

# Biogeochemical proxies

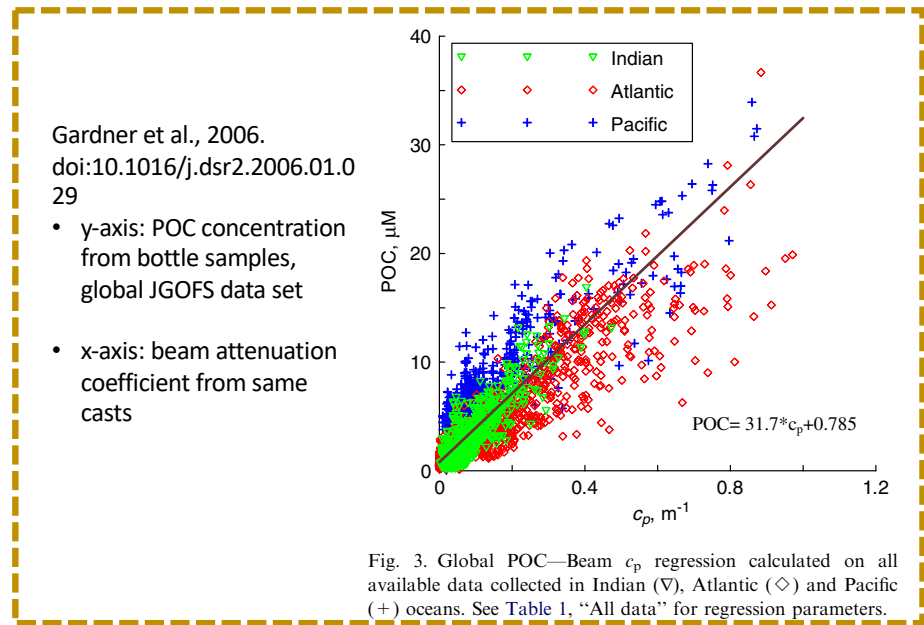
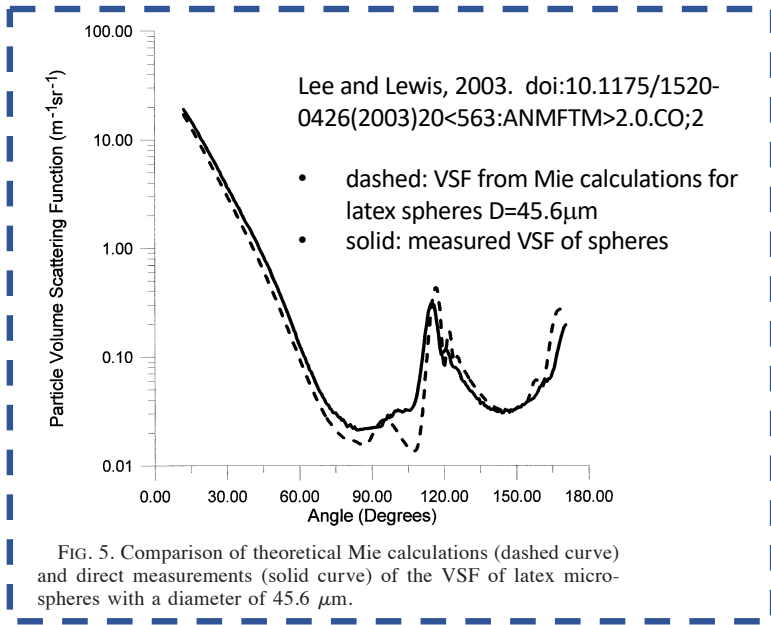
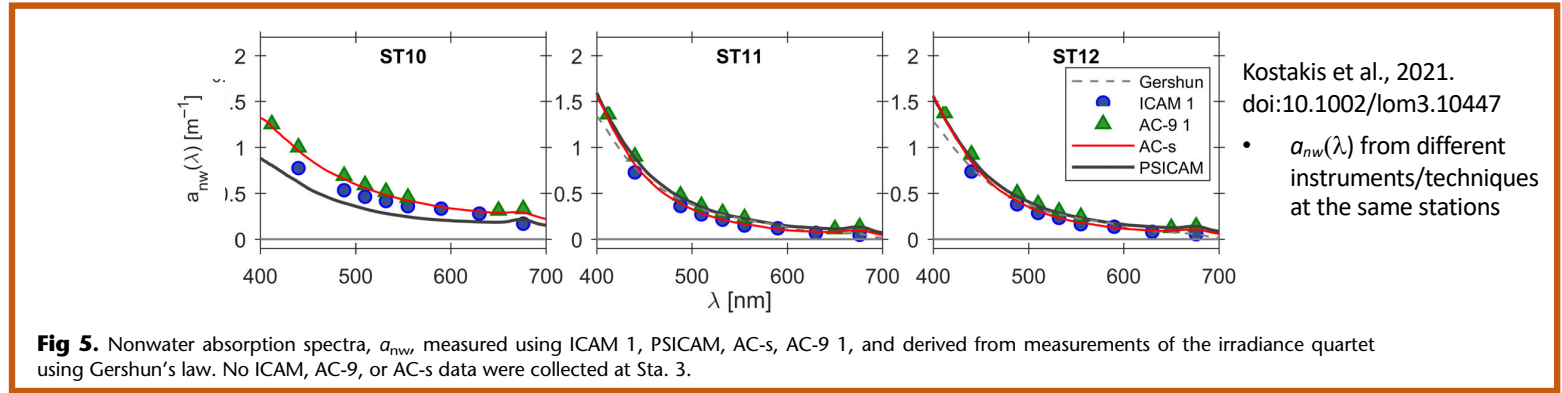
Meg Estapa

Ocean Optics class 2023

**Quick review:**

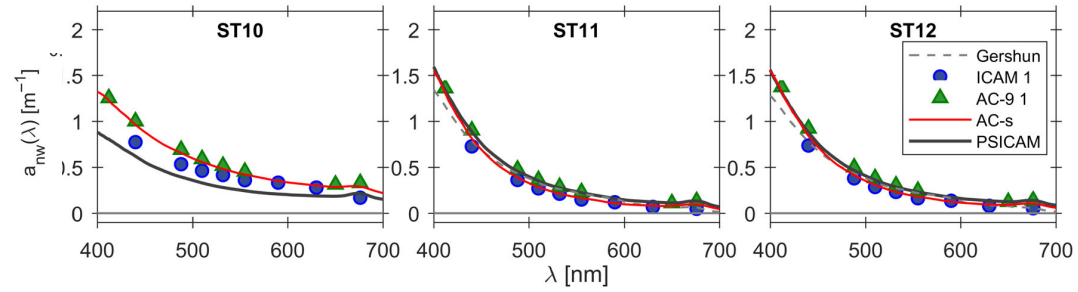
Which figure gives the best example of

- 1) proxy
- 2) calibration
- 3) validation



In the context of this class:

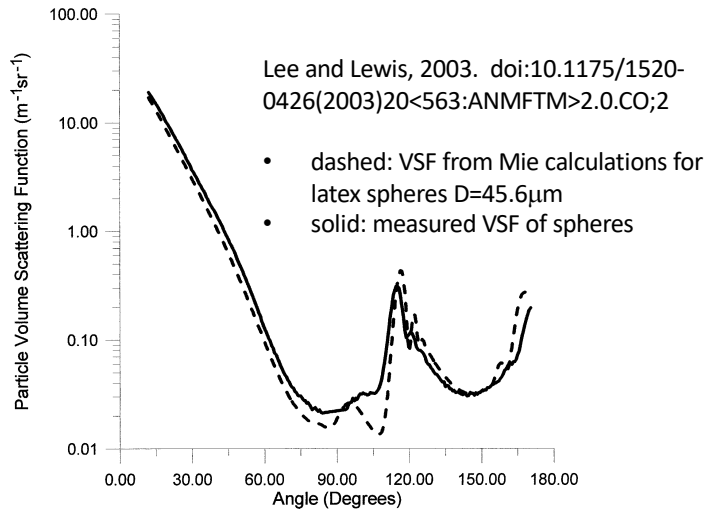
Proxy = use of an empirical relationship between two measurements to estimate one from the other.



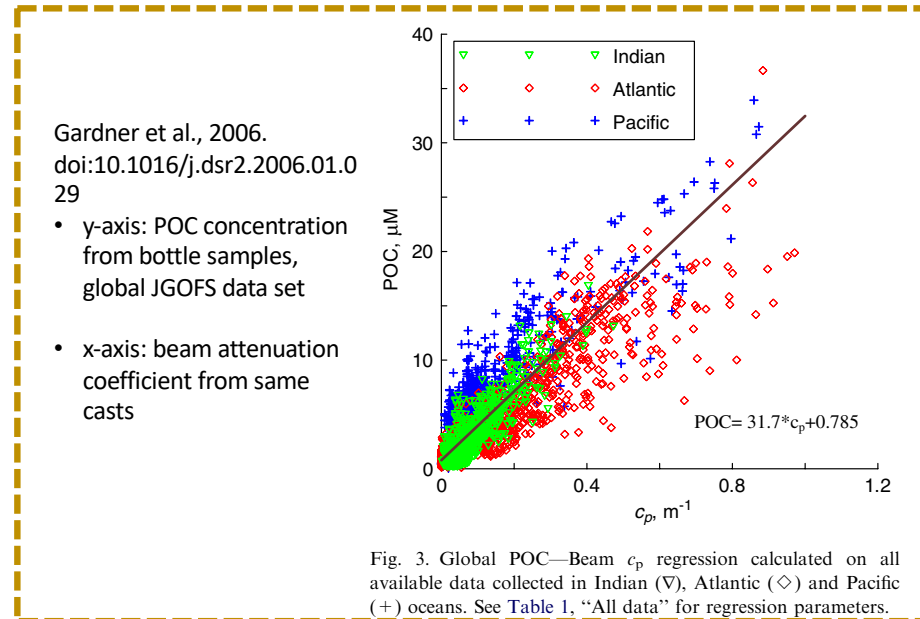
Kostakis et al., 2021.  
doi:10.1002/lom3.10447

- $a_{nw}(\lambda)$  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.



**Fig. 5.** Comparison of theoretical Mie calculations (dashed curve) and direct measurements (solid curve) of the VSF of latex microspheres with a diameter of  $45.6\ \mu\text{m}$ .



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

# Overview

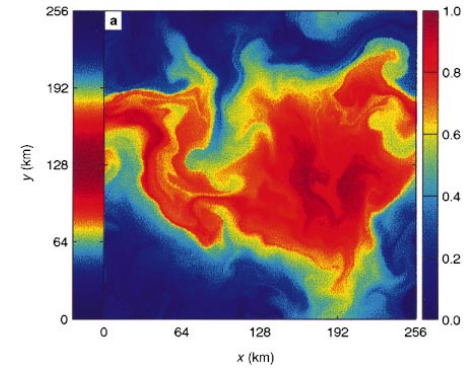
- ✓ Review of shared concepts across recent lectures
- Why create and use proxies?
- Some issues and cases to highlight
  - How many independent proxies can be extracted from observations?
  - Uncertainties in biogeochemical measurements
  - Restricted domain of the “training” dataset and extrapolation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

## Why create and use proxies?

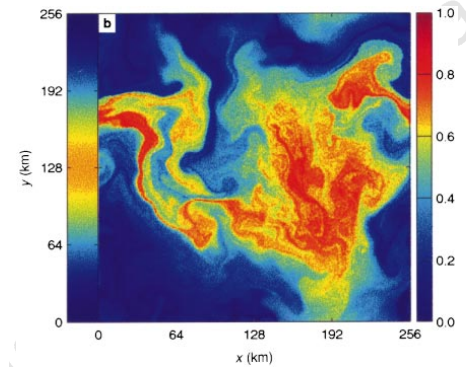
- Aquatic processes have different scales of variability and we need to match our observations to those scales
- Others?

What assumptions do we make when creating a proxy?

Carrying capacity  
( $F(\text{nutrients, light})$ )



Phytoplankton



Zooplankton

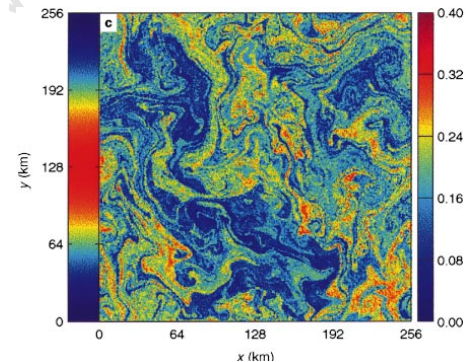


Figure: Abraham et al., 1998. *Nature* 391(6667): 577-580.

## 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.

The image displays four hand-drawn notes on proxy relationships in oceanography, each with various diagrams and equations.

**DOC (Dissolved Organic Carbon):** Importance: - Biological Pump, - Global Carbon Cycle, - Largest fraction of Organic C. Optical Proxy:  $a_g(440) [m^{-1}]$ . Instrument: Spectrophotometer (w/cuvette). Extra \$\$\$: Fluorometer: F vs. [DOC].

**POC (Particulate Organic Carbon):** Importance: FOR CARBON EXPORT, COMMUNITY METABOLISM, COMMUNITY COMPOSITION, Biological Carbon Pump - Particle Sinking. Optical Property: \* Scattering, \* Low relative backscatter, \* Absorption. MECHANISMS: Would use ACS for measuring the C and a determination identifying organic carbon based on pigment absorption. EXTRA \$\$\$: Hyper-BB.

**PHYTOPLANKTON BIOMASS:** → Total Carbon fixed in Oceans (↳ Biological Carbon Pump), → Total Energy available to Higher Trophic Levels (↳ Base of Food web). IMAGING: FLOW CYTOMETRY. MECHANISM: cells enumerated and imaged triggered by fluorescence. WITH MORE \$: FLOW CYTOMETER. MECHANISM: enumerated and identified using fluorescence.

**Bulk Particle Composition:** Importance: Dynamics of blooms + plumes, Phytoplankton, Biological carbon pumps. Optical Property:  $a_p(\lambda) = a_{chl} + a_{CDOM} + a_{scat} + a_{susp}$ . CDOM = Chlorophyll. EXTRA \$\$\$: Hyper-BB!

**PHYTOPLANKTON COMPOSITION:** Importance: Primary producers, Water quality (MSB), Food web dynamics, Climate & Carbon cycle, (Biological carbon pump). Optical Proxy: Absorption. Step 1: Total absorption spectra from ACS. Step 2: Correction for ACS filter factors (un-smoothing). Step 3: Spectral decomposition of pigments using e.g. PCA/Gaussian functions/neural networks. MECHANISMS FOR CO-VARIABILITY: Different phytoplankton have evolved unique pig. assemblages for optimizing light absorption resulting in different absorption spectra that can be used to characterize different phytoplankton populations. Additional proxy: IFCB → determine phytoplankton community composition.

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

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.

Absorption spectra have pretty similar shapes...

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

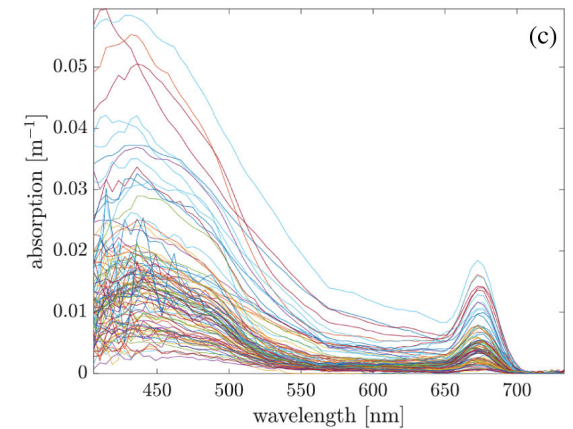
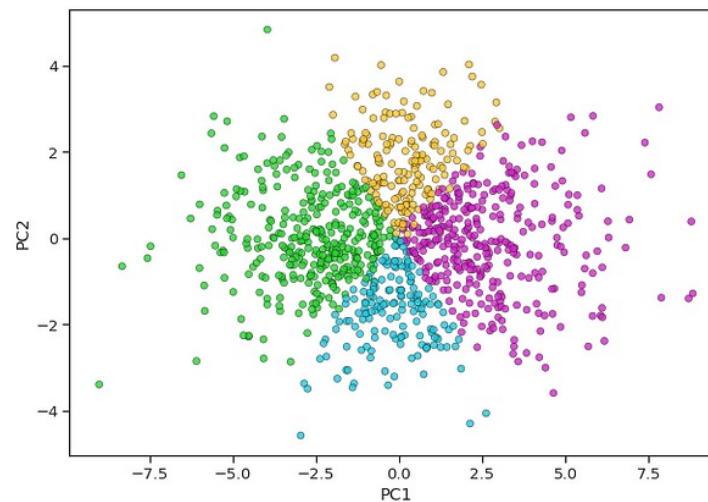
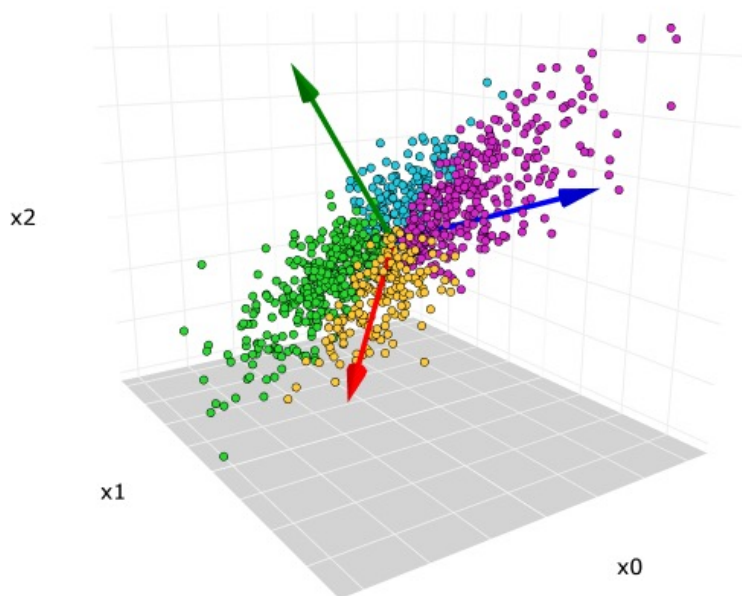


Figure: Cael et al. 2020, 10.1364/AO.389189



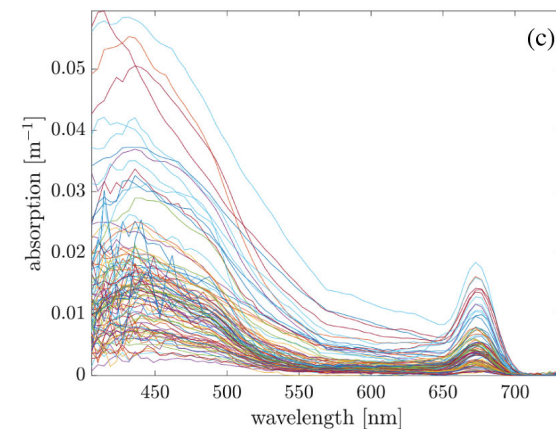
Figures: Cheng, C., 2022. <https://towardsdatascience.com/principal-component-analysis-pca-explained-visually-with-zero-math-1cbf392b9e7d>. Accessed 6/25/2023.

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

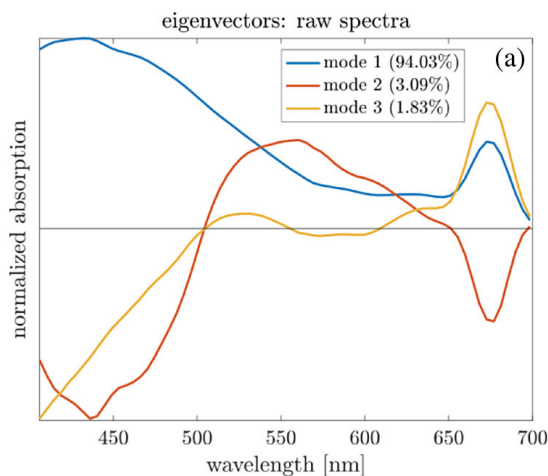
Absorption spectra have pretty similar shapes...

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

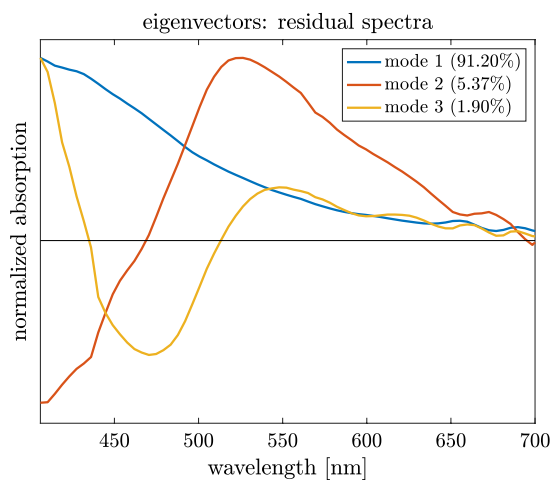
Principal component analysis (PCA): What are the basis vectors that sequentially describe the greatest amount of variance in the data?



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*



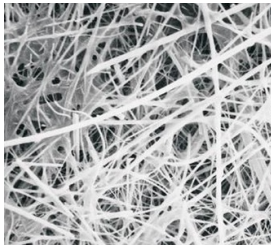
# Overview

- 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?

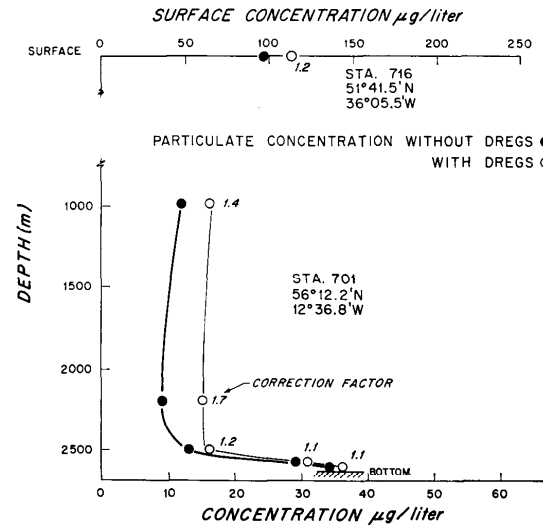
## Uncertainties in the “sea-truth” measurements used to build the proxy

### Example: Particulate organic carbon

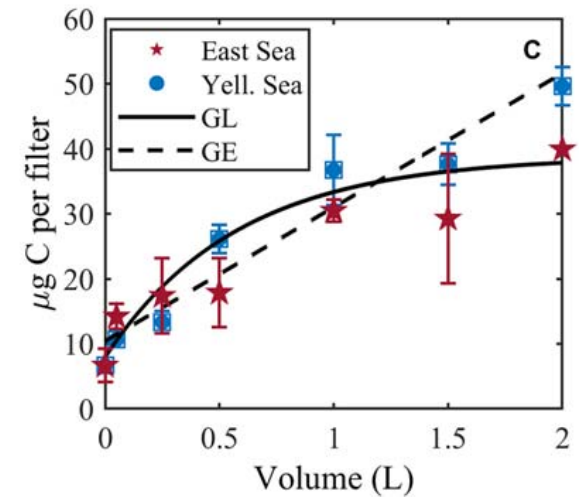
- “Dregs” in Niskin bottles
- Protection of sample from air contamination, handling contamination at all steps
- Low vacuum pressure
- Dissolved organic carbon blanks
  - DOC “adsorption sites” on GF/F filters
  - Adsorption is not instantaneous nor is it always linear with volume filtered



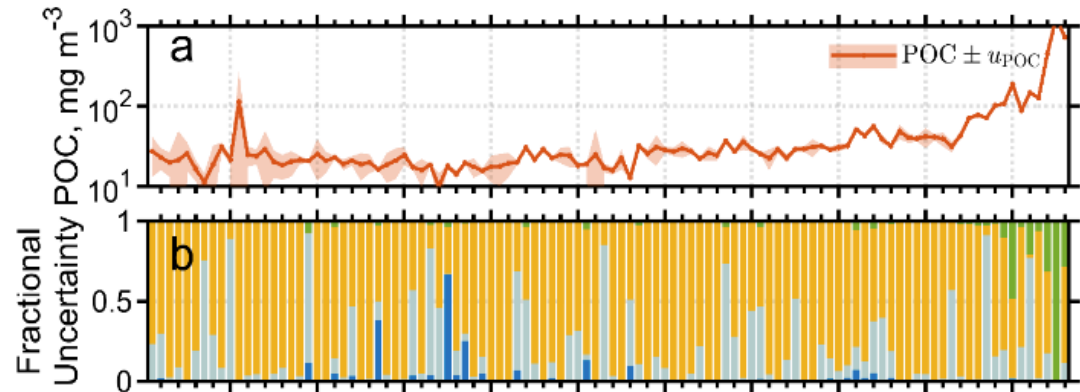
Chaves et al. 2021. 10.25607/OBP-1646



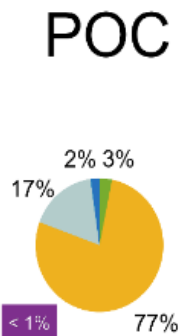
Gardner, 1977. 10.4319/lo.1977.22.4.0764



Novak et al. 2018. 10.1002/lom3.10248



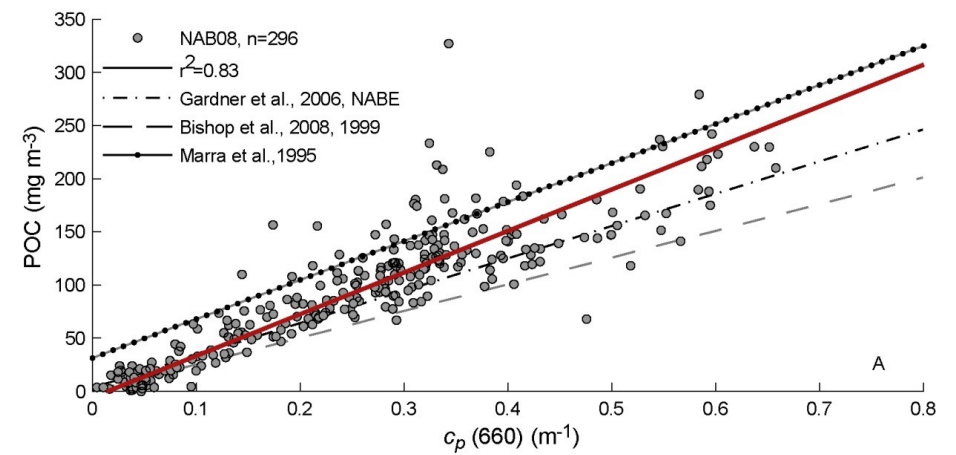
■ Sample packaging blank,  $M_{pk_f}^X$ 
■ Field filter blank,  $M_f^X$ 
■ Filtrate blank,  $M_{ads}^X$ 
■ Filtration volume,  $V$ 
■ Total unc. mass,  $M_T^X$



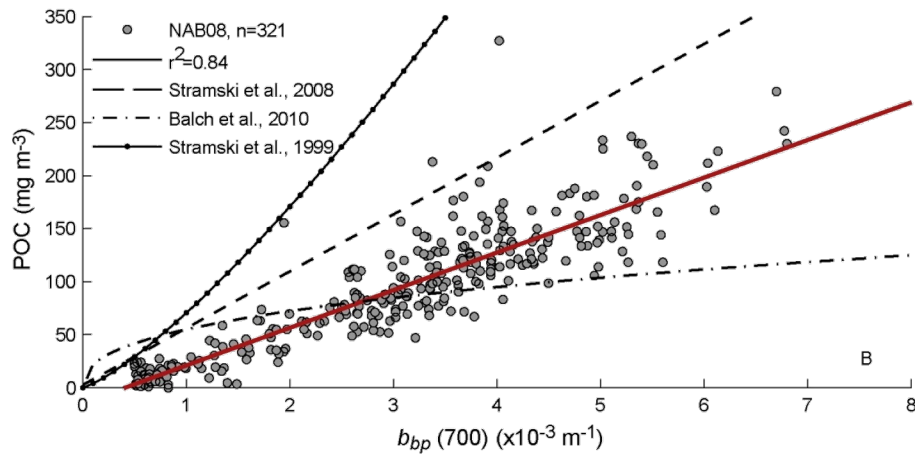
# Particulate Organic Carbon

Beam  
attenuation ( $c_p$ )

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

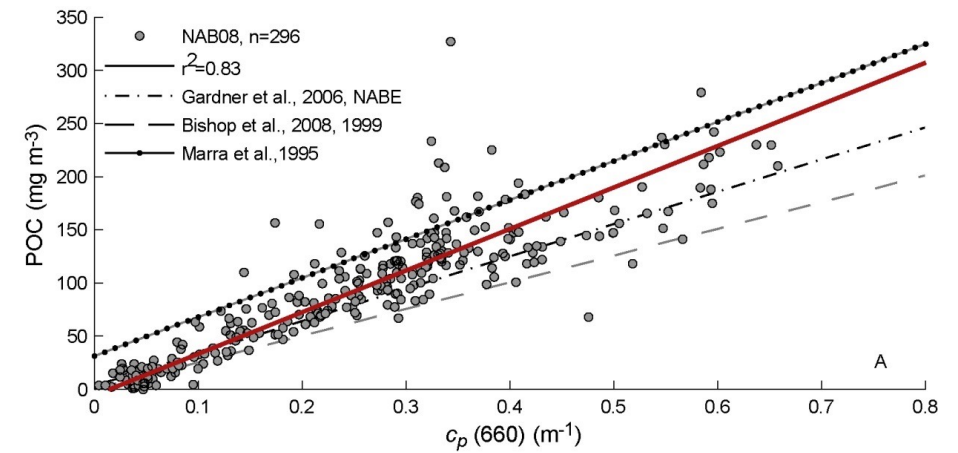


# Particulate Organic Carbon



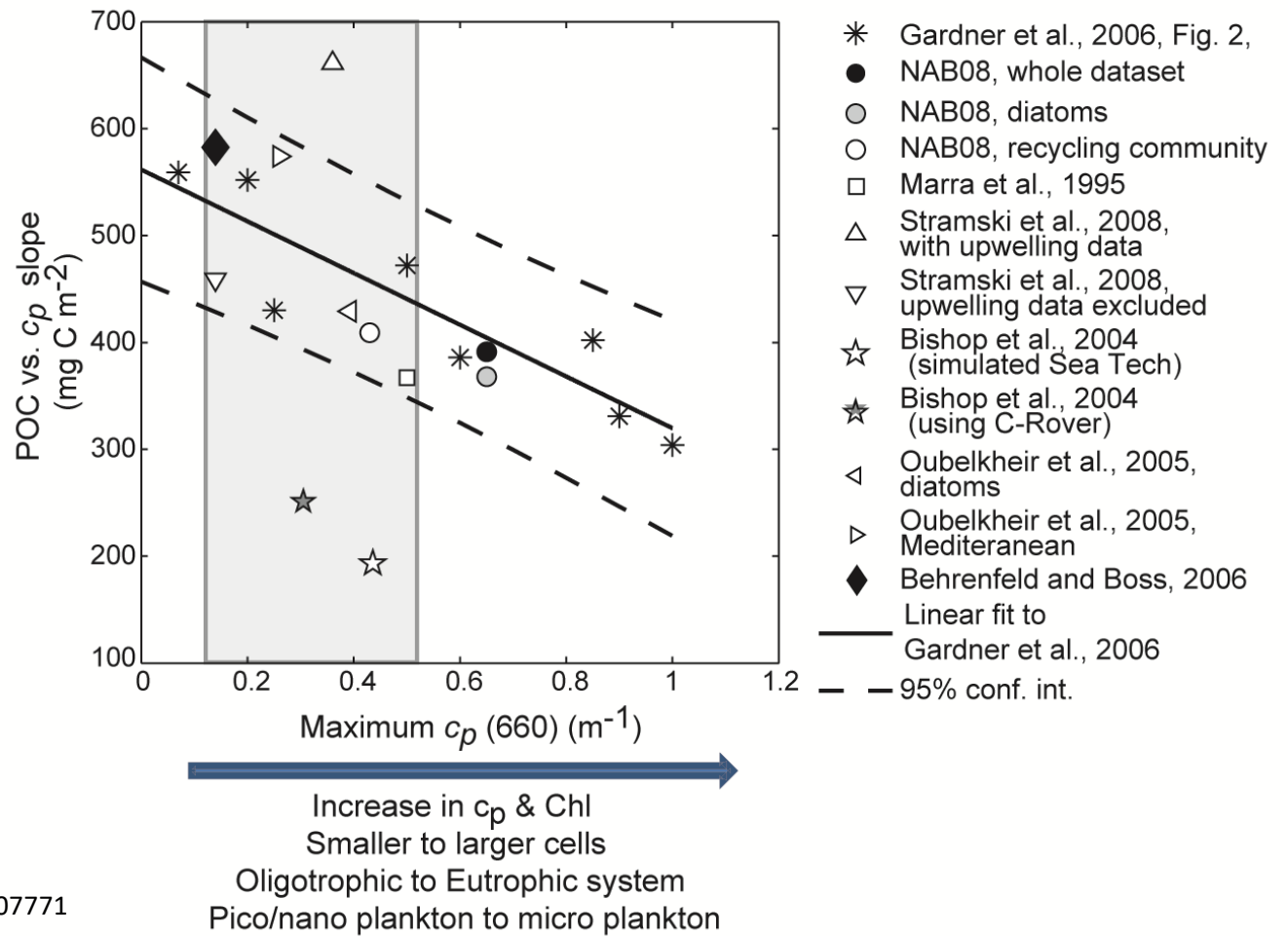
Particulate backscattering ( $b_{bp}$ )

Beam attenuation ( $c_p$ )



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

# POC/ $c_p$ slope comparison ( $\text{mg C m}^{-2}$ )



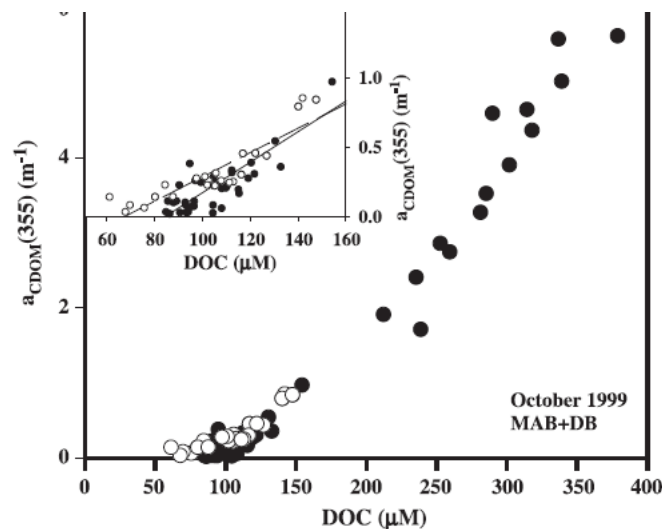
Figures: Cetinić et al., 2012, JGR. 10.1029/2011JC007771  
 Slide from I. Cetinić

# Overview

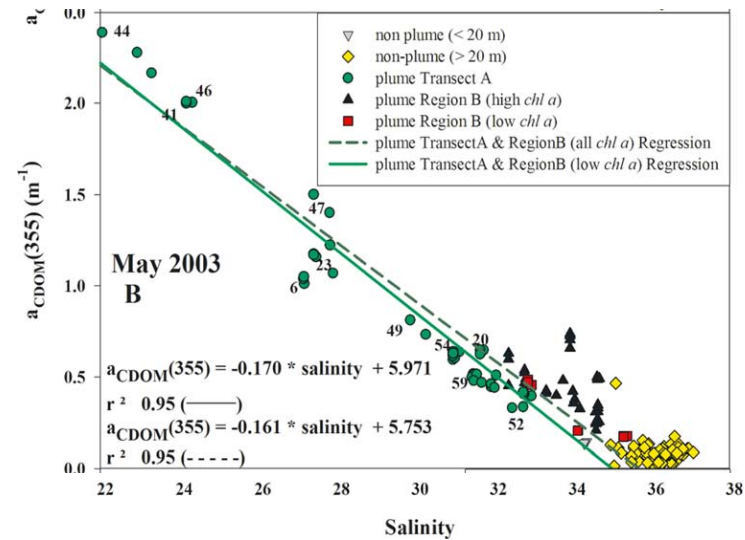
- Review of shared concepts across recent lectures
- Why create and use proxies?
- Some issues and cases to highlight
  - How many independent proxies can be extracted from observations?
  - Uncertainties in biogeochemical measurements
  - Restricted domain of the “training” dataset and extrapolation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

# Domain of training data: Dissolved Organic Carbon

## CDOM can approximate DOC in the coastal ocean



Del Vecchio and Blough, 2004



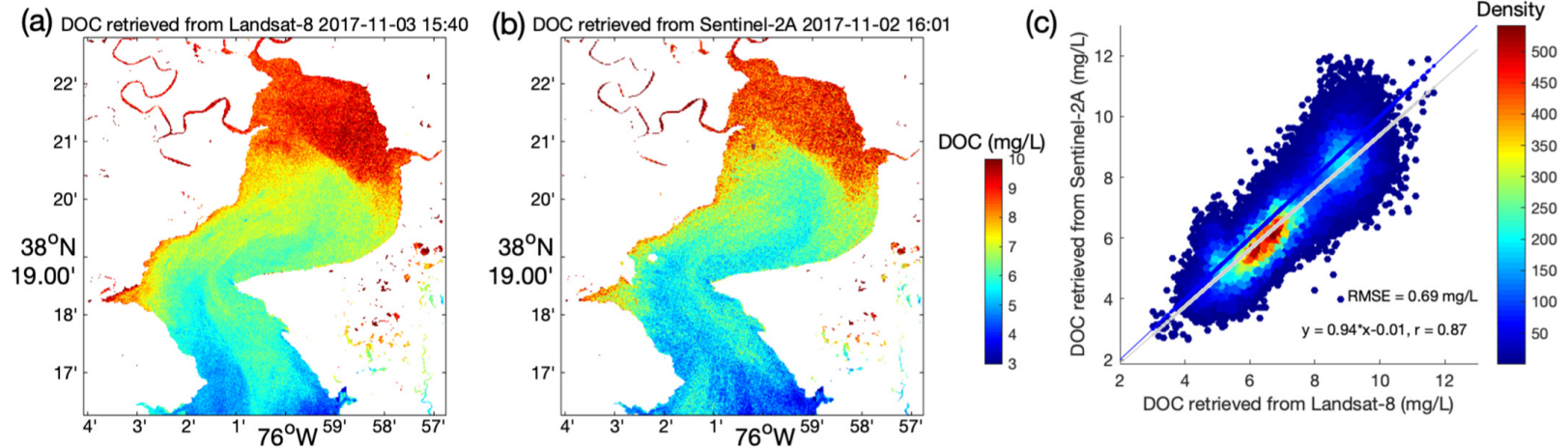
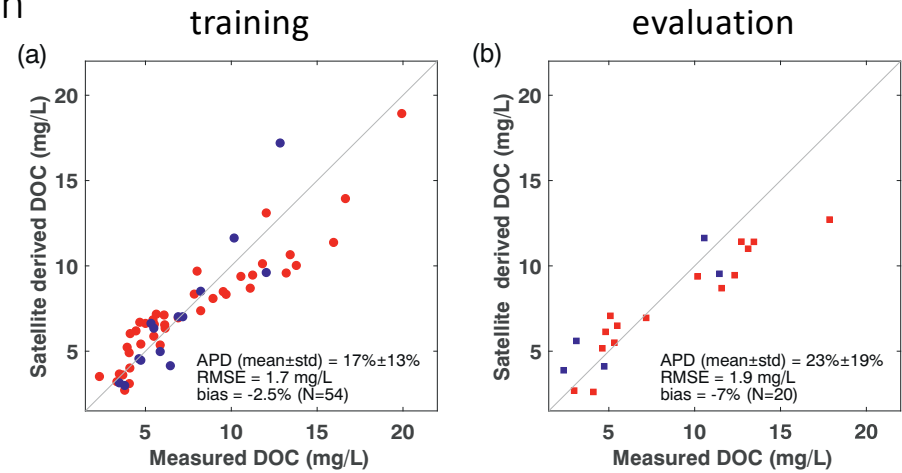
Del Vecchio and Subramaniam, 2004

- 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  $R_{rs}(\lambda)$

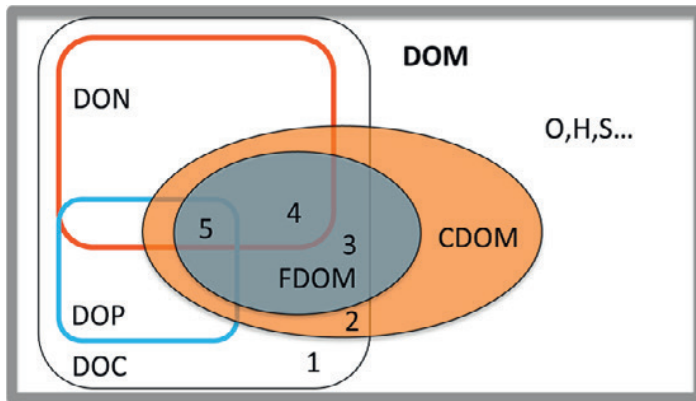
$$\text{DOC} = \exp(0.544 \times \log(R_{rs}(B1)) - 0.571 \times \log(R_{rs}(B2)) - 2.181 \times \log(R_{rs}(B3)) + 1.398 \times \log(R_{rs}(B4)) - 1.406)$$



**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.



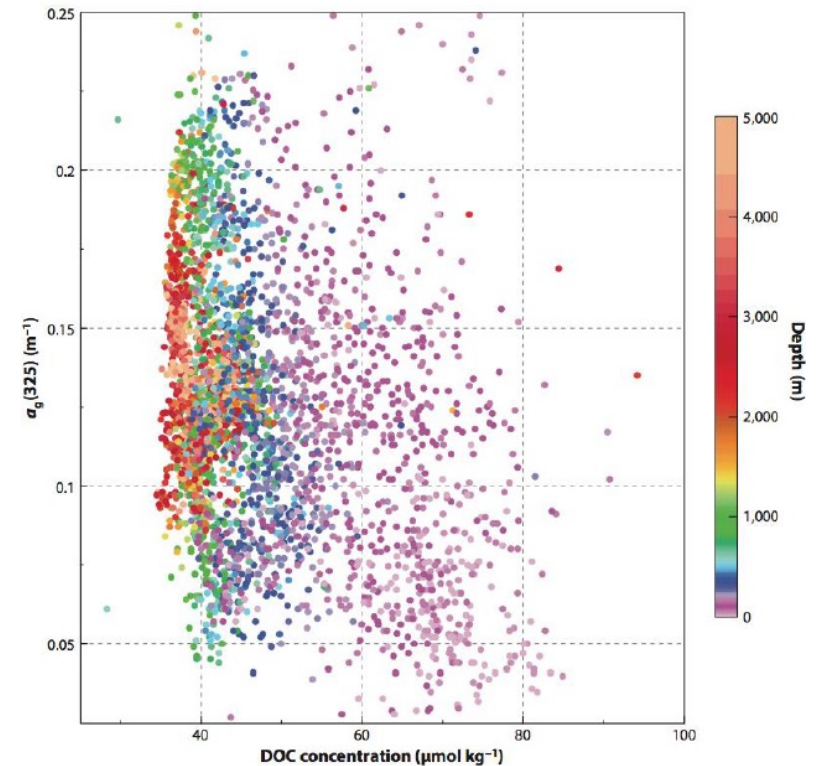
Domain of training data: Dissolved Organic Carbon



- 1 = No color, N and P absent (e.g., carbohydrates)
- 2 = Colored but N and P absent (e.g., carotenoids)
- 3 = Fluorescent but N and P absent (e.g., vanillic acid)
- 4 = Fluorescent and contains N (e.g., tyrosine)
- 5 = Fluorescent and contains both N and P (e.g., NADH)

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

Perhaps unsurprisingly, CDOM  $\neq$  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  $\mu\text{mol L}^{-1}$

# 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.

**SPECTROPHOTOMETER:  $a(\lambda)$**

**Description:**

- Wavelength: 300-850 nm
- Bandwidth: 2 nm
- Beam mode: double
- Pathlength: 1-10 cm
- No scattering effects

**Limits:**

- Discrete samples
- Careful use: fingerprints, condensation...

**Goal:** Measure absorption ( $A$ )

**Proxies:** From absorption, water components can be estimated: NAP, phytoplankton, CDOM, lithogenic components

- Ph (acidification of the oceans)
- Salinity, temperature

**VSF (BCo)**

**Laser In-Situ Scattering and Transmissometry**

**How it works:**

- scattering from sample @ angle  $\psi$  to radius
- On focal plane detector
- near-forward scattering  $15^\circ$  to  $90^\circ \rightarrow$  VSF

**Assumptions:**

- The theory based on homogeneous spheres
- smaller particles  $\rightarrow$   $b_b \uparrow$
- Particle size  $\parallel$
- $a \propto D^3$

**Limitations:**

- Flow obstruction due to large size
- Laser mount sensitivity to turbulence
- PSD (particle size distribution) for  $> 2 \mu m$

**Get a LI**

- for PSD through vertical column  $\rightarrow$  carbon export
- easy integration to existing biological & biogeochemical measurements (validation...)

**$b_b(660)$**

**HOW IT WORKS:** EMITS LIGHT @ 660nm w/  $\theta 120^\circ$

**PROXY:** Estimates Particle Size Dist. (PSD)

Exposure Flux  $\rightarrow$  Turbidity, Resuspension Events  $\rightarrow$  POC

Composition  $\rightarrow$  Bloom Dynamics  $\rightarrow$  Food Web Dynamics

**Limitations:**

- Low dependence on CDOM
- Higher signal w/ higher index of refraction

**Limitation:** Single channel @ red band wavelength

**Assumption:** Sullivan & Twardowski 2007 measurement of VSF @ one angle to est.  $S_{bp}$

**Extra \$\$\$: Fluorometer:  $F$  vs.  $[DOC]$**

**ACS**

**Goal:**  $a(\lambda)$ : collect all transmitted scattered light, reflected coating

- collects up to  $\sim 40^\circ$
- Limitation - not all scattered light is collected
- only transmitted light
- Limitation - Finite acceptance angle = scattering error

**PROXY MEASUREMENTS:**

- From a CDOM = DOC, SPM
- From  $c_p$ : chl in open ocean
- PSD: using power law slope distribution at  $c_p$
- With corrections for Temp and salinity

**Depends on:**

- $a_p$ : size, shape, refractive index, refractive composition
- $c_p$ : size, shape, real & imaginary index refr. composition
- $b = c - a$
- a CDOM: concentration, imaginary index refraction

**C-Star-c(660)**

**HOW IT WORKS:** takes a measurement of the attenuation at 660 nm.

$c = -\frac{1}{L} \ln\left(\frac{\Phi_t}{\Phi_0}\right)$

$L = 10$  or  $25$  cm

**IMPORTANCE:** attenuation is proportional to dissolved and particulate matter concentration. HAS a long history of measurements.

**LIMITATIONS:** pathlength limits range of  $C$  if good signal

- single  $\lambda$  only targeting some particle types (mostly miss color)
- only variable is attenuation
- limited by the acceptance angle

**PROXY FOR:**

- Water clarity
- particulate distribution (in space or vertical profile)
- vertical particulate processes

depth  $\rightarrow$  phyto, SHELF PROC, resuspension

## 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  $\lambda$ ))
- CDOM composition ( $S_{CDOM}$ )
- Particle size distribution ( $\gamma_{cp}$ ,  $\gamma_{bbp}$ )
- Particulate inorganic carbon (acid-labile  $b_{bp}$ , polarized  $c_p$ )
- Others?