NCAR | RESEARCH APPLICATIONS LABORATORY

NCAR is sponsored by National Science Foundation

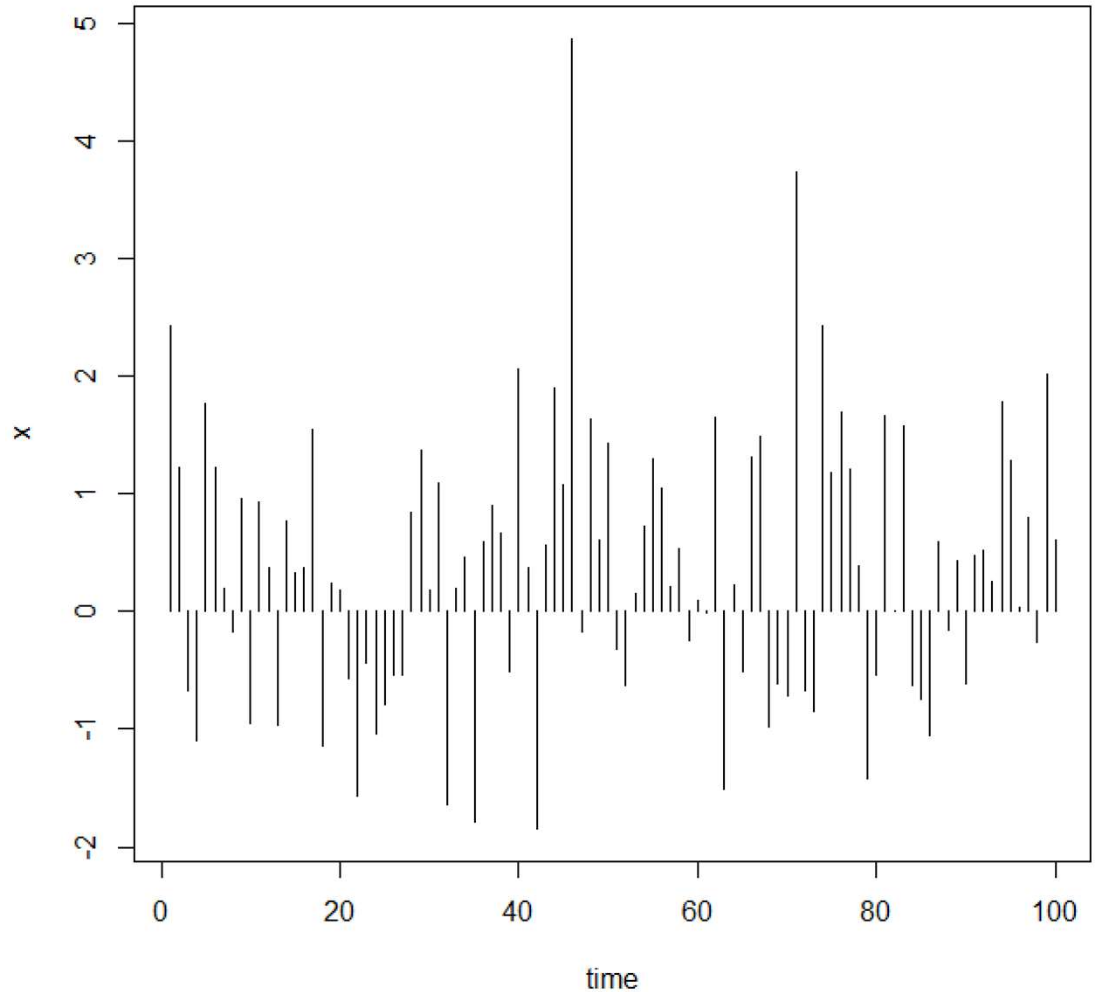
Earth System Extremes Affinity Group

extRemes: An R package for statistical extreme-value analysis (EVA)

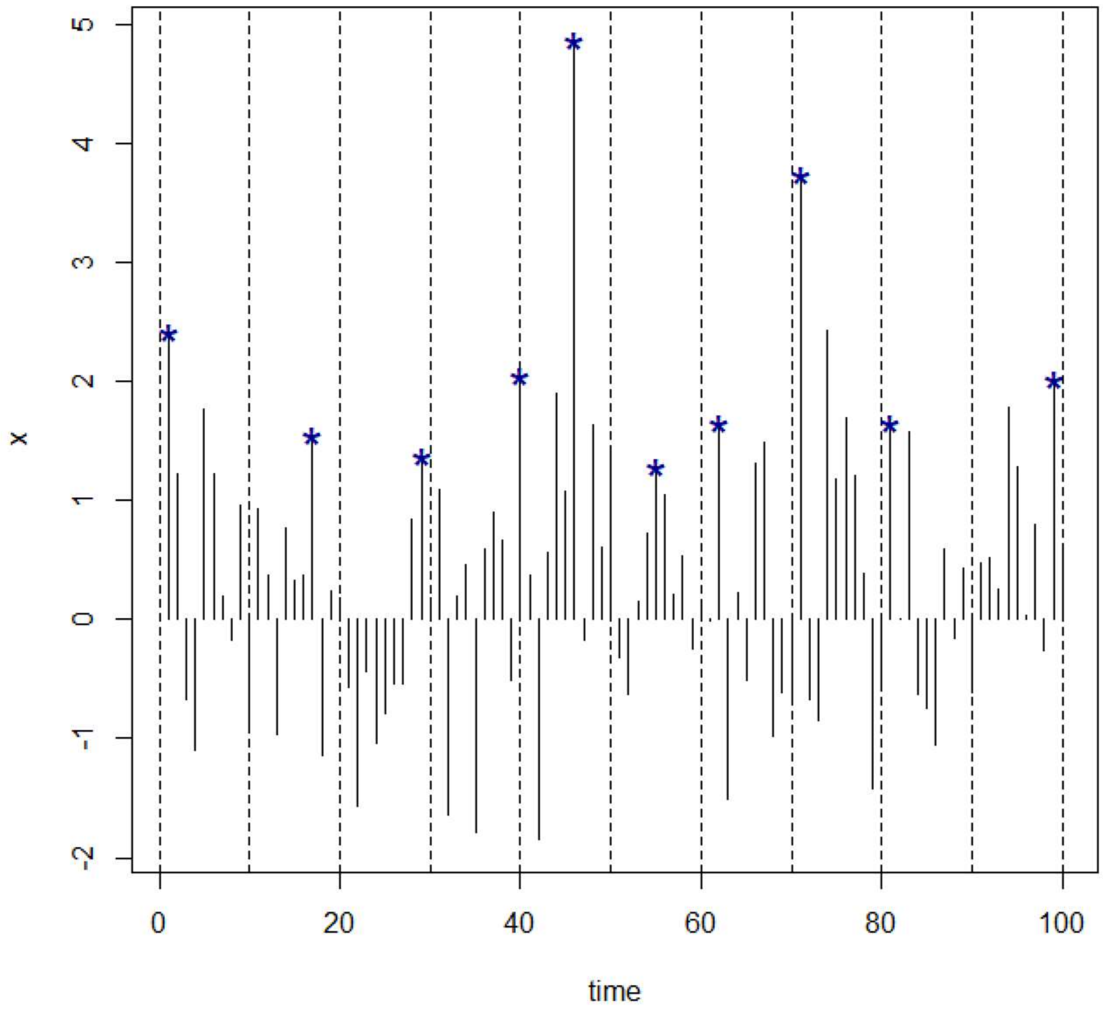
Eric Gilleland

7 April 2022

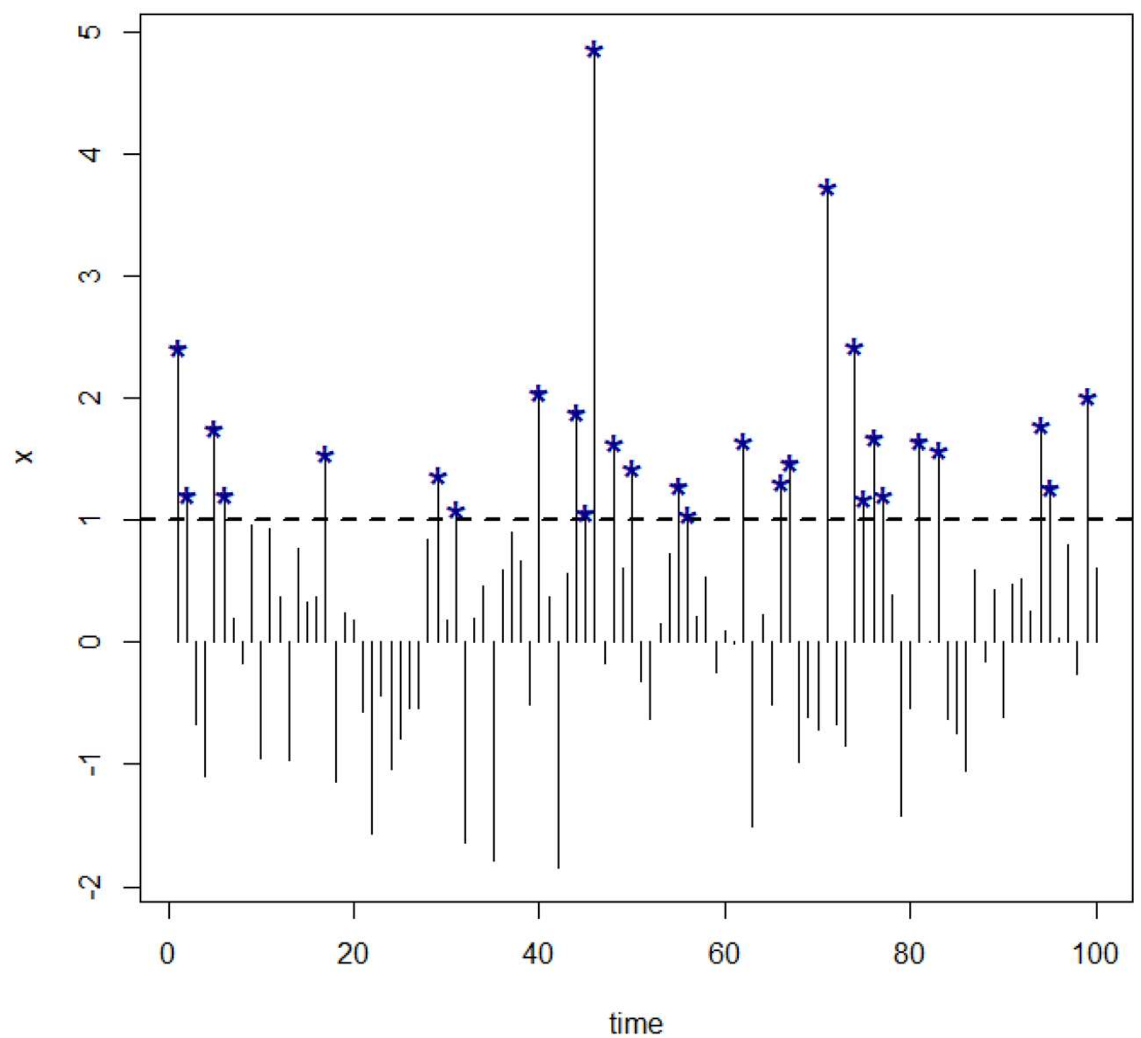
What is extreme?



What is extreme?



What is extreme?



What is extreme?

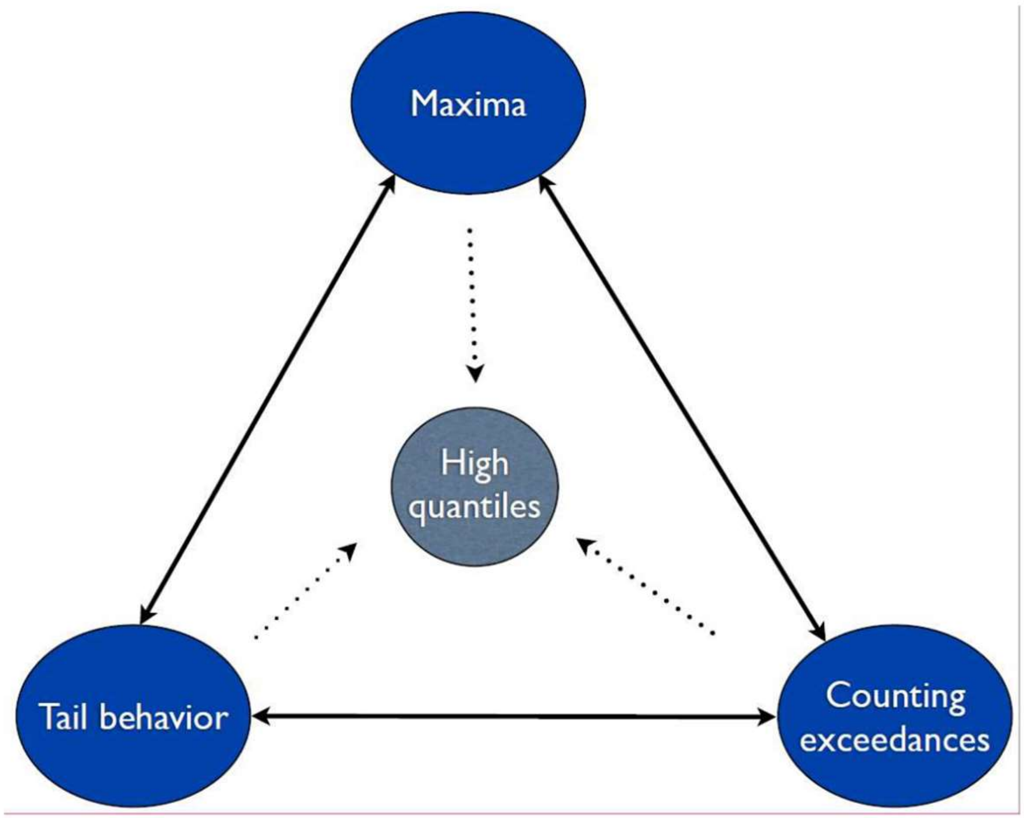


Figure from Philippe Naveau

Inference for the Maximum

- Only one maximum in a data set, but are we sure we've already recorded it? How much larger could it get?
- Let $X_1, \dots, X_n$ be iid with common (cumulative) distribution F, and let $M_n = \max\{X_1, \dots, X_n\}$.
- $\mathbb{P}[M_n \le z] = \mathbb{P}[X_1 \le z, \dots, X_n \le z] = \prod_{i=1}^{n} \mathbb{P}[X_i \le z] = F^n(z)$.

Problematic for large n!

Sum Stability

$$\frac{2+7+14+8+10+22}{6} = \frac{63}{6}$$

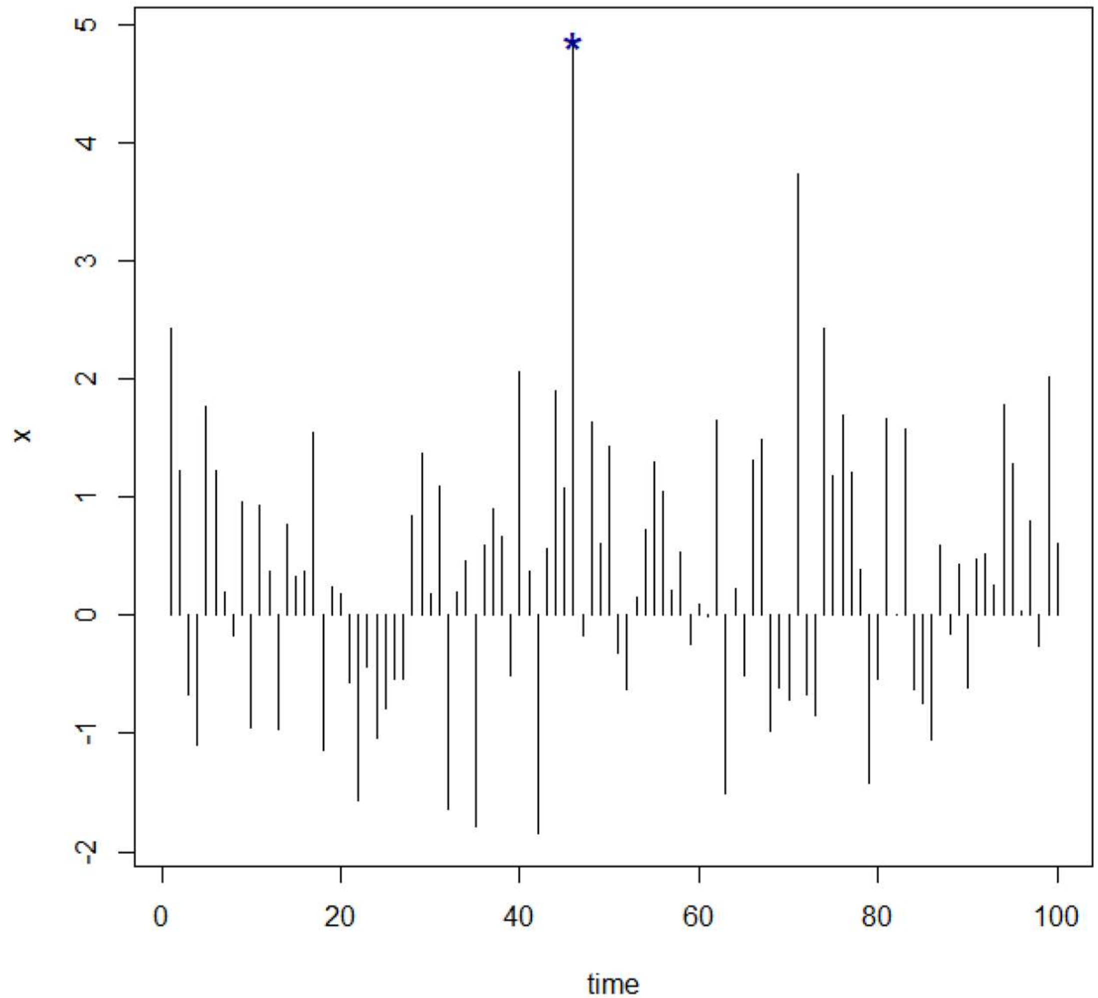
$$\frac{1}{2} \cdot \left(\frac{2+7+14}{3} + \frac{8+10+22}{3}\right) = \frac{1}{2} \cdot \left(\frac{23}{3} + \frac{40}{3}\right) = \frac{63}{6}$$

The two averages are the same. The first averages six numbers while the second averages only two.

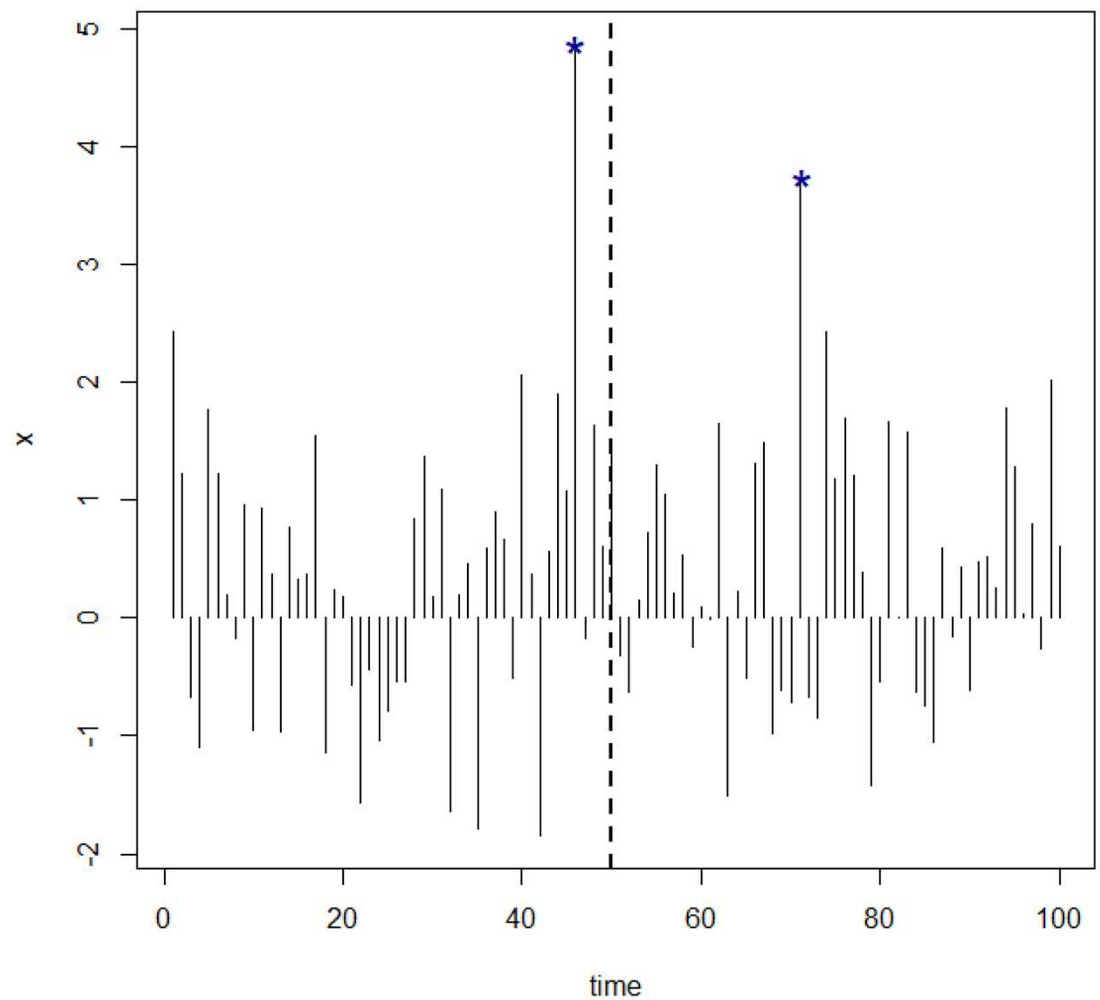
Let $X_1, X_2, \dots, X_n$ be iid random variables. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and let $\bar{Y}_m = \frac{1}{mk}\sum_{j=1}^{m} Y_j$, where $Y_1 = X_1 + X_2 + \cdots + X_k, Y_2 = X_{k+1} + X_{k+2} + \cdots + X_{2k}, \dots, Y_m = X_{n-k+1} + \cdots + X_n$.

The distribution of $\bar{X}_n$ should be the same as that of $\bar{Y}_m$. And sum stability says that it is! In this case, the normal distribution is sum-stable (among others). For example, if $X_i \sim N(\mu, \sigma^2)$ for all $i = 1, \dots, n$, then $\sum_i^n X_i \sim N(n\mu, n\sigma^2)$. Note the distribution is the same but for some re-scaling.

Max Stability



Max Stability



Max Stability

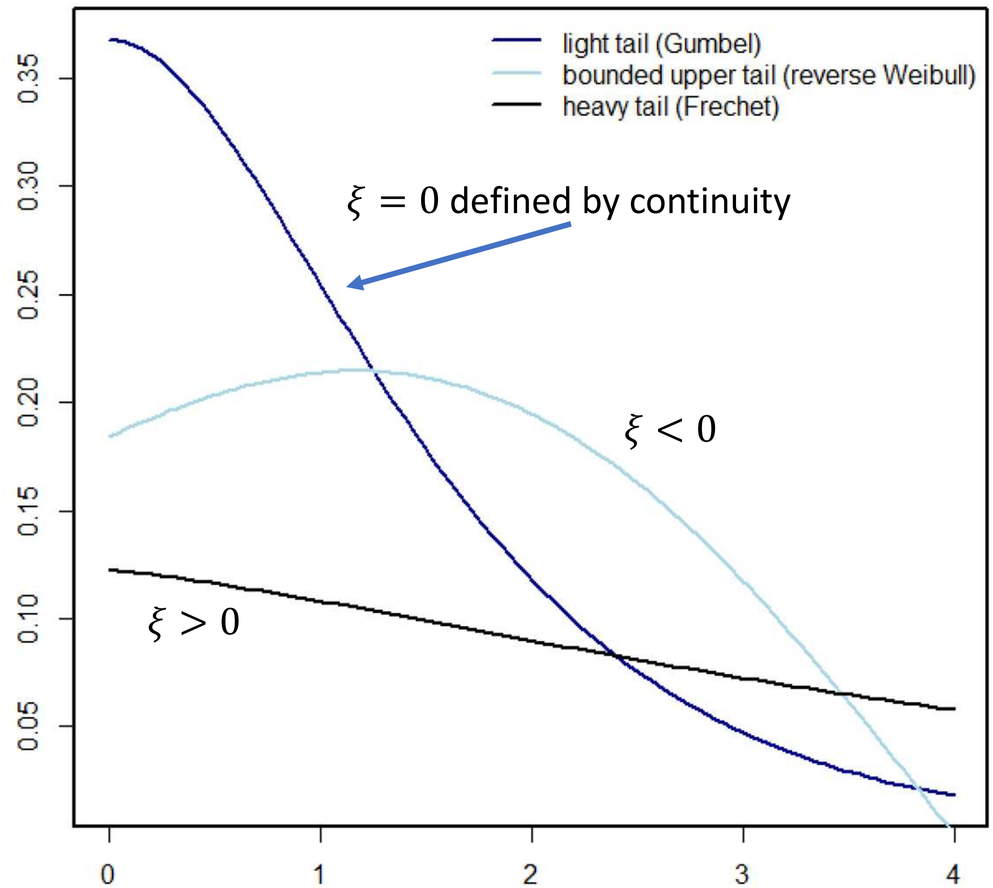
Similar to the sum-stable case,
$\max\{X_1, \dots, X_n\} = \max\left\{\max\left\{X_1, \dots, X_{\left\lfloor \frac{n}{2} \right\rfloor}\right\}, \max\left\{X_{\left\lfloor \frac{n}{2} \right\rfloor + 1}\right\}\right\}$ and so on.

So, apart from some re-scaling, we seek a distribution G such that $G(a_n z + b_n) = G^n(z)$ for some sequences of constants $a_n > 0$ and b_n.

Such a distribution is said to be max-stable, and there is only one (family) that fits the bill! The generalized extreme-value distribution (GEV).

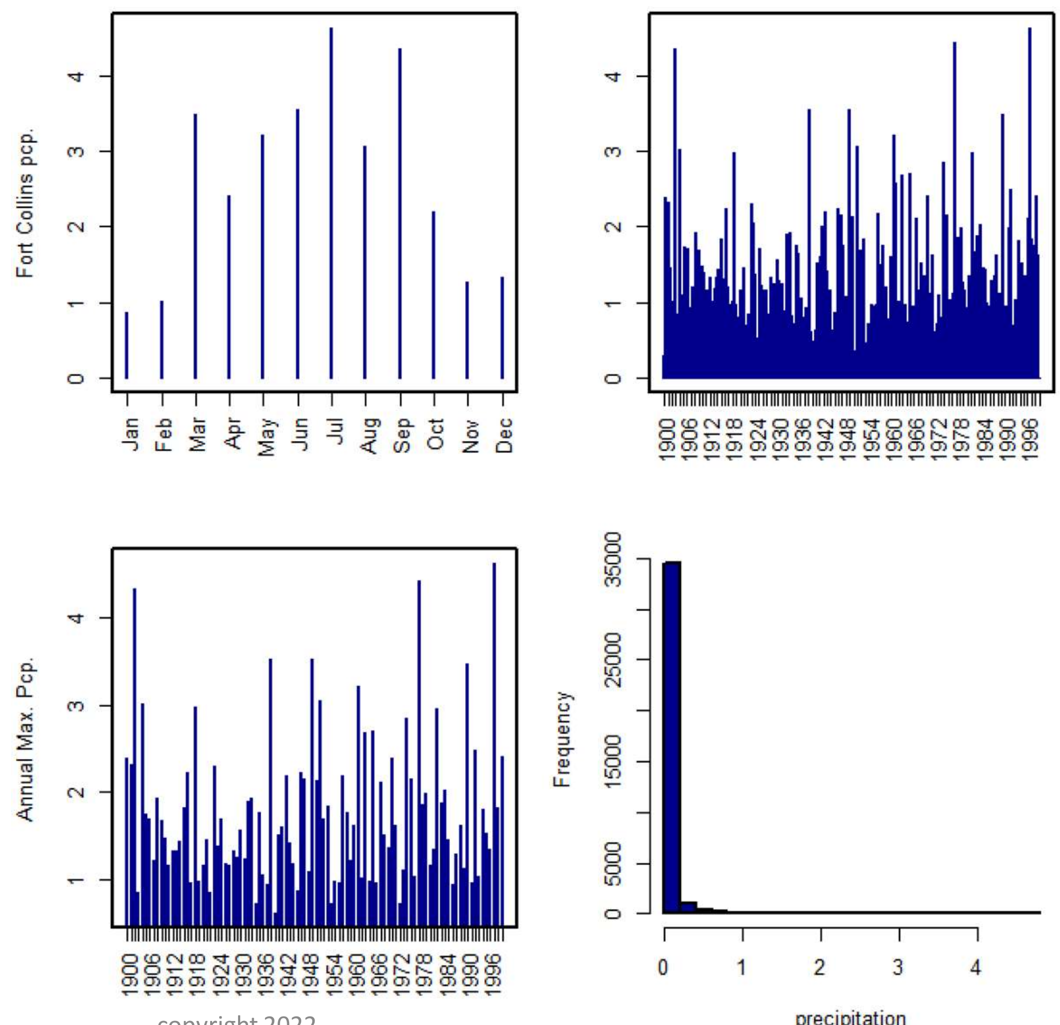
$$G(z) = \exp\left\{-\left[1 + \frac{\xi}{\sigma}(z - \mu)\right]_+^{-\frac{1}{\xi}}\right\}$$

Max Stability

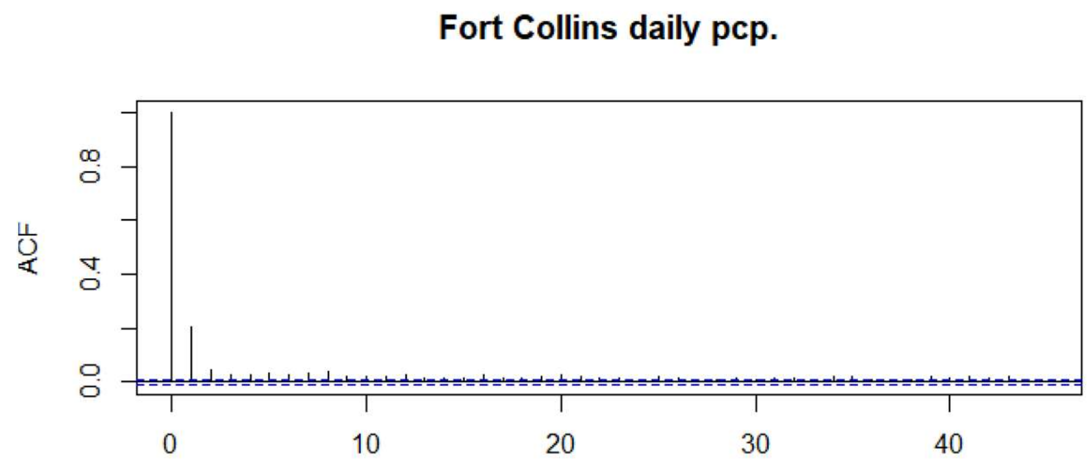


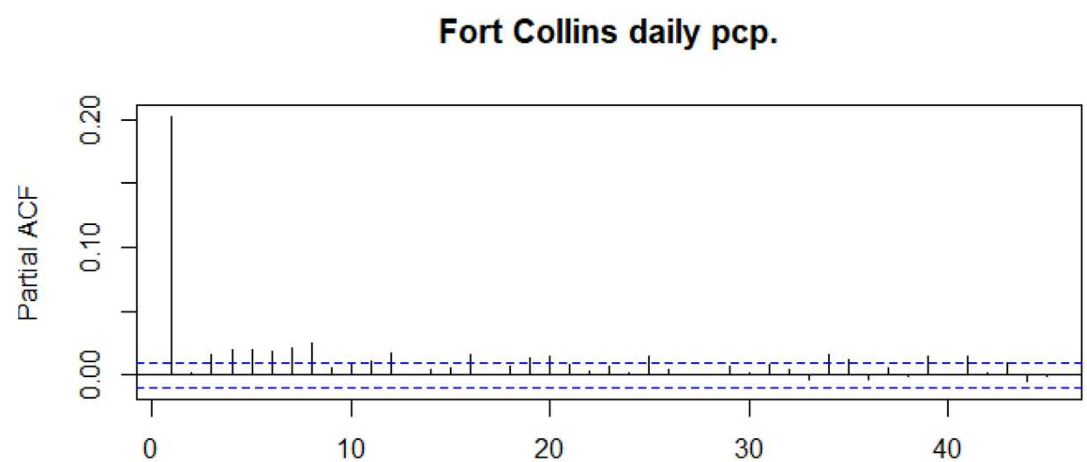
extRemes: An R software package

```
data( "Fort" )
```



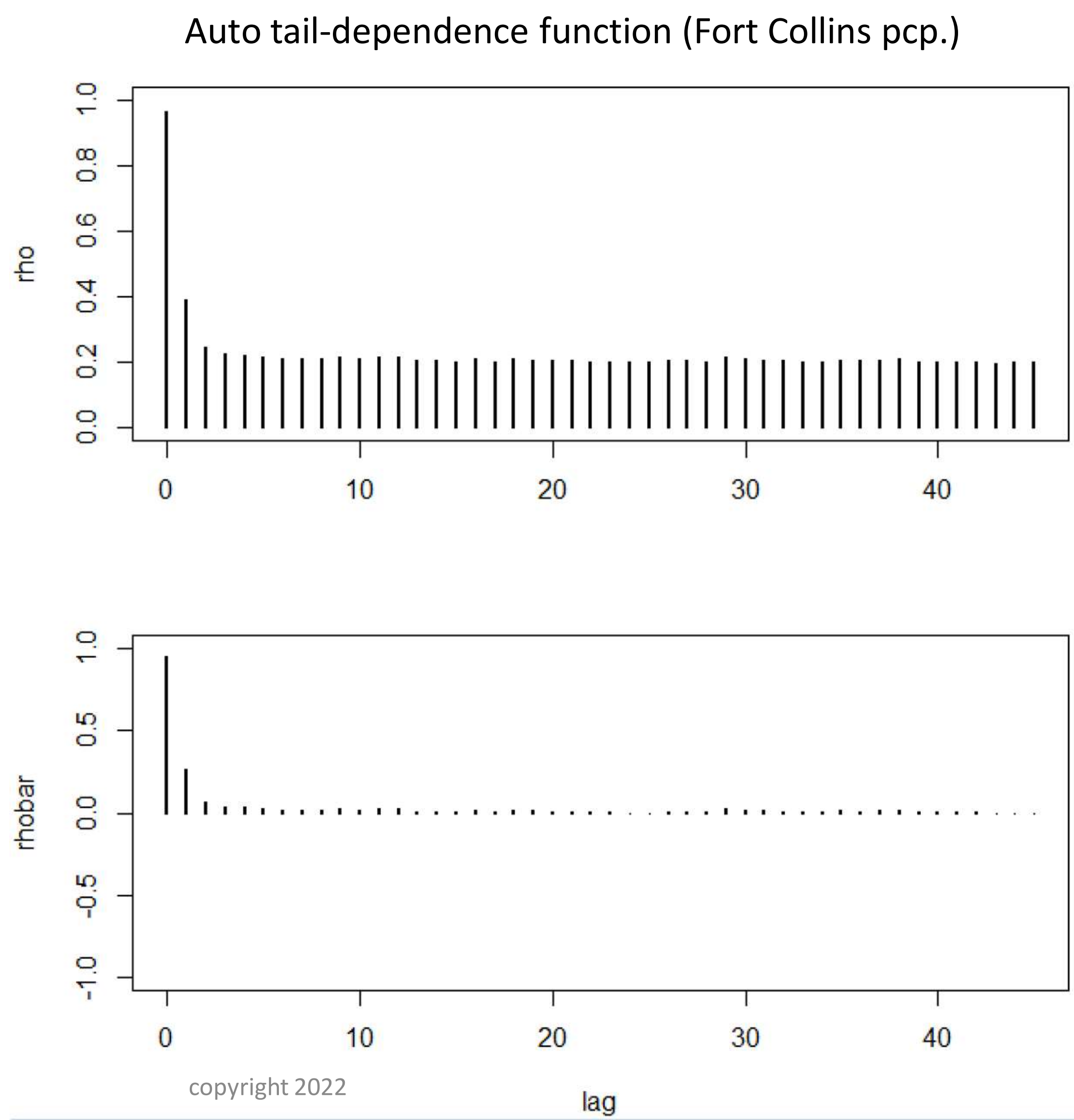
extRemes: An R software package





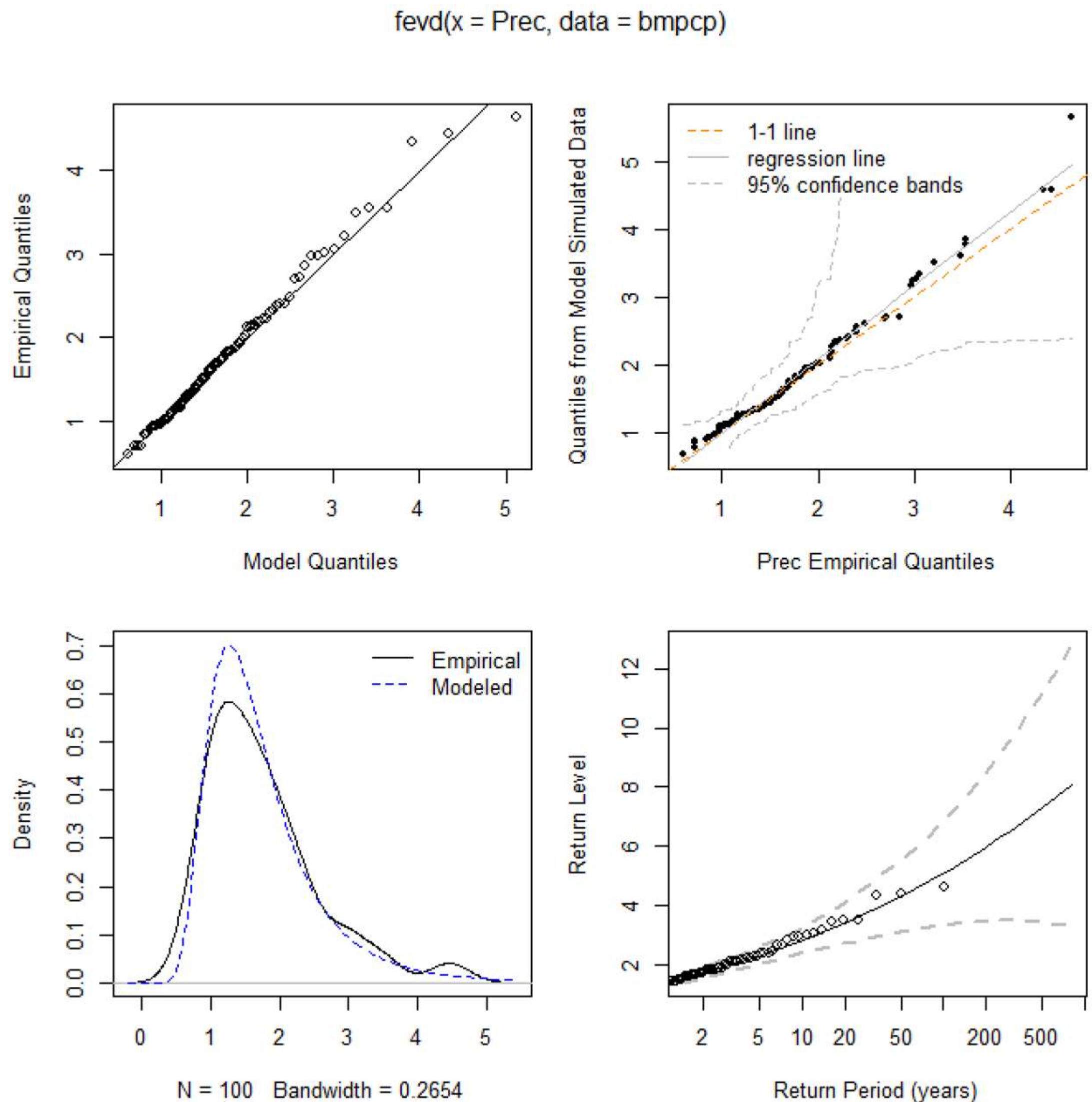
```
atdf( Fort$Prec, u = 0.8 )
```

extRemes: An R software package



```
bmpcp <- blockmaxxer(
    Fort, which = "Prec",
    blocks = Fort$year )

fit0 <- fevd( Prec,
    data = bmpcp )
fit0
plot( fit0 )
```

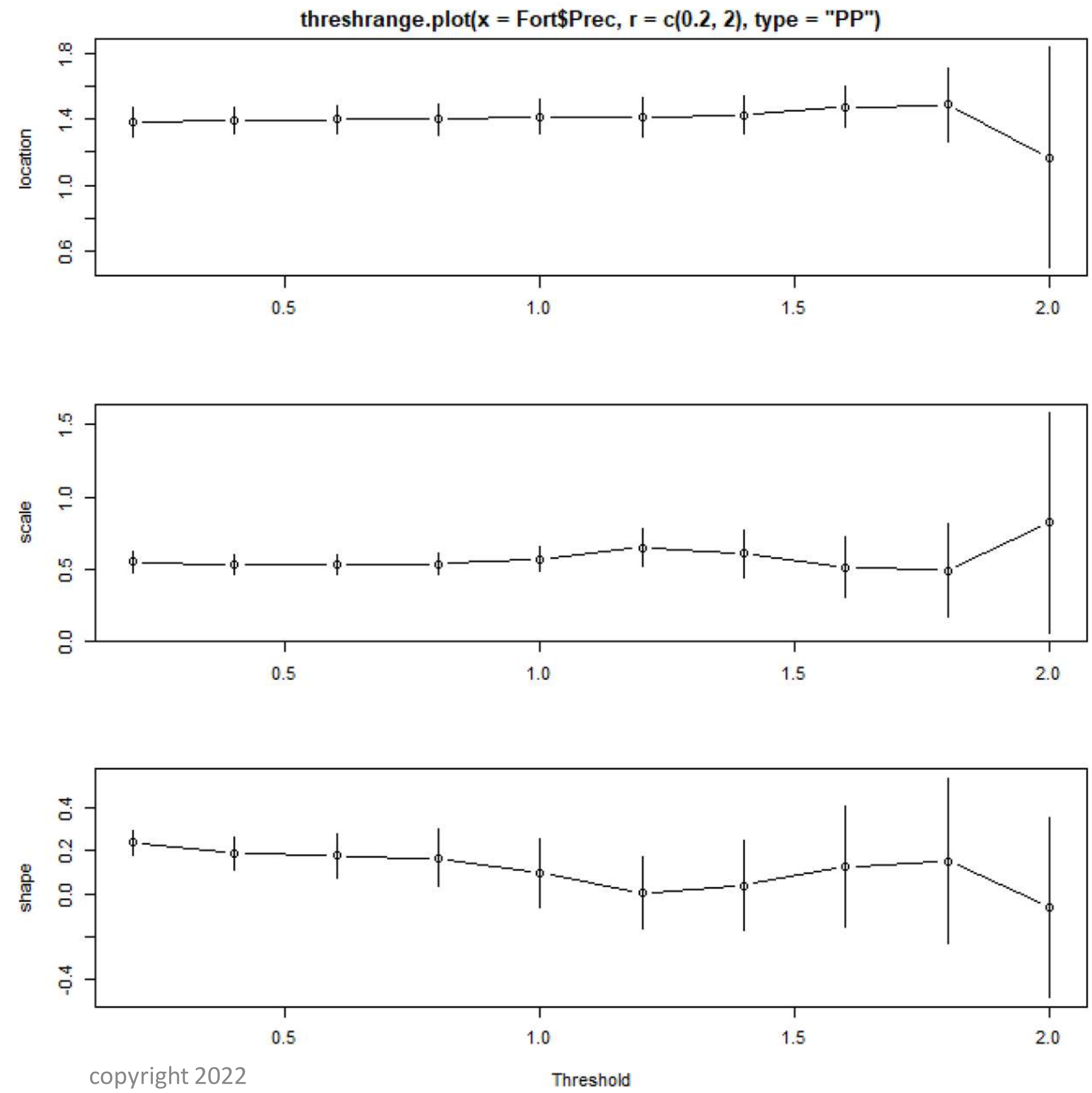
extRemes: An R software package



extRemes: An R software package

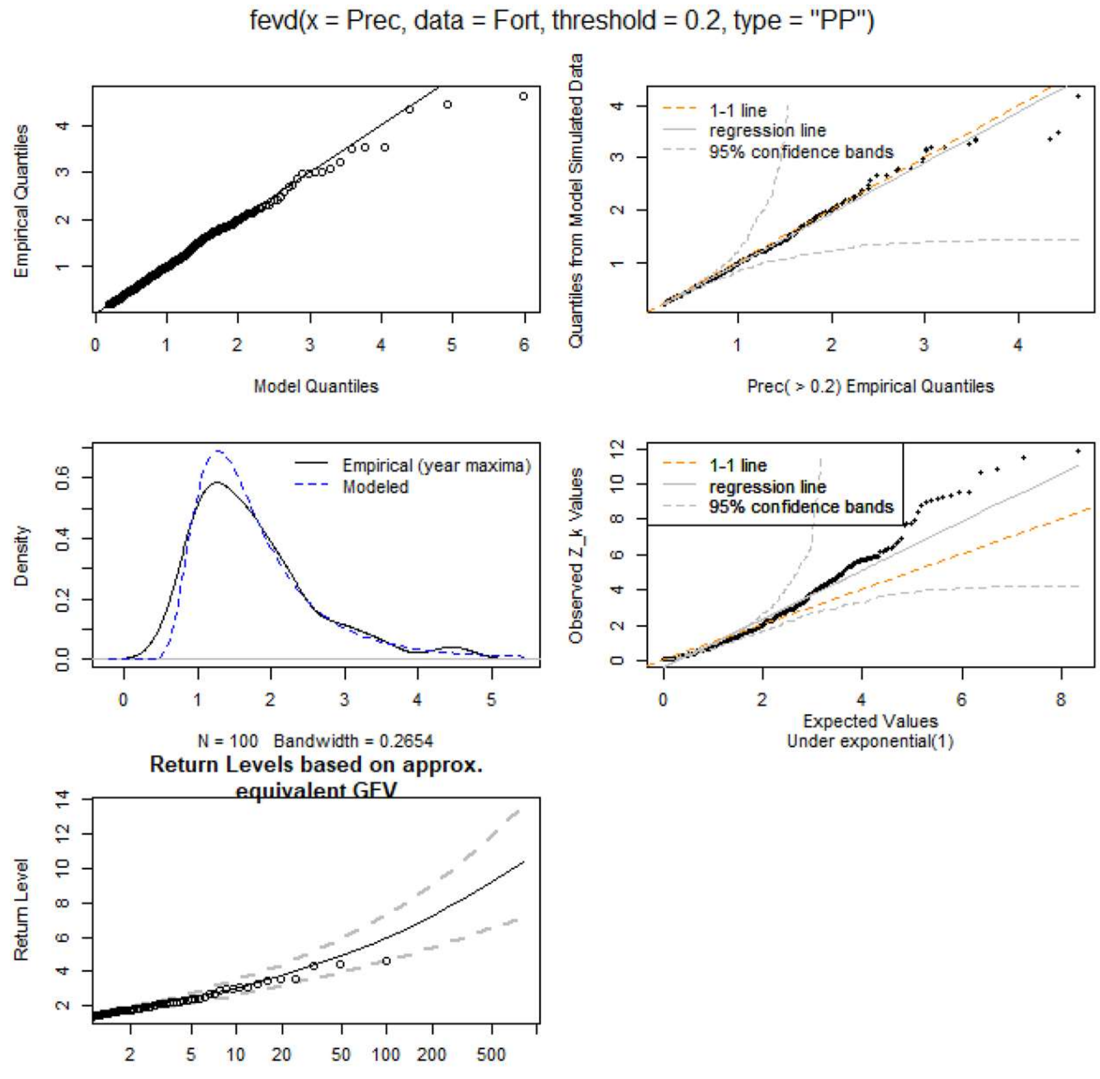
```
fevd(x = Prec, data = bmpcp)

[1] "Estimation Method used: MLE"


 Negative Log-Likelihood Value:  104.9645


 Estimated parameters:
 location     scale     shape
1.3466597 0.5328046 0.1736264

 Standard Error Estimates:
  location      scale      shape
0.06168793 0.04878843 0.09195458

 Estimated parameter covariance matrix.
           location        scale         shape
location  0.003805401  0.0017067043 -0.0020838301
scale     0.001706704  0.0023803113 -0.0008692638
shape    -0.002083830 -0.0008692638  0.0084556445

 AIC = 215.9291

 BIC = 223.7446
```

extRemes: An R software package



```
fit1 <- fevd( Prec,
    data = Fort,
    threshold = 0.2,
    type = "PP" )
fit1
plot( fit1 )
```

extRemes: An R software package



Copyright 2022

```
fit1 <- fevd( Prec,
    data = Fort,
    threshold = 0.2,
    type = "PP" )
fit1
plot( fit1 )
```

extRemes: An R software package

```
fevd(x = Prec, data = Fort, threshold = 0.2, type = "PP")

[1] "Estimation Method used: MLE"


 Negative Log-Likelihood Value:  -4417.258


 Estimated parameters:
 location     scale     shape
1.3832749 0.5476977 0.2383581

 Standard Error Estimates:
  location      scale      shape
0.04456320 0.03705234 0.02796421

 Estimated parameter covariance matrix.
             location        scale        shape
location 0.0019858784 0.0014822808 0.0008135954
scale    0.0014822808 0.0013728757 0.0009489163
shape    0.0008135954 0.0009489163 0.0007819970

 AIC = -8828.515

 BIC = -8811.593
```

extRemes: An R software package

```
fit2 <- fevd(x = Prec, data = Fort, threshold = 0.2, location.fun = ~year,
   type = "PP")

[1] "Estimation Method used: MLE"


 Negative Log-Likelihood Value:  -4417.386


 Estimated parameters:
        mu0           mu1         scale        shape
 1.5793476649 -0.0001005235  0.5478486937  0.2384873599

 Standard Error Estimates:
        mu0          mu1        scale        shape
4.453367e-02 2.000003e-08 3.702638e-02 2.794186e-02

 Estimated parameter covariance matrix.
              mu0           mu1         scale        shape
mu0    1.983248e-03 -1.935322e-12  1.480016e-03  8.118021e-04
mu1   -1.935322e-12  4.000013e-16 -1.557991e-12 -9.375119e-13
scale  1.480016e-03 -1.557991e-12  1.370953e-03  9.473656e-04
shape  8.118021e-04 -9.375119e-13  9.473656e-04  7.807476e-04

 AIC = -8826.772

 BIC = -8804.209
```

Higher values for both AIC and BIC, suggest annual trend not important

extRemes: An R software package

```
lr.test( fit1, fit2 )

        Likelihood-ratio Test

data:  PrecPrec
Likelihood-ratio = 0.25649, chi-square critical value = 3.8415,
alpha =
0.0500, Degrees of Freedom = 1.0000, p-value = 0.6125
alternative hypothesis: greater
```

Propinquity Model Example: Maximum Updraft Velocity × 0-6 km vertical wind shear (WmSh, m^2s^{-2})

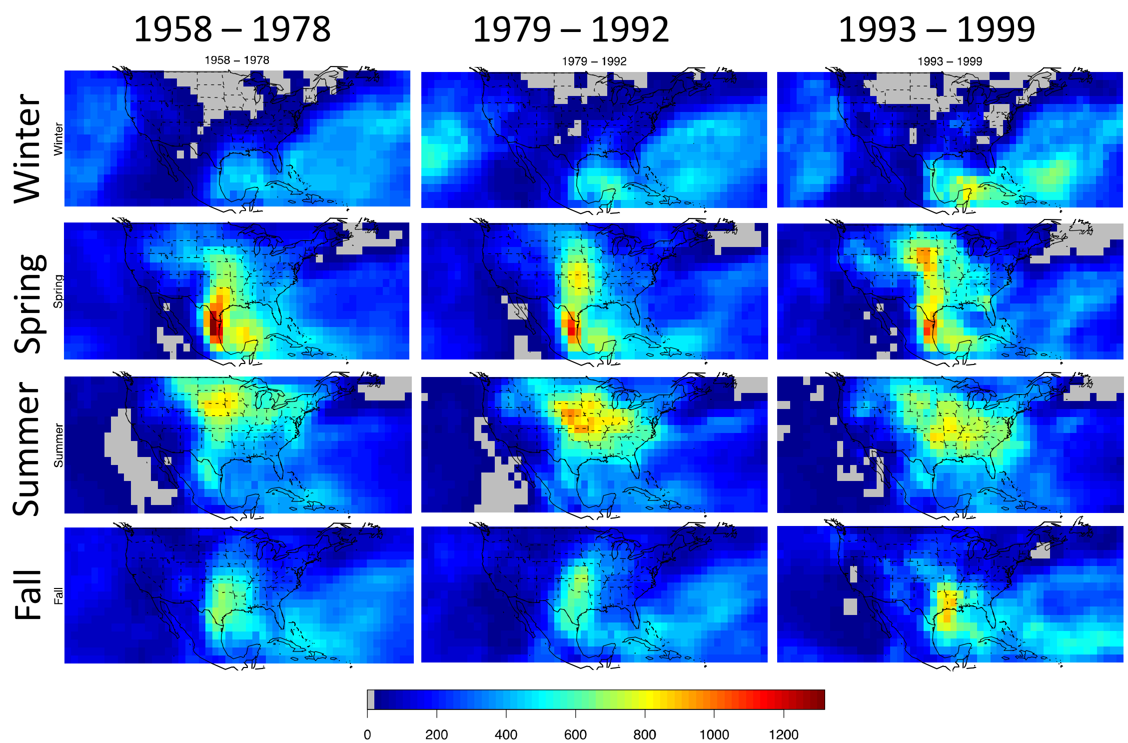
Mean of simulated WmSh at each grid point conditioned on high q75. Using the propinquity model with HT2004.

G. et al 2013. Spatial extreme value analysis to project extremes of large-scale indicators for severe weather. *Environmetrics*, **24** (6), 418 - 432, DOI: 10.1002/env.2234

Note: The R package **texmex** was used for fitting the HT2004 model in these slides.

Harry Southworth, Janet E. Heffernan and Paul D. Metcalfe (2018). texmex: Statistical modelling of extreme values. R package version 2.4.2.

$$\left[\mathrm{WmSh}_1,\ldots,\mathrm{WmSh}_n \,\middle|\, \mathrm{q75} > u\right]$$



Propinquity Model Example: Maximum Updraft Velocity × 0-6 km vertical wind shear (WmSh, m^2s^{-2})

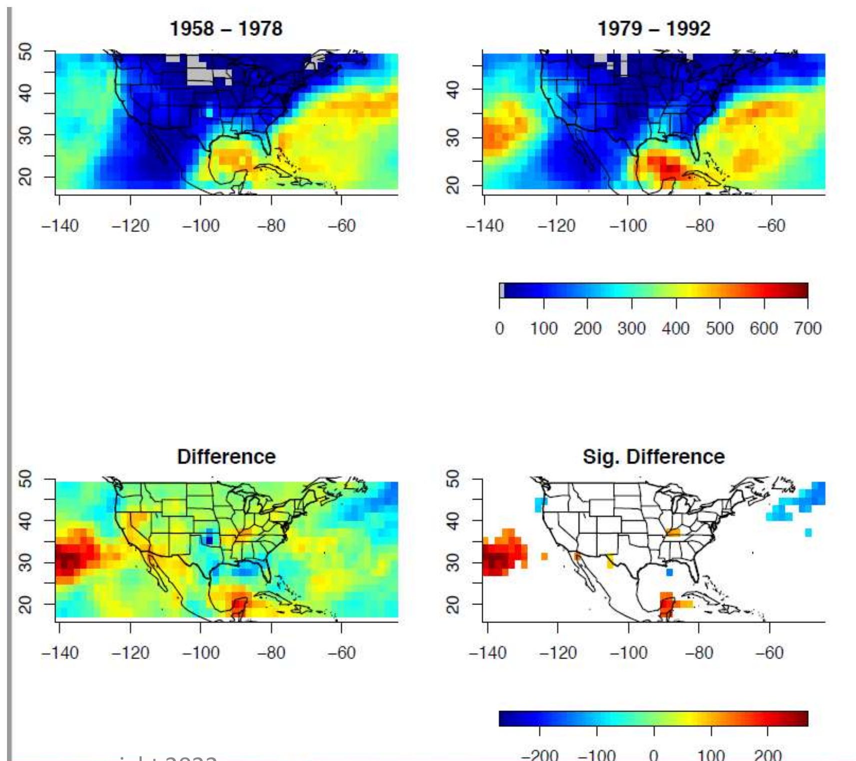
G. et al 2013. Spatial extreme value analysis to project extremes of large-scale indicators for severe weather. *Environmetrics*, **24** (6), 418 - 432, DOI: 10.1002/env.2234

Note: The R package **texmex** was used for fitting the HT2004 model in these slides.

Harry Southworth, Janet E. Heffernan and Paul D. Metcalfe (2018). texmex: Statistical modelling of extreme values. R package version 2.4.2.

$$\left[\mathrm{WmSh}_1,\ldots,\mathrm{WmSh}_n \mid \mathrm{q75} > u\right]$$

Fall (SON)



Propinquity Model Example: Maximum Updraft Velocity × 0-6 km vertical wind shear (WmSh, m^2s^{-2})

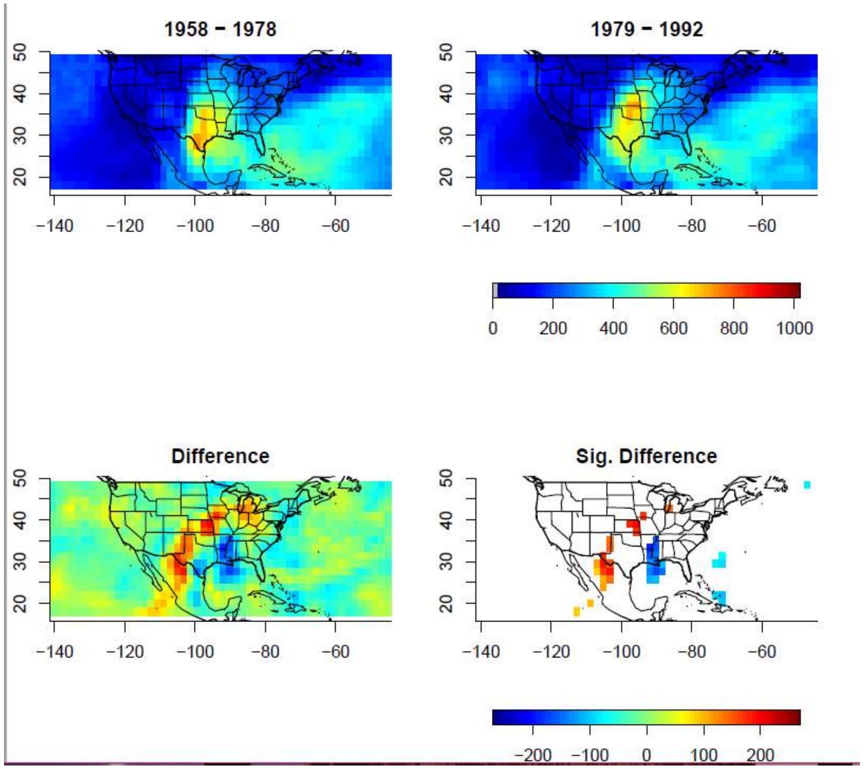
G. et al 2013. Spatial extreme value analysis to project extremes of large-scale indicators for severe weather. *Environmetrics*, **24** (6), 418 - 432, DOI: 10.1002/env.2234

Note: The R package **texmex** was used for fitting the HT2004 model in these slides.

Harry Southworth, Janet E. Heffernan and Paul D. Metcalfe (2018). texmex: Statistical modelling of extreme values. R package version 2.4.2.

$$\left[\mathrm{WmSh}_1, \ldots, \mathrm{WmSh}_n \middle| q75 > u\right]$$

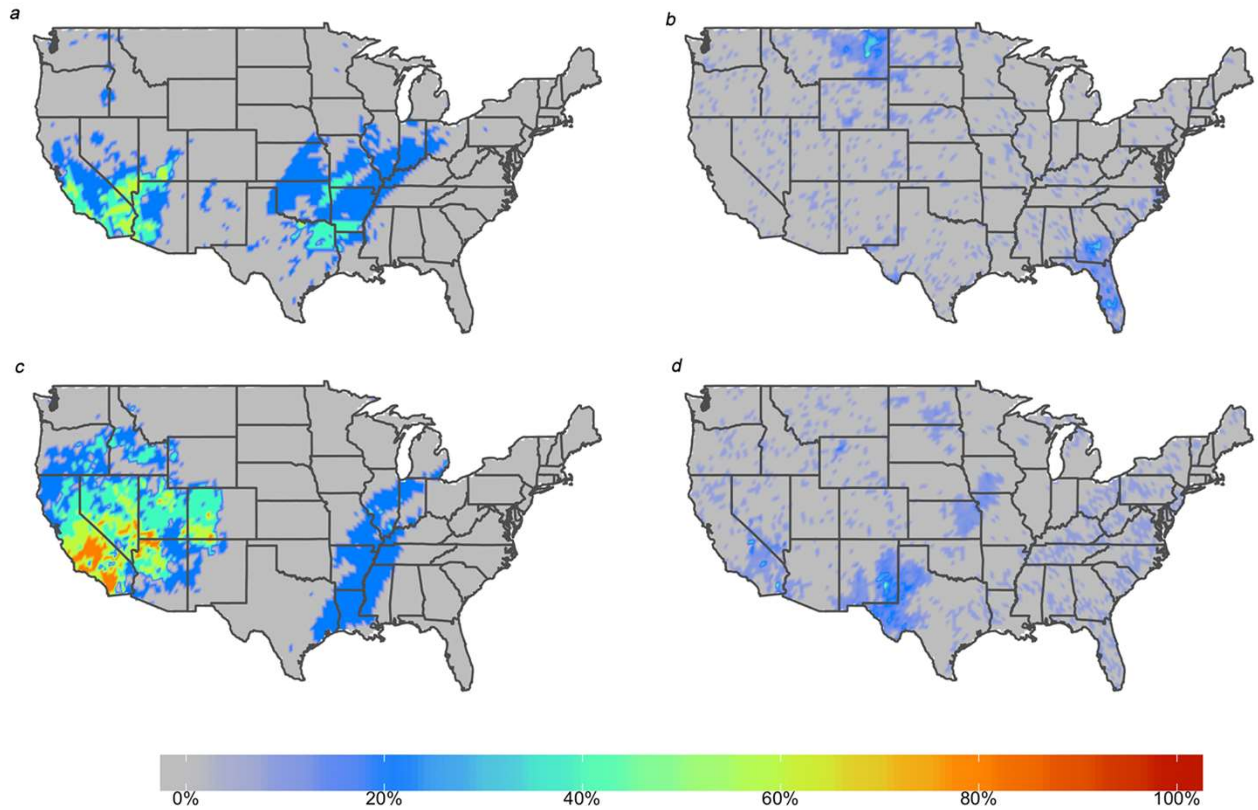
Winter (DJF)



Example: GHCN-Daily precipitation data

Kholodovsky, V. and X. Liang, 2021. A generalized spatio-temporal threshold clustering method for identification of extreme event patterns. *Adv. Stat. Clim. Meteorol. Oceanogr.*, 7, 35–52, doi: 10.5194/ascmo-7-35-2021.

Propinquity Model | Traditional (individual grid points)



Other References for further information

de Oliveira, M. M. F., J. L. F. de Oliveira, P. J F., Fernandes, P. J. F., E. Gilleland, and N. F. F. Ebecken, 2021. Extreme climate characteristics near the coastline of the southeast region of Brazil in the last 40 years. *Theoritcal and Applied Climatology*, **146**, 657 - 674, doi: 10.1007/s00704-021-03711-z.

Jha, S., M. Goyal, B. B. Gupta, C.-H. Hsu, E. Gilleland, and J. Das, 2021. A methodological framework for extreme climate risk assessment integrating satellite and location based datasets in Intelligent Systems. *International Journal of Intelligent Systems*, (in press), doi: 10.1002/int.22475.

Towler, E., D. Llewellyn, A. Prein, and E. Gilleland, 2020. Extreme-value analysis for the characterization of extremes in water resources: A generalized workflow and case study on New Mexico monsoon precipitation. *Weather and Climate Extremes*, **29**, 100260, doi: 10.1016/j.wace.2020.100260.

Gilleland, E., 2020. Bootstrap methods for statistical inference. Part II: Extreme-value analysis. *Journal of Atmospheric and Oceanic Technology*, **37** (11), 2135 - 2144, doi: 10.1175/JTECH-D-20-0070.1.

Gilleland, E., R. W. Katz, and P. Naveau, 2017. Quantifying the risk of extreme events under climate change. *Chance*, **30** (4), 30 - 36, doi: 10.1080/09332480.2017.1406757.

Gilleland, E. and R. W. Katz, 2016. extRemes 2.0: An Extreme Value Analysis Package in R. *Journal of Statistical Software*, **72** (8), 1 - 39, doi: 10.18637/jss.v072.i08.

Towler, E., B. Rajagopalan, E. Gilleland, R.S. Summers, D. Yates, and R.W. Katz, 2010. Modeling hydrologic and water quality extremes in a changing climate. *Water Resources Research*, **46**, W11504, doi: 10.1029/2009WR008876.