Biogeochemical proxies

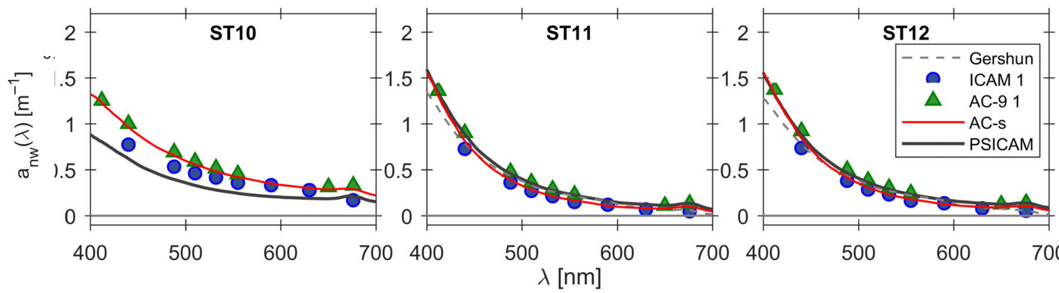
Meg Estapa

Ocean Optics class 2023

Quick review:

Which figure gives the best example of

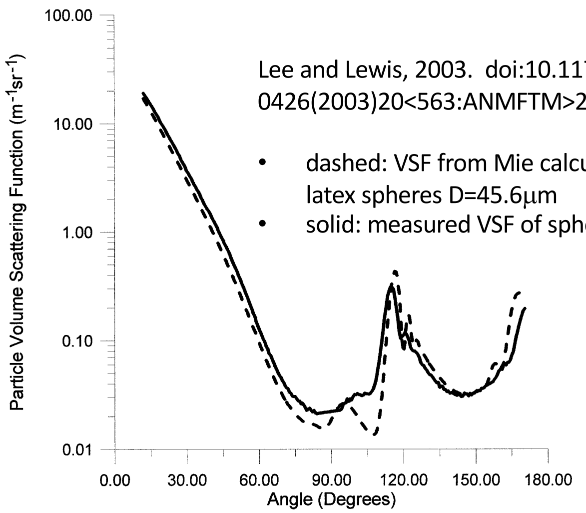
1) proxy
2) calibration
3) validation



Kostakis et al., 2021. doi:10.1002/lom3.10447

- $a_{nw}(\lambda)$ from different instruments/techniques at the same stations

Fig 5. Nonwater absorption spectra, a_{nw}, measured using ICAM 1, PSICAM, AC-s, AC-9 1, and derived from measurements of the irradiance quartet using Gershun's law. No ICAM, AC-9, or AC-s data were collected at Sta. 3.



Lee and Lewis, 2003. doi:10.1175/1520-0426(2003)20<563:ANMFTM>2.0.CO;2

- dashed: VSF from Mie calculations for latex spheres D=45.6μm
- solid: measured VSF of spheres

FIG. 5. Comparison of theoretical Mie calculations (dashed curve) and direct measurements (solid curve) of the VSF of latex microspheres with a diameter of 45.6 μm.

Gardner et al., 2006. doi:10.1016/j.dsr2.2006.01.029

- y-axis: POC concentration from bottle samples, global JGOFS data set
- x-axis: beam attenuation coefficient from same casts

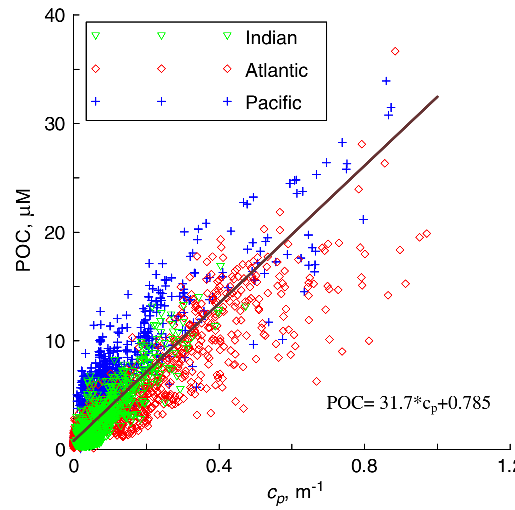


Fig. 3. Global POC—Beam c_p regression calculated on all available data collected in Indian (∇), Atlantic (◇) and Pacific (+) oceans. See Table 1, "All data" for regression parameters.

In the context of this class:

Proxy = use of an empirical relationship between two measurements to estimate one from the other.

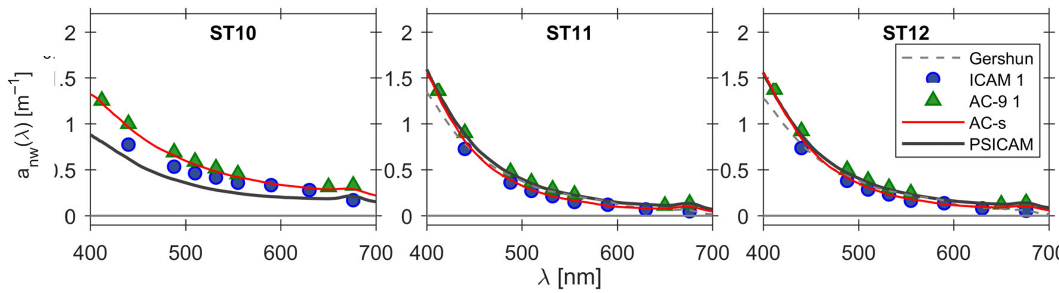
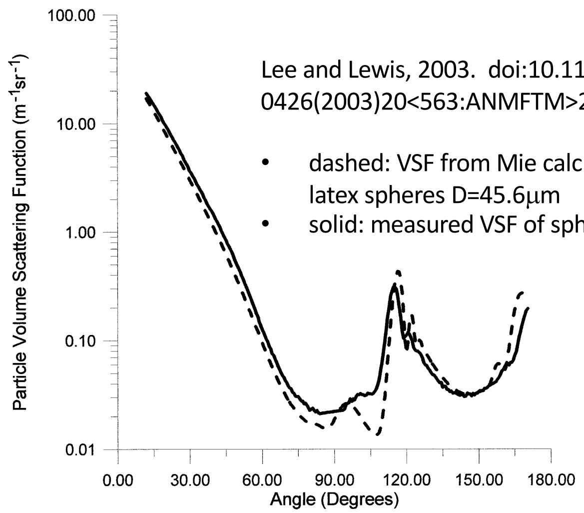


Fig 5. Nonwater absorption spectra, a_{nw}, measured using ICAM 1, PSICAM, AC-s, AC-9 1, and derived from measurements of the irradiance quartet using Gershun's law. No ICAM, AC-9, or AC-s data were collected at Sta. 3.

Kostakis et al., 2021. doi:10.1002/lom3.10447

- $a_{nw}(\lambda)$ from different instruments/techniques at the same stations



Lee and Lewis, 2003. doi:10.1175/1520-0426(2003)20<563:ANMFTM>2.0.CO;2

- dashed: VSF from Mie calculations for latex spheres D=45.6μm
- solid: measured VSF of spheres

FIG. 5. Comparison of theoretical Mie calculations (dashed curve) and direct measurements (solid curve) of the VSF of latex microspheres with a diameter of 45.6 μm.

Gardner et al., 2006. doi:10.1016/j.dsr2.2006.01.029

- y-axis: POC concentration from bottle samples, global JGOFS data set
- x-axis: beam attenuation coefficient from same casts

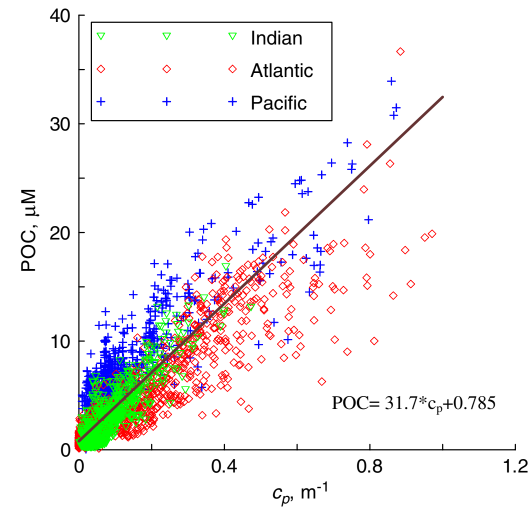


Fig. 3. Global POC—Beam c_p regression calculated on all available data collected in Indian (∇), Atlantic (◇) and Pacific (+) oceans. See Table 1, "All data" for regression parameters.

Overview

✓Review of shared concepts across recent lectures

- Why create and use proxies?
- Some issues and cases to highlight
  - How many independent proxies can be extracted from observations?
  - Uncertainties in biogeochemical measurements
  - Restricted domain of the “training” dataset and extrapolation beyond it
- Missing examples?

Why create and use proxies?

- Aquatic processes have different scales of variability and we need to match our observations to those scales
- Others?

What assumptions do we make when creating a proxy?

Carrying capacity (F(nutrients, light))

Phytoplankton

Zooplankton

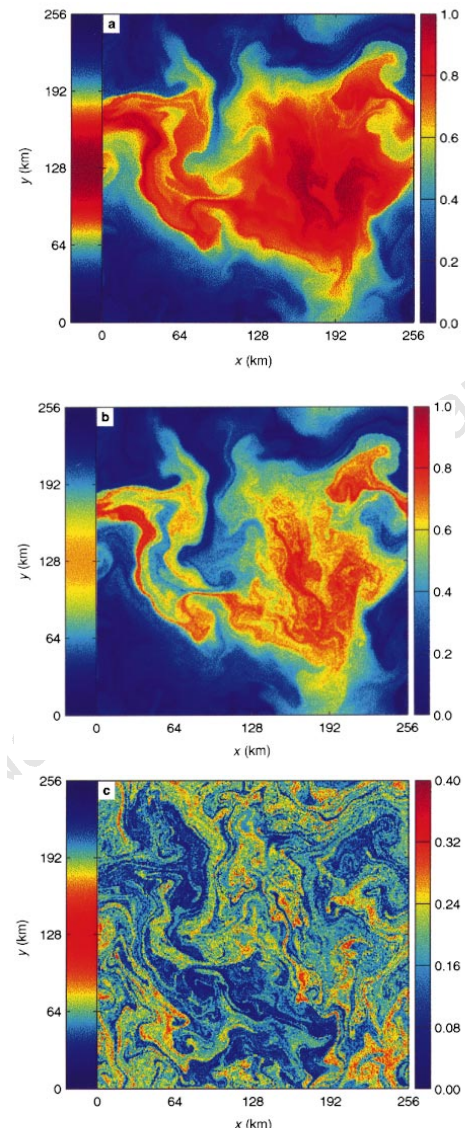
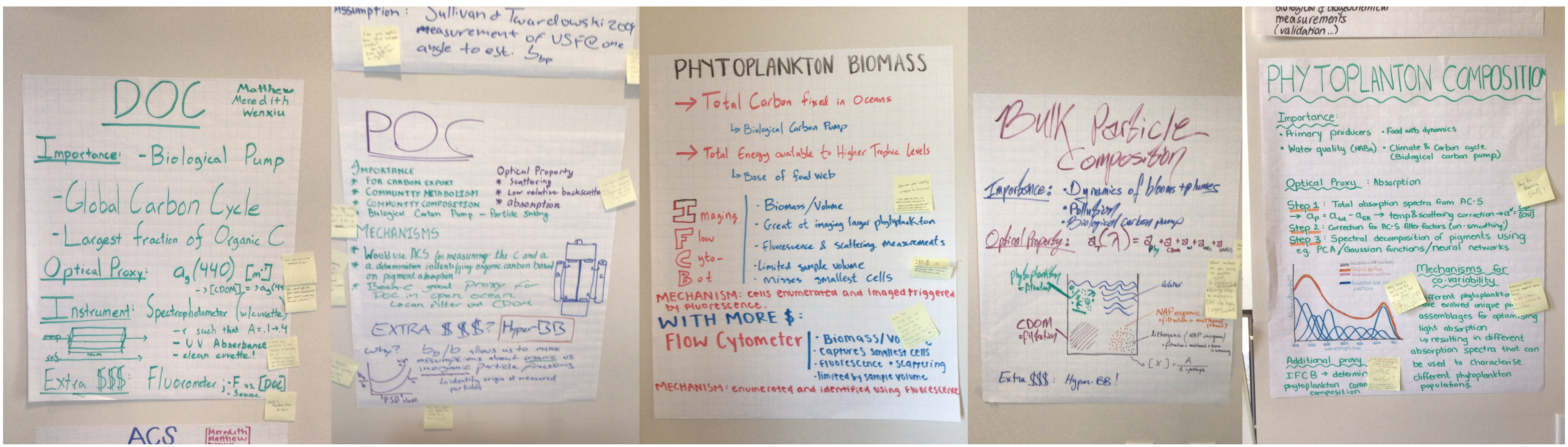


Figure: Abraham et al., 1998. *Nature* 391(6667): 577-580.

Proxy relationships have a physical basis

- DOC – $a_g(440)$
- POC – c_p
- Phytoplankton biomass – geometric area (IFCB images)
- Bulk particle composition – Beer's Law, components of $a_p(\lambda)$
- Phytoplankton composition – Beer's Law, pigment contributions to $a_p(\lambda)$

To first order, IOPs scale with concentration.
If we want to know *composition (size, etc)*, we need multiple, independently-varying wavelengths, angles, filtered/unfiltered, polarization, etc.



How many independent constituents can be extracted from an absorption spectrum?
(Cael, Chase, and Boss, 2020. *Appl Opt*)

Absorption spectra have pretty similar shapes...

"we are looking for small differences in noisy measurements to parse between covarying pieces of information."

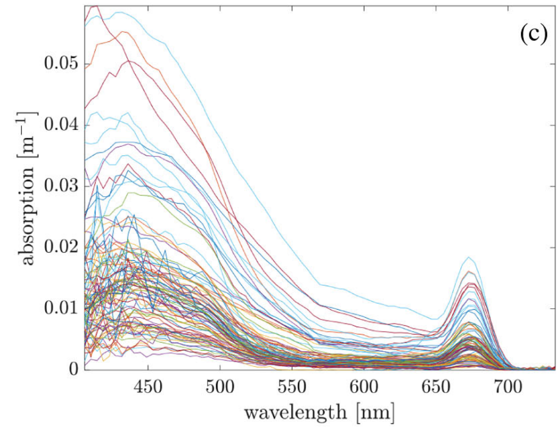
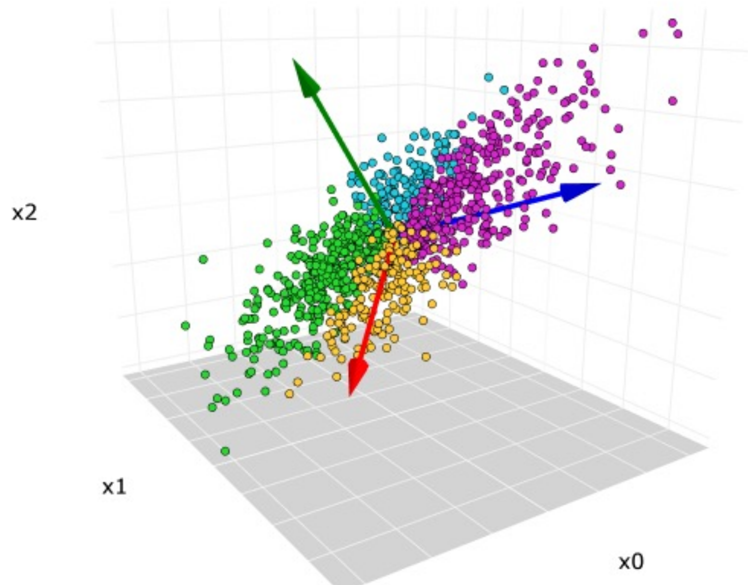
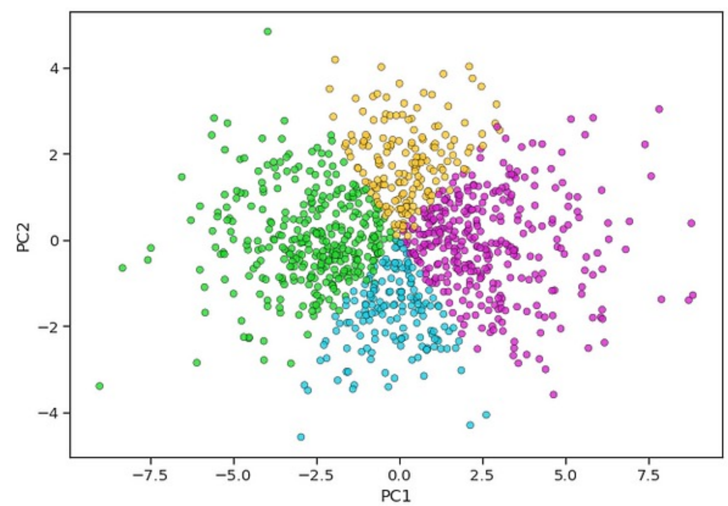


Figure: Cael et al. 2020, 10.1364/AO.389189

Principal component analysis (PCA): Linearly transform the data so that the greatest amount of variance lies along the first axis (first component), the next greatest amount along the second axis (second component), and so forth.



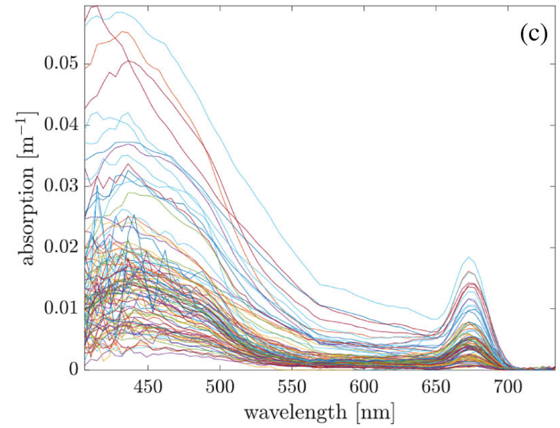


Figures: Cheng, C., 2022. https://towardsdatascience.com/principal-component-analysis-pca-explained-visually-with-zero-math-1cbf392b9e7d. Accessed 6/25/2023.

How many independent constituents can be extracted from an absorption spectrum? (Cael, Chase, and Boss, 2020. *Appl Opt*)

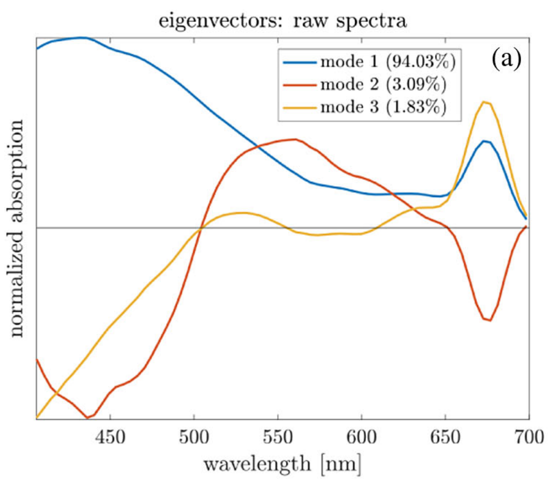
Absorption spectra have pretty similar shapes...

“we are looking for small differences in noisy measurements to parse between covarying pieces of information.”

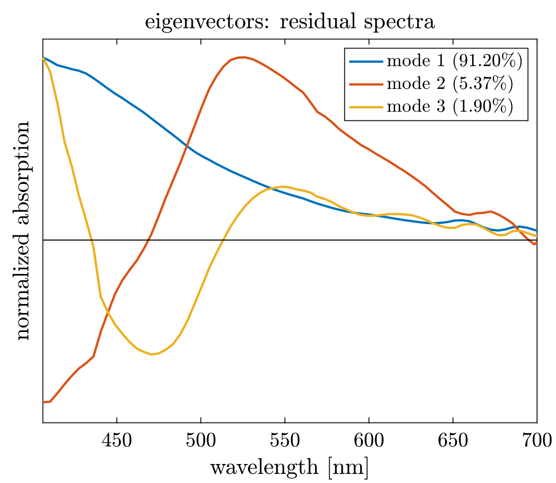


Principal component analysis (PCA): What are the basis vectors that sequentially describe the greatest amount of variance in the data?

PCA on $a_p(\lambda)$ from Tara Oceans...



Use a_{LH} to remove chl first, then PCA... now mode 1 looks like NAP



Take homes:

- There are 4-5 degrees of freedom (independently-covarying components) in hyperspectral $a_p(\lambda)$ observed *in situ*
- Overall amplitude/chlorophyll and NAP explain most of the variance
- To get more, you need really low uncertainty *or* **other sources of information**

Figures: Cael et al. 2020, 10.1364/AO.389189

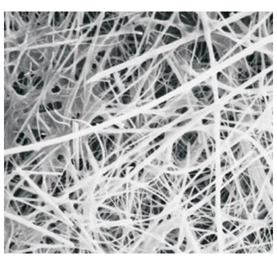
Overview

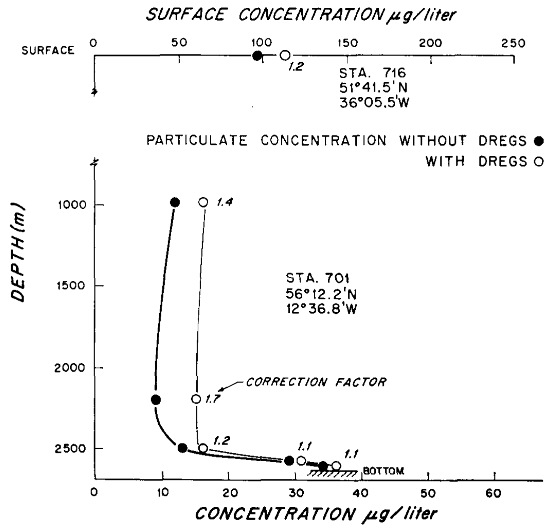
- Review of shared concepts across recent lectures
- Why create and use proxies?
- Some issues and cases to highlight
  - How many independent proxies can be extracted from observations?
  - Uncertainties in biogeochemical measurements
  - Restricted domain of the “training” dataset and extrapolation beyond it
- Missing examples?

Uncertainties in the "sea-truth" measurements used to build the proxy

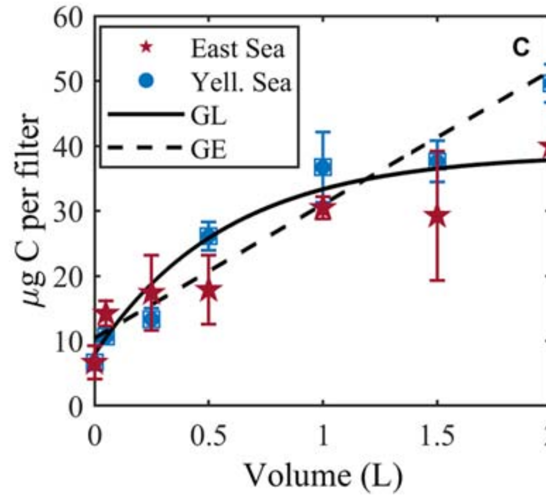
Example: Particulate organic carbon

- "Dregs" in Niskin bottles
- Protection of sample from air contamination, handling contamination at all steps
- Low vacuum pressure
- Dissolved organic carbon blanks
  - DOC "adsorption sites" on GF/F filters
  - Adsorption is not instantaneous nor is it always linear with volume filtered

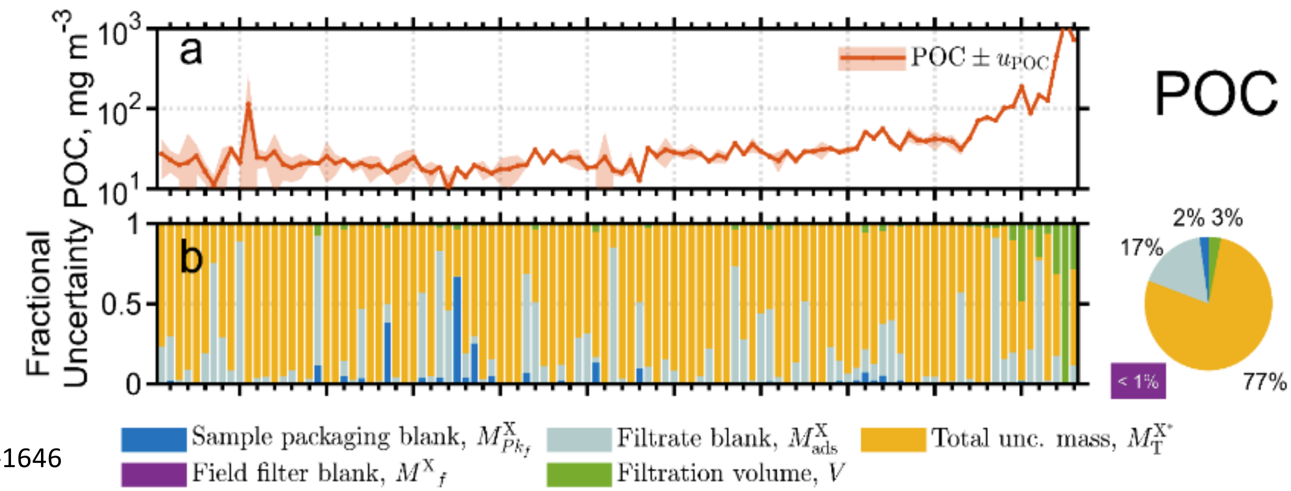




Gardner, 1977. 10.4319/lo.1977.22.4.0764



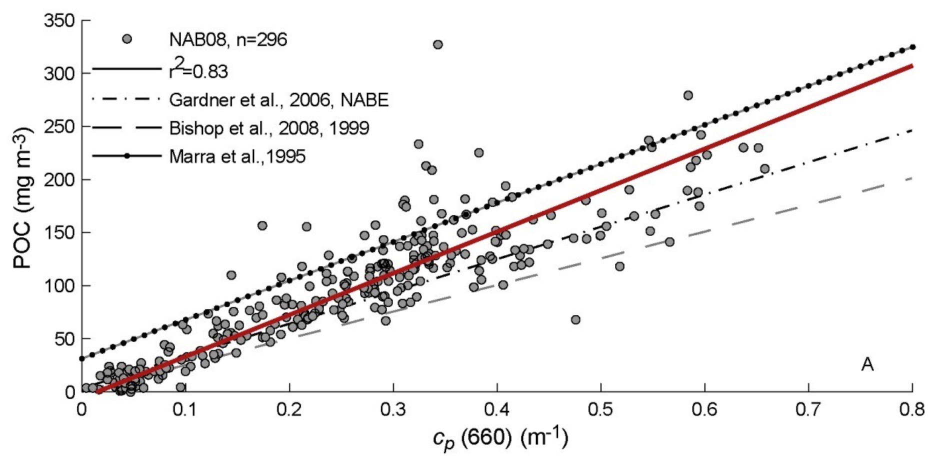
Novak et al. 2018. 10.1002/lom3.10248



Chaves et al. 2021. 10.25607/OBP-1646

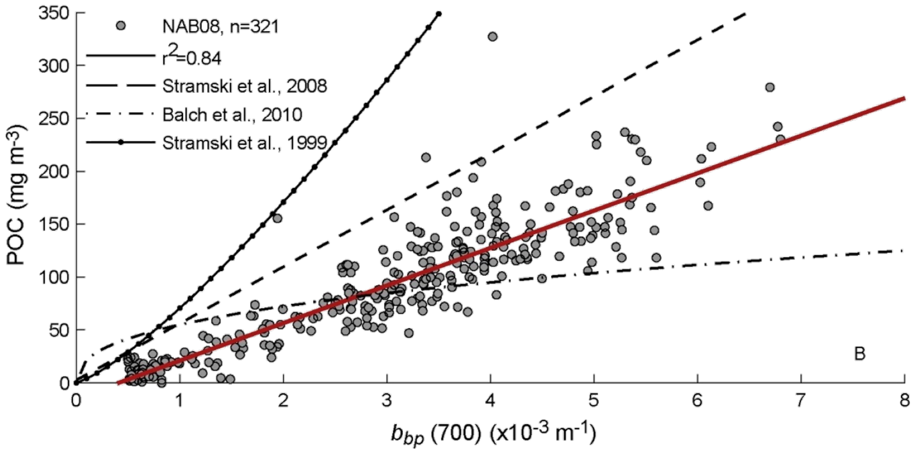
Particulate Organic Carbon

Beam attenuation (c_p)



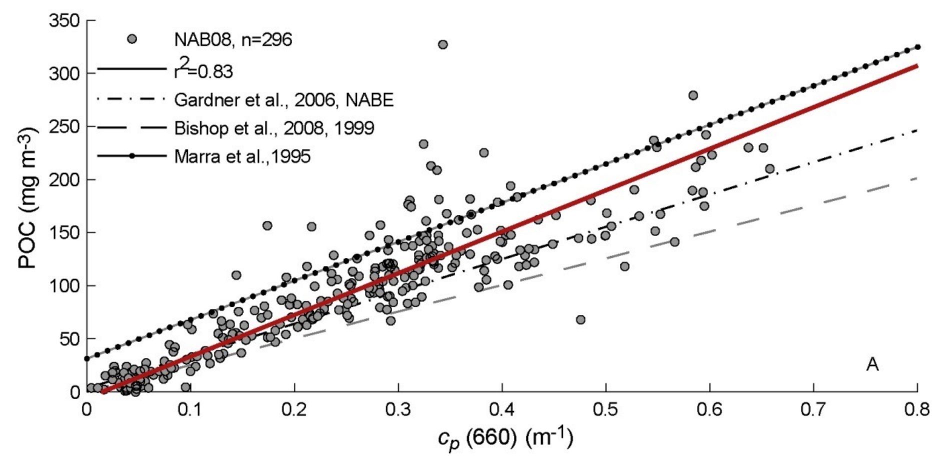
Figures: Cetinić et al., 2012, JGR. 10.1029/2011JC007771
Slide from I. Cetinić

Particulate Organic Carbon



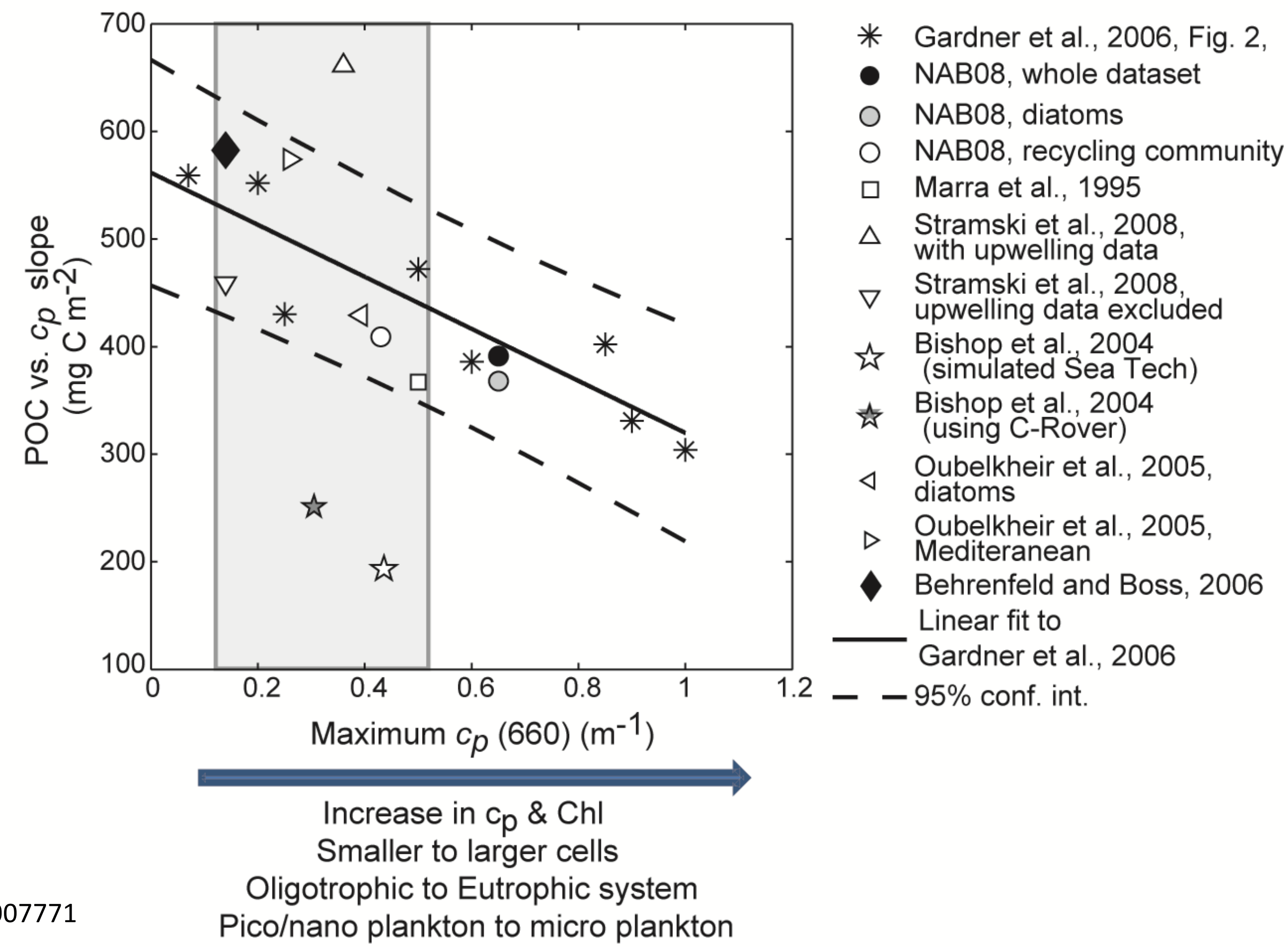
Particulate backscattering (b_{bp})

Beam attenuation (c_p)



Figures: Cetinić et al., 2012, JGR. 10.1029/2011JC007771
Slide from I. Cetinić

POC/c_p slope comparison (mg C m^{-2})



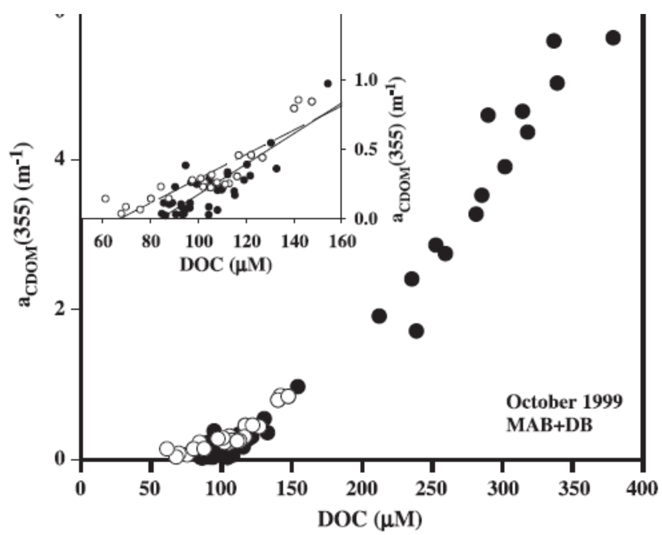
Figures: Cetinić et al., 2012, JGR. 10.1029/2011JC007771
Slide from I. Cetinić

Overview

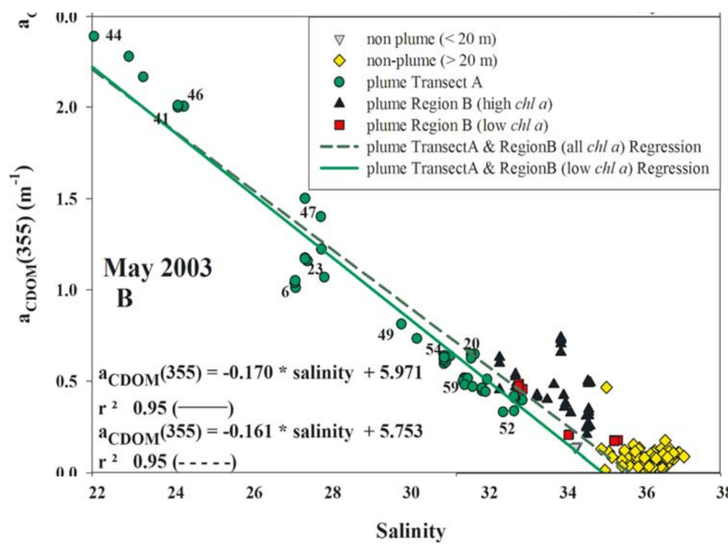
- Review of shared concepts across recent lectures
- Why create and use proxies?
- Some issues and cases to highlight
  - How many independent proxies can be extracted from observations?
  - Uncertainties in biogeochemical measurements
  - Restricted domain of the “training” dataset and extrapolation beyond it
- Missing examples?

Domain of training data: Dissolved Organic Carbon

CDOM can approximate DOC in the coastal ocean



Del Vecchio and Blough, 2004



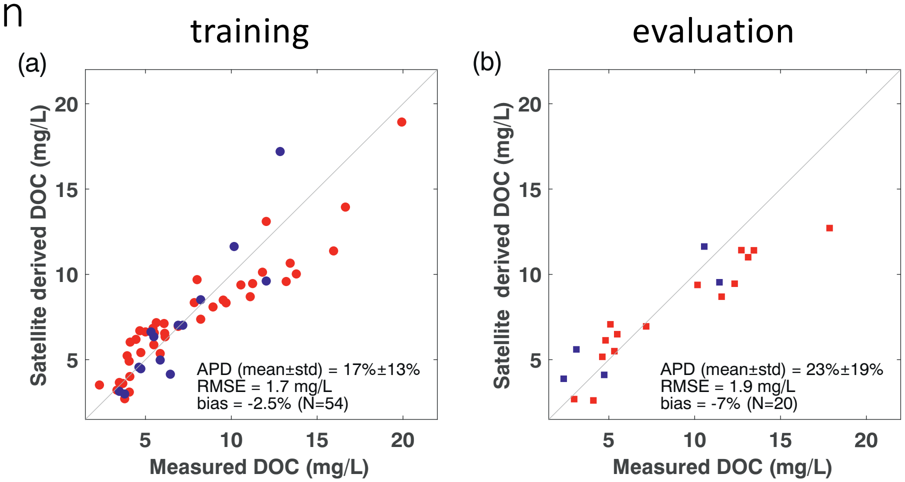
Del Vecchio and Subramaniam, 2004

- Every coastal region is different
- Better relationships when there is one strong source (e.g. river plume) and one major loss process (e.g. dilution into ocean)
- Photochemical loss ("bleaching") changes the DOC:CDOM relationship!

Domain of training data: Dissolved Organic Carbon

- Cao and Tzortziou (2020) – marsh system in Chesapeake Bay
- Missing *in situ* a_g so created empirical DOC algorithm straight from high res $R_{rs}(\lambda)$

$$\mathrm{DOC} = \exp\left(0.544 \times \log(R_{rs}(\mathrm{B1})) - 0.571 \times \log(R_{rs}(\mathrm{B2})) - 2.181 \times \log(R_{rs}(\mathrm{B3})) + 1.398 \times \log(R_{rs}(\mathrm{B4})) - 1.406\right)$$



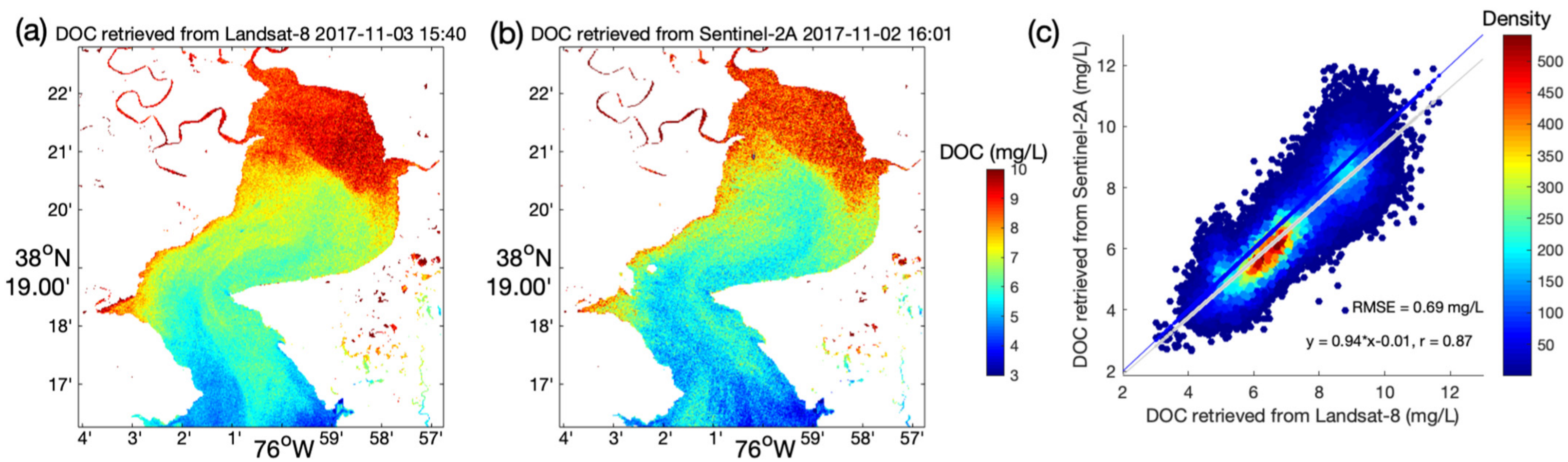


Fig. 7. DOC data consistency between (a) Landsat-8/OLI and (b) Sentinel-2/MSI in the BNWR-Fishing Bay marsh-estuary system. (c) Density scatterplot of Sentinel-2 versus Landsat-8 DOC retrievals. The regression fit and 1:1 line are shown as gray and blue lines, respectively.

Domain of training data: Dissolved Organic Carbon

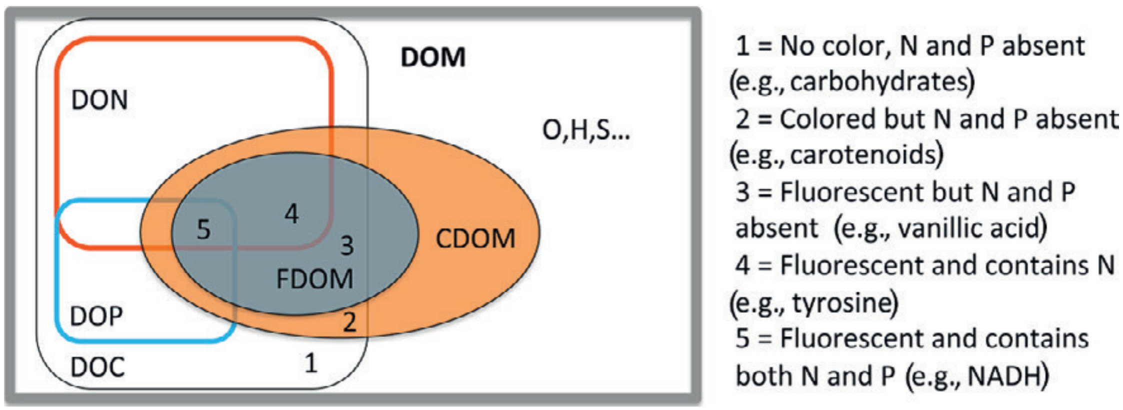
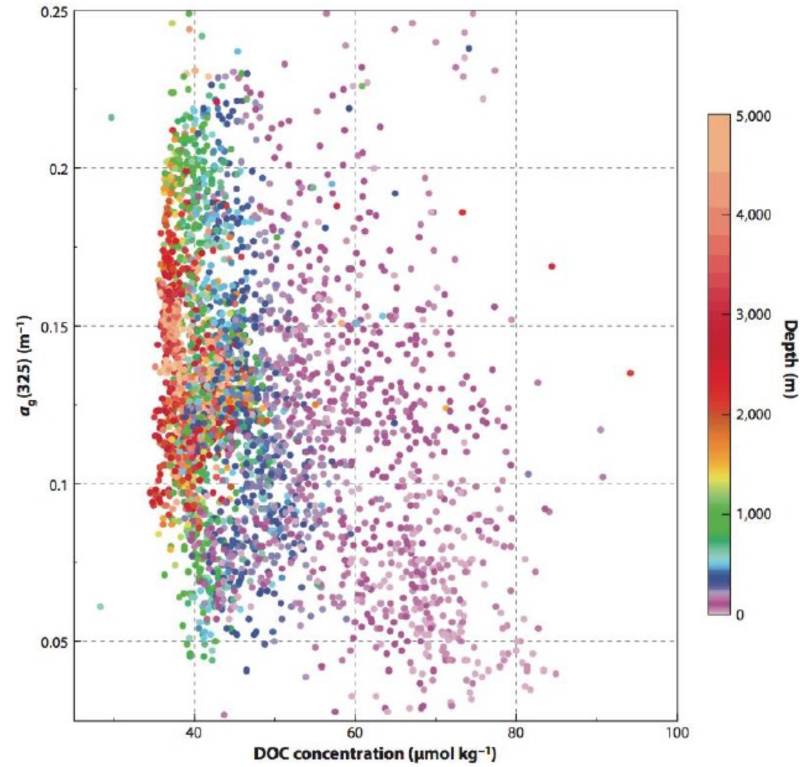


Figure: Stedmon and Nelson, 2015, *Biogeochemistry of Marine Dissolved Organic Matter, 2nd ed.*

Perhaps unsurprisingly, CDOM ≠ DOC in the open ocean



Nelson and Siegel 2013

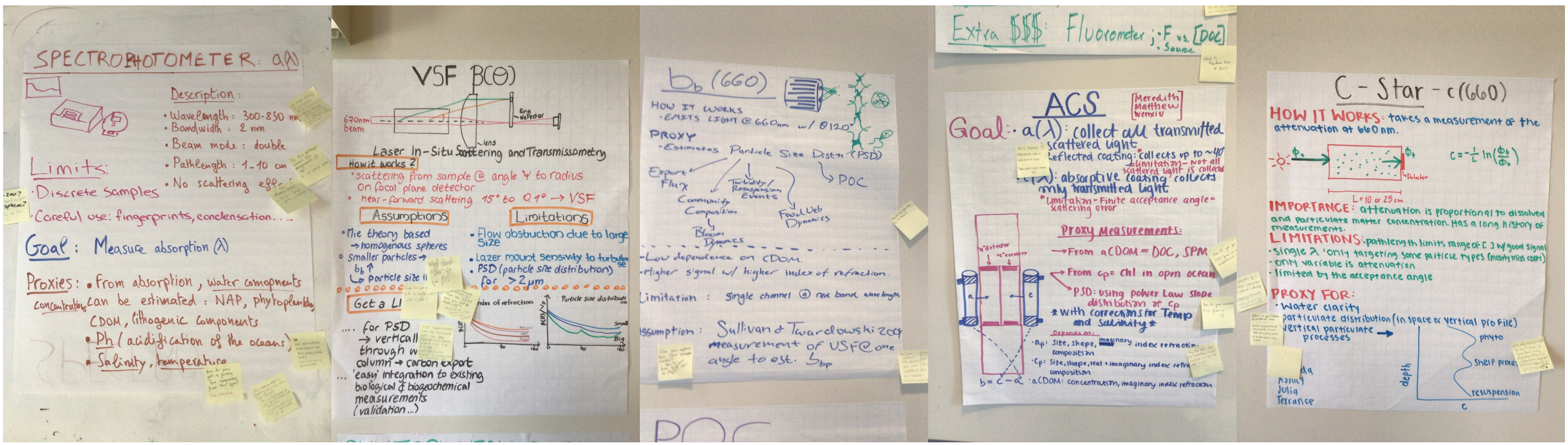
Eg. - global DOC from space – Aurin et al., 2018.
Empirical regression against 4 wavebands (440-555 nm) **+ salinity** required, RMSE still 27-29 μmol L^{-1}

Overview

- Review of shared concepts across recent lectures
- Why create and use proxies?
- Some issues and cases to highlight
  - How many independent proxies can be extracted from observations?
  - Uncertainties in biogeochemical measurements
  - Restricted domain of the “training” dataset and extrapolation beyond it
- Missing examples?

Discussion in poster working groups from last week:

- For your sensor, pick *one* of the proxies you listed
- Discuss (~15 min):
  - What type of samples should you collect to develop your proxy?
  - Where and when will you collect them?
  - How will you validate your proxy?
  - Unlike last week's exercise, you have a (finite, yet unspecified) budget... what are your biggest priorities?
- **Goal: Think through the *process* of proxy development considering the issues highlighted in this lecture so far.** We don't expect you to have read all the literature on your specific example proxy.



Proxies that have been mentioned in class so far...

- POC (c_p, b_{bp} sometimes)
- SPM (c_p, b_{bp}, turbidity)
- Chl (F_{Chl}, decomposition of $a_\phi(\lambda)$, a_{LH}, c_p anomalous dispersion)
- Large Chl or POC (high-freq. spikes in F_{Chl} or b_{bp})
- Other pigments (decomposition of $a_\phi(\lambda)$)
- DOC (a_{CDOM}(blue or UV λ))
- CDOM composition (S_{CDOM})
- Particle size distribution (γ_{cp}, γ_{bbp})
- Particulate inorganic carbon (acid-labile b_{bp}, polarized c_p)
- Others?