Case Study III: Functional ANOVA

Markov Random Fields and Regional Climate Models

Stephan R. Sain
Geophysical Statistics Project
Institute for Mathematics Applied to Geosciences
National Center for Atmospheric Research
Boulder, CO

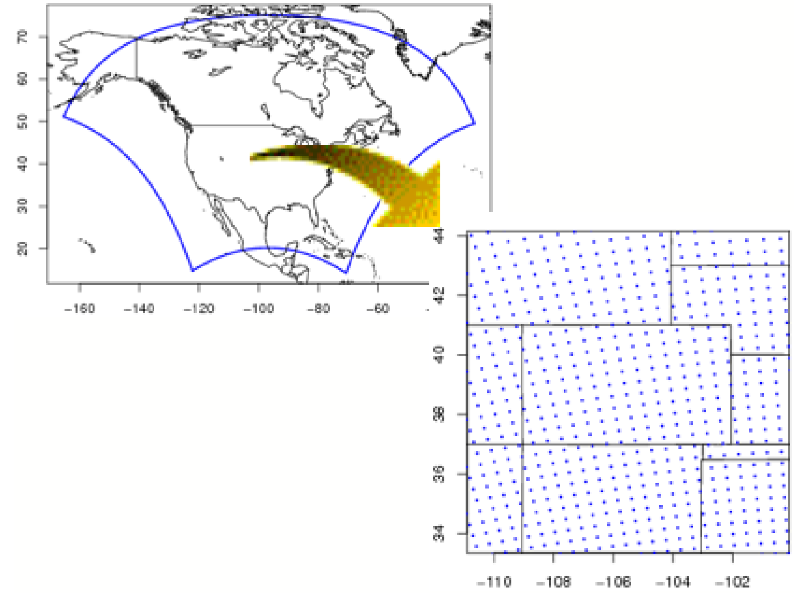
Supported by NSF ATM/DMS. Thanks also to Cari Kaufman.

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*.

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.

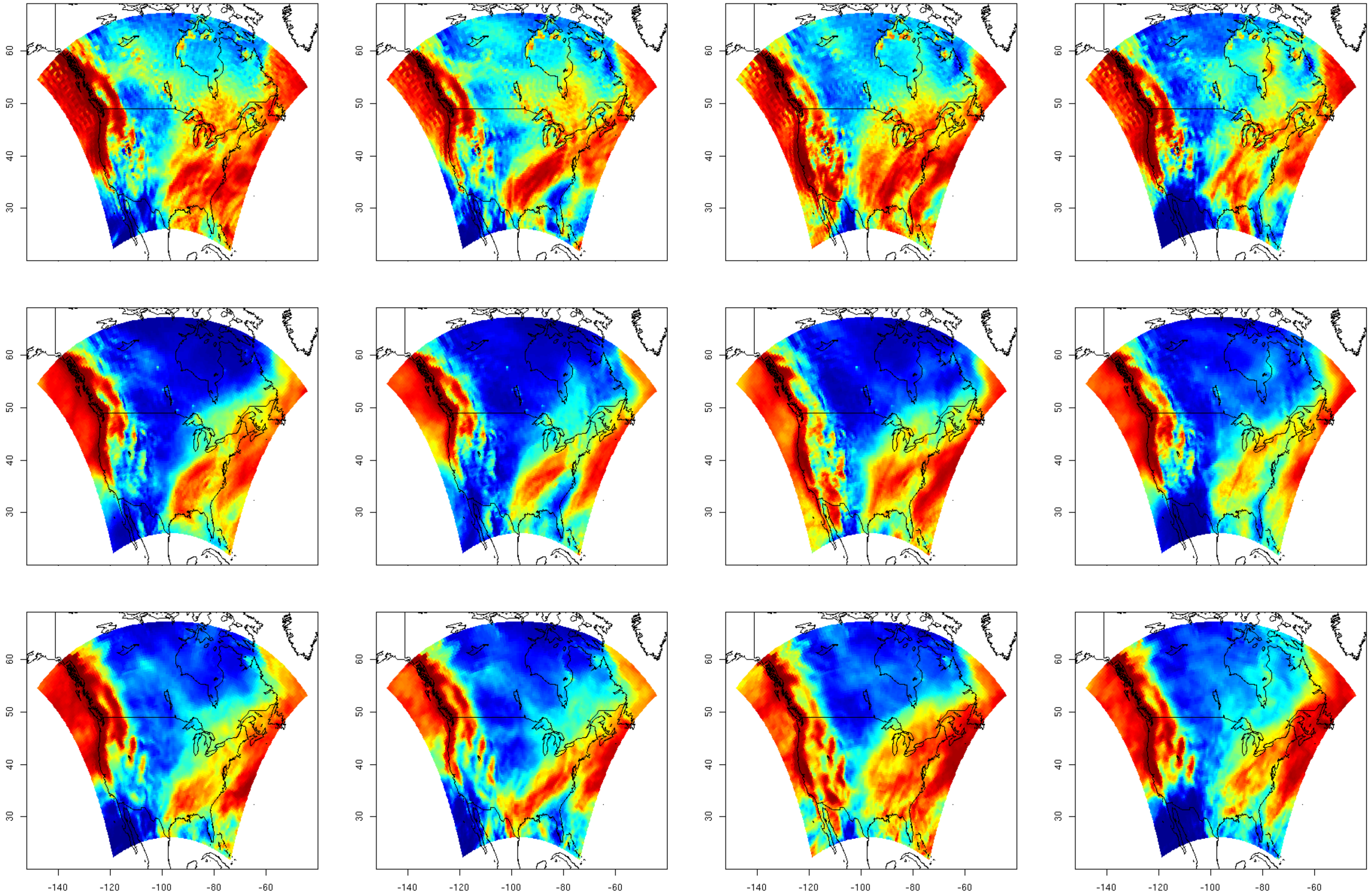
NARCCAP Design

- 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 correlation matrix $\mathbf{V}(\theta) = \mathbf{R}(\theta) \otimes \mathbf{C}(\theta)$ where $\mathbf{R}$ and $\mathbf{C}$ are parameterized through 1-D "stationary" Markov random fields:

$$\mathbf{R}(\theta) = \begin{bmatrix} 1 & -\theta & & & & & \\ -\theta & 1+\theta^2 & -\theta & & & & \\ & -\theta & 1+\theta^2 & -\theta & & & \\ & & & \ddots & & & \\ & & & -\theta & 1+\theta^2 & -\theta & \\ & & & & -\theta & 1+\theta^2 & -\theta \\ & & & & & -\theta & 1 \end{bmatrix}^{-1}$$

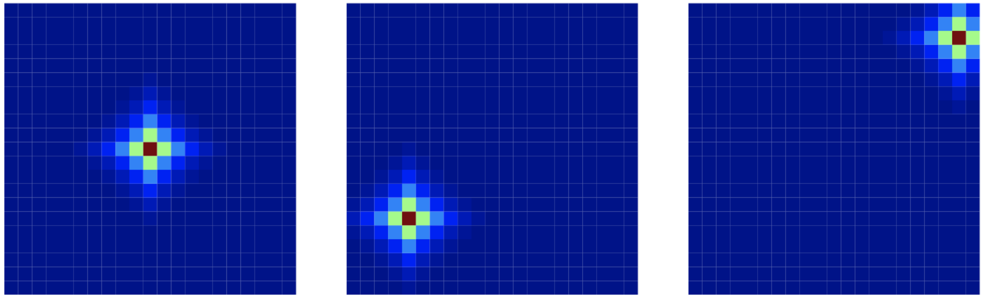
  - Computationally efficient: lower-dimensional + sparse precision matrices.
  - Other choices: tapering, reduced-rank kriging, etc.

MRF Formulation

- The conditional weight structure for an interior point:

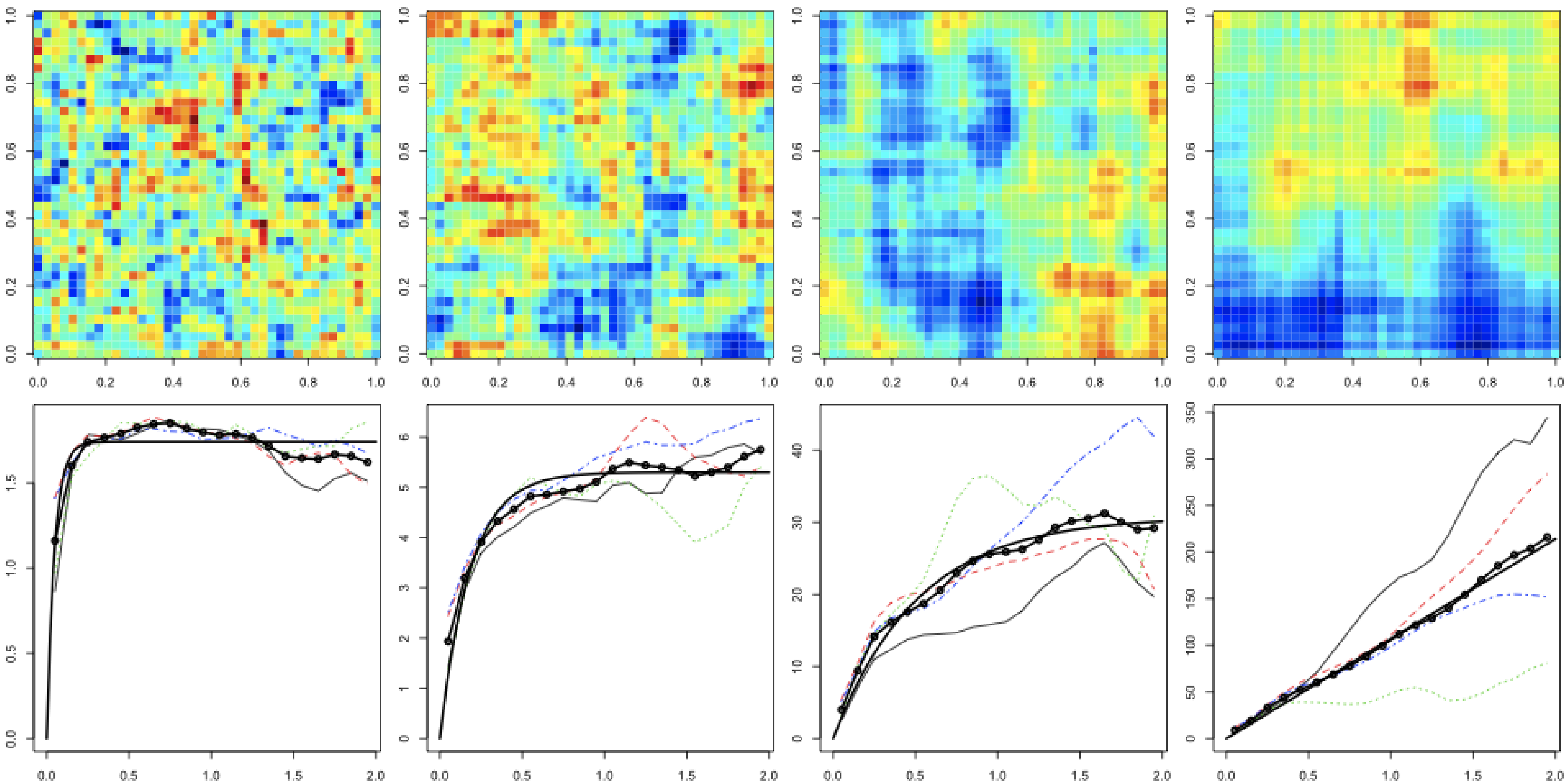
$$\frac{1}{(1+\theta^2)^2}\left\{\begin{array}{ccccccc} & & & \vdots & & & \\ & 0 & 0 & 0 & 0 & 0 & \\ & 0 & \theta^2 & -\theta(1+\theta^2) & \theta^2 & 0 & \\ \cdots & 0 & -\theta(1+\theta^2) & 0 & -\theta(1+\theta^2) & 0 & \cdots \\ & 0 & \theta^2 & -\theta(1+\theta^2) & \theta^2 & 0 & \\ & 0 & 0 & 0 & 0 & 0 & \\ & & & \vdots & & & \end{array}\right\}$$

- And the resulting correlation functions:



Another View

- Exponential-like behavior?



A Statistical Model

- A hierarchical construction:

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

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

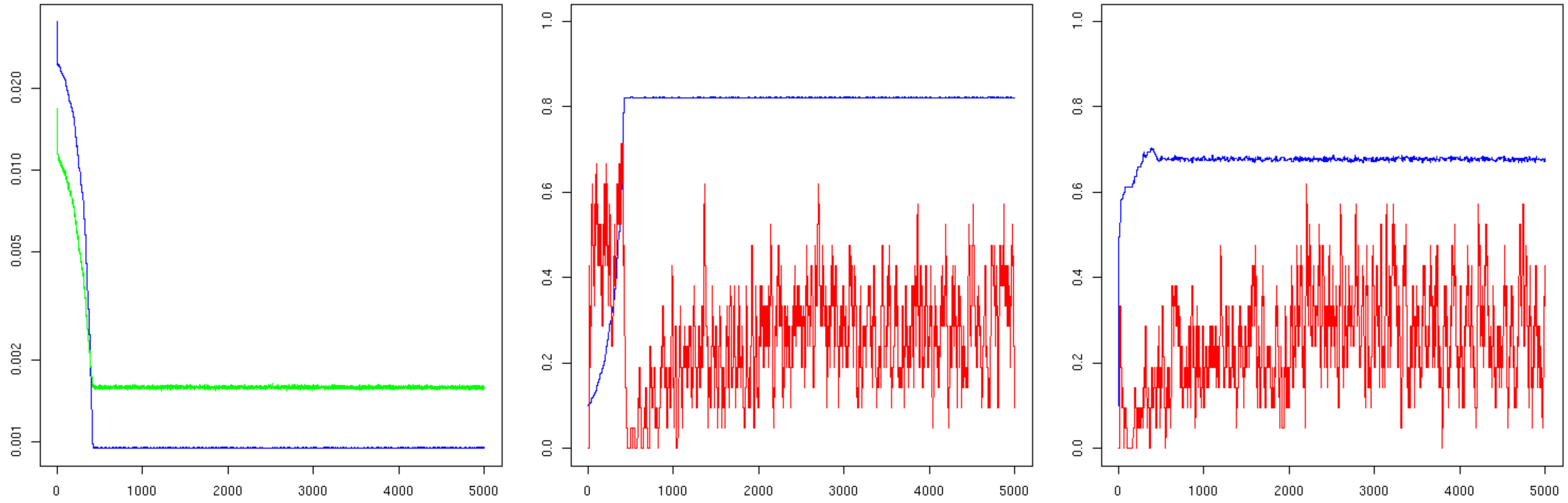
  Prior model: non-informative.

- An alternative formulation:

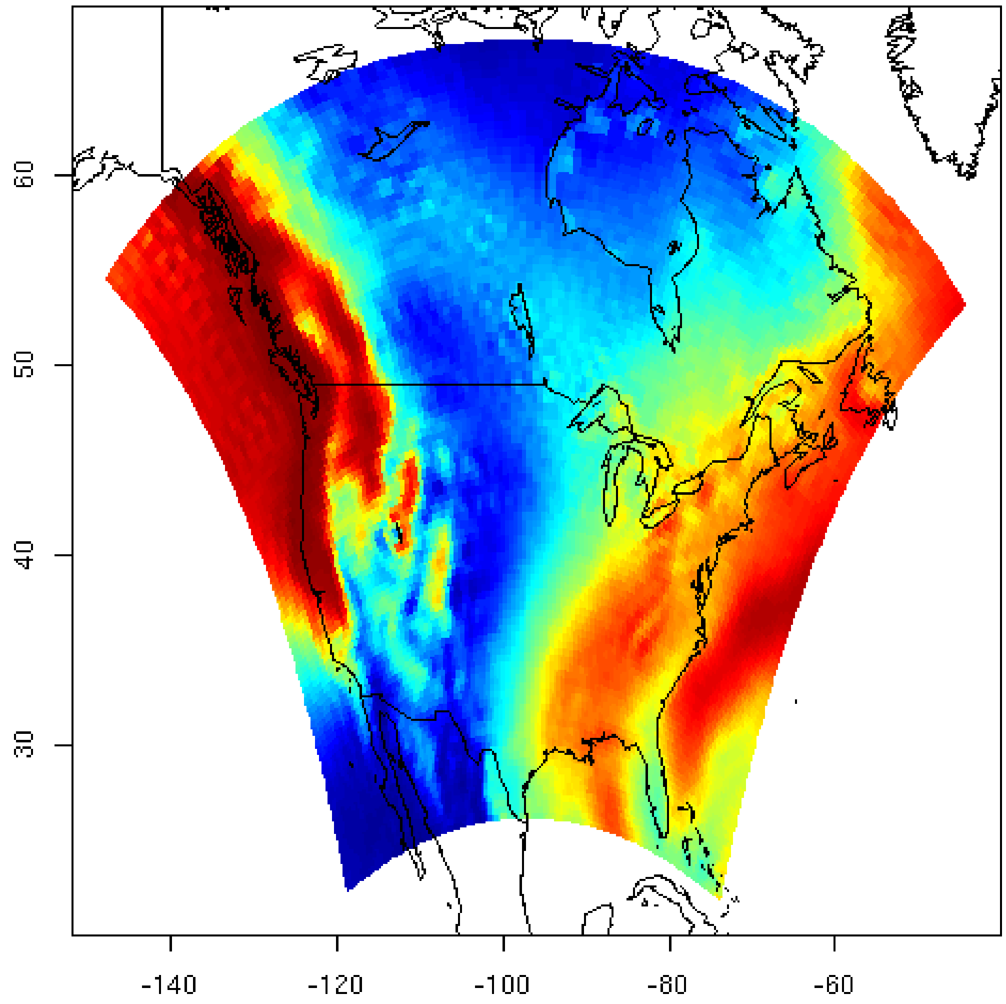
$$
\begin{aligned}
\mathbf{Y}_{ij} &= \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{ij} \\
&= \text{Common} + \text{RCM} + \text{Error}
\end{aligned}
$$

Parameter Estimation

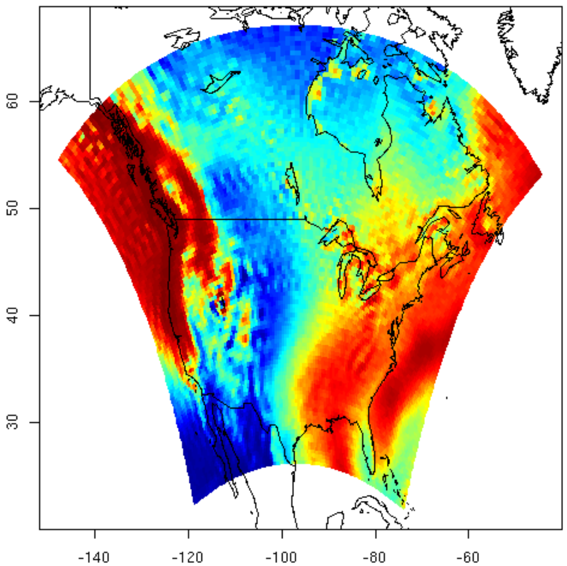
- MCMC to estimate parameters, posterior inference, etc.

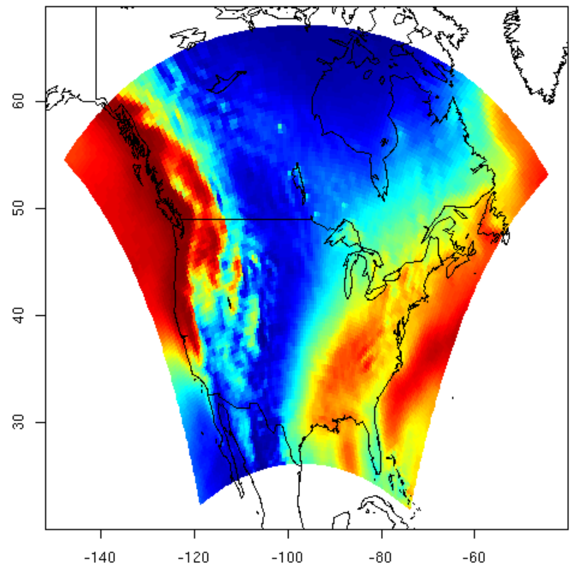


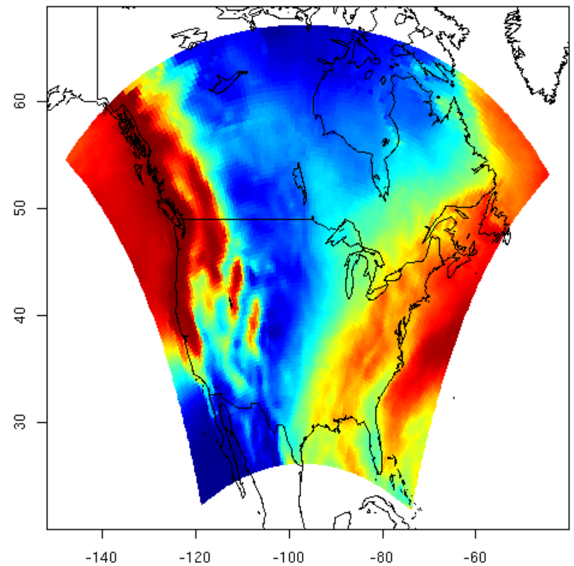
Posterior Means

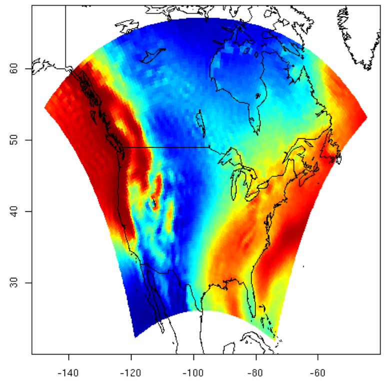


Posterior Means

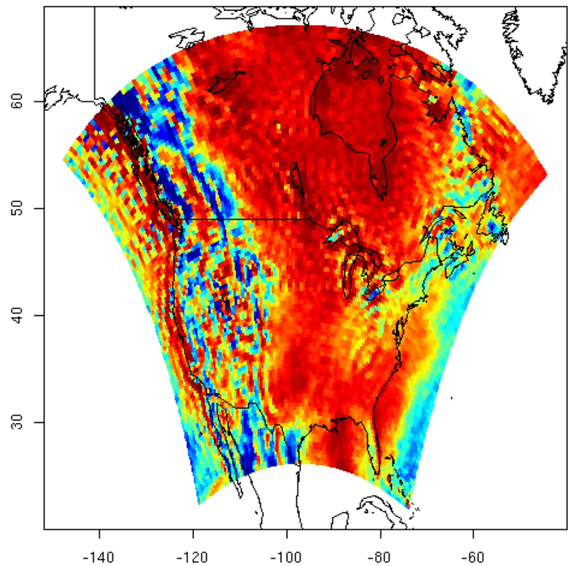


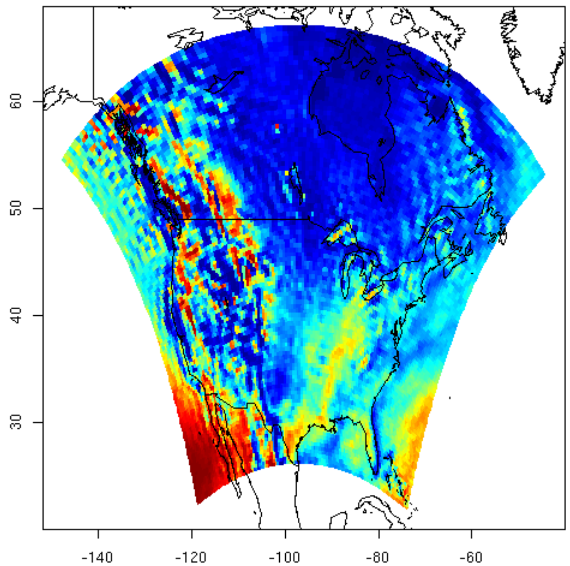


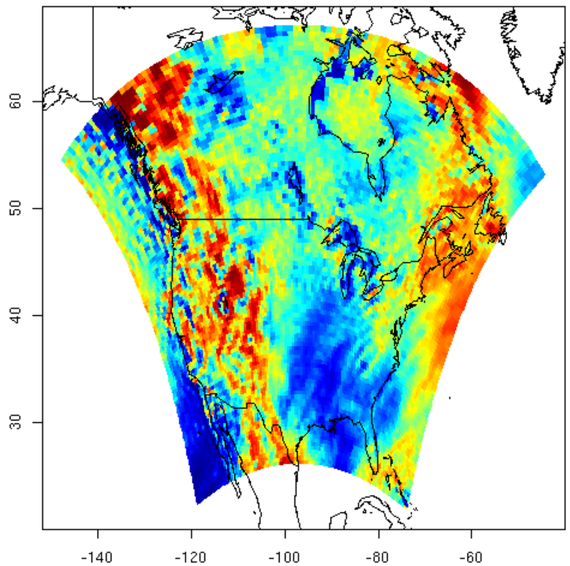


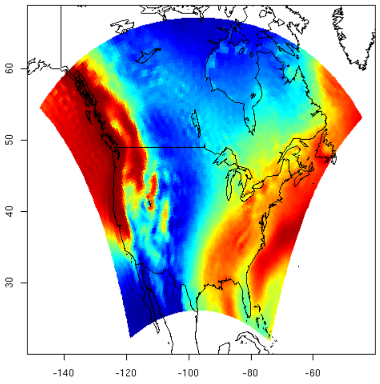


Posterior Means





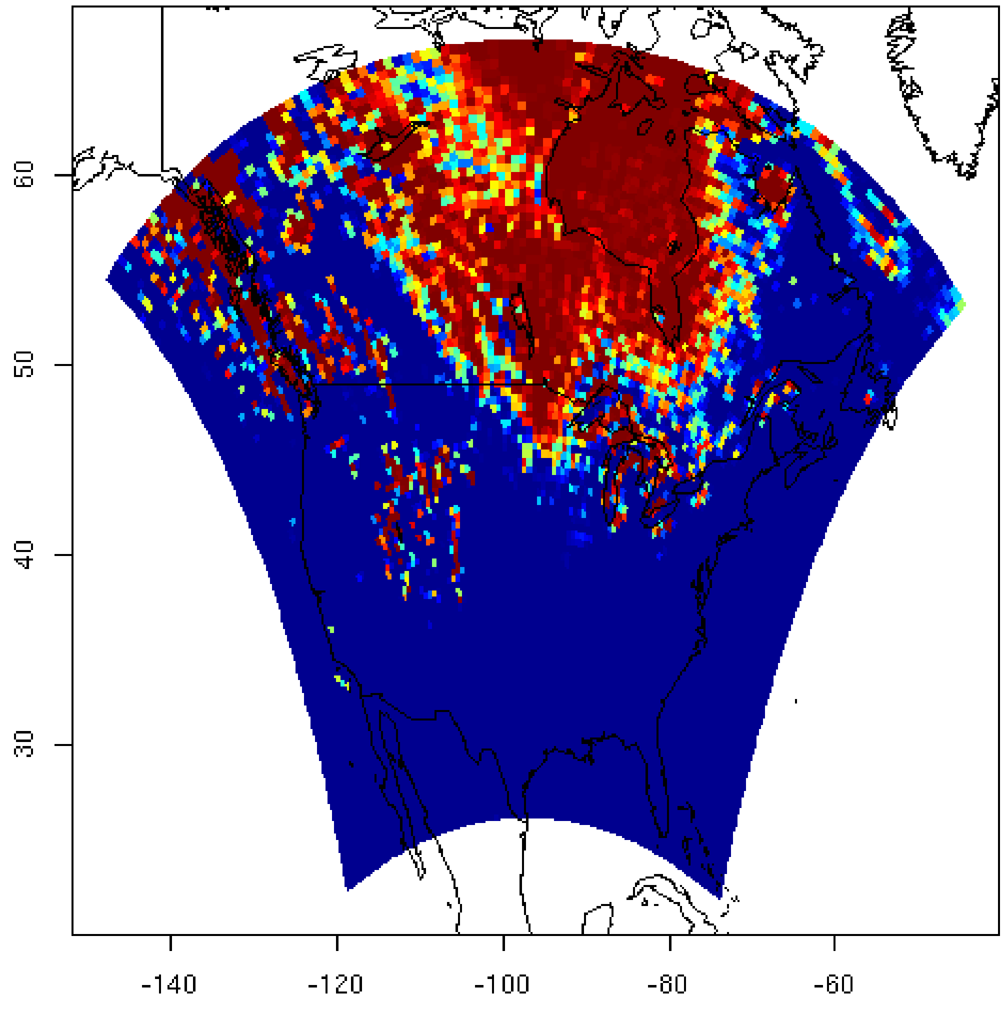




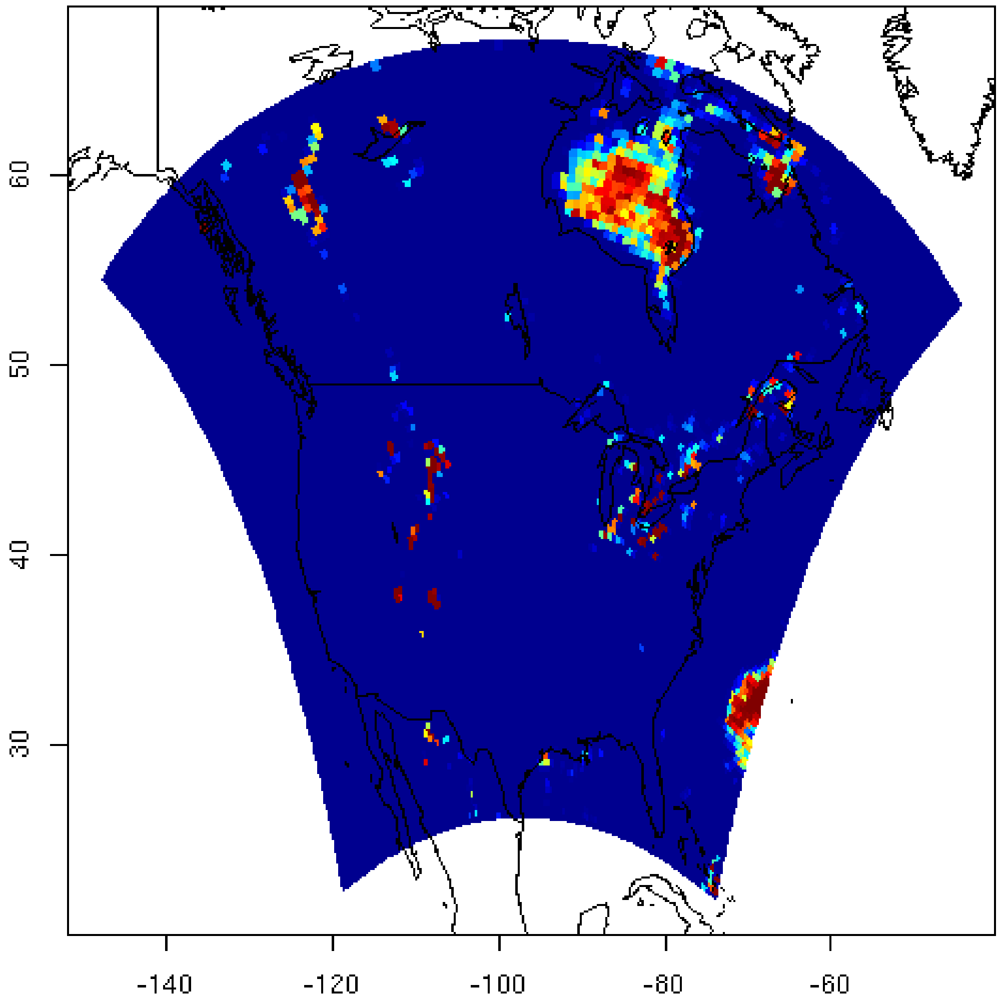
Inference

- 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

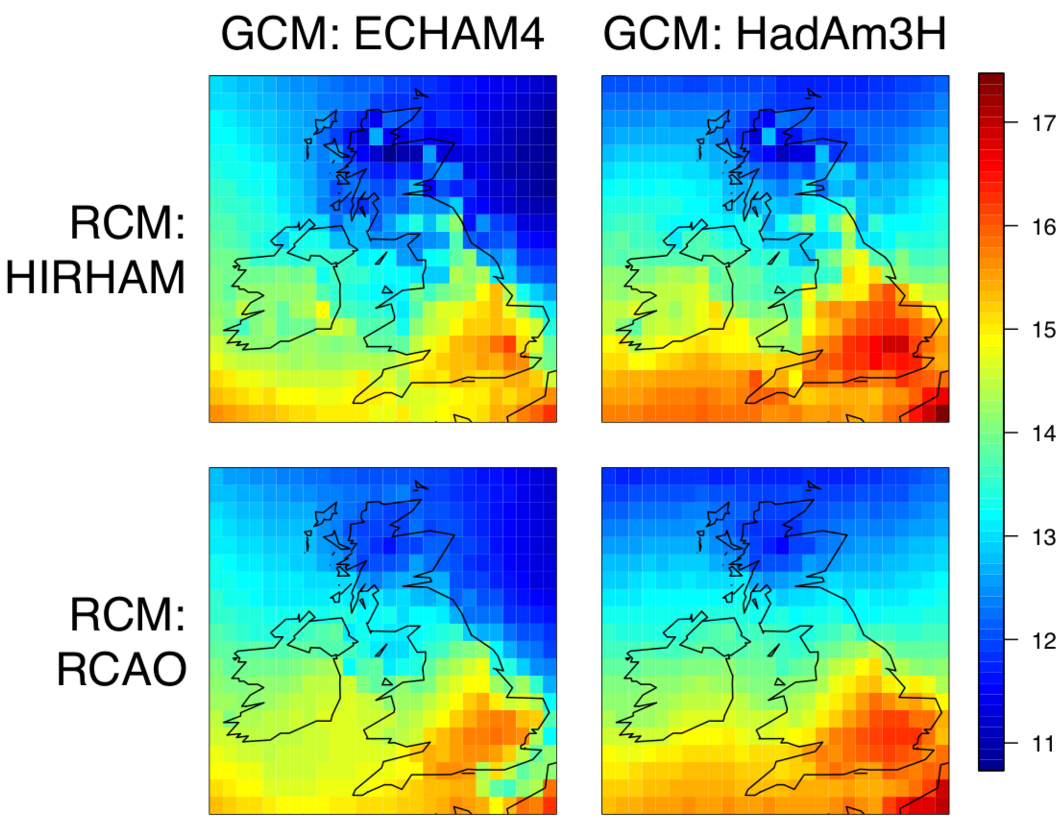
Winter Precipitation



Summer Precipitation

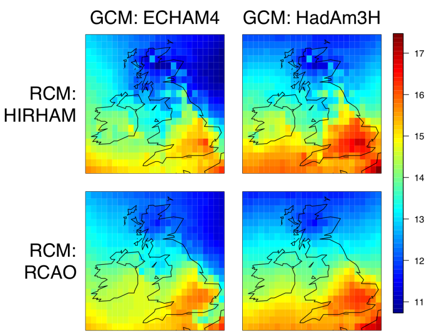


A PRUDENCE Example



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

A Two-Factor Model



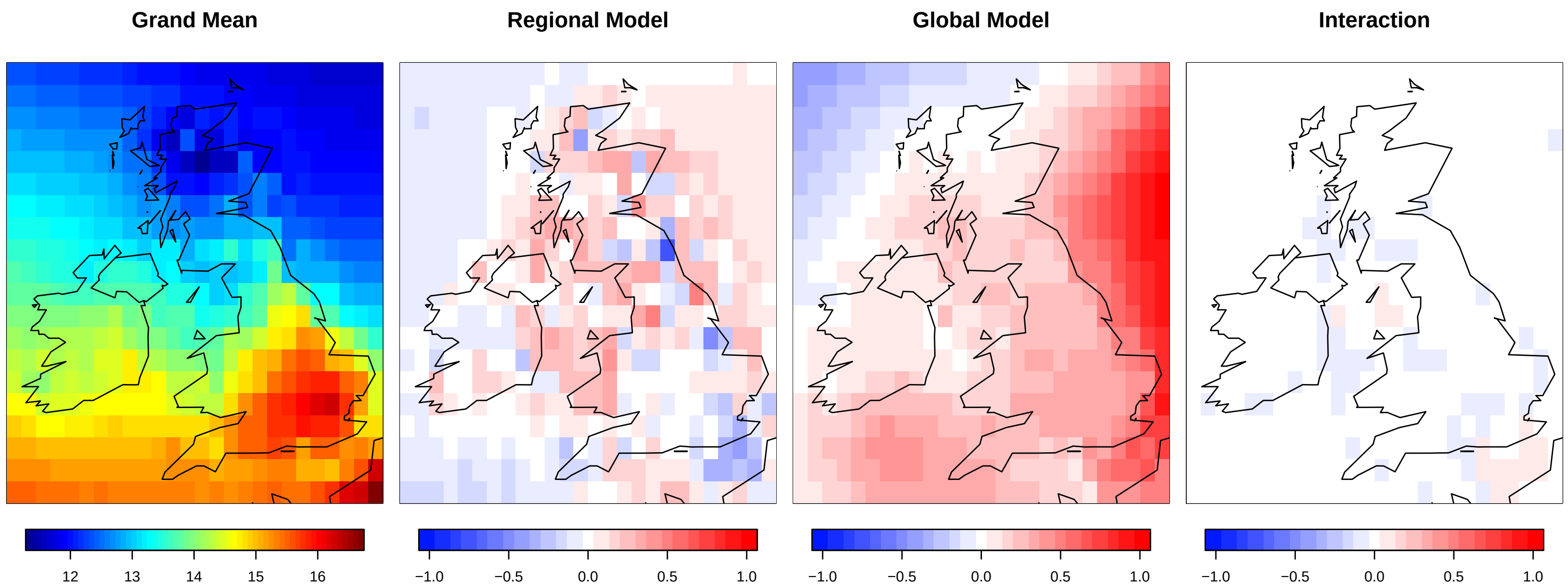
$$Z_{ijt}(s) = \mu_{ijt}(s) + \epsilon_{ijt}(s)$$

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

$$\begin{aligned} \mu_{ijt}(s) &= \mu(s) + i\alpha(s) + j\beta(s) + ij(\alpha\beta)(s) + \gamma t, \\ &= \text{Common} + \text{RCM} + \text{GCM} + \text{Interaction} + \text{Time} \end{aligned}$$

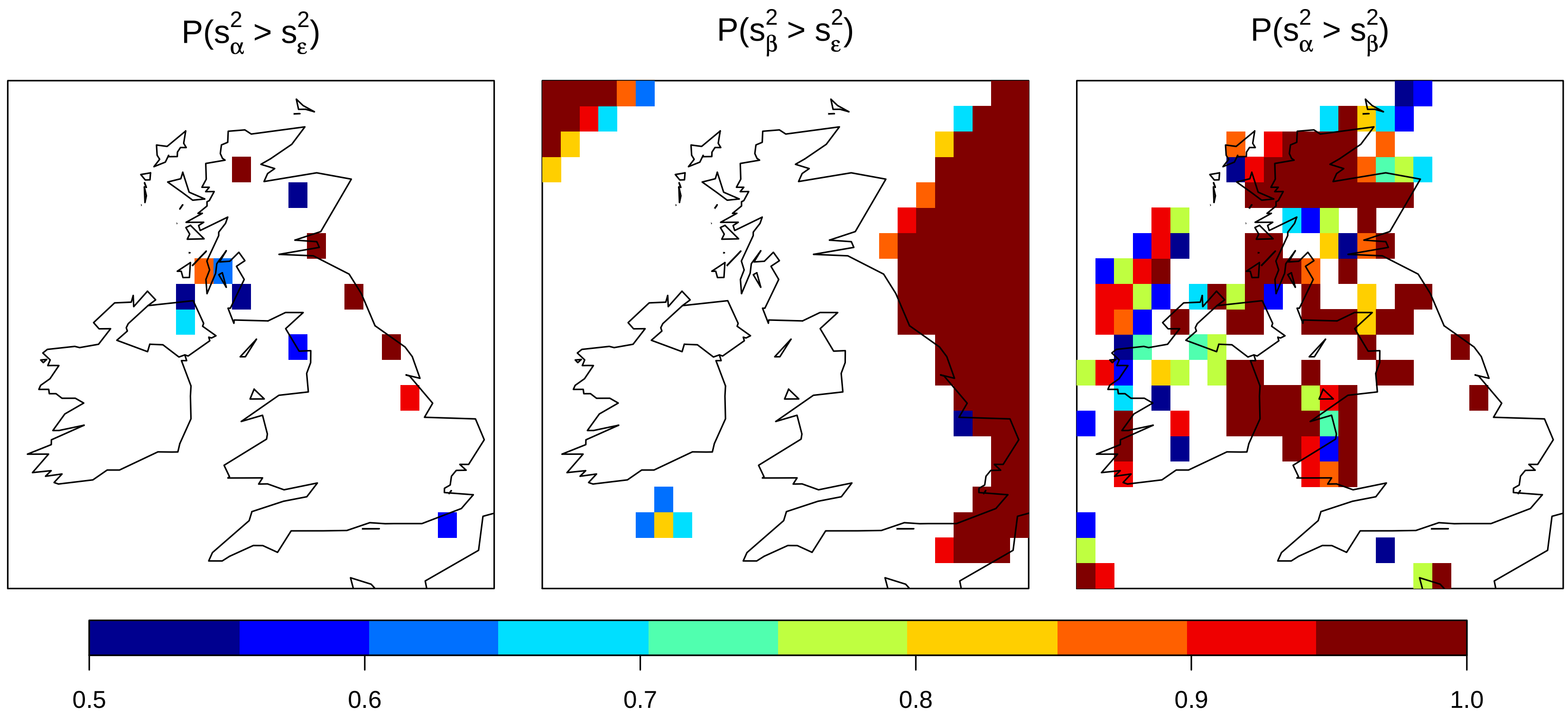
- $i, j = -1, 1$ (contrast coding)
- Hierarchical model with Gaussian process priors used for each effect.
- MCMC used to estimate parameters, posterior inference, etc.

Posterior Means



- Estimates of spatial effects.

Functional ANOVA



- Ratios of variances.