STATISTICS OF EXTREMES IN CLIMATE:

RECONCILING THEORY WITH OBSERVATIONS

Rick Katz

Institute for Study of Society and Environment
National Center for Atmospheric Research
Boulder, CO USA

email: rwk@ucar.edu

Home page: www.isse.ucar.edu/HP_rick/

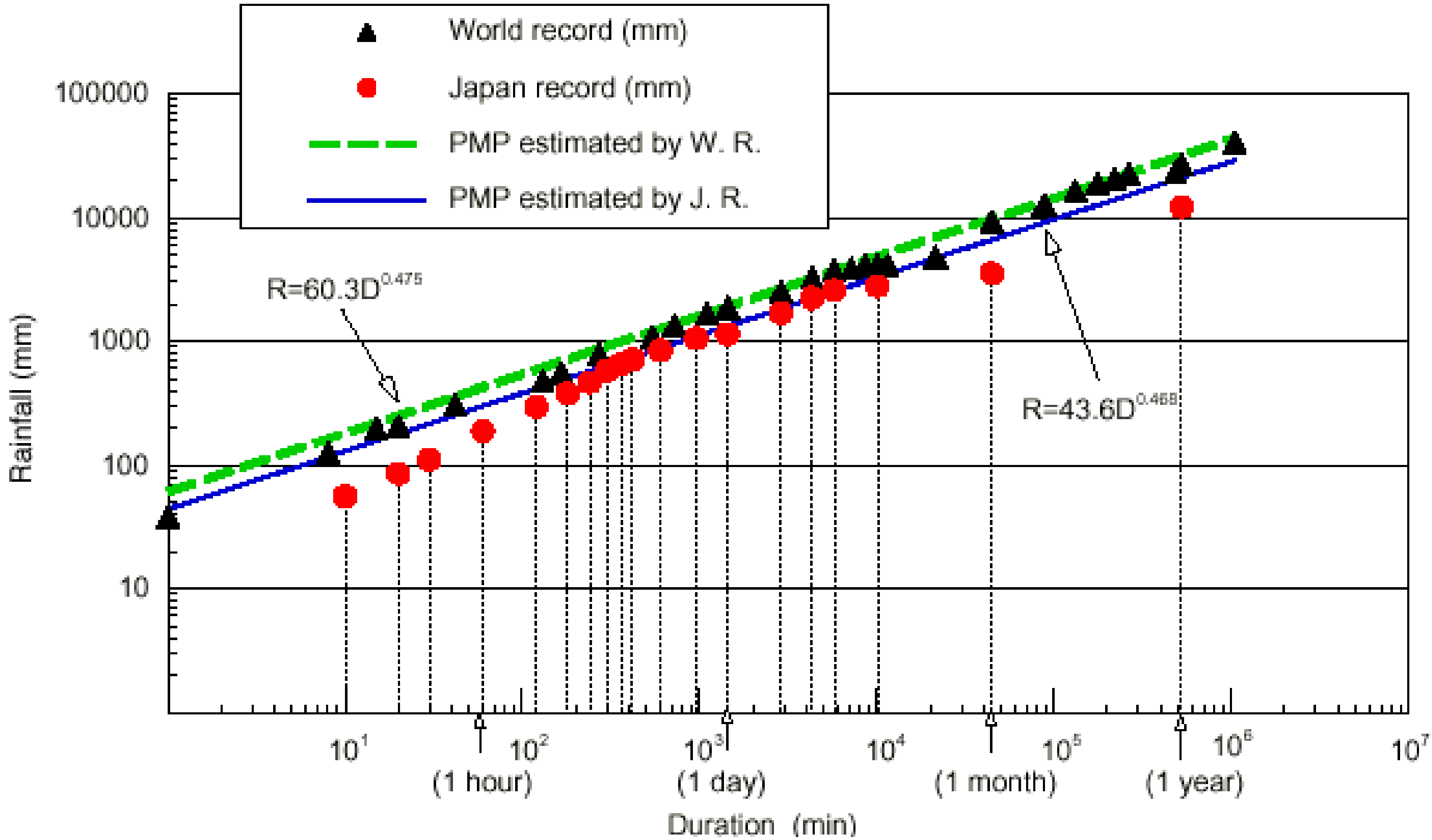
Lecture: www.isse.ucar.edu/HP_rick/pdf/katzcb08.pdf

QUOTE

- **Heraclitus (Greek philosopher)**

"*It is never possible to step twice into the same river.*"

Outline



World record (mm)
Japan record (mm)
PMP estimated by W. R.
PMP estimated by J. R.
100000
10000
1000
100
10
Rainfall (mm)
R=60.3D^0.475
R=43.6D^0.468
10^1
10^2
10^3
10^4
10^5
10^6
10^7
(1 hour)
(1 day)
(1 month)
(1 year)
Duration (min)

(1) Background

- **Use of Extremal Models in Climate Applications**

-- Non-Stationarity

Climate change?

-- Bounded Upper Tail?

-- Scaling / Aggregation

-- Stochastic weather generators

Improved treatment of extremes

(2) Observed Tail Behavior

- **Extreme Value Theory**

-- Generalized Extreme Value (GEV) distribution

-- Generalized Pareto (GP) distribution

-- Shape parameter ξ of GEV or GP distribution

$\xi > 0$ heavy-tailed distribution (Fréchet or Pareto)

$\xi < 0$ bounded distribution (Weibull or beta)

$\xi = 0$ light-tailed distribution (Gumbel or exponential)

- **Upper Bound**

-- Bounded upper tail (i. e., $\xi < 0$)

Daily minimum & maximum temperature

Hourly or daily wind speed

Pollutant concentration

Sea level

Wave height

-- Interpretation of upper bound

Physically meaningful?

Statistical artifact?

-- Heavy upper tail (i. e., $\xi > 0$)

Precipitation:

Typical estimates for daily totals $0.1 < \xi < 0.15$

Streamflow:

Estimates of ξ tend to be higher than for precipitation

(Effect of integrating precipitation across water basin?)

Economic damage from extreme events:

Estimates of ξ tend to be higher yet

Economic damage from hurricanes (estimate of $\xi \approx 0.5$)

- **Scaling / Aggregation**

-- **Apparent conflict with extreme value theory**

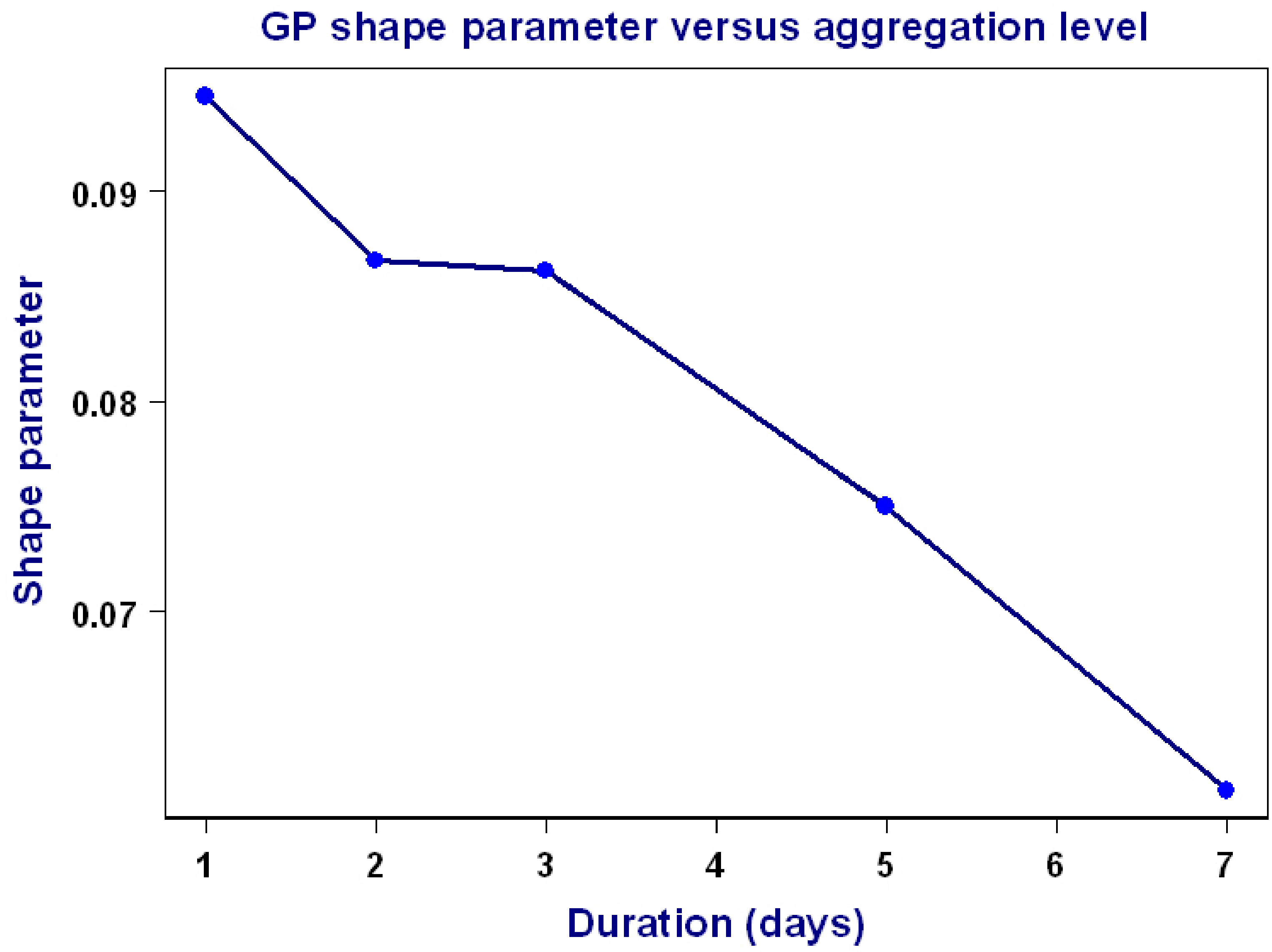
Observed precipitation extremes:

Shape parameter tends to decrease with aggregation over time (e. g., hourly vs. daily total amounts)

Example: Regional analysis of precipitation extremes in Texas

Extreme value theory:

Shape parameter should be invariant with respect to aggregation


GP shape parameter versus aggregation level
Shape parameter
0.09
0.08
0.07
1
2
3
4
5
6
7
Duration (days)

(3) Penultimate Approximations

- **"Ultimate" Extreme Value Theory**

-- GEV distribution as limiting distribution of maxima

$X_1, X_2, \ldots, X_T$ with common distribution function F

$M_T = \mathrm{Max}\{X_1, X_2, \ldots, X_T\}$

- **Penultimate Extreme Value Theory**

-- Suppose F in domain of attraction of Gumbel type (i. e., $\xi = 0$)

-- Still preferable in nearly all cases to use GEV as approximate distribution for maxima (i. e., act as if $\xi \neq 0$)

-- Expression for shape parameter ξ_T

"Hazard rate" (or "failure rate"):

$$H(x) = F'(x) \,/\, [1 - F(x)]$$

Then one choice of shape parameter is:

$$\xi_T = \; (1/H)'\,(x)\; |_{x=u(T)}$$

where the "characteristic largest value"

$$u(T) = F^{-1}(1 - 1/T)$$

Here $\xi_T \rightarrow 0$ as block size $T \rightarrow \infty$

- **Example:** ***Exponential Distribution***

-- **Exact exponential upper tail (unit scale parameter)**

$$1 - F(x) = \exp(-x), \quad x > 0$$

-- **Penultimate approximation**

Hazard rate: $H(x) = 1\,, \quad x > 0$

Shape parameter: $\xi_T = 0, \quad T = 1, 2, \ldots$

So no benefit to penultimate approximation

- Example: *Normal Distribution*

-- Fisher & Tippett (1928) proposed Weibull type of GEV as penultimate approximation

-- F Normal distribution (with zero mean & unit variance)

Hazard rate: $H(x) \approx x$ (for large x)

Characteristic largest value: $u(T) \approx (2 \log T)^{1/2}$ (for large T)

-- Penultimate approximation is Weibull type with

$$\xi_T \approx -1 / (2 \log T)$$

For instance: $\xi_{30} \approx -0.15$, $\xi_{365} \approx -0.085$

- **Example:** ***"Stretched Exponential" Distribution***

-- Traditional form of Weibull distribution (Bounded below)

***Note*: Weibull extremal type is reflected version**

$$1 - F(x) = \exp(-x^{c}), \quad x > 0, \quad c > 0$$

where c is shape parameter (unit scale parameter)

-- Shape parameter for penultimate approximation is:

$$\xi_T \approx (1 - c) / (c \log T)$$

(i) "*Superexponential*" ($c > 1$) $\xi_T \uparrow 0$ **as** $T \to \infty$

(ii) "*Subexponential*" ($c < 1$) $\xi_T \downarrow 0$ **as** $T \to \infty$

(4) Interpretation of Tail Behavior

- **Apparent Upper Bound**

-- **Maximum possible hurricane intensity**

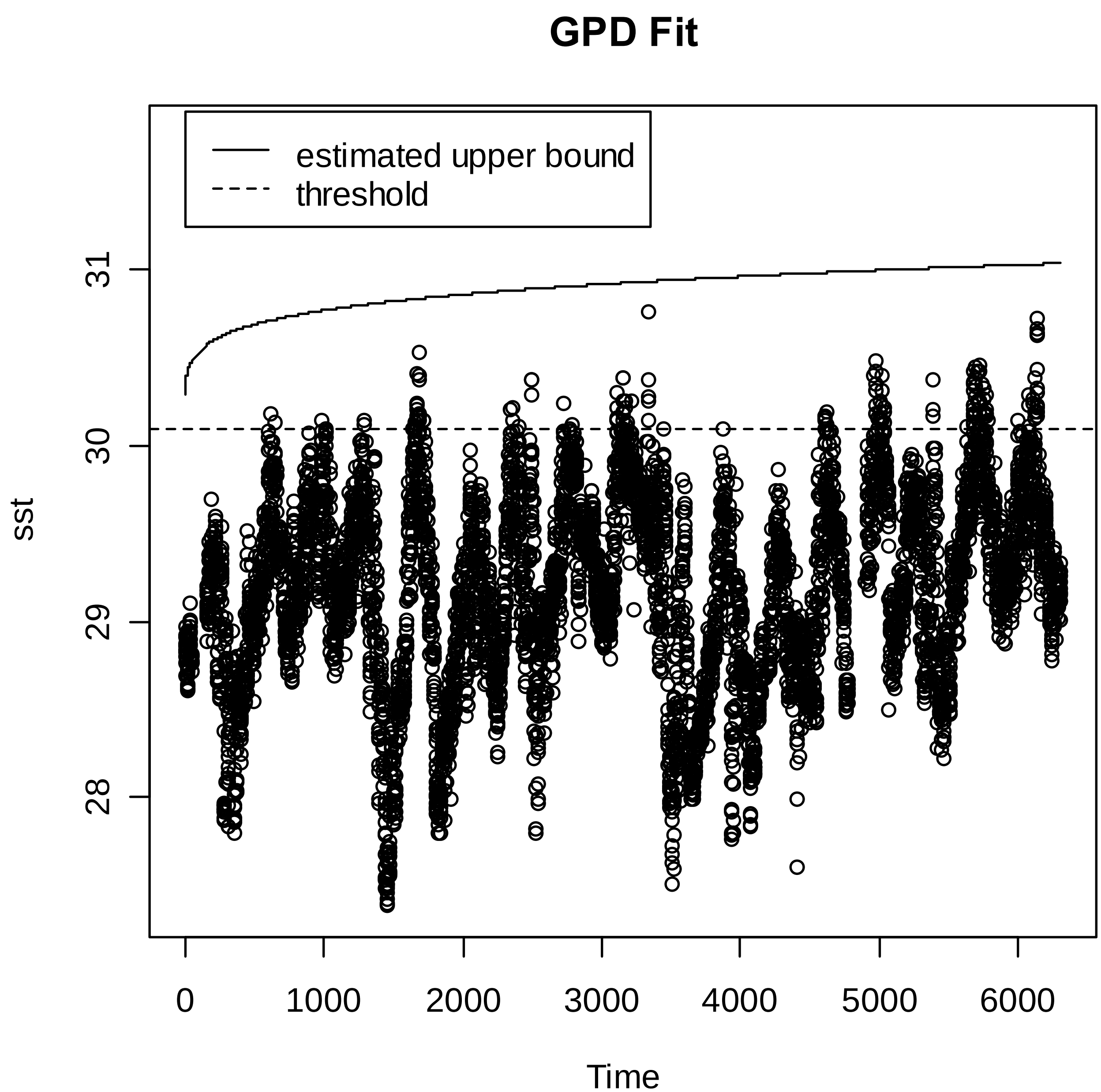
(e. g., pressure or wind speed)

Estimate upper bound (Trend due to global warming?)

-- **“Thermostat hypothesis”**

Hypothesized upper bound on sea surface temperature in tropical oceans (Even with warming trend in mean)

Implications for impact of global warming on coral reefs


GPD Fit
estimated upper bound
threshold
31
30
29
28
sst
0
1000
2000
3000
4000
5000
6000
Time

- **Apparent Heavy Tail**

-- Precipitation

(i) Penultimate approximation

Fréchet type of GEV can be obtained with *F* stretched exponential distribution (Shape parameter $c < 1$)

(ii) Physical argument

Wilson & Toumi (2005) gave heuristic argument for “universal” shape parameter of $c = 2/3$ for stretched exponential distribution for extreme high precipitation

-- Simulation experiment

Generate observations with stretched exponential distribution (with shape parameter $c = 2/3$)

Use block size of $T = 100$ to simulate maxima M_{100}

(Corresponds to daily precipitation occurrence rate about 27%, ignoring variation in number of wet days)

Penultimate approximation:

Should produce GEV shape parameter of $\xi_{100} \approx 0.11$

Fitted GEV distribution (40,000 replications):

Obtained estimate of $\xi_{100} \approx 0.10$

-- Aggregation Issue

Apparent decrease in shape parameter of GEV or GP distribution

Stretched exponential should be capable of resolving (at least qualitatively)

***Simulation experiment*:**

Sum of two independent stretched exponentials (each with $c = 2/3$)

Use block size of $T = 100$ to simulate maxima M_{100}

Fitted GEV dist. (40,000 replications): Estimate $\xi_{100} \approx 0.065$

But note that precipitation really a "random sum"

(5) Unified Modeling of Distributions

Goal: Model entire distribution, *not* just upper tail

- **Stochastic Weather Generators**

-- **Markov chain model for precipitation occurrence**

-- **Precipitation intensity**

Conditionally independent & identically distributed (e. g., gamma distribution)

-- **Multivariate autoregressive process for other daily weather variables (e. g., minimum & maximum temperature) conditional on precipitation occurrence**

- **Alternative approach (Popular in hydrology)**

-- Resampling using bootstrap scheme

-- Unrealistic for extremes

(Like stepping into same river twice!)

- **Fort Collins, CO, USA July daily precipitation intensities**

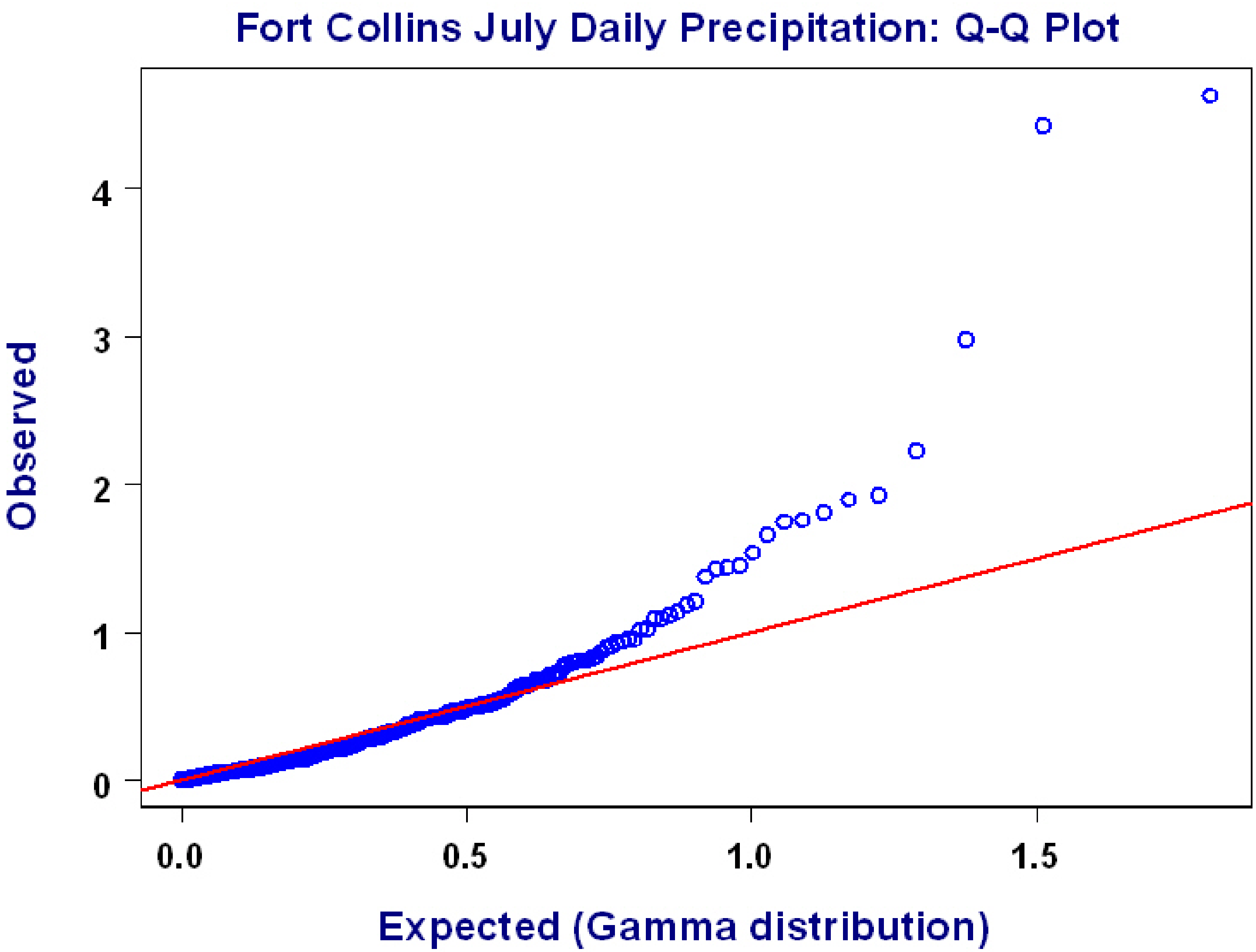
100 yrs. of data, 1900 – 1999

(i) Typical weather generator approach

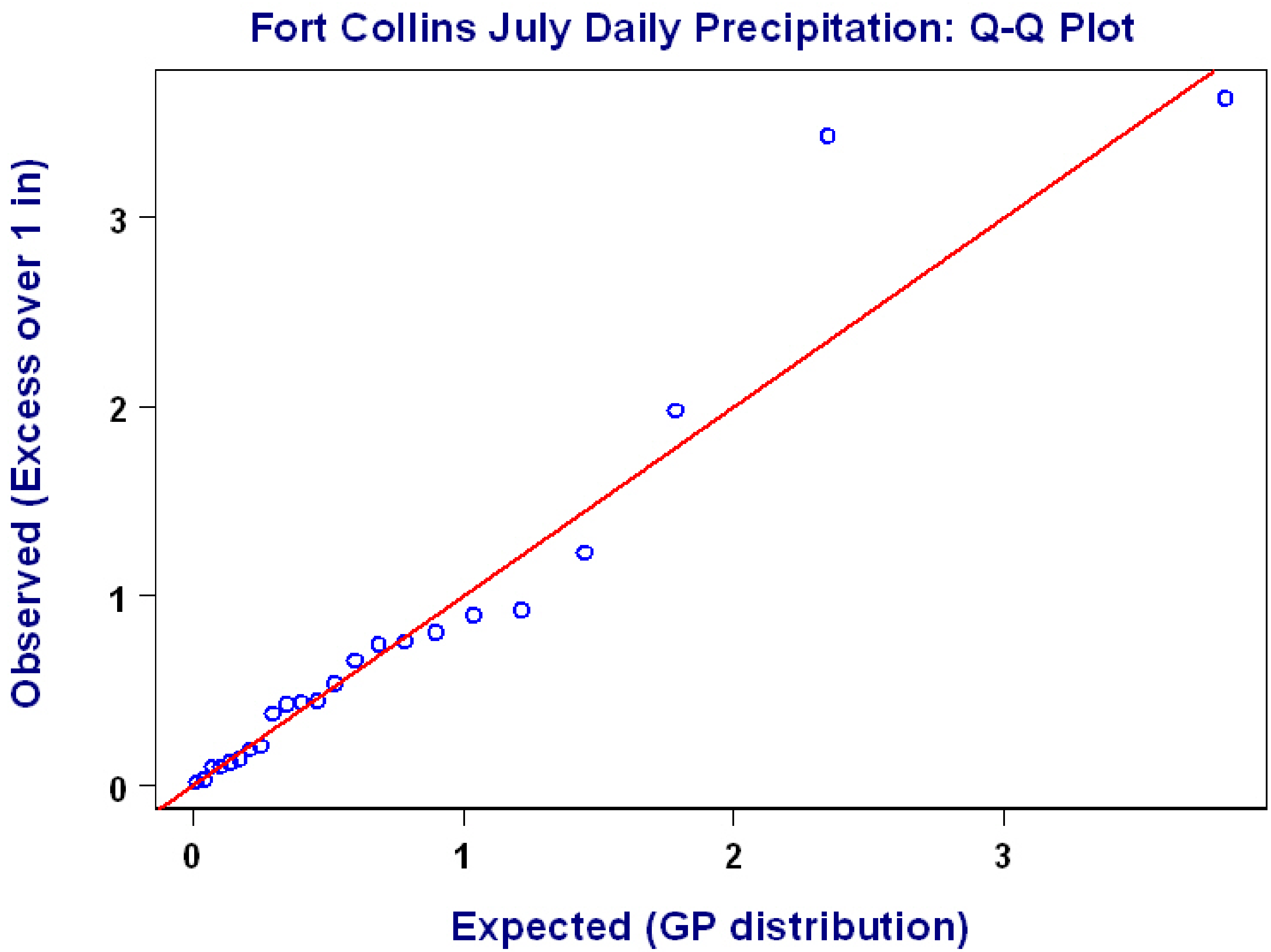
-- Fit gamma distribution to all data

(ii) Extreme value approach

-- Fit GP distribution to upper tail



Fort Collins July Daily Precipitation: Q-Q Plot
Observed
Expected (Gamma distribution)
0
1
2
3
4
0.0
0.5
1.0
1.5


Fort Collins July Daily Precipitation: Q-Q Plot
Observed (Excess over 1 in)
Expected (GP distribution)
0
1
2
3
0
1
2
3

- **GLM Approach for Stochastic Weather Generators**

-- Generalized Linear Model (GLM)

Can fit entire weather generator using only `glm` function in R

Easy to incorporate covariates such as annual cycles or El Niño phenomenon

***Reference*: Furrer & Katz (*Climate Research*, 2007)**

-- Extensions straightforward

Replace gamma distribution for precipitation intensity with heavier-tailed distribution such as stretched exponential

- **Unified treatment**

How to combine distributions?

-- Needs to be simple enough to include covariates in model

(e. g., annual cycles, ENSO index for Pergamino application)

-- Issue of parsimony (too many parameters?)

-- Issue of consistency (e. g., different models for annual cycles)

- **Threshold selection**

-- Extremely difficult for stretched exponential

Should *not* be too low

Should *not* be too high (unlike case of GP distribution)

- **Alternative approach**

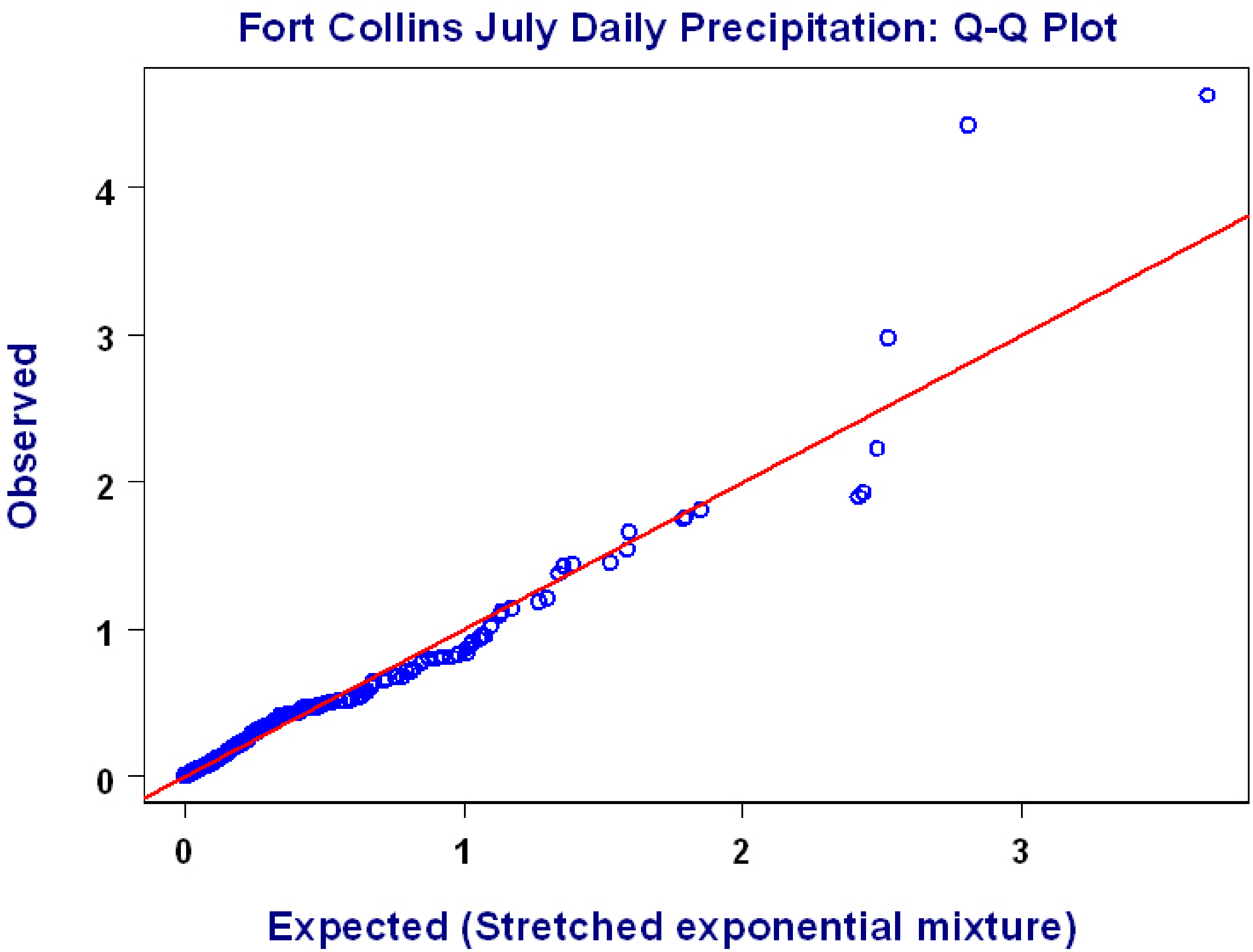
-- Extension of mixture of two exponential distributions (commonly used in hydrology to model precipitation intensity)

-- Outside GLM framework (But avoids threshold selection)

-- Mixture of two stretched exponentials (Fit to all precipitation intensity, *not* just high values)

-- Example

Mixture of exponential (i. e., $c = 1$) & stretched exponential ($c = 2/3$)


Fort Collins July Daily Precipitation: Q-Q Plot
Observed
4
3
2
1
0
0
1
2
3
Expected (Stretched exponential mixture)

- **“Hybrid” Approach**

(i) First Gamma distribution to all data

(ii) Then replace with GP distribution above high threshold

-- How to tie together two pdfs at threshold *u*?

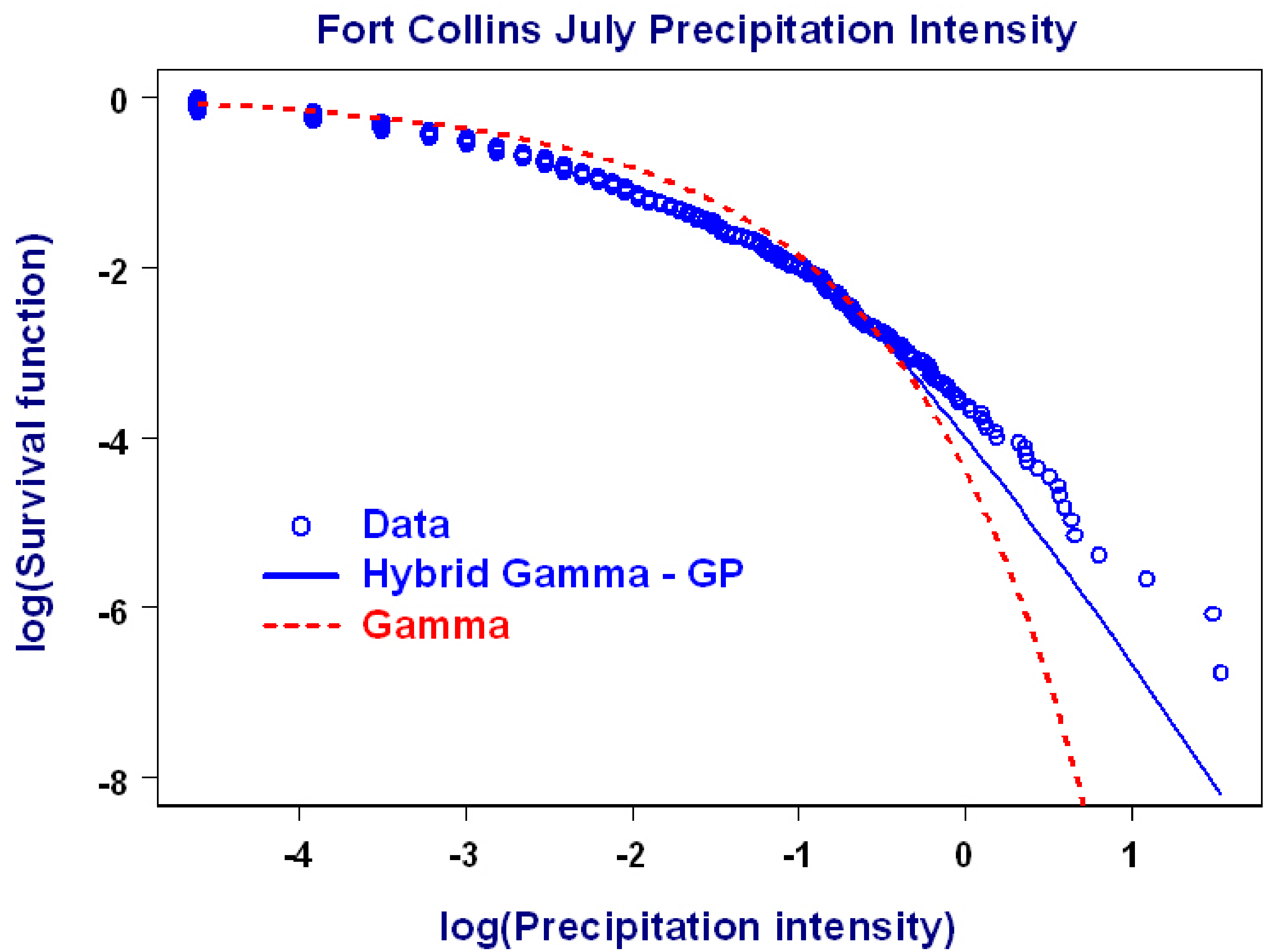
-- Ad hoc procedure

Adjust scale parameter σ of GP distribution:

$$\sigma = 1 / H(u)$$

where *H* is hazard rate for gamma distribution

-- Fort Collins July precipitation intensity, $u = 0.5$ in


Fort Collins July Precipitation Intensity
log(Survival function)
0
-2
-4
-6
-8
-4
-3
-2
-1
0
1
log(Precipitation intensity)
Data
Hybrid Gamma - GP
Gamma

(7) Resources

- **Statistics of Weather and Climate Extremes**

-- Application of statistics of extremes to weather & climate

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

- **Extremes Toolkit (`extRemes`)**

-- Open source software in `R` with GUIs

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

- **GLM Based Weather Generator**

`www.image.ucar.edu/~eva/GLMwgen/`