double clicking an input factor, it is possible to insert that factor in the expression visualized in ❸.
- ❷ **Available operations.** In this section of the panel the user can select some functions and basic operations to apply to the inputs factors. This section has the same functionality of a calculator with field ❸ as display.
- ❸ **Expression field.** This field displays the current expression; it is possible to insert the expression through “Available operations” or directly through the keyboard.

Discrete distribution

The density function of the discrete distribution has the following expression:

$$f(x) = \text{Weight}_i \cdot I_{(x_1 \dots x_n)}$$

where $I_{(x_1 \dots x_n)}$ is 1 in the x's points 0 otherwise.



Buttons description

Figure 100 Uniform distribution panel

- ❶ **LOAD button.** The user can select this button to load a set of points for the current discrete distribution. This list of points will overwrite any previous list of points present in the matrix.
- ❷ **SAVE button.** This button allows the user to save in a file the current points for the discrete distribution.
- ❸ **ADD button.** This button is used to insert a new point for the current discrete distribution. For each interval, the user has to insert the value of the point and the relative weight of the point. Points can't overlap and the sum of the weights must be equal to 1.
- ❹ **DEL button.** With this button the user may delete any point present in the matrix. The user must select one interval before pressing the DEL button. After deleting an interval, the weight of the remaining elements must be corrected.

Parameters area

This area contains the parameters that characterise the discrete distribution. Weight w_i is the weight of point i the sum of weights must be equal to 1.

By double clicking an item in the 'Parameters area' the user can modify its value .

Mean Value: $\sum_{i=1}^n w_i * x_i$ for each x in which the distribution is defined

Standard deviation: $\sqrt{\sum_i (x_i - \mu_x)^2 * w_i}$

APPENDIX B: THREE TEST CASES FOR SIMLAB

TEST CASE 1: GIS APPLICATION

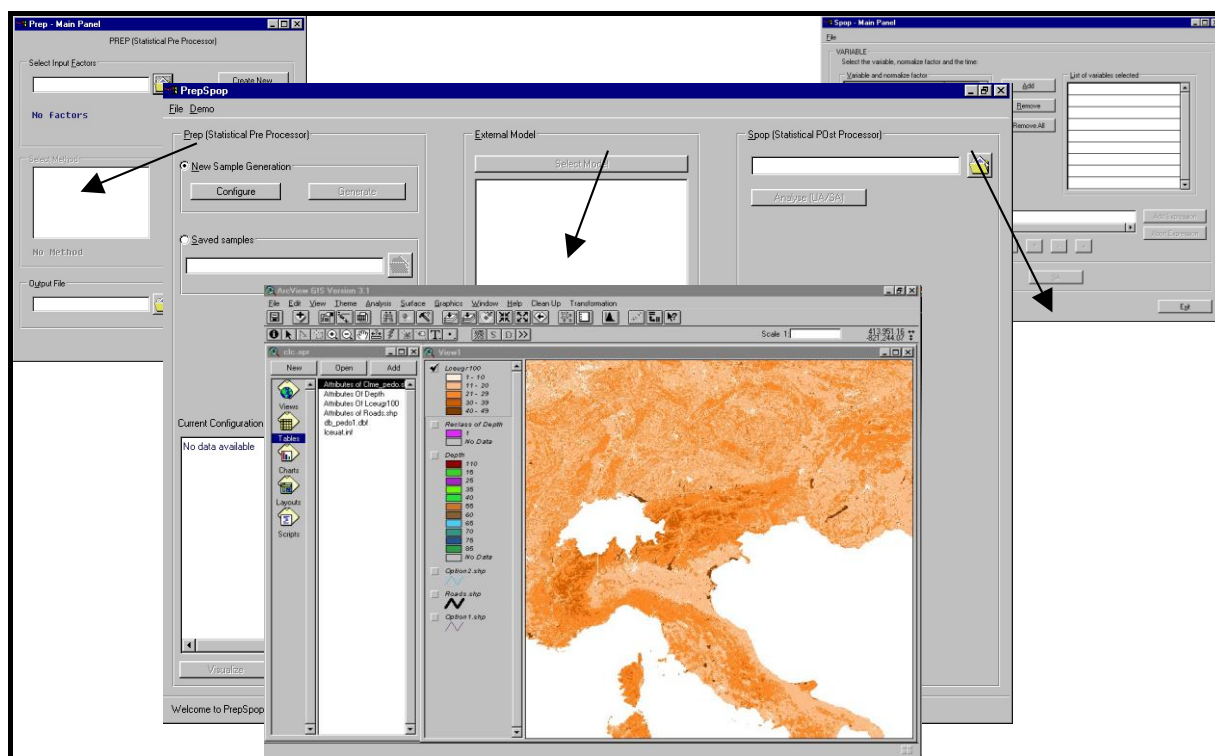


Figure 101 Interface between SimLab and GIS ArcView

DESCRIPTION OF THE TEST CASE

An interesting policy debate took place about the set up of a highway that link the Malpensa 2000 airport with the motorway Milan-Turin that is approximately 18 Km far from the airport. The area of interest is located in the Southern region of Malpensa 2000.

In road planning problems, the first stage is to establish a consensus upon the approximate layout of the new route. Assessing environmental issues upon two or more variants in the planning of the route is a necessary step to build such a consensus. This step might be seen as a preliminary strategic environmental assessment that could be conducted by calculating environmental indicators. In this test case we have selected the “Indicator for area loss and damage” I_{LD} . High values for this indicator denote high levels of area loss and damage.

It was assumed that two variants for the new highway are proposed. For each variant the “Indicator for area loss and damage” was evaluated and the option with the lower value for I_{LD} (e.i. I_{LD1} or I_{LD2}) was considered as preferable.

In order to quantify how much one variant is environmentally better than the other, a suitable function $Y = f(I_{LD1}, I_{LD2})$ was defined and considered as the target variable of the uncertainty and sensitivity analysis.

The indicator I_{LD} was evaluated by aggregating spatially distributed weights, which were assigned to the land parcels that belong to the region of interest. The weights are numerical values aimed at representing the potential economic/environmental costs to be sustained for building the highway. For instance, high-quality land coverage and productive soil capacity, and urbanised areas have high expropriating costs. The weight associated to watercourses is high because of the road infrastructures required to cross them. Also, the environmental cost of cutting protected areas with a highway must be taken into account.

OBJECTIVES

The first objective of this study is to investigate model relevance, which aims at distinguishing live components of the model — which drive model response and are hence relevant, from dead ones, which make no contribution to the variation in the model predictions. The second objective is to assess uncertainty importance in data resolution against uncertainty in subjective assumptions, model parameters, or other. To this purpose, all the uncertain factors need to be ranked in order of importance. This requires performing sensitivity analysis, which can be thought here as a tool for costs optimisation of data collection procedures. The third objective is to improve model robustness by means of sensitivity analysis. This consists in seeking for the most important factors and acting on them for reducing the overall level of uncertainty in model predictions.

In order to reach these objectives, PREP/SPOP software has been used. This is an instrument useful to the modeller to test the suitability of the underlying model to decision-making and as tools for assessment to policy decisions. In this particular the software permits to investigate how various resolution levels in the input data can affect the uncertainty in the indicators and in the target function. This is important for optimally allocating resources in GIS data acquisition.

INTERFACE BETWEEN PREPSPOP AND ARCVIEW

The processing of GIS data (vector and raster at different resolution levels) and related spatial analysis were performed, for the present test case, using ESRI ArcView GIS 3.1. An Avenue code was developed for the computation of the indicators I_{LD} from spatially distributed maps of land use, land coverage, soil properties and existing and planned networks. The value for land capability map (lc) has been obtained by simply adding three components: the land cover/land use weights, the weights for protected areas and the potential for cultivation values. Each cell value in the map of land capability is a numerical weight that yields the overall value/cost of the corresponding land parcel.

The indicator for area loss and damage I_{LD} has been evaluated by using the lc values and the road networks. Buffer regions around the existing and planned road networks were considered. The indicator I_{LD0} for the reference situation was defined as

$$I_{LD0} = \frac{\sum_{i \in \Omega_0} A_i lc_i}{\sum_i A_i}$$

where A_i and lc_i are the area and the values of land capability for pixel i . Ω_0 is the buffer area built around the existing road network.

The indicator I_{LD1} , which is aimed at quantifying the joint impact of the existing road network and the first variant, was defined as

$$I_{LD1} = \frac{\sum_{i \in \Omega_1} A_i lc_i}{\sum_i A_i}$$

where Ω_1 is the buffer area built around the existing road network and the first variant of the planned road, i.e. $\Omega_1 \supset \Omega_0$. The indicator I_{LD2} was computed using the same formula and adding the lc values over all $i \in \Omega_2$, where Ω_2 is the buffer area built around the existing road network and the second variant of the planned road.

The model output could be defined using two formal expressions of $Y = f(I_{LD1}, I_{LD2})$:

$$Y_1 = \log\left(\frac{I_{LD1} - I_{LD0}}{I_{LD2} - I_{LD0}}\right) \text{ and}$$

$$Y_2 = \log\left(\frac{I_{LD1}}{I_{LD2}}\right),$$

were given and considered in the study. Both expressions are in principle plausible to point out the less impact option, although the discrepancy between Y_1 and Y_2 can be considerable. In both expressions, positive values for the model output indicate that the second *variant* is preferable, and vice-versa. In the first expression, the subtraction of the reference indicator corresponds to the removal of the existing environmental burden over the region under study. The first expression is more sensible to portions of planned road with a high environmental impact in terms of area loss and damage. In the second expression, where the existing burden is not removed, the inclusion of a new road may change the situation very little, if the existing road concentration is high. There is a political dimension in the choice of the target.

PREP/SPOP was used to perform a number of uncertainty and sensitivity analysis tests. The software was interfaced with ArcView (see *Figure 101*) in order to allow the user to modify at her/his discretion ranges and distributions of the input factors considered in the study, as well as to use other techniques of sensitivity analysis that are available in PREP/SPOP (*Figure 102*). To inspect the complete set of results of this test case see the LOT 20 (Lot 20 of EUROSTAT SUP COM 1998) final report.

- Heuvelink G.B.M (1998). Error Propagation in Environmental modelling with GIS. Taylor & Francis, London.
- Iman, R. L., and Hora, S. C., (1990). A robust measure of uncertainty importance for use in fault tree system analysis. *Risk Analysis*, 10(3):401-406.
- Jones C.B . (1997). *Geographical Information Systems and Computer Cartography*. Longman, Harlow.
- Lodwick, W., Monson, W., Svoboda, L., 1990, Attribute error and sensitivity analysis of map operations in geographic information systems: suitability analysis. *Int. J. Geographical Information Systems* 4(4), 413-428.
- McKay M. D., Beckman R. J. and Conover W. J. (1979) A comparison of three methods of selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2), 239-245
- Pornon H., (1992). *Les SIG , mise en ouvre et applications*. Editions Hermès, Paris.
- PREP/SPOP (1999) *Software for Uncertainty and Sensitivity Analysis, User Manual*, POLIS-JRC/ISIS Ispra 1999.
- Saltelli, A., K. Chan, M. Scott, Editors, (2000), *Handbook of mathematical and statistical methods for sensitivity analysis of model output*, John Wiley & Sons publishers, Probability and Statistics series, forthcoming.
- Saltelli, A., S. Tarantola, K. P.-S. Chan, (1999), A quantitative model-independent method for global sensitivity analysis of model output, *Technometrics*, **41** (1), 39-56.
- Shi W., 1998, A generic statistical approach for modelling error of geometric features in GIS, *Int. J. Geographical Information Science*, 2(2), pp. 131-143.
- Sobol', I. M., (1993), Sensitivity analysis for nonlinear mathematical models, *Mathematical Modelling & Computational Experiment*, **1**, 407-414.
- Tarantola, S., L. Pastorelli, M. G. Beghi, C. E. Bottani, A dataless pre-calibration analysis in solid state physics, in Saltelli et al., Eds. 2000.

TEST CASE2: ECONOMIC APPLICATION

1. INTRODUCTION

One of the typical problems tackled in econometrics is forecasting the future value of an endogenous variable based on the past values of a number of exogenous ones. In the quite rare case of deterministic exogenous variables, the forecast is said to be unconditional, i.e. it does not depend on the probability density function (PDF) of the exogenous variables, given their deterministic character.

Unfortunately, in most econometric applications exogenous variables must be forecasted too, so the forecast becomes conditioned by the prediction of the exogenous variables. The stochastic nature of the input factors has two effects on the reliability of forecasts. First, the variance of the forecast is higher since the variance of the inputs increases the variance of the conditional expectations. Second, the choice of the most reliable forecasting model is far more difficult in the conditional case. A model

that shows a good fit in terms of unconditional forecasting may show a poor performance in the conditional case, because of the variance induced by the input factors. Often it may be complex to discriminate between several models with traditional econometric tests, because in most econometric applications the PDF of the forecast is not normally distributed. In particular it might be difficult to identify the relation between the variance of conditional expectation and the variance of any single input factor.

This application is a Variance based application to econometrics which aim is to show how the variance-based sensitivity analysis could improve econometric model screening and the analysis of forecast reliability.

FAST sampling method is used to analyze the sensitivity of the conditional forecast of two classical case studies: Solow's (1957) classic study of the technical change in U.S. Economy and the consumption function of Klein's (1950).

CASE STUDY 1: SENSITIVITY ANALYSIS OF SINGLE REGRESSION MODEL, SOLOW CLASSIC

Solow classic model suggests an aggregate production function where the aggregate output depends on capital and labor. One of the goals was to determine if aggregate output could be expressed as a function of labor and capital only, technological change being considered exogenous as stated in neoclassic growth theory. Solow (1957) tested five different models; in this exercise we will introduce and analyze the sixth one. A screening among the six models with the regression analysis on the sample 1909-1939 has been done. The series has been truncated in 1939 to avoid the World War II effect and the use of the dummy variable for the data after 1943, as suggested by Greene (1997). From these six models the sixth is the one with the best fit according to econometric tests. It will be use to forecast the outcome for 1940 and to perform sensitivity analysis that permits to determine which input factors influence the most the output variance. Forecast and sensitivity analysis will be performed in a Monte Carlo framework, i.e. we scan the entire PDF of model output instead of the neighborhood of the central tendency.

MODEL SELECTION

Solow (1957) suggested the following aggregate production function $q(t)=A(t)f[k(t)]$, where $q(t)$ is aggregate output per work-hour (\$/hours), $k(t)$ is the aggregate capital labor ratio, and $A(t)$ is the technology index. We estimate the following five models, proposed by Solow (1957), involving two coefficients b_0 and b_1 :

Model 1: $q/A = b_0 + b_1 * \ln(k)$ Eq. 1

Model 2: $q/A = b_0 - b_1/k$ Eq. 2

Model 3: $\ln(q/A) = b_0 + b_1 * \ln(k)$ Eq. 3

Model 4: $\ln(q/A) = b_0 - b_1/k$ Eq. 4

Model 5: $q(t)/A(t) - q(t - 1)/A(t - 1) = b_0 + b_1 * k(t - 1)$ Eq. 5

We have also consider the model:

Model 6: $\ln(q(t)/A(t)) - \ln(q(t - 1)/A(t - 1)) = b_0 + b_1 * (\ln(k(t)) - \ln(k(t - 1)))$ Eq. 6

Note that all models express a linear relationship between q/A or a transformation of it (natural logarithmic transformation, or first difference logarithmic transformation) and k or a transformation of it (natural logarithmic, first difference logarithmic or an inverse transformation).

All models can be estimated with a single regression. From previous analysis follows that:
 Model (5) can be eliminated because of the poor goodness of fit statistics (R^2);
 All models are affected by the 2nd order positive serial correlation, but Model (1) and (5) in weaker form;
 The best functional form is given by Model (6). In the following pages more attention is given to this model.

The forecast of the q/A ratio are performed using Model (6) where:

$$b_0 \sim N(\hat{b}_0, \hat{\sigma}_{b_0}) \quad \text{Eq. 7}$$

and

$$b_1 \sim N(\hat{b}_1, \hat{\sigma}_{b_1}) \quad \text{Eq. 8}$$

In order to forecast the ratio q/A at time $T+1$, we need to know the value of k in the same time period. Several model specification were investigated, the best one is given by an auto-regressive model computed on the first difference of $\ln(k)$.

Thus, the value of k in $T+1$ can be computed using the following model:

$$d\ln(k(t)) = a_1 d\ln(k(t-1)) \quad \text{Eq. 9}$$

where $d\ln(k(t))$ is defined as:

$$d\ln(k(t)) = \ln(k(t)) - \ln(k(t-1)) \quad \text{Eq. 10}$$

and

$$a_1 \sim N(\hat{a}_1, \hat{\sigma}_{a_1}) \quad \text{Eq. 11}$$

MONTE-CARLO FORECAST AND VARIANCE DECOMPOSITION

The final form of the 1-step ahead forecasting model can be found by substituting equation (Eq.9) into equation (Eq.6) and after some simplifications we have:

$$d\ln(q(t+1)/A(t+1)) = f(b_0, b_1, a_0) = b_0 + b_1 * (a_1 * d\ln(k(t))) \quad \text{Eq. 12}$$

The traditional way of forecasting $d\ln(q(t)/A(t))$ at time $T+1$, would be to take the expected value of the left hand side of equation (Eq.11). Far more complex would be the task of drawing its confidence intervals, because $d\ln(q(t+1)/A(t+1))$ is not normally distributed because of the product $b_1 * a_1$. Pindick and Rubinfeld (1997) suggested an estimate of the confidence interval for the forecast:

A sample of $N=9,987$ using FAST sampling scheme has been performed, where the three factors: a_1, a_2 and a_3 are Normal distributed with specified parameters.

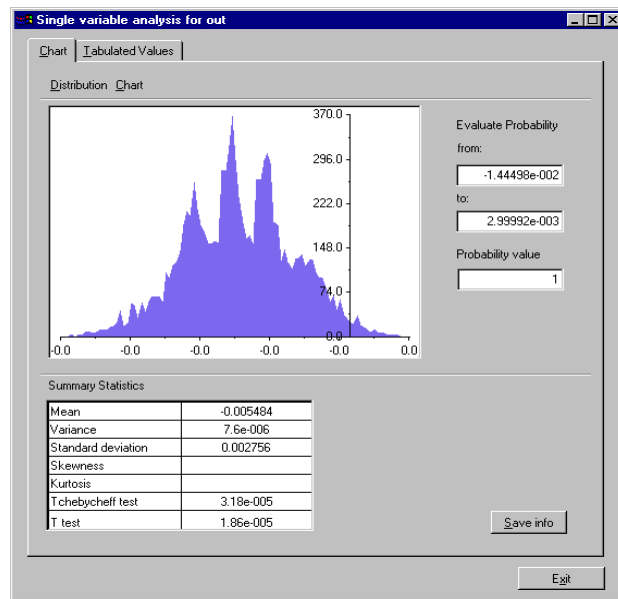


Figure 103: Probability Density Function of Solow Model Conditional Forecast

The expected value of the distribution can be estimated by:

$$\hat{E}[y(T + 1)] \cong \hat{f}_0 = \frac{1}{N} \sum_{n=1}^N f(b_{0i}, b_{1i}, a_{1i}) = -0.0055 \quad \text{Eq. 13}$$

and its estimated variance is given by:

$$\text{Var}[y(T + 1)] \cong \hat{D} = \frac{1}{N} \sum_{n=1}^N f^2(b_{0i}, b_{1i}, a_{1i}) - \hat{f}_0^2 = 7.87E - 06 \quad \text{Eq. 14}$$

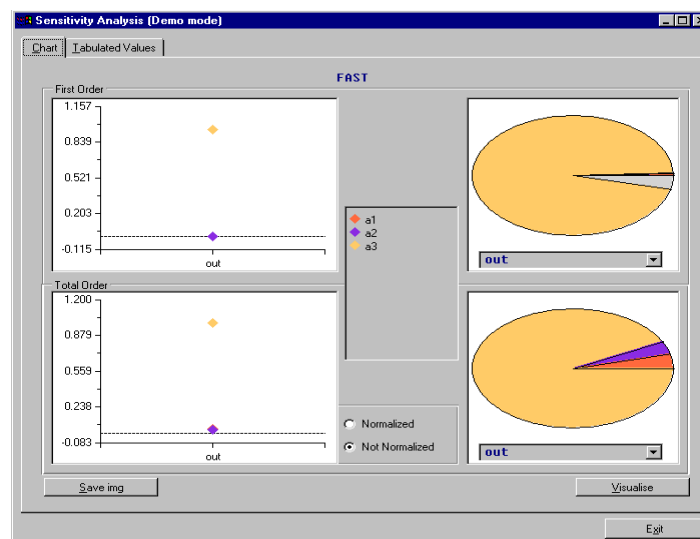


Figure 104: Sensitivity analysis of Solow growth model

It is also possible to draw the confidence intervals without assumption of normality, since it is possible to build a percentile distribution for any PDF.

The results obtained from FAST indices are the following table:

Factors	Fast first order indexes	Fast total order Indexes
a1	0.00	0.04
a2	0.00	0.04
a3	0.96	0.99

The most evident result is that the most important factor is *a3* that captures the 96% of the output variance. This is the coefficient of the autoregressive process of $dln(k(T+1))$. The other indices, first and second order, does not exceed 1%.

The overall result of the analysis is that the variance of the forecast depends almost exclusively on the volatility of the prediction of the capital labor ratio, while the production function plays almost no role in the explanation of the variance of the conditional forecast. So Solow' production function performed extremely well in the unconditional case, while the sensitivity analysis pointed out that in the conditional case the performance is rather poor. In economic terms, we are in the situation in which the model itself is perfect, but it would be better not to base your decisions on it. Nor are we in the condition to choose a better forecasting model since all the proposed models present only the capital labor ratio as exogenous variable. The forecast obtained with any of the other models would simply increase the variance of the output. So always from an economic point of view we can state that the growth of production cannot expressed as a function of the capital labor ratio, because we would not be able to predict the evolution of production, so Neoclassic growth theory does not hold. It must be mentioned that we could find the same results with the analysis of regression statistics.

CASE STUDY 2: SENSITIVITY ANALYSIS OF MULTIPLE REGRESSION MODEL, LINEAR VERSUS NON LINEAR MODELLING; KLEIN1 CONSUMPTION EQUATION

Klein (1950) Model 1 is a classic example of simultaneous equation model. Here we will focus only on the consumption function:

$$C_t = \alpha_0 + \alpha_1 P_t + \alpha_2 P_{t-1} + \alpha_3 (W_t^p + W_t^g) \tag{Eq. 15}$$

Equation (14) express Consumption C_t , as a linear function of prices P_t and P_{t-1} , private wages W_t^p and government wages W_t^g . Klein estimated the parameters using U.S. data for 1921 to 1941. We decided, once more, to truncate the series (the data corresponding to 1941 clearly show break comparing to the previous trend). Again we will try to identify the forecasting model that gives the more reliable results, but we will treat a different situation.

First we will consider a situation that involves uncertainty not only in the model functional form, but also in the selection of variables that should be included in the model. Second we will consider a forecasting horizon of 5 years and we will compute Fast indexes for each acceptable model in each time period.

MODEL SELECTION

In the first test case we analyzed several models, but all of them were linear models of a transformation of the capital/labor ratio. Here we considered all the combination of the three variables proposed by Klein and several transformations of them (first difference, first logarithmic difference), we also estimated non-linear models such as logarithmic, exponential and others non-linear models. While for Solow production function model (6) gave far better results then the other models, for Klein 1 we found three acceptable models:

$$\text{Model 1: } C_t = b_{01} + b_{11}P_{t-1} + b_{21}W_t^P \quad \text{Eq. 16}$$

$$\text{Model 2: } dC_t = b_{02} + b_{12}dP_t + b_{22}d(W_t^P + W_t^S) \quad \text{Eq. 17}$$

$$\text{Model 3: } C_t = b_{03} + P_{t-1} b_{13} + (W_t^P + W_t^S) b_{23} \quad \text{Eq. 18}$$

Estimation Statistics are given in Table 1.

Table 1: regression statistics of Klein 1 consumption function

	Ex. Var.	b0	S.E b0	t b0	prob. b0	b1	S.E b1	t b1	prob b1	b2	S.E b2	t b2	prob b2
Model 1	C	11.16	1.75	6.37	0.00	-0.30	0.08	-3.37	0.00	1.32	0.07	18.80	0.00
Model 2	d(C)	0.51	0.31	1.63	0.12	0.42	0.16	2.64	0.01	0.52	0.15	3.24	0.00
Model 3	C	9.08	1.41	6.41	0.00	0.64	0.07	8.15	0.00	0.98	0.01	82.14	0.00

	Ex. Var.	R-sq.	Adj. R-sq.	S.E. reg.	Sum sq. res.	Log lh	DW	Akaike.	Schwarz .	F-stat.	Prob(F-stat.)
Model 1	C	0.97	0.97	1.05	19.09	-27.91	2.28	3.09	3.24	295.00	0.00
Model 2	d(C)	0.82	0.79	1.27	27.72	-31.64	1.97	3.46	3.61	38.00	0.00
Model 3	C	0.97	0.96	1.18	25.36	-31.78	1.63	3.31	3.46	293.90	0.00

Legend:

Ex. Var. :	Exogenous Variable	Log lh	Log likelihood
b*:	Expected value of coefficient b*	DW	Durbin Watson statistics
S.E. b*:	Standard error of coefficient b*	Akaike.	Akaike. information criterion
t b*:	t statistic of coefficient b*	Schwarz .	Schwarz . Information criterion
prob. B*:	2 tailed probability of t statistics	F-stat.	F. Statistics
Adj. R-sq.	Adjusted R squared	Prob(F-stat.)	2 tailed probability of F statistics
S.E. reg.	Standard error of regression		
Sum sq. res.	Sum of squared residuals		

All three models are acceptable in terms of goodness of fit and serial correlation. The three models are rather different. The first one does not include government wages, the second one is defined on the first difference of all variables and the third is an inherently non-linear model where the parameters appear as exponents. As for Solow model all parameters follow a normal distribution (model 3 has been estimated via an iterative linearization method, so b_{13} and b_{23} too follow a normal distribution (See Pindick and Rubinfeld (1997) pp. 235-238).

Concerning the forecast of the exogenous variables the three following models gave the better results in term of goodness of fit:

$$d\ln(P_t) = a_{01} + a_{11}d\ln(P_{t-1}) \tag{Eq. 19}$$

$$d\ln(W_t^p) = a_{02} + a_{12}d\ln(W_{t-1}^p) \tag{Eq. 20}$$

$$d(W_t^g) = a_{03} + a_{13}d(W_{t-1}^g) \tag{Eq. 21}$$

Estimation outputs are reported in

Table 2

	Ex. Var.	a0	S.E a0	t a0	prob. a0	a1	S.E a1	t a1	prob a1	
Model 30	dln(p)		0.02	0.05	0.47	0.64	0.12	0.24	0.51	0.61
Model 31	dln(wp)		0.02	0.02	1.25	0.22	0.24	0.21	1.12	0.27
Model 32	dln(wg)		0.46	0.09	5.13	0.00	-0.63	0.18	-3.47	0.00

	Ex. Var.	R-sq.	Adj. R-sq.	S.E. reg.	Sum sq. res.	Log lh	DW	Akaike.	Schwarz .	F-stat.	Prob(F-sta	
Model 30	dln(p)		0.02	-0.04	0.23	0.90	2.00	1.86	0.00	0.09	0.27	0.61
Model 31	dln(wp)		0.06	0.01	0.08	0.12	20.82	1.40	-1.98	-1.88	1.26	0.27
Model 32	dln(wg)		0.41	0.37	0.32	1.73	-4.18	1.99	0.65	0.75	12.00	0.00

Legend:

Ex. Var. :	Exogenous Variable	Log lh	Log likelihood
b*:	Expected value of coefficient b*	DW	Durbin Watson statistics
S.E. b*:	Standard error of coefficient b*	Akaike.	Akaike. information criterion
t b*:	t statistic of coefficient b*	Schwarz .	Schwarz . Information criterion
prob. B*:	2 tailed probability of t statistics	F-stat.	F. Statistics
Adj. R-sq.	Adjusted R squared	Prob(F-stat.)	2 tailed probability of F statistics
S.E. reg.	Standard error of regression		
Sum sq. res.	Sum of squared residuals		

Table 2: Regression results of prices, private wages and public wages AR(1) models.

MONTECARLO FORECAST AND VARIANCE DECOMPOSITION

The three forecasting models are obtained by the substitution of (19), (20) and (21), in equation (22), (23) and (24), leading to:

Model 1:

$$C_t = f(b_{01}, b_{11}, b_{21}, a_{01}, a_{11}, a_{02}, a_{12}) = b_{01} + b_{11} \left(e^{a_{01}} \frac{P_{t-2}^{a_{11}+1}}{P_{t-3}^{a_{11}}} \right) + b_{21} \left(e^{a_{02}} \frac{W_{t-1}^P a_{12}+1}{W_{t-2}^P a_{12}} \right)$$

Eq. 22

Model 2:

$$\begin{aligned}
 dC_t &= f(b_{02}, b_{12}, b_{22}, a_{01}, a_{11}, a_{02}, a_{12}, a_{03}, a_{13}) = \\
 &= b_{02} + b_{12} \left(e^{a_{01}} \left(\frac{P_{t-1} a_{11}^{+1}}{P_{t-2} a_{11}} - \frac{P_{t-2} a_{11}^{+1}}{P_{t-3} a_{11}} \right) \right) + b_{22} \left(e^{a_{02}} \left(\frac{W_{t-1}^P a_{12}^{+1}}{W_{t-2}^P a_{12}} - \frac{W_{t-2}^P a_{12}^{+1}}{W_{t-3}^P a_{12}} \right) + a_{03} + a_{13} d(W_{t-1}^g) \right)
 \end{aligned}$$

Eq. 23

Model 3:

$$\begin{aligned}
 C_t &= f(b_{03}, b_{13}, b_{23}, a_{01}, a_{11}, a_{02}, a_{12}, a_{03}, a_{13}) = \\
 &= b_{03} + \left(e^{a_{01}} \frac{P_{t-2} a_{11}^{+1}}{P_{t-3} a_{11}} \right) b_{13} + \left(\left(e^{a_{02}} \frac{W_{t-1}^P a_{12}^{+1}}{W_{t-2}^P a_{12}} \right) + \left(a_{03} + (a_{13} + 1)W_{t-1}^g - a_{13}W_{t-2}^g \right) \right) b_{23}
 \end{aligned}$$

Eq. 24

where all $a_{..}$ and $b_{..}$ are normally distributed with the means and variances given in Table 1 and Table 2. Note that the values of prices and wages at the previous times periods are always obtained from equation (19), (20) and (21), so the input factors are the parameters $a_{..}$ and $b_{..}$ while prices and wages enter only as a deterministic constant valued at the last time period: 1939-1940 for both wages variable and 1938-1939 for prices values. It also rather evident that any forecast obtained by the three models does not present a normal distribution since they all involve products, exponential and even powers of normally distributed factors. So the goodness of fit statistics of the model estimates clearly does not apply in the conditional forecast case.

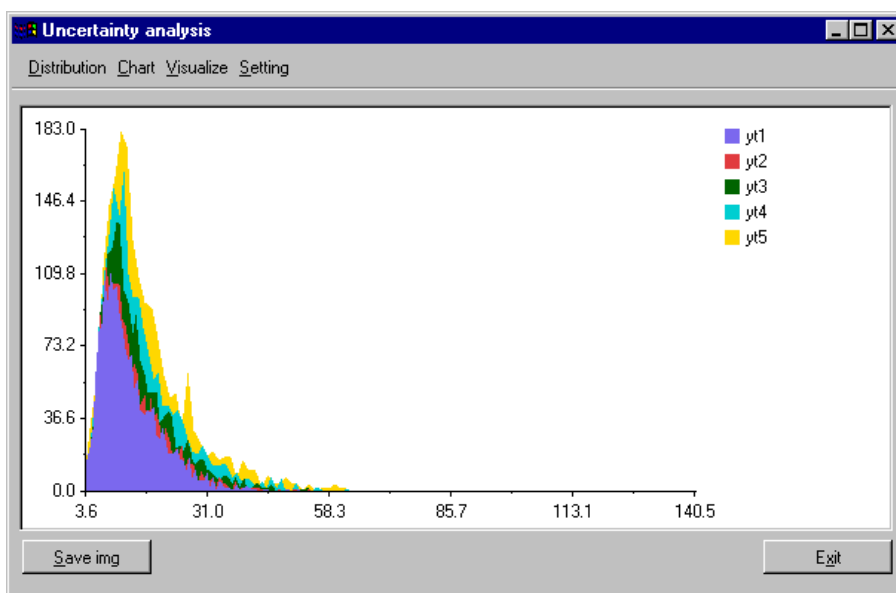


Figure 105: UA on Klein 3 model

FAST sampling method has been applied to the set of parameters of the three models. The first model has fewer factors (7) than the second and third ones (9). First and total order indices are computed by PREPSPOP.

Results are reported in *Figure 106*

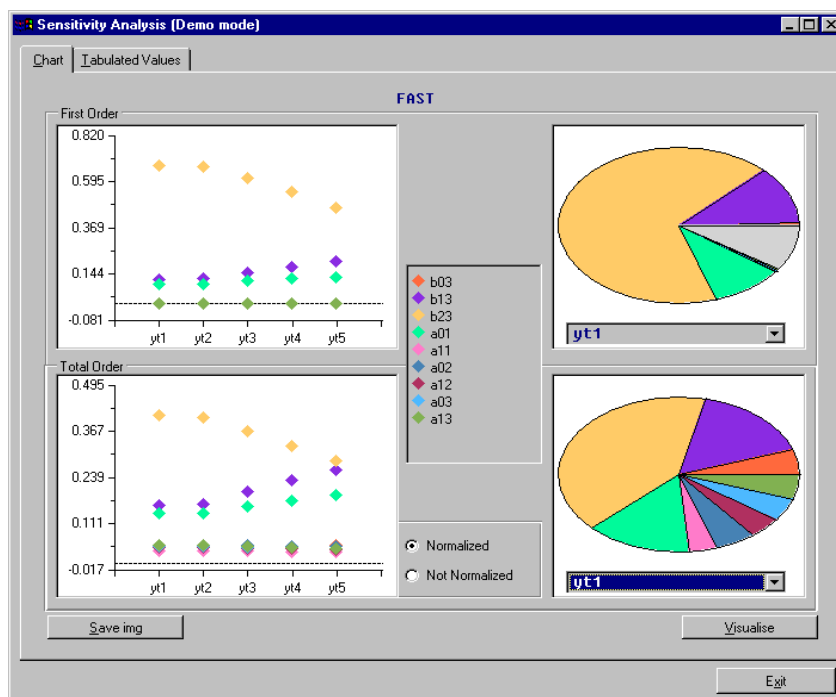


Figure 106: SA on Klein 3 model

Model1

Fast first order indexes

b01	0.05	0.04	0.04	0.03	0.03
b11	0.04	0.04	0.04	0.04	0.04
b21	0.09	0.08	0.07	0.05	0.04
a01	0.51	0.52	0.53	0.54	0.54
a11	0.00	0.00	0.00	0.00	0.00
a02	0.01	0.01	0.01	0.01	0.01
a12	0.01	0.01	0.01	0.01	0.00

Fast total order indexes

b01	0.17	0.17	0.17	0.17	0.17
b11	0.21	0.21	0.22	0.22	0.23
b21	0.23	0.23	0.21	0.20	0.19
a01	0.86	0.88	0.89	0.90	0.90
a11	0.21	0.22	0.22	0.23	0.23
a02	0.23	0.24	0.24	0.24	0.25
a12	0.23	0.23	0.23	0.24	0.24

Model2

Fast first order indexes

b02	0.02	0.00	0.00	0.00	0.00
b12	0.05	0.07	0.08	0.08	0.08
b22	0.05	0.02	0.02	0.01	0.01
a01	0.27	0.46	0.52	0.55	0.55
a11	0.03	0.01	0.00	0.00	0.00
a02	0.00	0.00	0.00	0.00	0.00
a12	0.00	0.00	0.00	0.00	0.00
a03	0.00	0.00	0.00	0.00	0.00
a13	0.47	0.07	0.06	0.01	0.00

Fast total order Indexes

b02	0.09	0.17	0.17	0.18	0.19
b12	0.21	0.36	0.36	0.38	0.38
b22	0.21	0.28	0.28	0.30	0.30
a01	0.36	0.83	0.85	0.90	0.90
a11	0.18	0.06	0.06	0.05	0.05
a02	0.06	0.08	0.07	0.07	0.07
a12	0.04	0.04	0.05	0.07	0.07
a03	0.07	0.09	0.09	0.09	0.09
a13	0.58	0.16	0.15	0.09	0.09

Model3

Fast first order indexes

b03	0.01	0.01	0.00	0.00	0.00
b13	0.12	0.12	0.15	0.18	0.21
b23	0.68	0.67	0.61	0.55	0.47
a01	0.10	0.10	0.11	0.12	0.13
a11	0.00	0.00	0.00	0.00	0.00
a02	0.00	0.00	0.00	0.00	0.00
a12	0.00	0.00	0.00	0.00	0.00
a03	0.00	0.00	0.00	0.00	0.00
a13	0.00	0.00	0.00	0.00	0.00

Fast total order Indexes

b03	0.10	0.10	0.10	0.10	0.10
b13	0.31	0.32	0.39	0.46	0.53
b23	0.78	0.78	0.72	0.65	0.58
a01	0.27	0.27	0.31	0.35	0.39
a11	0.07	0.07	0.07	0.07	0.07
a02	0.10	0.10	0.10	0.10	0.10
a12	0.09	0.09	0.08	0.08	0.08
a03	0.09	0.09	0.09	0.09	0.09
a13	0.10	0.10	0.10	0.09	0.09

Several interesting observations can be made for interpreting the results obtained from UA and SA applied to these three models, but the aim of this project is the evaluation of PREPSPOP in different

environments. In the test case examined PREPSPOP demonstrated various appealing features giving to the economist a proficuous help in the choice of the factors that better explain the output function.

REFERENCE

- Greene W. H. (1997), *Econometric Analysis*, Prentice-Hall
- Klein L. (1950), *Economic Fluctuations in the United States 1921-1941*. New York: John Wiley and Sons.
- Pindick R. e Rubinfeld D. (1997), *Econometric Models & Economic Forecast*, McGraw-Hill.
- Solow R. (1957), ‘Technical Change and the Aggregate Production Function’, *Review of Economics and statistics*, 39 pp. 312-320.

TEST CASE 3: FINANCIAL APPLICATION (BEI)

INTRODUCTION

Financial institutions such as Banks are continuously facing risks. Banks face *market risks*, associated with the unknown behaviour of the market variables, and *credit risks*, arising from the possibility of default by one of the counterparts involved in a financial transaction. It is essential for a proper functioning of a Bank to find strategies to manage risk.

To hedge the market risk deriving from a given contract, the Bank tries to enter into another contract that offset such risk. The final objective is that of building a portfolio, i.e. a collection of assets and liabilities, which has minimum risk (ideally risk-less).

This problem was faced in 1994 by JP Morgan, who introduced what is nowadays known as *Value at risk* (VaR). Value at risk is an attempt to provide a single number summarising the total risk in a financial portfolio. A Value at risk calculation aims at making a statement such as: “the Bank is X percent certain that will not lose more than Y Euro’s in the next N days”.

To build a portfolio minimising risk, the Bank needs to value each of the financial instruments included in the portfolio. Complex financial models have been developed to describe and value the sophisticated financial instruments used by the Bank to meet the needs of clients. Models aim at predicting the behaviour of market variables and quantify the maximum risk that a Bank is facing.

Uncertainty and sensitivity analyses are very useful to the purpose of managing risk, either it be market or credit risk. To build a risk-less portfolio it is essential to estimate the maximum loss that any given financial instrument in the portfolio might cause (uncertainty analysis) and what was this due to (sensitivity analysis). In credit risk problems, uncertainty and sensitivity analyses may estimate the maximum loss of the Bank in case of default by one of its clients and what factors are mostly responsible for the value reached by this loss.

The present document illustrates how uncertainty and sensitivity analyses can be helpful tools for Bank operators in two different problem settings. An application of uncertainty and sensitivity analysis to a financial model is also shown.

THE USE OF UA AND SA IN THE WORLD OF FINANCE

AN EXAMPLE OF USE OF UA AND SA TO MANAGE CREDIT RISK

In the world of finance, a *swap* is a private agreement between two counterparts to exchange cash flows in the future according to a prearranged formula. For example, a very common type of swap is “plain vanilla” interest rate swap. A agrees to pay B cash flows equal to interest at a predetermined fixed rate on a notional principal for a number of years. B agrees to pay A cash flows equal to interest at a floating rate on the same notional principal for the same number of years. The floating rate is usually the LIBOR, which is the rate of interest offered by banks on deposits from other banks in Eurocurrency markets. In this case, the swap is used to transform a floating rate loan into a fixed rate loan and the other way around.

Most of the times, two counterparts do not get in touch directly to arrange a swap. There is a financial institution, such as a Bank, that acts as intermediate. The Bank enters into two separate contracts, one with counterpart A and one with counterpart B, and it does that in such a way that, unless one of the two counterparts defaults, it is certain to make a profit. The profit is partly to compensate the default risk it is bearing (*credit risk* of a financial institution).

Imagine the Bank is the intermediate in a swap contract. It is natural for the Bank to find a way to minimise the credit risk caused by the possibility of default of one of the counterparts. In order to do so, the Bank needs to estimate the maximum value that a swap may reach at maturity and also what factors are mostly causing this value to be what it is, i.e. what factors are mostly responsible for variations in the value of the swap.

Uncertainty analysis provides an answer to the first question by providing an estimate of the maximum value reached by the swap. Sensitivity analysis indicates what factors are mostly affecting the value assumed by the swap and, if the method used is quantitative, what are the percentages of variance in the swap value accounted by each factor.

AN EXAMPLE OF USE OF UA AND SA TO MANAGE MARKET RISK

The goal of a Bank is that of building a portfolio, i.e. a collection of financial assets and liabilities, which is neutral with respect to market risk (i.e. risk-less). In other words, the Bank aims at hedging its portfolio so to manage market risk.

To hedge the exposure deriving from a certain contract, the Bank tries to enter into an opposite position in exactly the same contract. This would be a “perfect hedging” but, in practice, is not straightforward. The contracts included in the portfolio of a Bank are often designed to meet the needs of a particular client, and it is very hard to find another client who accepts entering into an offsetting strategy. If it is not possible to offset completely the risk of a contract, it is necessary to carry out a careful analysis. The objective is to determine which of the available contracts have values that are most closely correlated with the value of the asset being hedged. Therefore, the Bank needs to determine the value of each element of its portfolio and the factors mostly responsible for variations in this value.

Complex financial models are used to reproduce the behaviour of the market variables underlying the contracts in the portfolio, and to estimate the value of each portfolio’s element. Uncertainty and sensitivity analyses, when applied to these models, can be of help for a Bank operator making decisions on the portfolio.

The problem of minimising the risk is already addressed by classical financial techniques such as “portfolio stress testing”, a technique that tests the impact of extreme market moves on the portfolio’s value. Uncertainty and sensitivity analyses are tools that, used in conjunction with classical financial techniques, may provide additional information useful to manage risk.

Uncertainty analysis provides an estimate of the maximum loss that any given financial instrument in the portfolio might cause. Sensitivity analysis identifies what is the loss due to.

When a perfect hedging of the portfolio is not feasible, uncertainty and sensitivity analyses become precious tools. If a Bank operator knows that the portfolio contains a financial instrument, say A , that could reach an extreme value of M (uncertainty analysis), he/her will look on the market for a financial instrument that offsets the loss (now quantified) met by the Bank when A ’s value is M .

If he/her knows that this value is mostly depending on factors a , b , and c , the financial instrument he/her looks for will be one depending on the same factors. This should avoid that an unpredictable movement of one of these factors may cause a dramatic change in the value of one of the two contracts but not in the other.

AN APPLICATION OF UA AND SA

INTRODUCTION

Here we illustrate an example of application of uncertainty and sensitivity analyses to a financial model. We imagine that the model is used to price a financial instrument included in the Bank’s portfolio and that the objective is hedging the portfolio. The financial instrument is an *interest-rate derivative*.

In finance’s terminology, *derivatives* are financial instruments whose values depend on the prices of other more basic variables. *Interest-rate sensitive derivatives* are derivatives whose values depend on the interest rates.

The interest rate derivative priced in this example is an *interest rate option*. *Option* contracts are of two kinds: call options and put options. A *call option* gives the holder the right to buy an asset by a certain date for a certain price. A *put option* gives the holder the right to sell an asset by a certain date for a certain price. Options involve the payment of an up-front fee. The price written in an option contract is known as the *strike price*. The date in the contract is known as the *maturity date*. A *European option* can be exercised only at maturity date; an *American option* can be exercised at any time during its life (i.e. before maturity).

An interest-rate option is simply an option where payoffs are dependent in some ways on the level of interest rates. For example, it may be an option of buying a bond whose value is exactly determined by the knowledge of the interest rates.

Uncertainty analysis allows to estimating the price (or value) of the option contract at present time within some uncertainty bounds. Sensitivity analysis tells what factors are mostly responsible for variations in this price. If the option is included in the Bank’s portfolio, results of uncertainty and sensitivity analyses give strong indications on how to choose a financial instrument that can hedge the portfolio.

THE HJM MODEL

To price the interest rate option we have used a model that belongs to the class of the so-called HJM models (Heath, Jarrow and Morton, [1]), a class of models very common in finance because they can be used to price a number of different financial instruments.

The HJM models are suitable for the pricing of *interest-rate sensitive derivatives* and provide some distinct advantages:

An HJM model is consistent with the assumption of “absence of arbitrage”. This assumption excludes the possibility to develop a trading strategy that allows taking advantage of the fact that two or more financial securities are mispriced relative to each other. Intuitively, absence of arbitrage keeps out the possibility of making money without any risk.

An HJM model accommodates path-dependence injected by non-Markovian process.

An HJM model can handle a range of securities, including those with American features.

The model is developed in discrete time; the discrete interval between dates is denoted by h . The value of the option is based on forward rates, denoted by $f(t, T)$. A forward rate $f(t, T)$ is the rate determined at time t for a lending transaction over the interval $(T, T + h)$. Thus $f(t, T)$ denotes the one-period forward rate at time T in the future as seen at time t . The process of forward rates is represented as

$$f(t+h, T) = f(t, T) + \alpha(t, T, f(t, T))h + \sigma(t, T, f(t, T))X\sqrt{h}$$

where

$\alpha(t, T, f(t, T))$ is the drift which may be a function of current time, future forward rate time, and the level of forward rates;

$\sigma(t, T, f(t, T))$ is the volatility coefficient representing a measure of the uncertainty on the value of the forward rates;

and X is a binomial variable taking on a value of either +1 or -1 with equal probability.

The model is univariate and leads to a branching process with two branches emanating from each node, so assuming the form of a binomial tree. The branching process begins with an initial forward curve $f(0, T)$. For example, if $T=3$ and the discrete interval between dates is $h=1$, we have $f(0,0), f(0,1), f(0,2), f(0,3)$. Then, knowledge of $f(0, T)$ allows the computation of two values at time h , $f(h, T, \omega_u)$ and $f(h, T, \omega_d)$, where ω_u indicates a movement up of the forward rate while ω_d denotes a movement down of the forward rate. The two values are placed in the two nodes of the tree at time h . Note that at time h , i.e. after one step, the curve of forward rates has one dimension less than $f(0, T)$. For example when $T=3$ and $h=1$ we have $f(1,1), f(1,2), f(1,3)$. The element $f(1,0)$ would be meaningless.

In general, knowledge of the values $f(t, T)$ allows direct extension into the next period's nodes $f(t, T, \omega)$, and the curve at time $(t + h)$ has one component less than the curve at time t .

Once the forward rates have been computed at each node of the tree by forward induction, the price of the interest-rate option can be derived at each node by backward induction. At time T of maturity the value of the option is known. The option pays off if the spot rate at maturity is greater than the strike rate. Therefore, the option price at time T is computed as

$$C(T) = \text{Max}[0, (f(T, T) - K) \times N]$$

where $f(T, T)$ is the interest rate at time T (the spot interest rate); K is the strike rate, N is the notional value on which the contract has been written.

At an earlier time t , $t < T$, the exact market value (price) of the option is unknown. But we can calculate a forecast of it using the information available at time t . The forecast is calculated assuming that the fair market value $C(t)$ will equal a properly discounted value of $E[C(T)|I(t)]$, where $I(t)$ denotes the information available at time t , and the expectation is taken with respect to the distribution function that governs the price movements under the assumption of absence of arbitrage. Therefore, at a given time t , the knowledge of the spot forward rate curve, combined with the knowledge of the prices of the options at the next period's nodes, allows to calculating a forecast of the option value (or price).

The model outcome looks like a binomial tree indicating at each node the estimated price of the option. Uncertainty and sensitivity analyses on the HJM model

Uncertainty and sensitivity analyses have been applied to the model when used for pricing an interest-rate option with nominal value of \$100, and using the following data:

Strike rate = $K = 0.075$;

$h = 0.50$;

T	$f(0,T)$	σ_f
0	0.06	0.015
1	0.07	0.012
2	0.08	0.011
3	0.09	0.010

The output of interest was chosen to be the price of the option at time $t=0$. In other words the output analysed is that at time $t=0$. Note that, given the features of the model, that can accommodate even option with American features, the same analysis could have been applied at any time $t < T$.

The model input factors included in the analysis, and their distribution functions, all of them assumed to be normal, are the following:

Model input factors	Distributions Normal (μ , σ)
sigma0(2)	(0.012 , 0.0060)
sigma0(3)	(0.011,0.0055)
sigma0(4)	(0.010, 0.0050)
$f(0,1)$	(0.06, 0.015)
$f(0,2)$	(0.07, 0.012)
$f(0,3)$	(0.08, 0.011)
$f(0,4)$	(0.09, 0.010)

The sensitivity analysis method chosen for the analysis is that of Sobol'. A number of 9975 model evaluations were performed.

RESULTS AND CONCLUSIONS

Results of the uncertainty analysis are represented by the histogram in *Figure 107*. The statistics of interest are indicated on the right low margin of the in *Figure 107*. For example it is shown that, under

the assumption that the model input factors follow the distributions indicated above, the price of the option can reach up to a maximum value of \$8.46, and that its value is on average \$2.88.

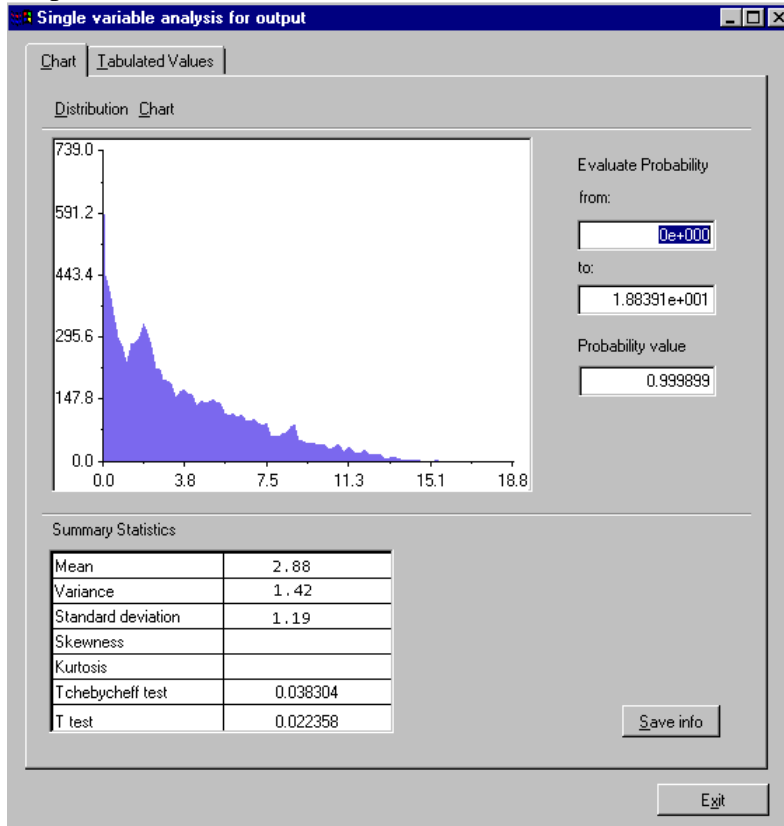


Figure 107 Uncertainty Analysis

The sensitivity analysis provides information on which of the 7 factors is mostly responsible for variations in the price. The Sobol’ sensitivity indices of the first order have a natural interpretation since they represent the fraction of the total output variance that is due to an individual factor. A graphical representation of the Sobol’ indices of first order is given in *Figure 108*. The pie represents the total output variance. The Sobol’ index relative to an input factor is represented as a slice of the pie.

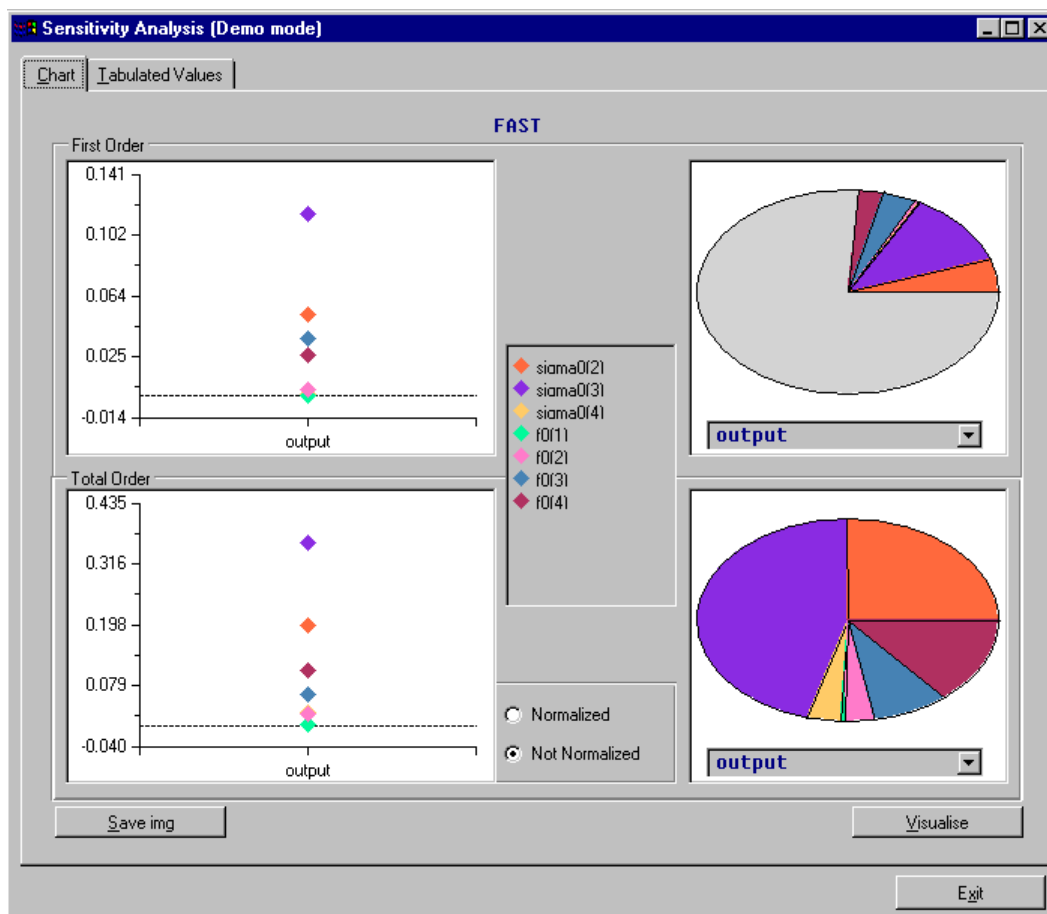


Figure 108 Sensitivity Analysis: Sobol's indices

For a linear model the Sobol' first order indices sum up to 1. If the model is non-linear some of the effects on the output are due to interactions among factors and are therefore not captured by first order sensitivity measures. The slice of the pie indicated by "rest" in the figure represents the percentage of output variance that is not due to linear effects and can be regarded as a measure of the model non-linearity.

From Figure 108 it is evident that the initial forward rates have a larger influence on the output than the volatility coefficients. The relative importance of these forward rates is inversely proportional to the distance from the maturity time T . The forward rate (as seen at time $t = 0$) relative to the maturity time T is the most influential factor on the option price. The forward rate (as seen at time $t = 0$) relative to the time $(T - 1)$ is the second most influential factor on the option price; and so on.

The information provided by uncertainty and sensitivity analyses of the option model is very useful for managing market risk. The maximum risk that the Bank is facing by holding the option contract in its portfolio is quantified by the uncertainty analysis. Sensitivity analysis provides indications on how to hedge the Bank's portfolio by highlighting the dependencies of the option value on other financial variables.

In order to apply UA/SA techniques for this kind of financial test case, it has been necessary interface PREP/SPOP with an executable file generated using MATLAB. In the follow figure:

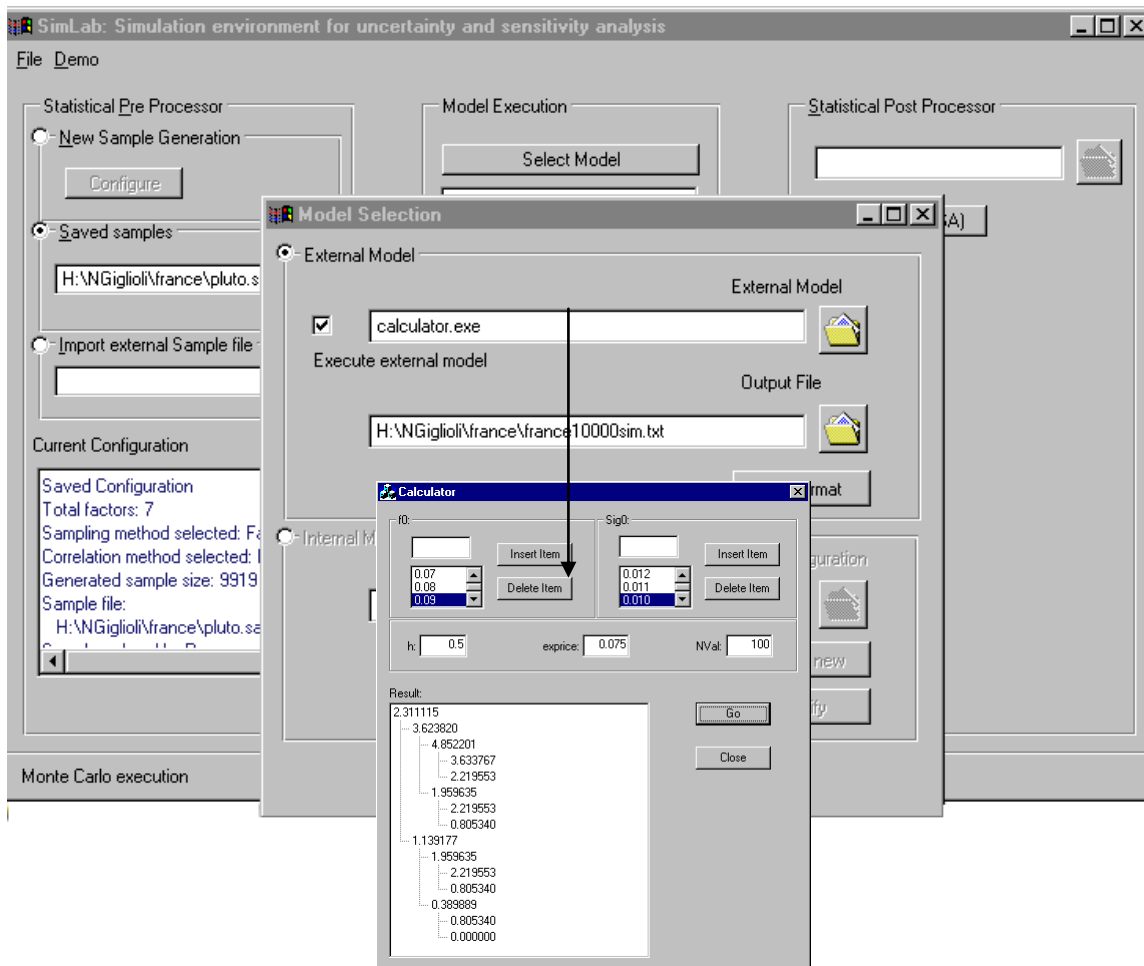


Figure 109 Interface between PREPSPOP and Calculator.exe

This section describes the file formats that are created during the SimLab working sessions.

INPUT FACTORS FILE FORMAT (.FAC)

Input factors are generated in the first step of the Monte Carlo analysis. For each factor the user selects a range and a distribution type with distribution specific parameters; these information are saved in a file having *.fac* extension.

The *.fac* file contains the following information:

default truncations values that the user can change using the default truncations window;

total number of inputs factors;

then, for each input factor, the file contains:

factor type;

factor name;

factor notes;

factor specific parameters (dependent on the factor type).

Again, for each factor, is listed the relative dependence tree correlation information:

Dependence tree correlation value for the factor;

Dependence tree parent of the factor in the correlation tree (-1 for item without parents);

number of children in the dependence correlation tree;

for each child: his **position** in the list of factors;

all the values in the **Iman Conover correlation matrix**.

Stein correlatin information

The correlation method selected by the user: 0 dependence tree, 1 Iman Conover, 2 stein correlation 3 no correlation

The following is an example of a file *.fac*:

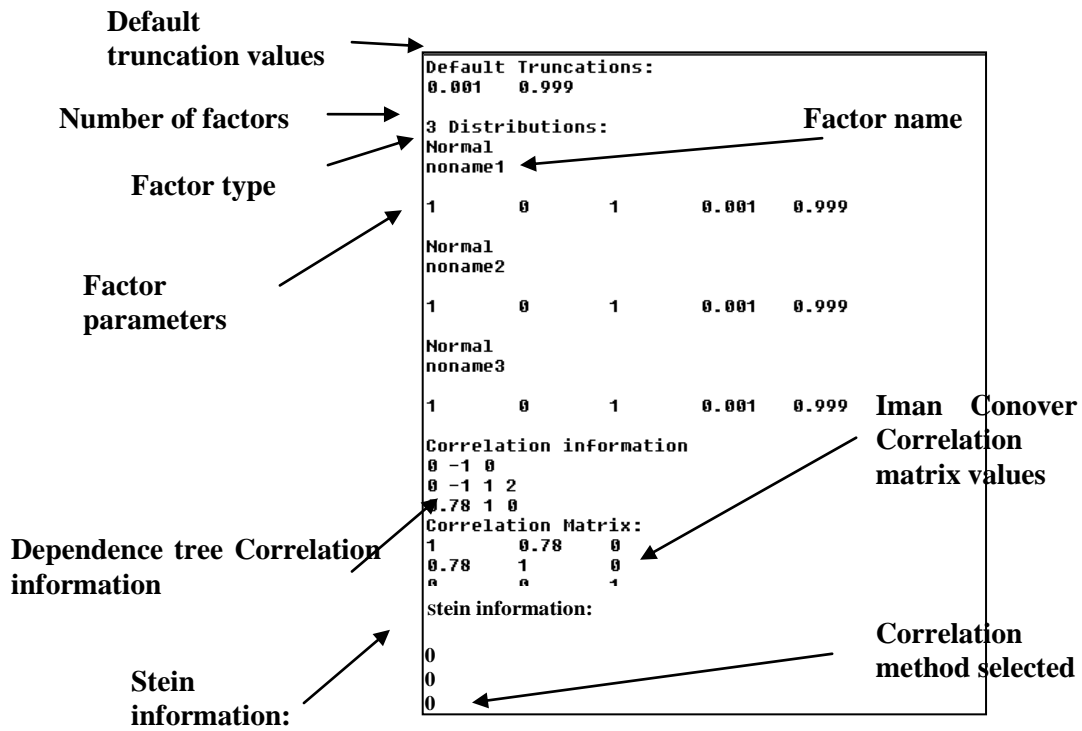


Figure 110 fac file example

SAMPLES FILE FORMAT (.SAM)

SimLab saves the samples generated using the current factors configuration and the selected sampling method in a file having *.sam* extension.

A *.sam* file contains the following information:

Configuration flag: 0 internal configuration, 1 imported configuration

total number of samples;

number of input factors;

a flag that indicates if the samples are ordered **by run** or **by factor**;

the **complete sample design**; if the samples are ordered by run each, row will contain the samples for all the factors in the current run; if the samples are ordered by factor, each row will contain the samples for the current factors in all the runs.

the first lines following the sample elements contain the information regarding the factors configuration recorded in the same way as in the description of the **fac file**;

id of the sampling method used for the generation;

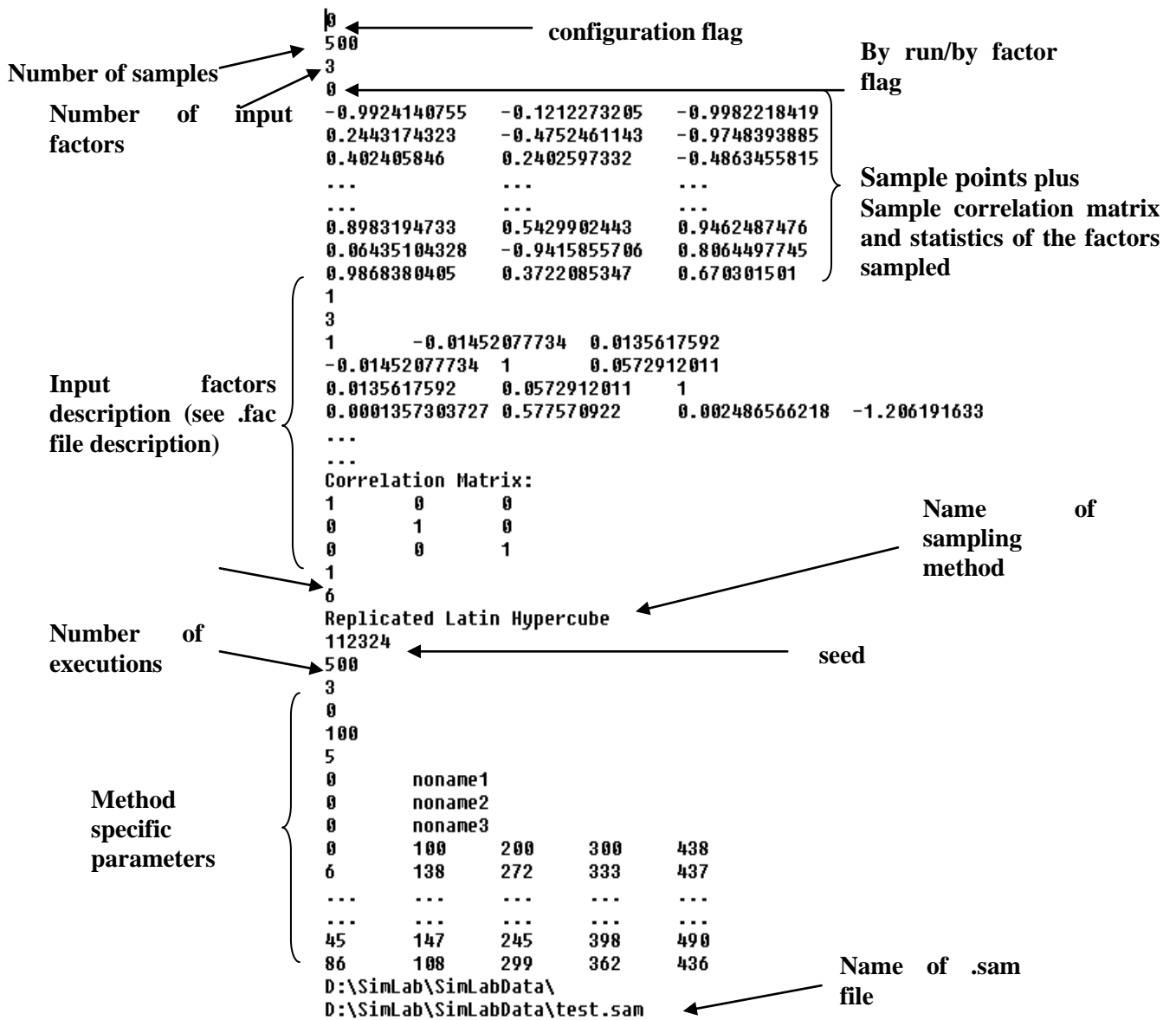
name of the sampling method used for the generation;

number of executions requested for the sampling method;

method specific parameters;

name and the path of the current sample file.

The following is an example of a *.sam* file:



EXTERNAL SAMPLE FILE FORMAT

The external file format allows to import an external sample file in the SimLab environment.

The external sample file must contain the following information:

Number of runs

Number of factors

For each factor the **Factor name** in a new line

The sample matrix order by run

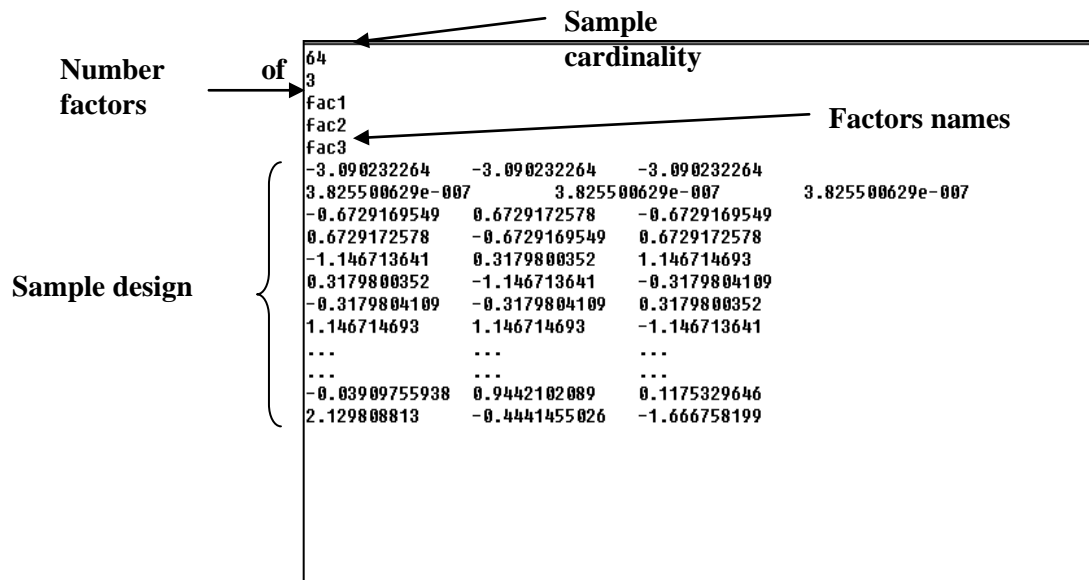


Figure 112: External sample file format

SIMLAB CONFIGURATION FORMAT FILE (.PSC)

The user can save the information about a working session in a file with .psc extension. This file contains all the data that allows the system to recreate the current condition in the future. The psc file contains the first part related to the Prep configuration (sam file), the middle related to the model configuration and the last to the UA/SA analysis (see spc file)

EXTERNAL MODEL OUTPUT FILE FORMAT

The external model output file will be generated by the execution of an external model. This output file must be in a format that SimLab could read.

In particular, the external model should use the following scheme when writing its output file:

- a) if the model is not time dependent:
 - the **number of variables** produced by the model;
 - for each variable, its **name** (on a different row);
 - a row containing: **time = no**;
 - the **number of model executions**;
 - for each execution:
 - for each output variable,
 - its **value**;

- b) if the model is time dependent:
 - the **number of variables** produced by the model;
 - for each variable, its **name** (on a different row);
 - a row containing: **time = yes**;
 - the **number of model executions**;
 - for each execution:
 - the word **RUN** followed by the **sequential run number**;
 - the **number of values for the time** in the current run;
 - for each value of the time:
 - the **value of the time point**;
 - for each output variable, its **value**;

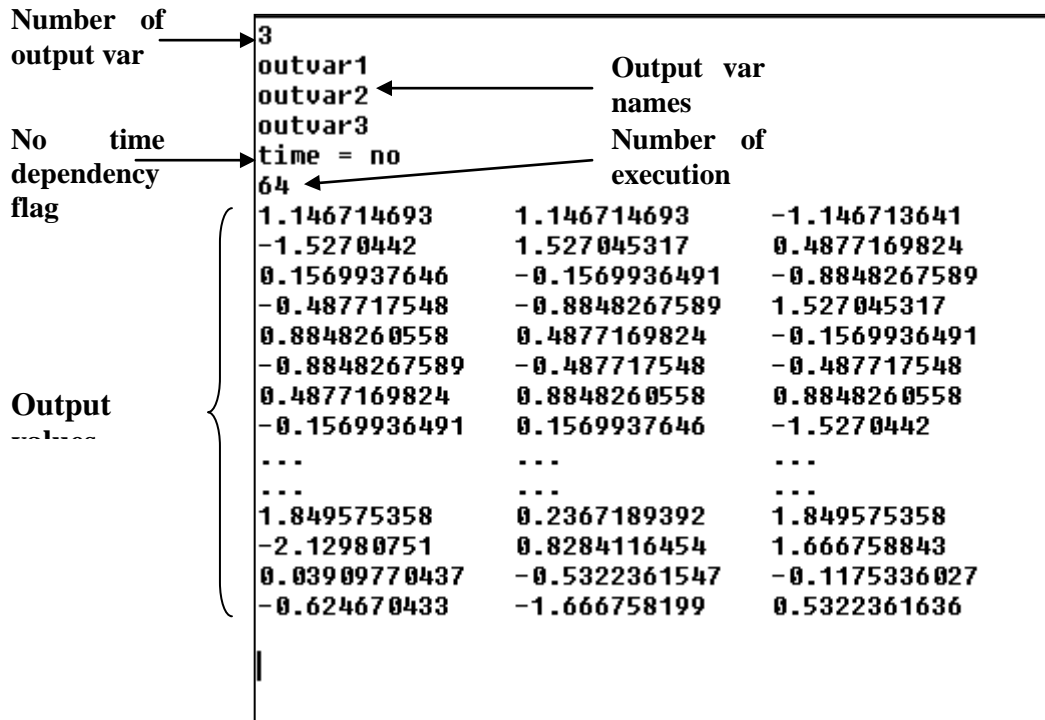


Figure 113: no time dependent model file format

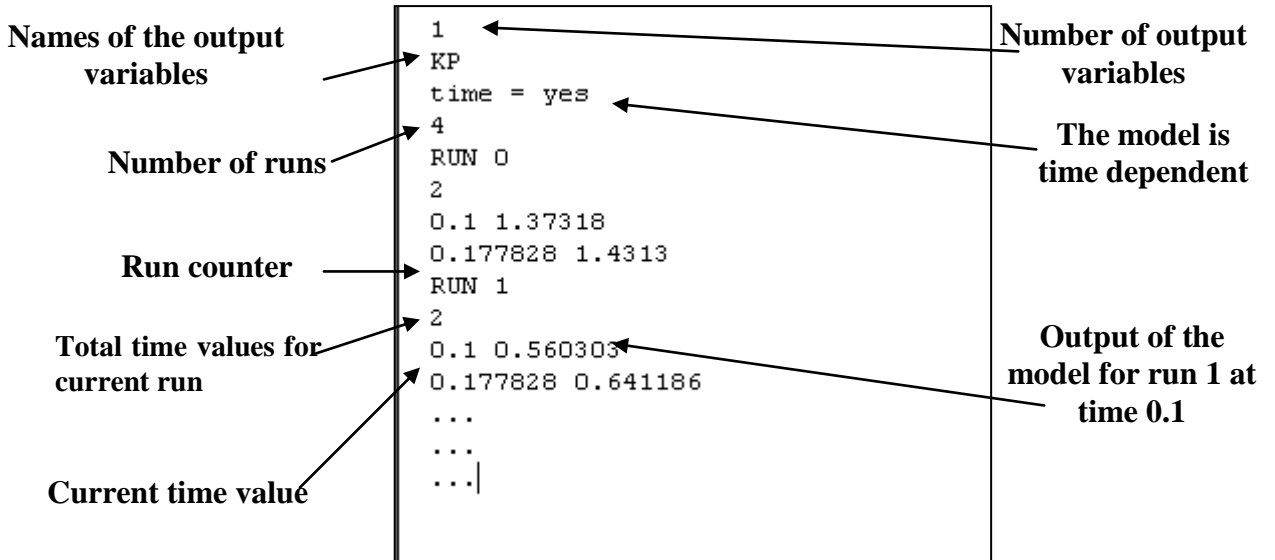


Figure 114 Time dependent output format.

POST PROCESSOR FILE FORMAT (.SPC)

The user can save the information about an uncertainty/sensitivity analysis session in a file with .spc extension. This file contains all the relevant data for the UA/SA analysis and its format depends on the sampling method selected using Prep. Moreover, to save time, it also contains the values of the sensitivity analysis indexes, once they have been computed the first time.

The general format of a Post Processor file is the following:

Samples information (the same explained in the discussion of the .sam file format);

Model configuration;

Number of variables selected for UA/SA analysis;

For each UA/SA variable, its **name**;

Number of variables produced by the model;

For each output variable, its **name** and (on the next line) the relative **weight** for the variable;

The remaining information depend on the sampling method selected by the user:

REFERENCES

- [1] Bedford T., Modelling with dependent uncertainties: a toolbox of techniques, 1997.
- [2] Draper, N. R., and Smith H., (1981). Applied Regression Analysis. John Wiley & Sons, New York.
- [3] T. Homma, A. Saltelli, Importance measures in global sensitivity analysis of model output, Reliability Engrg. System Safety 52 (1) (1996) 1–17.
- [4] Iman R. L. and Shortencarier M. J. (1984), A FORTRAN 77 program and user's guide for the generation of Latin hypercube and random samples for use with computer models. SANDIA Laboratory report SAND83-2365.
- [5] Iman R. L., and Conover W. J., (1982). “A distribution free approach to inducing rank correlation among input variables”, Comm. Statist.-Simula. Computa, 11(3), 311-334.
- [6] Iman R. L., Davenport J. M. and Zeigler D. K. (1980), Latin hypercube sampling (Program user's guide). SANDIA Laboratory report SAND79-1473.
- [7] Iman, R. L., and Helton, J. C., (1988). A comparison of uncertainty and sensitivity analysis techniques for computer models. Risk Analysis, 8(1):71-90.
- [8] Ishigami, T. and Homma, T. (1990). “An Important Quantification Technique in Uncertainty Analysis for Computer Models”, In the proceedings of the First International Symposium on Uncertainty Modelling and Analysis, pp398-403.
- [9] Meeuwissen A.M.H. & R.M. Cooke, Tree dependent random variables, Report 94-28 TWI, Delft University of Technology, 1994.
- [10] Morris M. D. (1987), Two Stage Factor Screening Procedures Using Multiple Grouping Assignments, Communications in Statistics - Theory and Methods, 16(10), 3051-3067
- [11] Morris M.D. (1991), Factorial Sampling Plans for Preliminary Computational Experiments. Technometrics, vol. 33, no. 2.
- [12] OECD/NEA PSAC User group, 1989, *PSACOIN Level E Intercomparison, An international code intercomparison exercise on a hypothetical safety assessment case study for radioactive waste disposal systems*, OECD/NEA publication. Paris.
- [13] Saltelli, K. Chan, M. Scott, Mathematical and statistical methods for sensitivity analysis, John Wiley & Sons publishers, Probability and Statistics series, to appear in 2000.
- [14] Saltelli A., Tarantola S., and Chan K., (1999). A quantitative, model independent method for global sensitivity analysis of model output. Technometrics, 41(1): 39-56.
- [15] Saltelli, A., and Bolado, R., (1998). An alternative way to compute Fourier Amplitude Sensitivity Test (FAST). Computational Statistic and Data Analysis, 26(4): 445-460.
- [16] Saltelli A., Tarantola S., and Chan K., 1997, *Sensitivity analysis of model output: an improvement of the FAST method*, EUR Report 17338 EN.
- [17] Saltelli and Sobol' (1995).”About the use of rank transformation in sensitivity analysis of model output”. Reliability Engineering and System Safety, 50: 225-239.

REFERENCES		
-------------------	--	--

- [18] Saltelli, A., Andres, T. H., and Homma, T., (1993). Sensitivity analysis of model output; An investigation of new techniques. *Computational Statistics and Data Analysis*, 15:211-238.
- [19] Saltelli, A., and Marivoet, J., (1990). Nonparametric statistics in sensitivity analysis for model output; A comparison of selected techniques. *Reliability Engineering and System Safety*, 28:229-253.
- [20] Shukhman B. (1993), Generation of quasi-random (LPt) vectors for parallel computation. *Computer Physics Communications* 78, 279-286.
- [21] Sobol' I.M. (1993) Sensitivity estimates for Non-linear Mathematical Models, *Mathematical Modelling and Computational Experiment* (1) 4 407-414.
- [22] Sobol' I.M., Turchaninov V.I., Levitan Y.L., and Shukhman B.V. (1992), Quasirandom sequence generators. Keldysh Institute of Applied Mathematics, Russian Academic of Sciences, Moscow.
- [23] Stein, M., (1987). Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, 29(2):143-151.
- [24] McKay, M. D., 1995, Evaluating prediction uncertainty, Los Alamos National. Laboratories report NUREG/CR- 6311, LA-12915-MS.

TABLE OF FIGURES		
------------------	--	--

TABLE OF FIGURES

Figure 1: SimLab main panel.....	9
Figure 2 Example of Simlab.ini	9
Figure 3: Main panel menu	10
Figure 4 Demo menu.....	10
Figure 5: Pre Processor frame	13
Figure 6: Current configuration area.....	15
Figure 7: New Sample Generation.....	16
Figure 8: Statistical Pre Processor - Main Panel.....	17
Figure 9 Select Input Factors frame	18
Figure 10 Input Factors Selection panel	19
Figure 11 Truncation values	20
Figure 12 Factors list area.....	21
Figure 13 Input factor summary.....	21
Figure 14: Correlation panel	23
Figure 20 Fast method panel.....	29
Figure 21 Fixed samples method panel.....	31
Figure 23 Morris method	34
Figure 24 Quasi-Random LpTau method panel.....	35
Figure 25 Random method.....	36
Figure 26: Replicated Latin Hypercube panel	37
Figure 27: Replicated Latin Hypercube subset selection Area	38
Figure 28 Setting output file panel.....	45
Figure 29 Output format selection	46
Figure 30 Samples generation.....	47
Figure 31 Pre Processor configuration loaded	48
Figure 32 Tabulated values panel	49
Figure 33 Correlation on sample data	49
Figure 35 Cobwebs graph panel.....	51
Figure 36 Cobwebs manual selection	52
Figure 37 Scatter Plot diagram.....	53
Figure 39 Histogram panel.....	55
Figure 40 External Model module panel.....	58
Figure 41 SimLab Main panel for Monte Carlo	59
Figure 42: External model execution schema	60
Figure 43: Model selection panel.....	61
Figure 44: External execution panel	62
Figure 45: Internal model selection	63
Figure 46: Internal model editor	64
Figure 47: Expression editor	65
Figure 48: Factor list editor.....	66
Figure 49: Factor modifier panel.....	66
Figure 50: Link factors panel	67
Figure 51 Statistical Post Processor panel	69
Figure 53 Post Processor File menu.....	70
Figure 54 Select variable for analysis panel	71
Figure 55 Expression area.....	72

TABLE OF FIGURES		
------------------	--	--

Figure 56 UA Main panel	76
Figure 57 Distribution Menu.....	76
Figure 58 Chart Menu	77
Figure 59 Visualise Menu	77
Figure 60: Variable selection panel	78
Figure 61: Post Processor panel with two variables	78
Figure 62: Setting Menu	79
Figure 63 UA setting panel	79
Figure 64: Single variable panel	81
Figure 65 Tabulated Value Panel.....	82
Figure 66 SA main Panel	90
Figure 67 Select Variables and factors Panel.....	91
Figure 69 Panel generated by clicking the header of output variable	93
Figure 70 Forecast panel	94
Figure 71 Plots panel.....	95
Figure 72 Cobwebs graph	96
Figure 73 Scatter plot panel	96
Figure 74 Sensitivity analysis for Morris.....	98
Figure 75 select Variables Panel	99
Figure 76 Tabulated values	99
Figure 77 Fast SA panel first orders effects.....	100
Figure 78 Fast panel tabulated values	101
Figure 79 Fast first and total order effects panel	101
Figure 80: Importance measures panel	102
Figure 81 select components Panel	103
Figure 82: Importance measures tabulated values panel	103
Figure 83 Sobol SA panel: Custom order calculation.....	104
Figure 84 Select variables Panel	105
Figure 85 Sobol SA panel: tabulated values tab	105
Figure 86 Default panel: Normal distribution.....	107
Figure 87 Distribution type selector.....	108
Figure 88 Quick Help example	108
Figure 90 Log-Normal distribution panel	110
Figure 91 Uniform distribution panel	111
Figure 92 LogUniform distribution panel	113
Figure 93 Weibull distribution panel	114
Figure 94 Constant distribution panel	115
Figure 95 Exponential distribution panel.....	116
Figure 96 Gamma distribution panel	117
Figure 97 Beta distribution panel.....	118
Figure 98 Triangular distribution panel	119
Figure 99 Relation panel	120
Figure 100 Uniform distribution panel	121
Figure 101 Interface between SimLab and GIS ArcView	123
Figure 102 UA and SA.....	126
Figure 103: Probability Density Function of Solow Model Conditional Forecast	130
Figure 104: Sensitivity analysis of Solow growth model	130
Figure 105: UA on Klein 3 model	135
Figure 106: SA on Klein 3 model	136

TABLE OF FIGURES		
------------------	--	--

Figure 107 Uncertainty Analysis143
 Figure 108 Sensitivity Analysis: Sobol’s indices144
 Figure 109 Interface between PREPSPOP and Calculator.exe.....145
 Figure 110 fac file example147
 Figure 111 Example of a .sam file.149
 Figure 112: External sample file format150
 Figure 113: no time dependent model file format.....153
 Figure 114 Time dependent output format.153