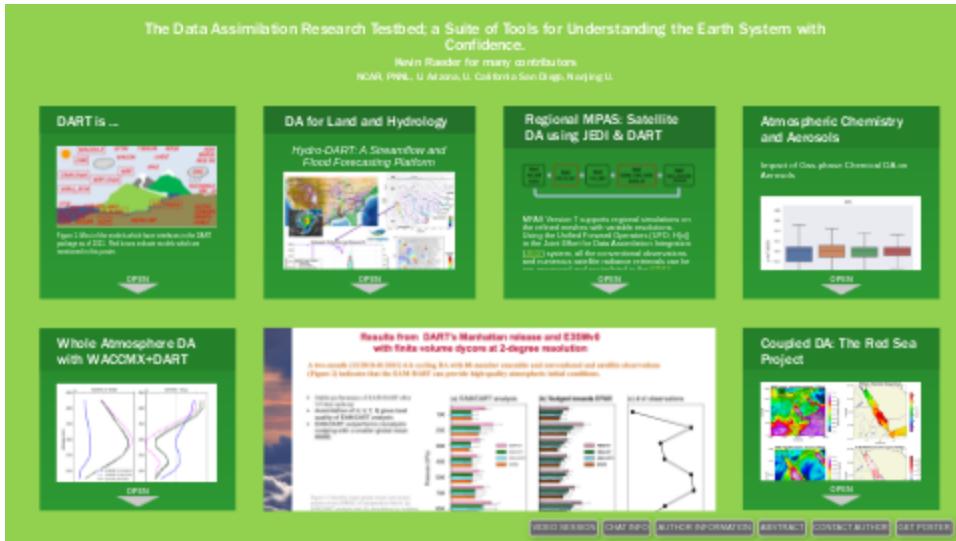


The Data Assimilation Research Testbed; a Suite of Tools for Understanding the Earth System with Confidence.



Kevin Raeder for many contributors

NCAR, PNNL, U. Arizona, U. California San Diego, Nanjing U.

PRESENTED AT:



DART IS ...

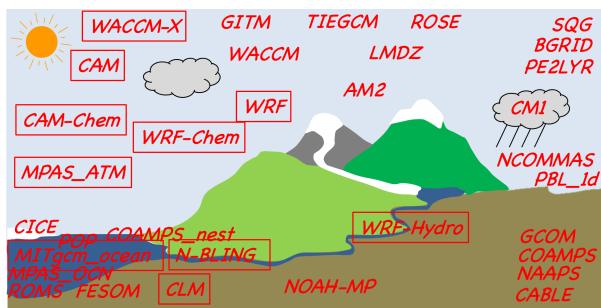


Figure 1: Most of the models which have interfaces in the DART package as of 2021. Red boxes indicate models which are mentioned in this poster.

DART (<https://docs.dart.ucar.edu/en/latest/README.html>) is a suite of open source, community developed, software tools hosted at the National Center for Atmospheric Research. The core of it is ensemble Kalman filter data assimilation, which blends the information from available observations and model hindcasts into a description of a physical system which is better than either input. The ensemble provides not only the "best estimate" (ensemble mean) of the state, but also an objective estimate of the confidence we should have in it (ensemble spread). The ensemble spread varies with physical quantity, space, and time, as a function of the distribution and estimated errors of the observations and the fidelity of the hindcast model. DART even has algorithms that take into account potential errors which have not been quantified, or even identified.

One of DART's design principles is to make interfacing to models and observations flexible and straightforward. DART already supports a wide range of Earth System models (see Fig. 1). Related observations available for use are illustrated in Fig. 2. New ones can be added in a matter of days to weeks, depending on the complexity.

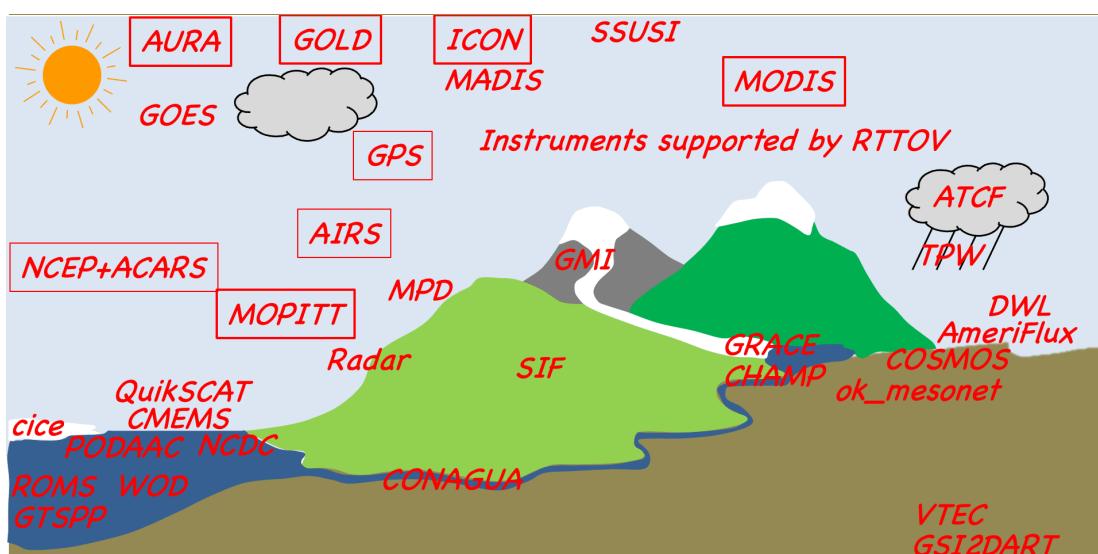


Figure 2: Most of the observations which have interfaces in the DART package as of 2021. Red boxes indicate observations which are mentioned in this poster.

DART provides numerous core assimilation algorithms, from the traditional "square-root" filters, such as the Ensemble Adjustment Kalman Filter (EAKF), to the Rank Histogram filter, particle filters, and the new Quantile Conserving Filter (QCF, see Jeff Anderson's talk (<https://agu.confex.com/agu/fm21/meetingapp.cgi/Paper/832977>)). There are numerous support algorithms to make ensemble data assimilation work well and efficiently in large (Earth system) models; e.g., localization, ensemble inflation, sampling error correction, highly parallel computation, efficient interprocess communication, ...

To analyse the output of assimilations effectively there are postprocessing tools to examine it in state space (on the model grid), in observation space (at the observation locations) and from an ensemble statistical perspective. The ensembles can be used to explore the evolution of patterns of sensitivity of a quantity of interest to all of the model state variables.

DART also provides a unique data set which is very useful in experiments involving components of the Earth system which have the atmosphere as one of their boundaries. The Community Atmosphere Model v6 (CAM6) was used to generate a multiyear, ensemble reanalysis, from which the ensemble of atmospheric forcing of surface components was archived at sub-daily frequencies (CAM6+DART Reanalysis (<https://www.nature.com/articles/s41598-021-92927-0>)).

This ensemble surface forcing helps to maintain the ensemble spread necessary in assimilations which use other model components of the Community Earth System Model v2 (CESM2): CLM, POP, CICE, etc., and provides good approximations of the actual weather to them.

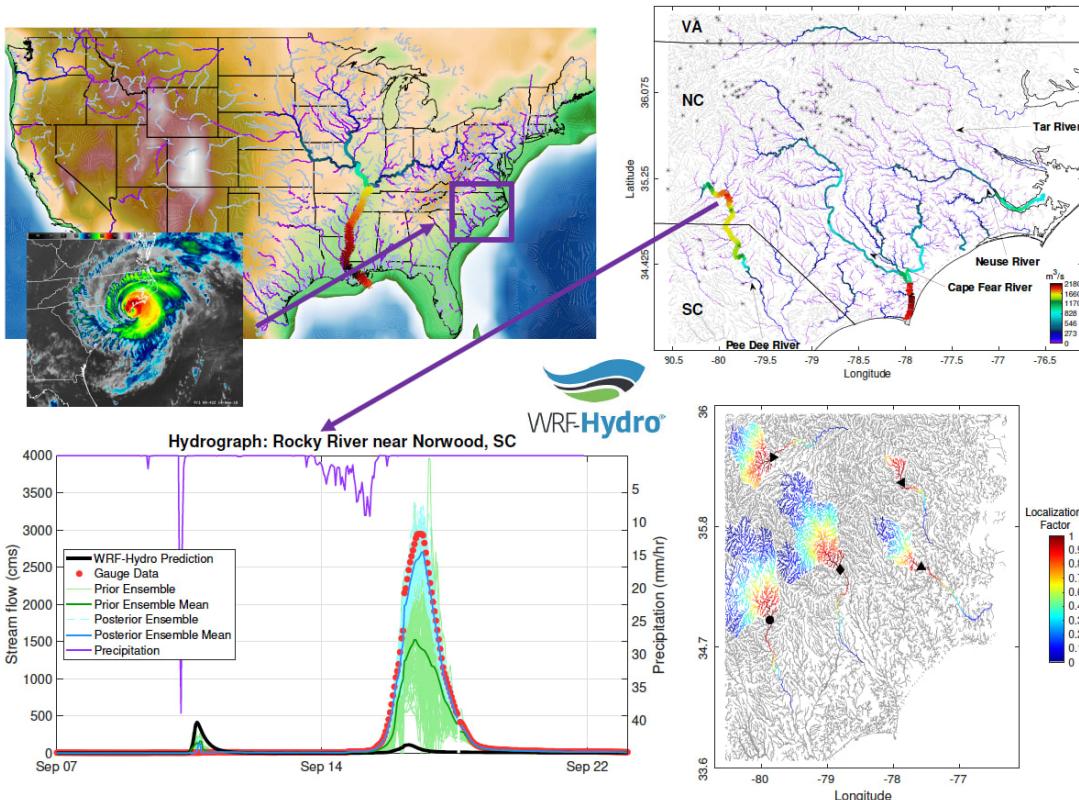
For more information:



<https://dart.ucar.edu>
dart@ucar.edu

DA FOR LAND AND HYDROLOGY

Hydro-DART: A Streamflow and Flood Forecasting Platform



The Weather Research and Forecasting Hydrologic model (WRF-Hydro) is a community modeling framework for hydrologic studies. Combined with DART, the facility is called Hydro-DART. The coupled system has been tested during Hurricane Florence which caused catastrophic flooding and damages in the Carolinas in September 2018 (El Gharamti et al., 2021 (<https://hess.copernicus.org/articles/25/5315/2021/hess-25-5315-2021.html>)).

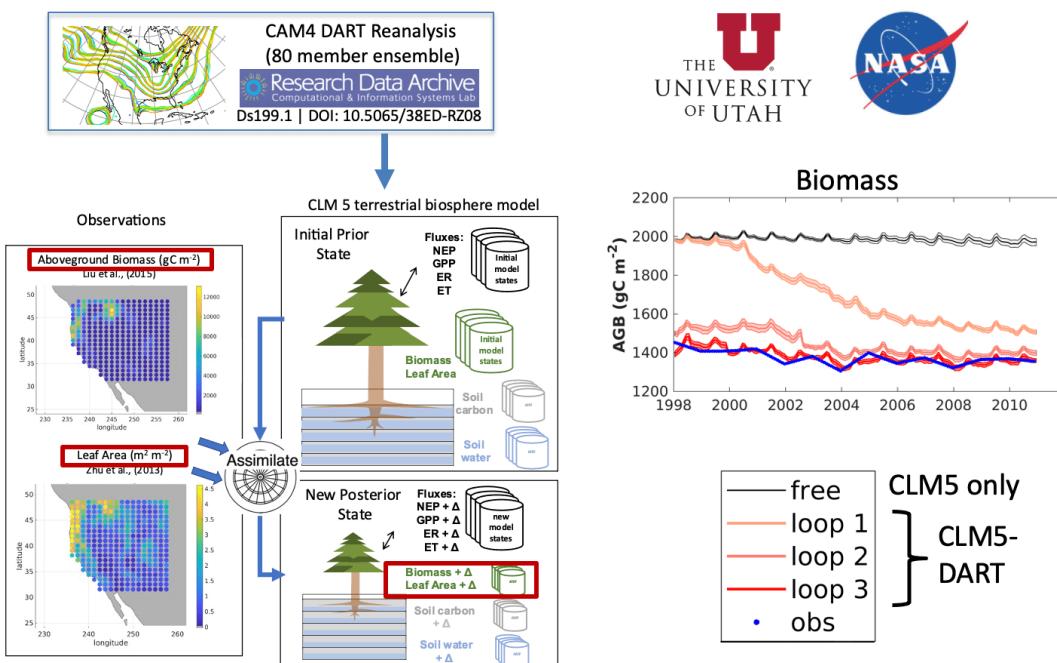
The figure shows in the top left panel the national water model (NWM) CONUS domain with an inset displaying the eye of the hurricane at the time of landfall. A regional subdomain located over the affected area is shown in the top-right panel. Major rivers such as Cape Fear in NC and Pee Dee in SC are annotated. Black markers depict the location of streamflow USGS gauges that are used for hourly assimilation. The study introduces a new topologically based ensemble covariance localization, named along-the-stream (ATS) localization. The method is aimed to mitigate not only spurious correlations, due to limited ensemble size, but also physically incorrect correlation between unconnected streams in the network. As such, distinct watersheds do not share information which results in tree-like localization shapes as shown in the bottom right panel. The use of ATS localization on top of adaptive covariance inflation makes it possible to alleviate model and forcing biases providing accurate streamflow estimates particularly along Pee Dee River in SC as shown in the bottom left panel.

Hydro-DART is currently being extended to support hybridized ensemble and climatological forecast error covariances. Ensemble covariances provide essential time-varying flow of information but suffer from sampling issues. Climatological background errors are static in nature and are statistically more robust than the ensemble ones. The ensemble and climatological background error covariances are

combined linearly using a scalar weighting factor. The weighting factor can be chosen adaptively in time, using the data, by assuming it to be a random variable (El Gharamti, 2021 (<https://journals.ametsoc.org/view/journals/mwre/149/1/mwr-d-20-0101.1.xml>)). This yields an efficient hybrid system that can adaptively operates as an EnKF, a pure EnOI or a mix of both depending on the prediction bias and the uncertainty in the background and the observations.

Contributors: M. El Gharamti, J. L. McCreight

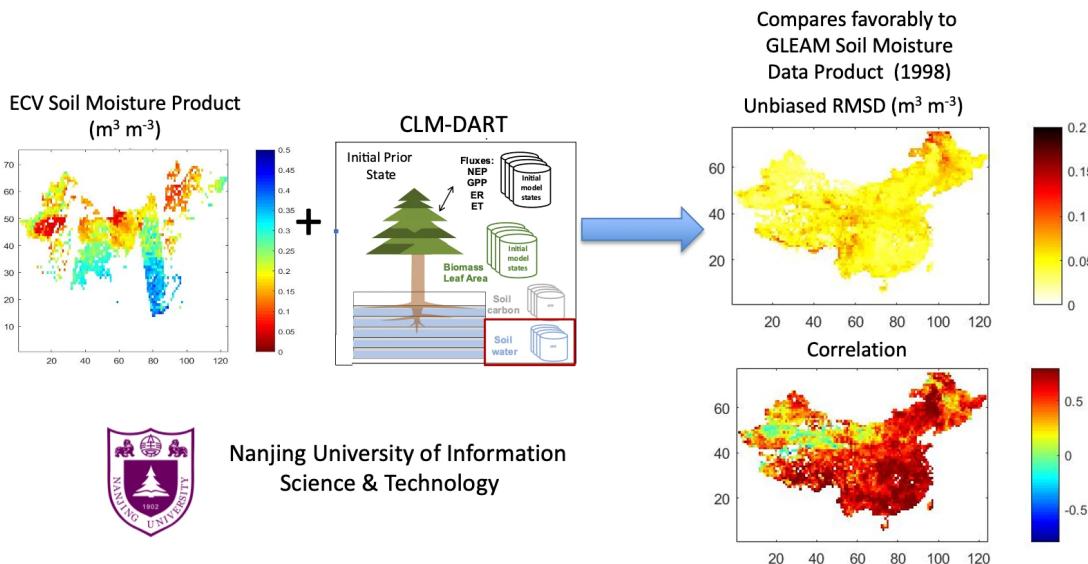
Improving CLM5.0 Biomass and Carbon Exchange across the Western US Using a Data Assimilation System



DART was used to adjust the Community Land Model (CLM5.0) with remotely sensed observations of leaf area and above-ground biomass. The adjusted simulation significantly reduced the above-ground biomass and leaf area, leading to a reduction in both photosynthesis and respiration fluxes. The reduction in the carbon fluxes mostly offset, thus both the adjusted and free simulation projected a weak carbon sink to the land. The results of this work are reported in this study; Raczka et al. 2021 (<https://doi.org/10.1029/2020MS002421>). More information can be found as part of NASA Carbon Monitoring System Project: B25G-1554 (<https://agu.confex.com/agu/fm21/meetingapp.cgi/Paper/921277>)

See also Xueli Hou's poster (<https://agu2021fallmeeting-agu.ipostersessions.com/Default.aspx?s=10-01-8F-2D-26-35-8A-55-10-94-EC-4E-33-57-A6-03>).

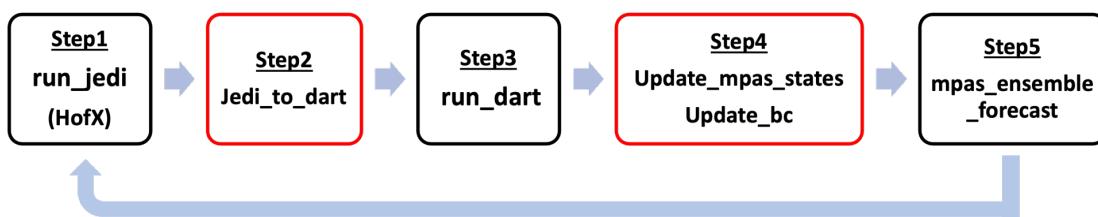
Gap-filling the ECV soil moisture product using CLM-DART



CLM-DART is used to improve the European Space Agency, Climate Change Initiative, Essential Climate Variable (ECV) soil moisture product across China. Soil moisture products are typically spatially incomplete because of limitations in the satellite return interval and data quality. Land surface models such as CLM can be used in conjunction with soil moisture products to both fill these gaps as well as reduce biases in the simulated soil moisture.

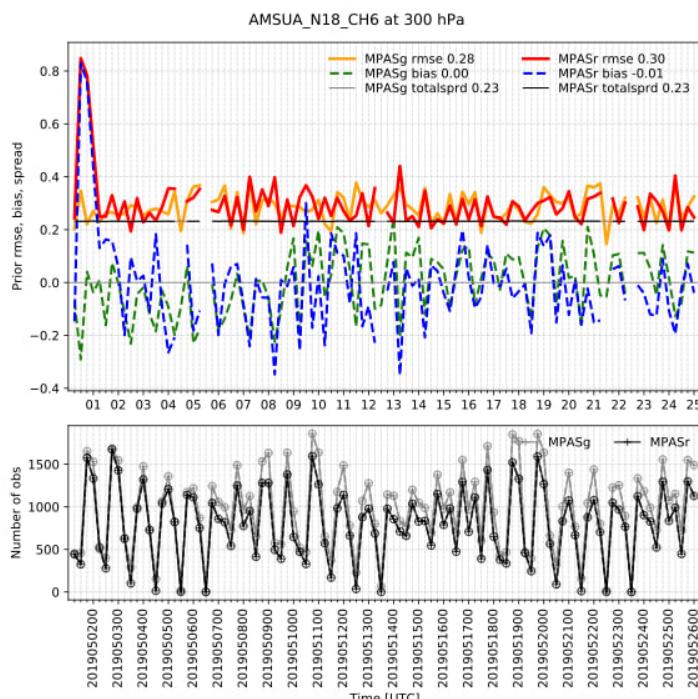
Contributor: Daniel Hagan of Nanjing University

REGIONAL MPAS: SATELLITE DA USING JEDI & DART



MPAS Version 7 supports regional simulations on the refined meshes with variable resolutions. Using the Unified Forward Operators (UFO; $H(x)$) in the Joint Effort for Data Assimilation Integration (JEDI (<https://www.jcsda.org/jcsda-project-jedi>)) system, all the conventional observations and numerous satellite radiance retrievals can be pre-processed and assimilated in the MPAS-DART (https://github.com/MPAS-Dev/mpas-dev.github.com/tree/master/dart_atmosphere) system. We recently extended this capability to regional MPAS meshes so that we can assimilate all meteorological data like JCSDA (Joint Center for Satellite Data Assimilation) (<http://jcsda.org>), including their bias correction and data quality check.

As shown in the schematic diagram of the workflow above, step 2 and 4 (in red box) are either updated or added for regional Voronoi meshes so that the same observations can be assimilated except for lateral boundaries.



Global and regional ensemble cycling on the 15-km quasi-uniform meshes showed that the HofX in JEDI can be used in DART for both global and regional unstructured

meshes, producing consistent results throughout the cycles for May 1-25 2019 (Fig. above). The regional mesh covered the entire CONUS domain on the same 15-km resolution.

Contributors: Soyoung Ha and Craig Schwartz

ATMOSPHERIC CHEMISTRY AND AEROSOLS

Impact of Gas-phase Chemical DA on Aerosols

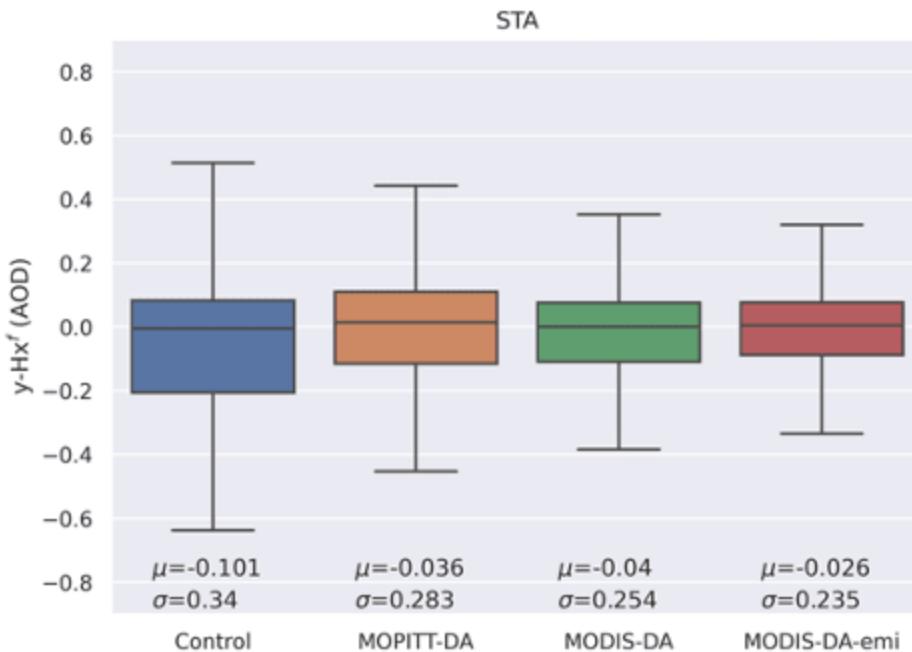


Figure 1: MODIS AOD observation-space Observation Minus Forecast distribution ($m=3311$) for 4 experiments, represented by whisker plots. The mean and standard deviation are also reported at the bottom.

Black Carbon (BC), Primary Organic Matter (POM) and Carbon Monoxide (CO) are co-emitted from incomplete combustion. The goal is to assess the added values of gas-phase chemical DA to improve aerosols, by updating initial conditions and emissions of POM and BC. We use a dynamical and physical ensemble to assimilate weather observations in the online Community Atmosphere Model with Chemistry (CAM-Chem). In addition, an adaptive inflation scheme is used to represent model error and prescribed emissions sources are perturbed. We performed 4 experiments in August 2015:

1. Control: Meteorological DA
2. MOPITT-DA: Same as Control, with assimilation of MOPITT V9J daytime (CO) profiles. Following MOPITT assimilation, initial conditions and emissions of CO, black carbon and primary organic matter are updated.
3. MODIS-DA: Same as Control, with initial conditions of aerosols optimized following the assimilation of Terra & Aqua combined MODIS C6 Aerosol Optical Depth.
4. MODIS-DA-emi: Same as MODIS-DA with the inclusion of an update of BC and POM emissions.

Fig. 1 shows the distribution of the Observation Minus Forecast (OMF) calculated for the 6 hours forecast against the MODIS data over Southern Tropical Africa (STA) for

10 days in August 2015 ($m=3311$ observations). Tropical Africa is the largest source of biomass burning aerosols globally, our control run shows average overestimation of MODIS observations. Interestingly the MOPITT assimilation actually leads to an improvement of AOD when comparing to the unassimilated MODIS. As expected, the MODIS-DA improves the AOD forecast and performs better than the MOPITT-DA forecast. Finally, slightly better performance is achieved when the emissions are updated in addition to the MODIS-DA.

Contributed by Ben Gaubert

WRF-Chem/DART Applied to FRAPPE

We have advanced WRF-Chem/DART to include forward operators for assimilation of OMI O₃, NO₂, and SO₂; TROPOMI CO, O₃, NO₂, and SO₂; and synthetic TEMPO NO₂ and O₃ retrievals. It also includes a revised emissions adjustment scheme that improves the adjustment of trace gases and aerosols in the lower troposphere as shown in Fig. 2. WRF-Chem/DART is being used at various universities and research institutions to assimilate CO₂, OCO₂, HCHO, and NH₃ retrievals. It is being used at NASA to document the potential impacts of TEMPO on air quality forecasting, and at NOAA to evaluate TEMPO's ability to dynamically adjust emissions.

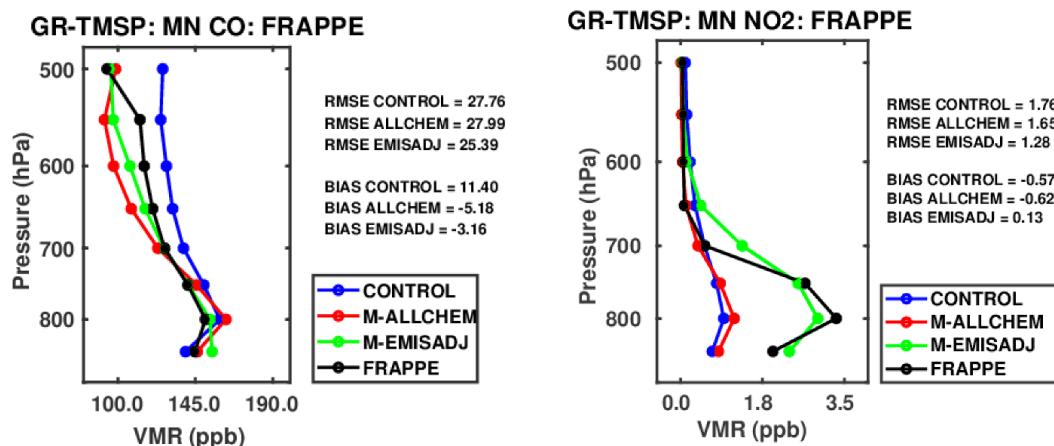


Figure 2: Results from applying WRF-Chem/DART with the revised emissions adjustment scheme to the FRAPPE field program. Skill metrics were calculated against independent observations from FRAPPE on July 17 – 21, 2014.

- CONTROL – assimilate only conventional meteorological observations.
- ALLCHEMA – same as CONTROL but assimilate chemistry observations.

- EMISSADJ – same as ALLCHEM but include emissions adjustment.

In these experiments, we assimilated MOPITT and IASI CO profile retrievals, OMI NO₂ tropospheric column retrievals, MODIS AOD total column retrievals, and AQS CO, O₃, NO₂, SO₂, PM10, and PM2.5 in situ observations.

Contributed by Arthur Mizzi

WHOLE ATMOSPHERE DA WITH WACCMX+DART

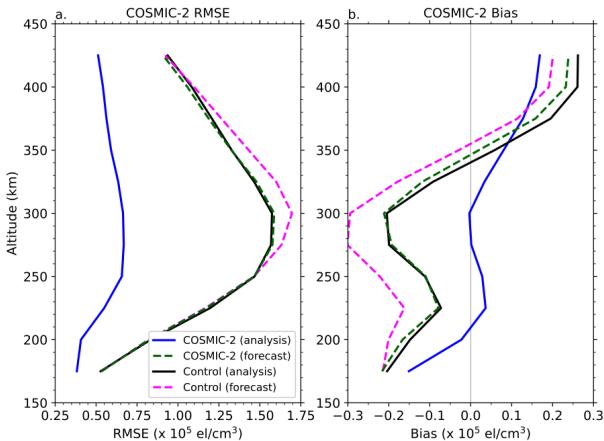


Figure 1: WACCMX+DART analysis and 1 hr forecast (a) root mean square error and (b) bias relative to COSMIC-2 electron density observations. Results are for 25-30 April 2020, and all longitudes between 60S to 60N.

WACCMX is a whole atmosphere model, spanning altitudes from the surface to the upper thermosphere (~500 km). Implementation of data assimilation (DA) in WACCMX with DART enables the study of middle-upper atmosphere variability as well as predictability of the near-Earth space environment.

Recent work has focused on assessing the impact of assimilating new satellite observations in the ionosphere-thermosphere:

- COSMIC-2 (SA33A-03 (<https://agu.confex.com/agu/fm21/meetingapp.cgi/Paper/861299>))
- ICON (SA23A-05 (<https://agu.confex.com/agu/fm21/meetingapp.cgi/Paper/820595>))
- GOLD (SA23A-07 (<https://agu.confex.com/agu/fm21/meetingapp.cgi/Paper/842477>))

An example is shown in Fig. 1, which shows a 6.8% and 24.9% reduction in ionosphere electron density forecast RMSE and bias at 300 km when assimilating COSMIC-2 observations.

Contributor: N. Pedatella



**An Assessment of Ensemble Hindcasts Using
the E3SM Atmosphere Model and the Data**



Assimilation Research Testbed (EAM-DART)

Shixuan Zhang, Kai Zhang, Hui Wan (PNNL)

Jeffrey Anderson, Kevin Raeder (NCAR)

¹Pacific Northwest National Laboratory Richland, WA 99352,
²National Center for Atmospheric Research, Boulder, CO 80305

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.

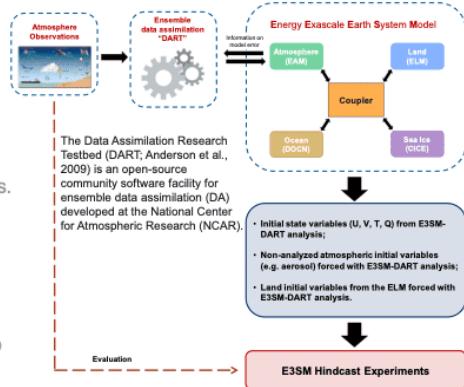
U.S. DEPARTMENT OF
ENERGY
BATTELLE
PNNL is operated by Battelle for the U.S. Department of Energy

Ensemble hindcast workflow with EAM-DART

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)

Our goal: develop and implement proper atmospheric initialization procedures and data assimilation infrastructure to provide high-quality ensemble hindcasts from self-consistent initial conditions to support the model evaluation in E3SM.



A schematic showing the procedure for generating initial conditions and EAM hindcast experiments with DART

2

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.

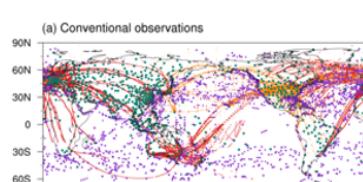
3

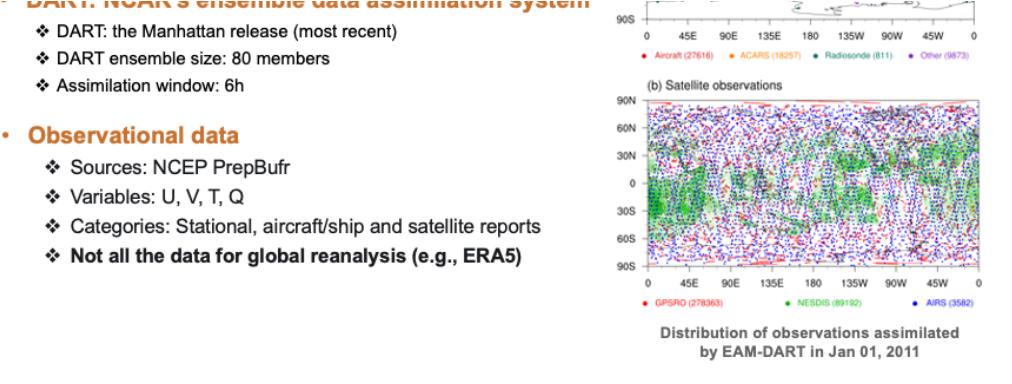
System Configuration

• EAM model configuration

- ❖ v0 configuration with finite-volume dynamical core
- ❖ AMIP simulation (atmosphere/land only)

• DART: NCAR's ensemble data assimilation system





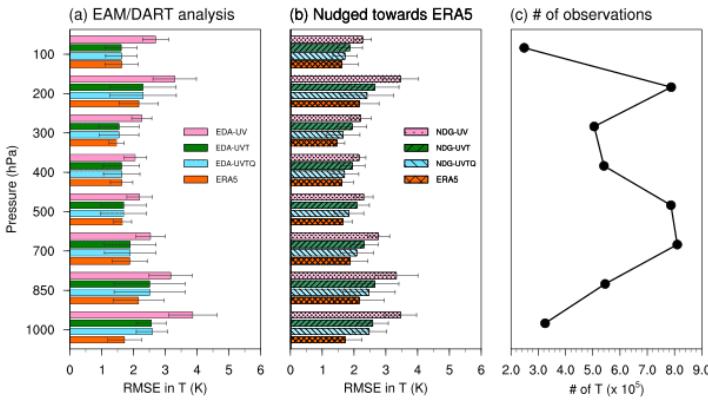
4

Results from DART's Manhattan release and E3SMv0 with finite volume dycore at 2-degree resolution

A two-month (12/2010-01/2011) 6-h cycling DA with 80-member ensemble and conventional and satellite observations (Figure 2) indicates that the EAM-DART can provide high-quality atmospheric initial conditions.

- Stable performance of EAM-DART after 3-5 days spin-up
- Assimilation of U, V, T, Q gives best quality of EAM-DART analysis
- EAM-DART outperforms reanalysis nudging with a smaller global mean RMSE

Figure 3: Monthly mean global mean root-mean-square errors (RMSE) of temperature field in (a) EAM/DART analysis and (b) simulations by nudging EAM toward ERA5 reanalysis during January 2011. The error metrics were derived by comparing the 6-h model output with the observations shown in Figure 1a. The ERA5 reanalysis (ERA5, orange bars) are also added as a reference to quantify the errors in the EAM/DART analysis and EAM nudged simulation.



5

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).

We performed three types of simulations to assess the benefit of DART DA system to the E3SMv0 hindcasts

- CLIM: 5-year (2011-2015) free-running simulations using the similar method for Atmospheric Model Intercomparison Project (AMIP).
- EAM-DART: series of 5-day free-running E3SM ensemble simulations (80-member) for January 2011 with all prognostic variables of the atmosphere and land models initialized using output from E3SM-DART analysis.
- Hybrid IC: series of 5-day free-running E3SM simulations for January 2011 initialized using a hybrid method including u, v, T, Q, PS from ERA5 reanalysis + other prognostic variables of the atmosphere and land models (e.g., clouds, aerosols, soil moisture) taken from nudged E3SM (atm+land) simulations. The experimental design followed Ma et al. (2015).

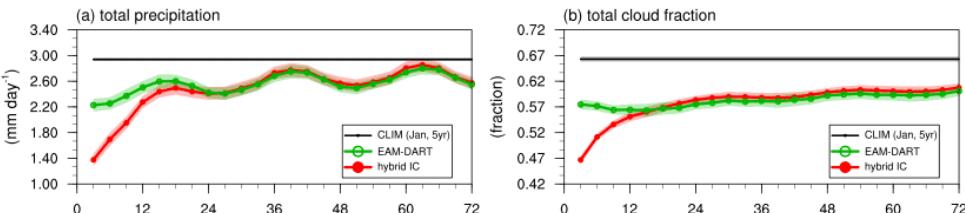


Figure 4: Ensemble means of global mean (a) total precipitation (mm day^{-1}) and (b) total cloud fraction (fraction) as a function of hindcast hours from the simulations initialized with hybrid IC (red dot lines) and EAM-DART (green circle lines) approaches, respectively. The shading denotes the standard deviation of 31 hindcasts initialized at 00Z of each day during January 2011. The horizontal lines in the plots are the 5-year mean climatology in January from the CLIM simulation.

6

Comparison of hindcast skills of simulations initialized with EAM-DART and hybrid IC (ERA5 reanalysis + Nudged simulations)

Evaluation of day 1 (0-24h) hindcast can be ideal for the regions with quick initial error growth (e.g., the boundary layer regions in Figure 6b) as the large initial condition uncertainty can already exist in the day 2-5 hindcasts and limit the robustness of the bias analysis.

- Validation data: 6-h profiling observations in Jan 2011 at ARM SGP site

Legend: EAM-DART (day 5) (green line with circles)

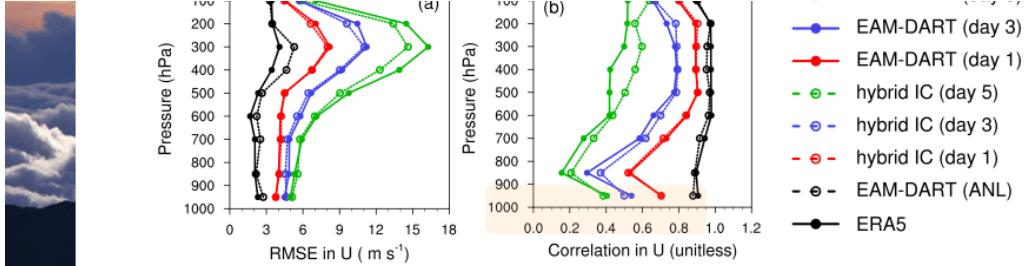


Figure 5: the root-mean-square errors (a) and temporal correlation coefficients (b) of zonal wind field (unit: m s^{-1}) compared against the radiosonde measurements collected by the ARM user facility at the Southern Great Plains (SGP) site. The dashed (solid) lines are for the simulations initialized with hybrid IC (EAM-DART). The 6-h model output during January 2011 was used to derive the error metrics.

7

Correspondence of total precipitation biases between short-term hindcast and long-term climate simulations

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 by the day1 EAM-DART hindcast

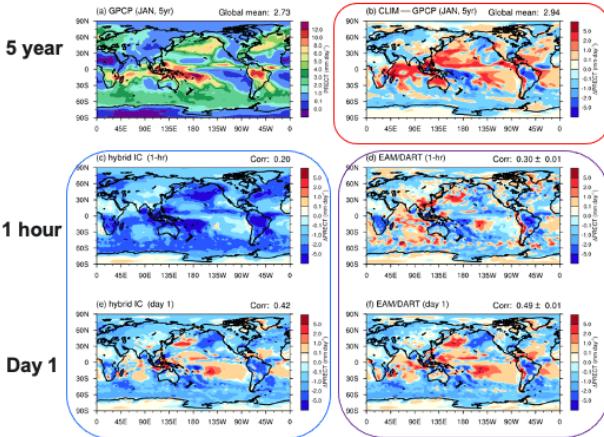


Figure 6: Example illustrating error correspondence between short-term hindcast simulations and long-term climate simulation. The first row shows the 5-year (2011–2015) mean total precipitation from GPCP (<https://psl.noaa.gov/data/gridded/data.gpcp.html>) and the mean biases in EAM model in January. The second and third rows show the model biases in the 1-hour and day-1 hindcasts for January 2011 from the hybrid IC (left column) and EAM-DART (right column).

Simulations:

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

8

Summary and future work

- **With a built-in DA system in E3SM:**
 - ❖ make best use of available observations to generate initial conditions
 - high-quality IC can be achieved by EAM-DART
 - ❖ facilitate analysis of model biases and uncertainty
 - more reliable short-term hindcast simulations
 - objective ensembles generated and trained by DA
 - ❖ assess the incremental benefit of new observations (e.g. ARM observations)
- **What we will do next**
 - ❖ further evaluation and completion of the DART interfaces with EAMv1/v2
 - ❖ collaboration with E3SM developers to apply EAM-DART for model development
 - evaluate model forecast skills on both weather and intraseasonal to interannual time scales
 - adequate manpower and expertise to handle the challenges of both weather and climate applications

9

COUPLED DA: THE RED SEA PROJECT

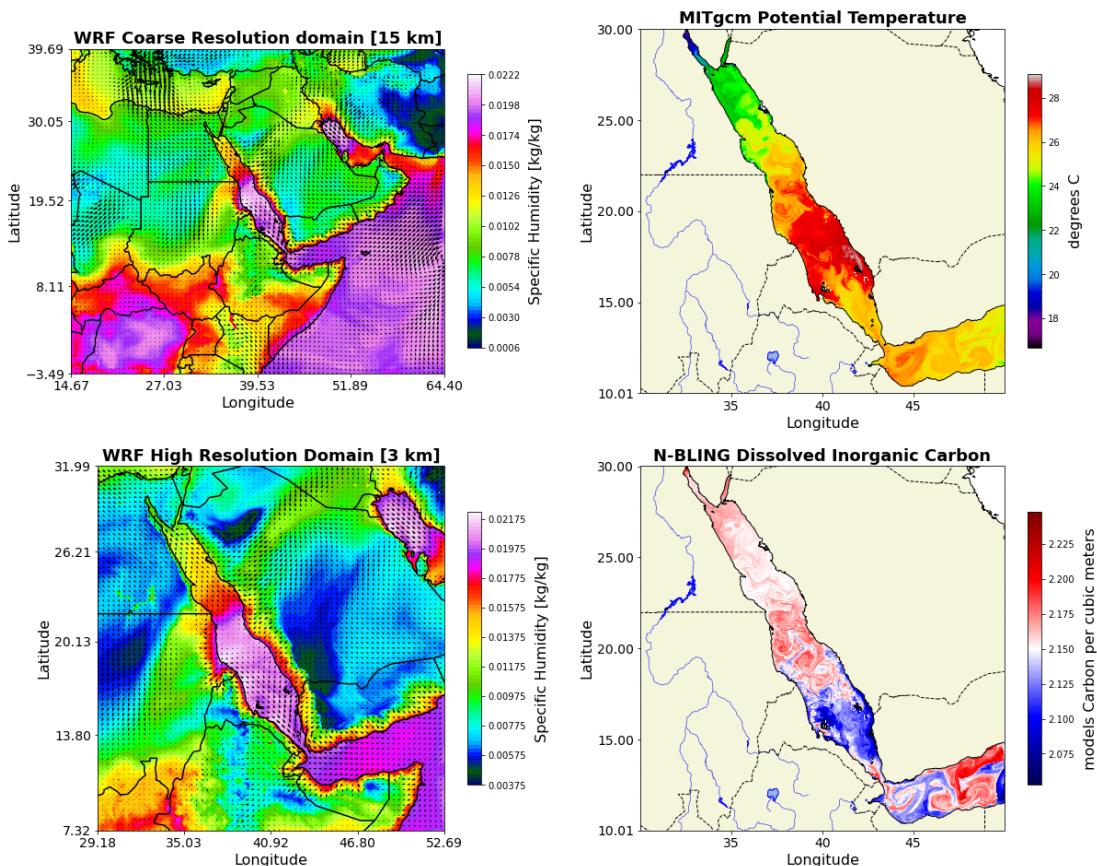


Figure 1: A sample of state-space diagnostic output from DART obtained on October 28th, 2015 [12Z]. The left panels show the prior ensemble means for specific humidity at the surface for the outer, coarse-resolution domain (top) and the inner, high-resolution domain (bottom). The right panels, on the other hand, show the prior ensemble means for sea surface Temperature (top) and dissolved inorganic carbon (bottom) at 1 km ocean resolution.

The Red Sea is a semi-enclosed, elongated marginal sea, lying between Africa and Asia, connected to the Indian Ocean through the Strait of Bab-al-Mandeb in the south and to the Mediterranean Sea through the Suez Canal in the north. It is home to the second-largest coral reef system in the world. To this end, building tools to monitor and predict the seasonal and decadal variability of the Red Sea's ecosystem indicators is of significant societal benefit.

The Red Sea Initiative is a collaborative effort between King Abdullah University of Science and Technology (KAUST), NCAR and Scripps Institute of Oceanography that is building a state-of-the-art regional coupled atmosphere-ocean-biogeochemical forecasting system. The atmospheric system is based on the Weather Research and Forecasting (WRF) model while the Red Sea hydrodynamics are modeled by the MIT general circulation (MITgcm) model. The rich and diverse ecosystem of the Red Sea is studied using the Nitrogen-version of the Biogeochemistry with Light, Iron, Nutrients and Gases (N-BLING) model, which is coupled to MITgcm. DART has been configured and interfaced with WRF, MITgcm and N-BLING for the Red Sea region in order to

1. improve our understanding of ocean and atmosphere interactions,
2. investigate complex ocean circulation and
3. study the impact of ocean biogeochemistry on the ecosystem's sustainability.

Ensemble atmospheric forecasts are constrained every 6 hours by conventional NCEP prepbufr observations in addition to GPS-RO refractivity data. Temperature and salinity data in addition to ocean color are used in DART to constrain the physical biogeochemical ocean drivers. In the near future, a fully coupled earth system data assimilation for the Red Sea region will be investigated.

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

AUTHOR INFORMATION

Kevin Raeder¹; raeder@ucar.edu

Moha Gharamti¹; gharamti@ucar.edu

Brett Raczka¹; bmraczka@ucar.edu

Xueli Huo²; huoxl90@email.arizona.edu

Benjamin Gaubert³; gaubert@ucar.edu

Soyoung Ha⁴; syha@ucar.edu

Craig Schwartz⁴; schwartz@ucar.edu

Nicholas Pedatella⁵; nickp@ucar.edu

Shixuan Zhang⁶; shixuan.zhang@pnnl.gov

Kai Zhang⁶; Kai.Zhang@pnnl.gov

Hui Wan⁶; Hui.Wan@pnnl.gov

Jeffrey Anderson¹; jla@ucar.edu

Ben Johnson¹; johnsonb@ucar.edu

Ibrahim Hoteit⁷; ibrahim.hoteit@kaust.edu.sa

Matt Mazloff⁸; mmazloff@ucsd.edu

James McCreight¹; jamesmcc@ucar.edu

Daniel Hagan⁹; dhagan@ucar.edu

¹ NCAR:CISL:DAReS

² U. Arizona

³ NCAR:ACOM

⁴ NCAR:MMM

⁵ NCAR:HAO

⁶ PNNL

⁷ KAUST

⁸ U. California at San Diego

⁹ Nanjing University

ABSTRACT

Society's ability to make wise decisions depends on an accurate understanding of the current state of Earth and on an ability to predict future states. The Data Assimilation Research Testbed (DART) is an example of a suite of tools designed to improve our understanding through the combination of observations with our theoretical understanding embodied in forecast models. DART's ensemble based data assimilation provides uncertainty quantification as a function of time, location, and variable.

Current research using DART includes:

- Improving streamflow prediction during intense rainfall events, which lead to flooding, using DART and the Weather Research and Forecasting model and the Noah-MP land model (WRF-Hydro).
- Building an integrated atmosphere and ocean forecasting system using DART and WRF for the Red Sea Initiative.
- Understanding air pollution using a global meteorology-aerosol-chemistry prediction system that assimilates aerosol optical depth, carbon monoxide, and weather observations into the Community Atmosphere Model with Chemistry (CAM-Chem).
- Assimilating observations of the Earth system from satellites into the Model for Prediction Across Scales (MPAS; regional and global) using observation operators from the Joint Effort for Data assimilation Integration (JEDI), bias correction for satellite retrievals from the Gridpoint Statistical Interpolation (GSI), and the assimilation environment of DART.
- Connecting the U.S. Department of Energy's E3SM atmospheric model with a broad spectrum of observations to perform short ensemble hindcast simulations for model development and evaluation.
- Generating atmospheric reanalysis data sets from CAM, which enables efficient data assimilation in other components of the Earth system; ocean, land, cryosphere, ...
- Improving DART by giving users more control over how observations are assimilated, and supporting the assimilation of additional observations, such as radiances through the use of the RTTOV software.