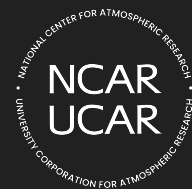


# CESM-DART

## Ocean Data Assimilation

Helen Kershaw  
DAReS NCAR



# Terminology

**DAReS** - the group of people

**DART** - the software we produce



# Team

## DAReS

Jeff Anderson  
Kevin Raeder  
Marlee Smith  
Brett Raczka  
Ben Johnson  
Helen Kershaw  
Moha Gharamti  
Anderson Chauphan

## CGD

Dan Amrhein  
Alper Altuntas  
Brian Dobbins  
Jim Edwards

# Ensemble Data Assimilation

# Ensemble Data Assimilation



Group of model forecasts

# Ensemble Data Assimilation



Group of model forecasts

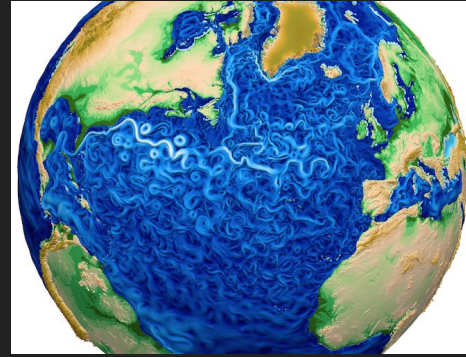
Measurements



# Ensemble Data Assimilation



Group of model forecasts



Improved estimate

Measurements



Featured project: Computational & Information Systems Lab & Research Applications Lab Collaboration

# PREDICTING FLOODS AND PROTECTING LIVES



## DATA ASSIMILATION FOR THE ENTIRE EARTH SYSTEM

Use ensemble DA techniques with geophysical models spanning the



## USE DATA FROM ANY SOURCE, TEST MANY ALGORITHMS

Assimilate any suitable observations. Swap out filter and



## LEARN ON LAPTOPS, RUN ON SUPERCOMPUTERS

Compile without MPI for conceptual models or with MPI for GCMs on

Featured project: NC State, UC San Diego, MIT & KAUST Collaboration

# UNDERSTANDING GULF OF MEXICO EDDY DYNAMICS



## DATA ASSIMILATION FOR THE ENTIRE EARTH SYSTEM

Use ensemble DA techniques with models spanning the



## USE DATA FROM ANY SOURCE, TEST MANY ALGORITHMS

Assimilate any suitable observations. Swap out filter and



## LEARN ON LAPTOPS, RUN ON SUPERCOMPUTERS

Compile without MPI for conceptual models or with MPI for GCMs on

Featured project: University of Michigan, NCAR, NASA & NRL Collaboration

# NEXT-GENERATION SPACE WEATHER PREDICTION



## DATA ASSIMILATION FOR THE ENTIRE EARTH SYSTEM



## USE DATA FROM ANY SOURCE, TEST MANY ALGORITHMS



## LEARN ON LAPTOPS, RUN ON SUPERCOMPUTERS

dart.ucar.edu



Existing **CESM** work

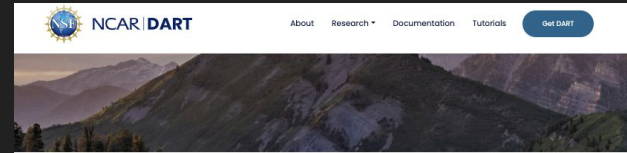
# Existing CESM work

2003 DART-CAM2 Breckenridge

# Existing CESM work

2003 DART-CAM2 Breckenridge

2021 DART-CAM6 Reanalysis



## The CAM6+DART Reanalysis for Earth System Science

The Earth system can be viewed as distinct but connected components: atmosphere, land, ocean, cryosphere, biosphere, et cetera (Fig. 1). Data assimilation can help us create the best available representation of the state of Earth, but it requires relevant observations and a forecast model which represents all of the components of interest.

Earth system components interact in many ways at the interfaces between them.

NCAR's Community Earth System Model (CESM) can run forecasts with a flexible choice of "active" components, in which the component model state evolves according to equations, and "data" components, in which the component state is read from a data file. For example, to generate atmospheric forecasts, the configuration could have active atmosphere and land components, but simply read sea surface temperatures (SSTs) from data files, instead of running an expensive, active, ocean component to generate SSTs. CESM has been developed at NCAR for decades, and has evolved to work effectively with DART through the efforts of the CESM Software Engineering Group (M. Vertenstein, S. Goldhaber, J. Edwards) and R. Montuoro.

Data assimilation has been extensively applied to the atmosphere for decades, but not to the surface components until more recently. One hurdle has been that the surface components tend to be more slowly varying, so they require atmospheric forcing over long time spans. It's expensive to run an atmospheric model, and many experiments may require the same atmospheric forcing, which would be stored in an appropriate patch files. In other words, you need an ensemble of surface models to maintain the necessary ensemble



Figure 1: Earth system components interact in many ways at the interfaces between them.

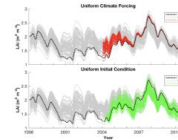


Figure 2: A single atmospheric forcing allows an ensemble to collapse (top). Multiple atmospheric forcings cause the spread to increase (bottom). Picture courtesy of A. Fox.

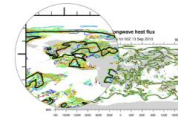


Figure 3: The downward longwave heat flux from 20 (of 80) ensemble members (various colors) and the ensemble mean (black).

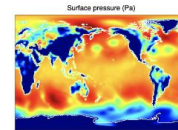


Figure 4: A 1-degree resolution model doesn't generate great hurricanes by itself, but here's hurricane Sandy shortly before landfall 02 10/29/2012 in the analysis, ensemble mean, surface pressure. Here's a comparison of central pressures: National Hurricane Center "best track" = 950 hPa, Knoff-Zehle-Courtesy pressure-wind relationship = 950 hPa. CAM6+DART ensemble mean

## Registry of Open Data on AWS

# CAM6 Data Assimilation Research Testbed (DART) Reanalysis: Cloud-Optimized Dataset

atmosphere climate climate model data assimilation forecast geoscience geospatial land meteorological weather zarr

## Description

This is a cloud-hosted subset of the CAM6+DART (Community Atmosphere Model version 6 Data Assimilation Research Testbed) Reanalysis dataset. These data products are designed to facilitate a broad variety of research using the NCAR CESM 2.1 (National Center for Atmospheric Research's Community Earth System Model version 2.1), including model evaluation, ensemble hindcasting, data assimilation experiments, and sensitivity studies. They come from an 80 member ensemble reanalysis of the global troposphere and stratosphere using DART and CAM6. The data products represent states of the atmosphere consistent with observations from 2011 through 2019 at 1 degree horizontal resolution and weekly frequency. Each ensemble member is

## Resources on AWS

Description  
Project data files

Resource type  
S3 Bucket

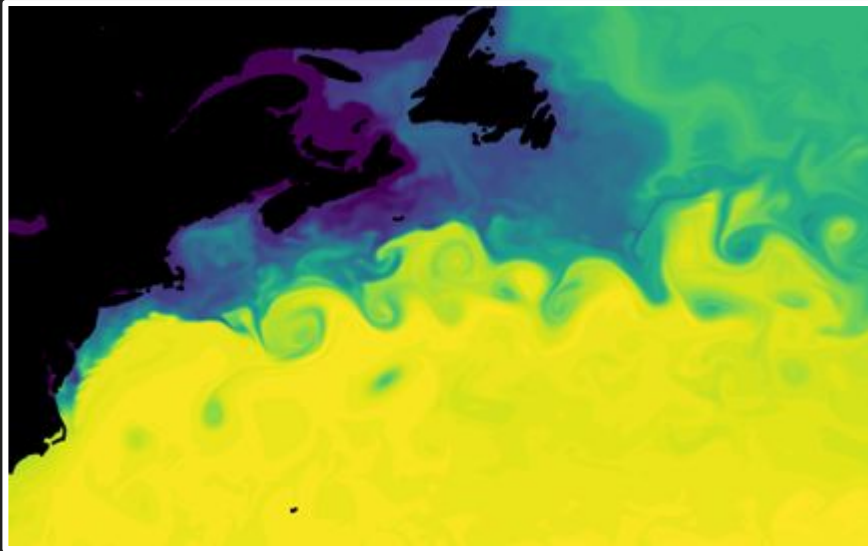
Amazon Resource Name (ARN)  
`arn:aws:s3::ncar-dart-cam6`

AWS Region  
`us-west-2`

[AWS CLI Access \(No AWS account required\)](#)



# POP Accelerated Scientific Discovery



**High-resolution ocean DA |**  
80-member ocean reanalyses  
spanning 2011-2017 at 1° and 0.1°  
*Ben Johnson, Moha Gharamti,  
Anna-Lena Deppenmeier, and Ian  
Grooms*

# MOM6 interface

The screenshot shows a web browser window displaying the NCAR DART documentation for the MOM6 interface. The browser's address bar shows the URL `docs.dart.ucar.edu/en/latest/models/MOM6/readme.html`. The page features a dark blue sidebar on the left with the NCAR DART logo and a search bar. The main content area is white and contains the following sections:

- Navigation:** Home icon, "/ MOM6", and "Edit on GitHub" link.
- Section Header:** "MOM6".
- Text:** "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."
- Text:** "This DART-MOM6 interface was developed for MOM6 within the CESM framework."
- Section Header:** "MOM6 checksum of restart files".
- Text:** "When reading in restart files, MOM6 verifies a checksum for each variable in the restart file. Data assimilation updates the data in the MOM6 restart file, which will cause the checksum verification to fail. To use DART-MOM6 with CESM turn off the checksum verification using the `user_nml_mom` namelist option:"
- Code Block:** A light gray box containing the code `RESTART_CHECKSUMS_REQUIRED = False`.

# MOM6 interface

The screenshot shows a web browser window displaying the MOM6 interface documentation. The browser's address bar shows the URL `docs.dart.ucar.edu/en/latest/models/MOM6/readme.html`. The page header features the NCAR DART logo and a navigation menu with the word "latest" and a search box labeled "Search docs". The main content area is titled "MOM6" and includes a link to "Edit on GitHub". The text describes the MOM6 model as a new ocean component model based on the Modular Ocean Model version 6, incorporated into CESM. It also mentions that 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. A paragraph states that this DART-MOM6 interface was developed for MOM6 within the CESM framework. A section titled "MOM6 checksum of restart files" explains that when reading in restart files, MOM6 verifies a checksum for each variable in the restart file. Data assimilation updates the data in the MOM6 restart file, which will cause the checksum verification to fail. To use DART-MOM6 with CESM, users should turn off the checksum verification using the `user_nml_mom` namelist option. A code block shows the configuration: 

```
RESTART_CHECKSUMS_REQUIRED = False
```

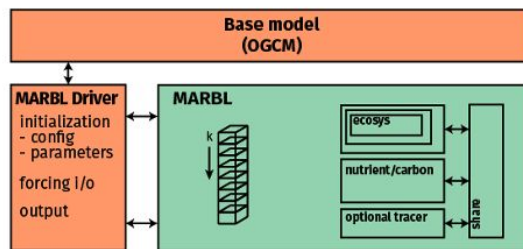
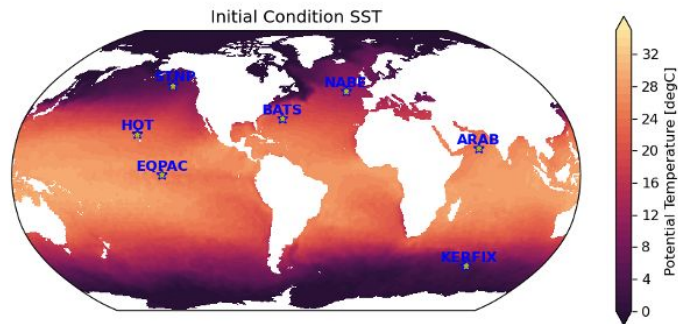
cesm2\_3\_alpha15a

# Parameter estimation MARBL

## DA and parameter estimation in MARBL-MOM6

Robin Armstrong (SIParCS summer student)

with Mike Levy, Kristen Krumhardt, Helen  
Kershaw, Alper Altuntas, Keith Lindsay, Moha  
Gharamti, Matt Long, and Dan Amrhein



### Project Goals

Compare 1-D column  
MOM6/MARBL to data  
at BATS, HOT, etc.

Estimate BGC parameters (nutrient  
uptake, productivity, predation, ...)  
using the ensemble Kalman filter

Evaluate parameters  
for global simulations

DART: observations and models



# DART: observations and models



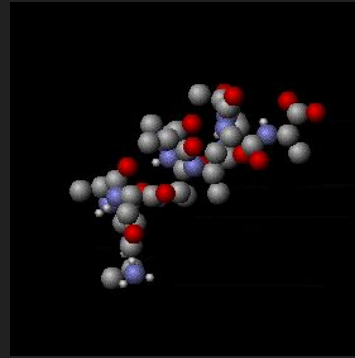
observation



# DART: observations and models

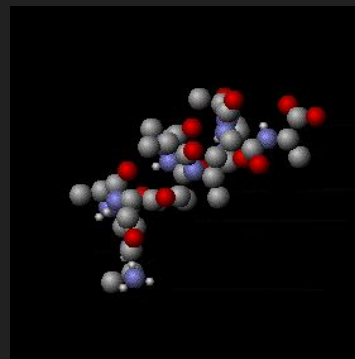


observation

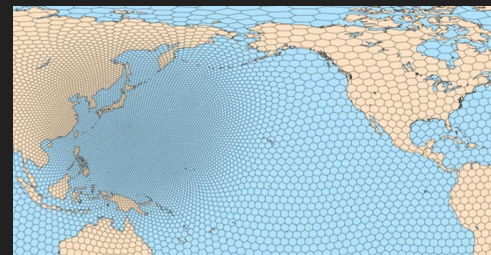


physical quantity

# DART: observations and models



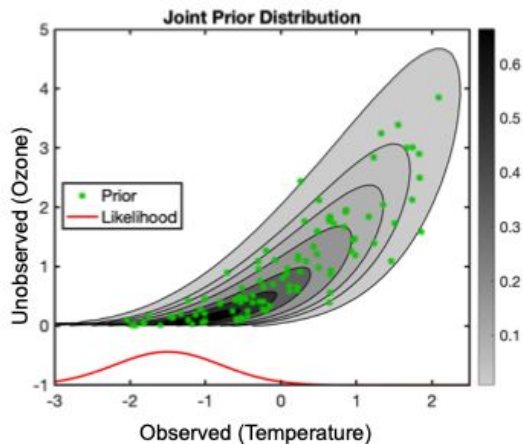
physical quantity



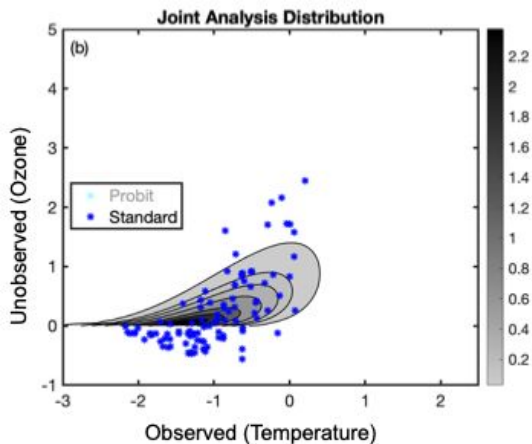
forecast model

## Standard EnKF: Challenged by Non-Gaussian and Nonlinear Relations

Prior for normal-gamma distribution with 100 member ensemble.

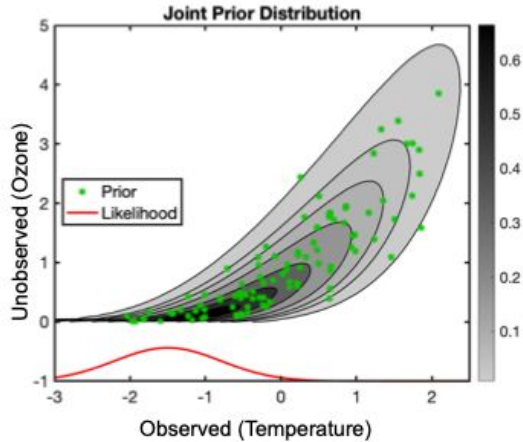


Posterior ensemble has problems.

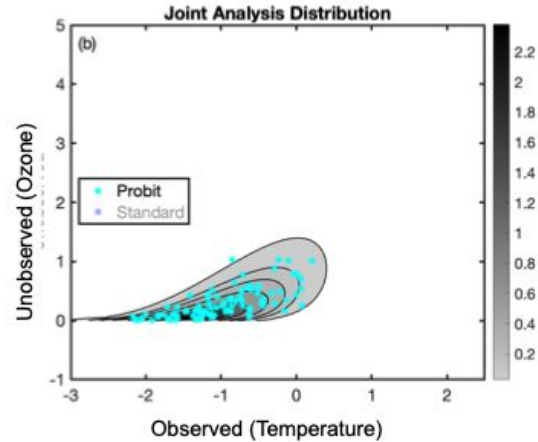


## DART: Novel, General Solutions for Nonlinear, Non-Gaussian Problems

Prior for normal-gamma distribution with 100 member ensemble.

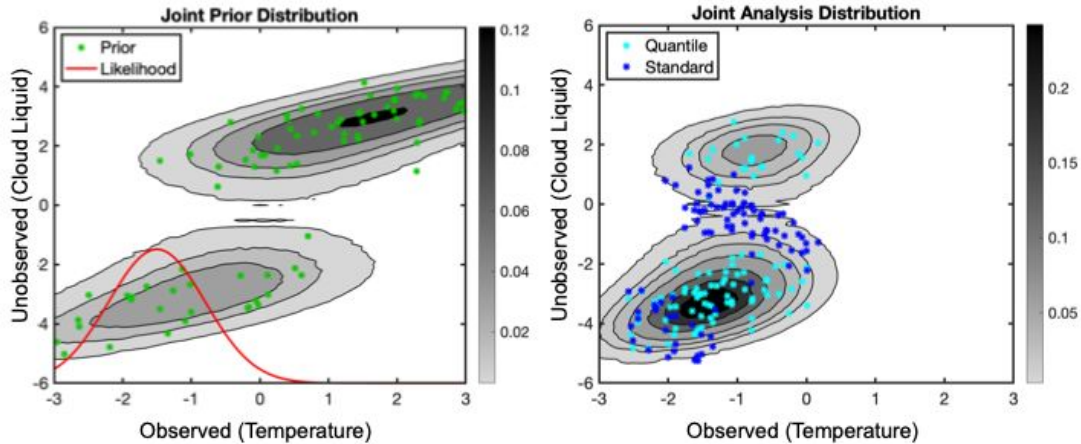


Bounds enforced. Nonlinear aspect respected.



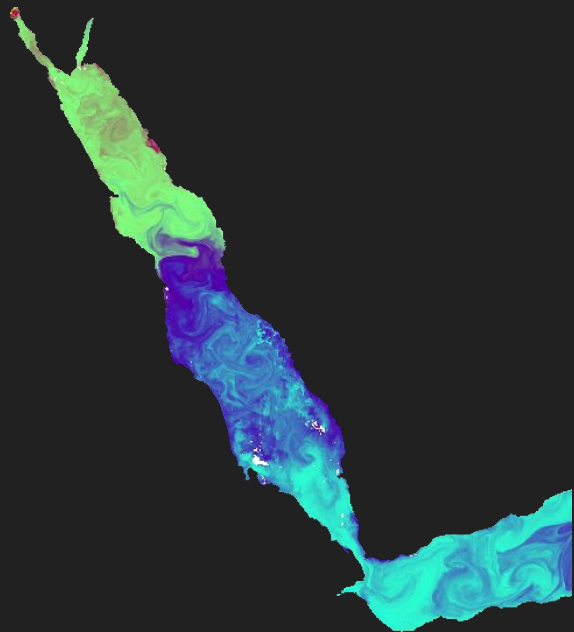
### Example 3: Normal observed, binormal unobserved

New methods move members from one mode to the other.  
Also adjusts ensemble in the modes.  
Relevant for initiating convection, for example.



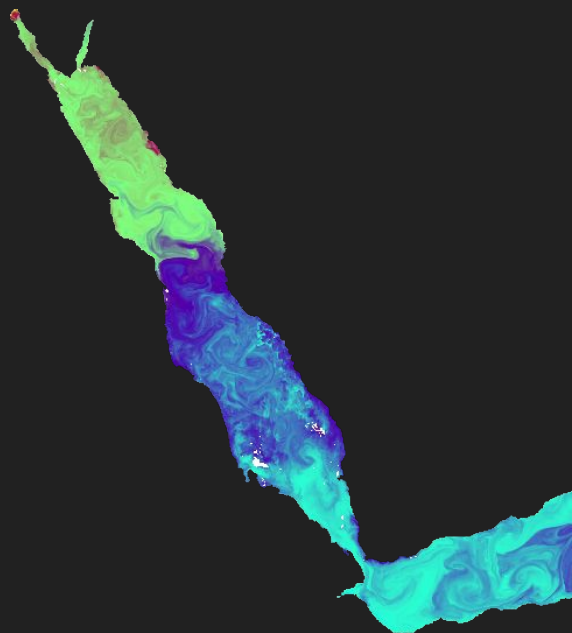
# Workflows

# Workflows

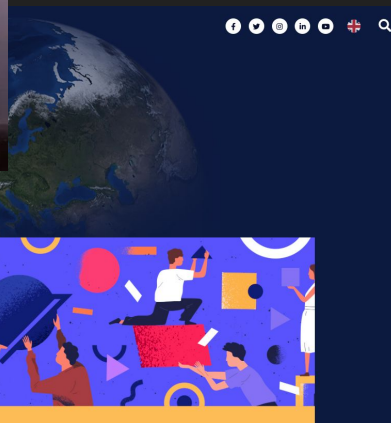
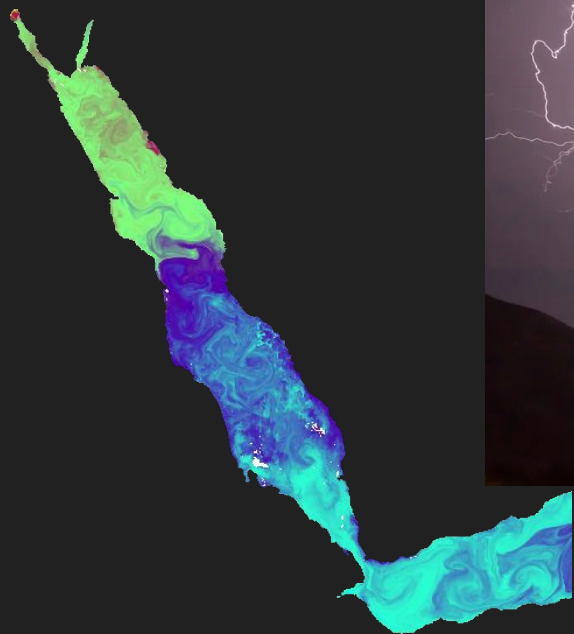




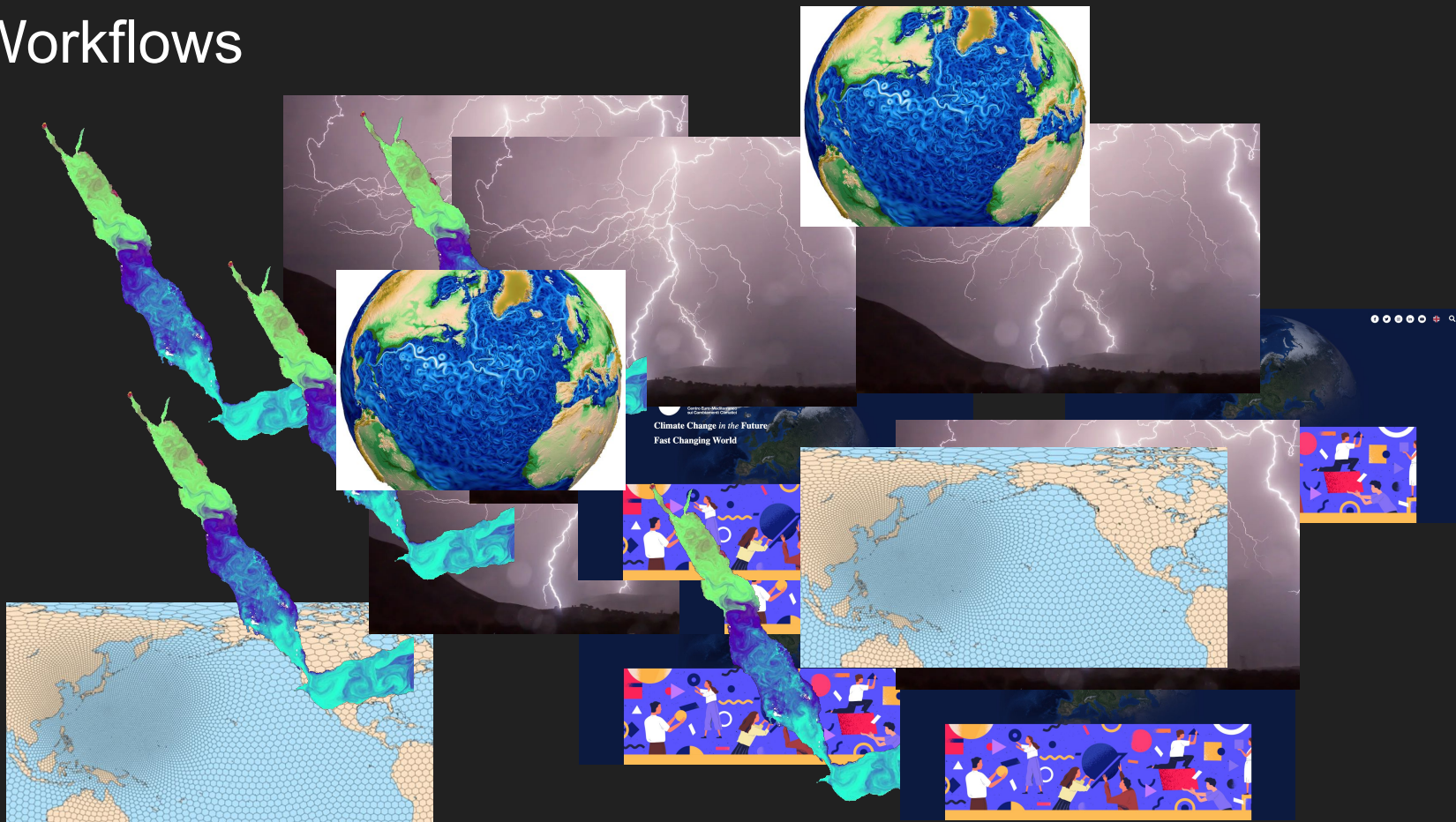
# Workflows



# Workflows



# Workflows



Integrate **DART** with **CESM**

# Key scientific motivation

How can we leverage observational constraints to improve model development and process representation and quantify Earth System Predictability?

# DART Compset

- **Leverage major existing capabilities** in CESM and the Data Assimilation Research Testbed (DART)
- **New community facility** for model-data comparison and **Earth System DA**
- Allow users to run CESM with DA in one or more of ocean, atmosphere, sea ice, and land components
- Leverage CESM's strengths in representing **complex coupled processes**

What is **tricky**?

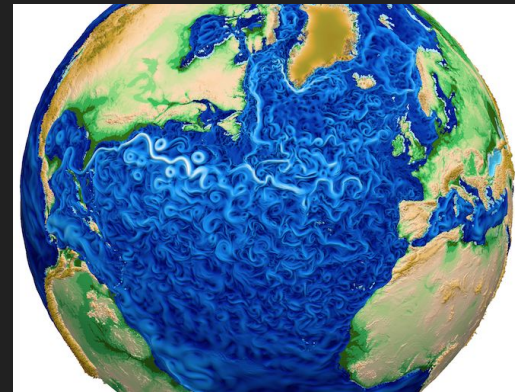
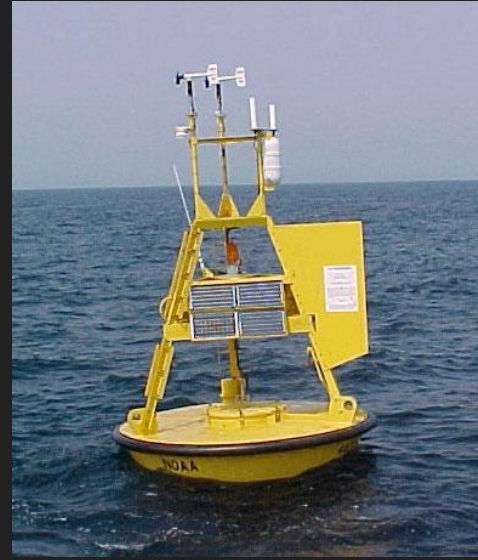
What **time** is it?



What **time** is it?



What **time** is it?

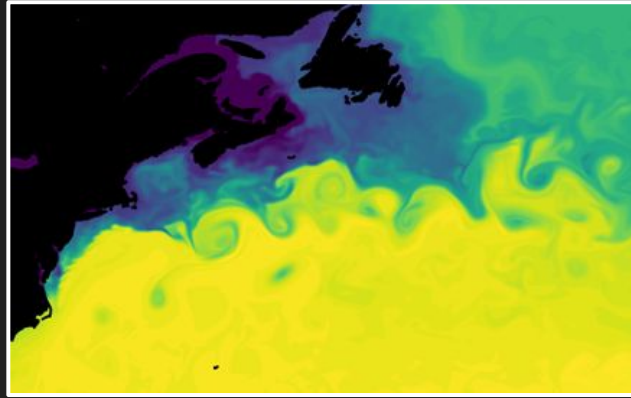


Ensemble

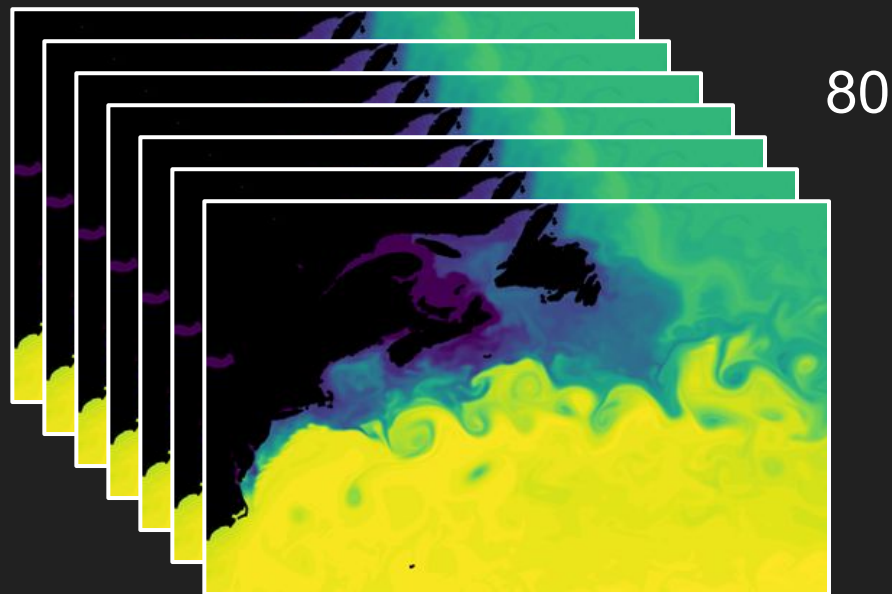
*How* many?

Ensemble

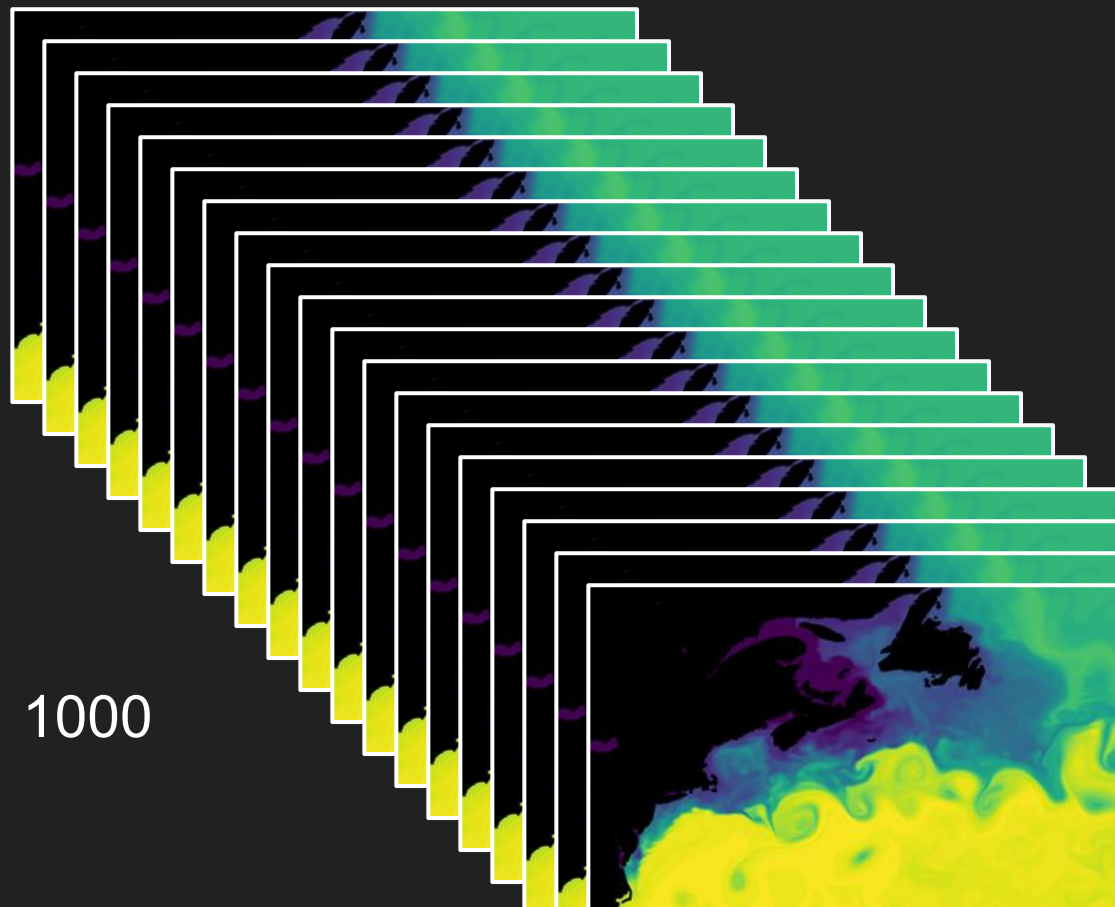
# Ensemble

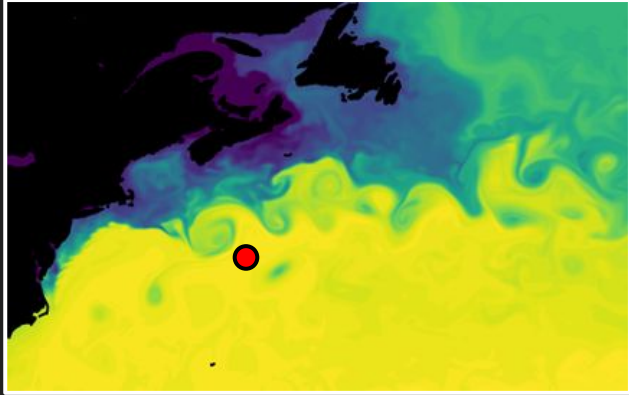


# Ensemble

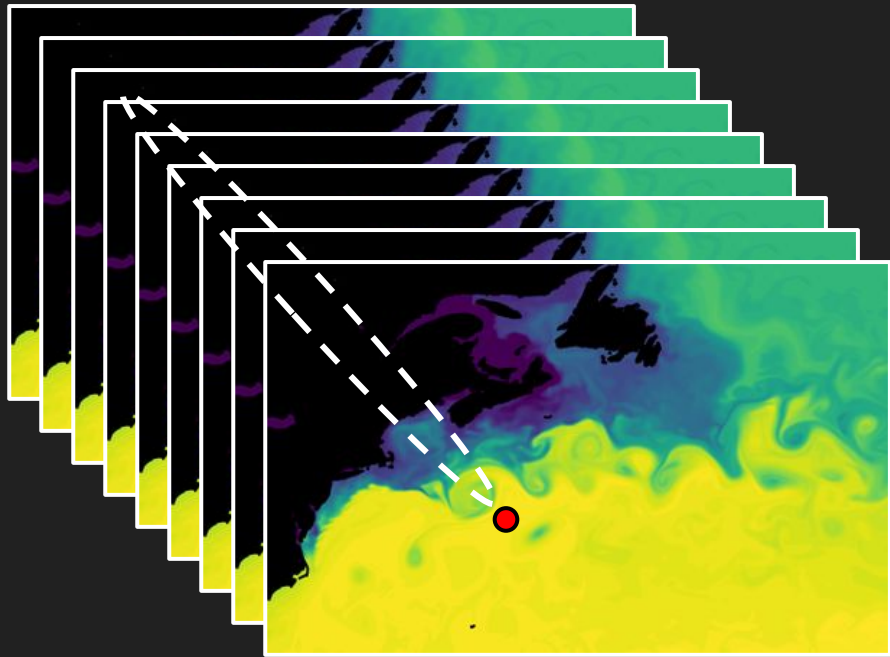


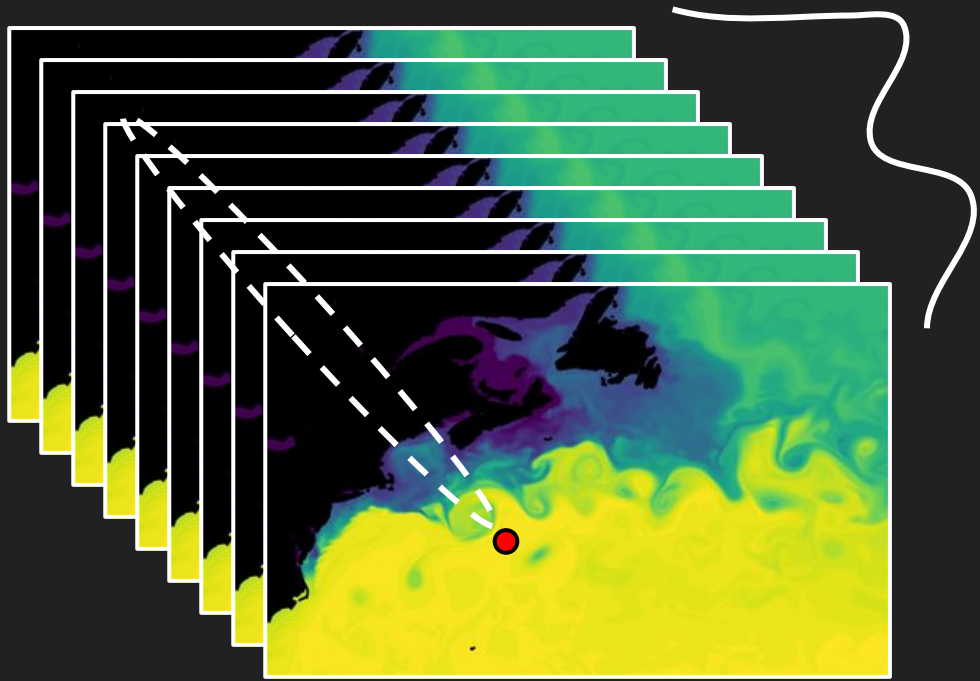
# Ensemble



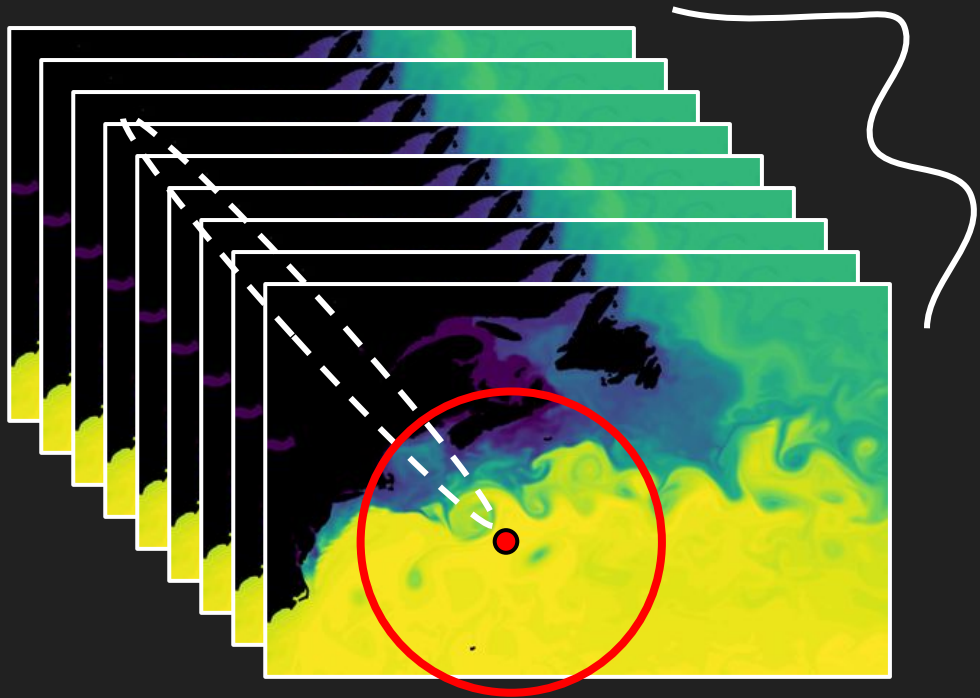








Look across the ensemble



Look across the ensemble

Localize

CESM-DART **Compset**

# CESM-DART **Compset**

MOM6 multi-instance

MOM6 interface for DART

Alper working on an CIME interface  
for DART

`./manage-externals`

Scripting (workflows!) for MOM6

Observations - wealth of existing  
obs on glade



Ensemble data assimilation in **CESM** is the *coolest* computational problem you can work on

# Join us

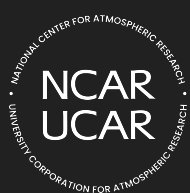
Collaborations

CISL Visitor Programs

Student Software Engineering - **Job Open**

Open Source

Challenging Problems: **Faster, Better, Stronger**



[dart@ucar.edu](mailto:dart@ucar.edu)  
[hkershaw@ucar.edu](mailto:hkershaw@ucar.edu)