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

Dan Cooley
Department of Statistics
Colorado State University

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; SS funded by DMS-0707069; NCAR is supported by the NSF.

Introductory Remarks

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.

NARCCAP

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

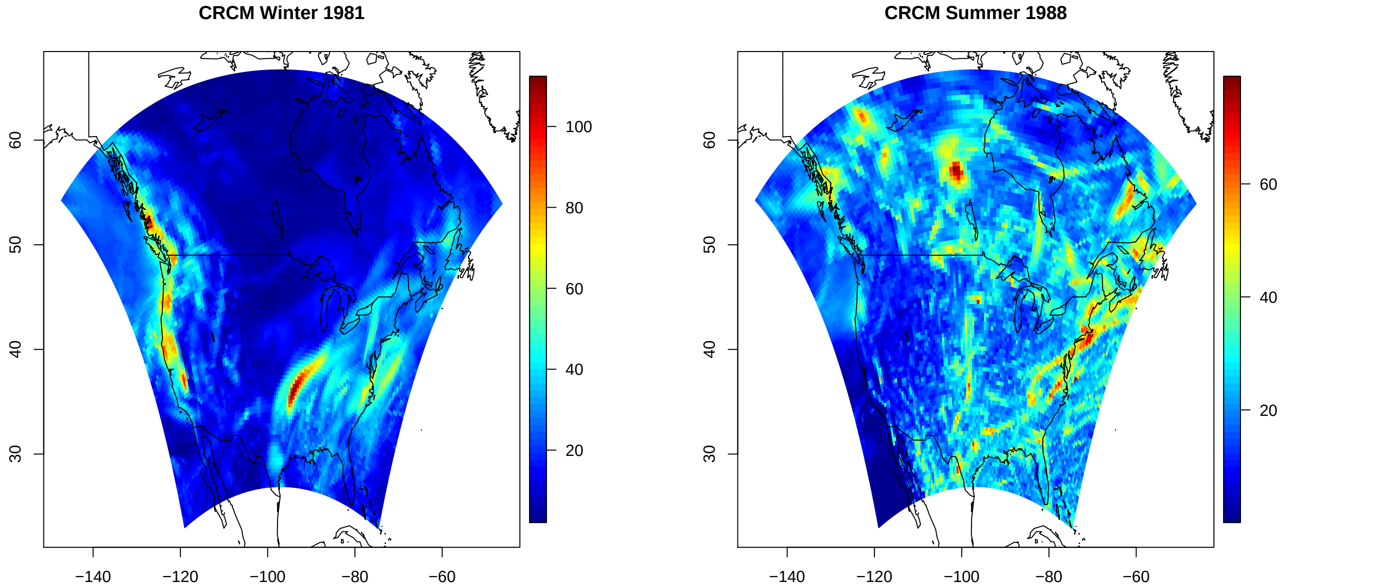
Abbr.	Model Name	Modeling Group
CRCM	Canadian Regional Climate Model	OURANOS / UQAM
ECPC	Experimental Climate Prediction Center	UC San Diego / Scripps
HRM3	Hadley Regional Model 3	Hadley Centre
MM5I	MM5 - PSU/NCAR mesoscale model	Iowa State University
RCM3	Regional Climate Model version 3	UC Santa Cruz
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 $120 \times 98 = 11760$.

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.

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 $Q = T \otimes Q_1$, $T : 3 \times 3$, $Q_1 : 11760 \times 11760$; 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:

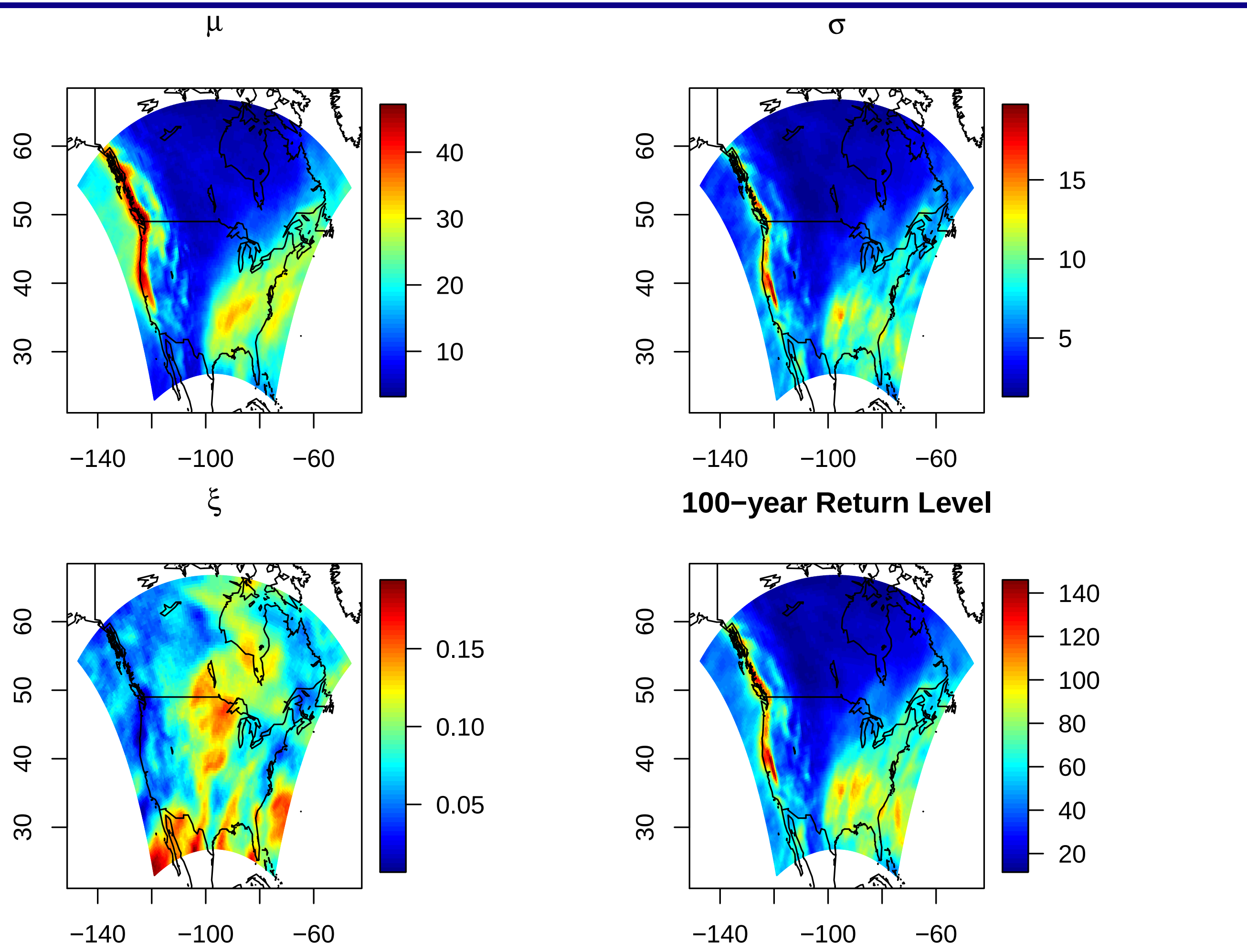
$$T \sim \text{Wishart prior}$$

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

Implementation

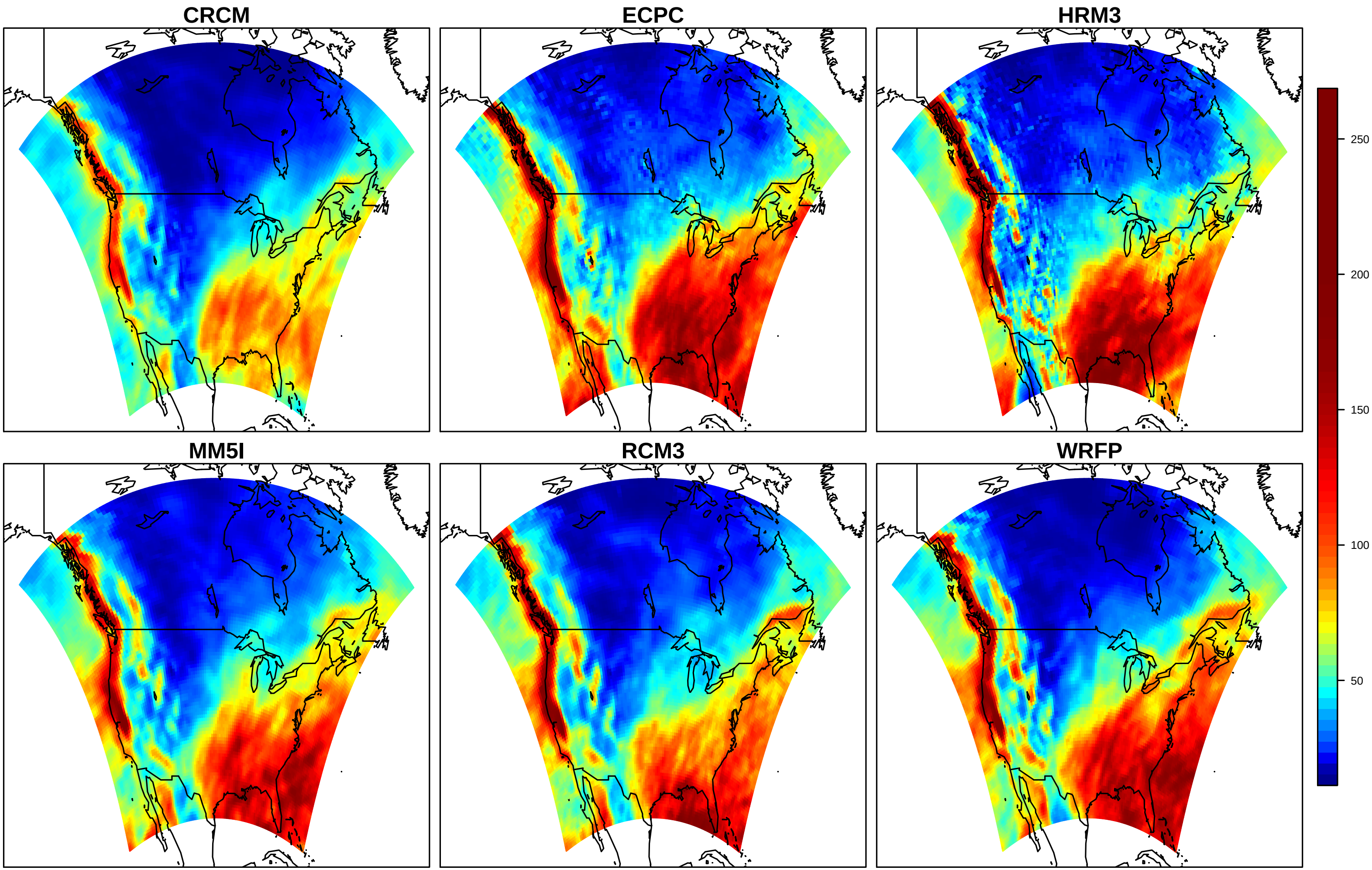
- 70569 non-indep parameters; effective number $\approx$ 10100
- MCMC via a Gibbs sampler
- $(\mu_{r,i}, \sigma_{r,i}, \xi_{r,i})$ updated cell-by-cell via Metropolis Hastings.
- All other parameters drawn directly.
- Take advantage of separability of Q and sparseness of Q_1.
- MCMC run of 6,000 iterations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

Winter Parameter Estimates: CRCM

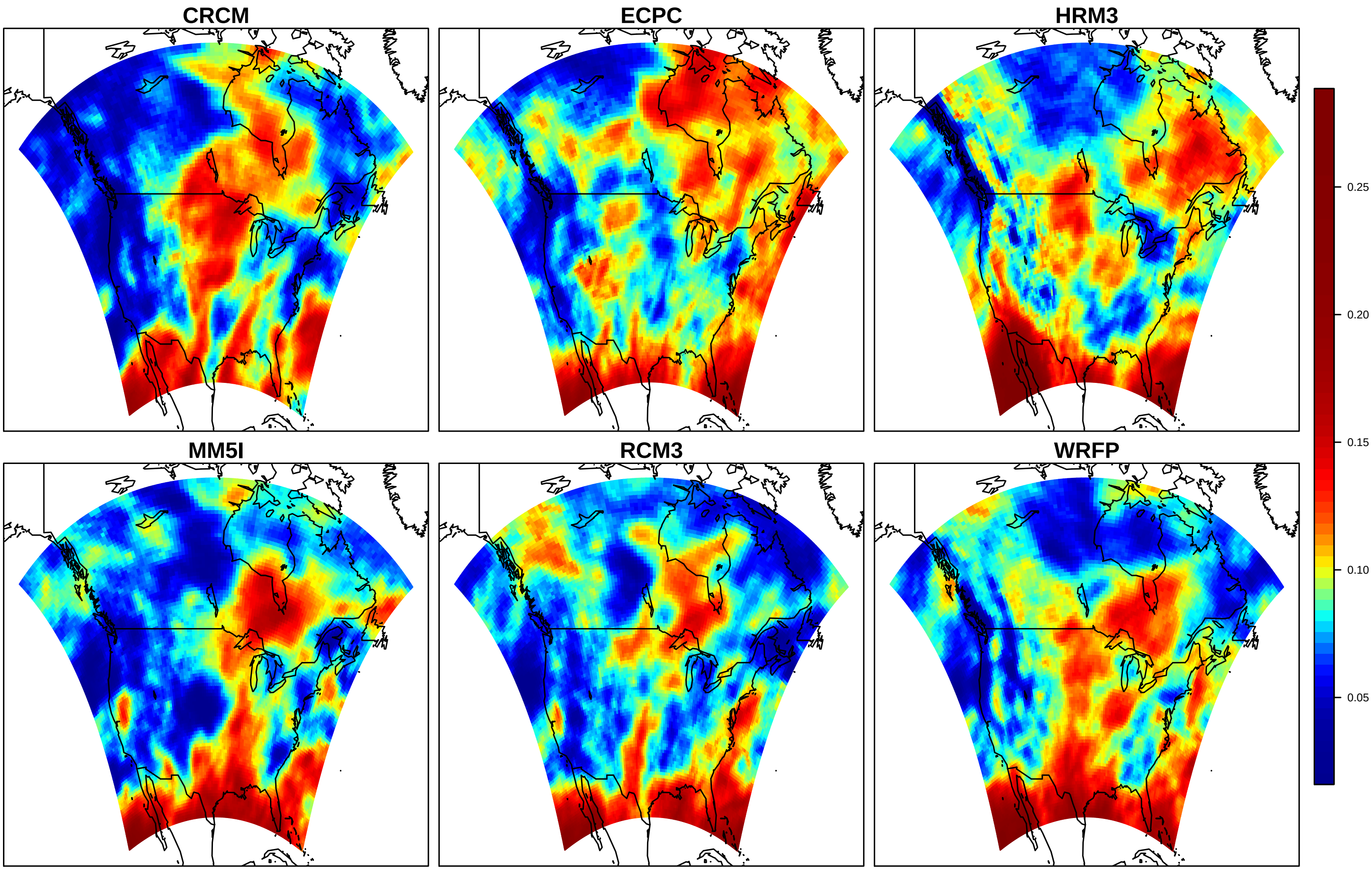


Note: estimation of *any* high quantile is straightforward.

Comparison of Winter 100-year Return Levels

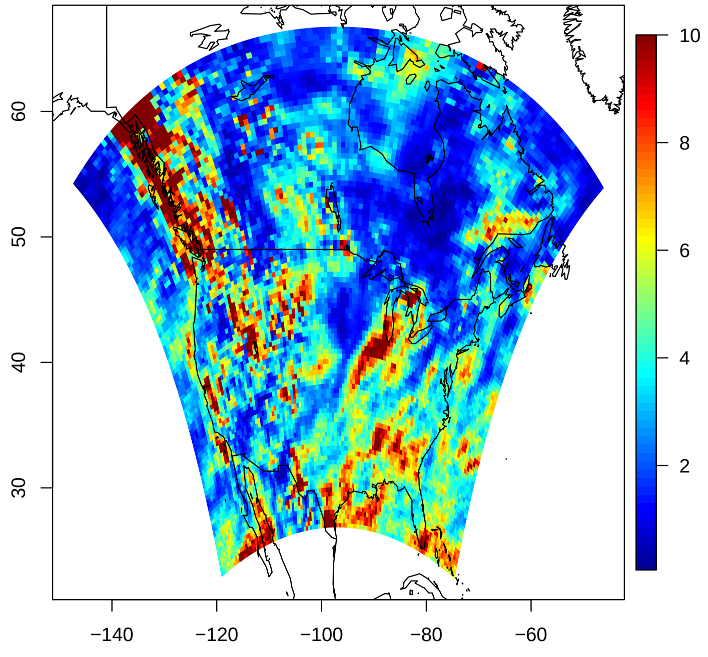


Examining ξ (Winter)

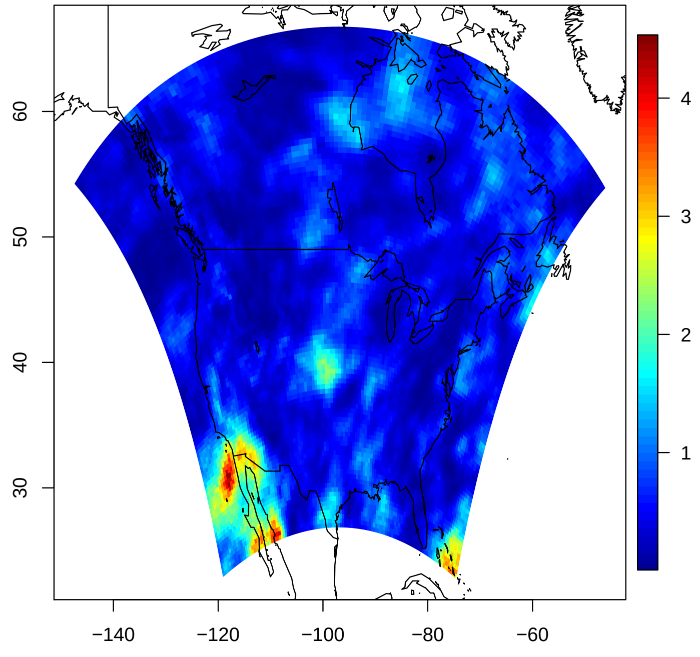


Significance (Winter)

Std Dev 100 RL CRCM Model

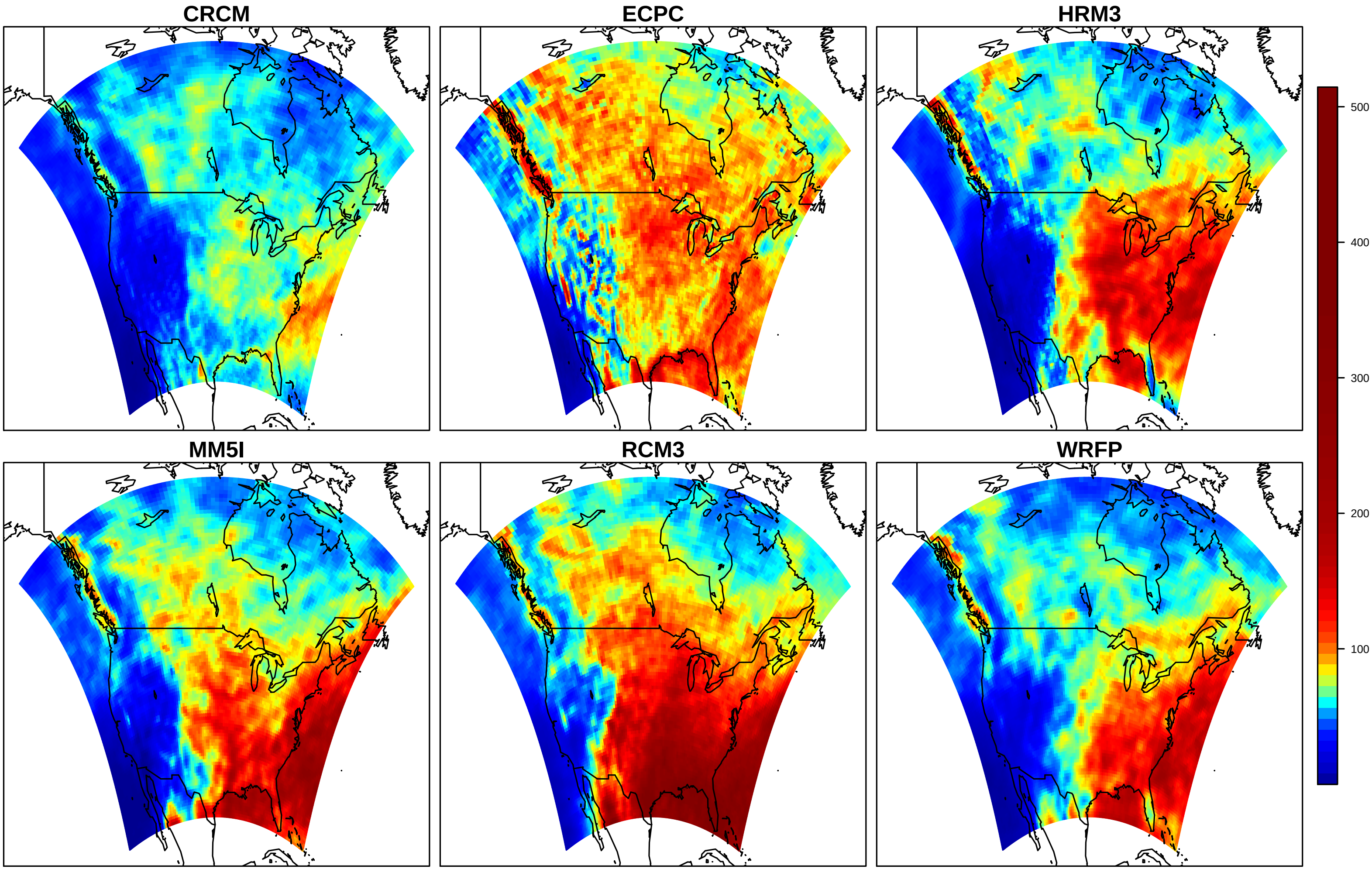


F-statistic

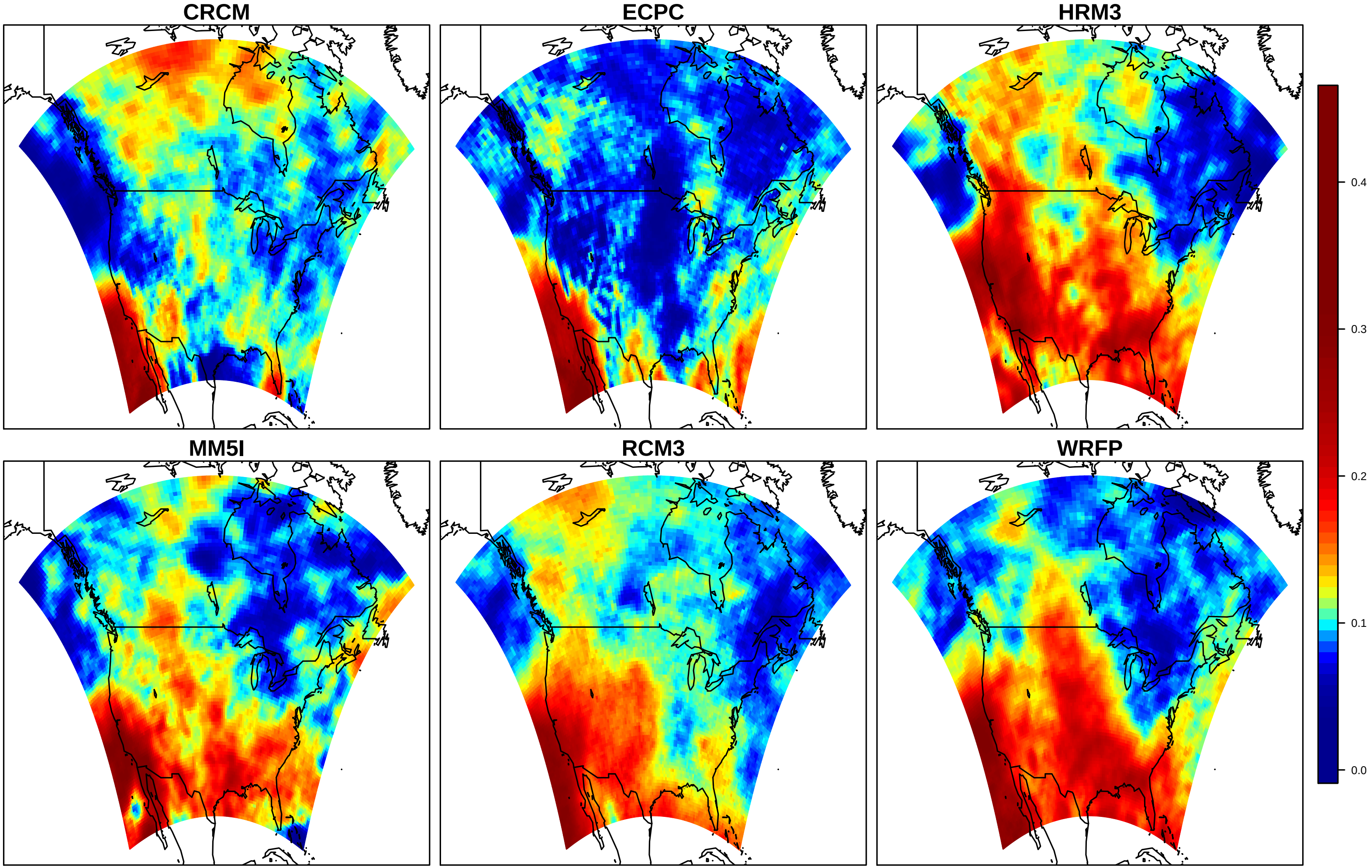


F-test for equality of means, significance level: 2.22
(disregards spatial dependence and multiple testing issues)

Comparison of Summer 100-year Return Levels

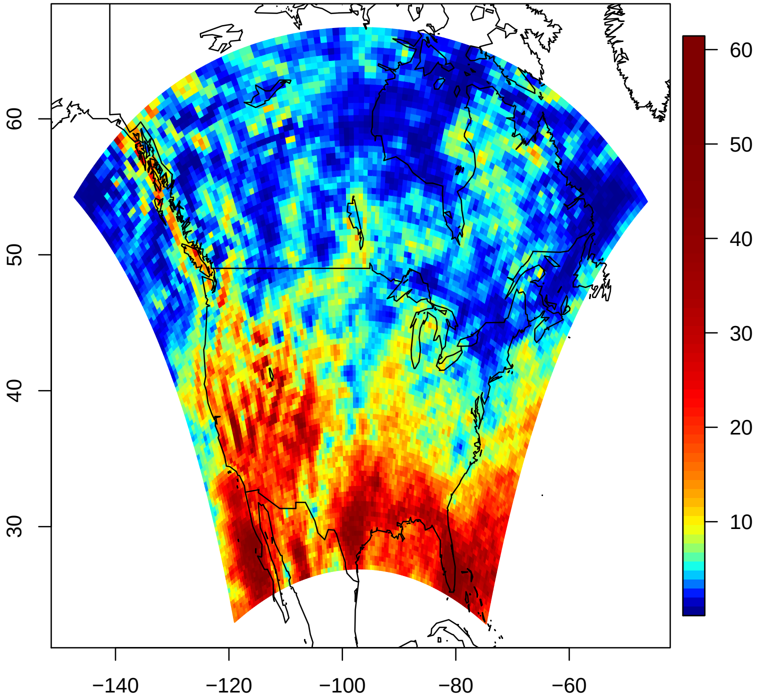


Examining ξ (Summer)

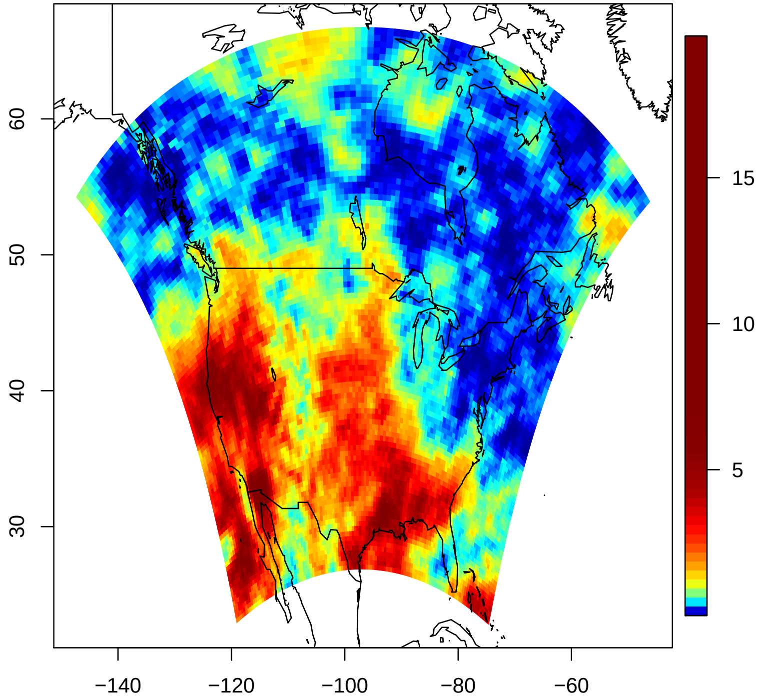


Significance (Summer)

100-year Return Level



ξ



F-test for equality of means, significance level: 2.22

Comparison to Ground Station (Summer)

Summer Estimates for Fort Collins, CO

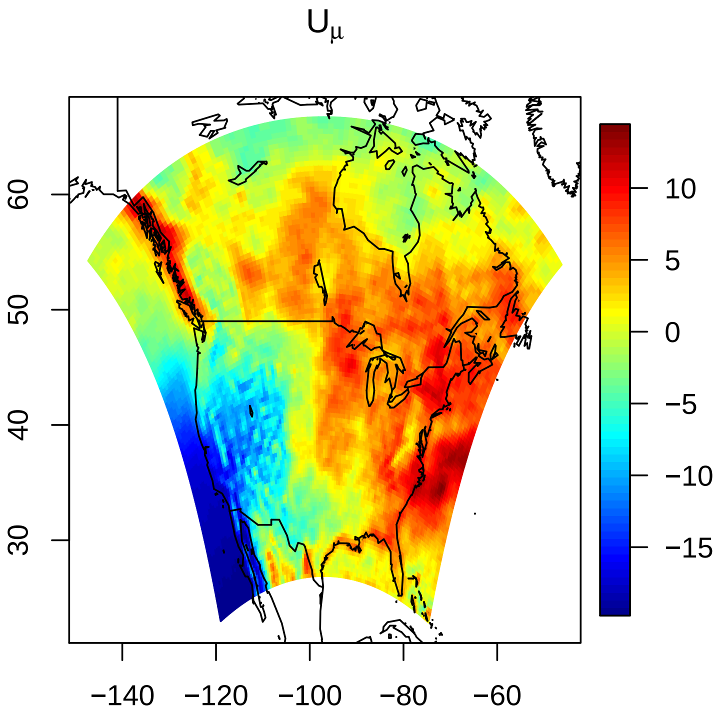
95% credible intervals

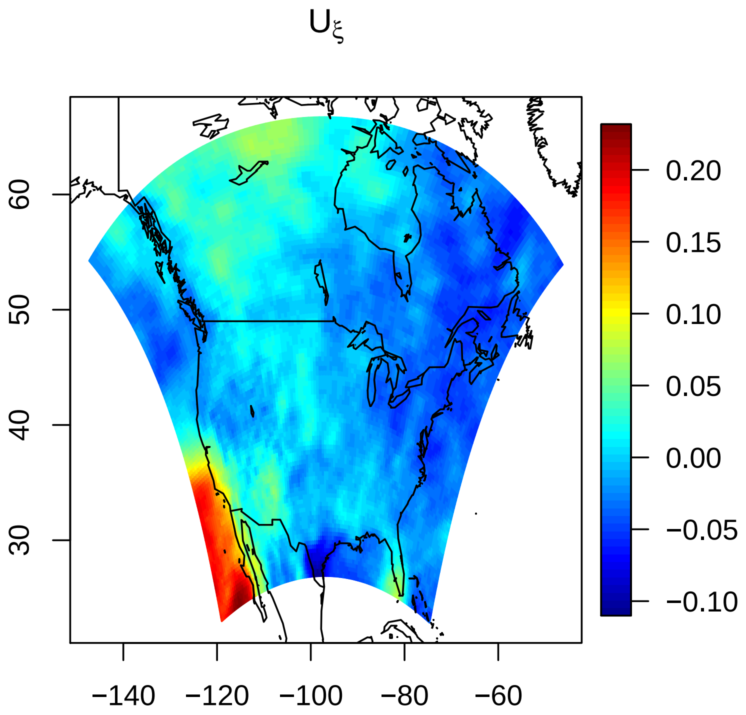
	ξ	100-yr RL
Weather Station[1]	(0.097, 0.144)	(9.01, 12.12)
CRCM	(0.040, 0.158)	(3.91, 5.63)
ECPC	(0.029, 0.145)	(6.70, 10.18)
HRM3	(0.080, 0.199)	(5.22, 8.40)
MM5I	(0.102, 0.224)	(6.76, 10.61)
RCM3	(0.100, 0.207)	(10.19, 15.52)
WRFP	(0.130, 0.240)	(3.54, 5.66)

[1] Weather station estimates from Cooley et al. (2007).

Is the spatial hierarchical model necessary?

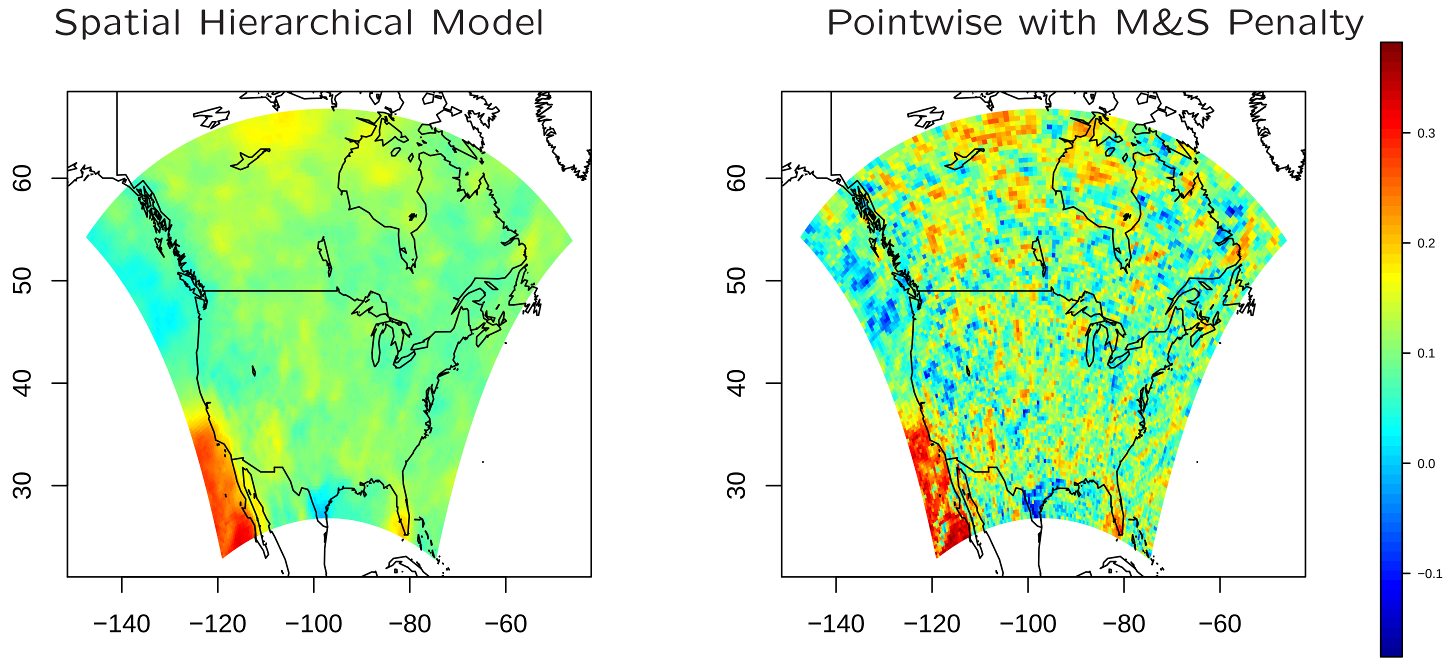
CRCM, Summer





What is the benefit of the spatial hierarchical model?

Estimates for ξ, CRCM Model, Summer



References

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102:824–840.

Cooley, D. and Sain, S. R. (2008). Spatial hierarchical modeling of precipiation extremes from a regional climate model. *accepted by JABES*. http://www.stat.colostate.edu/∼cooleyd/Papers/rcmPaper.pdf.

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.