# CORDEX-NA: Data Services

Seth McGinnis & Joe Barsugli

2013-02-20

# The Need for Data Services

**Data Services:** <span style="color:red">Analysis and transformation of data before transfer to end user</span>
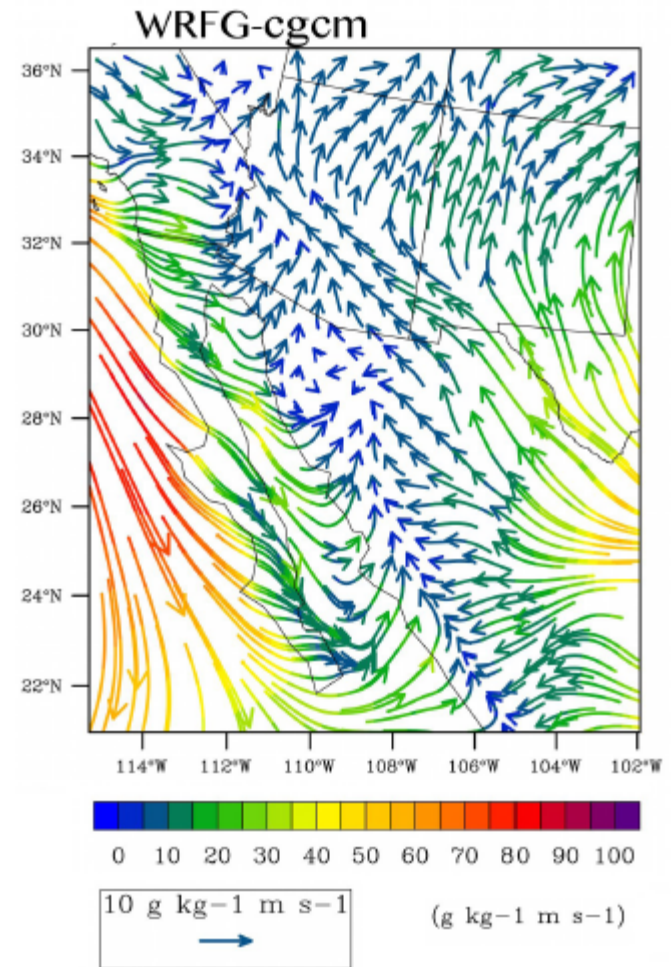
- Reduce need for large data downloads
  - Especially important as data gets bigger
- Improve usability for applic'ns, non-specialists
- Capture expertise as automated processing

Use-Case examples follow

# Process Analysis: Southwest Monsoon

Needs:

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

# Statistical Analysis: Heatwaves
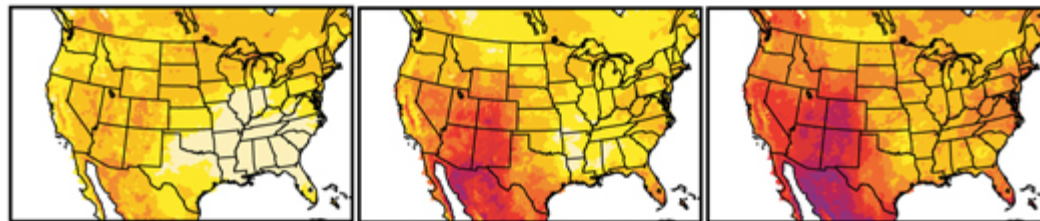
Needs:

- thresholds, sequences
- parallelization (current code takes 1 day/model run)

Number of Extremely Hot Seasons Per Decade

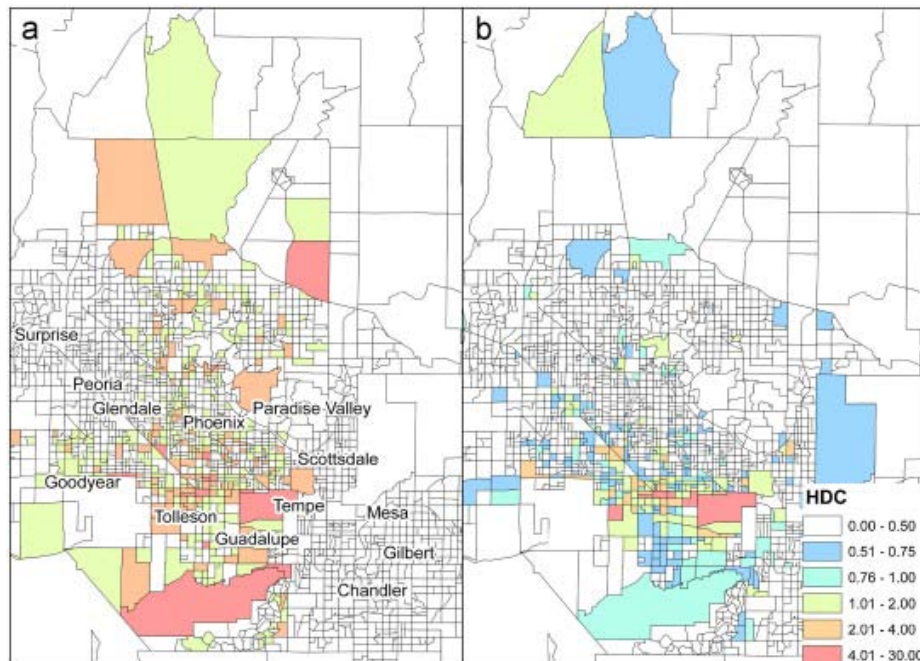2010-2019    2020-2029    2030-2039

events per decade

0    3    6    9

# Impacts Analysis: Urban Heat Stress

Needs:

- GIS compatible
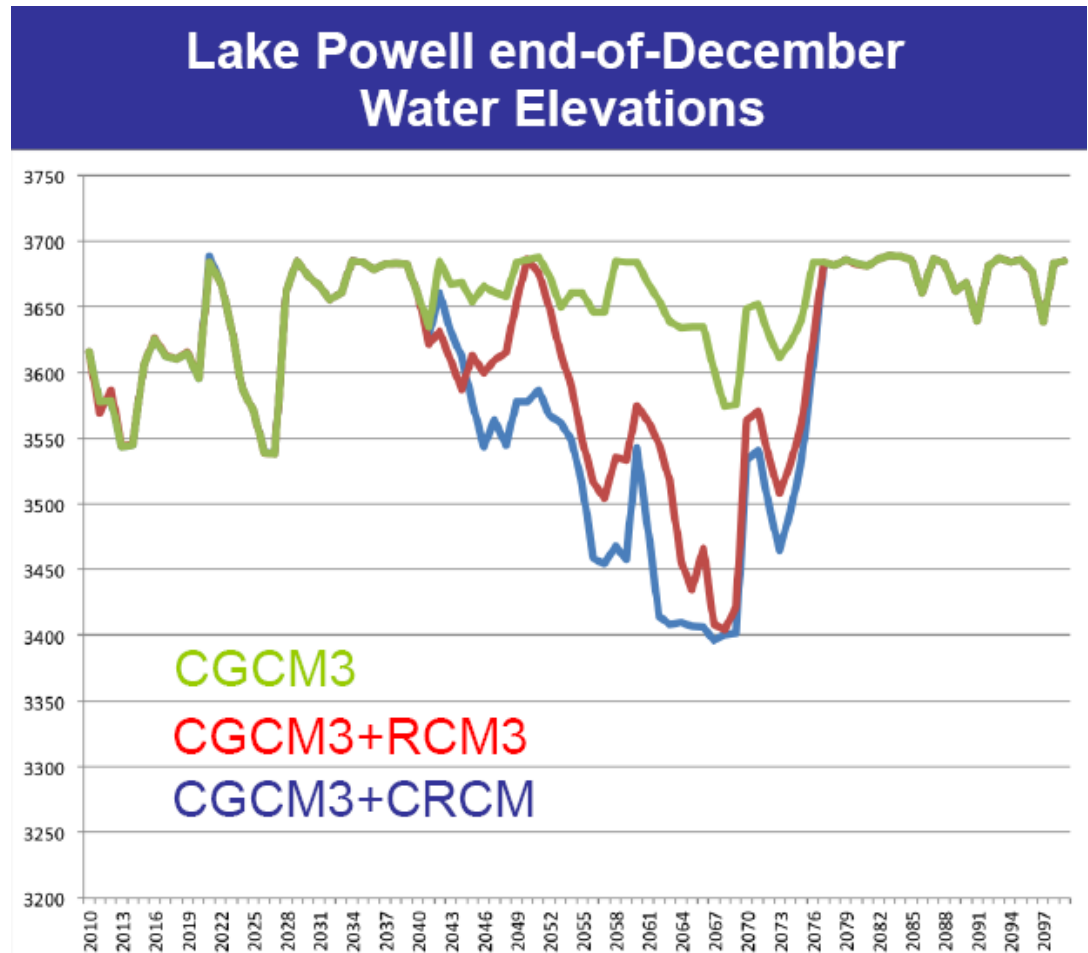
- spreadsheet-friendly

- error, uncertainty

# Decision Analysis:
# Water Resource Management

Needs:

- interpolation
- bias-correction



Lake Powell end-of-December Water Elevations

CGCM3
CGCM3+RCM3
CGCM3+CRCM
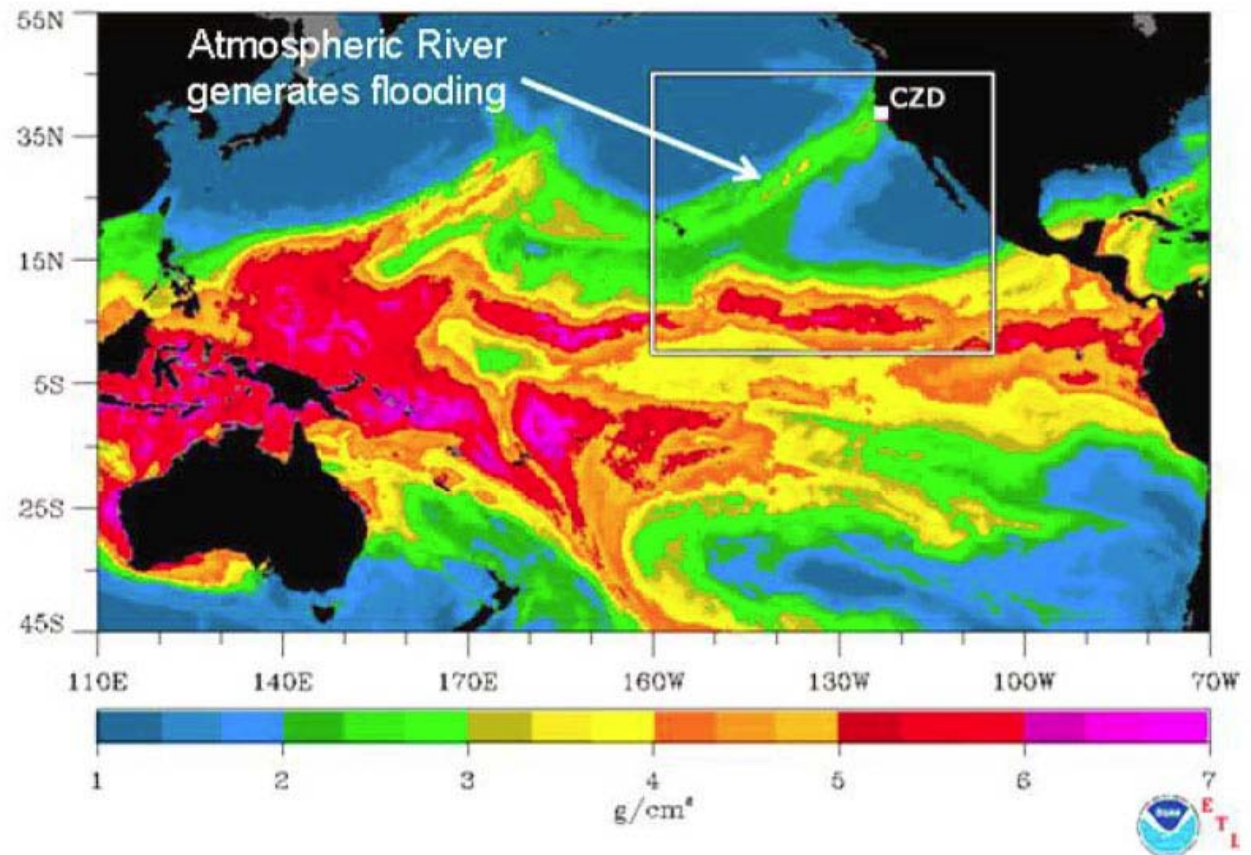
# Event Analysis: Atmospheric Rivers

Needs:

- object identification

- fast data search

# Target Audience Determines Service Requirements

**Applications Practitioners**

- Need highly-distilled data (bias-corrected climatology)
- Small data: regional, mostly 2D, especially "Big 7":
  Tmin, Tmax, precip, humidity, wind, incoming solar, pressure
- Data feeds into specialized tools, impacts models

**Climate Researchers**

- Need unmodified data for many related variables
- Big data: domain-wide fields, many variables, 3D
  – Further dynamical downscaling: 3D *and* high-frequency
- General-purpose analysis software reads NetCDF

# Data Service Categories for CORDEX

- Data access services
- Data transformation services
- Derived data products
- Visualization and interpretation services
- Data archiving*

  *discussed during afternoon session

- *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

Many tools currently in development
(e.g. OpenClimateGIS, new capabilities for ESG)

# 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

Efforts are underway to integrate data portals with analysis packages like UV-CDAT and NCL

# Derived Data Products

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

- 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

- Operate on raw data and on outputs from other services

- Visualization: maps, timeseries plots, transects
  - Low-hanging fruit -- lots of tools in development

- Customized services developed in collaboration with end user communities
  - Valuable but resource-intensive to develop

# Issues

- Pushing QC back to modelers
- Reducing unnecessary model diversity
- Increases in data volume
- Applications needs vs research needs
- Supported use-cases based on science, apps
- User-friendliness
- Engagement with users

- Existing tools vs new development
- Requirements (standards compliance) for interface with existing tools
- Separate, more capable portal for applications?
- Resource requirements (e.g., 5 FTEs for QC?)
- Prioritization!