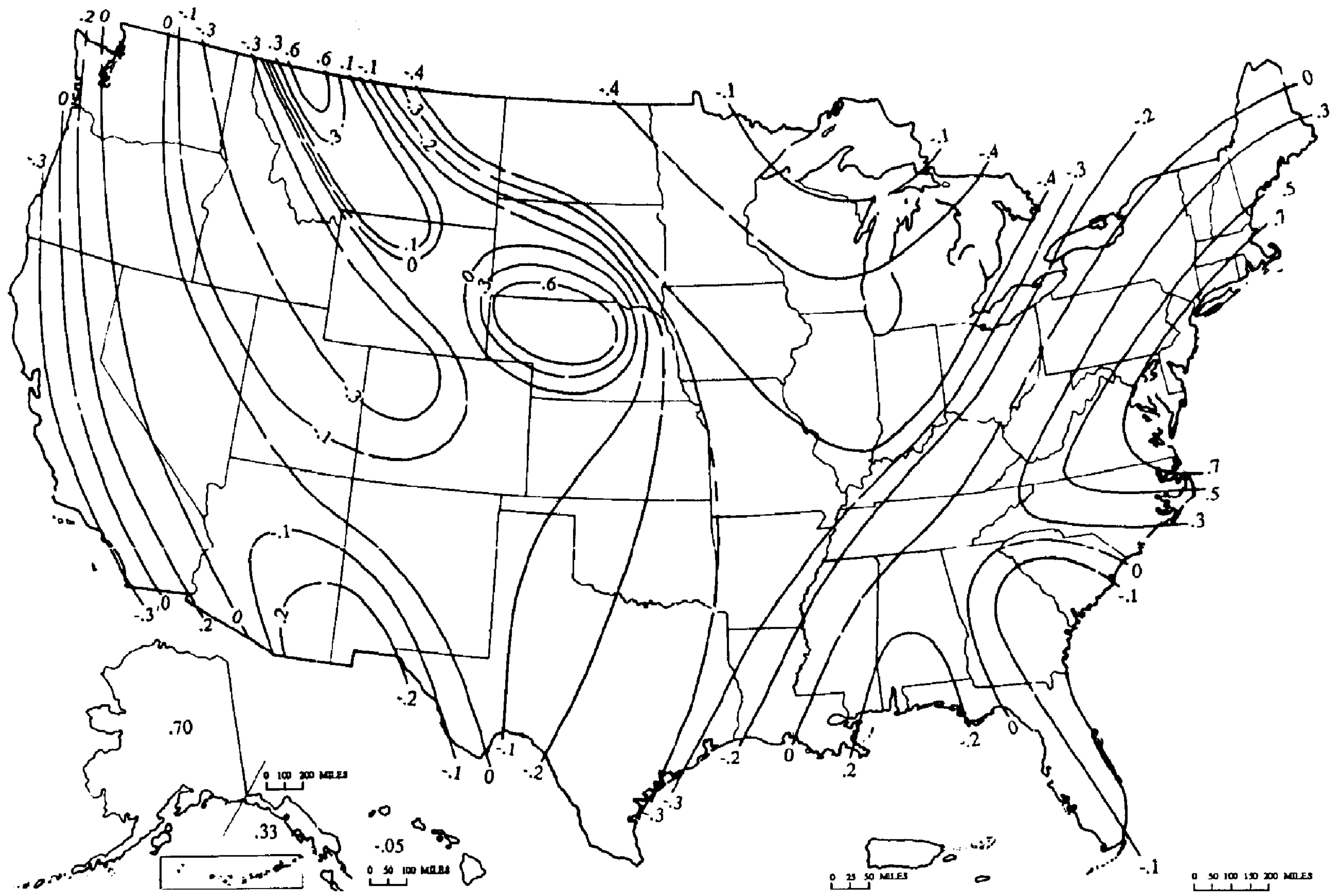
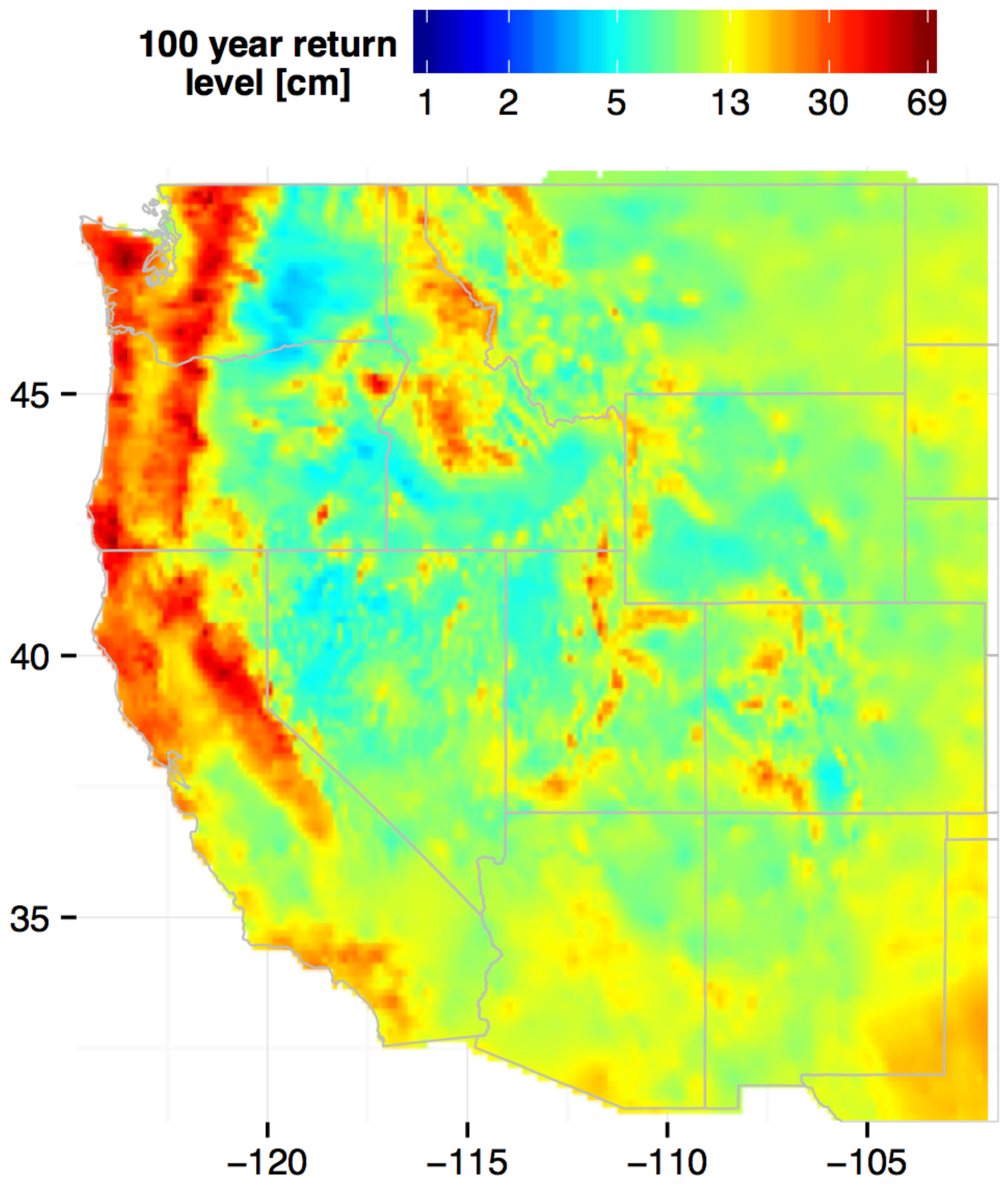
Hydrologic frequency analysis: History and contemporary research

Cameron Bracken
University of Colorado at Boulder





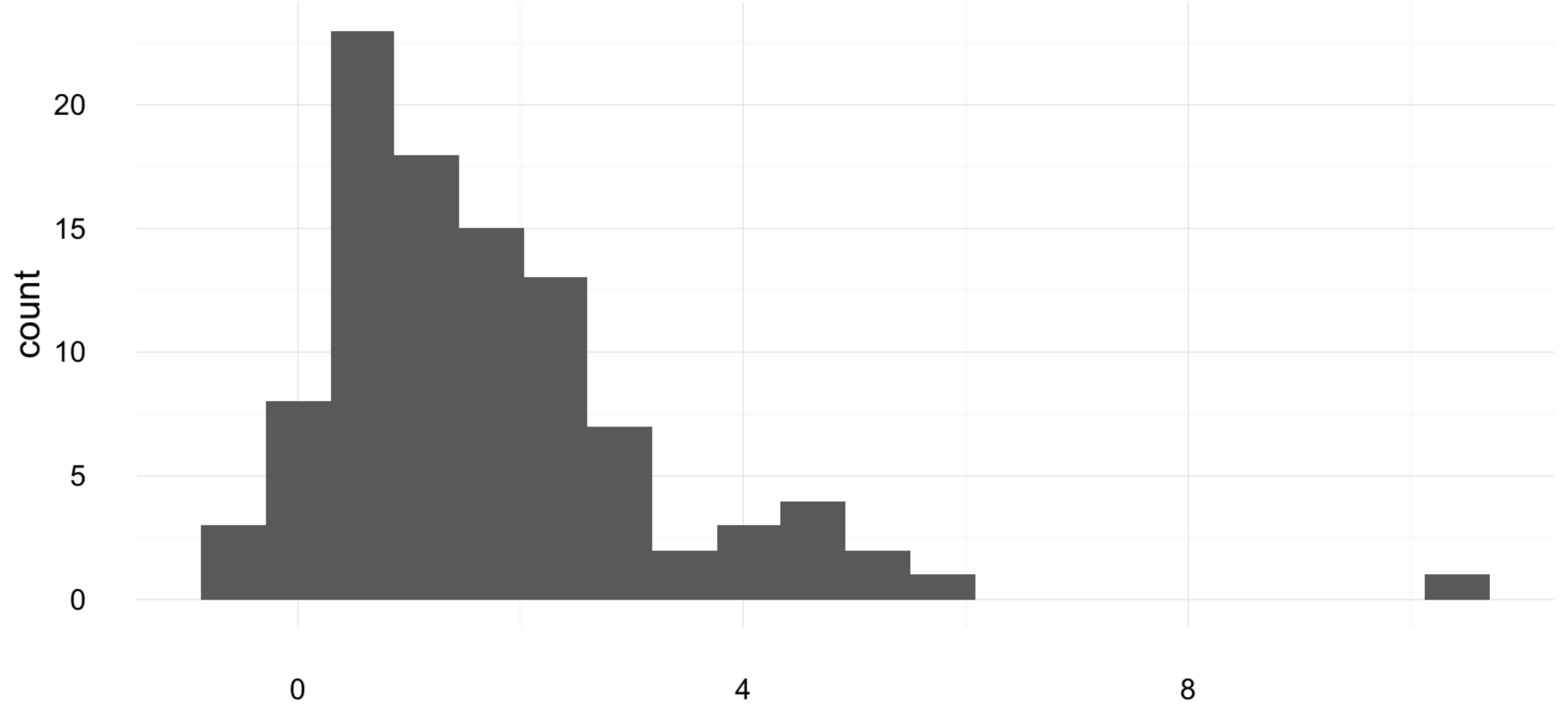
Outline

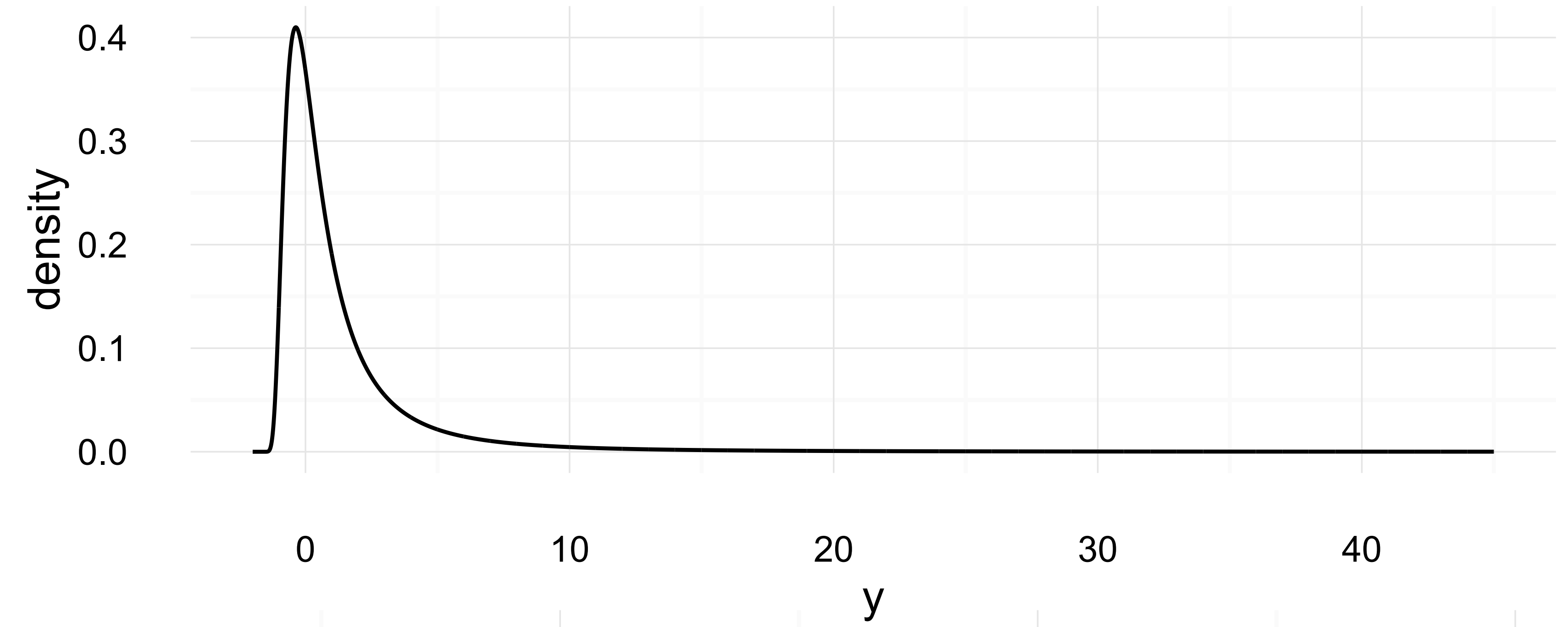
- Intro to frequency analysis
- History
- Federal government guidance
- Contemporary research
- Conclusions

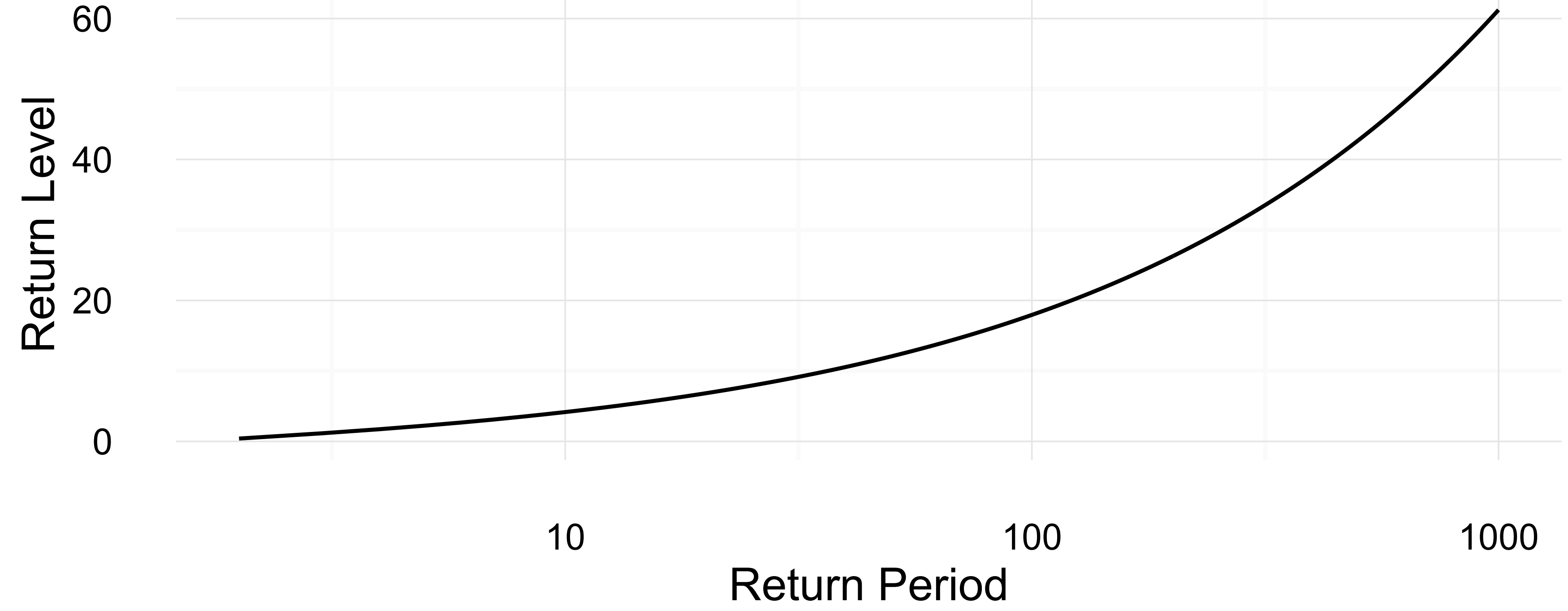
What is hydrologic frequency analysis?

Process of estimating recurrence probabilities of rare hydrologic events (floods, heavy rainfall, etc.).

- Generate extreme data. For example take the maximum daily flow value from each year from a daily flow dataset.
- Fit a probability distribution.
- Compute return levels (quantiles). A 100-year return level will be the (1-1/100)th quantile.







How is frequency analysis used?

- Engineering design and management: “Design flows”
- Probability of occurrence after n years: $p_n=1-(1-p)^n$
- Risk and safety assessments

RECLAMATION
Managing Water in the West

Friant Dam Hydrologic Hazard for Issue Evaluation

Central Valley Project, CA
Mid-Pacific Region

U.S. Department of the Interior
Bureau of Reclamation

September 2013

Extreme data can be developed two ways

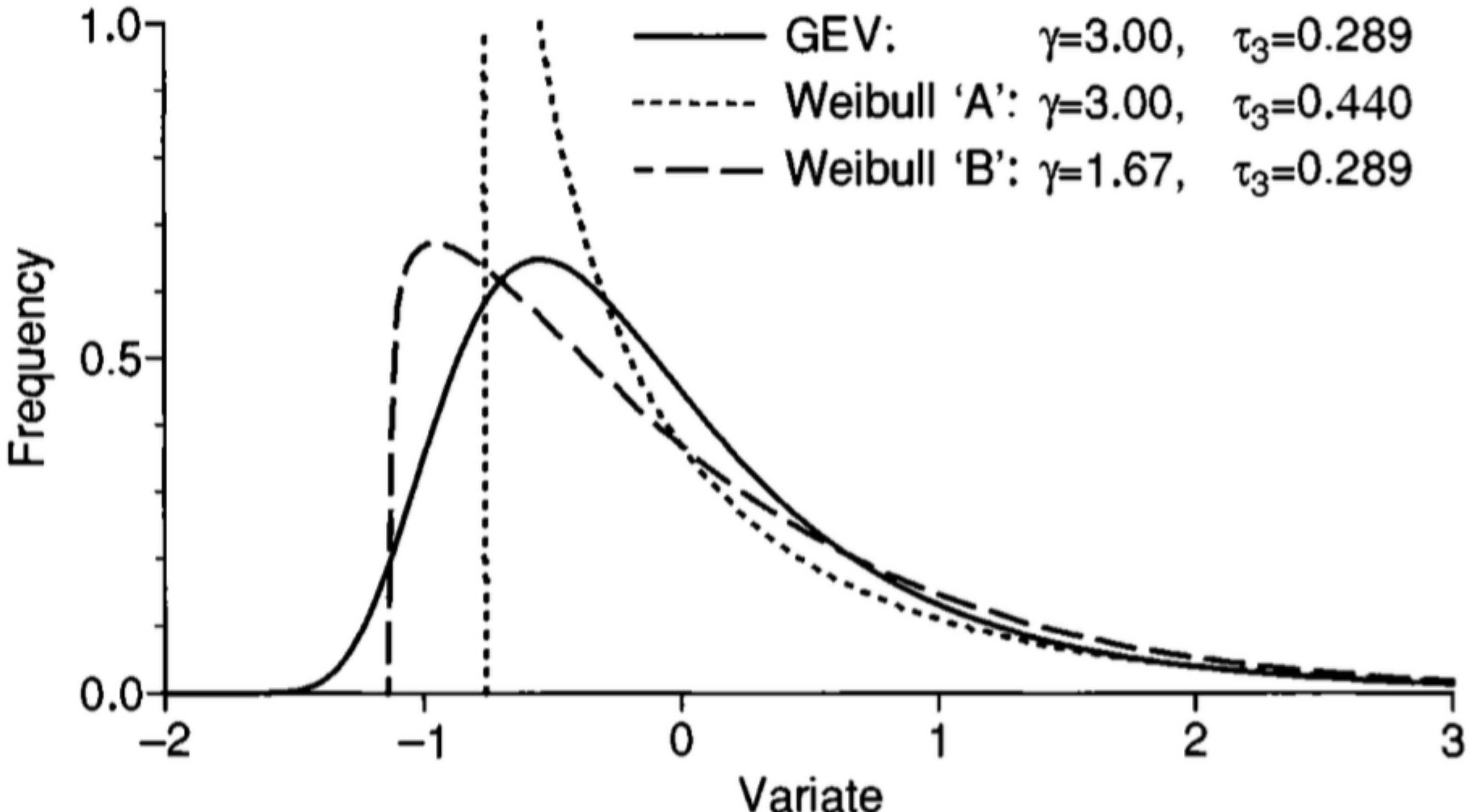
1. Annual or seasonal maximum of daily values

Theoretical Distribution:
Generalized Extreme Value (GEV)

Other options:
Gumbel
Log-Person Type 3 (LP3)
Weibull

2. All daily values above a threshold

Theoretical Distribution:
Generalized Pareto (GPD)



Hosting and Wallace, 1997

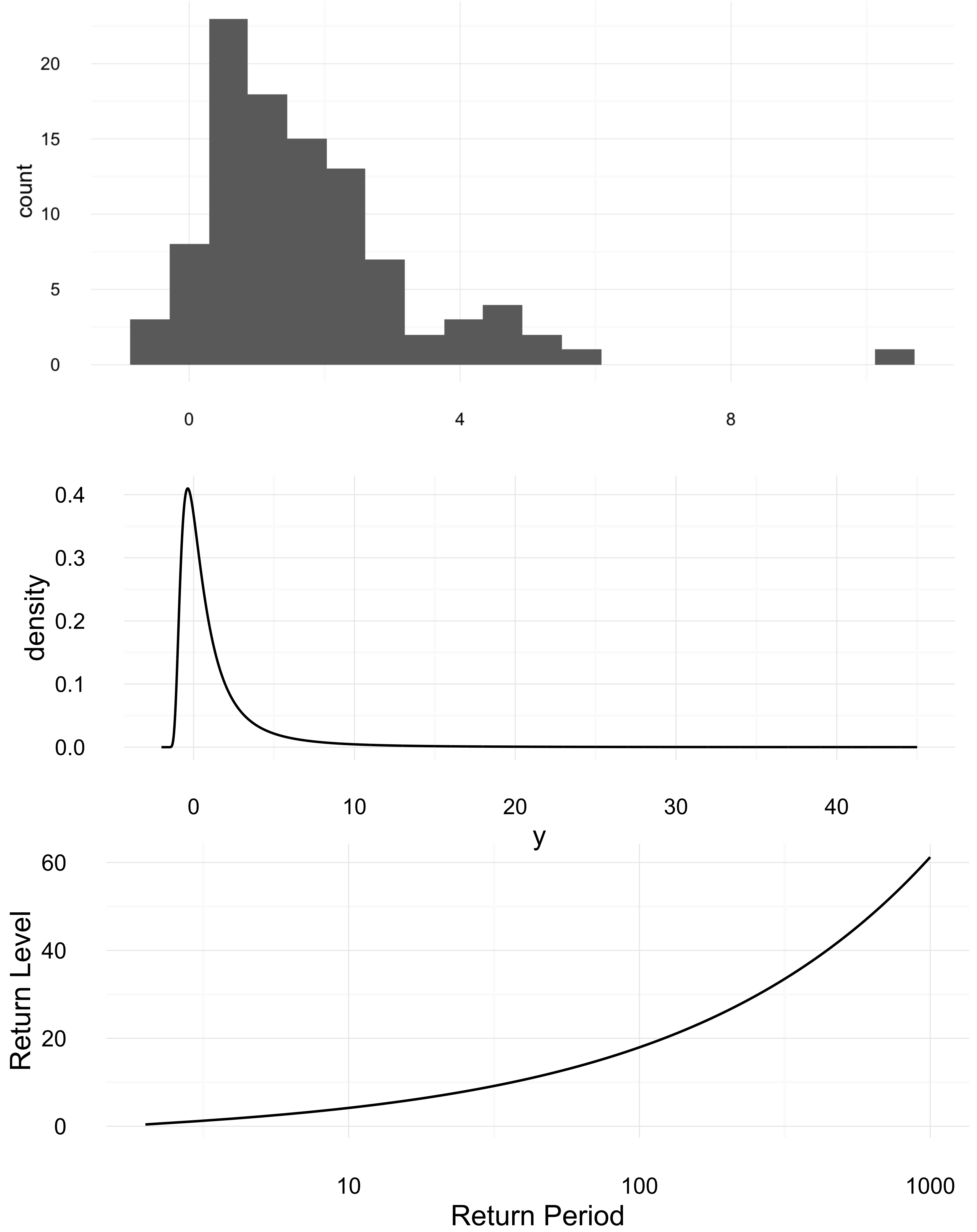
Part 1 - History of frequency analysis

"In order to apply any theory we have to suppose that the data are homogeneous, i.e. that ***no systematical change of climate and no important change in the basin have occurred within the observation period*** *and that no such changes will take place in the period for which extrapolations are made."*

- Gumbel (1941)

1941 - Stationary Single site models

$$y \sim GEV(\mu, \sigma, \xi)$$



Pros:
Very Simple

Cons:
Stationary
Non-spatial

1960 - Index flood method

$$Q_i(F) = \mu_i q(F)$$

Q: Flood return level
F: Flood quantile
μ: Index flood
i: Site
q: Regional flood curve

Assumptions:

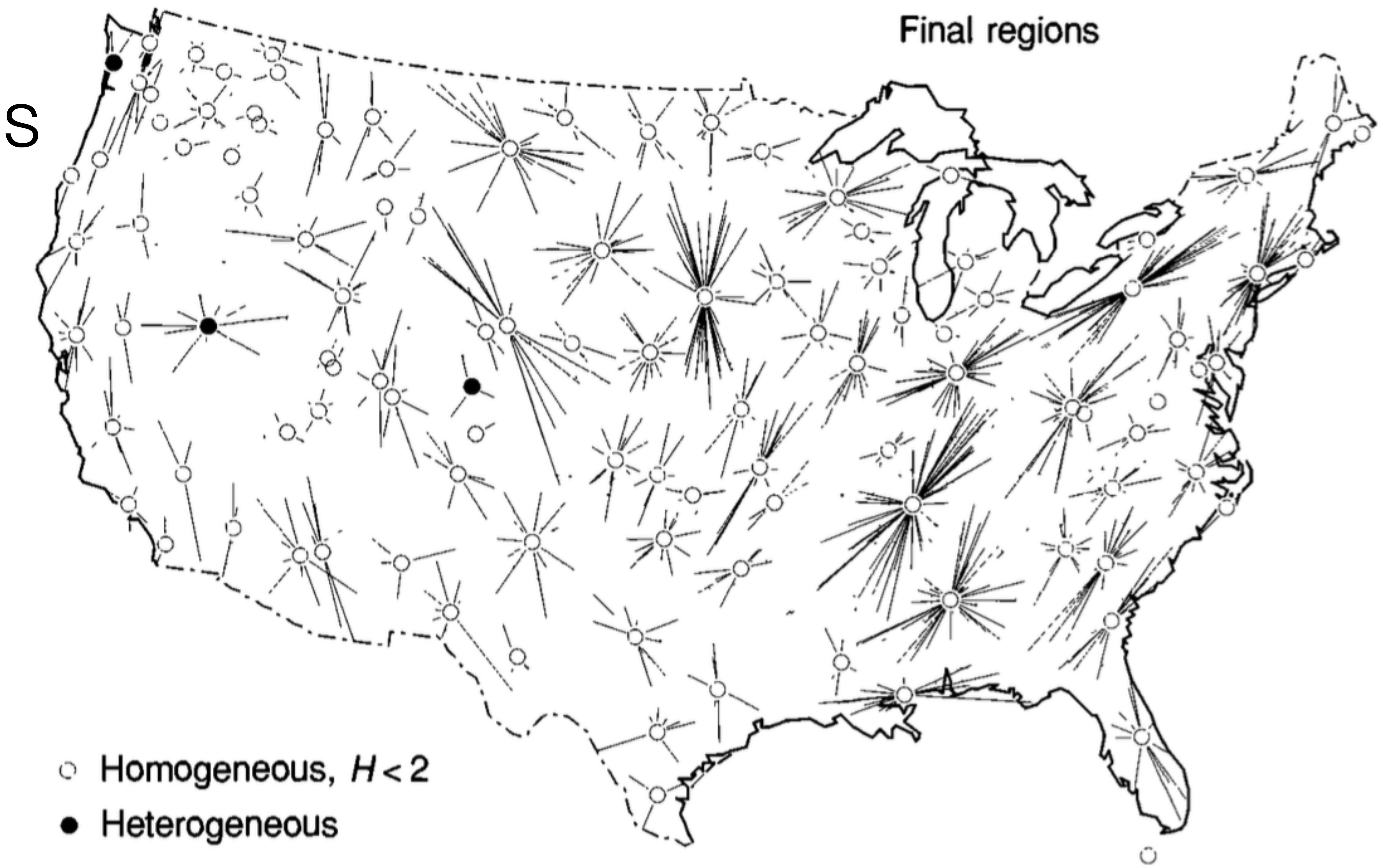
1. Data is iid at each site
2. Data at different sites are independent
3. Frequency distributions at different sites are identical apart from a scale factor
4. The regional distribution is correct

3 and 4 are particularly problematic

Pros: Very Simple **Cons:** Stationary, Non-spatial

1970s–1990s - Regional frequency analysis

General steps:
1. Screening of the data
2. Identification of homogeneous regions
3. Choose a frequency distribution
4. Estimate the distribution parameters
5. (Optional) Estimate frequency at ungaged locations



Cons: Delineation of regions is subjective, stationary, ad-hoc estimation at ungaged locations

Part 2 - Guidance from the Federal Government

- **1967 - 1977**: Bulletin 15 - 17a
- **1981**: Bulletin 17b
- **Dec 2015**: Bulletin 17c (DRAFT)

Guidelines For Determining

Flood Flow Frequency

Bulletin # 17B
of the
Hydrology Subcommittee

Revised September 1981
Editorial Corrections March 1982

INTERAGENCY ADVISORY COMMITTEE
ON WATER DATA

U.S. Department of the Interior
Geological Survey
Office of Water Data Coordination
Reston, Virginia 22092

Bulletin 17b

- "Guidelines for determining flood flow frequency"
- "Result of a continuing effort to develop a coherent set of procedures for accurately defining flood flow potentials"

Climatic Trends

There is much speculation about climatic changes. Available evidence indicates that major changes occur in time scales involving thousands of years. In hydrologic analysis it is conventional to assume flood flows are not affected by climatic trends or cycles. Climatic time invariance was assumed when developing this guide.

Bulletin 17b - Recommended procedure

1. Fit an LP3 distribution to annual max flow

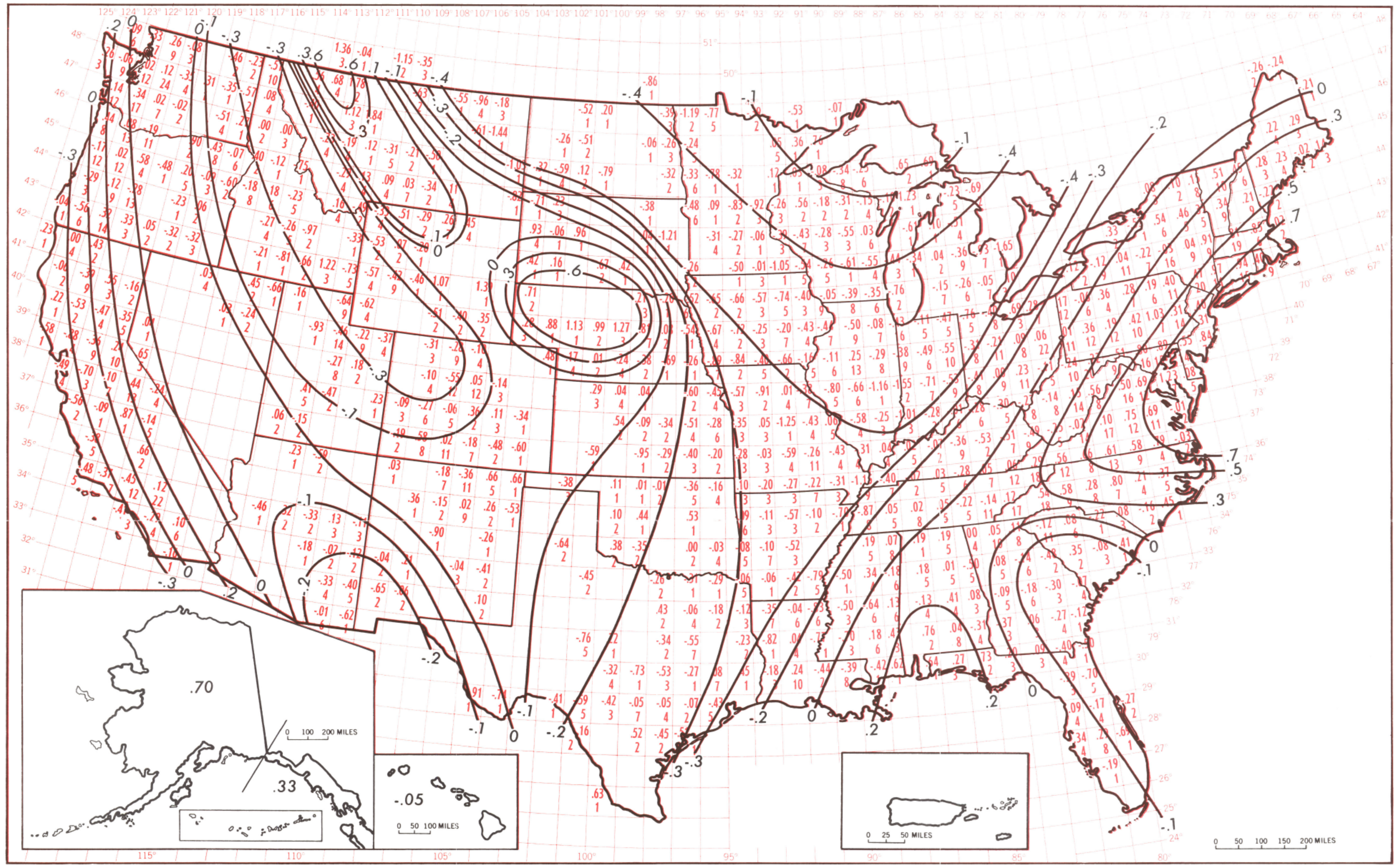
    1.1. Compute mean and standard deviation from the data

    1.2. Compute the regional skew parameter from regional data if available or if you're too lazy look it up from this handy graphic

    1.3. Compute parameters of the LP3 using the mean, sd, and skew

2. Compute flow quantiles

3. Profit



GENERALIZED SKEW COEFFICIENTS OF LOGARITHMS OF ANNUAL MAXIMUM STREAMFLOW
AVERAGE SKEW COEFFICIENT BY ONE DEGREE QUADRANGLES
Lower number in each quadrangle is number of stream gaging stations for which the average shown above it was computed

Pearson Type III Distribution

$$f(x|\tau,\alpha,\beta) = \frac{\left(\frac{x-\tau}{\beta}\right)^{\alpha-1} \exp\left(-\frac{x-\tau}{\beta}\right)}{|\beta|\Gamma(\alpha)}$$

$$x = \log(Q)$$

Why not Bulletin 17b?

- Stationary (no climate influence)
- Single site (basically)
- No general treatment of uncertainty
- Regional skew map may be non-physical

Bulletin 17c

Now in color!

Open for public comment until April 22!

USGS science for a changing world | US Army Corps of Engineers | ACWI Advisory Committee on Water Information

Guidelines for Determining Flood Flow Frequency
Bulletin 17C

Techniques and Methods 4–BXX

U.S. Department of the Interior
U.S. Geological Survey

DRAFT: December 29, 2015

---PROVISIONAL---
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PUBLIC COMMENT.
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

Bulletin 17c - What's New?

- Adoption of a generalized representation of flood data that allows for interval and censored data types
- A new method, called the Expected Moments Algorithm (EMA) that extends the method-of-moments so that it can accommodate interval data
- A generalized approach to identification of low outliers in flood data
- An improved method for computing confidence intervals.
- Bayesian estimation of regional skew parameter

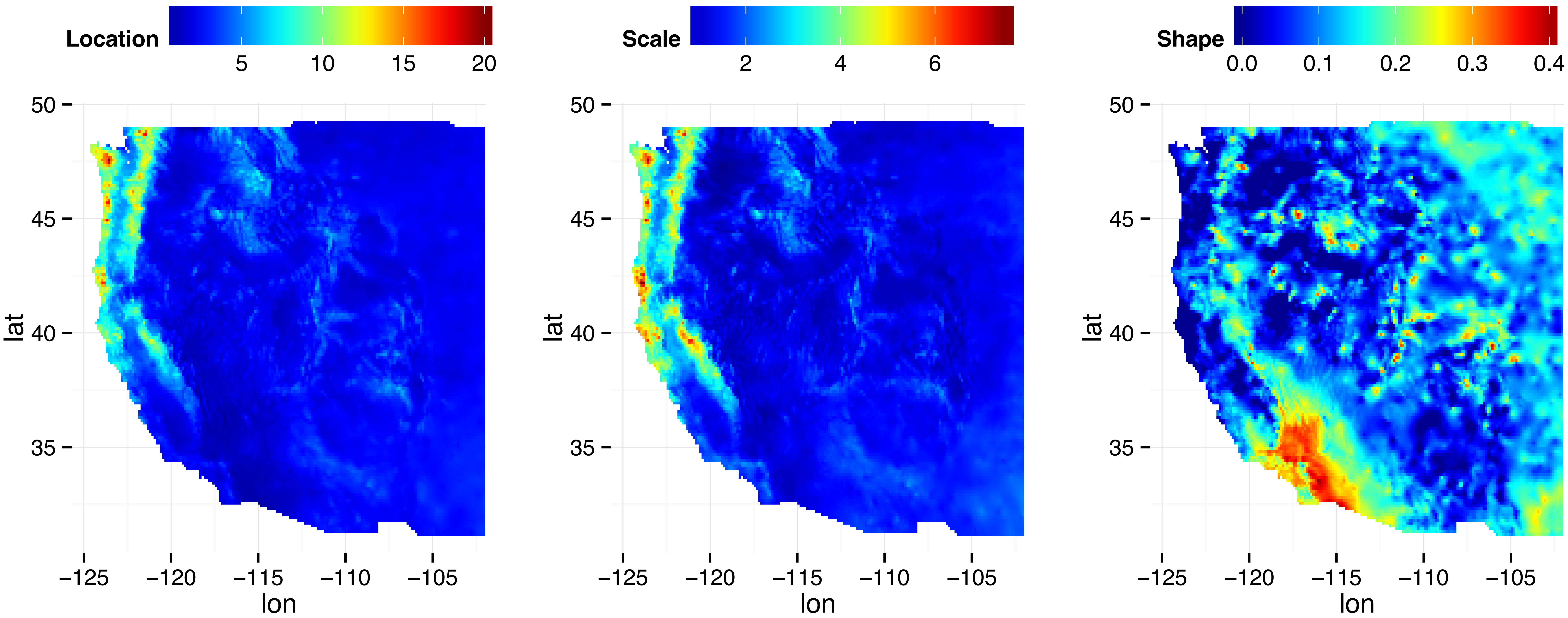
Bulletin 17c - What's the same?

- Still fit LP3
- Still use regional estimates of skew when possible

Bulletin 17c - What's still not addressed?

- Robust uncertainty quantification
- Nonstationarity and nonstationary risk
  - "*There is much concern about changes in flood risk associated with climate variability and long-term climate change.* ***Time invariance was assumed in the development of this guide.*** *…where there is sufficient scientific evidence … climate variability or change … should be incorporated in flood frequency analysis…*"
- Mixed distributions
- Ungaged locations
- Incorporation of precipitation data and other physical information
- Heavily altered watersheds

Part 3 - Modern frequency analysis



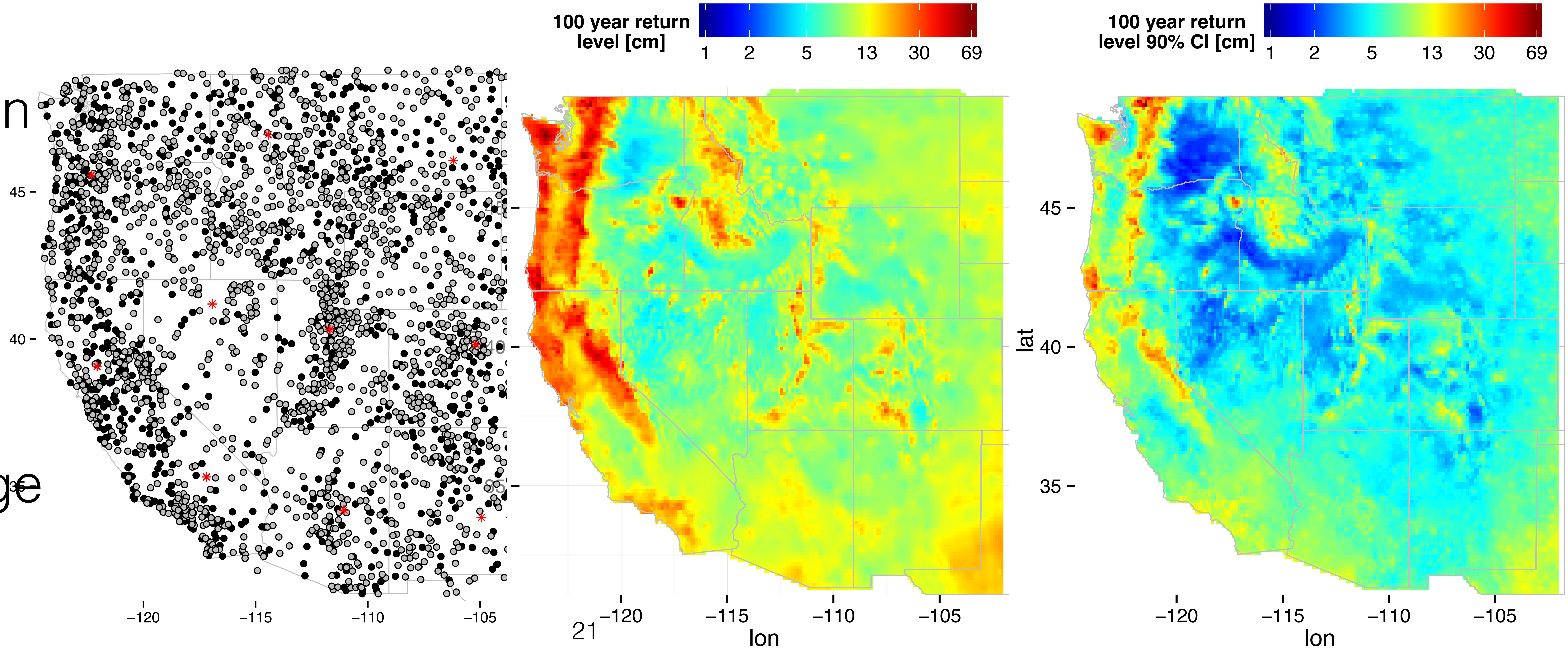
Recent improvements

- **Nonstationary:** Incorporating climate information
- **Spatial:** Allowing for estimation at ungaged locations and on grids
- **Multivariate:** Dependence between multiple variables
- **Uncertainty quantification:** Bayesian hierarchical models

Spatial frequency analysis

- Thousands of observation locations
- Gridded return levels
- Gridded uncertainty
- Arbitrarily large regions

$$y(s,t) \sim \mathrm{GEV}(\mu(s), \sigma(s), \xi(s))$$



Nonstationary frequency analysis

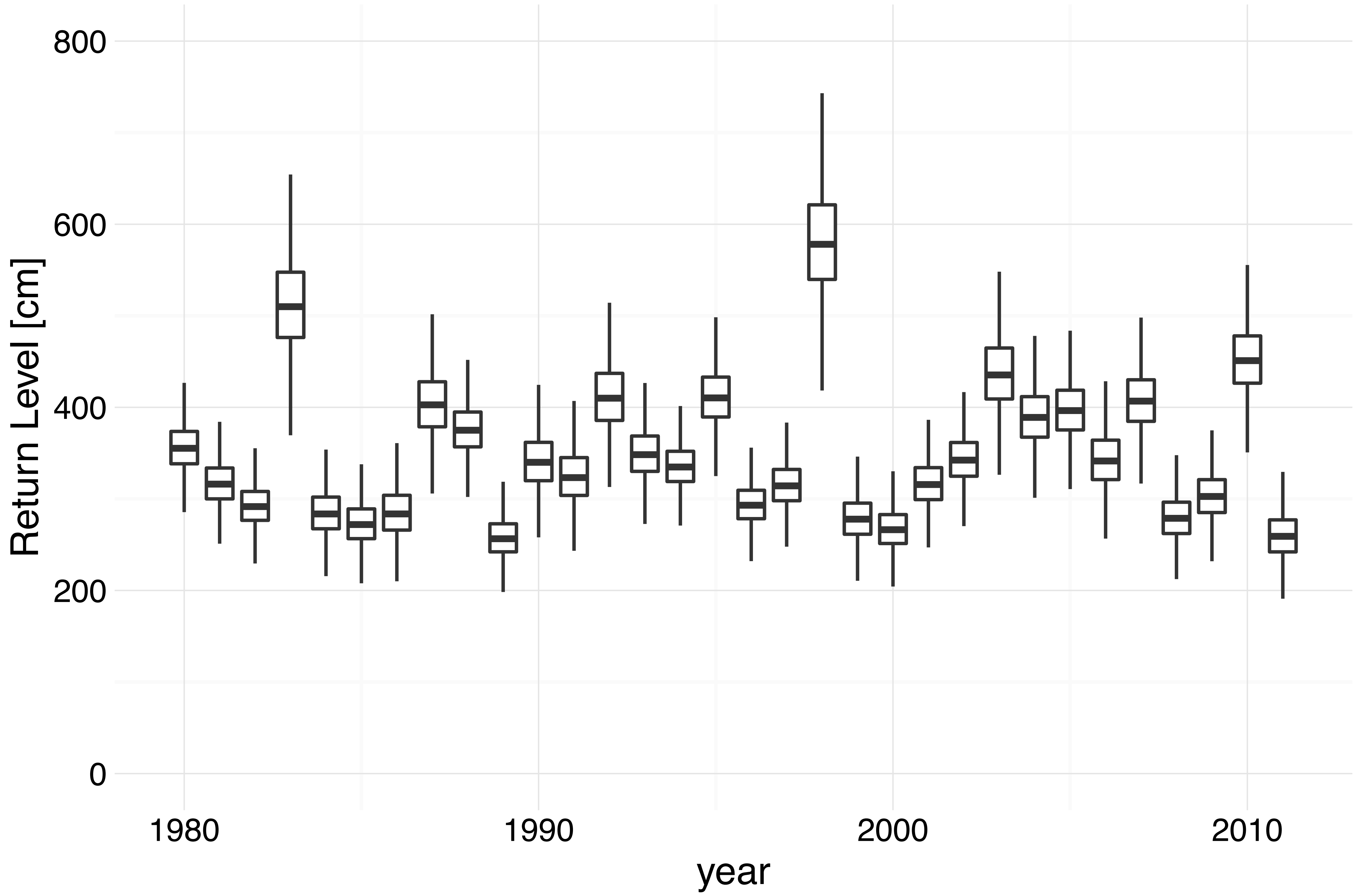
- Incorporate large scale climate information (e.g. ENSO)

$$y(s,t) \sim \mathrm{GEV}(\mu(t), \sigma(t), \xi(t))$$

- Nonstationary return levels

- Potential for forecasting, projections

- Robust uncertainty quantification



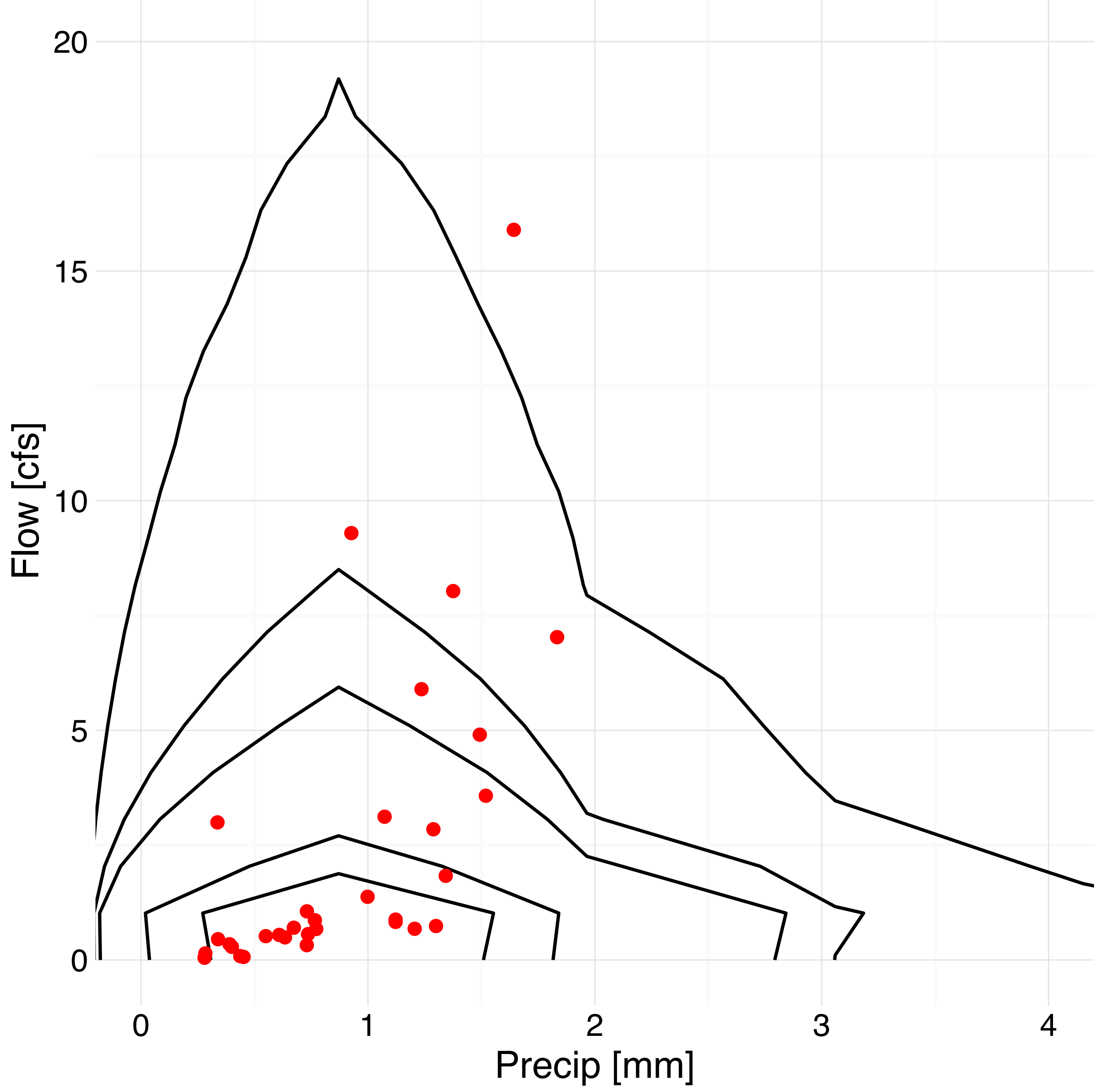
Multivariate frequency analysis

$$(y_1(s,t), y_2(s,t)) \sim \mathrm{C}(\Sigma | \{\mu_1(t), \sigma_1(t), \xi_1(t), \mu_2(t), \sigma_2(t), \xi_2(t)\})$$

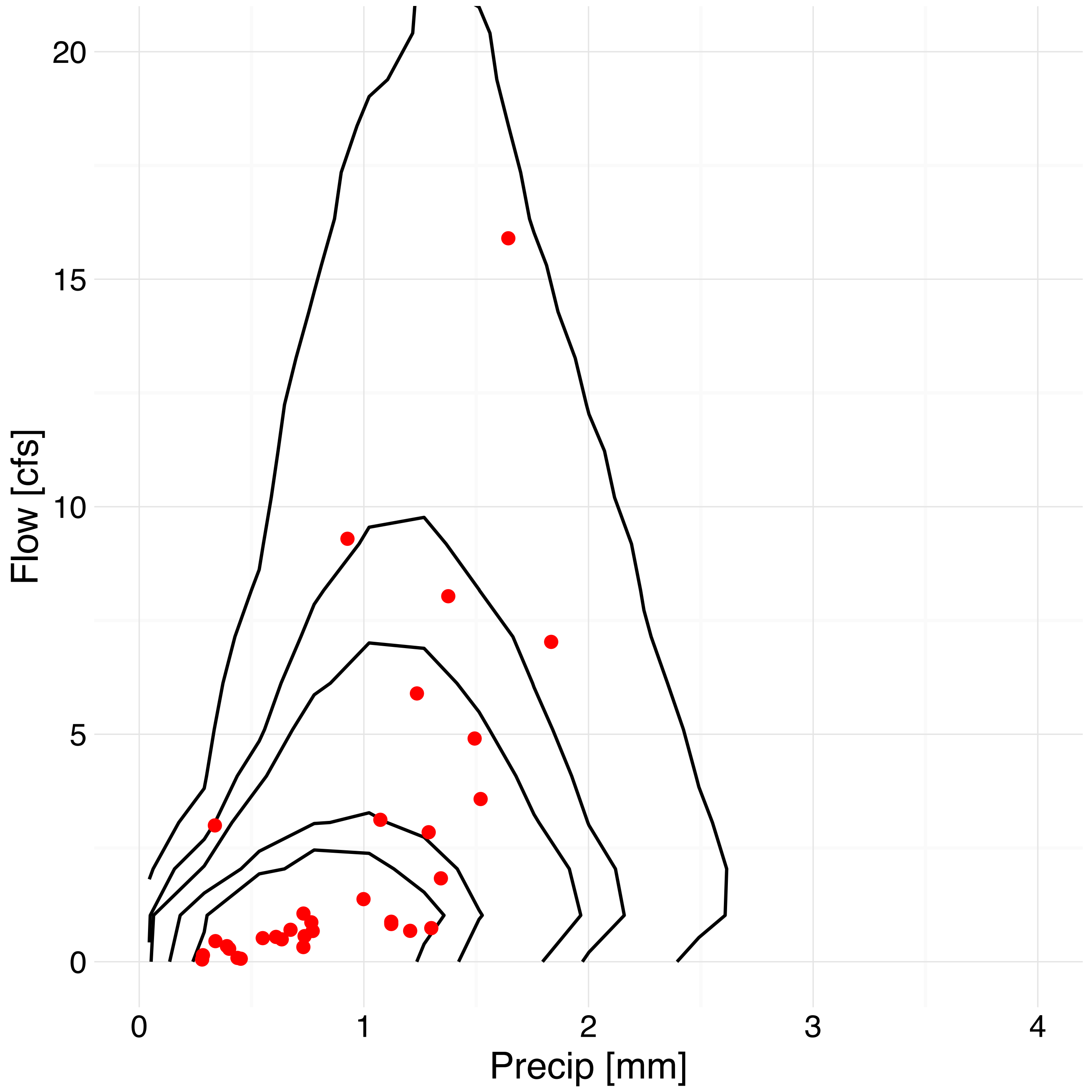
$$y_1(s,t) \sim \mathrm{GEV}(\mu_1(t), \sigma_1(t), \xi_1(t))$$

$$y_2(s,t) \sim \mathrm{GEV}(\mu_2(t), \sigma_2(t), \xi_2(t))$$

- Multivariate dependence
- Robust uncertainty quantification



Independent



Dependent

Conclusions

- History of frequency analysis 1940s - 1990s
  - Single site and regional
- Federal government guidance Bulletin 15 - 17b, 1960-1981
  - Go comment on the new Bulletin 17c!
- Recent research
    - Spatial, nonstationary and multivariate

The End!