NARCCAP UQ and Stat Stuff

Stephan R. Sain

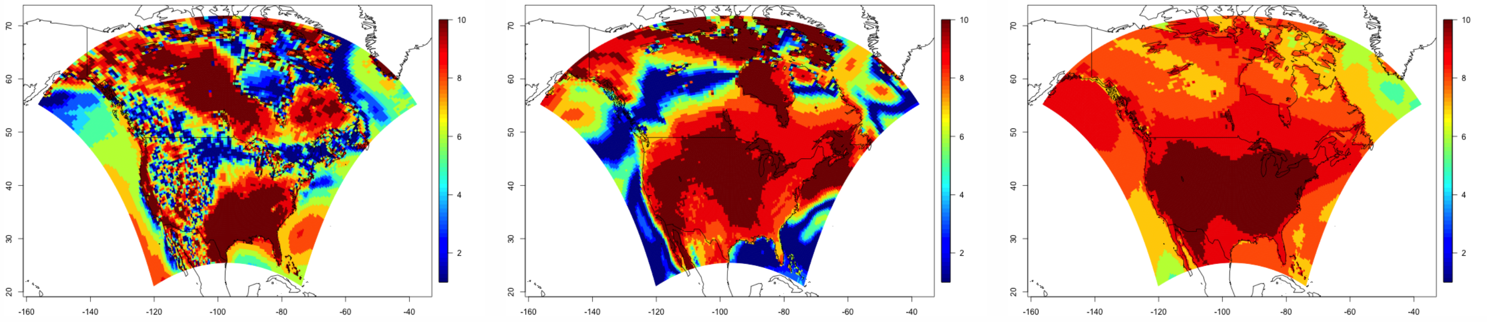
Geophysical Statistics Project
Institute for Mathematics Applied to Geosciences
National Center for Atmospheric Research
Boulder, CO

Tammy Greasby, Matt Heaton (NCAR); Reinhard Furrer, Steve Geinitz (Zurich); Cari Kaufman (Berkeley); Dan Cooley , Grant Weller (CSU)

Supported by NSF DMS/ATM.

Outline and Goals

- Understanding sources of variation – functional ANOVA.
  - Implications for design of future experiments, “completing” the table, etc.
- Delivering climate change information and uncertainty – temperature profiles.
- Combining information and model weighting.
- Others – extremes, conveying uncertainty, stat methods, etc.

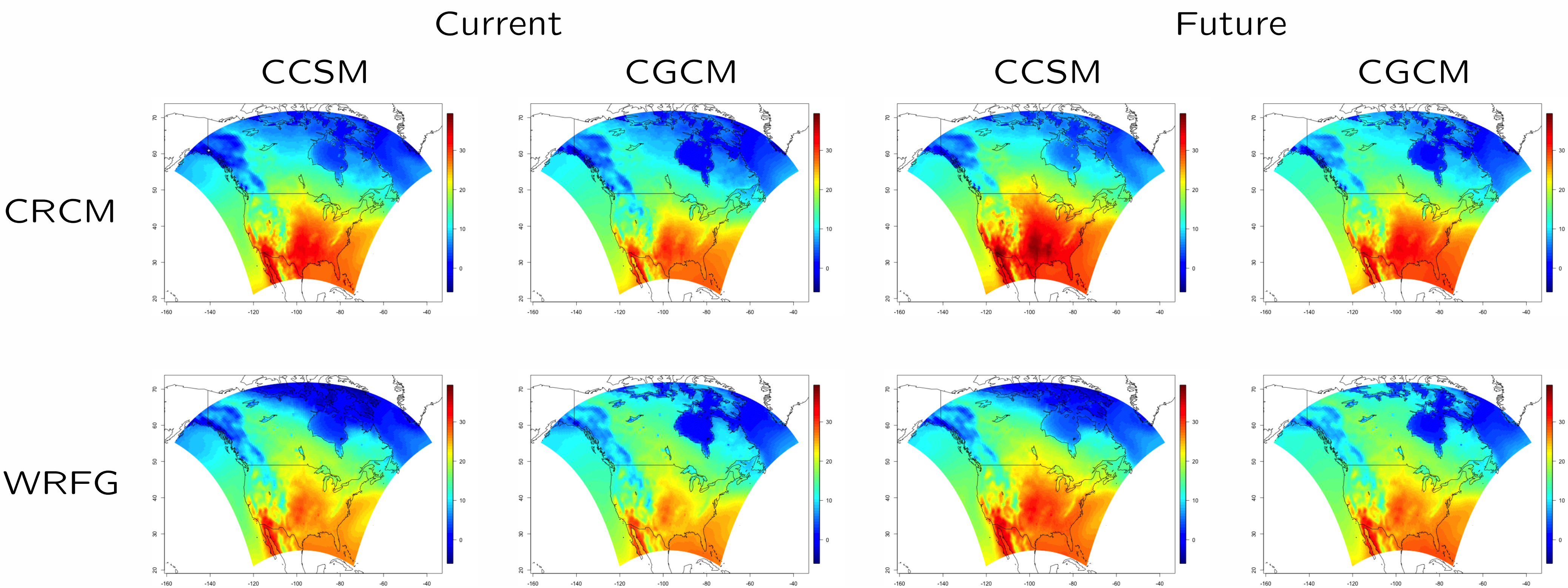


The NARCCAP Design

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

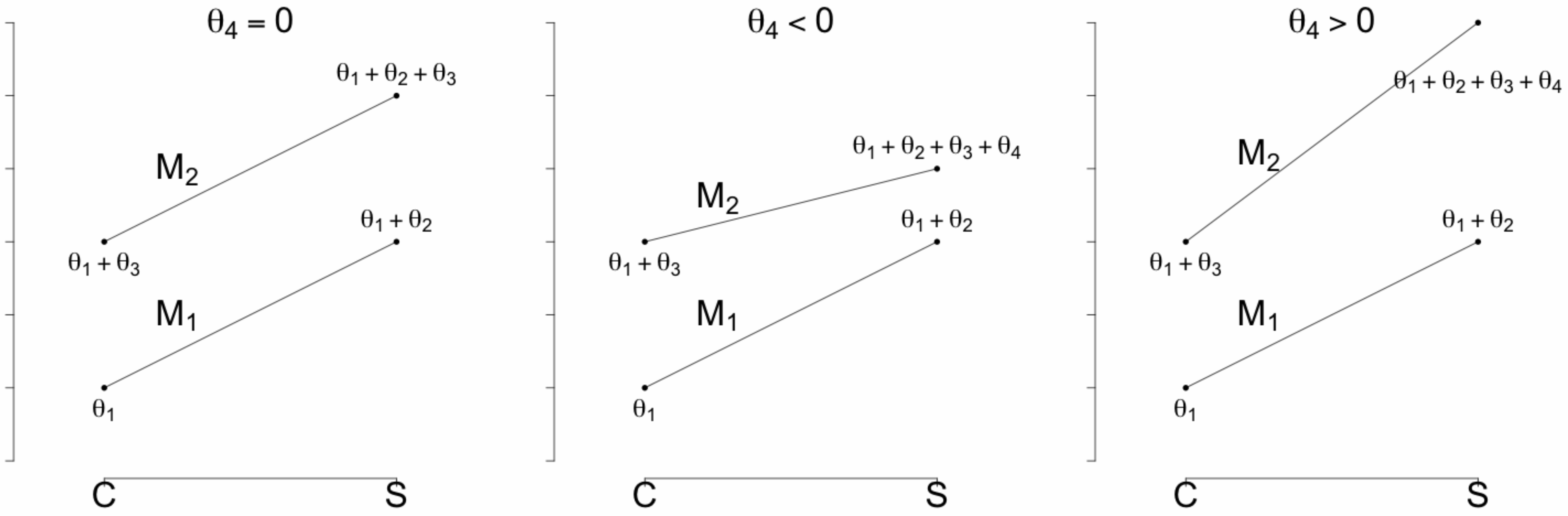
- Phase I: 1980-2000
- Phase II: 1971-2000 (Current), 2041-2070 (Future)
- All future runs use the A2 scenario
- Focus on seasonal summaries

A 2^3 example



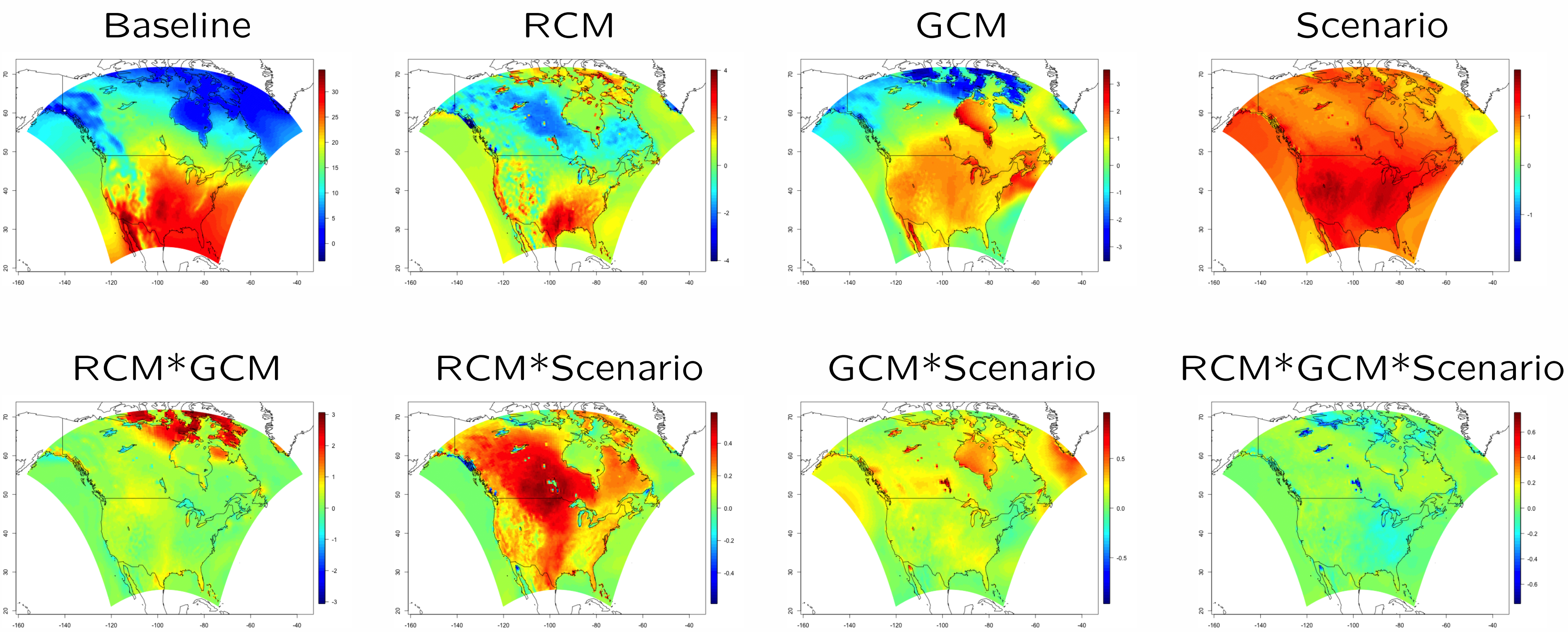
30-year average summer (JJA) temperature.

Reinhard Furrer, Steve Geinitz (Zurich), Kari Kaufman (Berkeley)



Does Model 2 respond to the forcing in the same way that Model 1 does ($\theta_4 = 0$)? Or does it respond in a way that is systematically different ($\theta_4 \neq 0$)?

A 2^3 example


Baseline
RCM
GCM
Scenario
RCM*GCM
RCM*Scenario
GCM*Scenario
RCM*GCM*Scenario

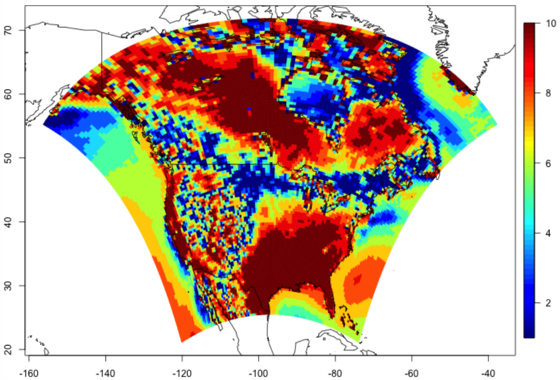
	DF	SS	MS	F	p-value	
GCM	1	5.51	5.51	2.01	0.16	
RCM	1	493.37	493.37	180.27	<2.2e-16	$\star\star\star$
Scenario	1	387.73	387.73	141.67	< 2.2e-16	$\star\star\star$
GCM:RCM	1	47.43	47.43	17.33	4.42e-05	$\star\star\star$
GCM:Scenario	1	53.92	53.92	19.70	1.4e-05	$\star\star\star$
RCM:Scenario	1	2.89	2.89	1.06	0.30	
GCM:RCM:Scenario	1	0.08	0.08	0.03	0.86	
Residuals	232	634.94	2.74			

An ANOVA table for a single grid near the S. Dakota, N. Dakota, Minnesota border.

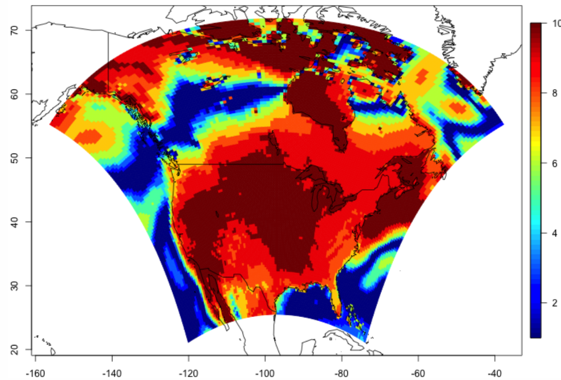
Interpretation complicated by spatial dependence, large numbers of tests, etc.

Issues can be addressed by treating the “effects” as spatial processes in a functional ANOVA framework embedded in a Bayesian hierarchical model.

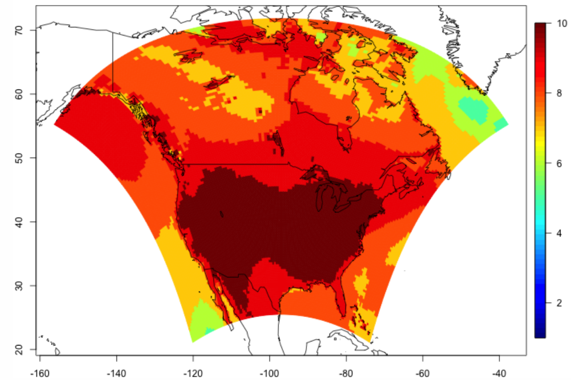
RCM



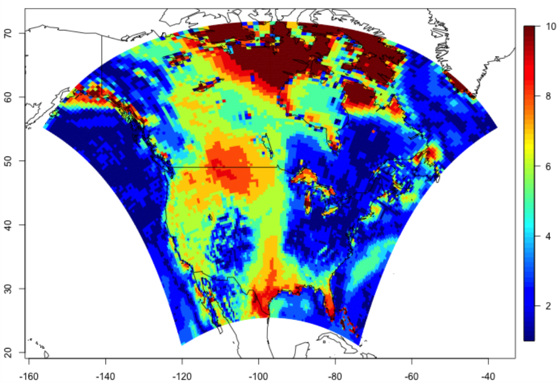
GCM



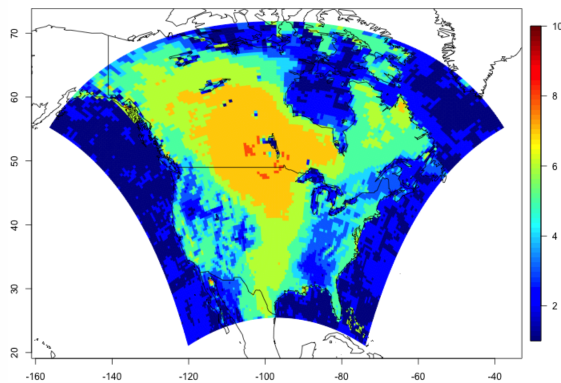
Scenario



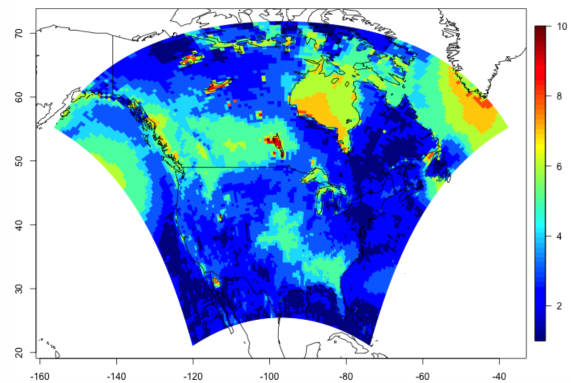
RCM$\star$GCM



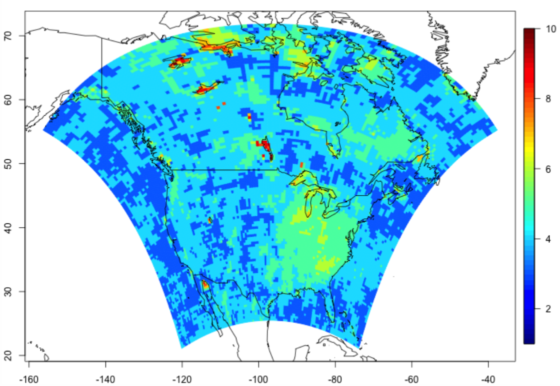
RCM$\star$Scenario



GCM$\star$Scenario

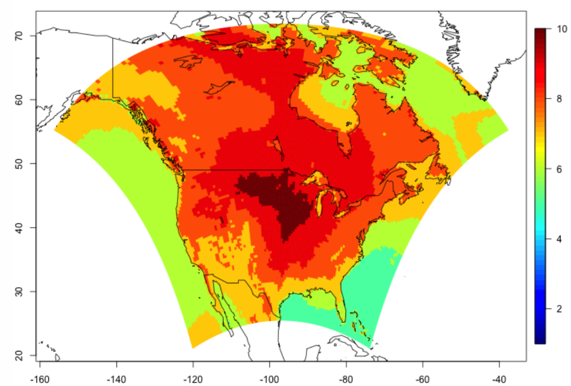


RCM$\star$GCM$\star$Scenario

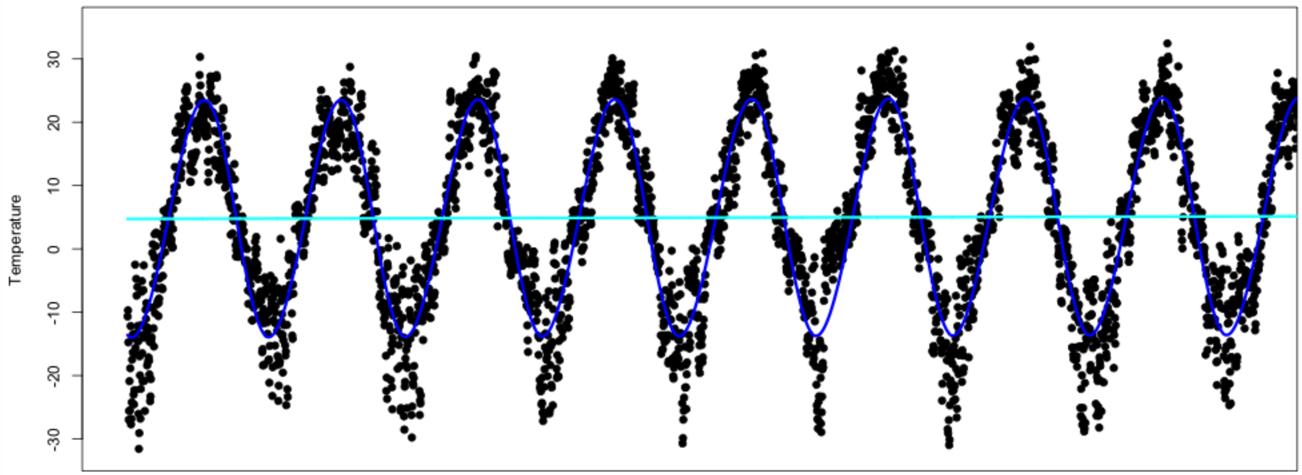


Finite-sample variances via a Bayesian hierarchical spatial model

Residual



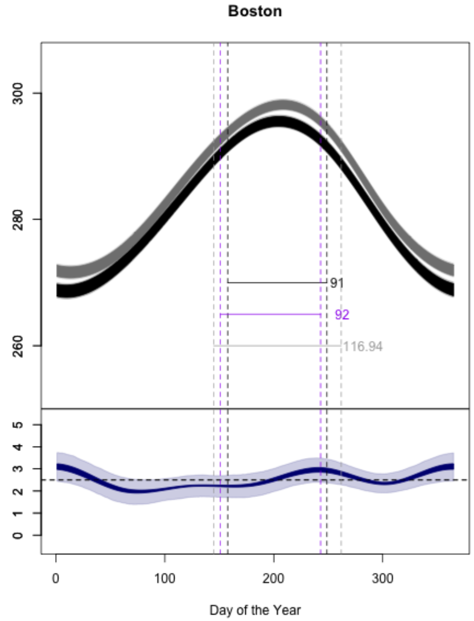
Annual Temperature Profiles

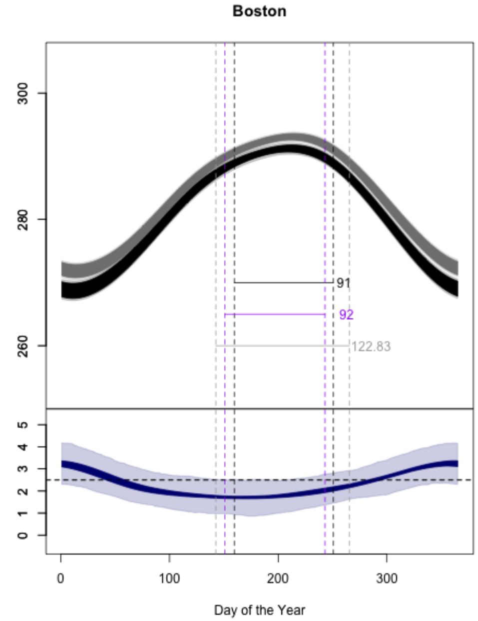


For each grid box, fit a periodic spline w/temporal trend to both current and future runs for a particular model. Allow coefficients to "borrow strength" through a multivariate spatial model on the coefficients of the spline.

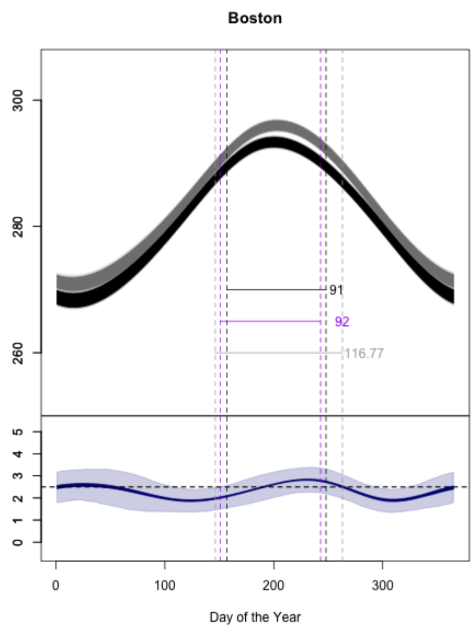
Tammy Greasby (NCAR)

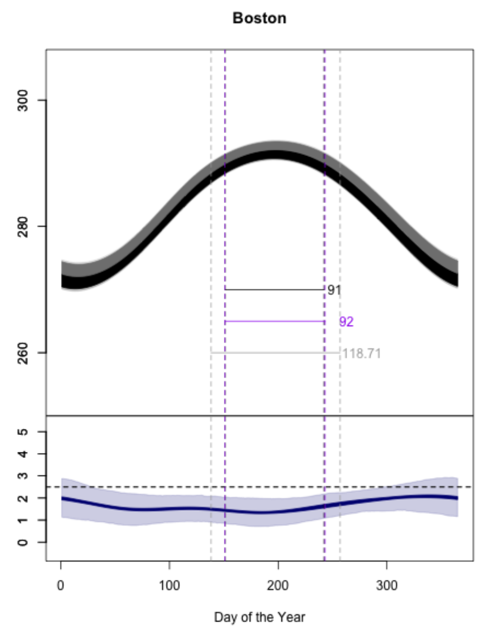
CCSM



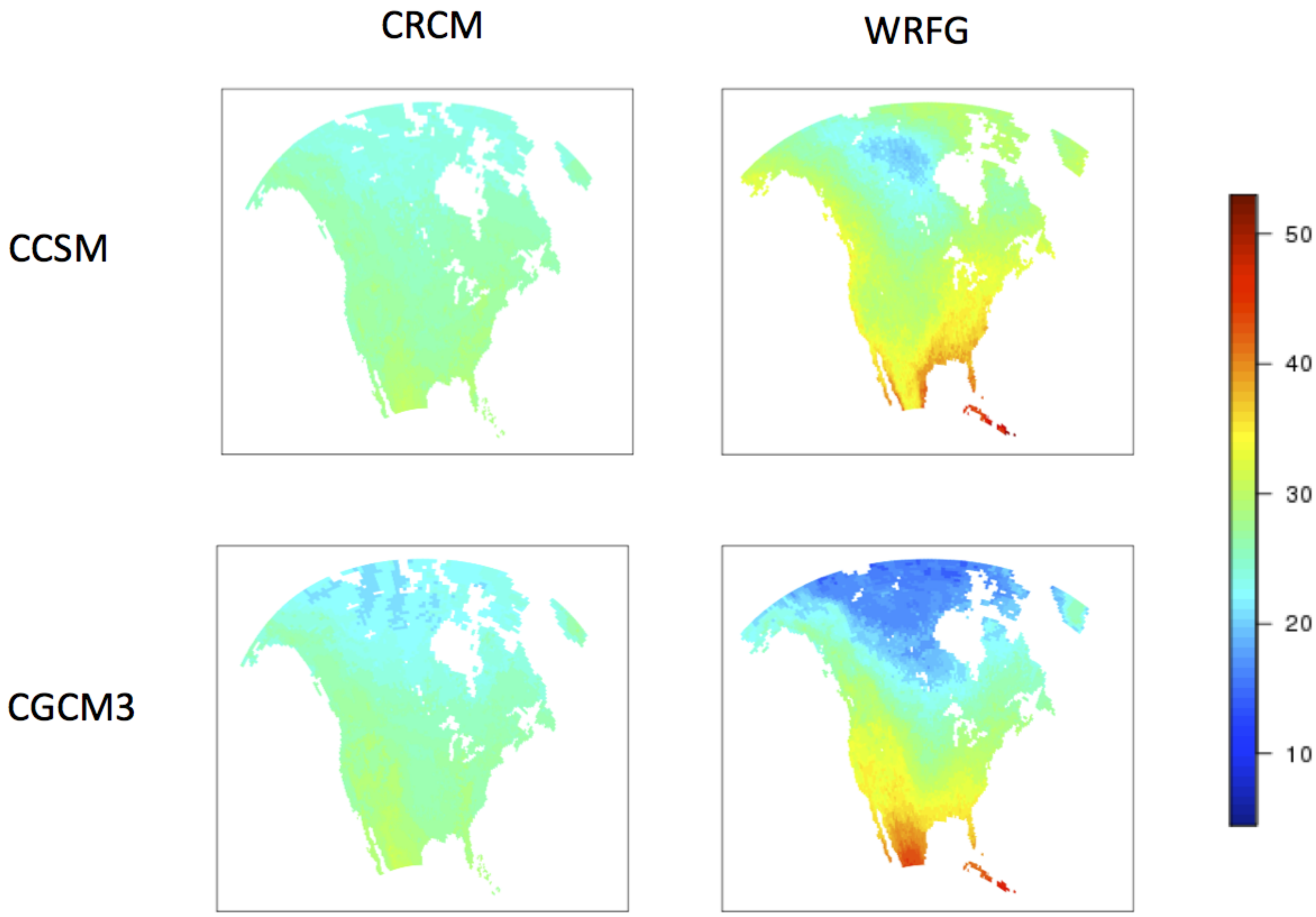


CGCM3

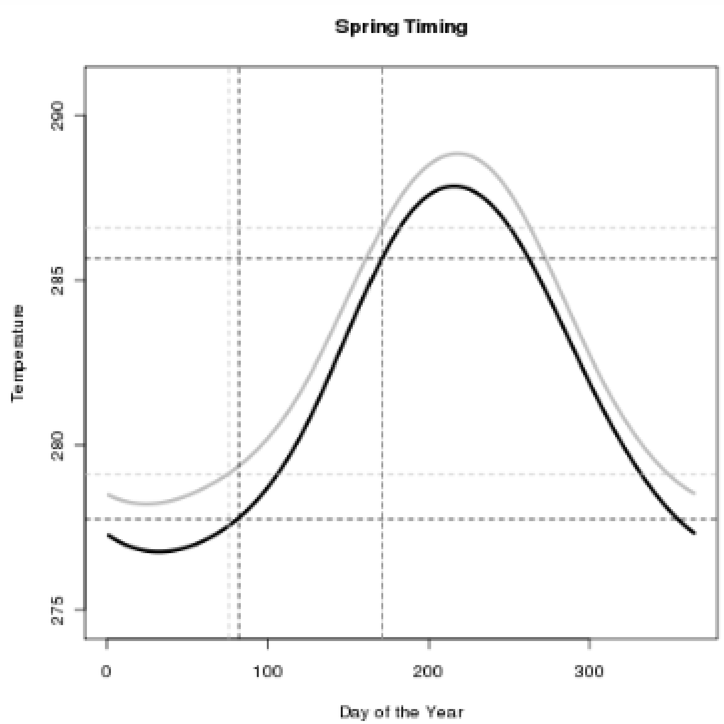


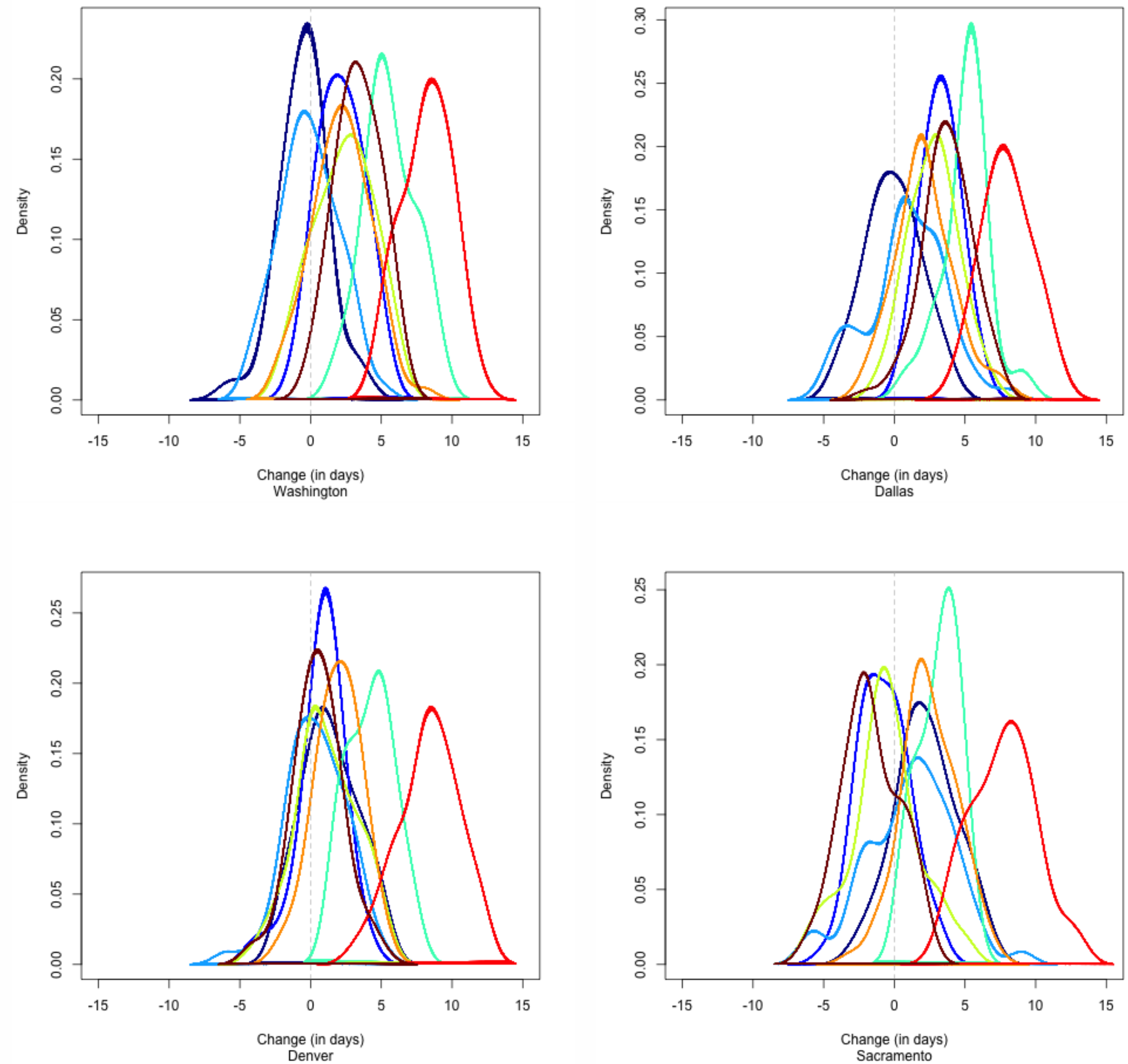


CRCM WRFG



Posterior mean length of increase (days) in summer-like temperatures. Other measures (e.g., growing-degree days, heating/cooling-degree days, etc.) can also be examined.





Onset of spring-like temperatures.

The Bayesian Paradigm

Postulate a model (pdf) for data that depends on some parameters:

$$Y_1, \dots, Y_n \sim \pi(Y_1, \dots, Y_n|\theta).$$

$\Rightarrow$ This forms the *likelihood*.

Postulate a model (pdf) for the parameters:

$$\theta \sim \pi(\theta)$$

$\Rightarrow$ This forms the *prior*.

Inference follows by examining of the posterior distribution:

$$\begin{aligned} \pi(\theta|Y_1, \dots, Y_n) &\propto \pi(Y_1, \dots, Y_n|\theta)\pi(\theta) \\ \text{posterior} &\propto \text{likelihood} \times \text{prior} \end{aligned}$$

$\Rightarrow$ From *Bayes' Theorem*.

A Simple Model

$$Y_1, \ldots, Y_n \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

$$\begin{aligned} \mu|\sigma^2 &\sim \mathcal{N}\left(\mu_0, \sigma^2/\kappa_0\right) \\ \sigma^2 &\sim \text{Inv} - \chi^2\left(\nu_0, \sigma_0^2\right) \end{aligned}$$

Posterior distribution for μ:

$$p\left(\mu|Y_1, \ldots, Y_n\right) = t_{\nu_n}\left(\mu_n, \sigma_n^2, \kappa_n\right)$$

$$\begin{aligned} \mu_n &= \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{Y} \\ \kappa_n &= \kappa_0 + n \\ \nu_n &= \nu_0 + n \\ \nu_n\sigma_n^2 &= \nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}\left(\bar{Y} - \mu_0\right)^2 \end{aligned}$$

A Simple Model

$$
\begin{aligned}
Y_1, \ldots, Y_n &\sim \mathcal{N}\left(\mu, \sigma^2\right) \\
\\
\mu | \sigma^2 &\sim \mathcal{N}\left(\mu_0, \sigma^2/\kappa_0\right) \\
\sigma^2 &\sim \text{Inv} - \chi^2\left(\nu_0, \sigma_0^2\right)
\end{aligned}
$$

Posterior predictive distribution:

1. Sample $\sigma^2 | \{Y_i\}$ from $\text{Inv} - \chi^2\left(\nu_n, \sigma_n^2\right)$.

2. Sample $\mu | \sigma^2, \{Y_i\}$ from $\mathcal{N}\left(\mu_n, \sigma^2/\kappa_n\right)$.

3. Sample Y^* from $\mathcal{N}\left(\mu, \sigma^2\right)$.

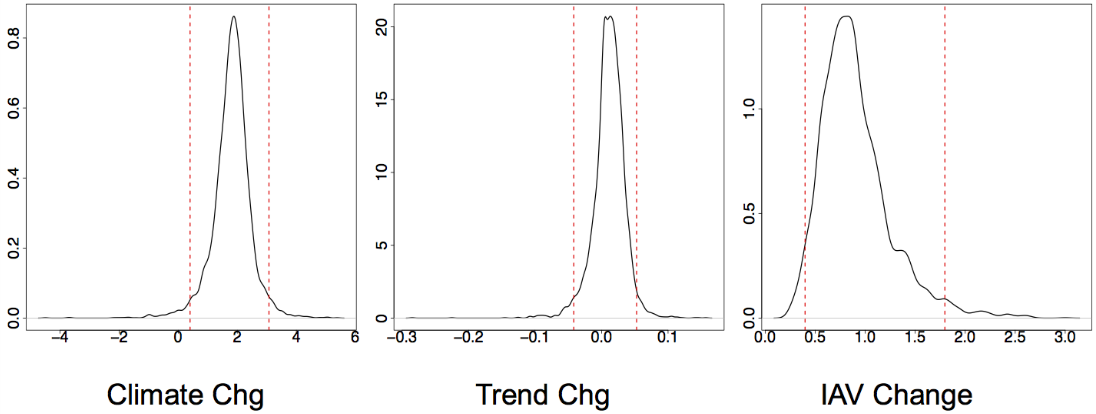
Model weighting

The Tebaldi model:

$$\begin{aligned} X_0 &\sim \mathcal{N}\left(\mu, \lambda_0^{-1}\right) \\ X_j &\sim \mathcal{N}\left(\mu, \lambda_j^{-1}\right) \\ Y_j &\sim \mathcal{N}\left(\nu + \beta\left(X_j - \mu\right), \left(\theta\lambda_j\right)^{-1}\right) \end{aligned}$$

- X_0 indicates an observed climate
- X_j indicates model output for the current time period $j = 1, \ldots, 13$.
- Y_j indicates model output for the future time period.
- μ is current mean temperature, ν is future temperature
- θ allows the climate model variance to change between time periods.
- β accounts for correlation between the current and future climate models. $\beta = 1$ is equivalent to modeling climate change directly.

- Modify Tebaldi model to:
  - Incorporate multiple observational datasets.
  - Model precisions (λs), which control the influence of a particular model on estimates of μ, ν and $\nu - \mu$, as a function of NARCCAP design (GCM, RCM, scenario).
  - *Tammy Greasby (NCAR)*
- Start from scratch (the kitchen sink model):
  - Incorporate multiple observational datasets.
  - Incorporate GCMs, NCEP-driven RCMs, GCM-driven RCMs.
  - Incorporate "familial" relationship between GCMs and RCMs.
  - Include model-to-model correlations and "bias" terms.
  - *Matt Heaton (NCAR)*



Pr(Fut>Cur) = 0.98　　Pr(Fut>Cur) = 0.76　　Pr(Fut>Cur) = 0.31

The Kitchen Sink Model:
Pacific Southwest winter temperatures.

Data	Current Weight	Future Weight
UDEL	0.245	NA
CRU	0.224	NA
CRCM	0.006	NA
ECP2	0.020	NA
HRM3	0.045	NA
MM5I	0.021	NA
RCM3	-0.000	NA
WRFG	0.010	NA
CCSM	0.081	0.190
CGCM3	0.083	0.186
GFDL	0.066	0.127
HADCM3	0.084	0.204
CRCM-ccsm	0.013	0.066
CRCM-cgcm3	0.012	0.034
ECP2-gfdl	0.017	0.048
HRM3-hadcm3	0.025	0.052
MM5I-ccsm	0.013	0.021
RCM3-cgcm3	0.011	0.021
RCM3-gfdl	0.006	0.005
WFRG-ccsm	0.006	0.005
WFRG-cgcm3	0.014	0.050

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

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.

Furrer, R., Geinitz, S., and Sain, S.R. (2011), Assessing variance components of general circulation model output fields, Environmetrics, to appear.

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

Greasby, T.A. and Sain, S.R. (2011), Multivariate spatial analysis of climate change projections, JABES, 16, 571- 585, doi: 10.1007/s13253-011-0072-8

Weller, G.B., Cooley, D.S., and Sain, S.R. (2012), An investigation of the pineapple express phenomenon via bivariate extreme value theory, Environmetrics, to appear.