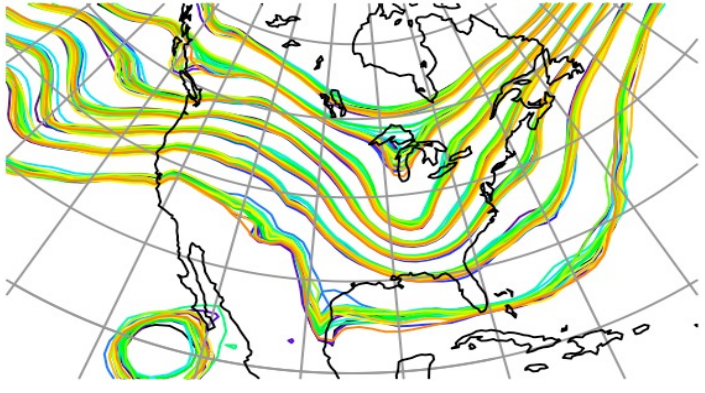
The Data Assimilation Research Testbed; a Suite of Tools for Understandingthe Earth System with Confidence

Kevin Raeder, Jeff Anderson, Moha El Gharamti, Helen Kershaw, Brett Raczka, Soyoung Ha, Craig Schwartz, Shixuan Zhang, Kai Zhang, Hui Wan, Benjamin K. Johnson, Ibrahim Hoteit, Matt Mazloff, Daniel Hagan, James McCreight, Tim Hoar, Nancy Collins



The National Center for Atmospheric Research is sponsored by the National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

NCAR | National Center for
UCAR | Atmospheric Research

DART Is:

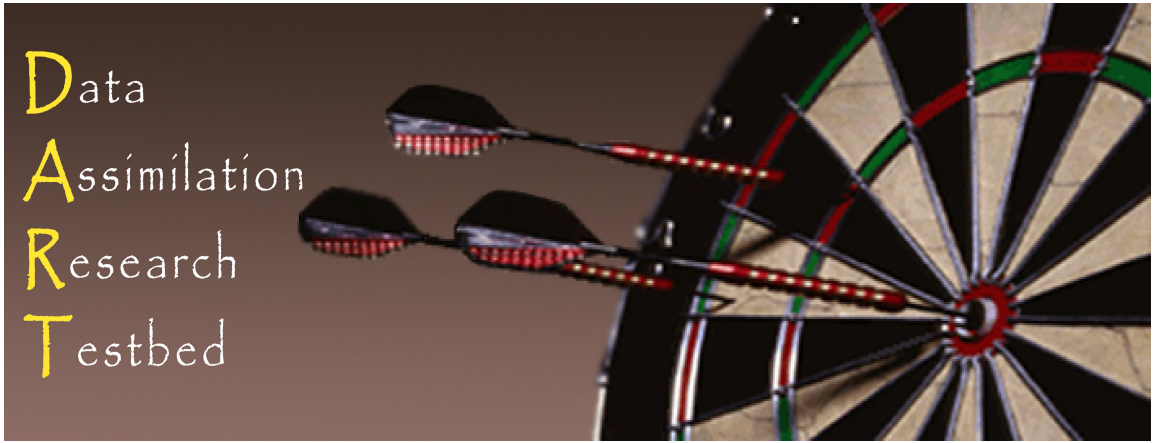
A flexible suite of software tools to accelerate Earth system research using ensemble Kalman filters.

Focused on

Educational Resource

Used at:

- 50+ Universities
- 100+ other sites
- 1500+ registered users

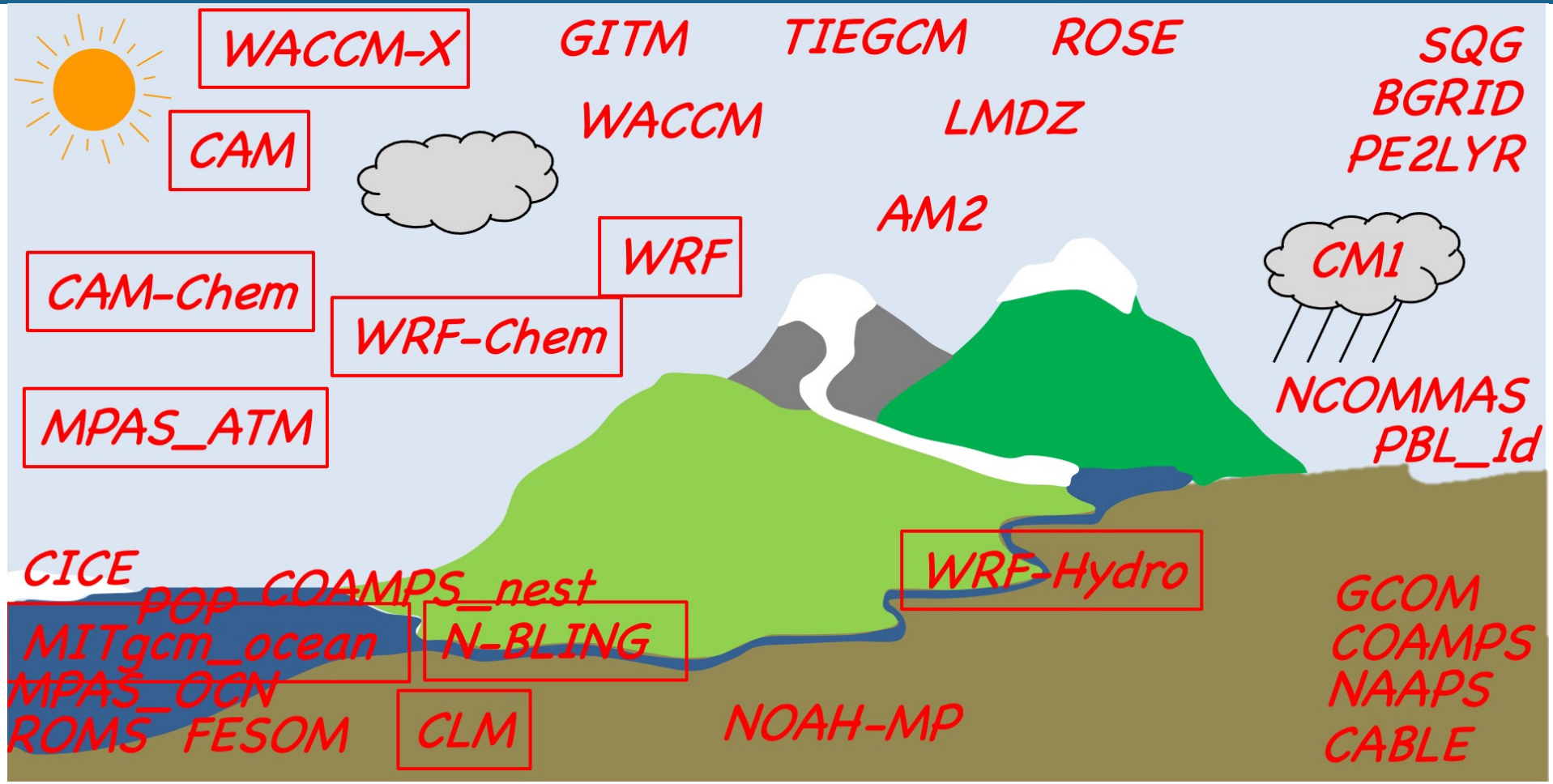


Open Source. Community members develop:

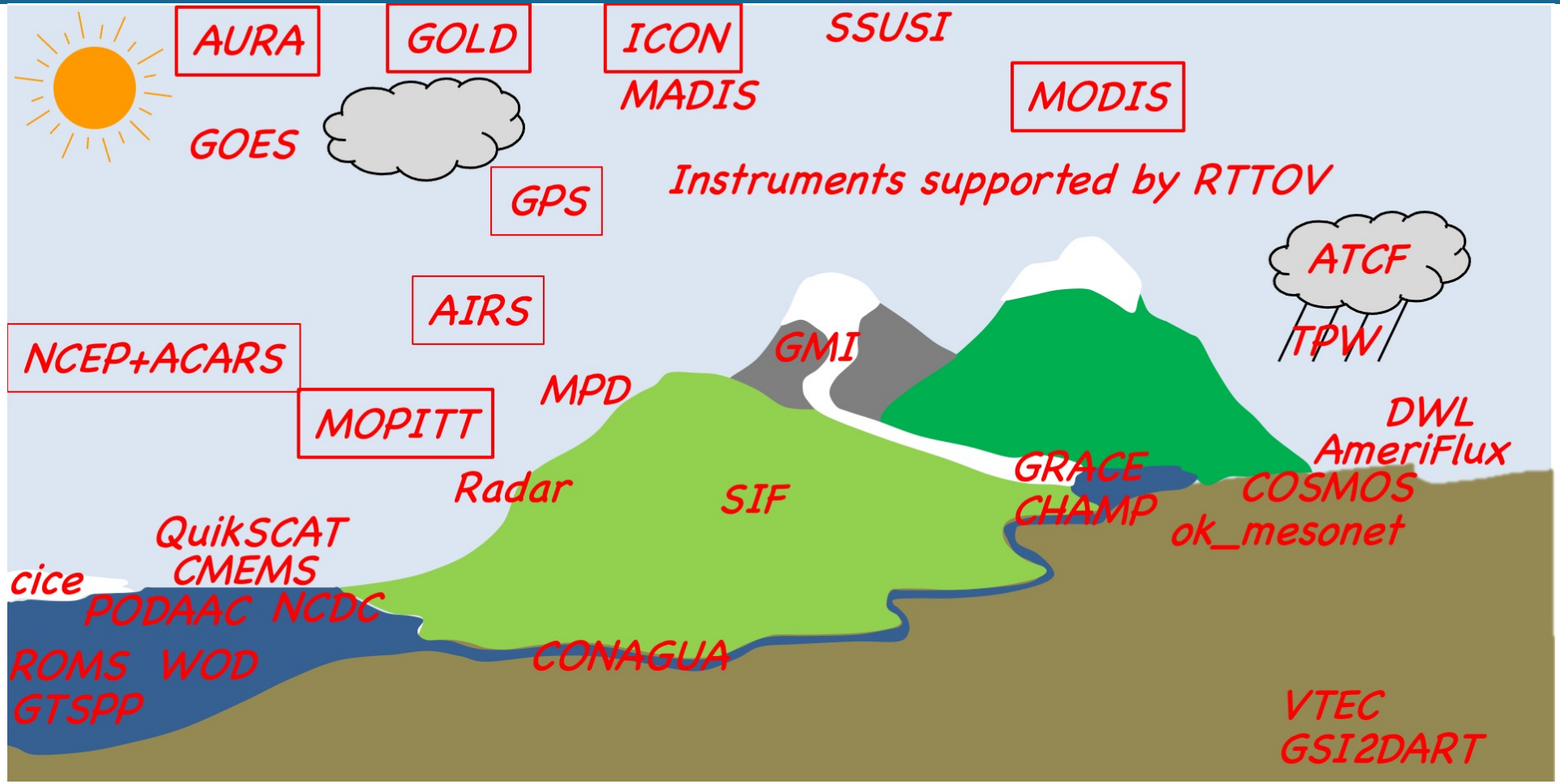
- model interfaces
- observation forward operators
- assimilation algorithms

Contributions are reviewed, streamlined, and tested before merging into the public DART.

Geophysical Models Interfaced to DART



Earth System Observations (others available)

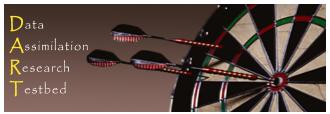


DART's Algorithms (a sampling)

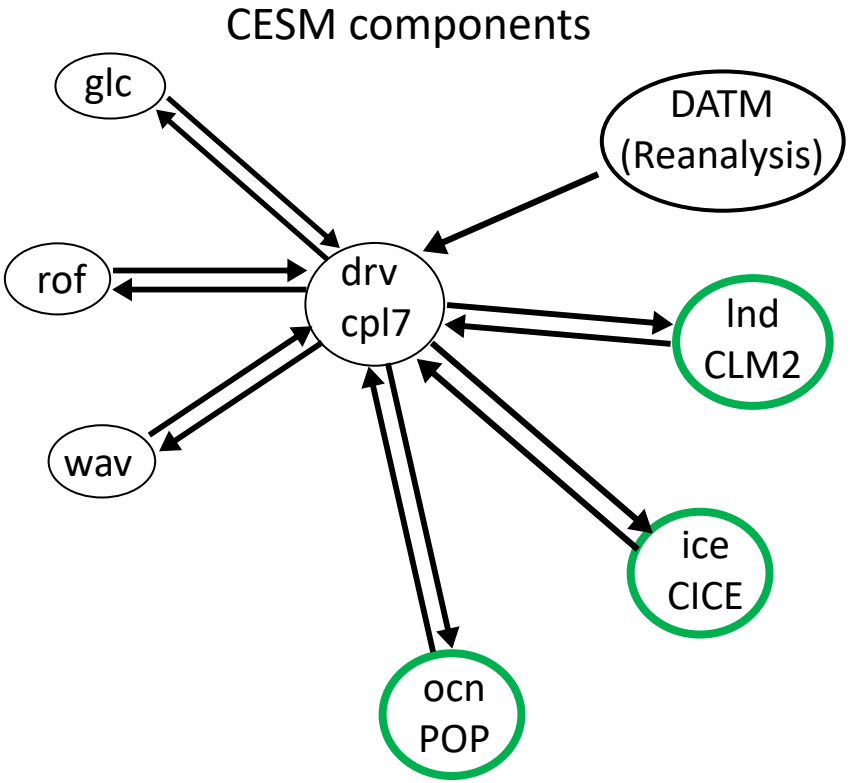
- Assimilation Flavors (9+):
  - ✓ Deterministic and stochastic ensemble Kalman Filters
  - ✓ Non-Gaussian rank histogram filters
  - ✓ Localized Particle Filter (Poterjoy)
  - ✓ Gamma/Inverse Gamma, Inverse Gamma/Gamma filters (Bishop)
  - ✓ Higher moment filters (Hodyss)
  - ✓ (coming; **Quantile Conserving Ensemble Filter** (Anderson))
- Ensemble Inflation; state-space, prior and posterior, adaptive, **inverse Γ**, damping
- "Localization"; spatial and **by variable** (esp. for chemistry)
- **Sampling Error Correction**, Spread Restoration, Sort Obs. Increments, Rank Regression
- Output 6 stages of assimilation in state space, plus observation space
- Quality Control; detailed reporting
- Compact enough for laptops, **scales to thousands of processors** (one-sided MPI, distributed states and mean)

Designed for flexible research and development,
including computationally intensive ideas.

NCAR | National Center for
UCAR | Atmospheric Research



Atmospheric Forcing of Surface Components; CAM6+DART Reanalysis



Lead; Raeder

Surface models in CESM2 (CLM, POP, CICE, ...) are forced by CAM6. DA using any of these can use an existing CAM6 reanalysis instead of re-running a CAM6 ensemble for each new case.
Reanalysis ≅ actual atmosphere.

Cpl history files:

- frequencies ranging from 1-6 hours
- ready to use in CESM in DATM mode
- 1 year, 1 member per file
- 2011-2019 (2020 soon)

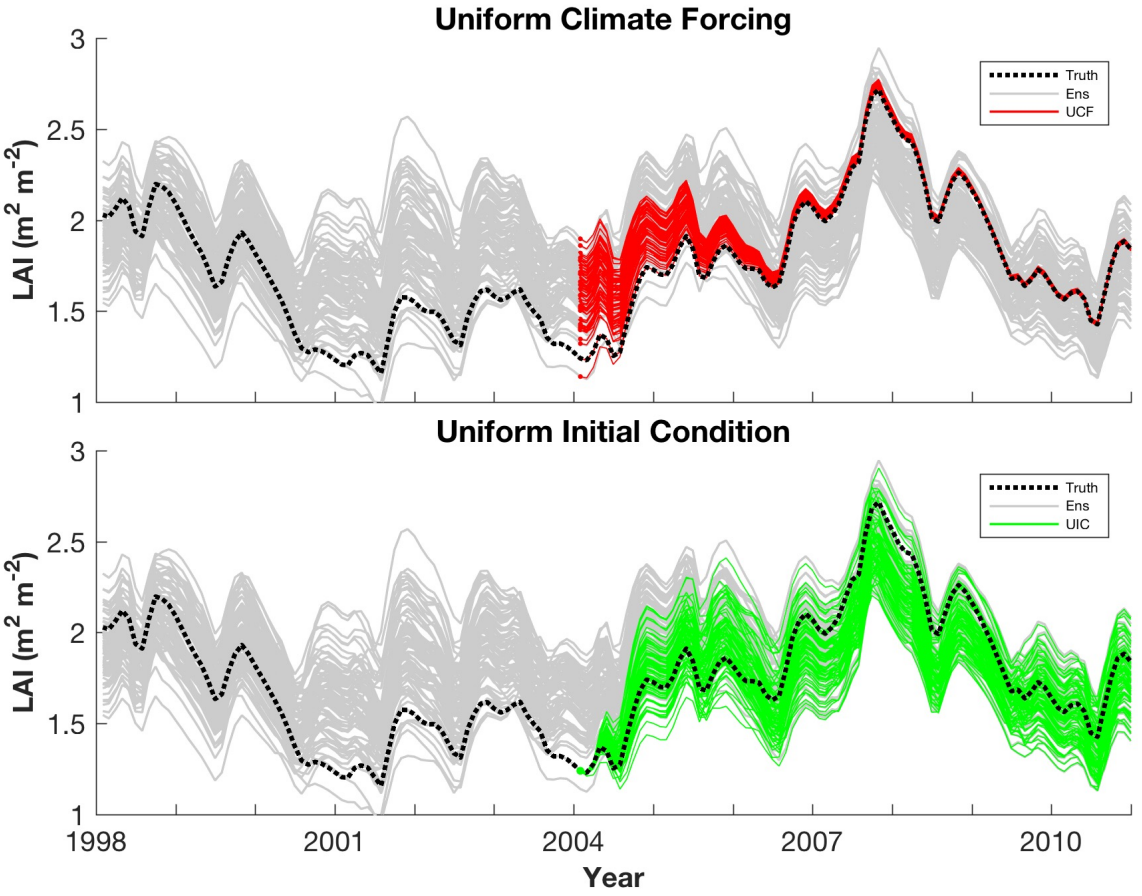
These models have DART interfaces for assimilation.

The CAM6+DART Ensemble Reanalysis

DA with surface models, such as CLM5 or WRF-Hydro, requires not only a good model, but good forcing from the atmosphere, both in the mean and ensemble spread.

Data available at https://rda.ucar.edu/datasets/ds345.0

Reanalysis description in Scientific Reports: https://rdcu.be/ctUVQ



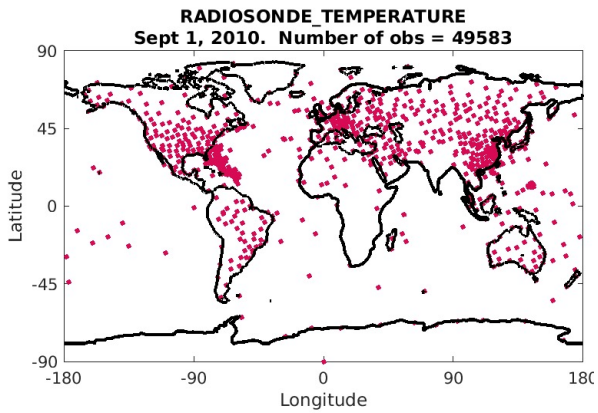
NCAR | National Center for Atmospheric Research
UCAR | Atmospheric Research

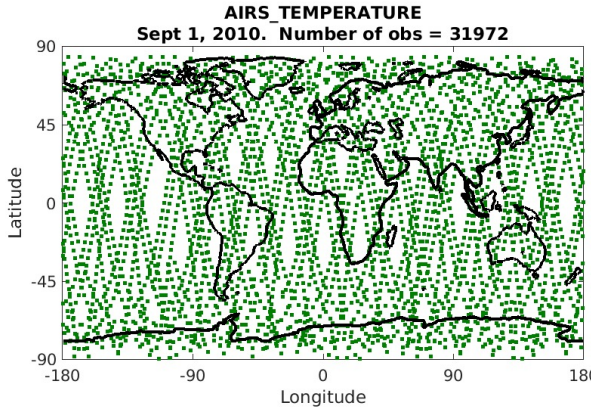
Data Assimilation Research Testbed

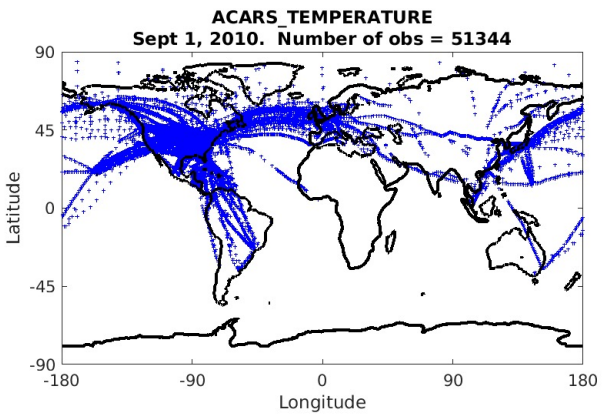
Reanalysis Quick Facts: Assimilation

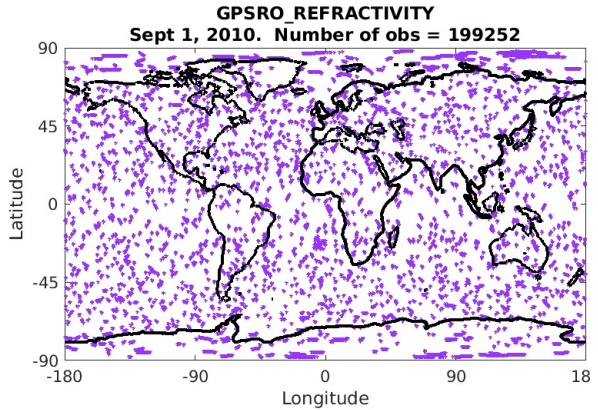
- DART Manhattan
- Ensemble Adjustment Kalman Filter (EAKF)
- 80 members with Sampling Error Correction
- 6-hour window
- Inverse Γ adaptive inflation
- Tuned parameters for localization, inflation, etc.
- Land state well spun up; in balance with atmosphere(s).

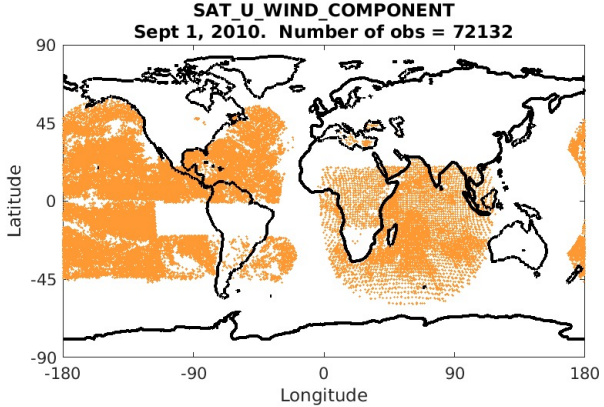
Reanalysis Quick Facts: Observations

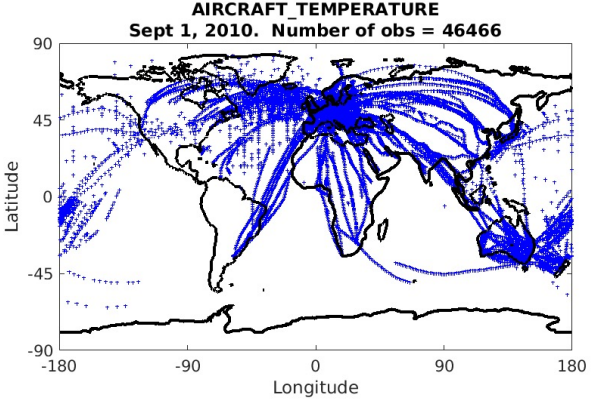






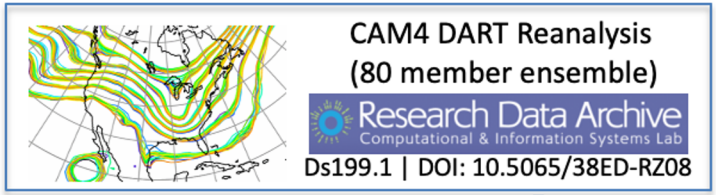


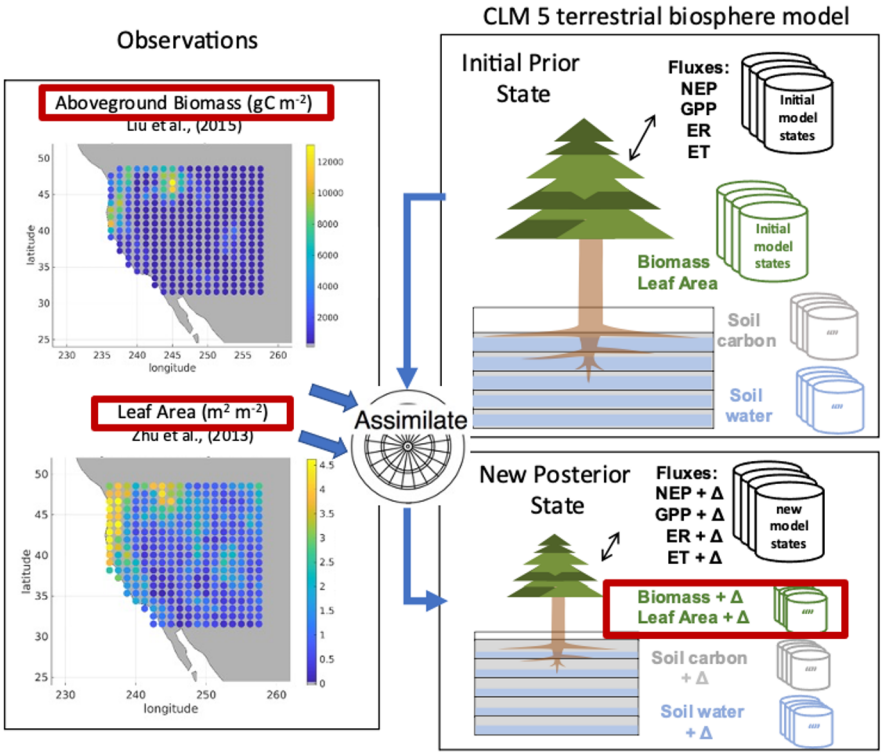




Example of observations used in 1 cycle; > 450,000 in this window.

CLM5.0 Biomass and Carbon Exchange



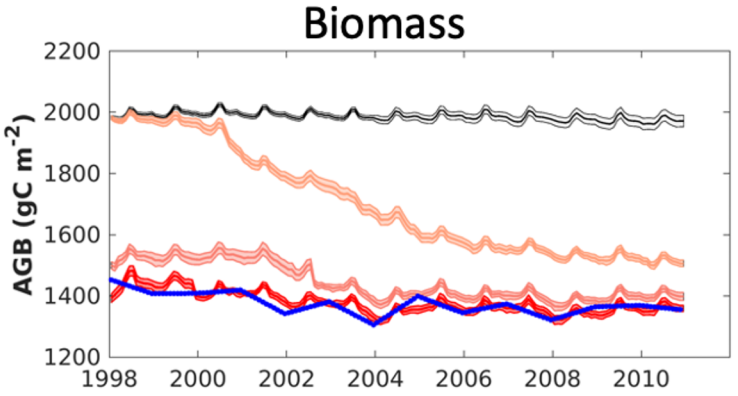


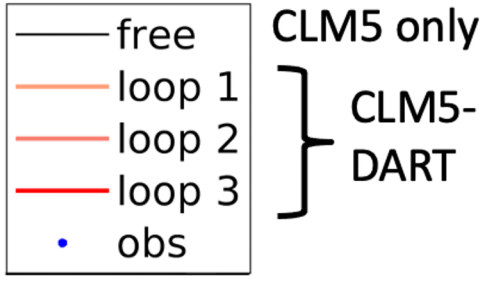
THE UNIVERSITY OF UTAH

NASA

Lead; Brett Raczka

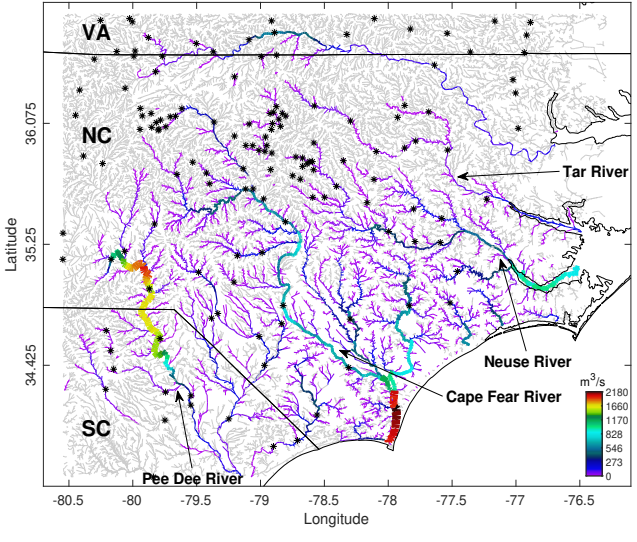
NCAR | UCAR | National Center for Atmospheric Research



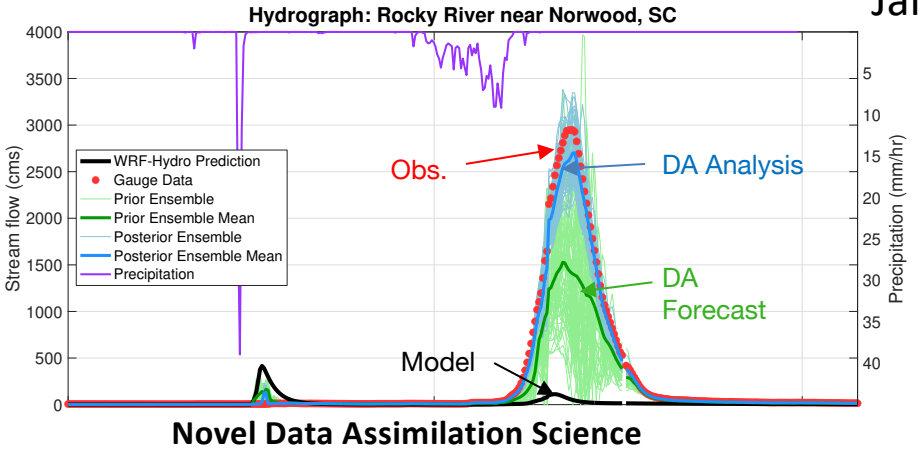


Flood Prediction: WRF-Hydro/DART for Hurricane Florence 2018

High-resolution stream network with USGS streamflow gauges.

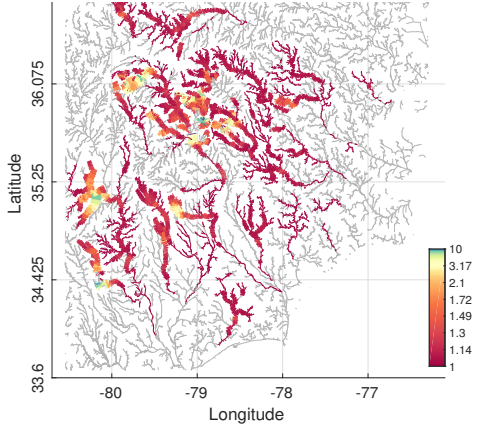


DA greatly improves analysis and forecasts of streamflow.

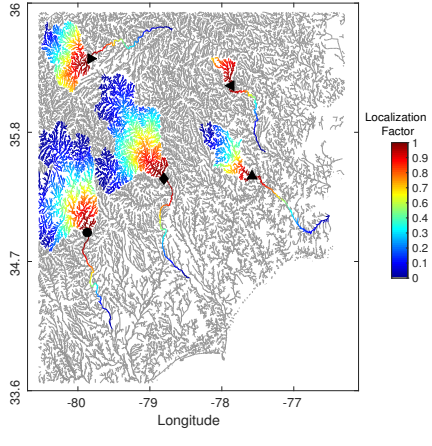


Novel Data Assimilation Science

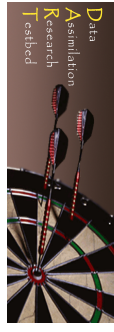
1. Prior and Posterior Adaptive Covariance Inflation



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



Moha el Gharamti
James McCreight

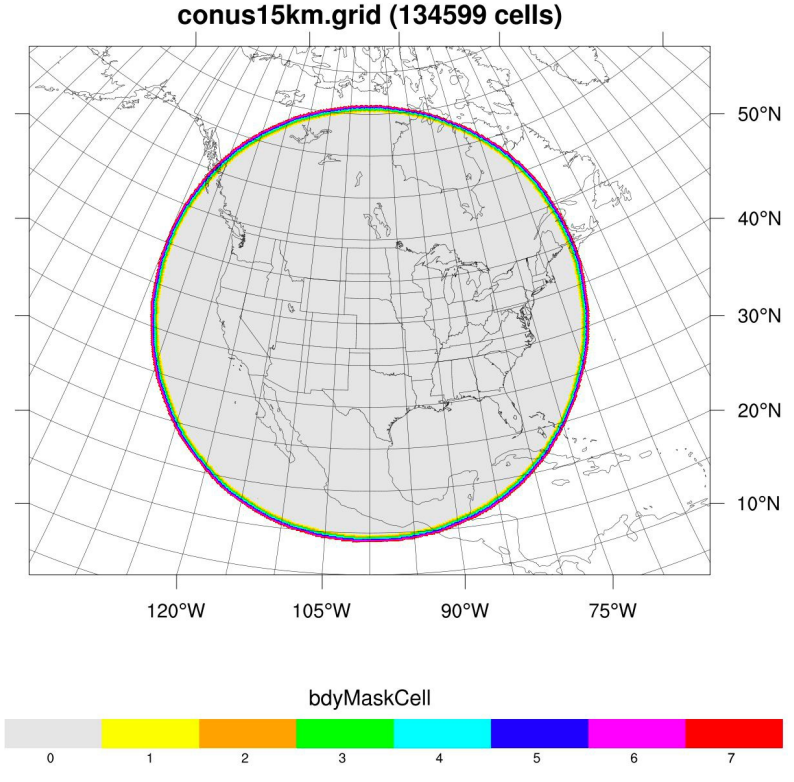


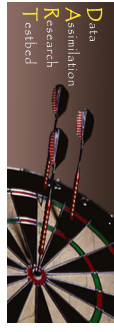
NCAR | National Center for Atmospheric Research
UCAR

Ensemble Data Assimilation using the Regional MPAS model

Lead: Soyoung Ha, Craig Schwartz

Generation of the Voronoi unstructured mesh over the region of interest (CONUS).
Ensemble IC/LBCs for the region from 80-member global MPAS/DART ensemble forecasts on the uniform 15-km mesh

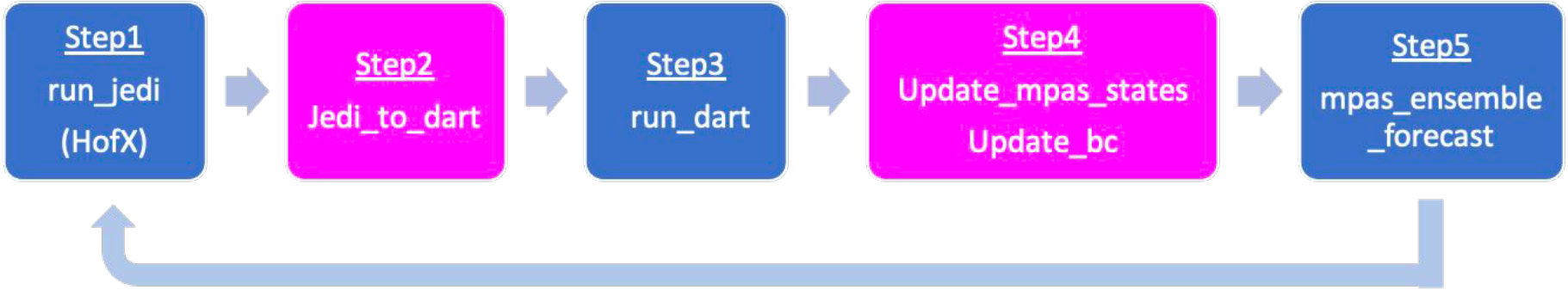




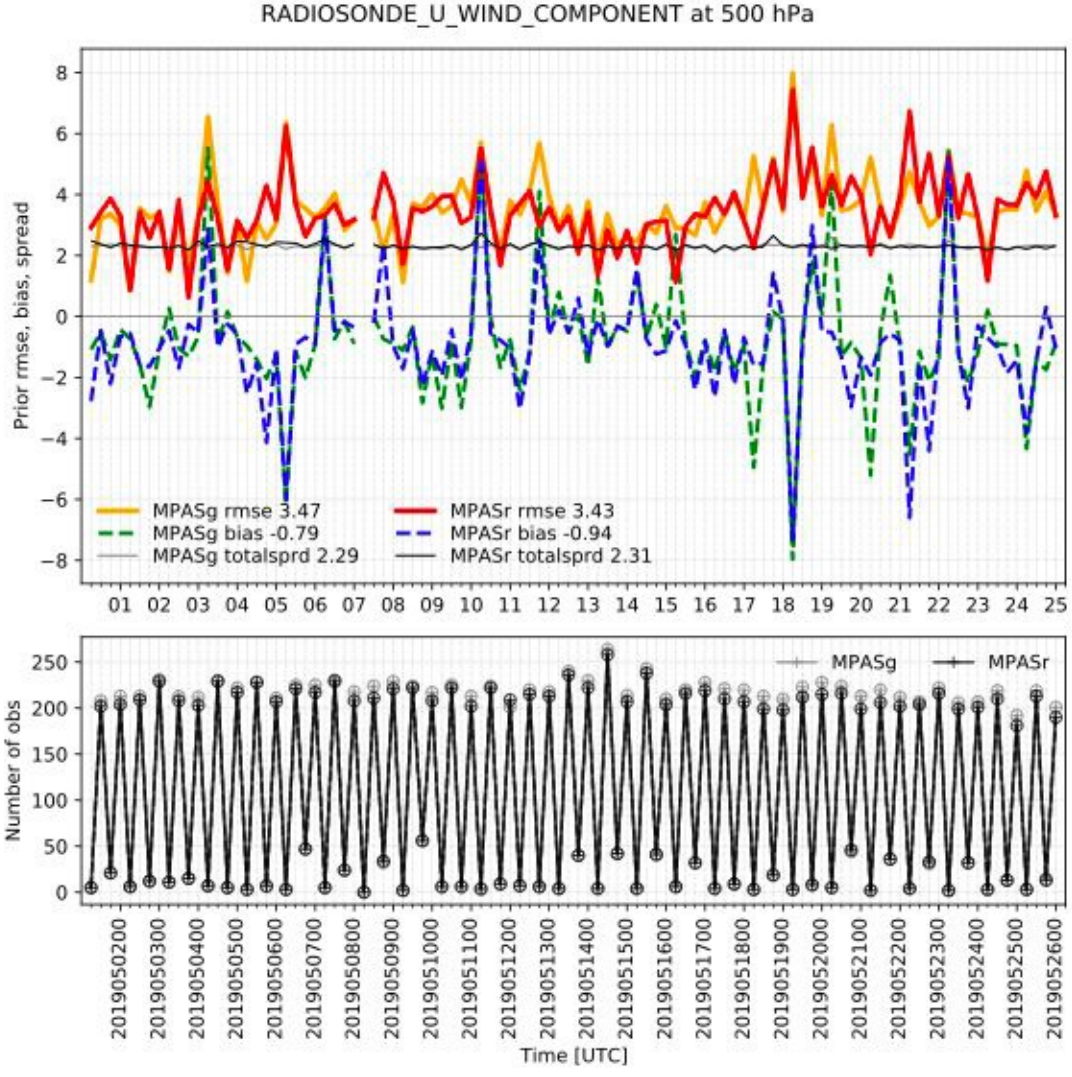
NCAR UCAR | National Center for Atmospheric Research

A Regional MPAS Ensemble DA system - Analysis (DART-JEDI)

- ✓ The Manhattan version of MPAS/DART (as of Jan 2021) includes Update_bc: After the analysis, the relaxation zone is updated through blending the posterior and the prior states
- ✓ JEDI/UFO for observation operators
- ✓ Assimilates multiple satellites and conventional observations in the global system
- ✓ Bias correction, adopted from NCEP GSI and implemented in DART

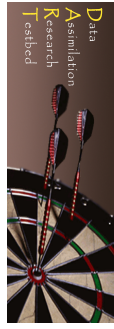


Glimpse Comparison of Regional and Global MPAS Assimilations



Essentially the same RMSE and bias relative to the radiosonde U wind observations.

Large spikes are almost all at times with few observations (lower panel); small number statistics.



NCAR | National Center for Atmospheric Research
UCAR

Shixuan Zhang, Kai Zhang, Hui Wan (PNNL)

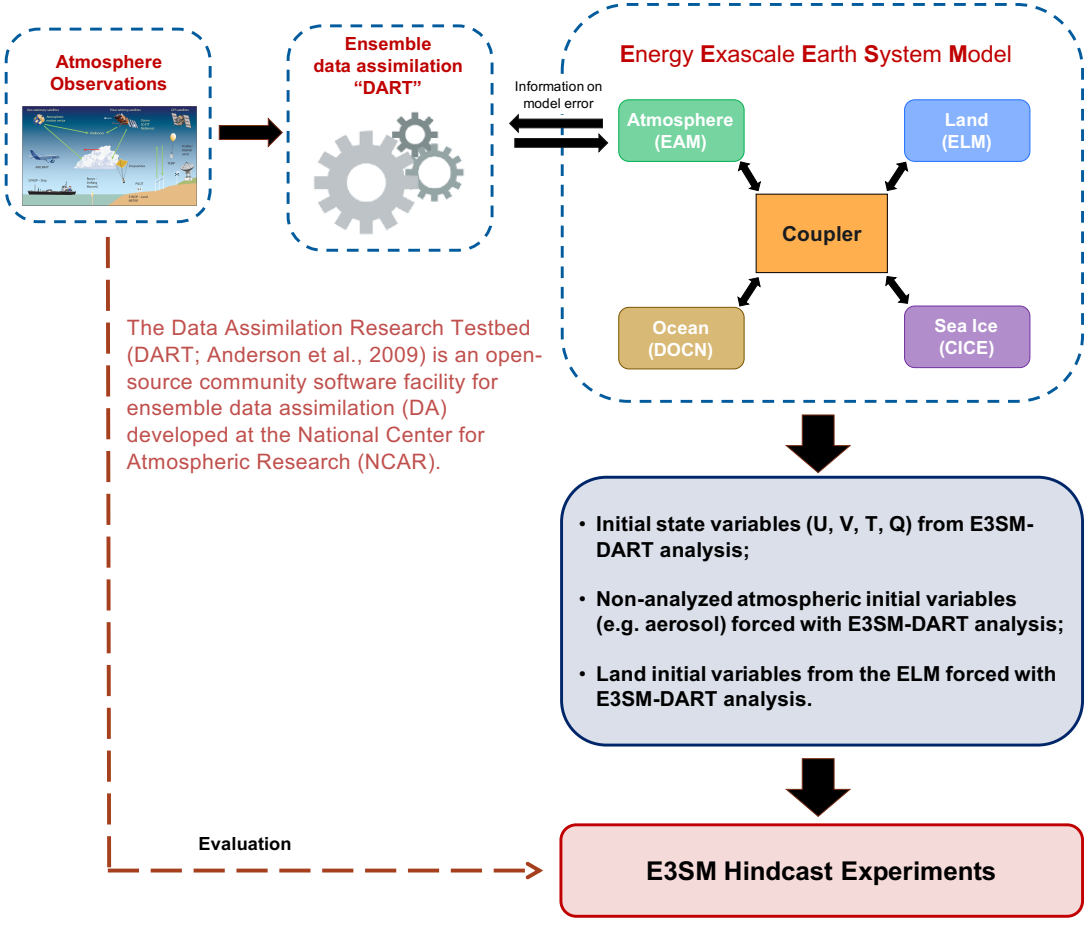
- **The need:** running climate model in “weather forecast” mode can facilitate more rigorous testing and improvement of parameterizations of “fast” physics process (Phillips et. al., 2004, Ma et al. 2015).

- **The gap:** existing capabilities using global reanalysis and nudging (Zhang et. al. 2014; Sun et. al. 2019, Ma et. al. 2015) to initialize climate model hindcast simulations have limitations. e.g. the “initial shock” problem because of the inconsistencies between reanalysis and climate model (Ma et. al., 2015)

Acknowledgements: *This research was supported by the Laboratory Directed Research and Development Program at Pacific Northwest National Laboratory (PNNL), a multiprogram national laboratory operated by Battelle for the U.S. Department of Energy under contract DE-AC05-76RL01830.*

NCAR | National Center for
UCAR | Atmospheric Research

Key messages

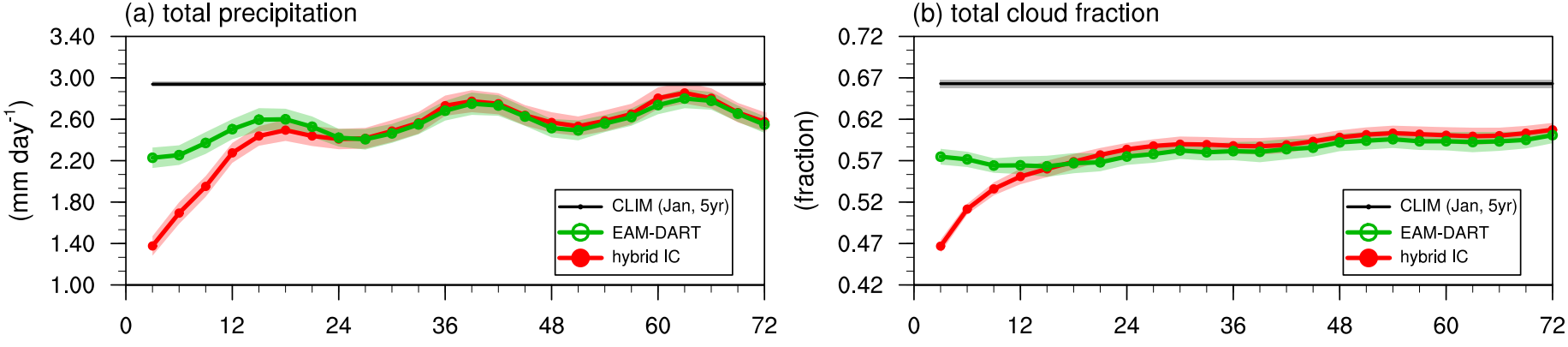


- High-quality initial conditions for the atmospheric component of E3SM can be achieved with EAM-DART.
- Self-consistent initial condition from EAM-DART reduces the "initial shock" in the short-term hindcast simulations.
- EAM-DART ensemble hindcast shows better error correspondence between short-term and long-term systematic errors in atmospheric component of E3SM.

Initial shock in Hindcasts

Three simulations to assess the benefit of DART DA system to the E3SMv0 hindcasts

- **CLIM:** 5-year (2011-2015) free-running simulations using a method similar to the Atmospheric Model Intercomparison Project (AMIP).
- **EAM-DART:** daily, 5 day, free running, E3SM, **ensemble** simulations for January 2011 with all prognostic variables of the atmosphere and land models initialized using output from E3SM-DART analysis.
- **Hybrid IC:** Same as EAM-DART, but not ensemble, and initialized using a hybrid method; u, v, T, Q, PS from ERA5 reanalysis + other prognostic variables (e.g., clouds, aerosols, soil moisture) from nudged E3SM (atm+land) simulations.



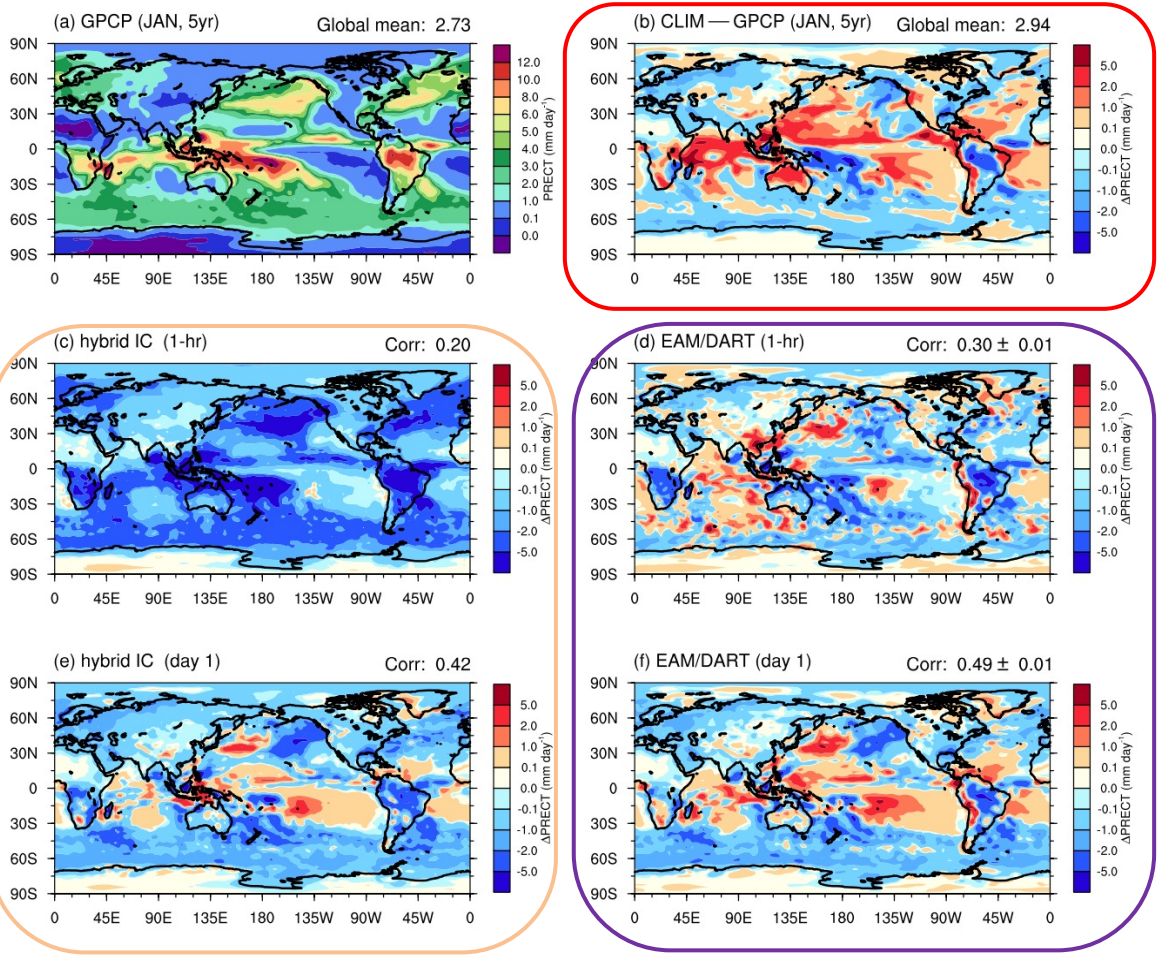
EAM-DART initialized hindcasts exhibit significantly less initial shock (0-24h) when compared to hindcast initialized with a combination of ERA5 reanalysis and output from the nudged simulations (hybrid IC).

Total Precipitation Biases; Short-term Hindcast and Long-term Climate Simulations

5 year

Hour 1

Day 1



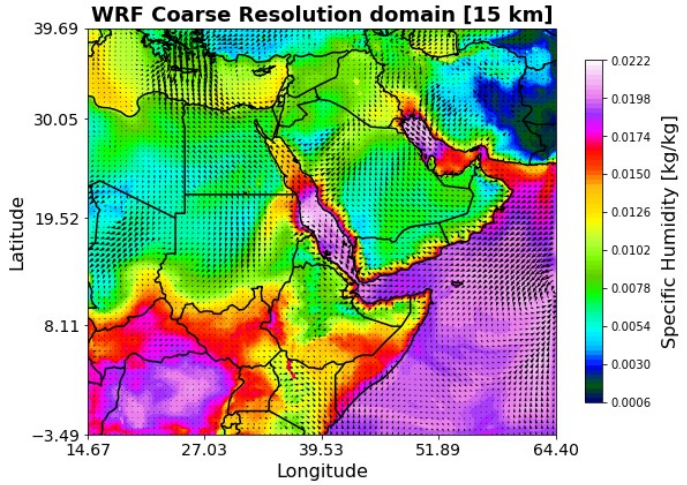
Simulations:

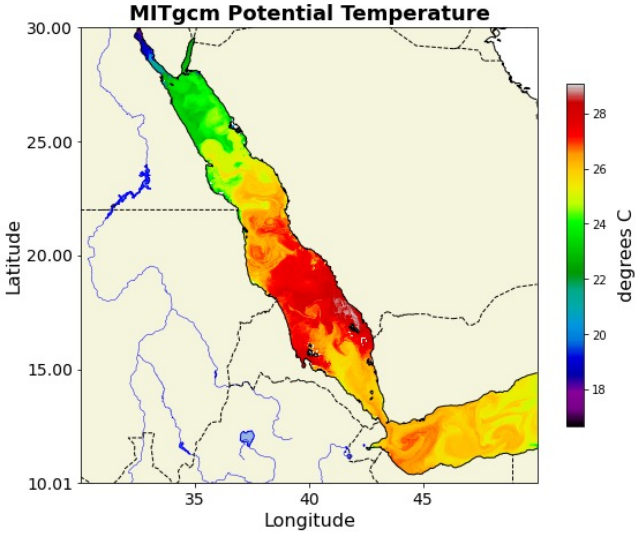
- **CLIM**: 5-year AMIP simulation
- **hybrid IC:** single hindcast
- **EAM-DART:** 80-member ensemble hindcasts

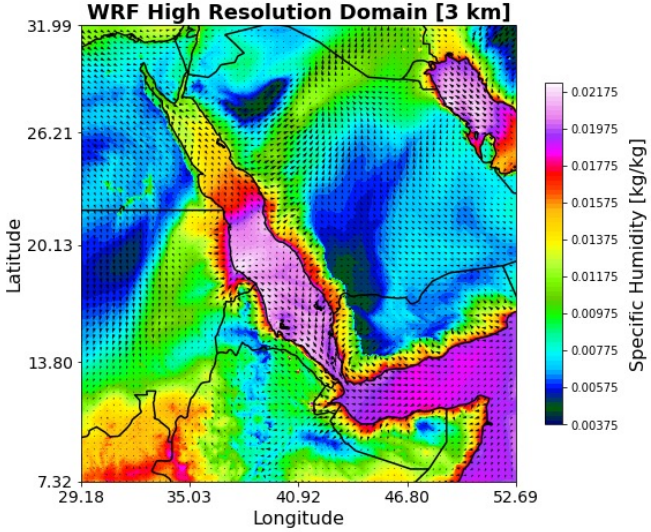
EAM-DART initialized hindcasts show better error correspondence between short-term hindcast and long-term climate simulations

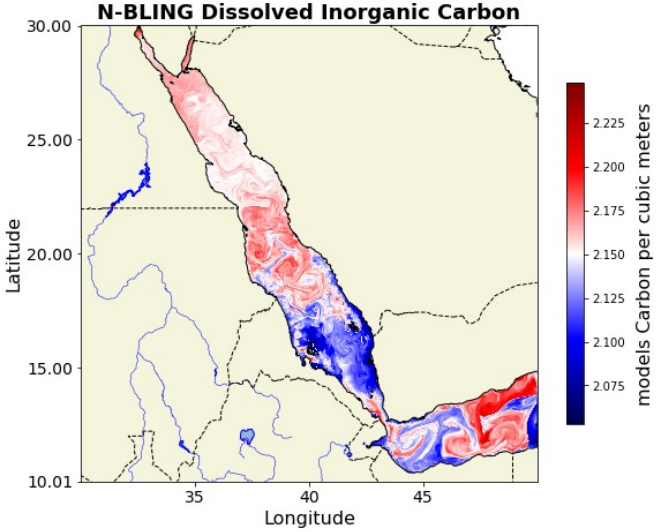
- **Most of the systematic model errors of precipitation can already be detected by EAM-DART hindcast after 1-h model integration.**
- **The patterns and amplitudes of climate error in precipitation are better captured even better by the day 1 EAM-DART hindcast.**

The Red Sea Initiative









M. El Gharamti, B. K. Johnson, I. Hoteit, M. Mazloff

KAUST+NCAR+Scripps regional coupled atmosphere-ocean-biogeochemical forecasting system

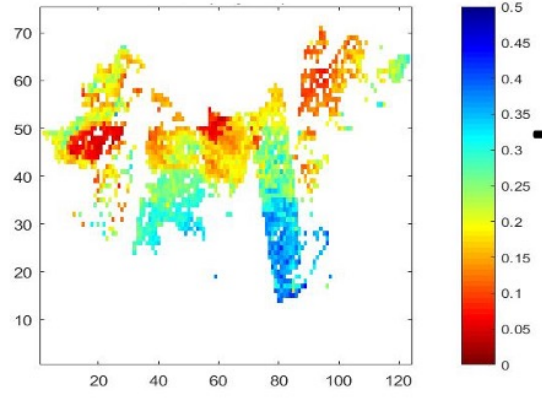
atm: WRF
ocn: MITgcm
ecosystem: N-BLING (Nitrogen-Biogeochemistry with Light, Iron, Nutrients and Gases)

NCAR UCAR | National Center for Atmospheric Research

Gap-filling the ECV soil moisture product using CLM-DART

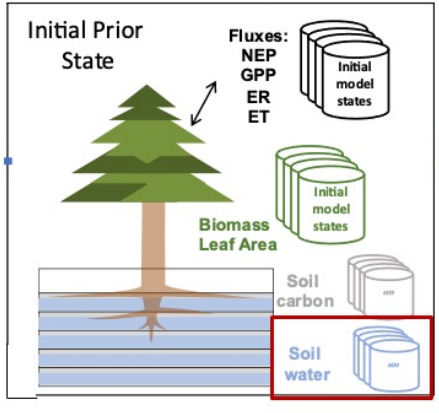
Daniel Hagan

ECV Soil Moisture Product ($m^3\ m^{-3}$)



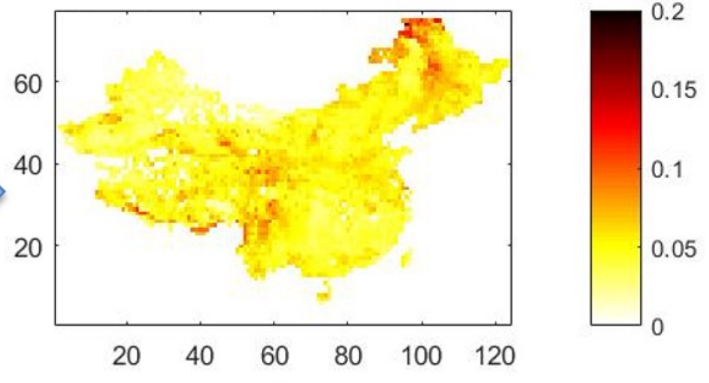
+

CLM-DART

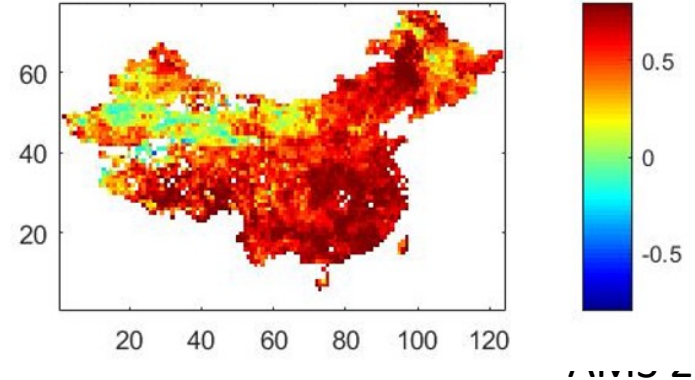


Compares favorably to GLEAM Soil Moisture Data Product (1998)

Unbiased RMSD ($m^3\ m^{-3}$)



Correlation



Nanjing University of Information Science & Technology

Other Recent DART Atmospheric Projects

AGU poster on ESSOAr (DOI 10.1002/essoar.10510010.1)

- Global air pollution
  Impact of Gas-phase Chemical DA on Aerosols using CAM-Chem (Gaubert)
  WRF-Chem and FRAPPE (Mizzi)
- Whole Atmosphere DA with WACCMX+DART (Pedatella).

Other Atmospheric Projects:

- Assimilation of radiance observations in RTTOV format at CMCC and KOPRI
- Toward Global Convection-Permitting Data Assimilation (Schwartz, Romine)
- Flow dependency of forecast errors in the tropics (Żagar)
- MPD Water Vapor Profile DA for Convective Weather Forecasts (Weckworth)

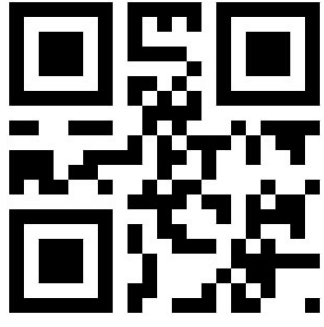
Many more are using DART without direct collaboration.

Summary and Resources

- DART is a flexible, research focused, community, ensemble DA system.
- It's used for a broad variety of Earth system research projects:
  - forecast model improvement; floods, weather, coupled models, ...
  - reanalyses (better picture of what actually happened)
- The CAM6+DART Reanalysis can accelerate research using non-atmospheric Earth system models at lower cost.
- DART ensembles provides objectively derived, realistic variability and uncertainty estimates.

https://dart.ucar.edu

dart@ucar.edu



CISL Computational & Information Systems Lab

We would like to acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation.

Extras

Other Motivations

2. Provide forcing for offline chemistry transport models or in a “nudging” framework.

3. Evaluate weather prediction capabilities of CAM.
- Confront climate model with observations.
- Identify systematic short-term forecast errors.
- Compare to earlier CAM reanalysis.

4. Very large, labeled data set of atmospheric observations + ensemble estimates; useful for machine learning.

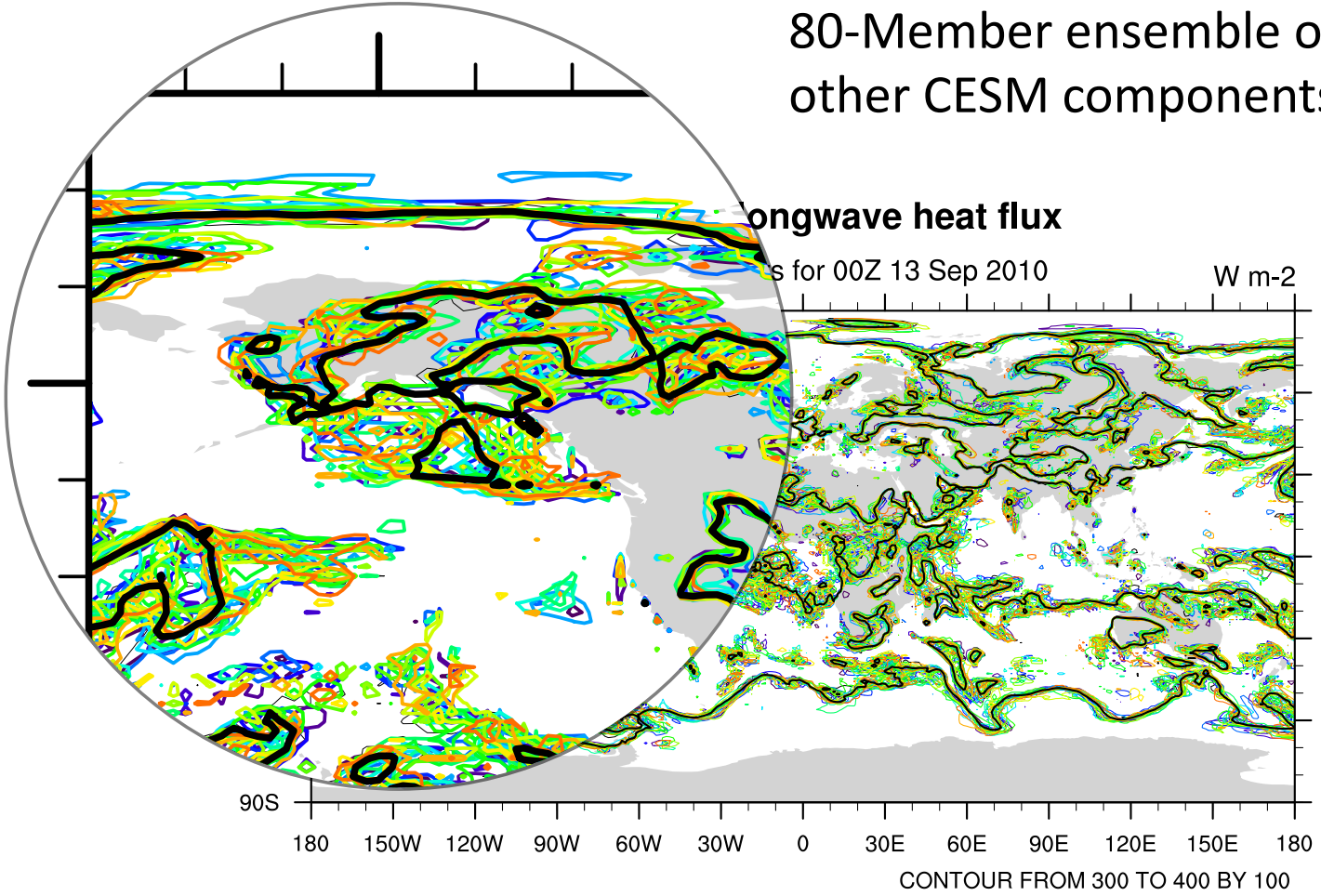
5. Ensemble of plant growth variables from CLM.

Research Data Archive: Contents

- O(120 Tbytes) of data
- Organized by CESM component (cpl, atm, esp, ...)
- Useful units of compressed data for easy download
- CESM gridded data
- "Observation space" data; ensemble *model estimates* of the observations at the obs locations

Ensemble of Atmospheric Forcing

80-Member ensemble of forcing files for other CESM components (20 shown).



Black is ensemble mean.

Reanalysis Quick Facts: Model

- CESM 2.1 release, also used for CMIP 6.
- Atmosphere: CAM6 0.9 degree latitude by 1.2 degree longitude, 32 levels.
- Land: CLM 5.0 BGC-CROP version, same grid as CAM.
- SST and Sea Ice *Coverage*: Specified daily 0.25 degree from AVHRR.
- Sea Ice *Thickness* from CICE model.
- Aerosols, greenhouse gases, volcanic forcing: from CESM when available.

Reanalysis Quick Facts: Observations

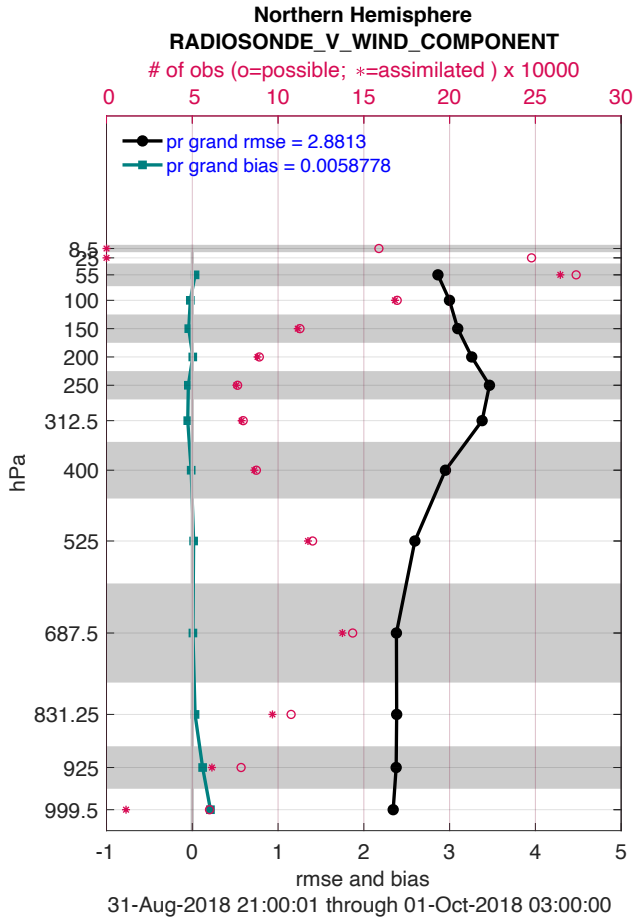
Observations assimilated:

- Temperatures and winds from radiosondes, ACARS and aircraft
- Cloud motion vector winds
- GPS radio occultation refractivity
- AIRS temperature retrievals

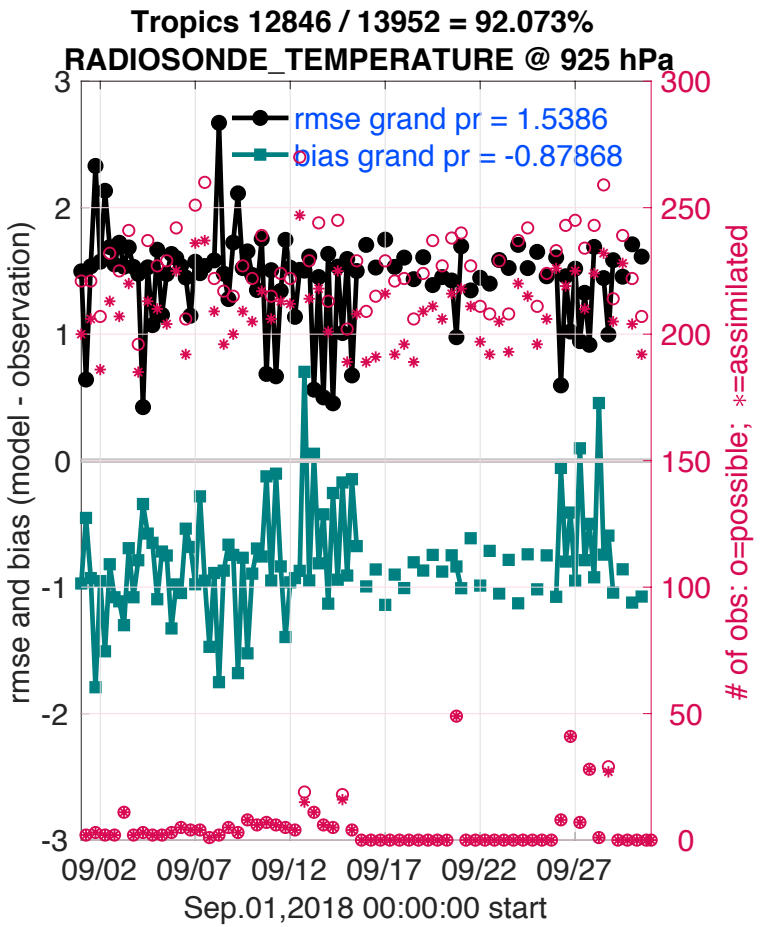
Observations evaluated (”withheld”):

- Radiosonde specific humidity
- AIRS specific humidity retrievals
- Radiosonde, land and marine altimeter

Observation Space Diagnostics

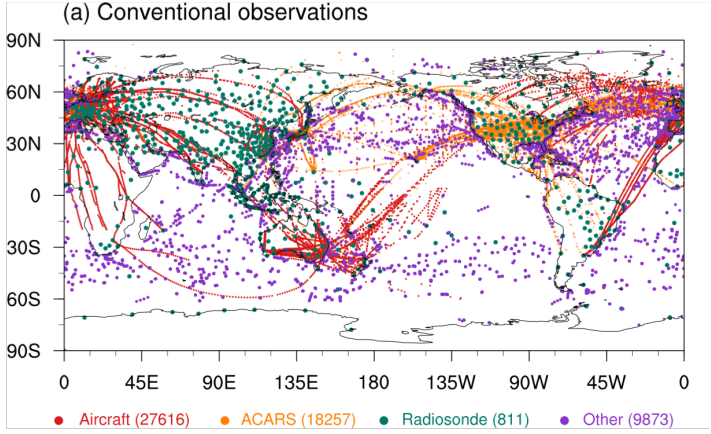


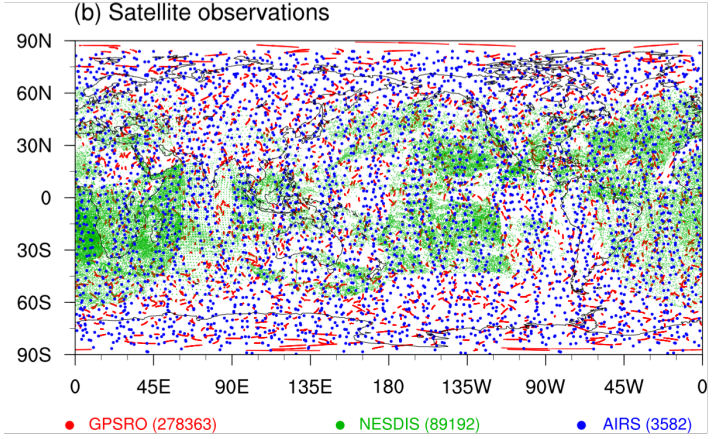
Assimilation status evaluated monthly relative to all obs types; RMSE, bias, totalspread, numbers of obs available (o) and used (*), time series, profiles, 3 regions. All archived.



System Configuration

- **EAM model configuration**
  - v0 configuration with finite-volume dynamical core
  - AMIP simulation (atmosphere/land only)
- **DART: NCAR's ensemble data assimilation system**
  - DART: the Manhattan release (most recent)
  - DART ensemble size: 80 members
  - Assimilation window: 6h
- **Observational data**
  - Sources: NCEP PrepBufr
  - Variables: U, V, T, Q
  - Categories: Stational, aircraft/ship and satellite reports
  - **Not all the data for global reanalysis (e.g., ERA5)**



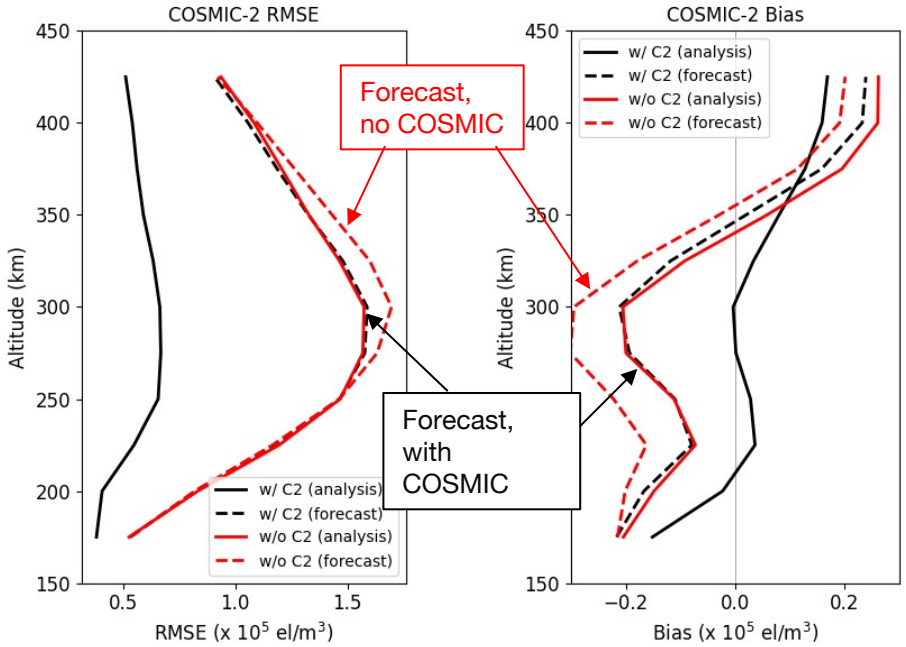


Distribution of observations assimilated by EAM-DART in Jan 01, 2011

DA for Space Weather and Earth's Upper Atmosphere

Lead; Nick Pedatella

- 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

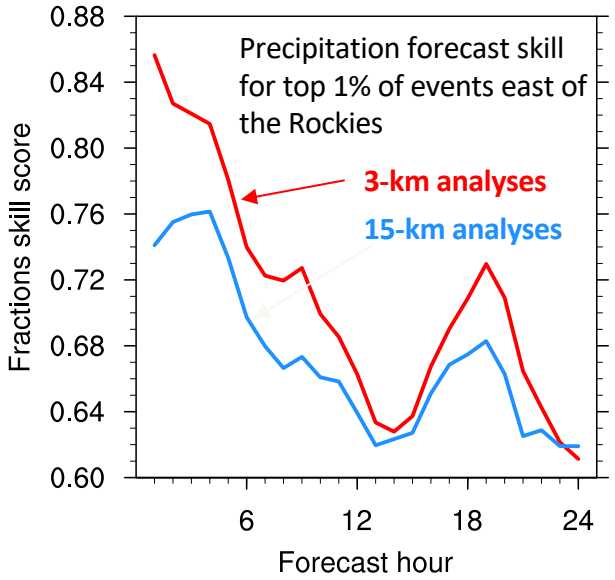
NCAR | National Center for
UCAR | Atmospheric Research

Data Assimilation Research Testbed

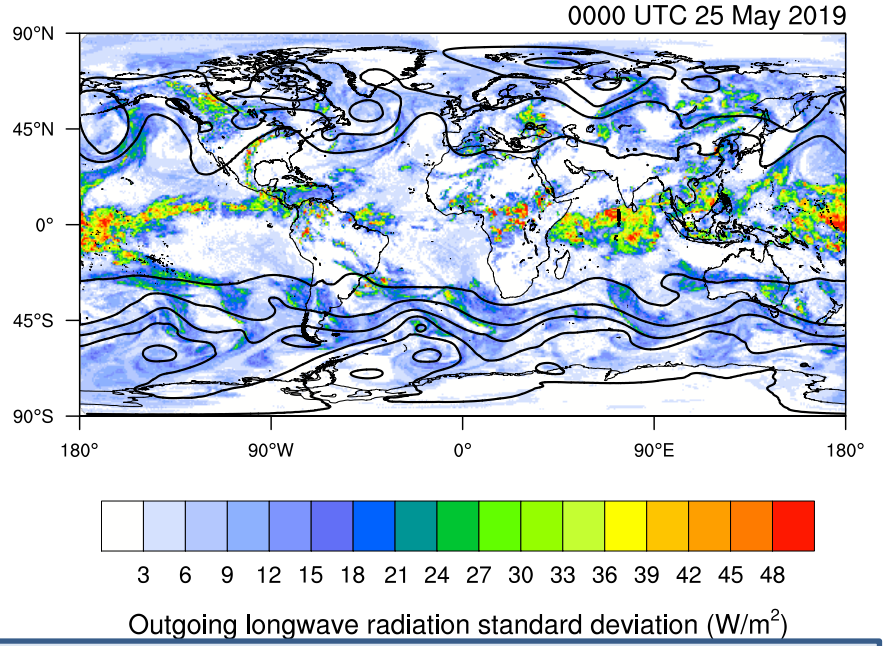
Toward Global Convection-Permitting Data Assimilation

Leads; Craig Schwartz and Glen Romine

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 → Global convection-permitting

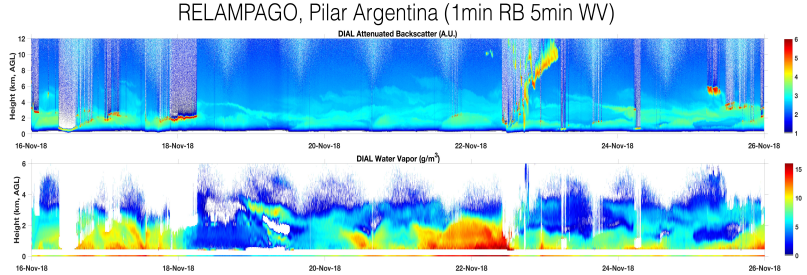
NCAR | National Center for
UCAR | Atmospheric Research

Data Assimilation Research Testbed

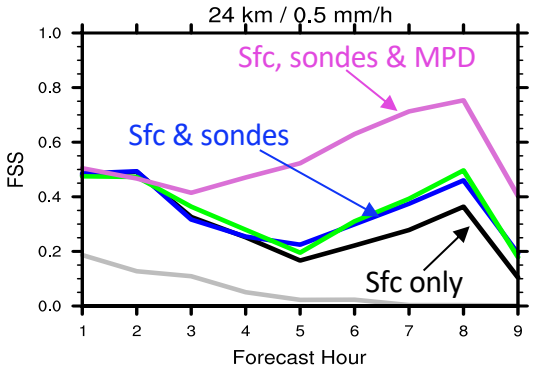
MPD Water Vapor Profile DA for Convective Weather Forecasts

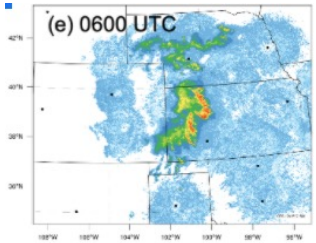
Lead; Tammy Weckwerth

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

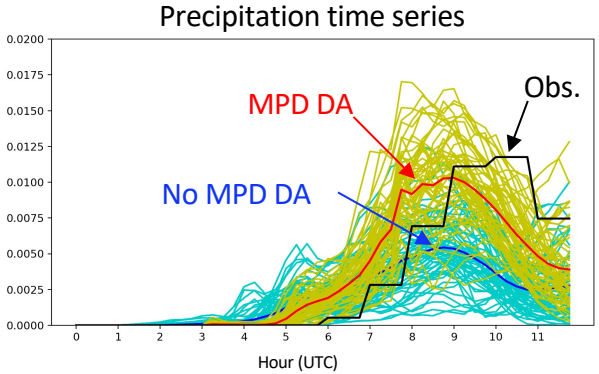


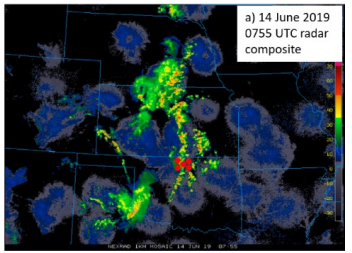
Observing System Simulation Experiment (OSSE)





Observing System Experiment (OSE)





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).

NCAR | National Center for
UCAR | Atmospheric Research

Data Assimilation Research Testbed