ML MINI-WORKSHOP

PLANKTON PROXIES

TEAM & COLLABORATORS: MONIQUE, BEN, FRANCISCO, MIKE M, KEVIN G, KATIE, YANWU, JOHN R

OVERARCHING GOAL

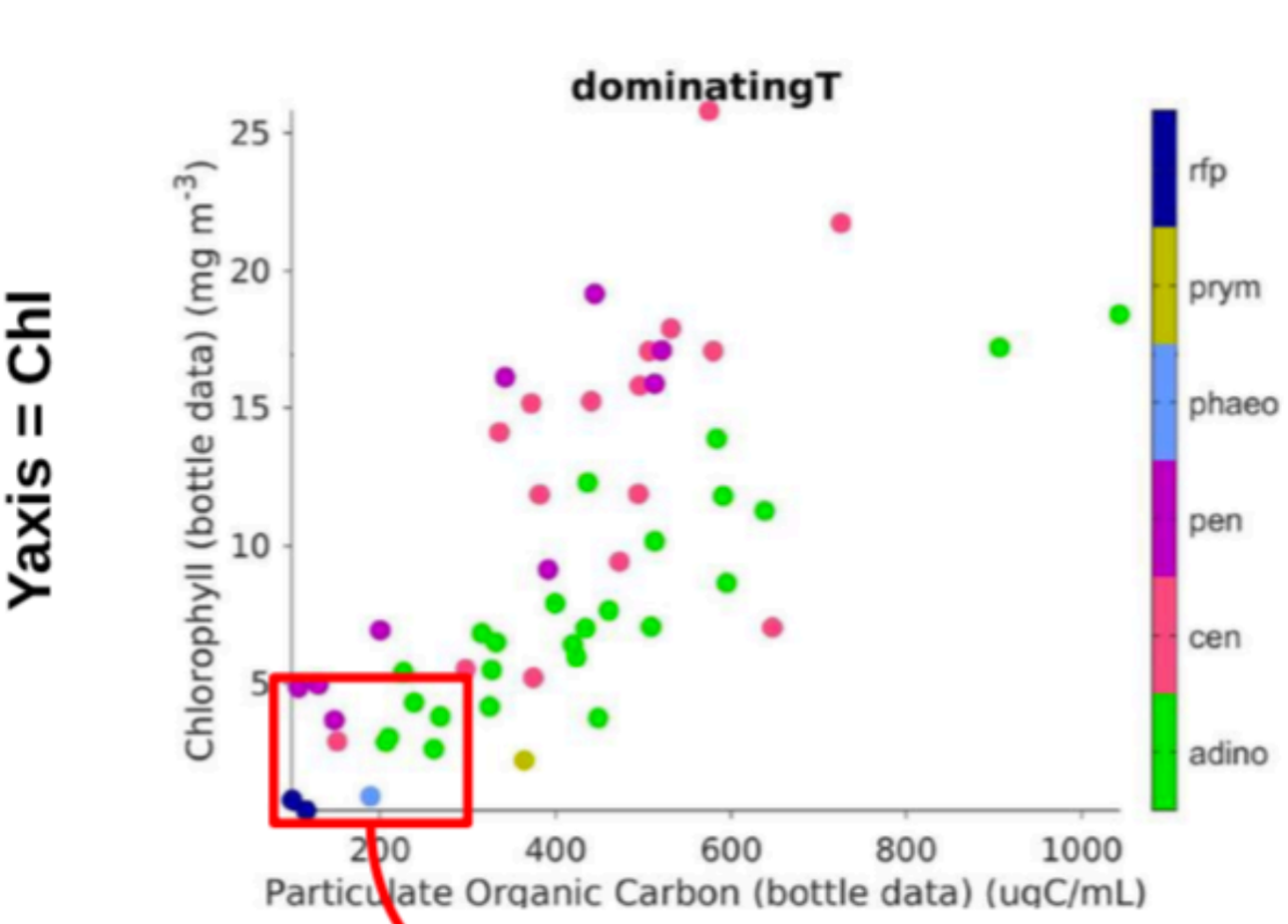
- Characterize phytoplankton communities using AUV sensor data
- More precisely, discriminate between:
  - Diatoms, Dinoflagellates and Picoplankton communities
- Using core payload sensors:
  - e.g., temperature, salinity, fluorescence, backscatter, and others

WHY IS THIS IMPORTANT

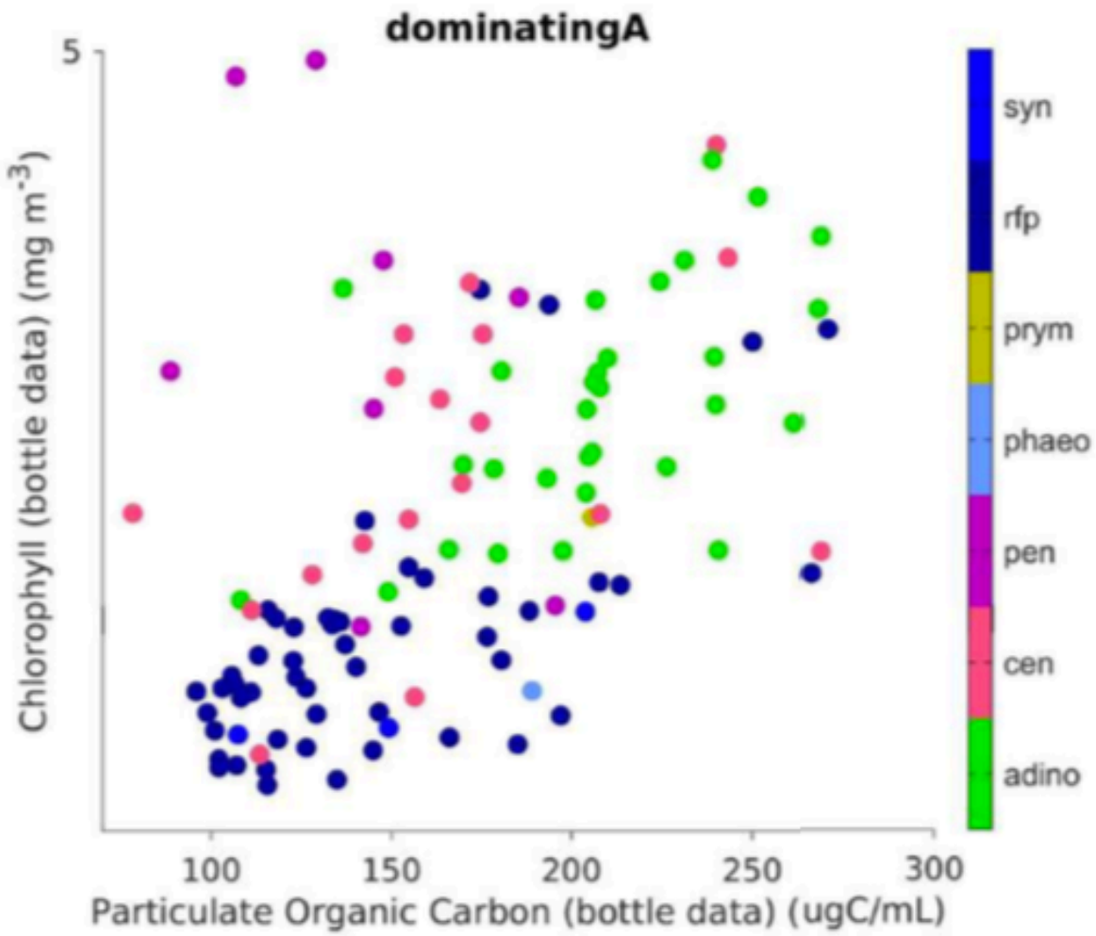
- MBARI currently relies on in situ sampling to get taxonomic information, whether from ship rosette, Dorado gulper, LRAUV sipper or ESP
- Reliable proxies for phytoplankton taxonomic groups derived from commonly used optical fluorescence/backscatter sensors:
  - investigate decades of historical data and would provide an unprecedented view of phytoplankton ecology in the region
  - real-time processing of phytoplankton taxa proxies onboard AUVs would enable more advanced targeted sampling

BASIS – C:CHL TAXA SPECIFIC

Dominating = **above 80%**



Dominating = **above 50%**



Xaxis = POC

zoom

PLANKTON PROXIES

- Fluorescence = proxy for chlorophyll
- Backscatter = proxy for carbon
- Labelled datasets:
    - CTD BOG count timeseris
    - bioluminescence nighttime labels (Dorado dataset)
    - wharf counts (S & N corners)
    - eDNA/QP (M1 & C1 corners)

MACHINE LEARNING PROCESSING PIPELINE

- Consolidate datasets overlapping with labels
- Cluster analysis for extracting data segments dominated by single taxon
- Develop classification models using Support Vector Machines (SVMs)
- Compute learning curves and performance metrics to support model compression and identify optimal predictors for each class

NEXT YEAR

- 1st 6 months:
  - assess LISST-100x dataset
  - load Dorado dataset into STOQS & create architecture for loading taxa data labels
- 2nd part:
  - generate the taxa labels from bioluminescence / wharf & load into STOQS
  - Perform supervised ML analysis on using 2 sets of labels