



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)

December 11, 2024

Goals

1. To build awareness of the many NCAR computing services, products, and resources available to help advance ***your*** science
1. 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 (Ben Kirk)
- Data services (Doug Schuster)
- Community analysis tools and visualization services (Nihanth Cherukuru)
- Data Assimilation (Brett Raczka)
- Machine Learning (David John Gagne)
- Trainings and Communities (Negin Sobhani)
- SIParCS (Negin Sobhani)
- Q&A (all)

- Slides
- Recordings
- Speaker bios
- More!



<https://www.cisl.ucar.edu/events/agu2024>



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

Ben Kirk
Consulting Services Group



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

HPC Systems

2017 SGI/HPE
 4032 Nodes, 145,152 Cores, 313 TB total memory, 4.79 PFlop/s
 #21 Supercomputer in the world at debut



2023

2023

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



+64 soon

Data Analysis & Visualization



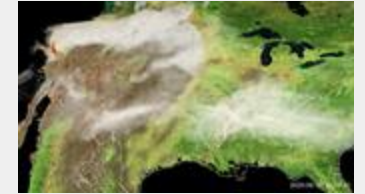
Casper: heterogeneous system for data analysis & viz.

- 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

<http://projectpythia.org>



CISL develops specialized visualization software & services for Earth Science applications



<https://geocat.ucar.edu>

High Performance Storage



Campaign Storage

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

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

Hierarchical Storage
 (HSM) Data Migration

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.
- The [NCAR HPC Users Group Slack workspace](#) is where the HPC community collaborates on timely issue reporting, seeking peer feedback, and networking with dedicated channels for *#derecho-users*, *#casper-users*, *#jupyterhub-users*, and more. ([join now!](#))
- User Groups & Training (*details later...*)

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

Search NSF NCAR Data Catalog (DASH Search)



<https://data.ucar.edu>



DASH Search

Home

Resources

Collections

About

DASH Search allows users to find, browse, and access digital assets created and published by NSF NCAR and UCAR Community Programs.

Search Data, Software, Models and Publications

Search...

Browse by Resource Type

collection

dataset

image

model

publication

software



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>

A screenshot of the NSF NCAR Research Data Archive website. The page features a header with the NSF and NCAR logos, navigation links for 'Contact Us', 'User Dashboard', and 'Sign Out', and a search bar. The main content area is titled 'Data for Atmospheric and Ocean Sciences Research' and includes a descriptive paragraph about the archive's collection and management. A prominent blue button labeled 'Search for Data' is positioned above a search input field with a magnifying glass icon.

Research Data Archive | Datasets | Resources | Support | News | Search | Search RDA

Data for Atmospheric and Ocean Sciences Research

The NSF NCAR Research Data Archive (RDA) contains a large and diverse collection of meteorological, atmospheric composition, and oceanographic observations, and operational and reanalysis model outputs, integrated with NSF NCAR High Performance Compute resources to support atmospheric and geosciences research. The RDA is managed by the Data Engineering and Curation Section (DECS) of the Computational and Information Systems Laboratory (CISL) at the NSF National Center for Atmospheric Research (NSF NCAR).

Search for Data

Data keyword search or dataset number

<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

[Home](#) / [Datasets](#) / D633006



ERA5 Reanalysis Model Level Data

d633006 | DOI: 10.5065/XV5R-5344 ☆

[ASK A QUESTION >](#)

DESCRIPTION

1

DATA ACCESS

CITATION

DOCUMENTATION

SOFTWARE

METRICS

Mouse over the underlined table headings for detailed descriptions

DATA DESCRIPTION	DATA FILE DOWNLOADS		CUSTOMIZABLE DATA REQUESTS	NCAR-ONLY ACCESS
	UNION OF AVAILABLE PRODUCTS	<u>Web Server Holdings</u>	<u>Globus Transfer Service (GridFTP)</u>	<u>Subsetting</u>
Web File Listing		Globus Transfer	Get a Subset	GLADE File Listing

NSF NCAR Research Data Archive - Data Proximate Compute



rda.ucar.edu

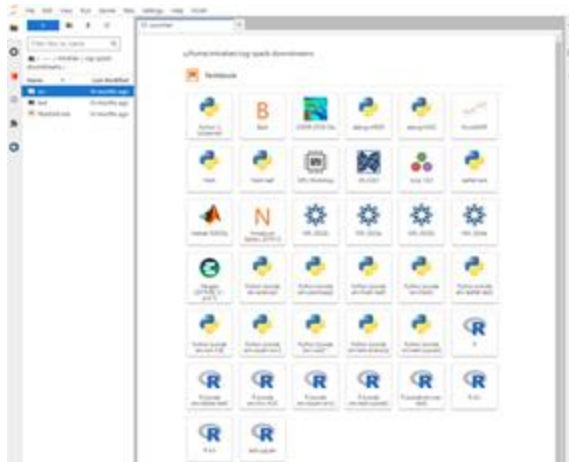
- **Popular Datasets**
 - **NCEP Products**
 - *Final Analysis, GFS, CFSR, CFS, In Situ Obs*
 - **ECMWF products**
 - *ERA-5 0.25 degree and model resolution*
 - **JMA products**
 - *JRA-55, JRA-3Q*
- **PI/NCAR produced datasets**
 - **USGS CONUS 404, CESM2 Large Ensemble**

Data 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



Allocation Opportunities

Submit new requests for projects here.
To manage one of your existing projects, select My Allocations above.

Small Allocation (University) START A NEW SUBMISSION DETAILS	Exploratory Allocation (University) START A NEW SUBMISSION DETAILS	Data Analysis Allocation (University) START A NEW SUBMISSION DETAILS
Classroom Allocation (University) START A NEW SUBMISSION DETAILS	NCAR External Projects START A NEW SUBMISSION DETAILS	

Questions?

rdahelp@ucar.edu

schuster@ucar.edu





Community Analysis Tools and Visualization Services

Nihanth Cherukuru, Orhan Eroglu
Visualization and Analysis Software Team

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”

Projects:



GeoCAT

Packages:

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



Contact GeoCAT



Orhan Eroglu



Anissa Zacharias



Cora Schneck



Julia Kent



Katelyn FitzGerald



Philip Chmielowiec



GeoCAT Software Products

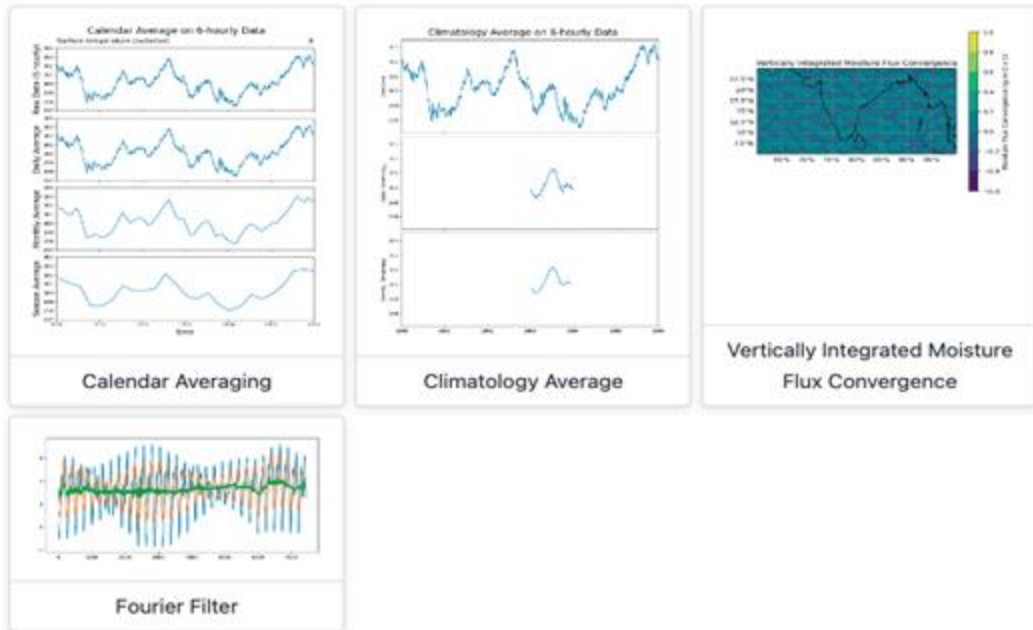
GeoCAT-comp: Analysis operators

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

Applications	
Plot Types	▼
Data Analysis	▼
Date and Time	▼

Usage Examples

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



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

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

NCL to Python

NCL Index

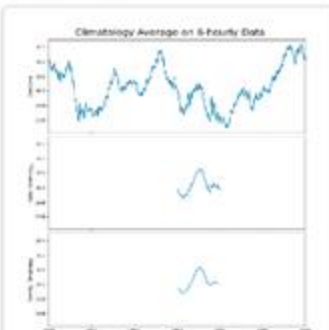
NCL Applications

Usage Examples

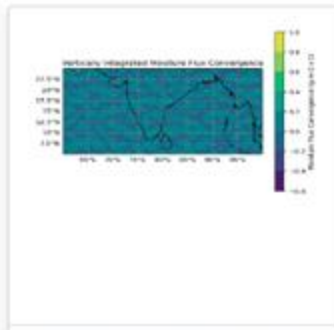
Here's some examples of how to use geocat-comp.



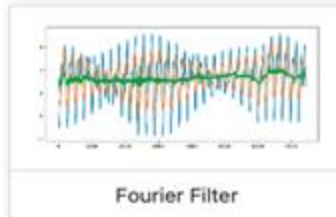
Calendar Averaging



Climatology Average



Vertically Integrated Moisture Flux Convergence



Fourier Filter

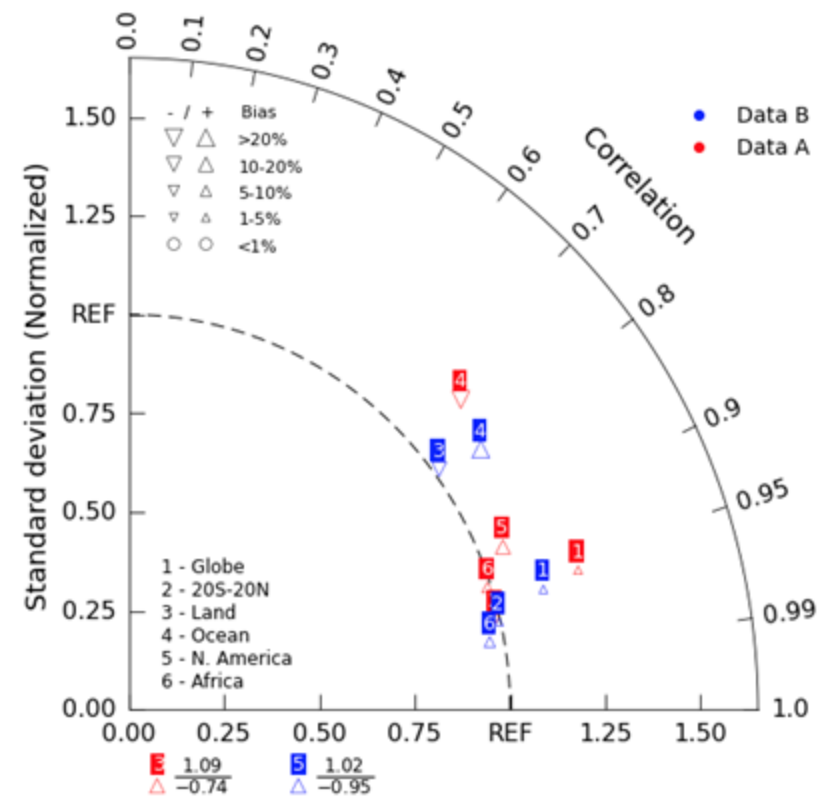
NCL Index

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

GeoCAT-examples: Visualization Gallery

The screenshot shows the GeoCAT Visualization Gallery website. At the top left are the NSF and NCAR logos. A search bar is present. A left sidebar lists various plotting categories such as Animate, Bar, Boxplots, Write CSV, Colors, Contour, Line, Map Projections, Masking, Meteograms, Overlays, Panels, Polygons, Probability Distributions, Regression, Scatter, and Shapefiles. The main content area is titled "Gallery" and contains text about the gallery's purpose and the software tools used (matplotlib, cartopy, geocat-datafiles, geocat-viz, Xarray, and numpy). Below this text are sections for "Animate" (with a thumbnail for "NCL_animate_1.py") and "Bar" (with four thumbnails for "NCL_bar_1.py", "NCL_bar_11.py", "NCL_bar_14.py", and "NCL_bar_2.py"). A right sidebar lists a "Contents" menu with categories like Animate, Bar, Boxplots, Write CSV, Colors, Contour, Line, Map Projections, Masking, Meteograms, Overlays, Panels, Polygons, Probability Distributions, Regression, Scatter, Shapefiles, Skew-T, Station Examples, Streamlines, Tables, Taylor Diagrams, Topography, Trajectories, Vector Plots, and WRF.

Special plots, e.g.



Taylor diagrams

Next generation, kilometer-scale climate and global weather models

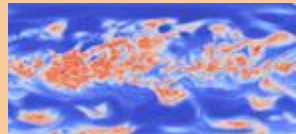
After nearly two decades of development and evaluation, the climate and global weather modeling communities are transitioning from simple structured grids to more complex, but more scalable **unstructured grids** upon which governing equations of state are solved.

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

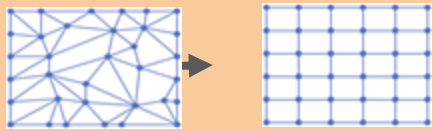
Computational operators

$$\mathcal{F}(x) \frac{\partial x}{\partial y} \mathcal{L}(x)$$

Plotting



Regridding



2. Sustainable and community owned



Visualization Services and Research (ViSR)

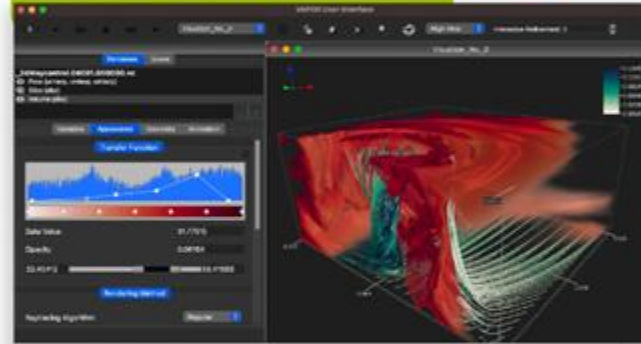
"Advanced Visualizations for Data Analysis and Communication"



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



Made with VAPOR



Nihanth Cherukuru



Stas Jaroszynski



Matt Rehme



Scott Pearse

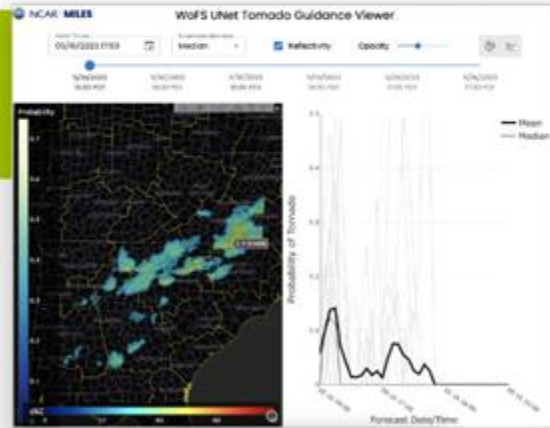


Tammy Zhang



Ian Franda

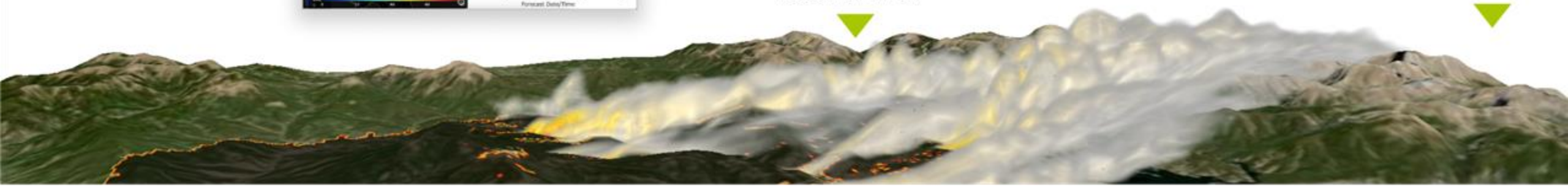
Visualizations for Decision Making



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

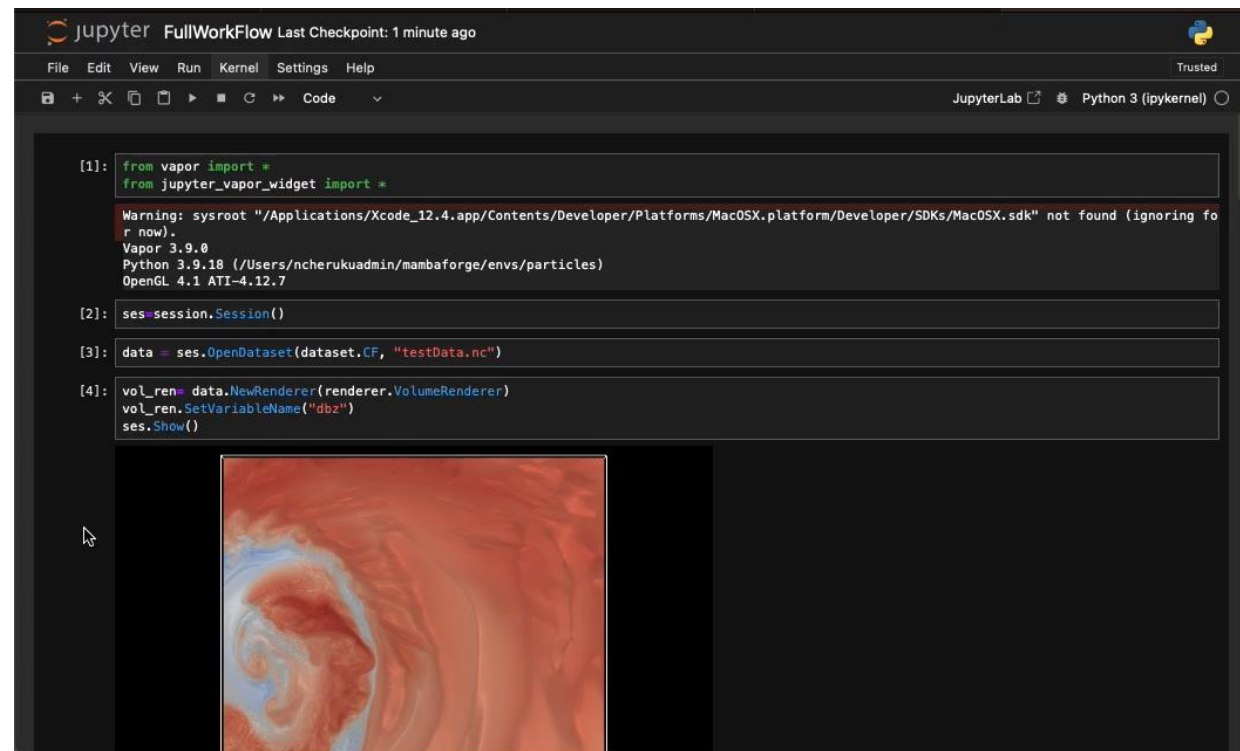
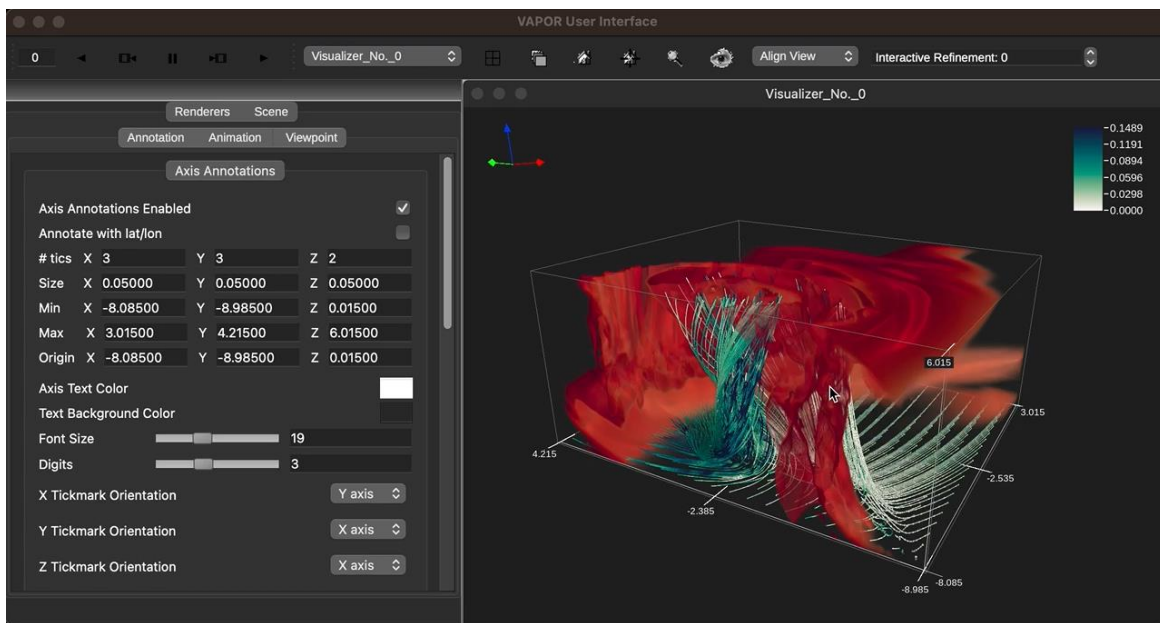
WRF Fire, East Troublesome Wildfire
Visualization: Scott Pearse

Science: Juliano T, DeCastro A, Kosovic B, Edgley C





- 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



Contact:

Nihanth Cherukuru
ncheruku@ucar.edu

CM-1 Simulation of a precipitating cumulus-congestus cloud

Visualization: Matt Rehme

Science: Hugh Morrison, Kamal Kant Chandrakar

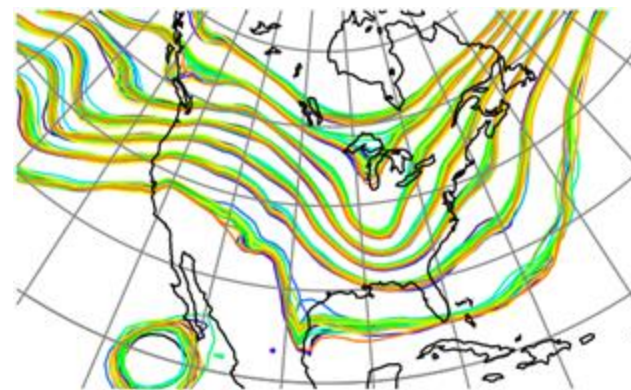




Data Assimilation

Brett Raczka

Data Assimilation Research Section



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

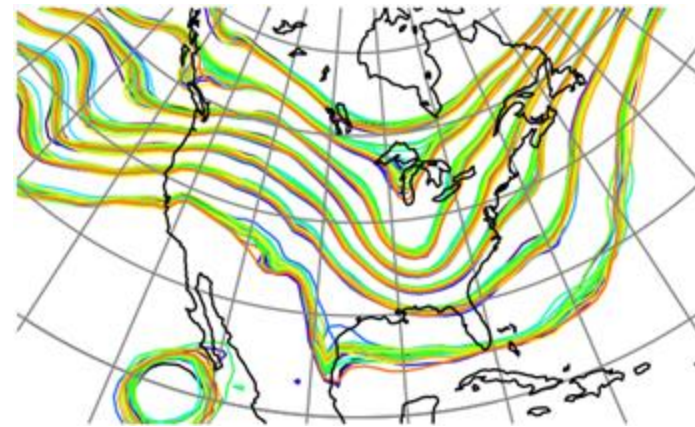


What is Data Assimilation?

Observations combined with a Model forecast...



+



...to produce an analysis
(best possible estimate).

The Data Assimilation Research Section (DAReS)

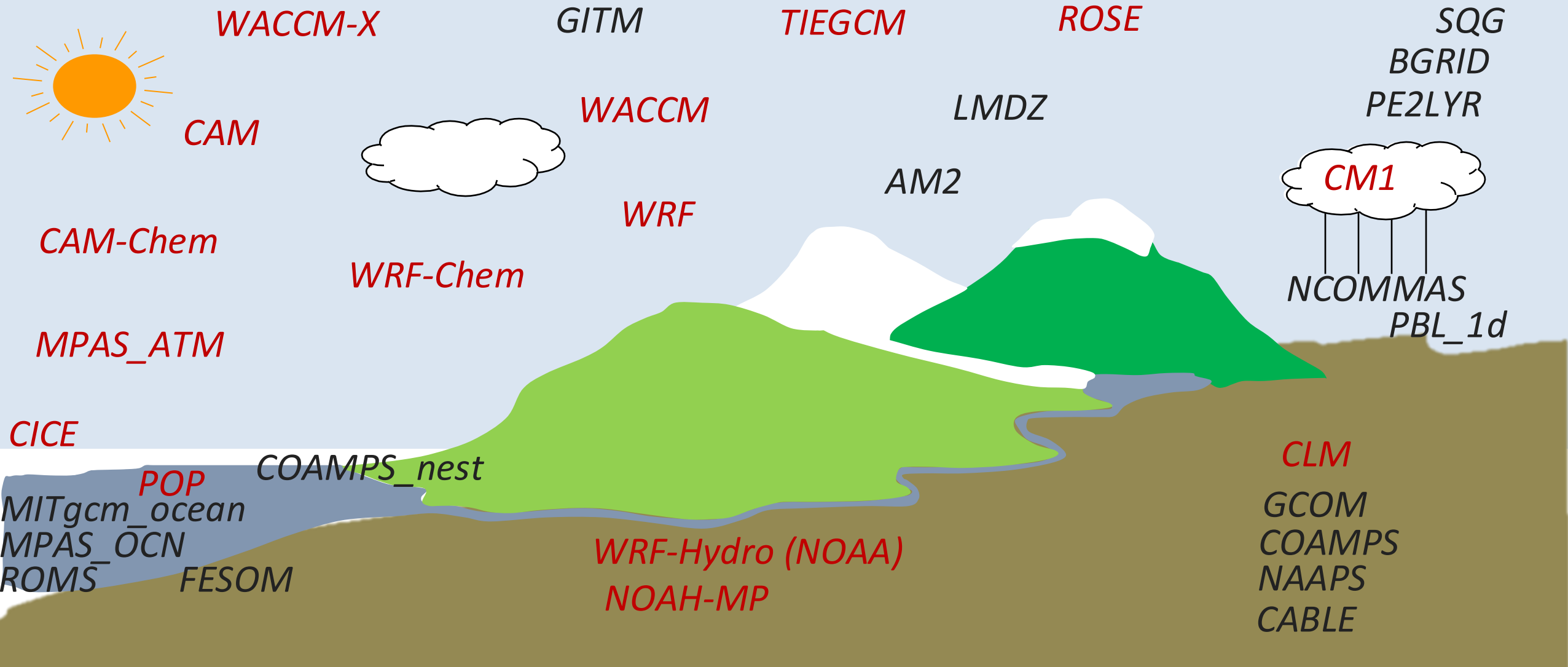
Mission: To accelerate progress in Earth System Science at NCAR, UCAR Universities, and in the broader science community by providing state-of-the-art ensemble DA capabilities.

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

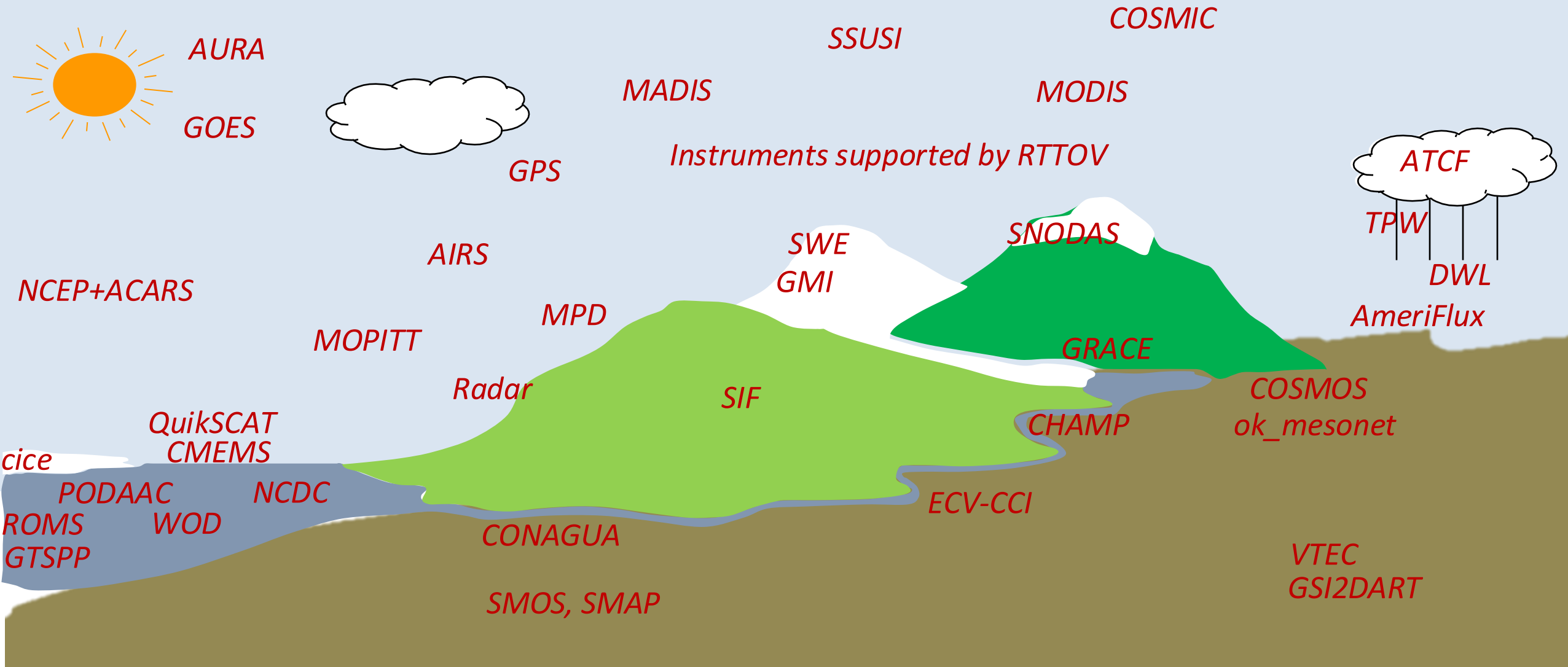
Data Assimilation Research Testbed (DART)

- 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

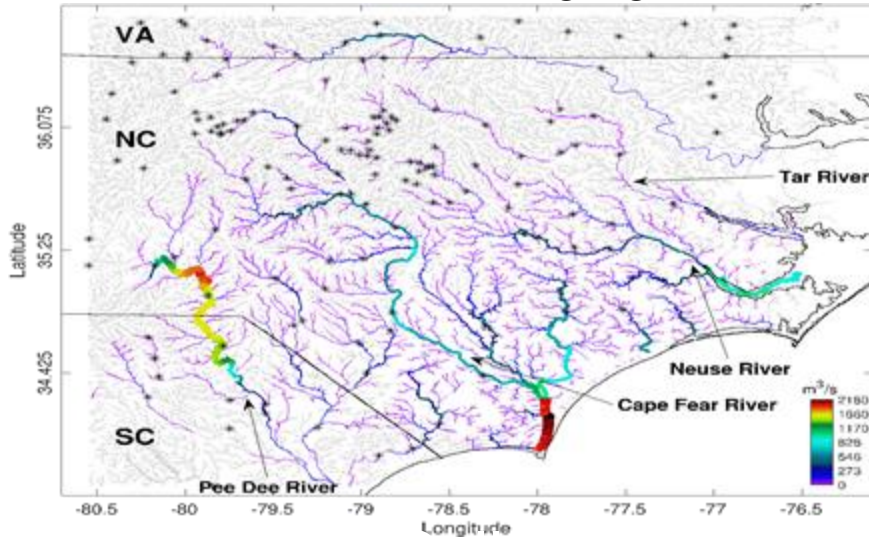


Supported Earth System Observations (others available)

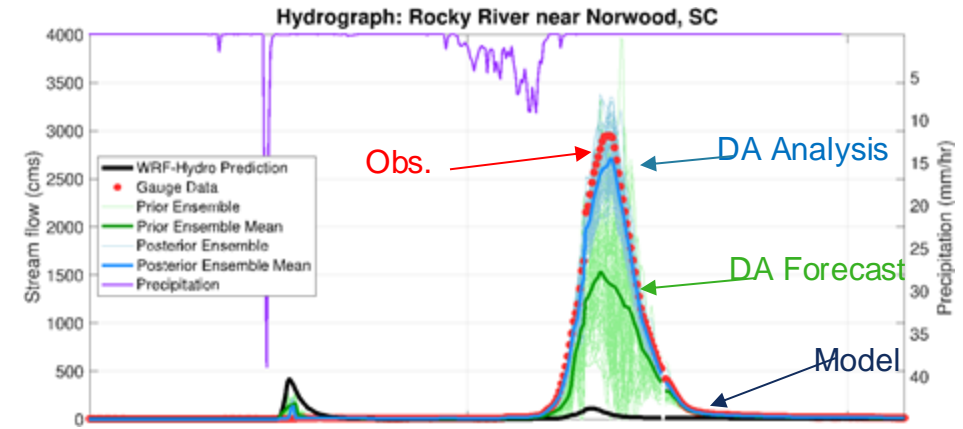


Flood Prediction: WRF-Hydro/DART for Hurricane Florence 2018

High-resolution stream network with USGS streamflow gauges.



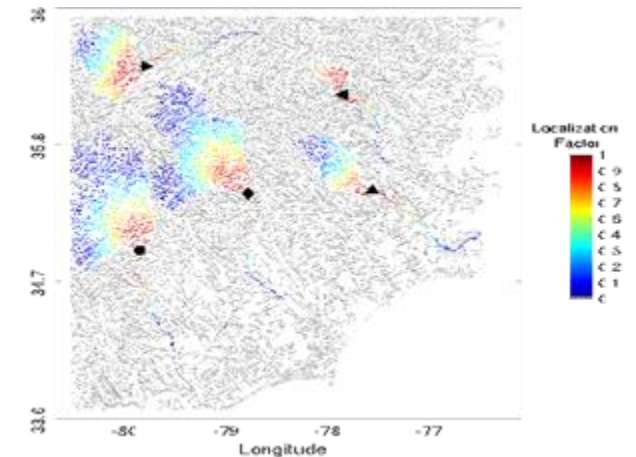
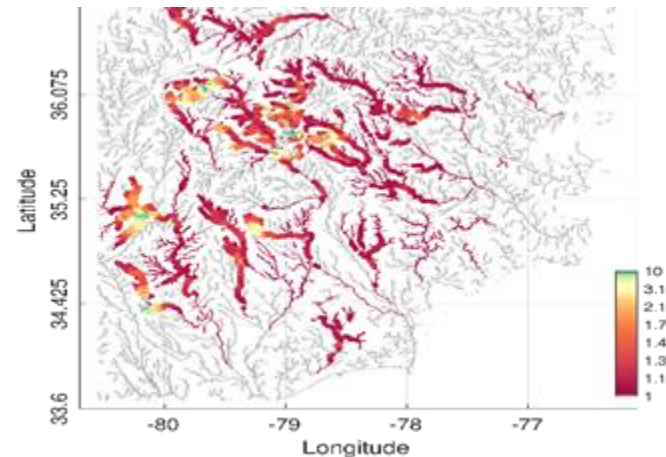
DA greatly improves analysis and forecasts of streamflow.



Novel Data Assimilation Science

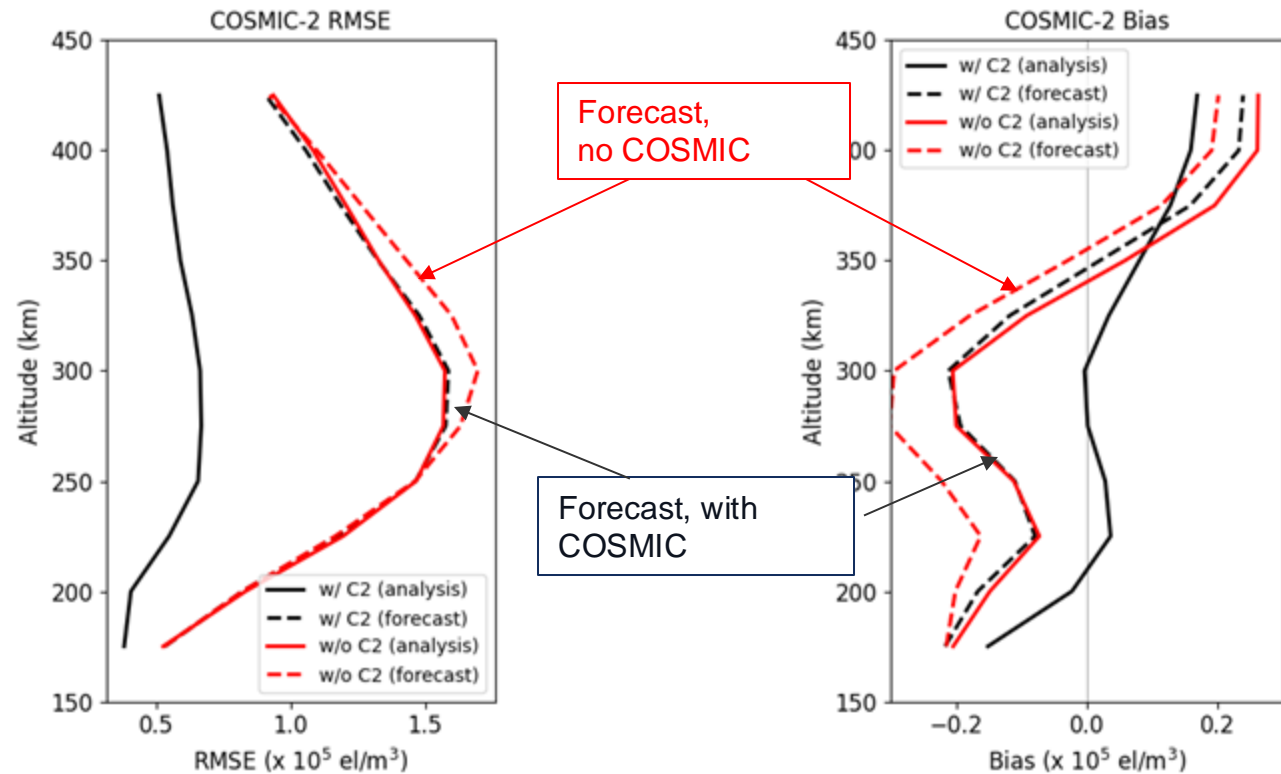
1. Prior and Posterior Adaptive Covariance Inflation

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.

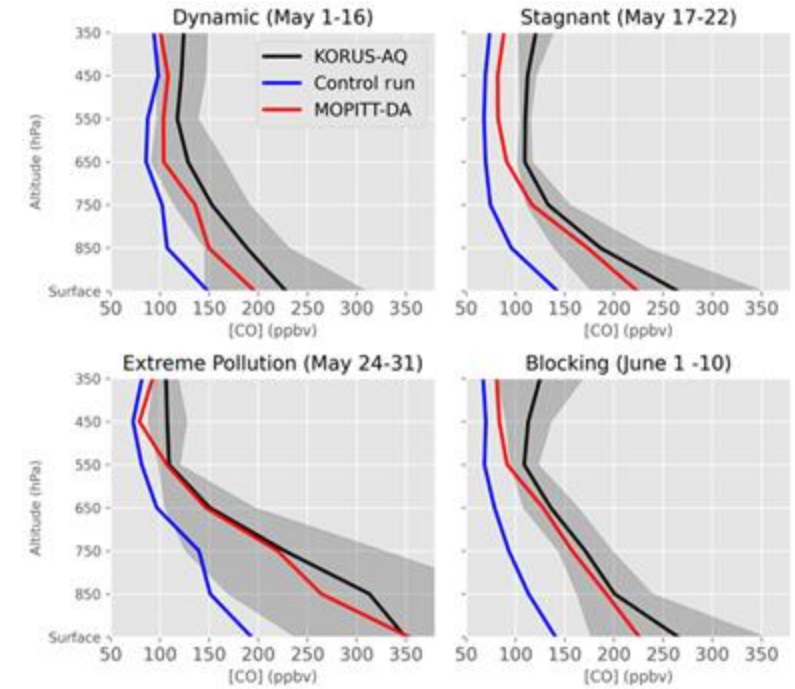


Forecast and analysis RMSE and bias compared to COSMIC-2 electron content observations.

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 O₃, OH, HO₂
 - Suggests underestimates of CO/VOCs in China

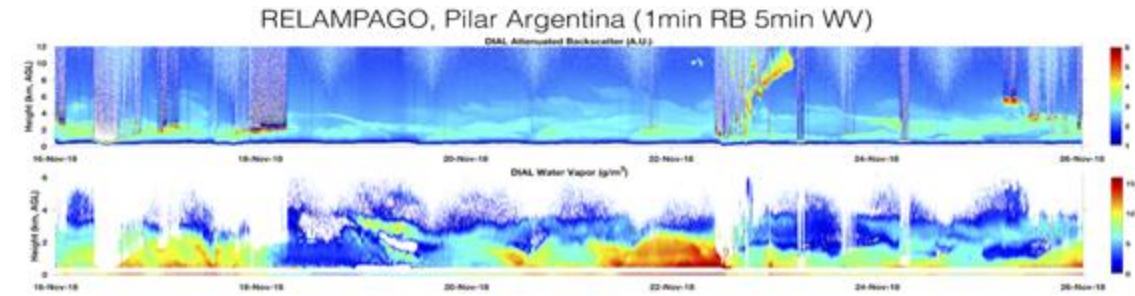


DA improves fit to NASA DC-8 aircraft CO measurements for all synoptic conditions: **DA** closer to obs than **no DA**.

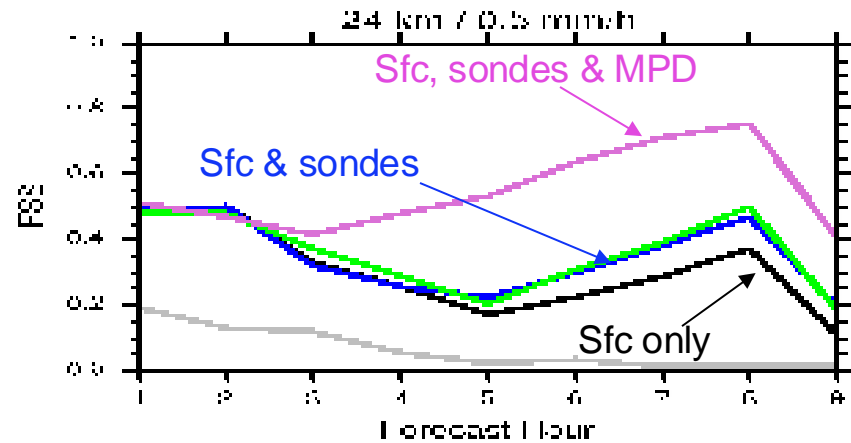
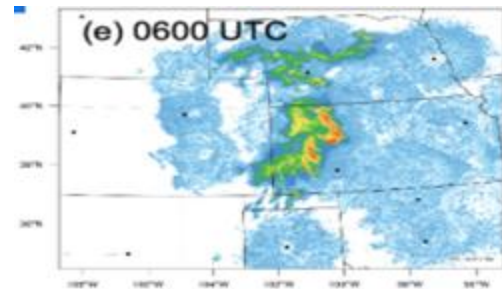
Novel Observations: MPD Water Vapor Profile for Convective Weather Forecasts



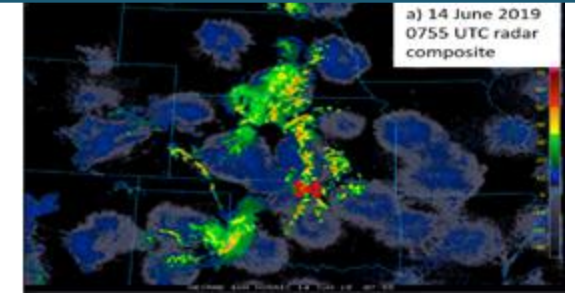
MicroPulse Differential absorption lidar (MPD) developed by Montana State University and EOL measures continuous relative backscatter and water vapor profiles.



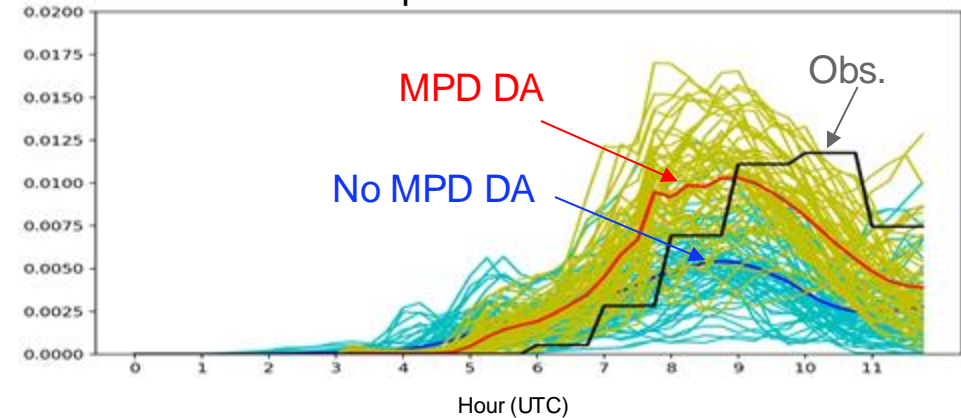
Observing System Simulation Experiment (OSSE)



Observing System Experiment (OSE)



Precipitation time series

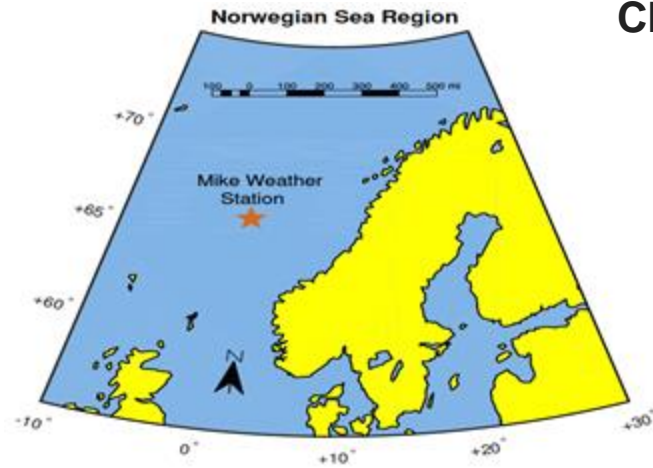


WRF/DART DA of MPD improves short-term forecasts of convection initiation and evolution compared to assimilating conventional observations (in the OSSE) and no DA (in the OSE).

Parameter Estimation: Ocean Biogeochemistry Modeling

CESM1: Community Earth System Model with ocean components:

- ❑ Physics: Miami Isopycnic Coordinate Ocean Model (MICOM, Bleck et al., 1992)
- ❑ Biogeochemistry: The HAMburg Ocean Carbon Cycle (HAMOOC5, Maier-Reimer et al. 2005)
 - NPZD-type BGC model with Carbon

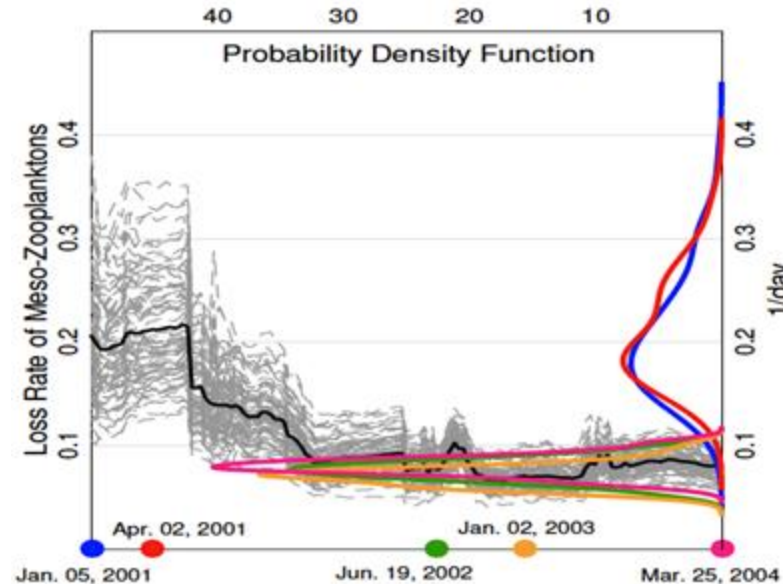


Study location (station M) in the Norwegian Sea

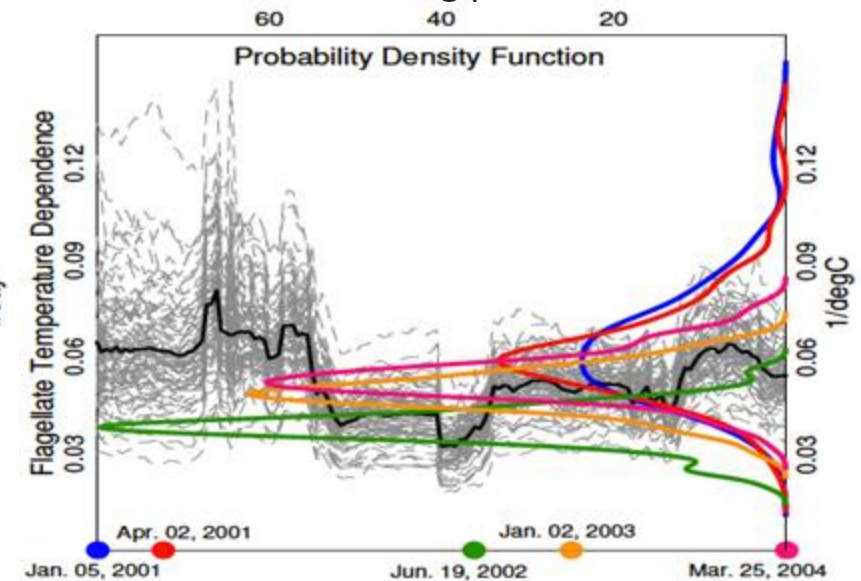
State and Parameters Estimation using the EnKF

- ❑ Weekly updates
- ❑ BGC state: 15 variables (tracers) & 11 parameters
- ❑ Data: nutrients & oxygen profiles, surface chlorophyll and partial pressure of CO₂

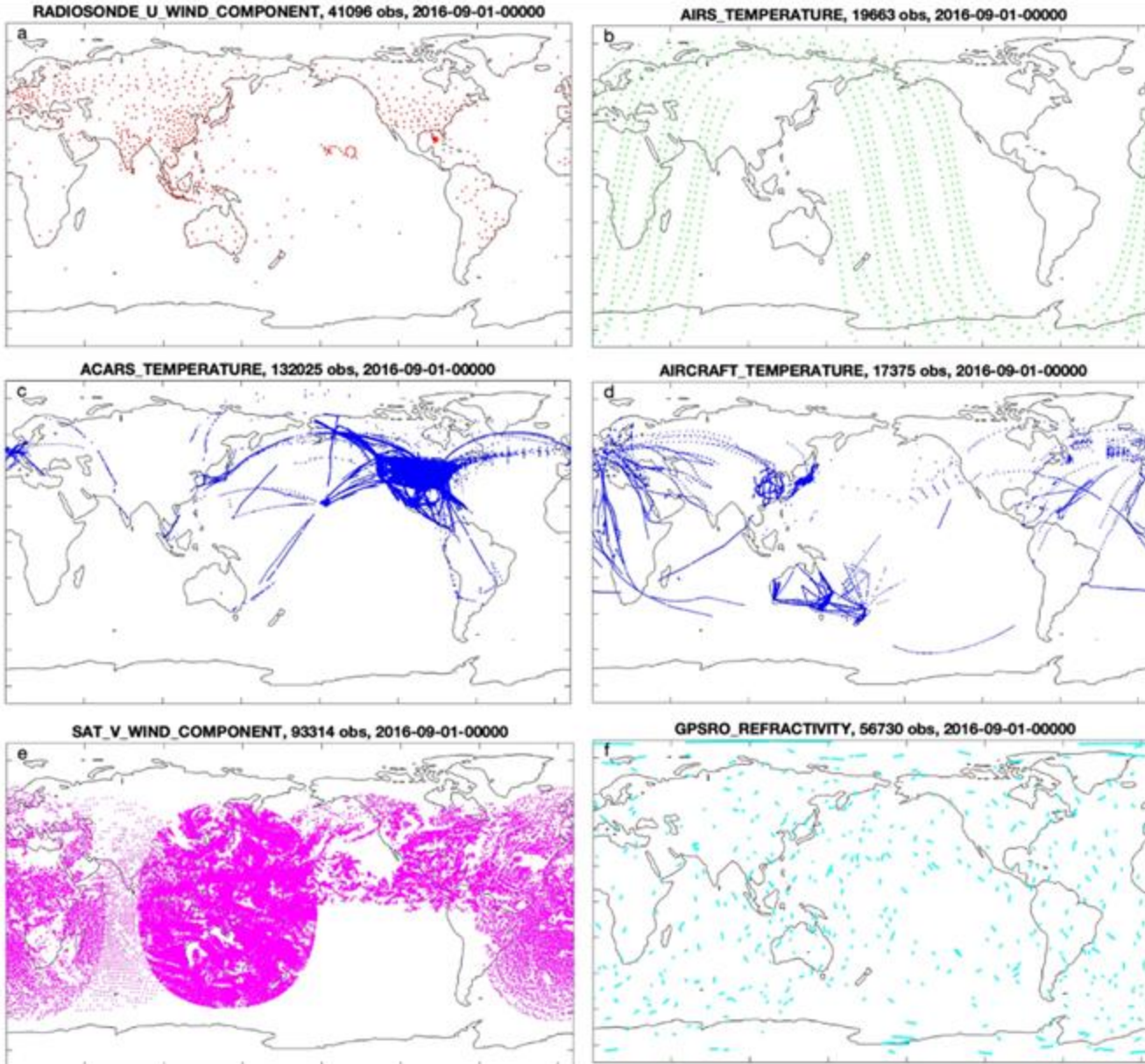
Example of a parameter that converged



Example of a parameter that shows a seasonal cycle closely related to the blooming periods



Atmospheric Reanalysis Product: Community Atmosphere Model (v6)

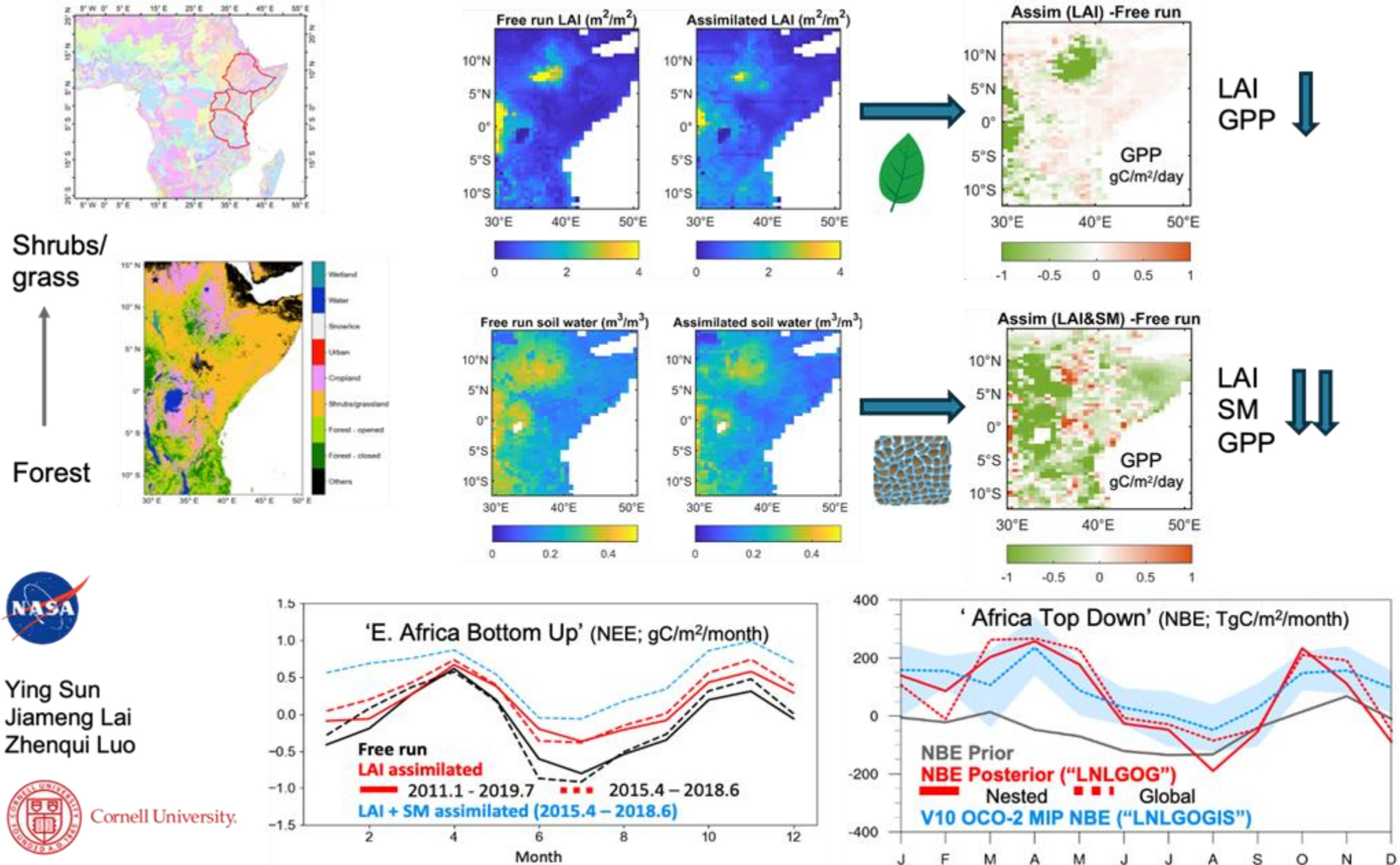


CAM6-DART 10 Year Reanalysis:

- 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/>

Land Surface Reanalysis: Carbon & Drought Monitoring (Africa)

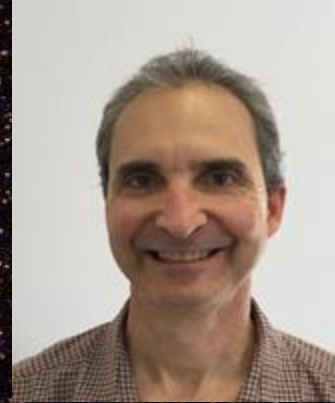
- Carbon and drought monitoring for remote, vulnerable ecosystems with MODIS leaf area (LAI) and SMAP soil moisture (SM) observations



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

David John “DJ” Gagne
Machine Learning Scientist II, MILES

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

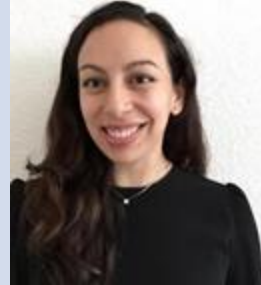
MILES Core



**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*

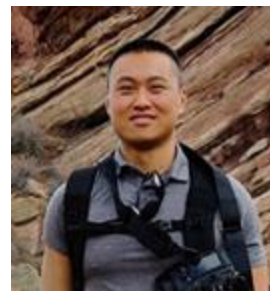


**Will
Chapman**
*Project
Scientist I*

MILES+LEAP



Da Fan
Postdoc



**Wayne
Chuang**
*Integration
Engineer*

MILES+WRaDS



Julie Demuth
*Project
Scientist III*



**Mariana
Cains**
*Project
Scientist I*



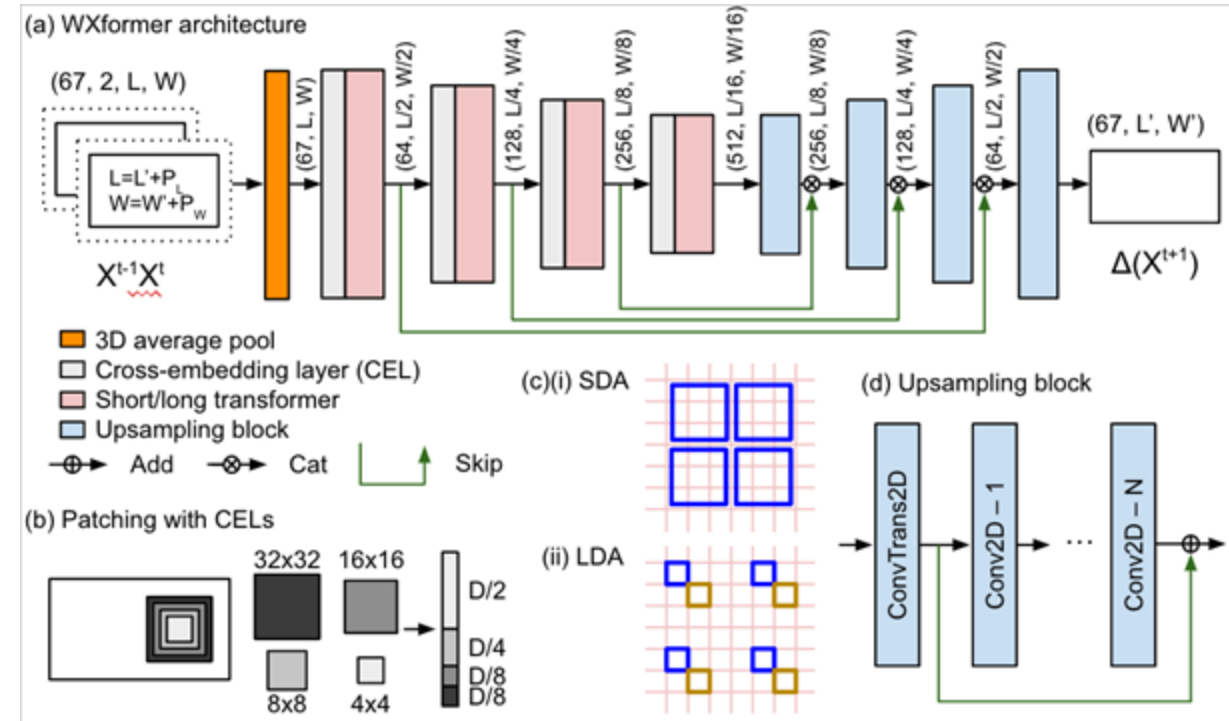
Chris Wirz
*Project
Scientist I*



Jacob Radford
*Research
Associate
CIRA*

NSF NCAR Community Research Earth Digital Intelligence Twin (CREDIT)

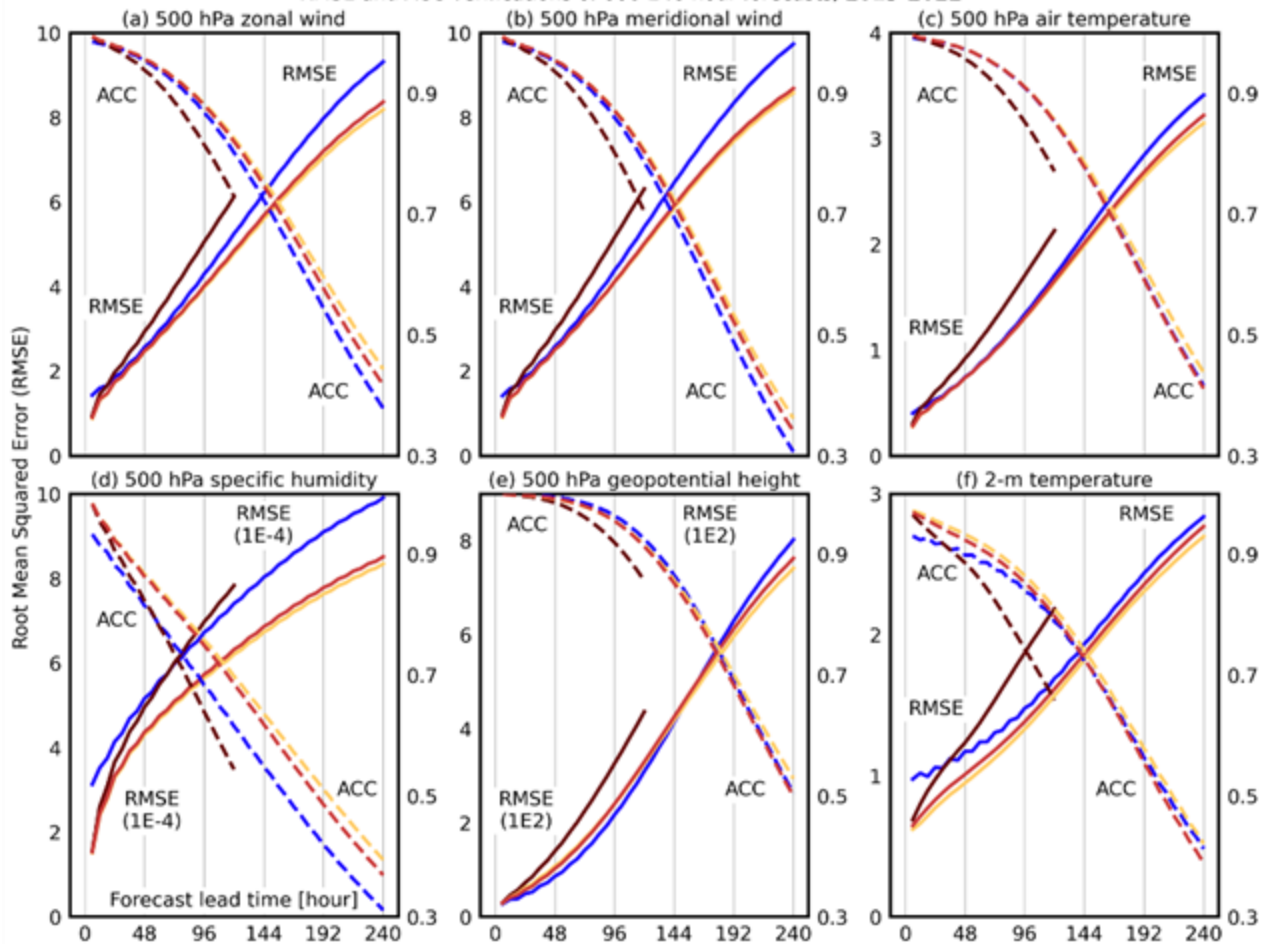
- Research platform for understanding of practices for training and operating global and regional AI weather prediction models
- Platform Features
 - Integrated pre-processing for reanalysis or reforecast data
 - Library of existing and new PyTorch neural network weather prediction architectures
 - Scalable 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

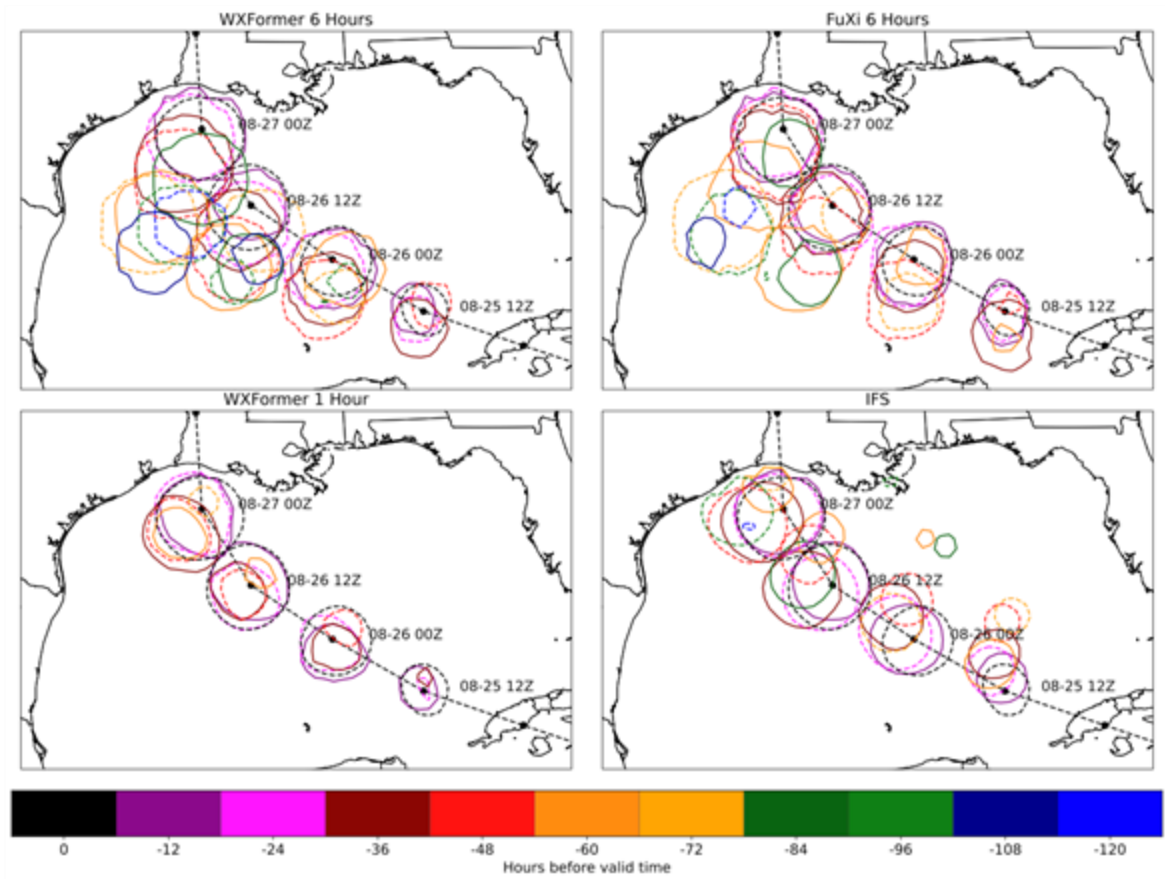
CREDIT Verification

RMSE and ACC verifications of 006-240 hour forecasts, 2019-2022^[*]

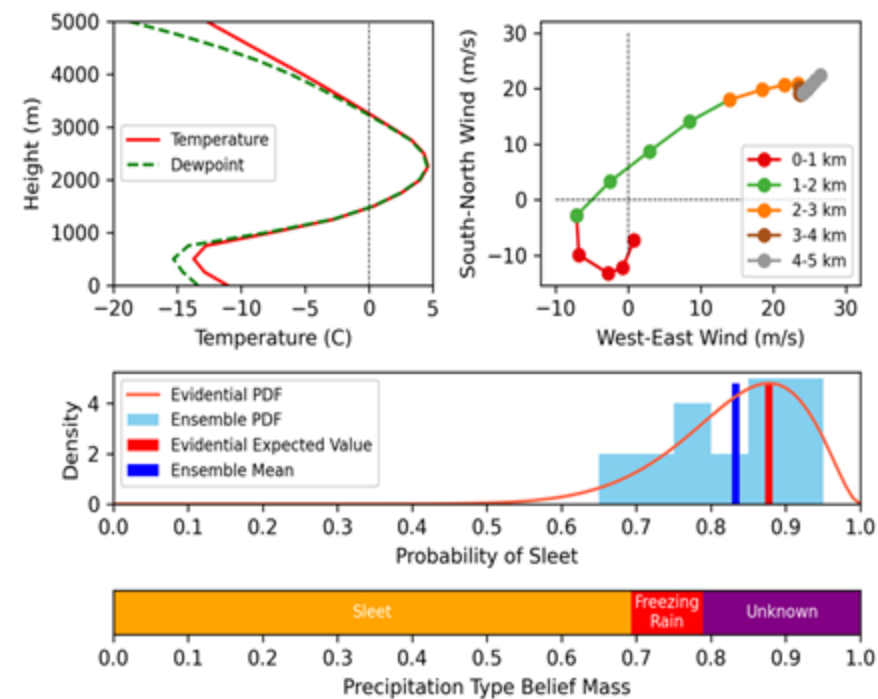


- RMSE - 6 hourly IFS
- ACC - 6 hourly IFS
- RMSE - 6 hourly FuXi
- ACC - 6 hourly FuXi
- RMSE - 6 hourly WXFormer
- ACC - 6 hourly WXFormer
- RMSE - hourly WXFormer^[**]
- ACC - hourly WXFormer^[**]

[*] Verification scores were computed using the ERA5 and averaged over 100 bootstrap replicates.
 [**] Hourly WXFormer was verified on 006-120 hour forecasts.

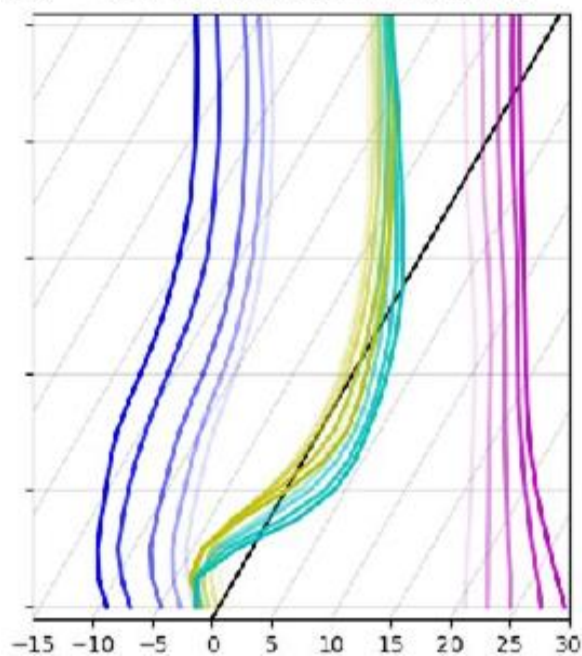


Machine Learning Uncertainty Quantification

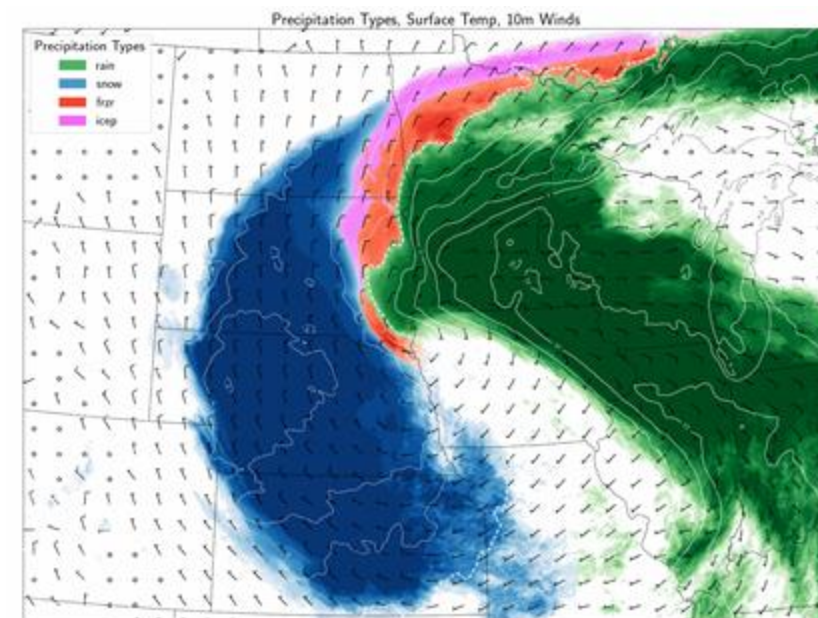


ML models can represent predictive uncertainty through ensembles or higher-order parametric distributions.

(iii) DST uncertainty, Evidential MLP



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>



Trainings and Communities

Negin Sobhani

CISL, Consulting Services Group (CSG)

Earth System Data Science (ESDS) Initiative

Vision

Profoundly **increase the effectiveness** of the NSF NCAR and UCAR workforce by promoting deeper **collaboration centered on analytics**, improving our capacity to deliver impactful, actionable, reproducible science and serve the university community by **transforming how geoscientists synthesize and extract information** from large, diverse datasets.

Approach

Cultivate a **community of practice** encouraging:

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

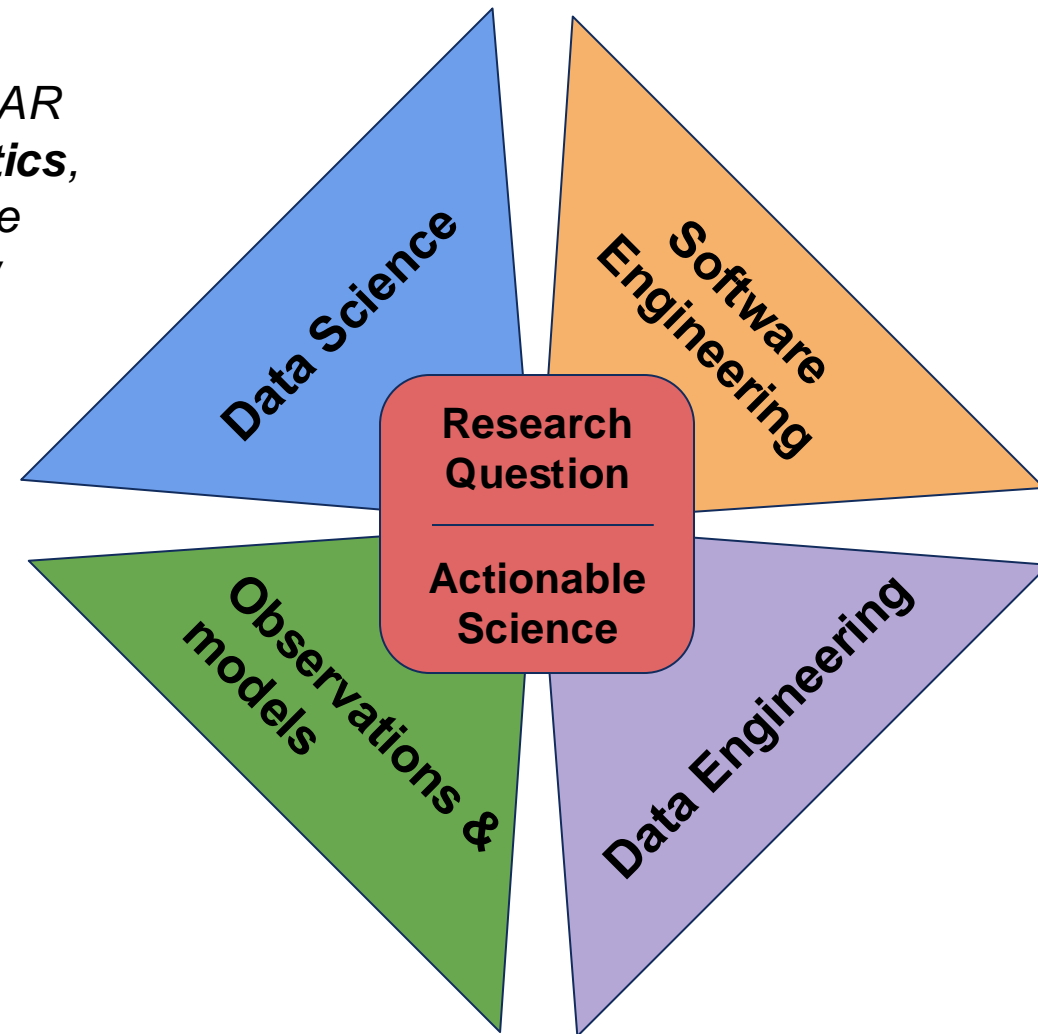
ESDS Communication and Resources

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, events, and more



UCAR Software Engineering Assembly (SEA)

The mission of the SEA is to build community, share knowledge, and foster collaboration among all UCAR/NCAR employees with an interest in **scientific software**.

Anyone, regardless of job category or LCPO, is welcome to join!



SEA Activities

- [Improving Scientific Software \(ISS\) Conference](#)
- Discussions on software development
- Tutorials on new hardware and software tools
- Getting involved in the wider software engineering/development community at UCAR

SEA Resources & Communication

- SEA site (*news and links*)
<https://sea.ucar.edu>
- SEA mailing list (*news and outreach*)
<https://mailman.ucar.edu/mailman/listinfo/sea/>
- SEA Committee (*ask us questions!*)
sea_committee@ucar.edu
- SEA Slack Workspace (*community discussion*)
<https://ucarsea.slack.com>

NCAR HPC User Group (NHUG)

NHUG is a dedicated community that aims to promote the ***productive use of high-performance computing (HPC) facilities at NCAR*** and increase collaboration among all of the NCAR HPC community.

NHUG Activities and Resource

- 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 Resources & Communication

- NHUG Page
<https://nhug.readthedocs.io/en/latest/>
- NHUG mailing list (*news and outreach*)
https://ncar.pub/NHUG_emailist
- NHUG Slack Workspace (*community discussion*)
https://ncar.pub/join_nhug_slack

Project Pythia and Pythia Cookoffs



- 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



Pythia Cookoffs

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



Supporting Python Learning



Dask Cookbook

DOI: [10.5281/zenodo.8157213](https://doi.org/10.5281/zenodo.8157213)



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](#).

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>

OpenACC Hackathons

Two hackathons hosted by NSF-NCAR in partnership with NVIDIA/OpenACC, NREL, and NOAA with goals of :



- Accelerate scientific software by porting code to GPUs.
- Enhance software performance on HPC, reduce core hour cost.
- Promote open science and collaboration across labs.
- Facilitate cohesive teams that build relationships between scientists and research software engineers.
- **Next event: Feb 18-27, 2025.**



NOAA/NCAR/NREL Open Hackathon

Dates: February 18-27, 2025

Event Focus: HPC+AI

Event Format: Hybrid Event

Region: North America/Latin America

Application Status: [Open](#)

SIParCS - Summer Internships in Parallel Computational Science

Mailing list



- 11-week Summer internship
- Open to: Upper division undergrads & Graduate students
- Domestic & International 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



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



Questions?

CISL_Roadshows@ucar.edu



<https://www.cisl.ucar.edu/events/agu2024>

Backup slides



SIParCS Internship Program

Summer Internships in Parallel Computational Science

TBD
CODE