Return period of high-dimensional compound events. Part I: Conceptual framework - Supplemantary information

Manuel Del Jesus[1], Diego Urrea Méndez[1], and Dina Vanessa Gomez Rave[1]

[1]IHCantabria - Instituto de Hidráulica Ambiental de la Universidad de Cantabria, Santander, Spain.

Correspondence: Manuel Del Jesus (manuel.deljesus@unican.es)

S1 Multivariate approaches

Various families of copula functions can be found in the literature. Each family is better suited to capture different types of dependency among variables and offers different dependence structures, symmetries and tail behaviors, thus collectively offering a great flexibility for modeling compound events. The selection of the copula family and type for modeling tail dependence is very important when modeling compound extreme events (Zscheischler et al., 2020). The copula families typically used in climate and hydrological studies are Meta-Elliptical, Archimedean, and Extreme-Value (Chen and Guo, 2019). In this section, we will describe the applicability of each copula family mentioned. Additionally, we will discuss which families are most effective for modeling dependencies in two or more dimensions, with particular attention to tail dependencies.

S1.1 Multivariate Meta-Elliptical Copula

The most commonly used Meta-Elliptical copulas in practice are the *Gaussian* and the *t-student* copulas. The *Multivariate Gaussian copula* has frequently been applied to model dependence in higher dimensions (greater than two) (Genest et al., 2007; Hao et al., 2017). However, it is less suitable for modeling hydroclimatic events that do not have radial symmetry (Jaser and Min, 2021). These copulas have certain limitations when it comes to modeling compound extreme events. One of the main reasons is that Gaussian copulas do not adequately capture tail dependence in the data. In fact, this means that they cannot accurately capture events in the tail of the copula. According to Genton (2004), the *Multivariate Gaussian copula* can be expressed as shown in Eq. (S1) as:

$$C(\boldsymbol{u};R) = \Phi_R\left(\Phi^{-1}\left(u_1\right),\ldots,\Phi^{-1}\left(u_d\right)\right) \tag{S1}$$

where Φ denotes the univariate standard normal cumulative distribution function, Φ^{-1} is the inverse of the function, and $\Phi_R\left(\Phi^{-1}\left(u_1\right),\ldots,\Phi^{-1}\left(u_d\right)\right)$ the multivariate standard normal distribution function with symmetric positive definite correlation matrix R, zero mean vector, and unit variances. The copula density is defined in Eq. (S2) as:

$$c(\boldsymbol{u};R) = |R|^{-\frac{1}{2}}\exp\left\{\frac{1}{2}\boldsymbol{x}^{op}\left(I_d - R^{-1}\right)\boldsymbol{x}\right\} \tag{S2}$$

where $\boldsymbol{x} = (x_1, \ldots, x_d)^{op} \in \mathbb{R}^d$, with $x_i := \Phi^{-1}(u_i), i = 1, \ldots, d$. More details and applications can be found in (Genton, 2004).

The *Multivariate Student-t copula* provides greater modeling flexibility in terms of tail dependence (Carreau and Bouvier, 2016), since, compared to the Multivariate Gaussian copula, it has an additional degree of freedom. Additionally, it presents heavy tails, meaning it can more accurately model compound extreme events and outliers. This type of copula can capture both linear and non-linear dependence between random variables, making it especially useful in situations where there is strong tail dependence in the marginal distributions. The *Multivariate Student-t copula* can be used to model dependence in higher dimensions (greater than two). According to Kotz and Nadarajah (2004), the *Multivariate Student's t copula* can be expressed as shown in Eq. (S3):

$$\begin{aligned} c(u_1, \ldots, u_d; R, \nu) &= \frac{\partial^d}{\partial u_1 \ldots \partial u_d} C(u_1, \ldots, u_d; R, \nu) \\ &= \frac{\partial^d}{\partial u_1 \ldots \partial u_d} \int_{-\infty}^{T_\nu^{-1}(u_1)} \cdots \int_{-\infty}^{T_v^{-1}(u_d)} \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{(\pi\nu)^{\frac{d}{2}} |R|^{\frac{1}{2}}} \left(1 + \frac{\mathbf{w}^T R^{-1} \mathbf{w}}{\nu}\right)^{-\frac{\nu+d}{2}} d\mathbf{w} \end{aligned} \tag{S3}$$

where R is the correlation matrix of X, R^{-1} is the inverse, d represents the dimension of variables, $\mathbf{w}$ represents the integral matrix $\mathbf{w} = [\mathbf{w_1}, ..., \mathbf{w_d}]^T$. More details and applications can be found in Kotz and Nadarajah (2004).

S1.2 Multivariate extreme-value copulas

The most well-known types of Extreme-Value copulas are Gumbel-Hougaard, Hüsler-Reiss, Galambos, Tawn, and t-EV. The Extreme-Value copulas arise from extensions of univariate extreme-value theory to higher dimensions. Extreme-value copulas are appropriate for modeling the dependence between extreme events and can also be a convenient choice for modeling general positive dependence structures (Gudendorf and Segers, 2010). For the d-dimensional case (greater than two), the dependence structure of a multivariate extreme-value distribution is realized in terms of a spectral measure. Details can be found in Ribeiro et al. (2020).

S1.3 Multivariate Archimedean copulas

Archimedean copulas are notable for their greater flexibility in capturing tail dependence (both lower and upper) (Nelsen, 2006), making them particularly suitable for modeling compound extreme events such as droughts and floods. The most common types of Archimedean copulas in hydroclimatic studies are Gumbel, Clayton, Frank, and Joe (Genest and Favre, 2007); however, other types with two parameters are also available , such as the BB1 and BB7 copulas. To model dependence in higher dimensions (greater than two), Archimedean copulas are divided into symmetric and asymmetric. The symmetric ones, denoted as exchangeable Archimedean copulas (EAC), are easy to construct but consider only a single parameter, meaning hydroclimatic variables share the same dependence structure, an assumption that is rarely valid (Zhang and Singh, 2019).

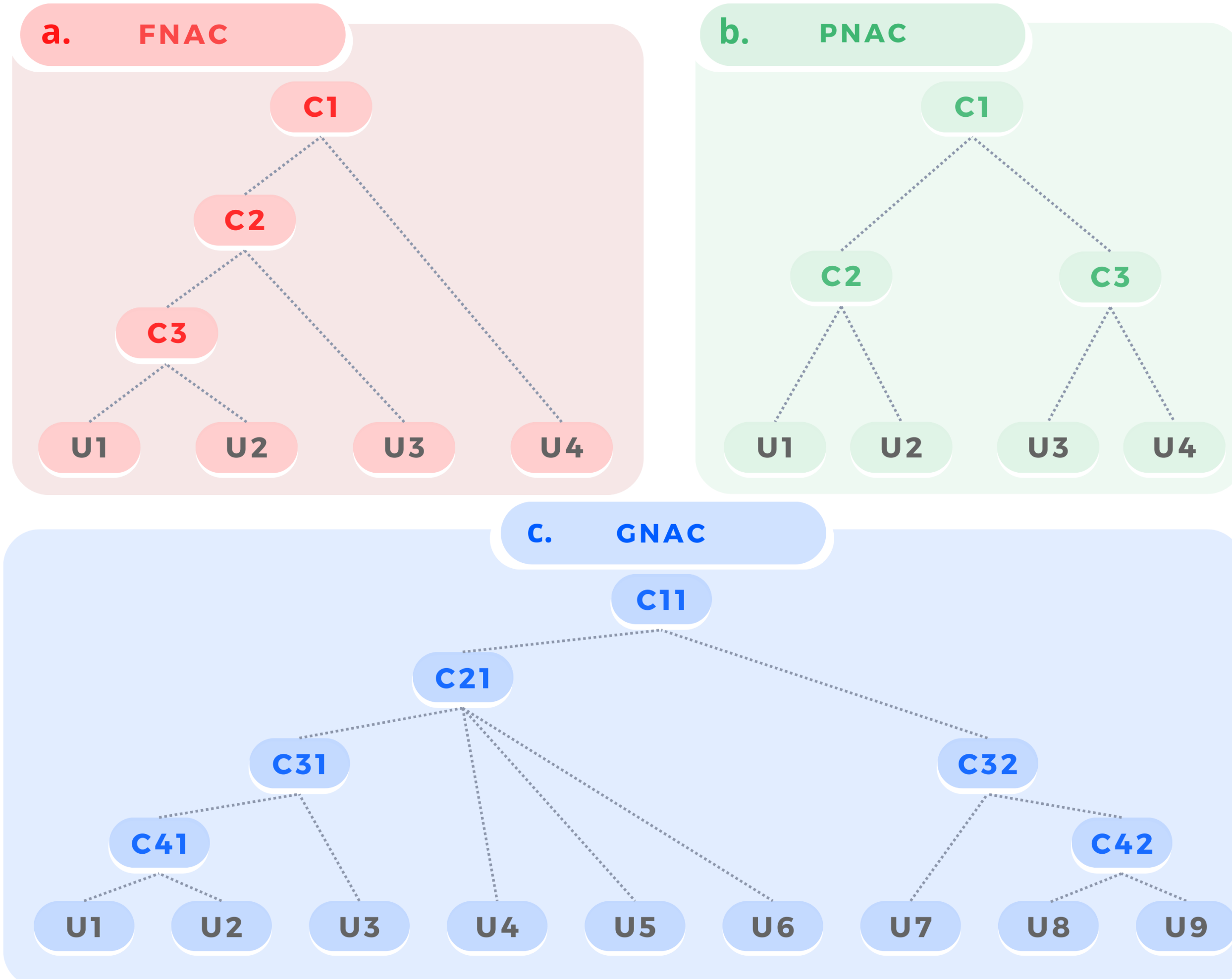

Figure S1. The construction of nested Archimedean copulas (NAC) involves different approaches: a) Fully Nested Archimedean Copulas (FNAC) in 4 dimensions, b) Partially Nested Archimedean Copulas (PNAC) in 4 dimensions, and c) the General Nested Archimedean copula (GNAC) structure in 9 dimensions. In these constructions, the variables are represented by U, while the nested copulas are represented by C.

To model diverse dependence structures, Grimaldi and Serinaldi (2006) implemented nested Archimedean copulas (NAC), also called asymmetric. This type of structure constitutes a significant improvement over EAC (Zhang and Singh, 2019). NACs are classified into Fully Nested Archimedean Copulas (FNAC), Partially Nested Archimedean Copulas (PNAC), and General Nested Archimedean copulas (GNAC) as seen in Fig. S1. For a more comprehensive explanation, see (Grimaldi and Serinaldi, 2006; Whelan, 2006; Aas and Berg, 2007; Savu and Trede, 2010; Zhang and Singh, 2019).

The mathematical representation of NACs can vary depending on the combination of variables and the proposed structure. Following (Joe, 1996; Whelan, 2006; Savu and Trede, 2010), a generalization of NAC can be written as shown in Eq. (S4):

$$
\begin{aligned}
C\left(u_1,\ldots,u_n\right) =& C_1\left(u_n, C_2\left(u_{n-1},\ldots,C_{n-1}\left(u_2,u_1\right)\cdots\right)\right) \\
=& \varphi_1^{-1}\left(\varphi_1\left(u_n\right)+\varphi_1\left(\varphi_2^{-1}\left(\varphi_2\left(u_{n-1}\right)+\cdots\right.\right.\right. \\
&\left.\left.\left. +\varphi_{n-1}^{-1}\left(\varphi_{n-1}\left(u_2\right)+\varphi_{n-1}\left(u_1\right)\right)\cdots\right)\right)\right)
\end{aligned}
\tag{S4}
$$

According to Aas and Berg (2007), this framework is not broad enough to model all possible mutual dependencies between variables, which is why other alternatives, such as Pair-Copula Construction (PCC), have been proposed in the literature.

S1.3.1 Vine copulas

An alternative to NACs is Pair-Copula Construction (PCC). In this type of copulas, the goal is to construct multivariate distributions using only bivariate building blocks. Compared to NACs, PCCs present a significant improvement by allowing the free specification of $d(d-1)/2$ copulas (Zhang and Singh, 2019), where d represents the number of variables. The two most cited types of PCC in the literature are Canonical vine (C-vine) and Drawable vine (D-vine). These types vary in how they structure a sequence of vine trees. Additionally, there is a third type called R-vine, which is a generalized form of structuring a sequence of vine trees and includes both D-vine and C-vine as special cases.

Here $f(|)$ and later $F(|)$ denote conditional densities and cumulative distribution functions of the variables. Let $(X_1, ..., X_d)$ be a set of variables with joint distribution $F_{1,\ldots,d}$ and density $f_{1,\ldots,d}$, respectively. According to Czado (2019), the density of a D-vine can be decomposed as show in Eq. (S5), and C-vine density can be decomposed as show in Eq. (S6):

$$
f_{1,\ldots,d}\left(x_1,\ldots,x_d\right) = \left[\prod_{j=1}^{d-1}\prod_{i=1}^{d-j} c_{i,(i+j);(i+1)\ldots,(i+j-1)}\right]\cdot\left[\prod_{k=1}^{d} f_k\left(x_k\right)\right]
\tag{S5}
$$

$$
f\left(x_1,\ldots,x_d\right) = \left[\prod_{j=1}^{d-1}\prod_{i=1}^{d-j} c_{j,j+i;1,\cdots,j-1}\right]\times\left[\prod_{k=1}^{d} f_k\left(x_k\right)\right]
\tag{S6}
$$

where index j identifies the trees, and i identifies the edges in each tree. $c_{i,j}$ is the copula associated with the bivariate conditional distribution. A more detailed and applied explanation can be found in Czado (2019).

Obtaining a joint cumulative distribution function from a vine copula is challenging from a theoretical perspective. Except in some special cases (such as when all pair copulas are Gaussian), there is no closed-form solution available. Therefore, two approaches are possible:

1. Using numerical integration. For this, you might need to adjust the pdf function so that it can be used as the integrand within Matlab's numerical integration methods. This approach can work in very low dimensions but definitely not for high-dimensional vine copulas.

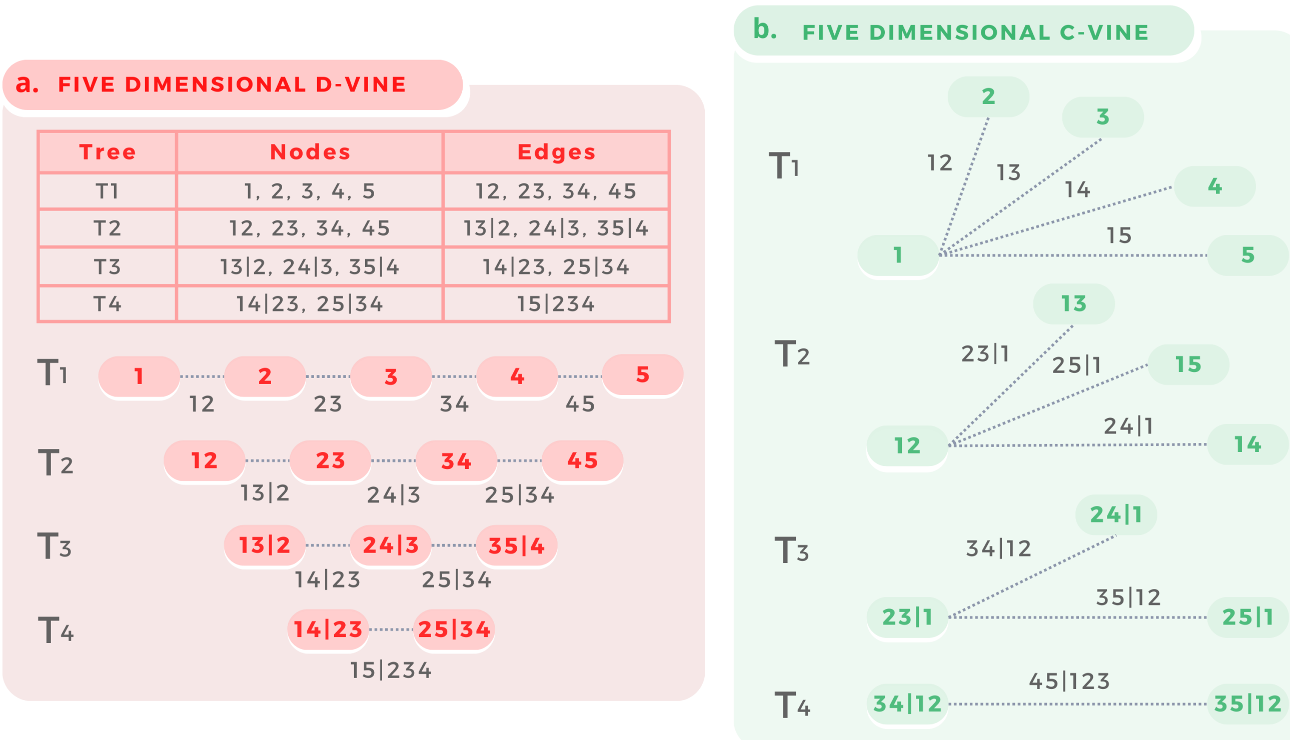

Figure S2. The figure shows the structure of a five-dimensional D-vine (left) and C-vine (right). The D-vine sequentially pairs variables through bivariate copulas across tree levels, while the C-vine centers each tree level on a key variable, conditioning dependencies on prior levels. Both structures are used to build multivariate distributions from bivariate copulas, capturing complex dependencies.

2. In the case of high dimension, a simulation-based approximation is proposed: simulate from your vine copula, which should be fast, even if you simulate a large sample. Use the empirical cumulative distribution function of the simulated sample as an approximation.

S1.3.2 Other copula families

Multivariate analysis and modeling in n dimensions offer various ways to address the dependence between variables. Among the most flexible structures are C-vine and D-vine, which serve as methods for modeling joint dependence between variables. On the other hand, theoretical copulas are not limited to the families mentioned. There are other types of copulas, such as the independence copula, used specifically for independent variables.

References

Aas, K. and Berg, D.: Models for Construction of Multivariate Dependence - a Comparison Study, The European Journal of Finance, 15, 639–659, http://www.tandfonline.com/doi/abs/10.1080/13518470802588767, 2007.

Carreau, J. and Bouvier, C.: Multivariate Density Model Comparison for Multi-Site Flood-Risk Rainfall in the French Mediterranean Area, Stochastic Environmental Research and Risk Assessment, 30, 1591–1612, https://doi.org/10.1007/s00477-015-1166-6, 2016.

Chen, L. and Guo, S.: Copulas and Its Application in Hydrology and Water Resources, Springer Water, Springer Singapore, Singapore, http://link.springer.com/10.1007/978-981-13-0574-0, 2019.

Czado, C.: Analyzing Dependent Data with Vine Copulas: A Practical Guide With R, vol. 222 of *Lecture Notes in Statistics*, Springer International Publishing, Cham, http://link.springer.com/10.1007/978-3-030-13785-4, 2019.

Genest, C. and Favre, A.-C.: Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask, Journal of Hydrologic Engineering, 12, 347–368, https://ascelibrary.org/doi/10.1061/%28ASCE%291084-0699%282007%2912%3A4%28347%29, 2007.

Genest, C., Favre, A.-C., Béliveau, J., and Jacques, C.: Metaelliptical Copulas and Their Use in Frequency Analysis of Multivariate Hydrological Data, Water Resources Research, 43, https://onlinelibrary.wiley.com/doi/abs/10.1029/2006WR005275, 2007.

Genton, M. G., ed.: Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality, Chapman and Hall/CRC, New York, 2004.

Grimaldi, S. and Serinaldi, F.: Asymmetric Copula in Multivariate Flood Frequency Analysis, Advances in Water Resources, 29, 1155–1167, https://www.sciencedirect.com/science/article/pii/S0309170805002307, 2006.

Gudendorf, G. and Segers, J.: Extreme-Value Copulas, in: Copula Theory and Its Applications, edited by Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T., Lecture Notes in Statistics, pp. 127–145, Springer, Berlin, Heidelberg, 2010.

Hao, C., Zhang, J., and Yao, F.: Multivariate Drought Frequency Estimation Using Copula Method in Southwest China, Theoretical and Applied Climatology, 127, 977–991, https://doi.org/10.1007/s00704-015-1678-5, 2017.

Jaser, M. and Min, A.: On Tests for Symmetry and Radial Symmetry of Bivariate Copulas Towards Testing for Ellipticity, Computational Statistics, 36, 1–26, https://doi.org/10.1007/s00180-020-00994-0, 2021.

Joe, H.: Families of m-Variate Distributions with Given Margins and m(m-1)/2 Bivariate Dependence Parameters, Lecture Notes-Monograph Series, 28, 120–141, http://www.jstor.org/stable/4355888, 1996.

Kotz, S. and Nadarajah, S.: Multivariate T-Distributions and Their Applications, Cambridge University Press, Cambridge, https://www.cambridge.org/core/books/multivariate-tdistributions-and-their-applications/463E067E32C230D3205563EDD195DA2D, 2004.

Nelsen, R. B.: An Introduction to Copulas, Springer Series in Statistics, Springer, New York, NY, http://link.springer.com/10.1007/0-387-28678-0, 2006.

Ribeiro, A. F., Russo, A., Gouveia, C. M., and Pires, C. A.: Drought-Related Hot Summers: A Joint Probability Analysis in the Iberian Peninsula, Weather and Climate Extremes, 30, 100 279, https://linkinghub.elsevier.com/retrieve/pii/S2212094720300062, 2020.

Savu, C. and Trede, M.: Hierarchies of Archimedean Copulas, Quantitative Finance, 10, 295–304, https://doi.org/10.1080/14697680902821733, 2010.

Whelan, N.: Sampling from Archimedean Copulas, Quantitative Finance, https://www.tandfonline.com/doi/abs/10.1088/1469-7688/4/3/009, 2006.

Zhang, L. and Singh, V. P.: Copulas and Their Applications in Water Resources Engineering, Cambridge University Press, 1 edn., https://www.cambridge.org/core/product/identifier/9781108565103/type/book, 2019.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., van den Hurk, B., AghaKouchak, A., Jézéquel, A., Mahecha, M. D., Maraun, D., Ramos, A. M., Ridder, N. N., Thiery, W., and Vignotto, E.: A Typology of Compound Weather and Climate Events, Nature Reviews Earth & Environment, 1, 333–347, https://www.nature.com/articles/s43017-020-0060-z, 2020.