Microsoft® Research
Faculty Summit 2010
Guarujá, Brasil | May 12 – 14 | In collaboration with FAPESP

Environmental eScience, *The Fourth Paradigm*, and the end of stationarity

Jeff Dozier
Professor of Environmental Science & Management
University of California, Santa Barbara

Changes in climate and hydrology are not just "global warming"

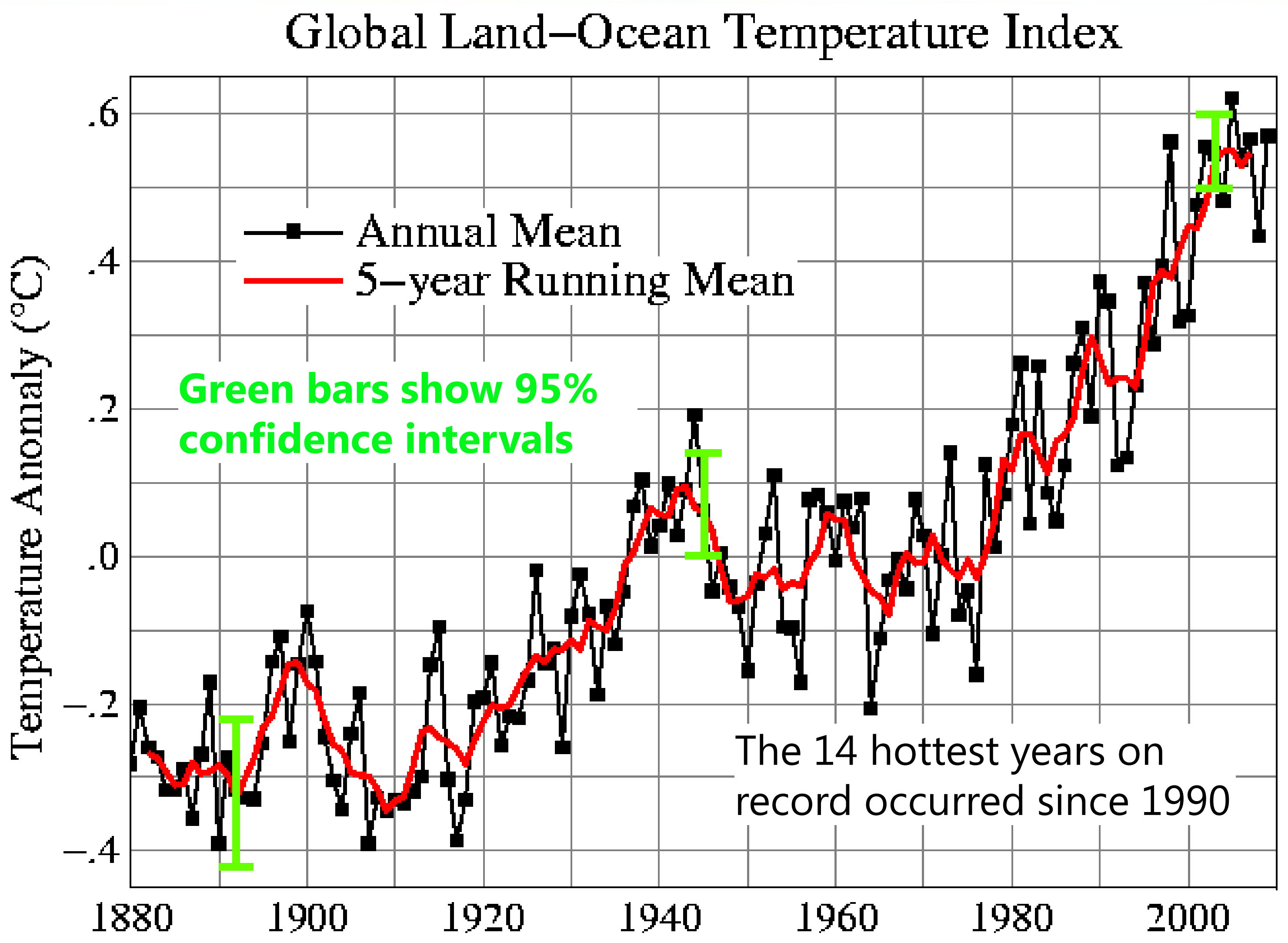
- The term "global warming" implies . . .
  - uniform across the planet
  - mainly about temperature
  - gradual
  - quite possibly benign

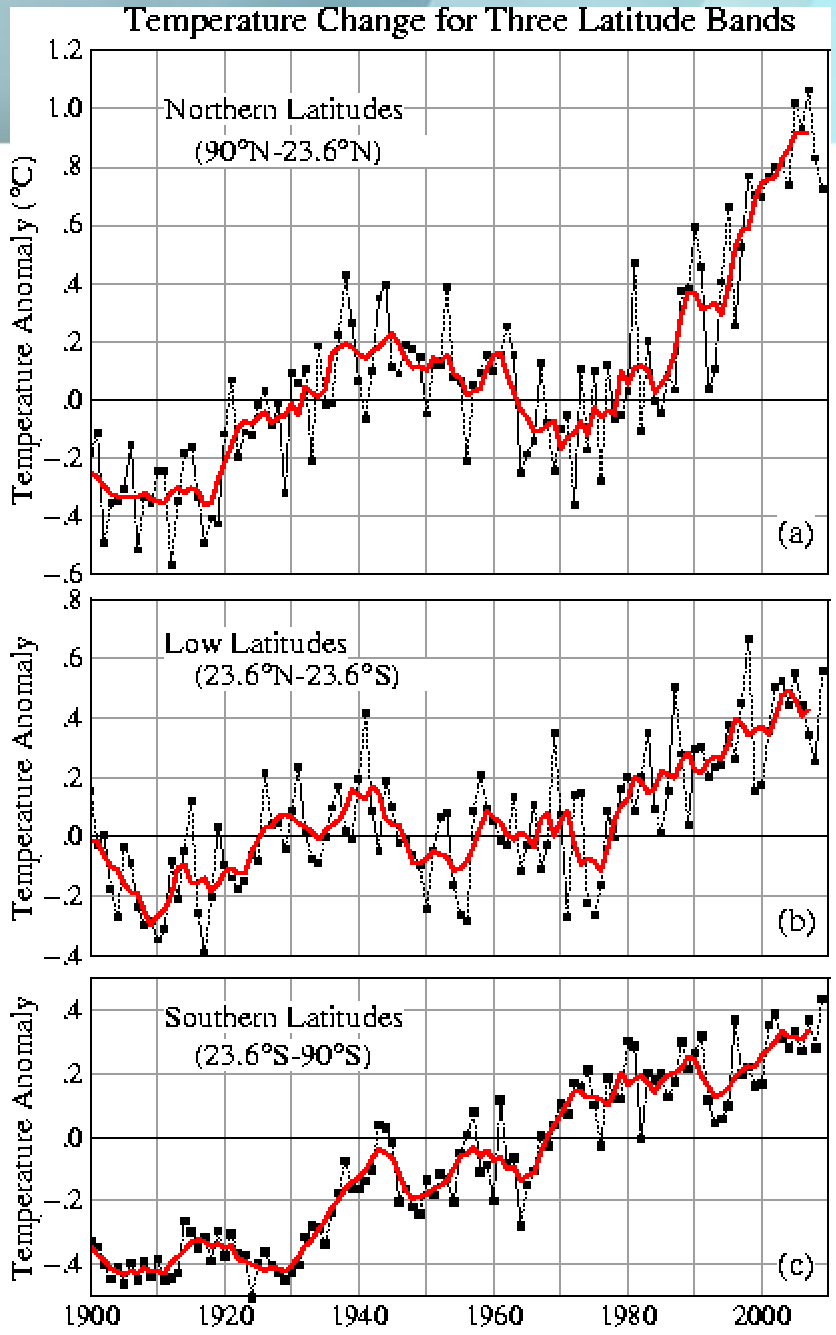
- What is actually happening is . . .
  - highly nonuniform
  - not just about temperature, but especially about water
  - rapid compared to capacities for adjustment
  - harmful for most places and times

We should call it "global climate disruption"

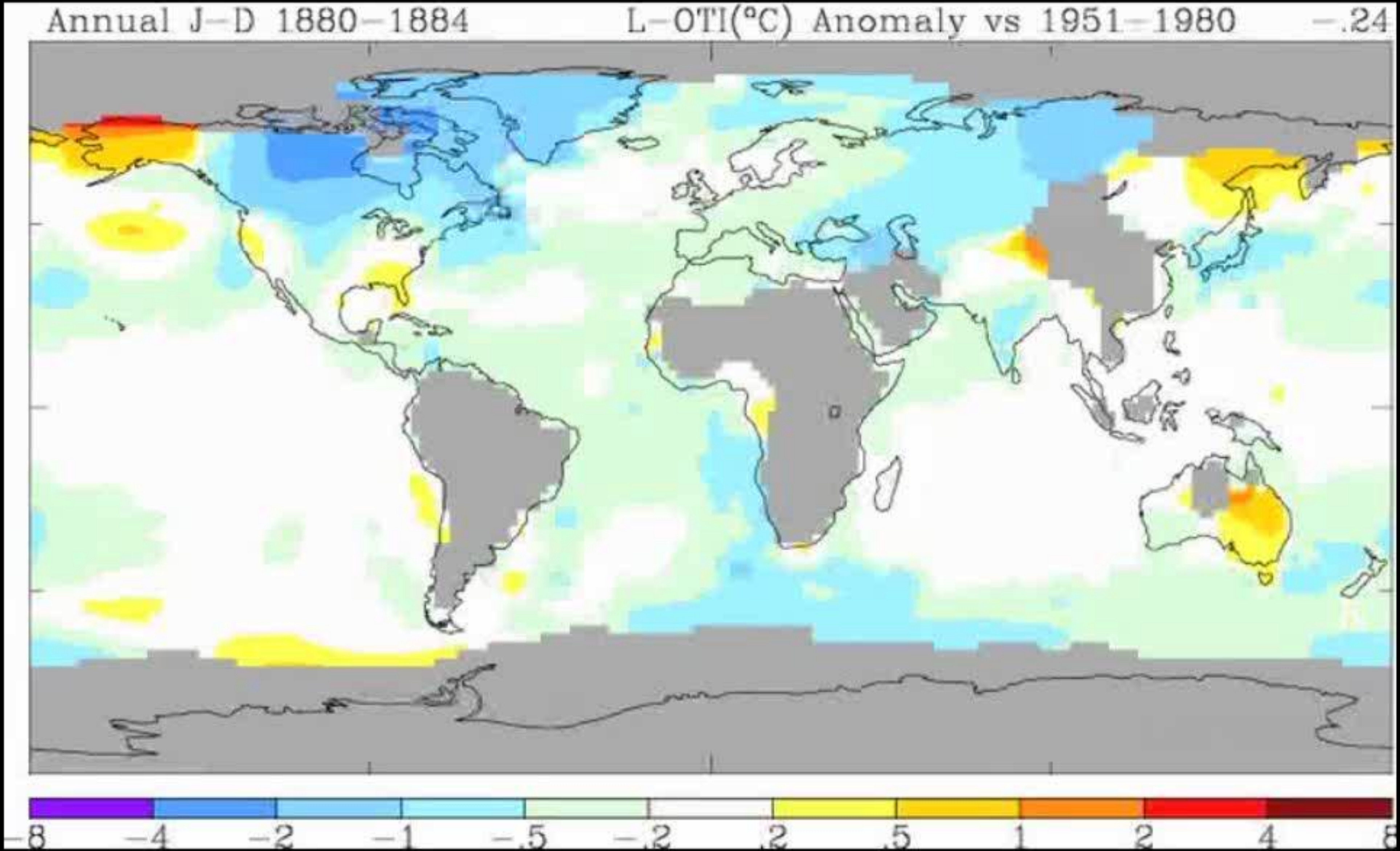
from John Holdren, US Presidential Science Advisor

The temperature record



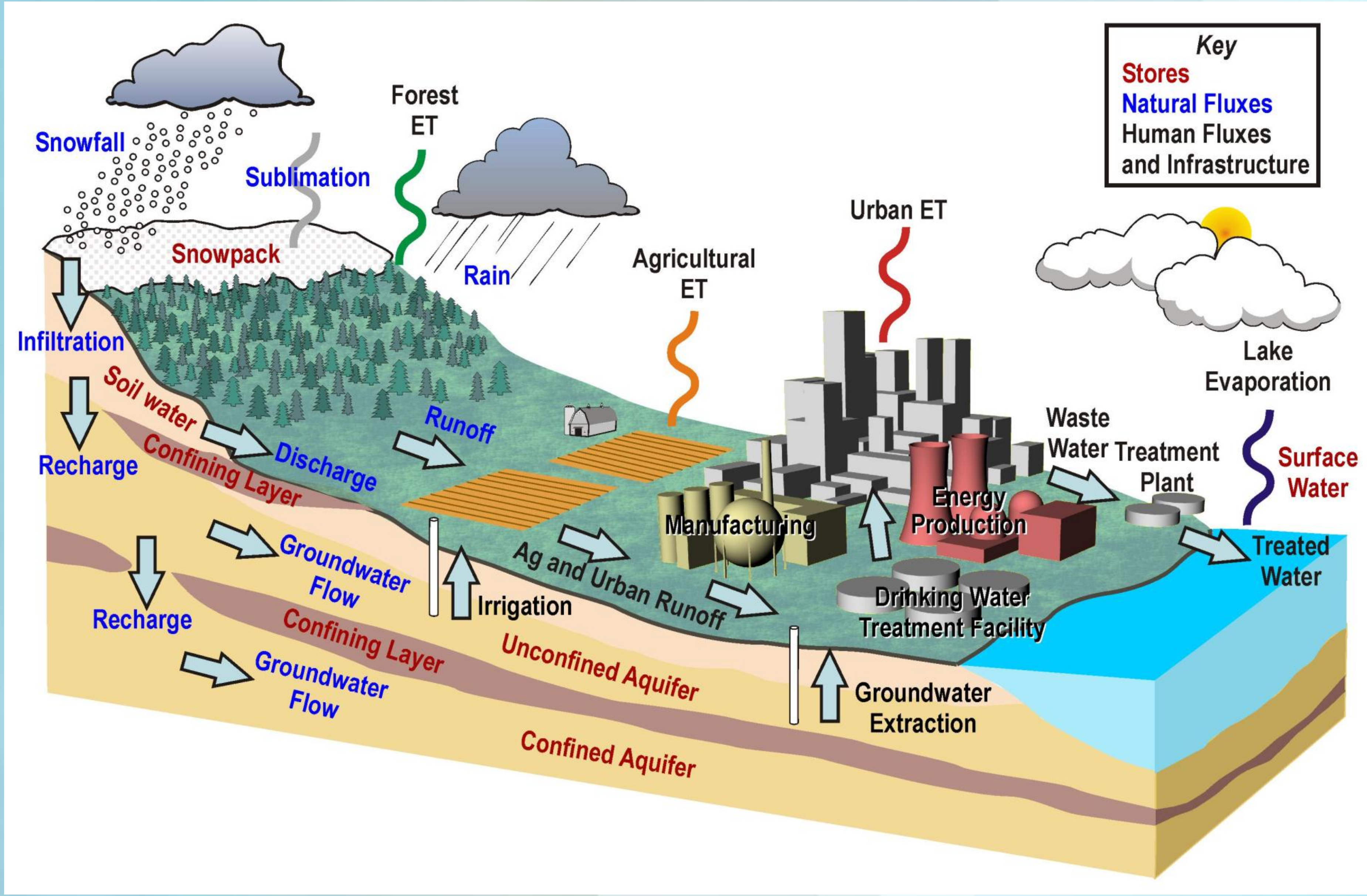


http://data.giss.nasa.gov/gistemp/graphs/

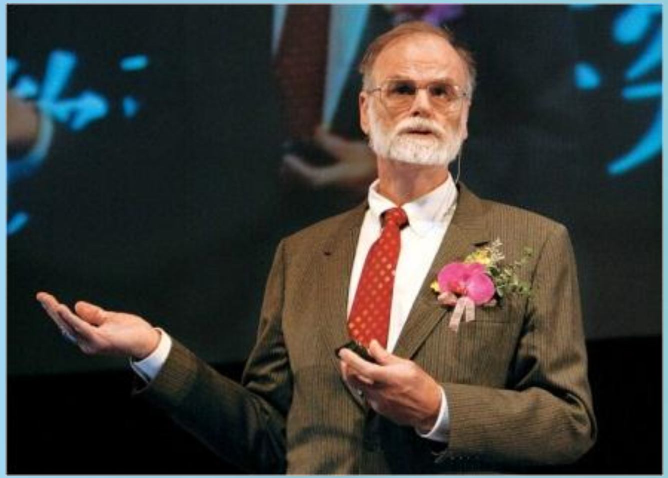

Annual J-D 1880-1884
L-OTI(°C) Anomaly vs 1951-1980
-.24
-8
-4
-2
-1
-.5
-.2
.2
.5
1
2
4
8

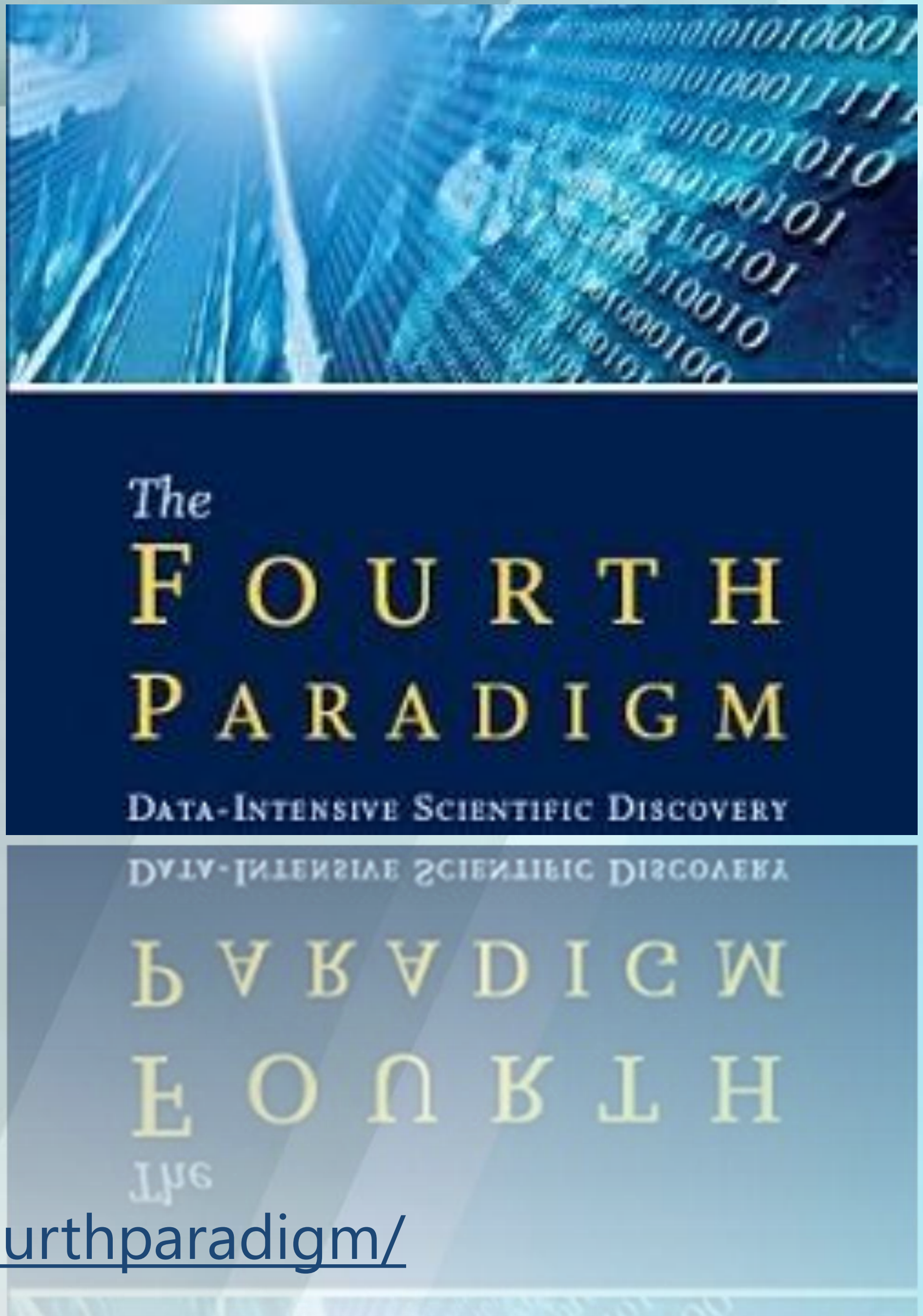
The human-impacted hydrologic cycle

How can we protect ecosystems and better manage and predict water availability and quality for future generations, given changes to the water cycle caused by human activities and climate trends?



The Fourth Paradigm



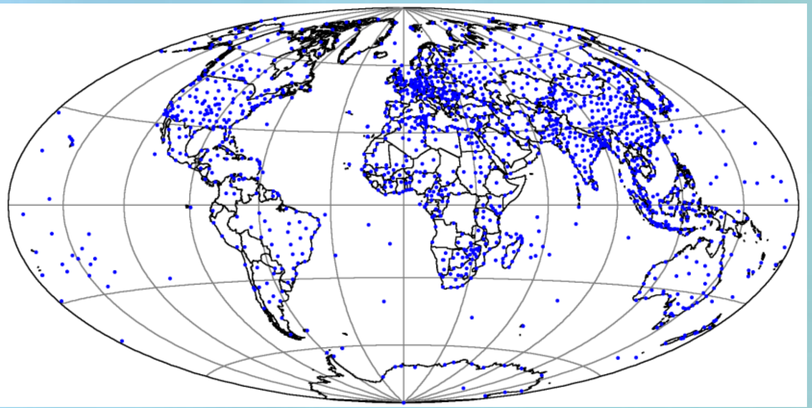


- **An "exaflood" of observational data requires a new generation of scientific computing tools to manage, visualize and analyze them**

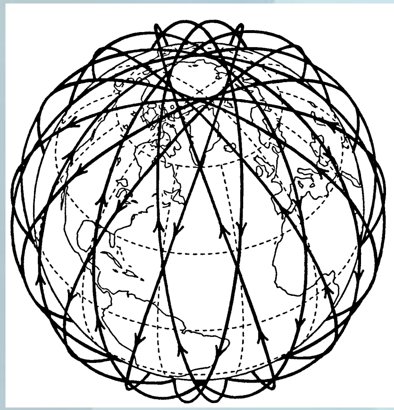
http://research.microsoft.com/en-us/collaboration/fourthparadigm/

A new synoptic view

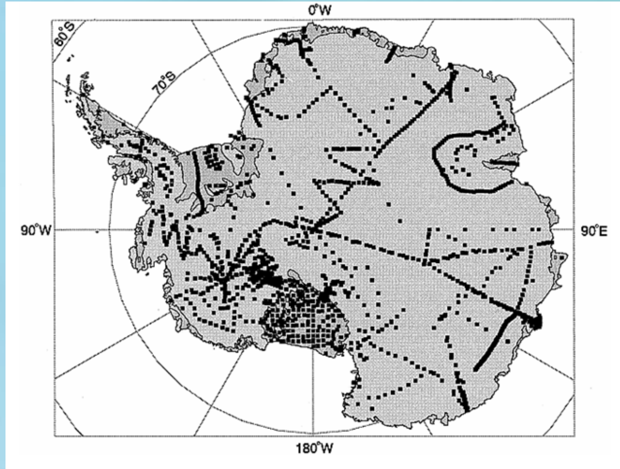
Earth Observations from Space, 2008



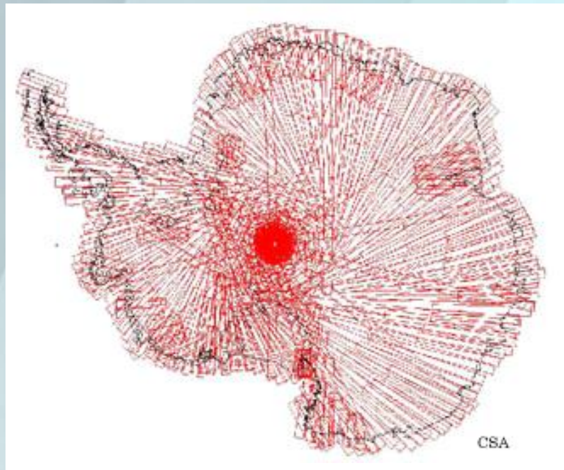
Ground-based weather observations



One day's satellite coverage

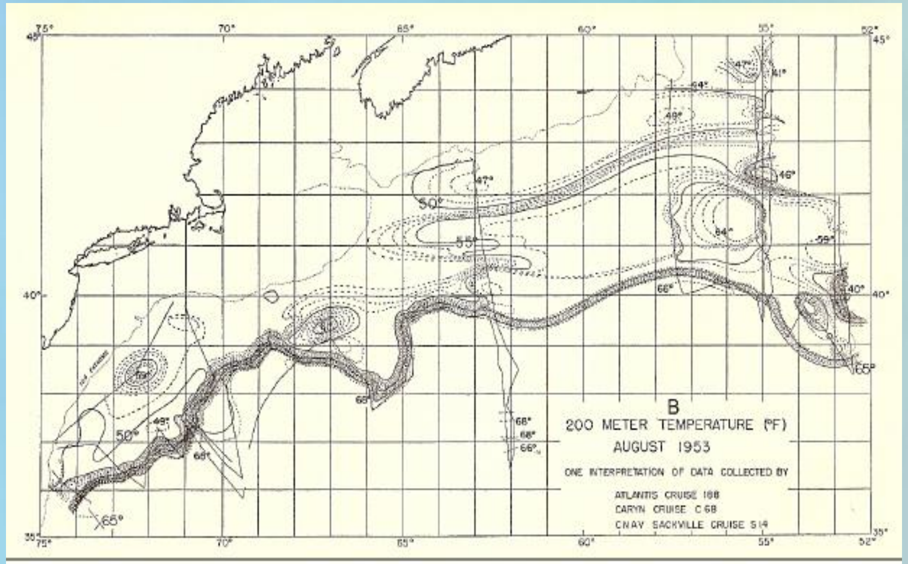


Antarctic ground-based observations since 1957

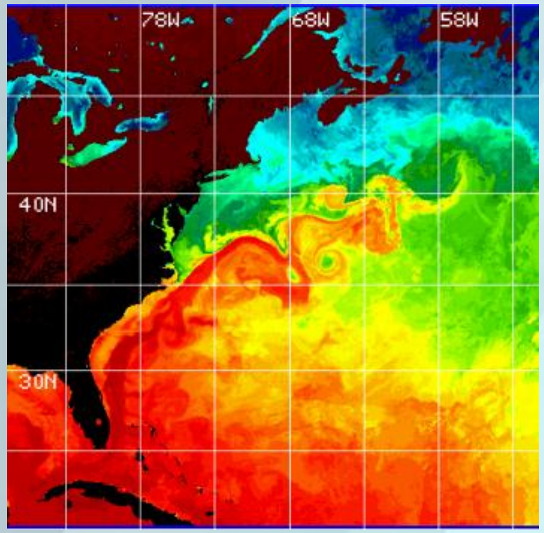


Satellite coverage

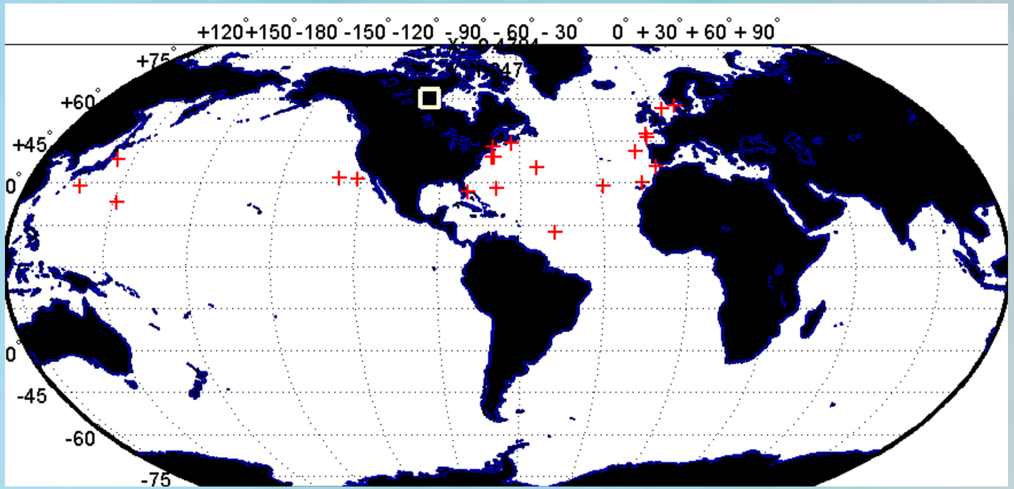
New understanding emerges



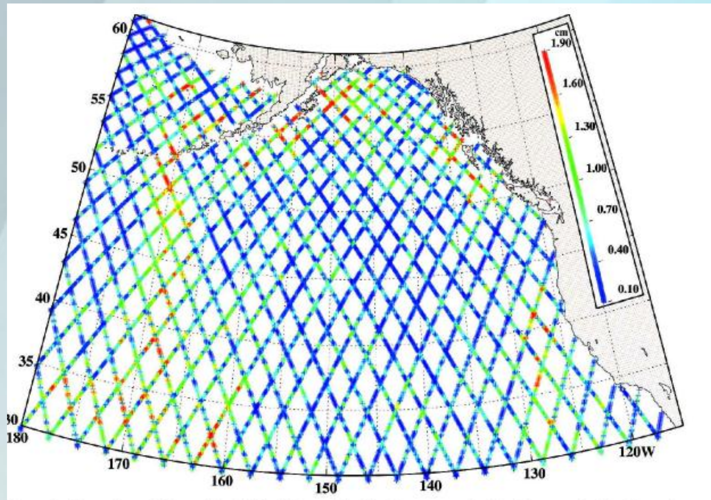
Stommel 1965: Multiple Gulf Stream hypothesis



1st satellite 1970s: Single current



Internal tide information 1975



Internal tides from altimetry

Along with *The Fourth Paradigm*, an emerging science of environmental applications

1. **Thousand years ago —experimental science**
   - **Description of natural phenomena**
2. **Last few hundred years —theoretical science**
   - **Newton's Laws, Maxwell's Equations . . .**
3. **Last few decades — computational science**
   - **Simulation of complex phenomena**
4. **Today — data-intensive science**

(from Tony Hey)

1. **1800s → ~1990 — discipline oriented**
   - **geology, atmospheric science, ecology, etc.**
2. **1980s → present — Earth System Science**
   - **interacting elements of a single complex system (Bretherton)**
   - **large scales, data intensive**
3. **Emerging today — knowledge created to target practical decisions and actions**
   - **e.g. climate change**
   - **large scales, data intensive**

What is different about applications science?

Core characteristics of environmental applications science

- Need driven vs curiosity driven
- Externally constrained
- Consequential and recursive
- Useful even when incomplete
- Scalable
- Robust
- Data intensive

New knowledge types and new tools for acquiring knowledge

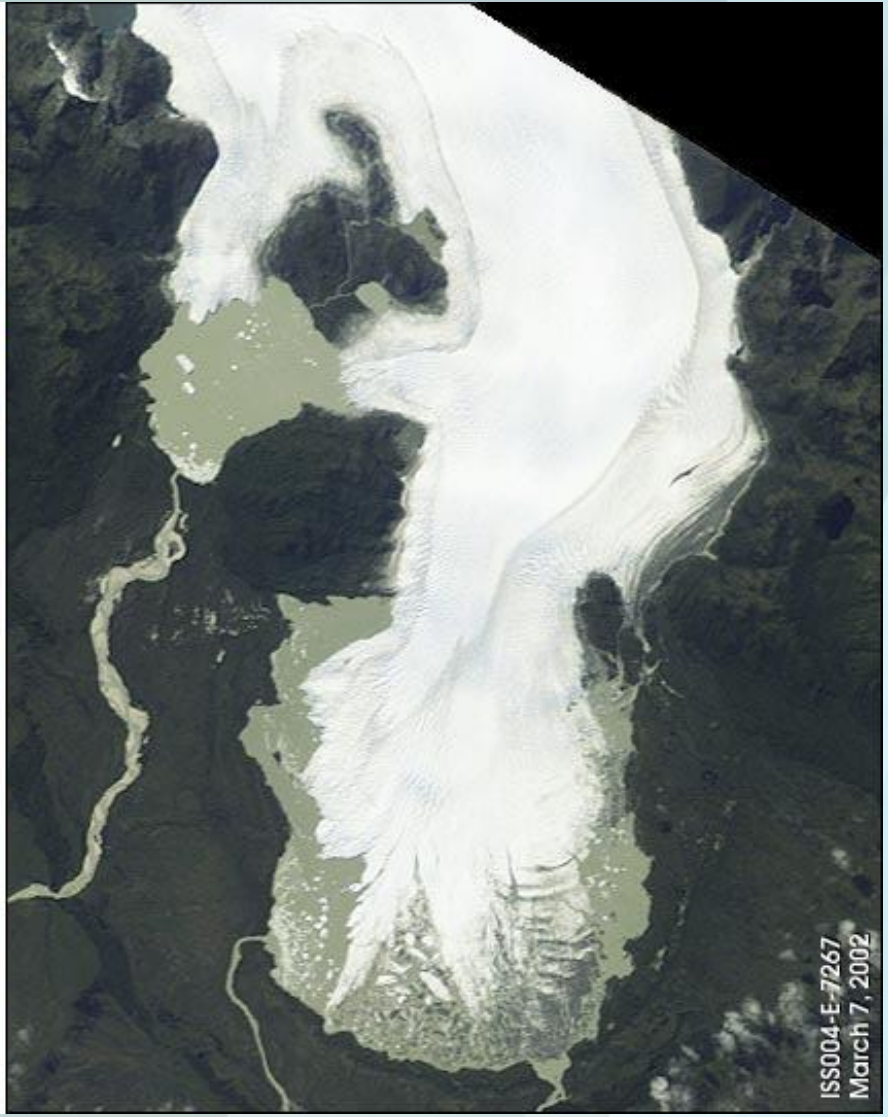
- Remote sensing
- Low-cost sensors and telemetry
- Social data to analyze decision making
- Cyberinfrastructure

“We seek solutions. We don't seek—dare I say this?—just scientific papers anymore”

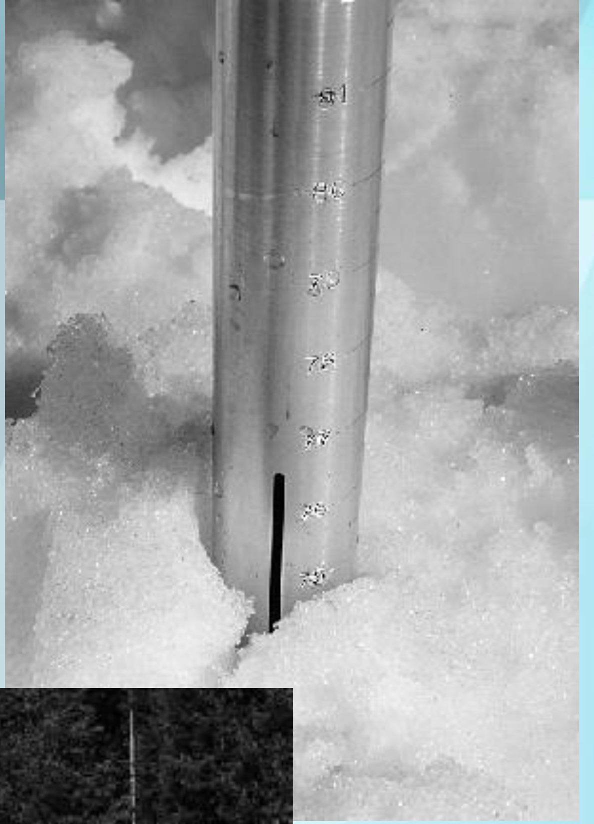
Steven Chu
Nobel Laureate
US Secretary of Energy

Shrinking ice, Muir Glacier (SE Alaska) in 1941 and 2004

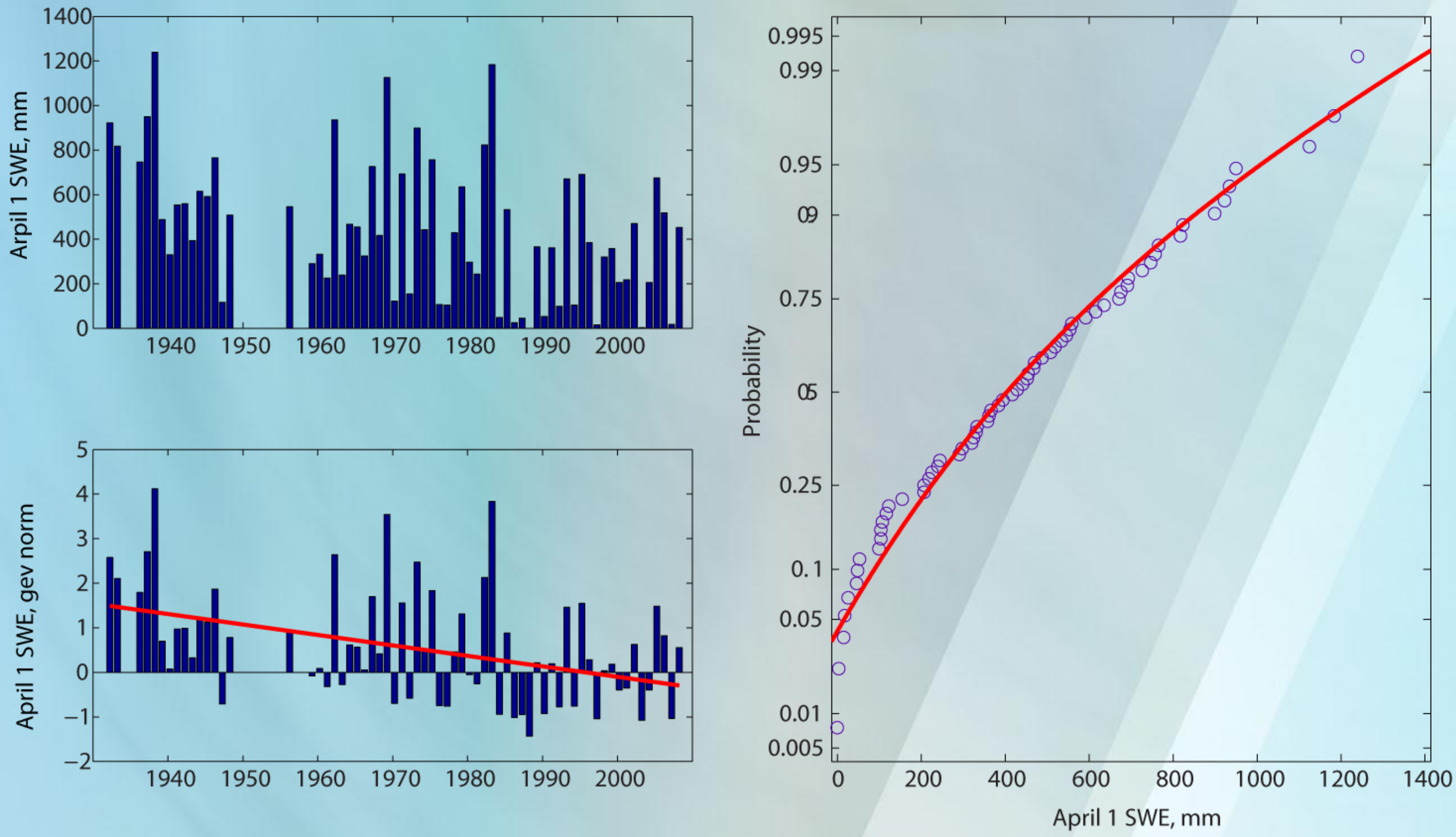
Shrinking ice, Chilean Patagonia, San Quintin Glacier, 1994 to 2002



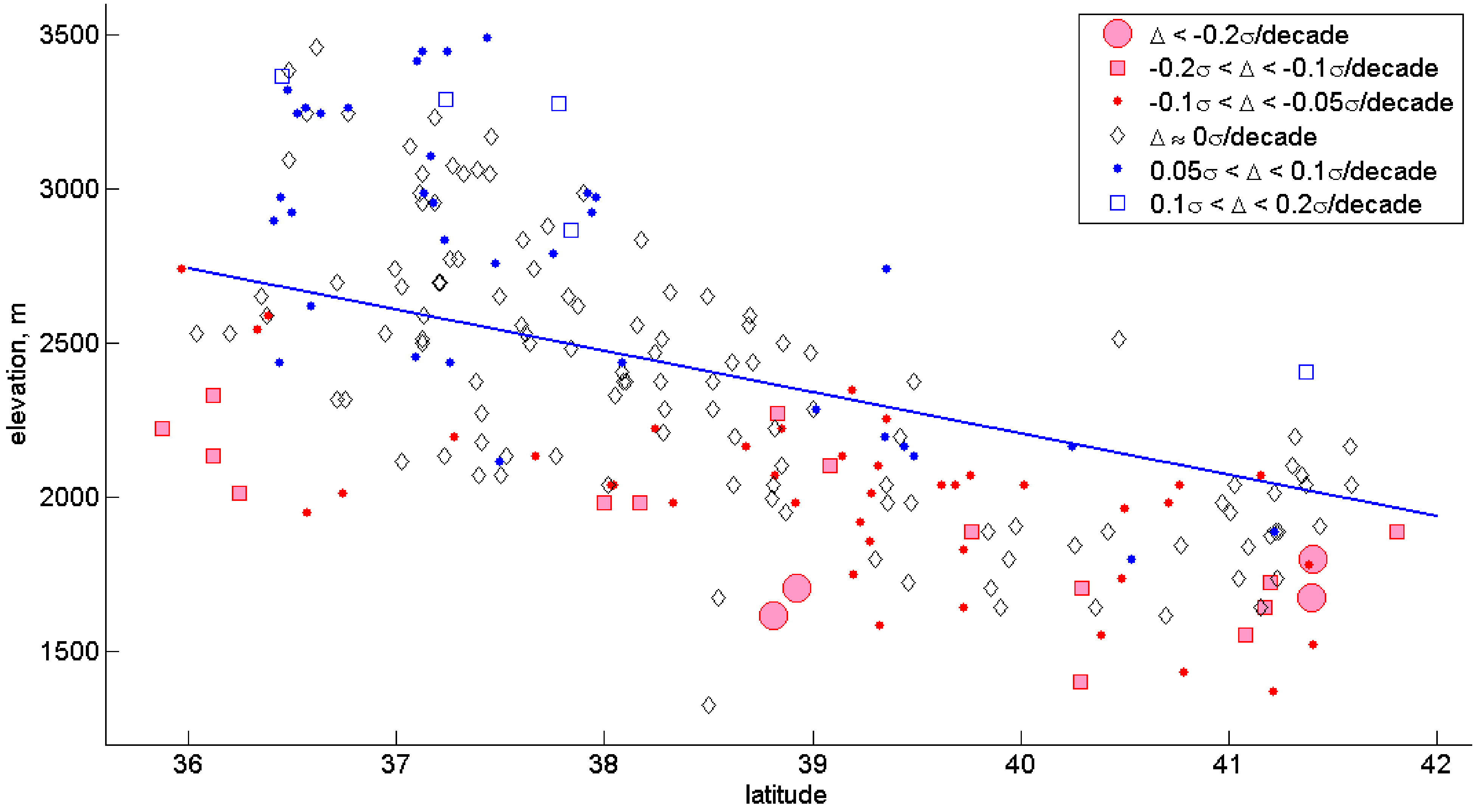
Manual measurement of SWE (snow water equivalent), started in the Sierra Nevada in 1910



Snow course Rubicon Valley, elev 1707m (American R)



Sierra Nevada, trends in 220 long-term snow courses (> 50 years, continuing to present)

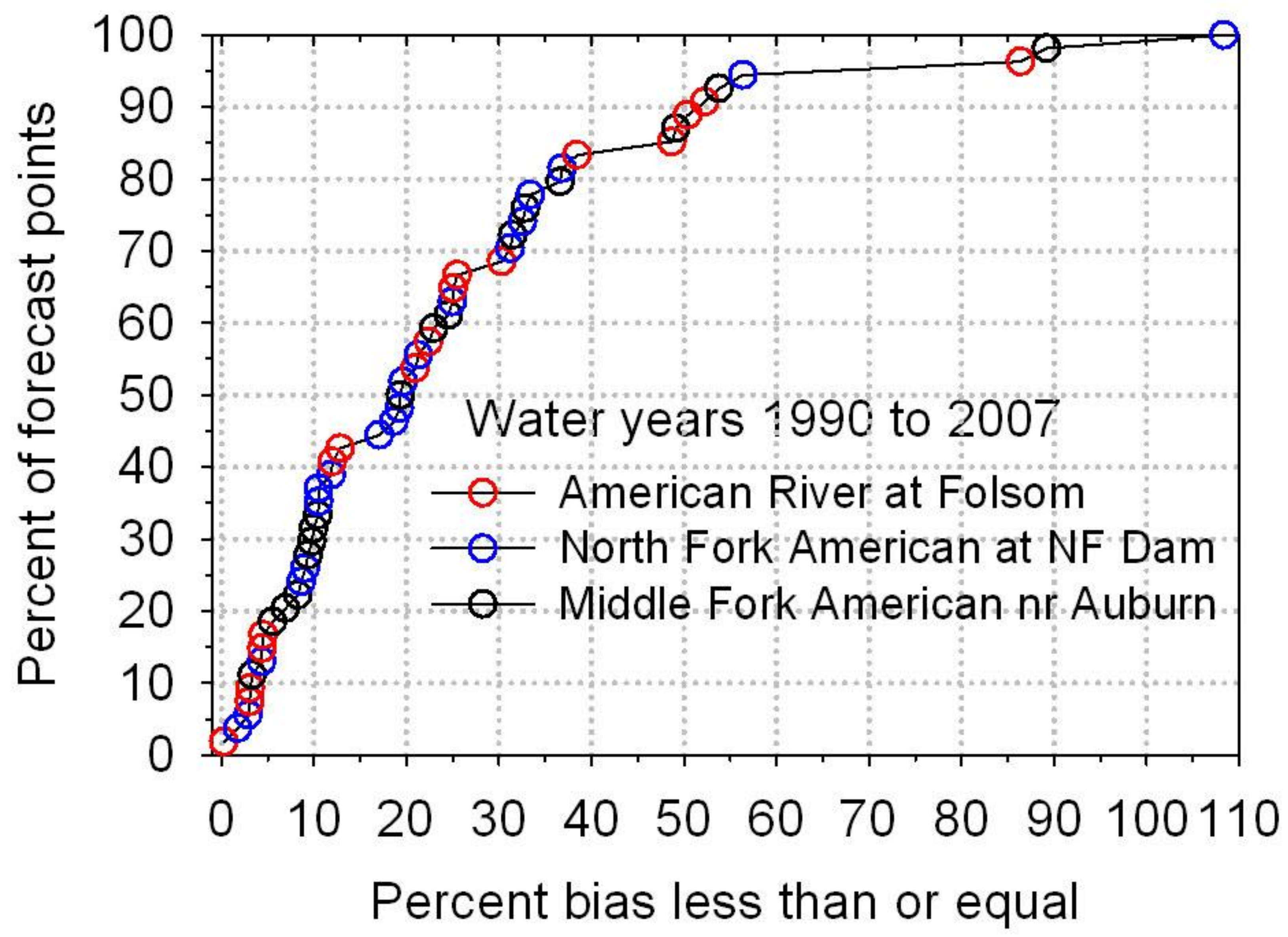


Example forecast, April 2010

- American River below Folsom Lake, April-July unimpaired runoff (units are km^3)

50-yr mean	Max	Min	This year	% of avg	80% prob range
1.530	3.792	0.282	**1.295**	85%	0.95–2.10

Errors in April-July forecast, American River



We manage water poorly . . .

- **We do not predict and manage water and its constituents well**
  - Despite large investments, we suffer from droughts, floods, stormwater, erosion, harmful algal blooms, hypoxia, and pathogens with little warning or prevention
- **Current empirical methods were developed over a period when human impacts were isolated and climate trends slower**
  - Drivers are climate change, population growth and sprawl, land use modification
  - Milly et al., *Science* 2008: Stationarity is dead: whither water management?
- **We need to better understand how/when to adapt, mitigate, solve, and predict**
  - More physically based, less empirical, methods are needed

The water information value ladder

Data >>> Information >>> Insight

>>> Increasing value >>>

Forecasting

Reporting

Analysis

Integration

Done poorly

Distribution

Aggregation

Done poorly to moderately

Quality assurance

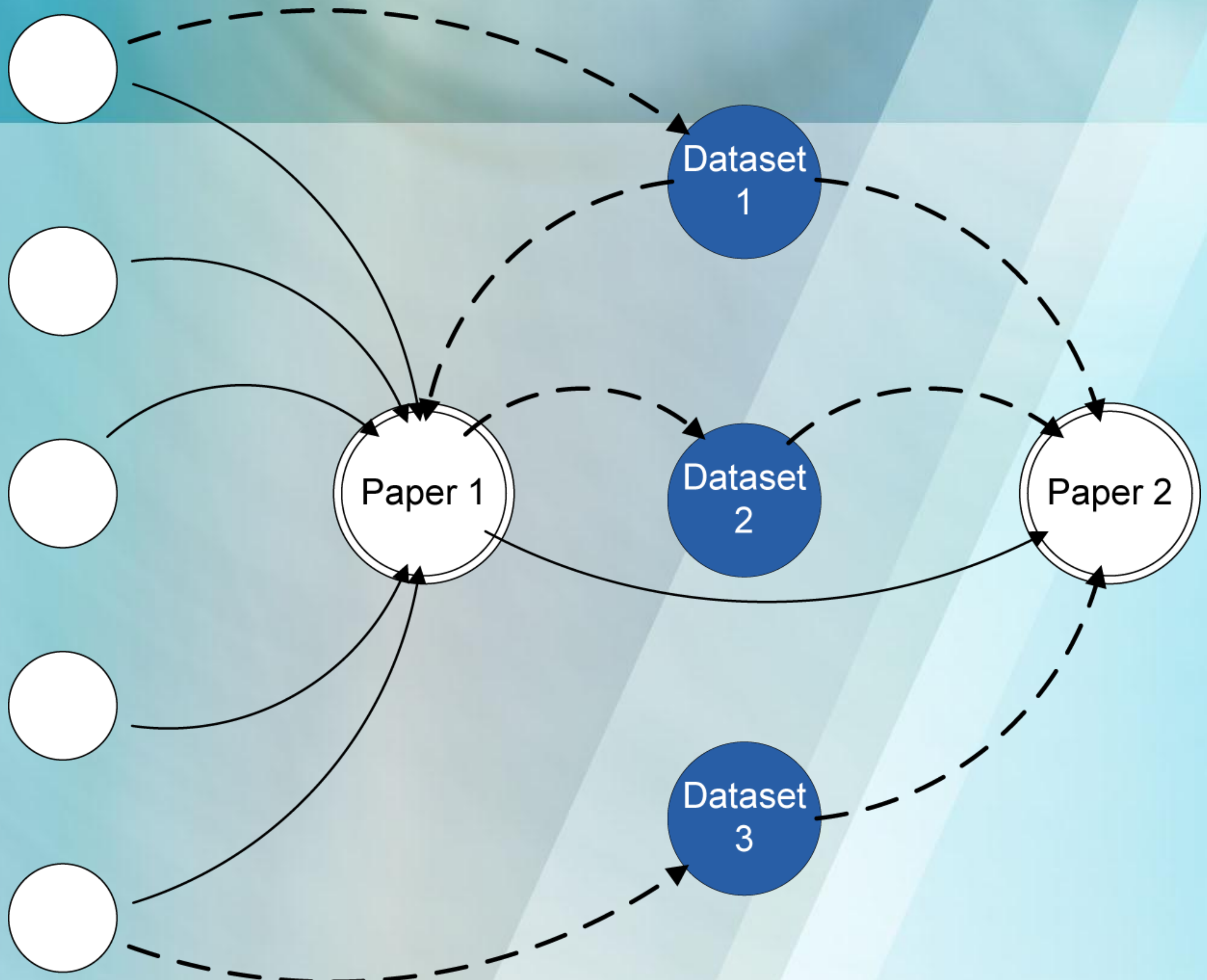
Collation

Monitoring

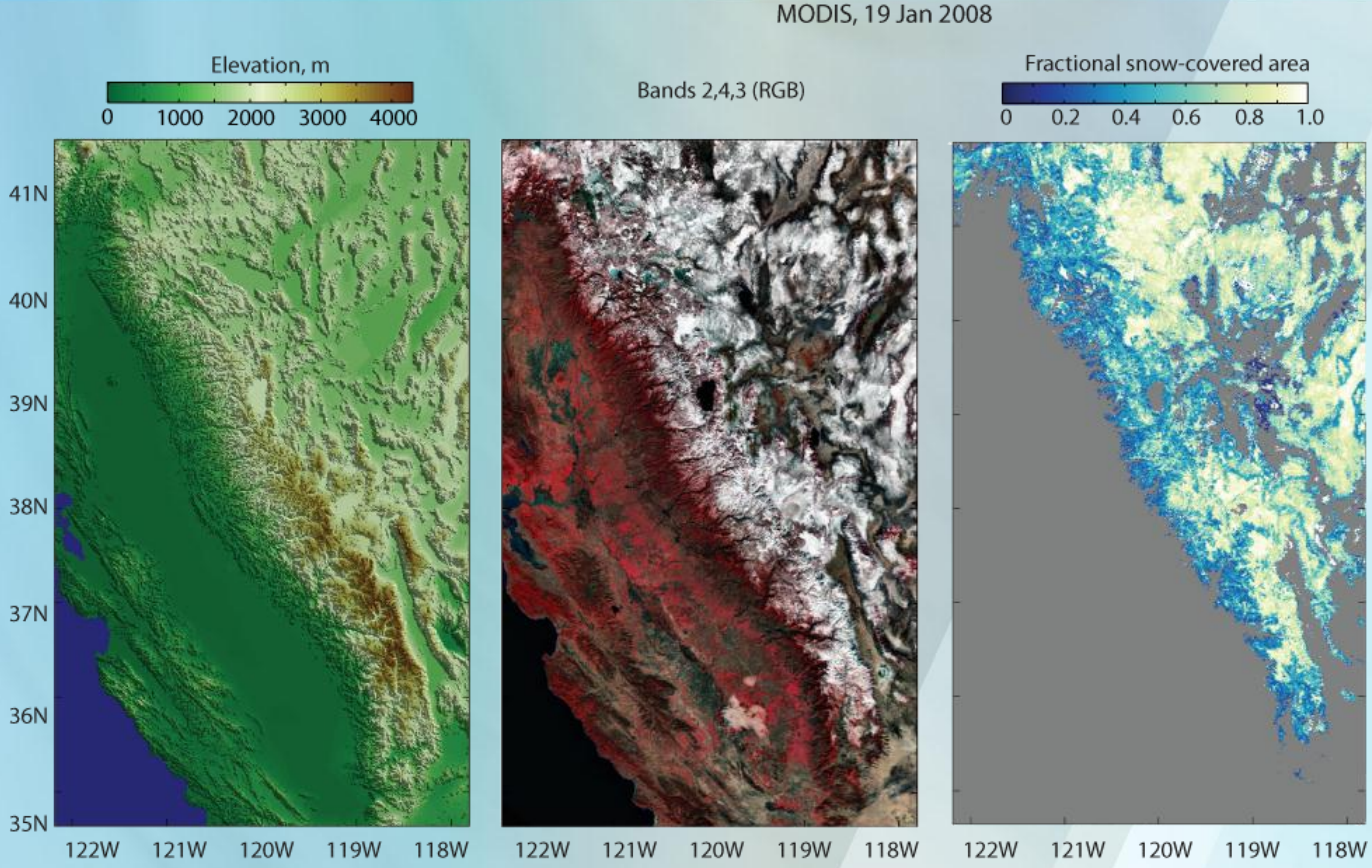
Sometimes done well, by many groups, but could be vastly improved

Slide Courtesy CSIRO, BOM, WMO

We cite papers, but not data



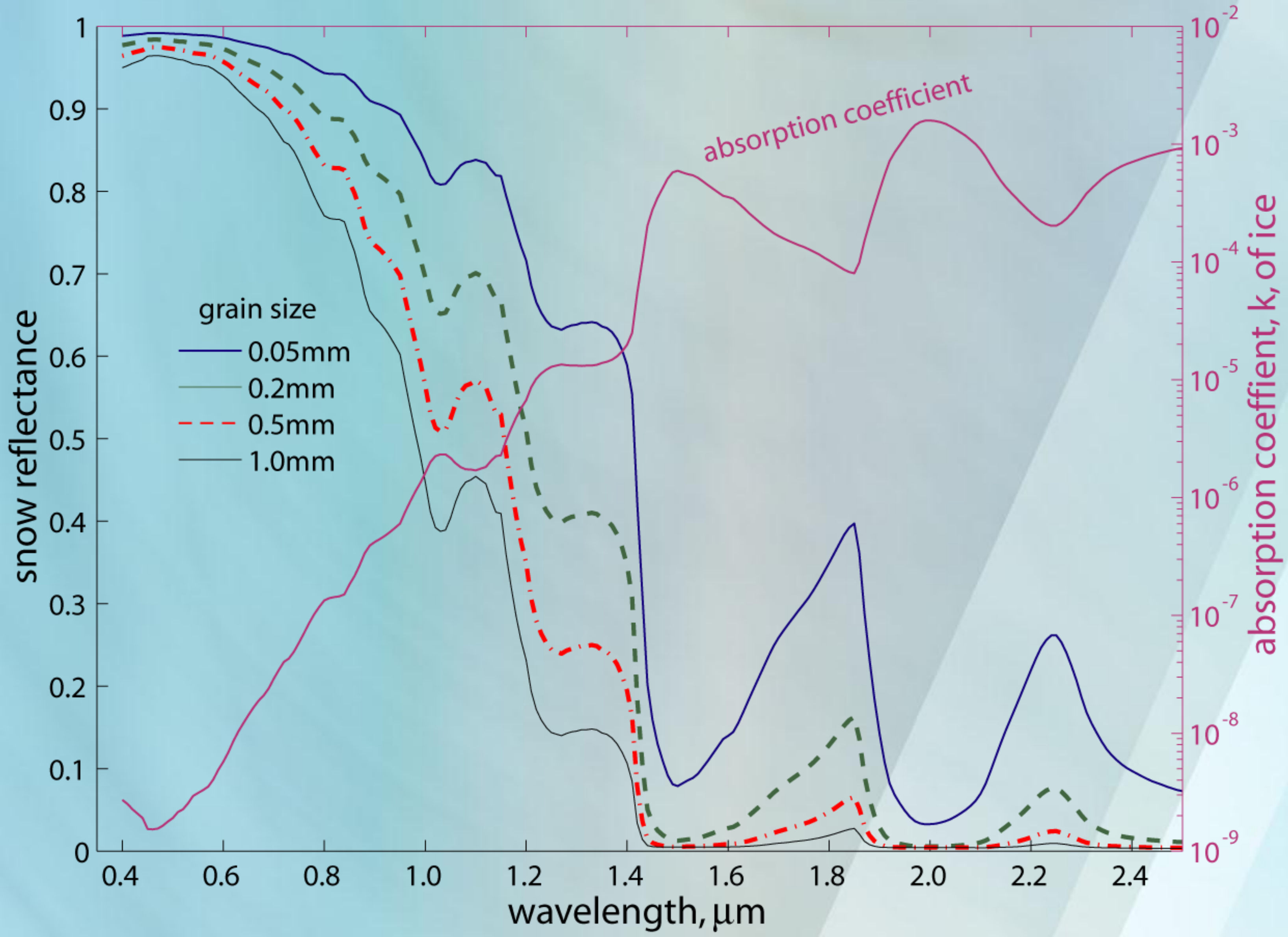
Example data product: fractional snow-covered area, Sierra Nevada



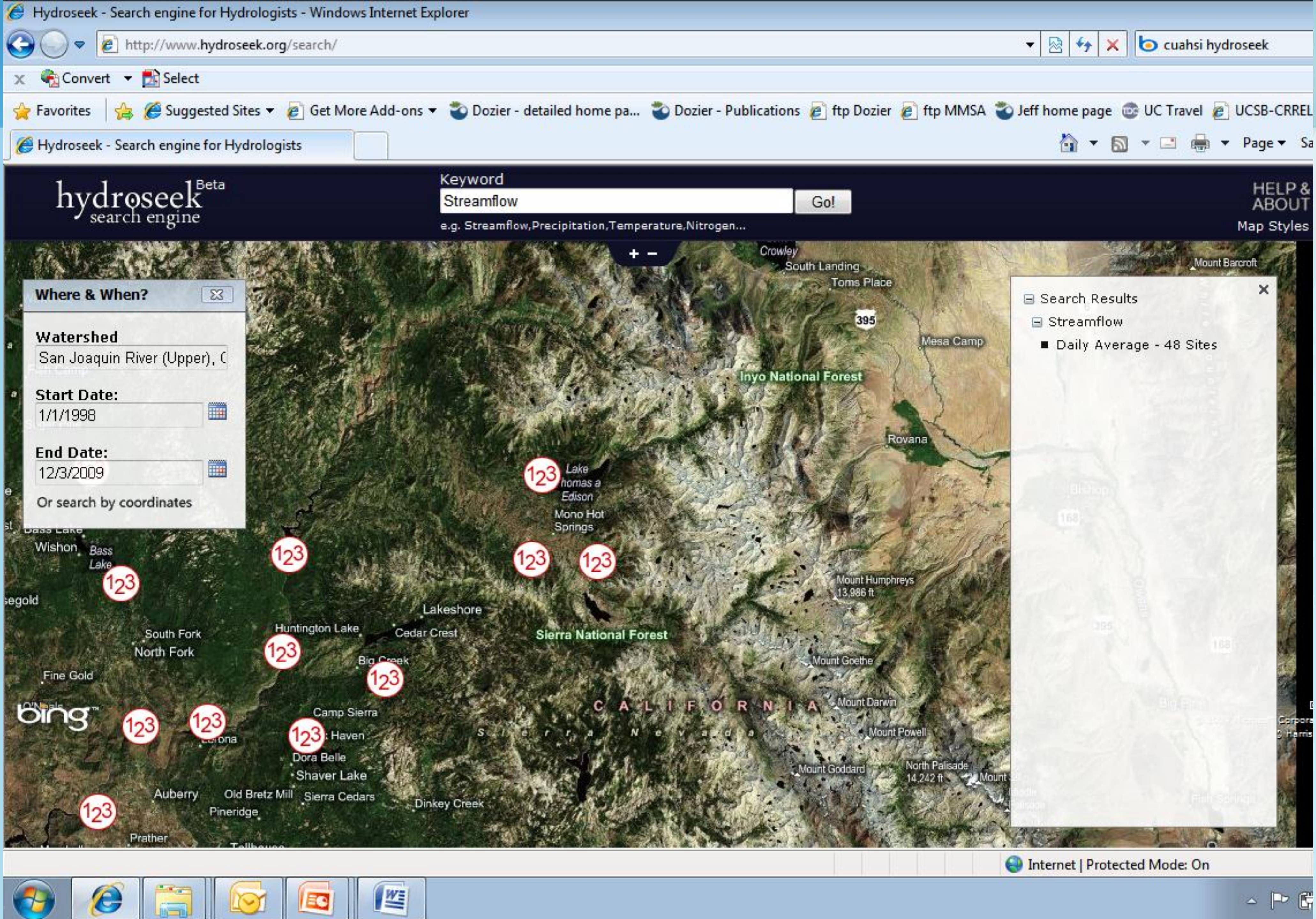
Snow is a collection of scattering grains

- One of my heroes, Moysés Nussenzveig, born São Paulo 1933, prof at Univ. Fed. do Rio de Janeiro. With Wiscombe developed complex angular momentum scattering.

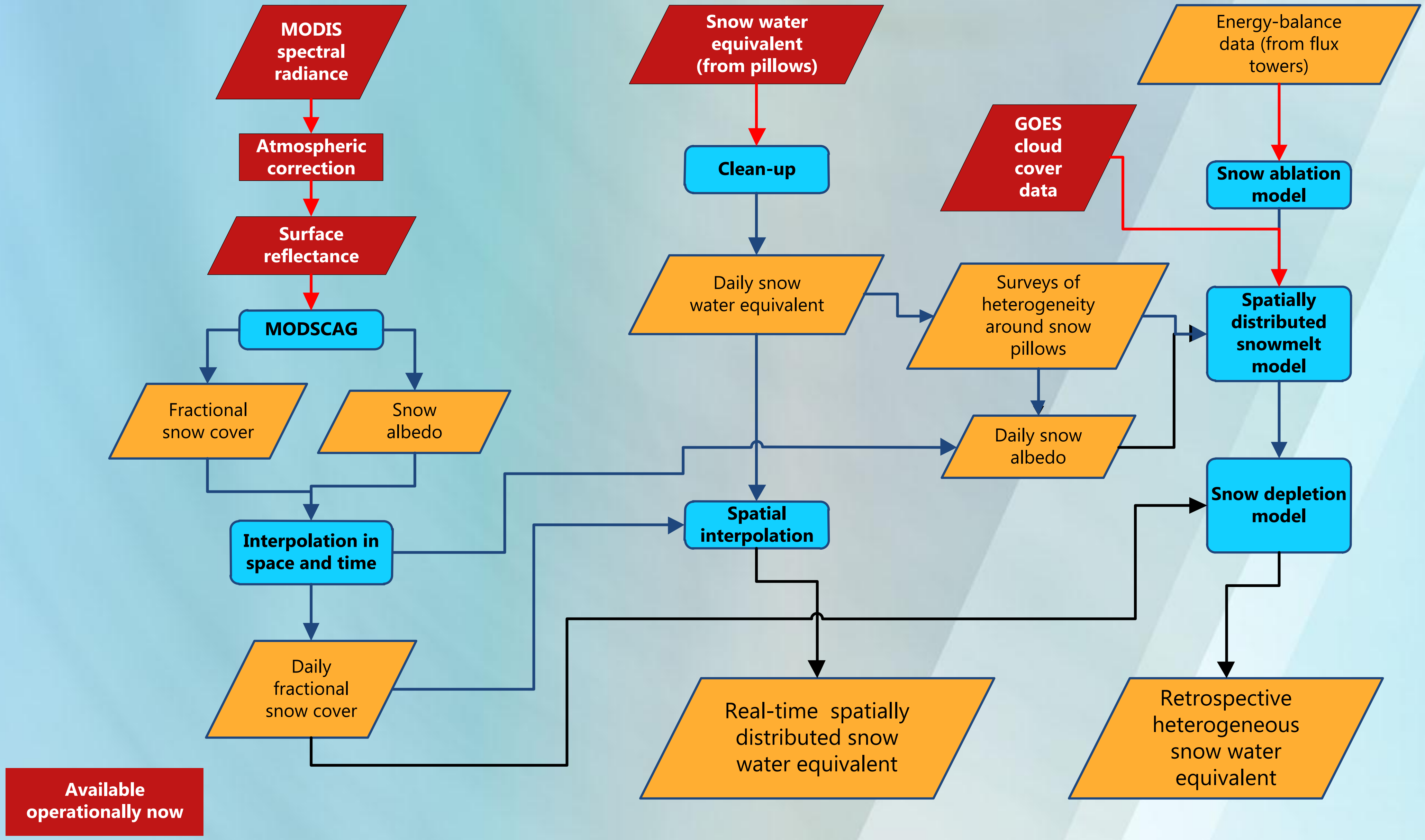
Snow spectral reflectance and absorption coefficient of ice



Measurement of snowpack energy exchange (Sierra Nevada)


Hydroseek - Search engine for Hydrologists - Windows Internet Explorer
http://www.hydroseek.org/search/
cuahsi hydroseek
Convert
Select
Favorites
Suggested Sites
Get More Add-ons
Dozier - detailed home pa...
Dozier - Publications
ftp Dozier
ftp MMSA
Jeff home page
UC Travel
UCSB-CRREL
Hydroseek - Search engine for Hydrologists
Page
hydroseek Beta
search engine
Keyword
Streamflow
Go!
e.g. Streamflow,Precipitation,Temperature,Nitrogen...
HELP & ABOUT
Map Styles
Where & When?
Watershed
San Joaquin River (Upper), C
Start Date:
1/1/1998
End Date:
12/3/2009
Or search by coordinates
Search Results
Streamflow
Daily Average - 48 Sites
Inyo National Forest
Sierra National Forest
CALIFORNIA
Sierra Nevada
Crowley
South Landing
Toms Place
Mesa Camp
Rovana
Mount Barcroft
Mono Hot Springs
Mount Humphreys 13,986 ft
Lakeshore
Huntington Lake
Cedar Crest
Big Creek
South Fork
North Fork
Fine Gold
Wishon
Bass Lake
Mount Goethe
Mount Darwin
Mount Powell
Mount Goddard
North Palisade 14,242 ft
Camp Sierra
Dora Belle
Shaver Lake
Auberry
Old Bretz Mill
Sierra Cedars
Pineridge
Dinkey Creek
Prather
bing
Internet | Protected Mode: On

Information products, hydrologic example



What to do about climate change?

There are only 3 options

- Mitigate – to avoid the unmanageable
  - reduce pace & magnitude of changes in global climate being caused by human activities
- Adapt – to manage the unavoidable
  - reduce adverse impacts to humans & ecosystems that result from climate change and related changes
- Suffer the consequences
  - from adverse impacts that are neither mitigated nor adapted to

from John Holdren, US Presidential Science Advisor

Real uncertainties in climate science

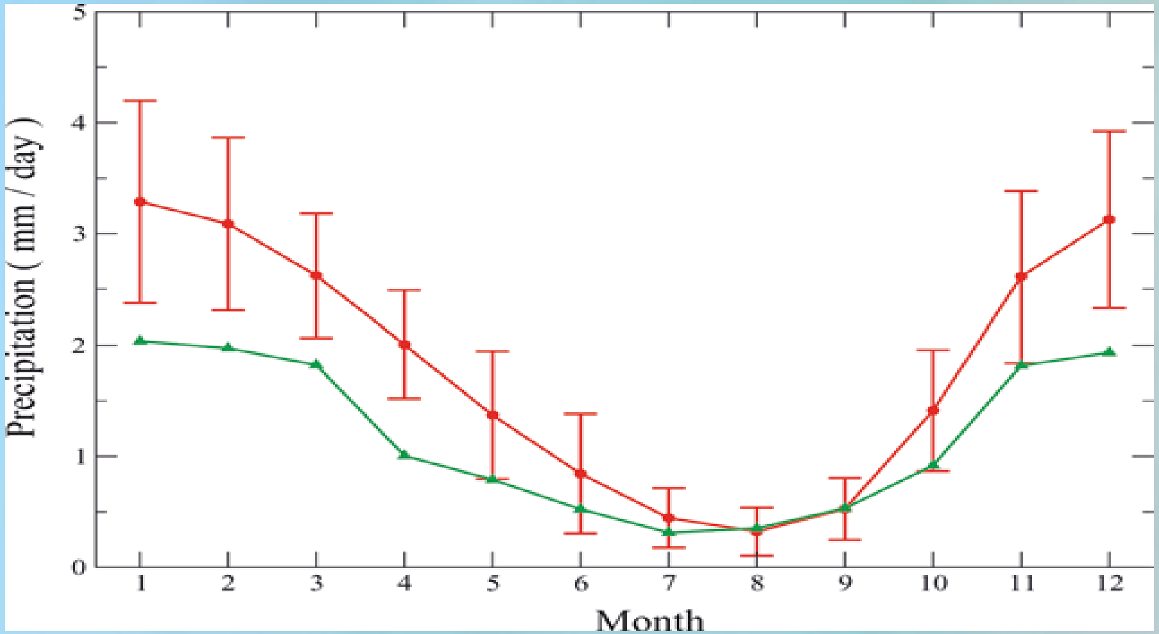
- Regional climate prediction
  - Adaptation is local
  - Problems w downscaling, especially in mountains
- **Precipitation**
  - Especially winter precipitation
- Response of vegetation
  - Photosynthesis in a warmer climate
  - Fires because plants are drier with more ET
- Ice sheet behavior

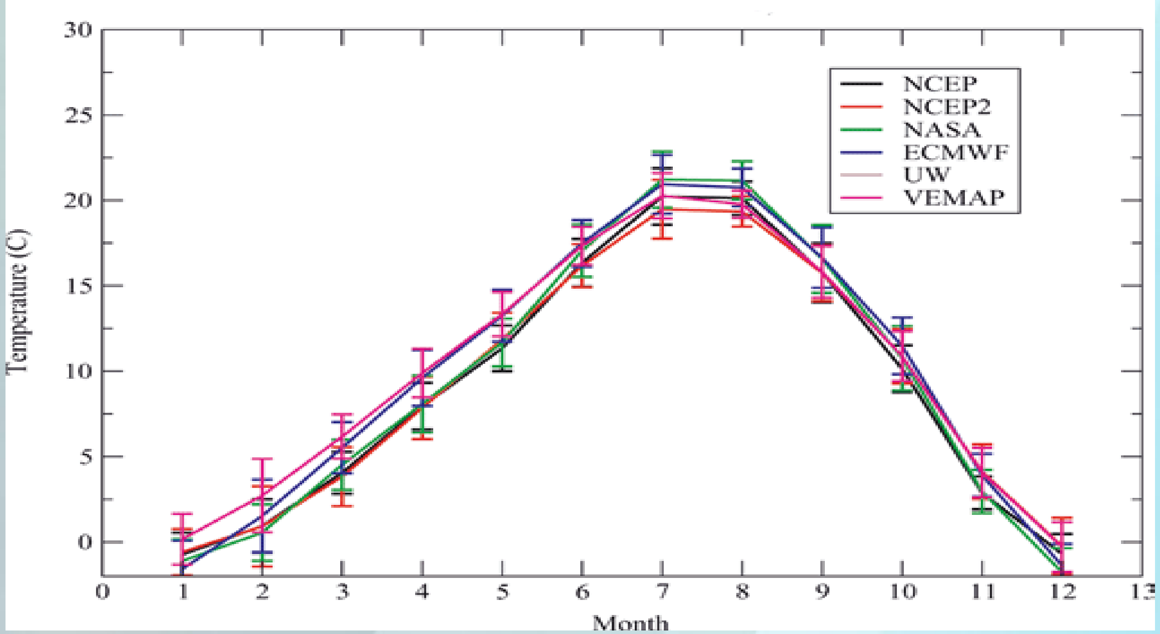
"This climate of suspicion we're working in is insane. It's drowning our ability to soberly communicate gaps in our science."

- Gavin Schmidt

Schiermeier, 2010, *Nature*

Regional models:
better results for temperature than for precipitation





Precipitation: mean of 15 models (red) vs observations (green)

Temperature: mean of 15 models (red) vs observations and reanalyses

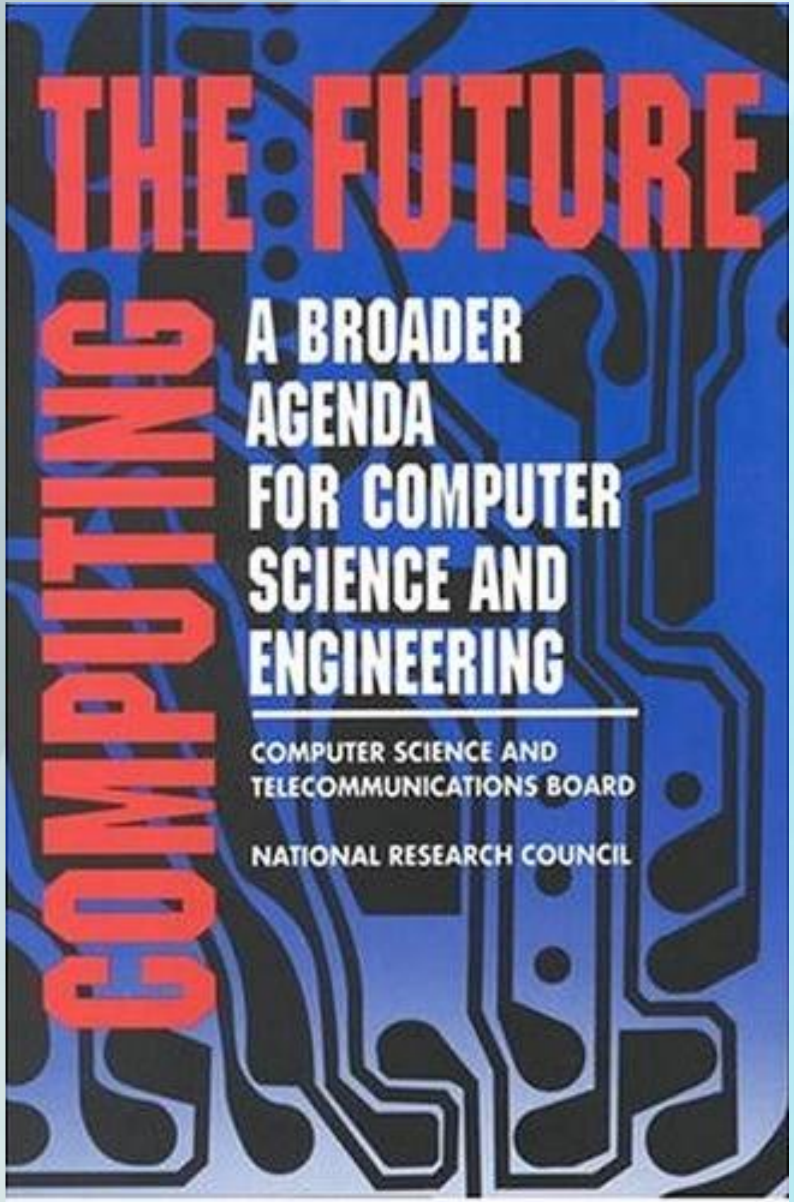
Vertical bars are ±1 standard deviation of model monthly results

Coquard et al., 2004, *Climate Dynamics*


2002-01-31

Computing the Future (1992, but still a good read)

- "Broaden the field . . . many intellectual opportunities at the intersection of CS&E and other problem domains"
- Predictions about the data deluge have come to pass
- Data collection often separated from their analyses, so infrastructure needed
- Science is increasingly collaborative, hence the need to support integration of disparate, distributed data with disparate models, among collaborators who are not co-located



And remember Gray's Laws

(Szalay & Blakeley, in *The Fourth Paradigm*, 2009)

- Scientific computing is becoming increasingly data intensive
- The solution is in a "scale-out" architecture
  - Bricks with local CPU and part of the data
    - Queries across heavily partitioned data not easy
- Bring computations to the data, rather than data to the computations
  - You need to enable the user to design and run operations at the data
- Start the design with the "20 queries"
  - If you can deal with the 20 most important, you can probably deal with the next 200
- Go from "working to working [a little better]"
  - No giant leaps without intervening milestones

The data cycle perspective, from creation to curation

- **The science information user:**
  - I want reliable, timely, usable science information products
- **The funding agencies and the science community:**
  - We want data from a network of authors
- **The science information author:**
  - I want to help users (and build my citation index)