Trends in observational and model-generated precipitation extremes: are they compatible?

Richard Smith

Department of Statistics and Operations Research
University of North Carolina, Chapel Hill

IDAG Meeting

Duke University

March 14 2007

Method for analyzing extreme precipitation events

(or, *What I've learned from SAP 3.3*)

Most popular method is the "NCDC approach" stemming from Karl and Knight (1998) and continued by Easterling, Groisman and others (see also Kunkel and co-authors, 1999, 2003a, 2003b, 2006,....). Calculate trends in exceedance counts of a high threshold at each station, then group and spatially smooth into climate regions.

A few authors have taken the "GEV" approach (Zwiers, Wehner and co-authors)

There are (at present) very few papers in the climate literature on the "threshold" approach to extremes, though there are papers starting to appear that analyze significant amounts of climate data (Cooley *et al.*, to appear in *JASA*; Shamseldin *et al.*, in preparation)

This talk is based primarily on work by R. Smith, A. Grady and G. Hegerl, preprint will shortly be available.

DATA SOURCES

- NCDC Rain Gauge Data (Groisman 2000)
  - Daily precipitation from 5873 stations
  - Select 1970–1999 as period of study
  - 90% data coverage provision —— 4939 stations meet that

- NCAR-CCSM climate model runs
  - 20 $\times$ 41 grid cells of side 1.4$^{\text{o}}$
  - 1970–1999 and 2070–2099 (A1B scenario)

- PRISM data
  - 1405 $\times$ 621 grid, side 4km
  - Elevations
  - Mean annual precipitation 1970–1997

EXTREME VALUES METHODOLOGY

If Y_t is daily rainfall on day t, assume for large threshold exceedances,

$$\Pr\{Y_t > y\} \approx \delta t \left(1 + \xi_t \frac{y - \mu_t}{\psi_t}\right)_+^{-1/\xi_t},$$

where δt is time unit ($= \frac{1}{365.25}$) and μ_t, ψ_t, ξ_t are (time-localized) parameters of Generalized Extreme Value distribution.

If μ, ψ, ξ are constants, this implies distribution of annual maxima is of GEV form

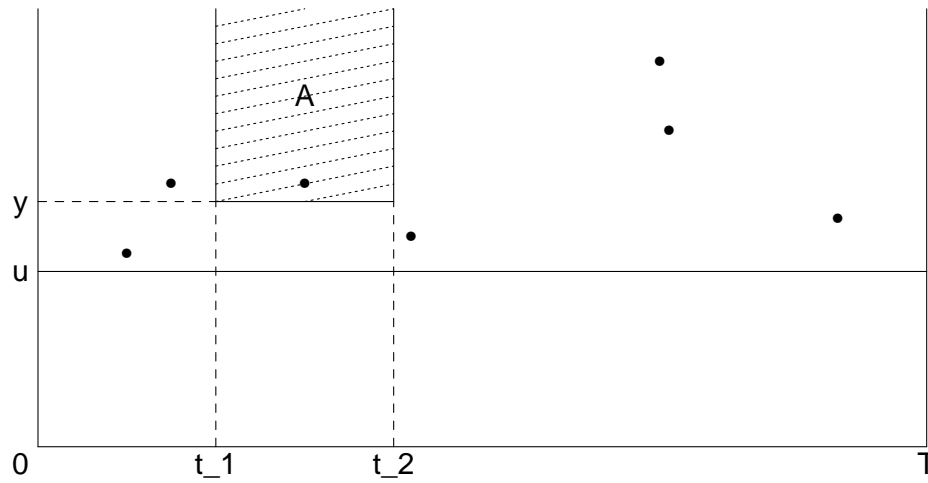$$\Pr\{Y_{\max} \leq y\} = \exp\left\{\left(1 + \xi \frac{y - \mu}{\psi}\right)_+^{-1/\xi}\right\}.$$

Illustration of threshold model

An aside: Some thoughts on the issue of "real-time attribution"

An analogy: A 55-year-old man has just died of a heart attack. It is known that he was a smoker. But he was also overweight, had high cholesterol, and a family history of heart disease. What proportion of his risk of heart attack may be "attributed" to smoking?

The traditional answer in biostatistics/epidemiology is *proportional hazard analysis* (D.R. Cox 1972): represent the hazard or risk function (i.e. infinitesimal probability of dying, conditional on not having died previously) as a function of each of the known risk factors. The function is developed by regression analysis over many individuals, but can be applied to a specific individual by inserting the individual values of smoking indicator, weight, cholesterol, etc. for that individual.

In this way it is possible to make sense of statements like "35% of this individual's risk of heart attack was due to smoking"

A possible analogy for risk of extreme events:

Suppose (for modern precipitation events) $\mu = 5$ cm., $\psi = 1$ cm., $\xi = 0.1$. We observe a rainfall level $y = 7$. We calculate the risk function as proportional to $\left(1 + \xi \frac{y-\mu}{\psi}\right)^{-1/\xi} = 0.162$.

However, from analysis of historical data we determine that in pre-industrial times μ was 3.5 (other parameters same). For that scenario, the risk of this extreme event was 0.050.

The ratio of risks is about 3.2. So we can say the risk of this particular event is about 3 times larger today than it was pre-global warming.

An extension

Suppose ENSO is also a risk factor. Specifically, suppose we perform a regression analysis of μ on ENSO and determine a regression coefficient of 0.8.

$$\mu \;=\; \mu_0 + 0.8 \times ENSO$$

where μ_0 is the base level (5 or 3.5)

Compare ENSO of $+1$ with ENSO of -1.

|	ENSO=+1	ENSO =−1
Pre-industrial	.085	.028
Present day	.322	.092

By analyzing log risks, we can argue that about 48% of the increased risk is due to greenhouse gases, the rest due to ENSO.

The actual analysis...

- Represent each of μ, $\log \psi$ and ξ as regression functions of seasonality and long-term trend (but not yet ENSO etc.)

- Perform this analysis for each station.

- Calculate an estimate of 25-year return values for each station, for 1970 and also for 1999. The look at ratios of return values.

- Spatially smooth across stations by kriging. This is carried out in 19 overlapping regions (overlapping to avoid boundary effects when we draw a map)
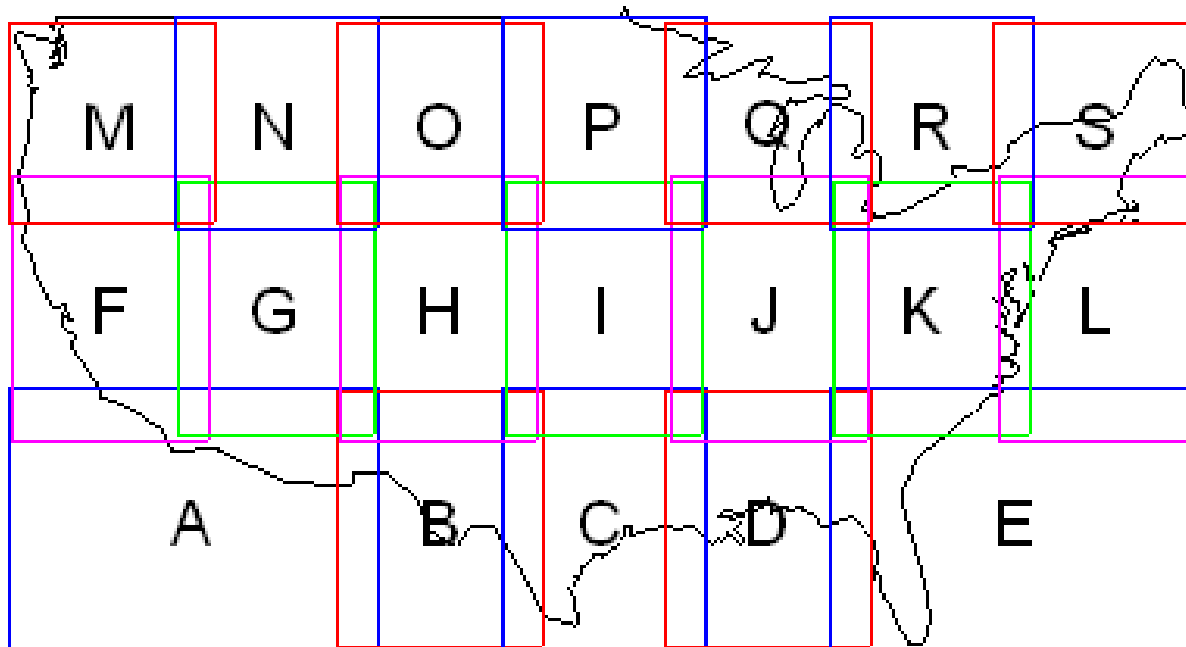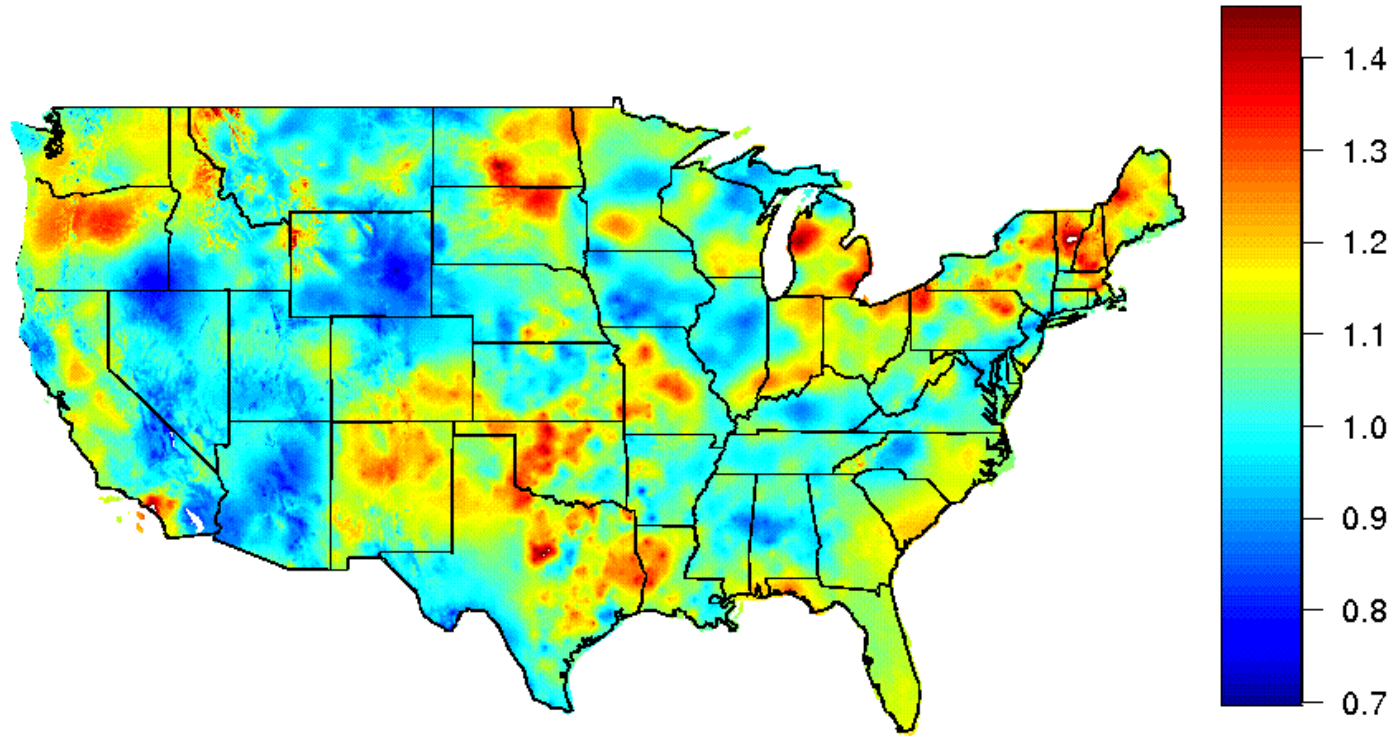
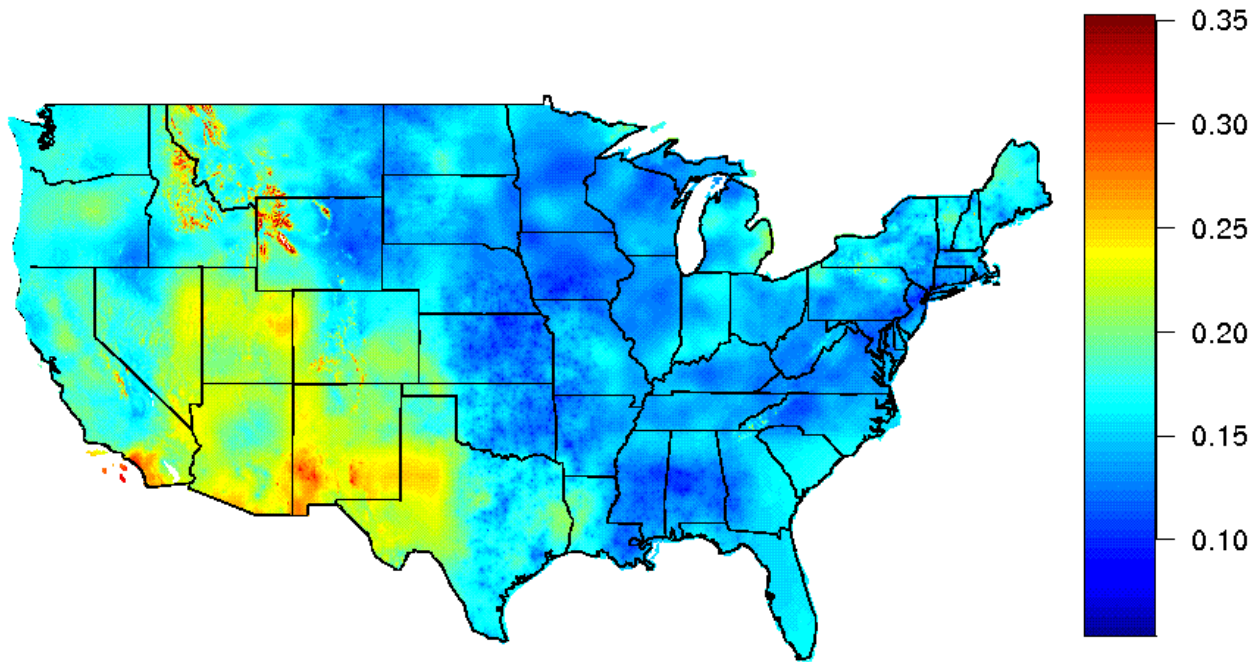Continental USA divided into 19 regions
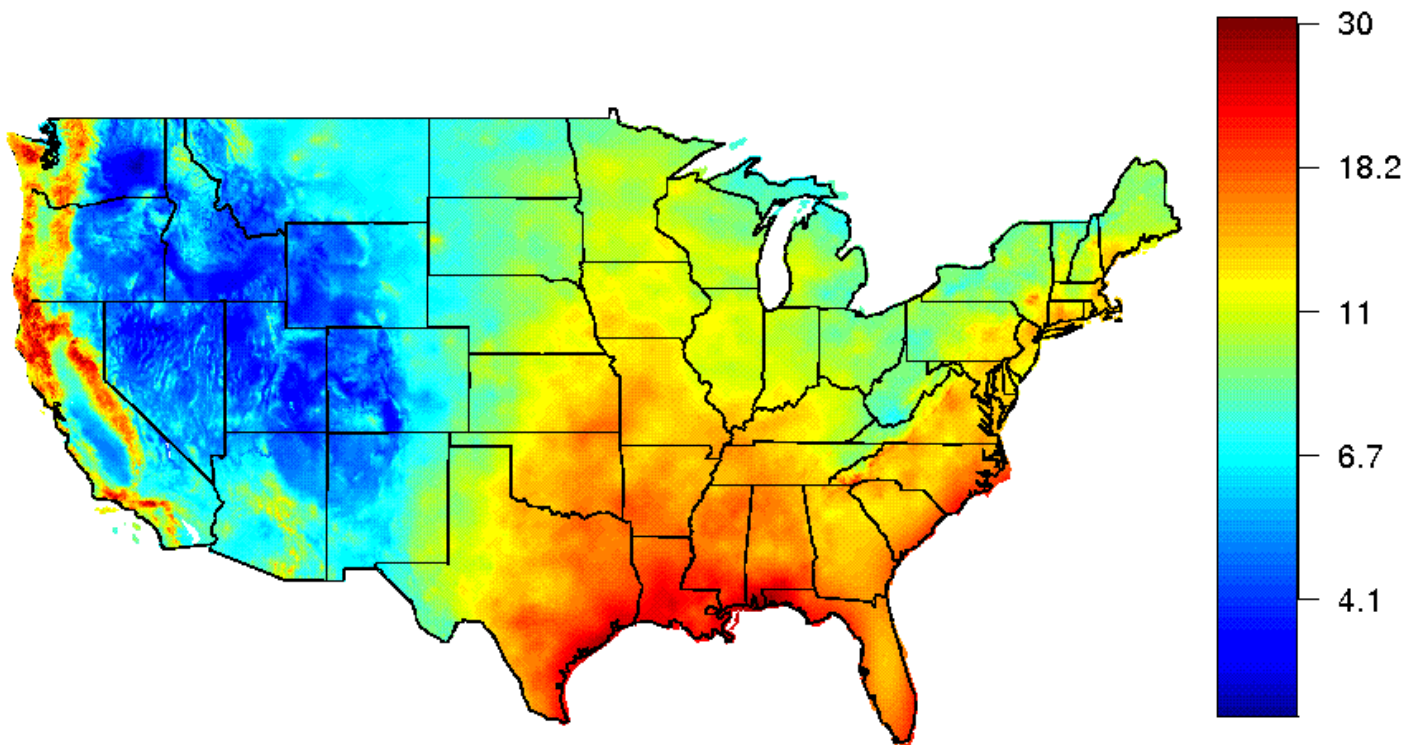
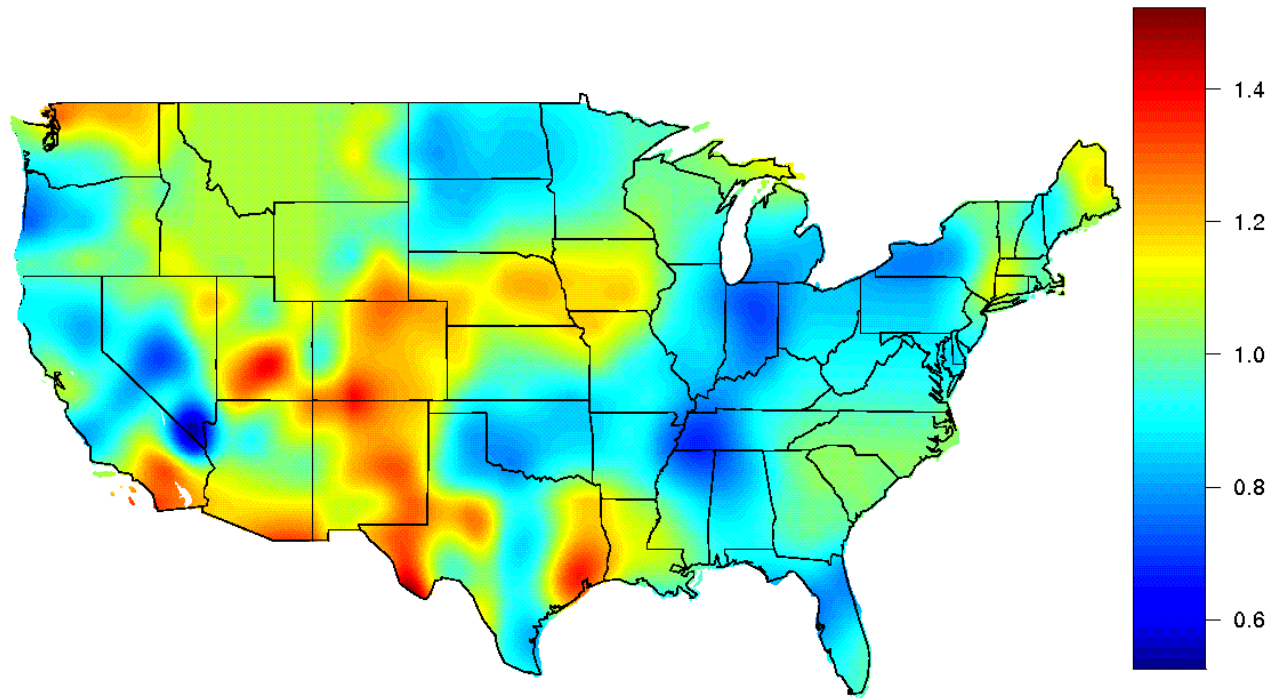Map of 25-year return values (cm.) for the years 1970–1999
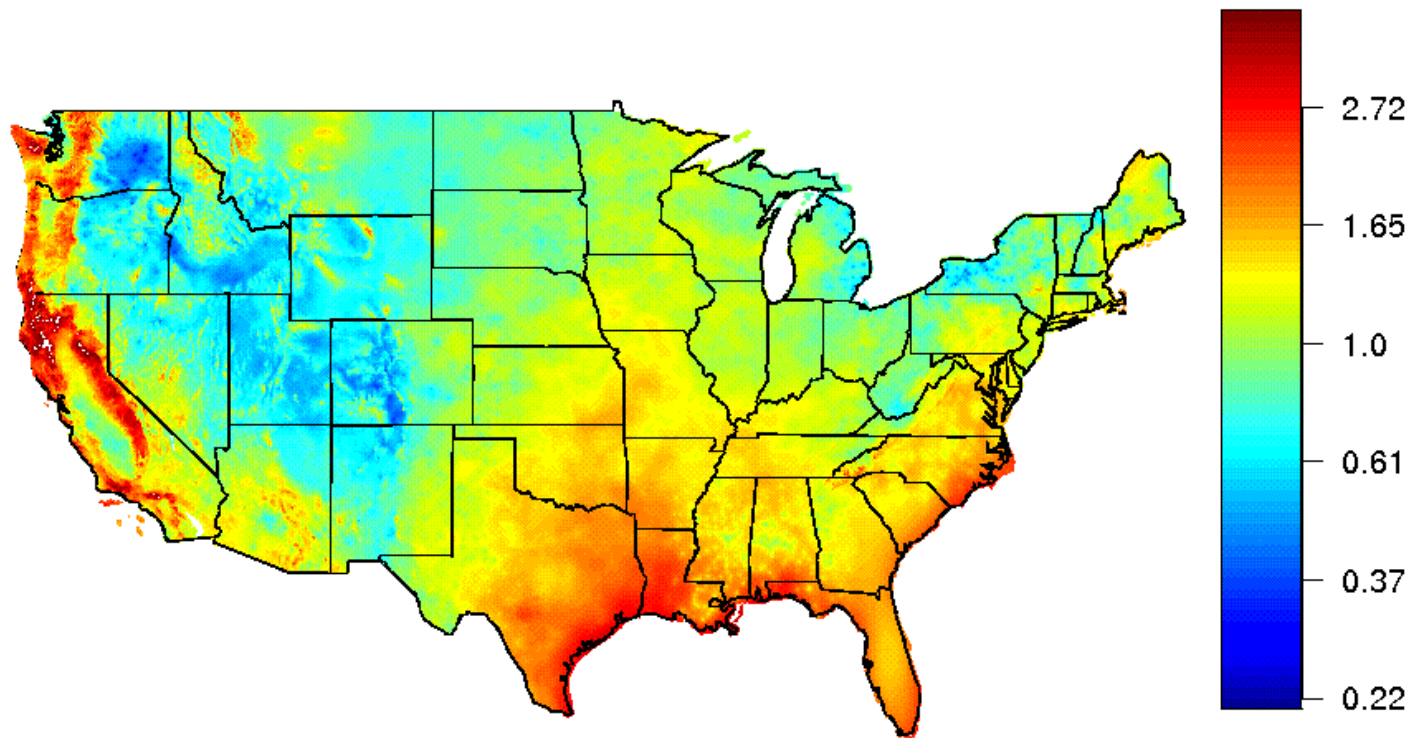
Root mean square prediction errors for map of 25-year return values for 1970–1999
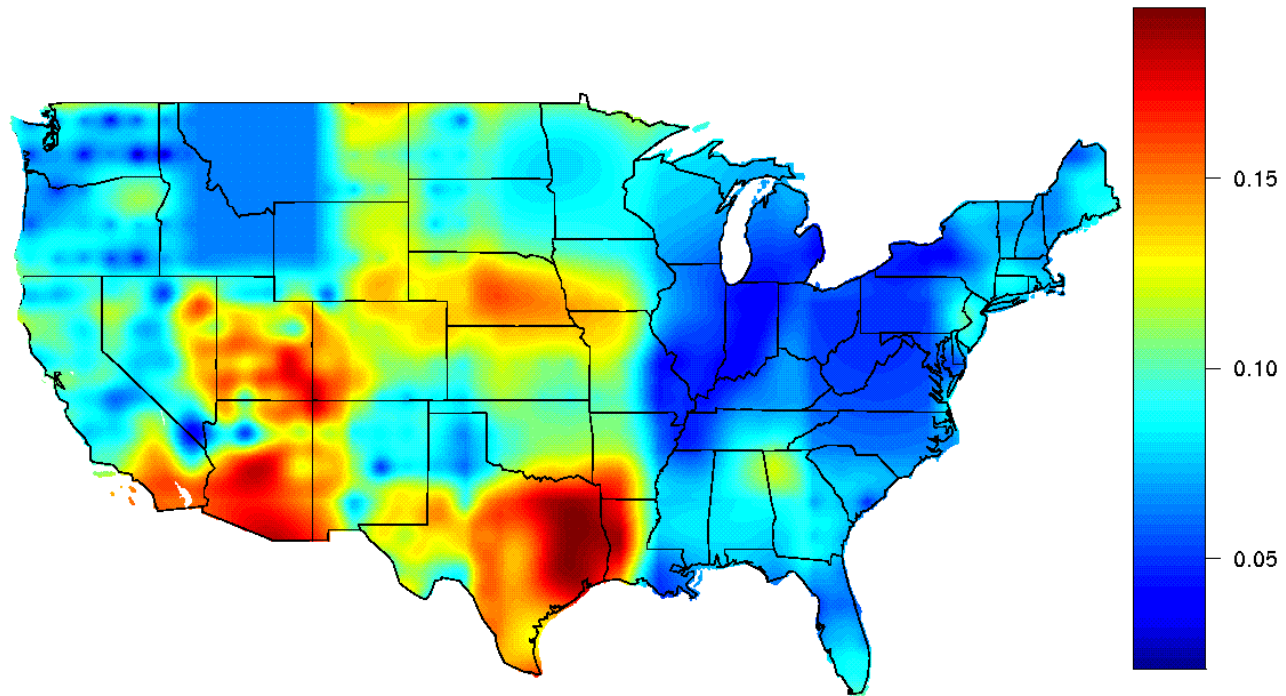
Ratios of return values in 1999 to those in 1970

Root mean square prediction errors for map of ratios of 25-year return values in 1999 to those in 1970

Extreme value model with trend: ratio of 25-year return value in 1999 to 25-year return value in 1970, based on CCSM data

RMSPE for map in previous slide

Region	Δ_1	SE_1	Δ_2	SE_2	Δ_3	SE_3	Δ_4	SE_4
A	−0.01	0.03	0.05**	0.05	0.16**	0.07	0.24**	0.10
B	0.07**	0.03	0.08***	0.04	0.14***	0.04	0.12***	0.06
C	0.11***	0.01	0.10	0.03	0.02	0.05	−0.14	0.11
D	0.05***	0.01	0.06	0.05	−0.06	0.04	−0.15	0.10
E	0.13***	0.02	0.14*	0.05	−0.07*	0.03	−0.09	0.08
F	0.00	0.02	0.05*	0.04	−0.07*	0.04	−0.03	0.05
G	−0.01	0.02	0.01	0.03	0.03	0.03	0.08*	0.04
H	0.08***	0.01	0.10***	0.03	0.11***	0.03	0.08	0.06
I	0.07***	0.01	0.12***	0.03	−0.02	0.04	−0.05	0.07
J	0.05***	0.01	0.08**	0.03	−0.15***	0.03	−0.16**	0.06
K	0.08***	0.01	0.09**	0.03	−0.08***	0.02	−0.11*	0.05
L	0.07***	0.02	0.07*	0.04	−0.04	0.04	−0.03	0.06
M	0.07***	0.02	0.10**	0.03	0.01	0.03	0.00	0.08
N	0.02	0.03	0.01	0.03	0.06**	0.02	0.05	0.06
O	0.01	0.02	0.02	0.03	−0.03	0.04	−0.06	0.07
P	0.07***	0.01	0.11***	0.03	−0.01	0.04	−0.07	0.07
Q	0.07***	0.01	0.11***	0.03	−0.04	0.04	−0.03	0.07
R	0.15***	0.02	0.13***	0.03	−0.17***	0.03	−0.06	0.08
S	0.14***	0.02	0.12*	0.06	0.00	0.04	0.02	0.05

Δ_1: Mean change in log 25-year return value (1970 to 1999) by kriging

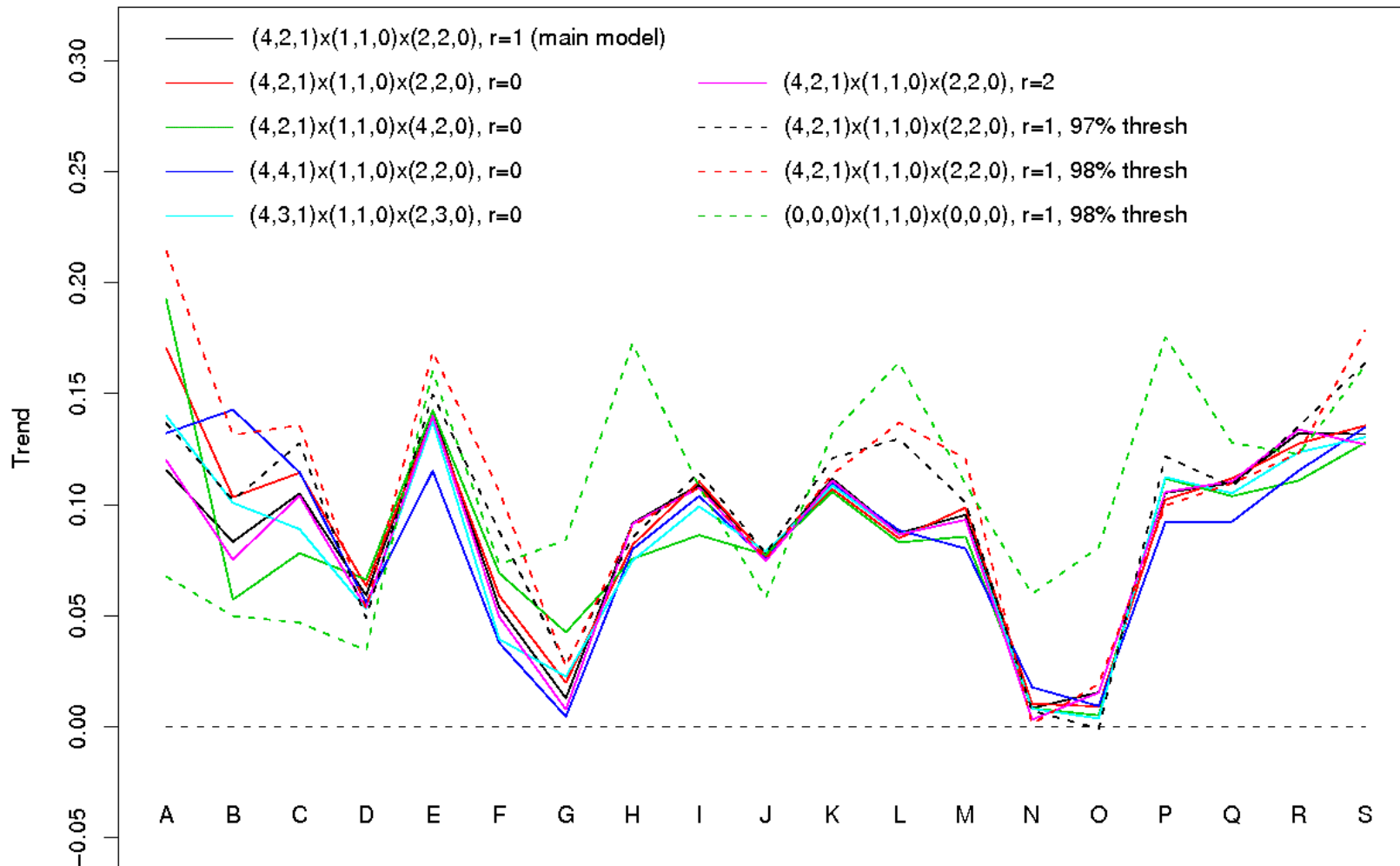SE_1: Corresponding standard error (or RMSPE)

Δ_2, SE_2: same but using geometrically weighted average (GWA)

Δ_3, SE_3: Results for CCSM using kriging

Δ_4, SE_4: Results for CCSM using GWA

Stars indicate significance at 5%*, 1%**, 0.1%***.

REGIONAL AVERAGE TRENDS FOR 9 EV MODELS (GWA METHOD)

Legend:
- (4,2,1)x(1,1,0)x(2,2,0), r=1 (main model)
- (4,2,1)x(1,1,0)x(2,2,0), r=0
- (4,2,1)x(1,1,0)x(4,2,0), r=0
- (4,4,1)x(1,1,0)x(2,2,0), r=0
- (4,3,1)x(1,1,0)x(2,3,0), r=0
- (4,2,1)x(1,1,0)x(2,2,0), r=2
- (4,2,1)x(1,1,0)x(2,2,0), r=1, 97% thresh
- (4,2,1)x(1,1,0)x(2,2,0), r=1, 98% thresh
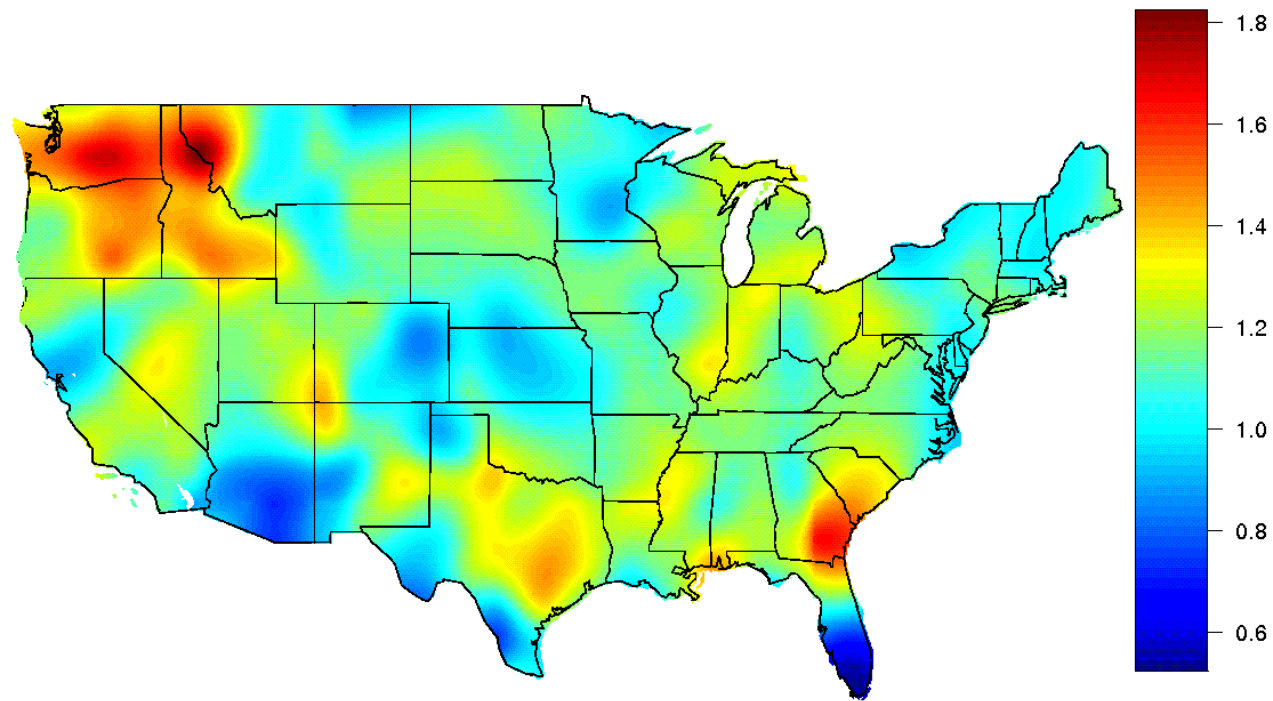- (0,0,0)x(1,1,0)x(0,0,0), r=1, 98% thresh

Trends across 19 regions (measured as change in log RV25) for 8 different seasonal models and one non-seasonal model with simple linear trends. Regional averaged trends by geometric weighted average approach.
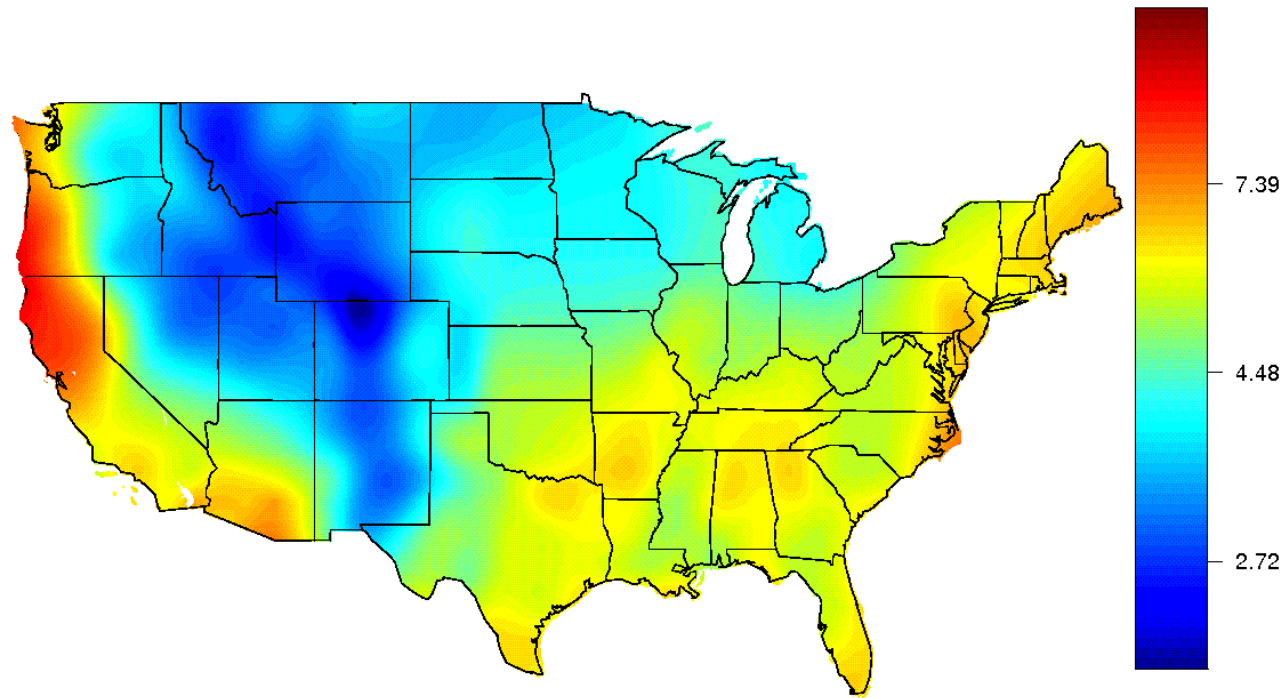
Return value map for CCSM data (cm.): 1970–1999
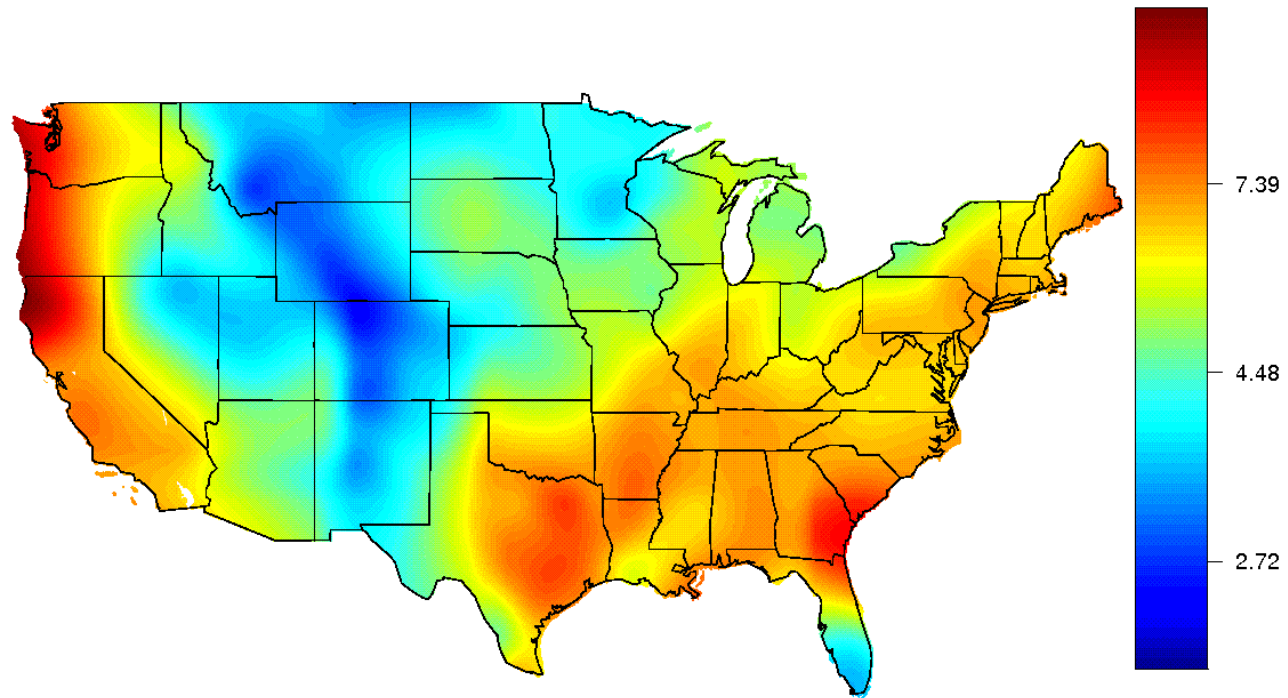
Return value map for CCSM data (cm.): 2070–2099

Estimated ratios of 25-year return values for 2070–2099 to those of 1970–1999, based on CCSM data, A2 scenario

CONCLUSIONS

1. Focus on N-year return values — strong historical tradition for this measure of extremes (we took $N = 25$ here)

2. Seasonal variation of extreme value parameters is a critical feature of this analysis

3. Overall significant increase over 1970–1999 except for parts of western states — average increase across continental US is 7%

4. Kriging better than weighting of "NCDC approach"

5. *But...* based on CCSM data there is a completely different spatial pattern and no overall increase

6. Projections to 2070–2099 show further strong increases but note caveat based on point 5

7. Decadal variations since 1950s show strongest increases during 1990s.