A Framework to Estimate Open-Ocean Diatom Carbon Biomass from Remote Sensing Observations

Alison Chase,[1*] Claire Berschauer,[1] Valentina Staneva,[2] Nils Haëntjens,[3] Charles Stern,[4] Emmanuel Boss,[3] Lee Karp-Boss,[3] Guillaume Bourdin,[3] Peter Gaube[1]

[1] Applied Physics Laboratory, University of Washington, Seattle, WA USA
[2] eScience Institute, University of Washington, Seattle, WA USA
[3] School of Marine Sciences, University of Maine, Orono, ME USA
[4] Lamont Doherty Earth Observatory, Columbia University, Palisades, NY USA
* contact: alichase@uw.edu

Summary

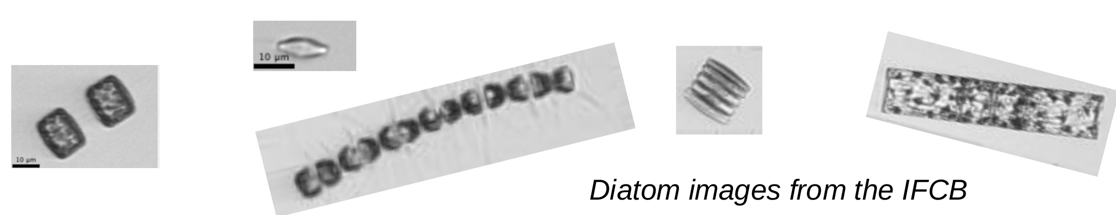
Spatial and temporal variability in diatom carbon biomass impacts carbon cycling and the flow of particulate biomass to higher trophic levels. To improve estimates of diatom carbon biomass on regional-to-global scales, we train a Random Forest regression model on in situ data. We then demonstrate model application to recent PACE OCI data, showing novel patterns in remote sensing-based diatom carbon biomass estimates.

THE FRAMEWORK

1. Plankton imagery data processing
2. DiatC model training and testing
3. Deployment of the model on PACE data
4. Model evaluation with in situ measurements

This framework is designed to be both iterated on and evaluated as additional in situ diatom carbon biomass ('DiatC') measurements are collected, thus improving the accuracy and robustness of the model for global application.

1. ifcbUTOPIA: Plankton imagery data processing



Diatom images from the IFCB

Phytoplankton images from an Imaging FlowCytobot (IFCB) are classified via a trained convolutional neural network (CNN). These images are then used to calculate diatom carbon biomass.

The diatom carbon biomass is subsequently used as the target parameter in a Random Forest model to estimate DiatC from environmental and optical parameters measured coincidently in situ.

ifcbUTOPIA

User-friendly Tools for Oceanic Plankton Image Analysis (UTOPIA) is for use with data from the Imaging FlowCytobot (IFCB)

Code and documentation: https://github.com/ifcb-utopia

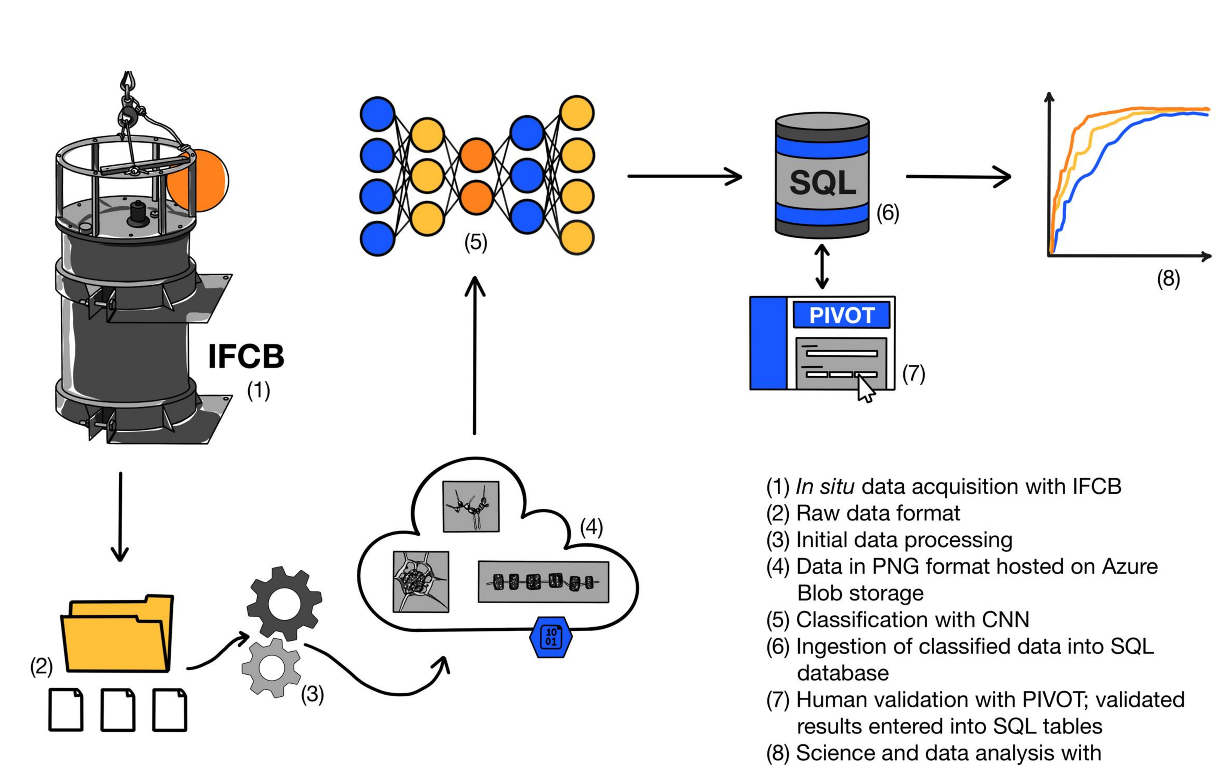
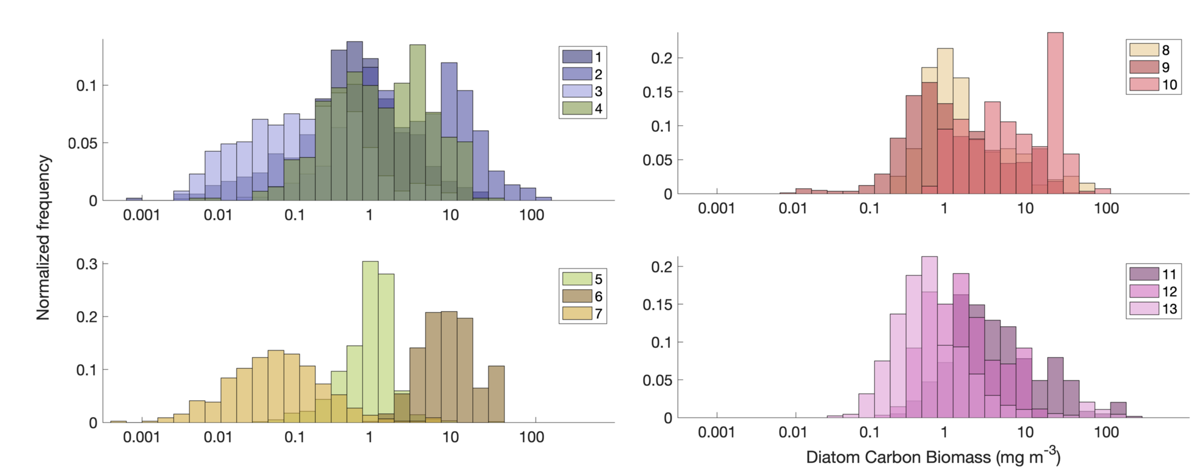


Diagram of the workflow to process IFCB images and calculate diatom carbon biomass for subsequent DiatC model development.

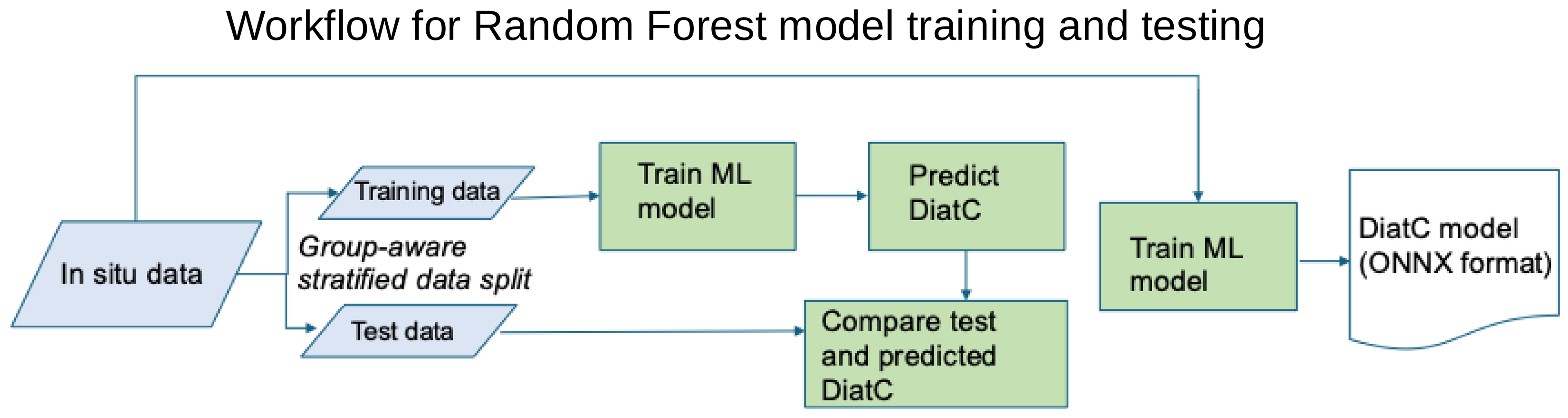
The ifcbUTOPIA workflow is designed for upcoming IFCB datasets collected during PACE mission validation, streamlining phytoplankton group biomass estimates for PACE product validation and algorithm refinement.



Distributions of diatom carbon biomass per sample calculated from IFCB images across 13 cruises or cruise segments (locations in map at bottom left of poster).

2. DiatC Model Training and Testing

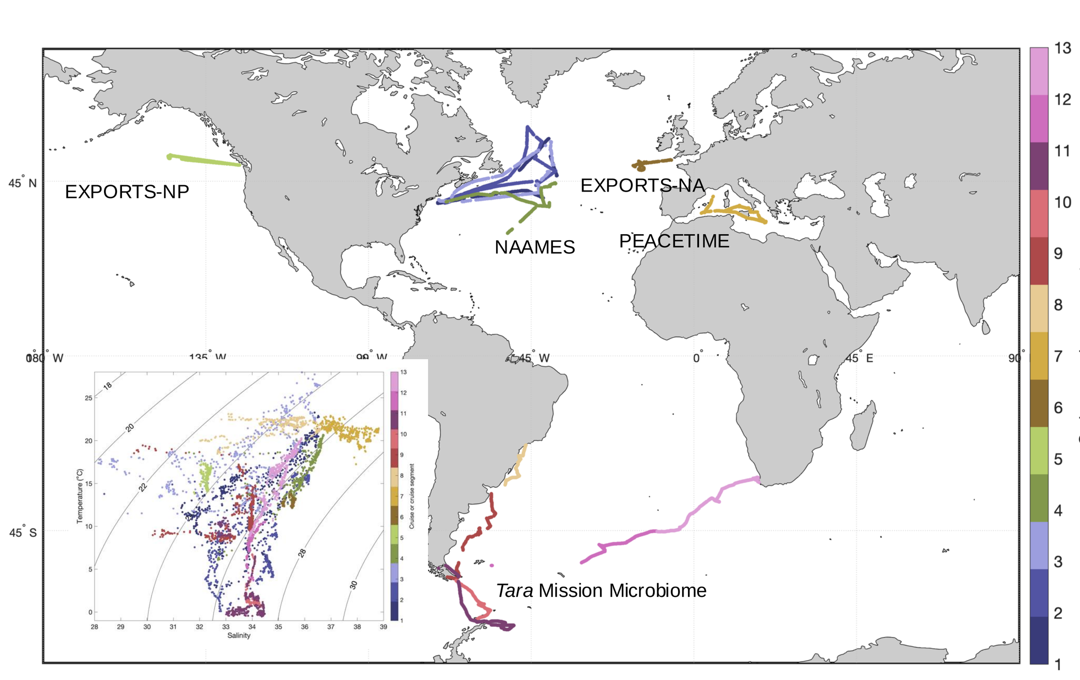
Workflow for Random Forest model training and testing



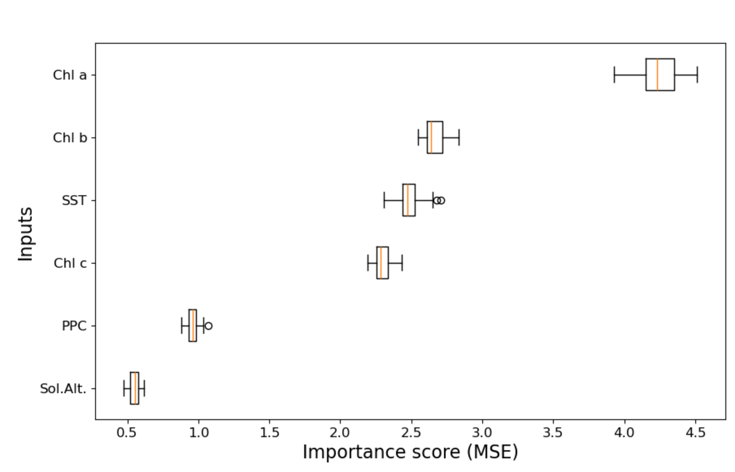
Six model inputs: Chl *a*, Chl *b*, Chl *c*, photoprotective carotenoids (PPC), SST, solar altitude
Model target: Diatom carbon biomass from IFCB images (calculated in step 1. above)
Pigments are estimated from spectral particulate absorption measurements (Chase et al. 2013)

Four cruises (2, 6, 7, 9 in map below) are held out as test data. We use group-aware stratified splitting during model training and testing to mitigate artificially high accuracy metrics that occur from 'data leakage' that can occur if random train/test data splitting is used (Stock 2022).

DiatC model metrics based on test data	
Median Abs. Error	1.59 mg m^{-3}
Mean Bias Error	-3.21 mg m^{-3}
63% of Error Abs. Val are ≦ 3 mg m^{-3}	



Locations and TS diagram of in situ data collected during 2015-2022 (n = 9256).

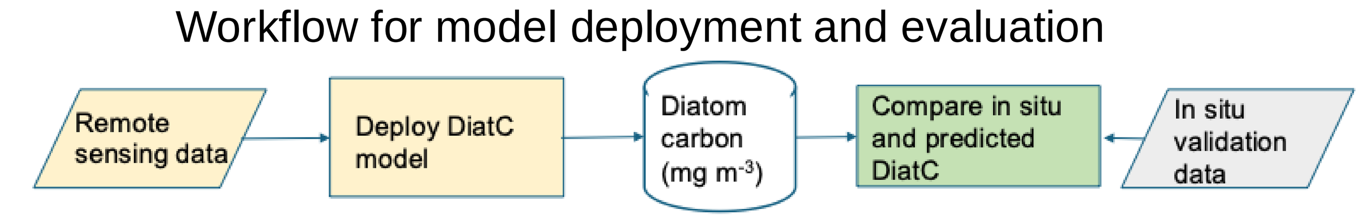


Permutation importances for the six input parameters, calculated on test data as mean squared error (MSE) and with 30 permutations.

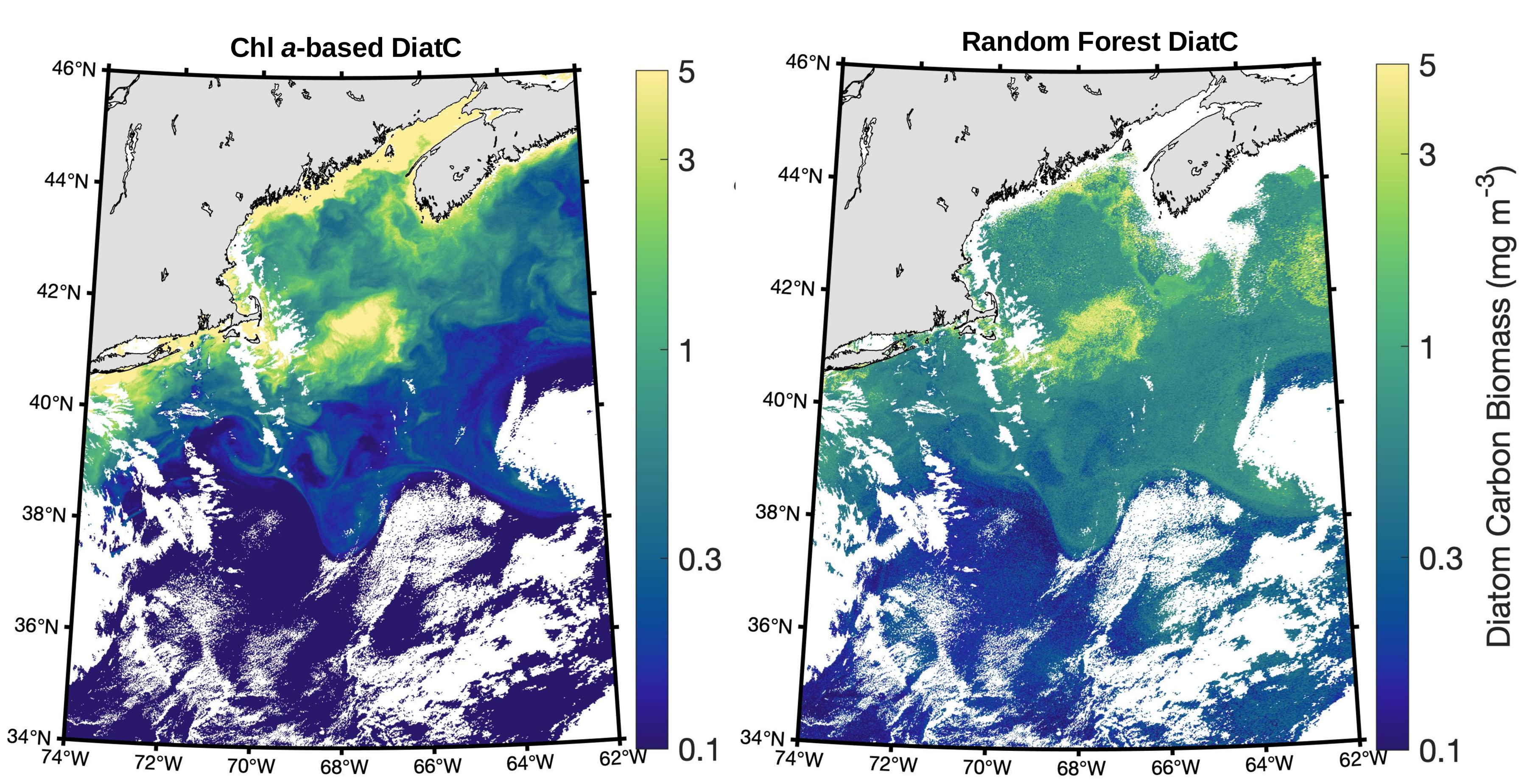
This work is funded by NASA grant 80NSSC20M0202. Thanks to all involved in collection of the data used in this study. Thank you to Patrick Gray and Steve Mussmann for helpful discussion on Machine Learning.

3. Diatom Carbon Biomass from PACE OCI

Workflow for model deployment and evaluation



The DiatC model is deployed on remote sensing measurements matching the inputs used during training. Pigments are calculated from PACE OCI Rrs following Chase et al. (2017), and SST is obtained from the GHRSST Level 4 MUR Global SST Analysis (v4.1).



Left: Chl a-based DiatC calculated as DiatC = 1.5[Chl a]$^{1.9}$ (baseline model from Chase et al. 2022). Right: Random Forest DiatC model deployed on remote sensing data showing PACE-derived diatom carbon biomass (mg m^{-3}). Differing spatial patterns indicate potential new information via the use of the Random Forest model that makes use of multiple input parameters during model training.*

Further model evaluation and iteration will be performed using IFCB data collected during ongoing and upcoming PACE mission validation team activities.

References:
Chase et al. 2013, Methods in Oce., dx.doi.org/10.1016/j.mio.2014.02.002; Chase et al. 2017, Journal Geophys. Res. – Oceans, doi.org/10.1002/2017JC012859; Chase et al. 2022, Geophys. Res. Lett., doi. org/10.1029/2022GL098076; Stock 2022, ISPRS Journal Photog. Rem. Sense, doi.org/10.1016/j.isprsjprs.2022.02.023