Ocean carbon and biogeochemistry

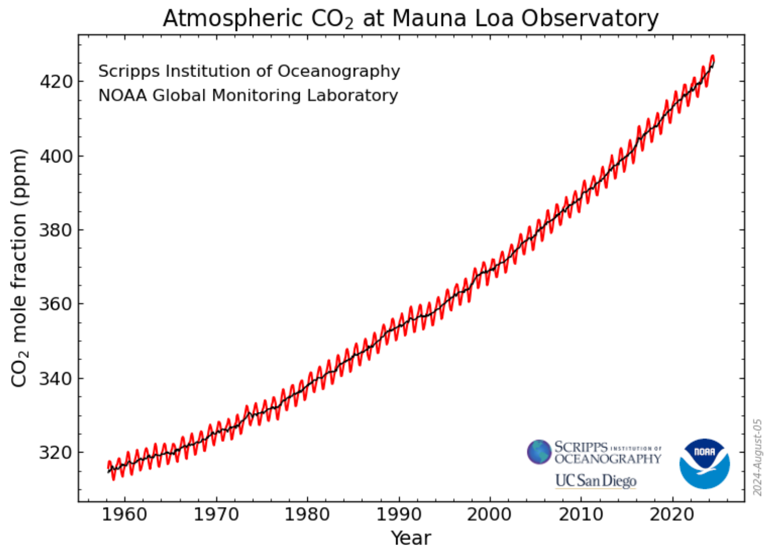
Galen A. McKinley

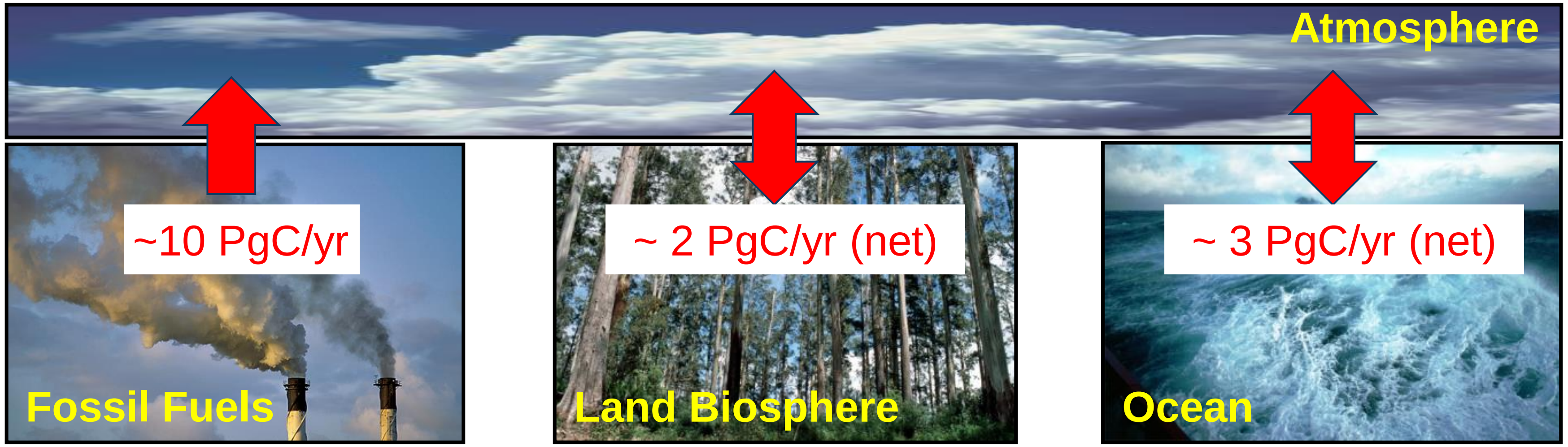
Professor of Earth and Environmental Sciences

Columbia University and Lamont-Doherty Earth Observatory

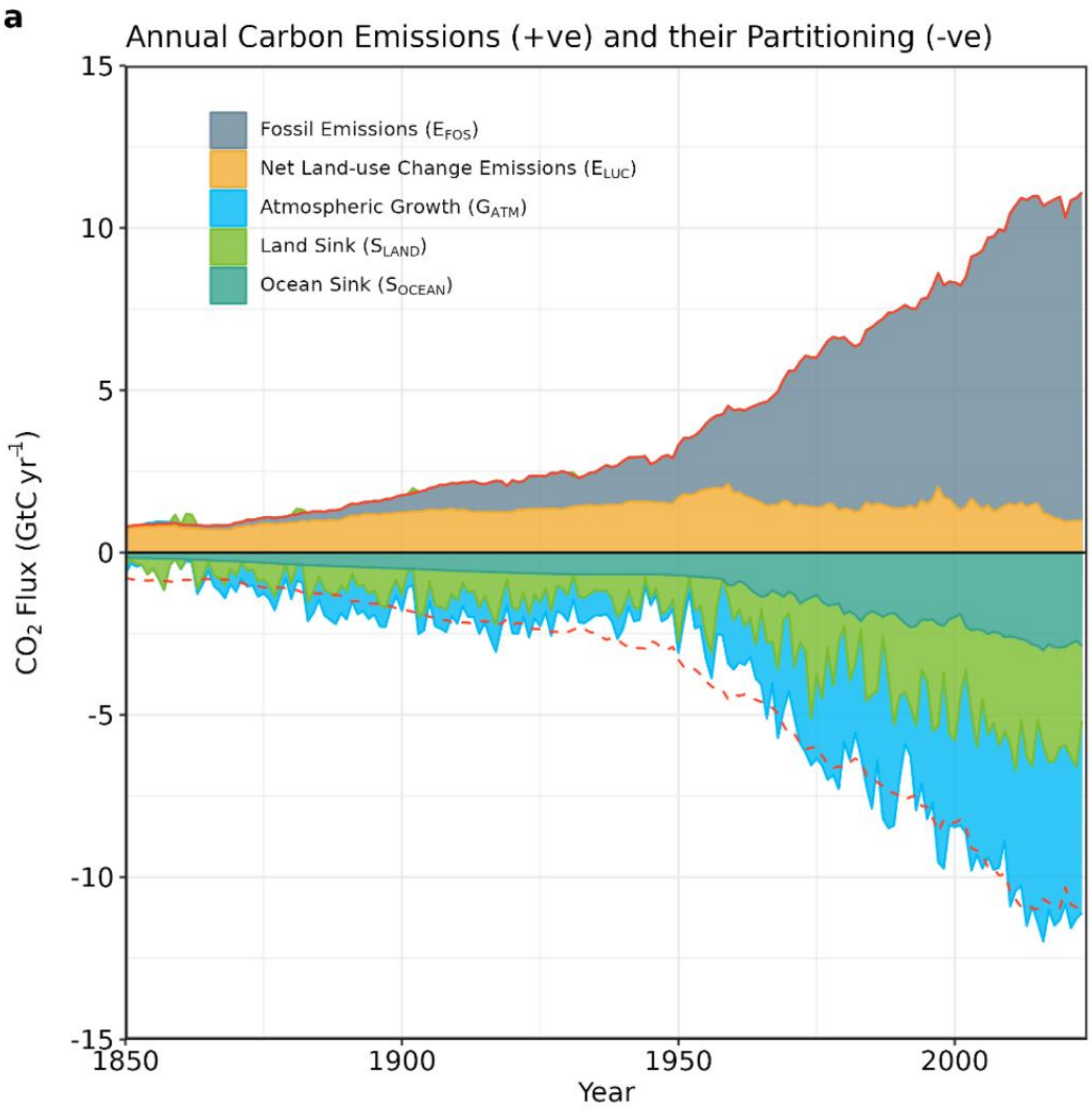
27 May 2025

Atmospheric CO_2 is rising due to human emissions
Ocean and terrestrial biosphere absorb ~50% of emissions





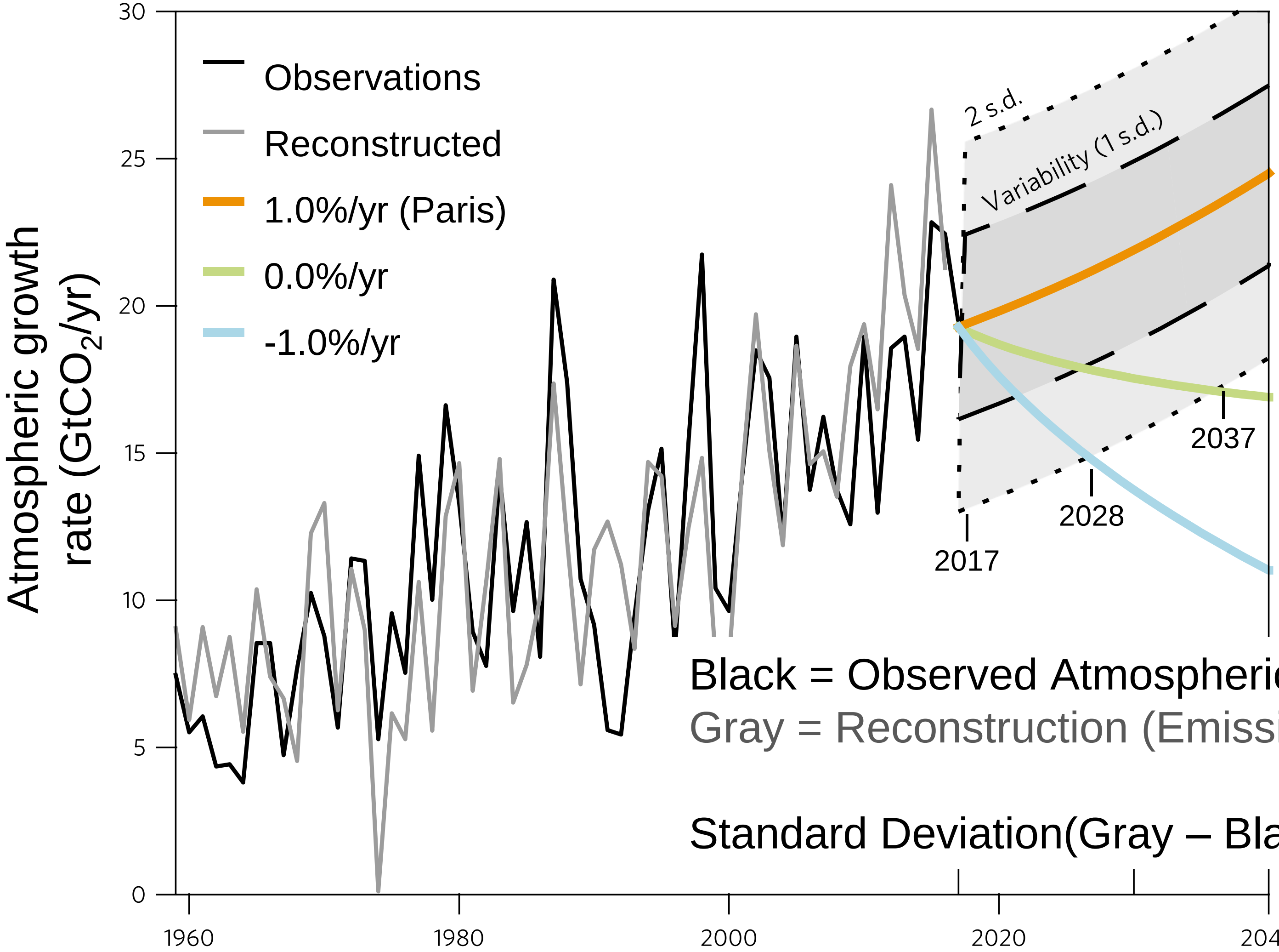
Global Carbon Budget



Friedlingstein et al. 2025, ESSD

Reduced uncertainty in ocean and land sinks required to usefully detect emissions mitigation

CO_2



Black = Observed Atmospheric Growth Rate

Gray = Reconstruction (Emission–Land–Ocean)

Standard Deviation(Gray – Black) = **3 $GtCO_2$/yr**

Peters et al. 2017 Nature Climate Change

Outline

- Key processes of the ocean carbon cycle

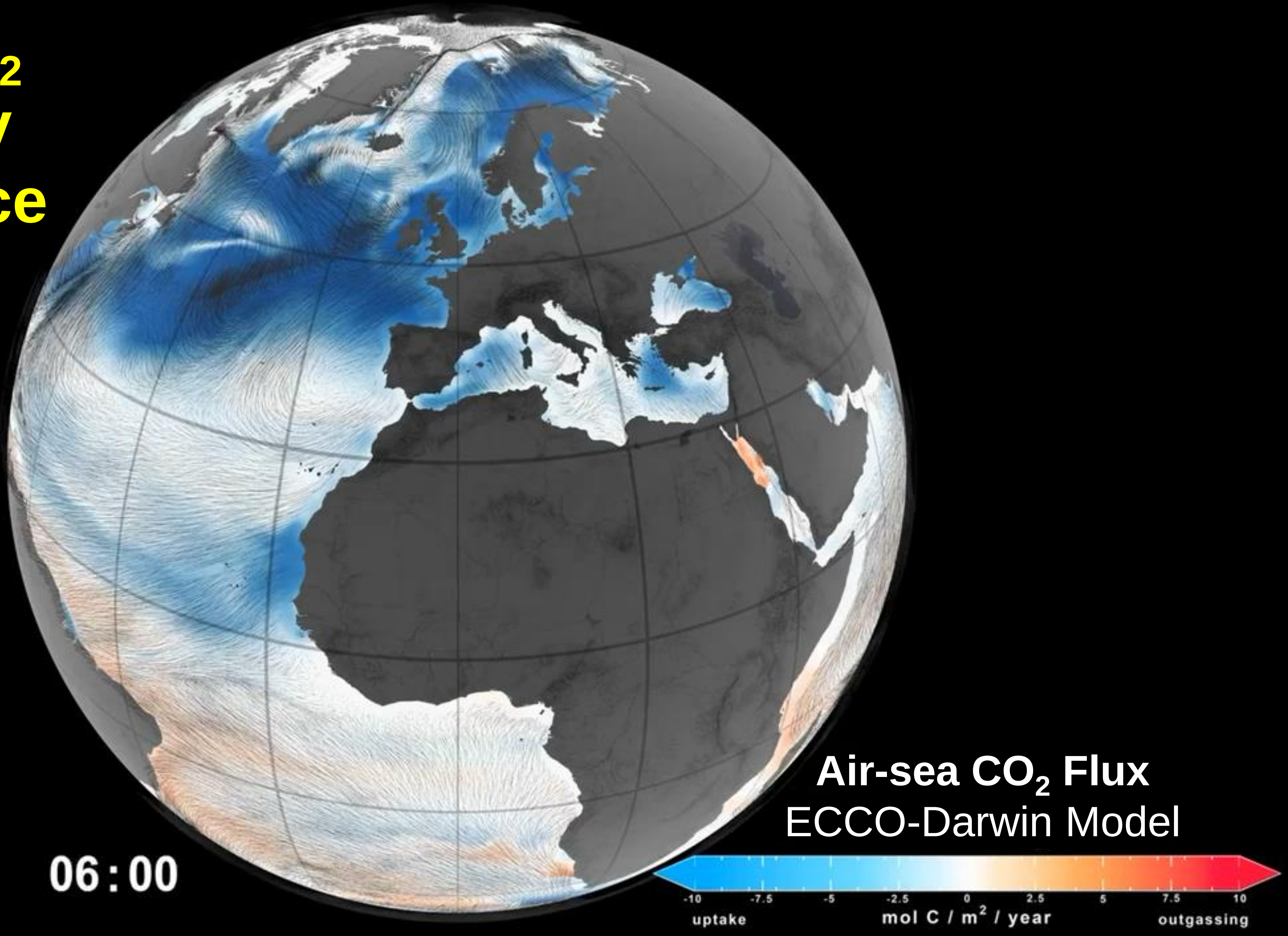
- Improving quantification of the ocean carbon sink
  - *Models*: evaluate ML-based data product skill, given sampling
  - *pCO_2 data products*: apply to identify model mean-state biases
  - *Models and data*: combine using ML
  - *marine Carbon Dioxide Removal (mCDR)* – quantifying "additionality"?

Outline

- **Key processes of the ocean carbon cycle**

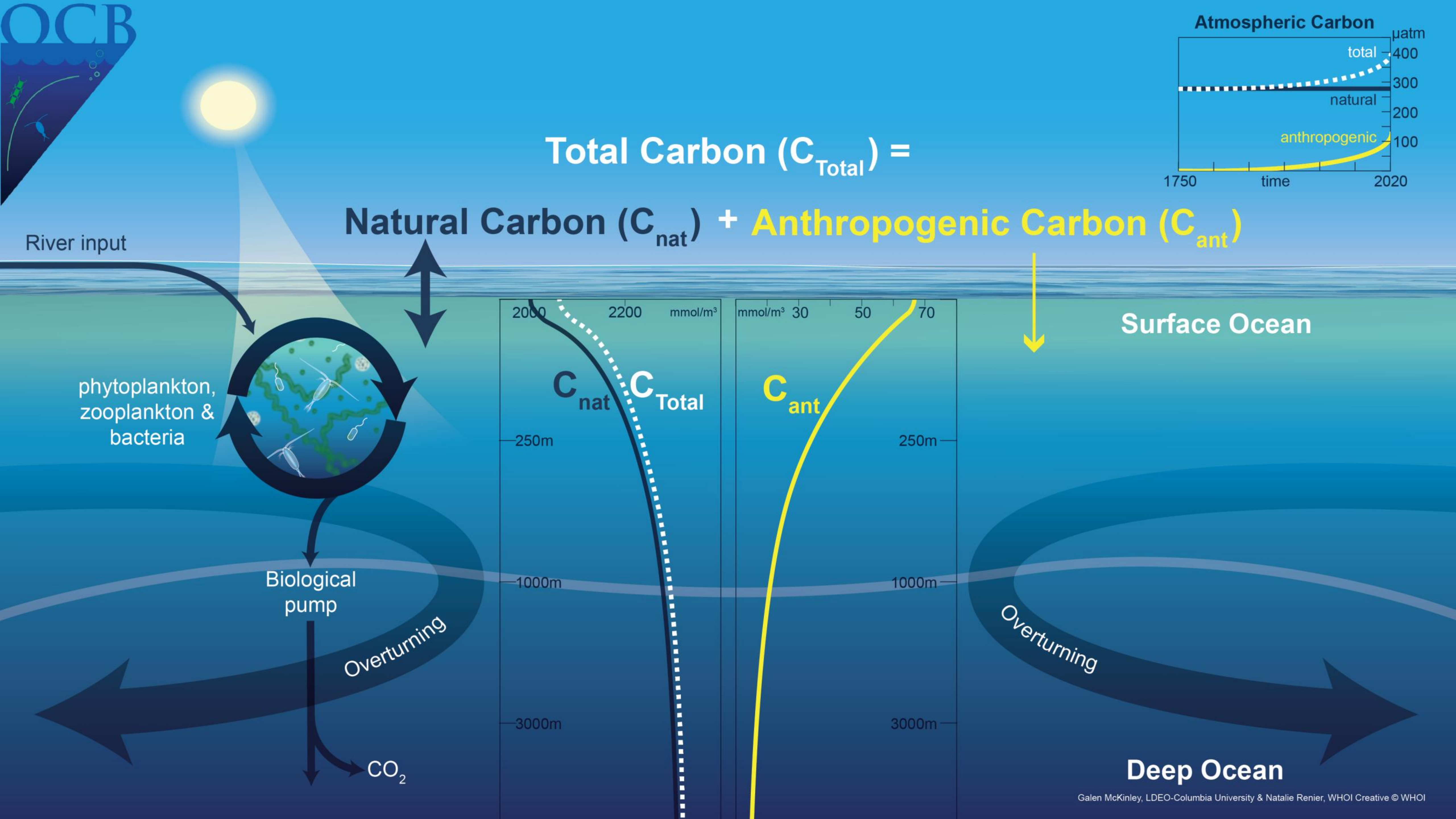
- Improving quantification of the ocean carbon sink
  - *Models*: evaluate ML-based data product skill, given sampling
  - *pCO_2 data products*: apply to identify model mean-state biases
  - *Models and data*: combine using ML
  - *marine Carbon Dioxide Removal (mCDR)* – quantifying “additionality”?

Air-sea CO_2 fluxes vary across space and time

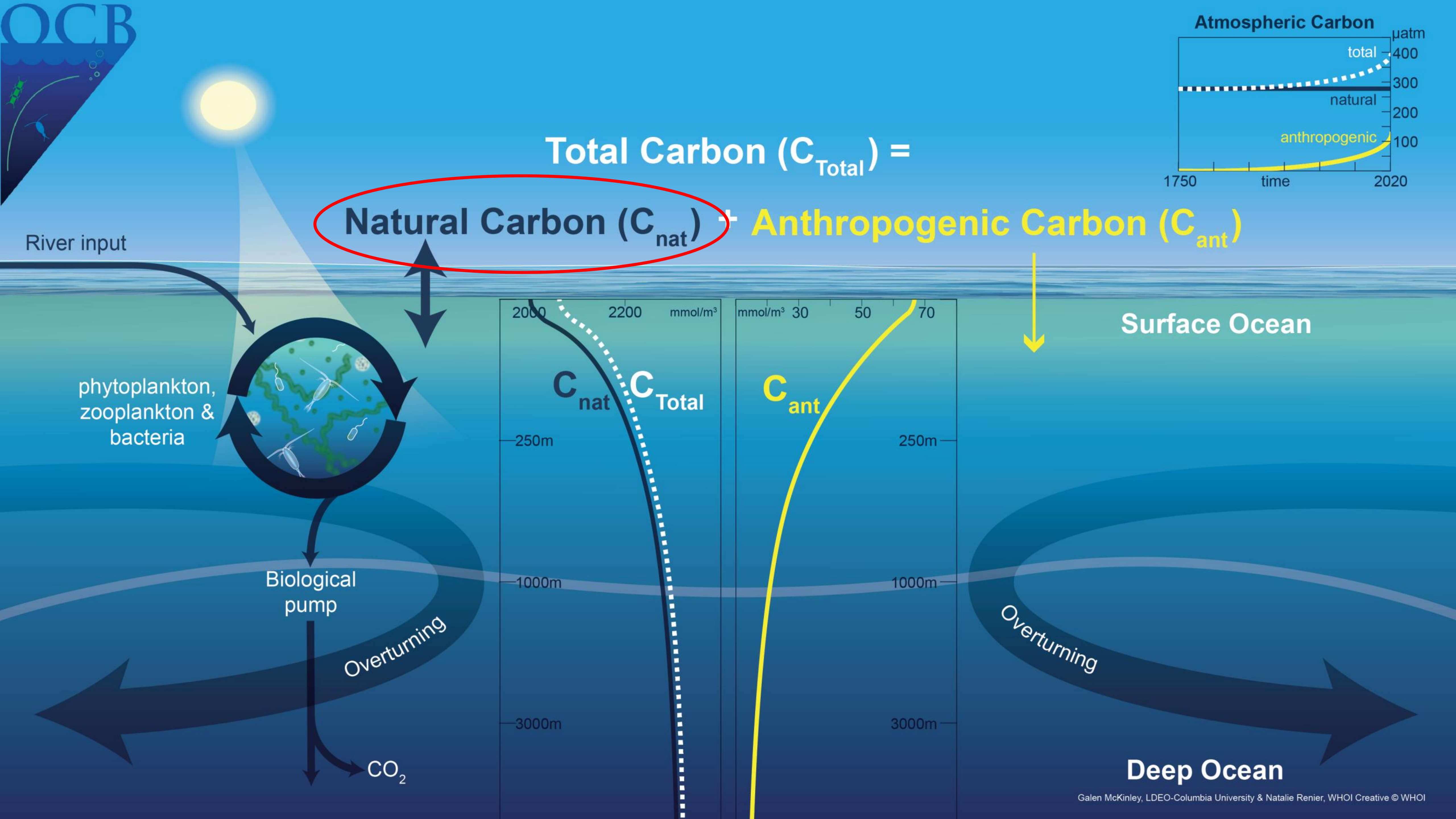


Carroll et al. 2020, 2022

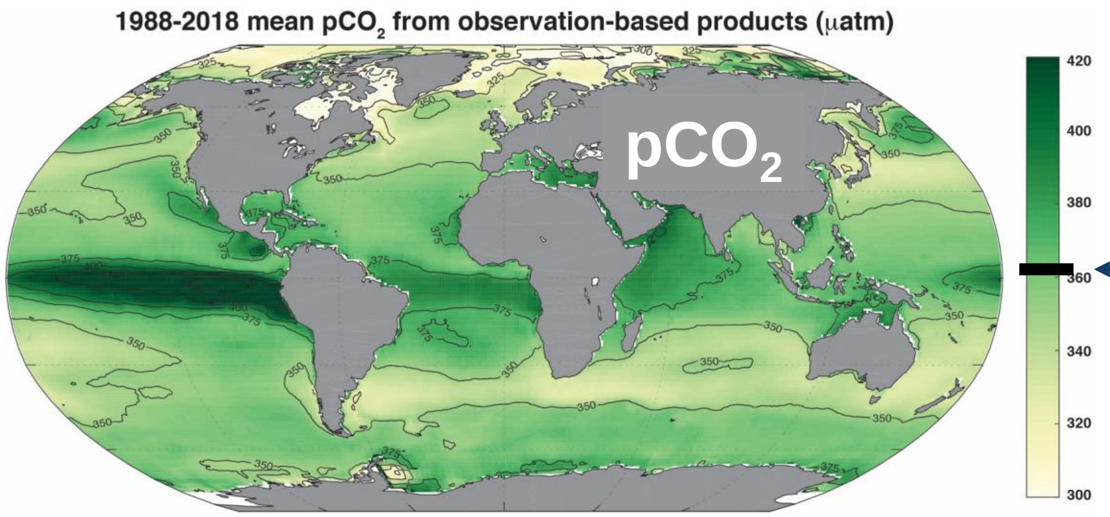
03 Jan 2012 06:00

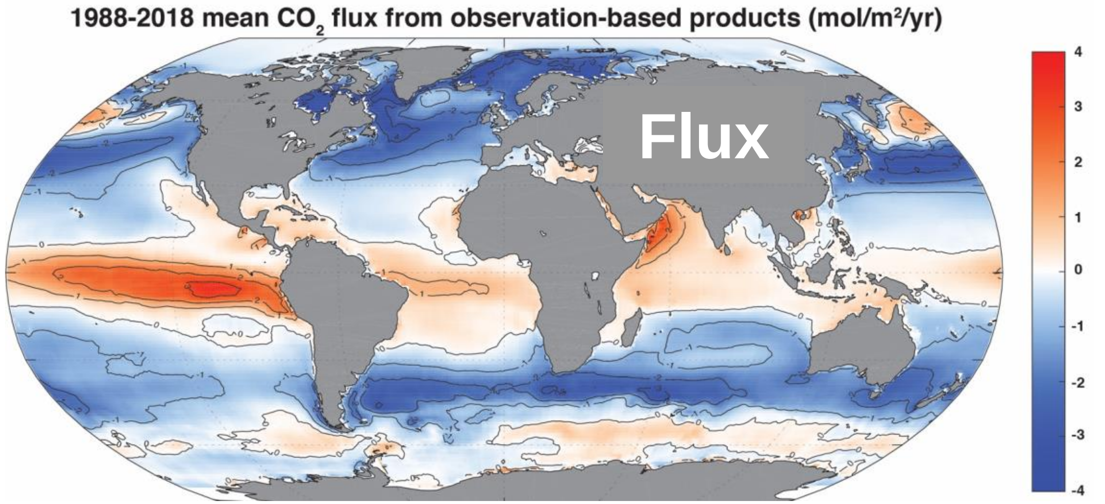


OCB
Atmospheric Carbon
µatm
total
natural
anthropogenic
400
300
200
100
1750
time
2020
Total Carbon (C_{Total}) =
Natural Carbon (C_{nat}) + Anthropogenic Carbon (C_{ant})
River input
phytoplankton, zooplankton & bacteria
Biological pump
CO_2
Overturning
Overturning
2000
2200
mmol/m³
C_{nat}
C_{Total}
mmol/m³
30
50
70
C_{ant}
250m
1000m
3000m
250m
1000m
3000m
Surface Ocean
Deep Ocean
Galen McKinley, LDEO-Columbia University & Natalie Renier, WHOI Creative © WHOI


OCB
Atmospheric Carbon
µatm
total
natural
anthropogenic
400
300
200
100
1750
time
2020
Total Carbon (C_{Total}) =
Natural Carbon (C_{nat}) + Anthropogenic Carbon (C_{ant})
River input
phytoplankton, zooplankton & bacteria
Biological pump
CO_2
Overturning
2000
2200
mmol/m³
C_{nat}
C_{Total}
250m
1000m
3000m
mmol/m³
30
50
70
C_{ant}
Surface Ocean
Overturning
Deep Ocean
Galen McKinley, LDEO-Columbia University & Natalie Renier, WHOI Creative © WHOI

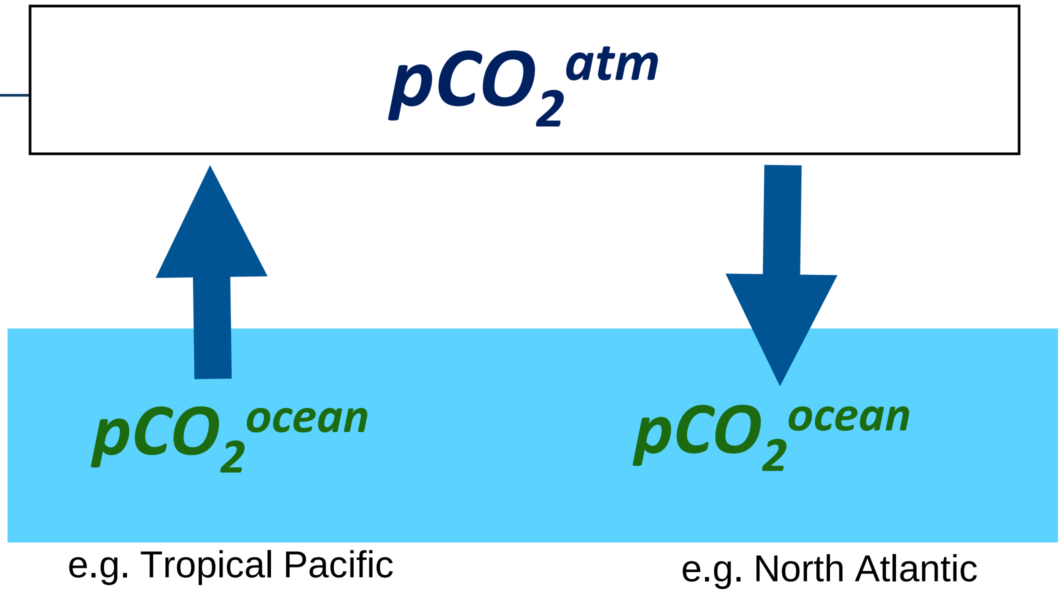
Surface ocean pCO_2 is primary control on air-sea CO_2 flux





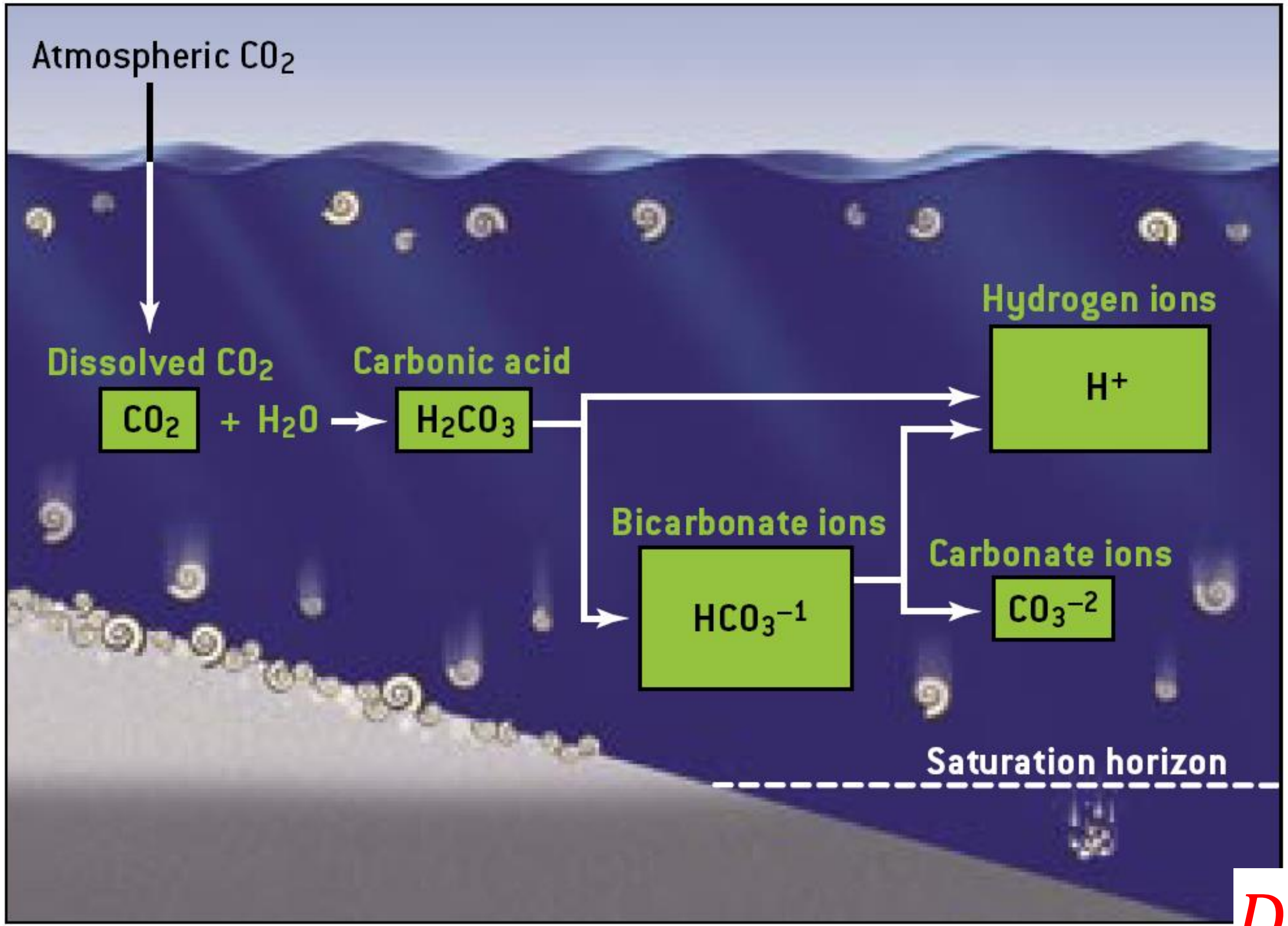
Crisp et al., 2022, Review of Geophysics

If $pCO_2^{ocean} > pCO_2^{atm}$ flux is out of ocean (positive)
If $pCO_2^{ocean} < pCO_2^{atm}$ flux is into ocean (negative)



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

Ocean Inorganic Carbon Cycle



Doney 2006 Scientific American

$$CO_2^* = H_2CO_3 + CO_2(aq)$$

$$CO_2^* + H_2O \Leftrightarrow HCO_3^- + H^+$$

$$K_1' = \frac{[HCO_3^-][H^+]}{[CO_2^*]}$$

$$HCO_3^- \Leftrightarrow CO_3^{2-} + H^+$$

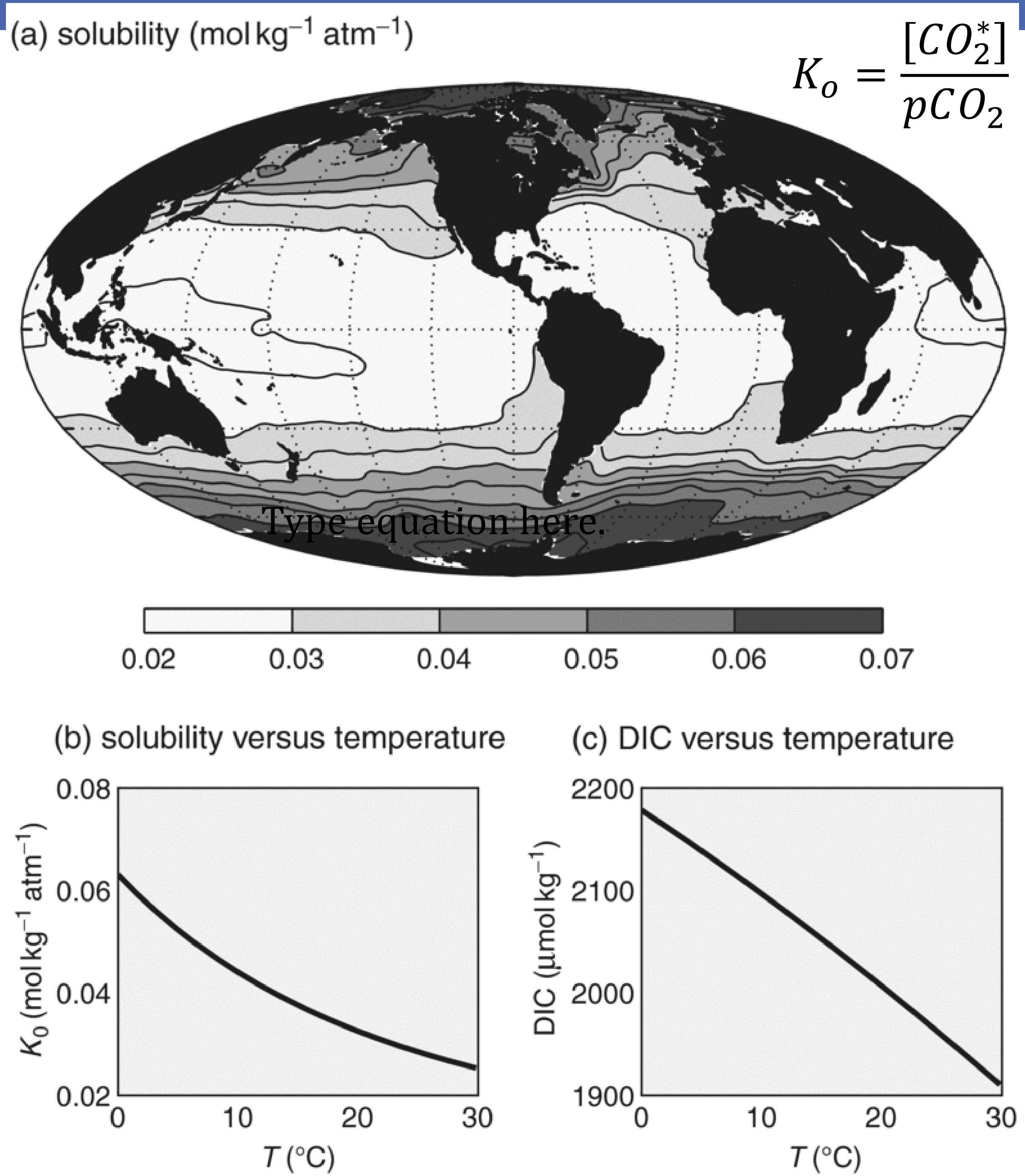
$$K_2' = \frac{[CO_3^{2-}][H^+]}{[HCO_3^-]}$$

$$DIC = CO_2^* + HCO_3^- + CO_3^{2-}$$

SOLUBILITY

$$pCO_2 = \frac{[CO_2^*]}{K_o}$$

$$K_o = \frac{[CO_2^*]}{pCO_2}$$

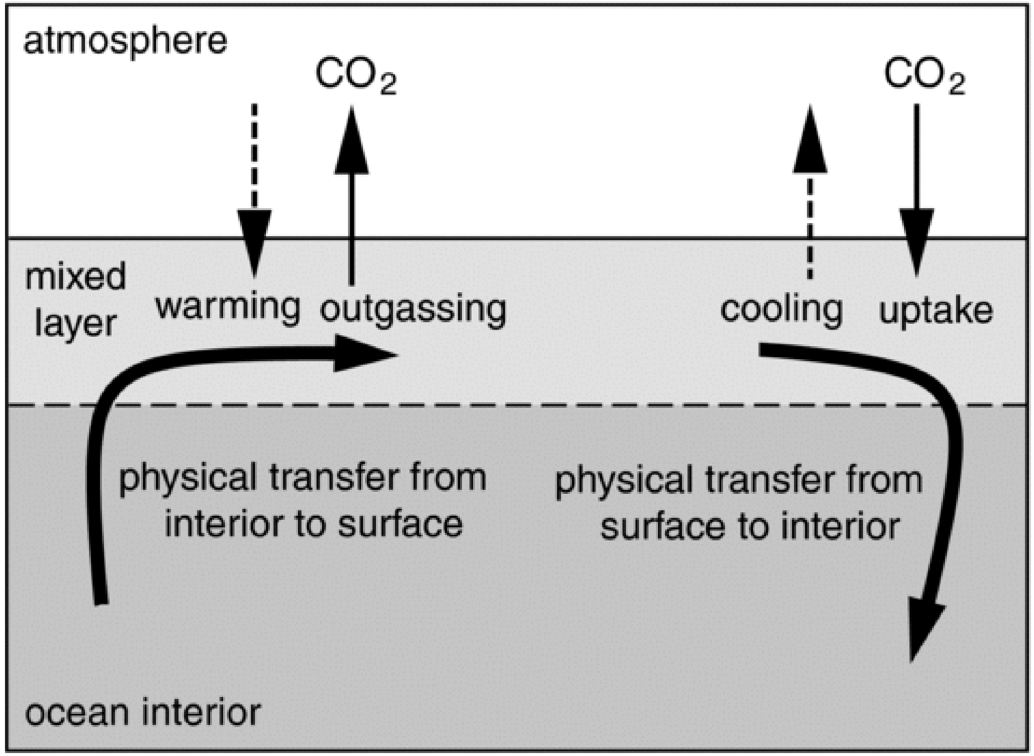


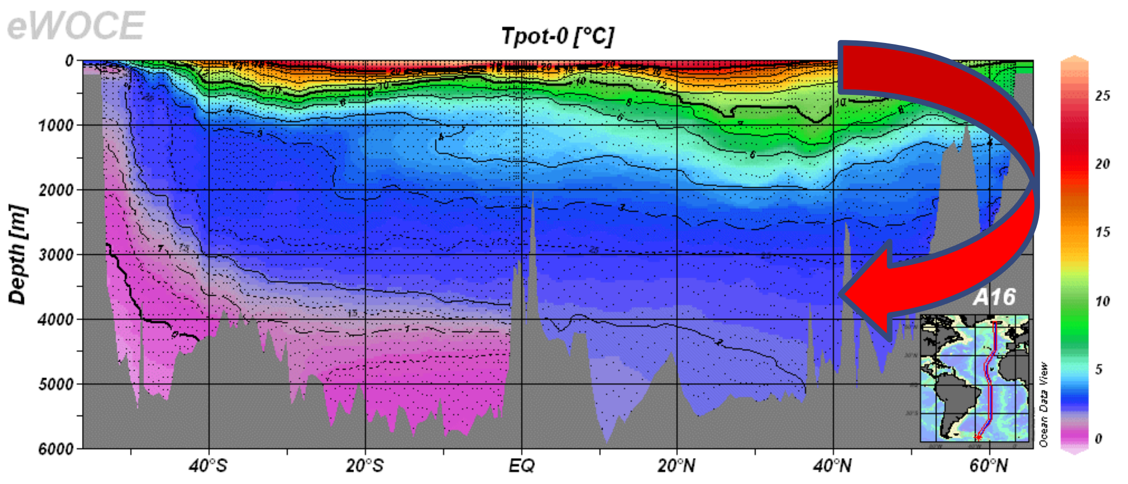
K_o is an inverse function of temperature

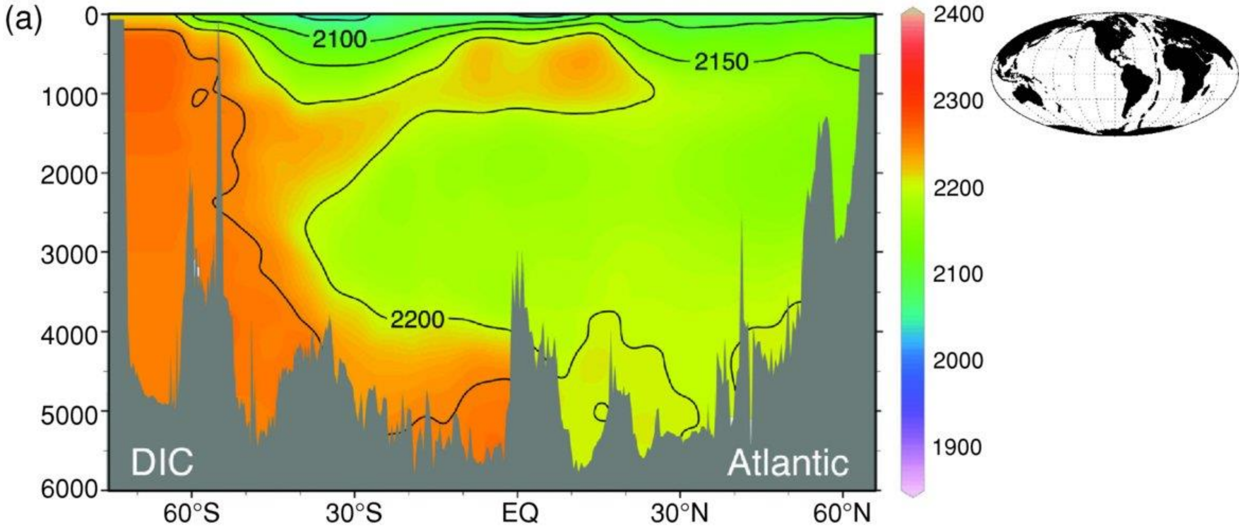
For same $[CO_2^*]$, pCO_2 is higher at high temperature, thus warm waters outgas and cold waters absorb CO_2

Solubility and physical transport

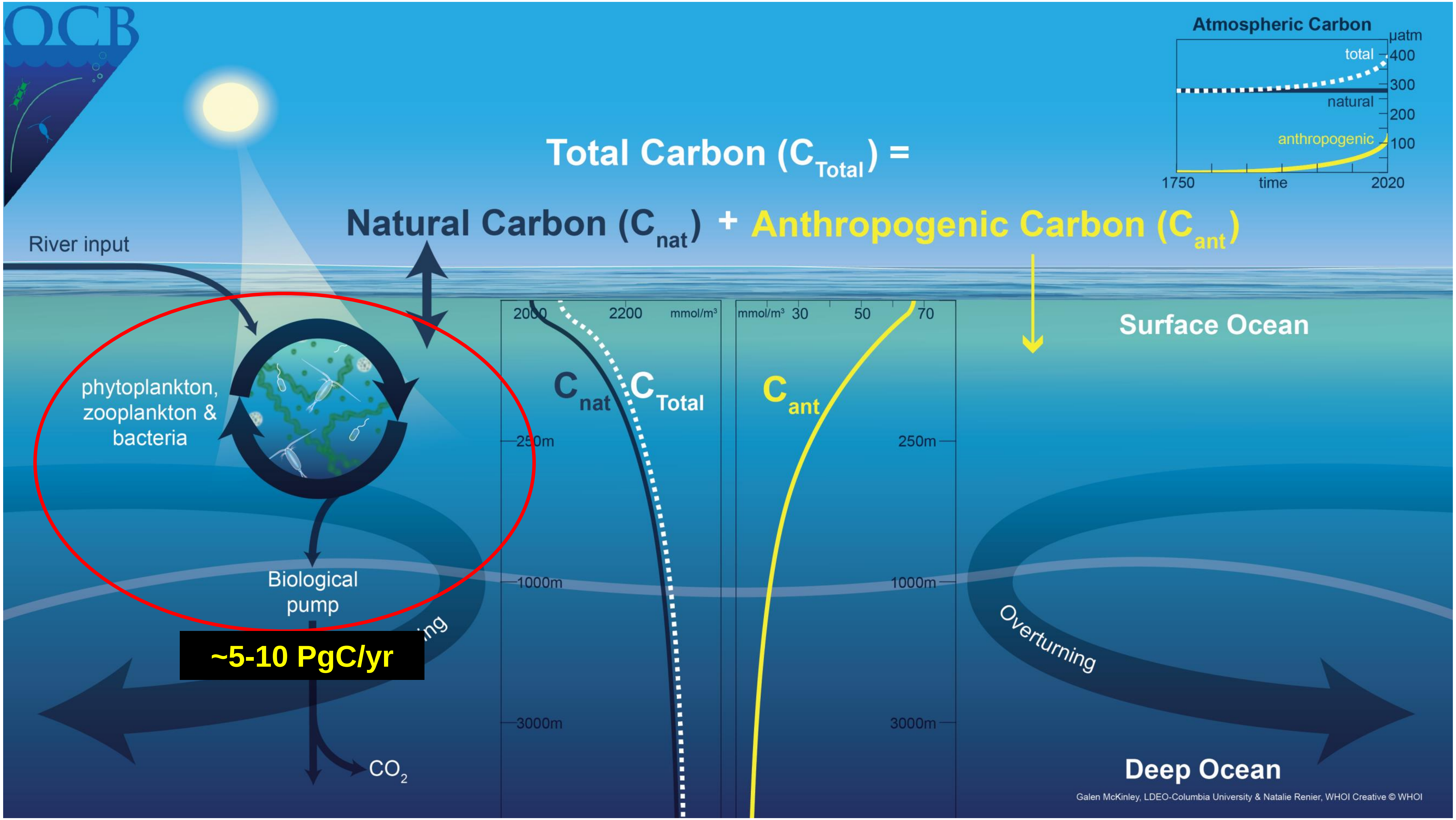
(a) physical transport and solubility



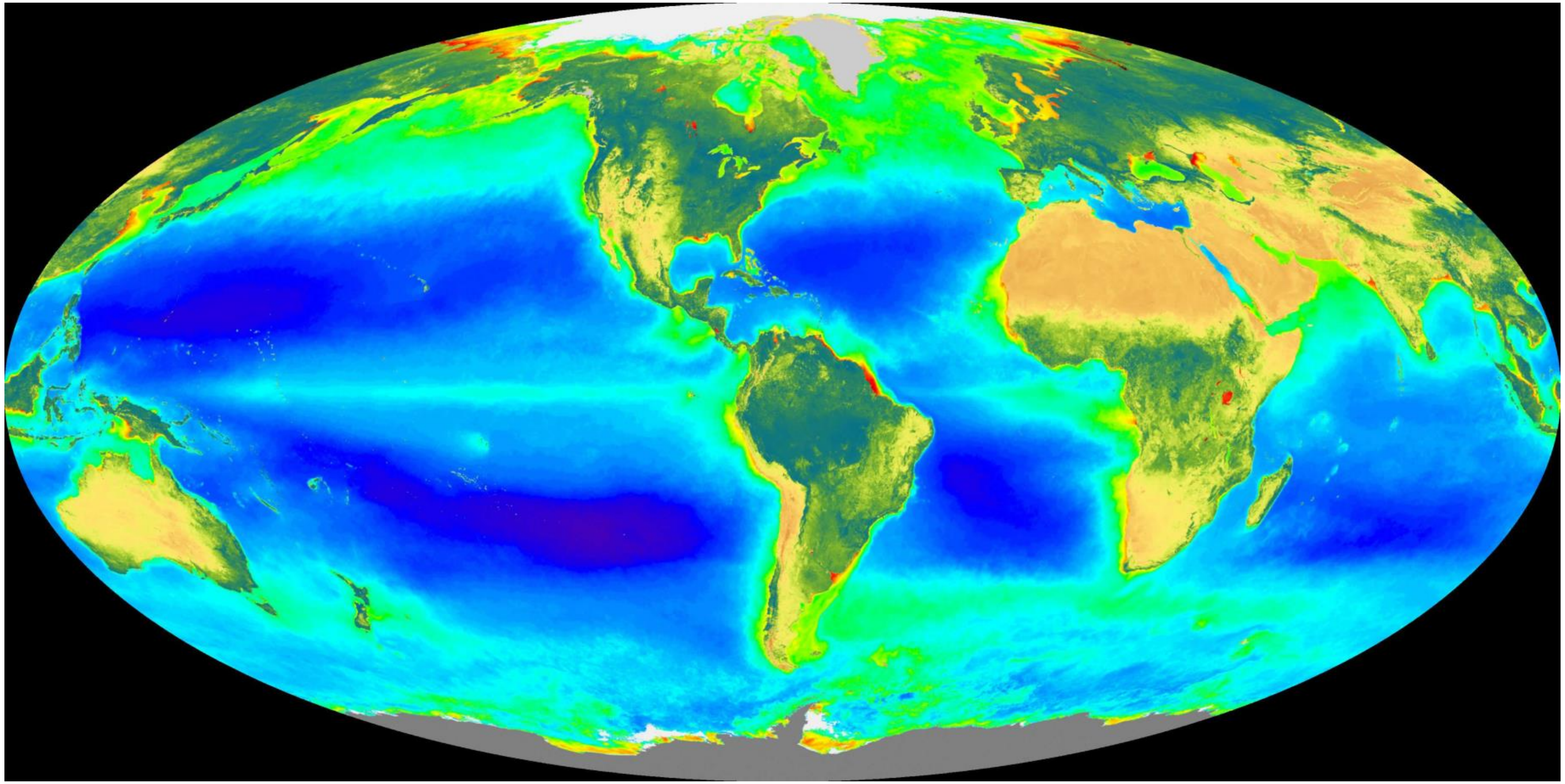




Williams_Fig. 2.24

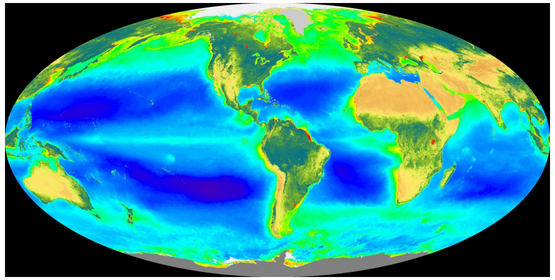

OCB
Atmospheric Carbon
μatm
total
natural
anthropogenic
400
300
200
100
1750
time
2020
Total Carbon (C_{Total}) =
Natural Carbon (C_{nat}) + Anthropogenic Carbon (C_{ant})
River input
phytoplankton, zooplankton & bacteria
Biological pump
~5-10 PgC/yr
CO_2
2000
2200
mmol/m³
C_{nat}
C_{Total}
mmol/m³
30
50
70
C_{ant}
250m
1000m
3000m
Surface Ocean
Overturning
Deep Ocean
Galen McKinley, LDEO-Columbia University & Natalie Renier, WHOI Creative © WHOI

Ocean primary productivity from NASA SeaWiFS satellite

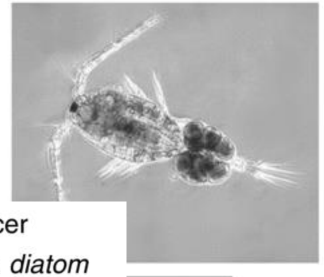


Dark blue = very low chlorophyll; green = moderate chlorophyll; red = very high

Biological carbon cycling in the ocean

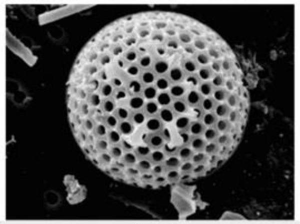


grazer
zooplankton, copepod

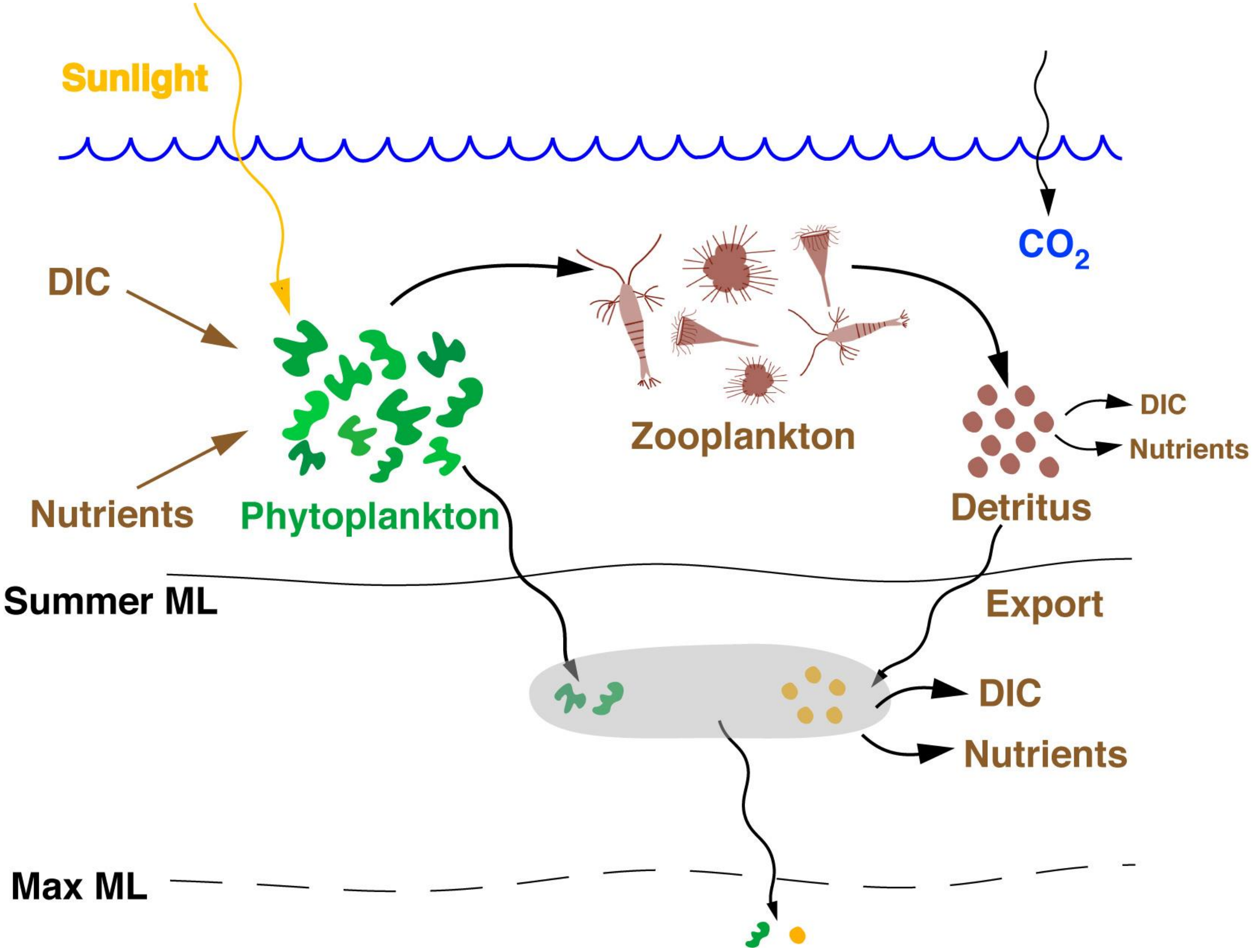


1 mm

primary producer
phytoplankton, diatom

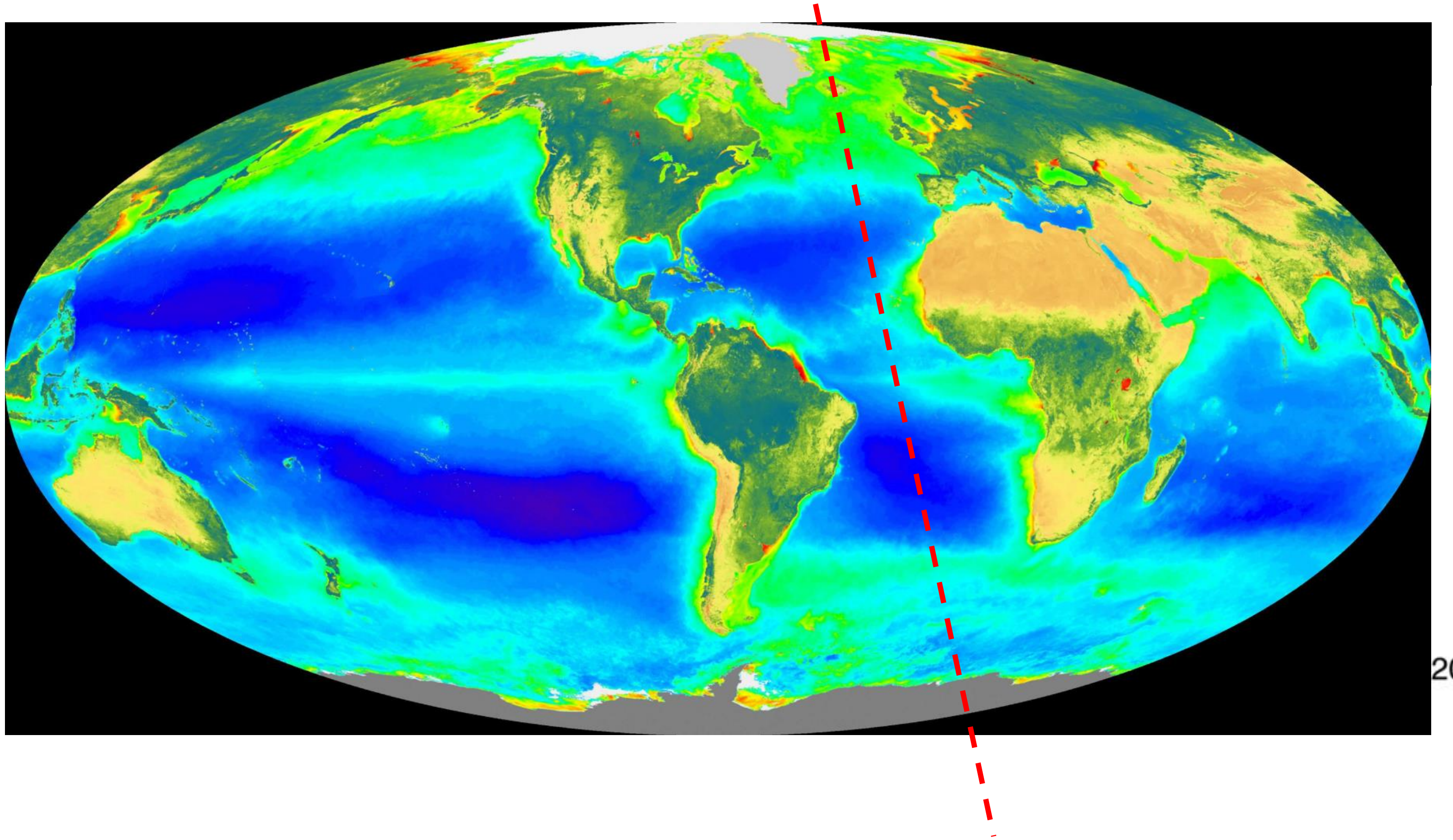


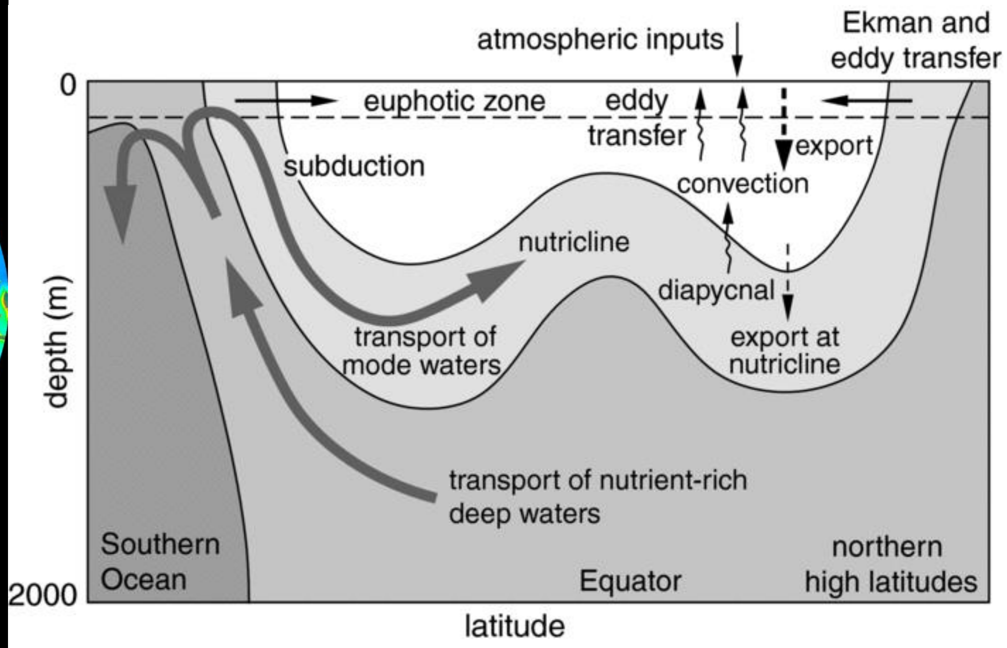
5 µm



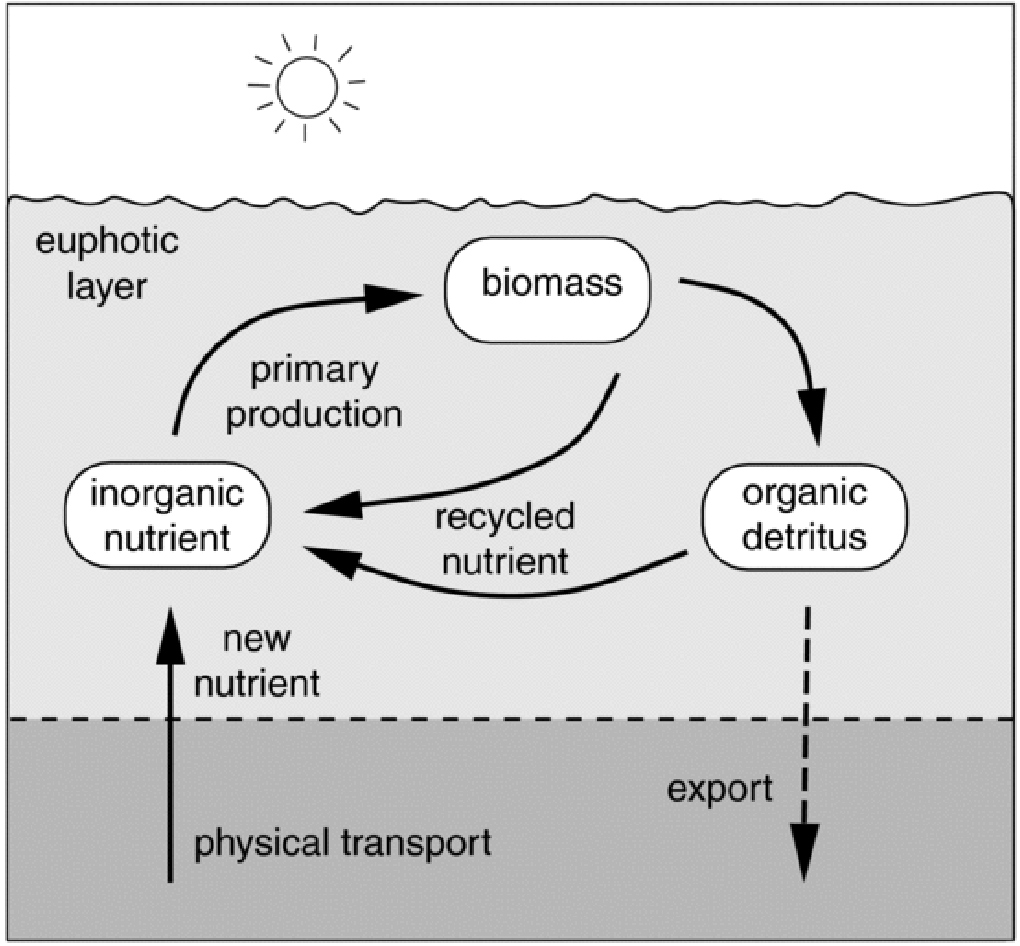
DIC = Dissolved Inorganic Carbon

Productivity is structured by nutrients distributions, in turn set by the circulation



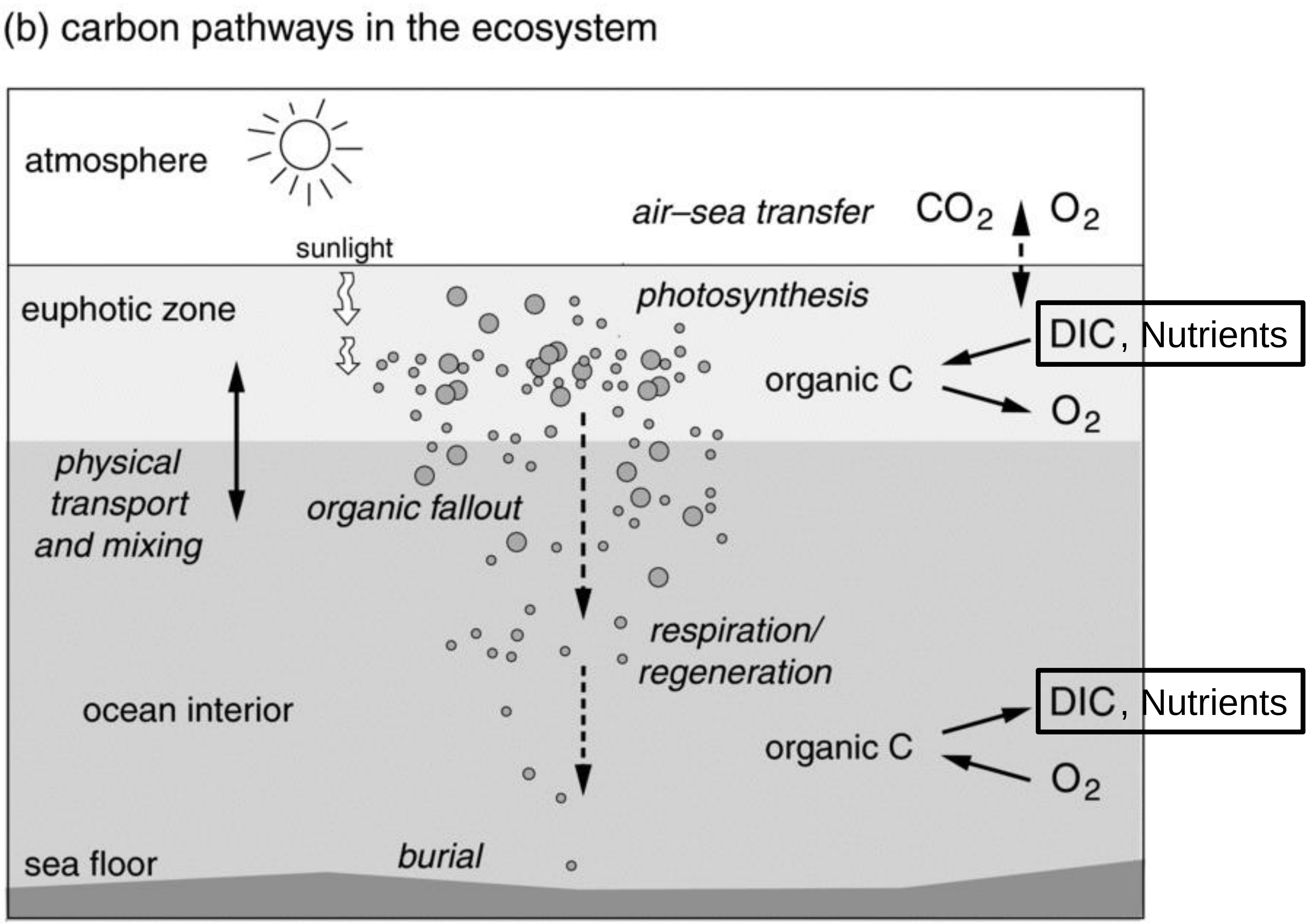


Nutrient cycling between biomass and inorganic nutrients



Williams_Fig. 5.15

Biogeochemical Cycling = Coupled carbon and nutrient cycles



DIC = Dissolved Inorganic Carbon

Williams_Fig. 5.15

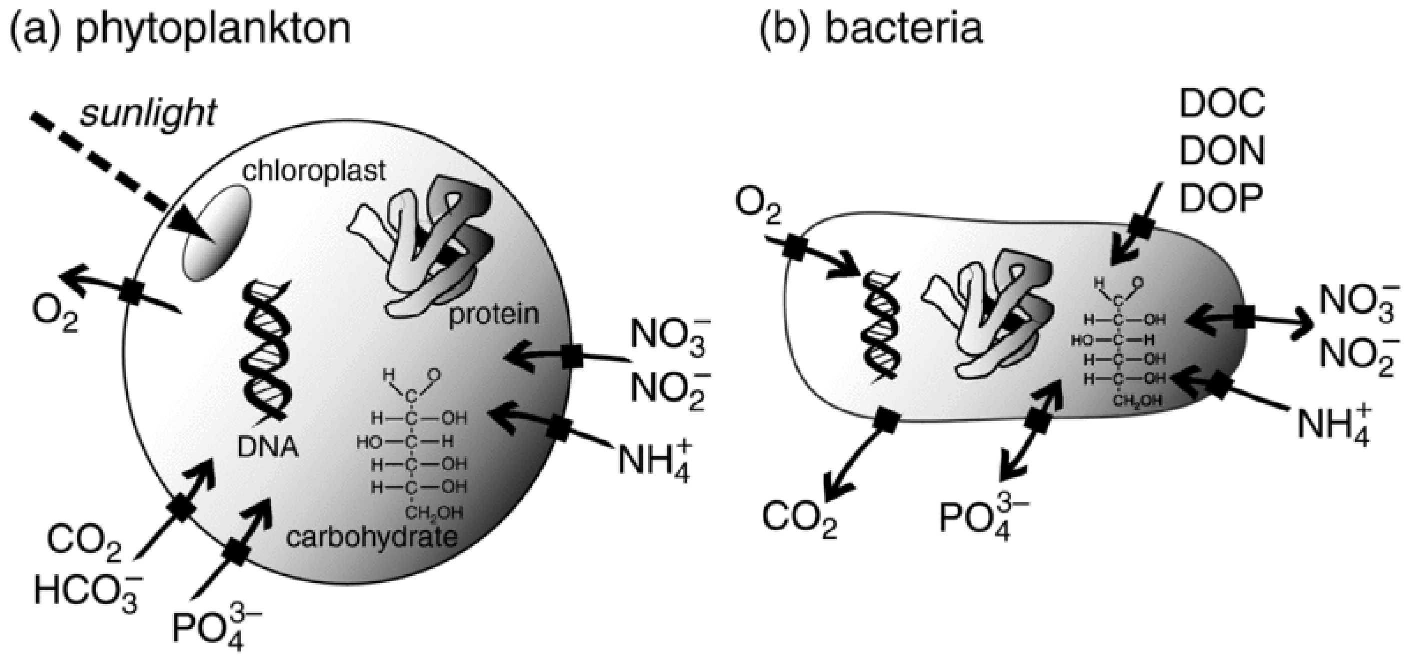
Simplified photosynthesis

$CO_2 + H_2O + \text{energy} \longrightarrow CH_2O + O_2$

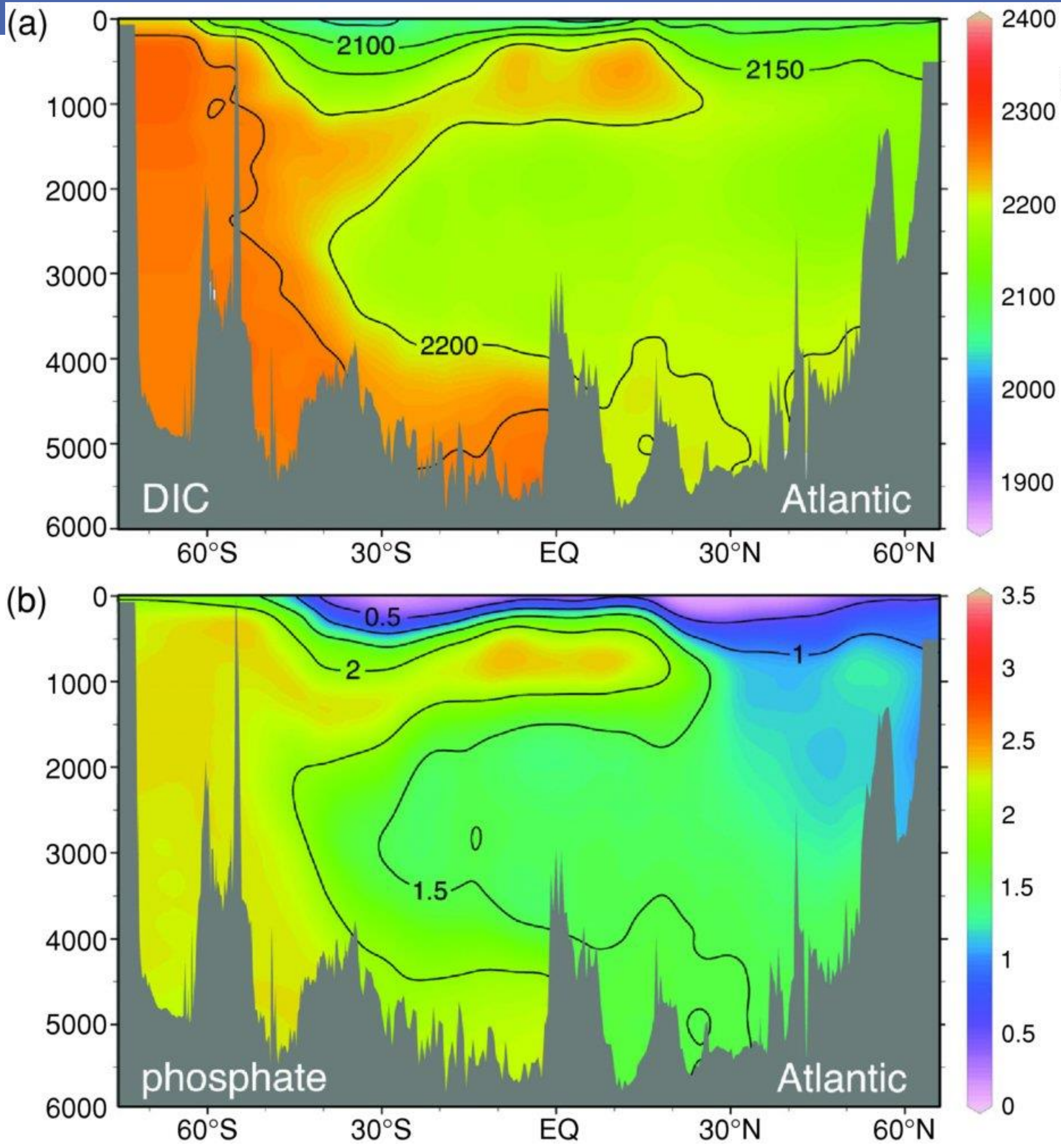
More realistic

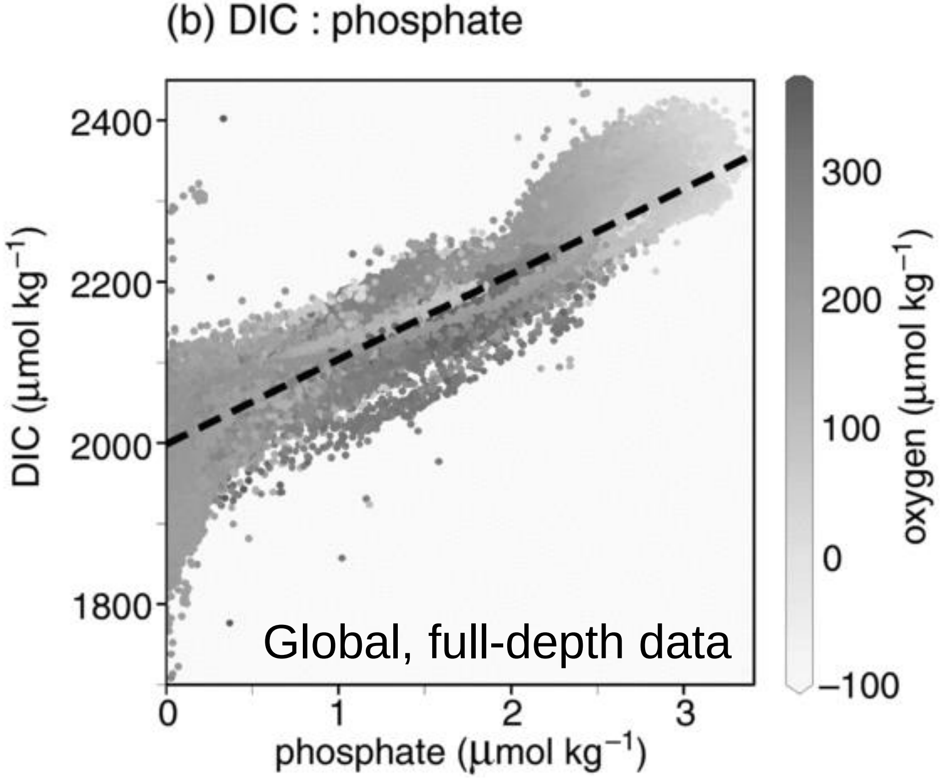
$106CO_2 + 16NO_3^- + H_2PO_4^- + 122H_2O + \text{energy}$
$\longrightarrow C_{106}H_{246}O_{110}N_{16}P + 138O_2$

Quantitative links between Carbon, Nutrients (N,P), Oxygen



Williams_Fig. 5.2





Williams_Plate-6

Table 5.2 The ratios of key elements in dissolved inorganic form in the water column and in marine phytoplankton and zooplankton ([a]Redfield *et al.*, 1963). The cultured phytoplankton data are the average of 15 species of estuarine, coastal and open ocean phytoplankton grown under controlled laboratory conditions ([b]Ho *et al.*, 2003). *Prochlorococcus* MED4 is a strain of tiny phytoplankton (about a micron in scale). Elemental ratios were measured in *Prochlorococcus* cultures grown under phosphorus replete and phosphorus limited conditions ([c]Bertilsson *et al.*, 2003). Ranges of values are also shown for isolated marine bacteria grown under various conditions of nutrient limitation ([d]Vrede *et al.*, 2002).

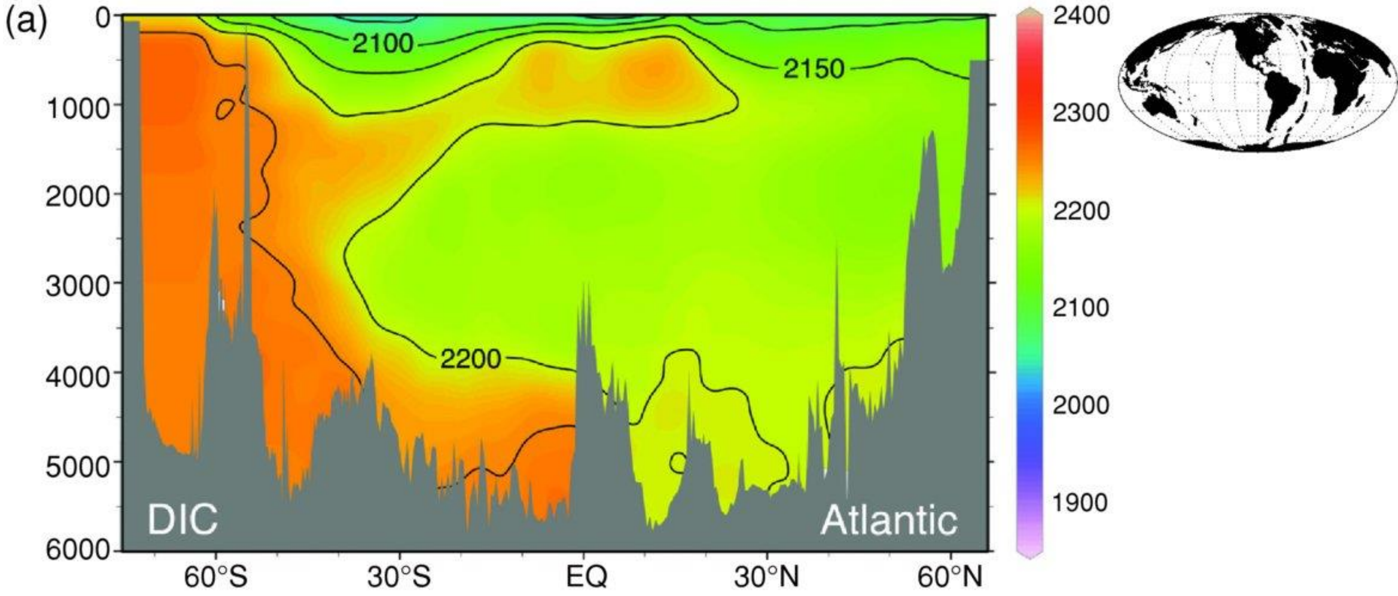
Reservoir	C	N	P
[a]Marine inorganic nutrients		15	1
[a]Bulk marine organic matter	106	16	1
[b]Cultured phytoplankton	147 ± 19	16 ± 2	1 ± 0.2
[c]*Prochlorococcus* MED4 (P-replete)	121 ± 17	21.2 ± 4.5	1
[c]*Prochlorococcus* MED4 (P-limited)	464 ± 28	62.3 ± 14.1	1
[d]Marine bacteria	35–178	7–18	1

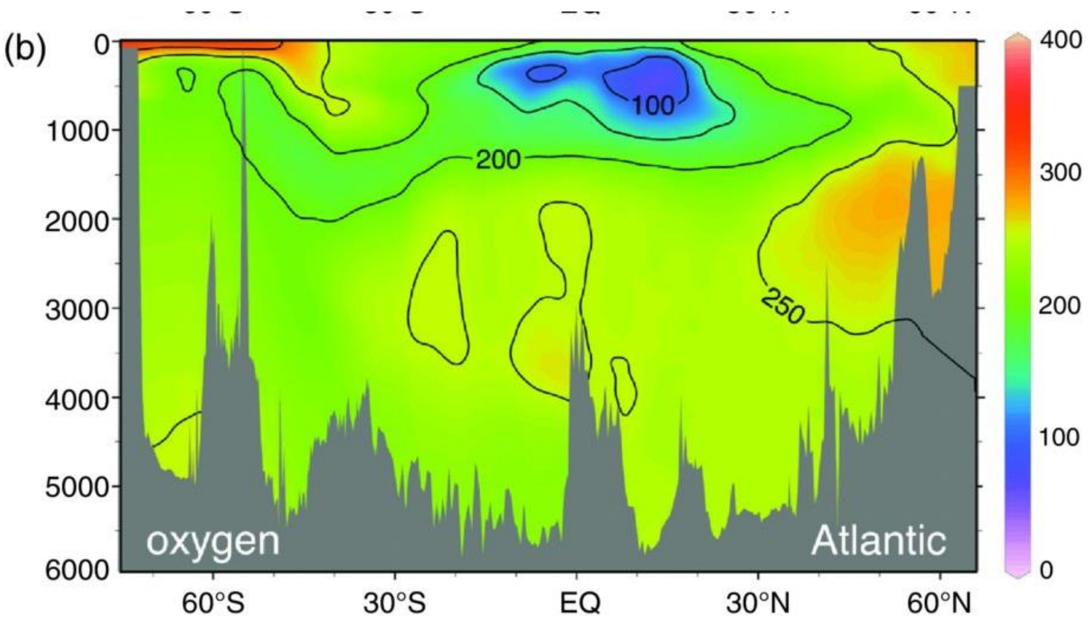
Updated Redfield

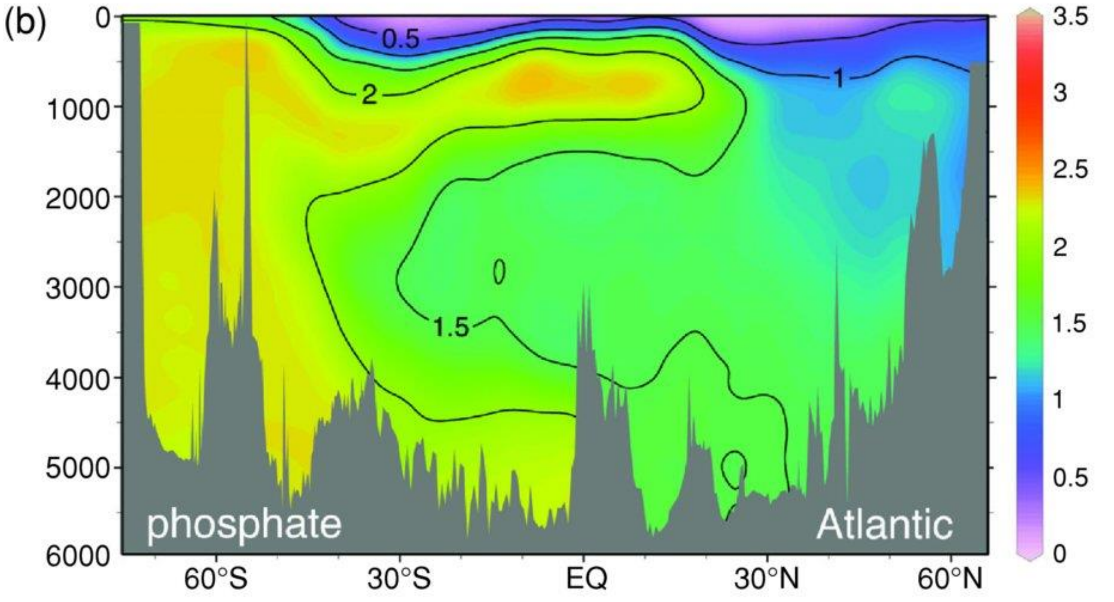
$$C:N:P:-O_2 = 117(\pm 14):16(\pm 1):1:-170(\pm 10)$$

Anderson and Sarmiento, 2004

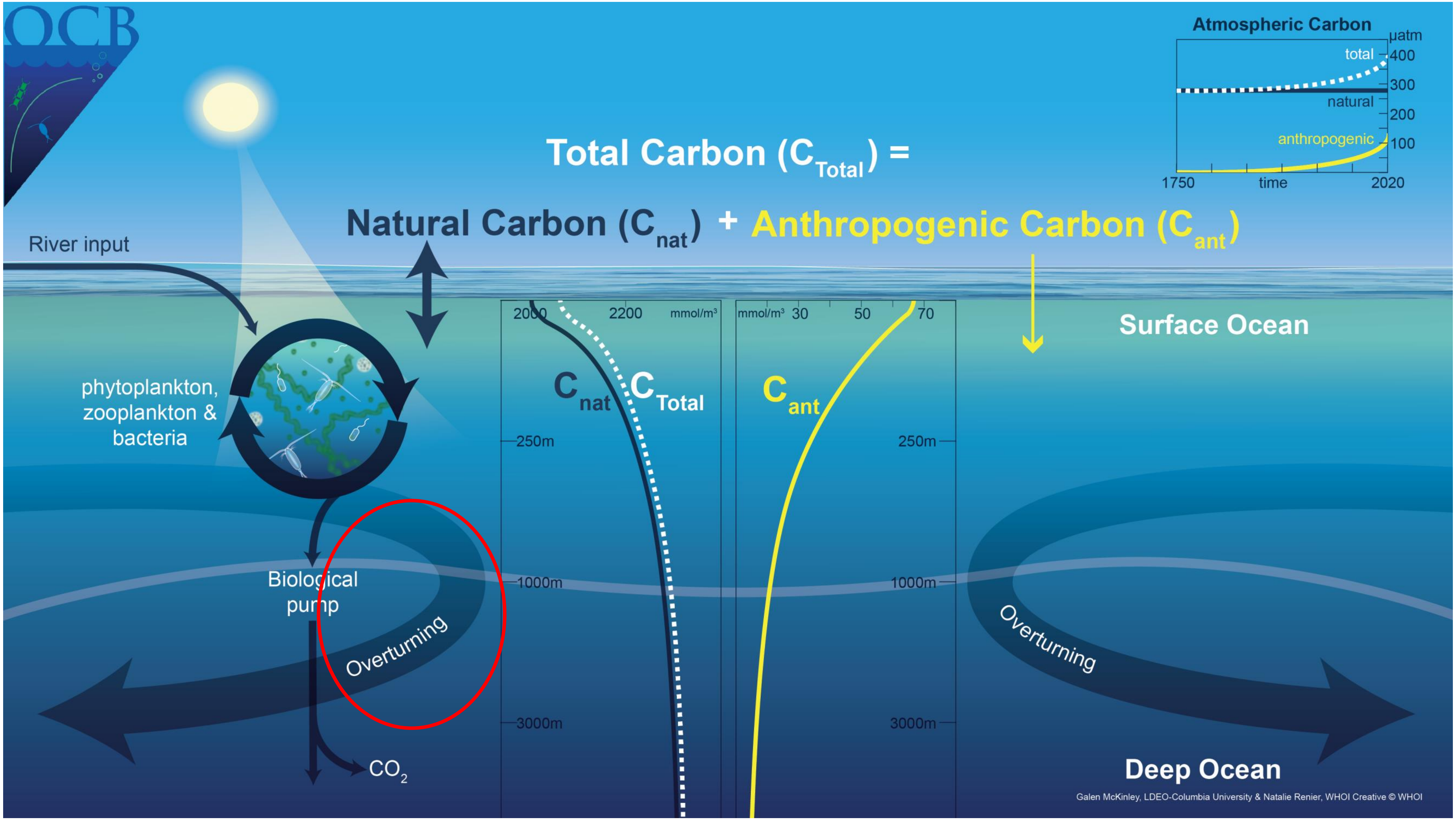
Oxygen distribution is inverse of nutrients and DIC







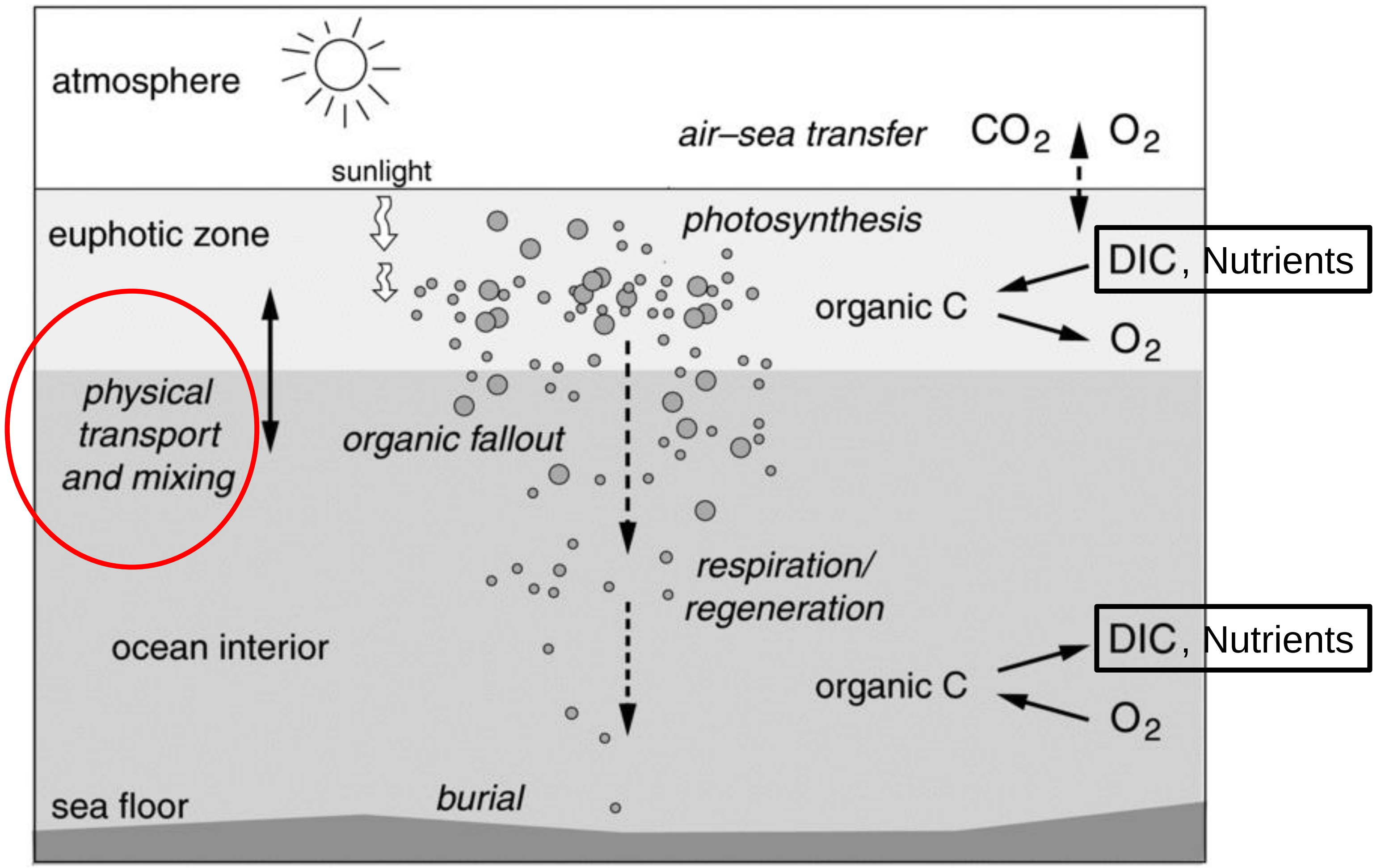
Williams_Plate-6



OCB
Atmospheric Carbon
µatm
total
natural
anthropogenic
400
300
200
100
1750
time
2020
Total Carbon (C_{Total}) =
Natural Carbon (C_{nat}) + Anthropogenic Carbon (C_{ant})
River input
phytoplankton, zooplankton & bacteria
Biological pump
CO_2
Overturning
2000
2200
mmol/m³
C_{nat}
C_{Total}
mmol/m³
30
50
70
C_{ant}
250m
1000m
3000m
Surface Ocean
Overturning
Deep Ocean
Galen McKinley, LDEO-Columbia University & Natalie Renier, WHOI Creative © WHOI

Biogeochemical Cycling = Coupled carbon and nutrient cycles

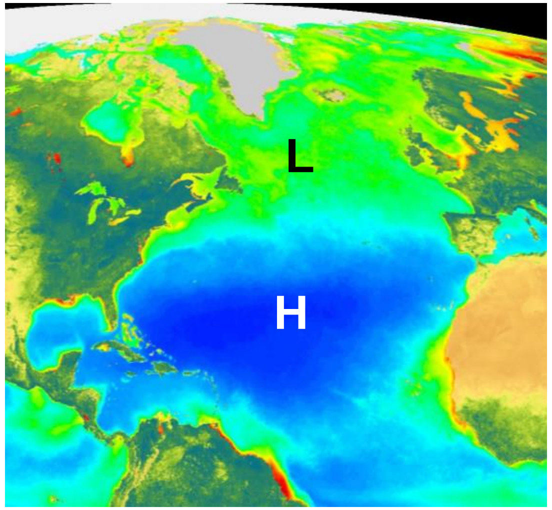
(b) carbon pathways in the ecosystem

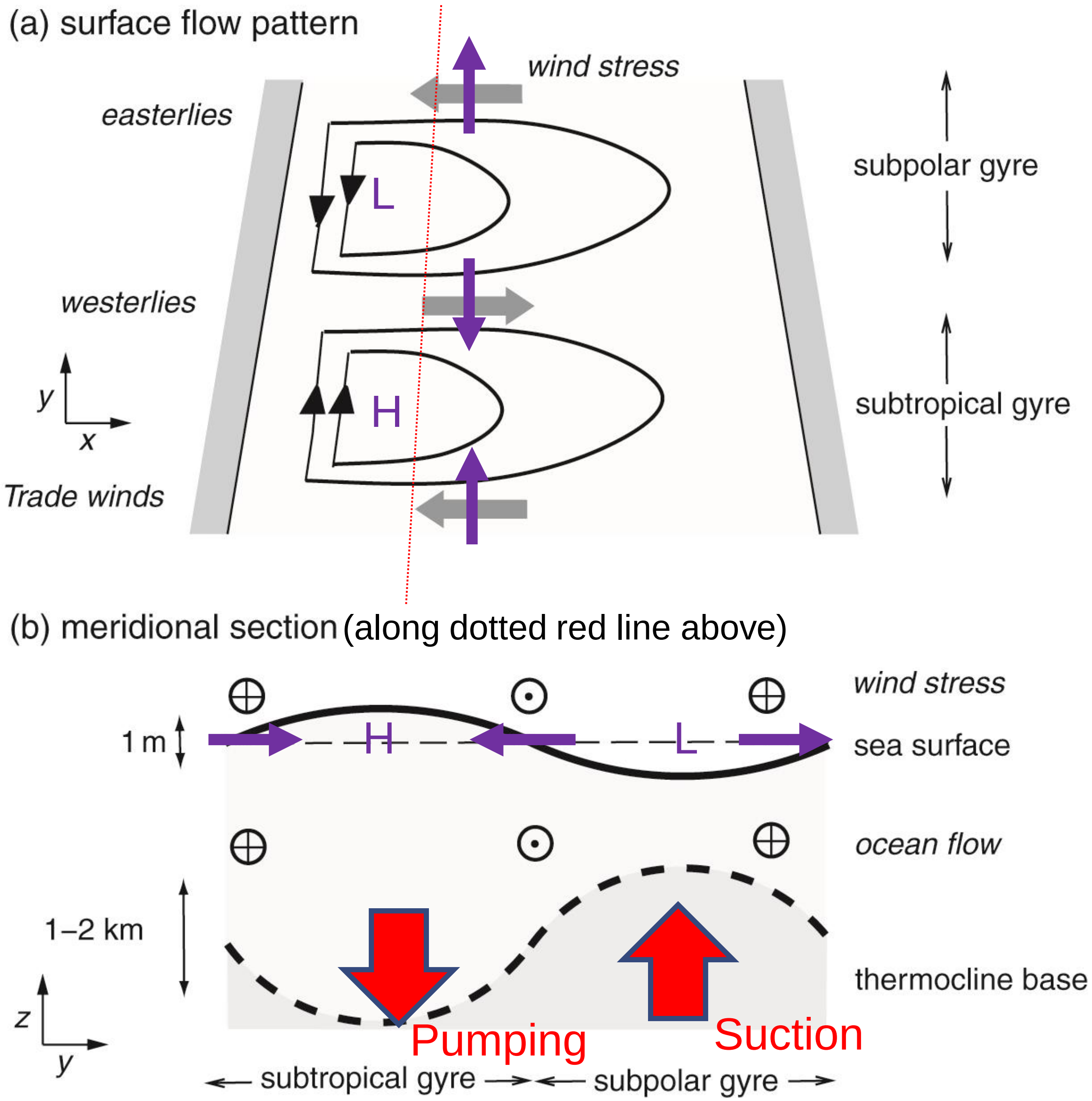


DIC = Dissolved Inorganic Carbon

Williams_Fig. 5.15

Circulation redistributes carbon and nutrients





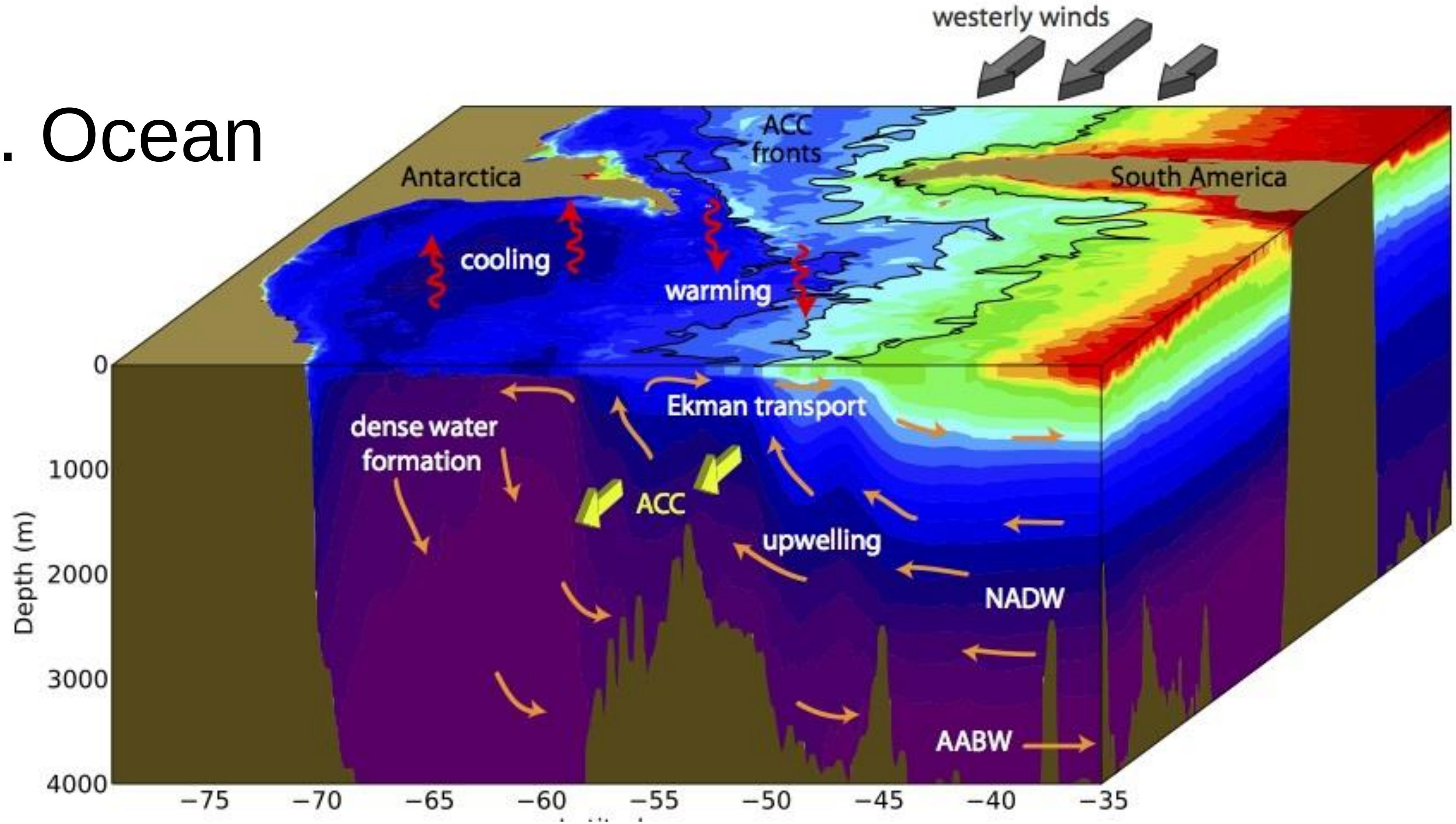
Winds + Coriolis force leads to mass convergence, divergence at surface; causing Highs and Lows in surface height

Horizontal flows lead to vertical motions; advecting carbon and nutrients

Williams and Follows Fig 8.1

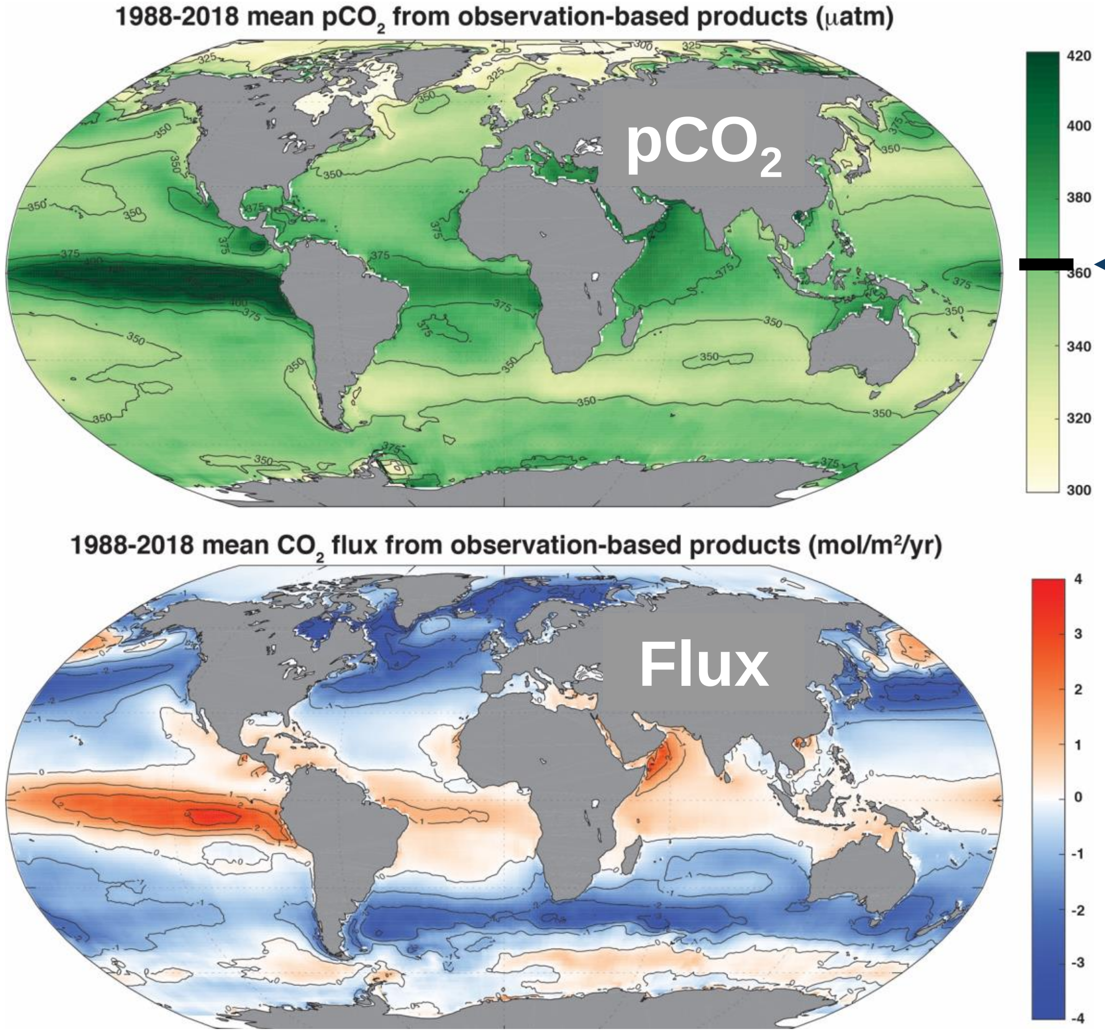
Circulation redistributes carbon and nutrients

S. Ocean



http://rses.anu.edu.au/research/projects/southern-ocean-circulation

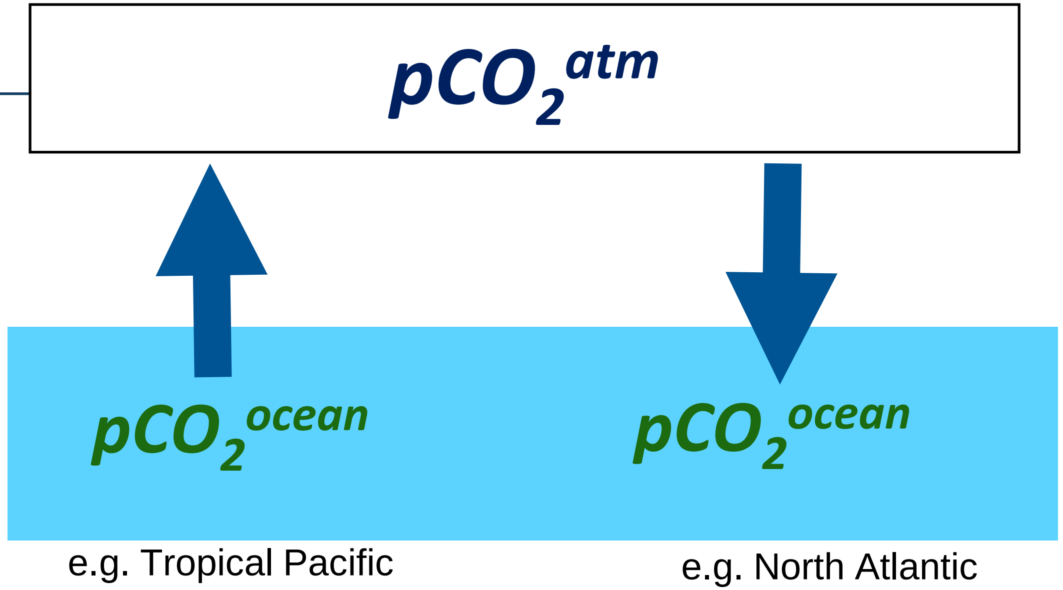
Surface ocean pCO_2 is primary control on air-sea CO_2 flux



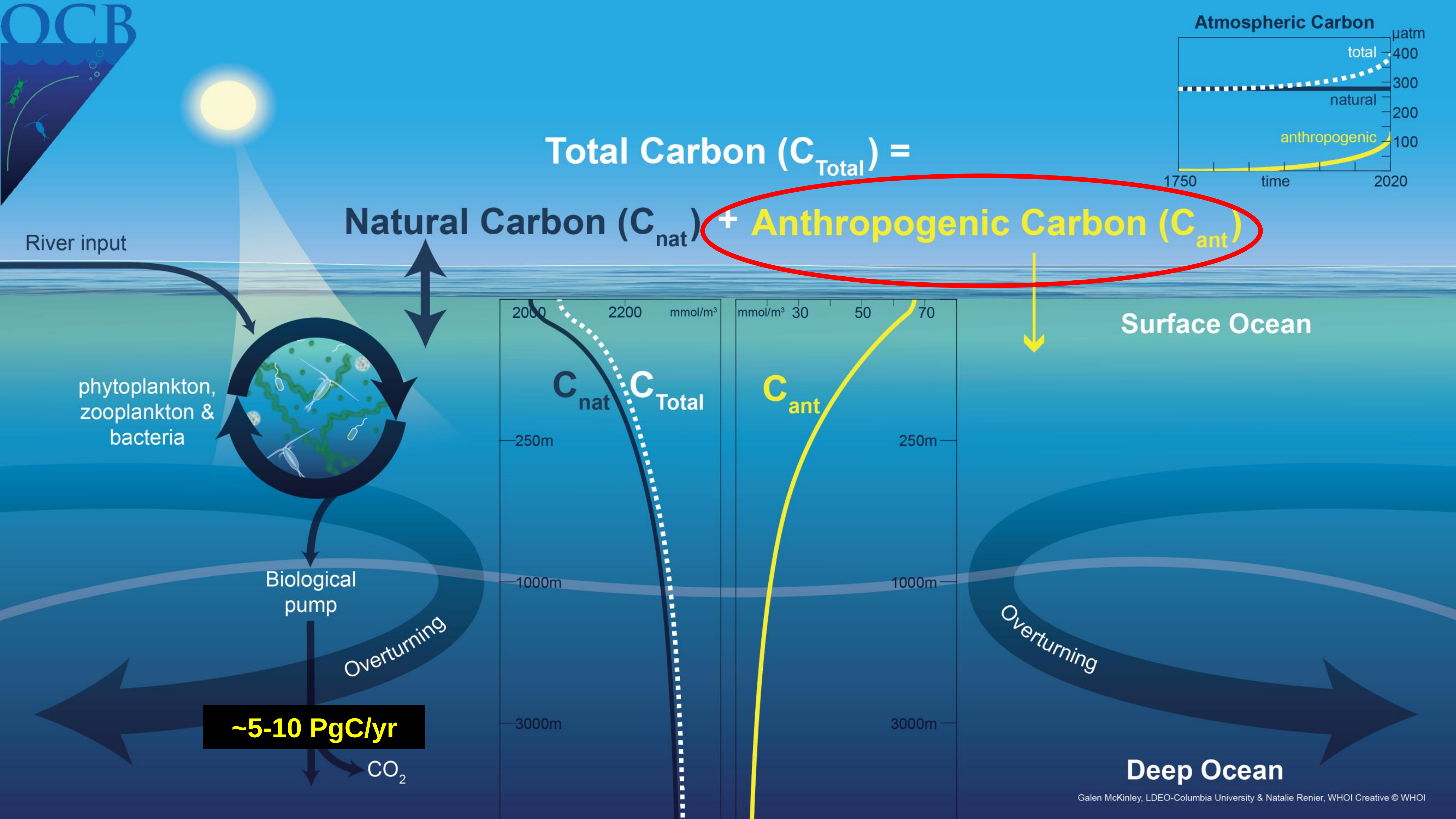
Crisp et al., 2022, Review of Geophysics

If $pCO_2^{ocean} > pCO_2^{atm}$ flux is out of ocean (positive)

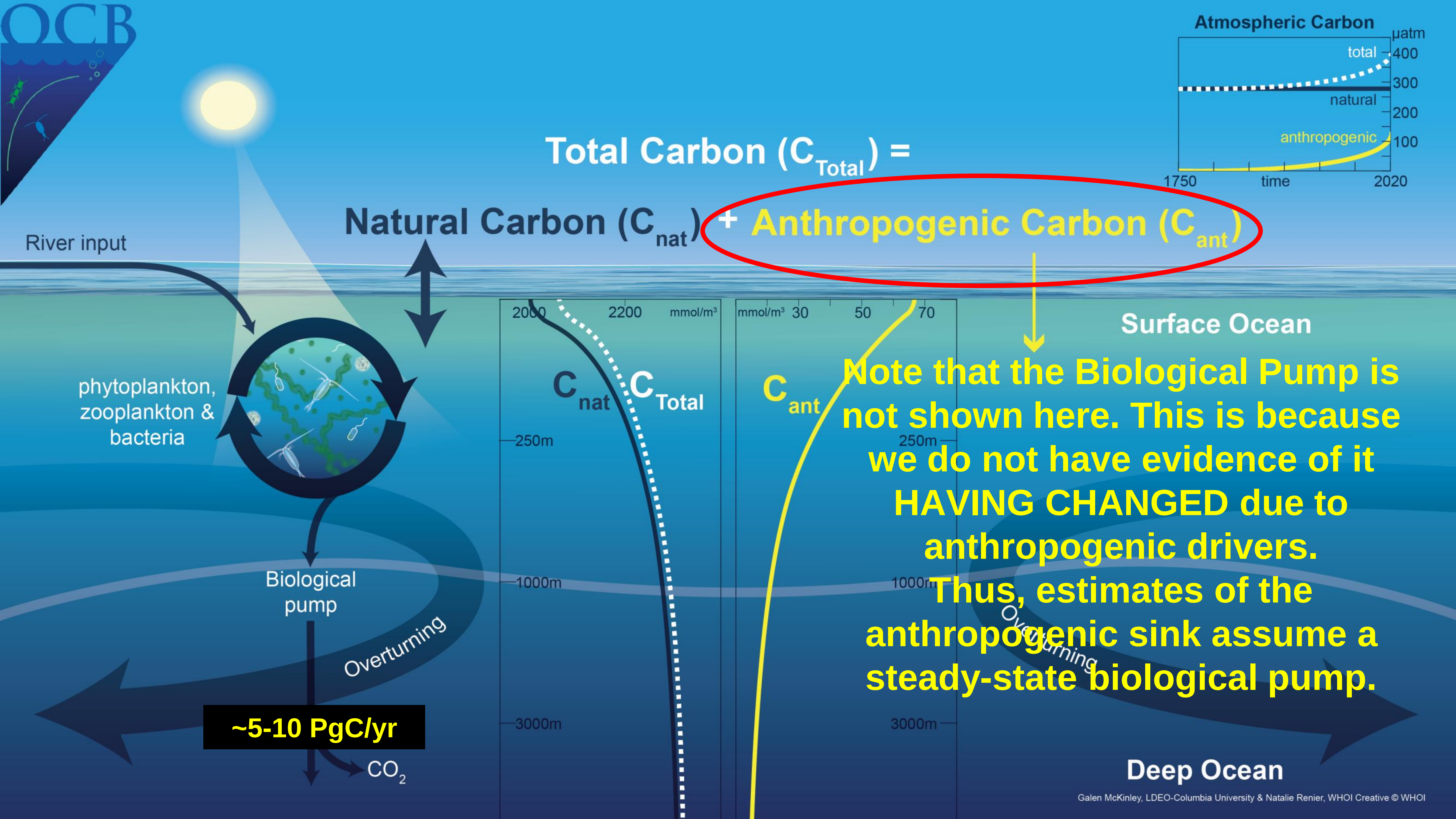
If $pCO_2^{ocean} < pCO_2^{atm}$ flux is into ocean (negative)



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



OCB
Atmospheric Carbon
µatm
total
natural
anthropogenic
400
300
200
100
1750
time
2020
Total Carbon (C_{Total}) =
Natural Carbon (C_{nat}) + Anthropogenic Carbon (C_{ant})
River input
2000
2200
mmol/m³
mmol/m³
30
50
70
Surface Ocean
C_{nat}
C_{Total}
C_{ant}
phytoplankton, zooplankton & bacteria
250m
250m
1000m
1000m
Biological pump
Overturning
Overturning
3000m
3000m
~5-10 PgC/yr
CO_2
Deep Ocean
Galen McKinley, LDEO-Columbia University & Natalie Renier, WHOI Creative © WHOI


OCB
Total Carbon (C_{Total}) =
Natural Carbon (C_{nat}) + Anthropogenic Carbon (C_{ant})
Atmospheric Carbon
μatm
total
natural
anthropogenic
400
300
200
100
1750
time
2020
River input
phytoplankton, zooplankton & bacteria
Biological pump
Overturning
~5-10 PgC/yr
CO_2
2000
2200
mmol/m³
C_{nat}
C_{Total}
mmol/m³
30
50
70
C_{ant}
250m
1000m
3000m
Surface Ocean
Note that the Biological Pump is not shown here. This is because we do not have evidence of it HAVING CHANGED due to anthropogenic drivers. Thus, estimates of the anthropogenic sink assume a steady-state biological pump.
Overturning
Deep Ocean
Galen McKinley, LDEO-Columbia University & Natalie Renier, WHOI Creative © WHOI

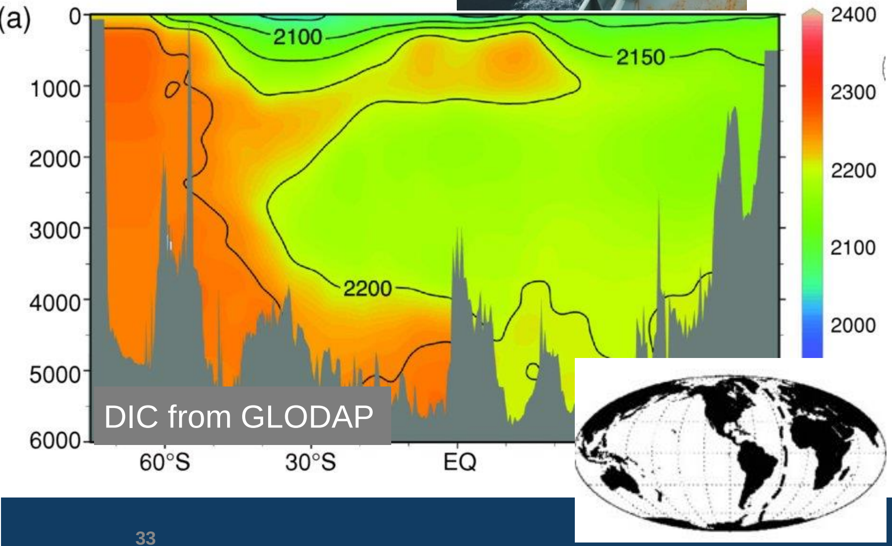
Three independent approaches constrain the ocean carbon sink

<u>1. Interior observations / products</u>

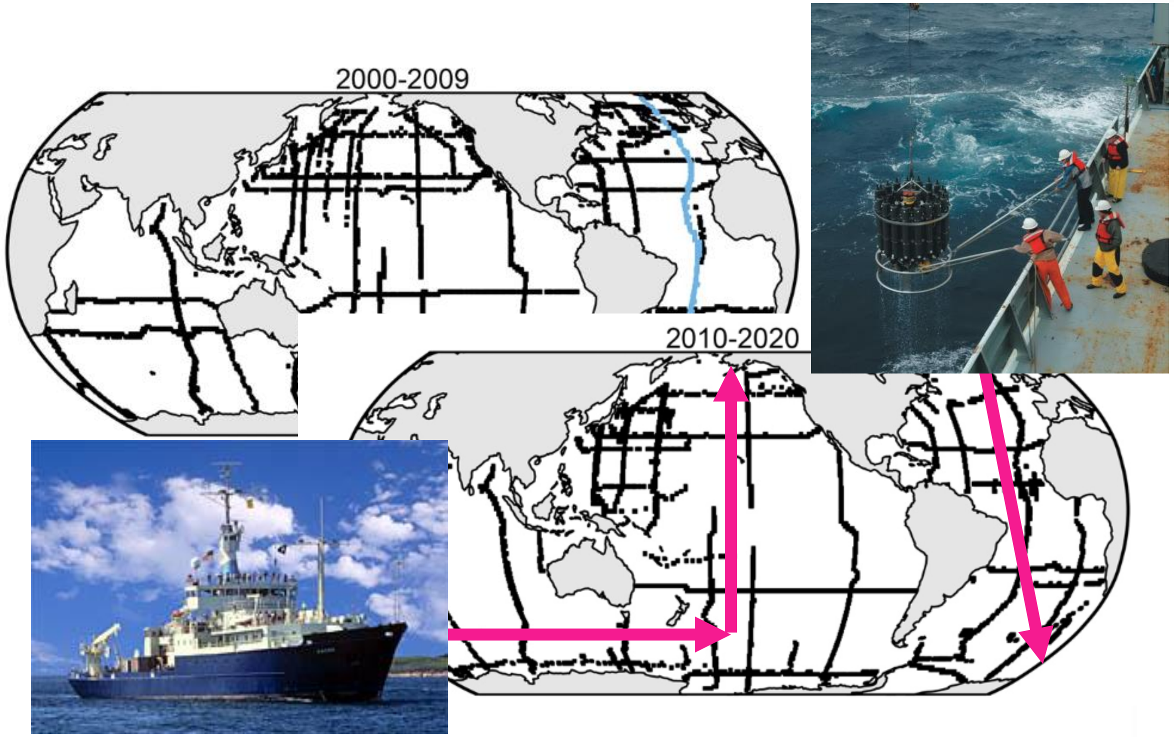
Interior carbon storage

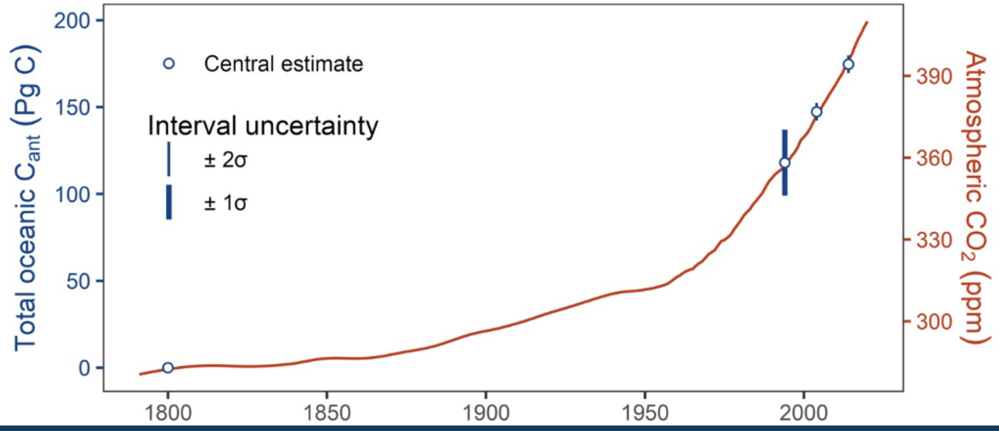
Decadal closure of global budget

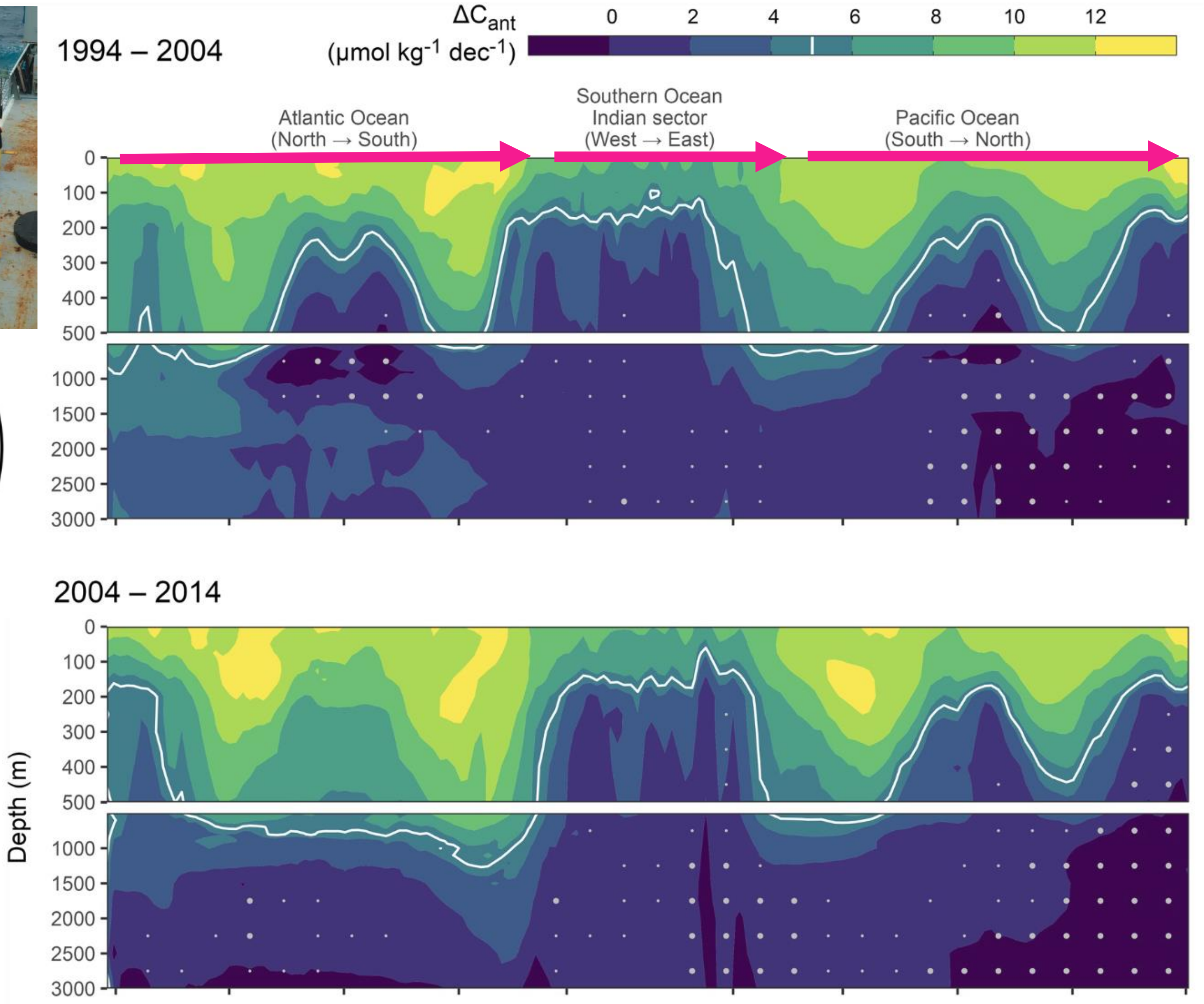
Model validation



Decadal Interior Data: Shows ocean *following atmospheric* pCO_2

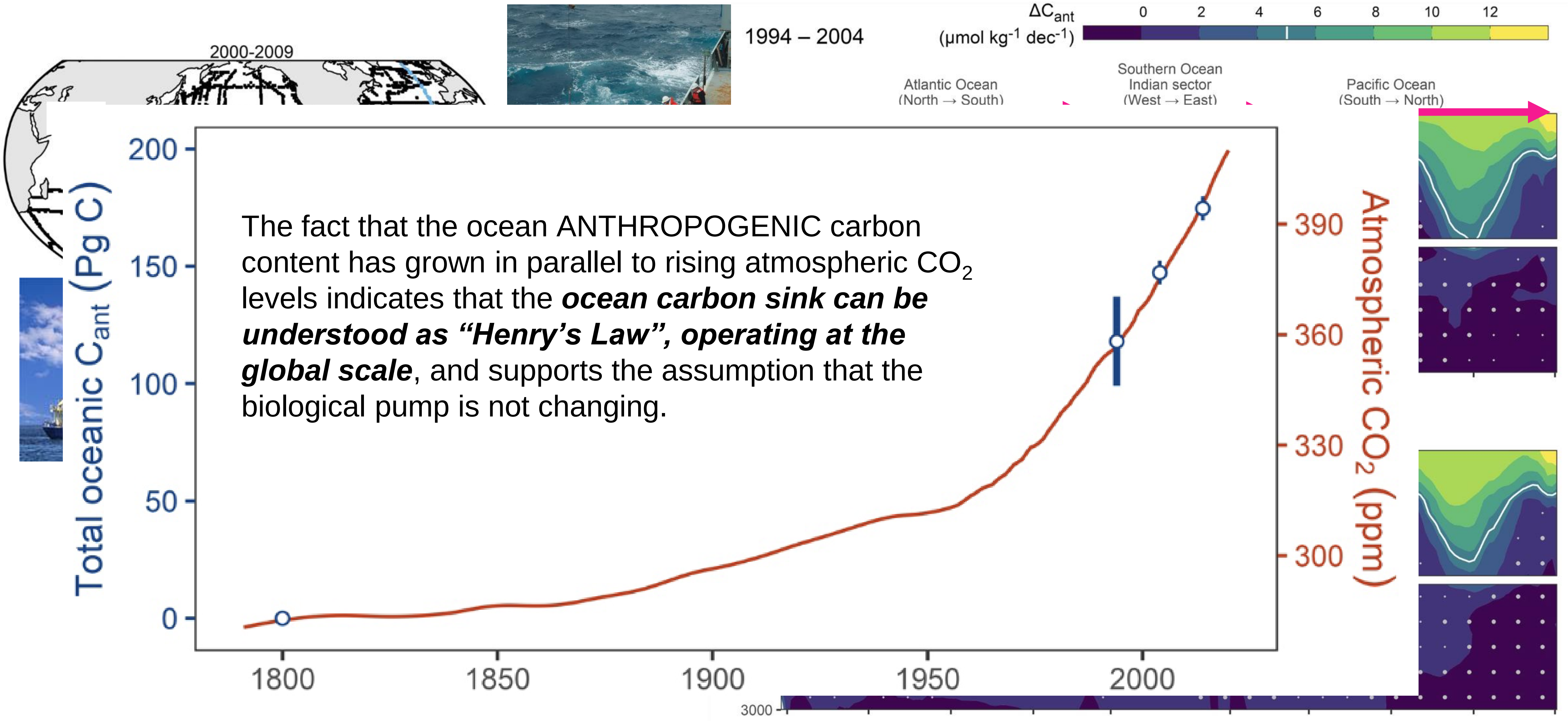






Müller, Gruber et al. 2023, AGU Advances

*Decadal Interior Data: Shows ocean **following atmospheric pCO_2***



The fact that the ocean ANTHROPOGENIC carbon content has grown in parallel to rising atmospheric CO_2 levels indicates that the ***ocean carbon sink can be understood as "Henry's Law", operating at the global scale***, and supports the assumption that the biological pump is not changing.

Müller, Gruber et al. 2023, AGU Advances

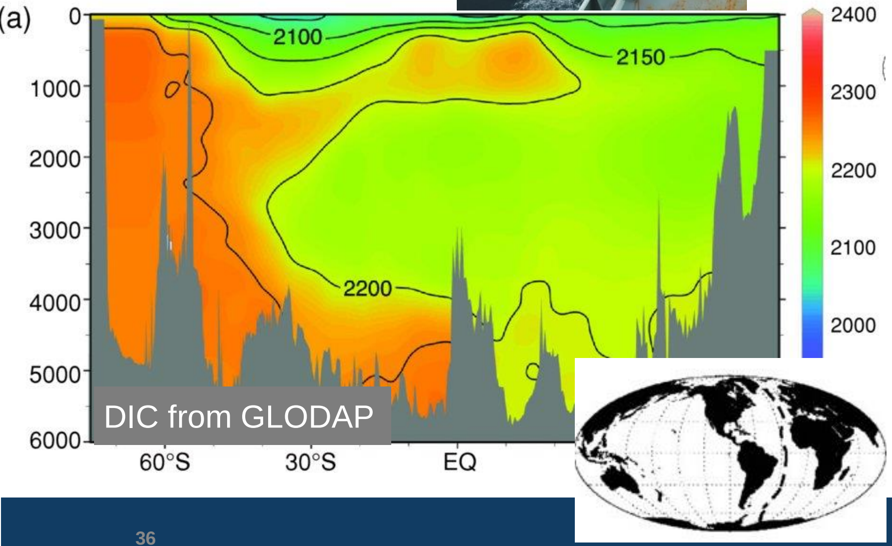
Three independent approaches constrain the ocean carbon sink

1. Interior observations / products

Interior carbon storage

Decadal closure of global budget

Model validation

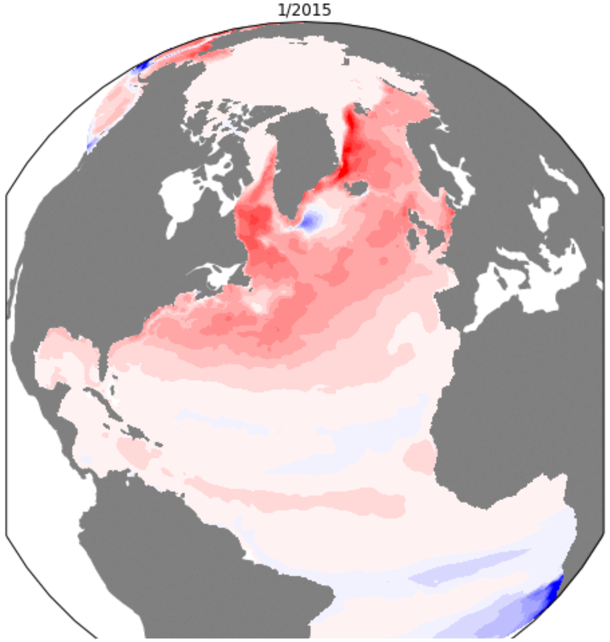


2. Modeling

Air-sea fluxes

Mechanisms

Projections



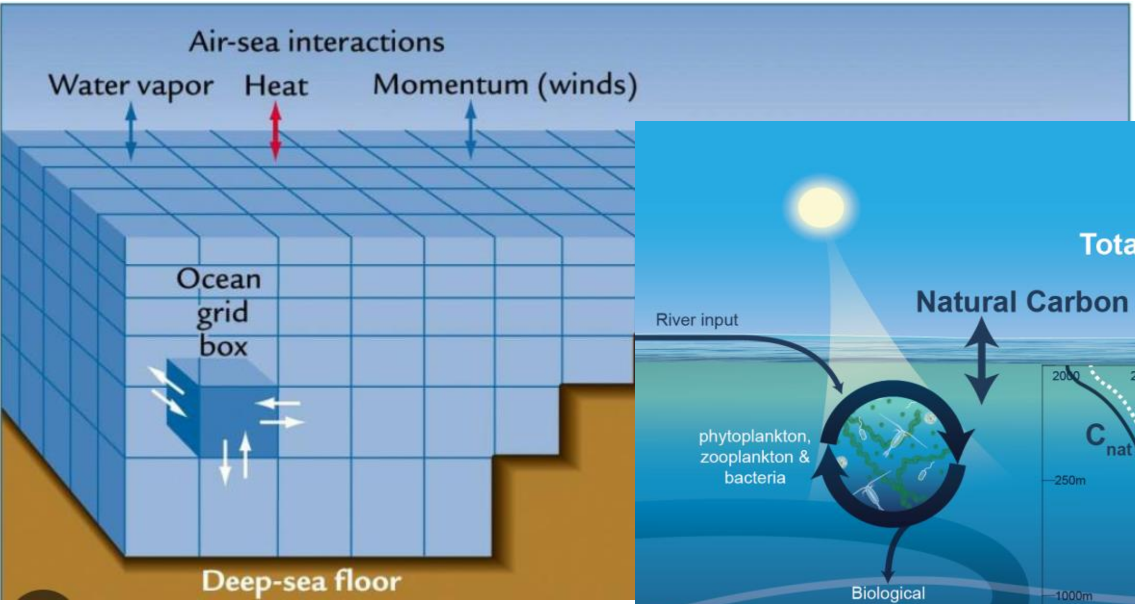
ASTE-BGC

air-sea CO_2 flux

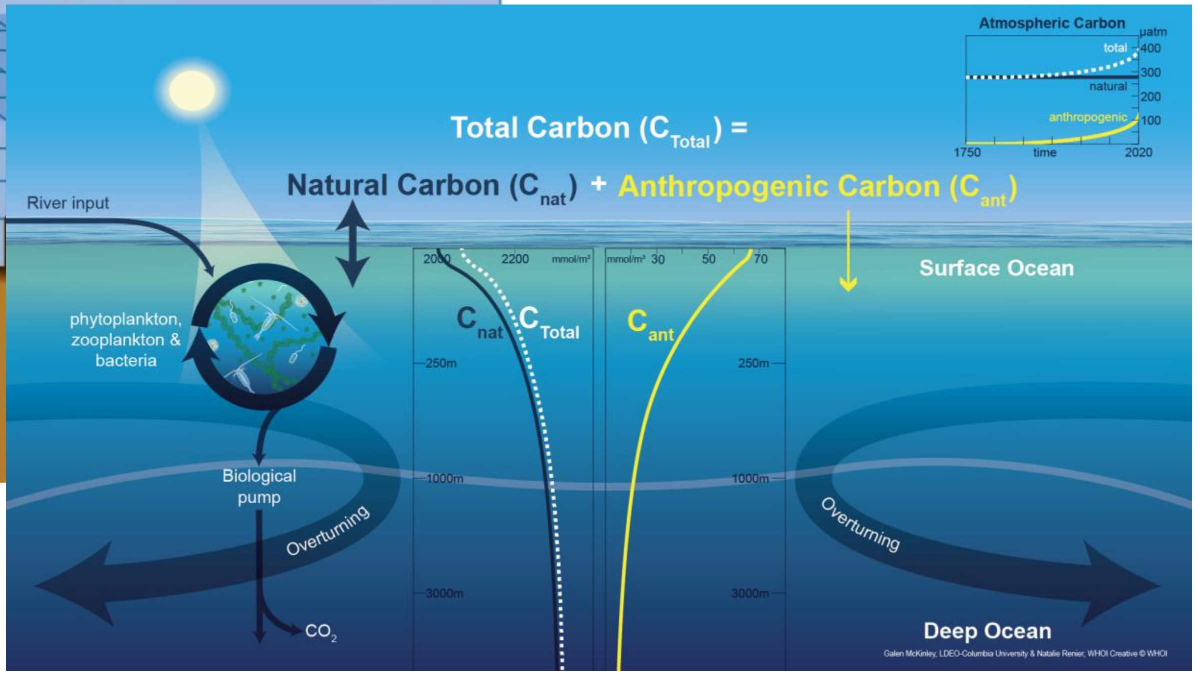
Moseley et al, in review

Ocean biogeochemical models

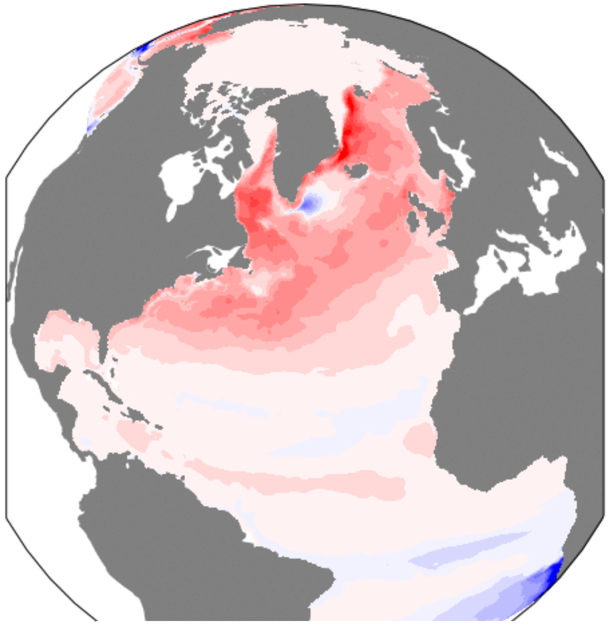
Circulation



+ Carbon Processes



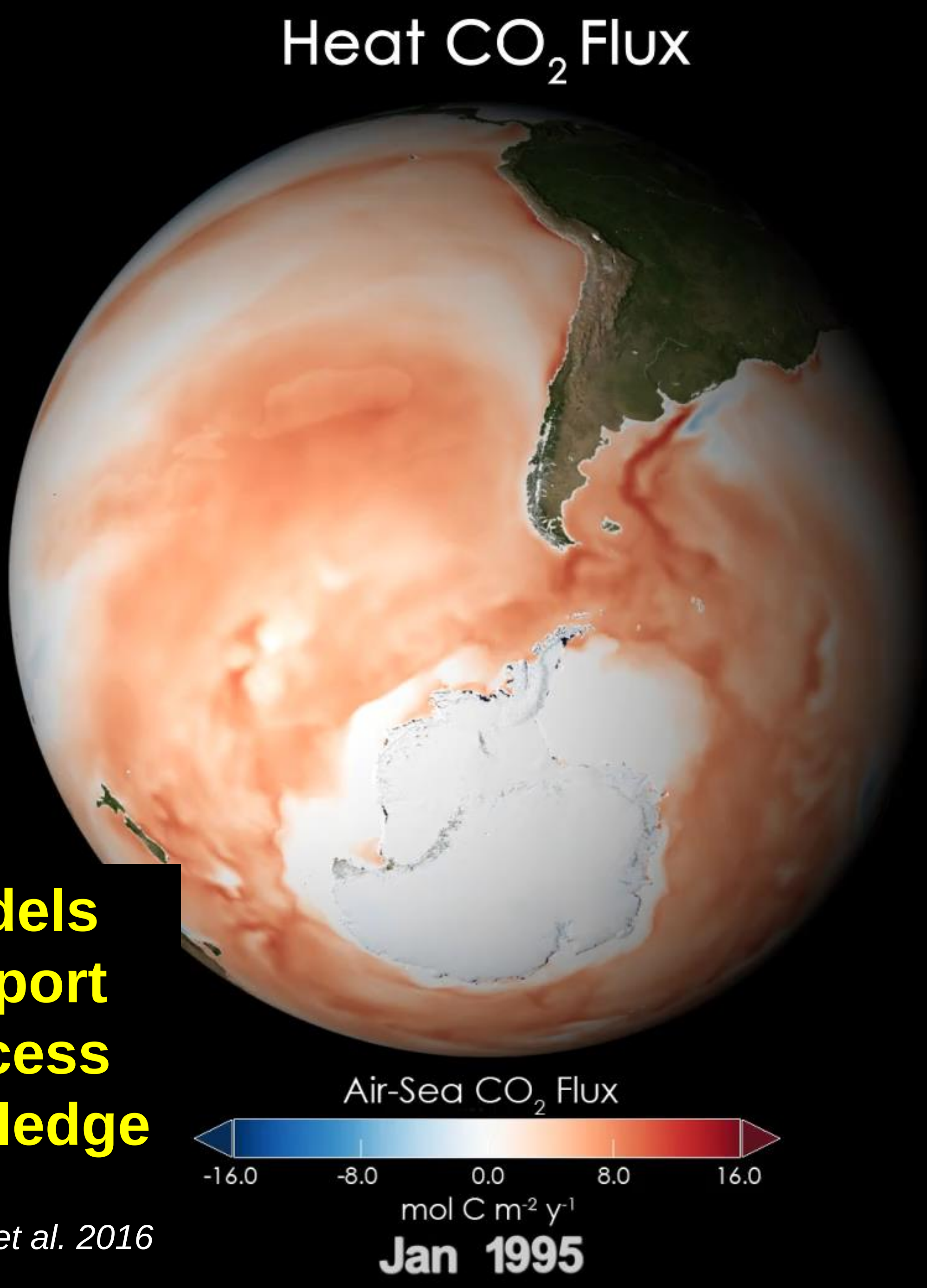
= Air-sea CO_2 fluxes



ASTE-BGC
air-sea CO_2 flux

Moseley et al, in review

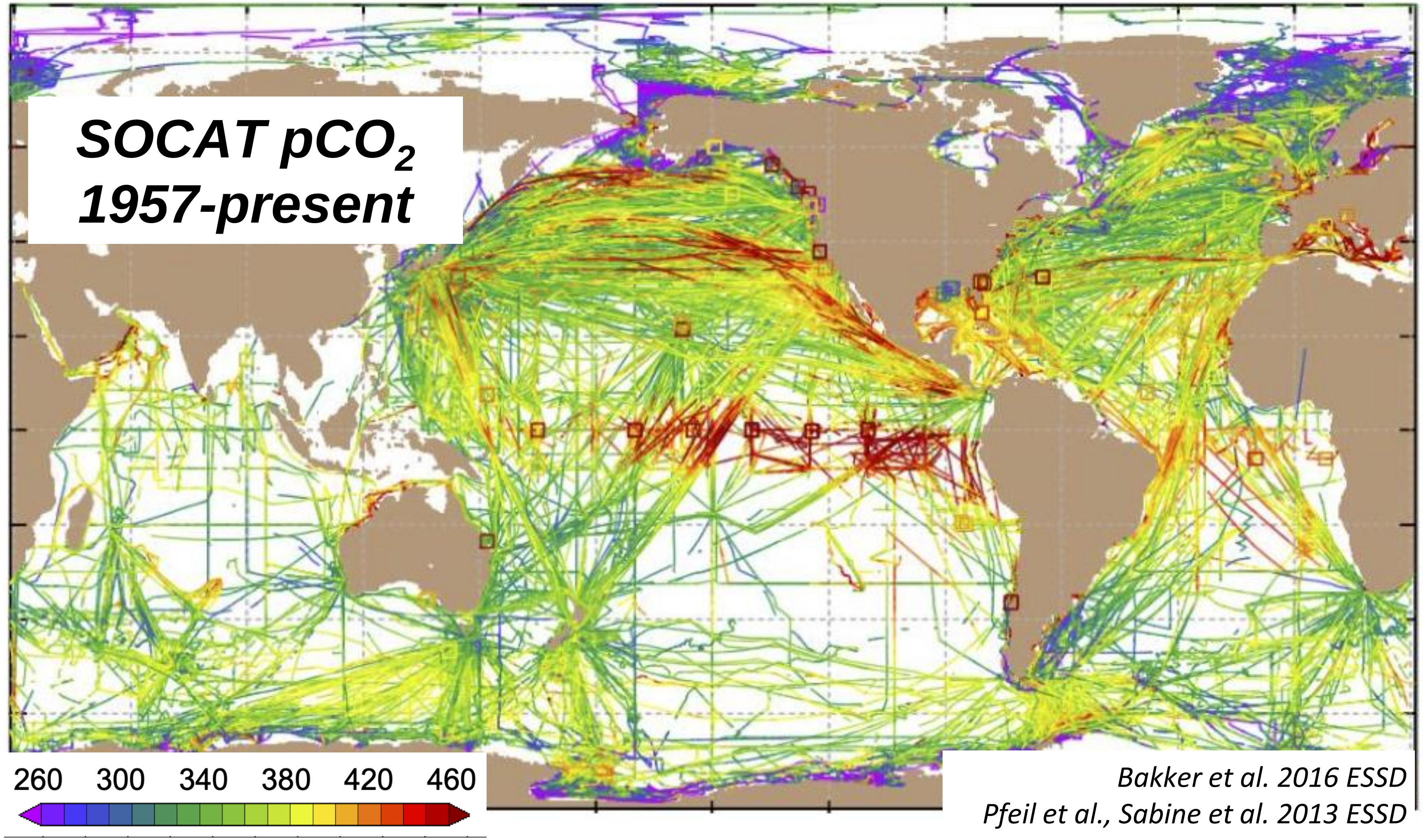
On a 3D grid, calculate circulation and carbon fluxes by integrating over time the equations of motion, conservation, biology and carbon chemistry



Models support process knowledge

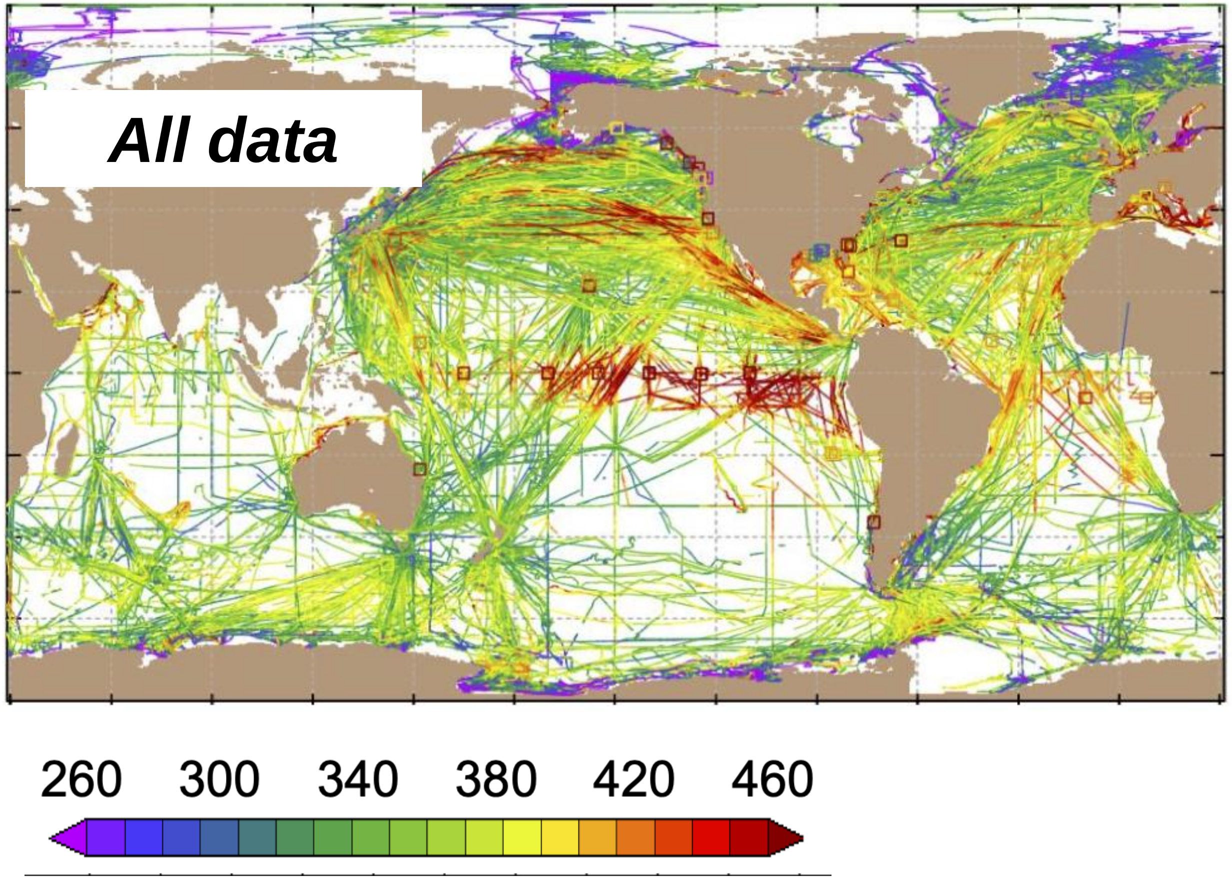
Lauderdale et al. 2016

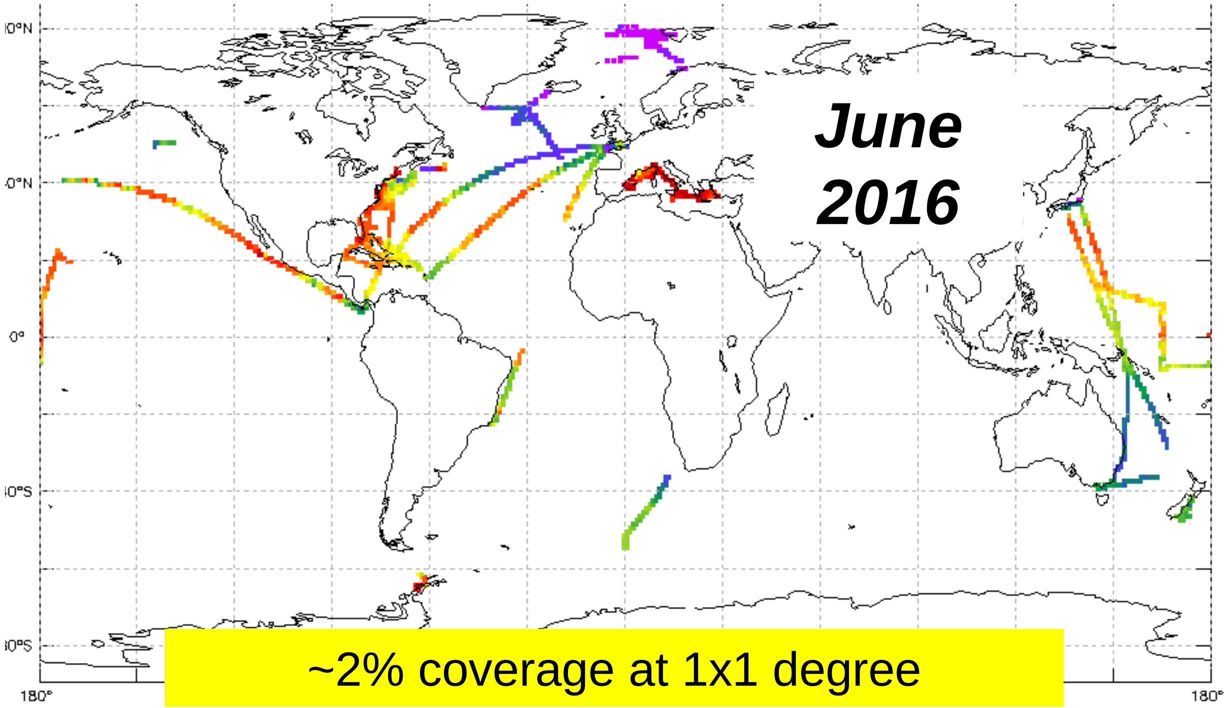
3. Surface ocean pCO_2 data used in reconstructions



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

3. Monthly pCO_2 data are very sparse …



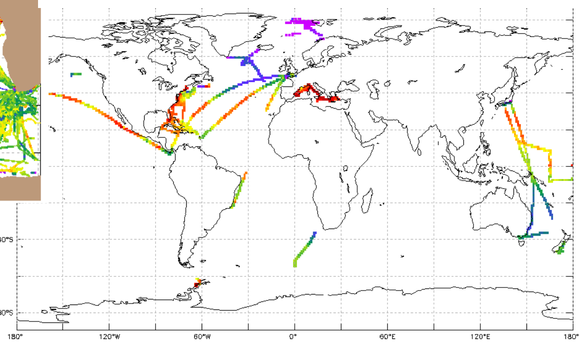


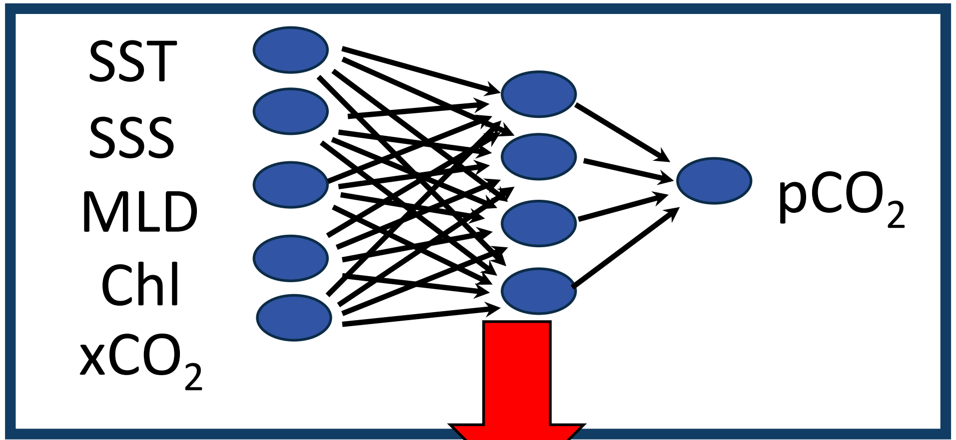
3. Data-based "Products" built with v. sparse pCO_2 data and ML

SOCAT pCO_2,1957-2023

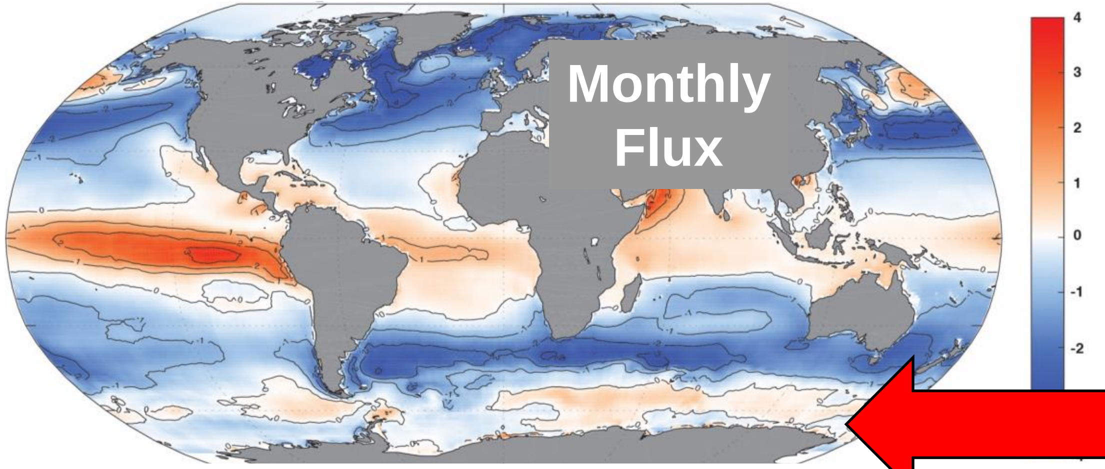
1. Train Machine Learning on sparse data

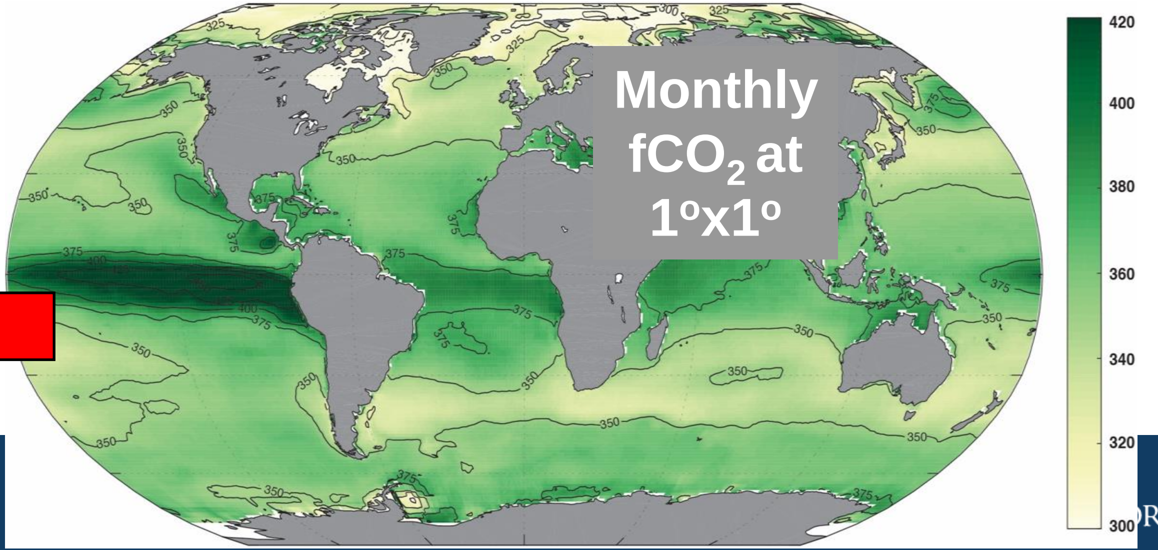
1 month ~ 2% coverage





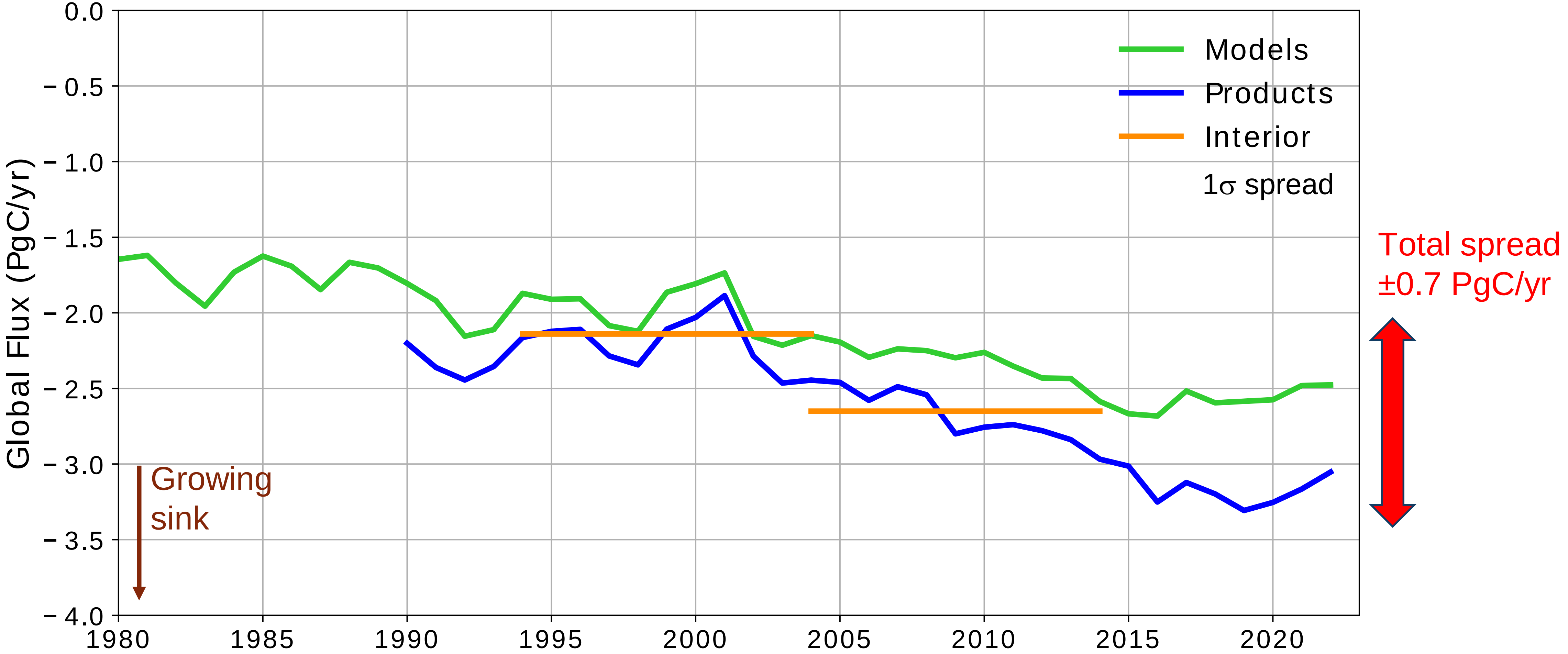
2. Predict pCO_2 = f(SST, SSS, MLD, Chl, xCO_2)





3. Calculate Flux from pCO_2

Globally integrated estimates agree, with ~30% uncertainty



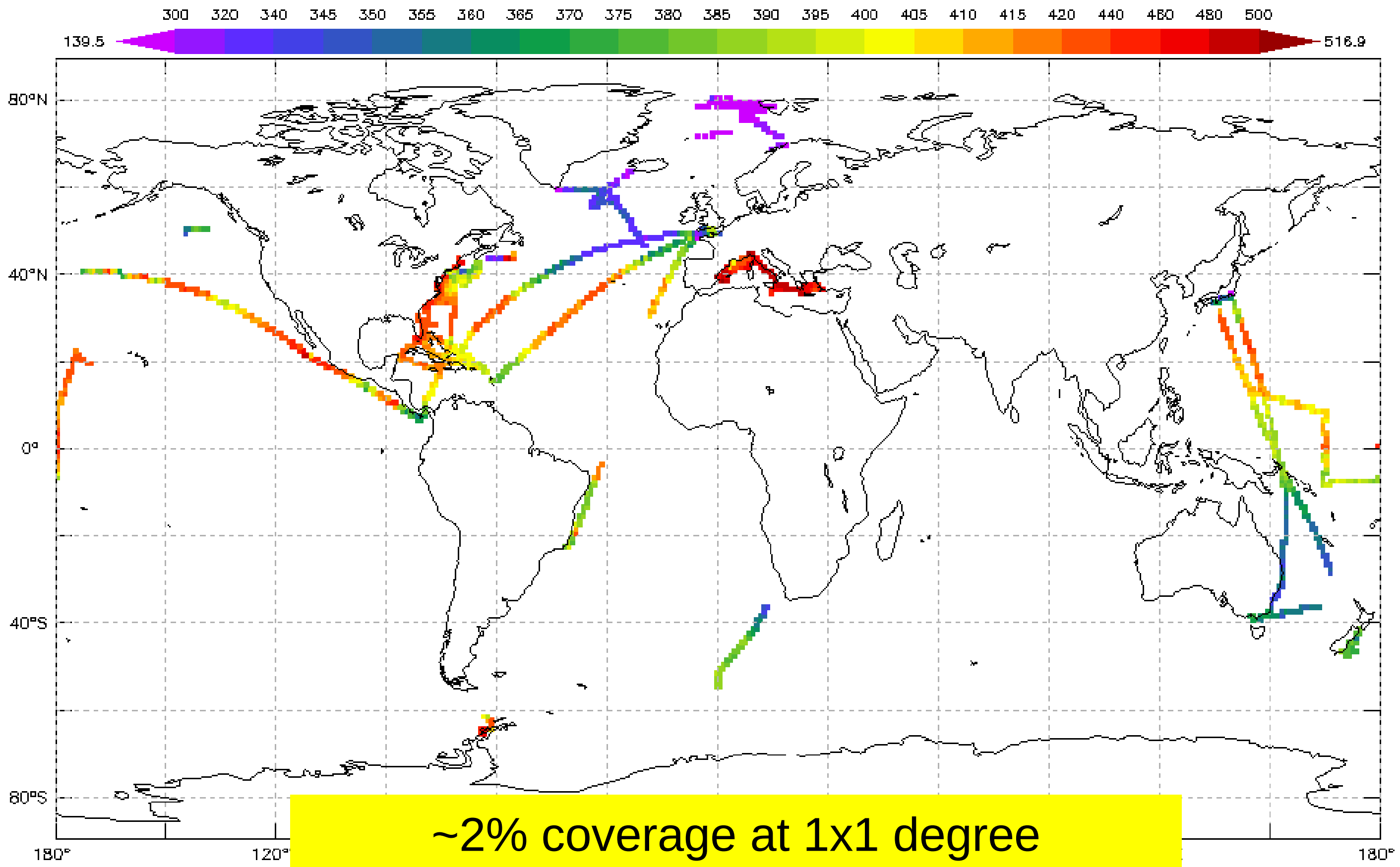
Models, products from Friedlingstein et al. 2023
Interior from Müller et al. 2023

Outline

- Key processes of the ocean carbon cycle

- Improving quantification of the ocean carbon sink
  - ***Models*: evaluate ML-based data product skill, given sampling**
  - *pCO_2 data products*: apply to identify model mean-state biases
  - *Models and data*: combine using ML
  - *marine Carbon Dioxide Removal (mCDR)* – quantifying “additionality”?

pCO₂ data are very sparse: Are reconstructions robust?

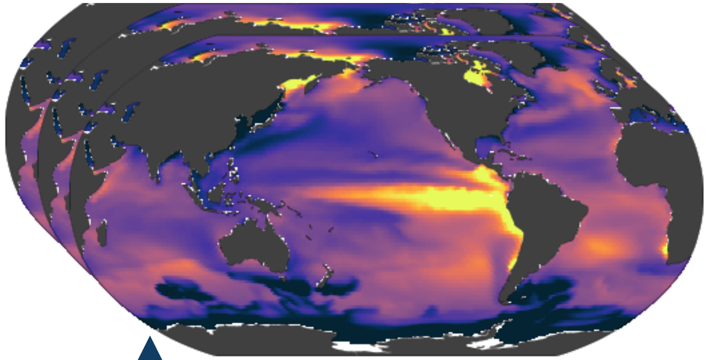


~2% coverage at 1x1 degree

ESM Large Ensembles as testbed for reconstruction skill

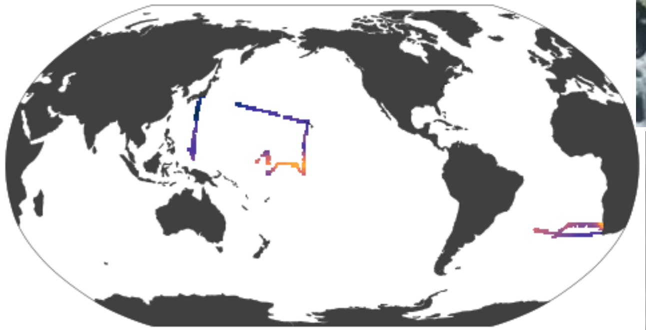
Earth System Models provide pCO_2 and drivers (SST, Chl, etc)

....

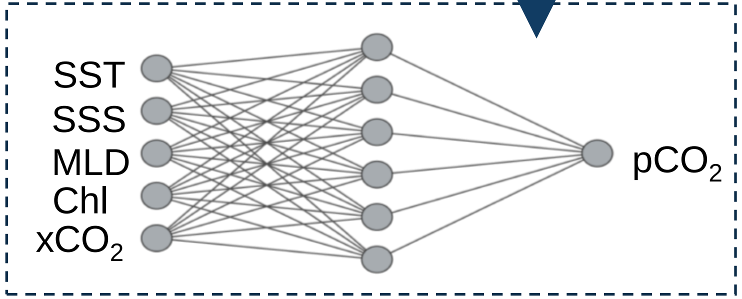


50 to 100 ensembles from 4 or 9 ESMs

1. Sample model member as SOCAT monthly pCO_2 product



2. Train, evaluate, test

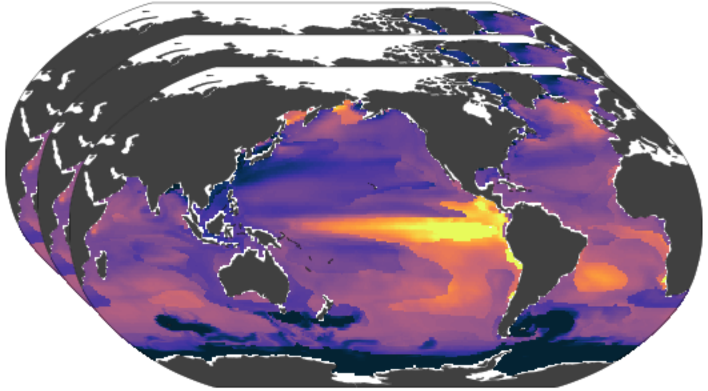


With various ML reconstructions
- *MPI-SOMFFN* (Gloege et al. 2021)
- NN, RF, XGB (Stamell et al. 2020)
- pCO_2-Residual (Bennington et al. 2022, Heimdal et al, 2024a,b; Heimdal et al. in press)

3. Estimate monthly varying pCO_2 on global scale using trained model, calculate flux

X ensemble members

....

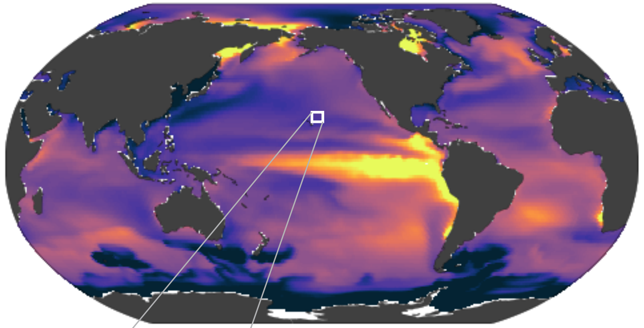


reconstruction

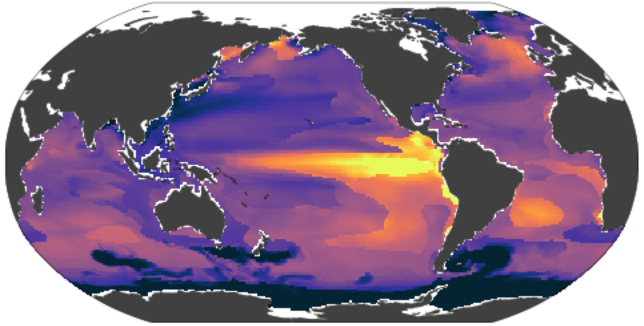
4. Statistically compare reconstructed CO_2 flux to model truth. Each spatial point is temporally decomposed

Skill evaluation: temporal decomposition

Ensemble member

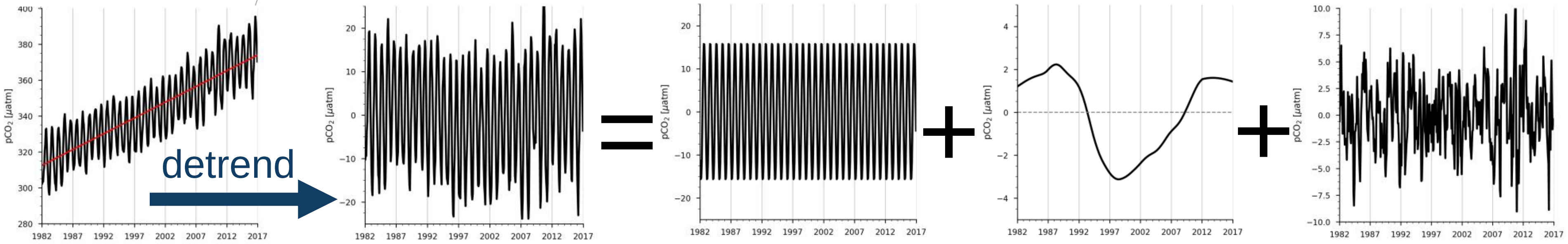


Reconstruction



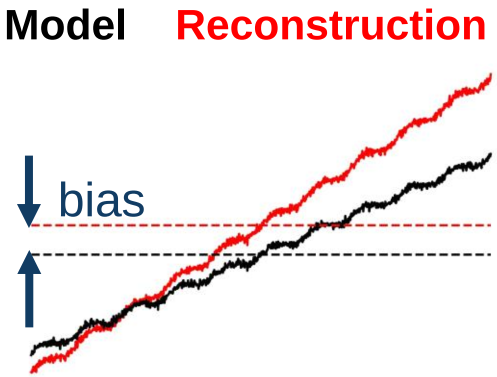
Each location in each member and reconstruction decomposed into seasonal, decadal, and sub-decadal variability

Full Signal | Detrended | Seasonal | Dec. var. | Sub-dec. var.

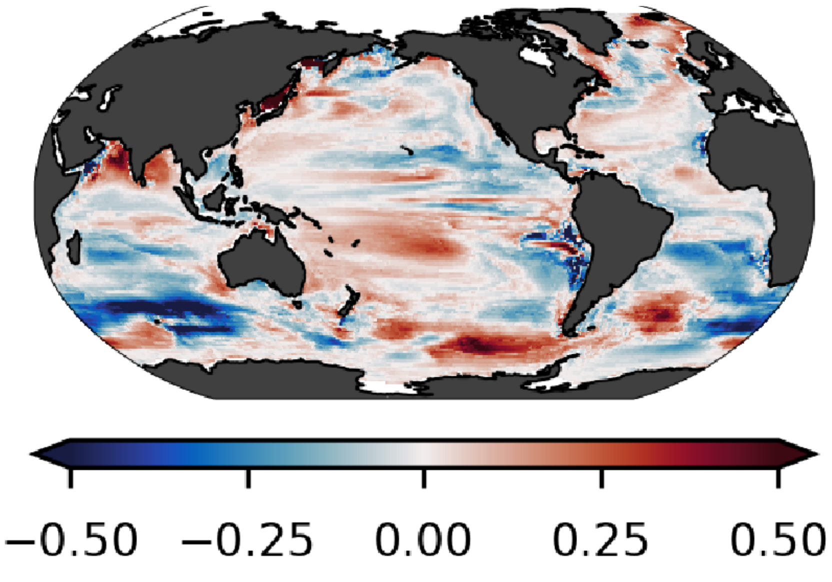


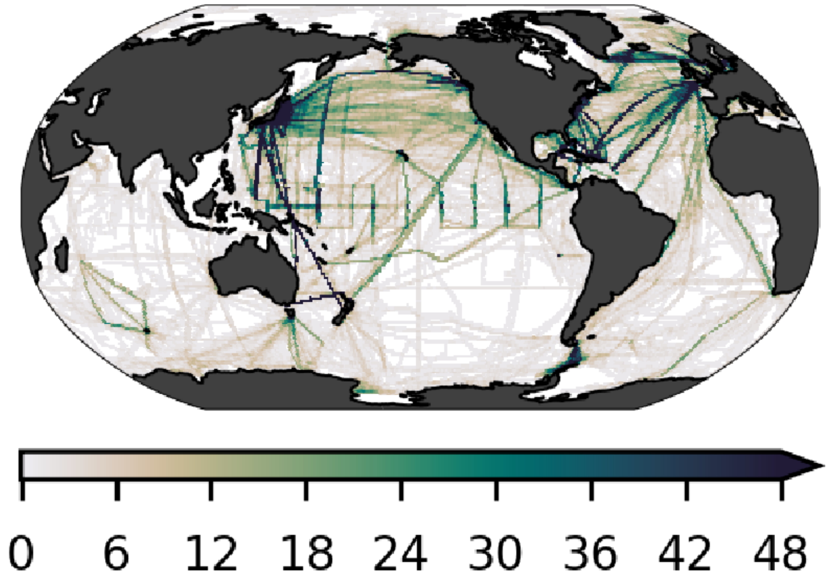
Gloege et al. 2021, GBC

Bias is small globally; larger where data sparse



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



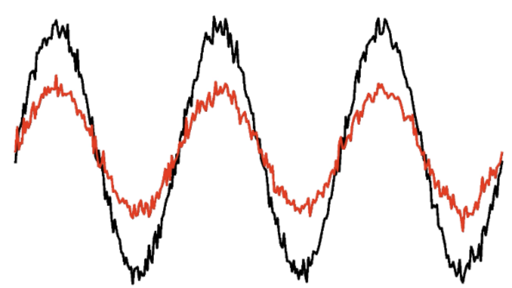


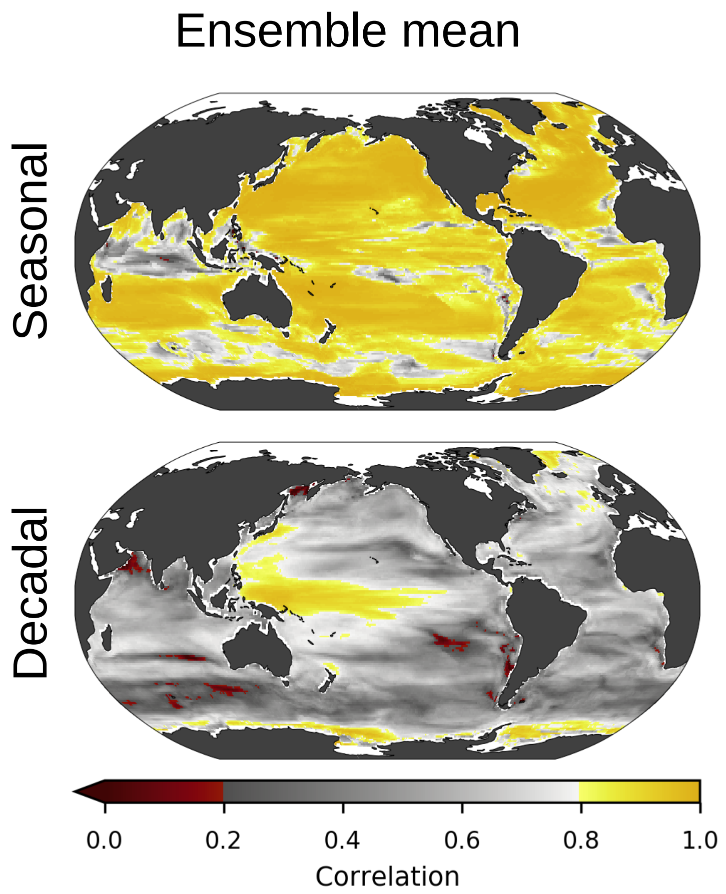
100-member mean
1985-2016
MPI-SOMFFN

Gloege et al. 2021, GBC

Gloege et al. 2021, GBC

Correlation
Is the reconstruction in phase with the original data?

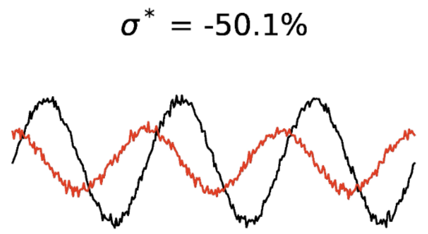




- Captures seasonal cycle well everywhere
- Decadal phase is more challenging to reconstruct

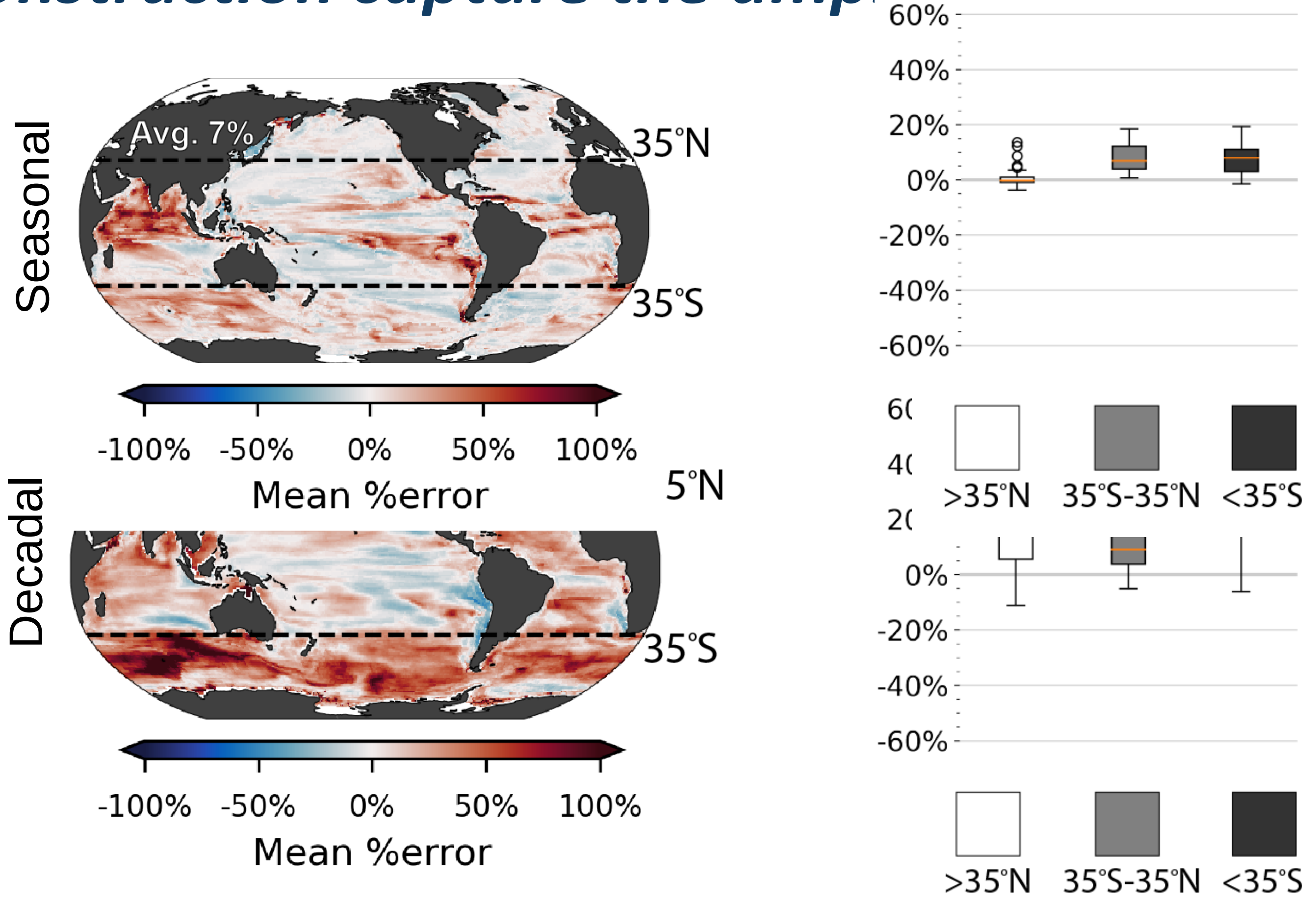
100-member mean
1985-2016
MPI-SOMFFN

Gloege et al. 2021, GBC



Standard deviation of % error

Does the reconstruction capture the amplitude of variability?

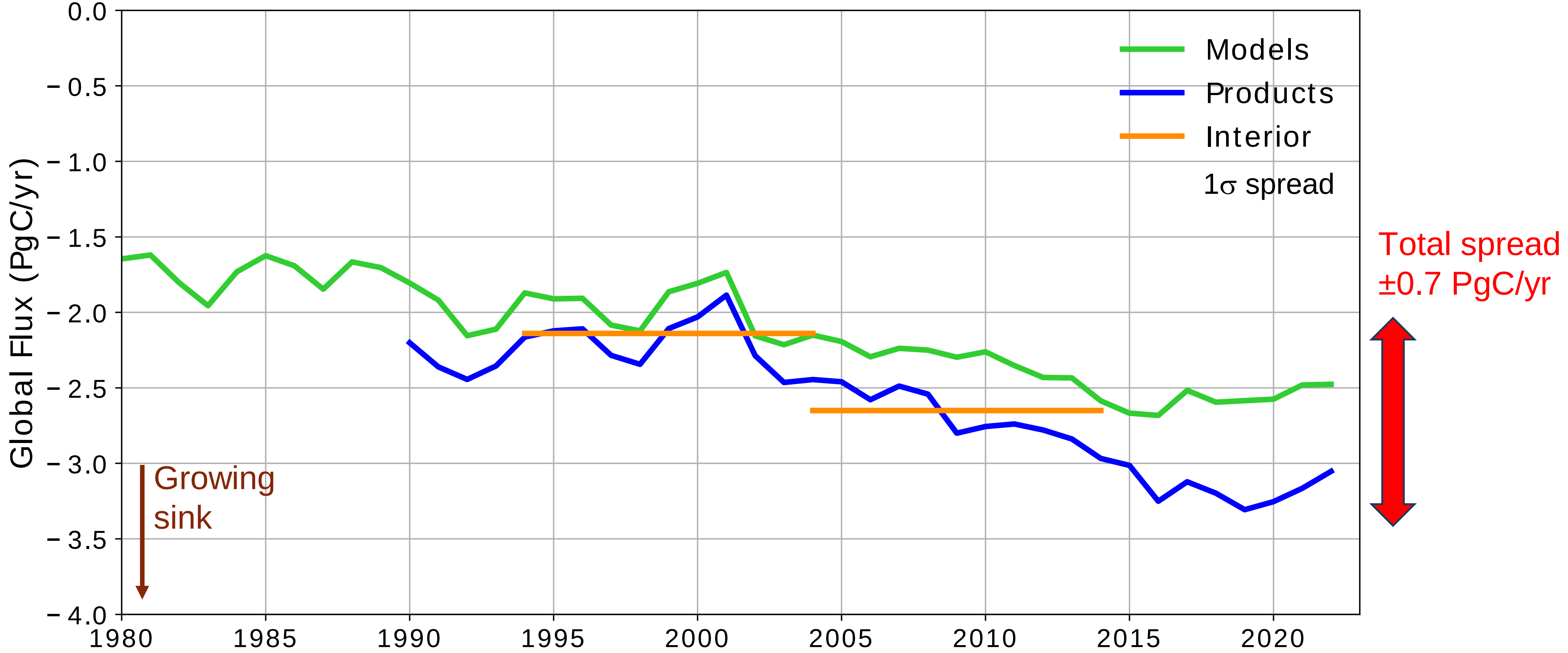


100-member mean
1985-2016
MPI-SOMFFN

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

Gloege et al. GBC 2021

Reduced decadal amplitude in products should increase agreement between products and models



Models, products from Friedlingstein et al. 2023
Interior from Müller et al. 2023

Take home messages

- Air-sea CO_2 flux mean and seasonality can be reconstructed from sparse pCO_2 data, but longer-timescale variability remains uncertain

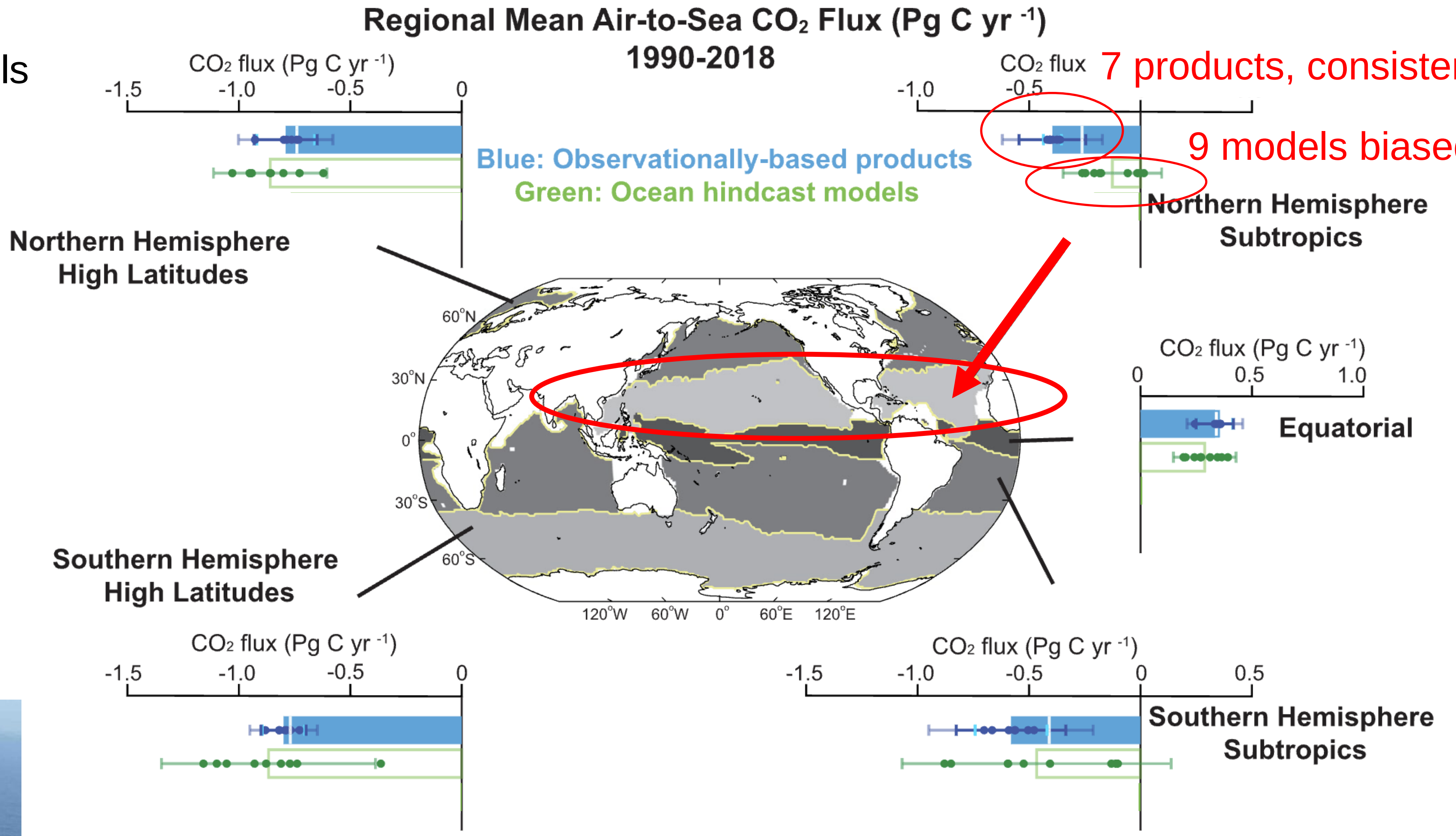
Outline

- Key processes of the ocean carbon cycle

- Improving quantification of the ocean carbon sink
  - *Models*: evaluate ML-based data product skill, given sampling
  - ***pCO_2 data products*: apply to identify model mean-state biases**
  - *Models and data*: combine using ML
  - *marine Carbon Dioxide Removal (mCDR)* – quantifying “additionality”?

Ocean models have significant mean-state errors

Products and hindcast models for real fluxes 1990-2018



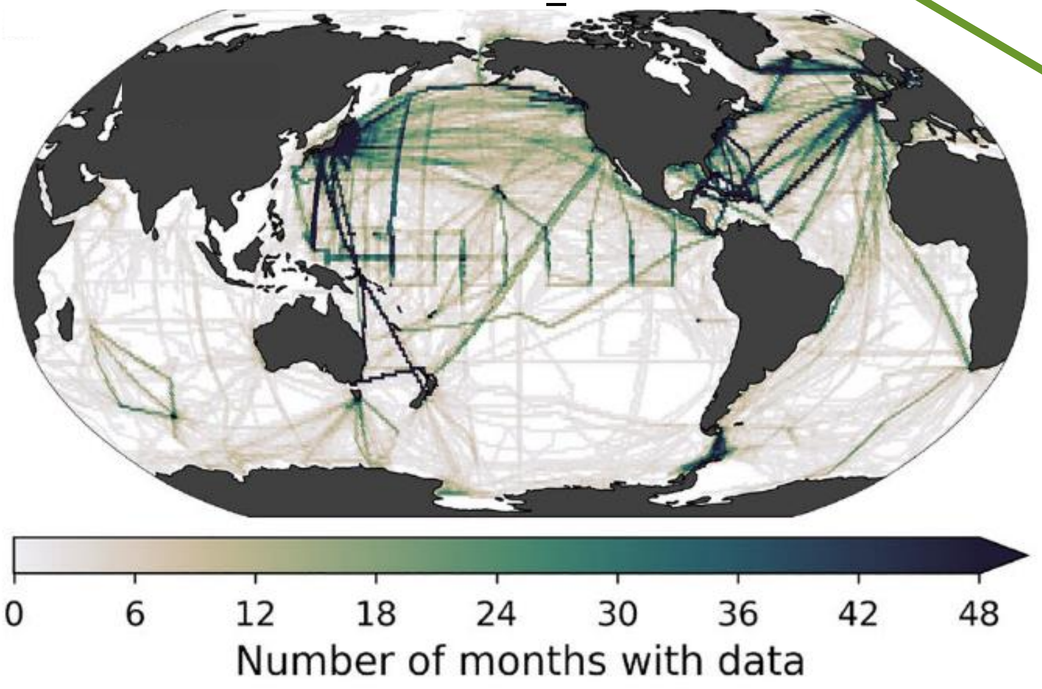
7 products, consistent long-term mean

9 models biased low, large spread

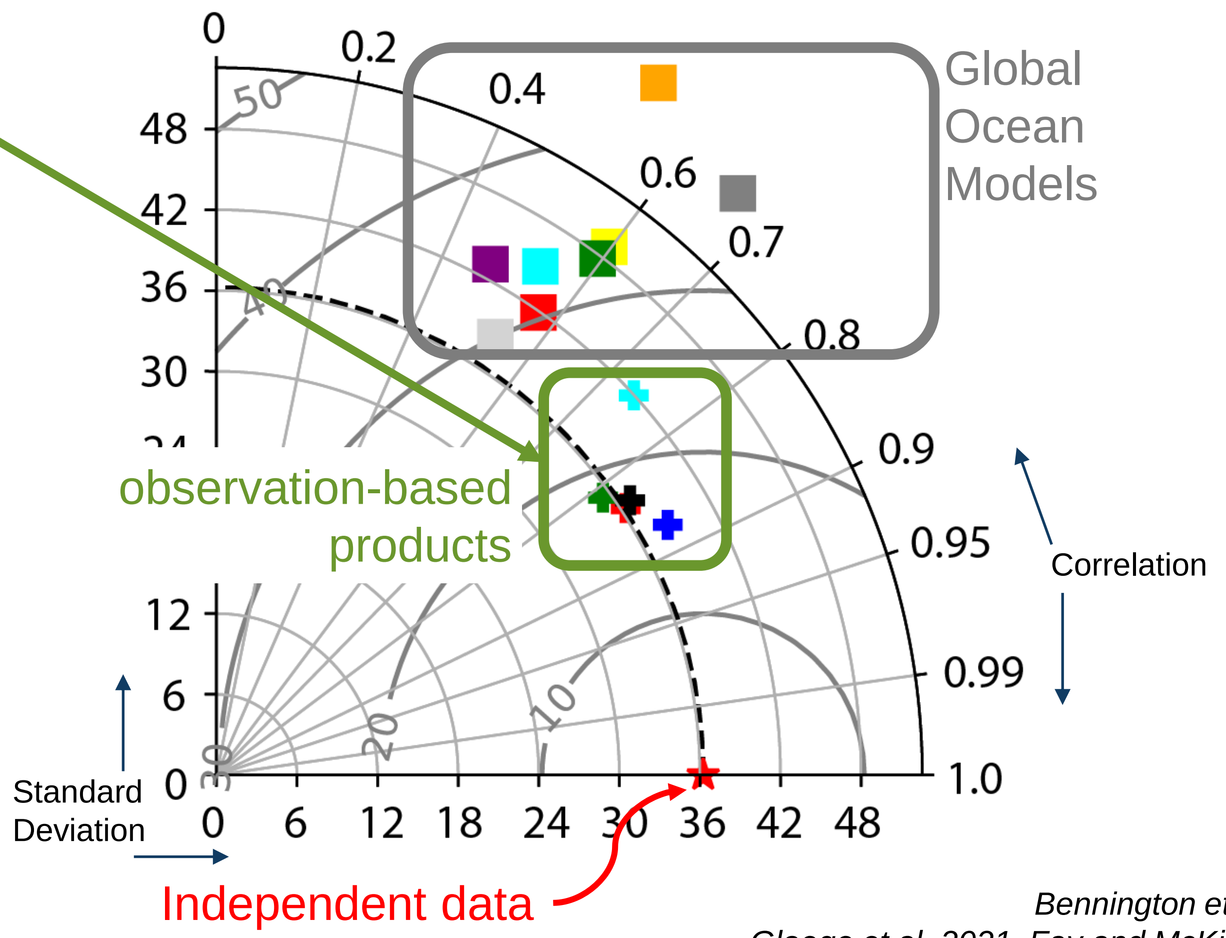
Fay and McKinley 2021

Ocean models have significant mean-state errors

Surface ocean pCO_2 data (SOCAT)



Model + product pCO_2 validation with GLODAP



Bennington et al., 2022
Gloege et al. 2021, Fay and McKinley 2021

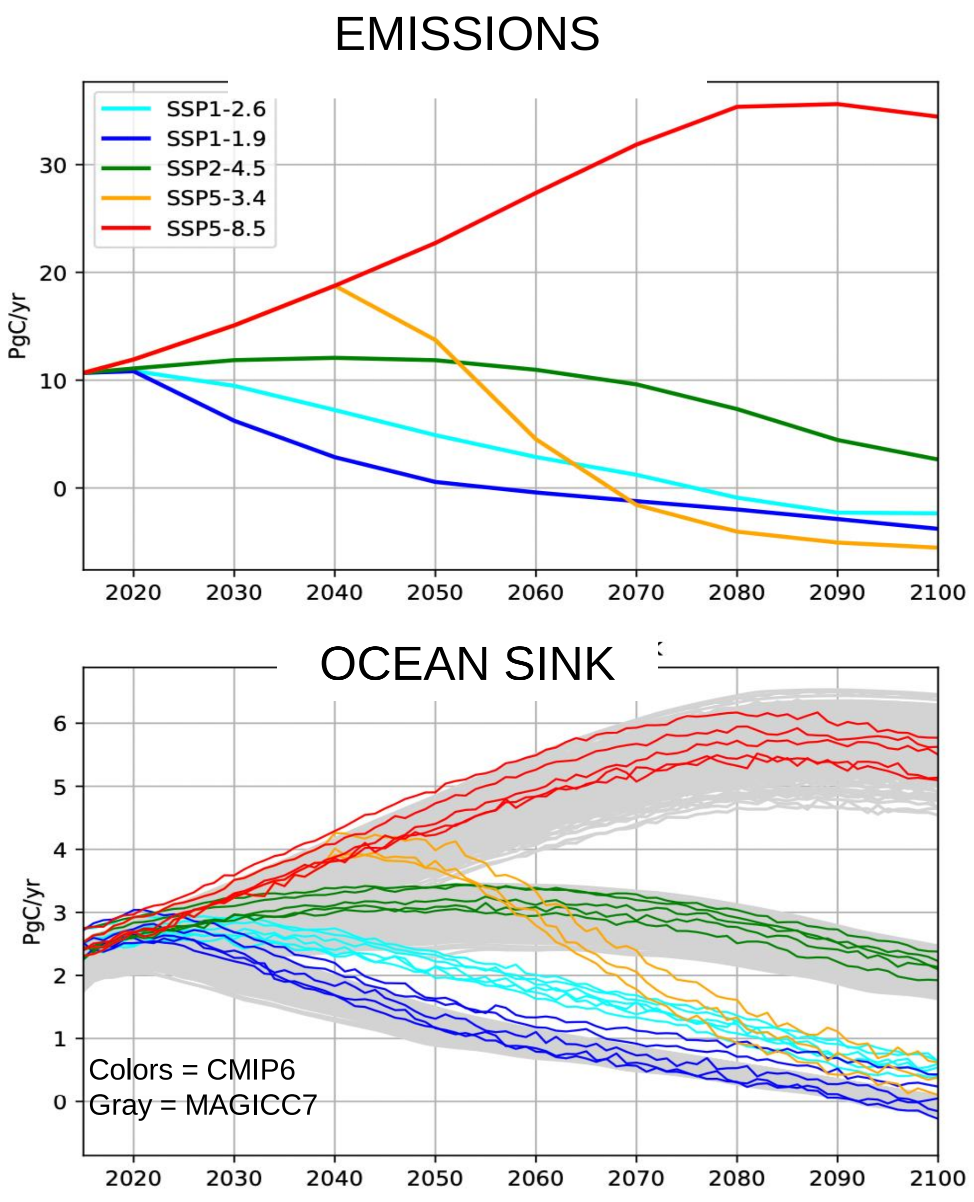
Mean-state biases are retained in CMIP6 projections

CMIP6

3-6 ensemble means
for each scenario

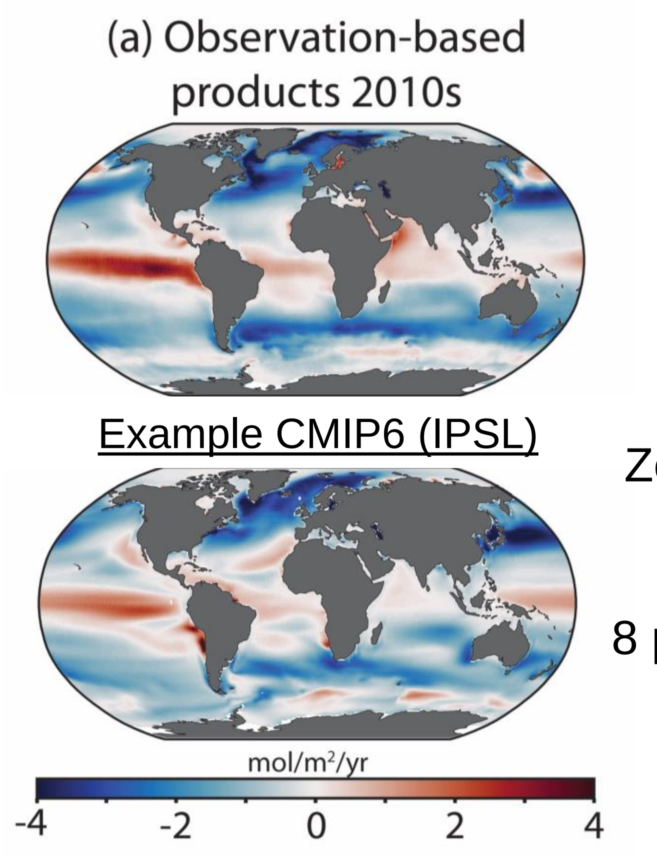
McKinley et al. 2023, ERL

Future Emissions and Ocean Sink



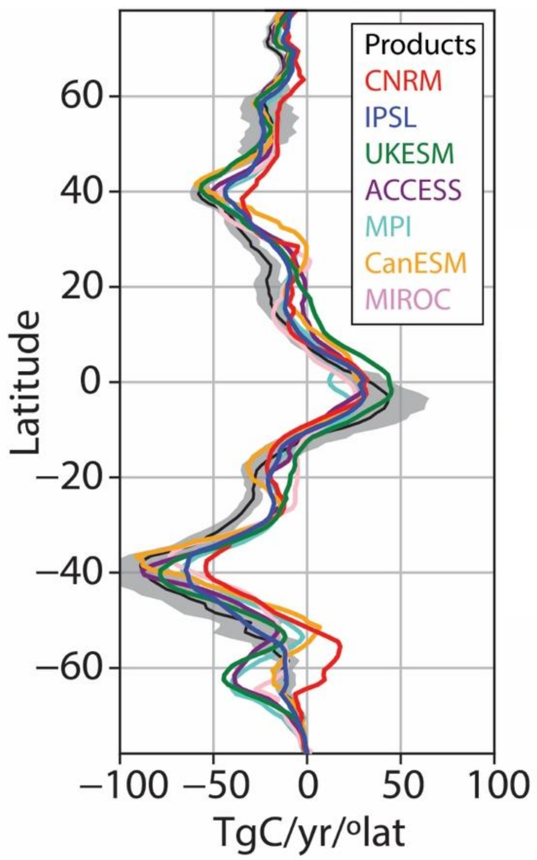
Mean-state biases are retained in CMIP6 projections

Modern = 2010-2020 Flux



Zonal average of this comparison:
7 models (colors)
8 products (black/gray)

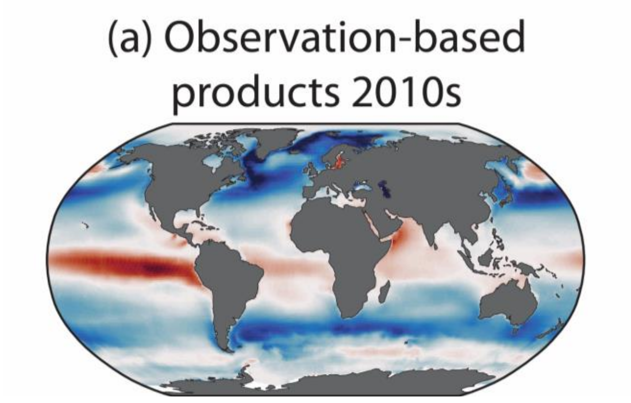
Modern Zonal-mean



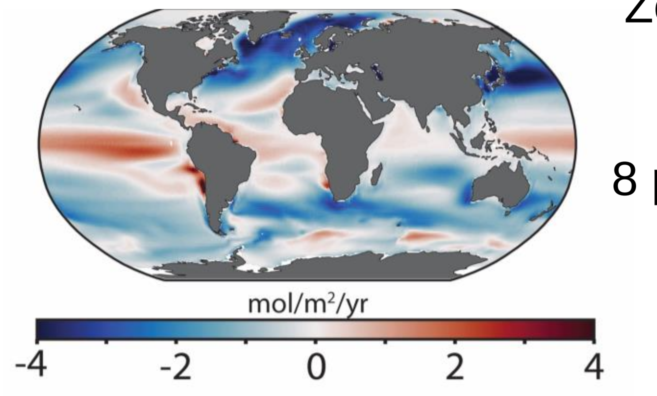
McKinley et al. 2023, ERL

Mean-state biases are retained in CMIP6 projections

Modern = 2010-2020 Flux

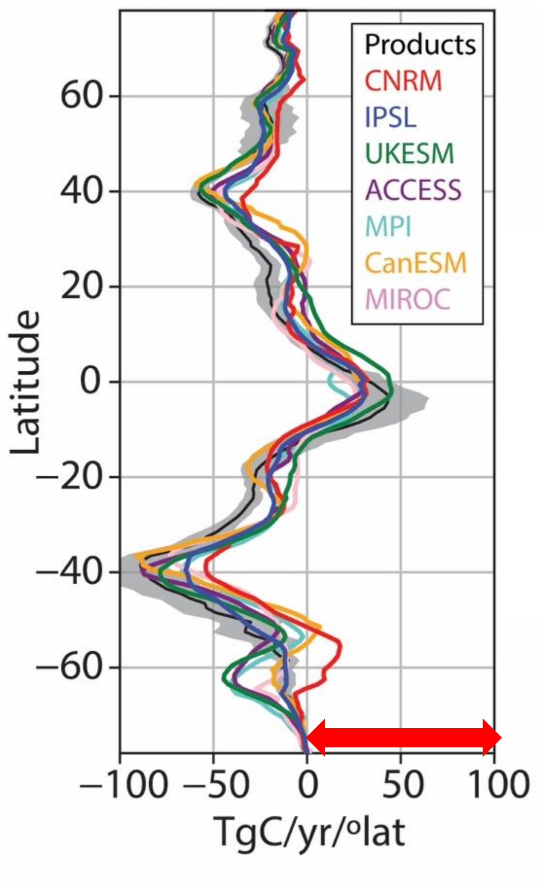


Example CMIP6 (IPSL)



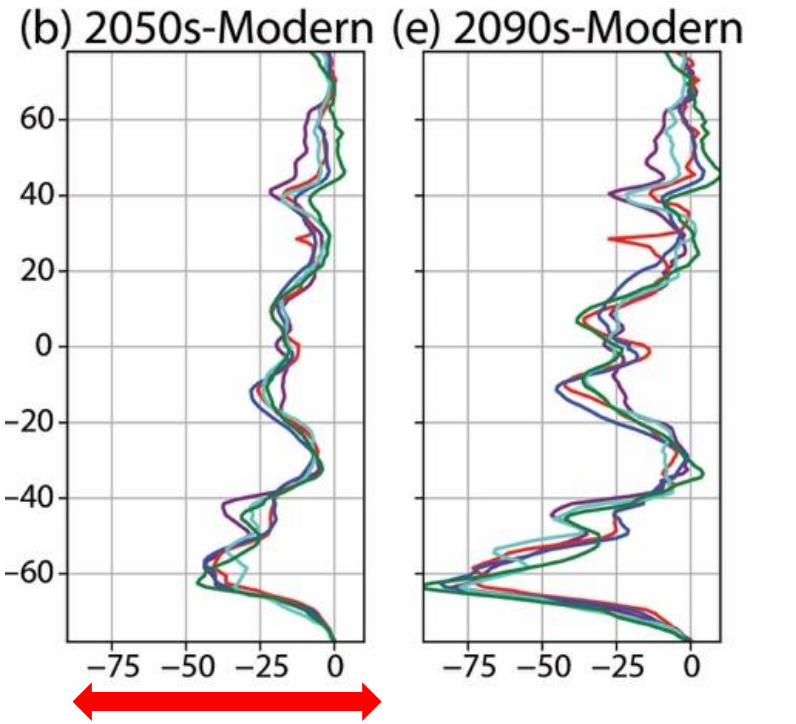
Zonal average of this comparison:
7 models (colors)
8 products (black/gray)

Modern Zonal-mean



Future Flux change

SSP5-8.5

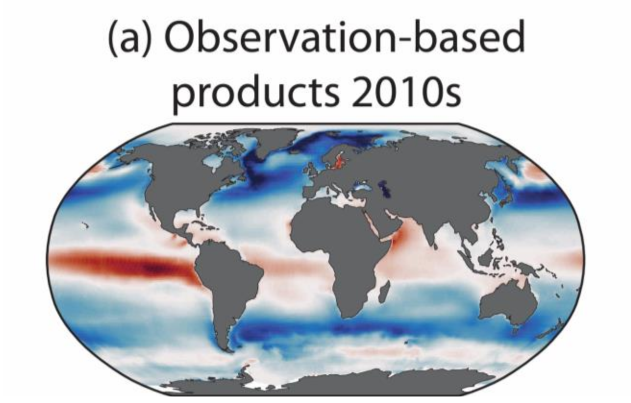


Note: expanded scale

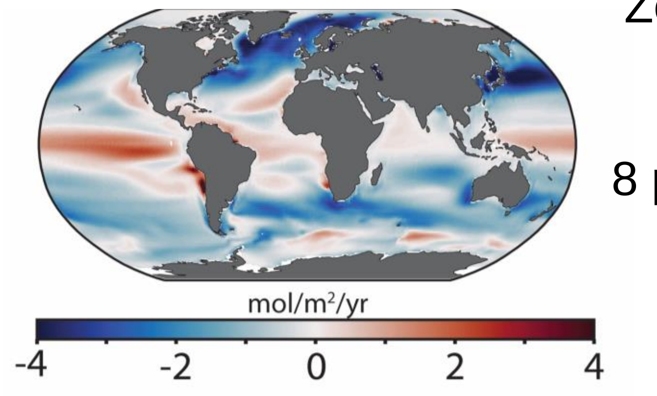
McKinley et al. 2023, ERL

Mean-state biases are retained in CMIP6 projections

Modern = 2010-2020 Flux

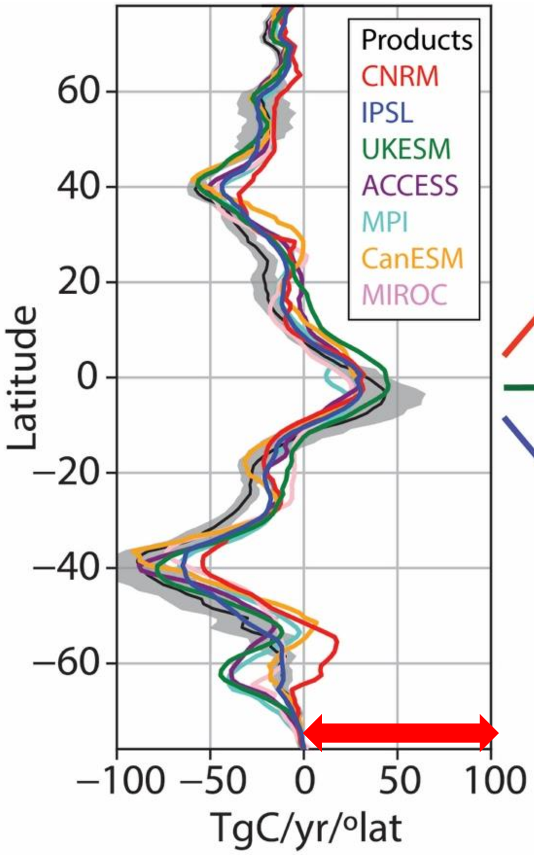


Example CMIP6 (IPSL)

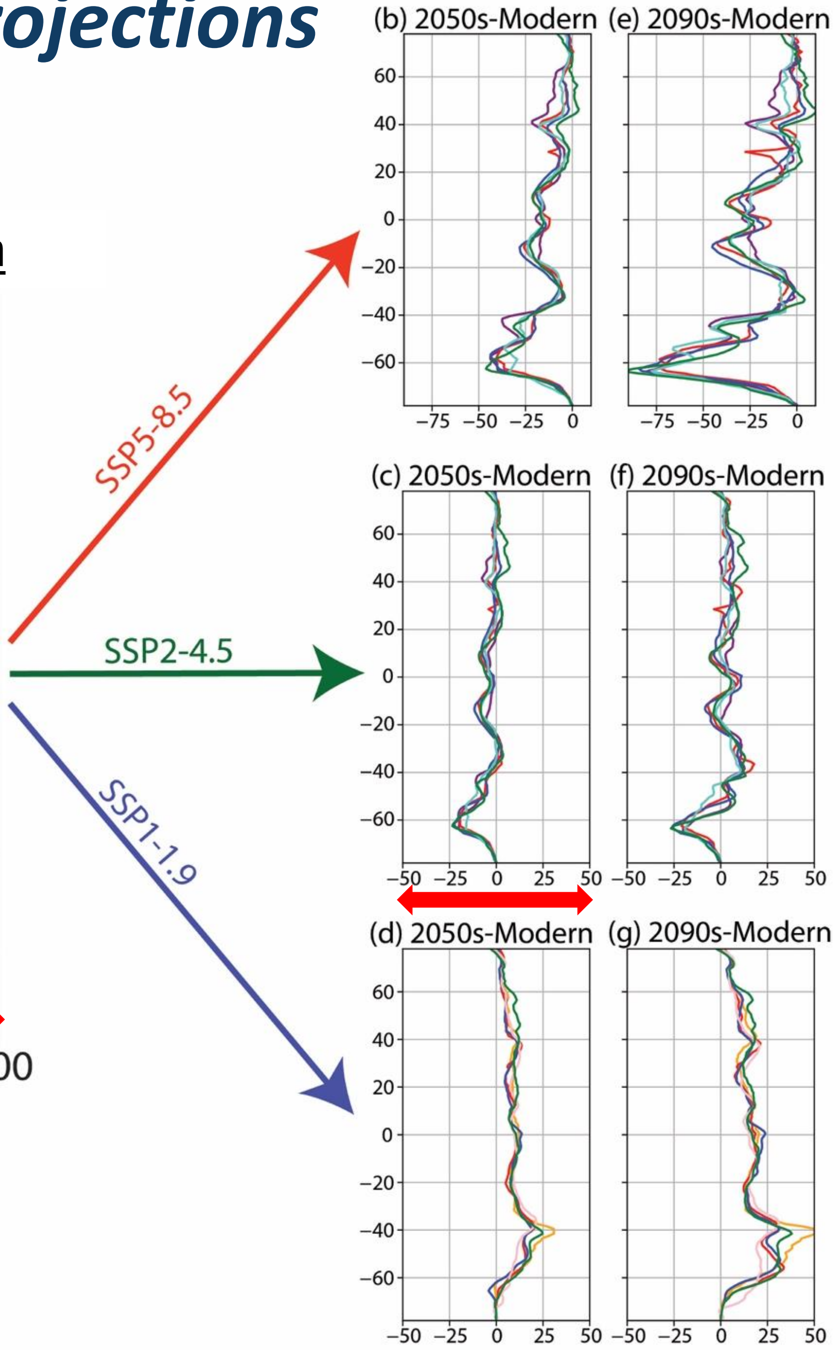


Zonal average of this comparison:
7 models (colors)
8 products (black/gray)

Modern Zonal-mean



Future Flux change

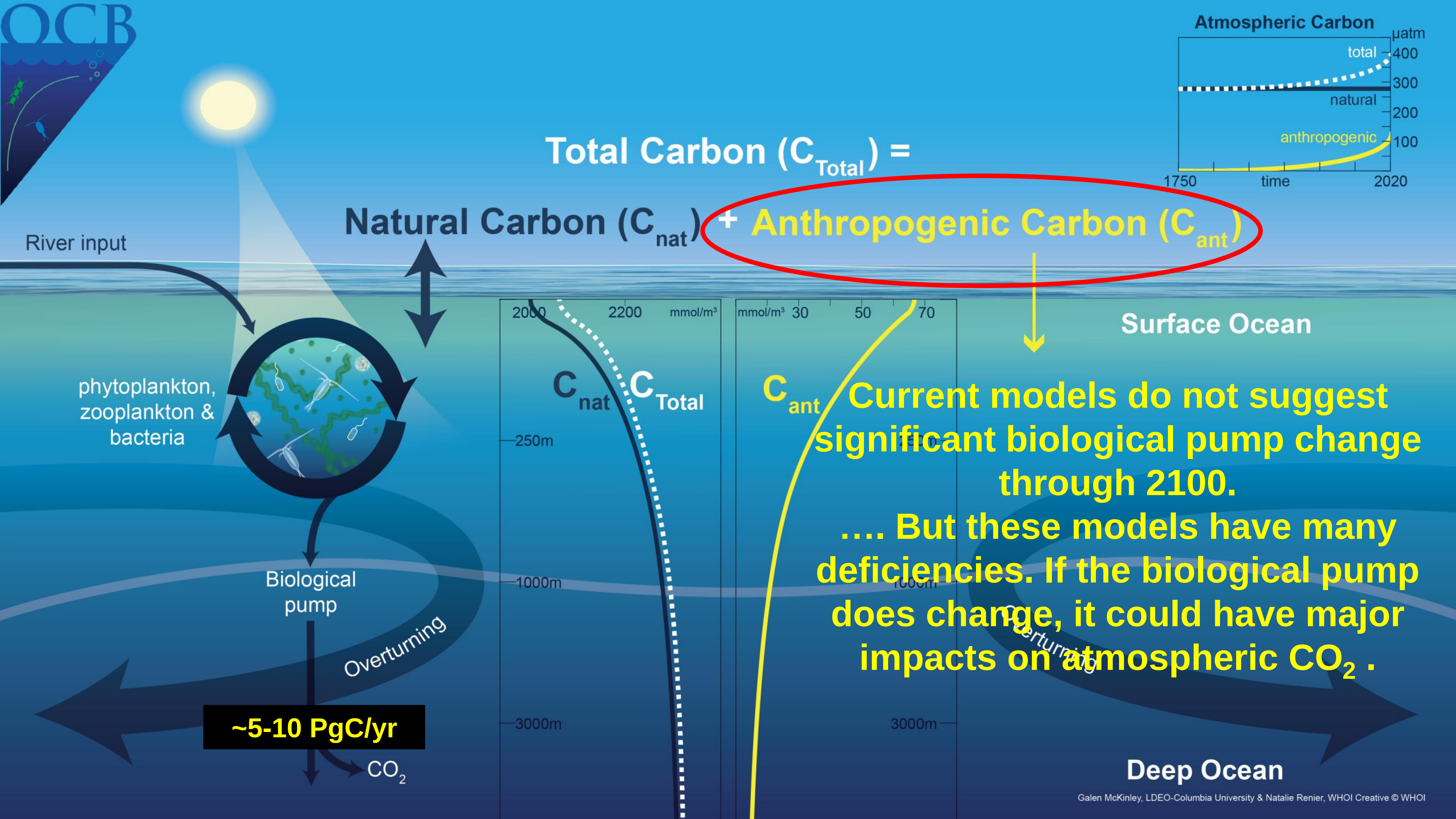


McKinley et al. 2023, ERL

Take home messages

- Air-sea CO_2 flux mean and seasonality can be reconstructed from sparse pCO_2 data, but longer-timescale variability remains uncertain

- Hindcast and Earth System Models have significant mean-state biases



Atmospheric Carbon
µatm
total
natural
anthropogenic
400
300
200
100
1750
time
2020
Total Carbon (C_{Total}) =
Natural Carbon (C_{nat}) + Anthropogenic Carbon (C_{ant})
River input
phytoplankton, zooplankton & bacteria
Biological pump
Overturning
~5-10 PgC/yr
CO_2
2000
2200
mmol/m³
C_{nat}
C_{Total}
250m
1000m
3000m
mmol/m³
30
50
70
C_{ant}
Surface Ocean
Current models do not suggest significant biological pump change through 2100.
…. But these models have many deficiencies. If the biological pump does change, it could have major impacts on atmospheric CO_2 .
Deep Ocean
Galen McKinley, LDEO-Columbia University & Natalie Renier, WHOI Creative © WHOI

Outline

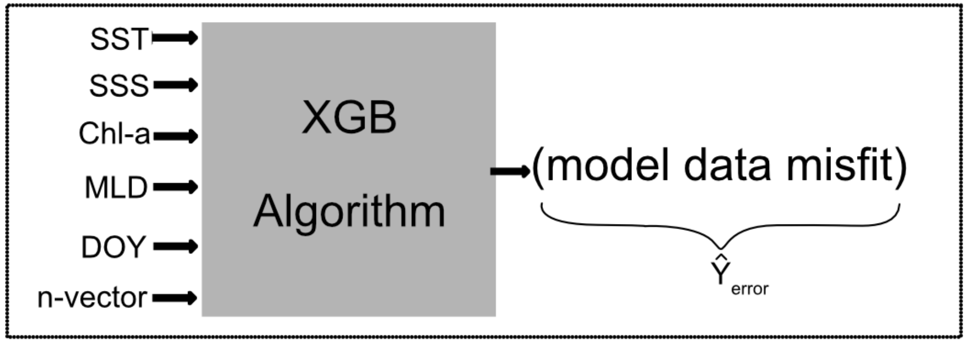
- Key processes of the ocean carbon cycle

- Improving quantification of the ocean carbon sink
  - *Models*: evaluate ML-based data product skill, given sampling
  - *pCO_2 data products*: apply to identify model mean-state biases
  - ***Models and data*: combine using ML**
  - *marine Carbon Dioxide Removal (mCDR)* – quantifying “additionality”?

Using ML to combine models and data improves surface flux estimates (LDEO-HPD)

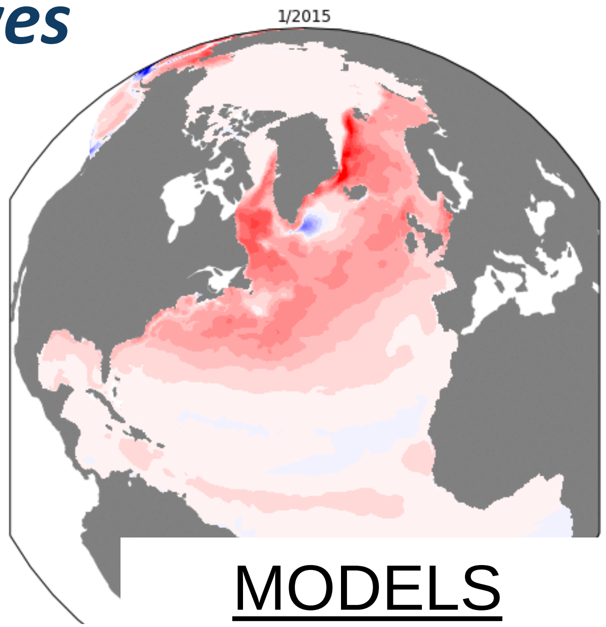
- pCO_2^{model} as priors
- Reconstruct $pCO_2^{misfit} = (pCO_2^{model} - pCO_2^{SOCAT})$

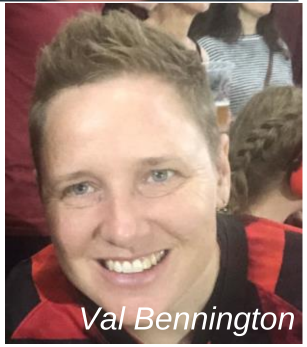
  (driver data: SST, SSS, Chl-a, MLD, time, location)
- eXtreme Gradient Boost (XGB) algorithm
- Misfits to correct N models; average for final pCO_2



$$\underbrace{\text{(model data misfit)}}_{\hat{Y}_{error}} + \underbrace{pCO_2^{model}}_{Y_{model}} = \widehat{pCO_2}$$

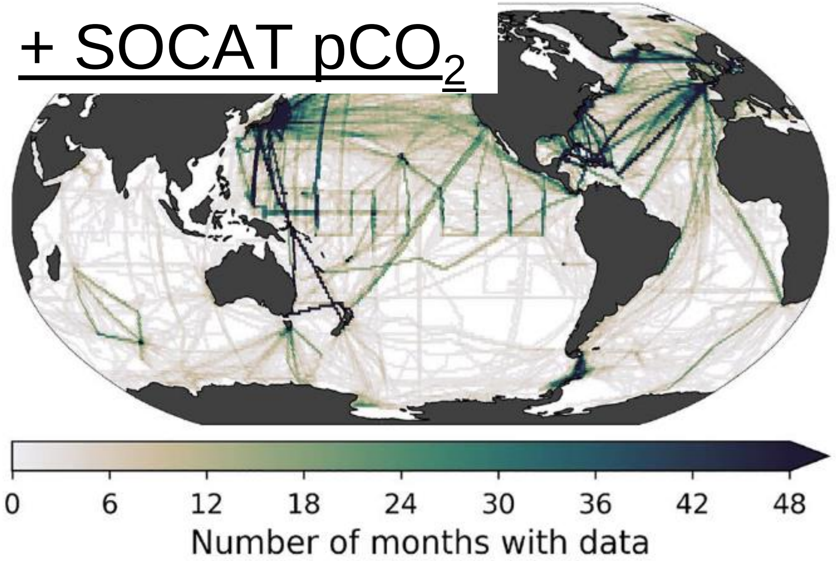
Gloege et al. 2022 JAMES (N=9), *Bennington* et al. 2022 GRL (N=8)



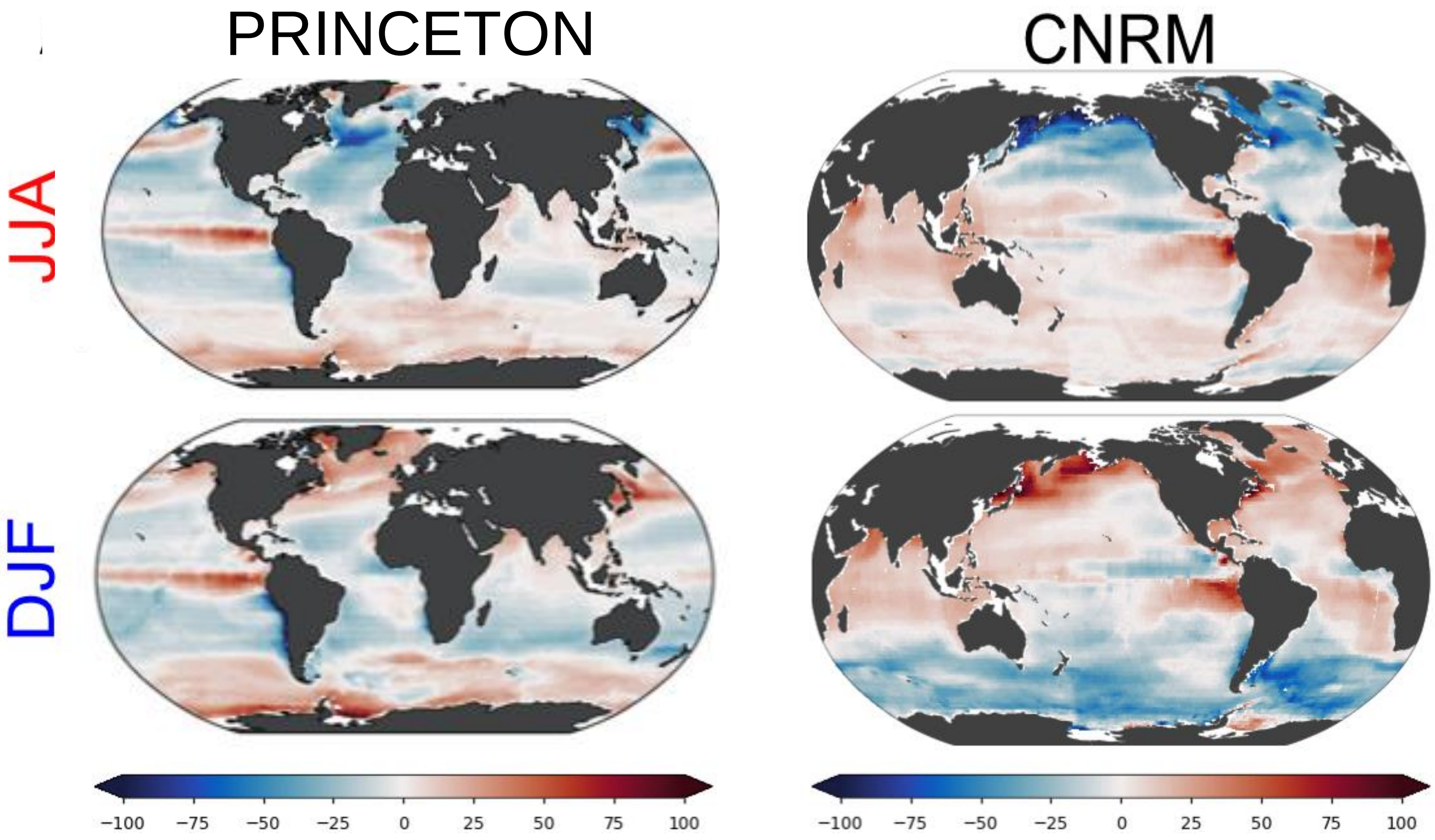


MODELS

+ SOCAT pCO_2



Full-coverage, monthly misfits illustrate substantial model biases

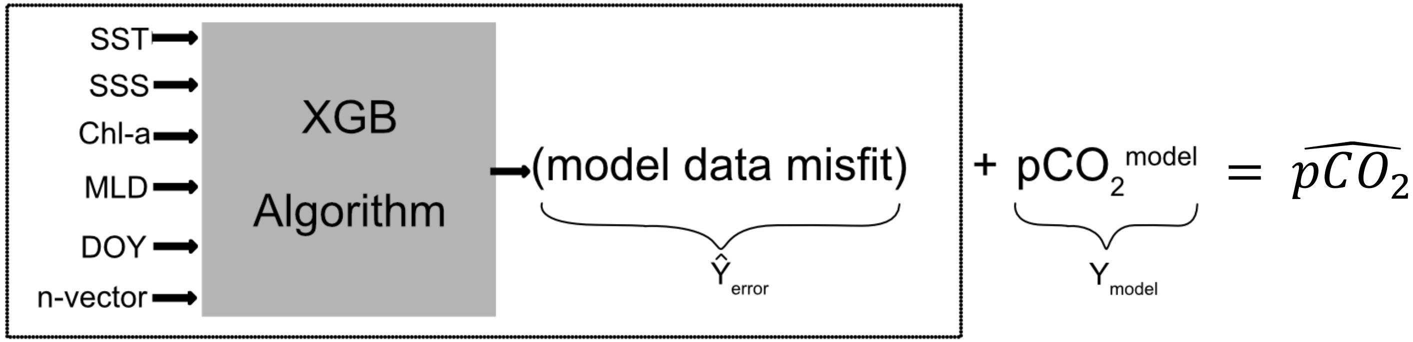


pCO_2^{misfit} (uatm)

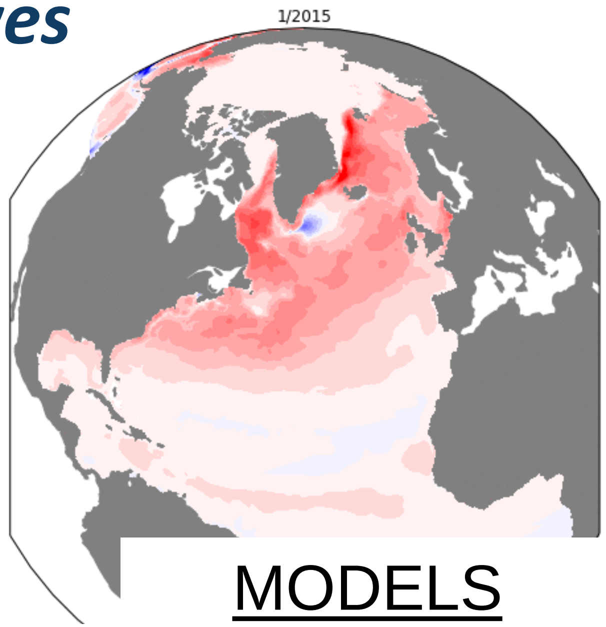
Gloege et al. 2022, JAMES

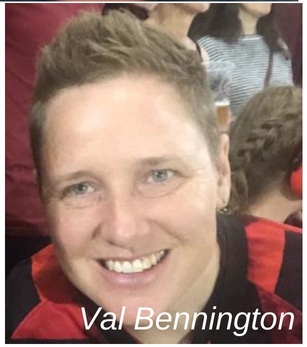
Using ML to combine models and data improves surface flux estimates (LDEO-HPD)

- pCO_2^{model} as priors
- Reconstruct $pCO_2^{misfit} = (pCO_2^{model} - pCO_2^{SOCAT})$
  (driver data: SST, SSS, Chl-a, MLD, time, location)
- eXtreme Gradient Boost (XGB) algorithm
- **Misfits to correct N models; average for final pCO_2**



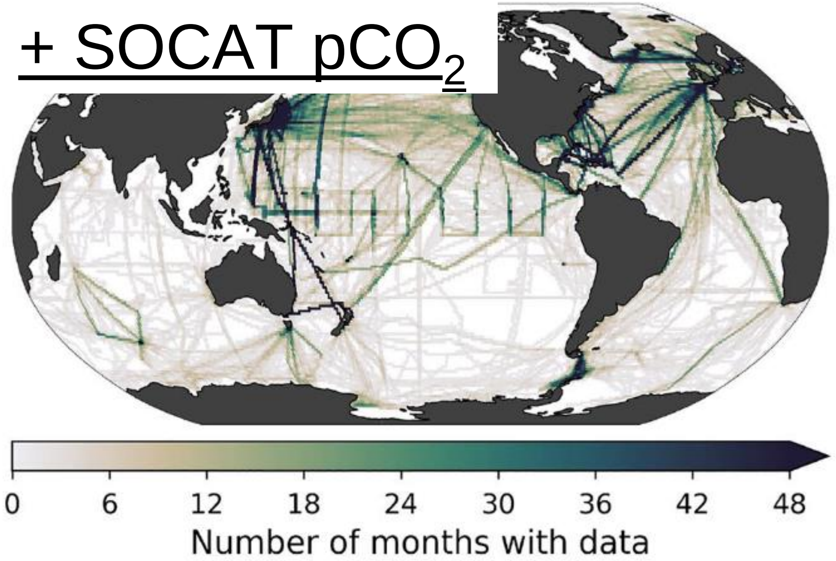
Gloege et al. 2022 JAMES (N=9), *Bennington* et al. 2022 GRL (N=8)



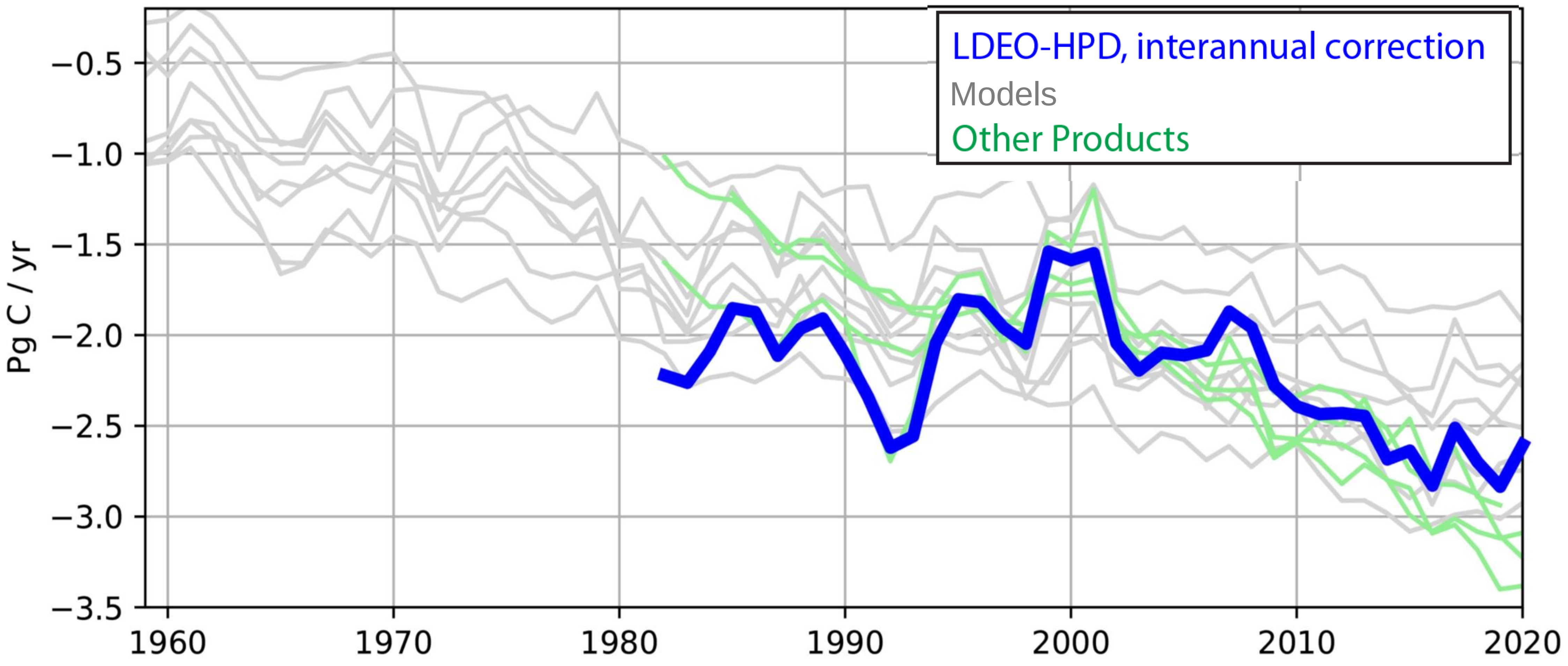


MODELS

+ SOCAT pCO_2

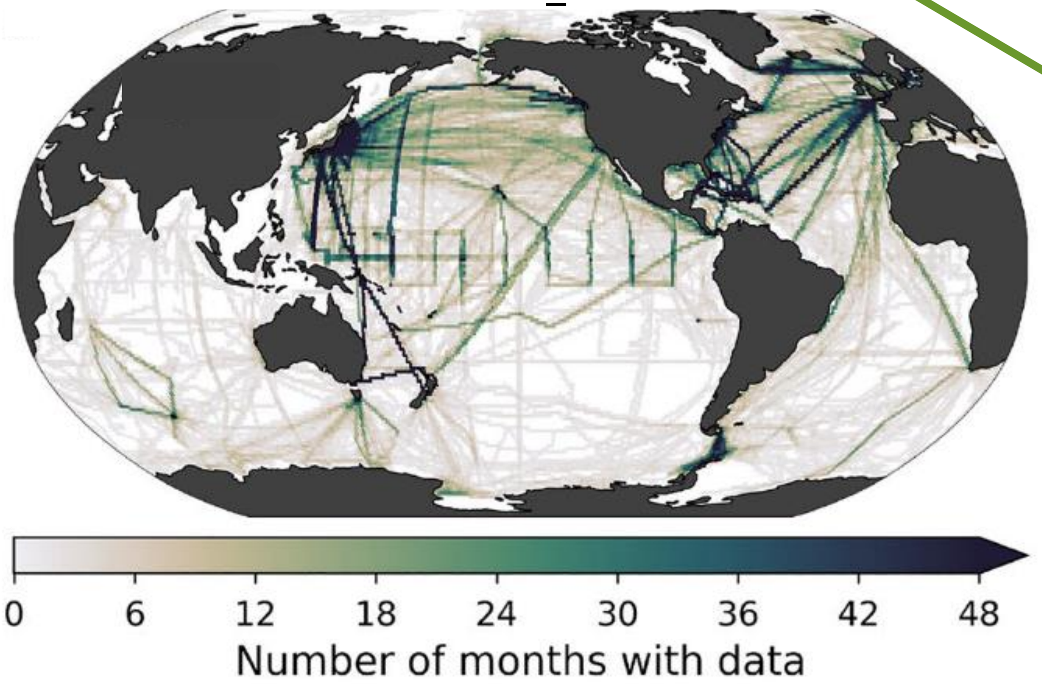


Global air-sea CO_2 flux

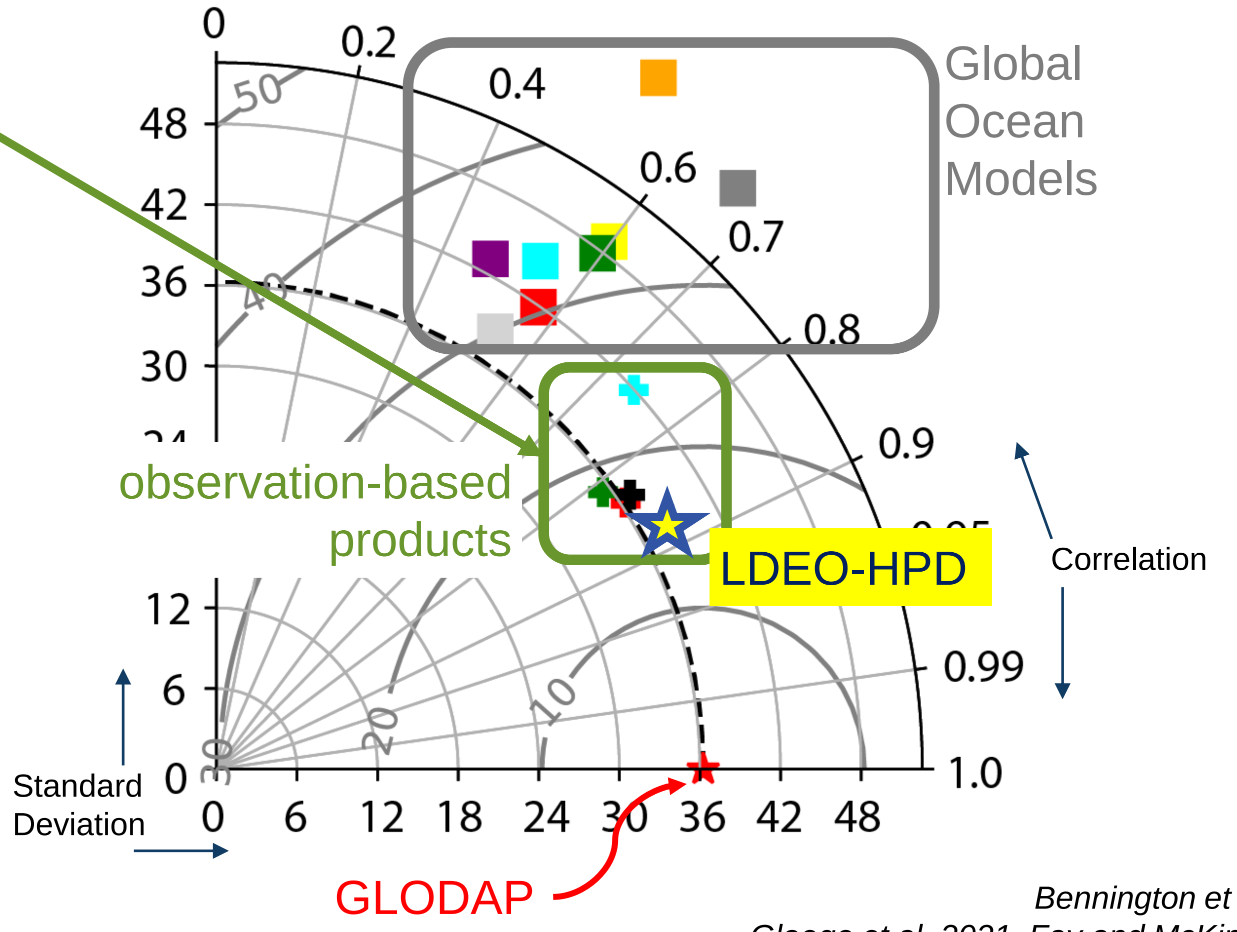


LDEO-HPD hybrid model-data approach offers improved skill

Surface ocean pCO_2 data (SOCAT)



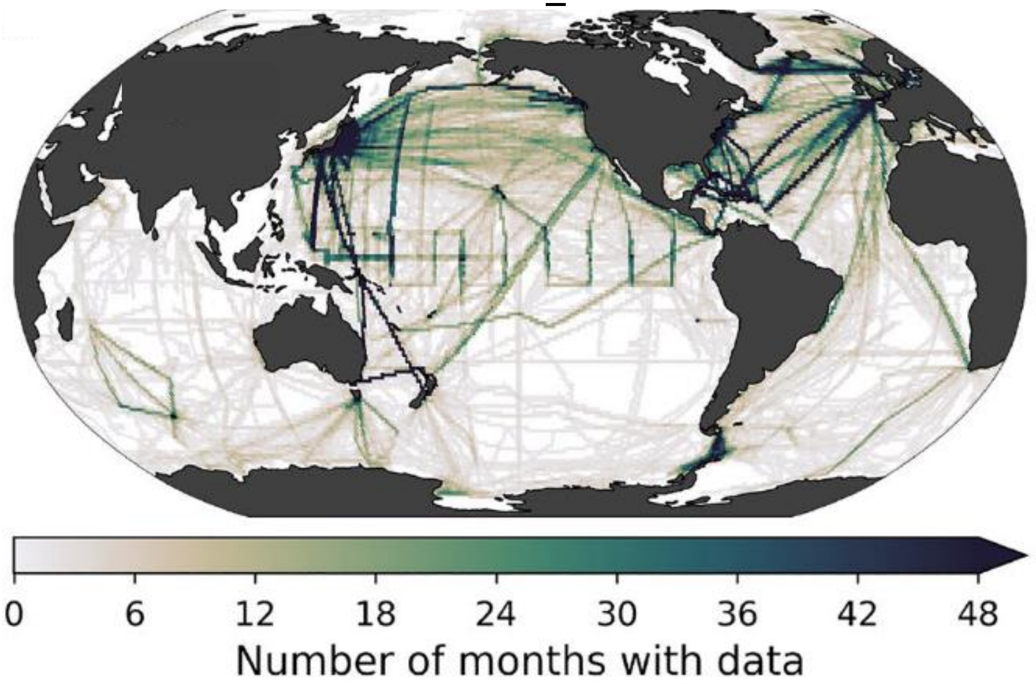
Model + product pCO_2 validation with GLODAP



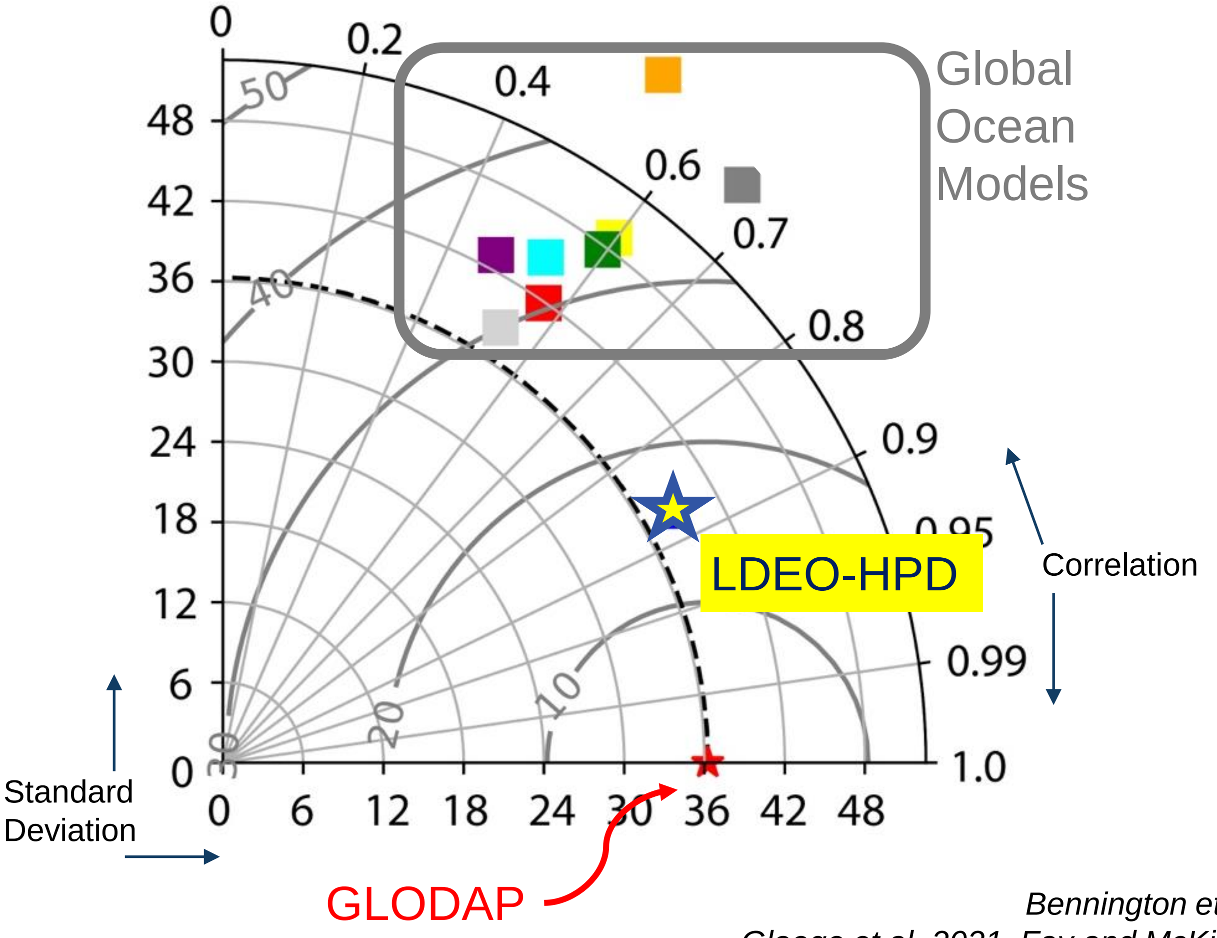
Bennington et al., 2022
Gloege et al. 2021, Fay and McKinley 2021

From where does this enhanced skill arise?

Surface ocean pCO_2 data (SOCAT)

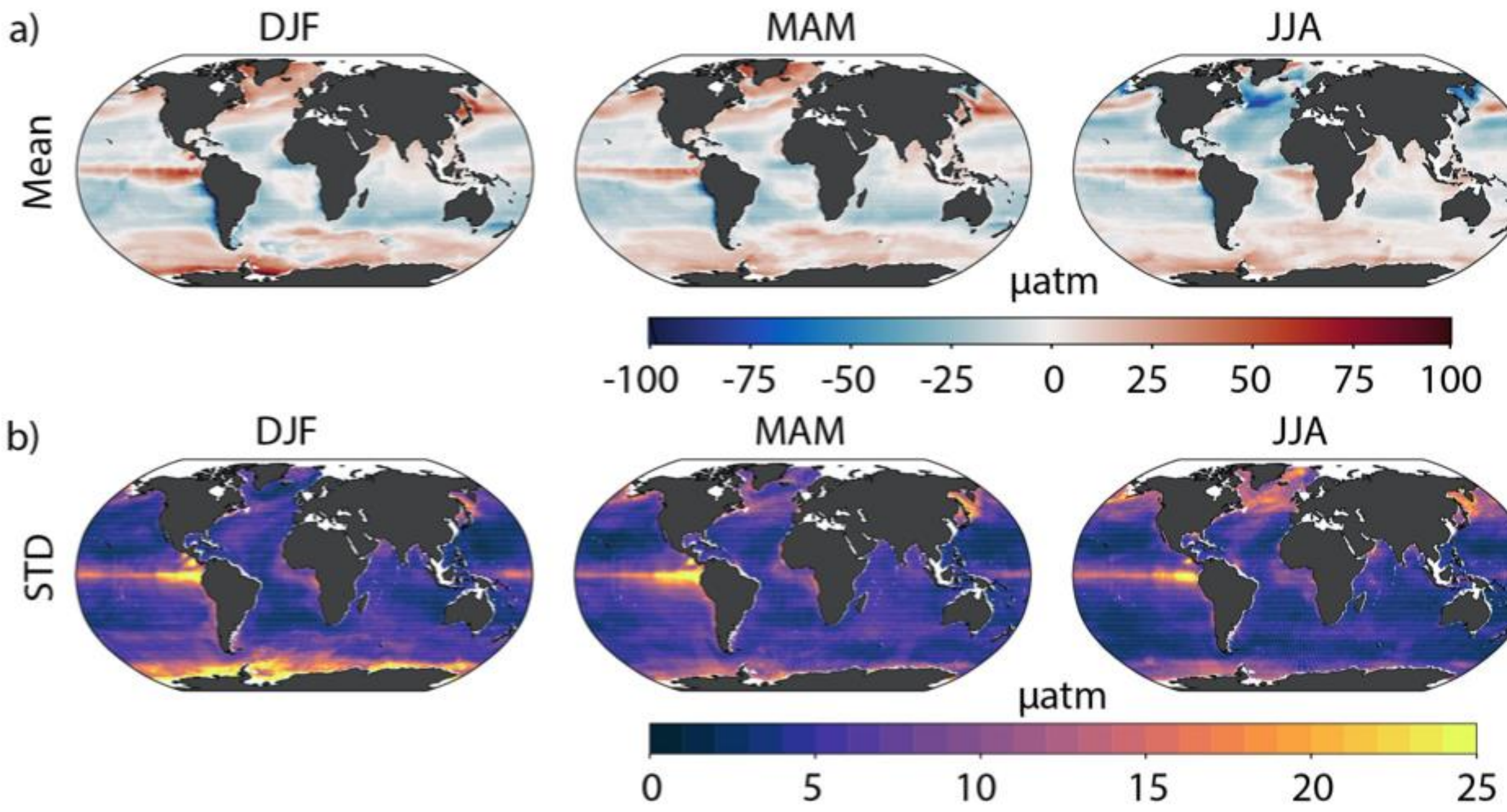


Model + product pCO_2 validation with GLODAP



Bennington et al., 2022
Gloege et al. 2021, Fay and McKinley 2021

Climatological model-data misfits much larger than interannual



Princeton Model, others similar

Bennington et al. 2022 GRL

Since climatological misfit dominates, how much skill is gained by applying only this as correction, as opposed to interannual correction?

- $HPD_{ClimTest}$ applies the 2000-2020 climatology of the model-observation misfit

LDEO-HPD = Model pCO_2 + Interannual Misfit

$HPD_{ClimTest}$ = Model pCO_2 + Climatological Misfit

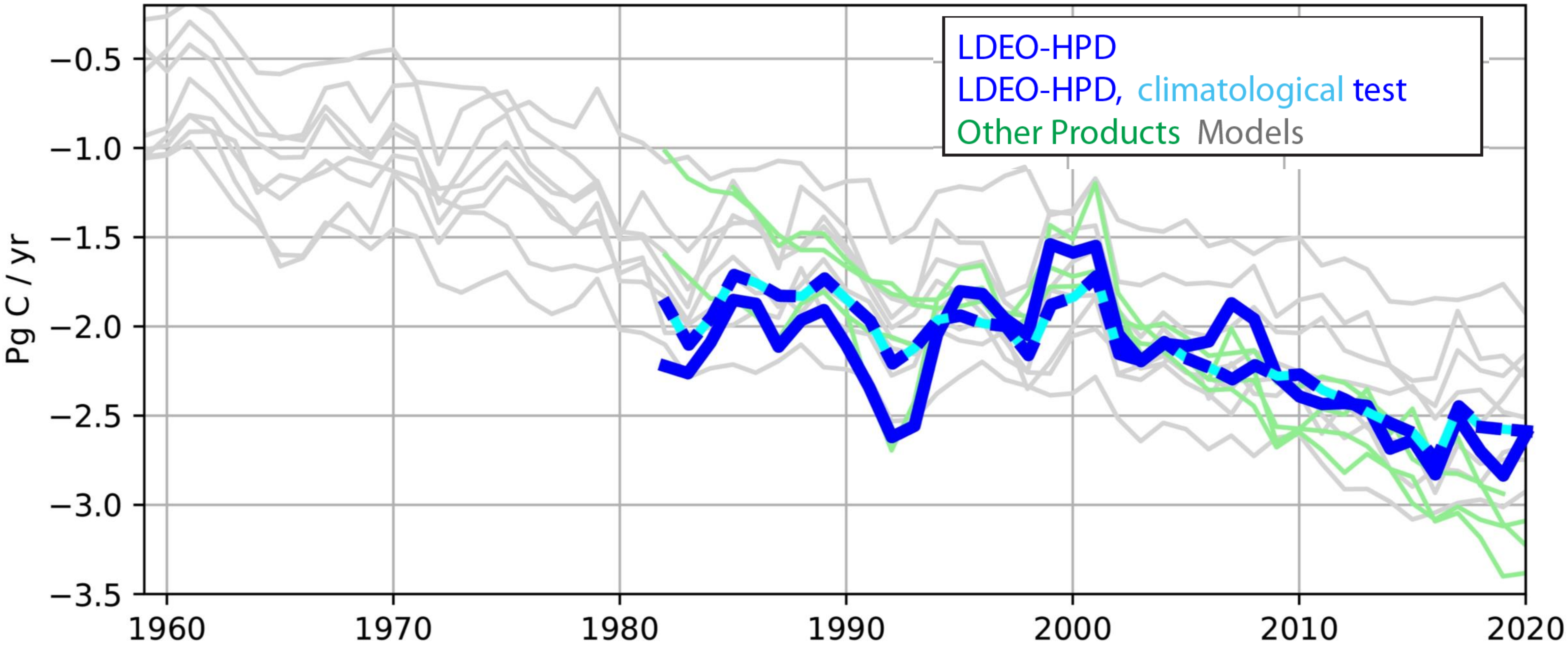
1982

Observations Begin

2020

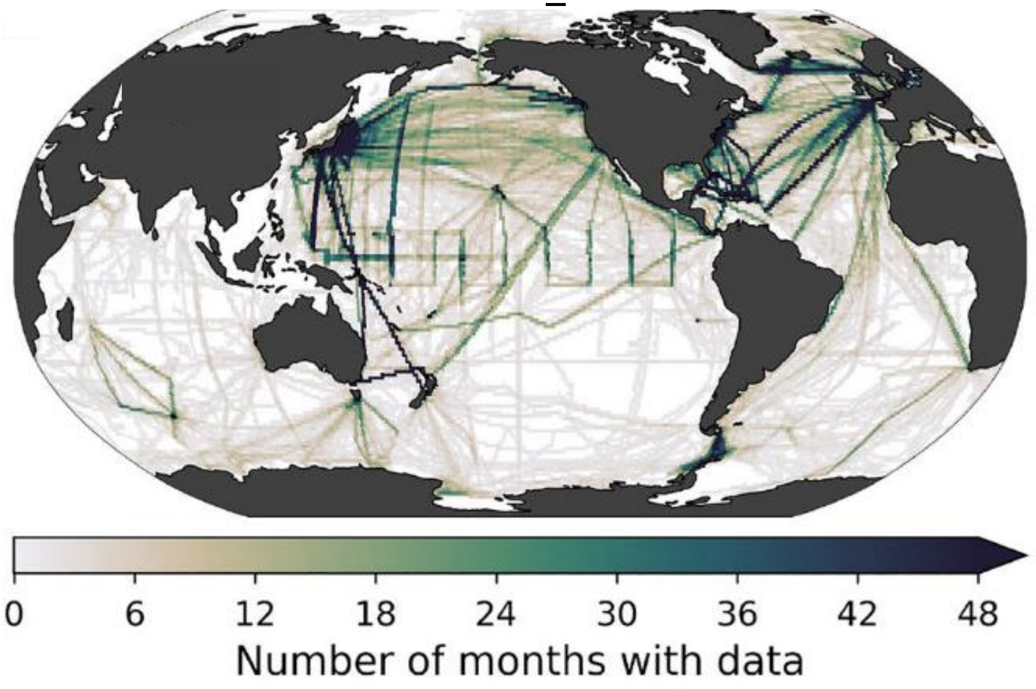
Bennington et al. 2022 GRL

Global air-sea CO_2 flux

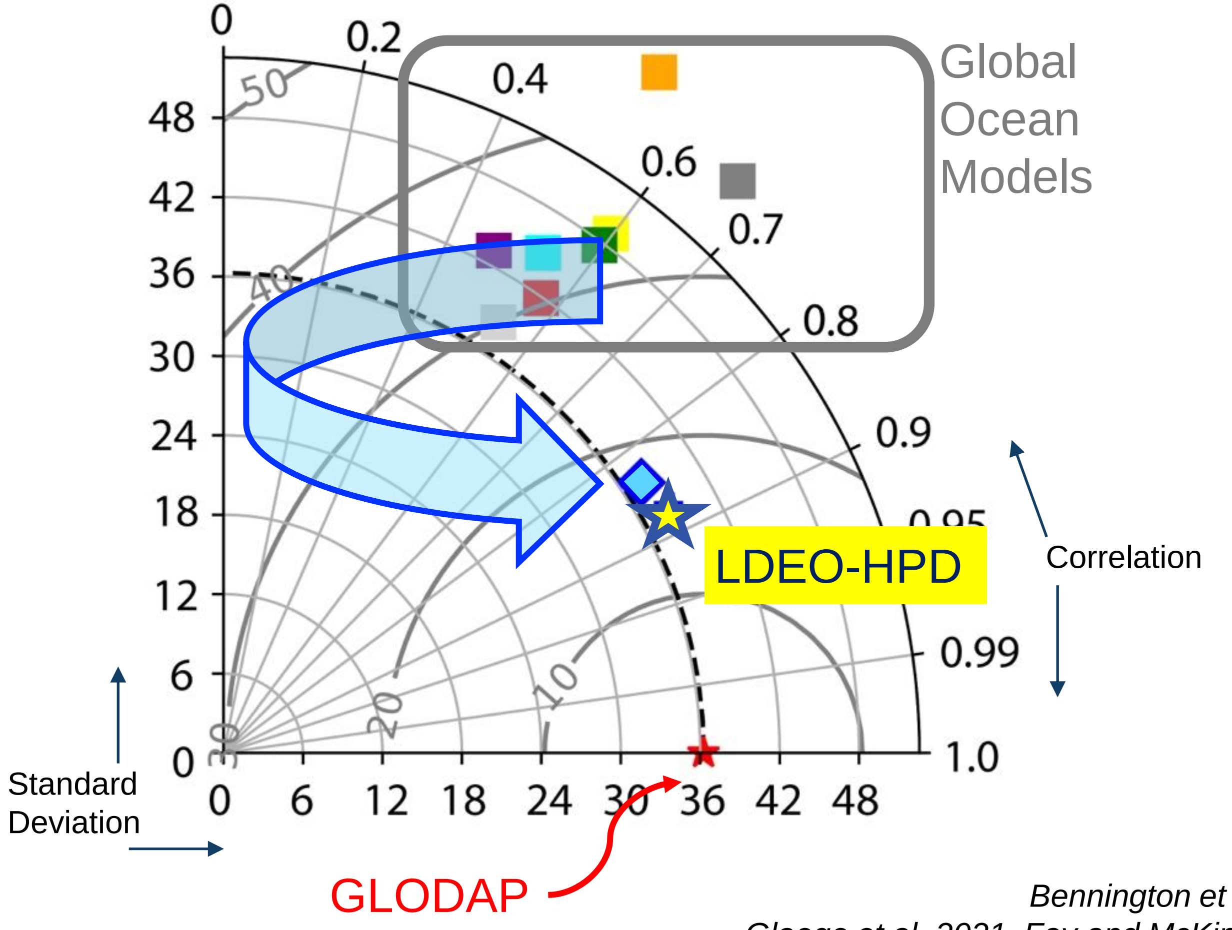


Most of the improvement comes from the climatological correction

Surface ocean pCO_2 data (SOCAT)



Model + product pCO_2 validation with GLODAP



Bennington et al., 2022
Gloege et al. 2021, Fay and McKinley 2021

COLUMBIA CLIMATE SCHOOL
LAMONT-DOHERTY EARTH OBSERVATORY

Taking advantage of the climatological misfit being the biggest source of model error, we extend back to 1959

- HPD$_{\text{ClimTest}}$ applies the 2000-2020 climatology of the model-observation misfit

- ***Since the climatological correction provides most of the additional skill, we use it to correct models in the pre-observed period***

LDEO-HPD = Model pCO_2 + **Climatological Misfit**

LDEO-HPD = Model pCO_2 + **Interannual Misfit**

1959
Model Period Begins

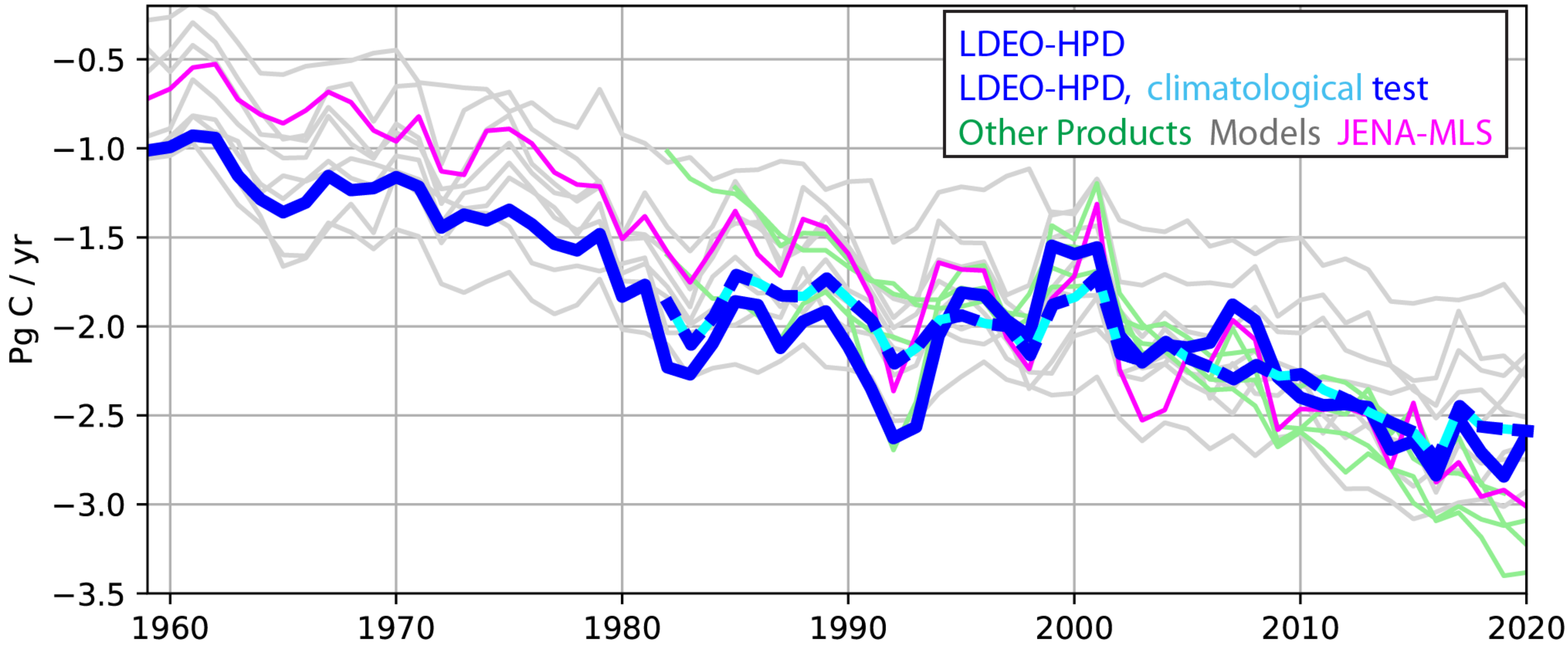
1982
Observations Begin

HPD$_{\text{ClimTest}}$ = Model pCO_2 + Climatological Misfit

2020

Bennington et al. 2022 GRL

Global air-sea CO_2 flux

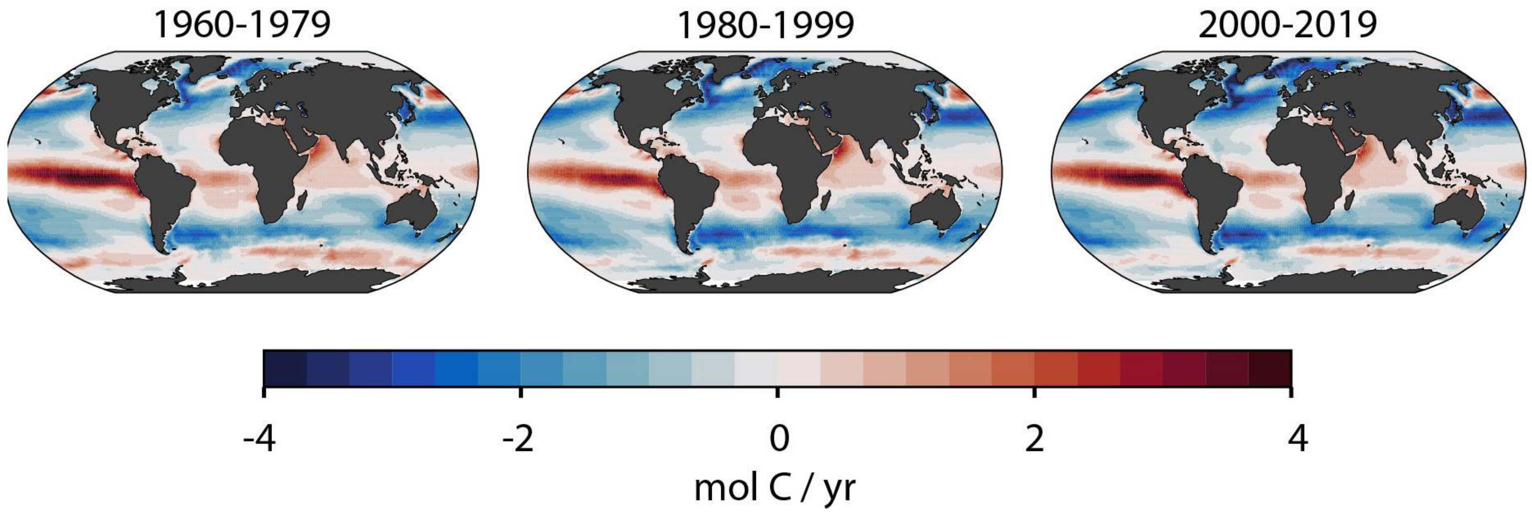


Climatological correction
1959-1982

Interannually varying correction
1982-2020

Bennington et al. 2022 GRL

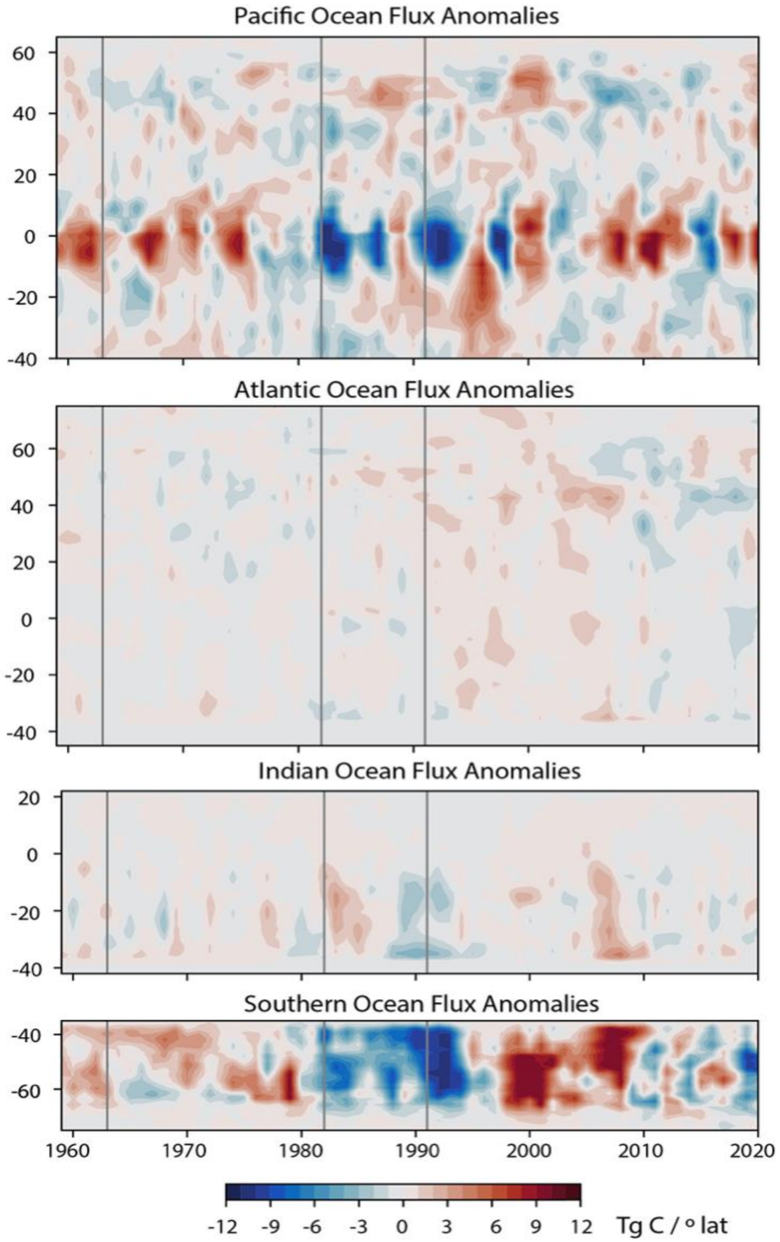
LDEO-HPD: 60+ years of air-sea CO_2 fluxes



- 60+ years of monthly 1x1 air-sea CO_2 fluxes

- Significant decadal variations; coherent between equatorial Pacific and Southern Ocean

Detrended Basin Flux Anomalies



Bennington et al. 2022 GRL

COLUMBIA CLIMATE SCHOOL
LAMONT-DOHERTY EARTH OBSERVATORY

Take home messages

- Air-sea CO_2 flux mean and seasonality can be reconstructed from sparse pCO_2 data, but longer-timescale variability remains uncertain

- Hindcast and Earth System Models have significant mean-state biases

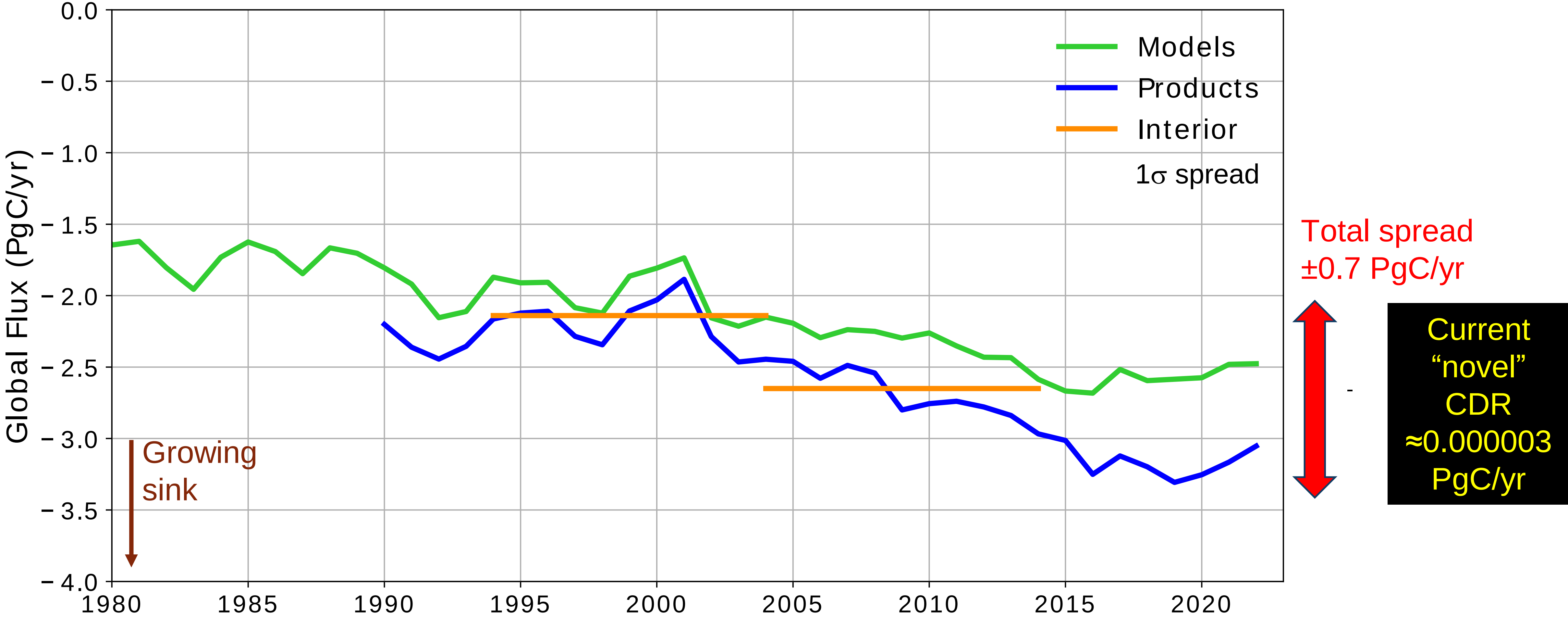
- Correcting models with data offers improved skill and 60+ years of reconstructed monthly surface ocean pCO_2

Outline

- Key processes of the ocean carbon cycle

- Improving quantification of the ocean carbon sink
  - *Models*: evaluate ML-based data product skill, given sampling
  - *pCO_2 data products*: apply to identify model mean-state biases
  - *Models and data*: combine using ML
  - ***marine Carbon Dioxide Removal (mCDR)*** **– quantifying “additionality”?**

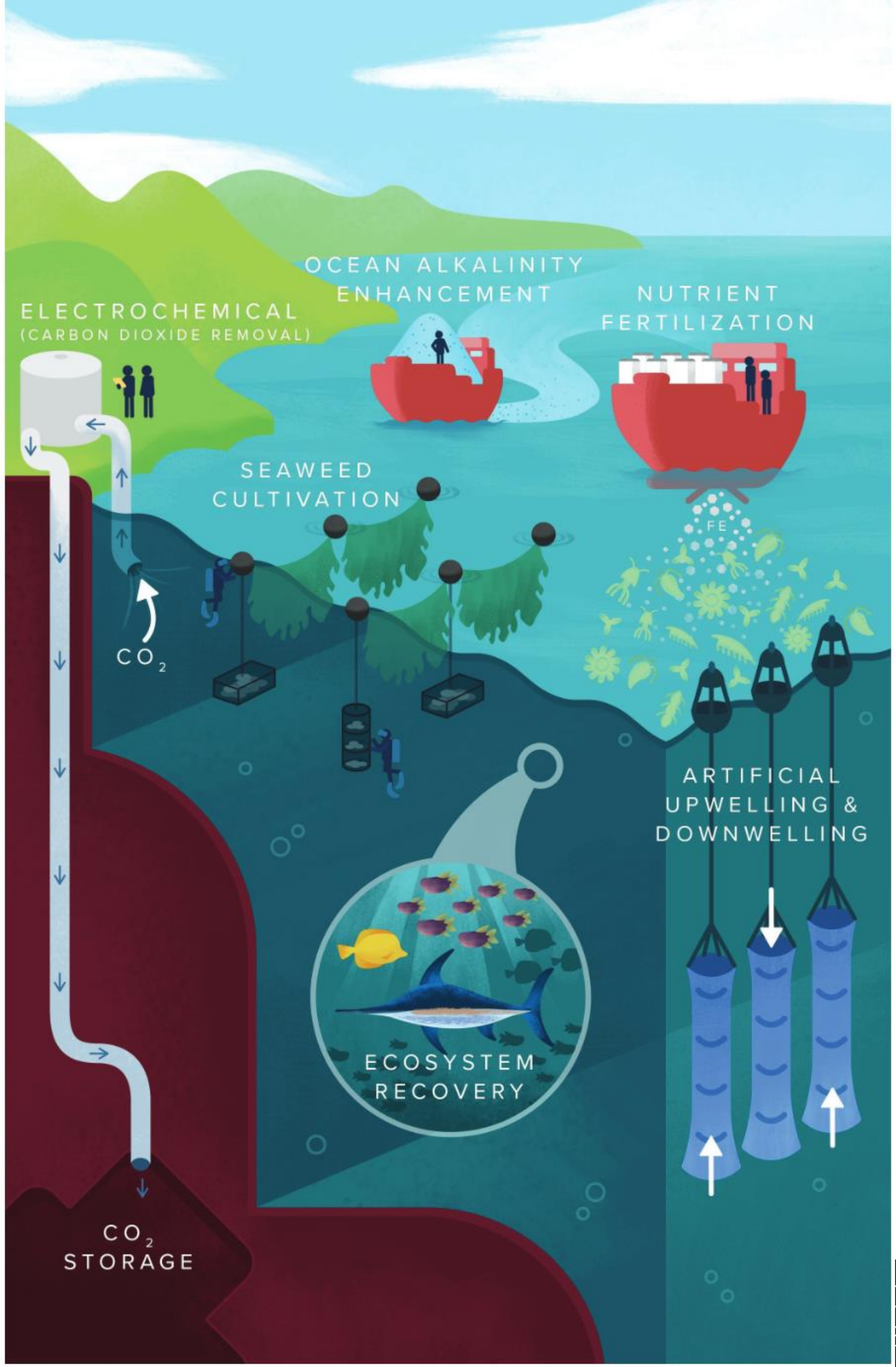
“novel” CDR is imperceptible with respect to the ocean sink



Models, products, CDR from Friedlingstein et al. 2023
Interior from Müller et al. 2023

Marine Carbon Dioxide Removal mCDR

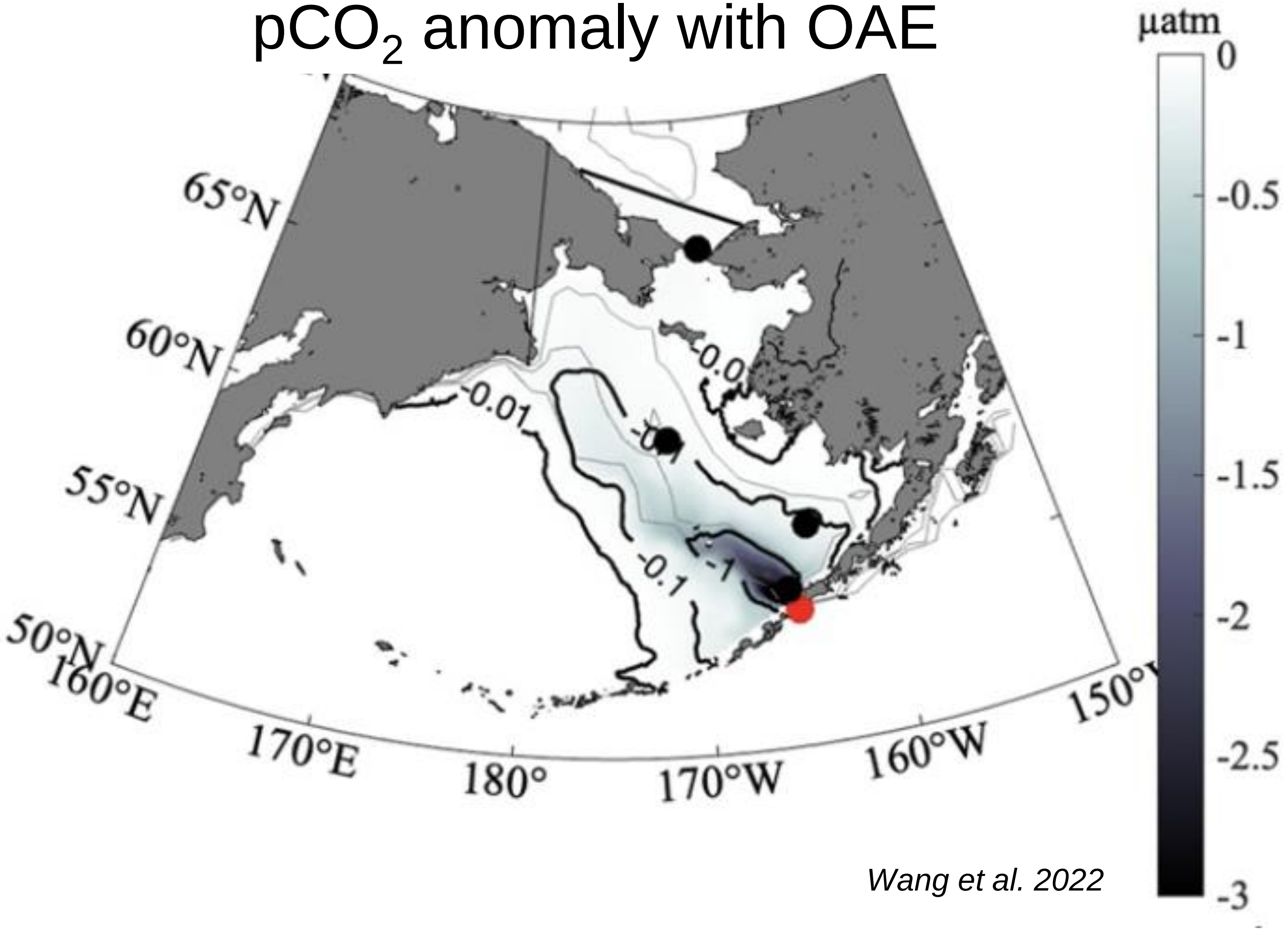
- Efficacy?
- Safety?
- Durability?
- **Additionality?**



NASEM 2021

mCDR pCO_2 impacts

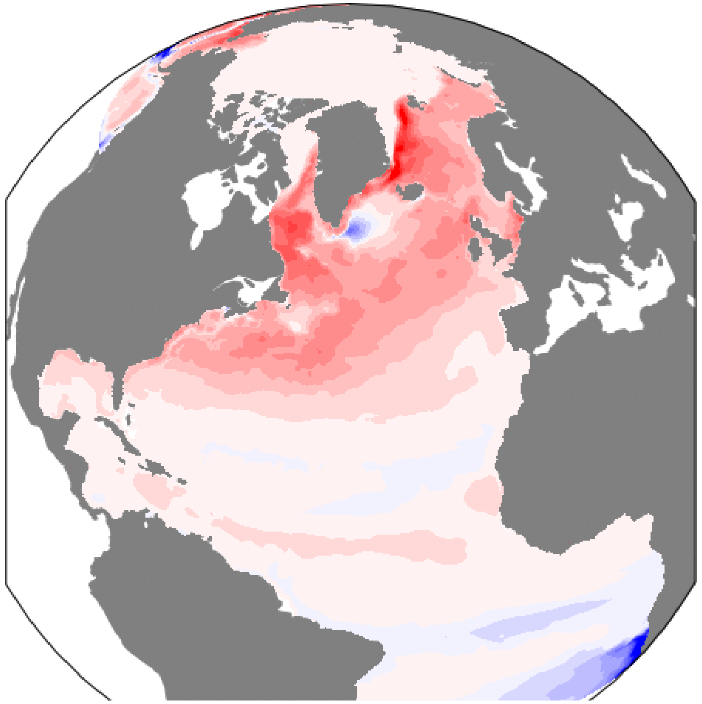
10-year mean modeled pCO_2 anomaly with OAE



Wang et al. 2022

Global approaches potential for time-space resolved baseline

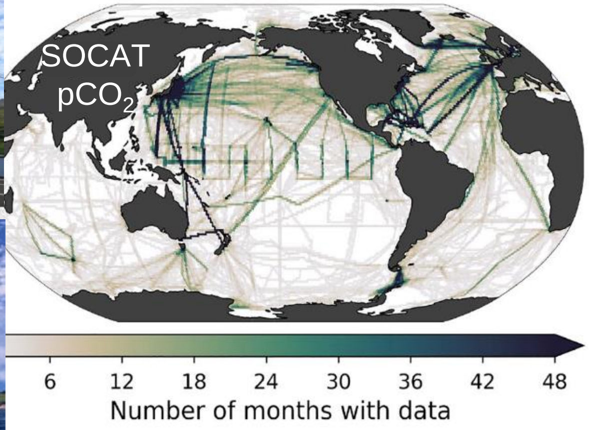
Models



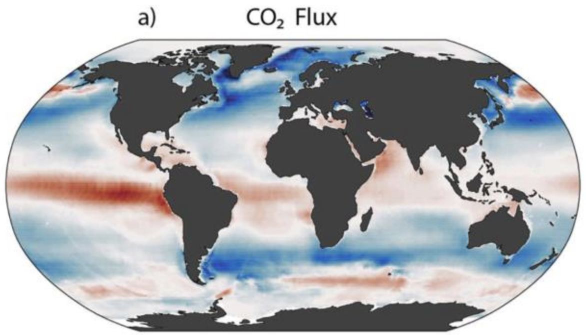
ASTE-BGC
air-sea CO_2 flux

Surface pCO_2 data products

Air-sea fluxes (~monthly)
Model validation



AI/ML

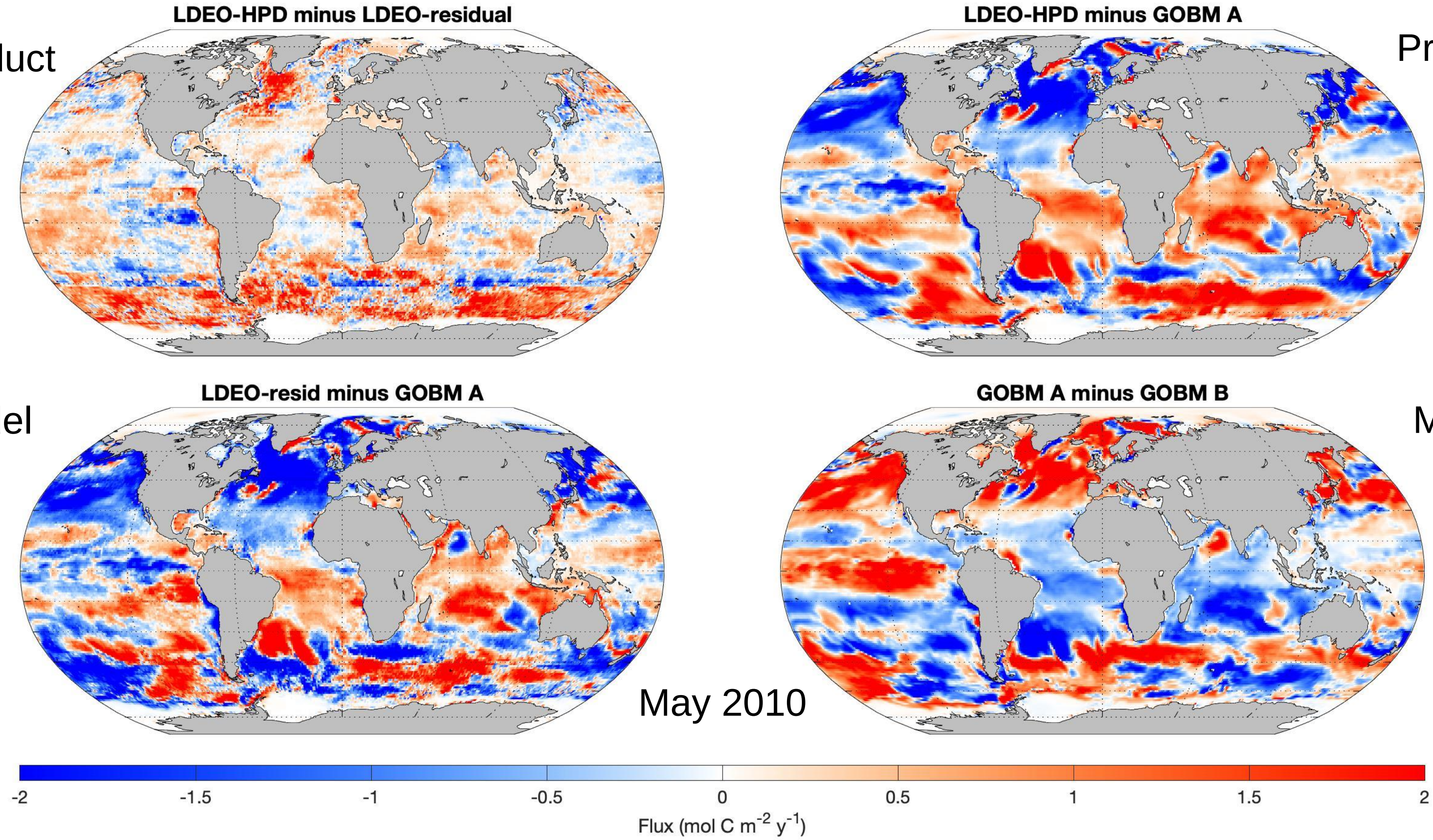


Air-sea flux uncertainties at smaller space-time scales are huge

Uncertainties are estimated here as difference between individual estimates

Product-product

Product-model



Product-model

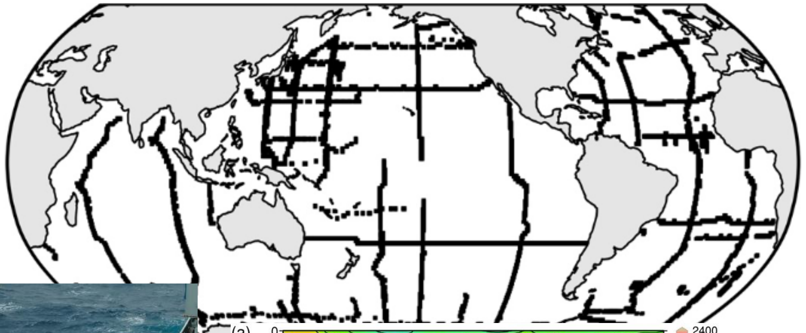
Model-model

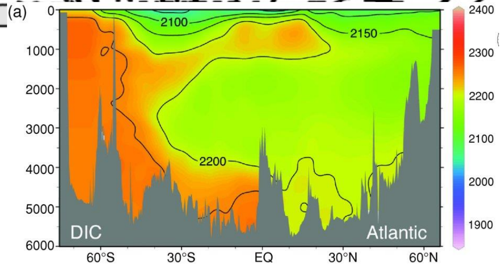
Local differences are same magnitude as long-term mean flux

Independent pCO_2 to assess product and model skill, 2000-2023

Hydrography

GLODAP

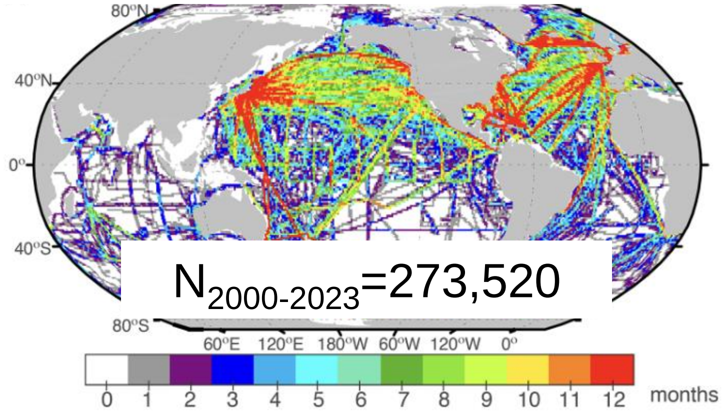




$N_{2000\text{-}2023}$=1,043

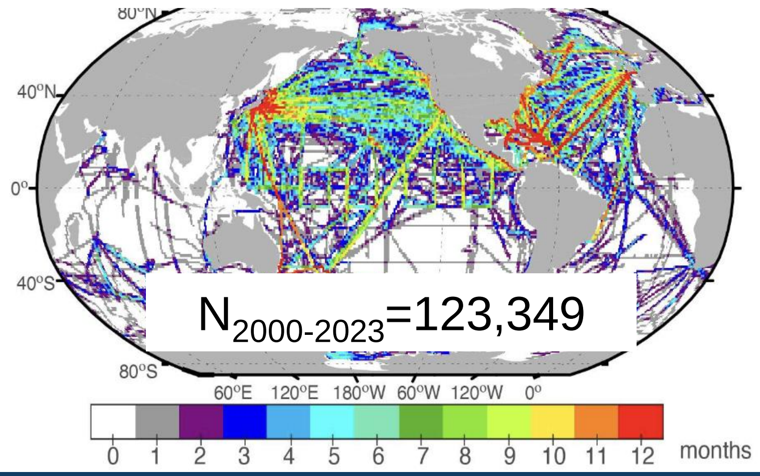
Surface pCO_2

SOCAT pCO_2,1980-present



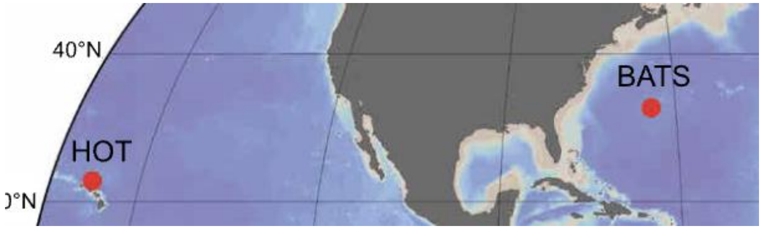
$N_{2000\text{-}2023}$=273,520

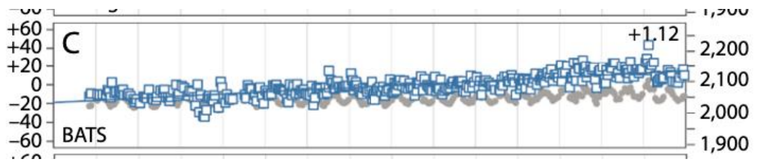
LDEO pCO_2,1980-2019



$N_{2000\text{-}2023}$=123,349

Timeseries





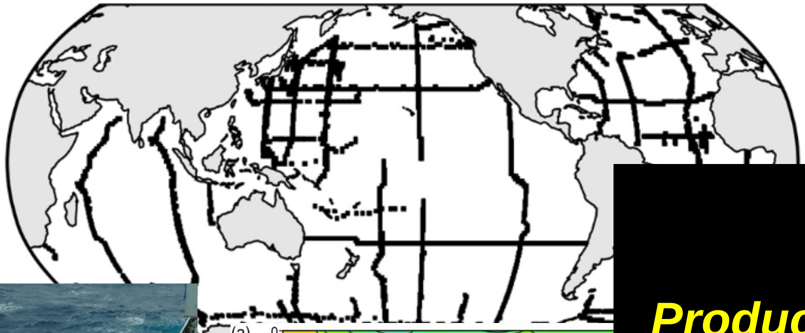
$N^{BATS}_{2000\text{-}2023}$=225

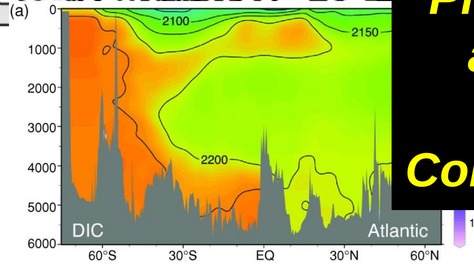
McKinley et al. in prep

Independent pCO_2 to assess product and model skill

Hydrography

GLODAP

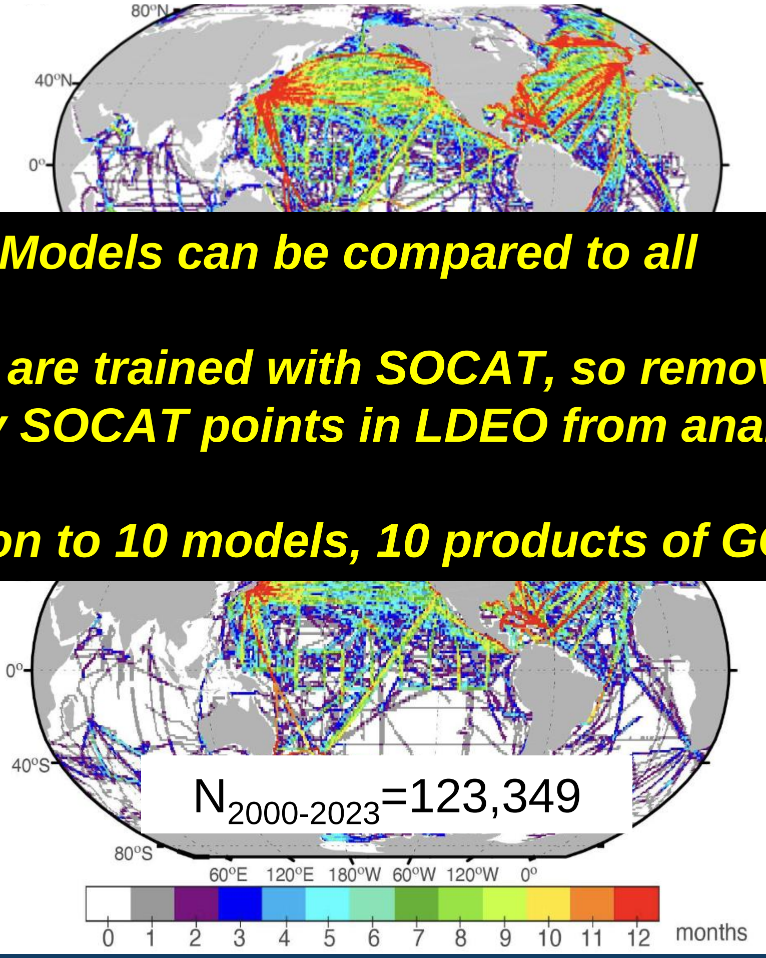




$N_{2000\text{-}2023}$=1,043

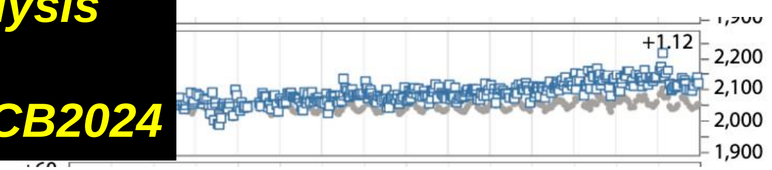
Surface pCO_2

SOCAT pCO_2,1980-present



$N_{2000\text{-}2023}$=123,349

Timeseries



$N^{BATS}_{2000\text{-}2023}$=225

Models can be compared to all

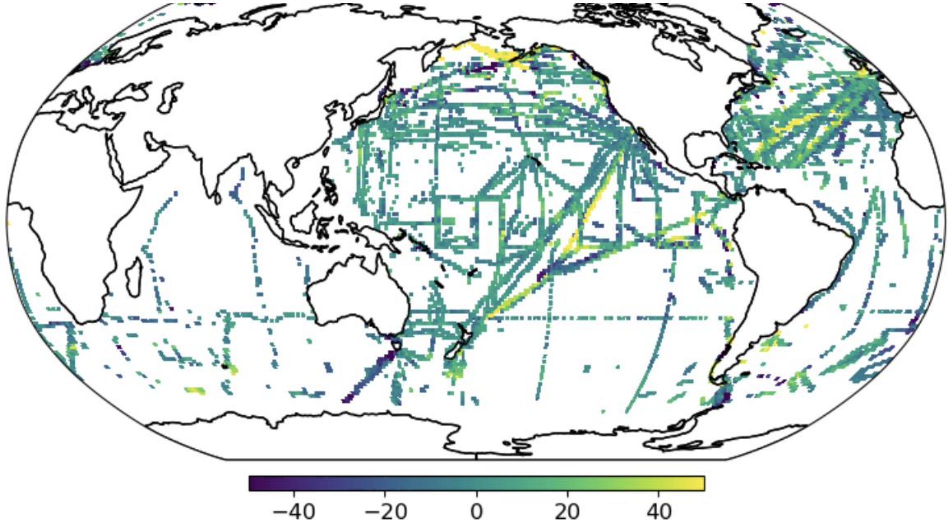
Products are trained with SOCAT, so remove this and any SOCAT points in LDEO from analysis

Comparison to 10 models, 10 products of GCB2024

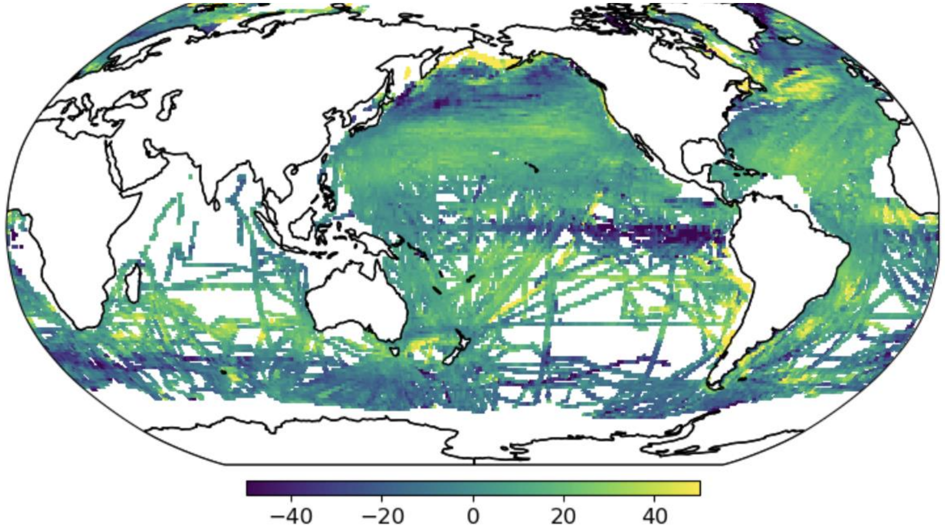
McKinley et al. in prep

Global pCO_2: Models vs. products, and vs. independent data

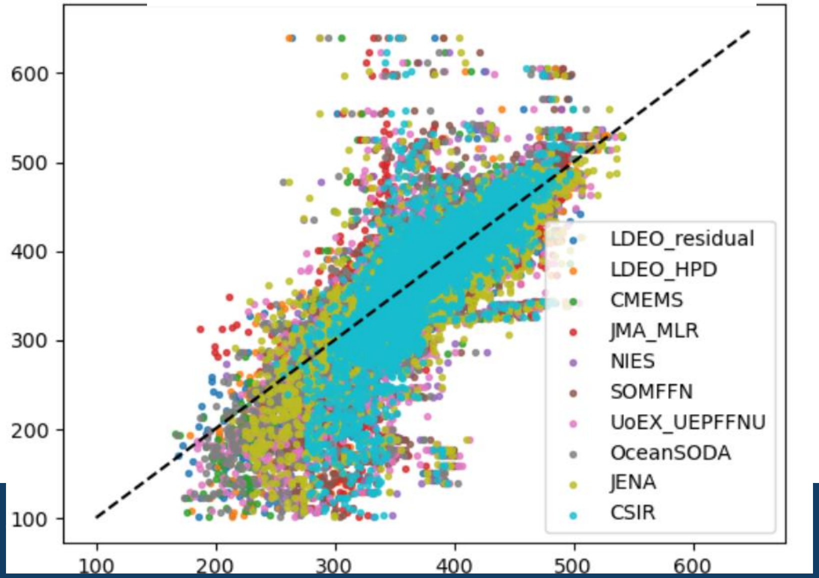
One Product - Data



One Model - Data



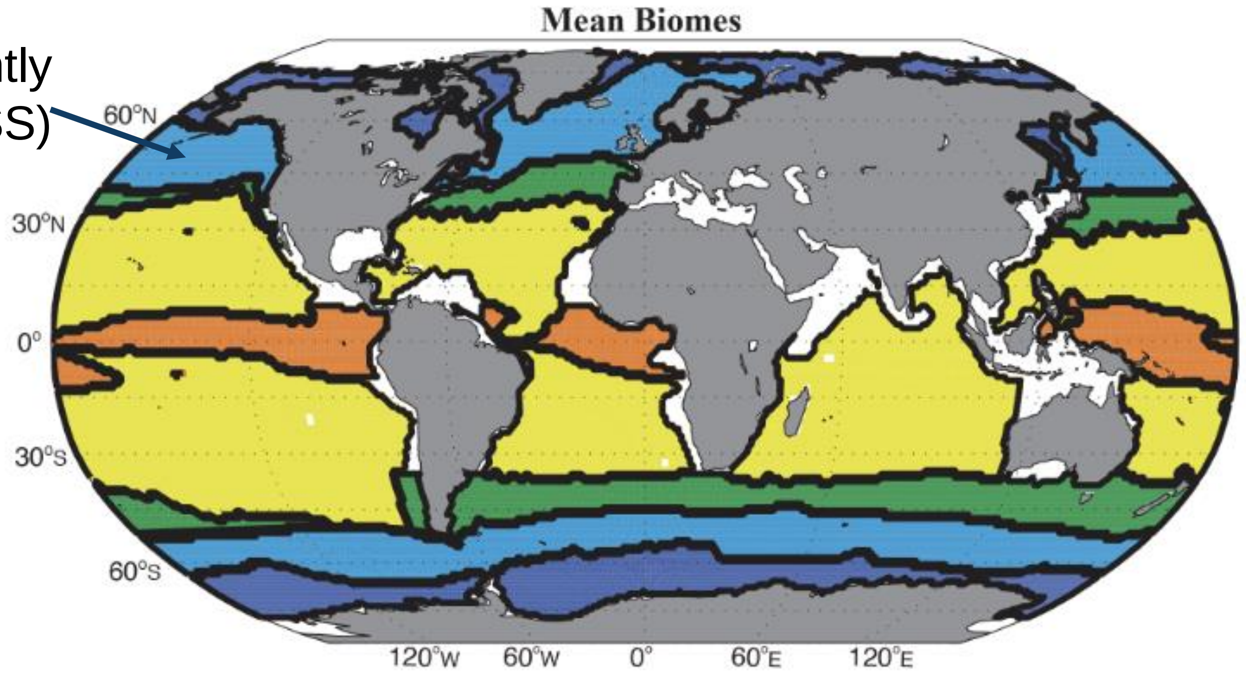
LDEO+GLODAP vs. Products



RMSE and correlation 2000-2023	10 Product mean	10 Model mean
LDEO + GLODAP (n=12,590)	**28.0 μatm** **r=0.80**	
SOCAT + GLODAP (n=285,256)		**32.9 μatm** **r=0.63**

Biomes are mechanistically similar; use these for temporal decomposition

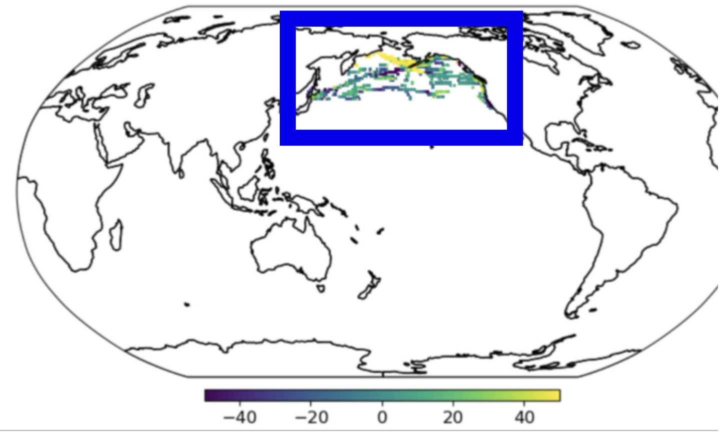
North Pacific Subpolar Permanently Stratified (NP SPSS)



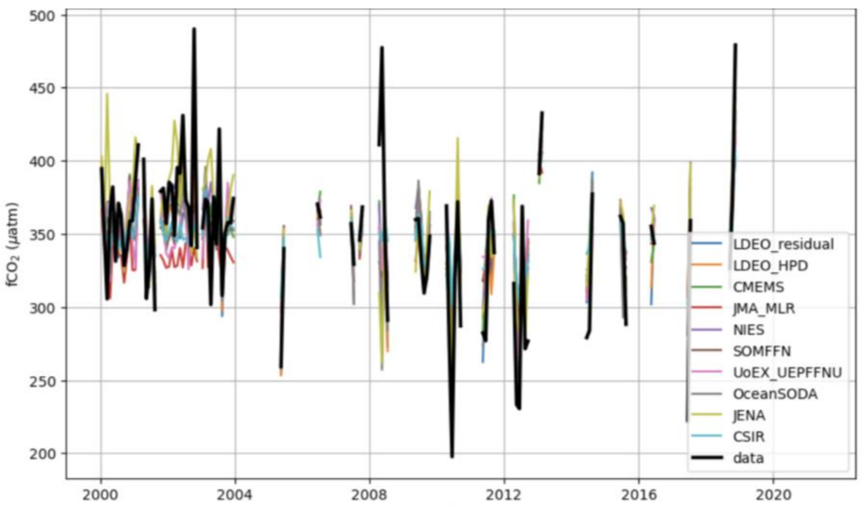
McKinley et al. in prep

Fay and McKinley, 2013

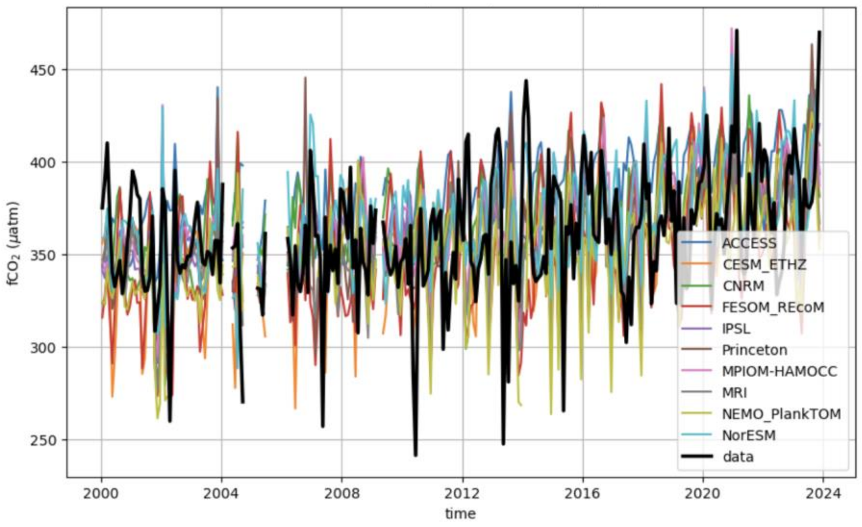
N. Pacific SPSS: Models and products vs. independent data



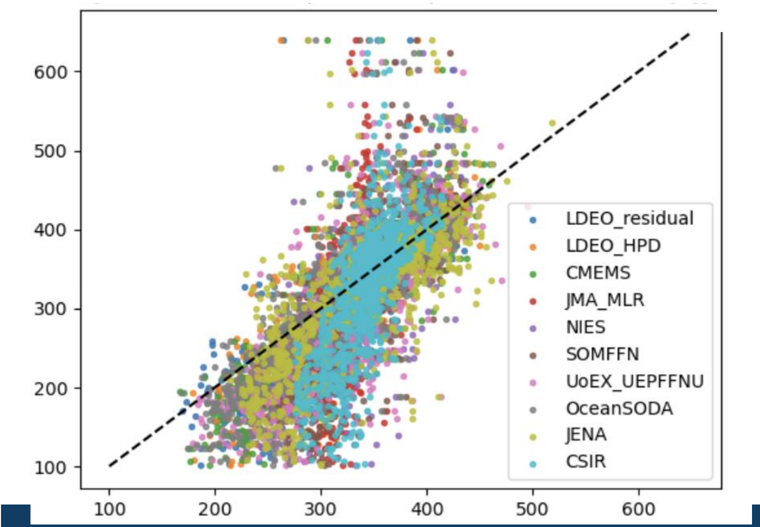
Products



Models



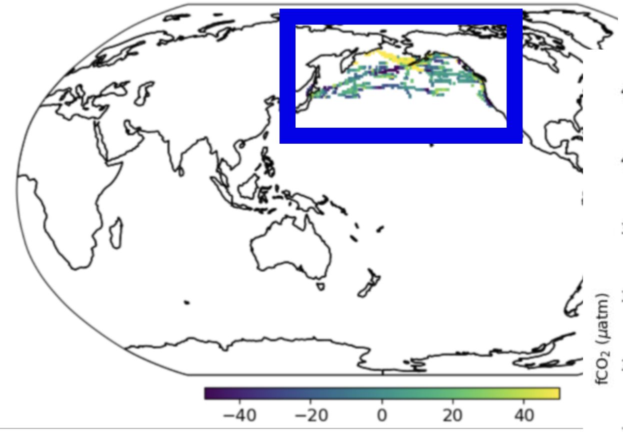
LDEO+GLODAP vs. Products



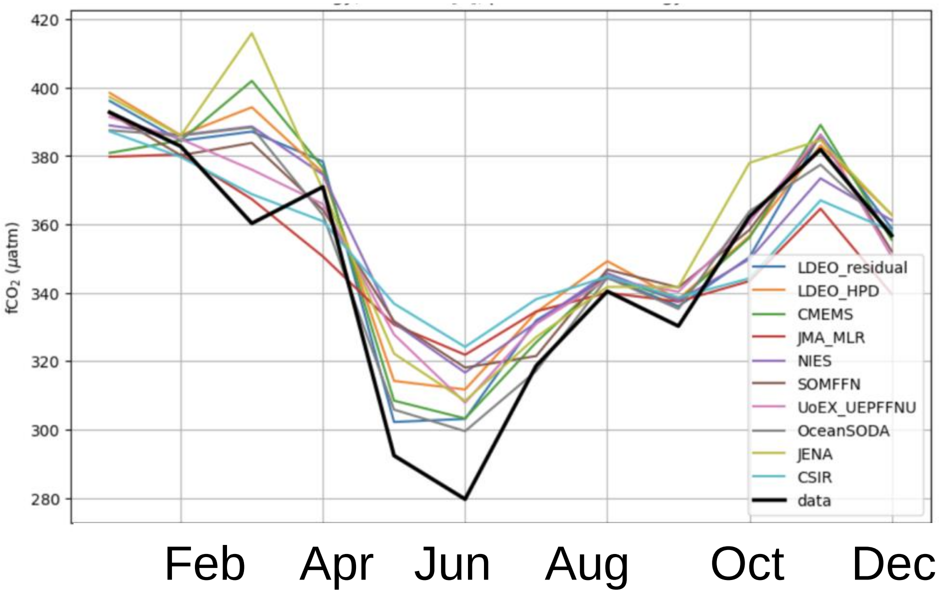
RMSE and correlation

2000-2023	10 Product mean	10 Model mean
LDEO + GLODAP (n=1,240)	**54.3µatm** **r=0.77**	
SOCAT + GLODAP (n=23,109)		**54.4µatm** **r=0.30**

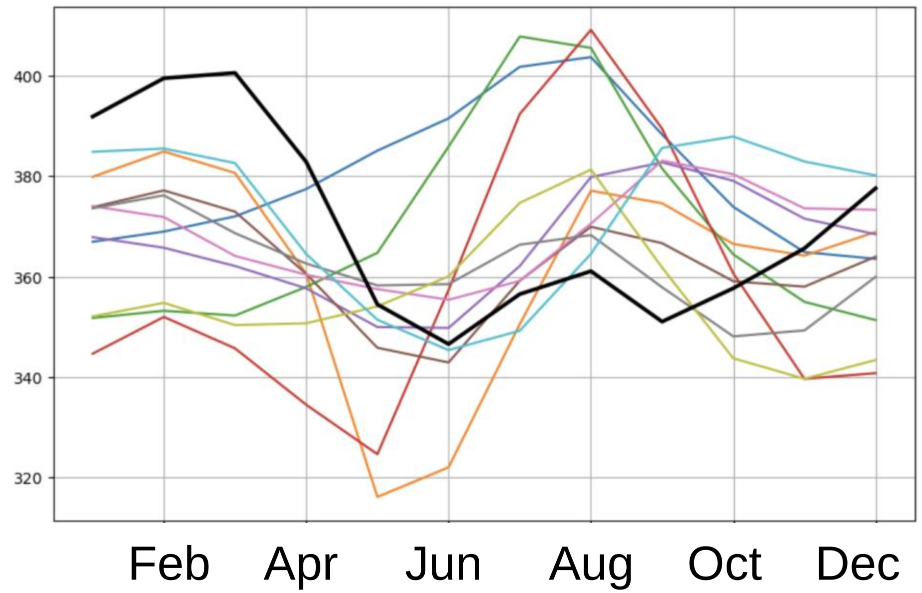
N. Pacific SPSS: Climatological



Products



Models

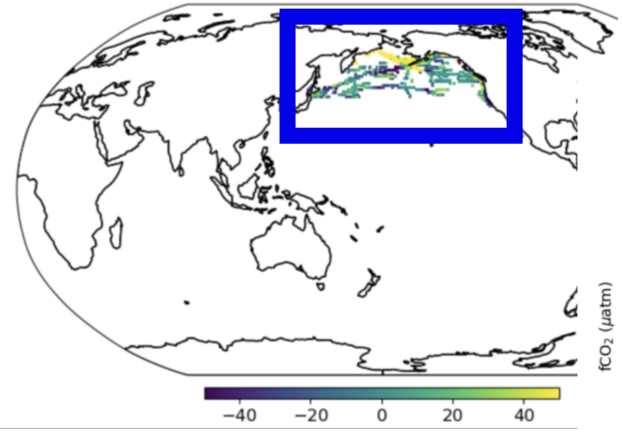


RMSE and correlation

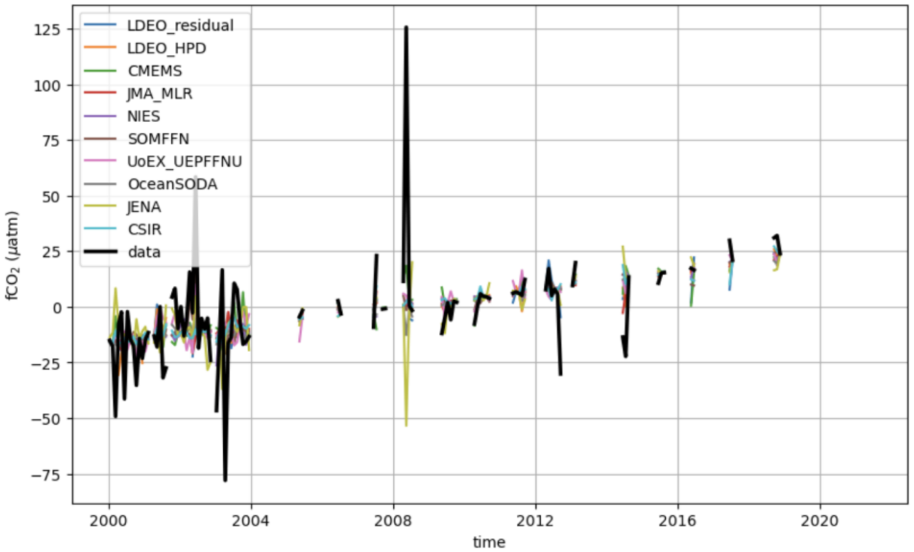
2000-2023	10 Product mean	10 Model mean
LDEO + GLODAP (n=1,240)	**50.7 µatm** **r=0.81**	
SOCAT + GLODAP (n=23,109)		**47.4 µatm** **r=0.28**

McKinley et al. in prep

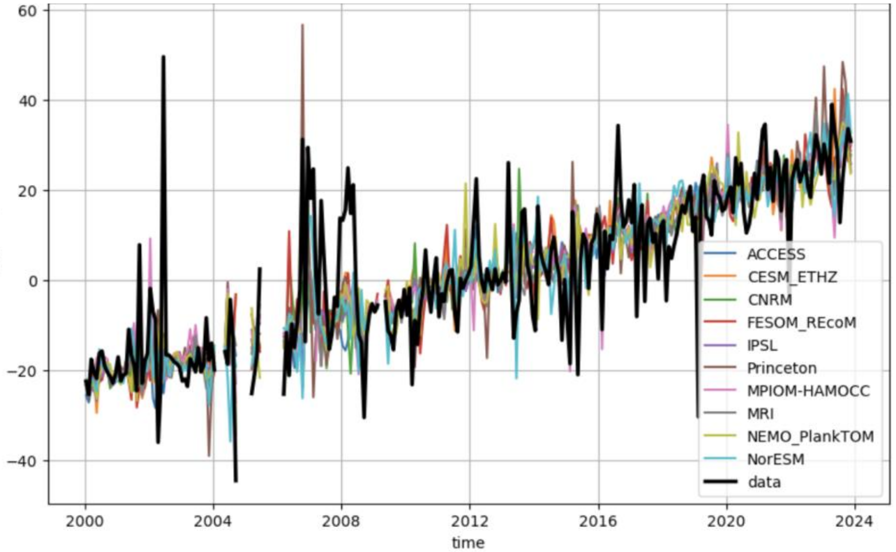
N. Pacific SPSS: Remove climatology



Products



Models



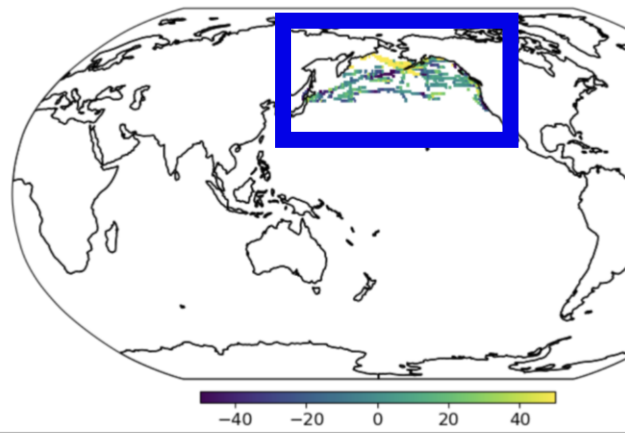
Product, Model, Datasets
Trends = 1.9-2.1 uatm/yr

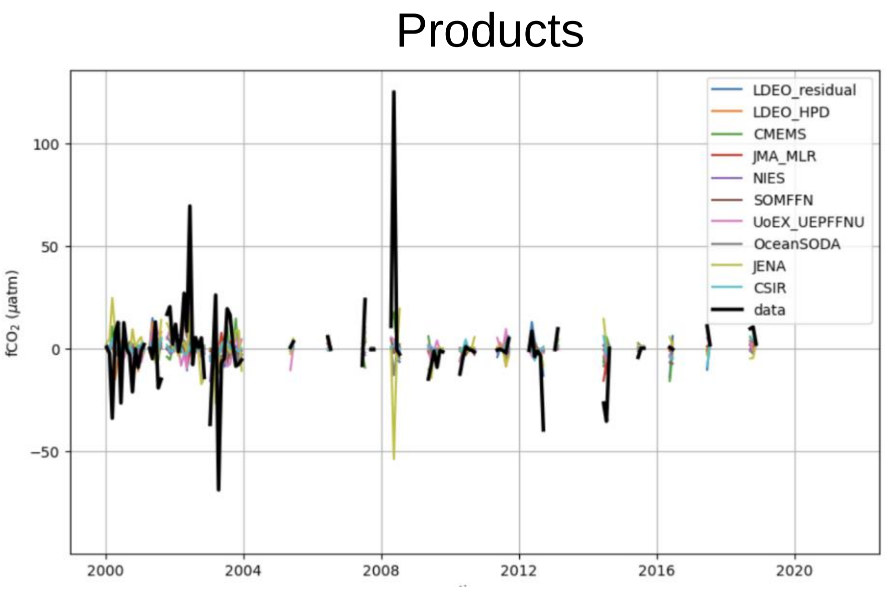
RMSE and correlation

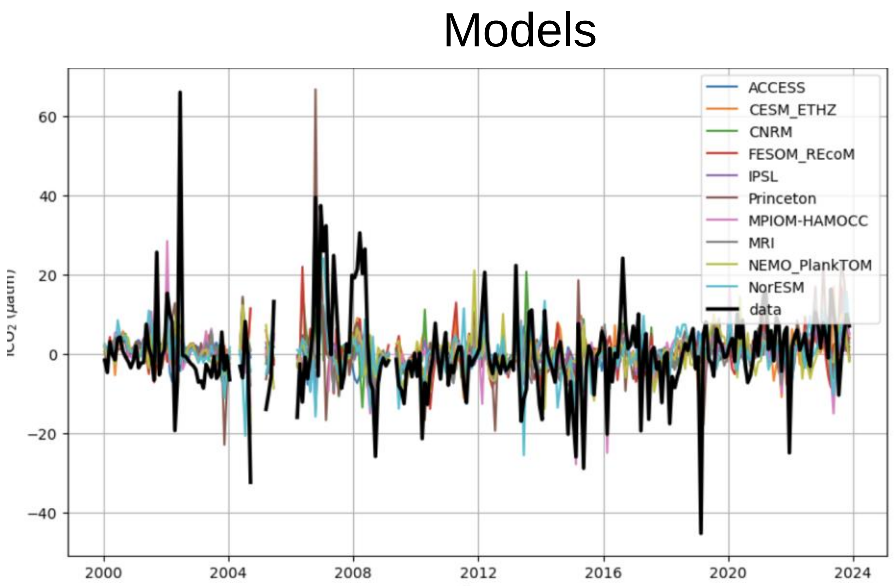
2000-2023	10 Product mean	10 Model mean
LDEO + GLODAP (n=1,240)	**18.9μatm** **r=0.50**	
SOCAT + GLODAP (n=23,109)		**23.2μatm** **r=0.54**

McKinley et al. in prep

N. Pacific SPSS: Remove trend







RMSE and correlation

2000-2023	10 Product mean	10 Model mean
LDEO + GLODAP (n=1,240)	**18.9µatm** **r=0.23**	
SOCAT + GLODAP (n=23,109)		**23.2µatm** **r=0.14**

McKinley et al. in prep

Take home messages

- Air-sea CO_2 flux mean and seasonality can be reconstructed from sparse pCO_2 data, but longer-timescale variability remains uncertain

- Hindcast and Earth System Models have significant mean-state biases

- Correcting models with data offers improved skill and 60+ years of reconstructed monthly surface ocean pCO_2

- Current models and data products have local uncertainties orders of magnitude larger than expected mCDR impacts

Thank you!

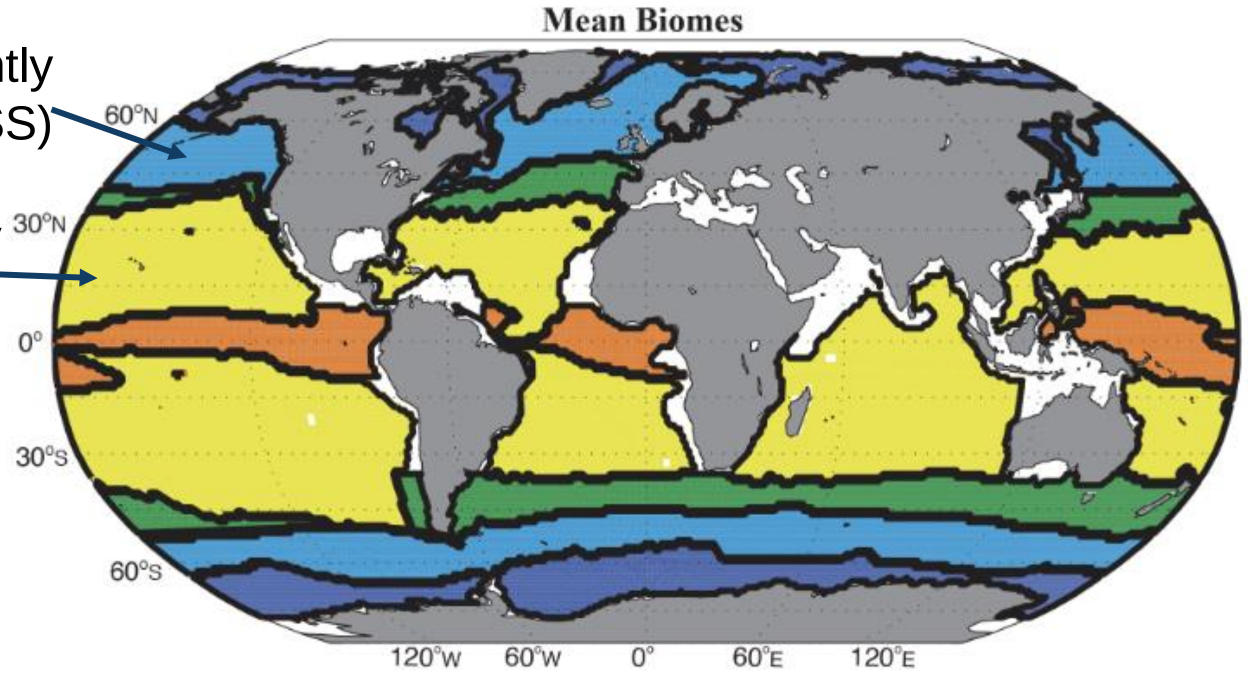
YEARS OF DISCOVERY
75
LAMONT-DOHERTY
EARTH OBSERVATORY
COLUMBIA UNIVERSITY

Viviana Acquaviva (professor, CUNY), Ce Bian (postdoc), Galen McKinley
Thea Heimdal (scientist), Abby Shaum (PhD student), Amanda Fay (senior scientist)

Biomes are mechanistically similar; use these for temporal decomposition

North Pacific Subpolar Permanently Stratified (NP SPSS)

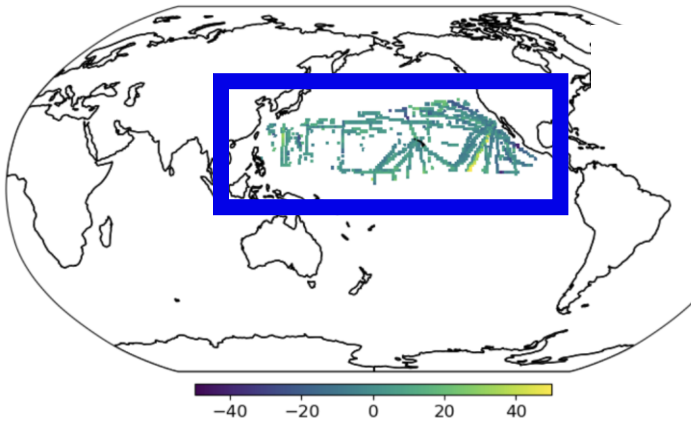
North Pacific Subtropical Permanently Stratified (NP STPS)

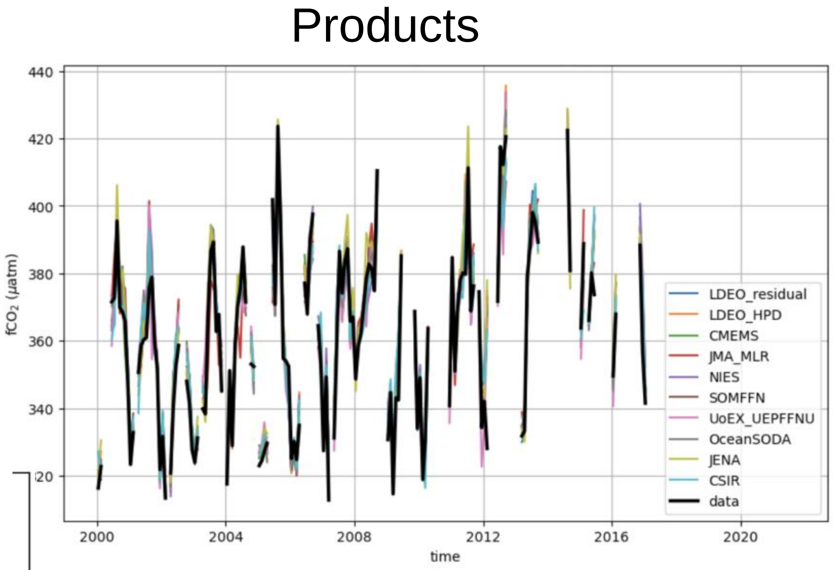


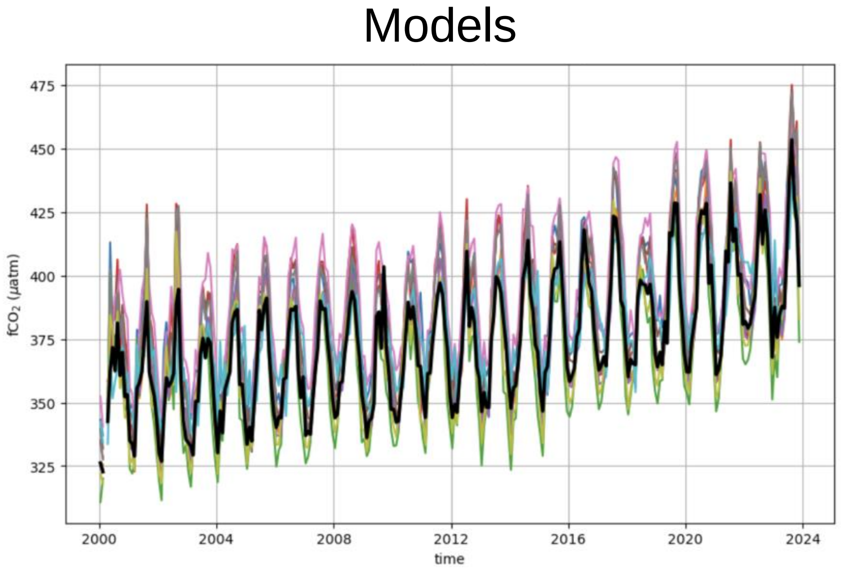
McKinley et al. in prep

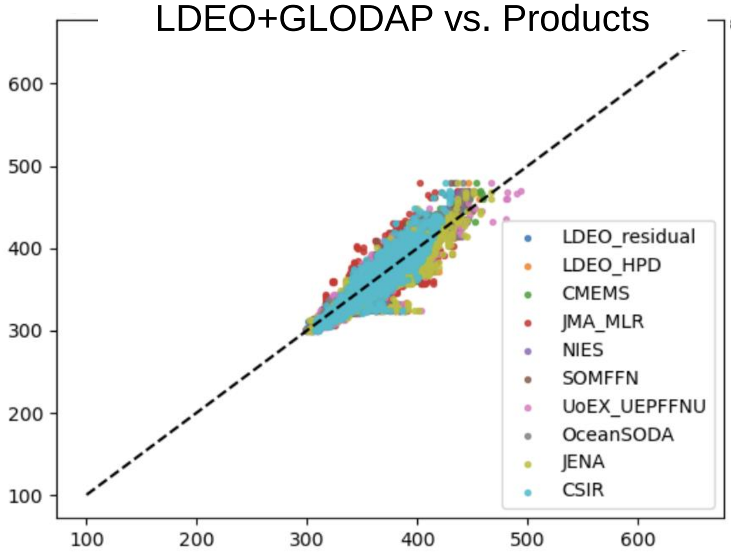
Fay and McKinley, 2013

N. Pacific STPS: Models and products vs. independent data



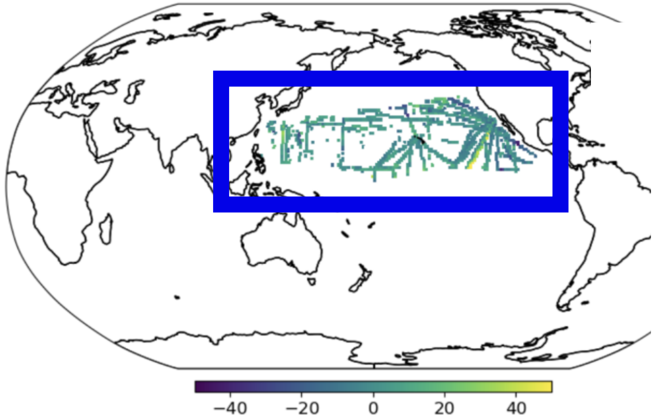




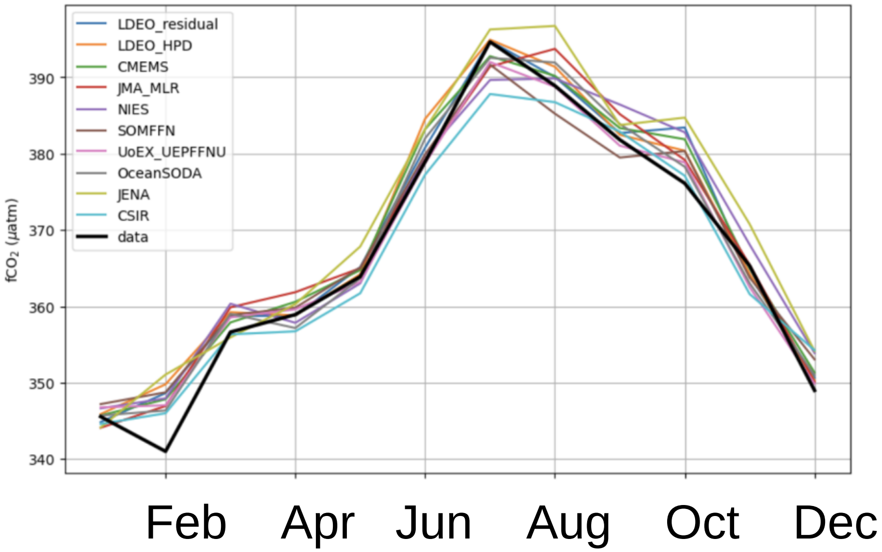


***RMSE* and correlation** 2000-2023	10 Product mean	10 Model mean
LDEO + GLODAP (n=1,931)	**12.3μatm** **r=0.91**	
SOCAT + GLODAP (n=41,820)		**18.3μatm** **r=0.88**

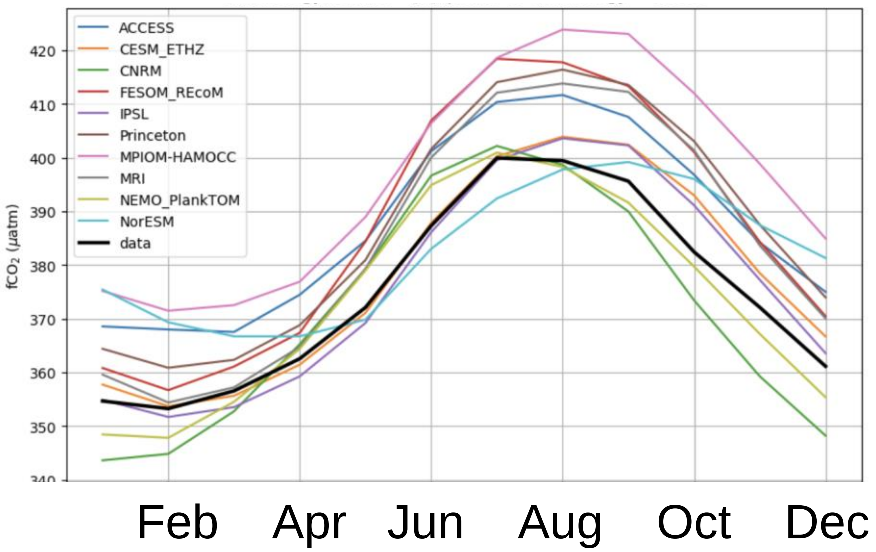
N. Pacific STPS: Climatological



Products



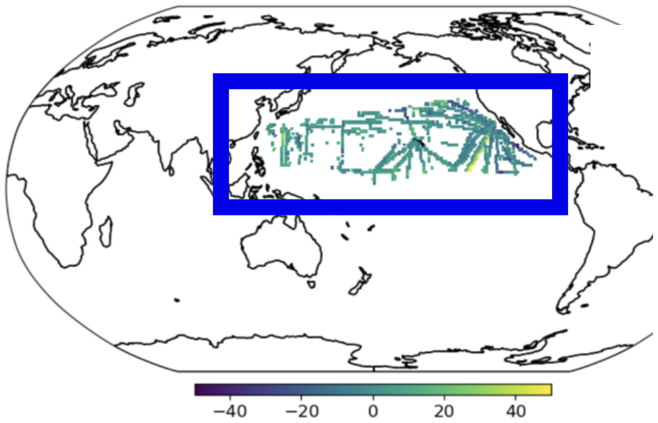
Models



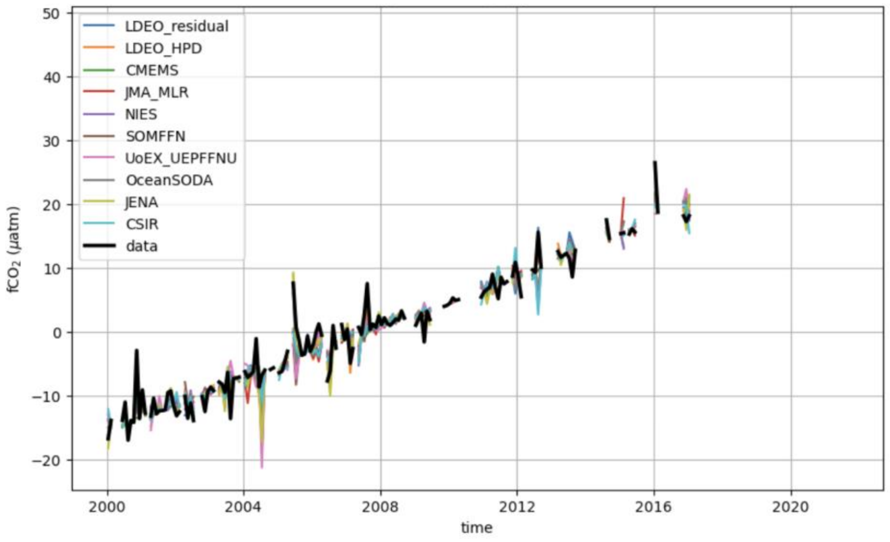
RMSE and correlation 2000-2023	10 Product mean	10 Model mean
LDEO + GLODAP (n=1,931)	**12.2μatm** **r=0.89**	
SOCAT + GLODAP (n=41,820)		**17.7μatm** **r=0.85**

McKinley et al. in prep

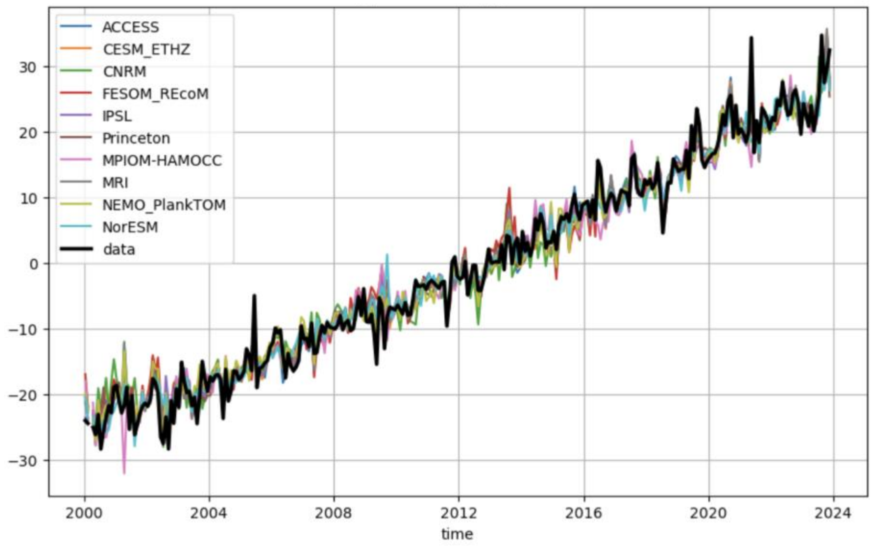
N. Pacific STPS: Remove climatology



Products



Models

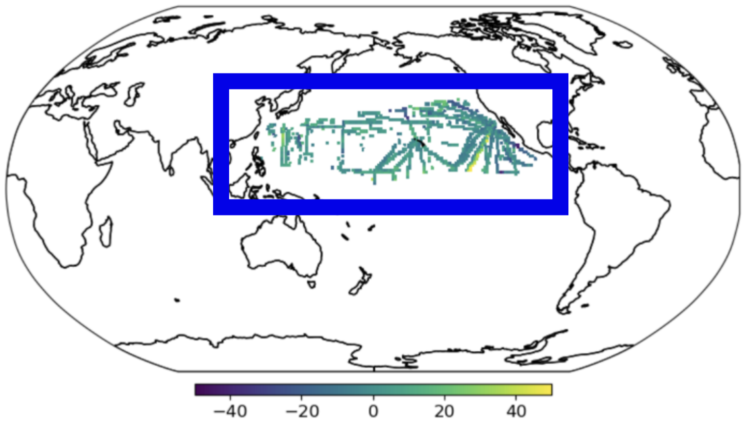


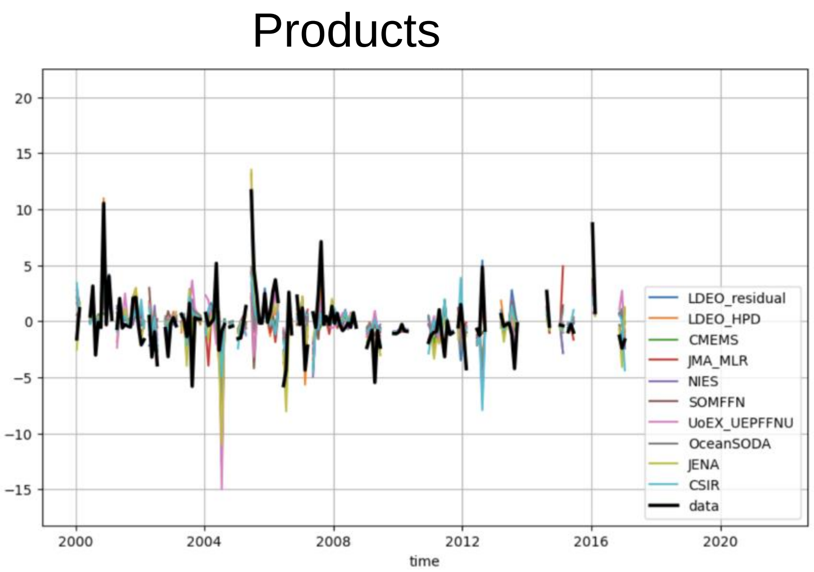
***RMSE* and correlation** 2000-2023	10 Product mean	10 Model mean
LDEO + GLODAP (n=1,931)	**3.7μatm** **r=0.92**	
SOCAT + GLODAP (n=41,820)		**7.2μatm** **r=0.89**

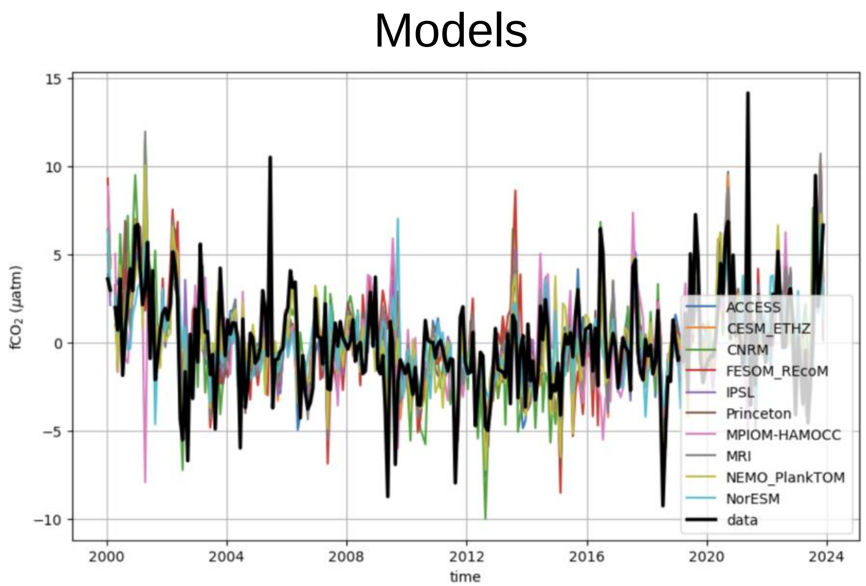
Product, Model, Datasets
Trend = 2.1-2.2 uatm/yr

McKinley et al. in prep

N. Pacific STPS: Remove trend







RMSE and correlation

2000-2023	10 Product mean	10 Model mean
LDEO + GLODAP (n=1,931)	**3.7µatm** **r=0.40**	
SOCAT + GLODAP (n=41,820)		**7.2µatm** **r=0.47**

McKinley et al. in prep