

# Extremes

*Extremes*

2022-08-31

Extremes

Benedikt Gräler



Univariate  
Extremes

Multivariate  
Extremes

Spatial Extremes

Hands-on

Benedikt Gräler

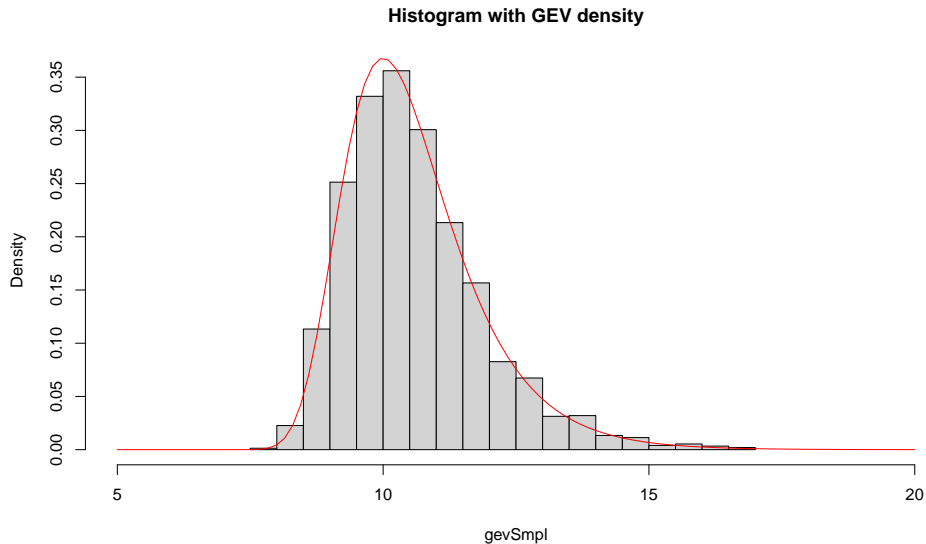
52°North Spatial Information Research GmbH

<https://52north.org>

## Univariate Extremes

- Extremes are rare events / measurements
  - few records in a spatial distributed scene
  - few records in a time series
  - both
- by definition, we will only observe few extremes (if any) in a given sample

# What is special about extremes?



## Extremes

Benedikt Gräler



Univariate  
Extremes

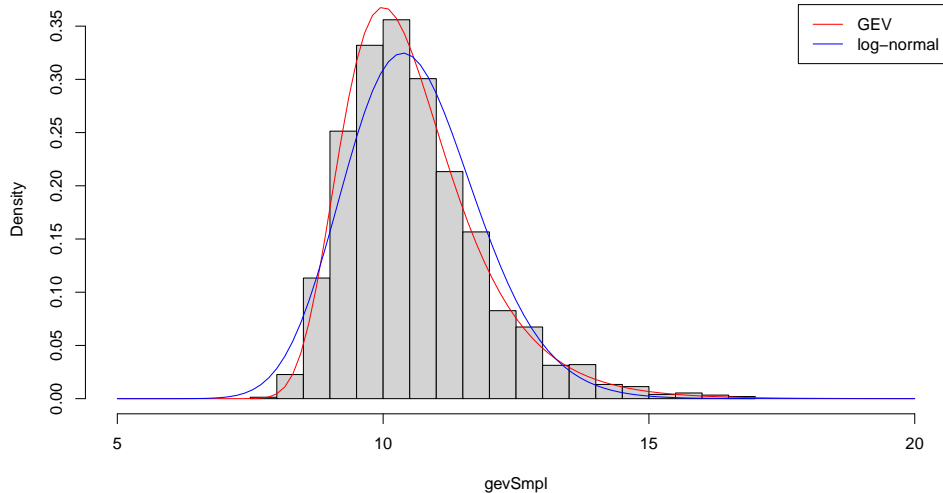
Multivariate  
Extremes

Spatial Extremes

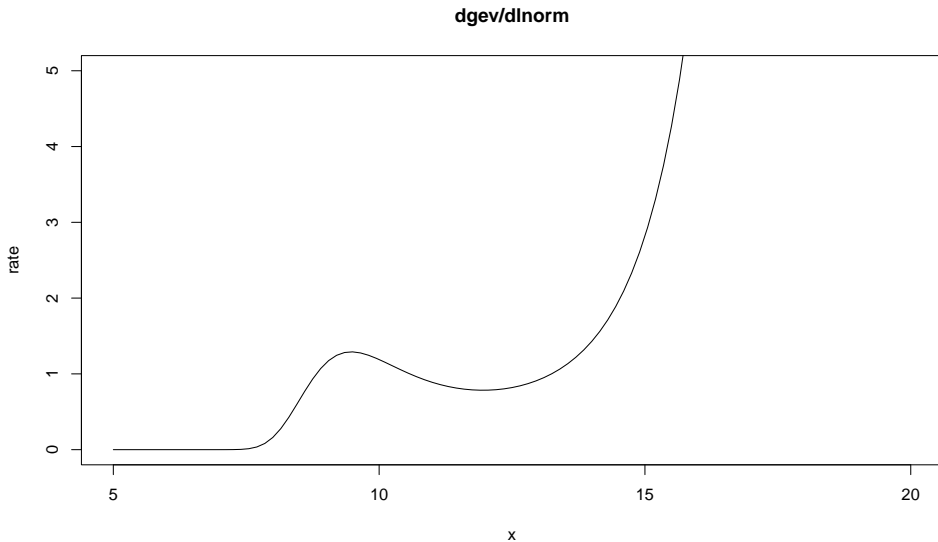
Hands-on

# What is special about extremes?

Histogram with GEV and log-normal densities.



# What is special about extremes?



## Extremes

Benedikt Gräler



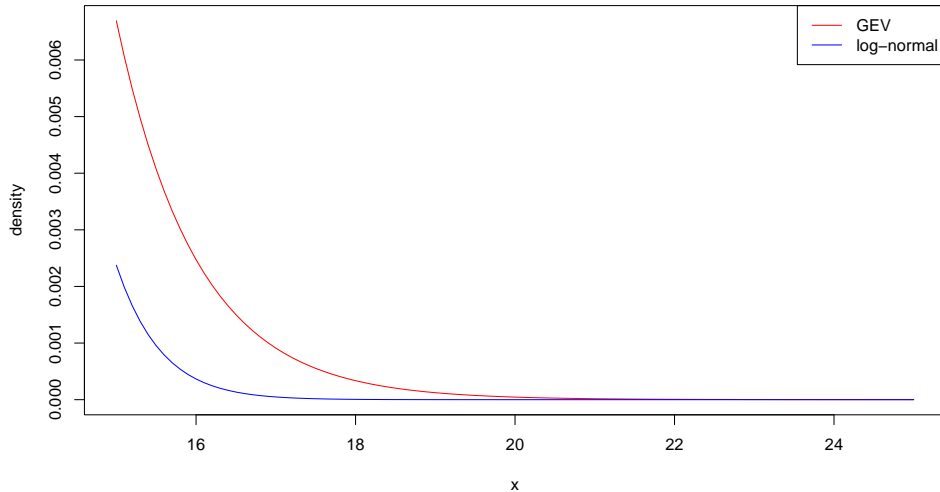
Univariate  
Extremes

Multivariate  
Extremes

Spatial Extremes

Hands-on

# What is special about extremes?



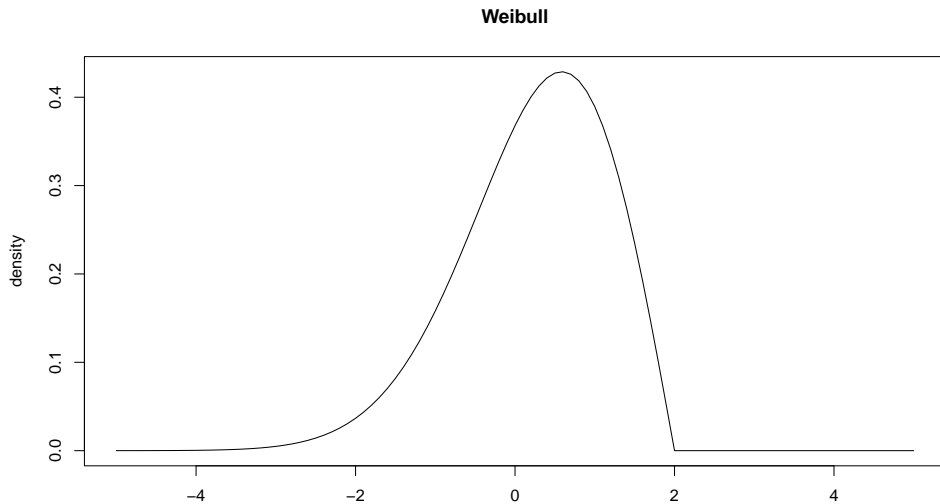
Let  $X_1, \dots, X_n$  be a sequence of independent and identically distributed random variables with cumulative distribution function  $F$  and let  $M_n = \max(X_1, \dots, X_n)$  denote the maximum.

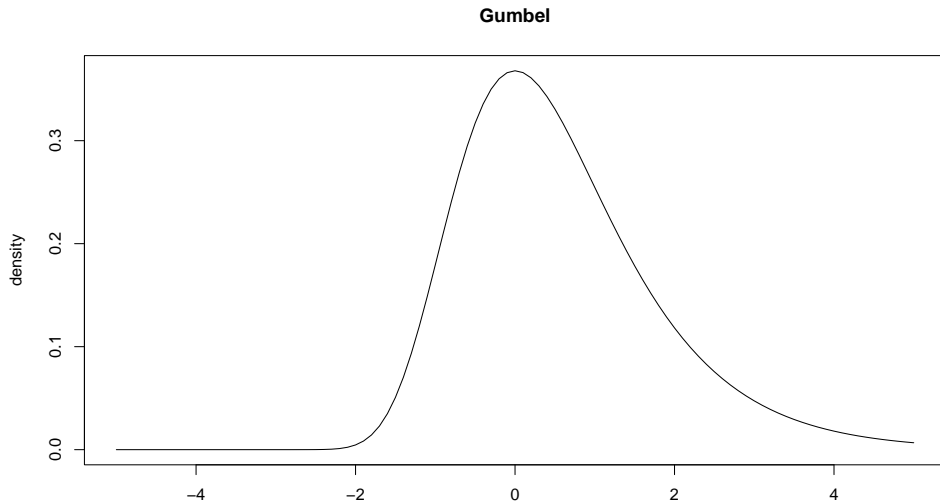
In theory, the exact distribution of the maximum can be derived:

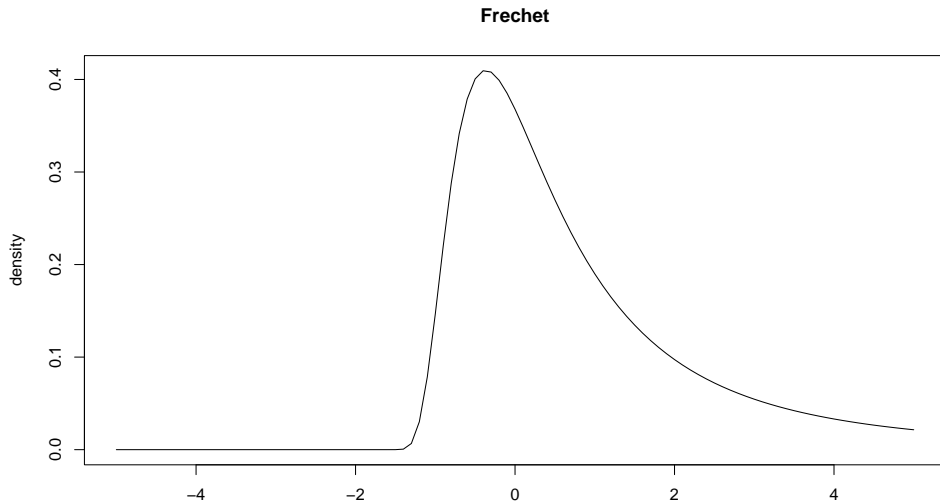
$$\begin{aligned}\Pr(M_n \leq z) &= \Pr(X_1 \leq z, \dots, X_n \leq z) \\ &= \Pr(X_1 \leq z) \cdots \Pr(X_n \leq z) = (F(z))^n.\end{aligned}$$

Source: [https://en.wikipedia.org/wiki/Extreme\\_value\\_theory](https://en.wikipedia.org/wiki/Extreme_value_theory)









$$G(z) = \exp \left( - (1 + s(z - a)/b)^{-1/s} \right)$$

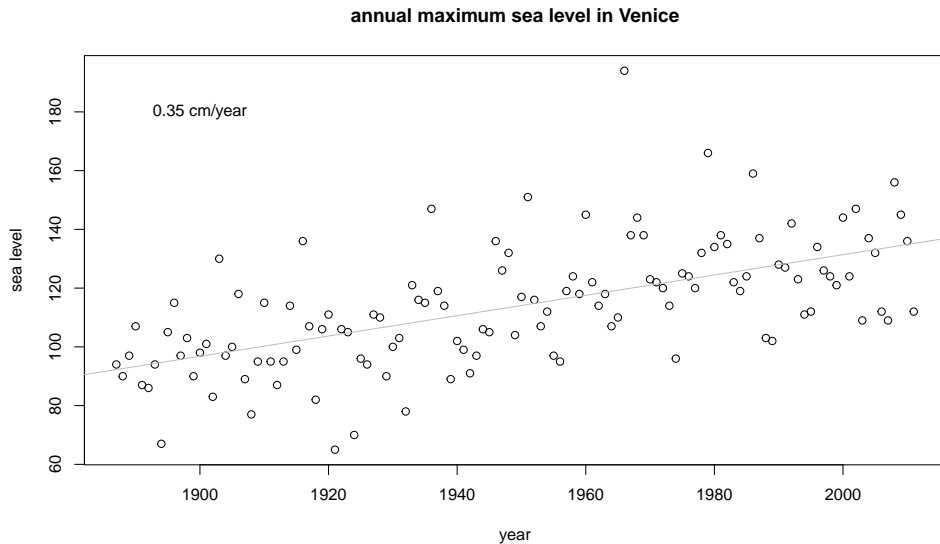
with  $\text{loc} = a$ ,  $\text{scale} = b$  and  $\text{shape} = s$ :

```
dgev(x, loc=0, scale=1, shape=0, log = FALSE)
pgev(q, loc=0, scale=1, shape=0, lower.tail = TRUE)
qgev(p, loc=0, scale=1, shape=0, lower.tail = TRUE)
rgev(n, loc=0, scale=1, shape=0)
```

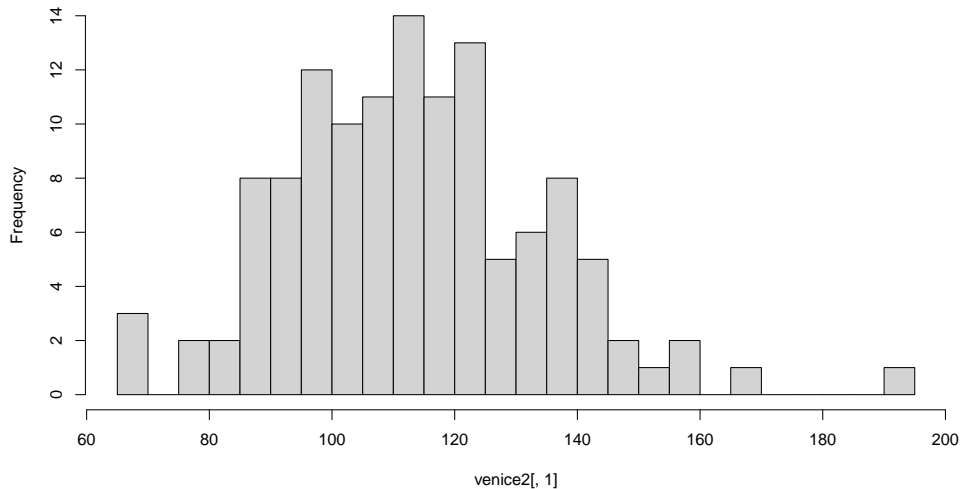
- Weibull:  $s < 0$
- Gumbel:  $s = 0$
- Frechet:  $s > 0$

select your maxima per block (typically temporal) to avoid (at least reduce) auto-correlation and be closer to an iid sample

- month
- annually
- daily
- ...



annual maximum sea levels 1887 – 2011

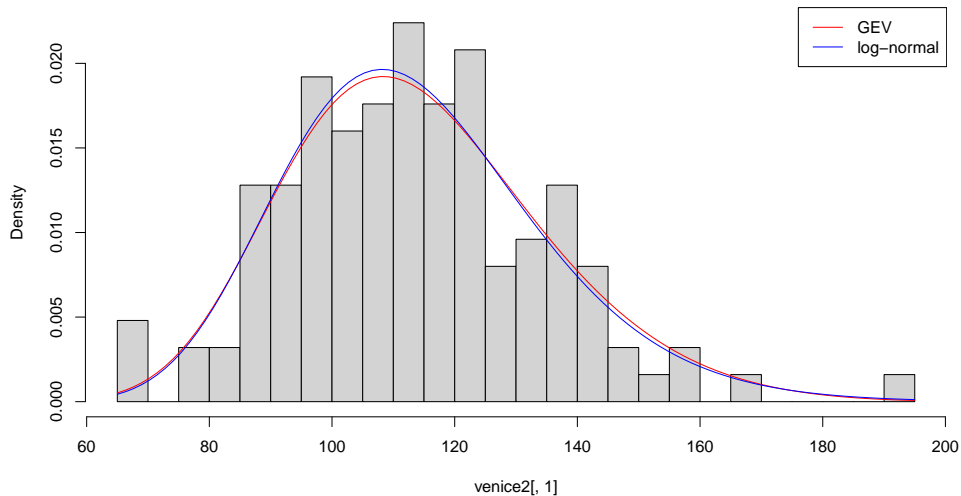


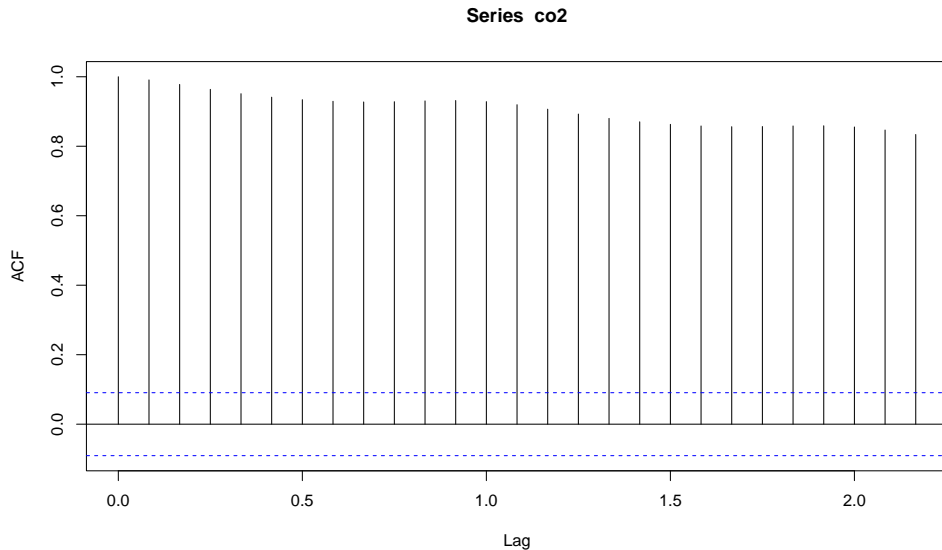
```
fgev(venice2[,1])

##
## Call: fgev(x = venice2[, 1])
## Deviance: 1111.223
##
## Estimates
##      loc      scale      shape
## 105.2995  19.3543  -0.1463
##
## Standard Errors
##      loc      scale      shape
## 1.87769  1.27804  0.04176
##
## Optimization Information
##   Convergence: successful
##   Function Evaluations: 27
##   Gradient Evaluations: 11
```

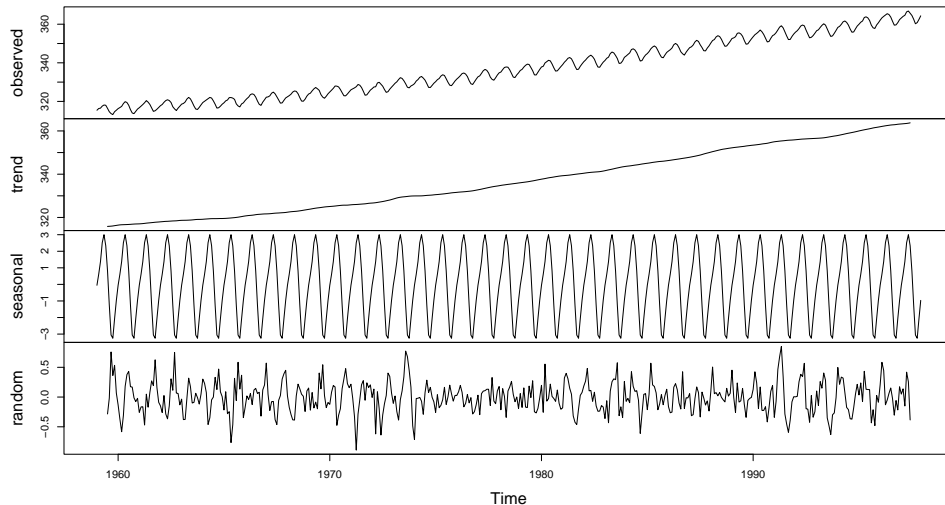


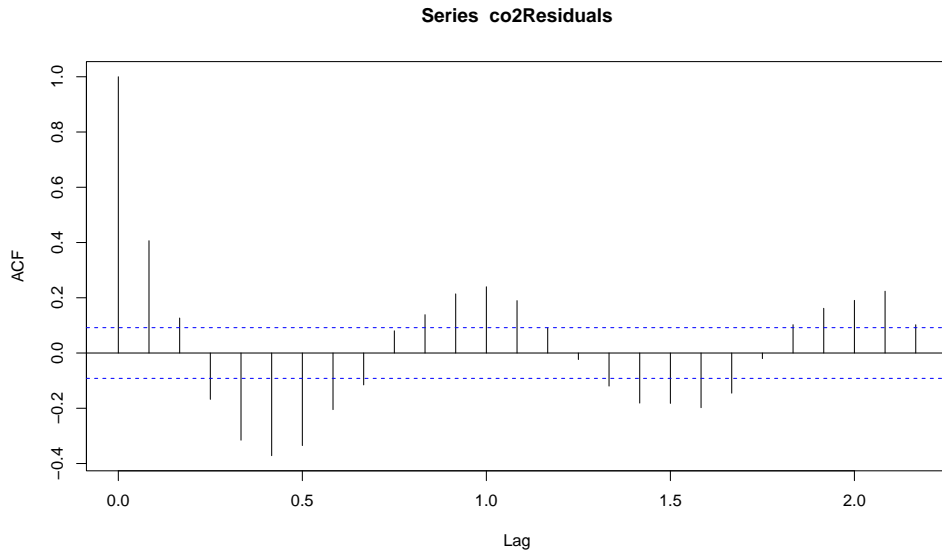
annual maximum sea levels 1887 – 2011

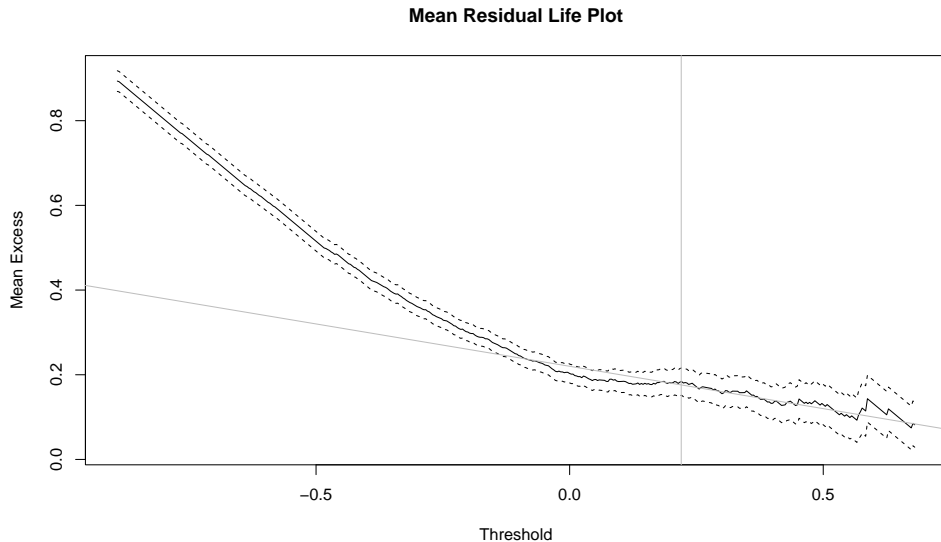




### Decomposition of additive time series

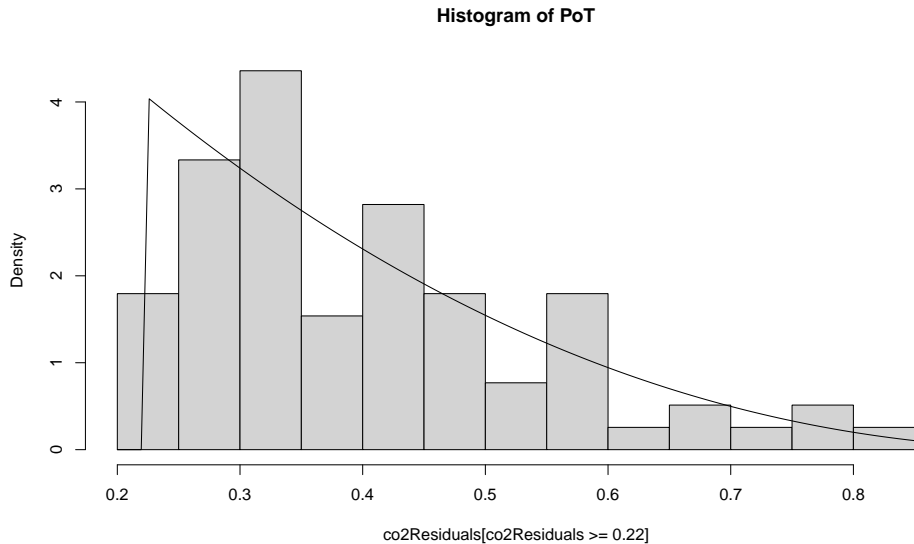






```
fspot(co2Residuals, threshold = 0.22)
```

```
##  
## Call: fspot(x = co2Residuals, threshold = 0.22)  
## Deviance: -113.6398  
##  
## Threshold: 0.22  
## Number Above: 78  
## Proportion Above: 0.1711  
##  
## Estimates  
##   scale   shape  
## 0.2436 -0.3161  
##  
## Standard Errors  
##   scale   shape  
## 0.03568 0.09961  
##  
## Optimization Information  
##   Convergence: successful  
##   Function Evaluations: 29  
##   Gradient Evaluations: 7
```

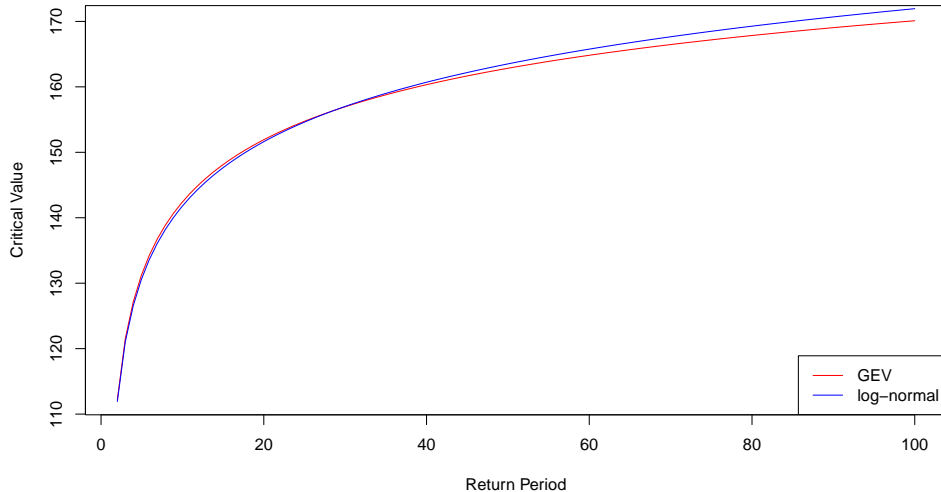


“On average, how many years will it take to observe an event this large/small?”

This is  $P(X \geq x) = 1/T$  where  $X$  is your annual maximum variable and  $T$  is the return period (typically 5, 10, 25, 50, 100, ... years),  $x$  is the “critical event”.



For a given return period, what is the annual maximum water level in Venice?



## Extremes

Benedikt Gräler

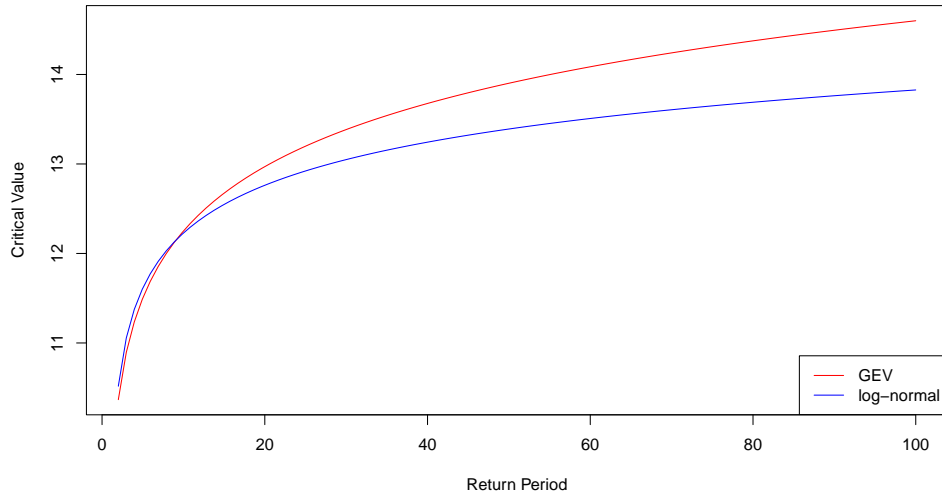
**52north**  
exploring horizons

Univariate  
Extremes

Multivariate  
Extremes

Spatial Extremes

Hands-on



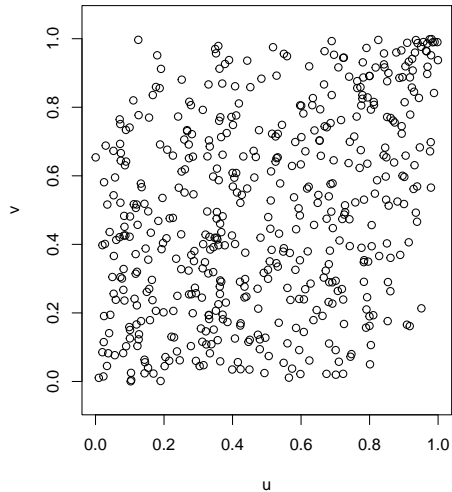
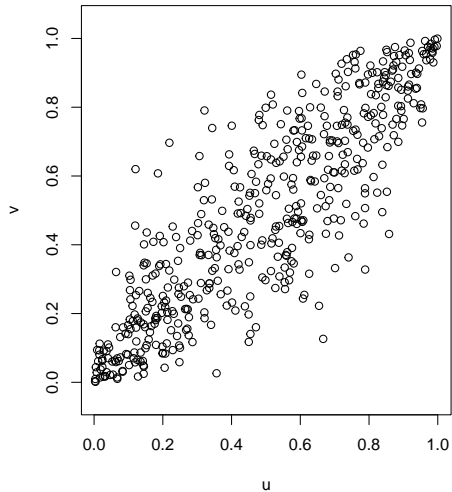
## Multivariate Extremes

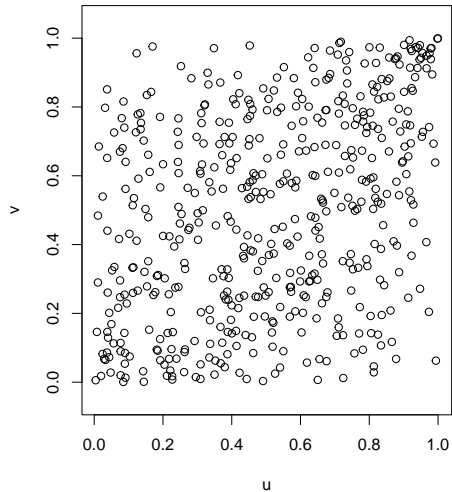
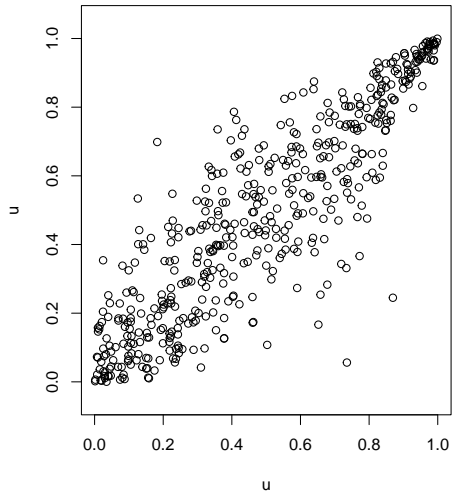
- Once we look at more than one variable, we have to think about dependencies.
- Correlation measures summarize dependence in a single number, but as for univariate statistics, a multivariate distribution has more characteristics than a single number
- while histograms are often the first step for univariate cases, scatter plots are the first choice for multivariate case

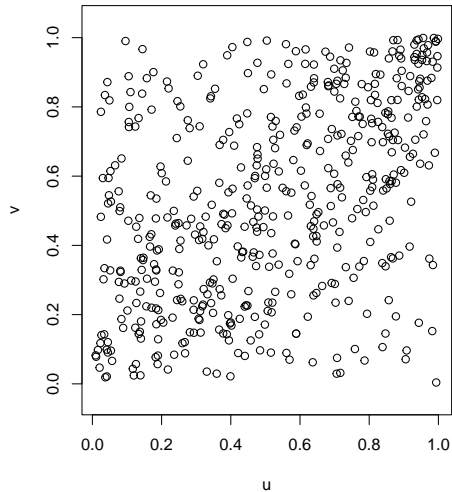
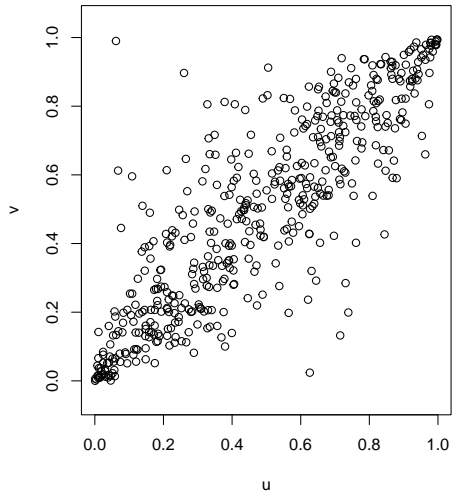
- Data is only sometimes Gaussian
- Most approaches based correlation/covariance matrices will assume a multivariate Gaussian distributions
- a wider concept are **copulas**. Any continuous multivariate distribution  $H$  with its  $X_1, \dots, X_n$  marginal random variables can be decomposed into its marginal distribution functions  $F_i$  and its copula  $C$ :

$$H(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n))$$

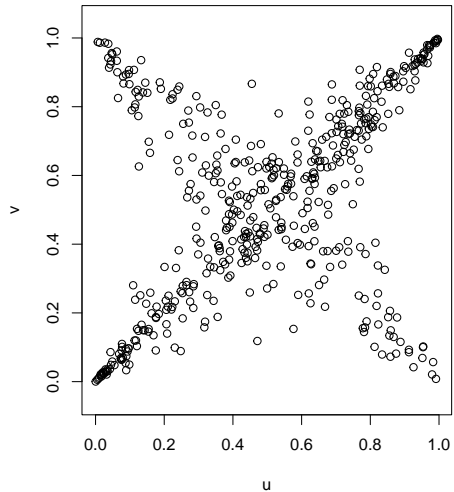
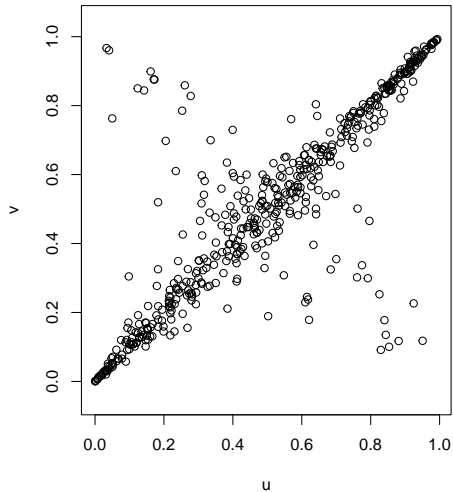
A copula can be perceived as a multivariate distribution on the unit hyper-cube  $[0, 1]^n$ .











<https://copulatheque.org>

- What happens with the dependence if both margins become large?
- Composite extremes is assessed via **upper/lower tail dependence**

```
lambda(normalCopula(iTau(normalCopula(), 0.7)))
```

```
## lower upper
```

```
##      0      0
```

```
lambda(gumbelCopula(iTau(gumbelCopula(), 0.7)))
```

```
##      lower      upper
```

```
## 0.0000000 0.7688556
```

```
lambda(tCopula(iTau(tCopula(), 0.7)))
```

```
##      lower      upper
```

```
## 0.6144008 0.6144008
```

```
lambda(tCopula(iTau(tCopula(df=0.3), 0.7), df=0.3))
```

```
##      lower      upper
```

```
## 0.8213634 0.8213634
```

```
lambda(normalCopula(iTau(normalCopula(), 0.3)))
```

```
## lower upper
```

```
##      0      0
```

```
lambda(gumbelCopula(iTau(gumbelCopula(), 0.3)))
```

```
##      lower      upper
```

```
## 0.0000000 0.3754952
```

```
lambda(tCopula(iTau(tCopula(), 0.3)))
```

```
##      lower      upper
```

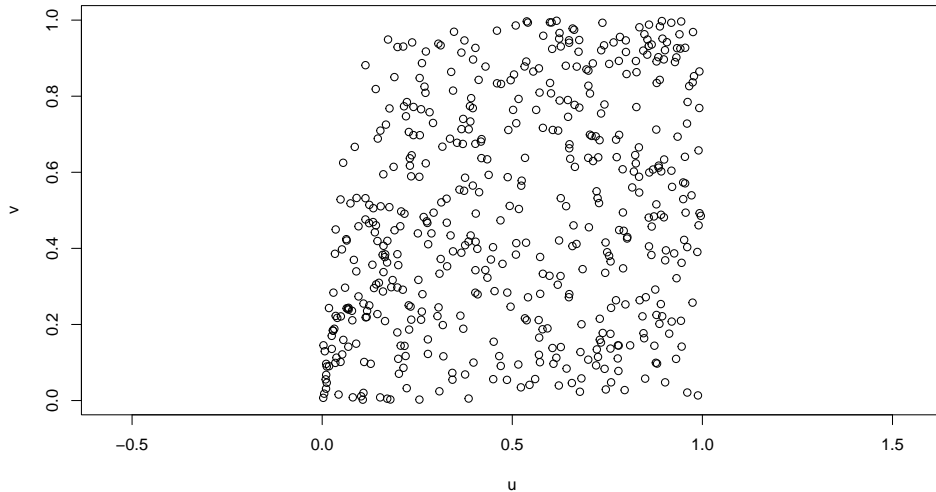
```
## 0.2289254 0.2289254
```

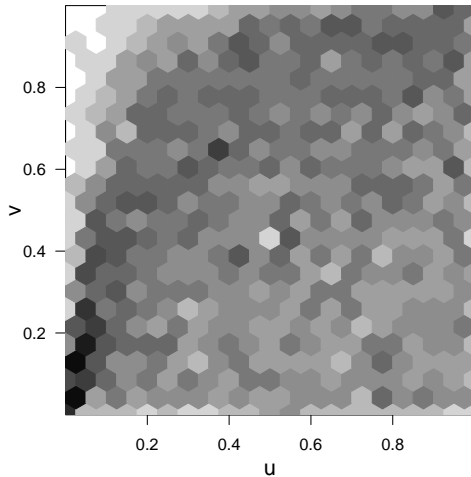
```
lambda(tCopula(iTau(tCopula(df=0.3), 0.3), df=0.3))
```

```
##      lower      upper
```

```
## 0.5884455 0.5884455
```

```
aqPseudoObs <- pobs(cbind(airquality$Solar.R,  
                           airquality$Temp))  
BiCopSelect(aqPseudoObs[,1], aqPseudoObs[,2])  
  
## Warning: In BiCopSelect: 7 observations (4.6%) contain NAs. Only complete  
## observations are used.  
  
## Bivariate copula: Rotated Tawn type 2 180 degrees (par = 2.46, par2 = 0.2, tau = 0.16)
```







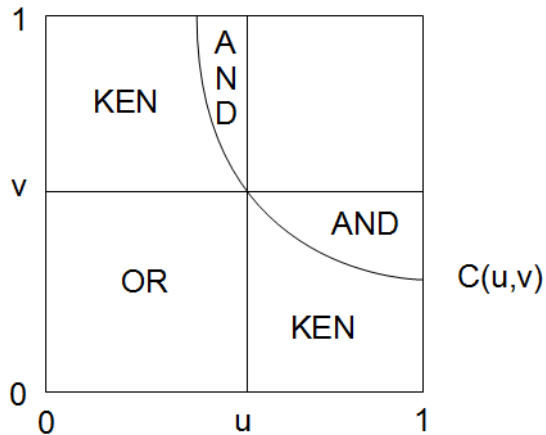
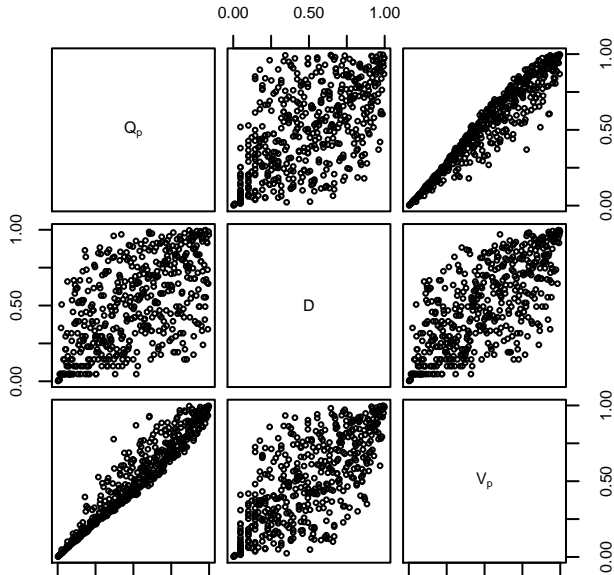
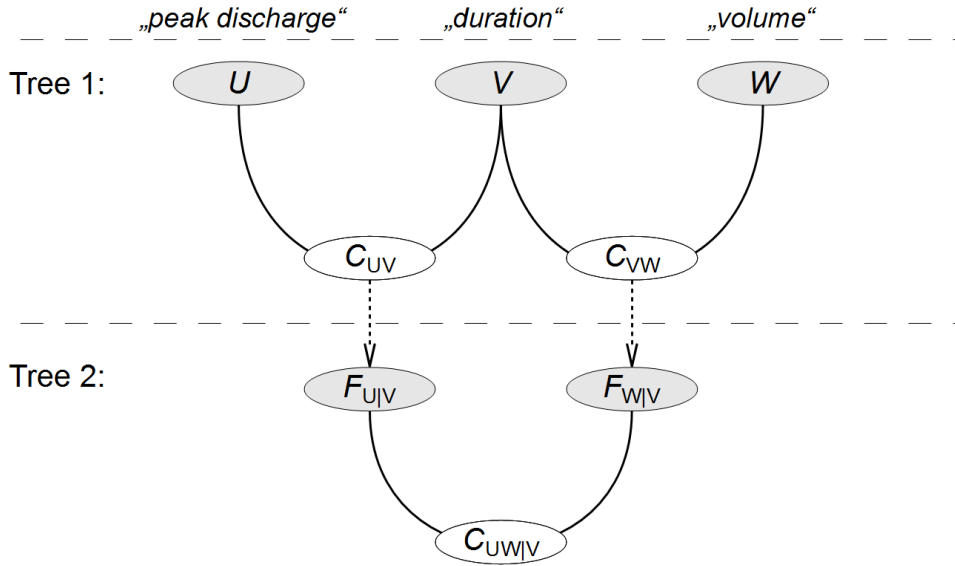


Figure 1: alt text here





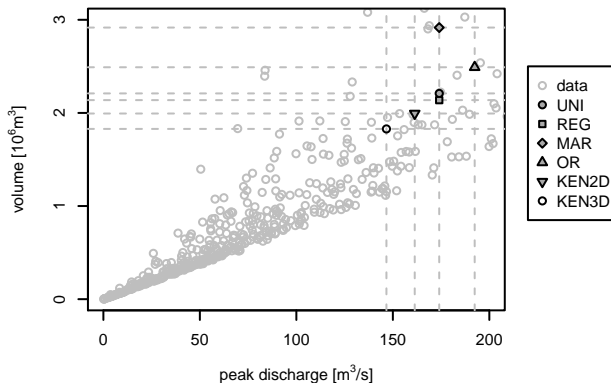


Figure 4: comparison of critical events

# Spatial Extremes

## Extremes

Benedikt Gräler



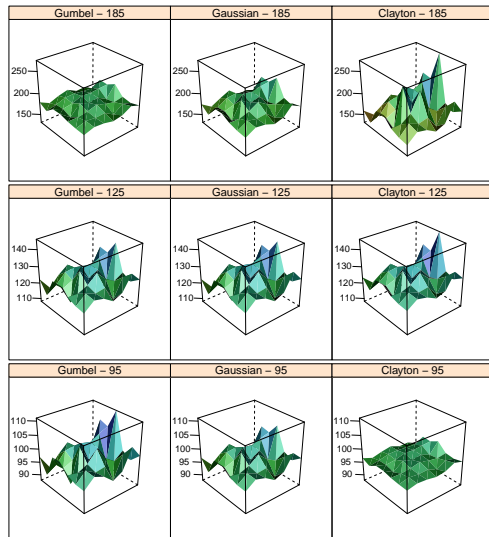
Univariate  
Extremes

Multivariate  
Extremes

Spatial Extremes

Hands-on

# Does that make any difference?



## Extremes

Benedikt Gräler

**52north**  
exploring horizons

Univariate  
Extremes

Multivariate  
Extremes

Spatial Extremes

Hands-on

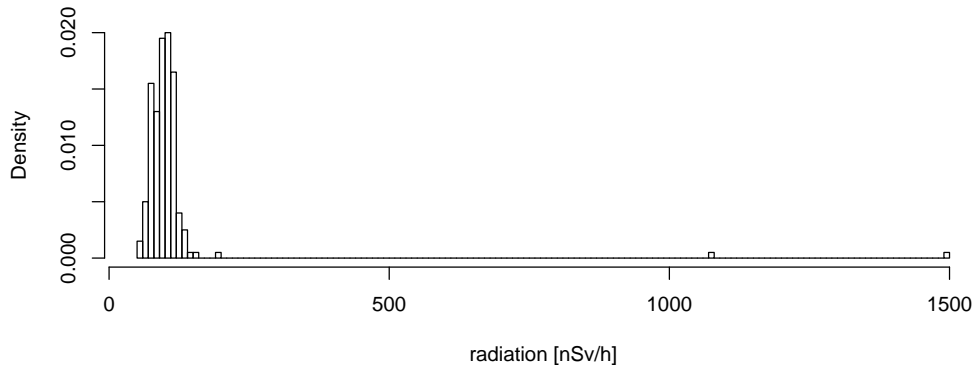


Figure 5: Marginal Distribution

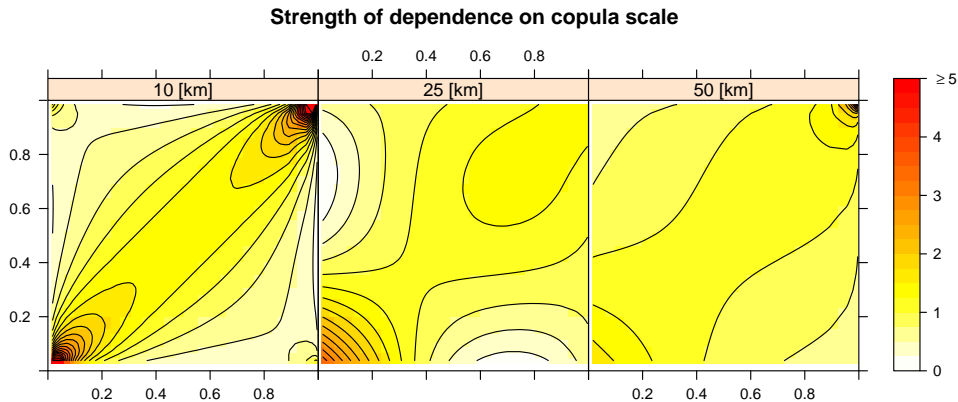
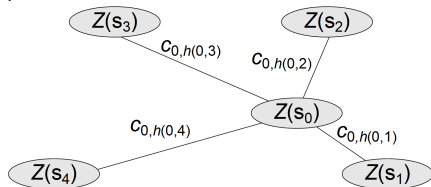


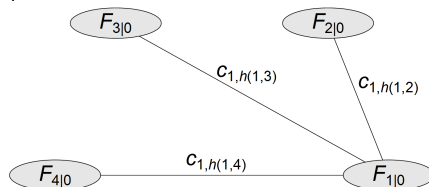
Figure 6: Spatially varying copulas



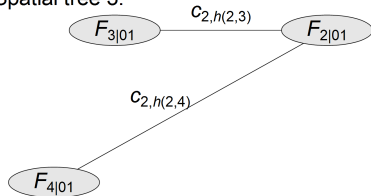
Spatial tree 1:



Spatial tree 2:



Spatial tree 3:



Spatial tree 4:

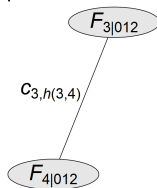
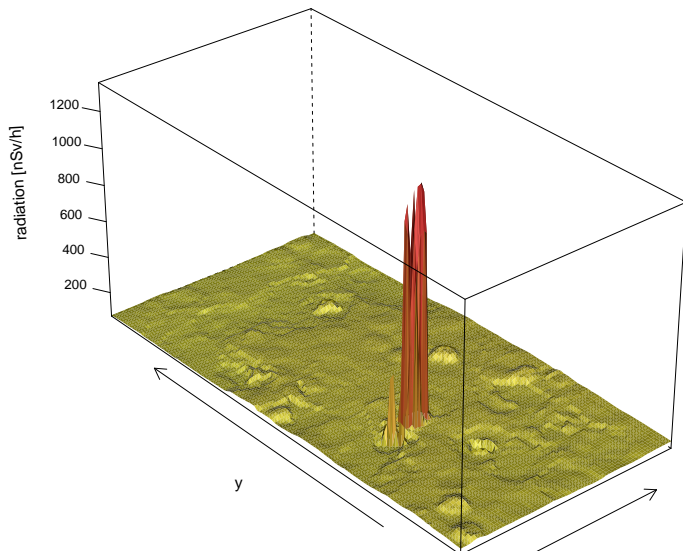


Figure 7: Spatial Vine



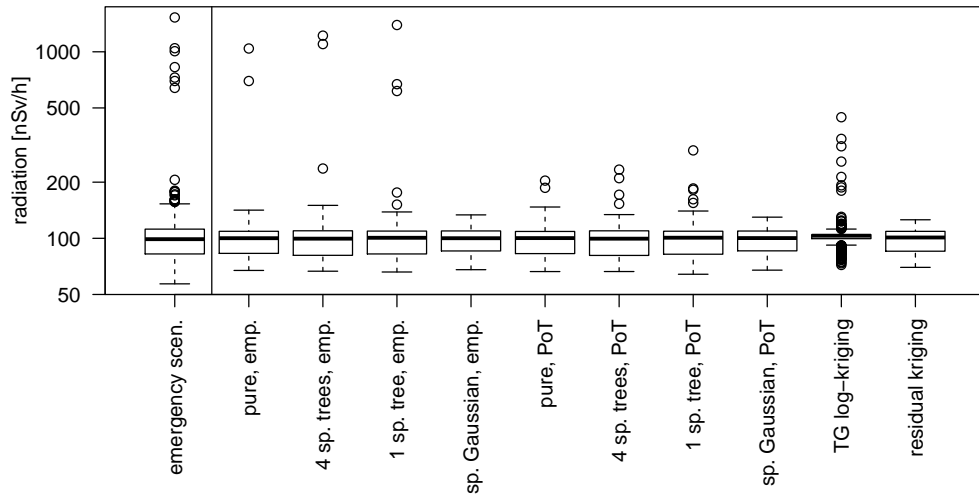
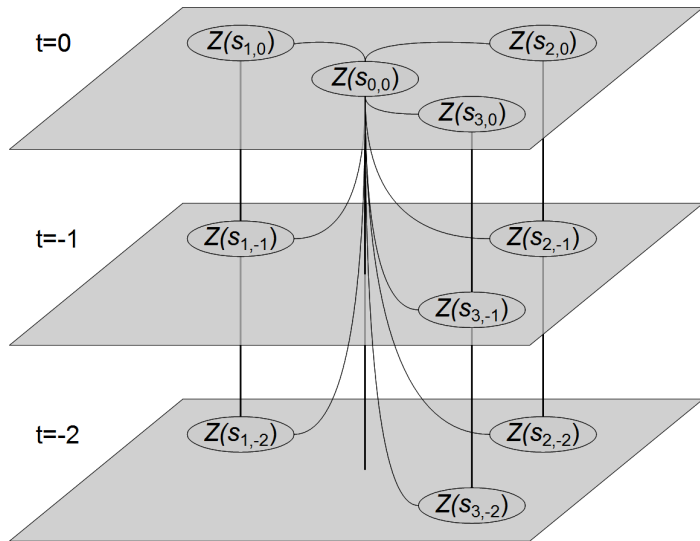


Figure 9: different approaches



# Hands-on

- copula
- VineCopula
- VC2copula
- spcopula (only on GitHub, slight workaround needed for VC2copula)

- retry the code snippets in the underlying Rmd-file
- use your own data set and check scatter plots of pseudo observations
- data set uranium of the copula package
- demo MRP of the spcopula package
- demo pureSpVineCopula of the spcopula package