Flanders
State of
the Art

00_144_3
FHR reports

Scientific Assistance towards a Probabilistic Formulation of Hydraulic Boundary Conditions

Copula Analysis Reference Guide

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.

Flanders Hydraulics Research
Flanders State of the Art

Legal notice

Copyright and citation

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_3. Flanders Hydraulics Research & IMDC: Antwerp.

Document identification

Customer:	Flanders Hydraulics Research	Ref.:	WL2019R00_144_3
Keywords (3-5):	Extreme Value; Flooding; Floodrisk		
Text (p.):	10	Appendices (p.):	/
Confidentiality:	☒ No	☒ Available online	

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

Control

	Name	Signature
Reviser(s):	Pereira, F.; Nossent, J.	Getekend door: Fernando Pereira (Signature) Getekend op: 2019-11-05 15:12:40 +01:00 Reden: Ik keur dit document goed Fernando Pereira Getekend door: Jiri Nossent (Signature) Getekend op: 2019-11-05 13:59:03 +01:00 Reden: Ik keur dit document goed Jiri Nossent
Project leader:	Pereira, F.	Getekend door: Fernando Pereira (Signature) Getekend op: 2019-11-05 15:13:01 +01:00 Reden: Ik keur dit document goed Fernando Pereira

Approval

Head of Division:	Mostaert, F.	Getekend door: Frank Mostaert (Signature) Getekend op: 2019-11-05 15:13:53 +01:00 Reden: Ik keur dit document goed Frank Mostaert

Abstract

This report gives an overview of the Copula theory including all the formulas necessary to perform a dependency analysis. This report is the base for a software tool developed for automated Copula analysis. A Copula function is a way to parameterize the dependence between variables. Once the dependence is known, it can be used for extrapolation.

Contents

List of figures

1 Introduction

1.1 Scope of contract

Flanders Hydraulic Research (FHR) has commissioned IMDC NV to adapt its standardized methodology for rendering composite hydrographs, developed at KU Leuven (Willems 2001 & 2002), to recent evolution (e.g. climate change), updated data series (e.g. recent measurements) and diversifying applications (e.g. coastal zone, flood risk calculations,…).

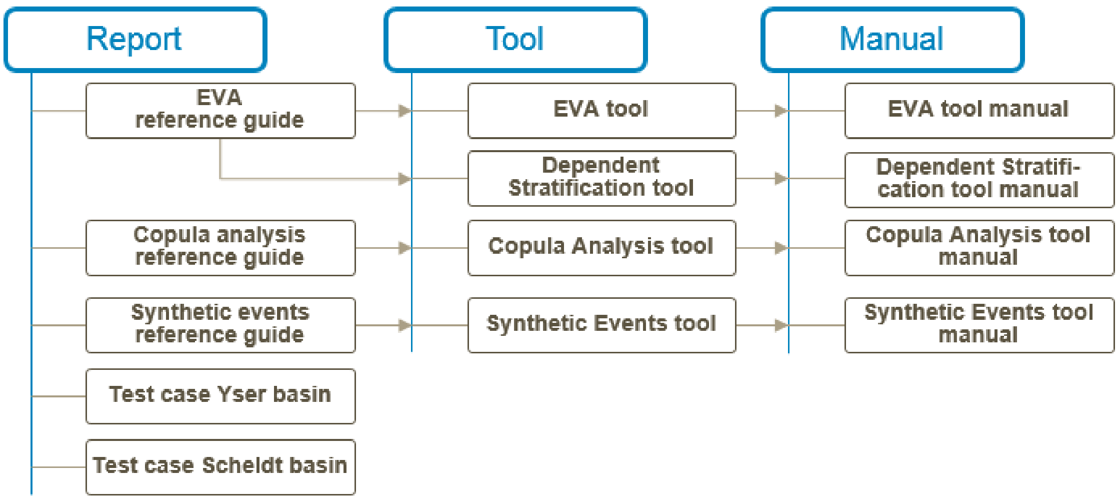
The project team consists of Sarah Doorme (advisor), Gert Leysen (jr. Eng.), Lorens Coorevits (jr. Eng.) and Joris Blanckaert (Sr. Eng. and project manager for IMDC). On behalf of FHR Eng. Fernando Pereira is in charge of the general supervision of the project. Eng. Toon Verwaest of FHR ministers scientific support towards coastal zone applications.

1.2 Overview

A new methodology is presented, which is based on extended literature review and expertise of the project team members. The methodology is described in a set of technical reports and is implemented in a suite of software tools for use in flood risk analysis and probabilistic design projects. The Graphical User Interfaces of the software tools are described in a set of manuals.

The new methodology is tested within two representative test cases, i.e. for the Yzer basin and the Scheldt basin (navigable waterways in Flanders). The test cases are described in two reports. Figure 1-1 presents an overview of the reports, tools and manuals.

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



1.3 This report

This report provides an overview of the Copula theory based on Nelsens standard work (2006) in combination with an intensive literature review. This overview is used to construct a software tool for Copulas. This tool is used to estimate the probability of the joint occurrence of different dependent variables, like wind and storm surge. A Copula function is a way to parameterize the dependence between variables. This parameterization allows for an extrapolation to higher return periods.

The second chapter gives an introduction in the Copula theory and the different methods to quantify the dependence between variables. The number of formulas has been kept to a minimum to ensure the readability. For a more extensive overview the reader is redirected to the different standard works. The third chapter gives an overview of the Copula functions that are implied in the tool.

2 Copula statistics: an overview

2.1 Introduction

The study of copulas and their applications in statistics is a recent phenomenon. The word copula is a Latin noun that means ‘a link, tie, bond’. The term Copula in a statistical sense can be traced back to Sklar (1996) in the theorem (Sklar’s theorem) describing the functions that ‘join together’ one-dimensional distribution functions to form multivariate distribution functions (Nelsen 2006). Copulas have been used in the survival analysis, actuaries sciences and risk analysis. More recent they have been applied in hydrological and meteorological application (Wong 2008; Genest 2007, Karmakar 2007, Blanckaert 2005).

Hydrological phenomena are often multidimensional and hence require the joint modelling of several random variables, like, wind and storm surge or surge and discharge. Traditionally the pair wise dependence between these variables have been described using classical families of bivariate distributions. The main limitation of this approach is that the individual behaviour of the two variables (or transformations thereof) must then be characterized by the same parametric family of univariate distributions (Genest 2007). Copulas avoid this restriction.

Copulas are now enjoying increasing popularity in other areas of risk analysis where one has considerable amounts of data. The rank order correlation employed by most Monte Carlo simulation tools is certainly a meaningful measure of dependence but is very limited in the patterns it can produce. Copulas offer a far more flexible method for combining marginal distributions into multivariate distributions and offer an enormous improvement in capturing the real correlation pattern.

2.2 Copula functions: general overview

Since Li (2000) first introduced copulas into default modelling, there has been increasing interest in this approach. Until that moment, the copula concept was used frequently in survival analysis and actuaries sciences.

Following to Li (2000) and Nelsen (2006), a copula is a function that joins or couples a multivariate distribution to their one-dimensional marginal distribution functions or a distribution function whose one-dimensional margins are uniform. For m uniform random variables U1, U2,…, Um the joint distribution function C is defined as:

$$C(u_1, u_2, \dots, u_m, \rho) = Pr[U_1 \leq u_1, U_2 \leq u_2, \dots, U_m \leq u_m]$$

where ρ is a dependence parameter, can also be called a copula function.

Copula can be used to link marginal distributions with a joint distribution. The copula function C can link univariate marginal or conditional distribution functions $F_1(x_1)$, $F_2(x_2), \dots, F_m(x_m)$ into a multivariate distribution function F with univariate marginal or conditional distributions specified by $F_1(x_1)$, $F_2(x_2), \dots, F_m(x_m)$.

$$C(F_1(x_1), F_2(x_2), \dots, F_m(x_m)) = F(x_1, x_2, \dots, x_m)$$

Sklar (1959) showed that any joint distribution function F can be seen as a copula function. He proved that if $F(x_1, x_2, \dots, x_m)$ is a joint multivariate distribution function with univariate marginal and conditional distribution functions $F_1(x_1), F_2(x_2), \dots, F_m(x_m)$, then there exists a copula $C(u_1, u_2, \dots, u_m)$ such that

$$F(x_1, x_2, \dots, x_m) = C(F_1(x_1), F_2(x_2), \dots, F_m(x_m))$$

If each F_i is continuous then C is unique (Sklar's theorem). Thus, copula functions provide a unifying and flexible way to study joint distributions. This report focuses on the bivariate copula functions C(u,v) for the uniform variables U and V, defined over the area $\{(u,v)|0 < u \leq 1, 0 < v \leq 1\}$.

2.3 Dependence and rank

2.3.1 Boundaries

Suppose that a random sample (X1 ,Y1) , … , (Xn ,Yn) is given from some pair (X,Y) of continuous variables, and that it is desired to identify the bivariate distribution F(x,y) that characterizes their joint behavior. If X and Y are stochastically independent C=u*v.

The other extreme is total dependency. If Y is a deterministic function of X, C must be either one of following copulas

$$W(u,v) = \max(0, u + v - 1) \; or \; M(u,v) = \min(u,v)$$

These two extreme cases which determine the boundaries of a bivariate distribution, are usually referred to as the Fréchet-Hoeffding bounds (Fréchet 1951; Nelsen 2006), where W is the lower bound an M is the upper bound.. When C=W, Y is a decreasing function of X while Y is monotone increasing in X when C=M (Genest 2007). In general, any copula C will lie between these extremes.

To get a feeling of the dependence between X and Y, a simple scatter plot of the pairs (X1 ,Y1) , …., (Xn ,Yn) will give a first impression.

2.3.2 Empirical copula

By associating a rank to every value X and Y a set of ranks $(R_1, S_1),\ldots,(R_n, S_n)$ is obtained. Where R_i stands for the rank of X_i among $X_1,\ldots,X_n$ and S_i stands for the rank of Y_i among $Y_1,\ldots,Y_n$. These rank couples can be used to obtain a so-called empirical copula.

$$C_{emp}(u,v) = \frac{1}{n}\sum_{i=1}^{n} l\left(\frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v\right)$$

Where $l(A)$ denoting the indicator function of set A. For any given pair (u,v), it may be shown that $C_{emp}(u,v)$ is a rank-based estimator of the unknown quantity C(u,v) whose large-sample distribution is centered at C(u,v) and normal (Genest 2007). This empirical Copula can be used as a goodness of fit measurement by comparing it with the Copula values.

2.3.3 Measuring dependence

These are two well-known nonparametric rank based measures availably to quantify the dependence, the Spearman's rho and the Kendall tau. Because these measures are rank based they have some advantages over the well-known Pearson correlation coefficient. The Spearman ρ is given by:

$$\rho_n = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(R_i - \bar{R})^2 \sum_{i=1}^{n}(S_i - \bar{S})^2}}$$

Where

$$\bar{R} = \frac{1}{n}\sum_{i}^{n} R_i = \frac{n+1}{2} = \frac{1}{n}\sum_{i}^{n} S_i = \bar{S}$$

The Spearman's ρ has the property that its expected value goes to zero when the variables are independent, like the classical Pearson r. However the Spearman's ρ has some advantages over the Pearsonn's r. E(ρ)=+-1 occurs if and only if X and Y are functionally dependent, i.e., whenever their underlying copula is one of the two Fréchet–Hoeffding bounds, M or W. In contrast, E(r)= ±1 if and only if X and Y are linear functions of one another, which is much more restrictive. Moreover ρ estimates a population parameter that is always well defined, whereas there are heavy-tailed distributions such as the Cauchy, for example for which a theoretical value of Pearson's correlation does not exist (Embrechts 2002).

The spearman ρ can be used to check the null hypothesis H0 of independence between X and Y The distribution of ρ is close to normal with zero mean and variance (1/(n-1) so one may reject H0 at the confidence level α=5% if $\sqrt{n-1}|\rho| > z_{a/2} = 1.96$. A standard p-test will give the minimal confidence level at which one may reject H0.

The second, well-known measure of dependence based on ranks is Kendall's tau, whose empirical version is given by

$$\tau = \frac{4}{n(n-1)} P_n - 1$$

where Pn is the number of concordant. Here, two pairs (Xi ,Yi), (Xj Yj) are said to be concordant when (Xi-Xj)(Yi-Yj)>0. It is obvious that τ is a function of the ranks of the observations only, since (Xi-Xj)(Yi -Yj)>0 if and only if (Ri-Rj)(Si -Sj)>0.

The Kendall τ can also be used to check the null hypothesis H0 of independence between X and Y. The distribution of τ under H0 is close to normal with zero mean and variance 2(2n+5)/(9n(n-1)) so one may reject H0 at the confidence level α=5 % if

$$\sqrt{\frac{9n(n-1)}{2(2n+5)}|\tau| > 1.96}$$

A standard p-test will give the minimal confidence level at which one may reject H0.

Kendall τ and Spearman ρ can be used as an estimator of the Archimedean Copula parameter (see §3).

2.4 Joint survival function

In many applications we are interested in the joint survival function. For a univariate distribution the survival probability is simply calculated by 1-exceedance probability or 1-CDF (cumulative density function).

$$P[X > x] = 1 - F_1(x) = \bar{F}_1$$

A joint survival function $\bar{F}$ is given by (Nelsen 2006)

$$\bar{F}(x,y) = P[X > x, Y > y) = 1 - F_1(x) - F_2(y) + F(x,y)$$

$$\bar{F}(x,y) = \bar{C}(u,v) = 1 - u - v + C(u,v)$$

3 Copula functions

We will describe 4 Copula function more in detail, the Normal Copula and three Archimedian Copulas, i.e. Gumbel, Clayton and Frank. These Copulas are implemented in a Copula tool (see Figure 1-1).

3.1 Normal copula

The normal copula is an elliptical copula given by:

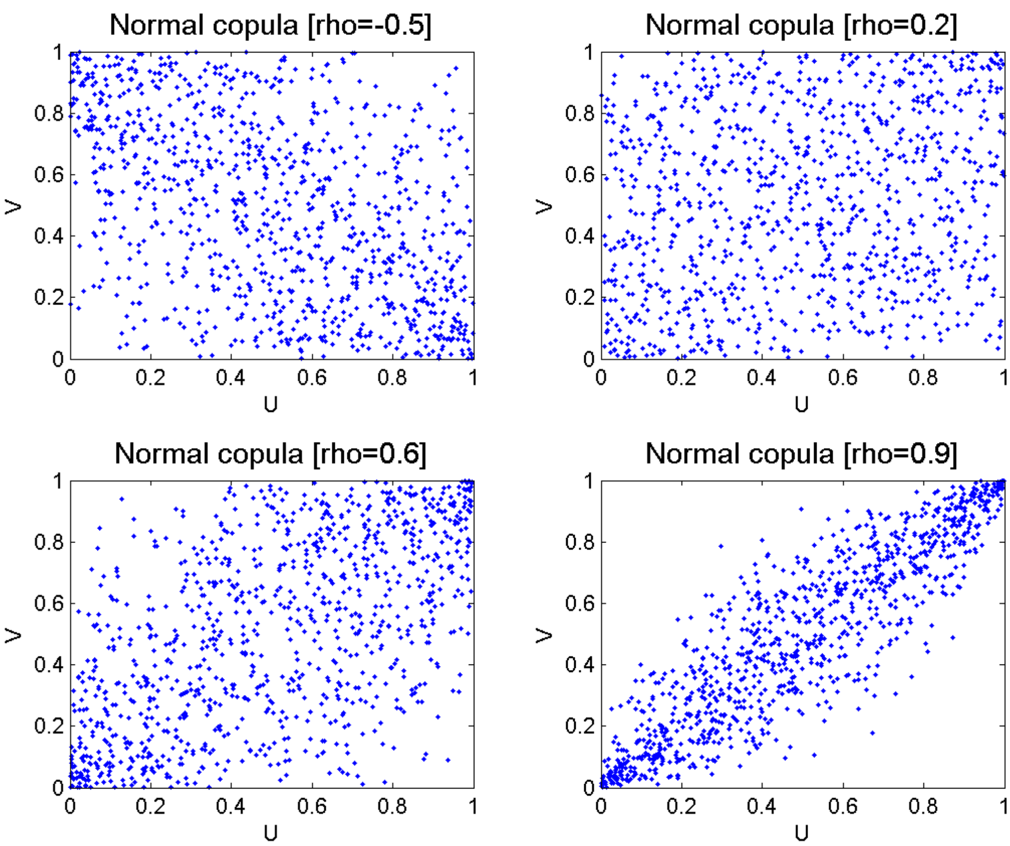
$$C_\rho(u,v) = N_\rho\left(\Phi^{-1}(\mathrm{u}), \Phi^{-1}(\mathrm{v})\right)$$

$$C_\rho(u,v) = \int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} exp\left(-\frac{(s^2-\rho st+t^2)}{2(1-\rho^2)}\right) dsdt$$

Where ρ ≠ -1, 0 or 1 and $\phi^{-1}(u) and\ \phi^{-1}(v)$ are the inverse cumulative distribution function of a standard normal distribution.

If ρ<0 this will imply a negative dependence and ρ>0 will imply a positive dependence. Some examples of random sampling with different values of ρ are visualised in Figure 3-1.

Figure 3-1: Random sampling of Normal Copula for 4 values of rho [-0.5; 0.2; 0.6; 0.9]



3.2 Gumbel Copula

This family of copulas was first discussed by Gumbel (1960) and is referred to as the Gumbel Copula. It should not be confused with the Gumbel-Hougaard Copula or the Gumbel extreme value distribution. The copula value is given by:

$$C(u,v) = exp\left(-[(-\ln(u)^{\alpha} + (-\ln(v)^{\alpha}]^{1/\alpha}\right)$$

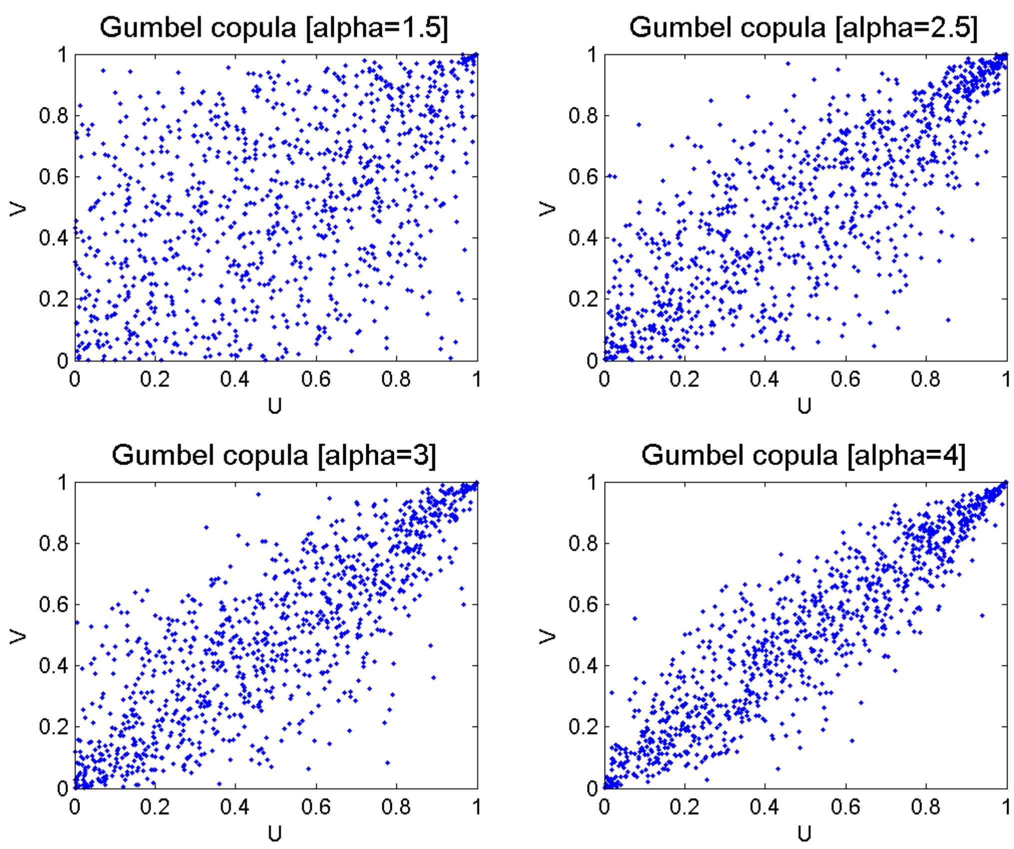
With $\alpha \in (1, \infty)$

And the random sampling generator by:

$$\varphi = \left(-ln(t)\right)^{\alpha}$$

The use of this generator is discussed intensively in Nelsen (2005) and Nelsen (2006). Some examples of random sampling with different values of α are visualised in Figure 3-2.

Figure 3-2: Random sampling of the Gumbel Copula for 4 values of alpha [1.5; 2.5; 3; 4]



3.3 Clayton Copula

This family of copulas was first discussed by Clayton (1978) and is defined as the Clayton family by Genest (1993) and Nelsen (2006). The copula value is given by:

$$C(u,v) = [max(u^{-\alpha} + v^{-\alpha} - 1,0)]^{-1/\alpha}$$

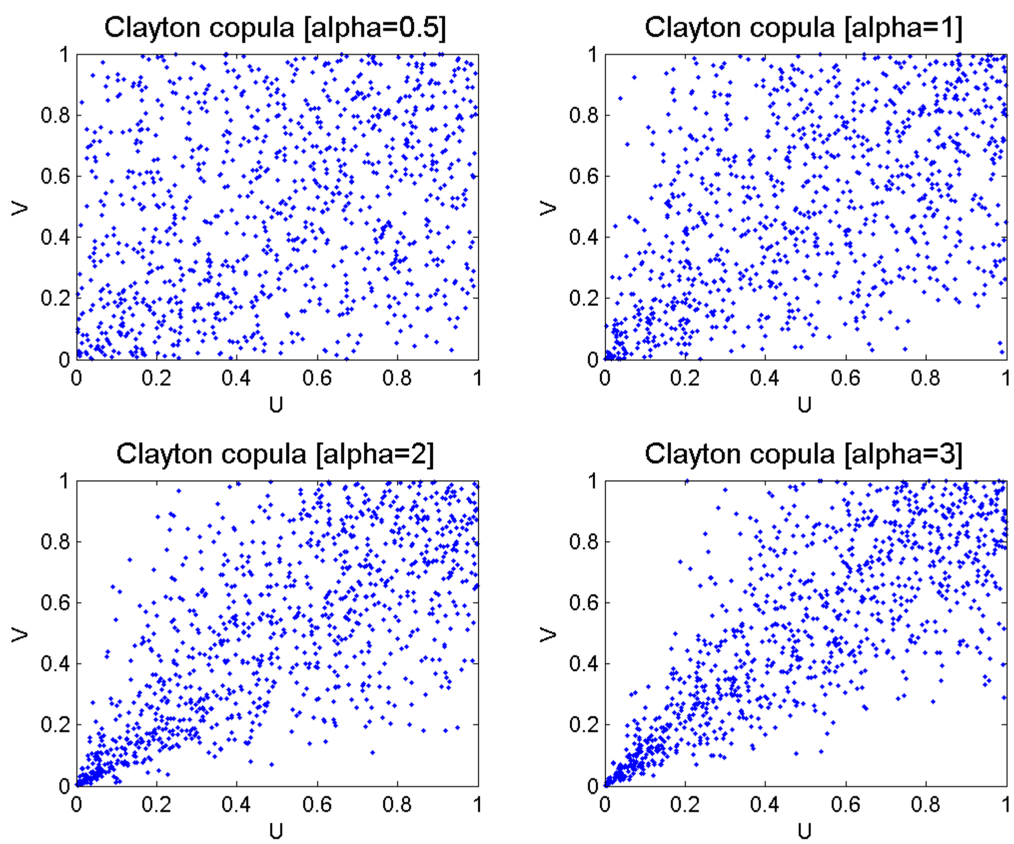
With $\alpha \in (-1,\infty)\backslash\{0\}$

And the random sampling generator by:

$$\varphi = \frac{1}{\alpha}(t^{-\alpha} - 1)$$

The use of this generator is discussed intensively in Nelsen (2005) and Nelsen (2006). Some examples of random sampling with different values of α are visualised in Figure 3-3.

Figure 3-3: Random sampling of the Clayton Copula for 4 values of alpha [0.5;1; 2; 3]



3.4 Frank Copula

This Copula first appeared in Frank (1979) in a non-statistical context. Its statistical properties are discussed in Nelsen (1986). The copula value is given by:

$$C(u,v) = -\frac{1}{\alpha} ln\left(1 + \frac{(e^{-\alpha u}-1)(e^{-\alpha v}-1)}{e^{-\alpha}-1}\right)$$

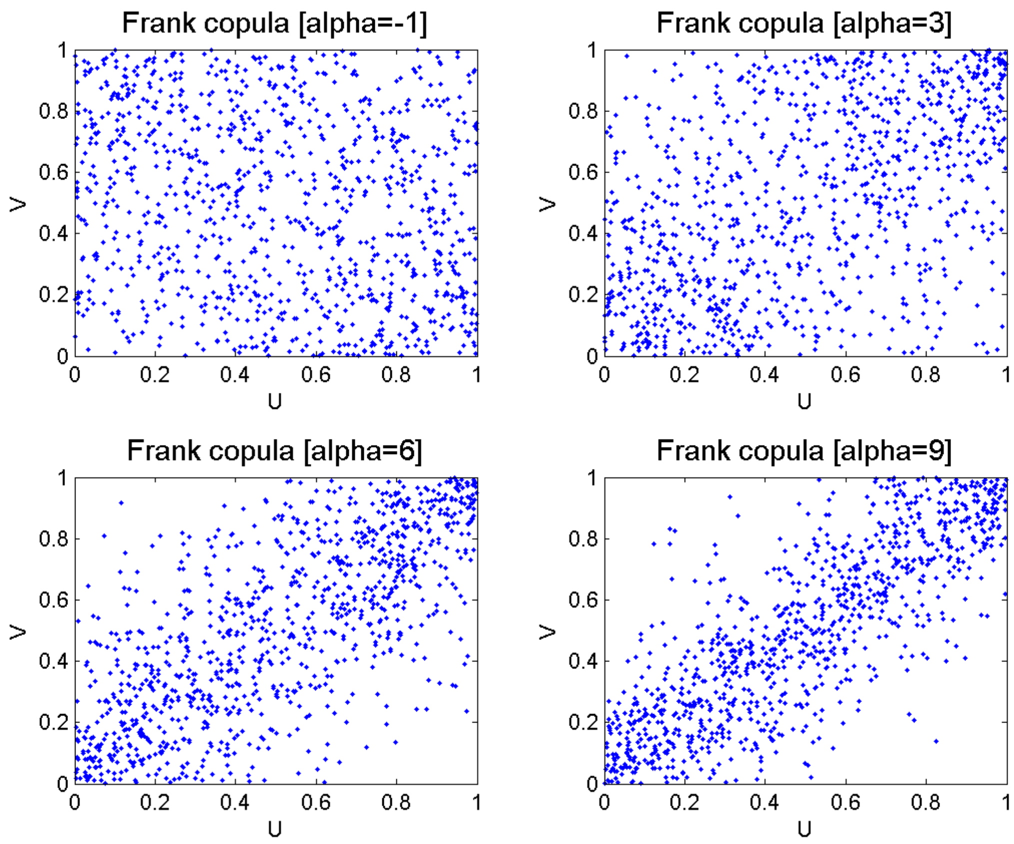
With $\alpha \in (-\infty, \infty)\backslash\{0\}$

And the random sampling generator by:

$$\varphi = -\, ln\frac{e^{-\alpha t}-1}{e^{-\alpha}-1}$$

The use of this generator is discussed intensively in Nelsen (2005) and Nelsen (2006). Some examples of random sampling with different values of α are visualised in Figure 3-4.

Figure 3-4: Random sampling of the Frank Copula for 4 values of alpha [-1 3 6 9]



4 References

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

Blanckaert J. Audoorn L. 2005, A probabilistic detailed level approach to flood risk assessment in the Scheldt estuary, ISSH-Stochastic Hydraulics Nijmegen The Netherlands

Clayton DG, 1978, A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65:141-151

Embrechts P., McNeil, A. J., and Straumann, D., 2002, “Correlation and dependence in risk management: Properties and pitfalls.” Risk management: Value at risk and beyond (Cambridge, 1998), Cambridge University Press, Cambridge, U.K., 176–223.

Frank MJ, 1979, On the simultaneous associativity of F(x,y) and x + y – F(x,y). Aequationes Math 19:194-226

Fréchet M., 1951, “Sur les tableaux de corrélation dont les marges sont données.” Ann. Univ. Lyon Sect. A. (3), 14(3), 53–77.

Genest C, Rivest L-P, 1993, Statistical inference procedures for bivariate Archimedean copulas. J Amer Statist Assoc 88:1034-1043

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

Gumbel E. J., 1960, Distributions des valeurs extrême en plusieurs dimensions. Publ Inst Statist Univ Paris 9: 171-173

Karmakar S., Simonovic S. P., 2007, Water resources research report: Flood frequency analysis using copula with mixed marginal distributions University of Western Ontario.

Li D., 2000, On default correlation: a Copula function Approach, working paper, RiskMetics Group, New York

Nelsen R.B., 1986, Properties of a one-parameter family of bivariate distributions with specified marginals. Comm Statist Theory Methods 15:3277-3285

Nelsen R.B., 2005, Dependence modeling with Archimedean copulas. Department of Mathematical Science, Lewis & Clark College.

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

Sklar A. 1959, "Fonctions de répartition à n dimensions et leurs marges", Publ. Inst. Statist. Univ. Paris 8: 229–231

Willems P., Berlamont, J., Qvick, A., Vaes, G., Feyen, J., Christiaens, K., 2001, Algemene methodologie voor het modelleren van de waterafvoer in bevaarbare waterlopen in Vlaanderen, studie i.o.v. AWZ, afd. Waterbouwkundig Laboratorium en Hydrologisch Onderzoek.

Willems P., Vaes G., Berlamont J., 2002, Onderzoek naar het effect van composiethydrogrammen voor de berekening van overstromingsvlaktes met eendimensionale hydraulische modellen, eindrapport, KULeuven

Wong G., Lambert M.F., Metcalfe A.V., 2008, Trivariate Copulas for characterization of droughts. Australian Methematical Soc.

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