Assessing Uncertainty in Regional Climate Experiments

Stephan R. Sain

Geophysical Statistics Project

Institute for Mathematics Applied to Geosciences

National Center for Atmospheric Research

Boulder, CO

Cari Kaufman, Doug Nychka, Linda Mearns (NCAR); Reinhard Furrer (CSM).

Supported by NSF ATM/DMS.

The Movie

(jan2002.mp4)

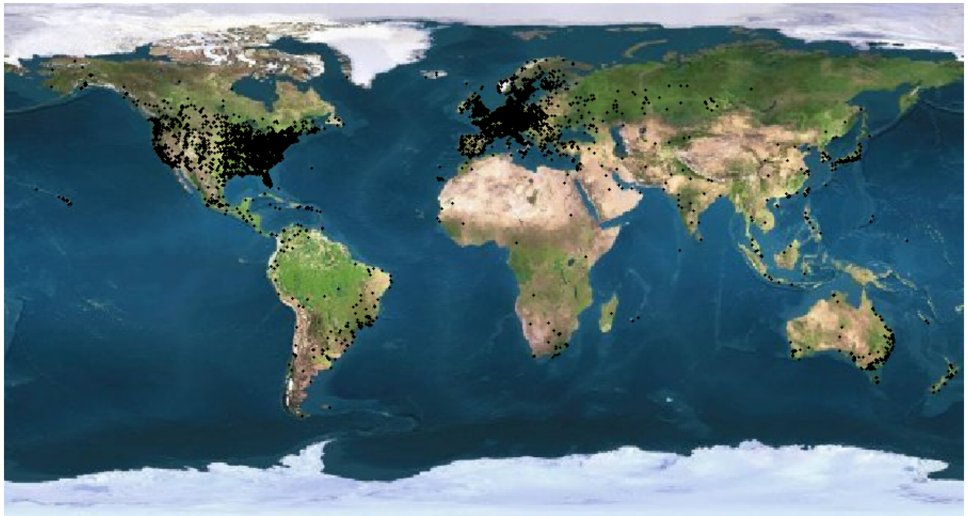
Or click here...

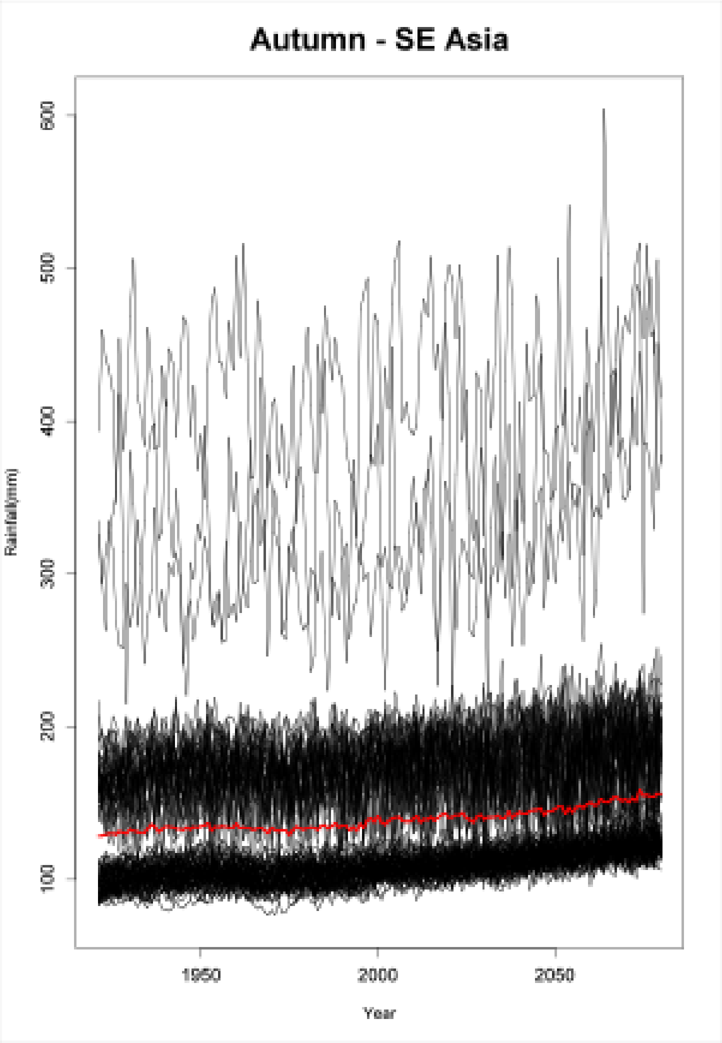
Goals

- Describe the distribution of (regional) climate model output.
- Understanding sources of variation.
  - NARCCAP/PRUDENCE: GCM, RCM, GCM$\times$RCM.
  - climateprediction.net: perturbed physics.
  - Others sources?
- Combining model output & weighting models.
- Recognizing model output represents spatial, temporal, or spatial-temporal fields $\Rightarrow$ *functional ANOVA*.
  - Gaussian process ANOVA (Kaufman and Sain, 2007).

climateprediction.net

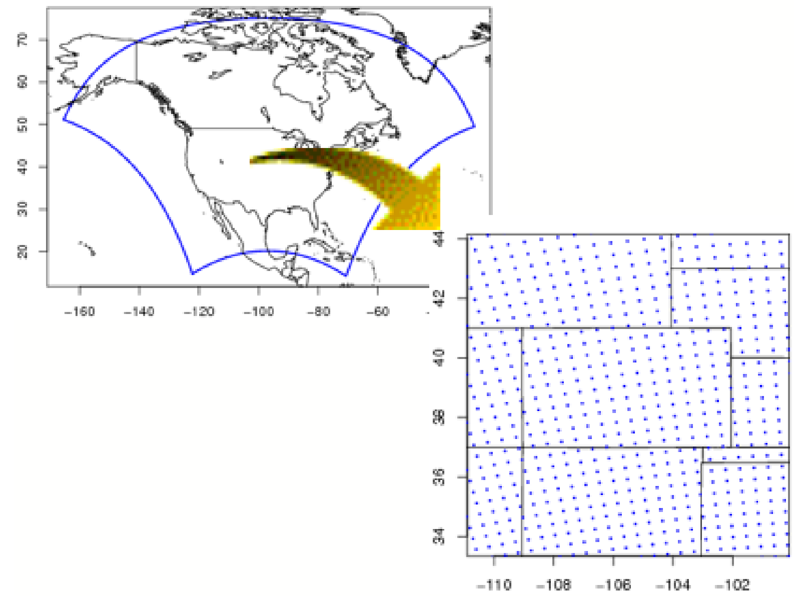
- Uses idle time on PCs to run a full-resolution AOGCM with varying input parameters.





NARCCAP

- North American Regional Climate Change Assessment Program (NARCCAP)
  - NCAR, ISU, CCCma, OURANOS, LLNL, GFDL, Hadley, Scripps, PNNL, USSC, UCDHSC, etc.
  - NSF, NOAA, DOE, etc.
  - www.narccap.ucar.edu



- Systematically investigate the uncertainties in regional scale projections of future climate.

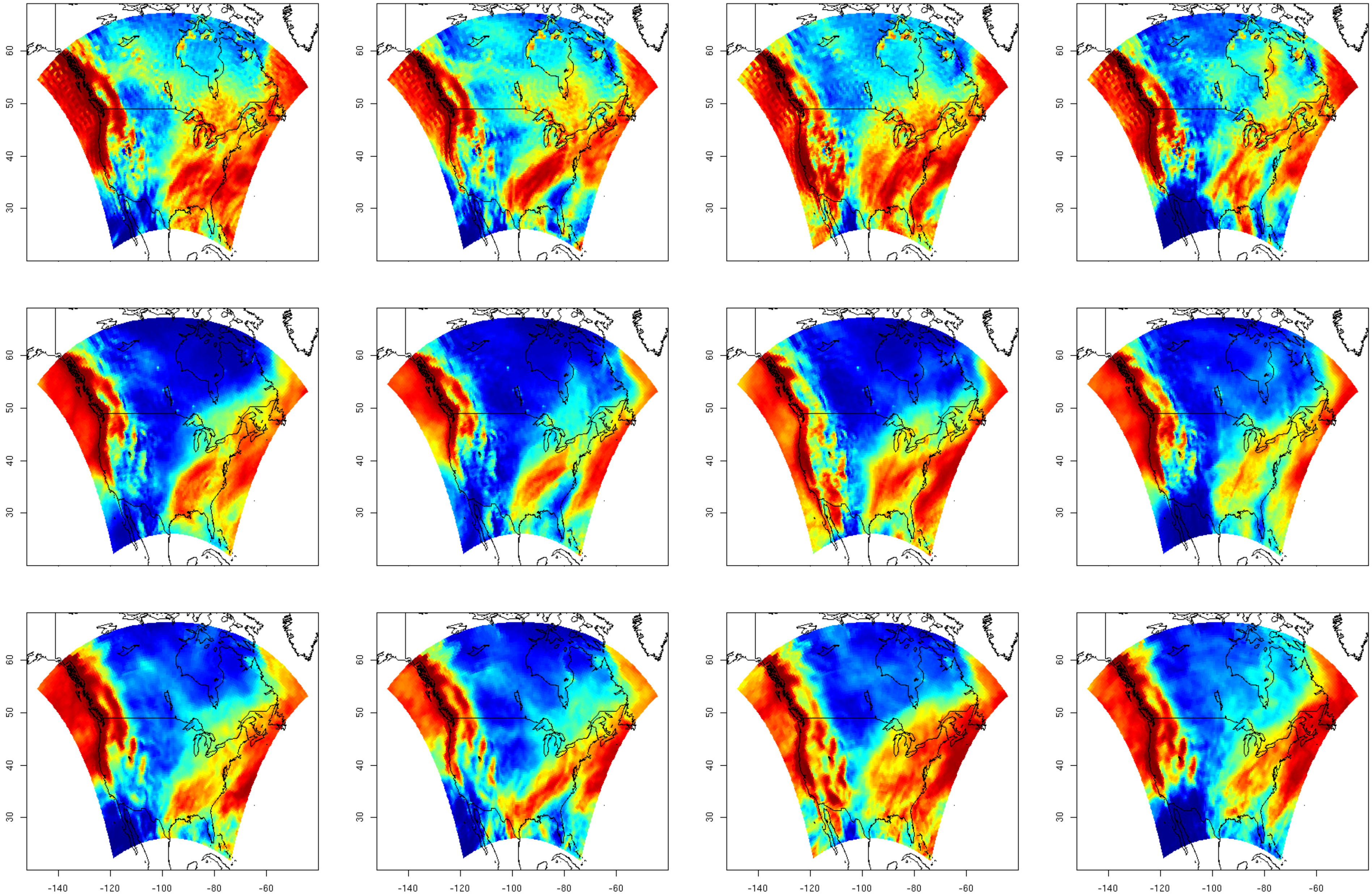
NARCCAP Design

- 4 GCMs provide boundary conditions for 6 RCMs

		GCM			
		GFDL	CGCM3	HADCM3	CCSM
	MM5			X	X
	RegCM3	X	X		
RCM	CRCM		X	X	
	PRECIS	X	X	X	X
	RSM	X			X
	WRF	X	X		X

A Work in Progress

- Three regional models – ECPC, MRCC, and RCM3
- Boundary conditions supplied by reanalysis.
- 1980-1999 (20 years)
- Total seasonal precipitation – winter (DJF) and summer (JJA)
- Common grid: $123 \times 101 = 12,423$ grid boxes


60
50
40
30
-140
-120
-100
-80
-60

A Statistical Model

- A hierarchical construction:

  Data model: $\mathbf{Y}_{ij} \sim \mathrm{N}\left(\boldsymbol{\mu}_i, \sigma_1^2 \mathbf{V}(\theta_1)\right),\ i = 1, 2, 3,\ j = 1, \ldots, 20$

  Process model: $\boldsymbol{\mu}_i \sim \mathrm{N}\left(\boldsymbol{\mu}, \sigma_2^2 \mathbf{V}(\theta_2)\right)$

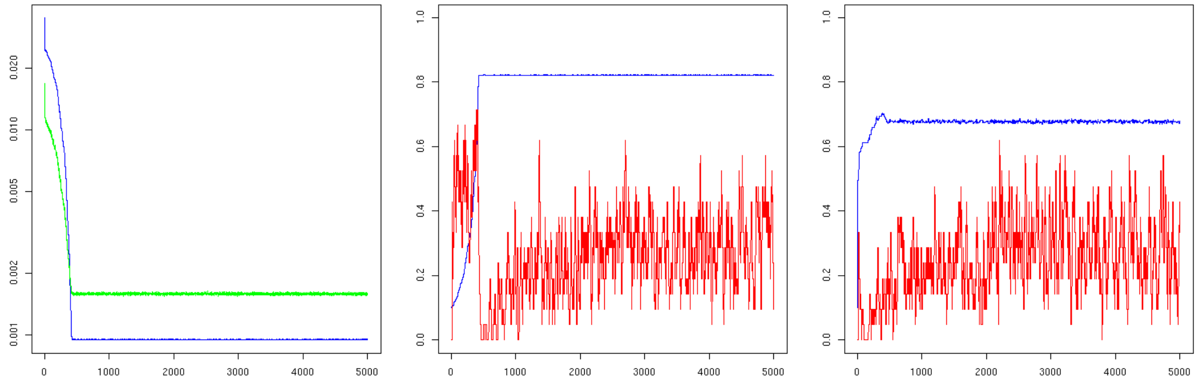
  Prior model: non-informative.

- An alternative formulation:

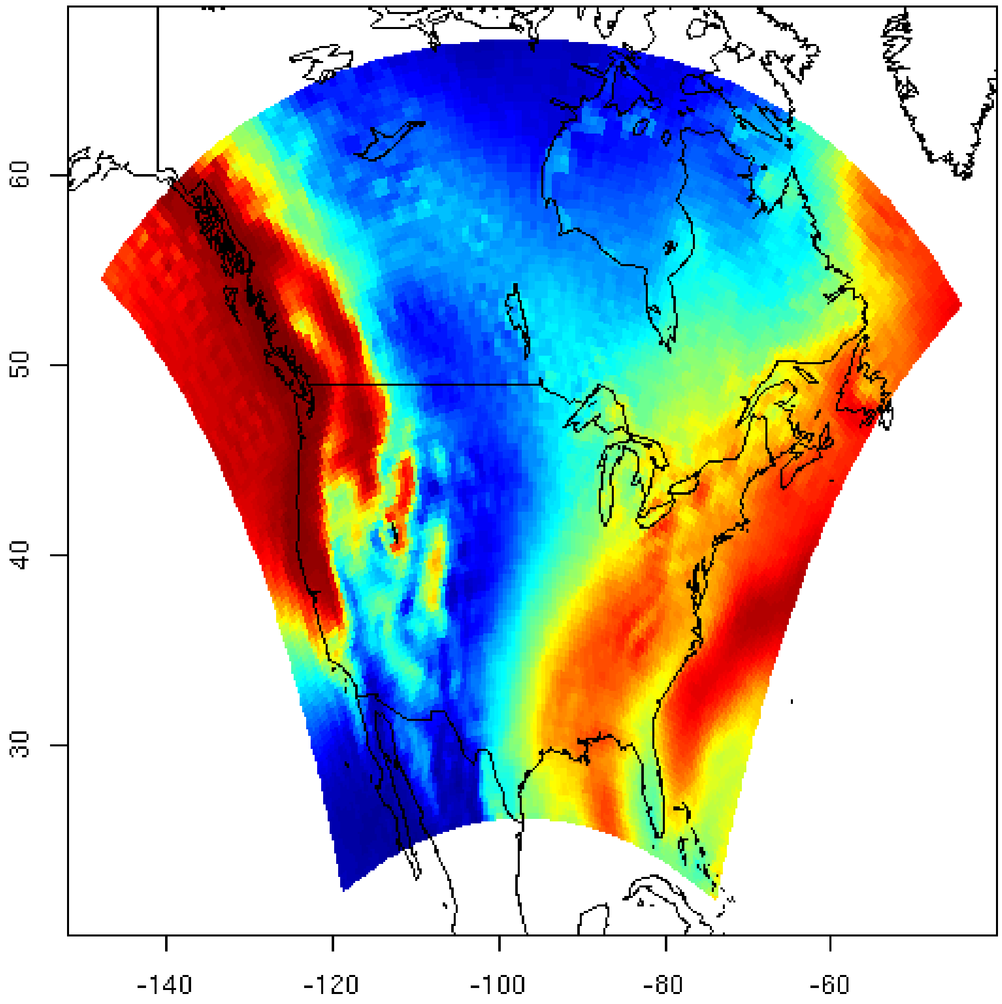
$$\begin{aligned} \mathbf{Y}_{ij} &= \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{ij} \\ &= \text{Common} + \text{RCM} + \text{Error} \end{aligned}$$

A Statistical Model

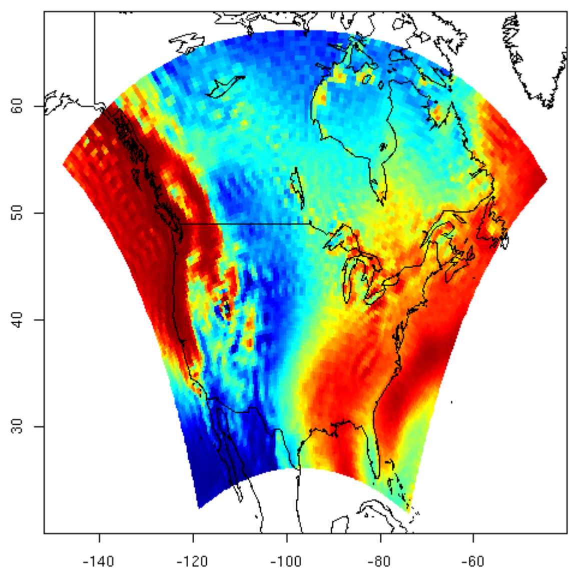
- Spatial covariance $\mathbf{V}(\theta) = \mathbf{R}(\theta) \otimes \mathbf{C}(\theta)$ where $\mathbf{R}$ and $\mathbf{C}$ are parameterized through 1-D "stationary" Markov random fields.
  - Computationally efficient: sparse precision matrices.
  - Other choices: tapering, nonstationary forms, etc.
- MCMC to estimate parameters, posterior inference, etc.

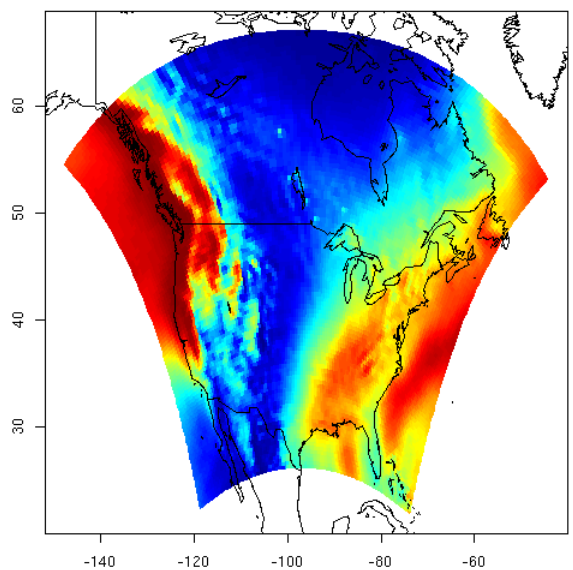


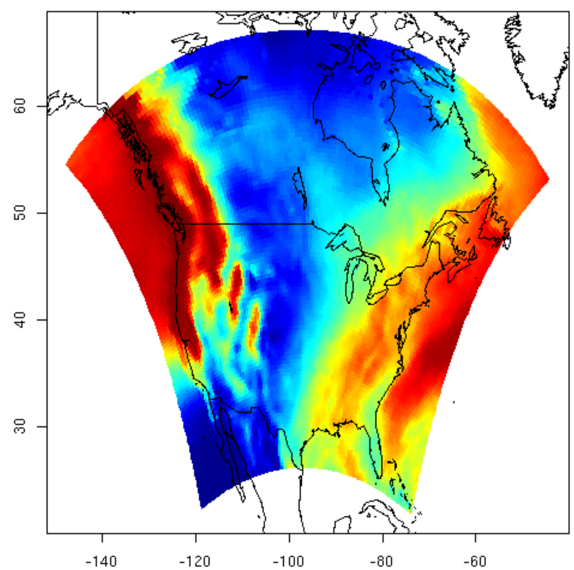
Posterior Means

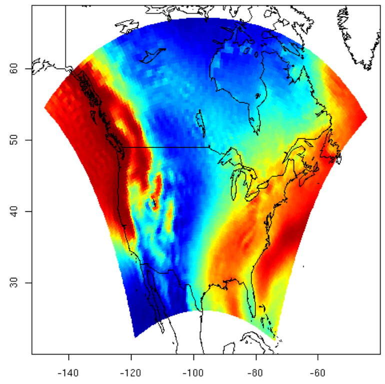


Posterior Means

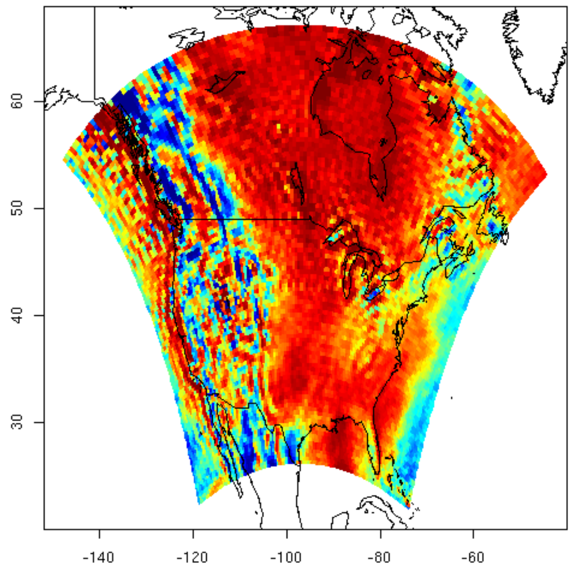


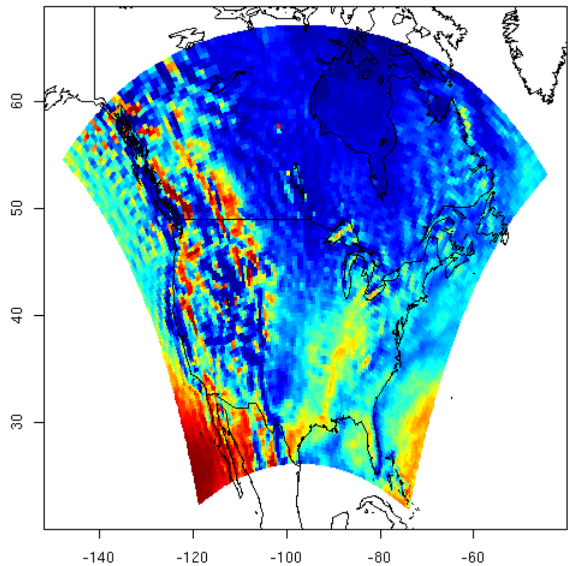


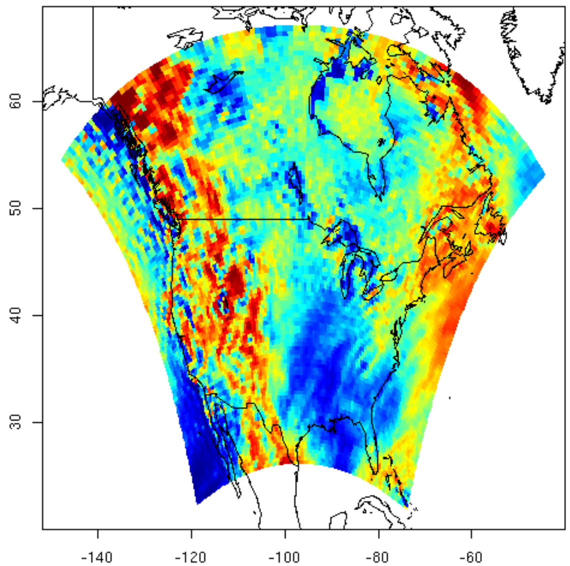


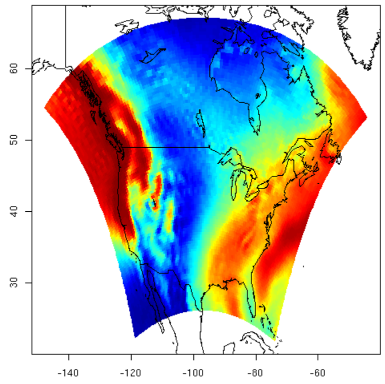


Posterior Means





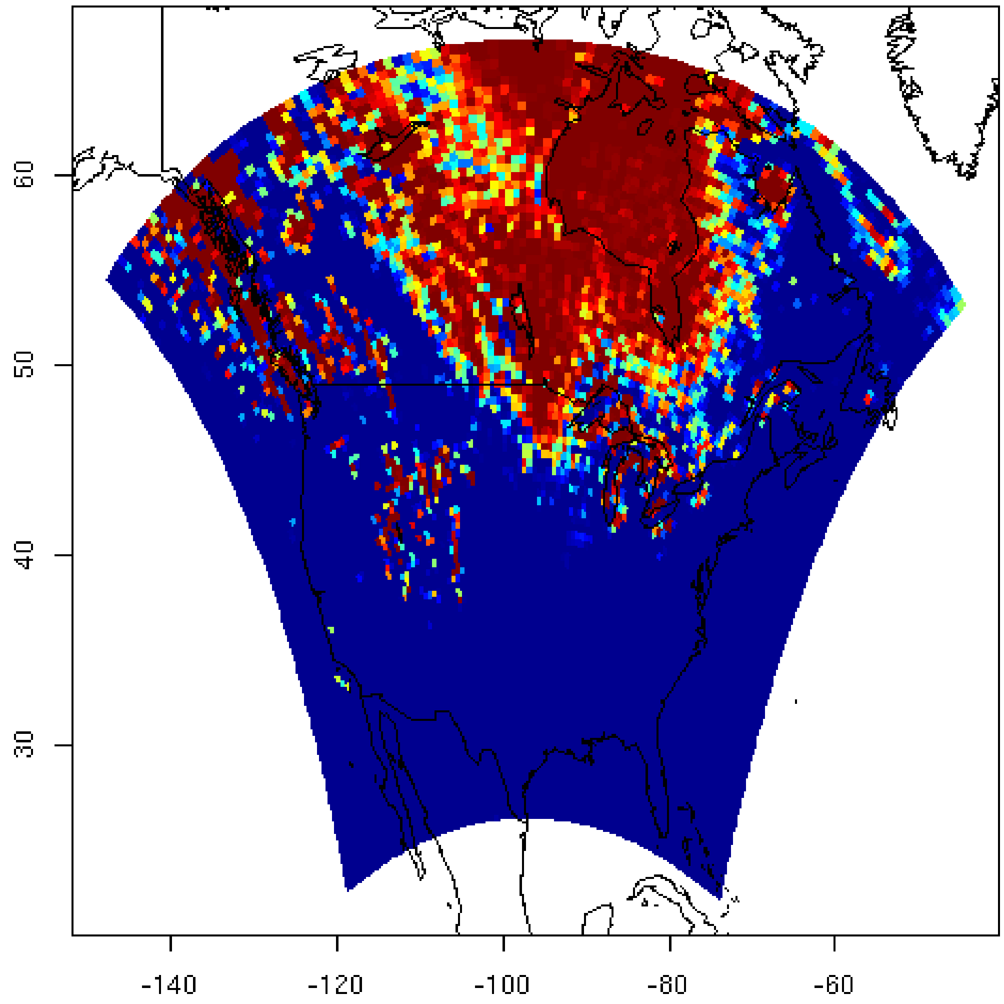




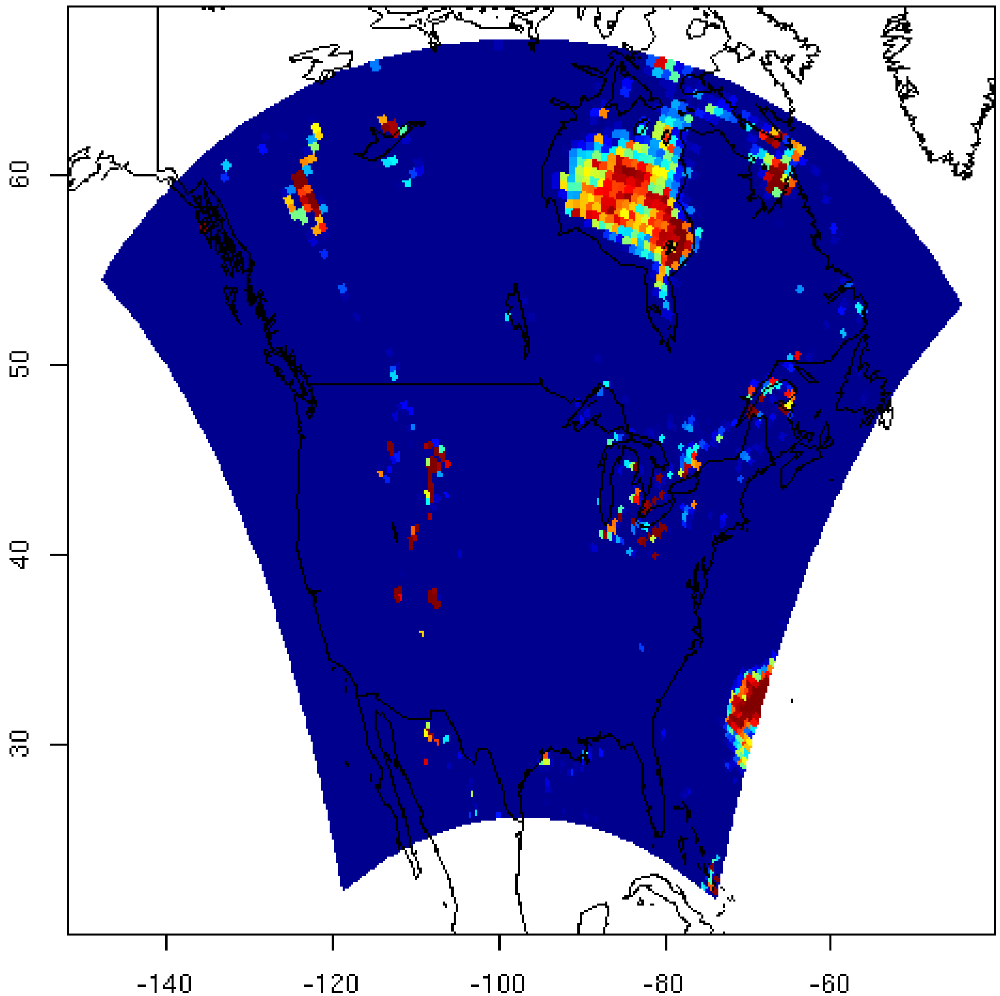
Inference

- for i in 1 to "a big number"...
  - sample $(\boldsymbol{\mu}^*, \boldsymbol{\mu}_1^*, \boldsymbol{\mu}_2^*, \boldsymbol{\mu}_3^*) \Rightarrow \boldsymbol{\alpha}^* = (\boldsymbol{\alpha}_1^*, \boldsymbol{\alpha}_2^*, \boldsymbol{\alpha}_3^*)$
  - construct (for each grid box):
    - $s_{\boldsymbol{\alpha}}^2$ (model-to-model variation)
    - s^2 (year-to-year variation)
  - identify and record grid boxes where $s_{\boldsymbol{\alpha}}^2$ is larger than s^2.
- compute $\widehat{P}[s_{\boldsymbol{\alpha}}^2 > s^2]$ for each grid box

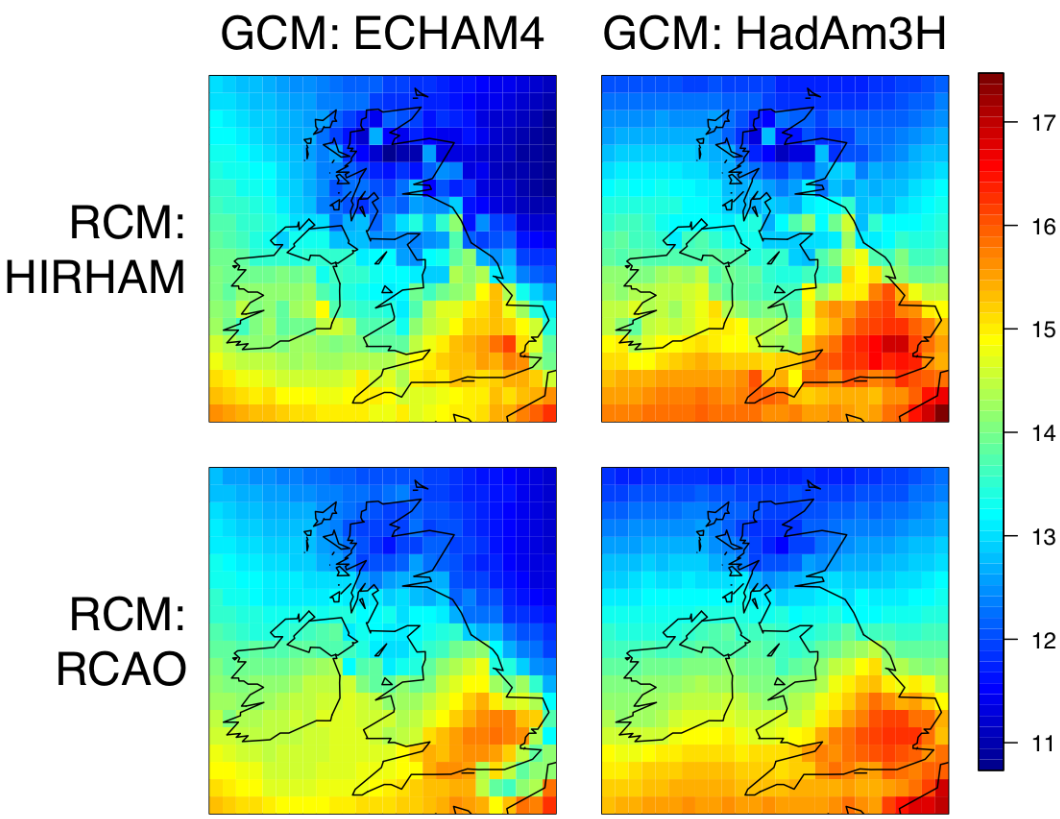
Winter Precipitation



Summer Precipitation

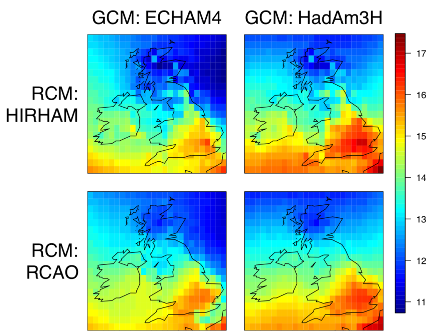


A PRUDENCE Example



- 2x2 “experiment”
  - 2 GCMs, 2 RCMs
  - PRUDENCE
- 1961-1990
- JJA average temp

A Two-Factor Model



$$Z_{ijt}(s) \quad = \quad \mu_{ijt}(s) \quad + \quad \epsilon_{ijt}(s)$$

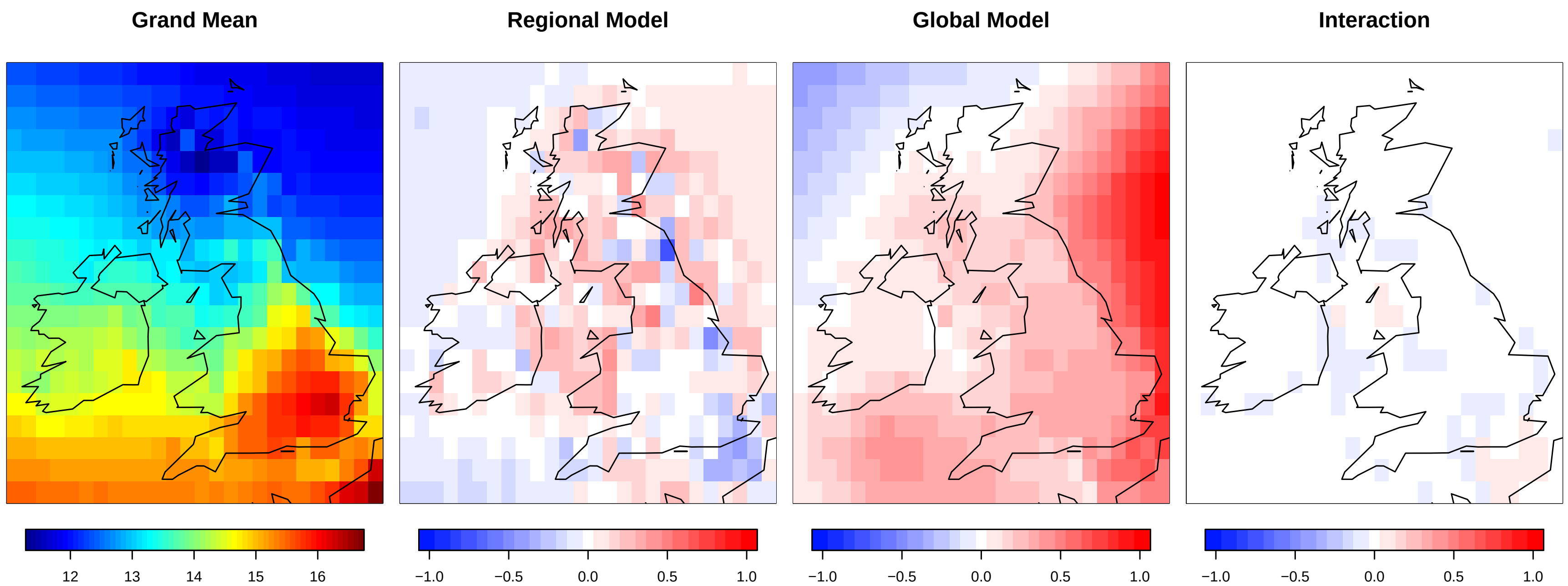
Output of RCM i, GCM j, at time t and location s	=	Expected/ "Climate" response	+	Spatially correlated residual/ "internal model variability"

$$\mu_{ijt}(s) = \mu(s) + i\alpha(s) + j\beta(s) + ij(\alpha\beta)(s) + \gamma t,$$

= Common + RCM + GCM + Interaction + Time

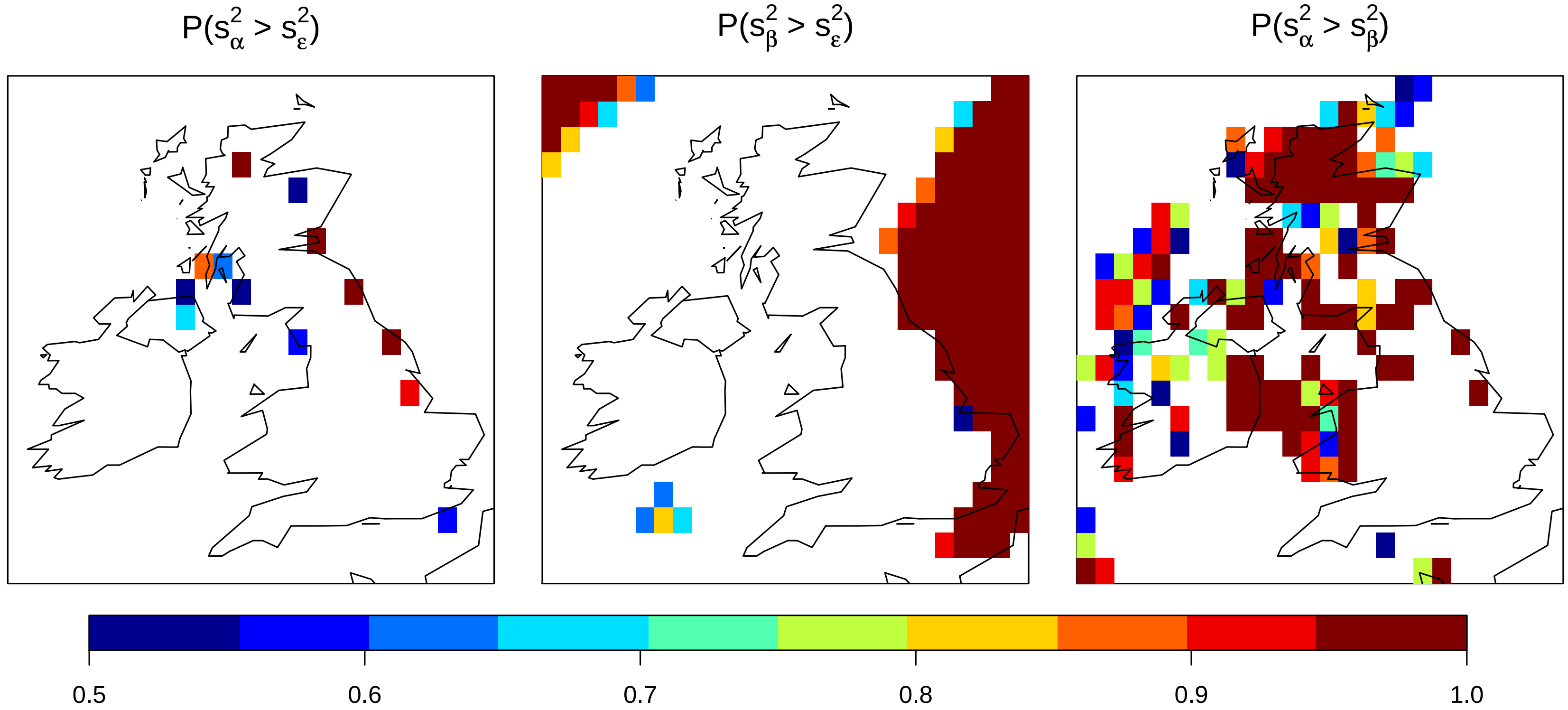
- $i, j = -1, 1$ (contrast coding)
- Hierarchical model with Gaussian process priors used for each effect.
- MCMC used to estimate parameters, posterior inference, etc.

Posterior Means



- Estimates of spatial effects.

Functional ANOVA



- Ratios of variances.

Final Thoughts

- Design issues become even more important as we move to a petascale computing environment...
- The posterior distribution is a good thing, but then what?
- Issues in model weighting...
  - What is the goal? Small error or span distribution?
  - What is the target? Model bias...
  - Likelihood–based weighting good, but where is the data?
  - Model correlations
  - Non-stationarity

Questions?

ssain@ucar.edu