HYDROLOGICAL FREQUENCY ANALYSIS USING HYFRAN-PLUS SOFTWARE (VERSION-V2.1)

Salaheddine El Adlouni[1] **Bernard BOBÉE**[2]

[1] **Professeur, Département de Mathématiques et de statistique, Université de Moncton**

[2] **Professeur Émérite INRS-ETE**

Version of the 13th of January 2015

Note: Documents listed in the references (page 61) and marked by * are available when installing the DEMO version of the software

Citation: El Adlouni, S. and B. Bobée (2015). Hydrological Frequency Analysis Using HYFRAN-PLUS Software. User's Guide available with the software DEMO http://www.wrpllc.com/books/HyfranPlus/indexhyfranplus3.html

TABLE OF CONTENTS

List of Tables

List of Figures

1. Main Menu

The HYFRAN-PLUS software is designed for Hydrological Frequency Analysis (HFA) especially for extreme value. Thus for flood analysis, we consider the maximum annual flow. However HYFRAN-PLUS can be used for any dataset of extreme values in other areas with different time steps, provided that observations are Independent and Identically Distributed (IID Hypotheses; cf 2.1.4 and Bobée and El Adlouni, 2015).

1.1. Interface

When the software is started, the following window (figure 1) appears and the menu bar contains the following items:

a) File
b) Edition
c) Sample
d) Decision Support System (DSS)
e) Fitting
f) Graphic
g) Display
h) Window
i) Help (?)

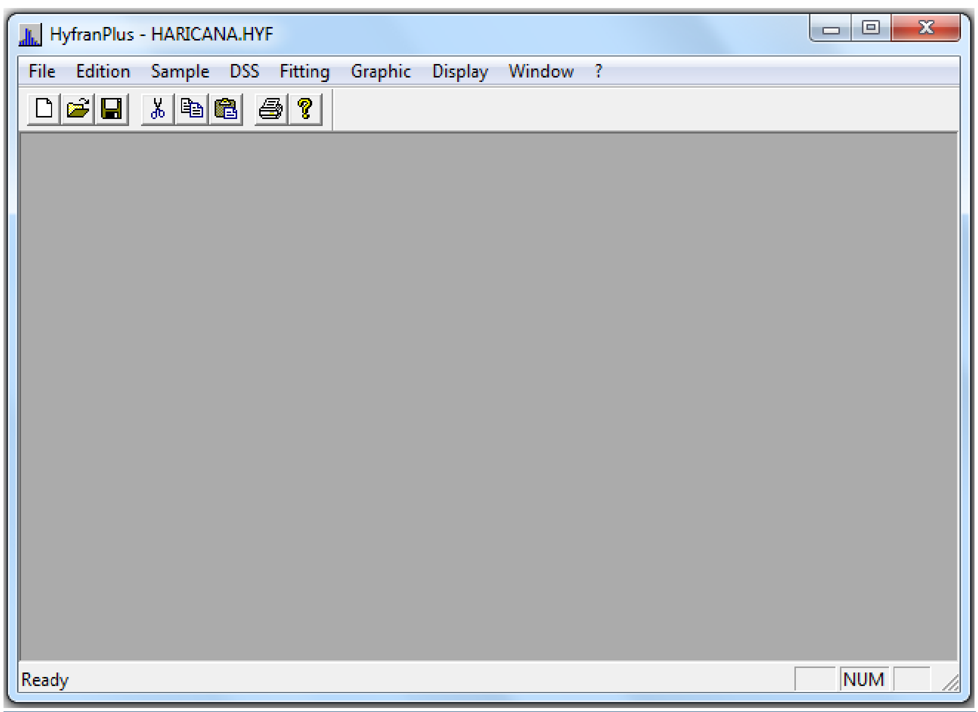


Figure 1: Opening Window of HYFRAN-PLUS

1.2. File

This menu (figure 2) contains the basic options for creation, opening and saving samples. It contains also the printing options.

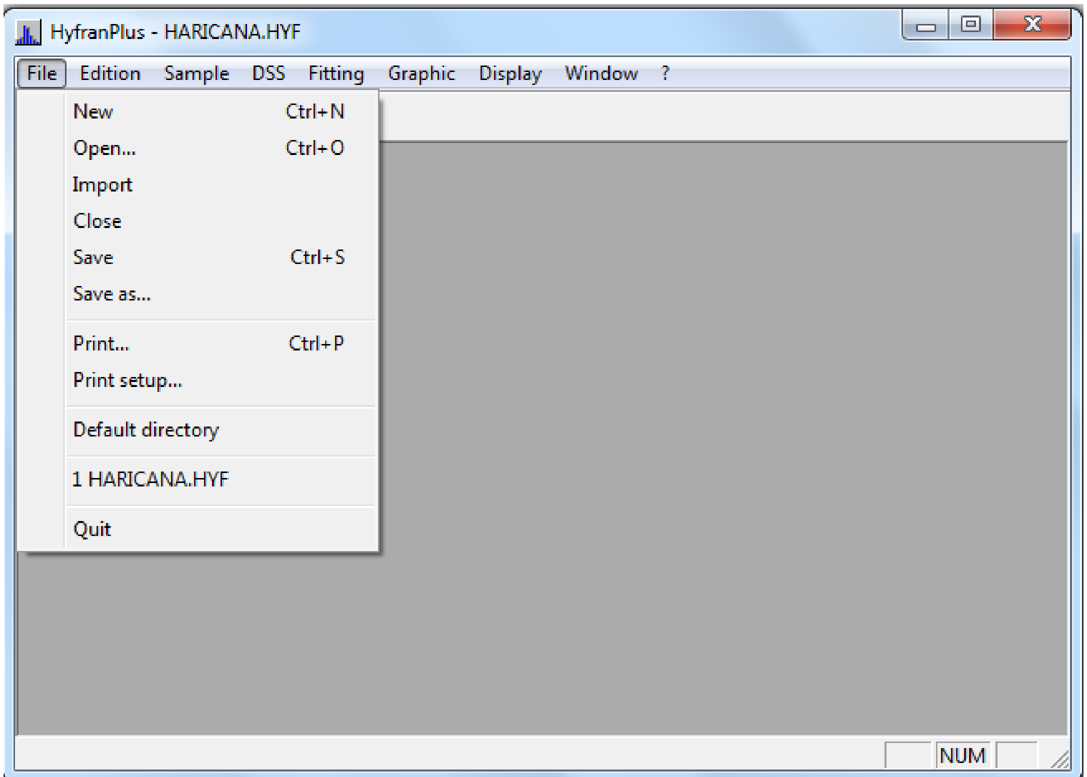


Figure 2: "File" menu

- **New:** This option allows creating a new project. It is possible to enter data manually or by using the clipboard.
- **Open:** When selecting this option, a dialog box appears allowing to select an existing project file. There are two choices: open a HYFRAN-PLUS file (*.hyf) or import different file types.
- **Importation:** This option allows importation of files from different formats. To use this option, select the file in his directory and click on the command button "open". If the file format is recognized, cursor will be positioned on the good file format in the new screen displayed. The HYFRAN- PLUS format file conversion will be then treated when the user enter a name for the new file.
- **Save:** This option allows to save modifications realized in the project file.
- **Save as:** The option "save as" allows to change the name or the access path of an existing project file.
- **Print:** The complete information of the whole windows can be printed, no matter the type of content (text or graphic).

- **Printer configuration:** This option allows to configure printer (choice of printer, paper, and orientation).
- **Directory by default:** This option allows to choose a directory by default for all data files.
- **Haricana :** The window corresponding to the opened project.
- **Exit:** By choosing this option, you quit HYFRAN-PLUS. If modifications are done on the current sample and they haven't been saved, a warning message will be appear.

1.3. Edition

This menu (Figure 3) contains the basic options for editing that is to say cancel, cut, copy, and paste. It is also possible to export graphics in another application by selecting "copy".

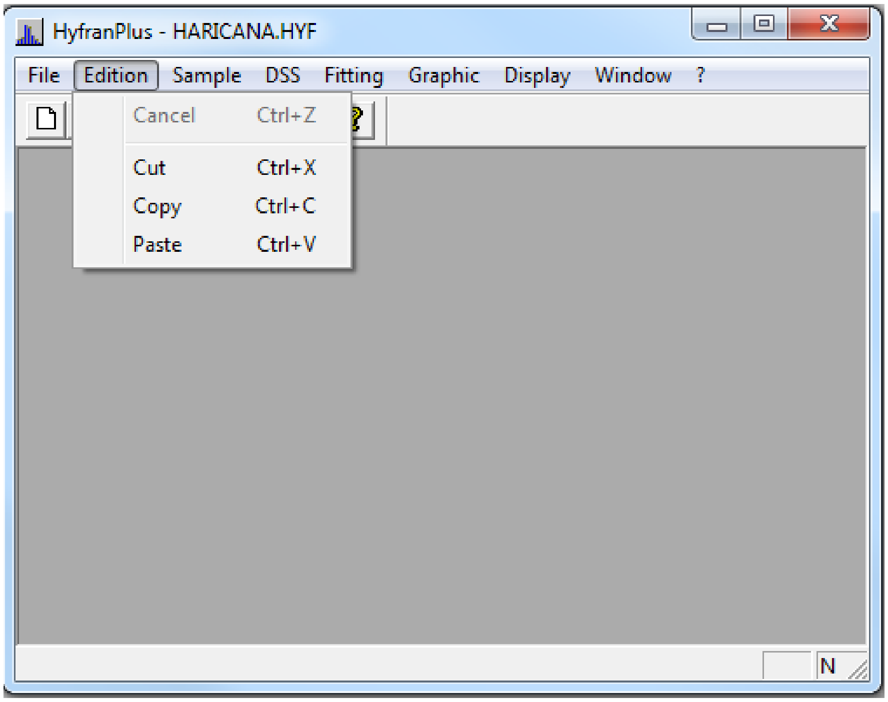


Figure 3: "Edition" menu

1.4. Sample

This menu (Figure 4) contains the options that are necessary to download all edit data.

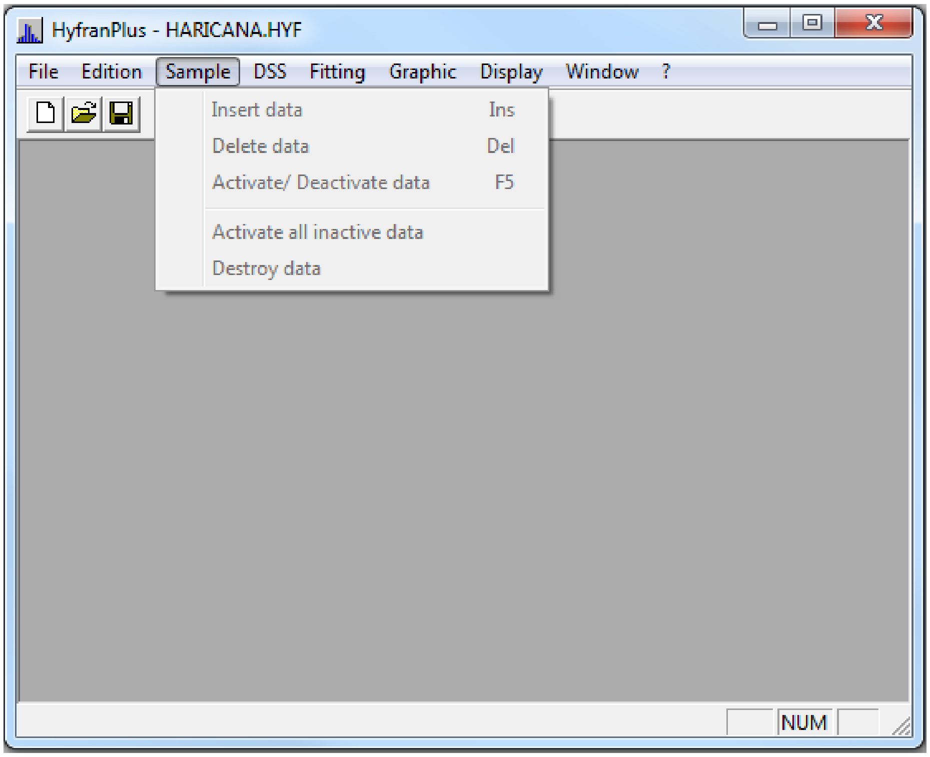


Figure 4: “Sample” menu

- **Insert data:** this option is selected to add a data.
- **Delete data:** this option allows deleting one or several data.
- **Activate/Deactivate data:** this option allows deactivating some active observations or activating some inactivate observations.
- **Activate all inactive data:** in order to activate all the inactive data
- **Destroy all inactive data:** In order to delete all the inactive data.

1.5. Decision Support System (DSS)

The DSS menu (Figure 5) is the main difference with the previous version of HYFRAN-PLUS (Version 2.0). In the HYFRAN-PLUS software (Version 2.1) the DSS allows the choice of the most appropriate class of distributions. The main elements of the DSS are presented in detail in El Adlouni et Bobée (2011) and in El Adlouni, Bobée and Samoud (2012).

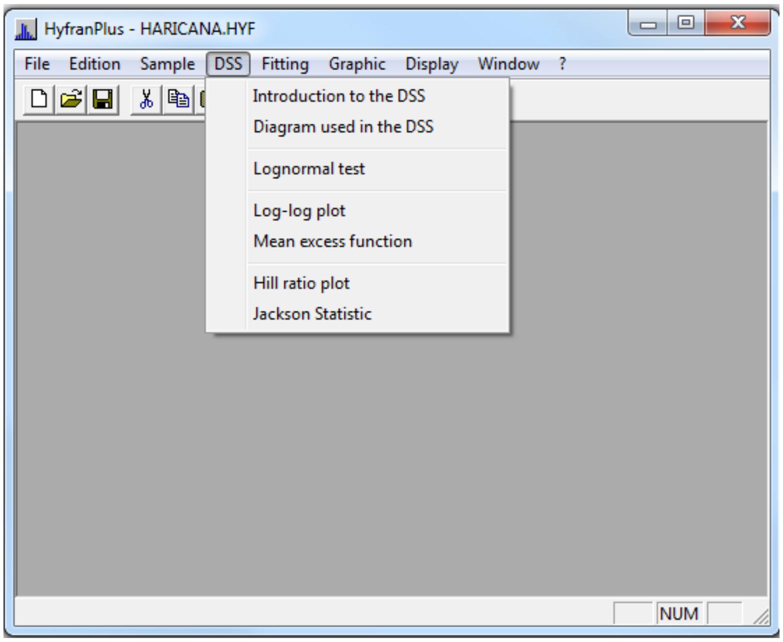


Figure 5: “DSS” menu

- **Introduction to DSS:** a brief presentation of DSS is available as well as a reference to papers related to the DSS.
- **Diagram DSS:** for a better understanding of the methodology of the DSS (Figure 5) a graphical illustration of the steps is provided in the interface of HYFRAN-PLUS (Figure 6). It represents the different steps of DSS.

DSS menu (Figure 5) also offers graphics and curves on which the decision support system is based, such as log-normal graph, the log-log graph, the mean excess function (MEF), the Hill ratio plot, the Statistics of Jackson (cf. El Adlouni and Bobée, 2011; El Adlouni, Bobée and Samoud, 2012).

- **Log-normal test** [(1)]**:** To test the log-normality using the Jarque-Berra test.
- **Log-log plot** [(1)] **:** To check the belonging to the class C (distributions to regular variations)
- **Mean Excess Function (MEF)** [(1)]**:** To check the belonging to the class D (sub-exponential distributions).
- **Hill ratio plot and Jackson’s Statistic** [(1)]**:** For a confirmatory analysis of the selected class.

Note: (1) These items will be developed and illustrated by examples in Section 2.2.

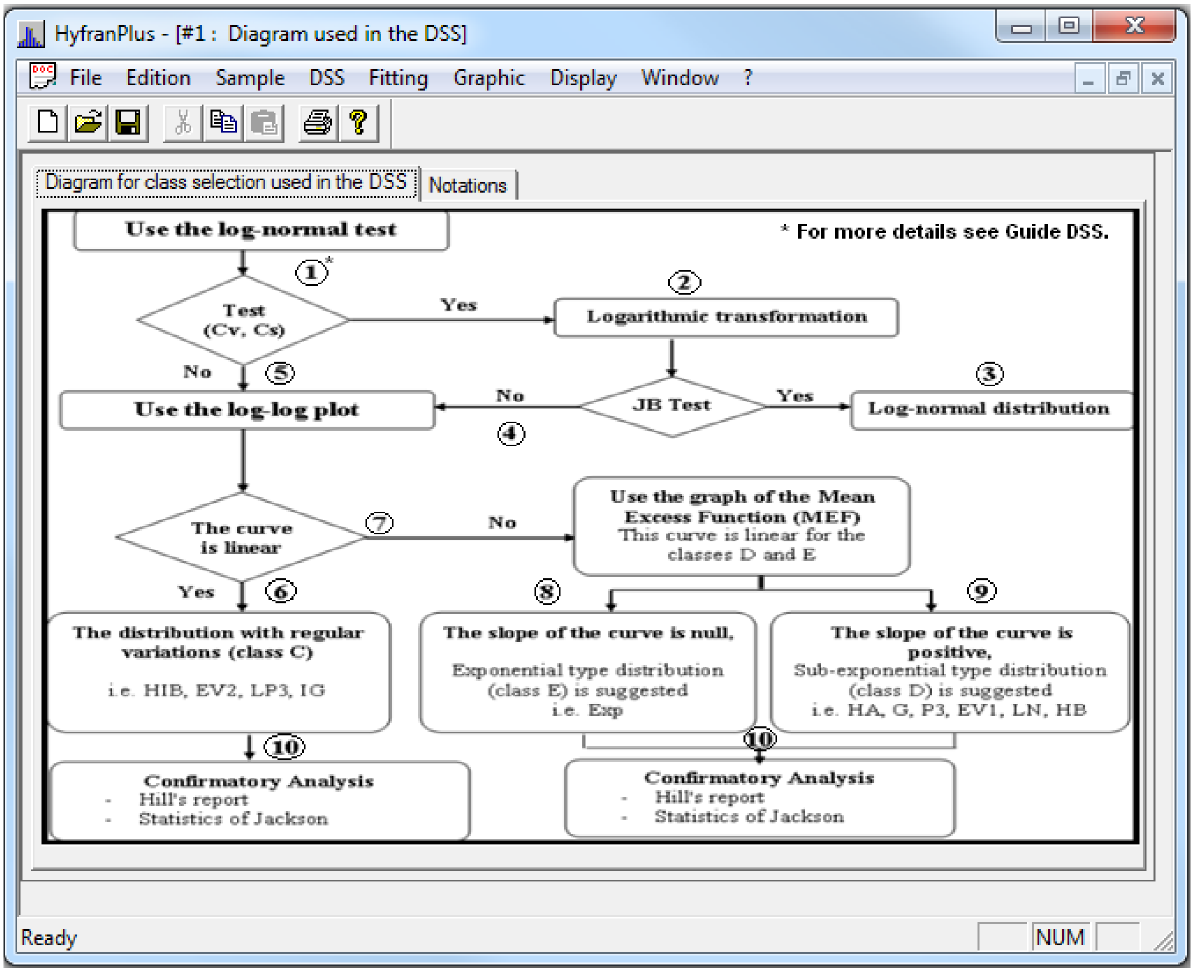


Figure 6: Diagram of the DSS

1.6. Fitting

HYFRAN-PLUS allows to fit (Figure 7) different statistical distributions (Compaore, El Adlouni et Bobee, 2014) to a random sample that satisfies IID [**I**ndependent and **I**dentically **D**istributed data, cf. Section 2.1.4] conditions with several estimation methods (cf. Bobée and El Adlouni, 2015; Bobee and Askar 1991).

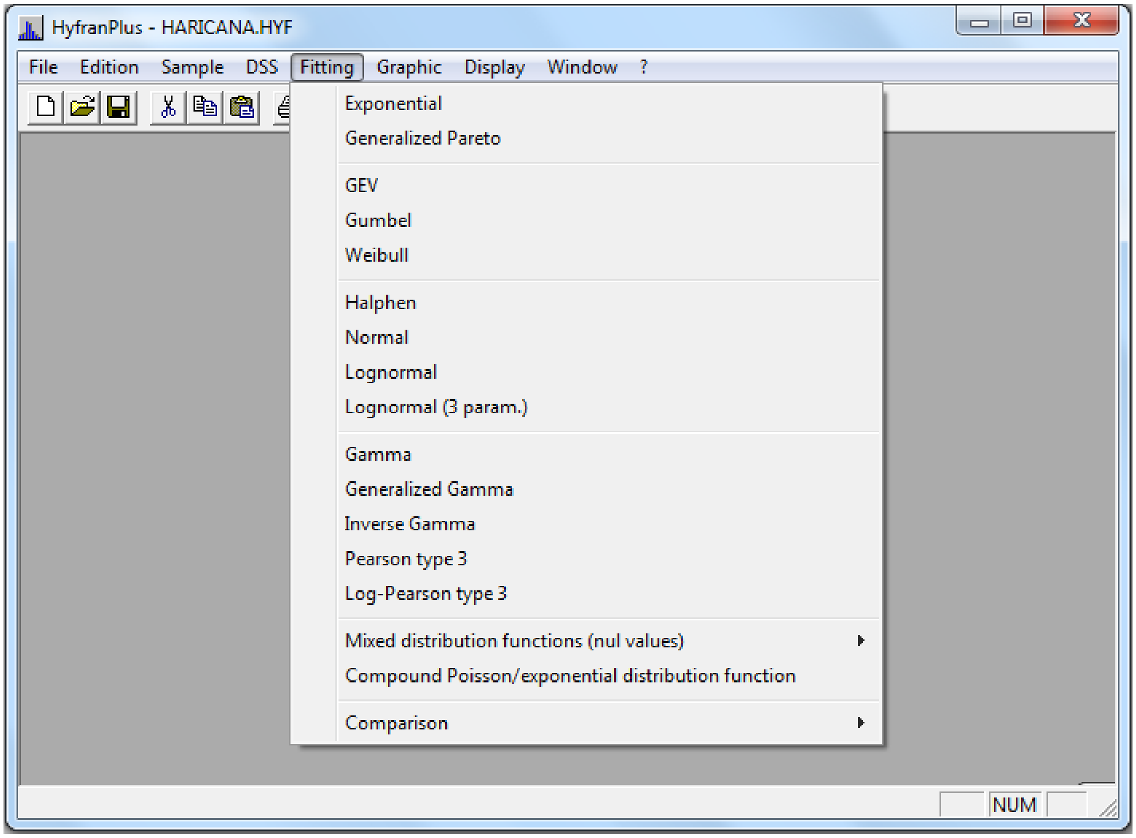


Figure 7: “Fitting” menu

The "comparison" option allows to compare several fittings to choose which is the most adequate to represent the studied dataset. We can compare the fittings using criteria or graphics.

- **Graphic:** it is possible to compare the results of several different fits (2 to 5) using either Normal or Gumbel probability paper.

- **Criteria:** Two criteria are available, these are the Akaike (AIC) and Bayesian information criteria (BIC), (see Ehsanzadeh, El Adlouni and Bobée, 2010).

Note:

- Halphen tab contains the three Halphen distributions (type A, type B and type Inverse B) with their limiting cases (Gamma and Inverse Gamma) (Morlat, 1956). However, in the fit the choice of one of three distributions is done automatically based on the characteristics of the sample and the theoretical properties of the Halphen family (Perreault, Bobée and Rasmussen, 1999).
- The GEV tab includes the three distributions Fréchet, Gumbel and Weibull. The choice is done automatically based on the estimate of the shape parameter. However, the user can decide to use the Gumbel or Weibull distributions (Figure 7).

1.7. Graphic

This menu (Figure 8) contains options that allow adding or deleting components on the graphics. It concerns the symbols, the curve, the confidence intervals and the legend.

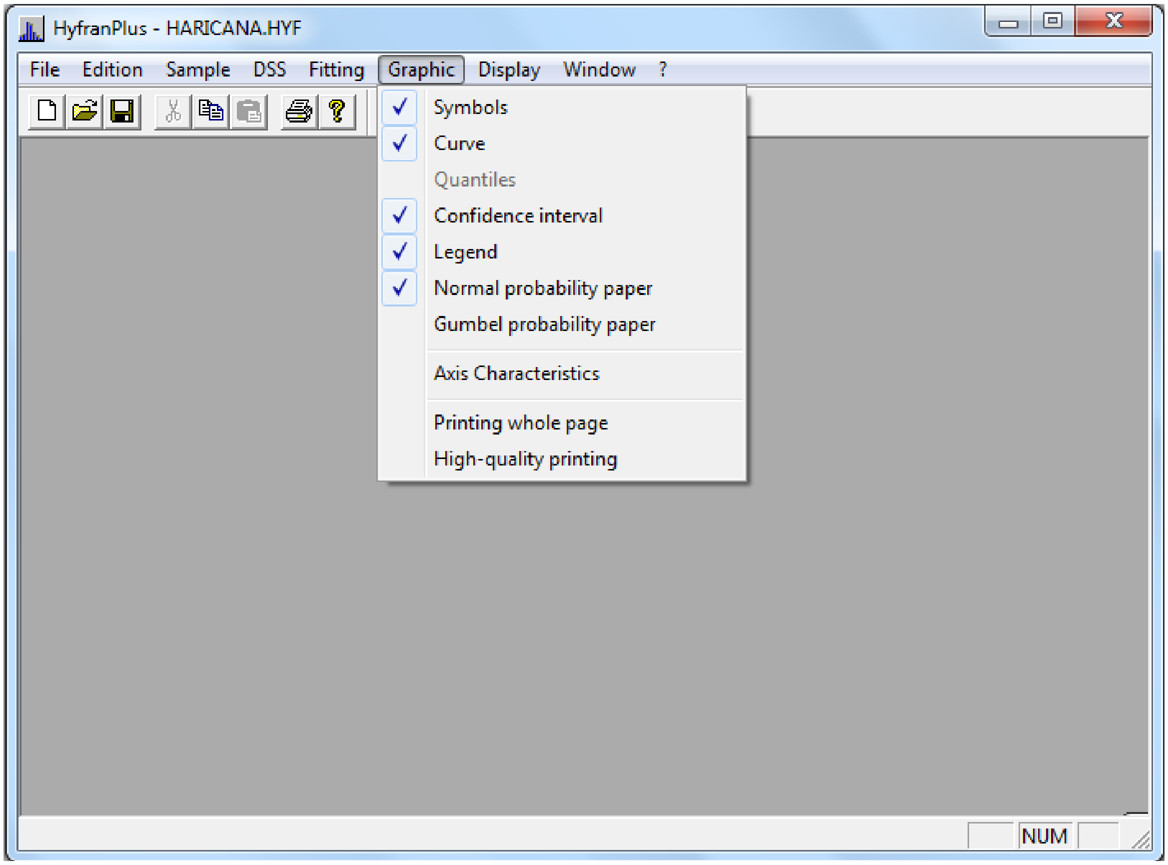


Figure 8: “Graphic” menu

In addition, it is possible to select here the type of probability paper on which the data and the fitting curves can be displayed. Two types of probability papers are available: Normal and Gumbel. A normal (Gumbel respectively) distribution would be represented as a linear curve on the normal (resp. Gumbel) paper.

The option “axis characteristics” allows to zoom on a more precise area of the graphic. The user can also decide to add or remove the legend.

The option “printing whole page” allows to print a graphic in full page. Select this option (a mark is inserted below this menu option), and then in the menu “File”, select “Print”. By default, printing will be done in landscape format, this option is suggested when you experiment printing problems with certain printers’ models.

The option “high-quality” printing allows to print a graphic with a higher precision.

Note: This option works with most current printers but not with all.

1.8. Display

This menu (figure 9) allows or no to display the tools and state bars.

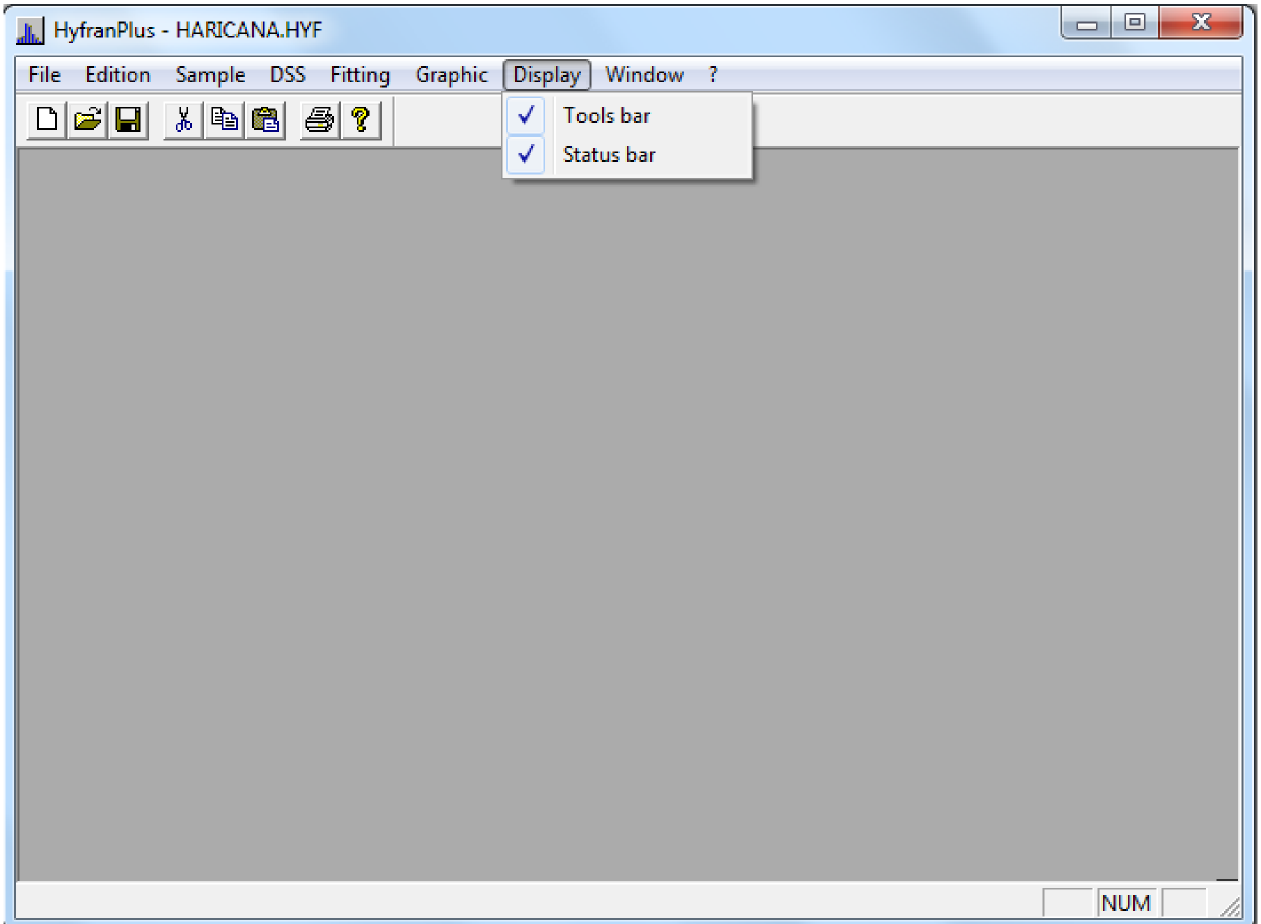


Figure 9: “Display” menu

1.9. Window

This menu (figure 10) allows to spot the different opened windows and to browse between them.

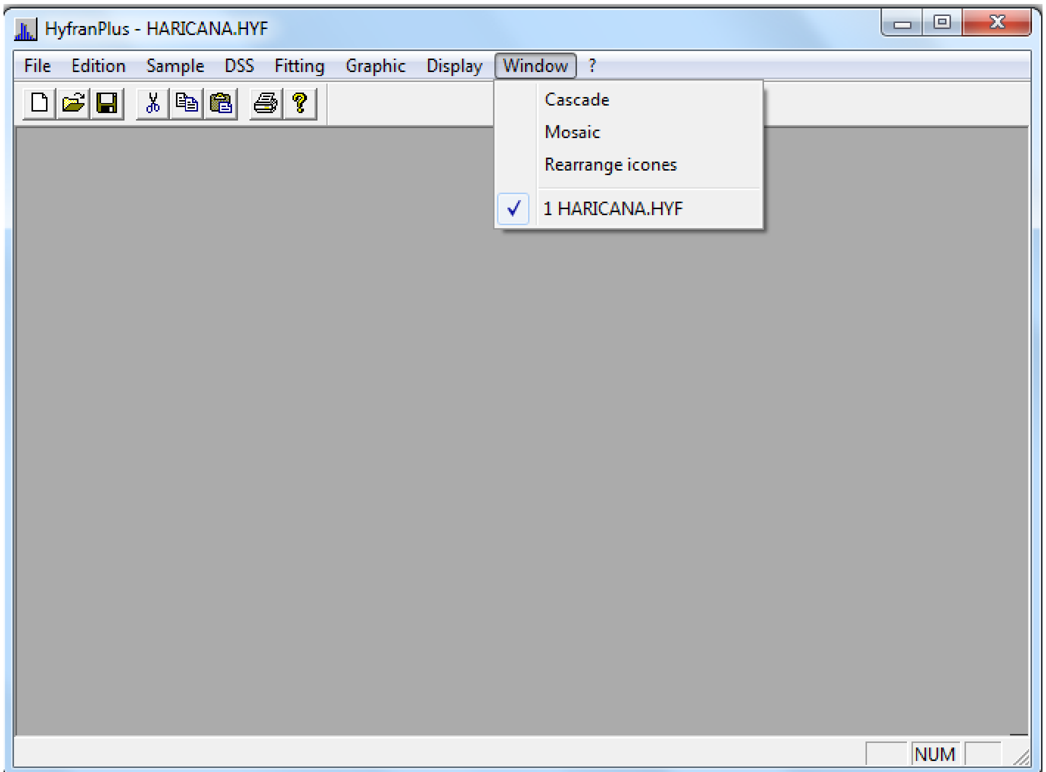


Figure 10: “Window» menu

2. Tutorial

HYFRAN-PLUS was developed in the Windows environment. This software is designed to make easiest the fitting of a statistical distribution to a random sample (IID, cf. section 2.1.4). The fitting steps can be grouped into two categories:

- Data editing and descriptive statistics to study the characteristics of a random sample (section 2.1);
- Fitting Procedures (sections 2.2 and 2.3).

HYFRAN-PLUS comes with a default project called Haricana which will be used in some parts of the tutorial.

For each function available in HYFRAN-PLUS, a corresponding tab box is provided with appropriate options. We present in what follows the different steps of a frequency analysis using statistical and graphical tools presented in HYFRAN-PLUS.

2.1. Data entry and Study of the statistical characteristics of a random sample

Creating a new project (Figure 11) or opening an existing project from the "File" menu a tab box appears. The tab box allows editing or modifying data, to evaluate some statistical characteristics of the random sample associated with the project to perform some statistical tests, as well as to produce several graphs.

To perform these tasks the user should browse between five different tabs:

1) Description
2) Data
3) Basic Statistics
4) Hypothesis Tests
5) Graphics

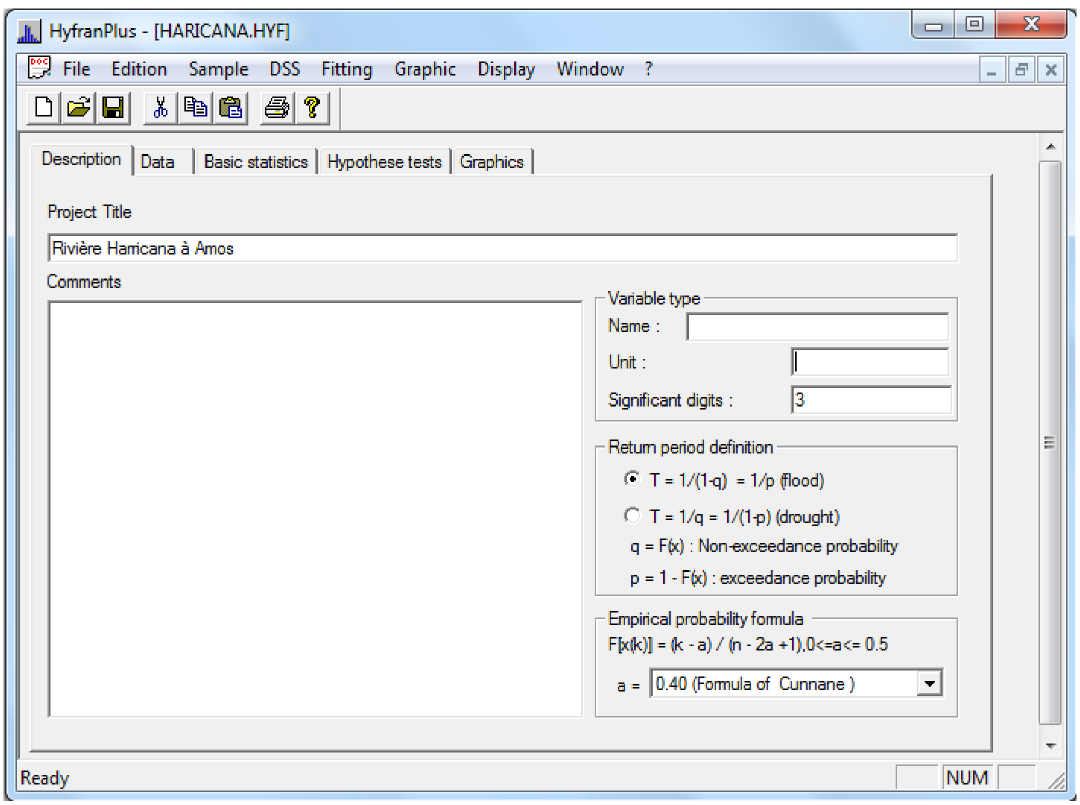


Figure 11: Creating a new project

2.1.1. Description

In the "Project Description" tab (Figure 12), we find information related to the project that is used to make graphs and tables to present the data and results:

a) Enter firstly the title of the project that will be the title of the graphics produced by HYFRAN-PLUS.
b) Then enter the name of the variable that will appear as the title of one of graph axes (abscissa or ordinate, according to the graphic)
c) Then enter the unit of measurement of the observations that will be specified after the variable name on the graphic.

d) Also we can specify the number of significant digits of data (between 1 and 18). This value is used to present the results with the correct number of significant digits but does not influence the accuracy of the calculations.
e) We can choose a definition for the concept of return period. It can be either:
   - the inverse of the probability of exceedance for flood data;
   - the inverse of the probability of non exceedance, for low flow data, (Bobée and El Adlouni, 2015).
f) We have to choose a Plotting Position (PP) (empirical probability formula), used to draw the observations on probability paper. In HYFRAN-PLUS, the PP formulas are used as follows: $P_k = (k - a)/(n - 2a + 1)$ (see Bobée and El Adlouni, 2015; Bobée and Ashklar, 1991). In the flood data case, P_k is the probability of non-exceedance of the observation X_k of order *k* ranked in ascending order in the sample of size n.

   The Cunnane formula (a = 0.4) is used by default in the HYFRAN-PLUS software but the other available formulas can be used depending on the user's choice (Bobée and Ashkar, 1991 - Table 1.3 on page 11).

g) We can finally enter any comments about the project. The information entered in the "Comments" section are not taken into account by HYFRAN-PLUS, however it can be useful for the user to describe the project in more detail (Figure 12).

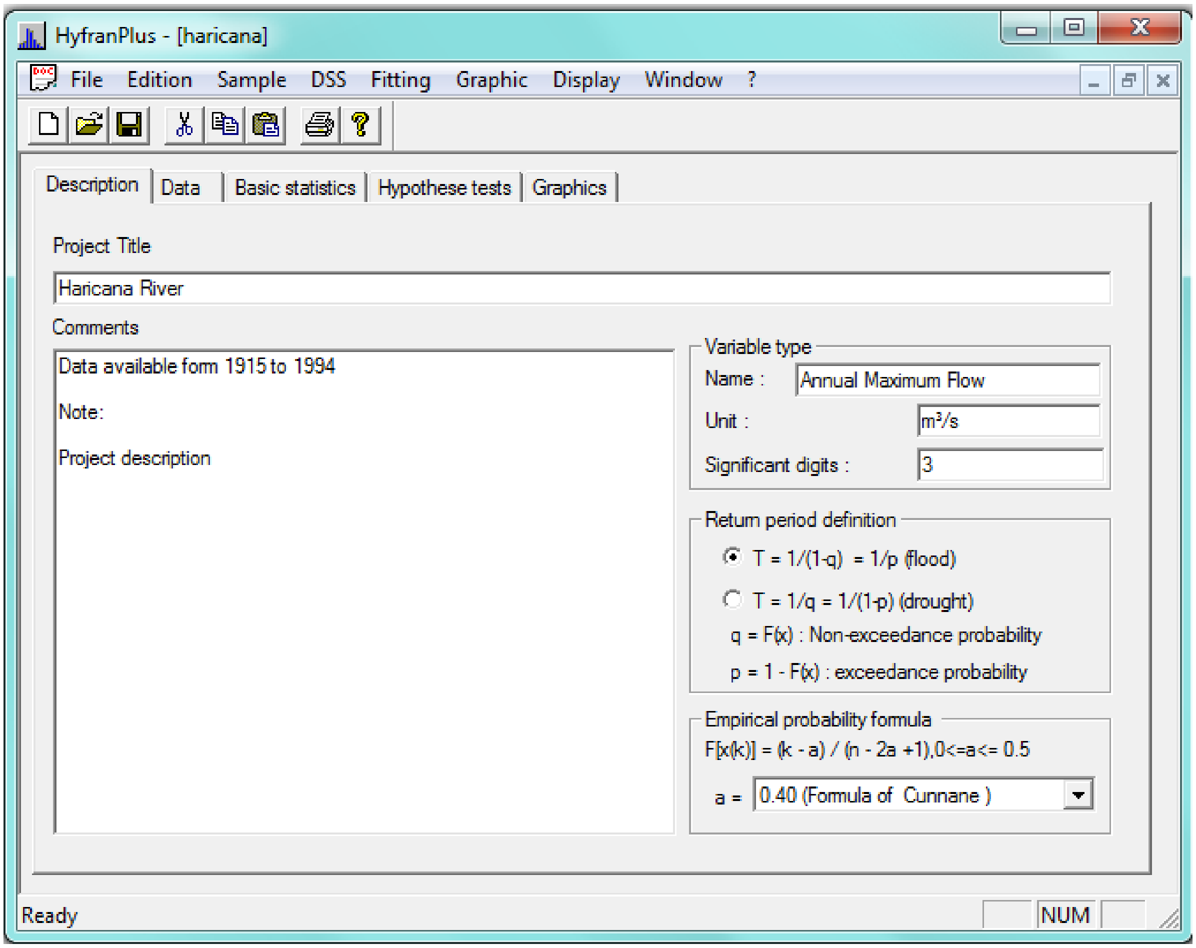


Figure 12: Example of description Haricana project

2.1.2. Data

There are three ways to enter data into the HYFRAN -PLUS software:

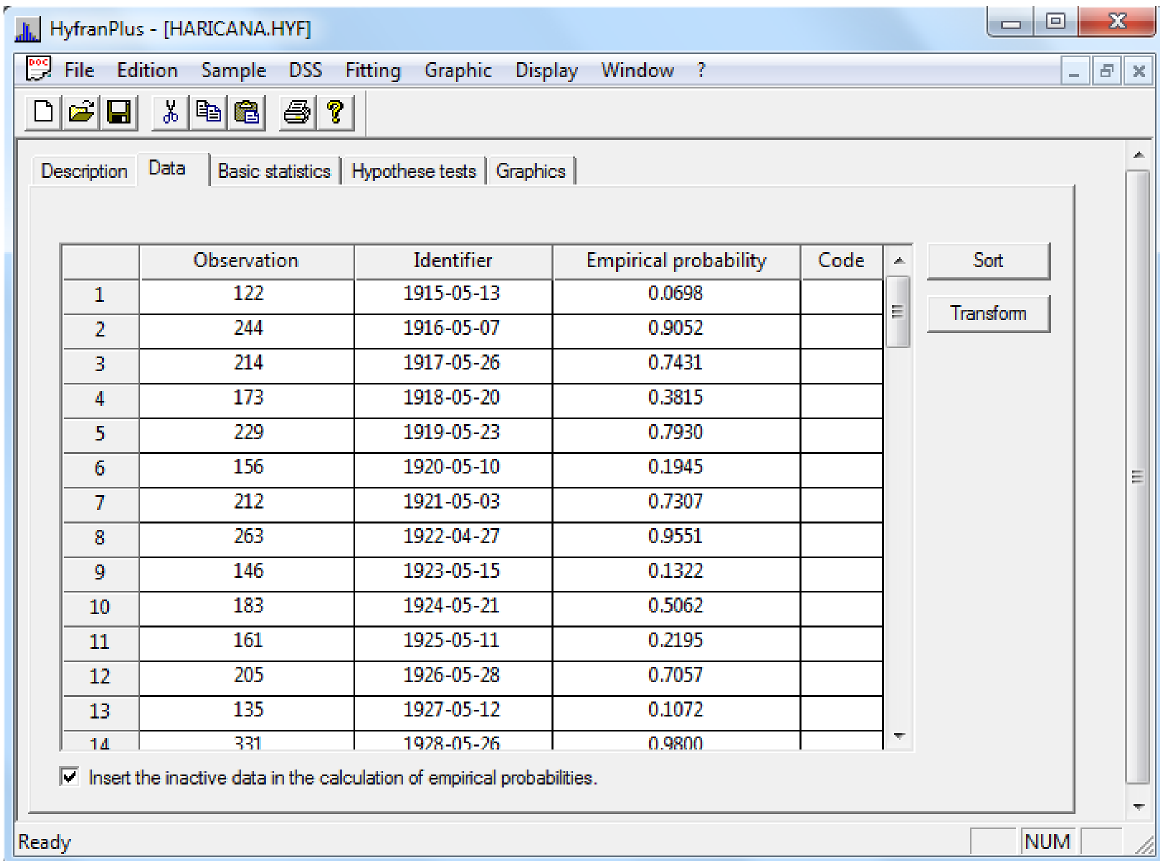
a) Use the built-in spreadsheet software (see Figure 13).
b) Import a data file (for formats that HYFRAN –PLUS knows; described in the following section).
c) Using the clipboard

a) Spreadsheet integrated software

HYFRAN -PLUS comes with a spreadsheet that allows you to enter and view data (Figure 13). Each line corresponds to an observation and is divided into four columns:

1. Observation: the numerical value of each observation must be entered in this column;

2. Identifier: you can assign a sequential number to each observation, if one wants to put a date, it should generally be entered in the format YYYY -MM -DD (cf. section 2.1.4-c), however the user can leave out the day and month and enter only the year (YYYY);
3. Plotting Position (PP): The PP (Empirical probability) associated with each observation is automatically calculated and displayed in this column, using the formula specified in the description page of the project (Figure 11);
4. Code: This column can be used to add only a single character in order to codify some observations; for example, add "M" for manually measured data and "R" for revised data.



	Observation	Identifier	Empirical probability	Code
1	122	1915-05-13	0.0698	
2	244	1916-05-07	0.9052	
3	214	1917-05-26	0.7431	
4	173	1918-05-20	0.3815	
5	229	1919-05-23	0.7930	
6	156	1920-05-10	0.1945	
7	212	1921-05-03	0.7307	
8	263	1922-04-27	0.9551	
9	146	1923-05-15	0.1322	
10	183	1924-05-21	0.5062	
11	161	1925-05-11	0.2195	
12	205	1926-05-28	0.7057	
13	135	1927-05-12	0.1072	
14	331	1928-05-26	0.9800	

Figure 13: Spreadsheet integrated software (Haricana project)

To insert new data (see Figure 4), use the "Insert" key on the keyboard or the equivalent option in the "Sample" menu. To delete one or more data, selects one or more data, then use the "Del" key on the keyboard or the equivalent menu option "Sample"(see Figure 5). It is also possible to copy all data from the spreadsheet HYFRAN-PLUS in the clipboard by pressing **Ctrl-A** ("Select All" from the "Edit" menu), then **Ctrl-C** ("Copy keys "in the" Edit "menu) and finally **Ctrl-V** (" Paste "from the" Edit "menu) [Figure 3]. In addition some options described in the following (deactivation, sorting and transformation of data) can be used from the spreadsheet.

- *Deactivated data*

One advantage of HYFRAN-PLUS is to allow deactivation of the data without destroying them (Figure 14). This allows among others to evaluate the sensitivity of statistical analysis with singular data such as possible outlier. Thus, extreme data can be represented on the graph of the fitted distribution but not be used in the fit of this distribution

To deactivate an active data or vice versa, you can press "F5" or choose the option of "Sample" menu (Figure 4).

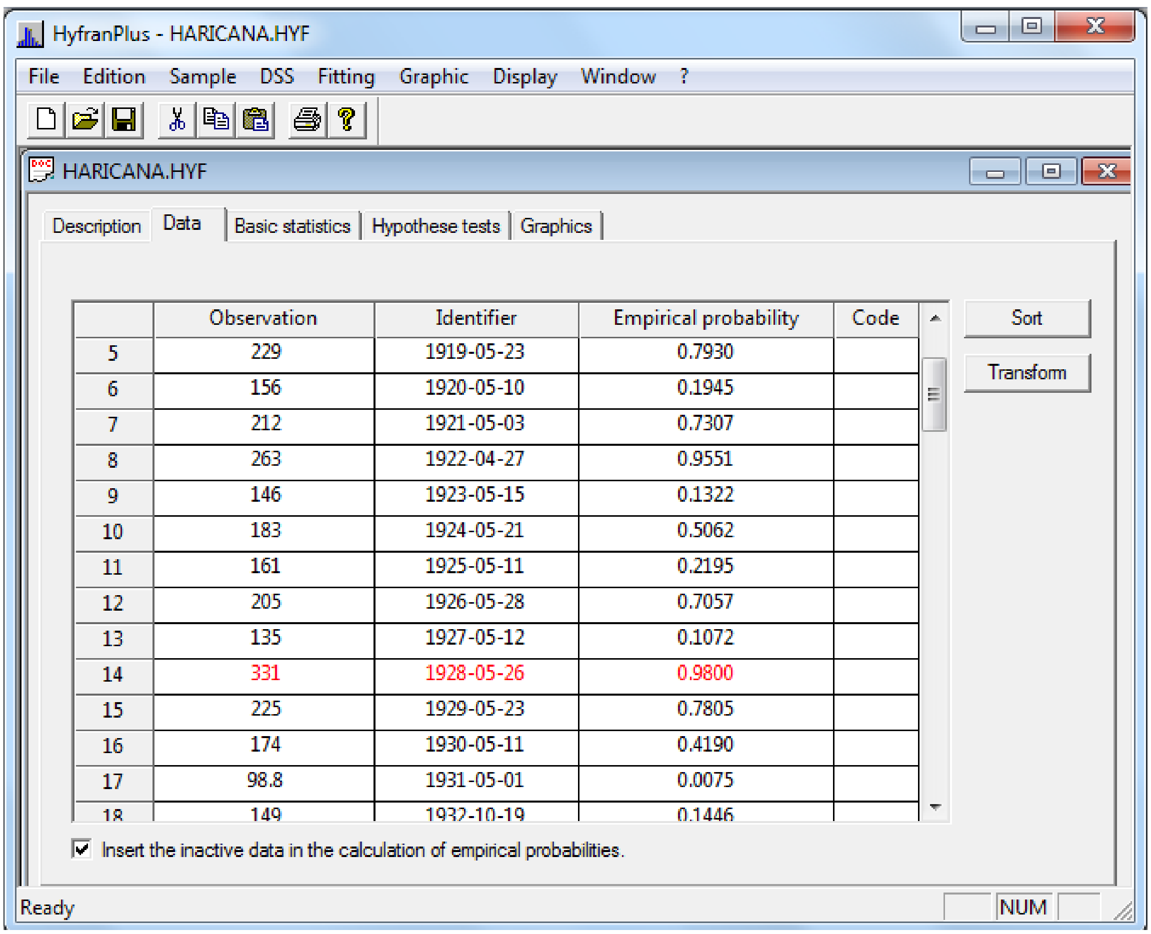


Figure 14: Example of deactivated data (line 14 of the Haricana project)

The deactivated data are not taken into account to perform various statistical tests of assumptions and fittings proposed by HYFRAN-PLUS. However in the graphics they are displayed with a different symbol (Figure 15).

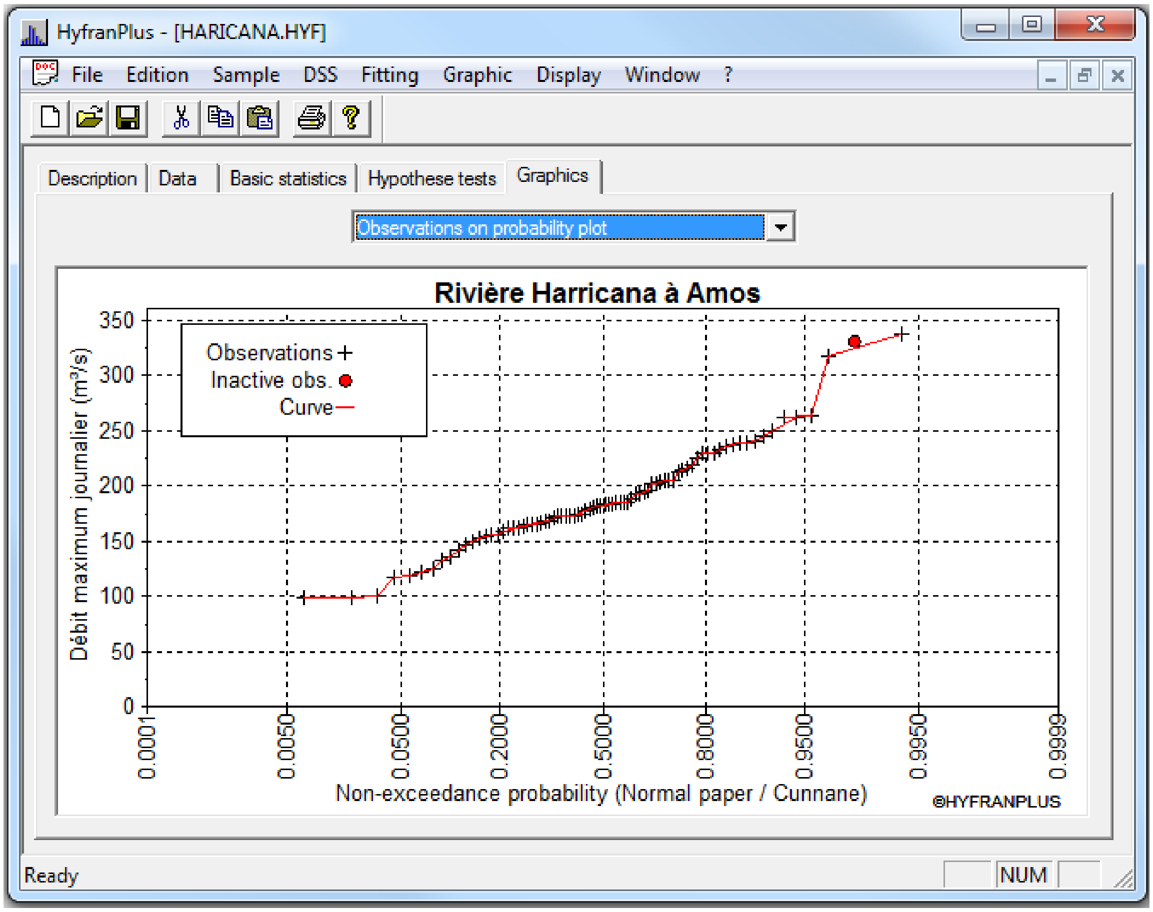


Figure 15: Graphic with an example of deactived data (Haricana project)

To calculate the PP, it is possible to include or exclude deactivated data (Figure 16). If the option "Activate all inactive data” is selected, all data (active or inactive) are treated in the same way to compute the PP. The empirical probability of deactivated data is always calculated in the same way, based on the full sample. Otherwise, the empirical probability of active observations is calculated without taking into account the existence of deactivated data.

To activate all inactive data 'Active all inactive data "menu" Sample " (Figure 4) is used. The destruction of inactive data is done using the option "Destroy all non-active data" menu "Sample".

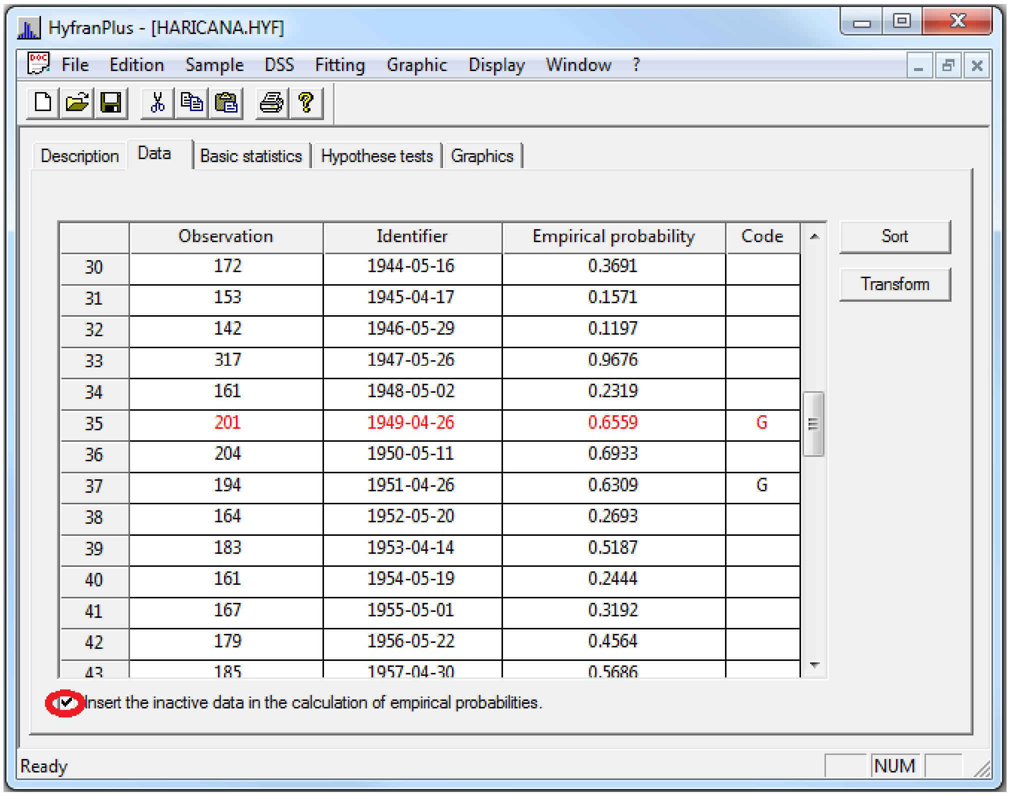


Figure 16: Possibility to include or exclude data

- *Sorting Data*

It is possible to sort the data in ascending or descending order, and based on either:

- the value of the observations (column 1)
- the identifier (column 2).

To sort the data, click on "Sort" button (Figure 16).

- *Transforming Data*

HYFRAN-PLUS allows to transform data using several simple functions. To transform data, you must press the "Transform" button (Figure 16). We opens a dialog box that provides the following transformations (Figure 17):

- Inverse: $1/x$
- Opposed: $-x$
- Absolute value: $|x|$
- Exponential: a^x
- Logarithmic: $\log_a(x)$
- Scale: $a \times x$
- Position: $a + x$
- Power: x^a

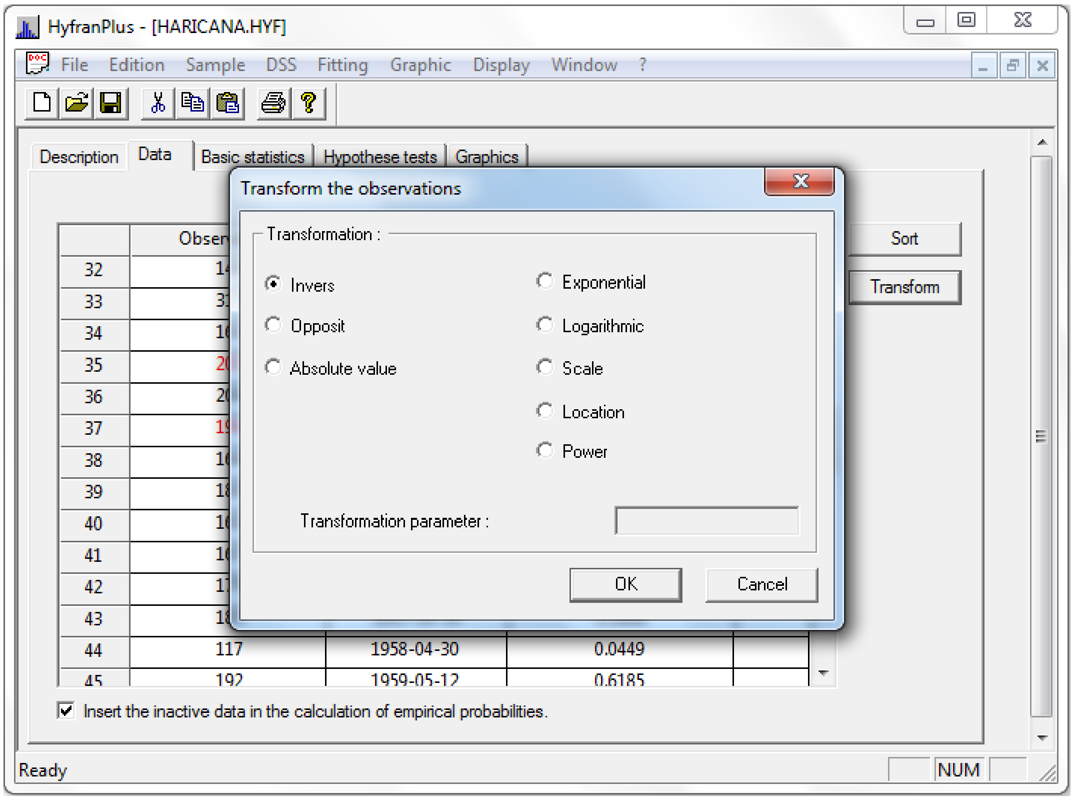


Figure 17: Transforming data

Note: the real number a corresponds to the value of the transformation parameter that can be specified in the same dialog box. In the case of a logarithmic transformation, a corresponds to the selected base. In practice, we consider: $a = 10$ (decimal logarithm) or $a = e$ (logarithm Napierian).

b) *Import of data files*

HYFRAN -PLUS allows importing various types of data format. To import data use the "File" menu (Figure 2) and click on the "Import" tab. Different importable files are listed below:

- HYDAT Formats
  - Export Extreme
  - Export Extreme Instantaneous
  - Export Mean
  - Extreme Print
  - Print Mean
- MATLAB Format: In this format, only the numeric data are accepted, the dates are missing. Data are separated from each other by one or more spaces. Each line in the file corresponds to a station.
- Text Format 1 column: Digital data is present in a single column, in this format, the dates are not entered. The user gets all the data in one file.

- Format HCDN ASCII Annual Mean
- Excel format with the same order of columns as that of HYFRAN-PLUS (First column: Observations and Second column is the identifier that is the date).
- Free Format: to import this type of data in the HYFRAN -PLUS software (may or not include dates) the following model must be respected:
  - 1st line: one enters a title on one line
  - 2nd line: then it writes the words “Free Form" without quotes but with a space between format and Free
  - Following lines: they are devoted to the data (one data per line). The data must not contain more than 19 digits including the decimal point. They contain a space, a date or a numeric identifier (optional). The date format is yyyy / mm / dd (year, month and day). The file must be saved in plain text format.

c) *Clipboard*

The clipboard can be used to paste data from another Windows application (word processor, spreadsheet, etc.). Just use the "Ctrl -C" key ("Copy" from the "Edit" menu) and "Ctrl -V" ("Paste" from the "Edit" menu, Figure 3) to insert the contents of the clipboard into the integrated spreadsheet to HYFRAN -PLUS.

2.1.3. Basic Statistics

The window of basic statistics (Figure 18) in addition to title of the study displayed the following statistics of the sample (see Bobée and El Adlouni, 2015).

- The sample size
- The minimum value
- The maximum value
- The mean
- The standard deviation
- The median
- The coefficient of variation
- The coefficient of skewness
- The coefficient of kurtosis

When there are inactive data, basic statistics are presented in two columns. The first column shows the statistics of the sample of active data only and the second column statistics for the full sample.

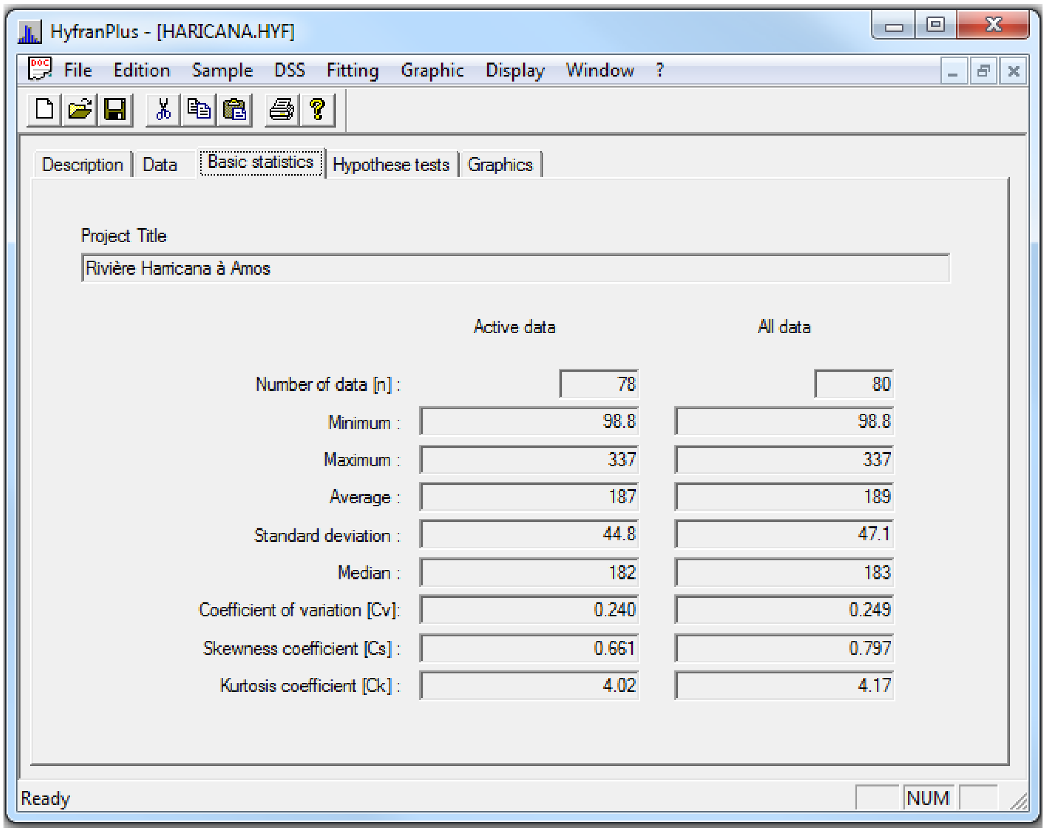


Figure 18: Basic statistics: Haricana dataset

2.1.4. Hypothesis test

Before fitting a sample using a statistical distribution, it is important to check if the data are Independent and Identically Distributed (IID) (Bobée and El Adlouni, 2015; Bobée and Ashkar, 1991). Indeed, the observations must be **I**ndependent realizations of a random variable from the same statistical distribution (i.e. **I**dentically **D**istributed). In HYFRAN-PLUS statistical tests are available to check the assumptions of independence, stationary and homogeneity. These are four hypothesis tests:

a) Independence test (Wald- Wolfowitz) ;
b) Stationary Test (Kendall);
c) Homogeneity at annual scale test (Wilcoxon);
d) Homogeneity at seasonal scale (Wilcoxon);

Note: The test of Wilcoson is also known under the name of Mann-Whitney test. All of these tests are described in Bobée and Ashkar (1991) and Bobée and El Adlouni (2015).

In the “Hypothesis Tests" tab the title of the project described (Figure 11) is shown and you can:

- first select the statistical test to perform;
- specify the null hypothesis (H0) and the alternative hypothesis (H1).

For each test we get:

- The value of the test statistic and the corresponding p-value (i.e. probability of exceedance of the statistic);
- The conclusion of the test (obtained from the p-value), i.e. the acceptance or rejection of the null hypothesis at a significance level of 5% or 1%

a) *Test of Independence (Wald-Wolfowitz test)*

The Wald-Wolfowitz test (Figure 19) allows to check if there is a significant first order autocorrelation between observations.

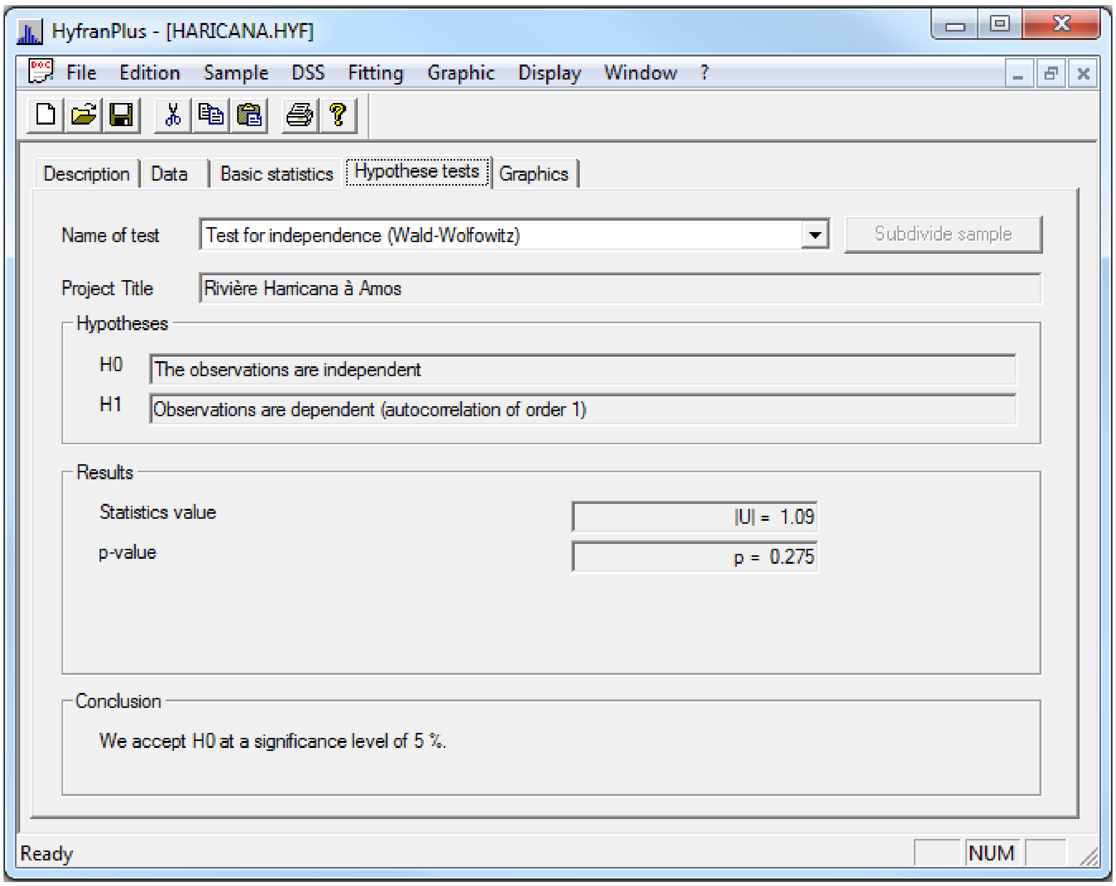


Figure 19: Independence test: Haricana project

b) *Stationary Test (Test of Kendall)*

The stationary test of Kendall (Figure 20) allows to check if there is a significant trend in the series.

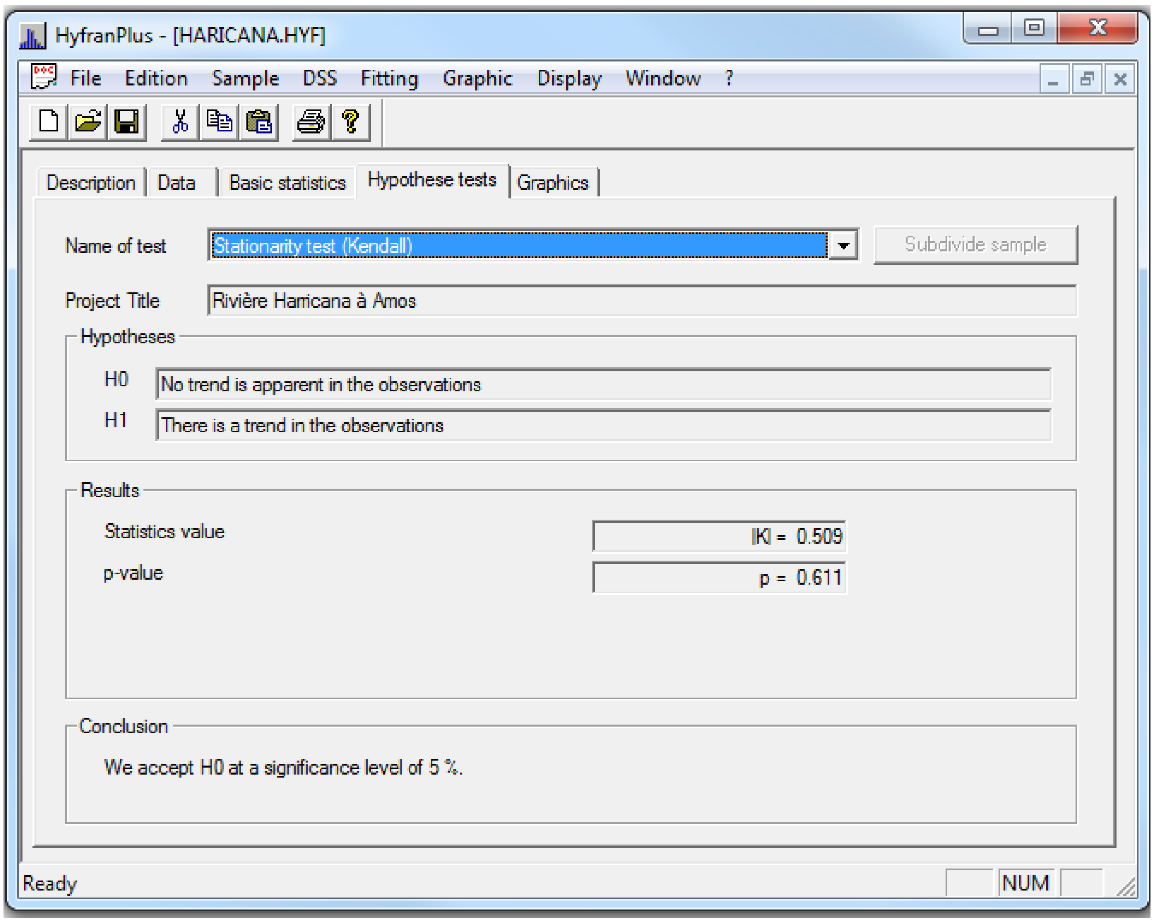


Figure 20: Stationary test: Haricana project

Note: When we chose the test of homogeneity (Figure 21) the button "Subdivide the sample" appear for the two homogeneity tests considered in the following. It is necessary in such cases to specify the two sub-samples for which the averages will be compared.

c) *Test of homogeneity at annual scale (Wilcoxon or Mann-Whitney test)*

The homogeneity test, on an annual scale (Figure 21), allows to check if the mean of the first sub-sample is significantly different from that of a second subsample. This test can be used, for example, in the case of shifting a flow recording station if one want to compare the mean of the data before and after the date of moving i.e. check whether the data belong to the same statistical population. The first sub-sample consists of observations from the earlier of the record to a cut-off year. The second sub-sample consists of observations of the year following the cut-off year. The cut-off year is specified by pressing the button "Subdivide the sample."

Note: To perform this test, it is necessary to specify the year for each observation when editing or importing the data (Figure 13).

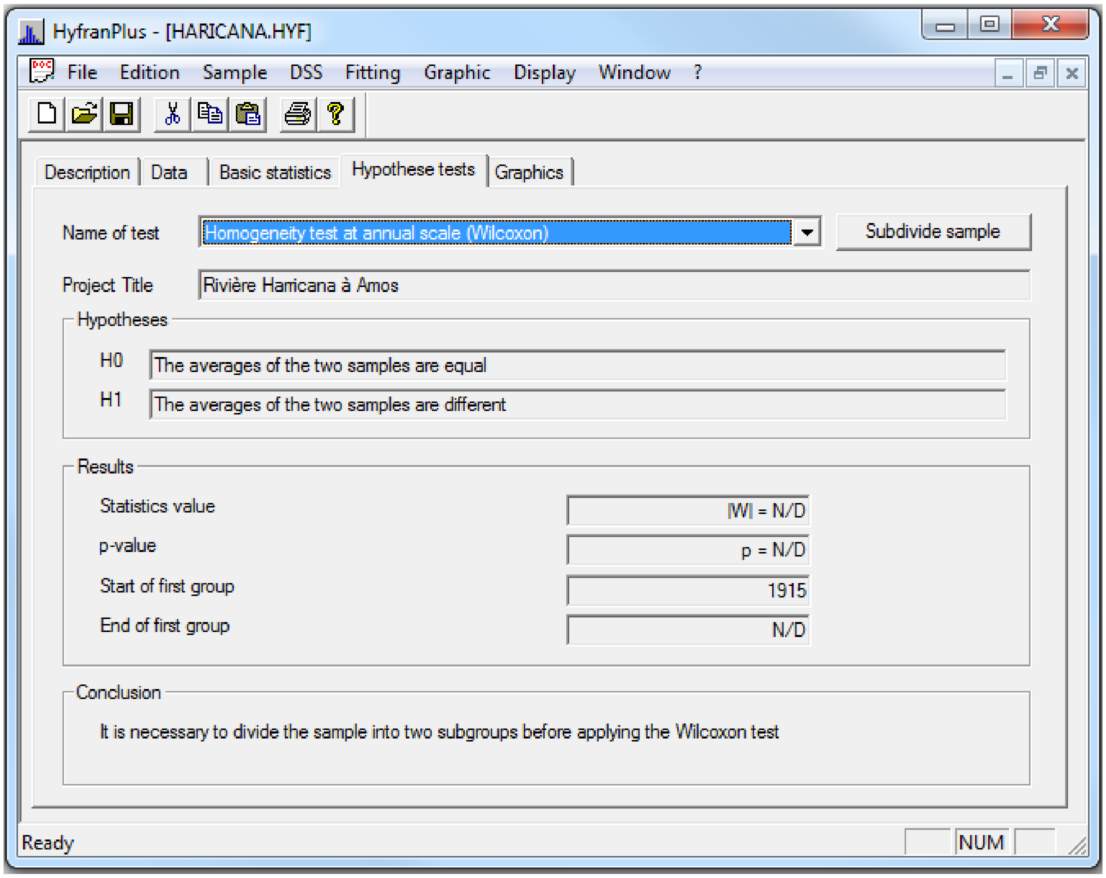


Figure 21: Homogeneity test at annual scale: Haricana project

d) *Test of homogeneity on a seasonal scale (Wilcoxon test or Mann-Whitney)*

As in the case of the homogeneity test on a seasonal scale the first sub-sample consists of observations included in the period from the month m (beginning of the first season) to the month n (end of the first season). The second sub-sample consists of observations included in the period from month n+1 (beginning of the second season) to month m-1 (end of the second season). Month m and n are specified by pressing the button "Subdivide the sample". This test can be used, for example, to check the homogeneity of spring floods (due to snowmelt) and fall flooding (due to rain) in order to check is if all the data can be grouped in the same sample.

Note: To perform this test, it is necessary to specify the month (Figure 22) for the edited or imported data (Figure 13). This test allows, for example, to check the homogeneity of the autumn floods caused by the precipitations (July-December) and spring floods caused by melting snow (January-June).

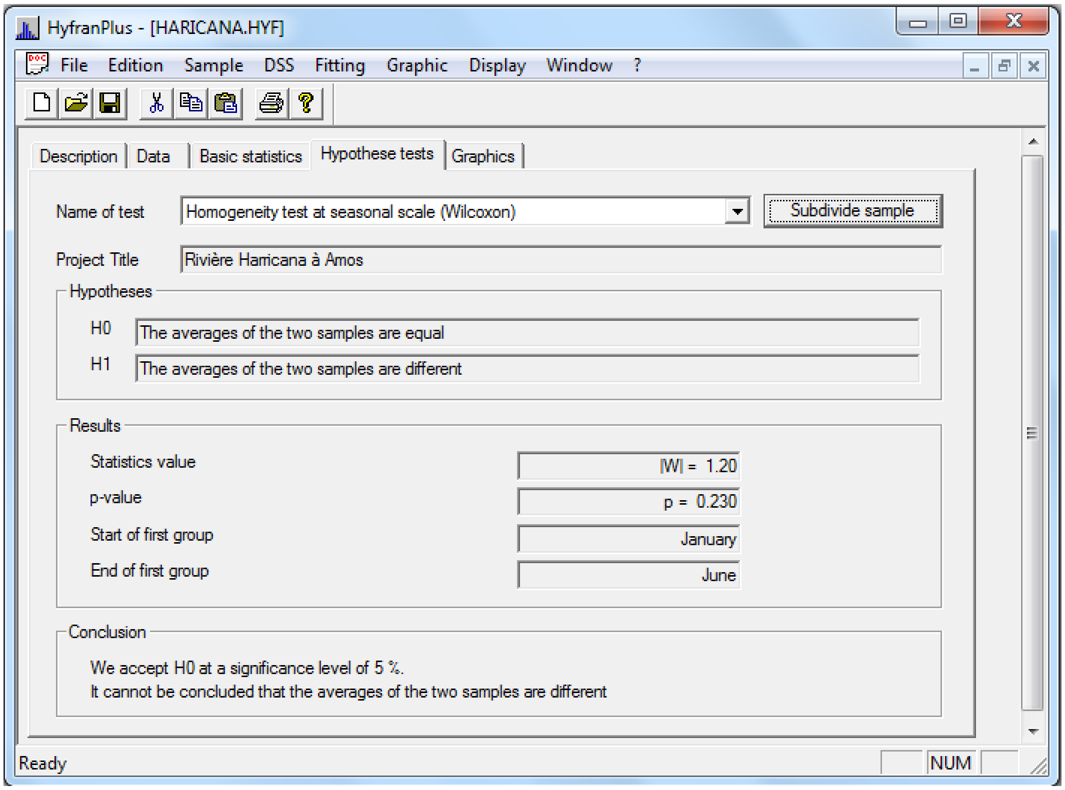


Figure 22: Homogeneity test at seasonal scale: Haricana project

2.1.5 Graphics

The "Graphics" tab of HYFRAN-PLUS (Figure 15) allows to view the data in different ways:

a) Observations on probability paper (normal or Gumbel);
b) Histogram of observations sorted by value;
c) Histogram of observations sorted by month;
d) Time curve.

In the following, these four points will be detailed and illustrated by examples.

a) *Observations on probability plot* (Figure 23)
This chart shows on normal or Gumbel probability paper (Figure 8), the PP (Plotting Position or empirical probabilities) associated with each observation of the sample.

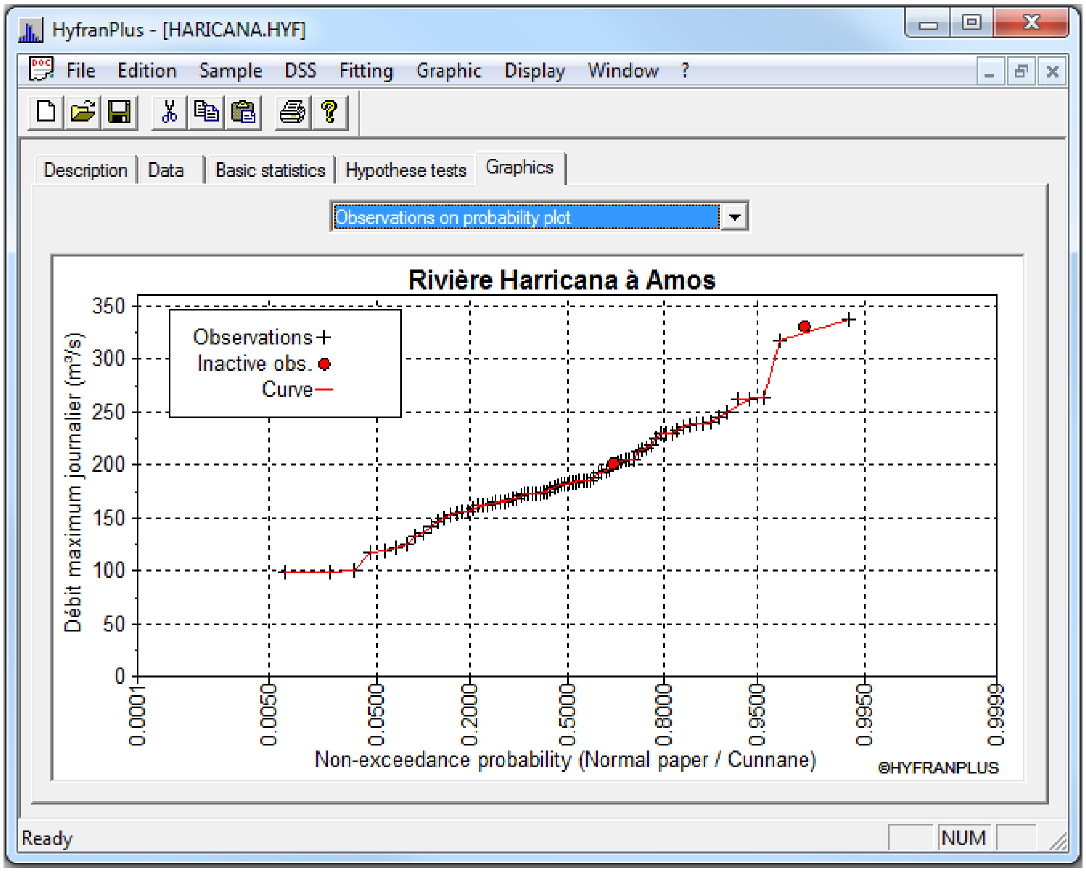


Figure 23: Observations on probability plot (normal paper): Haricana dataset

Note: The red points correspond to the deactivated values.

b) *Histogram of observations classified by value* (Figure 24): the histogram of the observations classified by value is a graph showing the number of observations per class value. This type of graph gives an illustration of the empirical probability density function of the data.
The number of equidistant classes is calculated using the following equation:

$$\text{Number of classes} = \left[5 \times log\left(n\right)\right]$$

Where n is the sample size and [.] is the integer value.

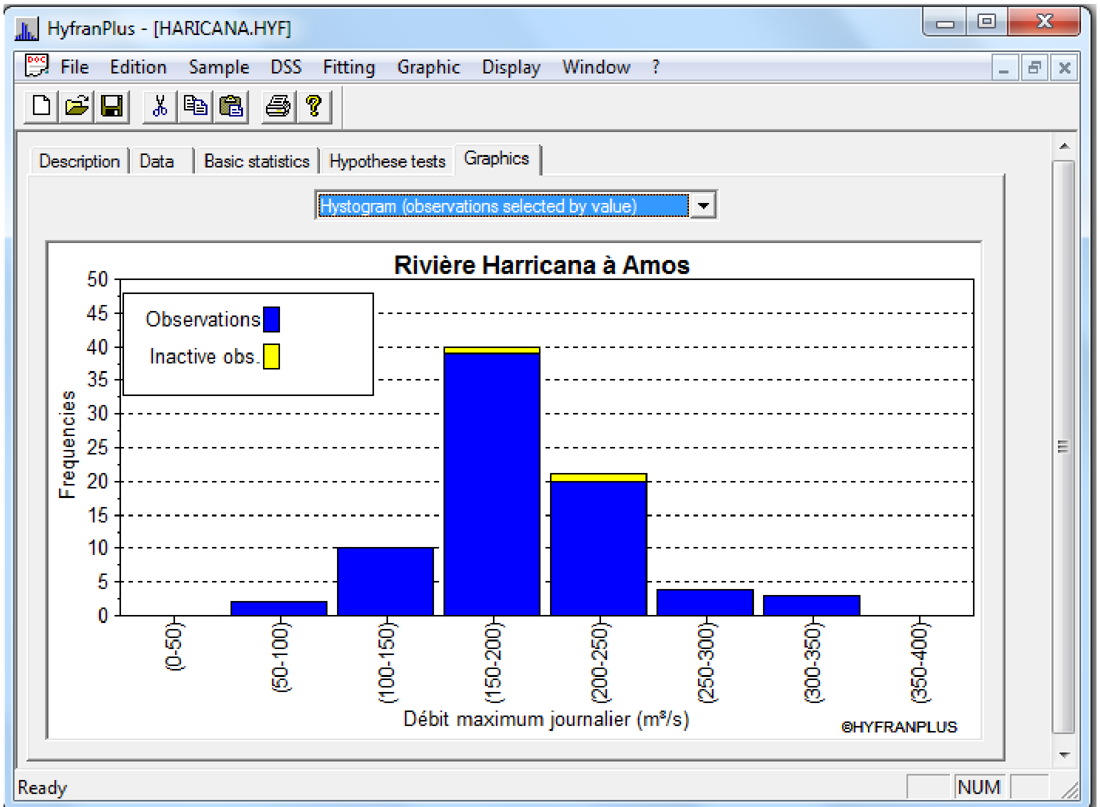


Figure 24: Histogram (observations classified by value): Haricana dataset

Note: The deactivated values are also shown in the histogram representation using different color (Figures 24 and 25).

c) *Histogram of observations classified by month* (Figure 25): the histogram of the observations listed by month is a graph showing the number of observations per month.

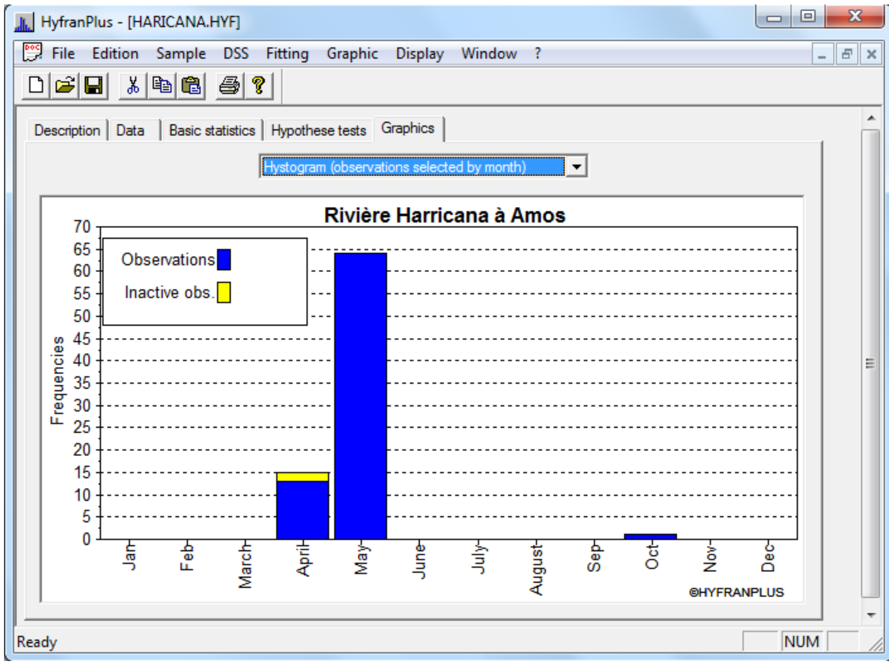


Figure 25: histogram (observations classified by month): Haricana dataset

d) *Time curve* (Figure 26): This graph shows the observations over time (year).

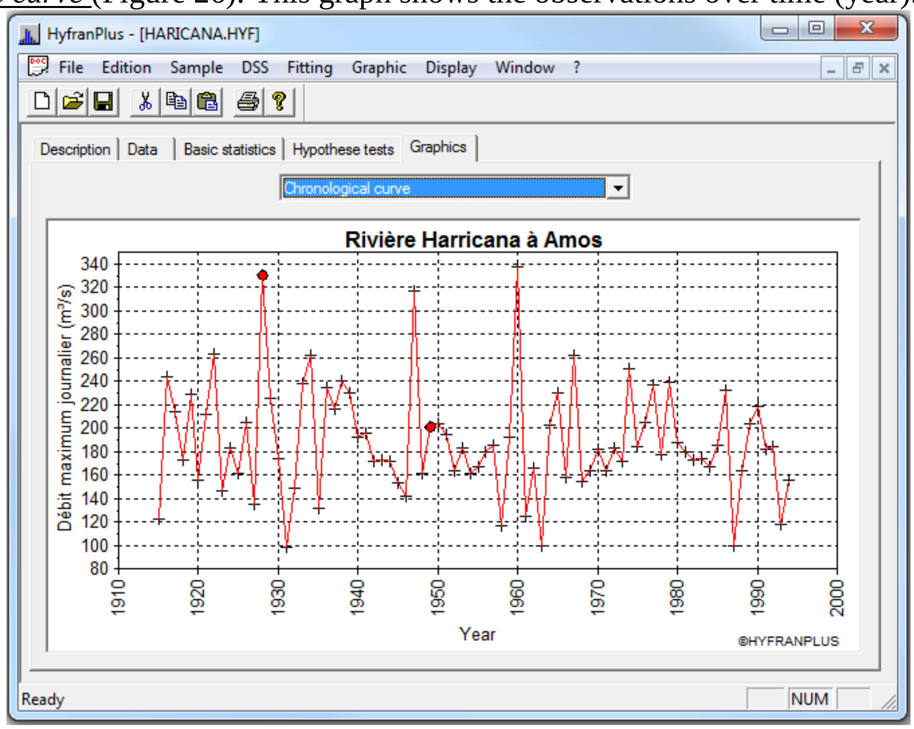


Figure 26: time curve: Haricana dataset

Note: The red points correspond to the deactivated values.

2.2. Decision Support System (DSS)

Before fitting a statistical distribution to the dataset, the DSS allows to select the most adequate class to represent the right tail behaviour (Figure B-1.Apendix B). For illustration purpose we'll use 3 datasets (see Annexe 1):

- The series corresponding to the Haricana project (Table A1) is available in the demo version of HYFRAN-PLUS. It will be shown (Figure 30) that the distribution that best fits this series belongs to the class C.
- A simulated series from the normal distribution (Table A2) then its logarithmic transformation which corresponds to lognormal distribution (Table A3). The log-normal distribution is a limiting case between classes C and D (El Adlouni, Bobée and Ouarda, 2008) case, and;
- A simulated dataset from a gamma distribution (class D) (Table A4).

In all these three cases the DSS is used as described in El Adlouni and Bobée (2011); El Adlouni, Bobée and Samoud (2012). Figure B-1 of the Appendix B represents a classification of the distributions, usually used in hydrology to fit extremes, with respect to their right tail behaviour.

a) **Dataset 1 (Annexe 1 – Table A.1) :**

- Test (Cv, Cs) (step 1 in the diagram DSS, Figure 6) is executed first to determine if you can test the log-normality. This is done using the "diagram log-normal" available in the "DSS" menu of HYFRAN-PLUS software (Figure 5). This option includes four tabs:
- " Graphics " in order to observe the diagram (Cv, Cs) (Figure 27),
- "Decision" gives the results of the test (Cv, Cs) (Figure 28),
- "JB test" performing and gives the conclusions of the JB test,
- "Help" which contains useful information on tests lognormal and Jarque-Bera (Martel, El Adlouni and Bobée, 2012).

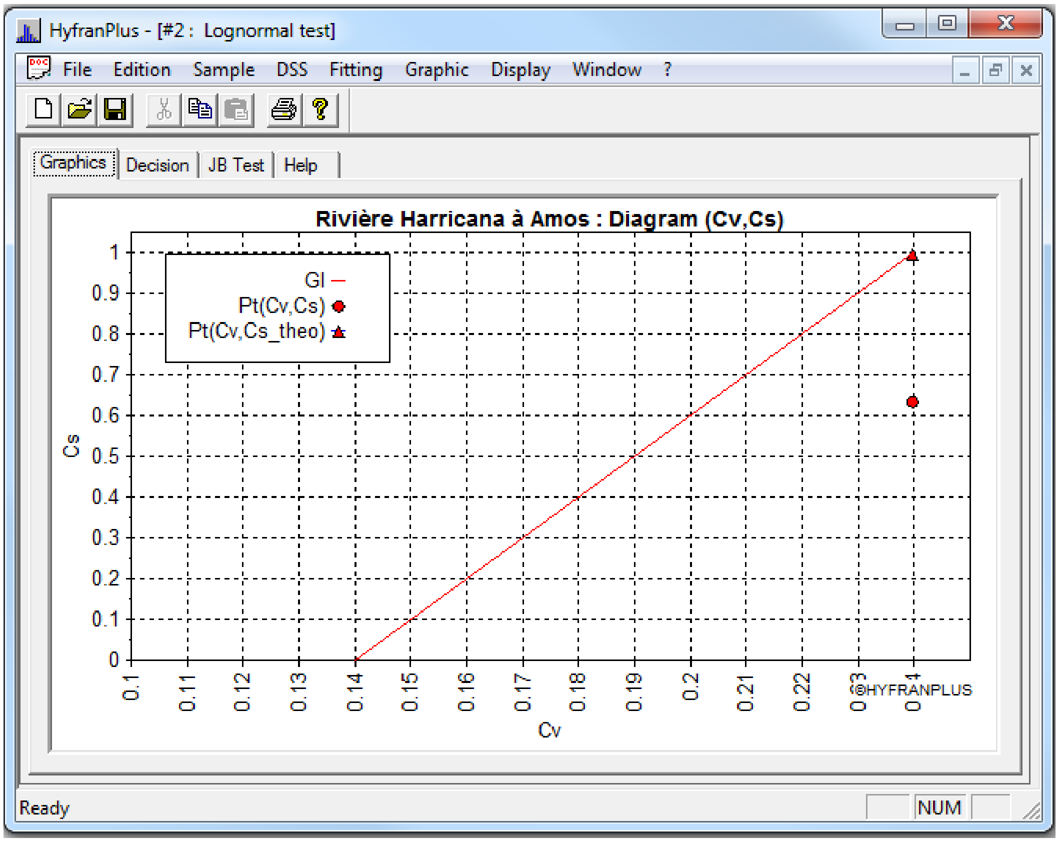


Figure 27: (Cv, Cs) diagram prior to the use of log-normality test (dataset 1)

- The observed point $\left(C_v, C_s\right)$ (●) is below the straight line (Figure 27), thus we can deduce (Figure 28) that we cannot use the test of Jarque-Bera (cf. Martel, El Adlouni and Bobée, 2012).

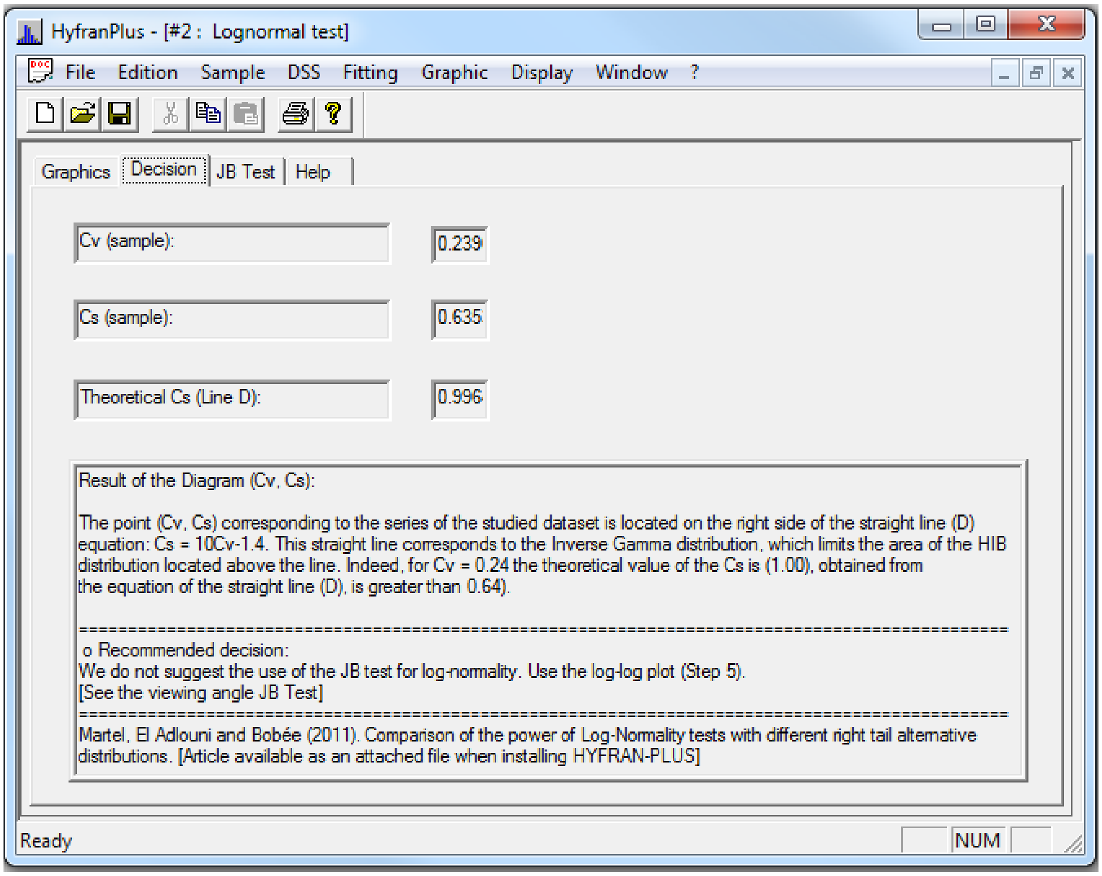


Figure 28: decision about the use of the log normality test (dataset 1)

- So we consider the log-log plot (step 5 of the DSS diagram, Figure 6) using the "log-log plot" option in the "DSS" menu (Figure 5). The selected option allows us to browse between three tabs:
  - o "Graphics" which allows to display the layout of the log-log diagram for the project studied (Figure 29),
  - o "Decision" to determine the class to which the data series and belongs (Figure 30) and,
  - o "Help" which contains useful information to understand the test.

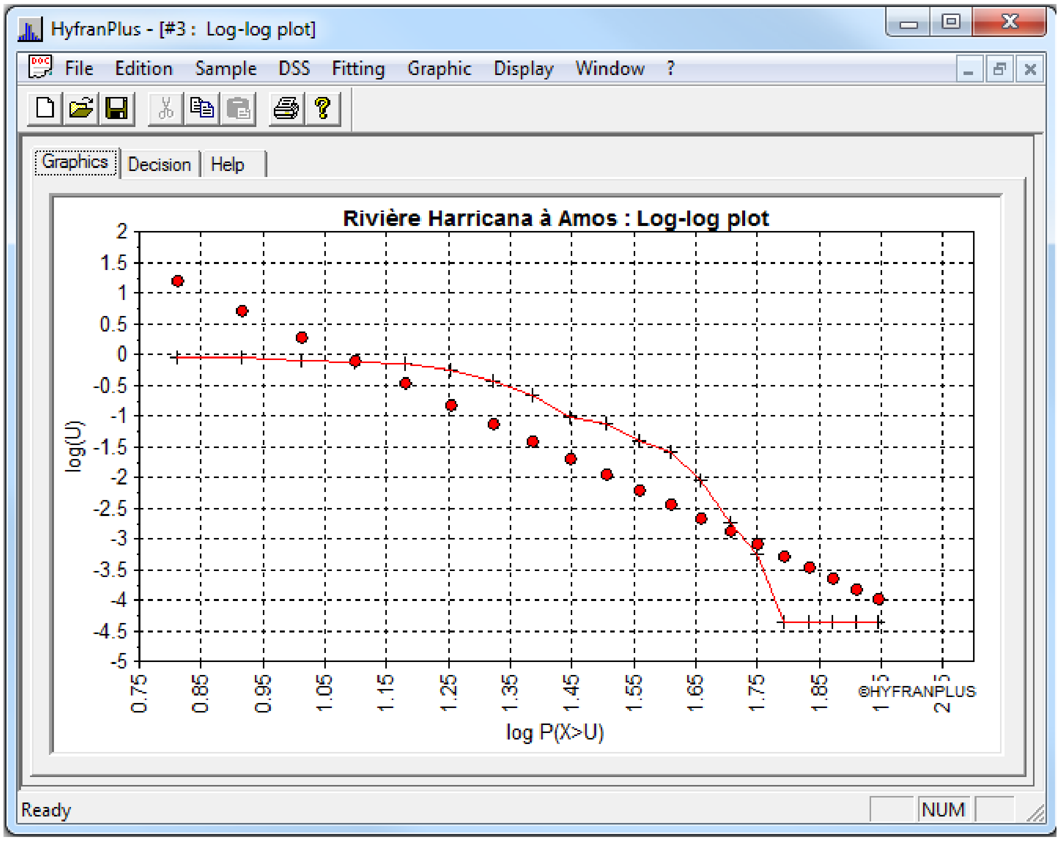


Figure 29: log-log plot (dataset 1)

Figure 29 shows the linearity of the curve is acceptable, we can deduce that the series belongs to the class C (step 6 of DSS diagram, Figure 6). Indeed (Figure 30), the observed correlation coefficient is greater than the critical value, it is not significantly different from 1. Therefore we can accept H0: the curve is linear (see El Adlouni, Bobée and Samoud, 2012).

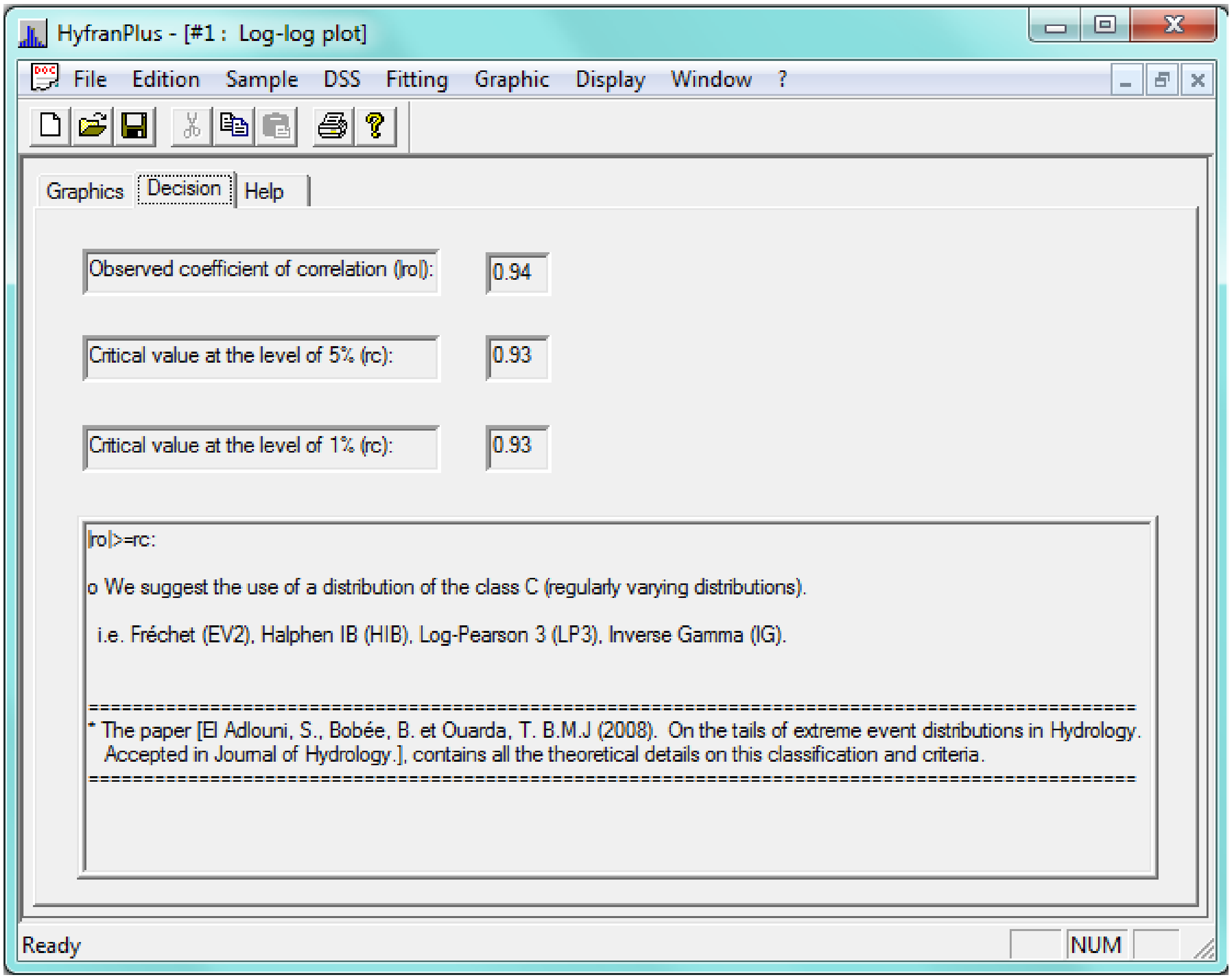


Figure 30: Decision related to the log-log test (dataset 1)

Note: The critical values for the log-log graph (Figure 30), were obtained by simulation (El Adlouni and Bobée 2011). We noticed that the values for the two thresholds 1% and 5% are identic even for 10,000 simulated samples.

- We use the Hill ratio and the Statistic of Jackson (Figure 5) to confirm the choice of the class C (step 10 of the DSS diagram, Figure 6). Figure 31 shows that the Hill statistic converges to a constant value different to zero, and Figure 32 corresponds to the statistic of Jackson which in this case converges towards 2. These results confirm that the series 1 may be represented by a distribution of class C (Fréchet (or EV2), Gamma Inverse, Halphen type Inverse B ... (Figure B-1, Appendix B).

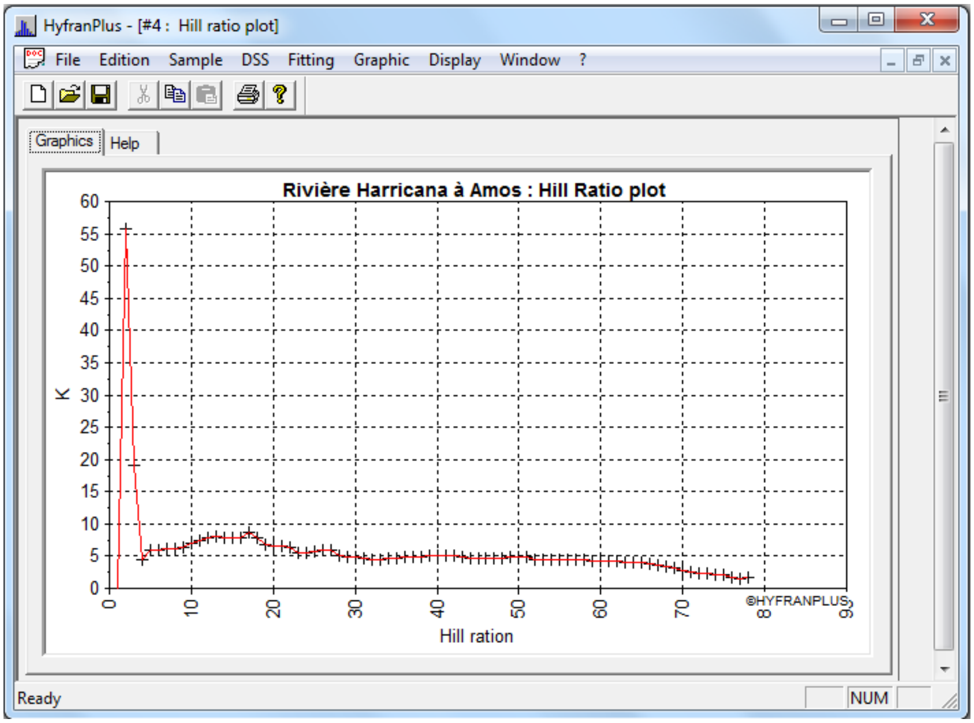


Figure 31: Hill Ratio plot (dataset 1)

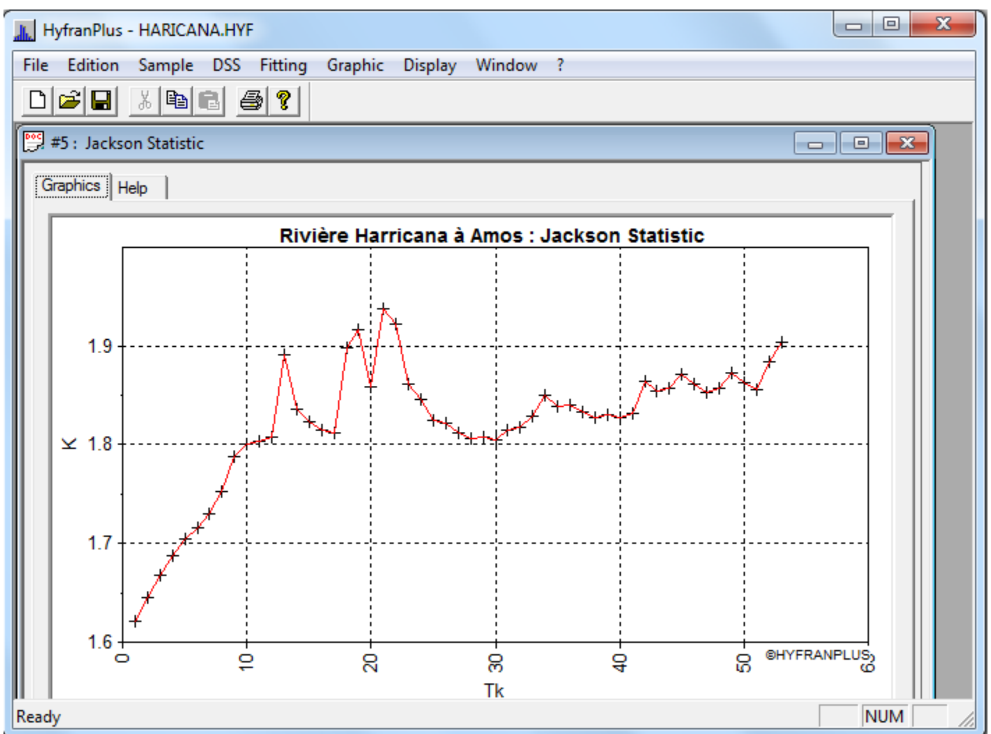


Figure 32: Jackson Statistic (dataset 1)

b) **Dataset 2 (Annexe 1 – Table A.2) :**

- The (Cv, Cs) plot is performed to determine whether it is possible to use the log-normal test diagram (step 1 in the DSS diagram, Figure 6).

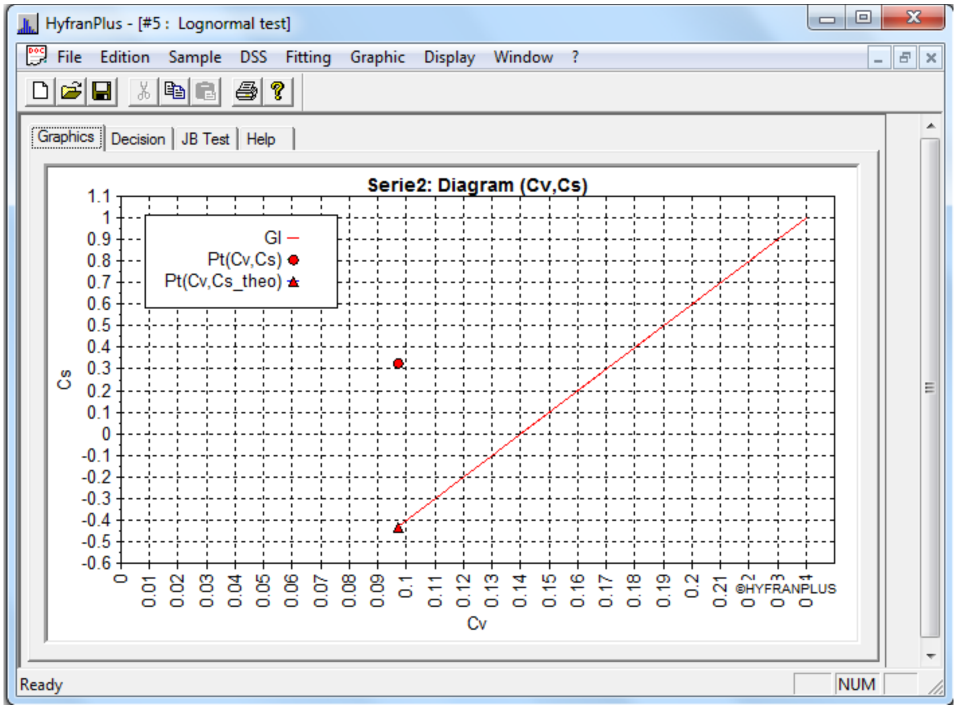


Figure 33: (Cv, Cs) diagram prior to the use of log-normality test (dataset 2)

The observed point (Cv, Cs) ( ) is above the straight line (Figure 33) and therefore belongs to the HIB area (see Martel, El Adlouni and Bobée, 2012), we can deduce (Figure 34) that the log-normality test is applicable (step 2 of the DSS diagram, figure 6). It therefore goes to the logarithmic transformation of the dataset 2 (Annexe 1 – Table A.3).

- The decision diagram (Cv, Cs) allows to test the log-normality so the Jarque-Bera test is performed (Figure 35).

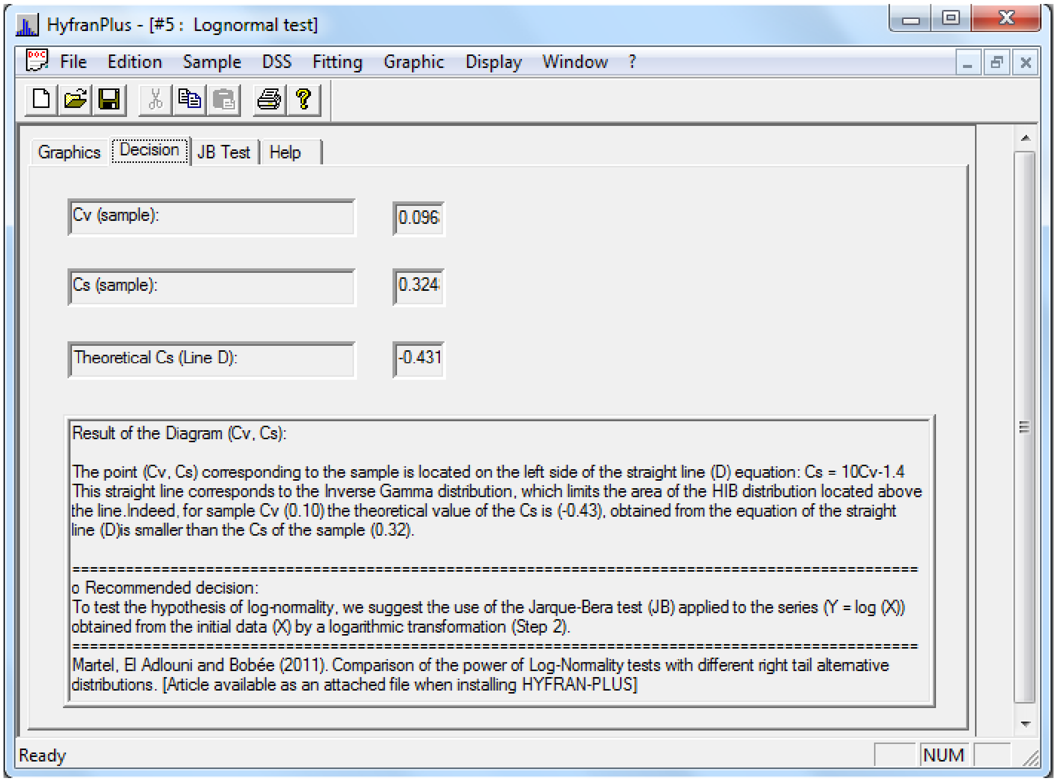


Figure 34: Decision concerning the use of the log-normality test (dataset 2)

JB test (Figure 35) shows that the assumption of log-normality is satisfactory, so we suggest the use of a log-normal distribution for fitting the dataset (step 3 of the DSS diagram, Figure 6).

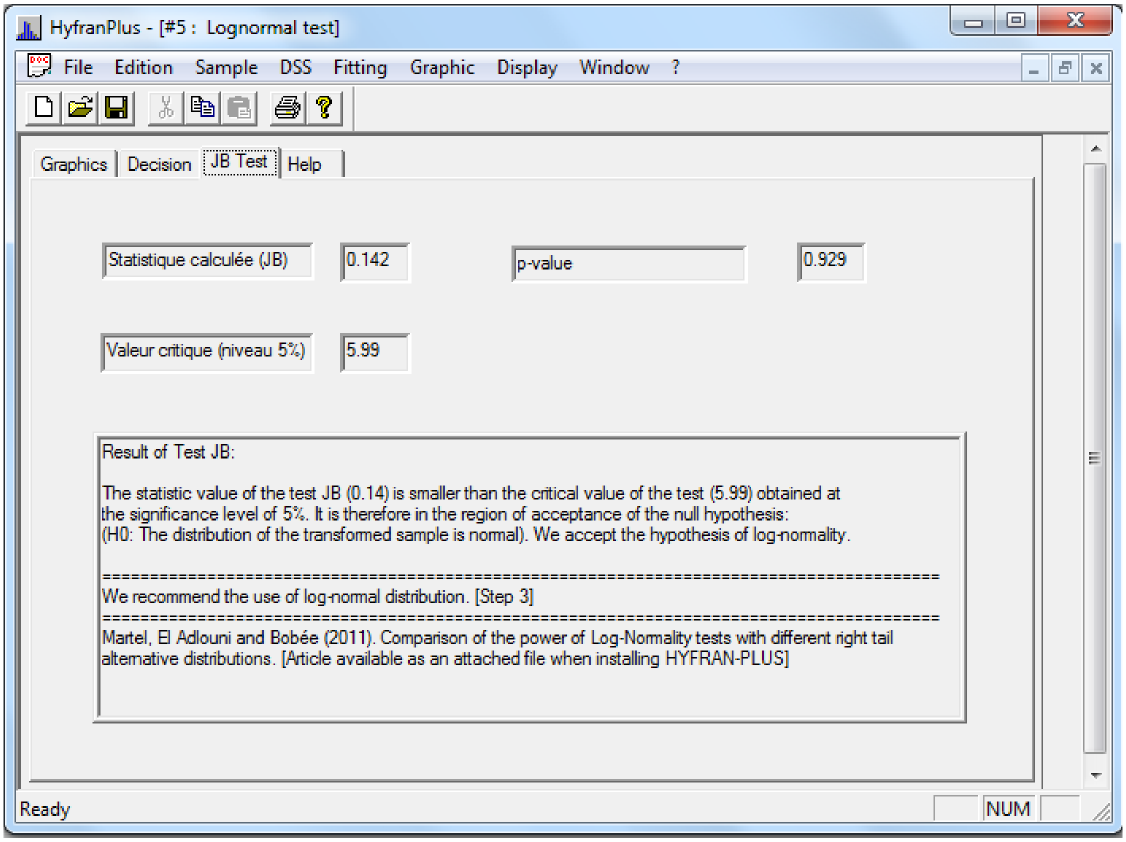


Figure 35: Jarque-Bera test decision (dataset 3)

Note that (Figure 36), if we choose to represent the dataset 3 (logarithmic transformation of the dataset 2) by the Log-normal distribution, the model curve will be a straight line on normal probability paper. This confirms the validity of the log-normal distribution to fit dataset 2. Indeed, if Y=log(X) is normally distributed, then X fellow a Log-normal distribution.

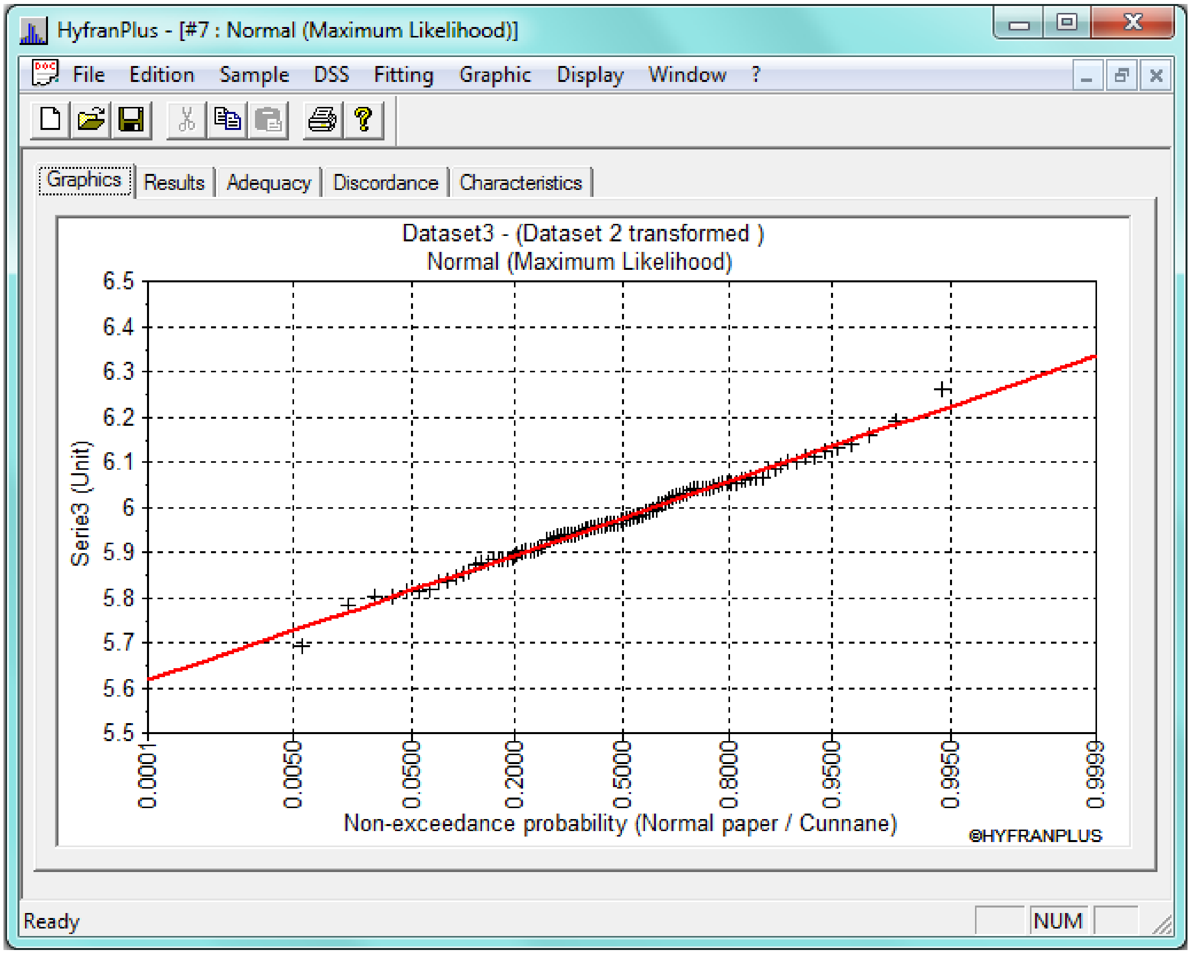


Figure 36: Graphical representation of the fitting of the dataset 3 by a normal distribution (logarithmic scale)

c) **Dataset 3 (Annexe 1 – Table A.4):**
   - Test (Cv, Cs) is performed to determine if we can test the lognormal (step 1 of the DSS diagram, Figure 6).

The observed point $\left(C_v, C_s\right)$ (●) is below the straight line (Figure 37), thus we can deduce (Figure 28) that we cannot use the test of Jarque-Bera (cf. Martel, El Adlouni and Bobée, 2012).

We then consider the log-log diagram (step 5 of the DSS diagram, Figure 6) using the option "log-log" option (Figure 5).

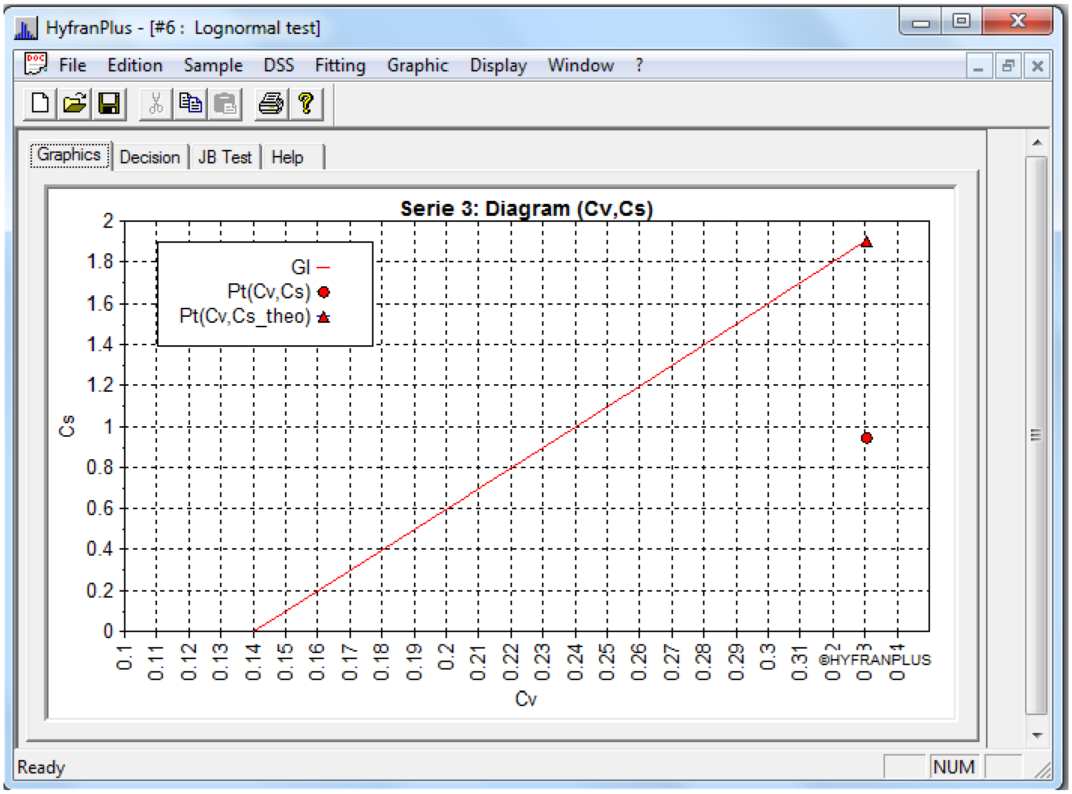


Figure 37 (Cv, Cs) diagram prior to the use of log-normality test (dataset 3)

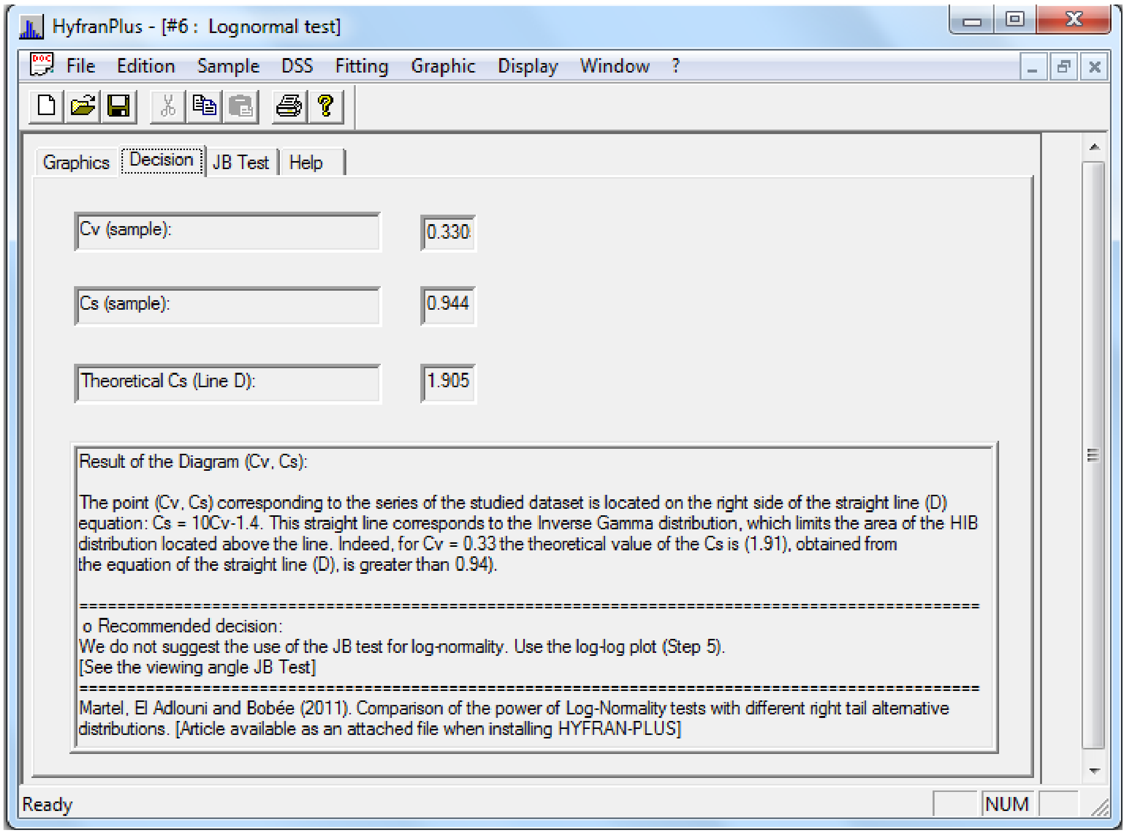


Figure 38: Decision related to the use of the log normality test (dataset 3)

We can see by considering the log-log diagram that the curve is not linear (El Adlouni, Bobée and Samoud, 2012) (Figure 39) we can then deduce (Figure 40) that the graph of the Mean Excess Function (MEF) should be used (Step 7 diagram DSS Figure 6).

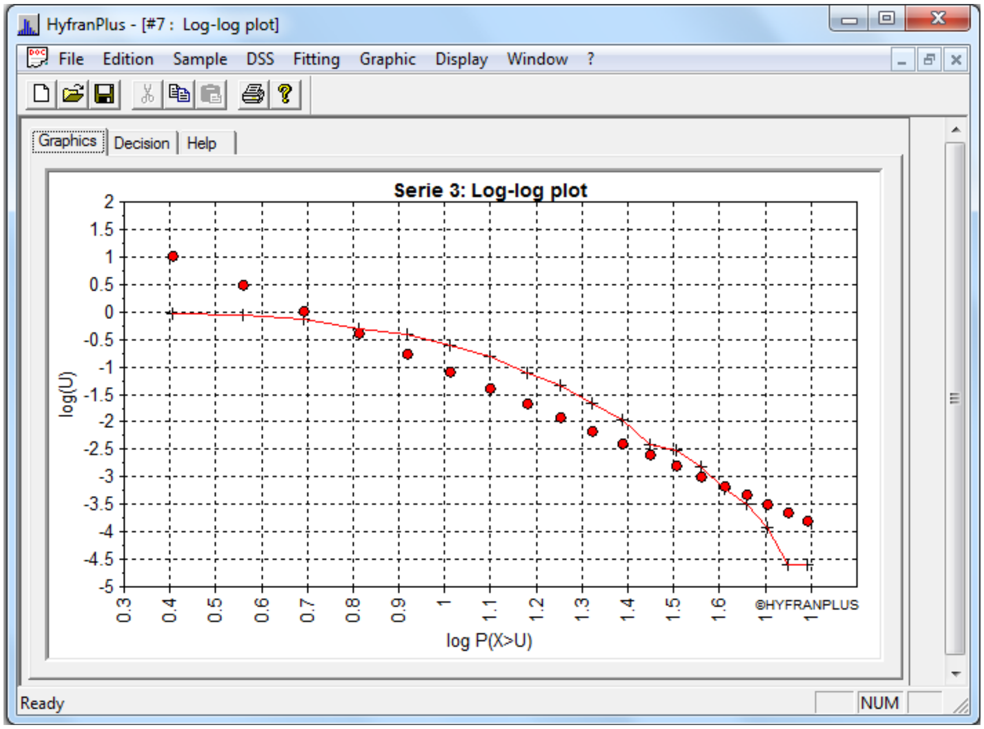


Figure 39: log-log plot (dataset 3)

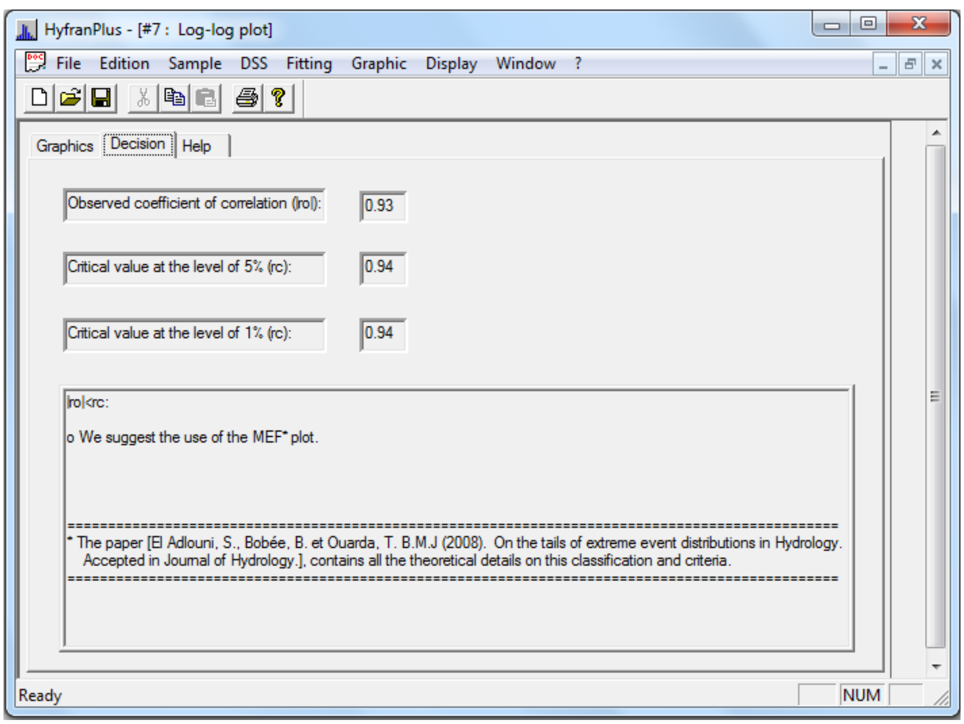


Figure 40: Log-log plot decision (dataset 3)

- To run the MEF test, select the 'Mean Function Excess" tab that can be found in the DSS menu. This option allows to navigate between 3 tabs:
  - "Graph" to observe the MEF curve (Figure 41),
  - "Decision" which provides the conclusions of the MEF test (Figure 42) and,
  - "Help" which contains useful information on the MEF test.

We note that the slope of the MEF curve is positive (Figure 41), we can then deduce that a distributions of the class D (FigureB-1, Appendix B) should be used to represent the dataset 3 (step 9 of DSS diagram, Figure 6). This conclusion is confirmed in Figure 42. Indeed, the observed slope *a0* is greater than the critical value *ac*, We reject H0 : The slope is null (Step 9 DSS diagram, Figure 6) (El Adlouni, Bobée and Samoud, 2012).

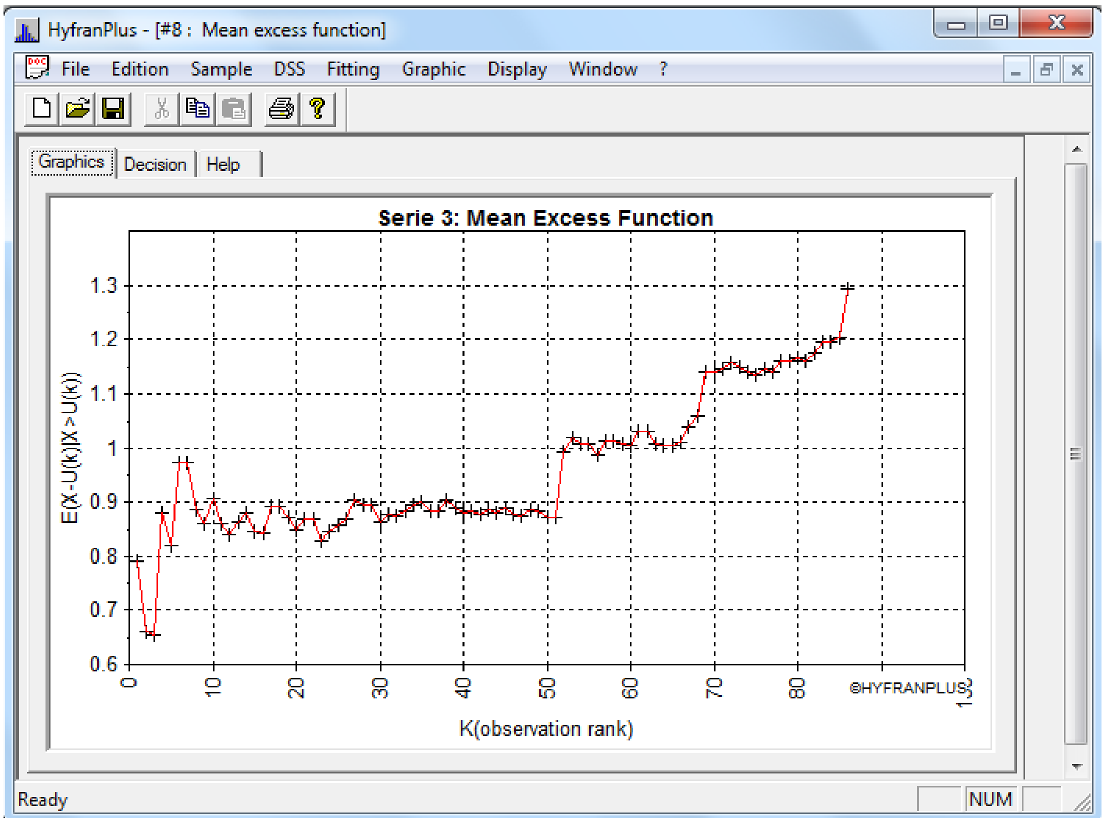


Figure 41: Mean Excess Function (dataset 3)

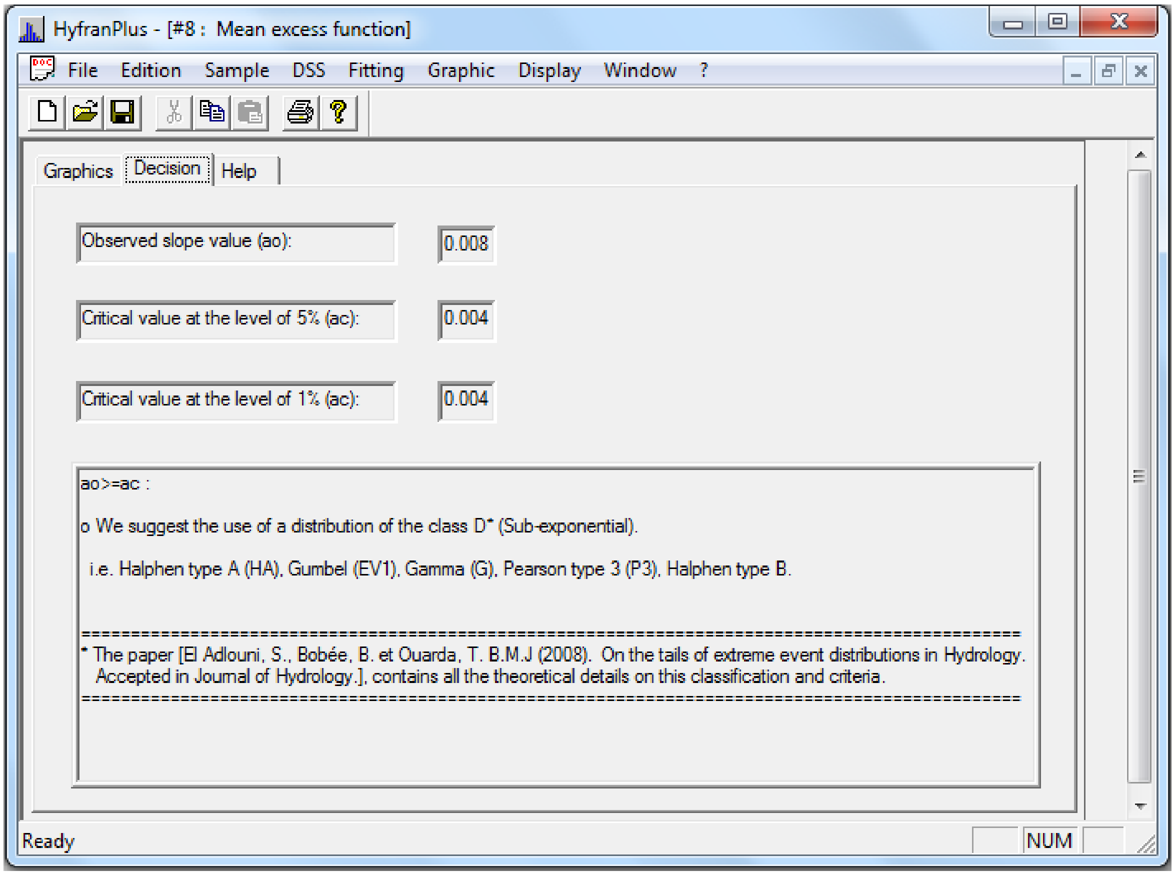


Figure 42: Mean Excess Function Decision (dataset 3)

Note: The critical values for the FME diagram (Figure 42), were obtained by simulation (El Adlouni and Bobée 2011). We noticed that the values for the two thresholds 1% and 5% are identic even when we run simulations with 10,000 samples.

- The Hill Ratio plot and Jackson Statistics are used to confirm the choice of the class D (step 10 of the DSS diagram, Figure 6). The figure 43 shows that the Hill statistic Hill converges towards zero and Figure 44 displays the Jackson Statistic which in this case has irregularities and does not converge to 2. We therefore deduce that (El Adlouni, Bobée and Samoud, 2012). the dataset 3 can be represented by a distribution of class D (Figure B-1, Appendix B) with sub-exponential tail: Gumbel, Halphen type B, Halphen type A ...

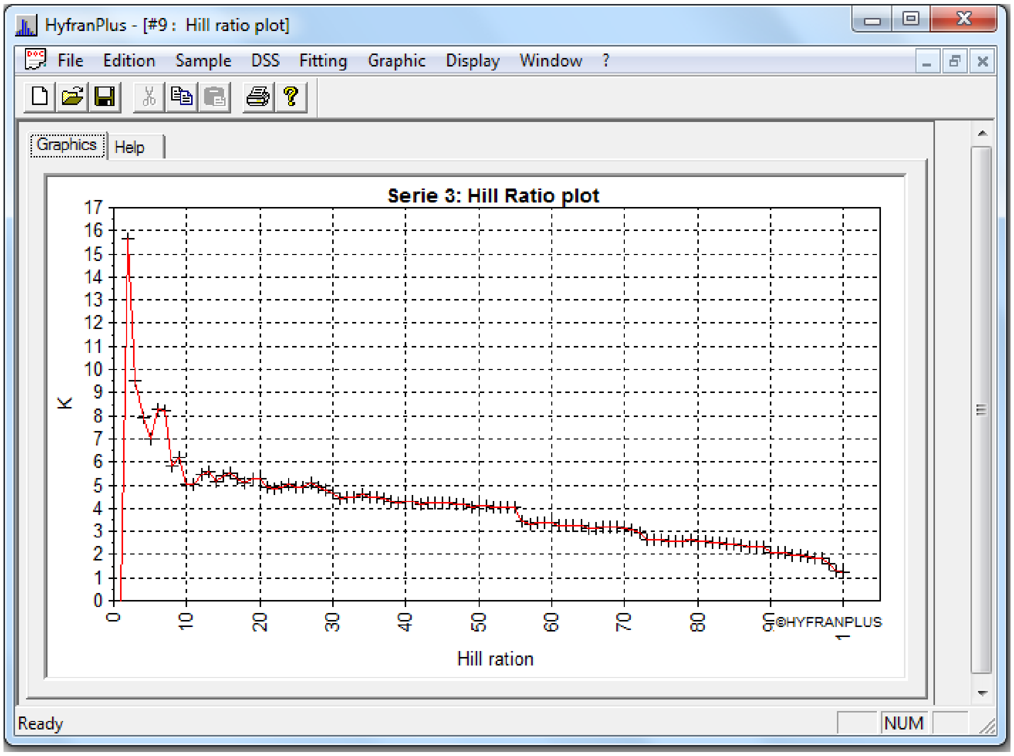


Figure 43: Hill Ratio plot (dataset 3)

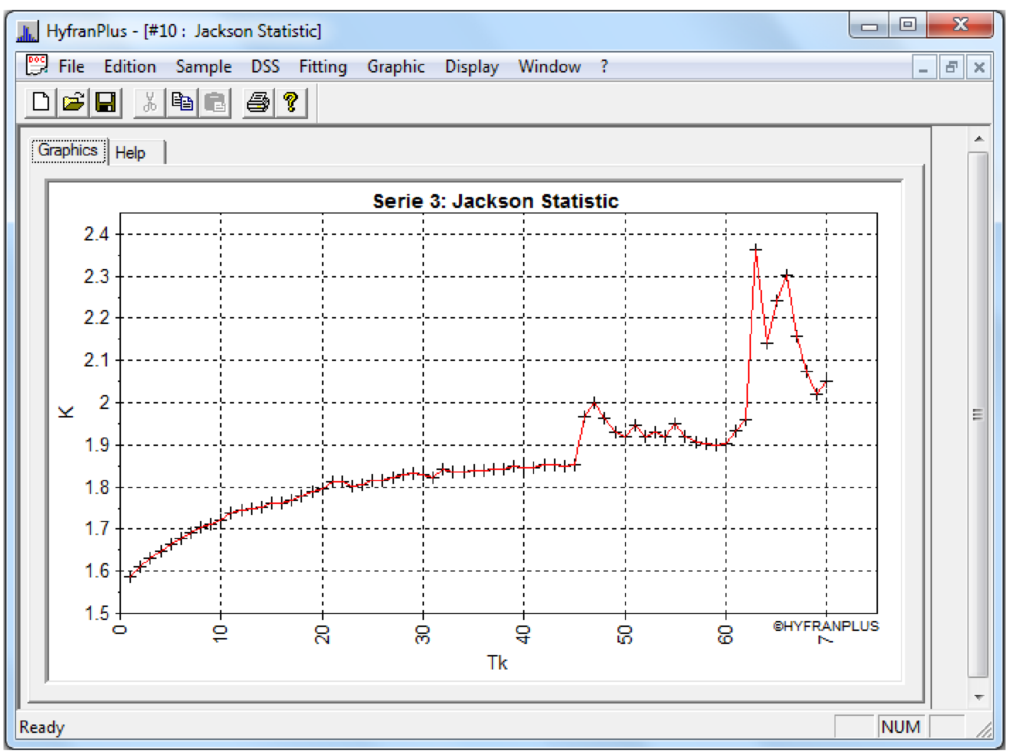


Figure 44: Jackson Statistic (dataset 3)

2.3. Fitting of a statistical distribution to a dataset

In the above, we used the DSS to select the most adequate class to represent the shape of the empirical distribution. Then we proceed to the fit of distributions belonging to the selected class. The classes C and D contain several distributions (Figure B-1, Appendix B). The choice of the most appropriate distribution can be done through graphical visualization or using information criteria (AIC and BIC).

In what follows, we will use, for illustrative purposes, the dataset 1 which may be represented by a distribution of class C (see Figure 30). HYFRAN-PLUS allows the fit of different statistical distributions to the IID dataset, considering different estimation methods.
To make a fit you have to follow these steps:

- Choose, alternately, each of the distributions of the selected class (Class C in the case of Series 1) in the "Fit" menu (Figure 7);
- Choose an estimation method, when more than one are available, by selection in the corresponding window that appears next;
- Press the "OK" button.

2.3.1. Selection of the most adequate statistical distribution

It has been shown (Figure 30) that the dataset 1 could be represented by a distribution of class C (Fréchet, Inverse Gamma, Halphen Inverste B, and Log-Pearson 3) (Figure B-1, Appendix B). HYFRAN-PLUS compares the fittings of several distributions to choose the most appropriate model to represent the dataset. In what follows discriminating distributions methods will be presented (graphics and information criteria). We will overfly fit for the requirements of this section.

Graphical comparison
We can compare the results of several different fits by using a normal or Gumbel probability graph plot. First make consecutive fittings of the distributions (five maximum) and compare by selecting "Compare" option “fitting" menu (Figure 7). A dialog box that allows to select the distributions appears.
After selecting the right distributions, a graphic is obtained with all distributions represented and the representation of empirical probabilities (Plotting Position) of the dataset 1 (Figure 45).

Note: We cannot fit of the Halphen type B inverse to the dataset 1 as the maximum likelihood equations system fails to converge in this case (Perreault, Bobée et Rasmussen, 1999).

Note that in figure 45 the inverse Gamma distribution (red), gives a more adequate fit than the Fréchet (green) and Log-Pearson 3 (blue) distributions.

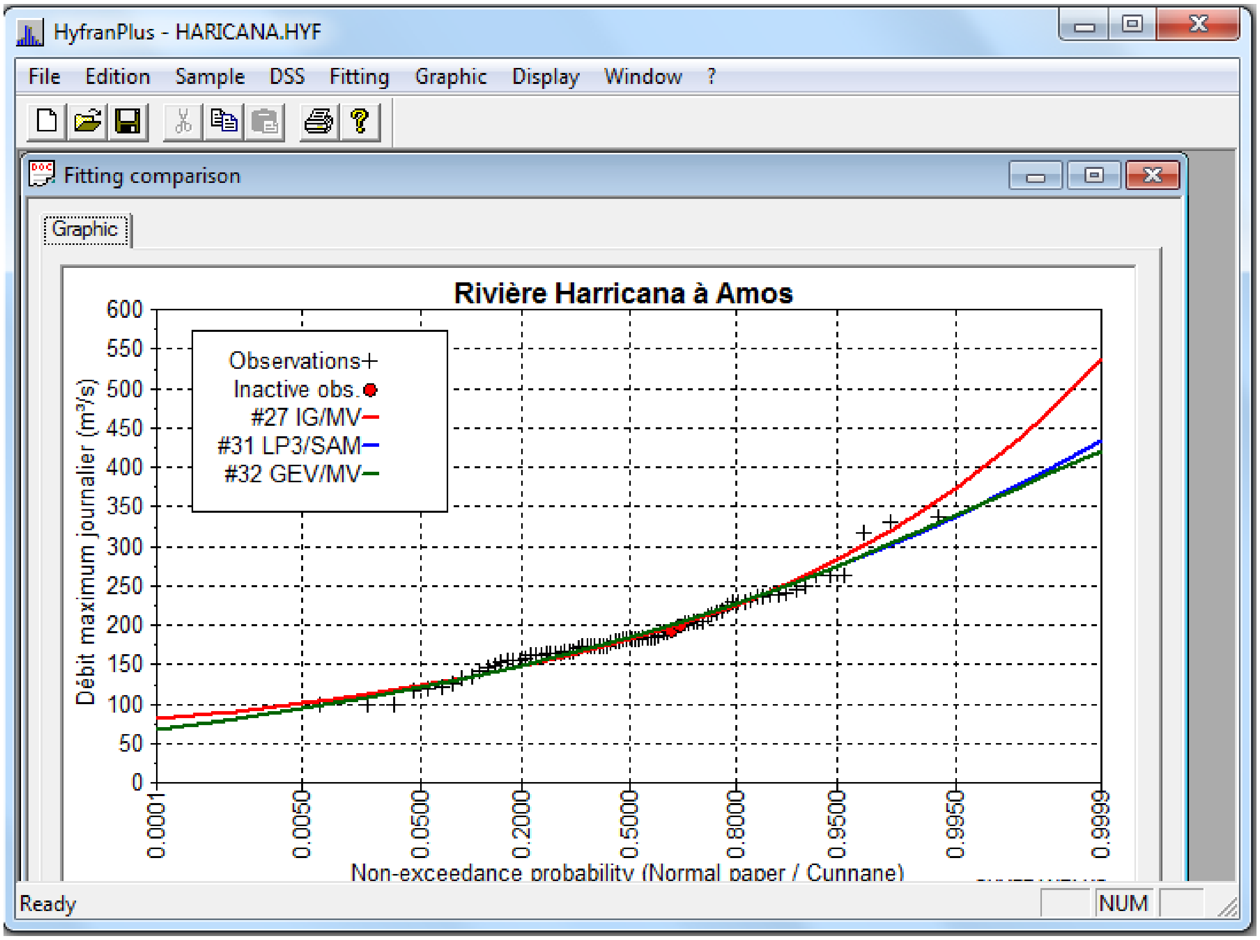


Figure 45: graphical comparison of dataset 1 fittings

Note: Note that the Frechet (EV2) is a particular case of the GEV (Generalized Extreme Value) distribution.

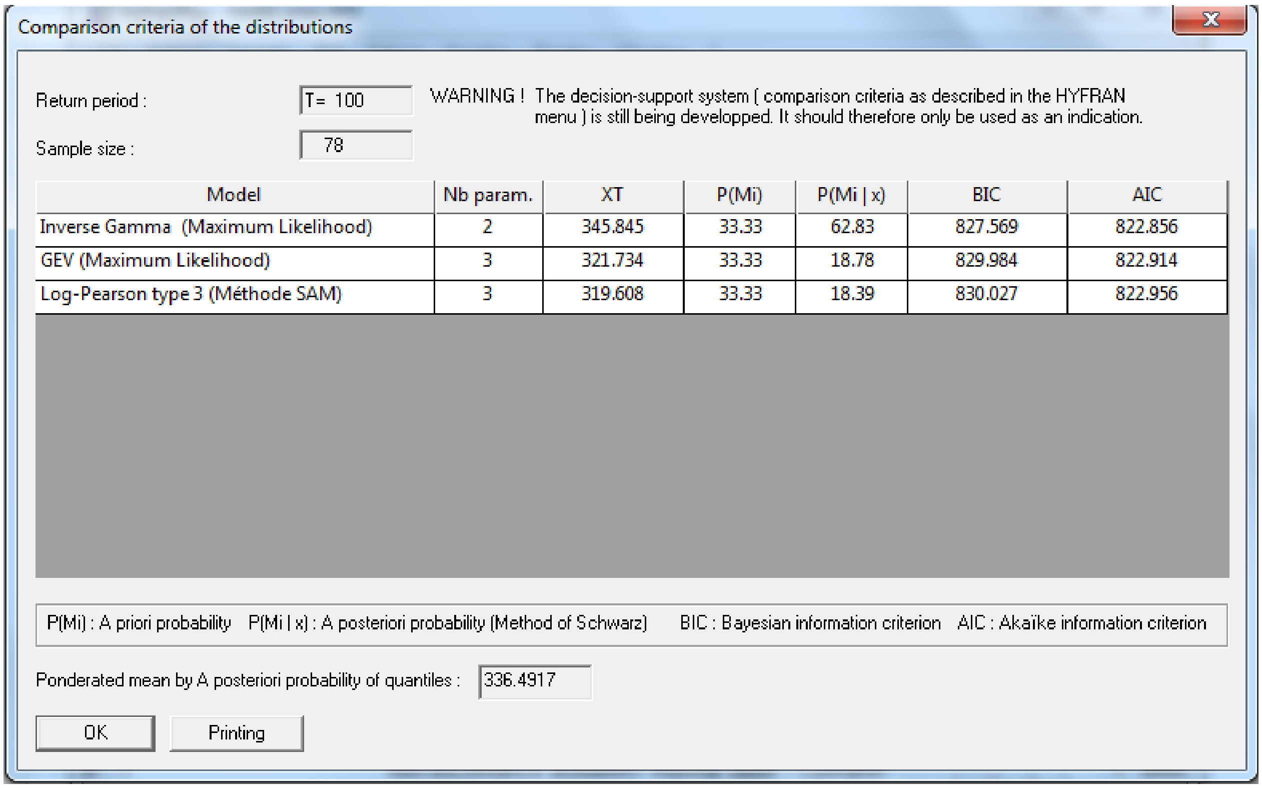
Comparison Criteria

To discriminate the different fits available the following two criteria are used in HYFRAN–PLUS (see Ehsanzadeh, El Adlouni and Bobée, 2010):

- Akaike Information Criterion (AIC)
- Bayesian Information Criterion (BIC)

These two criteria used consecutively (AIC and BIC) allows to build a classification of statistical models taking into account the principle of parsimony. Best fits correspond to lower values of criteria (see Ehsanzadeh, El Adlouni and Bobée, 2010).

To make the comparison, first fitting distributions must be made, and then the comparison is done by selecting the "Compare" option in "Fit" menu (Figure 7). A dialog box for selecting the different distributions (five maximum) appears. In the case of BIC, after selecting the right distributions, we specify the ratio of the prior probability (i.e. P(Mi) the weight given to each distribution) and the return period. The posterior probability P(Mi | x) is then determined by taking in account the observed dataset and then the BIC is deduced (see Ehsanzadeh El Adlouni and Bobée 2010). Results are presented in a new window that contains all the information relating to compared distributions (Figure 46).



Model	Nb param.	XT	P(Mi)	P(Mi | x)	BIC	AIC
Inverse Gamma (Maximum Likelihood)	2	345.845	33.33	62.83	827.569	822.856
GEV (Maximum Likelihood)	3	321.734	33.33	18.78	829.984	822.914
Log-Pearson type 3 (Méthode SAM)	3	319.608	33.33	18.39	830.027	822.956

Figure 46: compared fittings of dataset 1 using both information criteria

The criteria AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) and the notations of the table in Figure 46 are explained in detail in Ehsanzadeh El Adlouni and Bobée (2010).

Note: The AIC and BIC criteria suggest the use of distributions with the smallest number of parameters (principal of parsimony). However, in general to model annual maxima dataset, in hydrology, Morlat (1956) recommends the use of three-parameter distributions to take into account the shape (the skewness). Indeed, in the case of 2-parameter distributions shape is fixed (example: The coefficient of skewness for the Gumbel distribution is 1.137).

2.3.2. Fitting

In agreement with the graphical comparison (Figure 45) and the criteria AIC and BIC (Figure 46), the Inverse Gamma distribution has been selected to represent the dataset 1 (Rivière Haricana at Amos from 1915 to 1994, Table A.1). Therefore we represent the dataset 1 Haricana using the Inverse Gamma distribution for the tutorial part of this guide.

When performing the fitting of a statistical distribution to a given sample, a dialogue box appears (Figure 2 to 7), presenting the results of the fit as well as related information.

To access to the different steps of the fitting, we browse between different tabs (Figure 47):

- Results
- Graphic
- Adequacy
- Characteristics of the population
- Discordance test (appears only in the case of Normal and Log-normal distributions).

Fitting results

The window of the results of the fit (Figure 47) gives the following options:
The project, that is the name of the file containing the sample and its path;

- The title of the project;
- The size of the sample;
- The value of the parameters estimated of the Inverse Gamma distribution;
- The quantiles x_T for 21 predefined return periods. In the order from left to right we have:
    - the return period $(T = 1/p)$, where p is the probability of exceedance of x_T ;
    - the probability of non-exceedance $(q = 1 - p)$;
    - the value of the corresponding quantile x_T, the standard deviation of the quantile σ_{x_T} and finally,
    - the confidence interval CI of the quantile to a specified level of confidence (A summary of the theoretical approach to determine the quantile x_T, the standard deviation of the quantile σ_{x_T}, and the associated confidence interval is presented in the Appendix A);
- "Other return period" button,
- The confidence level (95% by default) of the confidence interval can be edited.

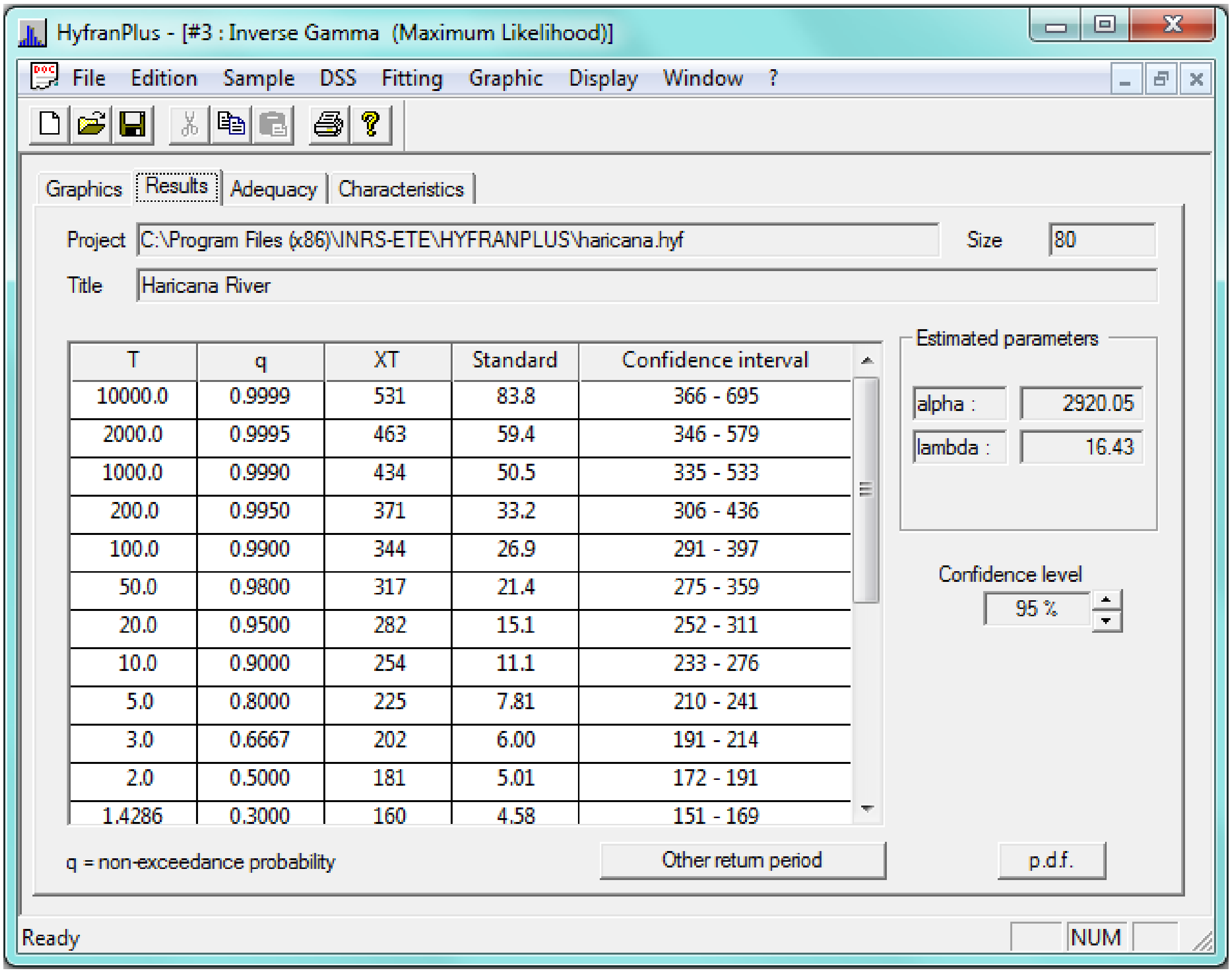


Figure 47: Results of the Inverse Gamma distribution fit to dataset 1

Graphic

It is appropriate to analyze the adequacy of a fitting through graphical representation. HYFRAN-PLUS allows viewing the fitting on Normal or Gumbel probability paper (Paper type selection is in the "Graphic" menu, Figure 8).

By default, the graph (figure 48) shows the sample data, the theoretical curve of the fitting (red line), the confidence interval (blue lines) whose level is specified in the "Results" tab and the legend. It is possible to remove or add some elements in the figure from the "Graphic" in the menu bar (Figure 8).

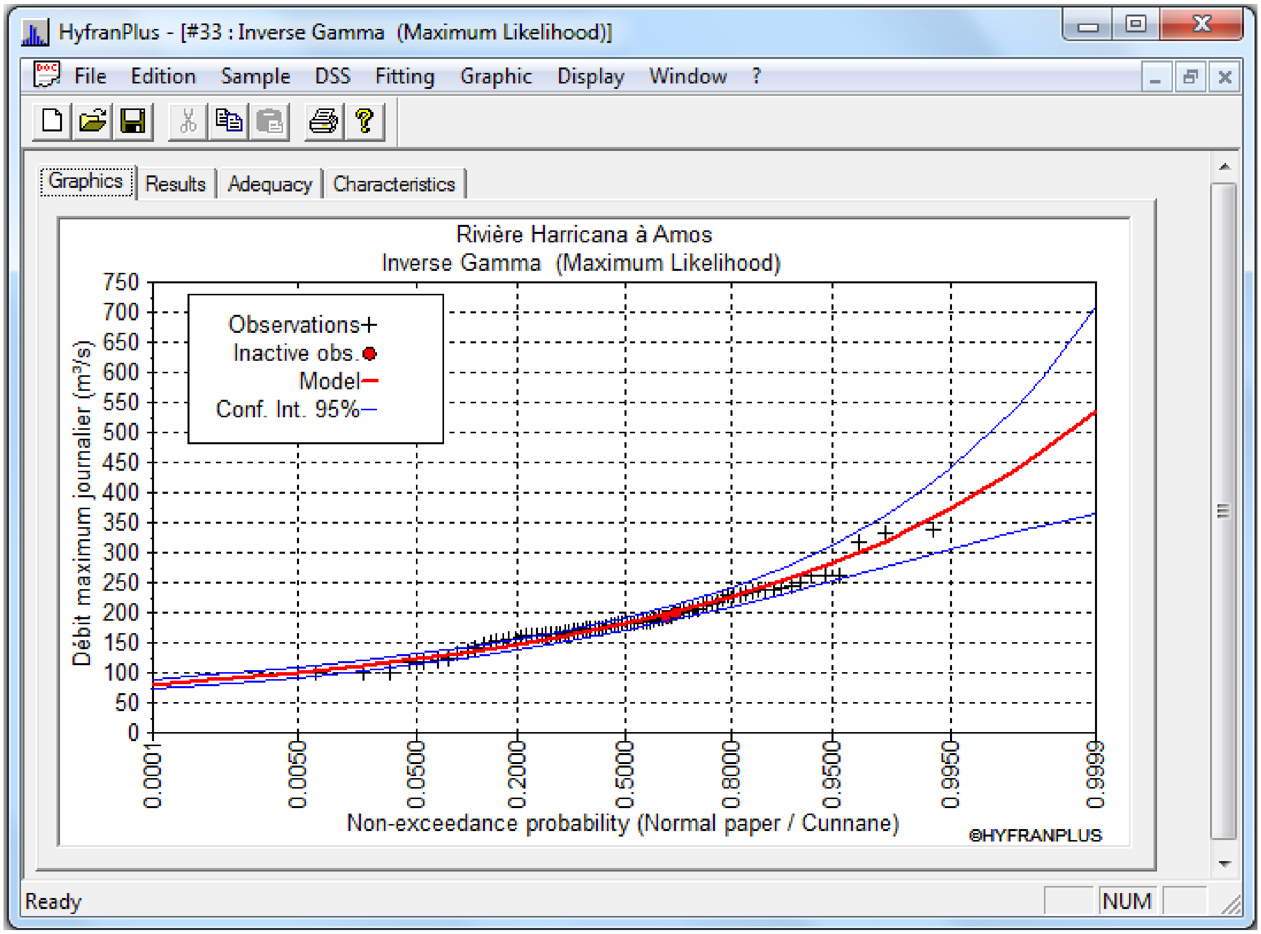


Figure 48: Graphical illustration of the Gamma Inverse distribution fit to dataset 1

Other return Period

For a given fit, HYFRAN-PLUS displays the results for 21 return periods, chosen for general use. However, the user may add many other return periods and get the corresponding estimated events.

To add a return period, from the results screen fitting proceed as follows:

- Press the "Other return period" button (Figure 47);
- The box "Other return period" dialog appears immediately (Figure 49)
- In the text box "Value T" enter the new return period you want to add.
- Press the "OK" button. The new return period is then inserted into the list.

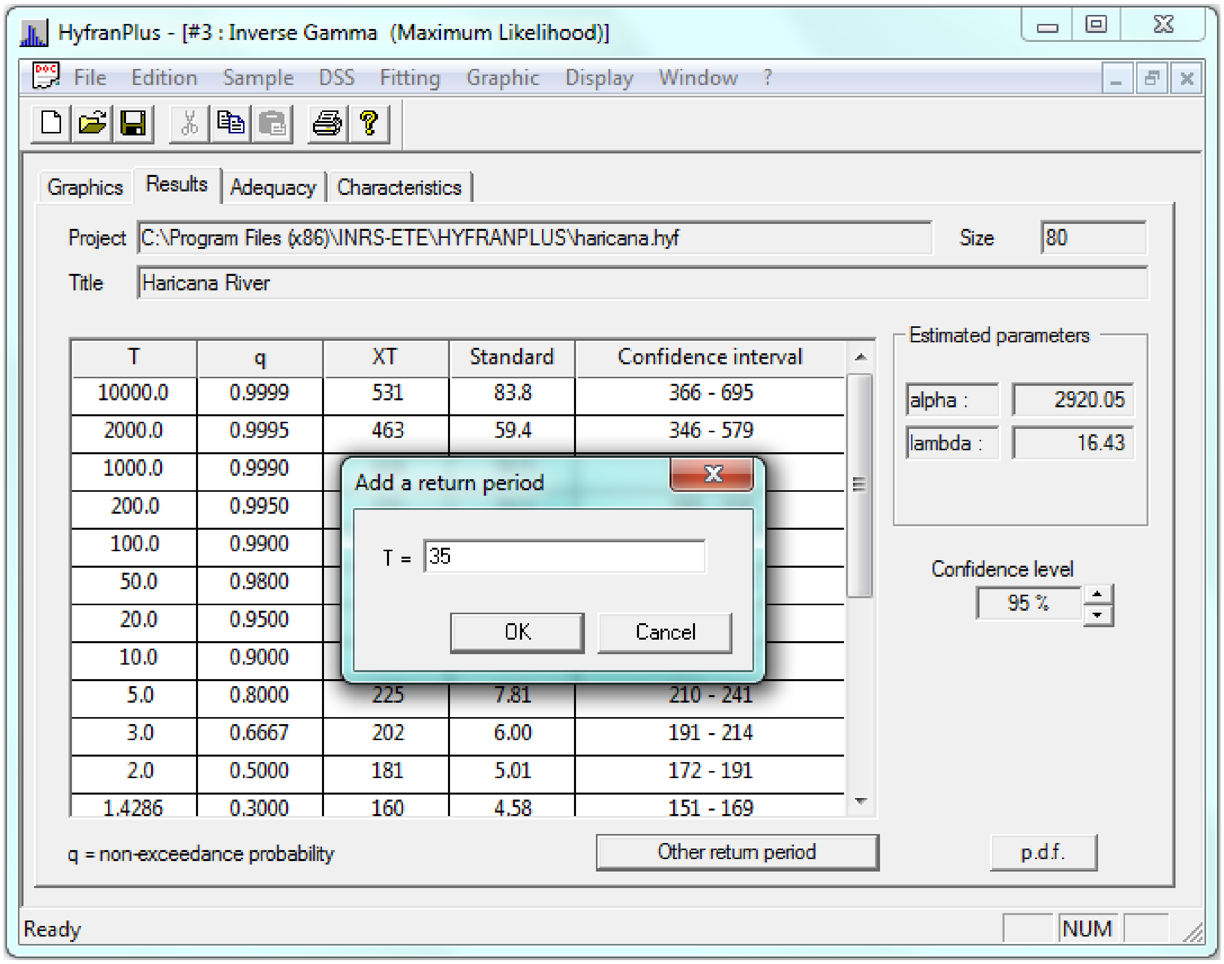


Figure 49: Example of the addition of another return Period

- Changing the level of the confidence interval

By default, the level of the confidence interval is 95%. To change it you need to use the arrows to the right of the text box "Confidence level". The confidence level ranges from 1% to 99%. HYFRAN-PLUS immediately displays the results for the new confidence level.

Adequacy

In order to judge objectively the quality of the fit to the data, there are various statistical adequacy tests. According to the distribution used and the sample size, HYFRAN-PLUS present the results for some of the following tests (see Bobée and El Adlouni, 2015 et Compaoré, El Adlouni et Bobée, 2013):

- Chi-square test (applicable for all statistical distributions)
- Test based on the sample moments (applicable only for normal and Log-normal distributions)

In the "Adequacy" tab (Figure 50), we find the following information:

- The project i.e. the name of the file that contains the sample and its path (Figure 11)
- The title of the project specified in the "Project Description" tab is shown (Figure 12)
- The null hypothesis (H_0) and the alternative hypothesis (H_1);
- Test results: found in this context the value of the test statistic and p-value of the statistic (Bobée and El Adlouni, 2015 et Compaoré, El Adlouni et Bobée, 2013).
- Finally, the conclusion of the test i.e. acceptance or rejection of the null hypothesis at a significance level of 5% or 1%

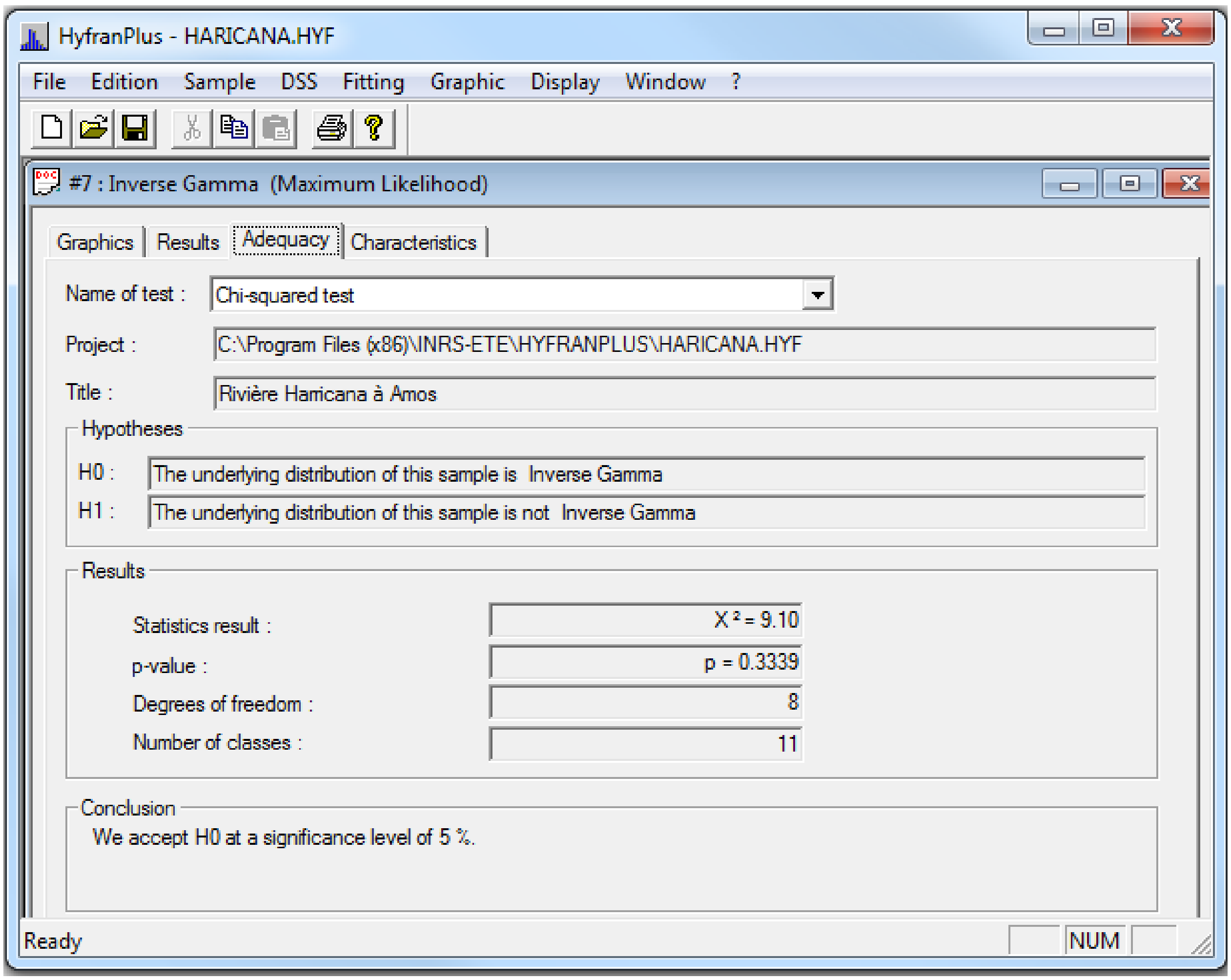


Figure 50: Adequacy for the fitting of dataset 1 by Inverse Gamma distribution

Statistical Characteristics for the fitted distribution

The window of the "characteristics of the population" (Figure 51) shows the statistics for the fitted distribution. In the first column we find the characteristics of the population and in a second the characteristics of the sample.

The size of the sample is given in column of sample characteristics. The project and the project title also appear in this window.

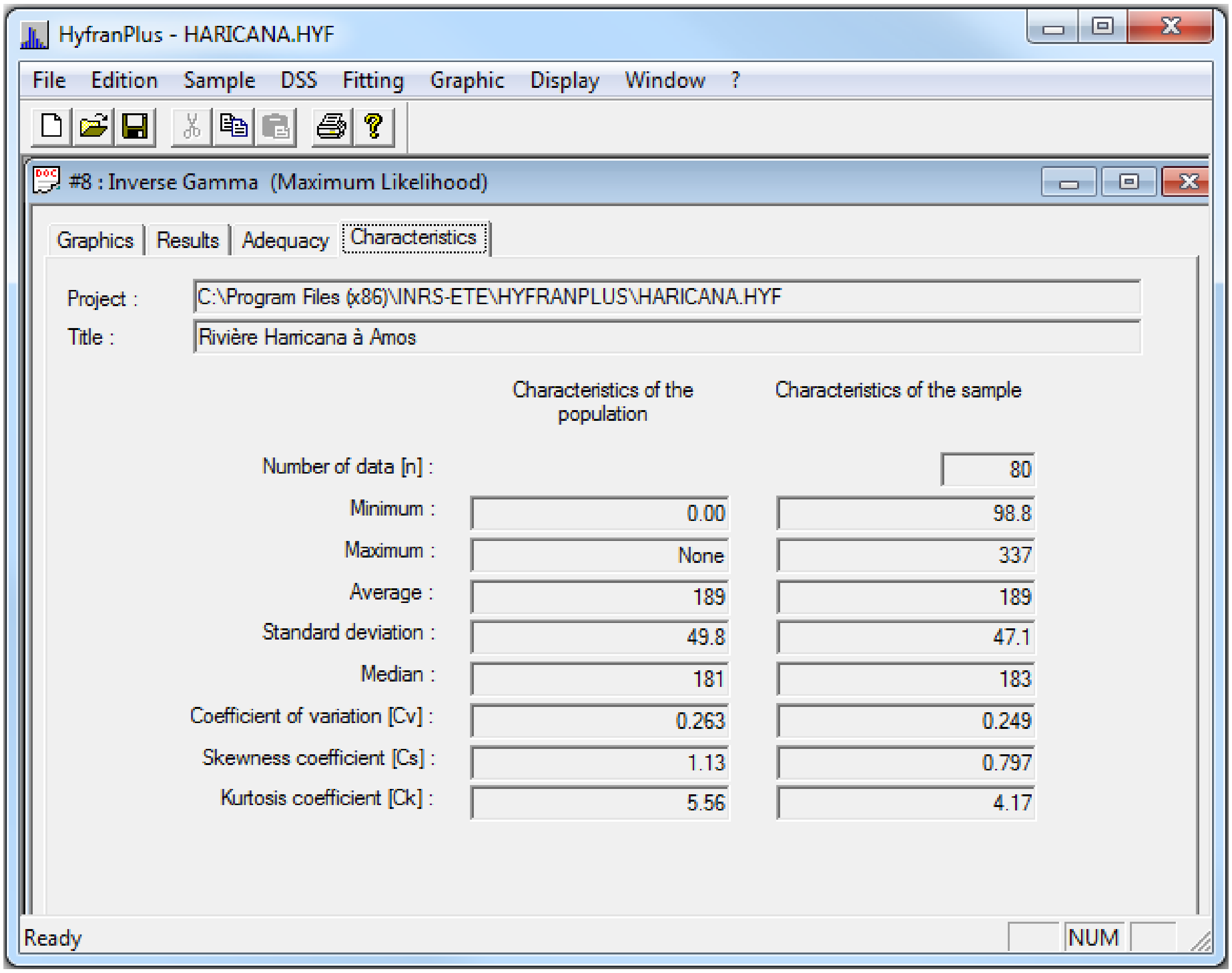


Figure 51: Characteristics for the fitting of dataset 1 using an Inverse Gamma distribution

Discordance

This option allows you to check in the case of Normal or Log-normal distribution, if the sample contains outliers i.e. the observations do not seem to come from the distribution used. To test this possibility, we use the test of Grubbs-Beck to detect unusual data suitable for normal and log-normal distributions (Bobée and El Adlouni, 2015).

Note: The majority of discordance tests are based on the assumption of normality (or log-normality by logarithmic transformation).

There are two tests, one for the smallest value and one for the largest value. We therefore check whether the smallest or the largest observation is discordant with the model proposed (Normal or Log-normal).

In the "Discordance" tab, we find the following information (Figure 52):
- The choice of the type of test: the smallest or the largest observation.
- The project i.e. the name of the file that contains the sample and its path
- The project title
- The null hypothesis (H_0) and the alternative hypothesis (H_1)
- Test results: found in this context the value of the test statistic and p-value of the statistic.
- Finally, the conclusion of the test i.e. acceptance or rejection of the null hypothesis at a significance level of 5% or 1%.

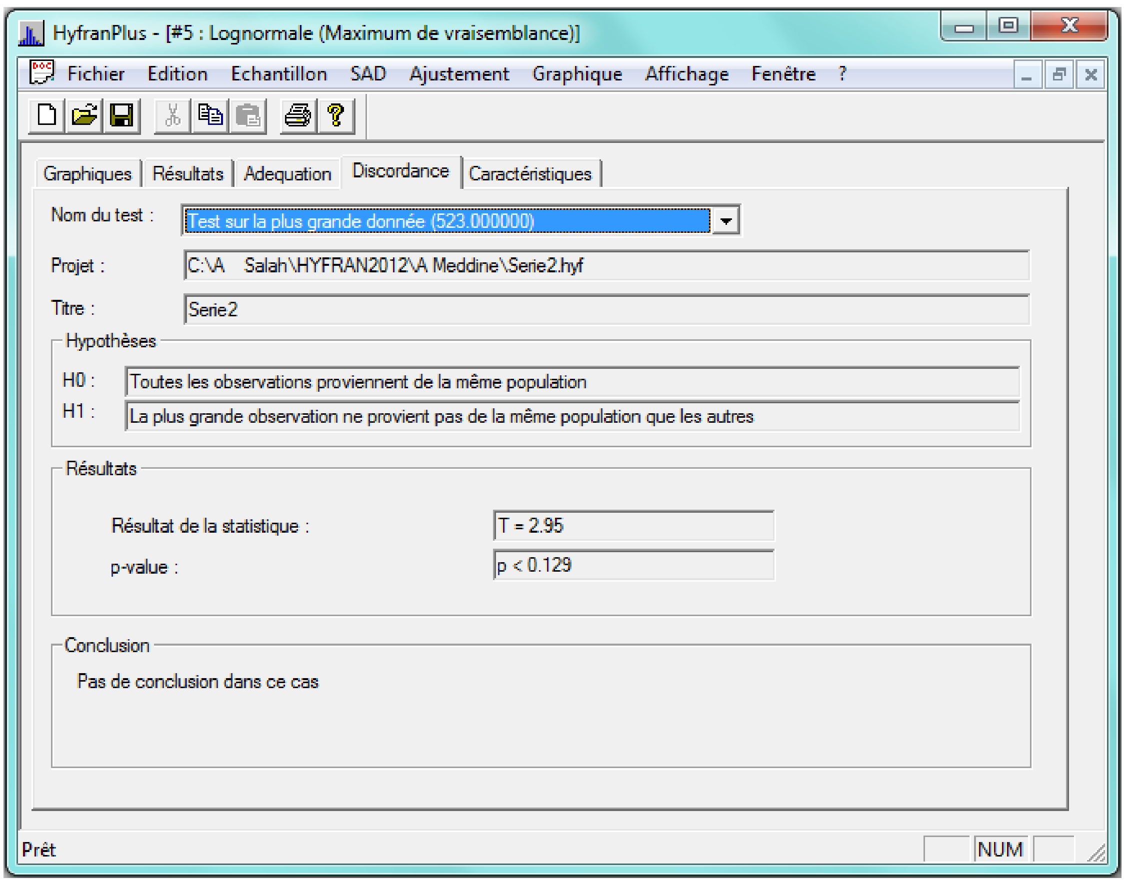


Figure 52: Discordance for log-normal fit of the largest observation (series 2)

Illustration through the fit of a log-normal distribution to series 2

When fitting a log-normal distribution to the series 2, we note that the largest observation is on the line corresponding to the upper limit of the quantile confidence interval (Figure 53).

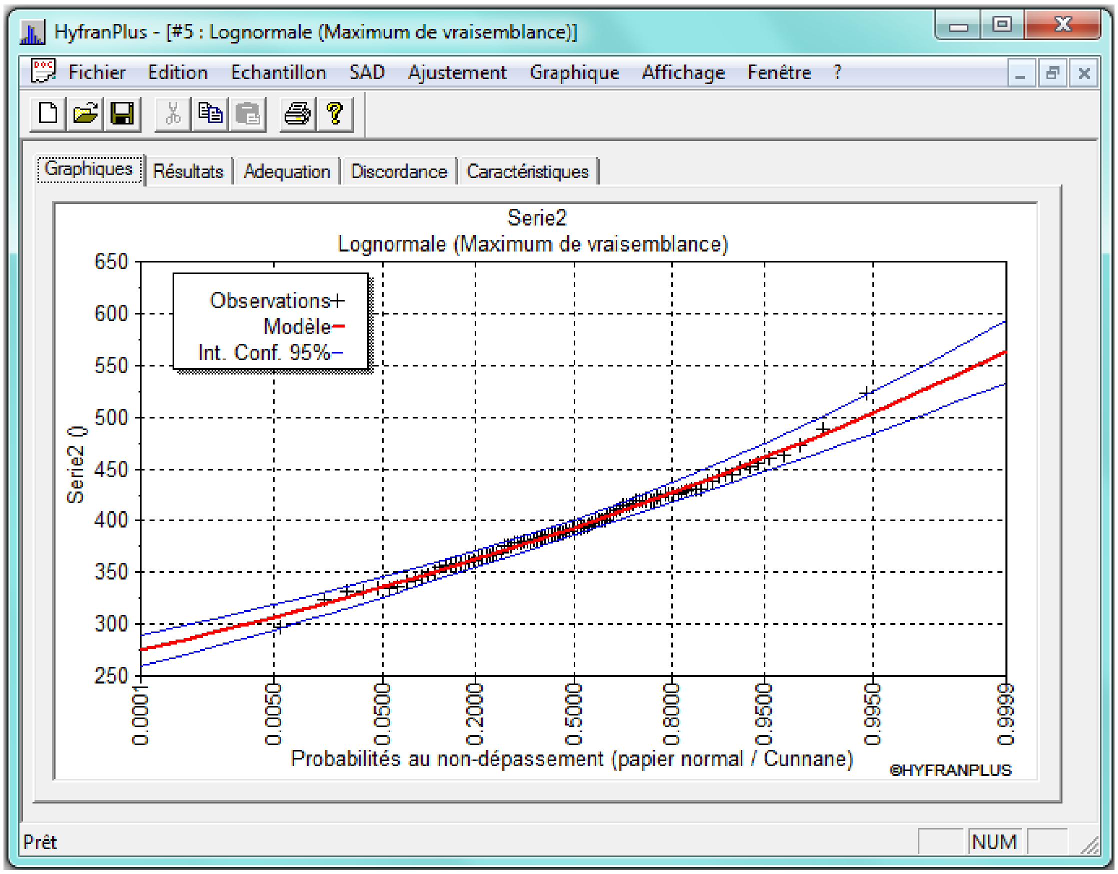


Figure 53: Fit of the log-normal distribution to the series 2 for the detection of discordant values under the assumption of log-normality

The application of the discordance test, to the series 2, shows that when considering the entire series, the test does not allow to conclude on the log-normality of the entire series and thus cannot confirm if the largest observation 523 belongs to the log-normal population (Figure 52). Indeed, in this case (Bobée and El Adlouni, 2015) we known only (Figure 52) that the p-value of the statistic t (t = 2.95) is less than 0.129; but it may then be higher or lower than the critical value of 5%.

When adding a new value (575) corresponding to 10% more than the highest value (523), the use of "Discordance" command shows that this new value is discordant with the log-normality hypothesis. Indeed, in this case (Figure 54), the p-value corresponding to t = 3.64 is such that p is below 0.008 which is smaller than the Type I error of 0.05.

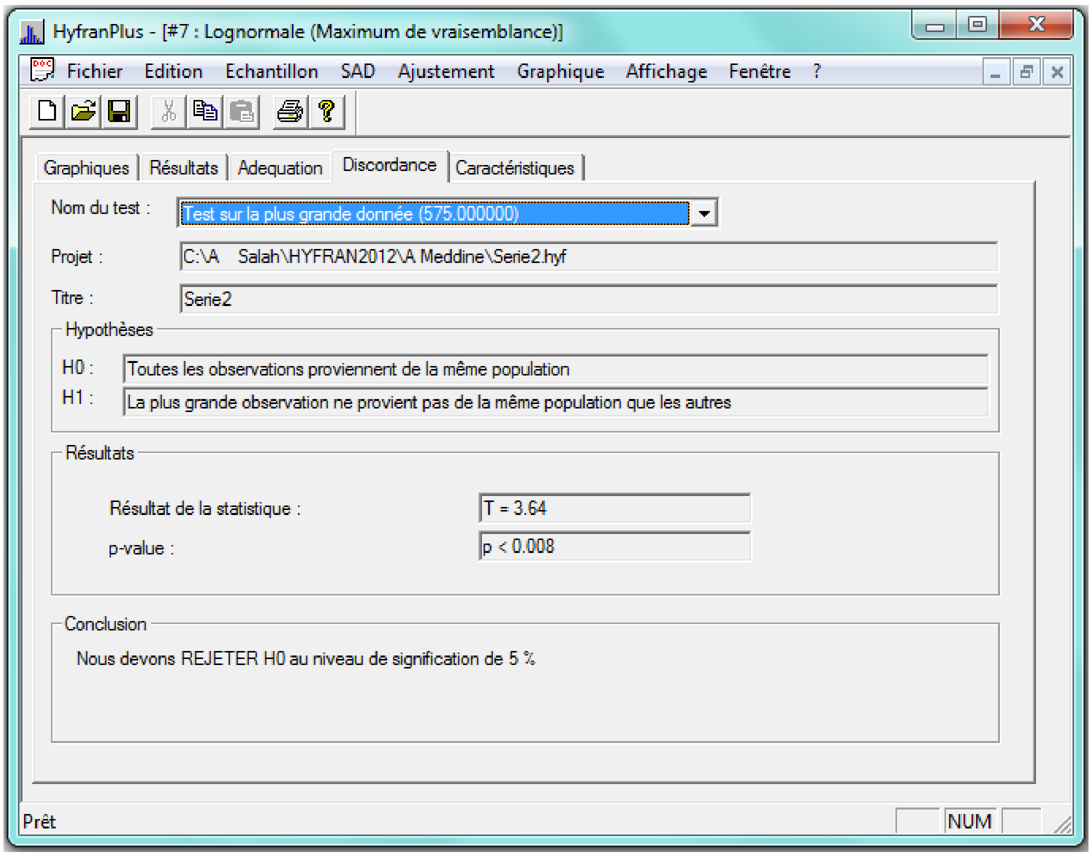


Figure 54 : Discordance for the fit of the log-normal distribution to series 2 when adding a new largest value

Figure 54 shows the result of this test to the new largest observation. The conclusion of the test is the rejection of the hypothesis "H0: All observations come from the same population." In conclusion, the test clearly shows (with a p-value less than 0.008) than the value 575 is not part of the population under the assumption of log-normality.

Note: When an outlier is detected, one must check if it is:

a. A real outlier (i.e. measurement error or input) that it should be eliminated or ,
b. A true and very important extreme value to keep. This validation should be performed from the hydro-meteorological context.

References

*Bobée B. et S. El Adlouni, (2015). Éléments d'Analyse Fréquentielle. Institut National de la Recherche Scientifique (INRS-ETE), Under review.

Bobee, B., et F. Ashkar (1991). The Gamma Family and Derived Distributions Applied in Hydrology, Water Resources Publications, Littleton, Colorado.

*Compaoré C., S. El Adlouni et B. Bobée (2013). Analyse fréquentielle en hydrologie : Logiciel HYFRAN-PLUS. Stage-Coop de l'Université de Moncton.

Ehsanzadeh E., S. El Adlouni et B. Bobée (2010). Frequency analysis incorporating a Decision Support System (DSS) for hydro-climatic variables. Journal of Hydrologic Engineering, Vol 15:11, pp. 861-881.

*El Adlouni, S. et B. Bobee (2011). Decision Support System for Flood Risk Assessment. Volume 1-2, p. 11-26.

*El Adlouni, S., B. Bobée et T.B.M.J. Ouarda (2008). On the tails of extreme event distributions. Journal of Hydrology, 355, 16-33.

*El Adlouni S., B. Bobée et O. Samoud (2012). Guide pour les méthodes du système d'aide à la décision (SAD) DE HYFRAN-PLUS. Water Resources Publication (http://www.wrpllc.com/books/HyfranPlus),

*Martel, B., S. El Adlouni et B. Bobée (2012). Comparison of the power of Log-Normality tests with different right tail alternative distributions. Journal of Hydrologic Engineering, (ASCE). 18(1), 1–9. doi: 10.1061/(ASCE)HE.1943-5584.0000595.

Morlat (1956). Les lois de probabilité de Halphen, Revue de statistiques appliquée, 4(3) : pp 21-46.

Perreault L., B. Bobée et PF Rasmussen (1999). Halphen distribution system. I: Mathematical and statistical properties. Journal of Hydrologic Engineering 4 (3), 189-199.

ANNEXE 1: DATASET

We present here the datasets 1, 2 and 3 used for the tutorial HYFRAN-PLUS.
It should be noted that the dataset 1 is the default dataset of Haricana project contained in the HYFRAN-PLUS software application example.
Datasets 2 and 3 are simulated data from Matlab software.

Dataset 1			
Observation	Empirical probability	Observation	Empirical probability
122	0.0698	167	0.3192
244	0.9052	179	0.4564
214	0.7431	185	0.5686
173	0.3815	117	0.0449
229	0.7930	192	0.6185
156	0.1945	337	0.9925
212	0.7307	125	0.0823
263	0.9551	166	0.3067
146	0.1322	99.1	0.0200
183	0.5062	202	0.6683
161	0.2195	230	0.8180
205	0.7057	158	0.2070
135	0.1072	262	0.9426
331	0.9800	154	0.1696
225	0.7805	164	0.2818
174	0.4190	182	0.4813
98.8	0.0075	164	0.2943
149	0.1446	183	0.5312
238	0.8678	171	0.3441
262	0.9302	250	0.9177
132	0.0948	184	0.5436
235	0.8429	205	0.7182
216	0.7556	237	0.8554
240	0.8928	177	0.4439
230	0.8055	239	0.8803
192	0.6060	187	0.5935
195	0.6434	180	0.4688
172	0.3566	173	0.4065
173	0.3940	174	0.4314

172	0.3691	167	0.3317
153	0.1571	185	0.5810
142	0.1197	232	0.8304
317	0.9676	100	0.0324
161	0.2319	163	0.2569
201	0.6559	203	0.6808
204	0.6933	219	0.7681
194	0.6309	182	0.4938
164	0.2693	184	0.5561
183	0.5187	118	0.0574
161	0.2444	155	0.1820

Table A.1: Observed dataset 1

Dataset 2			
Observation	Empirical probability	Observation	Empirical probability
426	0.8044	350	0.1158
346	0.0958	375	0.2854
395	0.5449	452	0.9341
384	0.3952	428	0.8244
419	0.7046	355	0.1257
420	0.7545	324	0.0160
420	0.7246	381	0.3653
389	0.4750	412	0.6547
380	0.3353	443	0.8942
380	0.3553	407	0.6347
439	0.8842	361	0.1956
335	0.0459	416	0.6946
395	0.5349	359	0.1557
414	0.6747	366	0.2355
378	0.3154	378	0.3054
423	0.7745	357	0.1457
401	0.5948	393	0.5150
367	0.2455	369	0.2655
410	0.6447	392	0.5050
393	0.5250	385	0.4052
387	0.4351	387	0.4451
376	0.2954	384	0.3852
406	0.6248	456	0.9441
335	0.0559	429	0.8343
388	0.4551	406	0.6148
382	0.3752	348	0.1058
368	0.2555	343	0.0858
431	0.8643	331	0.0359
375	0.2754	523	0.9940
426	0.7844	445	0.9042
445	0.9142	414	0.6647
397	0.5649	366	0.2255
415	0.6846	360	0.1756
430	0.8543	426	0.8144

400	0.5848	472	0.9741
426	0.7944	341	0.0758
356	0.1357	386	0.4152
451	0.9242	400	0.5749
365	0.2156	331	0.0259
336	0.0659	439	0.8743
463	0.9641	387	0.4251
401	0.6048	488	0.9840
422	0.7645	388	0.4651
389	0.4850	420	0.7445
364	0.2056	360	0.1856
297	0.0060	379	0.3253
430	0.8443	359	0.1657
392	0.4950	420	0.7146
396	0.5549	460	0.9541
420	0.7345	380	0.3453

Tableau A.2: Observed dataset 2

Dataset 2-Transformed			
Observation	Transformed Probability	Observation	Transformed Probability
6.05	0.8044	5.86	0.1158
5.85	0.0958	5.93	0.2854
5.98	0.5449	6.11	0.9341
5.95	0.3952	6.06	0.8244
6.04	0.7046	5.87	0.1257
6.04	0.7545	5.78	0.0160
6.04	0.7246	5.94	0.3653
5.96	0.4750	6.02	0.6547
5.94	0.3353	6.09	0.8942
5.94	0.3553	6.01	0.6347
6.08	0.8842	5.89	0.1956
5.81	0.0459	6.03	0.6946
5.98	0.5349	5.88	0.1557
6.03	0.6747	5.9	0.2355
5.93	0.3154	5.93	0.3054
6.05	0.7745	5.88	0.1457
5.99	0.5948	5.97	0.5150
5.91	0.2455	5.91	0.2655
6.02	0.6447	5.97	0.5050
5.97	0.5250	5.95	0.4052
5.96	0.4351	5.96	0.4451
5.93	0.2954	5.95	0.3852
6.01	0.6248	6.12	0.9441
5.81	0.0559	6.06	0.8343
5.96	0.4551	6.01	0.6148
5.95	0.3752	5.85	0.1058
5.91	0.2555	5.84	0.0858
6.07	0.8643	5.8	0.0359
5.93	0.2754	6.26	0.9940
6.05	0.7844	6.1	0.9042
6.1	0.9142	6.03	0.6647
5.98	0.5649	5.9	0.2255
6.03	0.6846	5.89	0.1756
6.06	0.8543	6.05	0.8144
5.99	0.5848	6.16	0.9741

6.05	0.7944	5.83	0.0758
5.87	0.1357	5.96	0.4152
6.11	0.9242	5.99	0.5749
5.9	0.2156	5.8	0.0259
5.82	0.0659	6.08	0.8743
6.14	0.9641	5.96	0.4251
5.99	0.6048	6.19	0.9840
6.05	0.7645	5.96	0.4651
5.96	0.4850	6.04	0.7445
5.9	0.2056	5.89	0.1856
5.69	0.0060	5.94	0.3253
6.06	0.8443	5.88	0.1657
5.97	0.4950	6.04	0.7146
5.98	0.5549	6.13	0.9541
6.04	0.7345	5.94	0.3453

Tableau A.3: Logarithmic transformation of the dataset 2 (Normaly distributed)

Dataset 3			
Observation	Empirical probability	Observation	Empirical probability
494	0.7146	572	0.8443
330	0.2754	580	0.8643
358	0.3154	467	0.6547
330	0.2655	549	0.8144
479	0.6946	524	0.7844
319	0.2255	360	0.3253
293	0.1357	459	0.6347
476	0.6747	369	0.3653
355	0.3054	308	0.1756
450	0.6148	488	0.7046
363	0.3553	412	0.4651
437	0.5649	381	0.4052
187	0.0060	440	0.5848
464	0.6447	543	0.8044
704	0.9341	454	0.6248
551	0.8244	428	0.5349
530	0.7944	439	0.5749
375	0.3952	410	0.4551
426	0.5250	320	0.2355
434	0.5549	511	0.7445
415	0.4850	616	0.8942
314	0.2056	344	0.2854
363	0.3453	719	0.9441
383	0.4251	383	0.4152
349	0.2954	375	0.3852
519	0.7645	308	0.1856
864	0.9840	318	0.2156
271	0.0958	303	0.1657
419	0.4950	449	0.6048

262	0.0758	651	0.9142
596	0.8743	473	0.6647
616	0.9042	521	0.7745
245	0.0359	603	0.8842
723	0.9541	191	0.0160
291	0.1257	384	0.4351
432	0.5449	449	0.5948
561	0.8343	500	0.7246
422	0.5150	273	0.1058
803	0.9741	298	0.1457
921	0.9940	511	0.7545
421	0.5050	390	0.4451
228	0.0259	476	0.6846
362	0.3353	302	0.1557
313	0.1956	247	0.0459
321	0.2455	507	0.7345
288	0.1158	577	0.8543
413	0.4750	658	0.9242
759	0.9641	255	0.0559
325	0.2555	270	0.0858
371	0.3752	259	0.0659

Tableau A.4: Observed dataset 3, simulated from Gamma distribution

Appendix A: Asymptotic Confidence intervals for quantiles

In Hydrological Frequency Analysis (HFA) a statistical distribution D is fitted to the annual maximum flow data using a method M , in order to get the estimation of $\hat{\underline{\theta}}$ of the the vector of the parameters $\underline{\theta}$, and then estimate the quantile: $\hat{X}_T = F^{-1}\left(1-1/T; \hat{\underline{\theta}}\right)$.

Note that the estimators of the parameters and the quantile, $\hat{\underline{\theta}}$ et $\hat{X}_T$, are random variables.

The distribution of the quantile estimator $\hat{X}_T$ is, in general, unknown except for some distributions such as exponential, normal or log-normal. However, for large sample size N , the quantile estimator $\hat{X}_T$ is asymptotically normal with:

- Mean X_T (the true unknown value)
- With variance $\operatorname{var}\hat{X}_T$. Thus

$$\hat{X}_T \sim N\left(X_T; \operatorname{var}\hat{X}_T\right) \Rightarrow u = \frac{\hat{X}_T - X_T}{\sqrt{\operatorname{var}\hat{X}_T}} \sim N(0,1)$$

For a given distribution $\operatorname{var}\hat{X}_T$ depends on the estimation method used (see Appendix D, Bobée and Ashkar (1991)) where we can find details of calculus for the MV and MM methods. We can deduce the confidence interval with a confidence level $(1-\alpha)$ of the true value from the equation E.1. We have (see Appendix E, Bobée and Ashkar (1991)) :

$$P\left[-u_{\alpha/2} \le \frac{\hat{X}_T - X_T}{\sqrt{\operatorname{var}\hat{X}_T}} \le u_{\alpha/2}\right] = 1-\alpha$$

Or

$$P\left[\hat{X}_T - u_{\alpha/2}\sqrt{\operatorname{var}\hat{X}_T} \le X_T \le \hat{X}_T + u_{\alpha/2}\sqrt{\operatorname{var}\hat{X}_T}\right] = 1-\alpha .$$

$u_{\alpha/2}$ is the quantile of the standard normal distribution corresponding to the probability of exceedance $\alpha/2$.

Appendix B : Distribution Classification

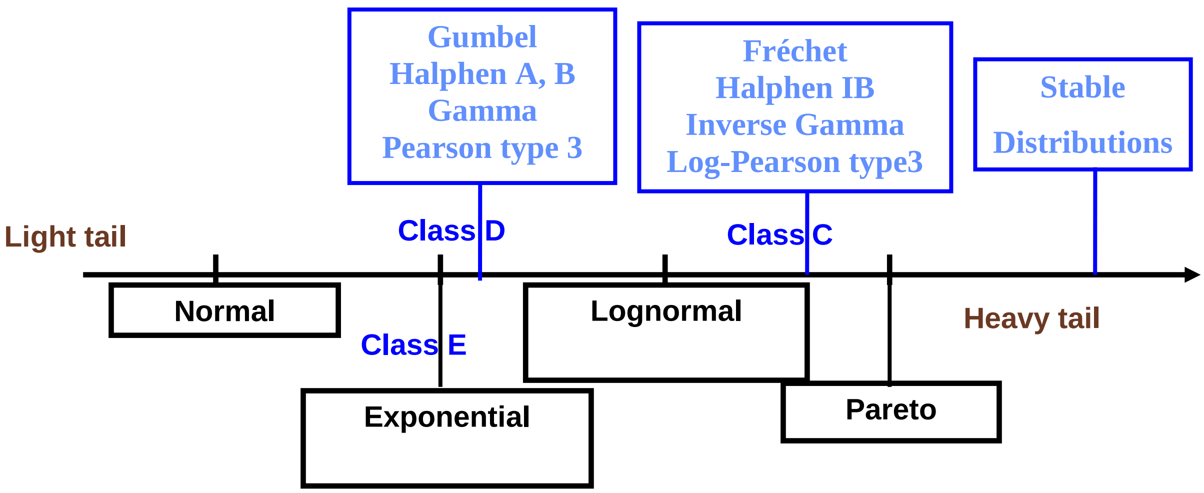


Figure B-1: Classification of statistical distribution with respect to their right tail behaviour (From El Adlouni, Bobée et Ouarda, 2008).