RECLAMATION

Managing Water in the West

Research to Develop Guidance on Extreme Precipitation Estimates in Orographic Regions

Andrew Verdin and Kathleen Holman

Flood Hydrology and Meteorology Group

Technical Service Center

U.S. Department of the Interior
Bureau of Reclamation

Outline

1. Motivation
2. Precipitation-Frequency analysis
   A. Frequency Analysis Methods
   B. Datasets
3. Case study in Tennessee River Valley
   A. L-moments
   B. Bayesian
4. Summary & Conclusions

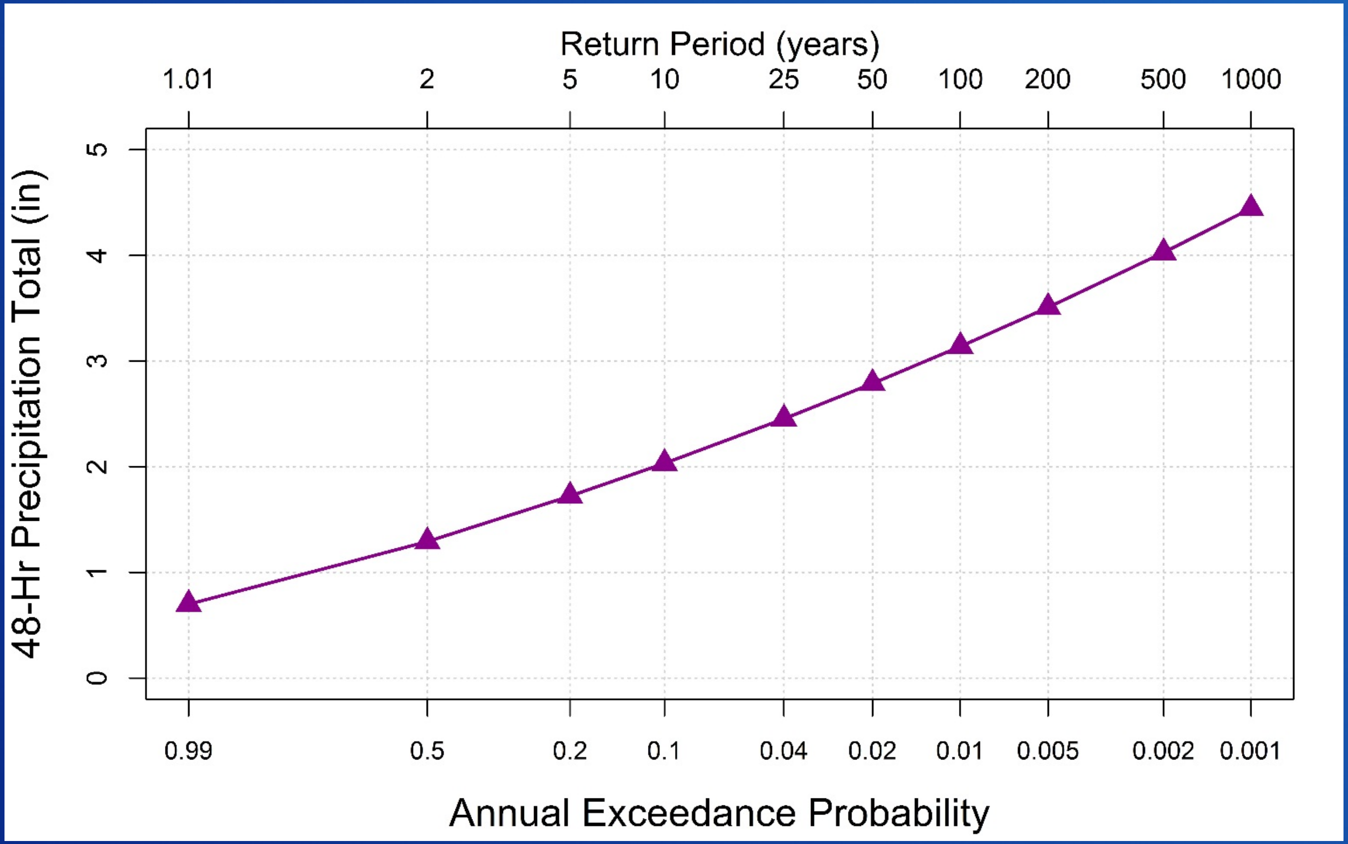
RECLAMATION

Motivation

RECLAMATION

Motivation

Q: What is a precipitation-frequency relationship?

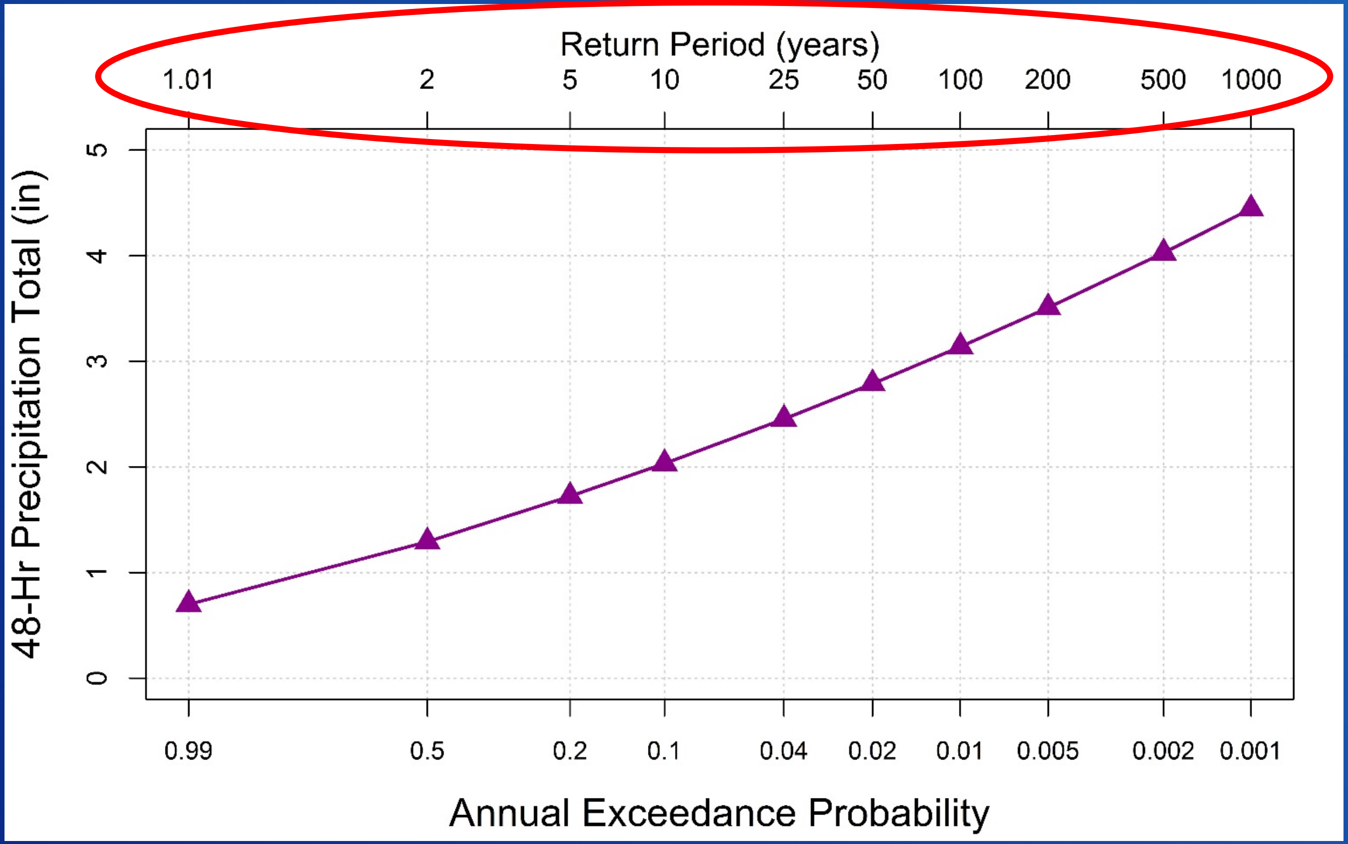


A: Statistical relationship relating precipitation depth to the probability of exceeding that depth.

RECLAMATION

Motivation

Probability reported as "return period" or "average recurrence interval"



Return period of: 100 years means the probability is 1-in-100 (0.01)

500 years means the probability is 1-in-500 (0.002)

1,000 years means the probability is 1-in-1000 (0.001)

Precipitation-Frequency Analysis

Define relevant precipitation duration

→ e.g., 1-day, 2-day, etc.

Extract annual/seasonal maxima from daily time series

QC annual/seasonal maxima for false maxima

Fit extreme value distribution to maxima

→ Estimate $\theta = (\mu, \sigma, \xi)$

Calculate quantiles of distribution

→ Precipitation magnitudes and associated probabilities

RECLAMATION

Regional Frequency Analyses

Assume observations within homogeneous region (HR) described by single distribution

→ Pool all annual/seasonal maxima within HR

Scale the annual/seasonal maxima by the at-site mean of the maxima

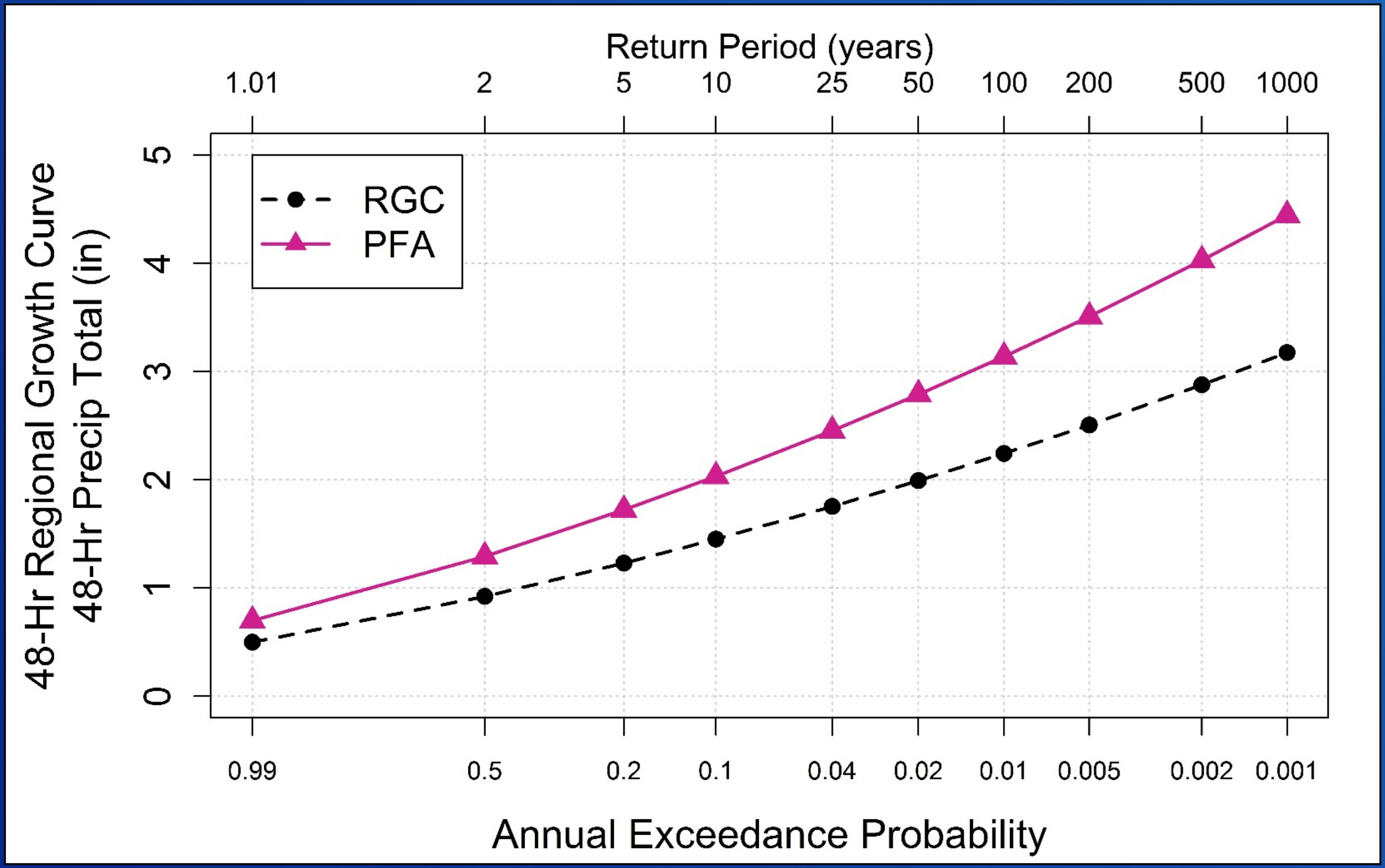
→ Mean at each site -> 1

Compute precipitation-frequency relationship, produce regional growth curve (RGC)

→ Scale by specific at-site mean for point estimates

Regional Growth Curve

Unitless curve describing all gauges in HR



Scale by site-specific ASM for site-specific PF curve

RECLAMATION

Frequency Analysis Methods

RECLAMATION

L-moments

RECLAMATION

L-Moments

1. Developed for ***regional*** frequency analysis
2. Identify weather stations (sites) within HR
3. Screen observations
   - annual/seasonal/monthly maxima
   - duration depends on meteorology
4. Quality control data
5. Compute L-statistics* for each site
   - L-mean, L-scale, L-skewness, L-kurtosis
6. Test for heterogeneity
   - Discordancy measures (e.g., $D_i \leq 3$)
7. Identify the "best" distribution
   - GEV, GPD, GNO, GLO, PE3, Wakeby
8. Calculate regional growth curve (weighted by POR)
   - Scale growth curve (point, basin, region)

Available R packages for L-moments:
library("lmom")
library("lmomRFA")

L-statistics

Alternative system of describing probability distribution functions based on linear combinations of moments

<u>L-moments:</u>

λ_1=L-location (mean)

λ_2=L-scale (variability or dispersion)

λ_3=L-skewness (asymmetry)

λ_4=L-kurtosis (thickness of tail)

<u>L-moment ratios (dimensionless):</u>

$T_r=\lambda_r/\lambda_2$

T=L-CV= λ_2/λ_1 (variability)

Hosking and Wallis (1997)

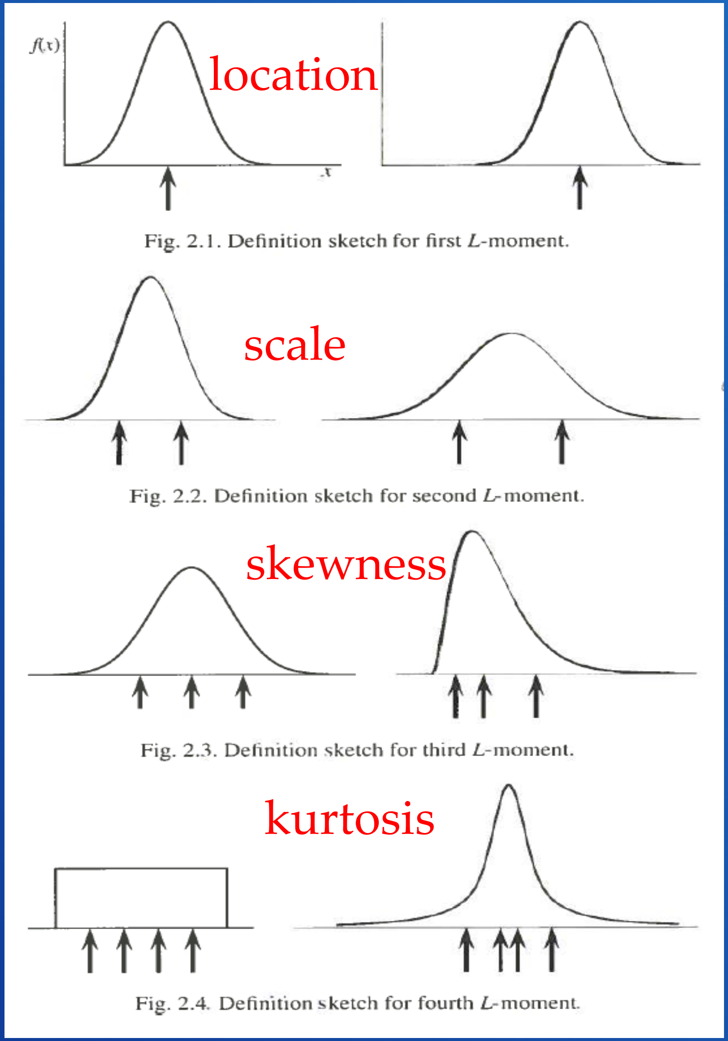


Fig. 2.1. Definition sketch for first *L*-moment.

Fig. 2.2. Definition sketch for second *L*-moment.

Fig. 2.3. Definition sketch for third *L*-moment.

Fig. 2.4. Definition sketch for fourth *L*-moment.

Available R packages for L-moments:
library("lmom")
library("lmomRFA")

Bayesian inference

Bayesian inference

Prior $p(\boldsymbol{\theta})$: the strength of our belief in θ without the data Y

Posterior $p(\boldsymbol{\theta}|\boldsymbol{Y})$: the strength of our belief in θ when the data Y are taken into account

Likelihood $p(\boldsymbol{Y}|\boldsymbol{\theta})$: the probability that the data Y could have been generated by the model with parameter values θ

Evidence p(Y): the probability of the data according to the model, determined by summing across all possible parameter values weighted by the strength of belief in those parameter values

→ typically unknown, can be ignored with proportionality

→ essentially a normalizing constant

→ does not enter into determining relative probabilities (models)

Available R packages for Bayesian inference:
library("rstan")
library("spBayes")

Bayesian inference

Bayes' Rule in a modeling framework:

$$p(\theta|\boldsymbol{Y}) = \frac{p(\boldsymbol{Y}|\theta)p(\theta)}{p(\boldsymbol{Y})} \propto p(\boldsymbol{Y}|\theta)p(\theta)$$

e.g., $\boldsymbol{Y} = (y_1, y_2, \dots, y_n); \theta = (\mu, \sigma, \xi)$

- Define *prior distributions* for model parameters θ (a priori knowledge)
- Can consider numerous likelihood functions (e.g., *GEV*, GNO, GLO, etc.)
- Monte Carlo, acceptance criteria, builds *posterior distributions* of θ

Bayesian inference derives the *posterior probability* as a consequence of a *prior probability* and a *likelihood function*

RECI

Available R packages for Bayesian inference:
library("rstan")
library("spBayes")

Regional Bayesian

Scale annual maxima by at-site mean

Assume scaled maxima within HR described by a single theoretical distribution

Generalized Extreme Value (GEV) distribution

Posterior distributions of $\theta = (\mu, \sigma, \xi)$

→ Quantification of one source of epistemic uncertainty

RECI

Available R packages for Bayesian inference:
library("rstan")
library("spBayes")

Datasets

Historical Observations

Global Historical Climatology Network:

Integrated database of daily climate summaries from land surface stations (100,000+) across the globe

Includes observations from multiple sources that have been subjected to a the same fully-automated quality control process (Durre et al. 2010)

- Duplication of records
- Exceedance of physical, absolute, climatological limits
- Temporal persistence
- Inconsistencies with neighboring observations

RECLAMATION

Historical Observations

Global Historical Climatology Network:

Integrated database of daily climate summaries from land surface stations (100,000+) across the globe

Includes observations from multiple sources that have been subjected to a the same fully-automated quality control process (Durre et al. 2010)

- Duplication of records
- Exceedance of physical, absolute, climatological limits
- Temporal persistence
- Inconsistencies with neighboring observations

Question:

How do we obtain PF estimates at ungauged locations?

RECLAMATION

Newman et al. (2015)

Gridded observation-based ensemble dataset of daily precipitation and temperature from 1980 to 2012

Ensemble generation:

*Locally-weighted regression models are used to produce "best estimate" values of precipitation and temperature at 1/8° lat×lon grid

* Regression residuals are used to perturb the best-estimate values with correlated random samples

Resulting dataset:

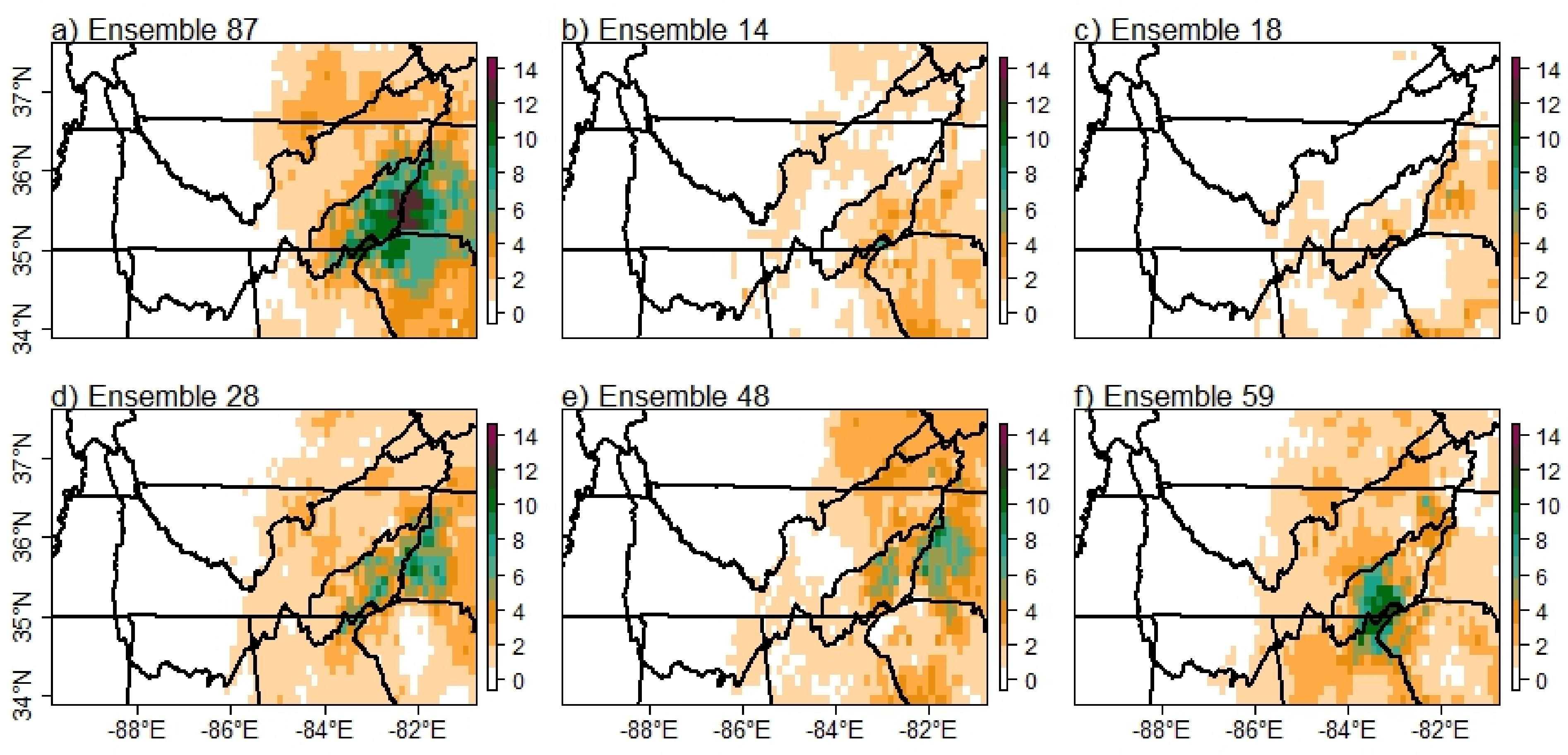
*100 plausible precipitation and temperature grids

*Each valid over same period

RECLAMATION

Newman et al. (2015)

Daily precipitation (inches) on September 15, 2004

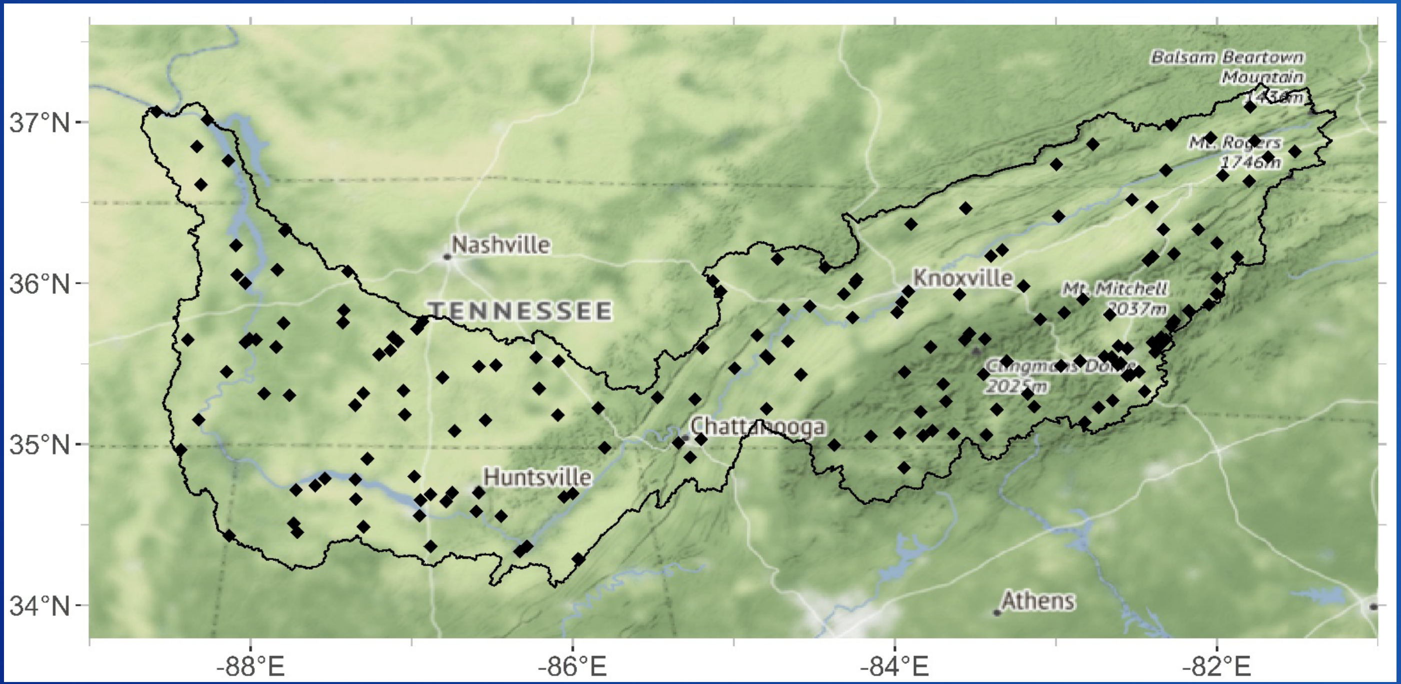


RECLAMATION

Case Study

Study Region

Tennessee River Valley watershed



GHCN-Daily gauges with 85% data availability for 10+ years period of record (POR)

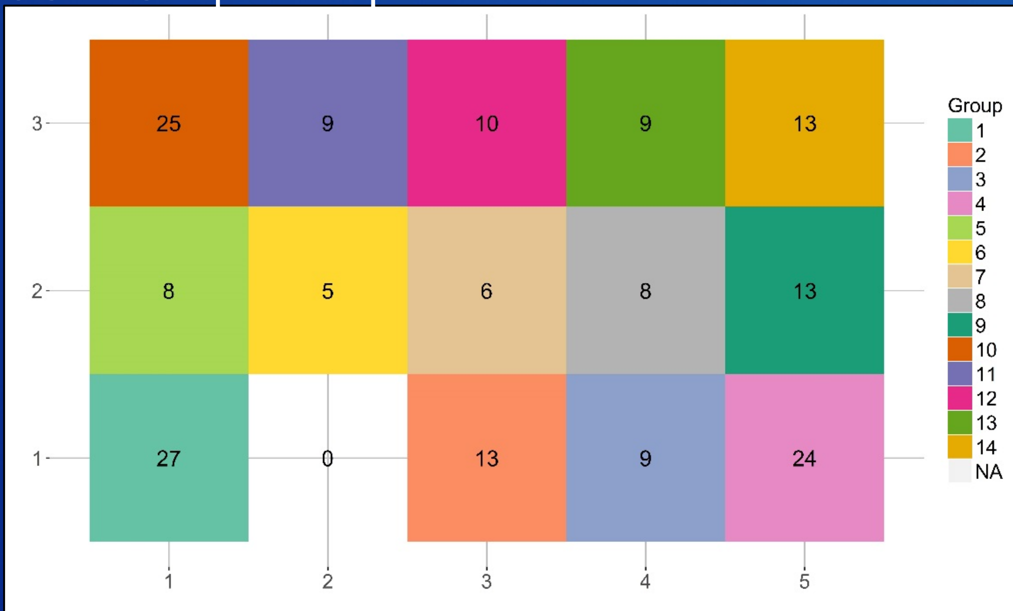
RECLAMATION

Homogeneous Regions

- Methods to define HR (Hosking and Wallis 1997)
  - Subjective methods
    - Geographical location
    - Seasonal timing of peak events
    - Mean annual precipitation (MAP)
    - Similar forcing mechanisms (synoptics)
  - Objective methods
    - Self-Organizing Maps (SOM)
    - Hierarchical clustering analysis (HCA)
    - Principle component analysis (PCA)
    - Heterogeneity measure

Self-Organizing Map

SOM Output Map



Clustering algorithm used to "group" stations with similar attributes

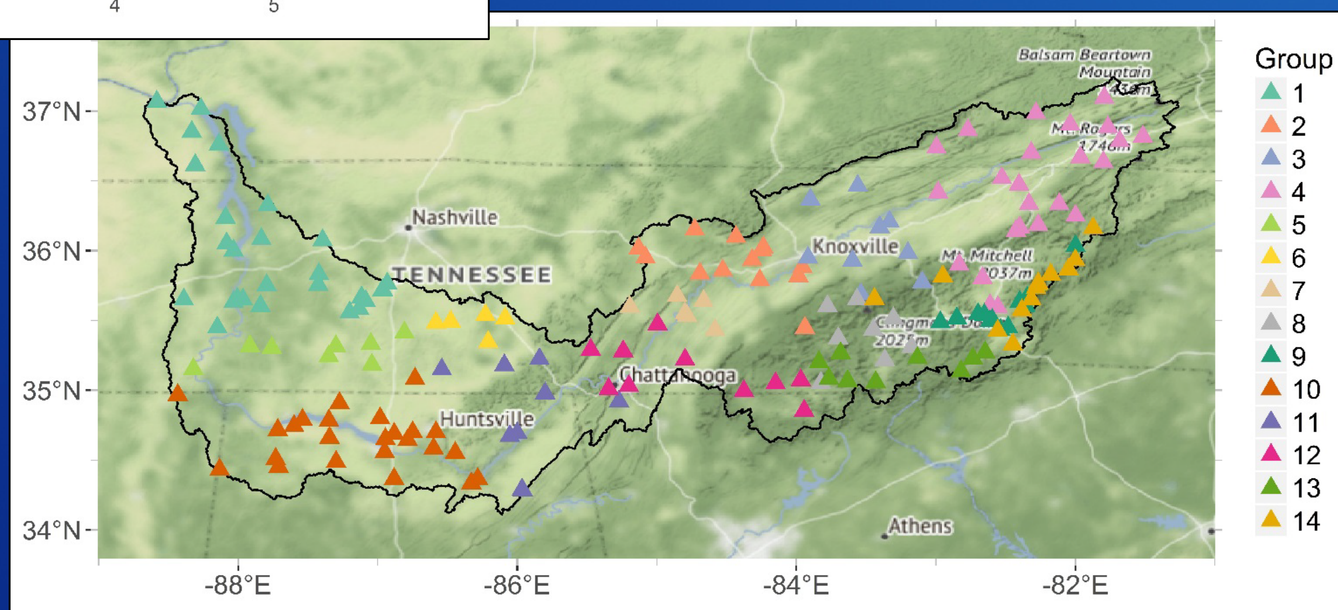
Apply SOM algorithm to :

- Latitude
- Longitude
- Elevation
- *Avg annual precipitation*
- *Avg annual max one-day precipitation*

Each station maps to a single SOM node

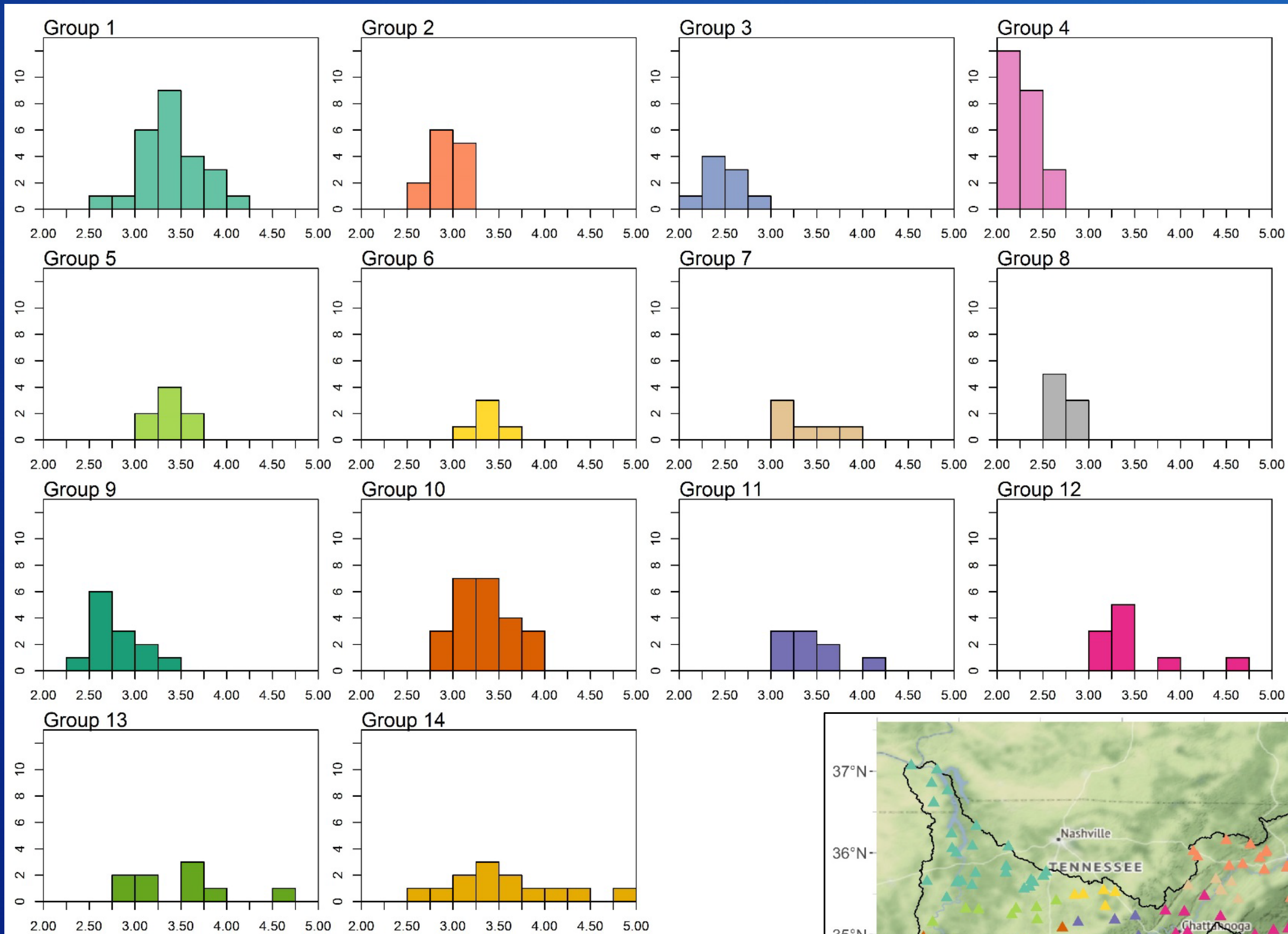
Gauges mapped to same node define homogeneous regions

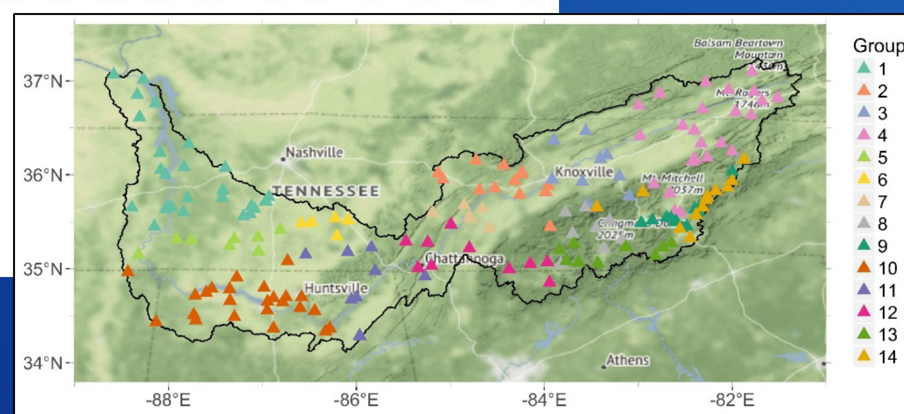
Homogeneous regions need not be contiguous



Available R packages for SOM analysis:
library("som")
library("kohonen")

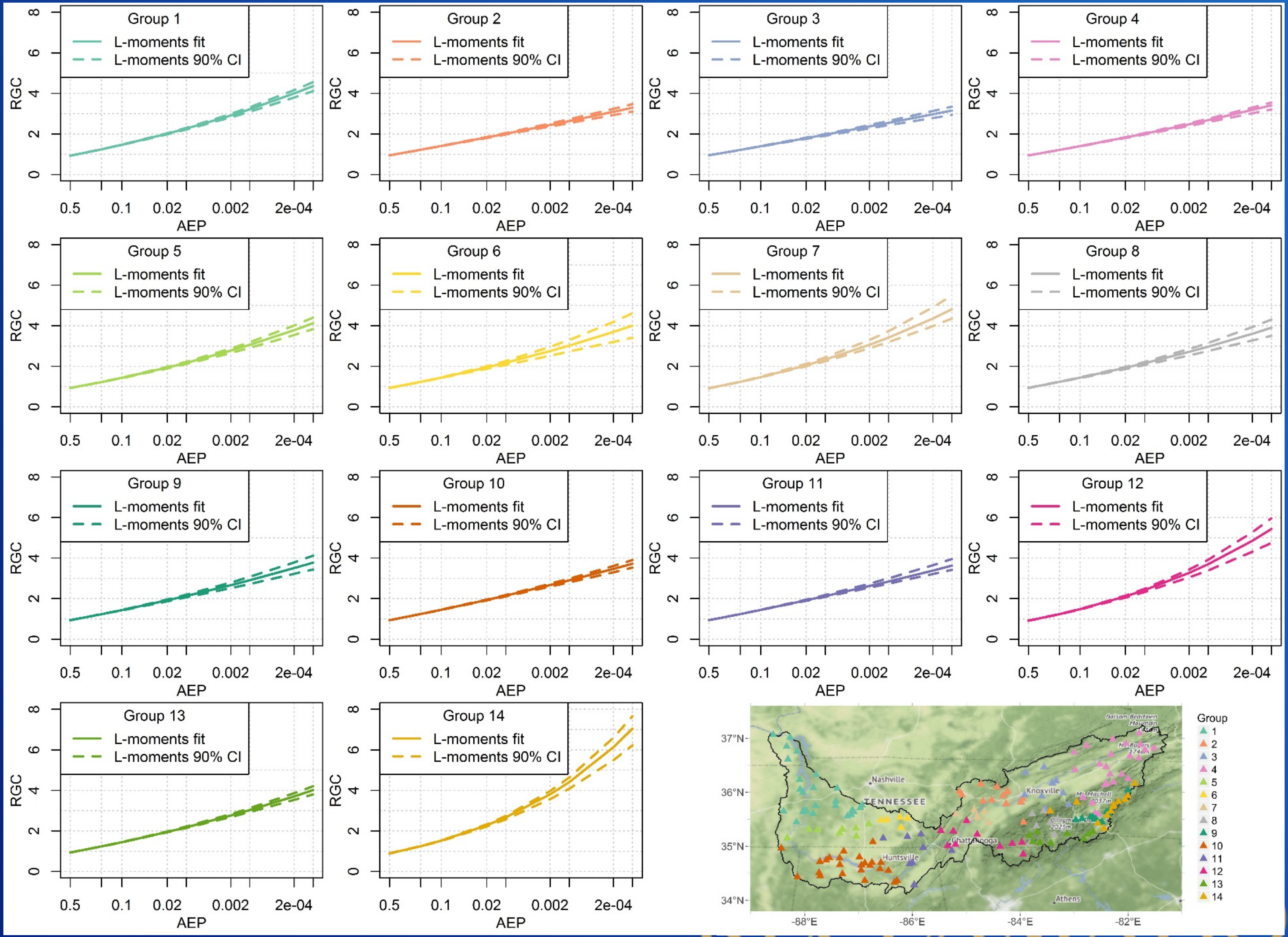
SOM Results – At-Site Means



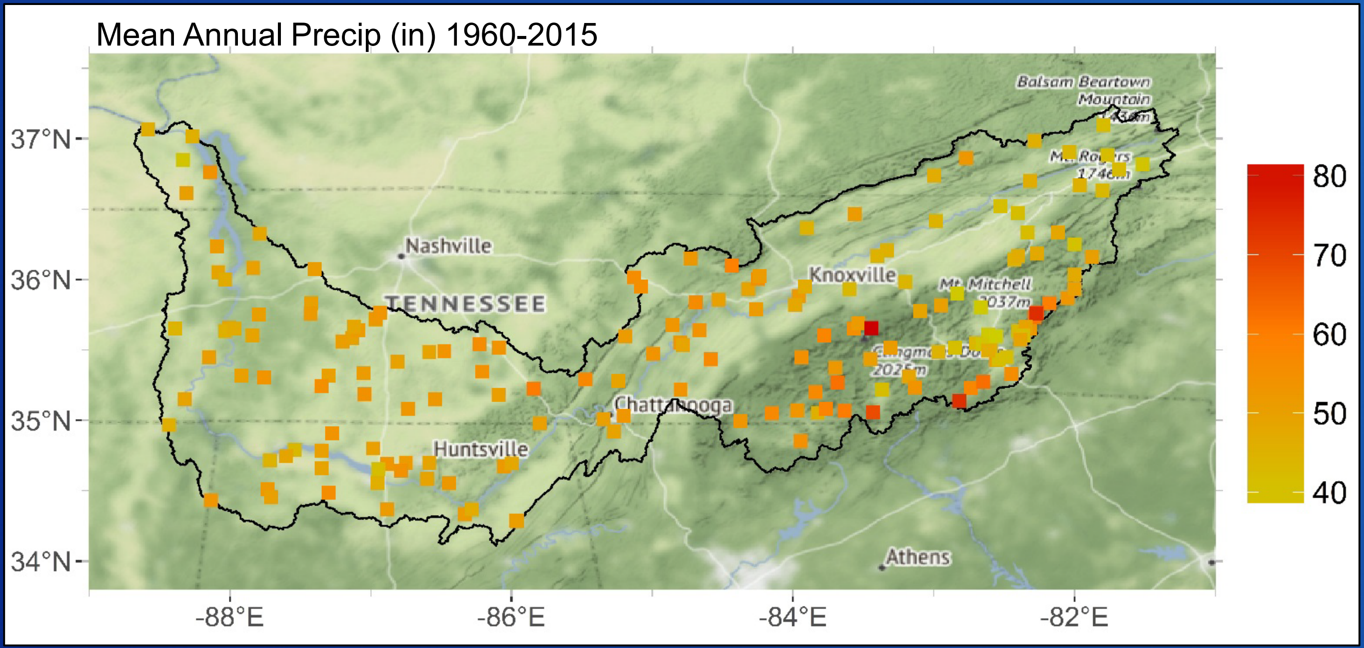


L-Moments

L-Moments RGCs

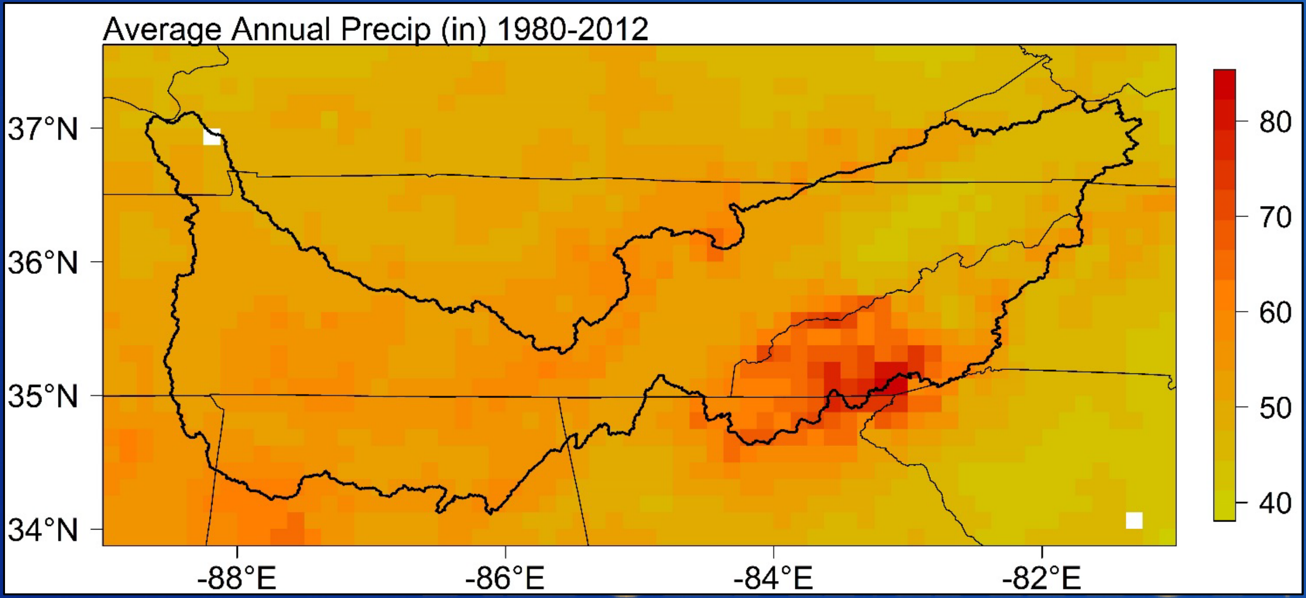


Data Availability

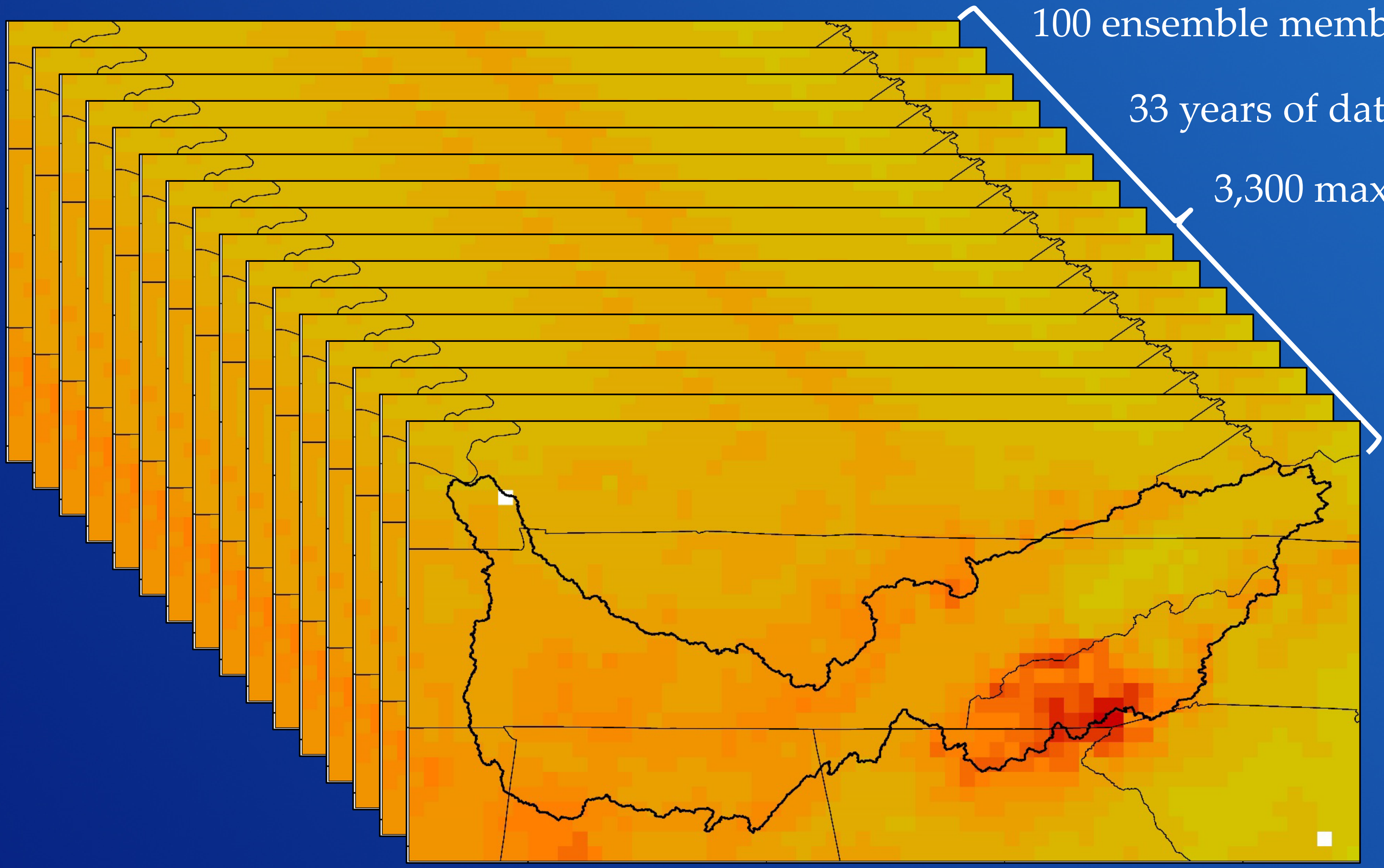


GHCN-Daily dataset

Newman et al. (2015) gridded ensemble



Combined Ensemble



100 ensemble members

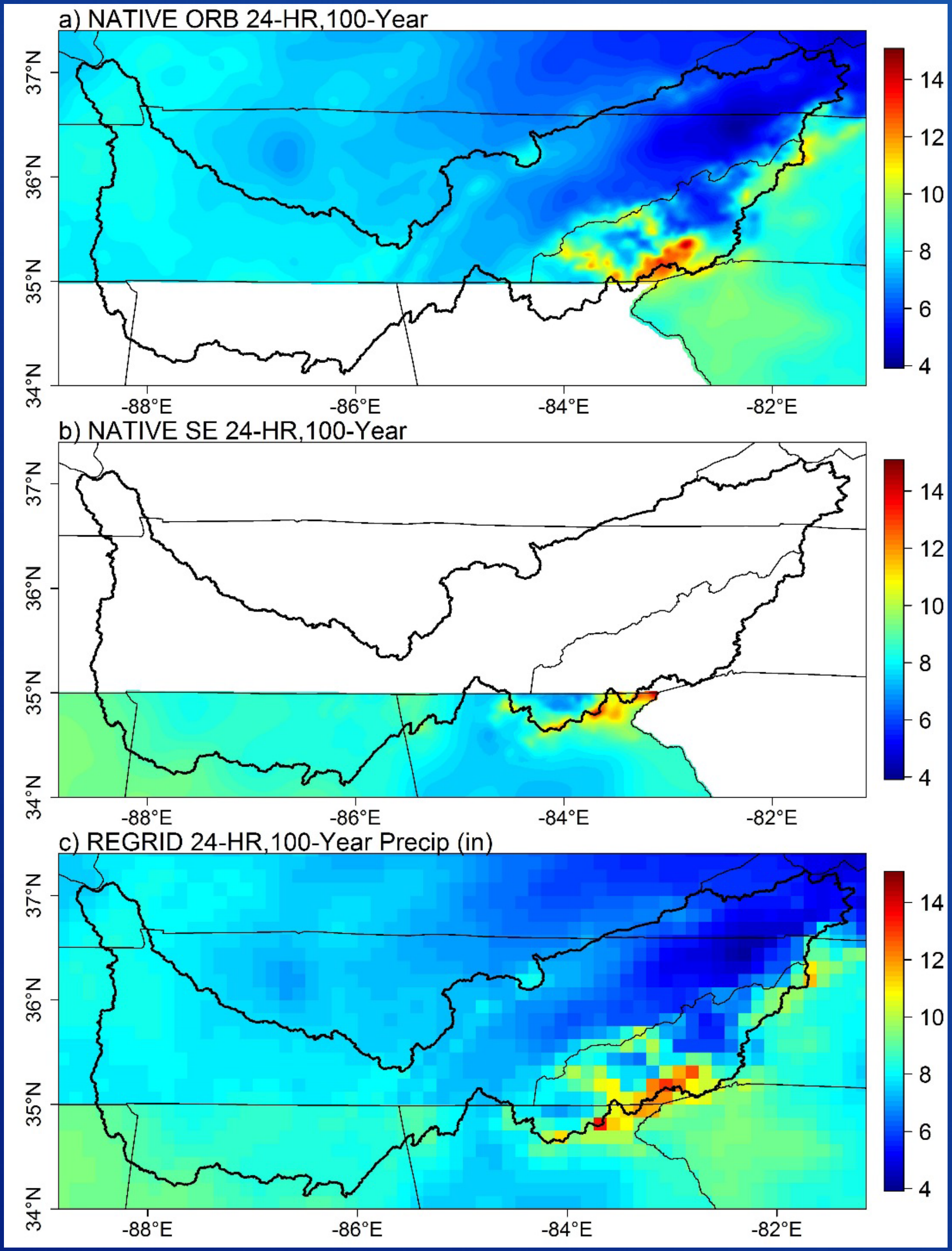
33 years of data

3,300 maxima/cell

Newman et al. (2015)

RECLAMATION

NOAA Atlas 14



Data from Ohio River Basin and Southeast domains use different lat/long grids

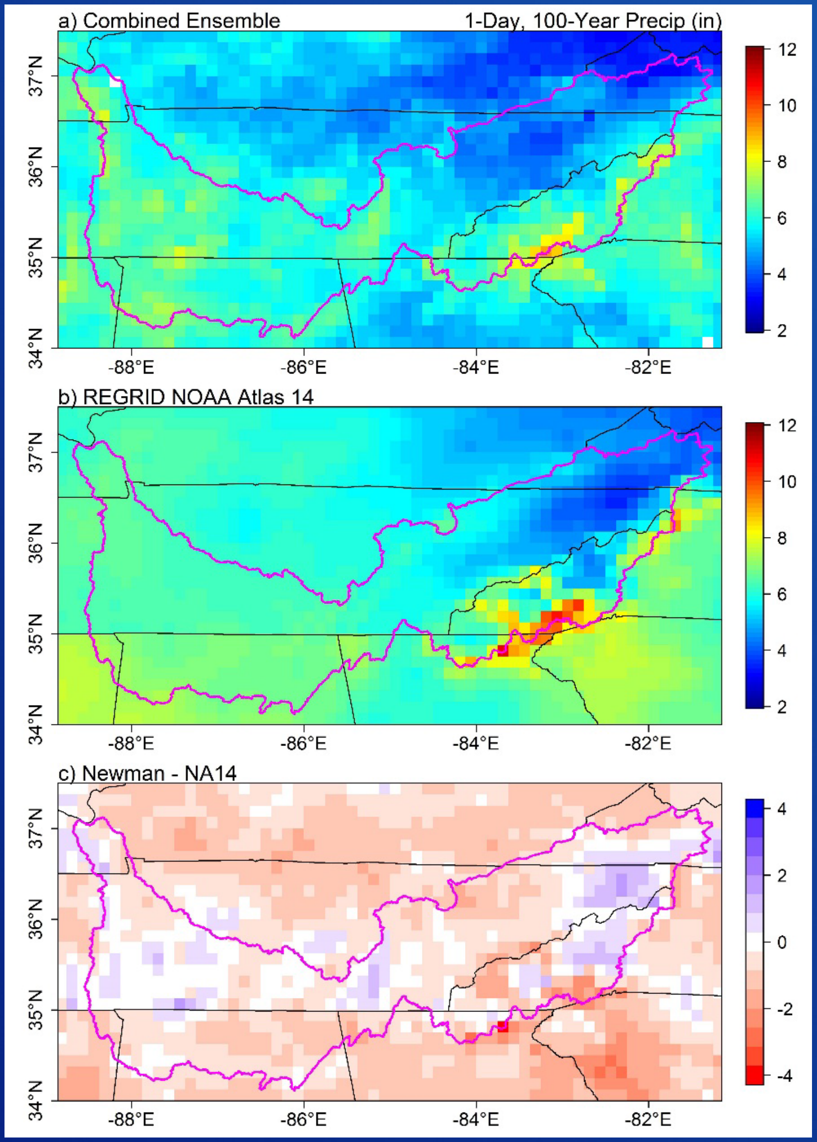
Regrid each field to the Newman resolution and then combine fields to produce a single field

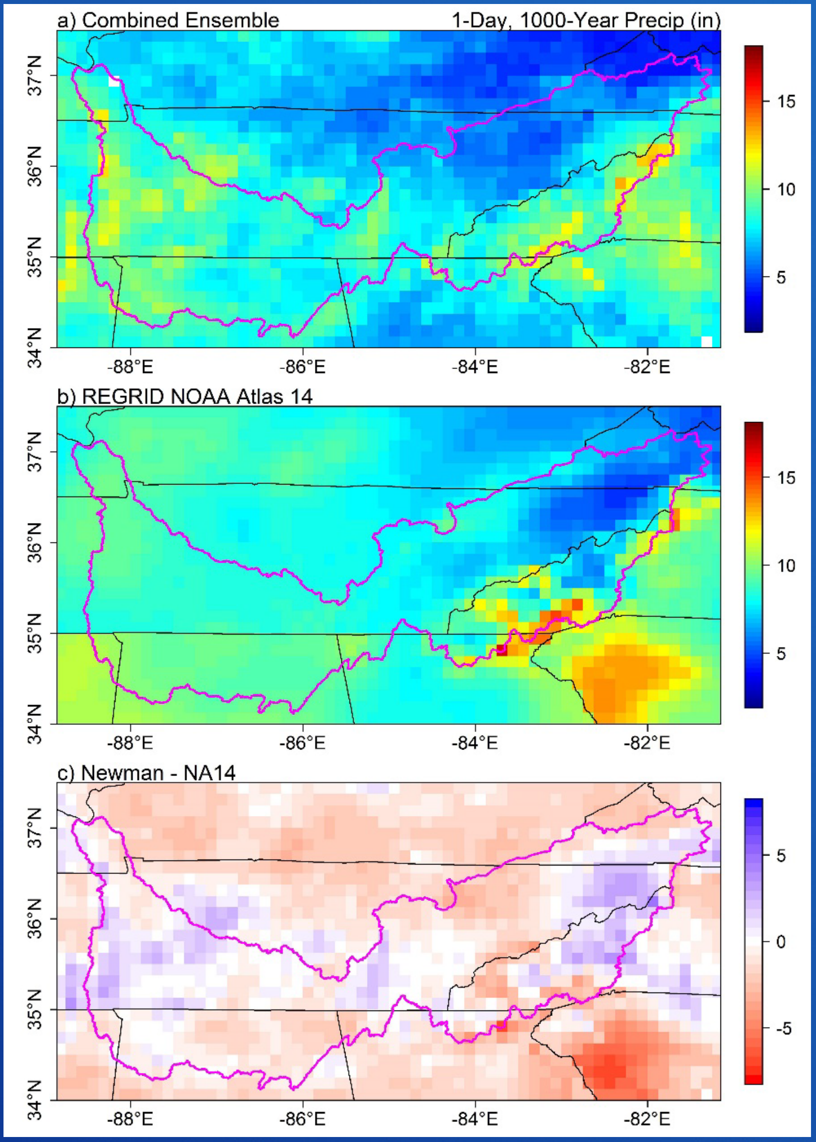
In order to compare with Newman estimates, applied two <u>scale factors</u>

1) Duration correction
2) Areal-reduction factor

RECLAMATION

Gridded L-moments



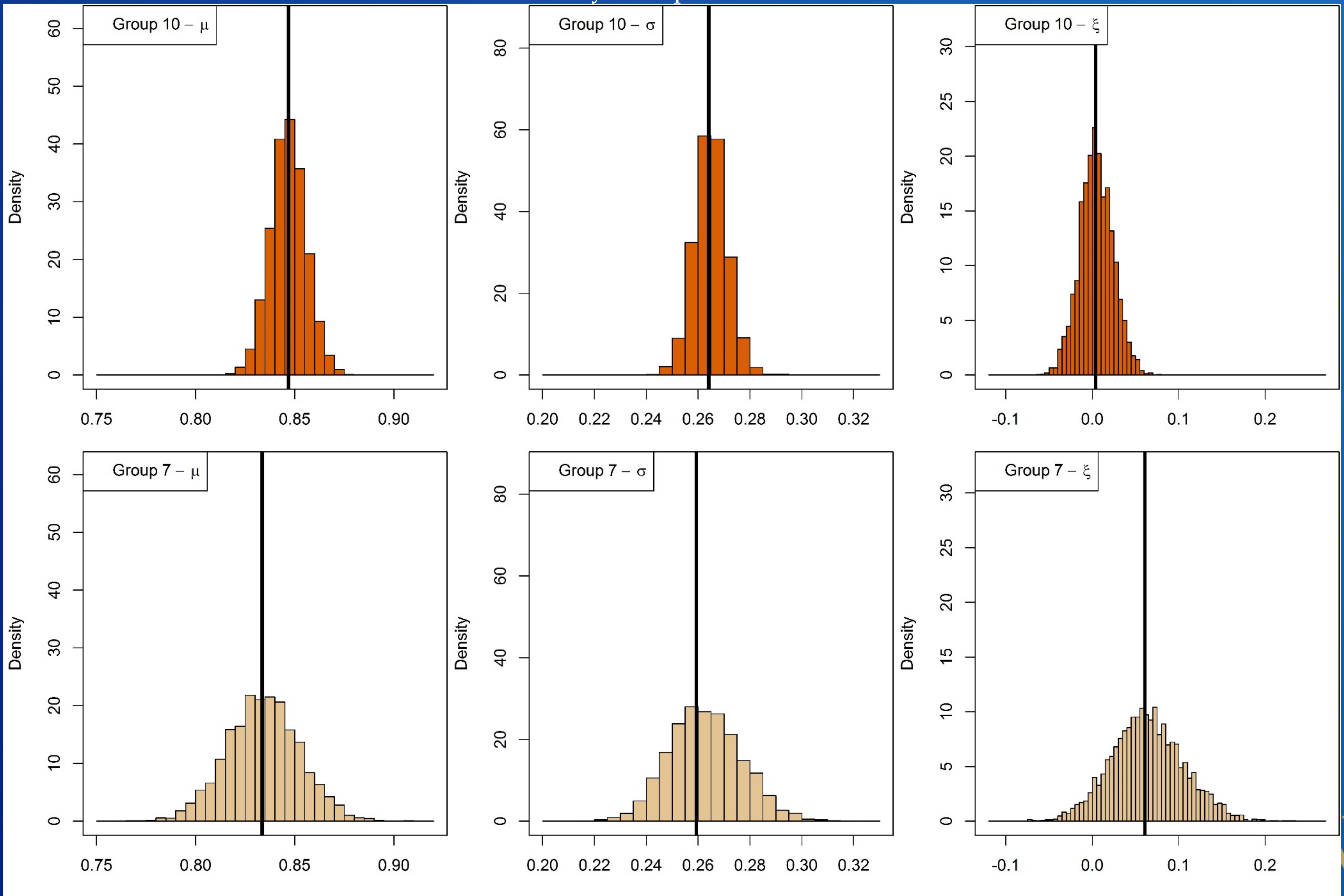


RECLAMATION

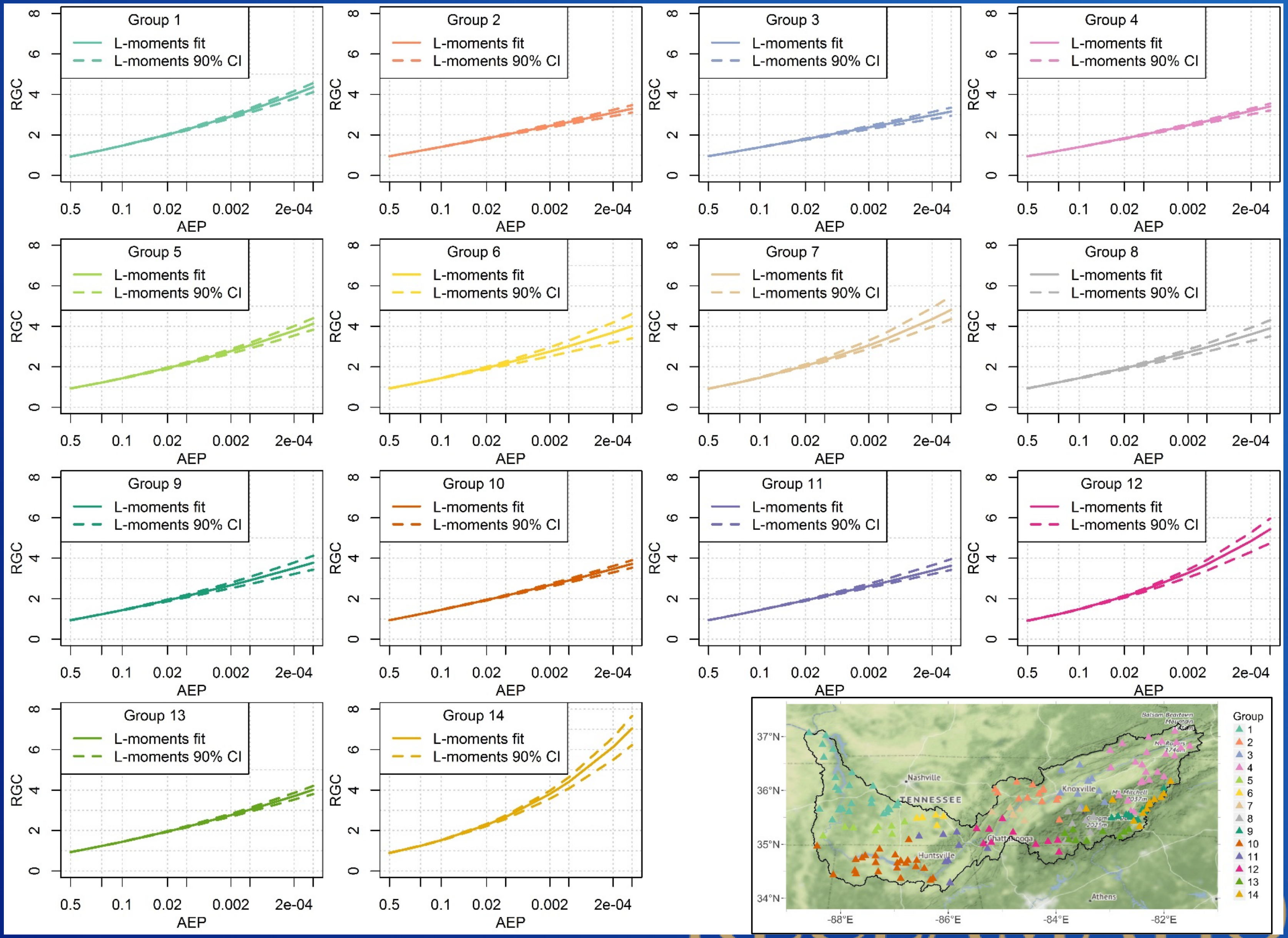
Bayesian

Posterior Distributions

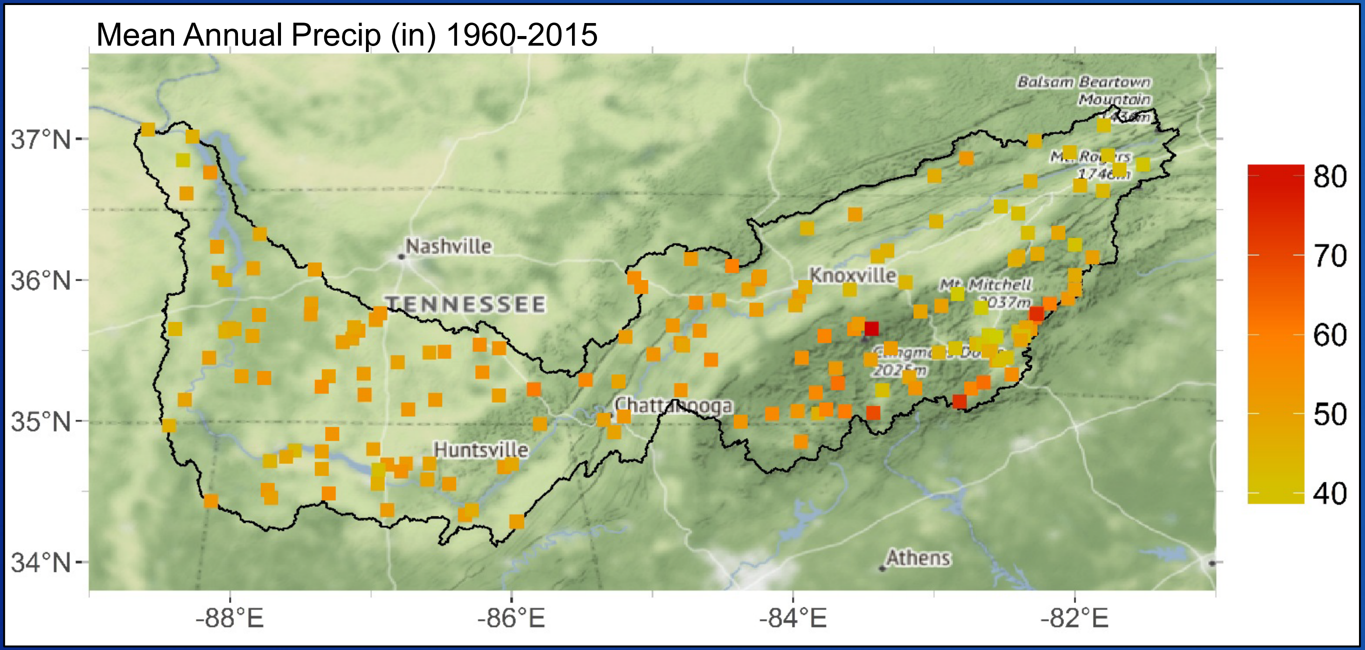
One-Day Precipitation Totals



Regional Growth Curves

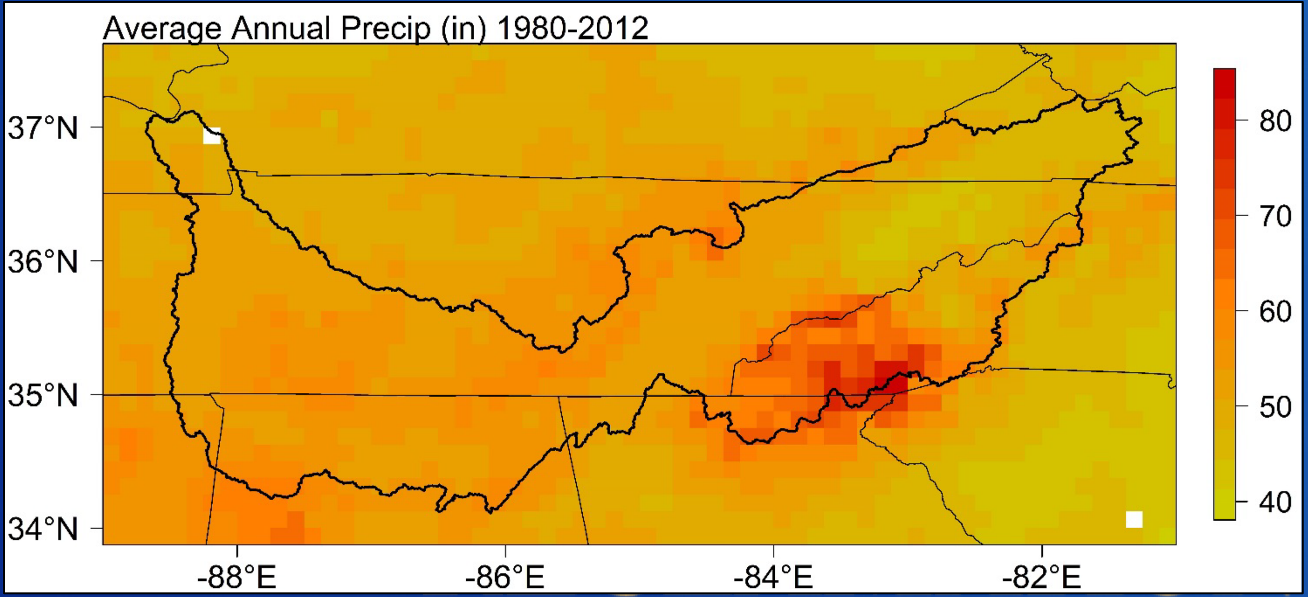


Data Availability

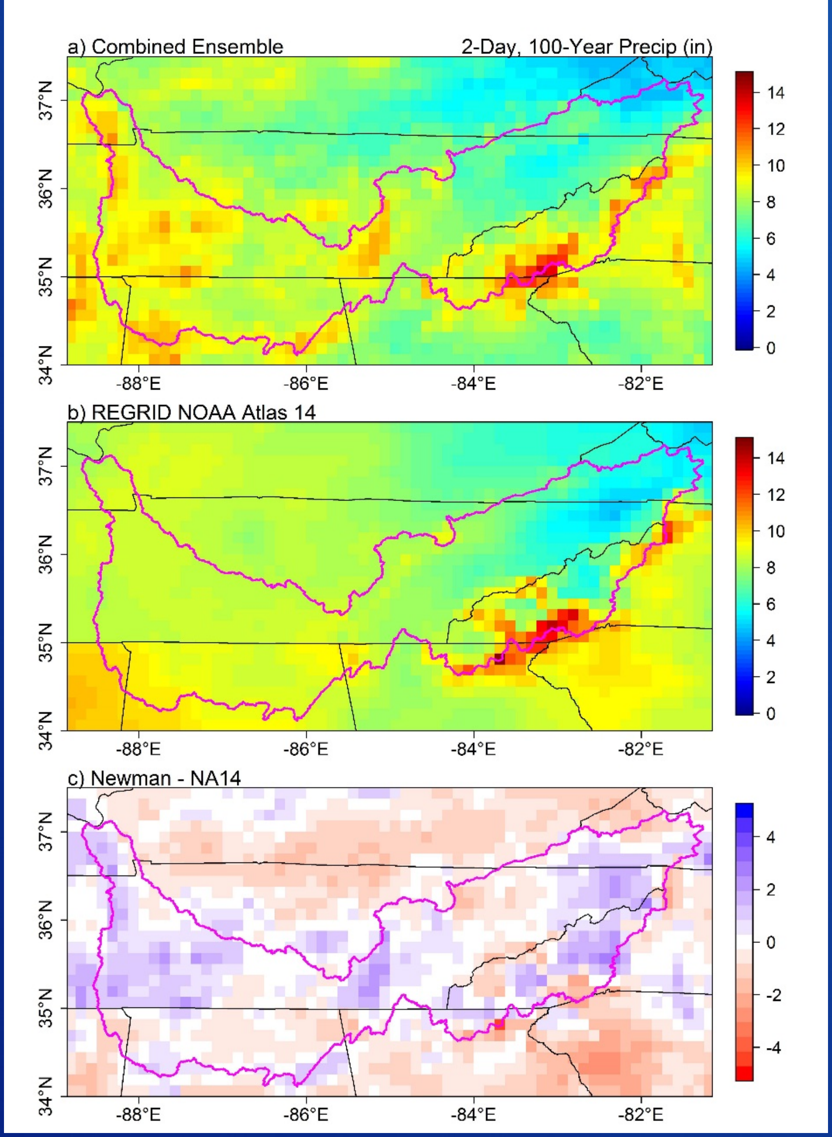


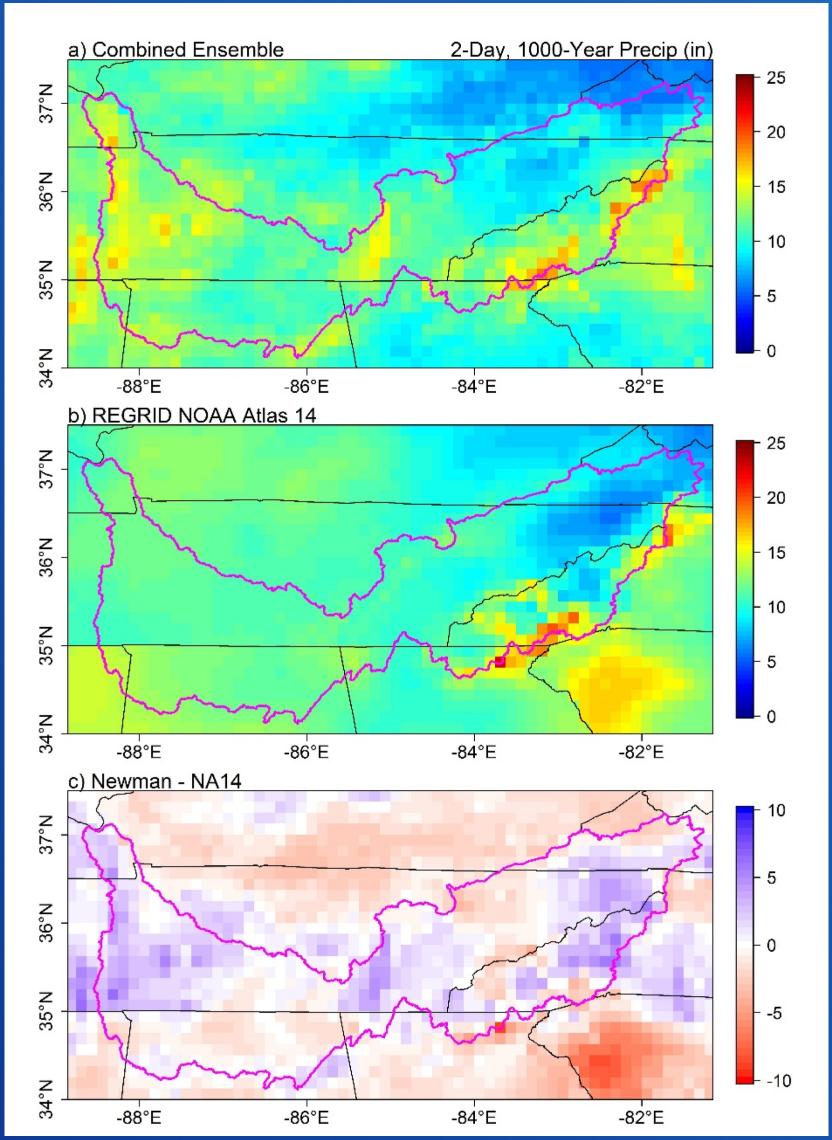
GHCN-Daily dataset

Newman et al. (2015) gridded ensemble



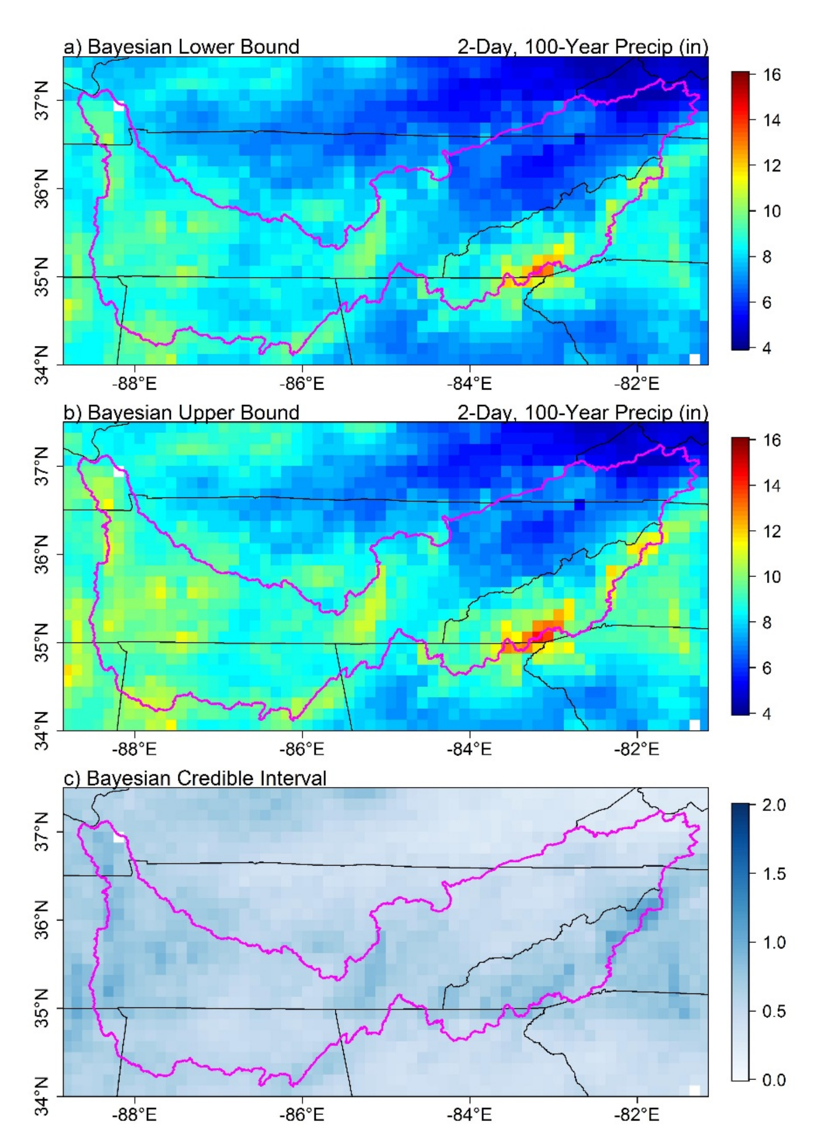
Gridded Bayesian - Medians

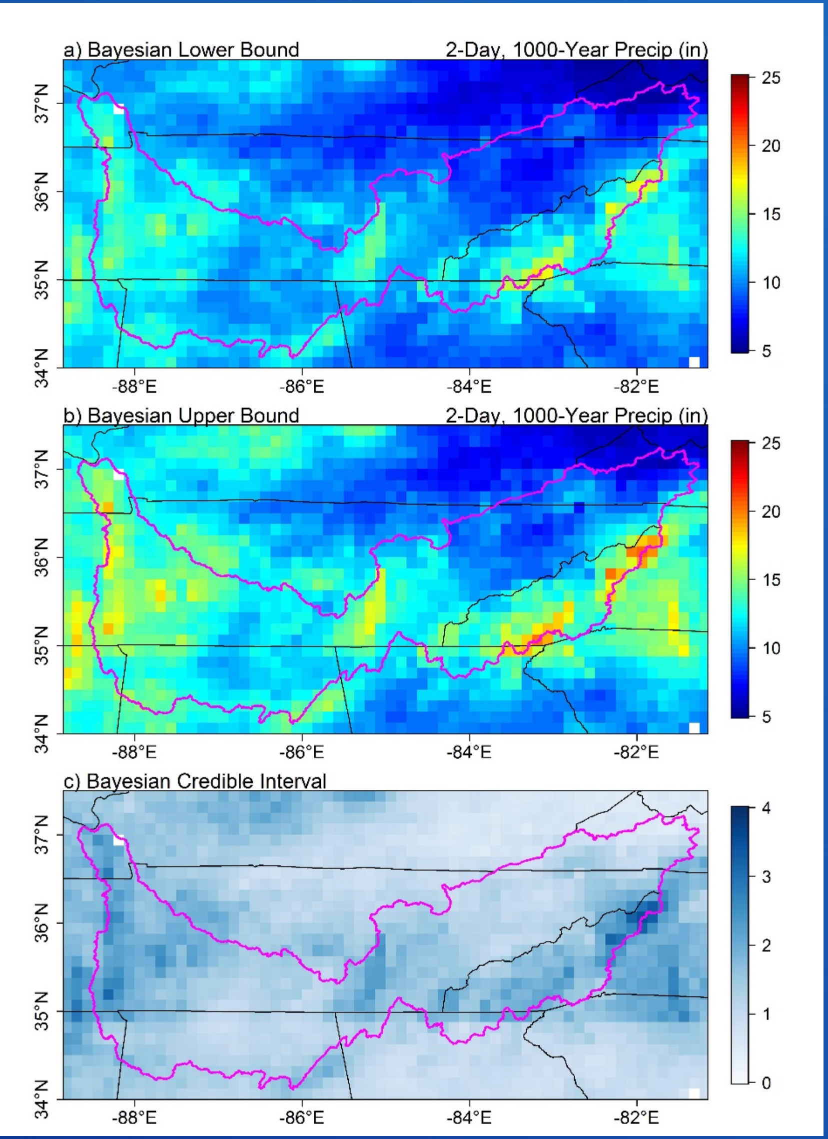




RECLAMATION

Gridded Bayesian - Uncertainty



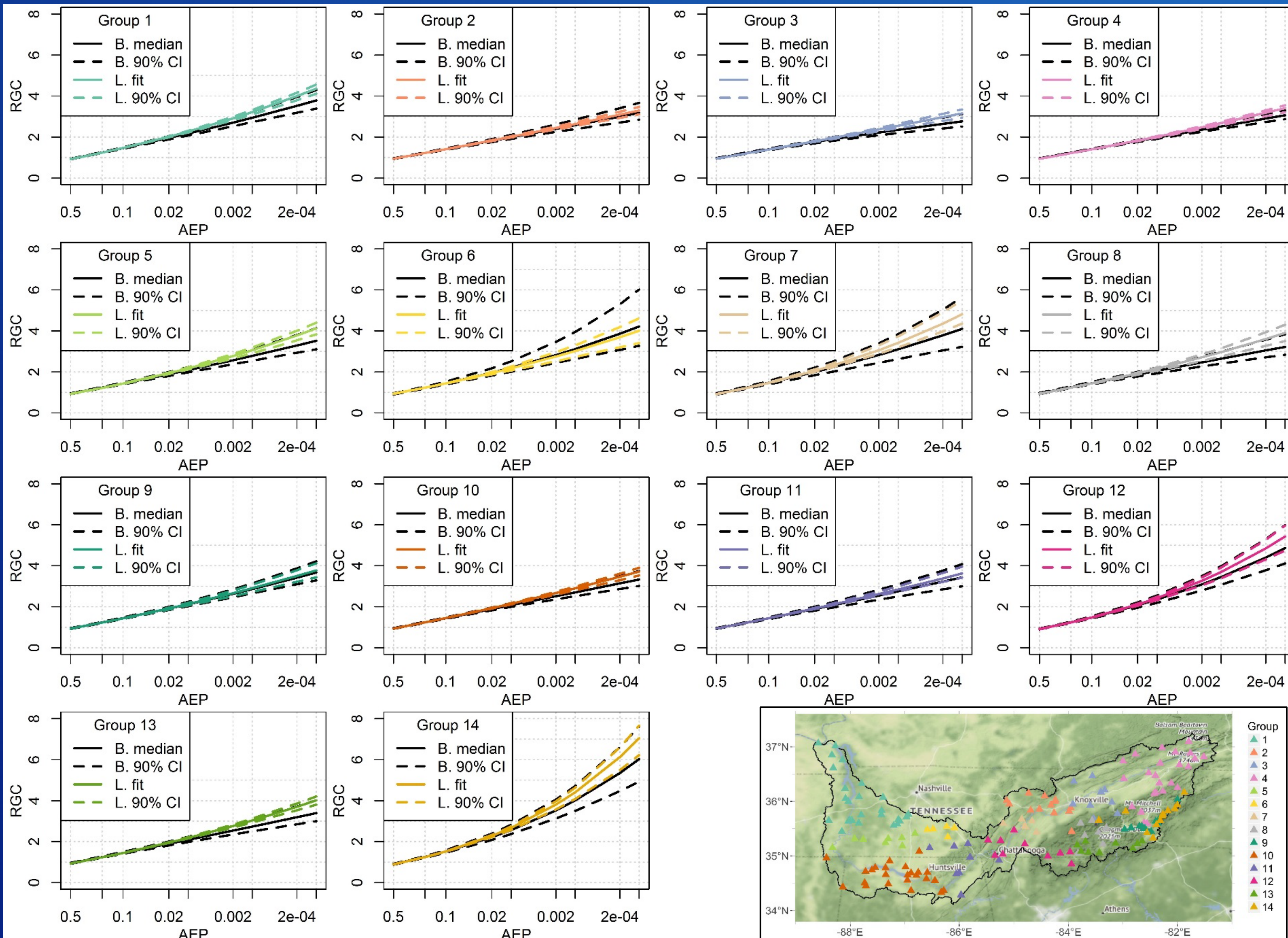


RECLAMATION

L-Moments vs. Bayesian

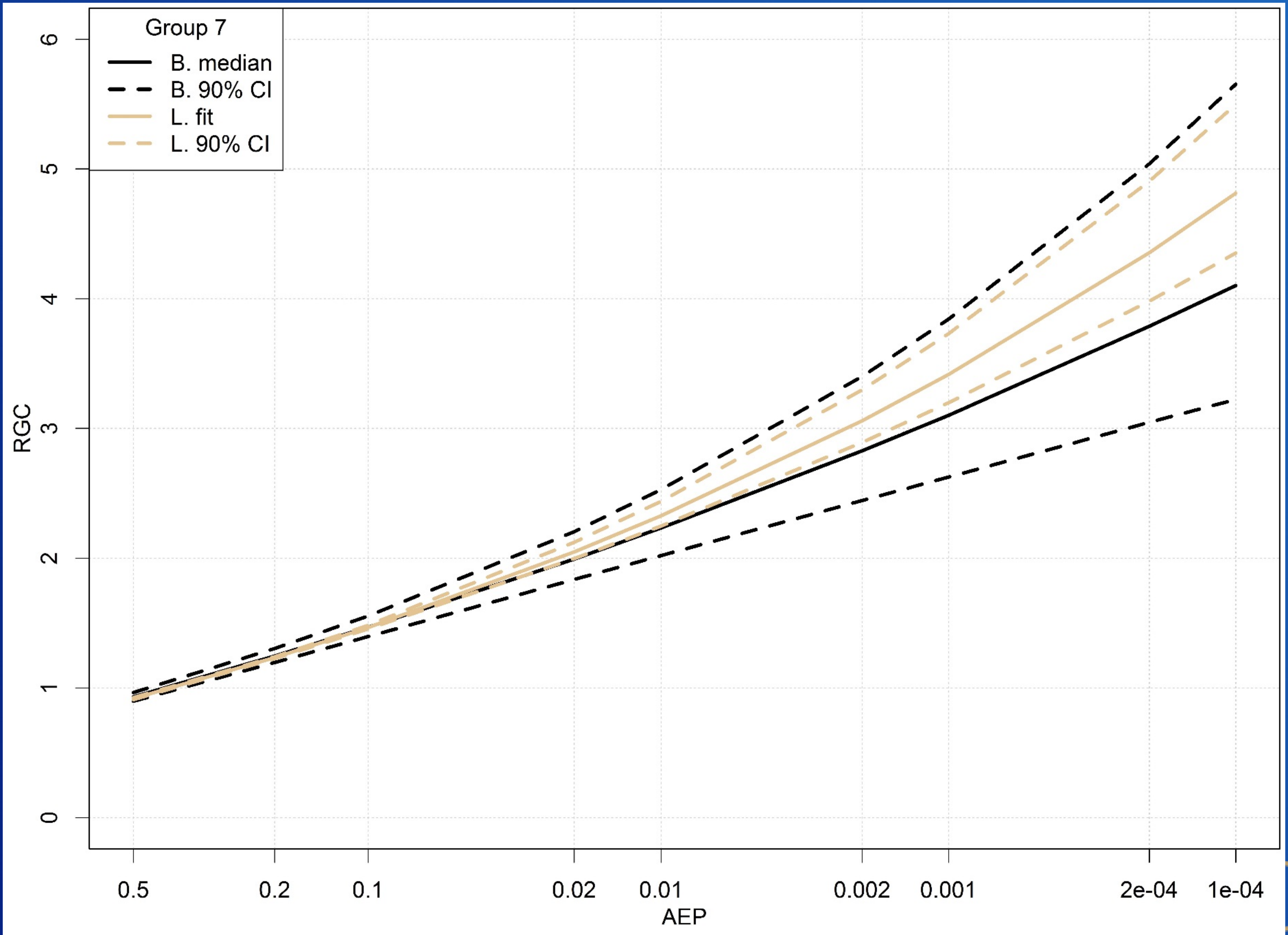
Regional Growth Curves

One-Day Precipitation Totals

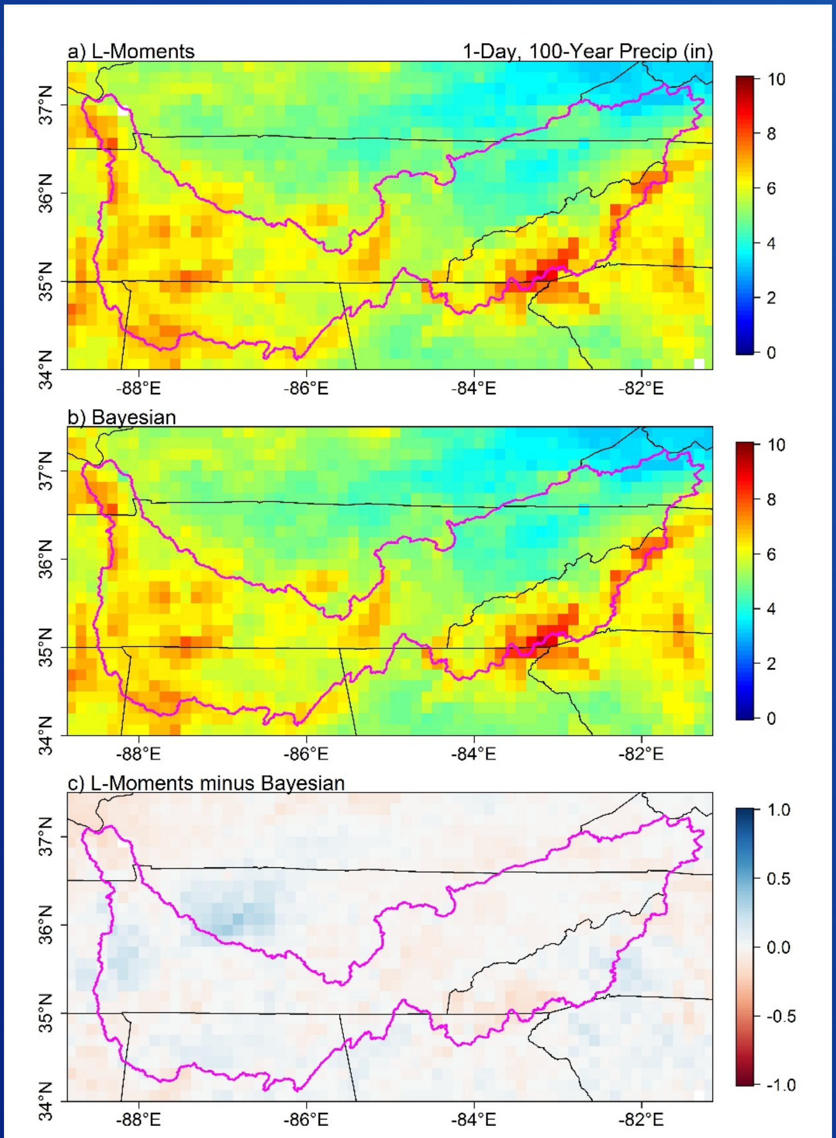


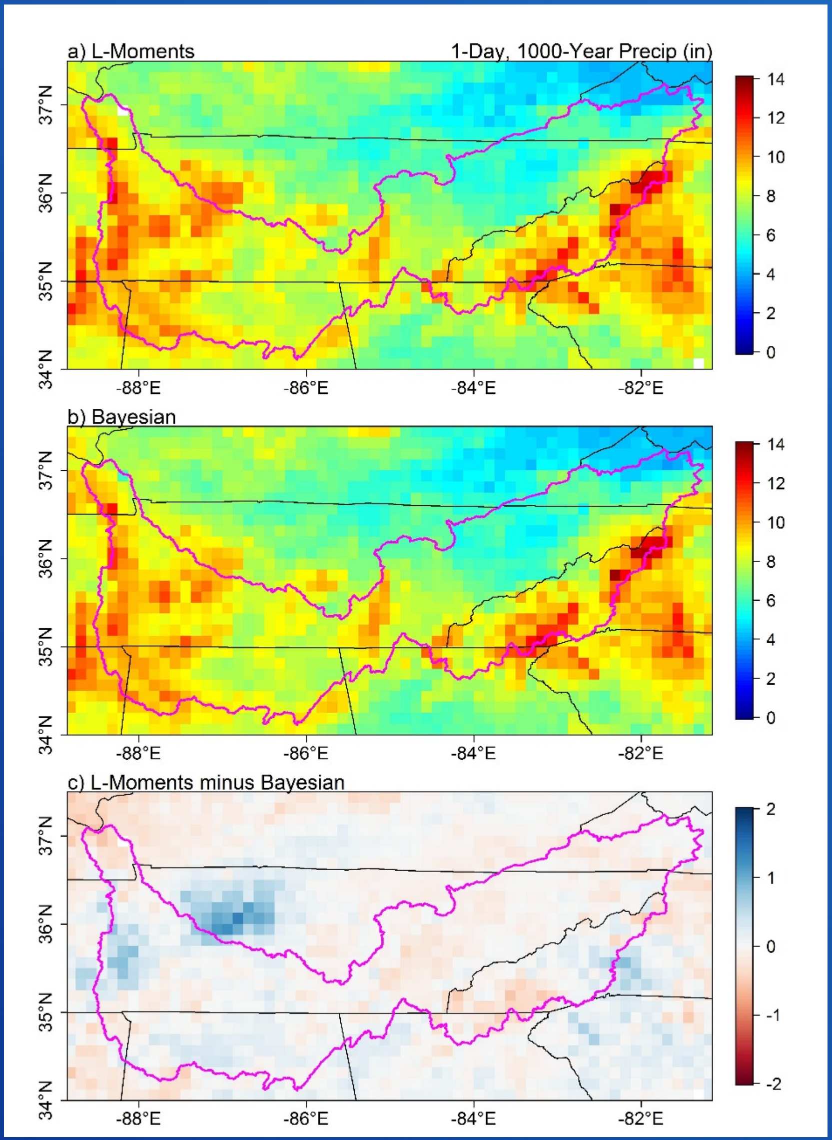
Regional Growth Curves

One-Day Precipitation Totals



Gridded Frequency Analysis





Summary

Regional frequency analysis using two methods

- L-moments and Bayesian inference

Homogeneous regions defined using semi-objective clustering algorithm, SOM (Self-Organizing Maps)

- Lon, lat, elev, mean annual precip, mean 1-day maxima
- Effectively accounts for orographics by clustering similar stations

Precipitation-frequency results and uncertainty bounds vary by method

- L-moments uses drop-10% bootstrap resampling
- Bayesian uses Monte Carlo, prior knowledge, likelihood function, acceptance criteria to build *posterior distributions*

Summary

Regional frequency estimation methods possible on any gridded dataset

- Newman used to address aleatory variability

L-moments and Bayesian produce similar estimates

- Bayesian median and L-moments fit show good agreement (< +/- 2" at 1,000-year return period)
- L-moments uncertainty estimation method lacking for large datasets (3,300 data points, drop-10% bootstrapping)
- Bayesian uncertainty also likely underestimated due to correlated data

Future work could entail multi-model combination

- Better quantify epistemic uncertainty via numerous gridded datasets, station-based datasets

RECLAMATION

Questions?

averdin@usbr.gov
+1.303.445.3647

kholman@usbr.gov
+1.303.445.2571