A Comparison Study of Extreme Precipitation from SixRegional Climate Models via Spatial Hierarchical Modeling

Dan Cooley
Department of Statistics
Colorado State University

Joint work with:
Erin Schliep, CSU
Steve Sain, NCAR

Acknowledgements:
Data provided by NARCCAP (NSF grants ATM-0502977 & ATM-0534173)
DC funded by NSF-DMS-0905315
ES supported by the Weather and Climate Impact Assessment Program

Extreme Value Analysis

EVA studies rare events from a *data* perspective.

- Fits asymptotically-justified distributions to characterize the tail.
- Often used for data that appear to have heavy tails.
- Uses only a subset of the data considered to be extreme.
  - block maximum approaches
  - threshold exceedance approaches
- Tail dependence described differently.
  - not via covariances/correlations.
  - one approach: angular (spectral) measure.

SAMSI Methodology Workshop:

Will be looking at 'rare events' from a few different perspectives.

RCM Project

Goal: To compare the extreme precipitation from six RCM's for North America.

Primary question: Are these RCM's telling the same story?

Relation to Impacts?

- A primary aim for developing RCM's is to model climate on a scale that is relevant for determining local impacts.
- Extreme precipitation events can have tremendous human and economic impact.

Audience for this work: Atmospheric scientists, particularly climate modelers, and statisticians.

Aim is similar to that of studies that have appeared in climate literature characterizing extremes produced by climate models, approach somewhat different.

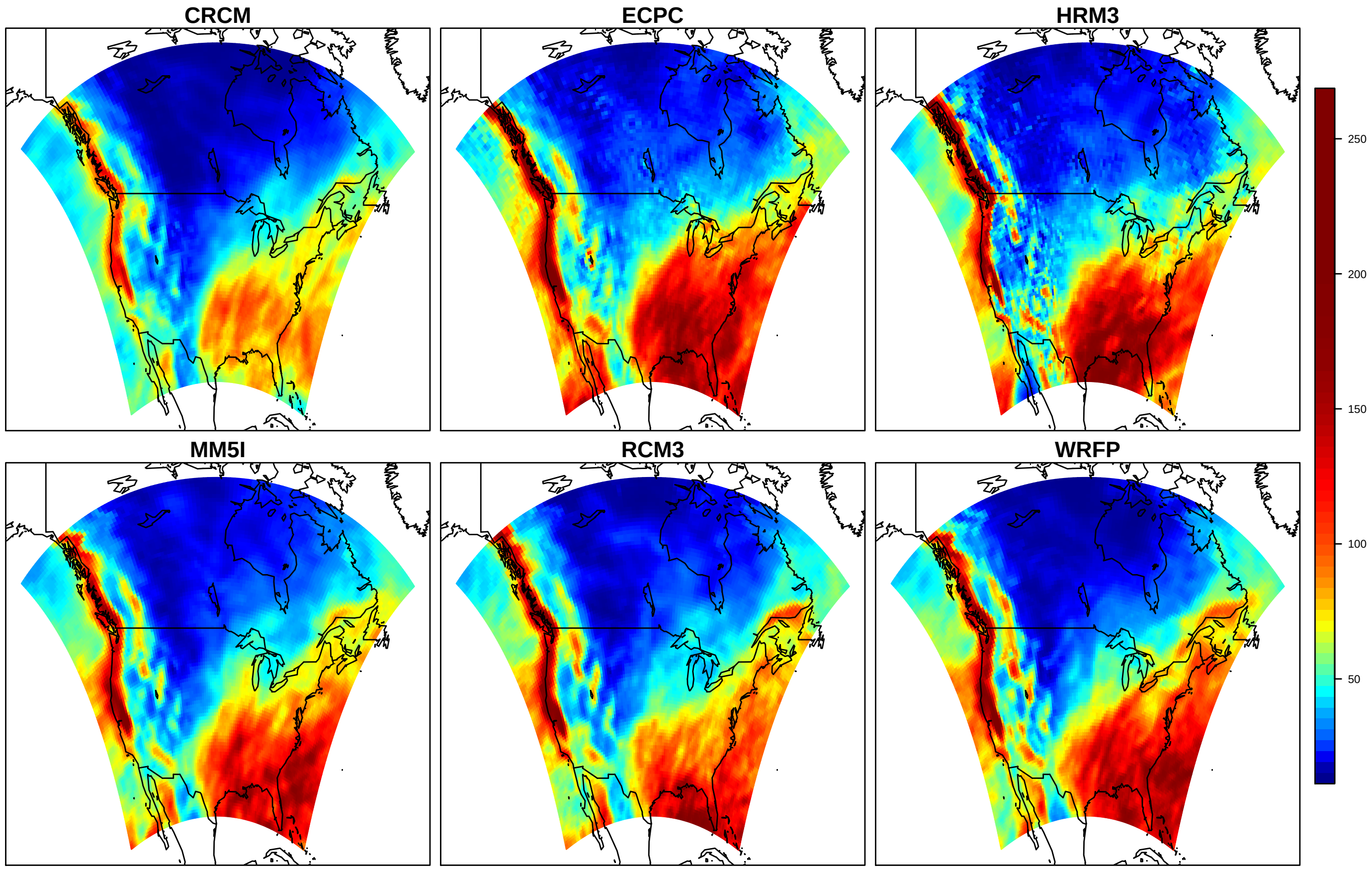
Overview of Methodology

We build a spatial hierarchical model for extreme data for each of the six RCM's.

- We separately model winter and summer seasons, and employ the annual maximum model output from each season.
- Spatial hierarchical approach allows us to borrow strength across locations when estimating parameters.
- Bayesian formulation allows for straightforward uncertainty quantification.

Preview of Results: Winter 100 Year Return Level

Interested in characterizing *marginal* behavior of extremes across study region.

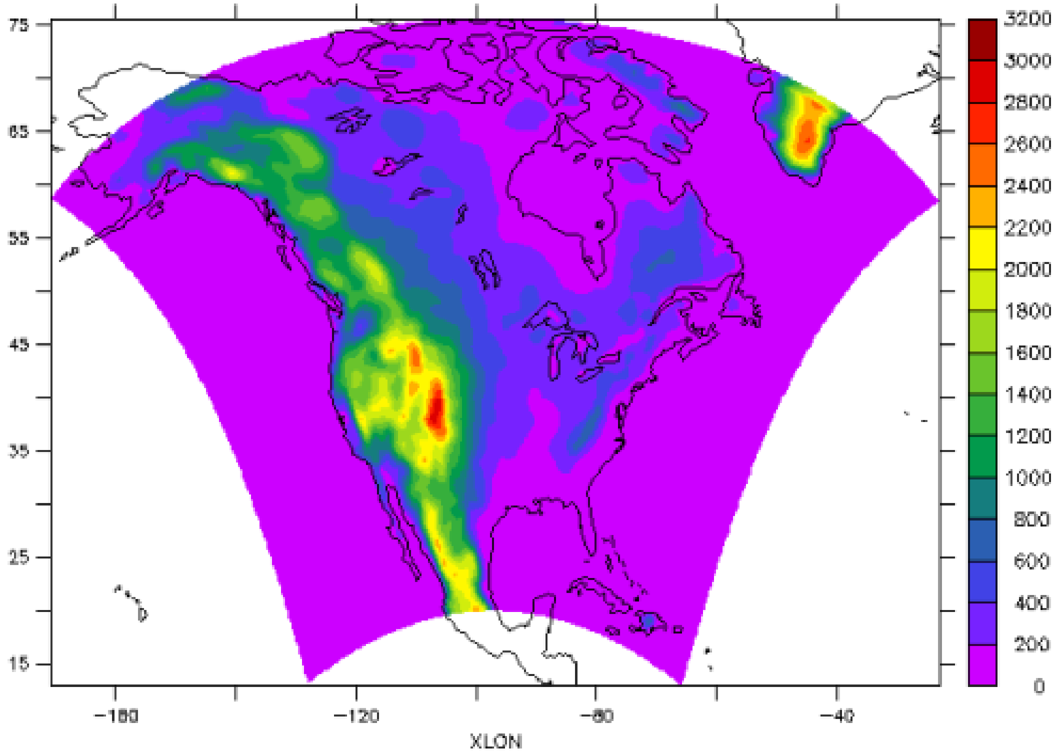


Outline

- NARCCAP and Model Output
- Background
  - Extreme Value Theory and Statistical Practice
  - Spatial Hierarchical Models
- Statistical Model for RCM Output
- Results
  - Winter Season
  - Summer Season
- UQ Program and Rare Events

NARCCAP

NARCCAP is a program which is producing a suite of high-resolution climate model runs for North America.



NARCCAP

NARCCAP is a program which is producing a suite of high-resolution climate model runs for North America.

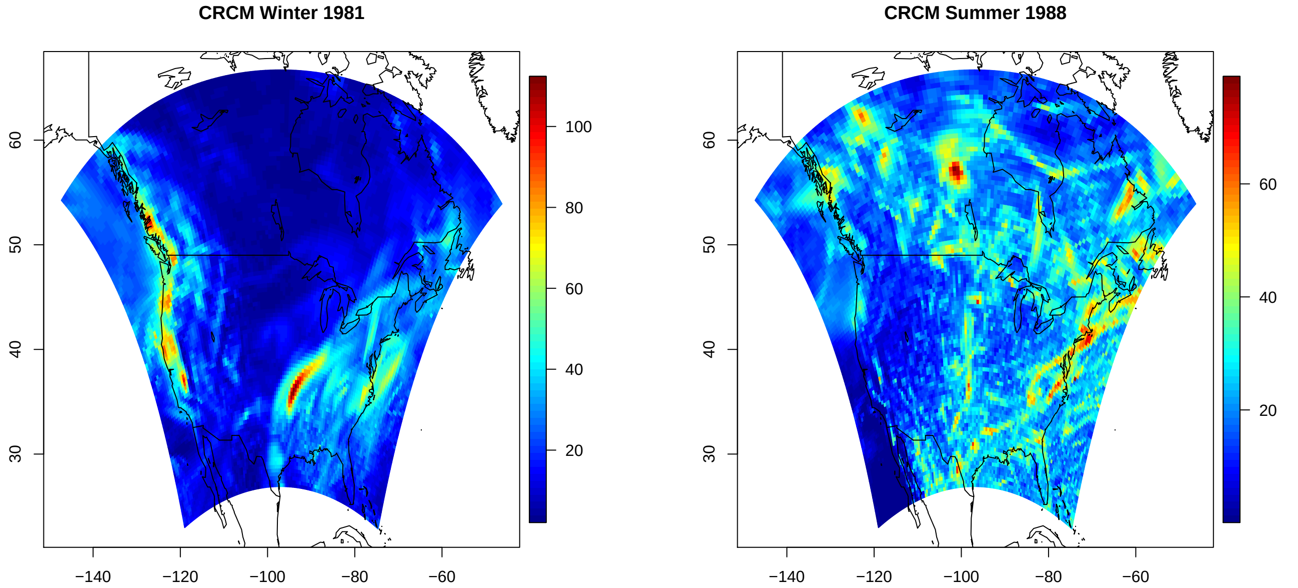
Abbr.	Model Name	Modeling Group
CRCM	Canadian Regional Climate Model	OURANOS / UQAM
ECPC	Experimental Climate Prediction Center	UC San Diego / Scripps
HRM3	Hadley Regional Model 3	Hadley Centre
MM5I	MM5 - PSU/NCAR mesoscale model	Iowa State University
RCM3	Regional Climate Model version 3	UC Santa Cruz
WRFP	Weather Research & Forecasting model	Pacific Northwest Nat'l Lab

---

Phase I : RCM's driven by NCEP reanalysis data.
Phase II: RCM's driven by a suite of AOGCM's.

Data: Annual Maxima

For each season (winter: DJF; summer: JJA), we fit a statistical model to the annual maxima for that season.



For each RCM, we have 20 fields of annual maxima for each season (1981-2002): data (model output) are spatially rich, temporally poor. Fields are $120 \times 98 = 11760$.

Statistics for Extremes

Extreme value theory provides distributions with which to model the upper tail. *It is not necessary to know the data's original distribution.*

Distributions are fit using only data considered to be extreme.

1. GEV(μ, σ, ξ): for block maximum data.
2. GPD($\tilde{\sigma}_u$, ξ, ζ): for exceedances over a threshold u.
3. PPE(μ, σ, ξ): for exceedances.

- ξ determines the tail behavior.
- In all cases, ξ is particularly difficult to estimate.
- Always data poor when studying extremes.
- Risk is often summarized in terms of return levels; basically a high quantile.

Spatial HM General Framework

Basic idea: Assume there is a latent spatial process that characterizes the behavior of the data over the study region.

Why bother? Latent process too complex to capture with fixed effects; covariates not rich enough.

Bayesian formulation, three levels.

Data level: Likelihood which characterizes the distribution of the observed data *given the parameters at the process level*. Often there is an assumption of *conditional independence*.

Process level: Where the latent process gets modeled by assuming a spatial model *for the data level parameters*.

Prior level: Ties up loose ends. Uses apriori information to put prior distributions on the parameters introduced in the process level.

Goals for Statistical Model for RCM Output

Develop a model which:

- spatially models extreme precipitation
- *sensibly pools data (trading space for time)* – different than some previous studies of extremes of climate model output
- could be viewed as an alternative to RFA
- can handle the large data set (11760 grid cells)
- gives uncertainty estimates for parameters and quantities of interest

Data level: GEV-based

Let Z_{ijt} be the max precip from RCM i, grid cell j, year t.

We assume

$$\mathbb{P}(Z_{ijt} \leq z) = \exp\left[-\left(1 + \xi_{ij}\frac{z - \mu_{ij}}{\sigma_{ij}}\right)^{-1/\xi_{ij}}\right],$$

and further *assume conditional independence.*

To stabilize ξ, we add a penalty (Martins & Stedinger 2000).

Our data level is comprised of the likelihood

$$\pi[\boldsymbol{z}_i|\boldsymbol{\mu}_i, \boldsymbol{\sigma}_i, \boldsymbol{\xi}_i] \;=\; K\prod_{j=1}^{d}\prod_{t=1}^{20}\exp\left\{-\left[1 + \xi_{ij}\left(\frac{z_{ijt} - \mu_{ij}}{\sigma_{ij}}\right)\right]^{-1/\xi_{ij}}\right\}$$
$$\times\frac{1}{\sigma_{ij}}\left[1 + \xi_{ij}\left(\frac{z_{ijt} - \mu_{ij}}{\sigma_{ij}}\right)\right]^{-1/\xi_{ij}-1}\frac{\Gamma(15)}{\Gamma(9)\Gamma(6)}(.5 + \xi_{ij})^{8}(.5 - \xi_{ij})^{5}.$$

Process level

We assume

$$\begin{aligned}
\mu_{ij} &\sim N(\boldsymbol{X}_j^T\boldsymbol{\beta}_{i\mu} + U_{ij\mu}, 1/\tau_\mu^2) \\
\log(\sigma_{ij}) &\sim N(\boldsymbol{X}_j^T\boldsymbol{\beta}_{i\sigma} + U_{ij\sigma}, 1/\tau_\sigma^2) \\
\xi_{ij} &\sim N(\boldsymbol{X}_j^T\boldsymbol{\beta}_{i\xi} + U_{ij\xi}, 1/\tau_\xi^2),
\end{aligned}$$

where $\tau_{\cdot}$ is a fixed precision.

Spatial model for $\boldsymbol{U}_i = (\boldsymbol{U}_{i\mu}, \boldsymbol{U}_{i\sigma}, \boldsymbol{U}_{i\xi})$: IAR, an improper GMRF.

- $\boldsymbol{U}_i$ has length $3 \times 11760 = 35280$.
- IAR defined by precision matrix Q. We assume $Q = T \otimes Q_1$, T : 3×3, Q_1 : 11760×11760; Q_1 based on a 1st-order neighborhood structure, very sparse.
- IAR is a simple, computationally-feasible spatial model that enables borrowing strength across locations.

Prior level

Conjugate priors:

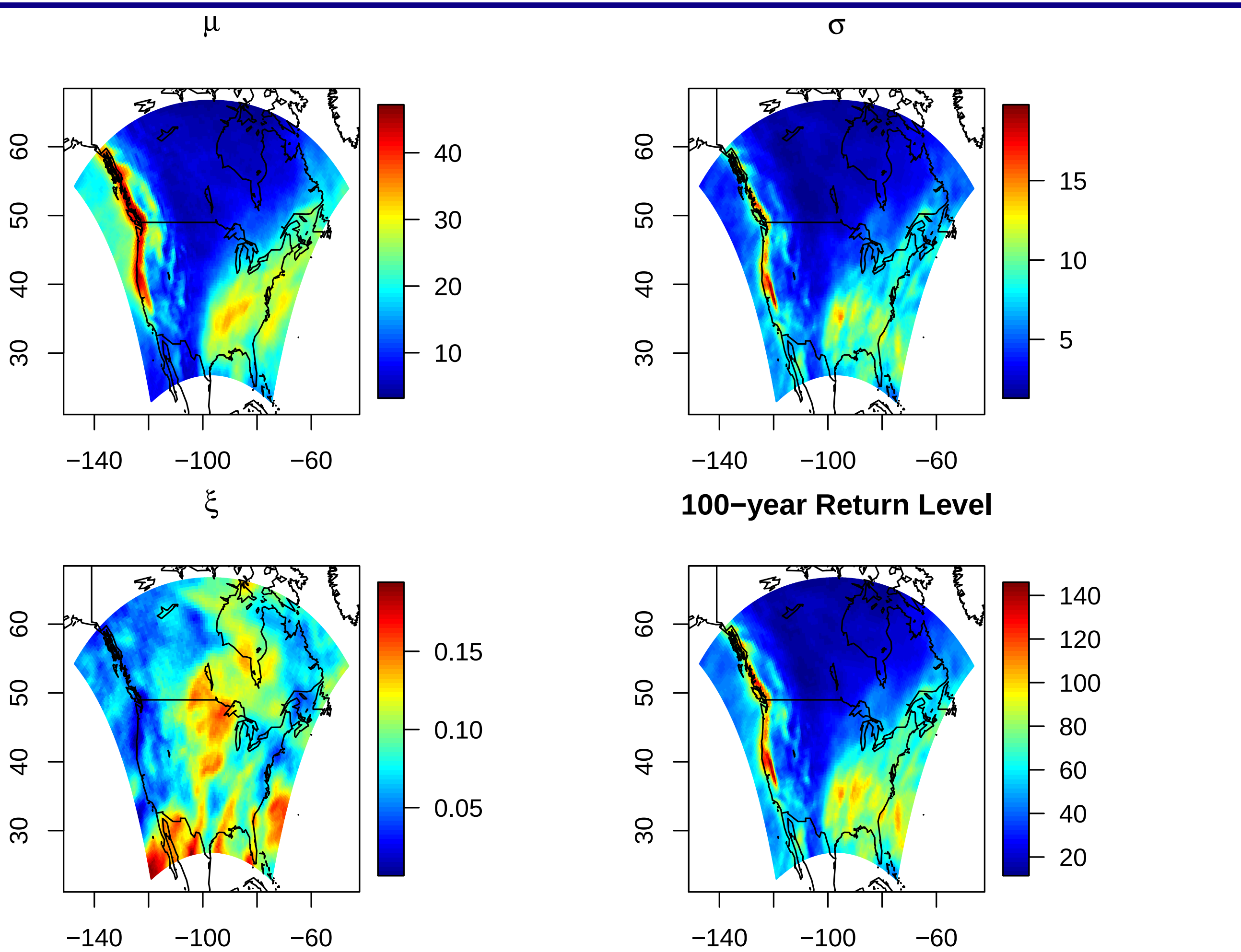
$$T \sim \text{Wishart prior}$$

$$\beta \sim \text{independent normal priors}$$

Implementation

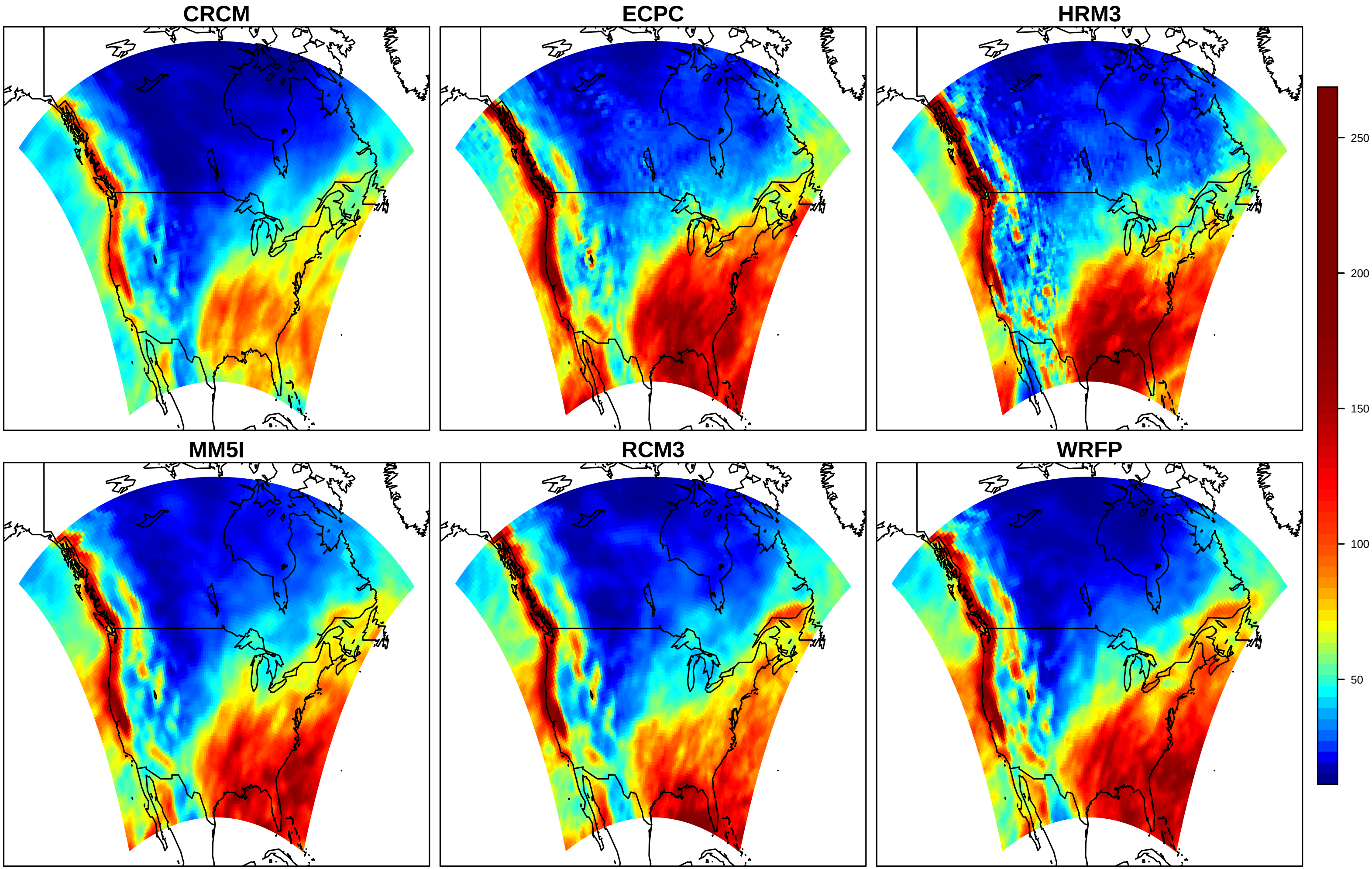
- 70569 non-indep parameters; effective number $\approx$ 10100
- MCMC via a Gibbs sampler
- $(\mu_{r,i}, \sigma_{r,i}, \xi_{r,i})$ updated cell-by-cell via Metropolis Hastings.
- All other parameters drawn directly.
- Take advantage of separability of Q and sparseness of Q_1.
- MCMC run of 6,000 iterations takes approx. 12 hrs.
- Four parallel chains for each RCM–convergence assessed.

Winter Parameter Estimates: CRCM

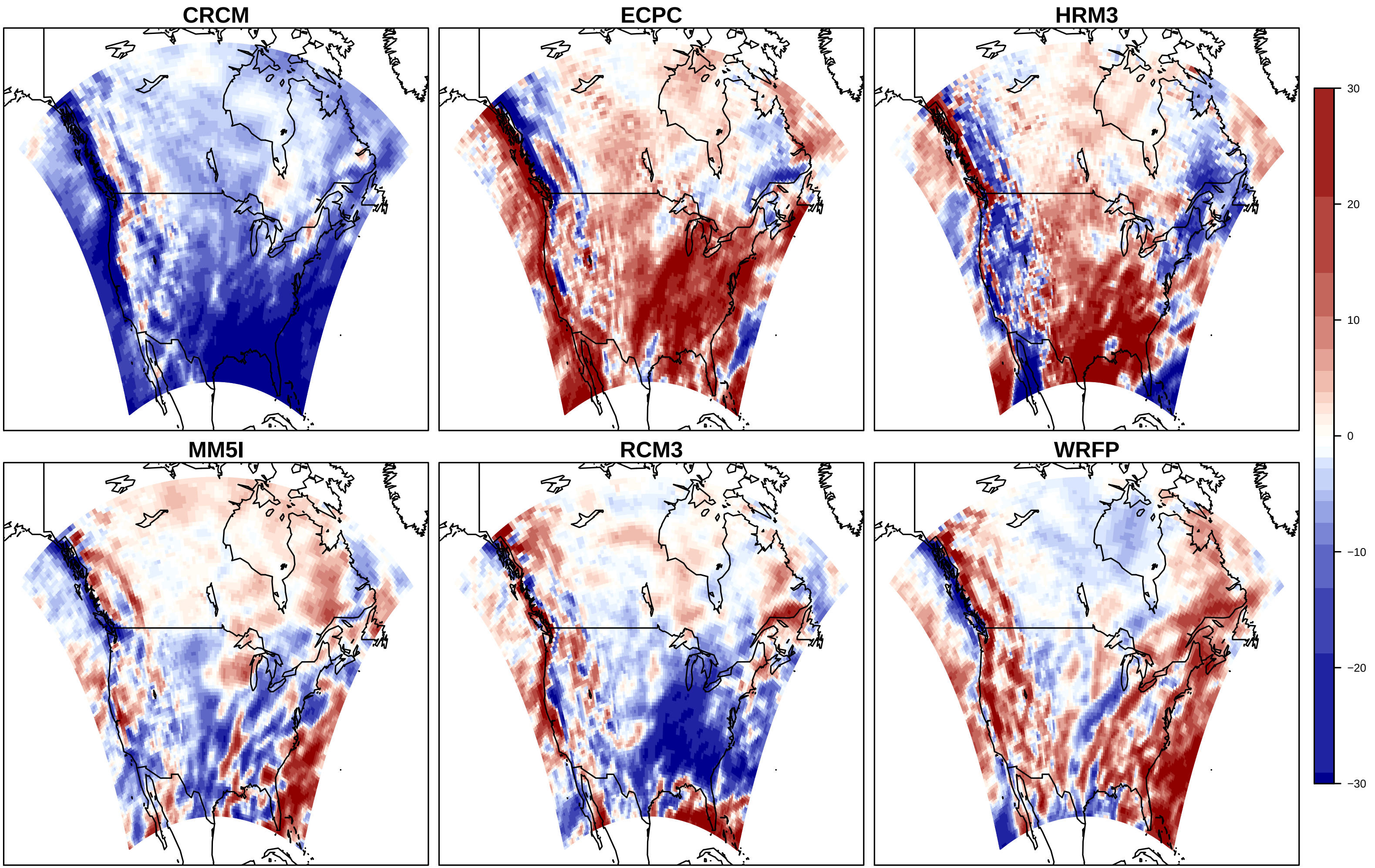


Note: estimation of *any* high quantile is straightforward.

Comparison of Winter 100-year Return Levels

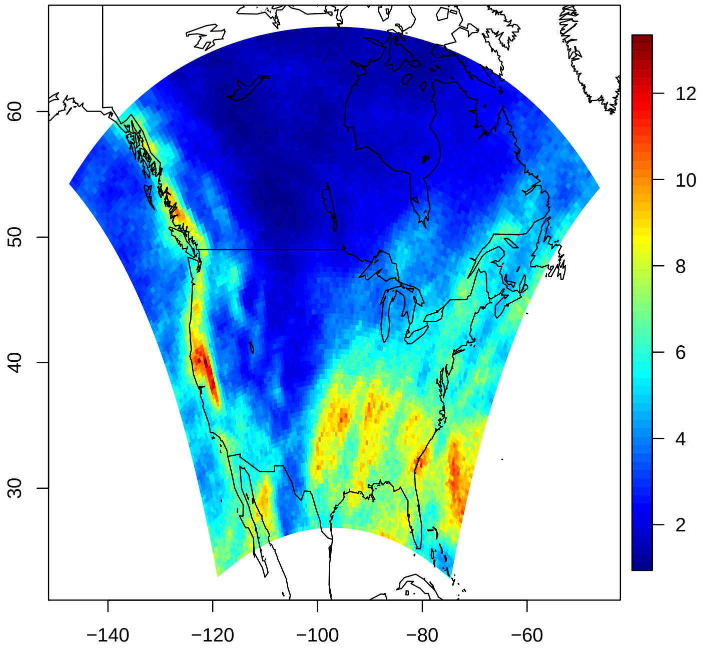


Comparison of Winter 100-year Return Levels

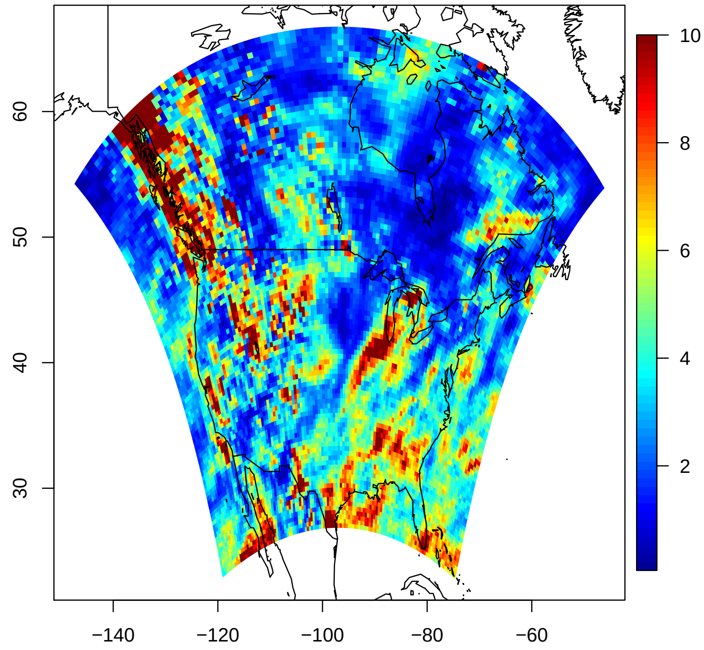


Uncertainty/Significance (Winter)

Std Dev 100 RL CRCM Model

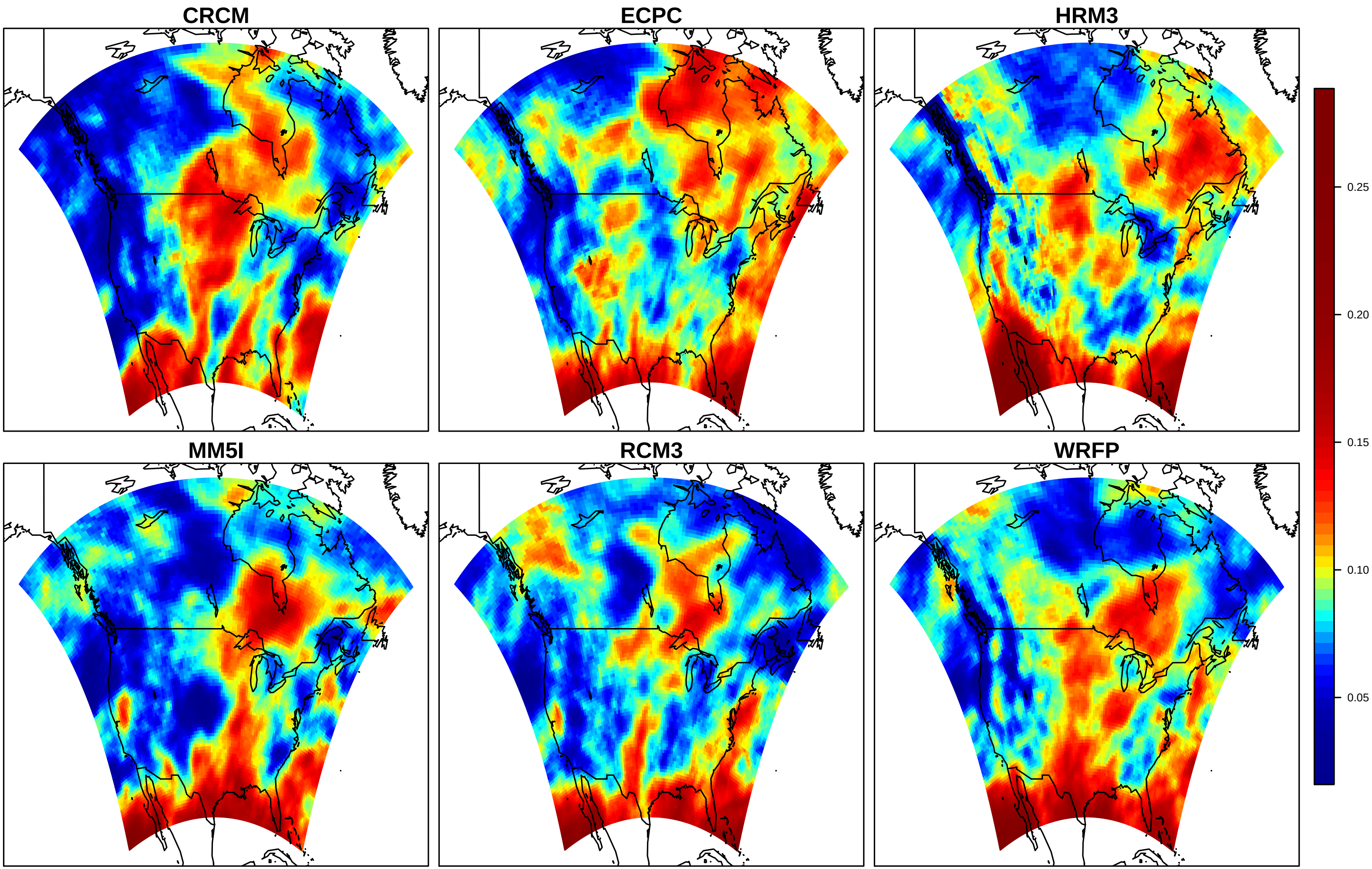


F-statistic

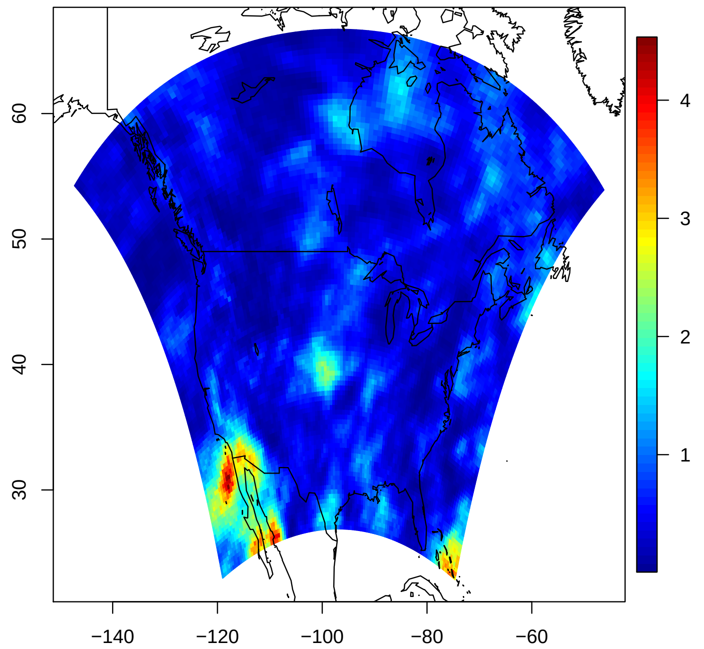


F-test for equality of means, significance level: 2.22
(disregards spatial dependence and multiple testing issues)

Examining ξ (Winter)



Significance for ξ (Winter)



F-test for equality of means, significance level: 2.22
Differences do not appear to be significant due to the uncertainty in estimating ξ.

Continued....

See part 2

References

Cooley, D. and Sain, S. R. (2010). Spatial hierarchical modeling of precipiation extremes from a regional climate model. *Journal of Agricultural, Biological, and Envrionmental Statistics*, 15:381–402.

Martins, E. and Stedinger, J. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36:737–744.

Schliep, E., Cooley, D., Sain, S. R., and Hoeting, J. A. (2010). A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling. *Extremes*, 13:219–239.