



*Introduction*

*Statistical Background*

*U.S. Precipitation, part 1*

*U.S. Precipitation, part2*

*Current work*

*Home Page*

*Title Page*



*Page 1 of 17*

*Go Back*

*Full Screen*

*Close*

*Quit*

# Extreme Value Theory for Global Climate Change and Atmospheric Pollution

National Institute of Statistical Sciences

Amy Grady

Funded by the

U.S. Environmental Protection Agency



# 1. Introduction

Research in global change (and atmospheric pollution) has become increasingly focused on extreme events. For example,

***General consensus on U.S precipitation:*** While there is an overall increasing trend in daily rainfall totals, the trend is concentrated in the more "extreme" levels and this pattern of behavior seems fairly uniform across the U.S.A.

***Method:*** Statistically more simplistic approaches – typically, trend analysis of summary statistics.

***Possible new emphasis:*** As attention becomes more focused on questions of statistical significance and agreement between model-generated and observational data, we expect greater attention to be paid to the statistical aspect of the research – and thus to more sophisticated statistical methodologies

Introduction

Statistical Background

U.S. Precipitation, part 1

U.S. Precipitation, part2

Current work

Home Page

Title Page



Page 2 of 17

Go Back

Full Screen

Close

Quit



Introduction

Statistical Background

U.S. Precipitation, part 1

U.S. Precipitation, part2

Current work

Home Page

Title Page



Page 3 of 17

Go Back

Full Screen

Close

Quit

## Our contribution:

---

In this research coop, we will make use of

- Established models and methodologies from Extreme Value Theory and Spatial Statistics, *to be outlined.*
- Recent developments concerning the joint distribution of sums and extremes, *to be outlined.* With further developments...
- Recent development of the combination of models for the distribution of extreme events at a single station with spatial models for the parameters across a network of stations, *to be outlined.* With further developments...



*Introduction*

*Statistical Background*

*U.S. Precipitation, part 1*

*U.S. Precipitation, part2*

*Current work*

*Home Page*

*Title Page*



*Page 4 of 17*

*Go Back*

*Full Screen*

*Close*

*Quit*

## (Possible) Research projects/applications:

- Daily rainfall series across the continental U.S, consisting of 5873 stations – **First up**
- Incorporate the Canadian daily rainfall series into the above investigation.
- Model-generated data. (with Dr. Gabrielle Hegerl)
- Corresponding temperature series.



*Introduction*

*Statistical Background*

*U.S. Precipitation, part 1*

*U.S. Precipitation, part2*

*Current work*

*Home Page*

*Title Page*



*Page 5 of 17*

*Go Back*

*Full Screen*

*Close*

*Quit*

## 2. Statistical Background

### OUTLINE OF E.V.T. METHODS

1. Annual maxima approach: Extreme Value Dist.

“Three types”: Gumbel, Fréchet, Weibull  $\Rightarrow$  GEV

2. Joint Density of (Annual) Maxima and Total

3. Threshold (POT) approaches

4. Point process viewpoint



# ANNUAL MAXIMA APPROACH

Introduction

Statistical Background

U.S. Precipitation, part 1

U.S. Precipitation, part 2

Current work

Home Page

Title Page



Page 6 of 17

Go Back

Full Screen

Close

Quit

*Generalized Extreme Value distribution:*

$$F_X(x) = \Pr\{X \leq x\} = \exp \left[ - \left\{ 1 + \xi \frac{x-\mu}{\sigma} \right\}_+^{-1/\xi} \right]$$

$(y_+ = \max(y, 0))$

$\mu$ : location parameter

$\sigma$ : scale parameter

$\xi$ : shape parameter

When  $\xi = 0$ ,

$$F_X(x) = \exp \left[ - \exp \left( - \frac{x-\mu}{\sigma} \right) \right].$$

$\xi > 0$ : long-tailed, Fréchet

$\xi = 0$ : exponential tail, Gumbel

$\xi < 0$ : short-tailed, Weibull



# Joint Density of Sum and Maxima

Under suitable regularity conditions,

$$f_{M_n^*, S_n^*}(v, w) \approx f_{M_n^*}(v) f_{S_n^*}(w) (1 + h_n(w, v))$$

where

$$\begin{aligned} h_n &= -\frac{b_n(e_{-v} - 1)w}{\sqrt{n\sigma^2}} \quad \text{Gumbel, } x_o = \infty \\ &= -\frac{(x_o - \mu)(e_{-v} - 1)w}{\sqrt{n\sigma^2}} \quad \text{Gumbel, } x_o < \infty \\ &= -\frac{a_n v (\frac{\alpha}{\alpha-1} v^{-\alpha} - 1)w}{\sqrt{n\sigma^2}} \quad \text{Fréchet} \\ &= -\frac{(x_o - \mu)((-v)^\alpha - 1)w}{\sqrt{n\sigma^2}} \quad \text{Weibull.} \end{aligned}$$

**In practice, Joint Density Model**

$$f_{S_n, M_n}(x, y) \approx \mathcal{N}'(x|y) H'(y)$$

where  $\mathcal{N}'(x|y)$  denotes the conditional normal density whose moments are  $E(S_n|M_n = y)$  and  $Var(S_n|M_n = y)$  and again  $H'(y)$  denotes the GEV distribution.

Introduction

Statistical Background

U.S. Precipitation, part 1

U.S. Precipitation, part 2

Current work

Home Page

Title Page

◀

▶

◀

▶

Page 7 of 17

Go Back

Full Screen

Close

Quit



*Introduction*

*Statistical Background*

*U.S. Precipitation, part 1*

*U.S. Precipitation, part2*

*Current work*

*Home Page*

*Title Page*



*Page 8 of 17*

*Go Back*

*Full Screen*

*Close*

*Quit*

# POT methods

---

Separate models for

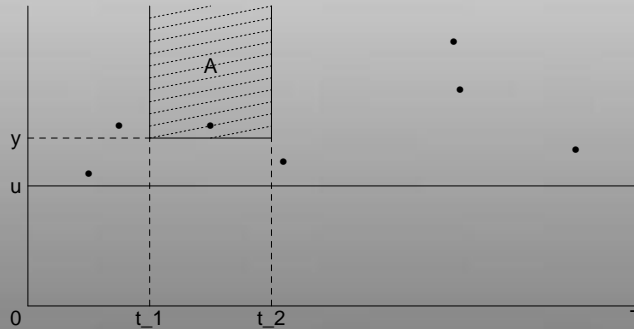
- times of crossing a threshold
- distribution of excess values

Simplest case (when raw data are independent, identically distributed)

- Times of crossing the threshold form a homogeneous Poisson process in one dimension (need to decluster in practice)
- Excess values: exponential, Pareto, generalized Pareto distributions (Cole, 2001)

# Point Process Approach (Cole, 2001)

Similar to POT approach:  
Homogeneous point process case



Let  $\Lambda(A)$  be observed number of points in the set  $A$ .  
Formula (homogeneous case):

$$\Lambda(A) = (t_2 - t_1)\Psi(y; \mu, \sigma, \xi),$$

where

$$\Psi(y; \mu, \sigma, \xi) = \left(1 + \xi \frac{y - \mu}{\sigma}\right)_+^{-1/\xi}.$$



Introduction

Statistical Background

U.S. Precipitation, part 1

U.S. Precipitation, part 2

Current work

Home Page

Title Page

◀

▶

◀

▶

Page 9 of 17

Go Back

Full Screen

Close

Quit



Introduction

Statistical Background

U.S. Precipitation, part 1

U.S. Precipitation, part2

Current work

Home Page

Title Page



Page 10 of 17

Go Back

Full Screen

Close

Quit

## Inhomogeneous case: Time-varying

Extension (Inhomogeneous case):

$$\Lambda(A) = \int_{t_2}^{t_1} \Psi(y; \mu_t, \sigma_t, \xi_t) dt,$$

where  $\mu_t, \sigma_t, \xi_t$  are allowed to vary with time  $t$  (possibly including other covariates)



Introduction

Statistical Background

U.S. Precipitation, part 1

U.S. Precipitation, part2

Current work

Home Page

Title Page



Page 11 of 17

Go Back

Full Screen

Close

Quit

## 3. U.S. Precipitation, part 1

### RAINFALL DATA EXPERIMENT

Preliminary work: Smith(1999)

**Data:** 187 stations of daily rainfall data from HCN network. Analysis is restricted to 1951–1997 during which coverage percentage is fairly constant.

**Model:** Time-varying Point Process

**Covariates:** Various covariates were tried (linear trend, seasonality, SOI, NAO, ...). *Important here* – The linear trend was always significant.



*Introduction*

*Statistical Background*

*U.S. Precipitation, part 1*

*U.S. Precipitation, part2*

*Current work*

*Home Page*

*Title Page*



Page *12* of *17*

*Go Back*

*Full Screen*

*Close*

*Quit*

## Next question:

This analysis is then run on all 187 stations.

**Question:** How to integrate the results from 187 stations in a meaningful way?



# SPATIAL INTEGRATION

---

$Z(s)$ : true but unobserved spatial field, indexed by location  $s$  (e.g. a long-term trend)

$\hat{Z}(s)$ : estimate of  $Z(s)$  at site  $s$ , e.g. from a regression analysis

2-stage model: universal kriging model with measurement error

$$Z \sim N[X\beta, \Sigma], \hat{Z} | Z \sim N[Z, W].$$

Combined:

$$\hat{Z} \sim N[X\beta, \Sigma + W].$$

In rainfall application,  $\Sigma$  taken as a exponential or Matérn spatial covariance function,  $X\beta$  represents quadratic spatial trend.

Introduction

Statistical Background

U.S. Precipitation, part 1

U.S. Precipitation, part2

Current work

Home Page

Title Page

◀

▶

◀

▶

Page 13 of 17

Go Back

Full Screen

Close

Quit



Introduction

Statistical Background

U.S. Precipitation, part 1

U.S. Precipitation, part2

Current work

Home Page

Title Page



Page 14 of 17

Go Back

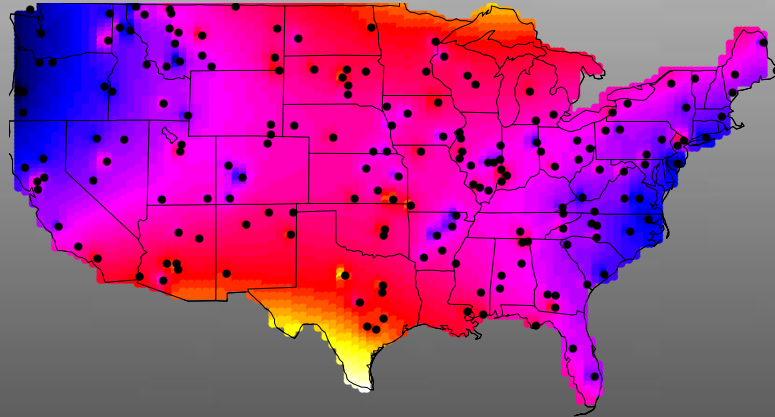
Full Screen

Close

Quit

# Perspective plots for spatial distribution of rainfall extreme trend coefficients

---



National averaged trend:  $.094$  with s.e.  $.007$



Introduction

Statistical Background

U.S. Precipitation, part 1

U.S. Precipitation, part2

Current work

Home Page

Title Page



Page 15 of 17

Go Back

Full Screen

Close

Quit

## 4. U.S. Precipitation, part2

Grady and Smith(2000)

**Data** Use the 187 stations from the HCN data set, except that we use time frame from 1901 to 1999.

**Model for individual stations** Joint Density of (Annual) Sum and Maxima [Note we have an adaptation for the missing data.]

**Trend Analysis** Ultimately, we used an exponential trend in location parameters of sum and maxima.

**Spatial analysis** Smith(1999) spatial model.

*Caution* Comparing “apples: Annual models” and “oranges: Point Process models”.

# Mapped Estimates



Introduction

Statistical Background

U.S. Precipitation, part 1

U.S. Precipitation, part 2

Current work

Home Page

Title Page



Page 16 of 17

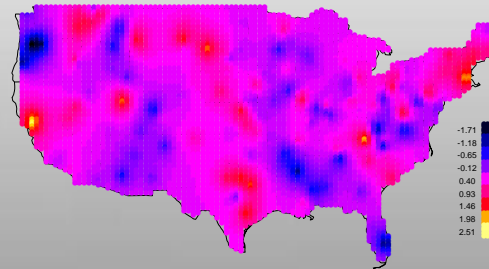
Go Back

Full Screen

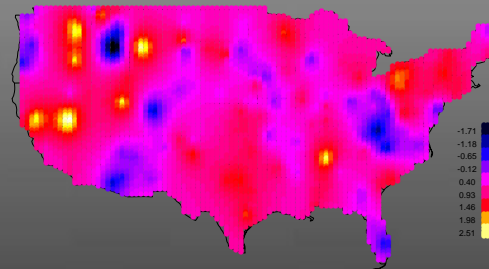
Close

Quit

(a.) Beta in E5



(b.) Gamma in E5



National average:  $.297$  ( $.013$ ) or approximately 3% increase in past century.

(GEV National average:  $.451$  ( $.013$ ) or approximately 4.5% increase in past century.



Introduction

Statistical Background

U.S. Precipitation, part 1

U.S. Precipitation, part2

Current work

Home Page

Title Page



Page 17 of 17

Go Back

Full Screen

Close

Quit

## 5. Current work

Analysis of extreme U.S. rainfall

**Data Description:** Data consists of 5873 stations.

**Improvement:** Denser network will allow for better spatial integration results, particularly important for rainfall data.

**Proposed analysis for each station:** The basic model is the time-varying point process model for extremes, with covariate (thresholds, etc.) considerations.

**Spatial model:** New developments of spatial integration model used in previous rainfall analysis.

**Improvements** to allow for non-stationary models and to expand to denser network (P. Caragea)