CESM-DART

Updates on Assimilation for Earth System Models

Helen Kershaw
DAReS NSF NCAR

SIParCS
DART NUOPC

CROCODILE
pyDART

Terminology

SIParCS Summer Internships in Parallel Computing

DART Data Assimilation Research Testbed

NUOPC National Unified Operational Prediction Capability

CROCODILE CESM Regional Ocean and Carbon cOnfigurator with Data assimilation and Embedding

CESM Community Earth System Model

Ensemble Data Assimilation

Ensemble Data Assimilation

Group of model forecasts

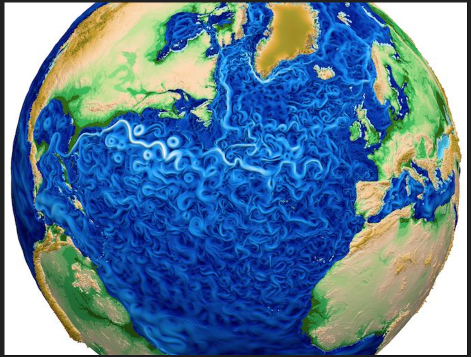
Ensemble Data Assimilation

Group of model forecasts

Measurements

Ensemble Data Assimilation

Group of model forecasts

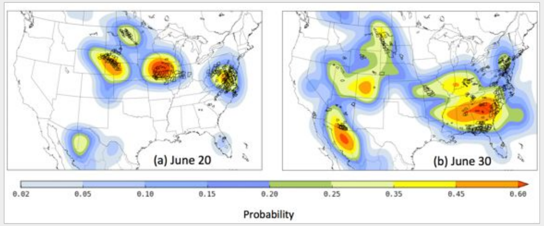


Improved estimate

Measurements

NCAR Real-time ensemble prediction system

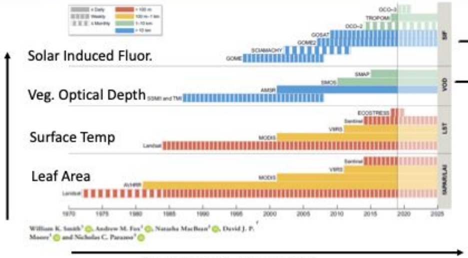
Severe weather forecast for *two* days compared to NWS warnings

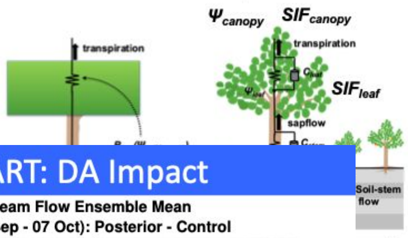


- WRF, 10 member ensemble, GFS for boundary conditions
- Continuous operation from April 2015 to December 2017
- 48 hour forecasts at 3km resolution
- First continuously cycling ensemble system for CONUS
- CISL Dedicated Queues and Computing Support were Vital

CISL Seminar: 6 Nov 2019 page 28

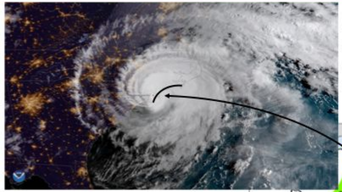
Advancing models & observations together





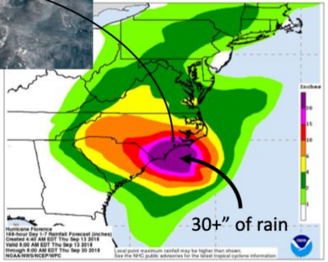
5

WRF-Hydro/DART: Florence 2018



Hurricane Florence made landfall near Wrightsville Beach, North Carolina at **7:15 a.m. ET September 14**. The GOES East satellite captured this geocolor image at 7:45 a.m. ET

Winds up to 150 mph (240 km/hr)
Damage: $24.23 billion
NOAA/NWS/NCEP/WPC

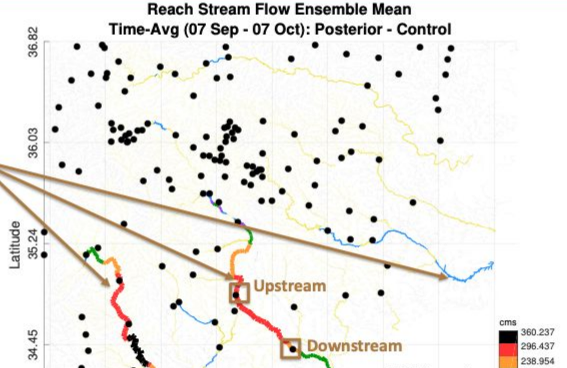


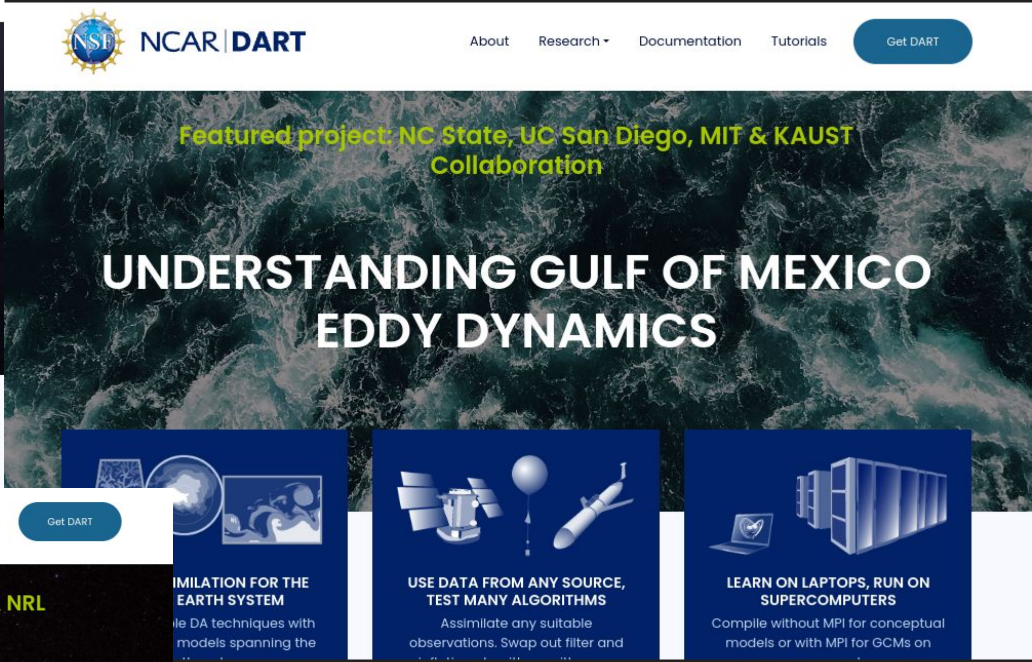
CISL Seminar: 6 Nov 2019 page 38

WRF-Hydro/DART: DA Impact

Assimilation happens every hour

Correction along major reaches. DA is adding water to the stream channels.



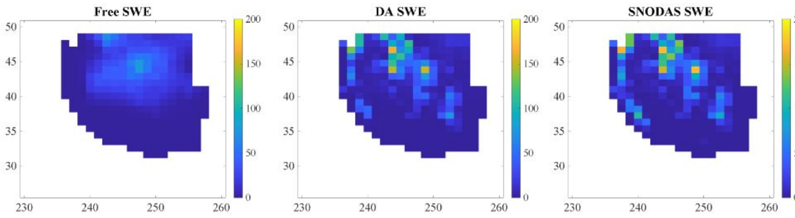


dart.ucar.edu

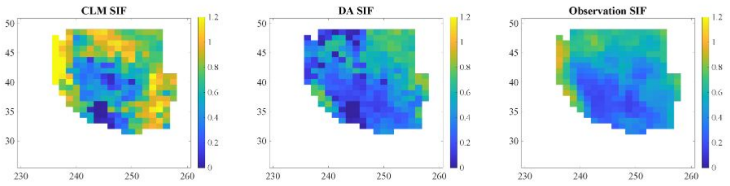
Shout out: Xueli’s talk

Monday, June 10th, 2024 CESM Land Agenda
2:08pm- **Assimilating SIF and SWE observations into CLM using DART to improve GPP over the high mountains in the Western US**
Xueli Huo

Despite challenges with snow input and melting, DA successfully corrects Snow Water Equivalent over the Western US

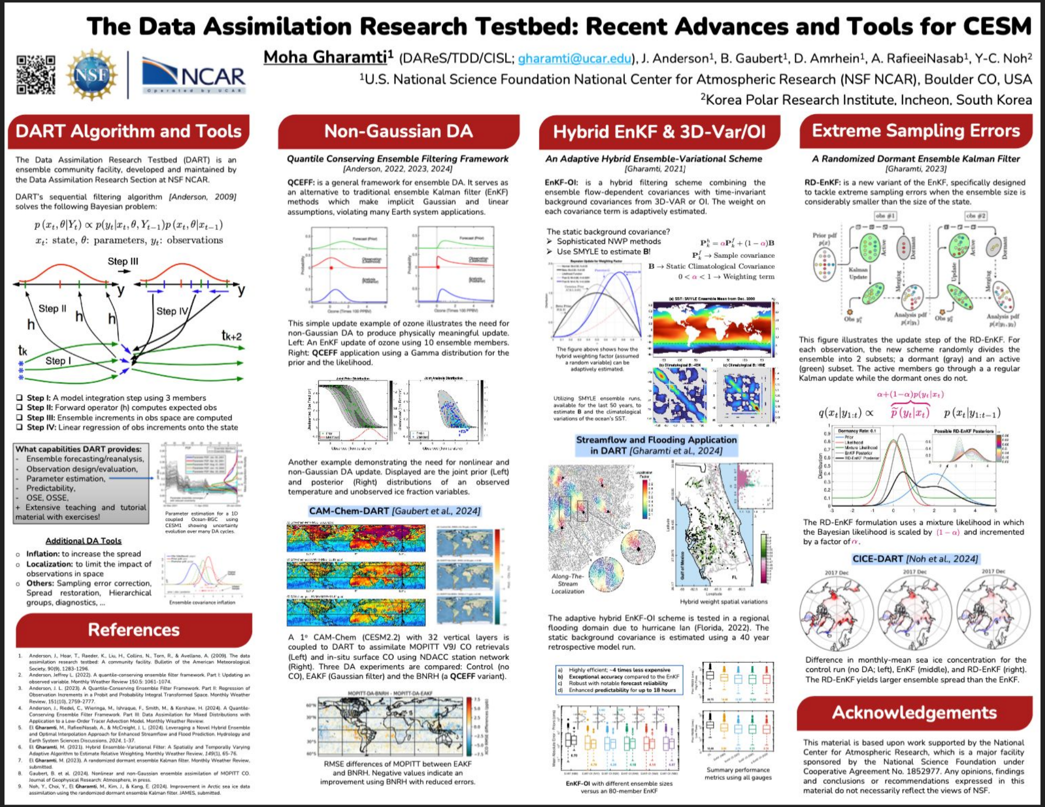


Though the regional average of SIF aligns well with the observation, SIF is overcorrected along the Cascade Range and in Montana



Need to refine the way of leaf nitrogen altered by DA to effectively impact GPP

Shout out: Moha & Kevin’s posters



Essential Tools for Predictability Studies Provided by the Data Assimilation Research Testbed

Importance of ensemble forcing to land assimilations and a couple showing what happened to the Reanalysis during COVID (aircraft obs disappeared, both biases increased).

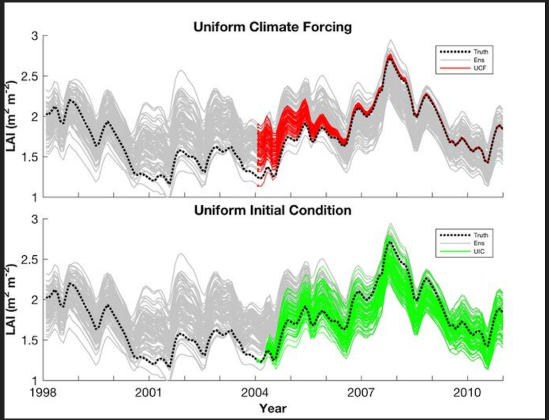


Figure courtesy of Andy Fox.

SIParCS

DART NUOPC

CROCODILE

pyDART

SIParCS project

"Optimizing ensemble data assimilation performance for coupled Earth System models" investigating the use of NUOPC/ESMF for DA, and comparing this to traditional 'offline' modes of Data Assimilation.

Anh Pham
Suman Shekhar

SIParCS project

"Optimizing ensemble data assimilation performance for coupled Earth System models" investigating the use of NUOPC/ESMF for DA, and comparing this to traditional 'offline' modes of Data Assimilation.

Anh Pham
Suman Shekhar

Mentors:
Dan Amrhein
Helen Kershaw

SIParCS project

"Optimizing ensemble data assimilation performance for coupled Earth System models" investigating the use of NUOPC/ESMF for DA, and comparing this to traditional 'offline' modes of Data Assimilation.

Anh Pham
Suman Shekhar

Mentors:
Dan Amrhein
Helen Kershaw

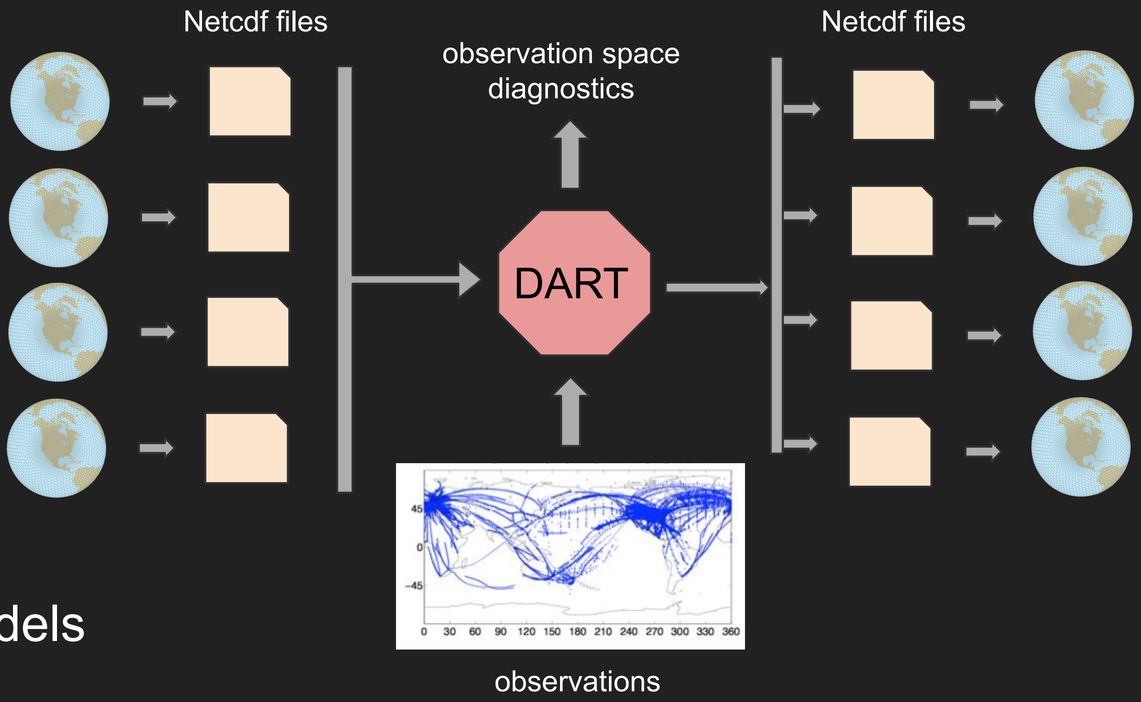
Thanks to the ESMF team and Alper Altuntas

SIParCS project

Every assimilation window:

- Advance the ensemble of models
- Every model writes their state to file
- DART reads all model states and observations for the time window
- Assimilation updates the ensemble of states
- DART writes updated state files
- Models restart with updated states

Lots of data movement
“**transpose**”



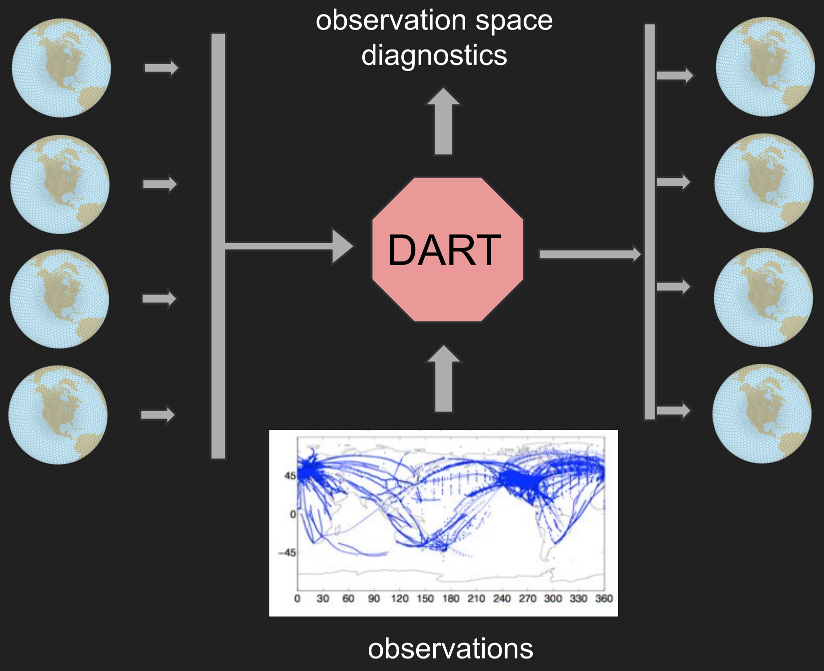
IO: Models -> DART -> Models

SIParCS project

Every assimilation window:

- Advance the ensemble of models
- DART reads the observations for the time window
- Assimilation updates the ensemble of states
- Models continue with updated states

Lots of data movement
“**transpose**”
Not going to disk



CROCODILE

5 year collaboration between NSF NCAR and WHOI

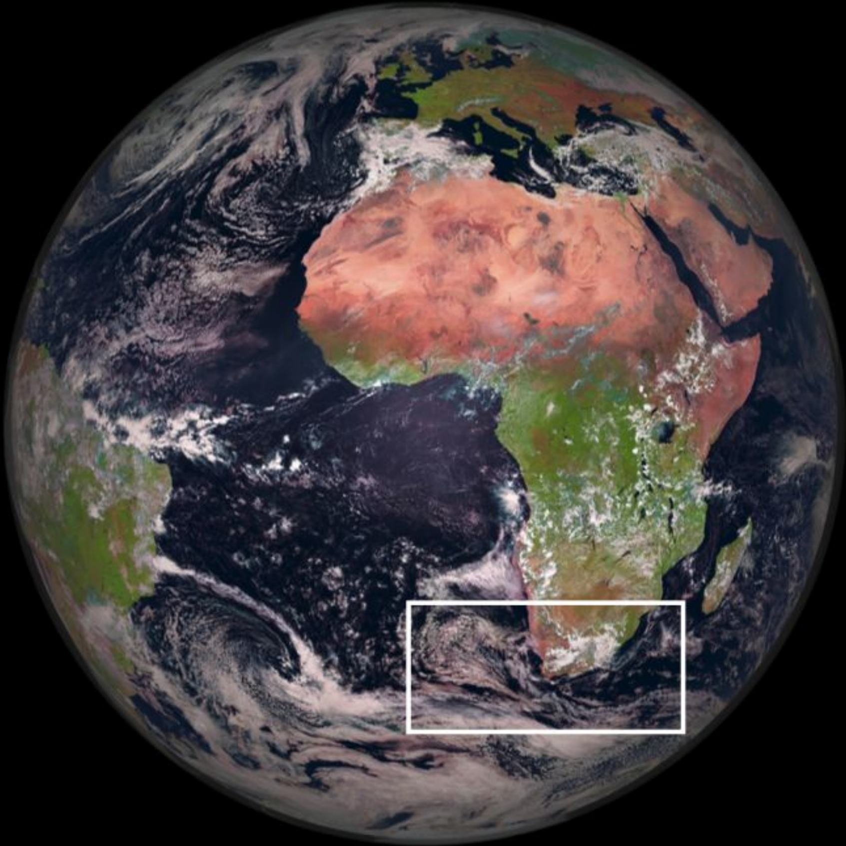
Motivation

How do we increase usability of infrastructure that translates global dynamics to human / actionable scales?

Regional model configuration requires setup and tuning.

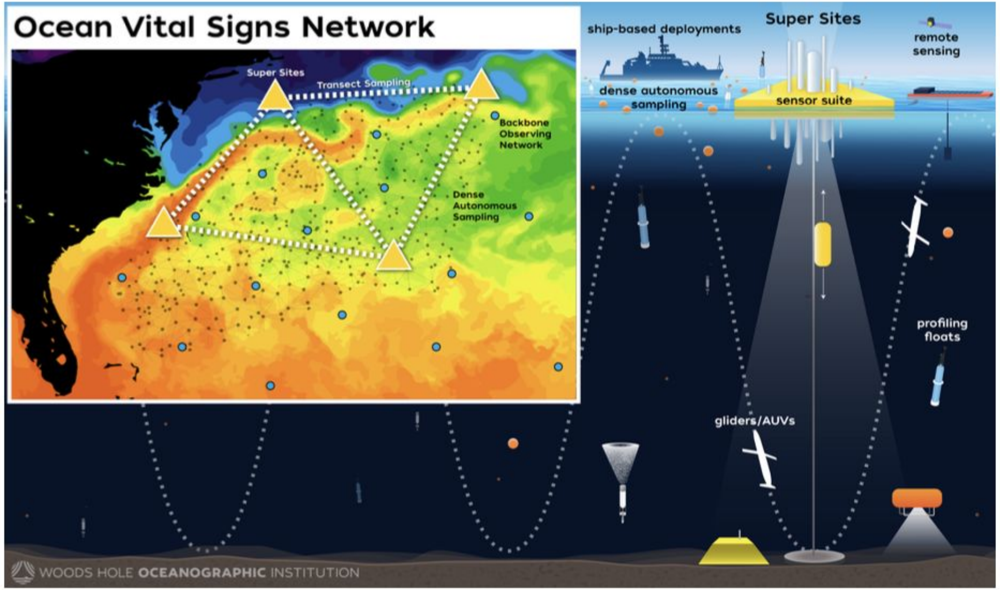
Data assimilation requires years of effort and technical capacity building.

Few have **access** to the computational resources and tools required for configuring and running these systems and then analyzing the relatively large data sets they generate.



Dan Amrhein

Use case 1: The Ocean Vital Signs Network (OVSN) and ocean CDR



Designing next-generation ocean observing systems

WOODS HOLE OCEANOGRAPHIC INSTITUTION

Dan Amrhein

CROCODILE

DART

5 year collaboration between NSF NCAR and WHOI

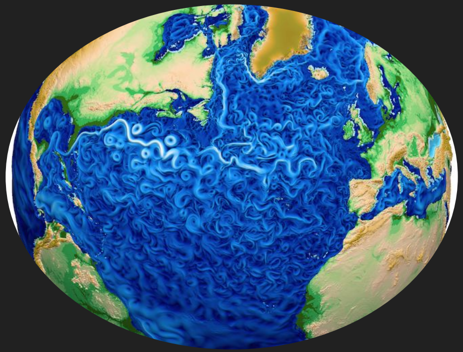
CROCODILE

(py)DART

5 year collaboration between NSF NCAR and WHOI

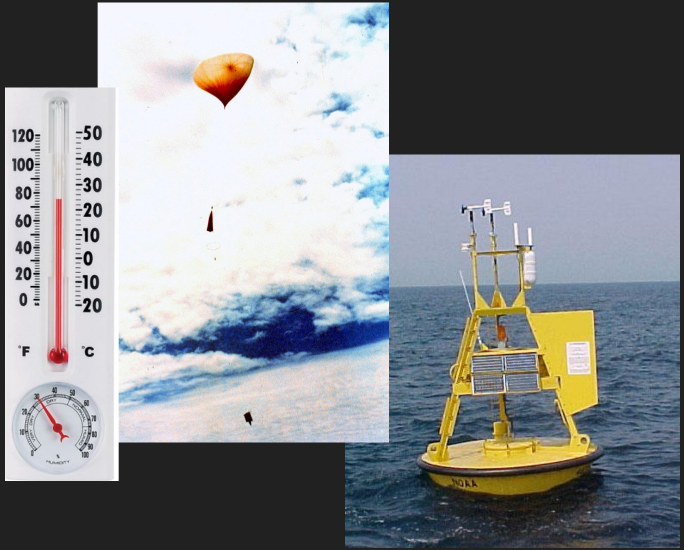
Data Assimilation Research Testbed

Models: MOM6



State Space

Observations

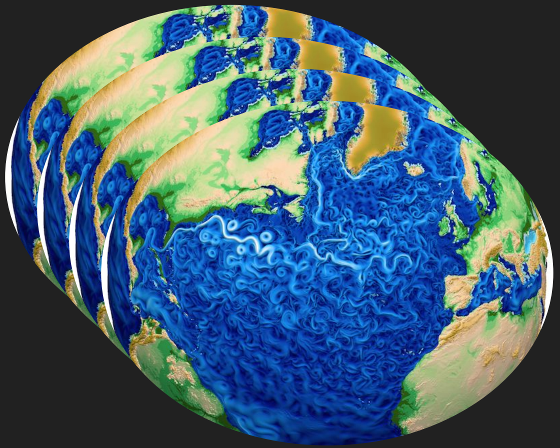


Observation Space

Data Assimilation Research Testbed

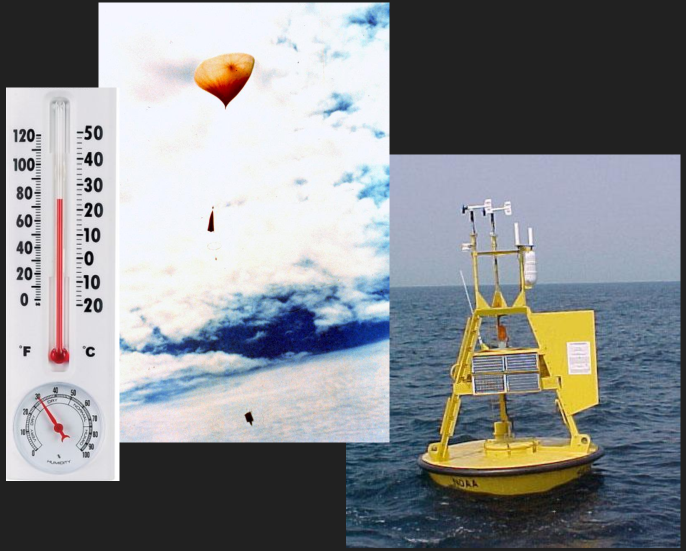
Models: MOM6

Ensemble



State Space

Observations



Observation Space

DART-MOM6

DART
latest

Search docs

MOM6

A new ocean component model based on the Modular Ocean Model version 6 (MOM6) has been incorporated into CESM and is anticipated to replace POP2 as the default ocean component in CESM3. An early functional release of the MOM6 ocean component has been made available to users beginning with CESM2.2. Instructions for using MOM6 in CESM are available on the MOM_interface GitHub Wiki.

This DART-MOM6 interface was developed for MOM6 within the CESM framework.

MOM6 time

The default in CESM is to run with no leap years. To assimilate real observations, we need to switch to the Gregorian calendar to account for leap years.

```
./xmlchange CALENDAR=GREGORIAN
```

To illustrate what happens if you do not set CALENDAR=GREGORIAN, here is an example where the RUN_STARTDATE is set to 2015-02-01 and MOM6 is run for 10 days.

```
./xmlchange RUN_STARTDATE=2015-02-01
```

The MOM6 restart file has the following meta data, where Time is days from year 1.

Observation Sequence

World Ocean Database example

DA Input: obs_seq.**out**

- The observation types in the file
- Locations of the observations
- Quality control value(s) for each observation
- Observation + any observation specific metadata
- Time and order of observations

```
 obs_sequence
obs_kind_definitions
          13
          15 FLOAT_SALINITY
          16 FLOAT_TEMPERATURE
          23 GLIDER_SALINITY
          24 GLIDER_TEMPERATURE
          27 MOORING_SALINITY
          28 MOORING_TEMPERATURE
          30 BOTTLE_SALINITY
          31 BOTTLE_TEMPERATURE
          32 CTD_SALINITY
          33 CTD_TEMPERATURE
          43 XBT_TEMPERATURE
          46 APB_SALINITY
          47 APB_TEMPERATURE
  num_copies:            1  num_qc:            1
  num_obs:        66643  max_num_obs:        66643
WOD observation
WOD QC
  first:            1  last:        66643
 OBS            1
  2.704750061035156E-002
  0.000000000000000E+000
          -1            2          -1
obdef
loc3d
     3.753021682990578        1.345058936474565        10.00000000000000      3
kind
          32
 43322     151163
  2.500000000000000E-007
 OBS            2
  2.704809951782227E-002
  0.000000000000000E+000
           1            3          -1
obdef
loc3d
     3.753021682990578        1.345058936474565        15.00000000000000      3
kind
          32
 43322     151163
  2.500000000000000E-007
 OBS            3
  2.704929924011230E-002
  0.000000000000000E+000
           2            4          -1
obdef
loc3d
     3.753021682990578        1.345058936474565        20.00000000000000      3
kind
          32
```

Observation Sequence

Radiance example

DA Output: obs_seq.**final**

- various ‘copies’ of each observation
  - Observation value
  - Truth
  - Prior and Posterior Forward Operators, mean, standard deviation

```
 obs_sequence
obs_type_definitions
           1
         216 GOES_16_ABI_RADIANCE
  num_copies:            4  num_qc:            2
  num_obs:        77042  max_num_obs:        77042
observation
prior ensemble mean
prior ensemble spread
prior ensemble member      1
GOES QC
DART quality control
  first:            1  last:        77042
 OBS           1
   2.54618100000000
   2.67347519169475
  8.220037240577385E-002
   2.70210881859020
  0.000000000000000E+000
  0.000000000000000E+000
          -1           2          -1
obdef
loc3d
     4.102617740631104        0.6268096566200256         35000.00000000000
kind
         216
 visir
   125.829100000000        68.2444100000000       -888888.000000000
  -888888.000000000
           4          16          44           2
  -888888.000000000
           1
 43072      152809
  6.250000000000000E-002
 OBS           2
   2.61012400000000
   2.72617267485070
  8.868187079127991E-002
   2.73930031819035
  0.000000000000000E+000
```

Observation Sequence

Radiance example

DA Output: obs_seq.**final**

- various 'copies' of each observation
  - Observation value
  - Truth
  - Prior and Posterior Forward Operators, mean, standard deviation

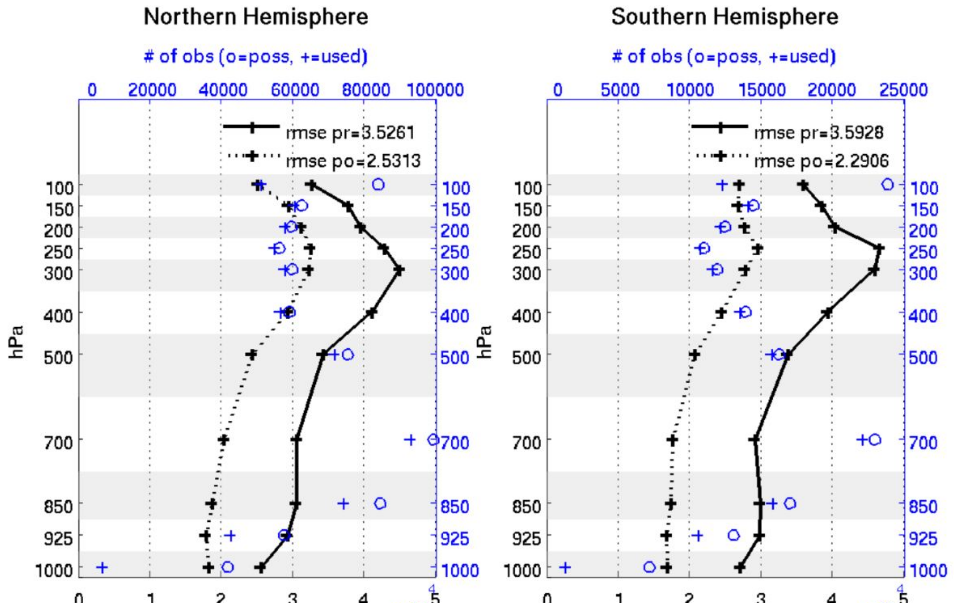
Observations can have additional metadata

```
 obs_sequence
obs_type_definitions
           1
         216 GOES_16_ABI_RADIANCE
  num_copies:            4  num_qc:            2
  num_obs:        77042  max_num_obs:        77042
observation
prior ensemble mean
prior ensemble spread
prior ensemble member      1
GOES QC
DART quality control
  first:            1  last:        77042
 OBS            1
   2.54618100000000
   2.67347519169475
  8.220037240577385E-002
   2.70210881859020
  0.000000000000000E+000
  0.000000000000000E+000
          -1           2          -1
obdef
loc3d
     4.102617740631104        0.6268096566200256         35000.00000000000
kind
         216
 visir
   125.829100000000        68.2444100000000       -888888.000000000
  -888888.000000000
           4          16          44           2
  -888888.000000000
           1
 43072       152809
  6.250000000000000E-002
 OBS            2
   2.61012400000000
   2.72617267485070
  8.868187079127991E-002
   2.73930031819035
  0.000000000000000E+000
```

Fortran + Matlab Observation Space Diagnostics

plot_profile.m plots the spatial and temporal average of any specified quantity as a function of height. The number of observations possible and used are plotted on the same axis.

```
fname      = 'POP11/obs_diag_output.nc';        % netcdf file produced by 'obs_diag'
copystring = 'rmse';                            % 'copy' string == quantity of interest
plotdat    = plot_profile(fname,copystring);
```

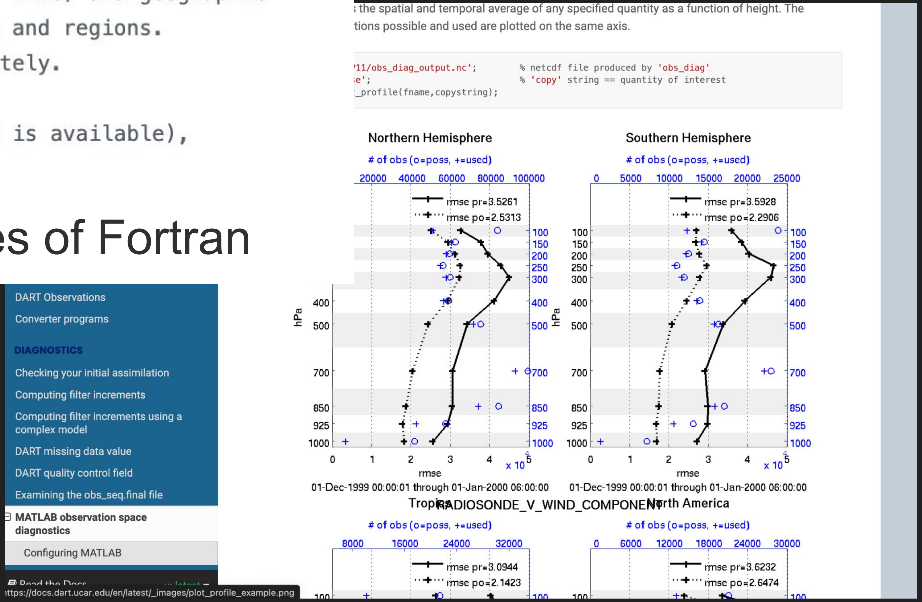


Fortran + Matlab Observation Space Diagnostics

```fortran
!> The programs defines a series of epochs (periods of time) and geographic
!> regions and accumulates statistics for these epochs and regions.
!> All 'possible' observation types are treated separately.
!> The results are written to a netCDF file.
!> If the rank histogram is requested (and if the data is available),
!> only the PRIOR rank is calculated.

program obs_diag
```

4370 lines of Fortran



Matlab
- not open source
- not freely available

Fortran + Matlab Observation Space Diagnostics

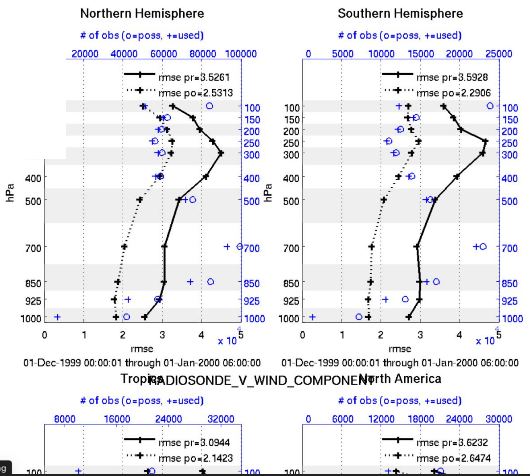
```
!> The programs defines a series of epochs (periods of time) and geographic
!> regions and accumulates statistics for these epochs and regions.
!> All 'possible' observation types are treated separately.
!> The results are written to a netCDF file.
!> If the rank histogram is requested (and if the data is available),
!> only the PRIOR rank is calculated.

program obs_diag
```

4370 lines of Fortran

s the spatial and temporal average of any specified quantity as a function of height. The
tions possible and used are plotted on the same axis.

```
'11/obs_diag_output.nc';        % netcdf file produced by 'obs_diag'
e';                             % 'copy' string == quantity of interest
_profile(fname,copystring);
```



DART Observations
Converter programs
DIAGNOSTICS
Checking your initial assimilation
Computing filter increments
Computing filter increments using a complex model
DART missing data value
DART quality control field
Examining the obs_seq.final file
MATLAB observation space diagnostics
Configuring MATLAB

https://docs.dart.ucar.edu/en/latest/_images/plot_profile_example.png

Matlab
- not open source
- not freely available

python™

pyDARTdiags

obs_seq.{out/final} -> pandas dataframe

github.com/NCAR/pyDARTdiags

```
 obs_sequence
obs_type_definitions
          22
           4 GPSRO_REFRACTIVITY
          51 RADIOSONDE_U_WIND_COMPONENT
          52 RADIOSONDE_V_WIND_COMPONENT
          55 RADIOSONDE_TEMPERATURE
          56 RADIOSONDE_SPECIFIC_HUMIDITY
          62 AIRCRAFT_U_WIND_COMPONENT
          63 AIRCRAFT_V_WIND_COMPONENT
          64 AIRCRAFT_TEMPERATURE
          66 ACARS_U_WIND_COMPONENT
          67 ACARS_V_WIND_COMPONENT
          68 ACARS_TEMPERATURE
          70 MARINE_SFC_U_WIND_COMPONENT
          71 MARINE_SFC_V_WIND_COMPONENT
          72 MARINE_SFC_TEMPERATURE
          73 MARINE_SFC_SPECIFIC_HUMIDITY
          80 SAT_U_WIND_COMPONENT
          81 SAT_V_WIND_COMPONENT
          83 AIRS_TEMPERATURE
          84 AIRS_SPECIFIC_HUMIDITY
          90 RADIOSONDE_SURFACE_ALTIMETER
          92 MARINE_SFC_ALTIMETER
          93 LAND_SFC_ALTIMETER
  num_copies:          83  num_qc:           2
  num_obs:      1091025  max_num_obs:      1091025
observation
prior ensemble mean
prior ensemble spread
prior ensemble member       1
prior ensemble member       2
```

Launcher | obs_seq_test.ipynb | matt_plot.ipynb

Code

[1]:
```python
import sys
import os
import pandas as pd
import numpy as np
import plotly.express as px
sys.path.append(os.path.abspath("/Users/hkershaw/DART/Projects/Diagnostics/pyDART/
import obs_sequence as dart_os
```

[2]:
```python
obs_seq = dart_os.obs_sequence('/Users/hkershaw/DART/Projects/Diagnostics/data/obs
```

[3]:
```python
obs_seq.df.head()
```

[3]:

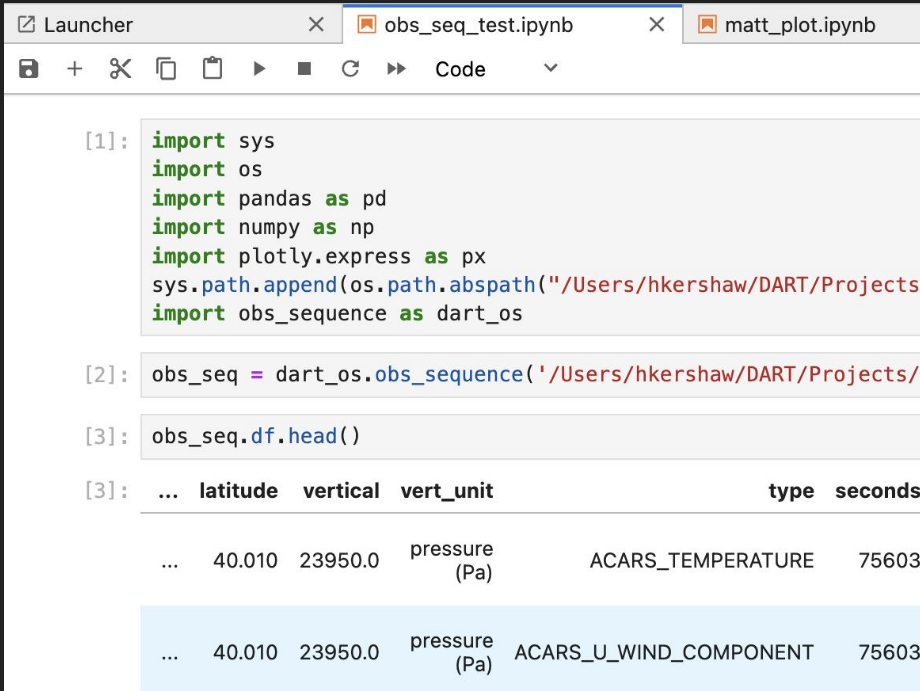
...	latitude	vertical	vert_unit	type	seconds	days	time	ob
...	40.010	23950.0	pressure (Pa)	ACARS_TEMPERATURE	75603	153005	2019-12-01 21:00:03	
...	40.010	23950.0	pressure (Pa)	ACARS_U_WIND_COMPONENT	75603	153005	2019-12-01 21:00:03	

pyDARTdiags

obs_seq.{out/final} -> pandas dataframe

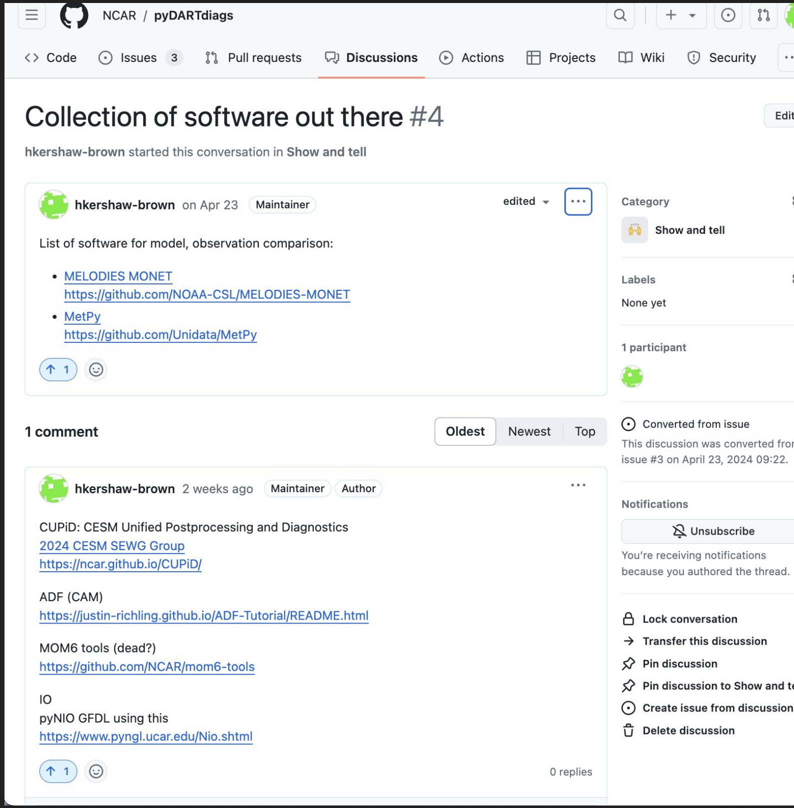
github.com/NCAR/pyDARTdiags

- Interactive python Observation Space Diagnostics: language of data science
- User extensible, & allows us to improve obs_seq format
- Observation preprocessing in be done in python?



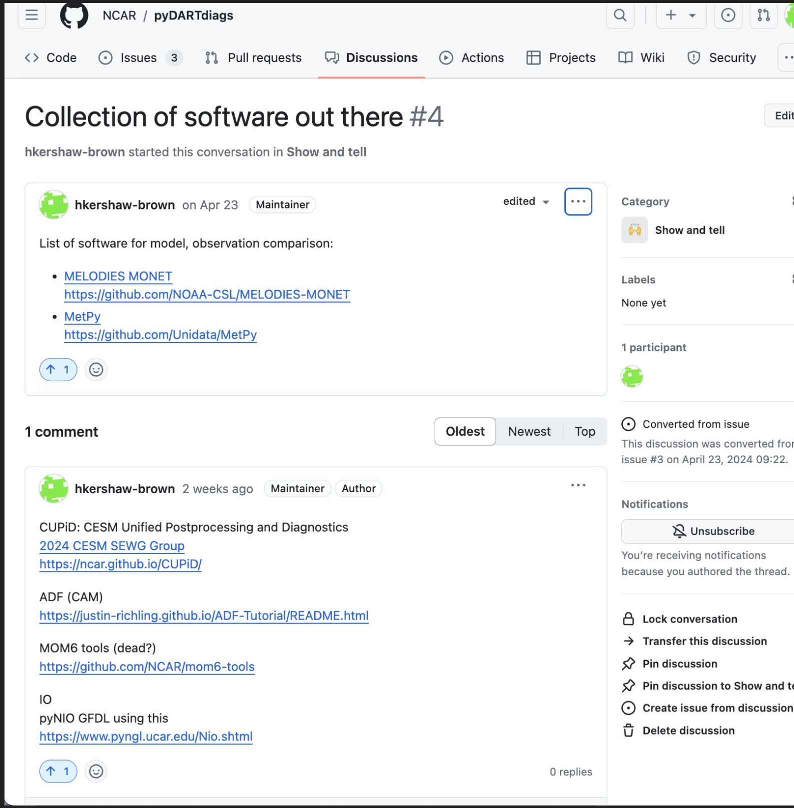
Next Steps

- What diagnostics are people interested in?
- What existing tools can we plug in to?
- Xarray + dask
    - Test case 20 year DART-CAM reanalysis
- pyDARTdiags as a preprocessing tool
- Pipeline of ocean data into CESM-DART



Next Steps

- What diagnostics are people interested in?
- What existing tools can we plug in to?
- Xarray + dask
    - Test case 20 year DART-CAM reanalysis
- pyDARTdiags as a preprocessing tool
- Pipeline of ocean data into CESM-DART



hkershaw@ucar.edu