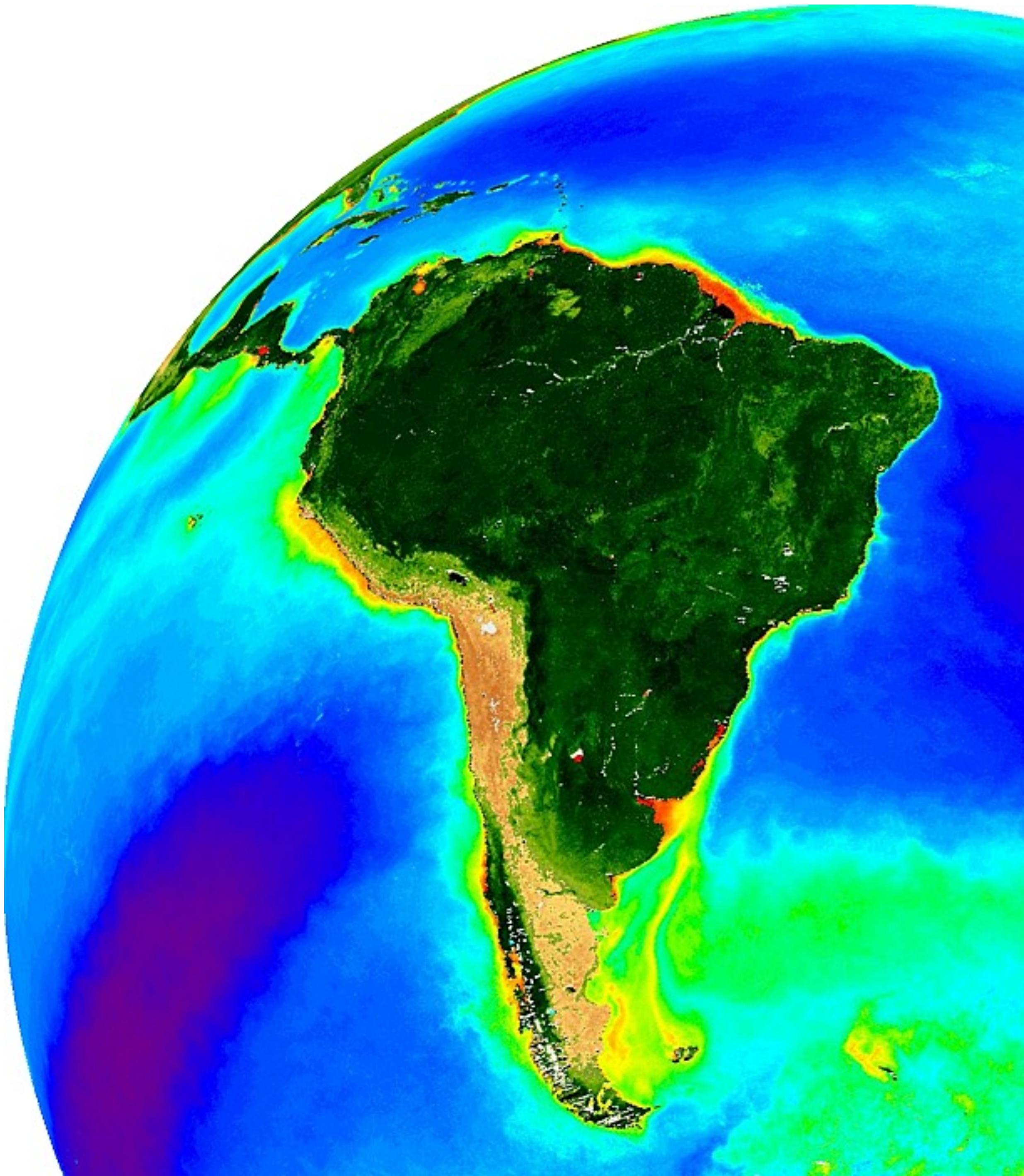
Your *in situ* data and you

Alternative title: *How* in situ *data are used informs on how a remote sensing product should be interpreted*

Jeremy Werdell
NASA Goddard Space Flight Center

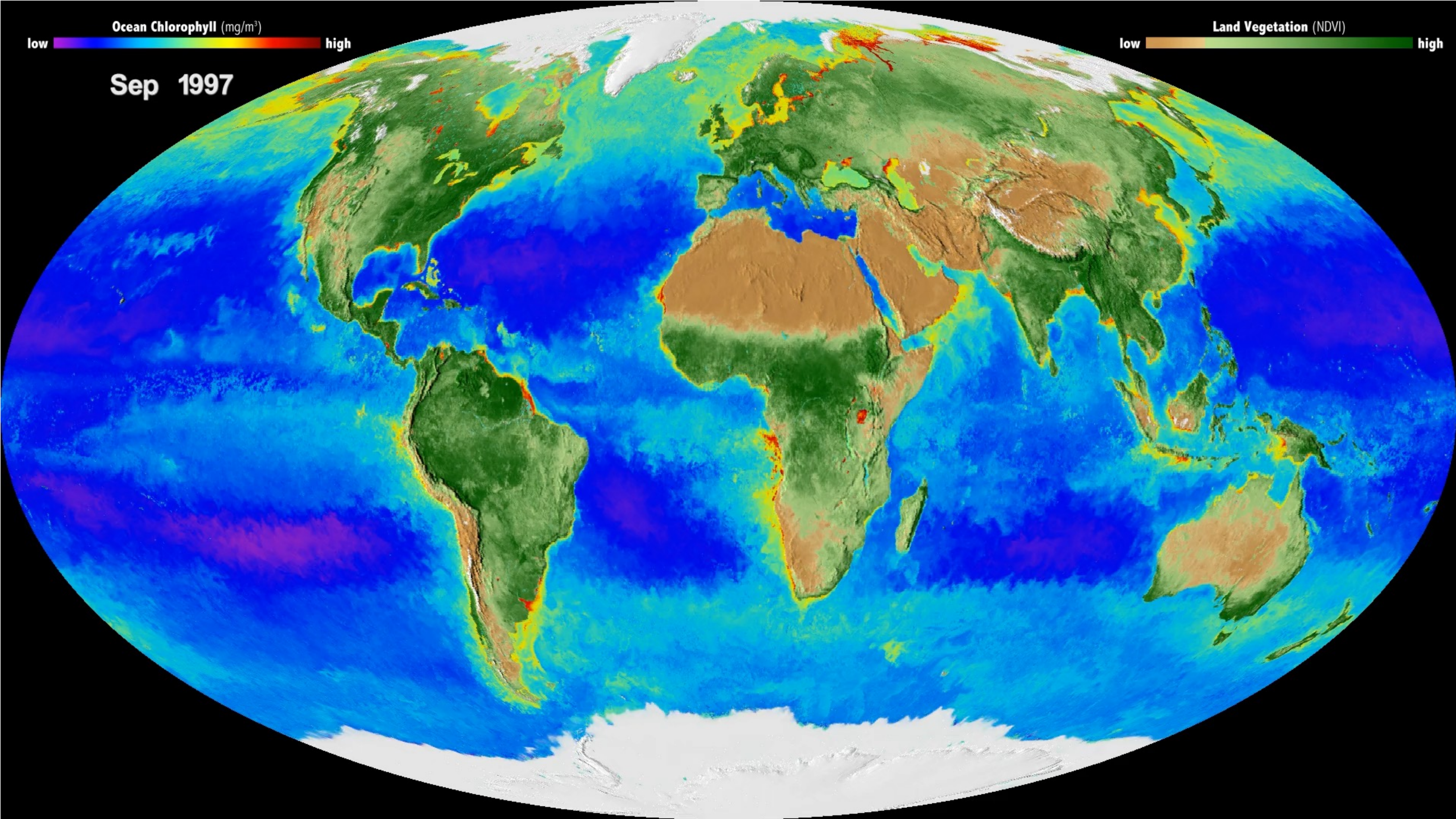
2024 IOCCG CETT training course



sharing experiences with *in situ* and satellite data with the goals of you …

leaving with an appreciation for the interconnectedness of *in situ* and remote sensing data (*in situ* data are pervasive!)

keeping this interconnectedness in mind when interpreting your results


Ocean Chlorophyll (mg/m³)
low
high
Land Vegetation (NDVI)
low
high
Sep 1997

In situ data inform empirical relationships

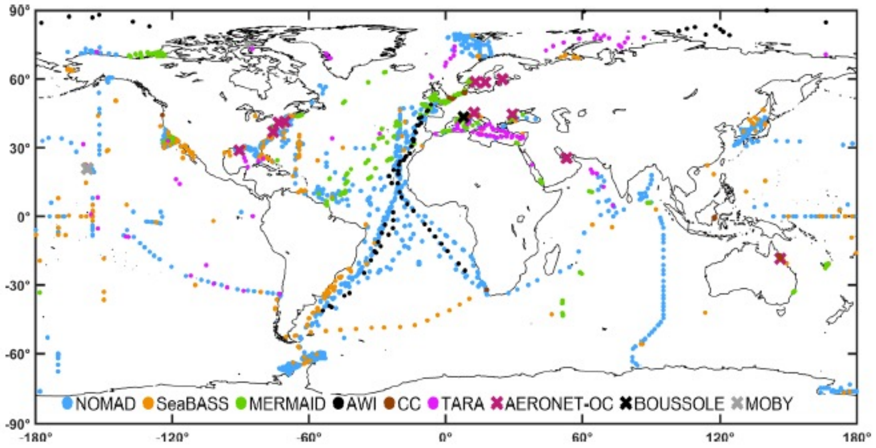
In situ data inform semi-analytic retrievals empirical (adjective): based on, concerned with, or **verifiable by observation or experience rather than theory or pure logic**

Rrs maximum band ratio

$X = \log_{10}[\ Rrs(443 > 490 > 510)\ /\ Rrs(555)\ ]$

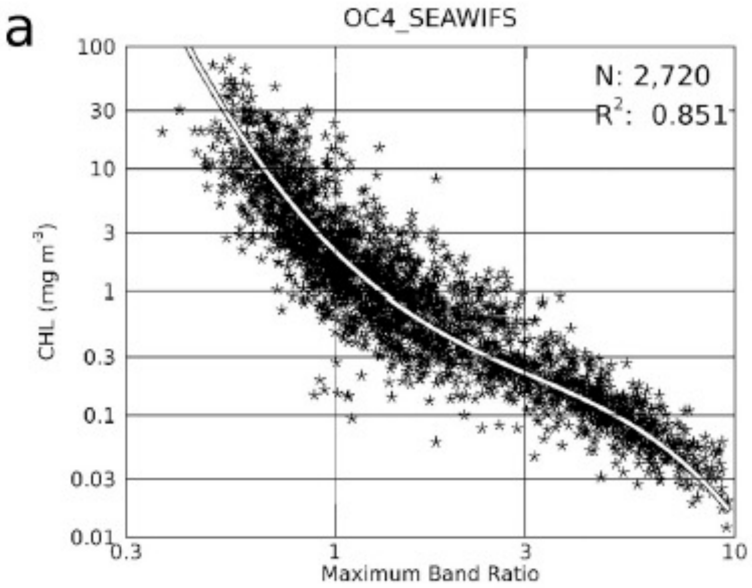
$\log_{10}(chl) = a_0 + a_1X + a_2X^2 + a_3X^3 + a_4X^4$

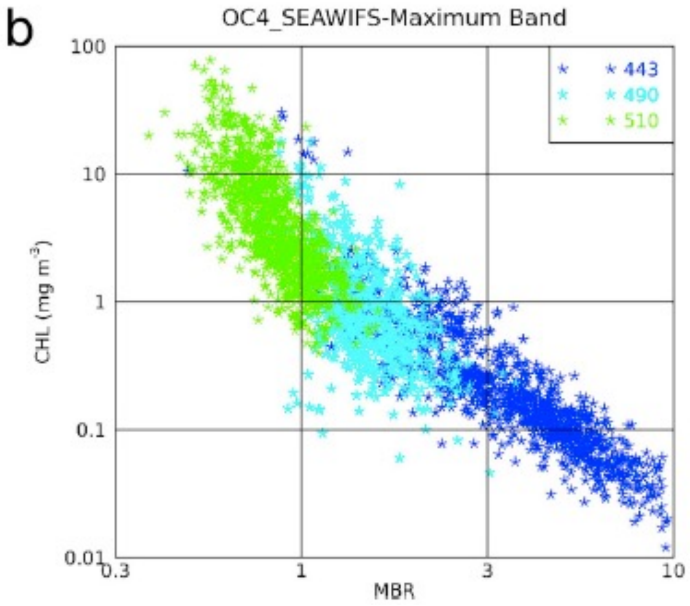
O'Reilly & Werdell 2019, Rem. Sens. Environ. [after O'Reilly et al. 1998, J. Geophys. Res.; O'Reilly et al. 2002, NASA TM, Werdell 2005, NASA OceanColor Web, others]

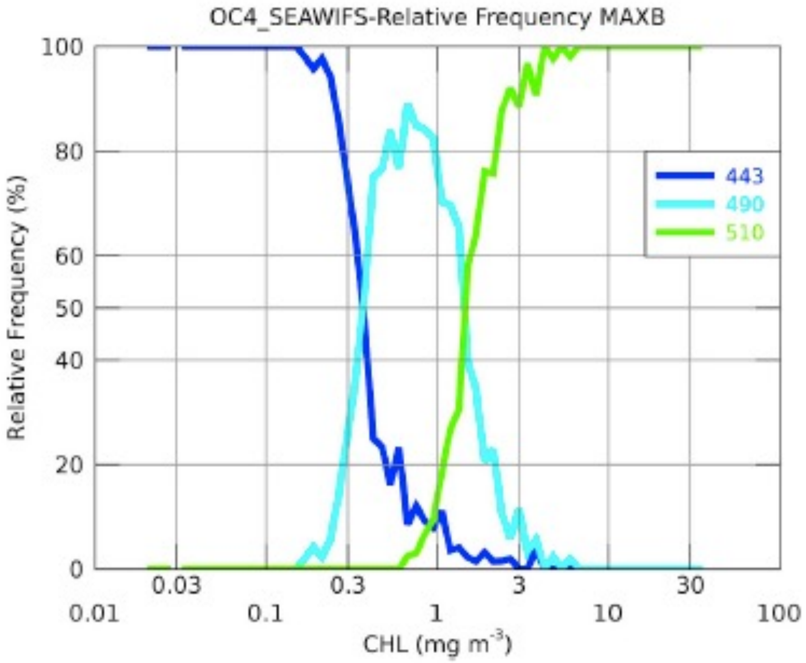


Valente et al. 2022, Earth Syst. Sci. Data

developed using a "global" dataset of in situ Rrs(λ) and chl

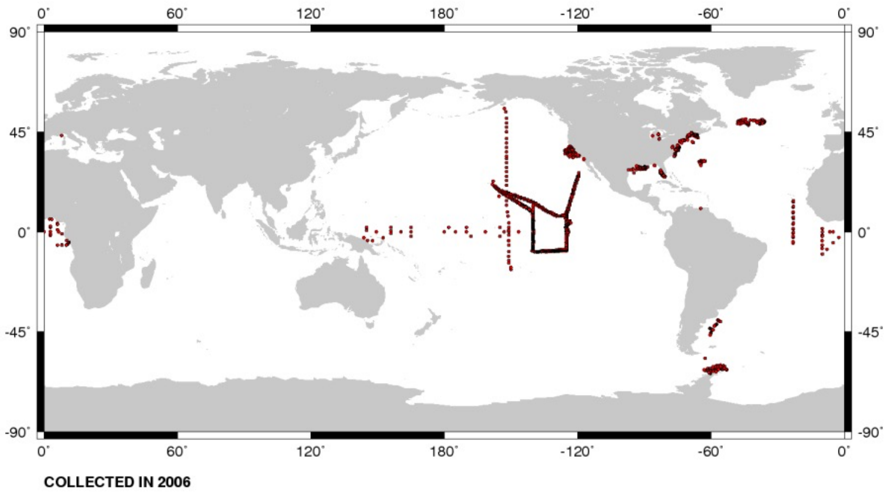




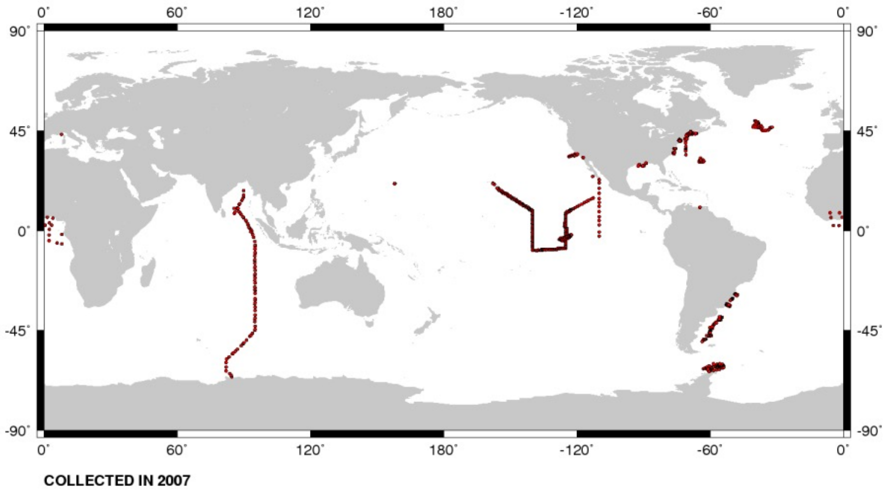


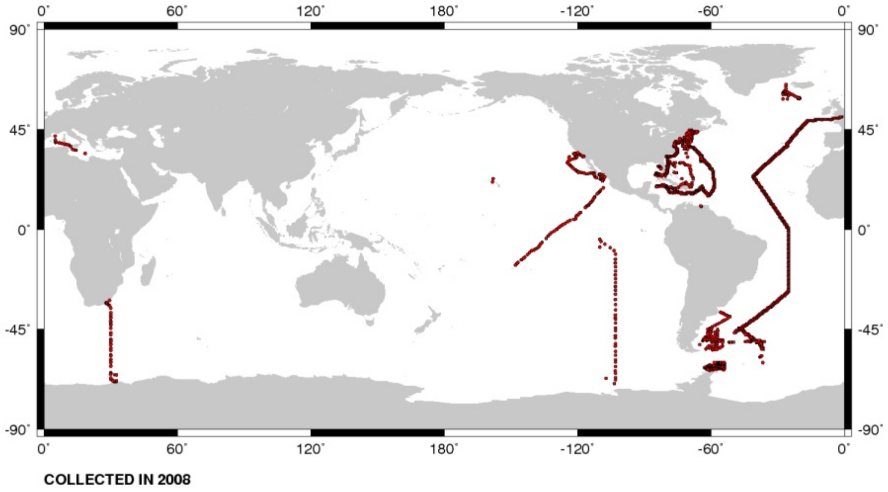
SeaBASS (NASA *in situ* archive) holdings by year: 2006-2009

2006

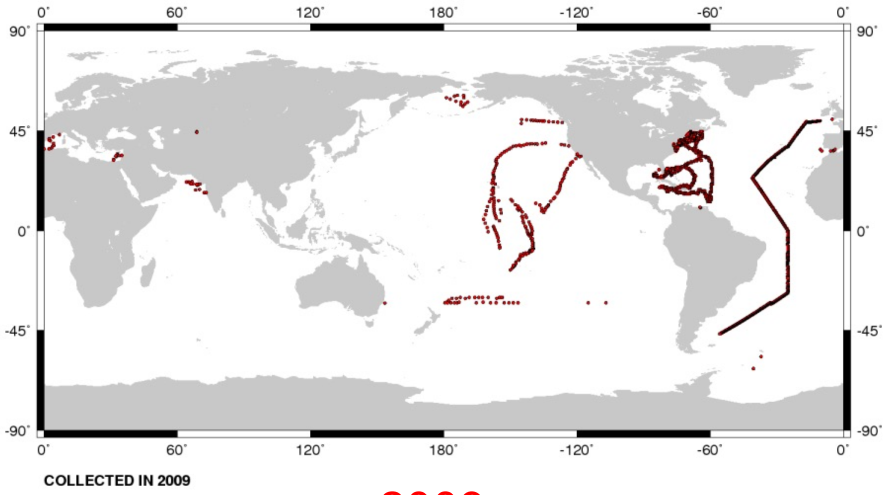


2007

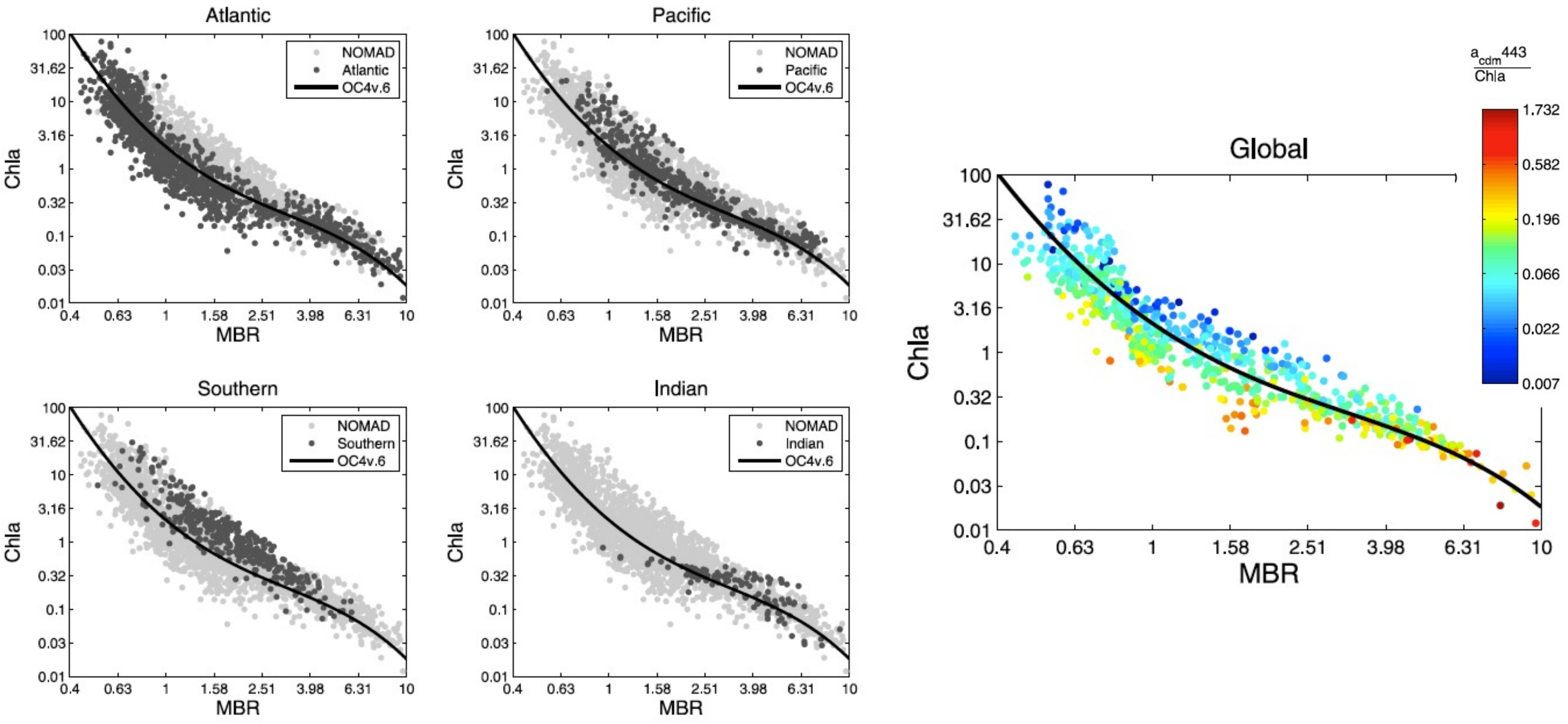




2008



2009

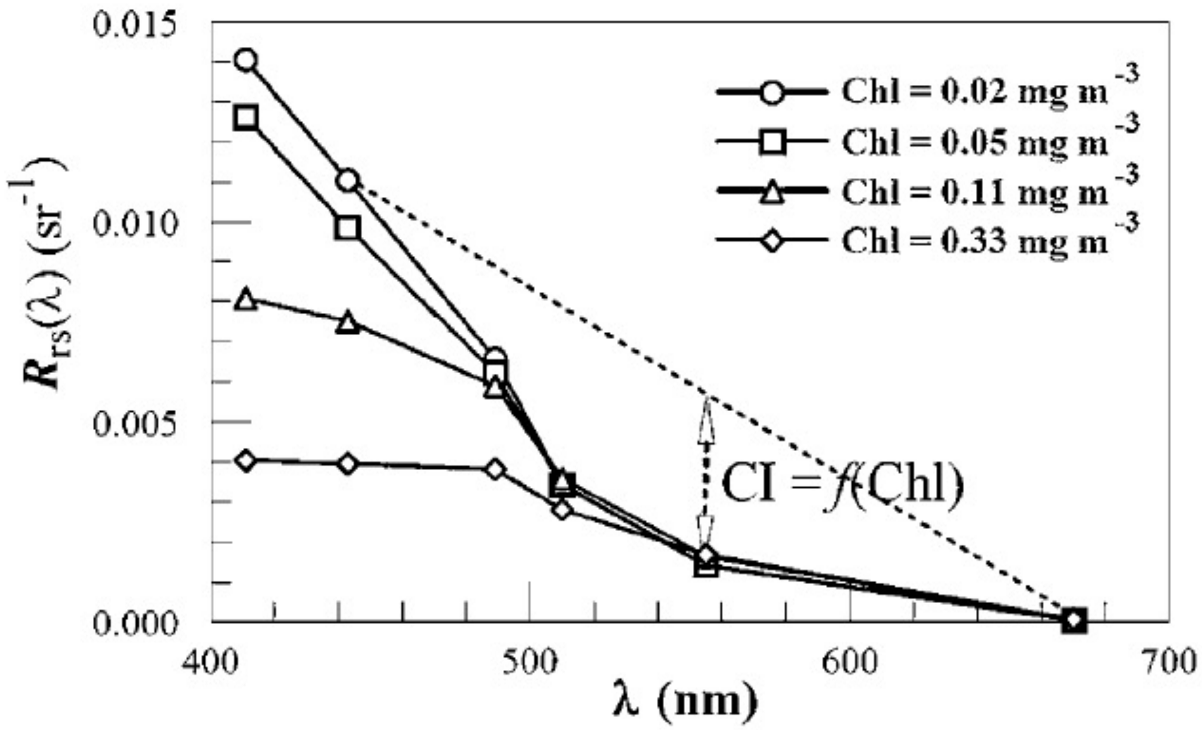


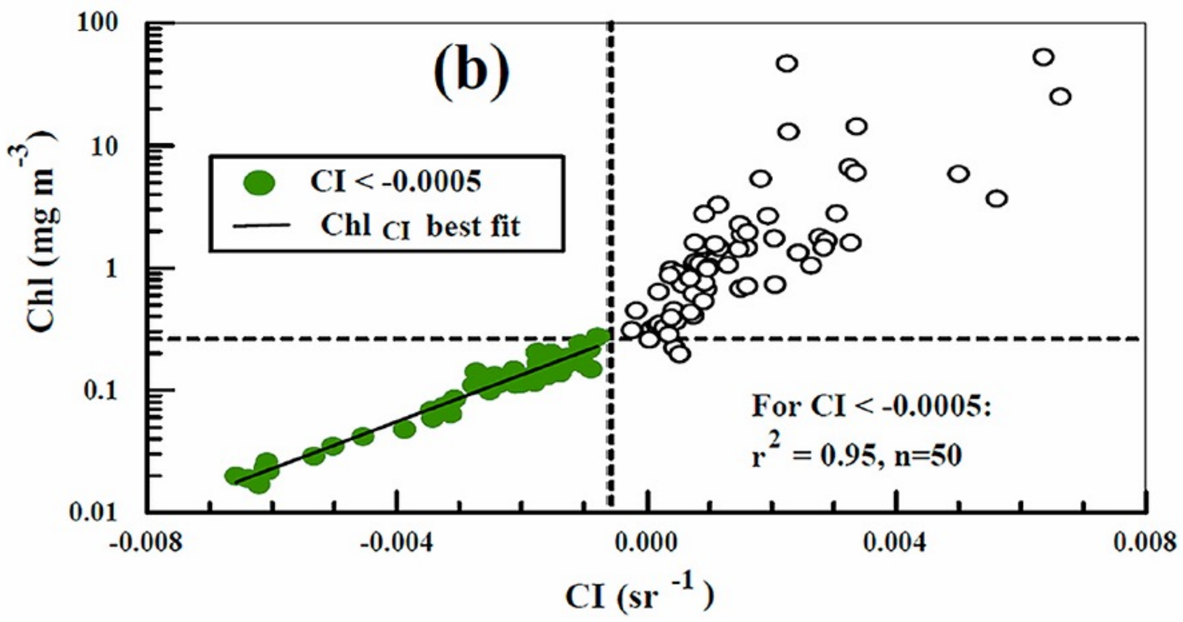
Szeto et al. 2011, J. Geophys. Res.

Rrs line height (baseline subtraction)

Chlorophyll Index (CI) from Hu et al. 2012, J. Geophys. Res. and Hu et al. 2019, J. Geophys. Res.

$$CI = R_{rs,555} - \left[R_{rs,443} + (555-443)/(670-443)\,x\,\left(R_{rs,670} - R_{rs,443}\right)\right]$$



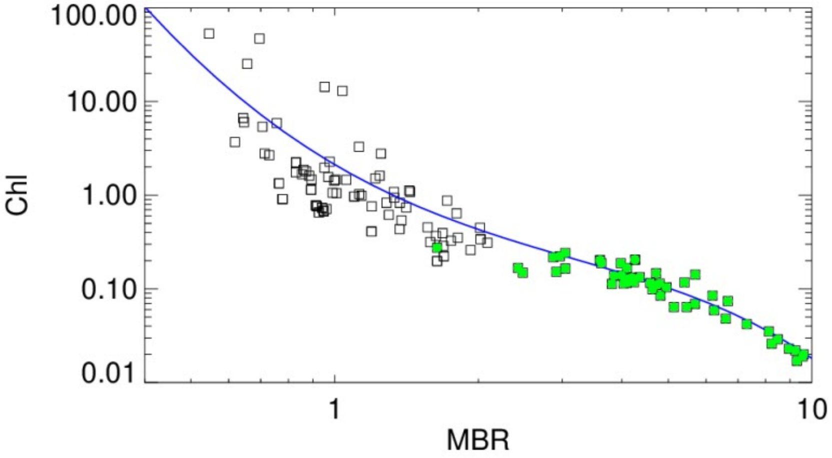


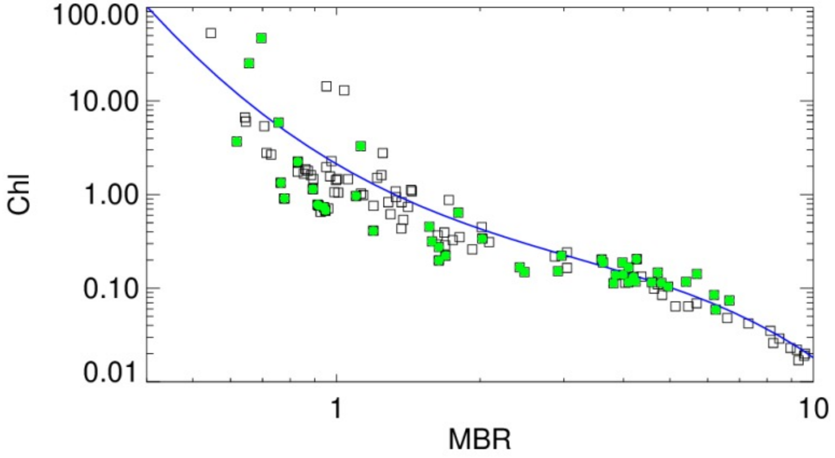
$$Log_{10}(Chl) = a\,CI + b$$

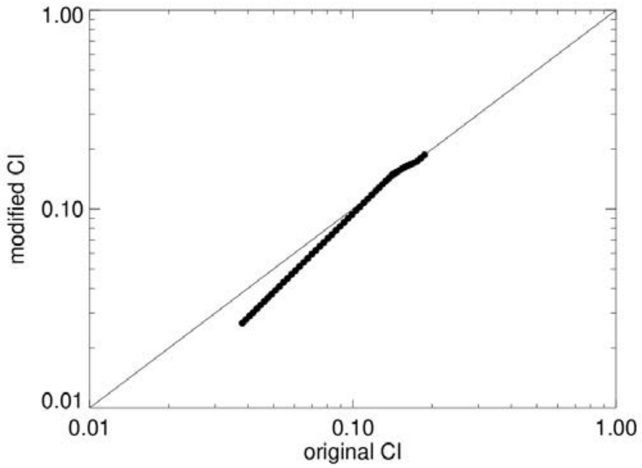
original

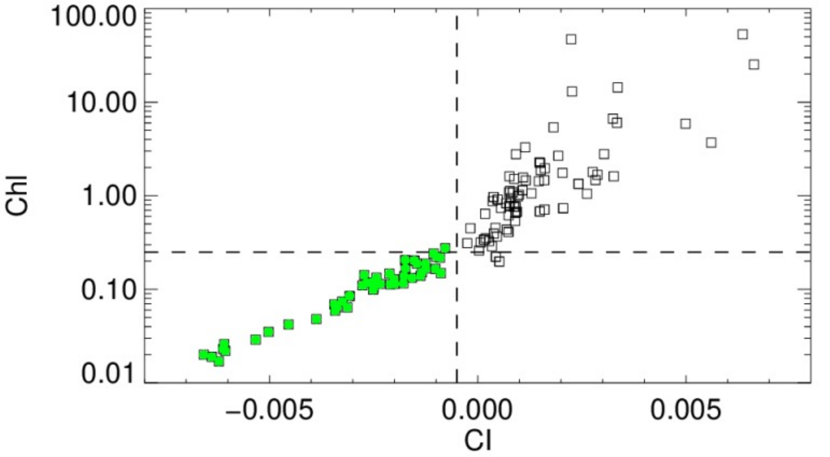
modified

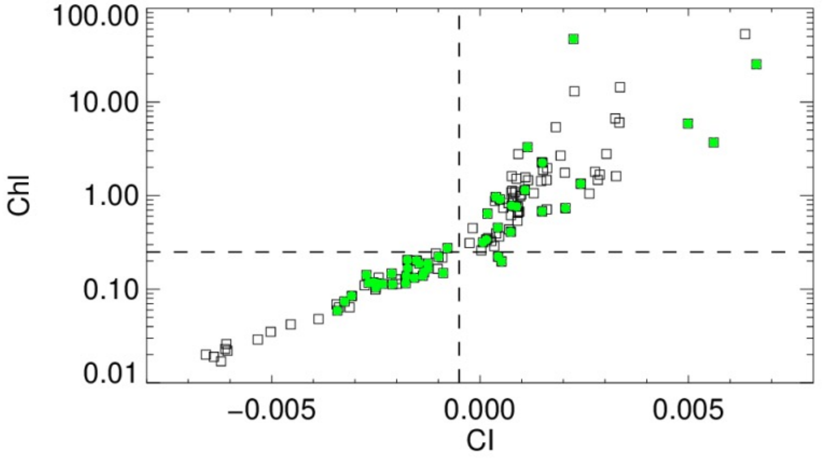
difference

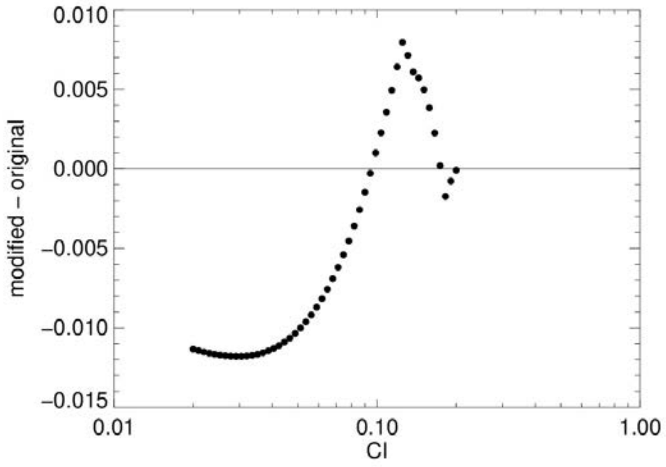












inverse (adjective): opposite or contrary in position, **direction**, **order**, or effect

one sentence summary of this inversion paradigm: How much of each absorbing and backscattering component is needed (in a least squares sense) to reconstruct the measured reflectance spectrum?

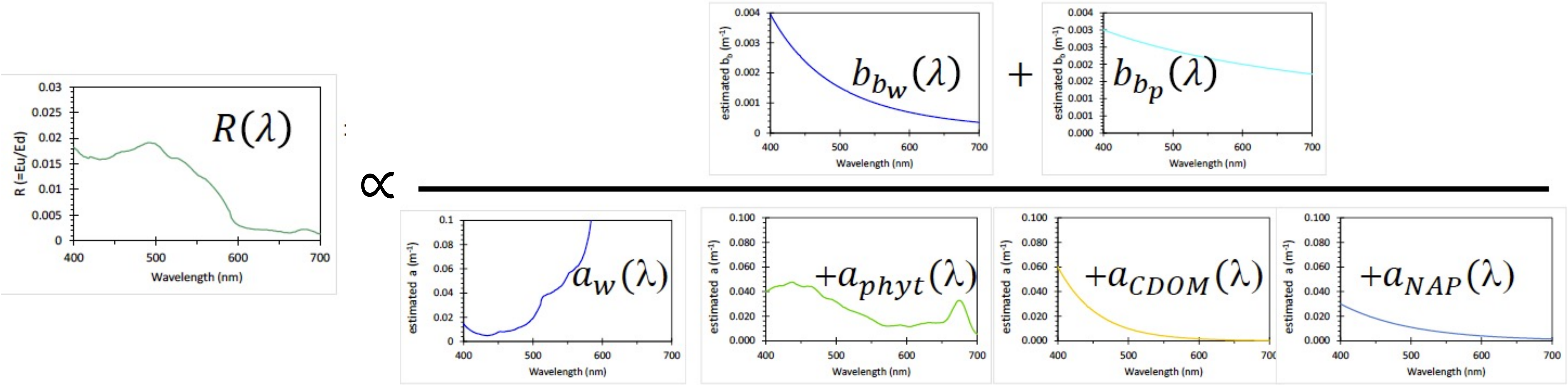
Rrs(λ) ⟵ *forward model* ⟵ IOP(λ)[chl, whatever]

$$r_{rs}(\lambda) = G(\lambda) \frac{b_{bw}(\lambda) + B_{bp} b_{bp}^{*}(\lambda)}{a_w(\lambda) + A_{ph} a_{ph}^{*}(\lambda) + +A_{dg} a_{dg}^{*}(\lambda)}$$

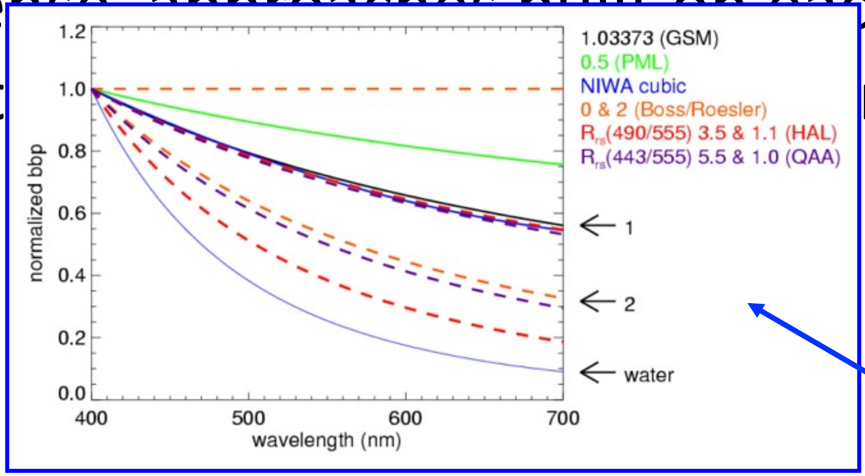
Rrs(λ) ⟶ *inverse model* ⟶ IOP(λ), chl, whatever

one sentence summary of this inversion paradigm: How much of each absorbing and backscattering component is needed (in a least squares sense) to reconstruct the measured reflectance spectrum?

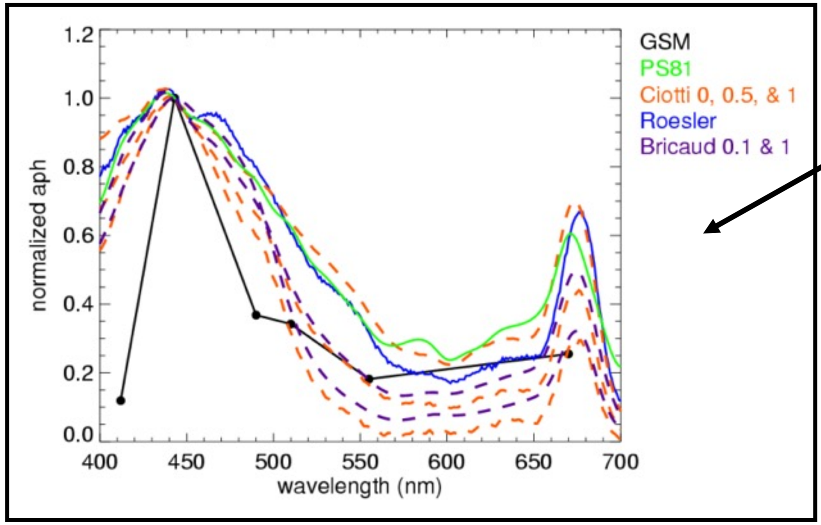
… can reconstruct this? what combination of these …

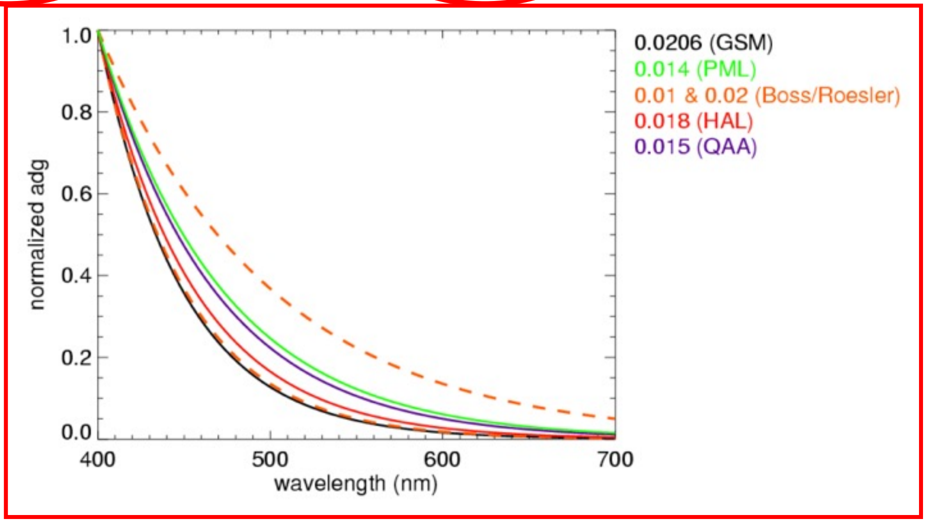


In the broadest se … ch other or experienced convergent evolut … n common and key places …



$$r_{rs}(\lambda) = G(\lambda) \frac{b_{bw}(\lambda) + B_{bp} b^*_{bp}(\lambda)}{a_w(\lambda) + A_{ph} a^*_{ph}(\lambda) + + A_{dg} a^*_{dg}(\lambda)}$$





you live in a consumer's market!

$$r_{rs}(\lambda) = G(\lambda) \frac{b_{bw}(\lambda) + B_{bp} b^*_{bp}(\lambda)}{a_w(\lambda) + A_{ph} a^*_{ph}(\lambda) + +A_{dg} a^*_{dg}(\lambda)}$$

Table 1
Example methods for deriving normalized $a_{ph}(\lambda)$.

Reference	Uses measured data (Y/N)	Input data required	Description
Prieur and Sathyendranath (1981)	Y	C_a	Single $a^*_{ph}(\lambda)$ vector
Roesler et al. (1989)	Y	C_a	Single $a^*_{ph}(\lambda)$ shape
Lee et al. (1996a, b)	N	C_a	Blends two Gaussian basis vectors
Bricaud et al. (1995)	Y	C_a	Blends two basis vectors
Bricaud et al. (1998)	Y	C_a	Blends two basis vectors
Hoge and Lyon (1996)	N	C_a	Single Gaussian basis vector (Hoepffner and Sathyendranath, 1993)
Sathyendranath et al. (2001)	Y	C_a	Blends $a^*_{ph}(\lambda)$ basis vectors for two phytoplankton populations
Ciotti et al. (2002)	Y	Size parameter, S_f	Blends $a^*_{ph}(\lambda)$ basis vectors for micro- and picophytoplankton

Werdell et al. 2018, Prog. Oceanography

you live in a consumer's market!

$$r_{rs}(\lambda) = G(\lambda) \frac{b_{bw}(\lambda) + B_{bp} b^*_{bp}(\lambda)}{a_w(\lambda) + A_{ph} a^*_{ph}(\lambda) + +A_{dg} a^*_{dg}(\lambda)}$$

Table 3
Example methods for deriving normalized S_{bp}.

Method	Uses measured data (Y/N)	Input data required	Description
Morel and Maritorena (2001)	Y	C_a	• S_{bp} defined in terms of $\tilde{b}$
Gordon et al. (1988)	N	C_a	• Defines $b_p(\lambda)$ from C_a • Assumes $\tilde{b} = F(Ca)$
Ciotti et al. (1999)	Y	C_a	• Logarithmic function from -2 where $C_a = 0.05\,\mathrm{mg\,m^{-3}}$ to 0 where $C_a = 20\,\mathrm{mg\,m^{-3}}$
Lee et al. (2002)	Y	$r_{rs}(\lambda)$	• Empirical relationship
Roesler and Boss (2003)	N	$\tilde{b}(\lambda)$, $c_p(\lambda)$, $a_p(\lambda)$	• Solves for for $b_{bp}(\lambda)$ and S_c
Antoine et al. (2011)	Y	C_a or $b_b(555)$	• Empirical relationship
Brewin et al. (2012)	Y	C_a	• Empirical relationship for $b_{bp}(\lambda)$

Werdell et al. 2018, Prog. Oceanography

you live in a consumer's market!

$$r_{rs}(\lambda) = G(\lambda) \frac{b_{bw}(\lambda) + B_{bp} b^*_{bp}(\lambda)}{a_w(\lambda) + A_{ph} a^*_{ph}(\lambda) + +A_{dg} a^*_{dg}(\lambda)}$$

Table 2
Example approaches for partitioning $a(\lambda)$ into $a_{ph}(\lambda)$ and $a_{dg}(\lambda)$.

Reference	Method and assumptions for parameterizing $a_{ph}(\lambda)$ and $a_{dg}(\lambda)$	Additional input data	Applied to ocean color data? (Y/N)
Roesler et al. (1989)	• $a_{dg}(\lambda)$ has fixed exponential slope, S_{dg} • $a_{ph}(\lambda)$ blue-to-red absorption peak defined using pigment data • Solves for $a_{dg}(\lambda)$ and $a_{ph}(\lambda)$	• C_a • Phaeophytin-a concentration	N
Lee et al. (2002)	• $a_{dg}(\lambda)$ has fixed exponential slope, S_{dg} • Empirical relationship uses $r_{rs}(\lambda)$ to parameterize band ratio of $a_{ph}(\lambda)$ • Solves for $a_{dg}(\lambda)$ and $a_{ph}(\lambda)$	• None	Y
Ciotti and Bricaud (2006) Method 1	• $a_{dg}(\lambda)$ assumed to be exponential • Empirical relationships uses C_a to parameterize band ratios of $a_{ph}(\lambda)$ • Solves for M_{dg}, S_{dg}, $a_{dg}(\lambda)$, and $a_{ph}(\lambda)$ algebraically (Bricaud and Stramski 1990)	• C_a	Y
Ciotti and Bricaud (2006) Method 2	• $a_{dg}(\lambda)$ assumed to be exponential • $a_{ph}(\lambda)$ parameterized through mixing of pico- and microphytoplankton contributions (Ciotti et al. 2002) • Solves for M_{dg}, S_{dg}, M_{ph}, and the size parameter of phytoplankton (S_f) via nonlinear optimization	• C_a	Y
Zheng and Stramski (2013b)	• $a_{dg}(\lambda)$ assumed to be exponential • $a_{ph}(\lambda)$ shape expressed through band ratios of 412:443 and 510:490 • Searches multiple speculative (feasible) solutions of M_{dg}, S_{dg}, $a_{dg}(\lambda)$, and $a_{ph}(\lambda)$ (Bricaud and Stramski, 1990) • Computes candidate and selects optimal solution for $a_{dg}(\lambda)$ and $a_{ph}(\lambda)$ using stacked inequality contstraints	• Pre-determined bounds of inequality constraints	Y
Zhang et al. (2015)	• $a_{dg}(\lambda)$ assumed to be exponential • $a_{ph}(\lambda)$ parameterized as the sum of mixing of pico-, nano-, and microphytoplankton contributions • Solves for M_{dg}, S_{dg}, $a_{dg}(\lambda)$, and $a_{ph}(\lambda)$ including contributions of pico-, nano-, and microphytoplankton using constrained least-squares optimization	• C_a-specific $a^*_{ph}(\lambda)$ for pico-, nano-, and micro-phytoplankton	N

Werdell et al. 2018, Prog. Oceanography

you live in a consumer's market!

$$r_{rs}(\lambda) = G(\lambda) \frac{b_{bw}(\lambda) + B_{bp} b_{bp}^{*}(\lambda)}{a_w(\lambda) + A_{ph} a_{ph}^{*}(\lambda) + +A_{dg} a_{dg}^{*}(\lambda)}$$

Table 5. Delta Statistics for the Sensitivity Analyses[a]

		MPD				Median				
Run	N	b_{bp}	a	a_{dg}	a_ϕ	ΔR_{rs}	Δb_{bp}	Δa	Δa_{dg}	Δa_ϕ
GIOP-DC	437	NA	NA	NA	NA	1.04	8.52	8.56	27.25	35.83
$S_{bp} - 33\%$	440	5.19	5.17	7.58	2.98	0.99	11.23	11.70	32.14	34.69
$S_{bp} + 33\%$	436	5.65	5.70	8.82	2.90	1.14	11.40	10.70	23.51	39.12
$S_{dg} - 33\%$	448	18.96	33.44	101.73	46.59	1.61	16.27	19.08	32.94	31.95
$S_{dg} + 33\%$	399	3.77	8.41	40.10	32.92	1.23	9.44	8.95	79.90	59.32
S_{dg} from [7]	439	3.20	5.33	20.40	14.58	1.10	8.65	9.80	22.25	34.42
$C_a - 33\%$ in [14]	419	2.02	2.92	1.48	7.25	1.19	8.79	8.83	28.62	31.10
$C_a + 33\%$ in [14]	437	1.56	2.28	1.14	5.90	1.10	8.12	9.17	26.79	40.09
Fixed C_a in [14]	369	4.57	7.89	2.60	21.68	1.46	11.30	11.53	30.97	26.70
a_ϕ^* from [17]	357	8.33	12.72	7.04	22.23	1.20	14.26	16.75	38.30	23.13
G from [22]	422	9.99	6.15	7.49	14.12	1.16	11.50	13.64	37.49	36.24
Matrix inversion	475	4.60	3.68	2.24	7.41	1.73	9.15	9.43	24.79	36.82
$400 \le \lambda \le 600$ nm	424	0.23	0.21	0.08	0.38	0.92	8.76	8.78	31.94	36.55

[a] N is the sample size. MPD is the average spectral median percent difference between GIOP-DC and each alternate run, as calculated in Tables 2 and 3. Medians of the ΔIOP frequency distributions are also presented, as presented in Table 4.

Werdell et al. 2013, Applied Optics

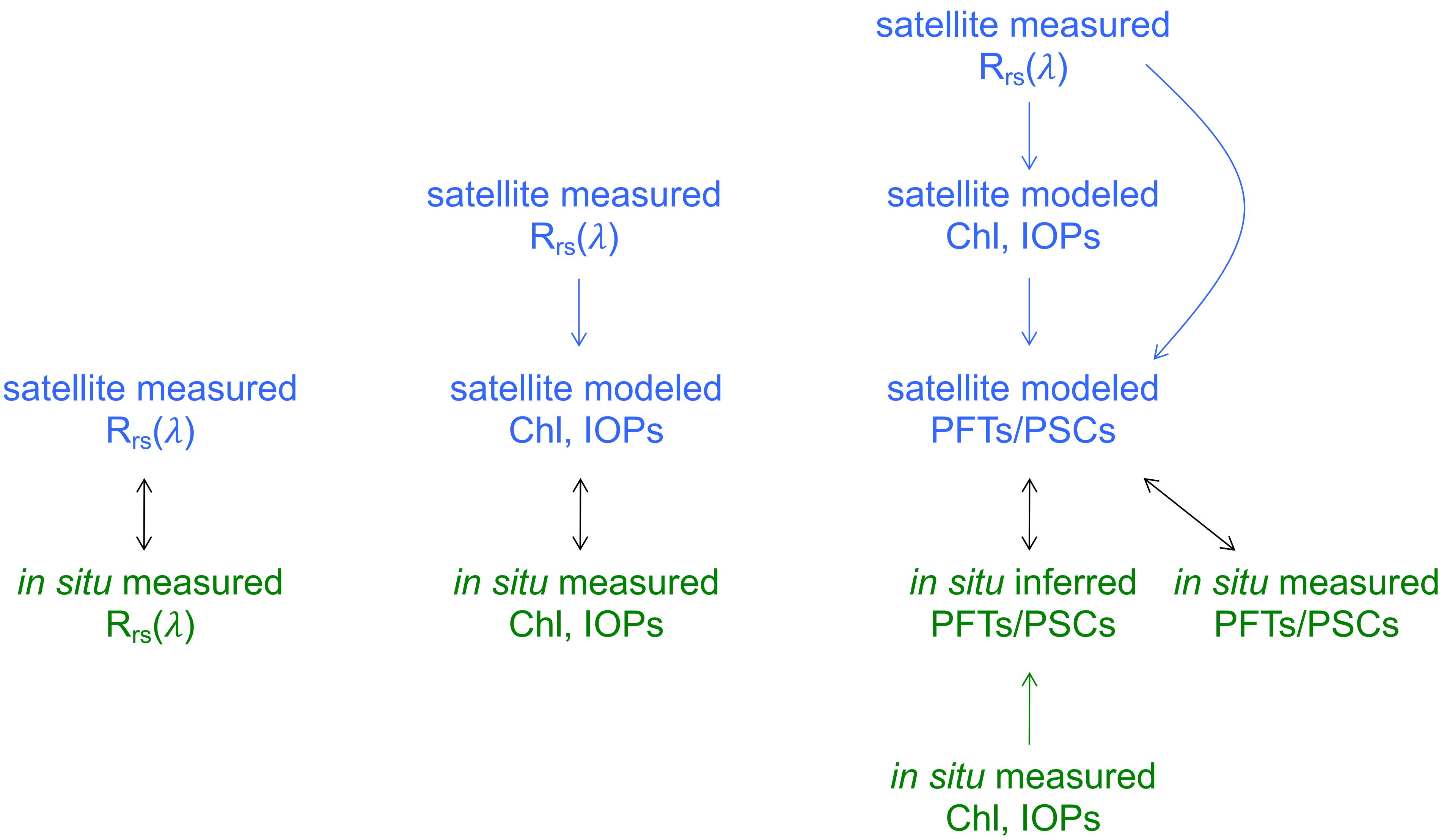
In situ data inform empirical relationships

In situ data inform semi-analytic retrievals

Degrees of separation between *in situ* and satellite data vary by product suite

one algorithm or dataset CANNOT always represent all conditions (and this is ok)

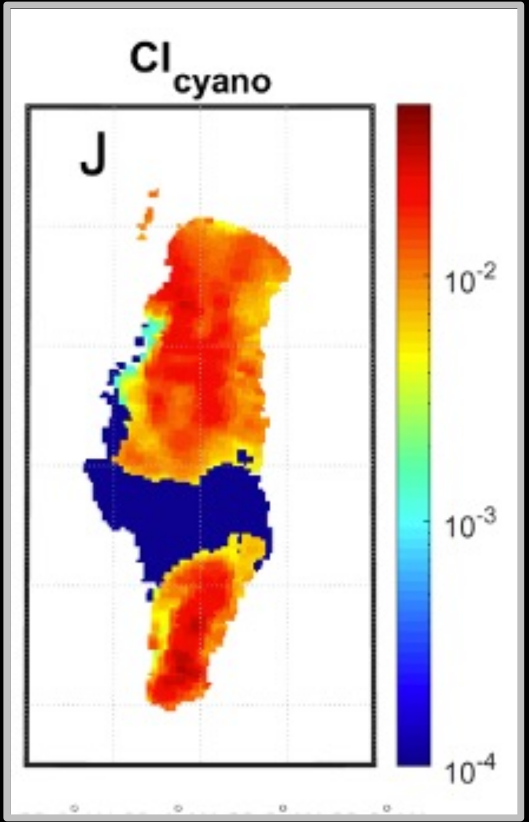
degrees of separation in data products



consider a question

Given what you know about the in situ *methods and the satellite algorithms, how would you prepare the* in situ *data for a validation satellite exercise to get as close to apples-to-apples comparisons as possible (e.g., common units, observational space, etc.)?*

microscopy
genetic/molecular methods
flow cytometry
coulter counters
video imaging (IFCB, FlowCam)
continuous plankton recorder
spectroscopy
optics (b_b, c spectral slopes)
HPLC pigment analyses
etc.



unitless (derived from $\rho_w(\lambda)$)

satellite measured $R_{rs}(\lambda)$

↓

satellite modeled Chl, IOPs

↓

satellite modeled PFTs/PSCs

↕

in situ inferred PFTs/PSCs | *in situ* measured PFTs/PSCs

↑

in situ measured Chl, IOPs

In situ data inform empirical relationships

one algorithm or dataset CANNOT always represent all conditions (and this is ok)

In situ data inform semi-analytic retrievals

Degrees of separation between *in situ* and satellite data vary by product suite

meaningfully relating *in situ* and satellite variables is an area of ongoing research (and this is ok)

Atmospheric correction is riddled with *in situ data*

NASA/TM–2016-217551

Atmospheric Correction for Satellite Ocean Color Radiometry

Curtis D. Mobley, Jeremy Werdell, Bryan Franz, Ziauddin Ahmad, and Sean Bailey

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

June 2016

aerosol lookup tables (AERONET)
NIR correction (black pixel assumption)
out-of-band correction
pure seawater
foam and whitecap mask/correction
BRDF correction
(others I'm forgetting)

<u>ancillary data</u>
atmospheric pressure
water vapor
relative humidity
wind speed
ozone
NO2
SST, SSS
sea ice

In situ data inform empirical relationships

In situ data inform semi-analytic retrievals

one algorithm or dataset CANNOT always represent all conditions (and this is ok)

Degrees of separation between *in situ* and satellite data vary by product suite

meaningfully relating *in situ* and satellite variables is an area of ongoing research (and this is ok)

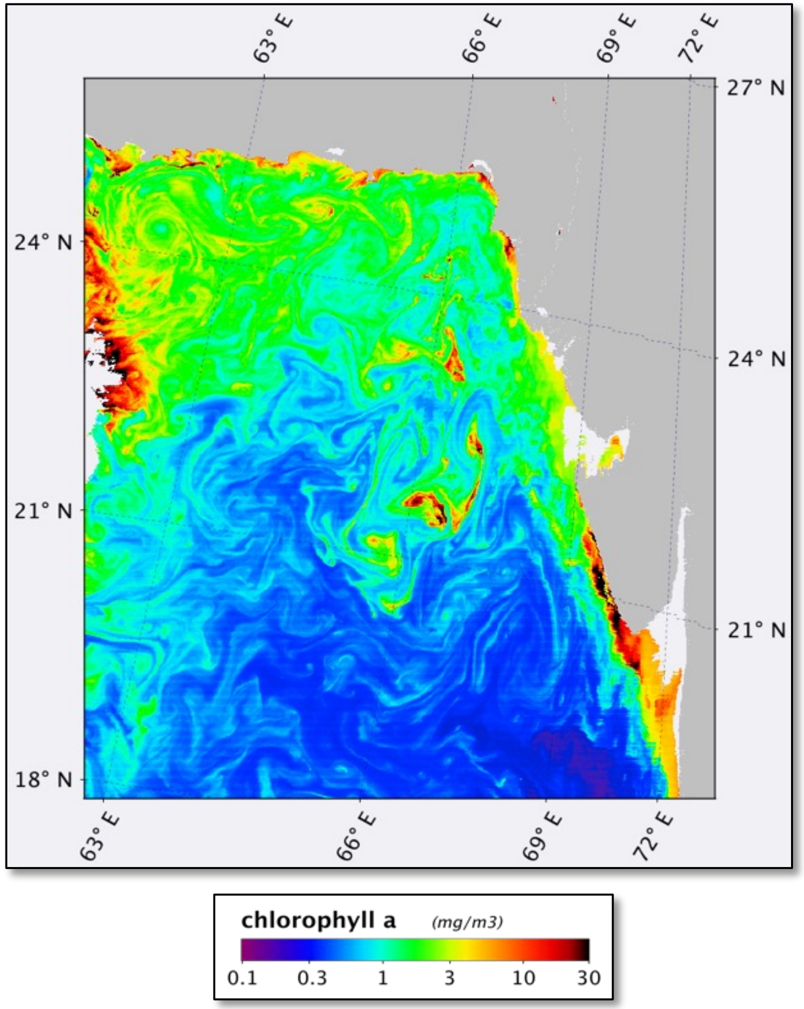
Atmospheric correction is riddled with *in situ data*

in situ data are embedded into almost 100% of ocean color

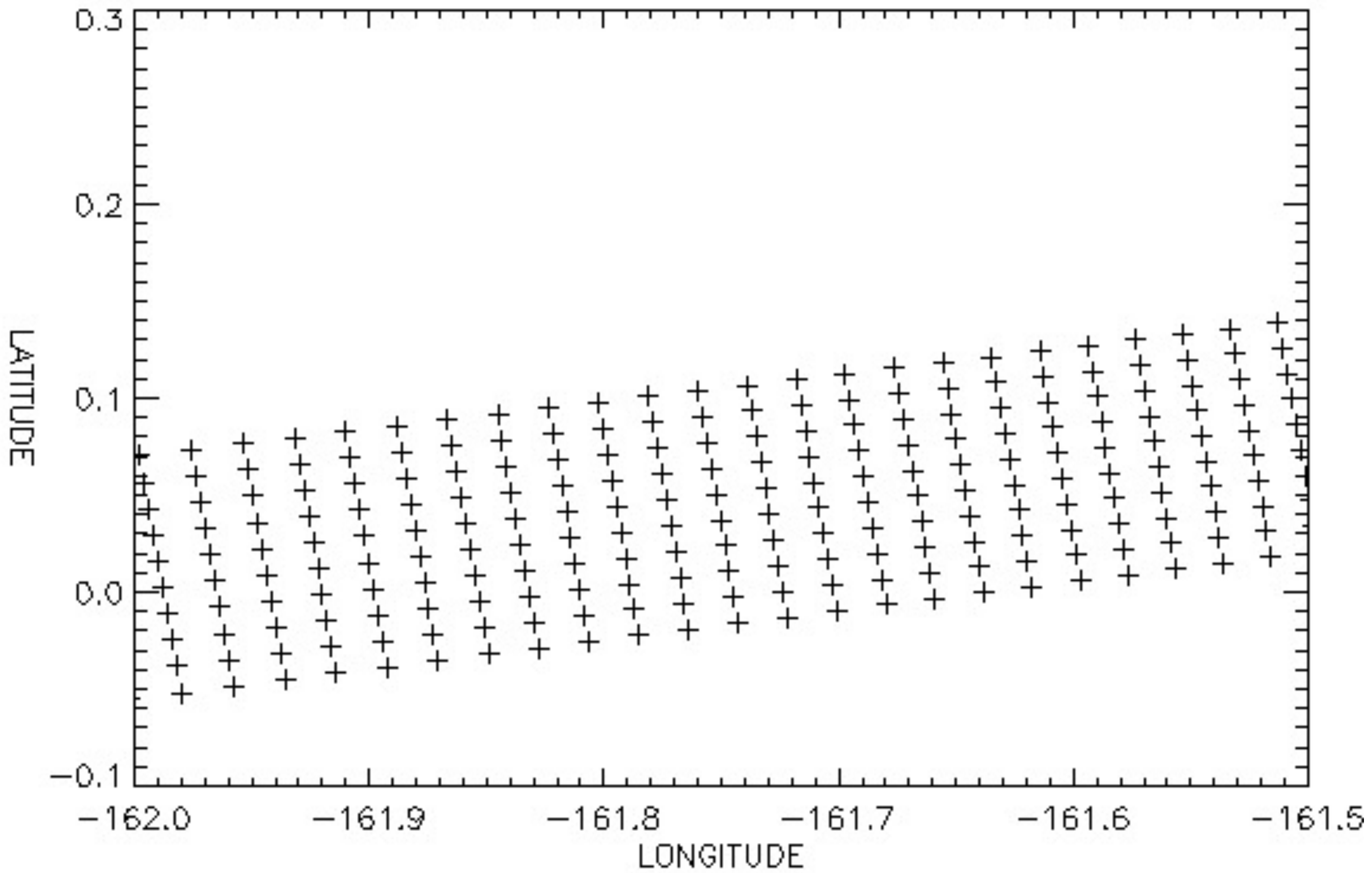
Data treatment / compositing changes the answer

understand how data processing changes the “answers”

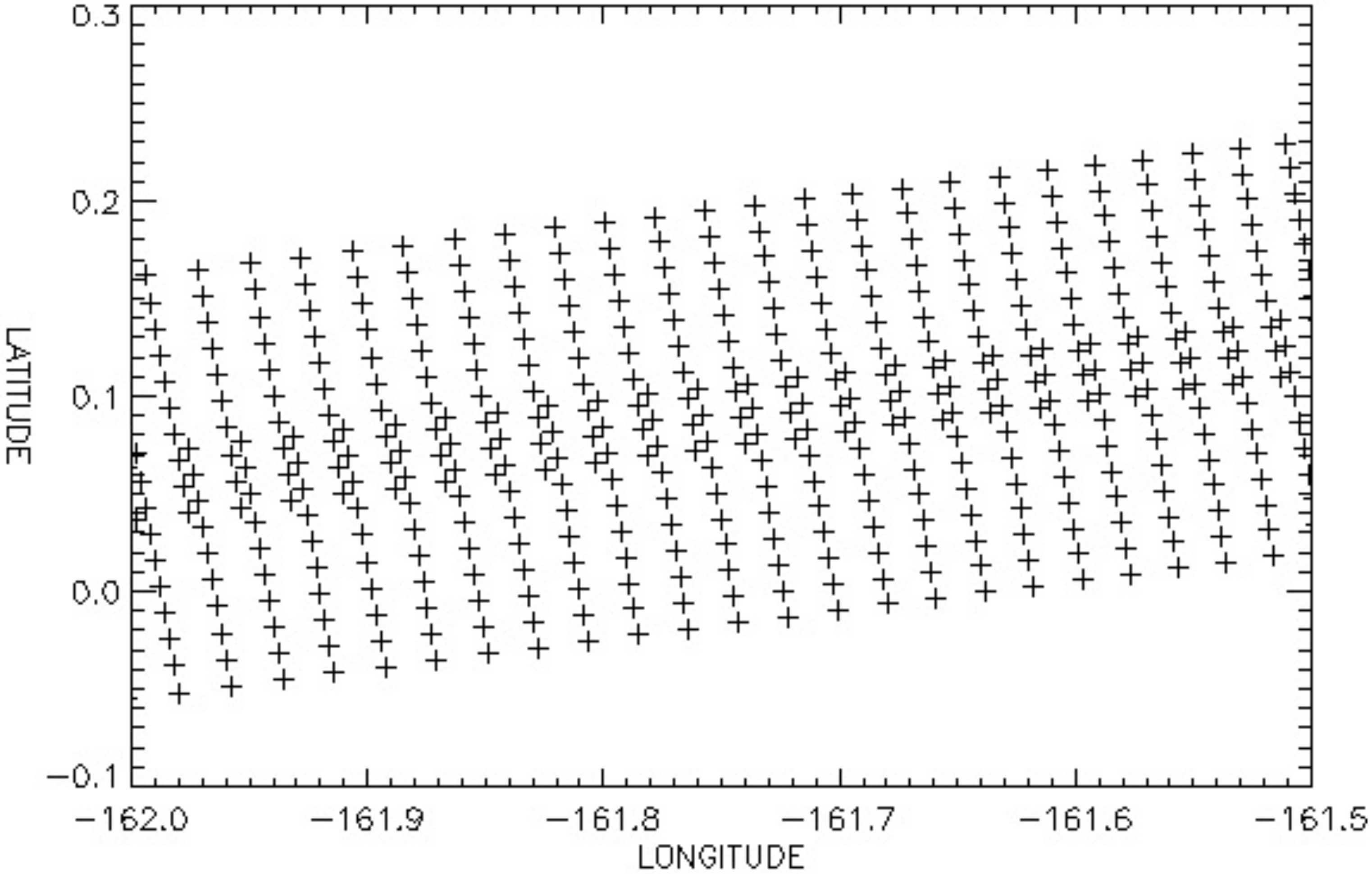
For your consideration:
- horizontal resolution
- temporal resolution
- vertical resolution



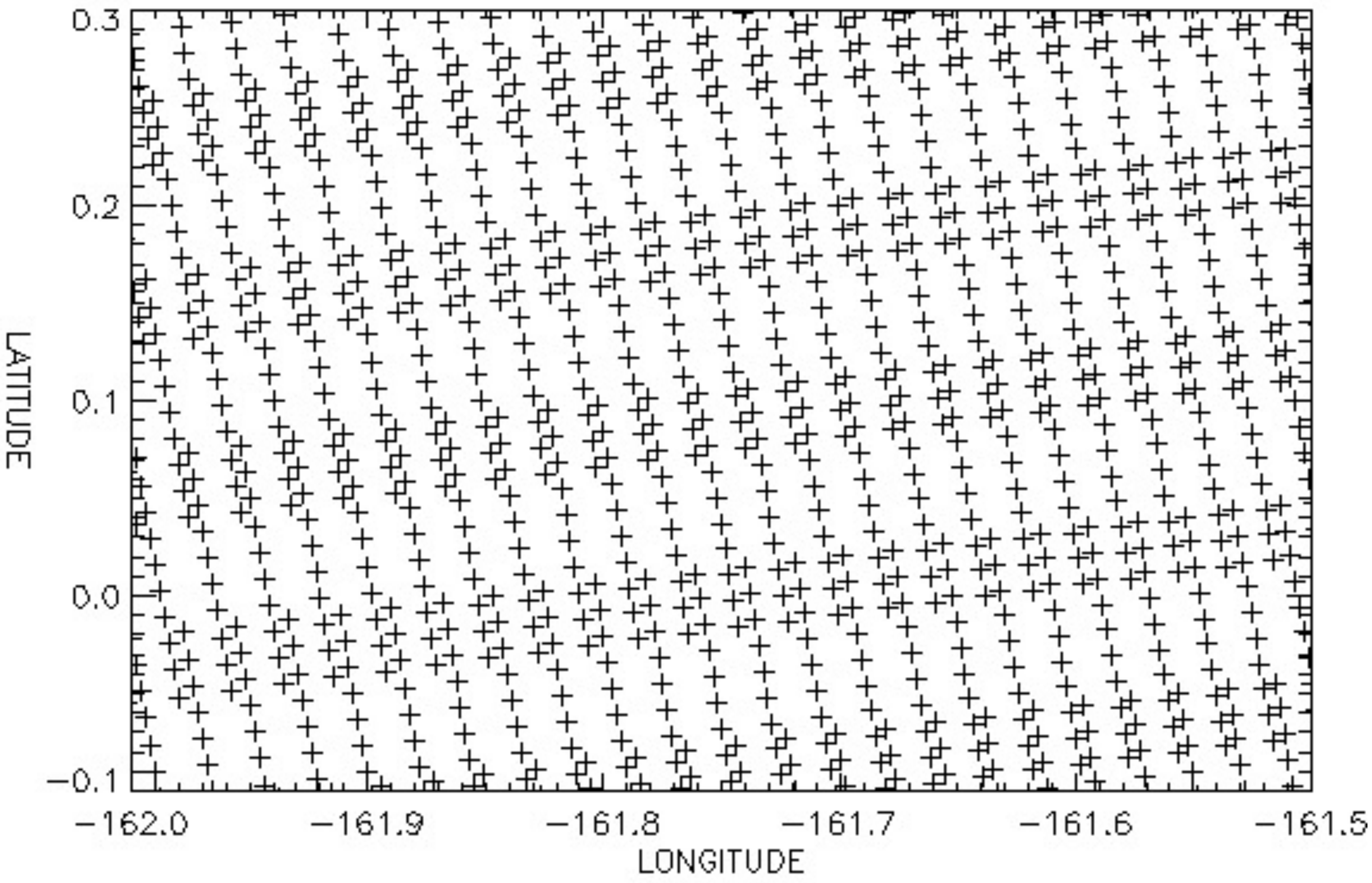
one MODIS scan at ~45 degrees scan angle



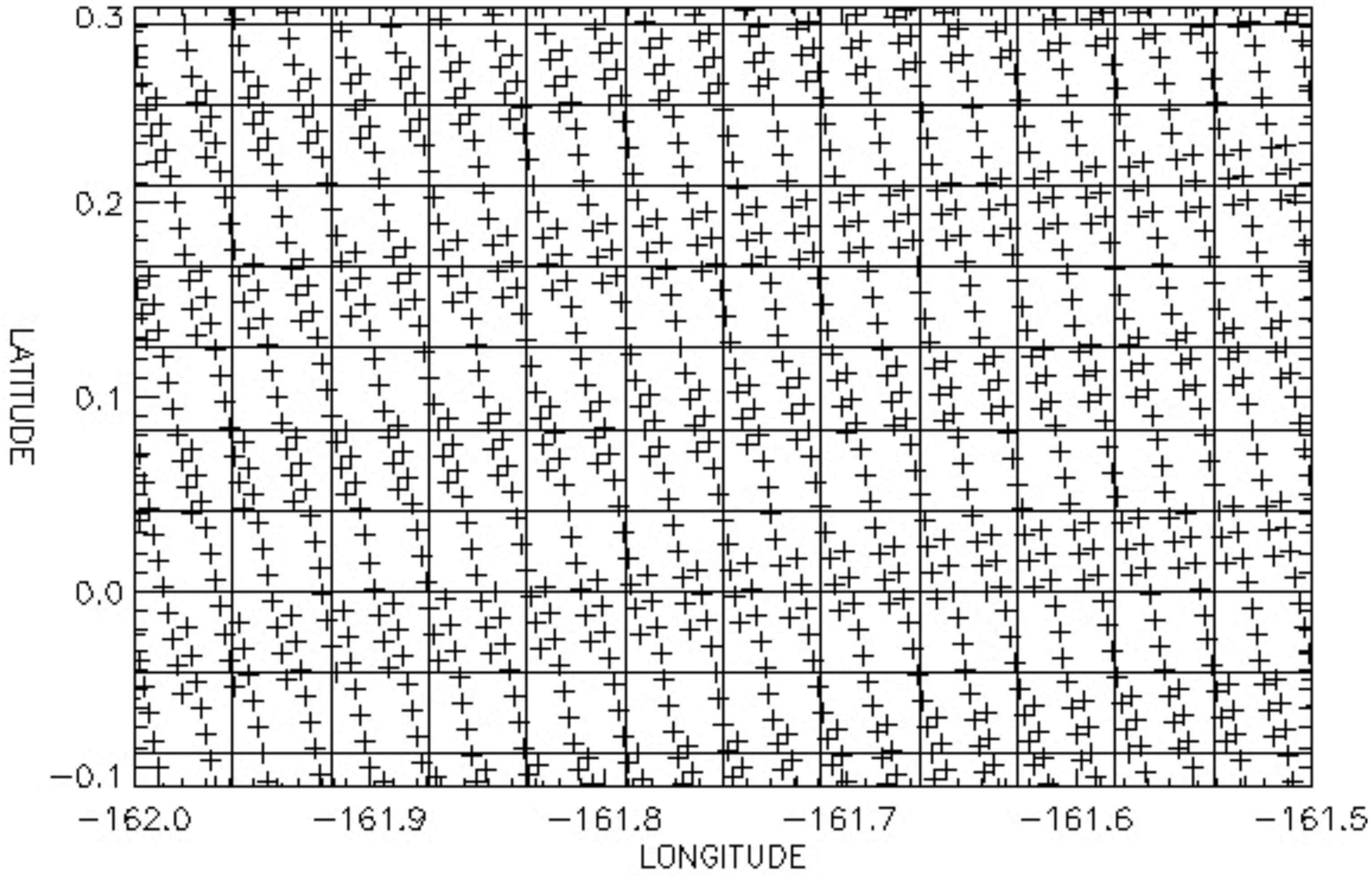
two MODIS scans showing overlap of pixels



multiple MODIS scans showing pixel overlap

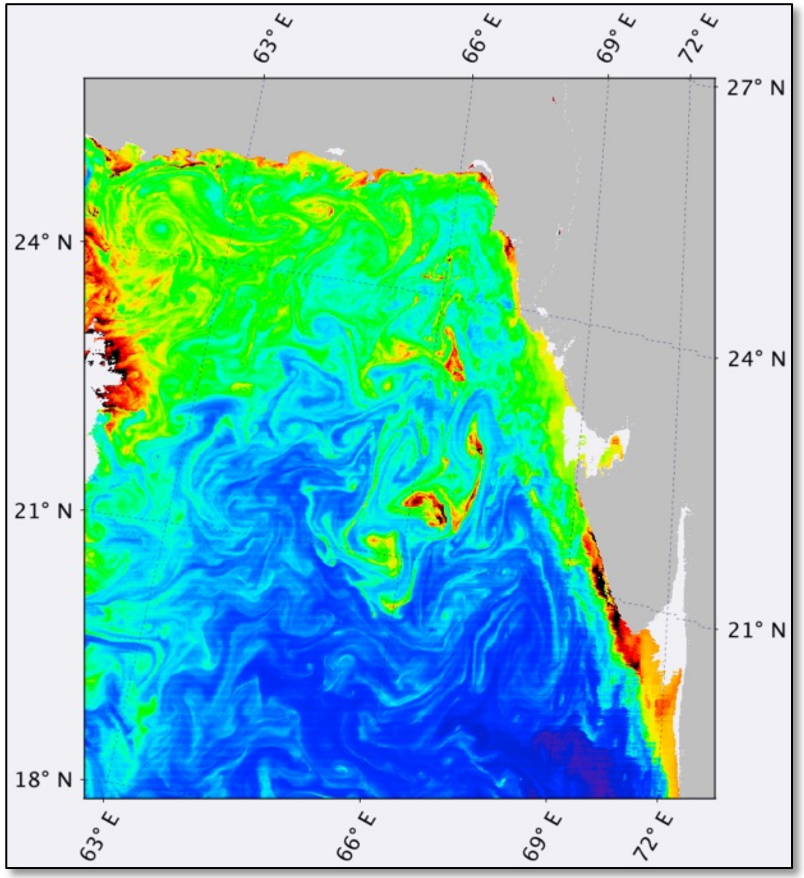


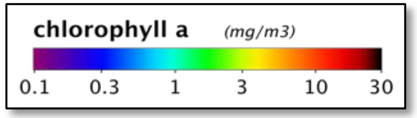
bin boundaries overlaid on pixel locations

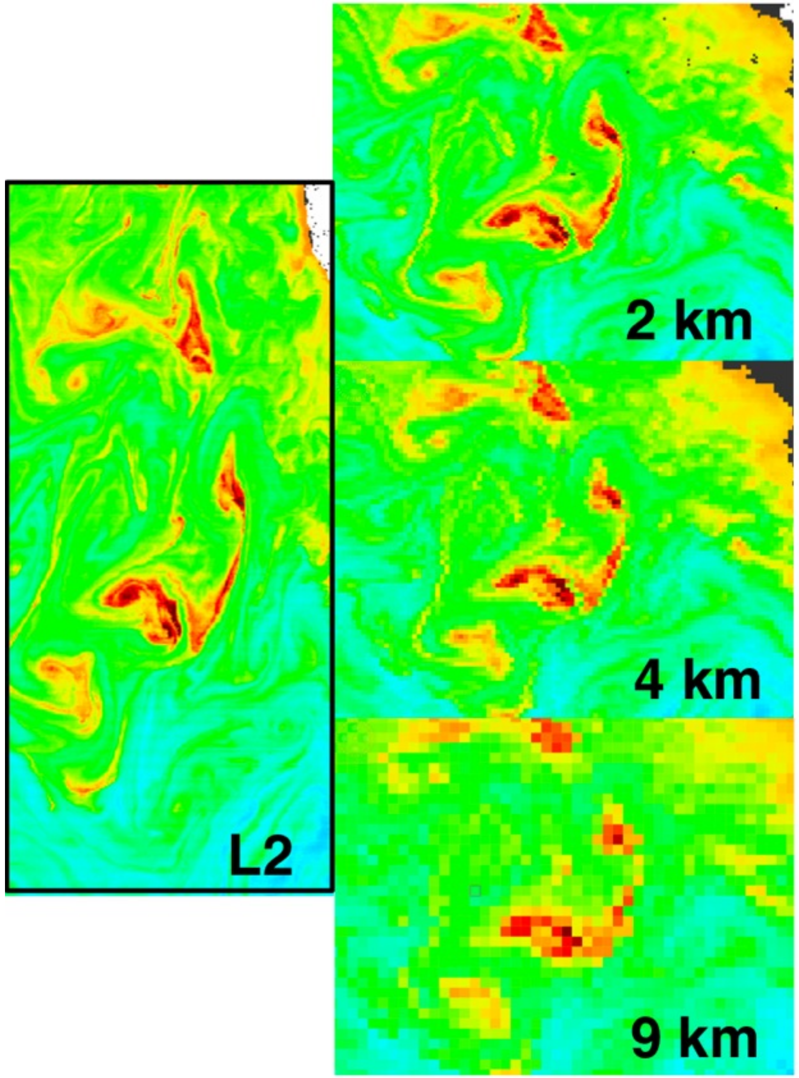


understand how data processing changes the "answers"

For your consideration:
- horizontal resolution
- temporal resolution
- vertical resolution

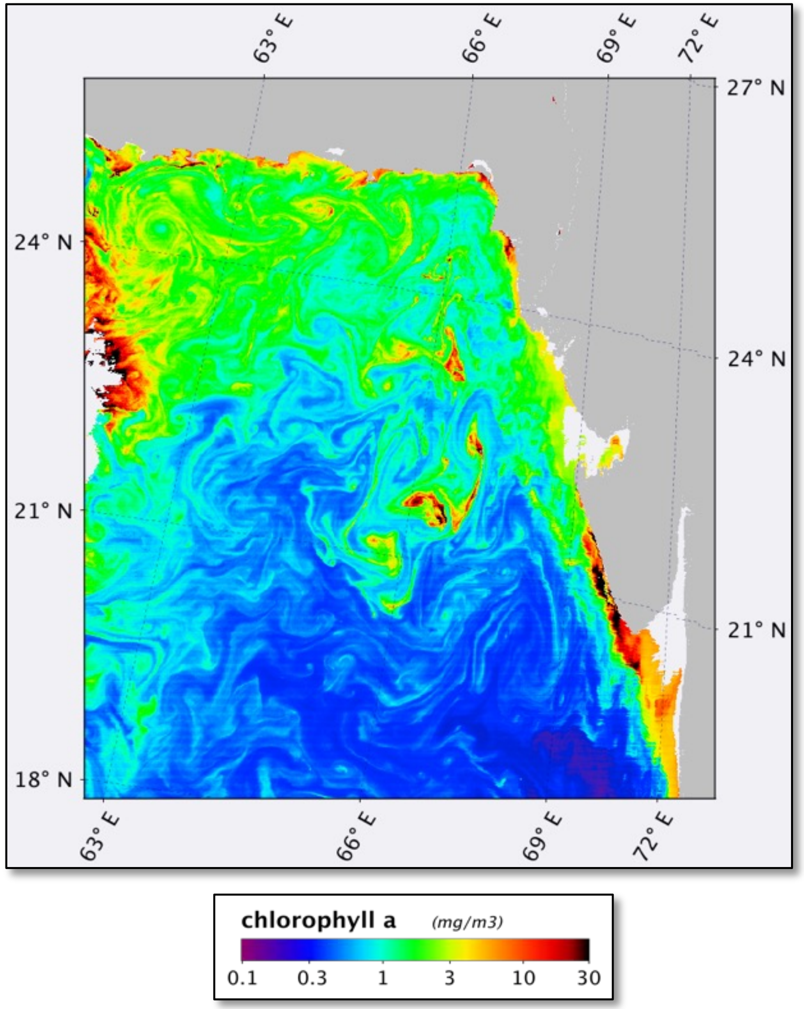


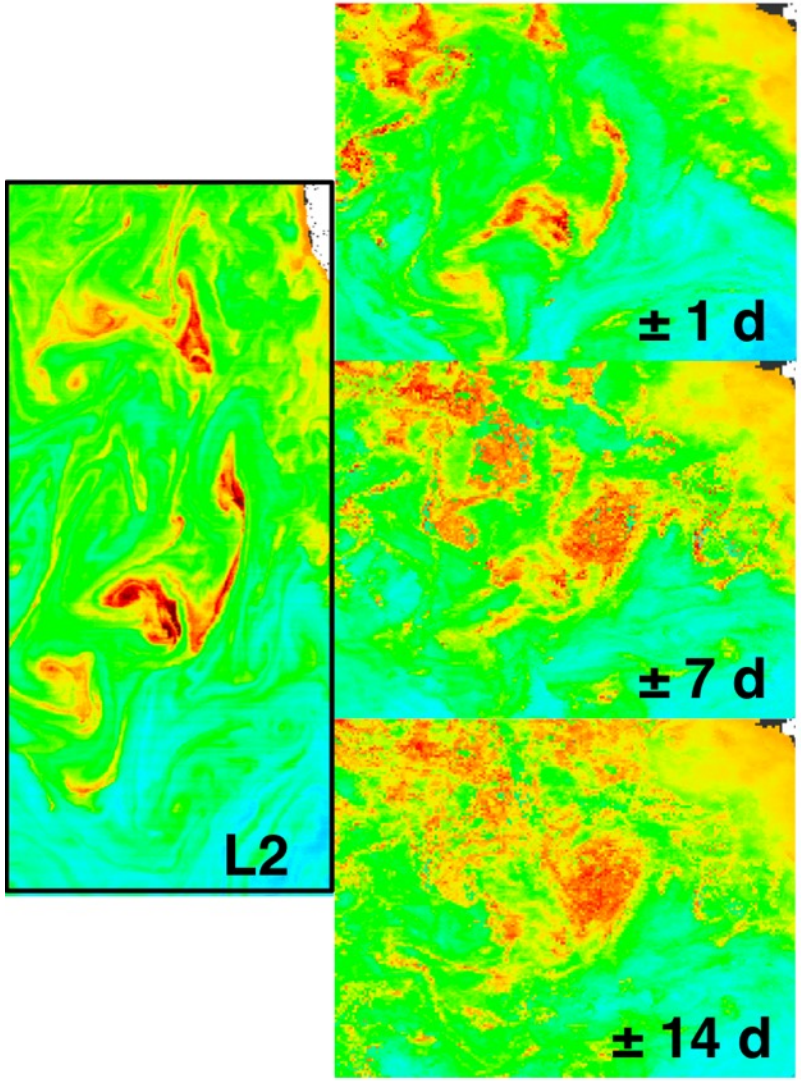




understand how data processing changes the "answers"

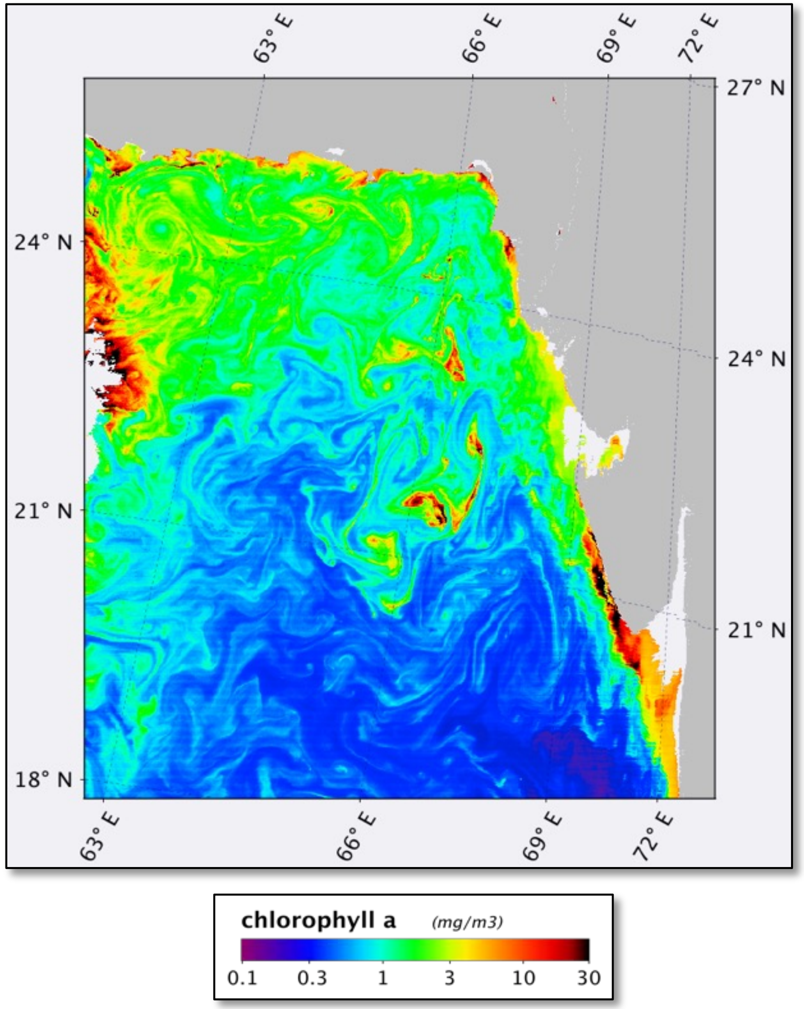
For your consideration:
- horizontal resolution
- temporal resolution
- vertical resolution

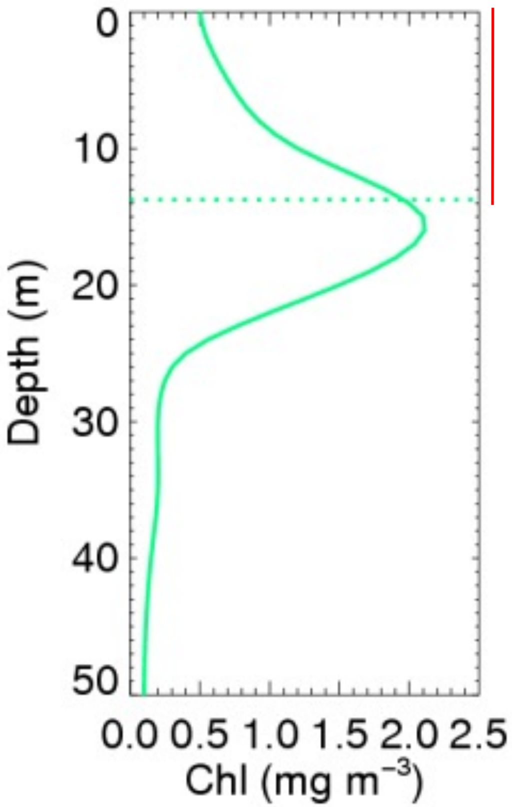




understand how data processing changes the “answers”

For your consideration:
- horizontal resolution
- temporal resolution
- vertical resolution





first optical depth
$0.37 = \exp(-K_d z)$
$-1 = -K_d z$

Estimation of the Depth of Sunlight Penetration in the Sea for Remote Sensing

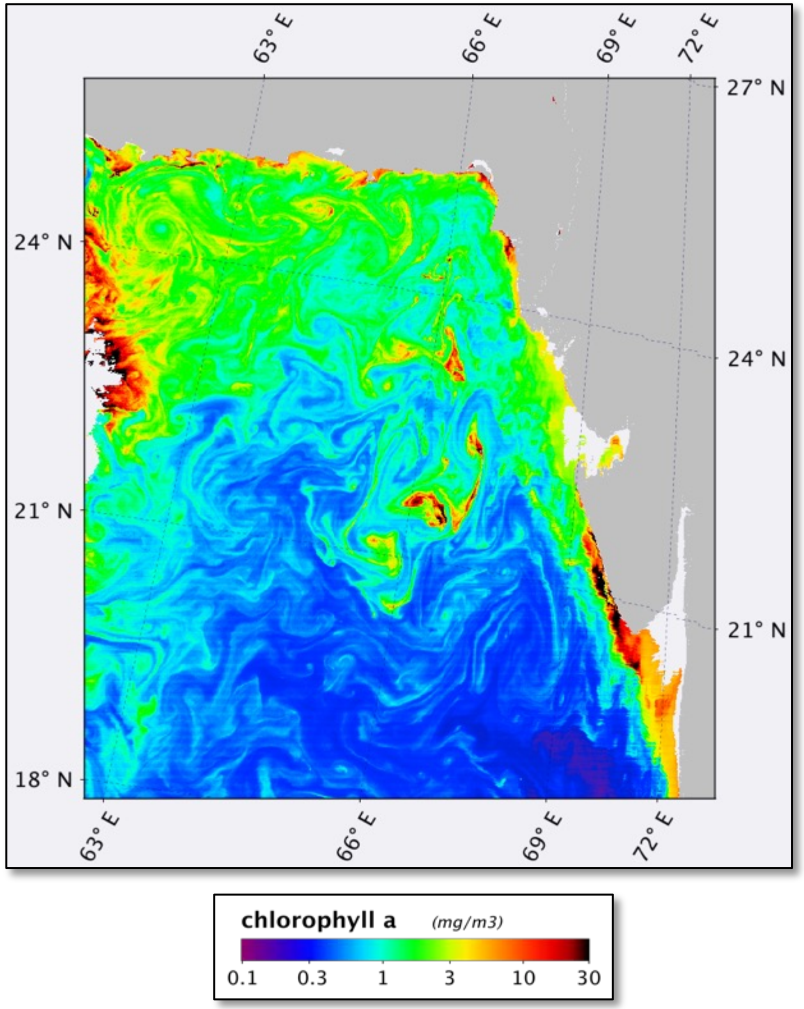
Howard R. Gordon and W. R. McCluney

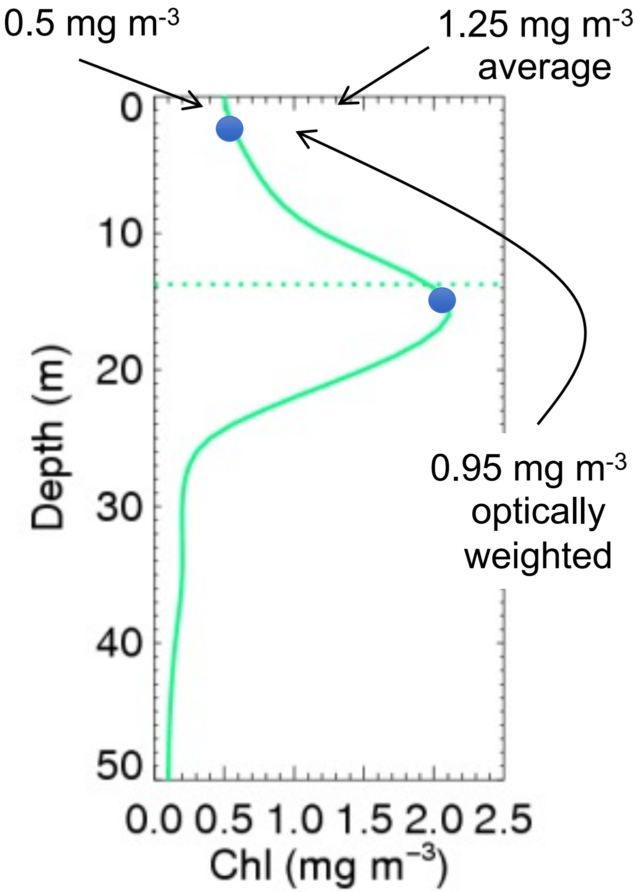
February 1975 / Vol. 14, No. 2 / APPLIED OPTICS 413

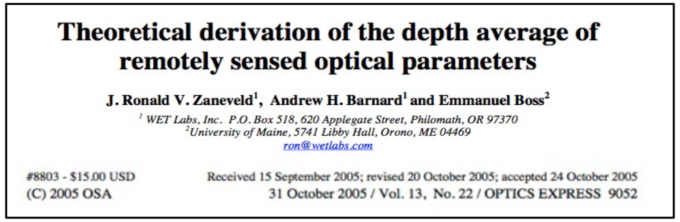
understand how data processing changes the “answers”

For your consideration:

- horizontal resolution
- temporal resolution
- vertical resolution







Theoretical derivation of the depth average of remotely sensed optical parameters

J. Ronald V. Zaneveld[1], Andrew H. Barnard[1] and Emmanuel Boss[2]

[1] WET Labs, Inc. P.O. Box 518, 620 Applegate Street, Philomath, OR 97370
[2] University of Maine, 5741 Libby Hall, Orono, ME 04469
ron@wetlabs.com

#8803 - $15.00 USD
Received 15 September 2005; revised 20 October 2005; accepted 24 October 2005

31 October 2005 / Vol. 13, No. 22 / OPTICS EXPRESS 9052

In situ data inform empirical relationships

In situ data inform semi-analytic retrievals

one algorithm or dataset CANNOT always represent all conditions (and this is ok)

Degrees of separation between *in situ* and satellite data vary by product suite

meaningfully relating *in situ* and satellite variables is an area of ongoing research (and this is ok)

Atmospheric correction is riddled with *in situ data*

in situ data are embedded into almost 100% of ocean color

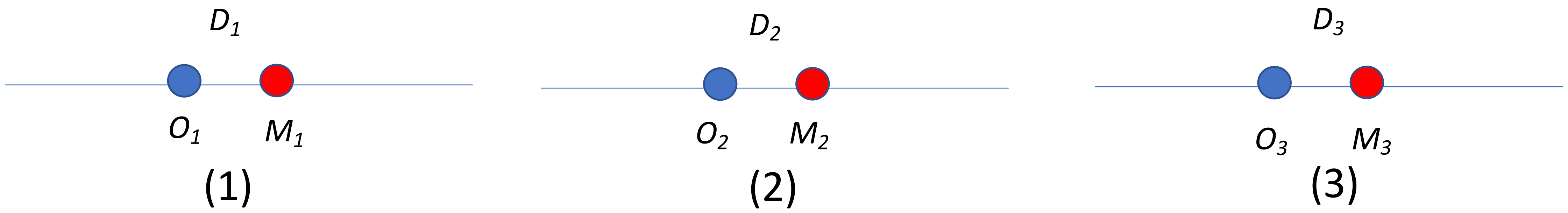
Data treatment / compositing changes the answer

Consideration of scales and resolution is critical to interpret differences

Validation should consider *in situ* uncertainties

Question:
Based on horizontal distance (*D*) between the blue and red dots, which pair(s) below would you consider to be different: 1, 2, or 3?

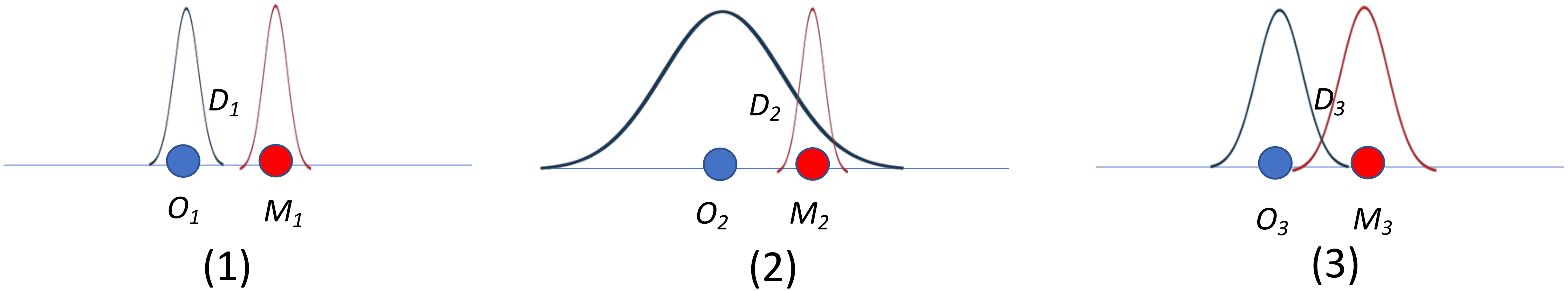
Answer: D is the same for all (stop wasting our time!)



Question:
Based on horizontal distance (*D)* between the blue and red dots, which pair(s) below would you consider to be different: 1, 2, or 3?

This time, we'll consider measurement uncertainty and draw a probability density function around each point …

Answer: (a) yes, (b) no, (c) somewhat.



Validation metrics

$$mean\ bias = \frac{1}{N}\sum_{i=0}^{N} M_i - O_i$$

$$MAE = \frac{1}{N}\sum_{i=0}^{N} |M_i - O_i|$$

MAE: mean absolute error

A method to account for overlapping PDFs

For mean bias and MAE, we compute the difference between the satellite observed (O_i) and in situ measurement (M_i) data pairs:

$$D_i = M_i - O_i$$

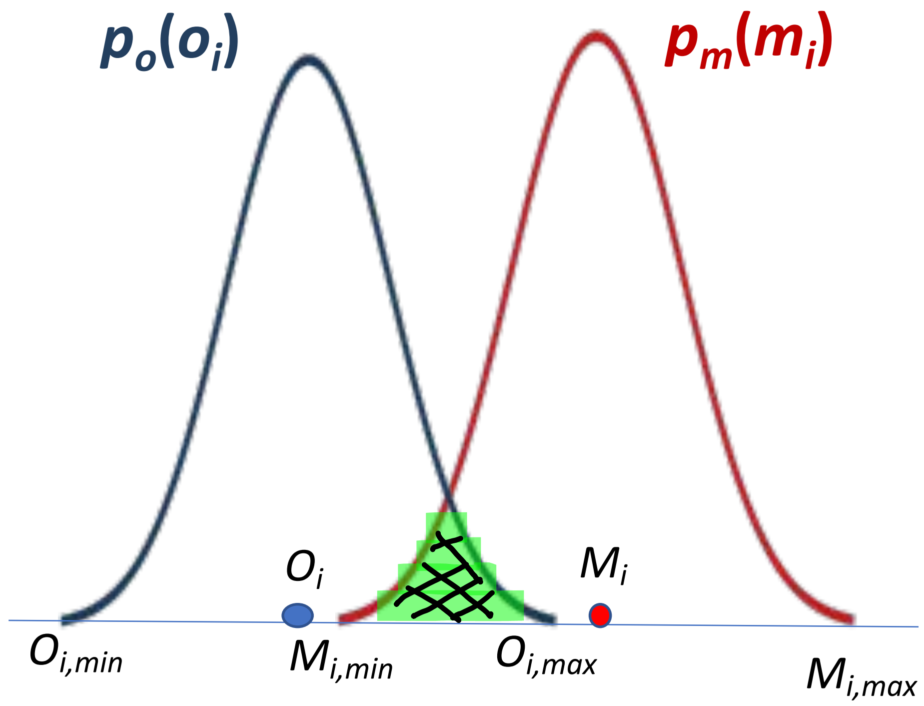
We correct difference with correction factor (CF):

$$CF_i = 1 - DO_i$$

(DO_i) is the *degree of overlap metric* proposed by Harmel et al (2010) (see paper for calculus).

Corrected difference is:

$$D'_i = CF_i D_i$$



A method to account for overlapping PDFs

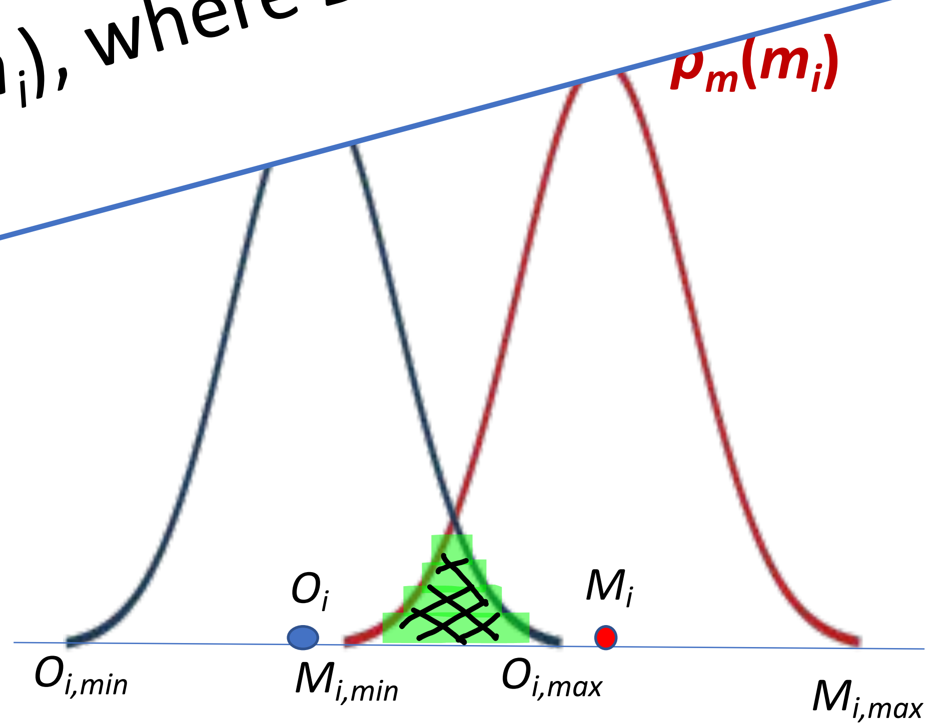
For mean bias and MAE, we compute the difference between the satellite observed (O_i) and in situ measurement (M_i) data pairs:

- Less weight is applied when DO_i approaches 1
- For completely overlapping $p_o(o_i)$ and $p_m(m_i)$, where $DO_i = 1$, no difference can be discerned

...tric proposed by ...ee paper for calculus).

C...rrected difference is:

$$D'_i = CF_i D_i$$



Validation metrics

$mean\ bias$ ...

Correcte...

JGR Oceans

RESEARCH ARTICLE

10.1029/2021JC017231

Key Points:

- A reflectance line height metric was used as a predictor of the particulate backscattering coefficient at 555 nm
- The degree of overlap metric was used to correct validation skill metrics for measurement

Development and Validation of an Empirical Ocean Color Algorithm with Uncertainties: A Case Study with the Particulate Backscattering Coefficient

Lachlan I. W. McKinna[1], Ivona Cetinić[2,3], and P. Jeremy Werdell[3]

[1]Go2Q Pty Ltd, Sunshine Coast, QLD, Australia, [2]GESTAR/USRA, Columbia, MD, USA, [3]NASA Goddard Flight Center,, Greenbelt, MD, USA

$$mean\ bias' = \frac{1}{N}\sum_{i=0}^{N} CF_i(M_i - O_i)$$

$$MAE' = \frac{1}{N}\sum_{i=0}^{N} |CFi(M_i - O_i)|$$

MAE: mean absolute error

In situ data inform empirical relationships

one algorithm or dataset CANNOT always represent all conditions (and this is ok)

In situ data inform semi-analytic retrievals

Degrees of separation between *in situ* and satellite data vary by product suite

meaningfully relating *in situ* and satellite variables is an area of ongoing research (and this is ok)

Atmospheric correction is riddled with *in situ data*

in situ data are embedded into almost 100% of ocean color

Data treatment / compositing changes the answer

Consideration of scales and resolution is critical to interpret differences

Validation should consider *in situ* uncertainties

in situ uncertainties contribute to the differences realized with satellites

Move beyond scatter plots

population statistics: long-term distributions and time series

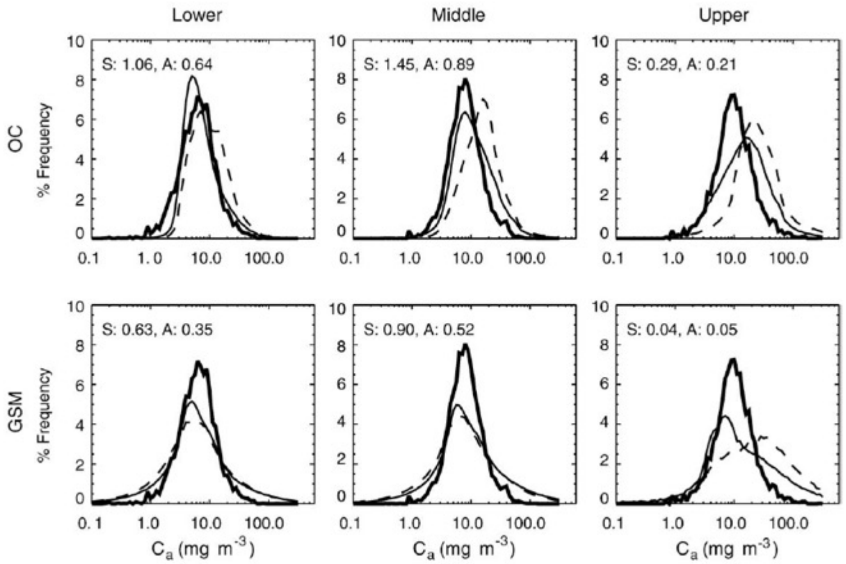


Fig. 4. *In situ* C_a distributions (thick lines) in the Lower, Middle, and Upper Bays compared to OC and GSM retrievals for SeaWiFS (thin solid) and MODIS-Aqua (thin dashed). OC4 was used for SeaWiFS and OC3 for MODIS-Aqua. Samples sizes (in million pixels) for satellite retrievals are provided in each panel, with SeaWiFS indicated by *S* and MODIS-Aqua indicated by *A*. *In situ* sample sizes are 7204, 5814, and 3660 for the Lower, Middle, and Upper Bay, respectively. Data from all four seasons are included.

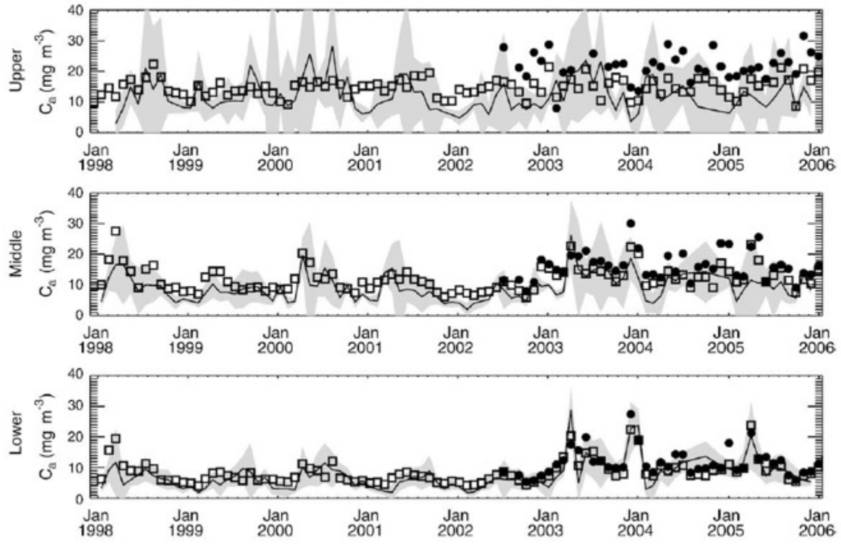


Fig. 5. Monthly averages of *in situ* C_a (thin lines) in the Upper, Middle, and Lower Bays compared to OC retrievals for SeaWiFS (OC4; empty squares) and MODIS-Aqua (OC3; filled circles). The grey shaded area represents one standard deviation about the *in situ* averages. The RPD reported in the text was calculated as 100% · (median($C_a^{satellite}/C_a^{in\ situ}$) − 1) using each monthly satellite and *in situ* pair.

residual histograms and scatter plots

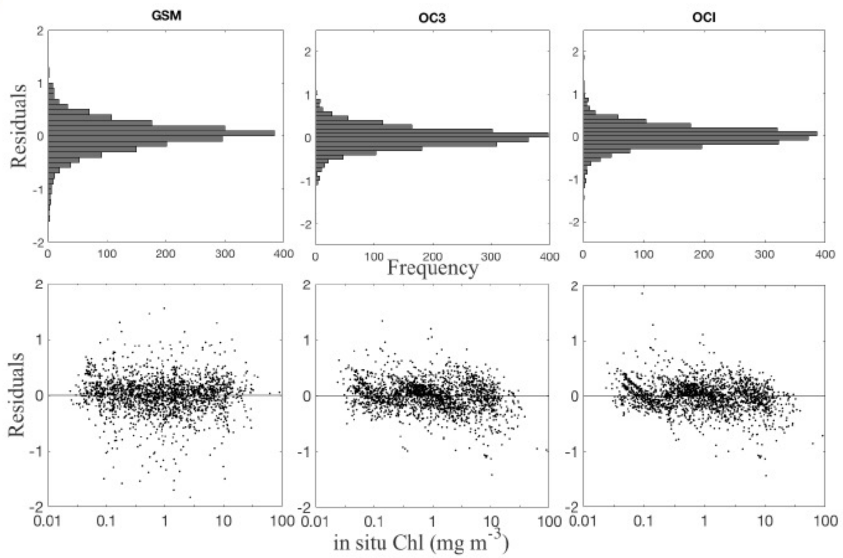


Fig. 3. Log_{10} residuals histograms and scatterplots the SeaWiFS-to-*in situ Chl* match-ups. The top row are histograms of log_{10} summarizing the error distribution of GSM, OC3, and OCI algorithms. The bottom panels are residual plots of the difference between model satellite *Chl* and the reference *in situ* values versus reference values. The plots were created with log_{10} values, but the axes are in *Chl* units (mg m^{-3}).

$$\zeta \frac{M_i - O_i}{\sqrt{u(M_i)^2 + u(O_i)^2}}$$

Zeta-score plots

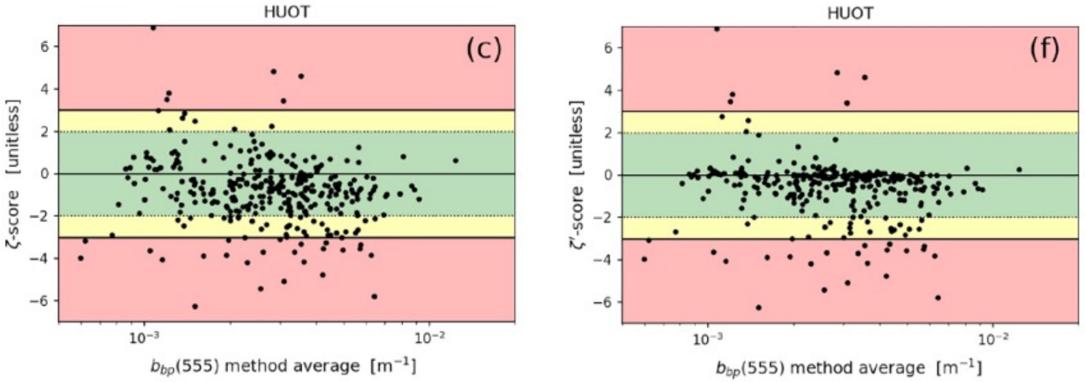


Figure 4. Zeta score plots comparing modeled and observed $b_{bp}(555)$ varying with the method average values of $b_{bp}(555)$. Subplots (a)–(c) correspond to LH, GIOP, and Huot models, respectively. Subplots (d)–(f) use corrected zeta scores for the LH, GIOP, and Huot models, respectively.

Taylor and target diagrams

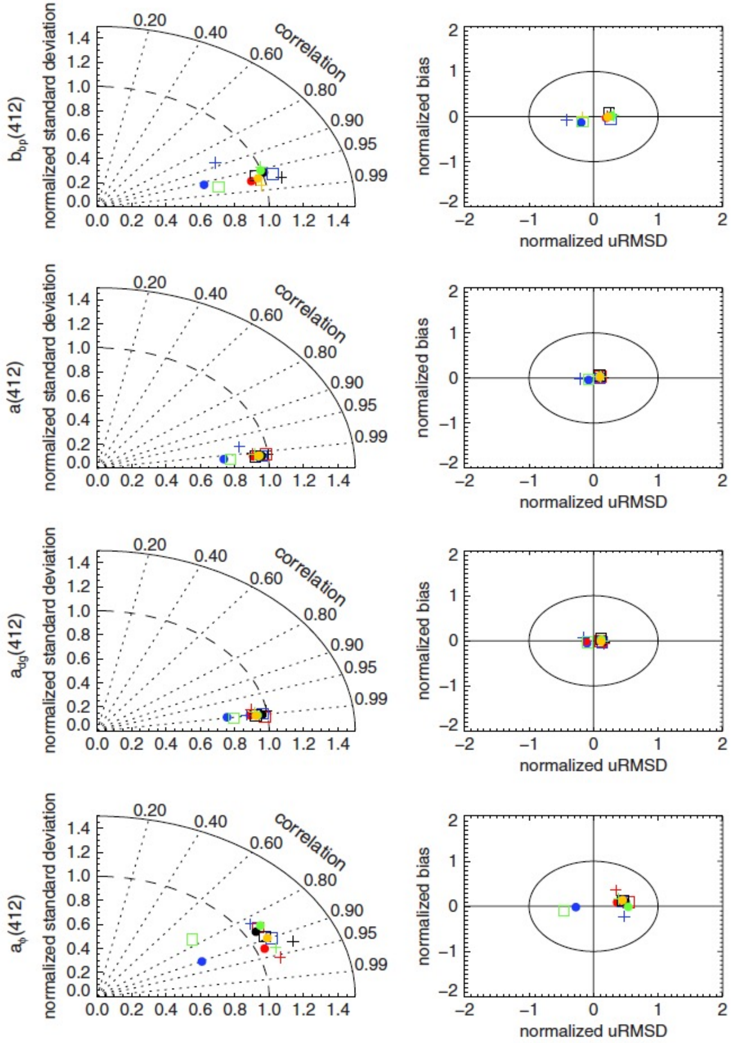


Fig. 5. Taylor and Target diagrams for IOPs at 412 nm from the IOCCG data set for the 12 alternate parameterizations of GIOP compared to GIOP-DC. *uRMSD* is the unbiased root mean square difference. Symbols indicate the following: blue cross = $S_{dg} - 33\%$ ($= 0.012\ \mathrm{nm}^{-1}$); red cross = $S_{dg} + 33\%$ ($= 0.024\ \mathrm{nm}^{-1}$); green circle = S_{dg} dynamically calculated using Lee *et al.* [7]; blue square = S_{bp} from Lee *et al.* [7] -33%; black square = S_{bp} from Lee *et al.* [7] $+33\%$; red circle = OC-derived $C_a - 33\%$ prior to input into Bricaud *et al.* [14]; black circle = OC-derived $C_a + 33\%$ prior to input into Bricaud *et al.* [14]; green square = $a^*_\phi(\lambda)$ from Bricaud *et al.* [14] with C_a fixed at $0.18\ \mathrm{mg\,m^{-3}}$; blue circle = $a^*_\phi(\lambda)$ from Ciotti and Bricaud [17] with a size fraction of 0.5; black cross = $G(\lambda)$ from Morel *et al.* [22]; orange cross = optimization using linear matrix inversion; and green cross = optimization considering only $400 \le \lambda \le 600$ nm.

confusion matrices

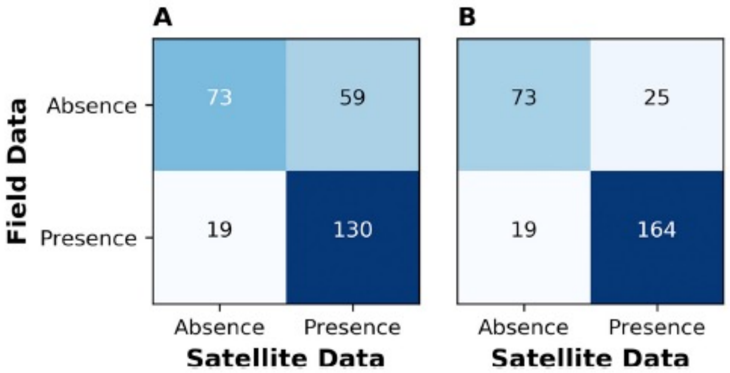


Fig. 3. Confusion matrix from the algorithm validation with (A) MC only, and (B) MC and cyanobacteria cell density as field CyanoHAB reference data. Numbers in the confusion matrix represent the sample count in each scenario. CyanoHAB class and the no-bloom classes were coded as 'Presence' and 'Absence' class.

star plots

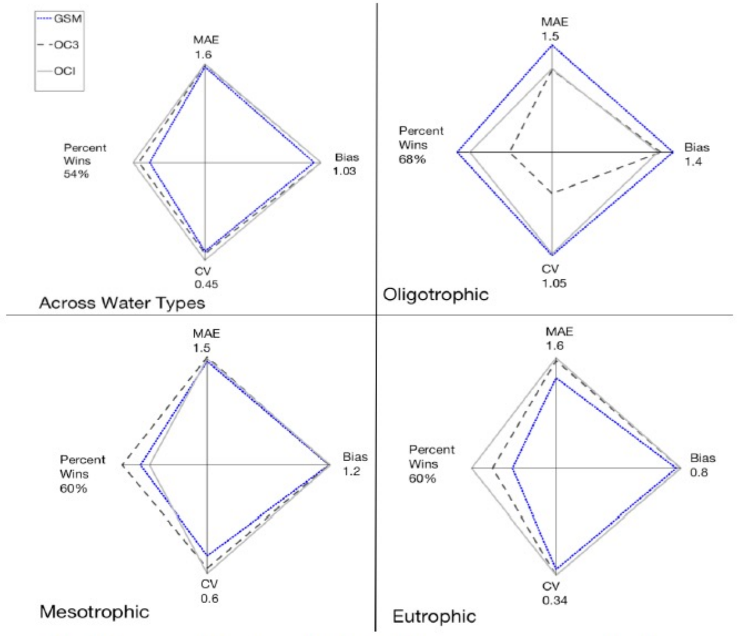


Fig. 6. Comparison of the metrics results of bias, MAE, pairwise percent wins and coefficient of variation summarized in star plots across all water types. The plot center represents values that indicate poor algorithm performance, while farthest from center represents the best performance. The numbers represent the value of the best performing algorithm value for each metric.

In situ data inform empirical relationships

In situ data inform semi-analytic retrievals

one algorithm or dataset CANNOT always represent all conditions (and this is ok)

Degrees of separation between *in situ* and satellite data vary by product suite

meaningfully relating *in situ* and satellite variables is an area of ongoing research (and this is ok)

Atmospheric correction is riddled with *in situ data*

in situ data are embedded into almost 100% of ocean color

Data treatment / compositing changes the answer

Consideration of scales and resolution is critical to interpret differences

Validation should consider *in situ* uncertainties

in situ uncertainties contribute to the differences realized with satellites

Move beyond scatter plots

any one figure CANNOT reveal all meaningful information

what about environmental conditions?

SPACE SCIENCE

Global
Biogeochemical Cycles

RESEARCH ARTICLE
10.1029/2018GB006118

Key Points:
- Globally, light availability in the water column is the most important parameter for phytoplankton size distribution
- Regionally, phytoplankton size distributions vary, responding to variable light and modes of nutrient delivery
- Cell size is increasing in the cold ocean and the dynamic regions in the warm ocean and declining in the warm ocean

A Satellite Assessment of Environmental Controls of Phytoplankton Community Size Structure

Colleen B. Mouw[1], Audrey B. Ciochetto[1], and James A. Yoder[1]

[1]Graduate School of Oceanography, University of Rhode Island, Narragansett, RI, USA

Abstract Phytoplankton play a key role as the base of the marine food web and a crucial component in the Earth's carbon cycle. There have been a few regional studies that have utilized satellite-estimated phytoplankton functional type products in conjunction with other environmental metrics. Here we expand to a global perspective and ask, what are the physical drivers of phytoplankton composition variability? Using a variety of satellite-observed ocean color products and physical properties spanning 1997–2015, we characterize spatial and temporal variability in phytoplankton community size structure in relation to

LIMNOLOGY AND OCEANOGRAPHY

ASLO
Association for the Sciences of Limnology and Oceanography

Article | Free Access

Identifying four phytoplankton functional types from space: An ecological approach

Dionysios E. Raitsos, Samantha J. Lavender, Christos D. Maravelias, John Haralabous, Anthony J. Richardson, Philip C. Reid

First published: 31 March 2008 | https://doi.org/10.4319/lo.2008.53.2.0605 | Citations: 75

ELSEVIER

Remote Sensing of Environment
Volume 240, April 2020, 111689

Incorporating environmental data in abundance-based algorithms for deriving phytoplankton size classes in the Atlantic Ocean

Timothy S. Moore [a, b], Christopher W. Brown [c]

Show more

+ Add to Mendeley   Share   Cite

https://doi.org/10.1016/j.rse.2020.111689   Get rights and content

NEVER GIVE UP

NEVER STOP TRYING TO EXCEED YOUR LIMITS. WE NEED THE ENTERTAINMENT.