EVIDENCE FOR CLUSTERING OF TEMPERATURES AT HIGH LEVELS BASED ON EXTREME VALUE THEORY

Rick Katz

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

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

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

“Recipe for Disaster: The Formula That Killed Wall Street”

Felix Salmon (*Wired Magazine*: 23 Feb. 2009):

$$\Pr[T_A < 1, T_B < 1] = \phi_2(\phi^{-1}(F_A(1)), \phi^{-1}(F_B(1)), \gamma)$$

- **Gaussian copula**

-- Statistical model for joint dependence

Marginal distributions are transformed to Gaussian (or uniform)

-- Proved inadequate for financial applications

What about for climate extremes?

Outline

(1) Background / Motivation

(2) Extremal Index for Clustering at High Levels

(3) First-order Gaussian Process

(4) Application to Daily Maximum Temperature

(5) Discussion

(1) Background / Motivation

- **Heat waves**

-- Meteorological phenomenon

Extreme event

Persistent (e. g., associated with blocking)

-- Statistical perspective

Extreme Value Theory

Concept of Extremal Index

Measure of clustering at high levels

(2) Extremal Index for Clustering at High Levels

- Extreme Value Theory (EVT)

-- Consider sequence of observations $\{X_1, X_2, \ldots\}$

Assumed independent (*for now*) and identically distributed (IID)

-- Extremal Types Theorem for $\max\{X_1, X_2, \ldots, X_n\}$

Approximate generalized extreme value (GEV) dist. for large n

-- Analogue for excess over high threshold $X_t - u$, given $X_t > u$

Approximate generalized Pareto (GP) dist. for high threshold u

- **EVT under temporal dependence**

-- Consider stationary process $\{X_1, X_2, . . . \}$

That is, possibly temporally dependent

-- Weak memory (Observations far apart in time)

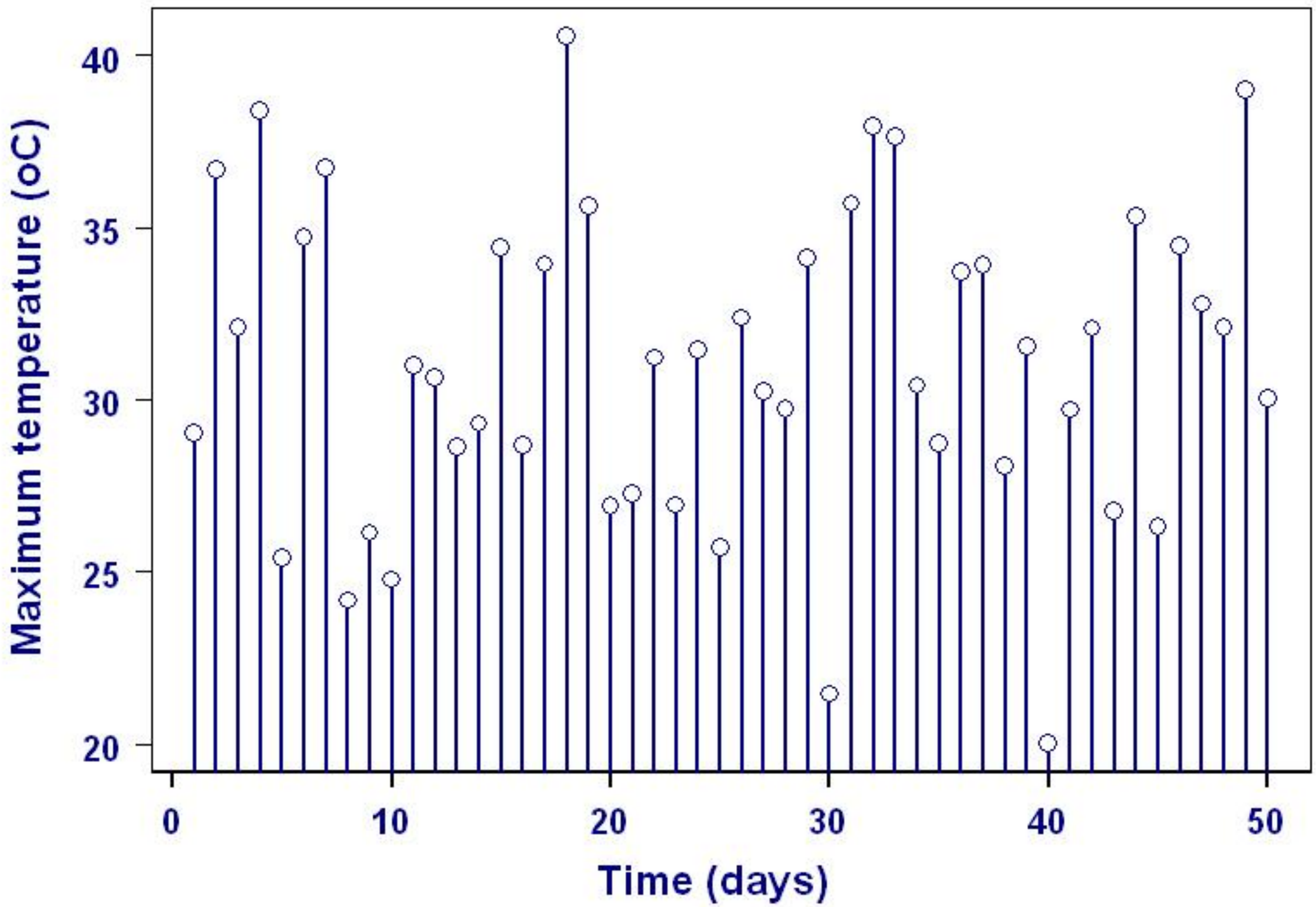
EVT still holds if dependence weakens at fast enough rate

-- Clustering at high levels (Observations close together in time)

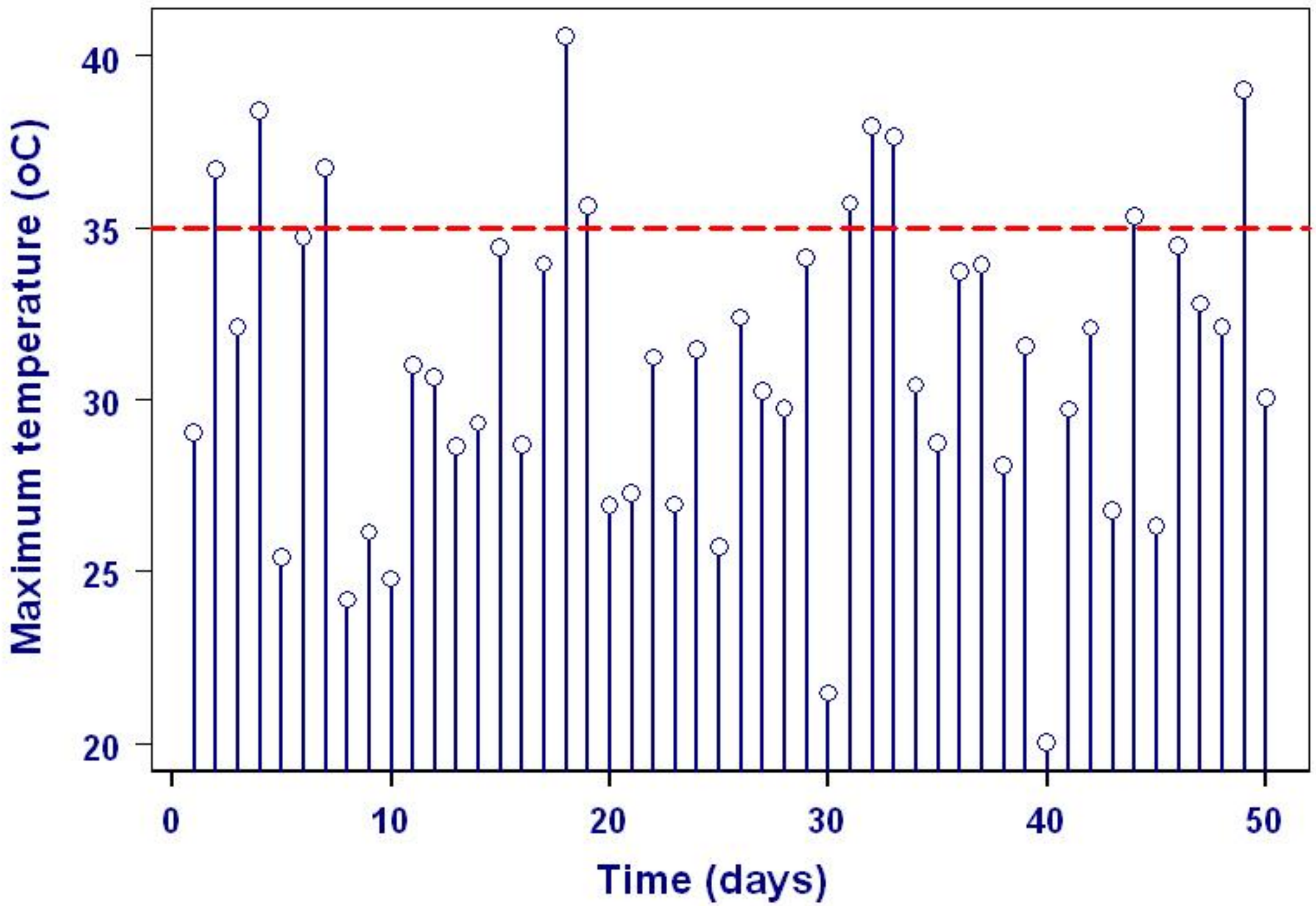
EVT still holds with adjustment to normalizing constants

Consider *cluster* (or *hot spell*): $X_{t+1} > u, . . ., X_{t+k} > u$

Starts at time $t + 1$, length k days ($k \geq 1$), where u is high threshold



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



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

- **Condition for *no* clustering at high levels**

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

- **Clustering at high levels (*Extremal Index* θ, 0 ≤ θ ≤ 1)**

(i) θ = 1

No clustering at high levels (Independent or Gaussian process)

(ii) 0 ≤ θ < 1

Clustering at high levels

***Interpretations*: (i) Mean Cluster Length ≈ 1 / θ**

(ii) Effective sample size (proportion)

- **Estimating Extremal Index**

(i) ***Intervals Estimator*** **(Ferro-Segers 2003)**

Does not require identification of clusters

Consider “interexceedance times” between exceedances of threshold

Coefficient of variation of these interexceeance times converges to simple function of Extremal Index θ

Resample (Bootstrap) interexceedance times to obtain confidence interval for Extremal Index θ

(ii) Runs

Approach typically used in "declustering"

Runs parameter $r = 1, 2, \ldots$

$r = 1$ corresponds to defining clusters as run of consecutive exceedances of threshold

$r = 2$ corresponds to terminating a cluster if two or more consecutive days fall below threshold

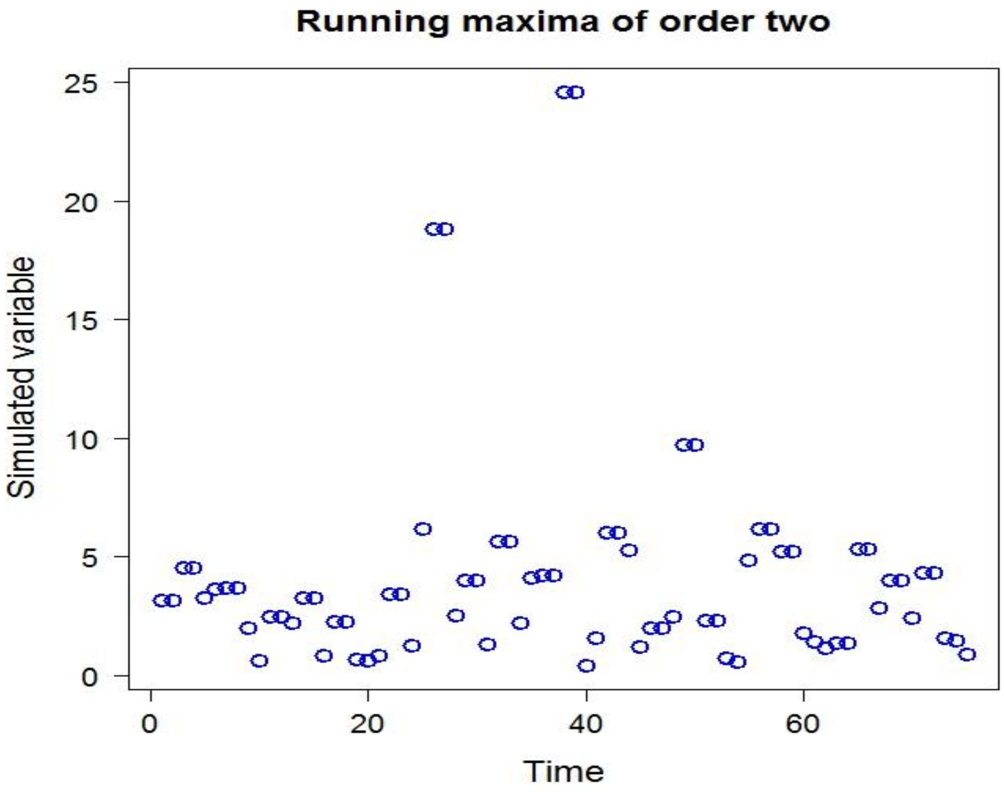
Identify clusters based on some value of r

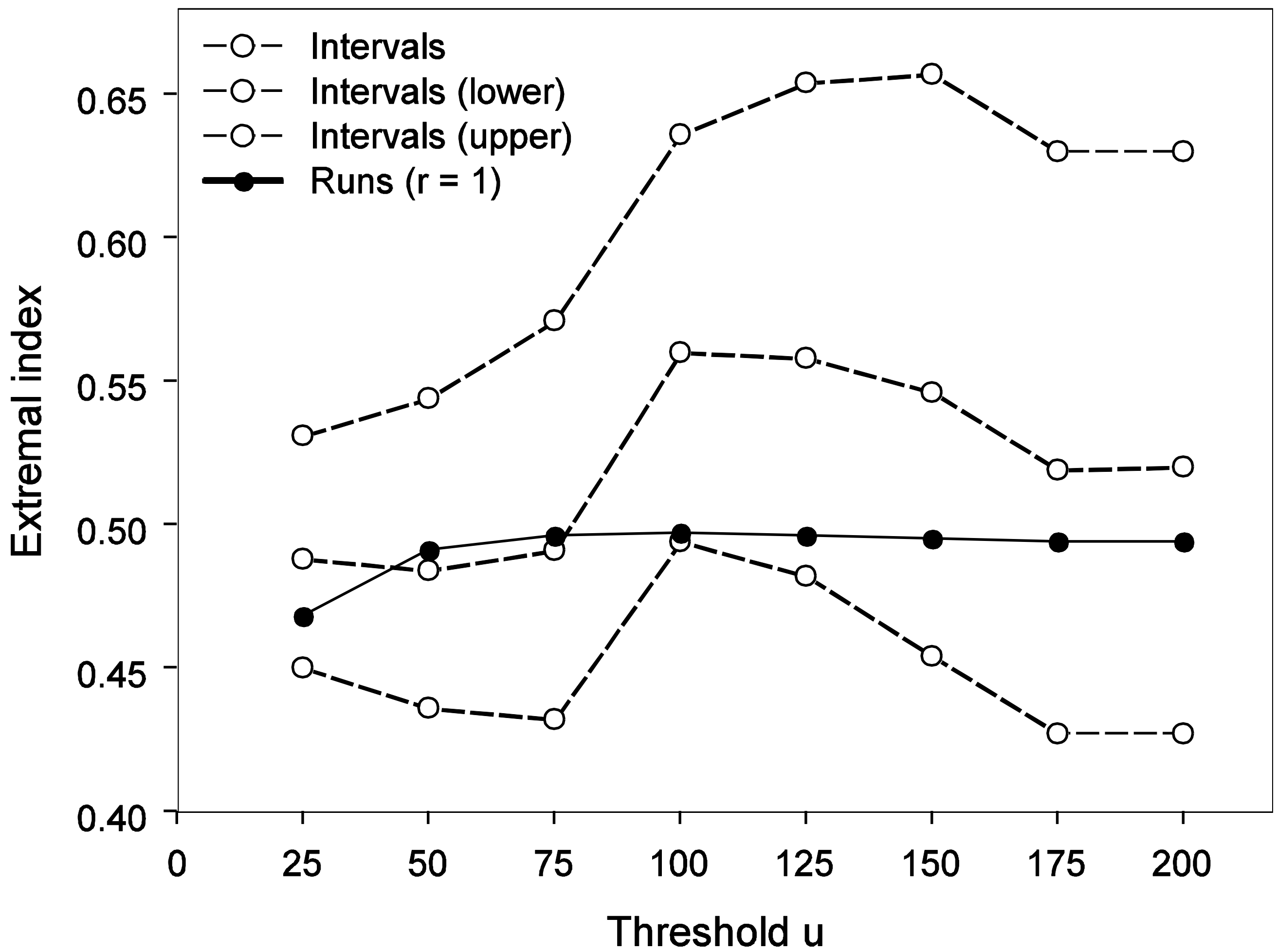
Then estimate θ by: 1 / (Mean Cluster Length)

- **Artificial example (Running maxima of order two)**

$Y_k = \max\{X_k, X_{k+1}\}, k = 1, 2, \ldots, \quad \{X_1, X_2, \ldots\}$ **IID**

Then extremal index $\theta = 0.5$ for $\{Y_k\}$ time series





Running maxima of order two

(3) First-Order Gaussian Process

- **AR(1) Process (zero mean, unit variance)**

$$X_t = \phi X_{t-1} + \varepsilon_t , \qquad \varepsilon_t \sim N(0, 1 - \phi^2)$$

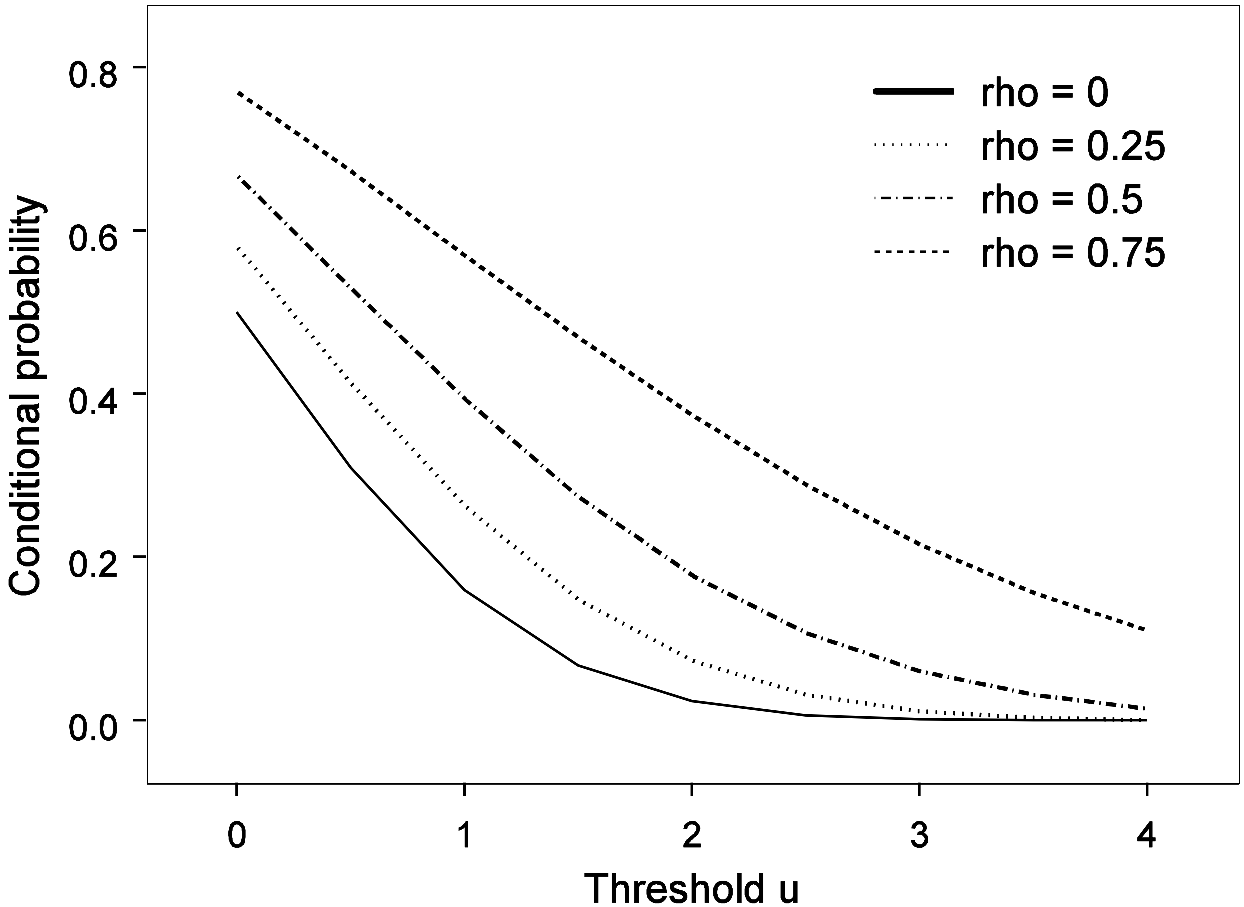
-- Autocorrelation function

$$\mathrm{Corr}(X_t , X_{t+k}) = \phi^k, \; k = 1, 2, \ldots$$

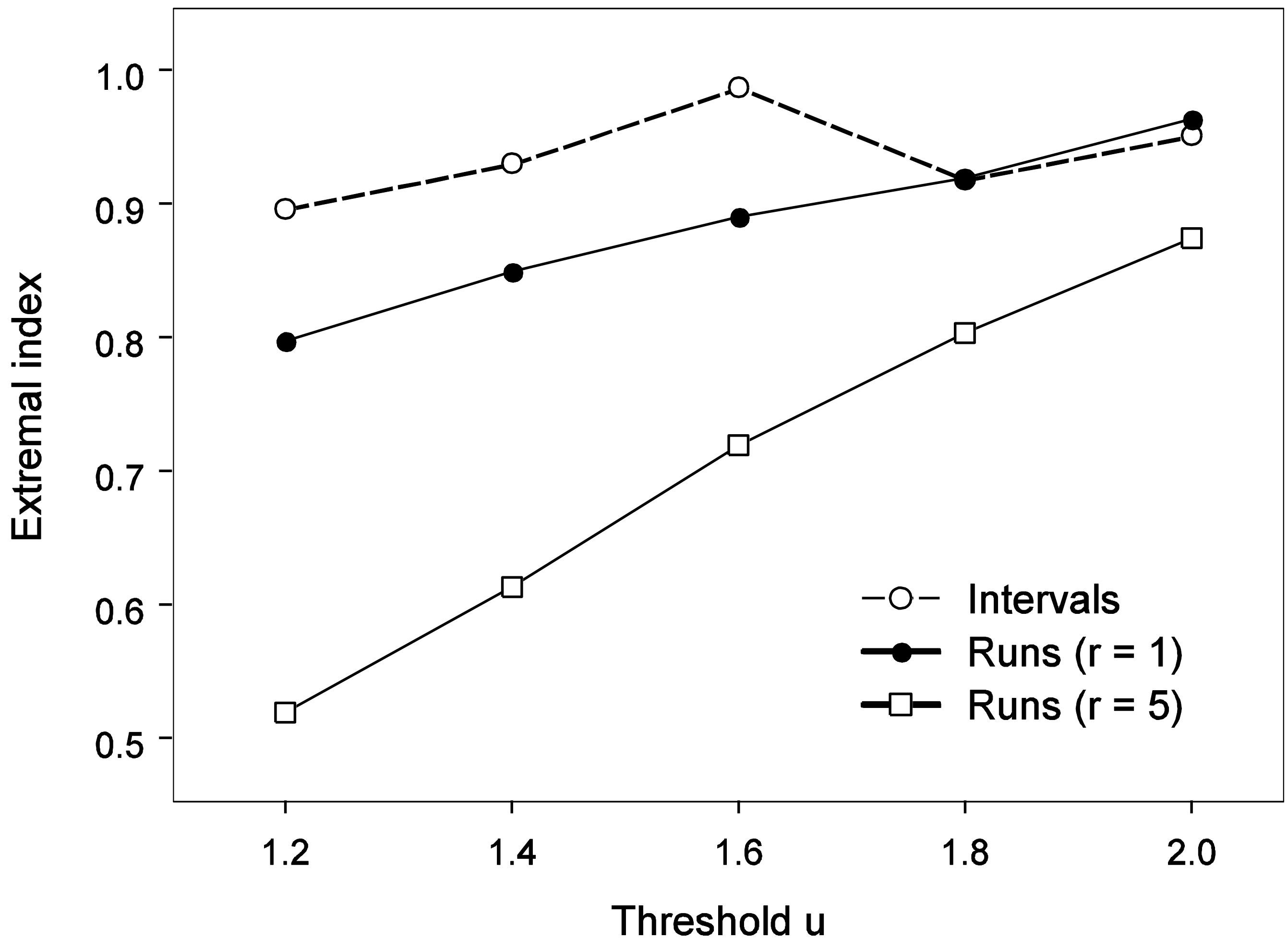
-- Joint distribution

$\{X_t , X_{t+k}\}$ ~ **Bivariate normal with correlation coefficient** $\rho = \phi^k$

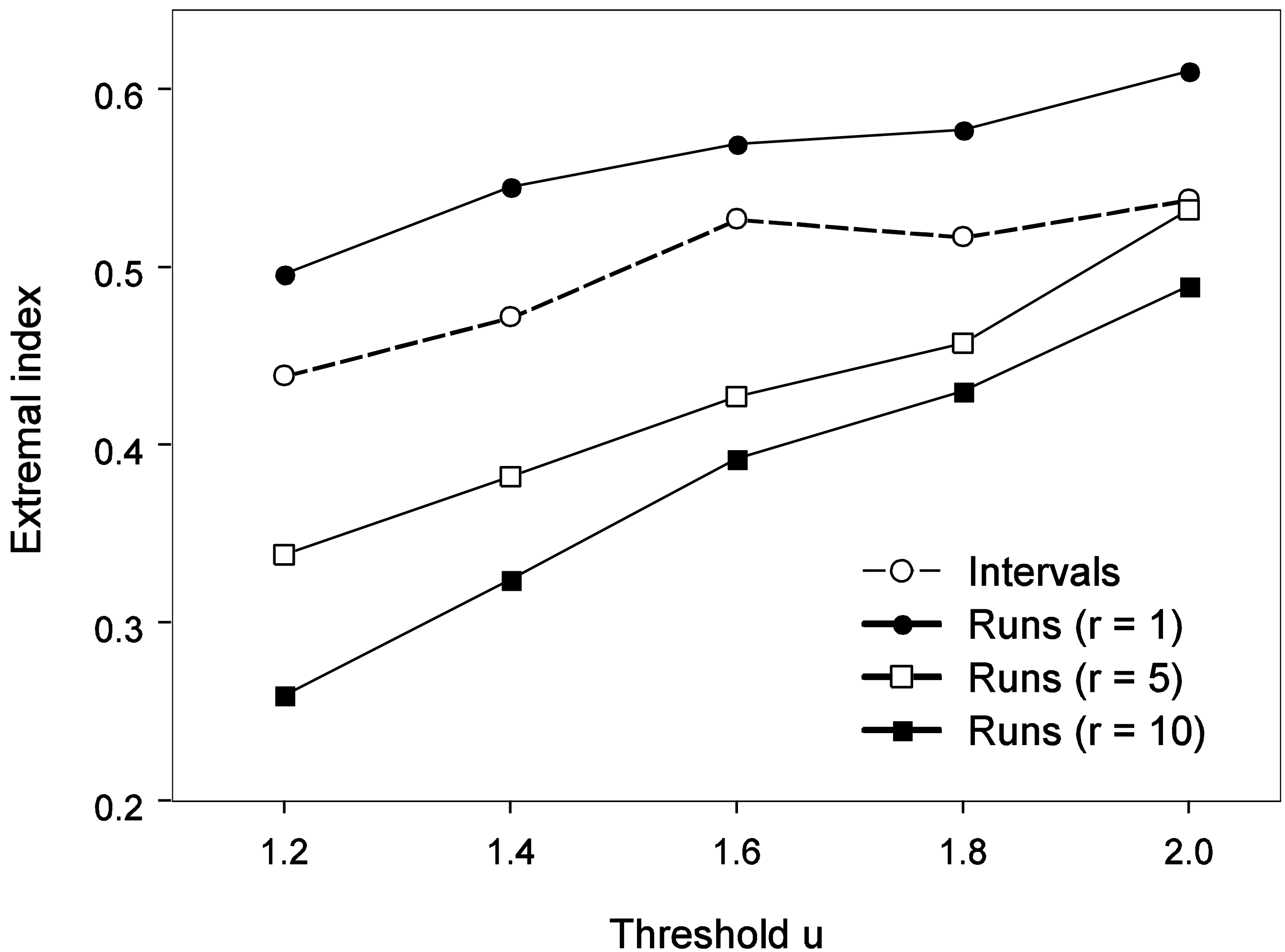
-- Straightforward to show that Extremal Index $\theta = 1$



Bivariate Normal Distribution: $\Pr\{X_{t+k} > u \mid X_t > u\}$



AR(1) Process: $\phi = 0.25$



AR(1) Process: $\phi = 0.75$

(4) Application to Daily Maximum Temperature

- **Estimate Extremal Index θ by two different methods**

(i) Intervals Estimator of θ

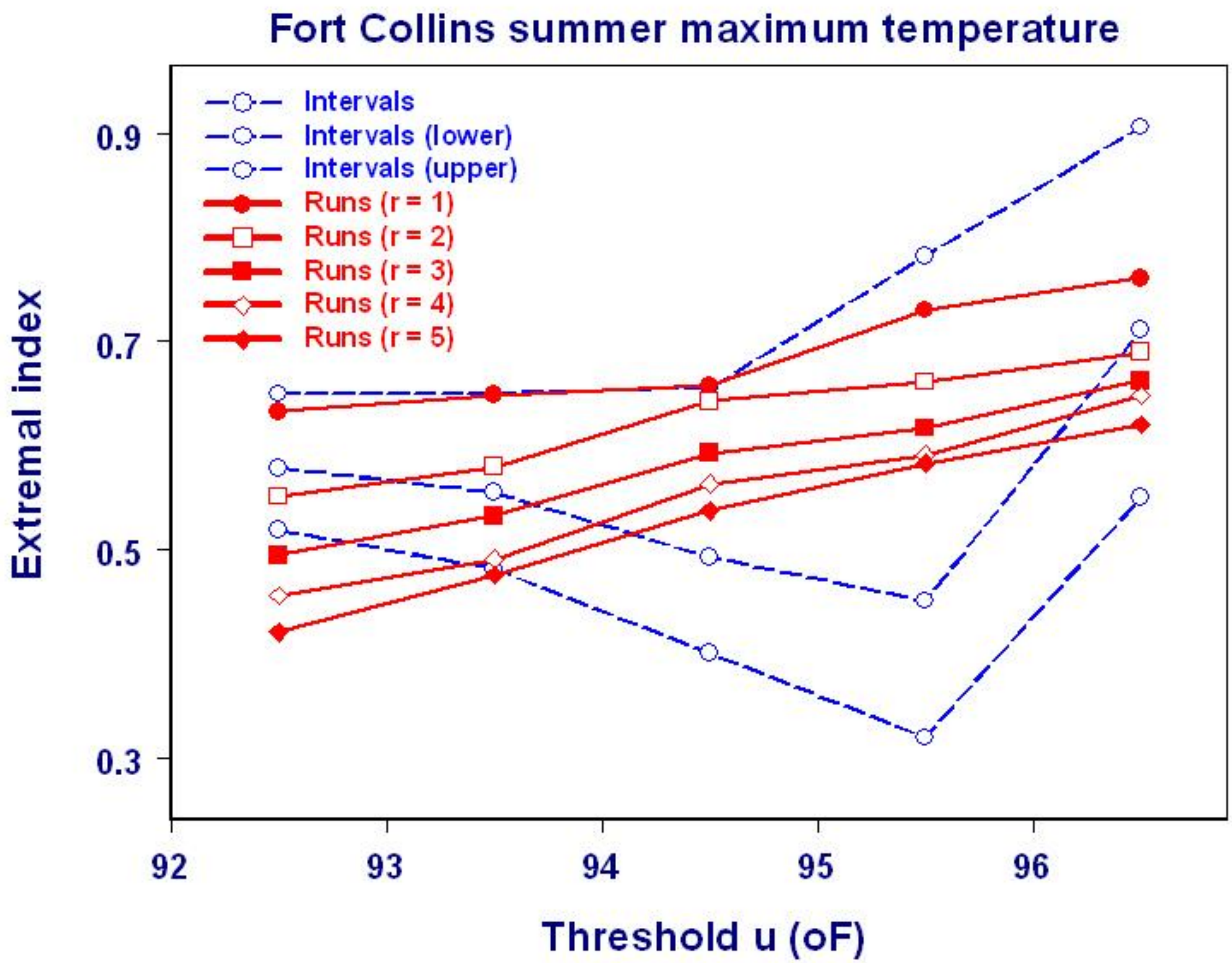
-- Vary threshold *u*

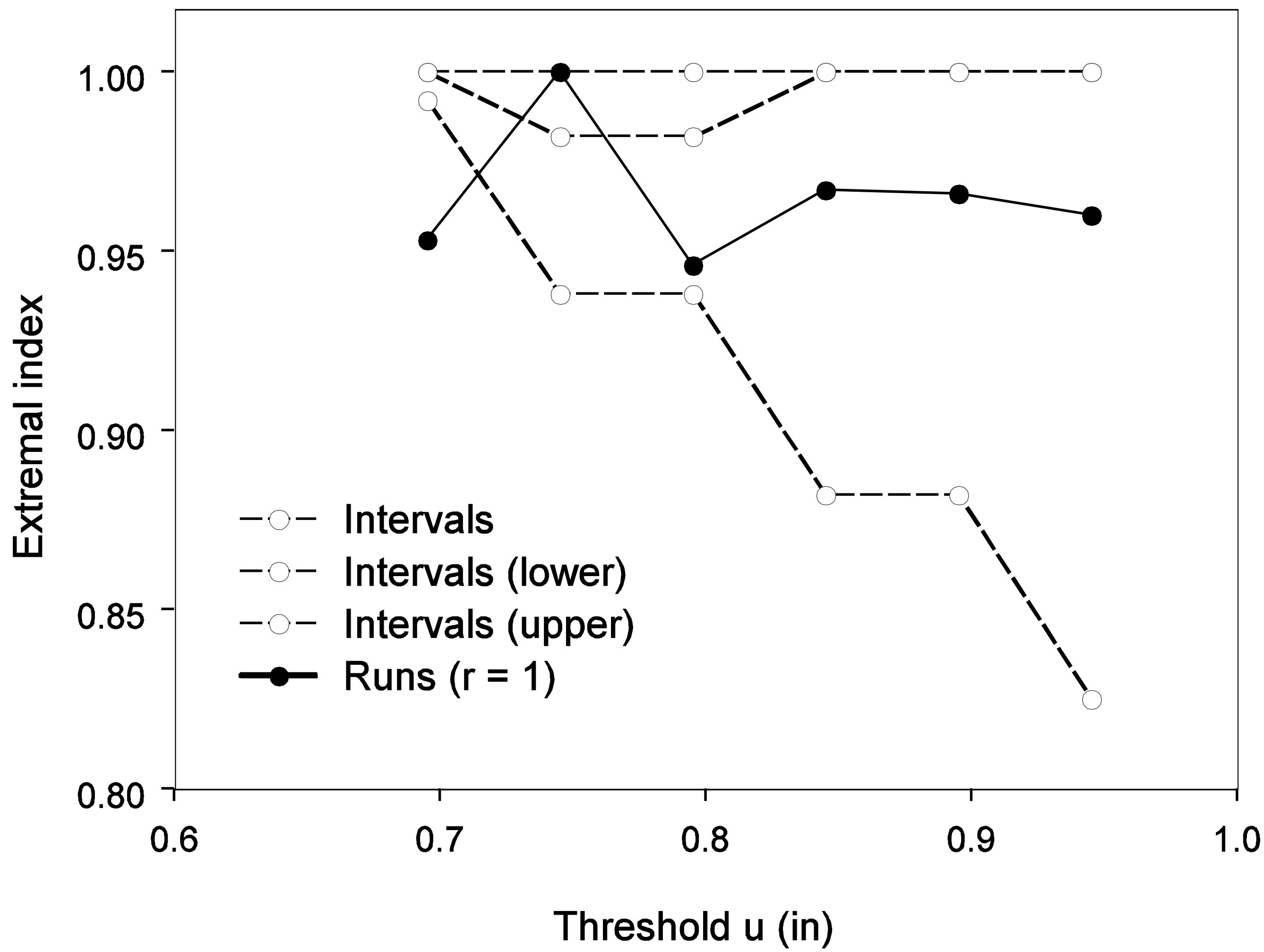
-- Attach confidence interval based on bootstrap

(i) Runs Estimator of θ

-- Runs parameter *r* = 1, 2, 3, 4, 5

-- Vary threshold *u*


Fort Collins summer maximum temperature
Intervals
Intervals (lower)
Intervals (upper)
Runs (r = 1)
Runs (r = 2)
Runs (r = 3)
Runs (r = 4)
Runs (r = 5)
0.9
0.7
0.5
0.3
Extremal index
92
93
94
95
96
Threshold u (oF)



Fort Collins July precipitation

(5) Discussion

- **Dilemma**

(i) *Financial statistics*

Strong evidence that Gaussian copula is inappropriate model for simultaneous occurrence of extremes

(ii) *Climate statistics*

Gaussian models are useful -- Should *not* be viewed as if produce no dependence of extremes

But do Gaussian models treat joint dependence of extremes realistically enough?