Using Machine Learning to Scale Scattered Biogeochemical Observations into Gridded Products

Jon Sharp (UW CICOES / NOAA PMEL)

Collaborators: Andrea Fassbender, Brendan Carter, John Lyman, Greg Johnson

2025 US CLIVAR Summit
Boulder, CO – 7/22/2025

1. discuss methods used to quantify the representativity of observations and process studies
2. review patterns of spatiotemporal patchiness and connectivity in physical and biological properties (e.g., correlation/decorrelation scales)
3. assess the potential of novel methods used to scale up observations into gridded regional and global products including uncertainty (e.g., AI/ML)

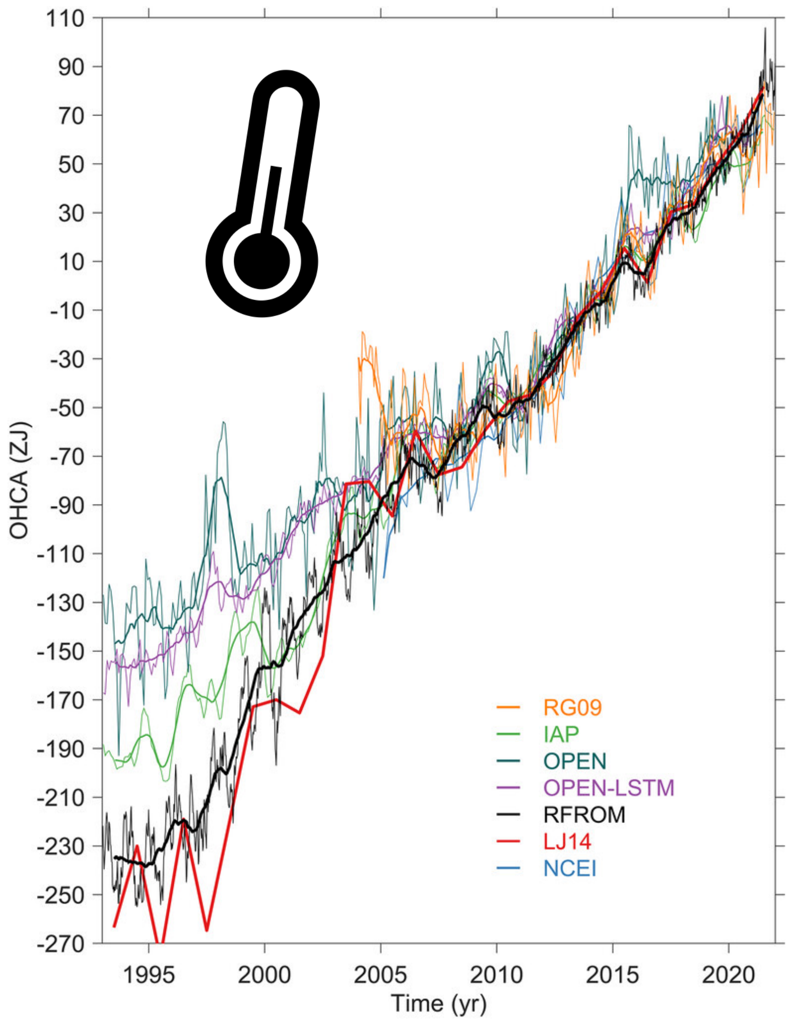
1. discuss methods used to quantify the representativity of observations and process studies
2. review patterns of spatiotemporal patchiness and connectivity in physical and biological properties (e.g., correlation/decorrelation scales)
3. assess the potential of novel methods used to scale up observations into gridded regional and global products including uncertainty (e.g., AI/ML)

THE OCEAN PROVIDES KEY CLIMATE SERVICES BY ABSORBING ANTHROPOGENIC HEAT AND CARBON

THIS RESULTS IN MAJOR IMPACTS ON MARINE ECOSYSTEMS AND BIOGEOCHEMISTRY

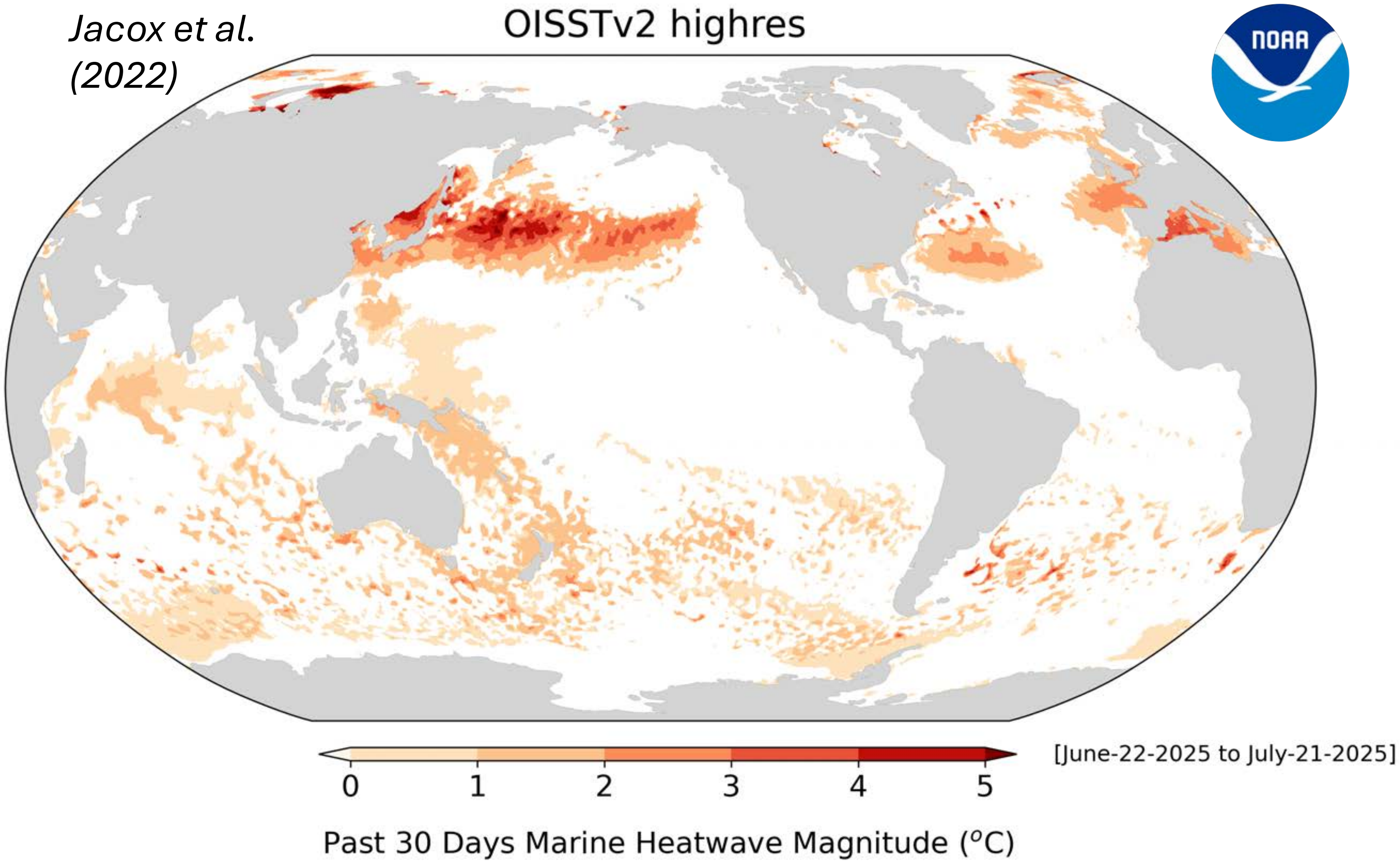
UNCERTAINTY REMAINS IN THE PAST AND FUTURE PROGRESSION OF THESE IMPACTS ON GLOBAL AND REGIONAL SCALES

ANTHROPOGENIC ACTIVITIES HAVE DRIVEN PHYSICAL AND BIOGEOCHEMICAL CHANGES IN THE GLOBAL OCEAN

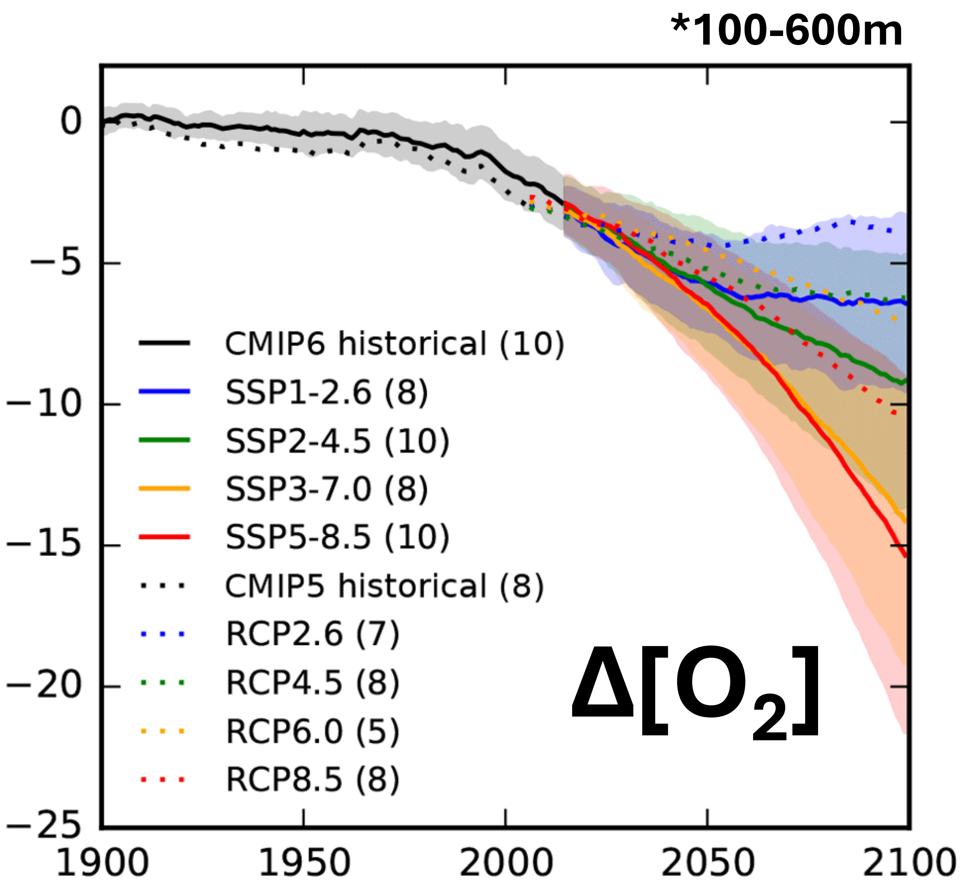


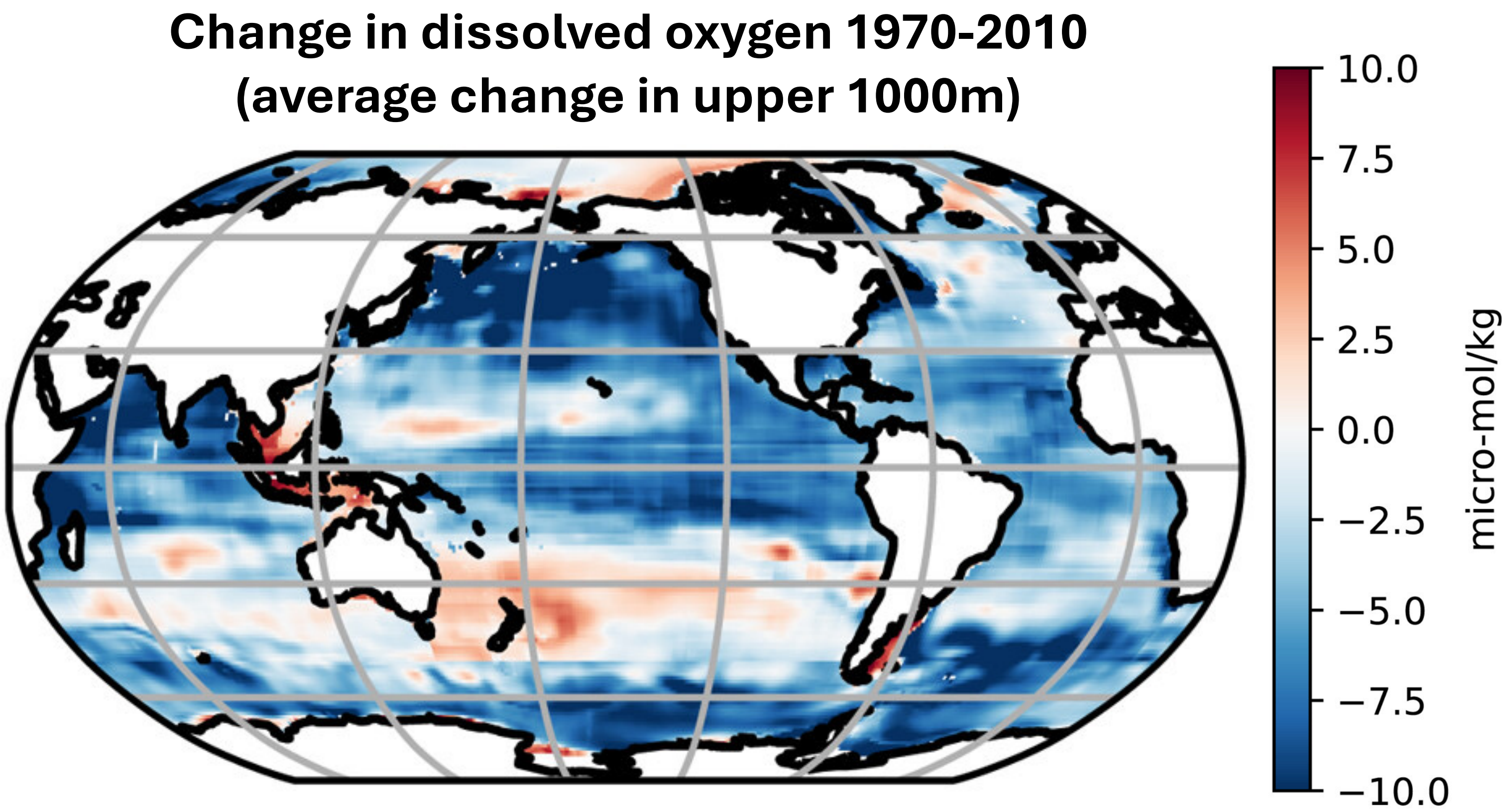
Lyman and Johnson (2023)

Jacox et al. (2022)



ANTHROPOGENIC ACTIVITIES HAVE DRIVEN PHYSICAL AND BIOGEOCHEMICAL CHANGES IN THE GLOBAL OCEAN



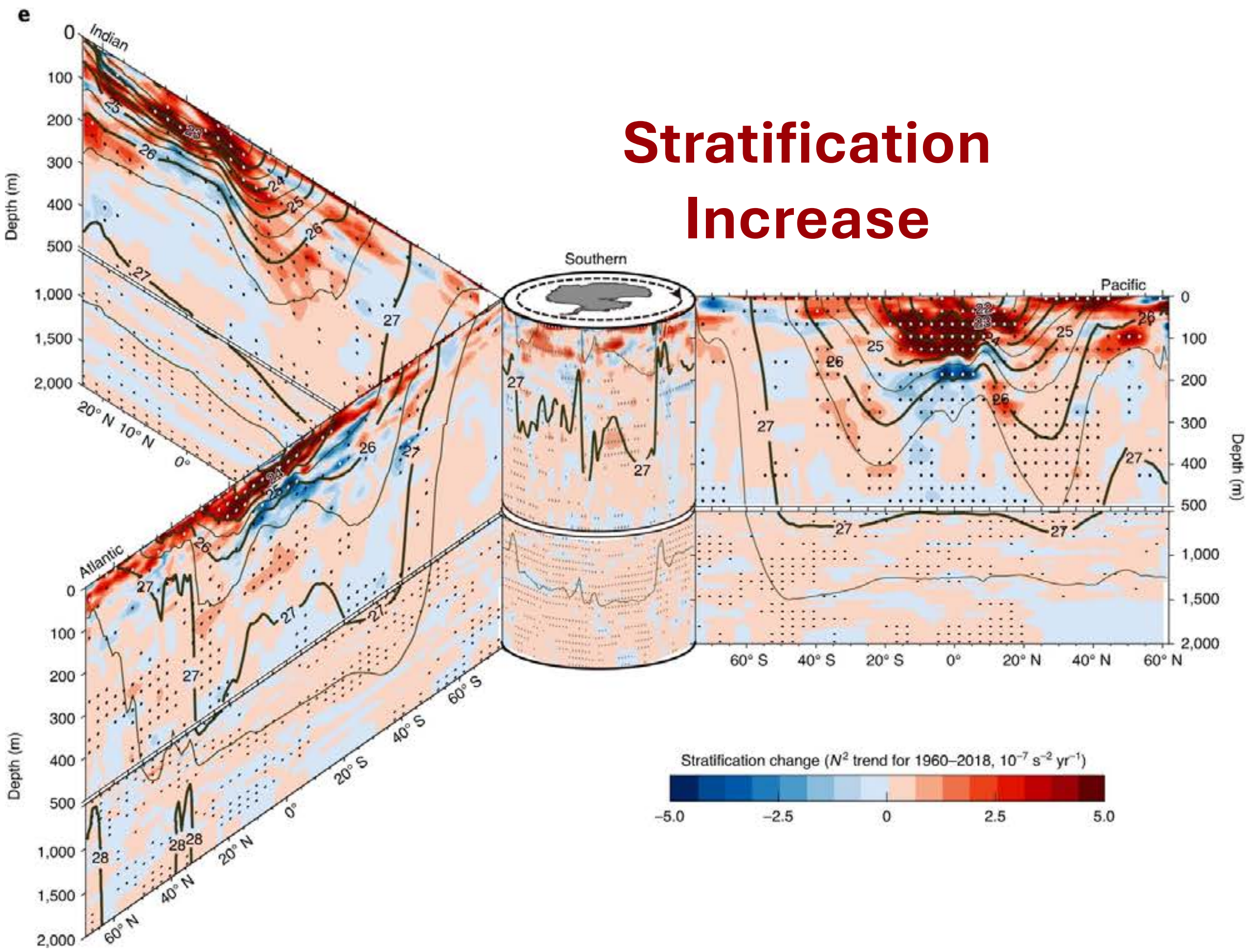


Kwiatkowski et al. (2020)

Ito et al. (2024)

THESE GLOBAL CHANGES HAVE SIGNIFICANT IMPLICATIONS FOR MARINE ECOSYSTEMS AND SPECIES

- Higher ocean temperatures result in greater stratification and **reduced supply of nutrients** to the euphotic zone (*Behrenfeld et al., 2006; Li et al., 2020*)

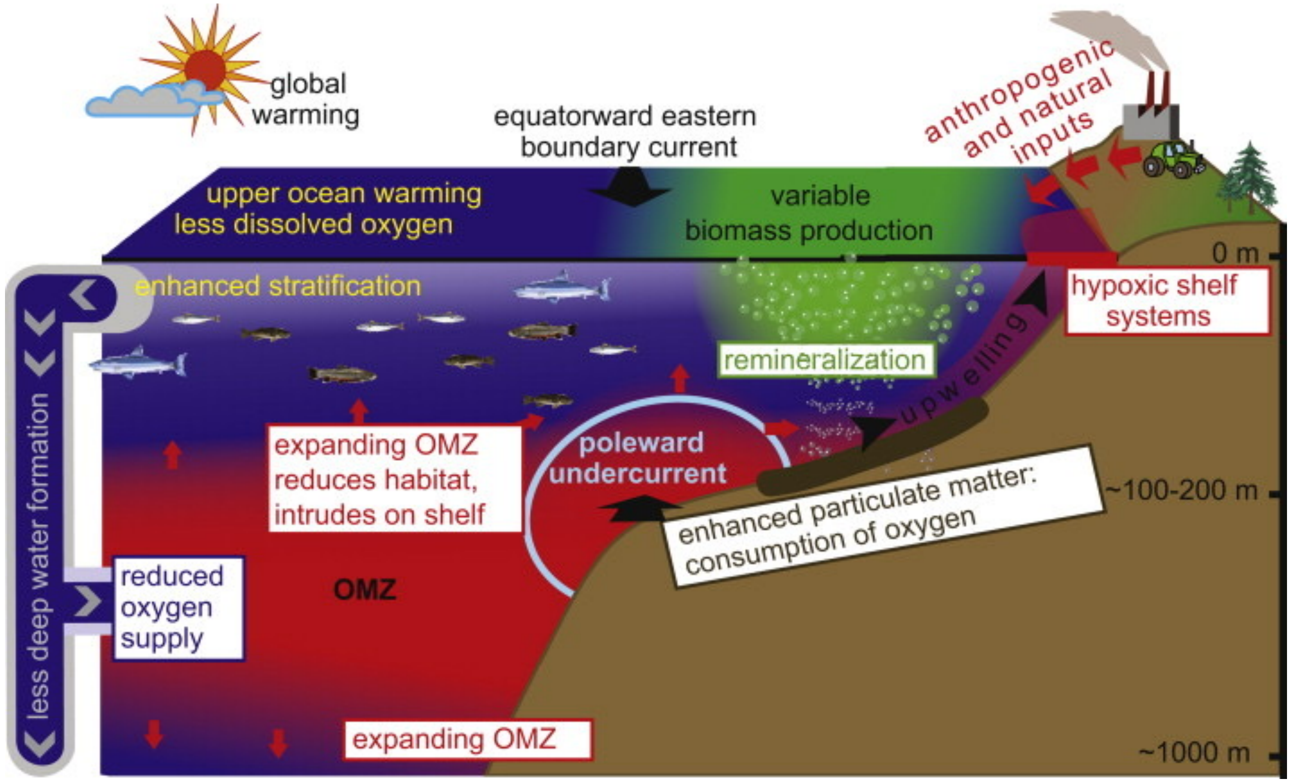


Stratification Increase

Li et al. (2020)

THESE GLOBAL CHANGES HAVE SIGNIFICANT IMPLICATIONS FOR MARINE ECOSYSTEMS AND SPECIES

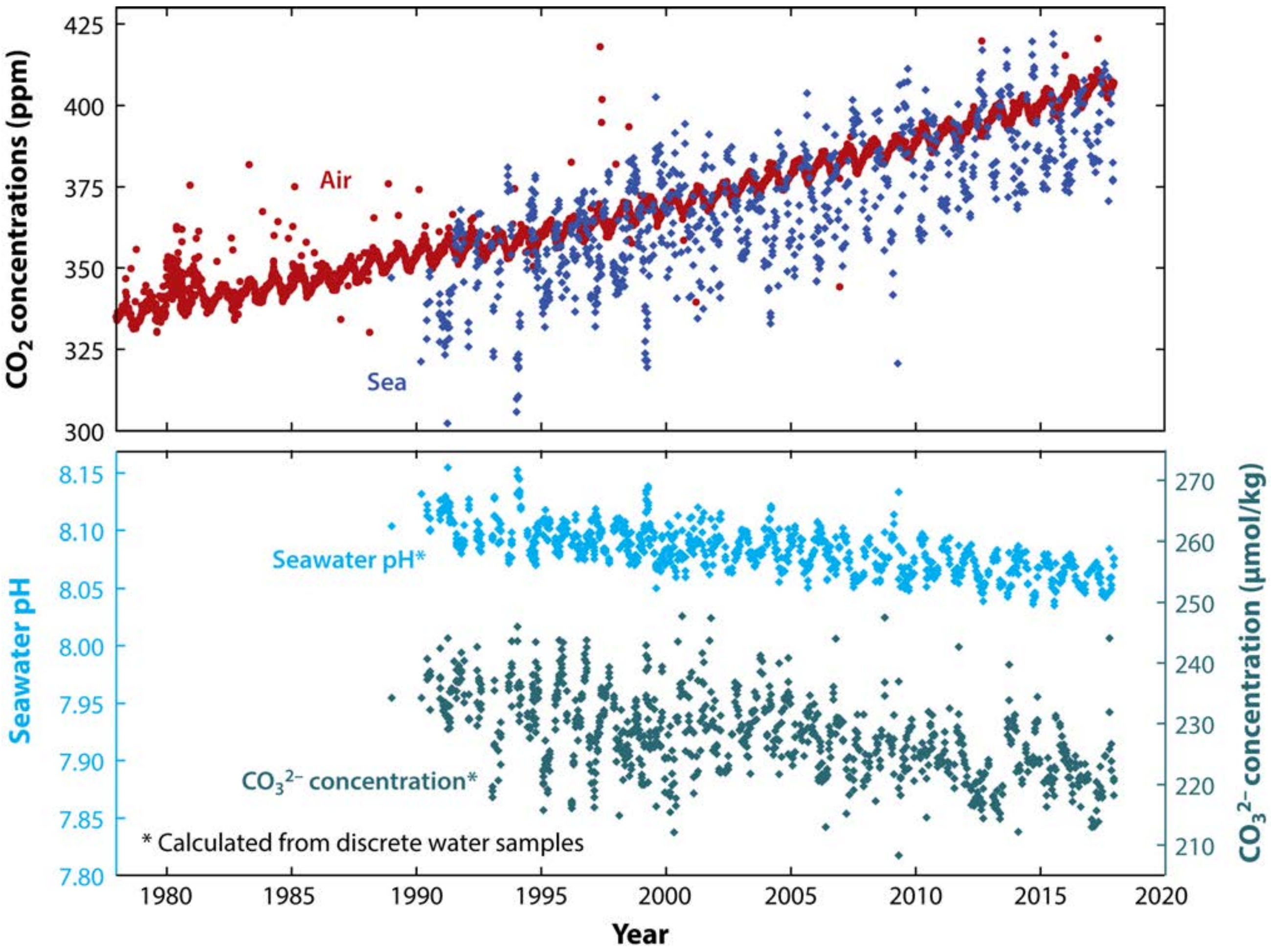
- Higher ocean temperatures result in greater stratification and **reduced supply of nutrients** to the euphotic zone (*Behrenfeld et al., 2006; Li et al., 2020*)
- Marine species respond directly to ocean temperature, especially in terms of **temperature-dependent oxygen tolerance** (*Pörtner, 2008; 2010*; *Deutsch et al., 2015*)



Stramma et al. (2010)

THESE GLOBAL CHANGES HAVE SIGNIFICANT IMPLICATIONS FOR MARINE ECOSYSTEMS AND SPECIES

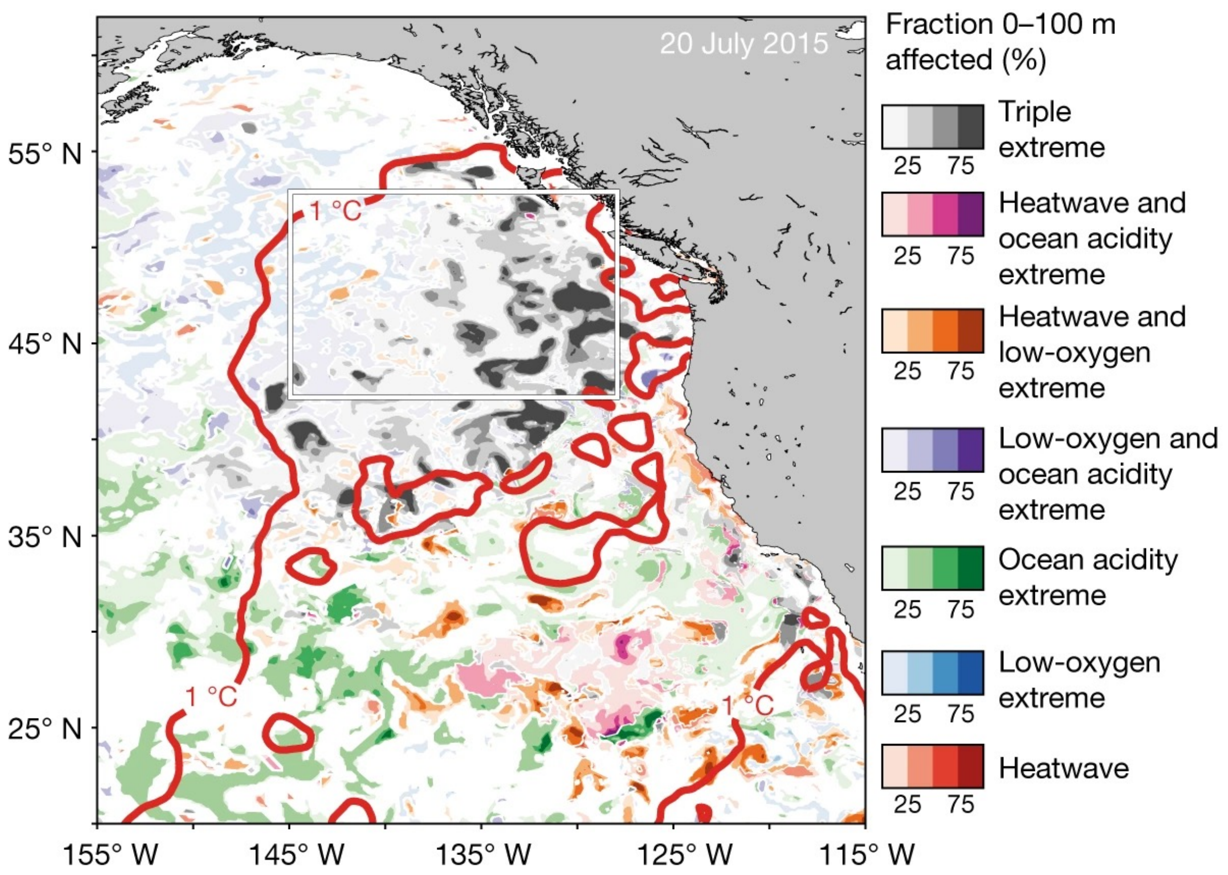
- Higher ocean temperatures result in greater stratification and **reduced supply of nutrients** to the euphotic zone (*Behrenfeld et al., 2006; Li et al., 2020*)
- Marine species respond directly to ocean temperature, especially in terms of **temperature-dependent oxygen tolerance** (*Pörtner, 2008; 2010*; *Deutsch et al., 2015*)
- **Ocean acidification** at the surface (*e.g., Jiang et al., 2023*) is detrimental to marine calcifiers and impacts $CaCO_3$ cycling and increased pCO_2 in the ocean interior (*Fassbender et al., 2023*) may influence other marine species



Doney et al. (2020)

THESE GLOBAL CHANGES HAVE SIGNIFICANT IMPLICATIONS FOR MARINE ECOSYSTEMS AND SPECIES

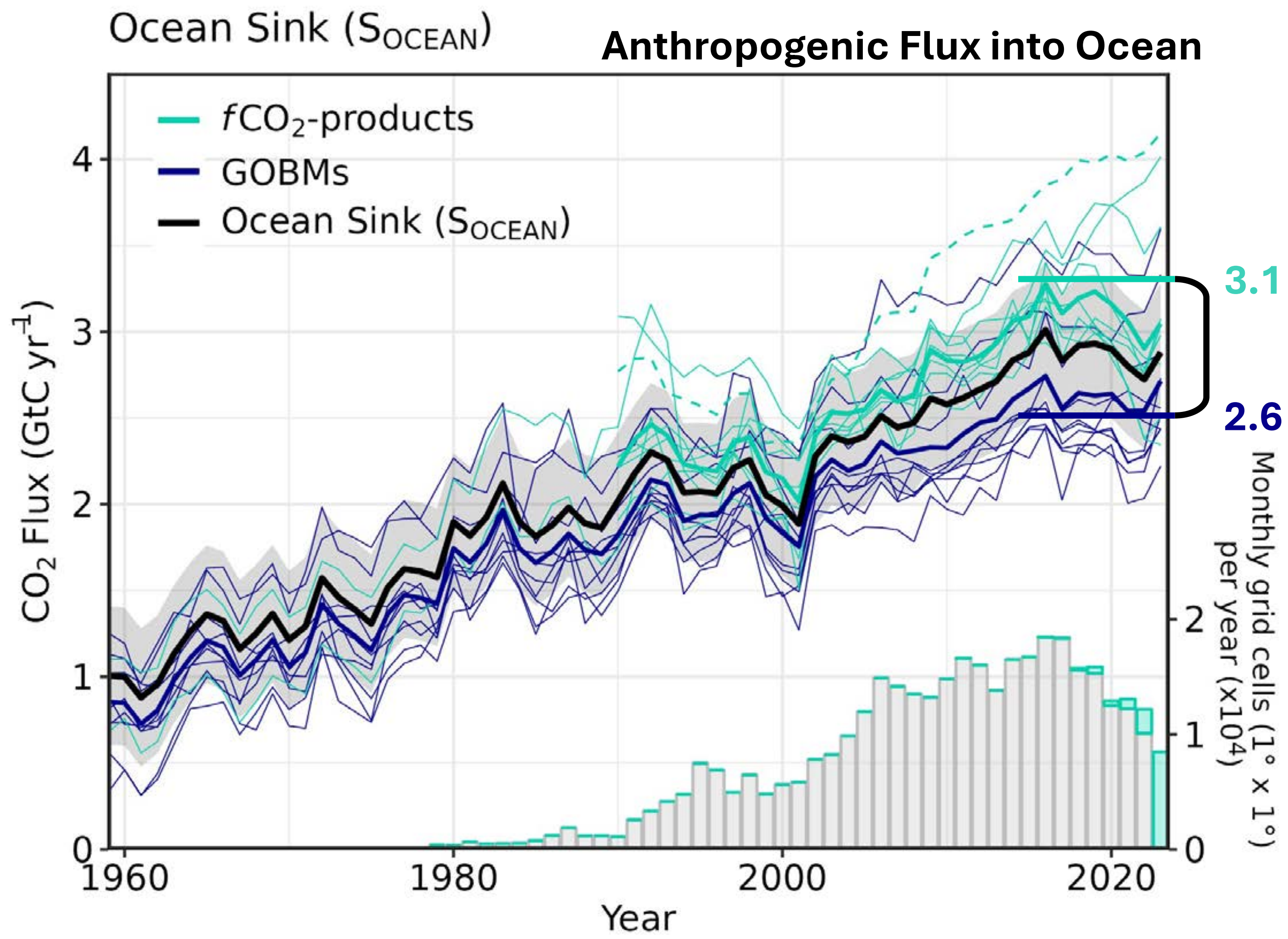
- Higher ocean temperatures result in greater stratification and **reduced supply of nutrients** to the euphotic zone (*Behrenfeld et al., 2006; Li et al., 2020*)
- Marine species respond directly to ocean temperature, especially in terms of **temperature-dependent oxygen tolerance** (*Pörtner, 2008; 2010*; *Deutsch et al., 2015*)
- **Ocean acidification** at the surface (*e.g., Jiang et al., 2023*) is detrimental to marine calcifiers and impacts $CaCO_3$ cycling and increased pCO_2 in the ocean interior (*Fassbender et al., 2023*) may influence other marine species



Gruber et al. (2021)

HOWEVER, THERE ARE SIGNIFICANT INTERMODEL AND MODEL-DATA DISCREPANCIES IN THESE CHANGES

Anthropogenic Flux into Ocean



3.1 GtC

Δ = 0.5 GtC per year discrepancy

2.6 GtC

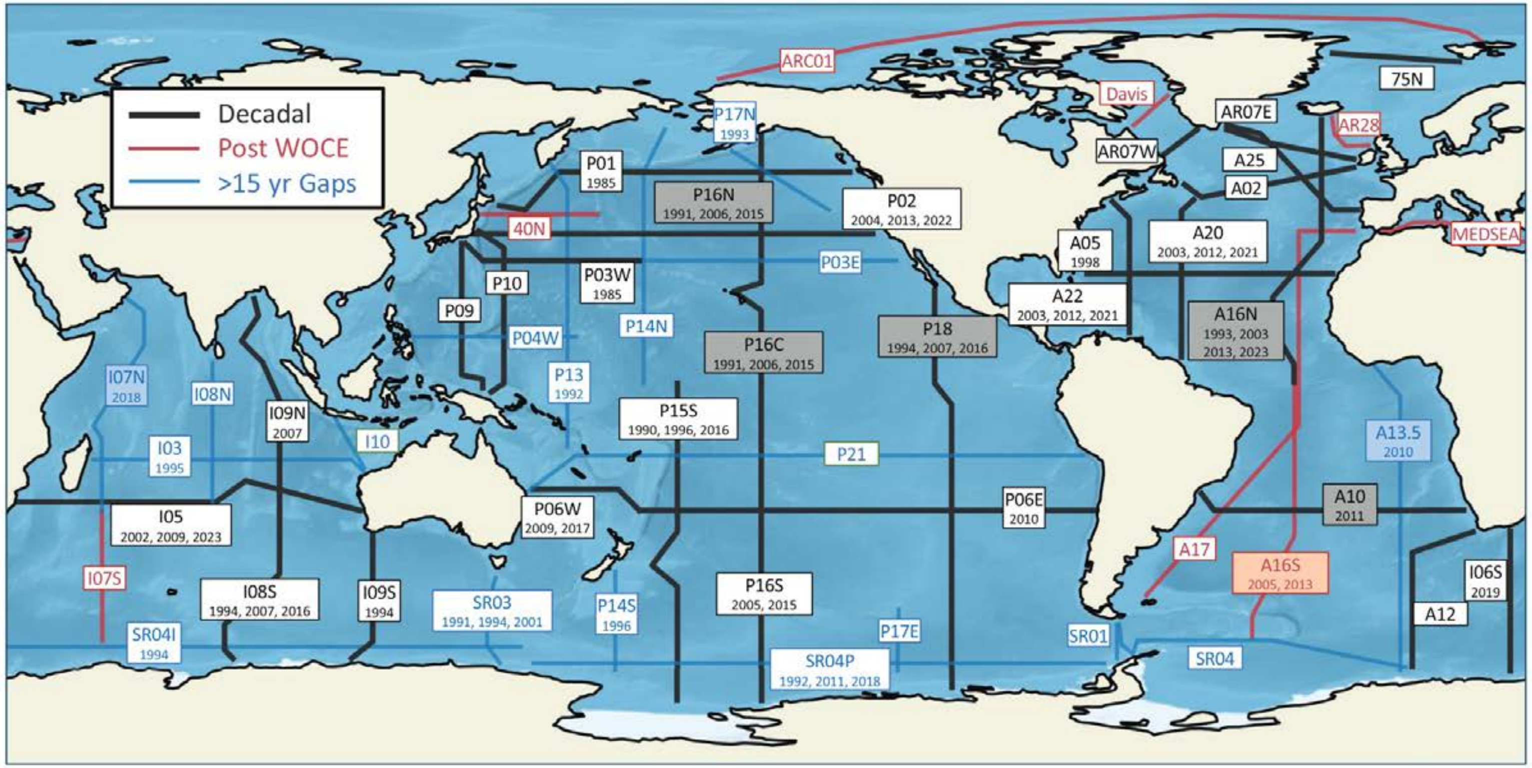
GLOBAL CARBON **BUDGET**

Compare to ~0.7 GtC per year cumulative annual emissions from the European Union

Friedlingstein et al. (2025)

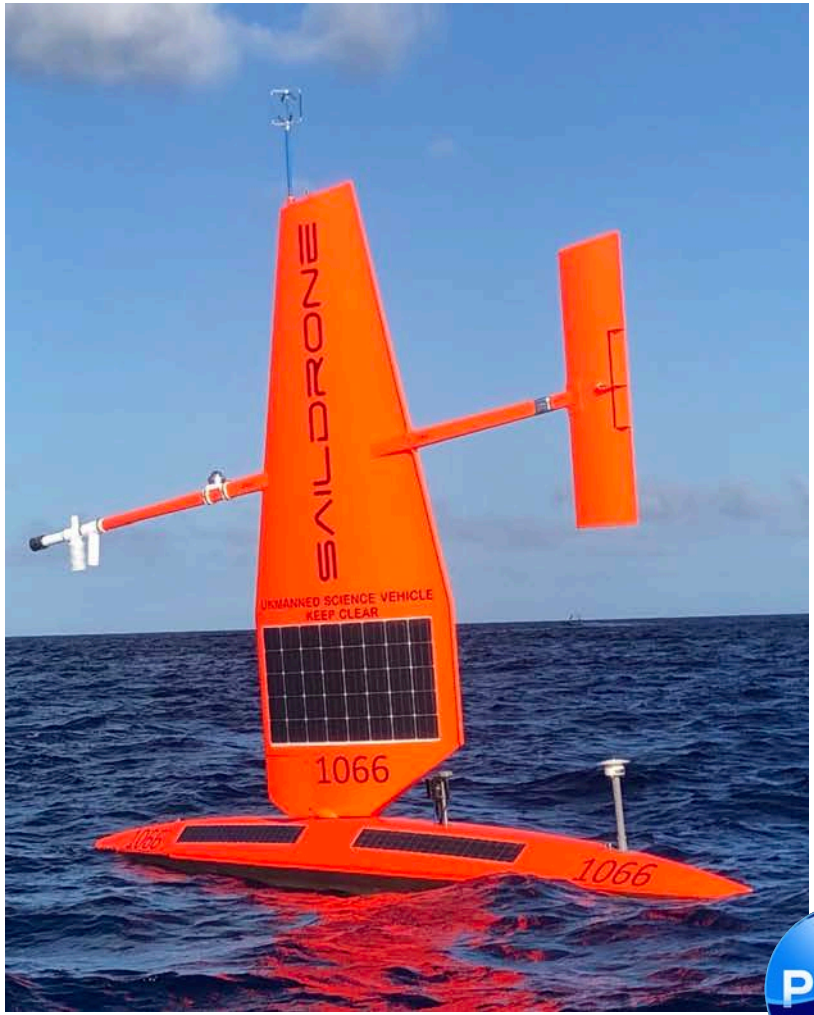
A FUNDAMENTAL SHIFT IN HOW WE MEASURE OCEAN INTERIOR PROPERTIES HAS PAVED THE WAY FOR NEW STRATEGIES TO INVESTIGATE GLOBAL OCEAN CHANGES FROM OBSERVATIONAL DATA

TRADITIONAL OCEAN OBSERVATIONS HAVE BEEN COLLECTED VIA DISCRETE SHIPBOARD SAMPLING OR SENSORS DEPLOYED ON CTD CASTS

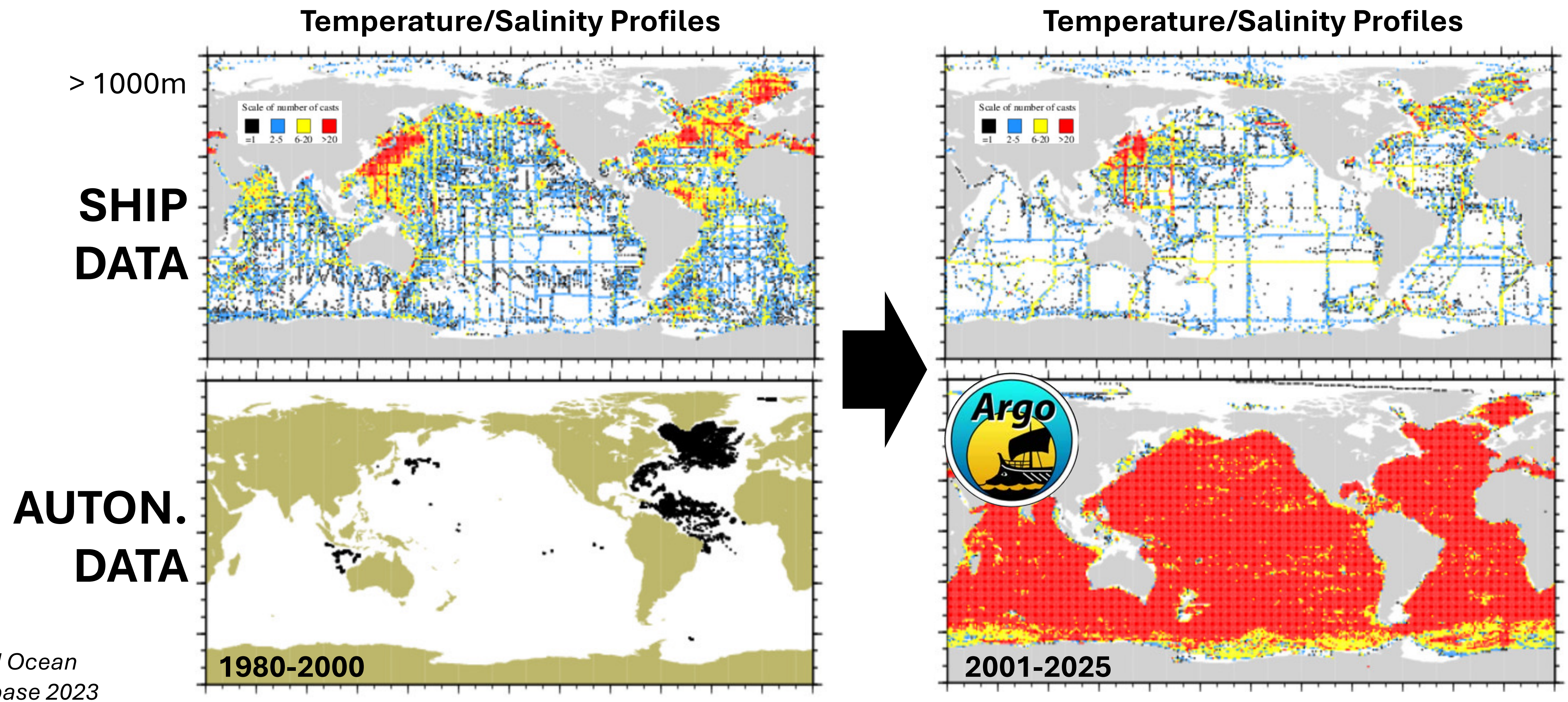


Erickson et al. (2023)

AUTONOMOUS IN SITU SENSORS OFFER A NEW WAY TO MONITOR OCEAN PROPERTIES



AUTONOMOUS SENSOR MEASUREMENTS NOW DOMINATE DISCRETE MEASUREMENTS IN GLOBAL OCEAN



World Ocean Database 2023

AUTONOMOUS SENSOR MEASUREMENTS NOW DOMINATE DISCRETE MEASUREMENTS IN GLOBAL OCEAN

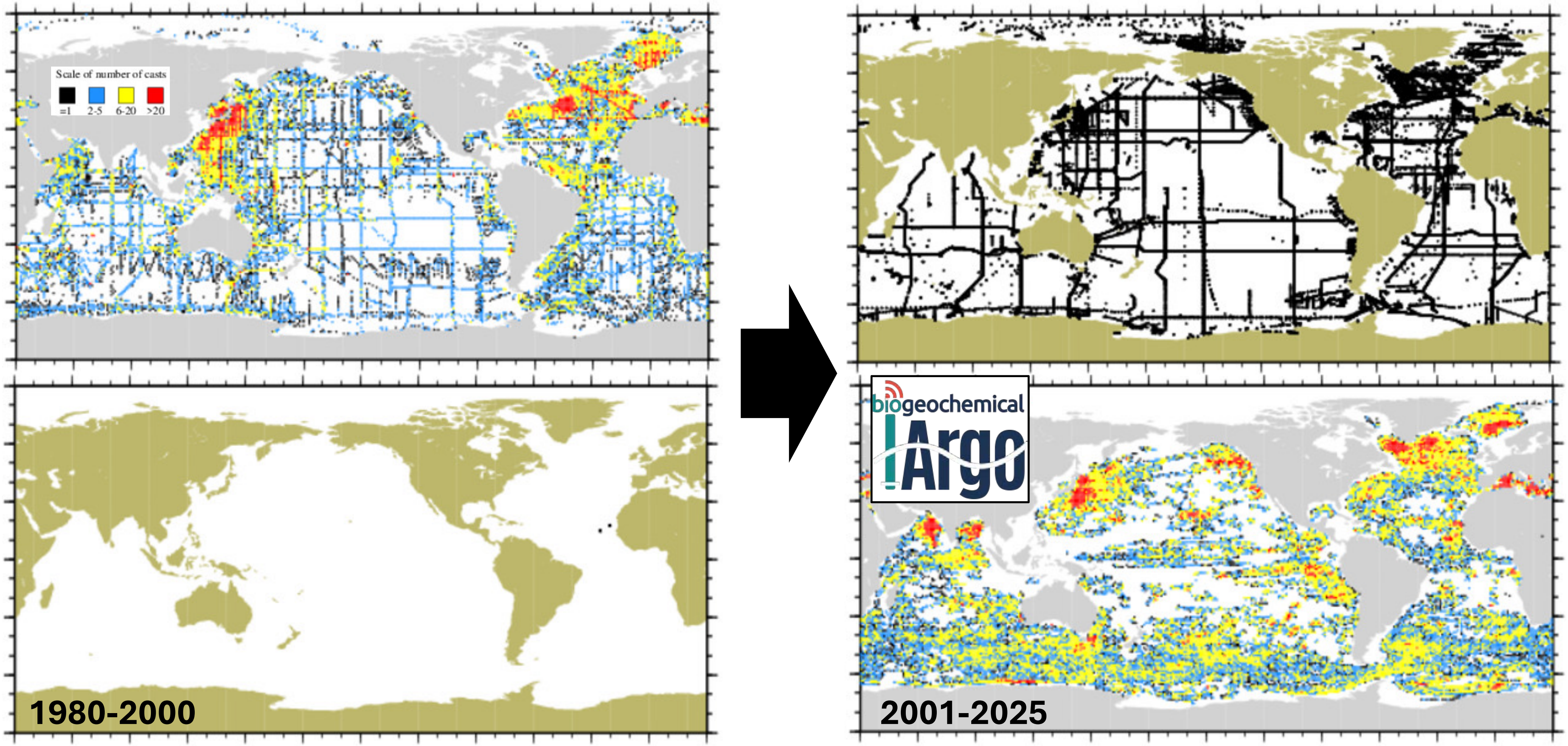
Temperature/Salinity/Oxygen Profiles

Temperature/Salinity/Oxygen Profiles

> 1000m

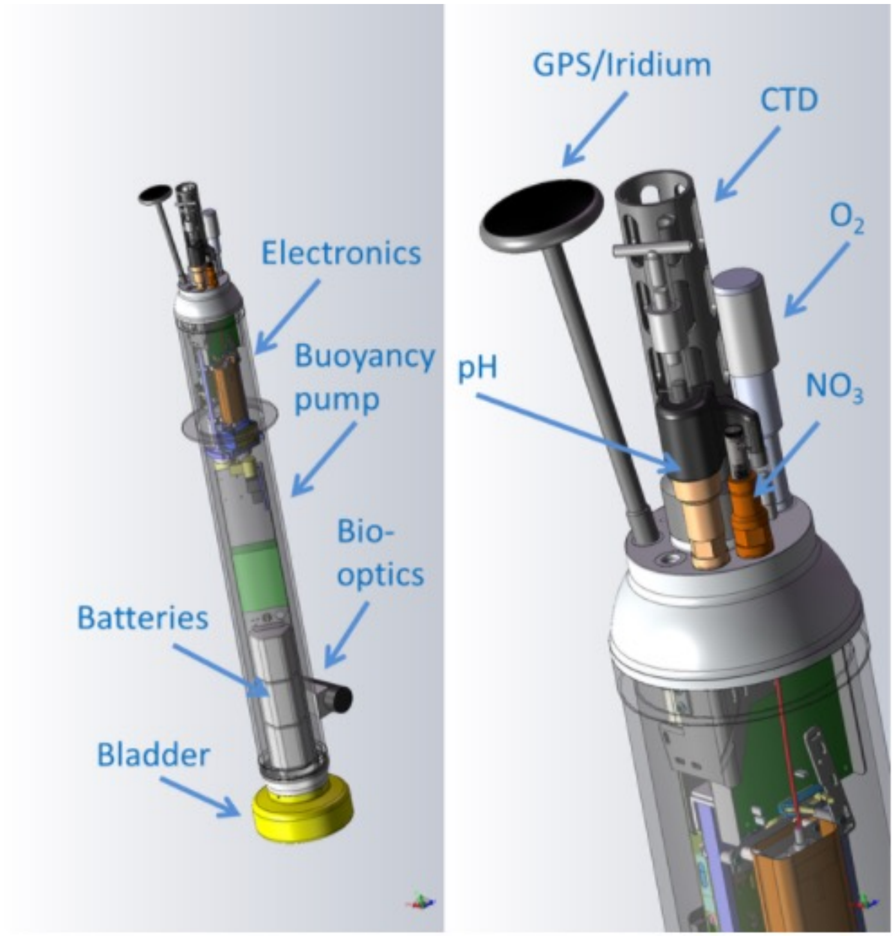
SHIP DATA

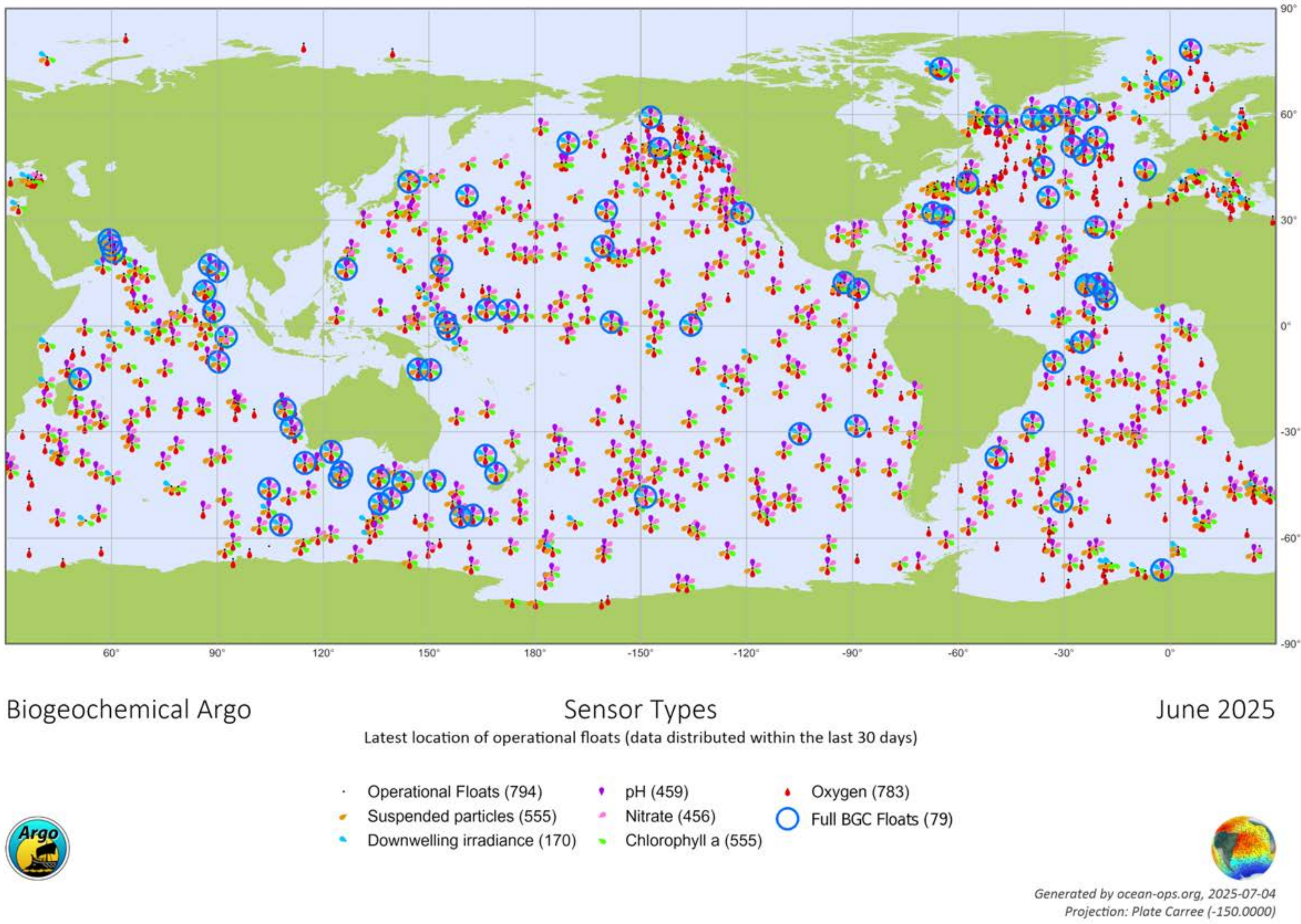
AUTON. DATA

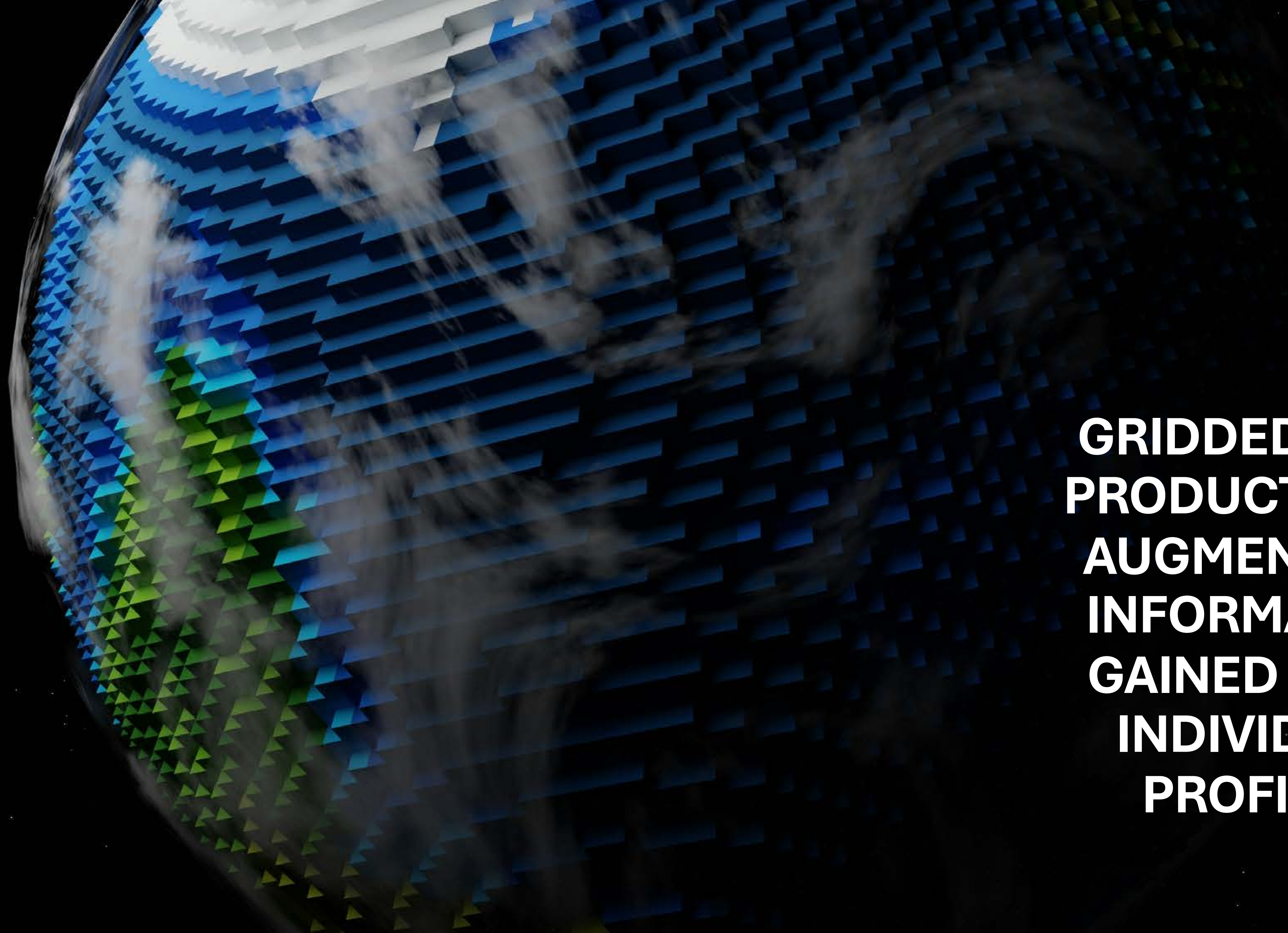


World Ocean Database 2023

ARGO FLOAT SENSORS



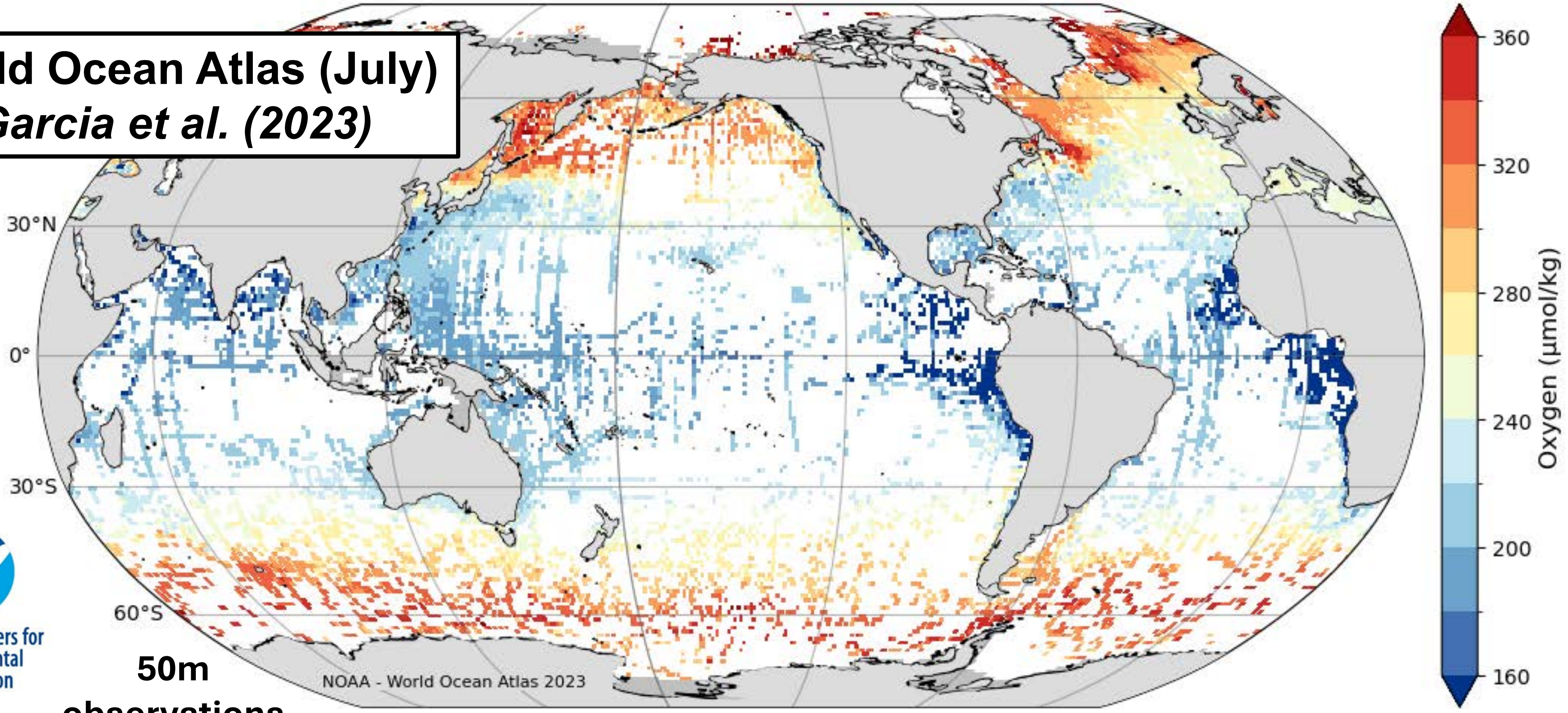




GRIDDED DATA PRODUCTS CAN AUGMENT THE INFORMATION GAINED FROM INDIVIDUAL PROFILES

THE TRANSLATION OF SCATTERED OBSERVATIONS TO REGULAR MAPS PROMOTES WIDESPREAD USE OF OBSERVATIONAL DATASETS

World Ocean Atlas (July)
Garcia et al. (2023)

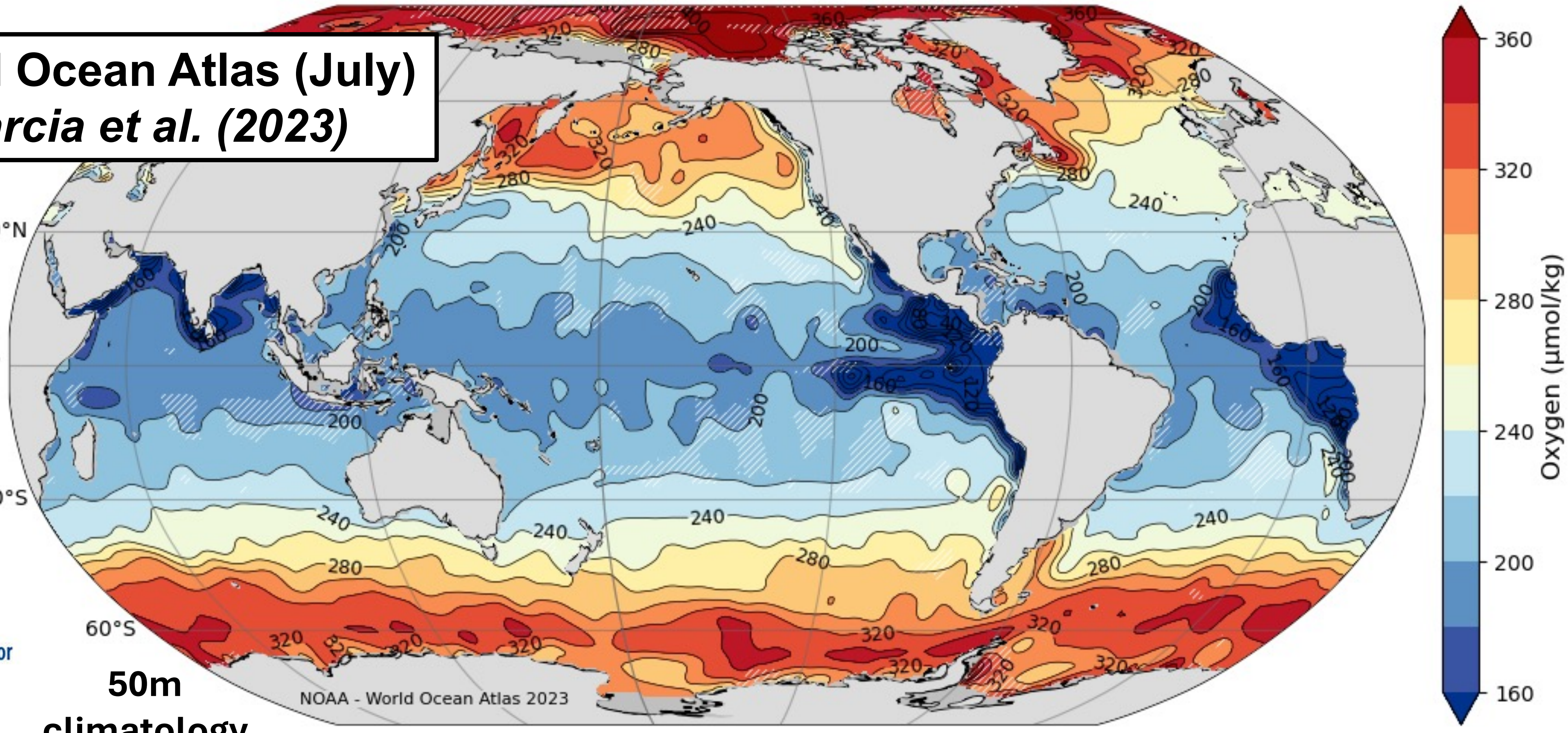


National Centers for Environmental Information

50m observations

THE TRANSLATION OF SCATTERED OBSERVATIONS TO REGULAR MAPS PROMOTES WIDESPREAD USE OF OBSERVATIONAL DATASETS

World Ocean Atlas (July)
Garcia et al. (2023)

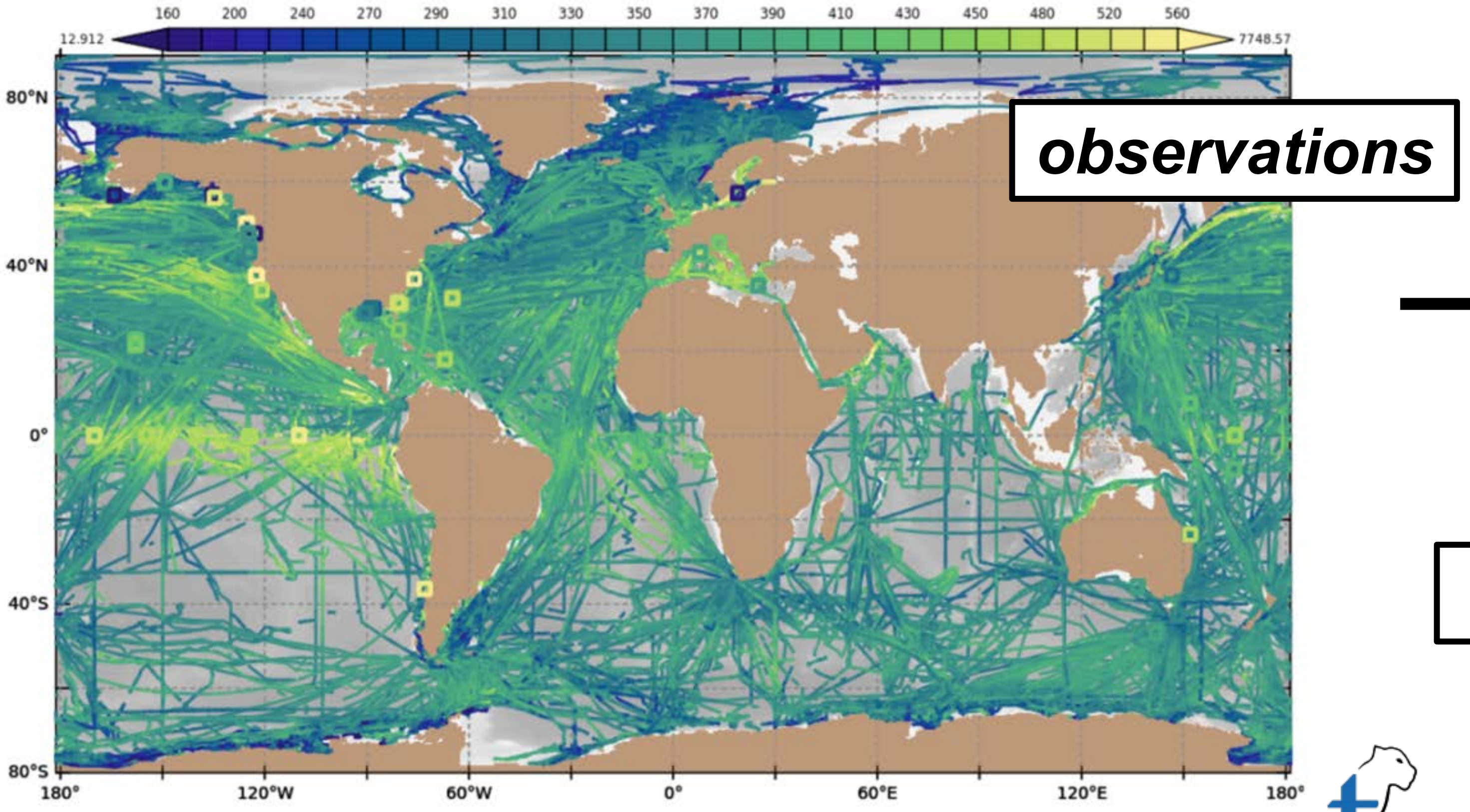


50m climatology

National Centers for Environmental Information

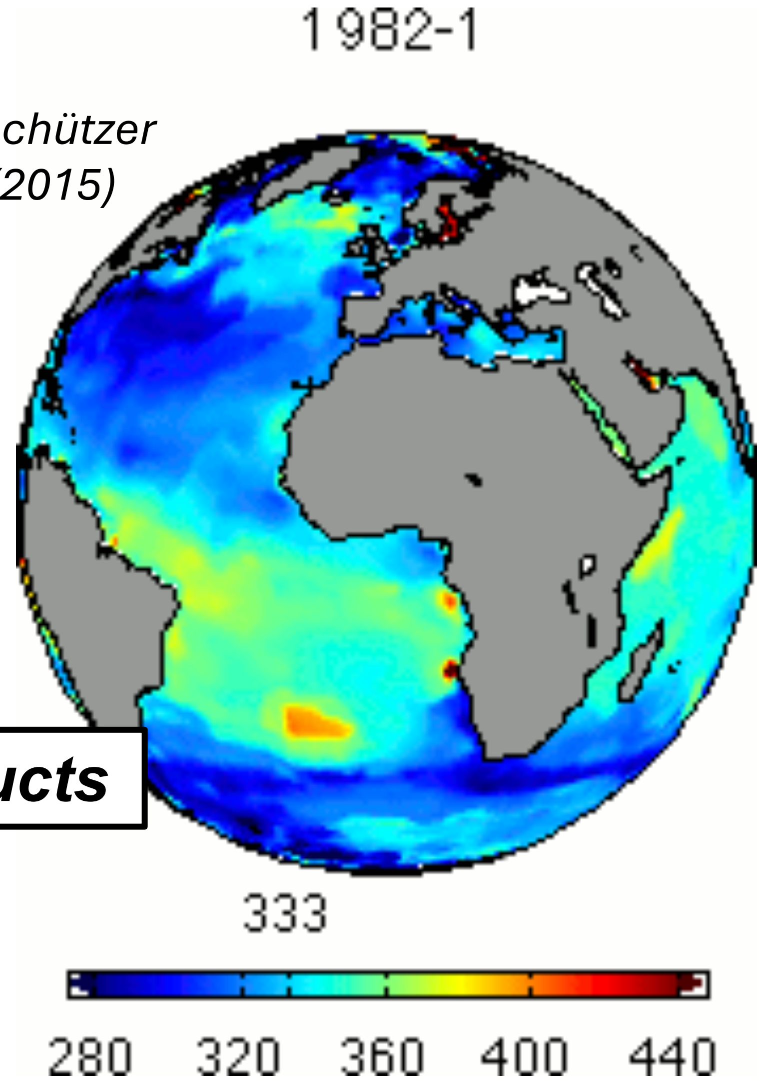
THE TRANSLATION OF SCATTERED OBSERVATIONS TO REGULAR MAPS PROMOTES WIDESPREAD USE OF OBSERVATIONAL DATASETS

Surface Ocean CO_2 data products built using machine learning contribute to Global Carbon Budget

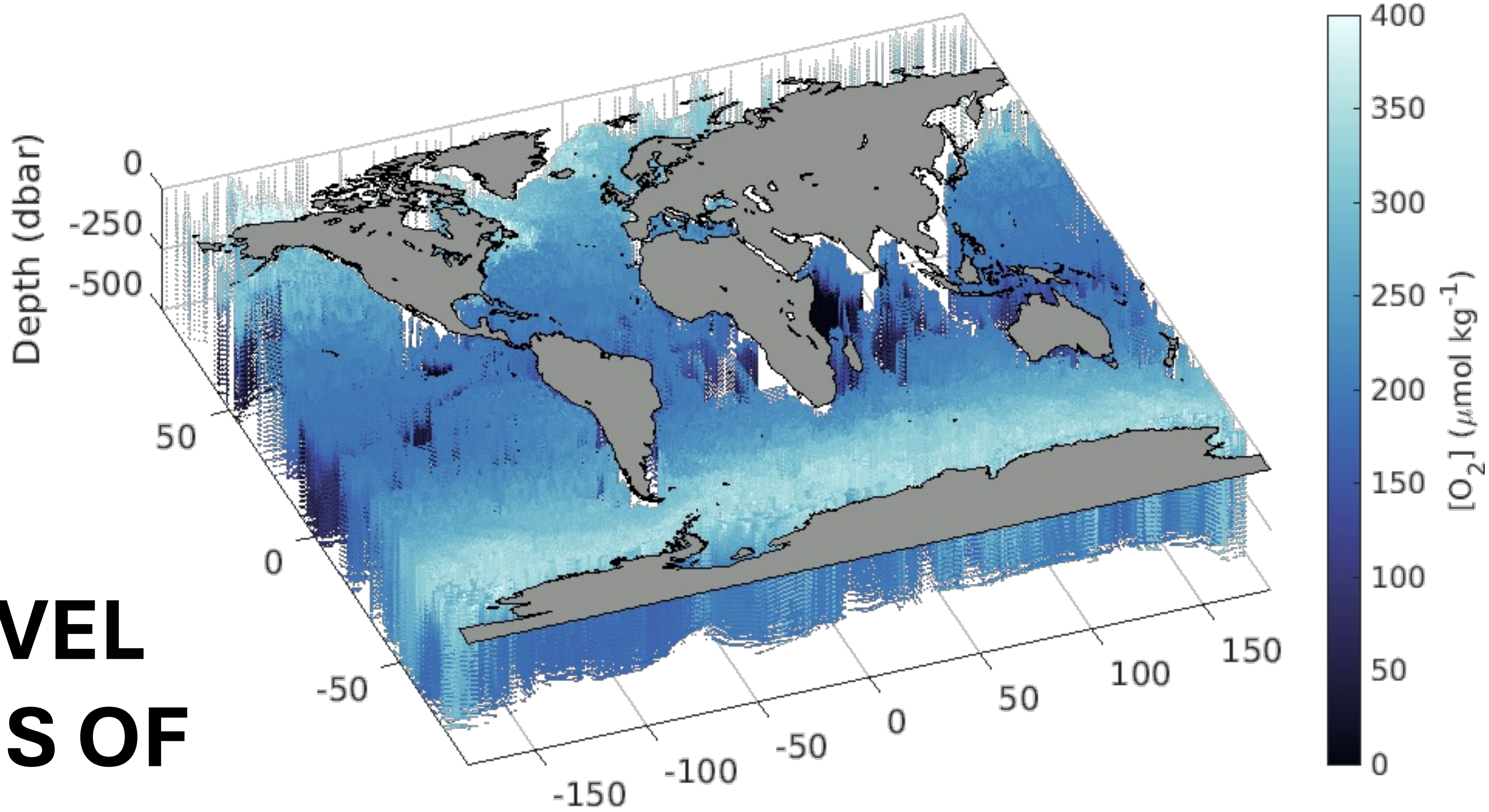


Bakker et al. (2016)

SOCAT · SURFACE OCEAN CO_2 ATLAS ·



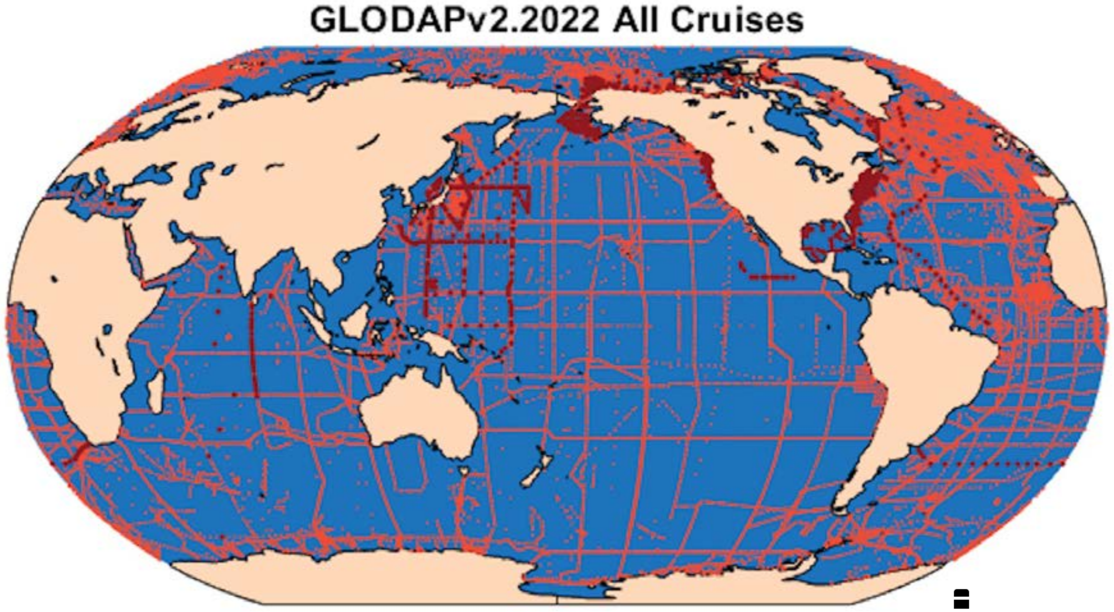
Landschützer et al. (2015)



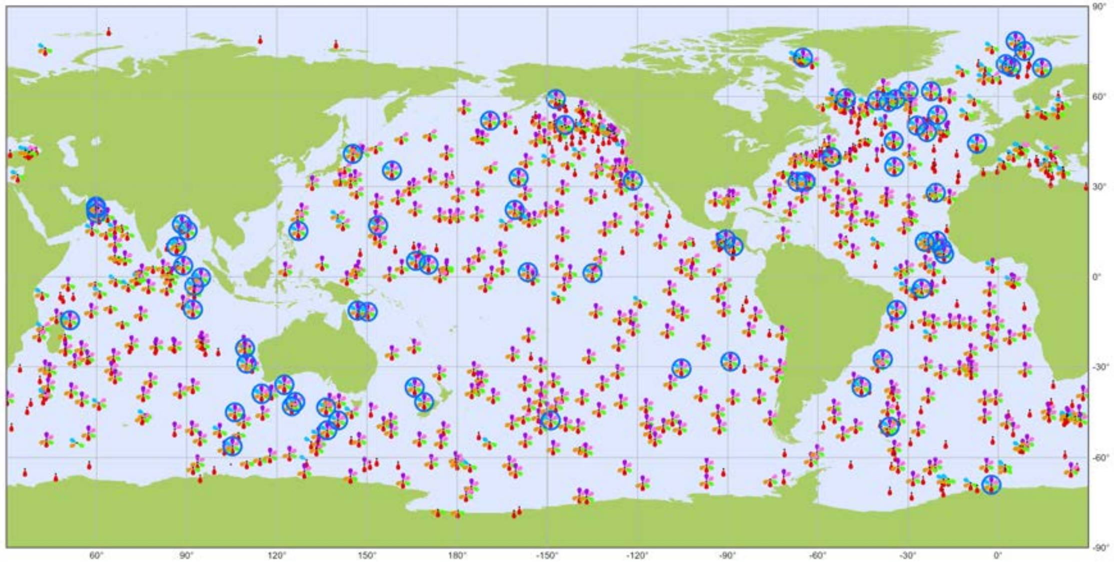
HOW CAN WE LEVERAGE NOVEL OBSERVATIONS OF OCEAN BIOGEOCHEMICAL PROPERTIES TO CREATE USEFUL, TIME-VARYING DATA PRODUCTS?

Sharp et al. (2023)
Sharp et al. (in prep)

OBSERVATIONAL DATA SOURCES

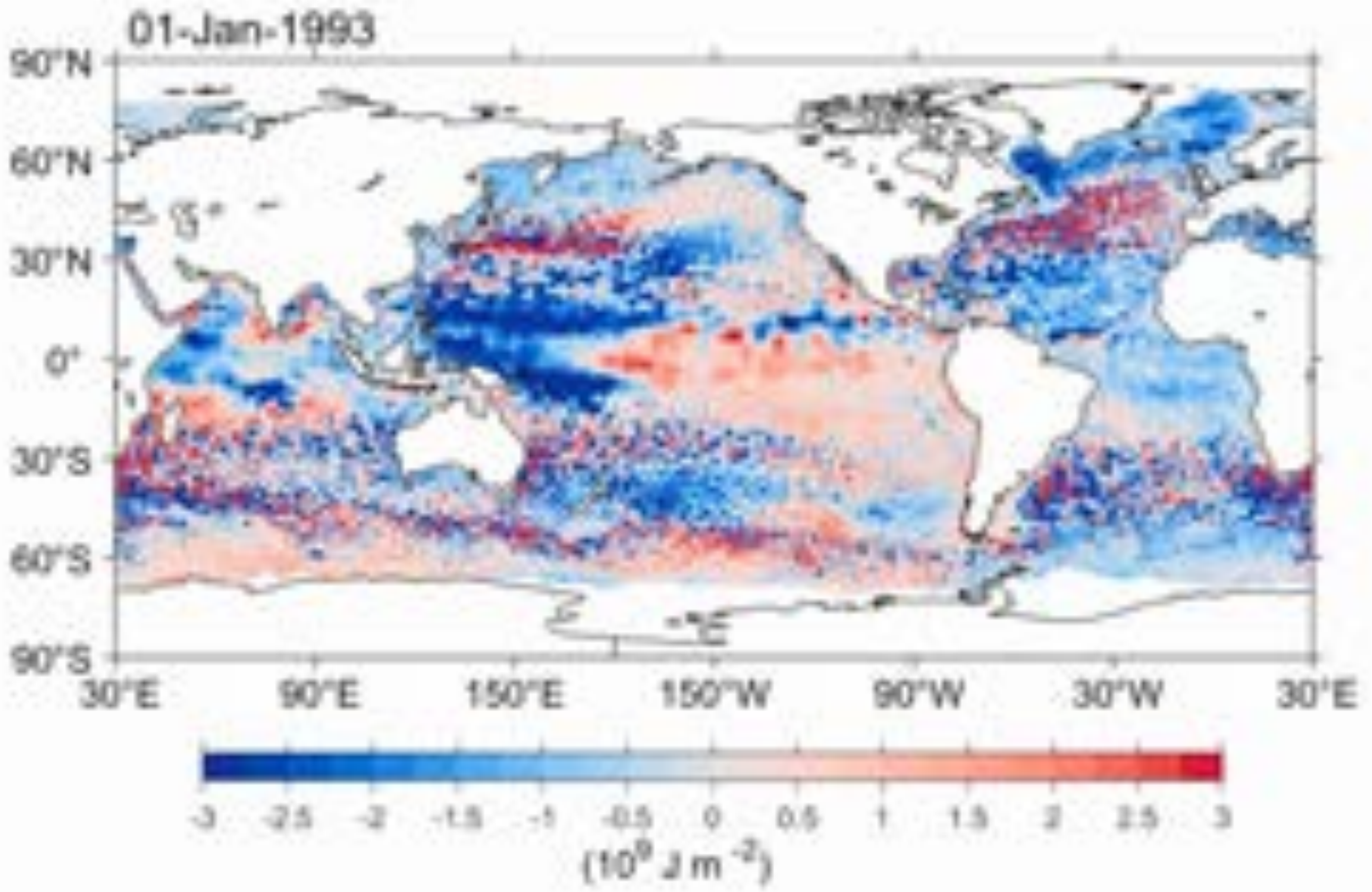


Global Ocean Data Analysis Project (GLODAP)

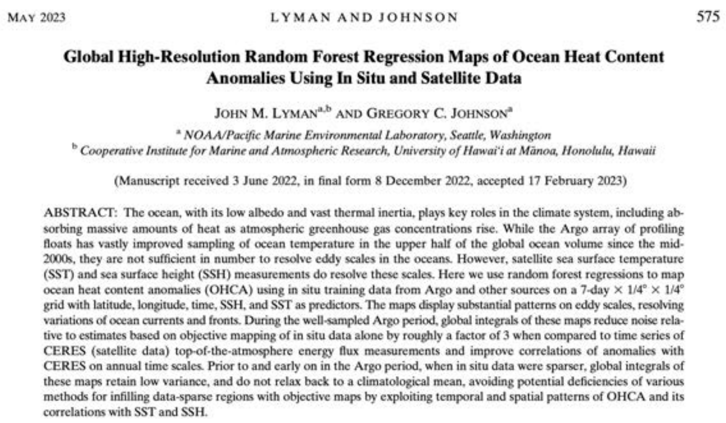


Biogeochemical Argo (BGC-Argo)

GRIDDED DATA SOURCE



- *Lyman and Johnson (2023)* introduce high resolution heat content maps (RFROM)
- We leverage the temp. and sal. RFROM ocean maps from *Lyman and Johnson (in prep)*
- This produces biogeochemical maps on 0.25°, weekly scales



MAY 2023 LYMAN AND JOHNSON 575

Global High-Resolution Random Forest Regression Maps of Ocean Heat Content Anomalies Using In Situ and Satellite Data

JOHN M. LYMAN[a,b] AND GREGORY C. JOHNSON[a]

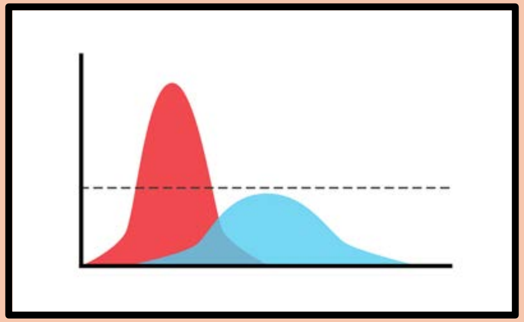
[a] *NOAA/Pacific Marine Environmental Laboratory, Seattle, Washington*
[b] *Cooperative Institute for Marine and Atmospheric Research, University of Hawai'i at Mānoa, Honolulu, Hawaii*

(Manuscript received 3 June 2022, in final form 8 December 2022, accepted 17 February 2023)

ABSTRACT: The ocean, with its low albedo and vast thermal inertia, plays key roles in the climate system, including absorbing massive amounts of heat as atmospheric greenhouse gas concentrations rise. While the Argo array of profiling floats has vastly improved sampling of ocean temperature in the upper half of the global ocean volume since the mid-2000s, they are not sufficient in number to resolve eddy scales in the oceans. However, satellite sea surface temperature (SST) and sea surface height (SSH) measurements do resolve these scales. Here we use random forest regressions to map ocean heat content anomalies (OHCA) using in situ training data from Argo and other sources on a 7-day × 1/4° × 1/4° grid with latitude, longitude, time, SSH, and SST as predictors. The maps display substantial patterns on eddy scales, resolving variations of ocean currents and fronts. During the well-sampled Argo period, global integrals of these maps reduce noise relative to estimates based on objective mapping of in situ data alone by roughly a factor of 3 when compared to time series of CERES (satellite data) top-of-the-atmosphere energy flux measurements and improve correlations of anomalies with CERES on annual time scales. Prior to and early on in the Argo period, when in situ data were sparser, global integrals of these maps retain low variance, and do not relax back to a climatological mean, avoiding potential deficiencies of various methods for infilling data-sparse regions with objective maps by exploiting temporal and spatial patterns of OHCA and its correlations with SST and SSH.

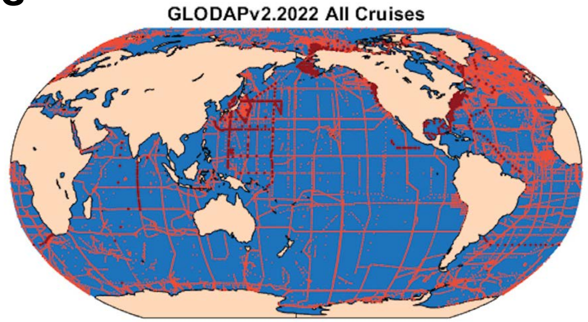
OBJECTIVE CLUSTERING OF DATA

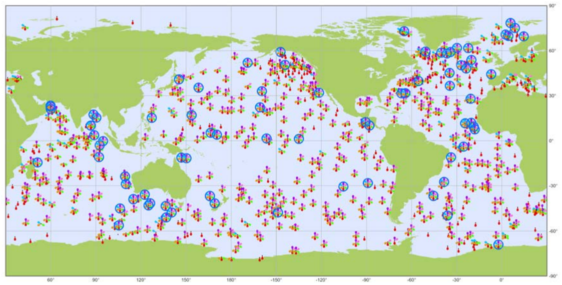
Gaussian Mixture Model with 15 Groups



Trained on Temperature, Salinity, and Depth

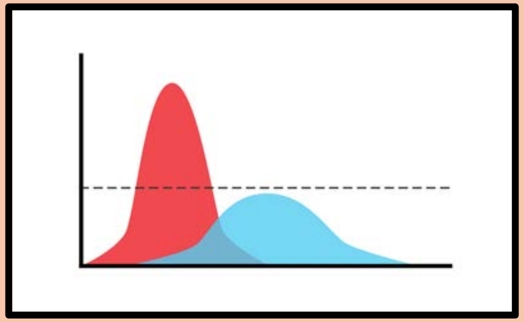
Profile Data



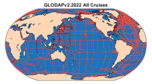


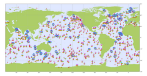
OBJECTIVE CLUSTERING OF DATA

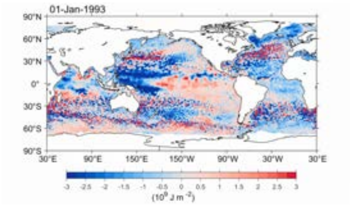
Gaussian Mixture Model with 15 Groups

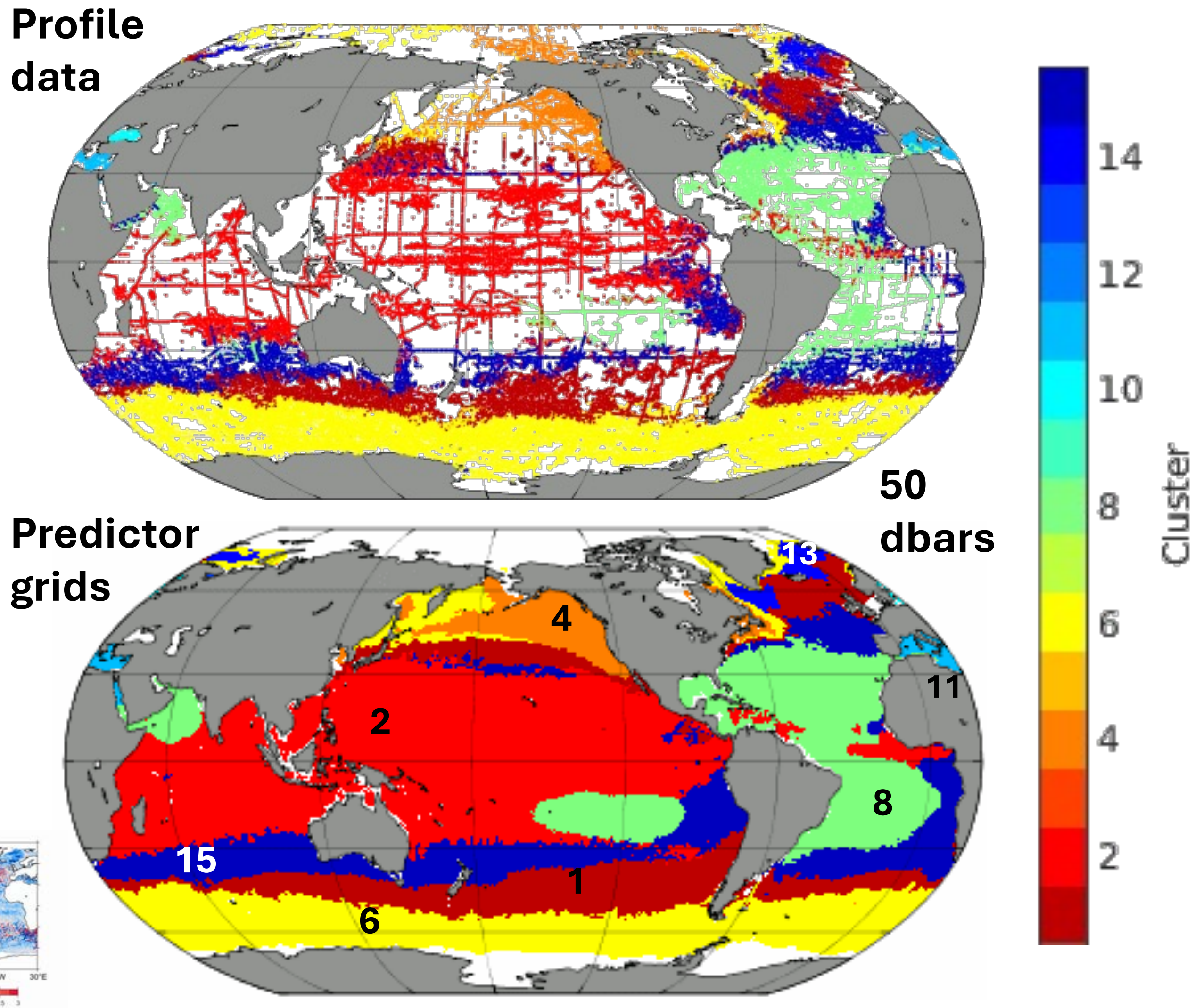


Trained on Temperature, Salinity, and Depth



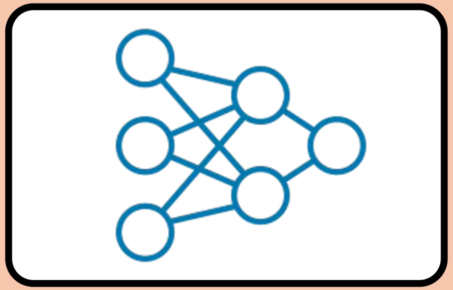






OBJECTIVE CLUSTERING OF DATA

Shallow Neural Network Committee

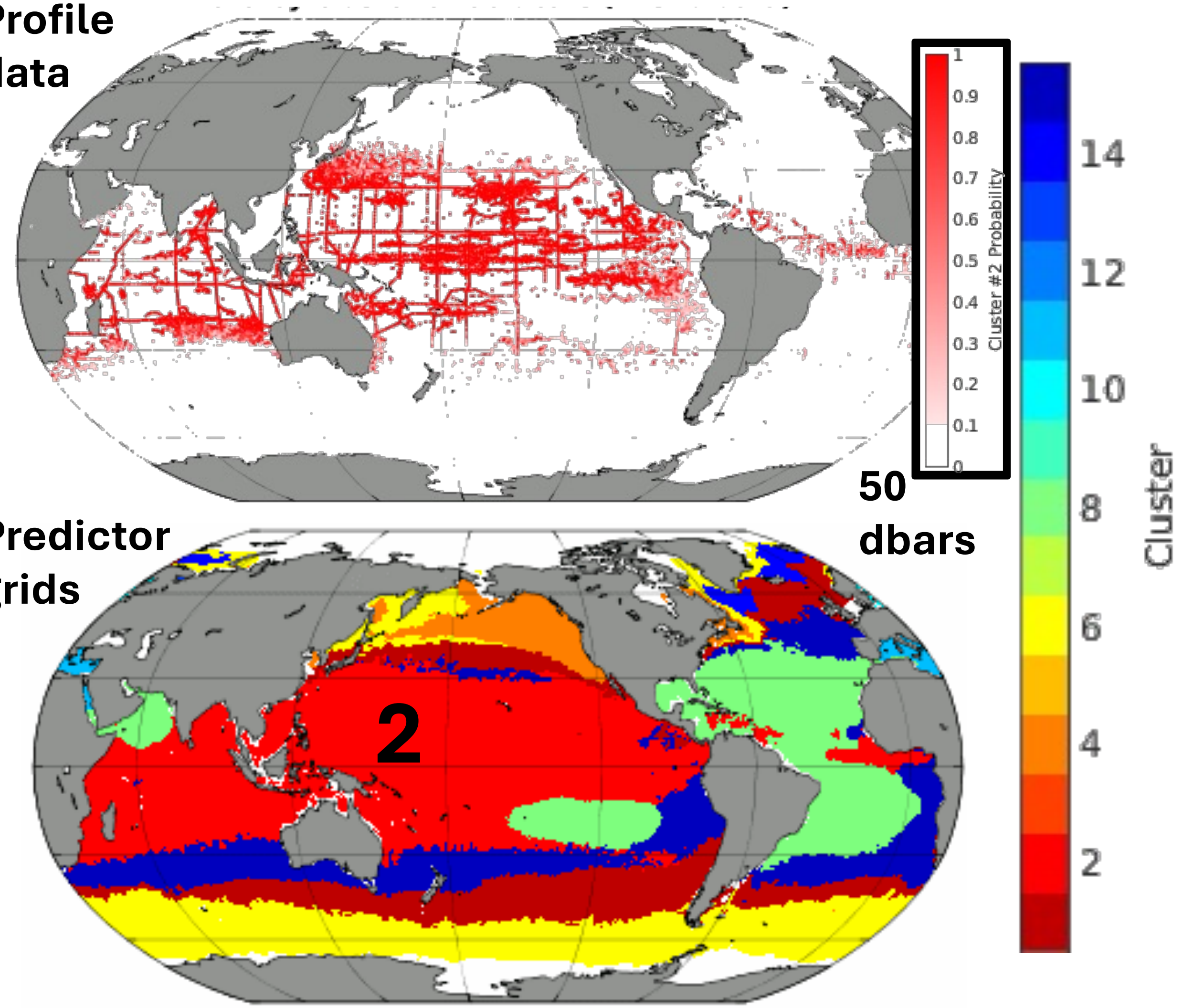


Trained on Temperature, Salinity, Potential Density, and Spatiotemporal Coordinates

Profile data

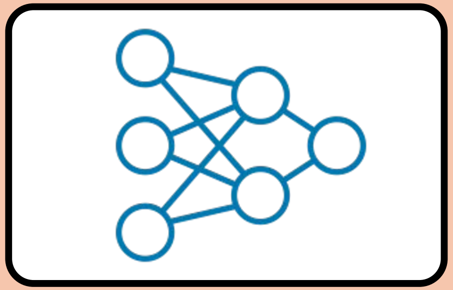
Predictor grids

50 dbars



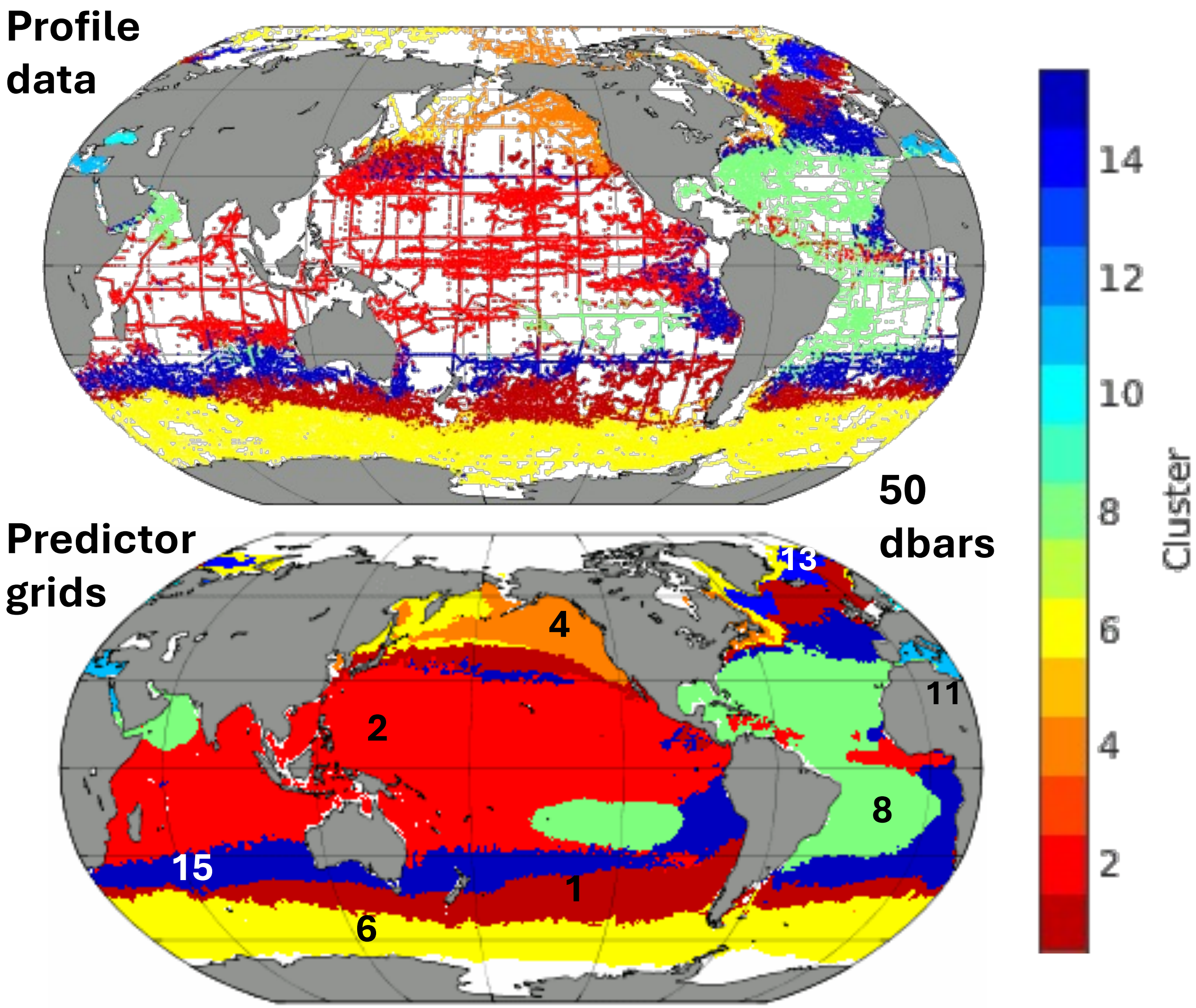
OBJECTIVE CLUSTERING OF DATA

Shallow Neural Network Committee



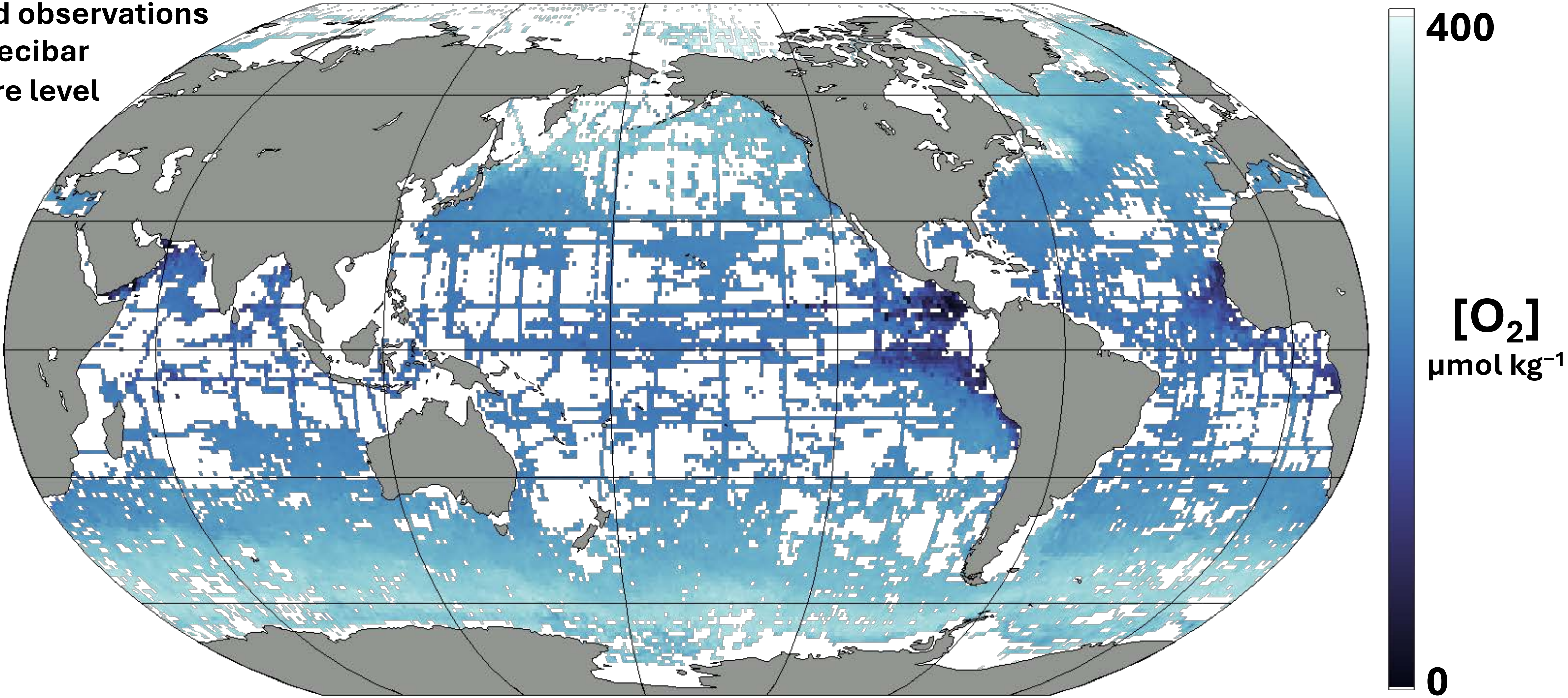
Trained on Temperature, Salinity, Potential Density, and Spatiotemporal Coordinates

15x



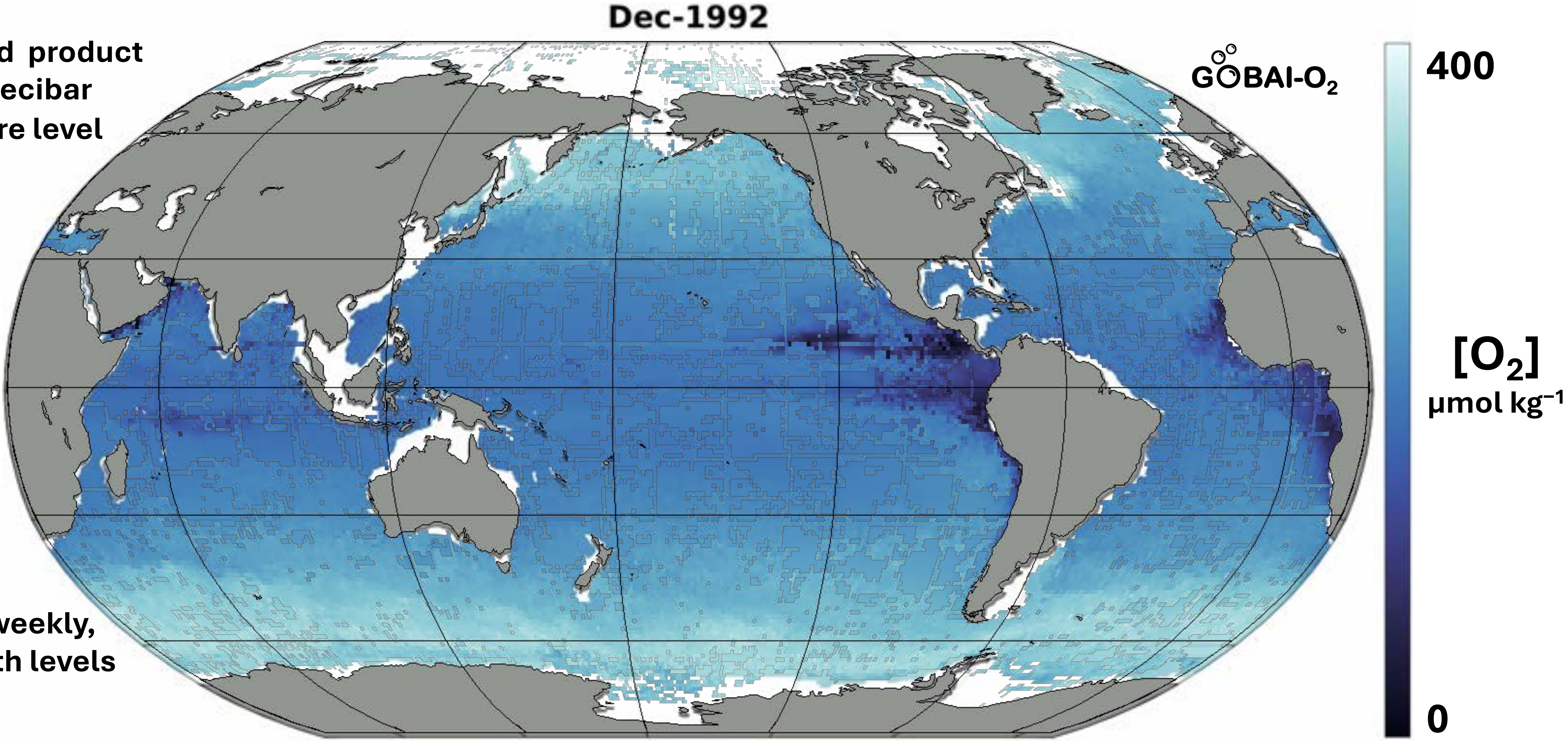
THIS ALLOWS US TO CREATE MAPPED, TIME-VARYING DATA PRODUCT FROM EMPIRICAL MODELS TRAINED ON OBSERVATIONS

Gridded observations on 50 decibar pressure level



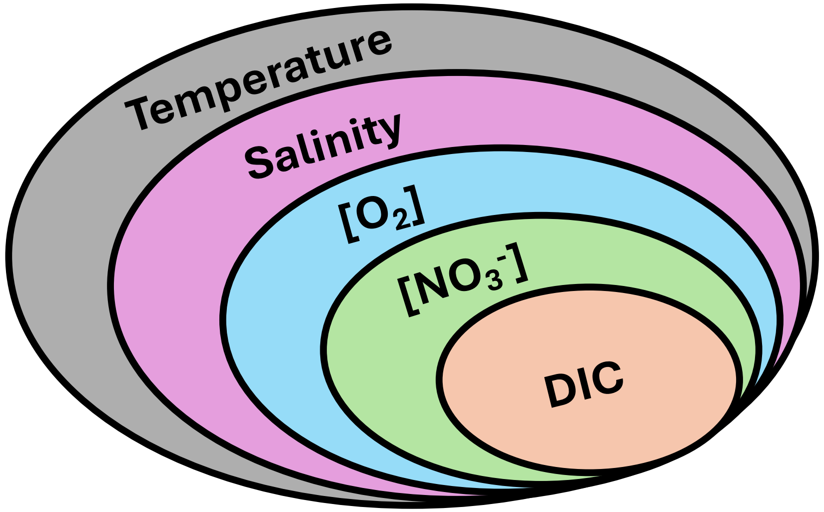
THIS ALLOWS US TO CREATE MAPPED, TIME-VARYING DATA PRODUCT FROM EMPIRICAL MODELS TRAINED ON OBSERVATIONS

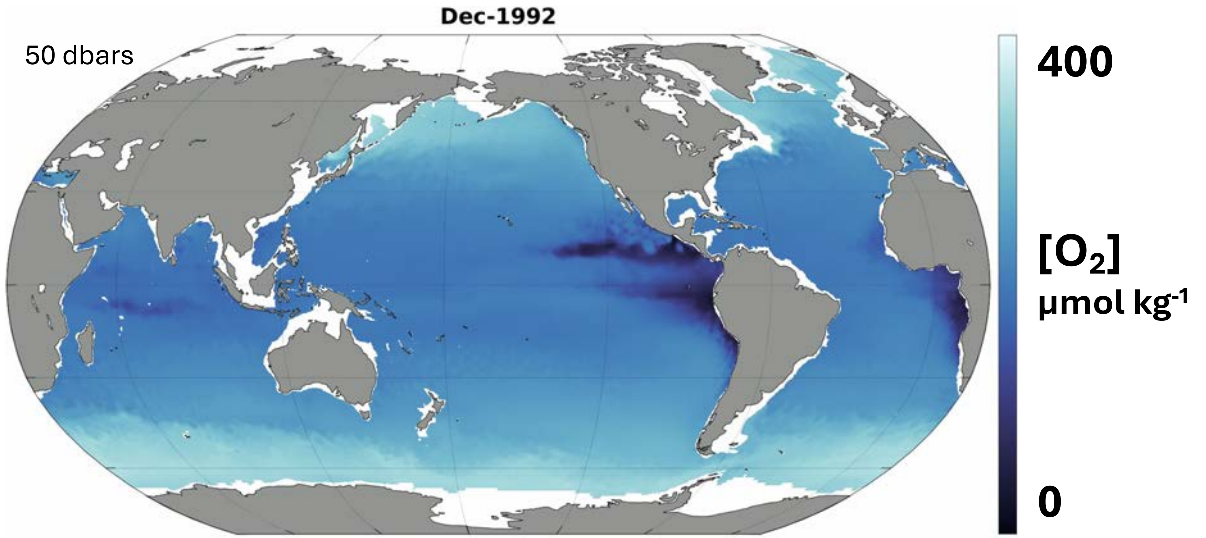
Mapped product on 50 decibar pressure level

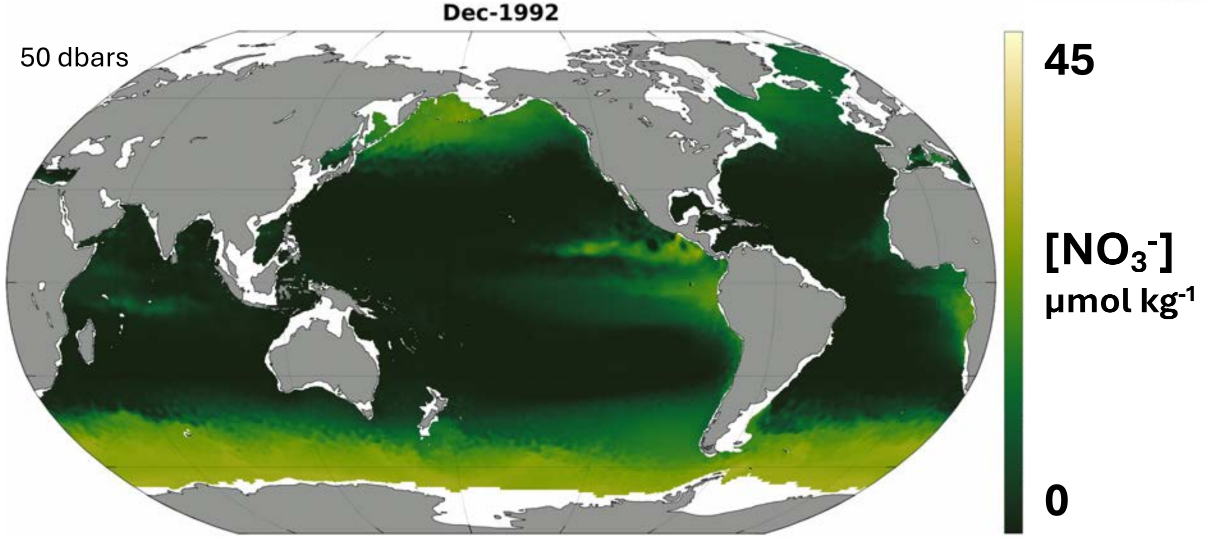


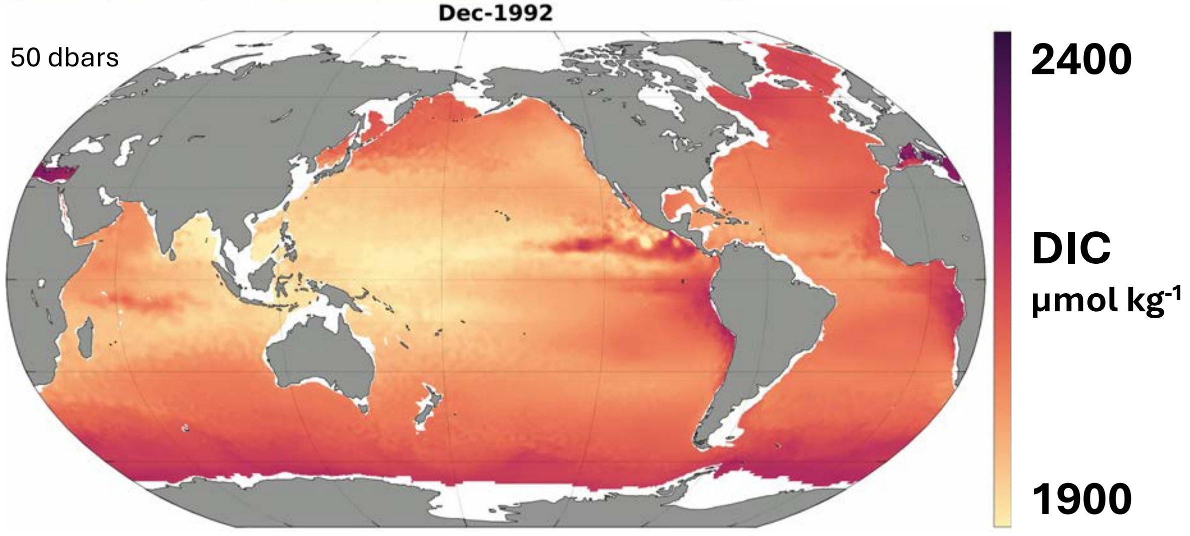
0.25°, weekly,
58 depth levels

REPEAT PROCESS FOR NITRATE CONCENTRATION AND DISSOLVED INORGANIC CARBON



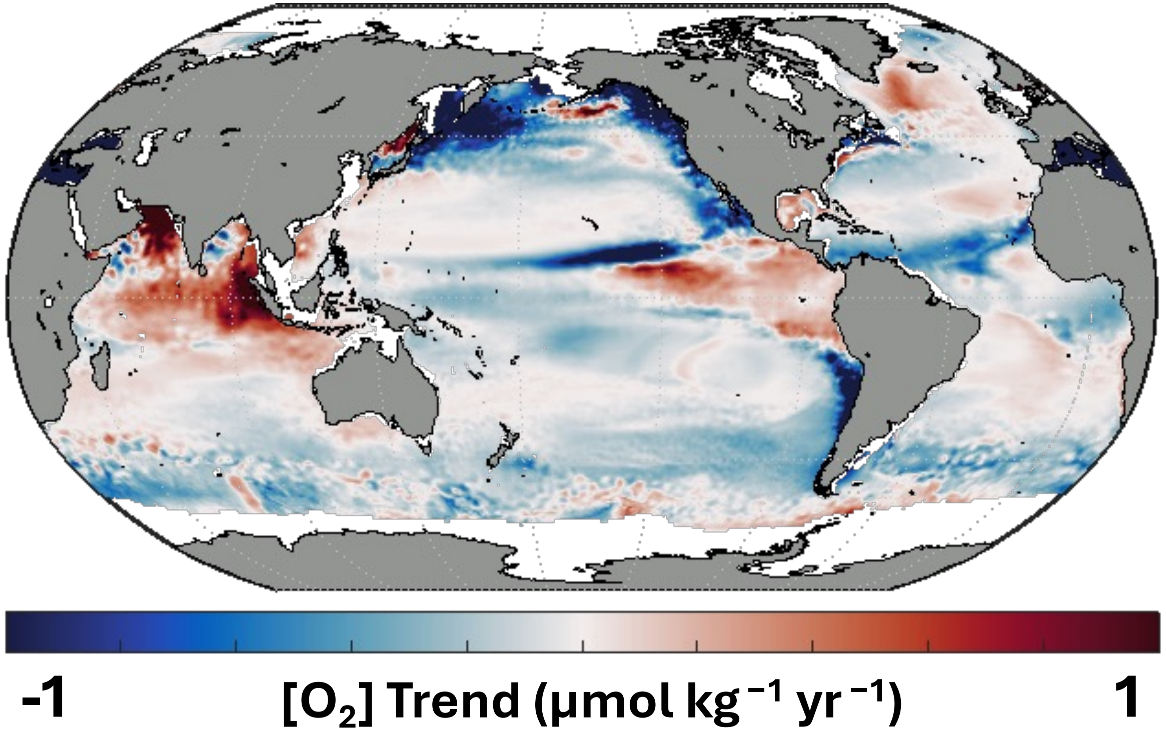




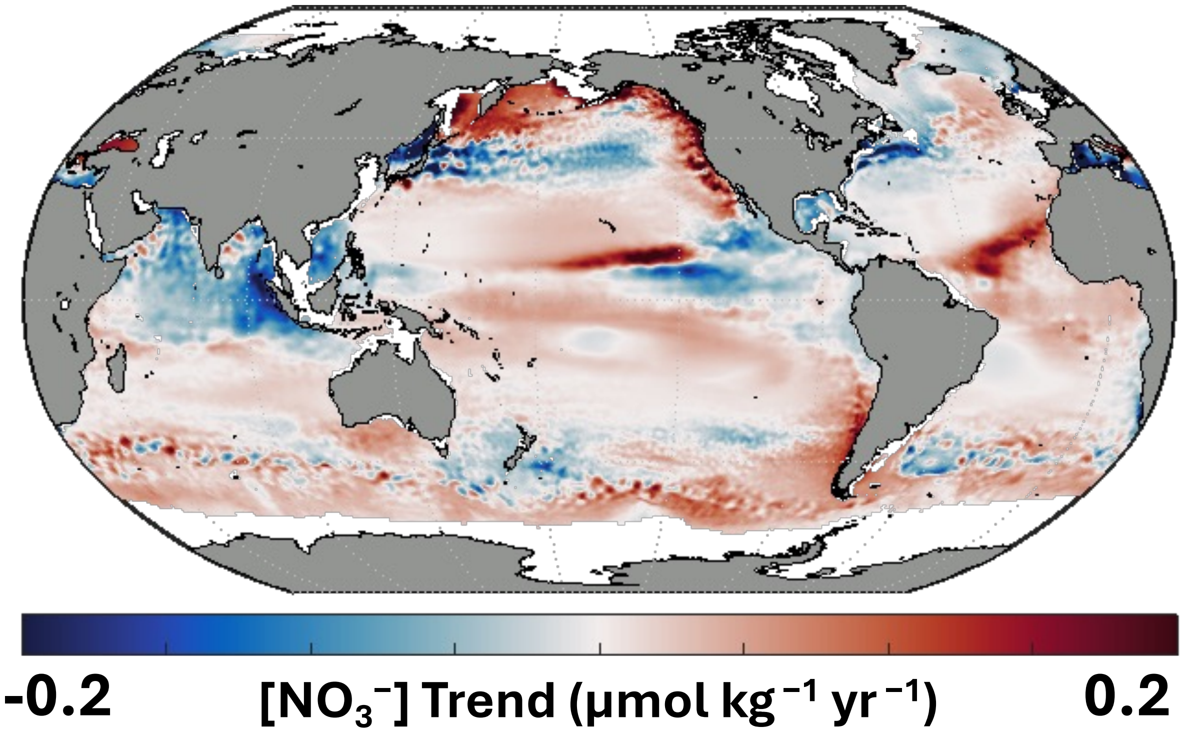


OXYGEN AND NITRATE DISPLAY COHERENT TRENDS

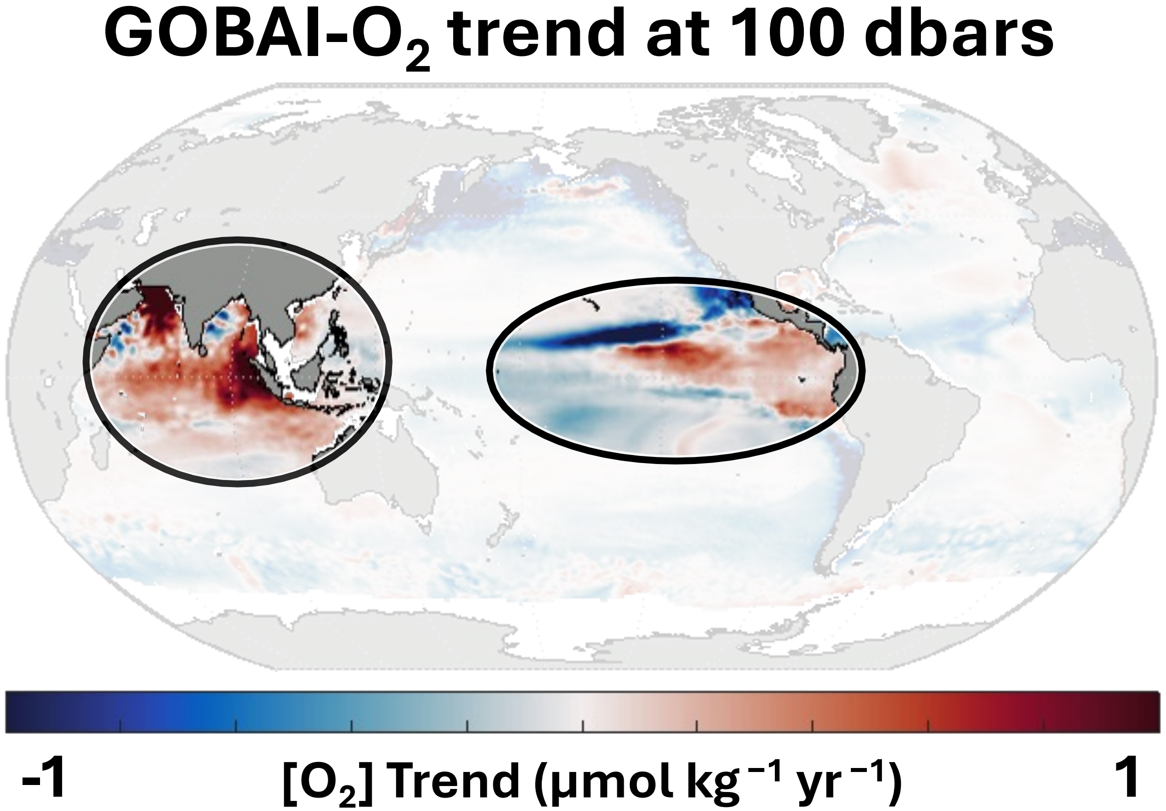
GOBAI-O_2 trend at 100 dbars

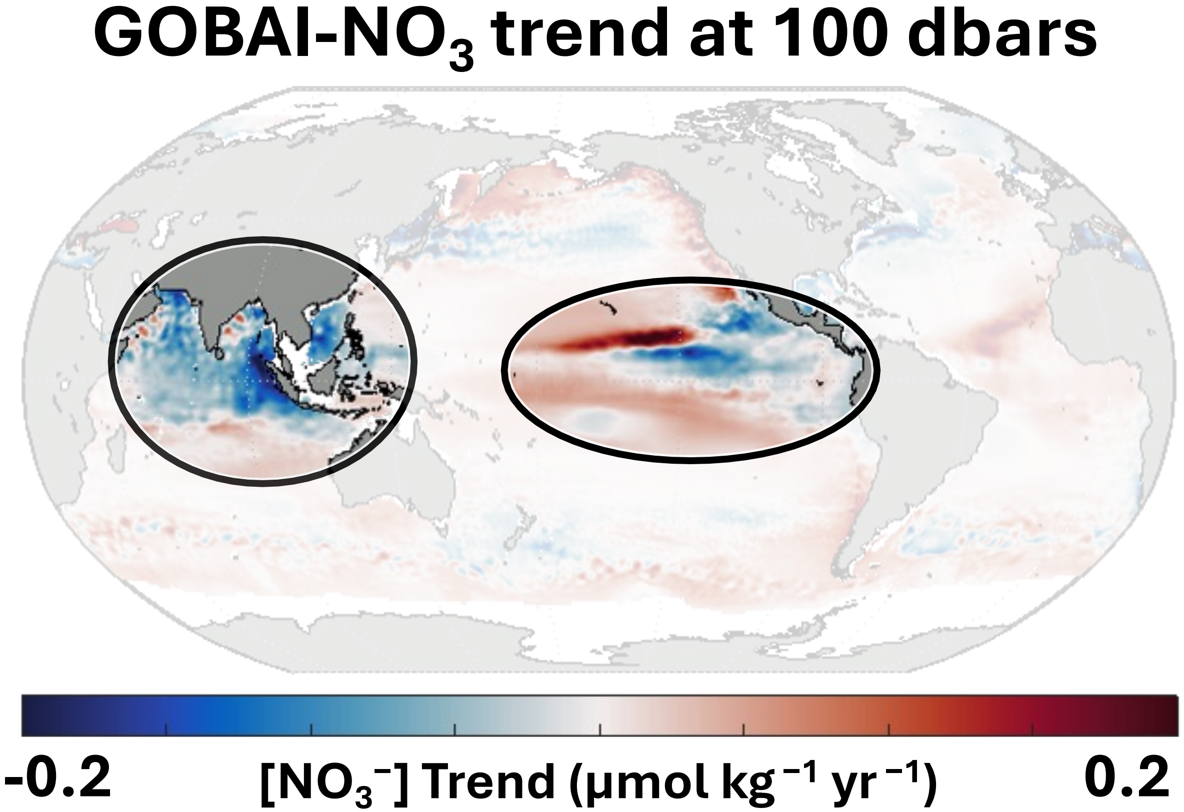


GOBAI-NO_3 trend at 100 dbars

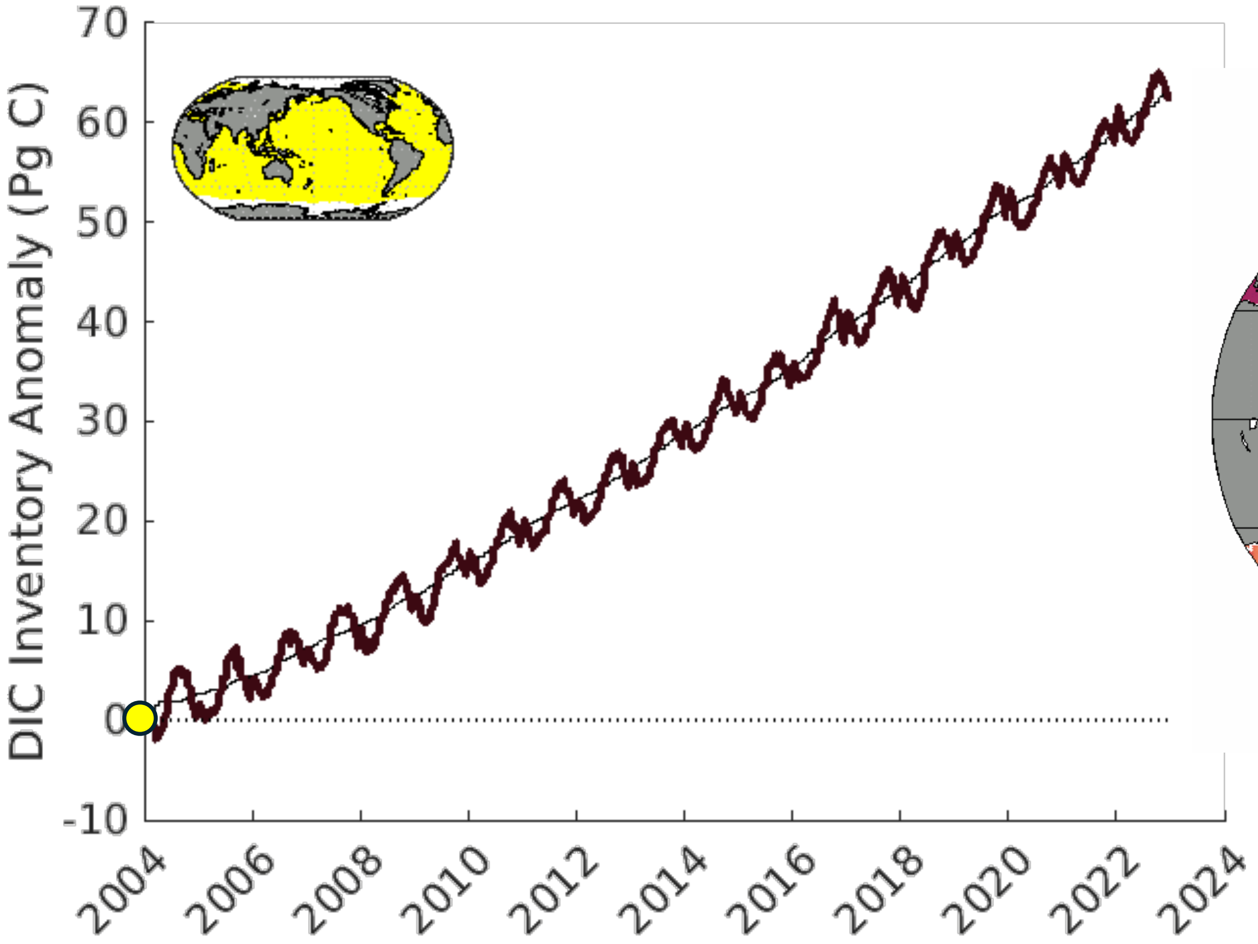


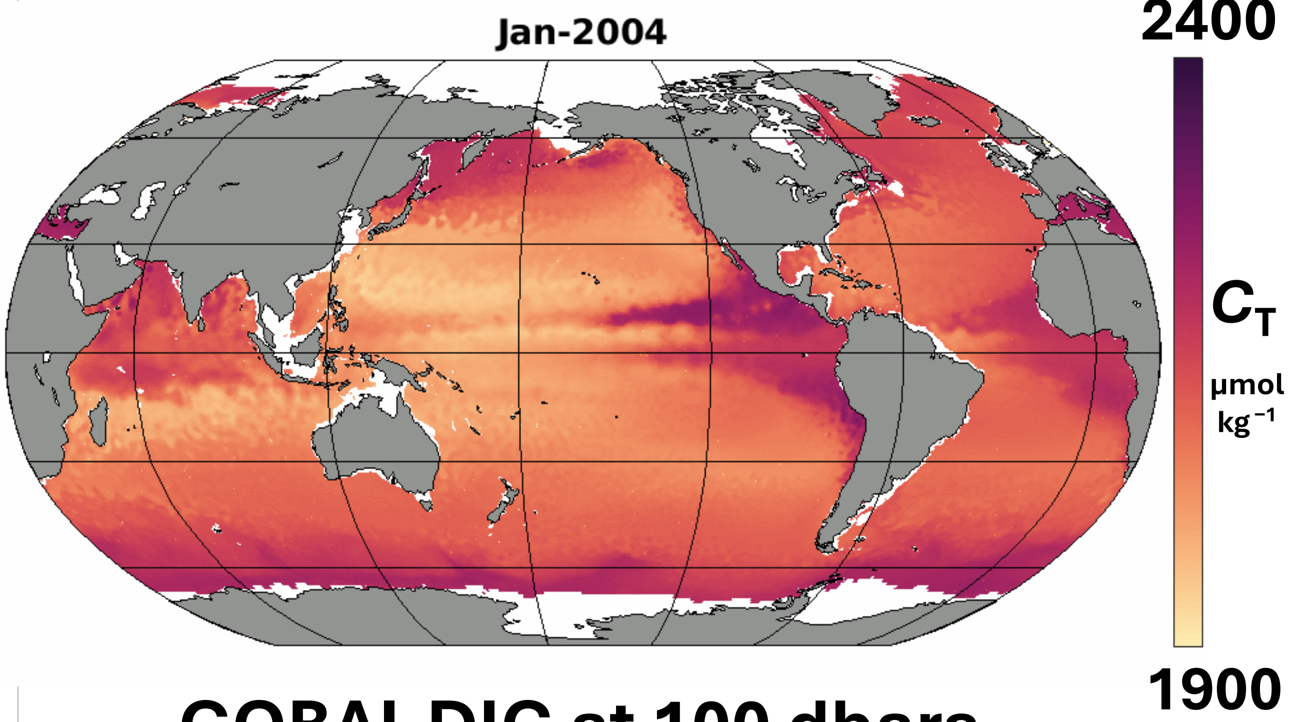
OXYGEN AND NITRATE DISPLAY COHERENT TRENDS





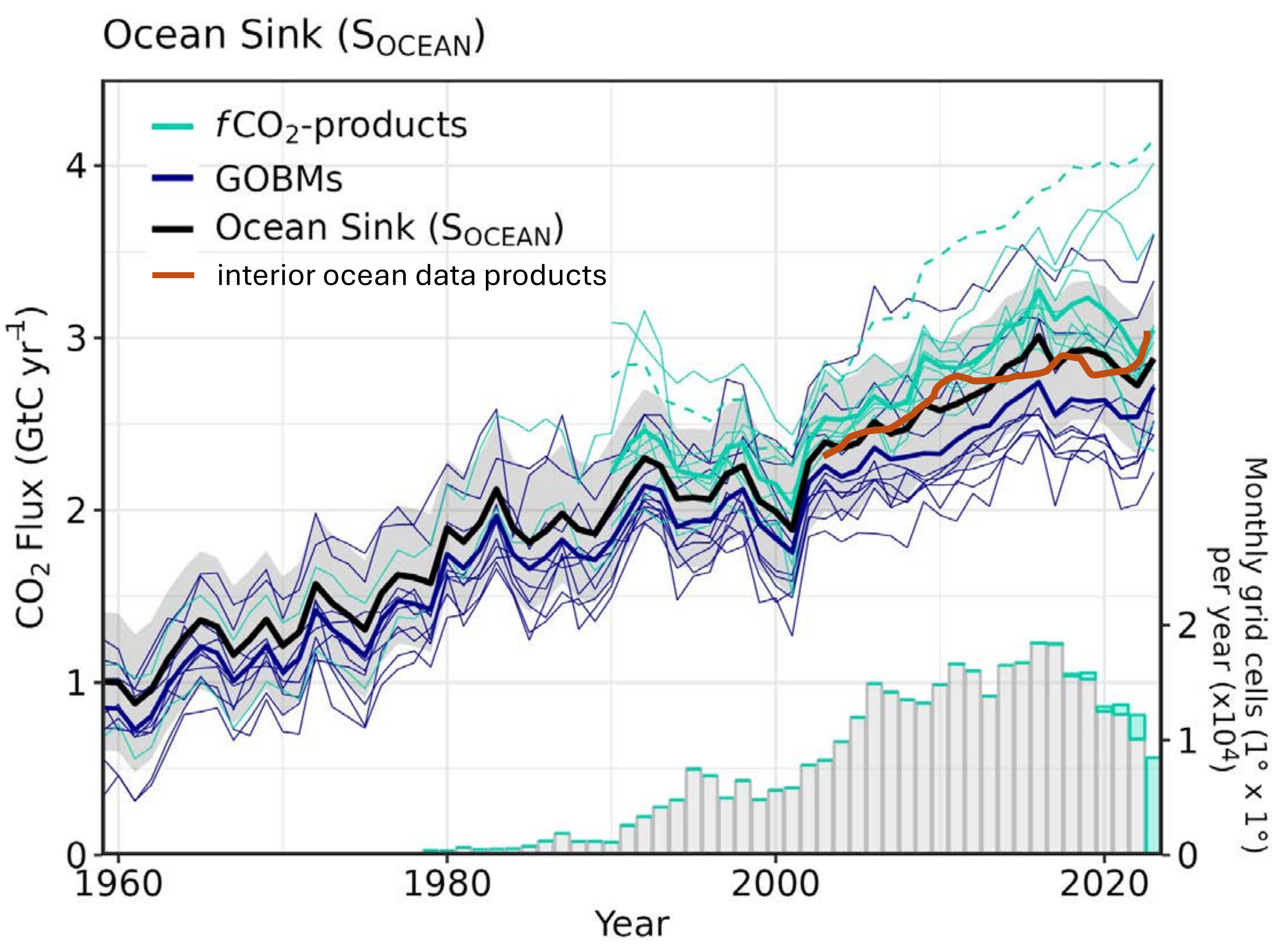
DISSOLVED INORGANIC CARBON SHOWS STEADY INCREASE OVER TWO DECADES

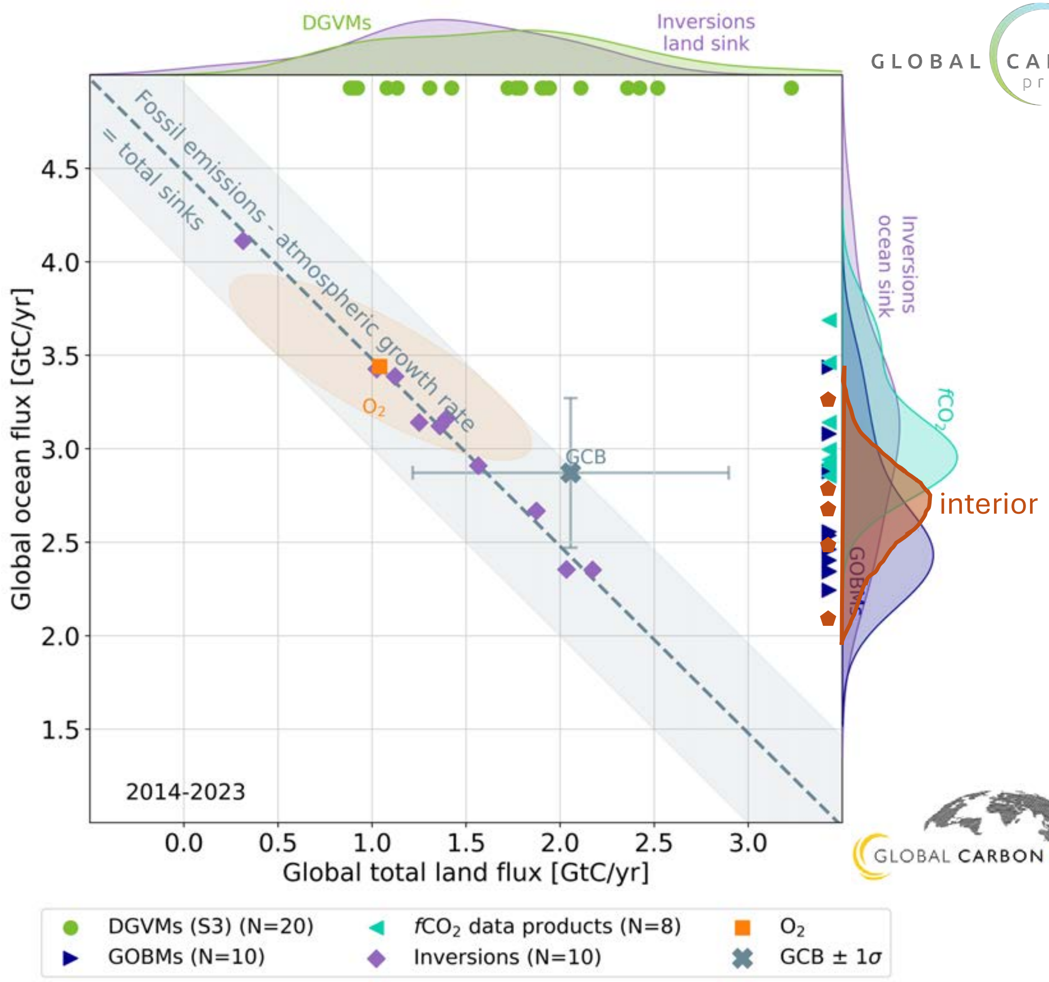




GOBAI-DIC at 100 dbars

MAPPED OCEAN INTERIOR DATA PRODUCTS MAY SOON OFFER ADDITIONAL CONSTRAINTS ON OCEAN CARBON SINK





KEY TAKEAWAYS

- Limitations remain in how well we can **constrain past changes and predict future changes** to ocean biogeochemistry across temporal and spatial scales
- Novel observational platforms present new opportunities for the **mapping of ocean biogeochemical properties** in four dimensions in near-real time
- Four-dimensional observation-based maps offer **brand new insights and additional constraints** on key ocean processes