netherlands eScience center

Special Interest Group (SIG)-statistics

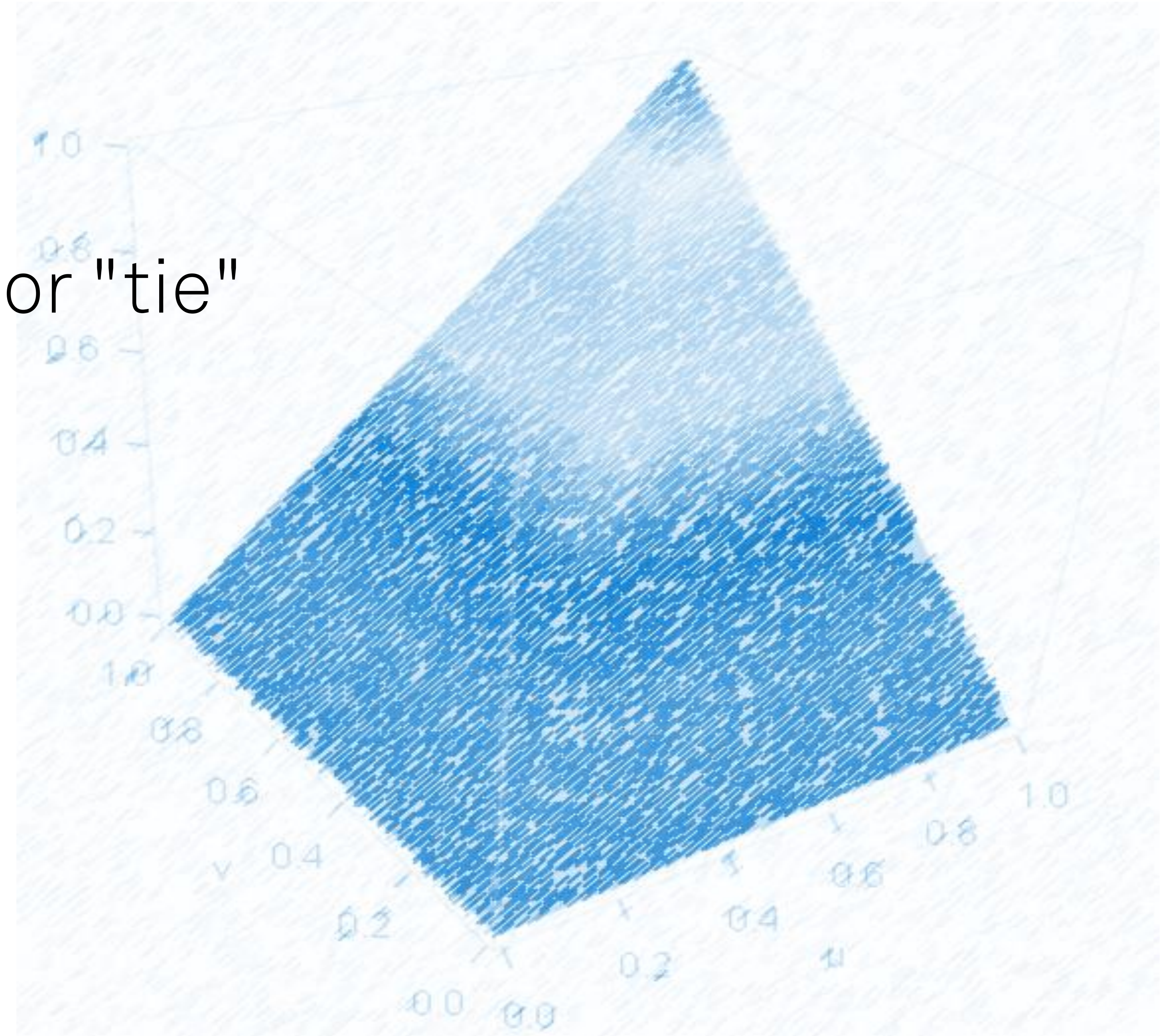
Fakhereh (Sarah) Alidoost

Amsterdam, 18/07/2019

netherlands eScience center

Copula /kɒpjʊlə/

The name comes from the Latin for "link" or "tie"



1) Introduction to copulas:

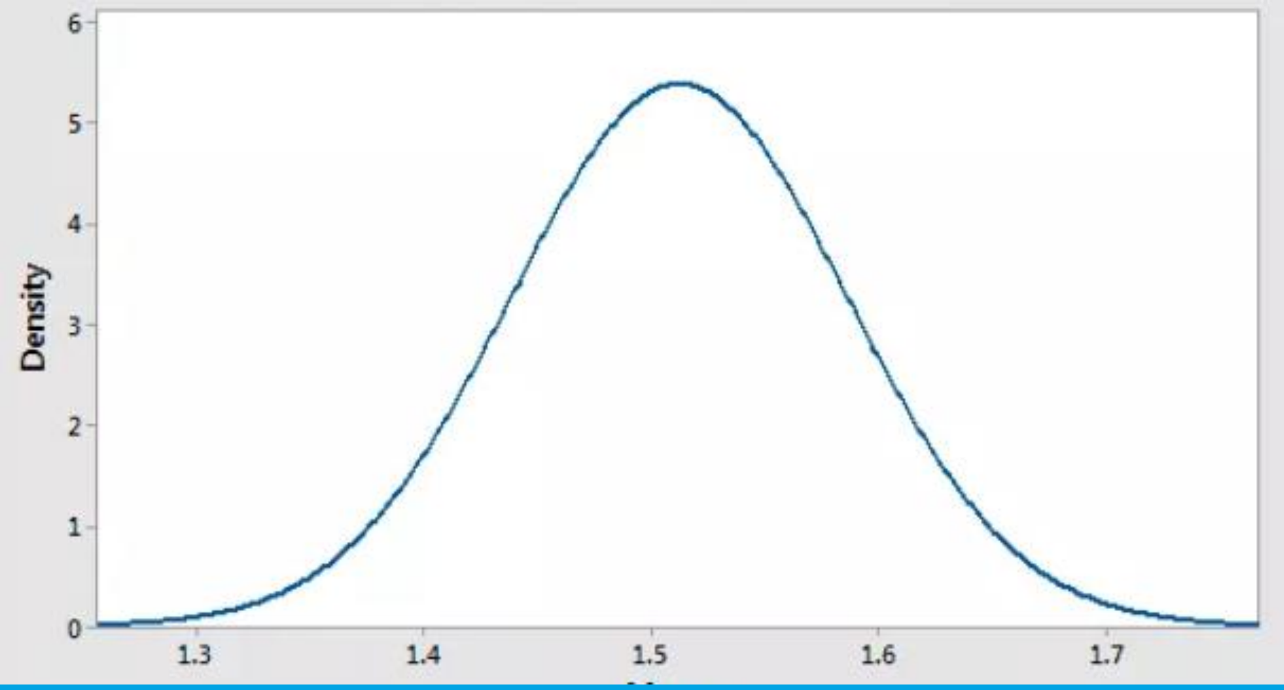
- Notation (Cumulative distribution function / probability density function)
- Dependence and joint distribution
- Distribution family
- **Estimation** vs. prediction
- Multivariate joint distribution: The world of vines

2) Application of copulas:

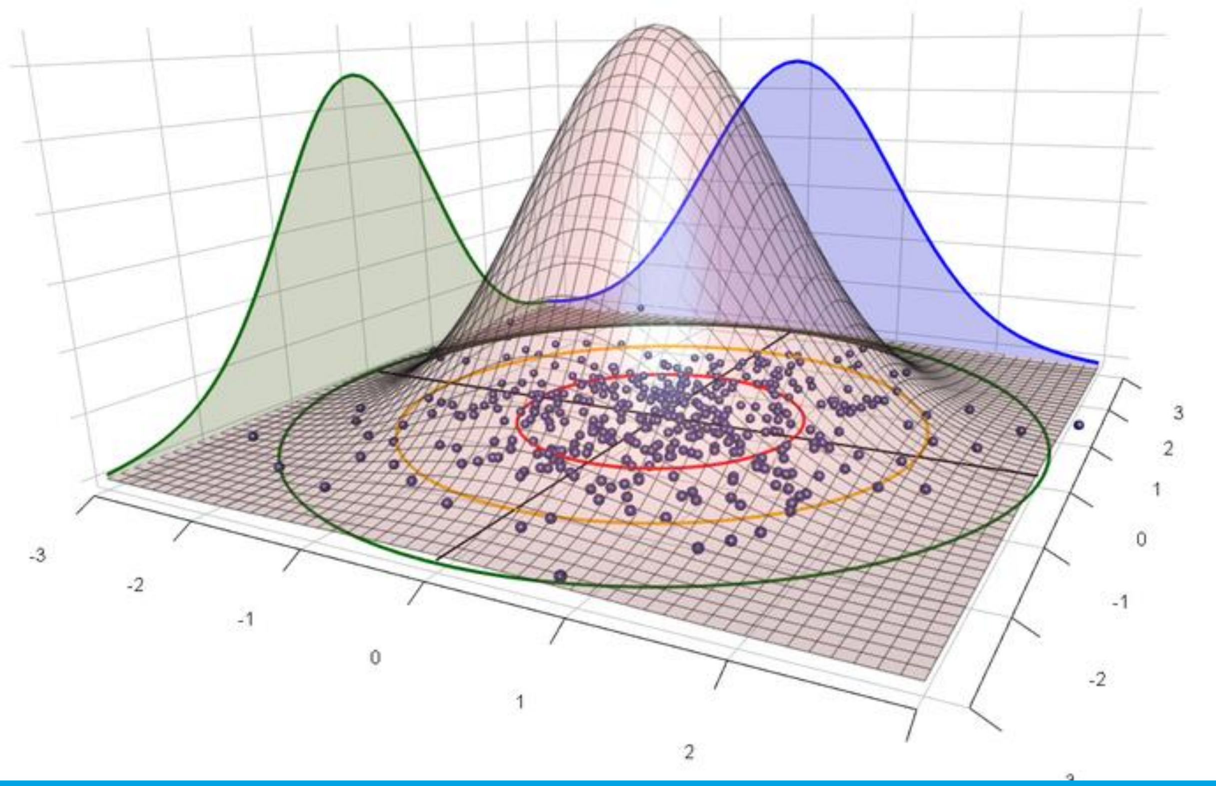
- Estimation vs. **prediction**
- Use-case
- Challenges (theoretical, technical)
- Domains and communities

Notation

- Probability density function (PDF), $R \to R$:

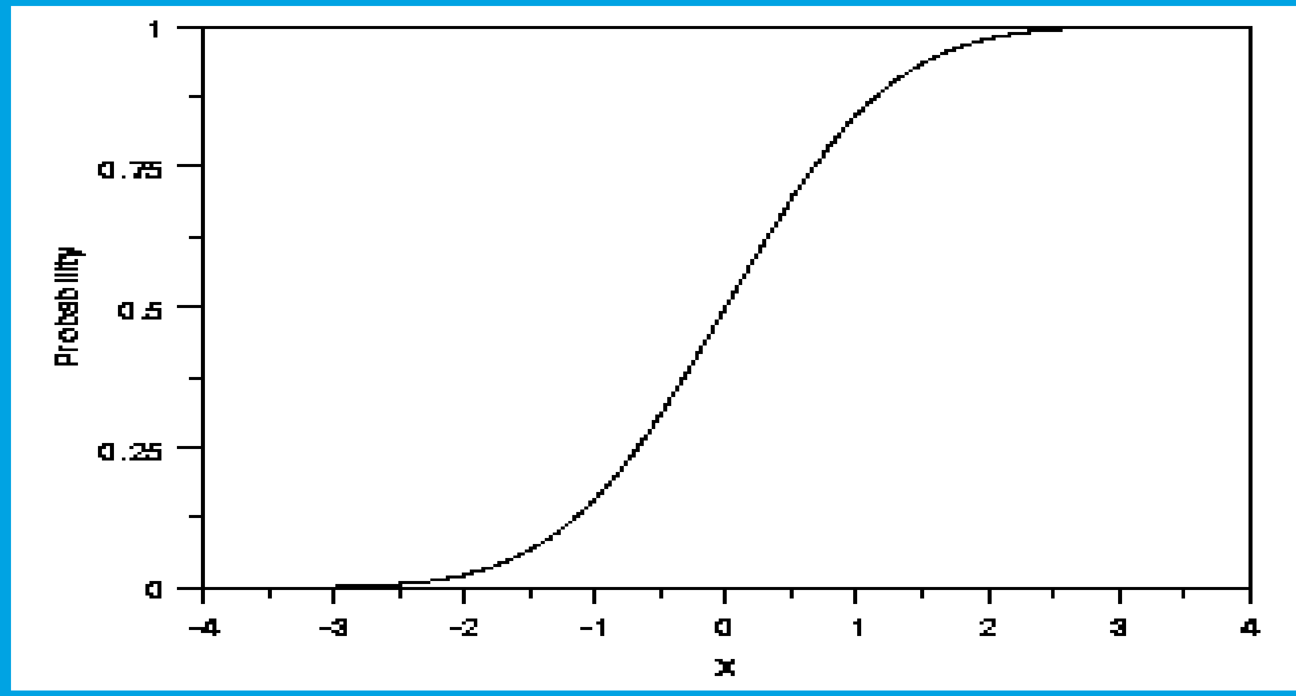


- Marginal PDF: $f(x)$

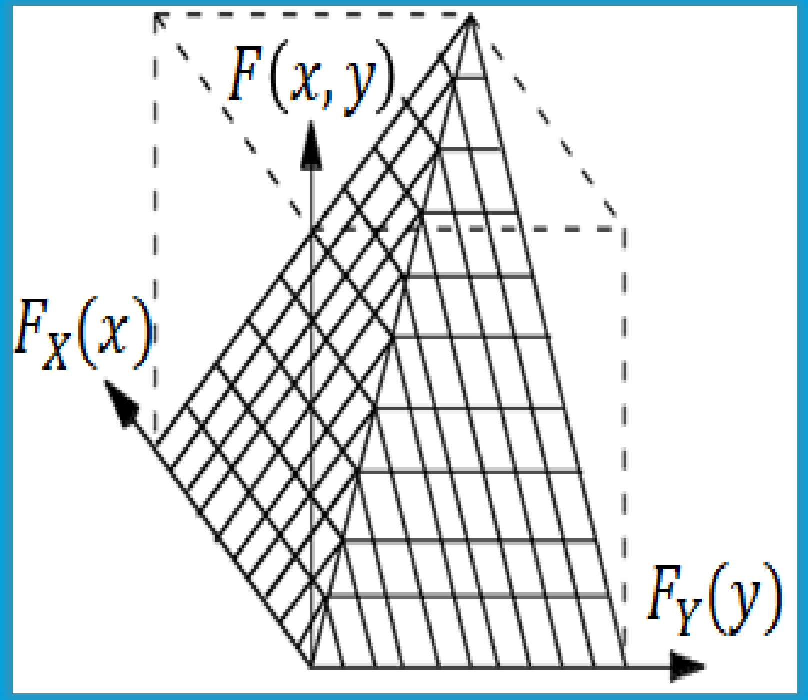


- Joint PDF: $f(x, y)$
- Conditional PDF: $f(x|\, Y = y)$

- Cumulative distribution function (CDF), $R \to [0,1]$:



- Marginal CDF: $F(x) = \int_{-\infty}^{x} f(x)dx$



- Joint CDF: $F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x, y)dxdy$
- Conditional CDF: $F(x|\, Y = y)$

- Sklar's theorem:

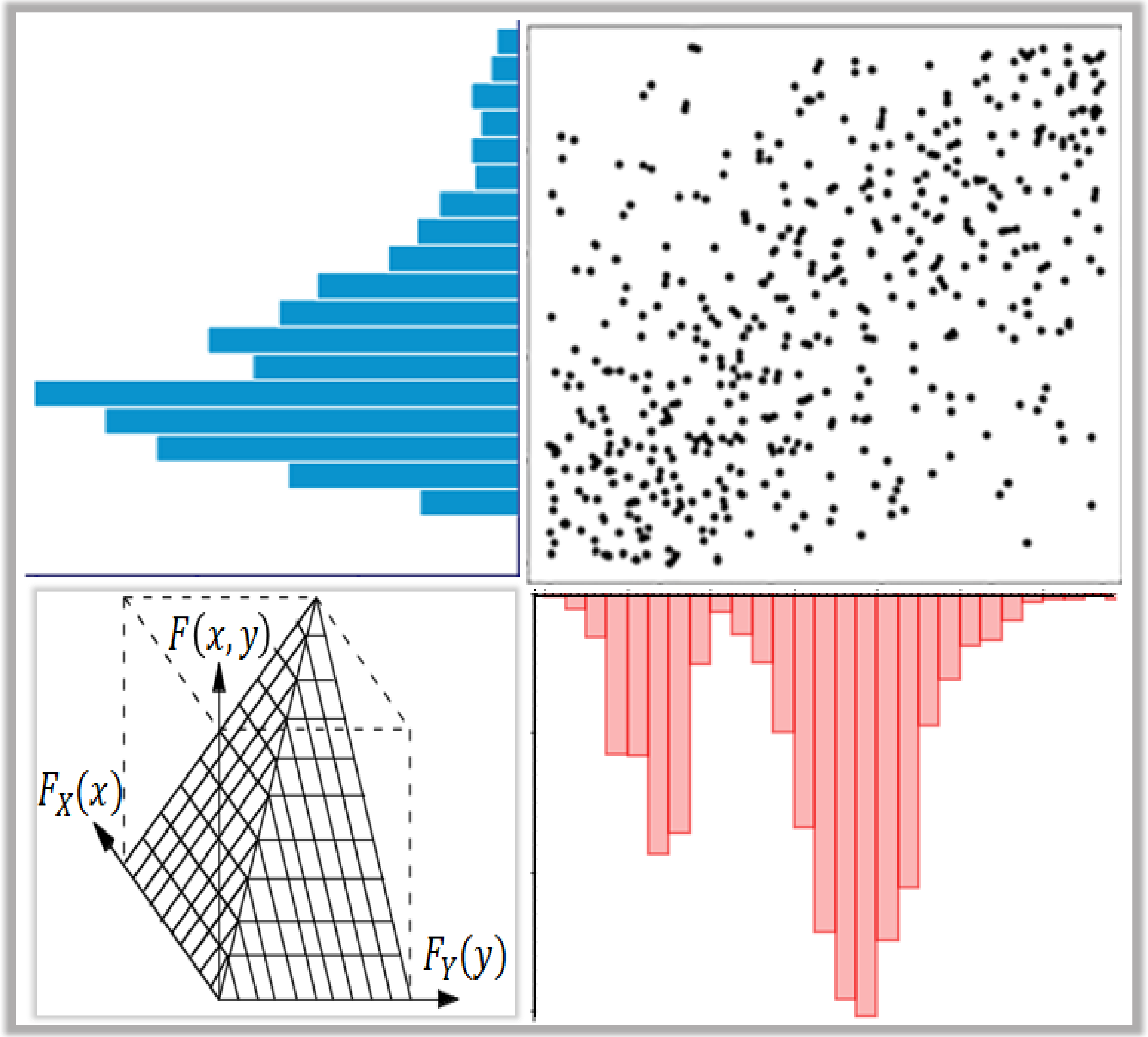
There is a function C that joins or ties these two variables: $F(x,y) = C\big(F_X(x), F_Y(y)\big)$

In the multivariate case: $F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n))$

Independent variables:
$f(x,y) = f_X(x) \times f_Y(y)$

Correlated variables:
$f(x,y) = c\big(F_X(x), F_Y(y)\big) \times f_X(x) \times f_Y(y)$

The function c can exhibit several types of non-linear negative or positive dependences.



What family does a copula belong to?

Copulas can be characterized using any distribution family that can be different from the marginal family of the involved variables.

1 = Gaussian copula
2 = Student t copula (t-copula)
3 = Clayton copula
4 = Gumbel copula
5 = Frank copula
6 = Joe copula
7 = BB1 copula
8 = BB6 copula
9 = BB7 copula
10 = BB8 copula

27 = rotated BB1 copula (90 degrees)
28 = rotated BB6 copula (90 degrees)
29 = rotated BB7 copula (90 degrees)
30 = rotated BB8 copula (90 degrees)
33 = rotated Clayton copula (270 degrees)
34 = rotated Gumbel copula (270 degrees)
36 = rotated Joe copula (270 degrees)
37 = rotated BB1 copula (270 degrees)
38 = rotated BB6 copula (270 degrees)
39 = rotated BB7 copula (270 degrees)
40 = rotated BB8 copula (270 degrees)

13 = rotated Clayton copula (180 degrees; "survival Clayton")
14 = rotated Gumbel copula (180 degrees; "survival Gumbel")
16 = rotated Joe copula (180 degrees; "survival Joe")
17 = rotated BB1 copula (180 degrees; "survival BB1")
18 = rotated BB6 copula (180 degrees; "survival BB6")
19 = rotated BB7 copula (180 degrees; "survival BB7")
20 = rotated BB8 copula (180 degrees; "survival BB8")
23 = rotated Clayton copula (90 degrees)
24 = rotated Gumbel copula (90 degrees)
26 = rotated Joe copula (90 degrees)

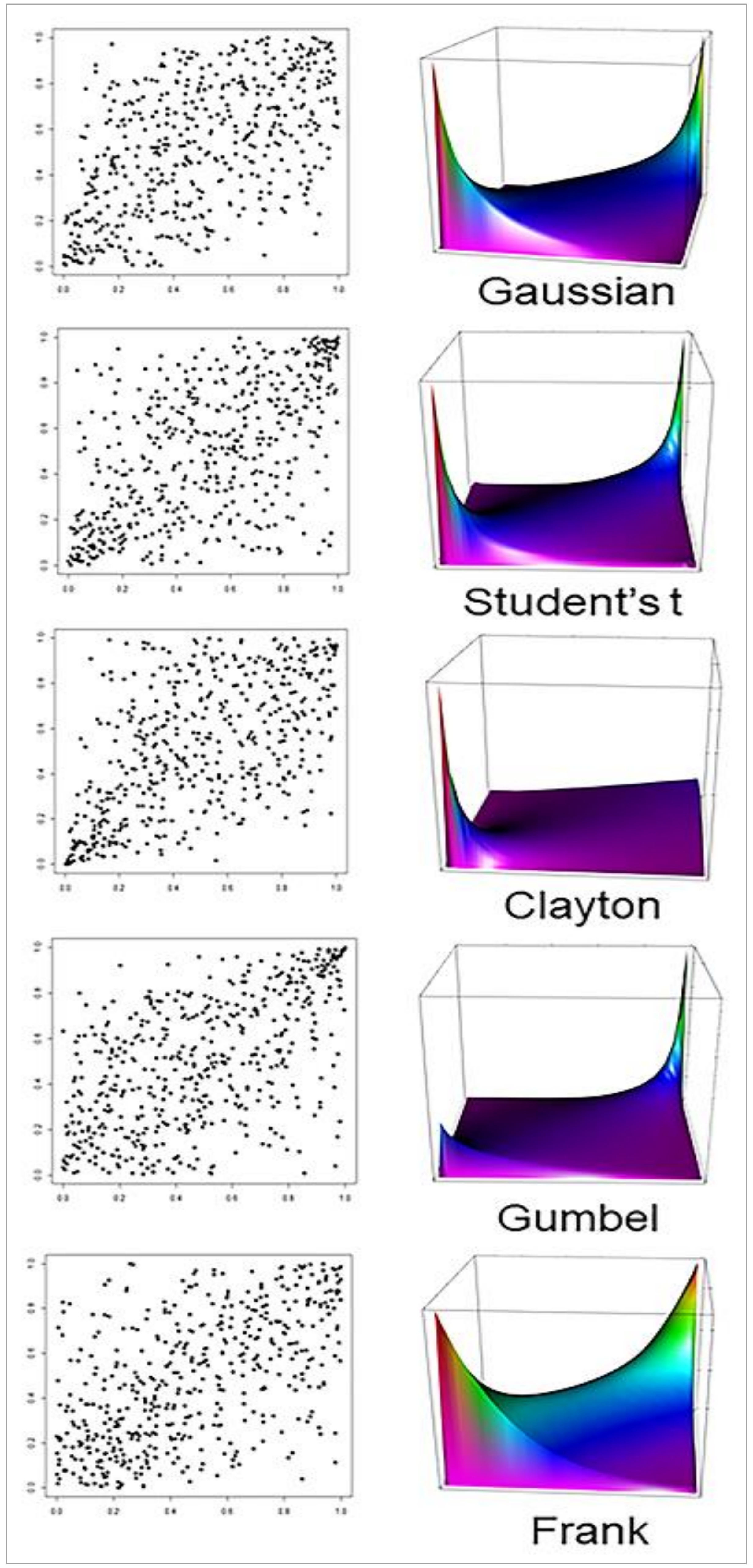
104 = Tawn type 1 copula
114 = rotated Tawn type 1 copula (180 degrees)
124 = rotated Tawn type 1 copula (90 degrees)
134 = rotated Tawn type 1 copula (270 degrees)
204 = Tawn type 2 copula
214 = rotated Tawn type 2 copula (180 degrees)
224 = rotated Tawn type 2 copula (90 degrees)
234 = rotated Tawn type 2 copula (270 degrees)

Bivariate copulas with one parameter $c_\theta(u,v)$:

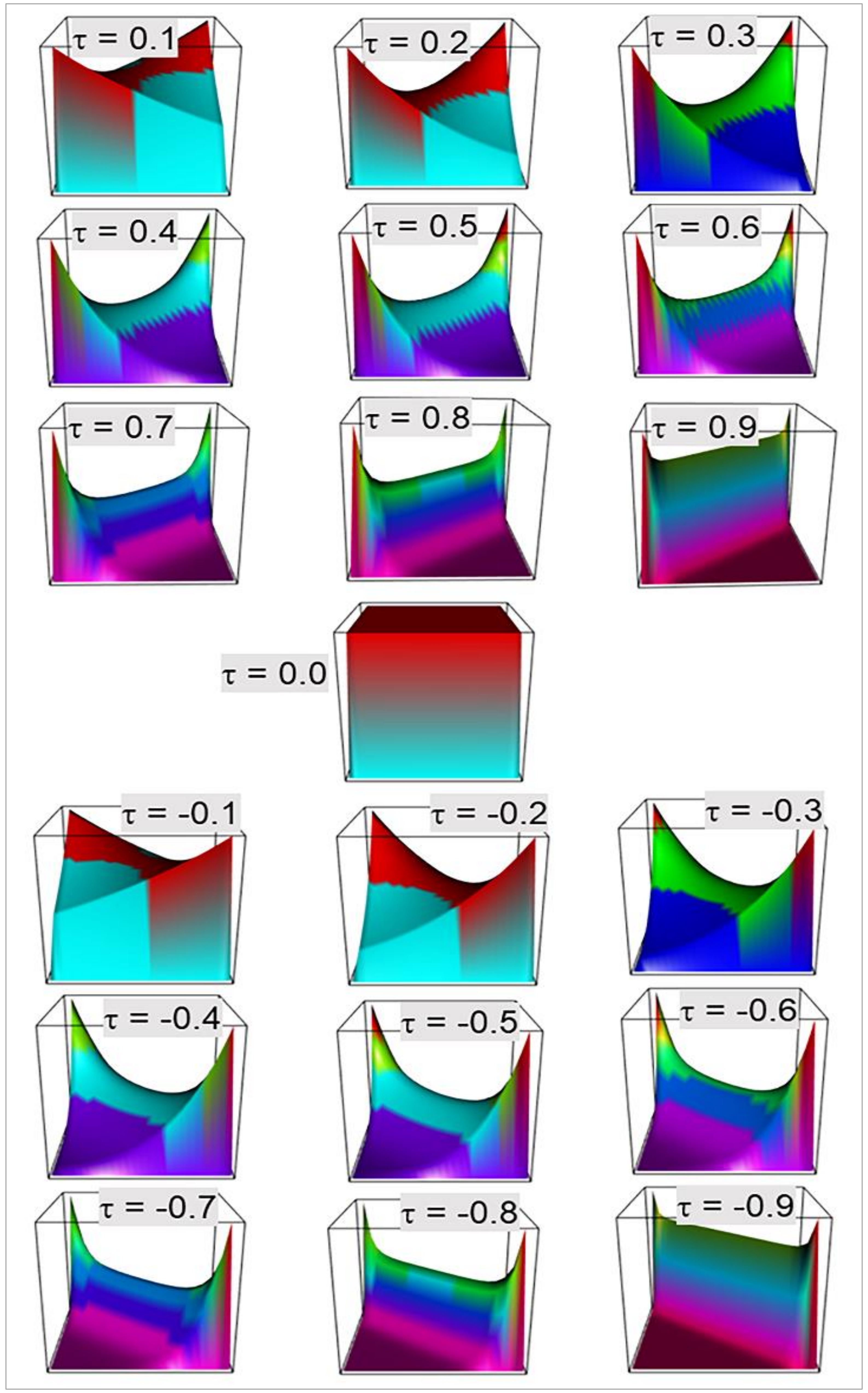
The parameter of copula is related to the Kendall's τ.

Five families of copulas are presented for several dependence structures between two variables while the Kendall's τ is equal to 0.4 in all dependences.

The horizontal axes are u and v and the third axes denote the density values.



The densities of Frank copula for several values of the Kendall's τ.



- Copula estimation (bivariate case):

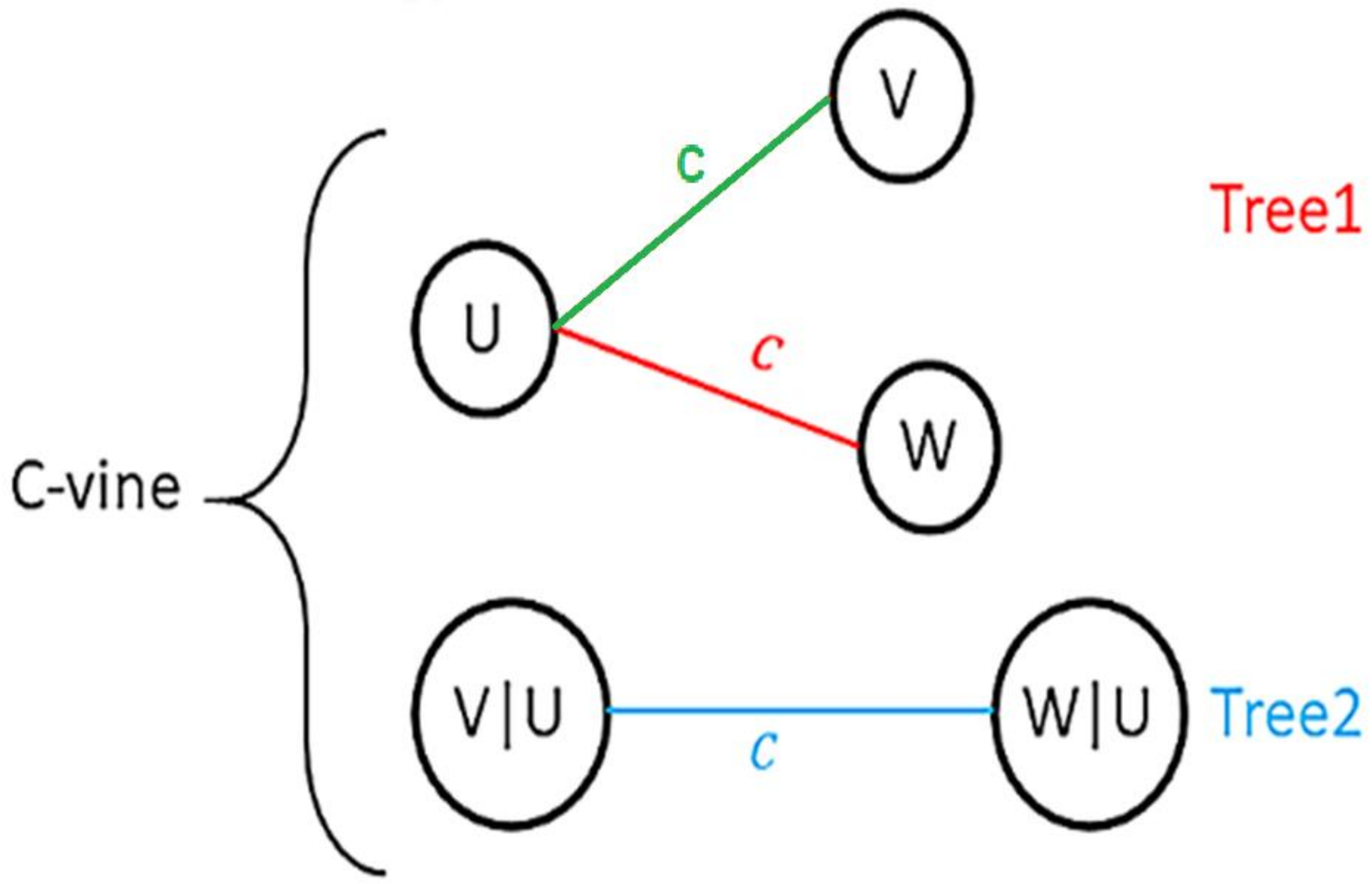
We apply maximum likelihood to estimate the parameter for each family using starting values obtained by Kendall's τ, being a measure for association between variables (Nelsen, 2006).

The most suitable family is selected according to Akaike's Information Criteria (AIC) (Akaike, 1974).

Let's consider three random variables *X*, *Y* and *Z* with a copula *C(U,V, W)*, where *X* is the target variable. The configuration of the structure is based upon c-vine, Sklar's theorem and the general decomposition rule of $f(x, y, z)=f(z)\times f(y|z)\times f(x|y,z)$.

In this example, the copula density c(U,V, W) is first decomposed into bivariate copulas as: c(U,W), c(U,V) and c(C(W|U),C(V|U)). Then, the copula density is the product of all bivariate copula densities in the structure:
$c(u,v, w)=c(u,w)\times c(u,v)\times c(C(W|U),C(V|U))$.

It follows that the dependence structure between those n=3 variables is described by a combination of n different families and in total $n\times(n-1)/2$ parameters.



The usefulness of copulas in real-world applications:

- Any joint distribution can be written in terms of a copula. This illustrates the growing interest its applications in diverse studies such as finance, image analysis, geostatistics, and in particular in the environmental sciences; hydrology, disasters, agriculture, weather and climate.

- The definition can be extended to higher dimensions including several random variables/fields: spatial dependences, temporal dependences, spatio-temporal dependences, and dependences between several variables at one point in time and space.

- The family distribution of C can be different from the family of F_X and F_Y. For example, both X and Y can follow Gaussian distributions, but C can be a non-Gaussian joint distribution.

A copula is a **cumulative joint distribution function** describing complex dependence structures.

1) Introduction to copulas:

- Notation (Cumulative distribution function / probability density function)
- Dependence and joint distribution
- Distribution family
- **Estimation** vs. prediction
- Multivariate joint distribution: The world of vines

2) Application of copulas:

- Estimation vs. **prediction**
- Use-case
- Challenges (theoretical, technical)
- Domains and communities

- The conditional probability:

$$\hat{x}_p = F^{-1}(p|.), \qquad p \in [0,1]$$

- The conditional expectation:

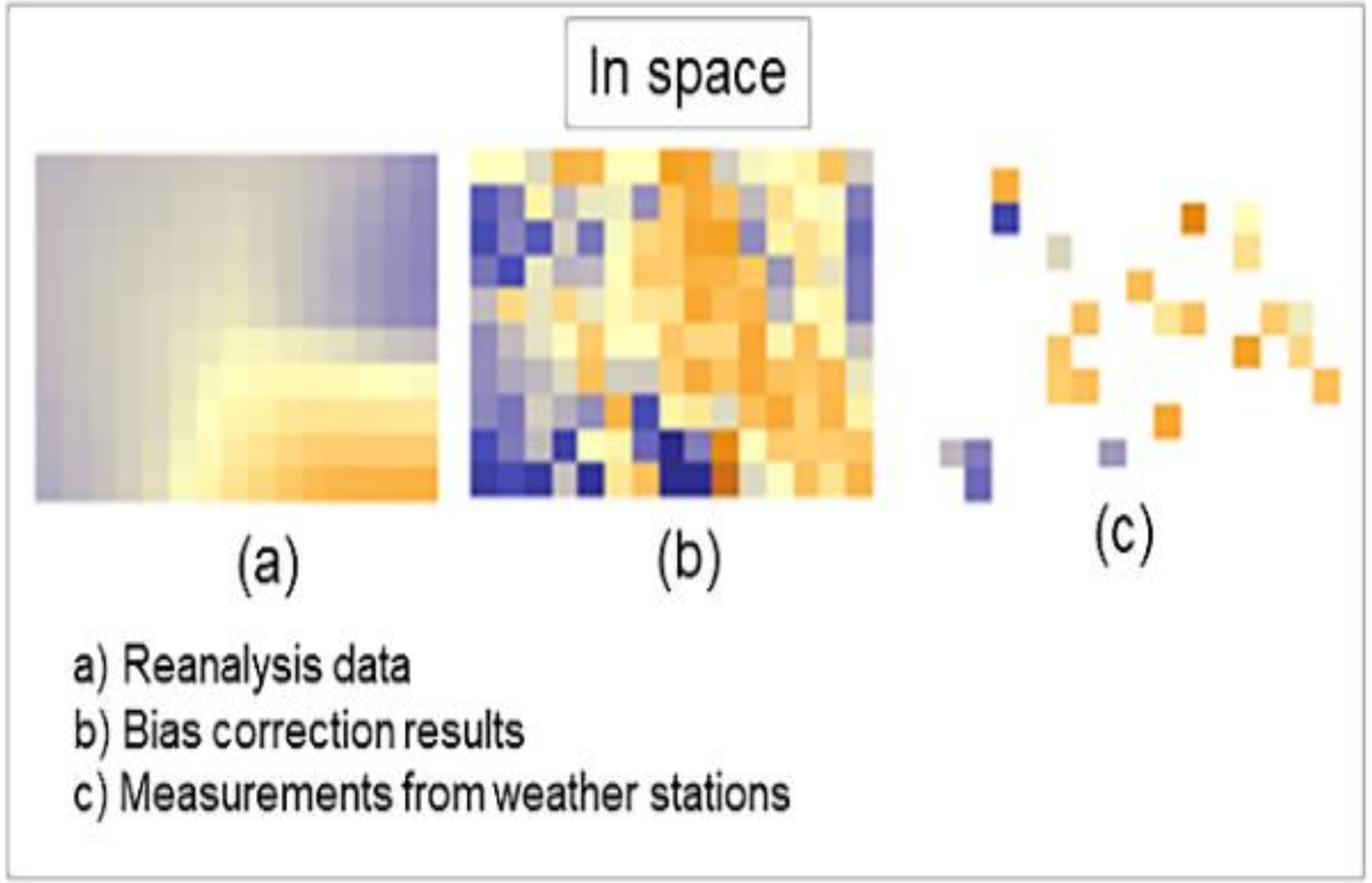
$$\hat{x}_{mean} = E[X|.] = \int_x x \cdot f(x|.)dx$$

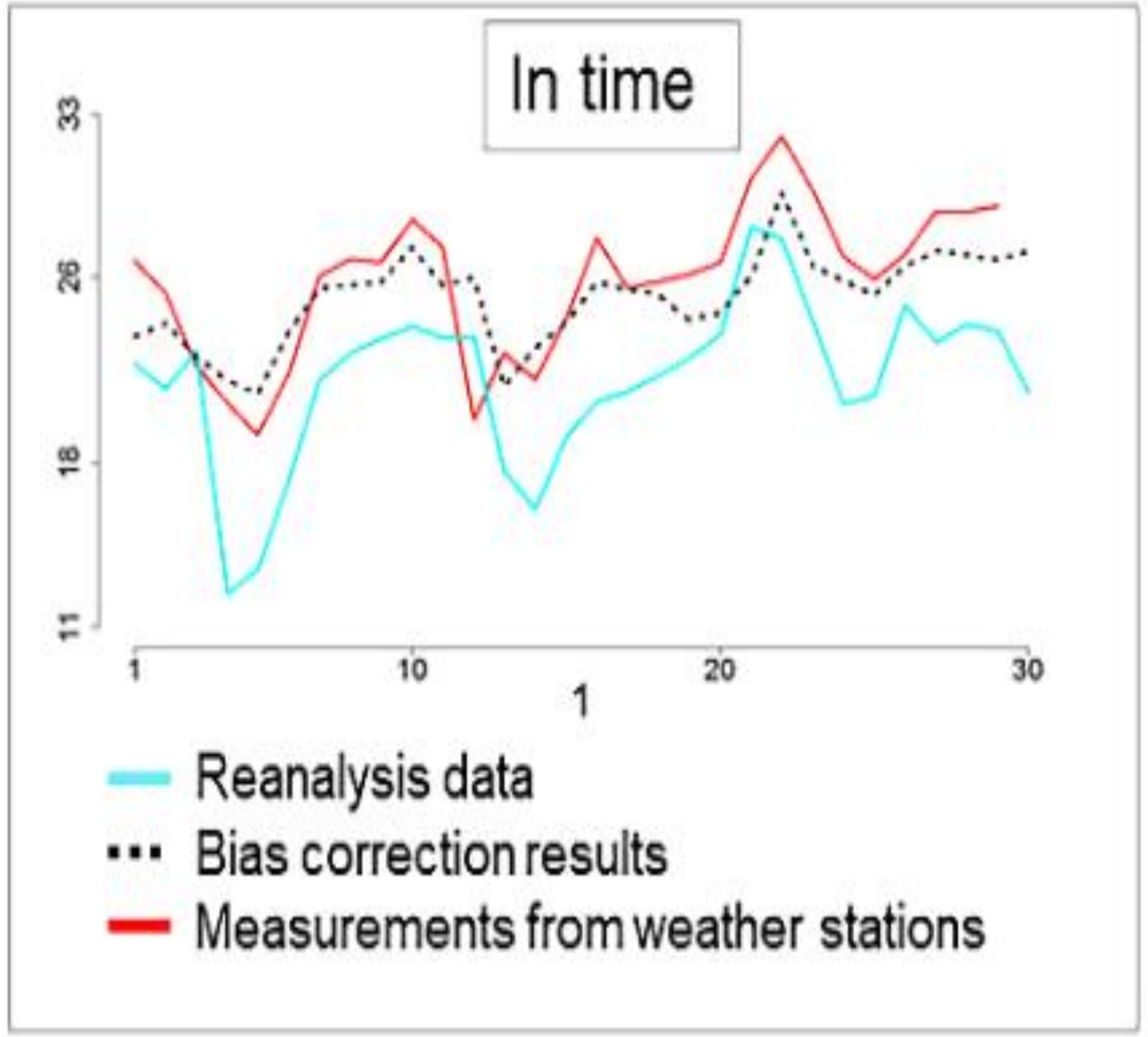
- The conditional median:

$$\hat{x}_{median} = F^{-1}(0.5|.)$$

A multivariate copula quantile mapping (MCQM) for bias correction

x=Air temperature measurement from weather stations, $u = F_X(x)$
y=Reanalysis ECMWF data, $v = F_Y(y)$
z= elevation, $w = F_Z(z)$





Question:
how to find $\hat{x}_i = F_X^{-1}(\hat{u}_i)$

Solution:
using MCQM, we have $\hat{u}_i = C^{-1}(C(v_i|W = w_i)|W = w_i)$.

A multivariate copula-based interpolator for downscaling

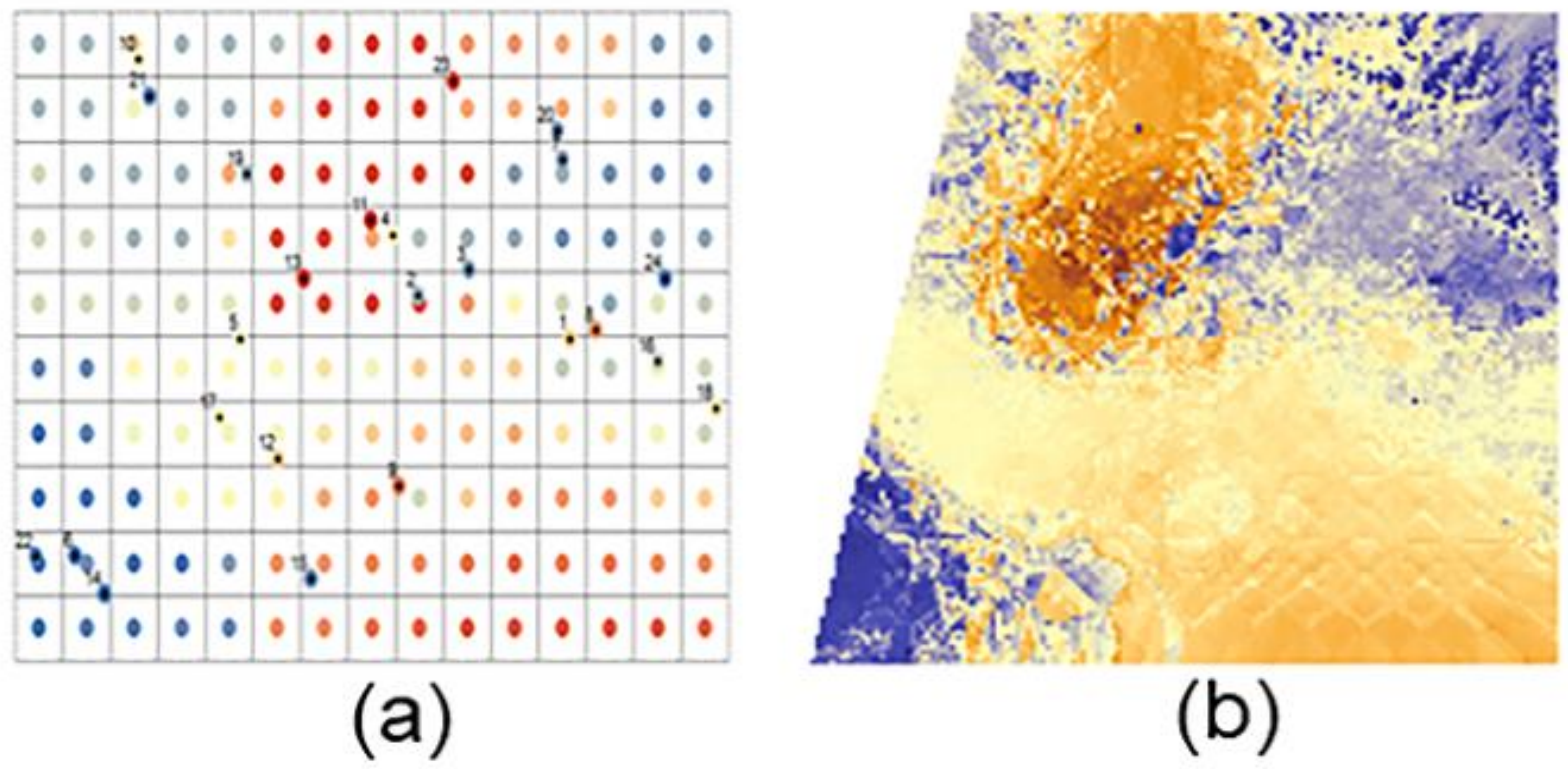
Question:

How to find $\hat{x}_0 = E[X|X = x_1, \dots, X = x_n] = \int_x x \cdot f(X|X = x_1, \dots, X = x_n)dx$

Solution

Copula based interpolator: $\hat{x}_0 = \int_0^1 F^{-1}(u) \cdot c(U|U = u_1, \dots, U = u_n)du$

where $f(X|X = x_1, \dots, X = x_n)$ is the conditional density distribution of the variable X at an unvisited location conditioned on its n nearest neighbours.



a) Observations
b) Interpolation results

Methods in geo statistic:

- Integration
- Spatio-temporal modeling
- Interpolation
- Bias correction
- Downscaling
- Data assimilation

Application in geo statistic:

- Groundwater quality parameters
- Predicting flood
- Trajectories and Sensor Observation Service
- Daily mean air temperature
- Evaluating the effects of climate extremes on crop variables

- Numerical evaluations concern the implementations of some mathematical/statistical operations for copula families, such as partial derivatives and inverse transformations.
- A d-dimensional joint probability, $d > 2$, is obtained using the numerical evaluations and simulations that are associated with uncertainty.
- For a deterministic approach, the main theories of copulas that are based upon probabilistic explanations need to be extended.

- Kjersti Aas
  The Norwegian Computing Centre
- Eike Brechmann
  Center for Mathematical Sciences, Technische Universität München
- Claudia Czado
  Center for Mathematical Sciences, Technische Universität München
- Benedikt Gräler
  52°North Initiative for Geospatial Open Source Software GmbH, Muenster, Germany
- Claus P. Haslauer
  University of Tuebingen, EKU Tübingen , Department of Geosciences

References

- Alidoost, F., (2019), *Copulas for integrating weather and land information in space and time* (Doctoral). University of Twente.
- Gräler, B., (2014), *Developing spatio-temporal copulas* (Doctoral). Westfälische Wilhelms-Universität Münster.
- Aas, K., et al. (2009). Pair-copula constructions of multiple dependence. Insurance: Mathematics and Economics, 44(2), 182–198.
- Nelsen, R. B. (2006), *An Introduction to Copulas.* United States of America: Springer.
- Genest, C. and Favre, A.C. (2007). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. Journal of Hydrologic Engineering, 12(4), 347-368.
- https://github.com/cran/copula
- https://github.com/topics/copula

Let’s stay in touch

f.alidoost@esciencecenter.nl

https://www.esciencecenter.nl/profile/sarah

https://nl.linkedin.com/in/fakhereh-sarah-alidoost-11a24a89

netherlands eScience center