# The R-package *CDVineCopulaConditional* for modelling Compound Events


E. Bevacqua (emanuele.bevacqua@uni-graz.at)[1], D. Maraun[1], I. Hobæk Haff[2], M. Vrac[3], M. Widmann[4], and C. Manning[4]

[1]Wegener Center for Climate and Global Change, University of Graz, Graz, Austria.
[2]Department of Mathematics, University of Oslo, Oslo, Norway.
[3]Laboratoire des Sciences du Climat et de l'Environnement, CNRS/IPSL, Gif-sur-Yvette, France.
[4]School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK.


## Introduction

Given a joint probability density function (pdf), copulas allows for modelling the dependence structure of the variables separately from their marginal pdfs. Copulas provide flexibility when modelling joint pdfs, and therefore have been largely used in science, e.g. in quantitative finance to model and minimize tail risk, in hydrology and more recently for modelling Compound Events (CEs). However, multivariate parametric copulas lack flexibility when modelling systems with high dimensionality, where "heterogeneous dependencies" exist among the different pairs. Pair-copula constructions (PCCs) decompose the pdf dependence structure in bivariate copulas and give greater flexibility in modelling generic high-dimensional systems compared to multivariate parametric copulas. Given the complex structure of some CEs, PCCs can be useful for their modelling. Moreover, joint conditional pdfs can be used to get information about the impact or the contributing variables of CEs, given proper predictors, which could represent for instance meteorological processes. The R-package *CDVineCopulaConditional* [Bevacqua (2017)] provides tools for modelling conditional joint pdfs decomposed via Pair-Copula Constructions.

## Pair-Copula Construntions

### *Copulas*

Consider a vector $\vec{Y} = (Y_1, ..., Y_n)$ of random variables, with marginal pdfs $f_1(y_1), ..., f_n(y_n)$, and continuous cumulative marginal distribution functions (CDFs) $F_1(y_1), ..., F_n(y_n)$. The joint pdf can be decomposed as:

$$f(y_1, ..., y_n) = f_1(y_1) \cdot ... \cdot f_n(y_n) \cdot c(u_1, ..., u_n) \quad (1)$$

where $c$ is the copula density and $U_i := F_i(Y_i)$.

- Copulas make it easy to construct a wide range of multivariate parametric distributions.
- However, the set of copula families having dimension $\geq 3$ is rather limited, and they lack flexibility in modelling "heterogeneous dependencies".

### *PCCs*

If the joint CDF is absolutely continuous, with strictly increasing marginal CDFs, PCCs allow to decompose the copula density. For example:

$$\begin{aligned} f_{12345}(y_1, y_2, y_3, y_4, y_5) = & f_1(y_1) \cdot f_2(y_2) \cdot f_3(y_3) \cdot f_4(y_4) \cdot f_5(y_5) \\ & \cdot c_{12}(u_1, u_2) \cdot c_{23}(u_2, u_3) \cdot c_{34}(u_3, u_4) \cdot c_{45}(u_4, u_5) \\ & \cdot c_{13|2}(u_{1|2}, u_{3|2}) \cdot c_{24|3}(u_{2|3}, u_{4|3}) \cdot c_{35|4}(u_{3|4}, u_{5|4}) \\ & \cdot c_{14|23}(u_{1|23}, u_{4|23}) \cdot c_{25|34}(u_{2|34}, u_{5|34}) \\ & \cdot c_{15|234}(u_{1|234}, u_{5|234}) \end{aligned} \quad (2)$$

PCCs provide higher flexibility in building high dimensional joint pdfs with respect to using the existing multivariate parametric copulas [Aas et al. (2009)].

### *Vines*

- When the dimension of the pdf is large, there can be many possible mathematically equally valid PCCs. *The regular vine* is a graphical model which helps to organize the possible decompositions.
- Here we concentrate on *C-* (or *canonical*) and *D-vines*, subcategories of the regular vines.

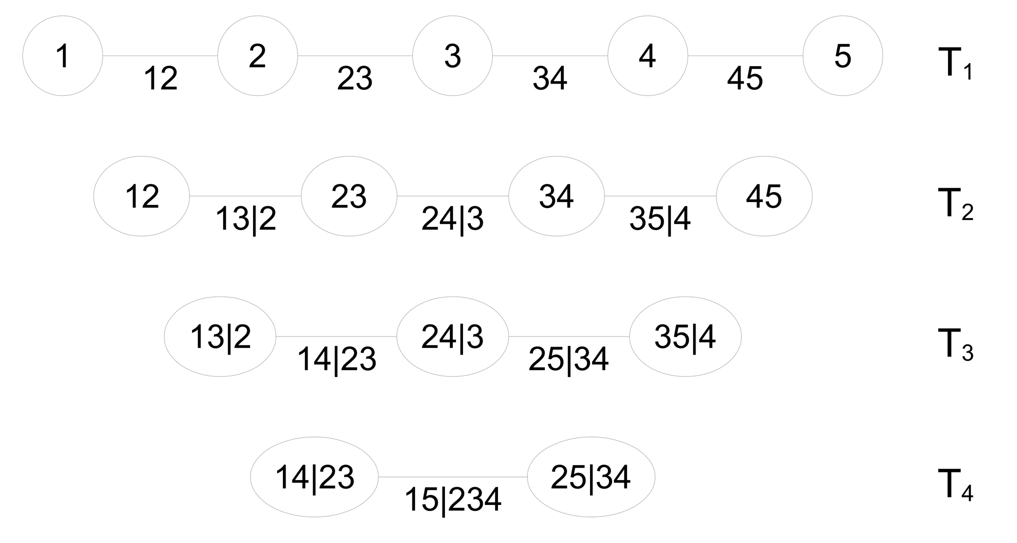


Figure 1: Representation of the 5-dim D-vine of eq. (2). There are 4 trees ($T_1, T_2, T_3, T_4$), and 10 edges. Each edge represents a pair-copula density, and the label indicates the subscript of the corresponding copula.

## Conditional Sampling

Not all of the PCCs associate to the joint pdf $f_{\vec{Y},\vec{X}}(\vec{Y}, \vec{X})$ can be used to sample from the conditional joint pdf $f_{\vec{Y}|\vec{X}}(\vec{Y}|\vec{X})$, i.e. to sample $\vec{Y}$, given fixed values of $\vec{X}$ . Here, the vines used for conditionally sampling are those which sample as first the conditioning variables (when following the sampling algorithms from [Aas et al. (2009)]). Given $\vec{X} = X_1, ..., X_{N_x}$ and $\vec{Y} = X_1, ..., X_{N_y}$, the number of such vines is:

- D-vines: $N_x! \cdot N_y!$.
- C-vines: $N_x! \cdot N_y!/2$ for $N_y > 1$; $N_x!$ for $N_y = 1$.

### *CDVineCopulaConditional functions*

- *CDVineCondFit*, *CDVineCondListMatrices*, *CDVineCondRank*: (1) PCC selection (according to AIC or BIC), among those which allow for such a conditional sampling; (2) pair-copula families fit.
- *CDVineCondSim*: conditional sampling based on the algorithms shown in [Bevacqua et al. (2017)].

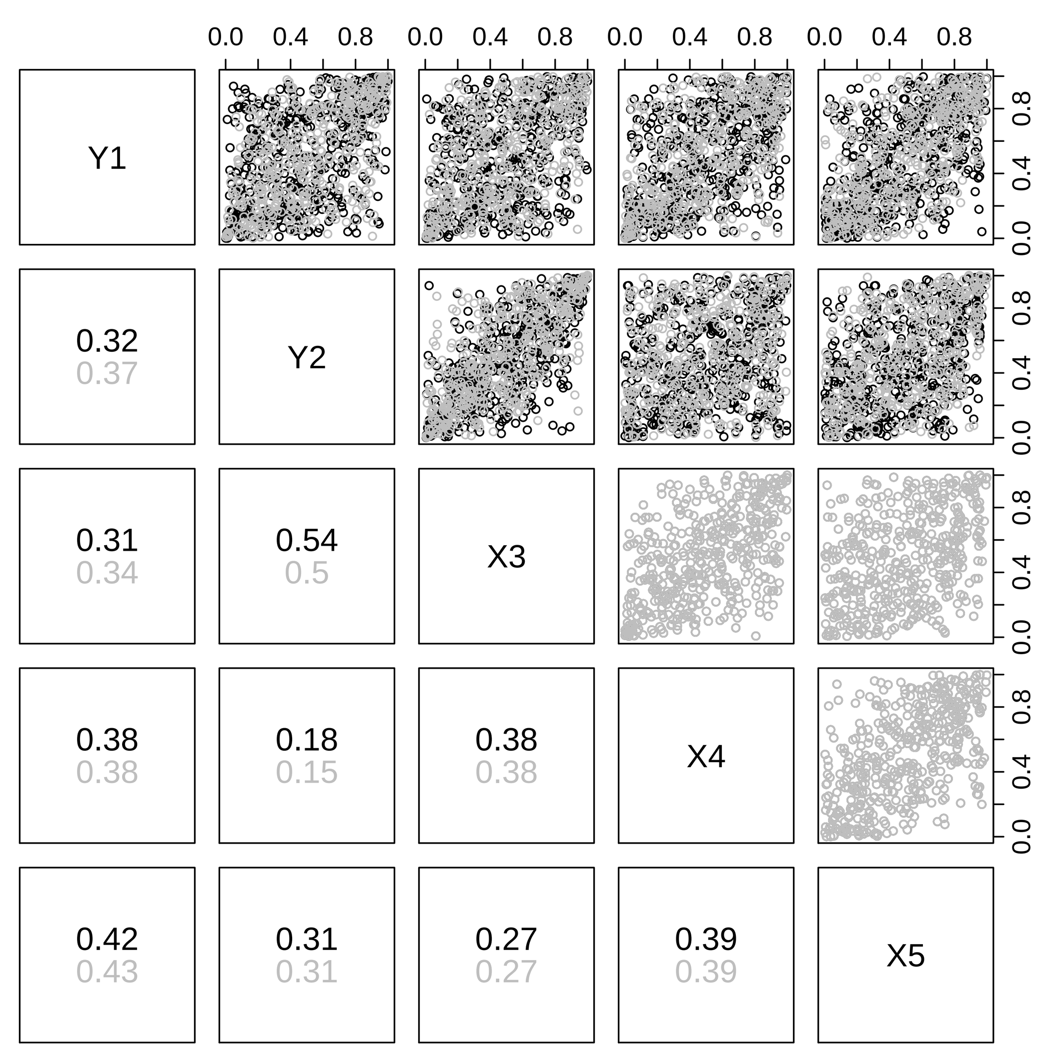


Figure 2: Scatterplot matrix of observed uniform data $(\vec{Y}^{obs}, \vec{X}^{obs})$ (black) vs simulated from $f_{\vec{Y}|\vec{X}}(\vec{Y}|\vec{X}^{obs})$ (grey). The numbers are the Kendall Tau correlation coefficients. Plot obtained via the *overplot* function.

## Applications to Compound Events

We define a conceptual model for CEs, which consists of:

1. Contributing variables $Y_i$ to the CE;
2. Impact of the CE: $h = h(Y_1, ..., Y_n)$, e.g. a river gauge level, agricultural yield or economic loss;
3. Predictors $X_i$ of $Y_i$, which provide insight into the physical processes underlying the CE.

PCC can be used to properly represent the dependencies of the variables and their marginal distributions. The conditional joint pdf can be used to implement such a conceptual model as follows:

- Compound floods in Ravenna (Italy). $f_{\vec{Y}|\vec{X}}(\vec{Y}|\vec{X})$ provides multivariate statistical downscaling [Bevacqua et al. (2017)];

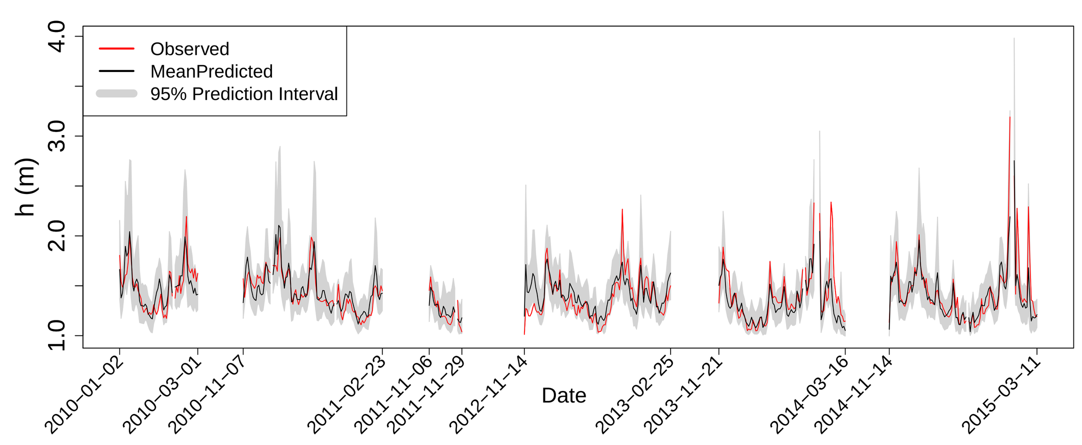


Figure 3: River gauge level $h = h(Y_{Sea}, Y_{River_1}, Y_{River_2})$, which is influenced both by the Sea and River levels. $h^{model}$ is the model 6-fold cross-validation time series.

- $f_{h|\vec{Y}}(h|\vec{Y})$ can be used to estimate $h$ if complex relations exist between the $h$ and the variables $Y$, and therefore it may be difficult to implement, e.g., a proper regression model to describe the impact $h$. See poster [Manning et al. (2017)] about Soil Moisture Drought.
- If the impact $h$ is not available, $f_{\vec{Y}|\vec{X}}(\vec{Y}|\vec{X})$ may be useful to estimate/downscale the variables $Y$, to then assess the risk of CEs through, e.g., multivariate return periods of the contributing variables $Y$.
- If the variables $Y$ are not available, $f_{h|\vec{X}}(h|\vec{X})$ may be used to directly estimate/downscale the impact $h$. In this case, depending on the physical system, it may be more or less complicated to calibrate the predictors.

### *References*


[Aas et al. (2009)] Aas, K., Czado, C., Frigessi, A. and Bakken, H.: Pair-copula constructions of multiple dependence, Insurance: Mathematics and Economics, 44(2), 182–198, doi:10.1016/j.insmatheco.2007.02.001, 2009.

[Bevacqua et al. (2017)] Bevacqua, E., Maraun, D., Hobæk Haff, I., Widmann, M., and Vrac, M.: Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy), Hydrol. Earth Syst. Sci., 21, 2701-2723, https://doi.org/10.5194/hess-21-2701-2017, 2017.

[Bevacqua (2017)] Bevacqua, E. (2017). CDVineCopulaConditional: Sampling from Conditional C- and D-Vine Copulas. R package version 0.1.0. https://CRAN.R-project.org/package=CDVineCopulaConditional.