Extreme value analysis with the R package extRemes

Eric Gilleland
Research Applications Laboratory
Weather and Climate Impacts Assessment Program
National Center for Atmospheric Research
28 August 2017

Environmental Risk Modeling and Extreme Events Workshop
28 – 31 August 2017
Centre de Recherches Mathématiques, Montréal, Québec, Canada

National Center for Atmospheric Research

Background Information

- Software funded by the Weather and Climate Impacts Assessment Science Program (http://www.assessment.ucar.edu)
- Project impetus, and continuing involvement, from Rick Katz (http://www.isse.ucar.edu/staff/katz)
- Primary goal is to shorten learning curve for atmospheric scientists to apply extreme value analysis (EVA) in their work when appropriate
- Two R (http://www.r-project.org) packages: extRemes (command-line) and in2extRemes (GUI for some extRemes functions)
- Web page for extRemes and in2extRemes (http://www.ral.ucar.edu/staff/ericg/extRemes)

Background Information

Tutorials for extRemes and in2extRemes

- 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 (https://www.jstatsoft.org/article/view/v072i08).
- Gilleland, E. and Katz, R. W., 2016: in2extremes: Into the R Package extremes - Extreme Value Analysis for Weather and Climate Applications. *NCAR Technical Note*, NCAR/TN-523+STR, 102 pp., DOI: 10.5065/D65T3HP2 (http://dx.doi.org/10.5065/D65T3HP2).

Background Information

Other EVA software (not just R packages, but mostly):

- List of EVA software at http://www.ral.ucar.edu/staff/ericg/softextreme.php
- Gilleland, E., 2016. Computing Software. Chapter 25 In *Extreme Value Modeling and Risk Analysis: Methods and Applications*. Edts. Dipak K. Dey and Jun Yan, CRC Press, Boca Raton, Florida, U.S.A., pp. 505 - 515.
- Gilleland, E. and Ribatet, M., 2015. Reinsurance and extremal events. In: *Computational Actuarial Science with R*. Ed. A. Charpentier, Chapman & Hall/CRC the R series, Boca Raton, Florida, U.S.A., pp. 257 - 286.
- Gilleland, E., M. Ribatet and A. G. Stephenson, 2013. A software review for extreme value analysis. *Extremes*, **16** (1), 103 - 119, DOI: 10.1007/s10687-012-0155-0 (available online at http://www.springerlink.com/openurl.asp?genre=article&id=doi:10.1007/s10687-012-0155-0).
- Stephenson, A. and E. Gilleland, 2005. Software for the Analysis of Extreme Events: The Current State and Future Directions. *Extremes*, **8**, 87 - 109.

fevd

NCAR

Main function for all (univariate) extreme value distribution (EVD) fitting

```
fevd(x, data, threshold = NULL, threshold.fun = ~1,
       location.fun = ~1, scale.fun = ~1, shape.fun = ~1,
       use.phi = FALSE, type = c("GEV", "GP", "PP",
              "Gumbel", "Exponential"),
       method = c("MLE", "GMLE", "Bayesian", "Lmoments"),
       initial = NULL, span, units = NULL,
       time.units = "days", period.basis = "year",
       na.action = na.fail, optim.args = NULL,
       priorFun = NULL, priorParams = NULL,
       proposalFun = NULL, proposalParams = NULL,
       iter = 9999, weights = 1, blocks = NULL,
       verbose = FALSE)
```

It's not as bad as it looks!

fevd

Fit GEV to block maxima using MLE

Samples of size 100 of maxima of standard normal distributed samples

```
Zmax <- matrix( rnorm( 100 * 1000 ), 1000, 100 )

dim( Zmax )

Zmax <- apply( Zmax, 2, max )

dim( Zmax )
```

fevd

Fit GEV to block maxima using MLE

```
library( extRemes )

fit <- fevd( Zmax )

fit
```

Load the library

Do the fit

See a summary of the fit

fevd

Fit GEV to block maxima using MLE

```
plot( fit )

ci( fit, type = "parameter" )

distill( fit )

strip( fit )
```

Look at diagnostic plots for the fit

Obtain CI's for the GEV parameters

See just the parameter estimates and some other information (useful for multiple fits, e.g., thousands of locations)

New (not yet available). Same as distill(), but simpler (only parameter estimates shown)

fevd

Fit GEV to block maxima using MLE

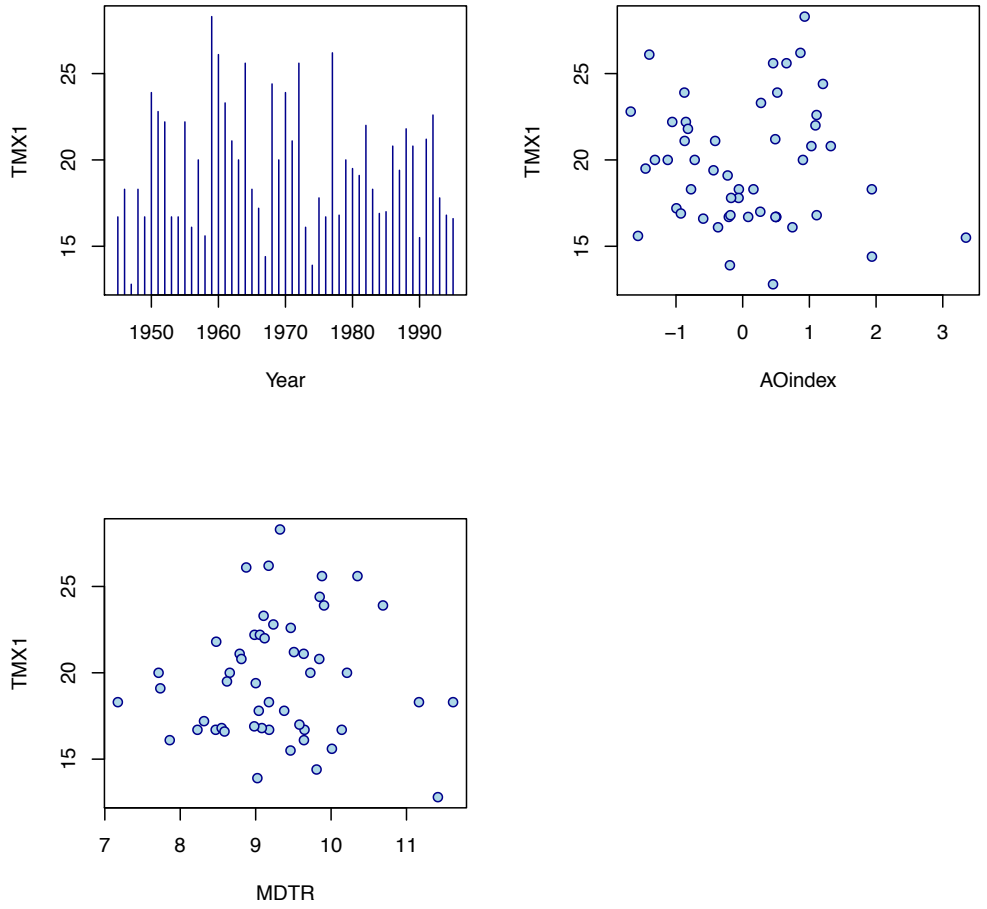
Maximum winter temperature (^{o}C) in Sept-Iles, Québec

```
data( SEPTsp )

?SEPTsp
```

fevd

Fit GEV to block maxima using MLE

Maximum winter temperature (°C) in Sept-Iles, Québec



fevd

Fit GEV to block maxima using MLE

Maximum winter temperature (°C) in Sept-Iles, Québec

```
fit0 <- fevd(TMX1, data = SEPTsp,
    units = "deg C")

fit0
```

fevd

Fit GEV to block maxima using MLE

Maximum winter temperature (°C) in Sept-Iles, Québec

`fit0`

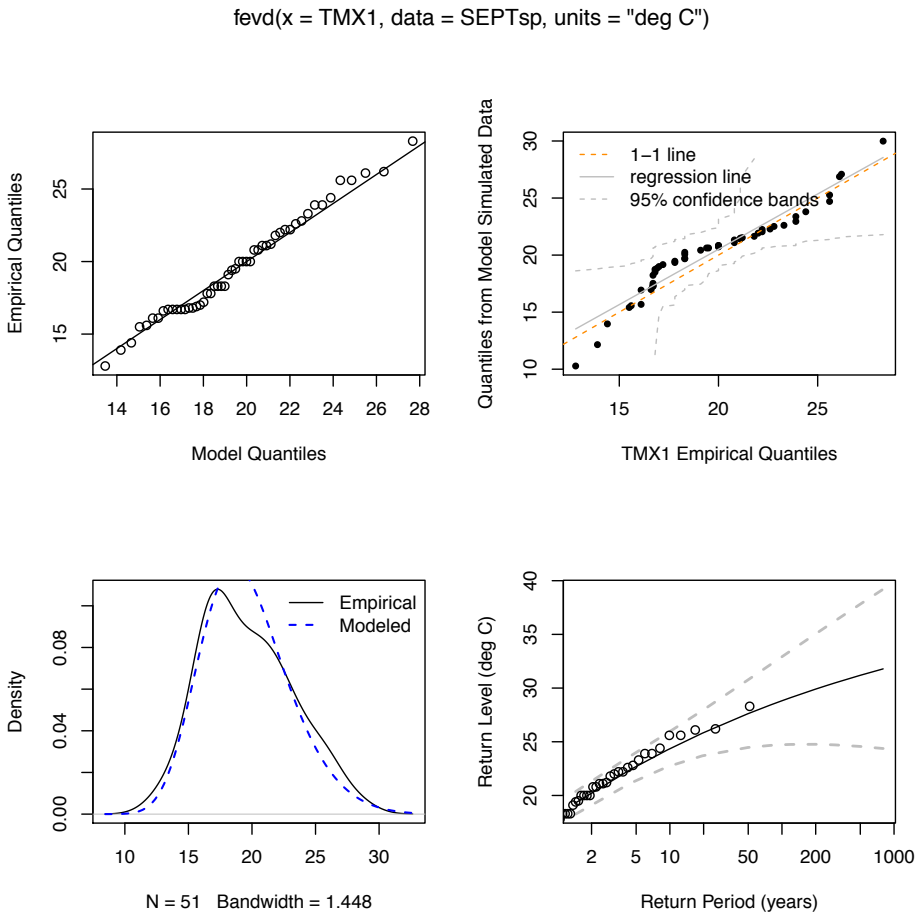
```
fevd(x = TMX1, data = SEPTsp, units = "deg C")

[1] "Estimation Method used: MLE" Negative Log-Likelihood Value:  134.9045

Estimated parameters:
  location      scale      shape
18.1978488  3.1266252 -0.1395647

Standard Error Estimates:
  location     scale     shape
0.4999587 0.3616231 0.1168080

Estimated parameter covariance matrix.
          location       scale       shape
location  0.24995872  0.04741458 -0.02468781
scale     0.04741458  0.13077124 -0.02121723
shape    -0.02468781 -0.02121723  0.01364411

AIC = 275.8091

BIC = 281.6045
```

fevd

Fit GEV to block maxima using MLE

Maximum winter temperature (°C) in Sept-Iles, Québec

```
plot( fit0 )
```



fevd

Fit GEV to block maxima using MLE

Maximum winter temperature (°C) in Sept-Iles, Québec

```
ci( fit0, type = "parameter" )
fevd(x = TMX1, data = SEPTsp, units = "deg C")

[1] "Normal Approx."
          95% lower CI   Estimate  95% upper CI
location      17.2179478 18.1978488  19.17774993
scale                    2.4178570  3.1266252    3.83539336
shape                   -0.3685042 -0.1395647    0.08937479
```

```
ci( fit0 )
fevd(x = TMX1, data = SEPTsp, units = "deg C")

[1] "Normal Approx."

[1] "100-year return level: 28.812"

[1] "95% Confidence Interval: (24.7221, 32.9011)"
```

fevd

Fit GEV to block maxima using MLE

Maximum winter temperature (°C) in Sept-Iles, Québec

```
fit1 <- fevd( TMX1, data = SEPTsp,
    location.fun = ~AOindex,
    units = "deg C")
fit1
Negative Log-Likelihood Value:  134.4556
Estimated parameters:
mu0         mu1       scale       shape
18.1781844 -0.4220587  3.0397157 -0.1043810
Standard Error Estimates:
mu0        mu1      scale      shape
0.4853334 0.4388729 0.3527318 0.1177925
…
AIC = 276.9112  BIC = 284.6385
```

Recall for fit0 that

AIC = 275.8091

BIC = 281.6045

Indicating fit0 is better!

Results shortened for space

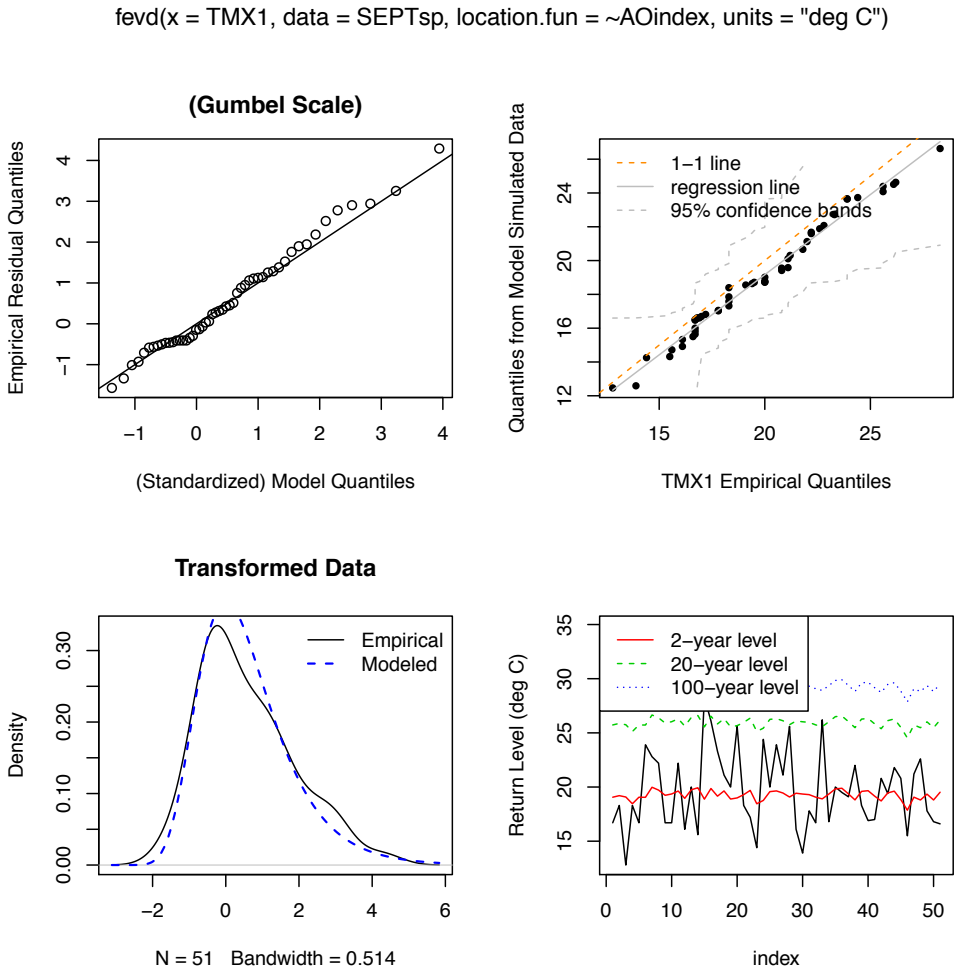
$\mu(\text{AOindex}) = \mu_0 + \mu_1 * \text{AOindex}$

fevd

Fit GEV to block maxima using MLE

Maximum winter temperature (°C) in Sept-Iles, Québec

```
plot(fit1)
```



fevd

Fit GEV to block maxima using MLE

Maximum winter temperature (°C) in Sept-Iles, Québec

```
lr.test(fit0, fit1)
```

```
Likelihood-ratio Test

data:  TMX1TMX1

Likelihood-ratio = 0.89789, chi-square critical value = 3.8415,
alpha =0.0500, Degrees of Freedom = 1.0000, p-value = 0.3433
alternative hypothesis: greater
```

Result agrees with AIC and BIC

fevd

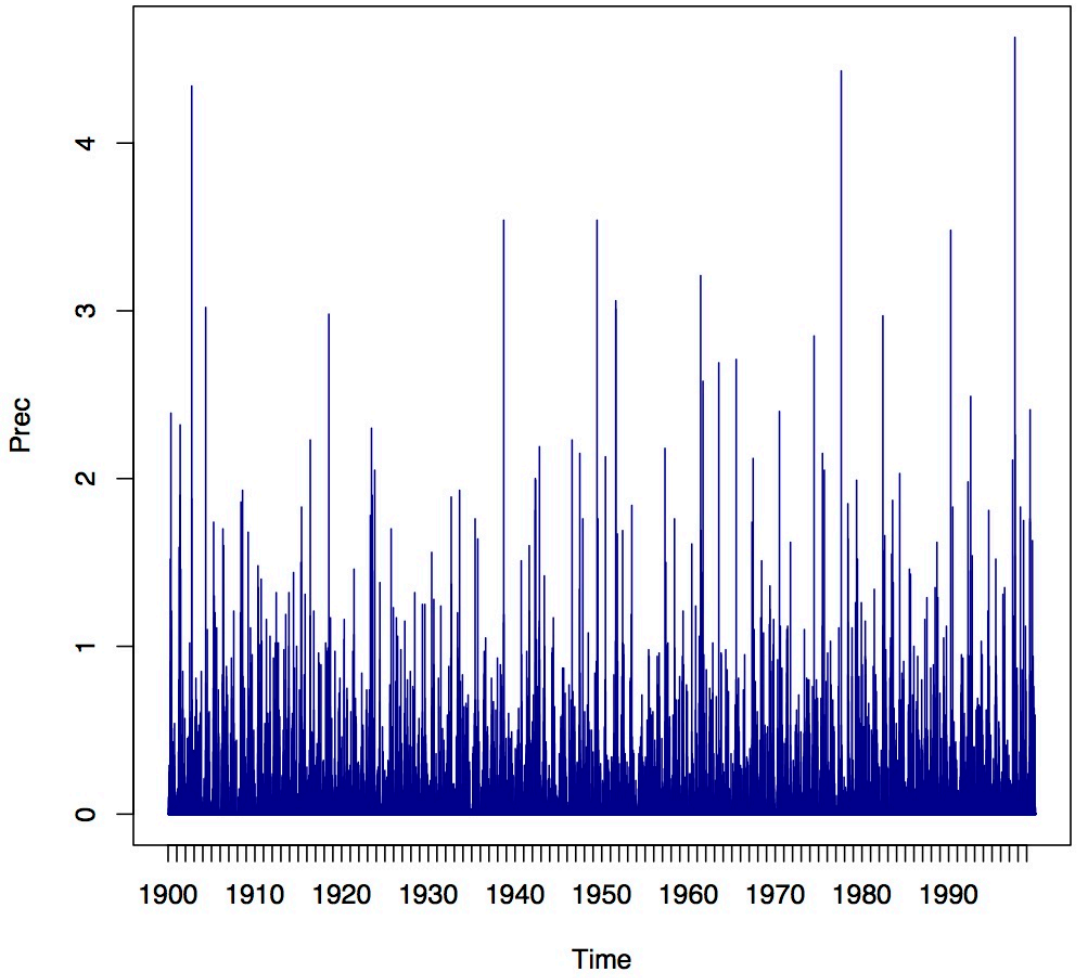
Fit GEV to minimum winter temperature (°C) using MLE

(Negative) Minimum winter temperature (°C) in Sept-Iles, Québec

```
fit0 <- fevd(-TMN0 ~ 1, data = SEPTsp,
    units = "neg. deg. C")
```

fevd

Fort Collins, Colorado daily precipitation (inches) 1900 to 1999



fevd

Fort Collins, Colorado daily precipitation (inches) 1900 to 1999

Fit a GP distribution to the data

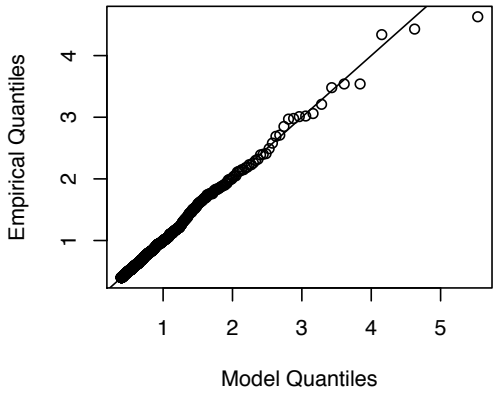
```
fit <- fevd( Prec, data = Fort,
     threshold = 0.395, type = "GP",
     units = "inches" )
```

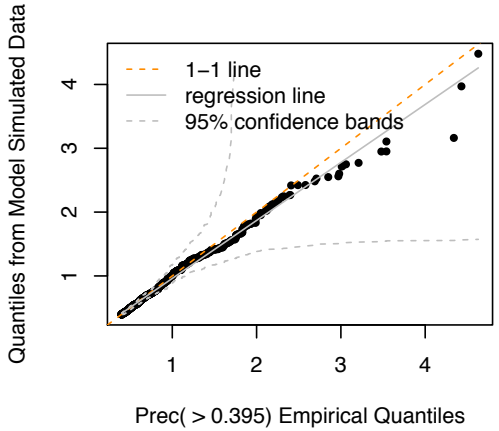
fevd

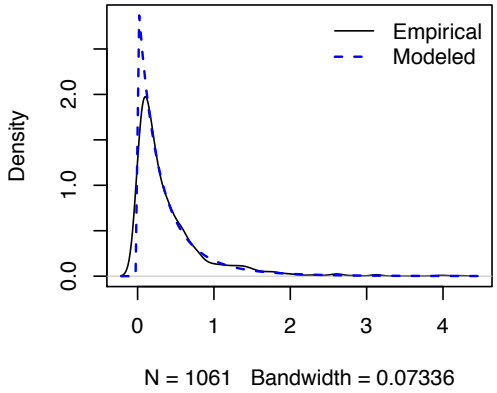
Fort Collins, Colorado daily precipitation (inches) 1900 to 1999

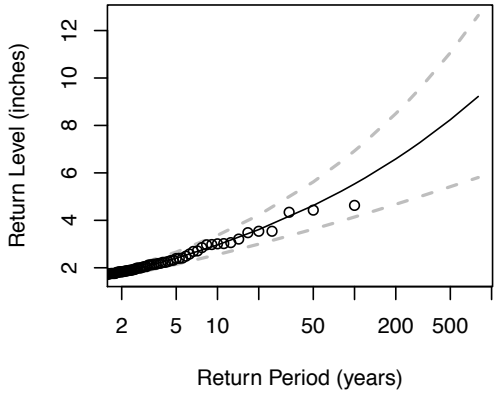
Fit a GP distribution to the data

```
plot( fit )
```





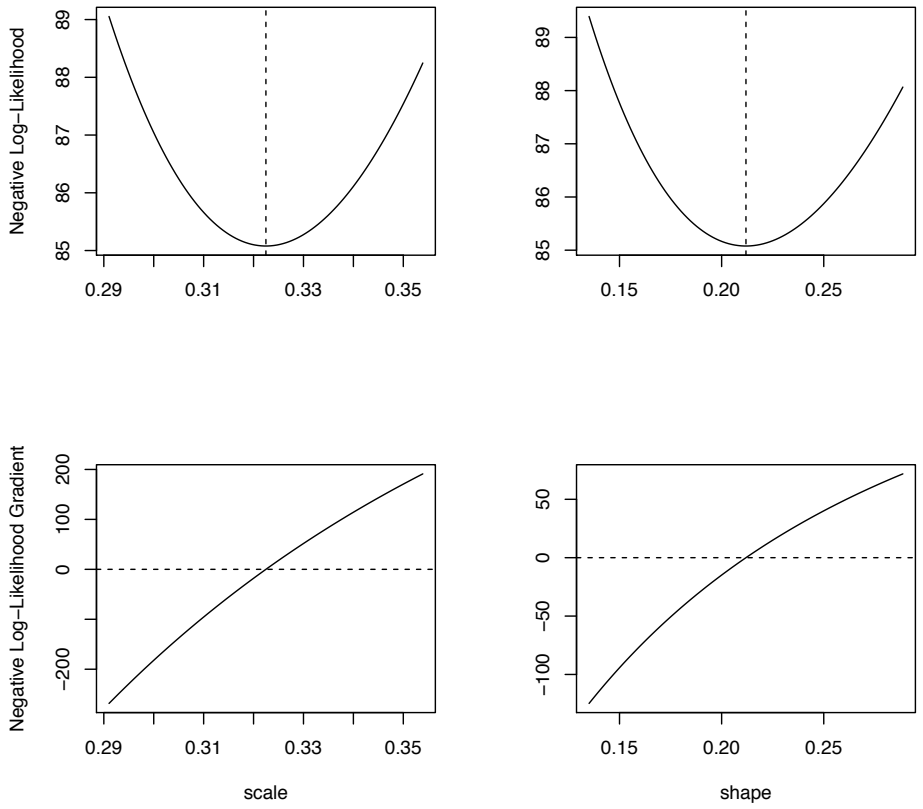




fevd

Fort Collins, Colorado daily precipitation (inches) 1900 to 1999

```
plot( fit, type = “trace” )
```



fevd

Fort Collins, Colorado daily precipitation (inches) 1900 to 1999

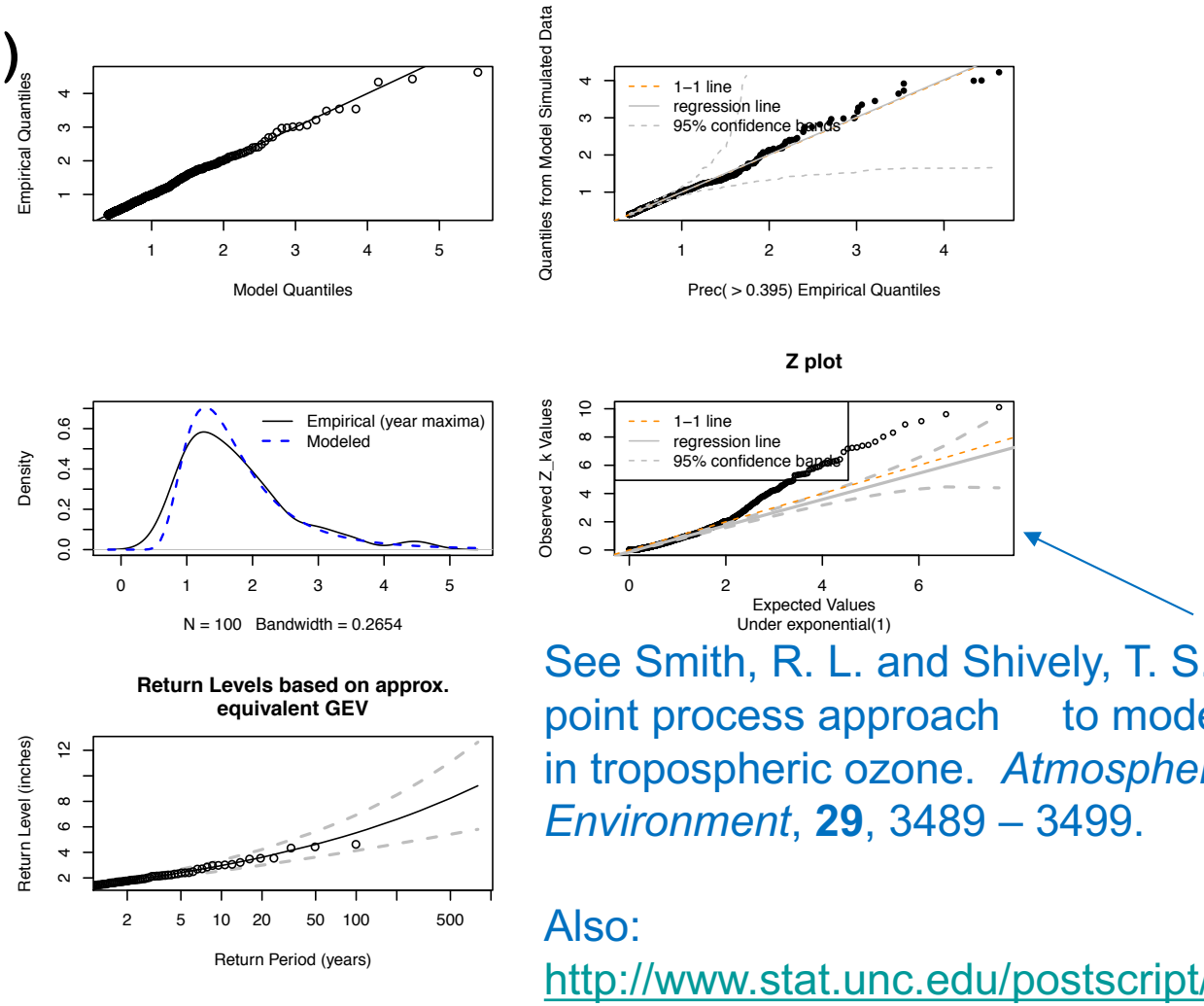
Fit a Poisson Point Process to the data

```
fit <- fevd( Prec, Fort,
     threshold = 0.395, type = "PP",
     units = "inches" )
```

fevd

Fort Collins, Colorado daily precipitation (inches) 1900 to 1999

```
plot( fit )
```



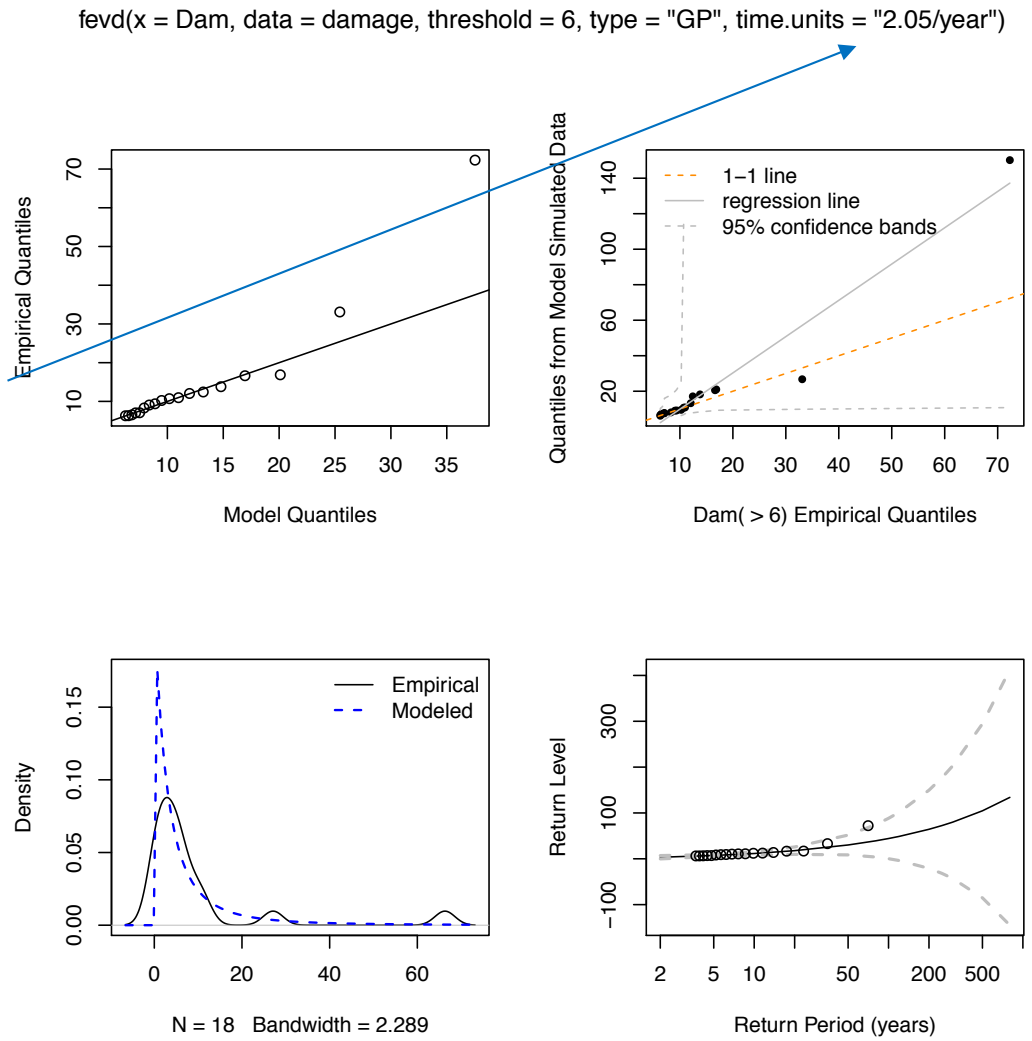
See Smith, R. L. and Shively, T. S., 1995. A point process approach to modeling trends in tropospheric ozone. *Atmospheric Environment*, **29**, 3489 – 3499.

Also:
http://www.stat.unc.edu/postscript/rs/var.pdf

fevd

Estimated economic damage (billions USD) caused by hurricanes

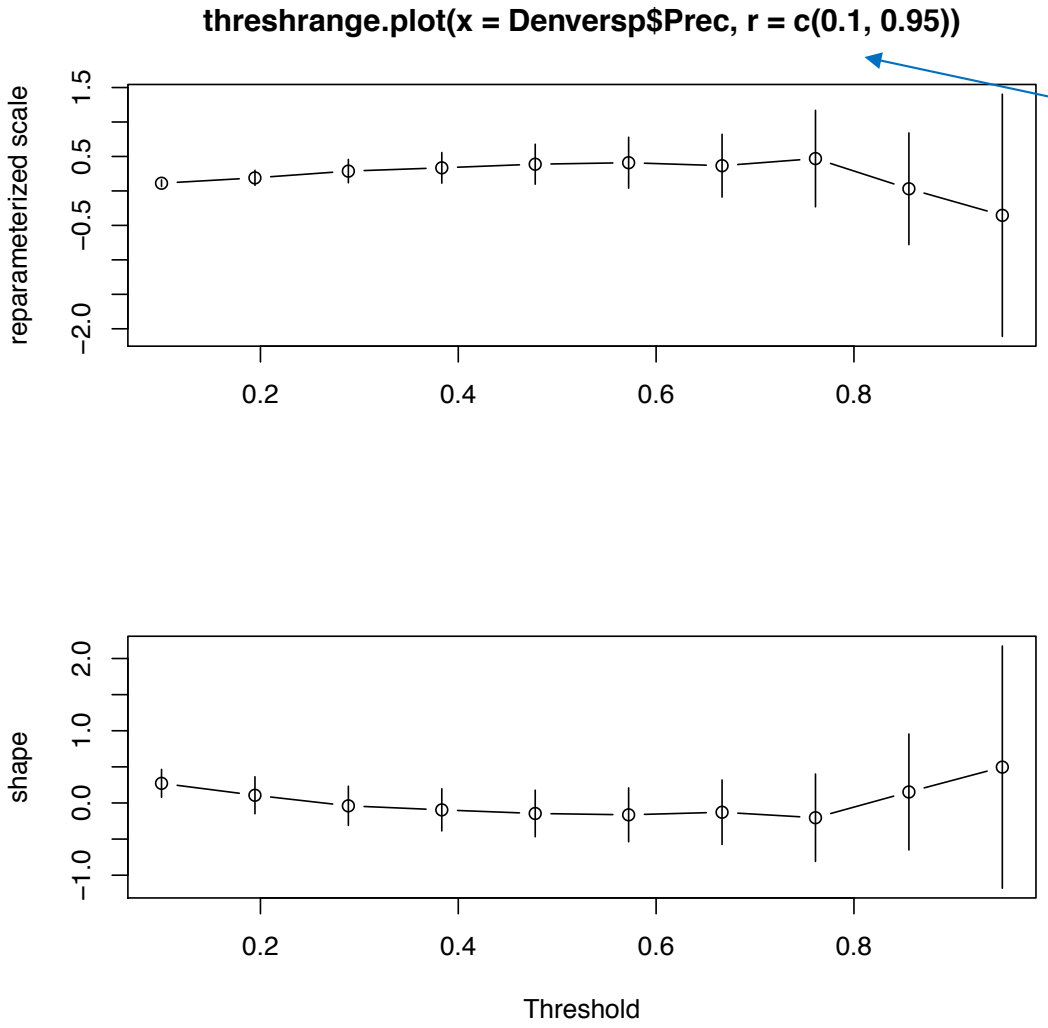
Data not taken every x time units, so must estimate an average number of events per year.



Threshold Selection

Plot parameter estimates over a range of thresholds

Generalized Pareto



Range of thresholds

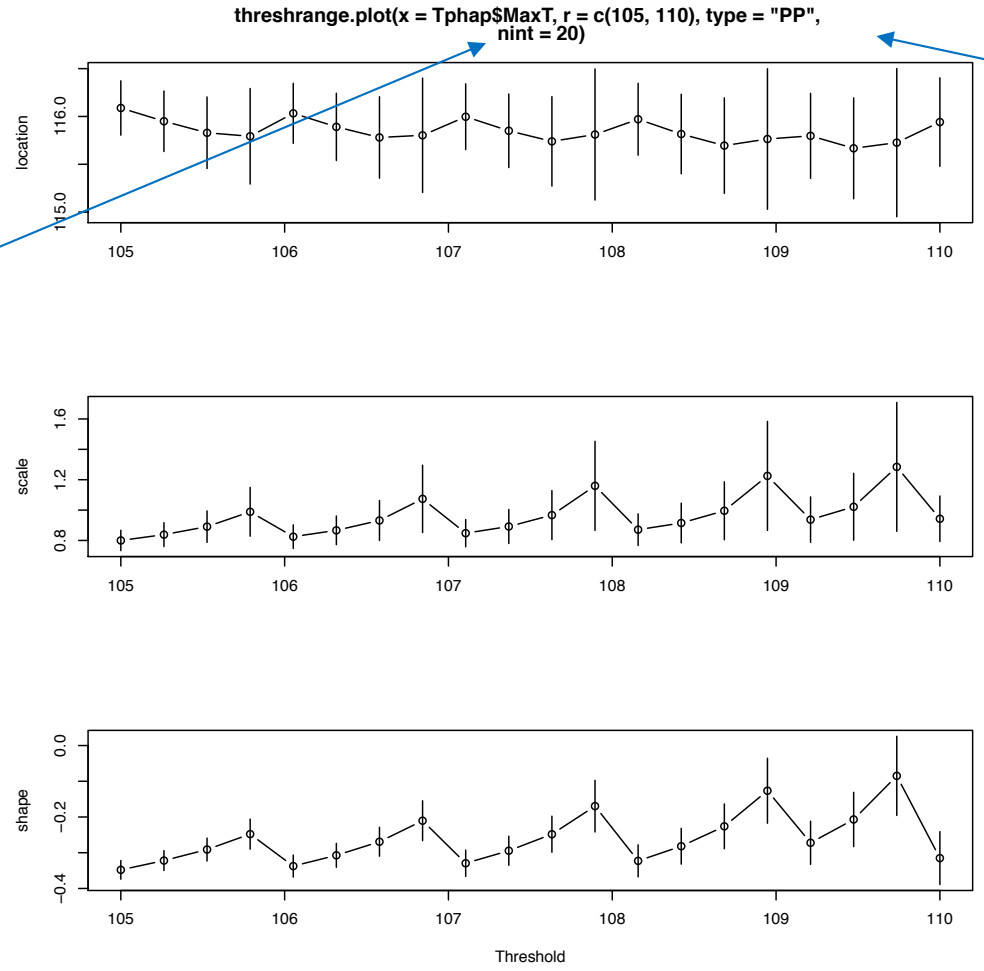
Threshold Selection

Plot parameter estimates over a range of thresholds

Poisson Point Process

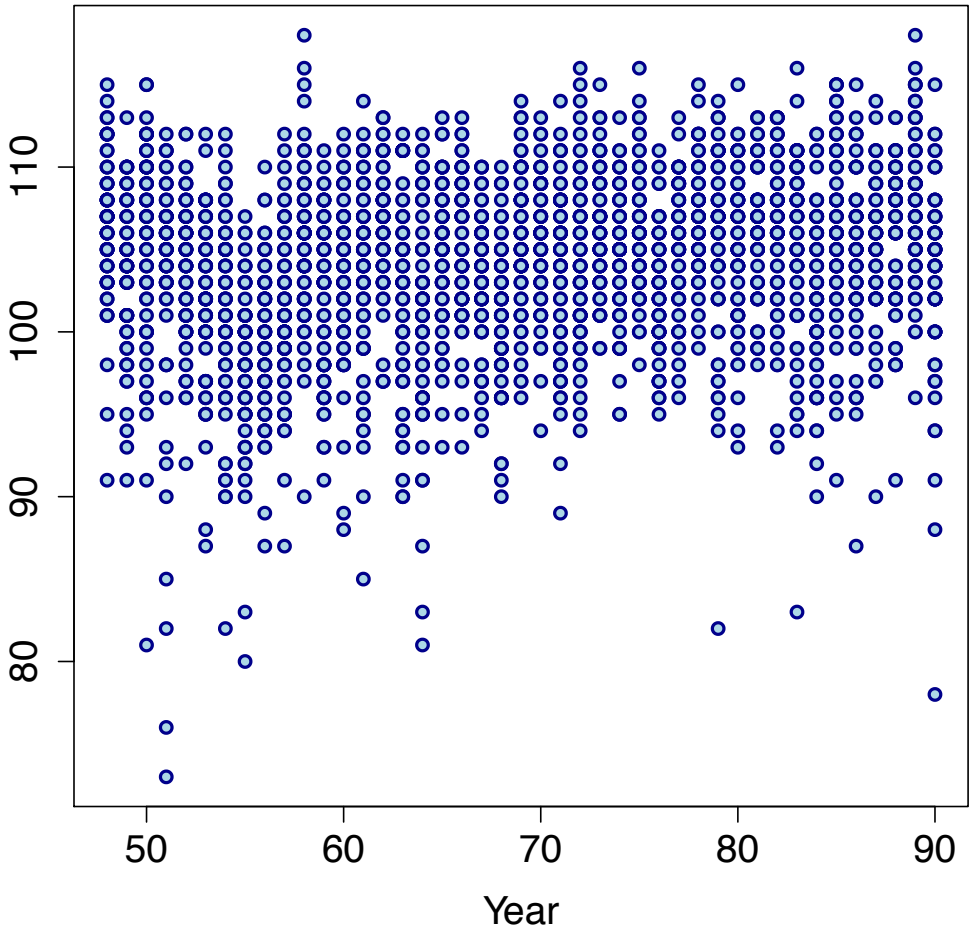
Specify type = “PP”

Change number of thresholds over the specified range



Declustering

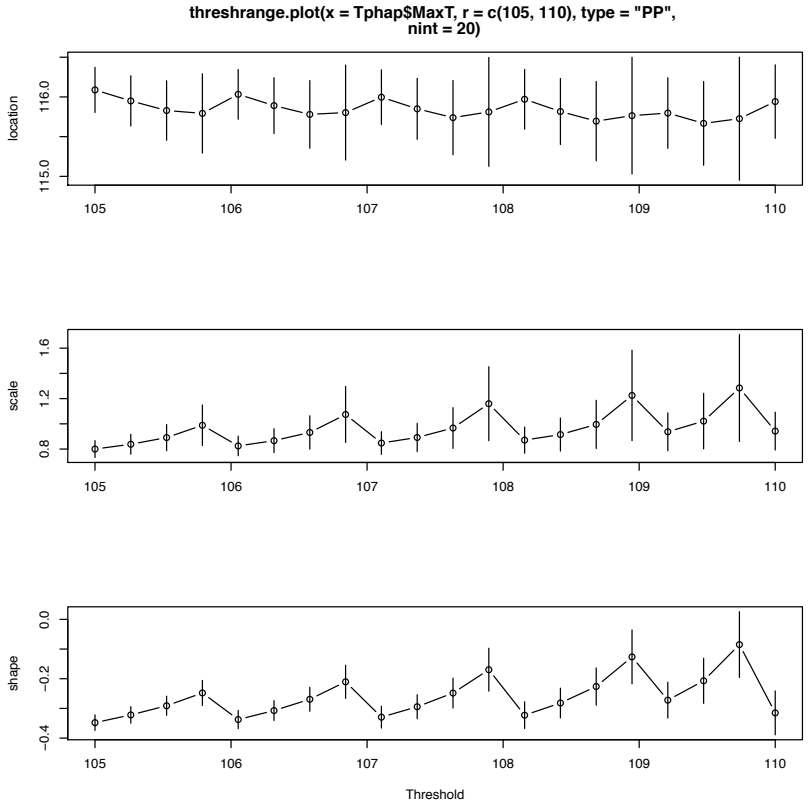
Sky Harbor airport, Phoenix, Arizona July to August maximum temperatures (^{o}F)



Declustering

Sky Harbor airport, Phoenix, Arizona July to August maximum temperatures (^{o}F)

```
threshrange.plot(Tphap$MaxT,
    r = c(105, 110), type = "PP")
```



Declustering

Sky Harbor airport, Phoenix, Arizona July to August maximum temperatures (^{o}F)

```
extremalindex( Tphap$MaxT, threshold = 105 )
```

θ	Number of Clusters	Run Length
0.21	234	2

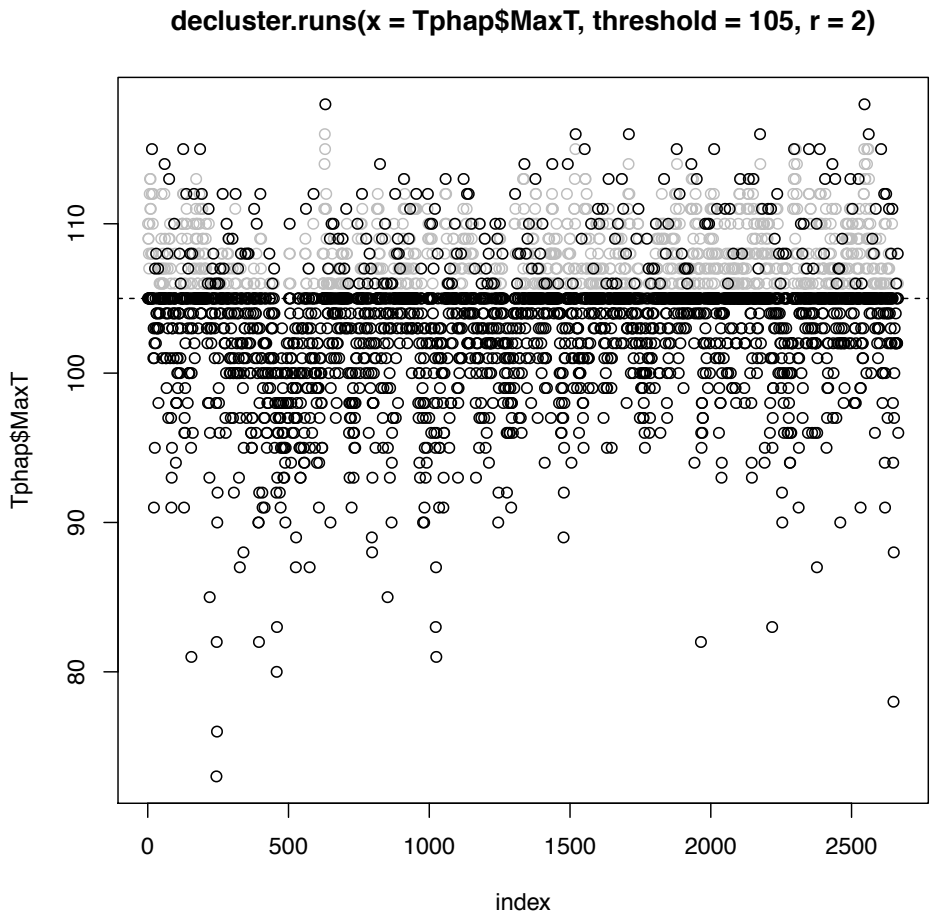
Declustering

Sky Harbor airport, Phoenix, Arizona July to August maximum temperatures (^{o}F)

```
y <- decluster( Tphap$MaxT,
     threshold = 105,
     r = 2 )

y

plot( y )
```

Declustering

Sky Harbor airport, Phoenix, Arizona July to August maximum temperatures (^{o}F)



Declustering

Sky Harbor airport, Phoenix, Arizona July to August maximum temperatures (^{o}F)

```
extremalindex( y, threshold = 105 )
```

θ	Number of Clusters	Run Length
1	229	3

Declustering

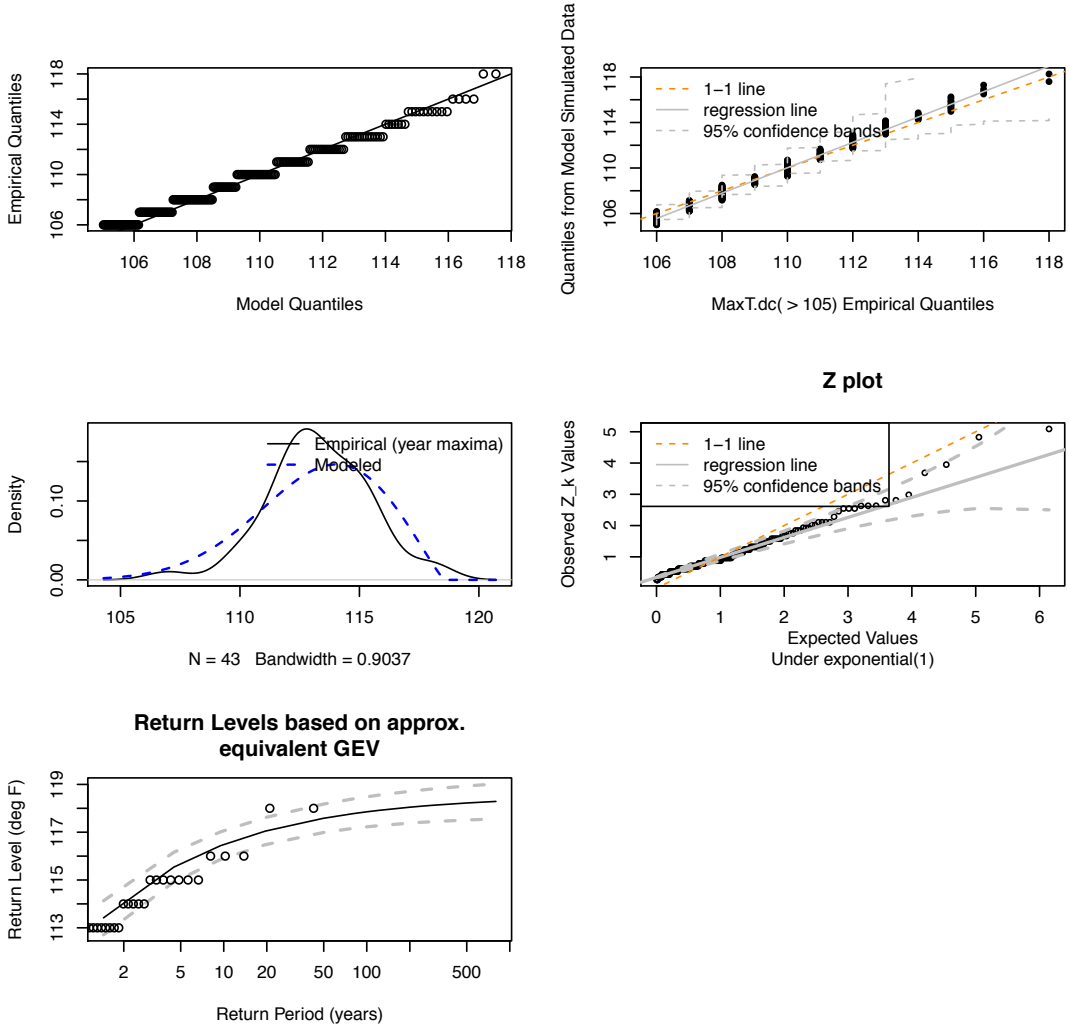
Sky Harbor airport, Phoenix, Arizona July to August maximum temperatures (^{o}F)

```
Tphap2 <- Tphap
Tphap2$MaxT.dc <- c(y)

fit <- fevd( MaxT.dc, threshold = 105,
     data = Tphap2, type = "PP",
     time.units = "62/year",
     units = "deg F" )
```

Declustering

NCAR

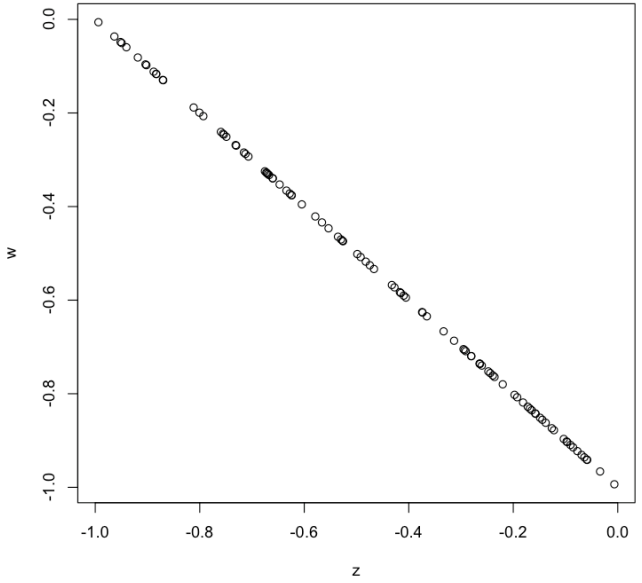
Sky Harbor airport, Phoenix, Arizona July to August maximum temperatures (^{o}F)



Tail dependence

Example where a random variable is completely dependent in terms of the variables, but completely tail independent (from Reiss and Thomas (2007) p. 75.

```
z <- runif( 100, -1, 0 )
w <- -1*(1 + z)
taildep( z, w, u = 0.8 )
taildep.test( z, w )
```



Reiss, R.-D. and Thomas, M., 2007. *Statistical Analysis of Extreme Values: with applications to insurance, finance, hydrology and other fields*. Birkhäuser, 530pp., 3rd edition.

Future Plans

- New bootstrap options (testing stage)
  - Currently only parametric bootstrap with percentile method is available via ci() function
  - Multiple options for regular bootstrap using the distillery package
    - m < n bootstrap
    - iid and block bootstrap options
    - Multiple choices for estimated intervals (e.g., BCa, basic, bootstrap-t, normal, etc.)
  - Test-inversion bootstrap also using distillery package
- New bivariate EVA functionality (with help from Dan Cooley; early stage)
- Other … (thinking stage; funding dependent)

Discussion Questions

- What functionality is missing from the software that would be most useful to include (that is not already on the docket)?
- Open-source software, such as extRemes, is use-at-your-own-risk. But, is a proprietary package better?

Thanks! Questions?