Progress on CESM data assimilation and regional modeling

Dan Amrhein

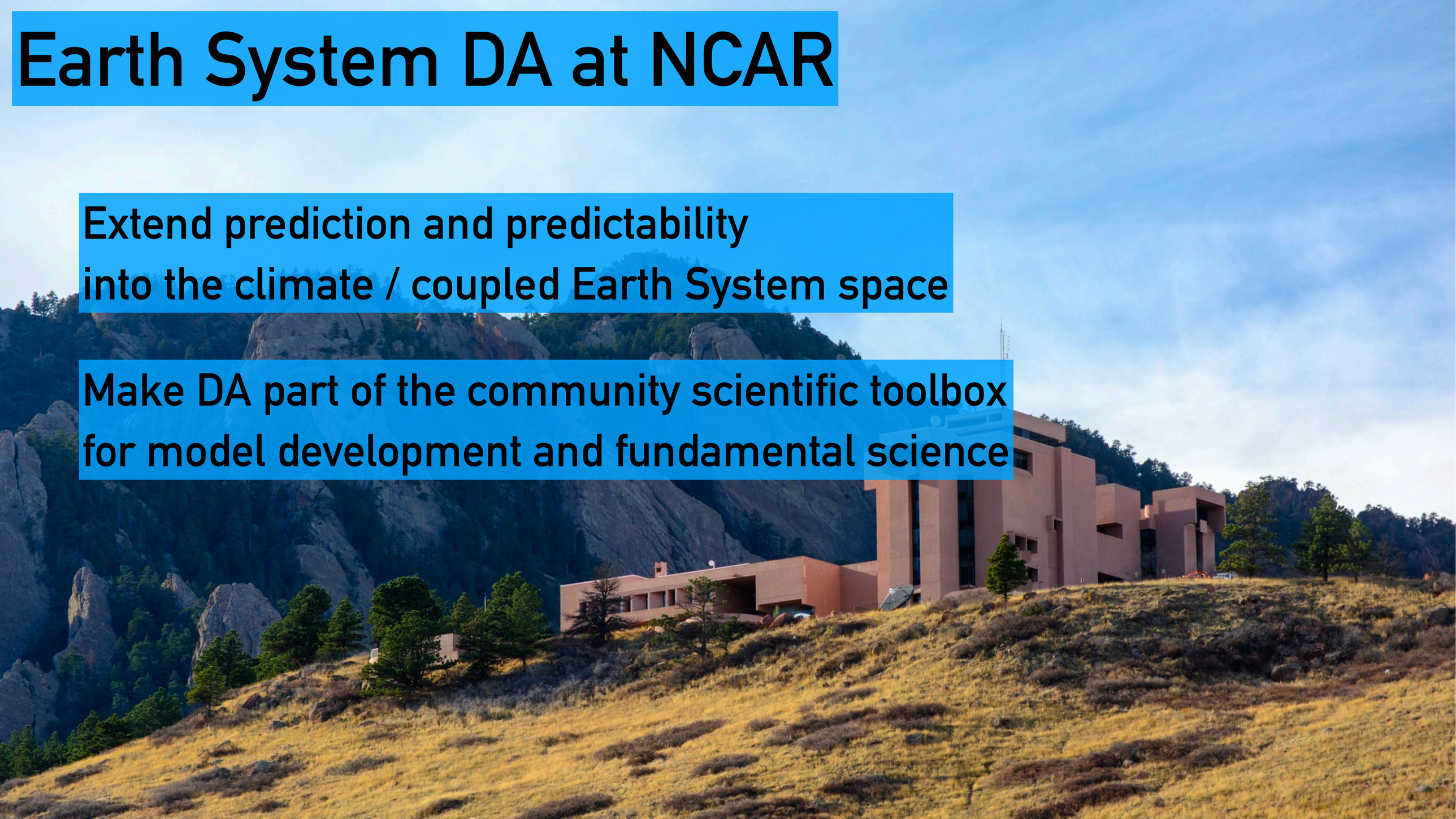
February 8th, 2024

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

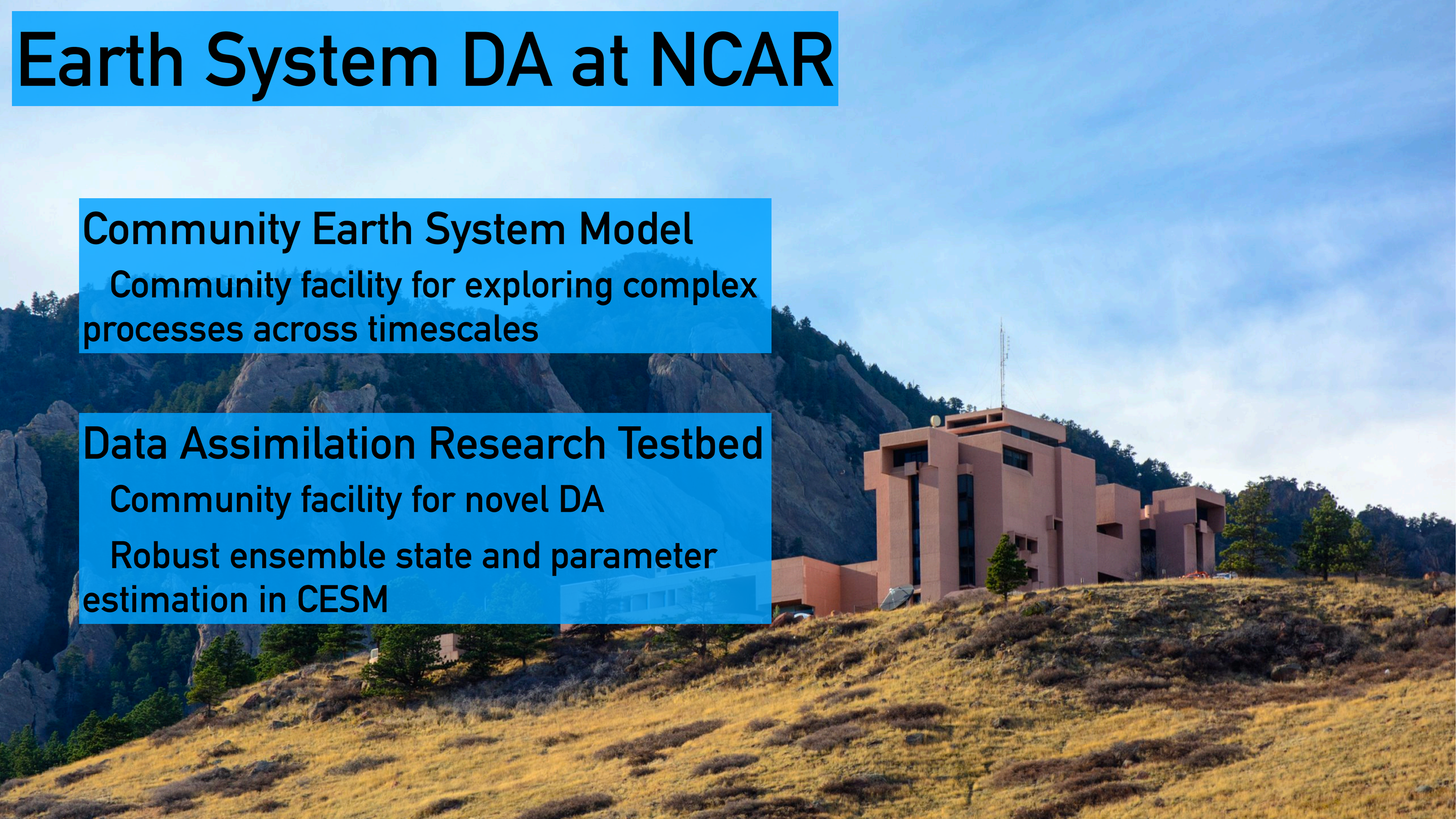
Earth System DA at NCAR

Extend prediction and predictability
into the climate / coupled Earth System space

Make DA part of the community scientific toolbox
for model development and fundamental science



Earth System DA at NCAR



Community Earth System Model

Community facility for exploring complex processes across timescales

Data Assimilation Research Testbed

Community facility for novel DA

Robust ensemble state and parameter estimation in CESM

Outline

NCAR base-funded efforts on CESM-DART in FY23

CROCODILE: Regional MOM6 with DA for BGC and climate impacts

Parameter estimation in MOM6-MARBL

Ocean DA in deep time paleoclimate

NCAR base-funded efforts on CESM-DART in FY23

(Helen Kershaw, Alper Altuntas)

Medium-range science goal: Capabilities for coupled DA science in CESM3.

1. *Multi-instance capabilities for MOM6 with NUOPC*
2. *DART-MOM6 integration and testing*
   - Advance DA capabilities for POP2 / CESM2 to MOM6 / CESM3 towards a CESM3 DA workhorse configuration enabling next-generation studies on initialized predictability, bias correction, carbon dioxide removal, and earth system state estimation.
   - Codependencies: None
3. *Testing MCT -> NUOPC*
   - Evaluate what needs to be changed in the CESM-DART interface for the NUOPC and make changes as necessary
   - Codependencies: None
4. *Building a test suite for DART DA components in CESM*
   - A set of minimal tests running DART across CESM components is important for maintaining DA capabilities as CESM evolves. It is a step towards making CESM-DART a community facility and extending usability beyond developers.
   - Codependencies: None
5. *Updating DART coupled DA code to CESM2 / Manhattan and extend beyond ocean / atmosphere*
   - Update the existing framework for coupled (cross-component) DA to include MOM6. Evaluate paths for other components (e.g., CLM) and perform initial tests.
   - Codependencies: 1., 2.
6. Gather existing DA capabilities across components into a **single workhorse compset for weakly coupled DA**. This work will include adding an Externals entry for DART and updating scripting (*months 1-3*), and performing basic tests of compset functionality (*month 4*).
7. Develop capabilities for hosting **existing pre-processed climate observations for ready ingestion into CESM.** A wealth of previously processed climate observations is currently hosted on glade; streamlining access to these removes a major DA roadblock for users. Altuntas will work with Amrhein to prioritize datasets (*month 1*) and implement hosting capabilities (*months 2-4*).

NCAR base-funded efforts on CESM-DART in FY23

(Helen Kershaw, Alper Altuntas)

This summary: Hiring two software engineer interns to work on a NUOPC cap for DART in CESM

Medium-range science goal: Capabilities for coupled DA science in CESM3.

1. *Multi-instance capabilities for MOM6 with NUOPC*
2. *DART-MOM6 integration and testing*
   - Advance DA capabilities for POP2 / CESM2 to MOM6 / CESM3 towards a CESM3 DA workhorse configuration enabling next-generation studies on initialized predictability, bias correction, carbon dioxide removal, and earth system state estimation.
   - Codependencies: None
3. *Testing MCT -> NUOPC*
   - Evaluate what needs to be changed in the CESM-DART interface for the NUOPC and make changes as necessary
   - Codependencies: None
4. *Building a test suite for DART DA components in CESM*
   - A set of minimal tests running DART across CESM components is important for maintaining DA capabilities as CESM evolves. It is a step towards making CESM-DART a community facility and extending usability beyond developers.
   - Codependencies: None
5. *Updating DART coupled DA code to CESM2 / Manhattan and extend beyond ocean / atmosphere*
   - Update the existing framework for coupled (cross-component) DA to include MOM6. Evaluate paths for other components (e.g., CLM) and perform initial tests.
   - Codependencies: 1., 2.
6. Gather existing DA capabilities across components into a **single workhorse compset for weakly coupled DA**. This work will include adding an Externals entry for DART and updating scripting (*months 1-3*), and performing basic tests of compset functionality (*month 4*).
7. Develop capabilities for hosting **existing pre-processed climate observations for ready ingestion into CESM.** A wealth of previously processed climate observations is currently hosted on glade; streamlining access to these removes a major DA roadblock for users. Altuntas will work with Amrhein to prioritize datasets (*month 1*) and implement hosting capabilities (*months 2-4*).

NCAR base-funded efforts on CESM-DART in FY23

(Helen Kershaw, Alper Altuntas)

Ahead of CESM3 release:

Testing MOM6 with cycling DART and any effects of increments on physics (Amrhein)

MOM6-enabled DA compset

Medium-range science goal: Capabilities for coupled DA science in CESM3.

1. *Multi-instance capabilities for MOM6 with NUOPC*
2. *DART-MOM6 integration and testing*
   - Advance DA capabilities for POP2 / CESM2 to MOM6 / CESM3 towards a CESM3 DA workhorse configuration enabling next-generation studies on initialized predictability, bias correction, carbon dioxide removal, and earth system state estimation.
   - Codependencies: None
3. *Testing MCT -> NUOPC*
   - Evaluate what needs to be changed in the CESM-DART interface for the NUOPC and make changes as necessary
   - Codependencies: None
4. *Building a test suite for DART DA components in CESM*
   - A set of minimal tests running DART across CESM components is important for maintaining DA capabilities as CESM evolves. It is a step towards making CESM-DART a community facility and extending usability beyond developers.
   - Codependencies: None
5. *Updating DART coupled DA code to CESM2 / Manhattan and extend beyond ocean / atmosphere*
   - Update the existing framework for coupled (cross-component) DA to include MOM6. Evaluate paths for other components (e.g., CLM) and perform initial tests.
   - Codependencies: 1., 2.
6. Gather existing DA capabilities across components into a **single workhorse compset for weakly coupled DA**. This work will include adding an Externals entry for DART and updating scripting (*months 1-3*), and performing basic tests of compset functionality (*month 4*).
7. Develop capabilities for hosting **existing pre-processed climate observations for ready ingestion into CESM.** A wealth of previously processed climate observations is currently hosted on glade; streamlining access to these removes a major DA roadblock for users. Altuntas will work with Amrhein to prioritize datasets (*month 1*) and implement hosting capabilities (*months 2-4*).

Advancing DA needs for critical NCAR capabilities

Earth System Predictability and Prediction (ESPP)

- Identify aspirational yet attainable science targets, noting readiness and specific gaps

Data assimilation, Infrastructure, and Modeling (AIM)

- Define unification and/or interoperability opportunities and create a roadmap to implement change

Software Engineering Across Labs (SEAL)

- Reimagine ongoing co-developed open-source community models in ways responsive to evolving software and hardware paradigms

Outline

NCAR base-funded efforts on CESM-DART in FY23

CROCODILE: Regional MOM6 with DA for BGC and climate impacts

Parameter estimation in MOM6-MARBL

Ocean DA in deep time paleoclimate

Dan Amrhein, Gustavo Marques,
Helen Kershaw, Keith Lindsay, Alper Altuntas,
Mike Levy, Matt Long,
Scott Bachman, Deepak Cherian

David (Roo) Nicholson, Susan Wijffels

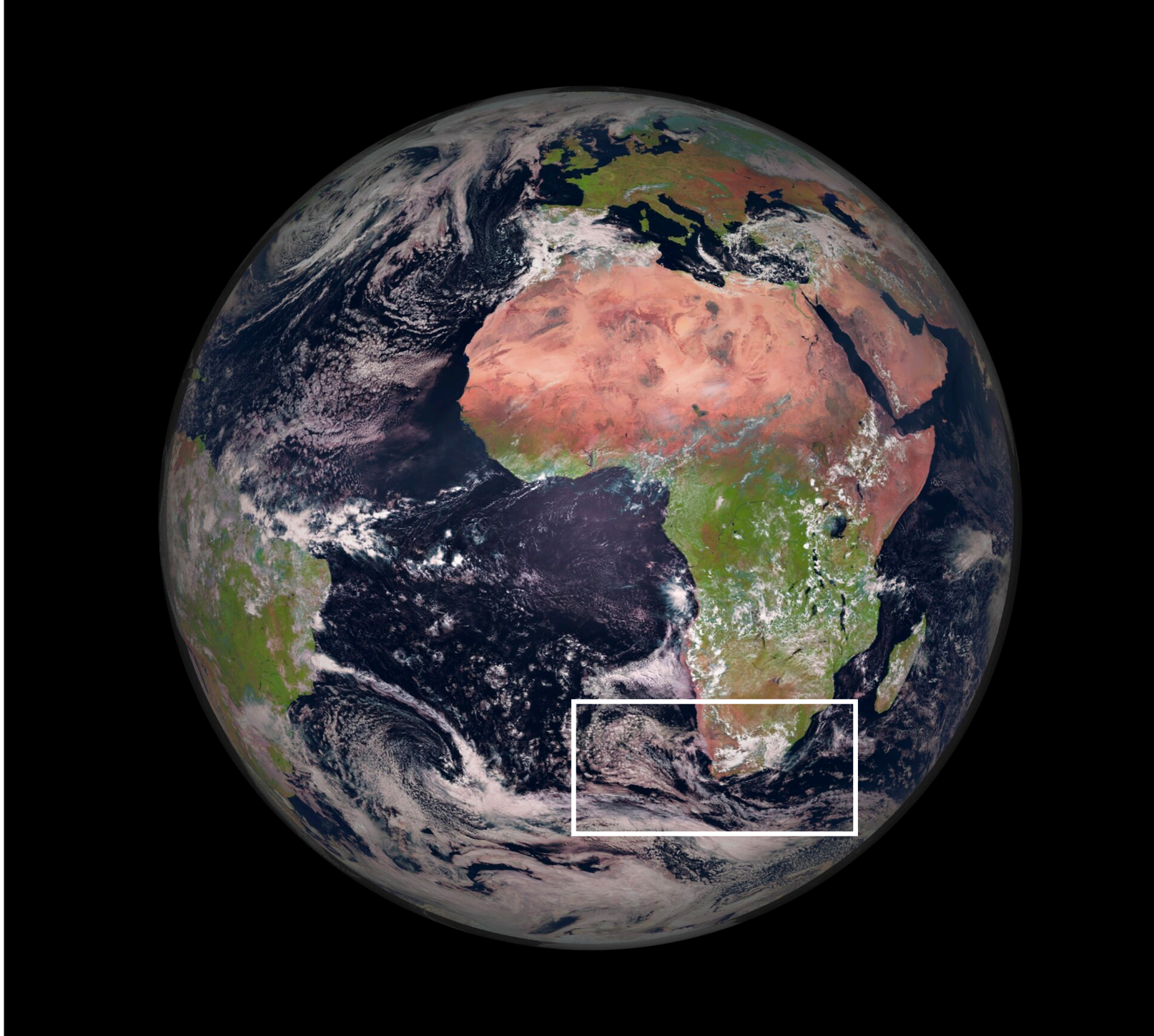
Motivation

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.



A workflow to broaden accessibility of regional climate modeling

regional_model_nbd.ipynb

Markdown

Python 3 (ipykernel)

Set up CESM / MOM6 / MARBL

```
[ ]: # Define bounds for a regional domain
     # Generate regional ocean grid
     # Extract lateral boundary conditions from (e.g.) a large ensemble simulation
```

Set up data assimilation

```
[ ]: # Define desired observations (including hypothetical obs for experimental design)
     # Download and stage observations for this region and time period
     # Define DART namelist options
```

Save this configuration for reproducibility

```
[ ]: # Define other xml options, etc.
     # Save experiment options
```

Run high-resolution regional CESM or CESM-DART and analyze output

```
[ ]: # Fire up DASK
     # Plot state-space metrics of interest (e.g., carbon uptake)
     # Plot observation-space metrics of interest using DART tools
```

Iterate to develop a scientifically useful model configuration!

A workflow to broaden accessibility of regional climate modeling

regional_model_nbd.ipynb

Markdown | Python 3 (ipykernel)

Set up CESM / MOM6 / MARBL

```
# Define bounds for a regional domain
# Generate regional ocean grid
# Extract lateral boundary conditions from (e.g.) a large ensemble simulation
```

Set up data assimilation

```
# Define desired observations (including hypothetical obs for experimental design)
# Download and stage observations for this region and time period
# Define DART namelist options
```

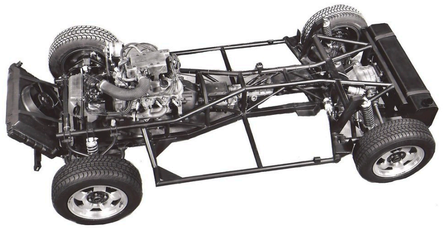
Save this configuration for reproducibility

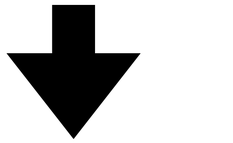
```
# Define other xml options, etc.
# Save experiment options
```

Run high-resolution regional CESM or CESM-DART and analyze output

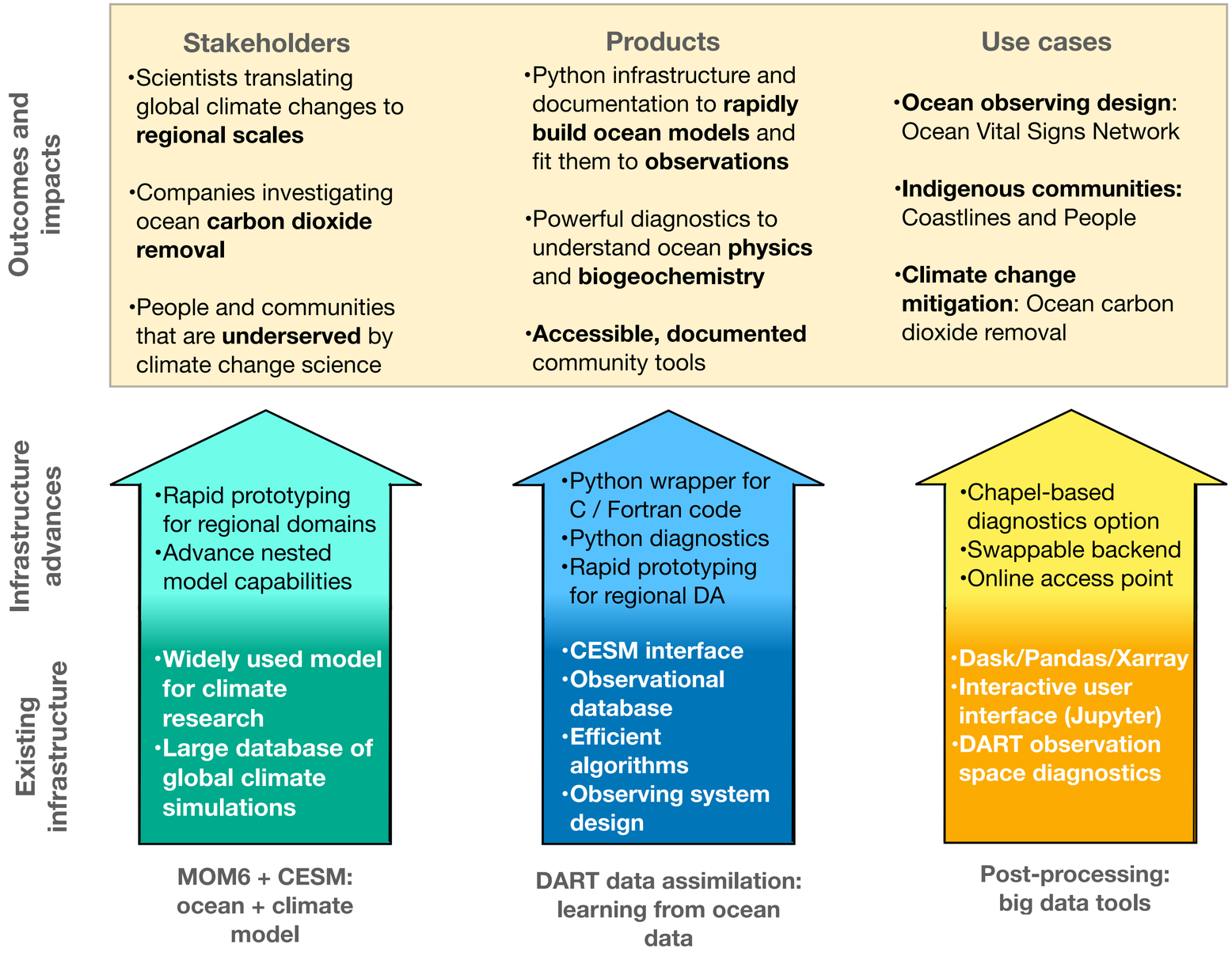
```
# Fire up DASK
# Plot state-space metrics of interest (e.g., carbon uptake)
# Plot observation-space metrics of interest using DART tools
```

Iterate to develop a scientifically useful model configuration!

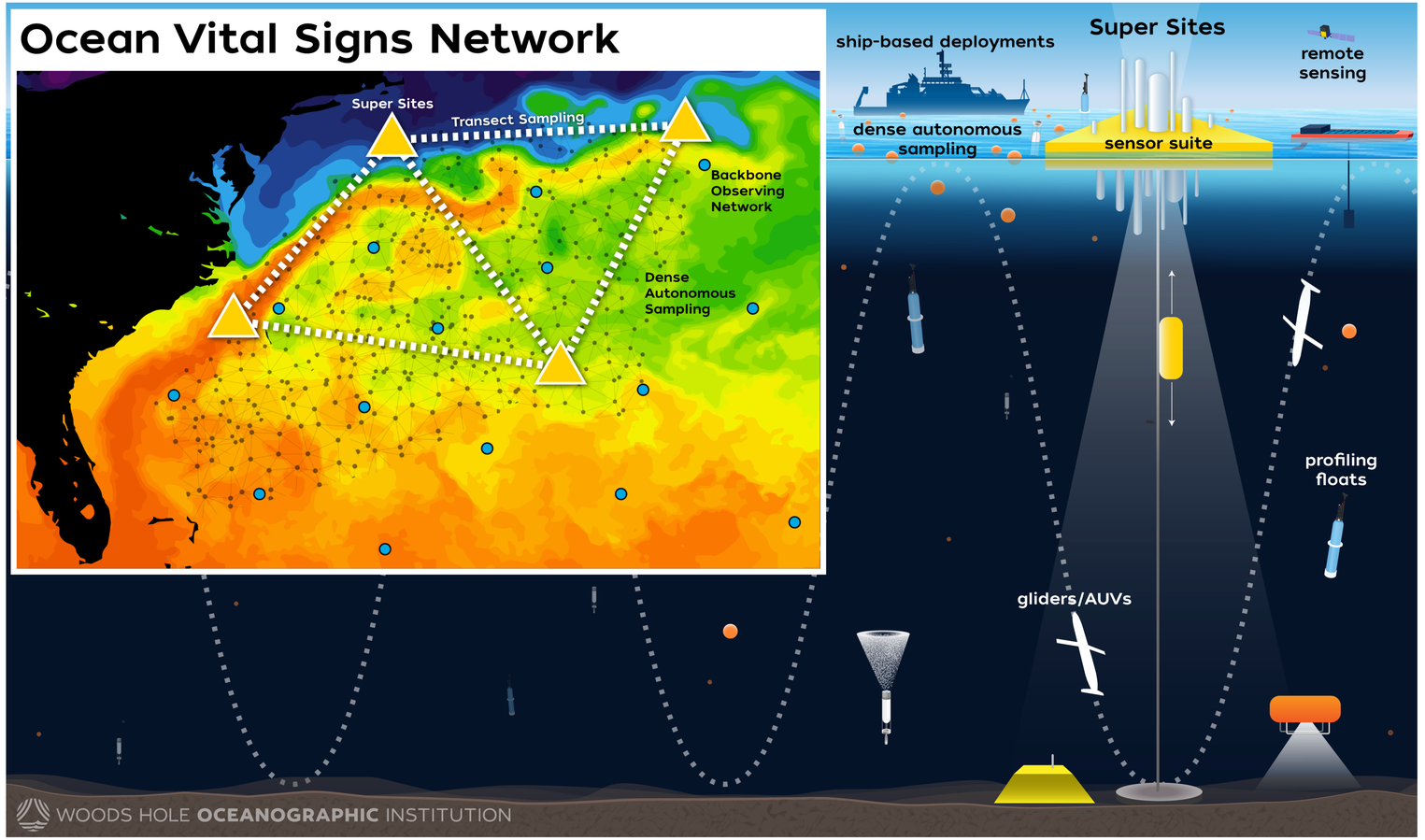




A community platform for accelerating observationally-constrained regional oceanographic modeling



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



Designing next-generation ocean observing systems

WOODS HOLE OCEANOGRAPHIC INSTITUTION

Use case 2: Co-development of regional modeling capabilities with Indigenous researchers and stakeholders

Making ocean change science more accessible to vulnerable populations and regions

Where We Work

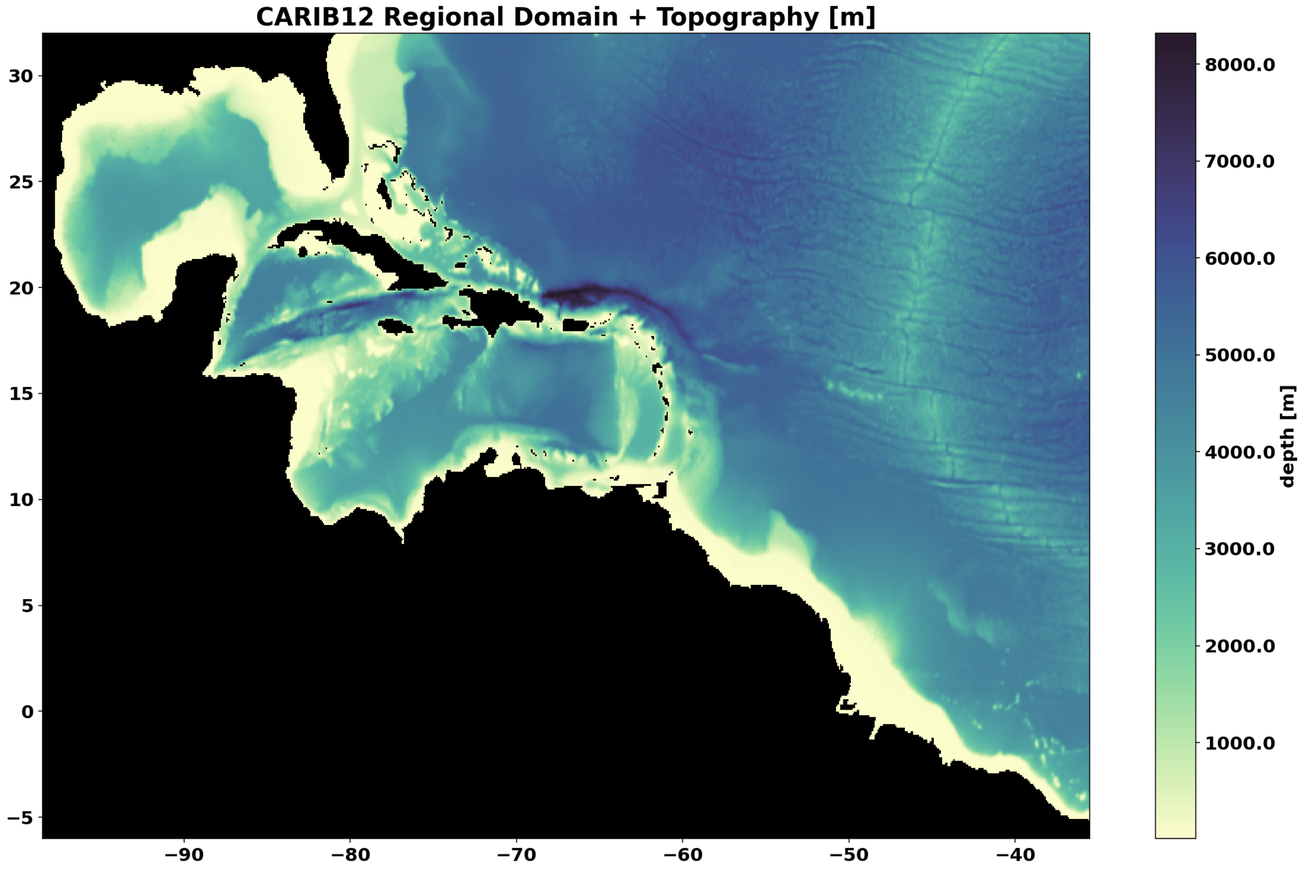
Led by the Haskell Foundation in Lawrence, KS, We are focused on impacts to four coastal regions in Indigenous territories

- Alaska (Arctic)
- Louisiana (Gulf of Mexico)
- Hawai'i (Pacific Islands)
- Puerto Rico (Caribbean Islands)

Use case 2: Co-development of regional modeling capabilities with Indigenous researchers and stakeholders

Making ocean change science more accessible to vulnerable populations and regions

(Welcome Giovanni Seijo as a postdoc on RVCC!)



Giovanni Seijo, CU Boulder

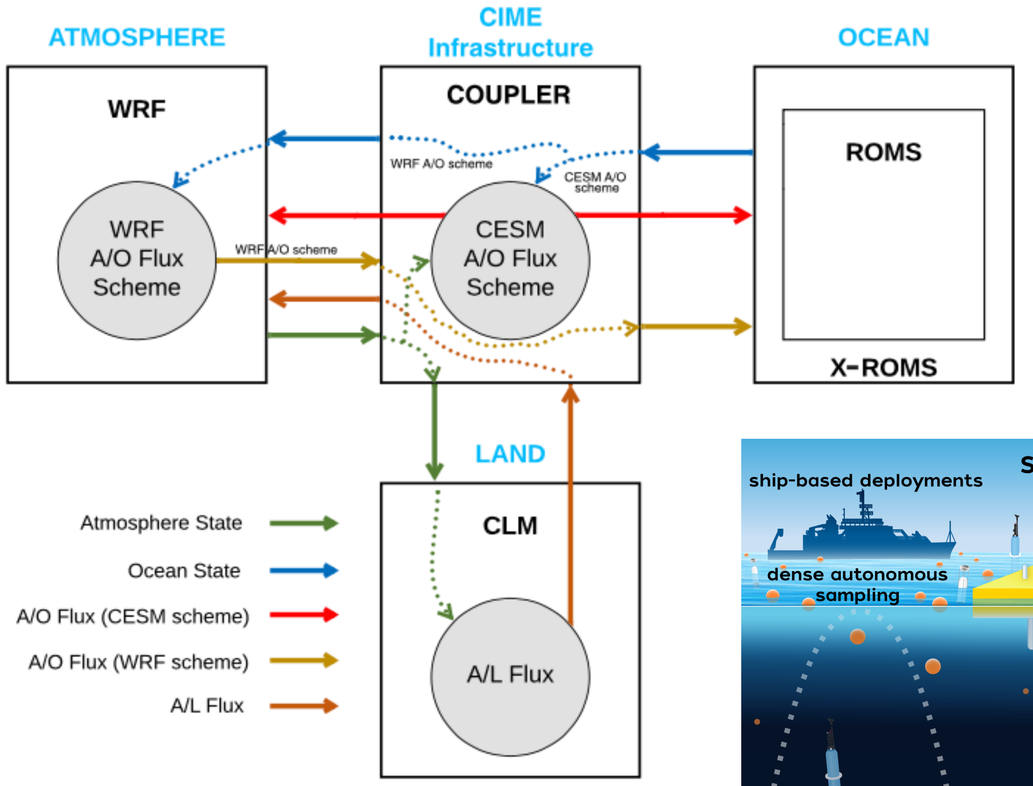
Year 1 work

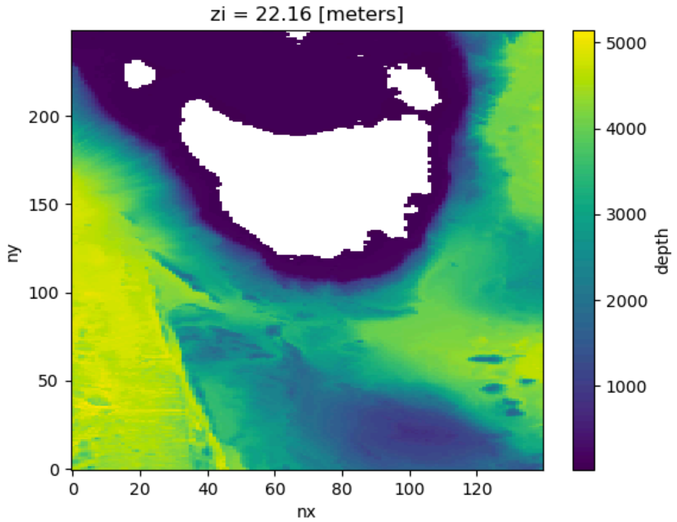
Develop regional coupling infrastructure to use with MOM6 and MARBL (new SE I hire at NCAR)

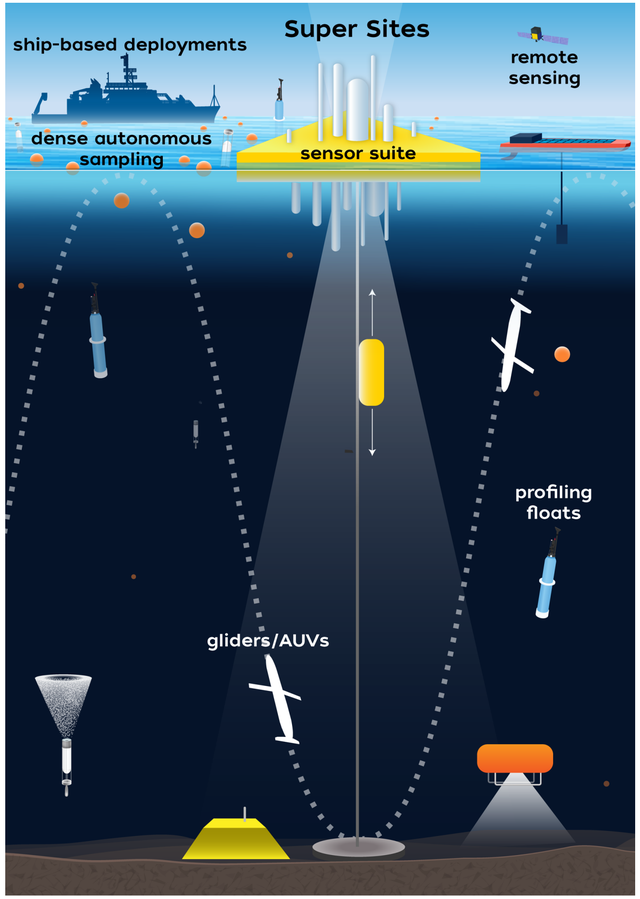
Automatic grid and ocean boundary condition generation (with Ashley Barnes, ANU)

Python wrapper for the Data Assimilation Research Testbed (DART; Helen Kershaw)

Implementation of biogeochemical "observation operators" at WHOI (new hire in progress)







Fu et al. 2020; COSIMA; WHOI

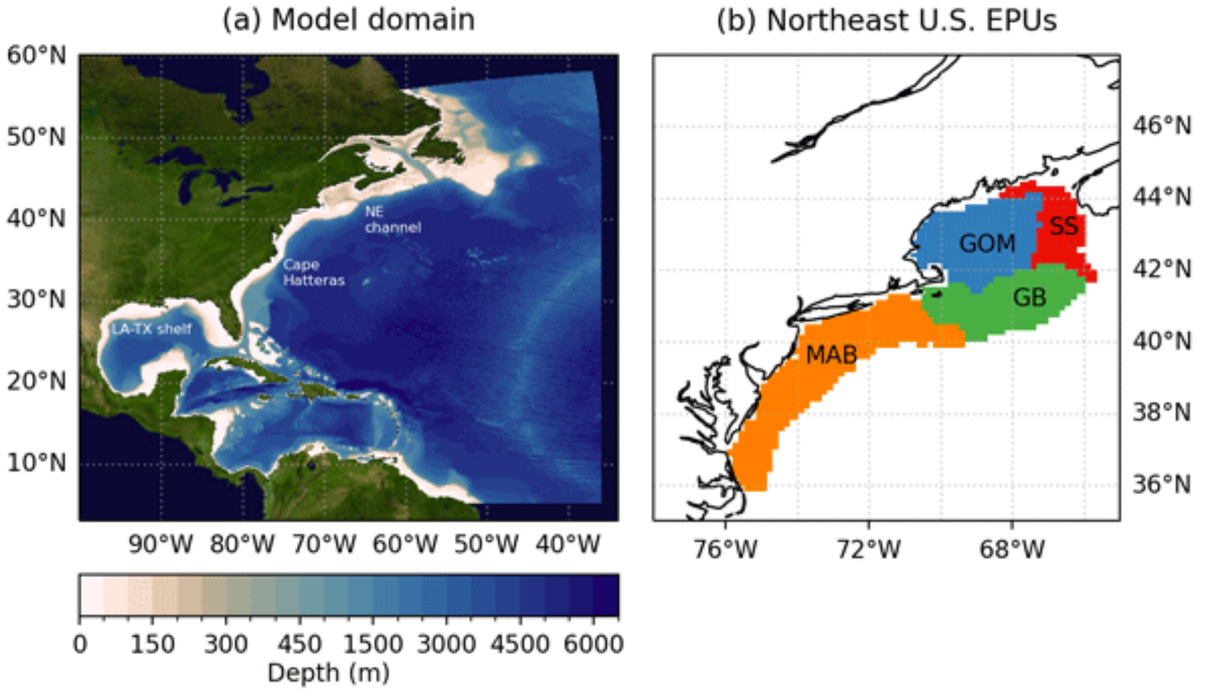
Develop regional coupling infrastructure to use with MOM6 and MARBL (new SE I hire at NCAR)

Automatic grid and ocean boundary condition generation (with Ashley Barnes, ANU)

Python wrapper for the Data Assimilation Research Testbed (DART; Helen Kershaw)

Implementation of biogeochemical "observation operators" at WHOI (new hire in progress)

Step 1: Run MOM6 NWA12 configuration on CESM 1/12 grid (Marques, Castruccio)



Ross et al. 2023

Outline

NCAR base-funded efforts on CESM-DART in FY23

CROCODILE: Regional MOM6 with DA for BGC and climate impacts

Parameter estimation in MOM6-MARBL

Ocean DA in deep time paleoclimate

Biogeochemical Parameter Estimation with MARBL-DART

Robin Armstrong (PhD Student, Cornell University, Center for Applied Mathematics)
Mentors: Moha Gharamti, Dan Amrhein, Kristen Krumhardt,
Mike Levy, Helen Kershaw, Keith Lindsay

December 12th, 2023

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

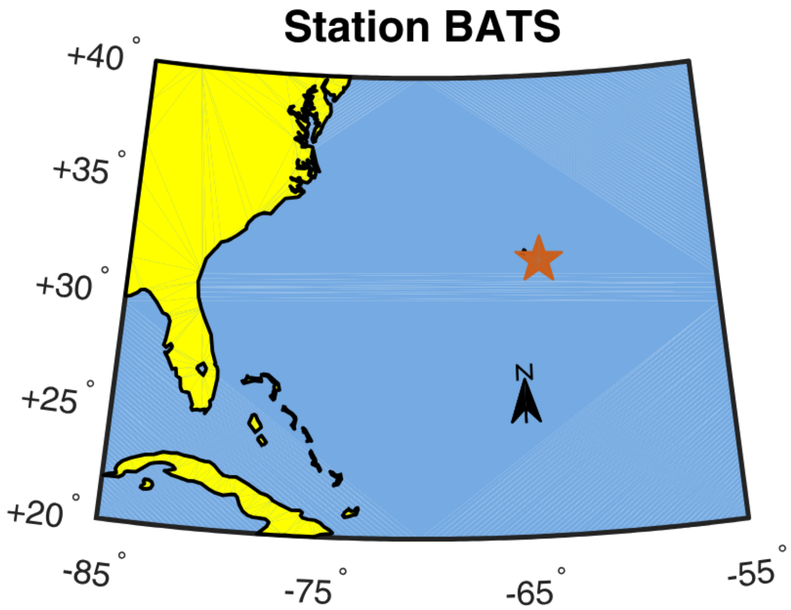
Our Objective:
To **optimize the parameters of MARBL** by assimilating data with **DART** (the Data Assimilation Research Testbed).

For more information see:

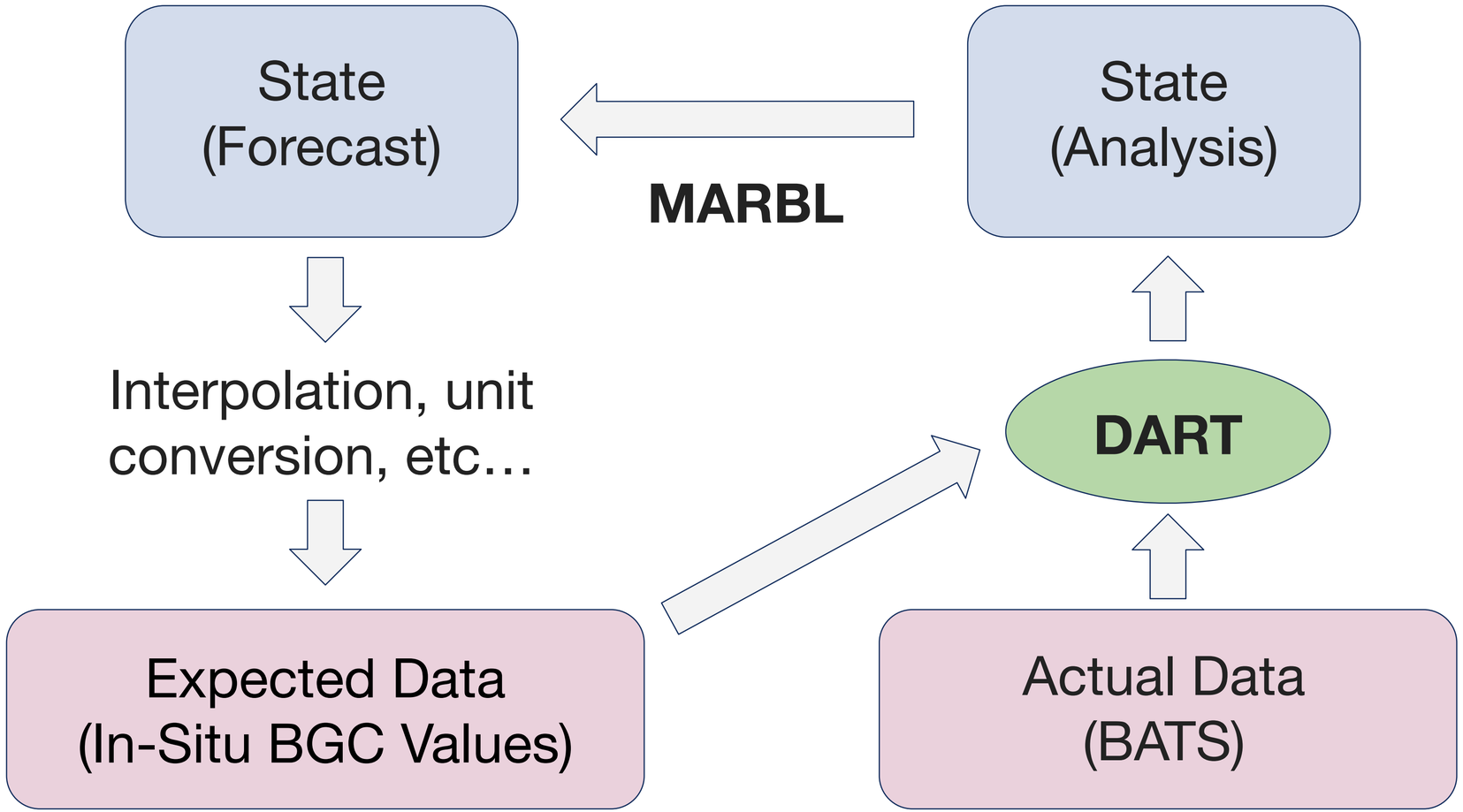
- Anderson et al, “The Data Assimilation Research Testbed: A Community Facility,” *American Meteorological Society* (2009).
- Long et al, “Simulations With the Marine Biogeochemistry Library,” *Journal of Advances in Modeling Earth Systems* (2021).

Data Preparation

- We use data from the Bermuda Atlantic Time-Series Study (**BATS**).
- More than 30 years of **in-situ biogeochemical data** from a single ocean column.
- Correspondence between data variables and model variables is nontrivial. **Mappings from model space to data space needed to be carefully determined**.
- **Time sparsity** is a challenge which calls for assimilation from multiple sources.

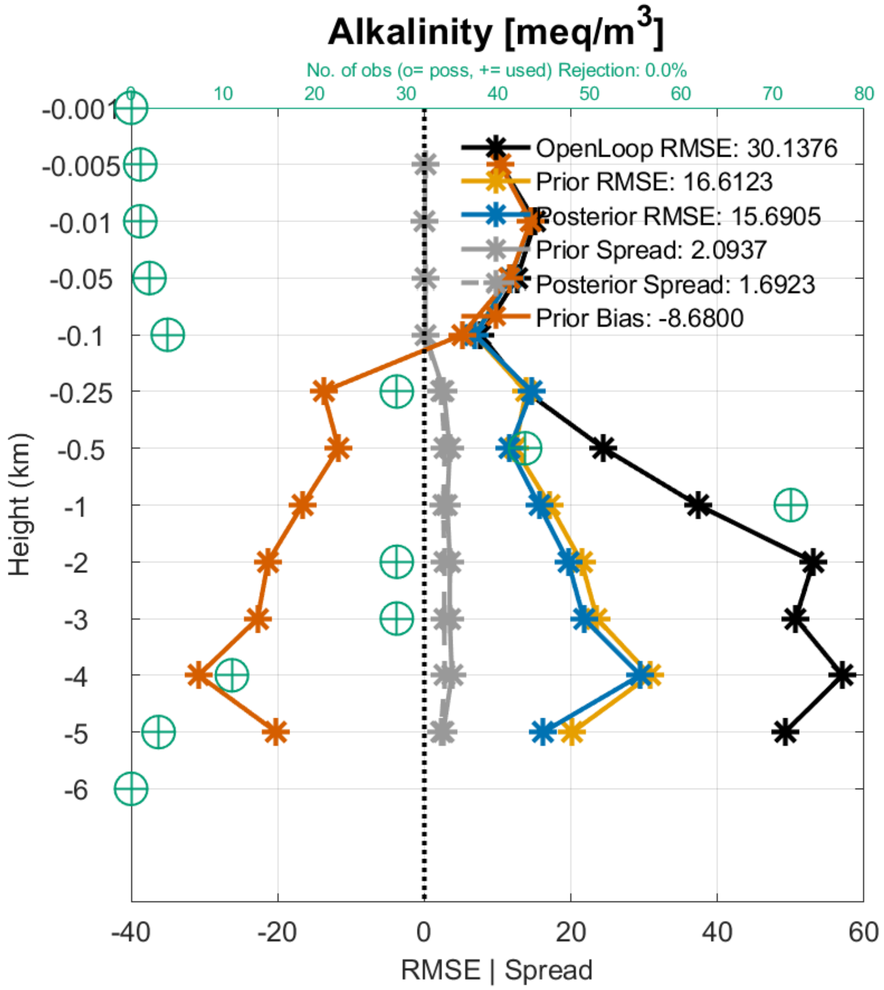


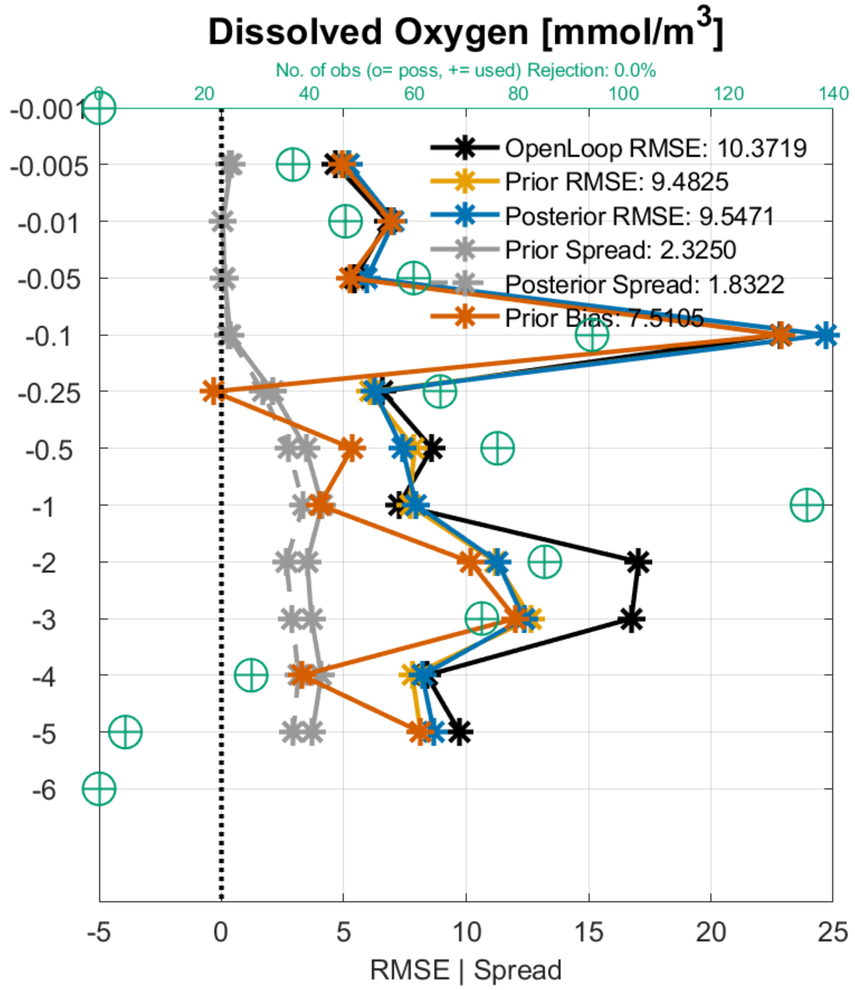
MARBL-DART, State-Estimation-Only Version (SIParCS 2023)



MARBL-DART, State-Estimation-Only Version (SIParCS 2023)

Assimilation Period: 24-Feb-2005 : 22-Feb-2006





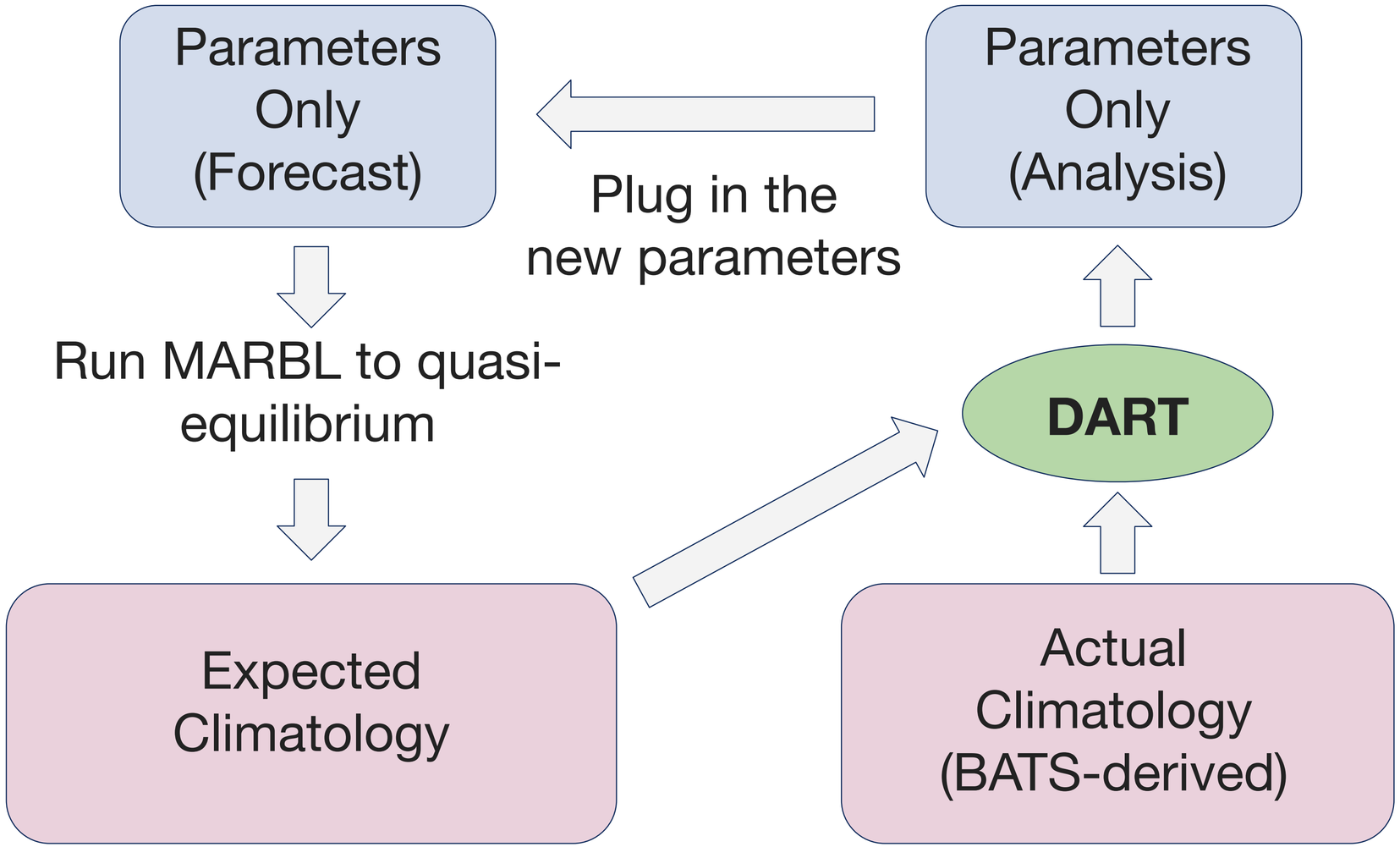
Looking Ahead

- MARBL-DART is currently a **state/parameter estimation system.**
  - Model time series is compared to a data time series.
  - DART adjusts the MARBL trajectory, and **simultaneously tunes MARBL parameters**.

- Advantage: potentially obtaining a better fit to data.
- Disadvantage: **time-varying parameters, with the danger of overfitting**.

- Our end goal: **MARBL-DART with parameters only**.
  - Parameters are thought of as time-constant.
  - Model climatology is compared to data climatology.
  - **DART fits a climatology rather than a time series**.

Parameter Only Estimation



Outline

NCAR base-funded efforts on CESM-DART in FY23

CROCODILE: Regional MOM6 with DA for BGC and climate impacts

Parameter estimation in MOM6-MARBL

Ocean DA in deep time paleoclimate

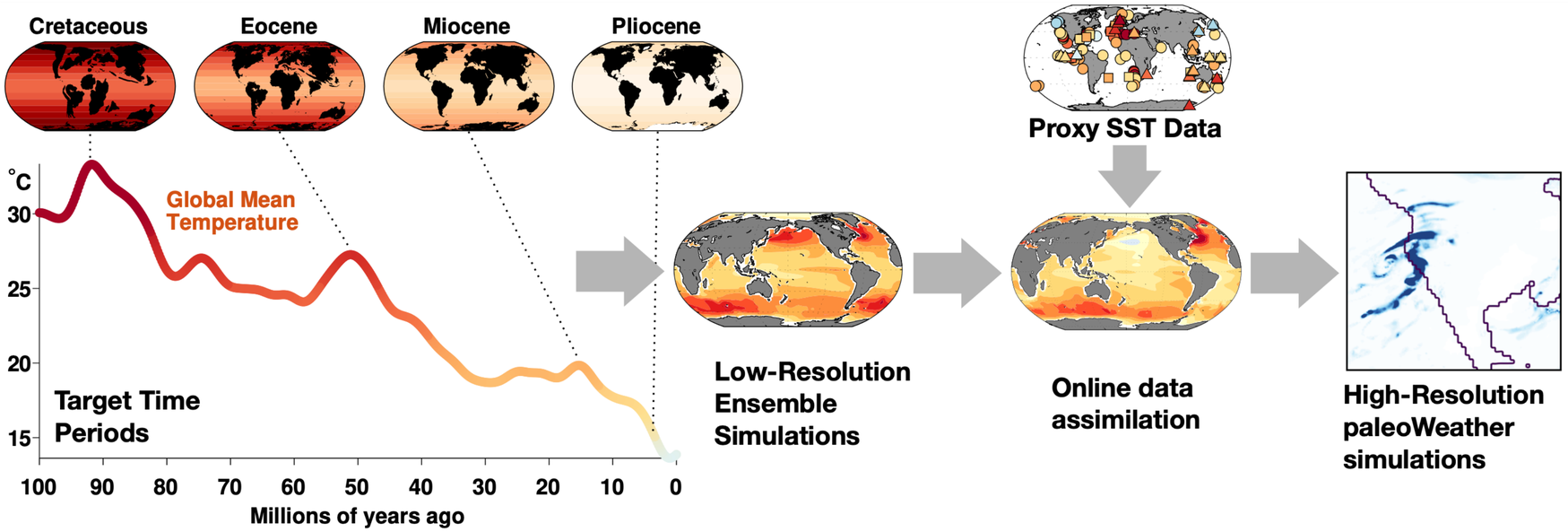
paleoWeather:
Hazards and habitability in hothouse climates

Jessica E. Tierney (UA), Bette L. Otto-Bliesner, Jiang Zhu, Sophia Macarewich, Dan Amrhein, Jane W. Baldwin (UCI), and Christopher J. Poulsen (UofO)

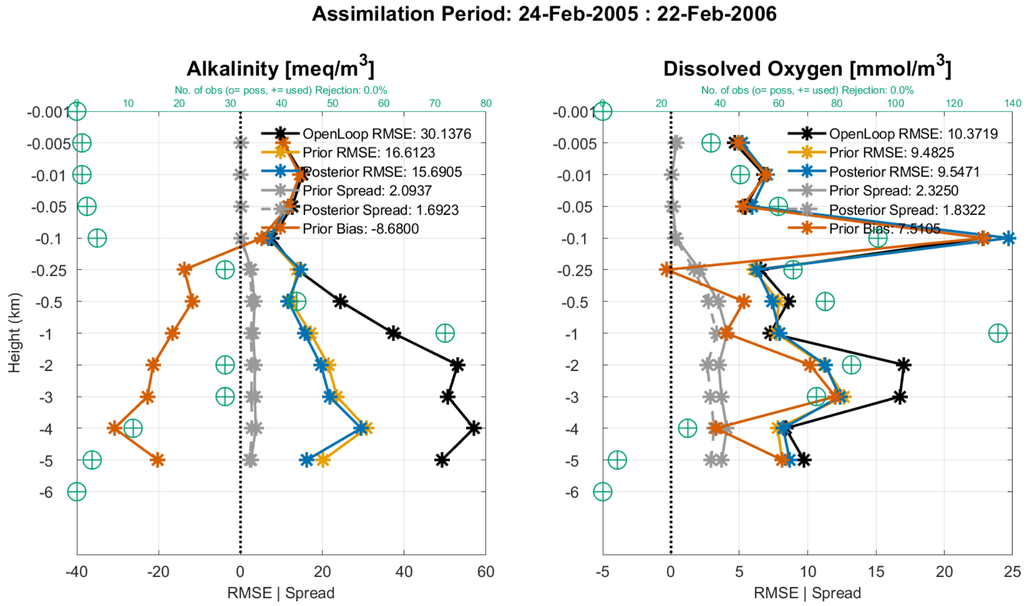
HEISING-SIMONS FOUNDATION

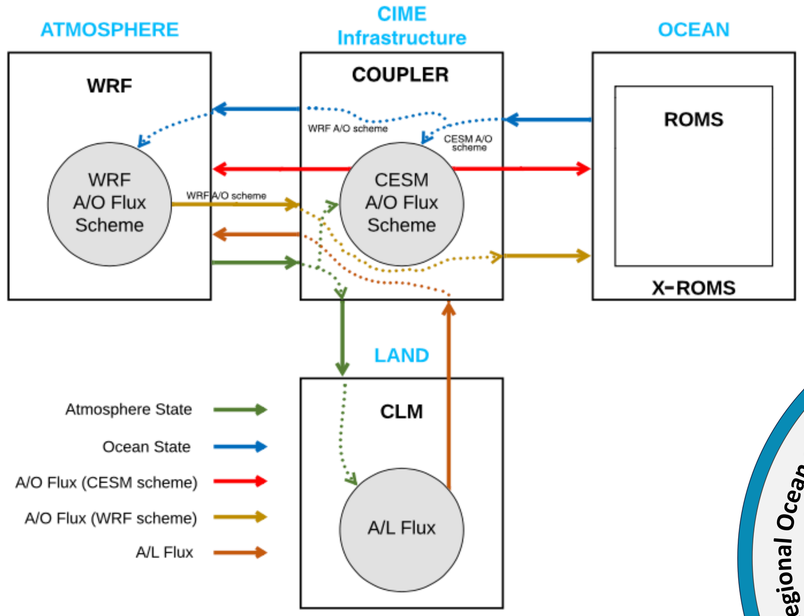
paleoWeather: Hazards and habitability in hothouse climates

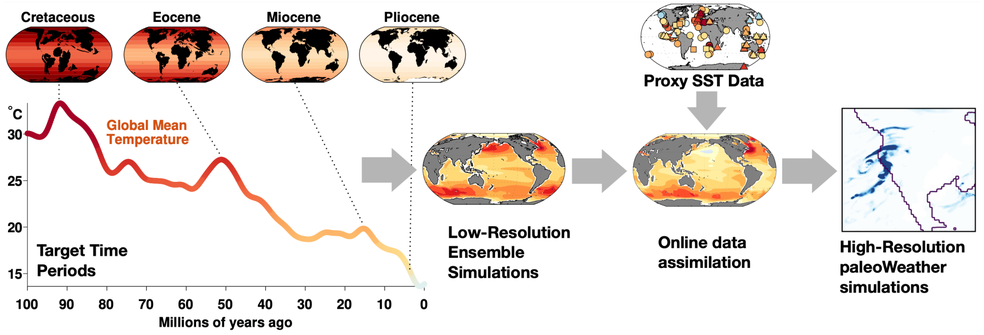
Jessica E. Tierney (UA), Bette L. Otto-Bliesner, Jiang Zhu, Sophia Macarewich, Dan Amrhein, Jane W. Baldwin (UCI), and Christopher J. Poulsen (UofO)



Dan Amrhein — damrhein@ucar.edu







Thanks!

damrhein@ucar.edu