

MARBL-DART: An Ensemble System for Biogeochemical Data Assimilation

Project 8, SIParCS 2023

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



August 1st, 2023

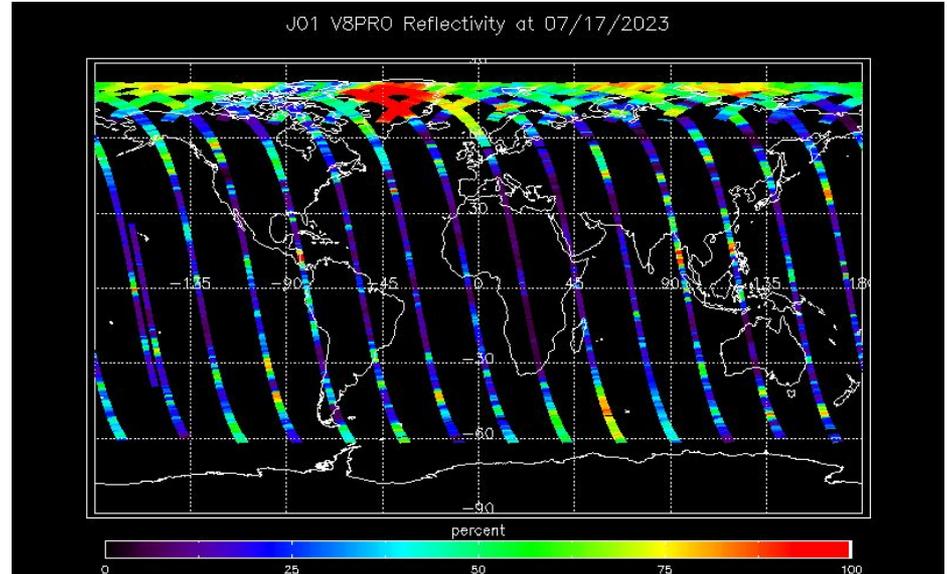


Part 1: Introduction



WHAT IS DATA ASSIMILATION?

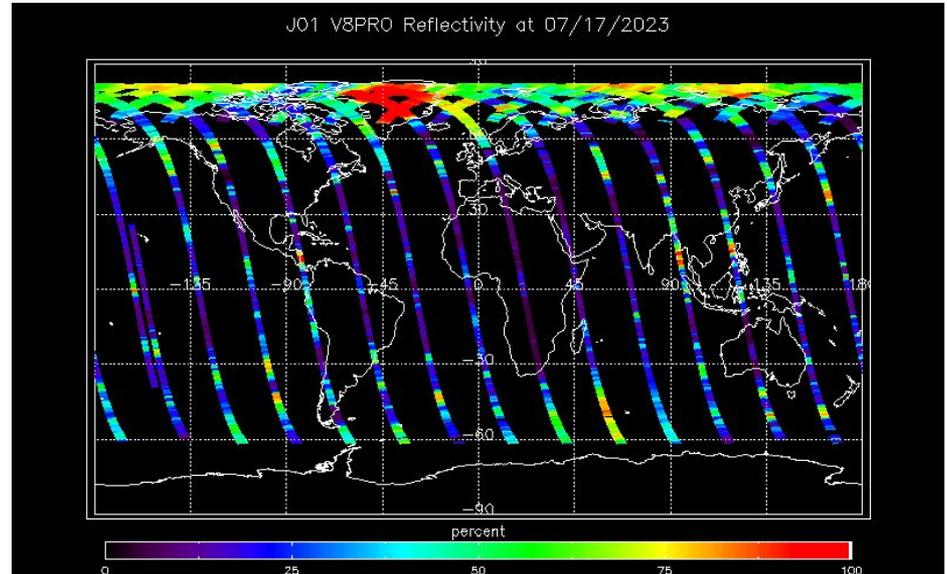
- Data assimilation is about combining **computer models** with **empirical data**.



Ozone data from the NOAA-20 satellite, taken on July 17th, 2023. Source: https://www.ospo.noaa.gov/data/atmosphere/ozone/Products_atmosphere_OMPS.html.

WHAT IS DATA ASSIMILATION?

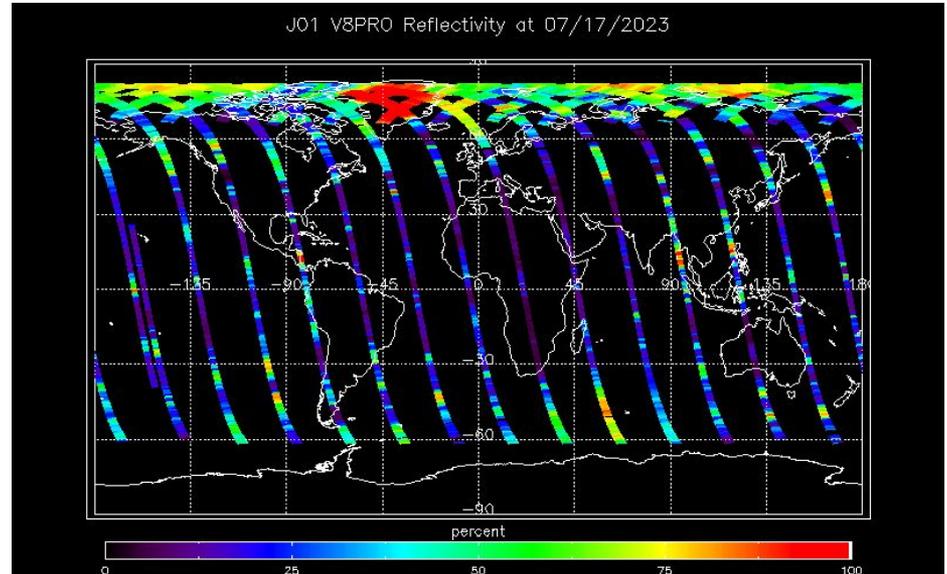
- Data assimilation is about combining **computer models** with **empirical data**.
- Computer models will usually be **biased**; they can't capture the full complexity of the Earth system!



Ozone data from the NOAA-20 satellite, taken on July 17th, 2023. Source: https://www.ospo.noaa.gov/data/atmosphere/ozone/Products_atmosphere_OMPS.html.

WHAT IS DATA ASSIMILATION?

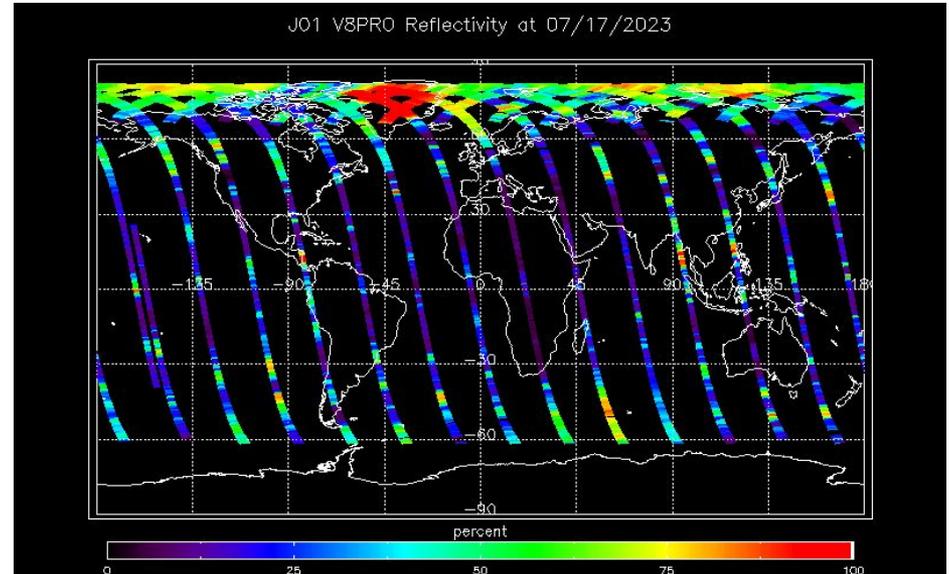
- Data assimilation is about combining **computer models** with **empirical data**.
- Computer models will usually be **biased**; they can't capture the full complexity of the Earth system!
- Data, on the other hand, comes with **measurement errors** and may be **sparse**.



Ozone data from the NOAA-20 satellite, taken on July 17th, 2023. Source: https://www.ospo.noaa.gov/data/atmosphere/ozone/Products_atmosphere_OMPS.html.

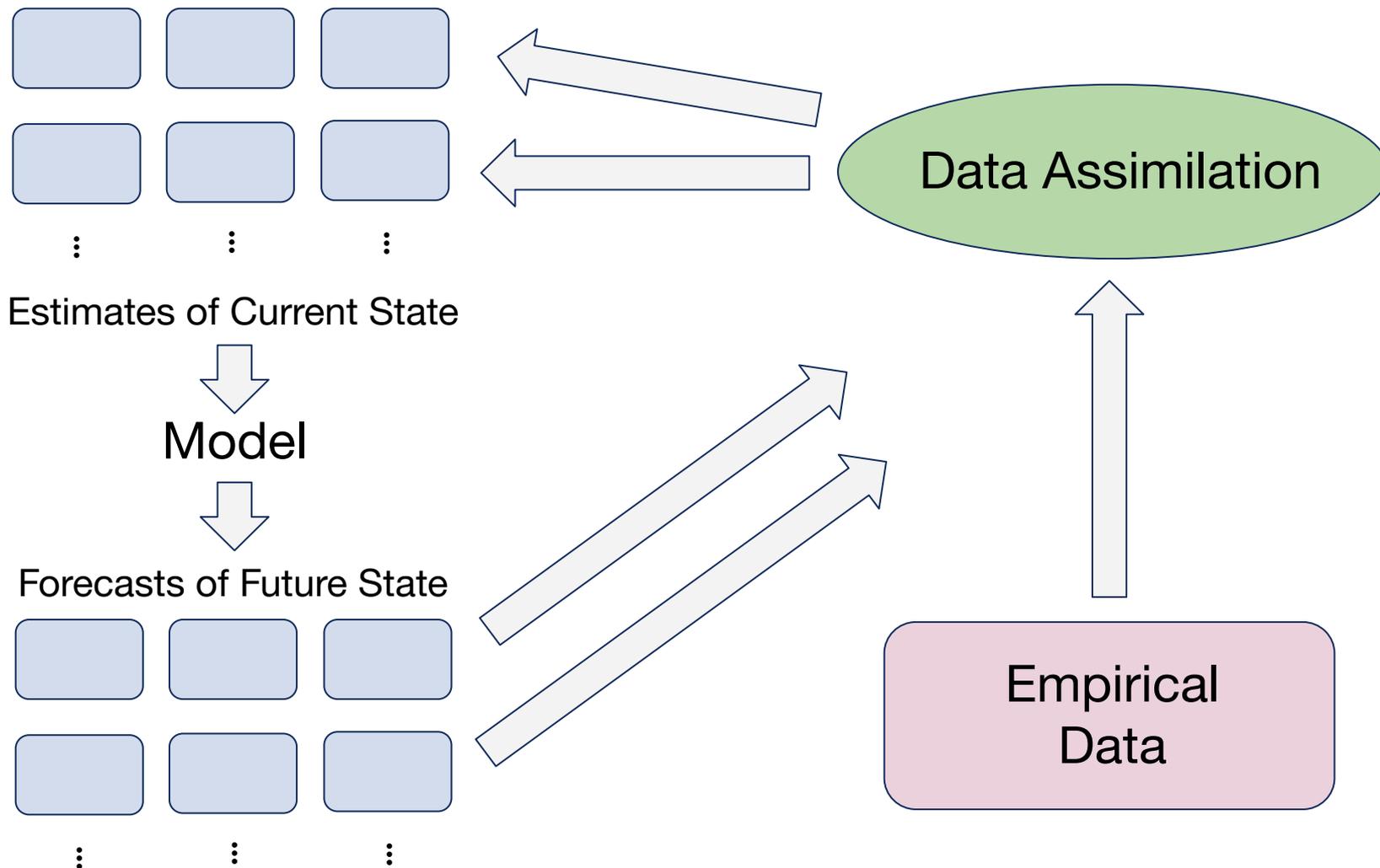
WHAT IS DATA ASSIMILATION?

- Data assimilation is about combining **computer models** with **empirical data**.
- Computer models will usually be **biased**; they can't capture the full complexity of the Earth system!
- Data, on the other hand, comes with **measurement errors** and may be **sparse**.
- **Data assimilation** balances the **uncertainty** in the data against the **uncertainty** in the model, finding a middle ground which is closer to the truth.

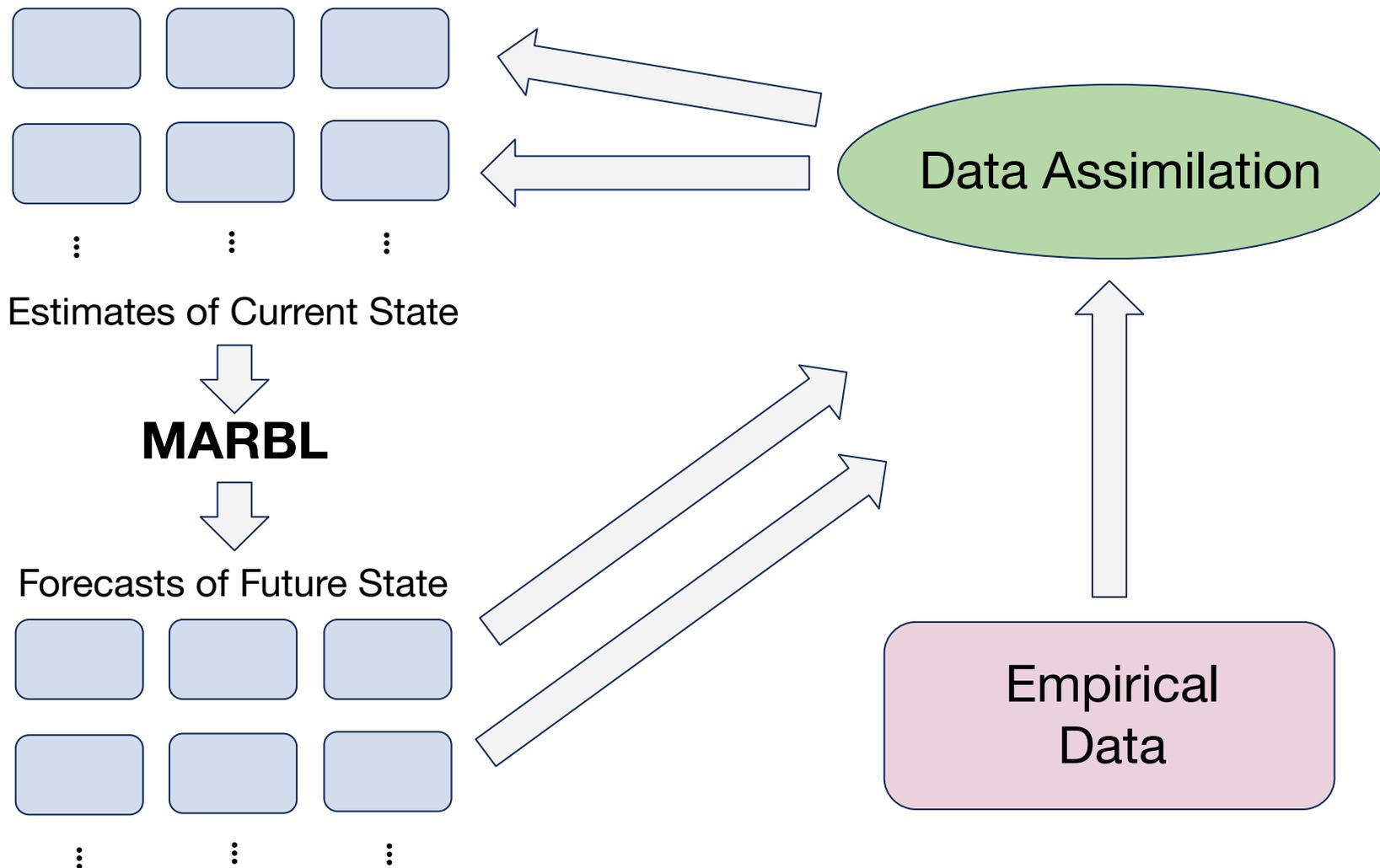


Ozone data from the NOAA-20 satellite, taken on July 17th, 2023. Source: https://www.ospo.noaa.gov/data/atmosphere/ozone/Products_atmosphere_OMPS.html.

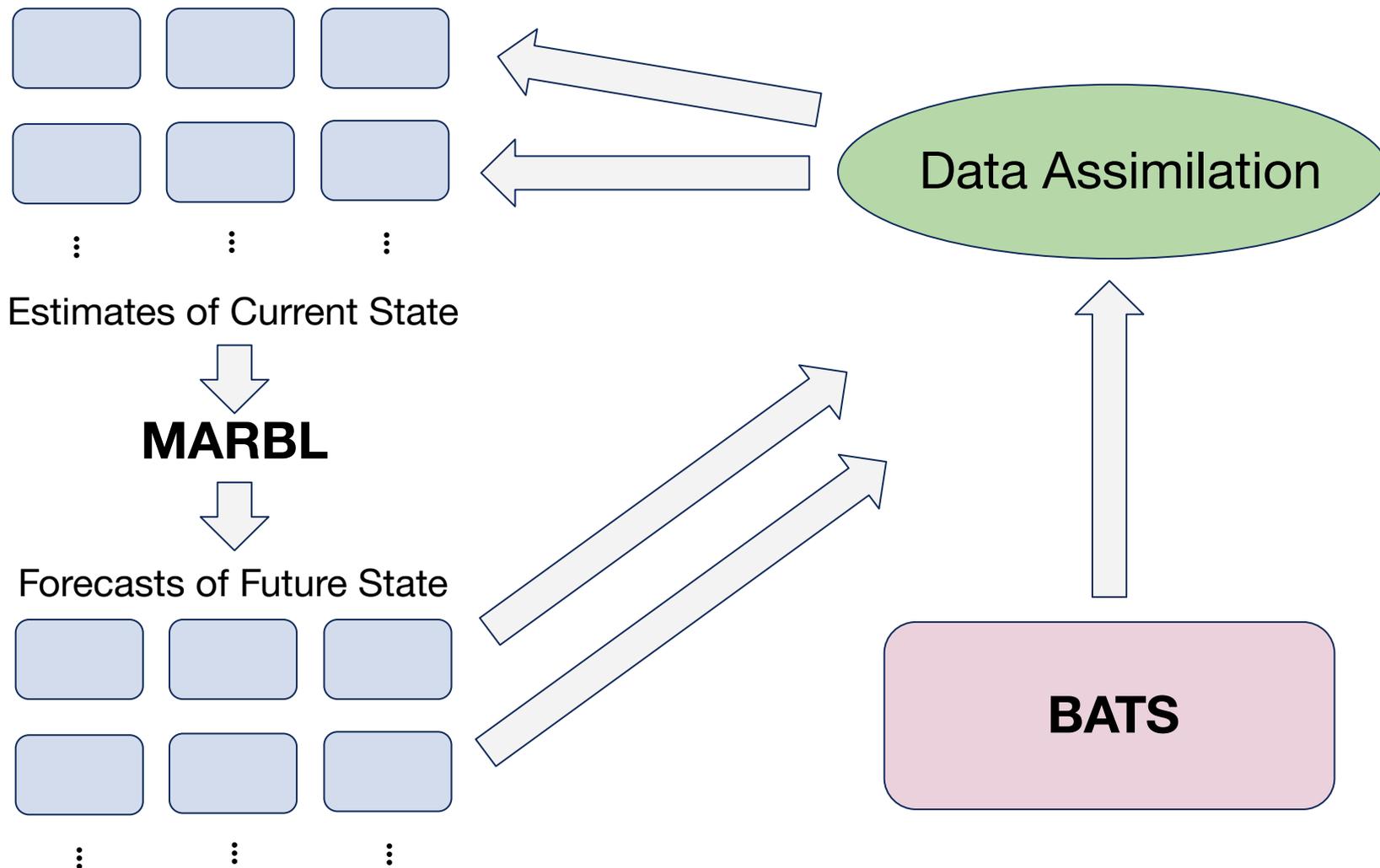
ENSEMBLE DATA ASSIMILATION



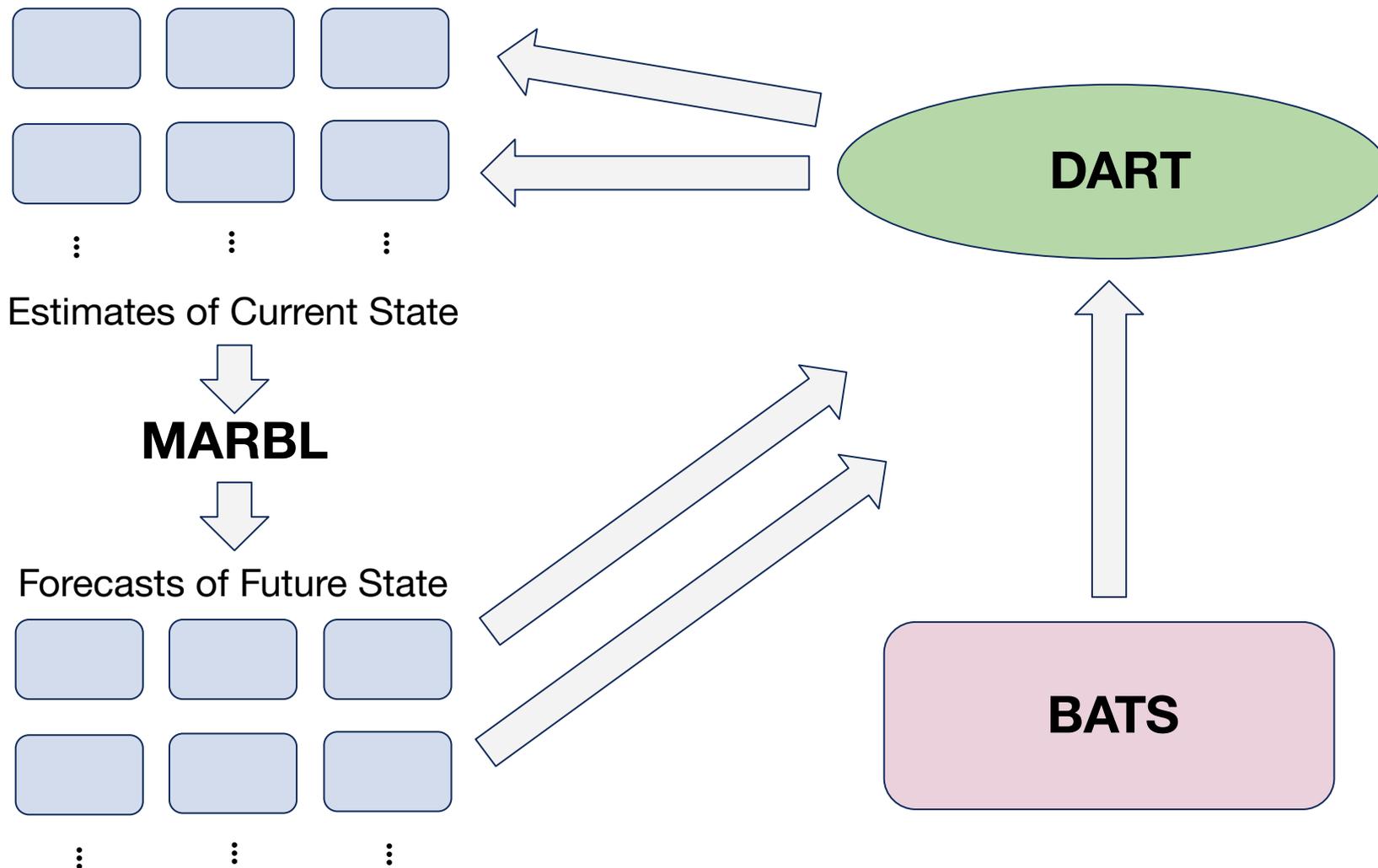
ENSEMBLE DATA ASSIMILATION



ENSEMBLE DATA ASSIMILATION



ENSEMBLE DATA ASSIMILATION



PROJECT OBJECTIVE

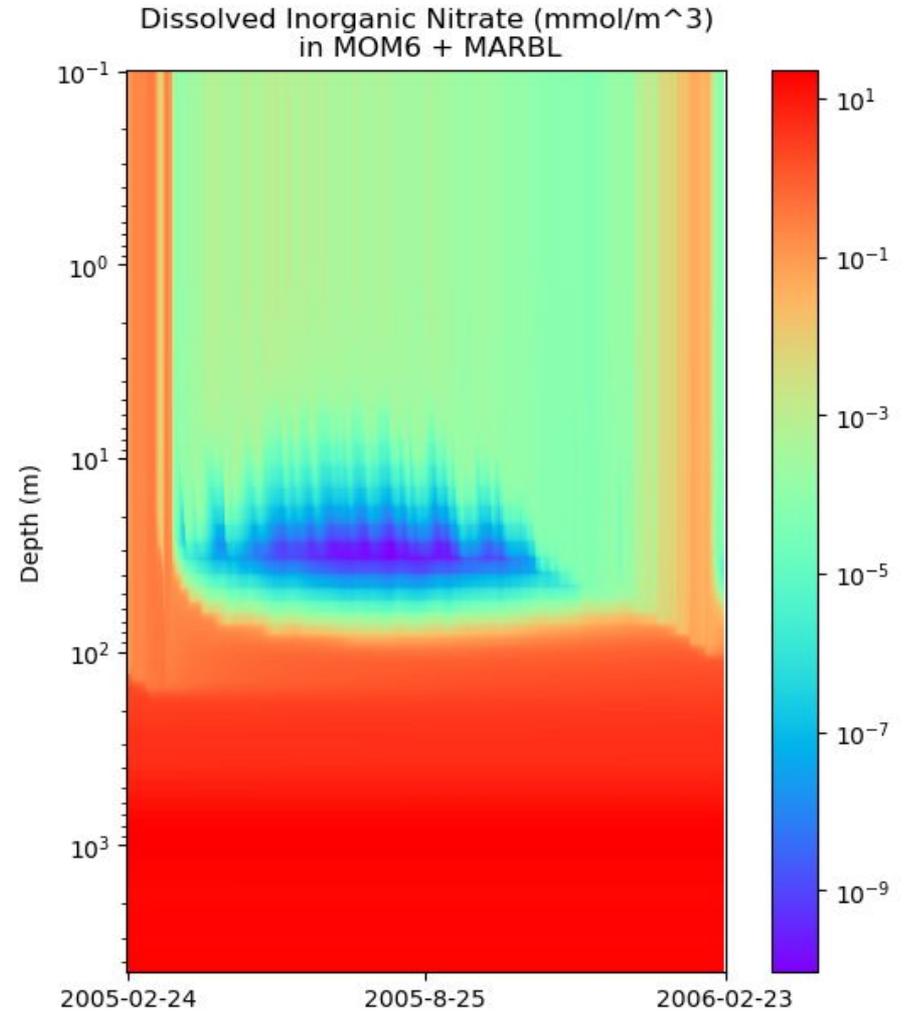
Project Objective: we build an **ensemble data assimilation system** for forecasting **ocean biogeochemistry** in a single-column ocean model. Our system allows forecasts of the Marine Biogeochemistry Library (**MARBL**) to be combined with real world measurements using the Data Assimilation Research Testbed (**DART**).

Part 2: Project Components



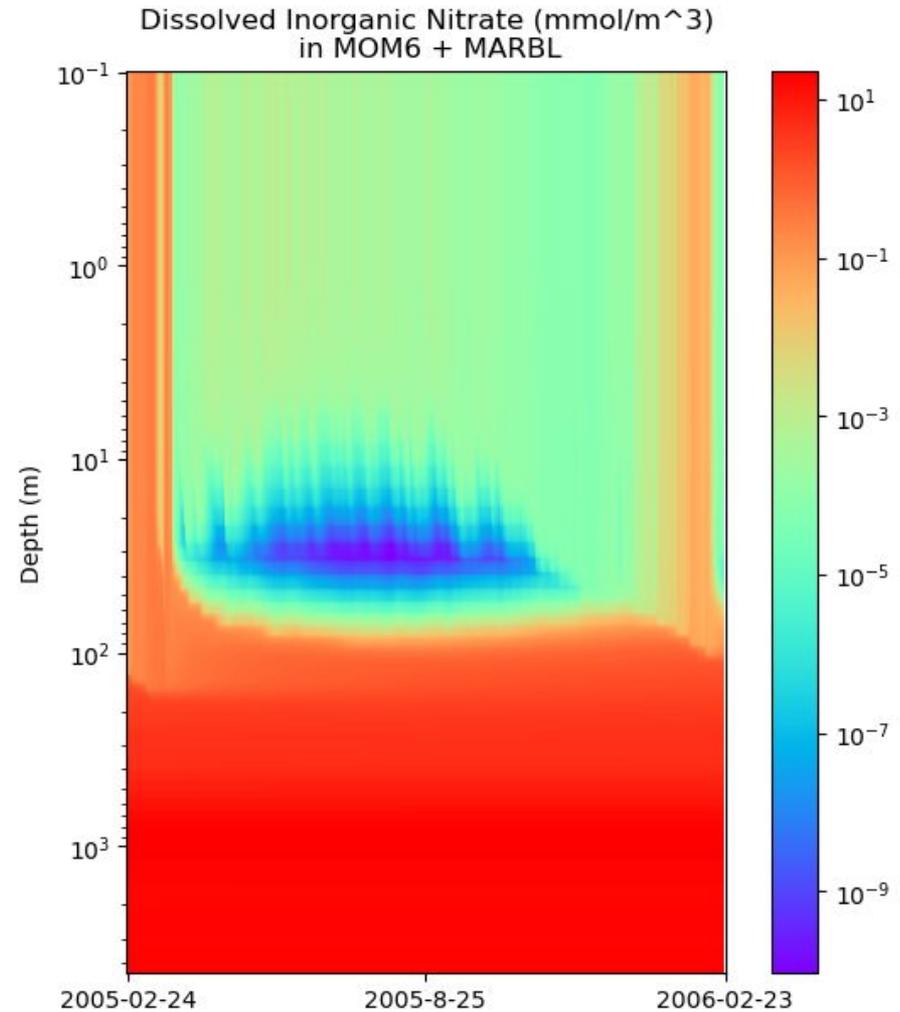
MOM6

- MOM6 is the **Modular Ocean Model**.



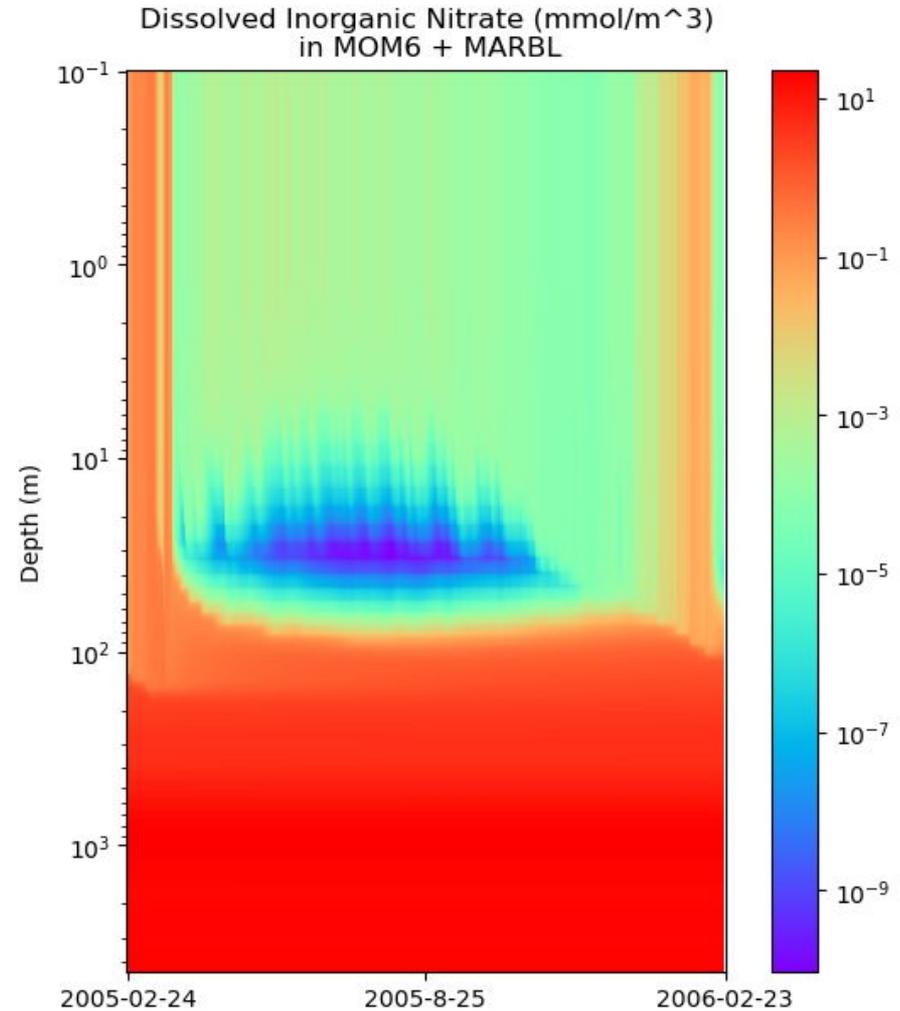
MOM6

- MOM6 is the **Modular Ocean Model**.
- Able to run with CESM, but we're running it **standalone**.



MOM6

- MOM6 is the **Modular Ocean Model**.
- Able to run with CESM, but we're running it **standalone**.
- Simulating only a **single ocean column** (BATS station).

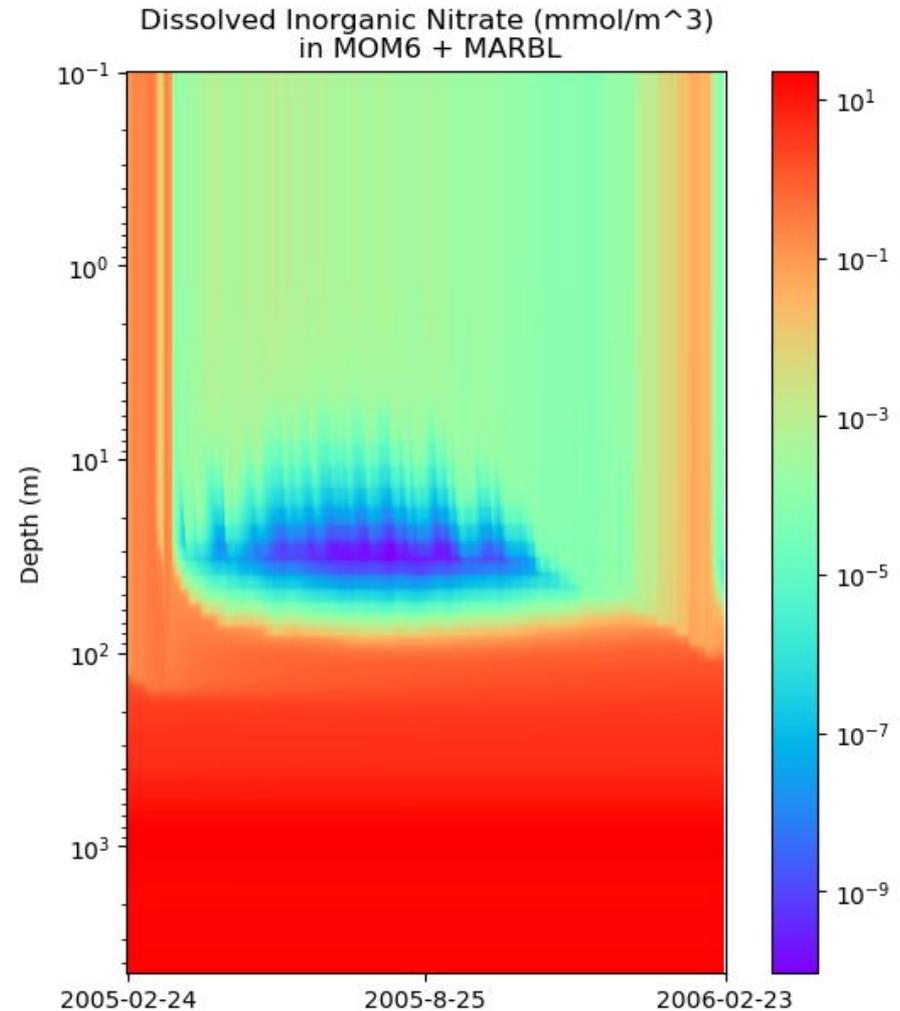


MOM6

- MOM6 is the **Modular Ocean Model**.
- Able to run with CESM, but we're running it **standalone**.
- Simulating only a **single ocean column** (BATS station).

MARBL

- MARBL is the **Marine Biogeochemistry Library** [3].

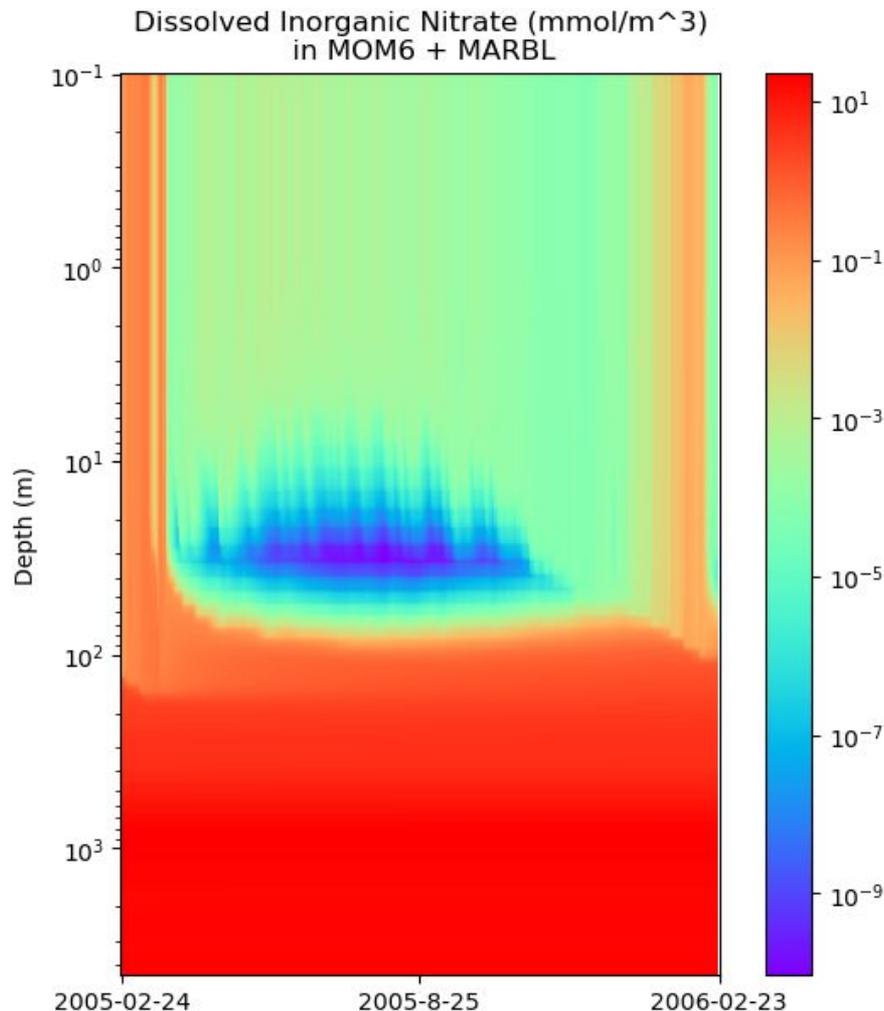


MOM6

- MOM6 is the **Modular Ocean Model**.
- Able to run with CESM, but we're running it **standalone**.
- Simulating only a **single ocean column** (BATS station).

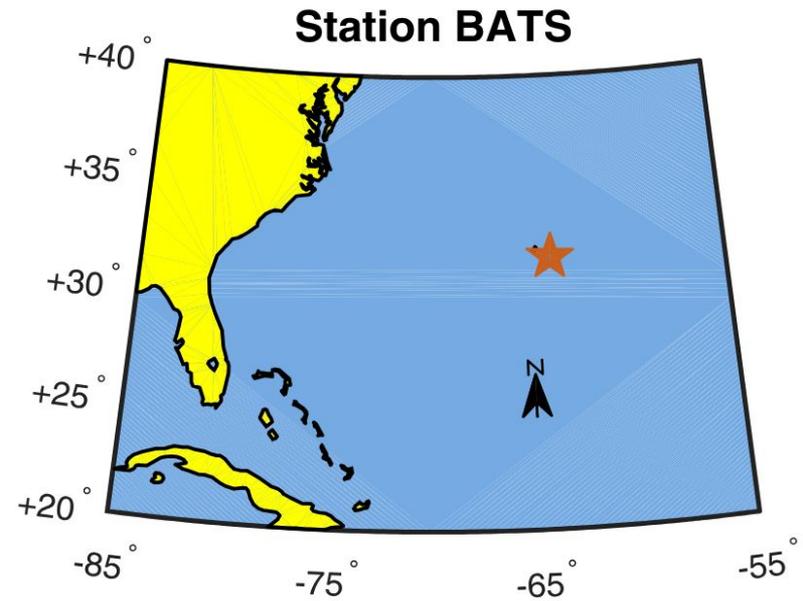
MARBL

- MARBL is the **Marine Biogeochemistry Library** [3].
- Simulates many BGC tracers, for example:
 - Nutrient concentrations (phosphate, nitrate, etc).
 - Concentrations of dissolved gasses (O₂ and CO₂).
 - Plankton concentrations.



BATS

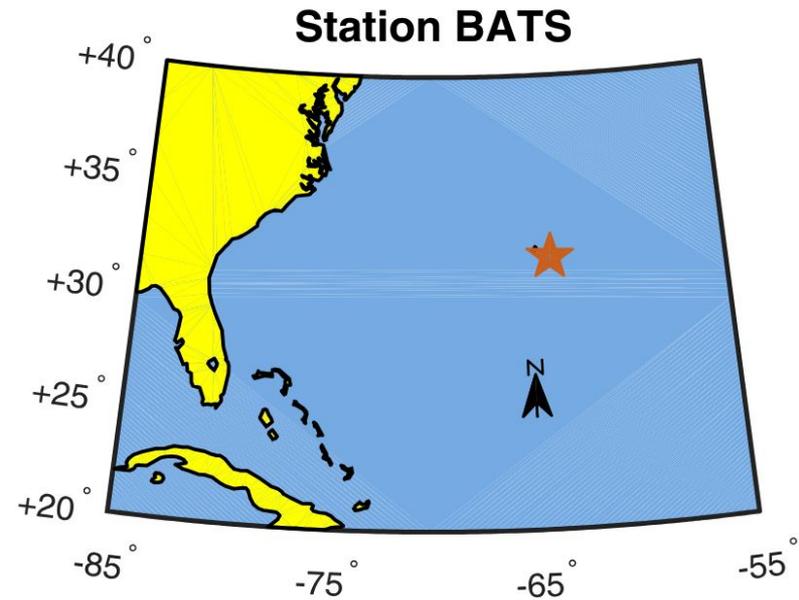
- BATS is the **Bermuda Atlantic Time-Series Study**.



Location where Bermuda Atlantic Time-Series data is taken [2].

BATS

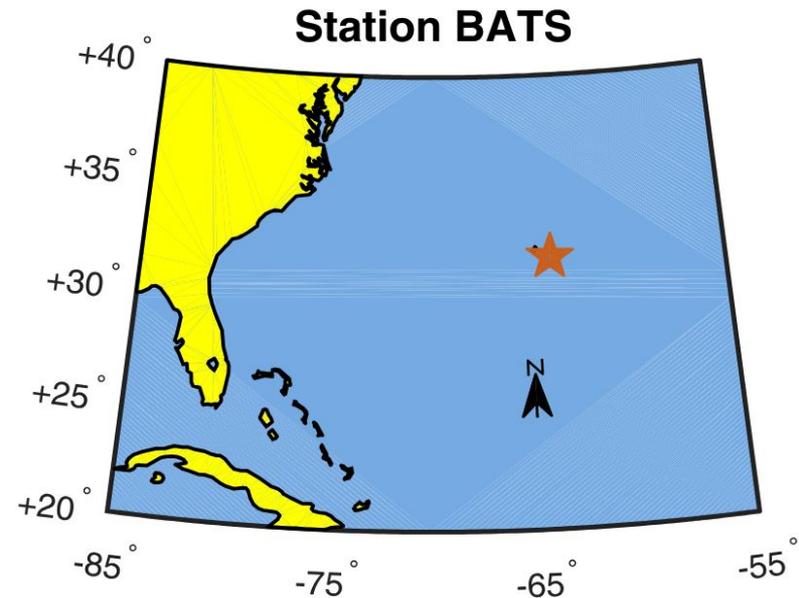
- BATS is the **Bermuda Atlantic Time-Series Study**.
- Measurements of physical and biogeochemical variables at various depths, taken from a water column in Bermuda.



Location where Bermuda Atlantic Time-Series data is taken [2].

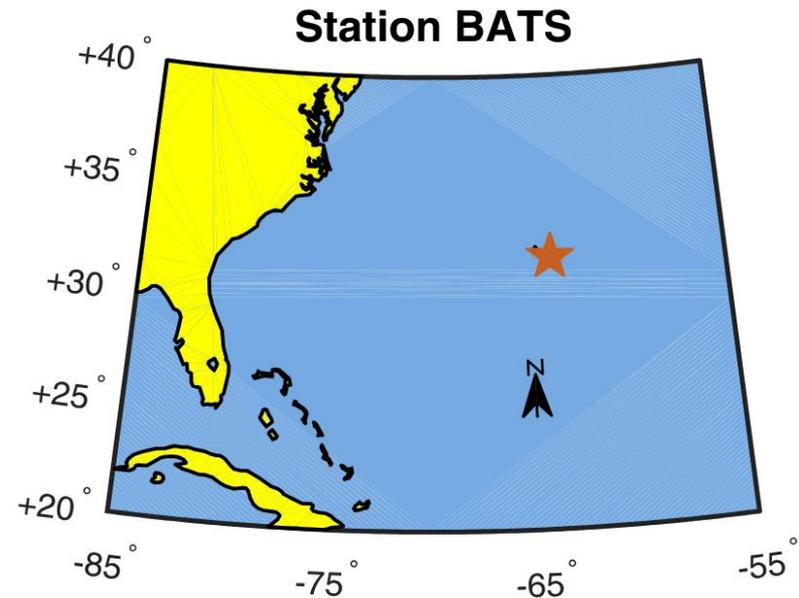
BATS

- BATS is the **Bermuda Atlantic Time-Series Study**.
- Measurements of physical and biogeochemical variables at various depths, taken from a water column in Bermuda.
- **Examples:**
 - Temperature
 - Salinity
 - Concentrations of O₂, CO₂
 - Nutrient concentrations (nitrate, phosphate, etc)



Location where Bermuda Atlantic Time-Series data is taken [2].

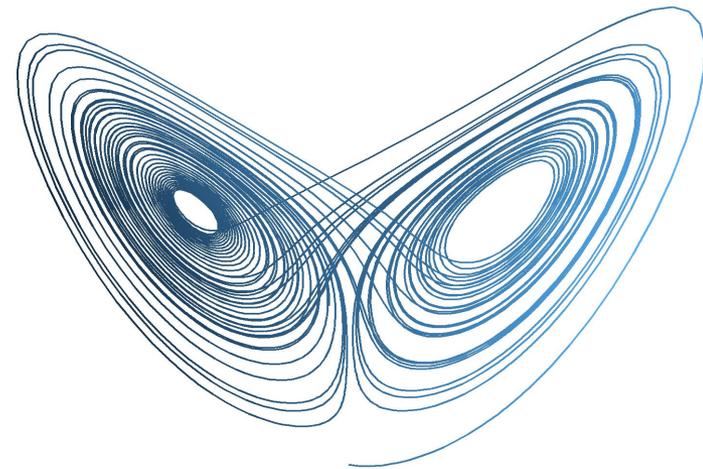
- BATS is the **Bermuda Atlantic Time-Series Study**.
- Measurements of physical and biogeochemical variables at various depths, taken from a water column in Bermuda.
- **Examples:**
 - Temperature
 - Salinity
 - Concentrations of O₂, CO₂
 - Nutrient concentrations (nitrate, phosphate, etc)
- 35 years of data, spanning from **1988 to present**.



Location where Bermuda Atlantic Time-Series data is taken [2].

DART

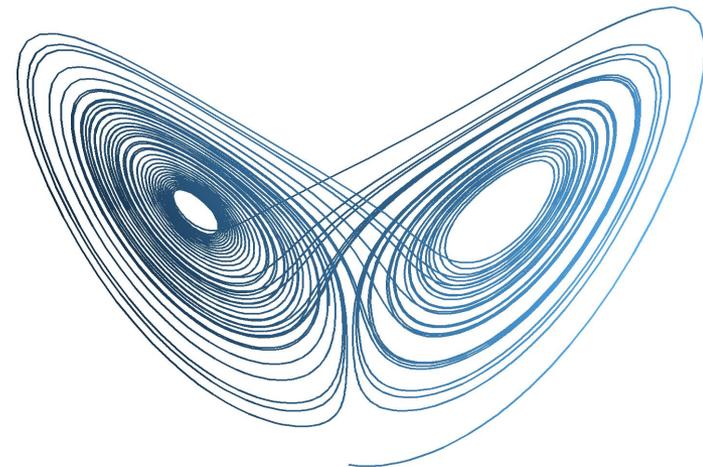
- DART is the **Data Assimilation Research Testbed** [1].



A trajectory of the Lorenz system, a small model which is used to test DART algorithms. Source: <http://imperfectlens.com/growing-a-lorenz-butterfly-in-r/>

DART

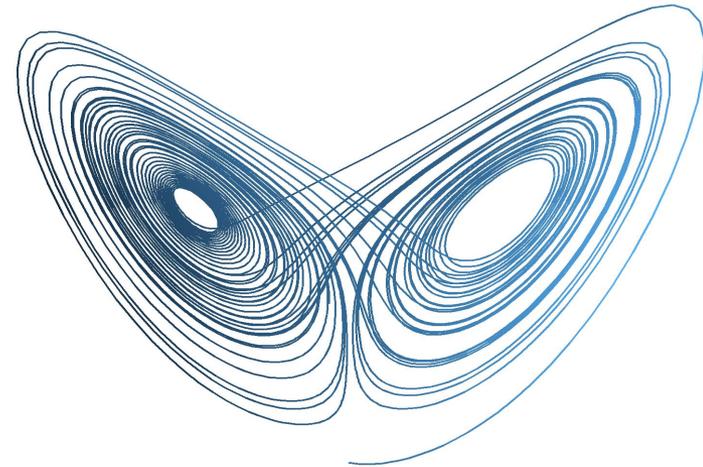
- DART is the **Data Assimilation Research Testbed** [1].
- Fortran software developed by NCAR in the **Data Assimilation Research Section (DAReS)**.



A trajectory of the Lorenz system, a small model which is used to test DART algorithms. Source: <http://imperfectlens.com/growing-a-lorenz-butterfly-in-r/>

DART

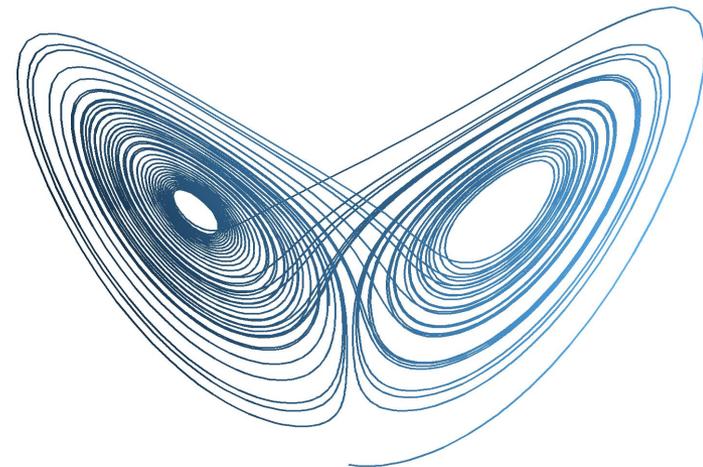
- DART is the **Data Assimilation Research Testbed** [1].
- Fortran software developed by NCAR in the **Data Assimilation Research Section (DAReS)**.
- Capable of running **many different algorithms** for **ensemble data assimilation**.



A trajectory of the Lorenz system, a small model which is used to test DART algorithms. Source: <http://imperfectlens.com/growing-a-lorenz-butterfly-in-r/>

DART

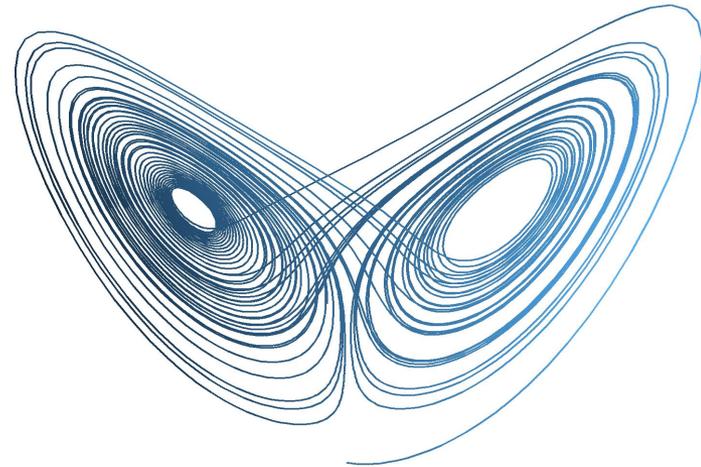
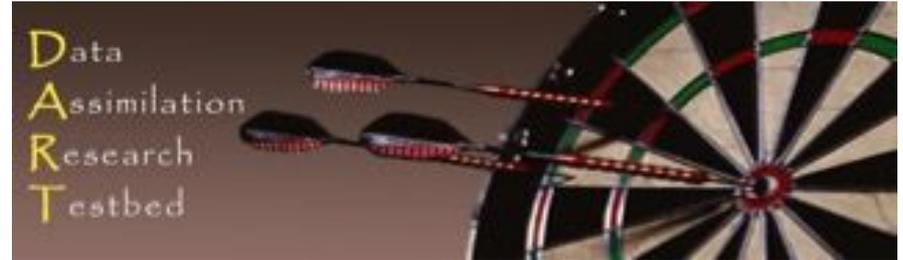
- DART is the **Data Assimilation Research Testbed** [1].
- Fortran software developed by NCAR in the **Data Assimilation Research Section (DAReS)**.
- Capable of running **many different algorithms** for **ensemble data assimilation**.
- Built for **portability**; can be interfaced to any model by implementing a few functions.



A trajectory of the Lorenz system, a small model which is used to test DART algorithms. Source: <http://imperfectlens.com/growing-a-lorenz-butterfly-in-r/>

DART

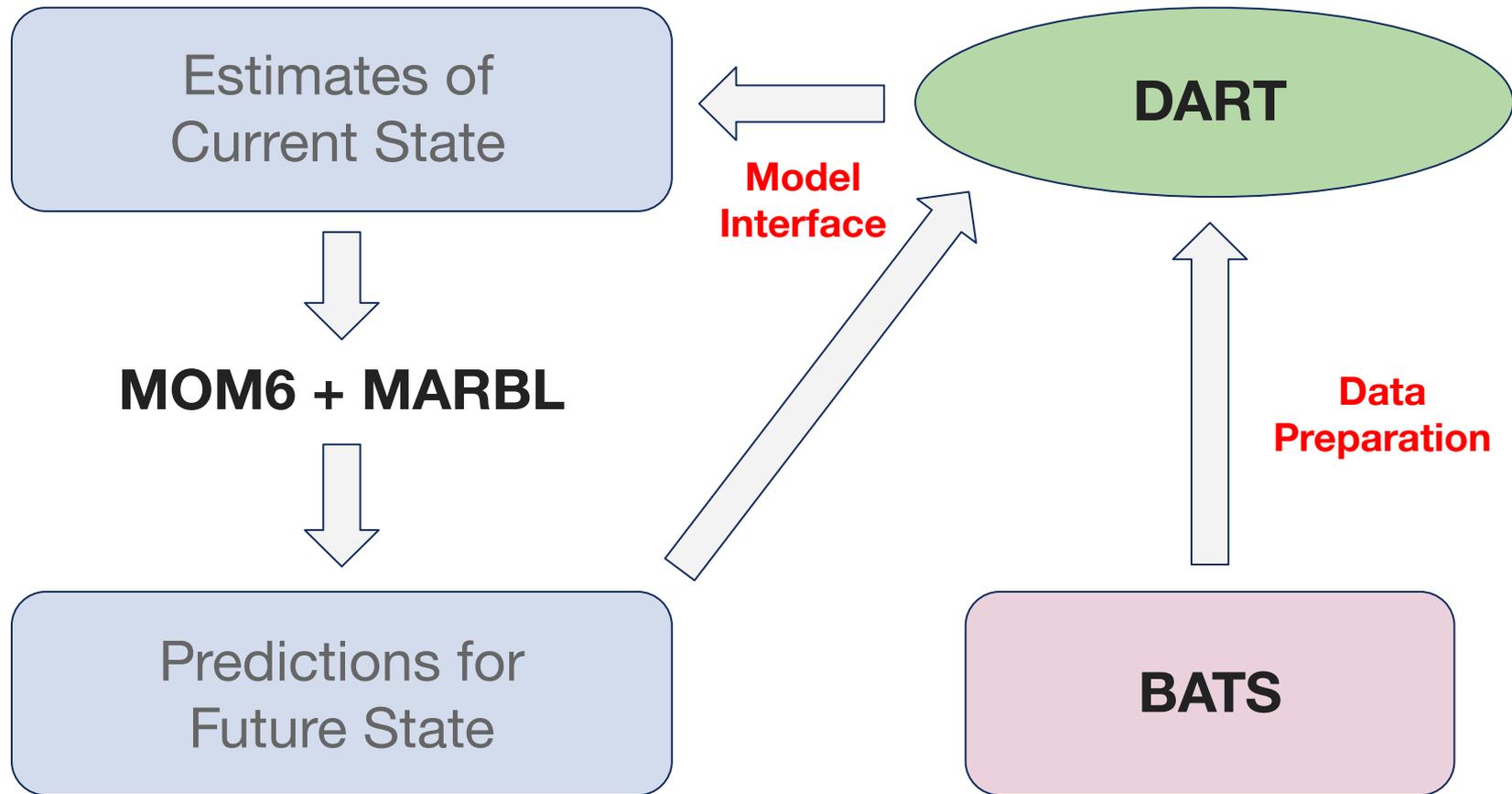
- DART is the **Data Assimilation Research Testbed** [1].
- Fortran software developed by NCAR in the **Data Assimilation Research Section (DAReS)**.
- Capable of running **many different algorithms** for **ensemble data assimilation**.
- Built for **portability**; can be interfaced to any model by implementing a few functions.
- Includes small, self-contained models which serve as **data assimilation tutorials**.



A trajectory of the Lorenz system, a small model which is used to test DART algorithms. Source: <http://imperfectlens.com/growing-a-lorenz-butterfly-in-r/>

Part 3: Building a Data Assimilation System

PUTTING THE PIECES TOGETHER



DATA PREPARATION

```
/Variables
  Id      ,  yyyyymmdd,   decy   ,  time ,  latN
1000100112  19881020  1988.80311  2230  31.783
1000100111  19881020  1988.80311  2230  31.783
1000100110  19881020  1988.80311  2230  31.783
1000100109  19881020  1988.80311  2230  31.783
1000100108  19881020  1988.80311  2230  31.783
1000100107  19881020  1988.80311  2230  31.783
1000100106  19881020  1988.80311  2230  31.783
1000100105  19881020  1988.80311  2230  31.783
1000100104  19881020  1988.80311  2230  31.783
1000100103  19881020  1988.80311  2230  31.783
1000100102  19881020  1988.80311  2230  31.783
1000100101  19881020  1988.80311  2230  31.783
1000108112  19881021  1988.80330   455  31.833
1000108111  19881021  1988.80330   455  31.833
1000108110  19881021  1988.80330   455  31.833
1000108109  19881021  1988.80330   455  31.833
1000108108  19881021  1988.80330   455  31.833
1000108107  19881021  1988.80330   455  31.833
1000108106  19881021  1988.80330   455  31.833
1000108105  19881021  1988.80330   455  31.833
1000108104  19881021  1988.80330   455  31.833
1000108103  19881021  1988.80330   455  31.833
1000108102  19881021  1988.80330   455  31.833
1000108101  19881021  1988.80330   455  31.833
```



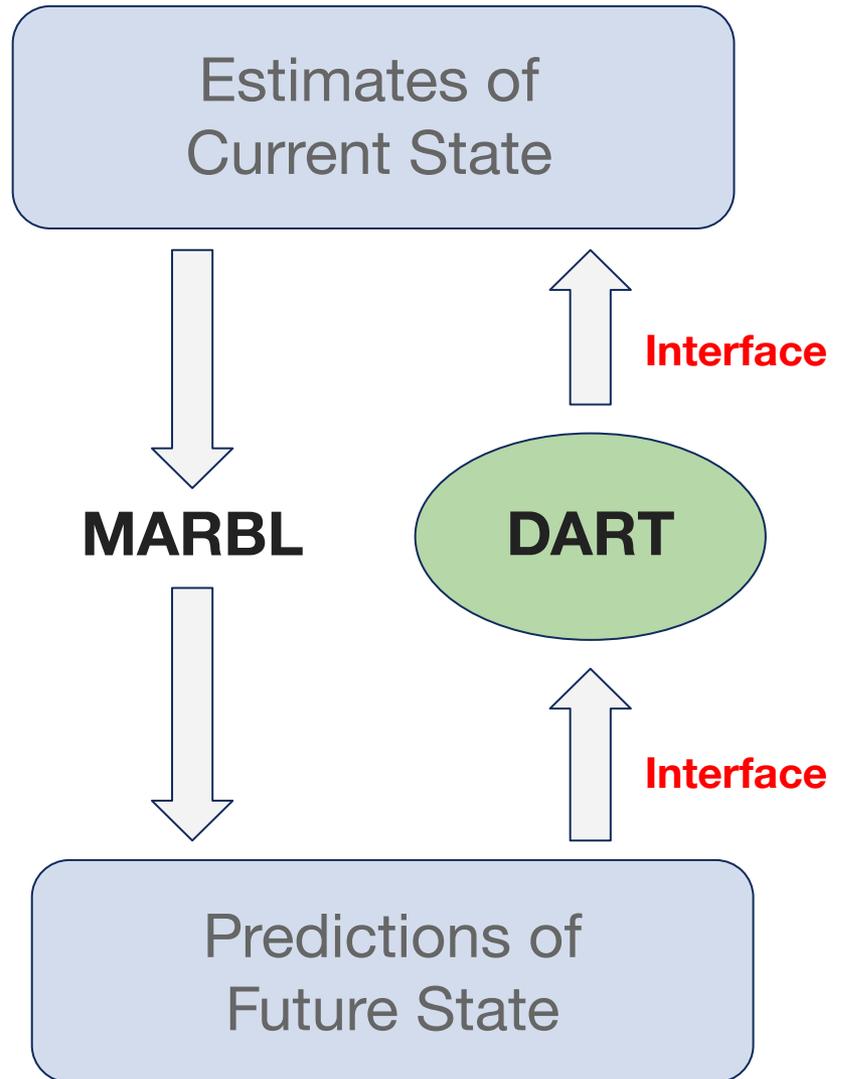
```
obs_sequence
obs_type_definitions
  7
  67 POLY_ELECTRODE_OXYGEN
  68 TITRATION_ALKALINITY
  69 CATALYTIC_CARBON
  70 UV_OXY_NITROGEN
  71 CFA_NITRATE
  72 CFA_SILICATE
  73 CFA_PHOSPHATE
  num_copies:          1  num_qc:          1
  num_obs:            113  max_num_obs:    113
observation
Data QC
  first:              1  last:            113
  OBS                1
  201.6000000000000
  0.000000000000000E+000
  -1                2                -1
obdef
loc3d
  1.117010721276371      0.5410520681182421
kind
  67
  83100      147611
  1625.702400000000
  OBS        2
  2384.200000000000
  0.000000000000000E+000
  1                3                -1
```

BATS Data Format

DART Data Format

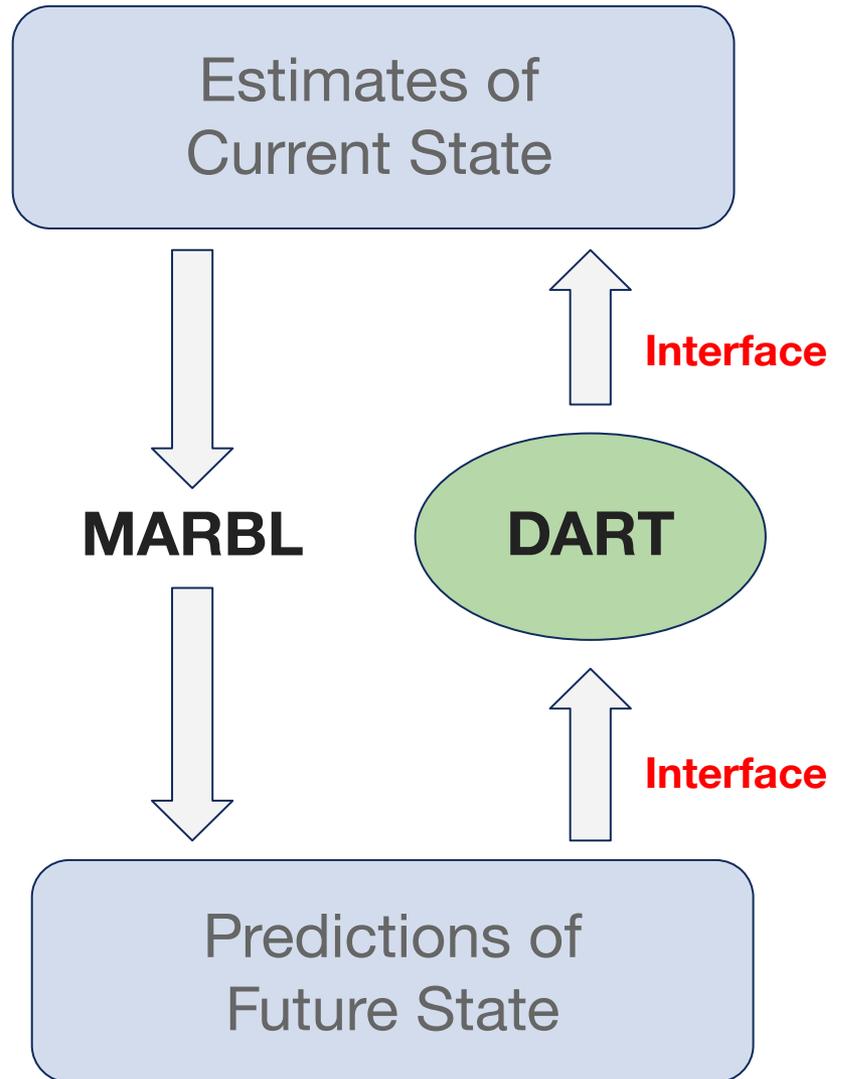
MODEL INTERFACE

- DART and MOM6 + MARBL are completely different softwares.
They need to talk to each other.



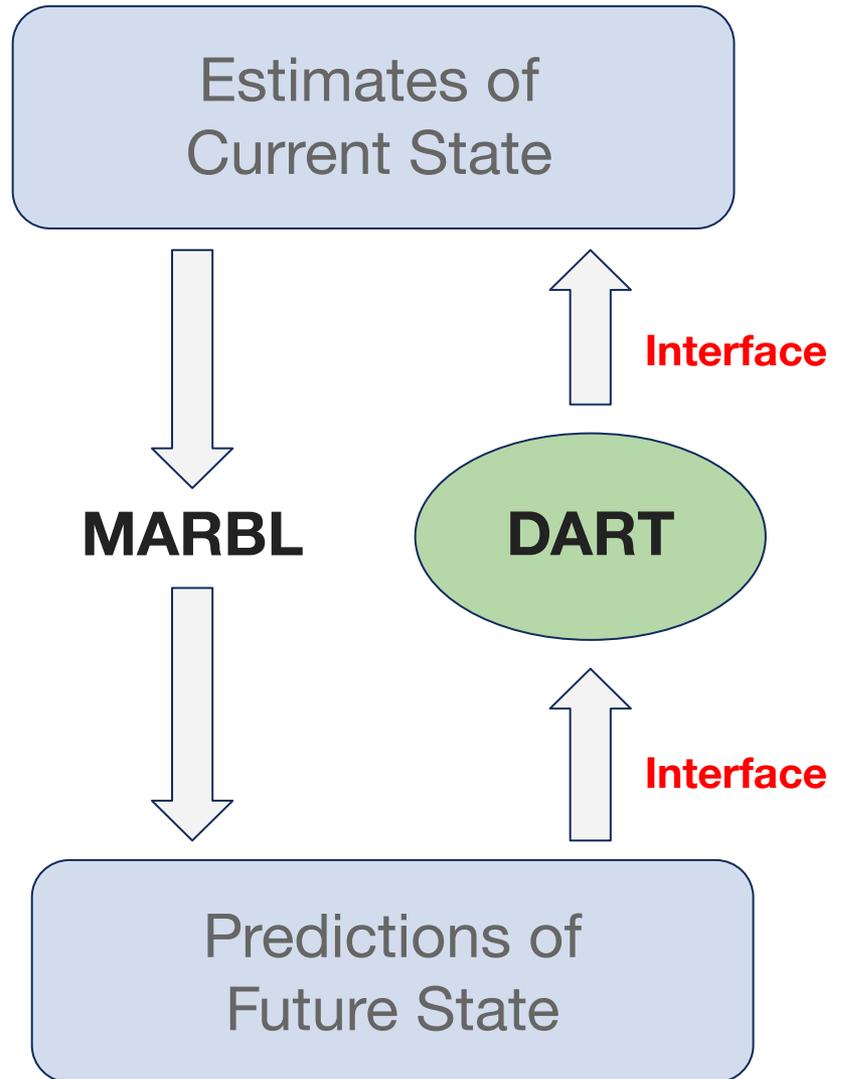
MODEL INTERFACE

- DART and MOM6 + MARBL are completely different softwares. **They need to talk to each other.**
- This requires a set of **interface functions** which tell DART how to:
 - **Read outputs** produced by the model.



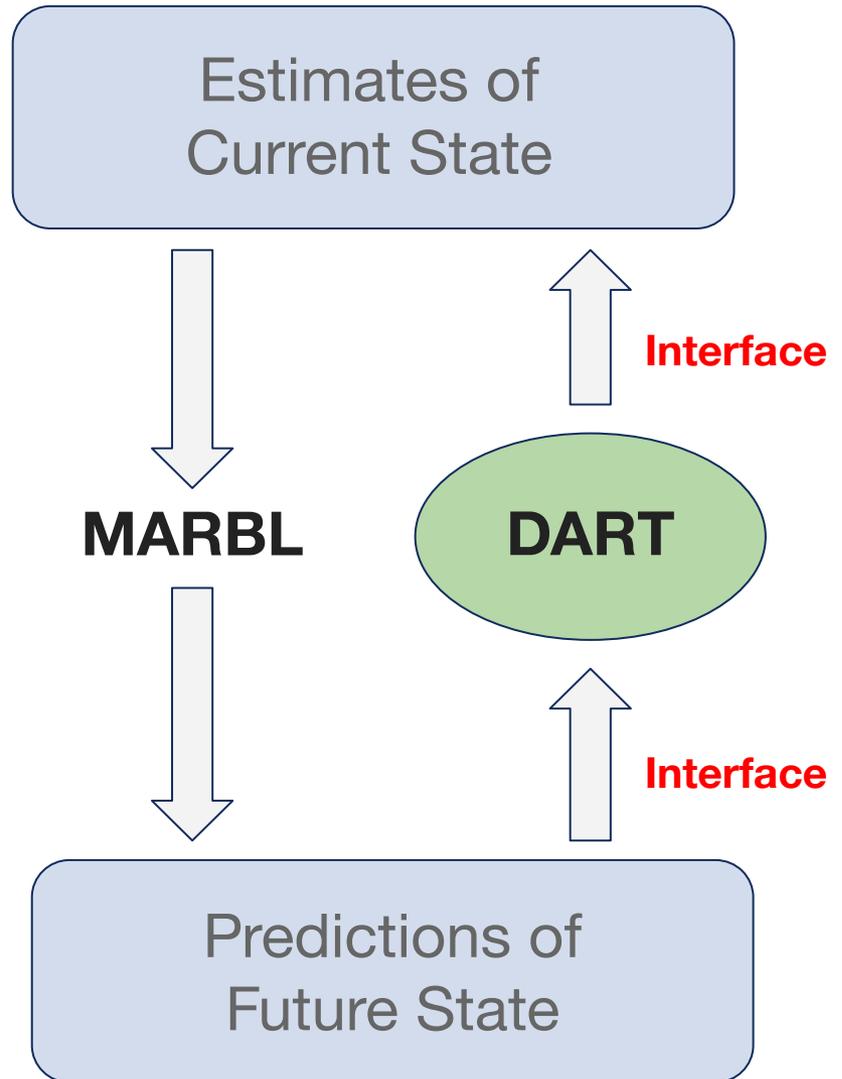
MODEL INTERFACE

- DART and MOM6 + MARBL are completely different softwares. **They need to talk to each other.**
- This requires a set of **interface functions** which tell DART how to:
 - **Read outputs** produced by the model.
 - **Write inputs** for the model to use.



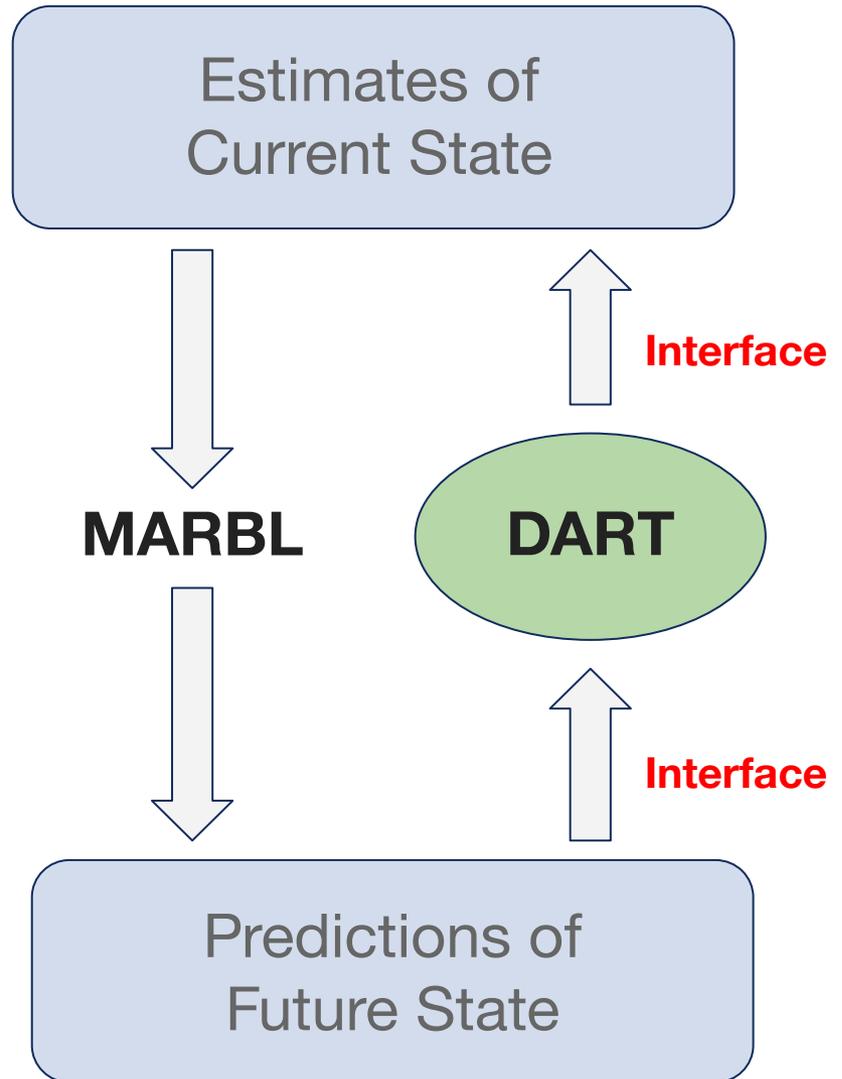
MODEL INTERFACE

- DART and MOM6 + MARBL are completely different softwares. **They need to talk to each other.**
- This requires a set of **interface functions** which tell DART how to:
 - **Read outputs** produced by the model.
 - **Write inputs** for the model to use.
 - **Interpolate** from the model grid onto the measurement location.



MODEL INTERFACE

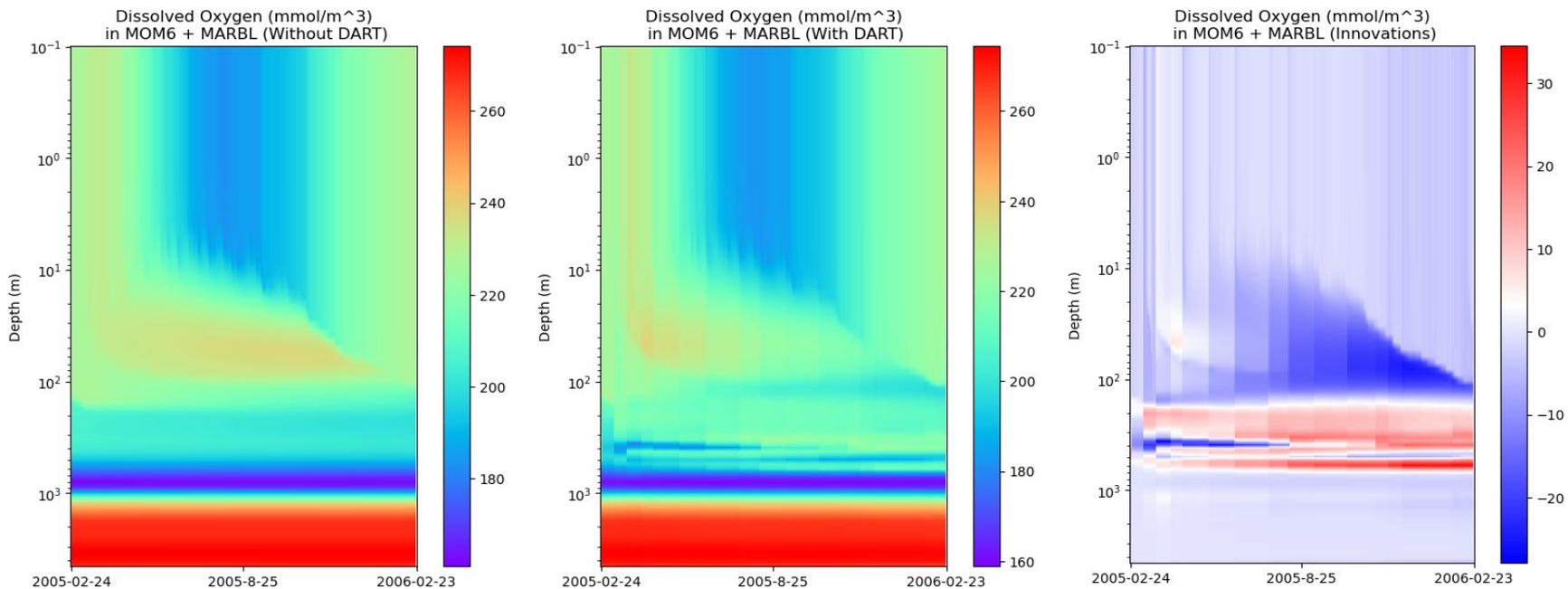
- DART and MOM6 + MARBL are completely different softwares. **They need to talk to each other.**
- This requires a set of **interface functions** which tell DART how to:
 - **Read outputs** produced by the model.
 - **Write inputs** for the model to use.
 - **Interpolate** from the model grid onto the measurement location.
- Writing this interface was the most complicated part of the implementation.



Part 4: Results



HOVMÖLLER DIAGRAMS

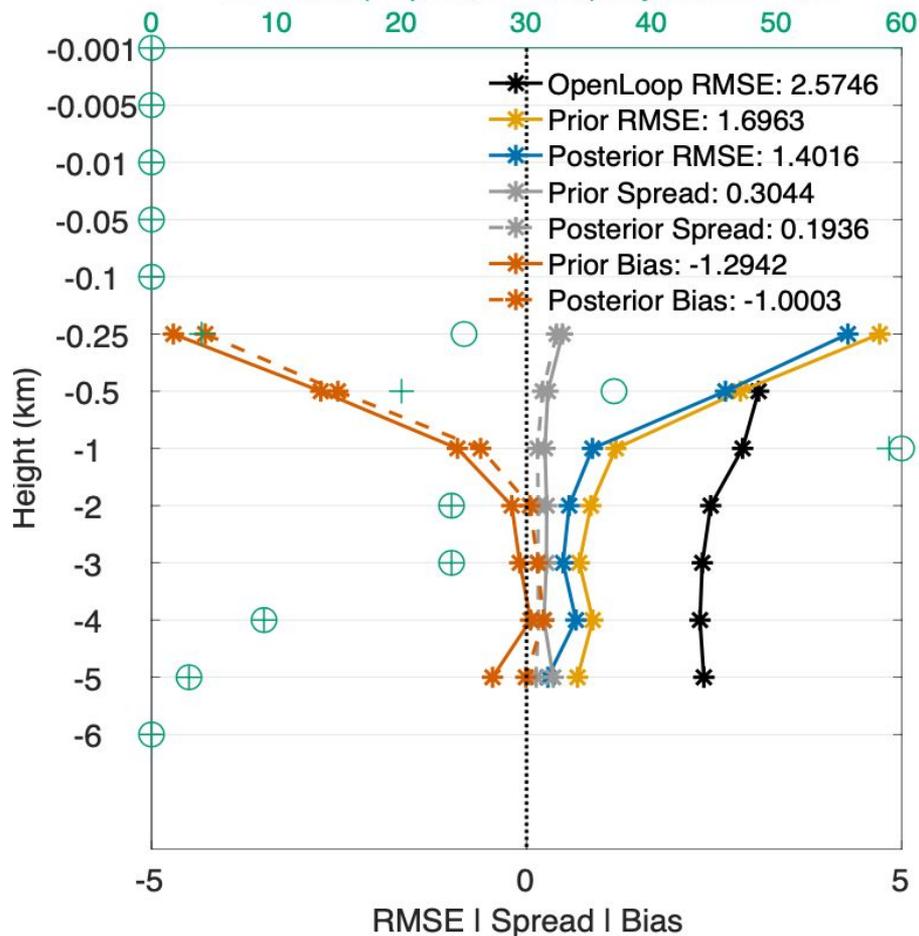


DART DIAGNOSTICS

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

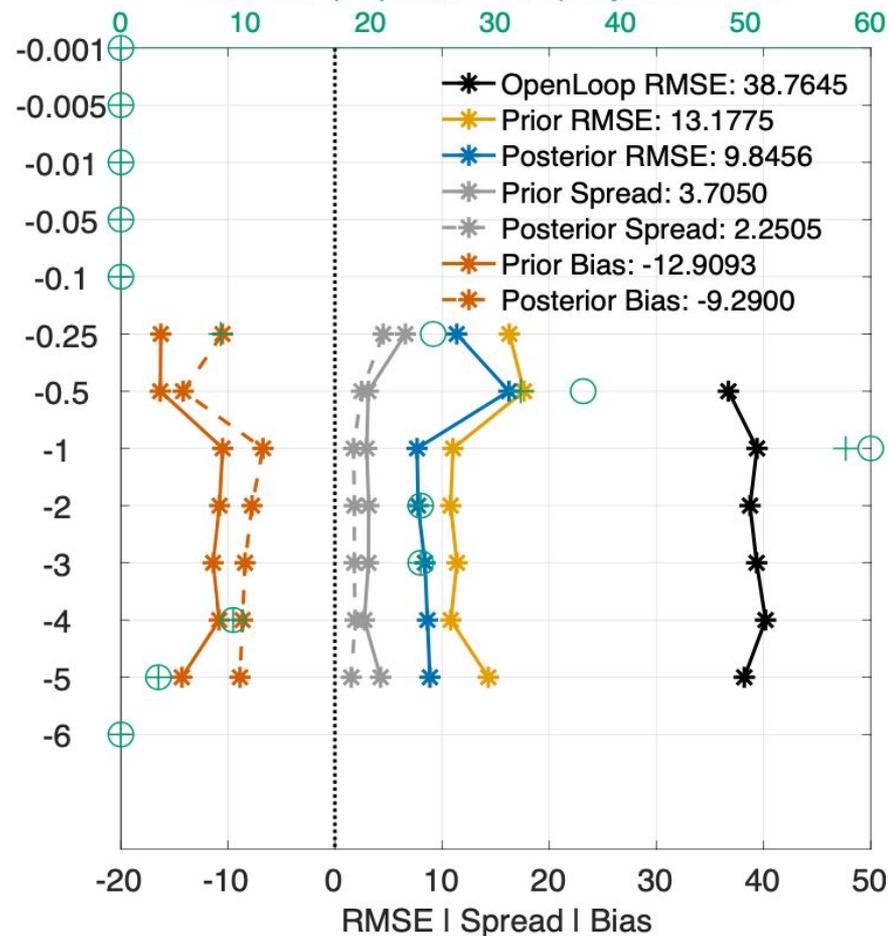
Dissolved Organic Nitrogen [mmol/m³]

No. of obs (o= poss, += used) Rejection: 21.4%

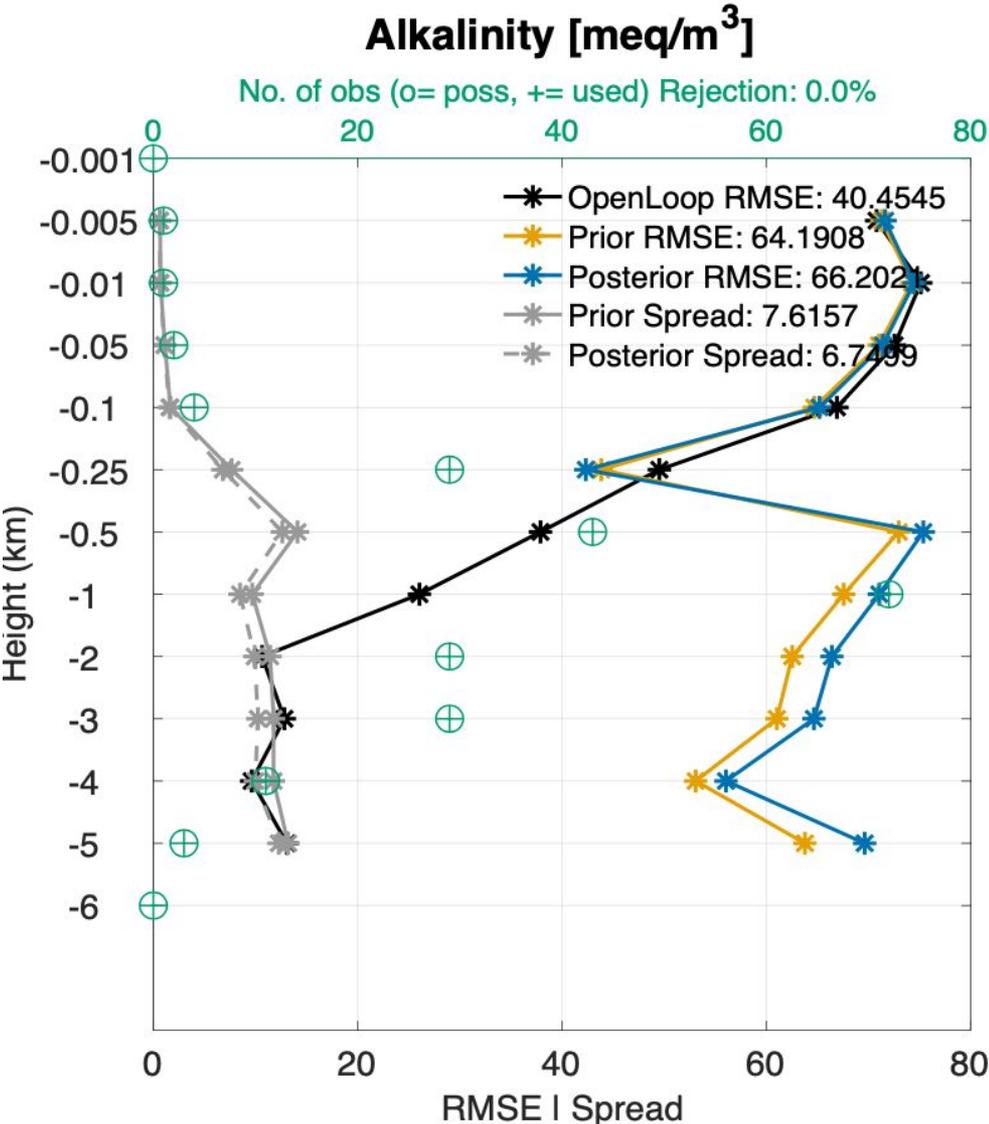


Dissolved Organic Carbon [mmol/m³]

No. of obs (o= poss, += used) Rejection: 13.2%



WHAT'S GOING ON WITH ALKALINITY?



Part 5: Conclusions



CONCLUSIONS AND FUTURE WORK

- MARBL-DART provides the **novel capability** to **assimilate data** into the **MARBL biogeochemistry model**.



CONCLUSIONS AND FUTURE WORK

- MARBL-DART provides the **novel capability** to **assimilate data** into the **MARBL biogeochemistry model**.
- Broadly, MARBL-DART produces forecasts that are **more accurate** than those produced by **MARBL alone**.



CONCLUSIONS AND FUTURE WORK

- MARBL-DART provides the **novel capability** to **assimilate data** into the **MARBL biogeochemistry model**.
- Broadly, MARBL-DART produces forecasts that are **more accurate than those produced by MARBL alone**.
- More work is needed to explain MARBL-DART's **anomalous alkalinity forecasts**.



CONCLUSIONS AND FUTURE WORK

- MARBL-DART provides the **novel capability** to **assimilate data** into the **MARBL biogeochemistry model**.
- Broadly, MARBL-DART produces forecasts that are **more accurate than those produced by MARBL alone**.
- More work is needed to explain MARBL-DART's **anomalous alkalinity forecasts**.
- **Future work:** we will build on MARBL-DART to create a data assimilation system for **biogeochemical parameter optimization**.

CONCLUSIONS AND FUTURE WORK

- MARBL-DART provides the **novel capability** to **assimilate data into the MARBL biogeochemistry model**.
- Broadly, MARBL-DART produces forecasts that are **more accurate than those produced by MARBL alone**.
- More work is needed to explain MARBL-DART's **anomalous alkalinity forecasts**.
- **Future work:** we will build on MARBL-DART to create a data assimilation system for **biogeochemical parameter optimization**.
- This means using data assimilation to **modify the underlying model itself**, instead of just modifying the model forecasts.

CONCLUSIONS AND FUTURE WORK

- MARBL-DART provides the **novel capability to assimilate data into the MARBL biogeochemistry model.**
- Broadly, MARBL-DART produces forecasts that are **more accurate than those produced by MARBL alone.**
- More work is needed to explain MARBL-DART's **anomalous alkalinity forecasts.**
- **Future work:** we will build on MARBL-DART to create a data assimilation system for **biogeochemical parameter optimization.**
- This means using data assimilation to **modify the underlying model itself**, instead of just modifying the model forecasts.
- To accomplish this we will assimilate new sources of data, e.g. **satellite chlorophyll data.**

CONCLUSIONS AND FUTURE WORK

- MARBL-DART provides the **novel capability** to **assimilate data** into the **MARBL biogeochemistry model**.
- Broadly, MARBL-DART produces forecasts that are **more accurate than those produced by MARBL alone**.
- More work is needed to explain MARBL-DART's **anomalous alkalinity forecasts**.
- **Future work:** we will build on MARBL-DART to create a data assimilation system for **biogeochemical parameter optimization**.
- This means using data assimilation to **modify the underlying model itself**, instead of just modifying the model forecasts.
- To accomplish this we will assimilate new sources of data, e.g. **satellite chlorophyll data**.
- This will also allow us to assimilate data **more frequently** than what we are currently capable of.

THANK YOU'S

Thank you to my mentors:



THANK YOU'S

Thank you to my mentors:

Moha Gharamti



THANK YOU'S

Thank you to my mentors:

Moha Gharamti, Dan Amrhein



THANK YOU'S

Thank you to my mentors:

Moha Gharamti, Dan Amrhein, Kristen Krumhardt



THANK YOU'S

Thank you to my mentors:

**Moha Gharamti, Dan Amrhein, Kristen Krumhardt,
Helen Kershaw**



THANK YOU'S

Thank you to my mentors:

**Moha Gharamti, Dan Amrhein, Kristen Krumhardt,
Helen Kershaw, and Mike Levy.**



THANK YOU'S

Thank you to my mentors:

**Moha Gharamti, Dan Amrhein, Kristen Krumhardt,
Helen Kershaw, and Mike Levy.**

Thank you to the internship organizers:



THANK YOU'S

Thank you to my mentors:

**Moha Gharamti, Dan Amrhein, Kristen Krumhardt,
Helen Kershaw, and Mike Levy.**

Thank you to the internship organizers:

Virginia Do



THANK YOU'S

Thank you to my mentors:

**Moha Gharamti, Dan Amrhein, Kristen Krumhardt,
Helen Kershaw, and Mike Levy.**

Thank you to the internship organizers:

Virginia Do, Jerry Cyccone



THANK YOU'S

Thank you to my mentors:

**Moha Gharamti, Dan Amrhein, Kristen Krumhardt,
Helen Kershaw, and Mike Levy.**

Thank you to the internship organizers:

Virginia Do, Jerry Cyccone, Julius Owusu Afriyie



THANK YOU'S

Thank you to my mentors:

**Moha Gharamti, Dan Amrhein, Kristen Krumhardt,
Helen Kershaw, and Mike Levy.**

Thank you to the internship organizers:

**Virginia Do, Jerry Cycone, Julius Owusu Afriyie,
and Ben Fellman.**



THANK YOU'S

Thank you to the organizations that support us:



THANK YOU'S

Thank you to the organizations that support us:



THANK YOU'S

Thank you to the organizations that support us:



THANK YOU'S

Thank you to the organizations that support us:



THANK YOU'S

Thank you to the organizations that support us:



REFERENCES

- [1] Anderson et al, “The Data Assimilation Research Testbed: A Community Facility,” *American Meteorological Society* (2009).
- [2] Gharamti et al, “Ensemble Data Assimilation for Ocean Biogeochemical State and Parameter Estimation at Different Sites,” *Ocean Modeling* (2017).
- [3] Long et al, “Simulations With the Marine Biogeochemistry Library,” *Journal of Advances in Modeling Earth Systems* (2021).