EVA Tutorial #2

PEAKS OVER THRESHOLD APPROACH UNDER NONSTATIONARITY

Rick Katz

Institute for Mathematics Applied to Geosciences
National Center for Atmospheric Research
Boulder, CO USA

email: rwk@ucar.edu

Home page: www.isse.ucar.edu/staff/katz/

Lecture: www.isse.ucar.edu/extremevalues/docs/eva2.pdf

Outline

(1) Tails of Distributions

(2) Return Levels

(3) Choice of Threshold

(4) Declustering

(5) Peaks over Threshold/Point Process Representation

(6) Point Process Approach under Stationarity

(7) Point Process Approach under Nonstationarity

(1) Tails of Distributions

- **Analogue to max stability**

-- X random variable

$Y = X - u$ "excess" over high threshold u, conditional on $X > u$

Then Y has an approximate *generalized Pareto* (GP) distribution for large u with cdf:

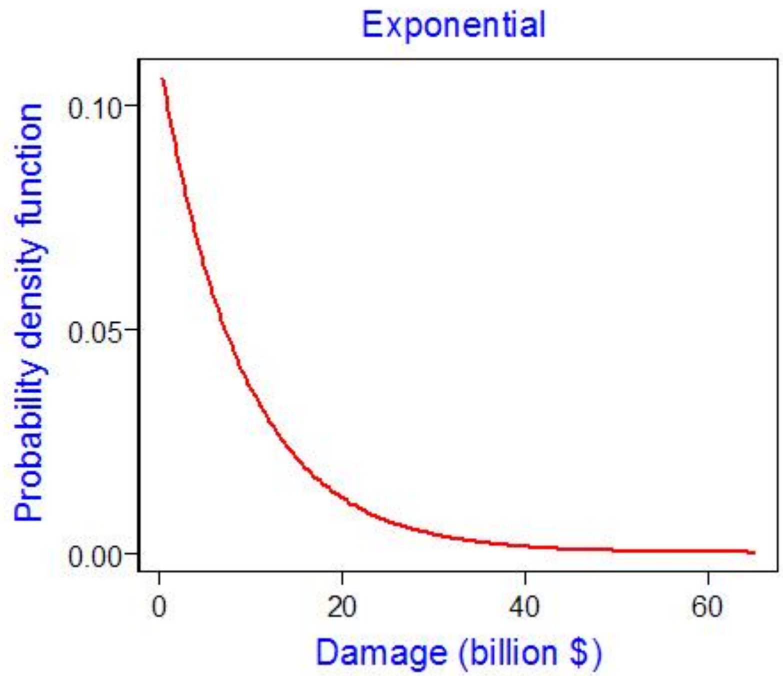
$$H(y;\, \sigma^*, \xi) = 1 - \left[1 + \xi\,(y/\sigma^*)\right]^{-1/\xi},\quad y > 0,\ 1 + \xi\,(y/\sigma^*) > 0$$

$\sigma^* > 0$ scale parameter [alternative notation $\sigma^*(u)$]

ξ shape parameter (same interpretation as for GEV dist.)

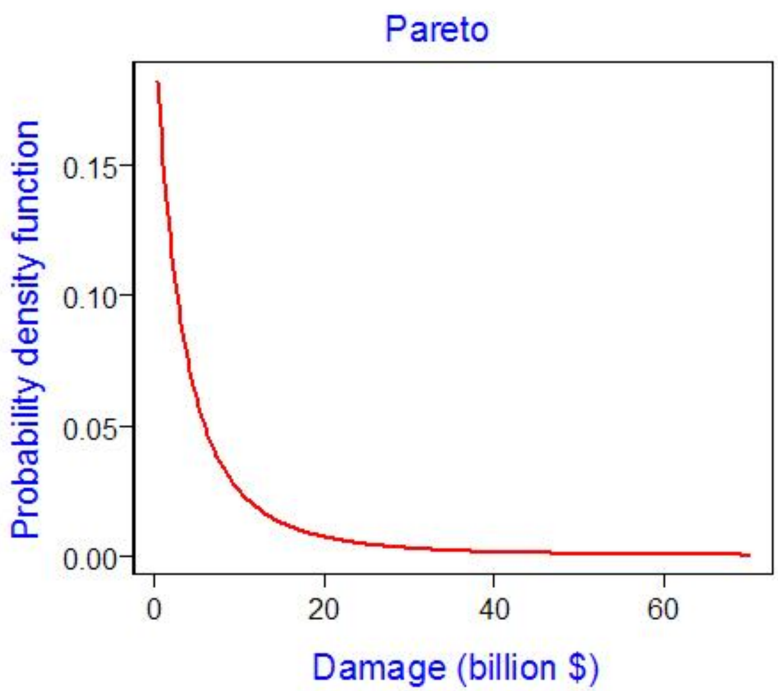
(i) $\xi = 0$ (*exponential type*)

“Light” tail



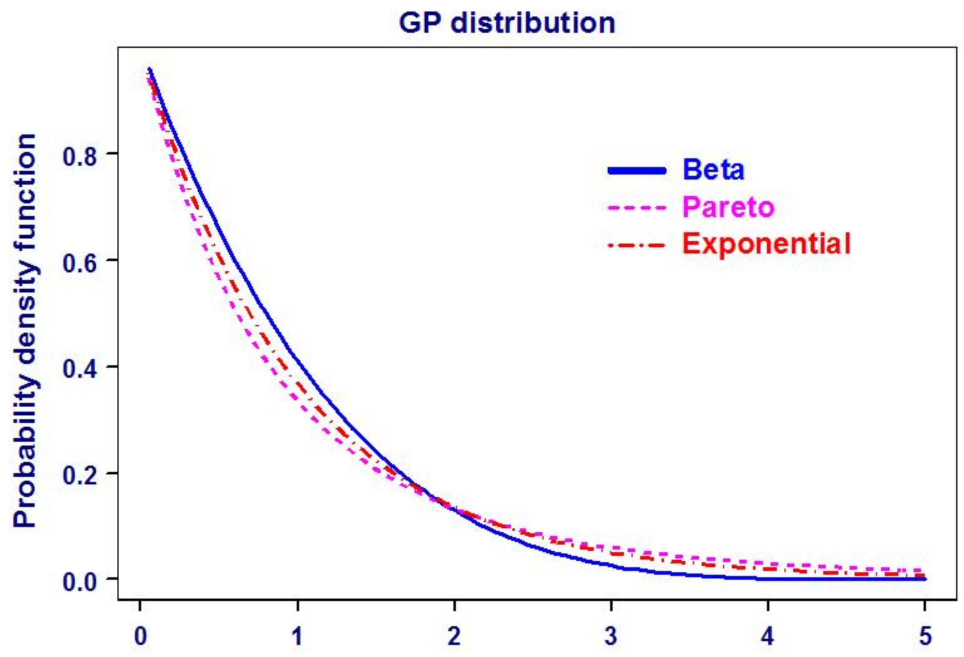
(ii) ξ > 0 (*Pareto type*)

“Heavy” tail



(iii) $\xi < 0$ (*beta type*)

Bounded tail [$y < \sigma^* / (-\xi)$ or $x < u + \sigma^* / (-\xi)$]



- **Connection between GP & GEV**

-- Maximum $M_n \leq u$ if none of X_t's exceeds u, $t = 1, 2, \ldots, n$

-- Assume X_t 's independent with common cdf F

-- Number of exceedances has binomial distribution with parameters:

No. of trials = n

Prob. of "success" = $1 - F(u)$

-- Using Poisson approximation to binomial

$$\Pr\{M_n \leq u\} \approx \exp\{-n\,[1 - F(u)]\}$$

for large n & u such that $n\,[1 - F(u)] \approx$ constant

- **Scaling**

-- “Memoryless” property of exponential distribution (parameter σ*)

$$\Pr\{Y > y + y' \mid Y > y'\} = \Pr\{Y > y\} = \exp[-(y/\sigma^*)], \quad y, y' > 0$$

Suppose *Y* represents life span & has exponential distribution:

Conditional distribution of future survival remains exponential with same scale parameter (No matter how long individual has already survived; i. e., no “aging” or becoming “younger”)

***Example*:**

Waiting time for next bus (if buses arrive at random)

-- Stability of GP distribution

Lose memoryless property (need to rescale)

Suppose excess *Y* over threshold *u* has an exact GP distribution with parameters ξ & σ*(*u*)

Then the excess over a higher threshold *u*' > *u* has GP distribution with parameters ξ & σ*(*u*')

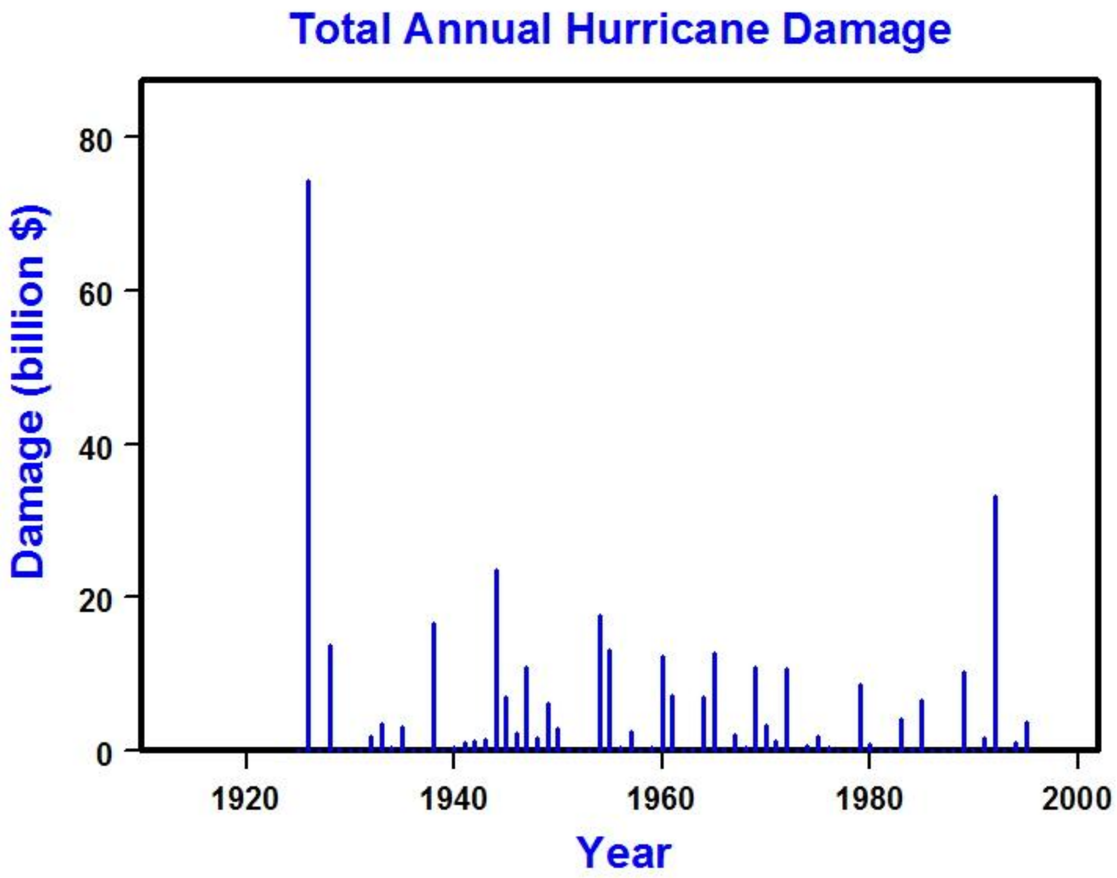
$$\sigma^*(u') = \sigma^*(u) + \xi(u' - u), \quad u' > u$$

(i) $\xi > 0$ implies $\sigma^*(u') > \sigma^*(u)$ (increase "size" of excesses)

(ii) $\xi < 0$ implies $\sigma^*(u') < \sigma^*(u)$ (decrease "size" of excesses)

- **Hurricane damage (1995 US$)**

-- Adjusted data (Remove trends in societal vulnerability), 1925-1995



- **Fit GP distribution to damage from individual storms**

-- Use threshold of *u* = 6 billion $

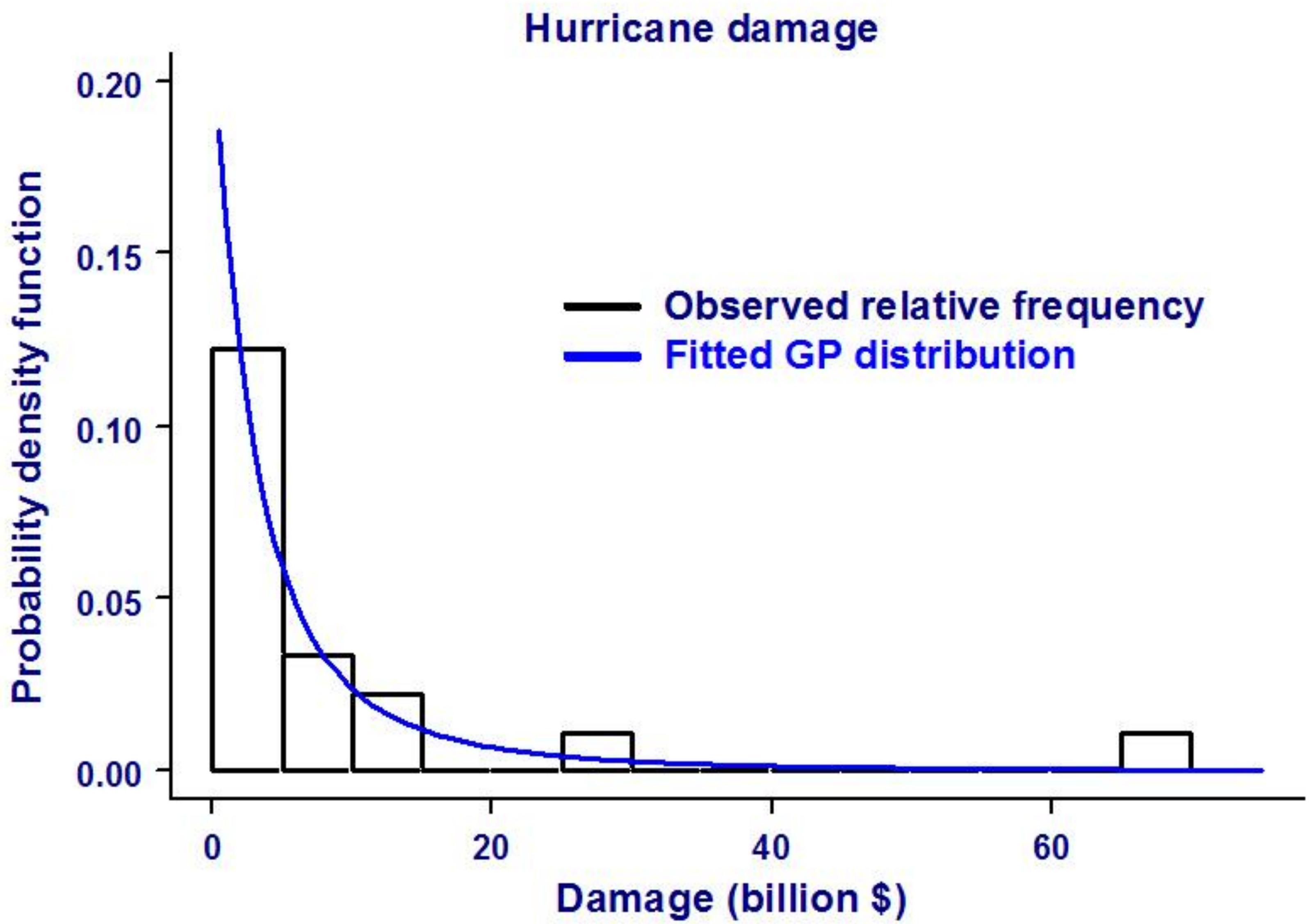
-- Parameter estimates and standard errors

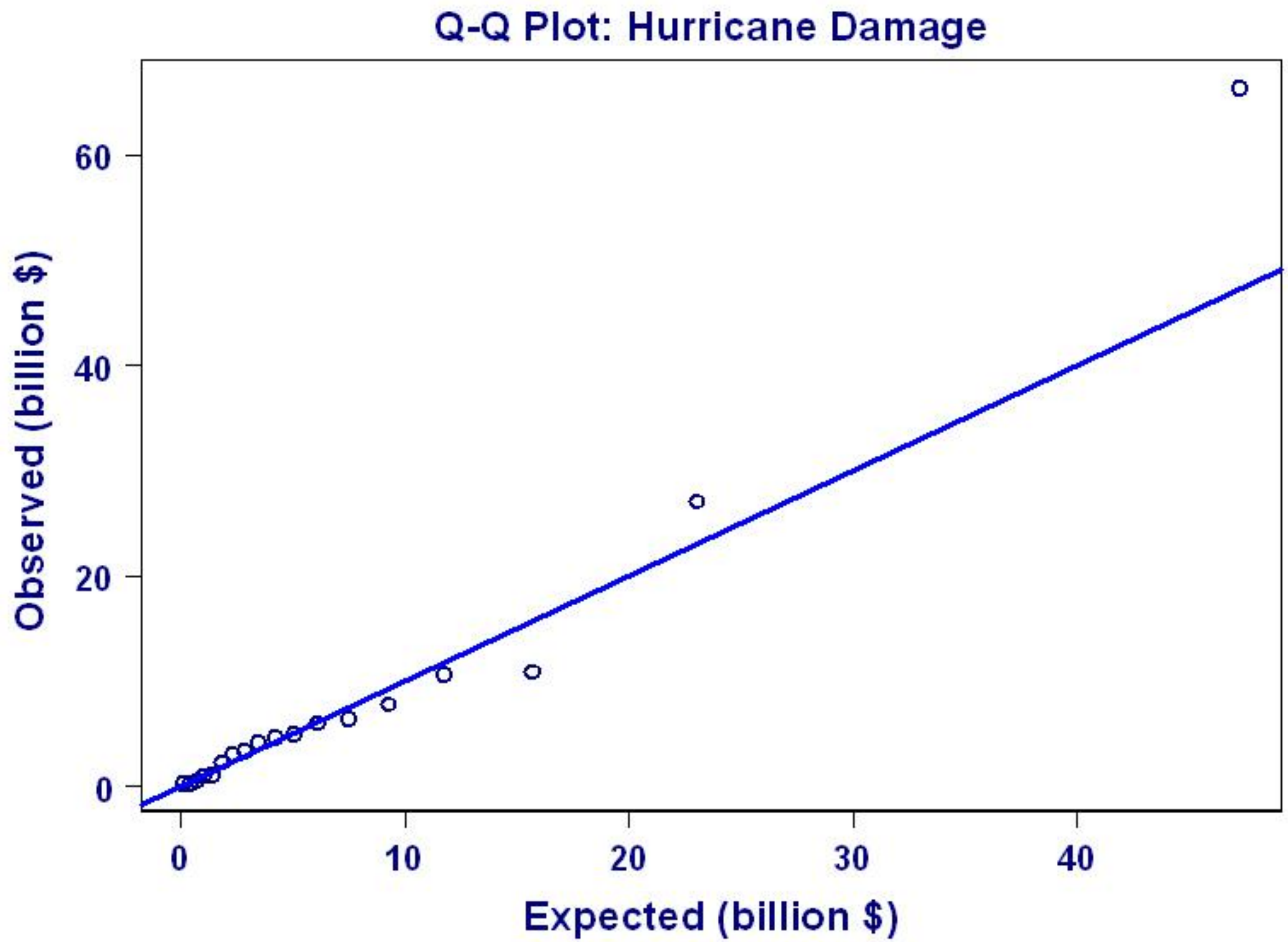
Parameter	Estimate	(Std. Error)
Scale σ^*	4.589	1.817
Shape ξ	0.512	0.341

-- Likelihood ratio test for $\xi = 0$ (*P*-value ≈ 0.018)

-- 95% confidence interval for shape parameter ξ (based on profile likelihood):

$$0.059 < \xi < 1.569$$


Hurricane damage
0.20
0.15
0.10
0.05
0.00
Probability density function
Observed relative frequency
Fitted GP distribution
0
20
40
60
Damage (billion $)


Q-Q Plot: Hurricane Damage
Observed (billion $)
Expected (billion $)
0
20
40
60
0
10
20
30
40

(2) Return Levels

- **GP distribution**

-- Quantile of GP distribution (Invert cdf)

$$H^{-1}(1 - p;\, \sigma^*, \xi) = (\sigma^* / \xi)\,(p^{-\xi} - 1), \quad 0 < p < 1$$

-- Complication

Need to take into account exceedance rate of threshold

(to provide return level corresponding to interpretable return period)

- **Return level calculation**

$$\Pr\{X_t > x\} = \Pr\{X_t > x \mid X_t > u\} \Pr\{X_t > u\}, \quad x > u$$

(i) GP distribution for excess over threshold

$$\Pr\{X_t > x \mid X_t > u\}$$

(ii) Binomial distribution for exceedance rate

$$\Pr\{X_t > u\}$$

Estimate by observed occurrence rate (mle)

Main source of uncertainty arises in estimating (i)

So ignore uncertainty in estimating (ii)

- **Hurricane example**

-- Lack of structure in data file

Reasonable to input average number of damaging hurricanes per year (in place of “no. of obs. per year”)

144 / 71 ≈ 2.03 hurricanes per year

Estimated 20-yr return level: 17.6 billion $

95% confidence interval for 20-year return level (based on profile likelihood):

12.2 billion \$ < x(0.05) < 35.6 billion \$

(3) Choice of Threshold

- **Invariance of GP above threshold**

-- Same shape parameter ξ

-- Reparameterize scale parameter, as threshold *u* varies:

$$\sigma^*(\text{adj}) = \sigma^*(u) - \xi\, u$$

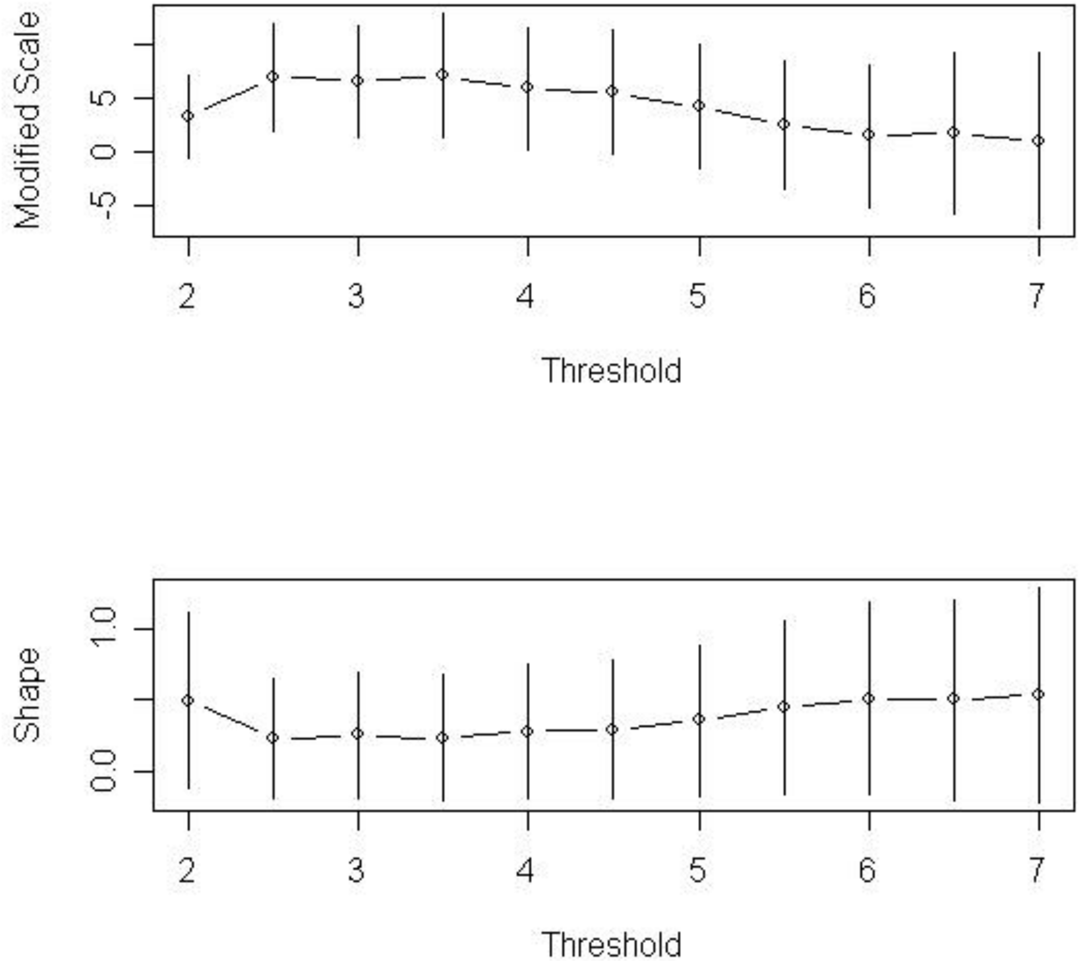
-- Check for stability in parameter estimates as vary threshold

- **Trade-off**

-- Better GP approximation for higher threshold

-- More reliable estimation for lower threshold

-- Lack of automatic procedure



Hurricane damage example

(4) Declustering

(i) Temporal dependence at long time scales

-- GEV and GP approximations still hold for short (or even long memory) processes

***Example*: Stationary Gaussian process**

Autocorrelation function: $\rho_k = \mathrm{Cor}(X_t, X_{t+k}),\ k = 1, 2, \ldots$

No effect on limiting distribution if

$$\rho_k \ln k \to 0 \text{ as } k \to \infty$$

Possible effect in terms of accuracy of approximation

(ii) Temporal dependence at short time scales

- **Concept of lack of "clustering" at high levels**

$$\Pr\{X_{t+k} > u \mid X_t > u\} \rightarrow 0 \text{ as } u \rightarrow \infty,\ k = 1, 2, \ldots$$

- **Stationary Gaussian processes**

-- Lack of clustering at high levels

- **Clustering at high levels**

-- Daily minimum & maximum temperature (strong evidence)

-- Daily precipitation (only weak evidence)

- **Lack of automatic procedure for declustering**

- **Declustering procedures**

-- Runs declustering

Clusters separated by at least *r* consecutive observations below threshold (*r* = 1, 2, . . .)

Model cluster maxima (instead of individual cluster members)

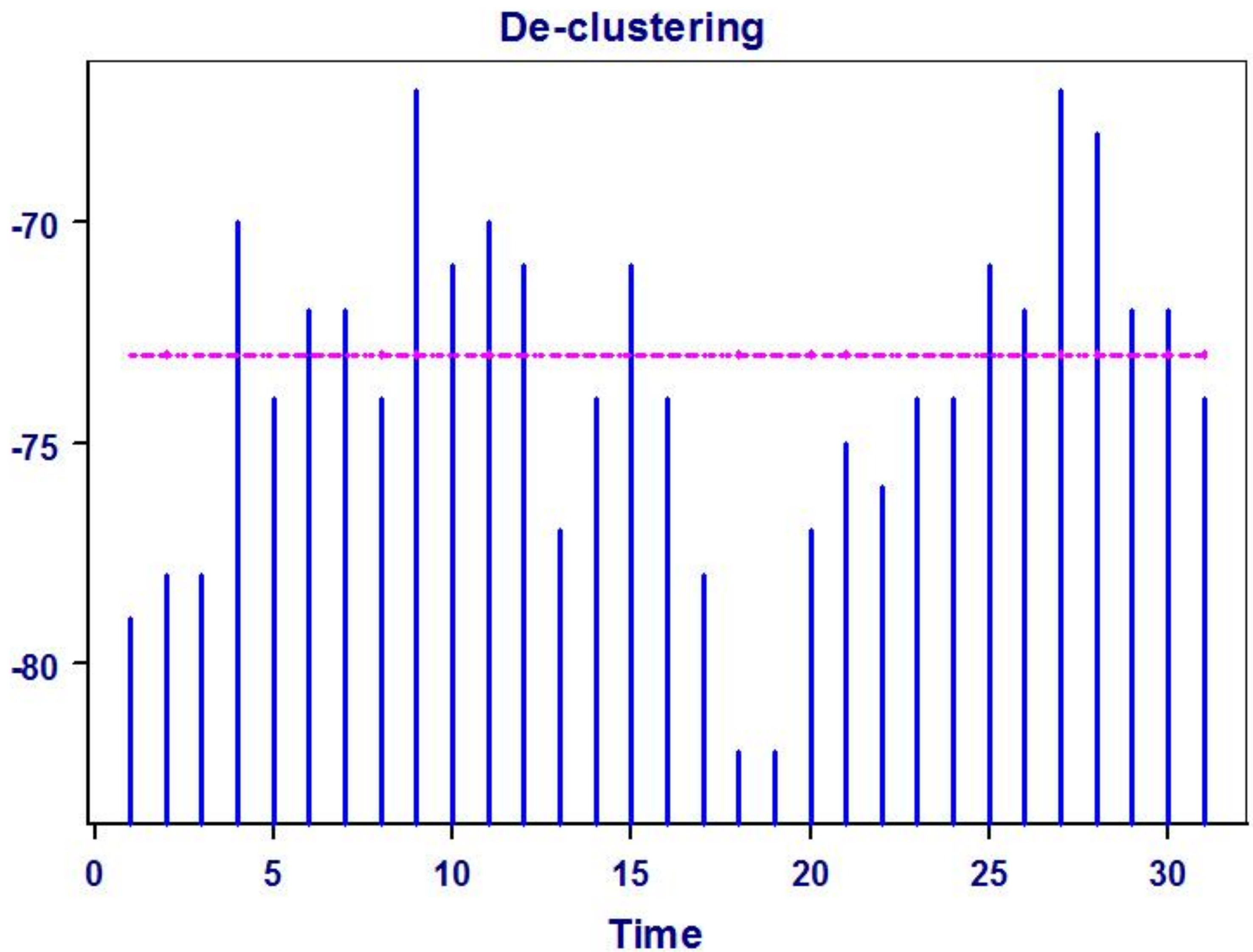
GEV & GP approximations still valid

- **Concept of "extremal index" θ, $0 < \theta \leq 1$**

$$1/\theta \approx \text{mean cluster size}$$

$\theta = 1$ corresponds to lack of clustering at high levels

Degree of clustering at high levels increases as θ decreases


De-clustering
-70
-75
-80
0
5
10
15
20
25
30
Time

- **Phoenix minimum temperature**

-- Phoenix, AZ, USA

Time series of daily minimum temperature (°F) for July-August, 1948-1990

But now consider daily time series, *not* just block minima

-- Lower tail vs. upper tail

Model lower tail as upper tail after negation

So consider $X^* = -X$, where X denotes daily minimum temperature

- **Fit GP distribution to declustered data (ignore any trend for now)**

-- Threshold $u = -73$ °F

-- Use runs declustering ($r = 1$)

De-clustering	Parameter	Estimate	(Std. Error)
None	Scale σ^*	3.915	(0.303)
$r = 1$		4.167	(0.501)
None	Shape ξ	−0.246	(0.049)
$r = 1$		−0.242	(0.079)

-- Mean cluster size

262 / 115 ≈ 2.3 days (very crude estimate of $\theta \approx 0.44$)

- **Alternatives to declustering**

-- Resampling to estimate standard errors

(Still fit GP distribution to original excesses even if clustered)

-- Explicit modeling of temporal dependence at high levels

(e. g., Markov model)

-- Revisit issue in more detail in EVA Tutorial #3

(e. g., method to estimate extremal index that does not require declustering)

(5) Peaks over Threshold/Point Process Representation

- **Rationale**

-- Make more use of information available about upper tail

(even if only interested in obtaining estimate for block maxima)

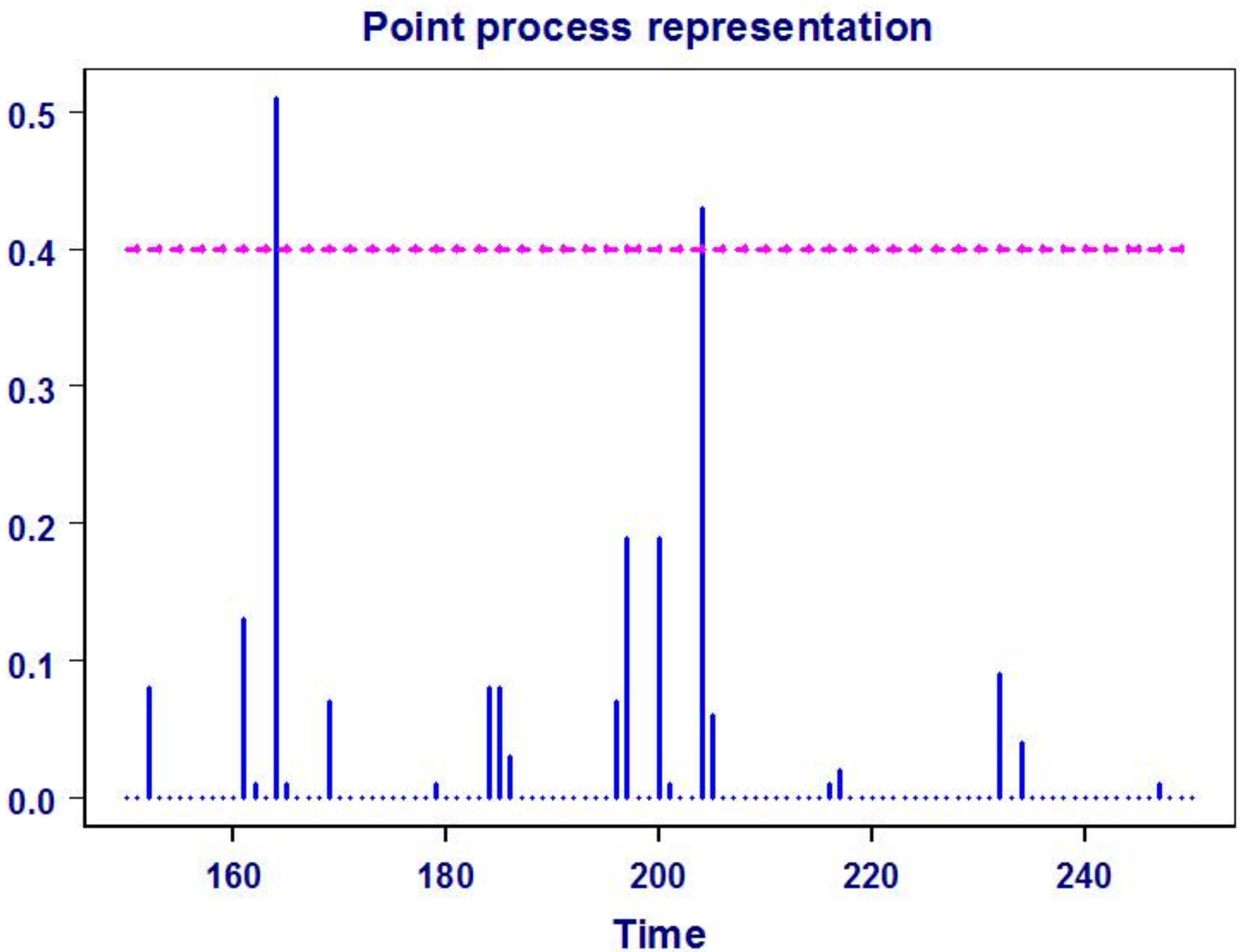
- **Consider process through which extremes arise**

-- *Occurrence*

(e. g., exceedance of high threshold)

-- *Intensity* (or severity)

(e. g., excess over threshold)



Point process representation
0.5
0.4
0.3
0.2
0.1
0.0
160
180
200
220
240
Time

(i) Poisson-GP Model

"Peaks over Threshold" or "Partial Duration Series"

- **Poisson process for exceedance of high threshold**

-- Event $X_t > u$

Rate parameter $\lambda > 0$

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 $X_t > u$

Parameters ξ & σ^*

(ii) Point process representation

-- Occurrence & intensity of extreme events

View as points in two-dimensional space

As threshold $u \longrightarrow \infty$, converges to two-dimensional, non-homogenous Poisson process (i. e., non-constant rate parameter)

Non-homogenous in vertical dimension because higher excesses less likely

Subsumes Poisson-GP model

Also connection to GEV distribution

-- GEV parameterization

Can relate parameters of GEV(μ, σ, ξ) to parameters of point process (λ, σ^*, ξ):

(i) Shape parameter ξ identical

(ii) $\ln \lambda = -(1/\xi) \ln[1 + \xi(u - \mu)/\sigma]$

(iii) $\sigma^* = \sigma + \xi(u - \mu)$

Change of block size for GEV distribution:

Time scaling constant *h* (annual max. of daily data, $h \approx 1/365.25$)

Time scale *h* for GEV(μ, σ, ξ) to time scale *h*' for GEV(μ', σ', ξ)

$$\sigma' = \sigma\, \delta^{\xi}, \quad \mu' = \mu + [\sigma'(1 - \delta^{-\xi})] / \xi, \quad \text{where } \delta = h/h'$$

- **Two different approaches to parameter estimation**

-- Poisson-GP (or orthogonal) approach

Fit Poisson & GP components separately

Convenient for estimation

More difficult to interpret (especially with covariates)

-- Point process (with GEV parameterization)

Fit two-dimension, non-homogenous Poisson process

More difficult to estimate

Easier to interpret (especially with covariates)

(6) Point Process Approach under Stationarity

- **Stationarity**

-- Poisson-GP and Point Process equivalent

Can obtain same parameter estimates (indirectly through use of relationships between two parameterizations)

-- Point process approach still has advantages

Scale parameter invariant with respect to threshold

Avoid need to combine uncertainty from two components as in Poisson-GP model

- **Fort Collins daily precipitation**

-- Now analyze daily data instead of only annual maxima (but ignore annual cycle for now)

Estimation Method	Parameter	Estimate
(i) Orthogonal	Rate λ	10.6 per yr.
(u = 0.395 in)	Scale σ^*	0.322
	Shape ξ	0.212
(ii) Point process	Location μ	1.383
(u = 0.395 in,	Scale σ	0.532
h = 1/365.25)	Shape ξ	0.212

-- Verify that two sets of parameter estimates are consistent

$\ln \lambda = -(1/\xi) \ln[1 + \xi(u - \mu)/\sigma] \approx 2.360$ [vs. $\ln(10.6) \approx 2.361$]

$\sigma^* = \sigma + \xi(u - \mu) \approx 0.323$ (vs. 0.322)

- **Diagnostics for point process approach**

-- Indirectly fitted GEV distribution for annual maxima

So one diagnostic would be Q-Q plot for annual maxima

(Looks fairly similar to one already shown for GEV distribution fitted directly to annual maxima)

(7) Point Process Approach under Nonstationarity

- **Flexibility**

-- Introduce annual cycles or other covariates on daily time scale

- **Poisson-GP model**

-- Introduce nonstationarity separately into two components

***Occurrence*: Poisson process with rate parameter $\lambda(t)$**

***Excess*: GP distribution with parameters $\sigma^*(t)$ & $\xi(t)$**

- **Point process approach**

-- Introduce nonstationarity in GEV parameters $\mu(t)$, $\sigma(t)$, $\xi(t)$

Threshold $u(t)$ can be time varying as well

- **Fort Collins precipitation example**

-- Threshold u = 0.395 in (could be time varying as well)

Length of year $T \approx 365.25$ days

-- Poisson-GP model

(i) Annual cycle in Poisson rate parameter

$$\ln \lambda(t) = \lambda_0 + \lambda_1 \sin(2\pi t / T) + \lambda_2 \cos(2\pi t / T)$$

Parameter		Estimate	(Std. Error)
Rate:	λ_0	−3.721	
	λ_1	0.221	(0.045)
	λ_2	−0.846	(0.049)

LRT for $\lambda_1 = \lambda_2 = 0$ (P-value ≈ 0)

(ii) Annual cycle in scale parameter of GP distribution

$$\ln \sigma^*(t) = \sigma_0^* + \sigma_1^* \sin(2\pi t / T) + \sigma_2^* \cos(2\pi t / T)$$

Parameter		Estimate	(Std. Error)
Scale:	σ_0^*	−1.238	
	σ_1^*	0.088	(0.048)
	σ_2^*	−0.303	(0.069)
Shape	ξ	0.181	

Likelihood ratio test for $\sigma_1^* = \sigma_2^* = 0$ (P-value $< 10^{-5}$)

Q-Q plot: Transform non-stationary GP to exponential dist.

$$\varepsilon_t = [1/\xi(t)] \ln\{1 + \xi(t) [Y_t / \sigma^*(t)]\}$$

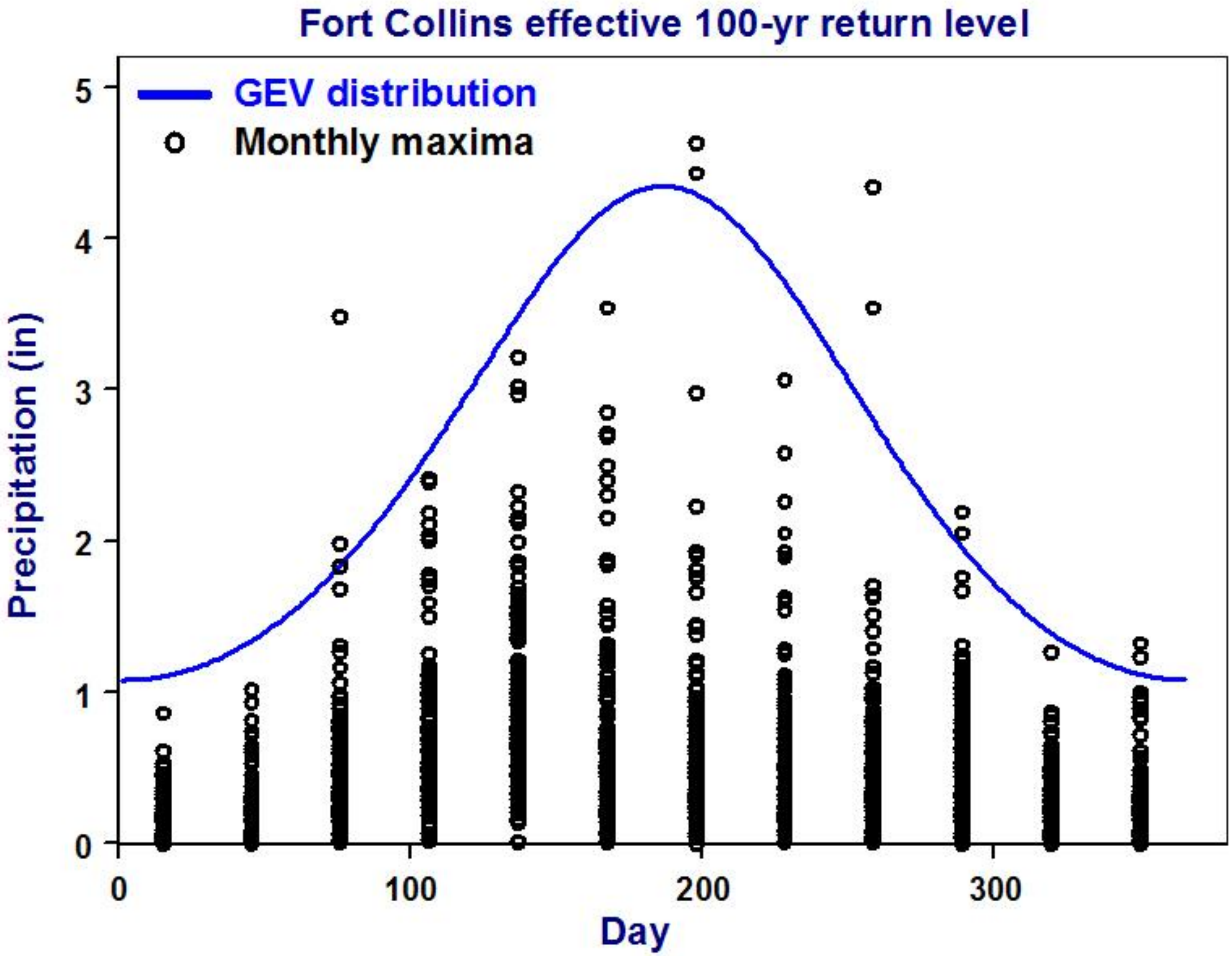
-- Point process approach (u = 0.395 in, h = 1/365.25)

Annual cycles in location & scale parameters of GEV distribution:

$$\mu(t) = \mu_0 + \mu_1 \sin(2\pi t / T) + \mu_2 \cos(2\pi t / T)$$

$$\ln \sigma(t) = \sigma_0 + \sigma_1 \sin(2\pi t / T) + \sigma_2 \cos(2\pi t / T)$$

Parameter		Estimate	(Std. Error)	LRT
Location:	μ_0	1.281		
	μ_1	−0.085	(0.031)	$\mu_1 = \mu_2 = 0$
	μ_2	−0.806	(0.043)	(P-value ≈ 0)
Scale:	σ_0	−0.847		
	σ_1	−0.123	(0.028)	$\sigma_1 = \sigma_2 = 0$
	σ_2	−0.602	(0.034)	(P-value ≈ 0)
Shape	ξ	0.182		


Fort Collins effective 100-yr return level
GEV distribution
Monthly maxima
Precipitation (in)
Day
0
1
2
3
4
5
0
100
200
300

- **Notes about fitted point process model**

-- Time varying threshold *u*(*t*) (i. e., higher in summer than in winter) would be more appropriate, but does not affect results much

-- Declustering (*r* = 1) does not affect results much (except for larger standard errors)

-- Return levels for annual maxima

Assume independence on daily time scale

Then can calculate return levels for annual maxima from GEV distribution with seasonally varying parameters

(But has little effect on return level estimates for annual maxima)

- **Comparison of Poisson-GP and point process models**

-- Lack of equivalence

Sine waves for parameters of Poisson and GP components:

Do *not* necessarily correspond to sine waves for GEV parameters in point process model

Annual cycle in terms of GEV parameters is more interpretable theoretically (arguably)

Homework

Generate pseudo random variables $X_1, X_2, \ldots, X_n$ iid $N(0, 1)$ and record $\max\{X_1, X_2, \ldots, X_n\}$, say with block size $n = 100$.

Repeat to obtain $T = 10{,}000$ sample maxima and fit GEV distribution to these data.

Are the results obtained consistent with the normal distribution being in the domain of attraction of the Gumbel (i. e., shape parameter $\xi = 0$)?