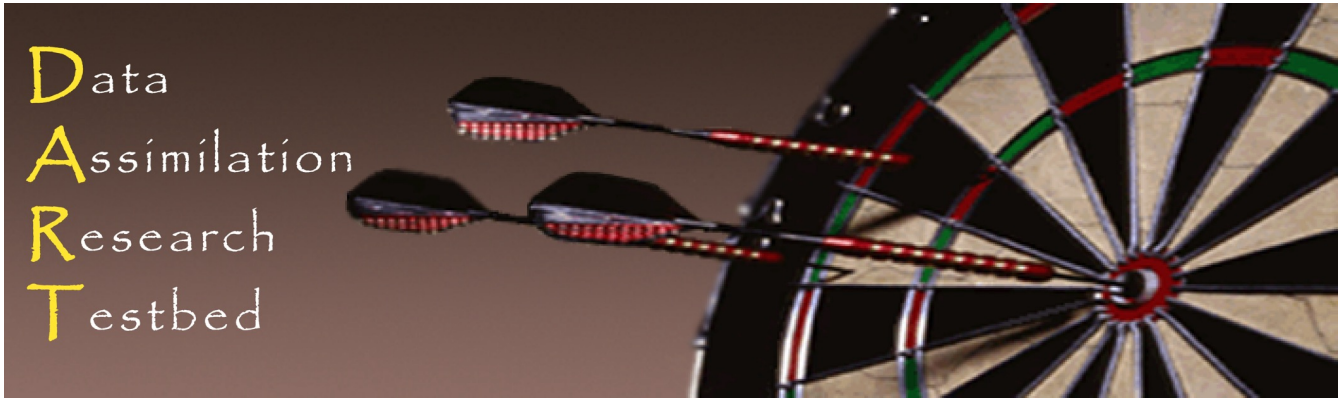
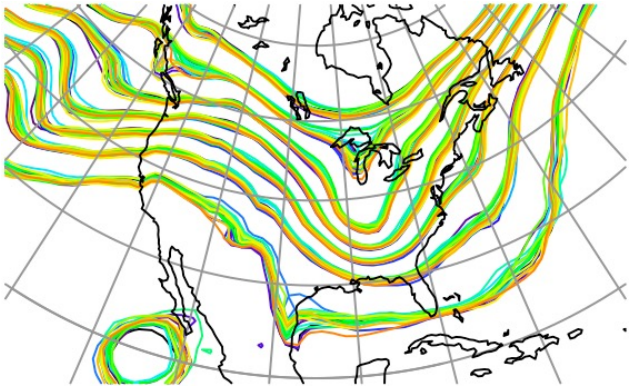


The Data Assimilation Research Testbed in 2022 and Beyond: Not Your Parents' DART

Moha Gharamti and Helen Kershaw, DAReS, CISL, NCAR



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.

Outline

1. DART: A brief background
2. Models and Data
3. Example Applications and Collaborative Projects
4. Algorithms and Diagnostics
5. Infrastructure and Implementation
6. Conclusion

1.1 What is DART?

DART: Data **A**ssimilation **R**esearch **T**estbed

- An open-source, flexible and freely available community facility for ensemble DA
- Developed and maintained by the Data Assimilation Research Section at NCAR
- Up and running; serving the community for nearly 2 decades

DART capabilities include:

- Ensemble forecasting/reanalysis,
- Model improvement,
- Predictability exploration,
- Sensitivity analysis,
- OSE, OSSE,
- Observation design/evaluation,
- Improving DA algorithms

Webpage
https://dart.ucar.edu/

New Improved Documentation
https://docs.dart.ucar.edu/

Easter, 2004 → **Manhattan, 2017**

Anderson, Jeffrey, T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, A. Arellano, 2009: The Data Assimilation Research Testbed: A Community Facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296

1.2 Getting to know DART

State-of-the-art Data Assimilation system for Geosciences:

- Well-tested, theory-based, widely applicable techniques
- Source code distributed on GitHub (https://github.com/NCAR/DART)
- 50+ universities, 100+ other sites and 1500+ registered users
- Extensive teaching and tutorial material with exercises

Simple and Easy-to-use framework:

- Works with many models and observations (real, synthetic, novel)
- No changes required for the forecast model
- Adding a new models can be done in weeks
- Modular: models, observations and assimilation tools are easily combined
- You don't have to be a DA expert to use DART!

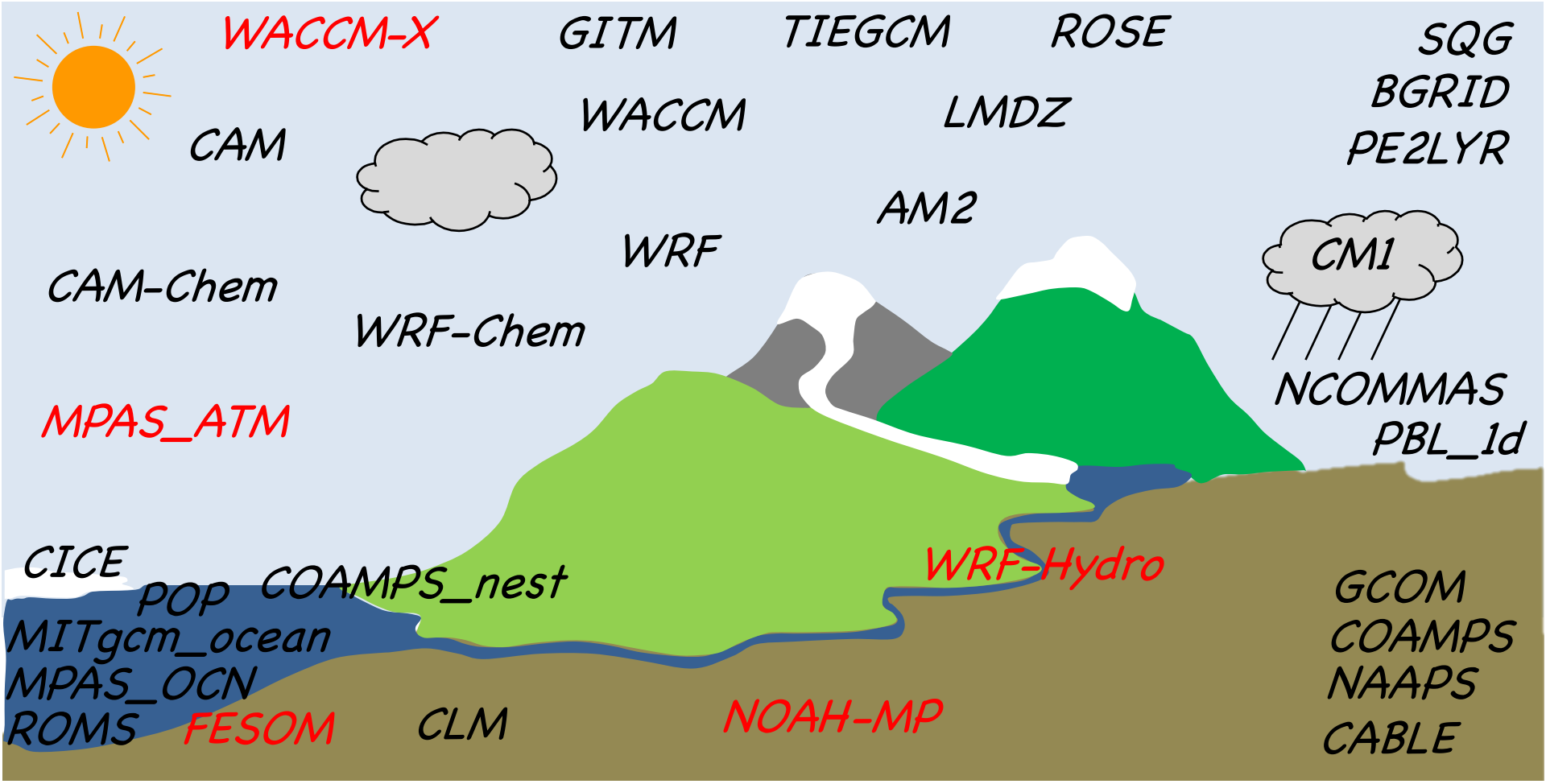
Professional Software Engineering:

- Fast and efficient: laptops to supercomputers
- Excellent performance and scalability (one-sided MPI)
- Quality control with detailed reporting
- Continuously updated with 30+ releases
- Users' contributions are always welcomed, reviewed, streamlined and tested

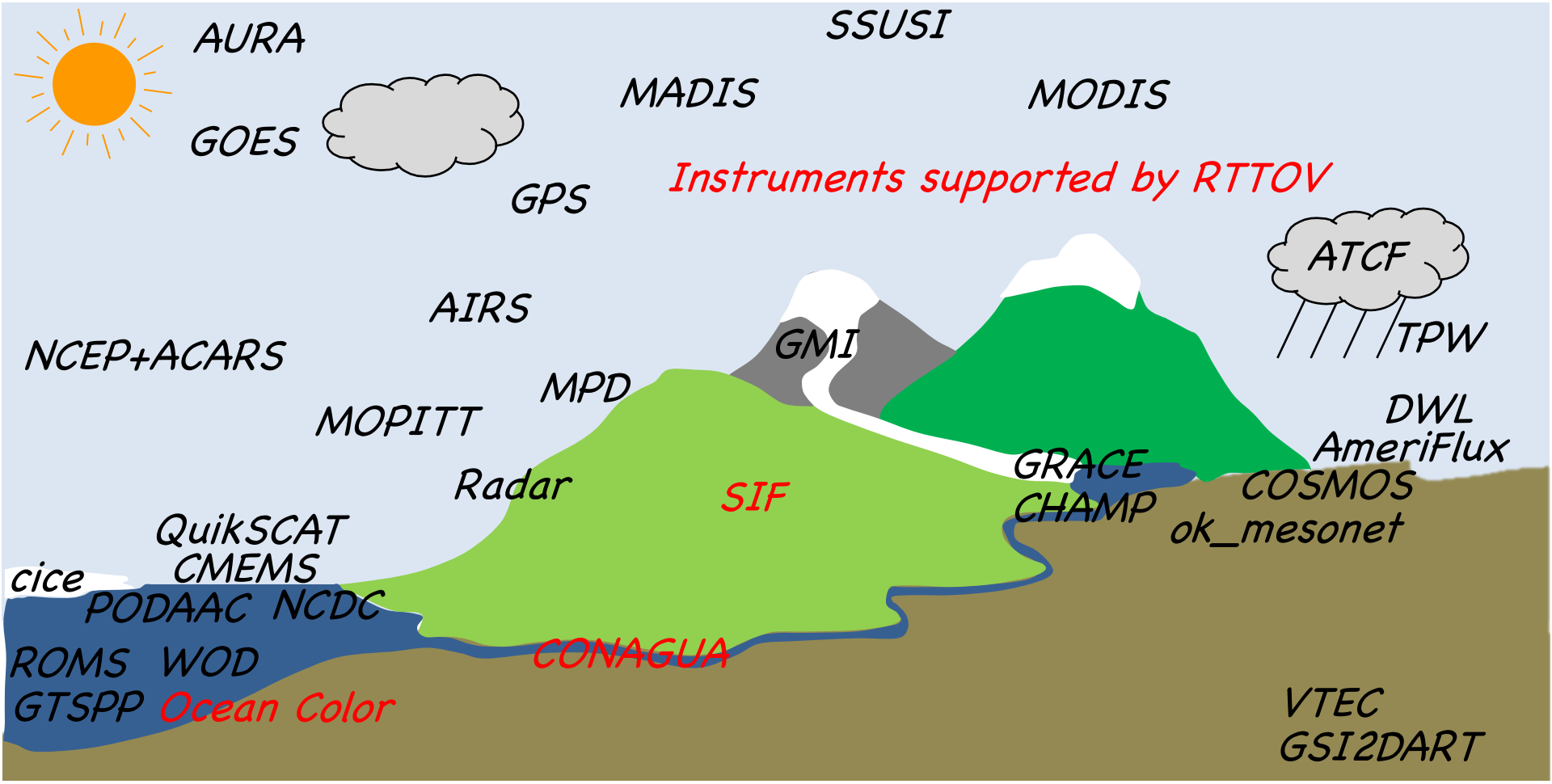
2.1 Some of the Models Interfaced to DART

1. Toy Models [theoretical development]: e.g., L63, L96, L96-2scale L84, L04, ikeda, …

2. Geophysical Models [realistic applications]:



2.2 Some Earth System Observations in DART



NEW

Radiance Observation Support:

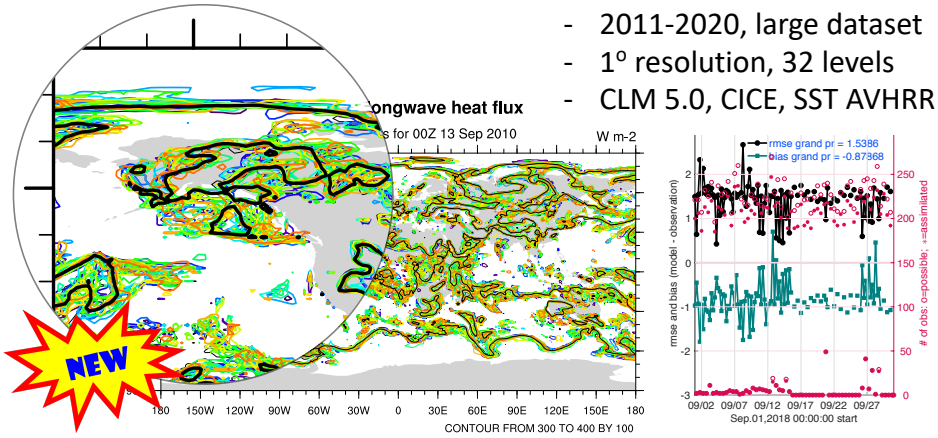
DART now includes the ability to use the RTTOV (radiative transfer for TOVs) forward operators for satellite radiance assimilation

3.1 Applications: Atmosphere

"20-year CAM6+DART Reanalysis"
Kevin Raeder, NCAR

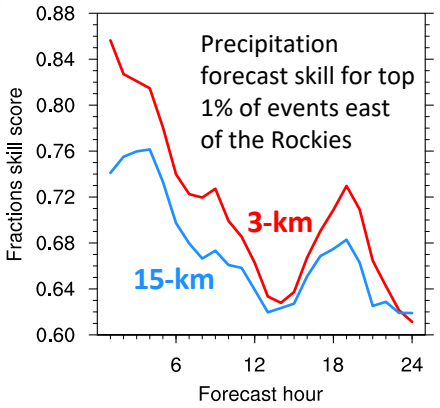
https://rda.ucar.edu/datasets/ds345.0

- CESM 2.1, 80 members
- 2011-2020, large dataset
- 1° resolution, 32 levels
- CLM 5.0, CICE, SST AVHRR

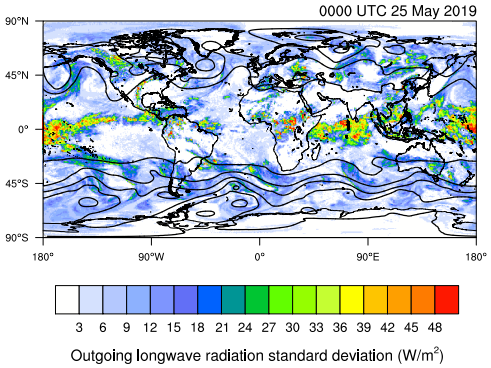


"Global Convection-Permitting DA"
Craig Schwartz & **Glen Romine**, NCAR

Regional 3-km WRF/DART



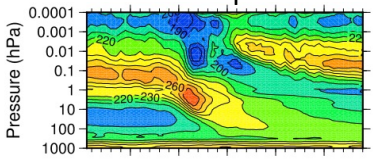
Global 15-km MPAS/DART

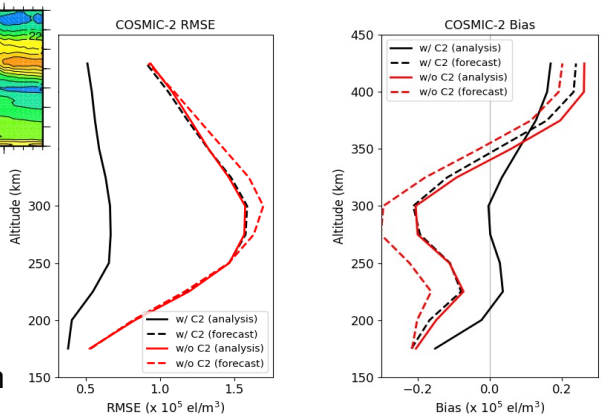


"DA for Space Weather & Upper Atmosphere"
Nick Pedatella, NCAR

WACCM-X + DART

Zonal Temperature



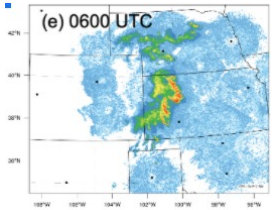


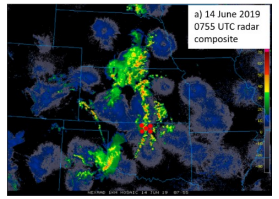
Whole atmosphere DA system

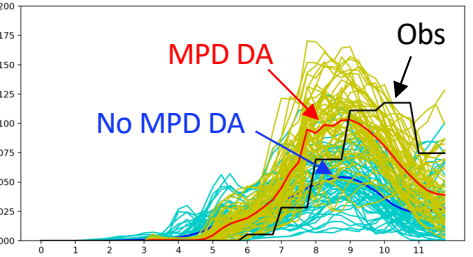
Assimilate obs from the surface to ~500 km

"MPD DA for Convective Weather Forecasts"
Tammy Weckwerth, NCAR

MicroPulse Differential absorption lidar (MPD) measures continuous relative backscatter and water vapor profiles







WRF/DART DA of MPD improves short-term forecasts of convection initiation and evolution

3.2 Applications: Ocean

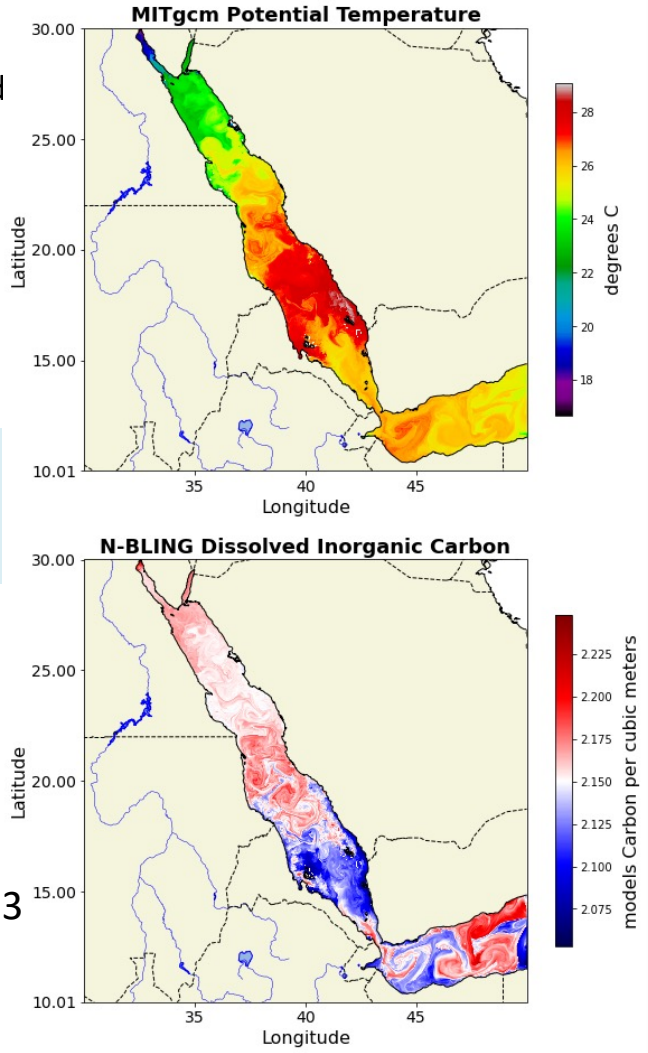
"The Red Sea Initiative"
NCAR, SCRIPPS, KAUST

Regional coupled atmosphere-ocean-biogeochemical forecasting system

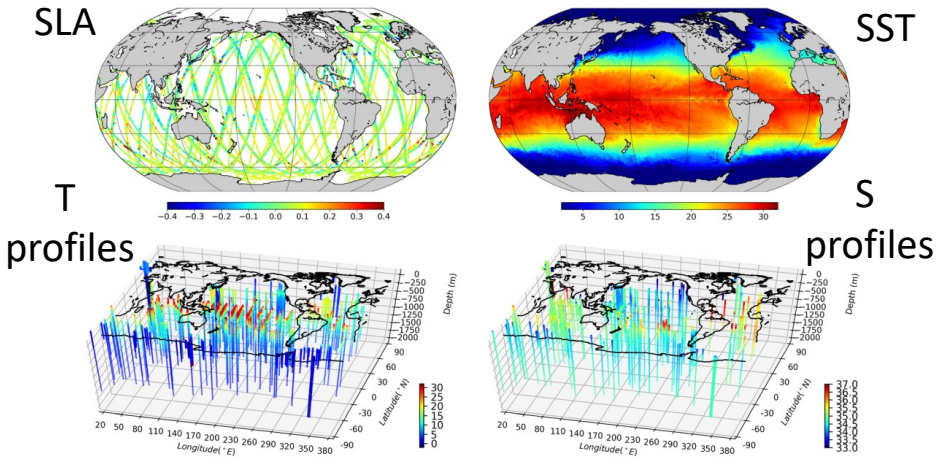
Goal: Fully coupled, high-resolution DA

Components:

- WRF
- MITgcm
- N-BLING
- WaveWatch3
- DART

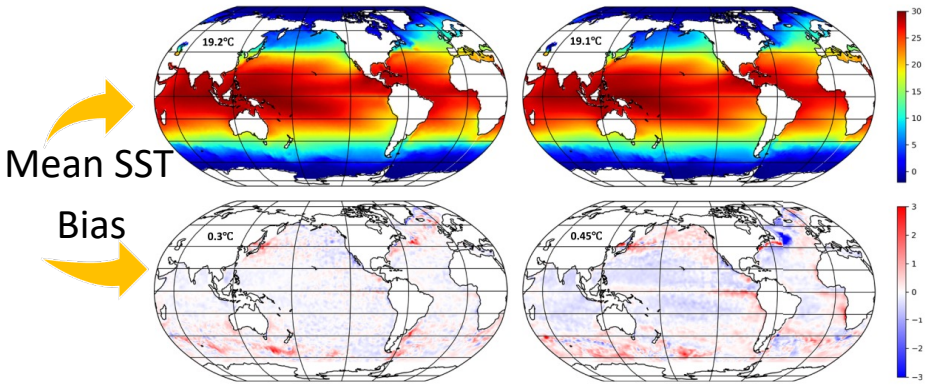


"High-Resolution Global POP+DART"
Fred Castruccio, NCAR



Observations: Daily assimilation [1 Mar. 2005]

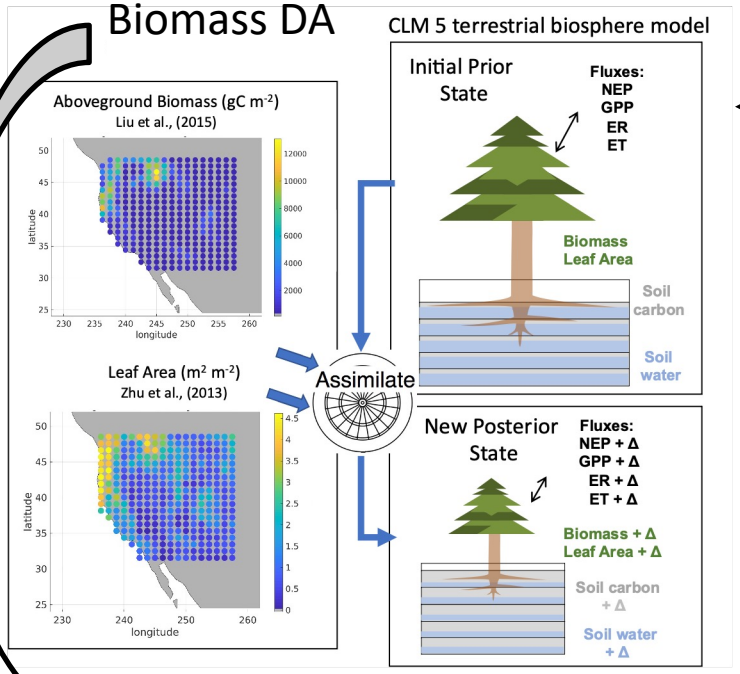
10th of a degree POP (CESM2)



3.3 Applications: Land and Hydrology

NEW

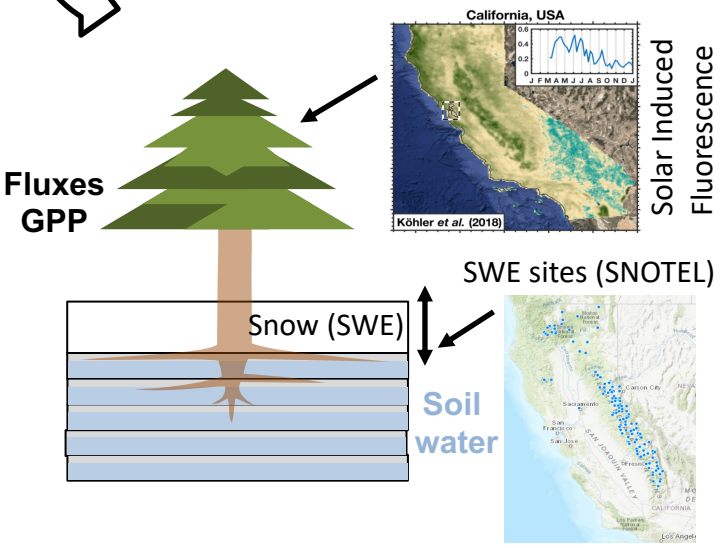
Biomass DA

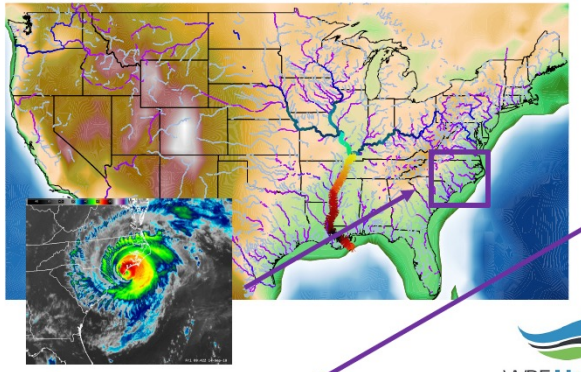


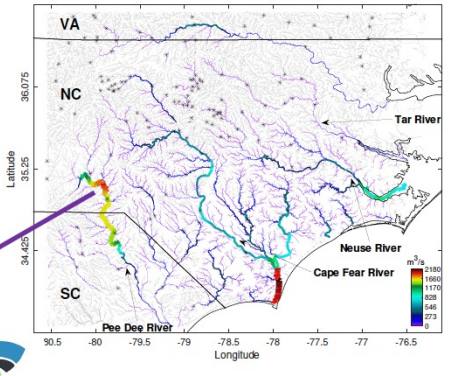
"Land DA: CLM5 + DART"
Brett Raczka, NCAR

"Streamflow and Flood Prediction: WRF-Hydro + DART"
El Gharamti & McCreight, NCAR

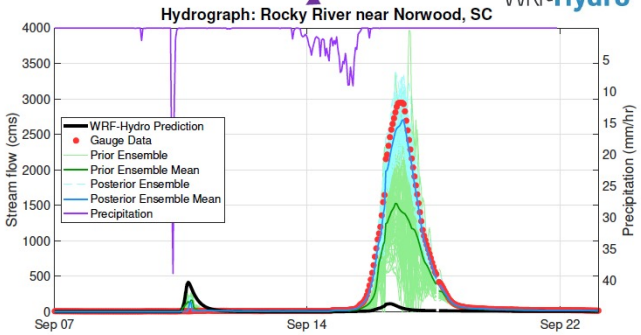
Water+SIF DA

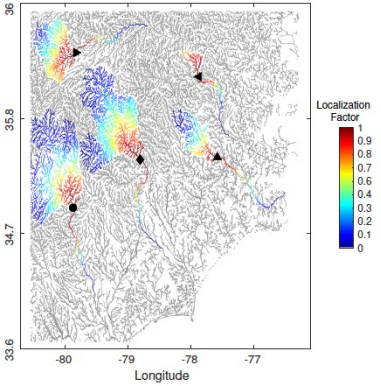






WRF-Hydro





4.1 Algorithms: Filters

An important design/engineering choice in DART:

- **Assimilate the observations serially** (i.e., one after the other)

More than 10 different filtering flavors available in DART:

Filters	Nature	Update	Reference
EnKF	Stochastic	Linear	*Evensen, 2003*
EAKF	Deterministic	Linear	*Anderson, 2003*
GIGG-EnKF	Deterministic	Nonlinear nonGaussian	*Bishop, 2016*
Quadratic EnKF	both	both	*Hodyss, 2012*
EnKF-esops	Stochastic	Linear	*Hoteit et al., 2015*
Localized PF	Deterministic	Nonlinear nonGaussian	*Poterjoy, 2016*
RHF/MARHF	Deterministic	Nonlinear nonGaussian	*Anderson, 2010, 2020*

Additional techniques include:
Gaussian anamorphosis, (fixed-lag) smoothers, group updates, parameter estimation, bias correction, …

Coming Soon:

- *Quantile Conserving Ensemble Filter (QCEF):* A nonlinear filter, especially suited for bounded quantities [*Anderson, 2021*]
- *Adaptive Hybrid EnKF-OI scheme:* Adaptively combine the ensemble statistics with variational / climatological flavors [*El Gharamti, 2021*]

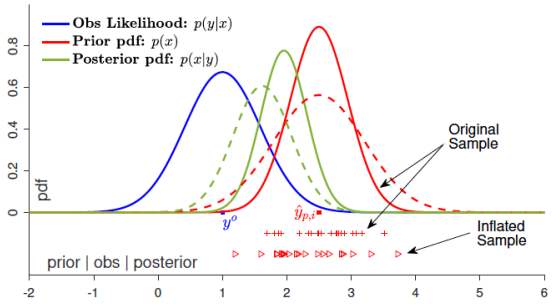
4.2 Algorithms: Other Enhancements

DART also consists of many algorithms that tackle sampling errors in the ensemble, including:

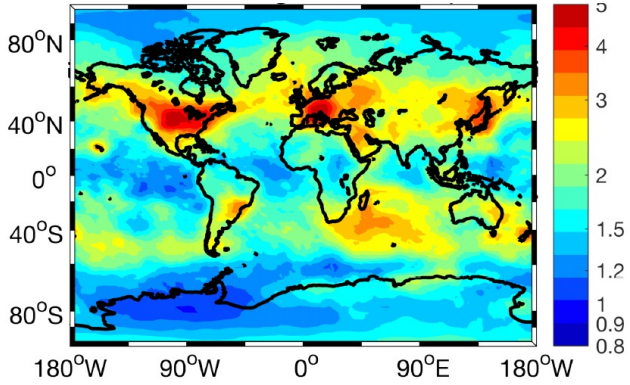
1. **Inflation:**
   a. Adaptive Prior covariance Inflation [*Anderson, 2007, 2009; El Gharamti, 2018*]
   b. Posterior inflation [*El Gharamti et al. 2019*]
   c. RTPS [*Whitaker and Hamill, 2012*]

2. **Localization:**
   a. Adaptive forms using ELFs [*Anderson and Lei, 2013; Lei et al., 2015*]
   b. Vertical and horizontal forms with different correlation functions
   c. Support for irregular grids (e.g., Along-The-Stream Localization)

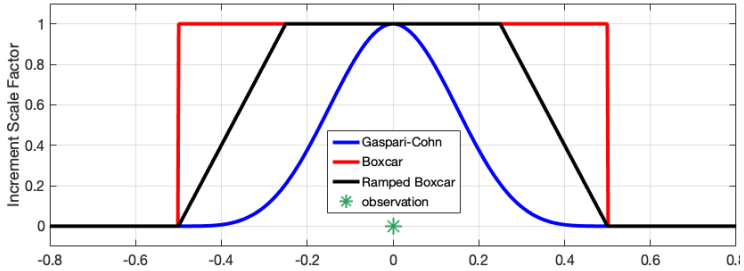
3. **Others:**
   a. Sampling Error Correction [*Anderson, 2012*]
   b. Hierarchical groups [*Anderson, 2007*]
   c. Spread Restoration [*Schwartz et al., 2014*]



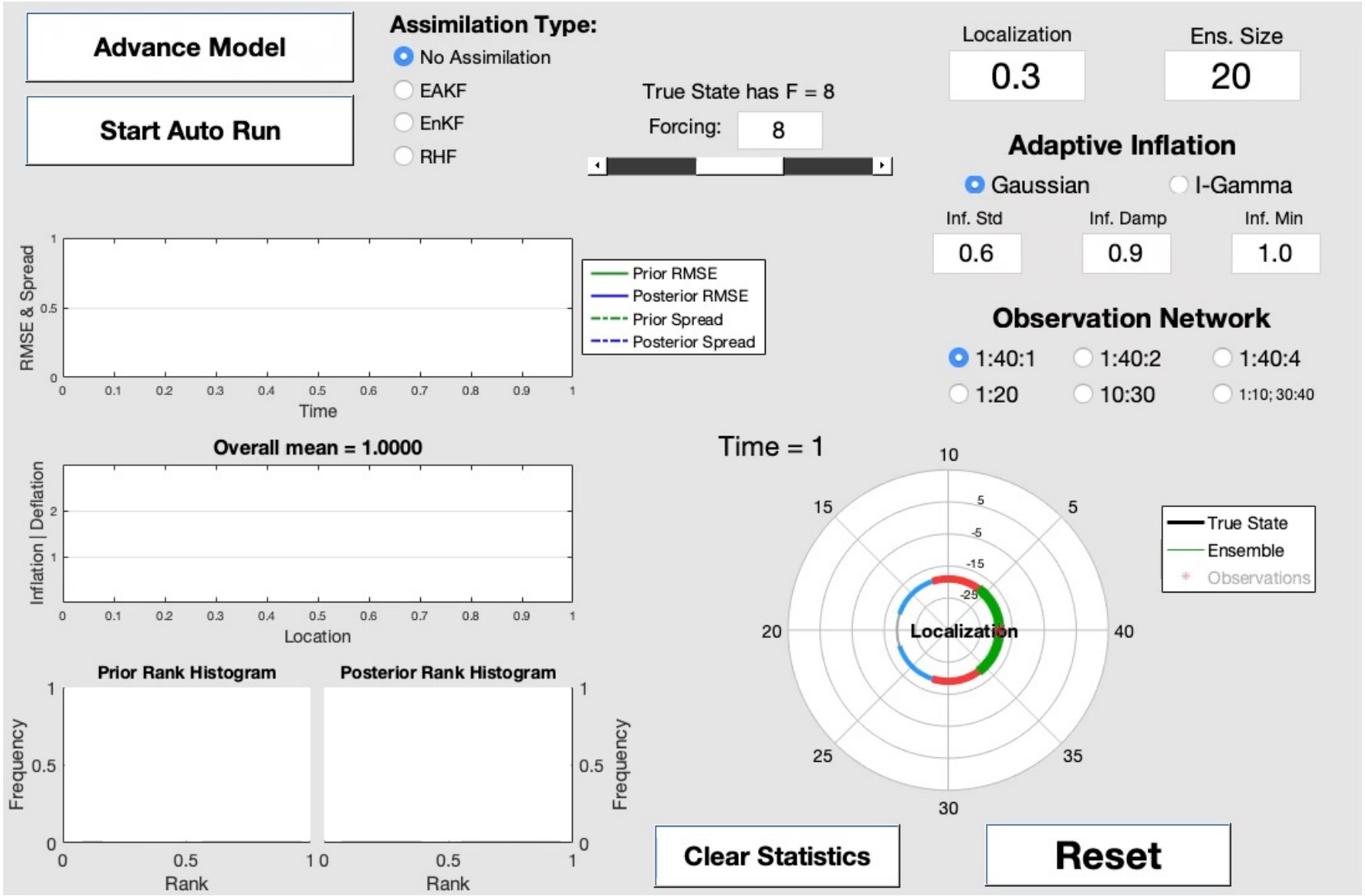
Spatially-varying inflation



Correlation Functions



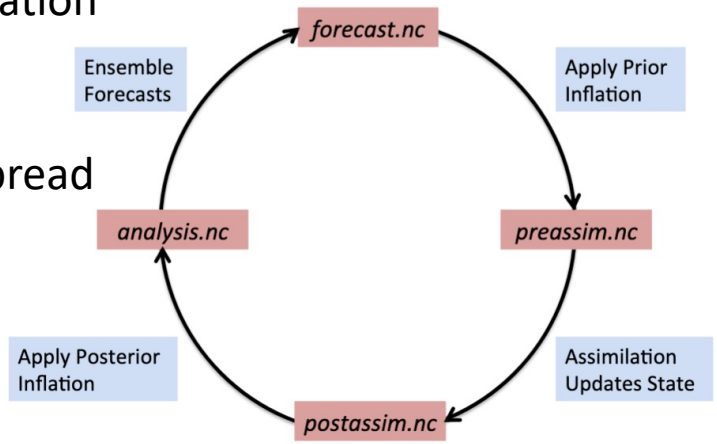
4.3 DART LAB


Advance Model
Start Auto Run
Assimilation Type:
No Assimilation
EAKF
EnKF
RHF
True State has F = 8
Forcing:
8
Localization
0.3
Ens. Size
20
Adaptive Inflation
Gaussian
I-Gamma
Inf. Std
0.6
Inf. Damp
0.9
Inf. Min
1.0
Observation Network
1:40:1
1:40:2
1:40:4
1:20
10:30
1:10; 30:40
RMSE & Spread
Time
Prior RMSE
Posterior RMSE
Prior Spread
Posterior Spread
Overall mean = 1.0000
Inflation | Deflation
Location
Time = 1
True State
Ensemble
Observations
Localization
Prior Rank Histogram
Posterior Rank Histogram
Frequency
Rank
Clear Statistics
Reset

4.4 Diagnostics

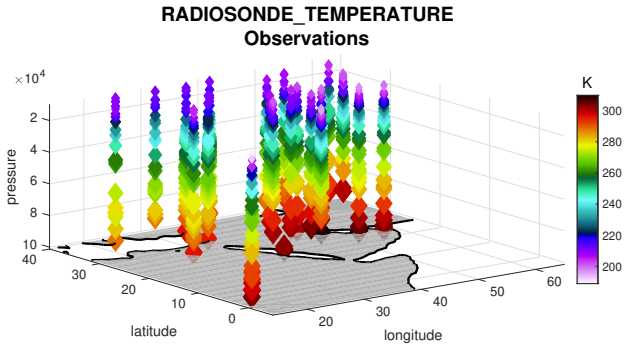
Output **4 stages** of assimilation in state space:

- member files
- Ensemble mean and spread
- Inflation files

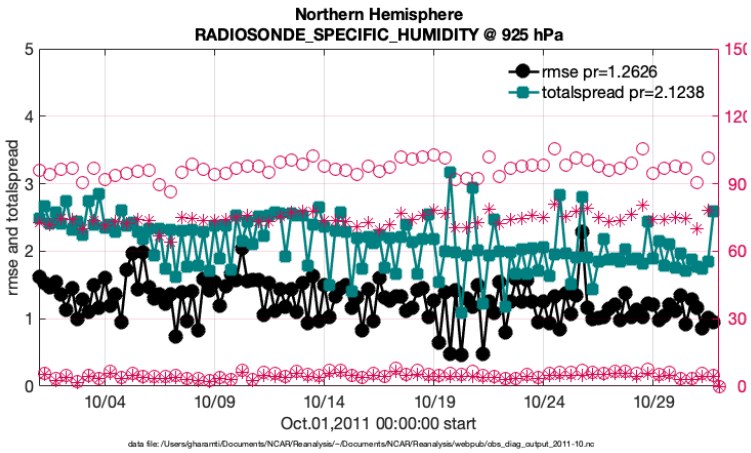


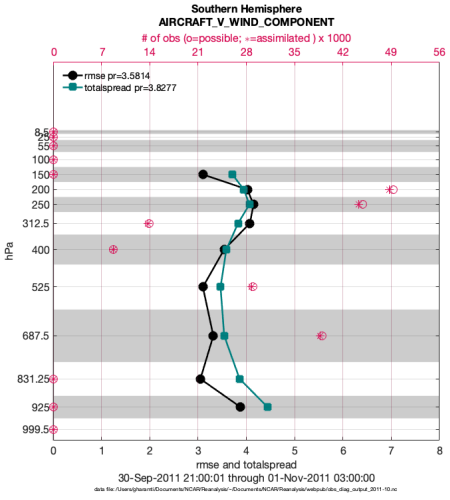
Other useful functionalities:

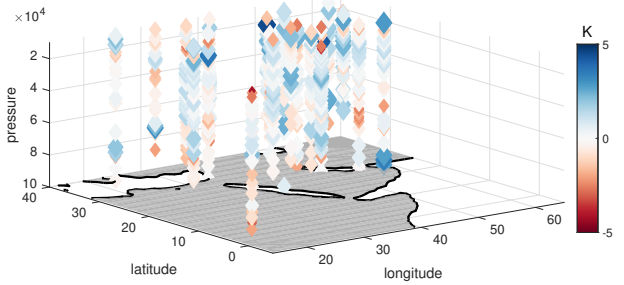
- Link-obs
- Compare experiments
- Common obs

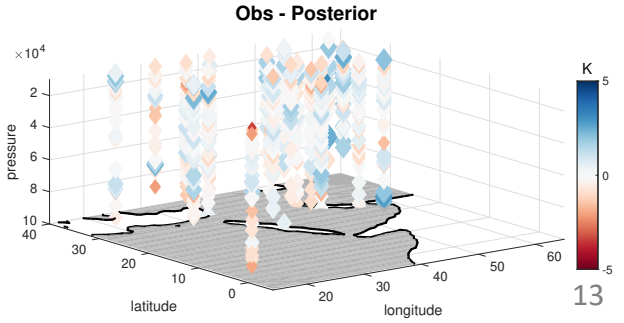


We also provide extensive obs-space diagnostic tools and functions that use MATLAB as part of the distributed code









5.1 Infrastructure and Implementation

Physical
quantity

QTY_U_WIND_COMPONENT

Observation
type

RADIOSONDE_U_WIND_COMPONENT
QKSWND_U_WIND_COMPONENT

Observation sequence

- Tools for manipulating observation sequences

Location

- Geometry
- Distance calculations

Forward Operators

- Observation metadata
- How to calculate a forward operator

Ensemble of State vectors

- IO
- Distributed memory

Model Interface

```
pert_model_copies
get_close_obs
get_close_state
convert_vertical_obs
convert_vertical_state
model_interpolate
read_model_time
write_model_time
end_model
```

```
static_init_model
get_model_size
init_conditions
init_time
adv_1step
get_state_meta_data
shortest_time_between_assimilations
```

Control

- Subroutine: DART calls the model
- Scripting: DART is a separate executable

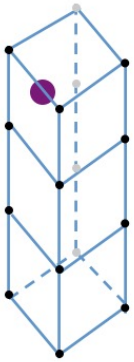
5.2 Distributed State

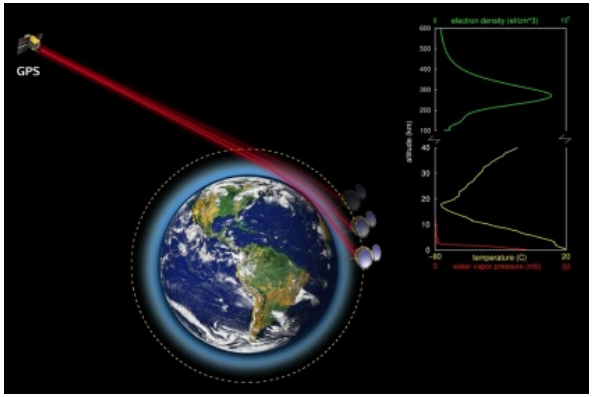
DART algorithm:

- Forward operators are calculated upfront
- Each observation is sequentially assimilated
- Assimilation updates the state and fwd operators for **not-yet assimilated** observations

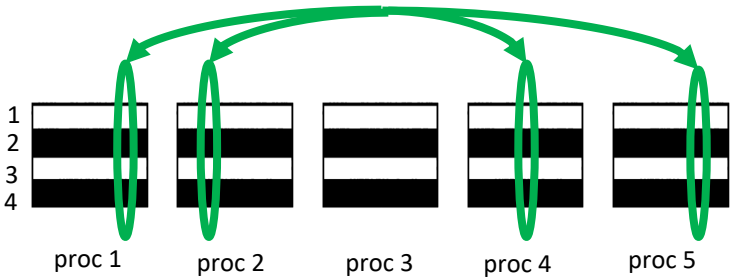
For assimilation, we want the whole ensemble for a subset of the state.
For forward operators, we potentially want the whole state vector.

Forward operator





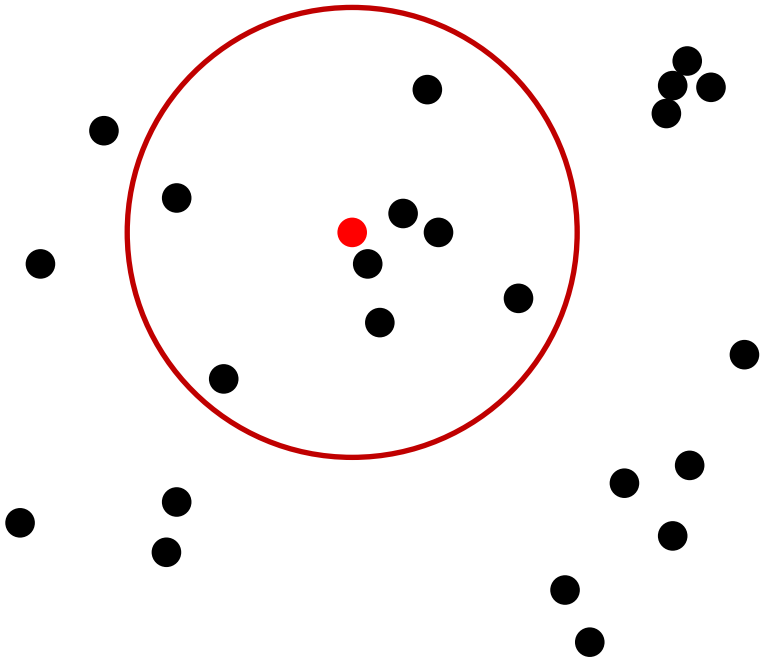
- Leave the ensemble distributed across processors
- One-sided MPI communication to get state elements when needed
- Vectorize the forward operator calculation across the ensemble



5.3 Initial GPU implementations

Localization

Find all state elements and observations within a radius from a given observation.



- NERSC Hackathon working with mentors from NVIDIA
- OpenACC implementation
- Mini-app: 10x speed up
- Upfront 2D calculation of distance

6. Conclusion

- DART is a flexible, research focused, community, ensemble DA system
- It is used for a broad variety of Earth system research projects
- From theory-based research (laptops) to large scale world applications (HPC)
- Excellent resource for learning DA

You have an interesting project?
We would like to work with you!
Please get in touch with the DART team and we'll be more than happy to collaborate.

Thank you!!

Moha: gharamti@ucar.edu
Helen: hkershaw@ucar.edu
DART team: dart@ucar.edu

CISL
Computational & Information Systems Lab

NCAR
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH