Extremes

Extremes
2022-08-31

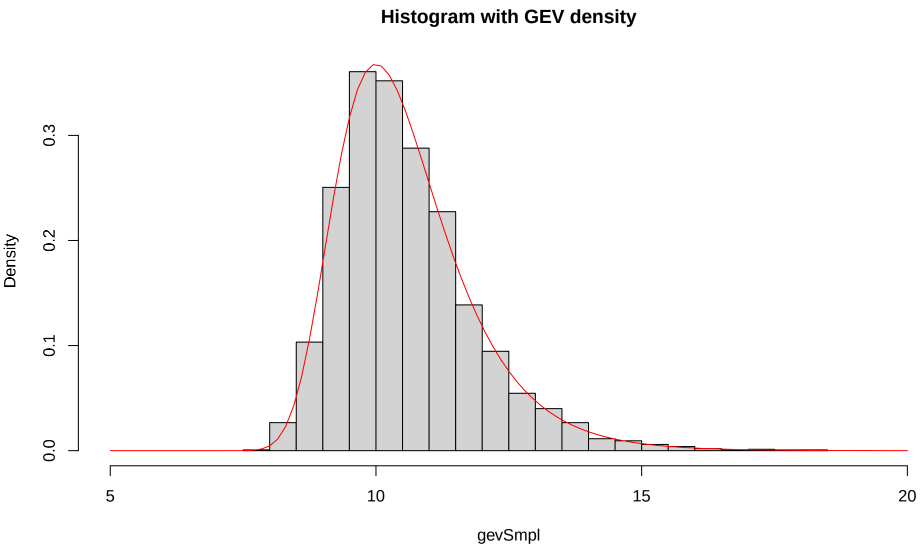
Benedikt Gräler

52°North Spatial Information Research GmbH
https://52north.org

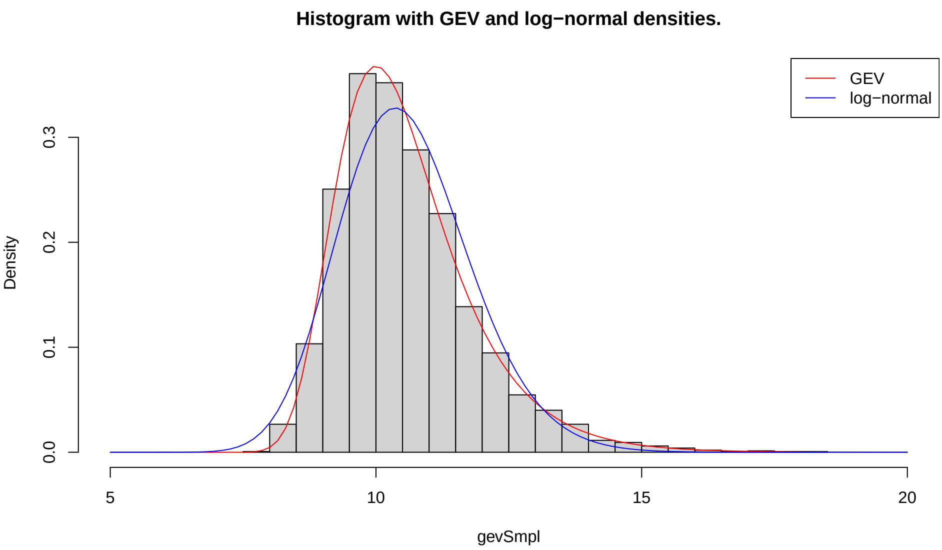
Univariate Extremes

What is special about extremes?

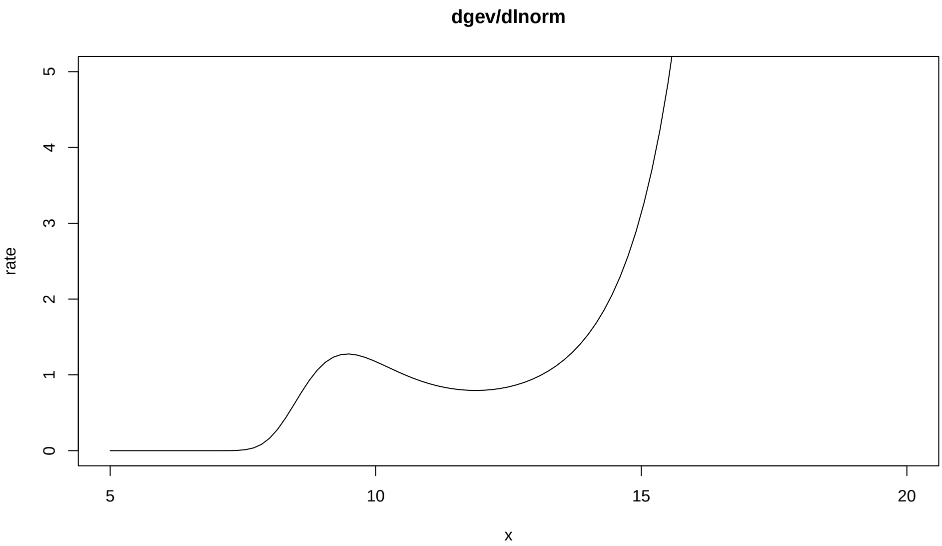
- Extremes are rare events / measurements
  - few records in a spatial distributed scene
  - few records in a time series
  - both
- by definition, we will only observe few extremes (if any) in a given sample


Histogram with GEV density
Density
0.0
0.1
0.2
0.3
5
10
15
20
gevSmpl

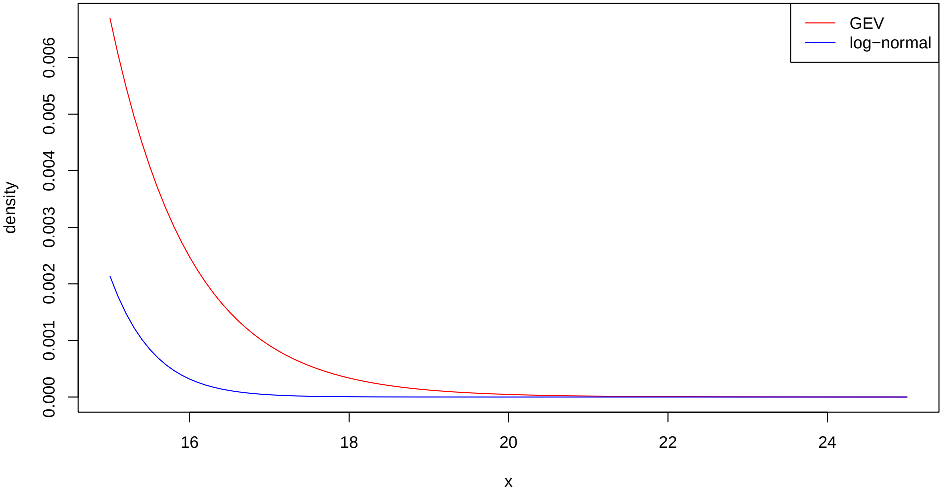
What is special about extremes?



What is special about extremes?



What is special about extremes?



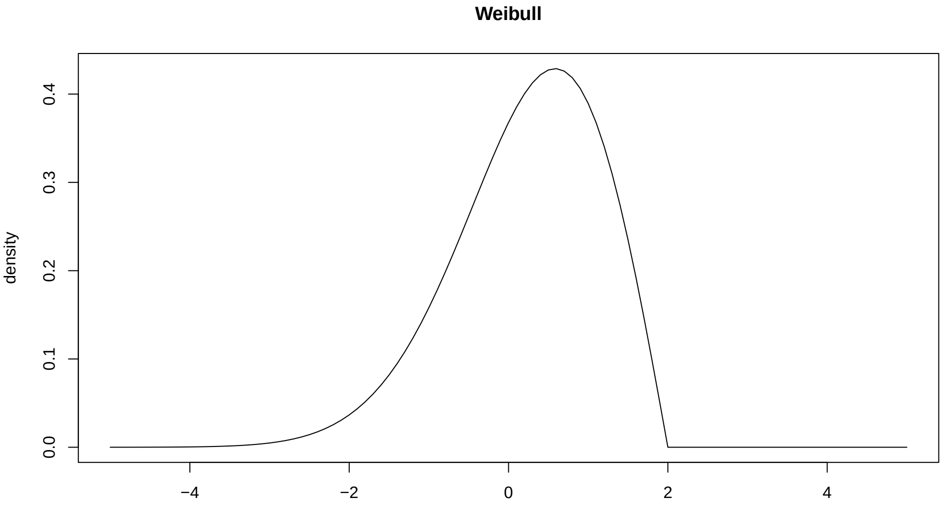
Let $X_1, \ldots, X_n$ be a sequence of independent and identically distributed random variables with cumulative distribution function F and let $M_n = \max(X_1, \ldots, X_n)$ denote the maximum.

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

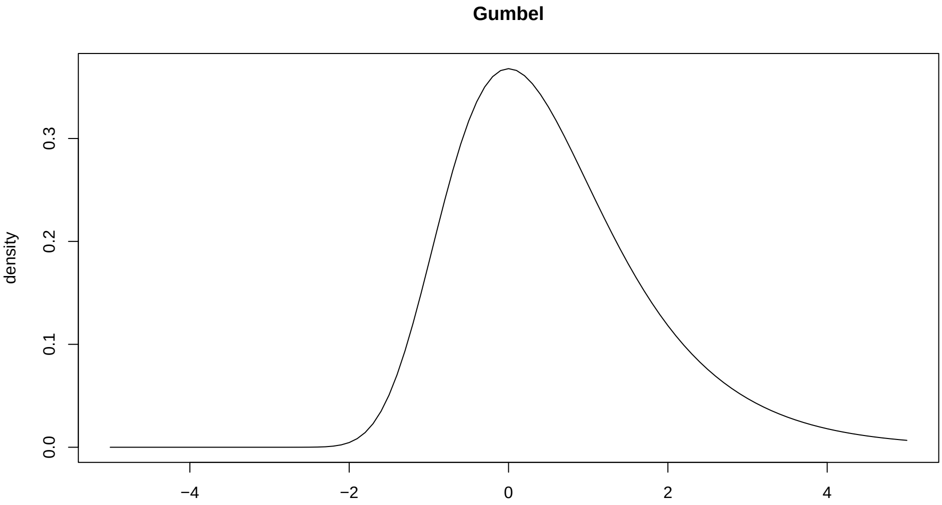
$$\begin{aligned}\Pr(M_n \le z) &= \Pr(X_1 \le z, \ldots, X_n \le z)\\ &= \Pr(X_1 \le z) \cdots \Pr(X_n \le z) = (F(z))^n.\end{aligned}$$

Source: https://en.wikipedia.org/wiki/Extreme_value_theory

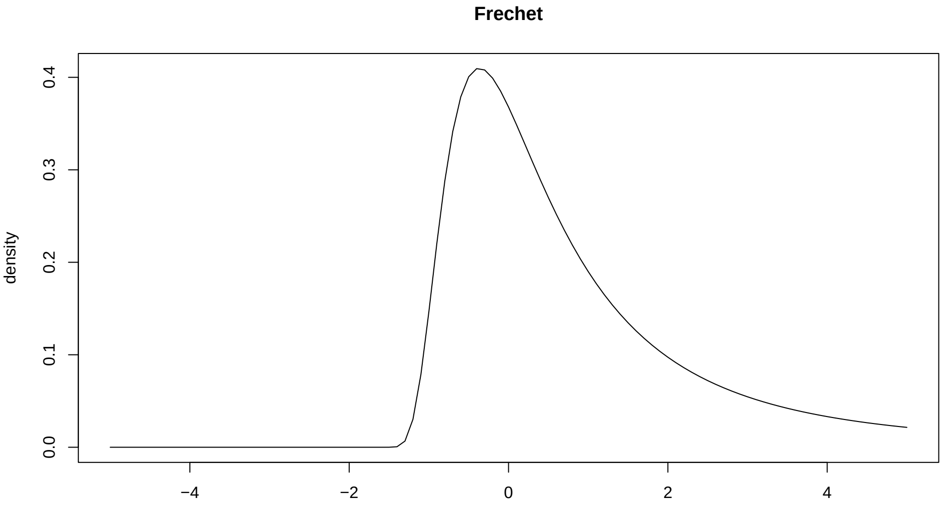
The Weibul



The Gumbel



The Fréchet



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

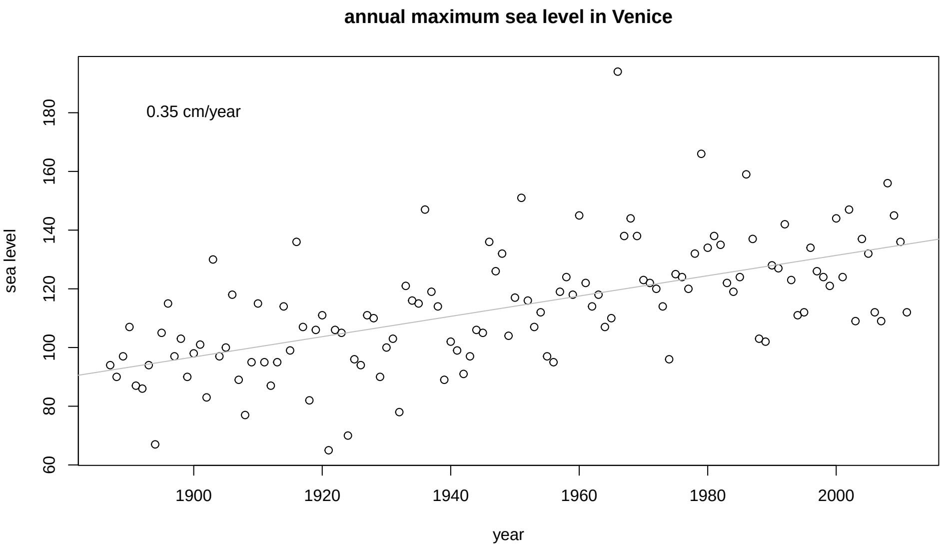
with `loc` $= a$, `scale` $= b$ and `shape` $= s$:

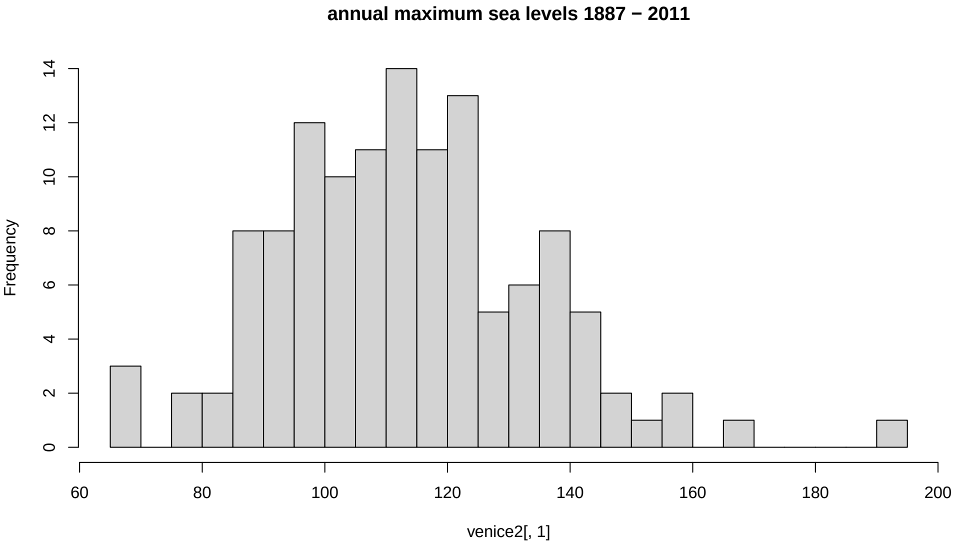
```
dgev(x, loc=0, scale=1, shape=0, log = FALSE)
pgev(q, loc=0, scale=1, shape=0, lower.tail = TRUE)
qgev(p, loc=0, scale=1, shape=0, lower.tail = TRUE)
rgev(n, loc=0, scale=1, shape=0)
```

- 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

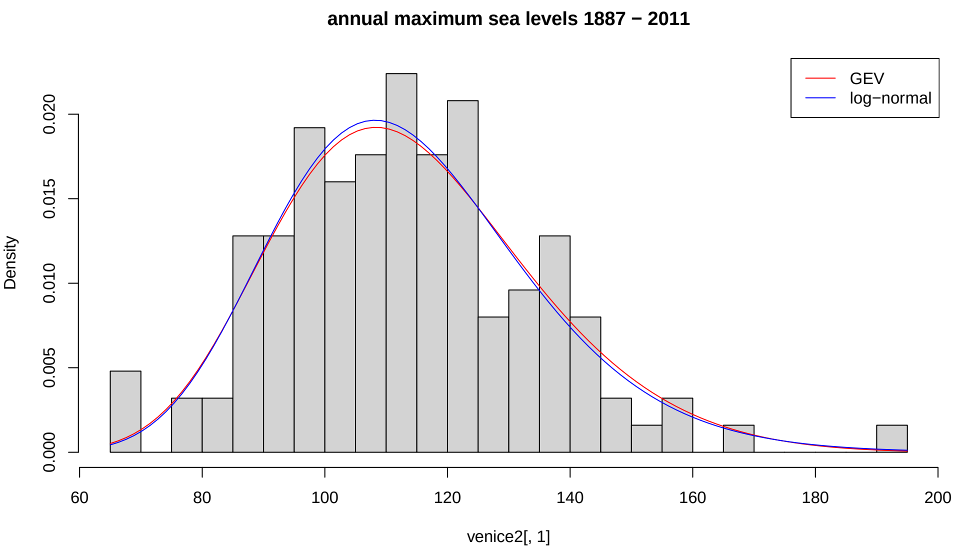
- month
- annually
- daily
- ...


annual maximum sea level in Venice
0.35 cm/year
sea level
60
80
100
120
140
160
180
1900
1920
1940
1960
1980
2000
year

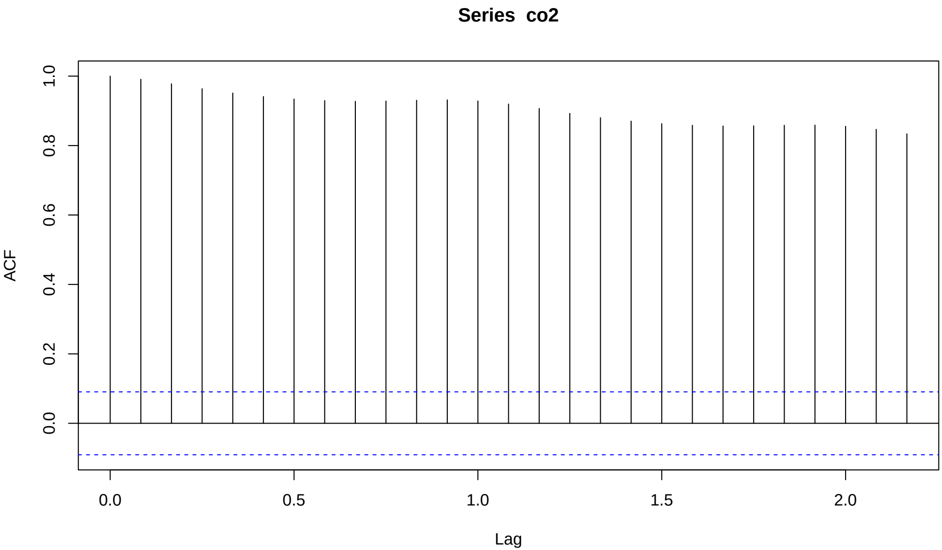

annual maximum sea levels 1887 – 2011
Frequency
0
2
4
6
8
10
12
14
60
80
100
120
140
160
180
200
venice2[, 1]

```r
fgev(venice2[,1])
```

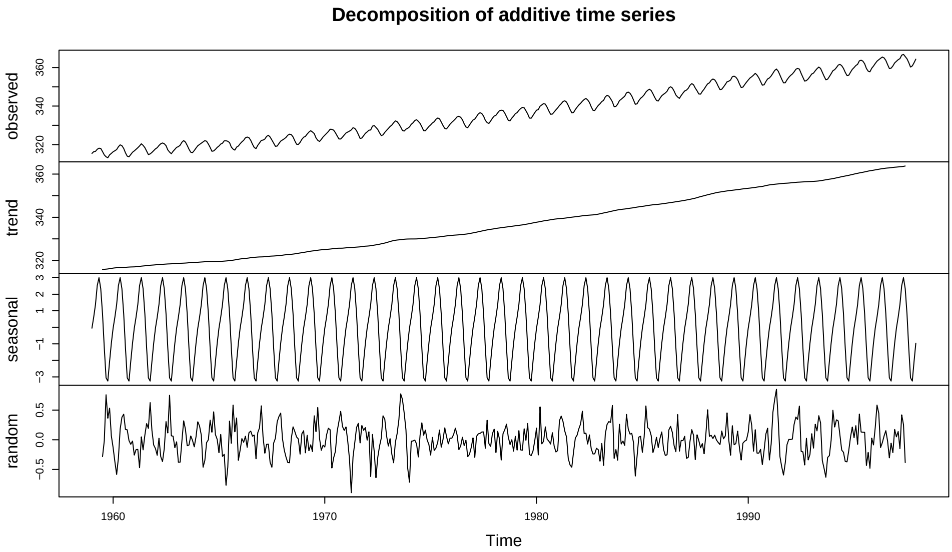
```
##
## Call: fgev(x = venice2[, 1])
## Deviance: 1111.223
##
## Estimates
##       loc     scale     shape
## 105.2995   19.3543   -0.1463
##
## Standard Errors
##      loc    scale    shape
## 1.87769  1.27804  0.04176
##
## Optimization Information
##   Convergence: successful
##   Function Evaluations: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


annual maximum sea levels 1887 − 2011
GEV
log−normal
Density
0.000
0.005
0.010
0.015
0.020
60
80
100
120
140
160
180
200
venice2[, 1]

PoT method

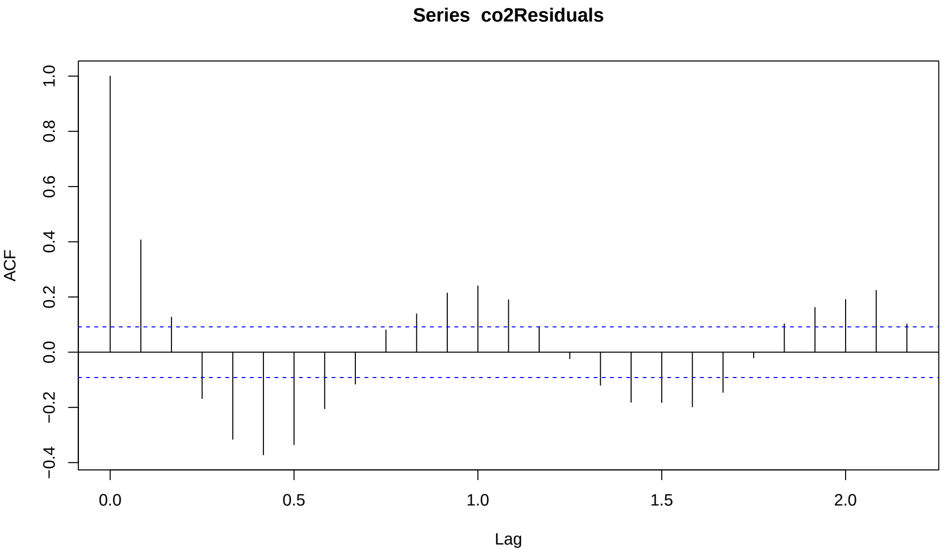


PoT method



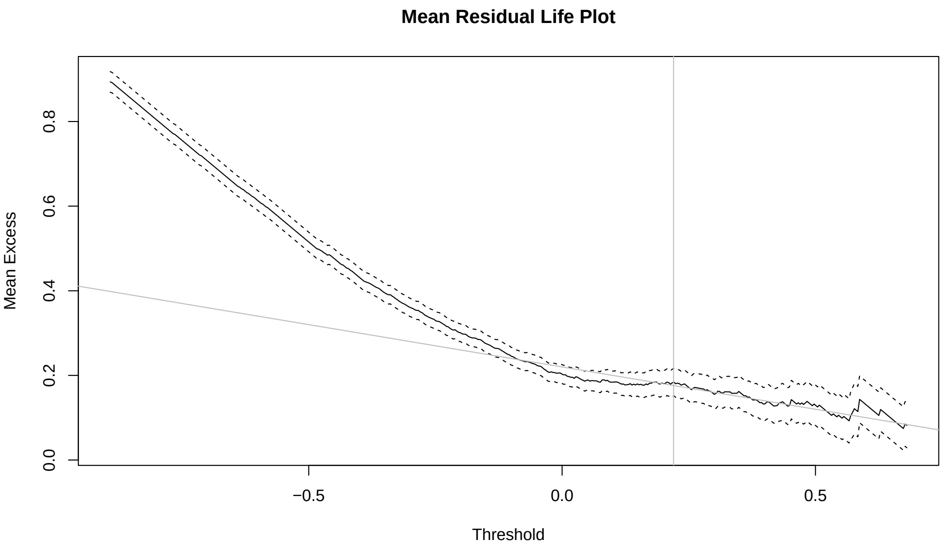
Decomposition of additive time series
observed
trend
seasonal
random
360
340
320
3
2
1
−1
−3
0.5
0.0
−0.5
1960
1970
1980
1990
Time

PoT method



Series co2Residuals
ACF
Lag
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0.0
0.5
1.0
1.5
2.0

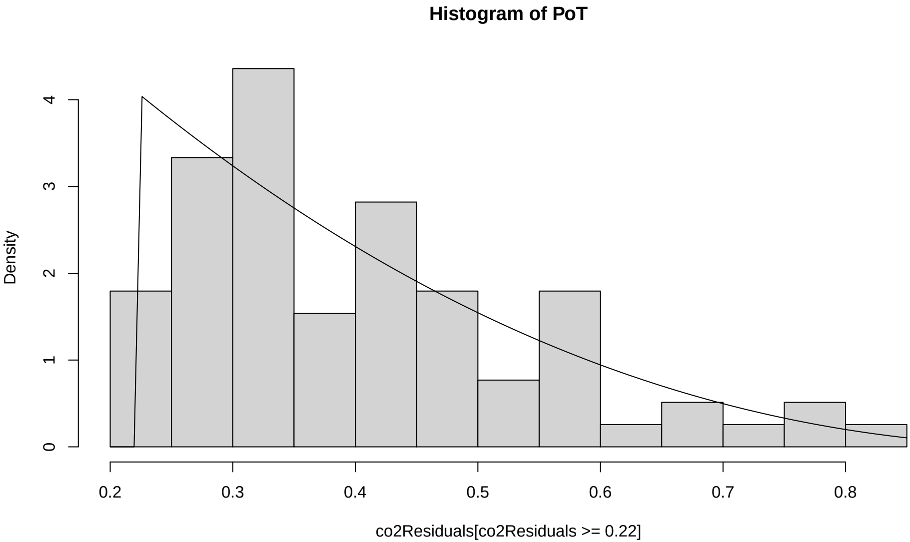
PoT method


Mean Residual Life Plot
Mean Excess
0.0
0.2
0.4
0.6
0.8
−0.5
0.0
0.5
Threshold

PoT method

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

```
##
## Call: fpot(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
```

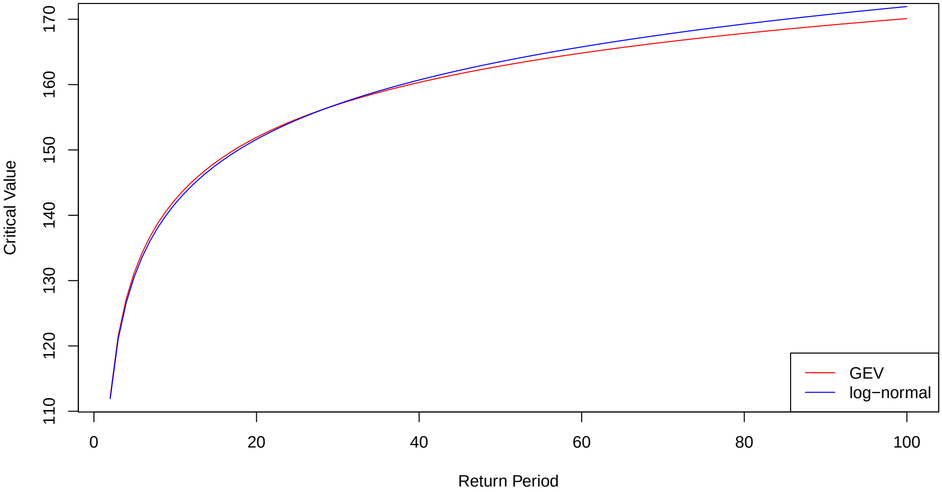

Histogram of PoT
Density
0
1
2
3
4
0.2
0.3
0.4
0.5
0.6
0.7
0.8
co2Residuals[co2Residuals >= 0.22]

What is an retrun period?

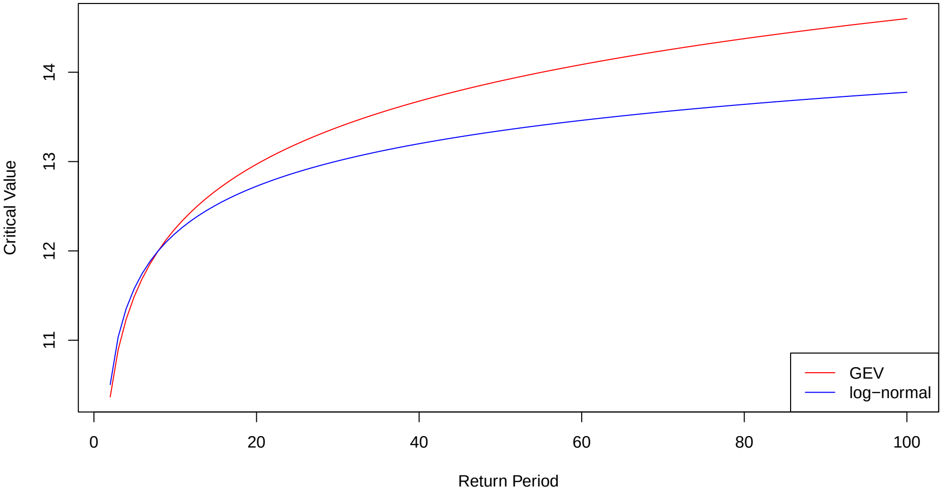
"On average, how many years will it take to observe an event this large/small?"

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 retrun period, what is the annual maximum water level in Venice?



Return Period - simulated example



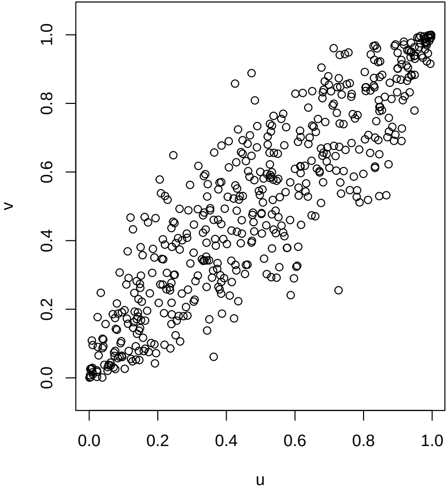
Multivariate Extremes

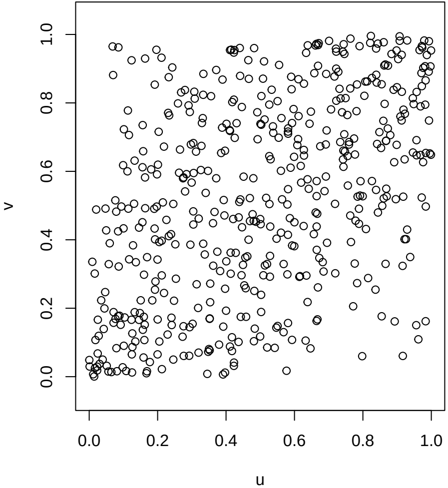
- Once we look at more than one variable, we have to think about dependencies.
- Correlation measures summarize dependence in a single number, but as for univariate statistics, a multivariate distribution has more characteristics than a single number
- while histograms are often the first step for univariate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

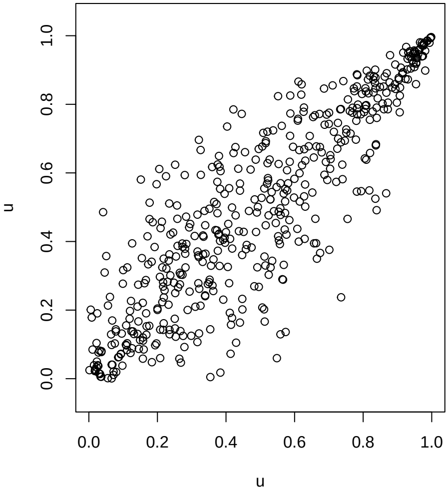
$$H(x_1, \dots, x_n) = C\left(F_1(x_1), \dots, F_n(x_n)\right)$$

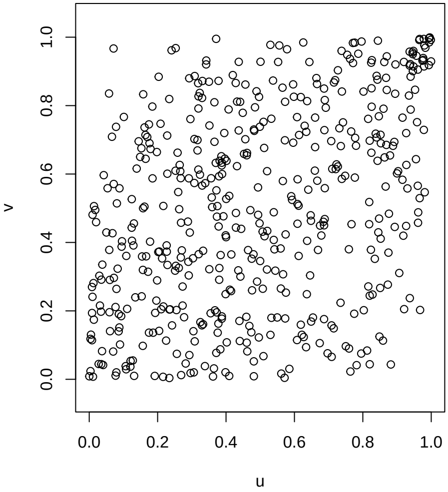
A copula can be perceived as a multivariate distribution on the unit hyper-cube $[0, 1]^n$.

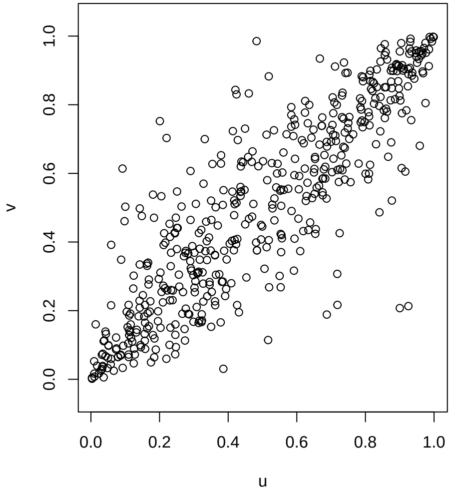

v
u
0.0
0.2
0.4
0.6
0.8
1.0

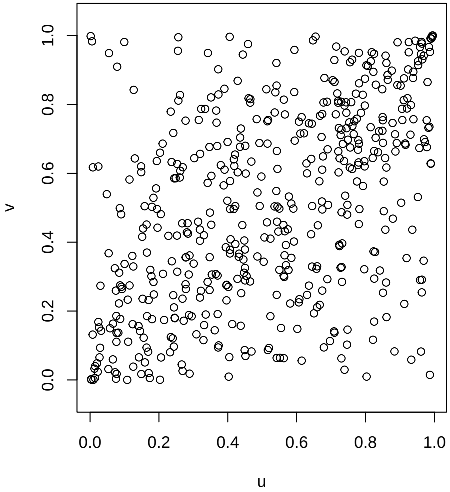

v
u
0.0
0.2
0.4
0.6
0.8
1.0

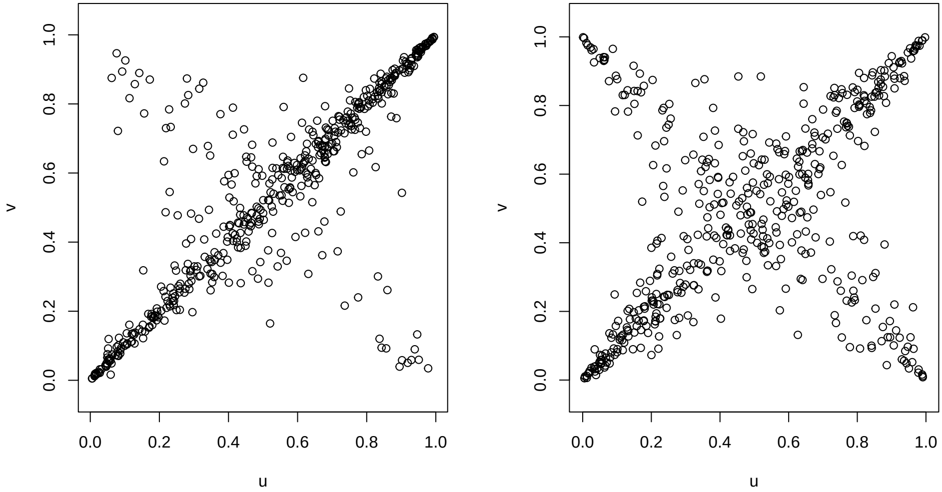
Copulas - Gumbel






v
0.0
0.2
0.4
0.6
0.8
1.0
u
0.0
0.2
0.4
0.6
0.8
1.0


v
0.0
0.2
0.4
0.6
0.8
1.0
u
0.0
0.2
0.4
0.6
0.8
1.0



v
0.0
0.2
0.4
0.6
0.8
1.0
u
0.0
0.2
0.4
0.6
0.8
1.0
v
0.0
0.2
0.4
0.6
0.8
1.0
u
0.0
0.2
0.4
0.6
0.8
1.0

https://copulatheque.org

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

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

```
## lower upper
##     0     0
```

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

```
##     lower     upper
## 0.0000000 0.7688556
```

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

```
##     lower     upper
## 0.6144008 0.6144008
```

```r
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
```

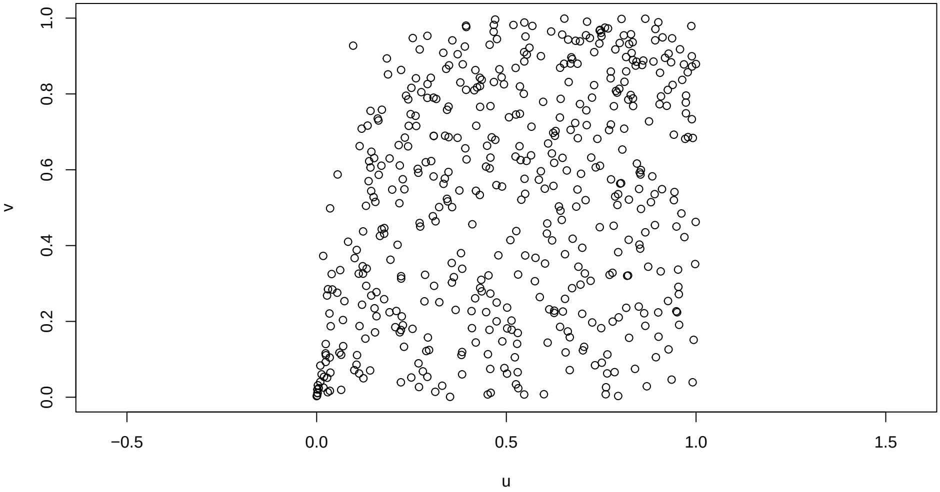
Fitting a Copula

```
aqPseudoObs <- pobs(cbind(airquality$Solar.R,
                          airquality$Temp))
BiCopSelect(aqPseudoObs[,1], aqPseudoObs[,2])
```

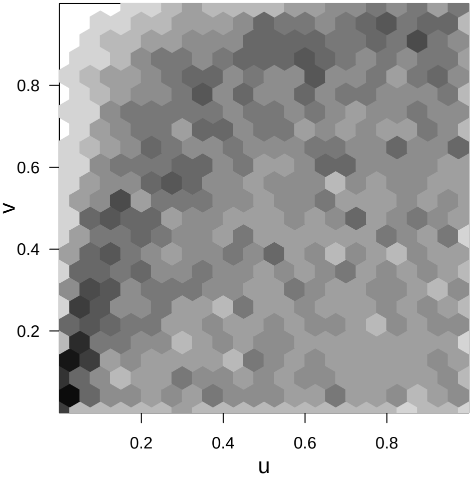
```
## Warning: In BiCopSelect: 7 observations (4.6%) contain NAs. Only complete
## observations are used.

## Bivariate copula: Rotated Tawn type 2 180 degrees (par = 2.46, par2 = 0.2, tau = 0.16)
```

Fitting a Copula



Fitting a Copula



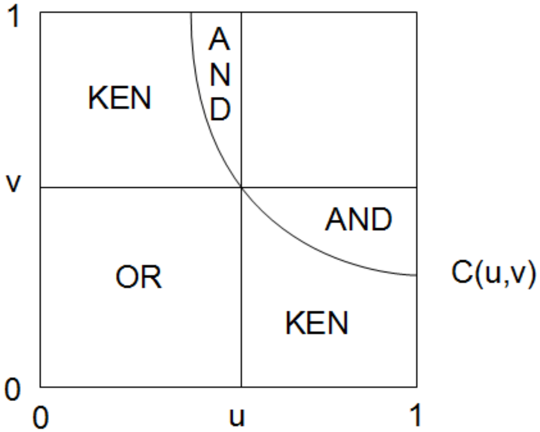
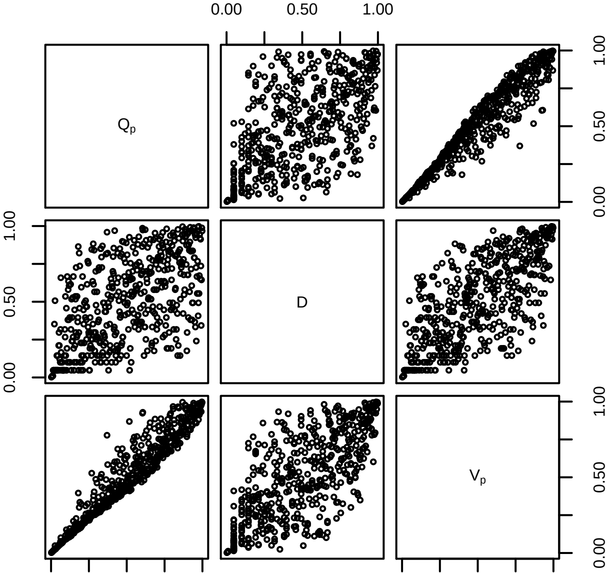
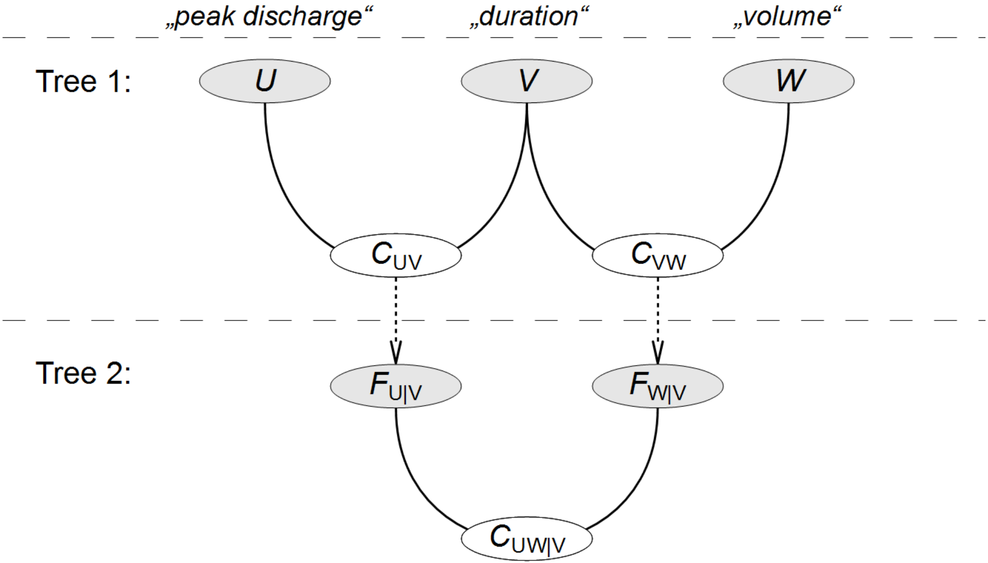


Figure 1: alt text here


Qp
D
Vp
0.00
0.50
1.00


„peak discharge“
„duration“
„volume“
Tree 1:
U
V
W
C_{UV}
C_{VW}
Tree 2:
$F_{U|V}$
$F_{W|V}$
$C_{UW|V}$

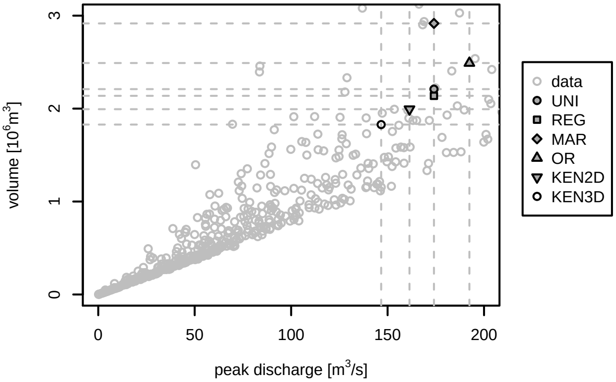
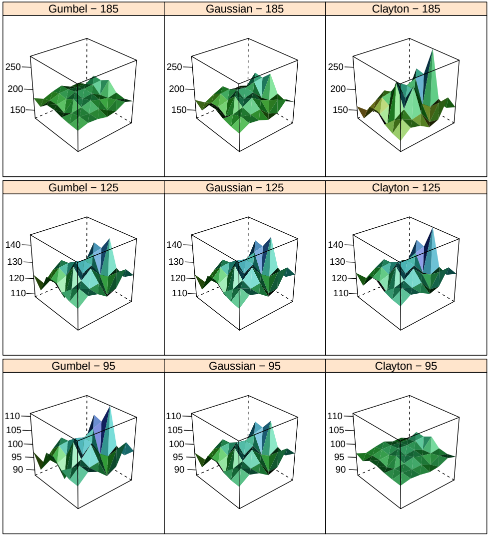


Figure 4: comparison of critical events

Spatial Extremes

Does that make any difference?



Spatial Copulas - Margins

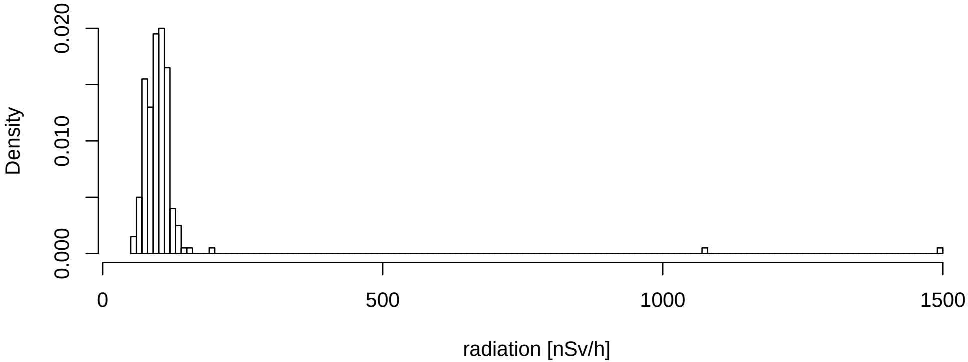


Figure 5: Marignal Distribution

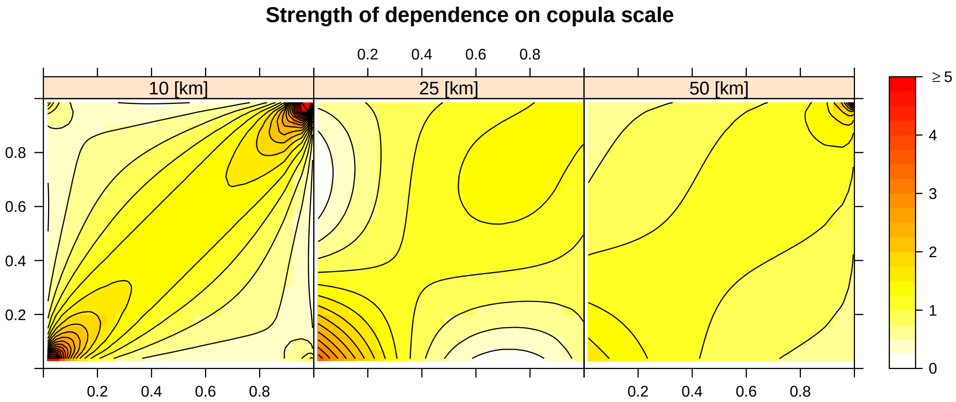


Figure 6: Spatially varying copulas

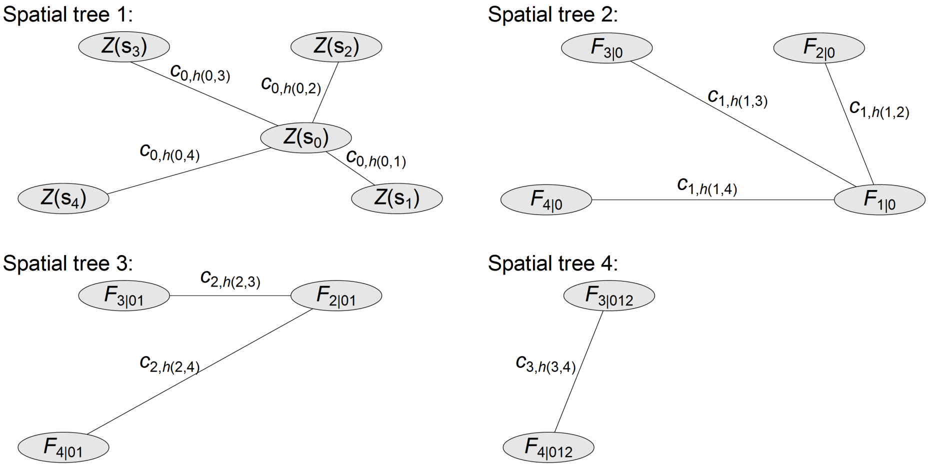
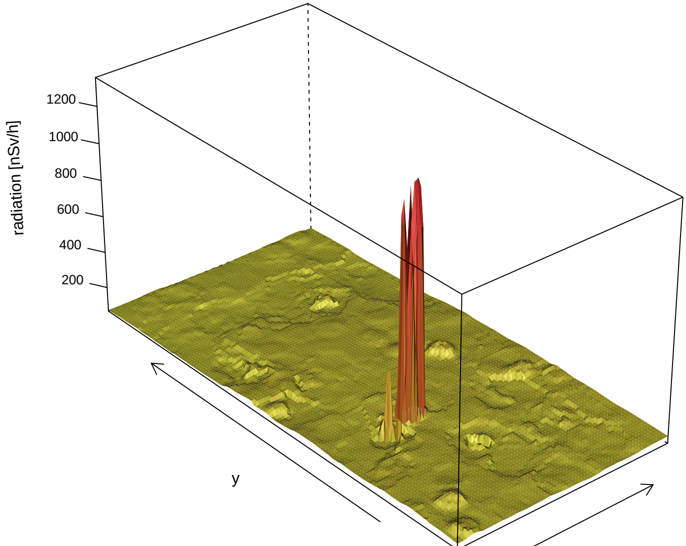


Figure 7: Spatial Vine

Spatial Copulas Interpolation



Spatail Copula - Comparison

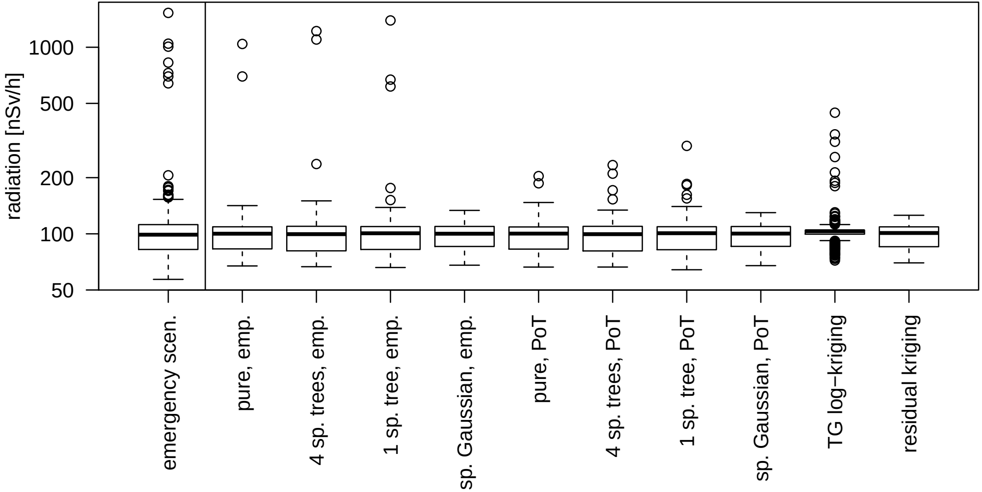
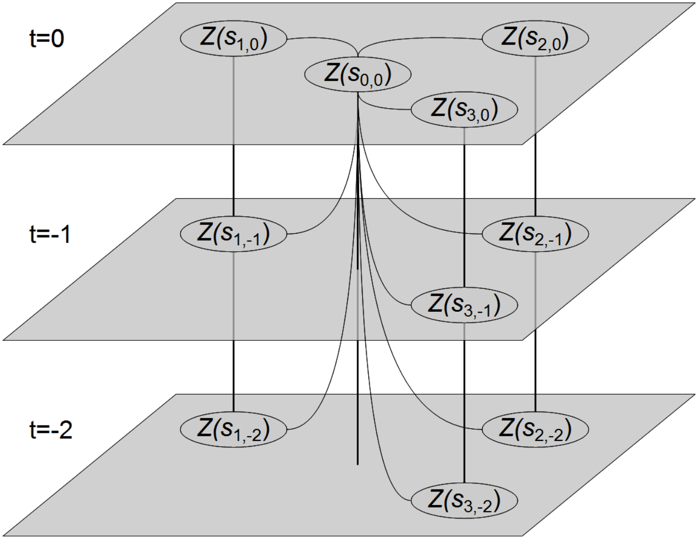


Figure 9: different approaches

Spatio-Temporal Copulas



Hands-on

Libraries to look at

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

Data sets to try and demos to follow

- 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