Impact of resolution in dynamic downscaling experiments

Stephan R. Sain

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

Doug Nychka (NCAR), Will Kleiber (CU), Ray Arritt (ISU),
Brian Reich, Ryan Parker (NCState)

Workshop on Stochastic Weather Generators; Avignon, France; 9/18/2014

CISL Computational & Information Systems Laboratory

NCAR

Supported by NSF (AGS, DMS), others.

Impact of resolution in dynamic downscaling experiments

Stephan R. Sain
Climatology Team
The Climate Corporation
San Francisco, CA / Boulder, CO

Doug Nychka (NCAR), Will Kleiber (CU), Ray Arritt (ISU),
Brian Reich, Ryan Parker (NCState)

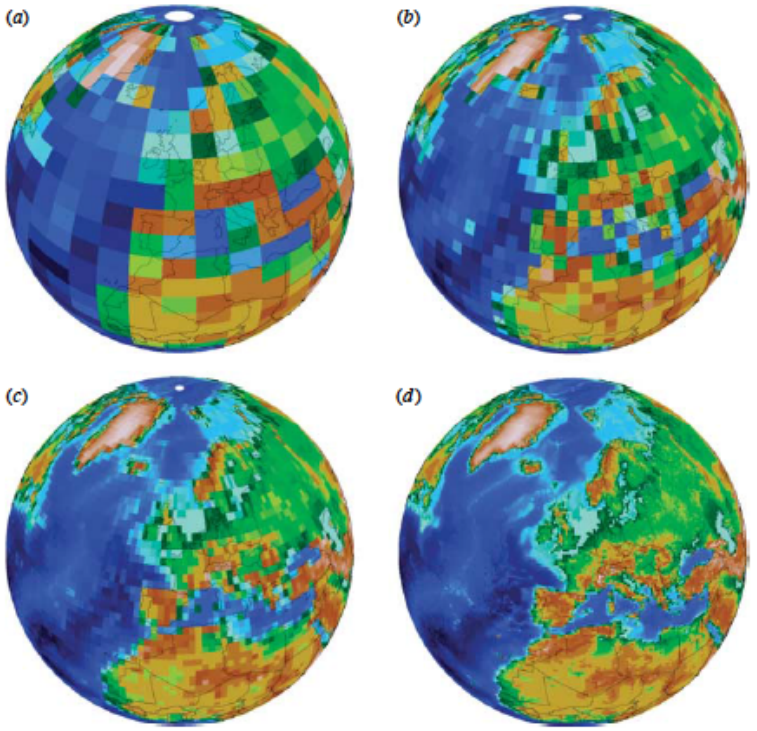
Workshop on Stochastic Weather Generators; Avignon, France; 9/18/2014

THE CLIMATE CORPORATION
www.climate.com

Uncertainty in projections of climate change

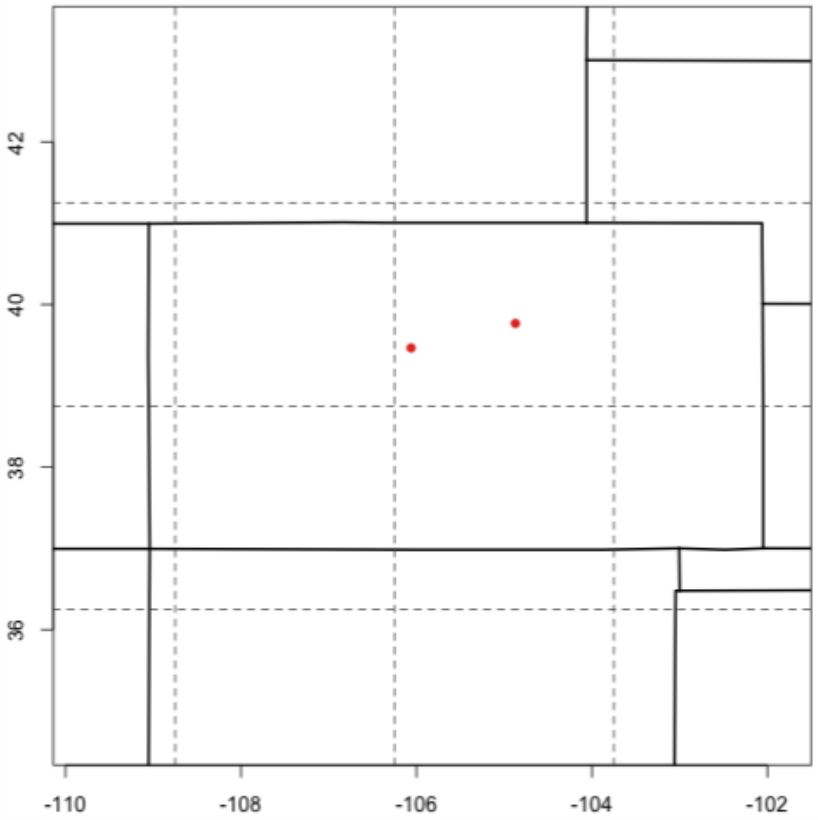
- Natural variability of climate system.
- Lack of knowledge about future emissions and aerosols as well as how the climate system will respond to these future forcings.

- Climate models:
  - Parametric uncertainty (e.g., sub-grid-scale approximations)
  - Structural uncertainty (e.g., unknown, not implemented, or poorly implemented physical processes)
  - ⋆ Spatial resolution???



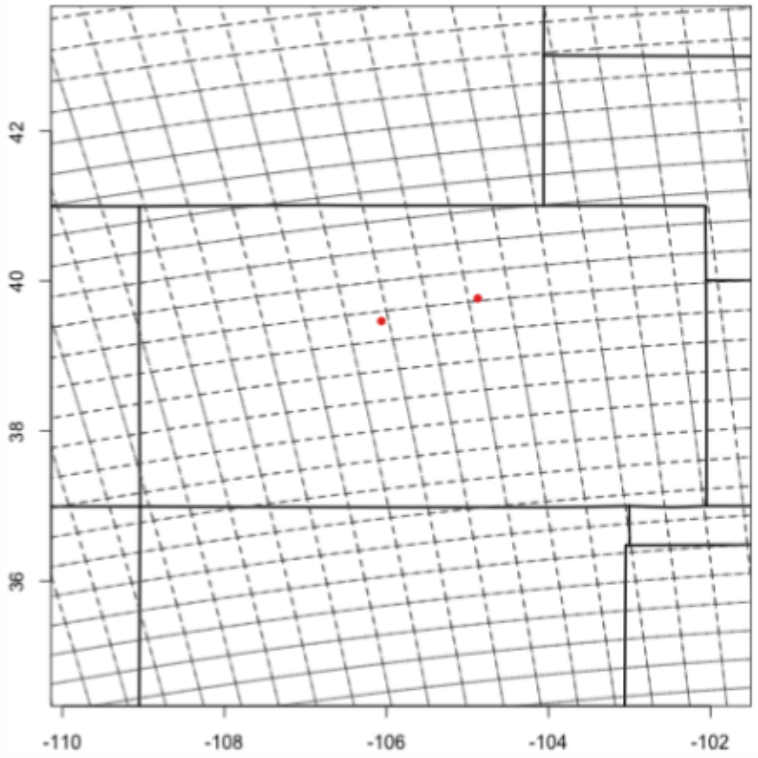
Downscaling

- GCMs are useful for investigating the large-scale circulation and forcings that affect the Earths global climate **but**, . . .
- There are limitations to their use for regional and local projections that might be of interest to the impacts community:
  - Heat waves, drought, floods, snowpack, etc.
  - Skiing, fire, water, crops, public health, power, etc.
- Generating regional climate information on the basis of GCMs is referred to as **downscaling**.



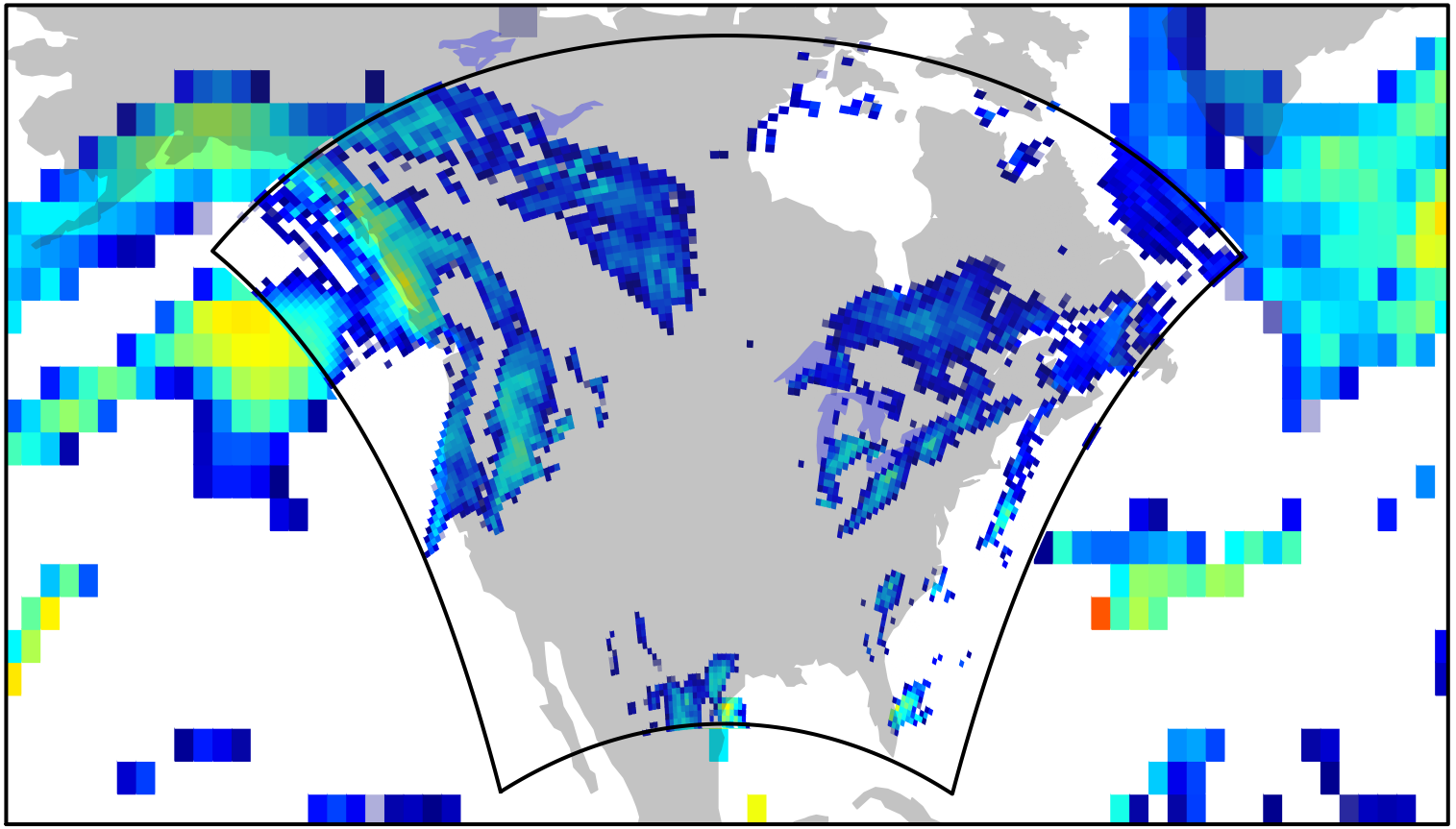
Downscaling

- Statistical downscaling is a computational efficient approach that uses empirical relationships to connect the coarse-resolution GCM output to regional and local variables.



- Dynamic downscaling uses high-resolution climate models, incorporating physical principles; shown to reproduce a broad range of climates around the world.

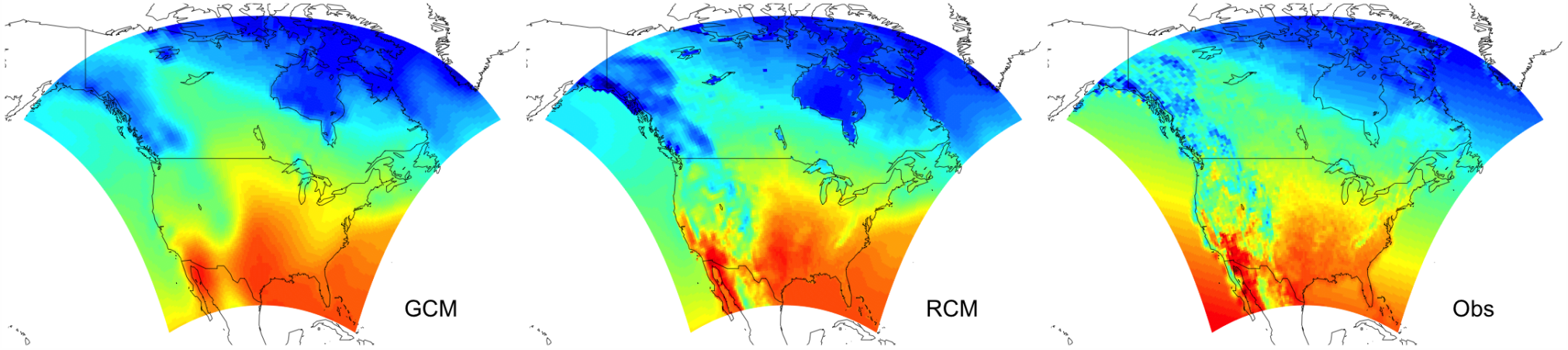
Regional climate models



Two questions

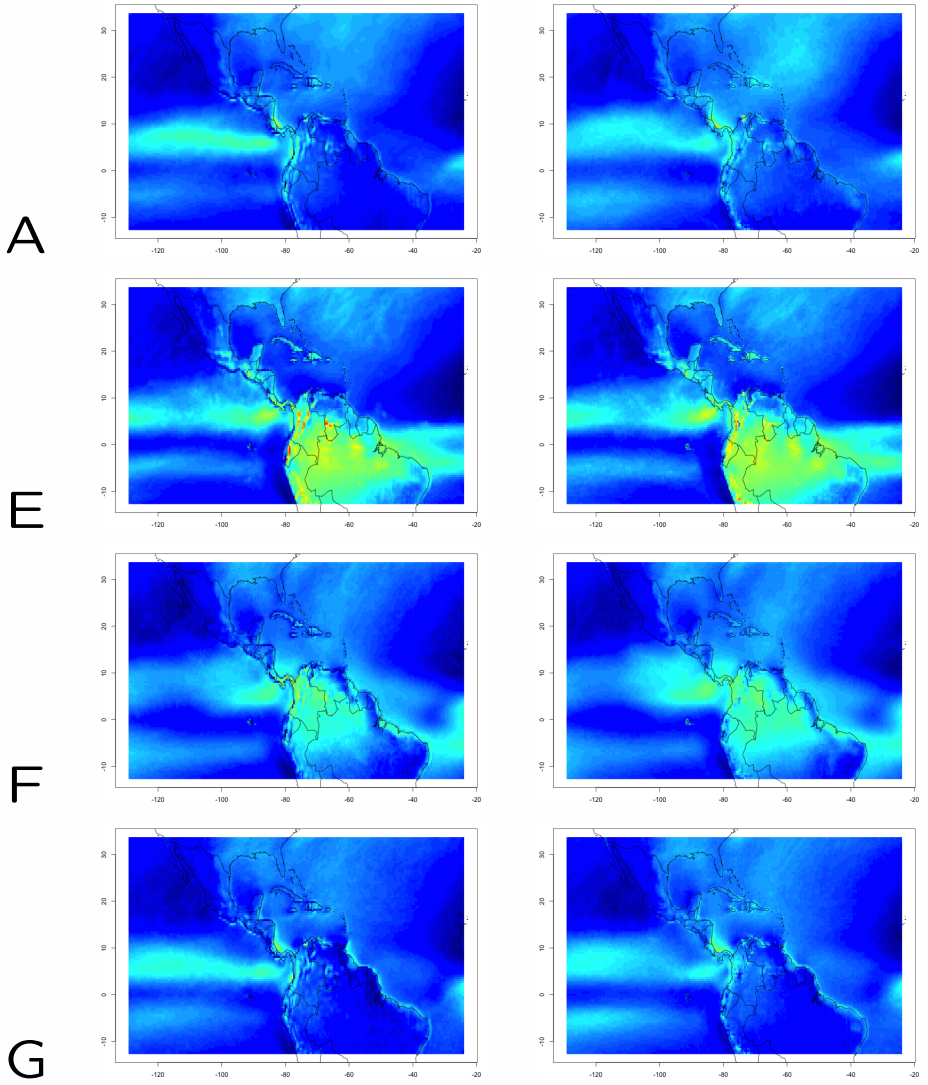
1. Moving to higher resolution requires certain tradeoffs and considerations. One question that arises is the impact of resolution on choice of physical parameterization.

2. Higher resolution model experiments (and climate model experiments in general) are very expensive. A second question that arises in how to assess the added value from downscaling methods like RCMs?



Resolution vs parameterization

- Part of the COordinated Regional climate Downscaling EXperiment (CORDEX).
- ICTP RegCM V4 w/boundary conditions providing by ERA-Interim reanalysis.
- 20 years (1989-2008) of monthly total precipitation.
- Two model resolutions (25km and 50km).
- Four convective parameterization schemes (A, E, F, G).



transformed monthly precip

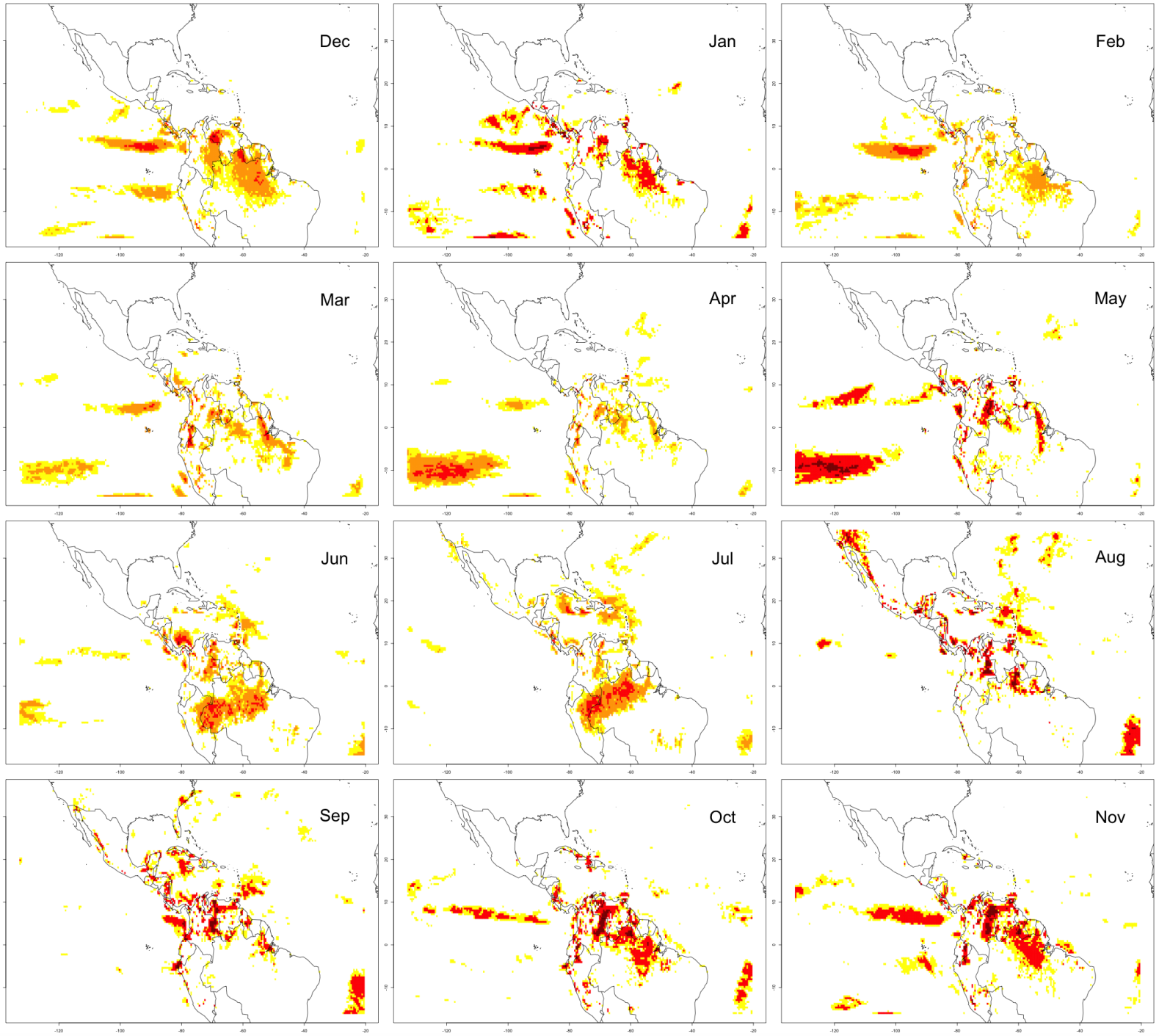
Local ANOVA

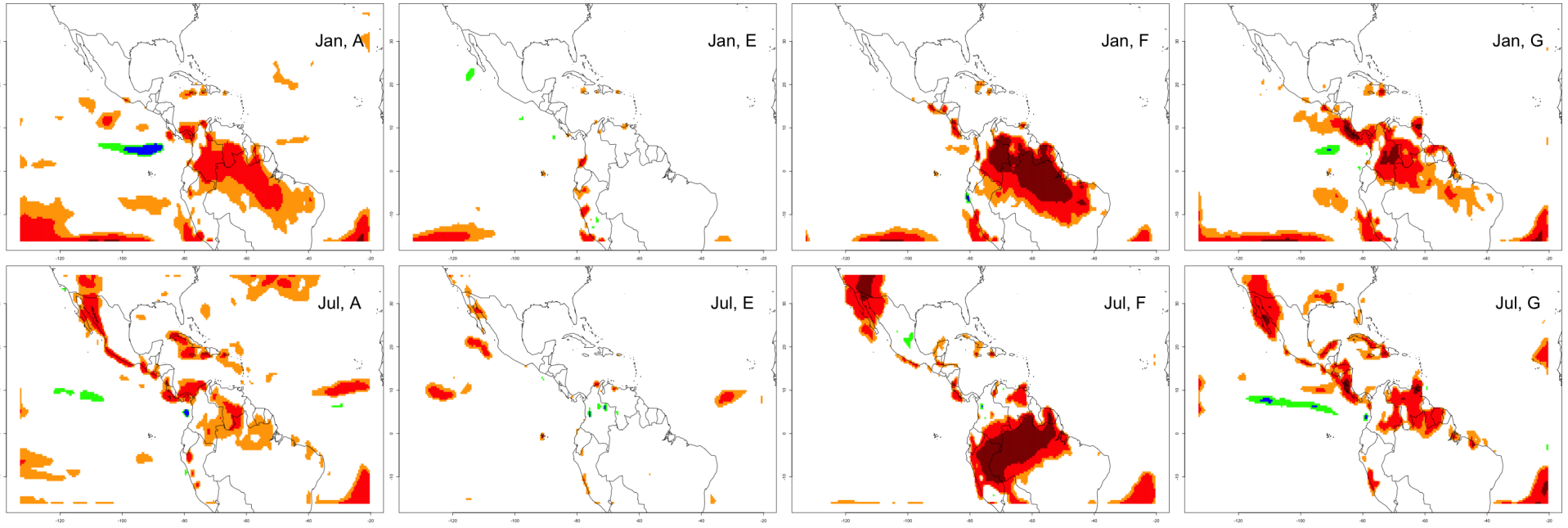
Fit a local analysis of variance model (ANOVA), i.e. find coefficients $\boldsymbol{\beta}$ that maximize

$$\sum_{i=1}^{n} K_h(\mathbf{s} - \mathbf{s}_i)\ell(y_i; \boldsymbol{\beta})$$

where $\ell(y_i; \boldsymbol{\beta})$ is based on the Gaussian likelihood $y_i \sim \mathcal{N}\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\right)$ with $\mathbf{X}$ containing linear terms for location, main effects of physics and resolution, and interactions between physics and resolution.

- A finite support kernel (biweight) is used with h chosen to include a 7x7 neighborhood.
- Each month considered separately.
- Impact of the interaction terms by measuring the "extra sums of squares" between a full and reduced model.


Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov



Initial results suggest that there are significant resolution interactions, but they are smaller than monthly differences, differences due to convective parameterizations, and year-to-year variability.

Connections to spatial statistics

Most spatial estimators are local smoothers:

$$\widehat{Y}_0 = \mathbf{w}'\mathbf{Y} = \sum_{i=1}^{n} w_i Y_i$$

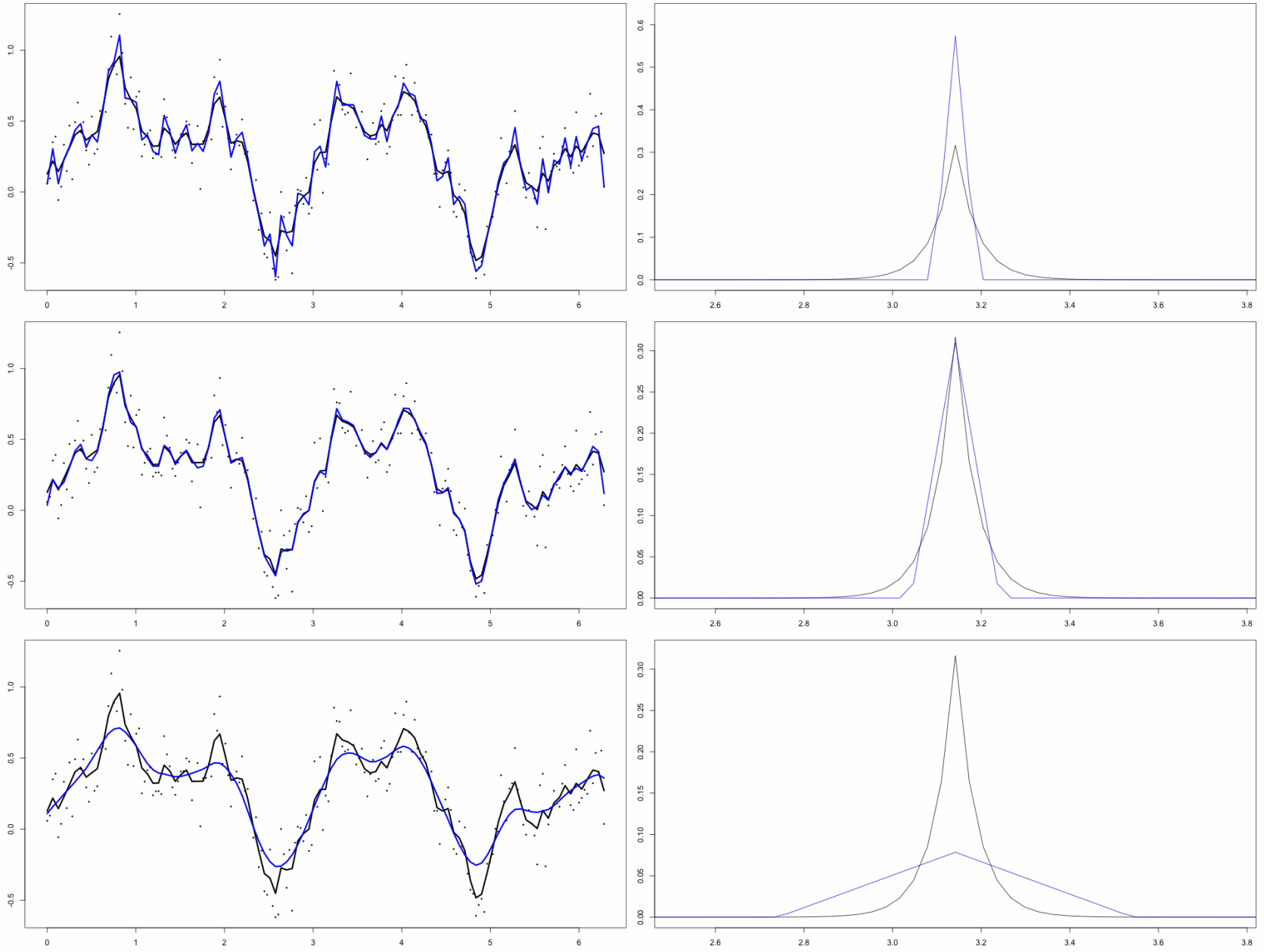
For example, Kriging yields $\mathbf{w} = \mathbf{c}_0'\boldsymbol{\Sigma}^{-1}\mathbf{Y}$.

Consider the local linear kernel estimator that is found by maximizing

$$-\frac{1}{2}\sum_{i=1}^{n} K_h(x - x_i)\left\{y_i - \alpha_0 - \alpha_1(x - x_i)\right\}^2$$

and yielding $\widehat{Y}(x) = \widehat{\alpha}_0$.

The local linear (as well as the more general local polynomial) estimators can also be written as a linear smoother.



Why do we care?

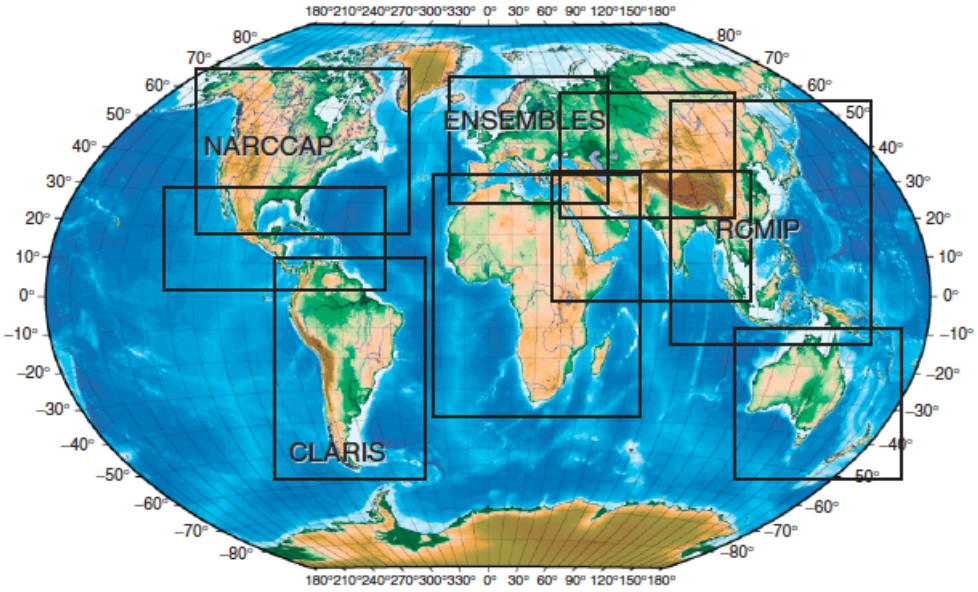
- The local linear approach is easy to implement on a cluster with a finite-support kernel.
- Properly calibrated, it is (approximately) equivalent to Kriging.
- As for calibration (i.e., picking the bandwidth/window) – still an open question, but we have some initial analytical results and some ideas based on sampling.
- Extensions – the local polynomial objective function is equivalent to a local likelihood, i.e. find the maximum of

$$\sum_{i=1}^{n} K_h(x - x_i)\ell(y_i; \theta),$$

which opens the doors for non-Gaussian data, multivariate (functional ANOVA?) and space-time data, etc.

NARCCAP – Added value

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



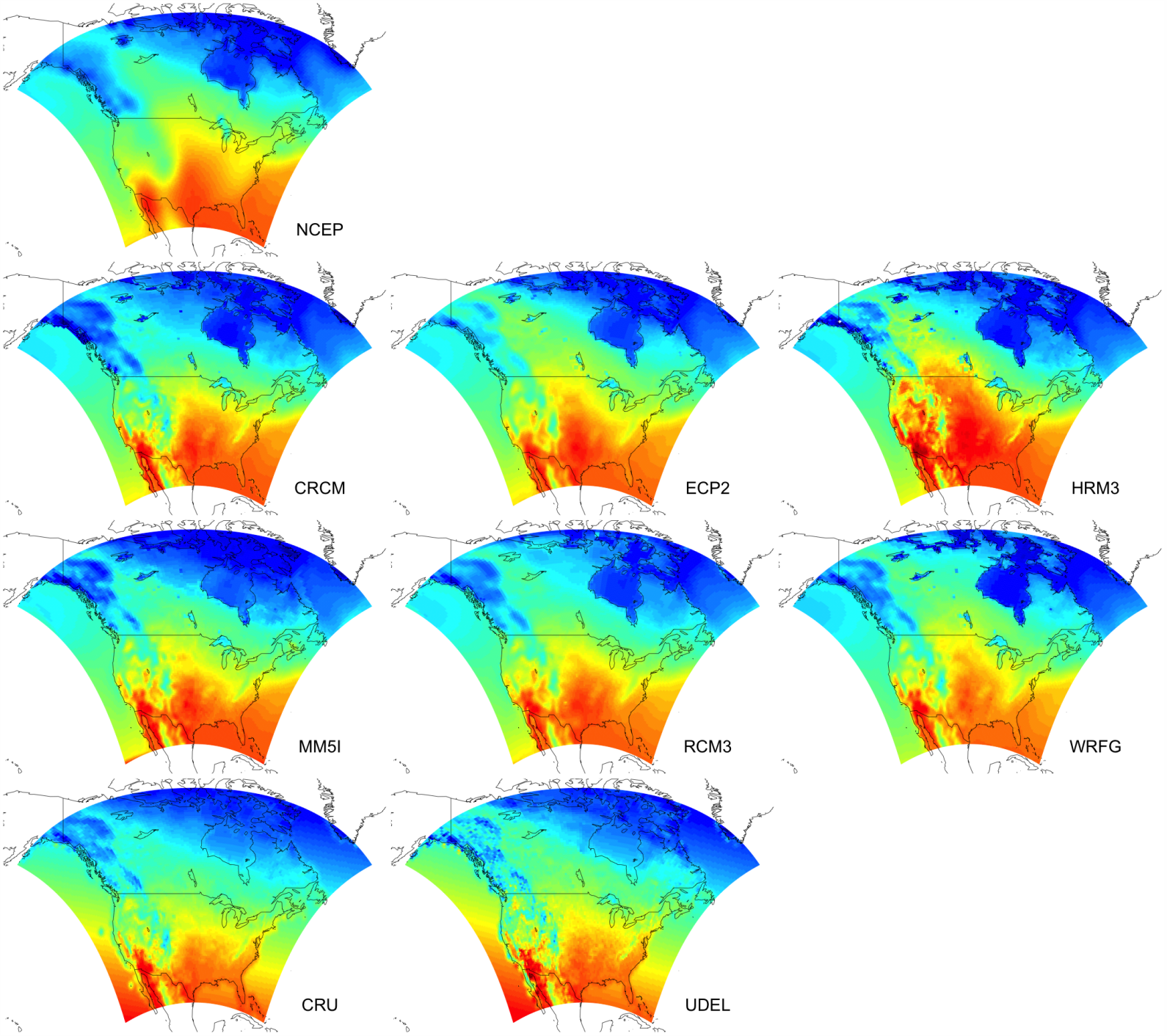
- *Systematically investigate the uncertainties in regional scale projections of future climate* and produce high resolution climate change projections using multiple RCM and multiple GCM simulations.

- 4 GCMs provide boundary conditions for 6 RCMs – balanced half-fraction

The NARCCAP design

	Phase I	Phase II			
	NCEP	GFDL	CGCM3	HADCM3	CCSM
CRCM	•		•		•
ECP2	•	•		??	
HRM3	•	•		•	
MM5I	•			•	•
RCM3	•	•	•		
WRFG	•		•		•

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


NCEP
CRCM
ECP2
HRM3
MM5I
RCM3
WRFG
CRU
UDEL

The DCT

Let $\mathbf{Y}_k$ denote an $n_1 \times n_2$ matrix of seasonal temp/precip for each of K data sources. Transform the $\mathbf{Y}_k$ to $\mathbf{Z}_k$ via the two-dimensional discrete cosine transformation (DCT), i.e.

$$Z(t_1, t_2) = a_{t_1} b_{t_2} \sum_{s_1=0}^{n_1-1} \sum_{s_2=0}^{n_2-1} \cos\left[\pi \frac{t_1}{n_1}\left(s_1 + \frac{1}{2}\right)\right] \cos\left[\pi \frac{t_2}{n_2}\left(s_2 + \frac{1}{2}\right)\right] Y_k(s_1, s_2).$$

- $Z_k(0,0)$ is the overall mean of $\mathbf{Y}_k$.
- $Z_k(t_1, t_2)$ corresponds to the feature with frequency $\boldsymbol{\omega} = (\omega_1, \omega_2)$ and $\omega_1 = t_1/(2n_1)$ the frequency of the rows and $\omega_2 = t_2/(2n_2)$ the frequency in the columns.

The DCT

Group the Z_k by resolution, i.e. let $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iK})'$ defined by (t_{i1}, t_{i2}) and $\boldsymbol{\omega}_i = (\omega_{i1}, \omega_{i2})$. Assume

$$\mathbf{Z}_i \sim \mathcal{N}\left(\mathbf{0}, d_i^{-1}\boldsymbol{\Sigma}(f_i)\right),$$

where d_i determines the relative variance of Z_i and $f_i = ||\boldsymbol{\omega}_i||$.

- Let

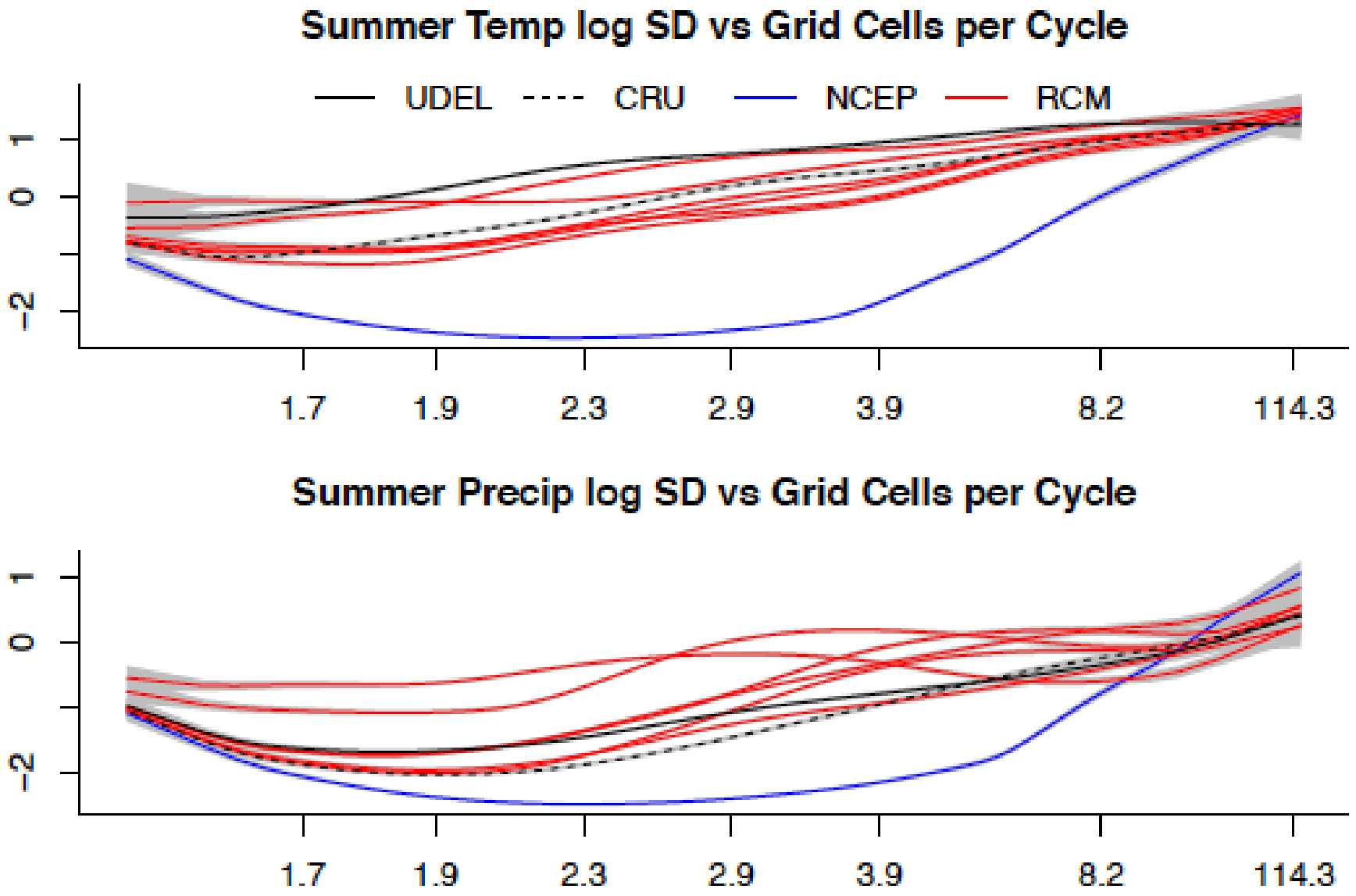
$$d_i = 4 - 2\left[\cos\left(t_{i1}\frac{\pi}{n_1}\right) + \cos\left(t_{i2}\frac{\pi}{n_2}\right)\right]$$

- Let

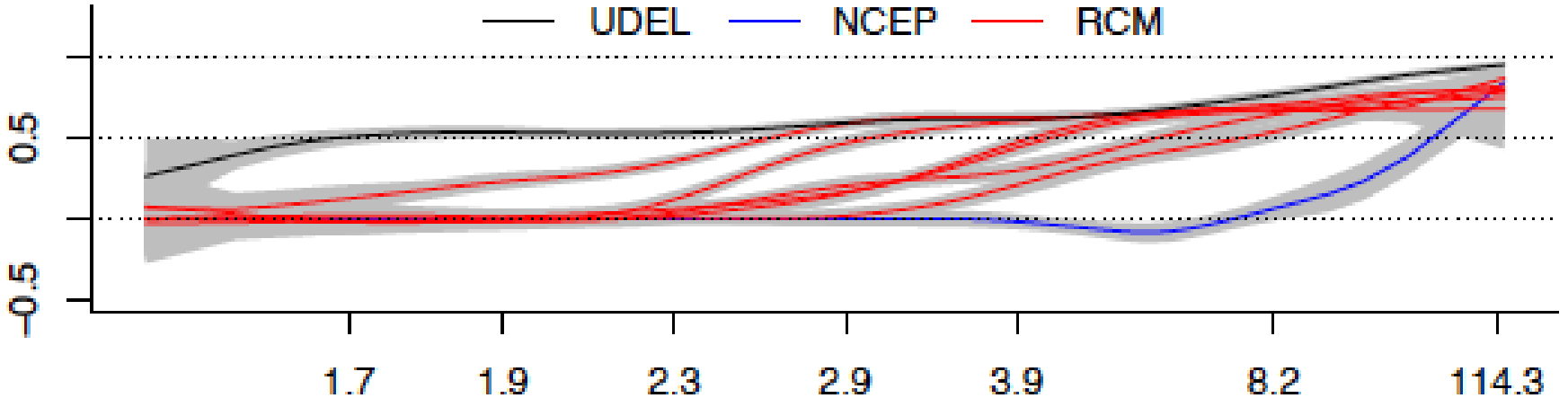
$$\boldsymbol{\Sigma}(f) = \sum_{\ell=1}^{L} w_\ell(f)\boldsymbol{\Omega}_\ell.$$

Data sources

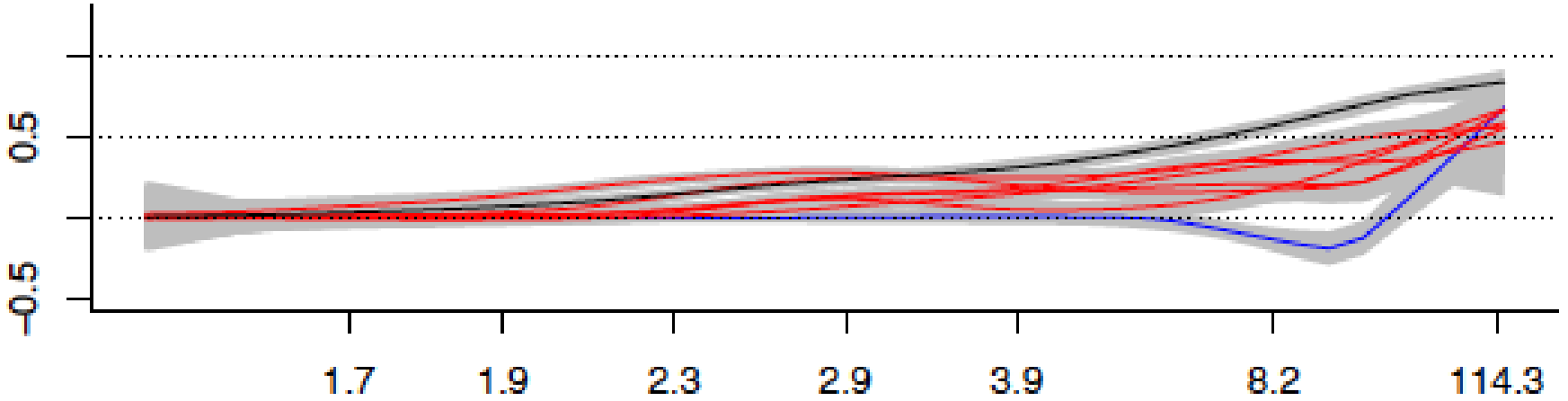
- NARCCAP Phase I model output
  - CRCM, ECP2, HRM3, MM5I, RCM3, WRFG.
- NCEP
- Observational datasets (CRU, UDEL).
- 20-year (1981-2000) averages of seasonal (JJA, DJF) temperature and precipitation.
- All data sources interpolated to common 50km grid.
- Measures of added value obtained by examining $\Sigma(f)$ at different frequencies.
  - Marginal variances, marginal correlations, conditional correlations...

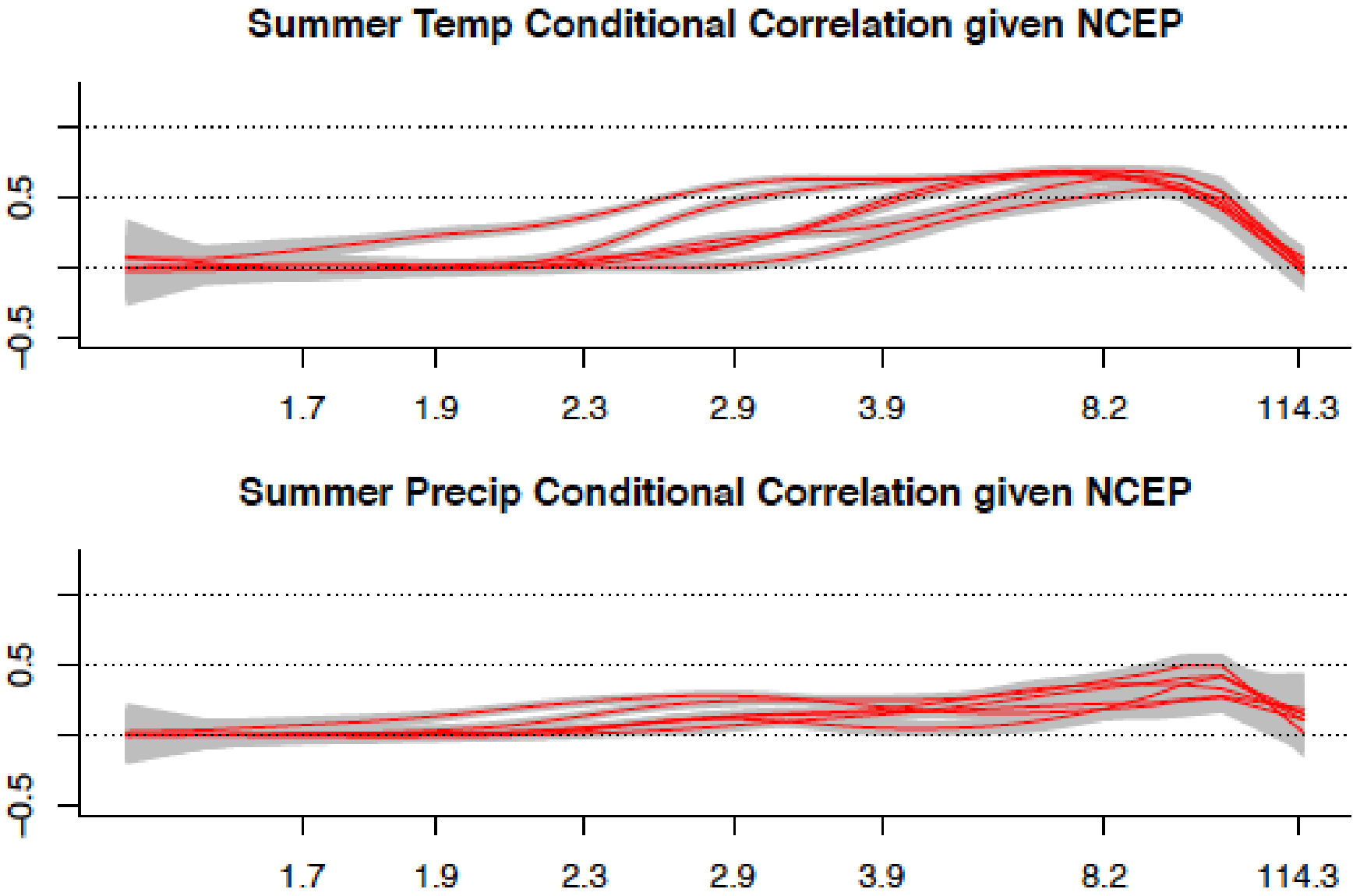

Summer Temp log SD vs Grid Cells per Cycle
UDEL
CRU
NCEP
RCM
1
0
−2
1.7
1.9
2.3
2.9
3.9
8.2
114.3
Summer Precip log SD vs Grid Cells per Cycle
1
0
−2
1.7
1.9
2.3
2.9
3.9
8.2
114.3

Summer Temp Correlation vs Grid Cells per Cycle



Summer Precip Correlation vs Grid Cells per Cycle




Summer Temp Conditional Correlation given NCEP
0.5
−0.5
1.7
1.9
2.3
2.9
3.9
8.2
114.3
Summer Precip Conditional Correlation given NCEP
0.5
−0.5
1.7
1.9
2.3
2.9
3.9
8.2
114.3

Connections to spatial statistics

With the prescribed d_i and by letting $\mathbf{\Sigma}(f)$ be constant for all f, then this model is equivalent to a multivariate conditional autoregressive (MCAR) model, i.e.

$$\mathbf{Y} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{Q}^{-1}\right)$$

where $\mathbf{Q}$ has a prescribed form based on a simple neighborhood structure (grid boxes that sharing edges).

The Climate Corporation aims to help farmers around the world protect and improve their farming operations with uniquely powerful software and insurance products.

Combine hyper-local weather monitoring, agronomic data modeling, and high-resolution weather simulations to deliver solutions that helps farmers make better informed operating and financing decisions.

To achieve these ambitious goals we use a wealth of data (gauge and other instruments, gridded data products, remote sensing, numerical model output, etc.) coupled with powerful applied statistical and machine learning tools.

THE CLIMATE CORPORATION
www.climate.com

Questions?

Thank You!

steve.sain@climate.com