

WATERBOUWKUNDIG LABORATORIUM



Flanders
State of
the Art

00_144_4
FHR reports

Scientific Assistance towards a Probabilistic Formulation of Hydraulic Boundary Conditions

Copula Analysis software tool Manual

DEPARTMENT
MOBILITY &
PUBLIC
WORKS

www.flandershydraulicsresearch.be

Scientific Assistance towards a Probabilistic Formulation of Hydraulic Boundary Conditions

Copula Analysis Reference Guide

Leyssen, G.; Blanckaert, J.; Pereira, F.; Nossent, J.; Mostaert, F.

Legal notice

Flanders Hydraulics Research is of the opinion that the information and positions in this report are substantiated by the available data and knowledge at the time of writing.
 The positions taken in this report are those of Flanders Hydraulics Research and do not reflect necessarily the opinion of the Government of Flanders or any of its institutions.
 Flanders Hydraulics Research nor any person or company acting on behalf of Flanders Hydraulics Research is responsible for any loss or damage arising from the use of the information in this report.

Copyright and citation

© The Government of Flanders, Department of Mobility and Public Works, Flanders Hydraulics Research 2019
 D/2017/3241/29

This publication should be cited as follows:

Leyssen, G.; Blanckaert, J.; Pereira, F.; Nossent, J.; Mostaert, F. (2019). Scientific Assistance towards a Probabilistic Formulation of Hydraulic Boundary Conditions: Copula Analysis Reference Guide. Version 3.0. FHR Reports, 00_144_4. Flanders Hydraulics Research & IMDC: Antwerp.

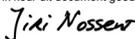
Reproduction of and reference to this publication is authorised provided the source is acknowledged correctly.

Document identification

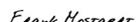
Customer:	Flanders Hydraulics Research	Ref.:	WL2019R 00_144_4
Keywords (3-5):	Extreme Value; Flooding ;Flood risk		
Text (p.):	19	Appendices (p.):	/
Confidentiality:	<input checked="" type="checkbox"/> No	<input checked="" type="checkbox"/> Available online	

Author(s):	Leyssen, G.; Blanckaert, J.
------------	-----------------------------

Control

	Name	Signature
Reviser(s):	Pereira, F.; Nossent, J.	 <small>Getekend door: Jiri Nossent (Signature) Getekend op: 2019-11-29 15:46:12 +01:00 Reden: Ik keur dit document goed</small> 
Project leader:	Pereira, F.	<small>Getekend door: Fernando Pereira (Signature) Getekend op: 2019-11-29 15:26:32 +01:00 Reden: Ik keur dit document goed</small> 

Approval

Head of Division:	Mostaert, F.	<small>Getekend door: Frank Mostaert (Signature) Getekend op: 2019-11-29 15:21:42 +01:00 Reden: Ik keur dit document goed</small> 
-------------------	--------------	--



Abstract

The Copula Analysis tool is a Matlab® software packet developed in commission of Flanders Hydraulic Research (FHR). The tool is a standalone executable which facilitates and automates the fit of a bivariate Copula and the selection of synthetic quantiles with corresponding joint frequencies. The tool is part of a suite of software tools to facilitate the probabilistic formulation of hydraulic boundary conditions. This report is a manual for the tool. It assists the user when installing the tool and gives an overview of the workflow. Furthermore it provides an in depth description of the workflow and the functionalities of the tool.

Contents

Abstract	III
Contents	V
List of figures	VI
1 Introduction.....	1
2 Software	2
2.1 Installation of MCR	2
3 Workflow	3
4 Synthetic extremes by stratification.....	4
5 Couples	6
5.1 Input	6
5.2 Couples	7
6 Copula.....	12
6.1 Definition of classes.....	12
6.2 Copula fitting	13
6.3 Output	15
7 Extra and info functions.....	18
8 References.....	19

List of figures

Figure 1-1: Overview of reports, tools and manuals 1

Figure 3-1: Workflow of the Copula Analysis tool 3

Figure 4-1: Area of occurrence of couples and uncoupled POT 4

Figure 4-2: Stratified Area of occurrence of couples and uncoupled POT 5

Figure 5-1: Input of the Copula Analysis tool 6

Figure 5-2: Input metadata dialog box 7

Figure 5-3: Control panel for the automated selection of couples 8

Figure 5-4: Example of the inspection of the input 8

Figure 5-5: Manual couples selection GUI 9

Figure 5-6: Distribution and couples check 10

Figure 5-7: Time differences between the POT values in the couples 11

Figure 5-8: Selected couples and uncoupled POT 11

Figure 6-1: Definition of the classes control panel 12

Figure 6-2: Definition of the class boundaries dialog boxes 13

Figure 6-3: Available Copulas in the Copula Analysis tool 14

Figure 6-4: Copula fitting parameters 14

Figure 6-5: Output menu 16

Figure 6-6: Dialog box with names of the output files 16

Figure 6-7: Synthetic extremes of variables 1 and 2 17

Figure 7-1: Extra and Info tab 18

1 Introduction

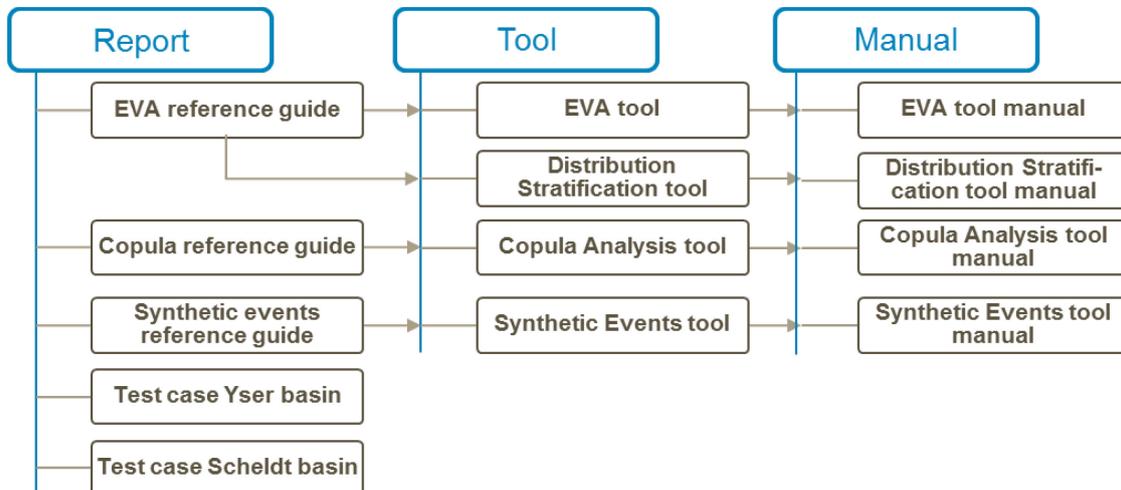
The Copula Analysis tool is a Matlab® software packet developed in commission of Flanders Hydraulic Research (FHR). The tool is a standalone executable which facilitates and automates the fit of a bivariate Copula and the selection of synthetic quantiles with corresponding joint frequencies. These joint frequencies give the frequency of the joint occurrence of a value X and Y of two dependent variables like for example a wind speed of 30 m/s and a storm surge of 1.5 m. To determine a Copula it is of great importance to select the appropriate couples. These couples will be the basis of the entire analysis.

The tool is part of a suite of software tools to facilitate the probabilistic formulation of hydraulic boundary conditions. An overview of the tools and corresponding reports and manuals is presented in Figure 1-1.

The tool has to be used in combination with the Extreme Value Analysis tool. The input for the tool are 2 files created by the Extreme Value Analysis tool, i.e. one for every variable, with the time series, the POT (peak over threshold) and the appropriate distribution.

The Copula Analysis reference guide gives an overview of the methodology and a summary of the applied formulae. The overview is based on an extensive literature review of the standard Copula literature (Nelsen, 2006; Genest, 2007) in combination with the standard Extreme Values literature (Coles, 2001; Kotz, 2000; Beirlant, 2004).

Figure 1-1: Overview of reports, tools and manuals



2 Software

The Copula Analysis tool has been developed by IMDC in a Matlab® environment and compiled into an executable, so there is no software license required to use the toolbox. The tool consists of three main visual interfaces which give access to numerous of Matlab® functions.

The user needs to install the Matlab Compiler Runtime (MCR) before the first execution of the EVA-tool. The MCR is a Matlab® copy without the graphical interface that can be deployed royalty free and possesses all the strengths of the full Matlab® environment. You must have administrative privileges to install the MCR on a target machine since it modifies both the system registry and the system path. Running the MCRInstaller after the MCR has been set up on the target machine requires only user-level privileges.

2.1 Installation of MCR

The installation of the MCR Installer is guided by an installation GUI which requires the following steps. The latest version of the MCR can be downloaded at the Matlab site (<http://www.mathworks.nl/products/compiler/mcr/index.html>)

- When the MCR Installer wizard appears, click Next to begin the installation. Click Next to continue.
- In the Select Installation Folder dialog box, specify the location where you want to install the MCR and whether you want to install the MCR for just yourself or others. Click Next to continue.
- Confirm your selections by clicking Next.
- The installation begins. The process takes some time due to the amount of files that are installed.

A more detailed explanation of the MCR can be found on Matlab (2011).

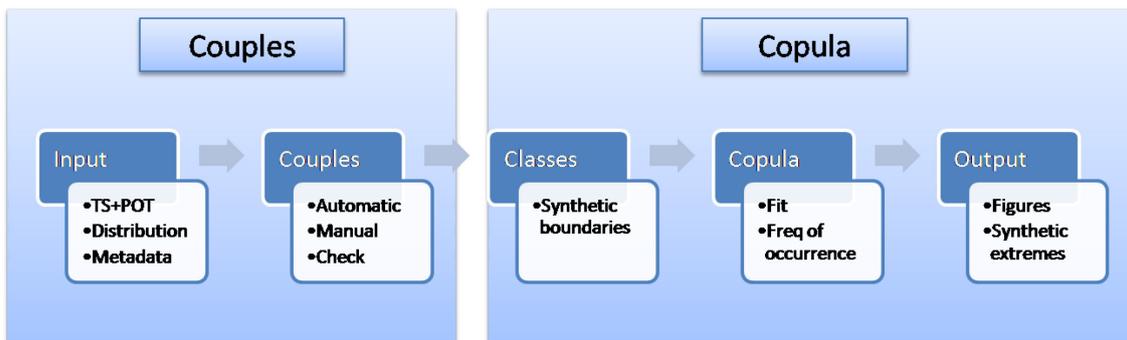
3 Workflow

The workflow of the Copula Analysis tool exists of 2 main parts, i.e. the selection and inspection of the couples and the actual fitting of the Copula with the creation of synthetic events (Figure 3-1). The files generated by the EVA-tool contain time series, POT values and a distribution.

Two of these files are necessary to fit a Copula and create synthetic extremes. The time series need to have a common time range from which the couples can be selected. This selection can be done automatically based on the time of occurrence of the POT but it is advised to manually check the automated selection. A special window is developed to facilitate this manual check.

Once the couples are selected the user has to define the boundaries of the classes used in the stratification. The chosen Copula will be fitted through the selected couples by a negative log Likelihood algorithm. The goodness of fit of the Copula is controlled by comparison with the data and the empirical Copula. In the end all the figures can be exported in png format and the synthetic extremes with yearly frequency of joint occurrences are exported. These can be combined with a unity profile to create synthetic events which serve as boundary conditions.

Figure 3-1: Workflow of the Copula Analysis tool

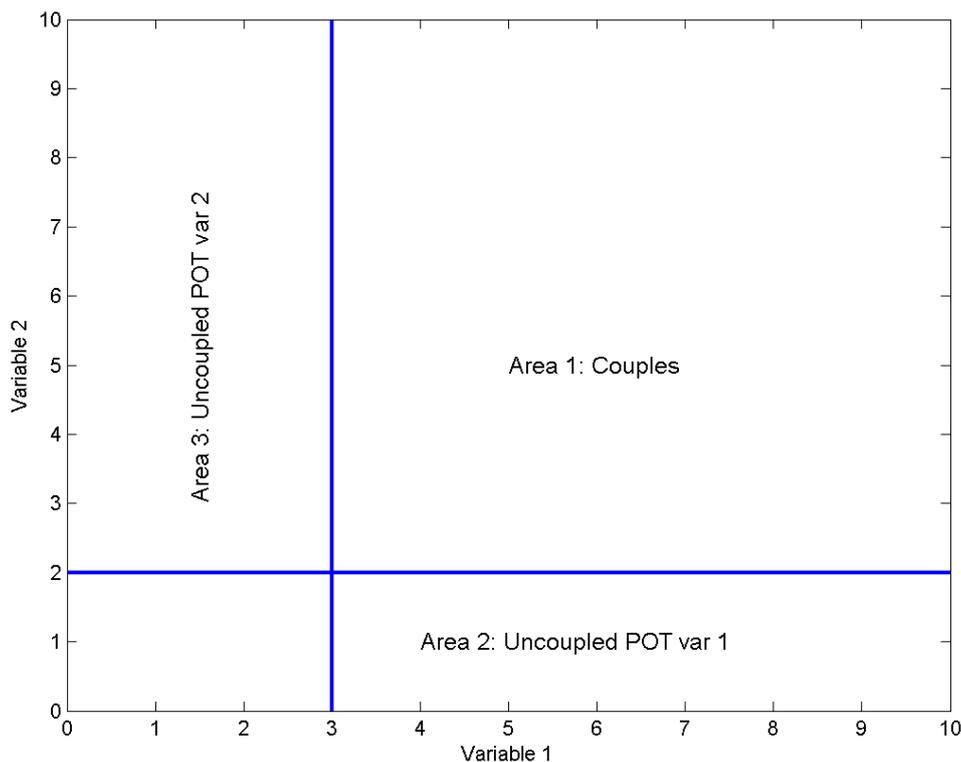


4 Synthetic extremes by stratification

The Copula Analysis tool combines 2 univariate distributions created by the EVA-tool into a multivariate distribution. The Copula theory is summarized in the Copula Reference Guide (see Figure 1-1), but the combination of the Copulas with stratification and synthetic extremes needs some more clarification.

Due to their specific properties Copulas are very useful to parameterize the dependence of dependent variables. To estimate the dependence one needs couples of simultaneous values. Because the univariate conditional distributions of the EVA-tool are only valid above their threshold, only the simultaneous extremes that exceed their corresponding threshold can be used as couples. This will obviously leave some extremes without a corresponding extreme of the other variable. The greater the dependence between the two investigated variables, the smaller this number of uncoupled POT. Figure 4-1 gives the areas in which the couples and the uncoupled POT's of variable 1 and 2 occur.

Figure 4-1: Area of occurrence of couples and uncoupled POT



The exceedance probability of the joint occurrences in a certain point of area 1 can be calculated by the Copula. If we stratify variable 1 and variable 2 both into 10 parts we will get 100 cells and 121 Copula values. The yearly frequency of occurrence of every cell can be calculated by

$$Frq(i, j) = \frac{k_1}{A_1} (C(i + 1, j + 1) - C(i, j + 1) - C(i + 1, j) + C(i, j))$$

Where C is the joint cumulative density function given by the Copula, k_1 is the number of couples and A_1 is the number of years. The value of variable 1 and 2 in the middle of each class will be assigned to that class (red dots in Figure 4-2).

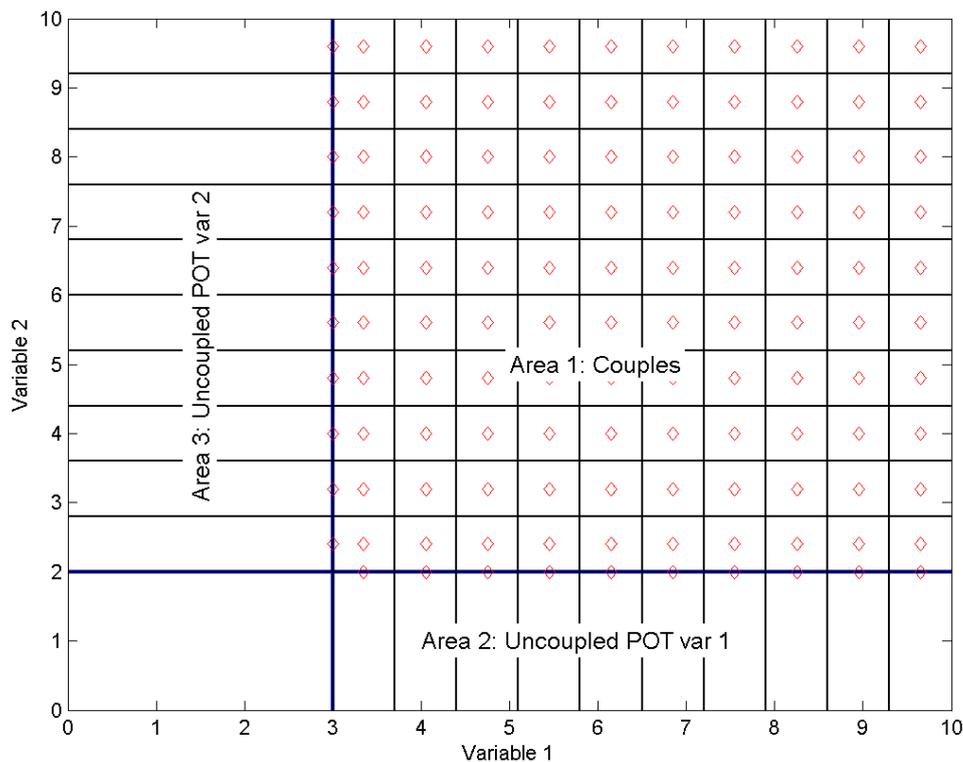
The yearly frequencies of occurrence of area 2 and 3 can be calculated by their univariate distributions.

$$Frq_{area2}(i) = \frac{k_2}{A_2} (CDF(i+1) - CDF(i))$$

$$Frq_{area3}(j) = \frac{k_3}{A_3} (CDF(j+1) - CDF(j))$$

Where k_2 , k_3 and A_2 , A_3 are the number of POT and years of data in the corresponding areas. By definition there is no corresponding value of variable 2 available for the uncoupled POT of variable 1, but we know it has to be a value lower than the threshold of variable 2. So it will be a secure assumption to assign the threshold value of variable 2 to the cells in area 1 together with the value of variable 1 in the middle of the class (red dots on the lowest horizontal blue line in Figure 4-2). The reasoning is similar for the uncoupled POT of variable 2 (red dots on the leftmost vertical blue line in Figure 4-2).

Figure 4-2: Stratified Area of occurrence of couples and uncoupled POT



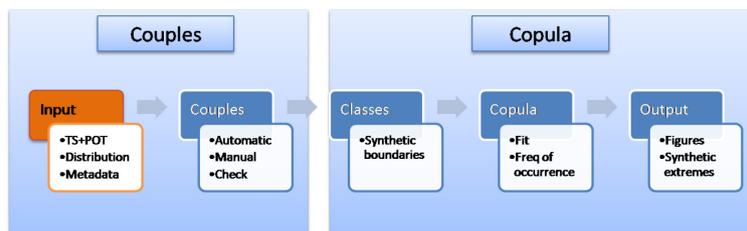
It is often preferred to have rounded strata values and so strata boundary values. Therefore the lower boundary of the strata will not be equal to the optimal threshold of the univariate extreme value distribution. The probabilities between the optimal threshold and the lowest strata are included in the probabilities of Area 2 and 3.

5 Couples

The Couples part of the Copula Analysis tool workflow comprises two steps, i.e.

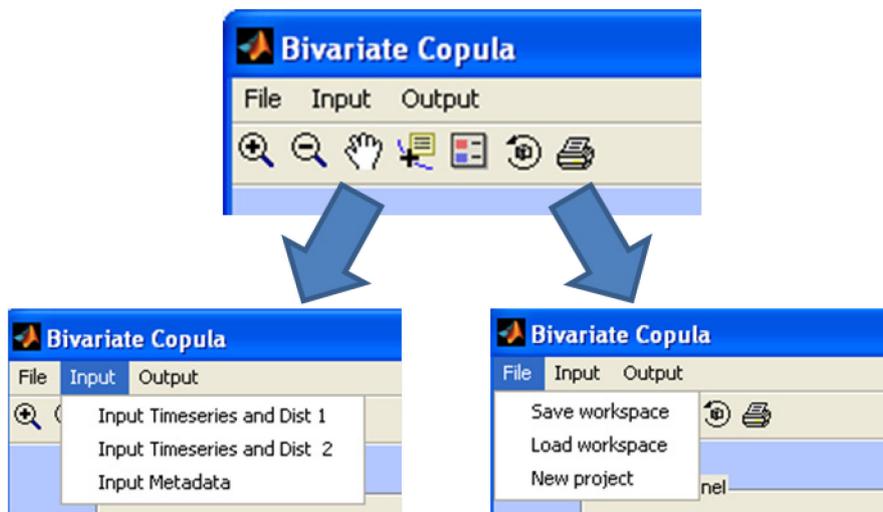
- the specification of the input
- the selection of the couples

5.1 Input



The input buttons of the Copula Analysis tool can be found in the upper right menu buttons, Input and File (Figure 5-1).

Figure 5-1: Input of the Copula Analysis tool



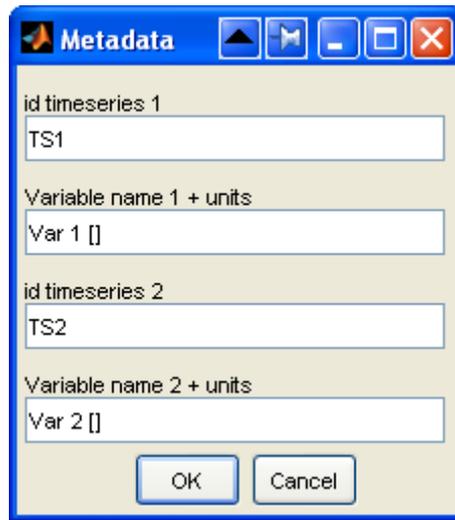
- Input time series and distribution:

The input of the Copula Analysis tool are mat files created by the EVA-tool. These files have a default name of 'TS_POT_Distr_“id”.mat' and contain the time series, the POT values, and the appropriate univariate distribution.

- Input metadata:

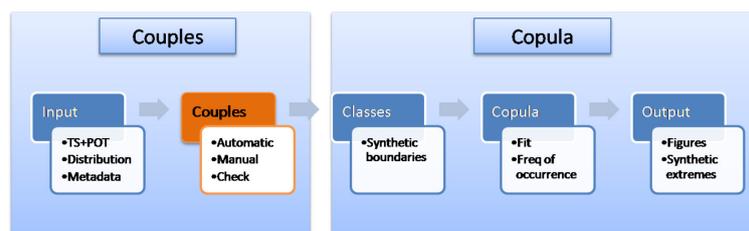
The input of the metadata is specified by means of the input dialog box in Figure 5-2. The Id of the 2 time series, for example the location of measurements, and the name of the variable with the units, for example discharge [m³/s]. These names will appear in the different figures and are interpreted by the Matlab editor which is similar to a TEX editor. So an underscore will give a subscript and a ^ will give a superscript.

Figure 5-2: Input metadata dialog box



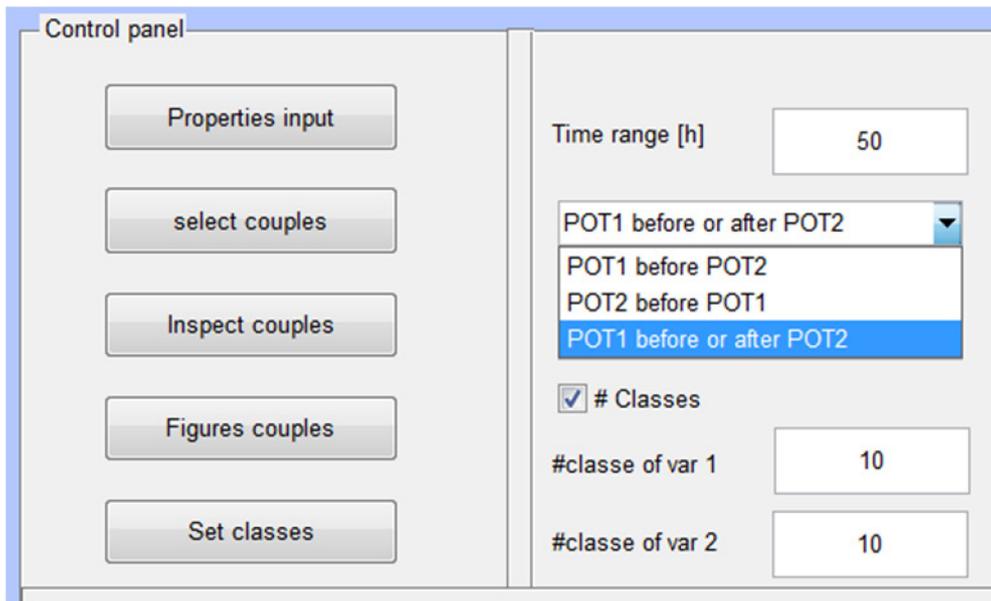
- Save and load workspace:
The entire workspace of the Copula Analysis tool can be saved during every step of the analysis and reloaded later.
- New project:
The new project button will clear the internal memory of the Copula Analysis tool and reset all the variables to their default values.

5.2 Couples



The couples are extreme values generated by joint events. They usually occur at almost the same time and will be used to determine the dependence between the two variables. An automated selection will be performed to link the POT values of the 2 variables into Couples based on their time of occurrence. The maximal time range between the two POT values of one couple is user input (Time range [h] in Figure 5-3). There are 3 selection modes available, “POT1 before POT2”, “POT2 before POT1” and “POT1 before or after POT2”.

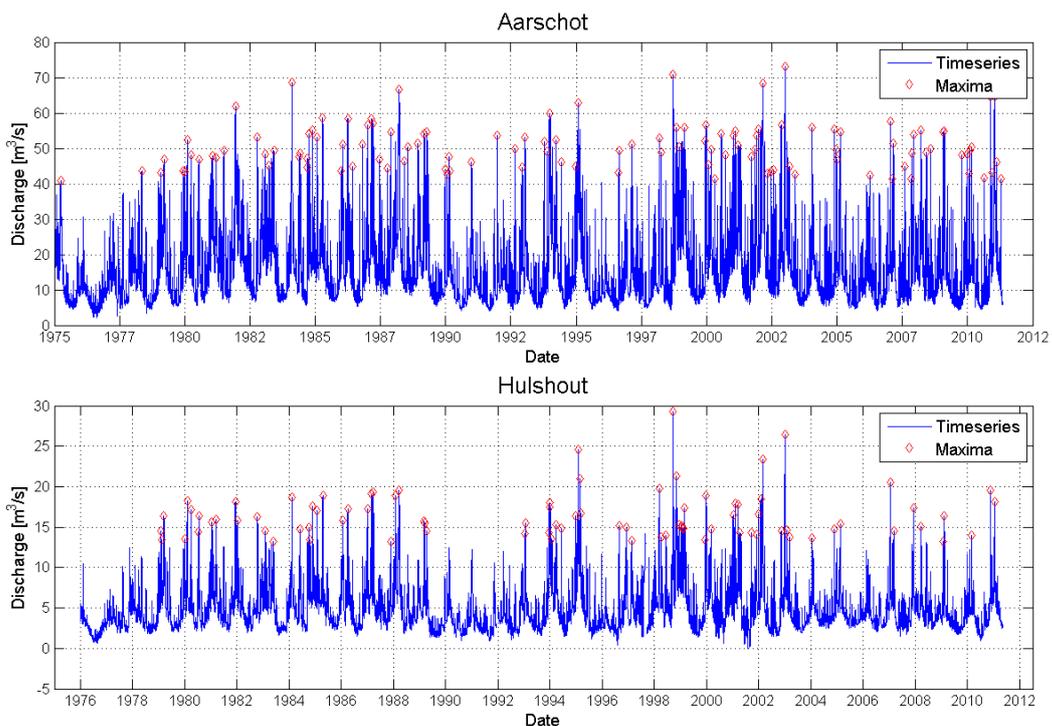
Figure 5-3: Control panel for the automated selection of couples



- Properties input:

A first step in the selection of the couples is the inspection of the input. The time series and the POT values of the two variables are visualized in 2 linked subplots (example in Figure 5-4). The x axes are linked. So by zooming on one of the two figures the x axis in the other figure will be zoomed correspondingly. It is advised to use the horizontal zoom (zoom – right click on the figure – zoom options – horizontal zoom (2-D plots only)).

Figure 5-4: Example of the inspection of the input



- Select couples**

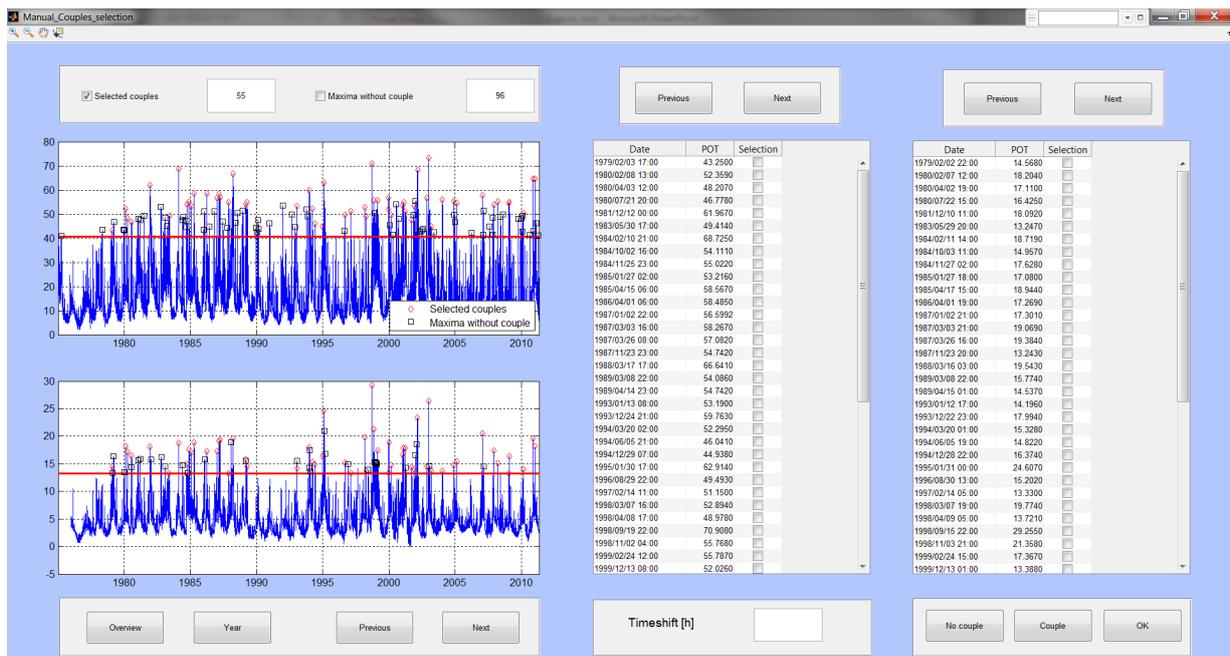
The automatic selection of the couples will use the time range and selection mode specified in the control panel in Figure 5-3 to link the POT values into couples. If there is more than one extreme of series Y that can be coupled with an extreme of series X, the one that's nearest in time is selected. The number of selected couples will be displayed in the log window in the lower left corner.
- Inspect couples:**

The automatically selected couples can be inspected in a sub GUI "Manual Couples selection" (Figure 5-5). This GUI will display the selected couples and the extremes without couple values in a figure and in 2 tables. The number of couples and uncoupled maxima are given in the upper left corner. The checkboxes "Selected couples" or "Maxima without couples" will enable the modification of respectively the couples or the maxima without couples. The entire time range can swiftly be checked by entering year mode (button "Year" under the figures). This displays one year of extremes and time series. The buttons "Previous" or "Next" will switch to the previous or next year. The button "Overview" will display the entire time range.
- Modify selected Couples:**

The selection of a couple can be done by clicking on a maximum in the figure or in the table. The corresponding maximum of the other variable will be automatically selected and a zoom will be performed. The time shift between the two maxima of the couples is displayed in the editable "Timeshift [h]". If the coupling is not correct it can be corrected by clicking on the button "No couple" in the lower right corner.
- Modify maxima without couple:**

The selection of an uncoupled maxima can be done by clicking on the maxima in the figure or in the table. An automated zoom will be applied on both figures. By selecting an appropriate couple maxima of the other variable the time shift between these two maxima is calculated. By clicking "Couple" in the lower right corner the two selected maxima are coupled. The button "OK" will close the Manual couples selection GUI and reenter the main Copula GUI.

Figure 5-5: Manual couples selection GUI



- Figures couples:
 - The button “Figures Couples” will generate three figures
 - a check of the distributions and the couples (Figure 5-6)
 - a fit of the time difference of the POT’s in a couple in function of the POT values (Figure 5-7)
 - a time plot of the couples and the uncoupled POT’s (Figure 5-8).
 These figures will be saved as png in the output directory.

Figure 5-6: Distribution and couples check

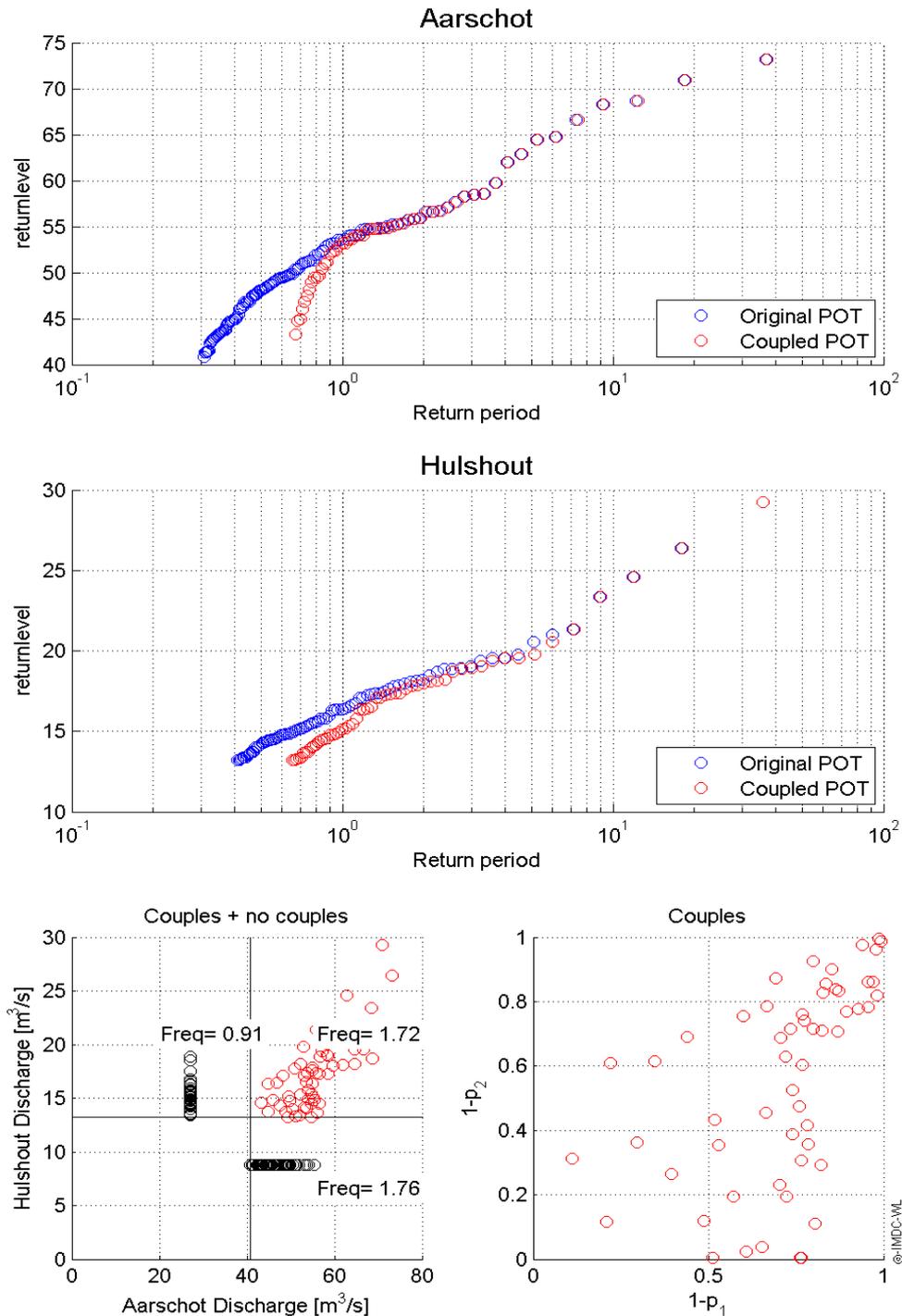


Figure 5-7: Time differences between the POT values in the couples

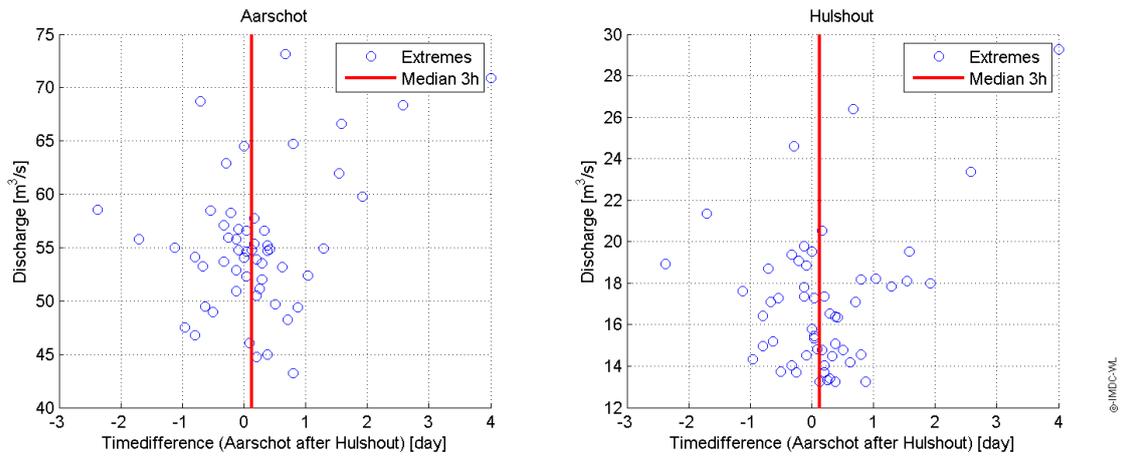
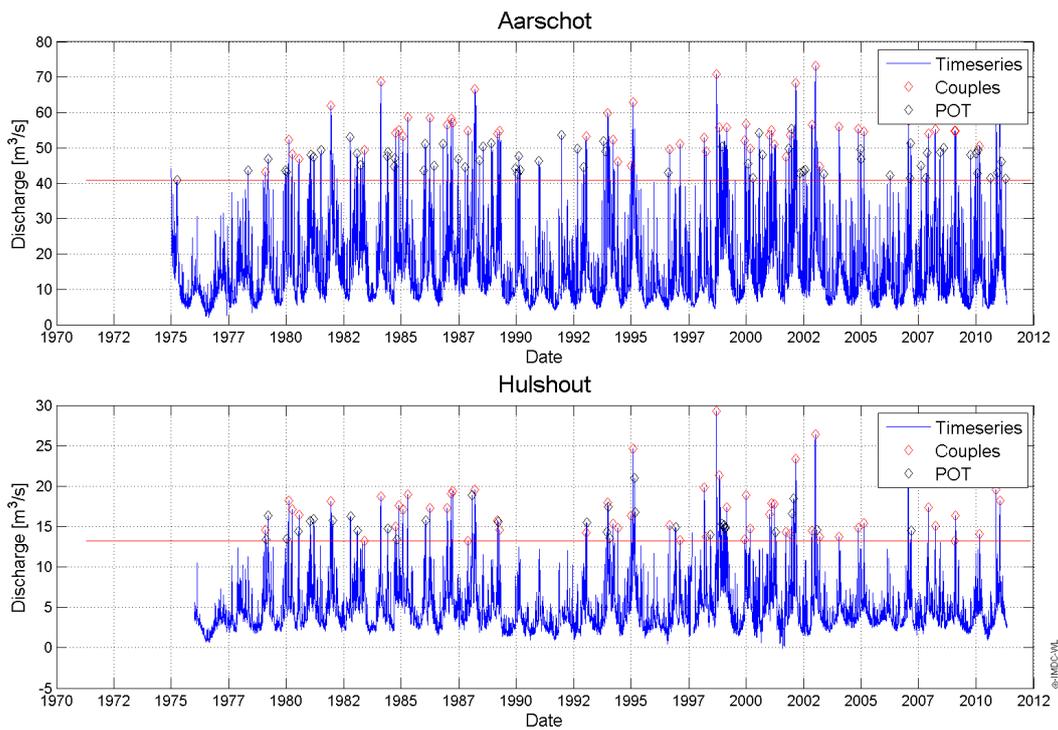


Figure 5-8: Selected couples and uncoupled POT

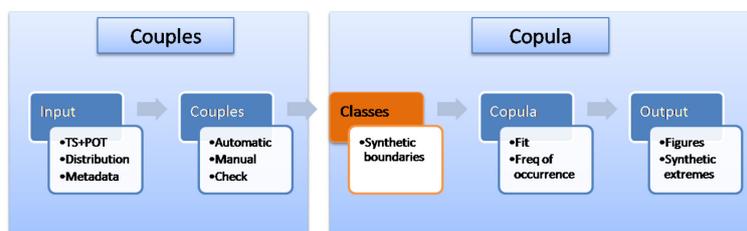


6 Copula

The Copula part of the Copula Analysis tool workflow comprises two steps, i.e.

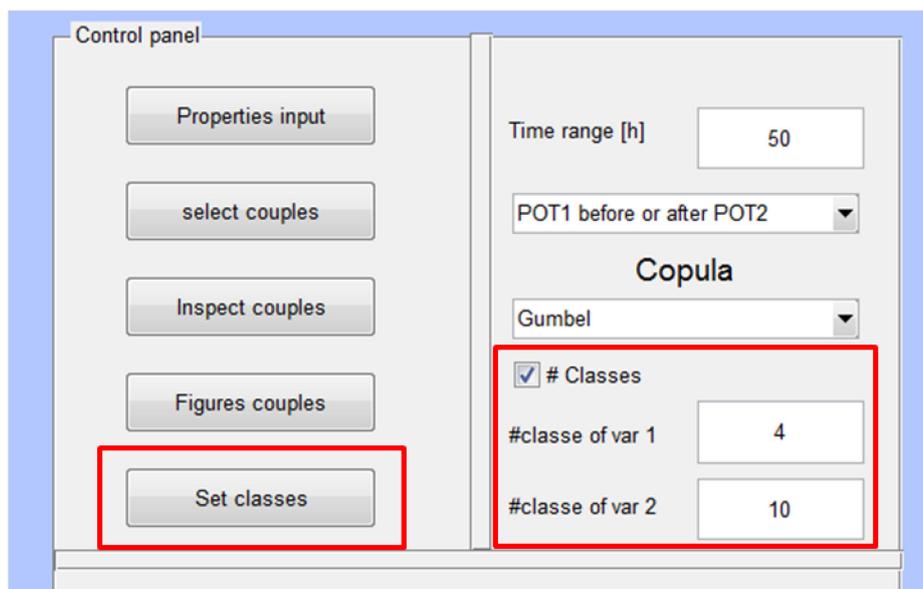
- the definition of classes
- the Copula fitting
- the output of the Copula

6.1 Definition of classes



The step before the fitting of a Copula is the definition of the classes that will be used in the stratification (Figure 6-1). This step is not strictly necessary to fit a Copula, but it is necessary input for the stratification procedure. The number of classes (per variable) is user determined with a default value of 11, which implies 12 class boundaries per variable and a total of 121 classes. The checkbox of the classes will be used in the visualization of the Copula's. If it is checked the Copula values will be visualized in classes, if it is unchecked the values will be smooth.

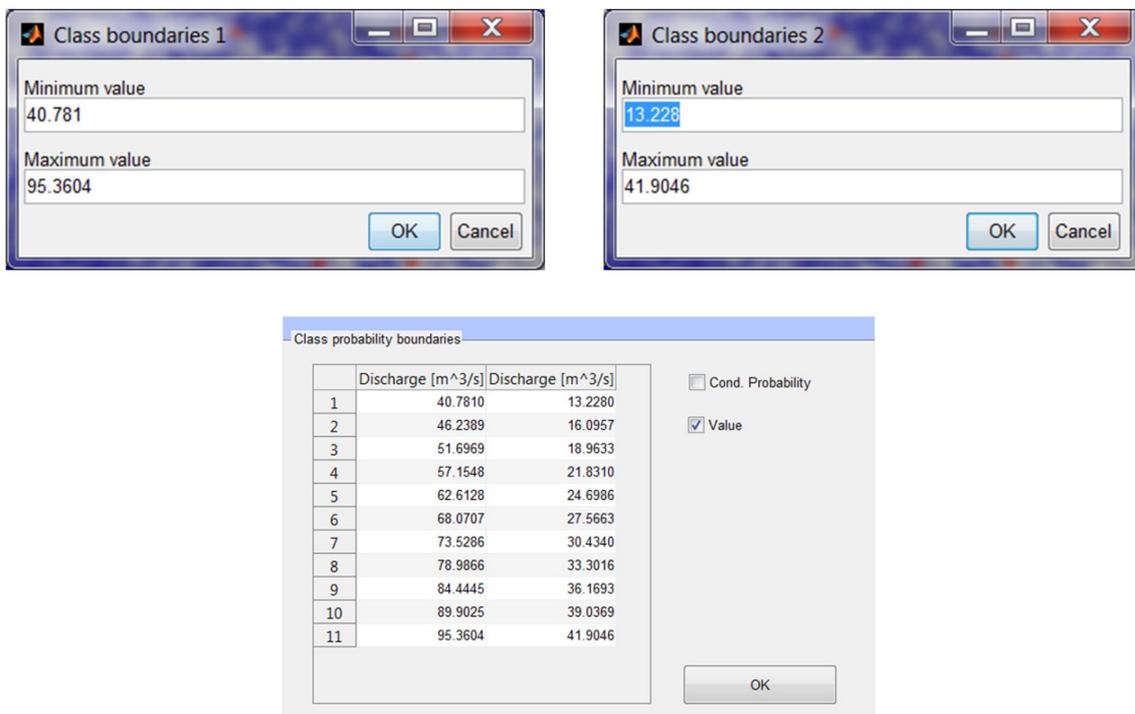
Figure 6-1: Definition of the classes control panel



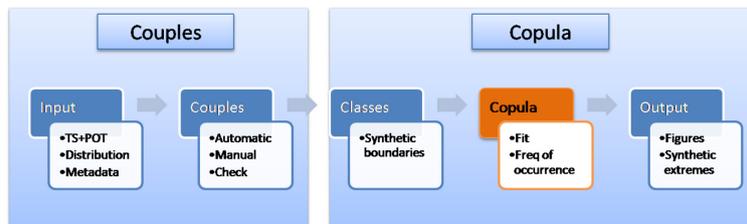
- Set class probability:

The button “Set classes” in Figure 6-1 will stratify the extreme domain of variables 1 and 2 in x classes. The minimum and maximum of the classes is asked in an input dialog box (Figure 6-2). The minimal value needs to be higher than the optimal threshold of the corresponding distribution. The default values of the minimal and maximal values are the threshold values and the return level corresponding with a return period of T 10000. The stratification will automatically result in x equal classes. The user can manually change the boundaries of the classes as long as they are not smaller than the thresholds. The corresponding conditional probability of non exceedance of the class boundaries can be visualized by checking the checkbox “Cond. Probability”.

Figure 6-2: Definition of the class boundaries dialog boxes

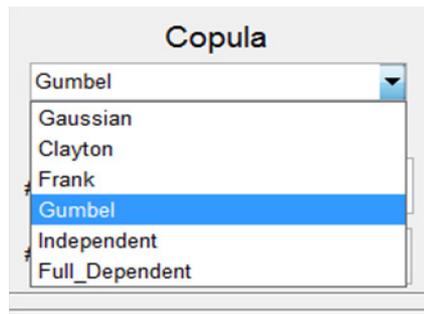


6.2 Copula fitting



There are 4 Copulas available (Figure 6-3) in the Copula Analysis Tool. The properties of these Copulas are described in the Copula Reference Guide (see Figure 1-1). The Clayton, Frank and Gumbel Copula are Archimedean Copulas. On top of these 4 copulas there is the option to combine the 2 univariate distributions in a fully independent and a fully dependent way.

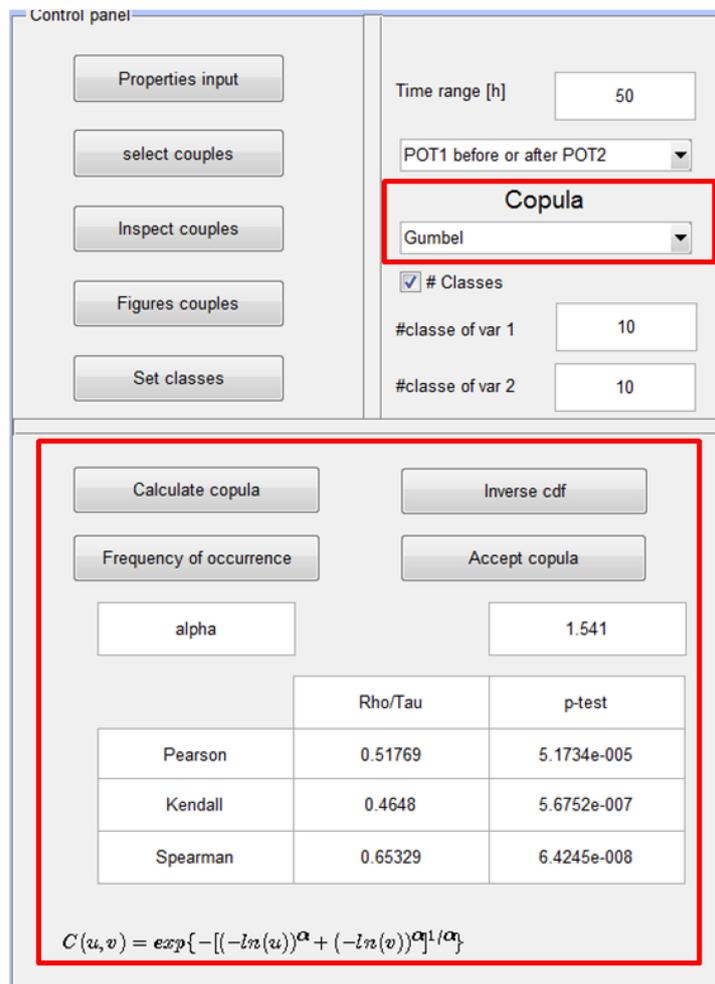
Figure 6-3: Available Copulas in the Copula Analysis tool



The parameters of the chosen Copula are determined through maximum likelihood estimation (Nelsen 2006). The dependency values (Pearson rho, Kendall tau and Spearman rho) are calculated and a statistical p-test is performed with zero hypothesis of independence. If the p-test value is less than the significance level (0.05) the null hypothesis is rejected. So if the p-test value is <0.05 the two variables are statistically dependent. In the other case one can still link the 2 variable as independent distributions. This can be the case if the values of the variables are independent but there are a significant number of joint events.

The fit of a Copula and the stratification procedure is controlled by 3 buttons: “Calculate Copula”, “Inverse cdf (survival function)” and finally “Frequency of occurrence” (Figure 6-4).

Figure 6-4: Copula fitting parameters



- Calculate Copula:

By clicking this button the actual fit of the Copula through the couples is performed. The Copula parameter and the Copula formula are visualized in the control panel. The Copula can be evaluated using 4 figures

- a scatter plot of the couples combined with random generated sample of the Copula (5000 values) (1)
- the values of the empirical Copula (Copula Reference Guide; see Figure 1-1) are compared with the chosen Copula in a QQ-plot (2); the R^2 value of this comparison is included in the figures header
- the Copula value in function of conditional probability of the two univariate distributions u and v , respectively in 3D (3) and 2D (4).

- Inverse cdf (=survival function)

“Inverse cdf” will transform u and v to their variables values by the inverse univariate cumulative density function. Again 4 evaluation figures are produced:

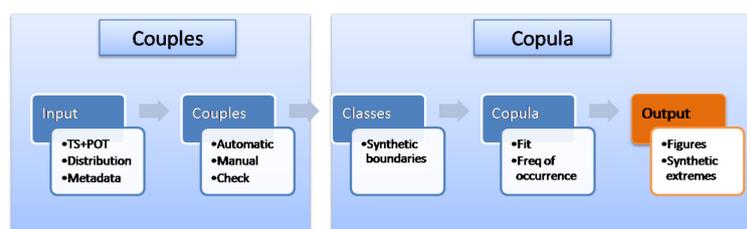
- A scatter plot of the couples and the random generated Copula values as a function of the variable values (1)
- the Copula values on the z-axis as a function of the values of variable 1 on the x-axis and the values of variable 2 on the y-axis (2)
- the survival value ($\Pr(X>x, Y>y)$) of the Copula as a function of variables 1 and 2, respectively in 3D (3) and 2D (4).

- Frequency of occurrence:

The button “Frequency of occurrence” will calculate the yearly frequency of occurrence of every class. This frequency is visualized in 2D and 3D. The 3D plot is turned 180° . So the x and y axes are shifted to. The fourth figure gives the inverse frequency of the classes with a maximum of 10^5 .

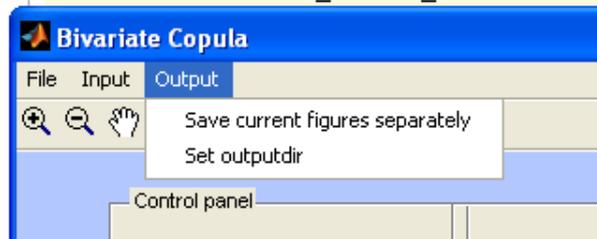
A comparison of the parametric Copula and the empirical frequency (count of observations in a class) is given as a separate figure. This is a good visual control of the performances of the Copula. A significant difference can be expected in the lowest class. This deviation is inherent to the use of joint POT above the optimal threshold as couples.

6.3 Output



The output last step in the Copula Analysis tool is the output of the Copula and all the figures. The menu bar contains an output menu with 2 submenus: “Save current figures separately” and “Set outputdir” (Figure 6-5). This allows the user to choose the directory to save all the output and to save the four figures displayed in the Copula fitting (see §6.2).

Figure 6-5: Output menu



The button “Accept Copula” is the last step in the Copula analysis. A dialog box will ask for the names of the output files (Figure 6-6). The first file is an ASCII file with the synthetic extremes (Figure 6-7). These extremes can be combined with a unity profile to create synthetic events (see Synthetic Events Tool; Figure 1-1). The second and third file are mat files with the synthetic extremes (extremes + joint probability of occurrence) of respectively variable 1 and 2. The fourth file is a binary mat file with all the synthetic extremes and the Copula parameters. Two figures are created, i.e. a 3D plot of the synthetic extremes with the frequency of occurrence and a univariate plot with the original univariate distribution and the recomposed distribution after stratification. The last ASCII file contains the corresponding values of the empirical Copula.

Figure 6-6: Dialog box with names of the output files

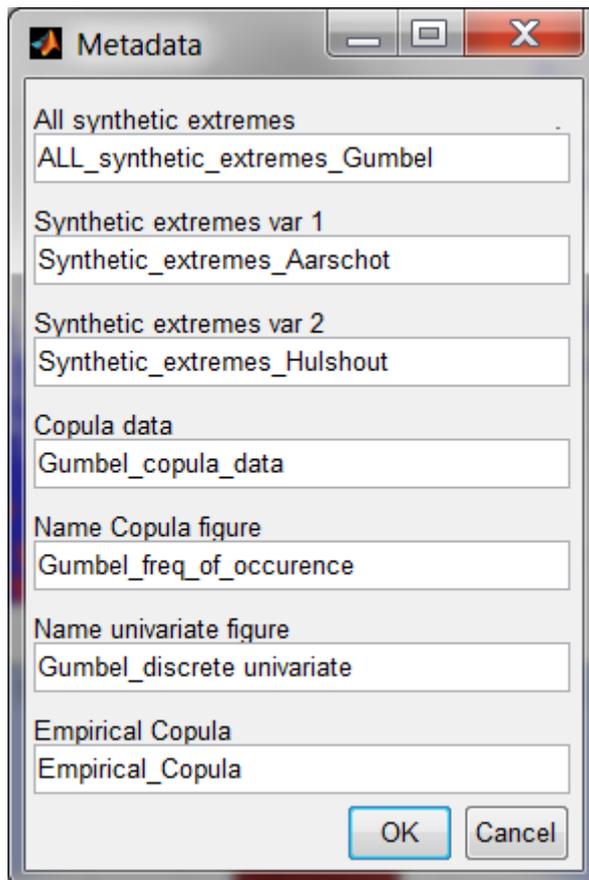


Figure 6-7: Synthetic extremes of variables 1 and 2

```

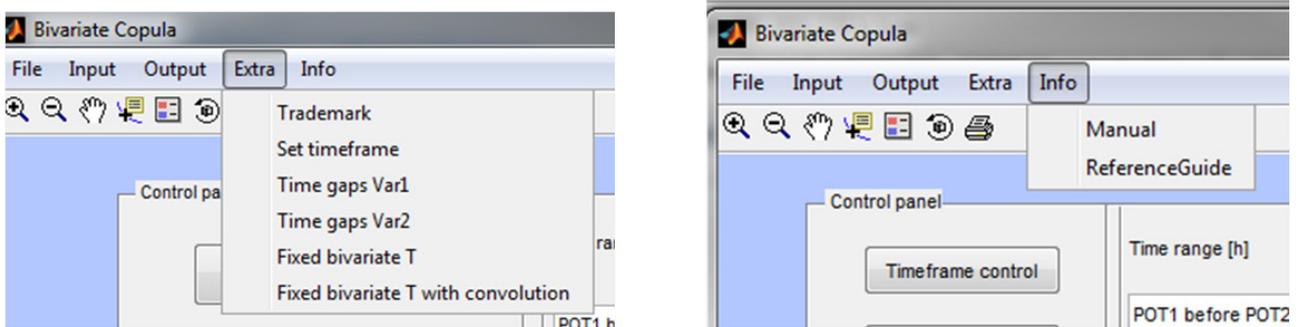
ALL_synthetic_extremes_Gumbel - Notepad
File Edit Format View Help
#Synthetic events (54) created by Bivariate Copula tool
#Gumbel copula
#2012/04/16 11:52
#Yearly frequency of joint occurrence      Discharge [m^3/s]      Discharge [m^3/s]
0.518109622462 40.781000      14.661830
0.236089542941 40.781000      17.529490
0.094341401630 40.781000      20.397150
0.035986531941 40.781000      23.264810
0.013328118042 40.781000      26.132470
0.004830169485 40.781000      29.000130
0.001720532529 40.781000      31.867790
0.000604136958 40.781000      34.735450
0.000209541756 40.781000      37.603110
0.000071900218 40.781000      40.470770
0.789249522548 47.603425      14.661830
0.285434341638 47.603425      17.529490
0.075602343000 47.603425      20.397150
0.017036295844 47.603425      23.264810
0.003610449360 47.603425      26.132470
0.000744133081 47.603425      29.000130
0.000150179113 47.603425      31.867790
0.000029738264 47.603425      34.735450
0.000005787402 47.603425      37.603110
0.000001108716 47.603425      40.470770
1.326102071324 47.603425      13.228000
0.099015656968 61.248275      14.661830
0.116640925853 61.248275      17.529490
0.081064722246 61.248275      20.397150
0.037410550421 61.248275      23.264810
0.012038612119 61.248275      26.132470
0.002917559330 61.248275      29.000130
0.000614298287 61.248275      31.867790
0.000122752530 61.248275      34.735450
0.000023932508 61.248275      37.603110
0.000004586471 61.248275      40.470770
0.395901681962 61.248275      13.228000
0.002234316324 74.893125      14.661830
0.003678297698 74.893125      17.529490
0.005432836538 74.893125      20.397150
0.007319131427 74.893125      23.264810
    
```

7 Extra and info functions

Extra functionalities will be put under the Menu tab Extra (Figure 7-1). The trademark of the output figures can be changed here. If one or both time series contain a lot of gaps it is advised to take only the overlapping periods and POT into account in the calculation of the copula. The functions 'Time gap Var1' and 'Time gaps Var 2' exclude the POT values of the other variable during a certain period. A period of 10 or more days is considered as a gap in a time series.

The functions 'Fixed bivariate T' and 'Fixed bivariate T with convolution' gives the bivariate return level line for a chosen return period. Because we work in a bivariate space this is not a point like for univariate distributions but a selection of point (Var1, Var2). In the function with convolution an extra input is required with the third variable. For example the astronomical high water level for a convolution of storm surge and astronomic high tide into high tide.

Figure 7-1: Extra and Info tab



8 References

Beirlant J., Goegebeur, Y., Teugels, J., 2004, Statistics of Extremes, Theory and Applications, John Wiley & Sons Ltd.

Coles S. 2001, An introduction to statistical modelling of extreme values. London: Springer-Verlag DC 2011

Genest C. , Favre A. 2007, Everything you always wanted to know about Copula modeling but where afraid to ask. Journal of hydrological engineering ASCE

Kotz, S. & Nadarajah, S. 2000, Extreme value distributions, theory and applications. London: Imperial College Press

Matlab 2011, <http://www.mathworks.com/help/toolbox/compiler/f12-999353.html#br2jauc-33> (retrieved on 15/05/2011).

Nelsen R.B., 2006, An Introduction to Copulas. New York, Springer Series in Statistics.

DEPARTMENT **MOBILITY & PUBLIC WORKS**
Flanders hydraulics Research

Berchemlei 115, 2140 Antwerp

T +32 (0)3 224 60 35

F +32 (0)3 224 60 36

waterbouwkundiglabo@vlaanderen.be

www.flandershydraulicsresearch.be