Blessing and Curse of Multivariate Return Periods

Multivariate Return Periods based on Vine Copulas

B. Gräler (1), S. Vandenberghe (2), M.J. van den Berg (2), S. Grimaldi (3,4,5), A. Petroselli (6), B. De Baets (7) & N.E.C Verhoest (2)

(1) University of Muenster, Institute for Geoinformatics, Münster, Germany, (2) Laboratory of Hydrology and Water Management, Ghent University, Coupure links 653, 9000 Ghent, Belgium, (3) Dipartimento per la innovazione nei sistemi biologici agroalimentari e forestali (DIBAF Department), University of Tuscia, Via San Camillo De Lellis snc, 01100 Viterbo, Italy, (4) Honors Center of Italian Universities (H2CU), Sapienza University of Rome, Via Eudossiana 18, 00184 Roma, Italy, (5) Department of Mechanical and Aerospace Engineering, Polytechnic Institute of New York University, Six MetroTech Center, Brooklyn, NY 11201, (6) Dipartimento di scienze e tecnologie per l'agricoltura, le foreste, la natura e l'energia (DAFNE Department), University of Tuscia, Via San Camillo De Lellis snc, 01100 Viterbo, Italy, (7) Department of Mathematical modelling, Statistics and Bio-informatics, Coupure links 653, 9000 Ghent, Belgium

(i) Multivariate Return Periods

Return periods describe the inter-arrival time T of extreme events usually denoting the border between the fraction of $(1 - 1/T)$ harmless and $1/T$ harmful events.

Extremes do not have a unique meaning in more than one-dimensional space as no natural ordering exists.

Design events cannot be chosen uniquely.

Copulas representing multivariate CDFs can provide multiple different critical events, resulting in *different sub-critical regions* (drawn in blue).

Kendall Return Periods, based on a copula's Kendall function, can be used to identify events defining a *critical layer*, which is defined through the cumulative distribution function as in the univariate approach identifying a *unique sub-critical region* (drawn as red hatching).

The Kendall function

$$K_C(t) = \int_{[0,1]^d} \mathbf{1}_{C(\mathbf{u}) \leq t}(\mathbf{u}) \, \mathbf{du}$$

Ensembles of possible design events might circumvent the drawback of non unique selection criteria.

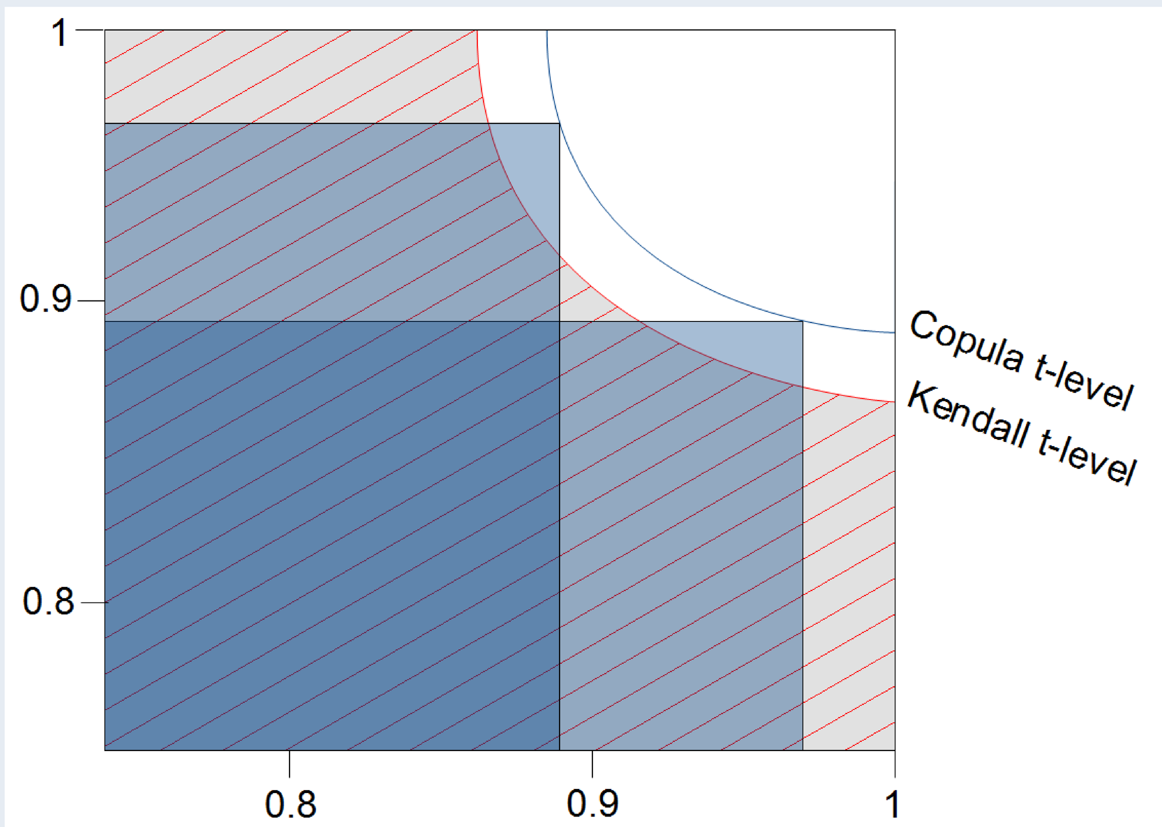


Figure: Illustration of sub-critical regions in the top right corner of the unit-square for the copula and Kendall approach. Blue rectangles indicate possible sub-critical regions for the copula approach; the red hatching indicates the unique sub-critical region of the Kendall approach. Possible design events lie along both level curves.

(ii) Vine Copulas . . .

- . . . approximate multivariate copulas, modeling *multivariate dependence structures*.
- . . . of dimension d rely on $d(d-1)/2$ bivariate copula building blocks.
- . . . allow to mix different copula families without limitations and are thus *very flexible*.
- . . . iteratively re-use well established estimation procedures for bivariate copulas.
- . . . combined with independently fitted marginal distributions yield a *full multivariate distribution* of the observed phenomenon.
- . . . allow to derive conditional densities from their multivariate density.

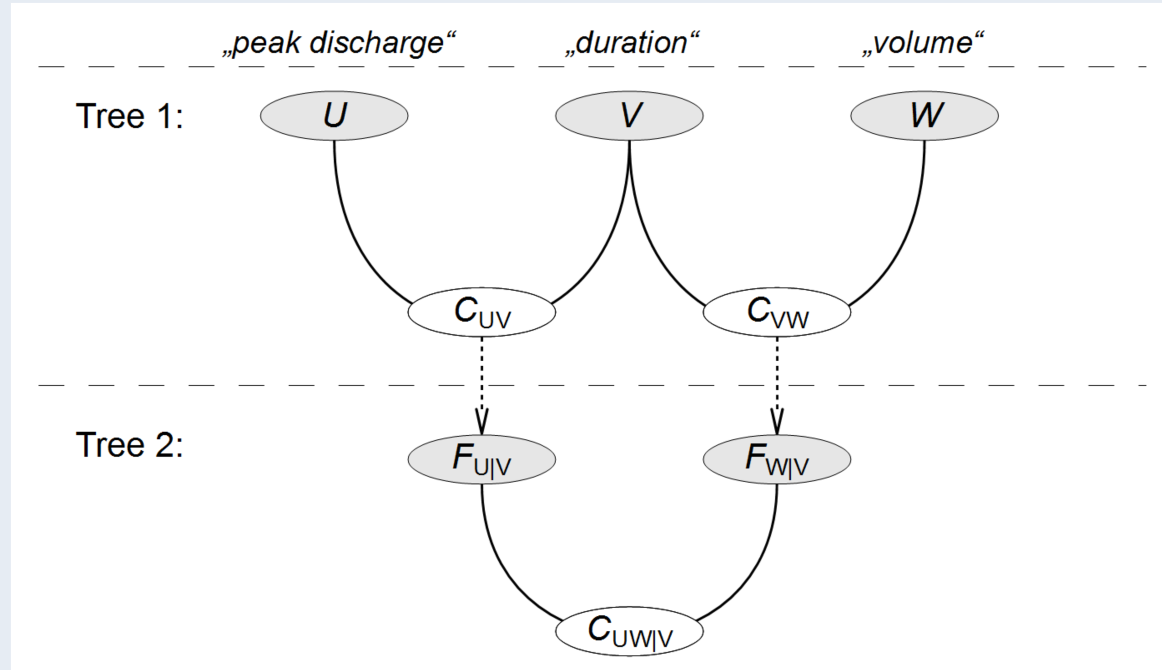


Figure: graphical representation of the 3D-vine copula

3D copula density

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

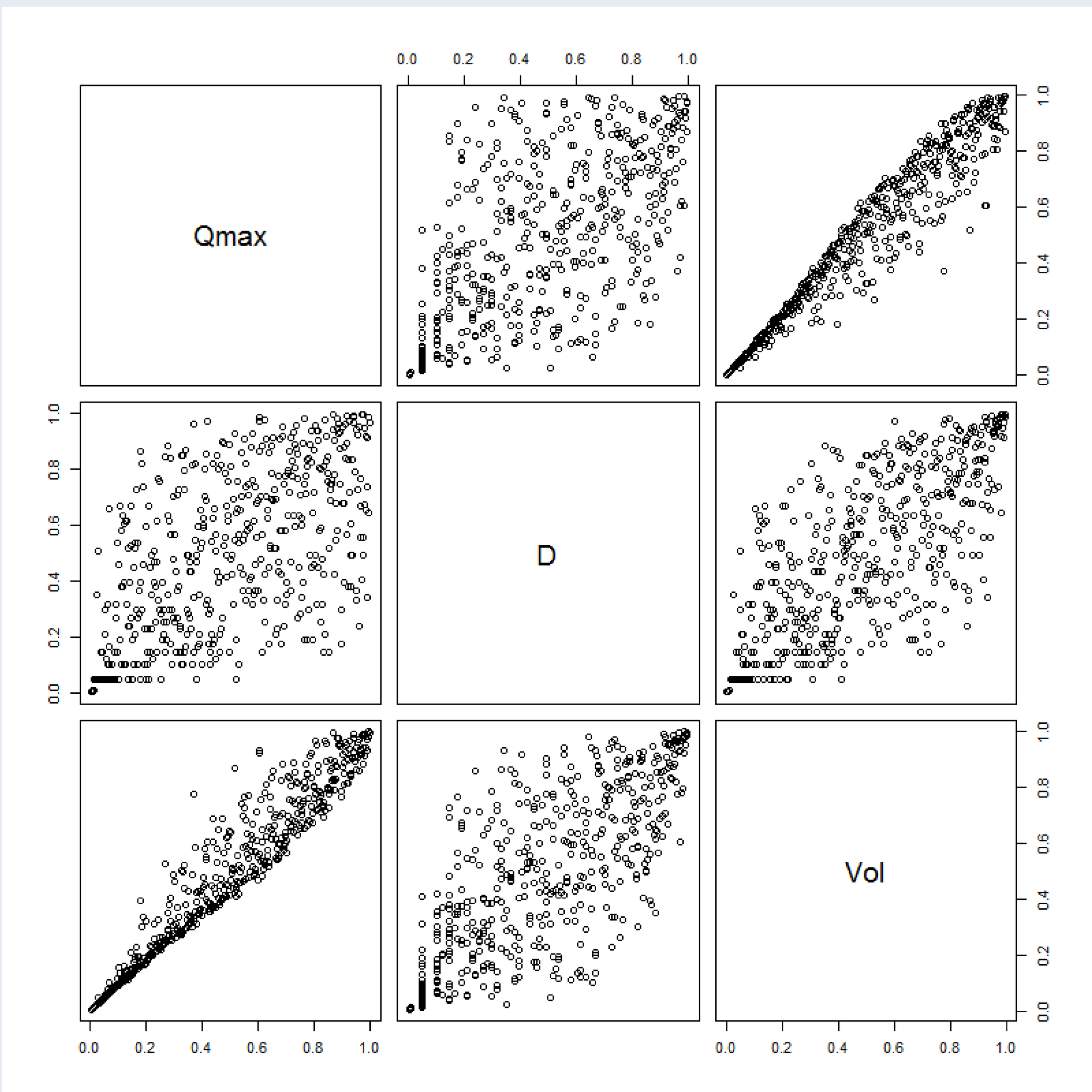
$$\text{where } F_{U|V}(u|v) = \frac{\partial C_{UV}(u,v)}{\partial v} \quad \text{and} \quad F_{W|V}(w|v) = \frac{\partial C_{VW}(v,w)}{\partial v}$$

3D full density

$$f(q_{max}, d, vol) = c\big(F_{Q_{max}}(q_{max}), F_D(d), F_{Vol}(vol)\big) \cdot f_{Q_{max}}(q_{max}) \cdot f_D(d) \cdot f_{Vol}(vol)$$

where $F_{Q_{max}}, F_D, F_{Vol}$ and $f_{Q_{max}}, f_D, f_{Vol}$ are the marginal CDFs and PDFs respectively.

(iii) Application



Data: 500 simulated triples (Q_{max}, D, Vol) of annual flood maxima at the Torbido River, central Italy

First vine tree: C_{UV}: Survival BB7 copula with parameters $\theta = 2.05$ and $\delta = 0.35$
C_{VW}: Survival BB7 copula with parameters $\theta = 2.25$ and $\delta = 1.09$

Second vine tree: $C_{UW|V}$: Student copula with parameters $\rho = 0.96$ and $\nu = 4$ degrees of freedom

Log-likelihoods: vine copula: 1047;
three-variate Gaussian copula: 935;
three-variate Archimedean copulas: 432 - 532

Three-variate return periods are derived through simulation and numerical integration.

Implementations and analysis are done in R, the developed tools can be accessed through the package *spcopula* on r-forge.

(iv) Conclusion & Outlook

CURSE

- The lack of a natural ordering involves choices affecting the results.
- Single design events from the Kendall approach might not well enough represent the sub-critical region.
- Different design events from the copula approach do not generally correspond to the same sub-critical region.

BLESSING

- Based on the log-likelihood, the vine copula is in favor of the 3-dimensional Gaussian or Archimedean copula.
- A three-variate return period improves the analysis of the phenomenon.
- The Kendall and the Copula approach can both be realized.

A comparison for a whole set of *uni- and multivariate approaches* will soon be published.
You are invited to attend the *talk Friday, 27 Apr, 15:30, Room: 34, session: HS7.5/NP8.3 (PReZI)*.

Contact

Benedikt Gräler
Institute for Geoinformatics
Weseler Str. 253
48151 Münster, Germany

ben.graeler@uni-muenster.de
`http://www.ifgi.de/graeler`

ifgi
Institute for Geoinformatics
University of Münster