



cmcc
Centro Euro-Mediterraneo
sui Cambiamenti Climatici

Annalisa Bracco

Qi Zhang and Taka Ito
Georgia Institute of Technology

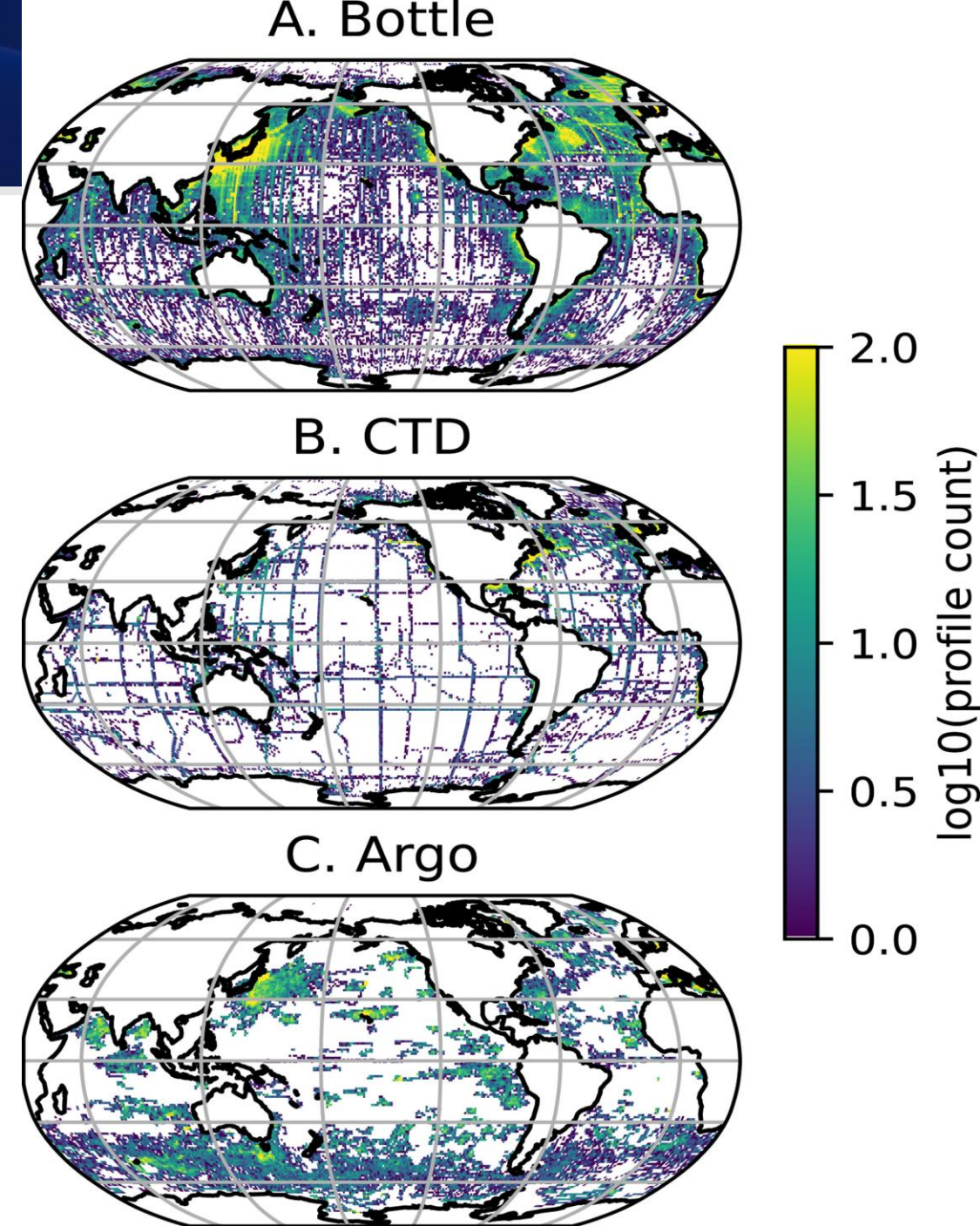
A physical informed emulator for ocean oxygen

www.cmcc.it

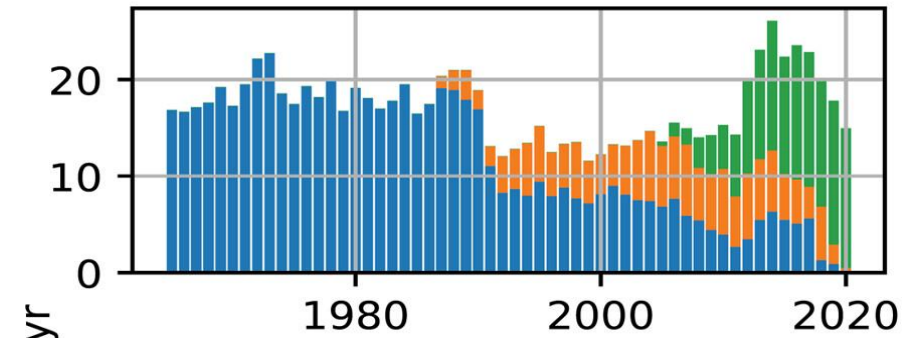


A cheap and fast emulator for **ocean oxygen concentrations**
aiming at global and decadal scales
Based on physical variables only (plus lon, lat and time)

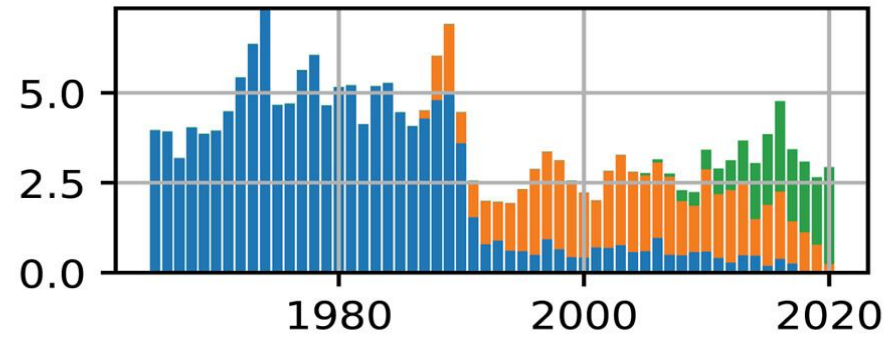
Sampling density (A,B) Logarithm (base 10) of the cumulative profile count within 154 each 1°x1° longitude-latitude cell for oxygen (O₂) based on the World Ocean Database 2018 155 (Boyer et al., 2018) downloaded in October 2023. The color saturates at 2 (more than 100 156 profiles) per cell since 1965. (C) The cumulative profile count for the BGC-Argo O₂ data.



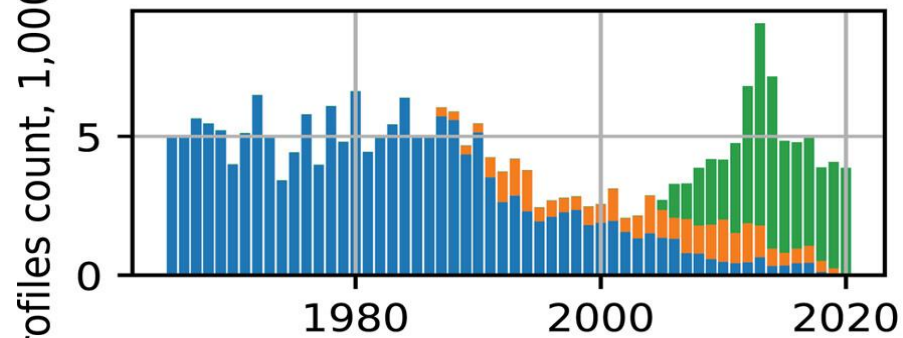
A. Global



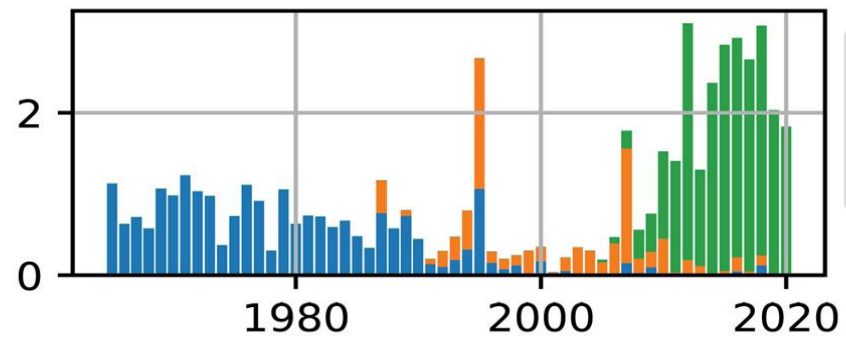
B. Atlantic



C. Pacific

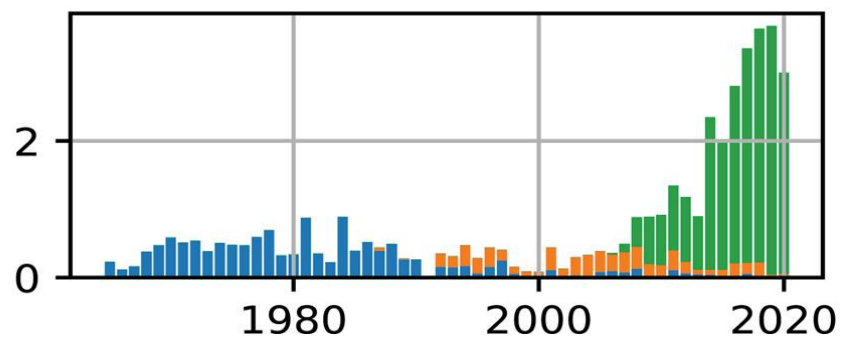


D. Indian

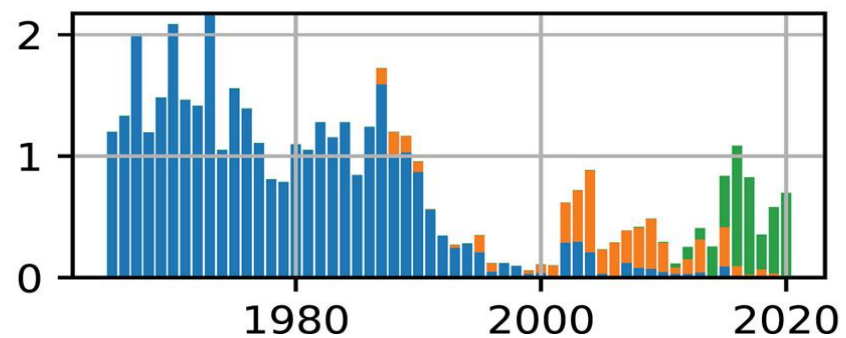


profiles count, 1,000s/yr

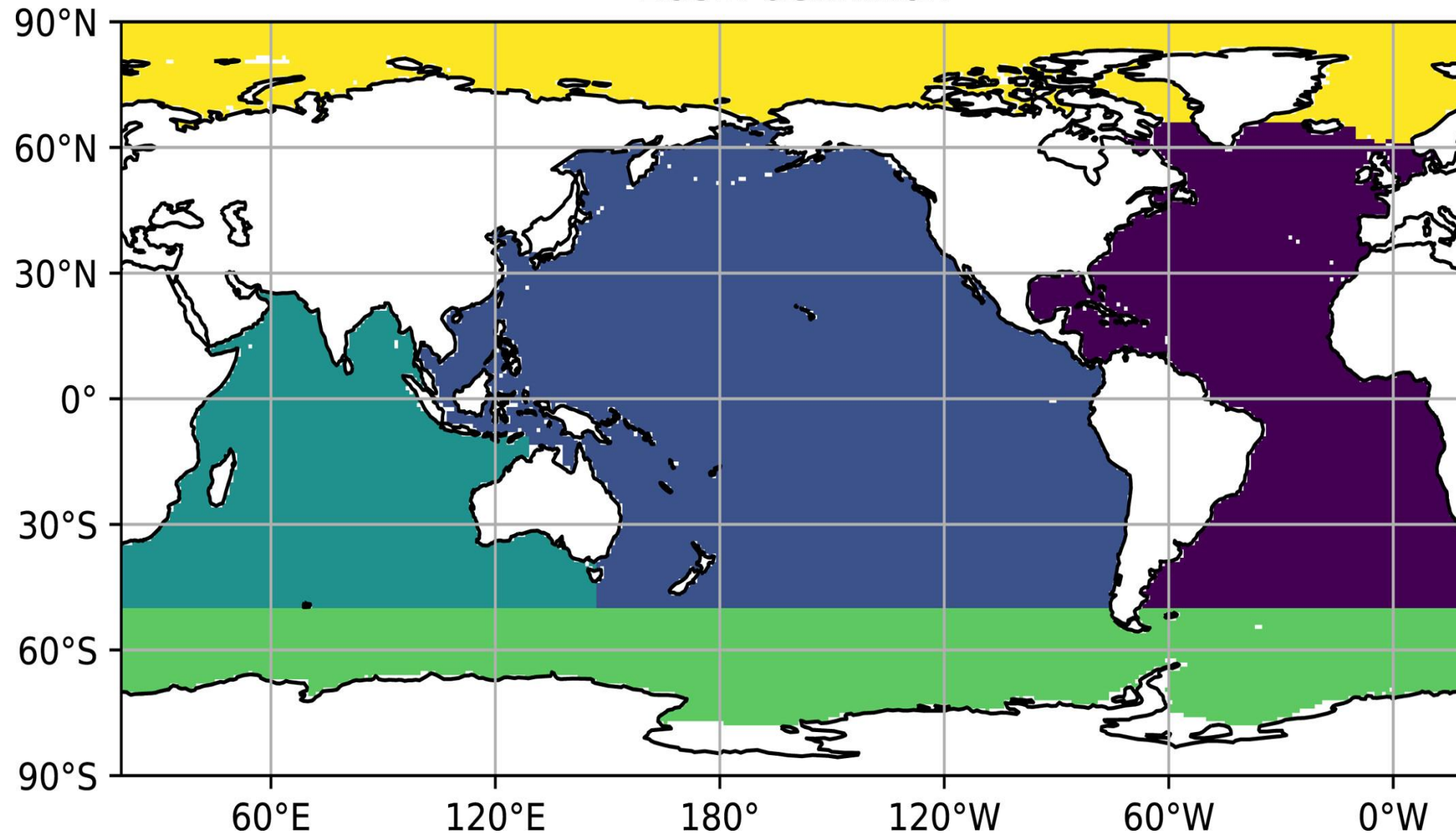
E. Southern Ocean



E. Arctic Ocean

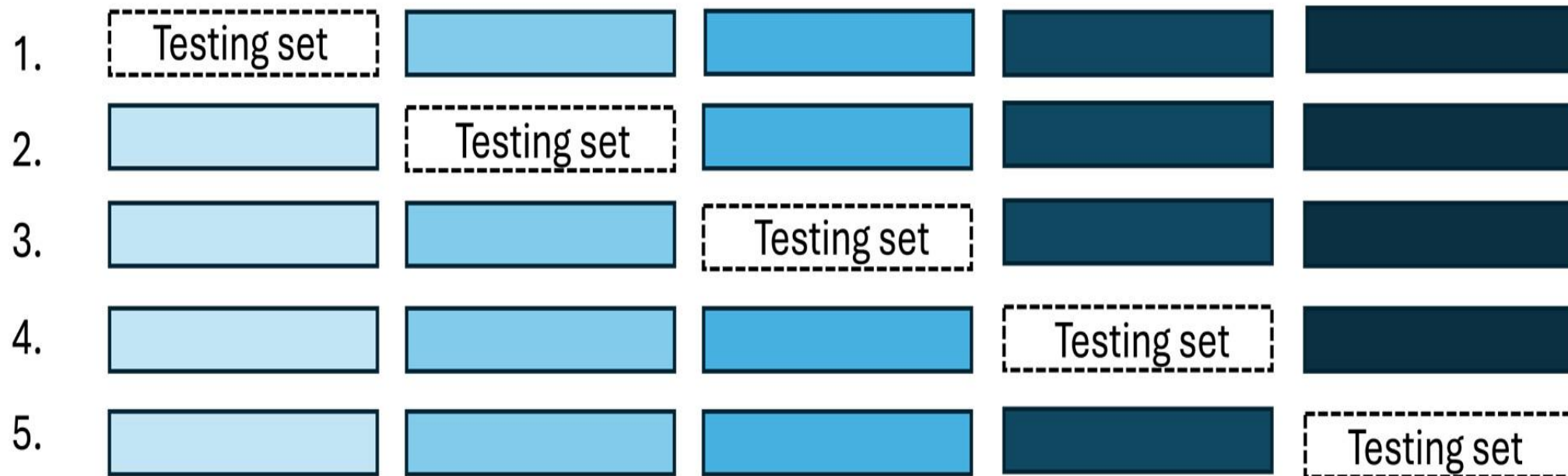


Basin definition

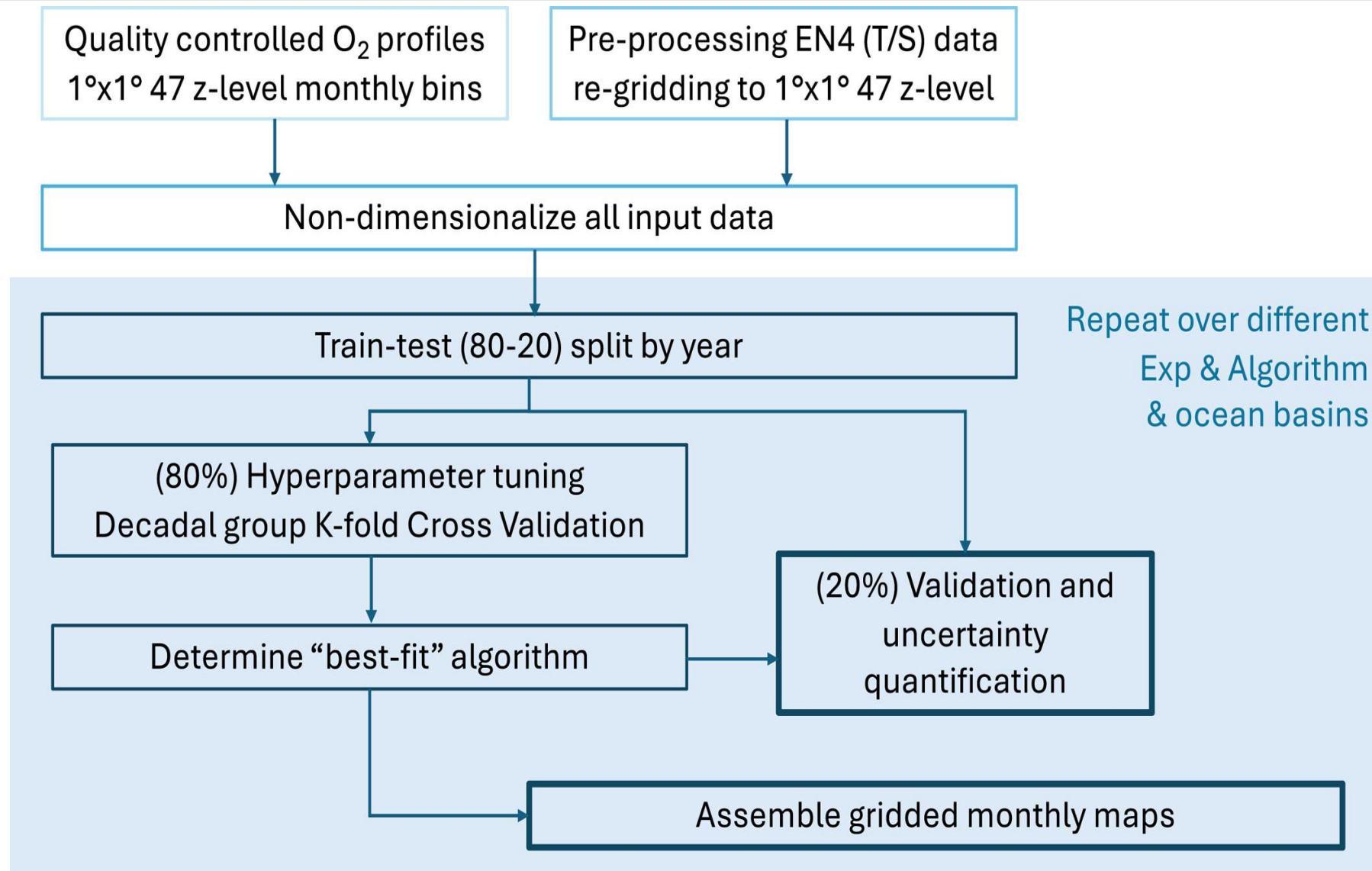


Whole training dataset

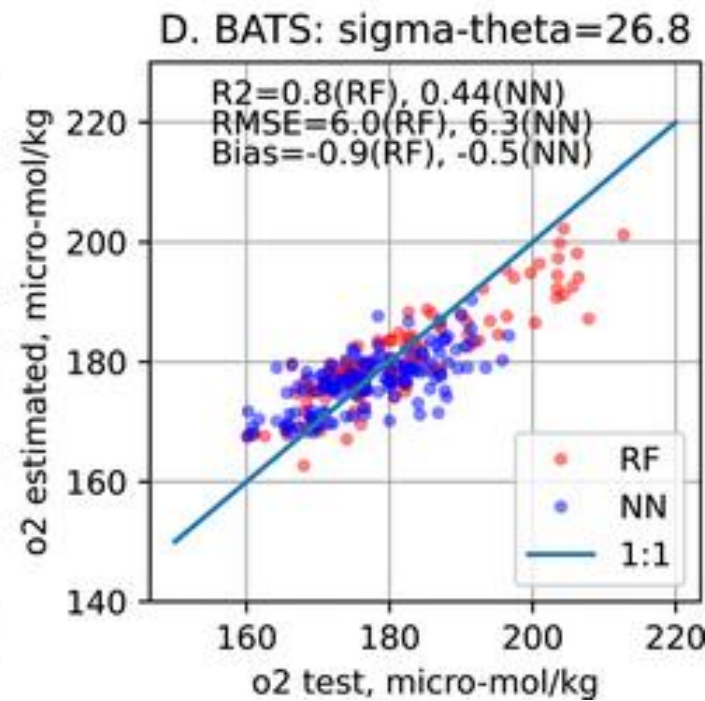
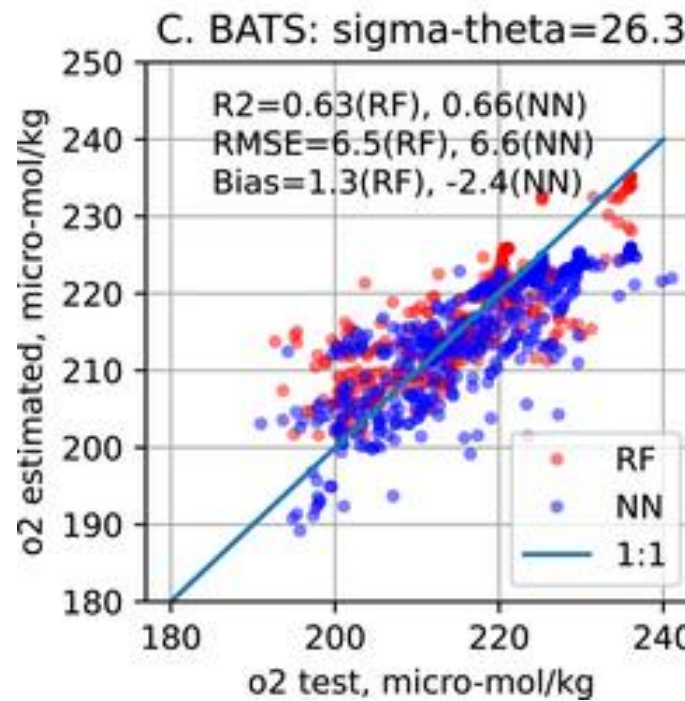
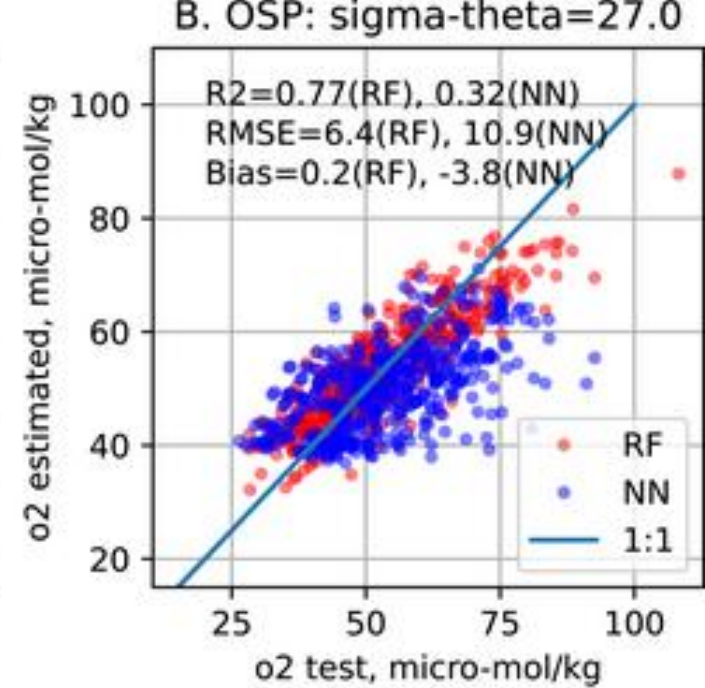
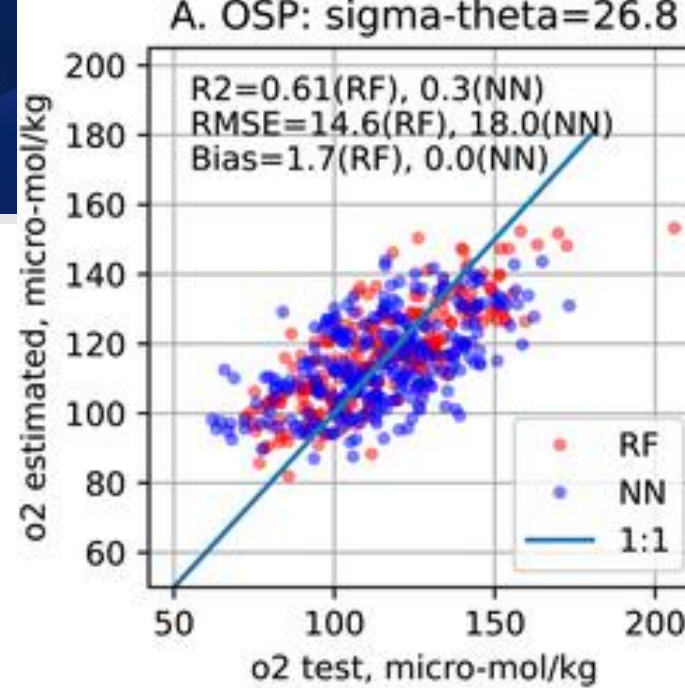
1965 1975 1985 1995 2005 2020



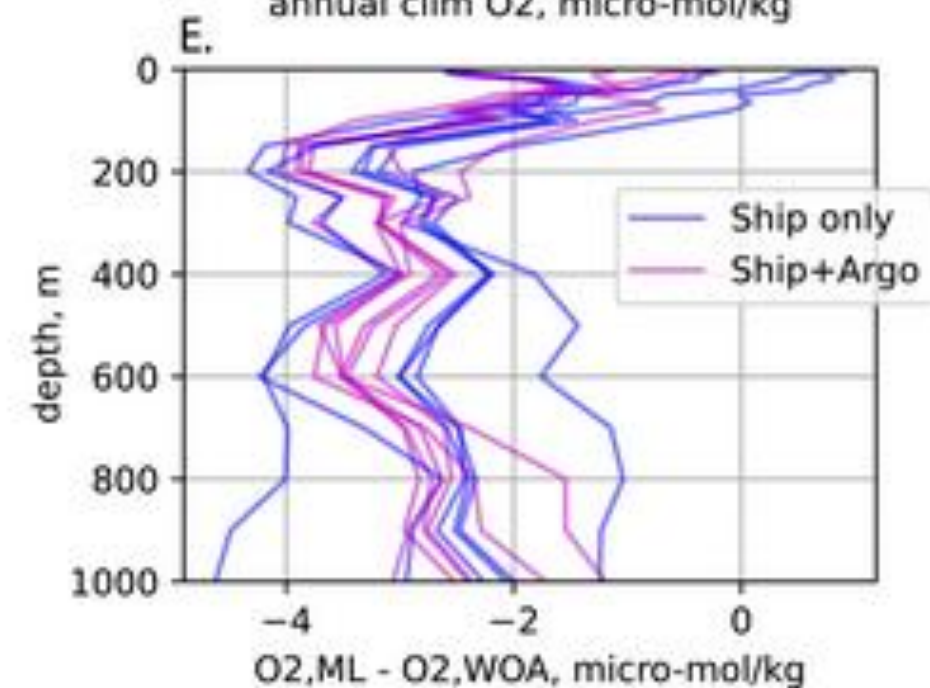
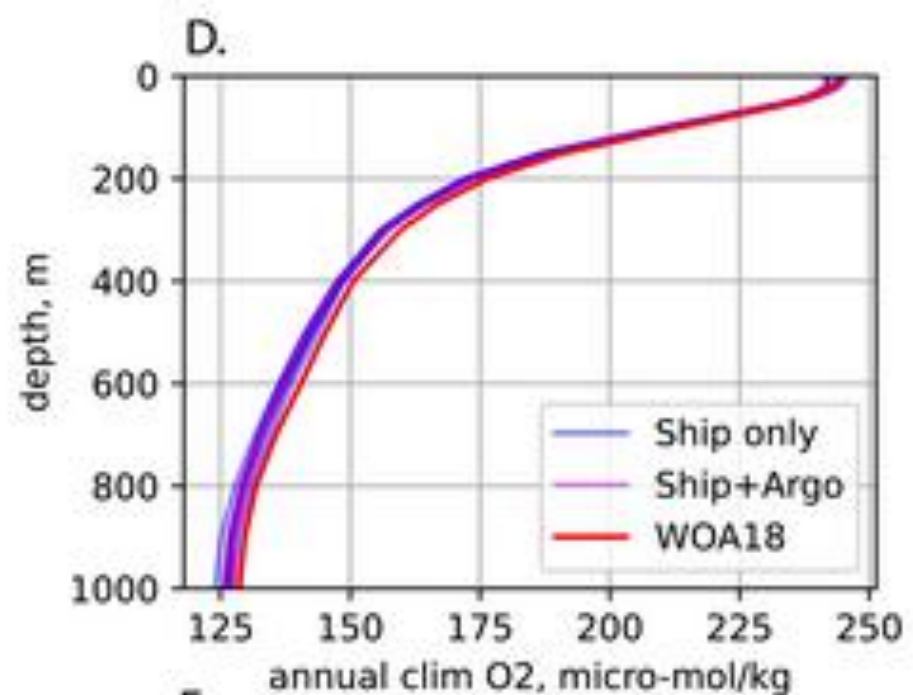
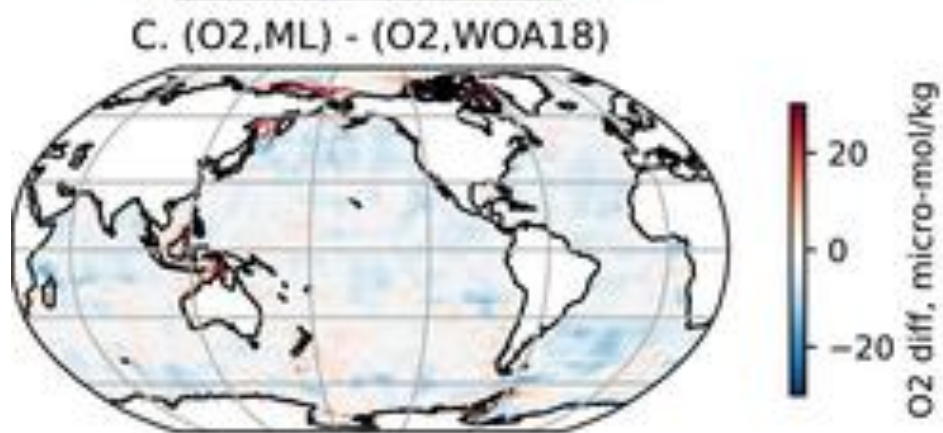
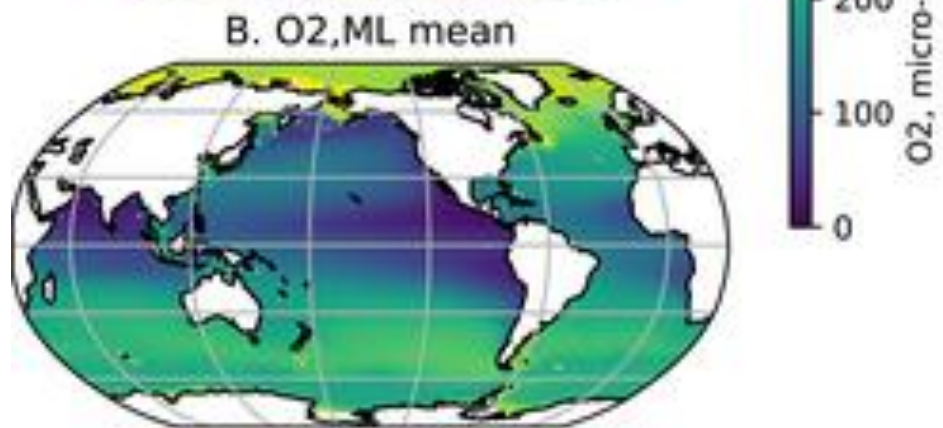
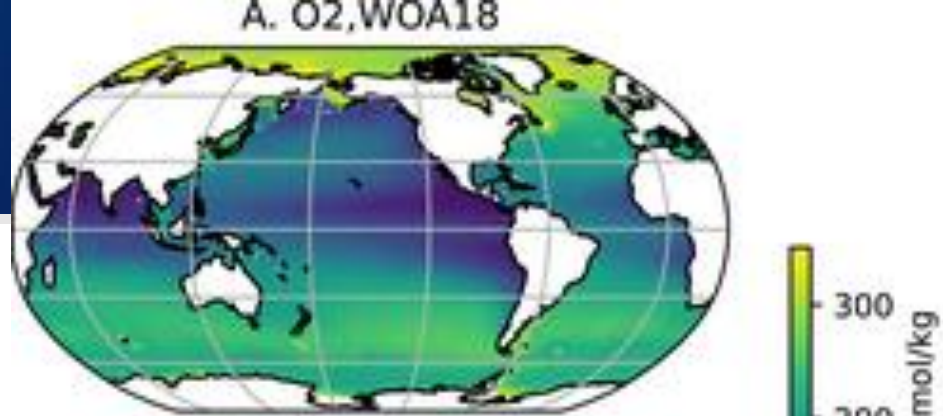
Decadal
Group
K-folds



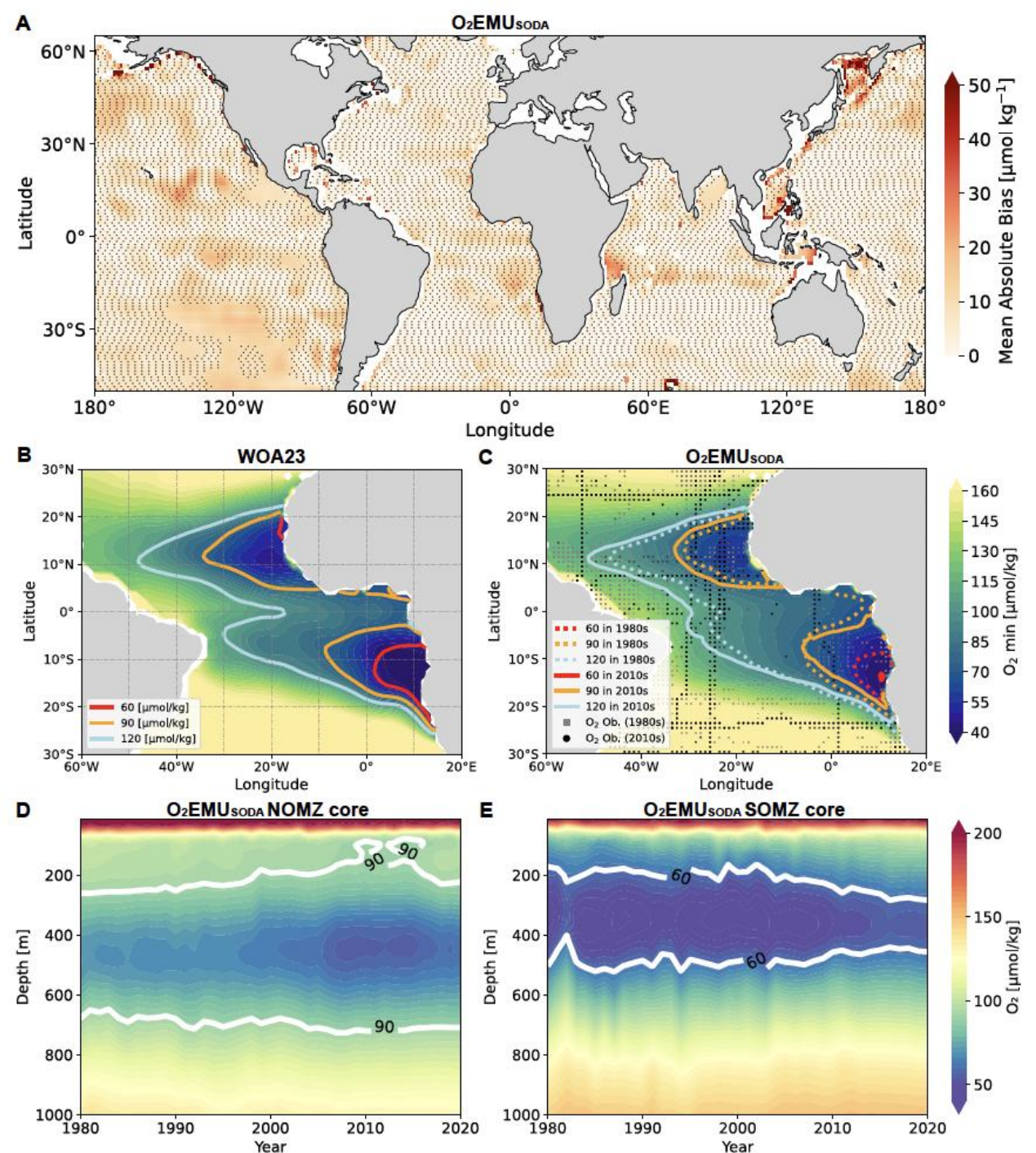
Validation of O₂ variability at two ocean time series stations (a, b) Ocean Station Papa and (c, d) Bermuda Atlantic Ocean Time Series. Blue dots are for the Neural Network algorithm, and red dots are for the Random Forest. The blue solid lines are 1:1 line.

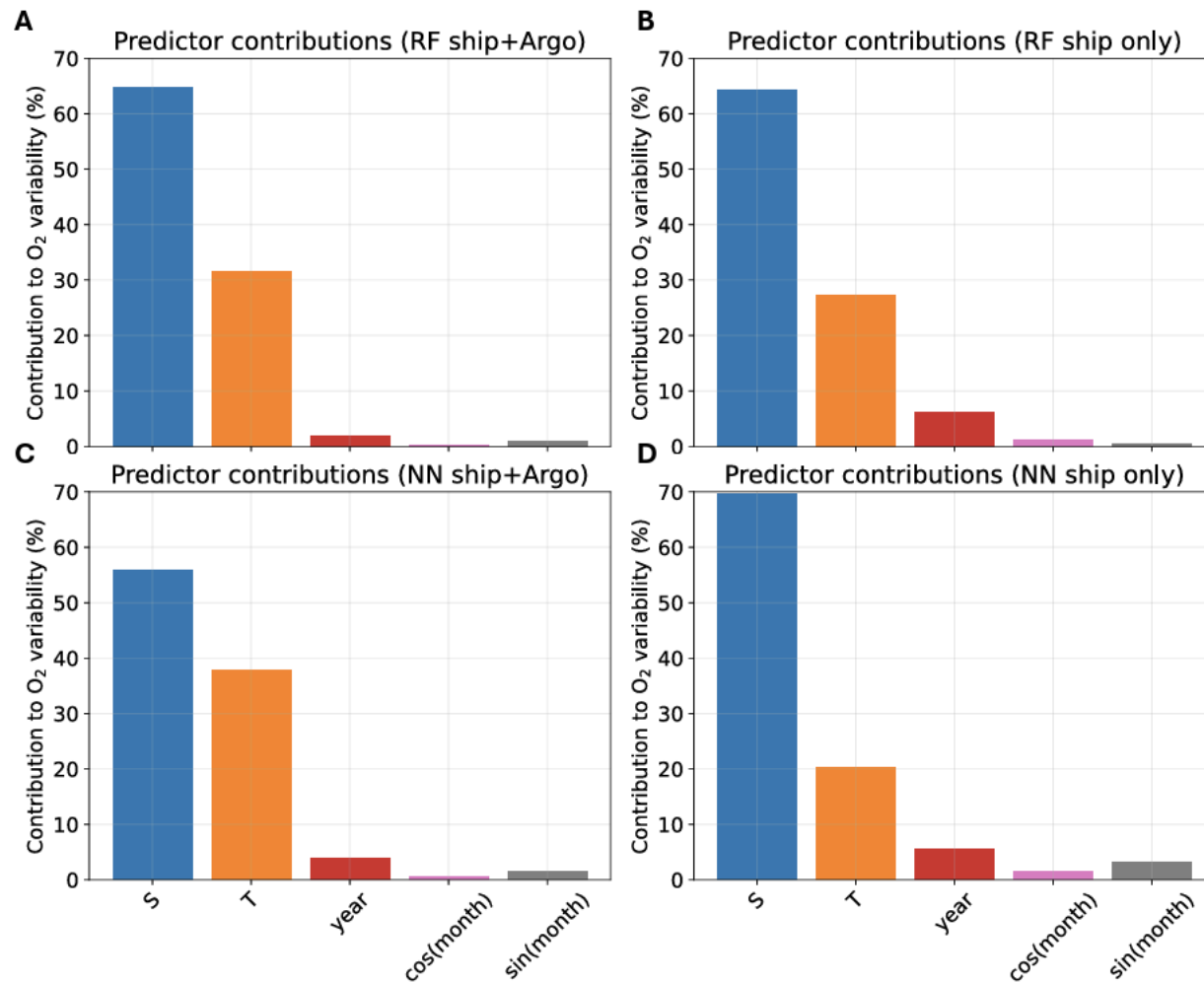


Climatology



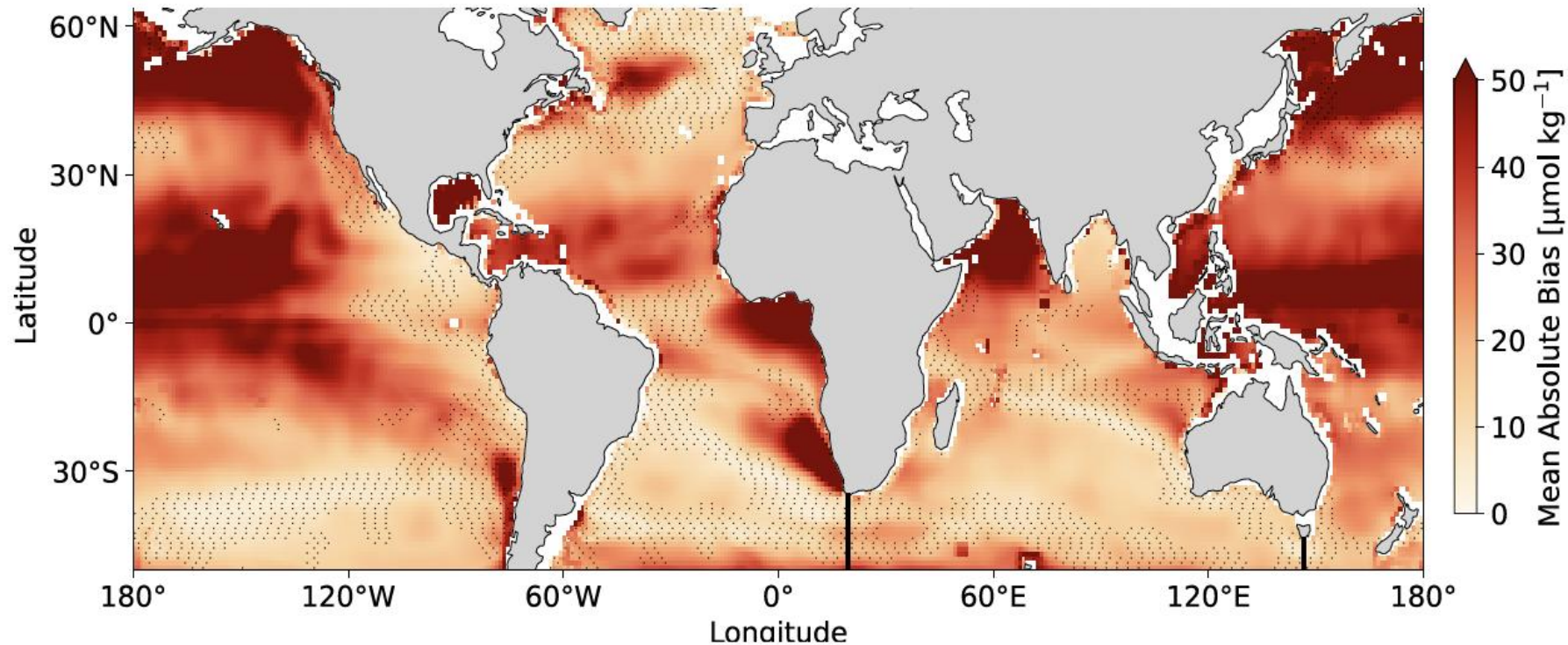
Climatology Atlantic Focus based on SODA reanalysis 1980-2020





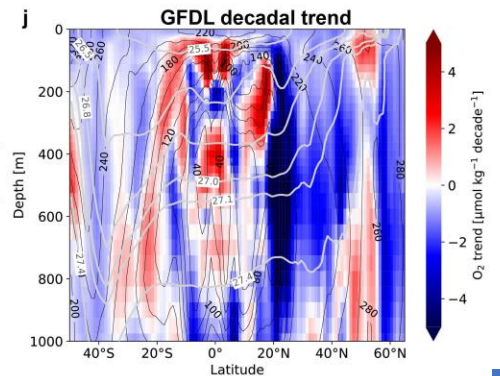
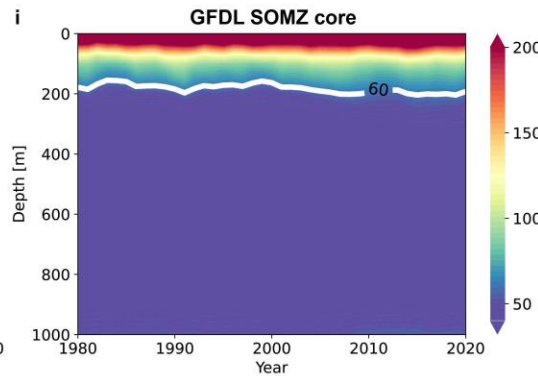
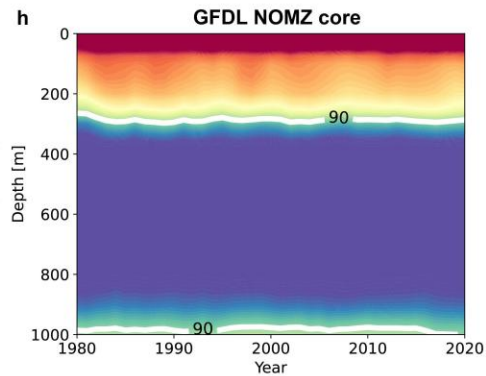
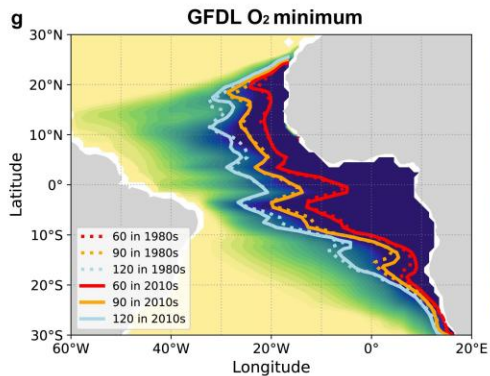
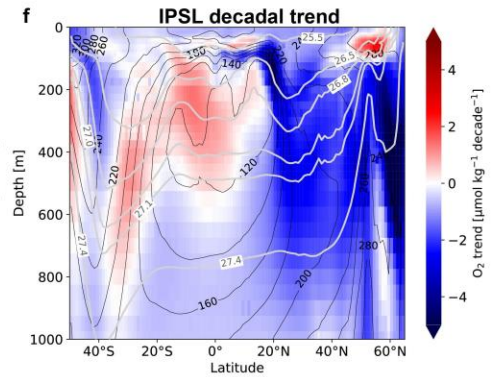
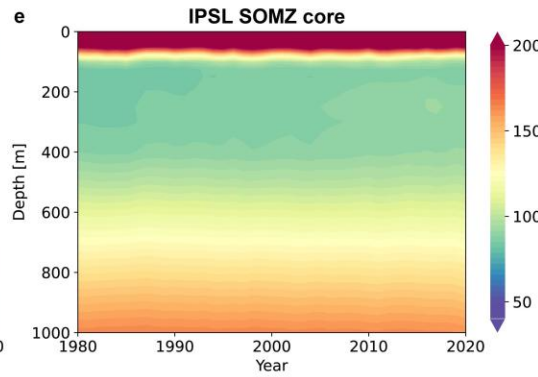
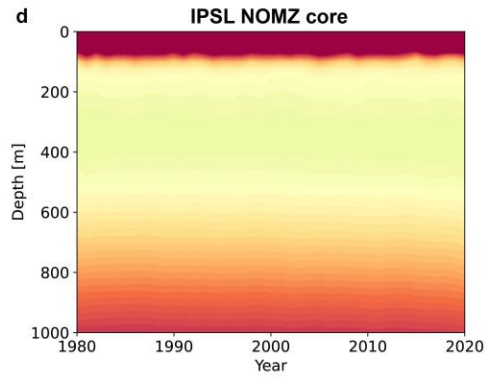
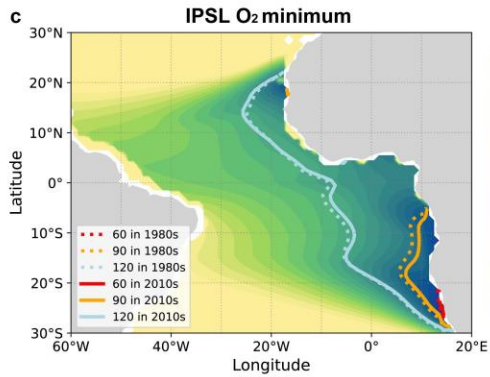
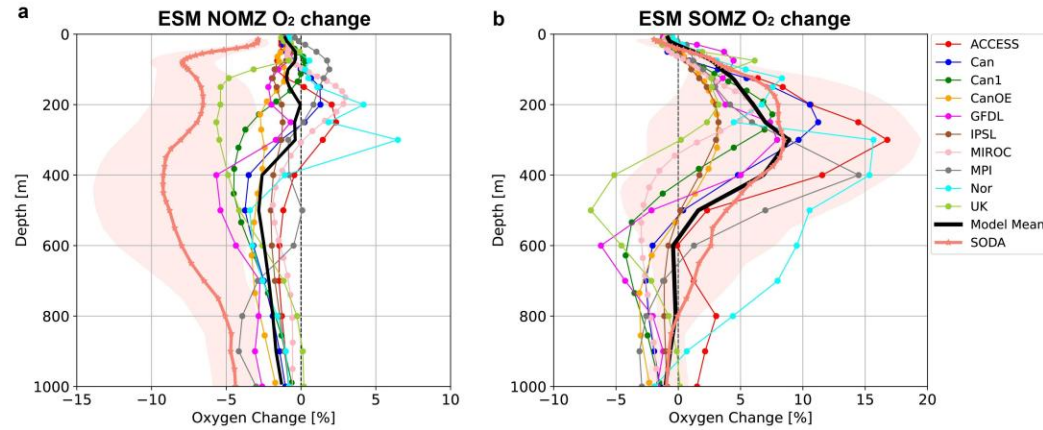
Relative contributions of salinity (S), temperature (T) and time (year, cos(month), sin(month)) predictors to ML-estimated O₂ sensitivity, using perturbation-based analysis at fixed spatial locations in each predictor for RF (A, B) and NN (C, D) models trained with ship+Argo (A, C) or ship-only (B, D) observations.

The CMIP6 problem: 10 ESMs. Ensemble mean O₂ bias

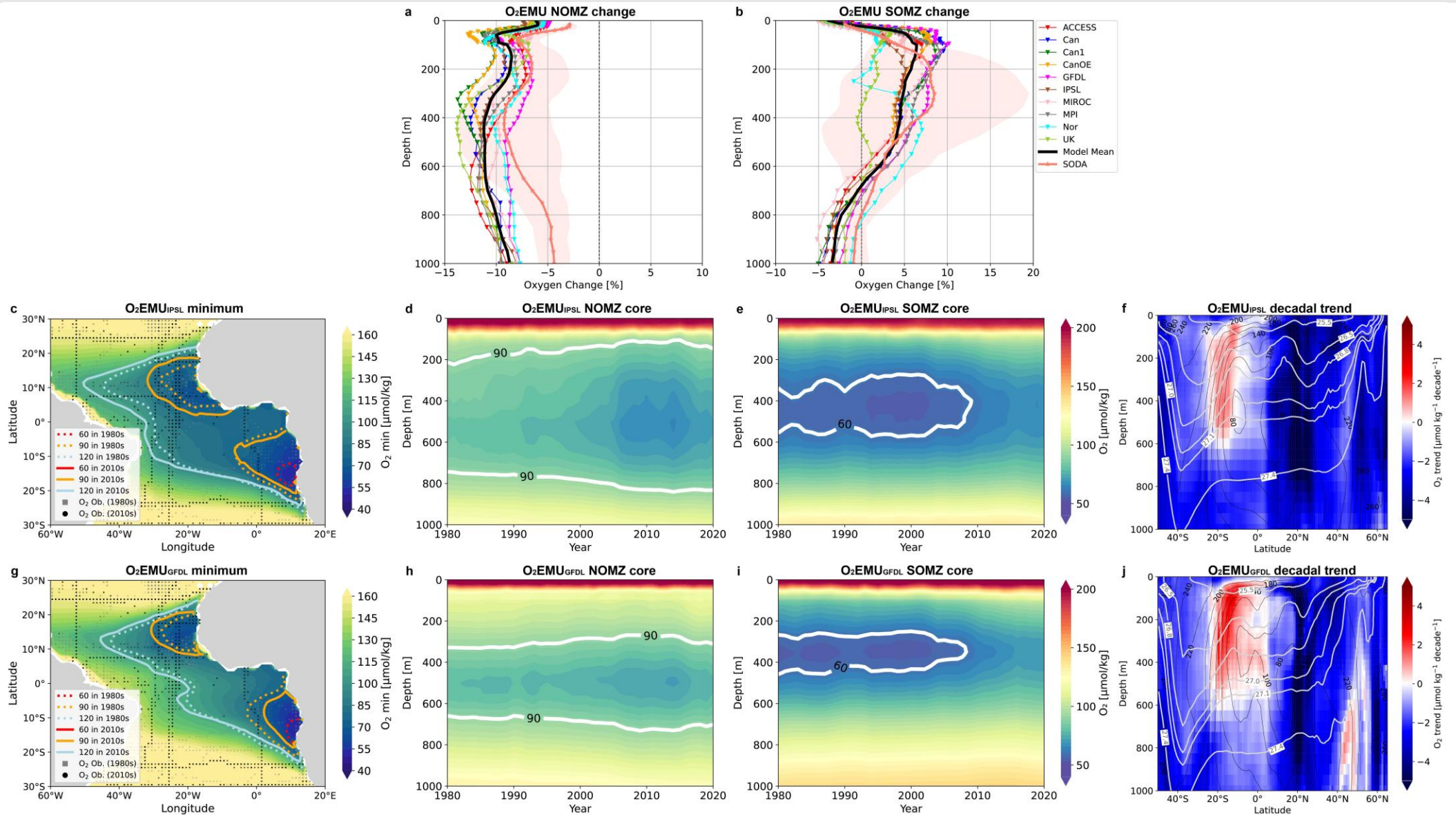


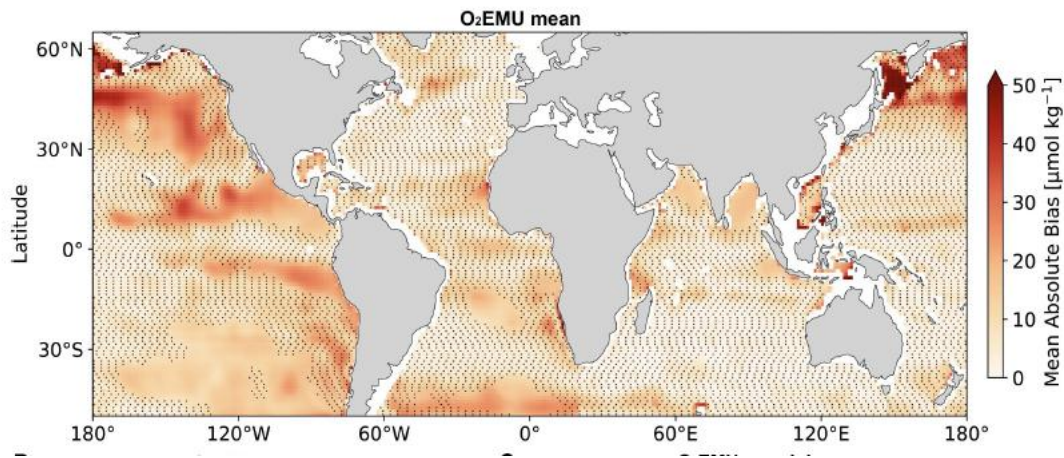
Mean Absolute Bias [$\mu\text{mol kg}^{-1}$] of 10 CMIP6 ESMs, compared to the WOA23 climatology for the depth interval 100 – 600 m. Gray dots indicate regions where the MAB falls within 2x of the objectively analyzed standard deviation of WOA23.

Ocean Oxygen in ESMs

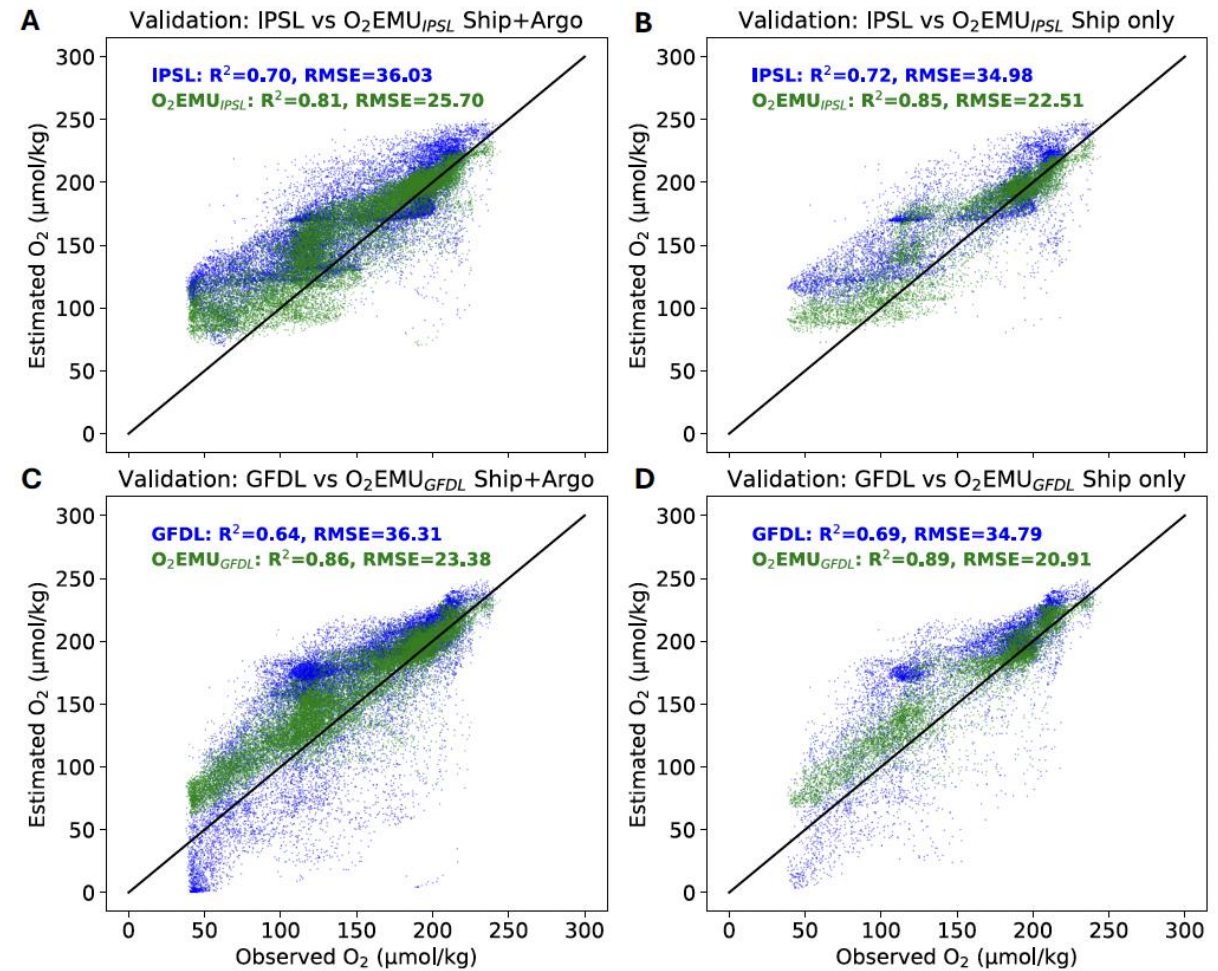


Ocean Oxygen in Emulator using T/S from ESMs





Mean Absolute Bias [$\mu\text{mol kg}^{-1}$] of 10 CMIP6 O_2EMU relative to WOA23 climatology for the depth interval 100 – 600 m. Gray dots indicate regions where the MAB falls within 2x of the objectively analyzed standard deviation of WOA23



Validation of future-predictions based on withheld WOD23 observations (period 2011-2014). Raw ESM O_2 (blue) and O_2EMU estimates (green) are compared against WOD23 O_2 observations from 2011-2014. O_2EMU trained using 1980-2010 data only. IPSL (A, B) and GFDL (C, D) under ship+Argo (A, C) and ship-only (B, D) training.

- O₂EMU does not just fill data gaps, but also estimates the variability of oxygen distributions from physical (T/S) variables.
- O₂EMU relies on the T/S variability to implicitly represent biologically-driven O₂ changes that are linked to physical circulation changes (e.g. upwelling), ventilation and temperature-regulated biogeochemical rates. ESM representation of biological variables (e.g. chlorophyll) is less reliable than T/S but there is typically a correlation between nutrient supply and T/S variability, which can be learned by ML algorithms.
- O₂ climatologies of CMIP6 models are better reproduced, and the observed relationships between physical and biogeochemical fields are captured. The O₂ evolution in CMIP6 models is now consistent with modeled physical variability. The standard deviation of O₂EMU reduces the ESM standard deviation by 70% and the intermodal spread by over 300%

- The O₂EMU emulator is a computationally efficient and scalable alternative to standard biogeochemical models. Most importantly, it has the potential to generalize to other tracers
- There remain issues with large changes and intermodal spread in T and S from ~2040. ML emulators cannot extrapolate to unseen conditions...



THANK YOU



www.cmcc.it