

Received: 10 May 2012 – Accepted: 16 May 2012 – Published: 31 May 2012
Correspondence to: S. Vandenberghe (sander.vandenberghe@ugent.be)
Published by Copernicus Publications on behalf of the European Geosciences Union.

HESSD

9, 6781–6828, 2012

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Abstract

Most of the hydrological and hydraulic studies refer to the notion of a return period to quantify design variables. When dealing with multiple design variables, the well-known univariate statistical analysis is no longer satisfactory and several issues challenge the practitioner. How should one incorporate the dependence between variables? How should the joint return period be defined and applied? In this study, an overview of the state-of-the-art for defining joint return periods is given. The construction of multivariate distribution functions is done through the use of copulas, given their practicality in multivariate frequency analysis and their ability to model numerous types of dependence structures in a flexible way. A case study focusing on the selection of design hydrograph characteristics is presented and the design values of a three-dimensional phenomenon composed of peak discharge, volume and duration are derived. Joint return period methods based on regression analysis, bivariate conditional distributions, bivariate joint distributions, and Kendal distribution functions are investigated and compared highlighting theoretical and practical issues of multivariate frequency analysis. Also an ensemble-based method is introduced. For a given design return period, the method chosen clearly affects the calculated design event. Eventually, light is shed on the practical implications of a chosen method.

1 Introduction

One of the most important aims of hydrological studies is to provide design variables for diverse engineering projects. Recently, there is an increasing interest in, and need for, simultaneously considering multiple design variables, which are likely to be associated with each other. In hydrology and hydraulics, several applications including sewer systems, dams and flood risk mapping require the selection of storm or hydrograph attributes with a predefined return period.

HESSD

9, 6781–6828, 2012

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Standard hydrological design methods are mostly based on well-established univariate frequency analysis methods. Notwithstanding this, methods to describe hydrological phenomena involving multiple variables have recently been proposed, aiding the practitioners to estimate multivariate return periods. In literature, as will be described later on, several methods have evolved over the years. However, it is not clear how these methods compare to each other and which one is the most appropriate.

Recent developments in statistical hydrology have shown the great potential of copulas for the construction of multivariate cumulative distribution functions (CDFs) and for carrying out a multivariate frequency analysis (Favre et al., 2004; Salvadori, 2004; Salvadori and De Michele, 2004, 2007; Salvadori et al., 2007, 2011; Genest and Favre, 2007; Vandenberghe et al., 2011). Copulas are functions that combine several univariate marginal cumulative distribution functions into their joint cumulative distribution function. Copulas describe the dependence structure between random variables and allow for the calculation of joint probabilities, independently of the marginal behaviour of the involved variables. For more theoretical details, we refer to Sklar (1959) and Nelsen (2006). Several studies have been dedicated to the frequency analysis of multivariate hydrological phenomena such as storms and floods, often within the context of design. However, limited applications have been developed with more than two variables (e.g., Vandenberghe et al., 2010; Pinya et al., 2009; Kao and Govindaraju, 2008, 2007; Genest et al., 2007; Serinaldi and Grimaldi, 2007; Zhang and Singh, 2007; Grimaldi and Serinaldi, 2006a,b). For a complete and continuously updated list of papers about copula applications in hydrology see the website of the International Commission on Statistical Hydrology of International Association of Hydrological Sciences.

Multivariate frequency analysis is becoming more and more widespread and several papers provide insight into generalizations of the univariate case and into new definitions of the multivariate return period (see e.g., Salvadori et al., 2011; Salvadori and De Michele, 2004; Shiau, 2003; Yue and Rasmussen, 2002). Since some of the proposed approaches are in contradiction and others are introduced within specific

contexts, there exists a need to clarify the definitions provided so far and to highlight their differences.

In this paper, the construction of multivariate distribution functions based on pair-copulas (Aas et al., 2009) is first briefly introduced (Sect. 2.2) followed by an overview of several methods for defining joint return periods (JRPs, Sect. 3). Subsequently, a case study focusing on the selection of a design hydrograph is presented, which will serve as a test case for evaluating the differences between the joint return period definitions. Section 4 provides all details on the practical context of this case study. Then, in Sect. 5, extreme discharge events are selected and their most important variables such as peak discharge, volume and duration are analyzed, as they form the basis of the analysis. Finally, Sect. 6 deals with evaluating the performance and differences between several JRP methods in quantifying design hydrograph characteristics and highlights important issues for practitioners concerned with multivariate frequency analyses in hydrology.

2 Constructing multivariate copulas

2.1 Choice of construction method

Most of the copula-based research in hydrology addresses the application of two-dimensional copulas, for which several fitting and evaluation criteria are becoming more and more widespread. In contrast, the use of multi-dimensional copulas remains a more challenging task. Only a few hydrological studies address this issue and almost always face severe (practical) drawbacks of the available high-dimensional copula families. Most work has been done in the trivariate analysis of rainfall (e.g., Zhang and Singh, 2007; Kao and Govindaraju, 2008; Salvadori and De Michele, 2006; Grimaldi and Serinaldi, 2006b), floods (e.g., Serinaldi and Grimaldi, 2007; Genest et al., 2007) and droughts (e.g., Song and Singh, 2010; Wong et al., 2010). Recently, a flexible construction method for high-dimensional copulas, based on the mixing of (conditional) two-dimensional copulas, has been introduced and has been shown to have a large

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



potential for hydrological applications. In literature, this construction is known as the pair-copula or vine-copula construction (see e.g., Kurowicka and Cooke, 2007; Aas et al., 2009; Aas and Berg, 2009; Hobæk Haff et al., 2010). The underlying theory for the vine-copula construction is described in Bedford and Cooke (2001, 2002). This construction method originates from work presented by Joe (1997) on which also the method of “conditional mixtures”, as applied by De Michele et al. (2007), is based. In this paper, the vine-copula method will be used to construct the three-dimensional copula for peak discharge Q_p , volume V_p and duration D . The construction and fitting is discussed in the next section.

2.2 Construction of a 3-D copula

In this paper, the focus will be on a three-dimensional vine-copula joining the three marginal distributions of three random variables X , Y and Z . In general, the approach can be extended to any number of dimensions, although limitations may be introduced by the computational power and data available. In the following, we assume that the samples of all three variables have each been transformed using the following rank-order-transformation S in order to obtain the marginal empirical distribution functions:

$$S(x) := \frac{\text{rank}(x)}{n + 1} \quad (1)$$

where n denotes the number of observations for the given variable. We denote the transformed variables by U , V and W so that all three variables are now approximately uniformly distributed on $[0, 1]$.

The basic idea of vine-copulas is to construct high-dimensional copulas based on a stagewise mixing of (conditional) bivariate copulas. This corresponds to decomposing the full density function into a product of low-dimensional density functions. At the base of the construction all relevant pairwise dependences are modelled with bivariate copulas. If all mutual dependences are with respect to the same variable, the construction

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



is called a canonical or C-vine. If all mutual dependences are considered one after the other, i.e. the first with the second one, the third with the fourth one, etc., this is called a D-vine. C- and D-vines are special cases of regular vines, the latter being all possible pairwise decompositions. In the three-dimensional case there is no difference between a C- or a D-vine, only the ordering of variables can be changed.

Figure 1 illustrates the concept for constructing a three-dimensional vine-copula. In the first tree, three variables U, V, W are given, and their pairwise dependences are captured by the bivariate copulas C_{UV} and C_{VW} . These bivariate copulas can be conditioned for a specific value of the second variable through partial differentiation (Aas et al., 2009). This conditioning is indicated by dashed arrows and results in the conditional cumulative distribution functions $F_{U|V}$ and $F_{W|V}$ (see Eq. 2).

$$F_{U|V}(u|v) = \frac{\partial C_{UV}(u, v)}{\partial v} \quad (2a)$$

$$F_{W|V}(w|v) = \frac{\partial C_{VW}(v, w)}{\partial v} \quad (2b)$$

In the second tree, the two conditional CDF values are calculated for all triplets (u_i, v_i, w_i) . These “conditioned observations”, which are again approximately uniformly distributed on $[0, 1]$, are then used to fit another bivariate copula $C_{UW|V}$ to. This copula can also be conditioned through partial differentiation to $F_{U|VW}$. This conditional CDF will be of use for simulation purposes (Aas et al., 2009). The full density function c_{UVW} of the three-dimensional copula is thus given by:

$$c_{UVW}(u, v, w) = c_{UW|V}(F_{U|V}(u|v), F_{W|V}(w|v)) \cdot c_{UV}(u, v) \cdot c_{VW}(v, w) \quad (3)$$

It should be noted that the choice of the conditioning variable (i.e. V) is not unique and different choices might lead to different results. In this paper, the ordering of variables is based on the two bivariate copulas C_{UV} and C_{VW} that fitted best considering the investigated copula families.

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Thus, in order to derive the building blocks of the three-dimensional copula, three bivariate copulas C_{UV} , C_{VW} and $C_{UW|V}$ need to be fitted. This is done stagewise and one can choose any of the available methods in literature. Here, each bivariate copula is fitted by means of the maximum likelihood method, considering different copula families. The best fit is determined by the highest log-likelihood value (see Sect. 5.3).

Several goodness-of-fit tests can be considered to validate the fitted bivariate copulas. In this paper, the chosen goodness-of-fit test is the A7 approach appearing in Berg (2009) and originating from Panchenko (2005). The advantage of this approach is that it estimates the distance between the two multivariate distribution functions without the need of any explicit dimension reduction, i.e. it is directly based on a comparison of observed pseudo-observations and simulated pseudo-observations under the null hypothesis. A bootstrap approach is taken to obtain the distribution of this test statistic under the null hypothesis. The original procedure as proposed by Berg (2009) is slightly altered in this paper as the test statistic of the hypothesis is averaged over the same number of simulations that are conducted during the bootstrap. A p-value estimate is derived from the fraction of test statistics exceeding this mean test statistic.

Combining the bivariate copulas as in Eq. (3) and substituting the marginal distribution functions F_X , F_Y and F_Z yields the three-dimensional distribution function of (X, Y, Z) . Let f_X , f_Y and f_Z denote the marginal density functions and define $u := F_X^{-1}(x)$, $v := F_Y^{-1}(y)$ and $w := F_Z^{-1}(z)$. The full density function f_{XYZ} of the distribution for any triplet (x, y, z) is then given by:

$$f_{XYZ}(x, y, z) := c_{UW|V}(F_{U|V}(u|v), F_{W|V}(w|v)) \cdot c_{UV}(u, v) \cdot c_{VW}(v, w) \cdot f_X(x) \cdot f_Y(y) \cdot f_Z(z) \quad (4)$$

The estimations in this paper have been done using R (R Development Core Team, 2011), a free software environment for statistical computing, and the package `spcopula`¹ building on the packages `copula` (Kojadinovic and Yan, 2010) and `CDVine` (Schepmeier and Brechmann, 2011). The R-scripts are available upon request from the authors. A demo related to this paper will be available in the `spcopula` package.

¹Under development, available at r-forge: <http://r-forge.r-project.org/projects/spcopula>.

3 Joint return periods (JRPs): definitions and methods

In literature, different ways for obtaining design events for a given design return period exist. The following sections provide a short overview of the most popular ways to define joint return periods, focusing on how the design event for a given return period should be calculated. An ensemble-based design approach, in contrast to a single design event, will also be introduced.

3.1 JRP based on a regression analysis

A first method is based on a univariate frequency analysis. First, the driving variable X , i.e. the variable with a prominent role in the design, is chosen. Then a design return period T is fixed, and given the marginal cumulative distribution of the design variable $F_X(x)$ the corresponding design quantile x_T is sought, based on Eq. (5), with μ_T the mean interarrival time (yr). In the case of yearly maxima, μ_T equals 1 yr. Then, based on a linear regression of X with the other design variable Y , the second design value y_T is obtained (see Eq. 6). This approach is taken by e.g., Grimaldi et al. (2012c):

$$T = \frac{\mu_T}{1 - F_X(x_T)} \iff x_T = F_X^{-1} \left(1 - \frac{\mu_T}{T} \right) \quad (5)$$

and

$$y_T = a \cdot x_T + b \quad (6)$$

with a and b constants.

3.2 JRP based on a bivariate conditional distribution

A second method consists in conditioning the bivariate cumulative distribution function (CDF) $F_{XY}(x,y)$ for the design quantile x_T corresponding to the chosen univariate design return period T (see Eq. 5). The resulting (univariate) conditional CDF

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



$F_{Y|X}(y|x = x_T)$ can then be used to calculate, similarly as in Eq. (5), the value y_T for the same univariate design return period T . Advantage will be taken of the bivariate copula $C_{XY}(x, y)$ to perform the calculation. With $u_T = F_X(x_T)$ and $v_T = F_Y(y_T)$ the procedure can be expressed as follows. We can rewrite the initial definition

$$5 \quad T = \frac{\mu_T}{1 - F_{Y|X}(y|x = x_T)}$$

in terms of a copula with $U := F_X(X)$ and $V := F_Y(Y)$ as

$$T = \frac{\mu_T}{1 - \left. \frac{\partial C_{UV}(u, v)}{\partial u} \right|_{u_T := 1 - \frac{\mu_T}{T}}}$$

$$\Leftrightarrow T = \frac{\mu_T}{1 - C_{V|U=u_T}(v_T)}$$

$$10 \quad \Leftrightarrow v_T = C_{V|U=u_T}^{-1}\left(1 - \frac{\mu_T}{T}\right)$$

Inverse transformation yields:

$$y_T = F_Y^{-1}(v_T) \tag{7}$$

15 It should be noted that this method does not result in a real bivariate design event having a joint return period in the strict sense. The bivariate distribution is conditioned for the quantile of interest to the practitioner (corresponding with a univariate return period). This conditioned distribution is then used to obtain the other quantile, again based on the principles of a univariate return period. Therefore, the two obtained design quantiles x_T and y_T should not be considered as a real joint design event.

3.3 JRP based on a bivariate joint distribution

20 Instead of using a conditional CDF, a widely used method to calculate a bivariate return period can be followed which exploits the full bivariate CDF $F_{XY}(x, y)$. This can easily

Selecting multiple design variables

S. Vandenberghe et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The use of the Kendall distribution function to define the probability measure for calculating a JRP is advocated by Salvadori et al. (2011) as it is a theoretically sound multivariate approach sharing the notion of a critical layer, defined through the cumulative distribution function, with the univariate approach. The definition of the return period in both the univariate and in the multivariate Kendall approach is characterized by making a distinction between super-critical and non-critical events based on a critical cumulative probability level. The only way to extend this to a multivariate context is by using the Kendall distribution function. Probability measures that are constructed differently always entail events that will have a joint cumulative distribution function value that is larger or smaller than the critical probability level, and thus fail in subdividing the space between super-critical and non-critical events with respect to the joint cumulative distribution function. Following this avenue, any critical probability level t uniquely corresponds to a subdivision of the space into super-critical and non-critical regions. This is different from the copula approach mentioned before, where in general different choices of critical events from the same critical probability level t subdivide the space differently.

For any given copula of any dimension, the Kendall distribution function can be calculated either analytically (e.g., for Archimedean copulas) or estimated numerically, and can thus be used to calculate the Kendall return period. Until now, only a very limited number of studies actually applied this kind of return period (e.g., Vandenberghe et al., 2010). In the following sections, the procedure for the two- and three-dimensional cases is outlined.

3.4.1 Two-dimensional case

After choosing the design return period T , the corresponding probability level t of the copula can be calculated by means of the inverse of the two-dimensional Kendall distribution function (Eq. 11). In 2-D this corresponds to finding an isoline on the copula.

Selecting multiple design variables

S. Vandenberghe et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



$$\begin{aligned}
 T &= \frac{\mu_T}{1 - K_C(t)} \\
 \Leftrightarrow K_C(t) &= 1 - \frac{\mu_T}{T} \\
 \Leftrightarrow t &= K_C^{-1} \left(1 - \frac{\mu_T}{T} \right)
 \end{aligned} \tag{11}$$

5 When no analytical expression for K_C is available, the inverse can be calculated numerically based on an extensive simulation algorithm, described in Salvadori et al. (2011). Once t is known, the most likely design event in the unit square (u_T, v_T) is selected on the corresponding isoline in the same way as described by Eq. (9). Through the use of the inverse of the marginal CDFs the corresponding design event (x_T, y_T) is then found.

10 3.4.2 Three-dimensional case

In three dimensions, the corresponding probability level t should be found again in the same way as in Eq. (11). To calculate the inverse of the function K_C , one needs to rely on a numerical method as for instance described by Salvadori et al. (2011). However, in contrast to the two-dimensional case, the probability level t corresponds to an isosurface, i.e. all triplets (u, v, w) on this surface have the same copula value t . Generally, for an n -dimensional copula an $n - 1$ dimensional isohypersurface exists that contains all n -dimensional points with the same copula level t . A single design event (u_T, v_T, w_T) should again be selected on this isosurface. Therefore the point (u, v, w) with the highest likelihood is selected. In fact this is the three-dimensional extension of the approach given in Eq. (9), i.e.:

$$(u_T, v_T, w_T) = \underset{C_{UVW}(u,v,w)=t}{\operatorname{argmax}} f_{XYZ}(F_X^{-1}(u), F_Y^{-1}(v), F_Z^{-1}(w)) \tag{12}$$

3.5 JRP and ensembles of design events

From Sects. 3.3 and 3.4 it should be clear that for a design event characterized by several variables, one has to select an event out of a range of events which all share the same JRP. The selection of merely one event sensibly reduces the amount of information that can be obtained by the multivariate approach chosen. The importance of an ensemble-based approach has already been stressed by Salvadori et al. (2011). Vandenberghe et al. (2010) provided a first attempt to benefit from the richness of an ensemble of critical values in a practical context. In this section, a new approach is introduced to obtain an ensemble of statistically similar design events, i.e. that have the same JRP, but that could affect the design in different ways.

Consider first the bivariate case, in which the JRP methods based on copulas (Sect. 3.3) and based on the Kendall distribution function (Sect. 3.4.1) result in the finding of a contour level $t = C(u, v)$ on which all pairs (u, v) have the same JRP.

Instead of using Eq. (9) to select the most likely point, the full likelihood function f_{XY} over the t -isoline could be seen as a univariate density function (PDF) out of which an ensemble of pairs can be sampled. Generally, not all pairs (u, v) on the t -isoline have the same likelihood, i.e. pairs on the edges are less likely than pairs closer to the center of the isoline. In this way, sampling according to f_{XY} makes more sense from a practical point of view than uniformly sampling over the isoline (as done by Vandenberghe et al., 2010).

Eventually, one will end up with an ensemble of (u_i, v_i) -pairs (with i ranging from 1 to n , the ensemble size). By means of the inverse marginal CDFs, these pairs are easily transformed to real values. This ensemble could then be used to run simulations from which detailed information on the uncertainty of specific design parameters can be assessed. As an example, one could route an ensemble of 1000 pairs of peak discharge and volume through a dam model and consider the water height in the reservoir. Using just one design event, only one single water height is obtained. However, using the ensemble, information on the range and likelihood of possible water heights for the

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



given design JRP is obtained, making it possible to incorporate the uncertainty in the design.

In the trivariate case (see Sect. 3.4.2) no isoline is obtained but an isosurface. Similar to the two-dimensional case, the full likelihood over this isosurface could be seen as a bivariate density function out of which an ensemble of triplets can be sampled. The higher the dimensionality of the design problem, the more advantageous the ensemble approach becomes: in three dimensions more information is lost than in two dimensions by selecting just one design event. The drawback of the ensemble approach is the increasing need for run time when higher dimensionalities are considered.

4 Comparison of JRP methods in the synthetic design hydrograph application

4.1 Experimental set-up

In order to provide an exhaustive comparison of the joint return period estimation methods described in the previous sections, a simulation experiment is set up and analyzed with respect to the synthetic design hydrograph (SDH) attributes.

The SDH is defined as a hydrograph with an assigned return period, which can be characterized by random variables such as the flood peak Q_p , the volume V_p and the duration D . Specifically, given an observed or simulated runoff time series from which a set of extreme hydrographs is selected, one can determine the SDH shape in several ways (see Serinaldi and Grimaldi (2011) and references therein). In a two-dimensional setup, two hydrograph parameters (peak-volume, peak-duration or volume-duration) should be fixed, while the third one is obtained from the chosen hydrograph shape distribution. In a three-dimensional set-up, the three characteristic parameters are obtained jointly.

In most common hydrological applications the interest is in the flood peak (Q_p) and volume (V_p). Consequently, the two-dimensional analyses in this paper focus on these variables. Serinaldi and Grimaldi (2011) illustrate a simple approach to obtain design

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



values for Q_p and V_p . They apply a univariate statistical analysis on Q_p to obtain the design value for the peak discharge with a given return period, and then define the corresponding design volume through a linear regression analysis between Q_p and V_p . This approach is equivalent to the method described in Sect. 3.1. However, as described in Sect. 3, there are several other approaches that lead to the design values for Q_p and V_p , including a three-dimensional approach.

The case study proposed in this paper consists of applying a continuous simulation model on a small, ungauged basin for which 500 yr of synthetic direct runoff time series at a 5 min resolution are simulated. From this series, the 500 maximum annual peaks are selected together with their corresponding hydrograph (identified as the continuous sequence of non-zero direct discharge values including the annual peak). Note that as direct discharge is considered, a zero discharge value does not imply a dry river. Consequently, 500 (Q_p, D, V_p) triplets are available to which the described JRP methods are applied. By considering a real case study, the obtained differences can be evaluated in a practical context. In order to simulate the 500 yr runoff time series, the COSMO4SUB model, described in the following section, is applied.

4.2 The COSMO4SUB framework

The synthetic data set on which the previously described JRP estimation methods are applied is obtained through the use of the COSMO4SUB framework (Grimaldi et al., 2012d,c). COSMO4SUB is a continuous model which allows the simulation of synthetic direct runoff time series using minimal input information from rainfall data and digital terrain support. Specifically, the watershed digital elevation model (DEM) with a standard resolution used in hydrological modelling, the soil use and type, daily (preferably at least 30 yr long) and sub-daily (preferably at least 5 yr long) rainfall observations are the only data necessary to run the model. COSMO4SUB includes three modules: a rainfall time series simulator, a rainfall excess scheme, and a geomorphological rainfall-runoff model. Next, the general principles are explained and in Sect. 5.1 specific details of the calibration are presented.

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The first module is based on a single-site copula-based daily rainfall generator (Seriali, 2009) and on the continuous-in-scale universal multifractal model (Schertzer and Lovejoy, 1987) for disaggregating the daily rainfall to the desired time scale (up to 5 min). The parameters included in this first module (six for each month for the daily rainfall simulator and three for the disaggregation model) are calibrated on the basis of the available rainfall observations (at two different scales).

The second module is related to the rainfall excess step. A new mixed Green Ampt-Curve Number (CN4GA Curve Number for Green Ampt) procedure was recently proposed (Grimaldi et al., 2012b) and included in the present version of the COSMO4SUB framework. The key concept is to use the initial abstraction (i.e. all the losses due to initial saturation, filling terrain gaps, interception, etc.) and the total SCS-CN excess rainfall volume to estimate the effective saturated hydraulic conductivity and the ponding time of the Green-Ampt model. Consequently, the CN4GA approach tries to appropriately distribute the volume estimated by the SCS-CN method over time. This module is characterized by five parameters (specified in Sect. 5.1) which are empirically assigned using the soil use and soil type map information. In addition, the event separation time (T_s) is included in this module since the continuous implementation of the SCS-CN method requires to fix a no-rain time interval for which the cumulative gross and excess precipitation can be reset to zero. As shown in Grimaldi et al. (2012d,c), this parameter has a limited influence on the final results and the value can be arbitrarily assigned in the range of 12–36 h.

The third module allows to carry out a continuous convolution of the rainfall excess for obtaining the direct runoff time series through an advanced version of the Width Function Instantaneous Unit Hydrograph (WFIUH). The adopted model, named WFIUH-1par (Grimaldi et al., 2010, 2012a), identifies the watershed IUH through the topographic information and needs only one parameter that can be quantified referring to the watershed concentration time (T_c), estimated using empirical equations. Following the application of the three described modules a continuous runoff scenario is obtained from which maximum annual hydrographs are selected.

5 Data and materials

5.1 Model set-up

In order to provide a realistic scenario that can be used to evaluate the different JRP methods, the COSMO4SUB model was applied on the Torbido River, a small tributary of the Tiber River located in Central Italy (watershed area: 61.67 km²). Basin elevations range from 85 to 625 m, the average slope is 22 % and the maximum distance between divide and outlet is 25.8 km. The watershed DEM at a 20-m spatial resolution was provided by the Italian Geographic Military Institute (IGMI, 2003), while land cover was extracted from the CORINE database (EEA, 2000).

Observed rainfall data, useful for calibrating the two-stage rainfall simulator parameters, are available from the Castel Cellesi rain gauge station for a period of 49 yr at a daily time scale and for a period of 10 yr at a 5-min resolution (Serinaldi, 2009, 2010). For a description and evaluation of the 500-yr rainfall synthetic time series, we refer to Grimaldi et al. (2012c).

As mentioned in Sect. 4.2, the excess rainfall time series was obtained by applying the CN4GA method and setting the percentage of potential retention that forms the initial abstraction equal to 20 % ($\lambda = 0.2$) and the Curve Number (CN) equal to 80 (estimated from NRCS (2008); Chapter 9, 2–3 pp.). Concerning Green-Ampt parameters, the following values were adopted based on the sandy loam average soil condition (EEA, 2000): 110 mm for the moving wetting front (h_f), 0.1 for the initial value of soil-water content (θ_i) and 0.5 for the field saturated soil-water content (θ_s); for the saturated hydraulic conductivity (K_s), an initial value of 10 mm h⁻¹ was chosen and later automatically calibrated by the model (Rawls et al., 1982, 1983).

The event separation time T_s is set to 36 h based on the physical hypothesis that in this region, the initial abstraction phenomena return to the pre-event condition in 1.5 days. The WFIUH-1par and the convolution of the excess rainfall series were computed as described by Grimaldi et al. (2010, 2012c,b). The only calibration parameter (the

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



channel velocity) was estimated based on the concentration time $T_c = 4.5$ h, which was quantified through an in Italy widely used empirical formula (Giandotti, 1934).

5.2 Annual extreme discharge events

Once the 500-yr synthetic direct runoff time series is determined, as described in Sect. 4.1, the 500 maximum annual events are selected and characterized by their peak discharge Q_p , volume V_p and duration D . For only six years the model provides a zero direct runoff, which is reasonable considering the limited size of the watershed. These values are excluded in the following analyses.

In first instance, the marginal distribution functions of Q_p , V_p and D need to be estimated. As these variables are annual extreme values selected from the 500 yr discharge series, the fit of several extreme value distributions is considered, i.e. the exponential, the Weibull and the Generalised Extreme Value (GEV) distribution functions. These distributions are, respectively, a one, two and three parameter distribution, allowing for various degrees of model complexity. Furthermore, the GEV distribution generally encompasses three different distributions, namely the Fréchet, the (reversed) Weibull and Gumbel distributions either directly, or through a transformation as in the case of the Weibull distribution which corresponds to a reversed Weibull distribution. These different distribution types each represent a different kind of tail behaviour, namely a light tail (Gumbel), a heavy tail with infinite variance (Fréchet) and a bounded upper tail (Weibull). These behaviours can be separated based on the shape parameter ξ of the GEV. Furthermore, the Weibull distribution is fitted separately as well, as it only corresponds to a GEV distribution after transformation. Finally, most extreme value distributions are of the exponential type, and cannot deal with an offset, i.e. when the smallest value of the variable in the CDF is larger than zero. However, as a result of censoring the zeros, the smallest value of the variables tends to be significantly higher than zero, leading to poor fits of the CDF. Therefore, the distributions are fitted to the data after subtraction of the minimum value to ensure a proper fit in the tails.

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



A first test to ascertain the appropriate distribution for the three marginal variables is to display the empirical CDFs together with the directly fitted distribution. This is shown in Fig. 2, in which only the upper tail of the CDF is shown, i.e. the interval [0.80, 1] as the focus is on the extremes. It can immediately be seen that not all the distributions fit these tails equally well. This is corroborated by the Akaike Information Criterion (AIC) computed for all different models, shown in Table 1, as well as the log-likelihood of each model (not shown). Based on these criteria, we would select the Weibull distribution for Q_p , the exponential distribution for V_p and the GEV for D . Despite this, the GEV does not fit D well, even though various statistics showing it as the best fit. Seemingly, the lower portion of the curve is reasonably well approximated by a GEV, resulting in the significance of its fit, despite the poor representation of the tail. As the focal point of this study is the tail, we chose to select the exponential distribution because of its better fit in this region. More in-depth testing through Q-Q plots (not shown here) indicates that this is indeed a better approximation of the distribution, despite its non-significance. Hence, the following models are selected:

- Q_p : Weibull (Anderson-Darling $p = 0.562$),
- V_p : Exponential (Anderson-Darling $p = 0.295$),
- D : Exponential (Anderson-Darling $p = 0.0443$).

Here, the Anderson-Darling test was used to determine whether the samples were significantly different from the fitted distributions. As mentioned earlier, only the duration was significantly different from its distribution. It should be understood that a consistency in marginal distribution functions across the different JRP estimation methods is far more important for comparison reasons than a perfect fit, considering the underlying data are simulations.

To analyse the association between the variables, which will be modelled by means of copulas, Kendall's tau is calculated and normalized rank scatterplots are evaluated for each pair of variables (Fig. 3). Evidently, there are strong positive associations.

Also, some ties are present, especially for D , but they will be neglected in the further analyses. The next section deals with the modelling of these associations.

5.3 Fitting of the 2-D and 3-D copulas

As described in Sect. 2.2, we use maximum likelihood estimation to fit a copula from each investigated family for every pair of variables and selected the best fitting one by the highest log-likelihood value. The copula families investigated include Normal, Student, Gumbel, Frank, Clayton, BB1, BB6, BB7, BB8 and the survival copulas of the 4 latter ones (details on all these families can be found in Nelsen, 2006; Joe, 1997). Table 2 gives an overview of the parameters and goodness-of-fit results. The p-values are estimated from 1000 iterations each.

The following JRP methods in the two-dimensional case make use of the fitted BB7 copula C_{13} which models the dependence between Q_p and V_p . It should be noted that this copula is not able to represent the boundary effect present in the rank-scatter plot (Fig. 3). As the BB7 copula family belongs to the class of Archimedean copulas, its Kendall distribution function can easily be obtained analytically. For the JRP method in the three-dimensional case, the three fitted bivariate copulas C_{12} , C_{23} and $C_{13|2}$ are then composed into the three-dimensional vine-copula as given in Eq. (3). For comparison purposes, three-dimensional copula fits for the 3 parameter Gaussian copula and the one parameter Clayton, Frank and Gumbel copulas have also been performed. The log-likelihoods show a 10% increase for the fitted vine-copula (1047) with respect to the Gaussian one (935), while the three one-parameter Archimedean copulas have far smaller values (432–532). Thus, the vine-copula yields the best fit within this set of copula families, i.e. a Gaussian distribution is outperformed by our approach. As no closed form exists for the cumulative distribution function of this vine-copula, a numerical evaluation based on a sample of 100 000 points was carried out in order to be able to calculate the (inverse of the) Kendall distribution function.

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



6 Results and discussion

6.1 Calculation of single design events

In this section, the design values for the SDH with a design return period of $T = 100$ yr are calculated based on the 2-D and 3-D JRP methods presented in Sect. 3. The triplet (Q_p, D, V_p) is considered for which the following transformations hold:

$$\begin{aligned}U &= F_{Q_p}(q_p) \\V &= F_D(d) \\W &= F_{V_p}(v_p)\end{aligned}\tag{13}$$

As a reference, the univariate case is first analyzed. Figure 4 shows the derivation of the design quantiles for Q_p , D and V_p based on the inverse of the CDFs F_{Q_p} , F_D and F_{V_p} , at a probability level of $1 - \frac{\mu_T}{T} = 1 - \frac{1}{T} = 0.99$. Design values $q_{p,T} = 302.9 \text{ m}^3 \text{ s}^{-1}$, $d_T = 28.37 \text{ h}$ and $v_{p,T} = 4.42 \times 10^6 \text{ m}^3$ are obtained. In the following, Table 3 and Fig. 9 provide a way to compare these and all further derived design events. In order to be able to compare design events with the data, the simulated pairs (Q_p, V_p) are visualized as gray dots in Fig. 9 that summarizes all described methods.

First the two-dimensional case is considered, in which the focus is on the couple (Q_p, V_p) . In the regression-based JRP method (Sect. 3.1) the starting point is the univariately derived quantile $q_{p,T}$, being usually the driving variable in many hydrological applications (see Eq. 5). Based on a linear regression between Q_p and V_p , as shown in Fig. 5, the design volume $v_{p,T}$ is easily estimated as $3.89 \times 10^6 \text{ m}^3$. This volume is lower than the one obtained by a purely univariate analysis.

The second two-dimensional method is the JRP method based on the conditional copula (Sect. 3.2). The conditioning of the bivariate copula C_{UW} (denoted as C_{13} in Sect. 5.3) for $u = 0.99$ results in the function $C_{W|U}(w, u = 0.99)$ as shown in Fig. 6. The value of $w = 0.9983$ corresponds with a probability level of 0.99. By means of

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



the inverse $F_{V_p}^{-1}(w)$, the design volume $v_{p,T}$ is calculated as $6.13 \times 10^6 \text{m}^3$, which is considerably larger than the former design volumes.

The JRP method based on the bivariate copula C_{UW} (Sect. 3.3) is the third 2-D approach. For $T = 100$ yr, the corresponding copula level t equals 0.99 and corresponds with an isoline. Using the marginal CDFs for Q_p and V_p Eq. (9) can be solved to find the point (u_T, w_T) with the highest likelihood, i.e. (0.9927, 0.9926). Using the inverse CDFs the design event is obtained: $(q_{p,T}, v_{p,T}) = (319.4 \text{m}^3 \text{s}^{-1}, 4.71 \times 10^6 \text{m}^3)$. Both the design peak discharge and the volume are larger than what is obtained in the univariate case.

The last two-dimensional approach is the one in which the JRP is calculated using the Kendall distribution function (Sect. 3.4.1). Here, the focus is on the inverse of K_C for a probability level of 0.99: $t = K_C^{-1}(0.99)$. Figure 7 shows the Kendall distribution function of the bivariate copula C_{UW} , which allows to calculate the t -level corresponding to a cumulative probability of 0.99, i.e. $t = 0.9823$. This level is smaller than the one obtained in the former copula-based JRP method. Again, Eq. (9) can be solved to obtain the most likely design event $(u_T, w_T) = (0.9871, 0.9870)$. Transformation to the real domain by means of the inverse CDFs results in the design event $(q_{p,T}, v_{p,T}) = (290 \text{m}^3 \text{s}^{-1}, 4.16 \times 10^6 \text{m}^3)$.

Besides the estimation of two-dimensional design events, also one method for estimating a three-dimensional design event is available. To do so, one should rely on the JRP method presented in Sect. 3.4.2 together with the fitted three-dimensional vine-copula (see Sect. 5.3). The three-dimensional vine-copula is used for simulating 100 000 triplets (u, v, w) as a basis for the numerical inversion of the Kendall distribution function. The latter is shown in Fig. 8: the probability level of 0.99 corresponds to a t -level of 0.9509 on the three-dimensional vine-copula. In contrast to the two-dimensional methods, this t -level corresponds to a surface. Using the marginal CDFs in combination with Eq. (12), the most likely point on this surface having a joint cumulative probability density of 0.9509 is found as $(u_T, v_T, w_T) = (0.9752, 0.9692, 0.9834)$. Using the inverse CDFs this results in the design event $(q_{p,T}, d_T, v_{p,T}) = (254 \text{m}^3 \text{s}^{-1}, 22.4 \text{h}, 3.93 \times 10^6 \text{m}^3)$.

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Note that the Kendall distribution function is a univariate representation of multivariate information and that its form is different in the two-dimensional and three-dimensional cases (Fig. 7 vs. 8).

6.2 Obtaining an ensemble of design events

5 The preceding analyses resulted in a single design event, however, as stated in Sect. 3.5 the generation of an ensemble would be preferable. For example, consider the JRP method based on the Kendall distribution function in the two-dimensional case. The t -level was found to be 0.9823 for a JRP of 100 yr (see Table 3). Figure 10 shows this t -level, together with the earlier identified most likely design event (u_T, w_T).

10 However, along this contour the occurrence of several other events is possible. A sampling across this contour according to the full likelihood function results in an ensemble of events all having a JRP equal to 100 yr. An ensemble size of 500 is chosen here. By means of the inverse marginal CDFs, this ensemble of (u, w) pairs can be transformed into (q_p, v_p) pairs. The result is shown in Fig. 11. All sampled events clearly lie on a contour, corresponding with the t -contour. According to the greyscale, the highest density of design events is sampled around the most likely realization whereas less design events are sampled on the two outer limits of the contour.

15 The density of the ensemble across this contour could be projected (and normalized) on both the Q_p and V_p axis, resulting in univariate PDFs for Q_p and V_p in the ensemble. These are shown in Figs. 12 and 13. The most likely design event ($q_{p,T}, v_{p,T}$) = ($290 \text{ m}^3 \text{ s}^{-1}$, $4.16 \times 10^6 \text{ m}^3$) (see Table 3) is naturally situated at the maximum of these PDFs. In general, these conditional distributions do not have to be bounded and extremely large events might possess a positive likelihood.

25 These PDFs hold a lot of information on the design events. For example, 90 % of all design events with JRP equal to 100 yr have a volume in the range of [$3.98 \times 10^6 \text{ m}^3$, $5.45 \times 10^6 \text{ m}^3$] and a peak discharge in [$279 \text{ m}^3 \text{ s}^{-1}$, $358 \text{ m}^3 \text{ s}^{-1}$]. Note from Fig. 11 that lower volumes occur together with higher peak discharge values and vice versa. As briefly mentioned in Sect. 3.5, the ensemble of design events can also be used to

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



calculate another design variable, such as the water height in a reservoir, for which again a PDF of possible design values can easily be obtained. However, this exercise is beyond the scope of this paper.

6.3 Some practical considerations

5 Table 3 and Fig. 9 clearly demonstrate that the choice of the JRP method influences the parameters of the design events. This evidently is something the practitioner should consider when designing a structure, e.g., a dam, based on a specific design hydrograph, as it directly influences the safety and the cost of the structure to be built.

10 With respect to the univariate design quantile $q_{p,T}$, only the copula-based JRP method provides a larger quantile, whereas for the JRP methods based on the Kendall distribution function a smaller quantile is found. The other JRP methods use the univariate quantile as a starting point, so these quantiles are the same. Considering the design quantile $v_{p,T}$, all JRP methods give a smaller quantile than the univariate quantile, except for the methods based on the (conditional) copula.

15 In fact, the presented JRP methods include the dependence between the design variables in two different ways: both by means of a linear regression, and by means of copulas. Nowadays, the use of copulas, which are able to model non-linear dependences, is preferred above the more simplistic linear regression, so practitioners should rely less on the latter method for JRP calculations. The remaining JRP methods all use
20 copula-based multivariate information. The method using the conditional copula is in principle different from the other three methods: the bivariate distribution function is conditioned for a univariate quantile which results in a univariate distribution function of the other design variable. Therefore a relatively larger design volume is found, compared to the standard univariate method. The three other JRP methods use information
25 of the full bivariate (copula and 2-D Kendall method) or trivariate (3-D Kendall method) distribution function. The Kendall function based methods have the advantage of using a mathematically consistent way of defining the probability of extremes or dangerous events relying on the CDF as in the univariate approach, unlike the JRP method based

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



functions. Of course, a good balance between the number of variables considered and the (numerical) complexity of the vine-copula should be sought, keeping in mind that all this also affects the eventual design.

Finally, a practitioner should also be very aware of the fact that the JRP methods based on the Kendall distribution function as presented here are only valid for variables that are positively associated, and with a focus on extremes in terms of large values. In all other cases, adaptations should be made in order to operate in the right “area” of the copula. Further applications of this JRP method in other case studies should provide more insight on this in the near future.

7 Conclusions

The aim of this study was to provide an overview of state-of-the-art methods for joint return period (JRP) estimation and to discuss their differences in a practical context. Therefore, an in-depth case study focusing on the estimation of design parameters for a synthetic design hydrograph (SDH) was considered. As they are the most important SDH variables, the peak discharge Q_p , the duration D and the volume V_p were chosen.

In first instance, a review of several JRP methods available in recent literature was provided focusing on how to apply these methods. As multiple variables were considered in the JRP methods, an important aspect is (the modelling of) the dependence between variables. In this context, the potential and use of copulas for the construction of multivariate distribution functions was stressed and illustrated. On the one hand a bivariate copula of (Q_p, V_p) was fitted. On the other hand, also the fitting of the trivariate copula of (Q_p, D, V_p) was elaborated in a comprehensive way by means of the promising pair-copula or vine-copula method.

Eventually, design events for a 100-yr joint return period were obtained considering a 2-D linear regression based, a 2-D conditional copula-based, a 2-D copula-based, a 2-D Kendall distribution function-based and a 3-D Kendall distribution function-based method. The traditional 1-D method is considered as a reference for comparison

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

I◀

▶I

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



purposes. Differences in design quantiles were discussed while also the theoretical appropriateness was explained. This paper warns practitioners for blind use of just one available JRP estimation method, and stresses the importance of good copula fitting and the effect on the eventual design event outcome. Based on the available literature and the case study in this paper, the JRP method based on the Kendall distribution function is probably the most valuable in a multivariate context, when applied correctly. For constructing multivariate copulas, the vine-copula method is advised.

Further (joint) research efforts should focus on a shift from one-design-event methods to ensemble-design-event methods, enabling to incorporate uncertainty in the design. A first valuable approach to this ensemble-based design was provided in this paper. The ultimate goal should be the elaboration of a useful and understandable framework for multivariate frequency analyses, with clear guidelines to practitioners.

Acknowledgements. The first author is a doctoral research fellow of the Research Foundation Flanders (FWO). The Special Research Fund of Ghent University financially enabled the research of M. J. van den Berg.

References

- Aas, K. and Berg, D.: Models for construction of multivariate dependence – a comparison study, *Eur. J. Finance*, 15, 639–659, doi:10.1080/13518470802588767, 2009. 6786
- Aas, K., Czado, C., Frigessi, A., and Bakken, H.: Pair-copula constructions of multiple dependence, *Insur. Math. Econ.*, 44, 182–198, doi:10.1016/j.insmatheco.2007.02.001, 2009. 6785, 6786, 6787
- Bedford, T. and Cooke, R.: Probability density decomposition for conditionally dependent random variables modeled by vines, *Ann. Math. Artif. Intell.*, 32, 245–268, 2001. 6786
- Bedford, T. and Cooke, R.: Vines – a new graphical model for dependent random variables, *Ann. Stat.*, 30, 1031–1068, 2002. 6786
- Berg, D.: Copula Goodness-of-fit testing: an overview and power comparison, *Eur. J. Finance*, 15, 675–701, 2009. 6788

- De Michele, C., Salvadori, G., Passoni, G., and Vezzoli, R.: A multivariate model of sea storms using copulas, *Coast. Eng.*, 54, 734–751, 2007. 6786
- EEA: CORINE (Coordination of Information on Environment) Database, a key database for European integrated environmental assessment, Programme of the European Commission, 2000. 6798
- 5 Favre, A. C., Adlouni, S. E., Perreault, L., Thiemonge, N., and Bobee, B.: Multivariate hydrological frequency analysis using copulas, *Water Resour. Res.*, 40, W01101, doi:10.1029/2003WR002456, 2004. 6784
- Genest, C. and Favre, A. C.: Everything you always wanted to know about copula modeling but were afraid to ask, *J. Hydrol. Eng.*, 12, 347–368, 2007. 6784
- 10 Genest, C., Favre, A. C., Beliveau, J., and Jacques, C.: Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data, *Water Resour. Res.*, 43, W09401, doi:10.1029/2006WR005275, 2007. 6784, 6785
- Giandotti, M.: Previsione delle piene e delle magre dei corsi d'acqua, *Istituto Poligrafico dello Stato*, 8, 107–117, 1934. 6799
- 15 Grimaldi, S. and Serinaldi, F.: Asymmetric copulas in multivariate flood frequency analysis, *Adv. Water Resour.*, 29, 1155–1167, 2006a. 6784
- Grimaldi, S. and Serinaldi, F.: Design hyetograph analysis with 3-copula function, *Hydrolog. Sci. J.*, 51, 223–238, 2006b. 6784, 6785
- 20 Grimaldi, S., Petroselli, A., Alonso, G., and Nardi, F.: Flow time estimation with variable hillslope velocity in ungauged basins, *Adv. Water Resour.*, 33, 1216–1223, 2010. 6797, 6798
- Grimaldi, S., Petroselli, A., and Nardi, F.: A parsimonious geomorphological unit hydrograph for rainfall runoff modeling in small ungauged basins, *Hydrol. Process.*, 57, 73–83, 2012a. 6797
- Grimaldi, S., Petroselli, A., and Romano, N.: Green-Ampt Curve Number mixed procedure as an empirical tool for rainfall-runoff modelling in small and ungauged basins, *Hydrol. Process.*, in press, doi:10.1002/hyp.9303, 2012b. 6797, 6798
- 25 Grimaldi, S., Petroselli, A., and Serinaldi, F.: Design hydrograph estimation in small and ungauged watersheds: continuous simulation method versus event-based approach, *Hydrol. Process.*, in press, doi:10.1002/hyp.8384, 2012c. 6789, 6796, 6797, 6798
- 30 Grimaldi, S., Petroselli, A., and Serinaldi, F.: A continuous simulation model for design hydrograph estimation in ungauged watersheds, *Hydrolog. Sci. J.*, accepted, 2012d. 6796, 6797
- Hobæk Haff, I., Aas, K., and Frigessi, A.: On the simplified pair-copula construction – simply useful or too simplistic?, *J. Multivar. Anal.*, 101, 1296–1310, 2010. 6786

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- IGMI: Raster (Matrix) numerical DEM of Italy, available at: http://www.igmi.org/pdf/info_matrix2003.pdf (last access: 29 May 2012), internal factsheet, 2003 (in Italian). 6798
- Joe, H.: Multivariate Models and Dependence Concepts, Chapman and Hall, 1997. 6786, 6801
- Kao, S. C. and Govindaraju, R. S.: A bivariate frequency analysis of extreme rainfall with implications for design, *J. Geophys. Res.*, 112, D13119, doi:10.1029/2007JD008522, 2007. 6784
- Kao, S. C. and Govindaraju, R. S.: Trivariate statistical analysis of extreme rainfall events via the Plackett family of copulas, *Water Resour. Res.*, 44, W02415, doi:10.1029/2007WR006261, 2008. 6784, 6785
- Kojadinovic, I. and Yan, J.: Modeling multivariate distributions with continuous margins using the copula R package, *J. Stat. Softw.*, 34, 1–20, available at: <http://www.jstatsoft.org/v34/i09/>, 2010. 6788
- Kurowicka, D. and Cooke, R. M.: Sampling algorithms for generating joint uniform distributions using the vine-copula method, *Comput. Stat. Data Anal.*, 51, 2889–2906, doi:10.1016/j.csa.2006.11.043, 2007. 6786
- Nelsen, R. B.: An Introduction to Copulas, Springer, 2006. 6784, 6801
- NRCS: Hydrology, National Engineering Handbook, part 630, US Department of Agriculture, Washington DC, 2008. 6798
- Panchenko, V.: Goodness-of-fit test for copulas, *Physica A*, 355, 176–182, 2005. 6788
- Pinya, M. A. S., Madsen, H., and Rosbjerg, D.: Assessment of the risk of inland flooding in a tidal sluice regulated catchment using multi-variate statistical techniques, *Phys. Chem. Earth*, 34, 662–669, doi:10.1016/j.pce.2009.03.001, 2009. 6784
- R Development Core Team: R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, available at: <http://www.R-project.org/>, packages: CDVine: 1-1.5 (last access: 30 December 2011), copula: 0.9-9 (last access: 02 December 2011), spcopula: 1.0.57 (last access: 13 March 2012), 2011. 6788
- Rawls, W. J., Brakensiek, D. L., and Saxton, K. E.: Estimation of soil water properties, *Trans. Am. Soc. Agr. Eng.*, 25, 1316–1328, 1982. 6798
- Rawls, W. J., Brakensiek, D. L., and Miller, N.: Estimation of soil water properties, in: Green-Ampt infiltration parameters from soil data, *J. Hydraul. Eng.-ASCE*, 109, 62–70, 1983. 6798
- Salvadori, G.: Bivariate return periods via 2-copulas, *J. R. Stat. Soc. Ser. B-Stat. Methodol.*, 1, 129–144, 2004. 6784, 6791

Selecting multiple design variables

S. Vandenberghe et al.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[◀](#)
[▶](#)
[◀](#)
[▶](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)


- Salvadori, G. and De Michele, C.: Frequency analysis via copulas: theoretical aspects and applications to hydrological events, *Water Resour. Res.*, 40, W12511, doi:10.1029/2004WR003133, 2004. 6784, 6791
- Salvadori, G. and De Michele, C.: Statistical characterization of temporal structure of storms, *Adv. Water Resour.*, 29, 827–842, 2006. 6785
- Salvadori, G. and De Michele, C.: On the use of copulas in hydrology: theory and practice, *J. Hydrol. Eng.*, 12, 369–380, 2007. 6784
- Salvadori, G., De Michele, C., Kottegoda, N., and Rosso, R.: *Extremes in Nature. An Approach using Copulas*, Springer, 2007. 6784, 6791
- Salvadori, G., De Michele, C., and Durante, F.: On the return period and design in a multivariate framework, *Hydrol. Earth Syst. Sci.*, 15, 3293–3305, doi:10.5194/hess-15-3293-2011, 2011. 6784, 6791, 6792, 6793, 6794, 6806
- Schepsmeier, U. and Brechmann, E. C.: CDVine: Statistical inference of C- and D-vine copulas, R package version 1.1-5., available at: <http://CRAN.R-project.org/package=CDVine>, packages: CDVine: 1-1.5 (last access: 30 December 2011), copula: 0.9-9 (last access: 02 December 2011), spcopula: 1.0.57 (last access: 13 March 2012), 2011. 6788
- Schertzer, D. and Lovejoy, S.: Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes, *J. Geophys. Res.*, 92, 9693–9714, 1987. 6797
- Serinaldi, F.: A multisite daily rainfall generator driven by bivariate copula-based mixed distributions, *J. Geophys. Res.-Atmos.*, 114, D10103, doi:10.1029/2008JD011258, 2009. 6797, 6798
- Serinaldi, F.: Multifractality, imperfect scaling and hydrological properties of rainfall time series simulated by continuous universal multifractal and discrete random cascade models, *Nonlinear Process Geophys.*, 17, 697–714, doi:10.5194/npg-17-697-2010, 2010. 6798
- Serinaldi, F. and Grimaldi, S.: Fully nested 3-copula: procedure and application on hydrological data, *J. Hydrol. Eng.*, 12, 420–430, 2007. 6784, 6785
- Serinaldi, F. and Grimaldi, S.: Synthetic design hydrographs based on distribution functions with finite support, *J. Hydrol. Eng.*, 16, 434–446, doi:10.1061/(ASCE)HE.1943-5584.0000339, 2011. 6795
- Shiau, J. T.: Return period of bivariate distributed extreme hydrological events, *Stoch. Environ. Res. Risk Assess.*, 17, 42–57, 2003. 6784
- Sklar, A.: Fonctions de répartition à n dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris*, 8, 229–231, 1959. 6784

- Song, S. and Singh, V. P.: Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data, *Stoch. Environ. Res. Risk Assess.*, 24, 425–444, doi:10.1007/s00477-009-0331-1, 2010. 6785
- 5 Vandenberghe, S., Verhoest, N. E. C., Buyse, E., and De Baets, B.: A stochastic design rainfall generator based on copulas and mass curves, *Hydrol. Earth Syst. Sci.*, 14, 2429–2442, doi:10.5194/hess-14-2429-2010, 2010. 6784, 6792, 6794
- 10 Vandenberghe, S., Verhoest, N. E. C., Onof, C., and De Baets, B.: Evaluation of Bartlett-Lewis rainfall models through a comparative copula-based bivariate frequency analysis of observed and simulated rainfall time series, *Water Resour. Res.*, 47, W07529, doi:10.1029/2009WR008388, 2011. 6784
- Wong, G., Lambert, M. F., Leonard, M., and Metcalfe, A. V.: Drought analysis using trivariate copulas conditional on climatic states, *J. Hydrol. Eng.*, 15, 129–141, doi:10.1061/(ASCE)HE.1943-5584.0000169, 2010. 6785
- 15 Yue, S. and Rasmussen, P.: Bivariate frequency analysis: discussion of some useful concepts in hydrological application, *Hydrol. Process.*, 16, 2881–2898, 2002. 6784
- Zhang, L. and Singh, V. P.: Gumbel-Hougaard copula for trivariate rainfall frequency analysis, *J. Hydrol. Eng.*, 12, 409–419, 2007. 6784, 6785

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Table 1. The values of the AIC for the various distributions of the respective variables.

	GEV	Exponential	Weibull
Q_p	5370.215	5358.230	5329.98
V_p	14640.998	14597.641	14612.409
D	2610.425	2646.081	2643.729

Selecting multiple design variables

S. Vandenberghe et al.

Table 2. An overview of the fitted bivariate copulas in the 2-D copula and 3-D vine-copula approach.

	Pairs of variables	id	$\tau_K (-)$	Copula family	Parameters		p-value
2-D	peak–volume	13	0.85	BB7	2.24	14.10	0.69
3-D	peak–duration	12	0.42	survival BB7	2.05	0.35	0.74
	duration–volume	23	0.54	survival BB7	2.25	1.09	0.75
	peak–volume duration	13 2	0.83	Student Copula	0.96	2.00	0.66

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[I◀](#)
[▶I](#)
[◀](#)
[▶](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)


Selecting multiple design variables

S. Vandenberghe et al.

Table 3. Overview of the calculated design event for $T = 100$ yr, based on several methods. The values are rounded to address the limited numerical precision and ease comparison.

JRP-approach	t (–)	K_C (–)	u_T (–)	v_T (–)	w_T (–)	$q_{p,T}$ ($\text{m}^3 \text{s}^{-1}$)	d_T (h)	$v_{p,T}$ ($\times 10^6 \text{m}^3$)
Univariate	x	x	0.9900	0.9900	0.9900	303	28.4	4.42
Lin.regr.	x	x	0.9900	x	0.9827	303	x	3.89
Cond.cop.	0.9900	x	0.9900	x	0.9983	303	x	6.13
Cop.	0.9900	x	0.9927	x	0.9926	319	x	4.71
2-D Kendall	0.9823	0.9900	0.9871	x	0.9870	290	x	4.16
3-D Kendall	0.9509	0.9900	0.9752	0.9692	0.9834	254	22.4	3.93

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[◀](#)
[▶](#)
[◀](#)
[▶](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)


Selecting multiple design variables

S. Vandenberghe et al.

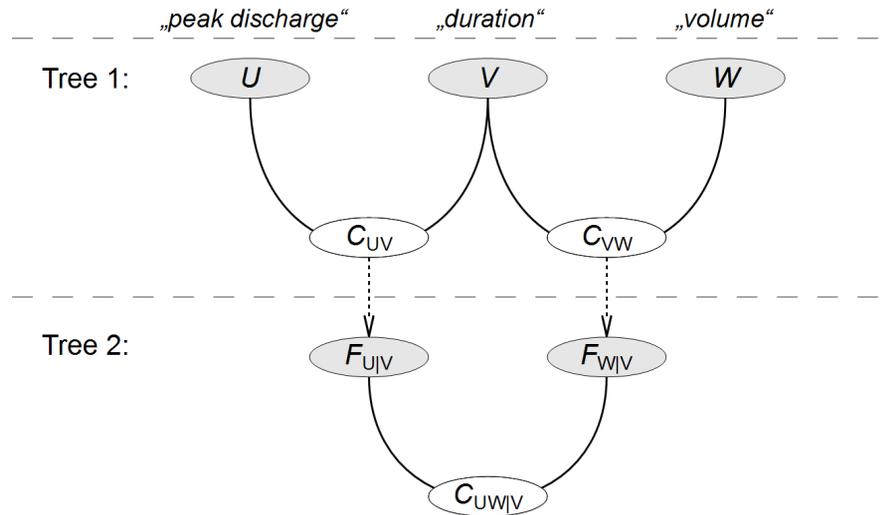


Fig. 1. The concept of the hierarchical nesting of bivariate copulas in the construction of a 3-D copula.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



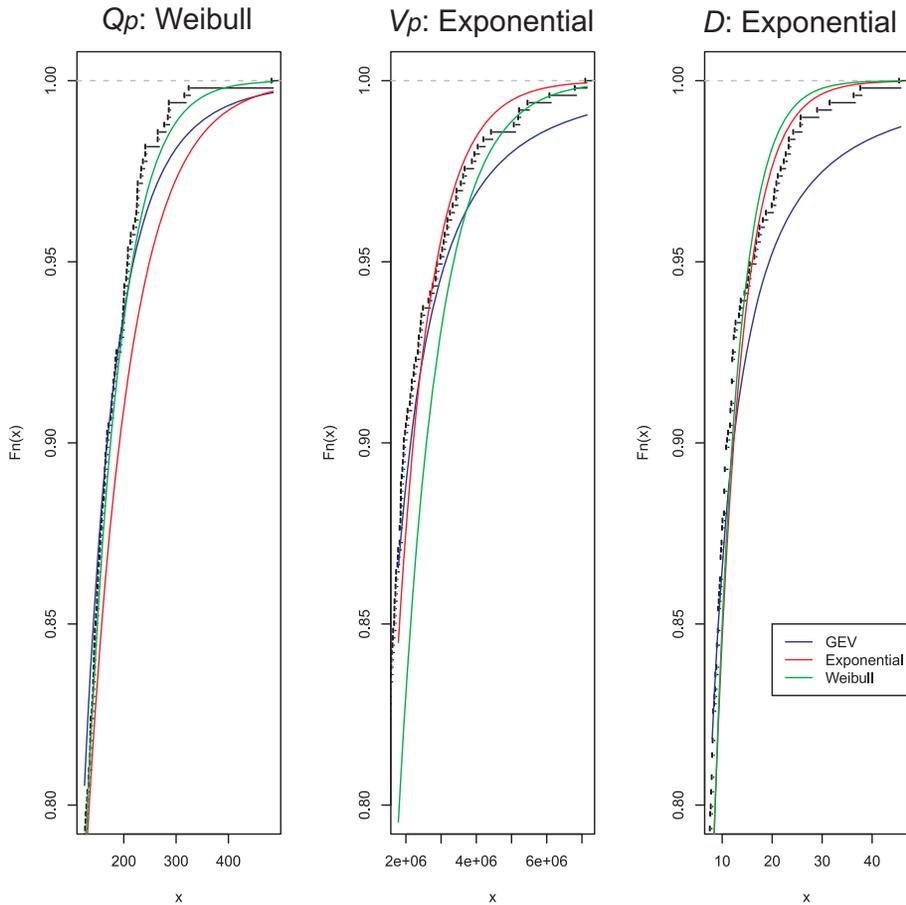


Fig. 2. The various EV distributions together with the empirical distribution for the three variables. The best fitting distribution is denoted in the title of each graph.

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



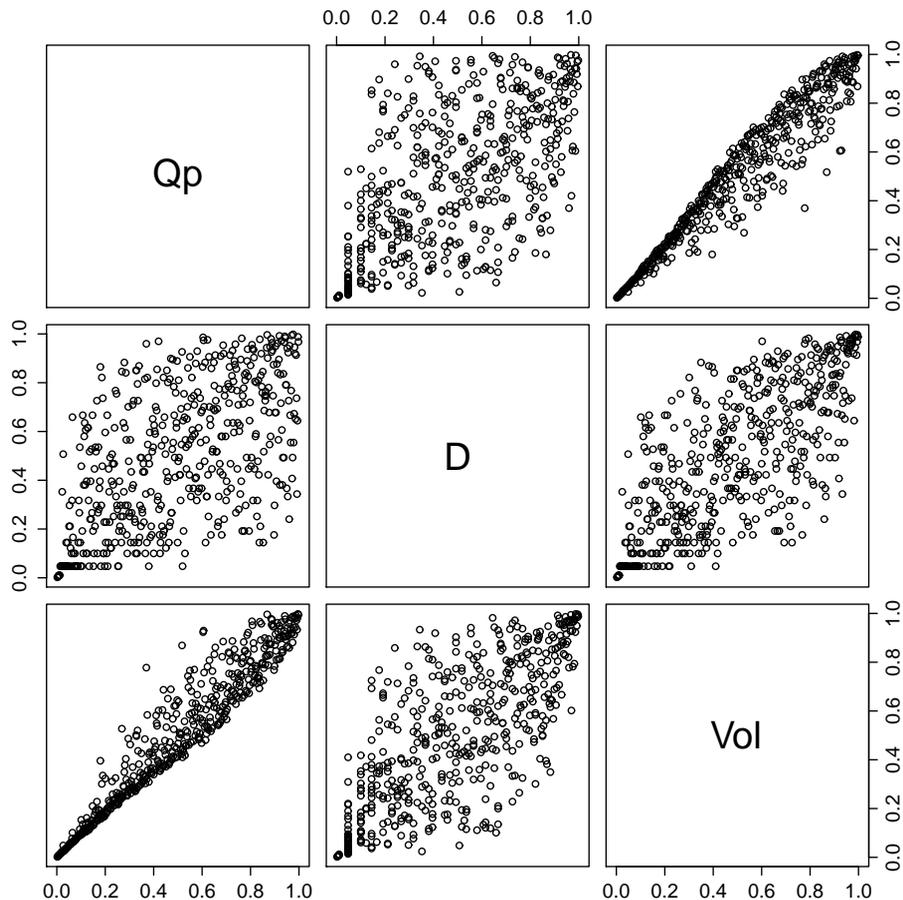


Fig. 3. Normalized rank scatter plots for all pairs of variables. Kendall's tau is 0.85 for (Q_p, V_p) , 0.42 for (Q_p, D) and 0.54 for (V_p, D) .

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



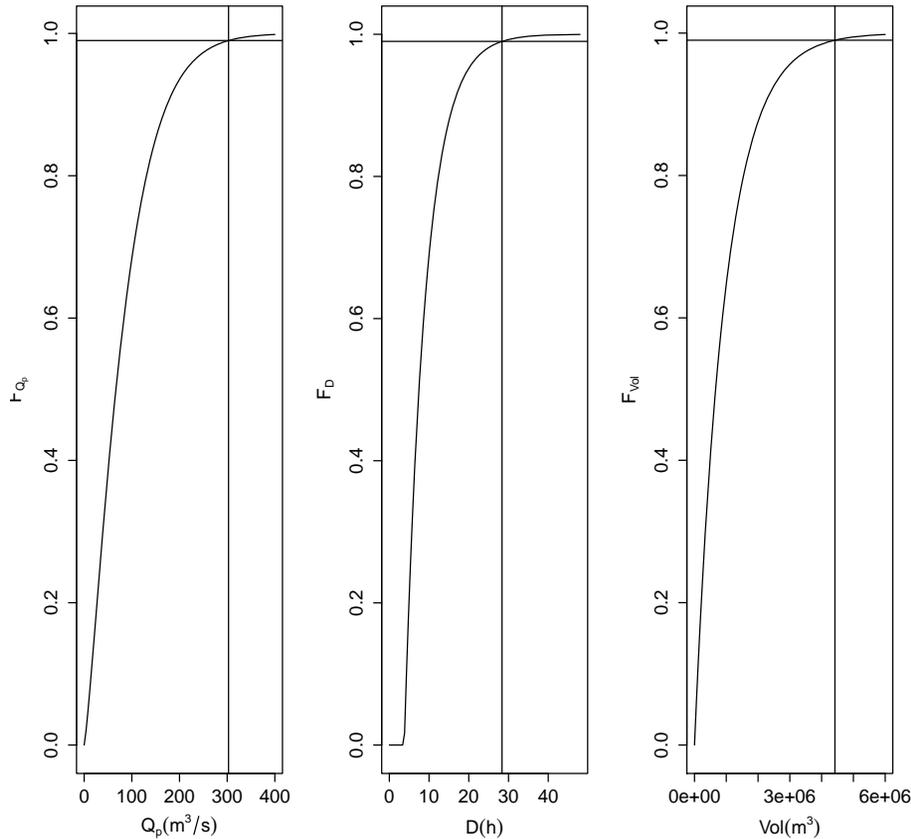


Fig. 4. The univariate design quantiles for Q_p and V_p for a return period of $T = 100$ yr ($p = 0.99$) are $302.9 \text{ m}^3 \text{ s}^{-1}$ and $4.42 \times 10^6 \text{ m}^3$, respectively.

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



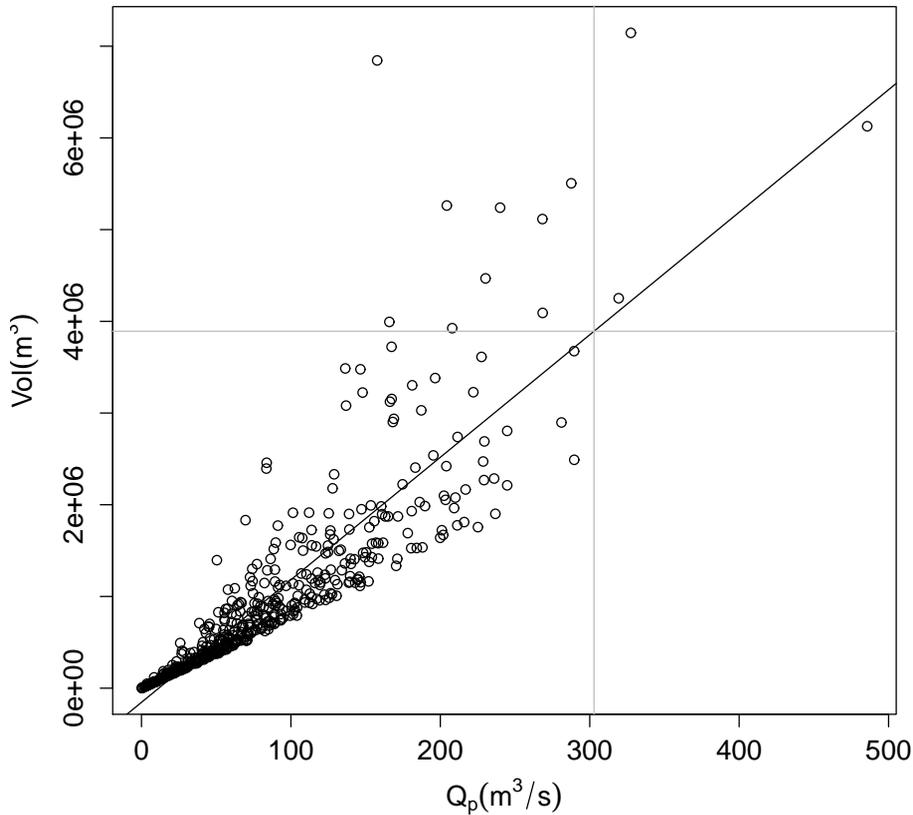


Fig. 5. Illustration of the derivation of the design quantiles based on the regression method.

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



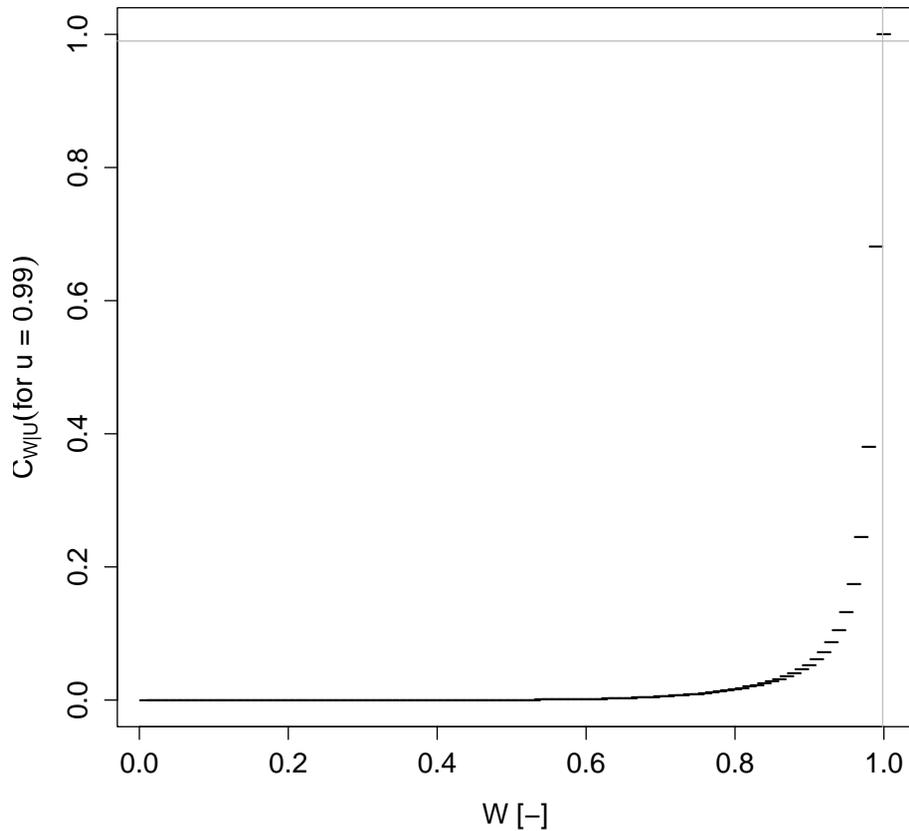


Fig. 6. Inversion of the copula-based conditional CDF $C_{W|U}$.

Selecting multiple design variables

S. Vandenberghe et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



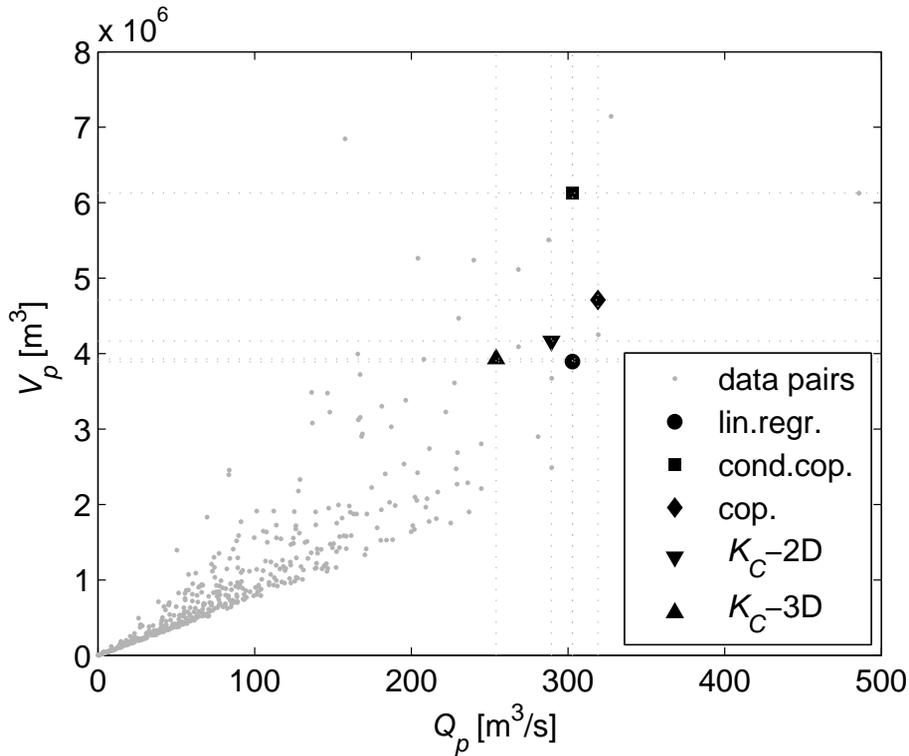


Fig. 9. An overview of the different design values for $T = 100$ yr obtained with the different methods.

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



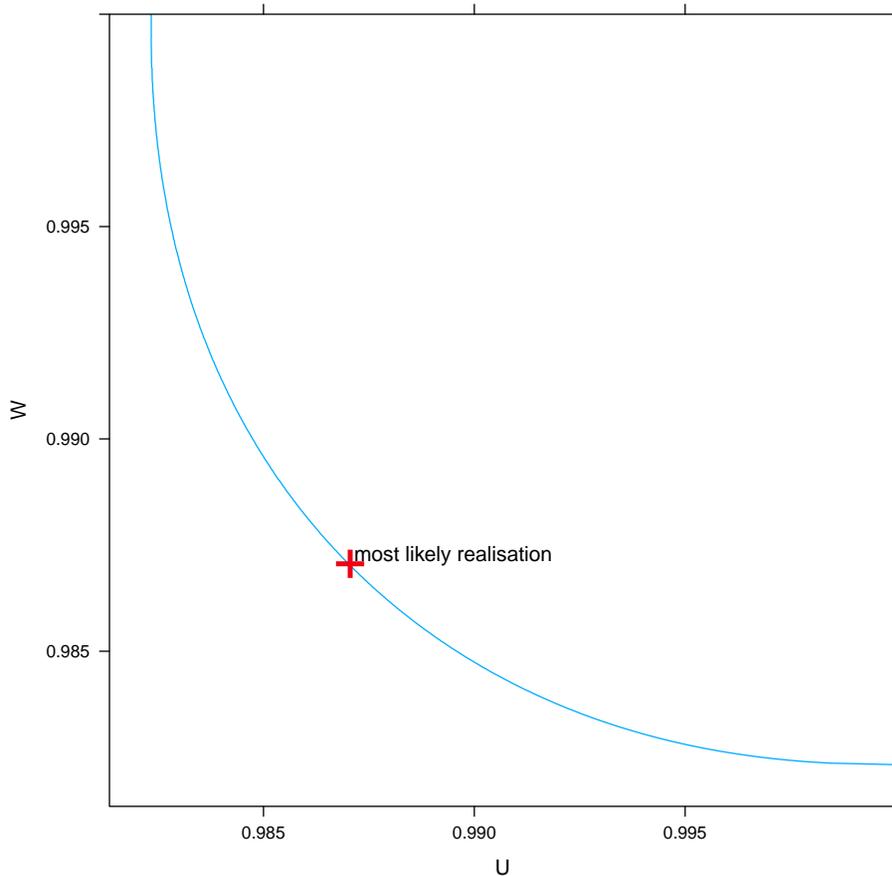


Fig. 10. t -level of the copula C_{UW} which corresponds to a JRP of 100 yr when considering the 2-D Kendall distribution function method, with indication of the most likely event.

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



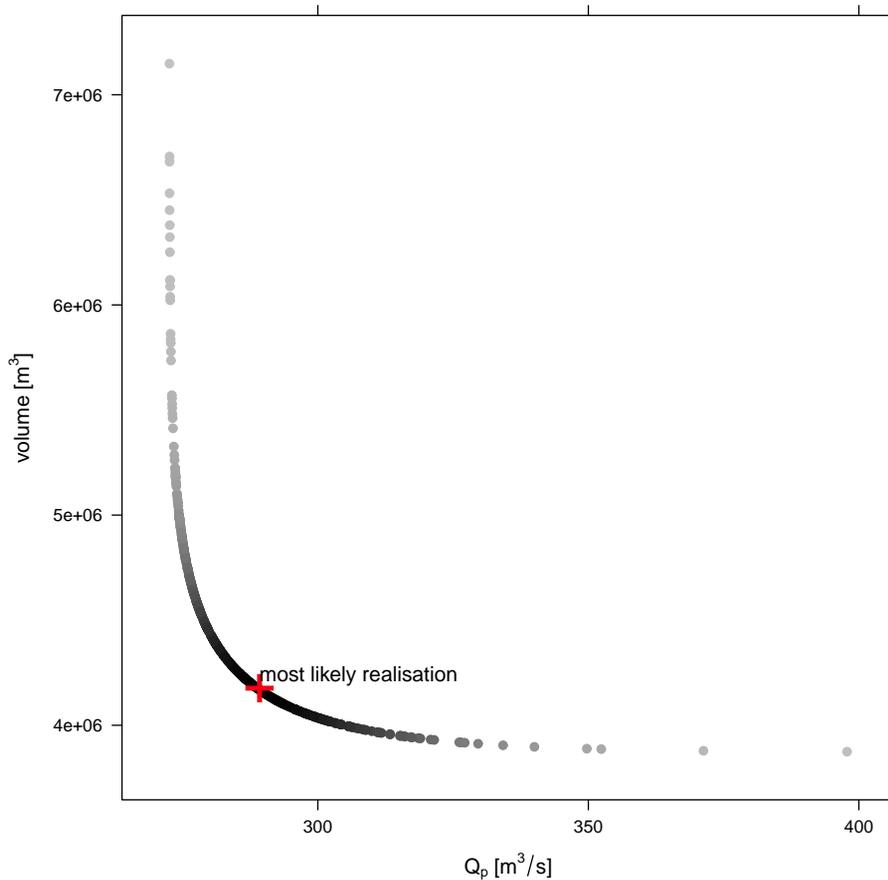


Fig. 11. An ensemble of 500 (q_p, v_p) pairs that all have a JRP of 100 yr. The density of the ensemble is given in greyscale: the more grey, the less events sampled. The most likely event is also indicated.

Selecting multiple design variables

S. Vandenberghe et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



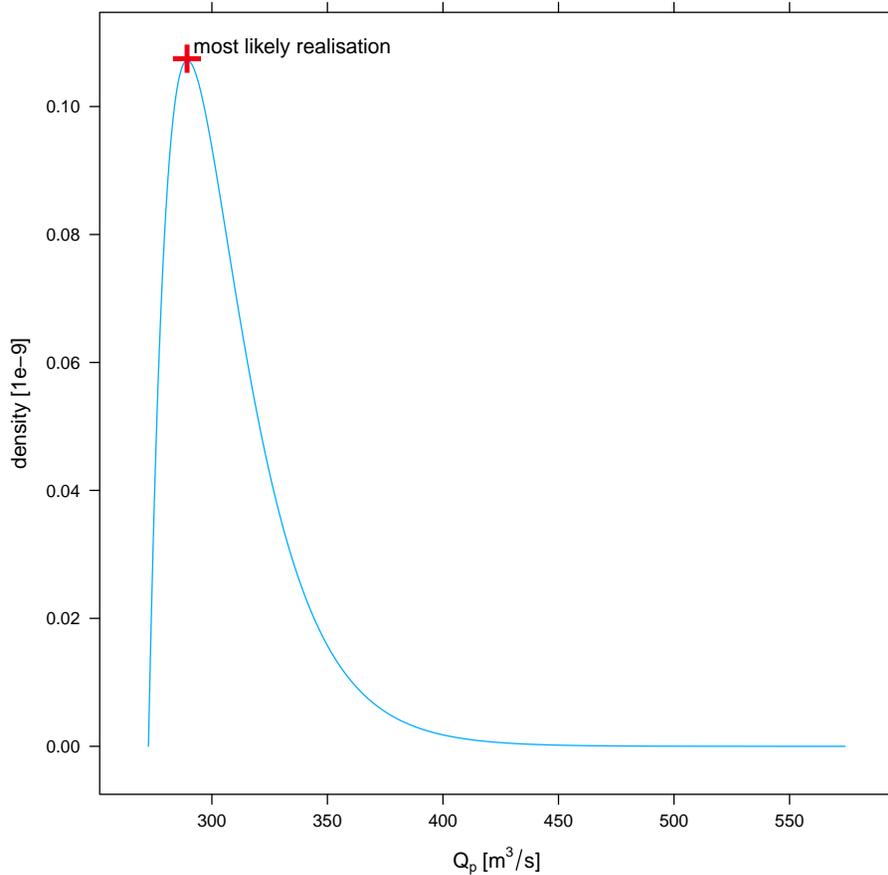


Fig. 12. PDF of Q_p in the ensemble. The most likely design discharge $q_{p,T}$ is indicated as well.

Selecting multiple design variables

S. Vandenberghe et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



