CORDEX-NA Data & Services

Seth McGinnis & Joseph Barsugli

2013-06-21

Outline

- CORDEX Data Context
- Quality Control
- Data Services
- Leveraging NCPP
- Bias Correction

CORDEX Data Context

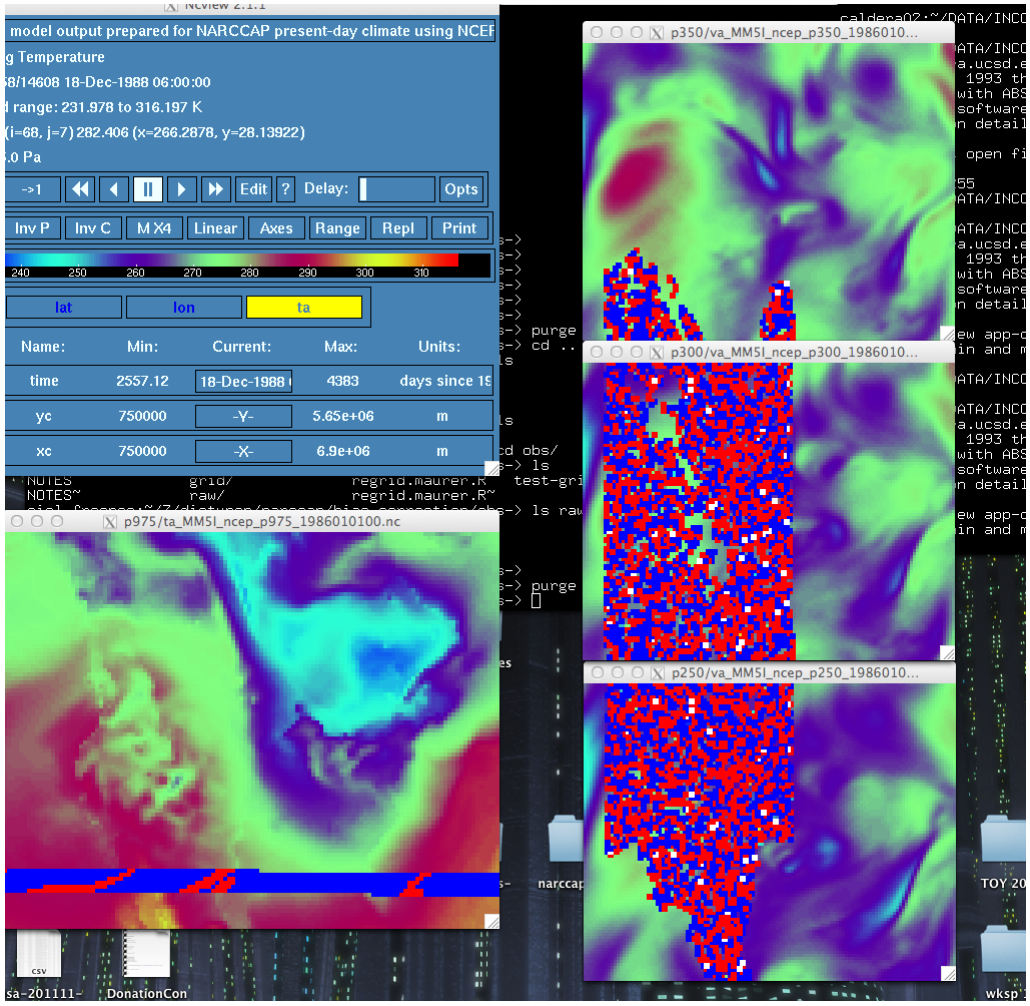
- Central Archive: 70-160 TB
  - Core: monthly and seasonal
  - Tier 1: daily
- Modeling Centers: 90-230 TB each
  - Tier 2: sub-daily
  - 3D: 3-hourly, on all available vertical levels
    not mandated by CORDEX
    needed for downscaling

Quality Control

Errors occur in data processing

- Metadata errors
- Mislabeled data
- Missing data
- Bad data: →

Need systematic checks to find and correct them



QC Process

- QC performed by modeling groups using tools and protocols from central archive
- QC output (data and graphics) published alongside data
- Central archive verifies checksums on data submitted to archive

QC Protocol

- Check metadata for standards compliance
- Check coordinates and ancillary data
- Plot summaries and slices
- Do basic statistical tests

Data Services

Analysis and transformation of data before transfer to end user

- Reduce the need for large data downloads
- Improve usability for applications, non-specialists
- Capture expertise as automated processing

Target Audience Determines Service Requirements

Applications Practitioners

- Need distilled data (e.g., bias-corrected climatology)
- Regional, text/GIS format, a few common variables
- Data feeds into specialized tools, impacts models

Climate Researchers

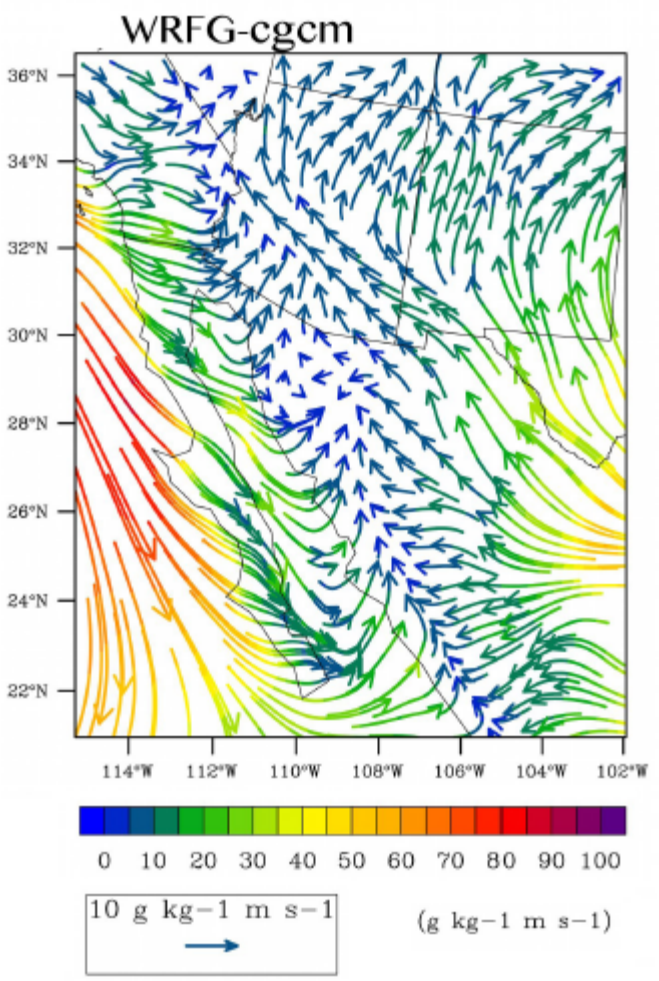
- Need unmodified data for many related variables
- Synoptic fields, long-tail variables, NetCDF, 3D data

Magnitude of dataset also affects service requirements

Process Analysis: Southwest Monsoon

Service Needs:

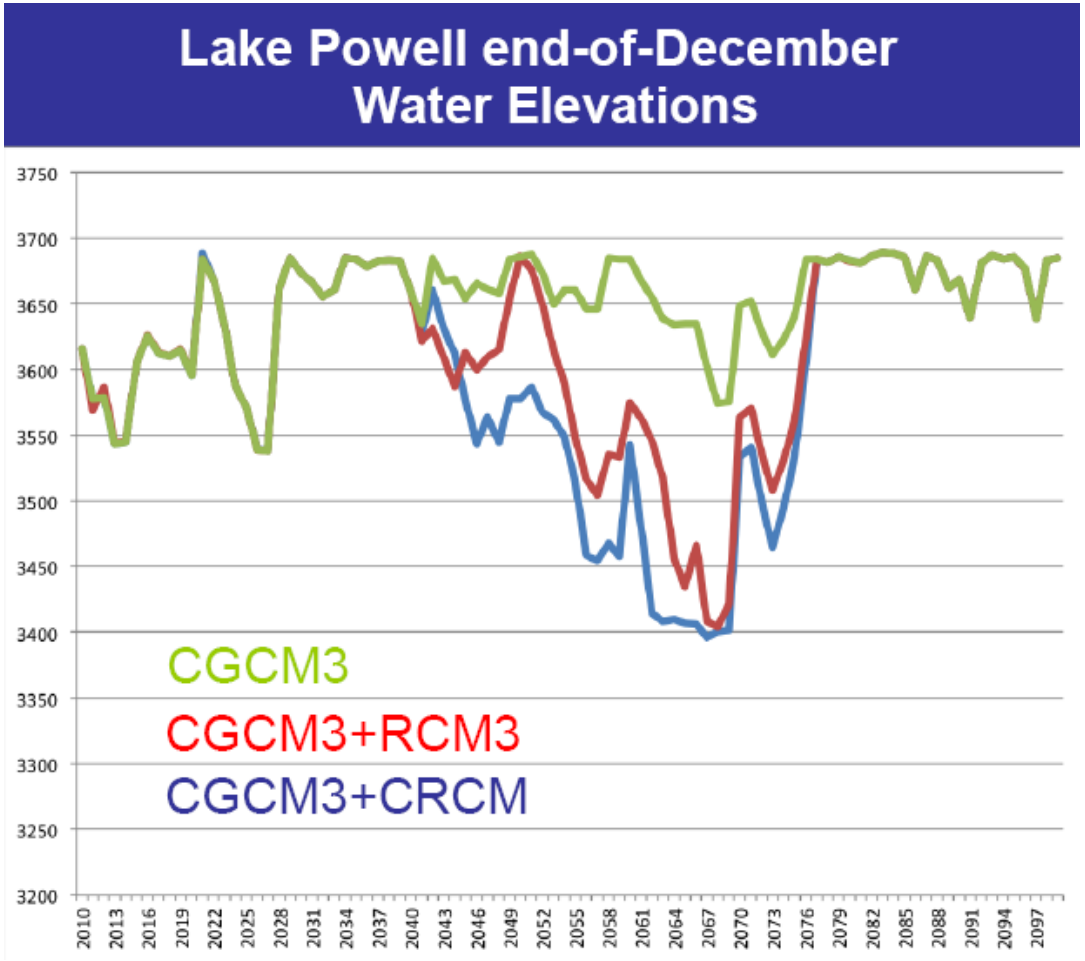
- spatial, temporal subsetting
- simple algebraic transforms
- combinations of variables



Decision Analysis:
Water Resource Management

Service Needs:

- interpolation
- bias-correction



Data Service Categories for CORDEX

- Data access services
- Data transformation services
- Derived data products
- Visualization and interpretation services
- *Provenance threaded through all services*

Data Access Services

Simple operations on the data that do not alter data values and are transparent to the user

- File aggregation
- Geographic and temporal subsetting
- Format conversion

Data Transformation Services

On-the-fly operations that transform data values

- Time and space averaging, extremes
- Interpolation and regridding
- New variables derived via simple algebra
  - e.g., moisture flux, relative humidity, °C to °F
- Downscaling on-demand?

Derived Data Products

Manipulations of data that are too expertise- or resource-intensive to generate dynamically

- Climatic indices (e.g., drought, heatwaves)
- Complex derived variables (e.g., CAPE)
- Evaluation metrics
- Bias-corrected data
  - Note dependence on method and observed data

Visualization and Interpretation Services

Non-data output based on data values. Also operates on output from other services.

- Maps, timeseries plots, transects
  - Low-hanging fruit -- lots of tools in development
- Statistical Analysis?
- Interactive data exploration?
- Customized services developed in collaboration with end-user communities

NCPP: OpenClimateGIS

- Can be part of CORDEX data services (running on archive data portal servers)
- Contributes to interoperability of services
- Plans to integrate with visualization & analysis systems like UVCDAT, Viztrails
- Sophisticated subsetting capabilities
- Currently in development; workshop in August

NCPP: Evaluation Activities

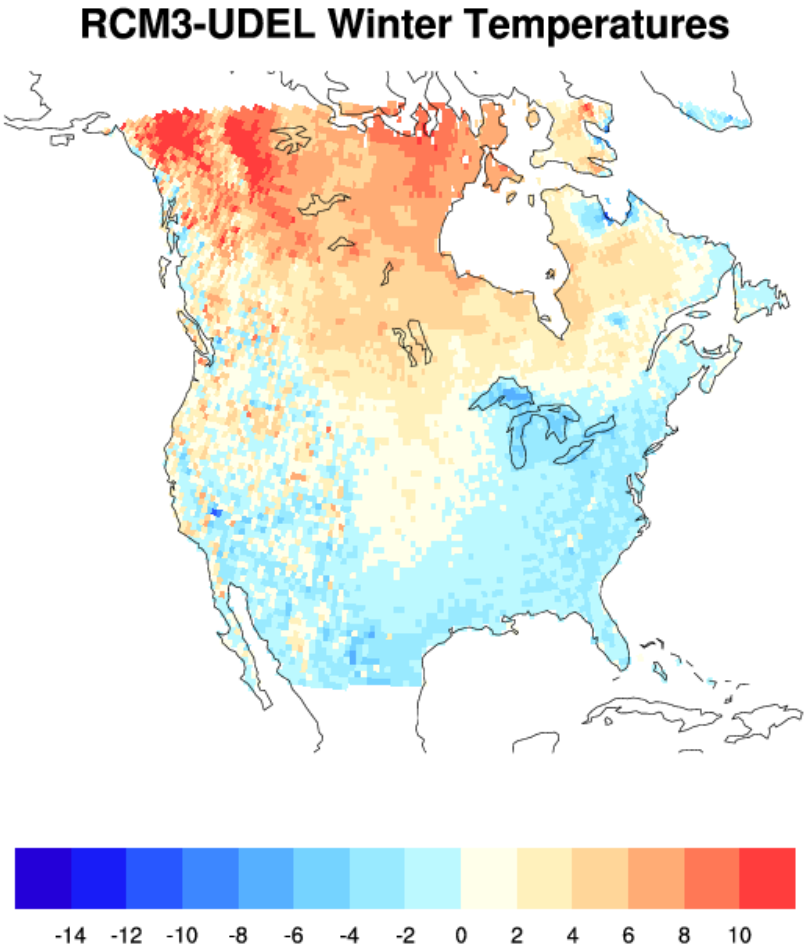
- Evaluation of downscaled data (statistical as well as dynamical)
- CORDEX should work with NCPP to serve evaluation results in ways meaningful to users
- Compute standardized evaluations and publish them alongside the data
- Need to ensure compatibility of CORDEX analysis with NCPP efforts / protocols
- Many of the NCPP evaluations are useful in QC as well

NCPP: CIM

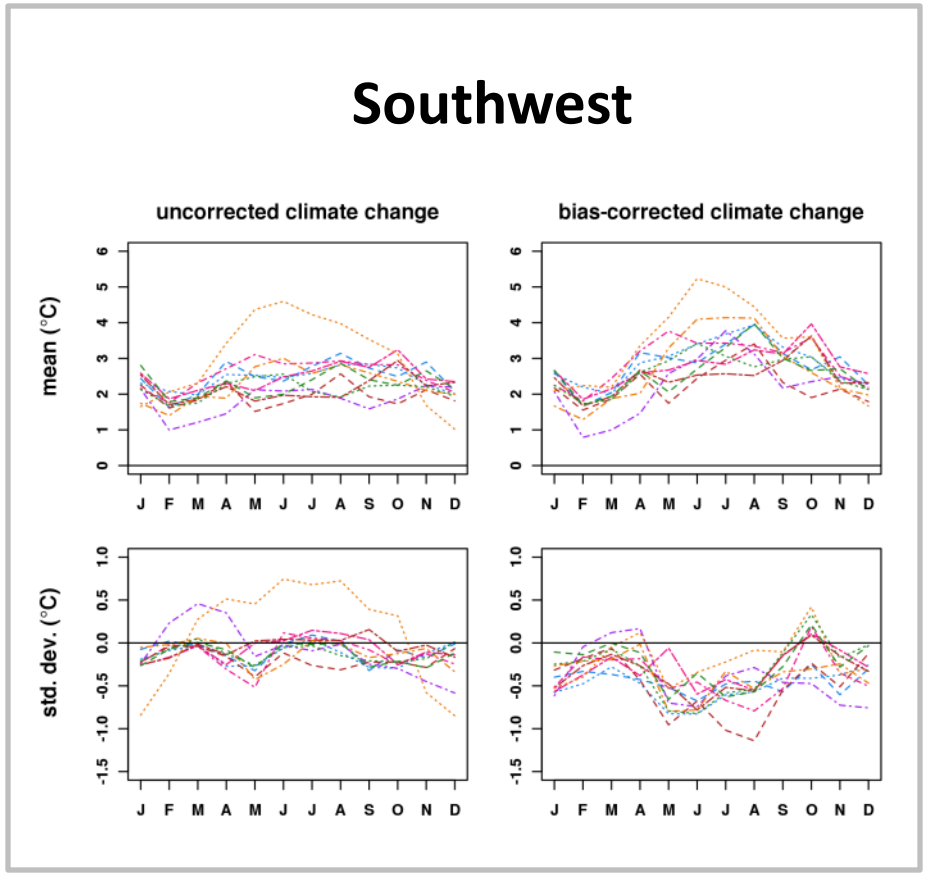
- Metadata standard for downscaling activities
- Customizable web form system for recording metadata (similar systems used in CMIP)
- Will improve CORDEX archive searchability, usability

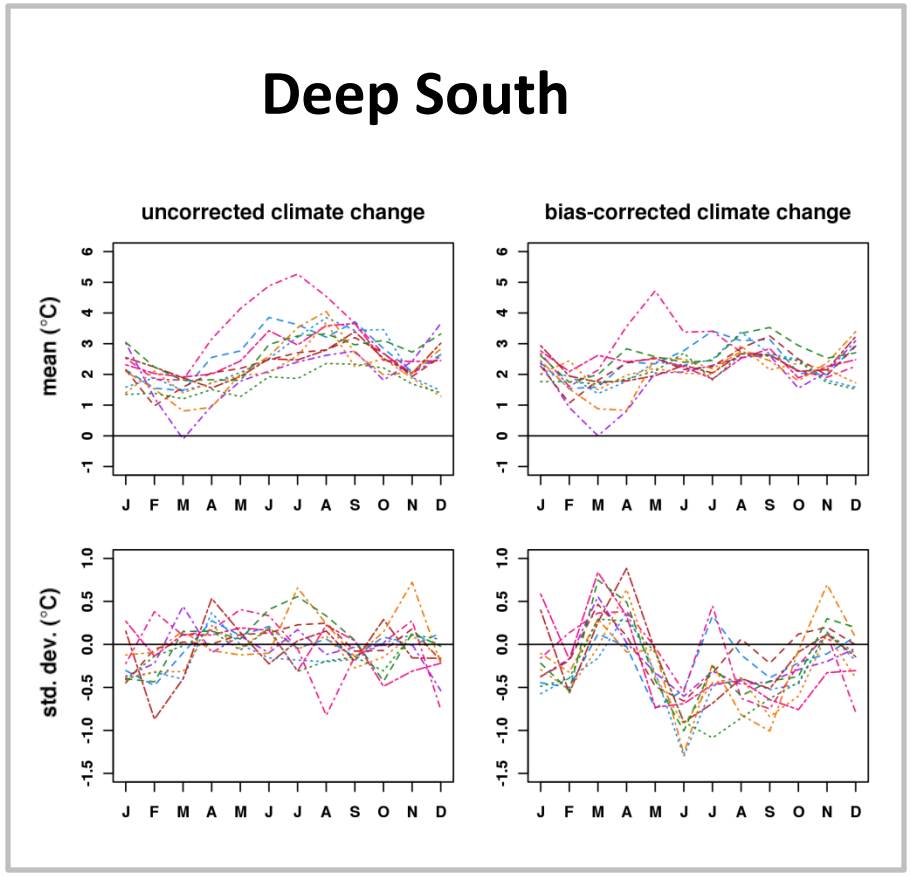
Bias Correction

- Climate models have bias
- Delta method is often used to correct mean bias*
  *assuming stationarity
- “Drizzle problem” requires correcting entire distribution



Bias correction can enhance (L) or diminish (R) climate change signal





Bias Correction Activities

- Using quantile mapping (QM)
  - Corrects entire distribution by construction
- Evaluation of methods in QM library
  - Best option: *non-parametric empirical quantiles*
- Regridding Maurer 1/8° observational dataset

- RCMES for delta-type correction of other vars?