NCAR

U.S. Department of Transportation
Federal Highway Administration [1]

GENEX SYSTEMS

Towards finding trends in extreme values of precipitation: A preliminary Analysis

Eric Gilleland
Research Applications Laboratory

24 – 25 October 2016

Co-authors: Geoff Bonnin, Jerry Shen, Michael St-Laurent

This research is preliminary and subject to FHWA final acceptance.

National Center for Atmospheric Research

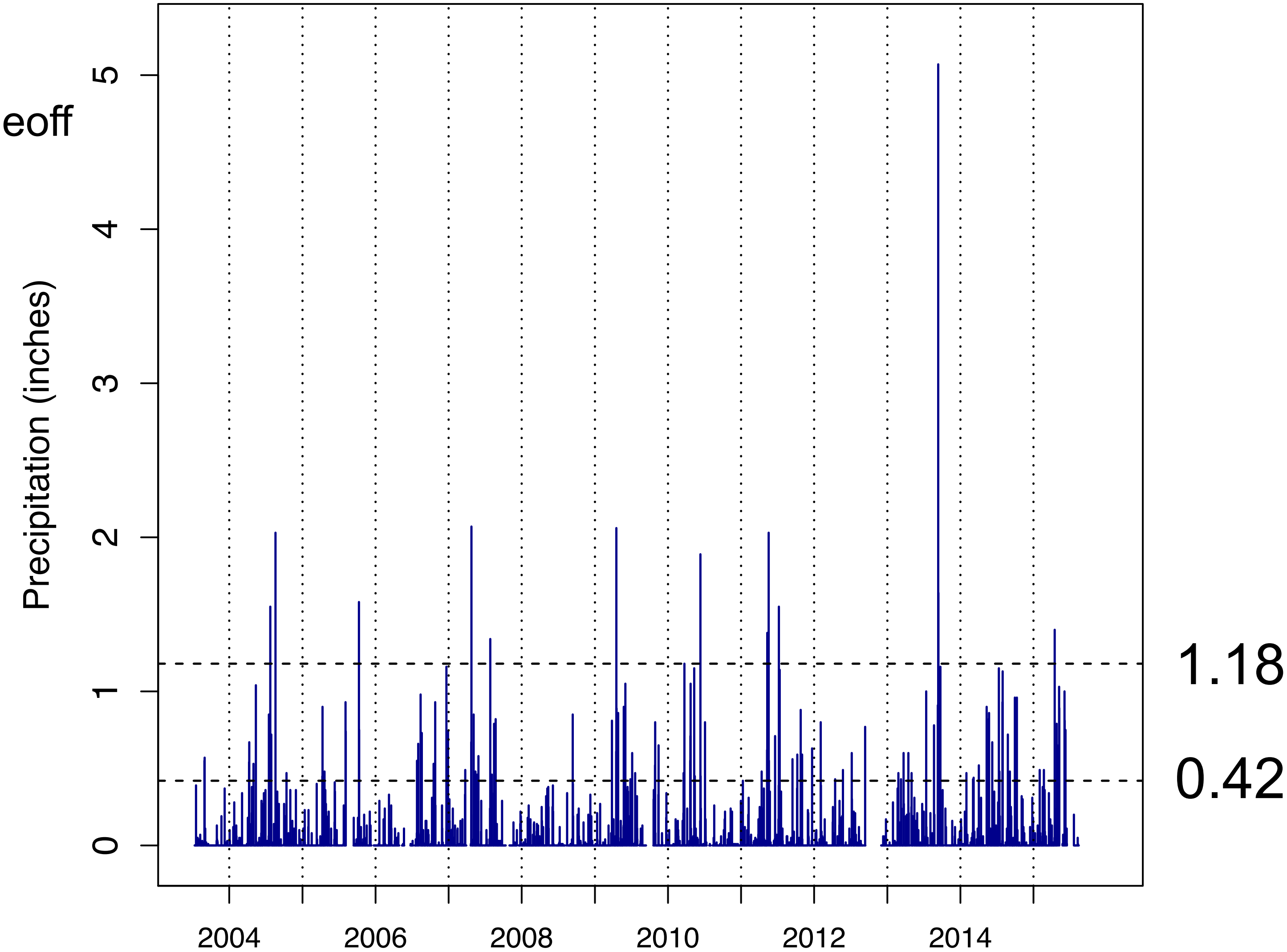
Data

NCAR

- Global Historical Climatology Network (GHCN)
  - https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/global-historical-climatology-network-ghcn
  - Daily data from 1893 to 2014
  - Two sub-regions of interest: New England area (NE) and semi-arid southwest (SW)
- Issues
  - Despite long temporal record, many stations moved during course of time leading to many stations with much shorter data records.
  - Occasional multi-day accumulations (instead of one day) resulting from human recordings not being taken over several days.

Threshold Selection

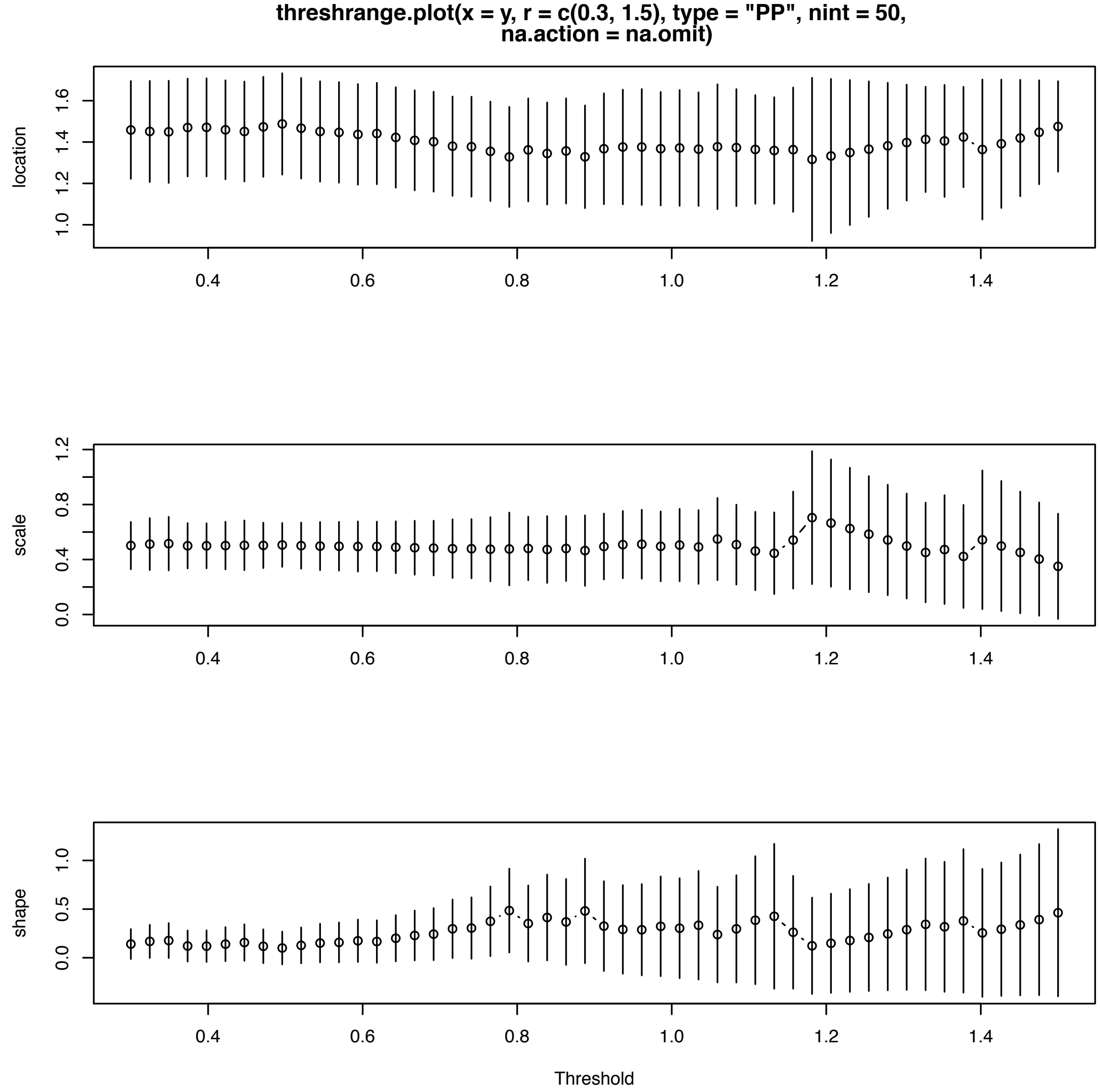
Variance/
Bias Tradeoff



Threshold Selection

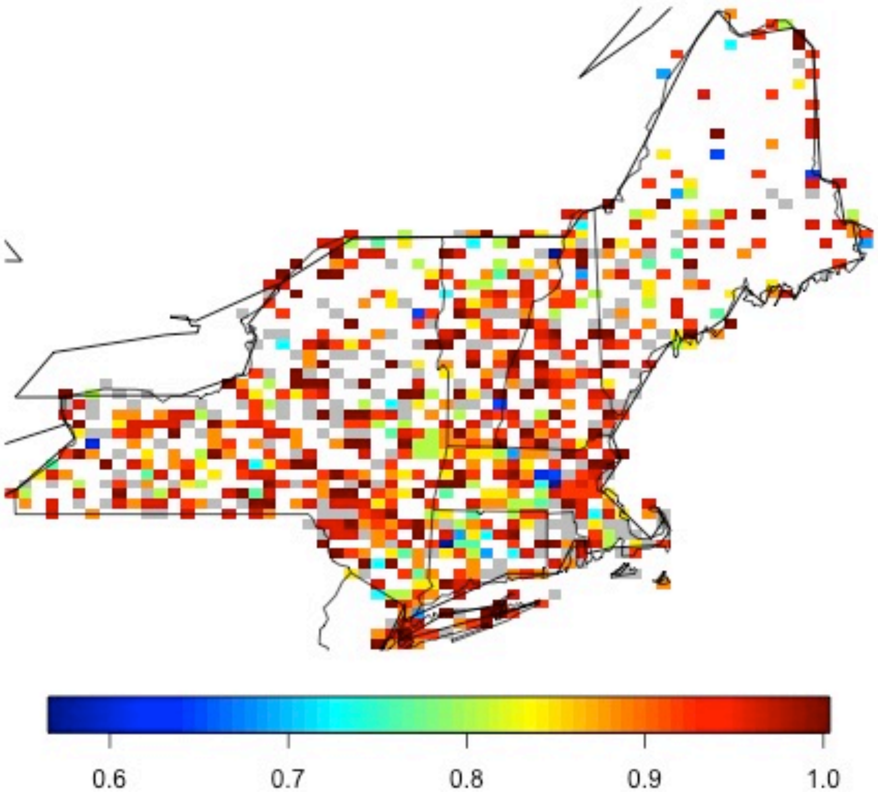
NCAR

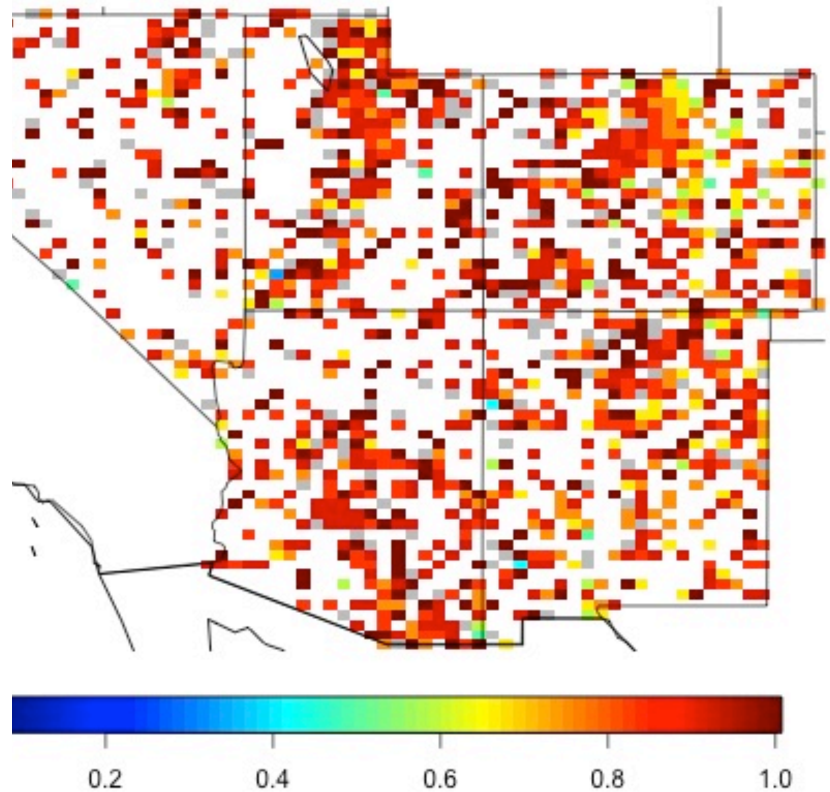
Variance/
Bias Tradeoff



Threshold Selection

Extremal Index estimates using threshold that yields equal number of excesses as years





PP – GEV (Northeast Region)

NCAR

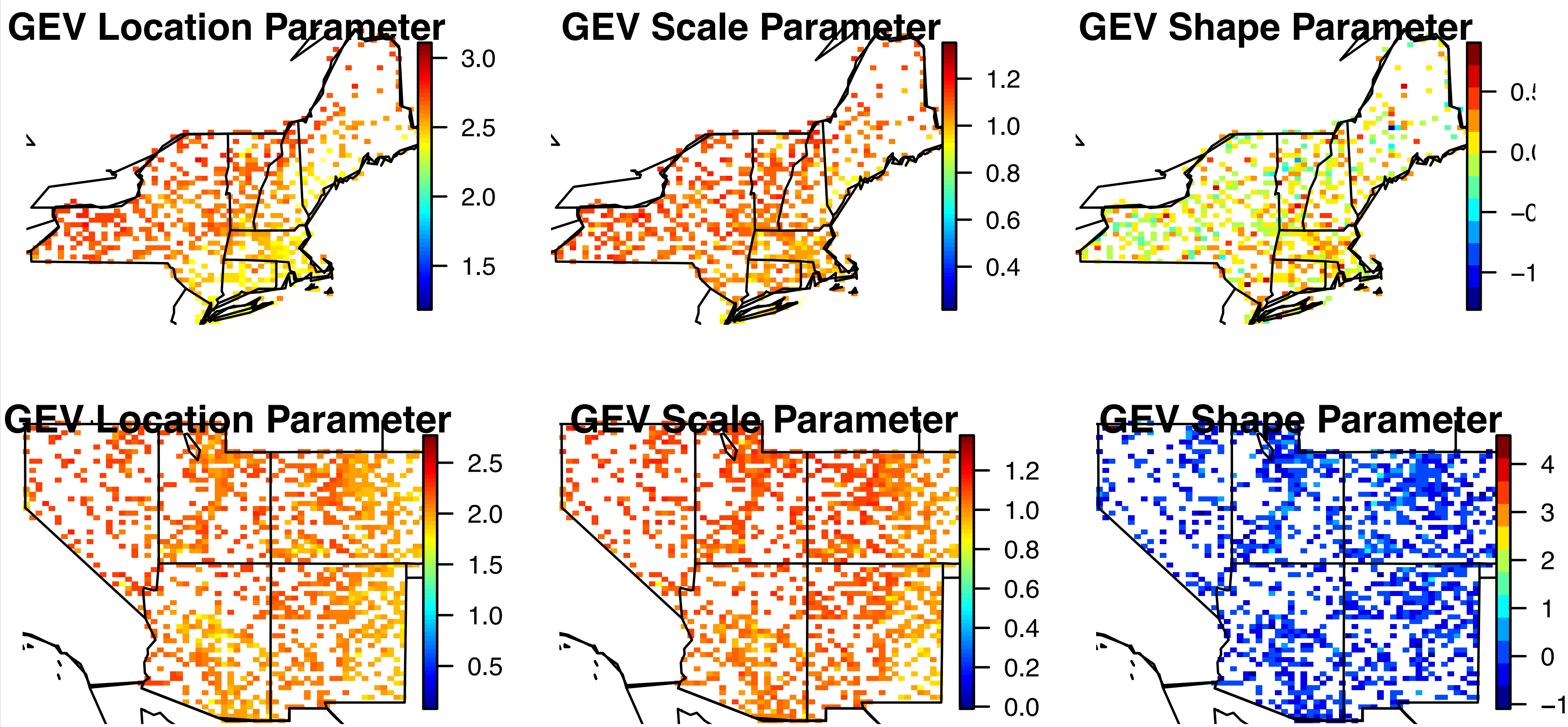
90th	location	scale	shape	500-year return level
Min	-0.36	-1.18	-0.82	-596.60
1st Quartile	0.02	-0.12	-0.09	-1.86
Median	0.09	-0.06	0.02	-0.08
Mean	0.10	-0.07	0.03	-1.95
3rd Quartile	0.17	0.01	0.15	1.04
Max	1.51	0.93	1.22	24.91
No. Missing	258	258	258	258

363 have shape with different signs. Total non-missing = 1000

95th	location	scale	shape	500-year return level
Min	-0.36	-1.41	-0.84	599.73
1st Quartile	0.02	-0.12	-0.08	-1.59
Median	0.09	-0.05	0.03	0.03
Mean	0.11	-0.06	0.03	-1.66
3rd Quartile	0.17	0.01	0.14	1.06
Max	1.67	1.32	1.20	85.55
No. Missing	258	258	258	258

350 have shape with different signs. Total non-missing = 1000

GEV fit to individual locations



Model 0

NCAR

GEV with one parameter set for entire region

GEV($\mu(\mathbf{s})$, $\sigma(\mathbf{s})$, $\xi(\mathbf{s})$)

$\mu(\mathbf{s}) = \mu$

$\sigma(\mathbf{s}) = \sigma$

$\xi(\mathbf{s}) = \xi$

Northeast (NE)

μ	σ	ξ
1.99	0.74	0.07

AIC = 144624.9, BIC = 144651.6

Southwest (SW)

μ	σ	ξ
1.03	0.52	0.15

AIC = 192141.1, BIC = 192169.5

Model 1

NCAR

GEV with different location and scale parameters for two sub-regions identified from individual fits

GEV($\mu(\mathbf{s})$, $\sigma(\mathbf{s})$, $\xi(\mathbf{s})$)

$\mu(\mathbf{s}) = \mu + \mathbf{1}_{\mathbf{s} \text{ in } A(\mu)} \delta_\mu$, $A(\mu)$ defined for the two regions

$\log \sigma(\mathbf{s}) = \phi(\mathbf{s}) = \phi + \mathbf{1}_{\mathbf{s} \text{ in } A(\phi)} \delta_\phi$, $A(\phi)$ defined for the two regions

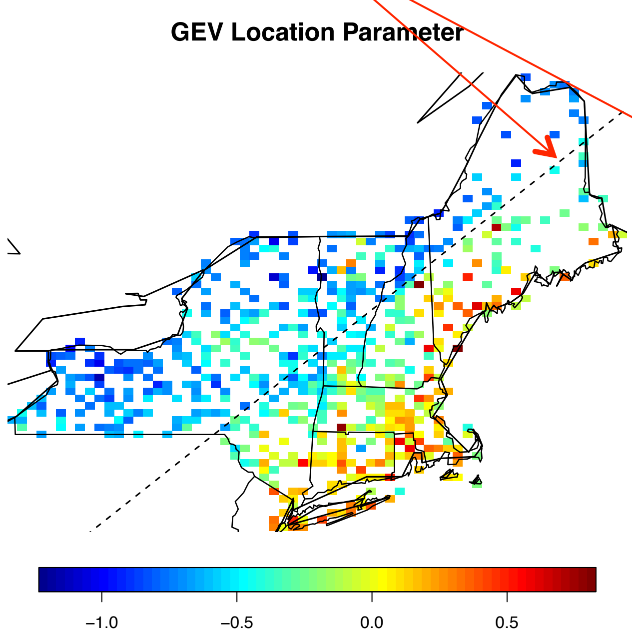
$\xi(\mathbf{s}) = \xi$

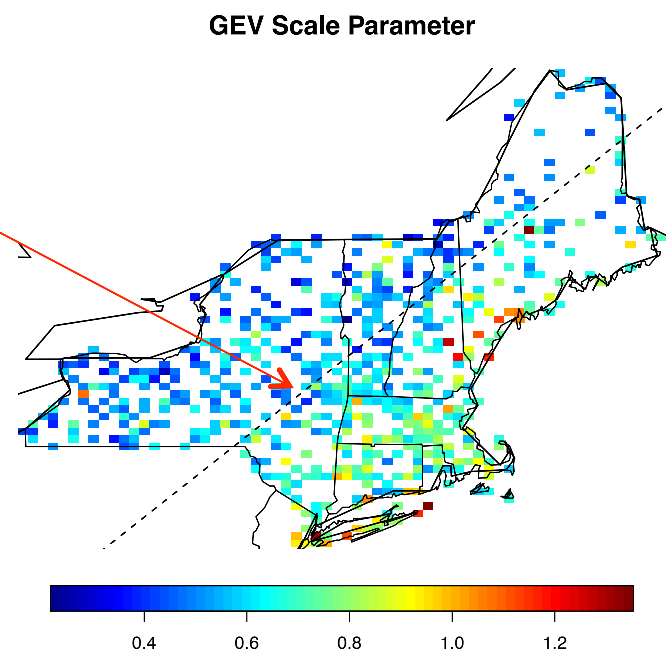
Model 1

Northeast Region

A(μ) = A(σ) and defined by being above or below the line

0.5710124 * longitude + 85.16103

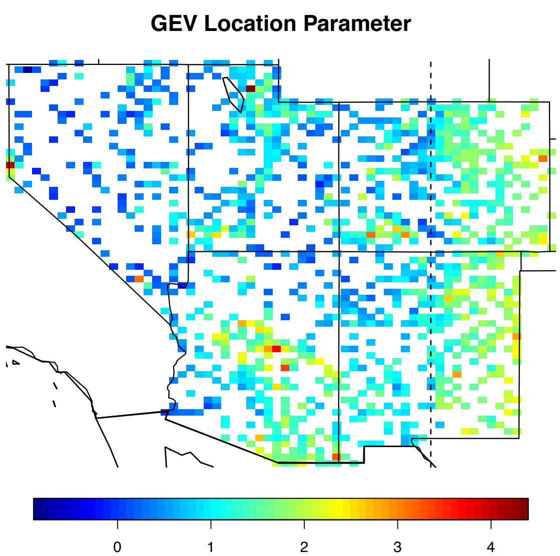


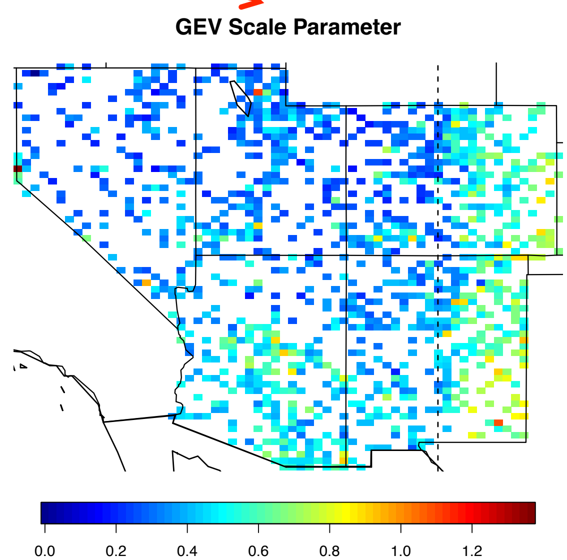


Model 1

Southwest Region

A(μ) = A(σ) and defined by being east of 106° W longitude

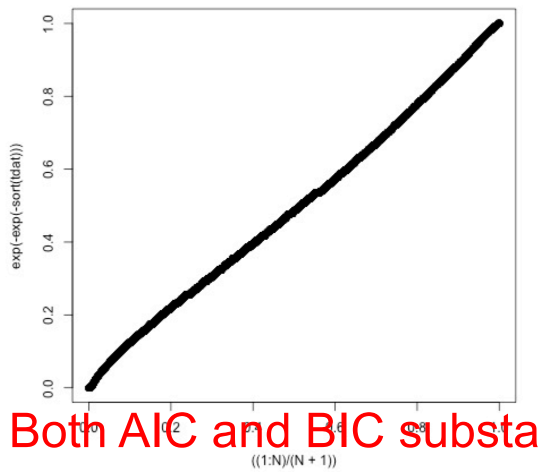




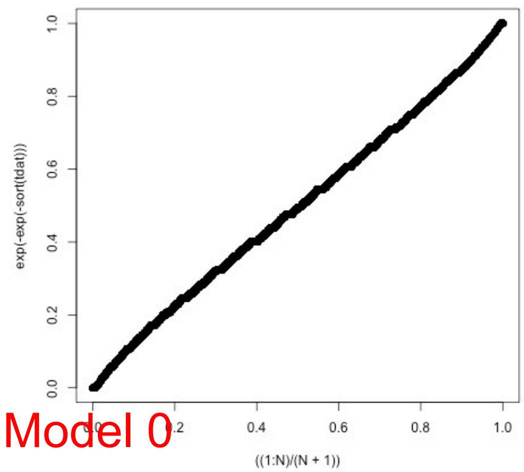
Model 1

NCAR

Northeast QQ-plot



Southwest QQ-plot



Both AIC and BIC substantially lower than Model 0

AIC = 137772, BIC = 137841.6

Northeast

μ	$\delta_\mu(1)$	$\delta_\mu(2)$	ϕ	$\delta_\phi(1)$	$\delta_\phi(2)$	ξ
1.99	-0.27	0.26	0.006	-0.54	-0.21	0.056

Southwest

μ	$\delta_\mu(1)$	$\delta_\mu(2)$	ϕ	$\delta_\phi(1)$	$\delta_\phi(2)$	ξ
1.08	0.20	-0.12	-0.251	-0.25	-0.50	0.145

AIC = 186365.3, BIC = 186438.8

1 = northwest / west, 2 = southeast / east

Model 2

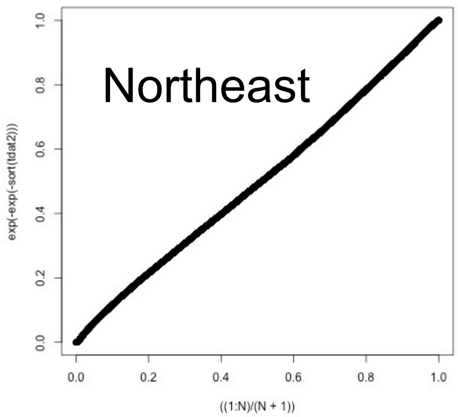
GEV with different location parameter at each location, and one scale parameter for each of the two sub-regions identified from individual fits

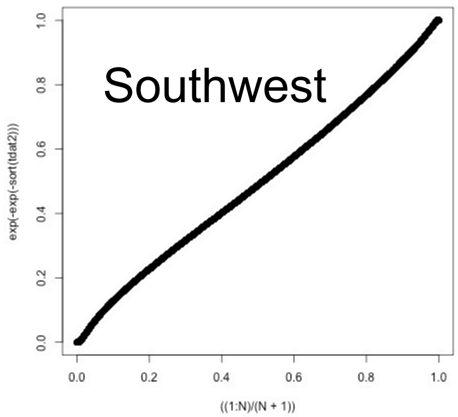
GEV(μ(**s**), σ(**s**), ξ(**s**))

$\mu(\mathbf{s}) = \mu + \mathbf{1}_{\mathbf{s} \text{ in } A(\mu)} \, \delta_\mu$, $A(\mu)$ defined as each individual station

$\log \sigma(\mathbf{s}) = \phi(\mathbf{s}) = \phi + \mathbf{1}_{\mathbf{s} \text{ in } A(\phi)} \, \delta_\phi$, $A(\phi)$ defined for the two regions

$\xi(\mathbf{s}) = \xi$





Model 2

NCAR

Northeast

μ	ϕ	$\delta_{\phi(1)}$	$\delta_{\phi(2)}$	ξ
2.03	-0.03	-0.56	-0.26	0.084

$\delta_{\mu(k)}$

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
-1.23	-0.29	-0.28	-0.02	0.006	0.28

AIC = 133926.4, BIC = 146492.6

Implies Model 2 better

AIC(Model 2) – AIC(Model 1) = -3845.583

BIC(Model 2) – BIC(Model 1) = 8650.942

Implies Model 1 better

Model 2

Southwest

μ	ϕ	$\delta_{\phi(1)}$	$\delta_{\phi(2)}$	ξ
1.10	-0.32	-0.26	-0.58	0.15

$\delta_{\mu(k)}$

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
-0.77	-0.21	-0.02	0.003	0.20	1.43

AIC = 165036.8, BIC = 188000.7

Implies Model 2 better

AIC(Model 2) – AIC(Model 1) = -21328.49

BIC(Model 2) – BIC(Model 1) = 1561.91

Implies Model 1 better

Model 3

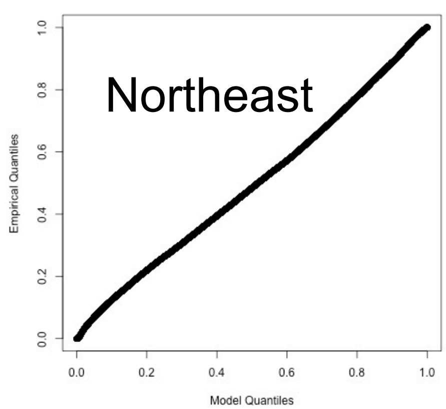
GEV with different location and scale parameters for two sub-regions identified from individual fits, and a temporal trend in the overall location parameter.

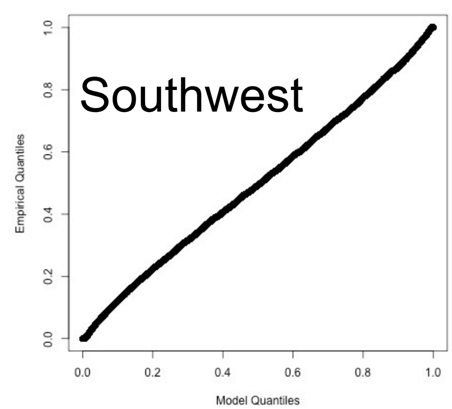
GEV($\mu(\mathbf{s}, \mathbf{year}), \sigma(\mathbf{s}), \xi(\mathbf{s})$), year = 1, 2, …

$\mu(\mathbf{s}, \mathbf{year}) = \mu_0 + \mu_1 \times \text{year} + \mathbf{1}_{\mathbf{s} \text{ in } A(\mu)}\, \delta_\mu$, $A(\mu)$ defined for the two regions

$\log \sigma(\mathbf{s}) = \phi(\mathbf{s}) = \phi + \mathbf{1}_{\mathbf{s} \text{ in } A(\phi)}\, \delta_\phi$, $A(\phi)$ defined for the two regions

$\xi(\mathbf{s}) = \xi$





Model 3

NCAR

	μ_0	μ_1	$\delta_{\mu(1)}$	$\delta_{\mu(2)}$	ϕ	$\delta_{\phi(1)}$	$\delta_{\phi(2)}$	ξ
NE	1.93	0.001	-0.30	0.23	-0.32	-0.21	0.12	0.05
SW	0.60	0.0005	0.65	0.33	-0.23	-0.27	-0.53	0.15

AIC favors Model 2 over Model 3
BIC favors Model 3 over Model 2

But, trend terms are negligible in both models (not likely to be significant).

Model 4

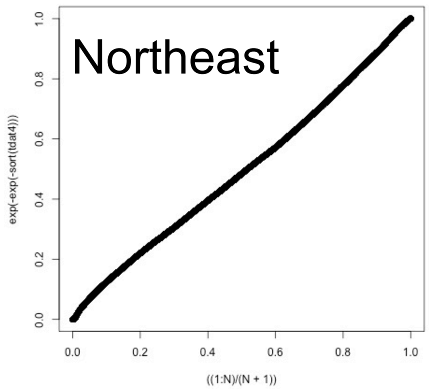
GEV with different location and scale parameters for two sub-regions identified from individual fits, and a temporal trend in the overall location parameter.

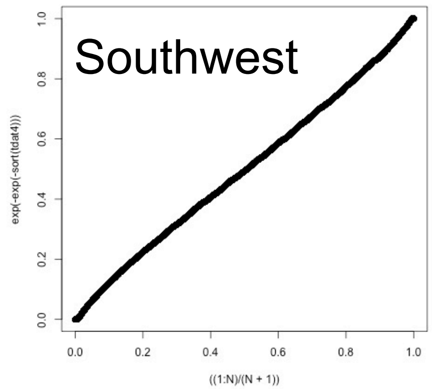
GEV(μ(**s, year**), σ(**s**), ξ(**s**)), year = 1, 2, …

$\mu(\mathbf{s}, \mathbf{year}) = \mu_0 + \mathbf{1}_{\mathbf{s} \text{ in } A(\mu)}\ [\ \delta_{\mu,0} + \delta_{\mu,1} \times \text{year}\ ]$, $A(\mu)$ defined for the two regions

$\log \sigma(\mathbf{s}) = \phi(\mathbf{s}) = \phi + \mathbf{1}_{\mathbf{s} \text{ in } A(\phi)}\, \delta_\phi$, $A(\phi)$ defined for the two regions

$\xi(\mathbf{s}) = \xi$





Model 4

NCAR

Northeast

Implies Model 2 better

μ_0	$\delta_{\mu,0}(1)$	$\delta_{\mu,0}(2)$	$\delta_{\mu,0}(1)$	$\delta_{\mu,0}(1)$	ϕ	$\delta_{\phi(1)}$	$\delta_{\phi(2)}$	ξ
1.97	-0.33	0.16	0.001	0.002	-0.003	-0.53	-0.20	0.05

AIC = 137574.4, BIC = 137663.9

Both AIC and BIC suggest Model 4 is better than Model 1

Southwest

μ_0	$\delta_{\mu,0}(1)$	$\delta_{\mu,0}(2)$	$\delta_{\mu,0}(1)$	$\delta_{\mu,0}(1)$	ϕ	$\delta_{\phi(1)}$	$\delta_{\phi(2)}$	ξ
1.01	0.23	-0.08	0.001	≈ 0	-0.34	-0.16	-0.42	0.15

AIC = 186269, BIC = 186363.5

Implies Model 4 better

Model 5+ (Future work?)

NCAR

Invoke a spatial process on the location (and scale?) parameters across entire region. Allow for a temporal trend in one or more parameter(s).

GEV($\mu(\mathbf{s, year})$, $\sigma(\mathbf{s, year})$, $\xi(\mathbf{s, epoch})$), year = 1, 2, …

($\mu(\mathbf{s, year})$, $\log \sigma(\mathbf{s})$) ~

Gaussian Process(($\text{mean}_{\text{location}}$(year), $\text{mean}_{\text{scale}}$(year)), Covariance)

$\xi(\mathbf{s, epoch}) = \xi$

Not feasible to allow shape parameter to vary every year, but may be good to allow it to vary every ten years (or more).

Model 6+ (Future work?)

NCAR

Following approach of Reich and Shaby (2012, doi: 10.1214/12-AOAS591) and Stephenson et al. (2015, doi:10.1175/JAMC-D-14-0041.1)

Let **A** = (A_1, …, A_K) be K independent random variables distributed according to a positive stable distribution with index equal to the spatial-dependence parameter α.

Define

$$\theta(s_i) = \left[\sum_{k=1}^{K} A_k w_k(s_i)^{1/\alpha} \right]^{\alpha}$$

kernel basis functions with $w_k \geq 0$

Precipitation | **A** ~ GEV($\mu^*(\mathbf{s}_i)$, $\sigma^*(\mathbf{s}_i)$, $\xi^*(\mathbf{s}_i)$) ← $\xi^*(\mathbf{s}_i) = \alpha\xi(\mathbf{s}_i)$

$\mu^*(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \sigma(\mathbf{s}_i)[\theta(\mathbf{s}_i)^{\xi(\mathbf{s})} - 1]$ $\sigma^*(\mathbf{s}_i) = \alpha\sigma(\mathbf{s}_i)\ \theta(\mathbf{s}_i)^{\xi(\mathbf{s})}$

Inference via Bayesian estimation

But, have $\xi^*(\mathbf{s}_i) = \alpha\xi(\mathbf{s}_i) = \alpha\xi$ (is this model still valid?)

Incorporate trend via **A**, perhaps by way of changing return level estimates

Summary

NCAR

- Threshold selection is challenging, but worthwhile endeavor in order to use PP model to obtain better estimates (less uncertainty) at greater expense of time.
- Model 4
  - reasonably parsimonious model
  - allows for pooling of data across locations
  - Shows promise in that AIC / BIC results are good
  - qq-plots reasonably linear
- Model 2 may be improved by imposing a spatial process on the parameter estimates (penalized likelihood problem / Bayesian)
- Model 3 incorporates temporal trend, but not significant for these regions
  - Consistent with other results, but …
    - Need to check Data Quality issues
    - Need more careful determination of sub-regions
- Estimated shape parameter consistent across models
- Not much variability in location/scale parameters within sub-regions
- Model 4 suggests small positive trend in NE, but not much trend in SW

Future Work?

NCAR

- Choose sub-regions more carefully
- Test for homogeneity of shape parameter in regions.
- Allow shape to vary some?
- More models (e.g., allow other parameters to vary in time).
- Incorporate covariates?
- Analyze resulting return levels
- Account for non-stationarity in return levels

Thank you for your attention.

Questions?