IMPROVING THE TREATMENT OF EXTREMES IN THE GENERATION OF CLIMATE CHANGE SCENARIOS

Rick Katz

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

Joint work with Eva Furrer (NCAR/IMAGE)

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

Talk: `···/staff/katz/docs/pdf/maltextr.pdf`

QUOTE

- **Heraclitus (Greek philosopher)**

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

OUTLINE

(0) Background / Motivation

(1) Resampling Approach to Weather Generation

(2) Parametric Approach to Weather Generation

(3) Generalized Linear Modeling (GLM) Framework

(4) Upper Tail Modeling of Precipitation Intensity

(5) Unified Modeling of Precipitation Intensity

(6) Hot Spells / Heat Waves

(7) Resources

(0) Background / Motivation

- Climate change impact assessment

-- Need scenarios of daily weather reflecting climate variability or change

- Statistical downscaling of numerical model output

-- Dynamical downscaling

-- Stochastic weather generators

- Performance of stochastic weather generators

-- Need improved treatment of extremes

(1) Resampling Approach to Weather Generation

- **Bootstrap**

-- Sample from observations with replacement

-- Advantages

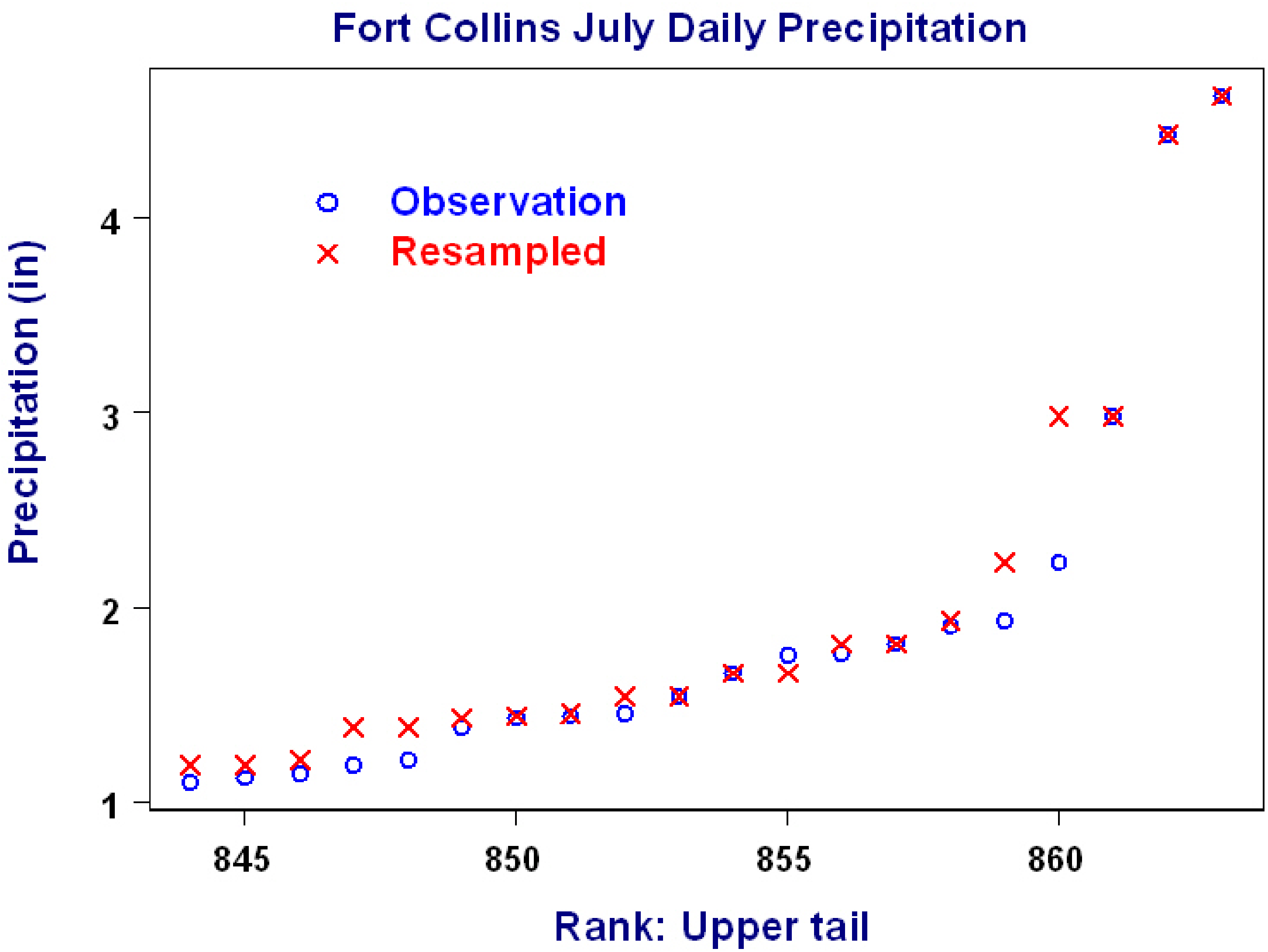
Nonparametric

(Capable of "reproducing" any desired statistic)

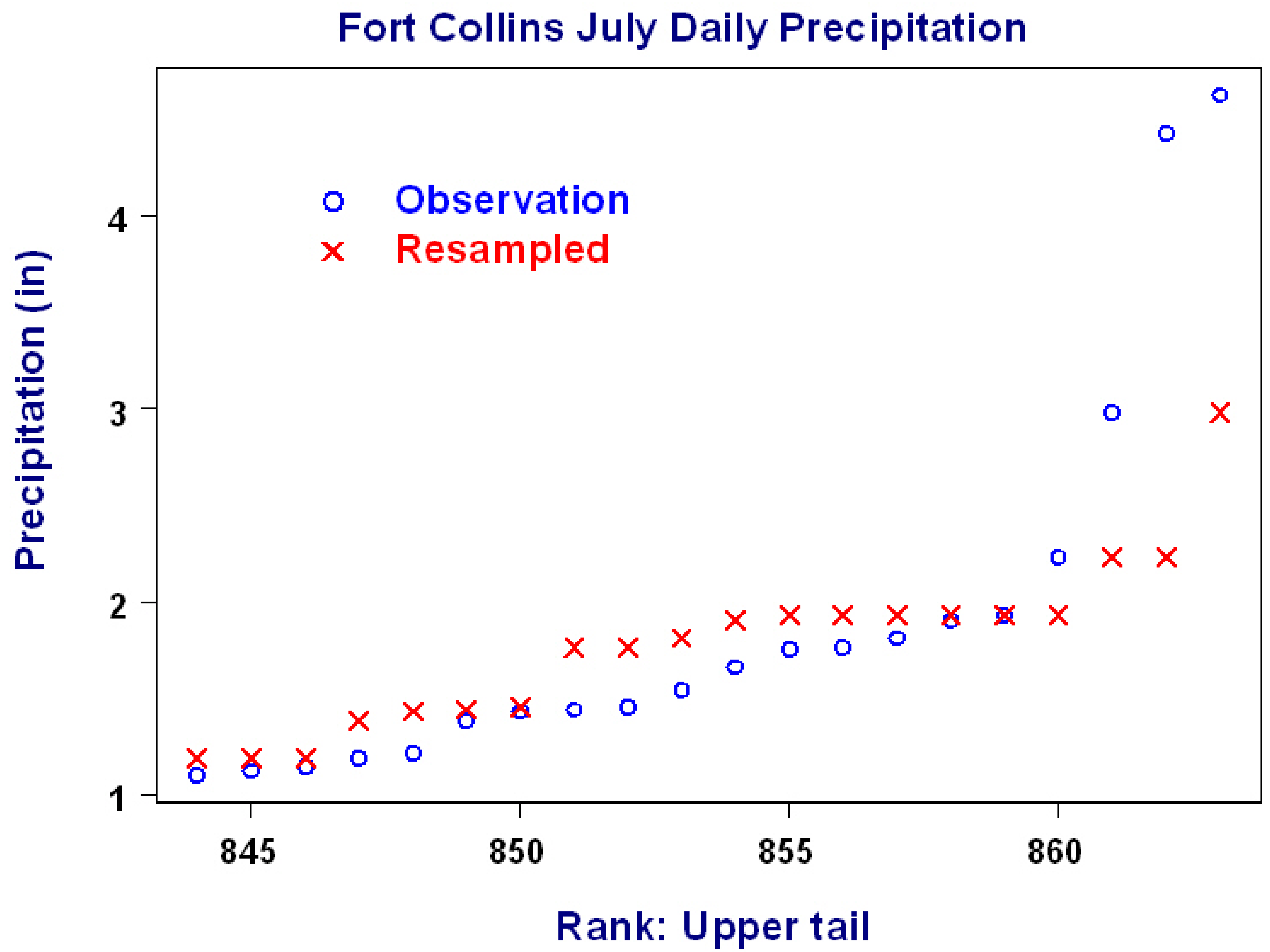
Automatic application

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

100 yrs of data, 1900 – 1999



Fort Collins July Daily Precipitation
Observation
Resampled
Precipitation (in)
1
2
3
4
845
850
855
860
Rank: Upper tail



Fort Collins July Daily Precipitation
Observation
Resampled
Precipitation (in)
1
2
3
4
845
850
855
860
Rank: Upper tail

- Theoretical properties of bootstrap

-- Unrealistic properties for extremes

Not possible to generate value greater than highest observed

Not possible to generate value between second highest & highest

Fails for estimation of maximum value

(Efron & Tibshirani: Need either "parametric knowledge" or "smoothing")

- Conclusion

-- Synthetic weather looks too much like observations

(2) Parametric Approach to Weather Generation

- **Stochastic model for daily weather**

-- Markov chain model for precipitation occurrence

-- Precipitation intensity

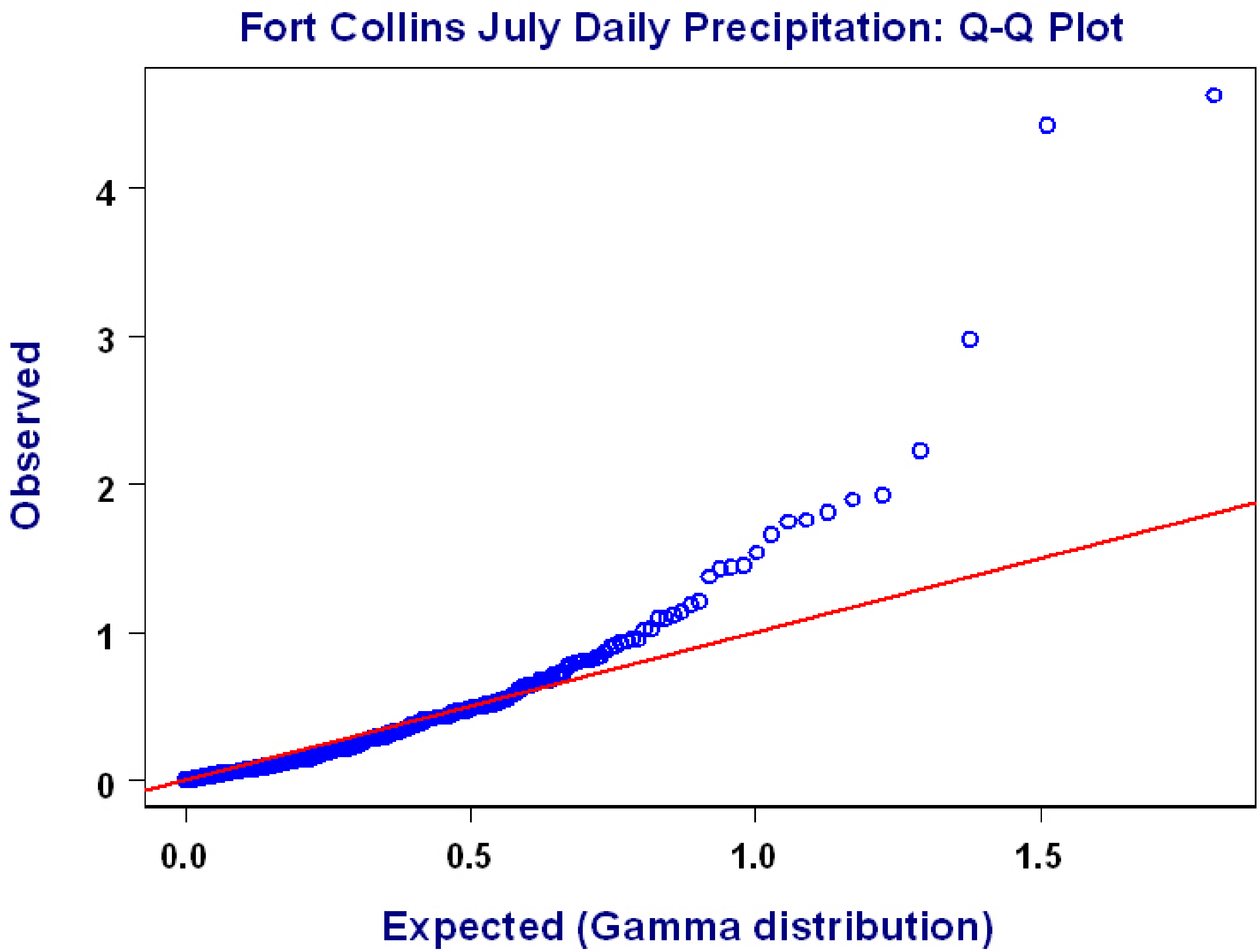
Amount of precipitation on day t, say X_t, given $X_t > 0$

c.d.f. $F(x) = \Pr\{X_t \leq x \mid X_t > 0\}, \; x > 0$

Conditionally independent & identically distributed

e.g., gamma distribution with p.d.f.

$$f(x;\, \alpha, \sigma) = \{1/[\sigma\, \Gamma(\alpha)]\}\, (x/\sigma)^{\alpha - 1} \exp[-(x/\sigma)], \; x, \alpha, \sigma > 0$$


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

-- **Model for other daily weather variables**

(e. g., minimum & maximum temperature)

Multivariate autoregressive process (AR) process

(Usually first-order multivariate AR process [AR(1)])

Conditional on precipitation occurrence (random standardization)

to link temperature to precipitation

-- **Other weather extremes**

Not as serious problem for temperature extremes (except for

spells)

-- **Conclusion**

Synthetic weather does not look enough like observations

(3) Generalized Linear Modeling (GLM) Framework

- Applied to daily weather data for Pergamino, Argentina

- Daily precipitation

-- Markov chain for precipitation occurrence

Annual cycles in transition probabilities, dependence on El Niño–Southern Oscillation (ENSO) index

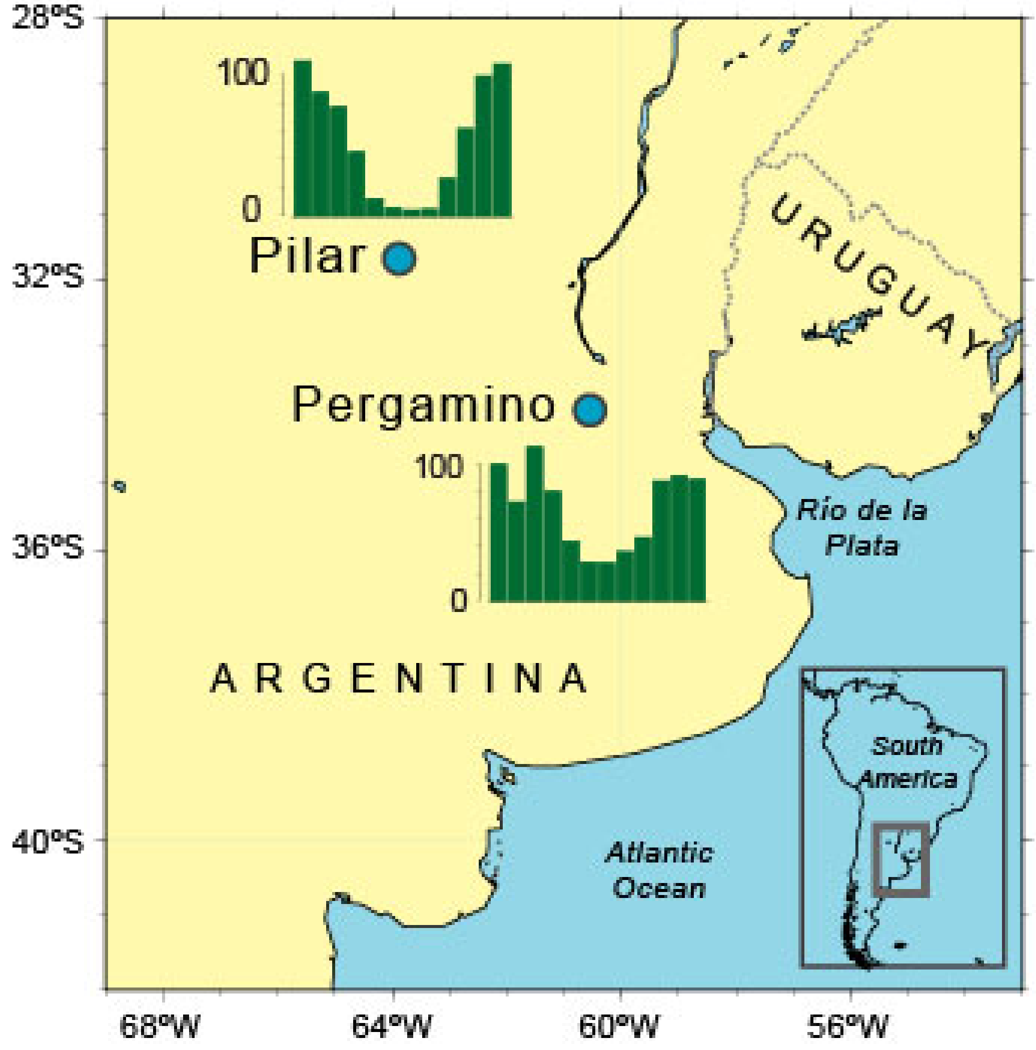
-- Gamma distribution for daily precipitation intensity

Annual cycle in mean intensity, dependence on ENSO index?

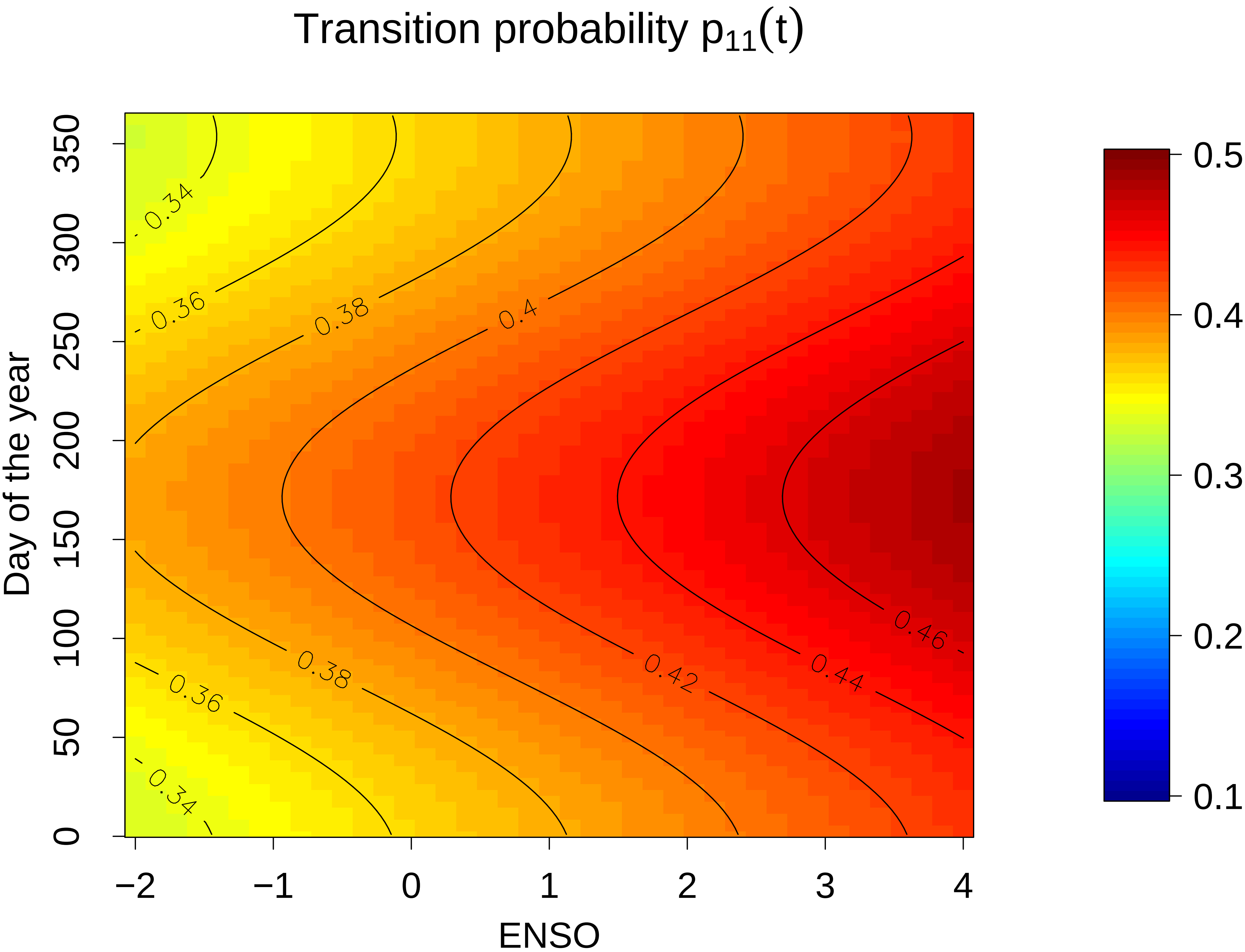
- Daily minimum & maximum temperature

-- Bivariate autoregressive process conditional on prec. occurrence

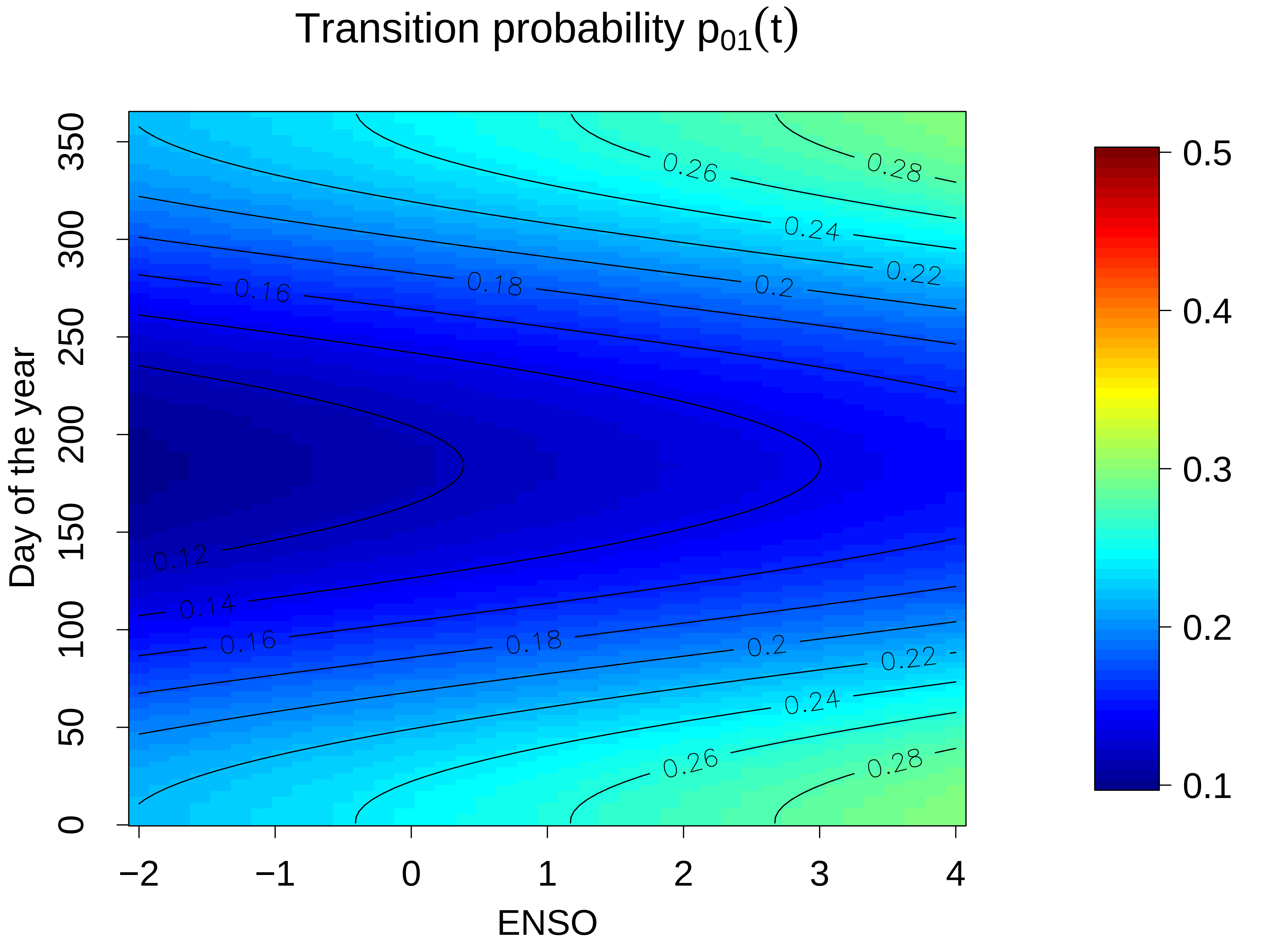
Annual cycles in means, dependence on ENSO index



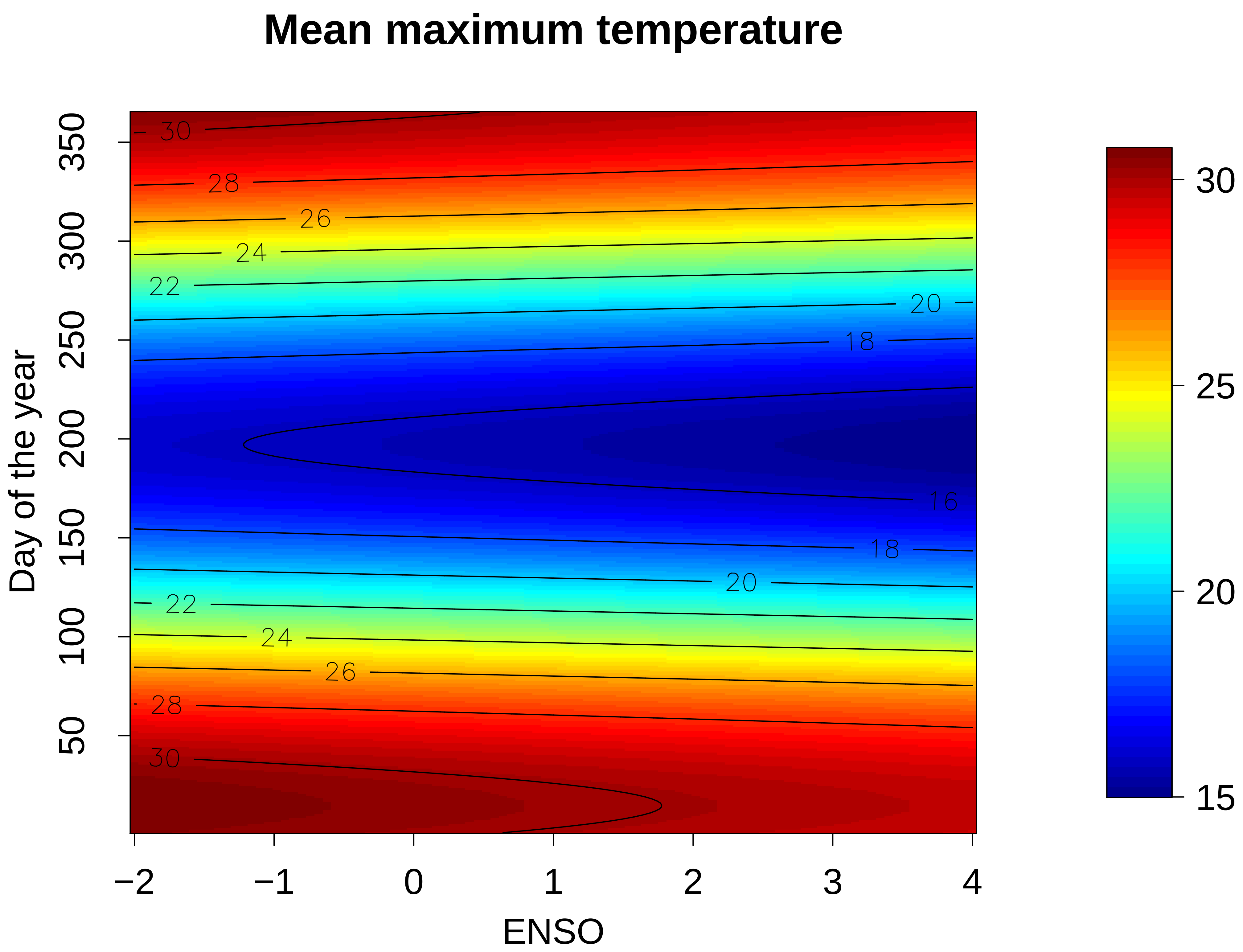
28ºS
32ºS
36ºS
40ºS
68ºW
64ºW
60ºW
56ºW
100
0
Pilar
Pergamino
100
0
URUGUAY
ARGENTINA
Río de la Plata
Atlantic Ocean
South America



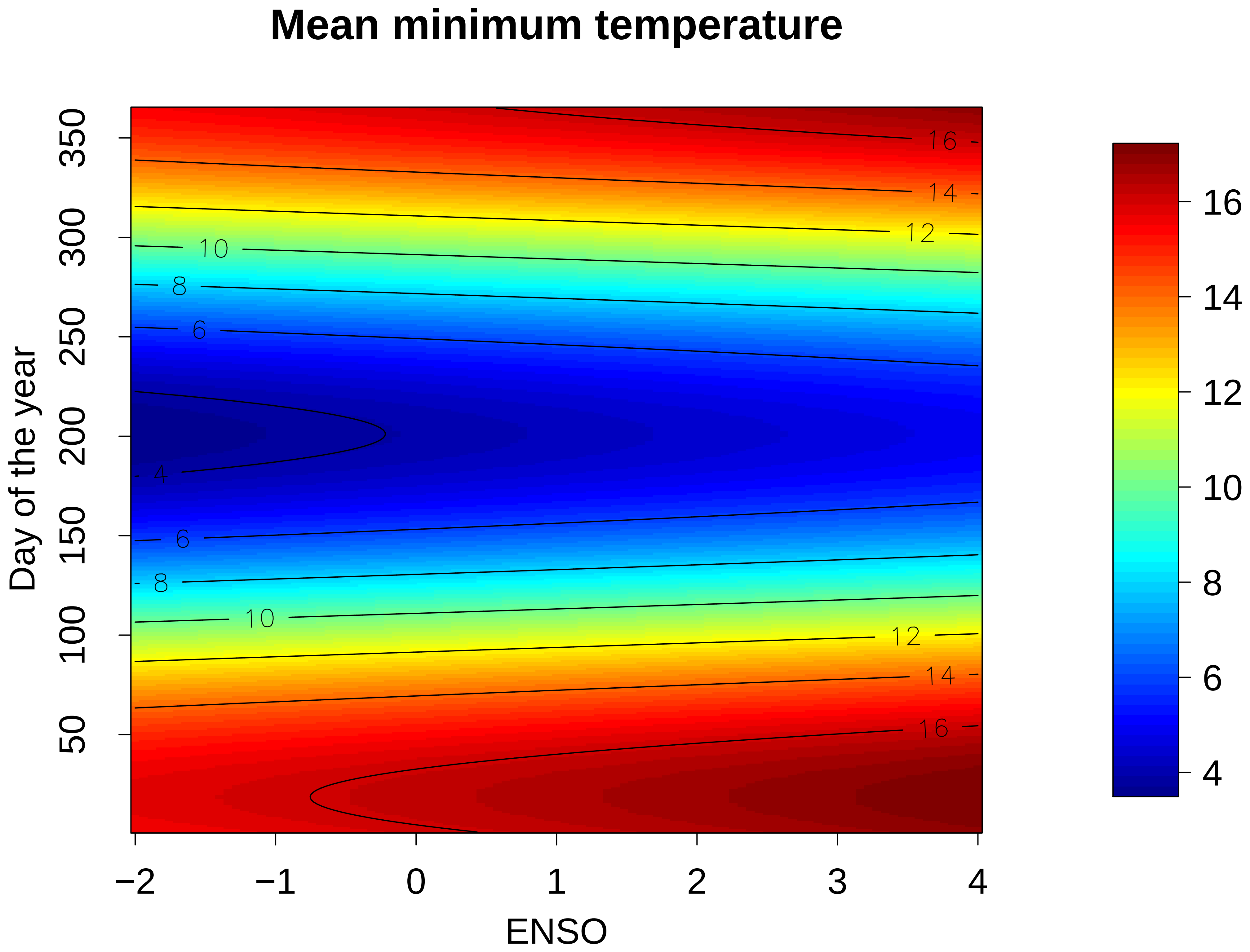
Transition probability $p_{11}(t)$
Day of the year
ENSO
0
50
100
150
200
250
300
350
−2
−1
0
1
2
3
4
0.34
0.36
0.38
0.4
0.42
0.44
0.46
0.1
0.2
0.3
0.4
0.5



Transition probability $p_{01}(t)$
Day of the year
ENSO
0
50
100
150
200
250
300
350
−2
−1
0
1
2
3
4
0.12
0.14
0.16
0.18
0.2
0.22
0.24
0.26
0.28
0.1
0.2
0.3
0.4
0.5



Mean maximum temperature
Day of the year
ENSO
350
300
250
200
150
100
50
−2
−1
0
1
2
3
4
30
28
26
24
22
20
18
16
18
20
22
24
26
28
30
30
25
20
15



Mean minimum temperature
Day of the year
ENSO
350
300
250
200
150
100
50
−2
−1
0
1
2
3
4
16
14
12
10
8
6
4

- **Software**

-- **R open source statistical programming language**

(function `glm`)

- **Advantages of GLM framework**

-- **Improved treatment of extremes**

(e. g., replace gamma distribution for precipitation intensity with heavier-tailed distribution)

- **Extension to multiple sites**

-- **Precipitation (Yang et al. 2005)**

(4) Upper Tail Modeling of Precipitation Intensity

- **Point process approach ("peaks over threshold")**

-- Poisson process for exceedance of high threshold

-- Generalized Pareto (GP) distribution for excess over high threshold u with c.d.f.

$$F(x;\, \xi, \sigma, u) = 1 - \{1 + \xi\, [(x - u)/\sigma]\}^{-1/\xi},$$

$$x > u,\ \sigma > 0,\ 1 + \xi\, [(x - u)/\sigma] > 0$$

Shape parameter ξ governs tail behavior (heavy tail if $\xi > 0$)

-- Include covariates in model

Annual cycles

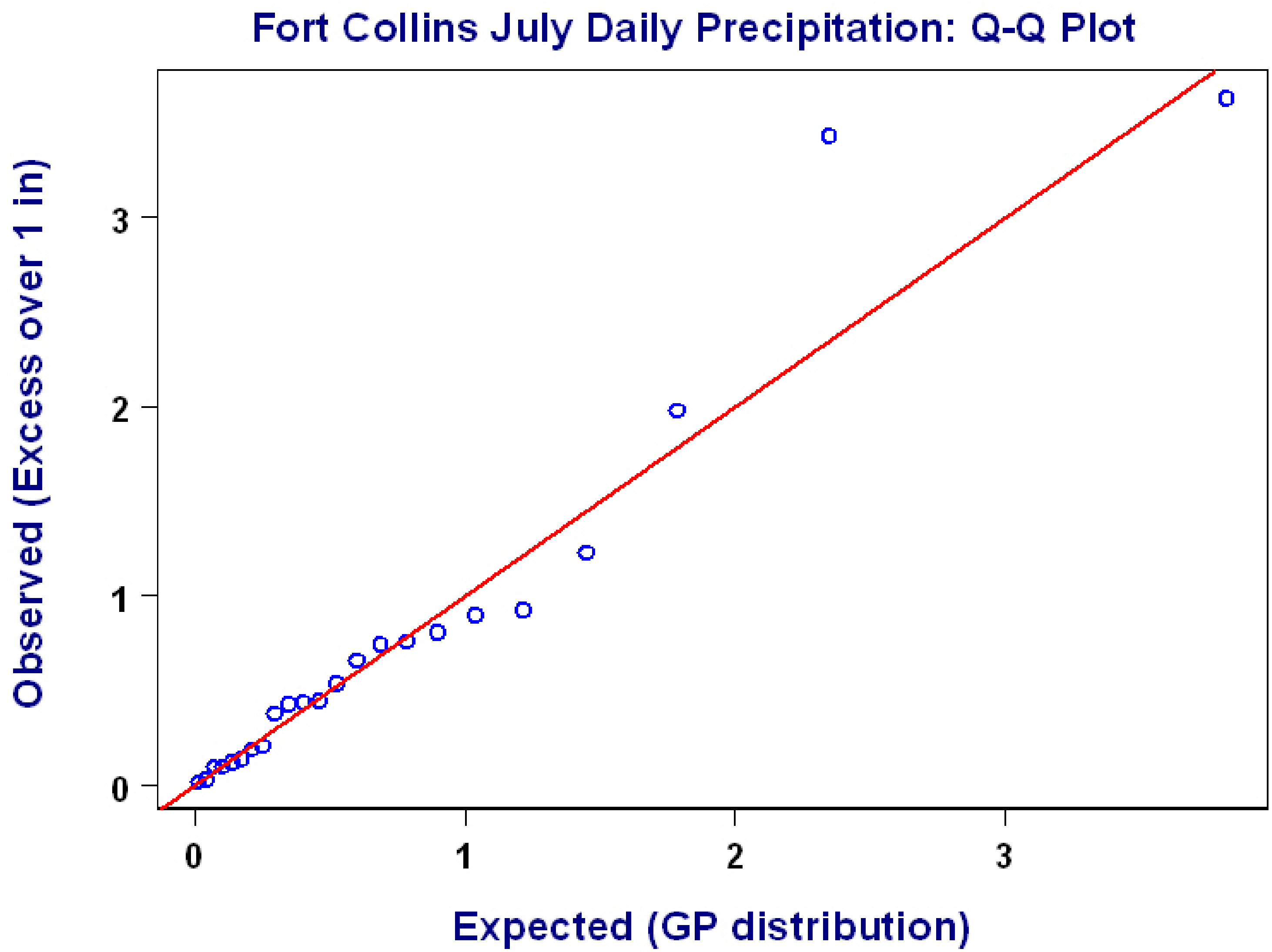
ENSO index (for Pergamino application)

-- Fort Collins example

Estimates of shape parameter ξ

(i) Annual maxima (GEV): $\xi = 0.17$

(ii) Annual cycles in location & scale parameters of GEV (point process approach, $u = 0.4$ in): $\xi = 0.18$


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

(5) Unified Modeling of Precipitation Intensity

- **Approaches to modeling entire distribution**

(i) Replace gamma with more flexible distribution

-- e.g., stretched exponential

(ii) Mixture of distributions

-- e.g., mixture of exponentials

(iii) Hybrid approach

-- e.g., gamma for low to moderate values

GP for high values

(i) **Replace gamma with more flexible distribution**

- **"Ultimate" extreme value theory**

-- GEV distribution as limiting distribution of maxima

$X_1, X_2, \ldots, X_T$ **with common c.d.f.** F

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

- **Penultimate extreme value theory**

-- Suppose d.f. 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 $\xi \neq 0$**)**

-- Corresponding interpretation of GP for excess over threshold

-- **Expression for shape parameter ξ_T**

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

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

One choice of shape parameter is:

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

where the "characteristic largest value"

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

***Note*:**

$$\xi_T \to 0 \text{ as block size } T \to \infty$$

- “Stretched exponential” distribution

-- Traditional form of Weibull distribution (Bounded below)

Note: Weibull extremal type is reflected version with c.d.f.

$$F(x; c, \sigma, u) = 1 - \exp\{-[(x - u)/\sigma]^{c}\}, \quad x > u, \sigma, c > 0$$

where c is shape parameter

-- Shape parameter of GEV for penultimate approximation

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

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

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

- Wilson & Toumi (2005)

-- Heuristic argument that $c = 2/3$

Assume block size $T = 100$

(i.e., precipitation occurs on about 27% of days within year)

Penultimate approximation: $\xi_T \approx 0.11$

- Simulation experiment

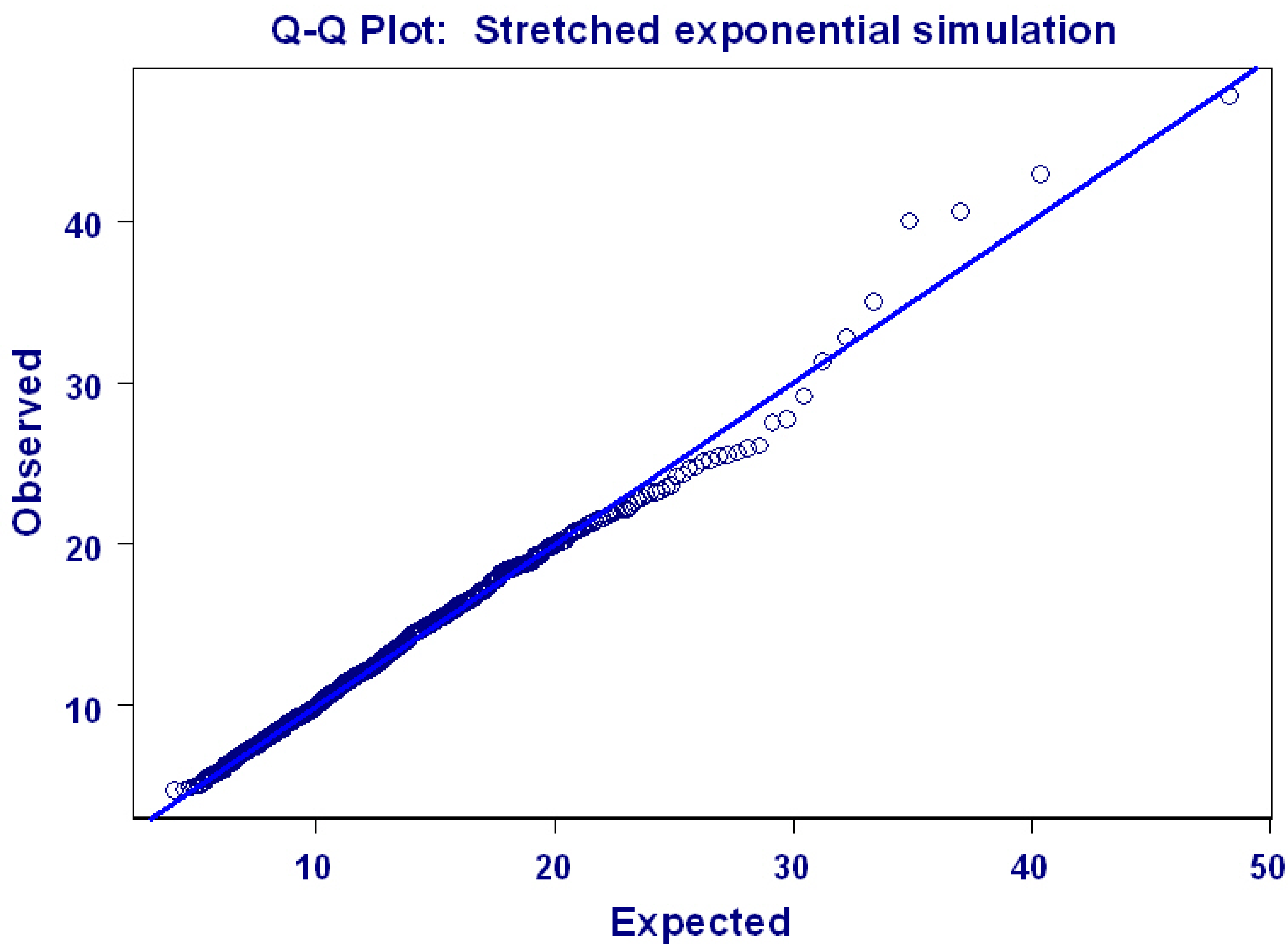
-- Generate observations with stretched exponential distribution

(with shape parameter $c = 2/3$)

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

Fitted GEV distribution (40,000 replications):

Obtained estimate of $\xi_{100} \approx 0.12$


Q-Q Plot: Stretched exponential simulation
Observed
Expected
10
20
30
40
10
20
30
40
50

- **Threshold selection**

-- Extremely difficult for stretched exponential

Should *not* be too low

Should *not* be too high

(unlike case of GP distribution)

-- Alternative of mixture of stretched exponentials

Avoid threshold selection

(ii) Mixture of distributions

- **Mixture of two stretched exponentials**

-- Fit to all precipitation intensity, *not* just high values

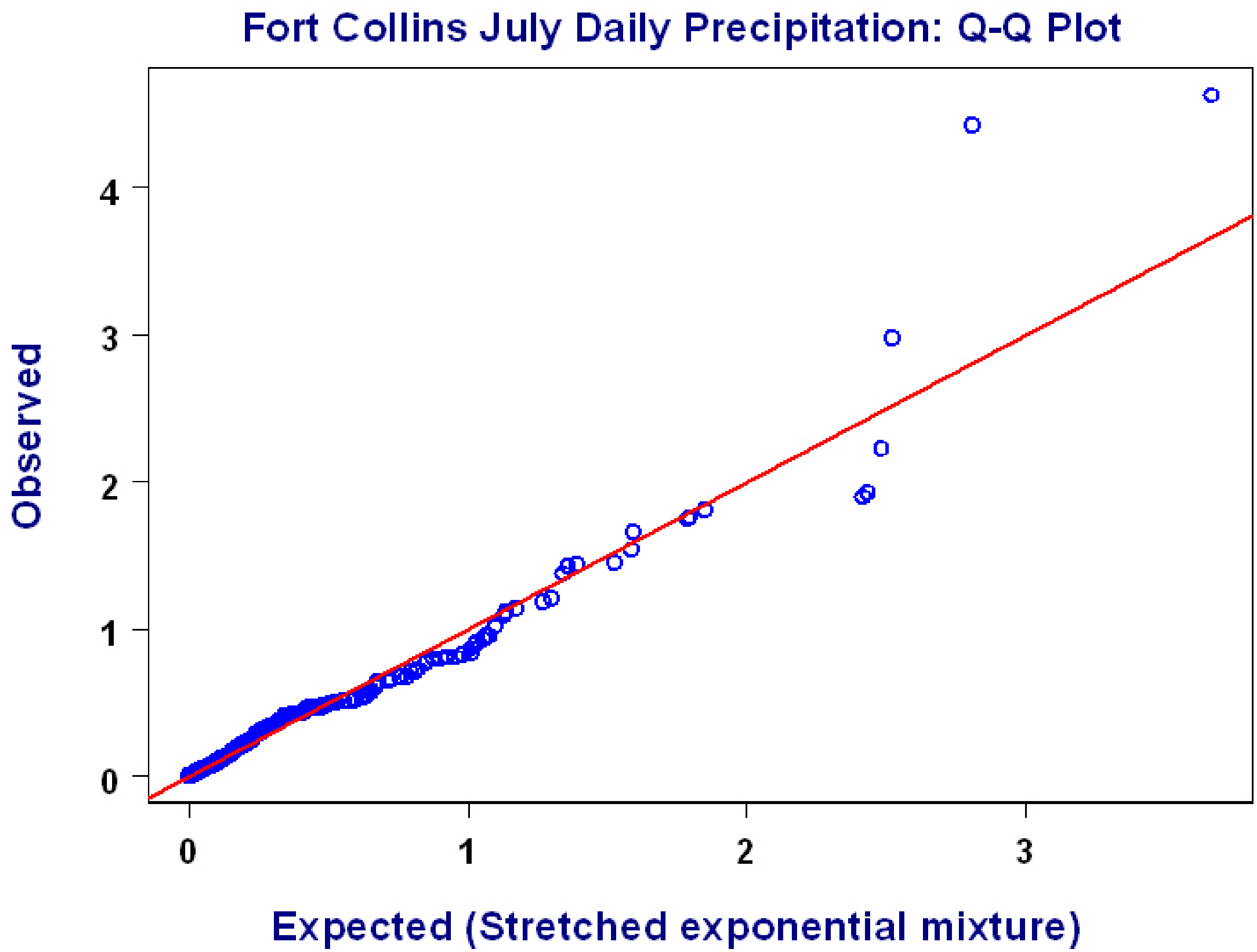
-- Example

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

c.d.f. $F(x; w, \sigma_1, \sigma_2, c) =$

$$1 - (1 - w) \exp[-(x/\sigma_1)] - w \exp[-(x/\sigma_2)^c],$$

$$x, \sigma_1, \sigma_2, c > 0, \; 0 < w < 1$$


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

(iii) Hybrid approach

- **Given gamma distribution (fit to all data)**

-- Replace with GP distribution above high threshold

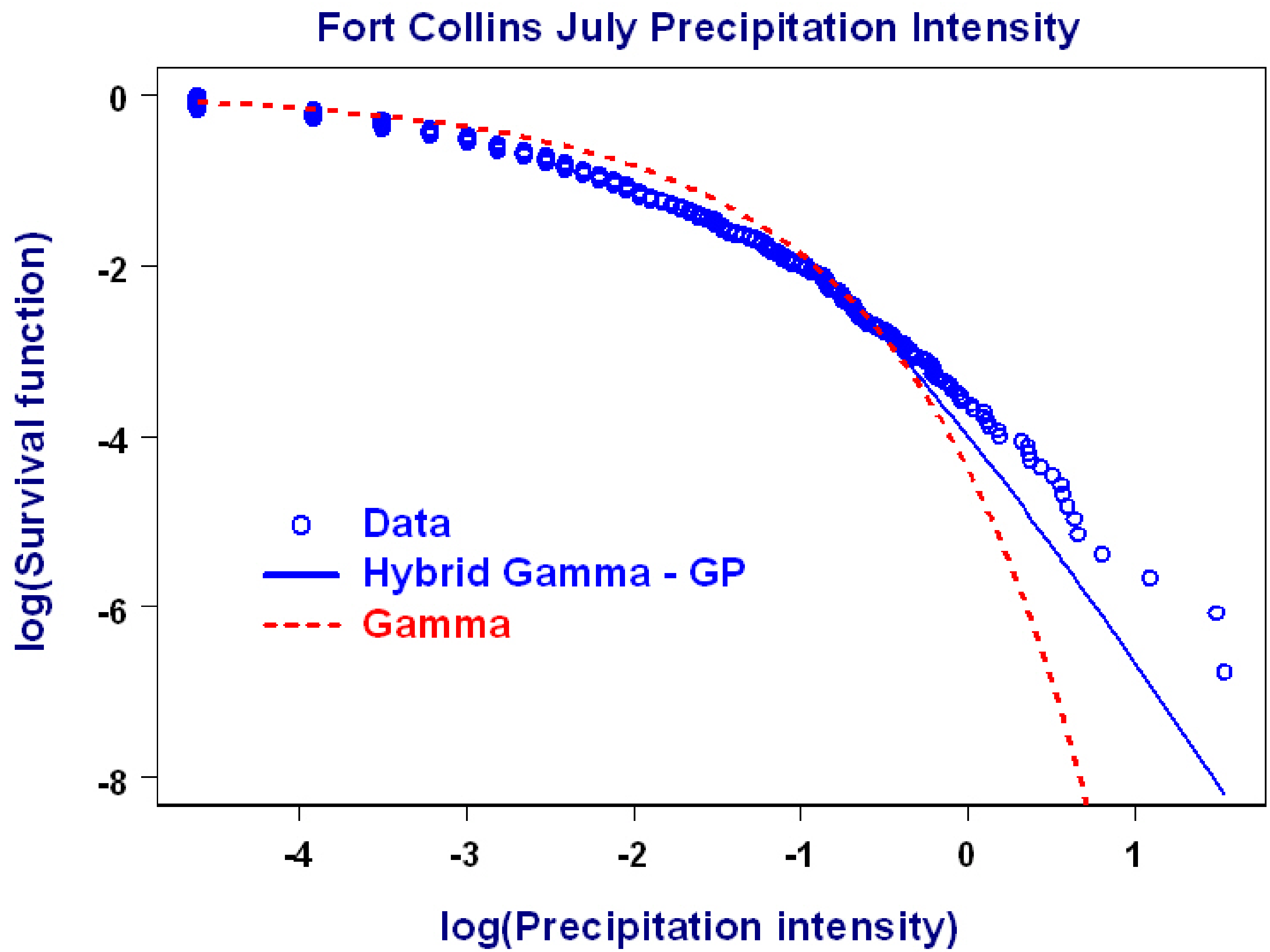
-- How to tie together two pdfs at threshold u?

-- 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

(6) Hot Spells / Heat Waves

- **Hot Spells / Heat Waves**

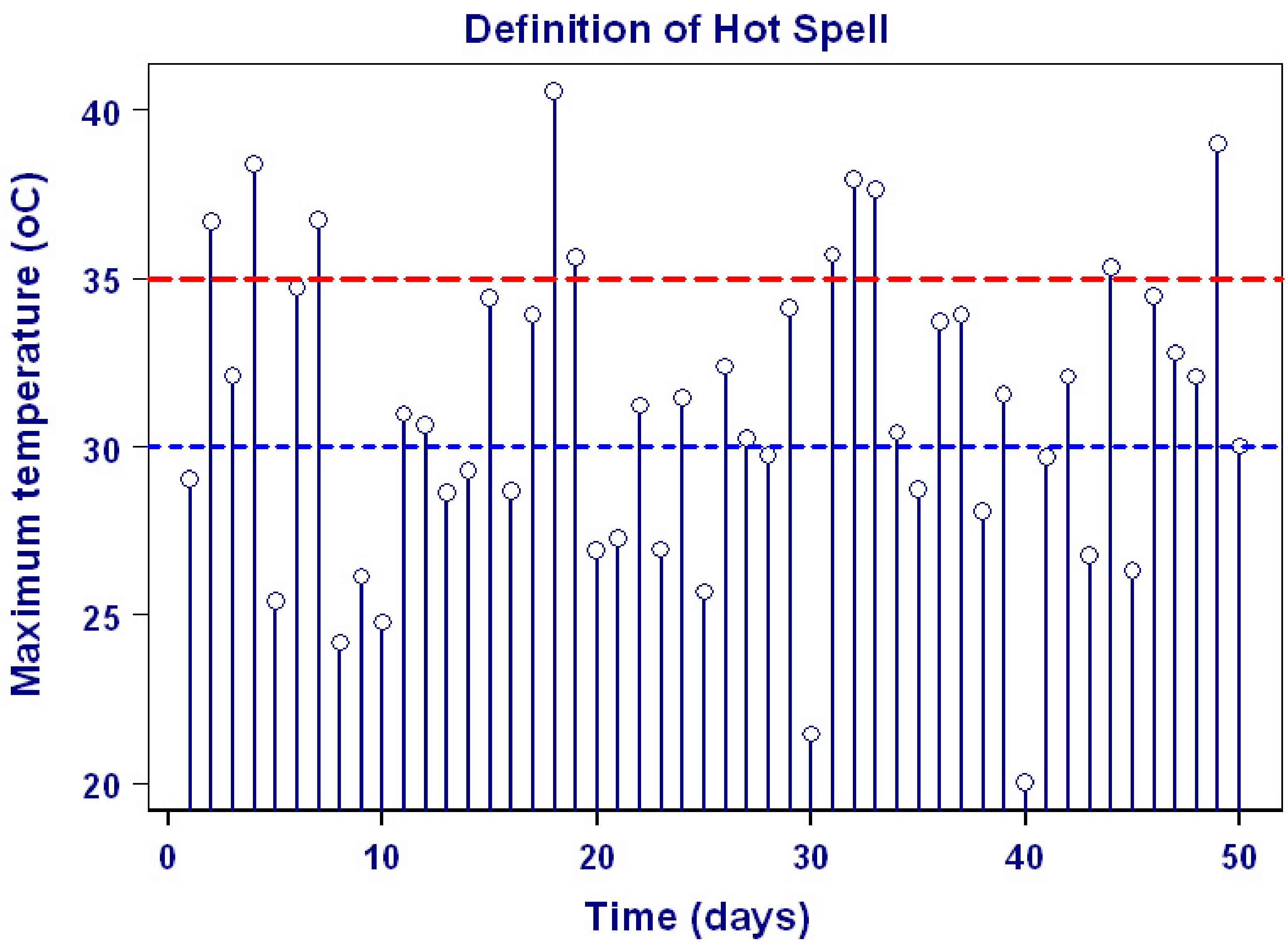
-- Approach based on extreme value theory

Extension of point process approach / De-clustering

-- Statistical modeling of clusters

Geometric distribution for cluster length (truncated?)

Conditional GP distribution for temporal dependence of excesses within cluster


Definition of Hot Spell
Maximum temperature (oC)
Time (days)
40
35
30
25
20
0
10
20
30
40
50

- **Covariate approach**

-- Advantage

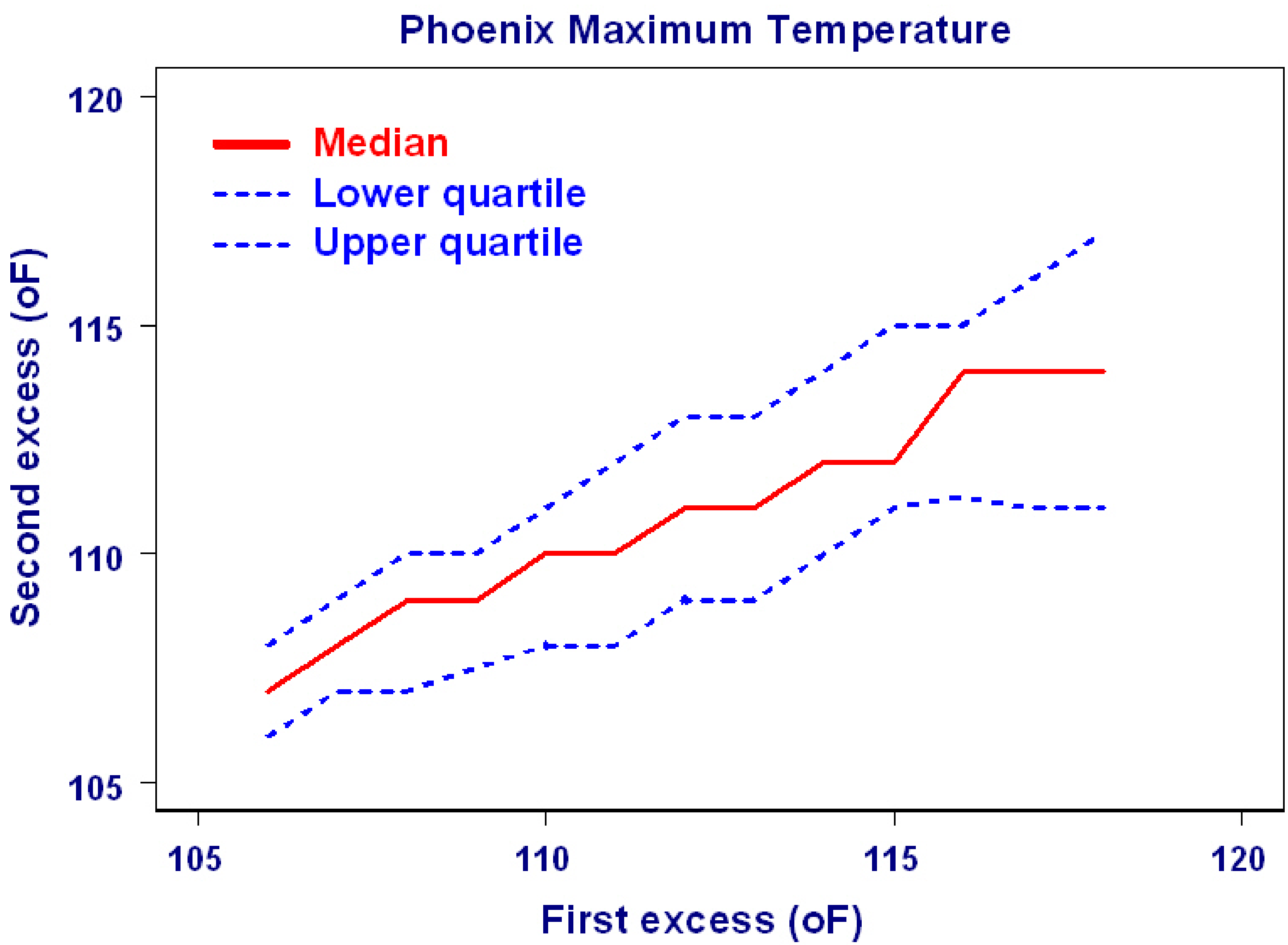
Requires only univariate extreme value theory (not multivariate)

-- Let $Y_1, Y_2, \ldots, Y_k$ denote excesses over threshold within given cluster / spell

Model conditional distribution of Y_2 given Y_1 as GP distribution with scale parameter σ depending on Y_1:

e.g., $\sigma(y) = \sigma_0 + \sigma_1 y$, given $Y_1 = y > 0$

Similar model for conditional distribution of Y_3 given Y_2 (etc.)



Phoenix Maximum Temperature
Median
Lower quartile
Upper quartile
Second excess (oF)
First excess (oF)
105
110
115
120
105
110
115
120

- **Stochastic weather generators**

-- Use hot spell model to evaluate performance in simulating heat waves

-- How to improve performance?

Not necessarily straightforward to implement changes consistent with hot spell model

Could at least increase order of multiple AR(1) process

(7) Resources

- `R` **(Open source statistical programming language)**

-- `www.r-project.org/`

- **GLM Weather Generator**

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

- **Statistics of Weather and Climate Extremes**

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

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

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

- **Papers**

-- ***Climate Research* (2007)**

"Generalized linear modeling approach to stochastic weather generators"

`www.isse.ucar.edu/staff/katz/docs/pdf/crglm.pdf`

-- ***Water Resources Research* (2008)**

"Improving the simulation of extreme precipitation events by stochastic weather generators"

`www.isse.ucar.edu/staff/katz/docs/pdf/penult.pdf`