NSF NCAR's advanced computing, data services, software and educational resources for earth system science

Thomas Hauser

Director NSF NCAR Computational Information Systems Laboratory (CISL)

January 14, 2025

This material is based upon work supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1852977.

Goals

1. To build awareness of the many NCAR computing services, products, and resources available to help advance ***your*** science

2. To build connections between NCAR staff and the scientific communities that we support

Format and agenda

Format
- 5-minute lightning talks with 1-minute for questions
- ~15 minutes for questions for all at the end

Agenda

- High Performance Computing (HPC): Resources, Access, and Support (Daniel Howard)
- Data services (Doug Schuster)
- Community analysis tools and visualization services (Orhan Eroglu)
- Data Assimilation (Jeff Anderson)
- Machine Learning (Charlie Becker)
- Training and Communities (Katelyn FitzGerald)
- SIParCS (Katelyn FitzGerald)
- Q&A (all)

NSF NCAR Town Hall event website

- Slides
- Recordings
- Speaker bios
- More!

https://www.cisl.ucar.edu/events/ams2025

High Performance Computing (HPC): *Resources, Access, and Support*

Daniel Howard
Consulting Services Group

This material is based upon work supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1852977.

NCAR's High-Performance Computing, Data, & Analysis Resources: 2024

NCAR's High-Performance Computing, Data, & Analysis Resources: 2024

HPC Systems

2017 — 2023

SGI/HPE
*4032 Nodes, **145,152 Cores**, 313 TB total memory, **4.79 PFlop/s***
#21 Supercomputer in the world at debut

Cray/HPE
*2570 Nodes, **323,712 CPU Cores**, 680 TB total memory*
3.5X performance vs Cheyenne
328 NVidia A100 GPUs** provides 20% overall performance, ~**20 PFlop/s

2023

NWSC-4
???

Data Analysis & Visualization

+64 soon

***Casper:** heterogeneous system for data analysis & viz.*
- 75 High-Throughput Computing nodes
- 15 visualization nodes with accelerated graphics
- 20 dense GPU nodes for AI/ML, Code Development
- 4 nodes for Research Data processing
- 6 1.5TB large memory nodes

jupyterhub

http://projectpythia.org

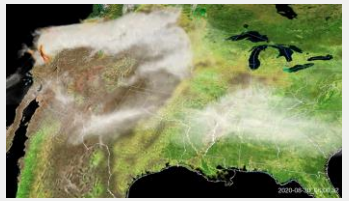
PROJECT PYTHIA

CISL develops specialized visualization software & services for Earth Science applications

GeoCAT

VAPOR
Visualization & Analysis Platform

https://geocat.ucar.edu



High Performance Storage

Campaign Storage
- **125PB** long-term, online storage
- ±17,464 hard drives
- 56 servers

Hierarchical Storage (HSM) Data Migration

Quasar Tape Library
- **35PB long term archival** storage
- 22 IBM TS1160 tape drives
- 1774 20TB tape cartridges
- 216 hard drives
- 2PB disk cache
- 5 data mover servers

Derecho 'scratch' Storage
- **55PB** short-term storage
- Principally supports HPC jobs
- 5,088 hard drives
- 28 servers

Stratus Object Storage
- **5PB object storage**
- 588 hard drives
- 6 servers

NSF NCAR HPC Access & Support Services

Compute and Storage Resource Allocations - https://ncar.pub/allocations

- NSF NCAR compute and storage resources are available to US-based university researchers in Earth System Science domains. The process depends on the scope of resource requirements:
  - **Exploratory Allocation**: Graduate students, post-doctoral researchers, or new faculty member at a U.S. university can request a one-time allocation at each career stage.
  - **Small Allocation**: Useful to complete small projects or to conduct initial runs in preparation for submitting a request for a large allocation.
  - **Data Analysis & Classroom Allocations**: are available to provide access to our extensive data holdings with minimal compute requirements.
  - **Large Allocation**: Proposals accepted every six months in March and September. Our HPC Allocations Panel reviews requests in April and October, and projects begin in May and November. No upper limit, within predefined resource capacity constraints.

Support Forums

- NSF NCAR's Consulting Services Group maintains **HPC user documentation**, fields **support tickets**, provides **virtual consulting services**, and maintains a **consultant on duty** during business hours to monitor and triage issues with the HPC systems.
- Users should join **NCAR HPC Users Group (NHUG) & Slack space** where the community collaborates on timely issues, seeks peer feedback, & networks with dedicated channels for `#derecho-users`, `#casper-users`, `#jupyterhub-users`, etc.
- **User Trainings, Software Hackathons, &** Support for **Educational/Workshop Uses** of HPC Resources

Allocation Types & Application Process

Award Type	Resource HPC limit	Frequency	Funding eligibility
Large	No upper limit (subject to availability)	Twice Annually (March & September)	NSF award required
Small	1M CPU / 2.5K GPU hours (Derecho)	Continuous	
Exploratory & Classroom	0.5M CPU / 1.5K GPU hours (Derecho)		N/A
Data Analysis	(Casper Only)		Any funding source

https://ncar.pub/allocations

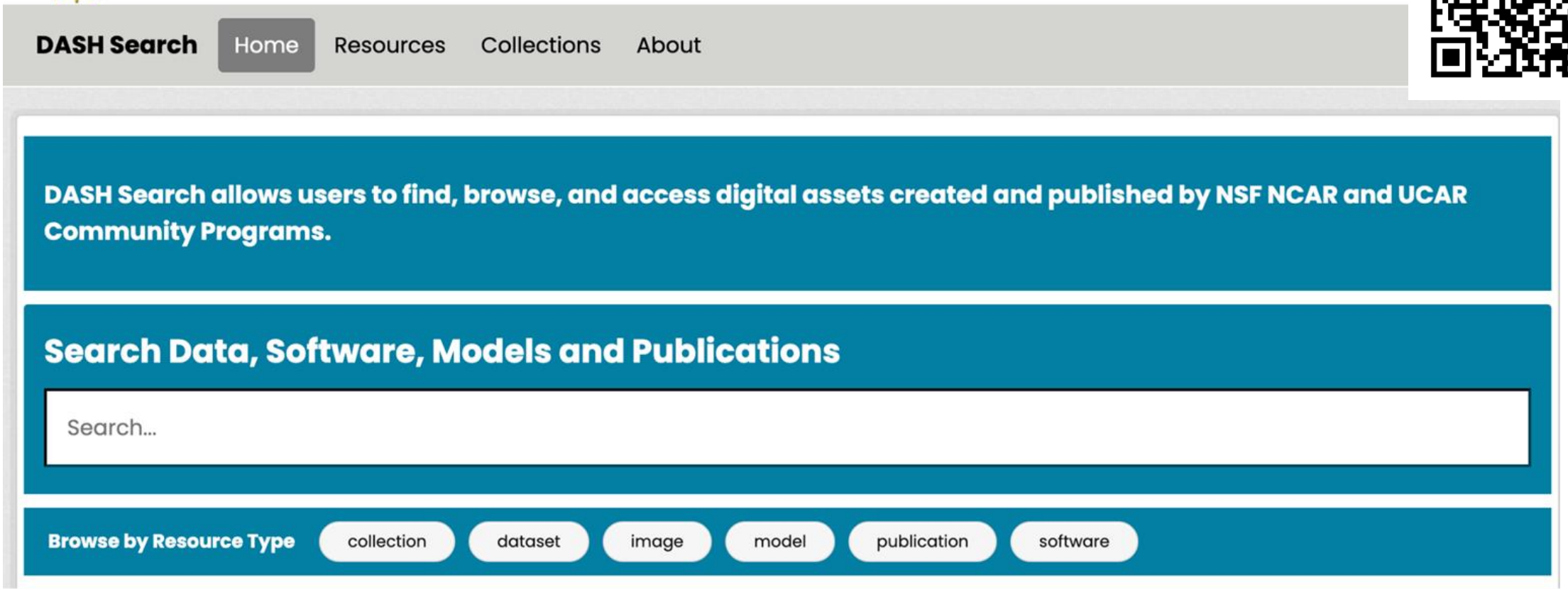
Data Services

Doug Schuster

This material is based upon work supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1852977.

Search NSF NCAR Data Catalog (DASH Search)

https://data.ucar.edu



GDEX Geoscience Data Exchange

Climate Data Gateway at NCAR

RDA Research Data Archive NCAR

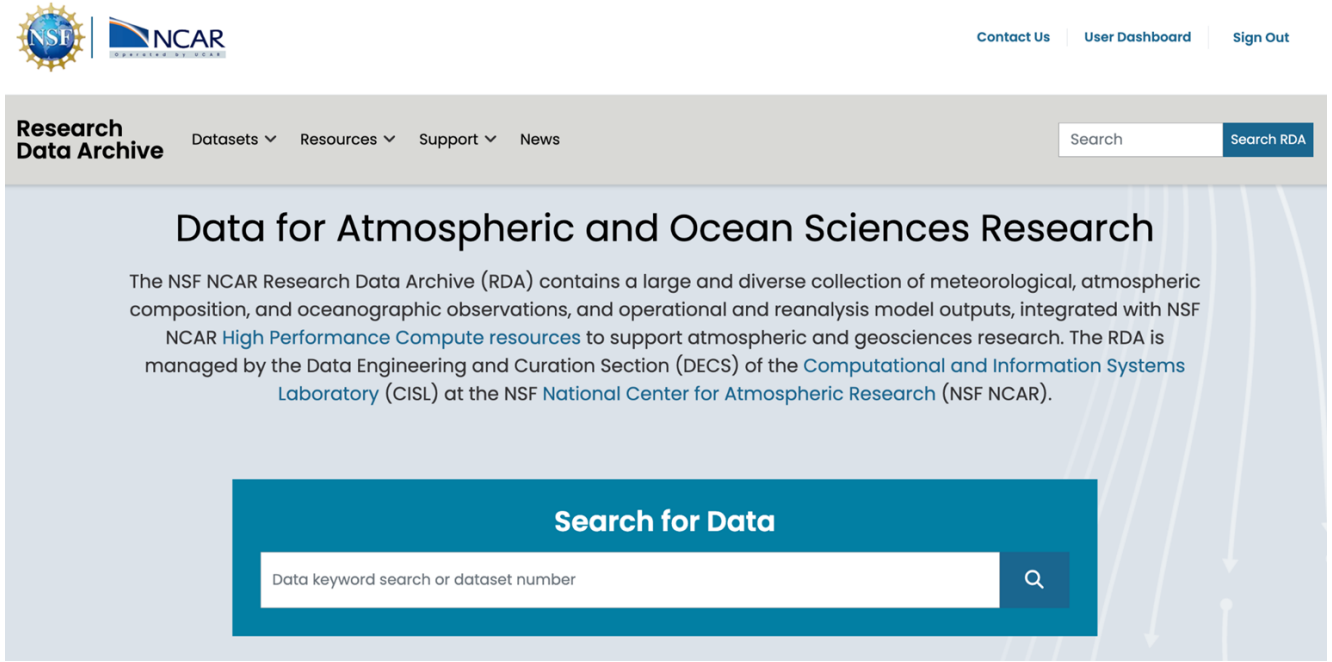
EOL data Earth Observing Laboratory

MAUNA LOA SOLAR OBSERVATORY

NSF NCAR Research Data Archive - Remote Access and Data Proximate Compute

- **History**
  - Established in the 1960s
- **Purpose**
  - Support atmospheric and ocean sciences research at NSF NCAR and US universities with community reference datasets
- **Collections**
  - Ocean & atmospheric observations, analyses, reanalyses, climate simulation outputs
  - 700+ datasets, 29M files, 9 PB
  - 70+ datasets growing daily-monthly
- **Free and open access**
  - 28,000+ Unique Web users in FY 2024
  - 22 PB Data Delivered in FY 2024
- **Science educated staff**

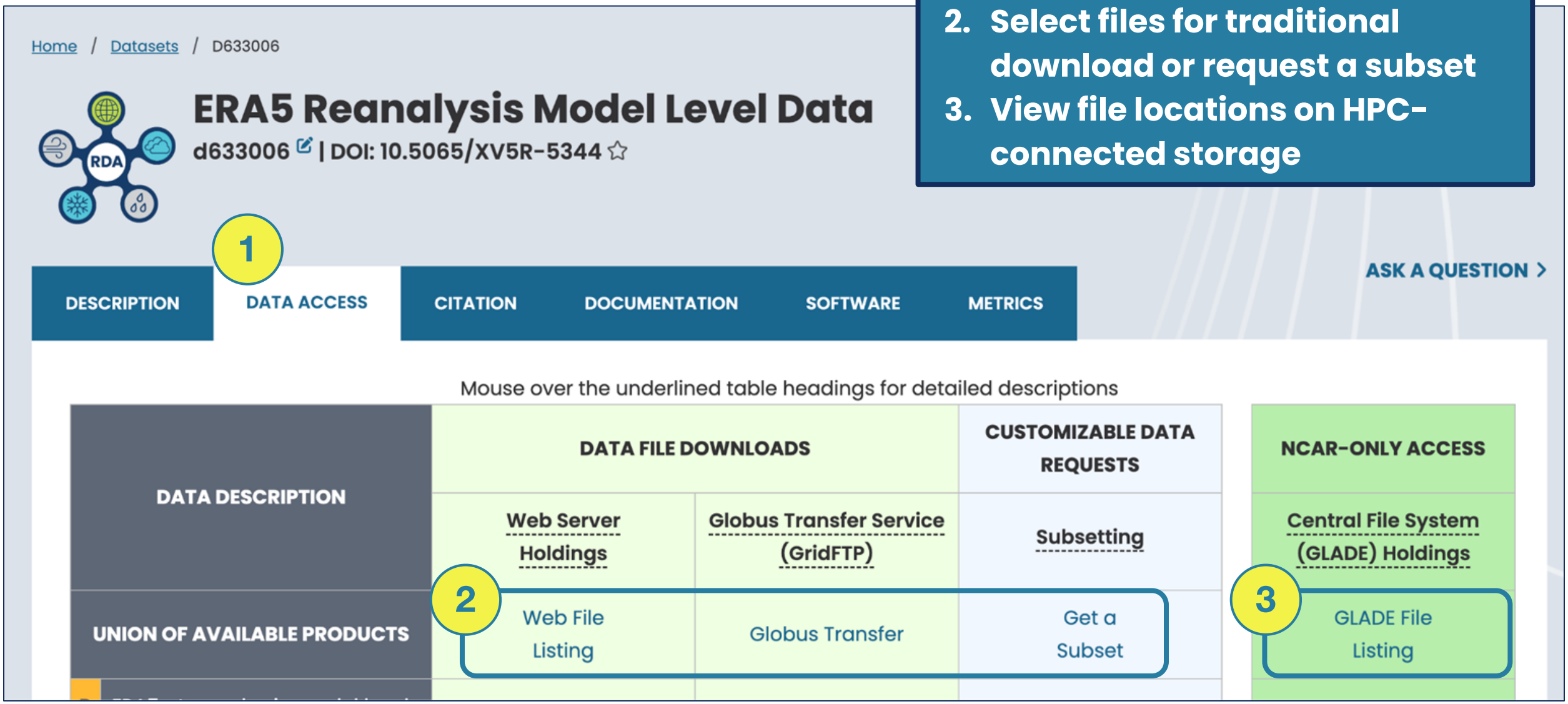
https://rda.ucar.edu



NSF NCAR Research Data Archive - Remote Access and Data Proximate Compute

https://rda.ucar.edu

1. Click on Data Access tab
2. Select files for traditional download or request a subset
3. View file locations on HPC-connected storage



NSF NCAR Research Data Archive - Data Proximate Compute

RDA Research Data Archive

NSF NCAR

rda.ucar.edu

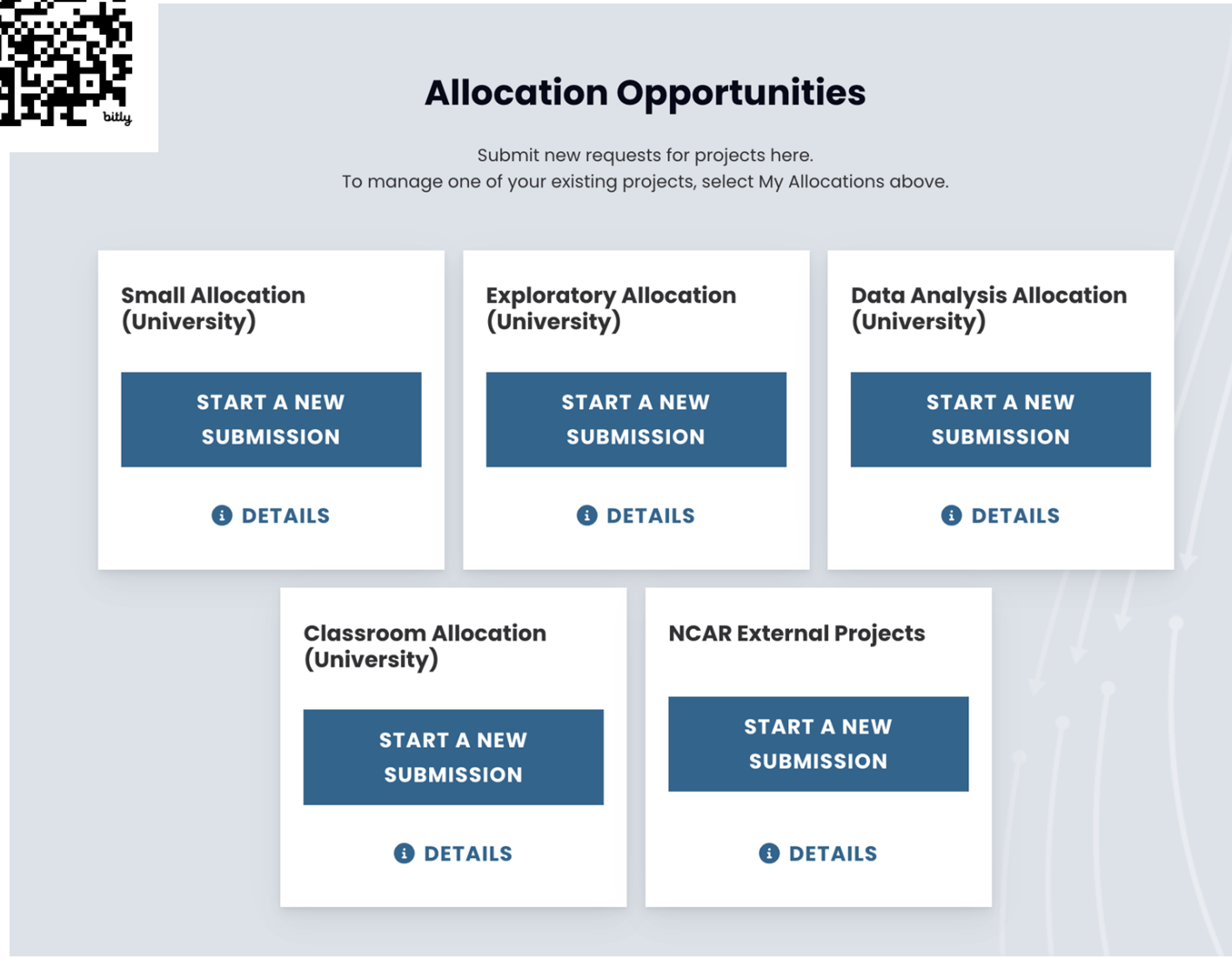
- **Popular Datasets**
  - **NCEP Products**
    - ***Final Analysis, GFS, CFSR, CFS, In Situ Obs***
  - **ECMWF products**
    - ***ERA-5 0.25 degree and model resolution***
  - **JMA products**
    - ***JRA-55, JRA-3Q***
- **PI/NCAR produced datasets**
  - **USGS CONUS 404, CESM2 Large Ensemble**

Data Proximate Compute at NSF NCAR

Data Analysis Allocation

Available to researchers from any eligible institution regardless of the source of funding for the planned analysis. Researchers are only required to identify the specific NCAR-hosted data sets that are essential to completing their science objectives.

https://arc.ucar.edu/xras_submit/opportunities

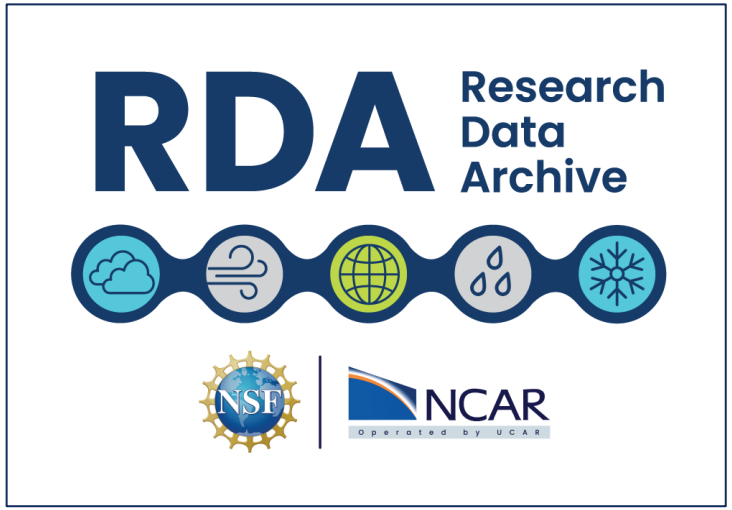


Questions and support

Questions?

rdahelp@ucar.edu

schuster@ucar.edu



Community Analysis Tools and Visualization Services

Orhan Eroglu, Nihanth Cherukuru
Visualization and Analysis Software Team

This material is based upon work supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1852977.

The GeoCAT (Geoscience Community Analysis Toolkit) Team and Projects

"Creating scalable data analysis and visualization tools for Earth System Science data to serve the geosciences community"


Orhan Eroglu


Anissa Zacharias


Cora Schneck


Julia Kent


Katelyn FitzGerald


Philip Chmielowiec

Projects:

GeoCAT

PROJECT raijin

Packages:

GeoCAT-comp
GeoCAT-applications
GeoCAT-examples
GeoCAT-viz

uxarray

GeoCAT

Contact GeoCAT

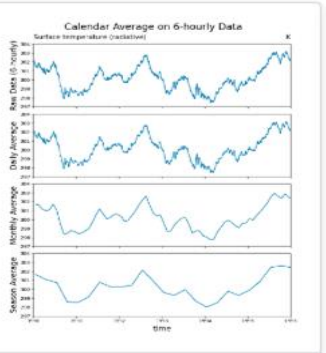
NCAR | GeoCAT

GeoCAT Software Products

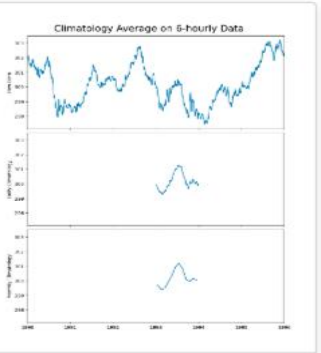
GeoCAT-comp: Analysis operators

Usage Examples

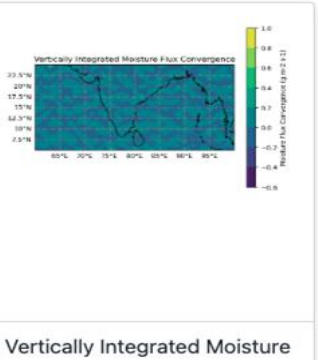
Here's some examples of how to use geocat-comp.



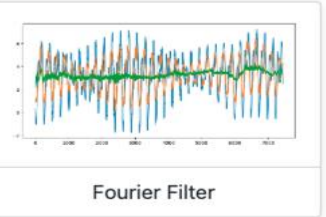
Calendar Averaging



Climatology Average



Vertically Integrated Moisture Flux Convergence



Fourier Filter

GeoCAT-applications: Further support for Pivot-to-Python

GeoCAT Applications

GeoCAT Applications is a community resource managed by the GeoCAT team. Inspired by the NCL Applications page, this is designed to be a quick reference demonstrating capabilities within the Scientific Python Ecosystem that may be relevant to your geoscience workflows.

New to Python or GeoCAT Applications? Check out the Getting Started guide!

Python Examples

Dates and Times

- Working with Date and Time

Data Analysis

- General Applied Math
- Calculating Climatologies
- Humid Heat Metrics

Regridding

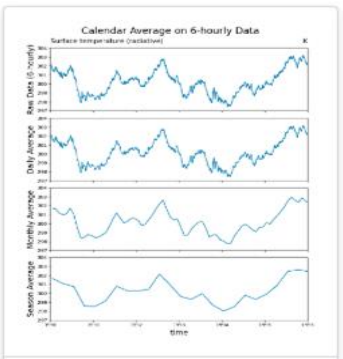
Plot Types

GeoCAT Software Products

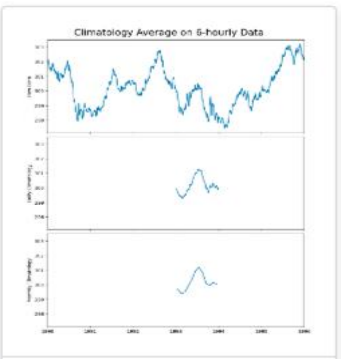
GeoCAT-comp: Analysis operators

Usage Examples

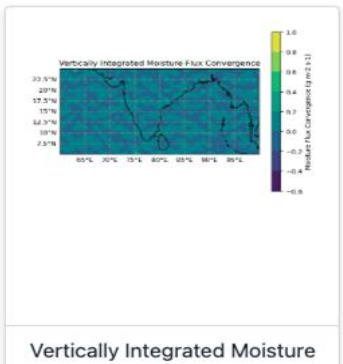
Here's some examples of how to use geocat-comp.



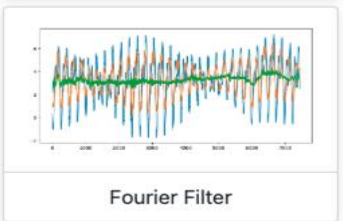
Calendar Averaging



Climatology Average



Vertically Integrated Moisture Flux Convergence



Fourier Filter

GeoCAT-applications: Further support for Pivot-to-Python

NCL to Python

NCL Index

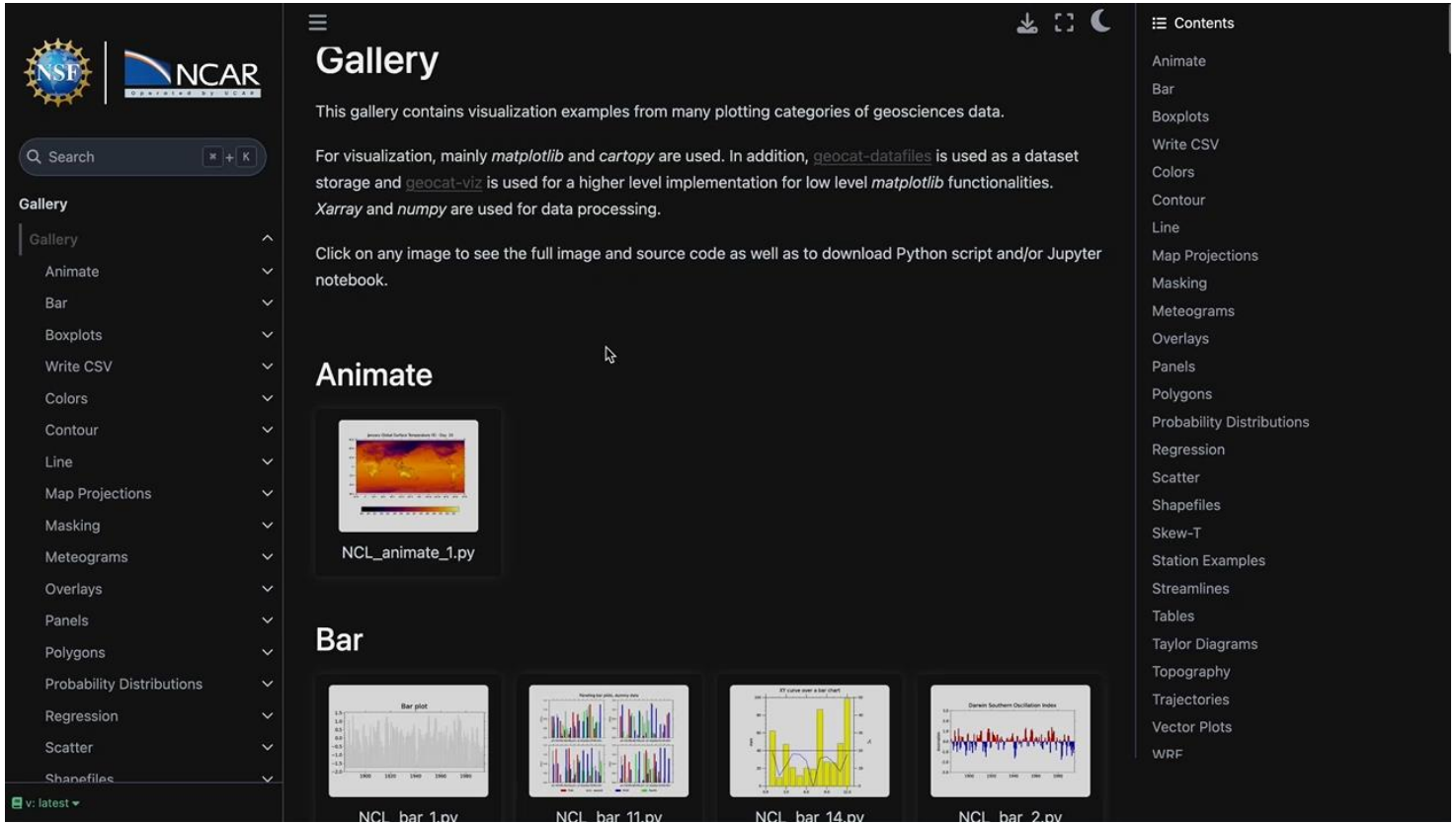
NCL Applications

NCL Index

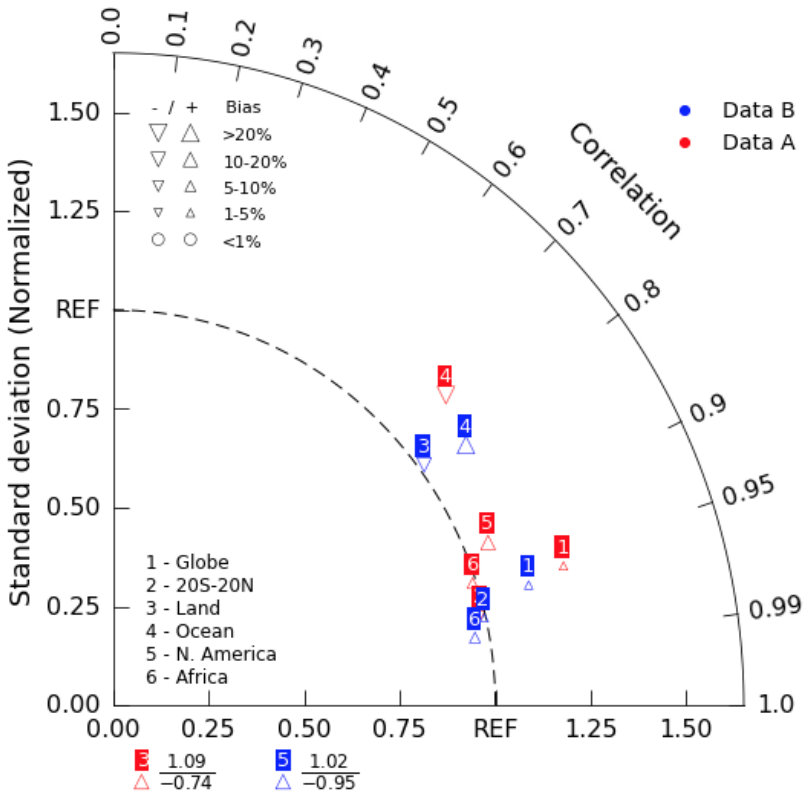
NCL Function	Description	Python Equivalent	Notes
abs	Returns the absolute value of numeric data	`abs()` or `numpy.abs()`	example notebook
asin	Computes the inverse sine of numeric types	`math.asin()` or `numpy.arcsin()`	example notebook
acos	Computes the inverse cosine of numeric types	`math.acos()` or `numpy.arccos()`	example notebook
atan	Computes the inverse cosine of numeric types	`math.atan()` or `numpy.arctan()`	example notebook
atan2	Computes the inverse tangent of (y/x) for numeric types	`math.atan2()` or `numpy.arctan2()`	example notebook
avg	Computes the average of a variable regardless of dimensionality	`numpy.average()` or `numpy.mean()`	example notebook
cos	Computes the cosine of numeric types	`math.cos()` or `numpy.cos()`	example notebook
cosh	Computes the hyperbolic cosine of	`math.cosh()` or `numpy.cosh()`	example

GeoCAT Software Products

GeoCAT-examples: Visualization Gallery



Special plots, e.g.



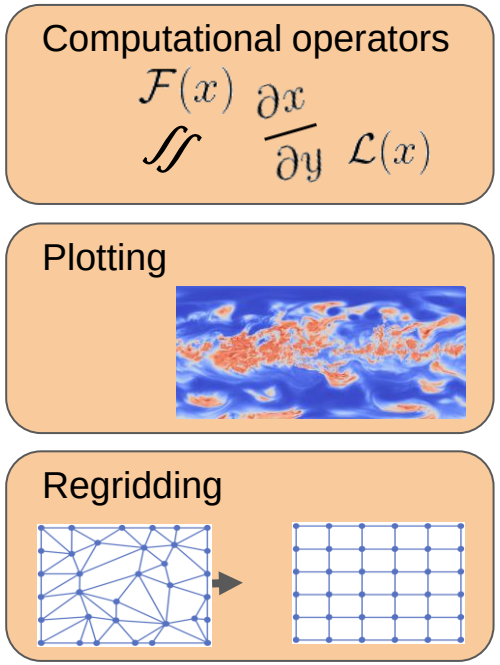
Taylor diagrams

UXArray : Community Geoscience Analysis Tools for Unstructured Mesh Data

Next generation, kilometer-scale climate and global weather models

After nearly two decades of development and evaluation, the climate and global weather modeling communities are transitioning from simple structured grids to more complex, but more scalable **unstructured grids** upon which governing equations of state are solved.

1. Develop extensible, scalable, open source software for data analysis and visualization on unstructured grids



2. Sustainable and community owned

Visualization Services and Research (ViSR)

"Advanced Visualizations for Data Analysis and Communication"

Nihanth Cherukuru

Stas Jaroszynski

Matt Rehme

Scott Pearse

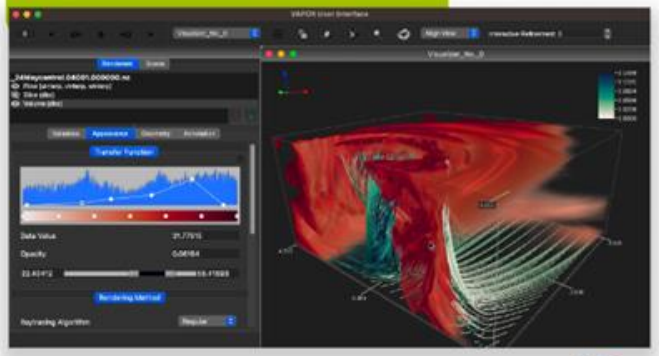
Tammy Zhang

Ian Franda

Visualizations for **Communication**

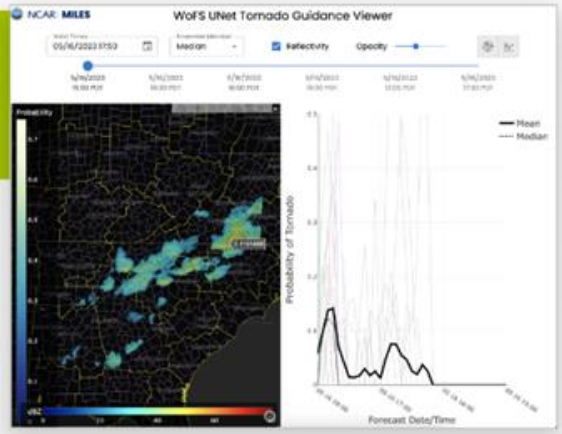
MPAS "Aqua-Planet"
Visualization: Matt Rehme
Science: Rosimar Rios-Berrios

Visualization Tools for **Discovery**



VAPOR is available both as a desktop graphical application and a python library

Visualizations for **Decision Making**

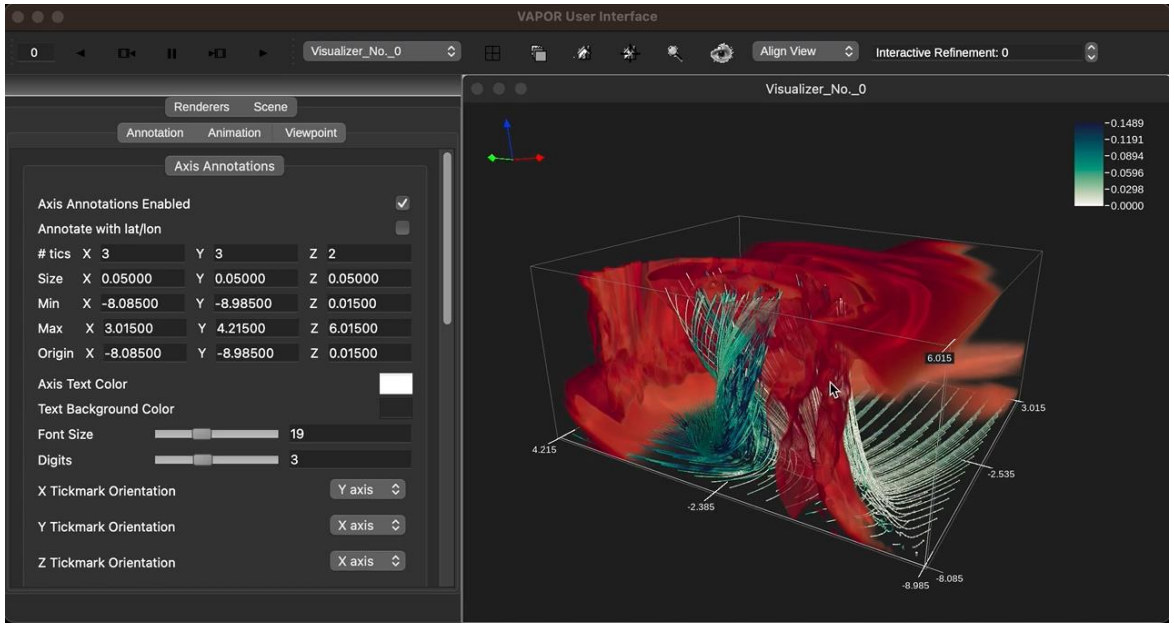


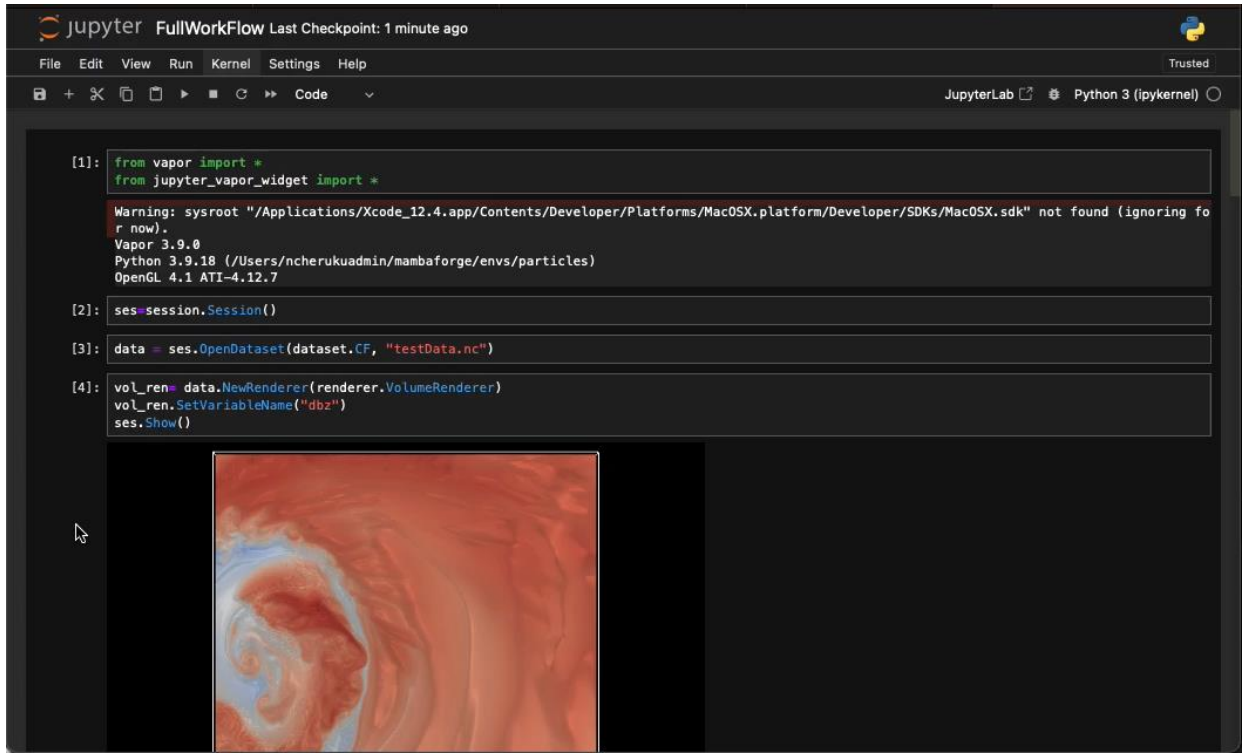
Tornado guidance dashboard
Visualization: Tammy Zhang
Science: David John Gagne

WRF Fire, East Troublesome Wildfire
Visualization: Scott Pearse
Science: Juliano T, DeCastro A, Kosovic B, Edgley C

Made with VAPOR

- Easy-to-use 3D visualization software
- Supports most of the commonly used data formats in geosciences
- Serving community ~20 years
- Open Source Software
- **Graphical Interface** and **Python library**





Learn more about VAPOR

Visualization Services and Research (ViSR)

We are always eager to explore new collaboration opportunities!

- We collaborate with researchers to design **production-quality visualizations**
- We develop and maintain **VAPOR** - A 3D data visualization application
- We design & develop **web interactive visualizations** for the broader audience
- We actively practice **UX methodologies** to understand societal needs and evaluate emerging technologies for visualization

Learn more about ViSR at the Visualization Gallery

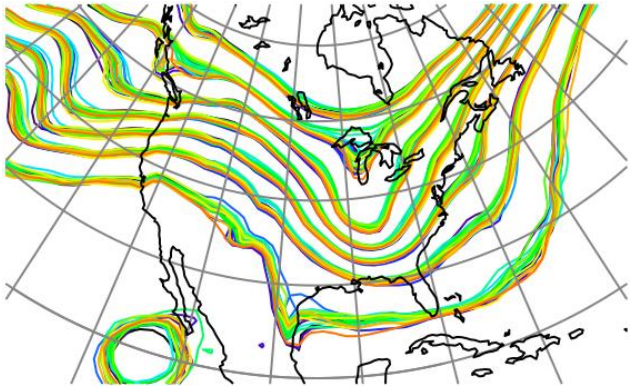
CM-1 Simulation of a precipitating cumulus-congestus cloud
***Visualization**: Matt Rehme*
***Science**: Hugh Morrison, Kamal Kant Chandrakar*

Contact: Nihanth Cherukuru
ncheruku@ucar.edu

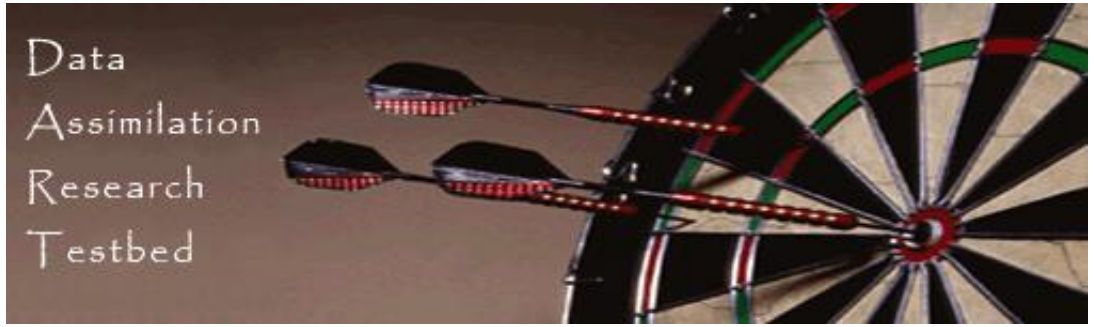
Data Assimilation

Jeff Anderson
Data Assimilation Research Section

This material is based upon work supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1852977.



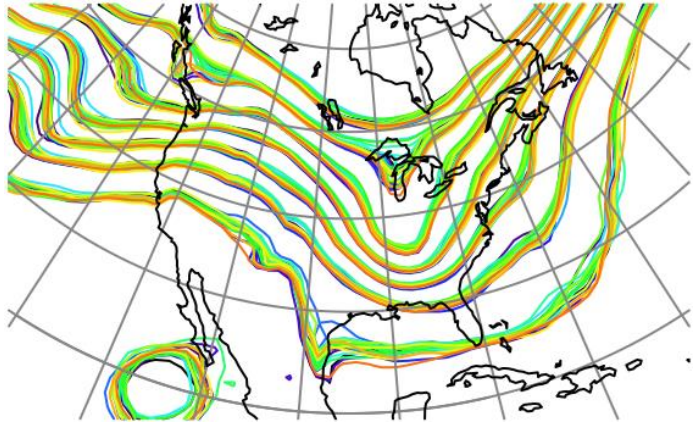
NCAR's Data Assimilation Research Section: Data Assimilation Tools for Earth System Science



What is Data Assimilation?

Observations combined with a Model forecast…

+



…to produce an analysis
(best possible estimate).

The Data Assimilation Research Section (DAReS)

Mission: To accelerate progress in Earth System Science at NCAR, UCAR Universities, and in the broader science community by providing state-of-the-art ensemble DA capabilities.

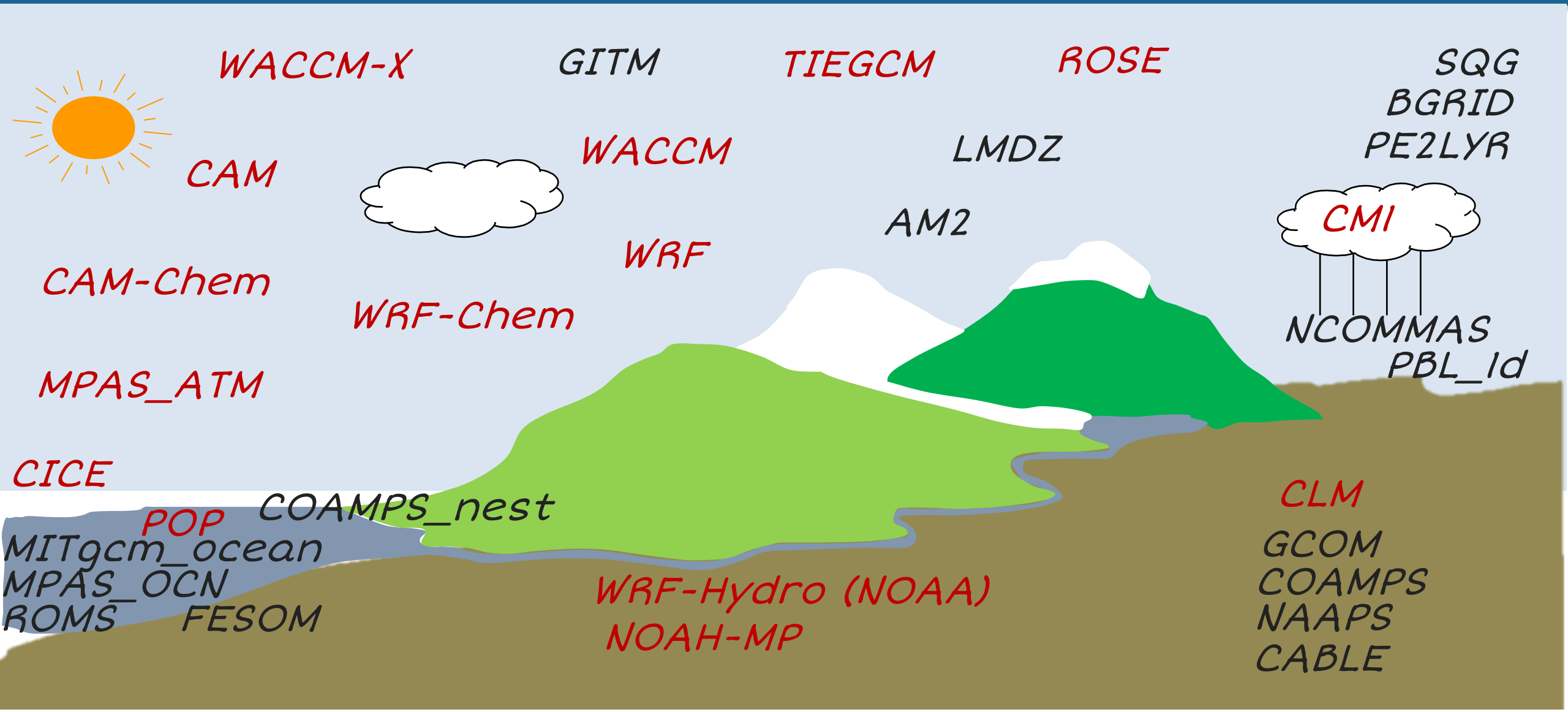
Method: DAReS develops and maintains the Data Assimilation Research Testbed (DART), a community facility for ensemble data assimilation.

Data Assimilation Research Testbed (DART)

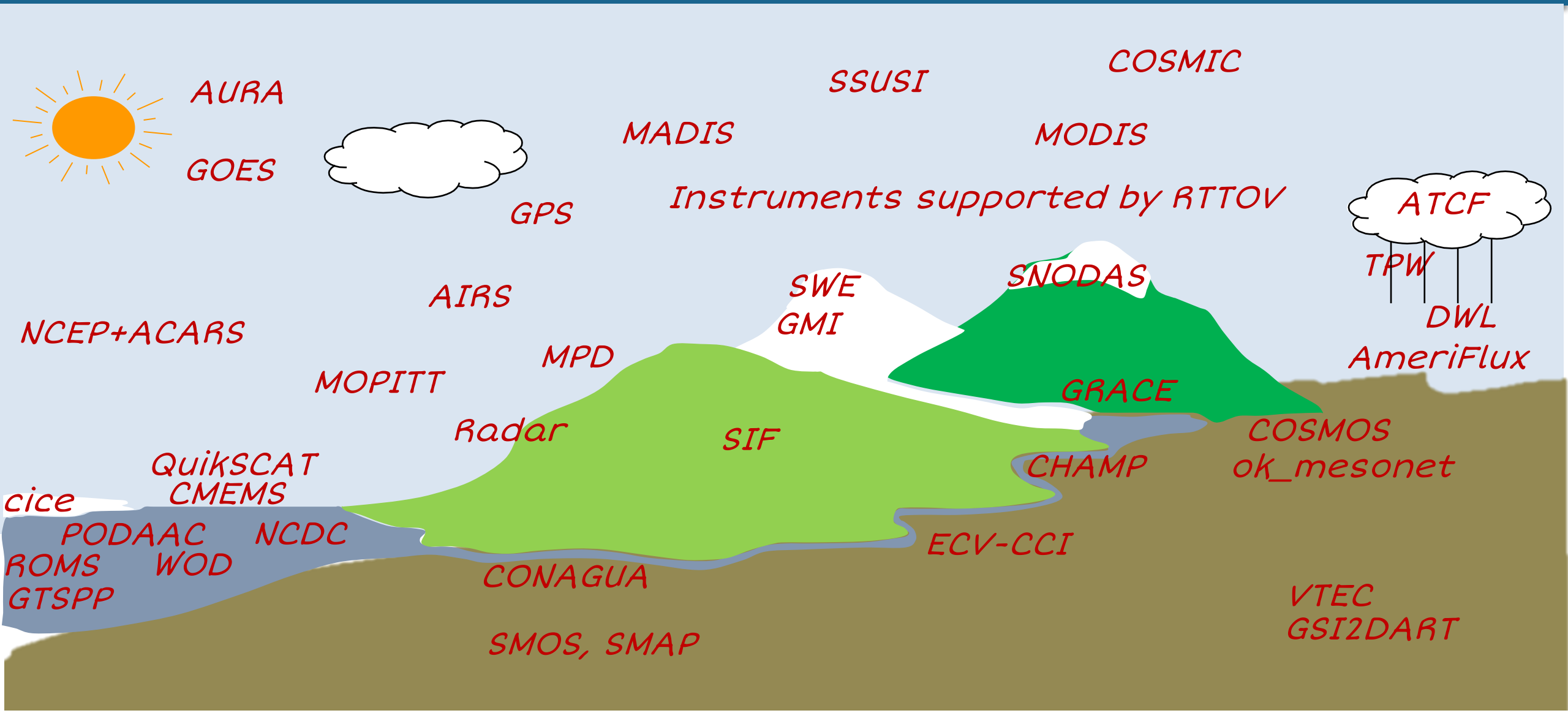
- A state-of-the-art Data Assimilation System for Geoscience
- A Data Assimilation Research System
- Professional software engineering
- People: The DAReS Team

- State-of-the-art ensemble DA is essential for research on:
  - Prediction
  - Predictability
  - Model development / parameter estimation
  - Observation system design / evaluation

Geophysical Models Interfaced with DART



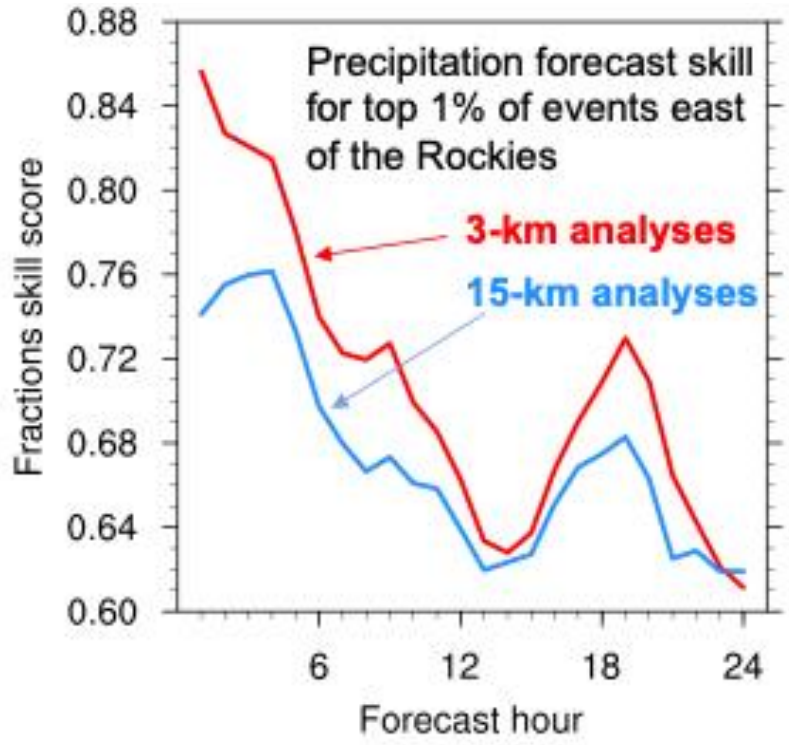
Supported Earth System Observations (others available)



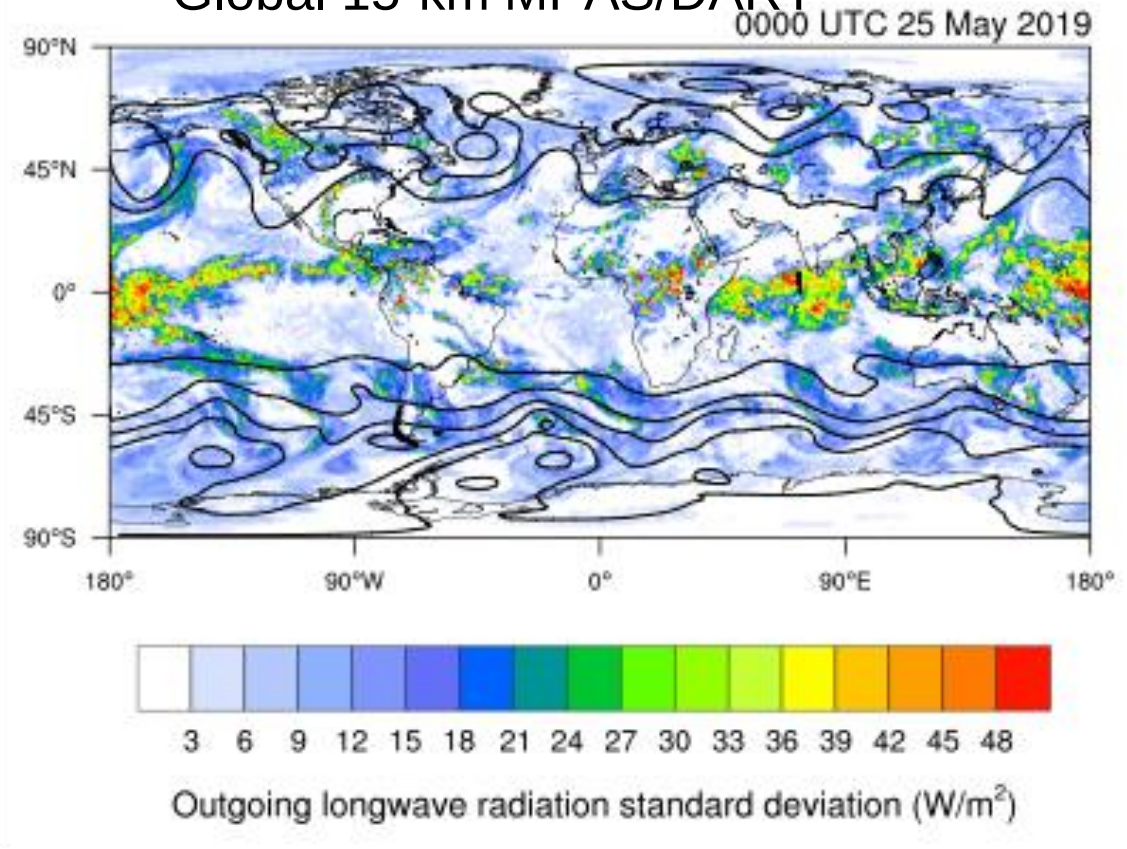
Toward Global Convection-Permitting Data Assimilation

- Current capabilities with 80 ensemble members:

Regional 3-km WRF/DART



Global 15-km MPAS/DART

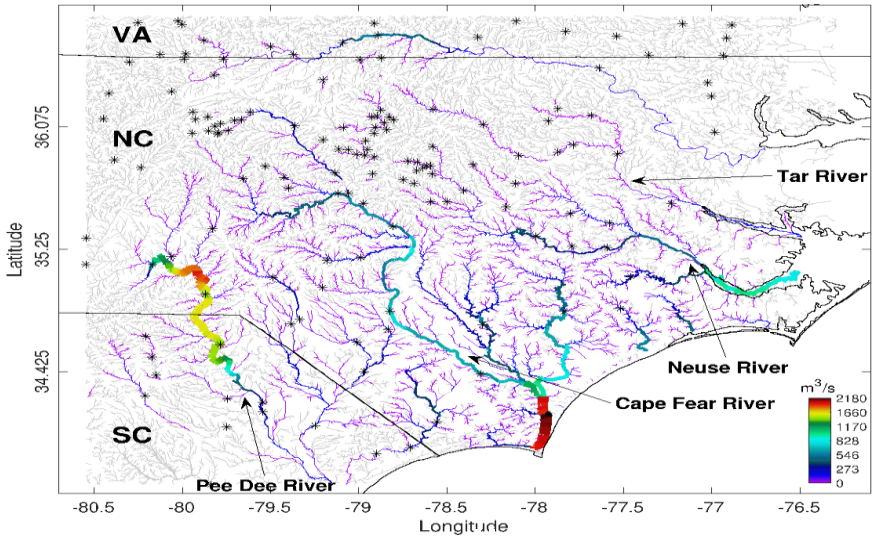


Gradual approach toward global convection-permitting ensemble-based DA

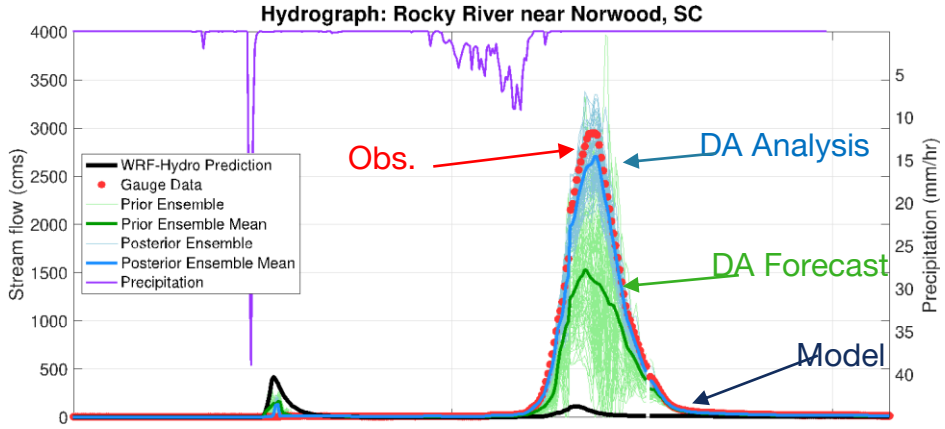
Variable-resolution meshes ⟶ "Dual-resolution" DA ⟶

Flood Prediction: WRF-Hydro/DART for Hurricane Florence 2018

High-resolution stream network with USGS streamflow gauges.

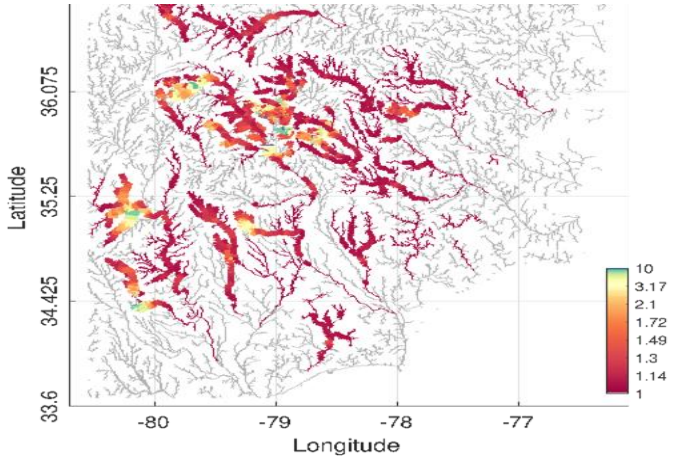


DA greatly improves analysis and forecasts of streamflow.

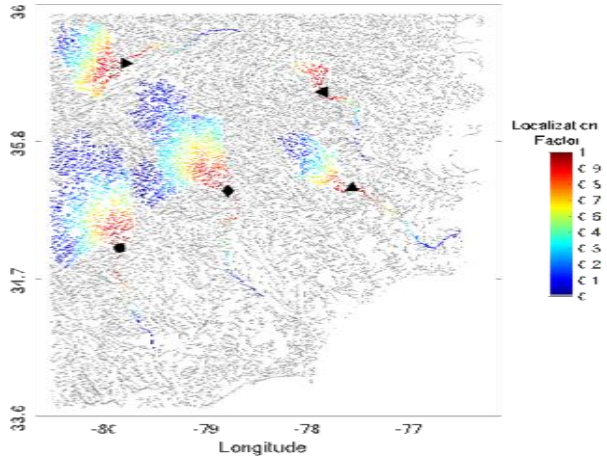


Novel Data Assimilation Science

1. Prior and Posterior Adaptive Covariance Inflation

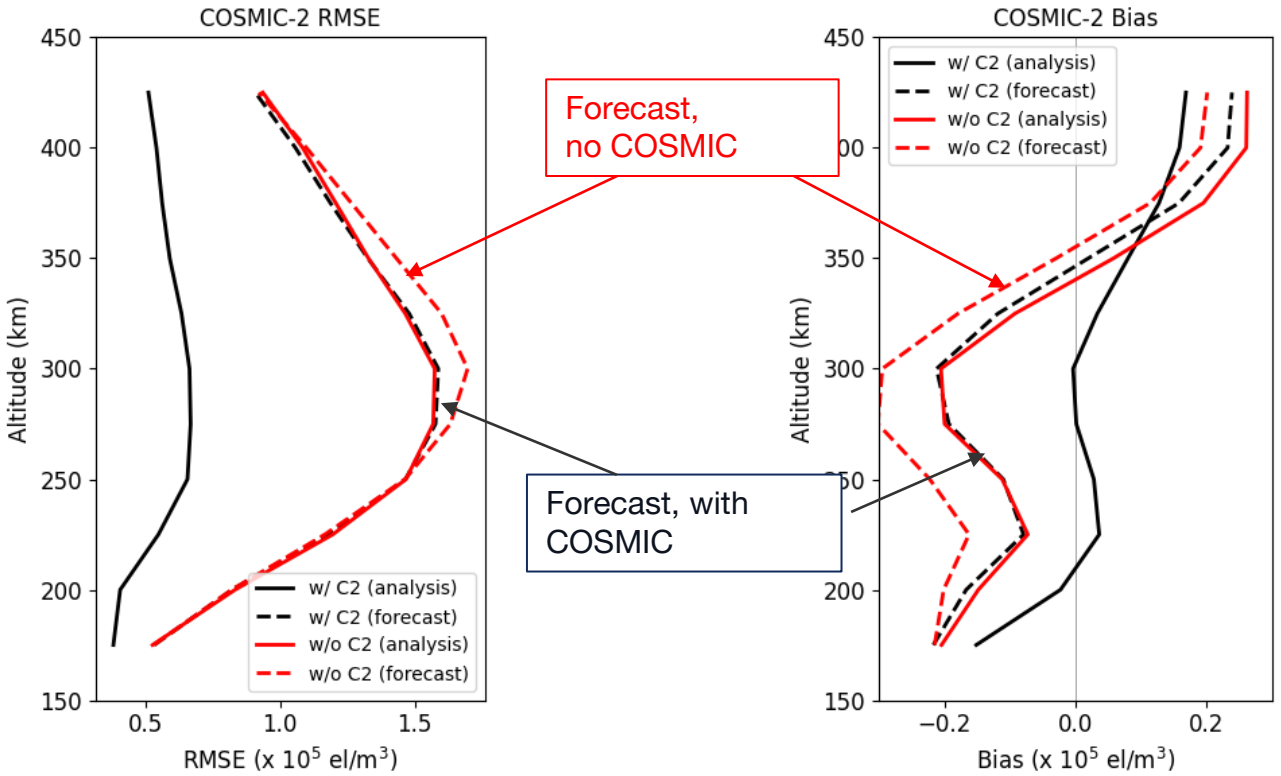


2. Along-The-Stream (topology-based) Localization



Space Weather: Data Assimilation Applied to Earth's Upper Atmosphere

- WACCMX+DART is first whole atmosphere DA system that can assimilate observations from the surface to ~500 km.
- Used to assess impact of new satellite missions (COSMIC2, NASA GOLD and ICON) on specifying and forecasting the space environment.
- Scientific applications:
  - Study middle-upper atmosphere variability forced by solar storms and lower atmosphere,
  - Predictability of the near-Earth space environment.



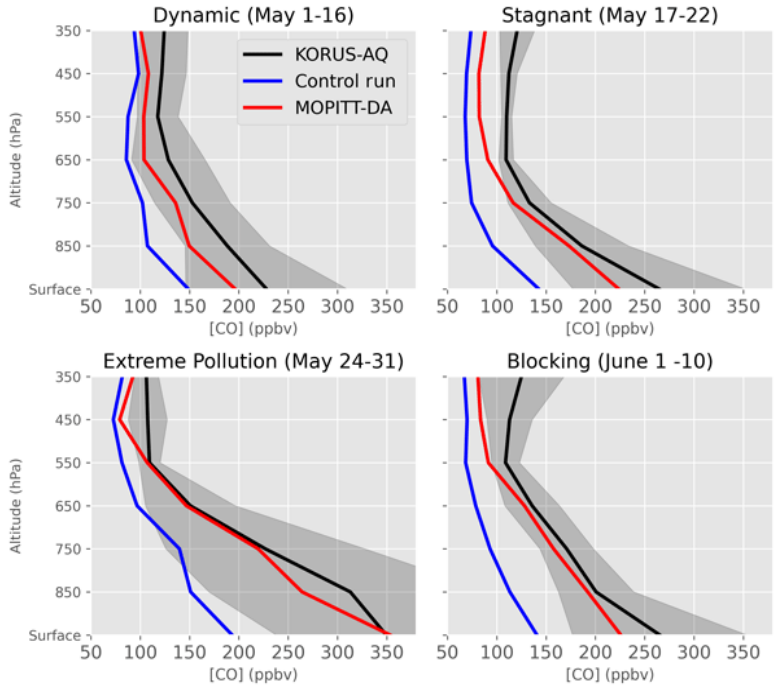
Forecast and analysis RMSE and bias compared to COSMIC-2 electron content observations.

Assimilating COSMIC-2 observations during April 25-30, 2020 reduces forecast RMSE and bias by 6.4% and 28.1% at 300 km

Air Quality: Field Campaign and Satellite Data for Pollution Emission Estimation

- **Aircraft measurements from KORUS-AQ field study in Korea 2016**
- **Satellite retrievals of CO from Terra/MOPITT**
- **Chemistry modeling with CAM-Chem DART** Ensemble Kalman Filter with:
  - Optimized CO initial conditions
  - Optimized CO emissions

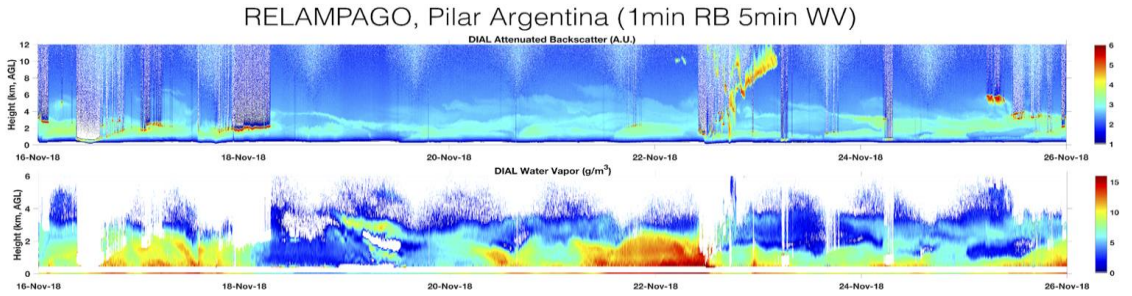
- **Inversion of MOPITT data updated emissions estimates, improved model performance**
  - Against the KORUS-AQ aircraft observations of CO (shown) and O_3, OH, HO_2
  - Suggests underestimates of CO/VOCs in China



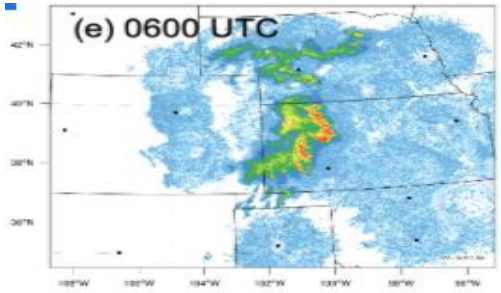
DA improves fit to NASA DC-8 aircraft CO measurements for all synoptic conditions: DA closer to obs than no DA.

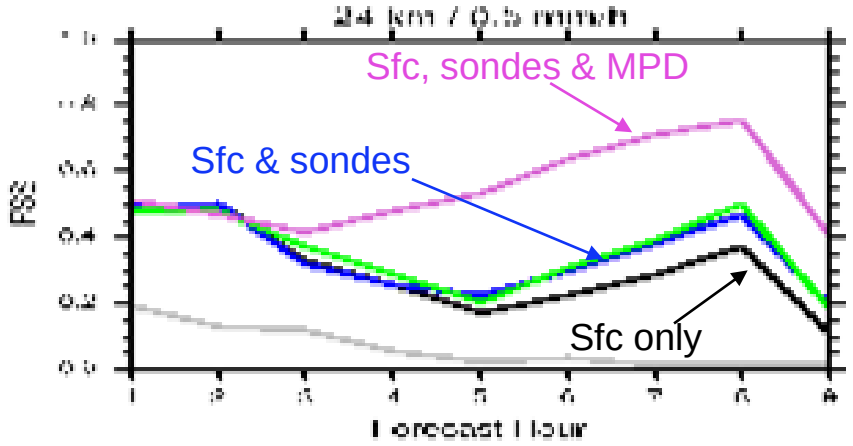
Novel Observations: MPD Water Vapor Profile for Convective Weather Forecasts

MicroPulse Differential absorption lidar (MPD) developed by Montana State University and EOL measures continuous relative backscatter and water vapor profiles.

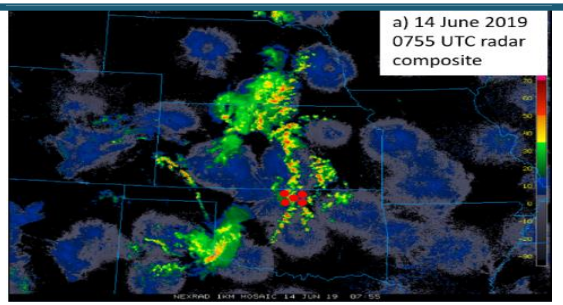


Observing System Simulation Experiment (OSSE)

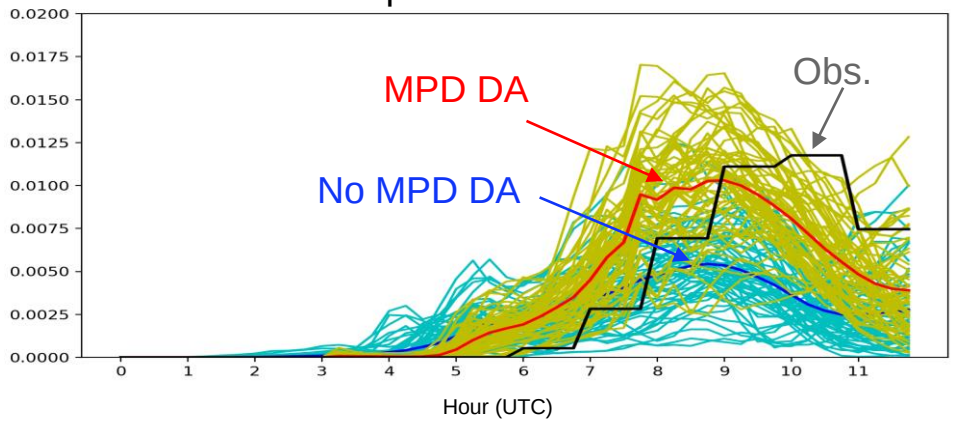




Observing System Experiment (OSE)

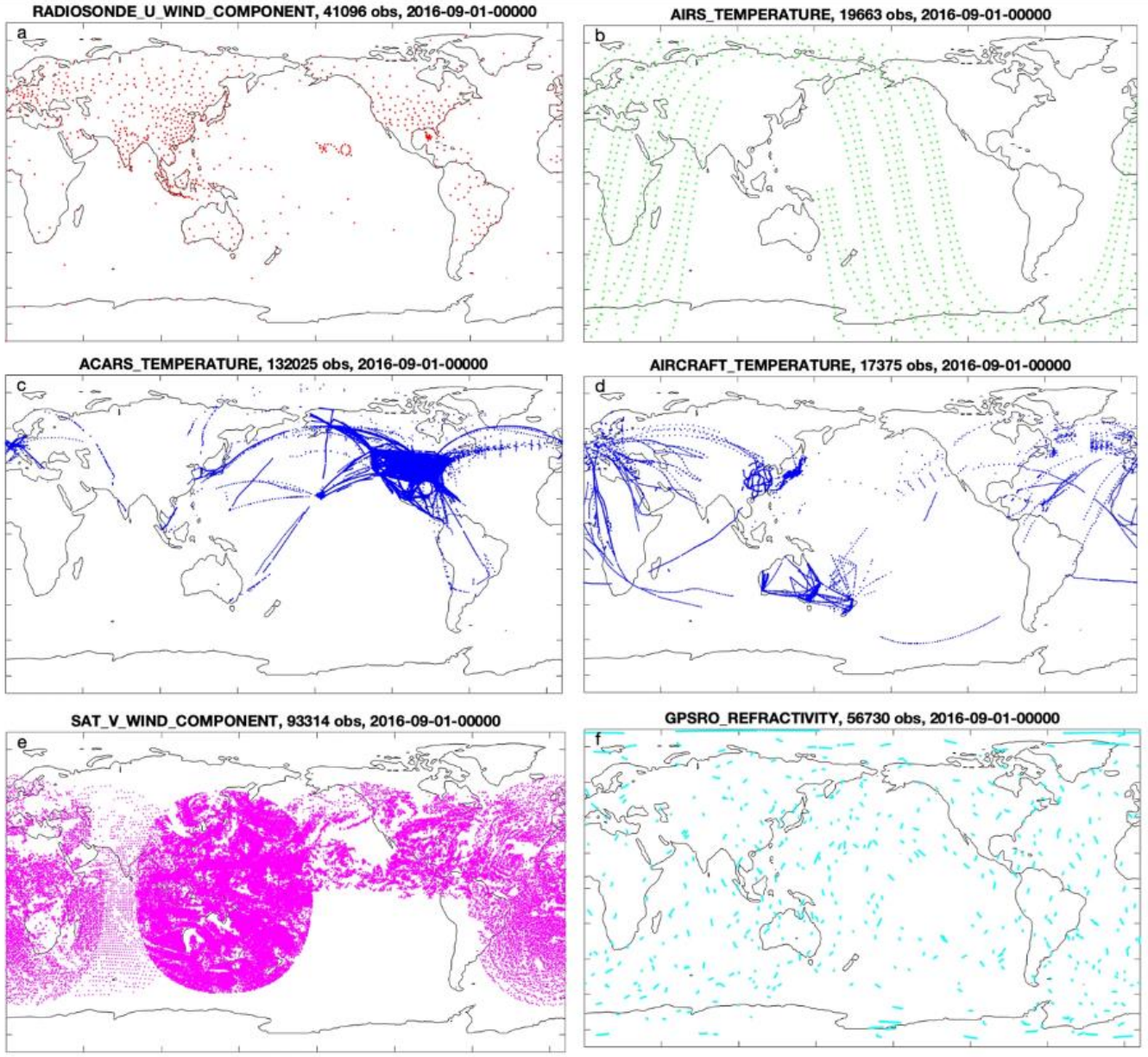


Precipitation time series



WRF/DART DA of MPD improves short-term forecasts of convection initiation and evolution compared to assimilating conventional observations (in the OSSE) and no DA (in the OSE).

Atmospheric Reanalysis Product: Community Atmosphere Model (v6)



CAM6-DART 10 Year Reanalysis:

- Observations: aircraft, radiosonde, satellite radiance, radio occultation
- 80 Ensemble Members,
- Provides forcing ensembles for other CESM models,
- Available from NCAR RDA. https://rda.ucar.edu/datasets/ds345.0/

DART: The Original System

Jeff Anderson
DA Algorithms

Kevin Raeder
Atmosphere DA
Reanalyses

Started in 2003 as an NCAR Director's Strategic Initiative

DART: The Next Generation

Our continuing Mission: To accelerate progress in Earth System Science at NCAR, UCAR Universities, and in the broader science community by providing state-of-the-art ensemble DA capabilities.

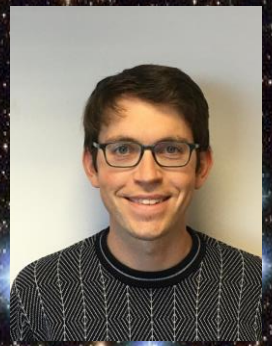
Moha Gharamti
DA Science
Hydrology DA

Ben Johnson
Ocean DA
Space Weather DA

Helen Kershaw
Software Engineer
HPC

Brett Raczka
Land Surface DA
Weather DA

Marlee Smith
Software Engineer



Dan Amrhein
Ocean, Paleo DA
Special CGD Guest Scientist

https://dart.ucar.edu

Artificial Intelligence/Machine Learning

Charlie Becker
Machine Learning Scientist II, MILES

This material is based upon work supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1852977.

Machine Learning Focus Areas

- The Machine Integration and Learning for Earth Systems (MILES) group in CISL/TDD is leading convergent machine learning research at NSF NCAR to push the boundaries of how AI/ML can be used in Earth System Science
- Research
  - Focus on emulation, uncertainty quantification, inverse problems, and process understanding
  - Engaged in both multi-year partnerships and focused collaborations
- Software
  - Develop open source research platforms for using ML in ESS contexts
  - Make cutting-edge ML easier to integrate with ESS observation and modeling systems
- Education
  - Workforce development: interns, visitors, postdocs

The MILES core team at the Research Aviation Facility. (DJ, John, Gabrielle, and Charlie, L->R)

The Machine Integration and Learning for Earth Systems (MILES) Group

MILES Core

David John Gagne
ML Scientist II
CISL/RAL

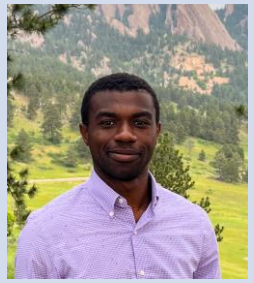
John Schreck
ML Scientist II
CISL

Gabrielle Gantos
Associate Data Scientist II
CISL

Charlie Becker
Associate Data Scientist II
CISL

Yingkai "Kyle" Sha
Postdoc
CISL

Dhamma Kimpara
Grad Intern
CISL/CU



Arnold Kazadi
Grad Intern
CISL/Rice

MILES+Unidata

Thomas Martin
Software Engineer

MILES+CGD

Kirsten Mayer
Project Scientist I

Will Chapman
Project Scientist I

MILES+LEAP

Da Fan
Postdoc

Wayne Chuang
Integration Engineer

MILES+WRaDS

Julie Demuth
Project Scientist III

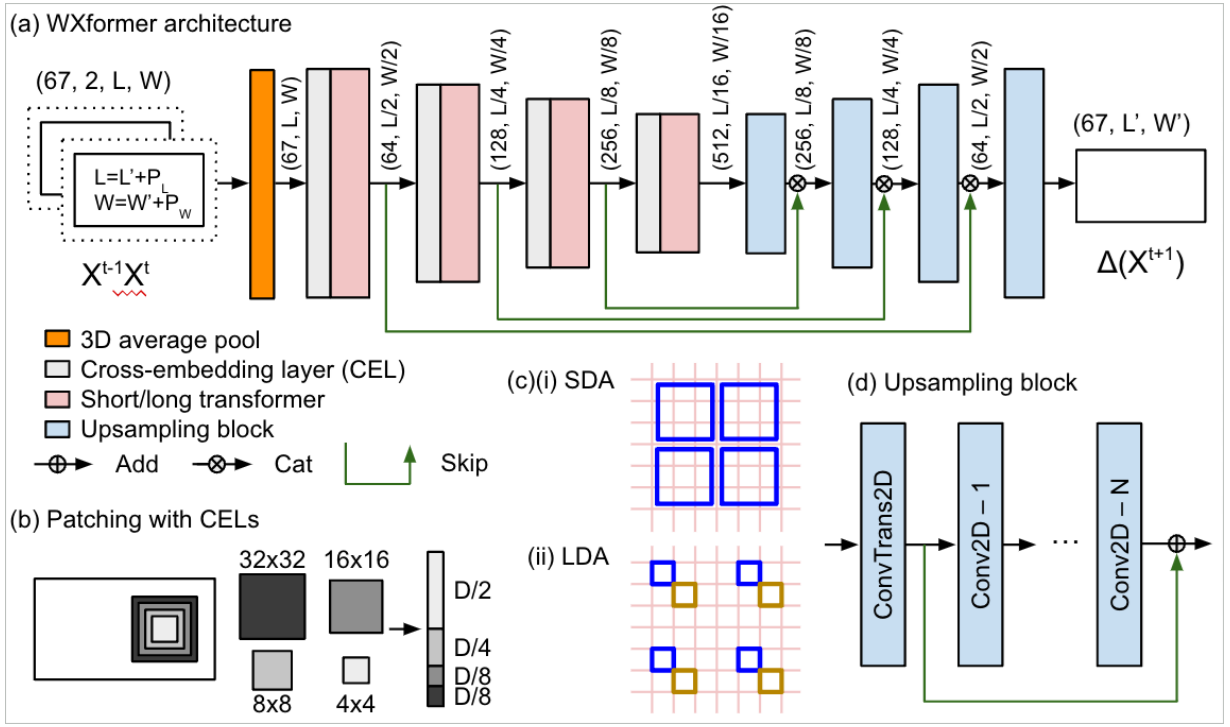
Mariana Cains
Project Scientist I

Chris Wirz
Project Scientist I

Jacob Radford
Research Associate
CIRA

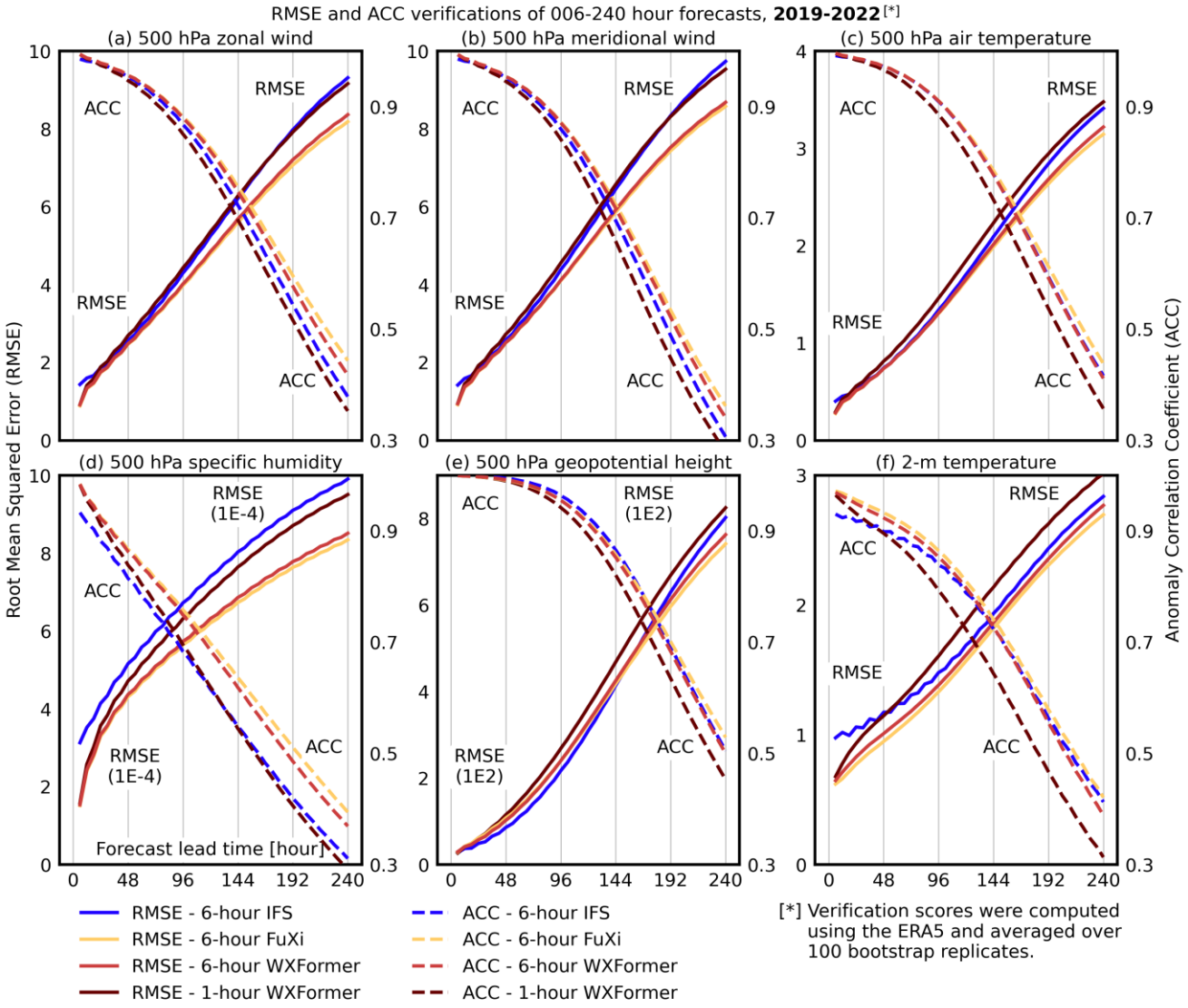
NSF NCAR Community Research Earth Digital Intelligence Twin (CREDIT)

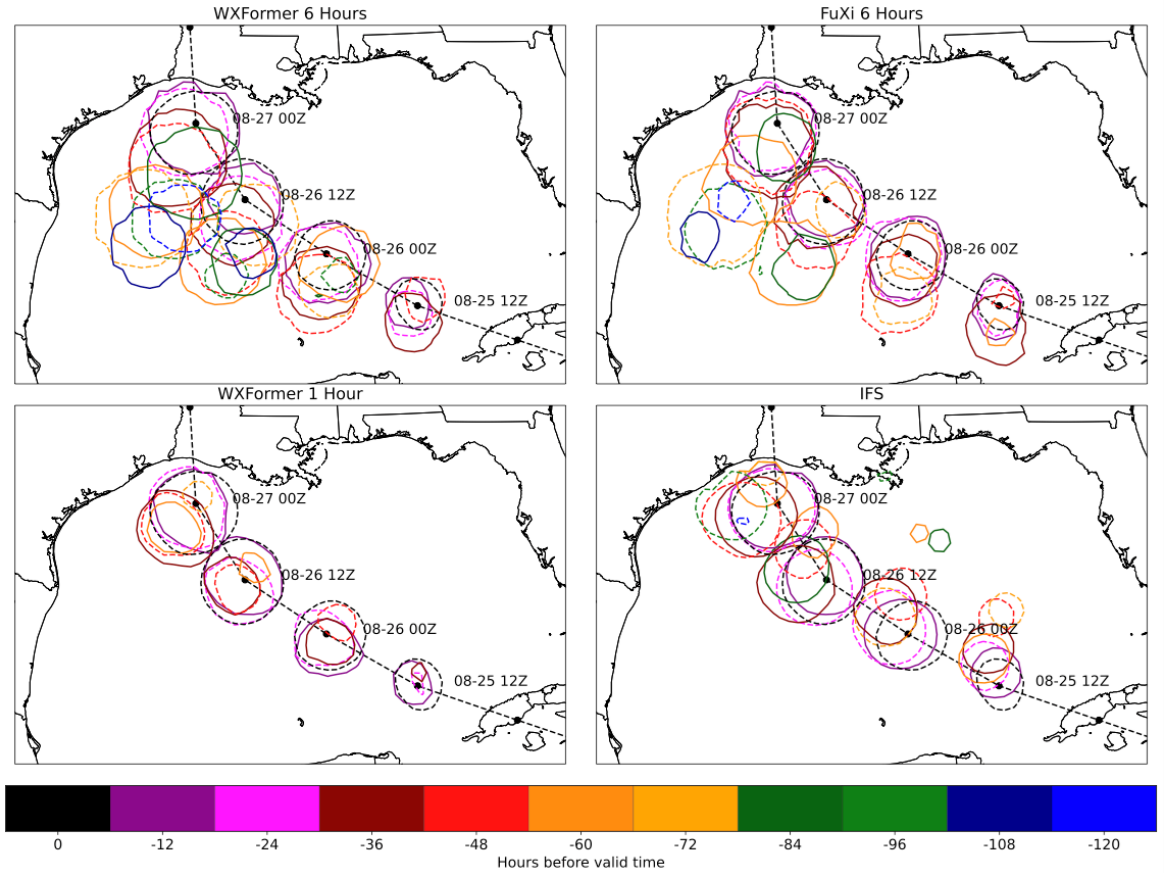
- Research platform for understanding of practices for training and operating global and regional AI weather prediction models
- Platform Features
  - Integrated pre-processing for reanalysis or reforecast data
  - Library of existing and new PyTorch neural network weather prediction architectures
  - Scalable training and inference on NCAR HPC
- Novel advances
  - Training on ERA5 model levels instead of pressure levels
  - WXFormer architecture
  - Stable hourly global model out to 5 days
  - Longer stable forecasts with multiple architectures



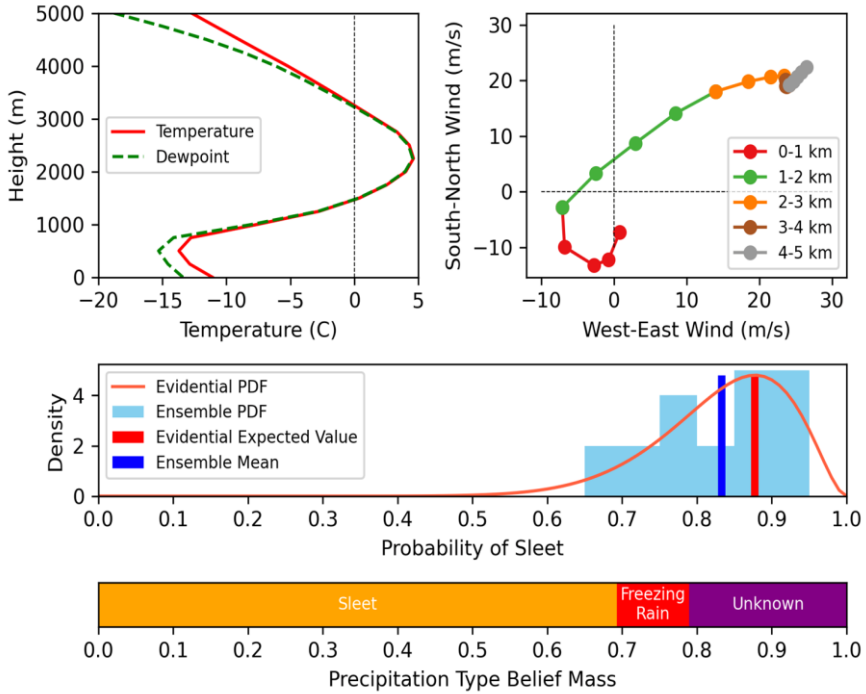
Team: John Schreck, Kyle Sha, Will Chapman, Dhamma Kimpara, Arnold Kazadi, Seth McGinnis, Negin Sobhani, Ben Kirk, Judith Berner, David John Gagne

CREDIT Verification

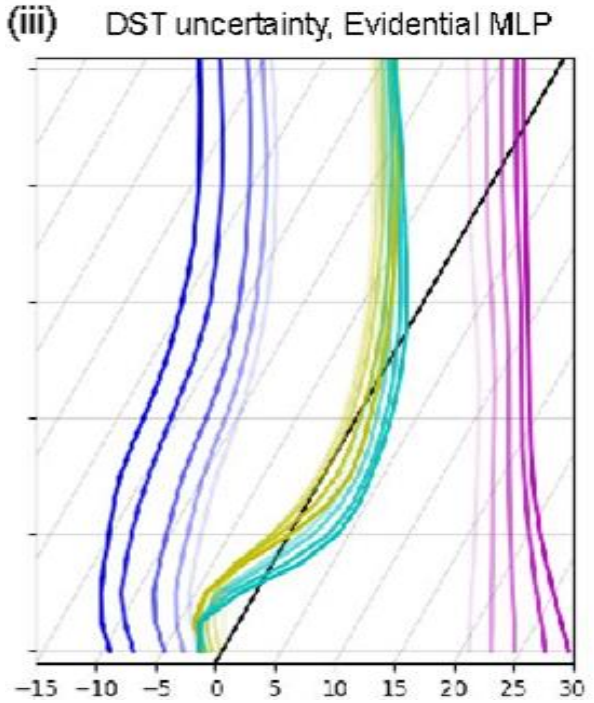




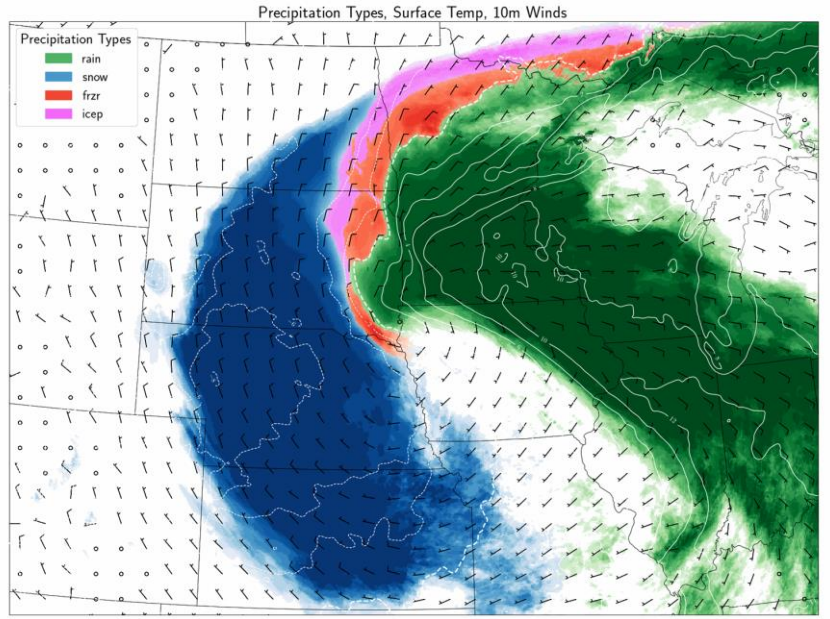
Machine Learning Uncertainty Quantification



ML models can represent predictive uncertainty through ensembles or higher-order parametric distributions.



Uncertainty can separate examples in physically realistic ways.



Higher uncertainty aligns with both terrain and weather features.

MILES-GUESS: machine learning uncertainty quantification with evidential and ensemble models

Paper: Schreck, J. et al., 2024: Evidential Deep Learning: Enhancing Predictive Uncertainty Estimation for Earth System Science Applications, https://arxiv.org/abs/2309.13207

MILES: What can we offer you?

- MILES collaborates with teams across NSF NCAR labs on cutting-edge, high impact AI/ML projects
- We support software for emulators, uncertainty quantification, process understanding, and inverse problems
- We scale AI/ML workflows on NSF NCAR HPC
- We co-mentor interns, visitors, and postdocs

Interested in collaborating? Email: milescore@ucar.edu

Website: https://miles.ucar.edu

Training and Communities

Katelyn FitzGerald

This material is based upon work supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1852977.

NCAR HPC User Group (NHUG)

NHUG is a community dedicated to **promoting the productive use of NCAR HPC resources** and increasing collaboration among NCAR HPC users including building partnerships with university collaborators.

NHUG Activities and Resources

- NHUG Monthly Meetings: Featuring different workflows on NCAR HPC and user groups.
- NHUG Community Blog : Tips and tricks from the community for the community.
- Tutorials on different technologies.

NHUG Page

ncar.pub/nhug_page

News; blogs, archive of past meetings; links to all resources.

NHUG Slack

ncar.pub/nhug-slack

Community discussions, issue reporting, peer networking.

Email List

ncar.pub/NHUG_emaillist

News, outreach, upcoming events.

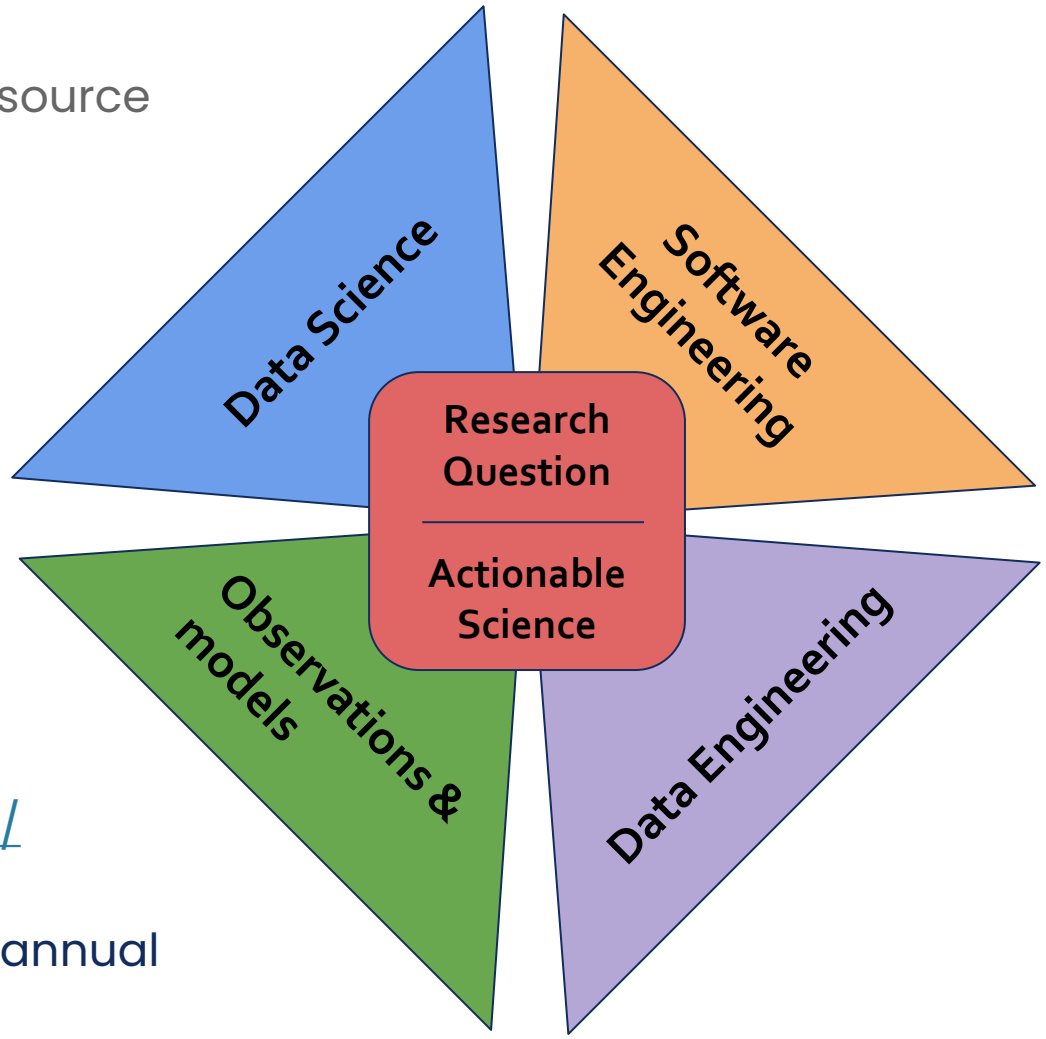
Earth System Data Science (ESDS) Initiative

ESDS is a community with a common interest in advancing geoscientists' ability to make effective use of modern, open-source tools and technologies to support their data science needs.

Approach

Cultivate a **community of practice** encouraging:

- Open development
- Reproducible science
- Training and education
- Community building



ESDS Communication and Resources

- **Website :** https ://ncar.github.io/esds/
- **Email list :** https://groups.google.com/a/ucar.edu/g/esds/
- **Zulip :** ncar.pub/esds-zulip
- **Activities :** discussion forum, office hours, biweekly forum, annual events, and more.

Project Pythia and Pythia Cookoffs

PROJECT PYTHIA

- An education and training hub for the geoscientific Python community
- Geoscience-focused Python tutorials, videos, examples & “cookbooks”
- Pythia Foundations: foundational skill building for scientific Python

See: projectpythia.org

EarthCube

Pythia Cookoffs

- Hackathon to create advanced geoscientific tutorials and workflows for Pythia Cookbooks.
- Collaborative, reproducible content creation using GitHub-based infrastructure.

Project Pythia: Supporting Python Learning

Dask Cookbook

This Project Pythia Cookbook provides a comprehensive guide to understanding the basic concepts and collections of Dask as well as its integration with Xarray. Dask is a parallel computing library that allows you to scale your computations to multiple cores or even clusters, while Xarray is a library that enables working with labelled multi-dimensional arrays, with a focus on working with netCDF datasets.

Motivation

The motivation behind this repository is to provide a clear and concise resource for anyone looking to learn about the basic concepts of Dask and its integration with Xarray. By providing step-by-step tutorials, we hope to make it easy for users to understand the fundamental concepts of parallel computing and distributed data processing, as well as how to apply them in practice using Dask and Dask+Xarray.

Authors

Negin Sobhani, Brian Vanderwende, Deepak Cherian, and Ben Kirk

Contributors

Note on Content Origin

This cookbook is derived from the extensive material used in the NCAR tutorial, "Using Dask on HPC systems", which was held in February 2023. The NCAR tutorial series also includes an in-depth exploration and practical use cases of Dask on HPC systems and best practices for Dask on HPC. For the complete set of NCAR tutorial materials, including these additional insights on Dask on HPC, please refer to the main NCAR tutorial content available here.

PROJECT PYTHIA

Blocked Algorithms

Dask Arrays use blocked algorithms to split large computations into smaller computations which operate on subsets of the data (called **chunks**).

Let's see what this means in an example:

```
import numpy as np
import dask.array as da

# A 4x4 numpy array that goes from 1 to 16

narr = np.array([
       [ 1,  2,  3,  4],
       [ 5,  6,  7,  8],
       [ 9, 10, 11, 12],
       [13, 14, 15, 16]
       ])

# -- convert numpy array to dask array with 4 chunks
darr = da.from_array( narr,chunks=(2, 2))
```

Now we can calculate the sum of this array using `darr.sum()` similar to numpy. **But how is it different from numpy?**

https://projectpythia.org/dask-cookbook/README.html

Improving Scientific Software Conference (ISS)

Since 2012, the Software Engineering Assembly (SEA) has hosted an annual ISS conference highlighting novel approaches to scientific software design, maintenance, and use.

Join us!

ISS 25 Improving Scientific Software

April 7–10, 2025 at NSF-NCAR (Boulder, CO) & Virtual

We welcome any abstracts related to improvements in scientific software, particularly those focused on the following topics:

- Modern Research Software Engineering (RSE) practices (e.g. CI/CD)
- AI/ML integration in software development
- Improved usability, accessibility, and documentation
- Data analysis & visualization tools
- Software optimized for HPC and specialized hardware.

Abstract submission deadline: January 24, 2025

To learn more ►

SIParCS - Summer Internships in Parallel Computational Science

Mailing list

- 11-week Summer internship
- Open to: Upper division undergrads & graduate students
- Domestic & International Students at U.S. Schools
- Project topics
  - Application Optimization
  - Data Science
  - Machine Learning
  - Numerical Methods
  - Software Engineering
  - Supercomputing Systems Operations
  - Visualization
- Support
  - Travel, Housing
  - Competitive salary

Deadline:

January 21, 2025

2025 SIParCS Undergraduate Technical Projects

Apply here ►

Project 1. Analysis of the Slingshot High-Speed Network

Project 2. CIRRUS - Cloud Infrastructure for Remote Research, Universities, and Scientists

Project 3. Developing Novel Capabilities for Reproducibility and Software Correctness in Regional Ocean and Climate Modeling

Project 4. HPC Outage Recovery Testing Tools

Project 5. Improving Data Center Visibility with AI

Project 6. Natural Language Discovery of NSF NCAR Scientific Data

Project 7. Project Pythia Infrastructure and Web Development

Project 8. Reimagining the UI/UX of Data Discoverability, Access, and Capabilities Within the Research Data Archive

2025 SIParCS Graduate Technical Projects

Project 8. Reimagining the UI/UX of Data Discoverability, Access, and Capabilities within the Research Data Archive

Project 9. Community-Driven LoRa Deployment, Open Source Integration, and Machine Learning Evaluation for Atmospheric Sensornets

Project 10. High-Performance Vector Analysis Operators in Python for Unstructured Grids

Project 11. Simulating atmospheric chemistry with python

Apply here ►

Note: Students may apply to up to two SIParCS projects.

Questions?

CISL_Roadshows@ucar.edu

Thank you to our lunch sponsors!

ALLIANCE TECHNOLOGY GROUP

NVIDIA

Hewlett Packard Enterprise

https://www.cisl.ucar.edu/events/ams2025

Backup slides