STATISTICAL METHODS FOR RELATING TEMPERATURE EXTREMES TO LARGE-SCALE METEOROLOGICAL PATTERNS

Rick Katz

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

(with help from Richard Grotjahn)

email: rwk@ucar.edu

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

Talk: www.isse.ucar.edu/staff/katz/docs/pdf/clivarrk.pdf

Outline

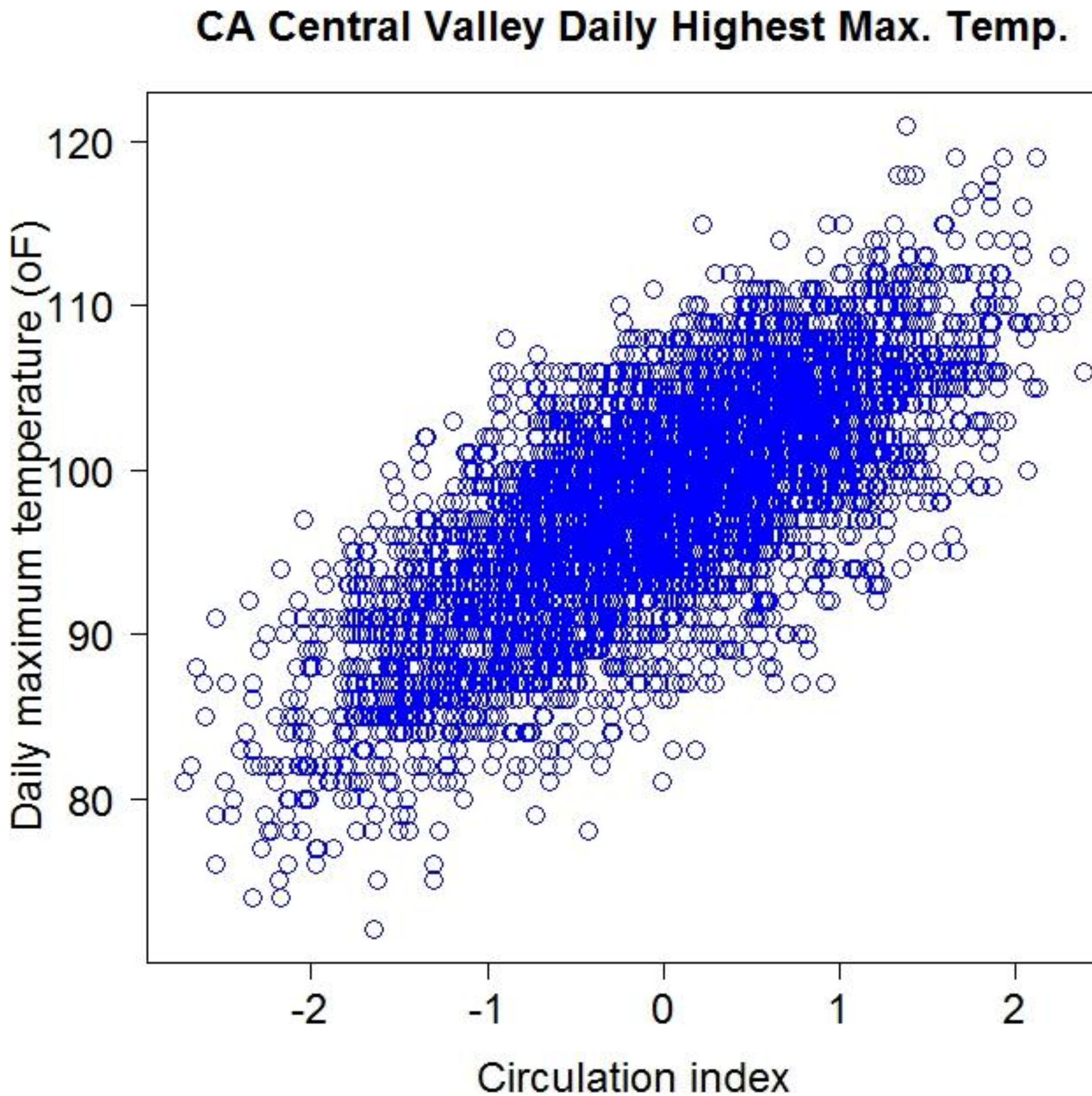
(1) Introduction

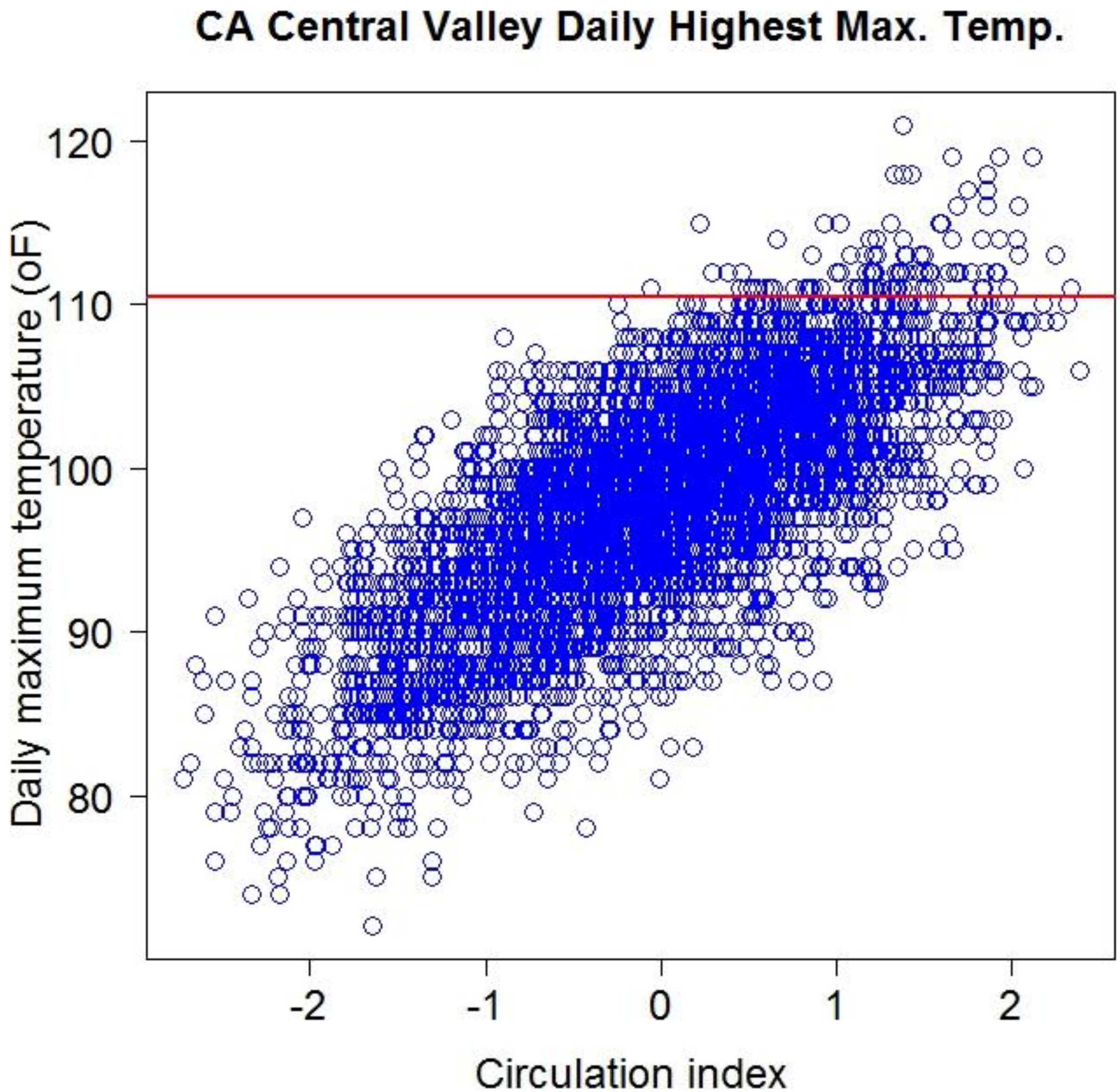
(2) Conditional Extreme Value Analysis: Block Maxima

(3) Conditional Extreme Value Analysis: Peaks over Threshold

(4) Application to California Temperature Extremes

(5) Remaining Work


CA Central Valley Daily Highest Max. Temp.
Daily maximum temperature (oF)
120
110
100
90
80
-2
-1
0
1
2
Circulation index


CA Central Valley Daily Highest Max. Temp.
Daily maximum temperature (oF)
120
110
100
90
80
-2
-1
0
1
2
Circulation index

(1) Introduction

- **Conditional extreme value analysis**

-- Block maxima approach

Condition extreme value distributions on covariate such as ENSO (e. g., on seasonal time scale)

-- Peaks over threshold approach

Condition extreme value distributions on index of large-scale meteorological circulation patterns (e. g., on daily time scale)

-- Software

e. g., "extRemes" package in R

(2) Conditional Extreme Value Analysis: Block Maxima

- **Extremal Types Theorem ("Max stability")**

-- Suppose limiting distribution of maximum of a sequence of random variables (suitably normalized) has cumulative distribution function (cdf) *G*

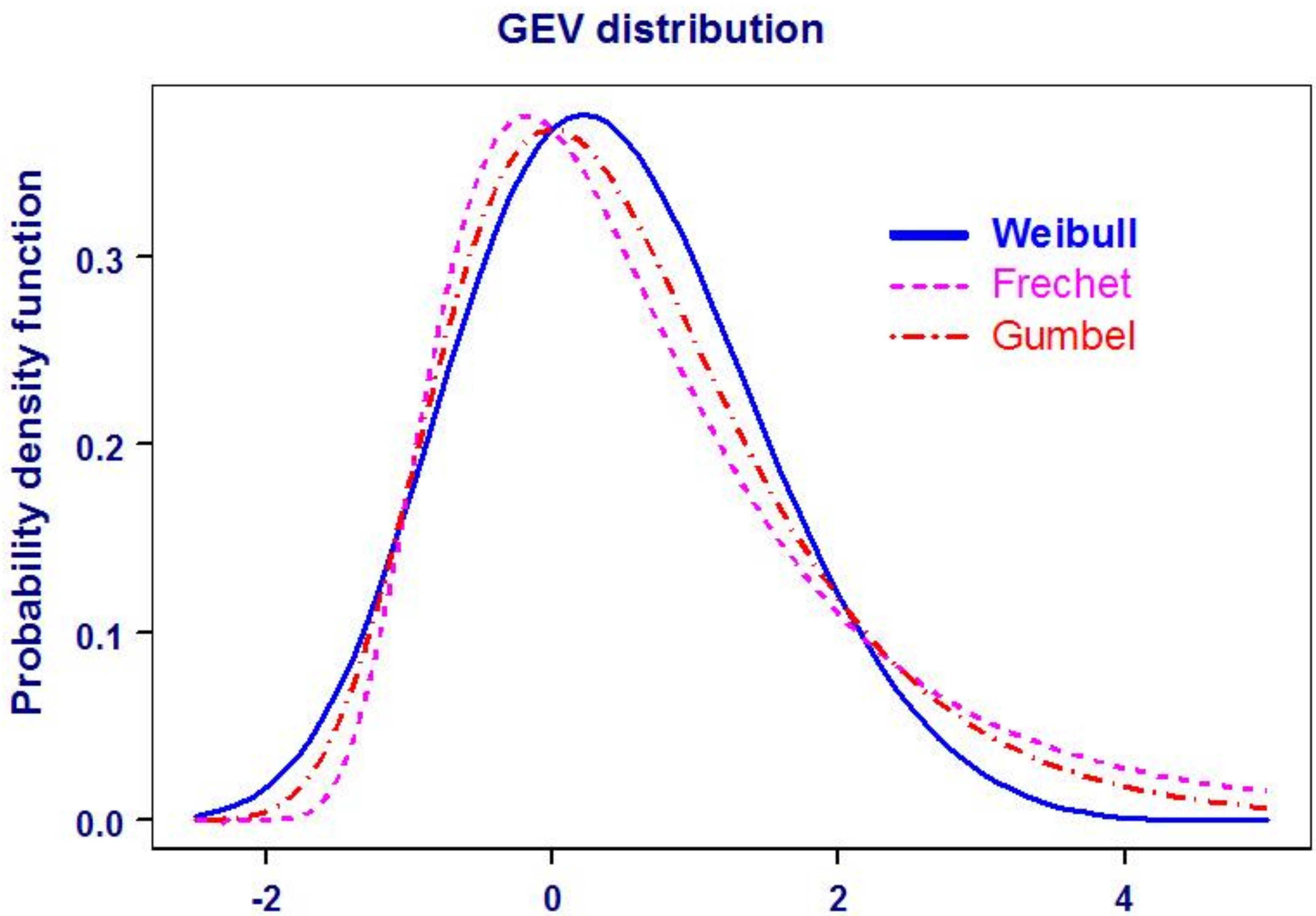
-- Then *G* must be *generalized extreme value* (GEV) cdf; that is,

$$G(x; \mu, \sigma, \xi) = \exp\left\{-[1 + \xi (x - \mu)/\sigma]^{-1/\xi}\right\},\ 1 + \xi (x - \mu)/\sigma > 0$$

μ location parameter ("center")

σ > 0 scale parameter ("spread")

ξ shape parameter


GEV distribution
Probability density function
0.0
0.1
0.2
0.3
-2
0
2
4
Weibull
Frechet
Gumbel

(i) Shape parameter $\xi < 0$ (*Weibull type*)

Bounded upper tail

Fits block maxima of temperature

(ii) Shape parameter $\xi > 0$ (*Fréchet type*)

“Heavy” upper tail

Fits block maxima of precipitation

- **Example of conditional analysis**

-- Winter maximum temperature at Port Jervis, NY, USA

Covariate *Z*: Winter index of Arctic Oscillation (AO)

(See Wettstein & Mearns, *J. Climate*, 2002)

Given *Z* = *z*, assume conditional distribution of winter maximum temperature is GEV with parameters:

$$\mu(z) = \mu_0 + \mu_1 z$$

$$\ln \sigma(z) = \sigma_0 + \sigma_1 z$$

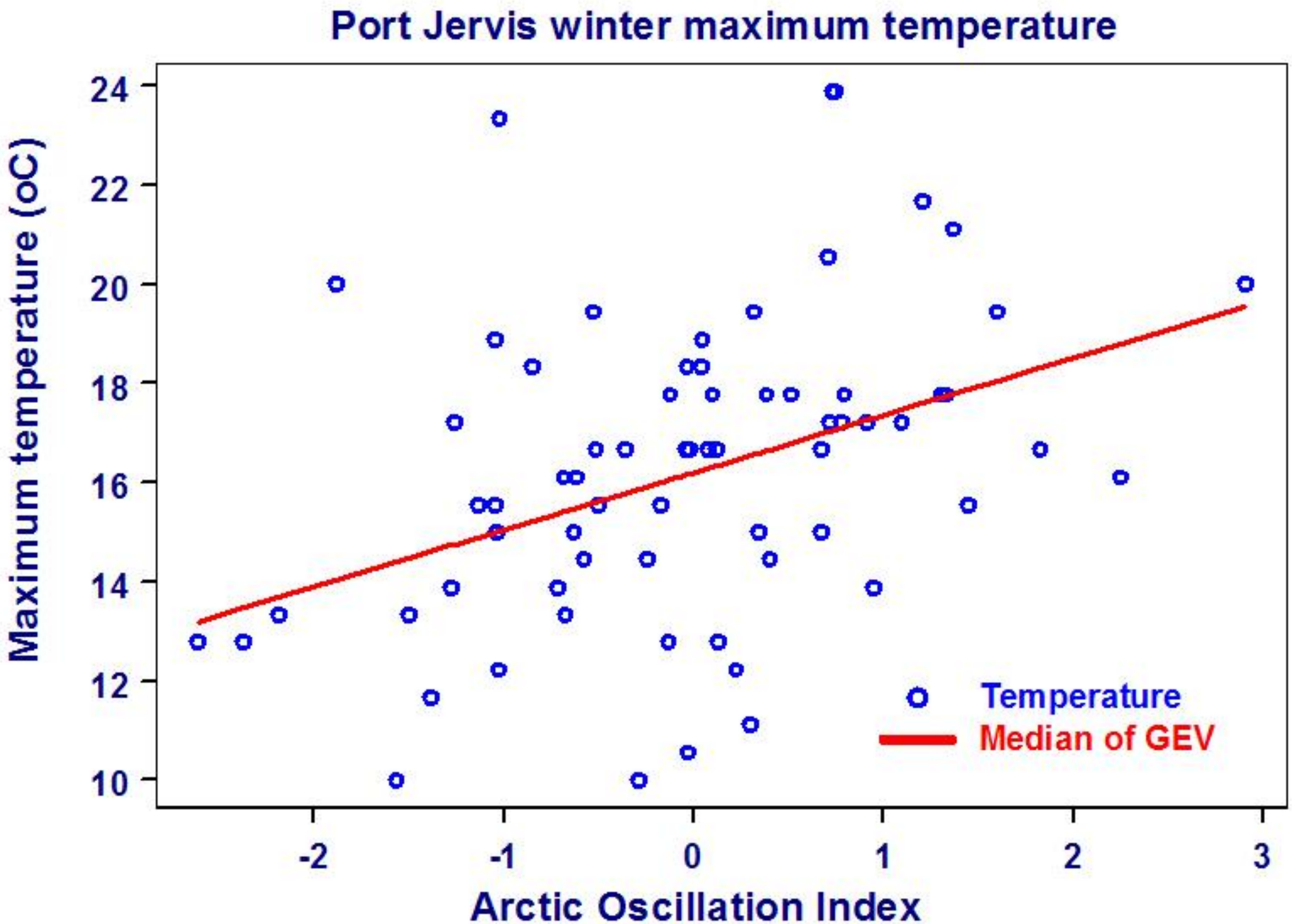
$$\xi(z) = \xi$$

- **Parameter estimates (maximum likelihood) and standard errors**

Parameter		**Estimate**	**(Std. Error)**
Location:	μ_0	**15.26**	
	μ_1	**1.175**	**(0.319)**
Scale:	σ_0	**0.984**	
	σ_1	**−0.044**	**(0.092)**
Shape:	ξ	**−0.186**	

-- Likelihood Ratio Test (LRT) for $\mu_1 = 0$ (*P*-value < 0.001)

-- LRT for $\sigma_1 = 0$ (*P*-value ≈ 0.635)


Port Jervis winter maximum temperature
Maximum temperature (oC)
24
22
20
18
16
14
12
10
-2
-1
0
1
2
3
Arctic Oscillation Index
Temperature
Median of GEV

(3) Conditional Extreme Value Analysis: Peaks over Threshold

- **Exceedance of high threshold *u***

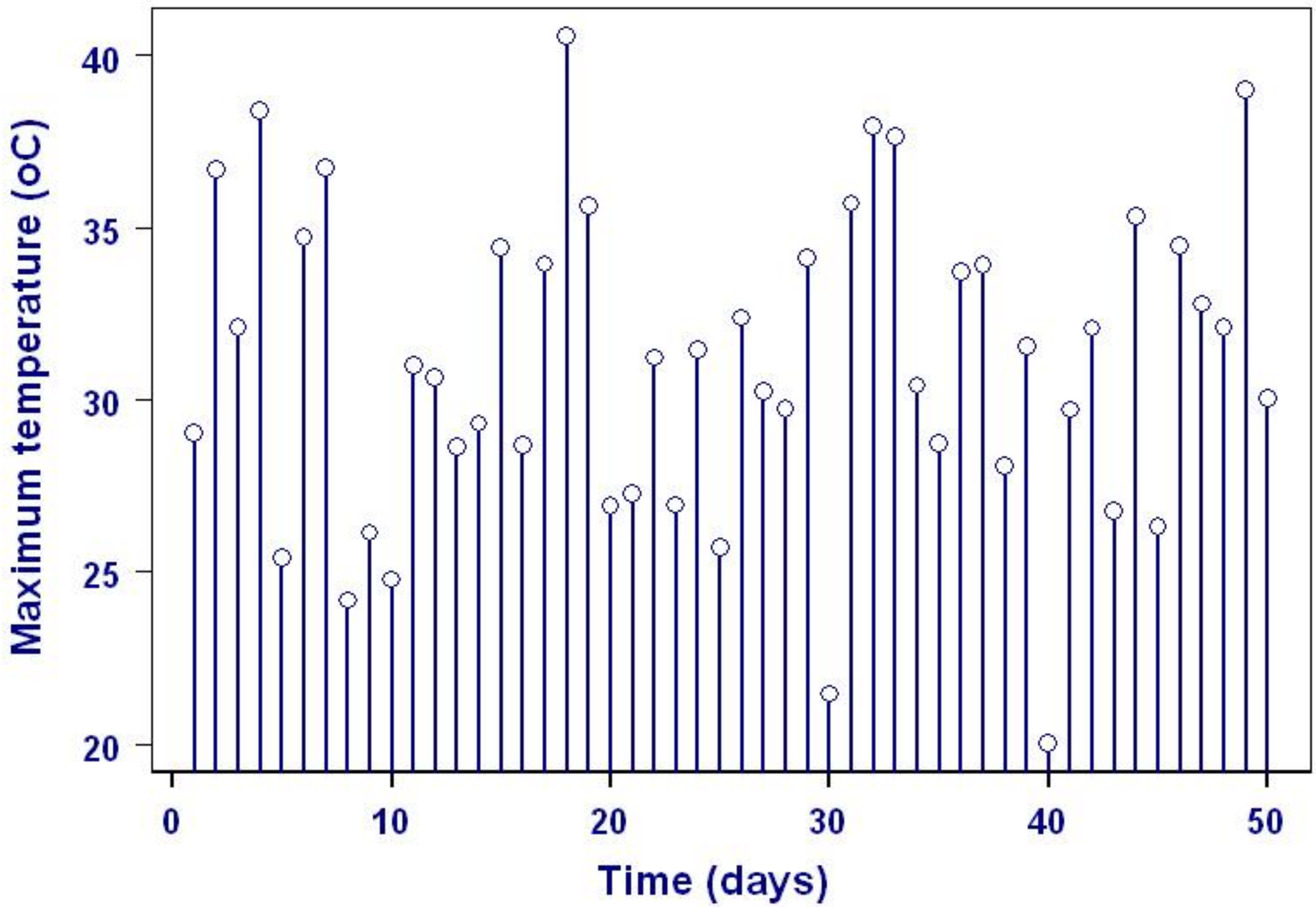
-- Law of Small Numbers (Poisson approximation to binomial)

Poisson distribution / process (Rate parameter $\lambda > 0$)

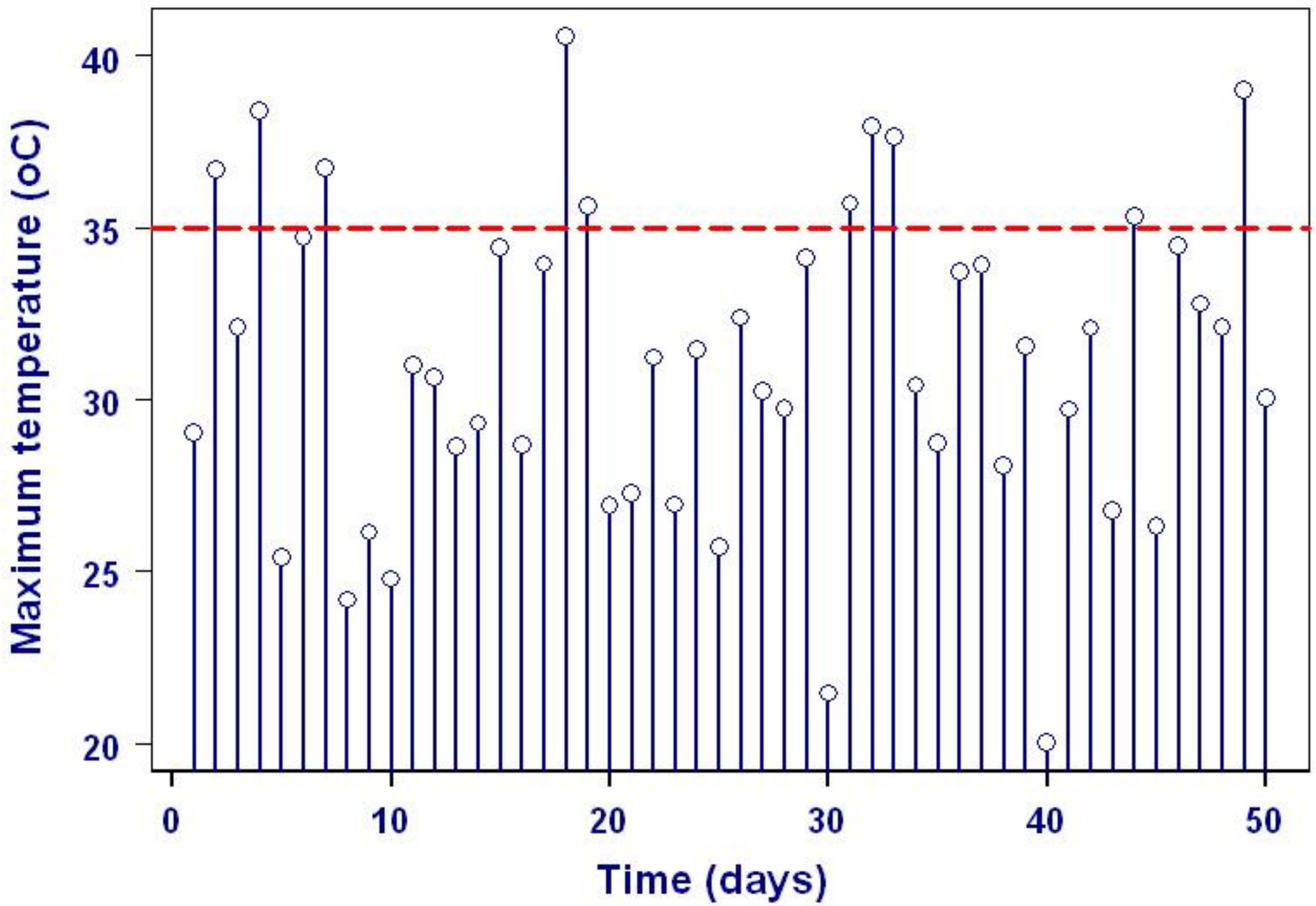
Given covariate $Z = z$ (i. e., circulation index on daily time scale), assume Poisson process with rate parameter depending on z:

$$\ln \lambda(z) = \lambda_0 + \lambda_1 z$$

(Called “Poisson regression”)



Maximum temperature (oC)
40
35
30
25
20
0
10
20
30
40
50
Time (days)


Maximum temperature (oC)
40
35
30
25
20
0
10
20
30
40
50
Time (days)

- **Excesses over high threshold**

-- "POT stability" (analogous to max stability)

Recall "memoryless" property of exponential distribution

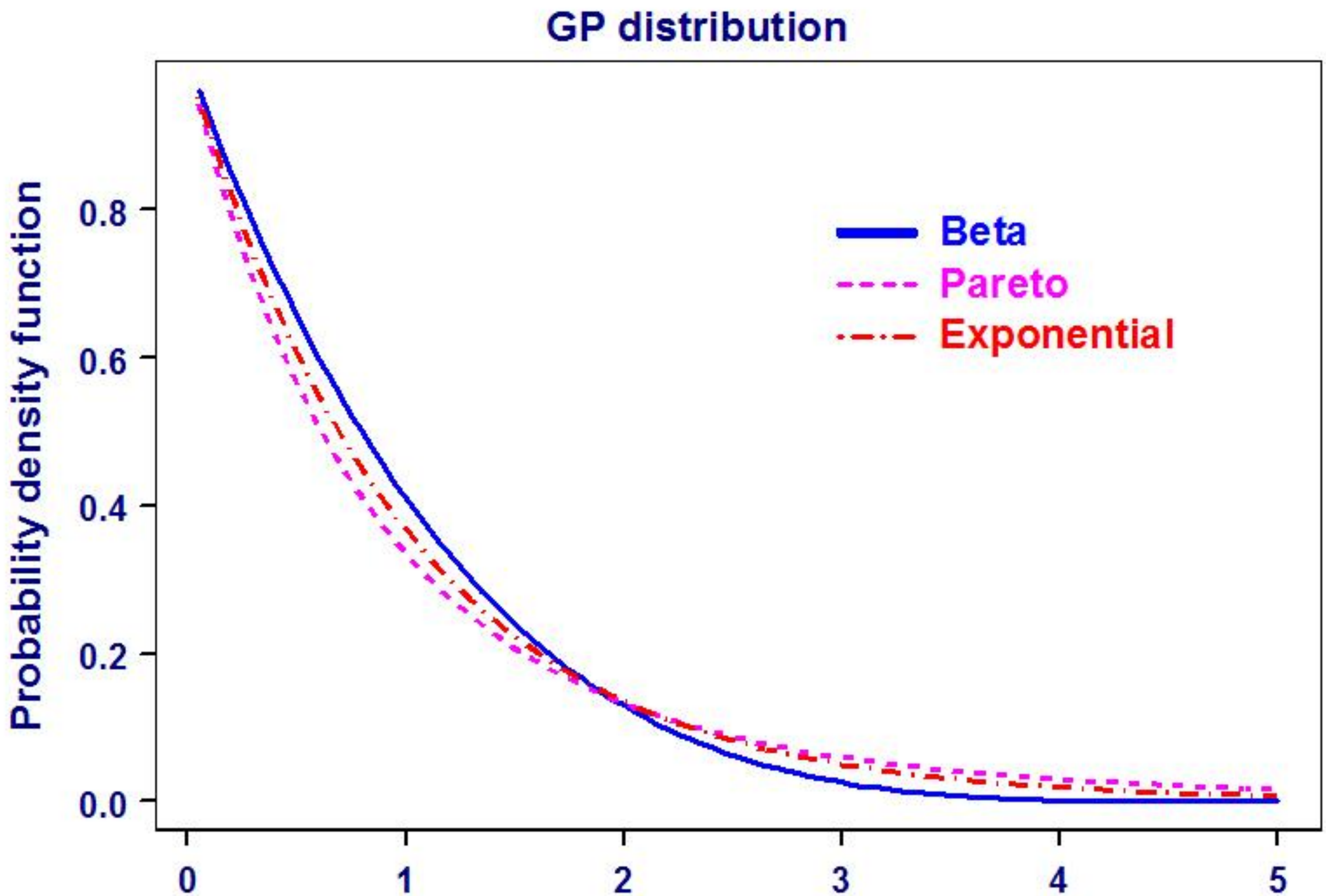
(In general, need to rescale)

-- Solution is Generalized Pareto (GP) distribution 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.)


GP distribution
Probability density function
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
Beta
Pareto
Exponential

(i) Shape parameter $\xi < 0$ (*Beta type*)

Bounded upper tail

(Analogous to Weibull type of GEV)

Fits excess temperature over high threshold

(ii) Shape parameter $\xi > 0$ (*Pareto type*)

Heavy upper tail

(Analogous to Fréchet type of GEV)

Fits excess precipitation over high threshold

- **Conditional approach**

-- Given covariate $Z = z$ (i. e., circulation index on daily time scale), assume conditional distribution is GP with parameters:

$$\ln \sigma^*(z) = \sigma_0 + \sigma_1 z$$

$$\xi(z) = \xi$$

That is, use circulation index on same day as excess

- **Challenges**

-- Choice of threshold

High (to satisfy theory), but *not* too high (need sufficient data)

-- Temporal dependence (Declustering)

"Runs declustering" with runs parameter

(For *r* = 1, cluster ends when one observation falls below *u*)

Only retain cluster maxima

Cluster ≈ hot spell

Other cluster / hot spell statistics

(e. g., duration)

(4) Application to California Temperature Extremes

- **CA Central Valley**

-- Daily maximum temperature at 3 sites

Bakersfield, Fresno, Red Bluff

Extract single highest temperature for each day

-- Period of record: 1951 – 2005

-- Summer season: 16 June – 15 September

-- Set threshold u = 110.5 ^{o}F

-- Runs declustering with r = 1

- **Declustering (*r* = 1)**

  101 exceedances of *u* = 110.5 °F, 61 clusters

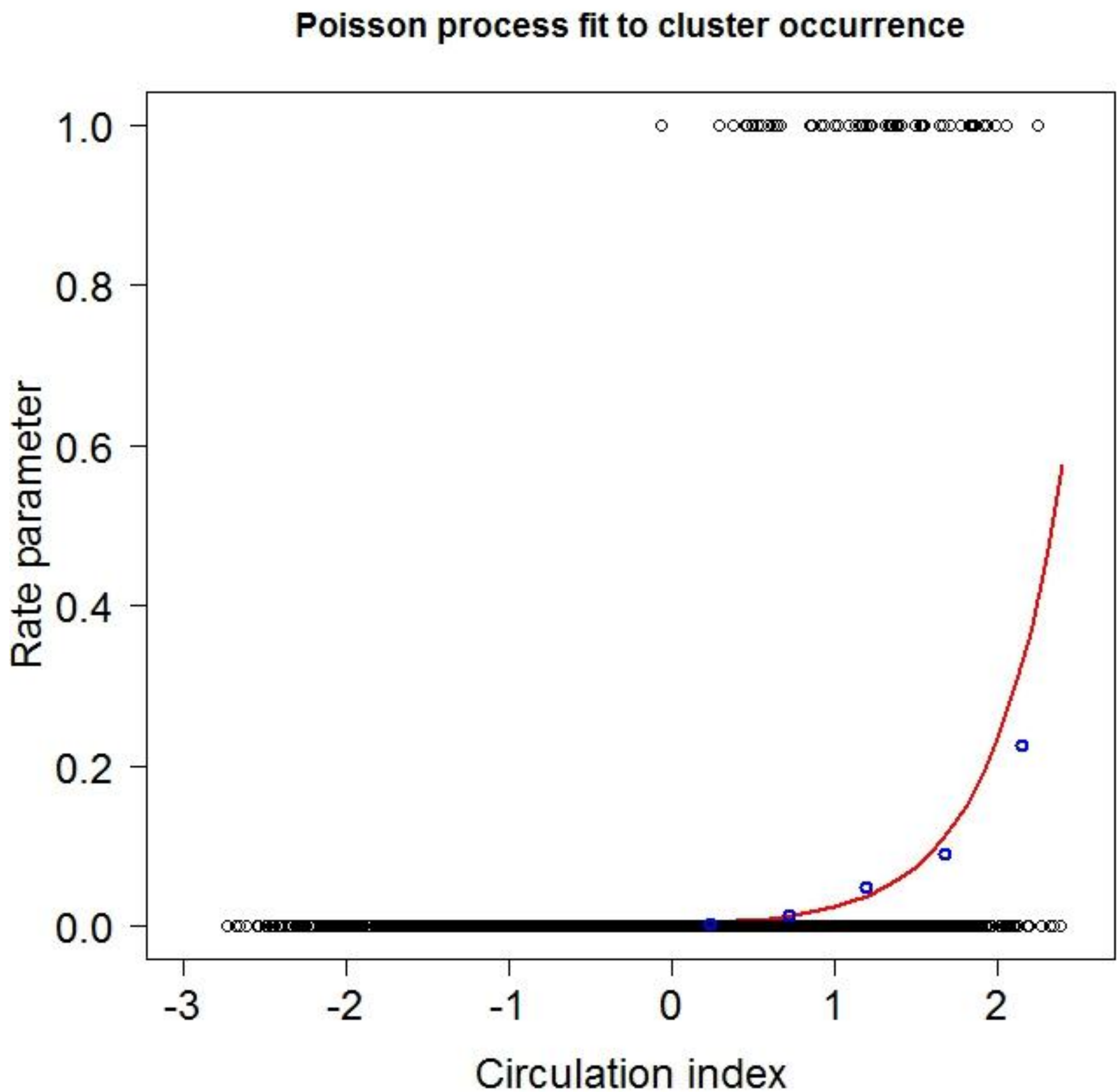
  Mean cluster length = 101 / 61 ≈ 1.656

  (Estimated extremal index ≈ 0.604)

- **Poisson process for cluster occurrence (*u* = 110.5 °F, *r* = 1)**

	Parameter	**Estimate**	**(Std. Error)**
Rate:	λ_0	**−5.950**	
	λ_1	**2.248**	**(0.203)**

-- LRT for $\lambda_1 = 0$ (*P*-value virtually zero)

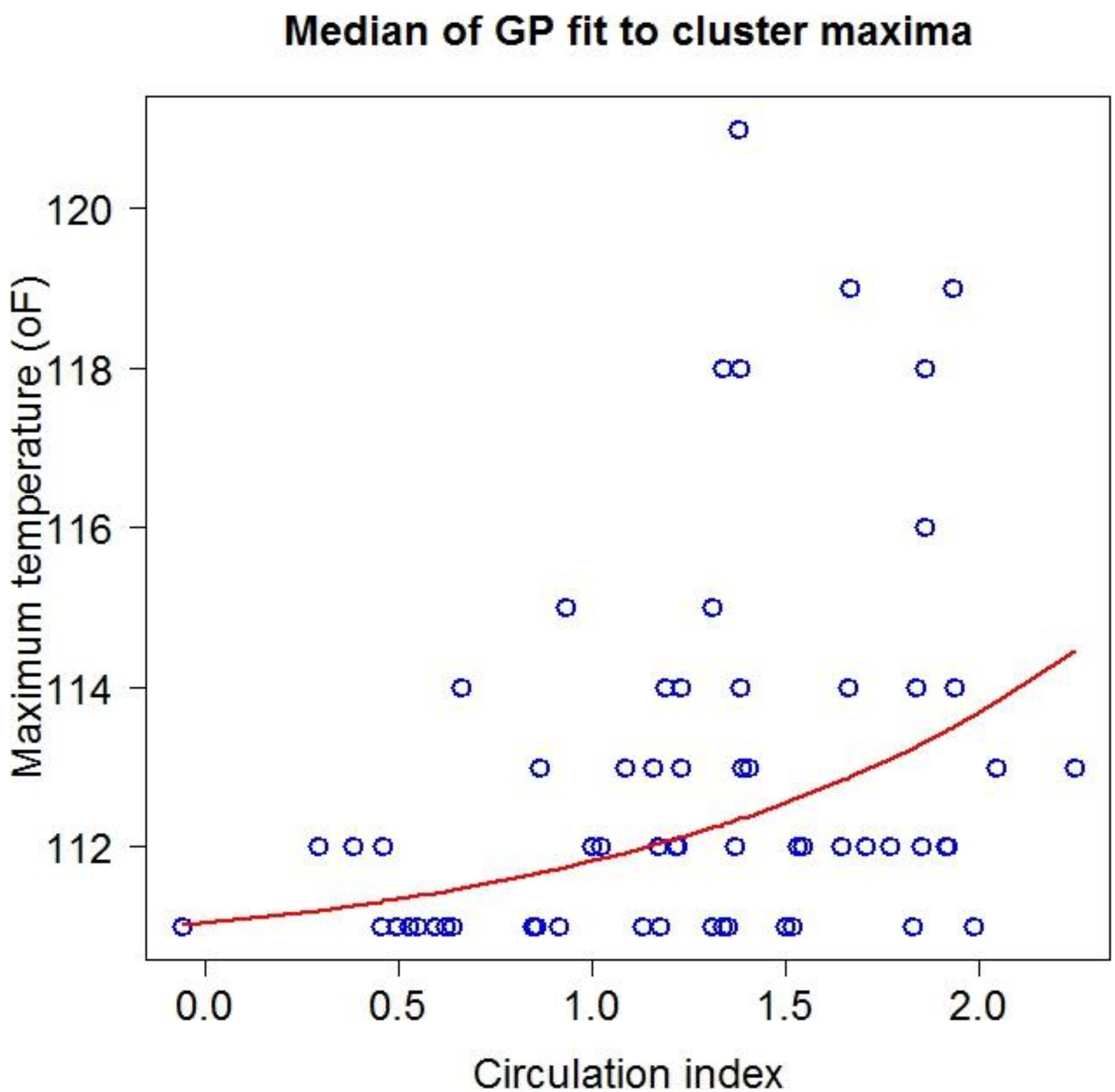


Poisson process fit to cluster occurrence
1.0
0.8
0.6
0.4
0.2
0.0
Rate parameter
-3
-2
-1
0
1
2
Circulation index

- **GP distribution for cluster maxima (u = 110.5 ^{o}F, r = 1)**

Parameter		Estimate	(Std. Error)
Scale:	σ_0	−0.180	
	σ_1	0.877	(0.258)
Shape:	ξ	−0.141	

-- LRT for $\sigma_1 = 0$ (*P*-value ≈ 0.002)


Median of GP fit to cluster maxima
Maximum temperature (oF)
120
118
116
114
112
0.0
0.5
1.0
1.5
2.0
Circulation index

- **Trends**

-- Could use trend (e. g., year) as covariate

Instead of (or in addition to) circulation index

-- Poisson process for cluster occurrence

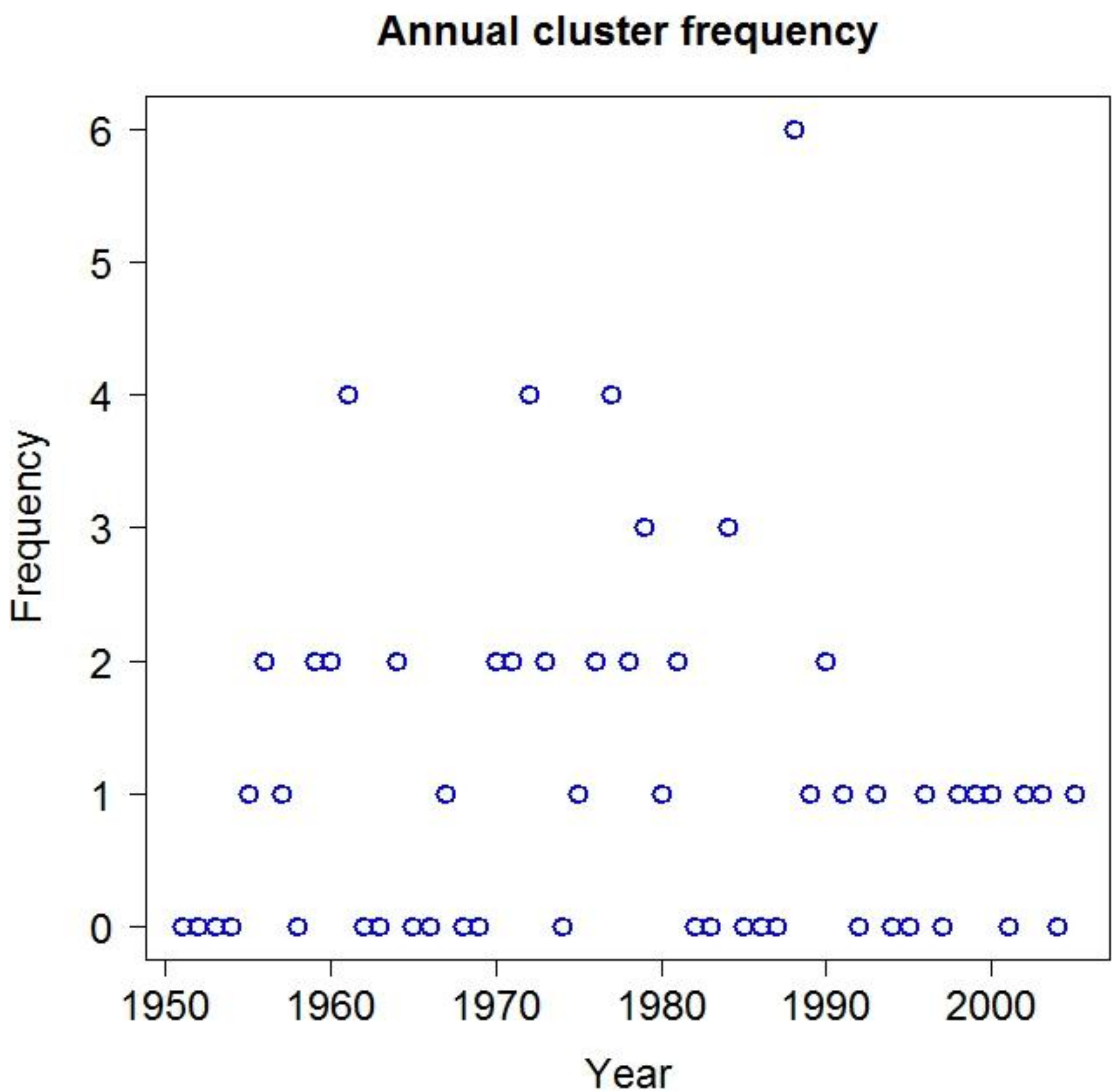
No evidence of trend (with/without circulation index)

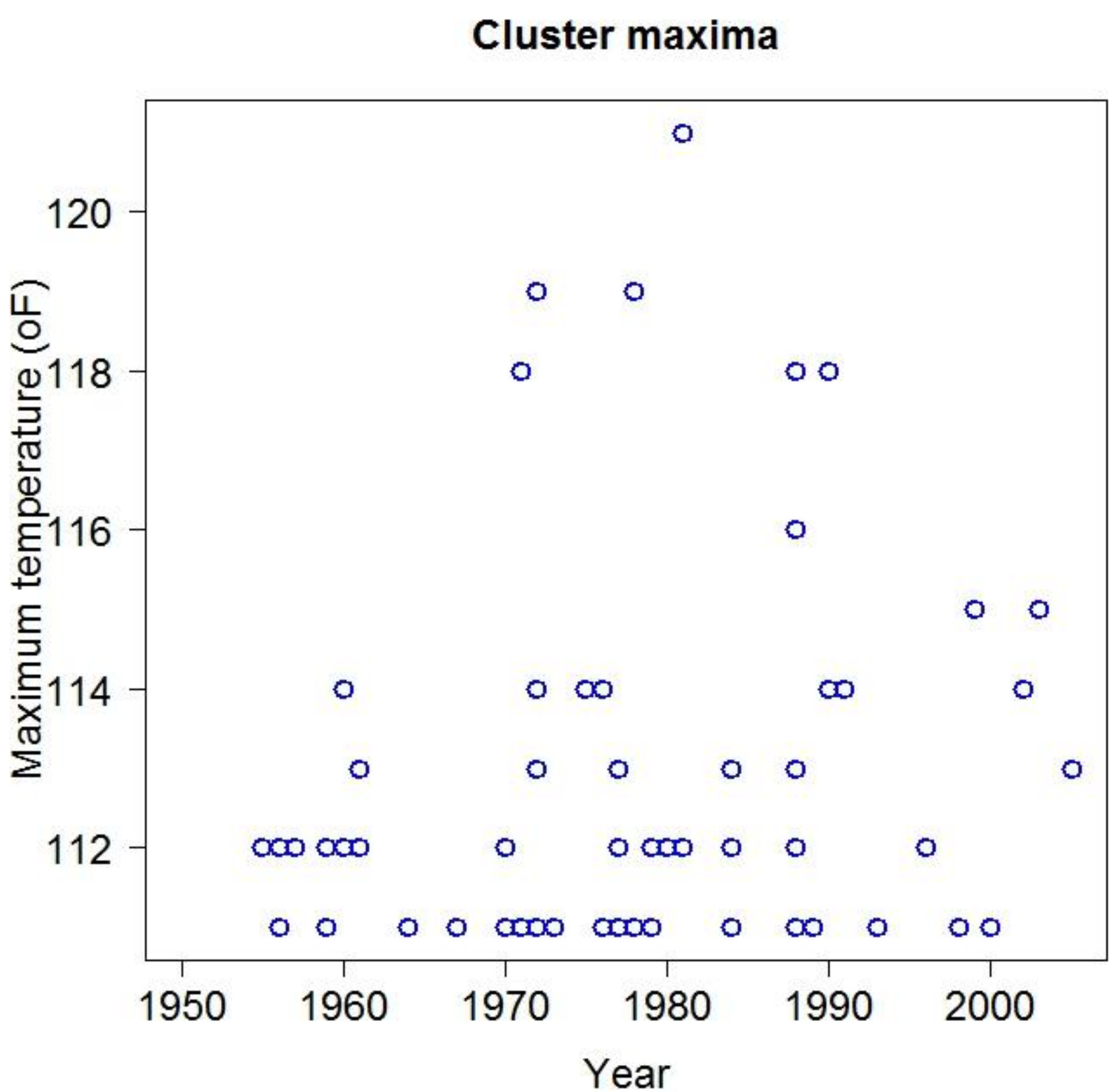
-- GP distribution for cluster maxima

Some evidence in support of at least weak increasing trend:

Trend vs. no trend: *P*-value ≈ 0.116

Circulation index & trend vs. circulation index: *P*-value ≈ 0.074


Annual cluster frequency
Frequency
0
1
2
3
4
5
6
1950
1960
1970
1980
1990
2000
Year


Cluster maxima
Maximum temperature (oF)
120
118
116
114
112
1950
1960
1970
1980
1990
2000
Year

(5) Remaining Work

- **CA temperature example**

-- Point process approach

Combine Poisson & GP components into single model

-- Cluster / hot spell statistics

Cluster duration vs. circulation index

Temporal dependence of excesses within a cluster

-- Temperature extremes at individual sites

Model extremes for all sites simultaneously

("Borrow strength" to reduce uncertainty in estimated effect of circulation index on extremes)

Resources

- **Statistics of Weather and Climate Extremes**

-- www.isse.ucar.edu/extremevalues/extreme.html

- **Extremes Toolkit**

-- www.isse.ucar.edu/extremevalues/evtk.html

-- cran.r-project.org/web/packages/extRemes/

- **Furrer et al. (*Climate Research*, 2010)**

-- Statistical model for hot spells

Trends in frequency, duration & intensity

www.isse.ucar.edu/staff/katz/docs/pdf/crheat10.pdf

- **Yun Li (CSIRO, Perth; unpublished)**

-- Daily index of large-scale circulation ("blocking")

Incorporation into statistical model for hot spells

(Frequency, duration, and intensity components)

-- Australian data

Both observations and model output