Catchment Characterization based on Runoff Copulas

Luis Samaniego (1), **András Bárdossy** (2)

(1) Helmholtz-Centre for Environmental Research - UFZ
(2) Institute of Hydraulic Engineering, University of Stuttgart

EGU General Assembly
Vienna, 19th April 2007

Outline

1. Introduction

Main objective:

Estimate time series characteristics using a set of catchment descriptors, **specially** where there is no discharge data (PUB)

Methods: Unsupervised/supervised clustering

1. Introduction

Main objective:

Estimate time series characteristics using a set of catchment descriptors, **specially** where there is no discharge data (PUB)

Methods: Unsupervised/supervised clustering

- ANN, k-NN, fuzzy-rules, spectral signatures...

1. Introduction

Main objective:

Estimate time series characteristics using a set of catchment descriptors, **specially** where there is no discharge data (PUB)

Methods: Unsupervised/supervised clustering

- ANN, k-NN, fuzzy-rules, spectral signatures...
- Generalized k-Nearest Neighbors + copulas

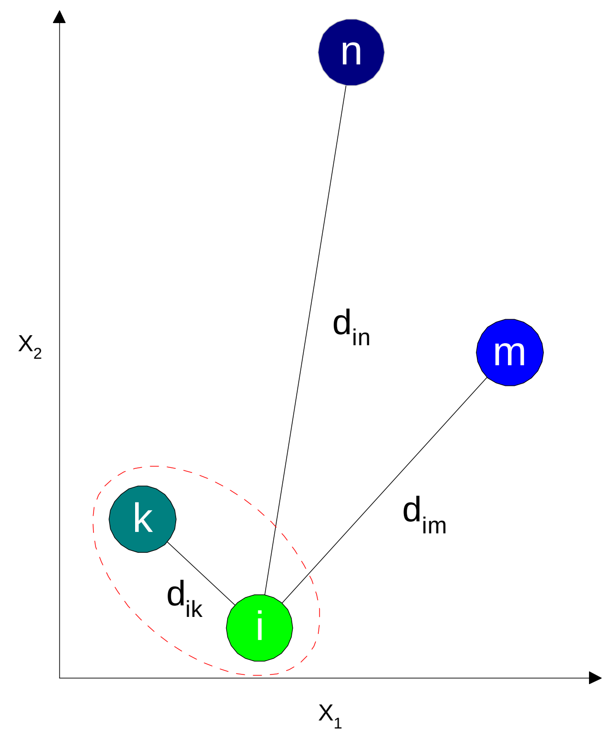
HELMHOLTZ
CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

2. Traditional Catchment Characterization Procedure

- **Data**:
  $\mathcal{D} = \{(\mathbf{x}_i, y_i) \quad i = 1, \ldots n\}$
  - Outputs: y $\longrightarrow$ Highly dynamic
  - Inputs: $\mathbf{x}$ $\longrightarrow$ Static or slightly dynamic

- **Similarity** $\Leftrightarrow$ **Metric**:
  Selected *a priori*, e.g. Euclidian distance

- **Results**:
  - Unsupervised approaches highly uncertain
  - Characterization is not unique
    $\sim$ runoff characteristic(s) y e.g. $\bar{q}$, q_5
    $\sim$ metric



HELMHOLTZ
CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

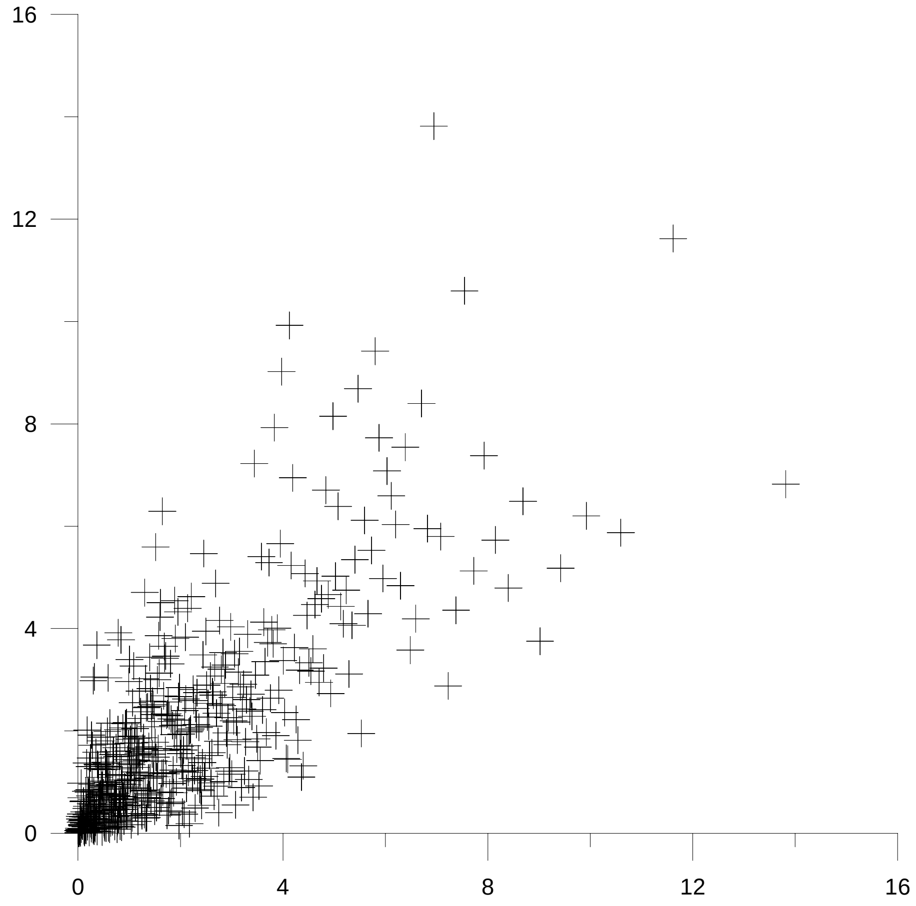
What is a Copula?

“In statistics, a **copula** is a multivariate distribution function on the unit hypercube with **uniform** marginals.”
Bárdossy, 2006, WRR

⇒ “Describes the dependence structure between random variables without information on the marginals.”

$$F(x_1, \ldots, x_n) = C\Big[F_1\big(x_1\big), \ldots, F_n\big(x_n\big)\Big]$$

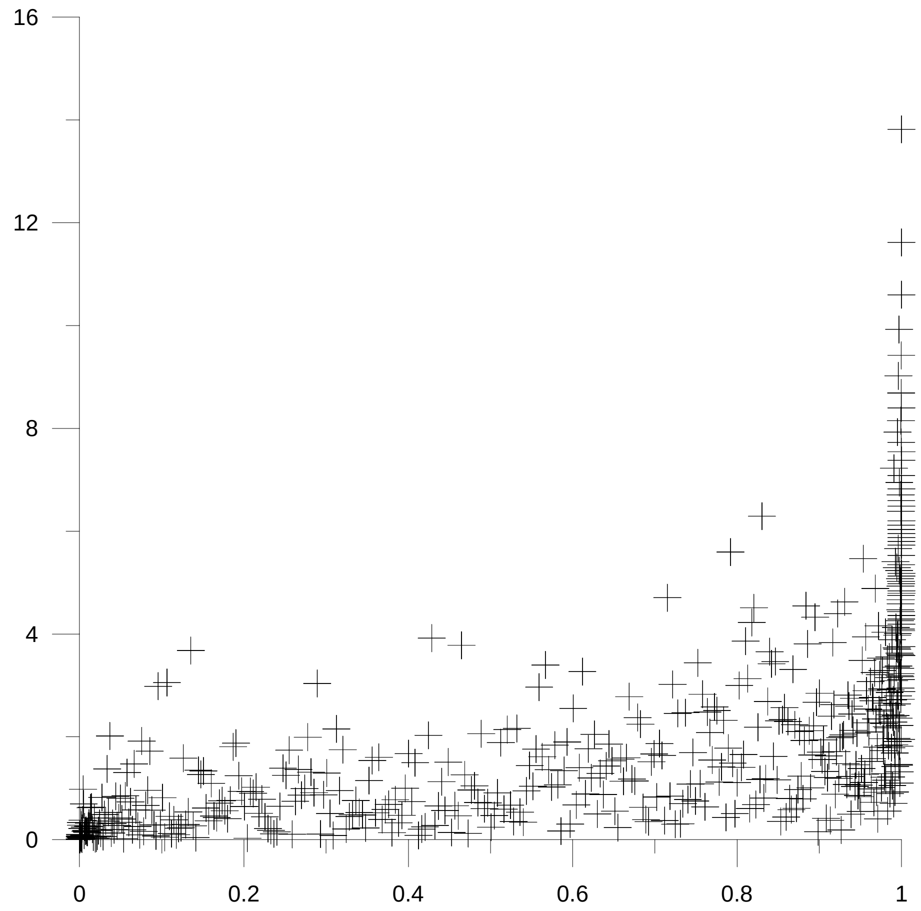
Effect of the Marginal Distributions



x_1 vs. x_2, marginals: F, G

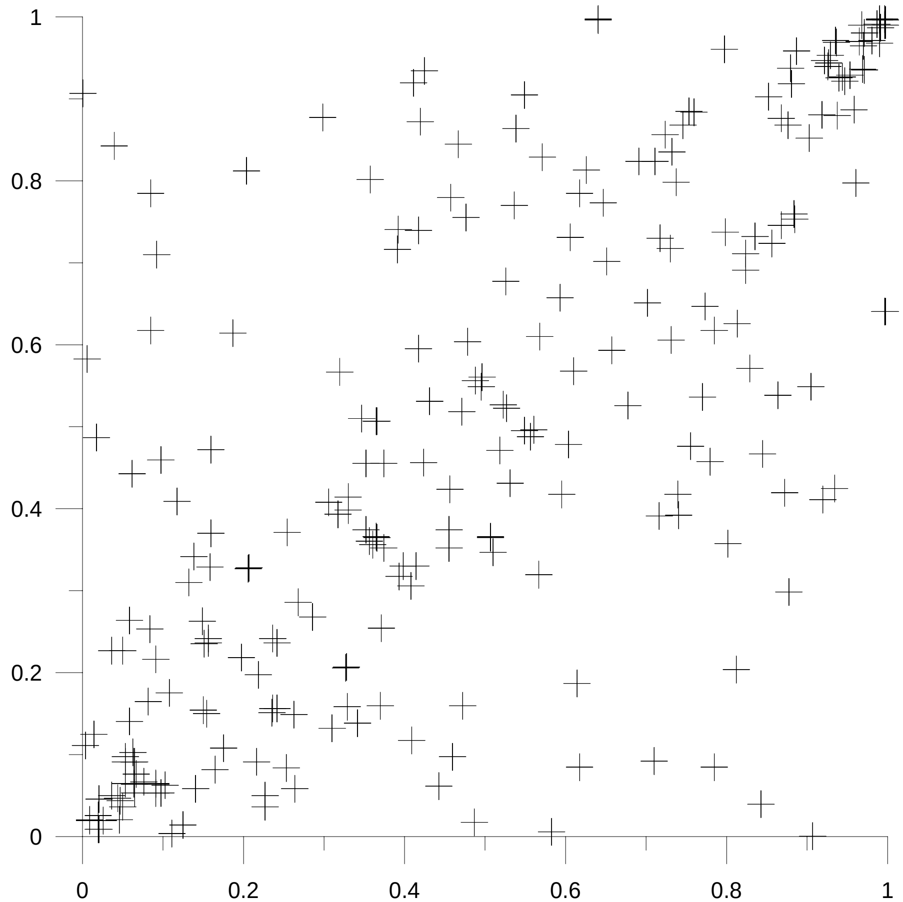
HELMHOLTZ
CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

Effect of the Marginal Distributions



x_1 vs. x_2, marginals: H, G

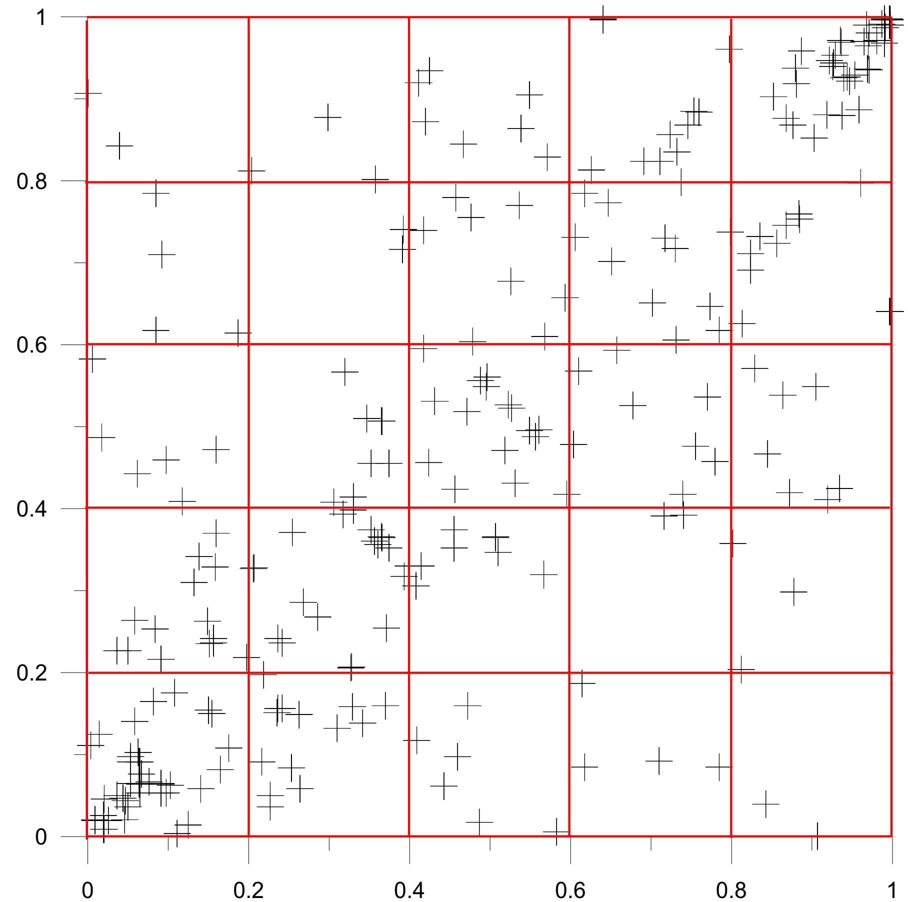
Effect of the Marginal Distributions



Empirical copula of x_1 vs. x_2

HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

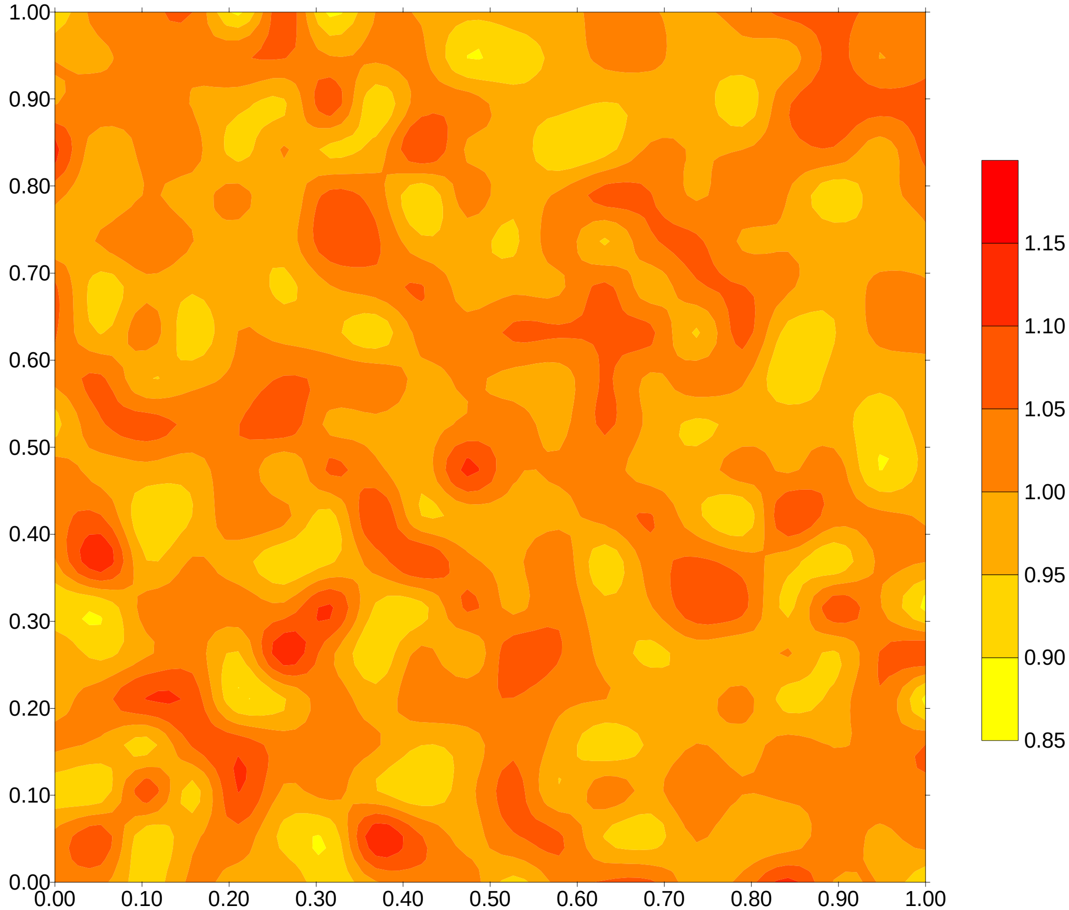
Effect of the Marginal Distributions



Empirical copula of x_1 vs. x_2

HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

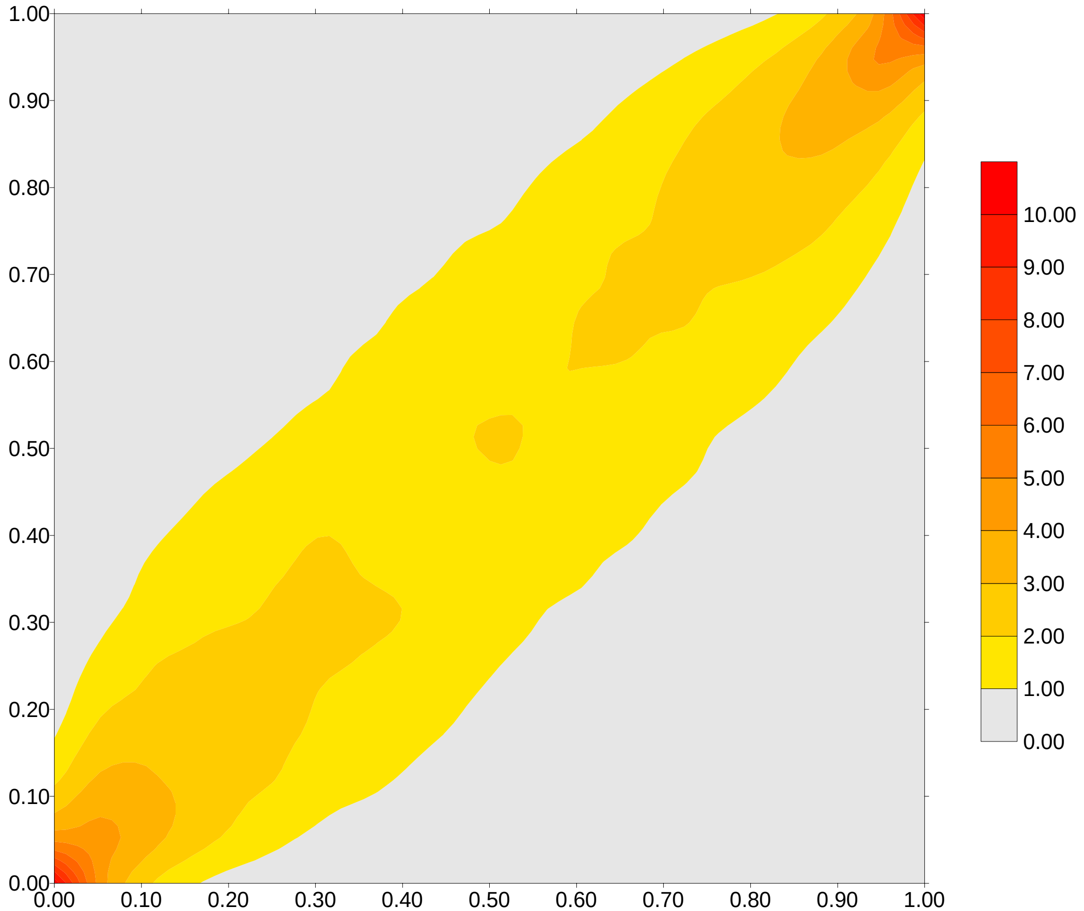
Examples of Copula Density



$x_1, x_2 \sim \mathrm{unf}$ Random variables

HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

Examples of Copula Density



$x_1, x_2 \sim \mathrm{N}(0, \sigma)$, r=0.8

HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

Empirical Runoff Copula

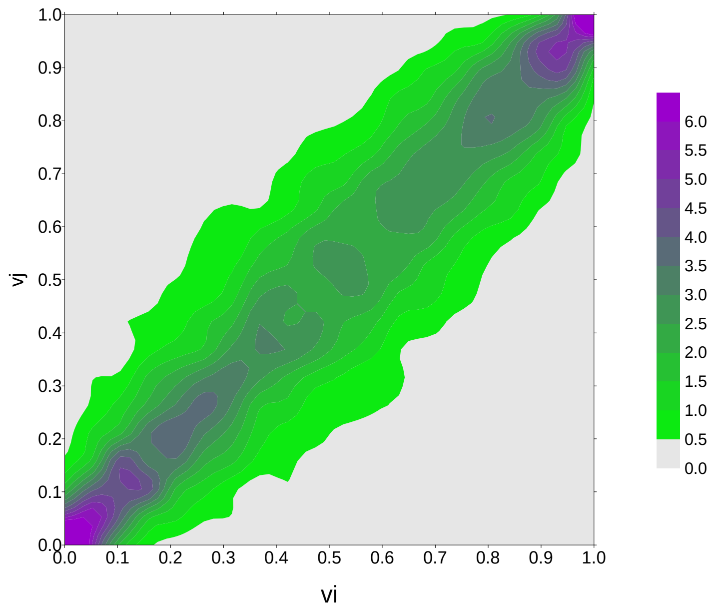
Given: $q_i^t, q_j^t \equiv$ daily time series (1908-2000)

$$\begin{aligned} C(v_i, v_j) &= P\Big[F_i(q_i) < v_i; F_j(q_j) < v_j\Big] \\ &= C\Big(F_i(q_i), F_j(q_j)\Big) \end{aligned}$$

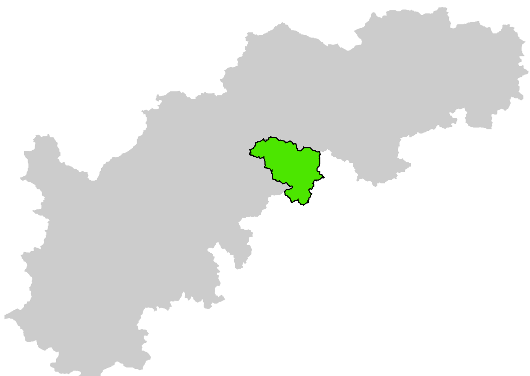
Empirical Runoff Copula

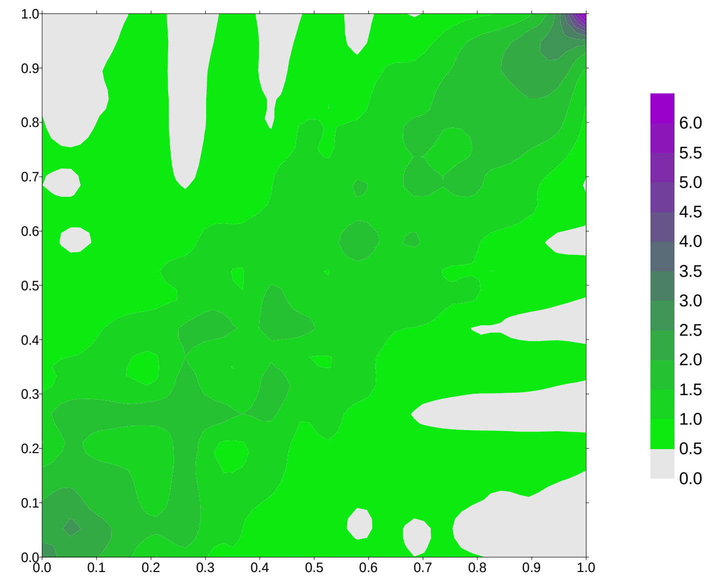
Given: $q_i^t, q_j^t \equiv$ daily time series (1908-2000)

$$
\begin{aligned}
C(v_i, v_j) &= P\Big[F_i(q_i) < v_i; F_j(q_j) < v_j\Big] \\
&= C\Big(F_i(q_i), F_j(q_j)\Big) \\
c(v_i, v_j) &= \frac{\partial^2 C(v_i, v_j)}{\partial v_i \partial v_j}
\end{aligned}
$$

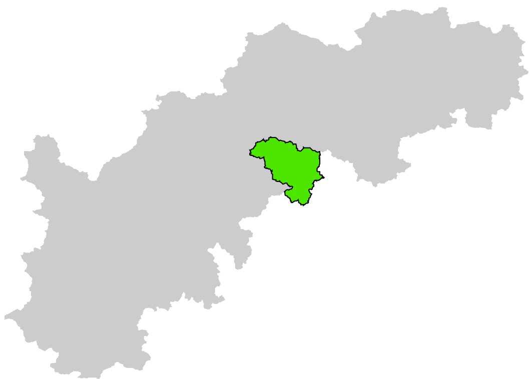


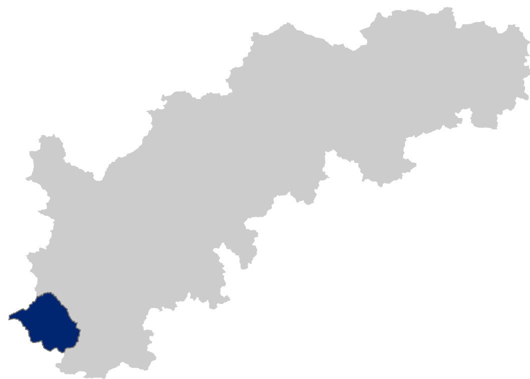
Distance Based on Runoff Copulas

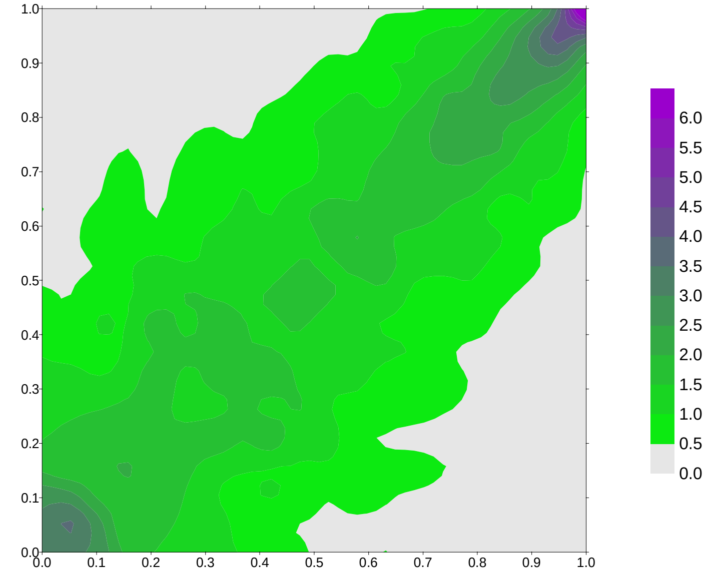




Distance Based on Runoff Copulas

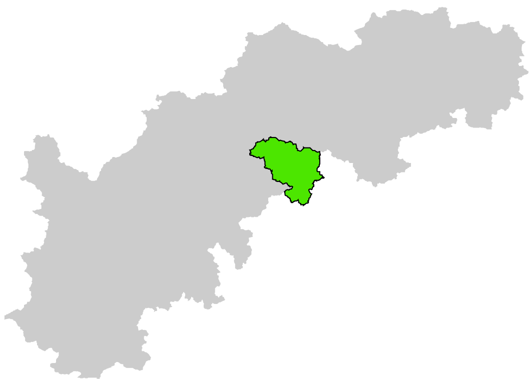


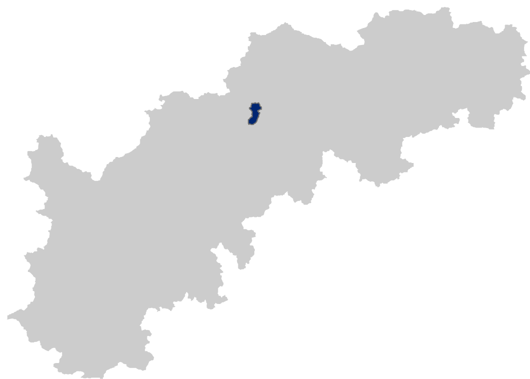


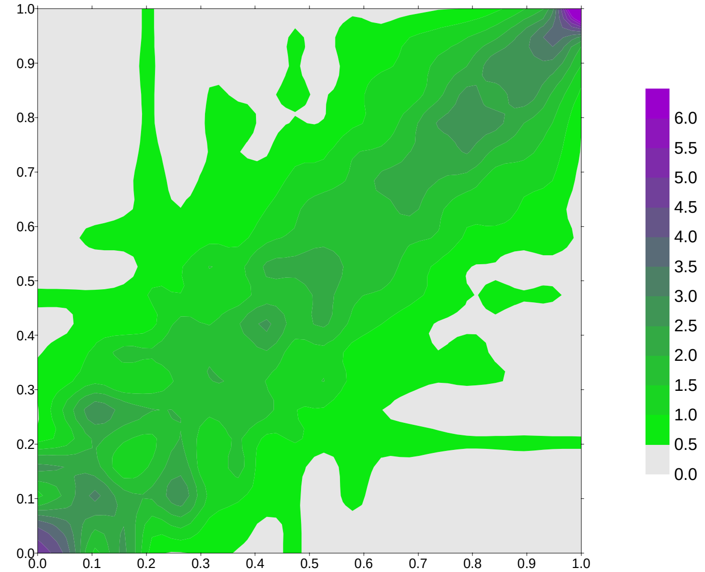


HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Distance Based on Runoff Copulas

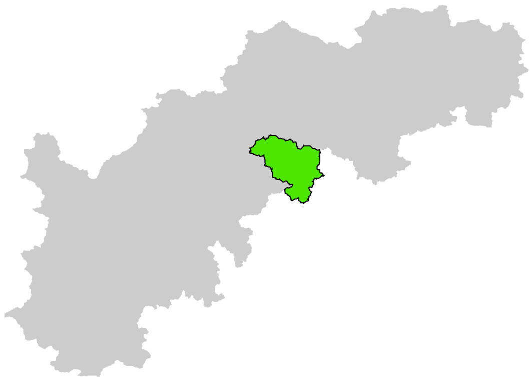


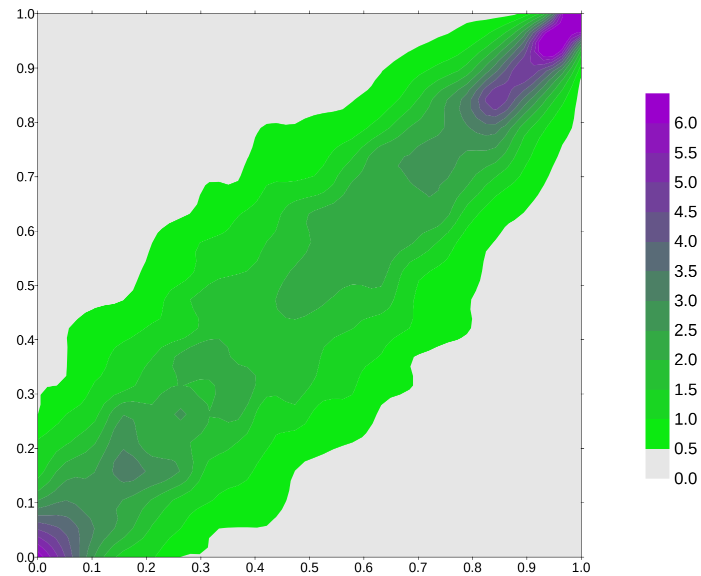




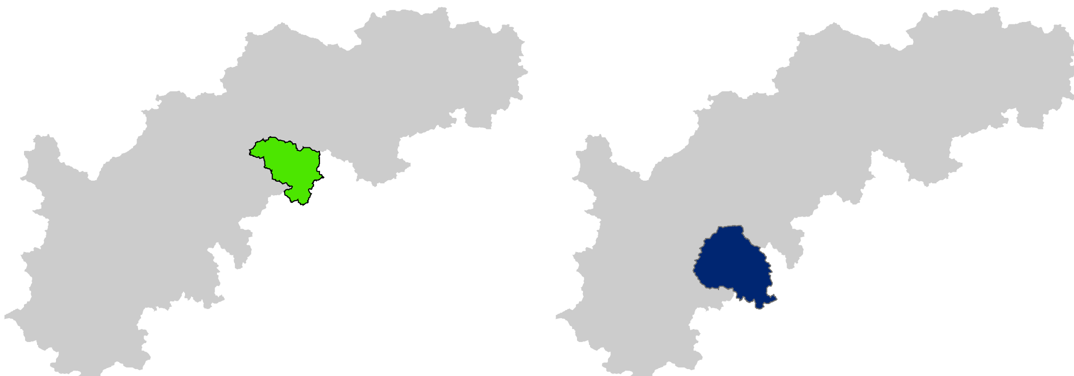
HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

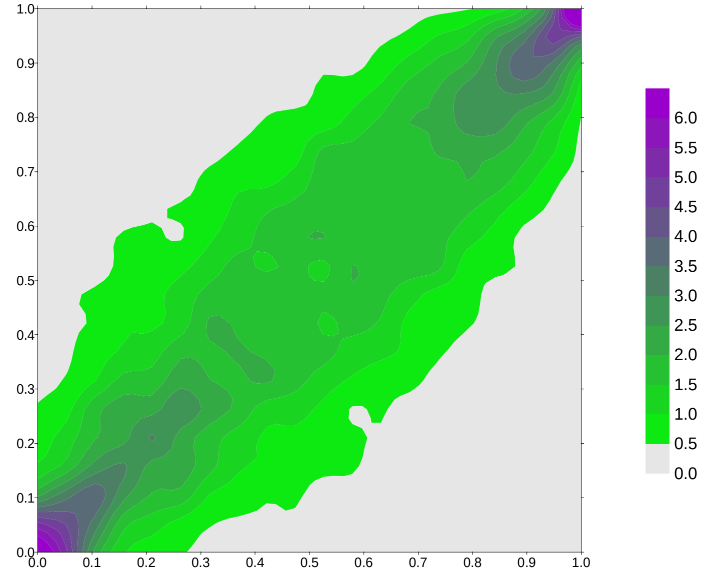
Distance Based on Runoff Copulas



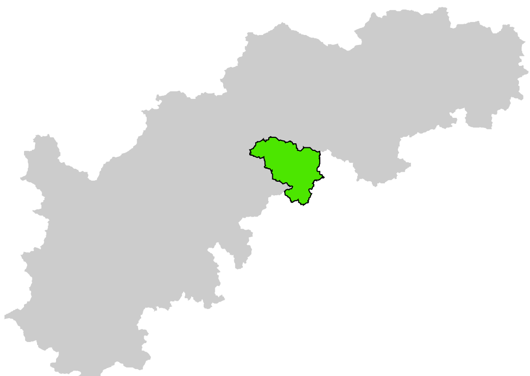


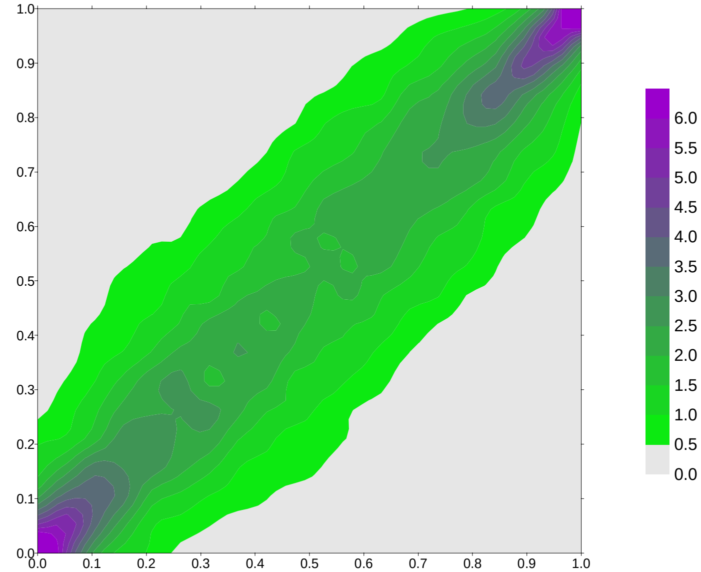
Distance Based on Runoff Copulas



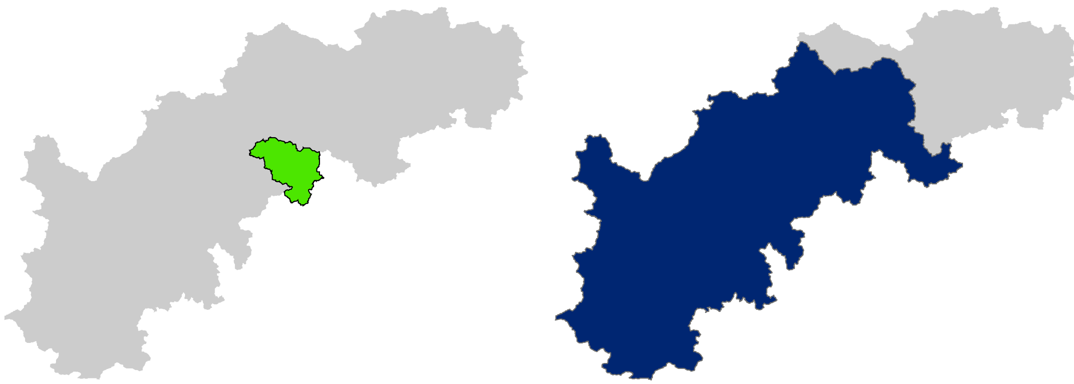


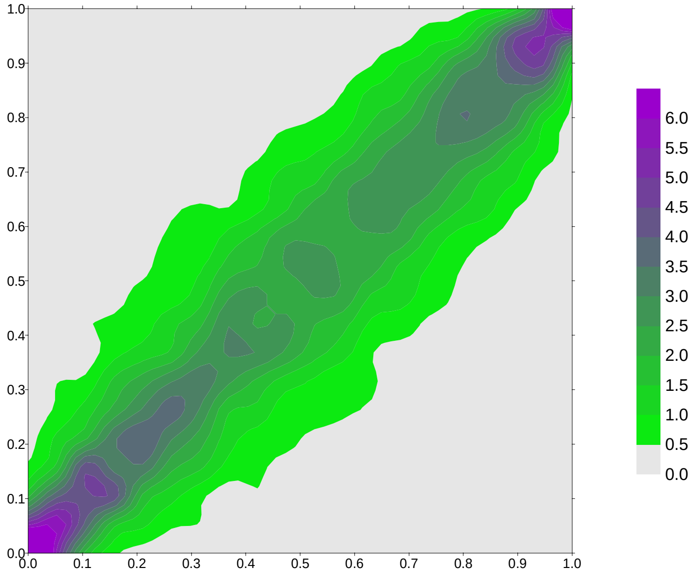
Distance Based on Runoff Copulas





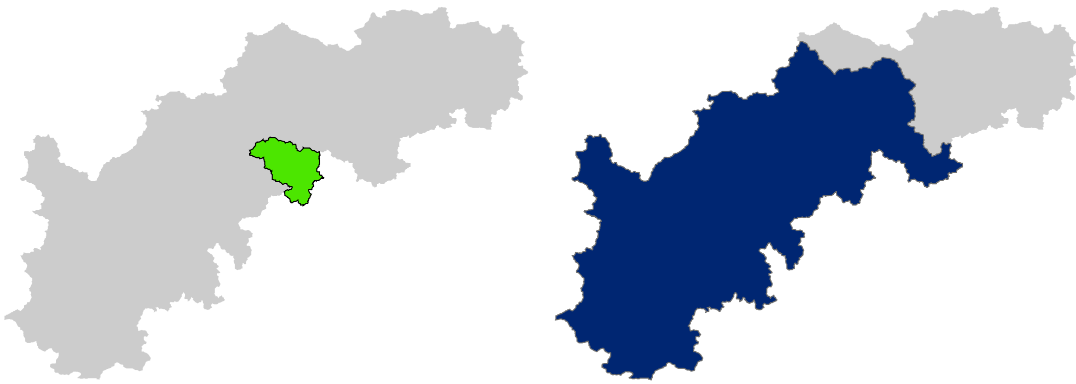
Distance Based on Runoff Copulas

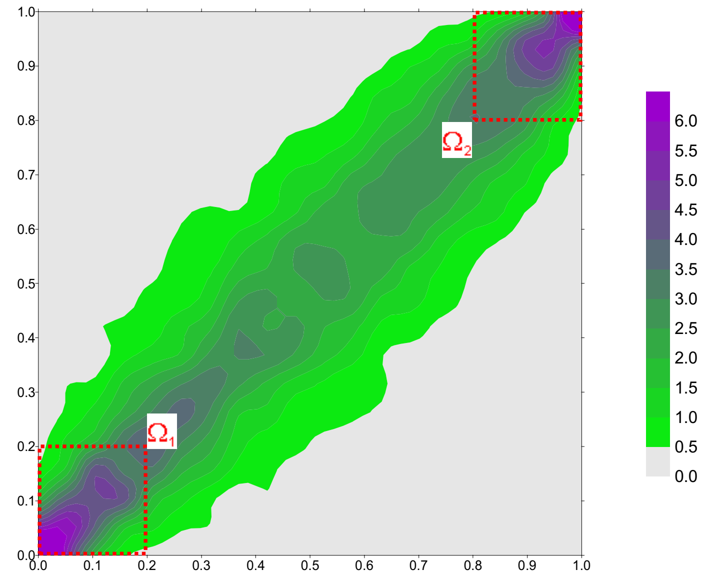




HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Distance Based on Runoff Copulas



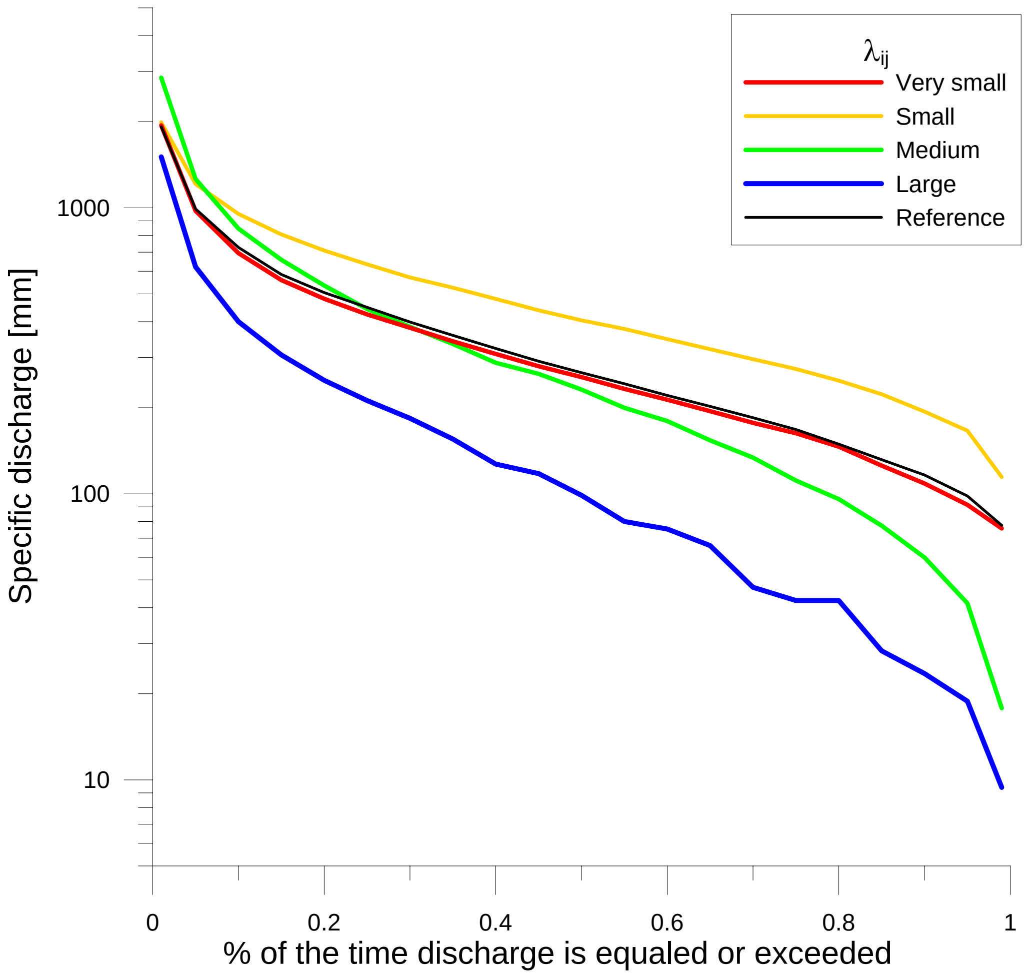


$$L_{ij} = \int\int_{\Omega_1} c(v_i, v_j) d\Omega$$

$$U_{ij} = \int\int_{\Omega_2} c(v_i, v_j) d\Omega$$

$$\lambda_{ij} = \frac{|U_{ij} - L_{ij}|}{U_{ij} + L_{ij}} + (p - L_{ij})$$

Distance λ and the FD Curves



HELMHOLTZ
CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

Proposed Characterization Method

Training $\longrightarrow$ $\mathcal{T} = \{\mathbf{x}_i, [\lambda_{ij}], y_i \quad \forall\, i, j \in n\}$

Embedding $\longrightarrow$ $\mathbf{u} = B[\mathbf{x}]$

Condition $\longrightarrow$ Small $d(\mathbf{u}_i, \mathbf{u}_j) \rightarrow$ Small λ_{ij}

$$d_{ij}^2 = (\mathbf{u}_i - \mathbf{u}_j)\mathbf{g}(\mathbf{u}_i - \mathbf{u}_j)^{\mathrm{T}}$$

Validation $\longrightarrow$ $\mathcal{V} = \{(\mathbf{x}, y)\}$

Prediction $\longrightarrow$ $$y = \frac{1}{N} \sum_{d_B(\mathbf{u},\mathbf{u_i})<D(N)} y_i$$

HELMHOLTZ
CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

Defining the Metric in the Predictor's Space

Euclidean Metric (Mahalanobis) $\mathbf{g} = \mathbf{I}_k = diag(1, 1, \ldots, 1)$

Riemannian Metric $\mathbf{g} = [g_{ij}]$

e.g. $[g_{ij}]$ is positive definite

$$g_{ij} = 1 + \alpha_{i,j} u_i u_j \quad \forall i = j$$

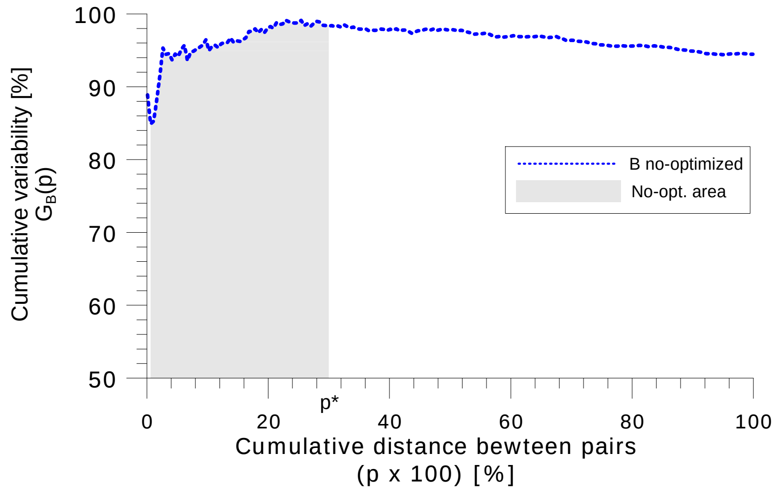
$$g_{ij} = \alpha_{i,j} u_i u_j \quad \forall i \neq j$$

Finding an Appropriate Embedding Space

Define a variance function:

Bárdossy, Pegram & Samaniego, 2005, WRR.

$$G_B(p) = \frac{1}{\mathcal{N}(D_B(p))} \sum_{d_B(i,j)<D_B(p)} \left(\lambda_{ij}\right)^2$$

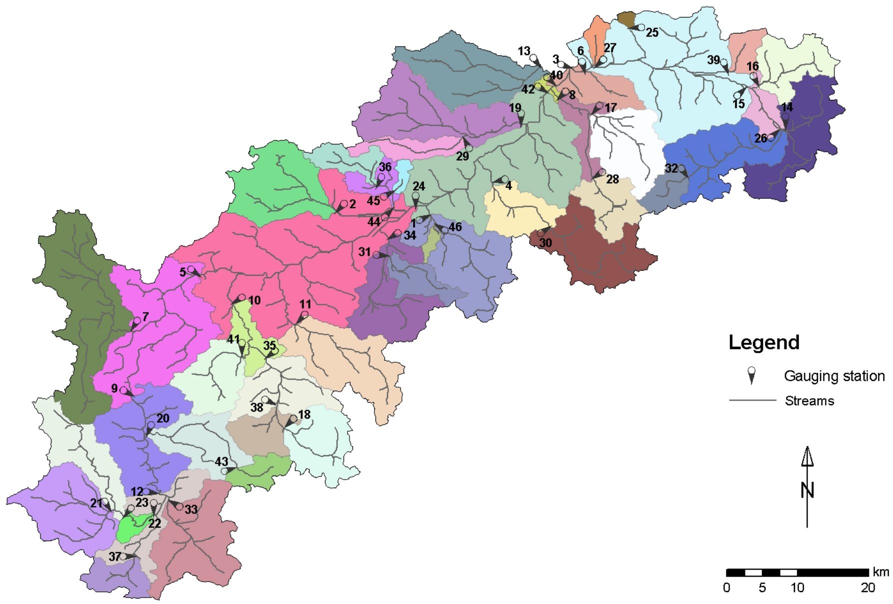


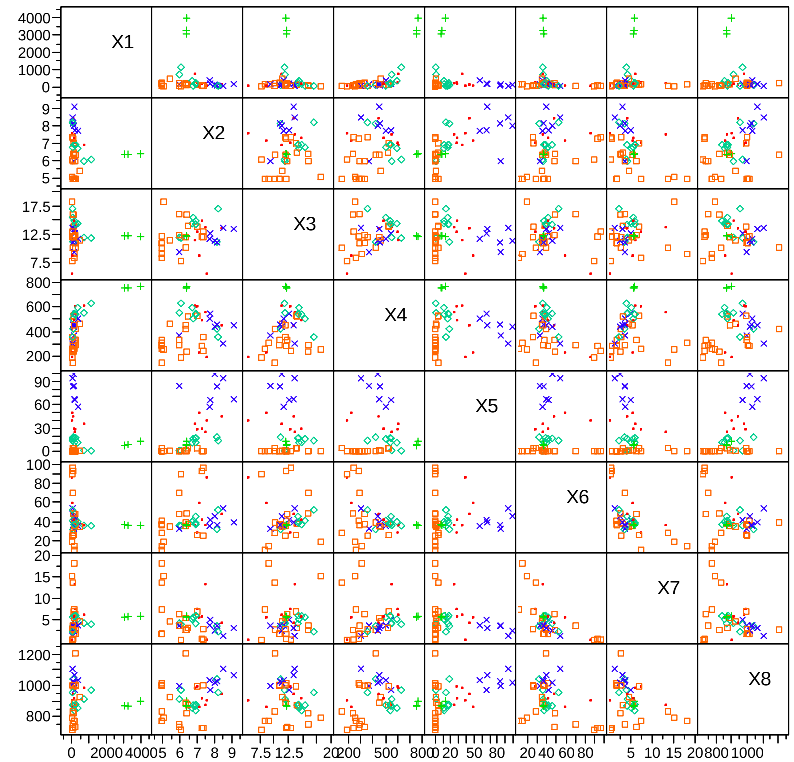
Find a transformation $B[\cdot]$ (e.g. $\mathbf{u} = \mathbf{Bx}$) and a metric $\mathbf{g}$ so that

$$\int_0^{p^*} G_B(p)dp \rightarrow \min$$

3. Application

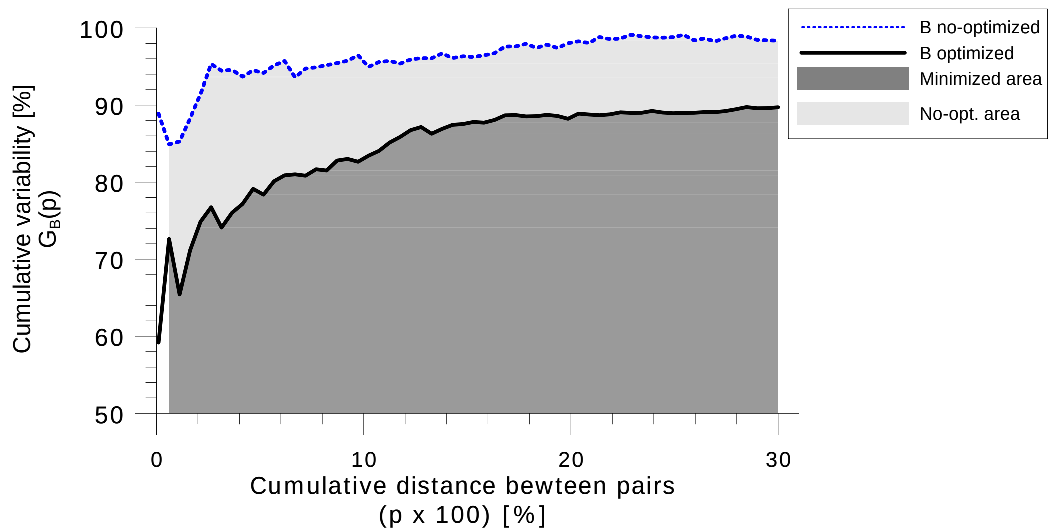
Study Area





x_1	area
x_2	trimmed mean slope $P_{15} - P_{85}$
x_3	north facing slopes
x_4	$h_{max} - h_{min}$ (elevation)
x_5	percentage of karstic formation
x_6	mean share of forest
x_7	mean share of impervious areas
x_8	30y-mean annual precipitation

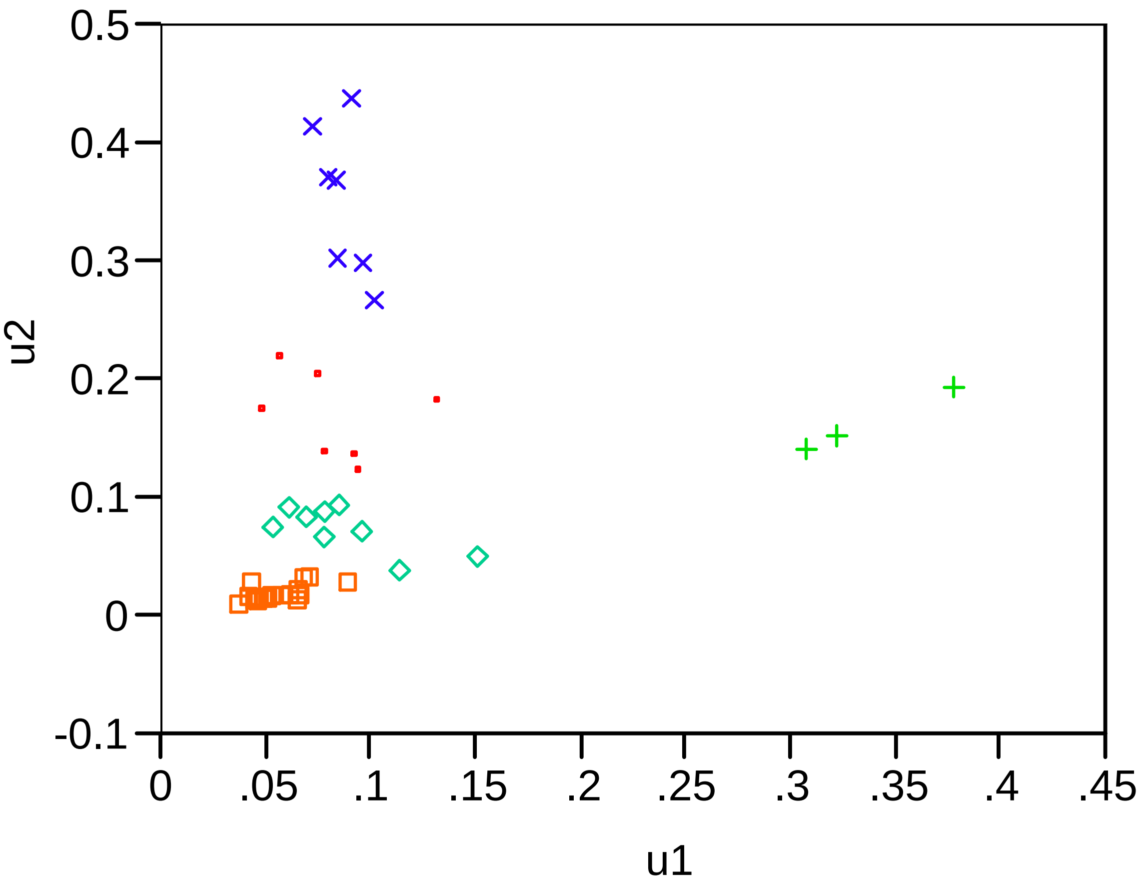
Results



$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 0.29 & 0.00 & 0.00 & 0.07 & 0.03 & 0.02 & 0.03 & 0.01 \\ 0.13 & 0.01 & 0.01 & 0.00 & 0.42 & 0.00 & 0.00 & 0.00 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{pmatrix}$$

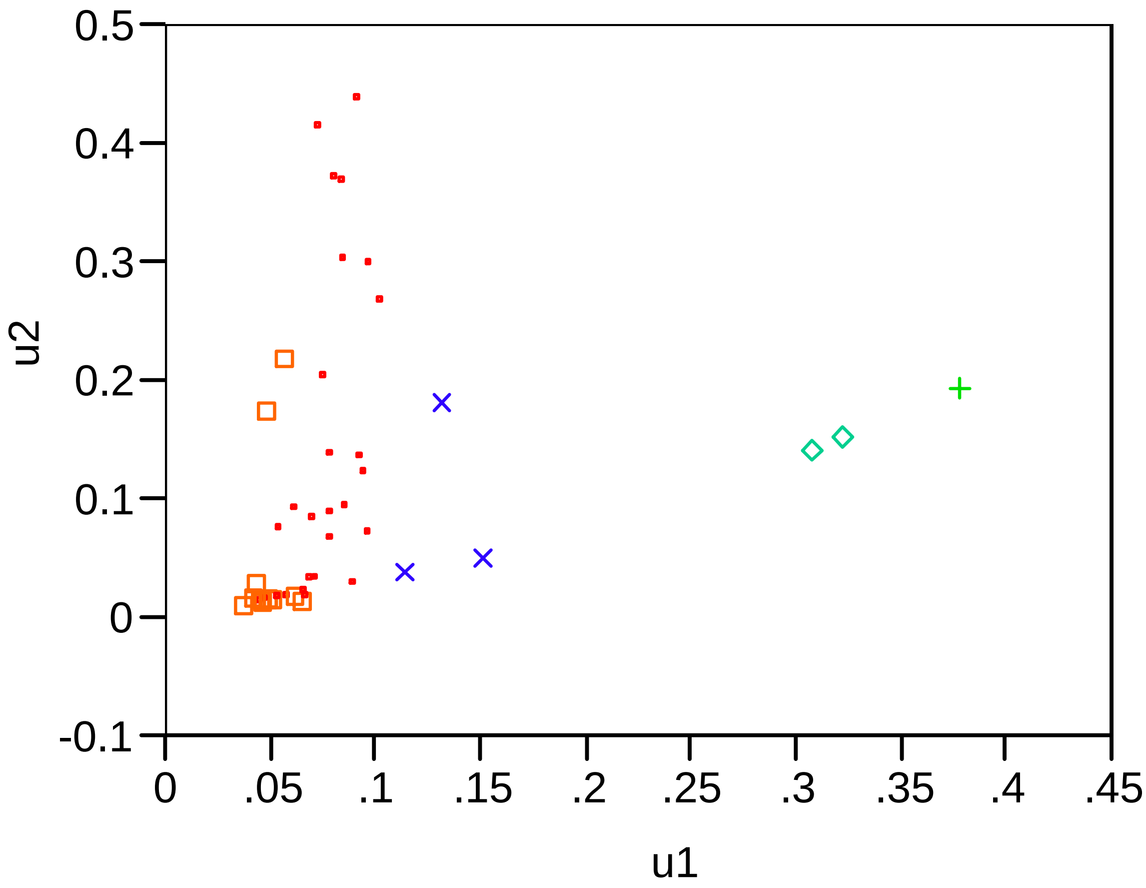
HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Basin Characterization



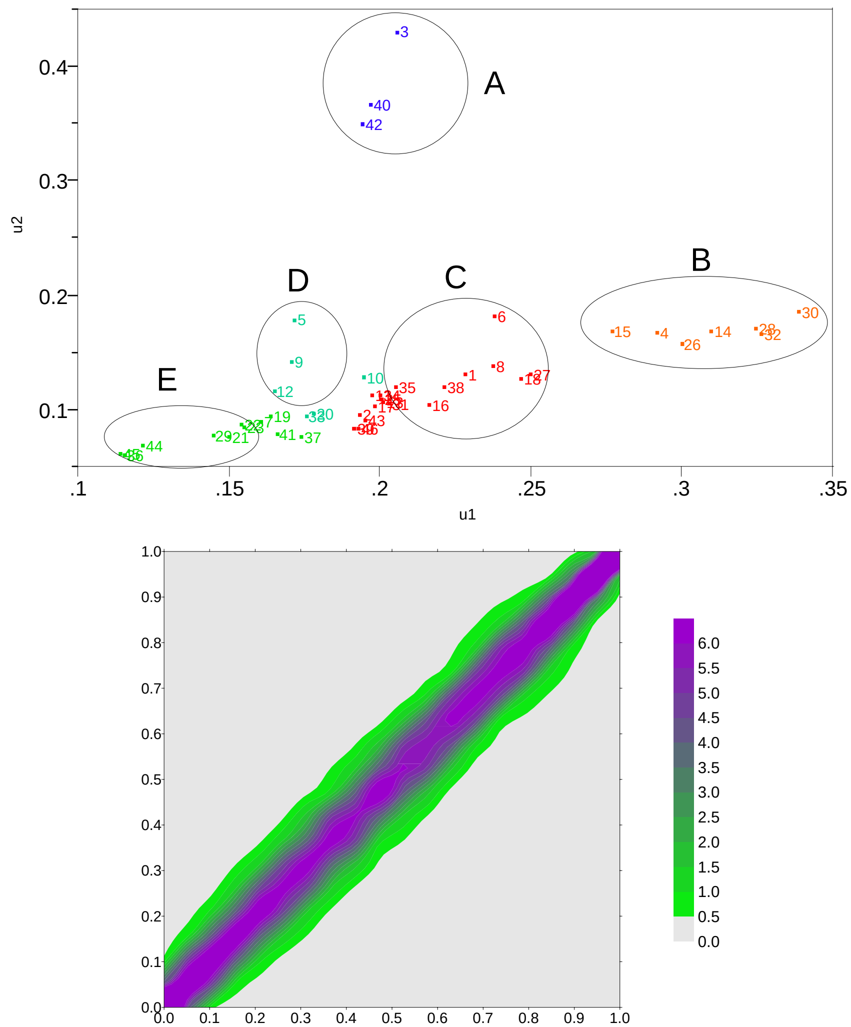
Based on distances in the embedded space $\mathbf{u}$

Basin Characterization



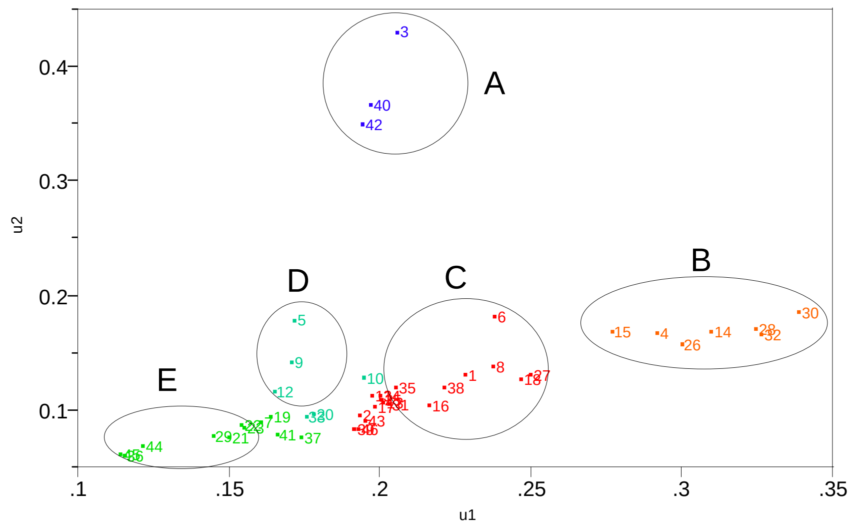
Based on distances in the input space $\mathbf{x}$

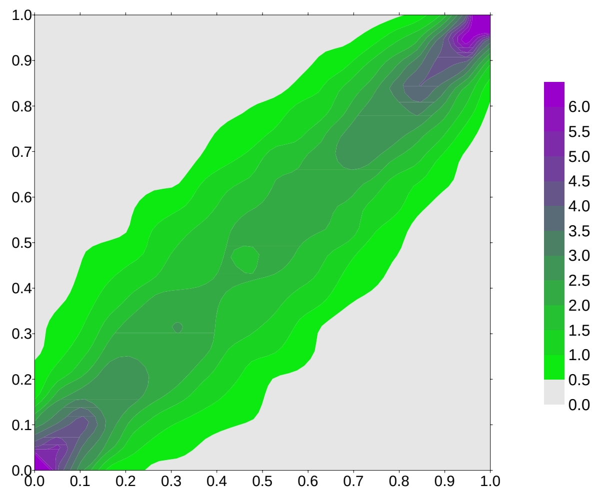
Typical Copulas within the Clusters



Cluster A

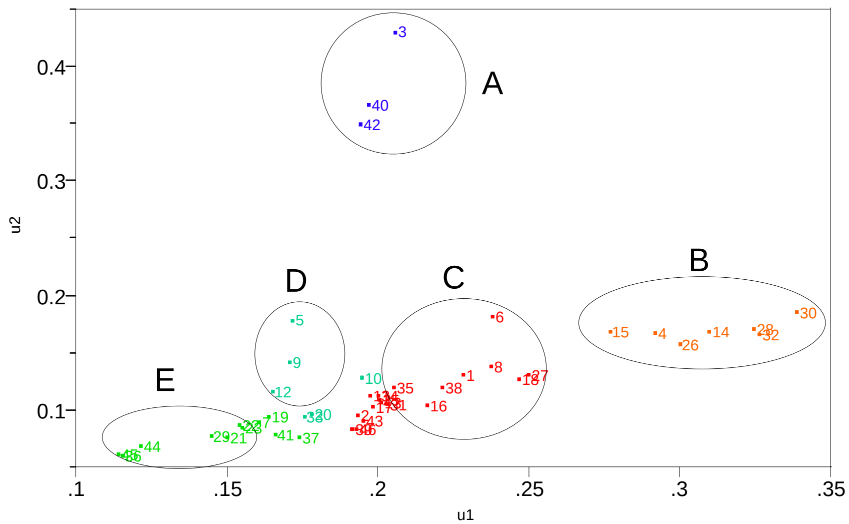
Typical Copulas within the Clusters

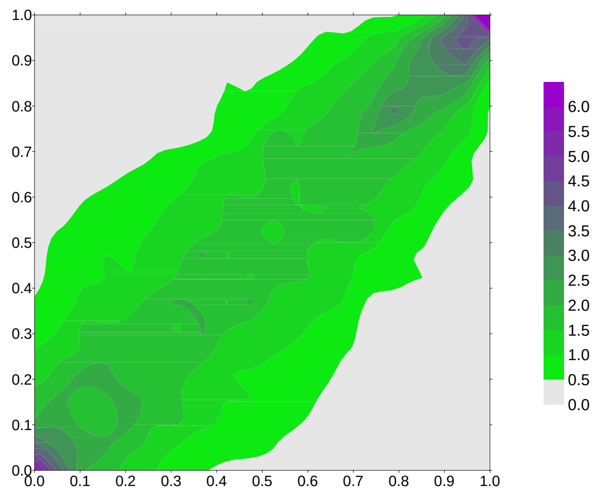




Cluster B

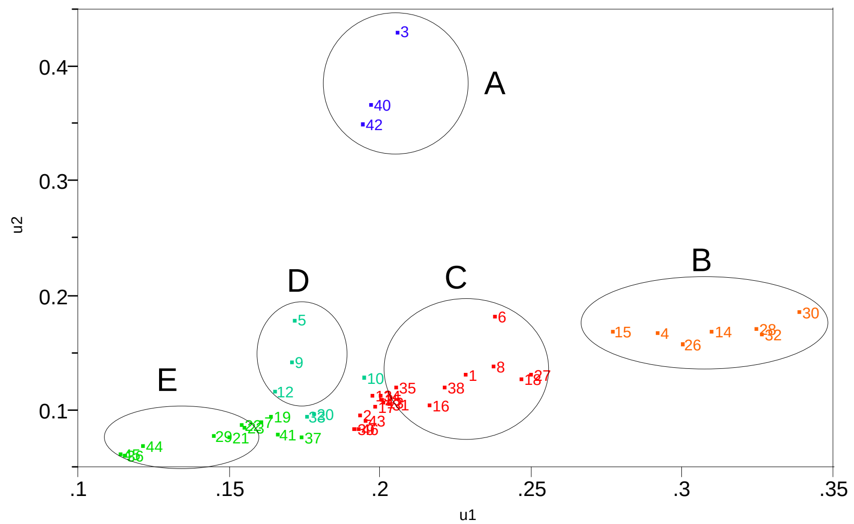
Typical Copulas within the Clusters

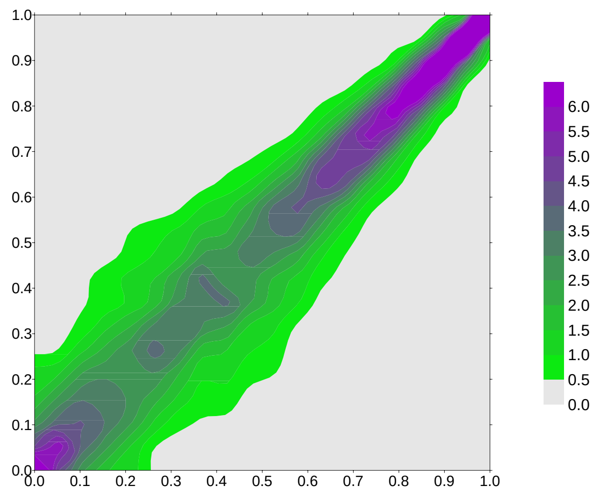




Cluster C

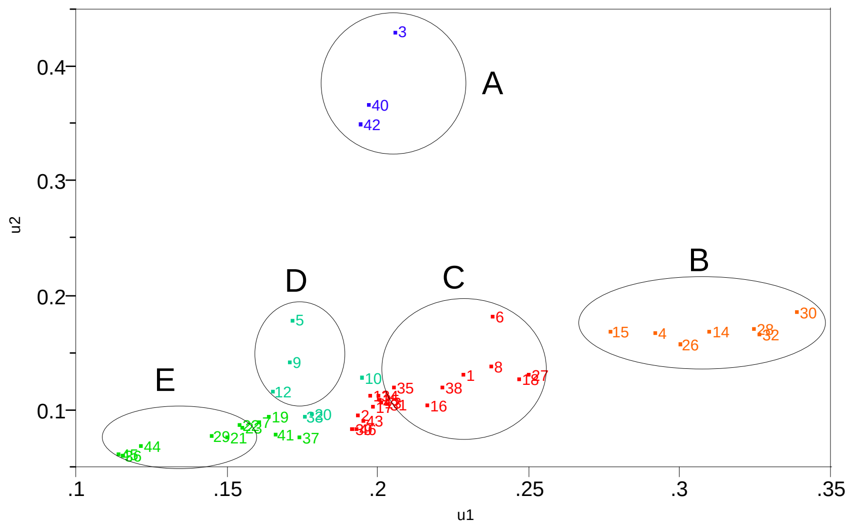
Typical Copulas within the Clusters

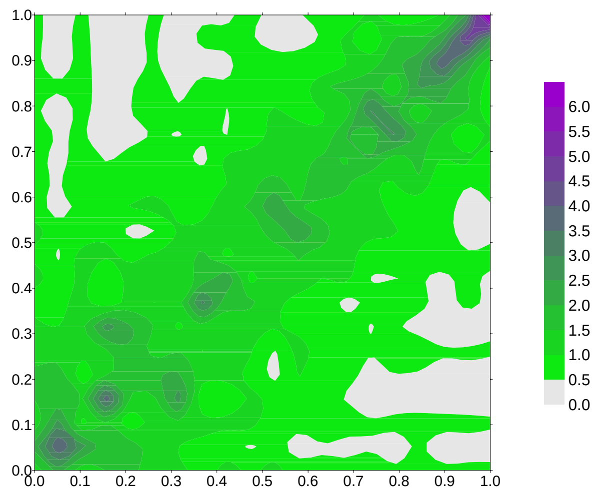




Cluster D

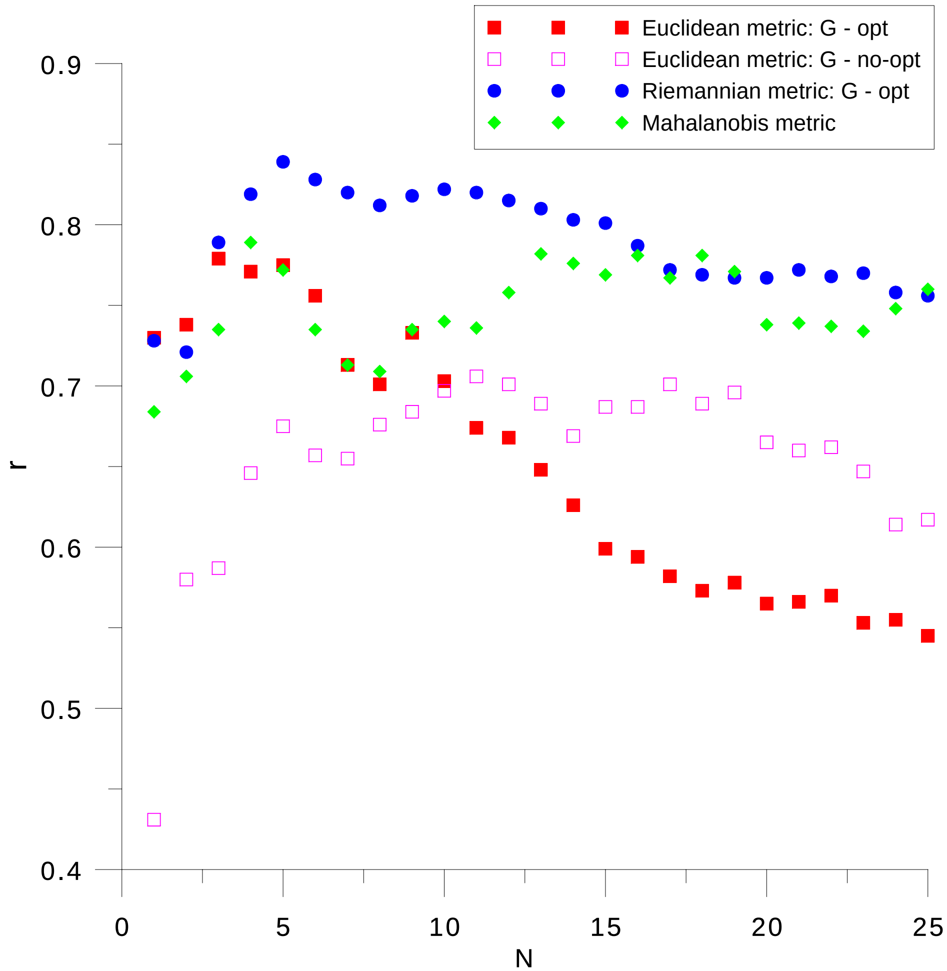
Typical Copulas within the Clusters





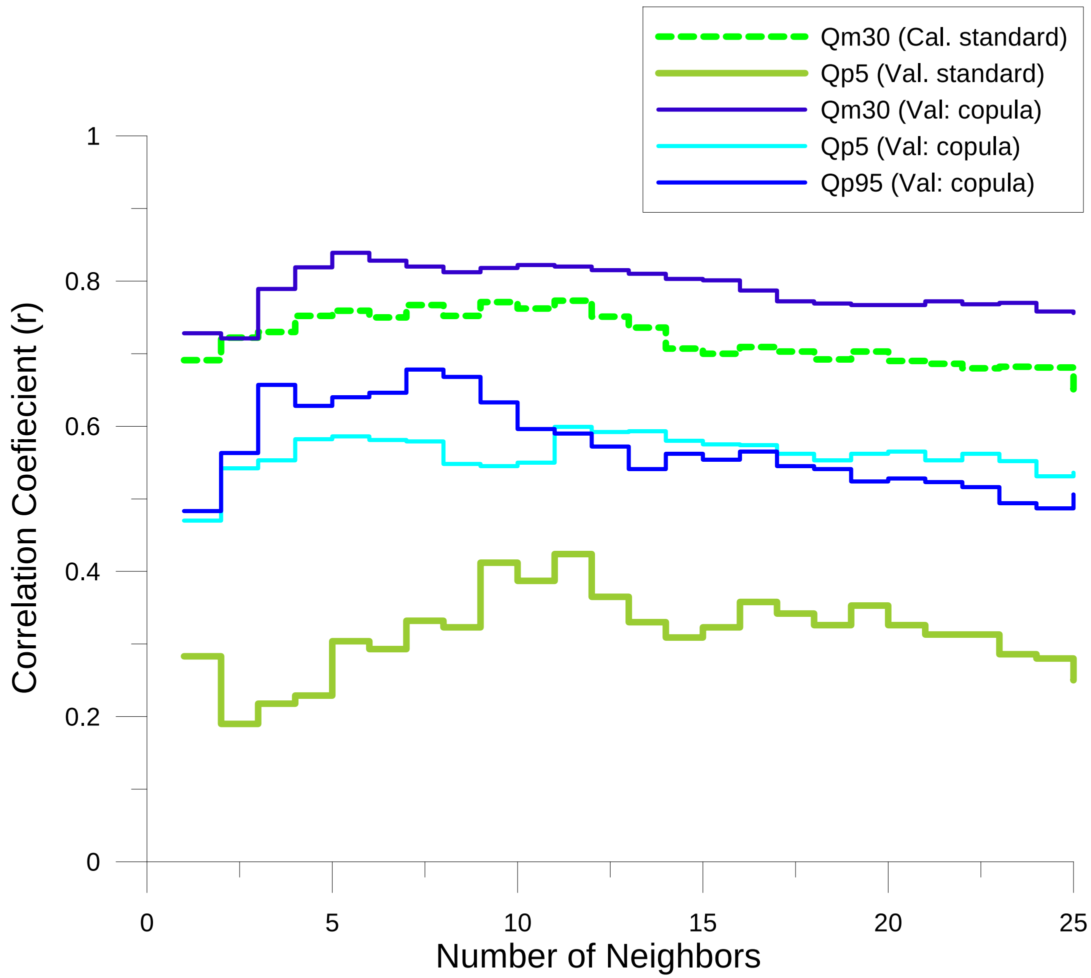
Cluster E

Validation



Performance of different metrics

Validation



Standard vs. copula based

4. Conclusions and Outlook

- The embedding space and the adaptive metric performed better than *a priori* selected standard metrics.

4. Conclusions and Outlook

- The embedding space and the adaptive metric performed better than *a priori* selected standard metrics.
- The distance based on density copulas lead to robust characterizations.

4. Conclusions and Outlook

- The embedding space and the adaptive metric performed better than *a priori* selected standard metrics.
- The distance based on density copulas lead to robust characterizations.
- Validation with several runoff characteristics (y) gives $r > 0.65$.

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

4. Conclusions and Outlook

- The embedding space and the adaptive metric performed better than *a priori* selected standard metrics.
- The distance based on density copulas lead to robust characterizations.
- Validation with several runoff characteristics (y) gives $r > 0.65$.

- Additional similarity (or dissimilarity) measures are still to be investigated.

4. Conclusions and Outlook

- The embedding space and the adaptive metric performed better than *a priori* selected standard metrics.
- The distance based on density copulas lead to robust characterizations.
- Validation with several runoff characteristics (y) gives $r > 0.65$.

- Additional similarity (or dissimilarity) measures are still to be investigated.
- Expand the data set to include a number of basins with different hydrological regimes.

Thank you

HELMHOLTZ
CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

Appendix