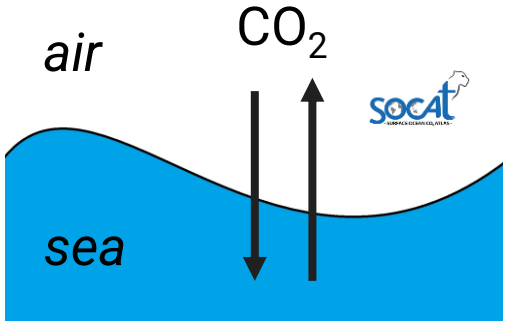
Machine Learning Approaches to Leverage the Marine Biogeochemistry Observing System

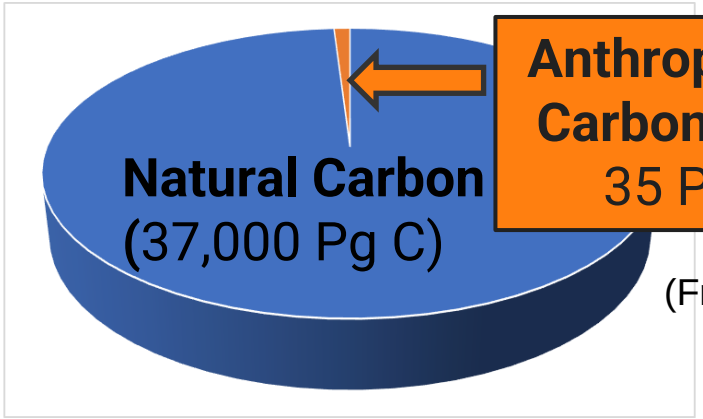
Larissa Dias
Brendan Carter, Andrea Fassbender, Jonathan Sharp
University of Washington & NOAA Pacific Marine Environmental Laboratory

GOMO Community Workshop, July 25-27, 2023, Silver Spring, MD

Quantifying Ocean Carbon



Total Ocean Inorganic Carbon

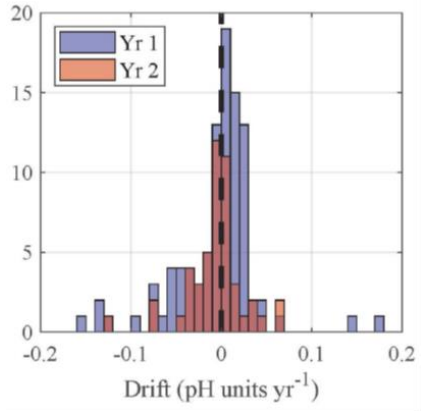


(Freidlingstein et al., 2022)

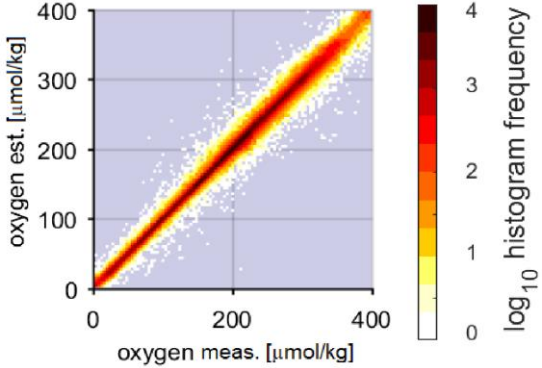
Estimation Algorithms

Empirical Seawater Property Estimation Routines (ESPERs):

- Aim: Realize the full potential of biogeochemical (BGC) float data
- Trained on ship-based measurements
  - Machine learning (ML) and locally interpolated regression (LIR) options



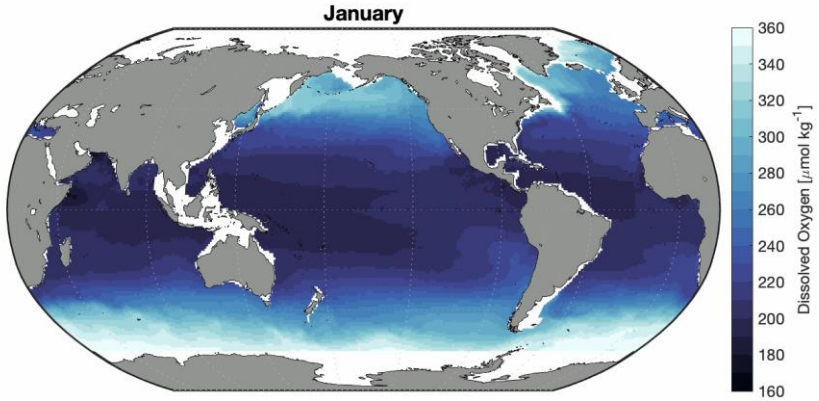
pH drift at 1500 m (*from* Mauer et al, 2021)



(*from* Carter, 2021)

Products

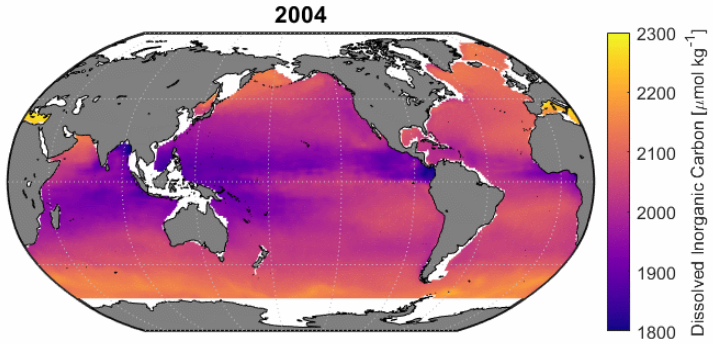
1. Python ESPER and updates
2. Gridded Ocean Biogeochemistry from Artificial Intelligence – Oxygen (GOBAI-O_2)
   - 4D product (lat., lon., depth, time)
   - ML algorithms applied to T, S from Argo



Surface [O_2] averaged over 2004–2022 from GOBAI-O_2 (from Sharp et al, *in press*)

Future Directions

- Extending 4D to other carbonate and biogeochemical properties
  - Combined float, hydrographic survey + ML
  - Applied to T and S product (Lyman and Johnson, *in prep*.)
  - OSSEs to assess data product skill



Monthly average surface DIC timeseries calculated with ESPER LIR and NN using climatology from Roemmich and Gilson (2009)

- Retrain ESPER with ship and float observations → enhanced temporal prediction ability
- 4D products + ESPER → dynamically updating

Societal Benefit

Near real-time ocean BGC information could make it possible to:

- Monitor strength of and variability in the ocean carbon sink and biological carbon pump seasonally
- Provide feedback during multiple stressor events (e.g., marine heat waves, low oxygen events, ocean acidification events)
- Improve boundary/initial conditions for regional model forecasts

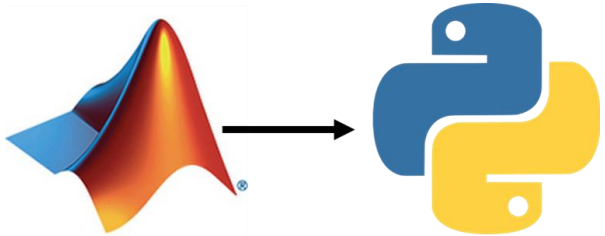
Partners, Stakeholders, and Beneficiaries:

- Ocean acidification researchers
- Fisheries researchers and managers
- Modelers and climate scientists
- Global carbon inventory tracking authorities

Data Management

FAIR (findability, accessibility, interoperability, and reusability) data practices

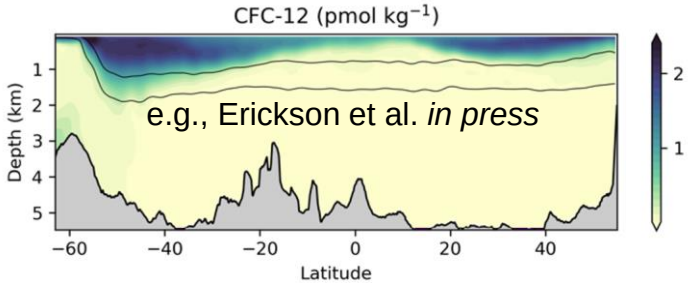
1. Publicly and freely-available
2. Multiple coding languages
3. Checkpoint doi



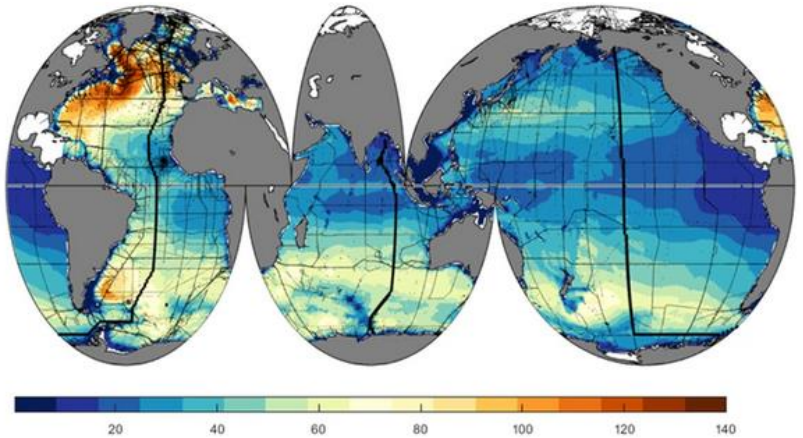
4. Data synthesis products for ESPER training data:
    - Internally consistent pH data product
    - Metadata data product

TAKEAWAY

BGC Argo floats (paired with machine learning methods) have the potential to **revolutionize the study and understanding of open ocean carbonate system and anthropogenic carbon stored in the ocean interior**. Our approaches provide a means of leveraging this data for this purpose and provide FAIR data tools for marine researchers to utilize. **More work needs to be done.**



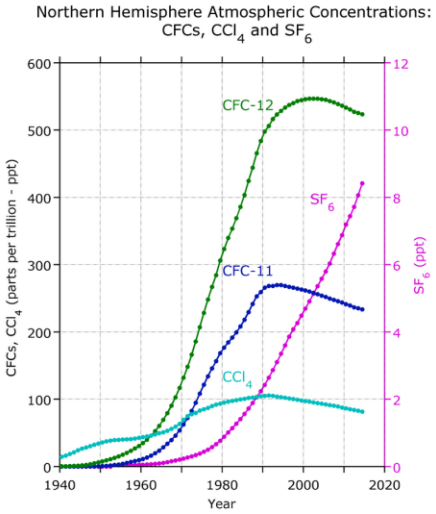
Lauvset et al., 2016



Oceanographic distributions ↑
with Atmospheric histories ↓

Transit time distributions

(Think: water mass ages)



TRACE:
Tracer-based
Rapid
Anthropogenic
Carbon
Estimation

Scale with time

ESPERv2023

ESPERv2021