



Copulas for hydroclimatic applications - A practical note on common misconceptions and pitfalls

Faranak Tootoonchi¹, Jan Olaf Haerter², Olle Råty³, Thomas Grabs¹, Mojtaba Sadegh⁴, Claudia Teutschbein¹

- 5 ¹ Department of Earth Sciences, Uppsala University, Uppsala, Sweden
² Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark
³ Finnish Meteorological Institute, Helsinki, Finland
⁴ Department of Civil Engineering, Boise State University, Boise, USA
Correspondence to: Faranak Tootoonchi (faranak.tootoonchi@geo.uu.se)

10 **Abstract.** For most hydroclimatic applications, precipitation and temperature are of particular interest as they strongly affect the water cycle, can easily be measured and are often readily available from many meteorological stations worldwide. To account for precipitation and temperature variability, their co-dependence and their correlation, several multivariate analysis methods have been adopted in the hydroclimatic literature in recent years. In line with the steadily rising number of publications on this topic, the notion of copula-based probability distribution has also attracted tremendous interest to deal with the
15 complexity of compound events in the multidimensional context. A copula is a function that connects a multivariate distribution to its one-dimensional margins, which allows for a joint distribution of random variables with great flexibility for the marginal distribution. However, there seems to be a lack of comprehensive understanding of the fundamental requirements of the copula concept such as the strength and significance of correlation between variables, autocorrelation effects and the choice of representative copula families, which potentially compromises the robustness of projections of future environmental
20 processes and natural hazards. Therefore, by combining a systematic literature review with a specific hydroclimatic case study in Sweden, we illustrate a practical approach to copula-based modeling, which (1) provides end-users with an overview of necessary requirements, statistical assumptions and consequential limitations of copulas, (2) highlights possible pitfalls and misconceptions, and (3) offers a decision support framework for the application of copulas to support researchers and decision makers in addressing climatological hazards and sustainable development, thereby demystifying what is currently an area of
25 great confusion.

1. Introduction

Water touches every aspect of our lives, from public health to safety, to the foundation of our economy. Alterations to the hydroclimatic drivers of the water cycle form a potential threat, as they influence socioeconomic, ecological and climate systems (Vogel et al., 2015). In this paper, we adopt the definition of hydroclimatology as the “study of the influence of climate
30 upon the waters of the land” originally proposed by Walter Langbein in the 1970’s and later further expanded to include hydrometeorology (i.e., the land-atmosphere interface) as well as the surface and near-surface water processes of evaporation, runoff, groundwater recharge and interception (Wendland, 1987). As such, hydroclimatology encompasses a large number of



stylistically, it would improve the readability if you either add a space between all paragraphs, or indent the leading line

35 applications, e.g. (1) the assessment of a changing climate on water resources (including water quantity and quality), (2) the analysis of feedback of the water cycle variations on climate change, and (3) the evaluation of anthropogenic impacts such as land-use change on local climate conditions and water resources.

40 For most hydroclimatic applications, precipitation and temperature are of main interest for modeling and future projections as both variables can easily be measured and are typically readily available from many meteorological stations worldwide (see e.g. NOAA Climate Data Online) or from global gridded databases such as NCEP/NCAR (Kalnay et al., 1996), ERA-interim (Dee et al., 2011) or JRA-55 reanalysis (Ebita et al., 2011). Both variables also strongly affect the water cycle: precipitation as a direct component of the water cycle, and temperature as a proxy for energy availability, which in turn is of key interest for estimating evapotranspiration as well as snow accumulation and melt, both of which exert influence on the amount and timing of runoff. As a result, the spatiotemporal trends of precipitation and temperature greatly affect environmental processes, control natural hazards such as floods and droughts, and influence water quality and availability (Alidoost et al., 2019; Ribeiro et al., 2019). Therefore, future socio-economic development and sustainable use of water resources is highly dependent on reliable representations and modeling of these two variables (Alexandrov and Hoogenboom, 2000; Phiri et al., 2019; Satgé et al., 2019).

I could not find saturation water vapor mixing ratio this paper

50 On top of the natural variability of these meteorological variables, anthropogenic climate change has resulted in more surface water being transferred to the atmosphere, increasing the saturation water vapor mixing ratio there (Ashraf Vaghefi et al., 2014). These processes have exacerbated droughts (Trenberth et al., 2014) and intensified rainfall and flooding events (Mangini et al., 2018). To account for precipitation and temperature variability, numerous studies have focused on assessing these two variables separately at different spatiotemporal scales (Hundecha and Bárdossy, 2005; Maftai et al., 2011; Teutschbein and Seibert, 2012). However, such hydroclimatic variables are interconnected and have a combined impact on the water balance. Focusing on these variables separately ignores their co-dependence, which may lead to an insufficient characterization of their variability and co-dependence (Hao et al., 2018). Hence, to consider the co-dependence and correlation between these two variables, there have been attempts to study them simultaneously in a compound context. For example, Benestad and Haugen (2007) examined the joint probabilities of high temperatures and heavy precipitation during spring in Norway to evaluate flood hazards. Piani and Haerter (2012) studied compound precipitation and temperature data derived from different stations in Germany to assess and correct biases in climate model data by using their joint distribution. Zscheischler, Orth, and Seneviratne (2017) used bivariate return periods of precipitation and temperature over Europe to investigate crop yield variability. Boeck et al. (2011) explored the combined influence of precipitation and temperature on plant communities and concluded that the negative effects of a dry spell combined with a heatwave on plant growth and functioning were considerably greater than the sum of their individual impacts.

65 All these studies highlight the importance of jointly considering the compound effects of precipitation and temperature dynamics in hydroclimatic studies. However, an accurate assessment of the joint distribution of these two variables – especially in the context of a changing climate - is not straightforward, as their degree of co-dependence can vary with changes in spatiotemporal scales and for extreme events (Li et al. 2014). For example, Trenberth et al. (2003) argue that at large spatial



scales the intensity of extreme precipitation events should increase exponentially with temperature. However, because of the influence of local atmospheric circulations at local scales, precipitation and temperature are likely to be weakly correlated in many regions (Chou and Neelin, 2004; Colacino and Dell'osso, 1978; Dai and Bloecker, 2019). The strength of precipitation-
70 temperature correlation is additionally affected by the temporal resolution under consideration. At the annual resolution AghaKouchak et al. (2014) found a significant negative correlation between anomalies of temperature and precipitation over California. At the monthly resolution, Trenberth (2018) found a strong negative correlation between temperature and precipitation in the mid-latitude continental regions during summer. However at the daily resolution there is typically weak statistical evidence of a link between temperature and precipitation in many regions (Ferrari et al., 2018).

75 Despite these limitations at relatively fine spatio-temporal resolutions, several multivariate analysis methods have been adopted in the hydroclimatic literature in recent years (Fig. 1(a)). A keyword search in Scopus (Burnham, 2006), resulted in 914 published papers since 1970 until the 1st of June 2020 when combining the terms '*multivariate*' and '*hydrology*', and 6,090 publications when combining the terms '*multivariate*' and '*climate*'. It should be noted that the total number of papers published per year has nearly doubled since 2010 (Fig. 1(a)). These papers utilize a number of different multivariate methods,
80 including, for instance, multivariate analysis of variance (Shukla and Gedam, 2019), multivariate regression (Abdulelah Al-Sudani et al., 2019) or principal component analysis (Wang et al. 2019).

In line with the increasing use of multivariate methods, the application of copula-based probability distributions - a specific type of multivariate approaches - has also attracted tremendous interest (Fig. 1(b)) to deal with the complexity of compound events in the multidimensional space. The copula concept is particularly advantageous, because it allows for constructing a
85 joint distribution of random variables without any constraint on their marginal distributions (Fig. 2) (AghaKouchak et al., 2010). A copula is a function that connects a multivariate distribution to its one-dimensional margins (Schweizer, 1991). It is based on the early works of Sklar (1959) and has been widely used in different fields of study such as in finance (Cherubini et al., 2013) or in the energy sector (Ghoddusi, 2017). Since the early 2000's, copula methods have been adopted in hydrological modeling, which was triggered by the study of Salvadori and De Michele (2010). Since then researchers have applied copulas
90 to study interactions between a variety of different hydroclimatic variables related to e.g. sediment transport (Shojaeezadeh et al., 2020), flood hazard (Didier et al., 2019) heatwave mortality (Mazdiyasnani et al., 2017) or to study hydrology and climate in general (Fig. 1(b)). **Bardossy, with Li and then Pegram, published copula-based papers in 2008 & 2009**

Copulas have been used in both theoretical and applied analyses of hydroclimatic data. Theoretical studies adopted the copula-based methodology for instance to gain a better understanding of the dependence structures of temperature and precipitation,
95 in different parts of the world (Cong and Brady, 2012; Lazoglou and Anagnostopoulou, 2019). Applied studies adopted the copula-based methodology to examine e.g. agricultural droughts (Wang et al. 2019; Ribeiro et al. 2019) or joint effects of temperature and precipitation extremes on vegetation growth (Alidoost et al., 2019; Cong and Brady, 2012). In addition, numerous climate-change impact studies applied copulas to bias-correct future projections by global or regional climate models (Piani and Haerter 2012; Li et al. 2014; Gennaretti, Sangelantoni, and Grenier 2015; Vrac and Friederichs 2015; Rana,



100 Moradkhani, and Qin 2017; Chen et al. 2018; Li and Babovic 2019; Mesbahzadeh et al. 2019) even though the usefulness of multivariate bias-correction methods for hydrological purposes has recently been put into question (Räty et al., 2018).

Despite the growing number of applied studies that adopt copulas, practitioners have to rely on statistical literature, which often focuses on the theoretical concept behind copulas (e.g., Genest, Rémillard, and Beaudoin 2009 or Berg 2009) while only a few papers also relate to practical challenges encountered in hydroclimatic research (e.g., Dupuis 2007). Hence, there seems to be a lack of comprehensive documentation in the scientific hydroclimatic literature that (1) provides an overview of necessary requirements, statistical assumptions and consequential limitations of copulas, (2) clarifies some of the common misconceptions around copulas, and (3) offers clear guidelines on how to implement copulas, in particular for studies of future hydroclimatic impacts.

Accordingly, this paper aims at filling this gap and serves as an overview of the state of the art of using copulas in hydroclimatology for practitioners interested in adopting this method for their research. We aim to illustrate a practical approach on copula-based modeling of hydroclimatic variables exemplified by precipitation and temperature. We introduce copula concept and bring forward a step-by-step copula approach (section 2). Based on a systematic literature review and a case study in Sweden we also explain common pitfalls and misconceptions (section 3), and provide a decision support framework for applying copulas (section 4) to support researchers and decision makers in addressing climatological hazards and sustainable development.

the

2. Step by step copulas

2.1. What is a copula?

A copula is a mathematical function that couples marginal distributions and represents the joint probability distribution (Fig. 2). Copulas have been introduced in the early work of Sklar (1959), who showed that any multivariate joint cumulative distribution function (CDF) could be specified as a copula function of the margins, assuming that the margins are continuous, time-independent and uniformly distributed on the [0, 1] interval. Therefore, if H is an m -dimensional cumulative distribution function with one-dimensional margins denoted as F_1, F_2, \dots, F_m , then an m -dimensional copula (C) exists such that:

$$H(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m) = C(F_1(\mathbf{x}_1), F_2(\mathbf{x}_2), \dots, F_m(\mathbf{x}_m)) \quad (1)$$

for which $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)$ are vectors of inputs.

125 If $F_1(\mathbf{x}_1) = u_1$, $F_2(\mathbf{x}_2) = u_2$ and $F_m(\mathbf{x}_m) = u_m$, the copula density function, c can be derived as (Shojaeezadeh et al., 2018):

$$c(u_1, u_2, \dots, u_m) = \frac{\partial^d C(u_1, u_2, \dots, u_m)}{\partial u_1 \dots \partial u_m} \quad (2)$$

Then, the joint density function can be decomposed as:

$$f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m) = c(u_1, u_2, \dots, u_m) \prod_{k=1}^m f_k(\mathbf{x}_k) \quad (3)$$

where f_k denotes the density function of the k^{th} dimension.



130 Although copulas can be defined for any number of variables n (so-called n -dimensional copulas), two-dimensional copulas that link only two variables, such as temperature and precipitation, are much more common in hydroclimatic applications and therefore **are** the focus of this paper. In the following sections, we give a brief overview of both empirical and theoretical copula distributions.

2.2. Empirical copula

135 To capture the underlying joint probability between two random variables, the joint rank correlation of the variables, termed *empirical copula* (Genest and Favre, 2007), needs to be derived. The empirical bivariate copula is defined as the discrete function C_n given by:

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{|\{r_{1i} \leq R_1 \wedge r_{2j} \leq R_2\}|}{n} \quad (4)$$

where r_1 and r_2 denote the order statistics of the sample and provides the cardinality of the subsequent set and i and j label the individual sample elements. The result would be the joint CDF of uniformly distributed marginal (Charpentier et al., 2007). According to Eq. (4), the empirical copula can be interpreted as the cumulative distribution of rank-transformed data.

2.3. Theoretical copula families

The governing pattern of the empirical copula can be used to guide decisions towards suitable theoretical copulas or so-called copula families. The two most widely used copula families in hydroclimatic modeling studies are *Elliptical* and *Archimedean* copulas. These two families are different in complexity, tail behaviour, dependence structure and their theoretical parameter estimation method (Sadegh et al., 2017). In order to demonstrate how these families behave in shape and tail dependence, 1000 normally-distributed points with Pearson correlation coefficient, $\rho = 0.75$ were generated. Thereafter, the five most widely used copulas were fitted to the data (Fig. 3).

It should be noted that the parameters of bivariate copulas reflect the strength of mutual dependence between two variables (e.g., between temperature and precipitation). To account for the degree of dependence between two variables, a critical assessment of their correlation is essential. The strength of dependence can be quantified through empirical measures, including Pearson's correlation coefficient ρ , Kendall's τ , or Spearman's S .

Pearson correlation (Pearson 1920) represents the linear correlation between x_1 and x_2 as follows:

$$\rho_{x_1x_2} = \frac{\text{Cov}(x_1, x_2)}{\sigma_{x_1}\sigma_{x_2}} \quad (5)$$

155 *Kendall's τ* (Kendall 1938) and *Spearman's S* (Spearman 1904) are measures of rank correlation between the ranks of two variables, which can be expressed as:

$$\tau = \frac{\text{number of concordant pairs} - \text{number of discordant pairs}}{\frac{n(n-1)}{2}} \quad (6)$$

$$S = 1 - \frac{6 \sum_1^n d_i^2}{n(n^2-1)} \quad (7)$$



where d_i is the difference between two ranks and n is the number of observations. These correlation measures can further be
160 used as a guide to estimate copula parameters for different copula families.

2.3.1. Archimedean copulas

Archimedean copulas are flexible tools to capture various dependence structures, e.g. concordance and tail dependence (Hofert,
2008). This makes them especially suitable for the modeling of extreme events.

The general form for Archimedean copulas is:

$$165 \quad C(u_1, u_2; \theta) = \psi^{-1}(\psi(u_1) + \psi(u_2)), \quad u \in [0, 1] \quad (8)$$

where θ is the selected Archimedean copula parameter, ψ is defined on $(0, 1]$ is a continuous strictly decreasing convex
function, known as the generator function and its pseudo-inverse ψ^{-1} (Genest and Rivest, 1993).

Aloui, Ben Aïssa, and Nguyen (2013) distinguished between three different types of Archimedean copulas (Fig. 3):

The Gumbel copula (Gumbel 1960) is an asymmetric copula with higher probability concentrated in the upper tail. It is given
170 by:

$$C(u_1, u_2) = \exp\left\{-\left[(-\ln u_1)^\theta + (-\ln u_2)^\theta\right]^{\frac{1}{\theta}}\right\} \quad (9)$$

The Clayton copula (Clayton 1978) has higher probability in the lower tail and is expressed as:

$$C(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}} \quad (10)$$

The Frank copula (Frank 1979) tends to have lower density in the tails compared to Clayton and Gumbel copulas. Also, it is
175 capable of retaining negative correlations. Frank copula is defined as:

$$C(u_1, u_2) = \frac{-1}{\theta} \ln\left(1 + \frac{(\exp(-\theta u_1) - 1)(\exp(-\theta u_2) - 1)}{\exp(-\theta) - 1}\right) \quad (11)$$

In the Archimedean copulas, the copula parameter θ can be derived from Kendall's τ (Allen et al., 2017). Depending on the
type of Archimedean copula, parameter θ takes different values and has different relationships to Kendall's τ (Table 1).

After estimating θ based on Kendall's τ (Table 1), the copulas can be computed using their respective formulas.

180 2.3.2. Elliptical copulas

Elliptical copulas are widely used in hydroclimatic modeling, as they are symmetric and capable of representing positive and
negative correlation. They are, however, less suitable to analyse and model hydroclimatic extremes, when there is evidence of
a sharp tail dependence.

The governing function for Elliptical copulas is:

$$185 \quad C(u_1, u_2) = \Phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \quad (12)$$

where Φ is a suitable multivariate distribution and Φ^{-1} denotes the inverses of univariate marginal distributions. Several
methods varying in accuracy and complexity have been suggested to estimate the associated copula parameters for Elliptical



copulas. For example, Aas (2004) suggested a Monte-Carlo-based method for parameter estimation, whereas Weiß (2011), used a Maximum likelihood approach.

190 The two most widely used types of Elliptical copulas are Gaussian and student-t copulas.

The *Gaussian copula* is expressed as:

$$C(u_1, u_2) = \Phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\theta^2}} \exp\left(-\frac{x_1^2 - 2\theta x_1 x_2 + x_2^2}{2(1-\theta^2)}\right) dx_1 dx_2 \quad (13)$$

where Φ is the standard bivariate normal distribution with linear correlation coefficient θ and is restricted to the interval $(-1, 1)$. There is no closed form to estimate the associated copula, but there are numerous ways to estimate it

195 numerically (Aloui et al., 2013).

The *Student-t copula* is given by:

$$C(u_1, u_2) = \Phi(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_2)) \\
 = \int_{-\infty}^{t_v^{-1}(u_1)} \int_{-\infty}^{t_v^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\theta^2}} \left(-\frac{x_1^2 - 2\theta x_1 x_2 + x_2^2}{v(1-\theta^2)}\right)^{-\frac{v+2}{2}} dx_1 dx_2 \quad (14)$$

200 The Student-t copula is characterized by a higher density in tails and an increased probability of joint extreme events compared to the Gaussian copula (Aas, 2004). This is reflected by the added v parameter, which represents the degrees of freedom. $t_v^{-1}(u)$ denotes the inverse of the CDF of the standard univariate Student-t distribution. By increasing the value of v , the co-dependence strength in the tails will decrease.

Embrechts, Mcneil, and Strauman (1999) estimated the Elliptical copula parameter from the rank-based Kendall's τ or Spearman's S . Later, Aas (2004) showed that the co-dependence structure for Elliptical copulas can be presented by linear
 205 (Pearson) correlation. Correspondingly, their copula parameter θ can either be estimated as being equal to the linear (Pearson) correlation or derived from Kendall's τ or Spearman's S . For more details on the corresponding equations, we refer the readers to Aas (2004).

2.4. Copulas as a basis for computing conditional distributions

210 In a bivariate case, copulas can be used to compute the conditional distribution of one variable given a specific value of second variable. An example is the estimation of the conditional probability distribution of precipitation, given one particular temperature probability and the known joint probability distribution.

The conditional probability of x_2 , given the occurrence of x_1 , can be shown as:

$$f(x_2|x_1) = \frac{f(x_1, x_2)}{f(x_1)} \quad (15)$$

215 where $f(x_1, x_2)$ gives the joint density of x_1 and x_2 , while $f(x_1)$ gives the marginal density of x_1 . Considering Eq. (3) the conditional copula function is reorganized as:

$$f(x_2|x_1) = \frac{c(u_1, u_2)f(x_1)f(x_2)}{f(x_1)} = c(u_1, u_2)f(x_2) \quad (16)$$



In practice, the conditional distribution can be obtained via Monte Carlo simulation, which is a numerical approach that relies on repeated random sampling (Aas, 2004). In order to compute the conditional probability from the copula, the following steps should be taken (Salvadori et al., 2007):

- 220
- Generate two independent sets of random variables u_1 and r on $(0,1)$ domain.
 - Set $u_2 = c_{u_1}^{-1}(r)$
 - Repeat until the latter is solved.

It is noteworthy to mention that in case of the Elliptical copulas, the conditional probability can be calculated by Cholesky decomposition of the Elliptical copula parameter (Li and Babovic 2019).

225 2.5. Performance measures for selecting a suitable copula family

After fitting copulas to the data, some performance measures are used to analyze the goodness of fit obtained from different copula families and parameters. This can be done by comparing cumulative distributions of empirical and theoretical copulas either parametrically or non-parametrically (Genest et al., 2009). The Cramér von Mises (S_n) (Eq. 17) and Kolmogorov Smirnov (D_n) (Eq. 18) statistics are two of the widely used performance measures for copula parameter estimation (Madadgar and Moradkhani, 2014):

$$S_n = \sum_{i=1}^n |C_n(x) - C_\theta(x)|^2 \quad (17)$$

where C_n is the Empirical joint distribution and C_θ is the fitted copula and

$$D_n = \sup |C_n(x) - C_\theta(x)| \quad (18)$$

in which \sup is the supremum of the distances at point x .

235 However, it is also important to evaluate the p-values of the applied performance measures (Berg, 2009; Genest et al., 2009). The p-value of Cramér von Mises or Kolmogorov Smirnov can be found by bootstrapping via Monte Carlo approach by following the steps proposed by Genest, Rémillard, and Beaudoin (2009). The null hypothesis of the test is the acceptance of the parametric copula. For a particular copula, the p-value is sufficient to determine the acceptance or rejection of the null hypothesis with the significance level of α , but in a group of different acceptable copulas, the best alternative is the one with the smallest S_n or D_n and the greatest p-value (Madadgar and Moradkhani, 2014).

In assessing the performance of copula models, using only one performance measure can be misleading, therefore it is recommended to use an ensemble of measures (Sadegh et al., 2017). The Nash-Sutcliffe Efficiency (E_{NS}) (Eq. 19) and Root Mean Square Error (E_{RMS}) (Eq.20) are two widely used performance measures in hydroclimatic applications, which can be adopted in copula model selection as well.

$$245 \quad E_{NS} = 1 - \frac{\sum_{i=1}^n |C_n(x) - C_\theta(x)|^2}{\sum_{i=1}^n |C_n(x) - \bar{C}_n(x)|^2} \quad (19)$$

$$E_{RMS} = \sqrt{\frac{\sum_{i=1}^n |C_n(x) - C_\theta(x)|^2}{n}} \quad (20)$$

very well explained and in-depth summary/tutorial of copulas and their benefits - nice job



3. Common issues, misconceptions and pitfalls

To identify common issues, misconceptions and pitfalls in the application of copulas in the existing hydroclimatic literature, a systematic literature review was conducted. We reviewed all articles returned from the Scopus search engine using the keywords “Copula + Precipitation + Temperature”. As of 1st of June 2020, these keywords resulted in a list of 85 papers, of which 16 were not accessible due to missing permissions. To examine the applied copula methods and underlying assumptions, the remaining 69 publications were investigated to evaluate how physical characteristics had been considered in these publications and if they fulfilled the statistical requirements necessary to adopt a copula framework. In order to do so, the following six aspects were assessed in detail.

255 *Physical characteristics:*

- 1- What is the **spatiotemporal scale** of the study?
- 2- Does the paper analyse and interpret the **correlation** between temperature and precipitation?

Statistical requirements:

- 3- Does the paper consider **autocorrelation** within each of the variables (i.e., autocorrelation of temperature respective to autocorrelation of precipitation)?
- 4- Does the paper consider zero-precipitation, e.g. by introduction of a **precipitation threshold** and deals with **data with the same rank**?
- 5- Does the paper reflect on the **significance of correlation** and **stationarity** issue?
- 6- Are several copula families fitted and **checked for errors**?

265

For each of these aspects, we first explain its significance. Thereafter, we demonstrate potential consequences, if not properly addressed, based on a real hydroclimatic case, namely the Vattholma river catchment in Southcentral Sweden (Fig. 4) close to the city of Uppsala and approximately 70 km north of the capital Stockholm. For this catchment, we constructed a copula of observed temperature and precipitation data for the climate normal periods 1961-1990. Daily precipitation and temperature data were downloaded from a freely available database (www.vattenwebb.smhi.se) provided by the Swedish Meteorological and Hydrological Institute (SMHI). According to this data, the catchment is characterized by a warm-summer continental climate with an average annual temperature of 5.2°C and an annual precipitation sum of 633 mm.

270

3.1. Spatiotemporal scale of the 69 papers

First, each **paper** was scanned for the spatial scale of the investigated study site(s). Roughly 56% of the reviewed papers were based on meso- to large-scale studies (i.e., > 300 km²), while the remaining 44% were conducted on small scales (i.e., < 300 km²). The spatial scale is of importance, because it affects some of the statistical requirements of copulas that are explained further below.

275



Our showcase catchment Vattholma has an area of 293 km^2 , thus, can be considered small-scale. The catchment is predominantly flat with altitudes ranging from 25 to 65 meters above sea level. It is mostly covered by forest (81%), while the remaining area includes 10% lakes/wetlands, 7% open land and 2% populated areas (Fig. 4).

In the second step, we checked for the temporal resolution (i.e., hourly, daily and monthly) of the published literature. We found that only 3% (2 publications) considered hourly values of temperature and precipitation. Most (66%) used daily values, whereas 31% considered monthly values. The temporal resolution, similar to the spatial scale, is crucial when it comes to some of the statistical properties such as correlation and autocorrelation.

3.2. Correlation

Copulas are built on the co-dependence between variables and it is crucial to examine the correlation between the variables (Eqs. 5-7). For precipitation and temperature, the co-dependence often varies regionally and depends on the spatial scale at which the data are recorded. It is also important to check for significance of correlation because it allows to examine if correlation strength is independent of the selected sample. But it is not always possible to find a significant statistical correlation between precipitation and temperature at a given scale (Serinaldi, 2016).

Out of the 69 reviewed papers, 29% did not indicate the strength or significance of their correlation by any of the aforementioned correlation measures. 25% mentioned that their data shows some degree of correlation, but did neither specify the strength nor the significance of any copula parameters calculated based on the correlation value. Less than half of all papers (46%) transparently stated the correlation strength and represented the correlation of the data either graphically or with one of the dependence empirical measures. Only 23% presented the strength and discussed the statistical significance of the correlation, e.g. by checking the p-value of their correlation measure.

For the Vattholma catchment, the correlation strength depends on the temporal resolution. We also checked the correlation significance and determined that in case of Vattholma catchment there are months in which the correlation is not significant at the 5% level, thus cannot be considered as reliable correlation between variables. Furthermore, in the daily resolution the correlation is generally weaker than monthly resolution (Fig. 5(a), Lines 1 and 2). As the Pearson correlation was weaker during winter, we considered the co-dependence structure to be addressable in the summer. We choose to use the daily resolution to further generate copula co-dependence structure, because daily data can readily be used for finer resolution hydrological purposes (e.g., bias correction, daily runoff estimation).

3.3. Stationarity of correlation

Copulas can also be used for multivariate frequency analyses to calculate joint return periods of extreme variables as e.g. rainfall intensity and duration of individual precipitation events or other variables (e.g., Salvadori and De Michele 2010 or AghaKouchak et al. 2014). In these cases, copulas are established for a set of available data within a certain time frame to draw conclusions about the frequency of certain events that depend on the combined influence of two variables such as temperature and precipitation in case of droughts. However, in these cases stationarity is of importance, because findings from

clever - a 'must read' paper - my best this year

one to determine whether the

did you try lagging the daily precipitation and streamflow data? the delay depends on size of catchment...



310 such an analysis are commonly used for future risk analyses. Whether or not this underlying stationarity assumption is actually met, is subject to debate (e.g., McCarl, Villavicencio, and Wu 2008; Vaze et al. 2010 or Lins and Cohn 2011). As McCarl et al. (2008) framed it, stationarity implies that “natural systems fluctuate within an unchanging envelope of variability”. However, given the scale of ongoing climate change and the projected changes in the climate system that we are likely to face in the future, the assumption of stationarity is possibly not met, which could affect the applied risk measures (Cooley, 2013).
315 In the case of copulas, if the underlying idea is to draw conclusion for potential futures, it is crucial to check for changes in the correlation strength over time and the statistical significance. In other words, when the correlation is highly sensitive to the selected time period, it is an indication of a non-stationary behaviour. Nevertheless, we found only 5 of all 69 reviewed papers (7%) touching upon the assumption of stationarity of correlation or discussing changes in correlation over time.

To check if the correlation was stationary in our Vattholma showcase, we followed a random selection technique: We first
320 selected a sample of 10 random years (preserving the order of days within a year but without necessarily selecting consecutive years) from the 30 years period. Thereafter, we computed the correlation between daily resolution variables within each month in the chosen sample. We repeated this procedure 1000 times and found that in many months, correlation was changing significantly across different sets of 10 years (Fig. 5(b)). For example, in October and November, the correlation varied from positive to negative values. We also assessed the significance of the correlation in each month by checking the average p-value
325 of the 1000 samples. In June and July, the correlation had the lowest average p-value and is relatively strong, therefore it could be concluded that correlation within these months was significant. We adopted a copula framework on July, because it has significant correlation at both daily and monthly resolutions (Fig. 5(a)-(b)).

probably due to rain in summer rather than snow in winter?

3.4. Autocorrelation

Autocorrelation describes how time series are dependent on a delayed copy of themselves. According to Cong and Brady
330 (2012), the presence of autocorrelation reduces the efficiency of copula, because it boosts the variances of residuals and estimated coefficients. In order to benefit from the copula concept, variable sets should be *time-independent* (i.e., not based on autocorrelation) (Mao et al., 2015).

In some studies, copulas have been applied to autocorrelated data using so-called serial copulas (Genest and Rémillard, 2004; Ghoudi and Rémillard, 2004) or autocopulas (Rakonczai et al., 2012). But the mathematical procedure to drive these copulas
335 is quite demanding and further uncertainties may emerge (Rakonczai et al., 2012). In contrast, other papers argued that copulas are still applicable in case of only low-degree (removable) auto-correlation in the time series (Laux et al., 2011; Vogl et al., 2012).

Therefore, it is important to consider the degree of autocorrelation in the studied data. However, 75% of the reviewed papers did not mention if their data exhibited autocorrelation. The remaining 25% did check for autocorrelation, but only about half
340 of those addressed it by removing it or by bootstrapping.

For the Vattholma catchment, the autocorrelation function (ACF) for precipitation (Fig. 6(a)) and temperature (Fig. 6(b)) in July indicates some autocorrelated behavior. In order to account for autocorrelation, Laux et al. (2011) suggested using a first-

to check the effect of autocorrelation you might like to look at: Sugimoto, Takayuki, Andrés Bárdossy, Geoffrey G.S. Pegram, and Johannes Cullmann (2016), Investigation of hydrological time series using copulas for detecting catchment characteristics and anthropogenic impacts, *Hydrol. Earth Syst. Sci.*, 20, 2705-2720, doi:10.5194/hess-20-2705-2016.



order autoregressive model AR(1) to synthesize time-dependence in the time series. Then the synthesized autocorrelated data is subtracted from the original set. We followed this procedure to create a data set free from autocorrelation, which we validate
345 by plotting the ACF of the residuals (Fig. 6(c)-(d)), indicating that the data is now time-independent and can be used in the copula approach.

3.5. Correlation for data with the same rank (dry days)

Some studies have concentrated on the joint distribution of precipitation and temperature on wet days (i.e., Piani and Haerter 2012; Li et al. 2014; Rätty et al. 2018). In other words, the joint distribution has been constructed between precipitation values
350 larger than zero and the associated temperature on the same day. As mentioned earlier, copulas must be built on continuous sets of variables, which is shown in Eq. (4) and is also reflected in the method for constructing empirical copula, which is based upon rank dependence between variables (see section 2.2). Therefore, data points with the exact same precipitation (i.e., most commonly dry days with zero precipitation) are problematic as they lead to several ranks associated with temperature but only one rank for zero precipitation. Depending on the region studied, there can be a considerable number of zero-
355 precipitation (dry) or drizzle days. Therefore, a precipitation threshold needs to be defined, below which all data points should be removed. Removing such data can often eliminate a large portion of the precipitation-temperature data set, especially in drier regions. Accordingly, some efforts have been made to control data which have the same rank by adopting some statistical methods. For instance, Salvadori, Tomasicchio, and D'Alessandro (2014) suggested a jittering algorithm before computing the copula parameters, in order to perturb the data that have the same rank. Regardless of the chosen method to deal with data
360 with the same rank, an added post-processing step is probably needed that also handles the temperature data points that are not considered in the construction of the copulas because they fall below the precipitation threshold. This can be considered as a source of added uncertainty, especially if the data is to be later used for bias correction or crop yield estimation. Therefore, we
recommend? suggest that each published study transparently explains what part of the data has been used for constructing the copula (e.g., only days with precipitation above zero or some threshold) and how the remaining data has been treated.

365 However, our literature review revealed that 76% of the publications did not mention this particular issue. Only 24% of the studies transparently stated that dry/drizzle days are excluded, and only 9% explained how the remaining dry days were handled. For example Li and Babovic (2019) later handled the eliminated data points by scaling the associated temperature distribution to the full range of temperature data. This can be considered an applicable alternative, but the added uncertainties should not be neglected. Of course, it should be mentioned that 21% of the reviewed papers focused on the extreme part of the
370 joint distribution, therefore there was no need to focus on the complete range of data and the effect of neglected variables.

3.5.1. Influence of geography, spatial scale and temporal resolution

It should be noted that whether or not the data includes zero-precipitation or drizzle days is largely dependent on the study region, the spatial scale and the temporal resolution, e.g. arid regions or smaller domains with few measurement stations. In arid regions the occurrence of data points with no precipitation is considerably higher. Similarly, at higher temporal resolution



375 (e.g., hourly or daily), more data points will fall below the threshold. For example, in the Vattholma catchment, the number of wet days varied from month to month, ranging from 42% to 64%. We here applied a precipitation threshold of 0.1 mm to remove drizzle days.

Excluding zero-precipitation or drizzle days affects the correlation between precipitation and temperature. Since a significant correlation is the basis for applying copulas, it is essential to re-check the correlation after removal. In our showcase Vattholma, 380 for daily resolution data, correlation in the entire 30-year period changed only for some months after removing drizzle days (Fig. 5(a), line 3). The correlation in June-August was highly affected by removing days without precipitation as there were more dry days in these months.

3.6. Selecting suitable copula families

As mentioned earlier, in order to assess the copula performance, a validation of the copula model compared to the empirical 385 joint distribution is vital. In the reviewed literature, 15% of the papers used the empirical copula as the co-dependence structure, therefore it was not necessary to validate the fitted copula families. However, 28% of all papers fitted a copula on the data without documenting any performance measures. 57% checked for the goodness of fit of different copulas using at least one performance measure.

We here demonstrate how such performance measures can be applied practically. For the Vattholma case study, different 390 copula families were adopted to construct the joint copula between daily temperature and daily precipitation data in July. Since these variables were negatively correlated, the Student-t, Gaussian and Clayton copulas were selected as potentially suitable copula families, because they are capable of addressing negative correlation as well. Thereafter, the fitted copulas were visually compared to the empirical copula at (0 1] scale (Fig. 7(a)).

After transforming to the original scale, corresponding precipitation and temperature and their density plots were estimated 395 (Fig. 7(b)). Thereafter, we computed the earlier-described set of performance measures: E_{RMS} , E_{NS} , S_n and D_n statistic (Table 2). We also checked for the significance of calculated S_n at a significance level of $\alpha = 0.05$. Three of these measures (i.e., E_{RMS} , E_{NS} and D_n) resulted in the same ranking of the copulas, namely the Student-t copula having the best fit and the Clayton having the worst. Only the S_n criterion ranked the Gaussian copula first (Table 2).

Here, evaluation of copulas by E_{RMS} or E_{NS} did not reveal the weak performance of the Clayton copula that is shown by the other 400 performance measures, underlining the need to assess the goodness of fit by multiple performance measures.

4. Decision support framework for applying copulas

Properly accounting for the joint distribution of precipitation and temperature is of great importance for a range of different hydroclimatic applications. Yet, we highlighted a number of common issues, misconceptions and pitfalls that need careful handling to ensure an adequate, consistent and transparent use of copulas in hydroclimatology. Below, we briefly outline a 405 stepwise decision support framework (Fig. 8) for hydroclimatic copula applications with the goal of preventing users from stepping into the many statistical traps that copulas conceal. The proposed decision support framework has been developed to



minimize the incorrect application of copulas and to ensure reliable and comparable results, by guiding the end user through six generic steps (Fig. 8) that should typically be adopted before and during implementation of the copula approach, and by elucidating different alternatives that can be undertaken.

410 For each step in the framework (Fig. 8), in-depth explanations, mathematical equations, common pitfalls and scientific references for the various associated tasks can be found in the previous sections. A short narrative of each step is provided here: Prior to constructing bivariate copulas, it is worth reflecting on whether the two involved variables are controlled by related driving forces that could, at a relevant spatiotemporal scale, lead to a dependence structure between them (section 3.1). If so, their correlation can further be assessed with a particular focus on underlying key assumptions (section 3.2). In cases
415 where stationarity is relevant, it is crucial to investigate the correlation over time (section 3.3). Prior to implementation of the copula approach, it is further vital to test for autocorrelation of the variables and potentially remove any existing serial correlation (section 3.4). To construct an empirical copula based upon rank dependence between variables, the variables need to be sorted and ranked, which allows to detect data points with the same ranks that cannot be included when constructing the copula and that need to be handled separately (section 3.5). Thereafter, the correlation value can be used to estimate copula
420 parameters for different families. The fitted copula families are then compared to the empirical copula with help of one or more performance measure to decide on the best-fitting copula family and type (section 3.6).

5. Concluding remarks

This paper provides an overview of the state of the art of using copulas in hydroclimatology and highlights necessary requirements, statistical assumptions and limitations of using copulas in hydroclimatic studies. Based on a systematic literature
425 review, we identified common pitfalls and misconceptions, and highlighted ways to avoid them by providing a showcase study. We would like to mention that copula modeling, similar to any other modeling effort, is associated with a range of uncertainties stemming from observation errors, limited observations, and copula model structural and parameter uncertainty, among others, that translate into joint probability estimates and other derived inferences that should be considered (Sadegh et al., 2018). In our study, two general issues emerged. Firstly, important steps for deriving copulas (e.g., handling of precipitation-free
430 days) are often not reported, which makes many studies not reproducible. Secondly, key statistical assumptions required by copulas (e.g., significance and stationarity of correlation or absence of autocorrelation) are often not tested, which could severely undermine the credibility of certain study conclusions. Considering the growing interest in copulas in hydroclimatic applications, we emphasize the need to address these issues and provide a visual decision support framework to foster the transparency and reproducibility of scientific publications, while at the same time reducing the risk for inconsistent or even
435 incorrect use of copulas in hydroclimatic research.

Although the framework has been specifically developed for temperature-precipitation copulas, it is anticipated that it can be easily applied to a wide range of hydroclimatic research studies. By breaking down the process into six generic steps and providing short and simple narratives, the framework can to a large extent easily be adopted for other pairs of (hydroclimatic)



variables to construct bivariate copulas, thereby supporting researchers and decision makers in addressing climatological
440 hazards and sustainable development.

Code and data availability. The codes are developed in MATLAB and are publicly available at <https://doi.org/10.5281/zenodo.3900001>. The observational data for daily precipitation and temperature in the climate normal periods 1961-1990 were downloaded from a freely available database (www.vattenwebb.smhi.se).

445 *Author contribution.* FT and CT designed the framework. FT developed the code. FT, CT and TG designed the visualization of the article. JOH and MS contributed to the statistical requirements of the framework. JOH, TB and OR contributed to the hydroclimatic implications of the results. FT developed the first manuscript. All co-authors contributed to interpret the final results. All co-authors contributed to edit the manuscript.

Competing interests. The authors declare that they have no conflict of interests.

450 *Acknowledgements.* This research was supported by the Swedish Research Council (VR, grant no. 2017-04970).



References

- Aas, K.: Modelling the dependence structure of financial assets: A survey of four copulas Gaussian copula Clayton copula Student's t-copula NR Norwegian Computing Center Title: Modelling the dependence structure of financial assets: A survey of four copulas. [online] Available from: <http://www.nr.no> (Accessed 2 December 2019), 2004.
- Abdulah Al-Sudani, Z., Salih, S. Q., sharafati, A. and Yaseen, Z. M.: Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation, *J. Hydrol.*, 573, 1–12, doi:10.1016/j.jhydrol.2019.03.004, 2019.
- AghaKouchak, A., Bárdossy, A. and Habib, E.: Copula-based uncertainty modelling: Application to multisensor precipitation estimates, *Hydrol. Process.*, 24(15), 2111–2124, doi:10.1002/hyp.7632, 2010.
- AghaKouchak, A., Cheng, L., Mazdiyasi, O. and Farahmand, A.: Global warming and changes in risk of concurrent climate extremes: Insights from the 2014 California drought, *Geophys. Res. Lett.*, 41(24), 8847–8852, doi:10.1002/2014GL062308, 2014.
- Alexandrov, V. A. and Hoogenboom, G.: The impact of climate variability and change on crop yield in Bulgaria, *Agric. For. Meteorol.*, 104(4), 315–327, doi:10.1016/S0168-1923(00)00166-0, 2000.
- Alidoost, F., Su, Z. and Stein, A.: Evaluating the effects of climate extremes on crop yield, production and price using multivariate distributions: A new copula application, *Weather Clim. Extrem.*, 26, 100227, doi:10.1016/j.wace.2019.100227, 2019.
- Allen, D., McAleer, M. and Singh, A.: Risk Measurement and Risk Modelling Using Applications of Vine Copulas, *Sustainability*, 9(10), 1762, doi:10.3390/su9101762, 2017.
- Aloui, R., Ben Aïssa, M. S. and Nguyen, D. K.: A Copula Approach to Dependence Structure in Petroleum Markets, in *Rethinking Valuation and Pricing Models*, pp. 317–329, Elsevier Inc., 2013.
- Ashraf Vaghefi, S., Mousavi, S. J., Abbaspour, K. C., Srinivasan, R. and Yang, H.: Analyses of the impact of climate change on water resources components, drought and wheat yield in semiarid regions: Karkheh River Basin in Iran, *Hydrol. Process.*, 28(4), 2018–2032, doi:<https://doi.org/10.1002/hyp.9747>, 2014.
- Benestad, R. E. and Haugen, J. E.: On complex extremes: flood hazards and combined high spring-time precipitation and temperature in Norway, *Clim. Change*, 85(3–4), 381–406, doi:10/bwsv8d, 2007.
- Berg, D.: Copula goodness-of-fit testing: An overview and power comparison, *Eur. J. Financ.*, 15(7–8), 675–701, doi:10.1080/13518470802697428, 2009.
- Boeck, H. J. De, Dreesen, F. E., Janssens, I. A. and Nijs, I.: Whole-system responses of experimental plant communities to climate extremes imposed in different seasons, *New Phytol.*, 189(3), 806–817, doi:10/fccgnq, 2011.
- Burnham, J. F.: Scopus database: a review, *Biomed. Digit. Libr.*, 3(1), doi:10/c4dc8g, 2006.
- Charpentier, A., Fermanian, J.-D. and Scaillet, O.: The Estimation of Copulas: Theory and Practice, *Copulas From Theory to Appl. Financ.*, 35–62, 2007.



- 485 Chen, J., Li, C., Brissette, F. P., Chen, H., Wang, M. and Essou, G. R. C.: Impacts of correcting the inter-variable correlation of climate model outputs on hydrological modeling, *J. Hydrol.*, 560, 326–341, doi:10.1016/j.jhydrol.2018.03.040, 2018.
- Cherubini, U., Luciano, E. and Vecchiato, W.: *Copula Methods in Finance*, Wiley Blackwell., 2013.
- Chou, C. and Neelin, J. D.: Mechanisms of Global Warming Impacts on Regional Tropical Precipitation*, *J. Clim.*, 17(13), 2688–2701 [online] Available from: <http://journals.ametsoc.org/doi/abs/10.1175/1520-0442%282004%29017%3C2688%3AMOGWIO%3E2.0.CO%3B2> (Accessed 27 November 2019), 2004.
- 490 Colacino, M. and Dell’osso, L.: The local atmospheric circulation in the Rome area: Surface observations, *Boundary-Layer Meteorol.*, 14(2), 133–151, doi:10.1007/BF00122615, 1978.
- Cong, R. G. and Brady, M.: The interdependence between rainfall and temperature: Copula analyses, *Sci. World J.*, 2012, doi:10.1100/2012/405675, 2012.
- 495 Cooley, D.: *Return Periods and Return Levels Under Climate Change*, pp. 97–114, Springer, Dordrecht., 2013.
- Dai, A. and Bloecker, C. E.: Impacts of internal variability on temperature and precipitation trends in large ensemble simulations by two climate models, *Clim. Dyn.*, 52, 289–306, doi:10.1007/s00382-018-4132-4, 2019.
- Dee, D. P., Alonso Balmaseda, M., Balsamo, G., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., M Beljaars, A. C., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J., Park, B., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J., Vitart, F., Acm, B., de Berg, van L., J-j, M., B-k, P. and Rosnay, de P.: 828 ftp The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.* *Q. J. R. Meteorol. Soc.*, 137, 553–597, doi:10.1002/qj.828, 2011.
- 500 505 Didier, D., Baudry, J., Bernatchez, P., Dumont, D., Sadegh, M., Bismuth, E., Bandet, M., Dugas, S. and Sévigny, C.: Multihazard simulation for coastal flood mapping: Bathhtub versus numerical modelling in an open estuary, Eastern Canada, *J. Flood Risk Manag.*, 12(S1), doi:10.1111/jfr3.12505, 2019.
- Dupuis, D. J.: Using Copulas in Hydrology: Benefits, Cautions, and Issues, *J. Hydrol. Eng.*, 12(4), 381–393, doi:10.1061/(ASCE)1084-0699(2007)12:4(381), 2007.
- 510 Ebita, A., Kobayashi, S., Ota, Y., Moriya, M., Kumabe, R., Onogi, K., Harada, Y., Yasui, S., Miyaoka, K., Takahashi, K., Kamahori, H., Kobayashi, C., Endo, H., Soma, M., Oikawa, Y. and Ishimizu, T.: The {Japanese} 55-year {Reanalysis} “{JRA}-55”: {An} {Interim} {Report}, *SOLA*, 7, 149–152, doi:10/dwmmvj, 2011.
- Embrechts, P., Mcneil, A. and Strauman, D.: *Correlation and Dependency in Risk Management*, New York., 1999.
- Ferrari, E., Coscarelli, R. and Sirangelo, B.: Correlation analysis of seasonal temperature and precipitation in a region of southern Italy, *Geosci.*, 8(5), 1–10, doi:10.3390/geosciences8050160, 2018.
- 515 Genest, C. and Favre, A.-C.: Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask, *J. Hydrol. Eng.*, 12(4), 347–368, doi:10.1061/(asce)1084-0699(2007)12:4(347), 2007.
- Genest, C. and Rémillard, B.: Tests of independence and randomness based on the empirical copula process, *Test*, 13(2),



335–370, doi:10.1007/bf02595777, 2004.

- 520 Genest, C. and Rivest, L.-P.: Statistical Inference Procedures for Bivariate Archimedean Copulas, *J. Am. Stat. Assoc.*, 88(423), 1034, doi:10.2307/2290796, 1993.
- Genest, C., Rémillard, B. and Beaudoin, D.: Goodness-of-fit tests for copulas: A review and a power study, *Insur. Math. Econ.*, 44(2), 199–213, doi:10.1016/j.insmatheco.2007.10.005, 2009.
- Gennaretti, F., Sangelantoni, L. and Grenier, P.: Toward daily climate scenarios for Canadian Arctic coastal zones with
525 more realistic temperature-precipitation interdependence, *J. Geophys. Res. Atmos.*, 120(23), 11,862–11,877, doi:10.1002/2015JD023890, 2015.
- Ghoddusi, H.: Price risks for biofuel producers in a deregulated market, *Renew. Energy*, 114(PB), 394–407, doi:10.1016/j.renene.2017.07.044, 2017.
- Ghoudi, K. and Rémillard, B.: Empirical processes based on pseudo-observations II: the multivariate case., 2004.
- 530 Hao, Z., Singh, V. P. and Hao, F.: Compound extremes in hydroclimatology: A review, *Water (Switzerland)*, 10(6), 16–21, doi:10.3390/w10060718, 2018.
- Hundecha, Y. and Bárdossy, A.: Trends in daily precipitation and temperature extremes across western Germany in the second half of the 20th century, *Int. J. Climatol.*, 25(9), 1189–1202, doi:10.1002/joc.1182, 2005.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J.,
535 Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R. and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, 77(3), 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.
- Laux, P., Vogl, S., Qiu, W., Knoche, H. R. and Kunstmann, H.: Copula-based statistical refinement of precipitation in RCM simulations over complex terrain, *Hydrol. Earth Syst. Sci.*, 15(7), 2401–2419, doi:10.5194/hess-15-2401-2011, 2011.
- 540 Lazoglou, G. and Anagnostopoulou, C.: Joint distribution of temperature and precipitation in the Mediterranean, using the Copula method, *Theor. Appl. Climatol.*, 135(3–4), 1399–1411, doi:10.1007/s00704-018-2447-z, 2019.
- Li, C., Sinha, E., Horton, D. E., Diffenbaugh, N. S. and Michalak, A. M.: Joint bias correction of temperature and precipitation in climate model simulations, *J. Geophys. Res. Atmos.*, 119(23), 13,153–13,162, doi:10.1002/2014JD022514, 2014.
- 545 Li, X. and Babovic, V.: Multi-site multivariate downscaling of global climate model outputs: an integrated framework combining quantile mapping, stochastic weather generator and Empirical Copula approaches, *Clim. Dyn.*, 52(9–10), 5775–5799, doi:10.1007/s00382-018-4480-0, 2019.
- Lins, H. F. and Cohn, T. A.: Stationarity: Wanted Dead or Alive?1, *JAWRA J. Am. Water Resour. Assoc.*, 47(3), 475–480, doi:10.1111/j.1752-1688.2011.00542.x, 2011.
- 550 Madadgar, S. and Moradkhani, H.: Spatio-temporal drought forecasting within Bayesian networks, *J. Hydrol.*, 512, 134–146, doi:10.1016/j.jhydrol.2014.02.039, 2014.
- Maftei, C., Barbulescu, A., Buta, C. and Serban, C.: Change points detection and variability analysis of some precipitation



- series, in Recent Researches in Computational Techniques, Non-Linear Systems and Control - Proc. of the 13th WSEAS Int. Conf. on MAMECTIS'11, NOLASC'11, CONTROL'11, WAMUS'11, pp. 232–237., 2011.
- 555 Mangini, W., Viglione, A., Hall, J., Hundecha, Y., Ceola, S., Montanari, A., Rogger, M., Salinas, J. L., Borzi, I. and Parajka, J.: Detection of trends in magnitude and frequency of flood peaks across Europe, *Hydrol. Sci. J.*, 63(4), 493–512, doi:10.1080/02626667.2018.1444766, 2018.
- Mao, G., Vogl, S., Laux, P., Wagner, S. and Kunstmann, H.: Stochastic bias correction of dynamically downscaled precipitation fields for Germany through Copula-based integration of gridded observation data, *Hydrol. Earth Syst. Sci.*, 19(4), 1787–1806, doi:10.5194/hess-19-1787-2015, 2015.
- 560 Mazdiyasi, O., AghaKouchak, A., Davis, S. J., Madadgar, S., Mehran, A., Ragno, E., Sadegh, M., Sengupta, A., Ghosh, S., Dhanya, C. T. and Niknejad, M.: Increasing probability of mortality during Indian heat waves, *Sci. Adv.*, 3(6), e1700066, doi:10.1126/sciadv.1700066, 2017.
- McCarl, B. A., Villavicencio, X. and Wu, X.: Climate Change and Future Analysis: Is Stationarity Dying?, *Am. J. Agric. Econ.*, 90(5), 1241–1247, doi:10.1111/j.1467-8276.2008.01211.x, 2008.
- 565 Mesbahzadeh, T., Miglietta, M. M., Mirakbari, M., Soleimani Sardoo, F. and Abdolhoseini, M.: Joint Modeling of Precipitation and Temperature Using Copula Theory for Current and Future Prediction under Climate Change Scenarios in Arid Lands (Case Study, Kerman Province, Iran), *Adv. Meteorol.*, 2019, doi:10.1155/2019/6848049, 2019.
- Phiri, D., Morgenroth, J. and Xu, C.: Long-term land cover change in Zambia: An assessment of driving factors, *Sci. Total Environ.*, 697, 134206, doi:10.1016/j.scitotenv.2019.134206, 2019.
- 570 Piani, C. and Haerter, J. O.: Two dimensional bias correction of temperature and precipitation copulas in climate models, *Geophys. Res. Lett.*, 39(20), 1–6, doi:10.1029/2012GL053839, 2012.
- Rakonczai, P., Márkus, L. and Zempléni, A.: Autocopulas: Investigating the Interdependence Structure of Stationary Time Series, *Methodol. Comput. Appl. Probab.*, 14(1), 149–167, doi:10.1007/s11009-011-9230-2, 2012.
- 575 Rana, A., Moradkhani, H. and Qin, Y.: Understanding the joint behavior of temperature and precipitation for climate change impact studies, *Theor. Appl. Climatol.*, 129(1–2), 321–339, doi:10.1007/s00704-016-1774-1, 2017.
- Räty, O., Räisänen, J., Bosshard, T. and Donnelly, C.: Intercomparison of Univariate and Joint Bias Correction Methods in Changing Climate From a Hydrological Perspective, *Climate*, 6(2), 33, doi:10.3390/cli6020033, 2018.
- Ribeiro, A. F. S., Russo, A., Gouveia, C. M. and Páscoa, P.: Copula-based agricultural drought risk of rainfed cropping systems, *Agric. Water Manag.*, 223, doi:10.1016/j.agwat.2019.105689, 2019.
- 580 Sadegh, M., Ragno, E. and AghaKouchak, A.: Multivariate Copula Analysis Toolbox (MvCAT): Describing dependence and underlying uncertainty using a Bayesian framework, *Water Resour. Res.*, 53(6), 5166–5183, doi:10.1002/2016WR020242, 2017.
- Sadegh, M., Moftakhari, H., Gupta, H. V., Ragno, E., Mazdiyasi, O., Sanders, B., Matthew, R. and AghaKouchak, A.: 585 Multihazard Scenarios for Analysis of Compound Extreme Events, *Geophys. Res. Lett.*, 45(11), 5470–5480, doi:10.1029/2018GL077317, 2018.



- Salvadori, G. and De Michele, C.: Multivariate multiparameter extreme value models and return periods: A copula approach, *Water Resour. Res.*, 46(10), doi:10.1029/2009WR009040, 2010.
- Salvadori, G., Michele, C. De, Kottegoda, N. T. and Rosso, R.: *Extremes in Nature: An Approach to Using Copulas*, Springer., 2007.
- Salvadori, G., Tomasicchio, G. R. and D'Alessandro, F.: Practical guidelines for multivariate analysis and design in coastal and off-shore engineering, *Coast. Eng.*, 88, 1–14, doi:10.1016/j.coastaleng.2014.01.011, 2014.
- Satgé, F., Hussain, Y., Xavier, A., Zolá, R. P., Salles, L., Timouk, F., Seyler, F., Garnier, J., Frappart, F. and Bonnet, M.-P.: Unraveling the impacts of droughts and agricultural intensification on the Altiplano water resources, *Agric. For. Meteorol.*, 595 279, 107710, doi:10.1016/j.agrformet.2019.107710, 2019.
- Schweizer, B.: *Thirty Years of Copulas*, in *Advances in Probability Distributions with Given Marginals*, pp. 13–50, Springer Netherlands., 1991.
- Serinaldi, F.: Can we tell more than we can know? The limits of bivariate drought analyses in the United States, *Stoch. Environ. Res. Risk Assess.*, 30(6), 1691–1704, doi:10.1007/s00477-015-1124-3, 2016.
- Shojaeezadeh, S. A., Nikoo, M. R., McNamara, J. P., AghaKouchak, A. and Sadegh, M.: Stochastic modeling of suspended sediment load in alluvial rivers, *Adv. Water Resour.*, 119(June), 188–196, doi:10.1016/j.advwatres.2018.06.006, 2018.
- Shojaeezadeh, S. A., Nikoo, M. R., Mirchi, A., Mallakpour, I., AghaKouchak, A. and Sadegh, M.: Probabilistic hazard assessment of contaminated sediment in rivers, *Sci. Total Environ.*, 703, 134875, doi:10.1016/j.scitotenv.2019.134875, 2020.
- Shukla, S. and Gedam, S.: Evaluating Hydrological Responses to Urbanization in a Tropical River Basin: A Water Resources Management Perspective, *Nat. Resour. Res.*, 28(2), 327–347, doi:10.1007/s11053-018-9390-7, 2019.
- Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, *J. Hydrol.*, 456–457, 12–29, doi:10.1016/j.jhydrol.2012.05.052, 2012.
- Trenberth, K. E.: Climate change caused by human activities is happening and it already has major consequences, *J. Energy Nat. Resour. Law*, 36(4), 463–481, doi:10/ggc8pb, 2018.
- Trenberth, K. E., Dai, A., Rasmussen, R. M. and Parsons, D. B.: The changing character of precipitation, *Bull. Am. Meteorol. Soc.*, 84(9), 1205–1217, doi:10.1175/BAMS-84-9-1205, 2003.
- Trenberth, K. E., Dai, A., Van Der Schrier, G., Jones, P. D., Barichivich, J., Briffa, K. R. and Sheffield, J.: Global warming and changes in drought, *Nat. Clim. Chang.*, 4(1), 17–22, doi:10.1038/nclimate2067, 2014.
- Vaze, J., Post, D. A., Chiew, F. H. S., Perraud, J.-M., Viney, N. R. and Teng, J.: Climate non-stationarity – Validity of calibrated rainfall–runoff models for use in climate change studies, *J. Hydrol.*, 394(3–4), 447–457, doi:10.1016/J.JHYDROL.2010.09.018, 2010.
- Vogel, R. M., Lall, U., Cai, X., Rajagopalan, B., Weiskel, P. K., Hooper, R. P. and Matalas, N. C.: Hydrology: The interdisciplinary science of water, *Water Resour. Res.*, 51(6), 4409–4430, doi:10.1002/2015WR017049, 2015.
- Vogl, S., Laux, P., Qiu, W., Mao, G. and Kunstmann, H.: Copula-based assimilation of radar and gauge information to



derive bias-corrected precipitation fields, *Hydrol. Earth Syst. Sci.*, 16(7), 2311–2328, doi:10.5194/hess-16-2311-2012, 2012.

Vrac, M. and Friederichs, P.: Multivariate—Intervariable, Spatial, and Temporal—Bias Correction*, *J. Clim.*, 28(1), 218–237, doi:10.1175/JCLI-D-14-00059.1, 2015.

625 Wang, J., Chen, G., Kang, W., Hu, K. and Wang, L.: Impoundment intensity determines temporal patterns of hydrological fluctuation, carbon cycling and algal succession in a dammed lake of Southwest China, *Water Res.*, 148, 162–175, doi:10.1016/j.watres.2018.10.032, 2019a.

Wang, L., Yu, H., Yang, M., Yang, R., Gao, R. and Wang, Y.: A drought index: The standardized precipitation evapotranspiration runoff index, *J. Hydrol.*, 571, 651–668, doi:10.1016/j.jhydrol.2019.02.023, 2019b.

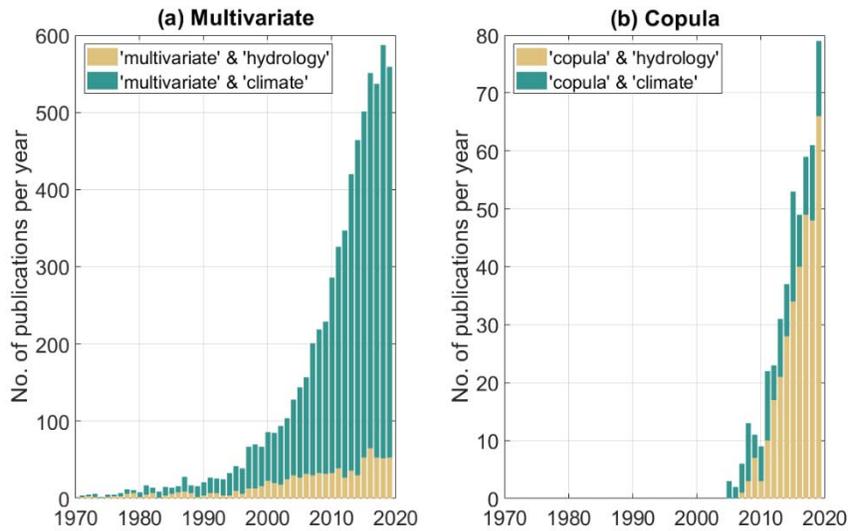
630 Weiß, G.: Copula parameter estimation by maximum-likelihood and minimum-distance estimators: A simulation study, *Comput. Stat.*, 26(1), 31–54, doi:10.1007/s00180-010-0203-7, 2011.

Wendland, W. M.: *Hydroclimatology*, in *Climatology*, pp. 497–502, Kluwer Academic Publishers., 1987.

Zscheischler, J. ; Orth, R. ; and Seneviratne, S. I.: Bivariate return periods of temperature and precipitation explain a large fraction of European crop yields, , doi:10.3929/ethz-b-000190400, 2017.



635 **Figures**



640 **Figure 1.** Number of annually published peer-reviewed papers listed by Scopus (a) when combining the search term ‘multivariate’ with either ‘hydrology’ (beige bars) or ‘climate’ (turquoise bars), and (b) when combining the search term ‘copula’ with either ‘hydrology’ (beige bars) or ‘climate’ (turquoise bars).

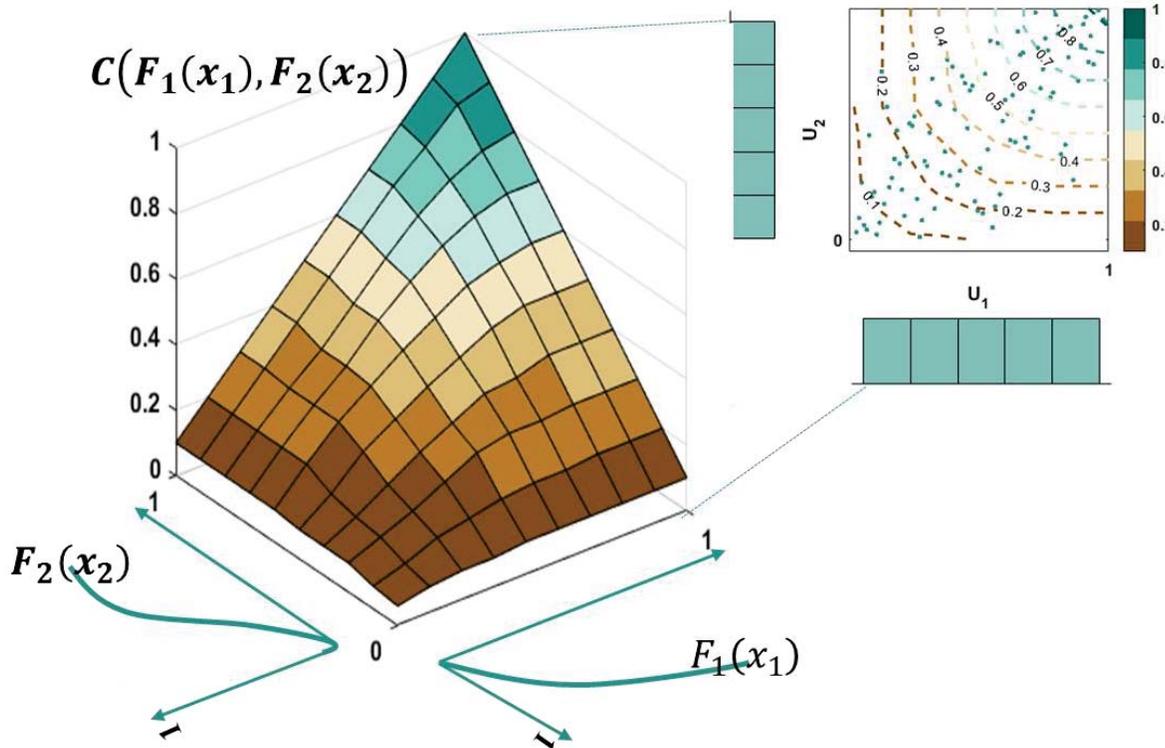
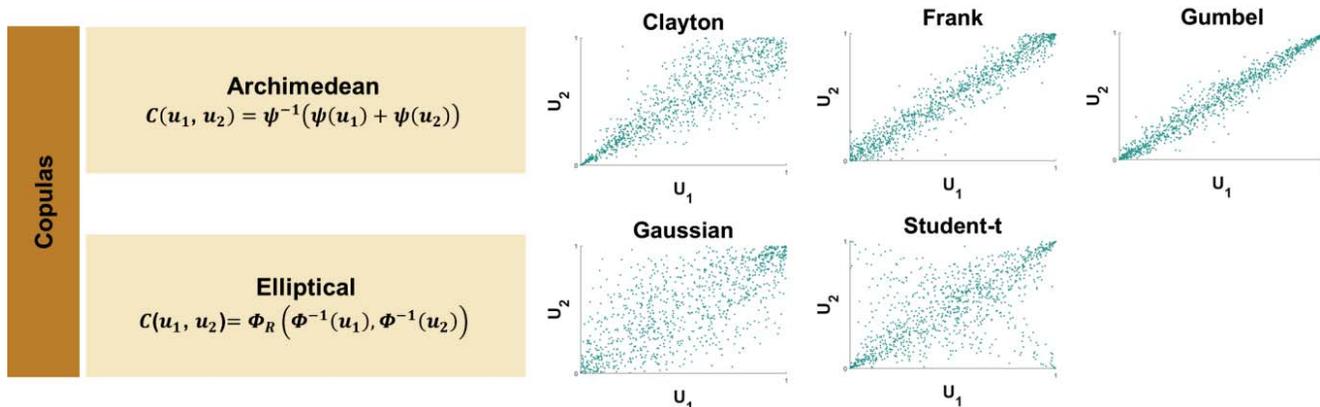
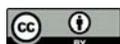
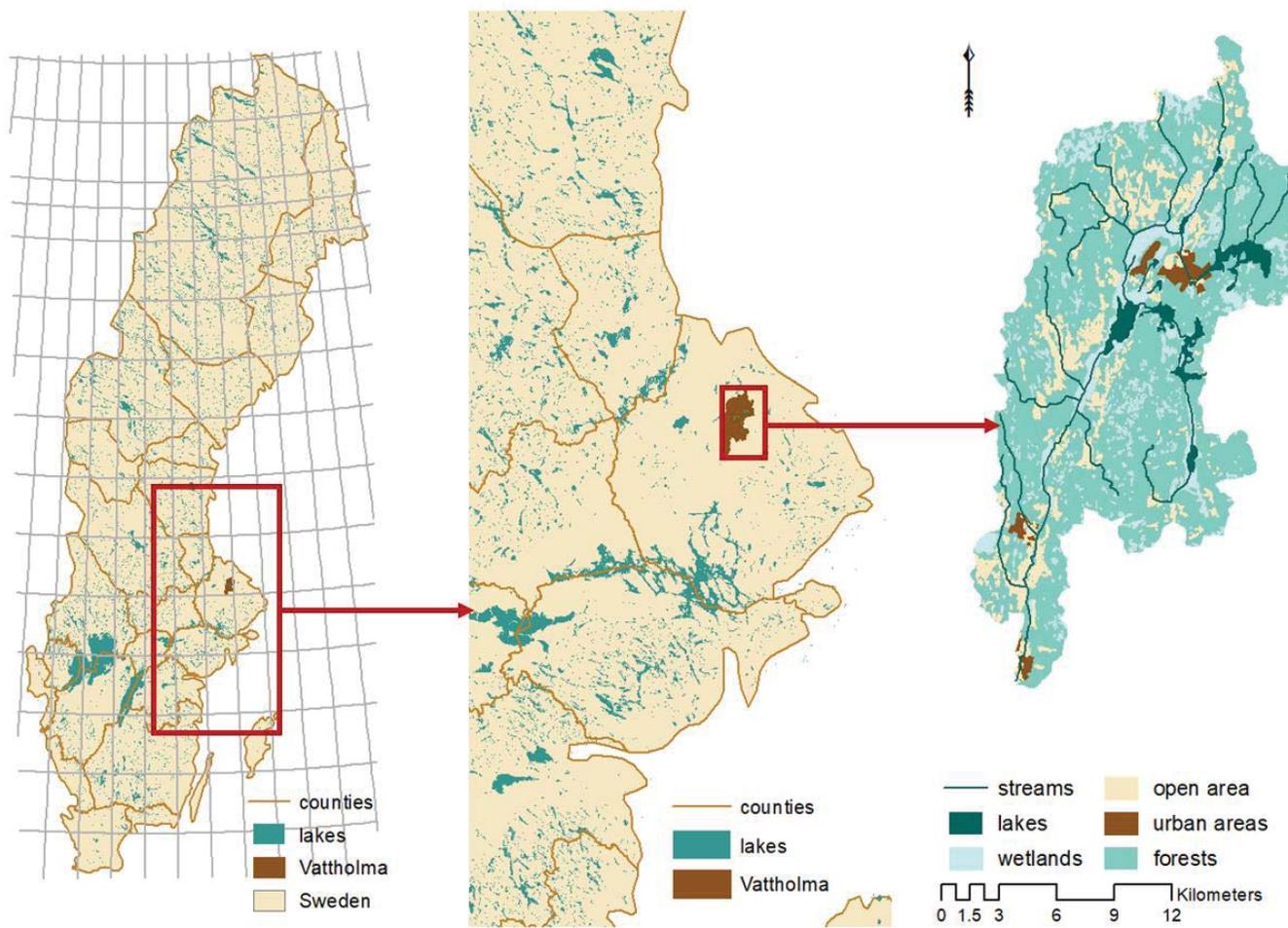


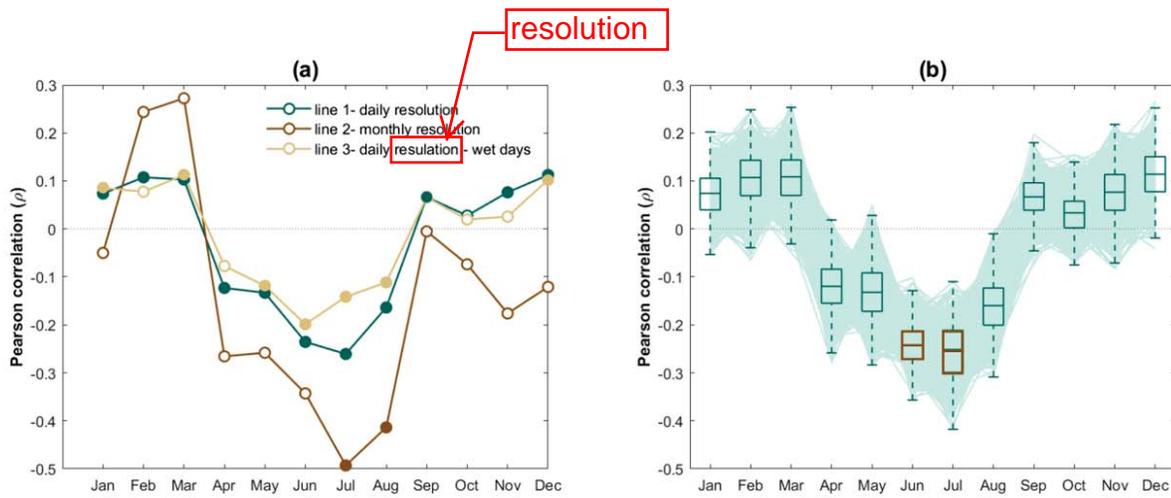
Figure 2. Schematic visualization of the empirical two-dimensional copula (surface) and its relation to the marginal cumulative distribution functions.



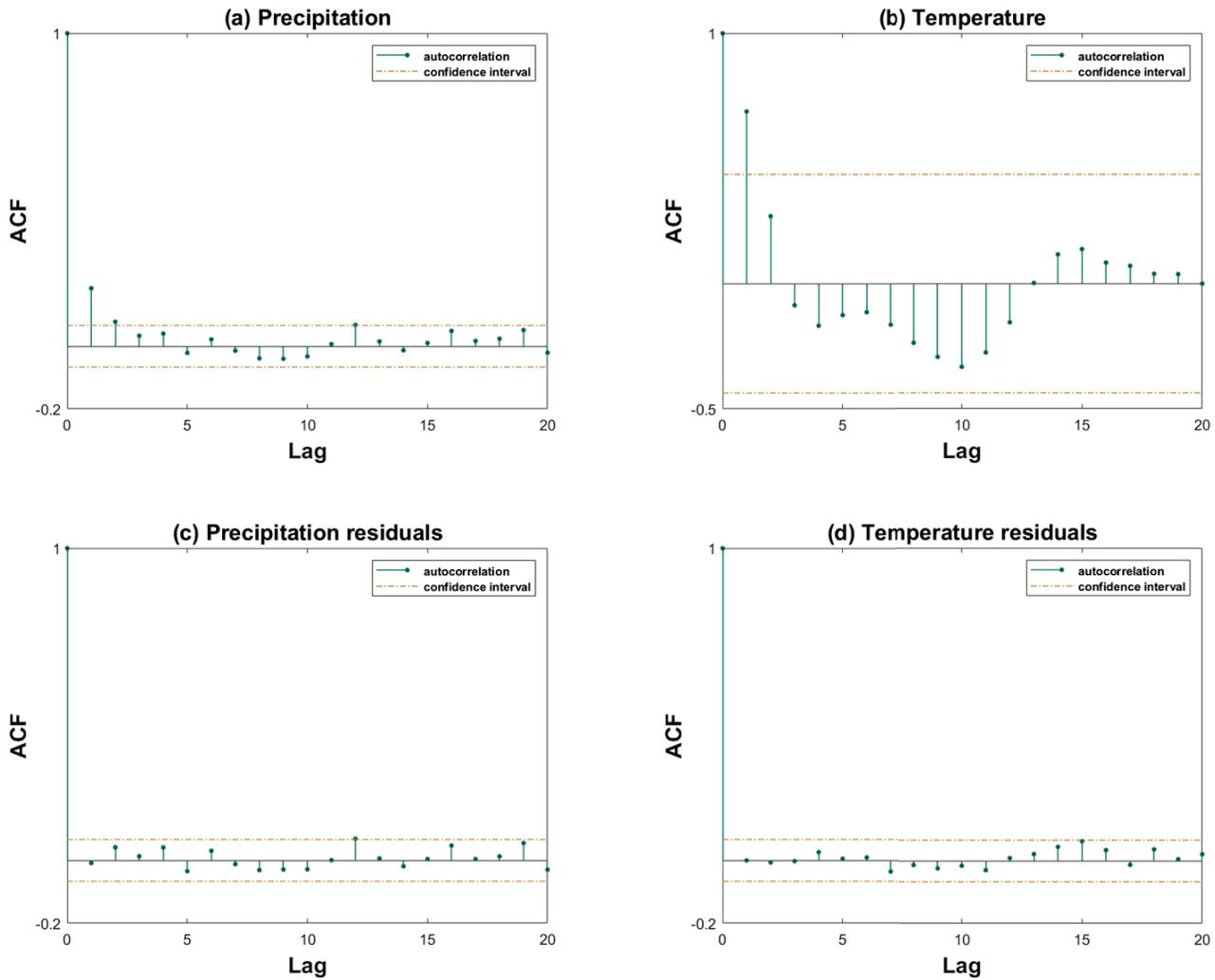
645 Figure 3. Illustration of two widely used copula families (Archimedean and Elliptical) and their common sub-types. 1000 normally-distributed points with Pearson correlation of $\rho=0.75$ were randomly generated and transformed into uniform margins U_1 and U_2 and plotted in two-dimensional space.



650 Figure 4. Overview of the Vattholma catchment, with its outlet being located at 60.0194°N and 17.7307°E.



655 **Figure 5. Linear correlation between temperature and precipitation: (a) Comparison of different temporal resolutions. Filled markers indicate significance at 0.05% significance level. (b) Daily resolution correlation based on 1000 samples of 10 randomly drawn years from a 30 years period. Boxes in brown represent months with the lowest average p-value (here below 0.001).**

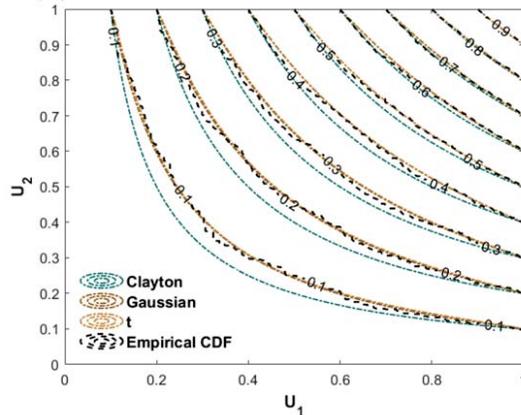


660 Figure 6. The autocorrelation function (ACF) for daily precipitation (a) and temperature (b) in July compared to the precipitation residuals (c) and temperature residuals (d) after removing autocorrelation from the data.

what is the spread of the
confidence intervals - 95%?
in (d) it looks like 100%



(a) Empirical and theoretical copulas



(b) Empirical and theoretical

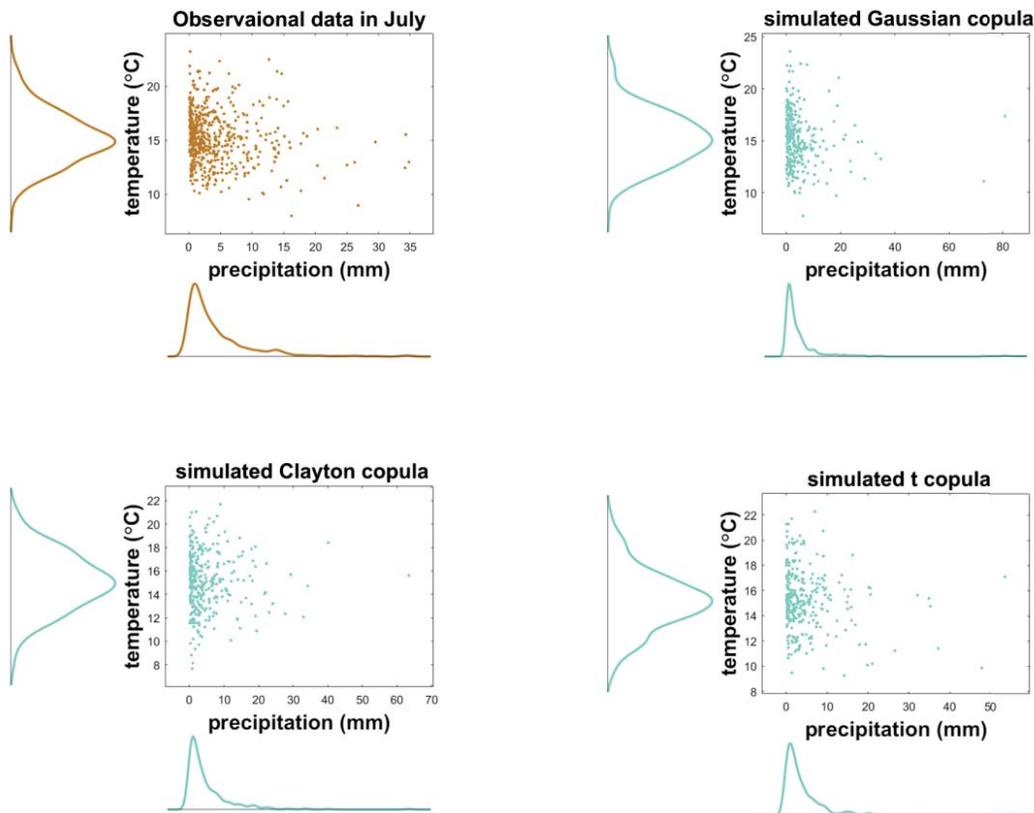


Figure 7. Copulas for daily precipitation and temperature in July in the Vattholma catchment. (a) Comparison of three fitted copulas with the joint empirical cumulative distribution function (CDF) of transformed values. The numeric labels on the curves correspond to the cumulative joint probability of U_1 and U_2 . (b) Joint copula distribution of the original data and different copula families. Each panel shows a scatterplot of daily precipitation (mm) and daily average near-surface temperature ($^{\circ}\text{C}$) and the associated marginal probability density function.

665

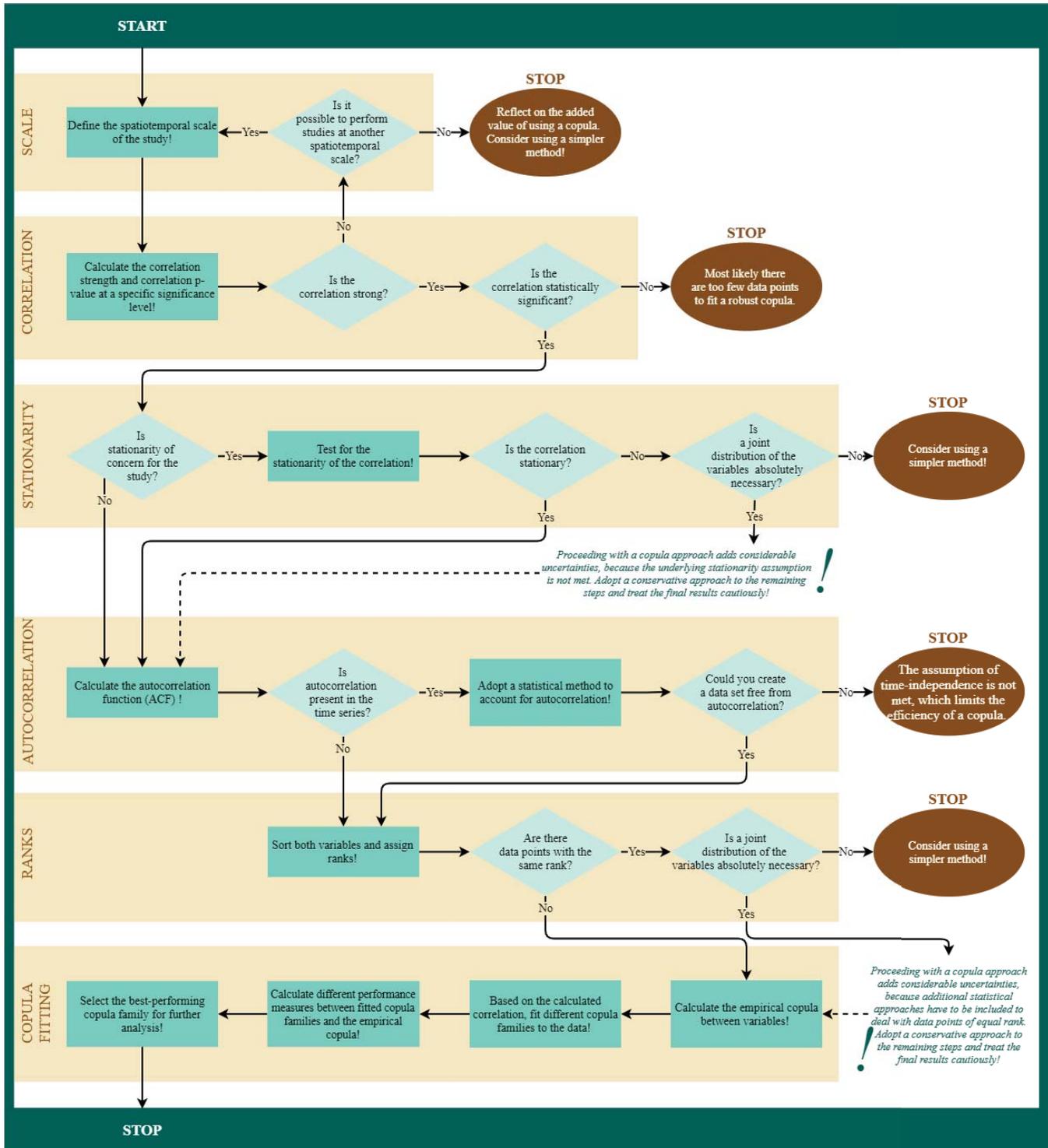


Figure 8. Decision support framework for adopting the copula approach. Each step (beige boxes) refers to a subheading of section 3, which brings up the statistical requirements and general assumptions that should be fulfilled.

that's a very helpful flow-chart - especially the !!



670

Tables

Table 1. Archimedean copula parameter range and relationship to Kendall's τ .

	θ	τ
Gumbel	$[1, +\infty)$	$= \frac{\theta-1}{\theta}$
Clayton	$[0, +\infty)$	$= \frac{\theta}{\theta+2}$
Frank	$(-\infty, +\infty) \setminus 0$	$= 1 - \frac{4}{\theta} \left[\frac{1}{\theta} \int_0^\theta \frac{t}{e^t-1} dt - 1 \right]$

675 **Table 2. Different copula families and their performance measures as well as their ranking (numbers in parenthesis) in being able to represent the empirical copula. Families with the highest rank are shown in bold.**

Performance measure	copula family		
	Gaussian	Student-t	Clayton
E_{RMS} (ranking)	0.3252 (2)	0.318 (1)	0.4062 (3)
E_{NS} (ranking)	0.9956 (2)	0.996 (1)	0.9931 (3)
S_n (ranking)	8.2100 (1)	9.5648 (2)	62.171 (3)
D_n (ranking)	0.0199 (2)	0.0197 (1)	0.0341 (3)

please define the symbols in the caption to help the reader:
 E_{RMS} , E_{NS} , S_n and D_n :