Open and Free Community Resources for Environmental Research

Lower Mekong Initiative Workshop

August 18-22, 2014; Hanoi, Vietnam

Don Middleton, Chief Technologist for Data and eScience
U.S. National Center for Atmospheric Research (NCAR)
Computational & Information Systems Laboratory
Technology Development Division (CISL/TDD)
Boulder, Colorado, USA

On behalf of many people and projects: NCAR Science Gateways, GIS Program, Climate Data Guide, CESM, WRF, DART, NCL/PyNGL, Vapor; UCAR's Unidata Program for IDV and the LDD/IDM; the global Earth System Grid Federation (ESGF); and the Globus Project

The National Center for Atmospheric Research

Managed by the University Corporation for Atmospheric Research

A National Science Foundation FFRDC

Mesa Campus; Boulder, Colorado, USA

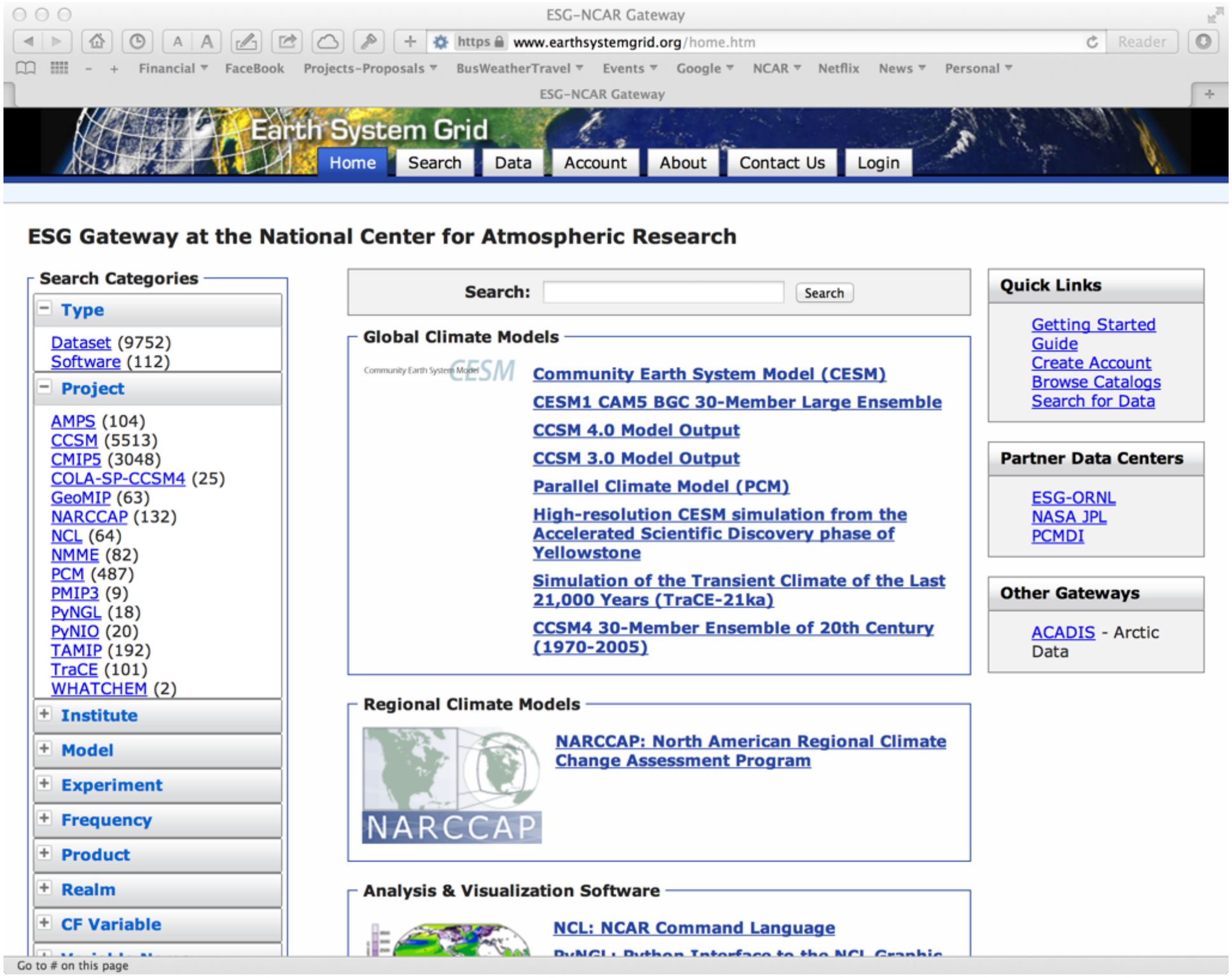
NCAR

Topics

- Data Resources
- Models and Data Assimilation
- Data Analysis, Visualization, and Transfer Tools
- NSF's EarthCube Program

Data Resources

NCAR



www.earthsystemgrid.org

NCAR

A Few ESG Metrics

- Total registered users: ~40,000
- Average monthly registrations: 569
- Avg. monthly download volume: 110TB
- Avg. monthly download users: 1,190
- Est. 5 petabytes downloaded over time

NCAR's ESG Gateway feeds into the ESGF

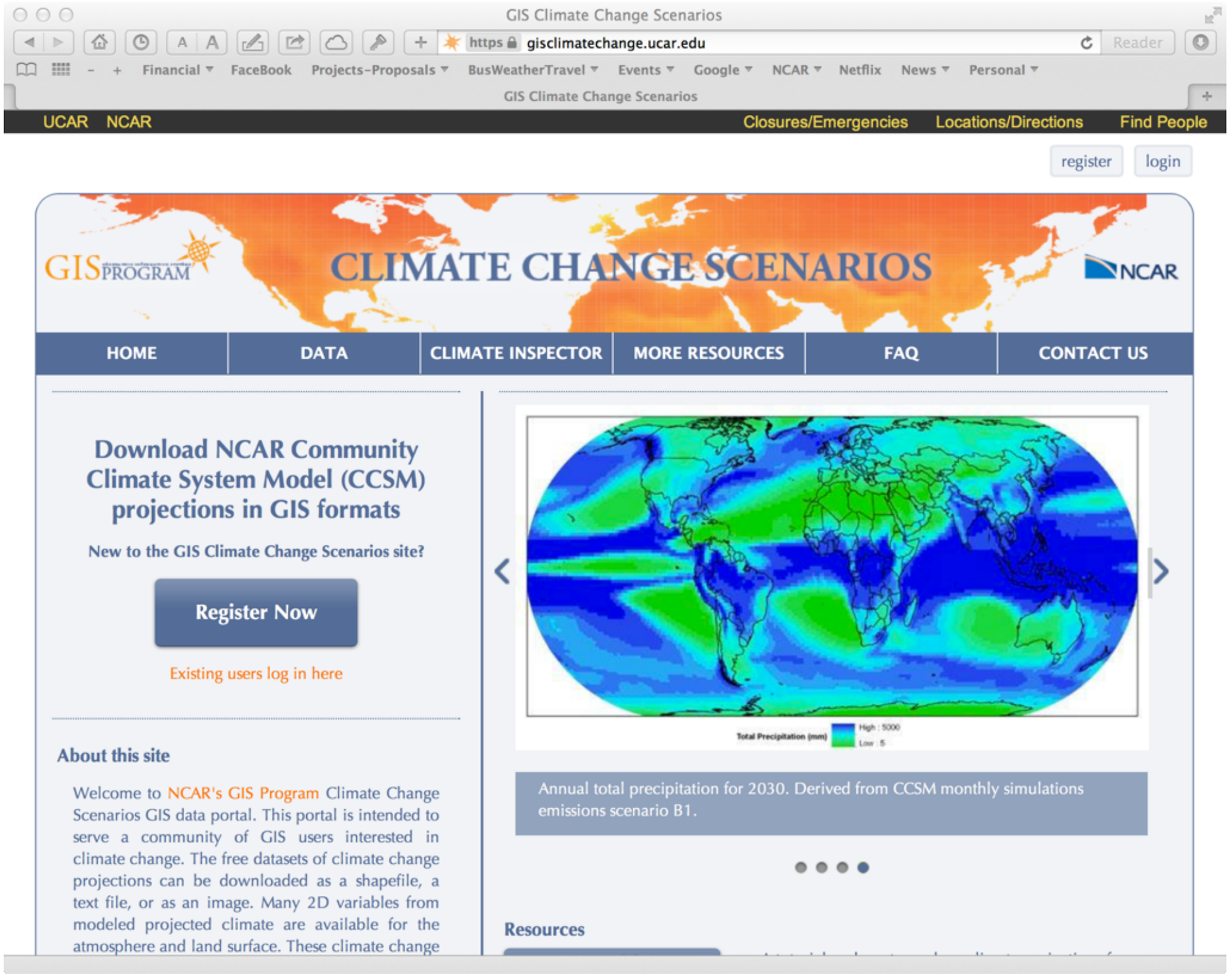
http://pcmdi9.llnl.gov

NCAR

NSF

Research Data Archive (RDA) at NCAR

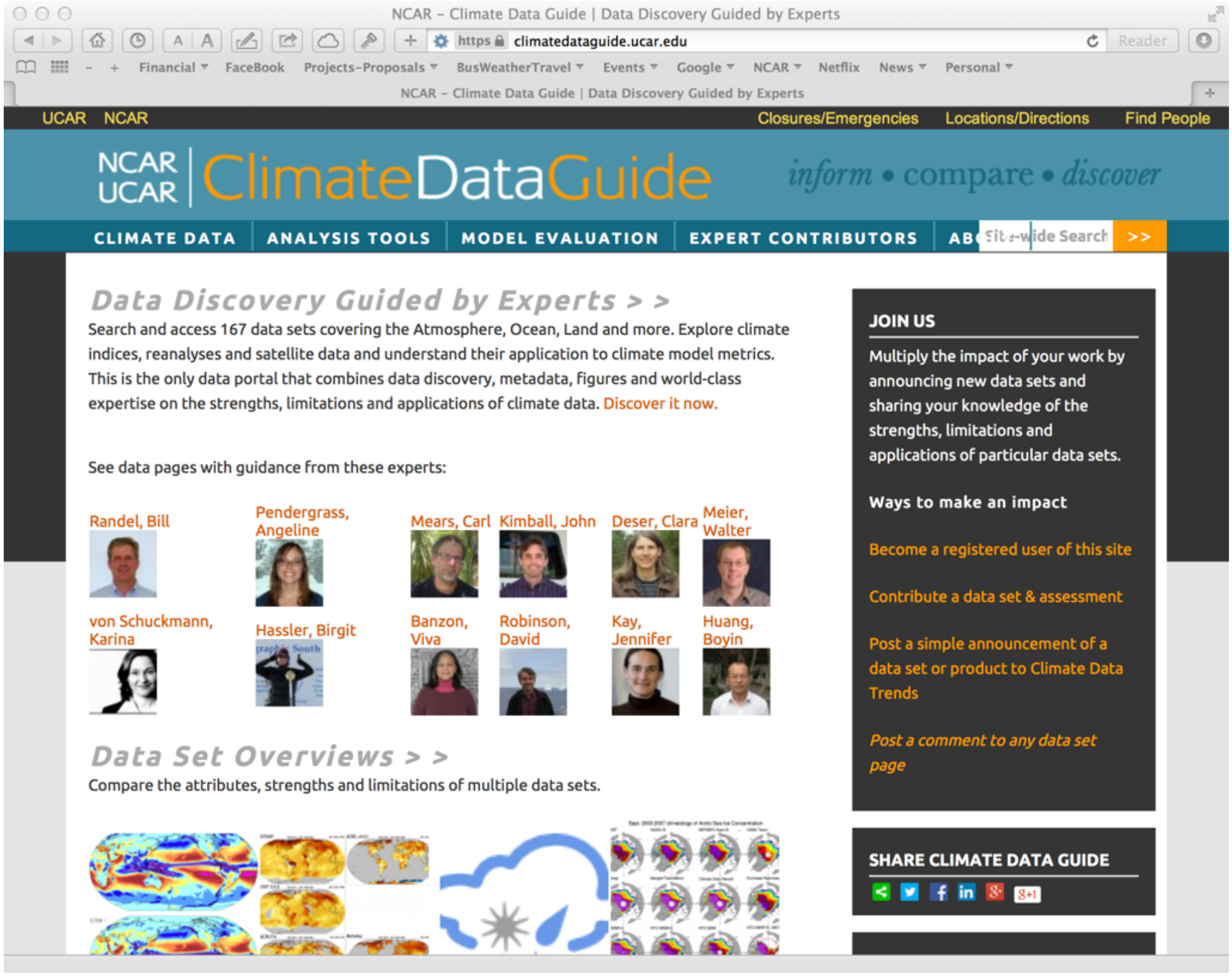
- Purpose: Support climate & weather research
- Collections: Ocean & atmosphere observations, analyses, reanalyses, operational NWP outputs
- Basic Metrics
  - Established in 1960s
  - 600+ datasets, 8M files, 1.8 PB
  - +70 datasets growing daily to monthly
- Science educated staff
  - Expert consultants and data engineers
- Free and open access
- Web Address: http://rda.ucar.edu/



https://gisclimatechange.ucar.edu

NCAR

NSF



https://climatedataguide.ucar.edu

NCAR

NSF

Data Citation, and Related

- NCAR/UCAR have a Data Citation Initiative, led by the NCAR Library (with all major data providers)
- It's not just for data
- We want to form an interlinked web of citable data collections, tools, models, visualizations, and scholarly publications
- We're using DOI's via EZID/Datacite

Examples

- Mearns, L.O., et al., 2007, updated 2012. *The North American Regional Climate Change Assessment Program dataset*, National Center for Atmospheric Research Earth System Grid data portal, Boulder, CO. Data downloaded 2013-08-12. [doi:10.5065/D6RN35ST]

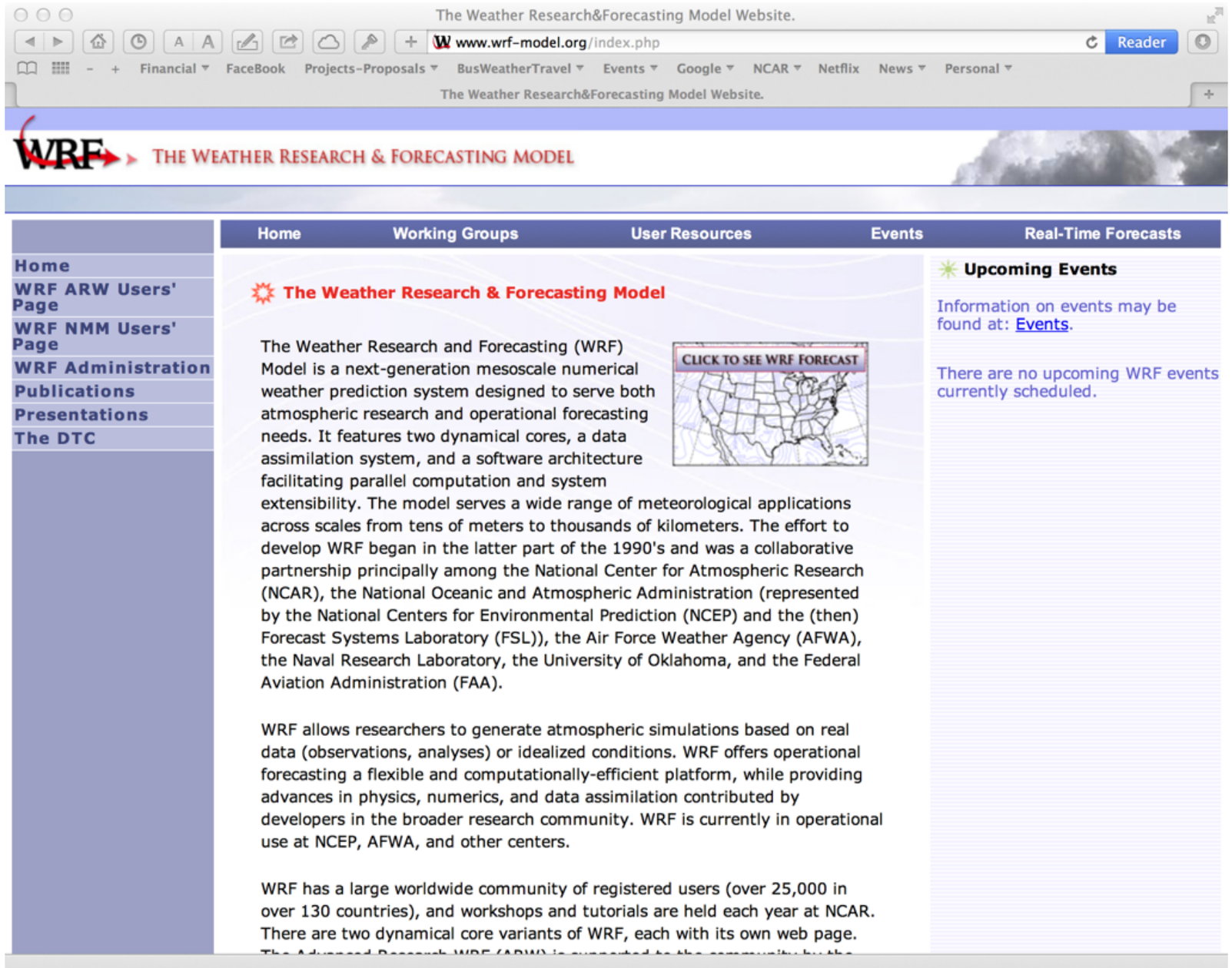
- Pelto, Mauri S. (2013). Juneau Icefield Glacier Mass Balance. UCAR/NCAR – CISL – ACADIS. http://dx.doi.org/10.5065/D6NZ85N3

- The NCAR Command Language (Version 6.1.1) [Software]. (2013). Boulder, Colorado: UCAR/NCAR/CISL/VETS. http://dx.doi.org/10.5065/D6WD3XH5

Models and Data Assimilation

NCAR

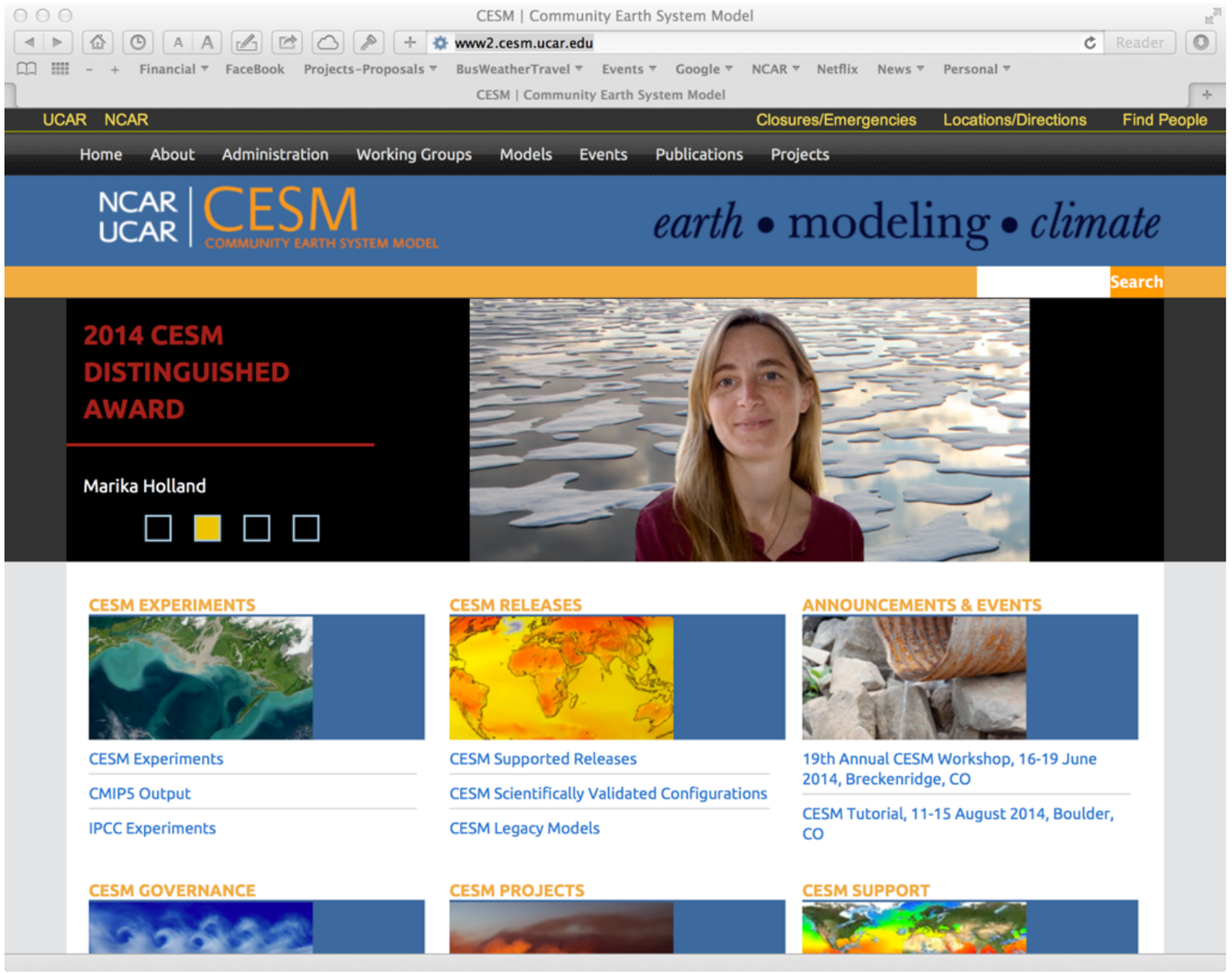
WRF: The Weather Research and Forecast Model



http://www.wrf-model.org

NCAR

CESM: The Community Earth System Model



http://www2.cesm.ucar.edu

NCAR

UCAR The Data Assimilation Research Testbed -- DART

www.image.ucar.edu/DAReS/DART/

Reader

Financial FaceBook Projects-Proposals BusWeatherTravel Events Google NCAR Netflix News Personal

UCAR The Data Assimilation Research Testbed -- DART

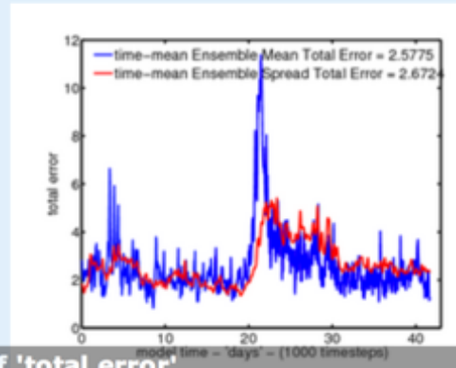
UCAR NCAR Find People Contact/Visit

CISL IMAGe DAReS DAReS Search advanced

NCAR IMAGe: Data Assimilation Research Section

DART Why DART? Research Getting Started Documentation Diagnostics Miscellany

DART board



time series of 'total error'
Quantify the effects of observations or parameter settings.

Welcome to the Data Assimilation Research Testbed - DART

DART is a community facility for ensemble DA developed and maintained by the Data Assimilation Research Section (DAReS) at the National Center for Atmospheric Research (NCAR). DART provides modelers, observational scientists, and geophysicists with powerful, flexible DA tools that are easy to implement and use and can be customized to support efficient operational DA applications. DART is a software environment that makes it easy to explore a variety of data assimiliation methods and observations with different numerical models and is designed to facilitate the combination of assimilation algorithms, models, and **real** (as well as synthetic) observations to allow increased understanding of all three. DART includes extensive documentation, a comprehensive tutorial, and a variety of models and observation sets that can be used to introduce new users or graduate students to ensemble DA. DART also provides a framework for developing, testing, and distributing advances in ensemble DA to a broad community of users by removing the implementation-specific peculiarities of one-off DA systems.

http://www.image.ucar.edu/DAReS/DART

NCAR

NSF

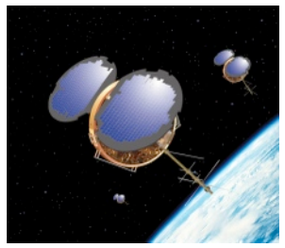
Some Current **DART** Research Projects

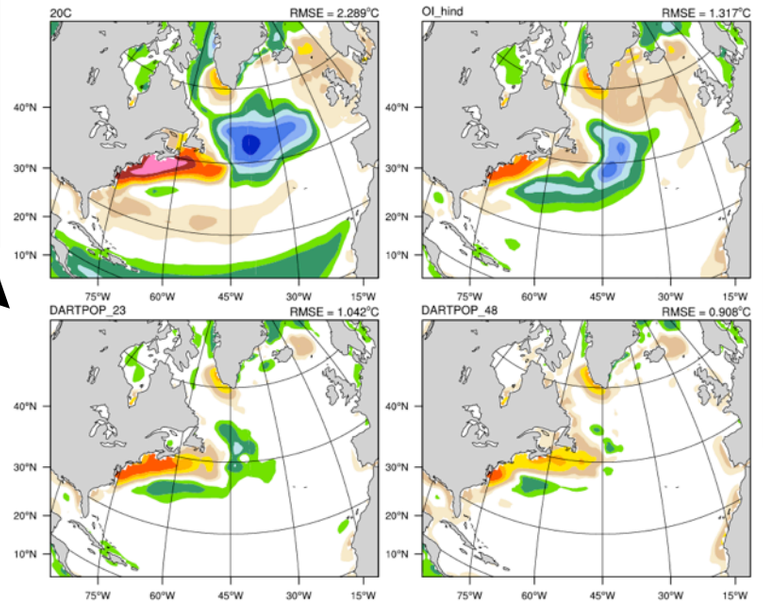
DART/CAM
Global DA

Ensemble of
Boundary Conditions

DART/POP
Coupled Ocean
DA

Observations

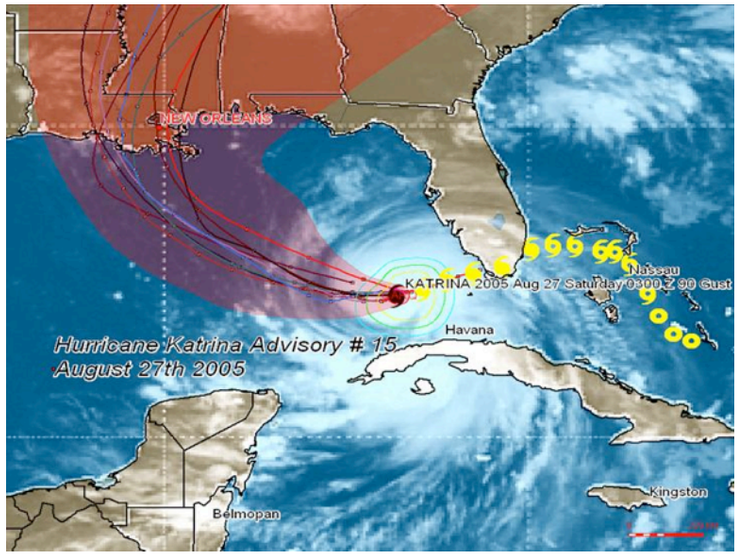




Ensemble Ocean Reanalysis
IPCC Decadal Forecast ICs

Ensemble Reanalysis Datasets
(1998-2010 available)

DART/WRF
Regional DA

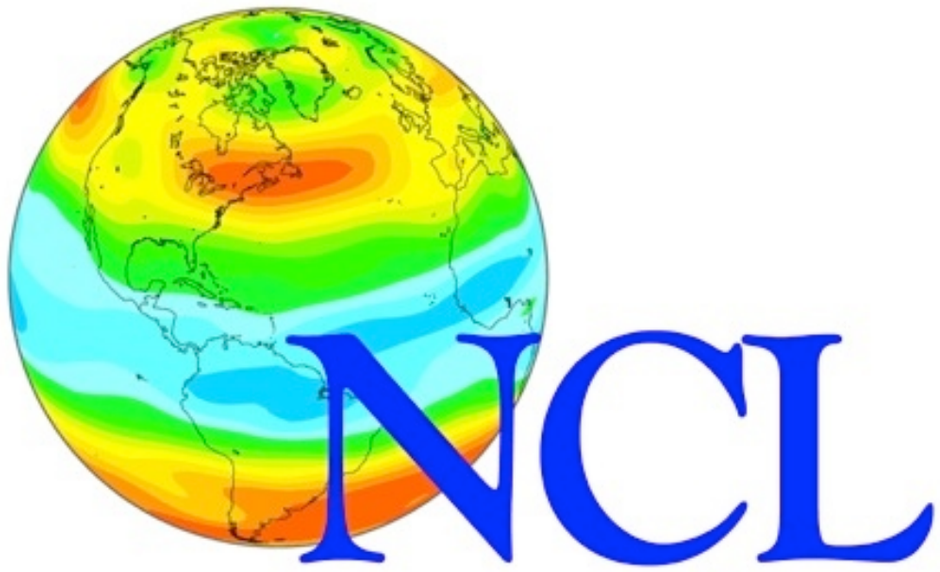


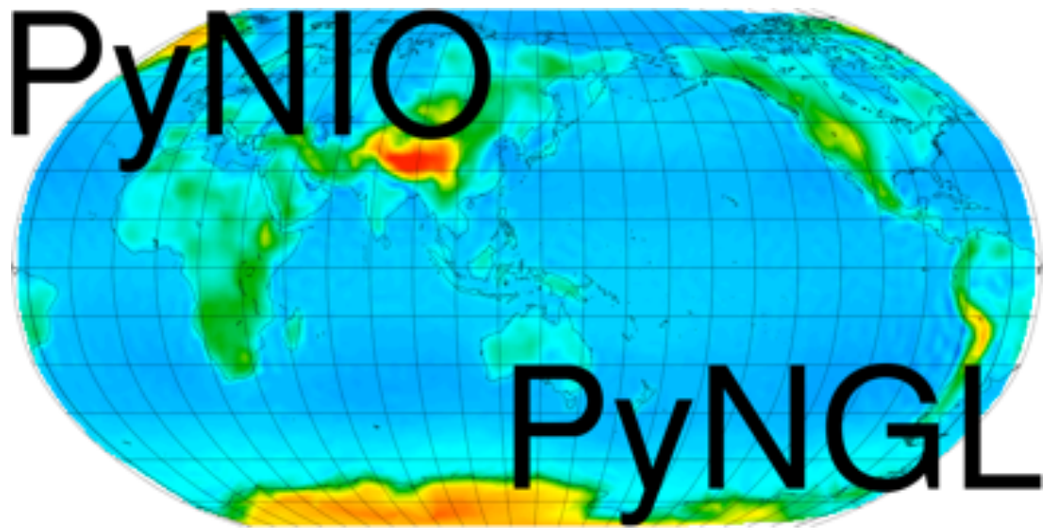
Hurricane and Severe
Weather Prediction

Data Analysis, Visualization, and Transfer Tools

NCL & Python material courtesy of Mary Haley, NCAR/CISL/TDD

Community Tools for the Analysis and Visualization of Geoscientific Data





CISL

NESL

NSF

NCAR

The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Dennis Shea
Science guy
Data expert
Trainer

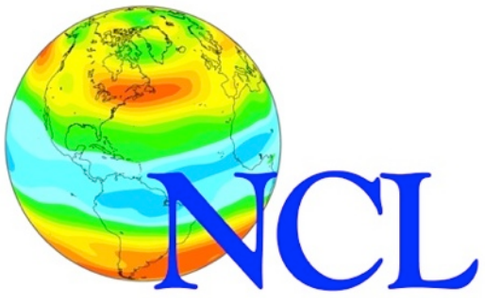
Dave Brown
NCL Tech Lead

Mary Haley
Product owner
Trainer

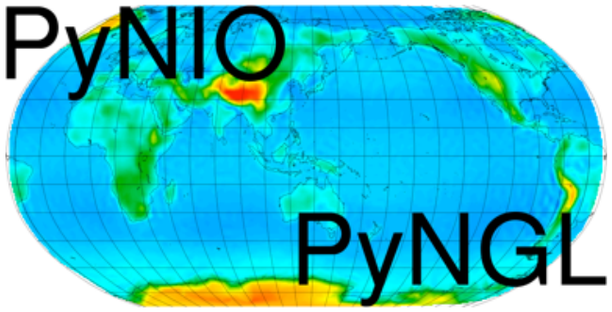
Rick Brownrigg
Developer
Research

Wei Huang
Developer
Data Formats

Adam Phillips
Science guy
Graphical expert



What we do



- Develop geoscientific analysis and visualization software in close collaboration with NCAR scientists
- Document software and provide extensive examples
- Answer user questions on a daily (hourly!) basis
- Offer hands-on training workshops 4-6x a year
- *Collaborate with researchers world-wide to enhance software*

NCAR Command Language (NCL)

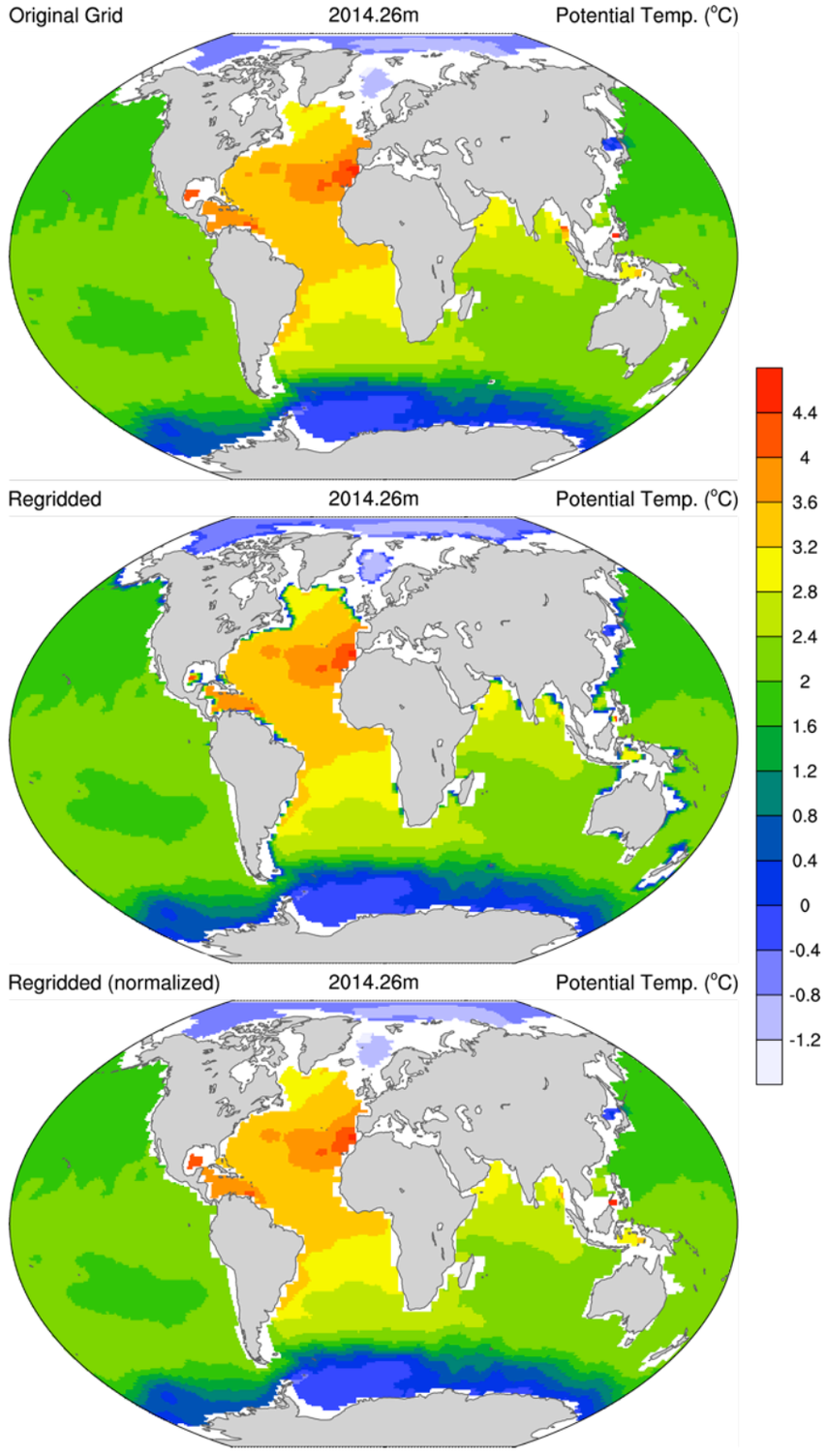
A scripting language tailored for the analysis and visualization of geoscientific data

- Simple, robust file input and output
- Hundreds of analysis (computational) functions
- Visualizations (2D) are publication quality and highly customizable

- Users range from graduate students working on doctoral theses to programmers in scientific organizations working on large research projects
- UNIX binaries & source available, free
- Extensive website, **regular workshops**

http://www.ncl.ucar.edu/

http://dx.doi.org/10.5065/D6WD3XH5

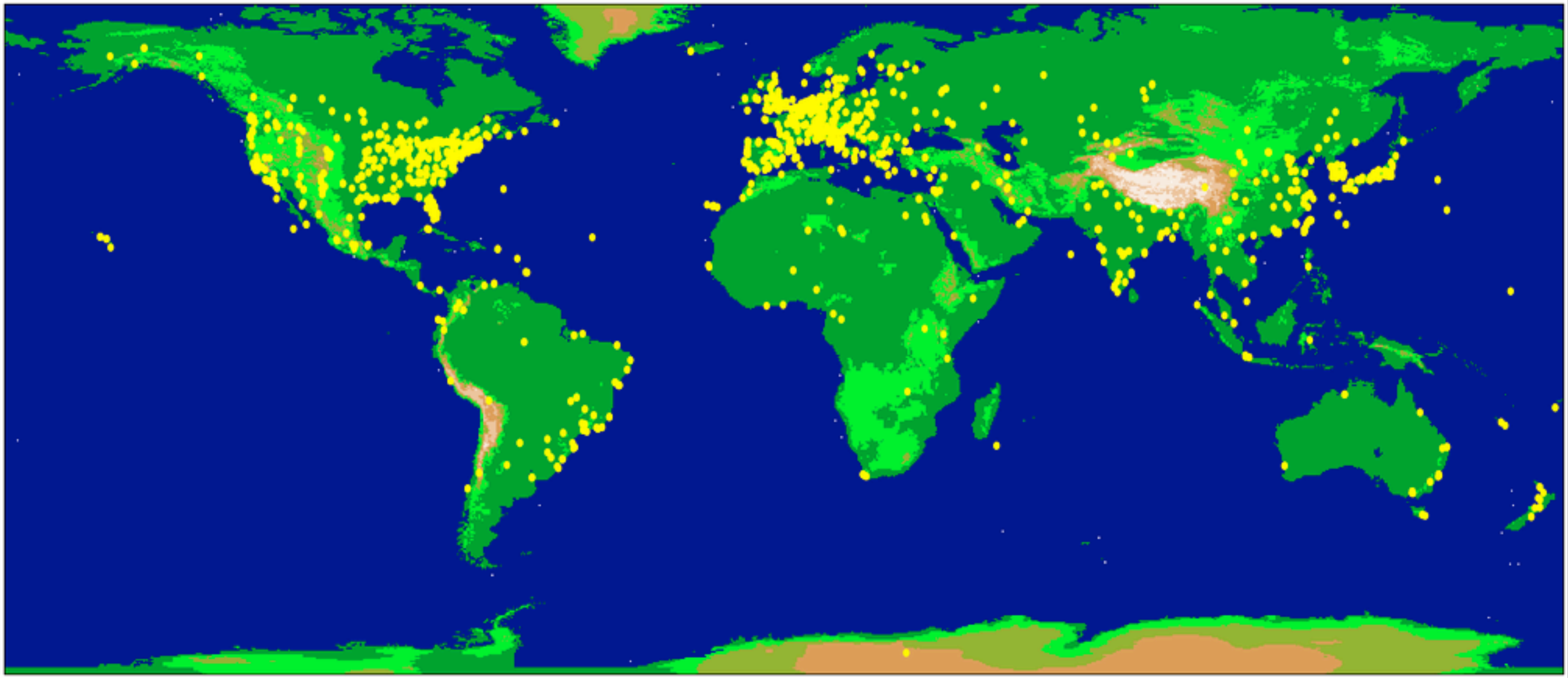


Metrics and scope of software usage

– 20,000+ registered users in 127 countries

– Average of 1700 downloads a month

– Average of 300 emails a month on email list (includes responses)



NCL's File Input and Output

- Handles many geoscientific data formats:
  - NetCDF (versions 3 and 4)
  - GRIB (versions 1 and 2)
  - HDF4, HDF5, HDF-EOS (versions 4 and 5)
  - ESRI Shapefiles
- OPeNDAP-enabled
- ASCII (text files, CSV files)
- Fortran and C binary

NCL's computational analyses

- Array-based math (no need to loop across dimensions) – similar to Matlab, Python, IDL

- Hundreds of computational functions
  - Spherical harmonics, empirical orthogonal functions (EOFs)
  - Scalar and vector regridding
  - Interpolation
  - Climatologies
  - Functions specific to CESM and WRF models

- Many tailored to climate and weather

- Some are highly-specialized

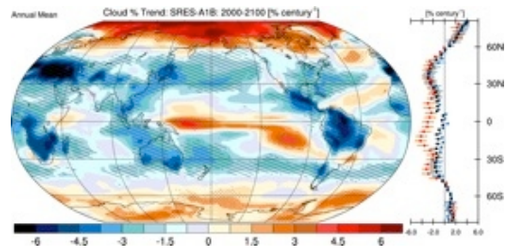
- Most handle missing data

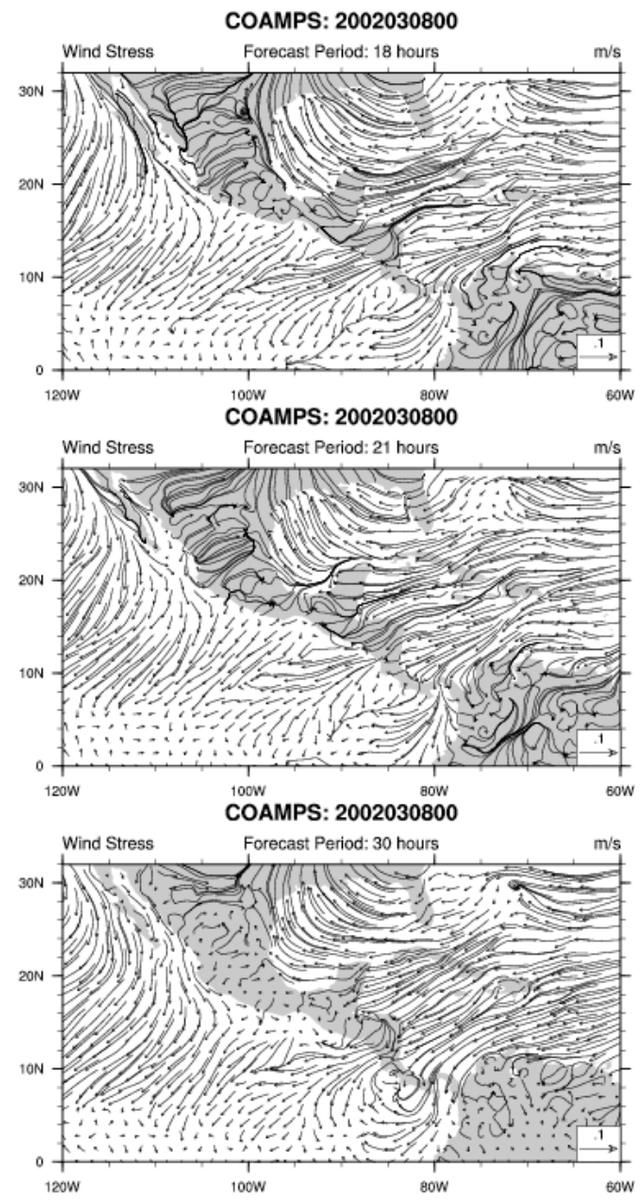
- Can call C and Fortran routines from NCL and PyNGL

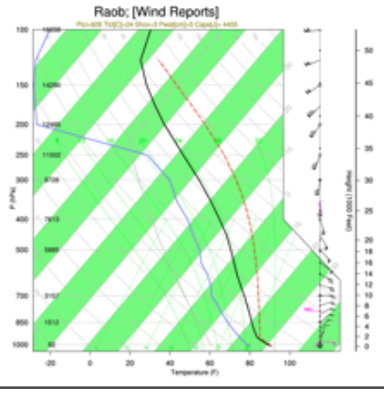
- Many contributions from users, they drive priorities

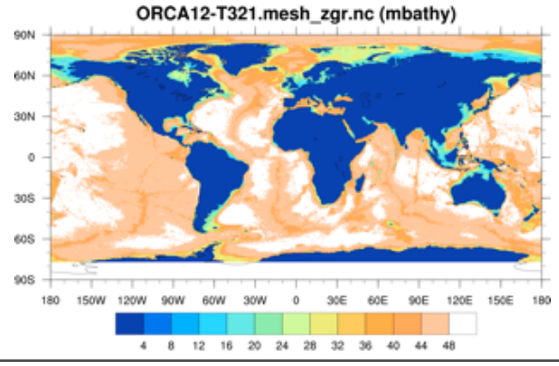
NCL Visualizations

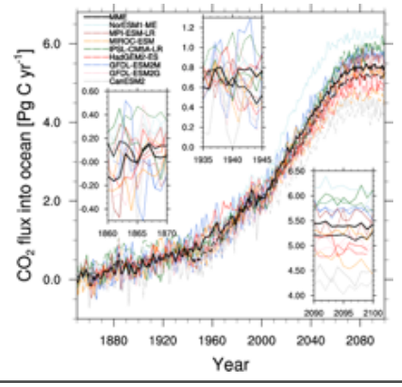
- High-quality and customizable visualizations
- Contours, XY, vectors, streamlines
- Maps with common map projections
- Handles data on regular and irregular grids, triangular meshes
- Specialized scripts for climate diagnostics, skew-T, wind roses, histograms, Taylor diagrams, panels, bar charts
- High-level interfaces: simplifies graphics code
- Over 1,400 visualization “options”

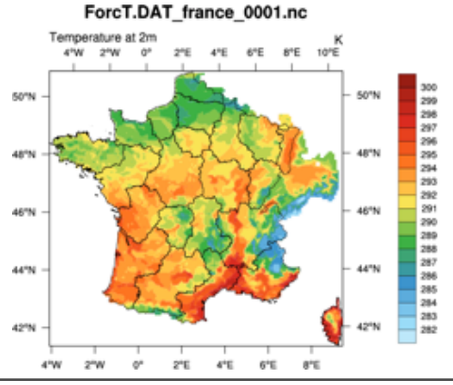






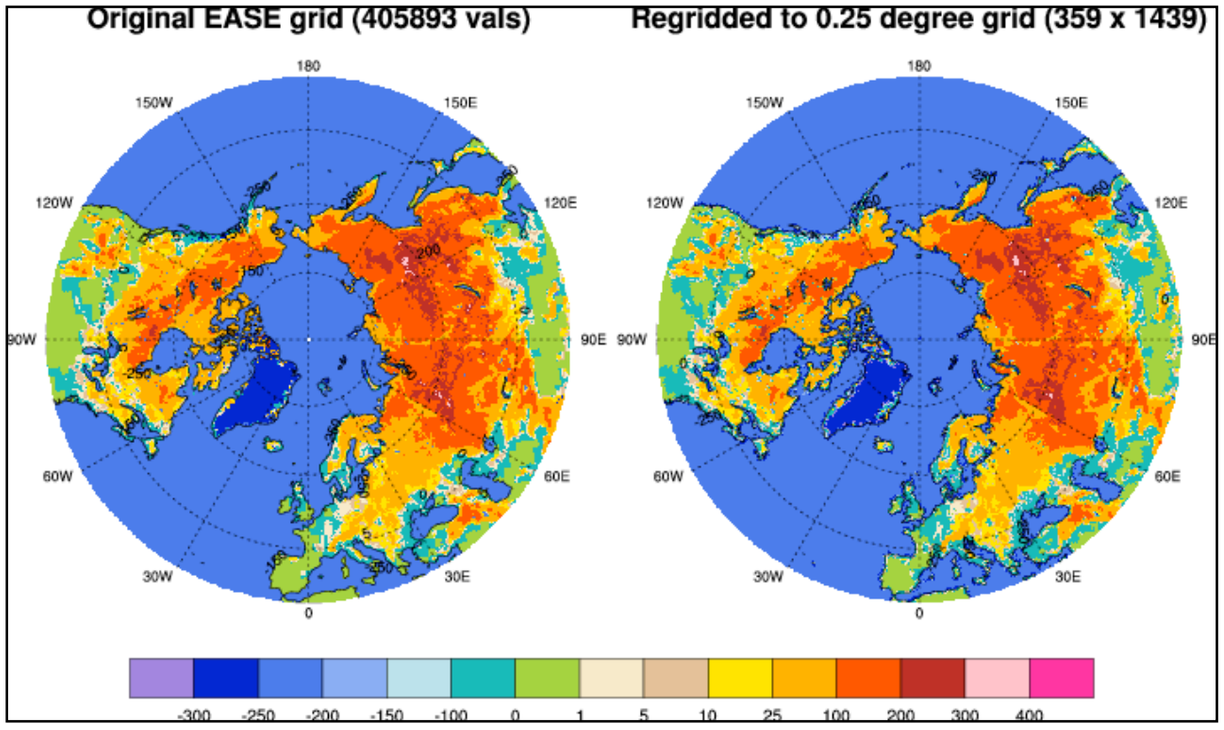


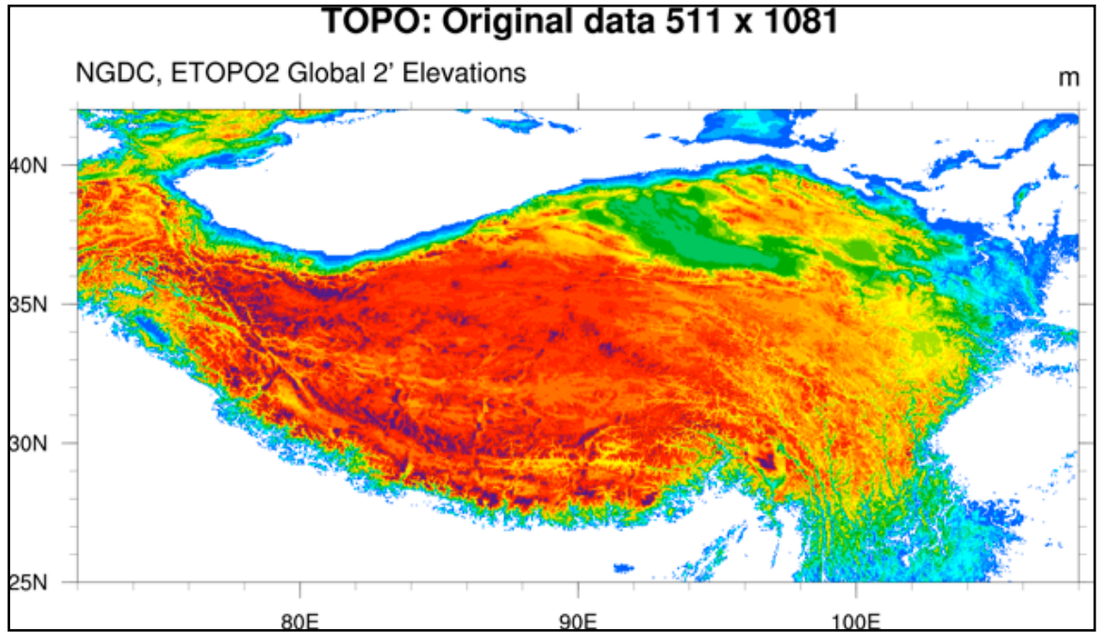


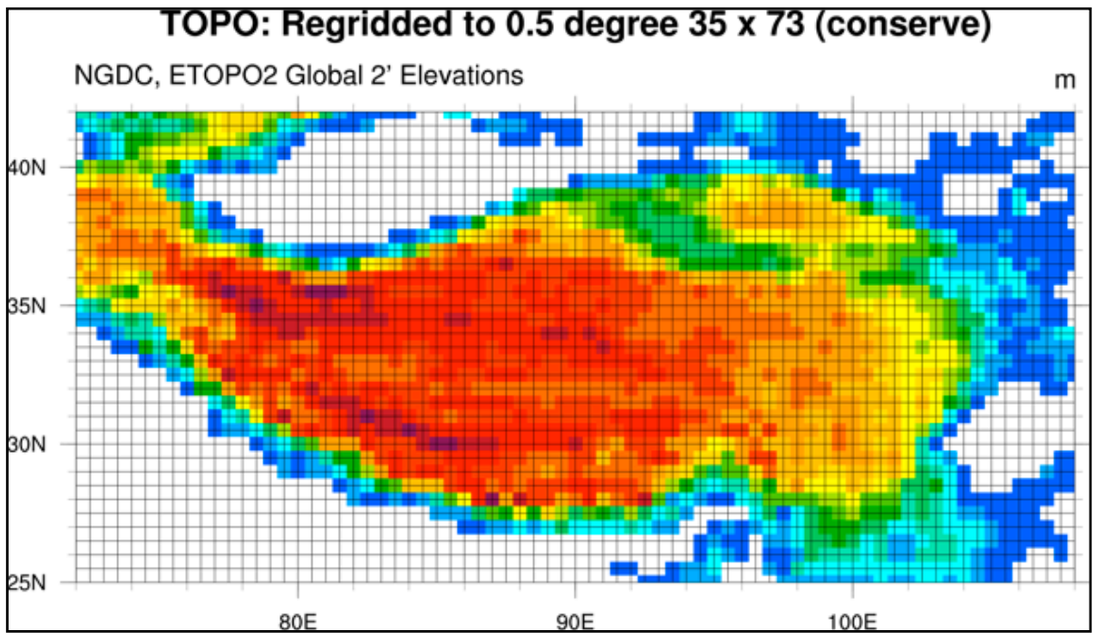


ESMF regridding software built into NCL

Mohammad Abouali, SIParCS intern
Dave Brown & Mary Haley, mentors
Robert Oehmke, collaborator
ESMF Group, NOAA/ESRL/NESII







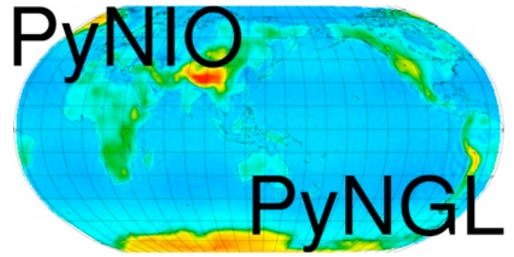
Regridding is the process of interpolating data from one grid (rectilinear, curvilinear, unstructured) to another while preserving the qualities of the original data.

ESMF regridding has multiple interpolation methods; a crucial feature for climate model intercomparison.

http://www.ncl.ucar.edu/Applications/ESMF.shtml

PyNIO and PyNGL

Python modules based on a subset of NCL's capabilities



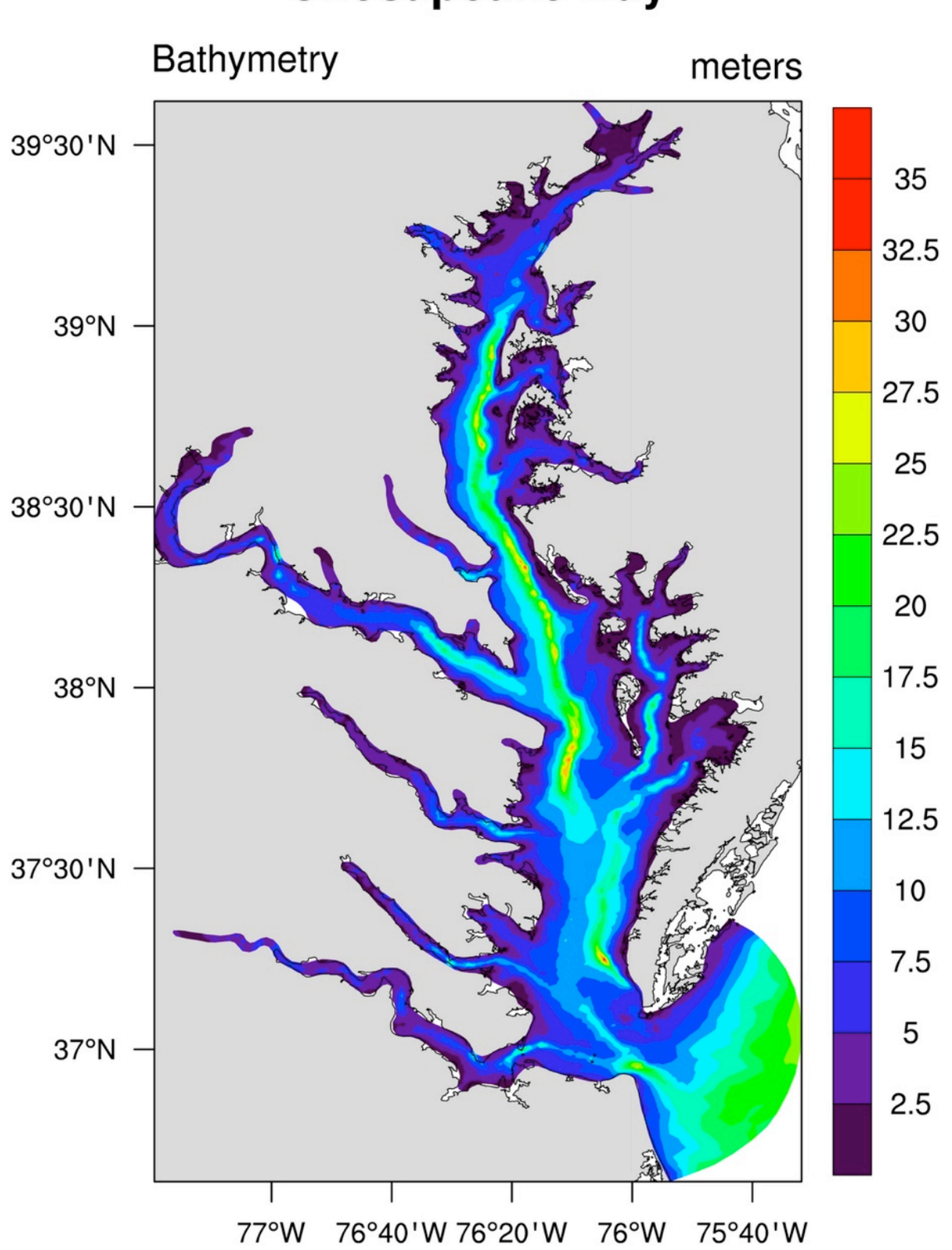
PyNIO

- Reads and writes same data formats as NCL
- Has special syntax for complex data subsetting (not available in NCL)

PyNGL

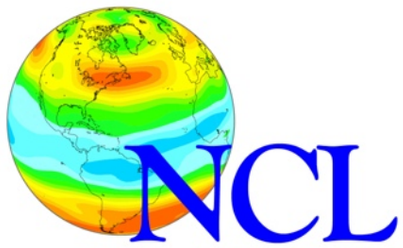
- Same publication-quality graphics as NCL
- Utilizes existing Python modules and development tools (NumPy, swig, f2py, ScientificPython)
- Some climate-specific computations

http://www.pyngl.ucar.edu/



pNetCDF

MOAB



agoda

swift

parVis

3-year joint DOE project with Argonne, Sandia, and Pacific Northwest National labs and UC-Davis to develop parallel analysis tools and new visualization techniques for ultra-large climate datasets.

- High focus on Community Earth System Model data
- Special (and fast) algorithms for calculating vorticity/ divergence on HOMME grids
- **ParGAL** (Parallel Gridded Analysis Library) - will vastly improve the speed of climate data analysis compared to the current serial tools.
- **ParNCL** – NCL with file I/O and computational components parallelized

http://trac.mcs.anl.gov/projects/parvis/wiki

Argonne NATIONAL LABORATORY

Pacific Northwest NATIONAL LABORATORY

NCAR

Sandia National Laboratories

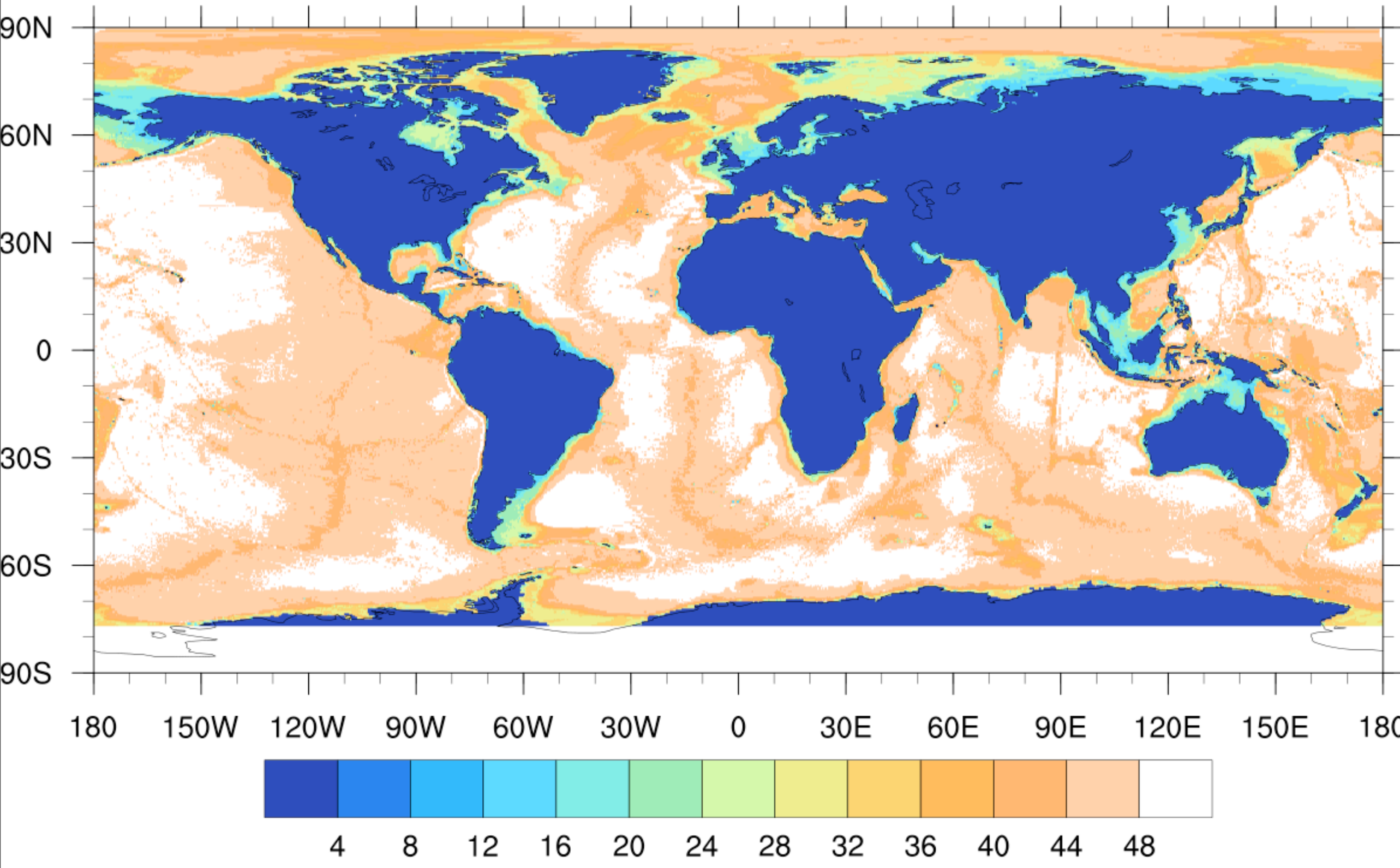
NCL and PyNGL Visualizations

Hundreds of example and real world scripts available:

http://www.ncl.ucar.edu/Applications/
http://www.pyngl.ucar.edu/Examples/gallery.shtml

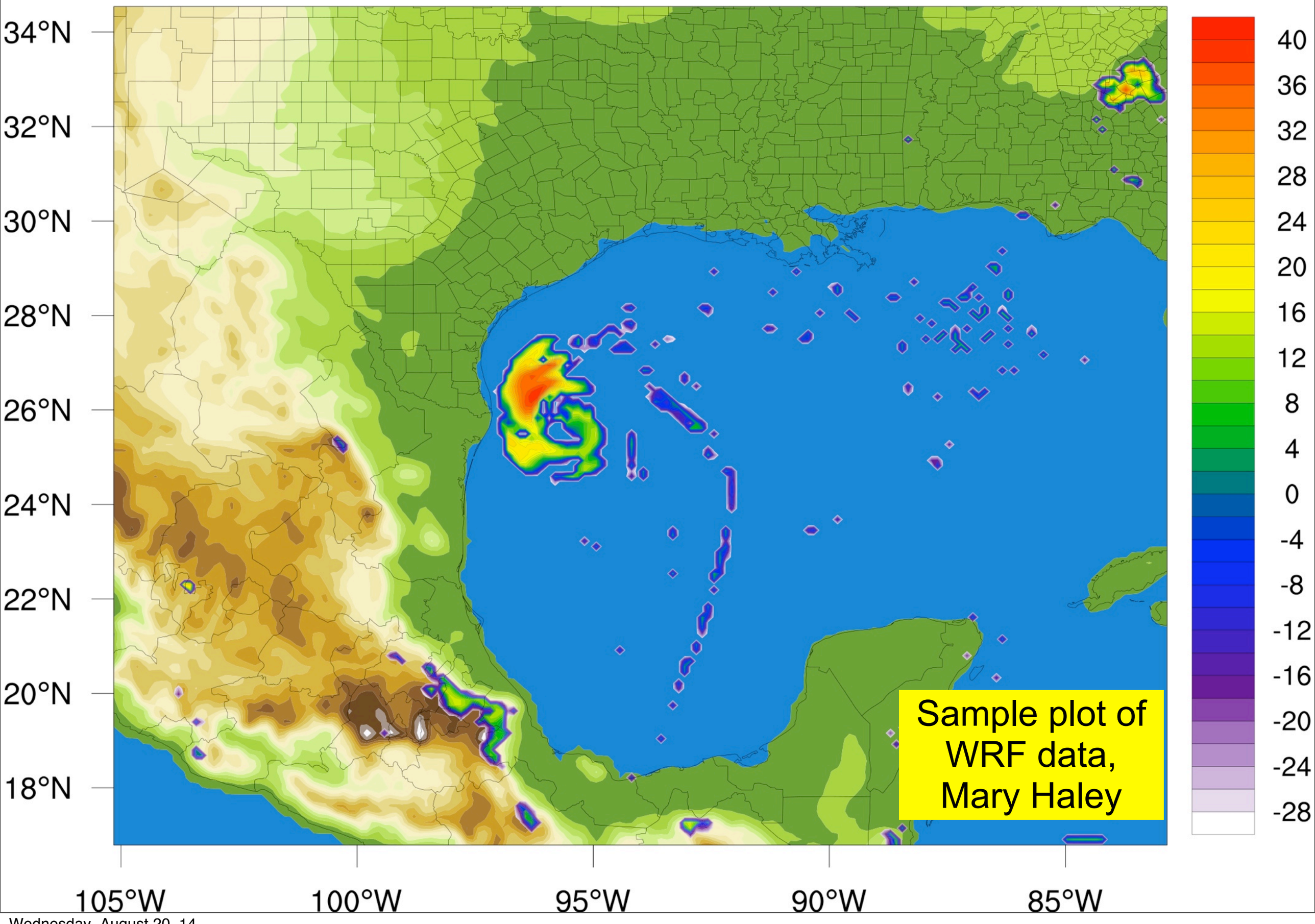
Most visualizations that follow are taken from above URLs and have scripts you can download to generate them.

ORCA12-T321.mesh_zgr.nc (mbathy)



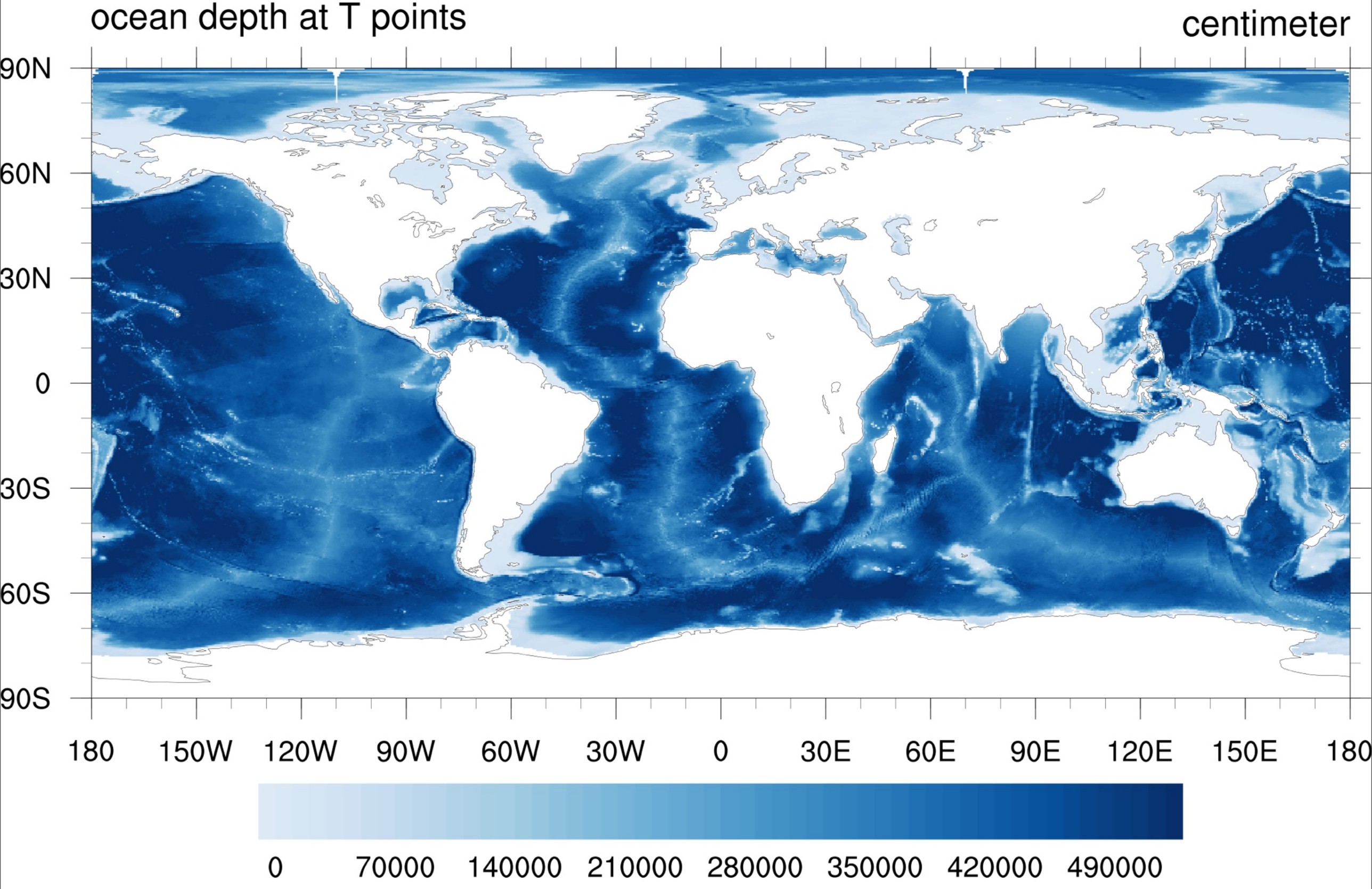
Bathymetry from an ORCA12 grid file provided by Romain Bourdalle-Badie of Mercator Ocean. The variable is dimensioned 3059 x 4322

Reflectivity (dBZ) at level = 0.996

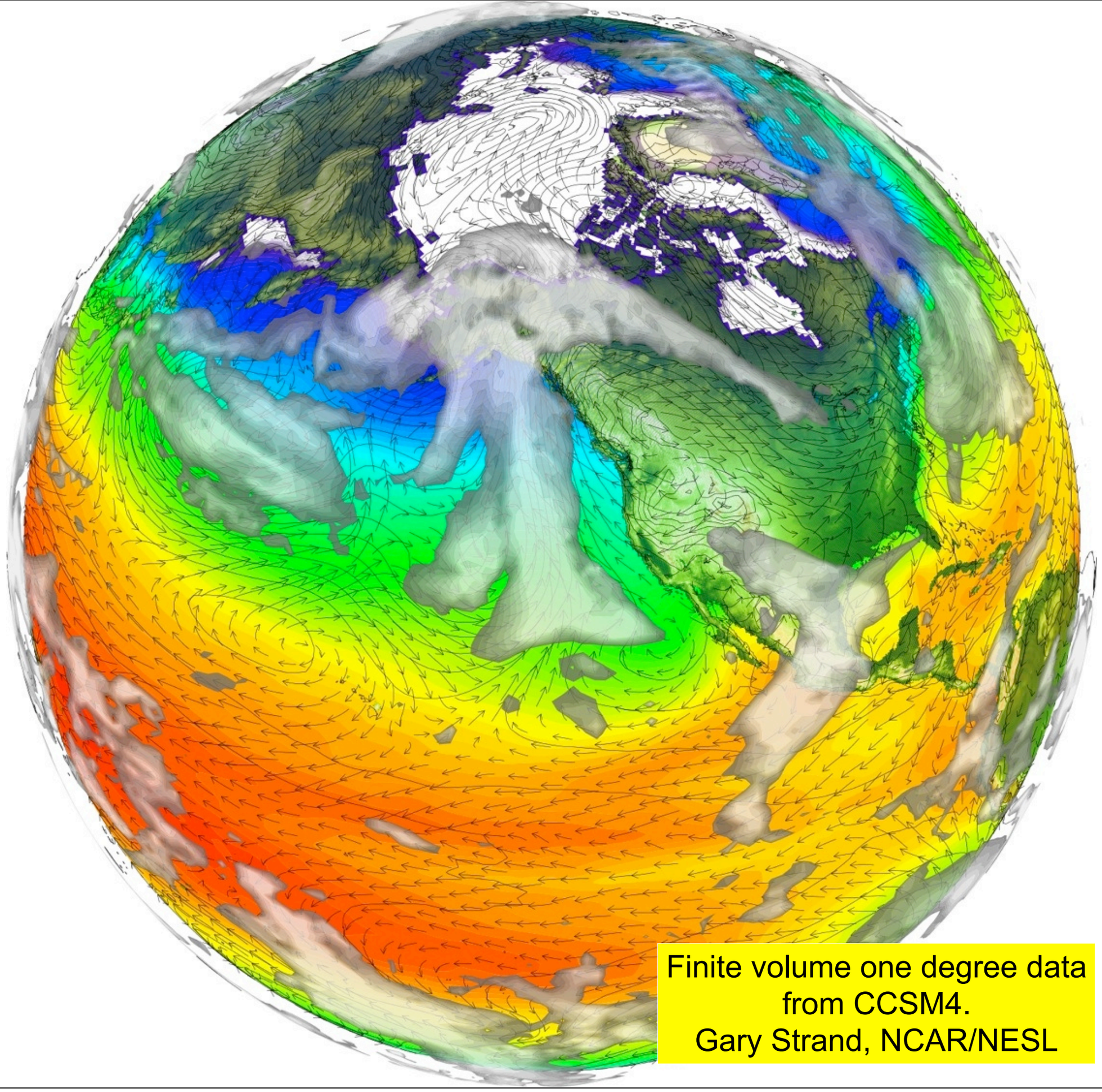


Wednesday, August 20, 14

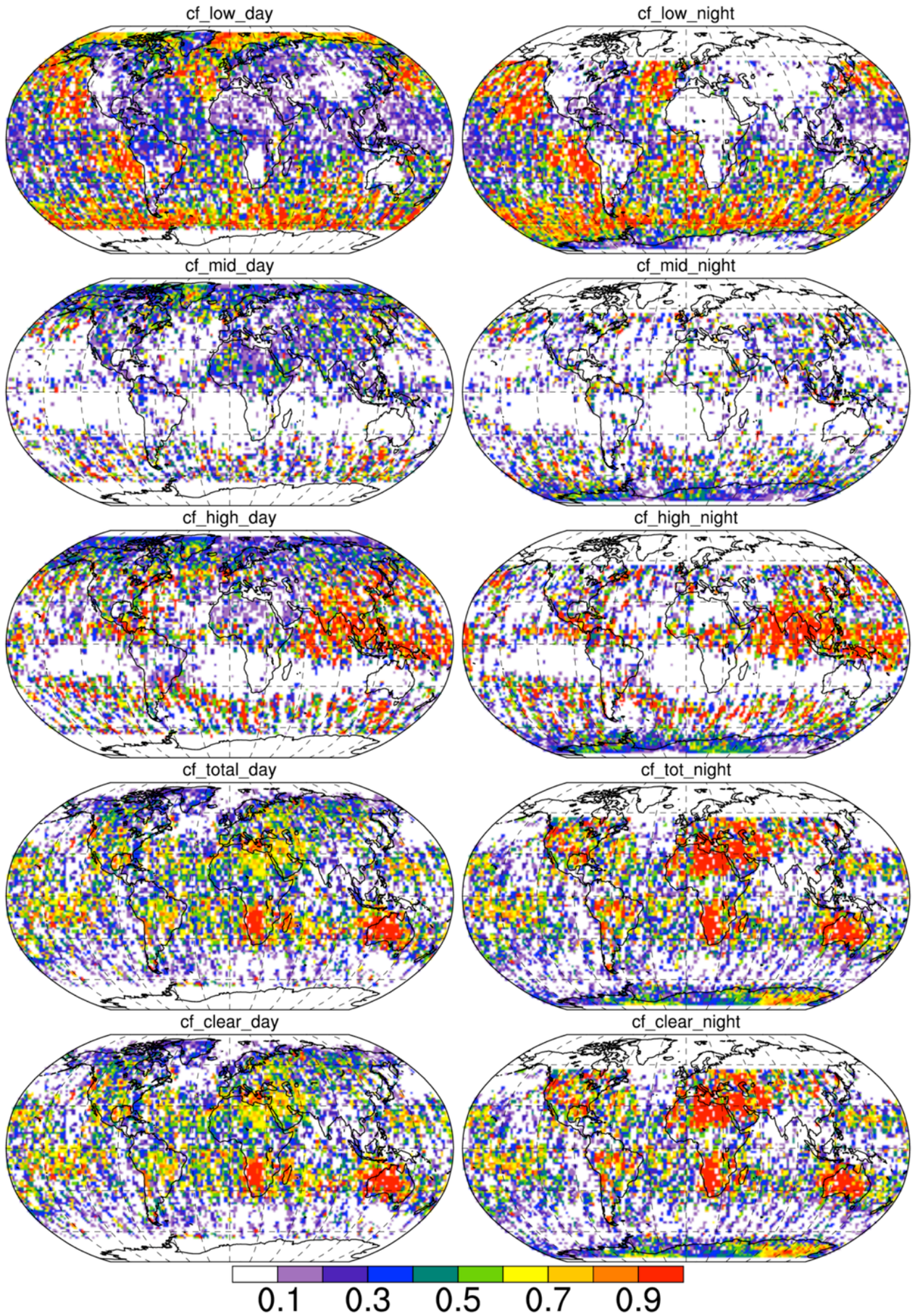
grid.tx01_62l.2013-07-13.nc (2400 x 3600, raster)



High-resolution POP grid from Steve Yeager, NCAR/CGD, 44 CPU seconds



Finite volume one degree data
from CCSM4.
Gary Strand, NCAR/NESL



CALIPSO was launched on April 28, 2006 to study the impact of clouds and aerosols on the Earth's radiation budget and climate.

It flies in formation with five other satellites in the international "A-Train" constellation for coincident Earth observations. The CALIPSO satellite comprises three instruments, the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), the Imaging Infrared Radiometer (IIR), and the Wide Field Camera (WFC).

CALIPSO is a joint satellite mission between NASA and the French Agency, CNES

NCL/PyNGL visualizations of shapefile data

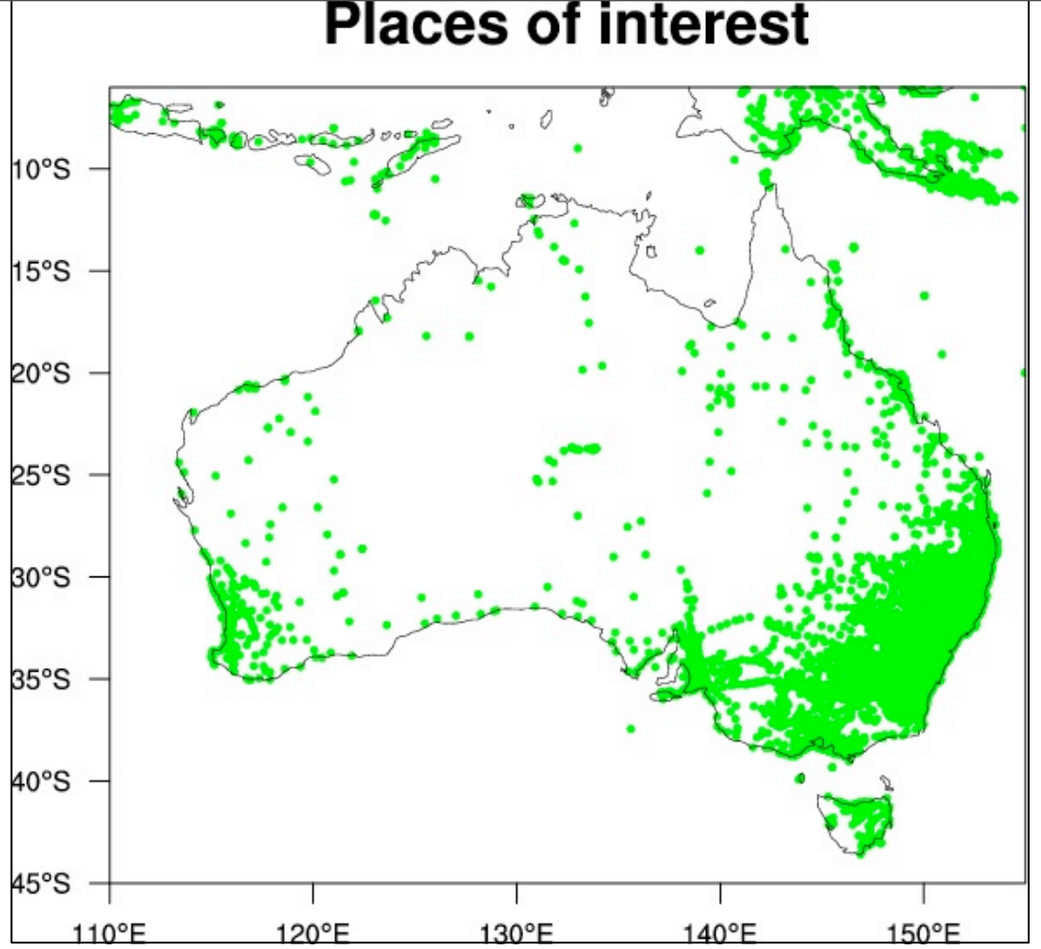
The ESRI Shapefile is a popular geospatial vector data format for geographic information systems software. It is developed and regulated by ESRI as a (mostly) open specification for data interoperability among ESRI and other software products.

Shapefiles describe a homogeneous set of geometrical features comprised of either points, polylines, or polygons.

Numerous (and free) shapefiles can be found by googling on the web.

Global Administrative Areas database (http://www.gadm.org) offers consistent administrative boundaries at 3 levels. The level 0 database (nations) is good to use for global or mesoscale results, level 1 is the first level of sub-national administration (typically states/provinces and territories) while level 2 offers the second level of administration and is potentially useful for high-resolution plots. The global shapefiles are large but it's possible to download individual countries separately.

http://www.ncl.ucar.edu/Applications/shapefiles.shtml

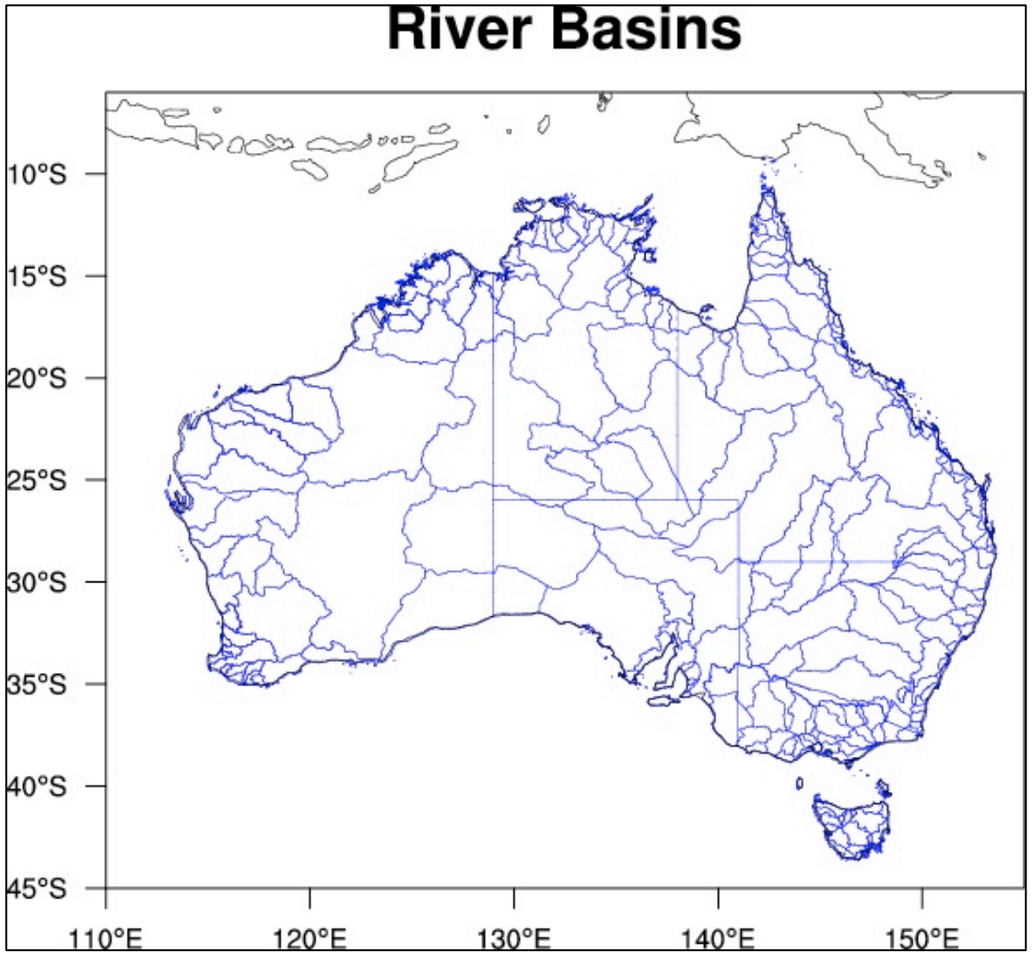


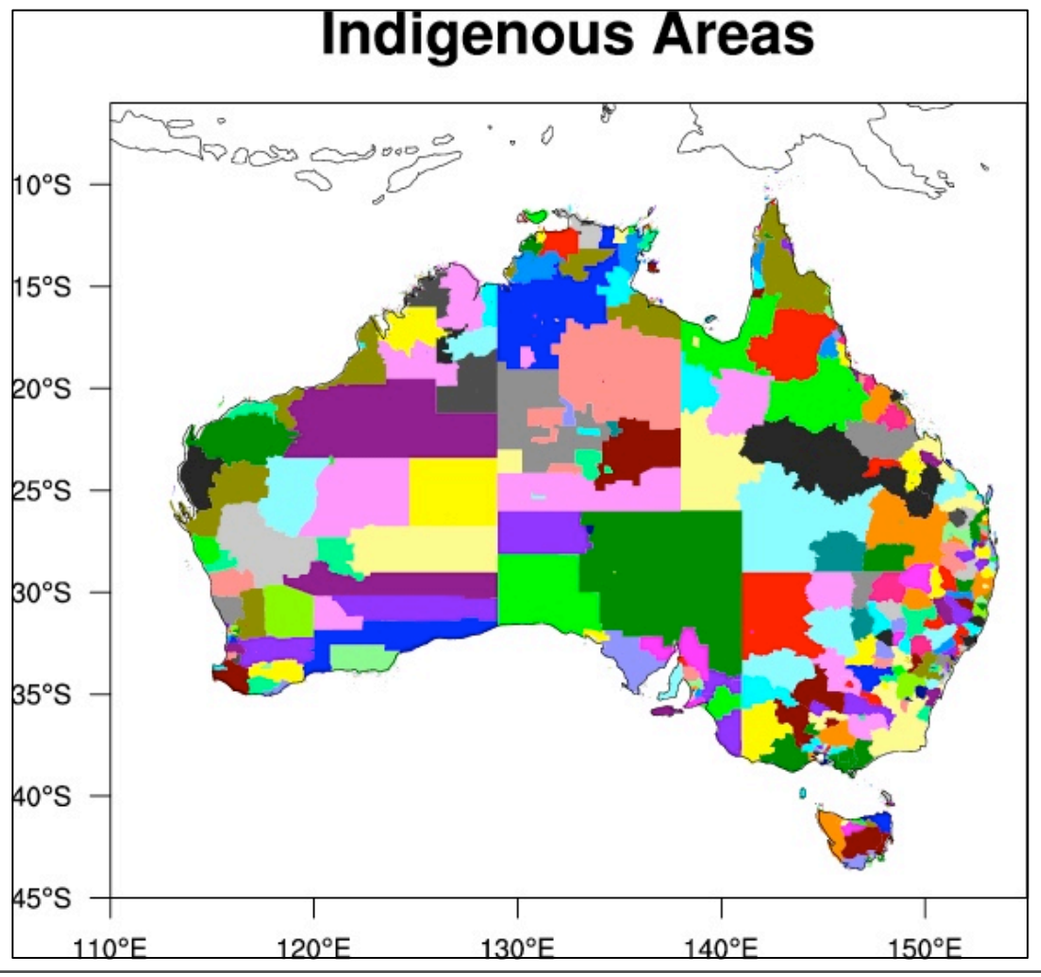
The three types of shapefiles supported by NCL:

Point – locations of cities, population data, election results ☺

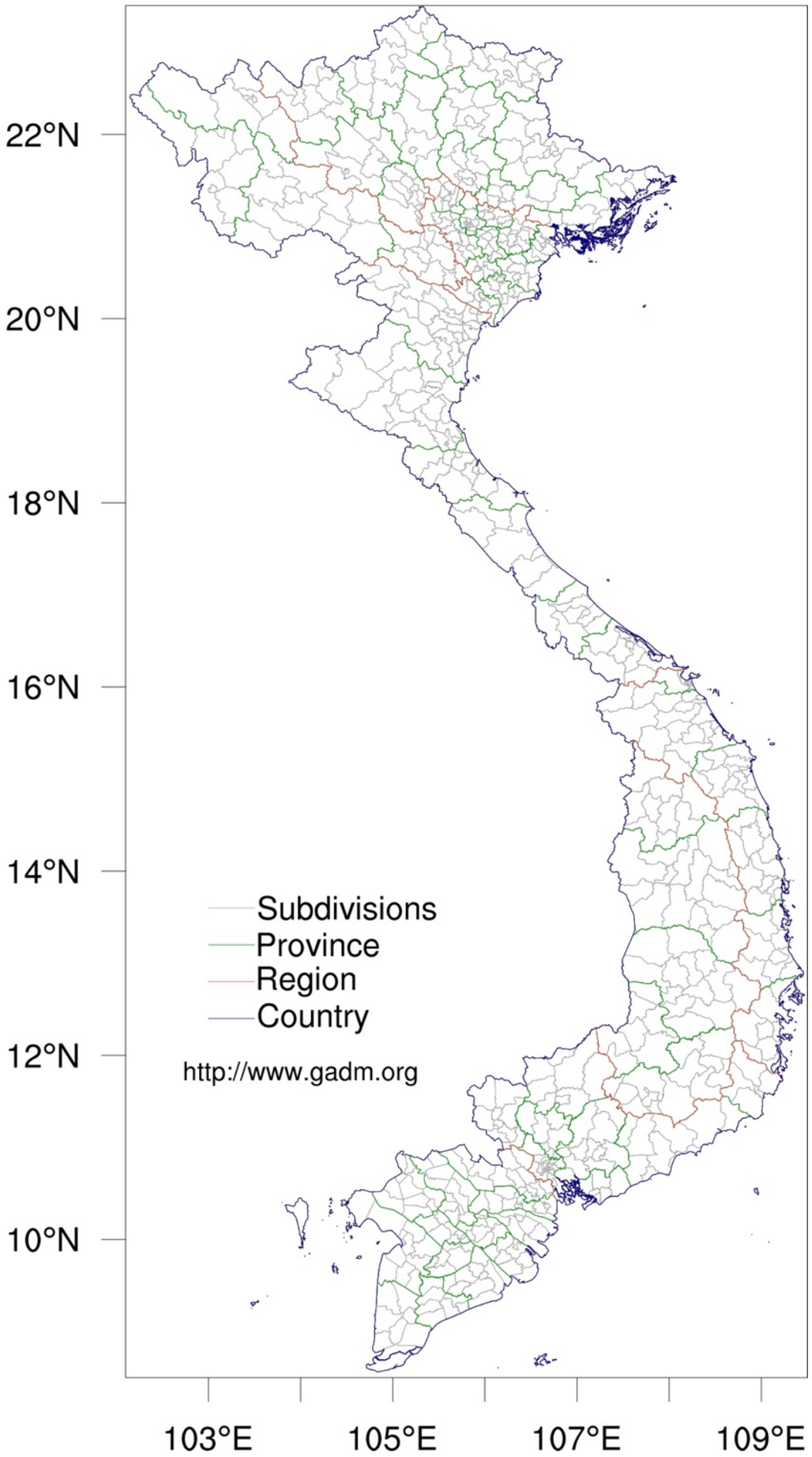
Line – rivers, roads, trails

Polygon – counties, lakes

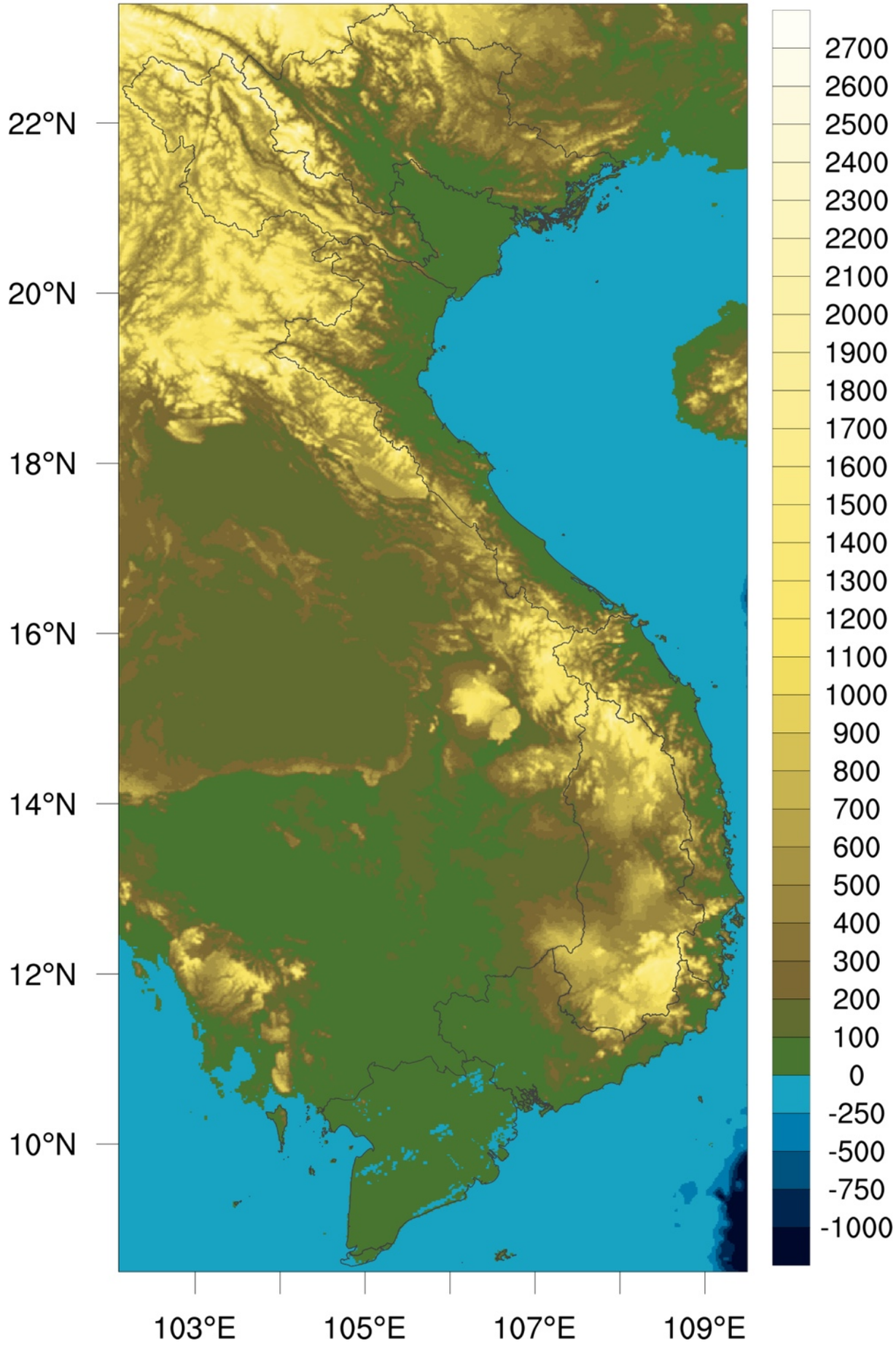




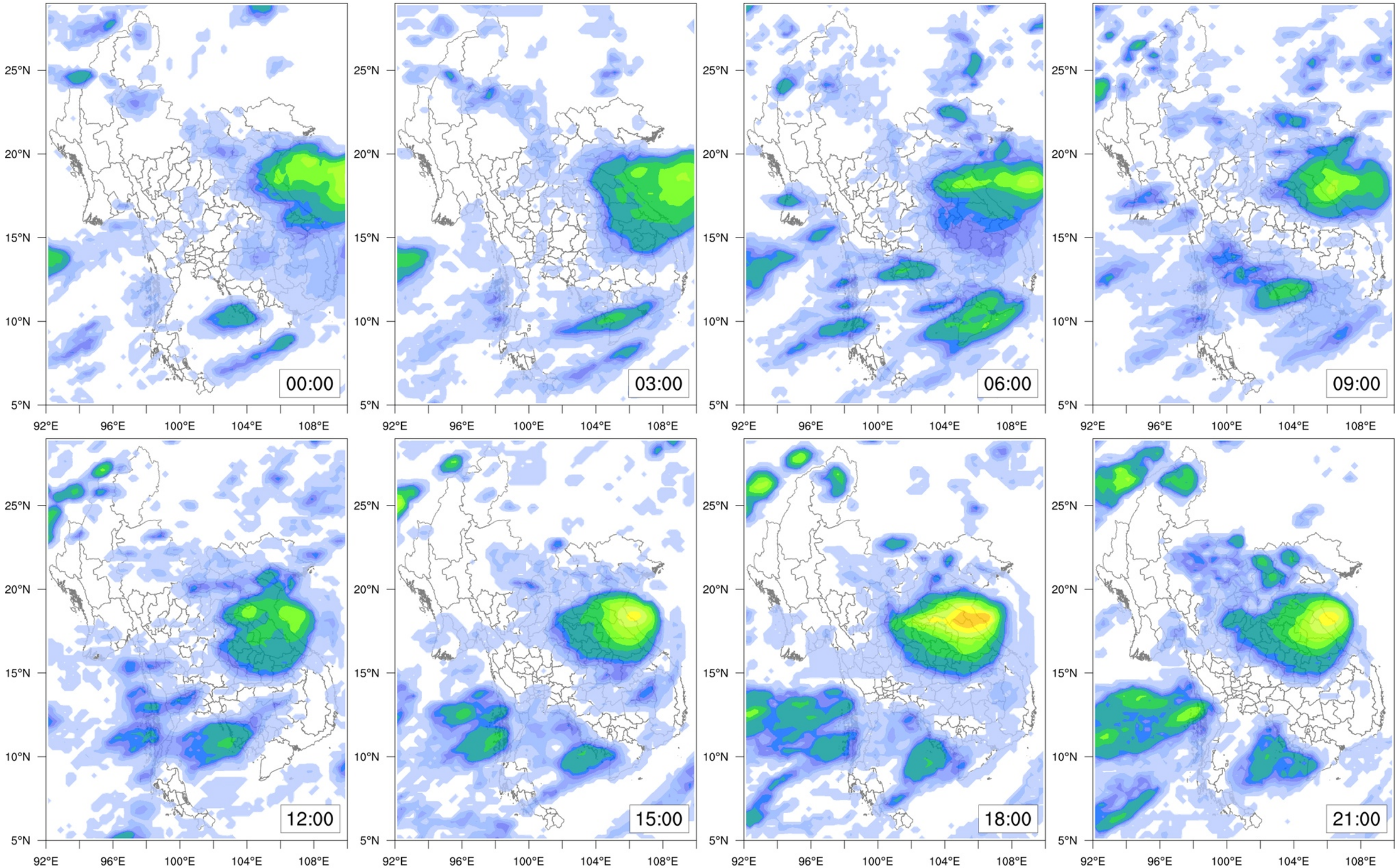
Administrative outlines for Vietnam



ETOPO1_Bed_c_gmt4.grd.nc



Lower Mekong: PERSIANN precipitation (mm/3hr) Jun 22, 2013



High-resolution precipitation data: http://chrs.web.uci.edu/persiann/
Shapefile outlines: http://www.gadm.org/country

NCL scripts for analyzing and visualizing WRF-ARW data (WRF-NCL)

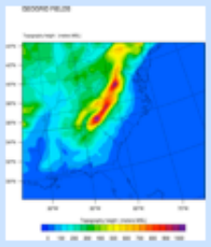
The Weather Research and Forecasting (WRF) Model is a next-generation mesoscale numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs.

http://www.mmm.ucar.edu/wrf/OnLineTutorial/Graphics/NCL/

WRF ARW OnLineTutorial

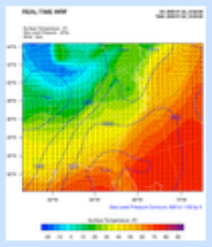
WRF-ARW online tutorial is maintained by Cindy Bruyère and Abby Jaye NCAR/MMM

Basic Plots



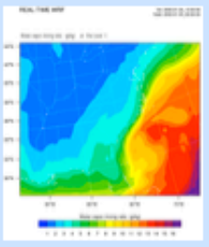
Basic Plot Setup
(This series of examples takes users though same basic steps in generating plotting scripts.)
Get and plot a single field
Multiple input files

Basic Surface Plots



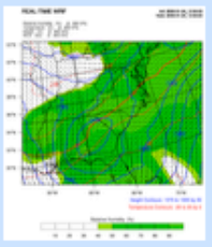
Surface 1
Surface 3
Surface 2

Plots on Model Levels



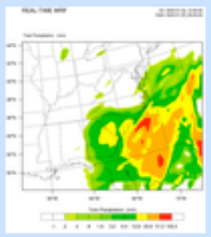
Clouds
Levels from wrfout files
Levels from metgrid files

Plots on Interpolated Levels



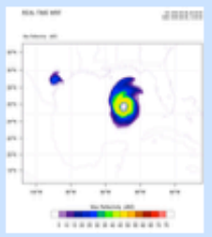
Height Levels
Pressure Levels

Plotting Precipitation



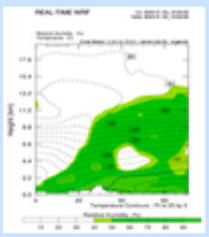
Precipitation

Diagnostics



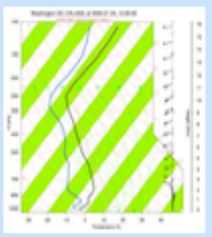
CAPE
dBZ
Vorticity
(More diagnotics are available, shown are only some newer/special diagnostics)

Cross-section Plots



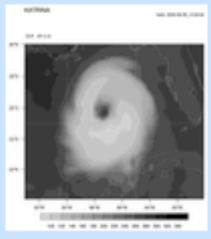
Height - Through a Pivot Point
Height - Point A to Point B
Pressure
Limited Vertical Extent
For 2D fields

Skew_T Plots



Skew_T

Speciality Plots



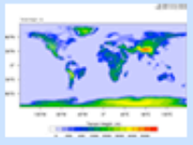
Overlay
Zoom
Overlay & Zoom
Panel 1
Panel 2
Meteograms
WRF Time Series data
All fields in a file

Preview Domain

This functionality, although available in NCL version 5.0.1, is still experiential.

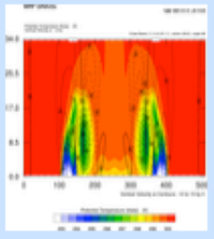
Preview

Global WRF



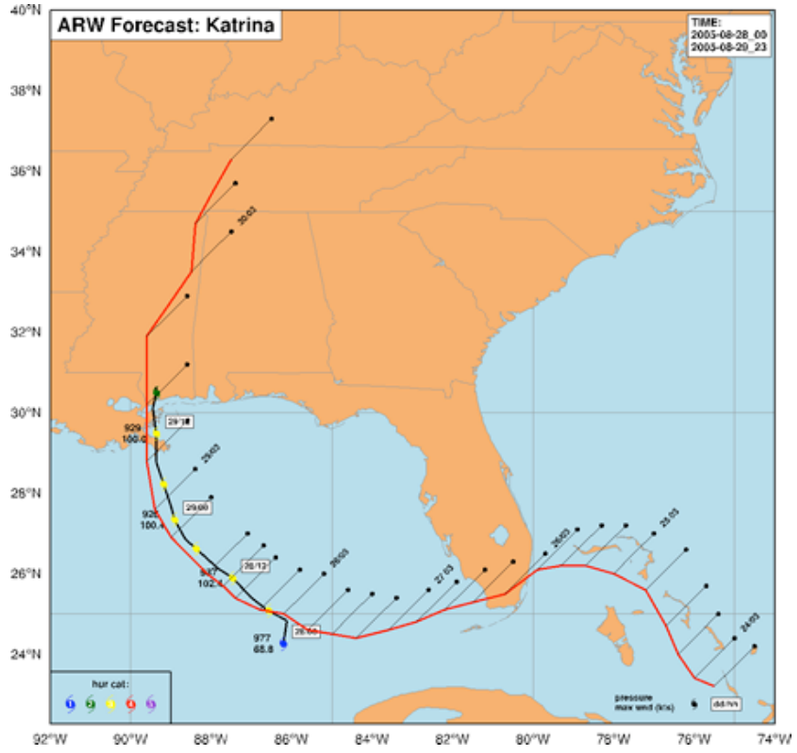
gWRF_merc

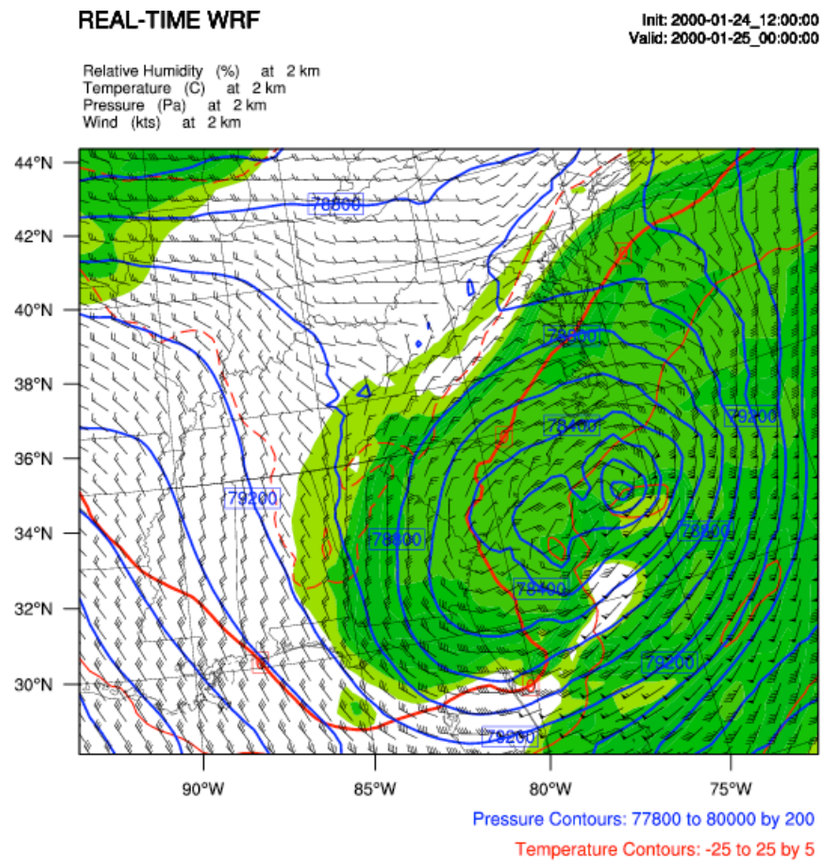
Idealized cases

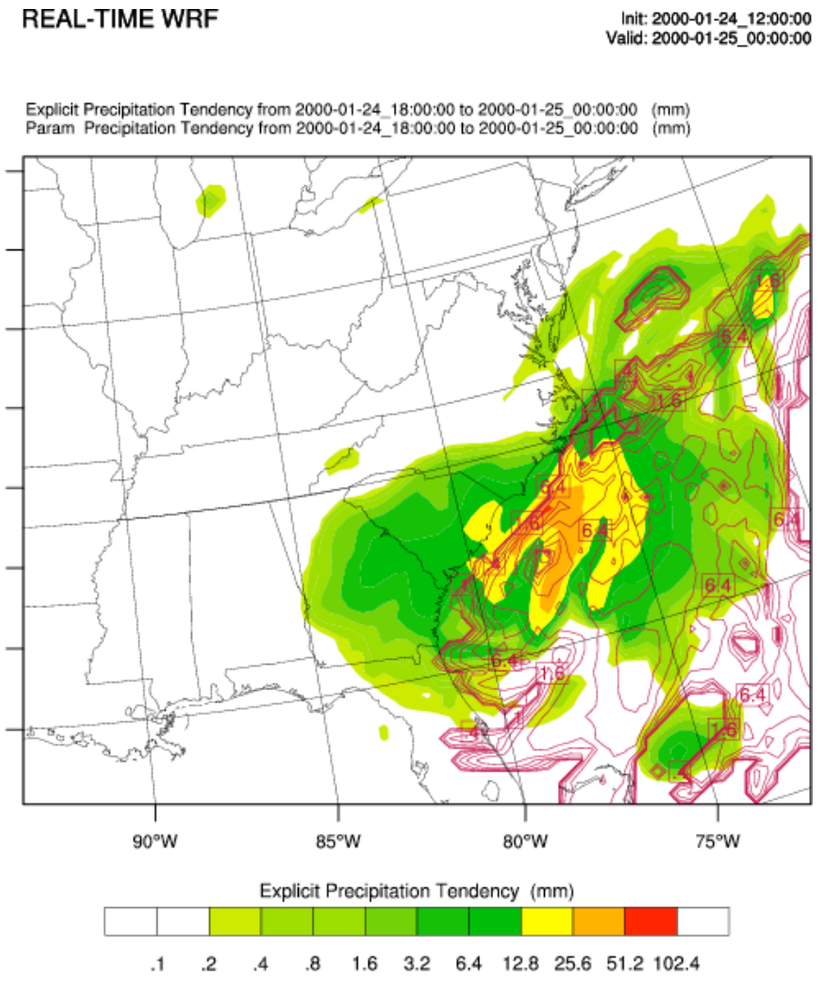


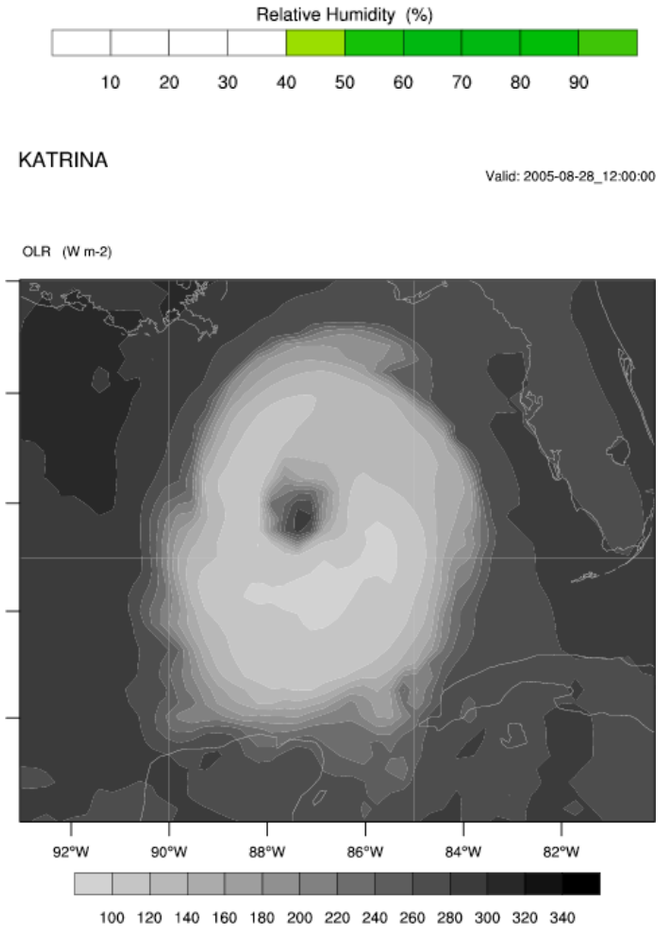
wrf_Grav2x
wrf_Hill2d
wrf_Squall_2d_x
wrf_Squall_2d_y
wrf_Seabreeze2x
wrf_BWave
wrf_QSS

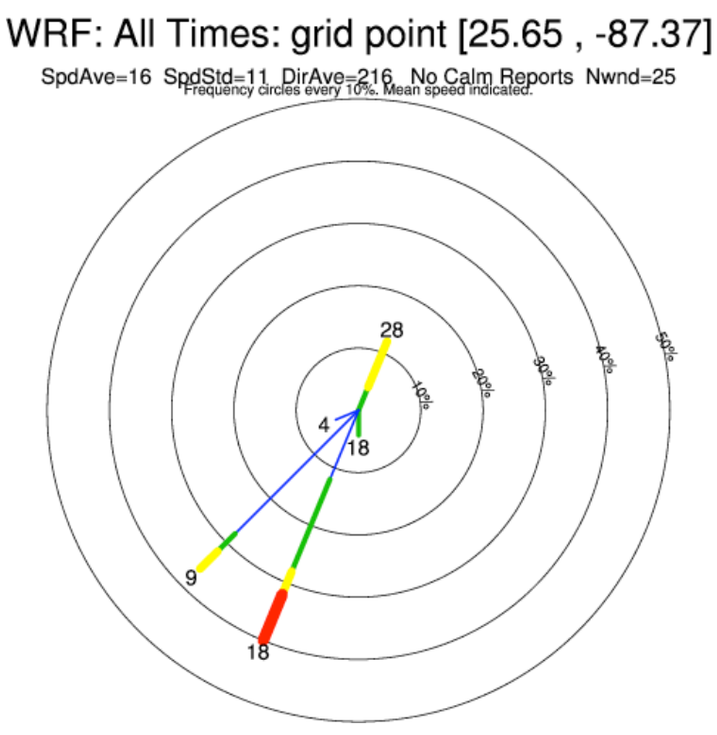
Sample WRF graphics generated by NCL





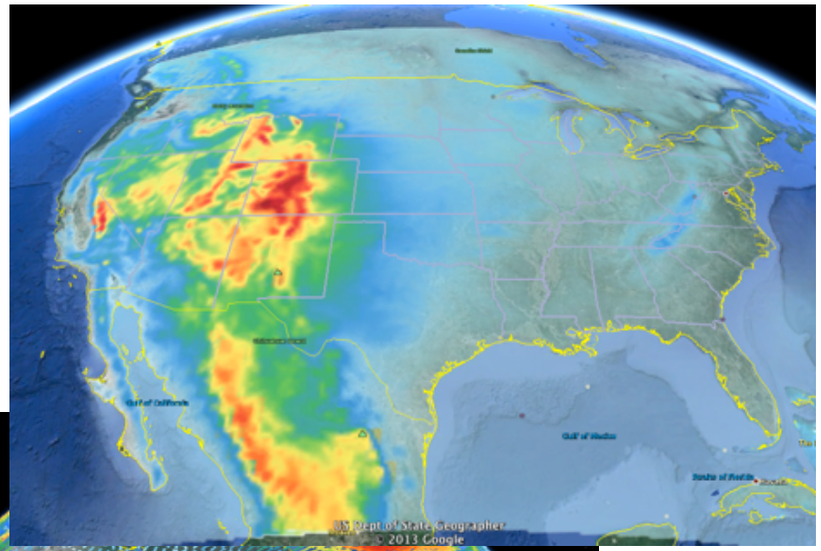


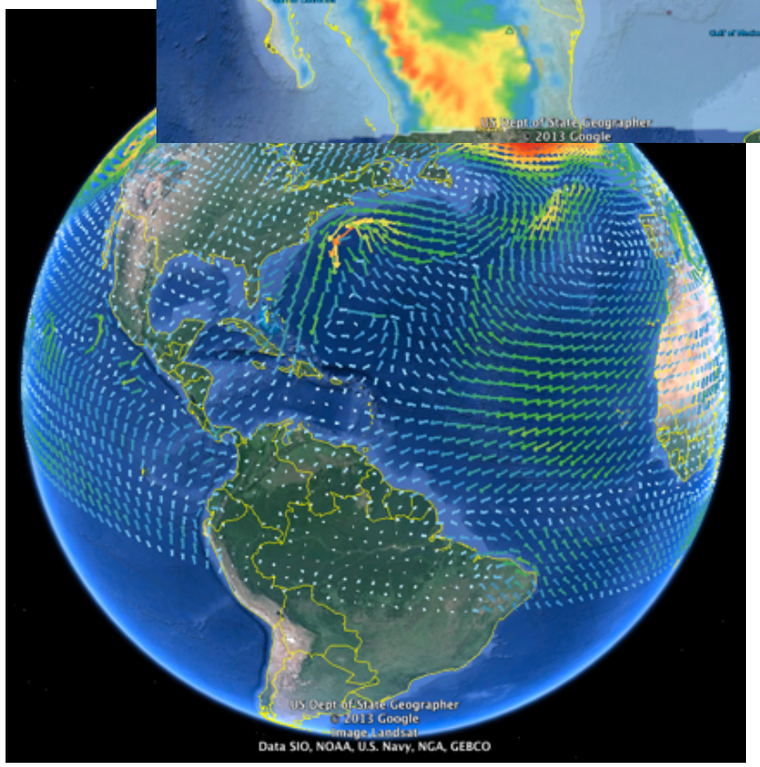


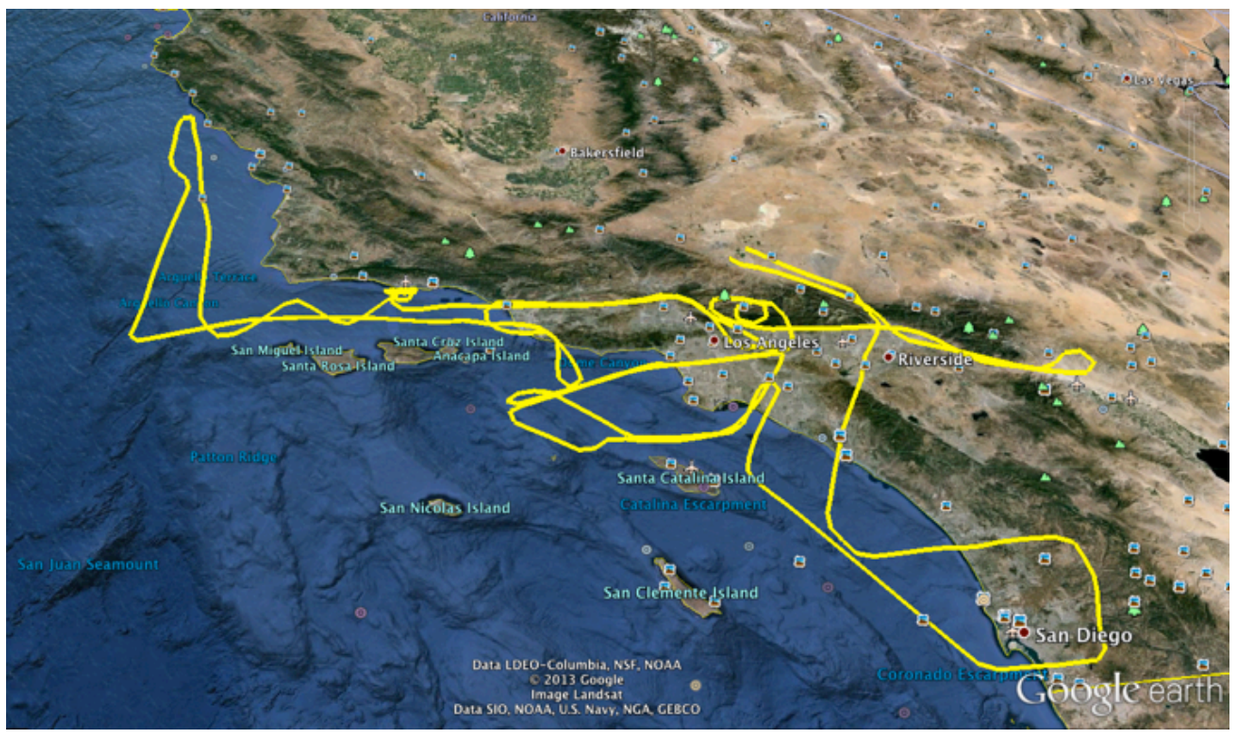


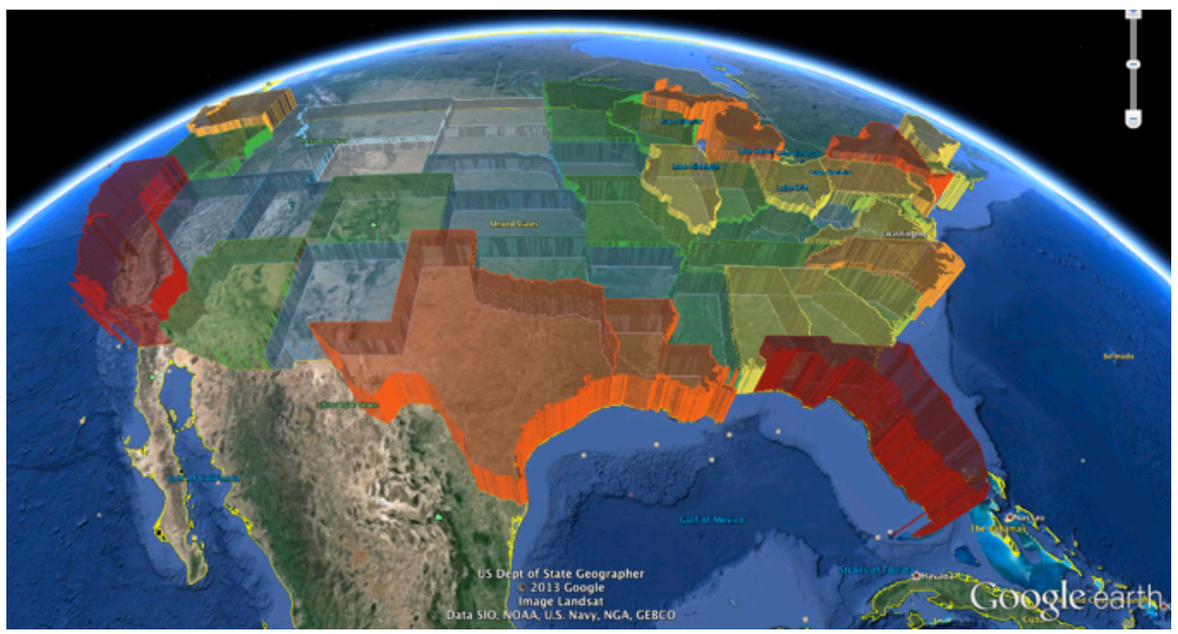
NCL and GoogleEarth™

- Summer intern project in CISL
- Google Earth™ offers a highly intuitive and usable visualization environment for displaying Earth science data in its geospatial context.
- A library of NCL routines has been developed to enable earth scientists to easily convert geo-referenced model output and other data to KML for display in Google Earth.







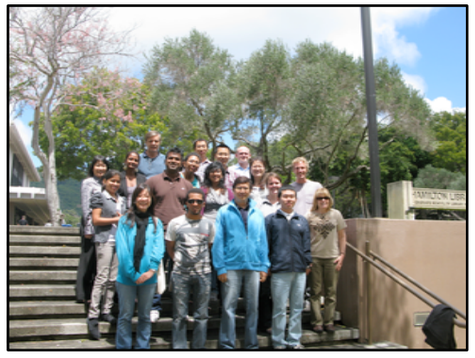


Discovering, accessing, and analyzing diverse data sources can be difficult and time consuming! Training can be very beneficial.

“The Complexities of individual data sources or models tend to absorb individual scientists and science teams. This limits the synergistic combination of data sources and science teaming."
-- Chris Elvidge, earlier at this meeting

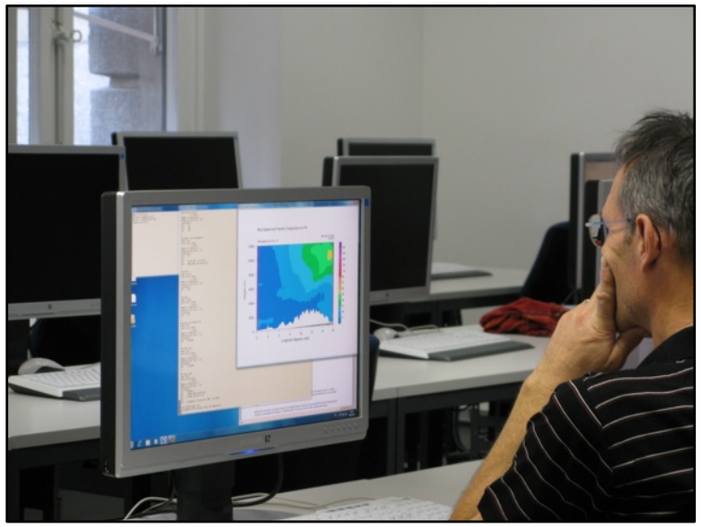
NCL Data Analysis and Visualization Workshops

- Primarily for climate and weather researchers
- 3.5 days with lectures and intensive hands-on lab sessions
- Co-taught by NCAR scientific and engineering staff
- Students encouraged to BYOD (Bring Your Own Data)
- Free for all students; full funding provided for EPSCoR and MSI students; partial funding provided for U.S. universities
- Workshops given locally and at universities and research institutions world-wide
- 73 workshops taught to 1139 students since February 2000



International Workshops

MPI Hamburg 2008-2013



ETH Zürich 2010

UNSW Sydney 2011

BoM Melbourne 2011

CERFACS Toulouse 2012

UFRN Natal 2013

APEC Busan 2006

VAPOR

Courtesy of John Clyne, NCAR/CISL/OSD

NCAR

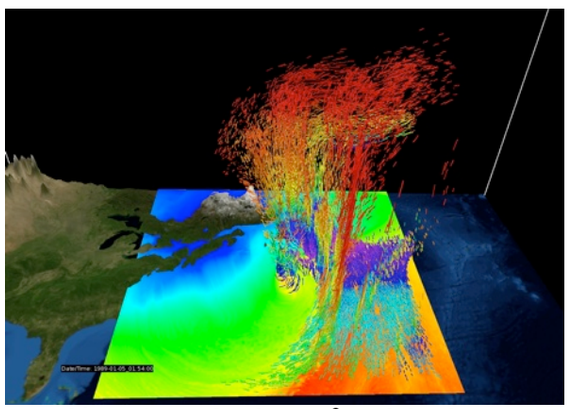
VAPOR

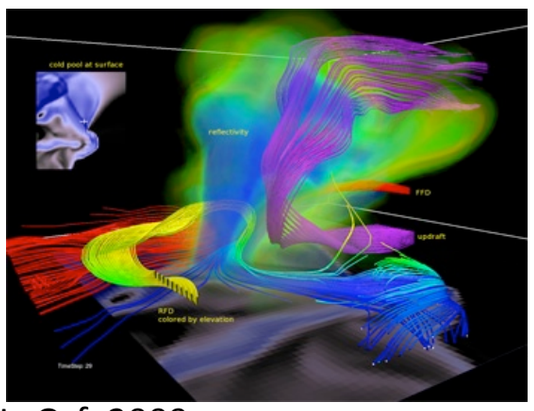
Visualization aided data analysis for the earth sciences

- A **domain-focused**, **open source** visual data analysis package targeted for researchers working in **numerical weather prediction**, **atmospheric, oceanic, and related sciences**
- By leveraging a wavelet-based *intelligent data storage* model VAPOR enables highly interactive exploration of the **largest numerical simulation** outputs using only **commodity computing** resources.
- A **community-driven** feature set guided by an international steering committee of computational scientists working in a broad gamut of earth science disciplines
- Metrics:
  - ~6000 registered users since January, 2011
  - ~4000 unique VAPOR web site visitors per month in 2013 (up from 1000 in 2012, and 500 in 2011)
  - ~100 scholarly citations for VAPOR

I. Grooms, et al. 2010

S. Wedemeyer-Böhm, et al. 2012


M. Shapiro; S. Grønås, 2012



L. Orf, 2009

www.vapor.ucar.edu

VAPOR is funded in part through U.S. National Science Foundation grants 03-25934 and 09-06379, 14-40412, a TeraGrid GIG award, and the Korean Institute for Science and Information Technology

NCAR Computational & Information Systems Laboratory

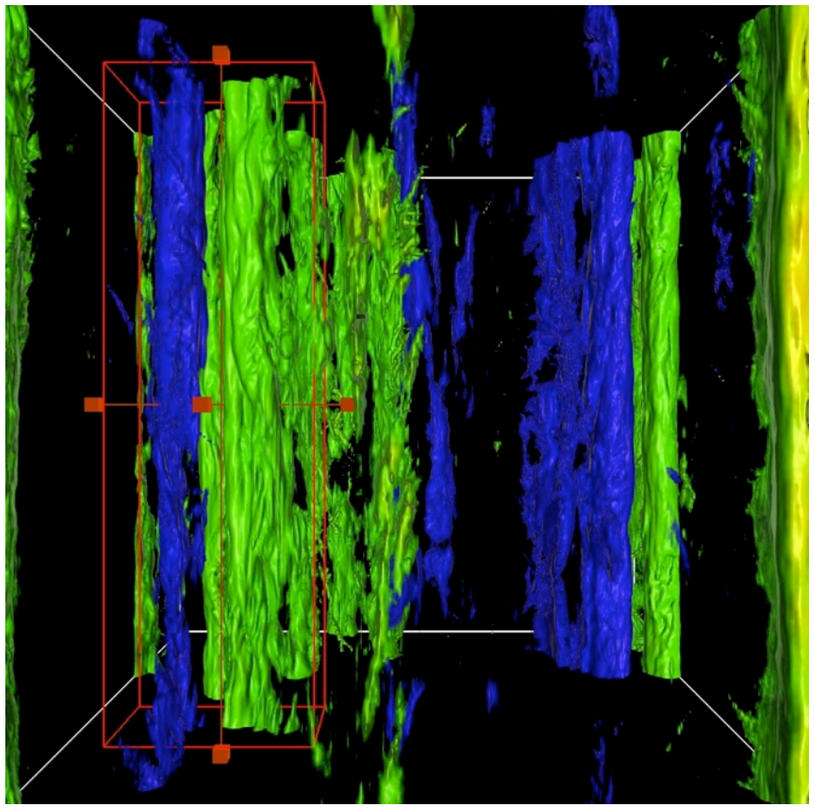
KiSTi www.kisti.re.kr

NSF

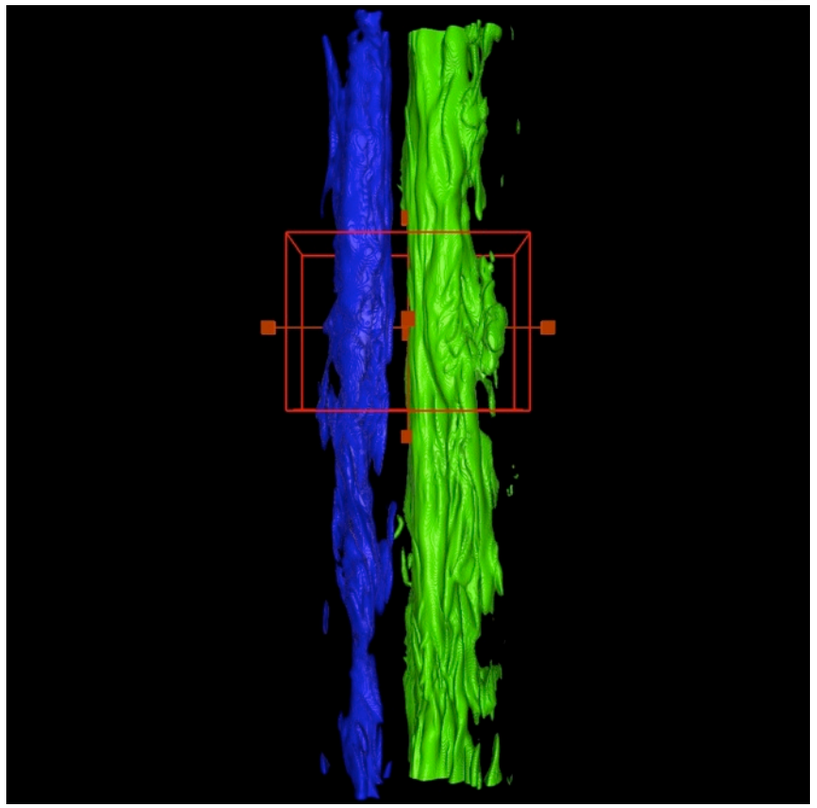
Wavelet Enabled Progressive Data Refinement

Exploring a 1536^3 simulation output on a laptop

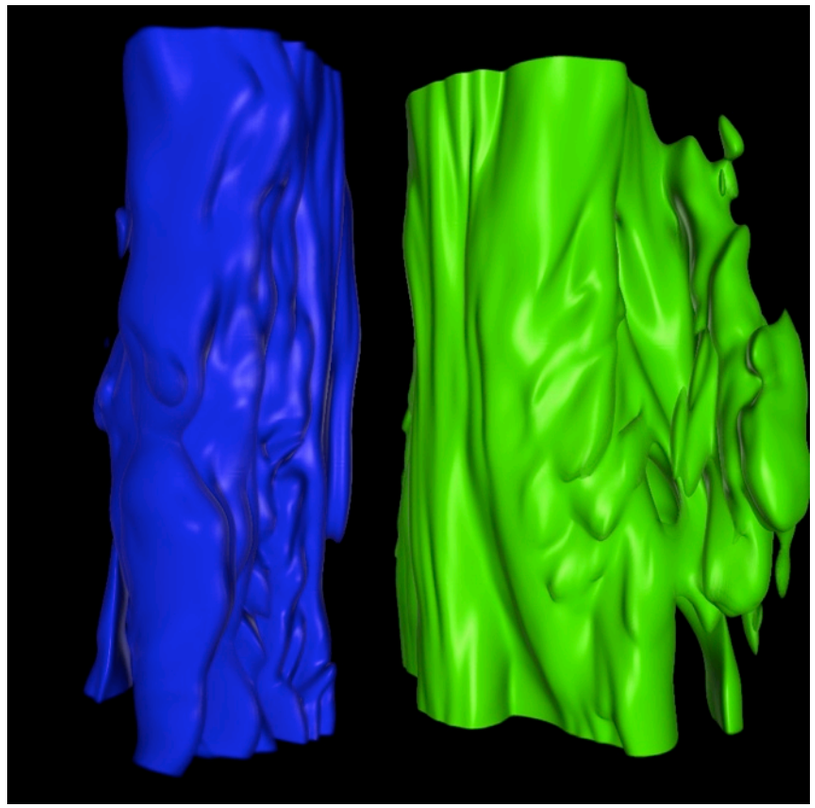
Browse highly compressed data, subset, refine, repeat.



Full domain, compressed 500:1
14GBs -> 29MBs



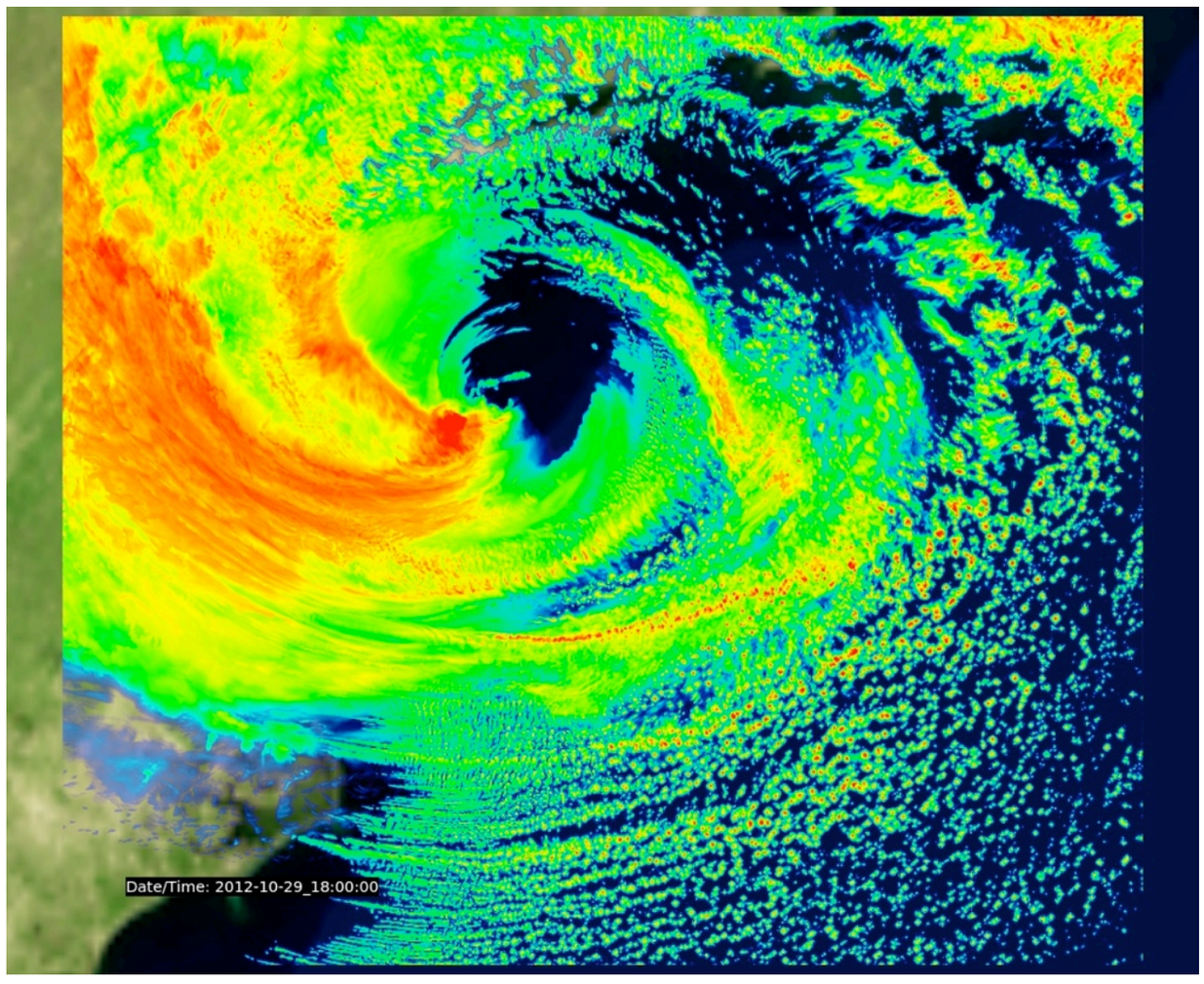
Region of interest (ROI), 100:1
800MBs -> 8MBs



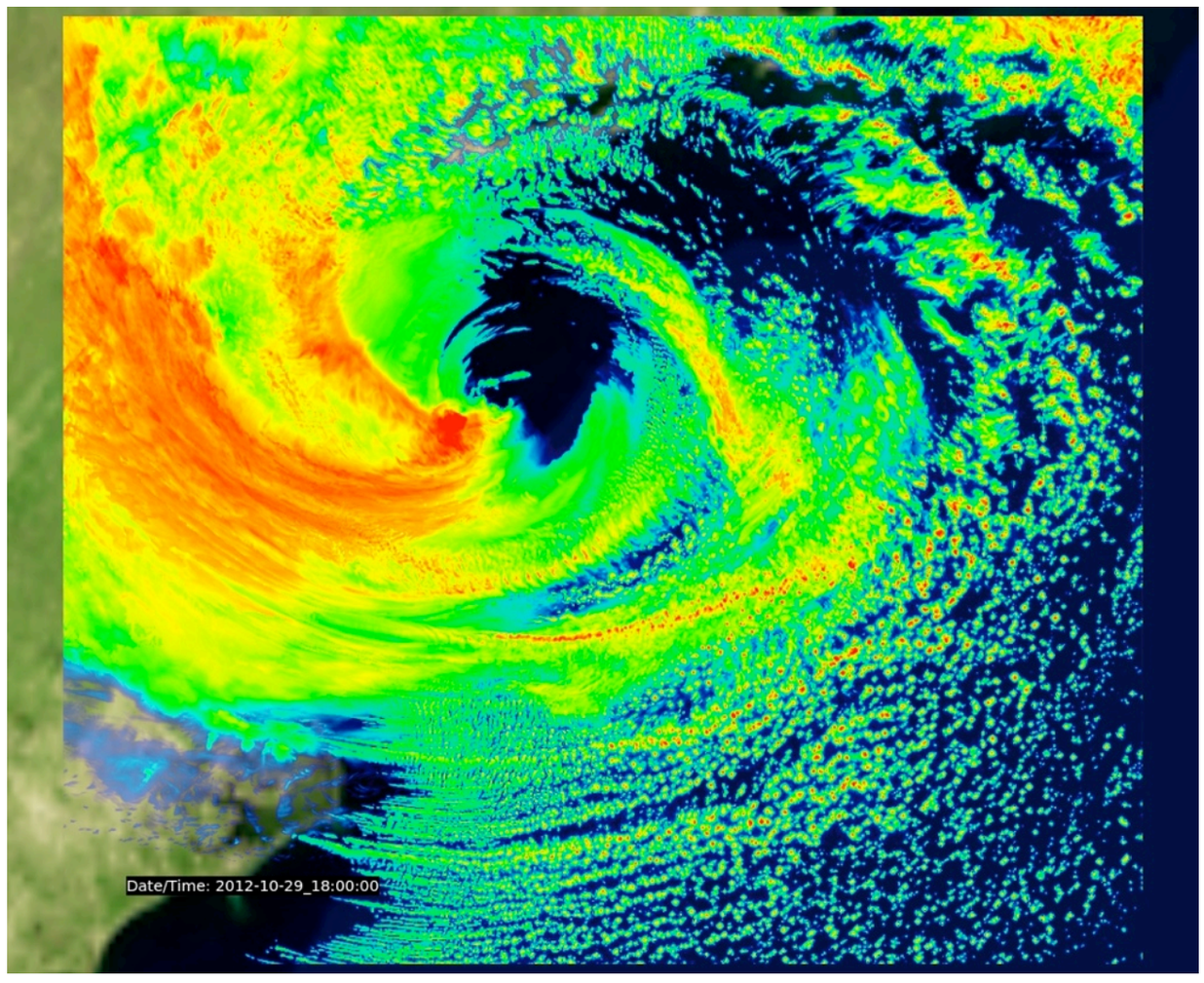
ROI, No compression
182MBs

Radar reflectivity derived from a 500m Weather Research Forecast simulation of Hurricane Sandy, the largest NWP model ever run [Johnsen 2013]

DBZ computed from original data (203 GBs)



DBZ computed from variables compressed 10:1 (20.3 GBs)



Supported file formats & model outputs

- Climate Forecast (CF) convention NetCDF files
  - MOM4
  - POP
  - ROMS
  - CAM
- WRF outputs
- NetCDF (structured grid data)
- Flash AMR
- Raw (binary) data

Features

- GPU enabled volume rendering and isosurfaces
- Cutting planes
- Isolines
- Flow visualization
  - Streamlines (steady flows)
  - Pathlines (unsteady flows)
  - Image Based Flow Vis.
  - Hedgehog plots (wind barbs)
- Geo-referenced data support
- NumPy/SciPy calculation engine
- Key-frame animation
- Quantitative analysis
  - Probes
  - Statistics

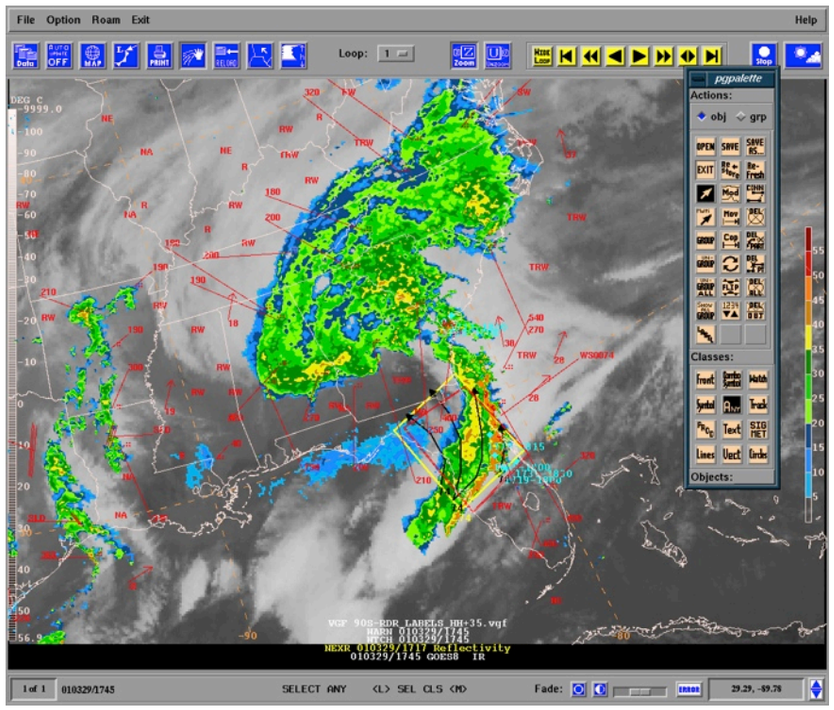
UCAR/Unidata Tools

Courtesy of Mohan Ramamurthy, UCAR/Unidata

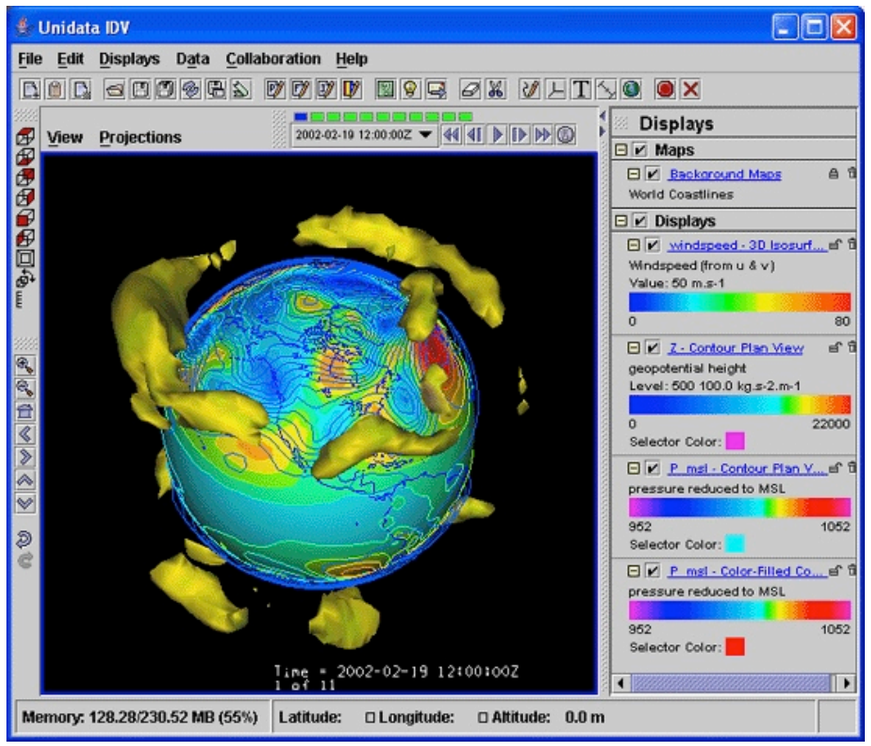
http://www.unidata.ucar.edu

Unidata Visualization Software

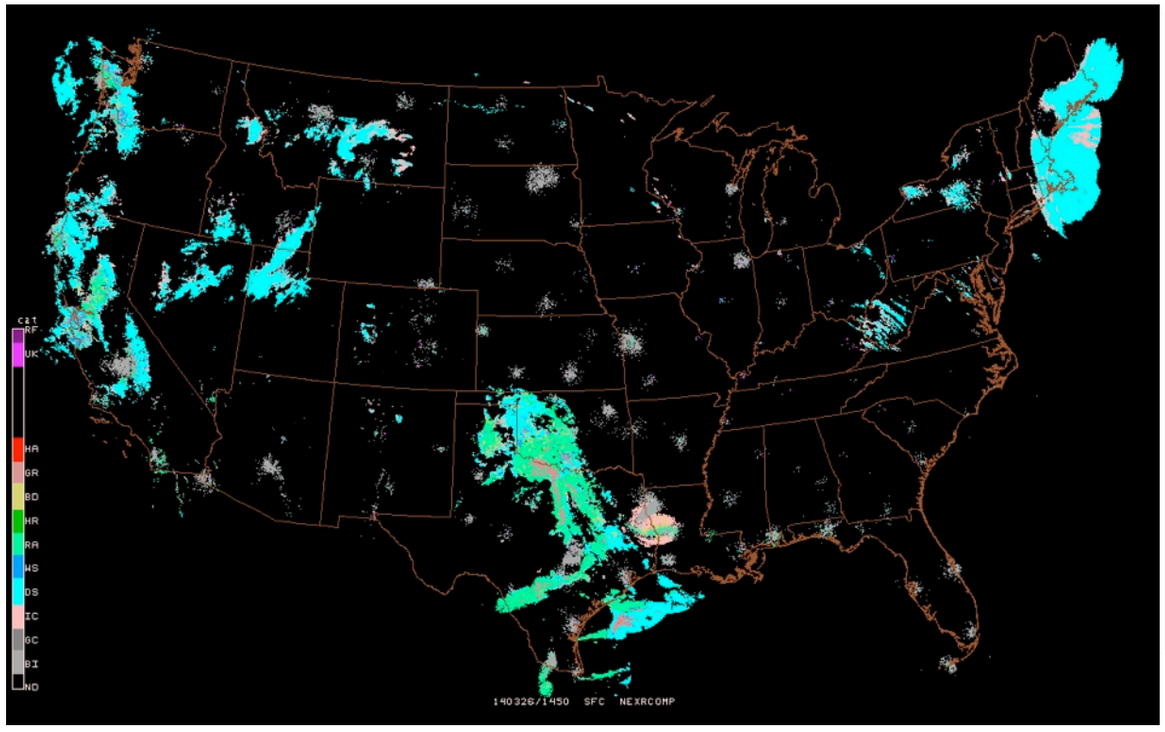
GEMPAK



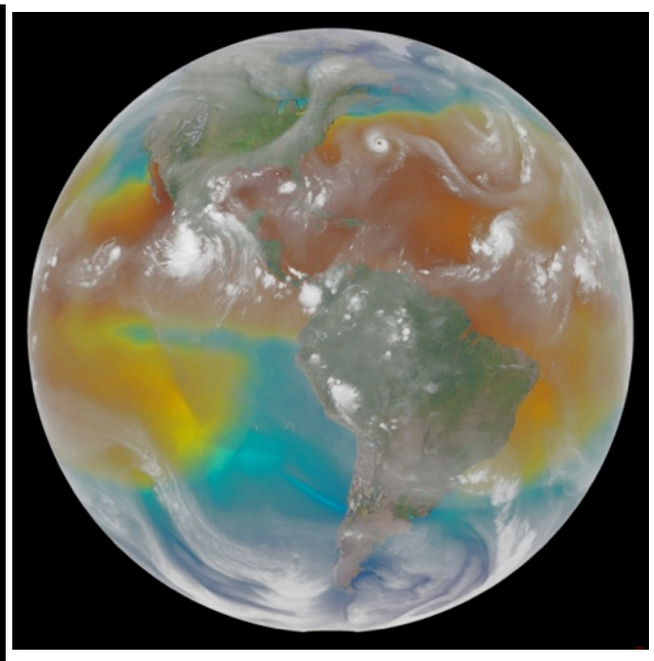
IDV



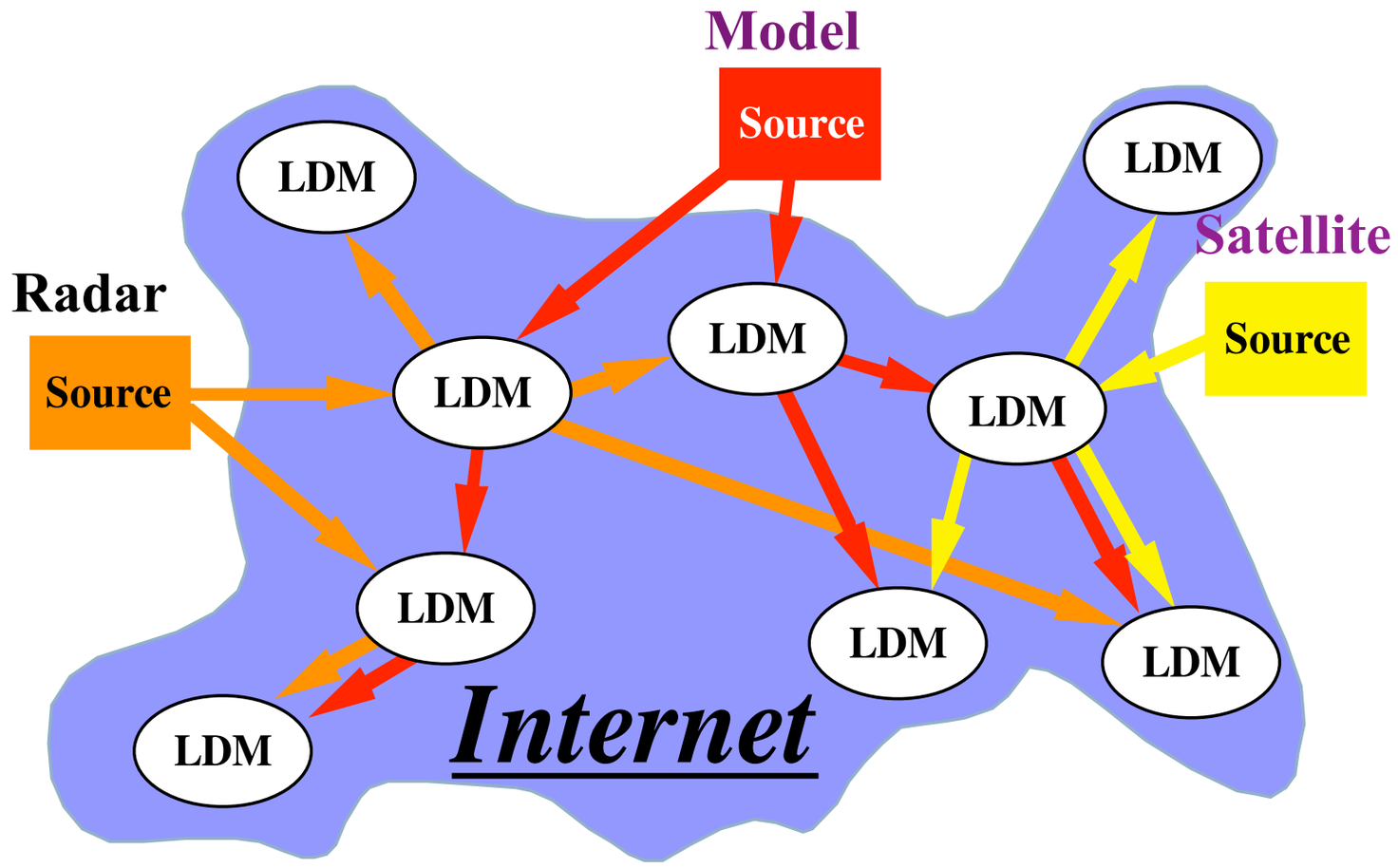
AWIPS II



McIDAS-X



Real-time Data Distribution



About 30 different streams of real-time weather data from diverse sources are provided to a global community.

Unidata systems move more data (~30 Terabyes/week) via Internet 2, more than any other advanced application.

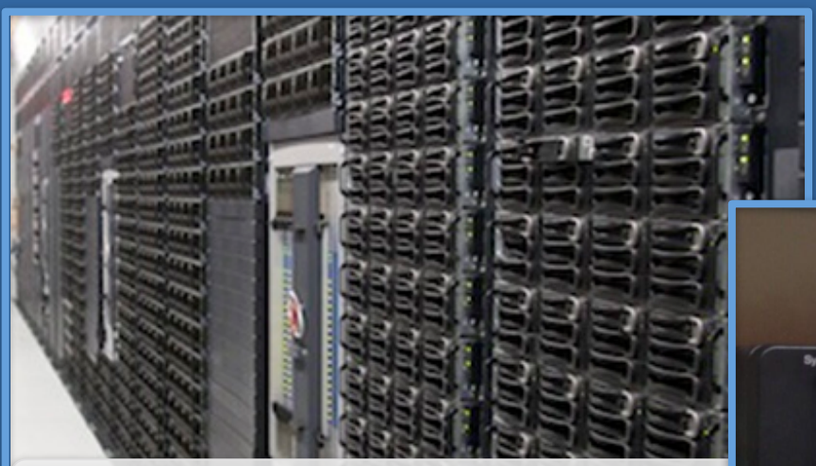
Globus Transfer

Courtesy of Rachana Ananthakrishnan and Ian Foster,
Computation Institute, Univ. of Chicago

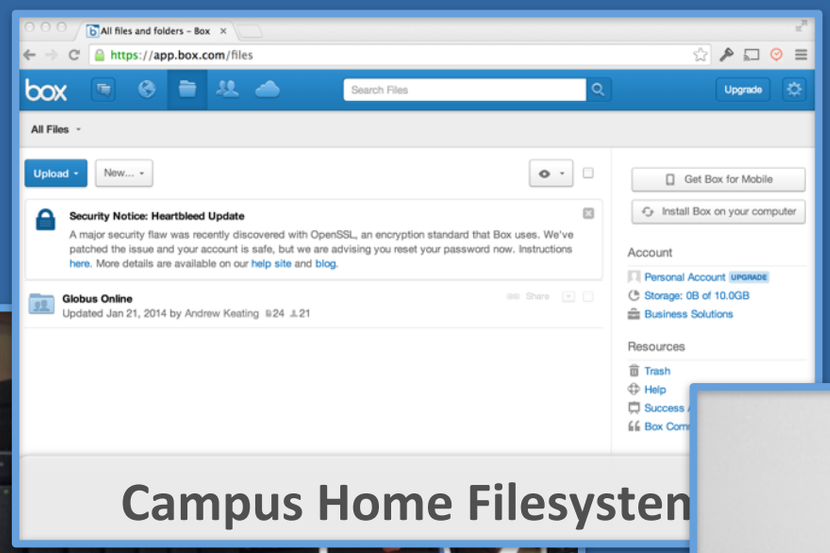
https://www.globus.org

NCAR

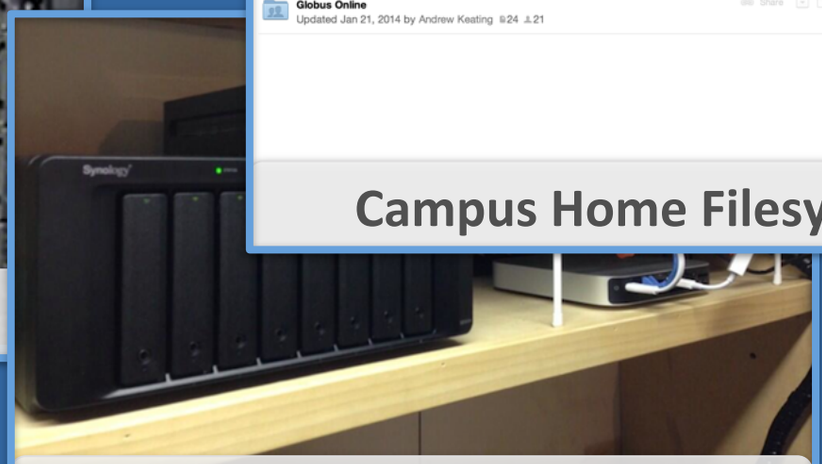
“I need to easily, quickly, & reliably move or mirror portions of my data to other places.”



Research Computing HPC Cluster



Campus Home Filesysten



Lab Server

Personal Laptop

Desktop Workstation

XSEDE Resource

amazon web services | S3

red cloud

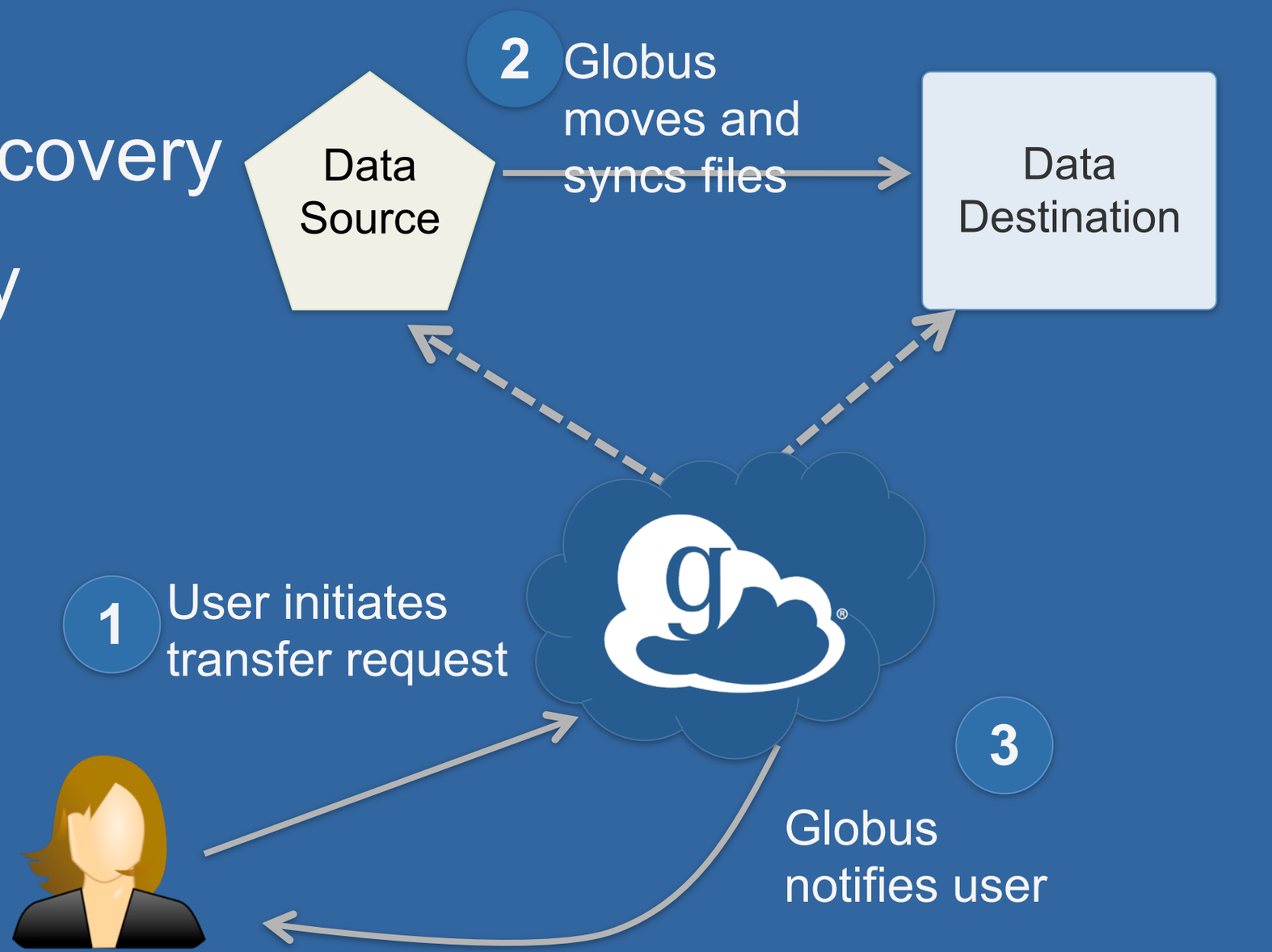
Windows Azure

Google Cloud Platform

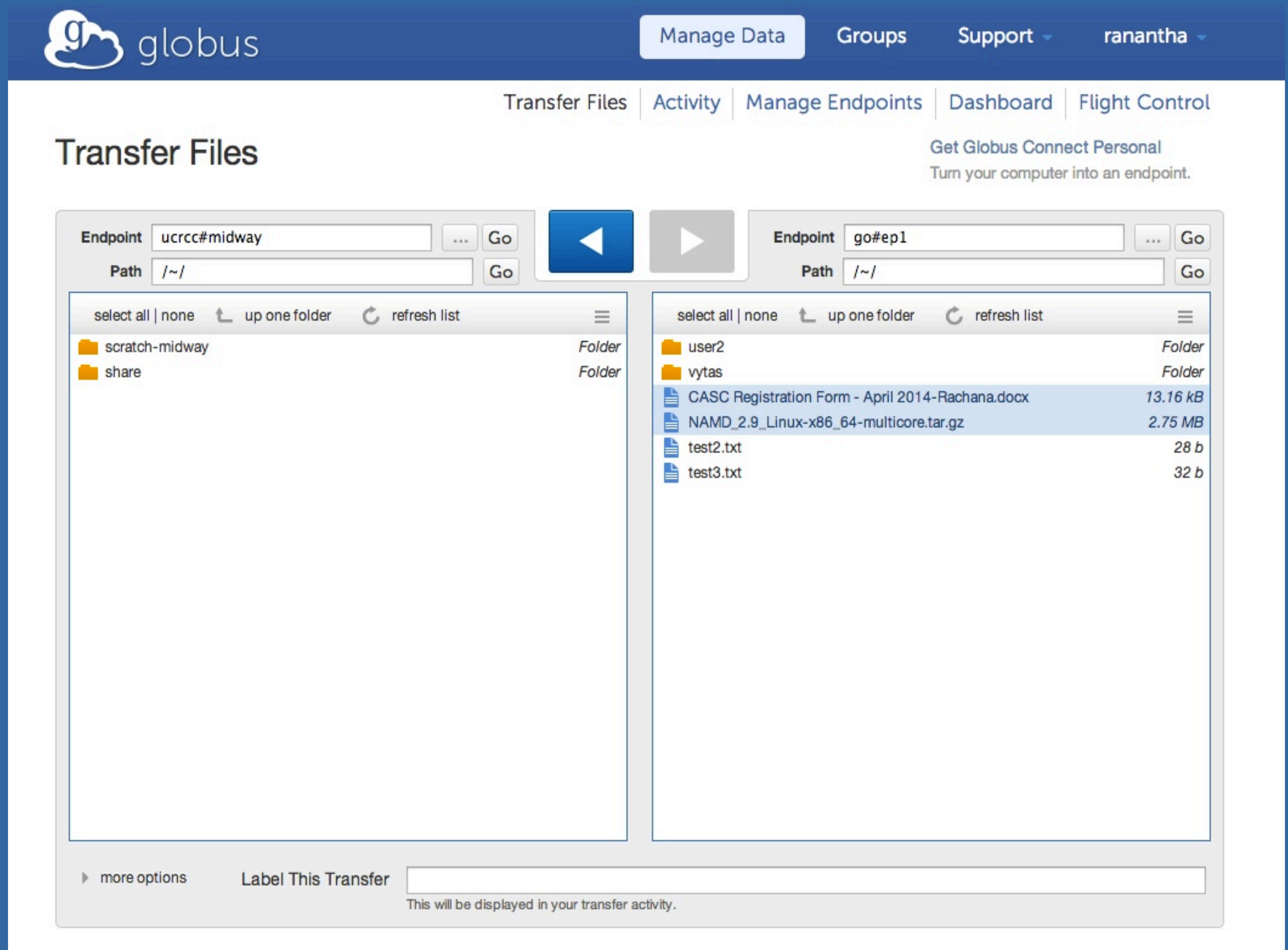
Public Cloud

Reliable, secure, high-performance ***file transfer*** and ***synchronization***

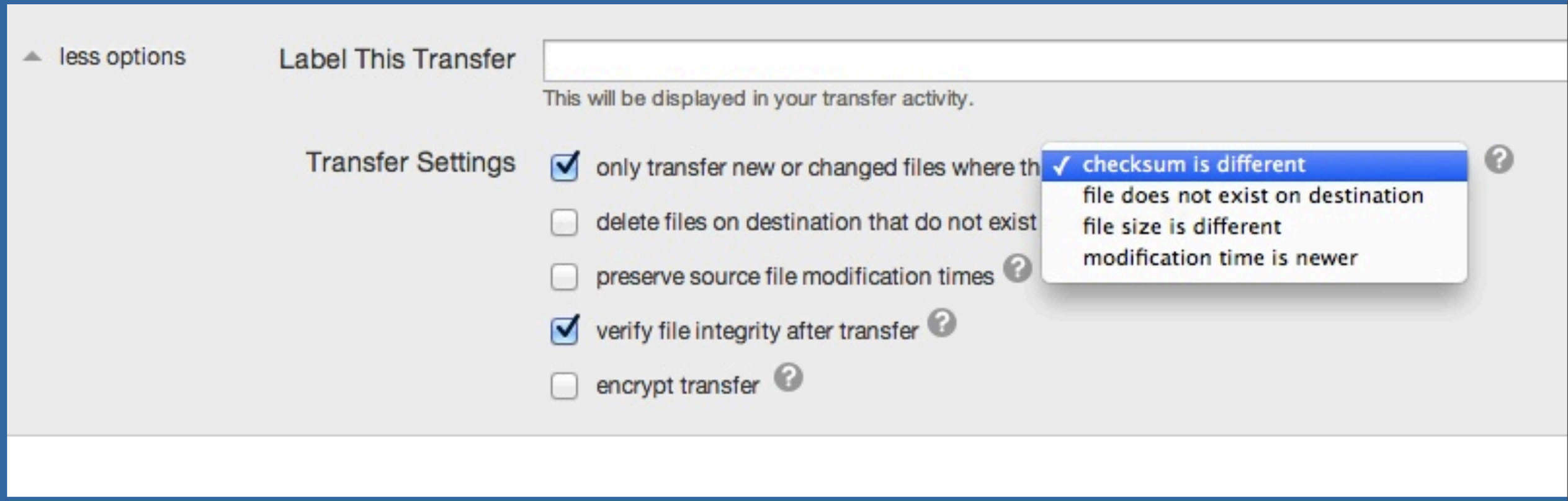
- "Fire-and-forget" transfers
- Automatic fault recovery
- Seamless security integration
- Powerful GUI and APIs



Transfer Files



Transfer Options

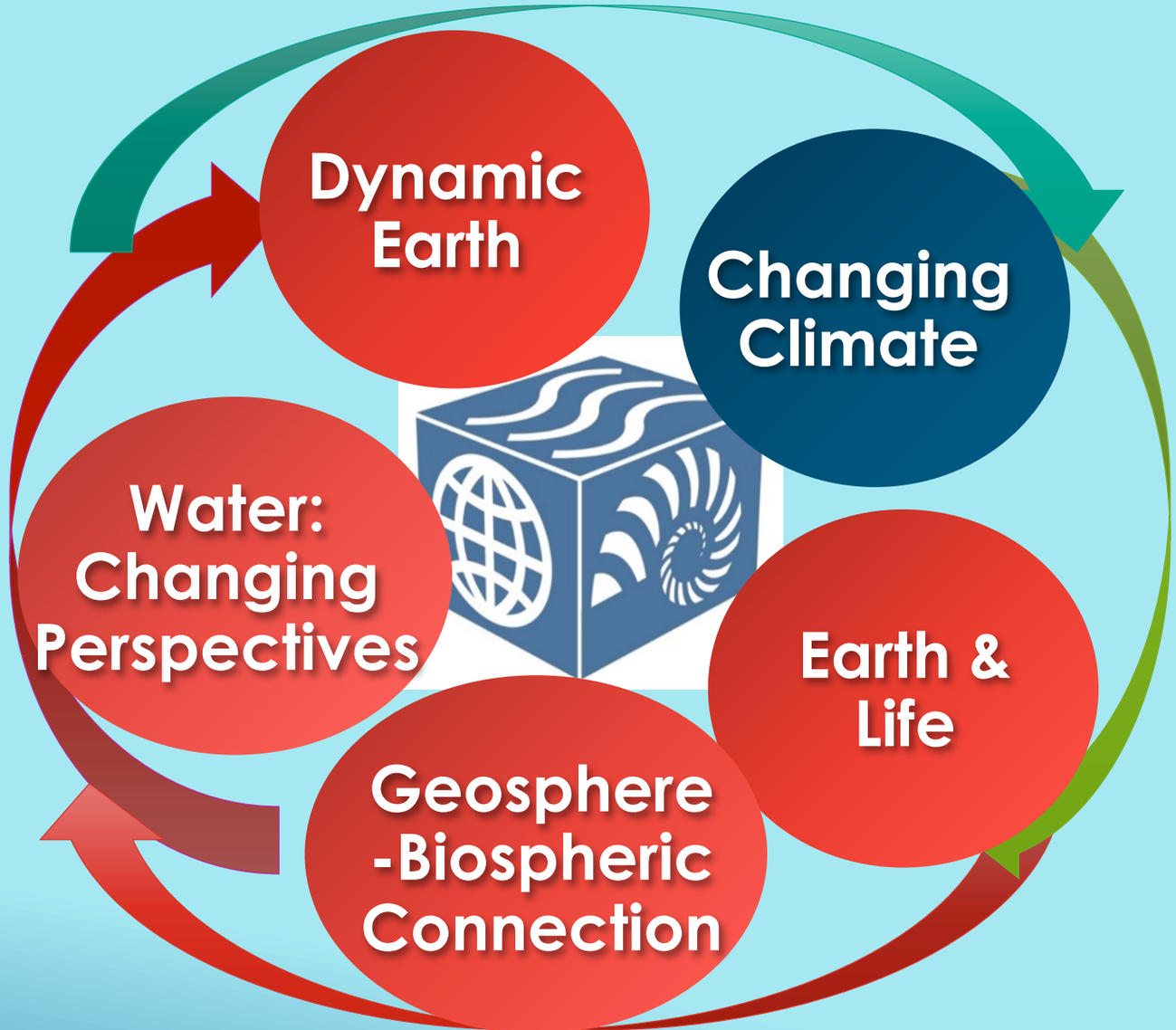


NSF's EarthCube Initiative

Courtesy of Rachael Black, Lee Allison, and others from the Arizona Geological Survey, serving EarthCube's Governance effort. Search for "EarthCube".

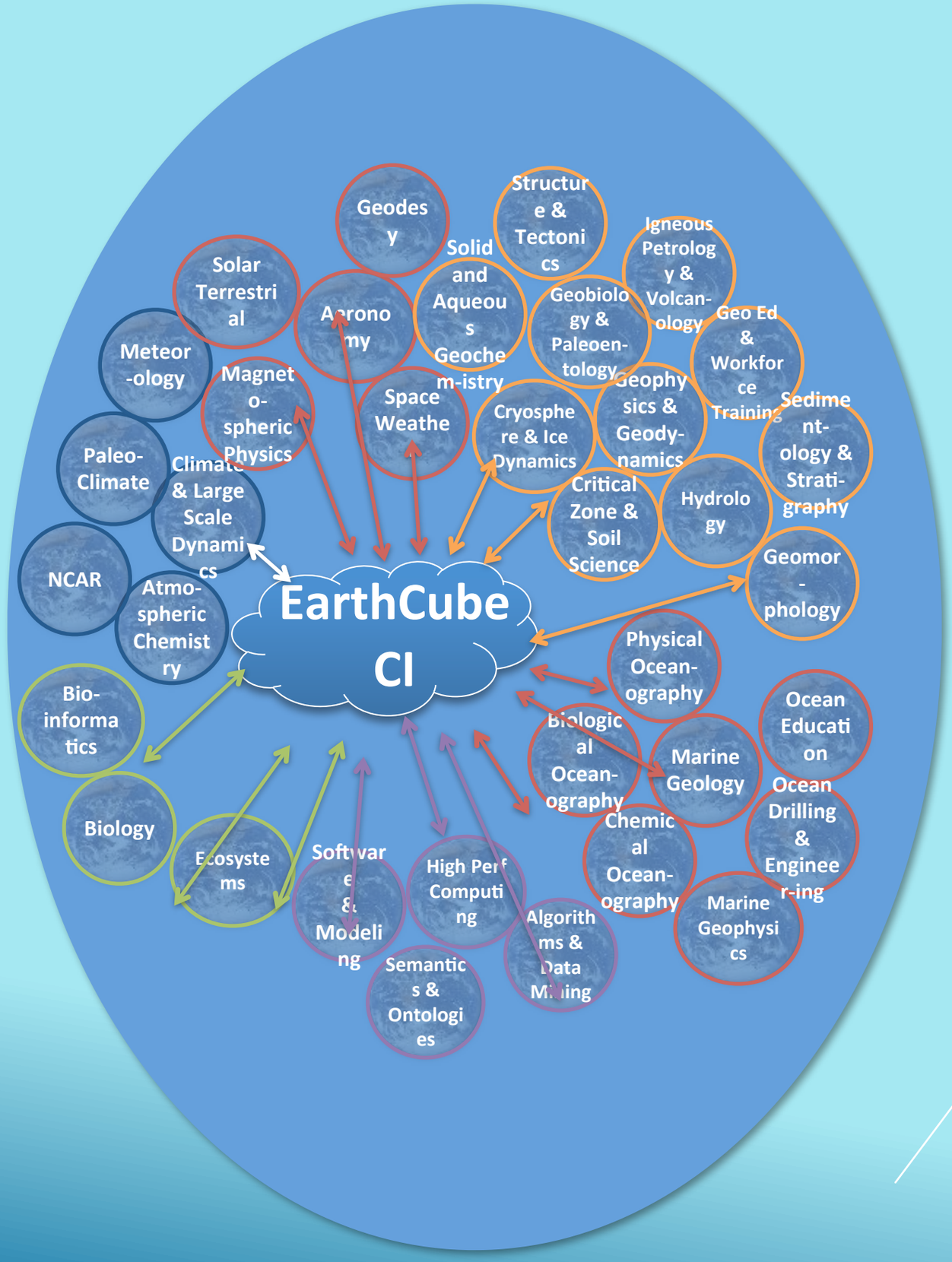
NCAR

EARTHCUBE VISION



- **Transform** the **conduct** of data-enabled geoscience-related **research**
- Create effective **community-driven cyberinfrastructure**
- Allow **global data discovery** and **knowledge management**
- Achieve **interoperability** and data integration **across disciplines**

Who is EarthCube?



Enables transformative geoscience by fostering a community committed to providing unprecedented discovery, access, and analysis of geoscience data.

Academic Geoscience Researchers in

- **Earth**
- **Oceans**
- **Atmosphere**
- **Polar**

The EarthCube

2011 Present

Dear Colleague Letter & Charrettes

2012

White Papers & Expressions of Interest *(Geo & CI)*

Roadmaps & Concept Designs *(technical roadmaps and small prototype designs)*

2013

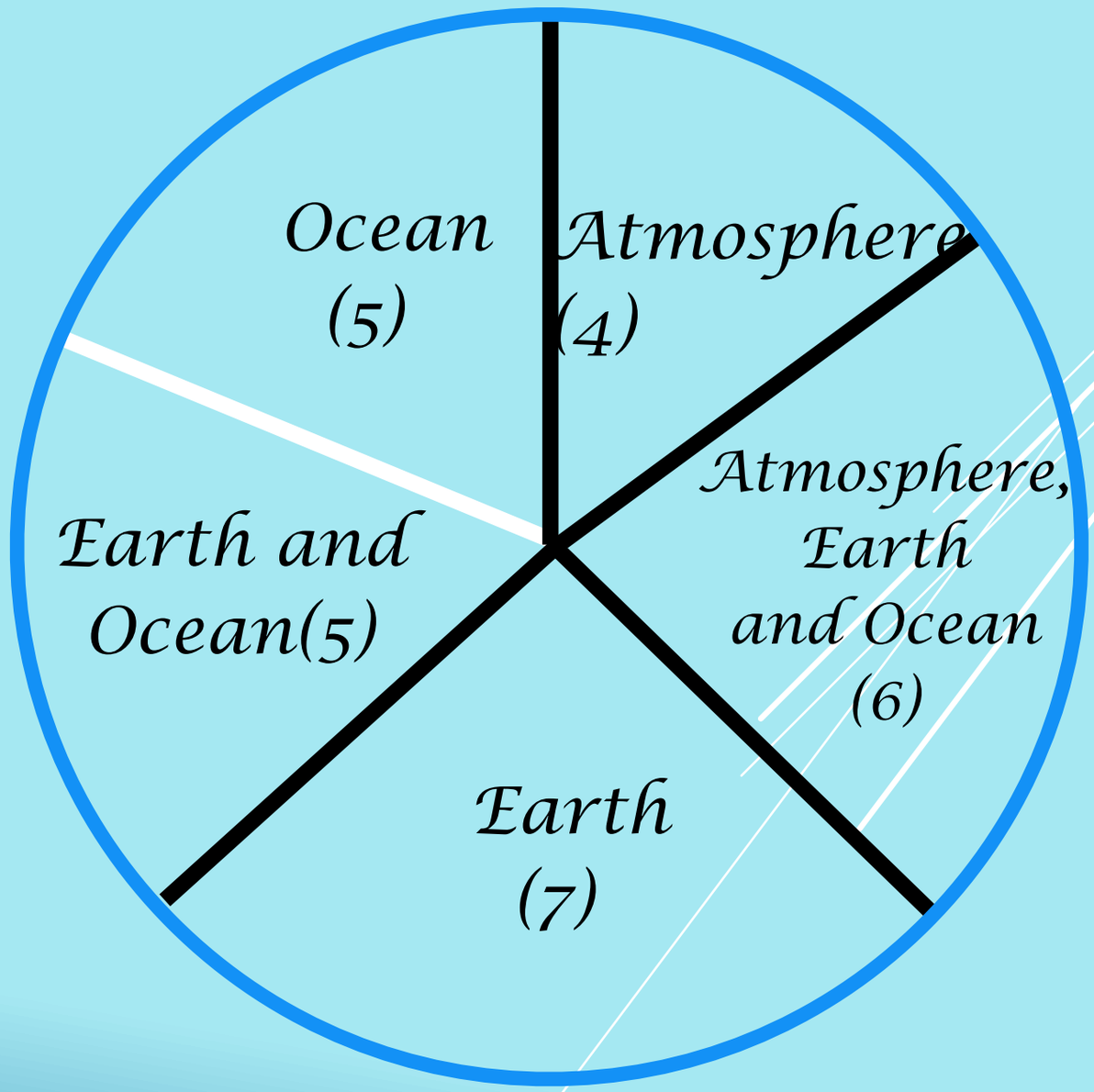
End-User Workshops & Stakeholder Alignment

Building Blocks, RCNs, and Conceptual Design Awards *(current funded projects)*

Full List of Awards at: http://earthcube.org/page/earthcube-

27 End-User Workshops: 2 pending

~2,000 participants, multiple agencies (NOAA, NASA, USGS, USDA, NRL, +)



Questions?

don@ucar.edu

NCAR

How we reach our users

Workshops

Hundreds of website examples

http://www.ncl.ucar.edu/Applications/

CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations) was launched on April 28, 2006 to study the impact of clouds and aerosols on the Earth's radiation budget and climate. It flies in formation with five other satellites in the international "A-Train" constellation for coincident Earth observations. The CALIPSO satellite comprises three instruments, the Cloud-Aerosol LIdar with Orthogonal Polarization (CALIOP), the Imaging Infrared Radiometer (IIR), and the Wide Field Camera (WFC). CALIPSO is a joint satellite mission between NASA and the French Agency, CNES

CALIPSO data are available as Level-2 and as part of the CFMIP-OBS (Cloud Feedback Model Intercomparison Program) a protocol to evaluate clouds in climate and weather prediction models based on satellite observations has been designed.

calipso_1.ncl: CALIPSO cloud fractions (low, middle, high, total and clear) for June 6, 2006.

The file used in this example was obtained from:
ftp://ftp.climserv.ipsl.polytechnique.fr/cfmip/goccp/3D_CloudFraction/grid_2x2xL40/

One-on-one correspondence with users

Active email lists

I am regridding a 12km by 12km lambert conical conformal grid of sparse emissions over the U.S. To a 1/2 by 2/3 degree grid over the same domain. If I choose the grid corners to be rounded to zero decimal places, I get a slightly different answer in my final emissions than if I round the grid corner to 2 decimal places which is actually slightly (1%) less than the original emissions. Why would this occur, and what is 'best practices' in choosing the grid corners for this scenario?