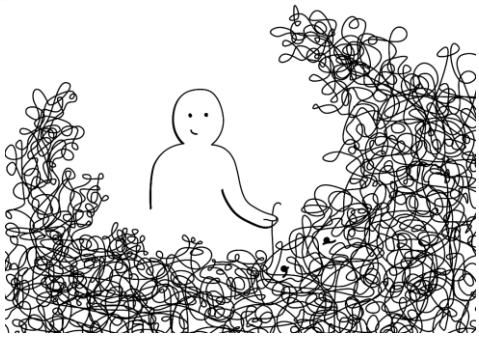


# The state of the art of using Copulas in hydroclimatic applications,

Faranak Tootoonchi
*PhD student,*
*Environmental Impact Assessment*

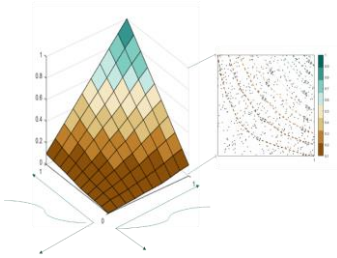
**An article by:** Faranak Tootoonchi, Jan Haerter, Olle Raty, Thomas Grabs, Mojtaba Sadegh, Claudia Teutschbein

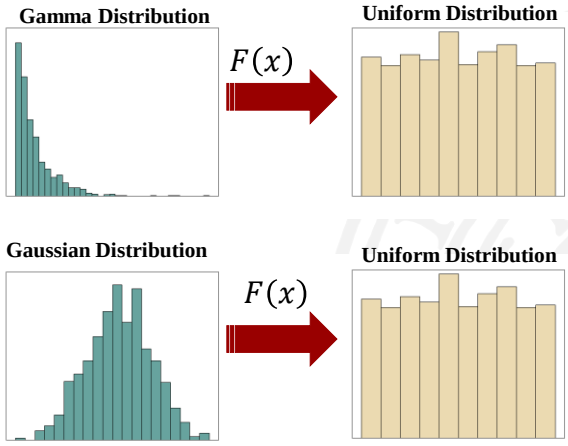
## Why to use multi-variable methods?

- Hydrological and climatological variables are interdependent, and we often use multivariate methods to understand their interactions and associations.
- Examples include storm duration and intensity, flood peak and volume, concurrent drought and heat waves, drought duration and severity, and precipitation and soil moisture.
- Univariate distributions may not be sufficient to describe hydrological variables (or events) that bear intrinsic multivariate characteristics
- The concept of copula, as one approach from the cohort of several multivariate analysis methods, is widely used to model the dependence structure of two (or more) random variables

# What is Copula?

$$H(x_1, x_2, \dots, x_d) = C\big(F_1(x_1), F_2(x_2), \dots, F_d(x_d)\big)$$






# Empirical Copula

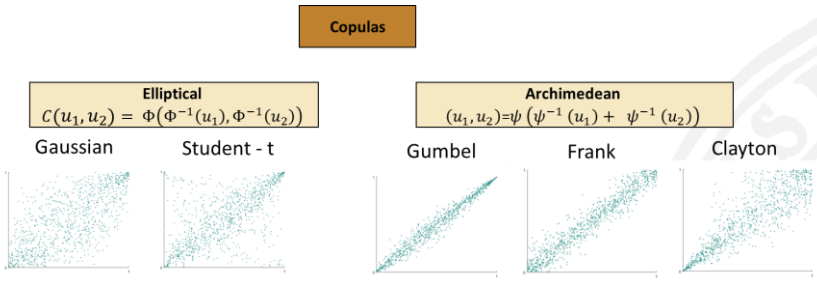
In order to perceive the underlying co-dependent probability between two *random* variables, the **joint rank correlation** of the variables should be investigated. The empirical Copula is **cumulative distribution of the rank transformed variables** (Charpentier, Fermanian and Scaillet, 2007).

The empirical (bivariate case) Copula is defined as the discrete function $C_n$ given by:

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) = \#\frac{\{(x_1, x_2) | x_1 \leq X_1 \; and \; x_2 \leq X_2\}}{n}$$

Where $x_1$and $x_2$ denote the order statistics of the sample and # provides the cardinality of the subsequent set. The result would be the joint CDF of uniformly distributed marginals.

# Theoretical Copula




# Copula parameter estimation

## Empirical correlation

- **Pearson correlation**:

$$\rho_{x_1 x_2} = \frac{\mathrm{Cov}(x_1, x_2)}{\sigma_{x_1}\sigma_{x_2}}$$

- **Kendall's $\tau$**:

$$\tau = \frac{\#\ concordant\ pairs - \#\ discordant\ pairs}{\frac{n(n-1)}{2}}$$

- **Spearman's S**:

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

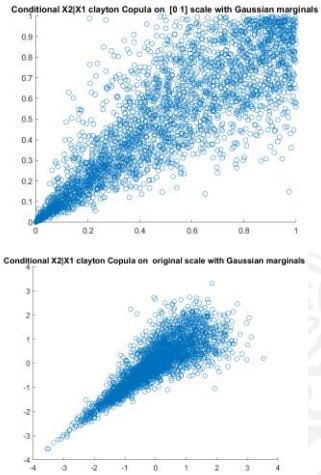
## Theoretical Copula parameter

- In Elliptical Copulas, $\rho$ can be considered as the Copula parameter.
- In Archimedean Copulas, $\tau$ or $s$ are used to be build the Copula parameter.

# Simulating from Copulas (Conditional Copula )

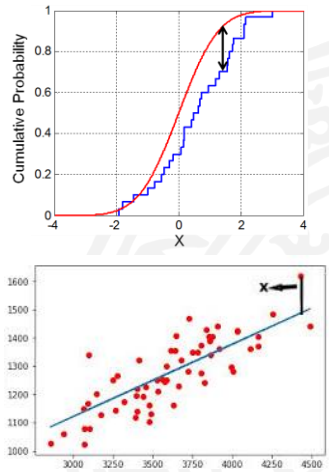
- Copulas can be used to compute the conditional distribution of one variable given that the distribution of the other variable is known. An example is the estimation of the conditional probability distribution of precipitation, given one particular temperature probability and the known joint probability distribution.

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




# Goodness of Fit Measures for choosing proper Copula family

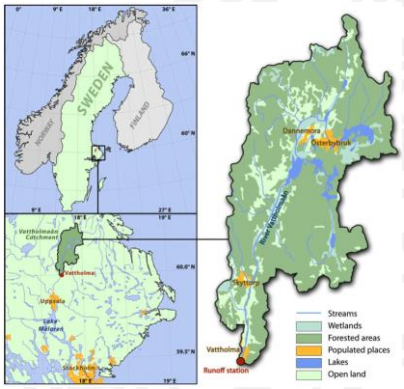
- **Cramer Von-Mises** $S_n n \sum_{i=1}^{n} |C_n - C_\theta|^2$
- **Kolmogorov–Smirnov** $D_n = \sup|C_n(x) - C_\theta(x)|$
- **Nash-Sutcliffe Efficiency** $NSE = 1 - \frac{\sum_{i=1}^{n}|C_n(x)-C_\theta(x)|^2}{\sum_{i=1}^{n}|C_n(x)-\bar{C}_n(x)|^2}$
- **Root mean square error** $RMSE = \sqrt{\frac{\sum_{i=1}^{n}|C_n(x)-C_\theta(x)|^2}{n}}$



# Common issues, misconceptions and pitfalls

Scopus : Precipitation + Temperature + Copula (Returned 58 papers)

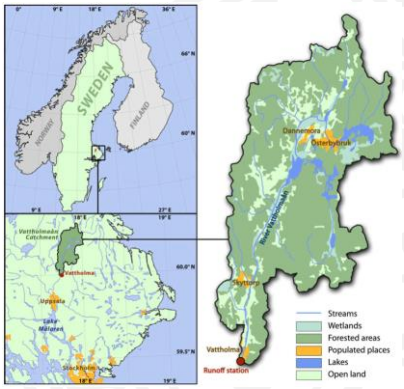
- What is the **spatial scale** of the study site(s)?
- What is the **temporal resolution** of temperature and precipitation data?
- Does the paper check the **correlation between temperature and precipitation**?
- Does the paper consider **autocorrelation** within each of the variables?




# Common issues, misconceptions and pitfalls

Scopus : Precipitation + Temperature + Copula (Returned 58 papers)

- Does the paper consider zero-precipitation, e.g. by introduction of a **precipitation threshold**?
- Does the paper reflect on the **stationarity** issue?
- Were several copula families fitted and tested for **goodness of fit**?



## Common issues, misconceptions and pitfalls

### Spatial scale

- Almost half of the paper were on large scale areas.

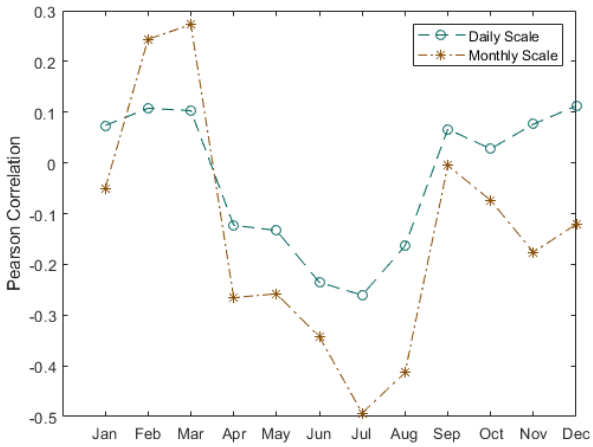
### Temporal resolution

- 65% of the papers were on daily scale

### Correlation between variables

Less than half of all papers (45%) transparently stated the correlation strength and represented the correlation of the data either graphically or with one of the correlation statistical measures.

## Common issues, misconceptions and pitfalls

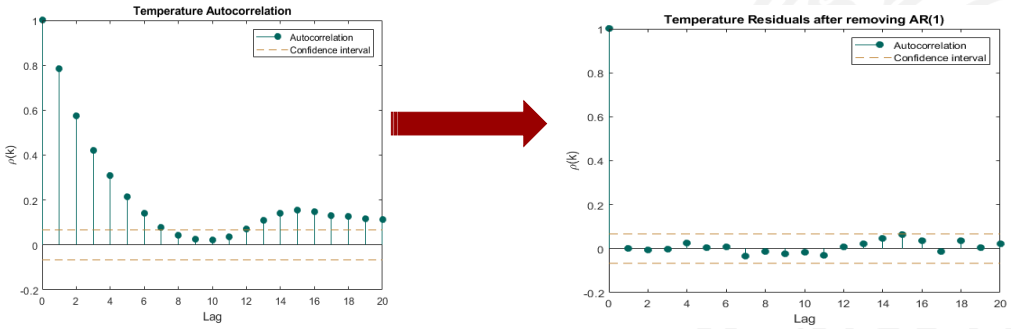



Correlation between temperature and precipitation at different temporal scales

## Common issues, misconceptions and pitfalls

### Autocorrelation

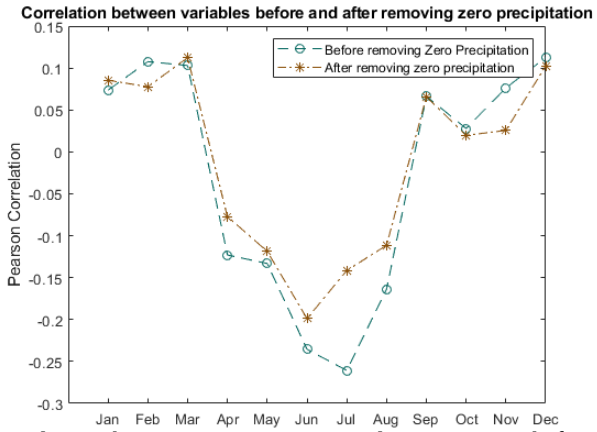
75% of the reviewed papers did not mention if their data exhibits autocorrelation.




## Common issues, misconceptions and pitfalls

### Precipitation Threshold

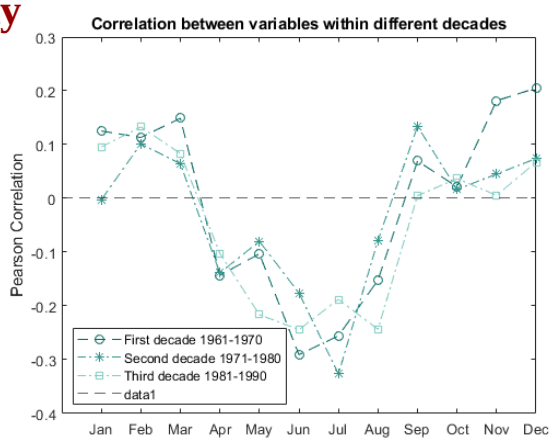
76% of the publications did not even mention this particular issue.




Correlation between precipitation and temperature before (brown) and after (blue) removing drizzle days

## Common issues, misconceptions and pitfalls

### Stationarity



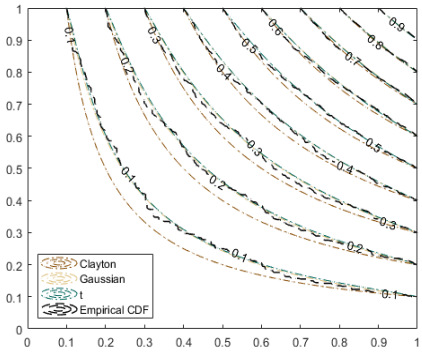

The variability of correlation between precipitation and temperature for each month (January – December) based on assuming different time-spans.

## Common issues, misconceptions and pitfalls

### GoF

Almost half of the papers (55 %) checked for Goodness of their Copula fits

| Goodness-of-fit measure | Copula family | | |
|---|---|---|---|
| | Gaussian | t | Clayton |
| RMSE (ranking) | 0.3252 (2) | 0.318 (1) | 0.4062 (3) |
| NSE (ranking) | 0.9956 (2) | 0.996 (1) | 0.9931 (3) |
| CVM (ranking) | 1121 (1) | 1356 (2) | 3436 (3) |
| Dn (ranking) | 0.0199 (2) | 0.0197 (1) | 0.0341 (3) |



## Concluding Remarks

Copula is not a magical tool that can handle all uncertainties in a multidimensional context.
If one wishes to benefit from this framework, there are important credential to take into account; most importantly the reliability of **correlation** between variables.