Science Mentoring Opportunity at American Museum of Natural History for PhD students and recent PhDs

This year, they are particularly looking for ML/AI projects.

Join SRMP: Become a Mentor

The **S**cience **R**esearch **M**entoring **P**rogram (SRMP) is a one-year, after school mentored research internship for NYC high school students. Our goal is to increase access to science fields and careers by providing authentic, high quality science research opportunities and meaningful mentorship, with a focus on students from backgrounds traditionally underrepresented in science.

What does a mentor do?

- **Be present:** Support & guide mentees afterschool, 2x a week from September thru May (~ 125 hrs)
- **Conduct research:** Share the joys of your science, train mentees in new skills, and assist students in communicating their research
- **Share your professional journey -** ups & downs - and assist your mentees in theirs.
- **Commit to learning**: develop effective mentor practice through formal workshops (3 to 5) and peer reflection

Why join?

- **Add value to your career:** Mentoring high school students provides personal fulfillment as well as professional development. Build skills in communication, mentorship and team management, making scientists more competitive in a wide range of career pathways.
- **Change the face of science:** Equity and justice are at the heart of SRMP as we support BIPOC and students from marginalized communities as they start their STEM journey
- **Advance knowledge:** Not all SRMP research are publishable, but they are all real. SRMP projects are great for pilot studies, subsets of your research agenda, or engaging with side-projects

Mentors receive $9,000 stipend + $500 for supplies OR TA credit through RGGS

Contact Dr. Maria Strangas; Senior Manager to learn more srmp@amnh.org

Support for the Science Research Mentoring Program at the American Museum of Natural History is provided by Christopher C. Davis; The Shelby Cullom Davis Charitable Fund; The Pinkerton Foundation; and the Adolph and Ruth Schnurmacher Foundation.

Climate Prediction Challenges

Project 3: Quantifying the ocean carbon sink with machine learning

Galen A. McKinley

Earth and Environmental Sciences

Lamont Doherty Earth Observatory

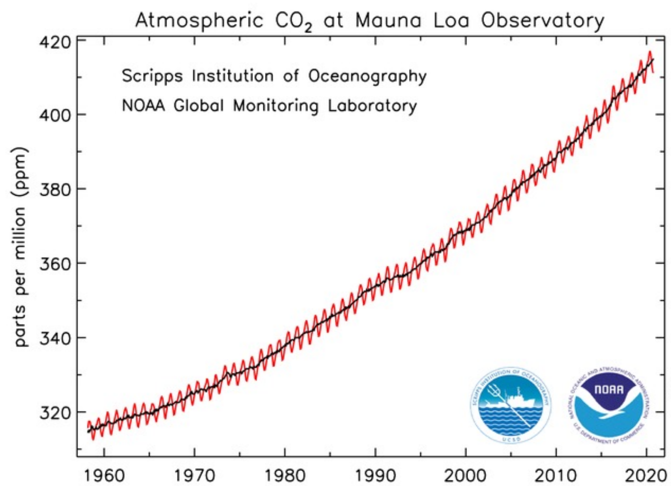
STAT GR5243 / GU4243

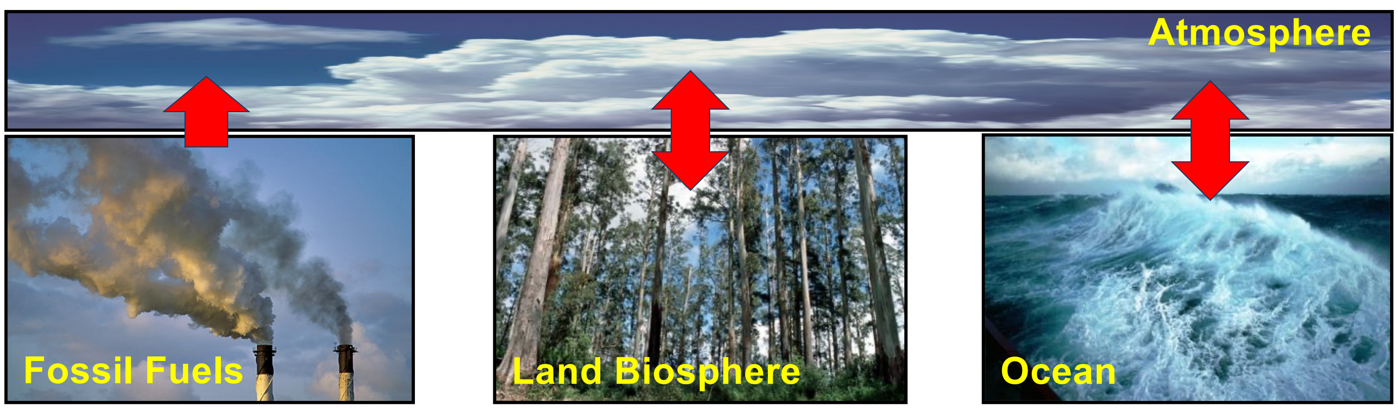
22 March 2022

Lamont-Doherty Earth Observatory

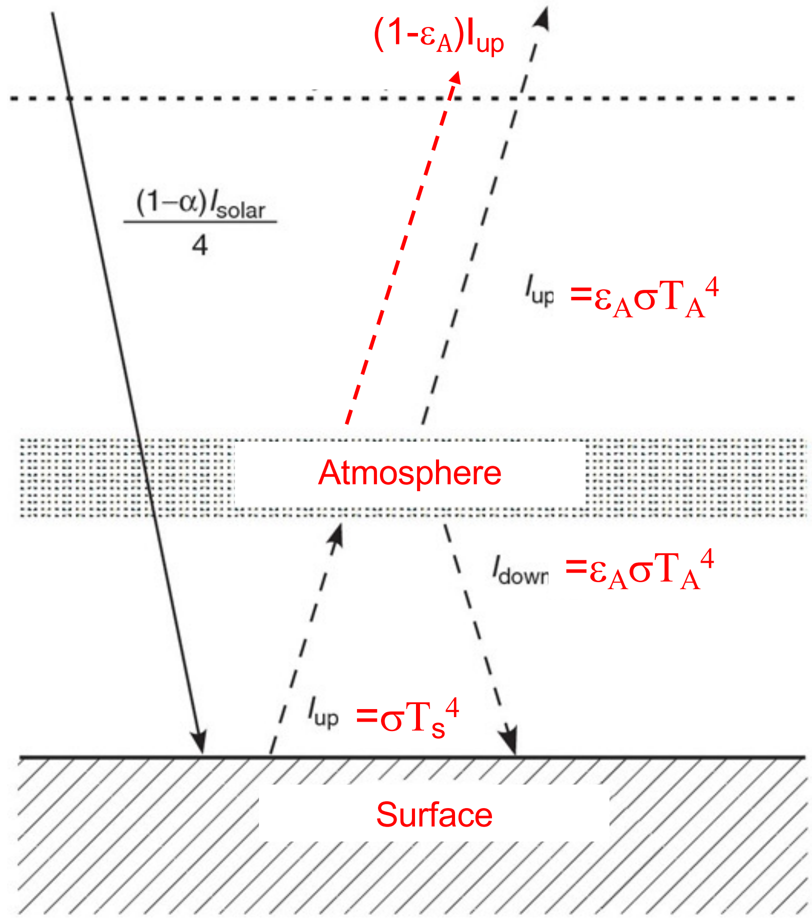
COLUMBIA UNIVERSITY | EARTH INSTITUTE

All our emitted carbon doesn't remain in the atmosphere



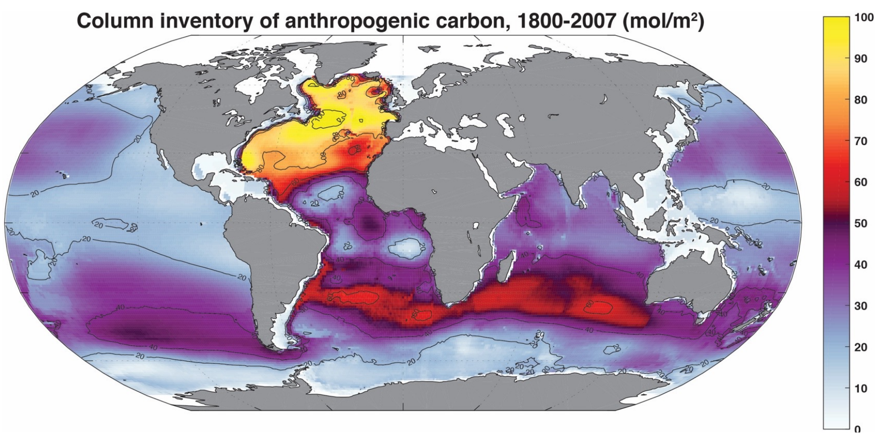


The Greenhouse Effect. Terrestrial radiation absorbed by atmosphere is re-radiated back to surface. If ε increases, less radiation escapes, thus surface warms. **Increasing CO_2 raises ε.**

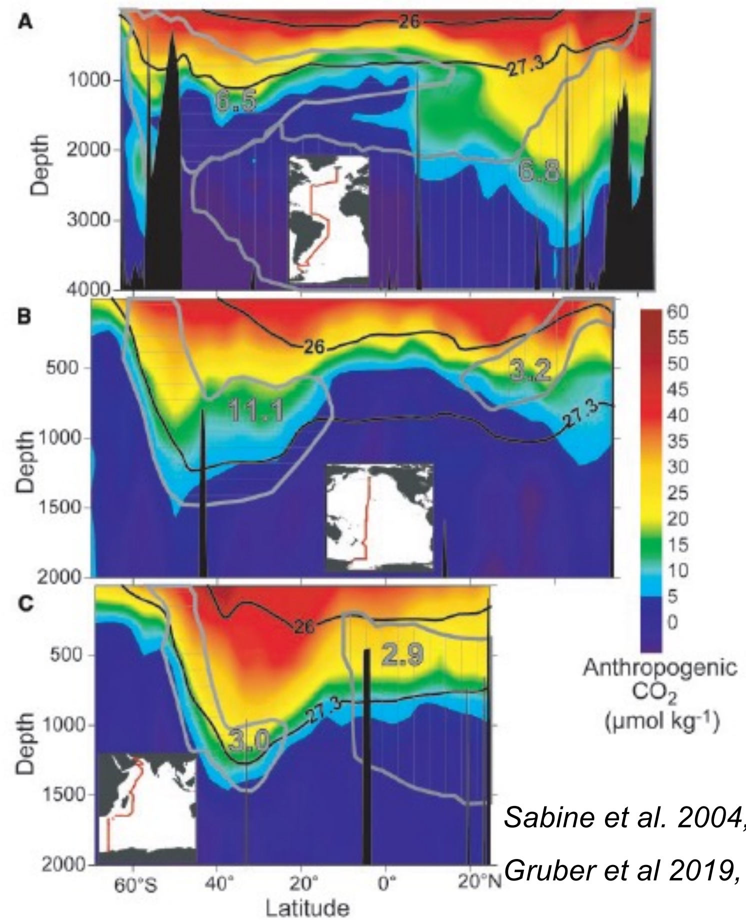


Archer Figure 3-4

Total ocean anthropogenic carbon uptake (1800-2007)



~40% of all fossil carbon has gone into the ocean (140 PgC)

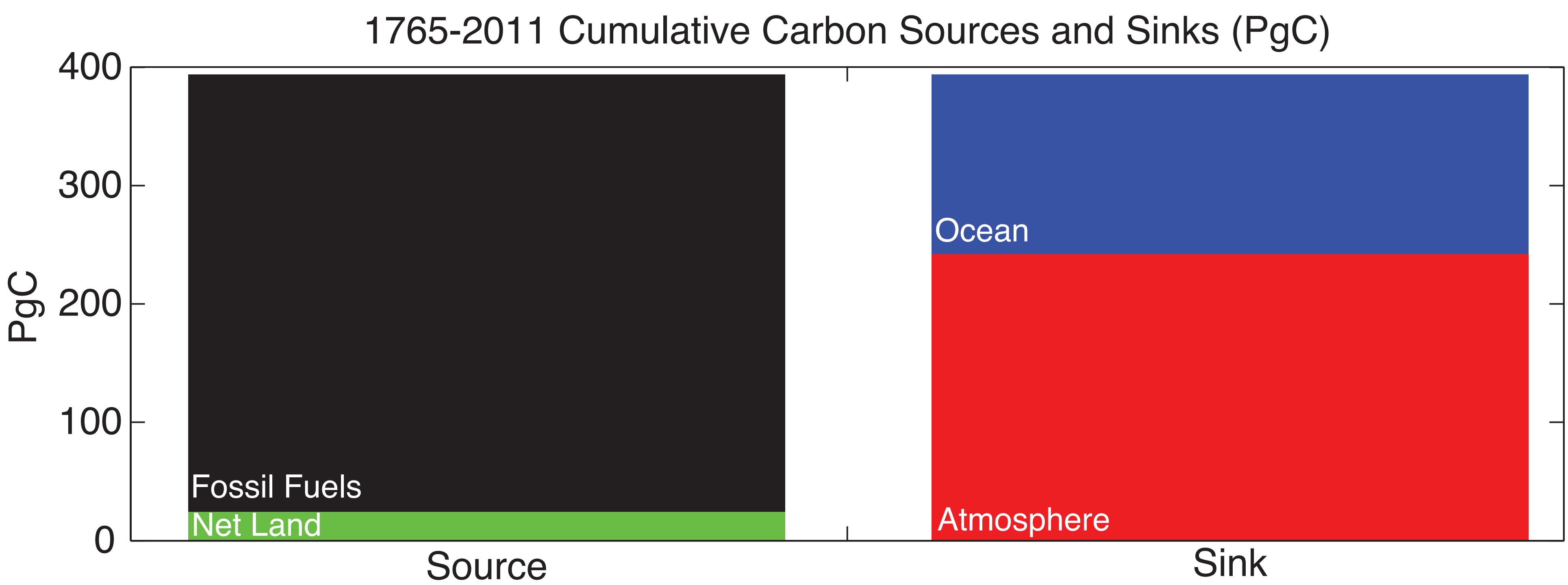


Sabine et al. 2004, Science

Gruber et al 2019, Science

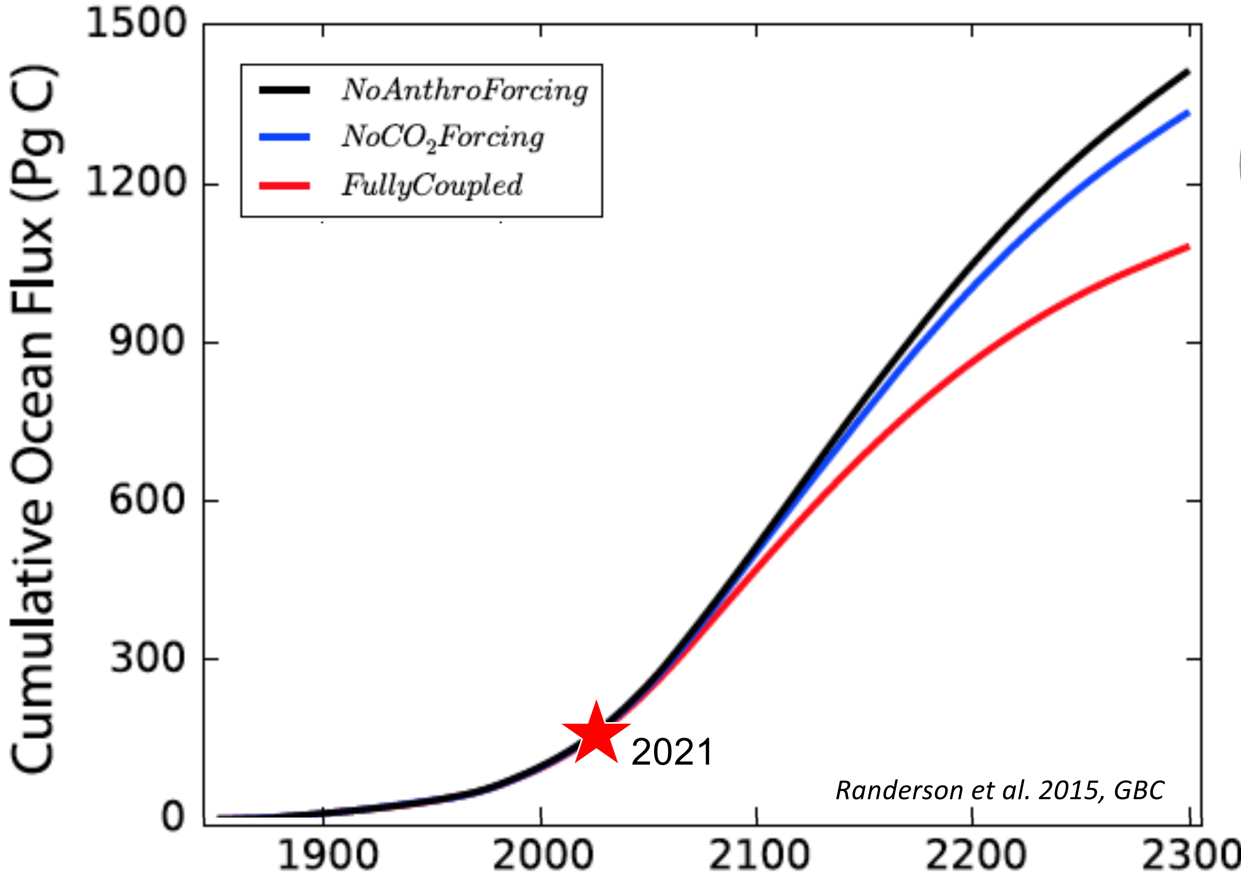
Based on high quality repeat-hydrography data

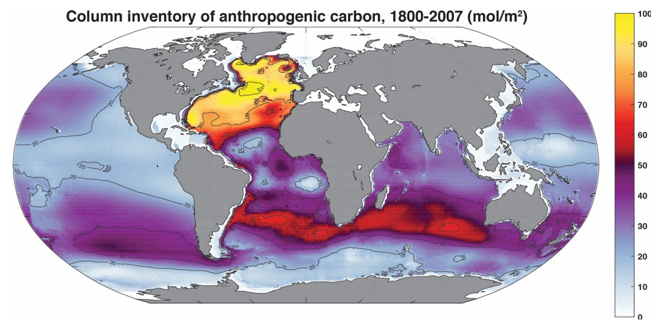
The ocean has absorbed ~40% of all fossil fuel CO_2



Data of Khatiwala et al. 2009 Nature, updated

Climate model projection, cumulative anthropogenic carbon uptake

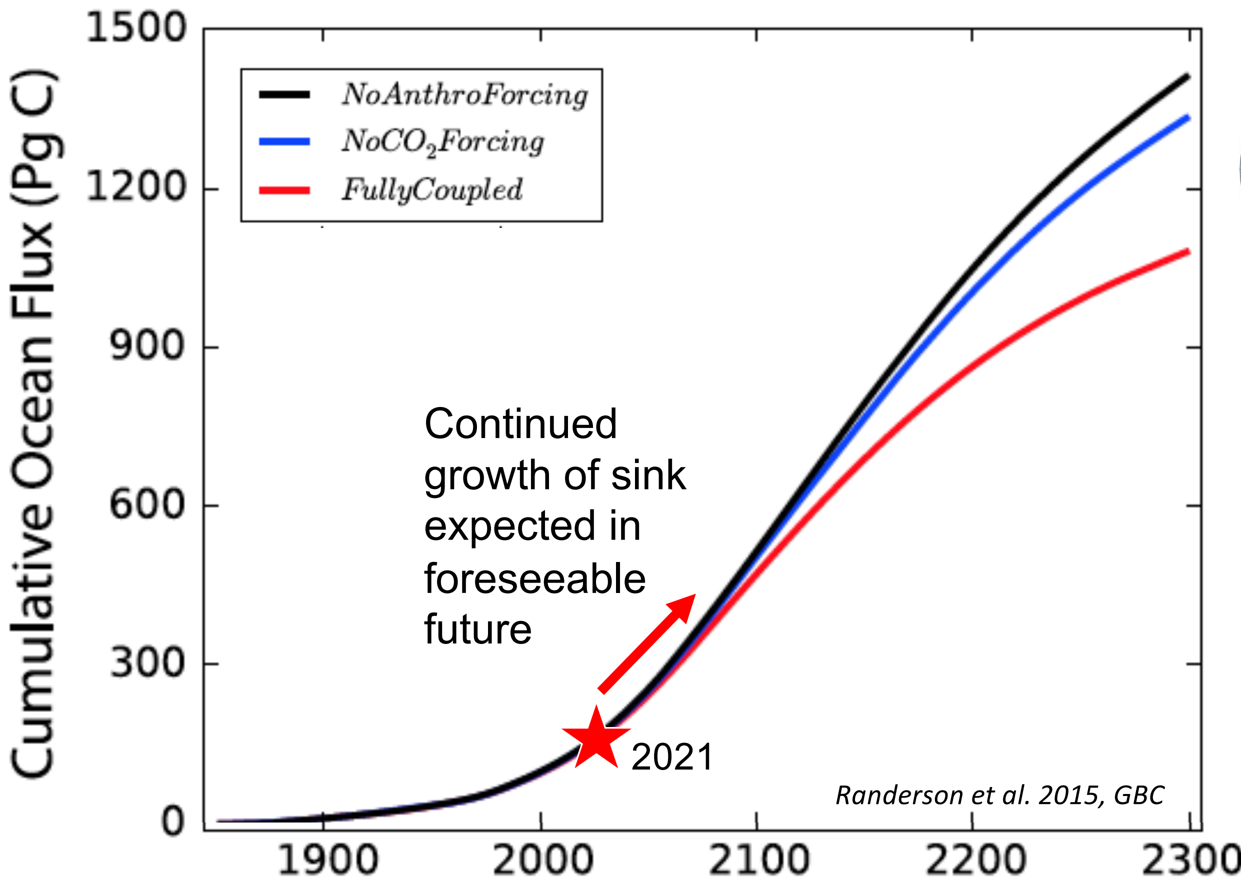


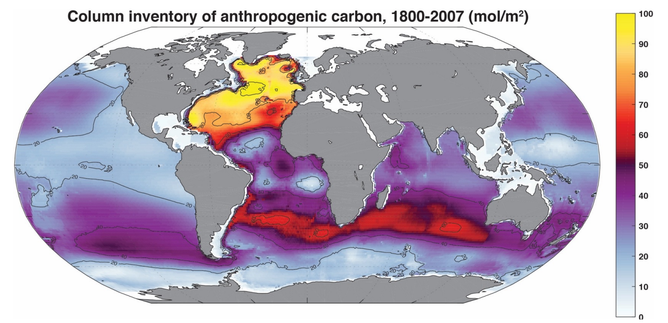


Growing sink

RCP8.5 scenario

Climate model projection, cumulative anthropogenic carbon uptake

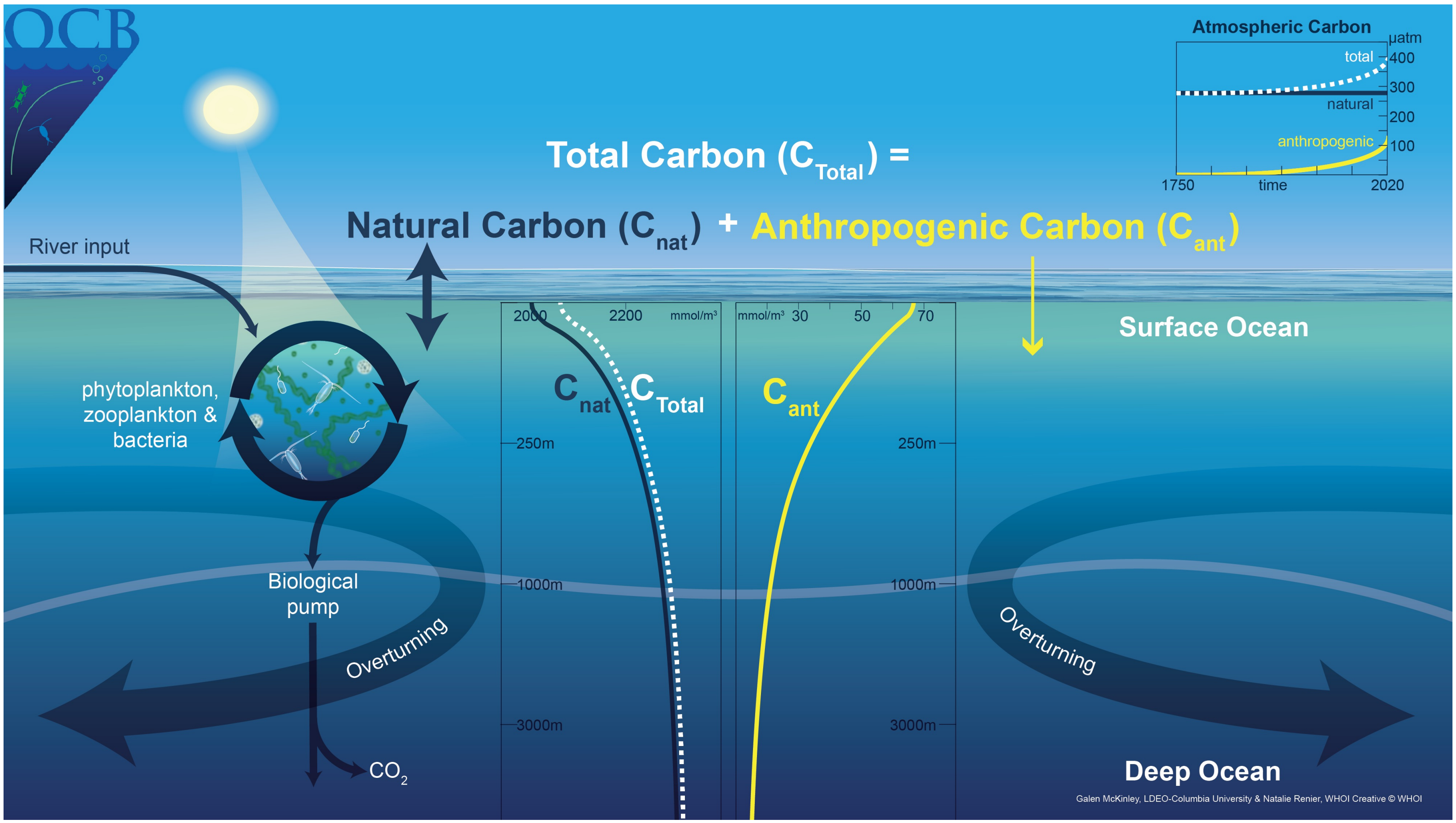




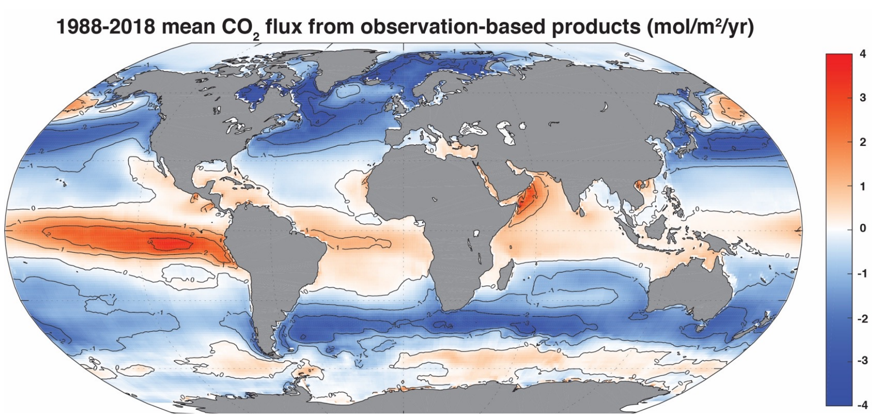
Growing sink

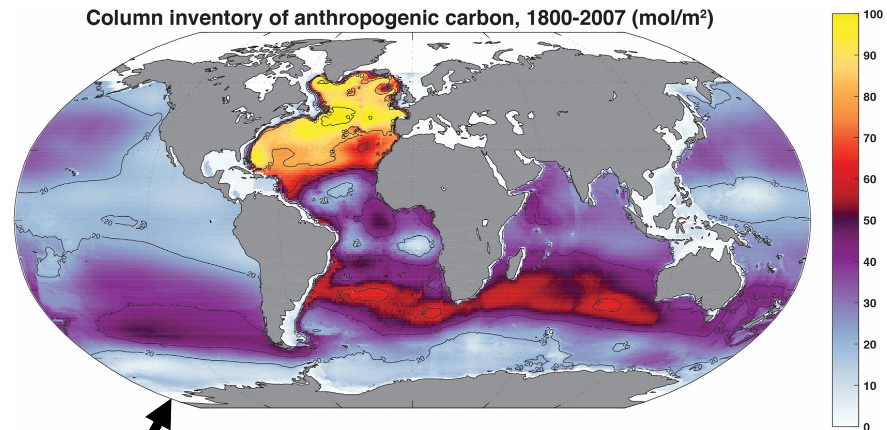
RCP8.5 scenario

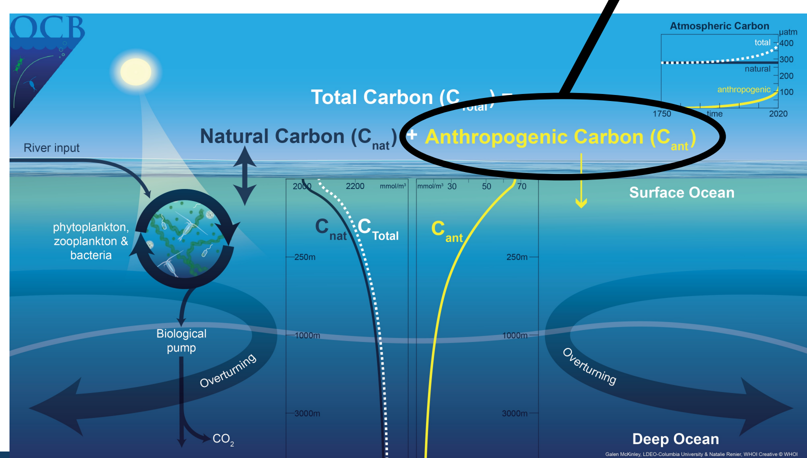
Natural carbon cycling and the anthropogenic ocean carbon sink



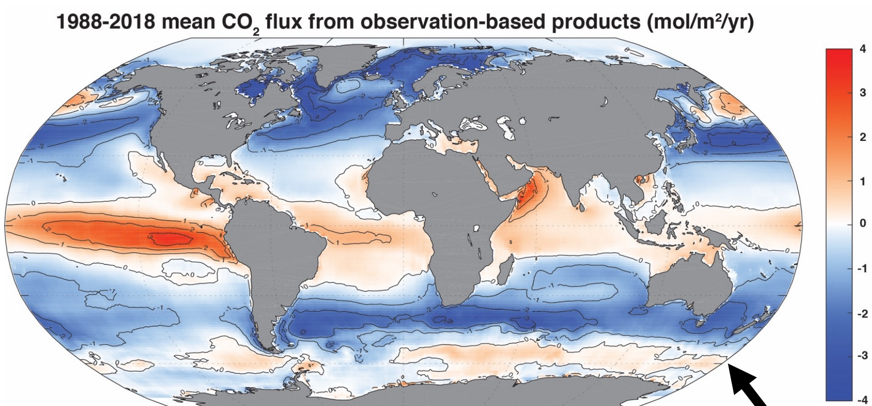
Natural carbon cycling and the anthropogenic ocean carbon sink

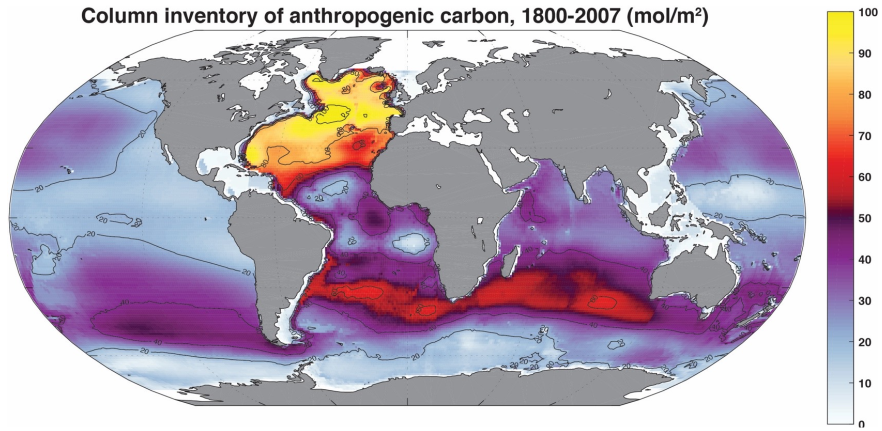


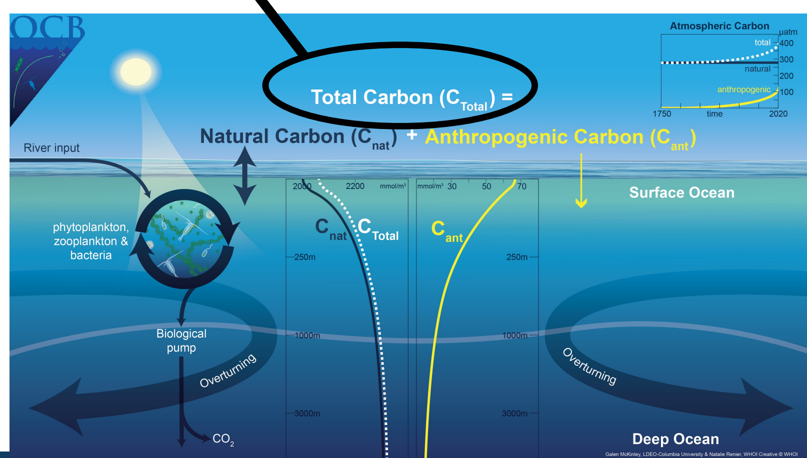




Natural carbon cycling and the anthropogenic ocean carbon sink

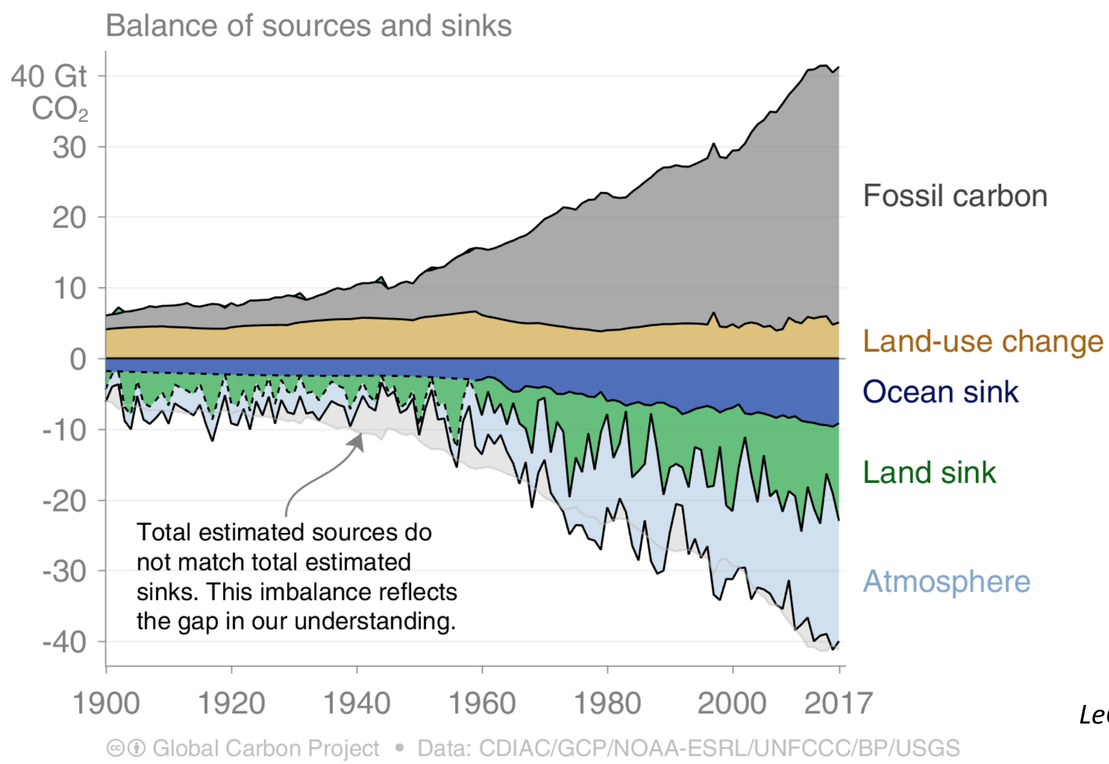






Lamont-Doherty Earth Observatory
Columbia University | Earth Institute

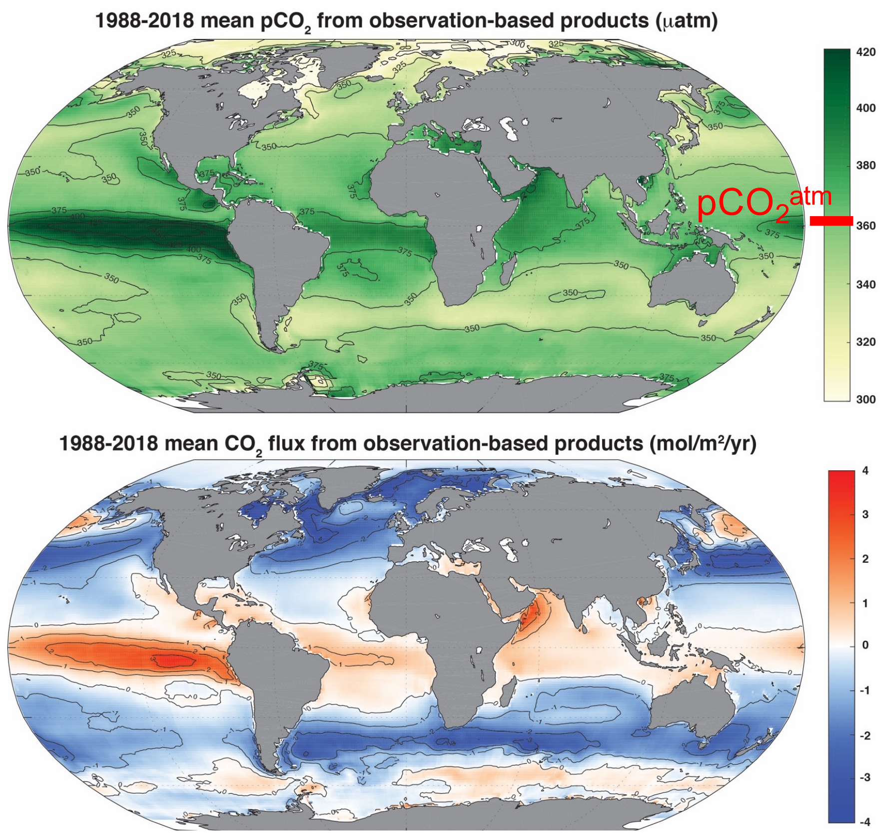
Variability of carbon sinks is generally understood; but remaining lack of closure will prevent rapid identification of change in atmospheric CO_2 trajectory due to mitigation actions



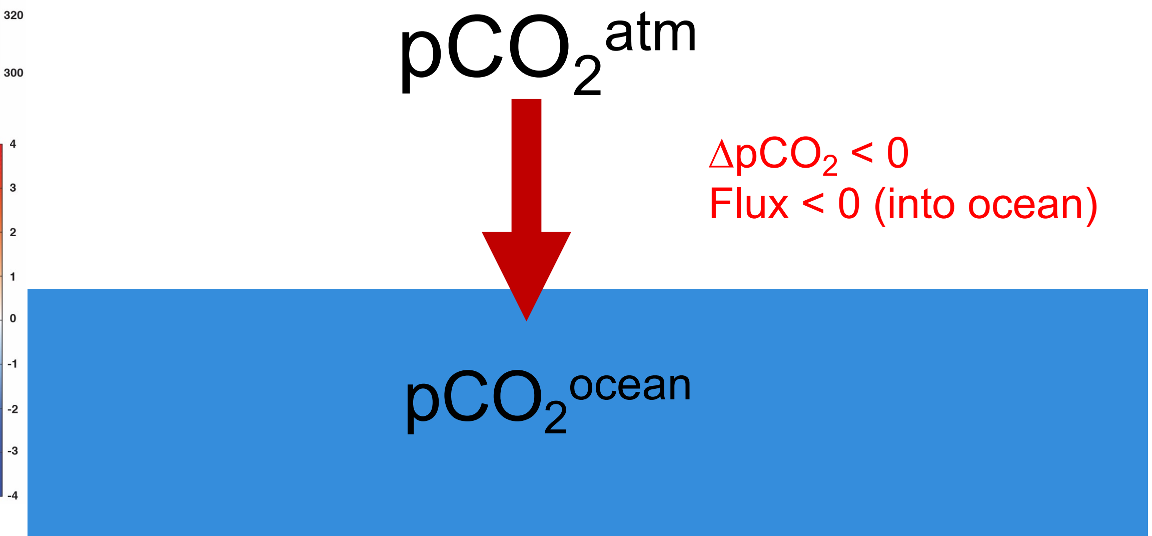
LeQuere et al 2018 ESSD

How do we quantify the ocean sink variability?

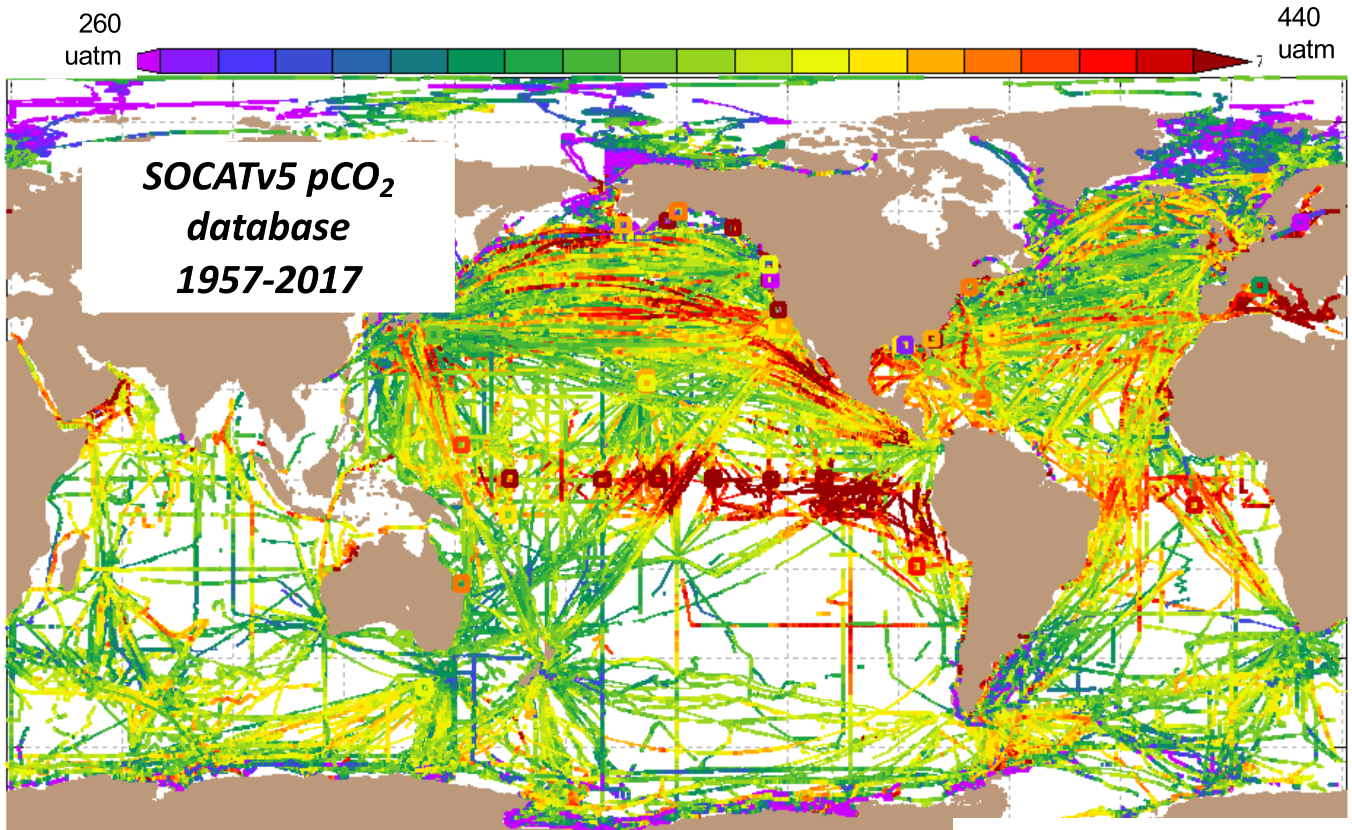
$$Air\ sea\ CO_2\ flux = k_w S_{CO_2}(1 - f_{ice})\underbrace{(\boldsymbol{pCO_2^{ocean}} - \boldsymbol{pCO_2^{atm}})}_{\boldsymbol{\Delta pCO_2^{atm}}}$$



Crisp et al., *in review*

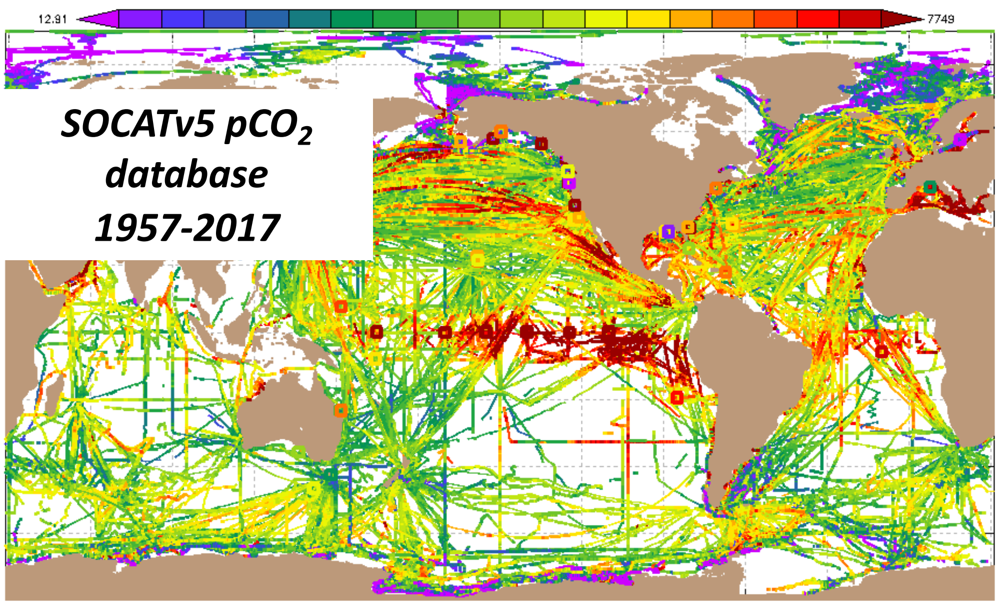


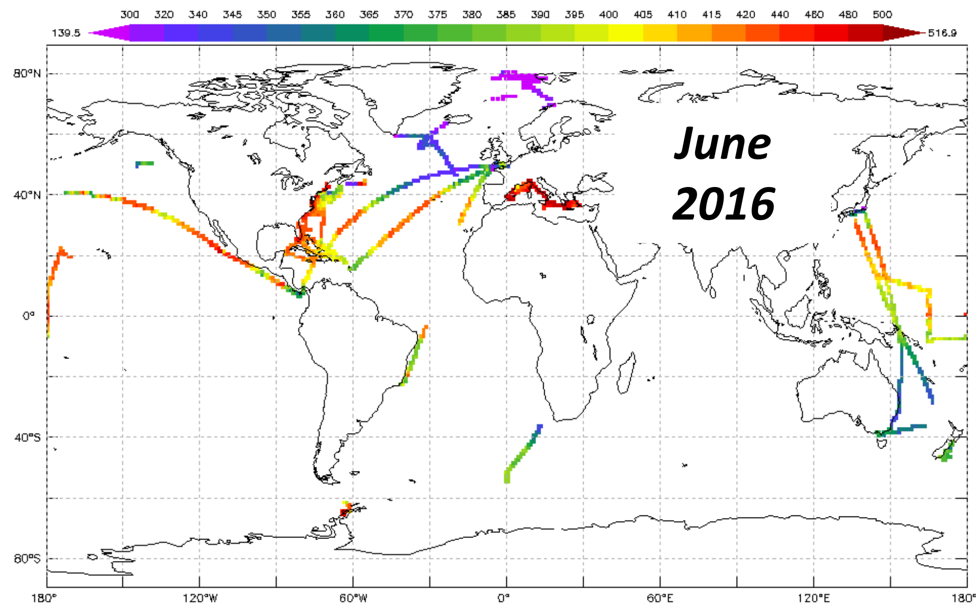
Surface ocean pCO_2 observations



Bakker et al. 2016 ESSD
Pfeil et al., Sabine et al. 2013 ESSD

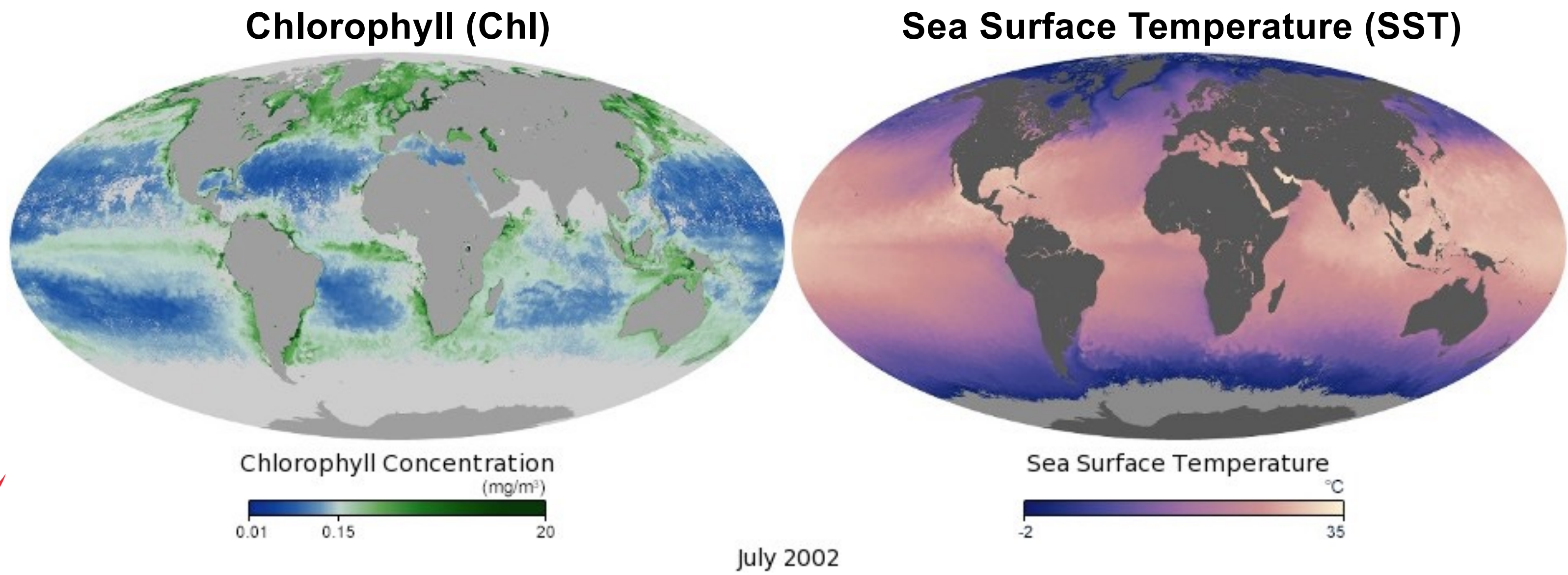
pCO_2 data are very sparse …
How to turn sparse data into full coverage, time-evolving flux maps?





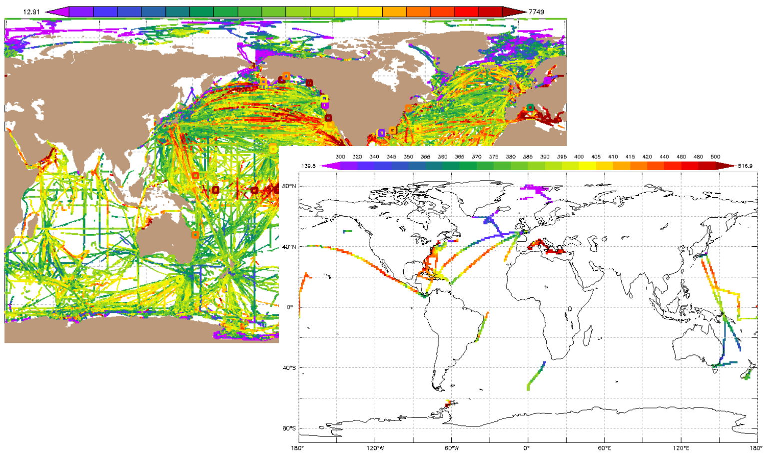
 |

Full coverage data related to pCO_2 do exist

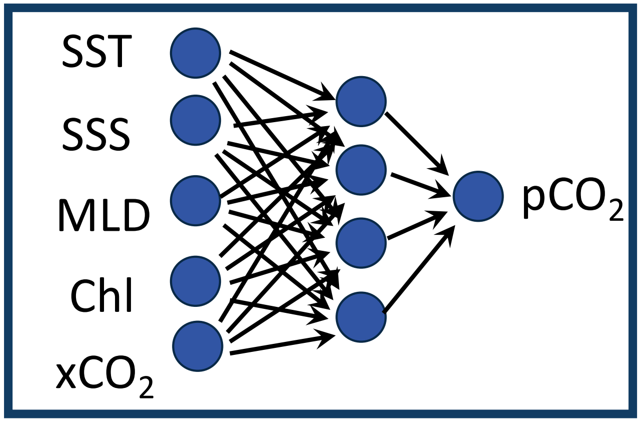


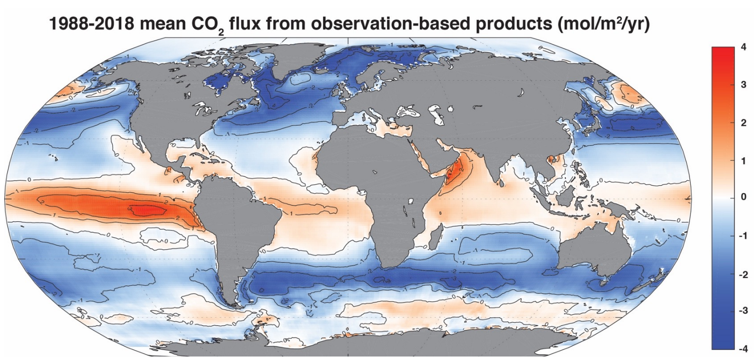
Lamont-Doherty Earth Observatory
COLUMBIA UNIVERSITY | EARTH INSTITUTE

Machine learning now common to extrapolate pCO_2 to global



1. Train algorithm on sparse data



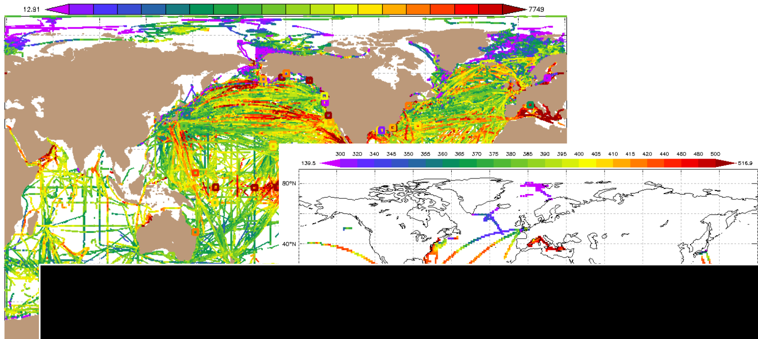


2. Predict pCO_2 = f(SST, SSS, MLD, Chl, xCO_2)

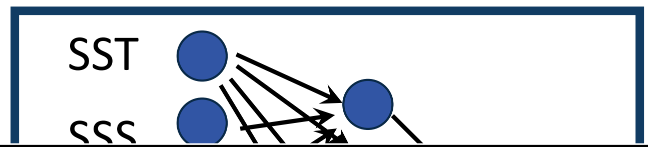
3. Calculate air-sea flux from pCO_2 for all locations and all months

Lamont-Doherty Earth Observatory
Columbia University | Earth Institute

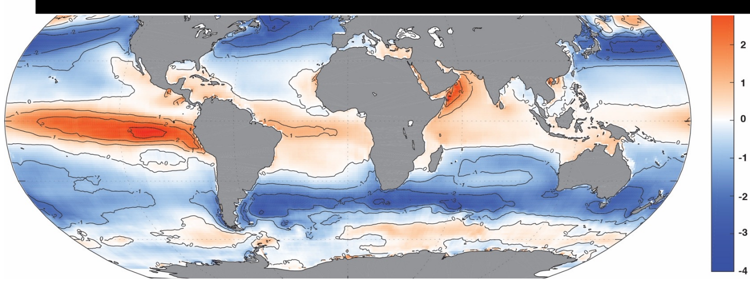
Machine learning now common to extrapolate pCO_2 to global



1. Train algorithm on sparse data



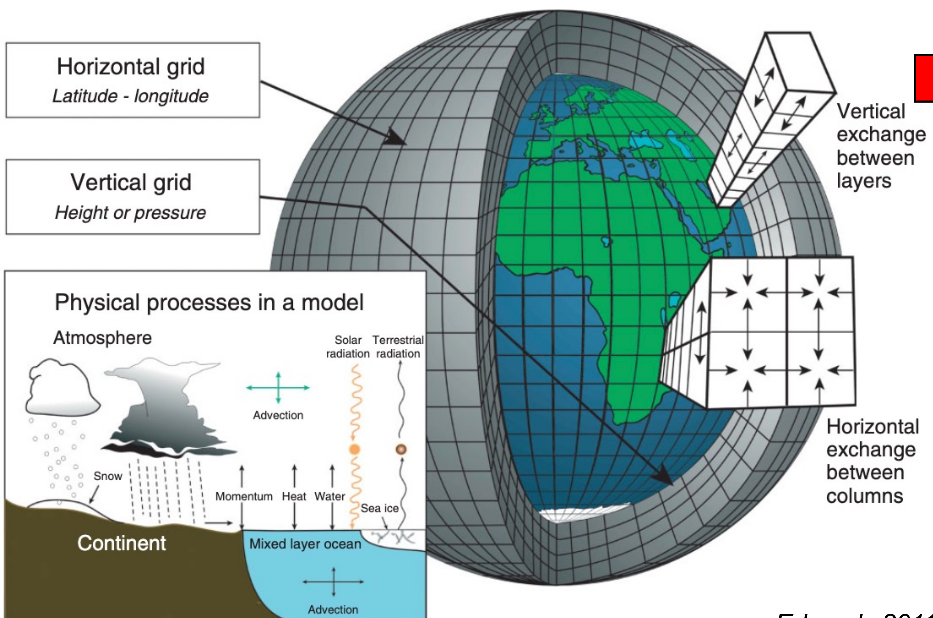
How accurate are the resulting products?
How can we test them when the truth is unknown?



2. Predict pCO_2 = f(SST, SSS, MLD, Chl, xCO_2)
3. Calculate air-sea flux from pCO_2 for all locations and all months

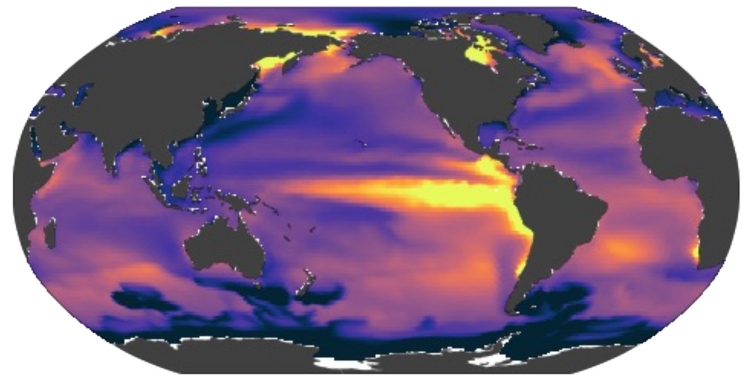
Use ocean models as <u>testbed</u> for extrapolation skill
Sample the model as observations, reconstruct, compare to original model

Earth System Model (ESM)



Edwards 2011

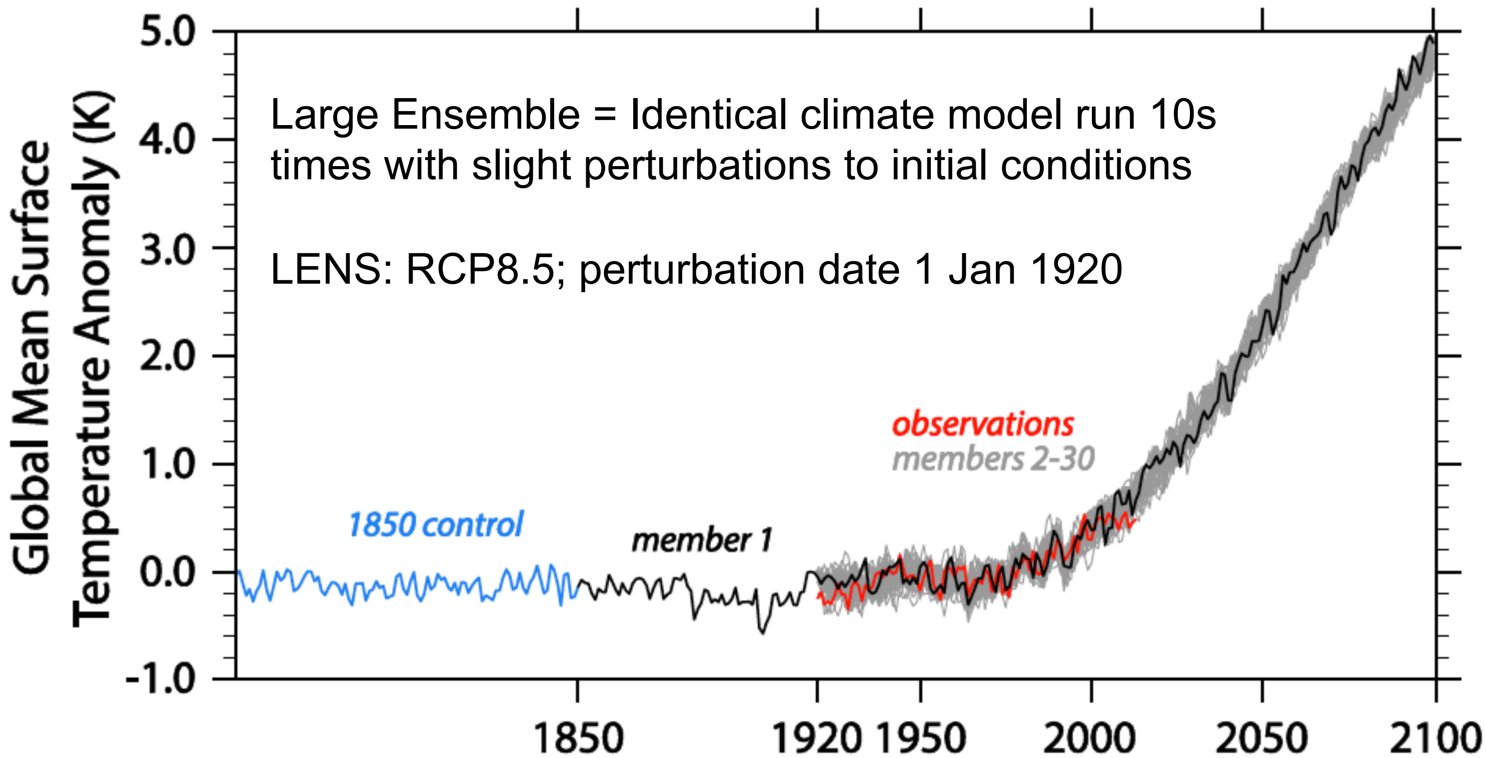
Ocean pCO_2



- 4 ESM Large Ensembles (CESM, CanESM, GFDL, MPI)
- 25 members x 4 = 100 members
- 1985-2016

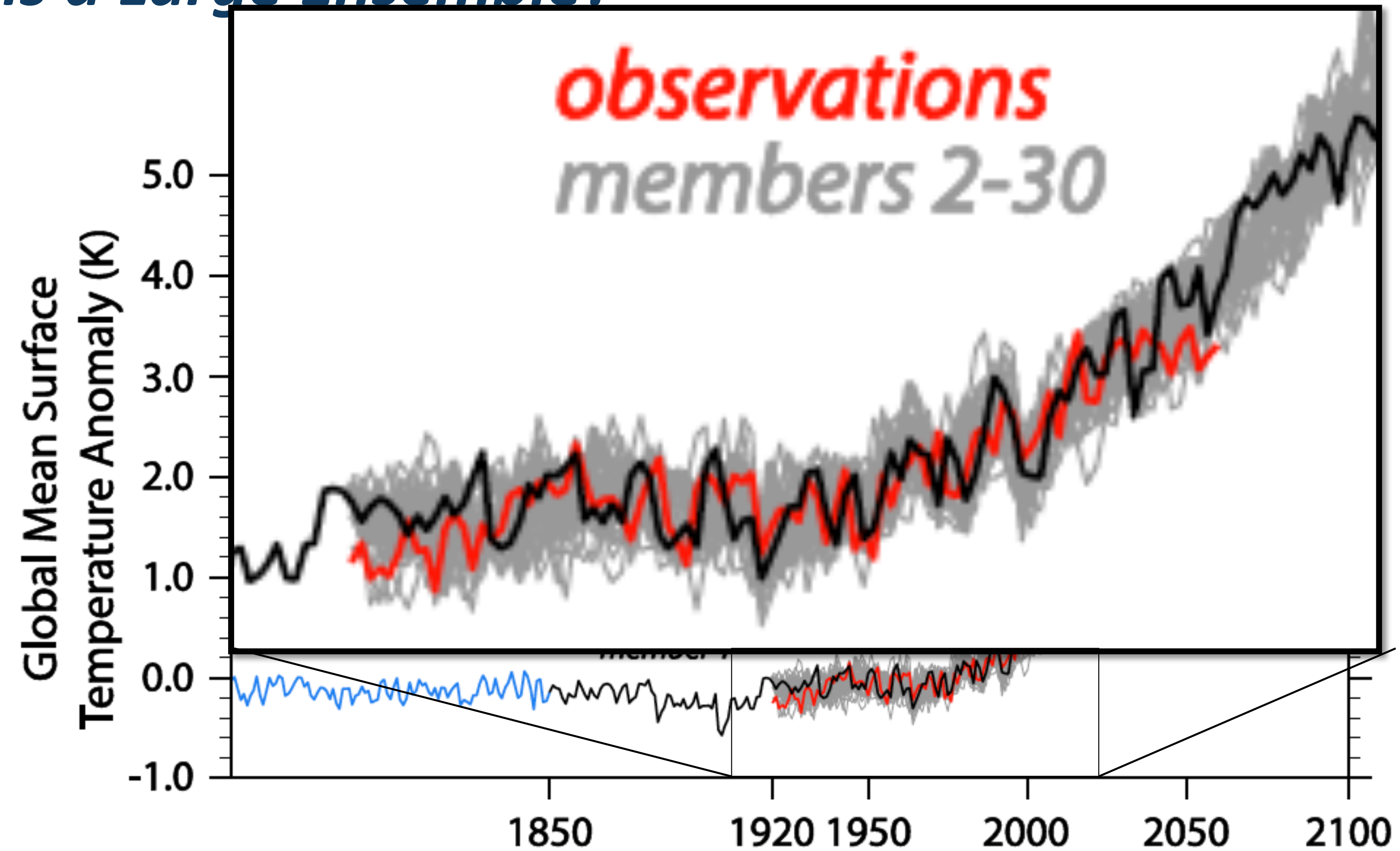
Gloege et al. GBC 2021

What is a Large Ensemble?



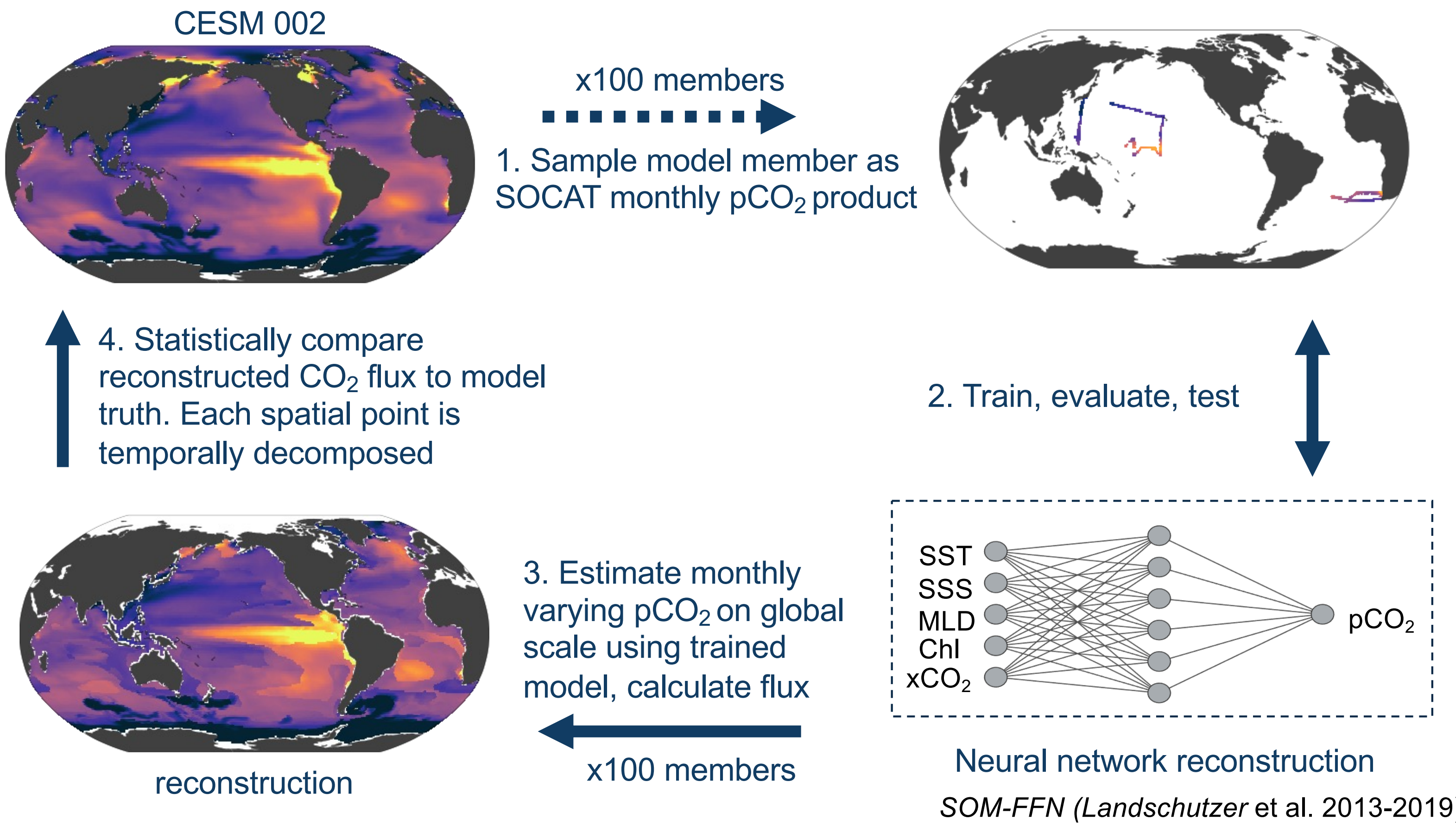
Kay et al. 2014 BAMS

What is a Large Ensemble?



Kay et al. 2014 BAMS

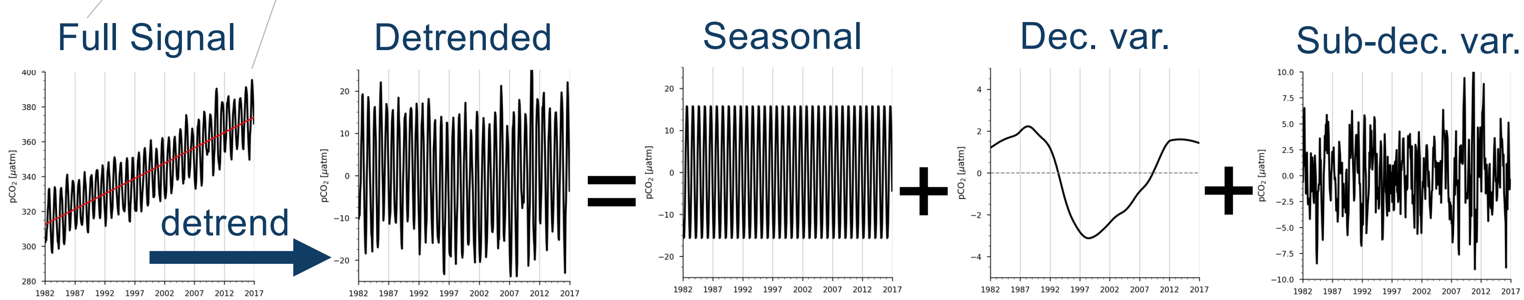
Large ensemble CO_2 testbed



Skill evaluation: temporal decomposition

Ensemble member **Reconstruction**

Each location in each member and reconstruction decomposed into seasonal, decadal, and sub-decadal variability



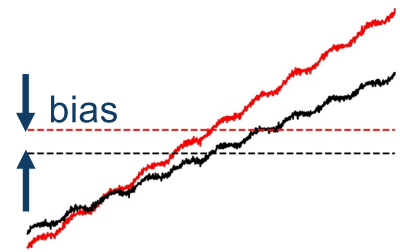
Skill evaluation: statistical metrics

Model **Reconstruction**

Long-term mean

Is there a systemic offset?
<u>Calculated on full signal</u>

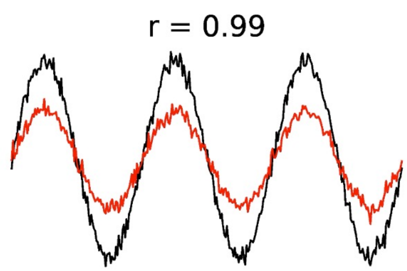
$$Bias = \bar{m} - \bar{r}$$



Phasing

Is reconstruction in phase?

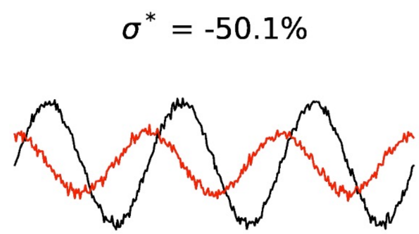
$$Correlation\ (\mathrm{r}) = \frac{cov(m, r)}{\sigma_r \sigma_m}$$



Amplitude

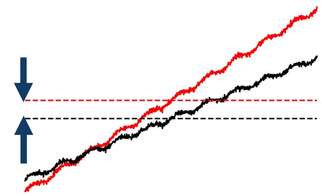
Does reconstruction capture amplitudes of variability?

$$\%error_{STD}\ \sigma^* = \left(\frac{\sigma_r - \sigma_m}{\sigma_m}\right) * 100\%$$

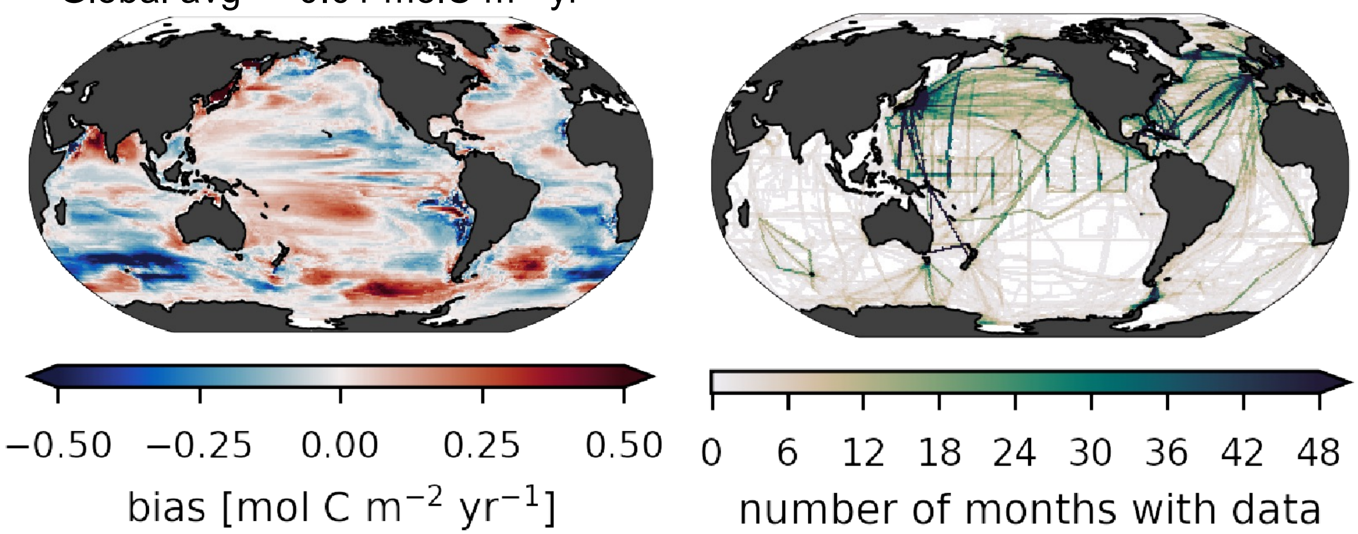


 |

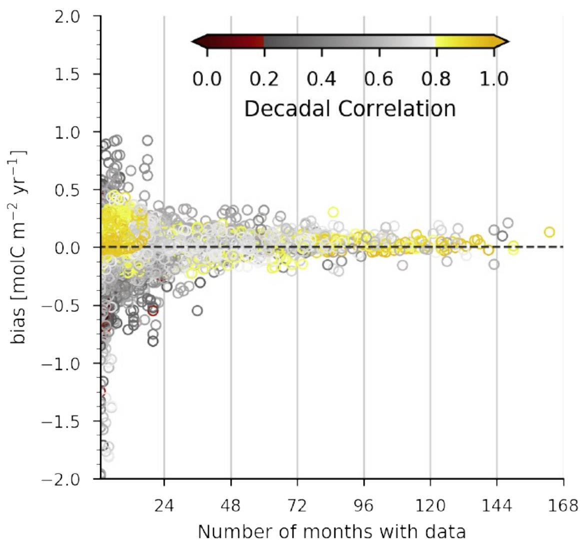
Bias is small globally; larger where data sparse



Global avg = -0.01 molC m^{-2} yr^{-1}

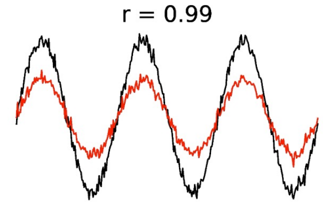


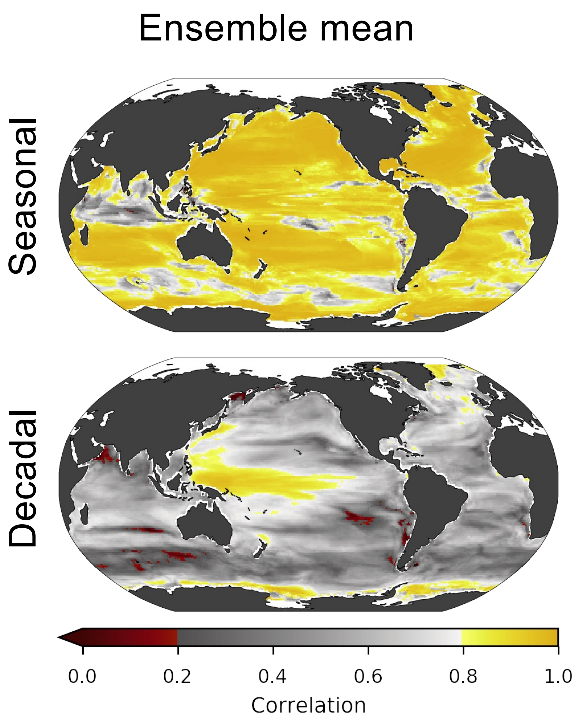
100-member mean



Gloege et al. GBC 2021

Correlation
Is the reconstruction in phase with the original data?



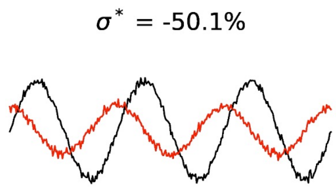


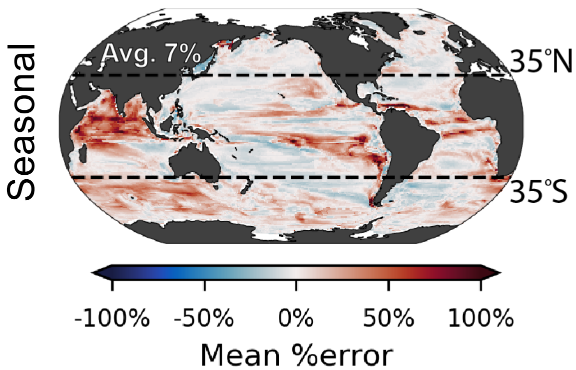
- Captures seasonal cycle well everywhere
- Decadal phase is more challenging to reconstruct

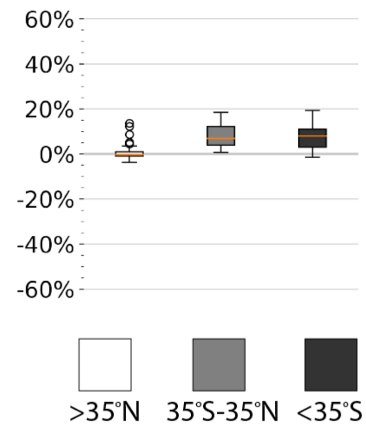
Gloege et al. GBC 2021

Standard deviation of % error

Does the reconstruction capture the amplitude of variability?



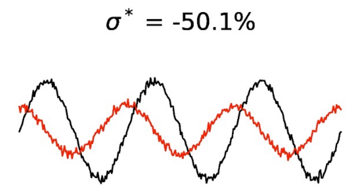


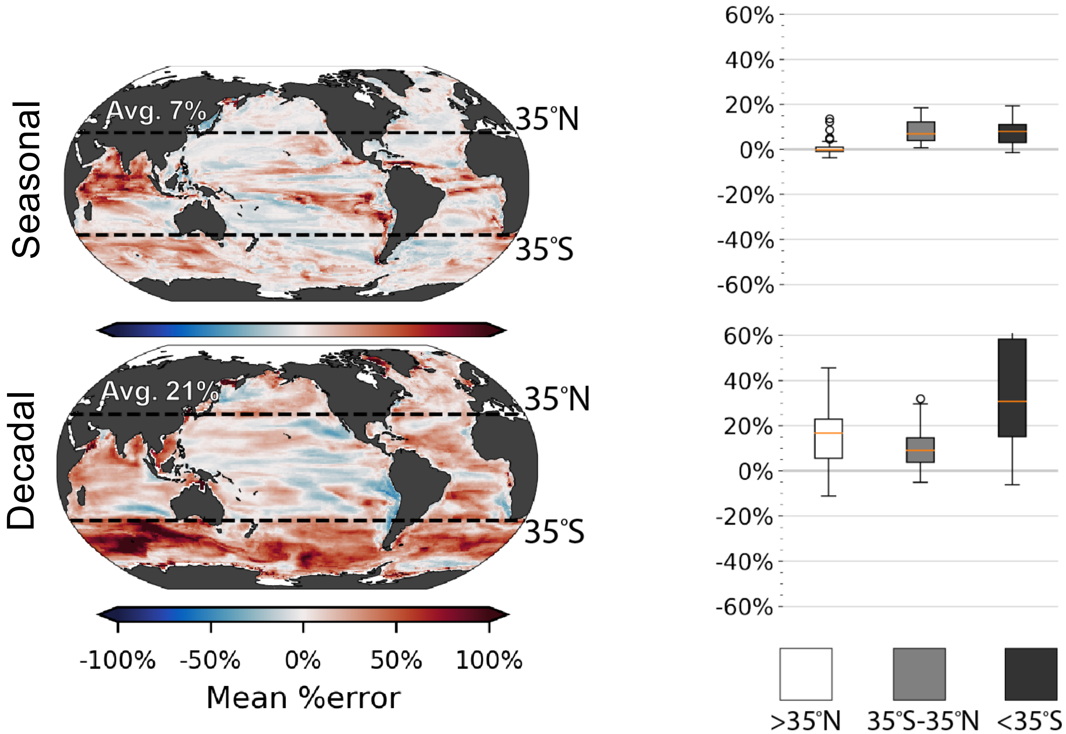


Gloege et al. GBC 2021

Standard deviation of % error

Does the reconstruction capture the amplitude of variability?

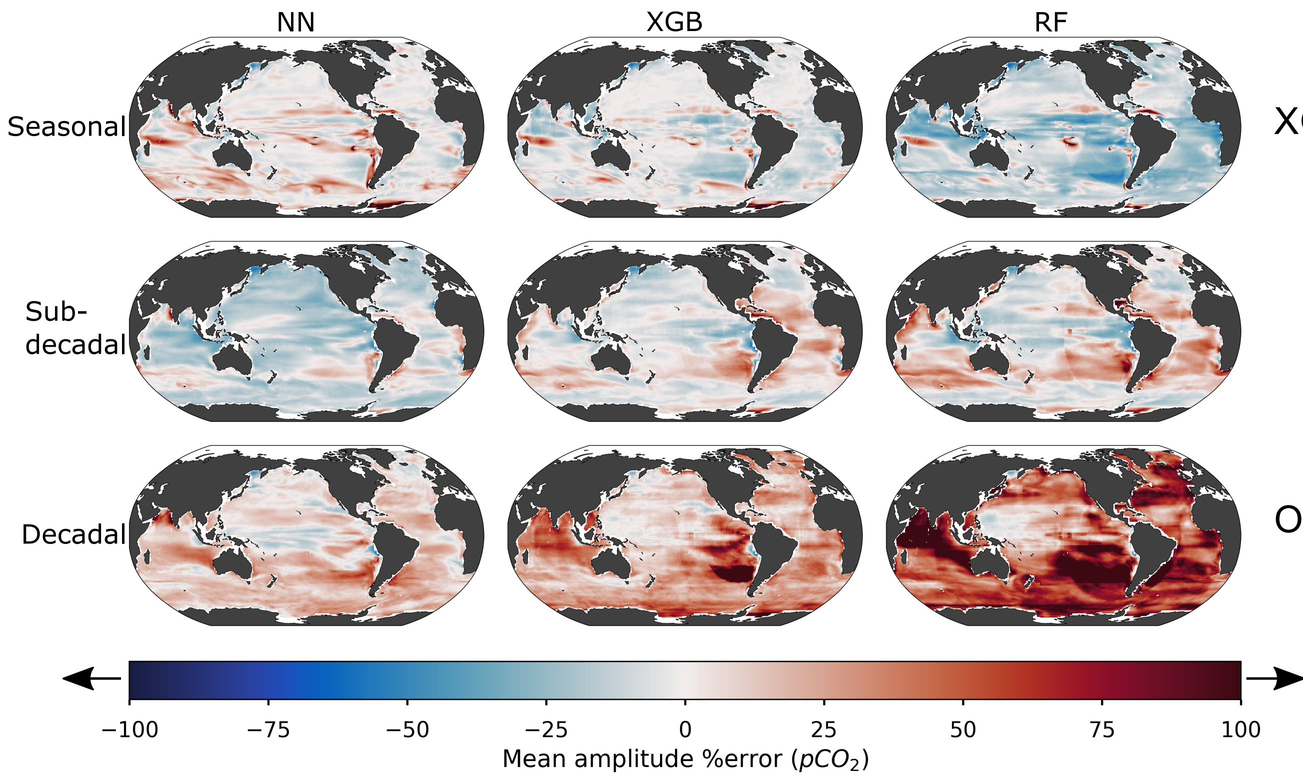




- Globally the seasonal cycle is overestimated by ~7%
- Overestimates decadal variability in Southern Ocean by ~39%

Gloege et al. GBC 2021

Other standard ML methods do well at seasonality, but overestimate decadal variability

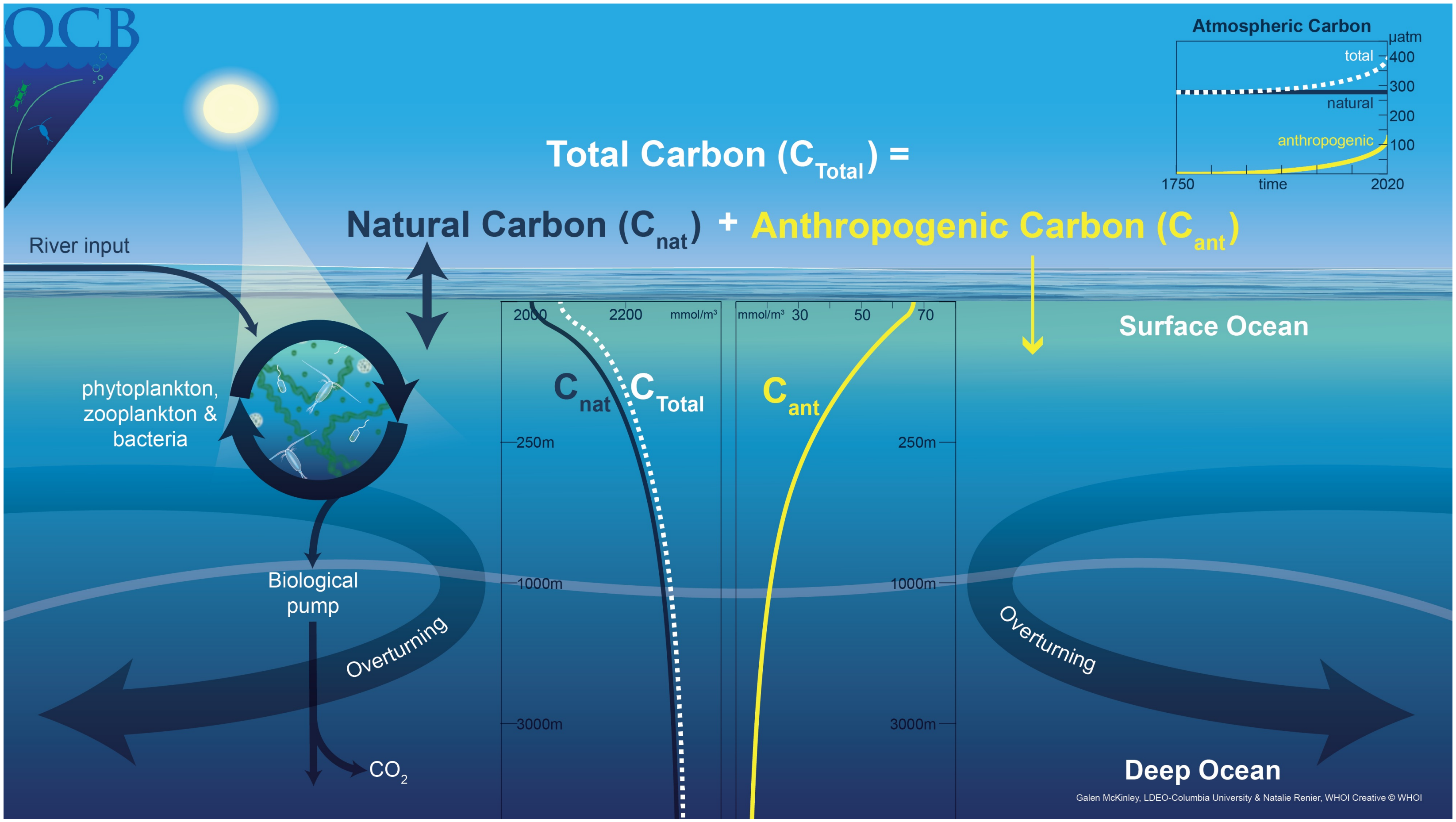


XGB best at seasonality

Our Neural Network best at decadal

Stamell et al. 2020

Natural carbon cycling and the anthropogenic ocean carbon sink



QUESTIONS?

Professor Galen McKinley

mckinley@ldeo.columbia.edu

Office hours: by appointment