



NCAR



Statistical Extreme Value Theory (EVT) Part II

Eric Gilleland

Research Applications Laboratory

21 July 2014, Trieste, Italy

National Center for Atmospheric Research



Peaks Over Threshold (POT)



Poisson distribution applicable for modeling the *frequency* of exceeding a high threshold (i.e., low probability, rare, event).

What about the intensity of values that exceed the threshold?

Peaks Over Threshold (POT)



- X random variable
- Let $Y = X - u$, conditional on $X > u$, where u is a high threshold.
- Model the “excesses” over the threshold.
- Y has an approximate generalized Pareto (GP) distribution for high u with cdf:

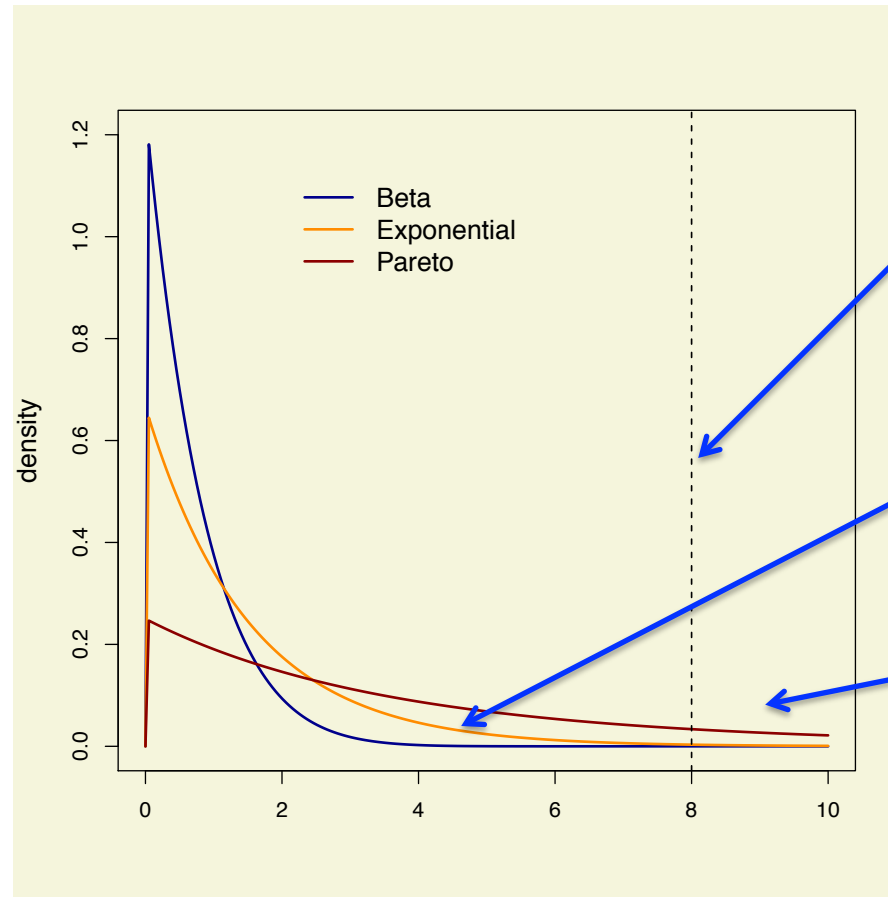
$$H(y; \sigma(u), \xi) = 1 - \left[1 + \xi \frac{y}{\sigma(u)} \right]^{-1/\xi}, \quad y > 0, \quad \xi \frac{y}{\sigma(u)} > 0$$

$$\sigma(u) > 0$$

Peaks Over Threshold (POT)



Three types
of Extreme
Value df's



Beta df;
bounded
upper tail

Exponential;
light tail

Pareto;
heavy tail

Peaks Over Threshold (POT)



- $\xi < 0$ yields the Beta cdf (bounded upper tail)
upper bound at: $u - \sigma(u) / \xi$
- $\xi = 0$ yields the exponential cdf (“light” upper tail)
- $\xi > 0$ yields the Pareto cdf (“heavy” upper tail)

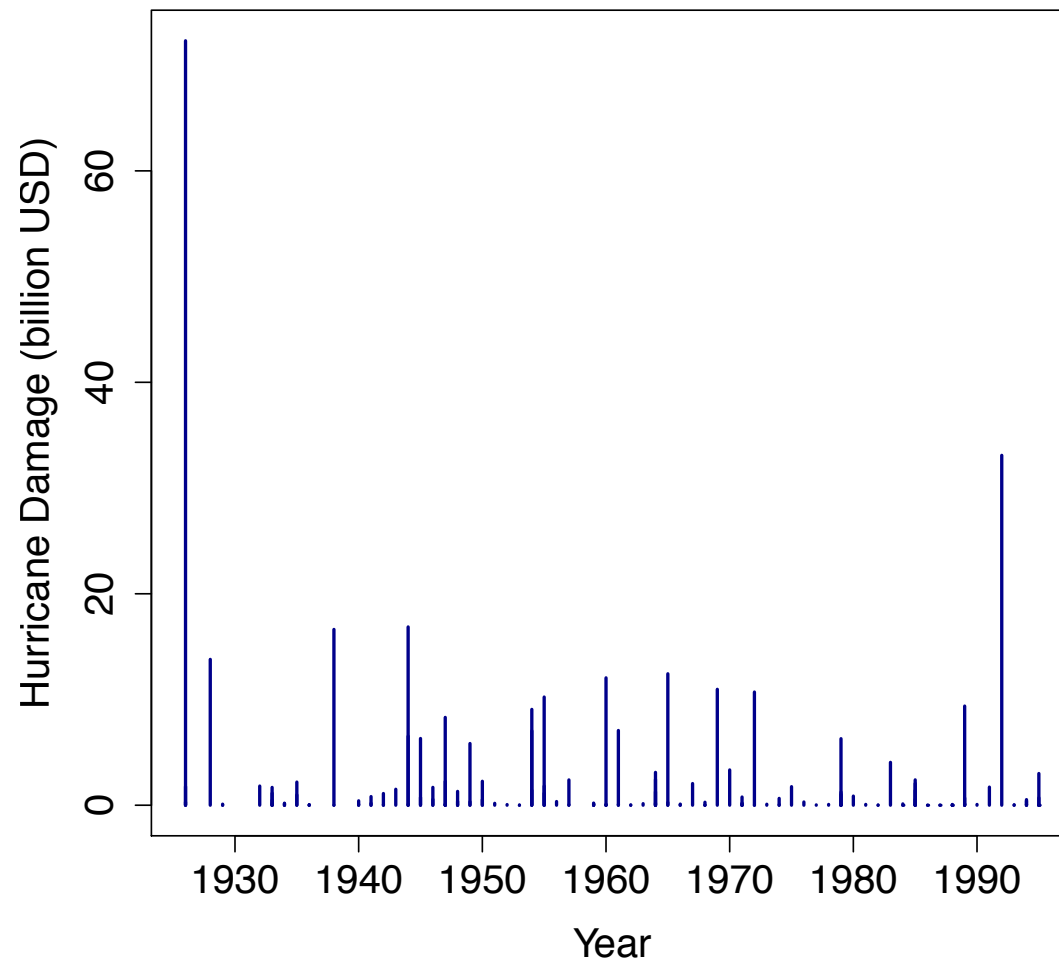
Peaks Over Threshold (POT)

Connection between GEV and GP families

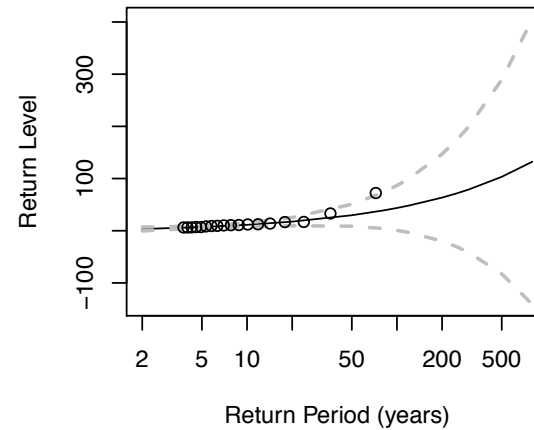
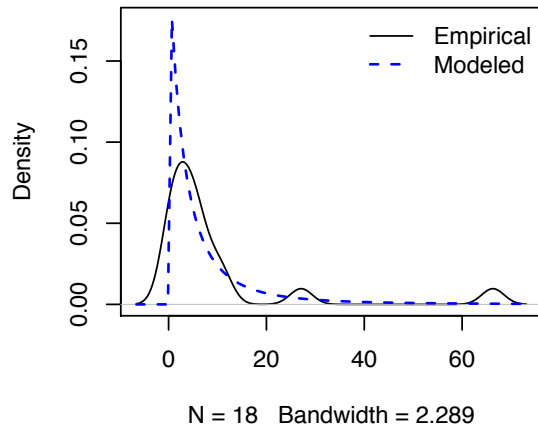
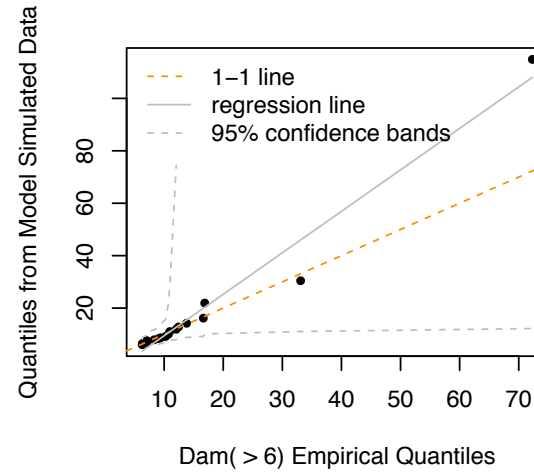
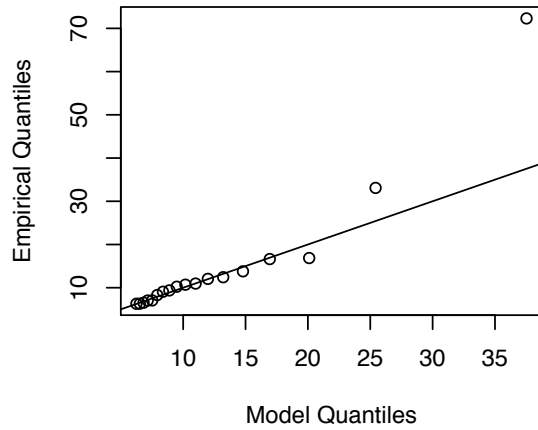
- Maximum $M_n \leq u$ if no $X_i > u$, $i = 1, 2, \dots, n$
- “Memoryless” property of exponential distribution
 $\Pr\{ Y > y + y' \mid Y > y' \} = \Pr\{ Y > y \}$
- Stability of GP distribution
 - Lose memoryless property (need to rescale)
 - If $Y = X - u$ ($X > u$) has exact GP distribution with parameters $\sigma(u) > 0$ and ξ , then excess over a higher threshold $u' > u$ follows the GP distribution with parameters $\sigma(u') > 0$ and ξ , where

$$\sigma(u') = \sigma(u) + \xi(u' - u), \quad u' > u$$

Peaks Over Threshold (POT)



Peaks Over Threshold (POT)



Peaks Over Threshold (POT)



	95% lower CI	Estimate	95% upper CI
$\sigma(u = 6 \text{ billion USD})$	1.03 billion USD	4.59 billion USD	8.15 billion USD
ξ	-0.15	0.51	1.18
100-year return level (billion USD)	0.65 billion USD	43.64 billion USD	86.64 billion USD

Peaks Over Threshold (POT): GP return levels



- First need quantile of GP cdf
 - $x_p = H^{-1}(1 - p; \sigma(u), \xi)$
 - $= (\sigma(u) / \xi) \times (p^{-\xi} - 1), 0 < p < 1$
- Complication: must account for the rate, ζ , of exceeding the threshold for interpretability, as well as the number of points per year, n_y .
 - replace p^{-1} with
 $m = \text{return period of interest} \times n_y \times \zeta$

Peaks Over Threshold (POT): GP return levels



Hurricane example

Lack of structure in data: some years have no hurricanes, some one, some two, etc.

Reasonable to use an average number of points per year in place of n_y .

Peaks Over Threshold (POT): Threshold Selection



- The Bias-Variance Trade-Off
- Want a high threshold to obtain better GP approximation
- Want a lower threshold for more reliable estimation (more data!)
- Difficult to automate selection

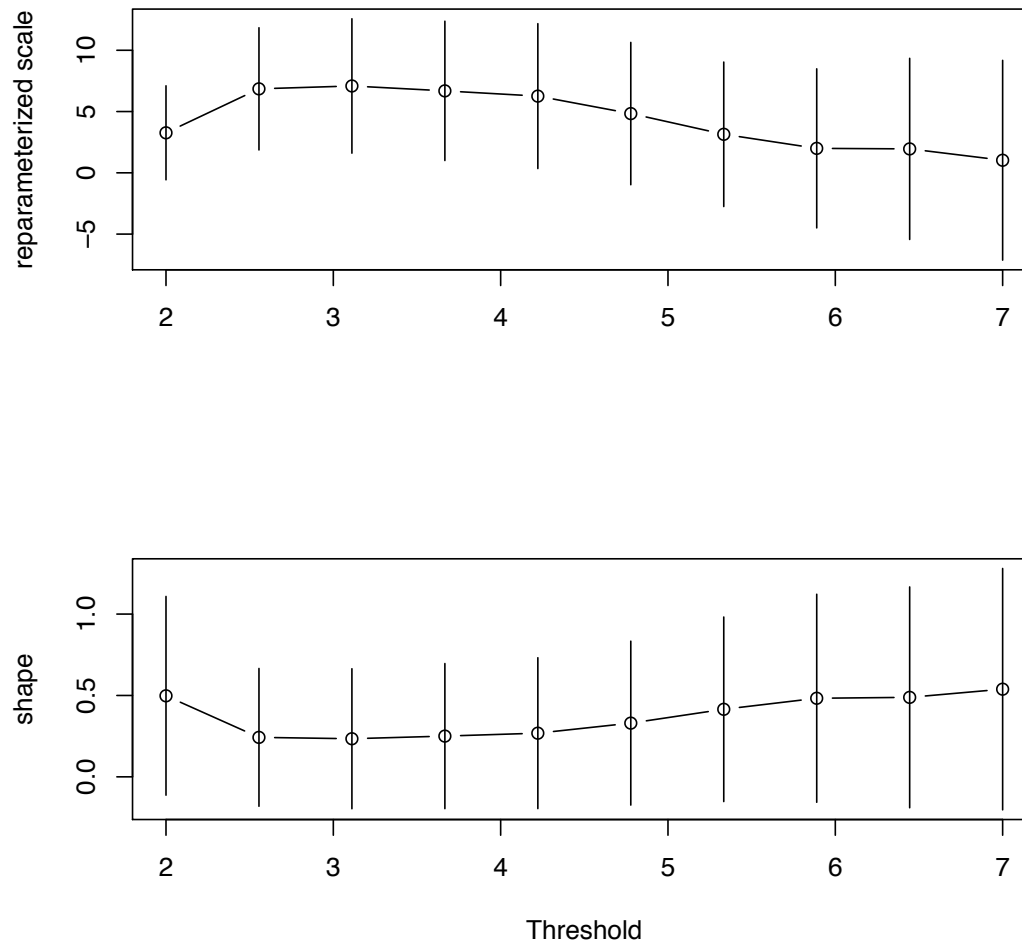
Peaks Over Threshold (POT): Threshold Selection



Invariance of GP above threshold

- ξ does not change
- $\sigma(u)$ is a function of the threshold
- The modification, $\sigma^* = \sigma(u) - \xi u$, is no longer a function of the threshold, and does not change
- Check for stability in parameter estimates as the threshold varies.

Peaks Over Threshold (POT): Threshold Selection



Peaks Over Threshold (POT): Dependence in Threshold Excess Data



- Remove dependence (e.g., decluster)
 - Runs declustering perhaps the simplest
- Model the dependence (e.g., through bivariate extreme value models)
- Concept of “extremal index”

Peaks Over Threshold (POT): Dependence in Threshold Excess Data

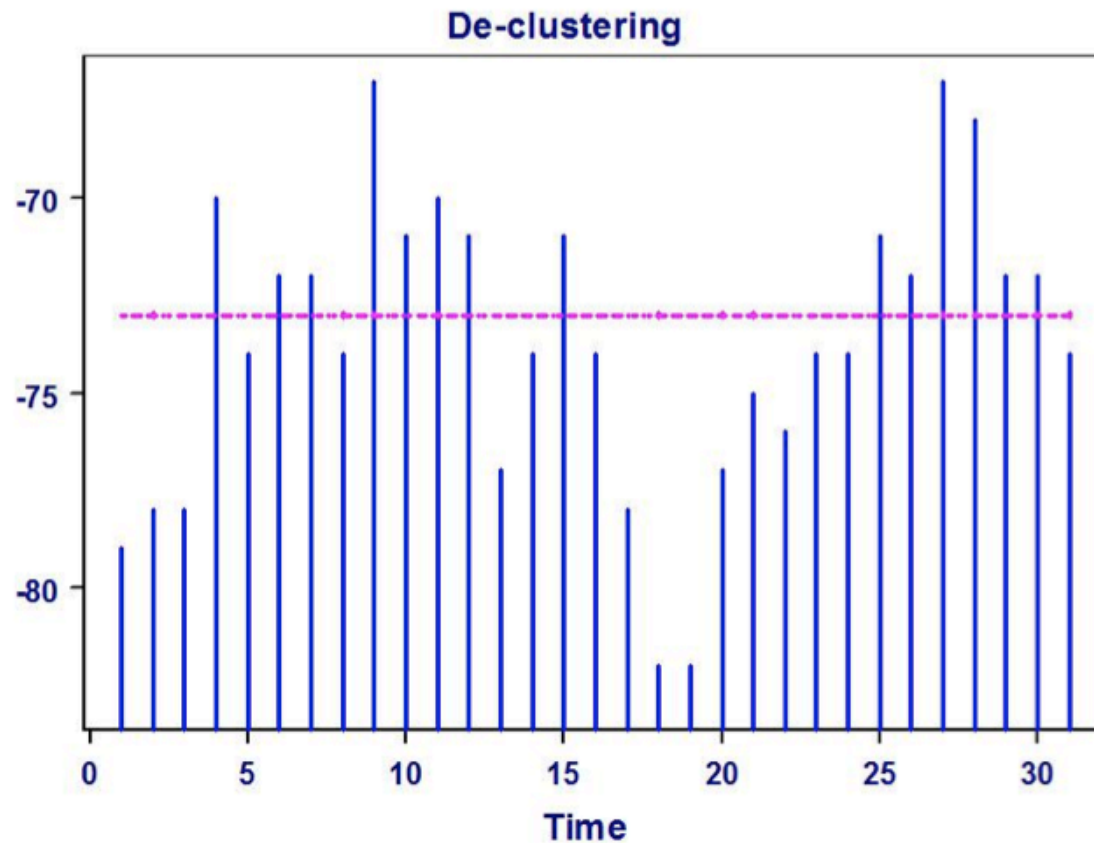


Extremal Index, θ , $0 < \theta \leq 1$

Mean cluster size $\approx 1 / \theta$

$\theta = 1$ implies a lack of clustering at high levels,
and degree of clustering at high levels increases
as θ decreases

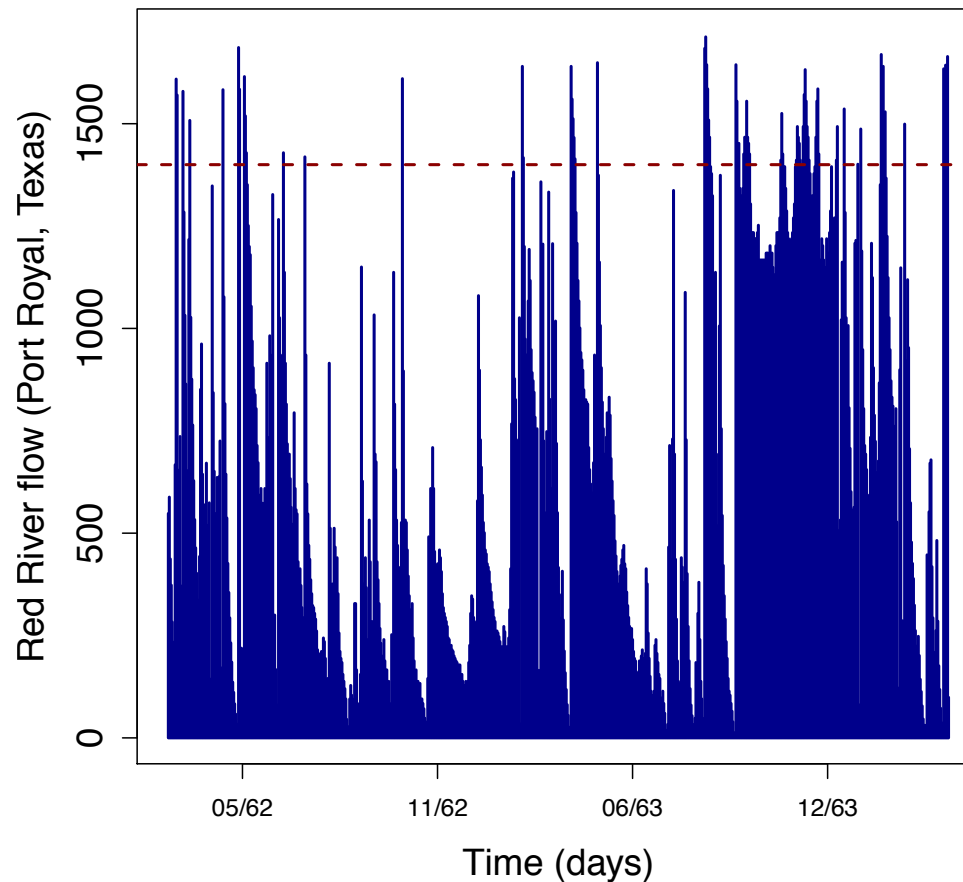
Peaks Over Threshold (POT): Dependence in Threshold Excess Data



Peaks Over Threshold (POT): Dependence in Threshold Excess Data



Full time
range is
1 August
1961 to 28
September
2010

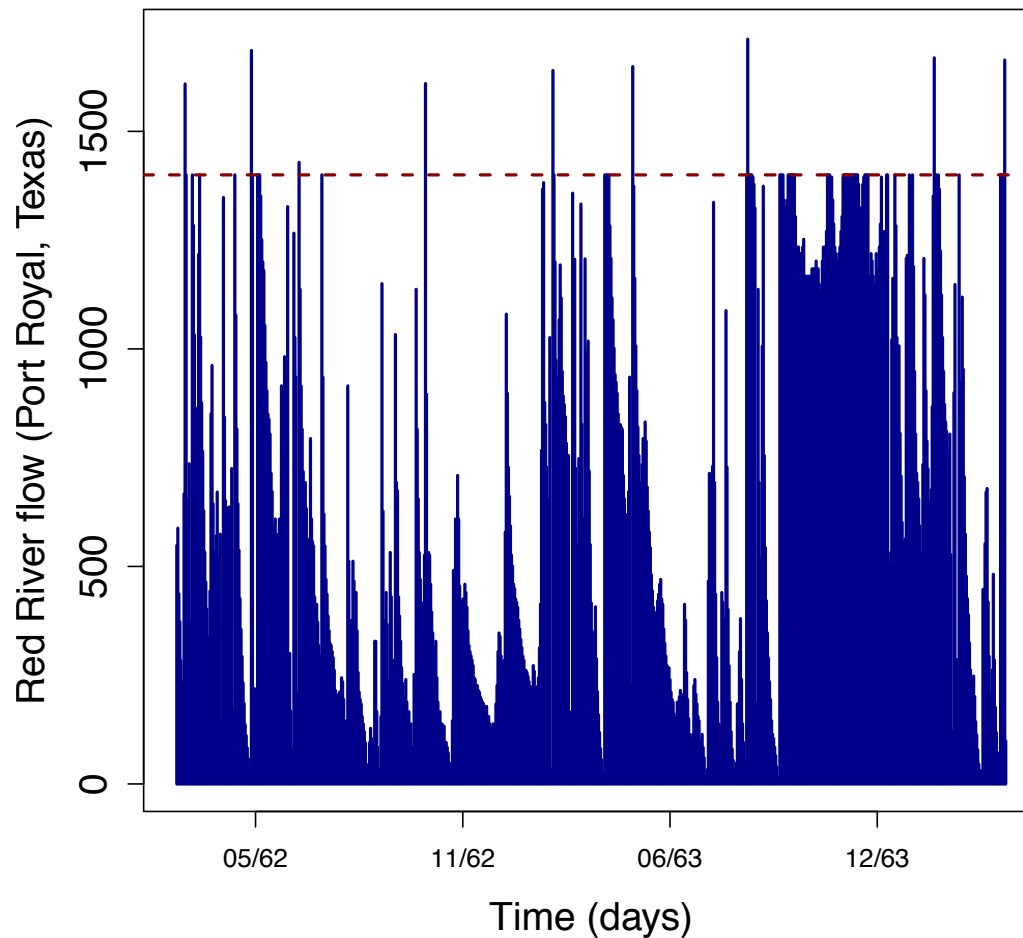


$$\theta \approx 0.08$$

$$r = 29$$

164 clusters

Peaks Over Threshold (POT): Dependence in Threshold Excess Data



$\theta \approx 0.85$

Peaks Over Threshold (POT): Modeling Frequency and Intensity



Poisson-GP Model

- Poisson process for exceeding a high threshold
 - Event: $X_t > u$
 - Rate parameter: λ
 - Number of events in $[0, T]$ has Poisson distribution with parameter λT
- GP distribution for excess over threshold
 - Excess $Y_t = X_t - u$ given $X_t > u$
 - Scale and Shape parameters

Peaks Over Threshold (POT): Modeling Frequency and Intensity



Point Process (PP) representation

- Subsumes Poisson-GP model
- GEV parameterization (connection to GEV)
- Can relate parameters of $GEV(\mu, \sigma, \xi)$ to those of the $PP(\lambda, \sigma(u), \xi)$
 - Shape parameter, ξ , is identical
 - $\ln \lambda = -(1/\xi) \ln(1 + \xi(u - \mu) / \sigma)$
 - $\sigma(u) = \sigma + \xi(u - \mu)$
- Need time scale, h , to take account of the block size for GEV (e.g., $h \approx 1/365.25$ for daily data and annual blocks)

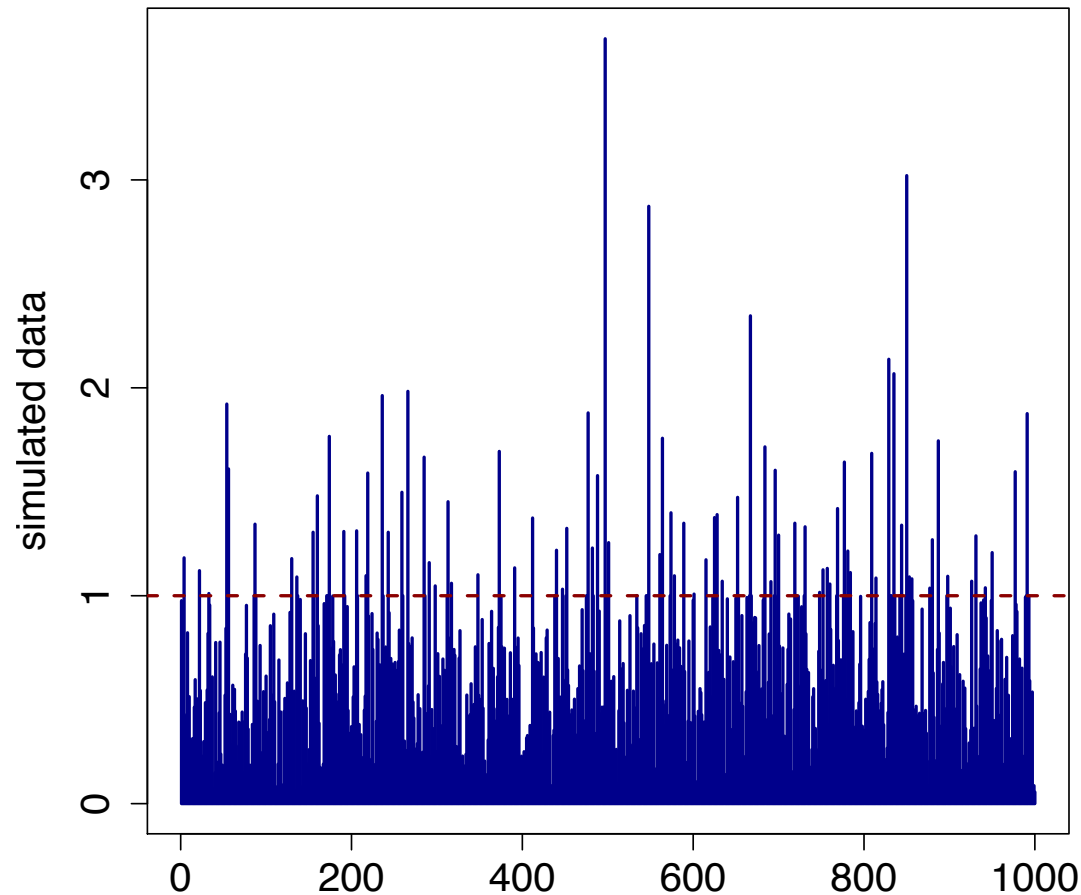
Peaks Over Threshold (POT): Modeling Frequency and Intensity



Point Process (PP) vs Poisson-GP

- Approaches are equivalent
- Poisson-GP model
 - Can obtain the same parameter estimates indirectly through the relationships between the two parameterizations
- Point Process
 - More convenient to quantify total uncertainty
 - Easier to interpret (eliminates dependence of parameters on the threshold)

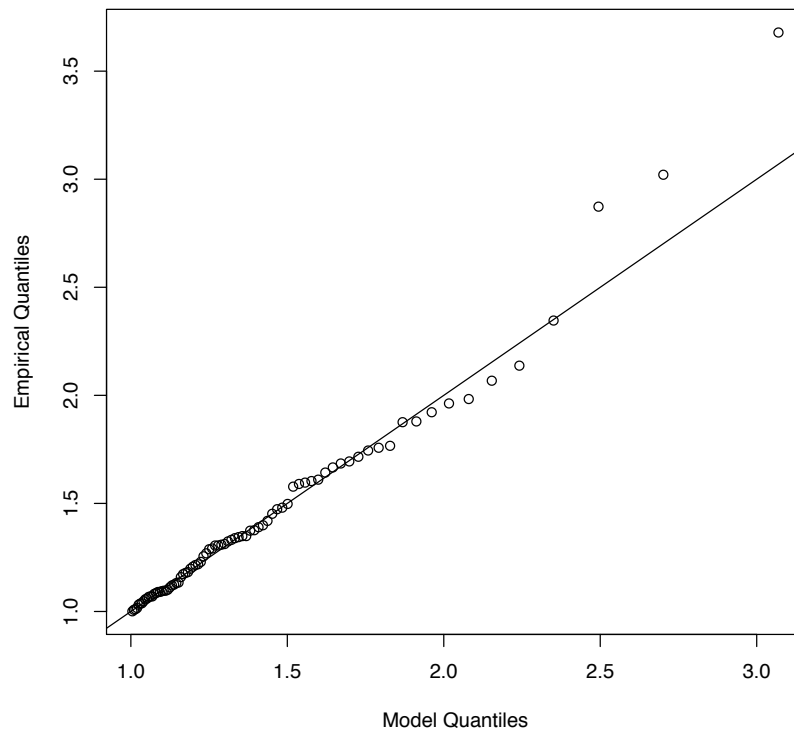
Peaks Over Threshold (POT): Modeling Frequency and Intensity



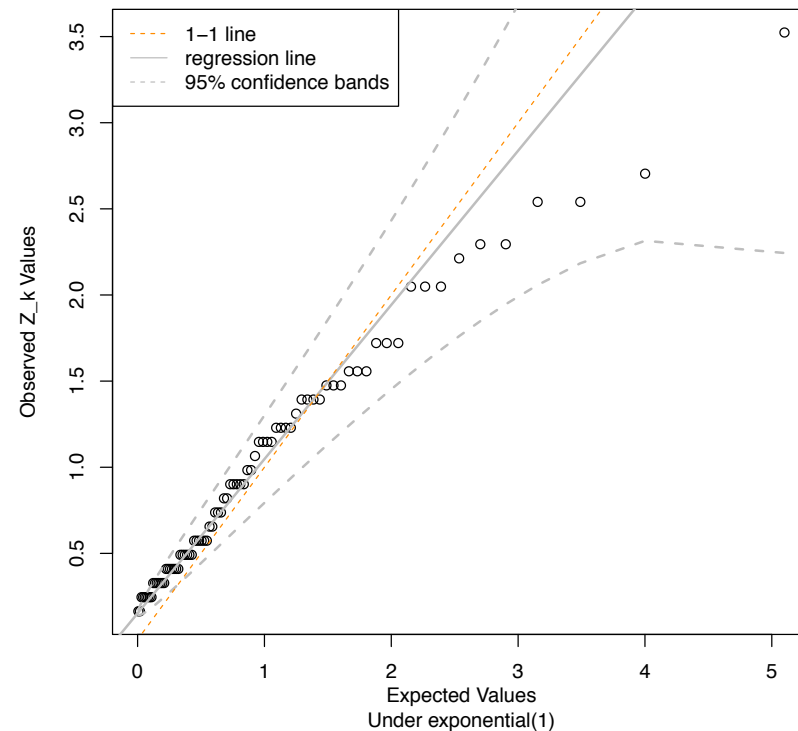
Peaks Over Threshold (POT): Modeling Frequency and Intensity



fevd(x = y, threshold = 1, type = "PP")



Intensity



Frequency

Peaks Over Threshold (POT): Modeling Frequency and Intensity



Poisson-GP estimation method

	95% lower CI	Estimate	95% upper CI
λ	0.07	0.08	0.10
$\sigma(u)$	0.28	0.40	0.53
ξ	-0.16	0.07	0.29

PP estimation method

	95% lower CI	Estimate	95% upper CI
μ	2.06	2.53	3.01
σ	0.20	0.51	0.82
ξ	-0.16	0.07	0.30

Peaks Over Threshold (POT): Modeling Frequency and Intensity



Verify that the two sets of parameter estimates are consistent

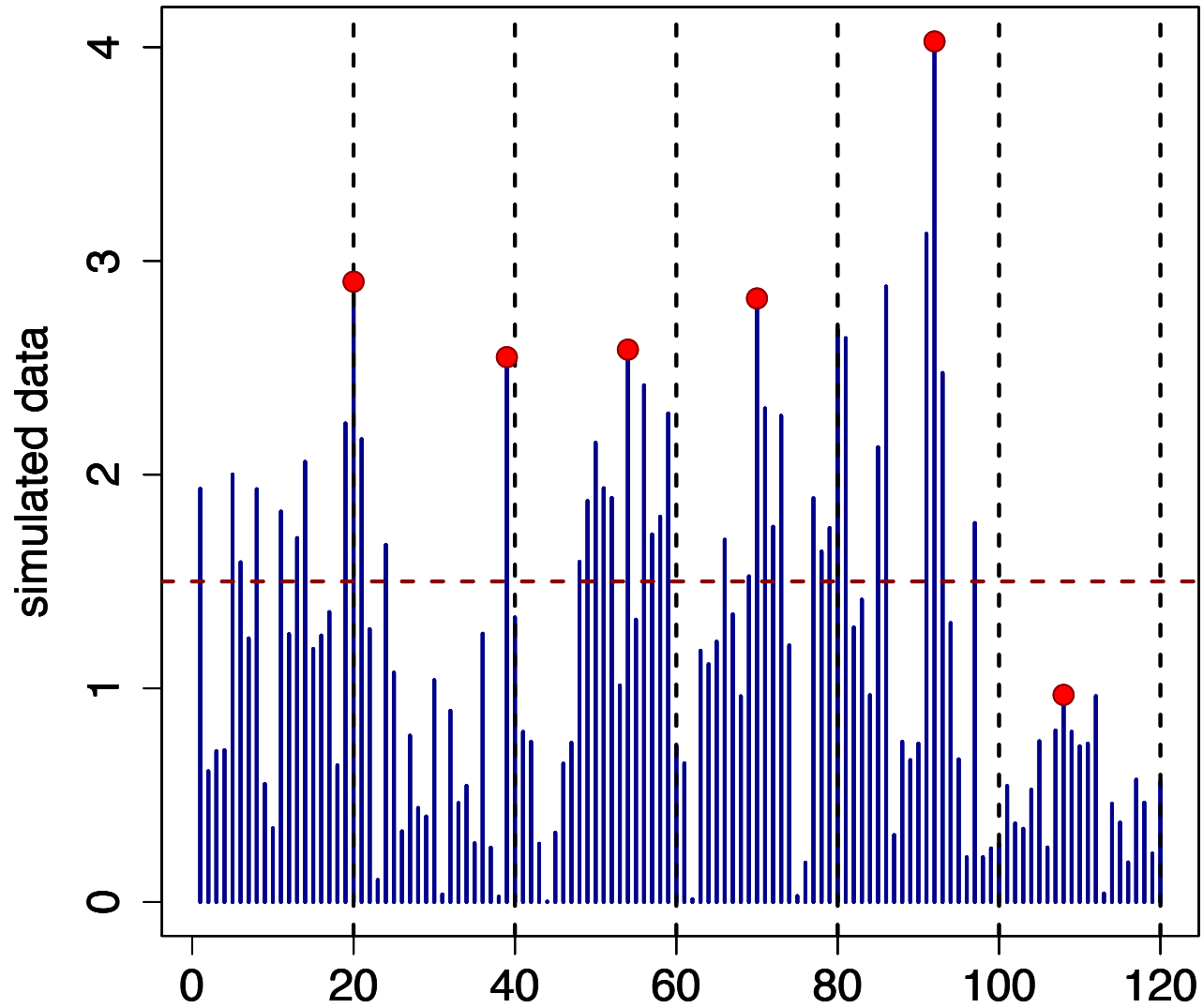
$$\ln \lambda = -(1 / \xi) \ln(1 + \xi(u - \mu) / \sigma) \approx 0.009$$

$$\lambda \approx 29.93 \text{ (estimated on an annual maximum scale)}$$

$$0.08 * 365.25 \approx 29.22$$

$$\sigma(u) = \sigma + \xi(u - \mu) \approx 0.4$$

POT vs Block Maxima





Conclusion of Part II: Questions?