Uncertainty and regional climate experiments

Stephan R. Sain

Geophysical Statistics Project

Institute for Mathematics Applied to Geosciences

National Center for Atmospheric Research

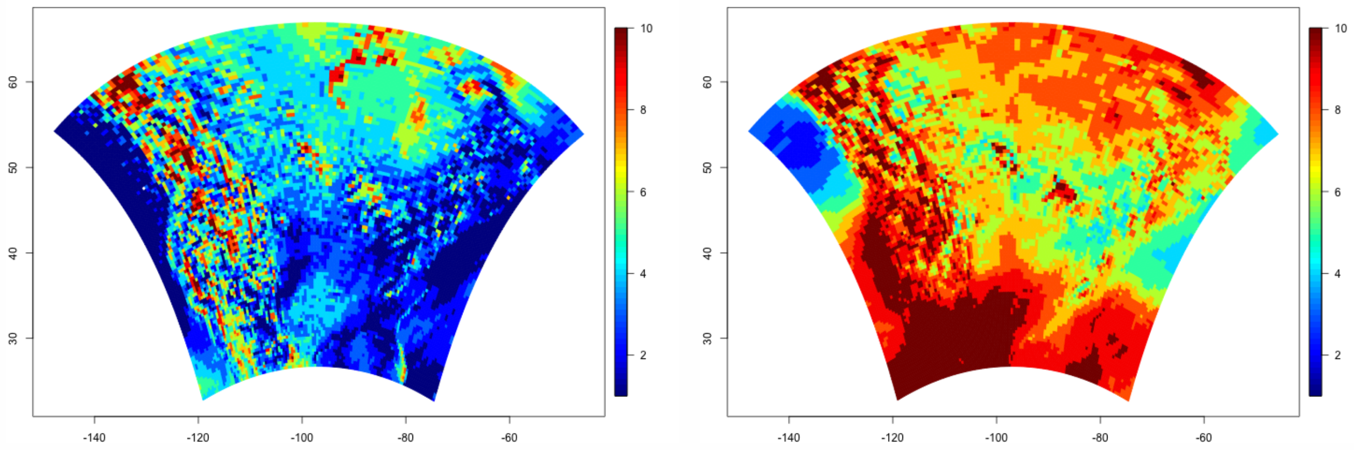
Boulder, CO

Linda Mearns, Doug Nychka, Tammy Greasby (NCAR); Reinhard Furrer (Zurich); Cari Kaufman (Berkeley); Dan Cooley (CSU); The NARCCAP Team; et al.

Supported by NSF DMS/ATM.

Goals

- Examining sources of variability/uncertainty
  - GCM, RCM, scenario, interactions, physical processes, etc.
- Projections of climate change – combining across models
- Survey of other projects
  - Multivariate, extremes, correlated models, etc.

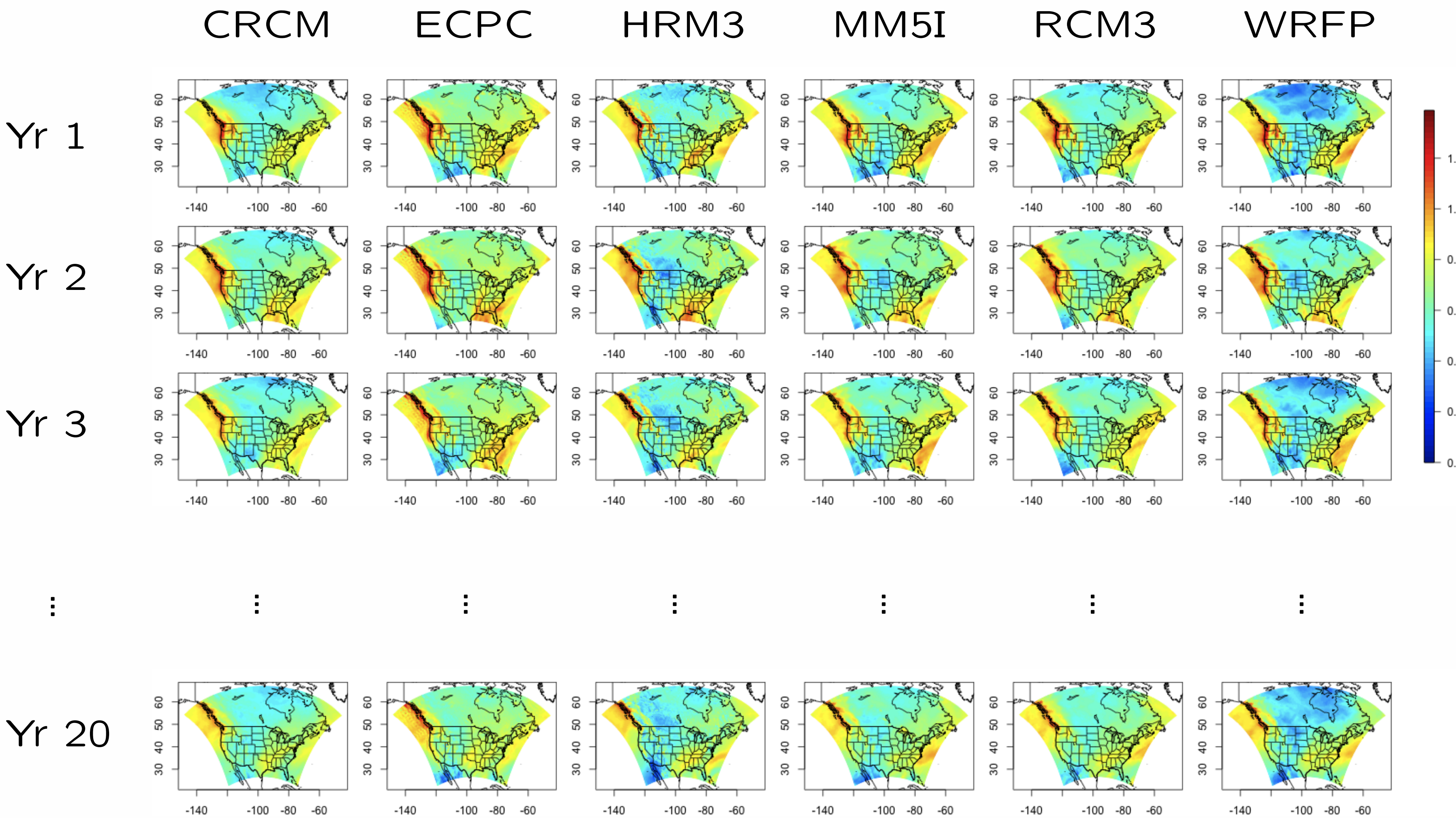


The NARCCAP Design

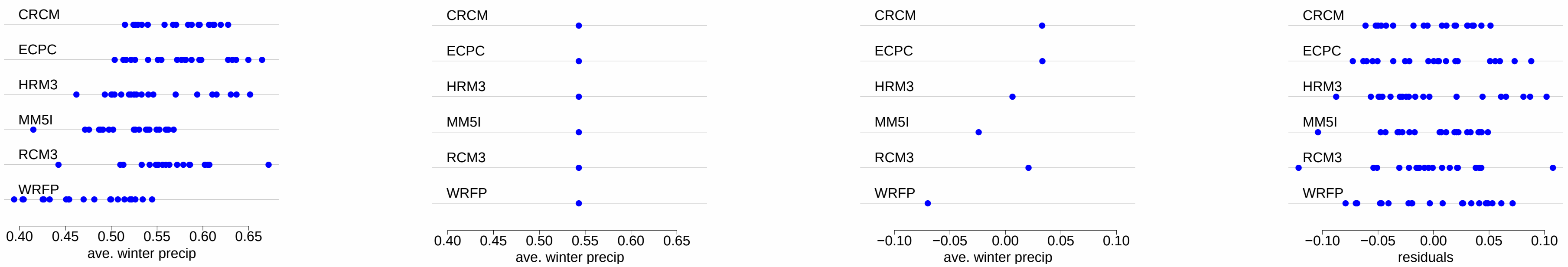
	Phase I	Phase II			
	NCEP	GFDL	CGCM3	HADCM3	CCSM
CRCM	**finished**		finished		finished
ECP2	**finished**	running		planned	
HRM3	**finished**	planned		finished	
MM5I	**finished**			planned	finished
RCM3	**finished**	finished	finished		
WRFG	**finished**		running		finished

Phase I:

- 1981 – 2000 (20 years)
- Average daily precipitation (mm) – winter (DJF)
- Interpolated to a common grid: $120 \times 98 = 11,760$ grid boxes


CRCM
ECPC
HRM3
MM5I
RCM3
WRFP
Yr 1
Yr 2
Yr 3
Yr 20
-140
-100
-80
-60
30
40
50
60
0.0
0.2
0.4
0.6
0.8
1.0
1.2

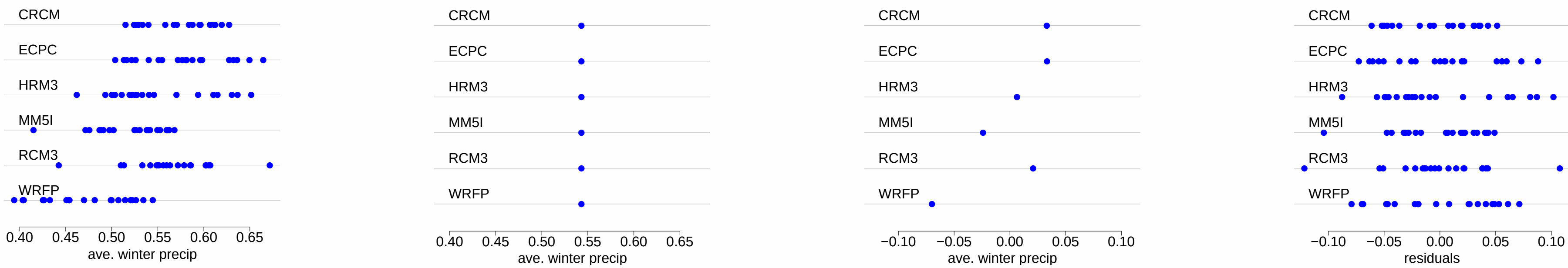
Analysis of Variance



$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

- For every grid box (this grid-box is in eastern Nebraska):
  - Y_{ij} is the (transformed) precipitation for the ith model and the jth year.
  - μ is a common mean
  - α_i is a RCM-specific effect
  - ϵ_{ij} is the error or residual

Analysis of Variance



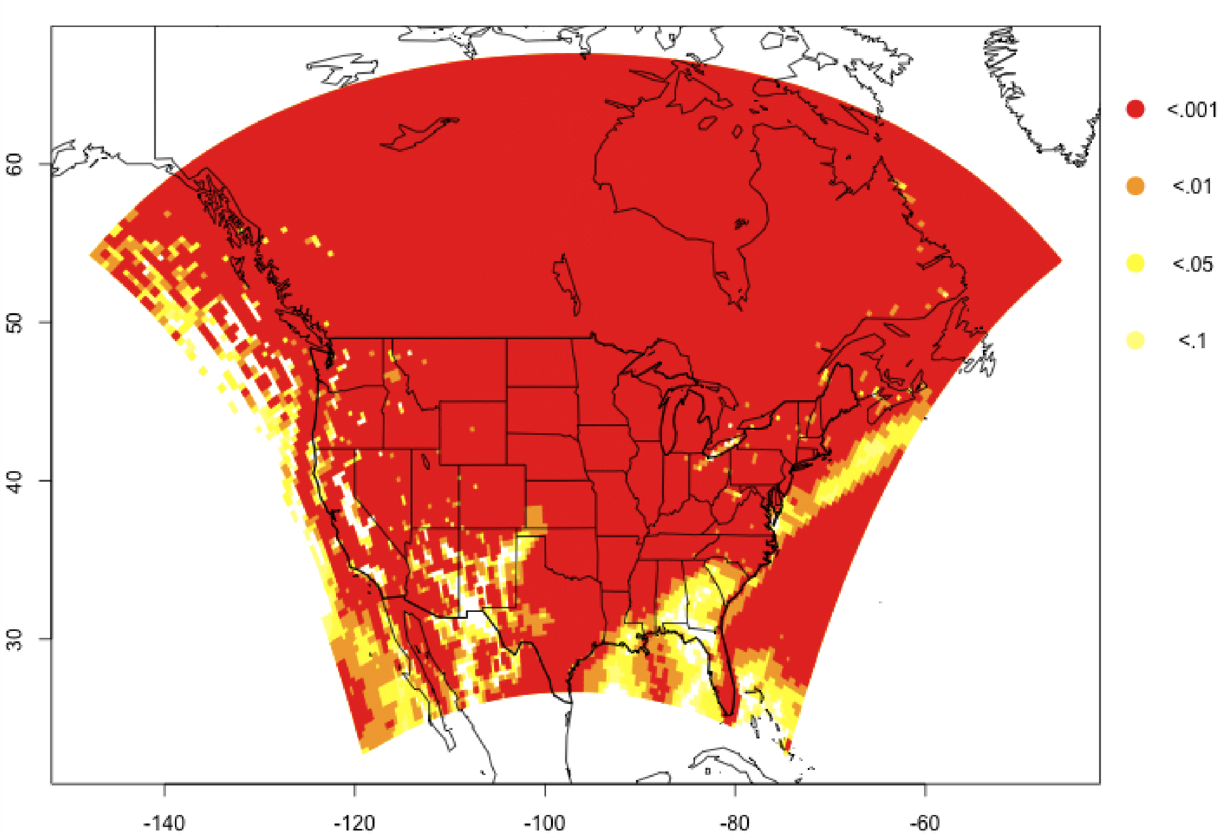
$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

- Testing the null hypothesis $H_0 : \alpha_1 = \ldots = \alpha_6 = 0$:

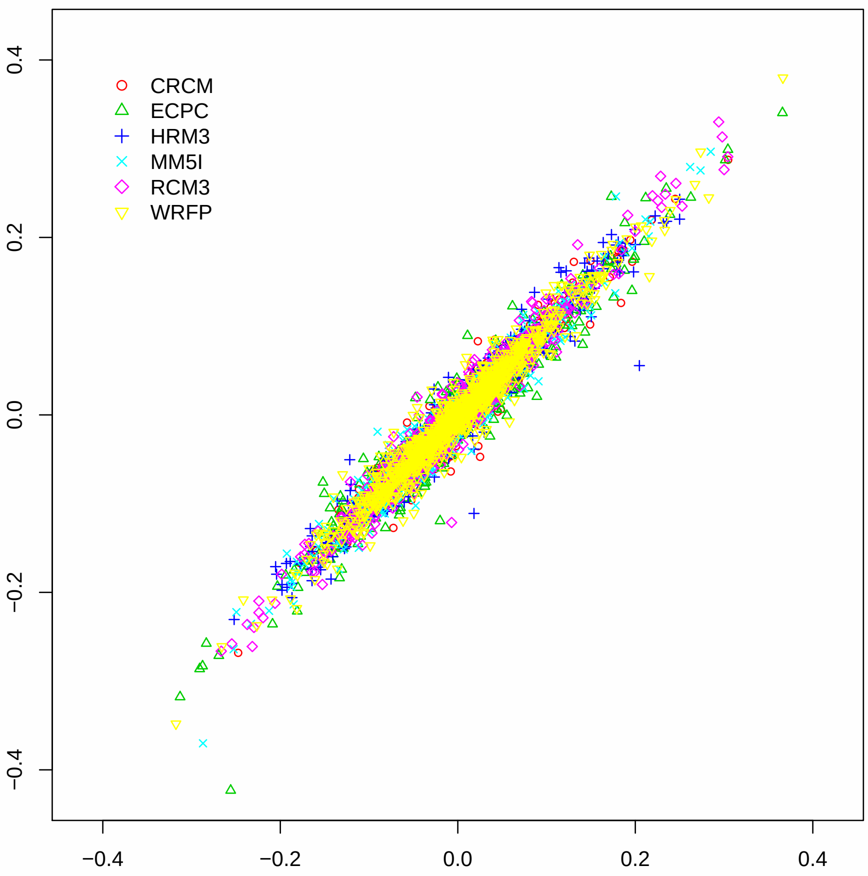
	df	SS	MS	F	p-value	
RCM	5	0.163	0.0326	15.3	1.75e-11	$\star\star\star$
Residual	114	0.243	0.00213			

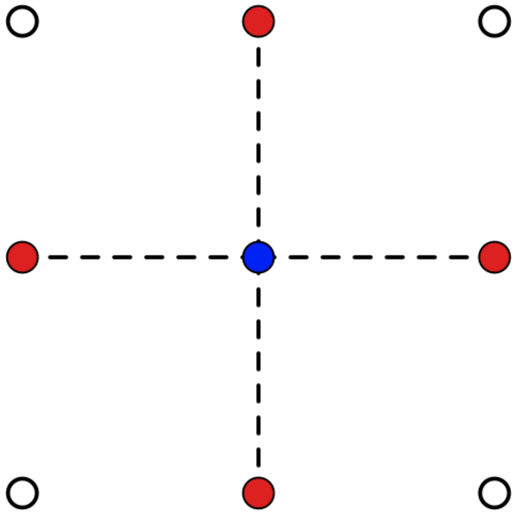
- Conclusion: strong evidence of differences in the RCM means.

Analysis of Variance



Map of pointwise p-values: strong evidence of differences in RCM means over nearly every grid box in the domain ???

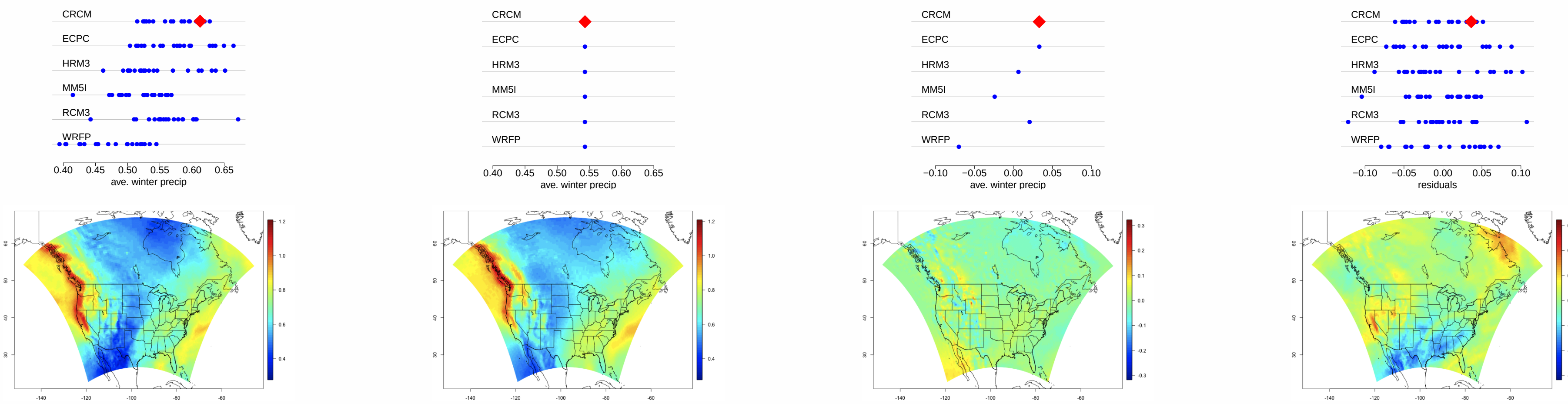




- Problem: correlated residuals at neighboring grid-boxes.

★ Result: invalid inference – any conclusions based on the p-value map are suspect.

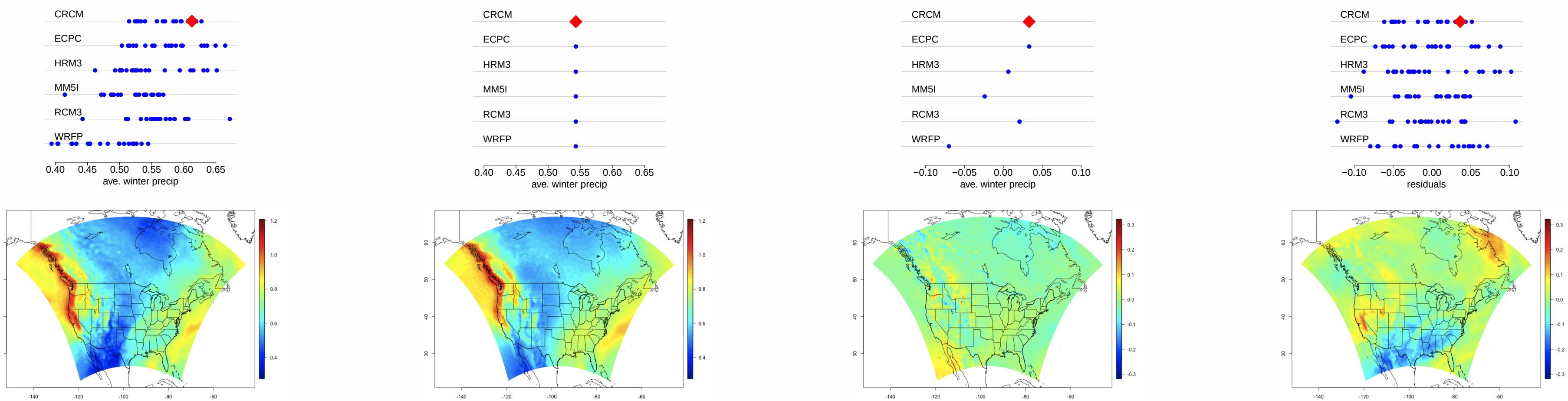
Functional Analysis of Variance



$$\mathbf{Y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{ij}$$

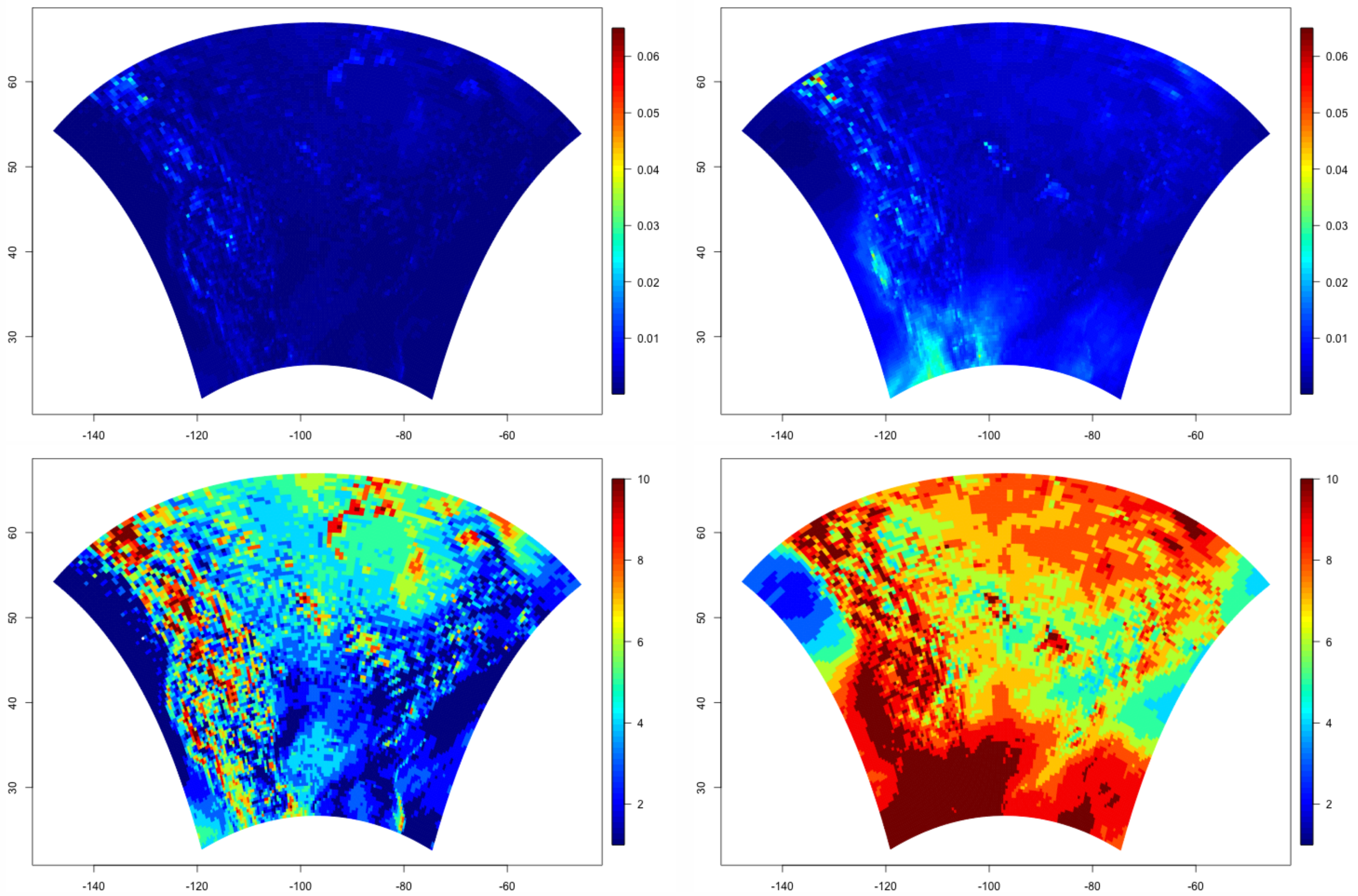
- $\mathbf{Y}_{ij}$ is the vector of (transformed) precipitation for the ith model and jth year.
- $\boldsymbol{\mu}$ is the vector mean common to all RCMs
- $\boldsymbol{\alpha}_i$ is the vector RCM-specific effect
- $\boldsymbol{\epsilon}_{ij}$ is the vector residual.

Functional Analysis of Variance

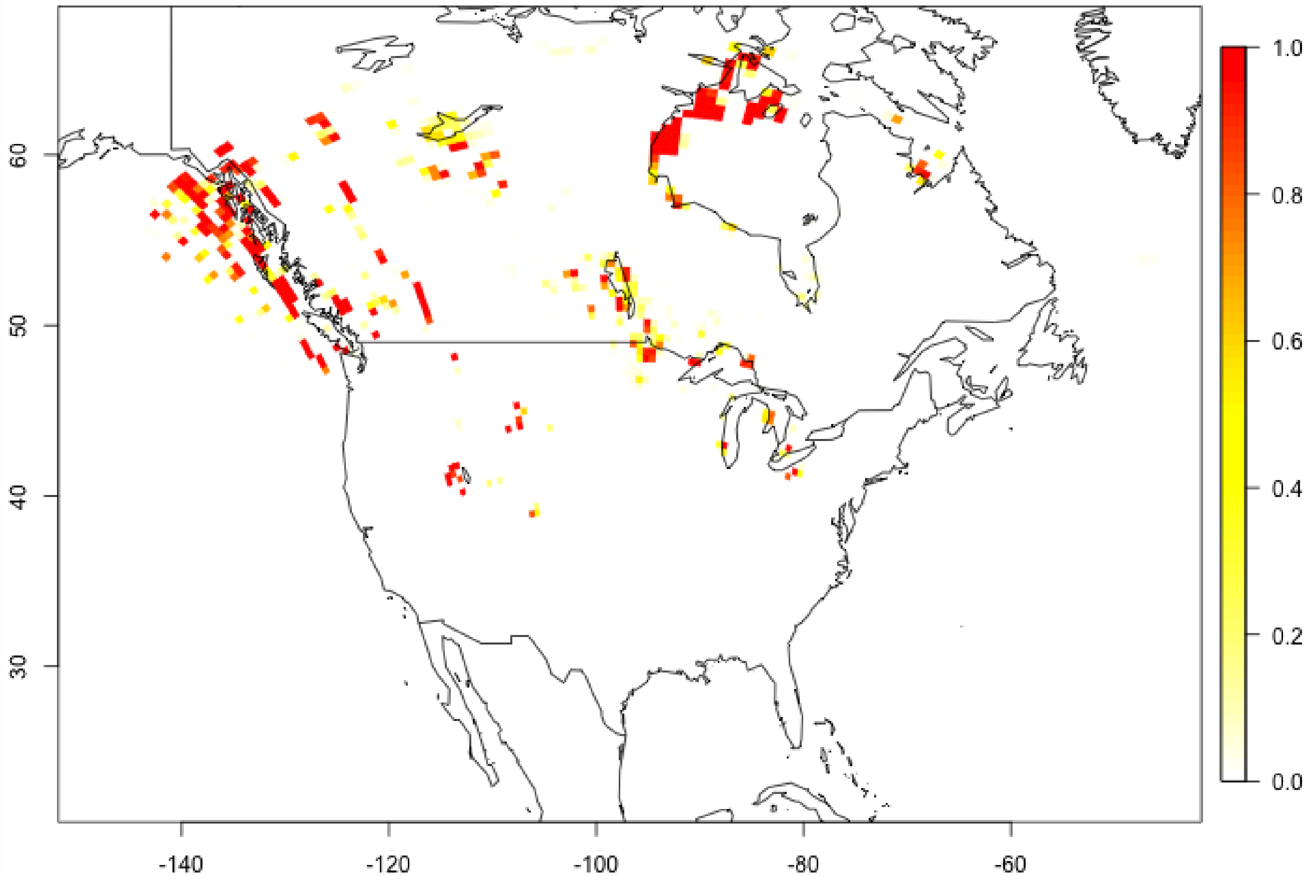


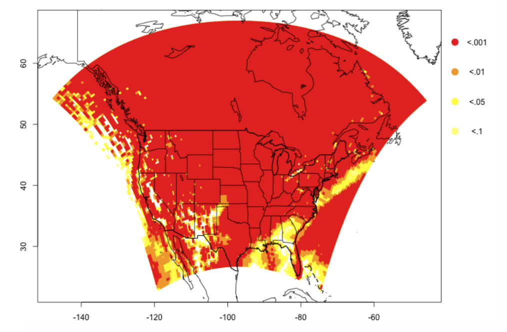
$$\mathbf{Y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{ij}$$

- The innovation is that each of these effects is a *surface*.
- Each effect is considered a realization from a random process.
- Gaussian fields are often used as prior distributions; inferences about the effects involve conditioning on the observed output fields.
- Kaufman and Sain (2010), Sain, Nychka and Mearns (2010).



Posterior means of model-to-model variation (left column) and residual or year-to-year variation (right column). Color scheme for bottom row based on quantiles.





$\widehat{P}[s_\alpha^2 > s_\epsilon^2]$

Pointwise probabilities that the model-to-model variation is larger than the year-to-year variation (analogous to small p-values in a traditional ANOVA).

Another example

- Two datasets and three regional models.
- Summer (JJA) average temperature.
  - (Seasonal temp/precip, extreme precip, heat stress, bivariate...)
- Current: 1971-2000; Future: 2041-2070.
- All models use A2 scenario for future emissions.
- All data/model output interpolated to common 114×102 grid.
  - 11628 gridboxes

Another example

Datasets:

- **CRU**: UEA Climate Research Unit's Global Climate Dataset (1970-2002).

- **UDEL**: Data from Willmott, Matsuura, and Collaborators at the University of Delaware (1979-2006).

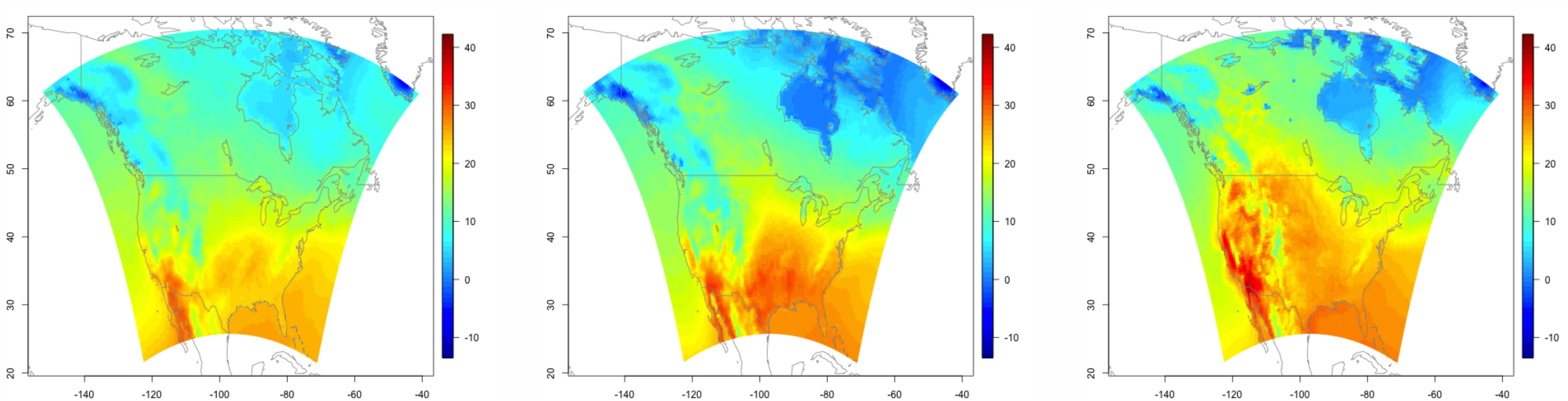
Models:

- **GFDL/RCM3**: UC Santa Cruz's Regional Climate Model; driven by NOAA's Geophysical Fluid Dynamics Laboratory GCM
- **CGCM3/CRCM**: OURANOS' Canadian Regional Climate Model; driven by CCCma's Third Generation Coupled Global Climate Model.
- **HadCM3/HRM3**: Hadley Centre's Hadley Regional Model; driven by Hadley Centre Coupled Model.

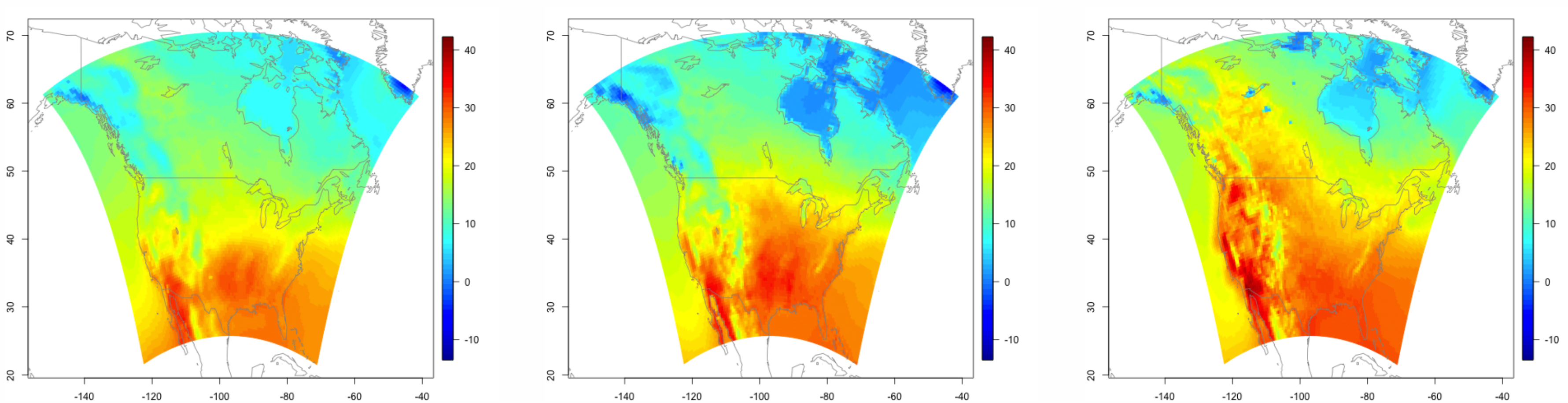
A Preview

RCM3 CRCM HRM3

2000



2070



A Hierarchical Model

Data Model:

$$\mathbf{Y}_{it} \sim \mathcal{N}\left(\mathbf{H}_i \mu_i, \boldsymbol{\Sigma}_{\mathbf{Y}_i}\right), \quad i = 1, 2; t = 1, \ldots, N_i$$

$$\mathbf{Z}^0_{it} \sim \mathcal{N}\left(\mu^0_i, \boldsymbol{\Sigma}_{\mathbf{Z}^0_i}\right), \quad i = 1, 2, 3; t = 1, \ldots, 30$$

$$\mathbf{Z}^1_{it} \sim \mathcal{N}\left(\mu^1_i, \boldsymbol{\Sigma}_{\mathbf{Z}^1_i}\right), \quad i = 1, 2, 3; t = 1, \ldots, 30$$

Process Model:

$$\boldsymbol{\mu}_i \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{\mu}_i}\right), \quad i = 1, 2$$

$$\boldsymbol{\mu}^0_i \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{\mu}_0}\right), \quad i = 1, 2, 3$$

$$\boldsymbol{\mu}^1_i \sim \mathcal{N}\left(\boldsymbol{\mu}^0_i + \boldsymbol{\Delta}, \boldsymbol{\Sigma}_{\boldsymbol{\mu}_1}\right), \quad i = 1, 2, 3$$

Prior Model:

$$\boldsymbol{\mu} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\text{NCEP}}, \boldsymbol{\Sigma}_{\boldsymbol{\mu}}\right), \quad \boldsymbol{\Sigma}_{\boldsymbol{\mu}} = \sigma^2_{\boldsymbol{\mu}} \mathbf{I}, \quad \sigma^2_{\boldsymbol{\mu}} >> 0$$

$$\boldsymbol{\Delta} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\Delta}}\right), \quad \boldsymbol{\Sigma}_{\boldsymbol{\Delta}} = \sigma^2_{\boldsymbol{\Delta}} \mathbf{I}, \quad \sigma^2_{\boldsymbol{\Delta}} >> 0$$

$$
\begin{aligned}
\mathbf{Y}_{it} &= \mathbf{H}_i\,(\boldsymbol{\mu}+\boldsymbol{\alpha}_i)+\boldsymbol{\epsilon}_{it} \\
\mathbf{Z}^0_{it} &= \boldsymbol{\mu}+\boldsymbol{\beta}_i+\boldsymbol{\epsilon}^0_{it} \\
\mathbf{Z}^1_{it} &= \boldsymbol{\mu}+\boldsymbol{\Delta}+\boldsymbol{\gamma}_i+\boldsymbol{\epsilon}^1_{it} \\
&= \boldsymbol{\mu}+\boldsymbol{\Delta}+\boldsymbol{\beta}_i+\boldsymbol{\eta}_i+\boldsymbol{\epsilon}^1_{it}
\end{aligned}
$$

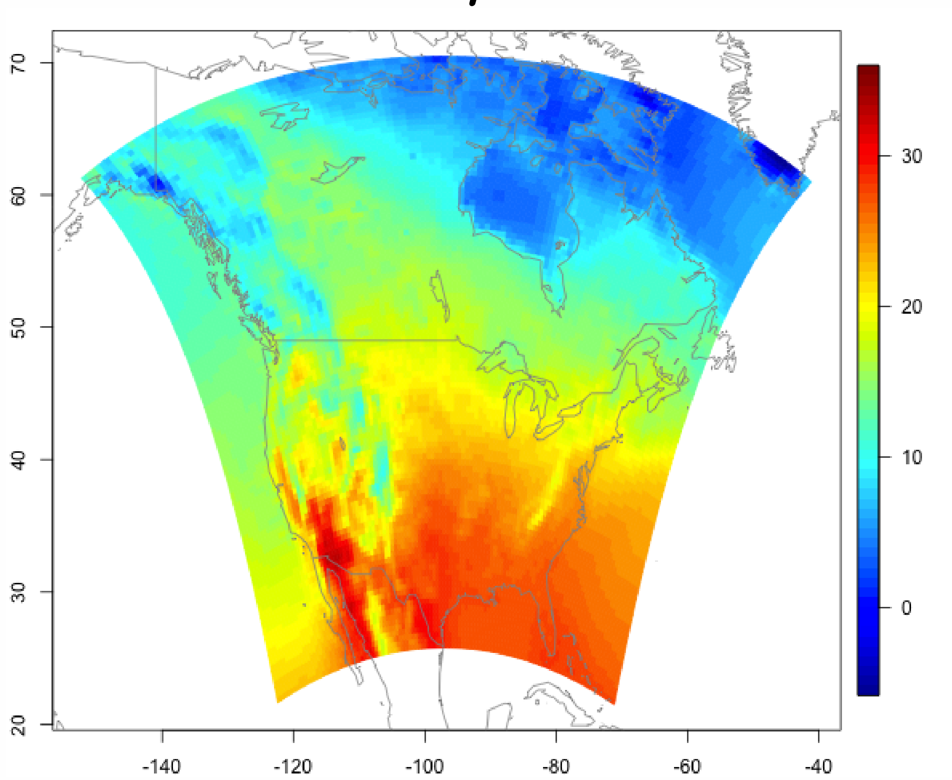
- Each yearly season of a dataset or current run of an RCM has a common "climate" ($\boldsymbol{\mu}$) plus individual model-specific deviations ($\boldsymbol{\alpha}_i/\boldsymbol{\beta}_i$) plus year-to-year variation ($\boldsymbol{\epsilon}_{it}/\boldsymbol{\epsilon}^0_{it}$).

- Each yearly season of a future run of an RCM has a common "climate" ($\boldsymbol{\mu}$) plus a common deviation or change ($\boldsymbol{\Delta}$), plus individual model-specific deviations ($\boldsymbol{\gamma}_i$), plus year-to-year variation ($\boldsymbol{\epsilon}^1_{it}$).

- Note that $\boldsymbol{\gamma}_i$ can be thought of as a model-specific deviation plus an interaction.

Posterior Mean Fields (Current)

μ

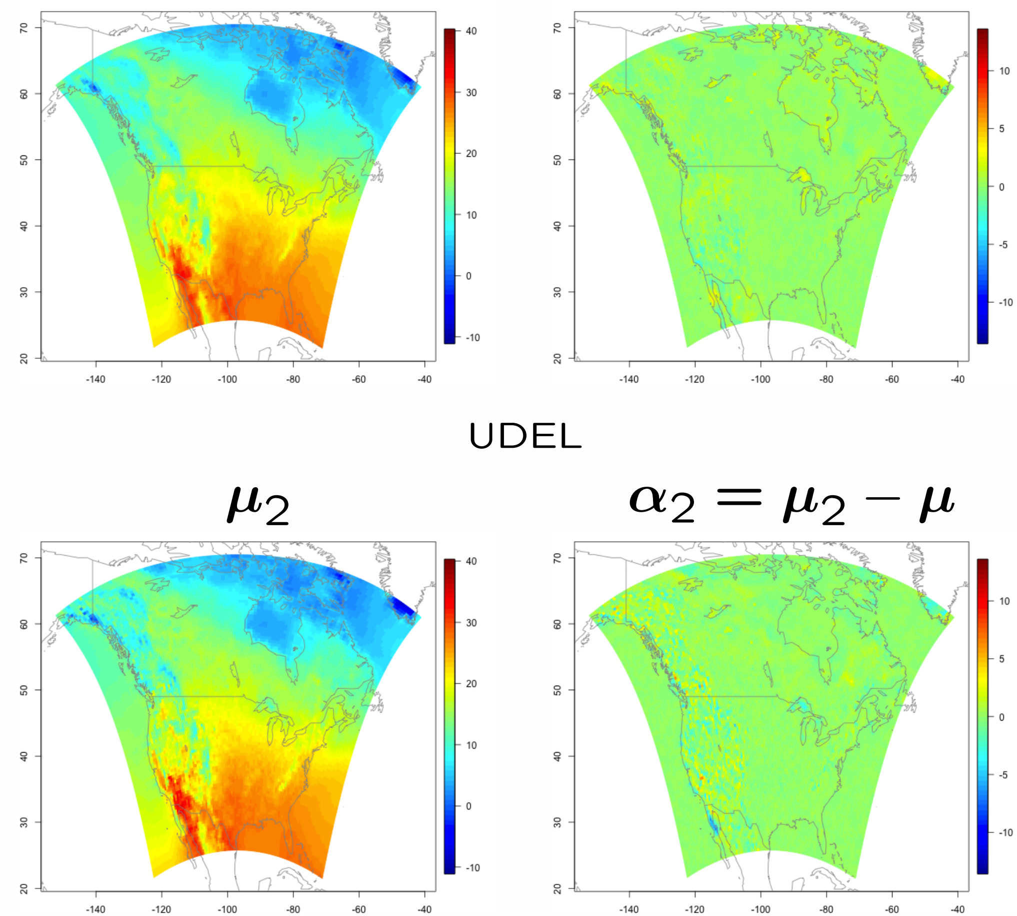


CRU

μ_1 $\qquad$ $\alpha_1 = \mu_1 - \mu$

UDEL

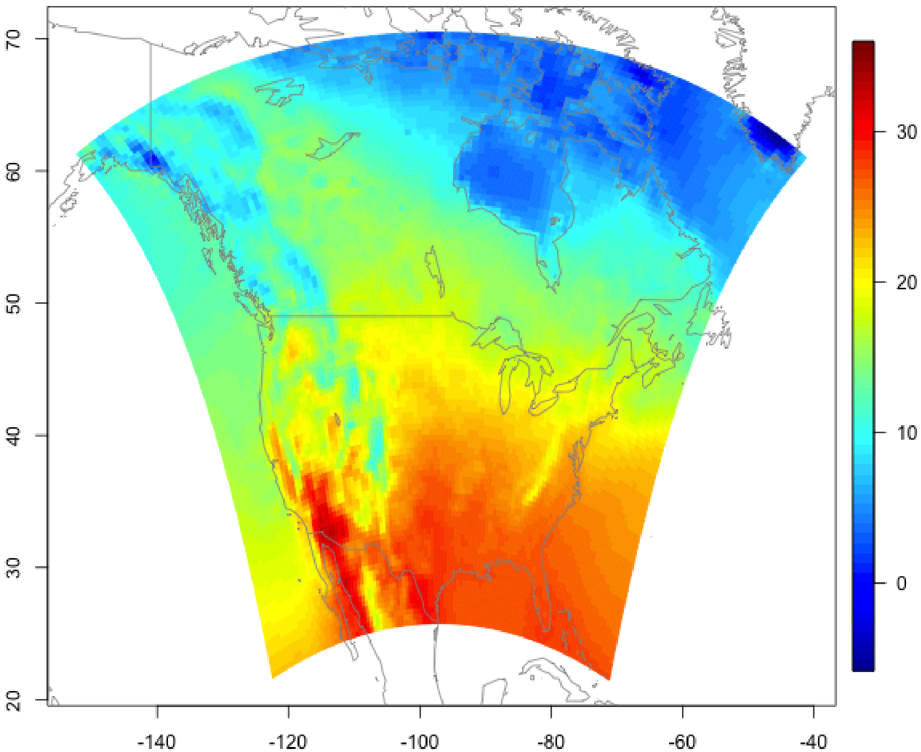
μ_2 $\qquad$ $\alpha_2 = \mu_2 - \mu$



Differences in observational datasets.

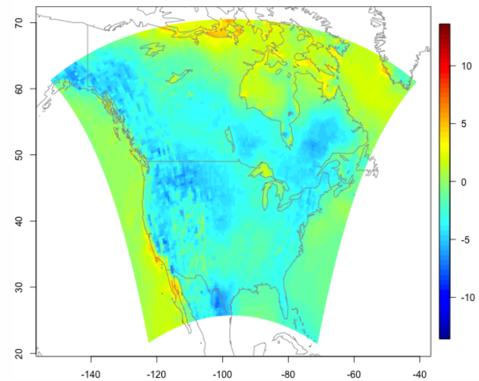
Posterior Mean Fields (Current)

μ



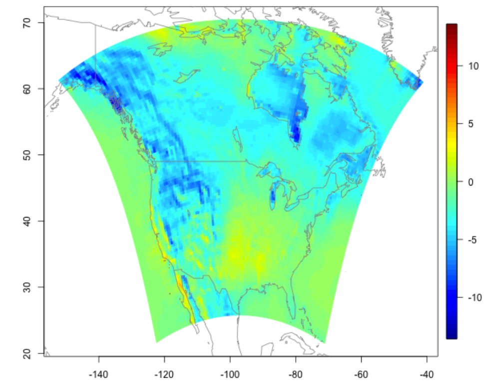
RCM3

$\beta_1 = \mu_1^0 - \mu$



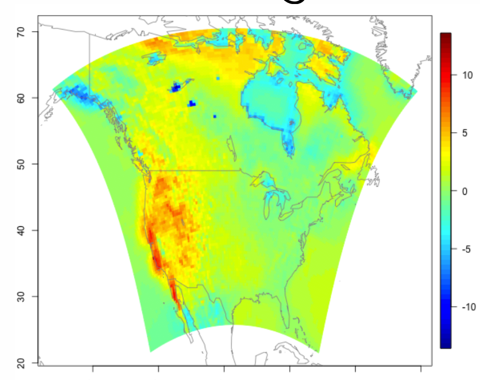
CRCM

$\beta_2 = \mu_2^0 - \mu$



HRM3

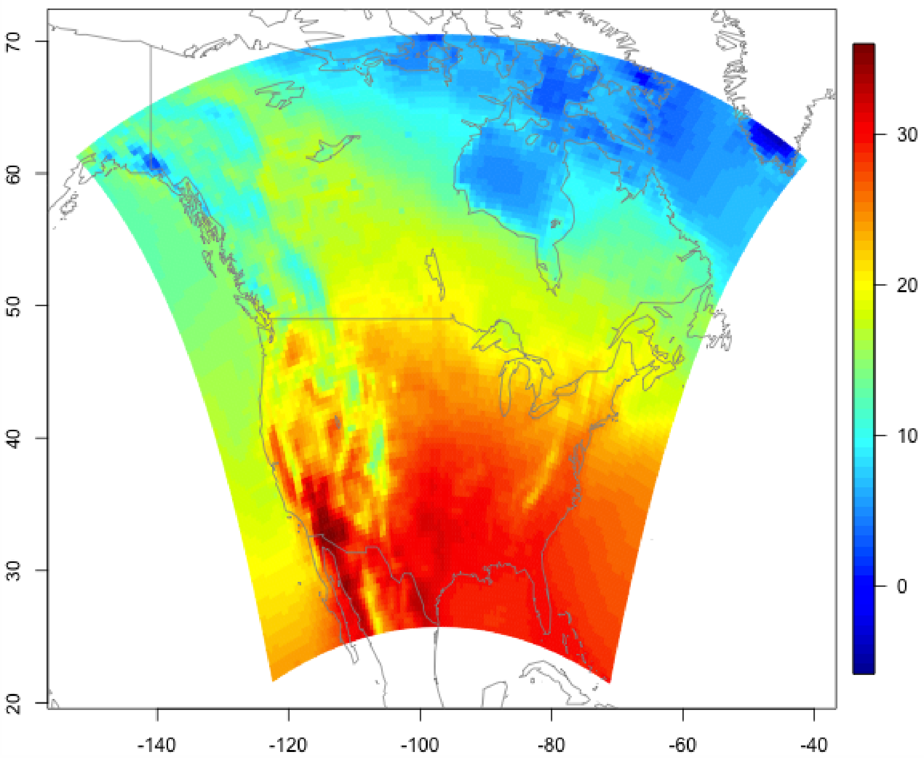
$\beta_3 = \mu_3^0 - \mu$



Differences in current runs.

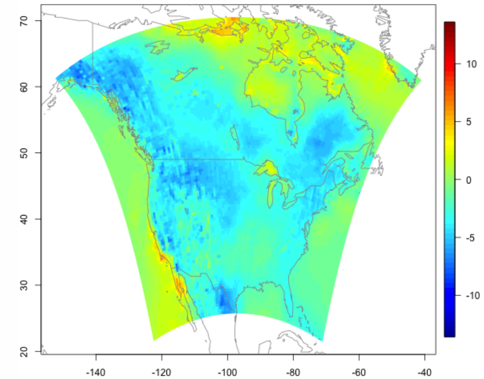
Posterior Mean Fields (Future)

$\boldsymbol{\mu}^1$



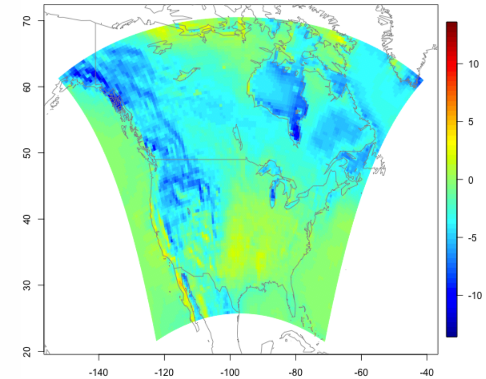
RCM3

$\gamma_1 = \boldsymbol{\mu}_1^1 - \boldsymbol{\mu}^1$



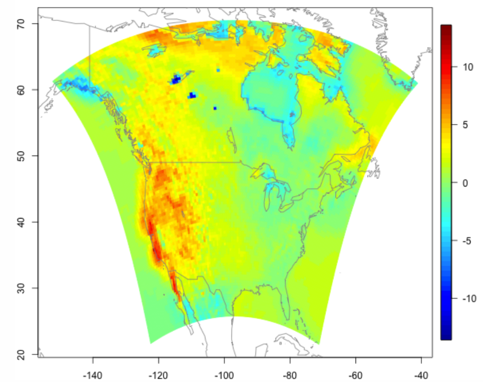
CRCM

$\gamma_2 = \boldsymbol{\mu}_2^1 - \boldsymbol{\mu}^1$



HRM3

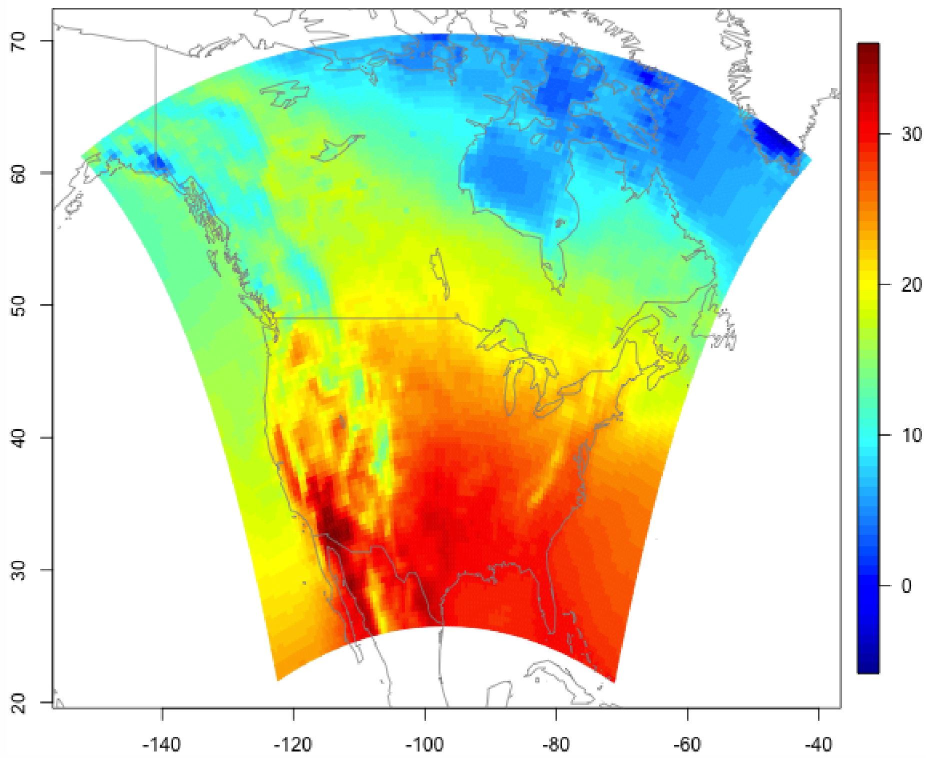
$\gamma_3 = \boldsymbol{\mu}_3^1 - \boldsymbol{\mu}^1$



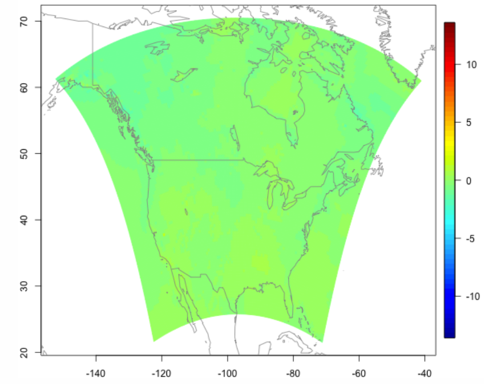
Differences in future runs.

Posterior Mean Fields (Future)

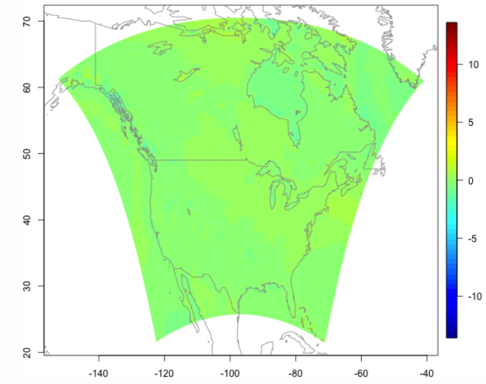
μ^1



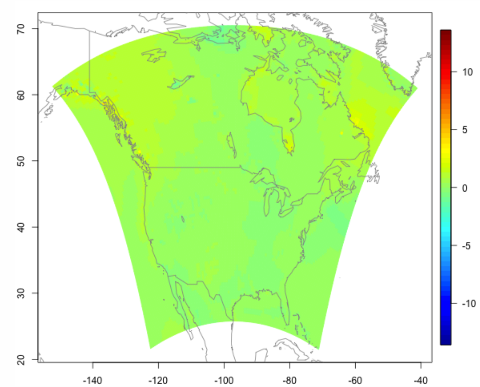
η_1 (RCM3)



η_2 (CRCM)

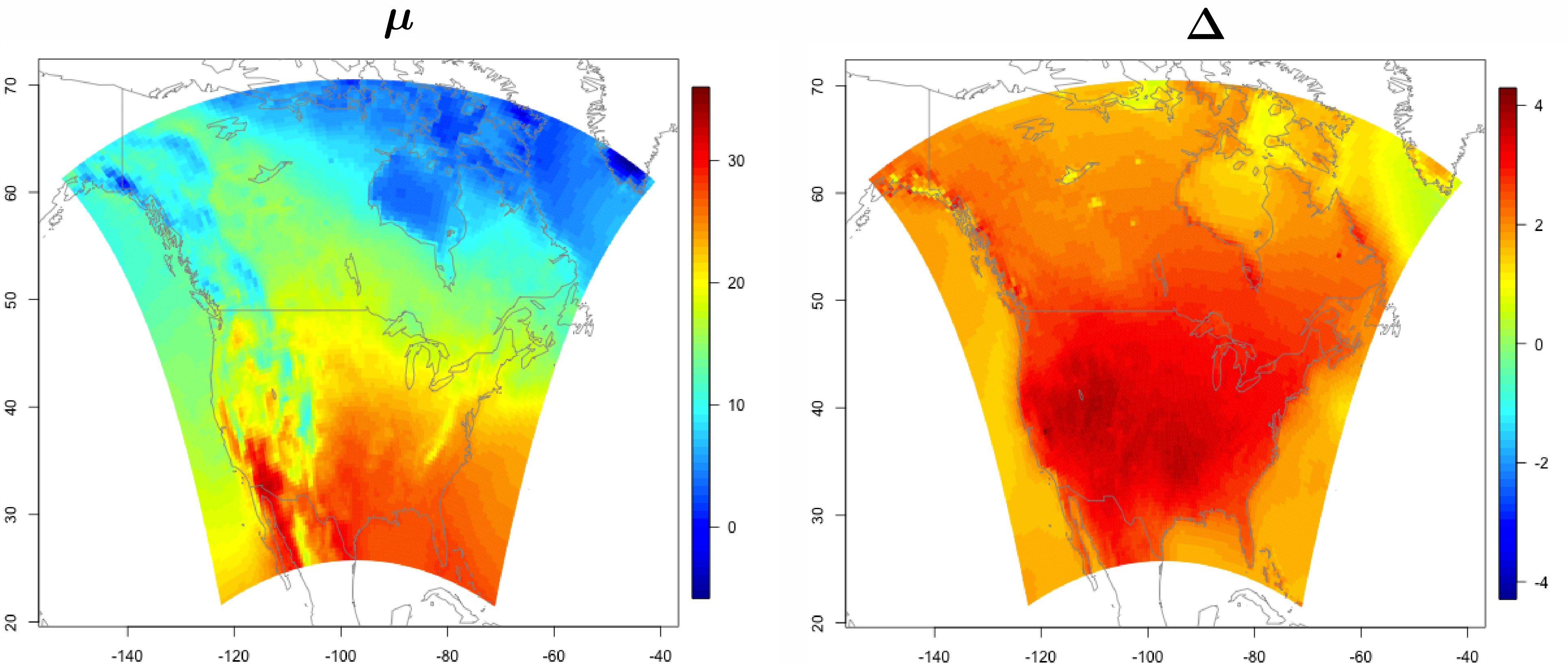


η_3 (HRM3)



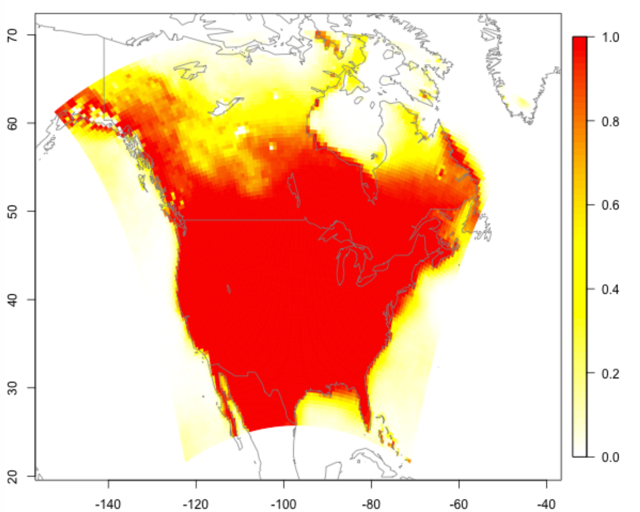
Interactions - RCMs responding to scenario forcing in different ways.

Posterior Mean Fields

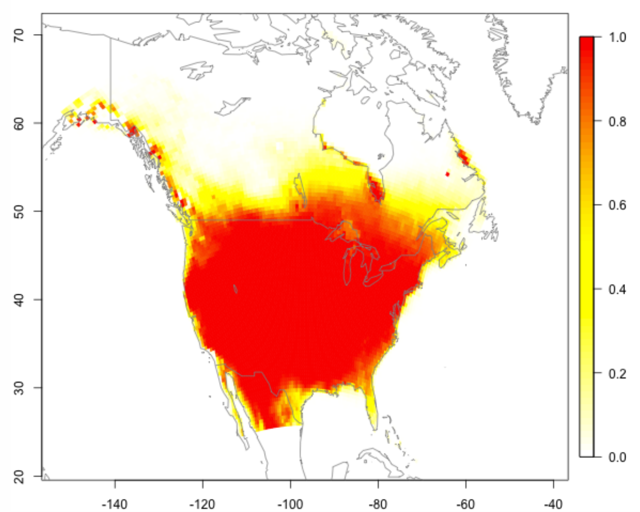


$\triangle$ Uncertainty

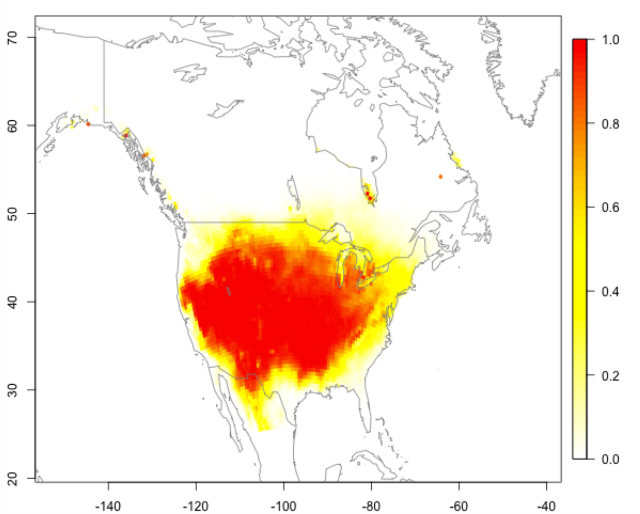
$P[\Delta > 2.0]$



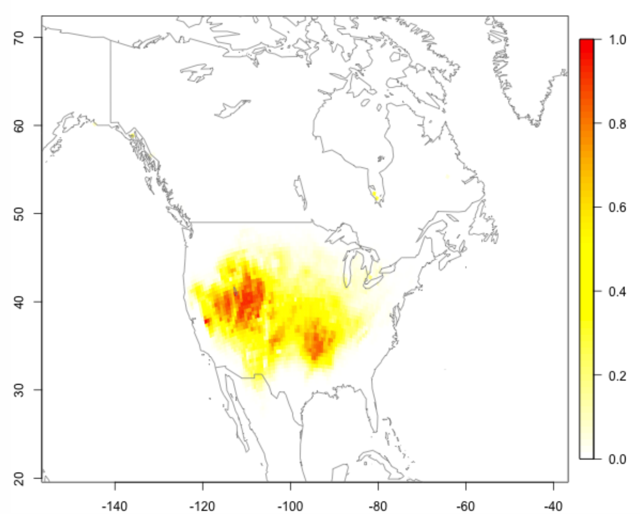
$P[\Delta > 2.5]$



$P[\Delta > 3.0]$

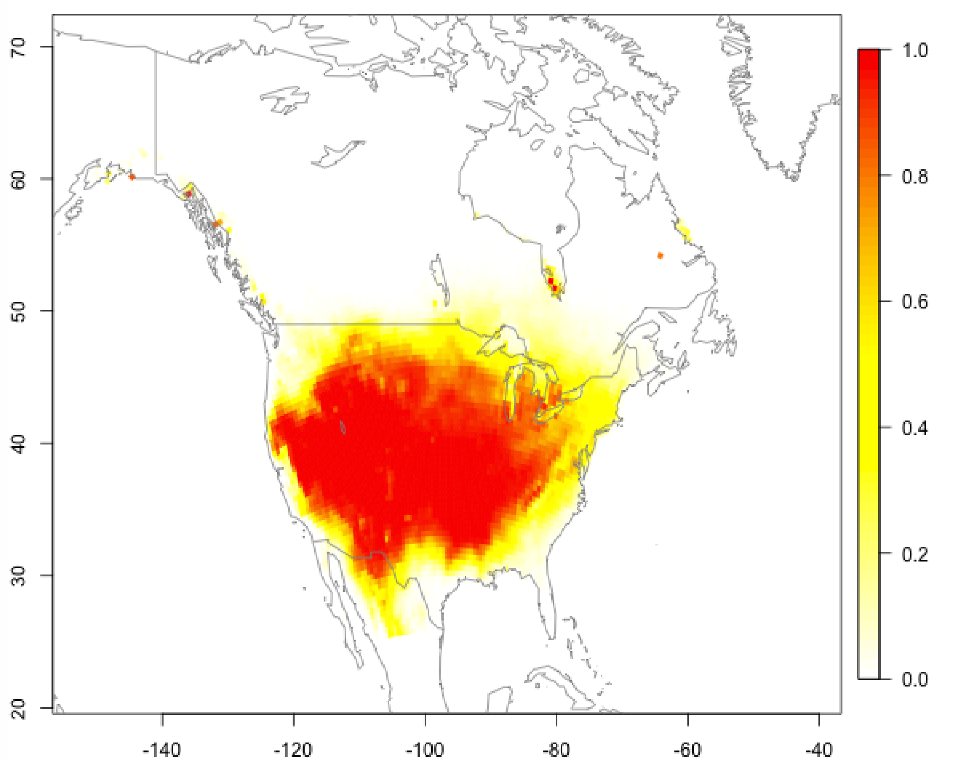


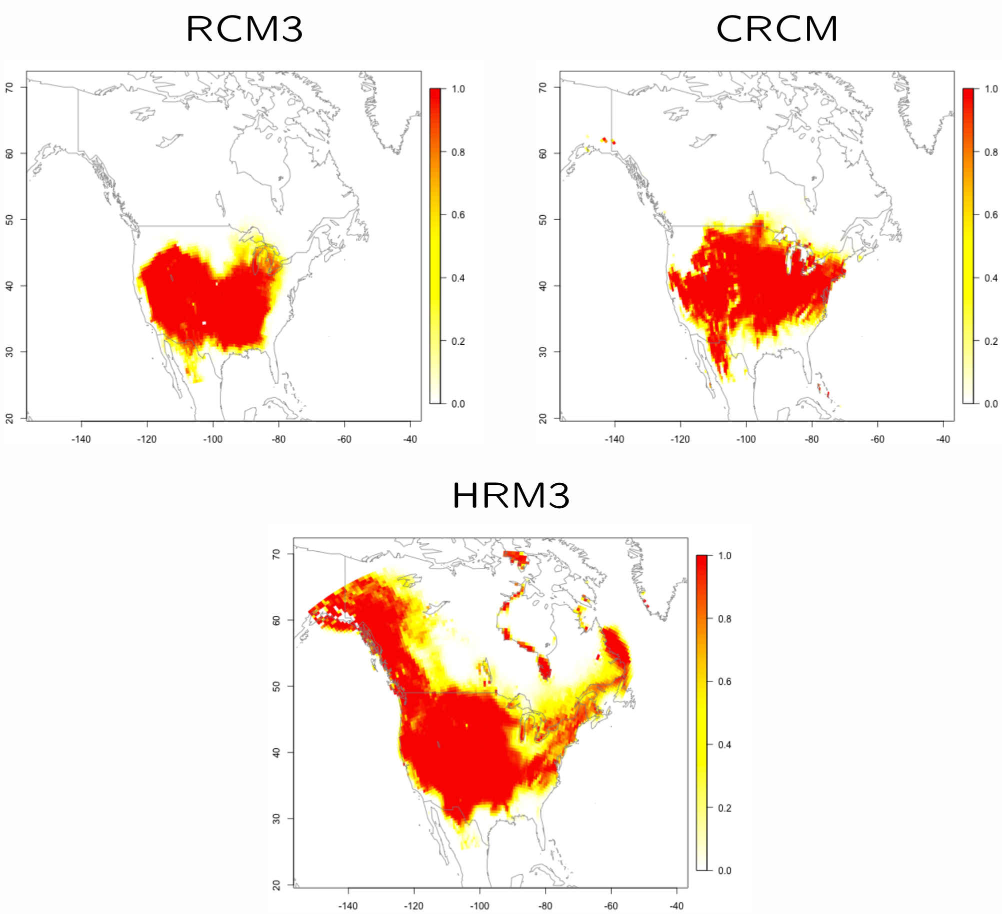
$P[\Delta > 3.5]$



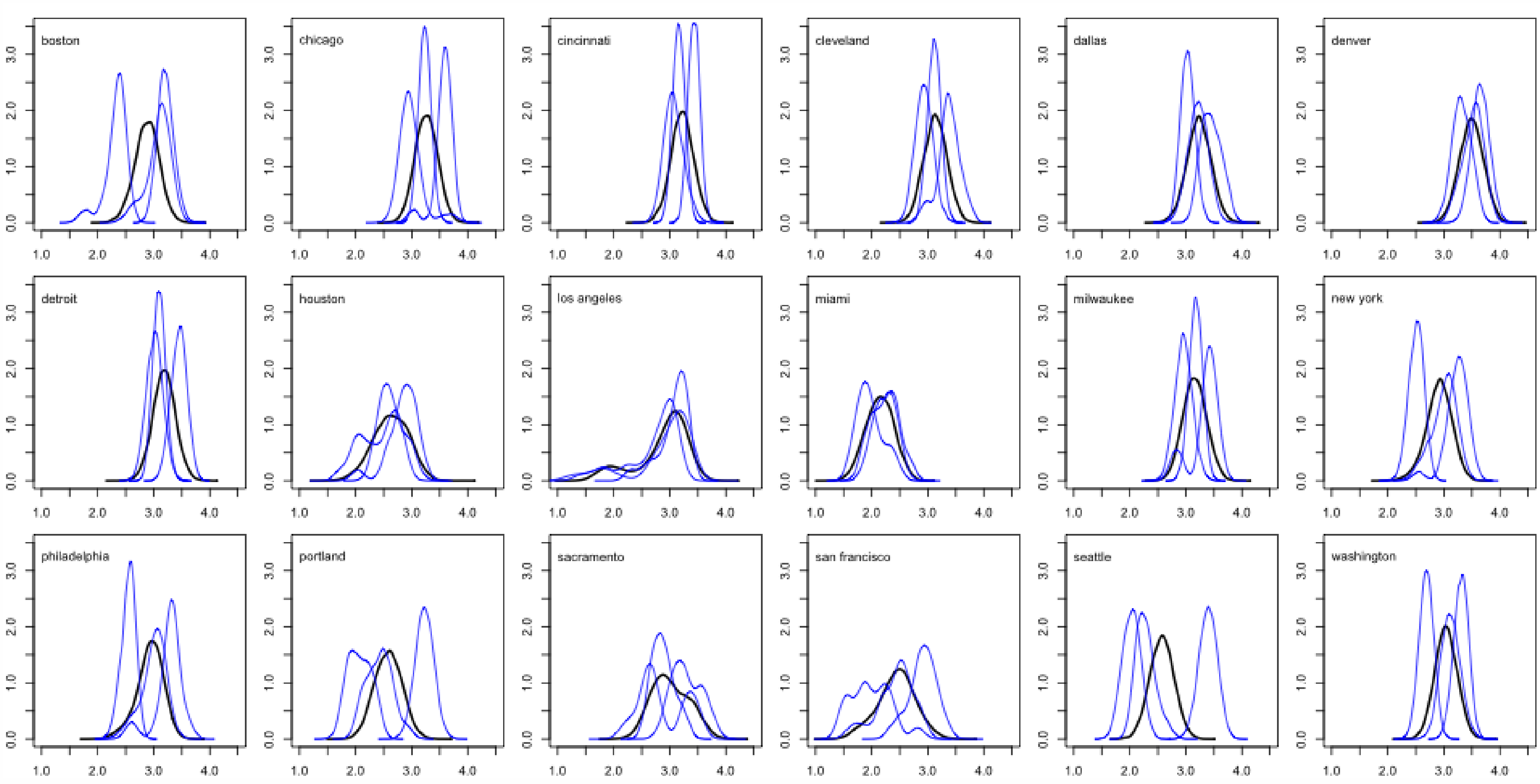
Δ Uncertainty

$P[\Delta > 3]$

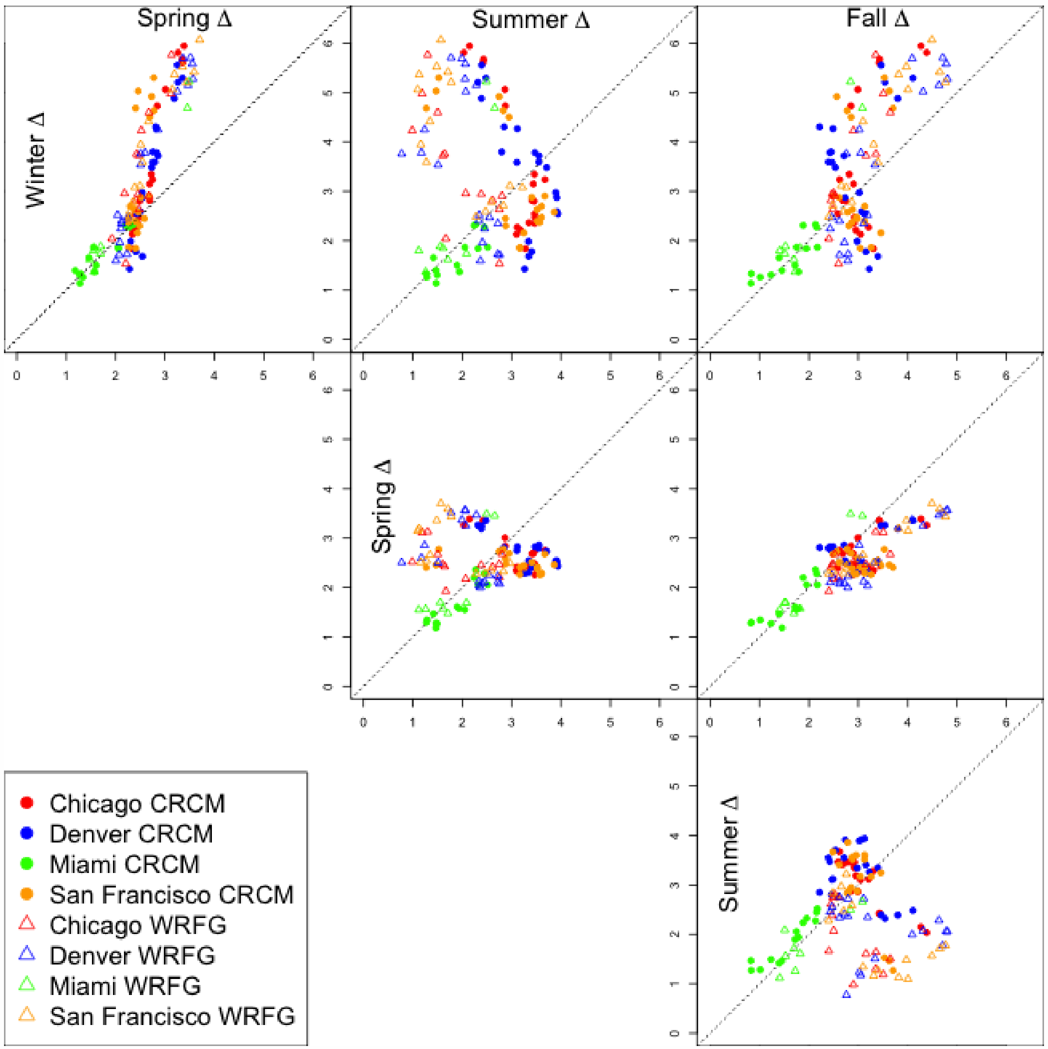




△ Uncertainty



Other Topics: Multivariate



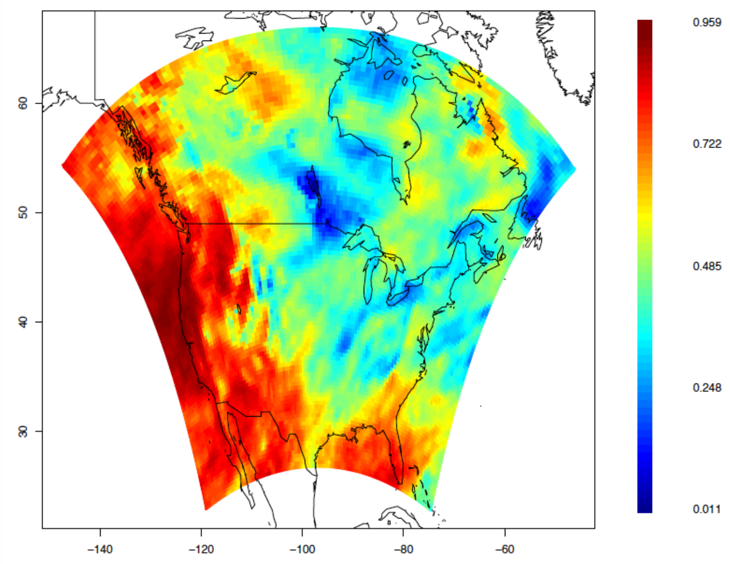
Seasonal temperature changes for selected CMSAs based on a multivariate spatial model. Extensions focused on representing a profile for whole year.

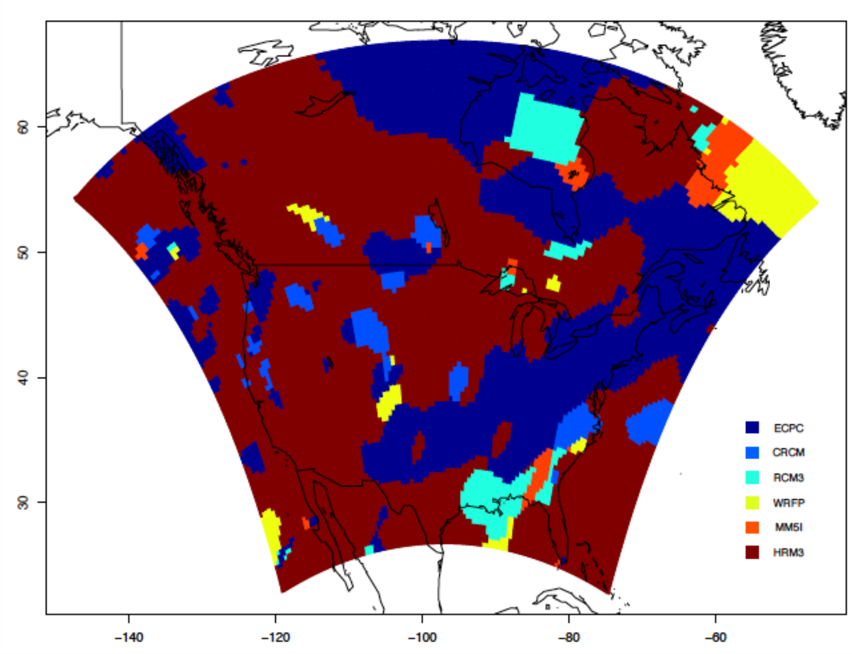
With T. Greasby, NCAR.

Other Topics: Correlated Models

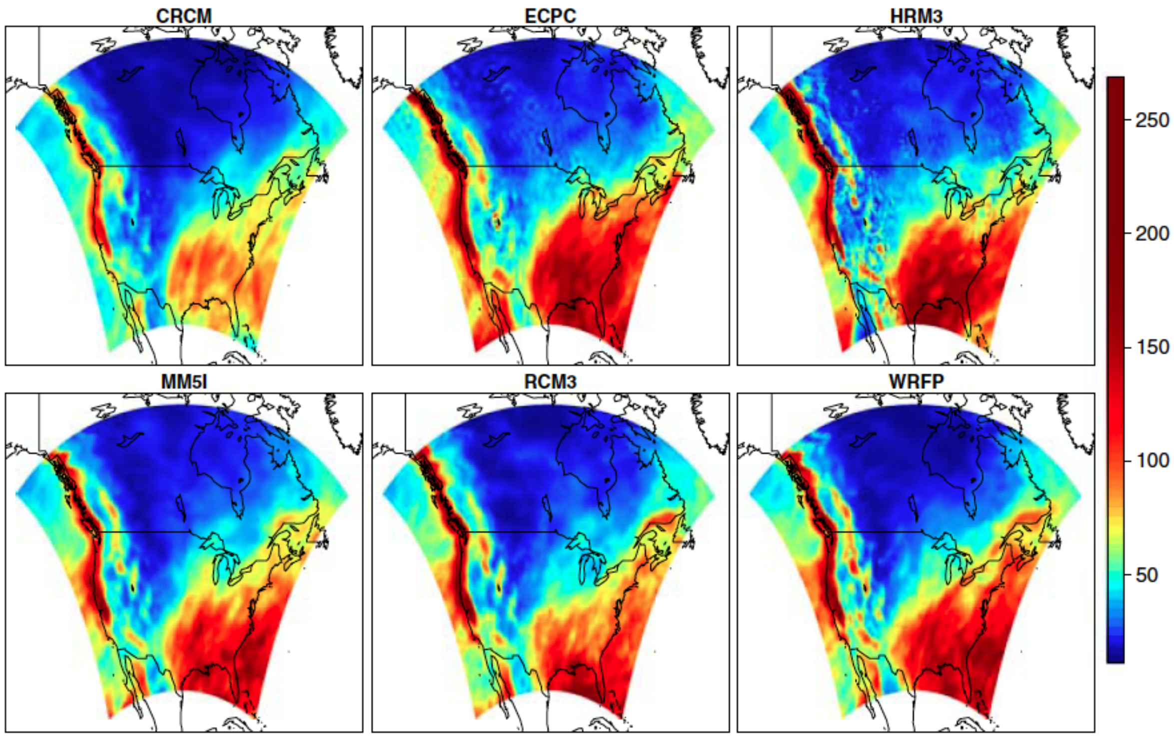
A fundamental concern with combining model output from different models is model-to-model correlations. Latent variable modeling provides a unique view into these correlations and how to combine models.

With W. Christensen, BYU.



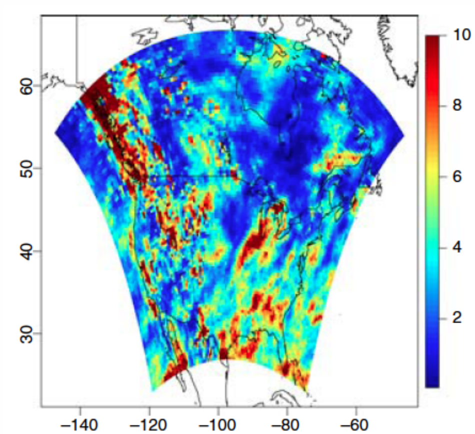


Other Topics: Extremes



100-year return levels for winter precipitation based on a hierarchical Bayesian spatial model based on the GEV representation for extremes.

With D. Cooley, CSU.



Questions?

Many opportunities for visits and collaboration: ASP, RSVP, SIParCs, GSP, IMAGe, Theme-of-the-Year,...

`ssain@ucar.edu`
`http://www.image.ucar.edu/~ssain`

Thank You!

Kaufman and Sain (2010), “Bayesian functional ANOVA modeling using Gaussian process prior distributions,” Bayes Anal, 5, 123-150, doi:10.1214/10-BA505.

Schliep, Cooley, Sain, and Hoeting (2010), “A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling,” Extremes, 13, 219-239, doi:10.1007/s10687-009-0098-2.

Sain and Furrer (2010), “Combining climate model output via model correlations,” Stoch Env Res Risk A, 24, 821-829, doi:10.1007/s00477-010-0380-5.

Cooley and Sain (2010), “Spatial hierarchical modeling of precipitation extremes from a regional climate model,” JABES, 15, 381-402, doi:10.1007/s13253-010-0023-9

Furrer and Sain (2010), “spam: A sparse matrix R package with emphasis on MCMC methods for Gaussian Markov random felds,” J Stat Softw, `http://www.jstatsoft.org/v36/i10`.

Sain, Furrer, and Cressie (2011), “A spatial analysis of multivariate output from regional climate models,” AOAS, 5, 150-175, doi:10.1214/10-AOAS369.

Christensen and Sain (2010), “Spatial latent variable modeling for integrating output from multiple climate models,” Math Geosci, doi:10.1007/s11004-011-9321-1.

Sain, Nychka, and Mearns (2010), “Functional ANOVA and regional climate experiments: A statistical analysis of dynamic downscaling,” Environmetrics, doi:10.1002/env.1068

Kang, Cressie, and Sain (2011), “Combining outputs from the NARCCAP regional climate models using a Bayesian hierarchical model,” Applied Statistics, under revision.