NCAR

Distilling Regional Climate Model Data from NARCCAP for Use in Impacts Analysis

Seth McGinnis

IMAGe – CISL – NCAR

mcginnis@ucar.edu

2013-09-09

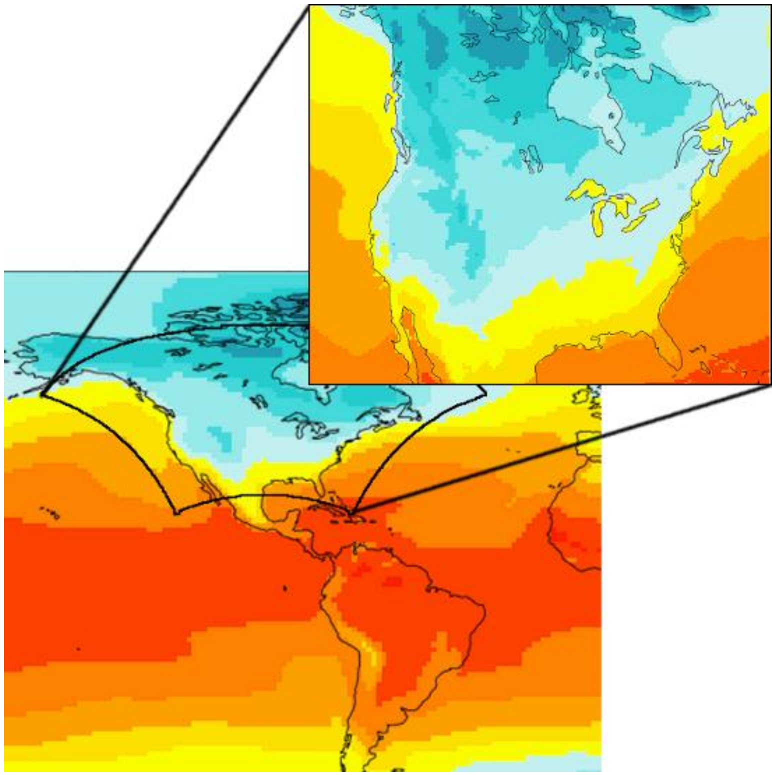
National Center for Atmospheric Research

v.5

Outline

- Introduction
- Overview of NARCCAP
- Supporting impacts users
  - Aggregation
  - Interpolation
  - Bias correction
- Looking forward

NARCCAP: North American Regional Climate Change Assessment Program



Nest high-resolution regional climate models (RCMs) inside coarser global models (GCMs) over North America

NARCCAP Collaborators

NCAR – Linda Mearns, Seth McGinnis, Melissa Bukovsky, Larry McDaniel, Doug Nychka, Steve Sain, Josh Thompson

GFDL – Isaac Held, Bruce Wyman

Hadley Centre – Richard Jones, Simon Tucker, Erasmo Buonomo, Wilfran Moufouma-Okia

Iowa State University – Bill Gutowski, Ray Arritt, Dave Flory, Daryl Herzmann, Gene Takle

LLNL – Phil Duffy, Dave Bader, Dean Williams

OURANOS – Sebastien Biner, Daniel Caya, Rene Laprise

PNNL – Ruby Leung, James Correia, Yun Qian

Scripps – Ana Nunes (also UFRJ), John Roads (deceased)

UC Santa Cruz – Lisa Sloan, Mark Snyder

Experimental Design

	25 years	Two 30-year runs, current & future			
	NCEP	**GFDL**	**CGCM3**	**HADCM3**	**CCSM**
CRCM	**X**	**--**	**X**	**--**	**X**
ECP2	**X**	**X**	**--**	**X**	**--**
HRM3	**X**	**X**	**--**	**X**	**--**
MM5I	**X**	**--**	**--**	**X**	**X**
RCM3	**X**	**X**	**X**	**--**	**--**
WRFG	**X**	**--**	**X**	**--**	**X**
Timeslices		**X**	**--**	**--**	**X**

6 RCMs x 4 GCMs
+ NCEP and timeslices
= 34 runs total

Simulation Output Archive

- 3-hourly frequency
- 50-km gridcells
- Avg domain size: 139×112 gridpoints
- 2D variables: 35
- 3D variables: 7
- Vertical levels: 28
- NetCDF format

34 runs × 30 years × 365 days × 8 timesteps × 139 X × 112 Y × (35 + 7×28 vars) × 4 bytes =

~40 TB total data volume

NARCCAP Program Goals

- Evaluate model performance and uncertainty
- Support further dynamical downscaling experiments
- Generate high-res climate change scenario data for impacts analysis

Supporting Impacts Users

Real-world example:

 # days w/ T_{max} ≥ 90°, 100° F for Austin, TX?

(i.e., boil it all down to a few spreadsheet cells)

Requires:

- Time aggregation
- Interpolation
- Bias correction

Time Aggregation Is Tricky

Model output is 3-hourly

Users need averages / climatologies

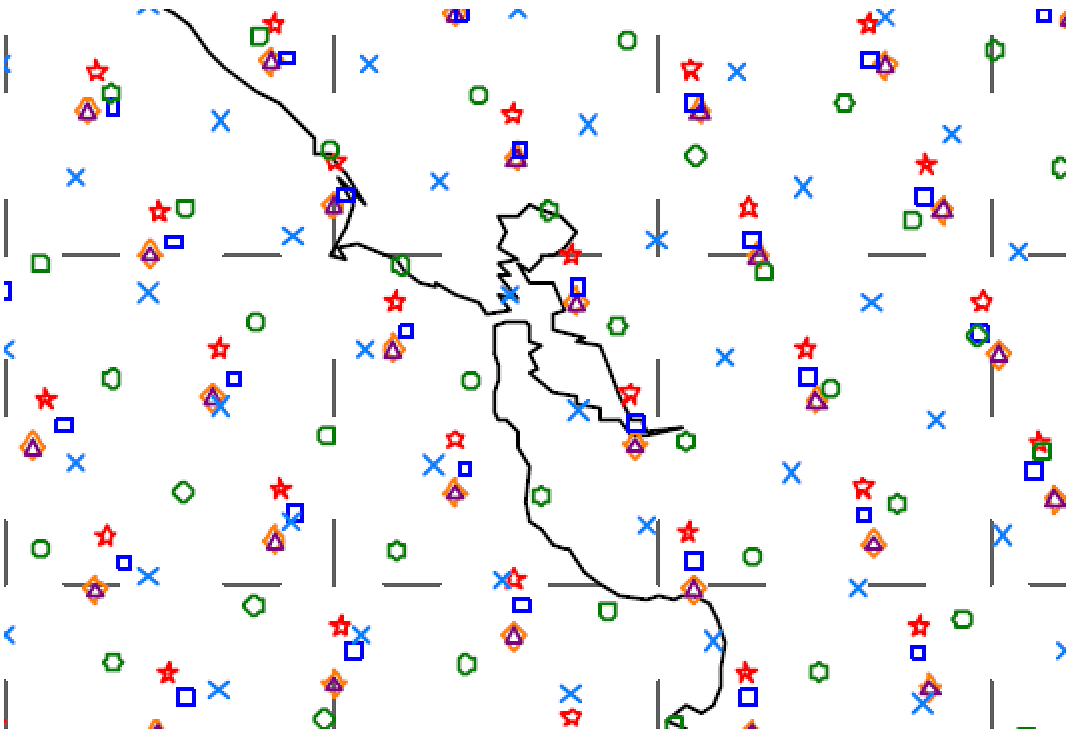
Theoretically straightforward, BUT…

- Different calendars
- Endpoint variations
- Gaps in data

Easy to make small errors with big effect

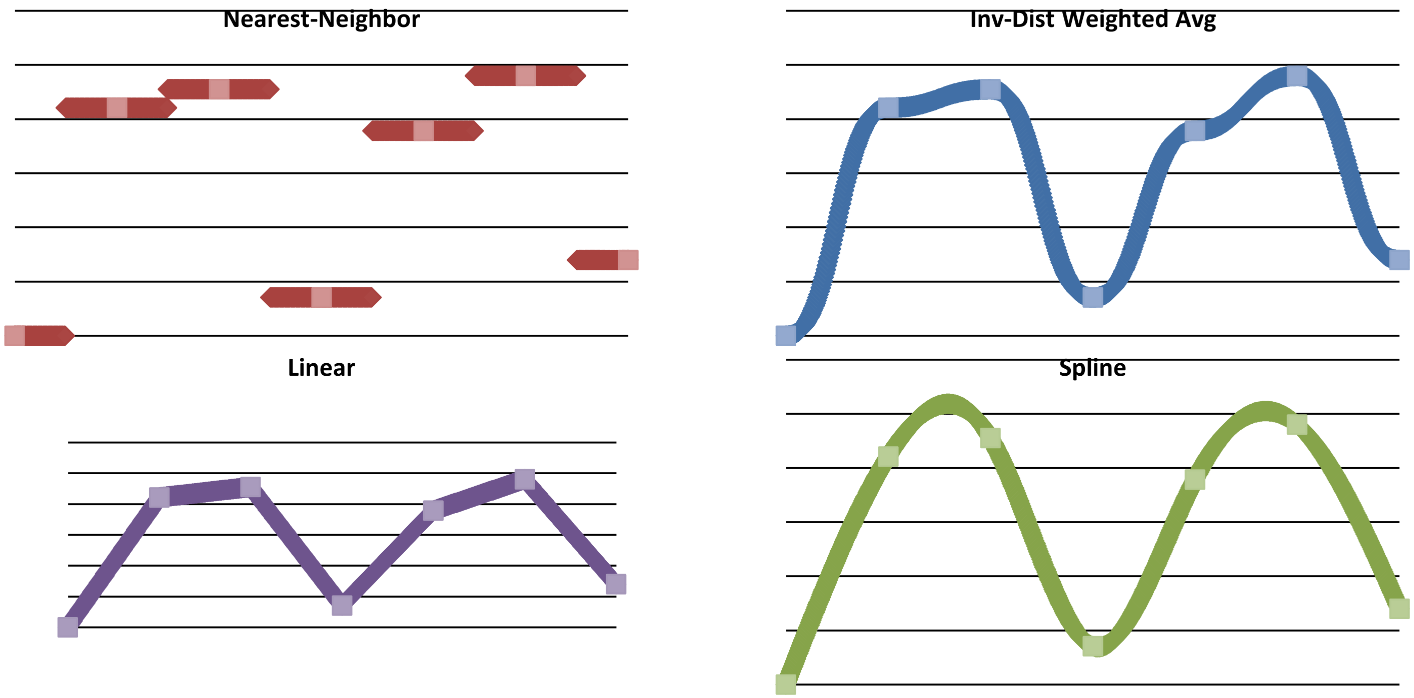
Interpolation

Model gridpoints are seldom conveniently located



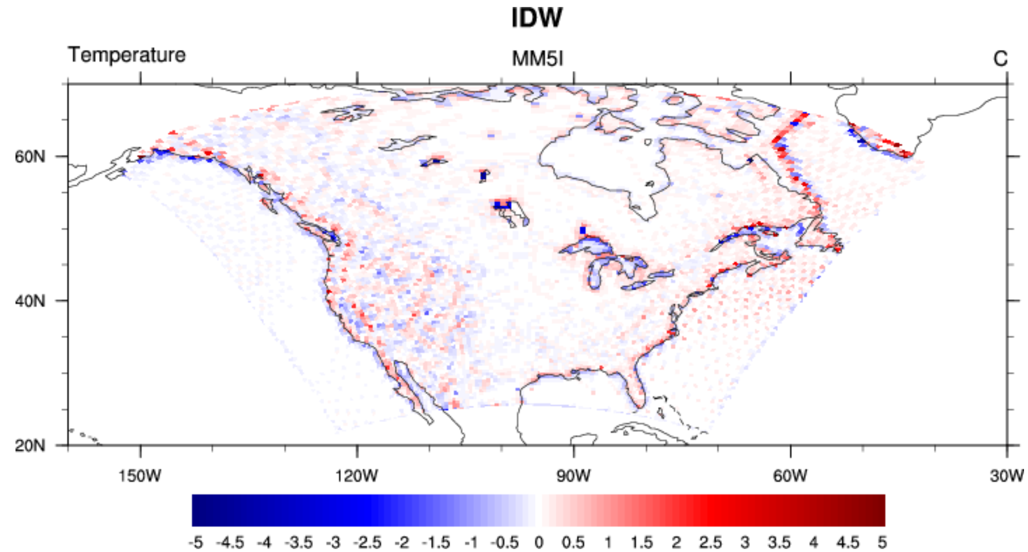
Many Interpolation Methods

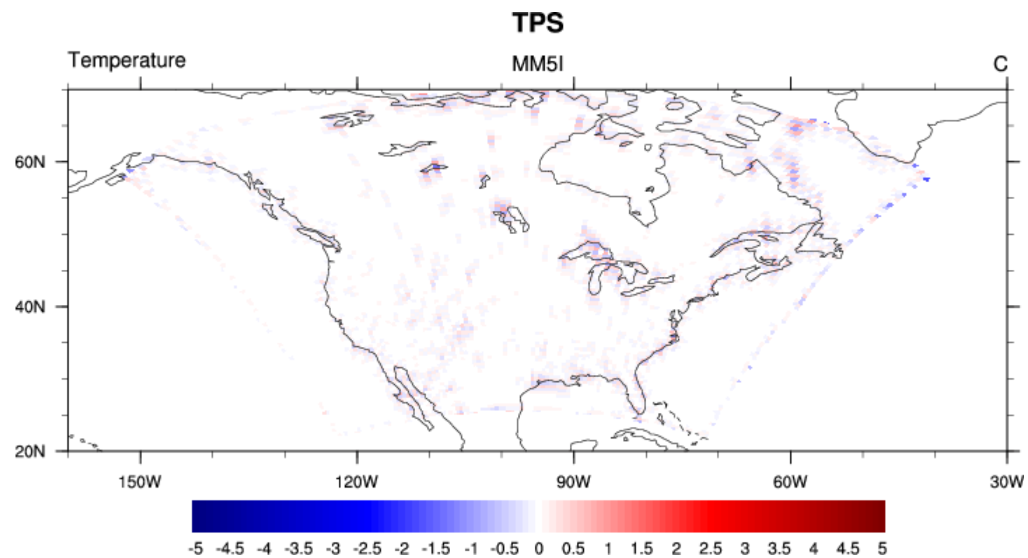
Does it matter which algorithm you use?



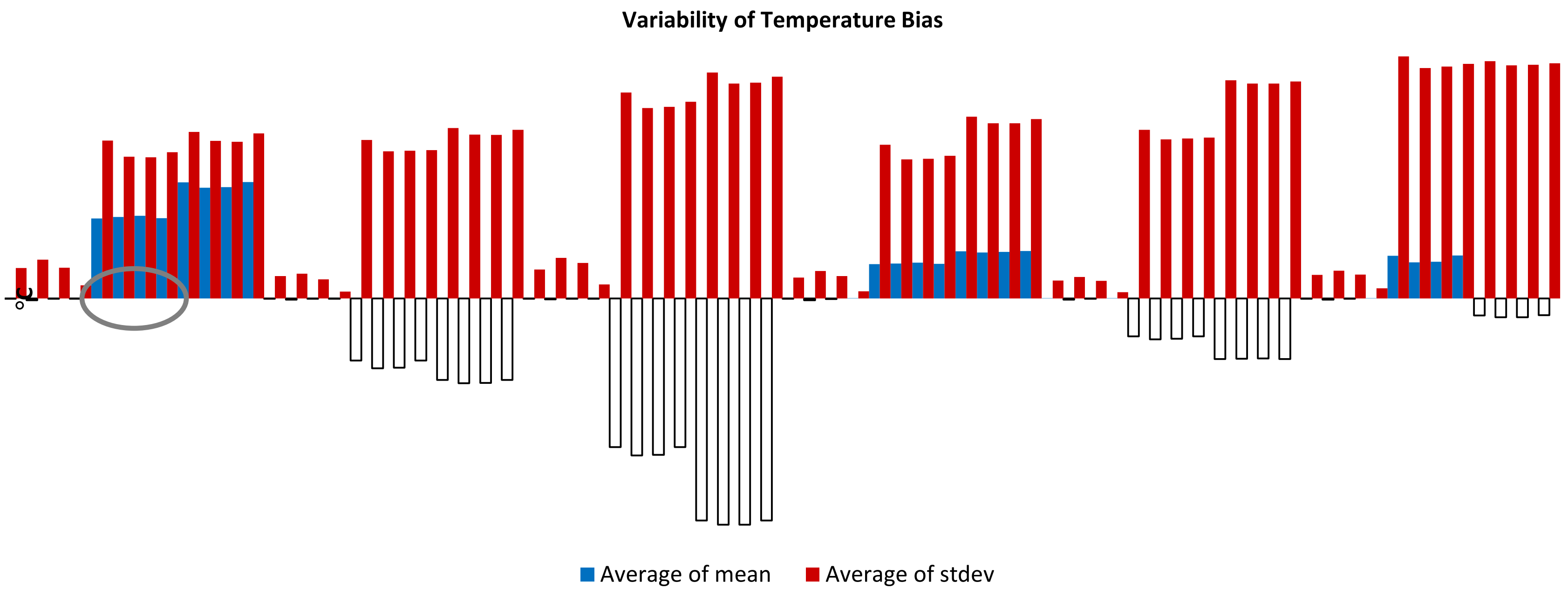
Interpolation Error

Estimate error by interpolating to new grid and back to original





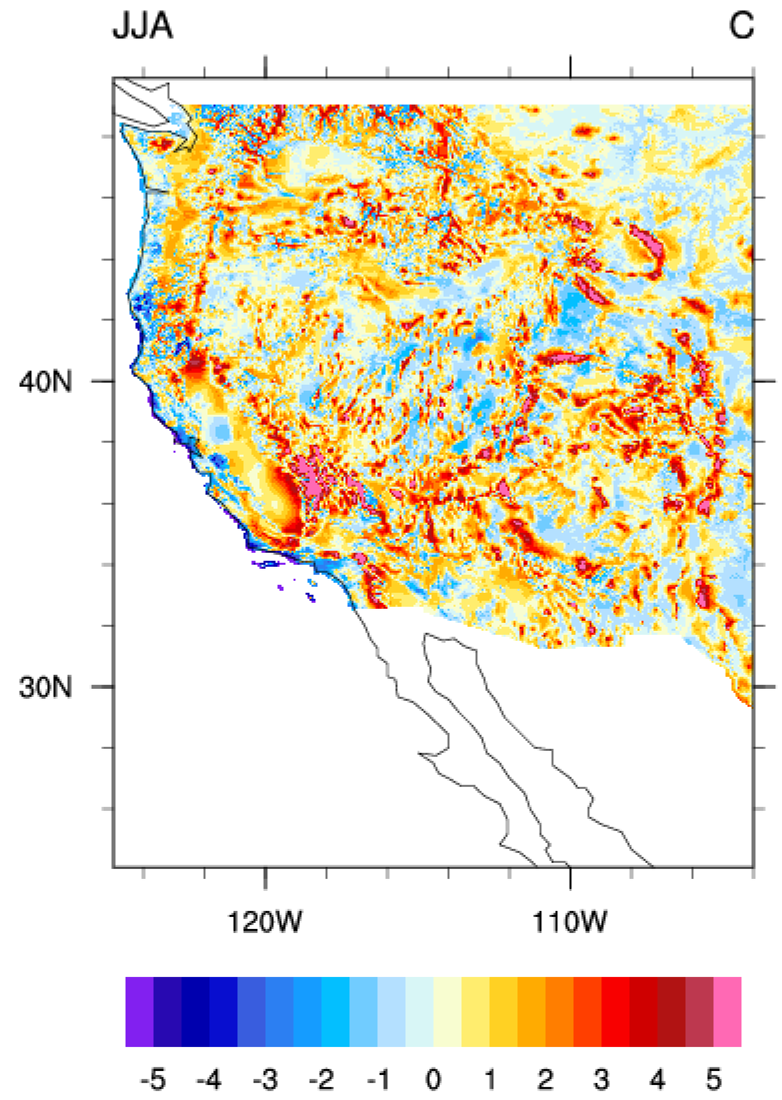
Interpolation Error vs Variability Range of Bias



Interpolation error (short bars) is noticeable on the same scale as temperature bias (long bars)

Reduction in Bias Due to Elevation Correction

- NCEP-driven ensemble compared to PRISM
- Interpolate via kriging w/ elevation covariate
- No significant effect east of Rocky Mtns



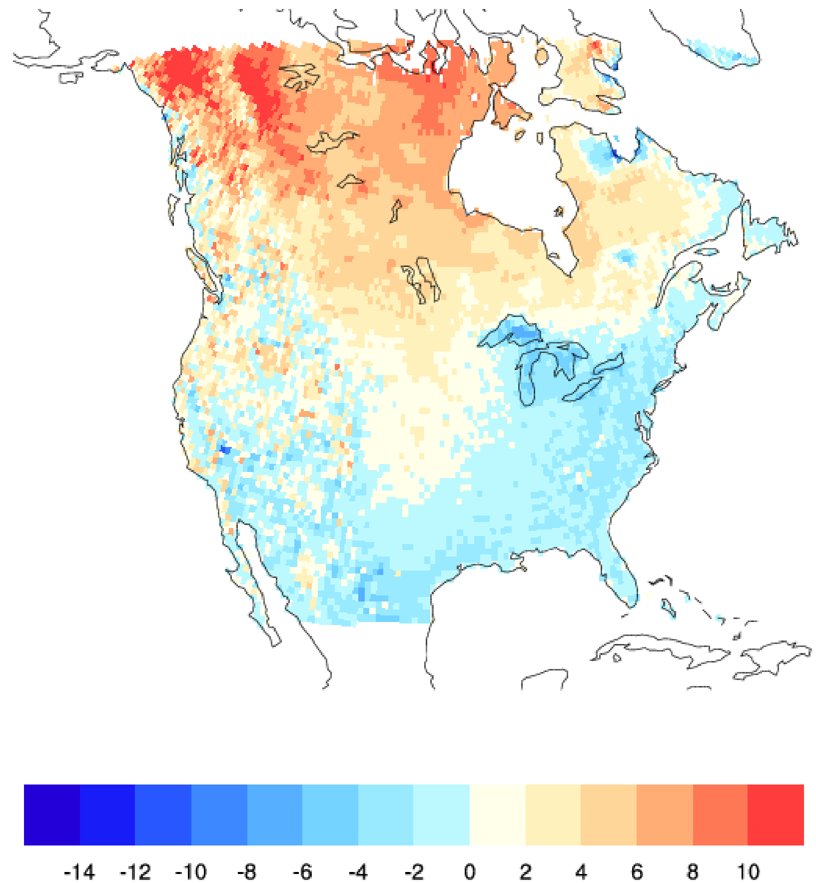
Interpolation Is Difficult

- More sophisticated methods perform better in complex terrain
- Simplistic methods may smooth away features of interest
- Need to provide both interpolation tools and interpolated data

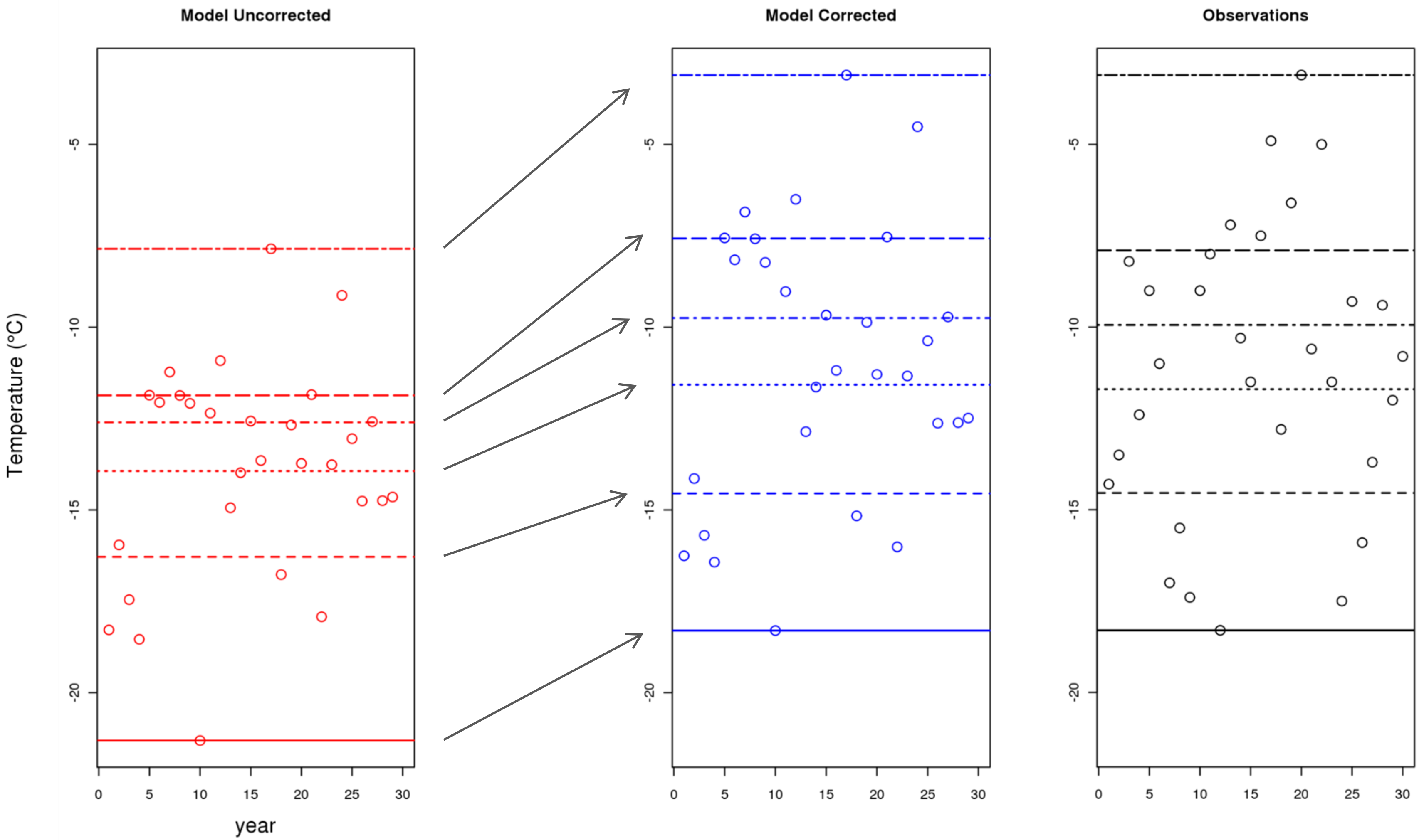
Bias Correction

- Climate models have bias
- Delta method often used to correct mean bias*
  *assuming stationarity
- What about the rest of the distribution?

RCM3-UDEL Winter Temperatures



Quantile Mapping Corrects Entire Distribution

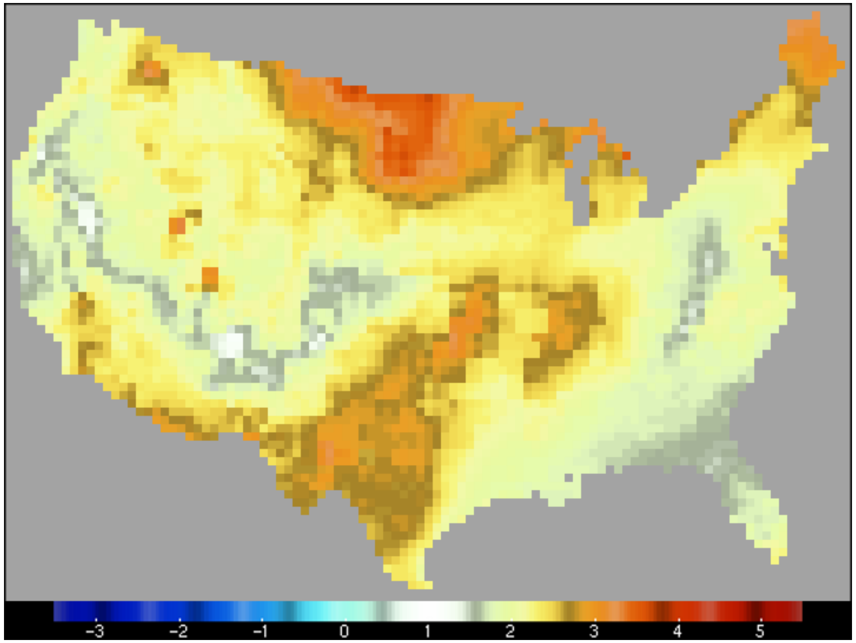


Quantile Mapping Methodology

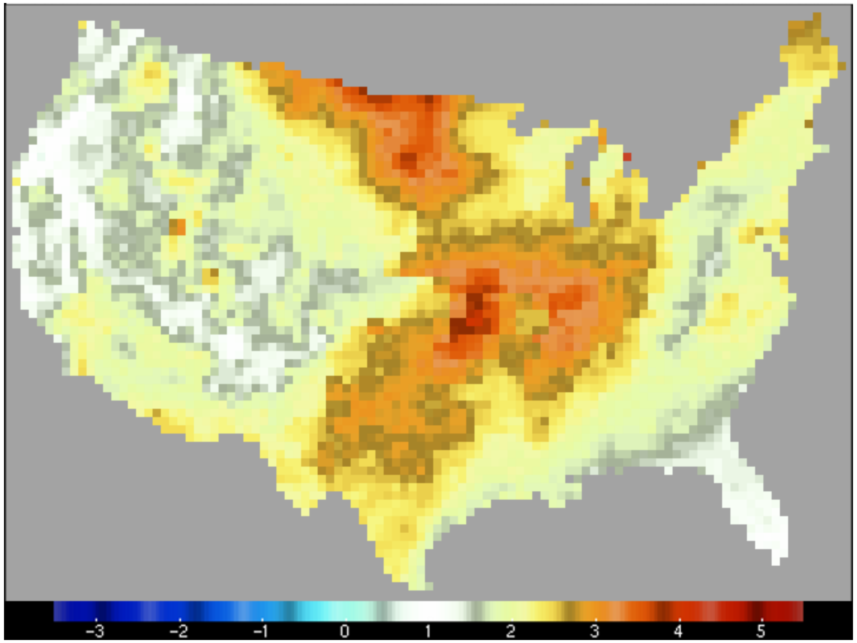
- Operate on daily data using `qmap` library for R
- Use Maurer 1/8° daily gridded data for obs
- 15-day moving window, correct center day
- Correct each grid-cell separately
- Empirical quantiles with linear extrapolation
- # quantiles = # inputs (CDF mapping)
- Assume stationarity to correct future data

Change in Winter T_{max} (°C, CRCM-ccsm)

Uncorrected

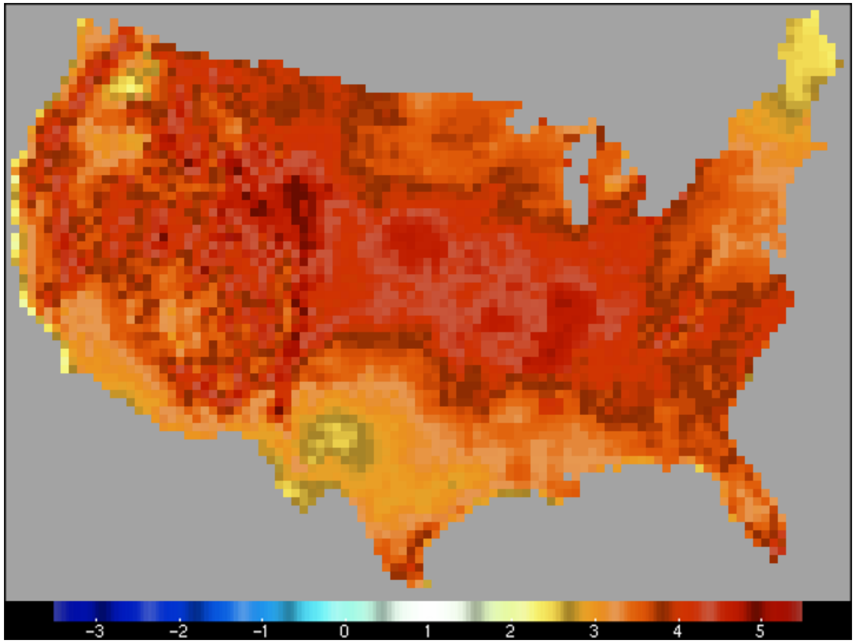


Bias-corrected

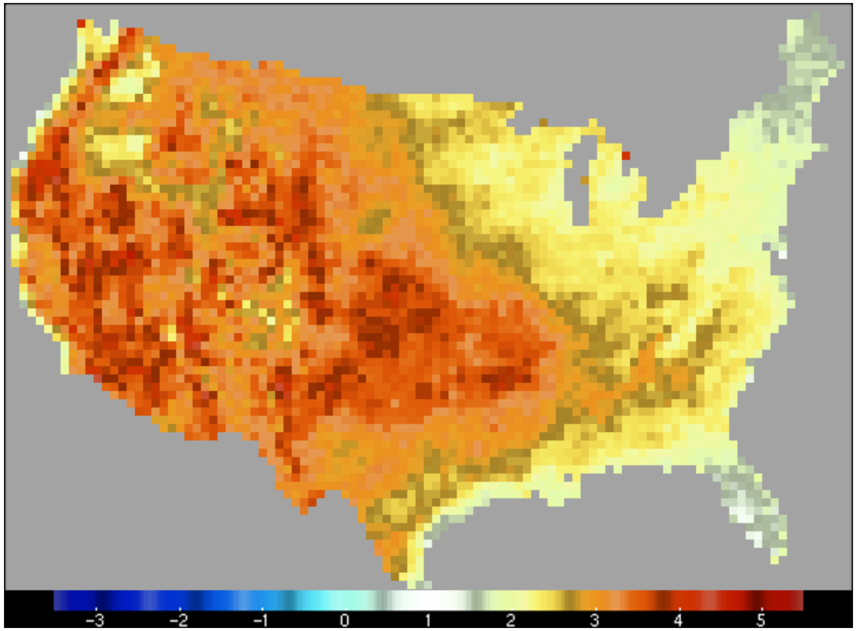


Change in Summer T_{max} (°C, CRCM-ccsm)

Uncorrected



Corrected



Bias Correction is Complicated AND Expensive

- Regridding obs data takes ≈ 20 hours per RCM
  - More I/O- than CPU- or memory-dependent
- Bias-correcting current run takes 2.5 hours
- Bias-correcting future run takes < 1 minute
- Entire process is embarrassingly parallel

Further Complications: Uncertainty and Ensembles

Although users would prefer a crystal ball, uncertainty is important to robust analysis

- Obs are uncertain – use multiple sources
- Package uncertainty as multiple realizations

Many next-generation data products will have ensemble form

So what does all this mean?

- Downloading data to process on desktop wastes resources, especially for impacts
- Big Data needs processing *before* download
- Significant expertise needed to properly distill data into meaningful information
- Experts are a limited resource

→ We Need Data Services

Data Services

Analyze and transform data before transfer to end user

- Reduces the need for large data downloads
- Improves usability for non-specialists, applications
- Captures expertise as automated processing

➢ *Need provenance threaded through all services*

➢ *Intimately related to data archiving & publication*

➢ *Capabilities needed depend on target audience*

A Taxonomy of Data Services

Access services

Transparent; don't alter data

- Subsetting
- Format conversion
- File spanning

Transformation services

On-the-fly changes to data

- Averages, extremes
- Regridding
- Simple math (e.g., vector winds to speed, °C to °F)

Derived data products

Expensive/tricky to generate

- Climatic indices
- Complex calculations (e.g., CAPE)
- Evaluation metrics
- Bias-correction

Viz. & interpretation

Non-data output

- Maps, plots, transects
- Statistical analysis
- Custom services