UNIVERSITY OF CALIFORNIA
SANTA CRUZ

ADVANCES IN BIO-OPTICS FOR OBSERVING AQUATIC ECOSYSTEMS

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

OCEAN SCIENCES

by

Henry F. Houskeeper

June 2020

The Dissertation of Henry F. Houskeeper
is approved:

Professor Raphael M. Kudela, Chair

Professor Christopher A. Edwards

Dr. John P. Ryan

Dr. Stanford B. Hooker

Quentin Williams
Acting Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures v

List of Tables viii

Abstract ix

Dedication x

Acknowledgments xi

Introduction 1
References . . . 3

1 Atmospheric correction effects on coastal ocean satellite datasets 5
1.1 Introduction . . . 6
1.2 Methods . . . 9
1.2.1 Site selection . . . 9
1.2.2 Atmospheric dataset . . . 10
1.2.3 Biological field data . . . 11
1.2.4 Satellite data . . . 12
1.2.5 Remote estimation of phytoplankton biomass . . . 13
1.2.6 Matchups, climatological averages, and error statistics . . . 13
1.3 Results . . . 14
1.3.1 Association between RBD and phytoplankton biomass . . . 14
1.3.2 Climatology of atmospheric correction flags . . . 15
1.3.3 Impact of atmospheric correction masks on RBD datasets . . . 17
1.4 Discussion . . . 19
1.4.1 Performance of satellite products at matchup sites . . . 19
1.4.2 Variability of atmospheric constituents and ocean color AC flags . . . 22
1.4.3 Potential for user evaluation of Level 2 and 3 datasets . . . 22
1.5 Conclusion . . . 23
References . . . 24

2 Remote sensing of colored dissolved organic matter for oceanic, coastal, and inland waters **30**
2.1 Introduction . . . 31
2.2 Methods . . . 33
2.2.1 Derivation of synthetic case-1 bio-optical datasets . . . 33
2.2.2 Biogeochemical parameterizations of $K_d(\lambda)$ . . . 34
2.2.3 Biogeochemical parameterizations of $\left[L_W(\lambda)\right]_{\mathrm{N}}$ . . . 35
2.2.4 Description of above- and in-water data products . . . 35
2.2.5 Quality control and partitioning of datasets . . . 37
2.2.6 Description of airborne datasets . . . 38
2.2.7 Derivation of algorithm fits for $\left[L_W(\lambda)\right]_{\mathrm{N}}$ EMA . . . 39
2.2.8 Summary of algorithm validation statistics . . . 40
2.3 Results . . . 41
2.3.1 Case-1 evaluation of a $K_d(\lambda)$ end-member approach . . . 41
2.3.2 EMA algorithms derived for $\left[L_W(\lambda)\right]_{\mathrm{N}}$ pairs . . . 43
2.3.3 Validation of $a_{\mathrm{CDOM}}(440)$ algorithms . . . 44
2.3.4 Efficacy of $a_{\mathrm{CDOM}}(440)$ retrievals in non-conservative water bodies . . . 48
2.3.5 Analysis of airborne imagery . . . 49
2.4 Discussion . . . 51
2.4.1 Consistency of EMA with case-1 parameterizations . . . 51
2.4.2 Performance of EMA algorithms . . . 55
2.5 Conclusion . . . 59
References . . . 60

3 Estimating chlorophyll *a* from hyperspectral particulate beam-attenuation in oceanic and coastal ecosystems **72**
3.1 Introduction . . . 73
3.1.1 Overview of absorption band effects on beam-attenuation spectra . . . 74
3.1.2 Relevance of absorption band effects to phytoplankton composition . . . 75
3.2 Methods . . . 77
3.2.1 Description of bio-optical dataset . . . 77
3.2.2 Reduction of $c'_{\mathrm{p}}(\lambda)$ spectra using empirical orthogonal functions . . . 79
3.3 Results . . . 80
3.3.1 Effects of autocorrelation . . . 80
3.3.2 Interpretation of the eigenfunctions . . . 81
3.3.3 Interpretation of the predictors . . . 83
3.4 Discussion . . . 85
3.4.1 Relevance of absorption band effects to phytoplankton dynamics . . . 85
3.4.2 Potential for application of absorption band effects . . . 88
3.5 Conclusions . . . 89
References . . . 90

Conclusions **97**
References . . . 98

List of Figures

1.1 *MODIS Aqua imagery of phytoplankton blooms obtained on 26 October 2016 in Monterey Bay, California (upper) and 27 September 2011 near Cape Columbine, Western Cape (lower).* ***(a)*** *Pseudo-true color images with clouds masked in grey;* ***(b)*** *Red Band Difference algorithm (a proxy for phytoplankton biomass) with clouds masked in grey; and* ***(c)*** *Red Band Difference algorithm with clouds and suspect atmospheric correction (defined as maximum aerosol iterations reached and low water-leaving radiances) masked in grey.* 8

1.2 *Location of biological and atmospheric measurements used in this study for* ***(a)*** *Monterey Bay, California and* ***(b)*** *St. Helena Bay, Western Cape. Chl a and fluorometer measurements were obtained at sites denoted with red circles, and atmospheric measurements at sites denoted with orange triangles. Regions of satellite L2 datasets used in this analysis are indicated by dashed black lines.* 11

1.3 *Climatology of atmospheric and satellite products near Monterey Bay and St. Helena Bay, shifted for the phase timing of northern and southern hemispheres (Monterey: solid black line, March-February; Simonstown: dashed black line, September-August). Panels indicate:* ***(a)*** *Aerosol optical depth;* ***(b)*** *assignment frequency of AC flags;* ***(c)*** *Ångstrom exponent; and* ***(d)*** *Red Band Difference.* 16

1.4 *Fraction of pixels assigned atmospheric correction (AC) flags as a function of Red Band Difference, with Monterey Bay and St. Helena Bay shown with solid and dashed black lines, respectively. The range shown encompasses greater than 98% of the data for both regions.* 18

1.5 *Standardized bias in Red Band Difference from MODISA composites (L3) due to the removal of underlying (L2) pixels, shown for standard map grids within:* ***(a)*** *a transect within the central California Current System;* ***(b)*** *a transect within the southern Benguela Current System;* ***(c)*** *within Monterey Bay; and* ***(d)*** *within St. Helena Bay.* 19

1.6 *Standardized bias in Red Band Difference composites (L3) across longitudinal transects due to the removal of underlying (L2) pixels assigned AC flags. The California Current System transect is indicated by a solid line and the Benguela Current System transect is indicated by a dashed line.* 20

1.7 *Standardized bias in Red Band Difference for L3 composites, shown as a function of L2 AC flag assignments for data within the California Current System (solid line) and Benguela Current System (dashed line) transects and within Monterey (dashed-dotted) and St Helena (dotted) Bays.* 21

2.1 *Locations of C-AERO flight lines and data products after glint correction:* ***(a)*** *SFBD in California;* ***(b)*** *LT in California and Nevada; and* ***(c)*** *MB in California. Twin Otter flight lines are indicated in solid black, with C-AERO data products shown as red circles, and water sampling matchup sites indicated as orange triangles. Grey triangles indicate water sampling sites that were not considered for matchups due to spatial or temporal separation.* . . 39

2.2 $K_d(\lambda_1)/K_d(\lambda_2)$ *relationships to* $a_{\mathrm{CDOM}}(440)$*, with the dataset of Hooker et al., 2020 shown in red and overlaid, respectively, as follows:* ***(a)*** *GSyn with* $S = 0.018\,nm^{-1}$ *in solid grey, MSyn in dashed black, and EMA from Hooker et al., 2020 in solid black; and* ***(b)*** *GSyn with* $S = 0.018\,nm^{-1}$ *in solid grey, MSyn in dashed black, and a mean absolute deviation power-law fit in solid black.* R^2 *is reported for* $\log_{10}$*-transformed products. Panel inlays show relationships in linear space, with sensitivity testing for GSyn with* $S = 0.01, 0.03\,nm^{-1}$ *included in dashed grey.* 42

2.3 *OCEAN dataset (red dots), EMA algorithm (solid black line), and PSyn case-1 dataset (solid grey line) for:* ***(a)*** Λ_{780}^{320}*;* ***(b)*** Λ_{670}^{412}*;* ***(c)*** Λ_{555}^{443}*; and* ***(d)*** Λ_{625}^{465}*.* R^2 *is reported for* $\log_{10}$*-transformed products.* 44

2.4 *GLOBC dataset (red dots), EMA algorithm (solid black line), and PSyn case-1 dataset (solid grey line) for:* ***(a)*** Λ_{780}^{320}*;* ***(b)*** Λ_{670}^{412}*;* ***(c)*** Λ_{555}^{443}*; and* ***(d)*** Λ_{625}^{465}*.* R^2 *is reported for* $\log_{10}$*-transformed products.* 45

2.5 *Algorithm versus measured* $a_{\mathrm{CDOM}}(440)$ *from the GLOBC dataset for:* ***(a)*** Λ_{780}^{320}*;* ***(b)*** Λ_{670}^{412}*;* ***(c)*** *GSM; and* ***(d)*** *QAA. Linear axes are included in subpanels, with a one-to-one line shown in solid black for all panels.* 47

2.6 *The* Λ_{780}^{320} *EMA Algorithm applied to C-AERO measurements obtained at LSA and under clear-sky conditions at the following sites:* ***(a)*** *SFBD with larger San Pablo Bay on the left and smaller Grizzly Bay on the right (the northwest* $2\,m$ *bathymetric contour is overlaid for San Pablo Bay with a dashed grey line);* ***(b)*** *LT; and* ***(c)*** *MB.* 52

2.7 ***(a)*** *NOMAD dataset shown in red for the* Λ_{670}^{412} *products, with the PSyn synthetic case-1 dataset overlaid in solid grey, the* Λ_{670}^{412} *EMA algorithm fit to the OCEAN dataset overlaid in dashed black, and a least absolute deviation power-law fit of the NOMAD data in solid black;* ***(b)*** *NNet dataset shown in red for synthetic* $K_d(412)/K_d(670)$ *products, with the MSyn synthetic case-1 dataset overlaid in solid grey and a least absolute deviation power-law fit of the NNet data in solid black;* ***(c)*** *Validation results from the NOMAD* Λ_{670}^{412} *EMA with a one-to-one line indicated in solid black; and* ***(d)*** *Validation results from the NNet* $K_d(412)/K_d(670)$ *EMA with a one-to-one line indicated in solid black.* R^2 *values for panels* ***(a)*** *and* ***(b)*** *are derived from* $\log_{10}$*-transformed variables.* 54

2.8 *The relative percent uncertainty from the western USA portion of the GLOBC dataset (positive indicates overestimation) as a function of* a_{CDOM} *spectral slope as follows: (**a**)* $K_d(320)/K_d(780)$ *EMA using a linear fit described in Hooker et al., 2020 ; (**b**)* $K_d(412)/K_d(670)$ *EMA using an absolute deviation power-law fit; (**c**)* Λ_{780}^{320} *EMA fit to the OCEAN dataset; (**d**)* Λ_{670}^{412} *EMA fit to the OCEAN dataset; (**e**) GSM with failed retrievals shown as open circles at the top of the y-axis and the defined* a_{CDOM} *slope shown as a black diamond on the x-axis; and (**f**) QAA with the median derived* a_{CDOM} *slope shown as a black diamond on the x-axis. Data points which satisfy the 25% RMSLE criteria are shown in black, and those which exceed the criteria are shown in red.* 57

3.1 ***(a)** Anomalous diffraction approximation for* $Q_{\mathrm{ext}}(\lambda)$ *at a narrow absorption band as a function of light frequency for various* ρ *values, recreated from van de Hulst, 1957 using a lookup table; and **(b-d)** illustrative examples of* $Q_{\mathrm{ext}}(\lambda)$ *anomalies for various small-sized phytoplankton (diameters 8, 5, and 1* μ*m, respectively), with a fixed* n *of 1.0344 and spectral* n' *with a maximum value of 0.0024 at the Chl a red absorption peak. The sizes presented in panels **(b-d)** are sensitive to the selection of real and imaginary refractive index. For example, Aas, 1996 illustrates an anomalous dispersion curve for* $Q_{\mathrm{b}}(\lambda)$ *using a 1* μ*m absorbing sphere and spectral dependencies in both* n *and* n'*.* . 76

3.2 *Sampling locations for the bio-optical measurements obtained from the Tara Oceans Expedition archive in SeaBASS.* 78

3.3 *Principal components of the* $c_{\mathrm{p}}'(\lambda)$ *covariance matrix, with the nominal location of the Chl a red absorption maximum indicated by the vertical line.* . . 82

3.4 *Relationships between* a_{LH} *and the* $\boldsymbol{S}$ *matrix predictors **(a)** $\boldsymbol{P}_1{}^{(+)}$; **(b)** $\boldsymbol{P}_1{}^{(-)}$; **(c)** $\boldsymbol{P}_2{}^{(+)}$; **(d)** $\boldsymbol{P}_3{}^{(+)}$; and **(e)** $\boldsymbol{P}_3{}^{(-)}$. **(f)** Residuals against predicted* a_{LH} *values from a multivariate linear regression of the* $\boldsymbol{S}$ *matrix predictors.* . . . 84

3.5 *Median validation scatterplot relating measured (vertical) and predicted (horizontal)* a_{LH} *values from the* $\boldsymbol{S}$ *matrix predictors, with both axes on the* $\log_{10}$ *scale.* 85

3.6 *Relationships between* γ *and the* $\boldsymbol{S}$ *matrix predictors **(a)** $\boldsymbol{P}_1{}^{(+)}$; **(b)** $\boldsymbol{P}_1{}^{(-)}$; **(c)** $\boldsymbol{P}_2{}^{(+)}$; **(d)** $\boldsymbol{P}_3{}^{(+)}$; and **(e)** $\boldsymbol{P}_3{}^{(-)}$, with locally-weighted scatterplot smoothing (`lowess`) functions overlaid in solid black; the horizontal scales are* $\log_{10}$*. **(f)** Histogram estimate of the probability density function of the* γ *values derived from the* $c_{\mathrm{p}}(\lambda)$ *dataset.* 86

3.7 $\boldsymbol{S}$ *matrix predictors derived from the Tara Oceans Expedition data (red and green dots) overlaid onto satellite Chl a imagery (represented by blue shading). The coast of the Yucatan Peninsula (MODIS Aqua) is shown in panels **(a)** and **(c)**, and the Western Cape (MODIS Terra) in panels **(a)** and **(c)**. Pie charts indicate phytoplankton size fractions determined using pigment analysis. The* $\boldsymbol{S}$ *matrix predictor* $\boldsymbol{P}_2{}^{(+)}$ *is shown in panels **(a)** and **(b)**, and* $\boldsymbol{P}_3{}^{(-)}$ *is shown in panels **(c)** and **(d)**. The satellite matchup shown in panels **(b)** and **(d)** precedes the Tara measurements by an average of 5 days due to cloud cover.* 96

List of Tables

1.1 *Matchup statistics for three MODIS Aqua phytoplankton biomass proxies in Monterey Bay, California, USA and St. Helena Bay, Western Cape, South Africa where* n *is the number of valid matchups, r is the Pearson correlation coefficient and* nRMSE *is the root mean square error of the linear fit of the satellite data to the in situ data, normalized by the in situ data range.* . . . 15

2.1 *The* $\left[L_W(\lambda)\right]_{\mathrm{N}}$ *EMA power-law coefficients derived from in situ datasets.* . 46

2.2 *Performance of* $\left[L_W(\lambda)\right]_{\mathrm{N}}$ *end-member algorithms within various water body classifications. Bold* R^2 *values indicate satisfaction of* $P \leq 0.01$*, and bold RMSLE values indicate satisfaction of the 25% algorithm uncertainty target for next generation calibration and validation activities in inland waters.* . . 50

Abstract

Advances in bio-optics for observing aquatic ecosystems

by

Henry F. Houskeeper

The generalized bio-optical transition from oligotrophic blue to mesotrophic green waters has enabled remote measurement of oceanic ecosystems under conditions in which the optically relevant water mass constituents covary with the primary phytoplankton photopigment, chlorophyll *a* (Chl *a*). This assumption, known as case-1, is generally incorrect within coastal and inland water bodies, where external factors, such as riverine discharge and sediment resuspension, also modify the concentrations of organic and inorganic constituents beyond the ranges anticipated for natural oceanic ecosystems. New approaches to characterize aquatic ecosystems without relying on case-1 criteria are proposed here using measurements of apparent and inherent optical properties. The methods applied in this work include bio-optical modeling and analyses of *in situ*, airborne, and satellite radiometric measurements of oceanic, coastal, and inland water ecosystems. Key findings are that: satellite phytoplankton datasets underestimate phytoplankton biomass in coastal ocean regions when default atmospheric correction methods are applied (Chapter 1); the expansion of spectral range for above-water radiometric measurements improves the accuracy of colored dissolved organic matter algorithms (Chapter 2); and hyperspectral beam-attenuation meters can accurately predict Chl *a* without relying on empirical relationships between organic and inorganic particle populations (Chapter 3). The bio-optical advances presented here improve measurement capabilities for optically complex waters and are relevant to global (i.e., oceanic, coastal, and inland) water bodies. Finally, this work supports the perspective that bio-optical approaches should be consistent with electromagnetic theory or other fundamental principles — rather than requiring a partition between waters that satisfy versus fail empirical open ocean criteria — to incorporate a greater diversity of global aquatic ecosystems.

To Jamie, Robert, and Sam.

Acknowledgments

I am grateful to my advisor Raphael Kudela and to my other committee members Chris Edwards, Stan Hooker, and John Ryan. Their feedback helped me to organize my research ideas and complete the work presented here. In particular, I appreciate the unique learning opportunities that occurred during field campaigns and the valuable collaborations that followed these activities.

I also received useful advice on data science from David Draper, science communication from Rob Irion, and bio-optics from the instructors at the NASA 2019 Ocean Optics Summer Course. I benefited from the kindness and expertise of my peers, including the past and current members of the Kudela Lab, my fellow mentors at the NASA Student Airborne Research Program, and my classmates at the NASA 2019 Ocean Optics Summer Course. Finally, I'm grateful to Christina Ravelo, who was a role model early in my graduate education.

This work was made possible by support from (in alphabetical order): the California State Water Resources Control Board (grant number A17-0553), the Friends of the Long Marine Lab (Student Research and Education award), the International Ocean Colour Coordinating Group (IOCCG; student travel award), the NASA C-HARRIER campaign (grant number NNX17AK89G), the NASA Ocean Carbon and Biogeochemistry program (OCB; student travel award), the NASA Ocean Optics Course (grant number 80NSSC19K0812), and the NOAA Central and Northern California Ocean Observing System (CeNCOOS; grant number NA16NOS0120021).

The first chapter of this dissertation includes a reprinted version of previously published work:

> **H.F. Houskeeper** and R.M. Kudela. Ocean color quality control masks contain the high phytoplankton fraction of coastal ocean observations. *Remote Sensing, Special Issue: Coastal waters monitoring using remote sensing technology*, vol. 11, no. 18, pp. 2167, 2019. `https://doi.org/10.3390/rs11182167`

Raphael M. Kudela supervised this project and is listed as a co-author, but the analysis and writing of this paper were substantively my own.

The second chapter of this dissertation contains a reprint of submitted work:

> **H.F. Houskeeper**, S.B. Hooker, and R.M. Kudela. Spectral range within global $a_{\mathrm{CDOM}}(440)$ algorithms for oceanic, coastal, and inland waters with application to airborne measurements. Submitted February 2020 to *Remote Sensing of Environment*.

Raphael M. Kudela and Stanford B. Hooker supervised this project and are listed as co-authors. Again, the analysis and writing of this manuscript were substantively my own.

Finally, the third chapter of this dissertation contains a reprint of submitted work:

> **H.F. Houskeeper**, D. Draper, R.M. Kudela, and E. Boss. Chlorophyll absorption and phytoplankton size information inferred from hyperspectral particulate beam-attenuation. Submitted May 2020 to *Applied Optics*.

Raphael M. Kudela and Emmanuel Boss supervised this project and are listed as co-authors. David Draper advised on data science and is also listed as a co-author. The analysis and writing of this manuscript were substantively my own.

When possible, symbols and abbreviations have been chosen to be uniform across separate chapters of this dissertation. However, the range of topics discussed and the desire to use common symbols (e.g., the surface reflectance, the Pearson correlation coefficient, and the phase lag parameterization for anomalous diffraction theory are all commonly denoted ρ), require that the definitions and symbology presented within each chapter are separately defined.

Introduction

The basic premise that oligotrophic blue waters transition to mesotrophic green waters [1] has enabled remote measurement of phytoplankton concentrations and led to the development of satellite instruments dedicated to recording ocean color [2, 3]. Stable relationships between the surface ocean's color and its biogeochemistry depend on the covariance of the ocean's optically relevant constituents, a condition defined in optical oceanography as case-1 or optically simple [4]. Satellite remote sensing often adds an additional criteria – that the ocean is a dark target at near-infrared (NIR) wavelengths, i.e., the black-pixel assumption [5] – which helps to separate the atmospheric and oceanic signals.

The success of the case-1 and black-pixel approximations has enabled ocean color tools to measure the open ocean, but not coastal or inland (i.e., case-2) environments, in which the natural range in the concentrations of suspended particles is expanded. For example, rivers and wave energy add suspended sediments to the coastal ocean, and increased nutrients fuel dense phytoplankton blooms. Elevated near-surface particle loads brighten satellite NIR measurements and prevent accurate atmospheric correction, often resulting in non-physical (negative) derived reflectances. Chapter 1 evaluates the decrease in satellite-recorded biomass due to the loss of high-phytoplankton pixels by comparing fluorescence between pixels that fail versus satisfy standard atmospheric correction processing. Differences in fluorescence, a biomass proxy that doesn't require atmospheric correction [6], reveal that high phytoplankton measurements are underrepresented in satellite datasets from two productive coastal regions.

Legacy remote sensing methods for measuring the primary phytoplankton photopigment chlorophyll a (Chl a) target the high absorption of blue light by Chl a, which manifests in the ocean color perspective as the transition from oligotrophic blue to mesotrophic green waters. This simplification requires that the concentrations of other constituents that also

strongly absorb blue light are either invariant in the marine environment or are covarying with Chl a. Colored dissolved organic matter (CDOM) strongly absorbs blue light and is correlated with Chl a in the marine environment (its main oceanic sources are phytoplankton secretion and decomposition). In coastal waters, CDOM is also added from rivers or land, and uncertainty in its absorption at 440 nm, $a_{\mathrm{CDOM}}(440)$, limits the usefulness of standard ocean color Chl a algorithms. Chapter 2 describes a band-ratio algorithm for remote sensing of $a_{\mathrm{CDOM}}(440)$ that is based on the perspective that increasing spectral separation of the band-ratio wavelengths increases the dynamic range and improves the algorithm's accuracy in global (i.e., oceanic, coastal, and inland) waters [7, 8]. The proposed algorithm is shown to be consistent with case-1 empiricisms but also to be valid for case-2 conditions. The results suggest that algorithms using more spectrally separate wavebands outside of the visible domain (e.g., ultraviolet and NIR wavelengths) are less sensitive to the operational partition of case-1 and case-2 water types.

In-water instruments equipped with an independent light source can directly measure the absorption, scattering, and attenuation properties of seawater, i.e., the inherent optical properties [9]. These instruments are generally deployed as packages of combined sensors (e.g., packages that contain separate beam-attenuation and absorption meters) in order to adequately characterize the marine environment. For example, spectral absorption meters can reliably estimate Chl a [10] but cannot describe the particle size distribution. Spectral beam-attenuation (beam-c) meters enable estimation of the particle size distribution but cannot distinguish between organic (e.g., phytoplankton) and inorganic (e.g., sediment) particle types [11]. The combination of multiple sensors raises instrument costs and increases the risk that an instrument fails (e.g., absorption meters are more prone to fouling than beam-c meters). Chapter 3 describes a new method to accurately estimate Chl a absorption using only hyperspectral beam-c measurements based on changes in the complex index of refraction in the spectral vicinity of the Chl a red absorption peak [12].

Future technological advances are anticipated that will expand the spectral range and resolution of *in situ*, airborne, and satellite sensors to mitigate the current challenges for remote sensing of coastal and inland waters, but corresponding algorithm improvements that move beyond the black-pixel and case-1 assumptions will be required. The following dissertation outlines fundamental challenges to remote sensing of coastal and inland waters, and it proposes new methods in optical oceanography intended to overcome some of the current problems.

References

[1] C. S. Yentsch, "The influence of phytoplankton pigments on the colour of sea water," *Deep Sea Research (1953)*, vol. 7, no. 1, pp. 1 – 9, 1960.

[2] M. R. Abbott and D. B. Chelton, "Advances in passive remote sensing of the ocean," *Reviews of Geophysics*, vol. 29, no. S2, pp. 571–589, 1991.

[3] S. B. Hooker and W. E. Esaias, "An overview of the seawifs project," *Eos, Transactions American Geophysical Union*, vol. 74, no. 21, pp. 241–246, 1993.

[4] A. Morel and L. Prieur, "Analysis of variations in ocean color," *Limnology and Oceanography*, vol. 22, no. 4, pp. 709–722, 1977.

[5] D. A. Siegel, M. Wang, S. Maritorena, and W. Robinson, "Atmospheric correction of satellite ocean color imagery: the black pixel assumption," *Applied Optics*, vol. 39, no. 21, pp. 3582–3591, 2000.

[6] J. F. Gower and S. A. King, "Use of satellite images of chlorophyll fluorescence to monitor the spring bloom in coastal waters," *International Journal of Remote Sensing*, vol. 33, no. 23, pp. 7469–7481, 2012.

[7] S. B. Hooker, J. H. Morrow, and A. Matsuoka, "Apparent optical properties of the canadian beaufort sea–part 2: The 1% and 1 cm perspective in deriving and validating aop data products," *Biogeosciences*, vol. 10, no. 7, pp. 4511–4527, 2013.

[8] S. B. Hooker, A. Matsuoka, R. M. Kudela, Y. Yamashita, K. Suzuki, and H. F. Houskeeper, "A global end-member approach to derive $a_{\mathrm{CDOM}}(440)$ from near-surface optical measurements," *Biogeosciences*, vol. 17, no. 2, pp. 475–497, 2020.

[9] R. W. Preisendorfer and J. E. Tyler, "The measurement of light in natural waters; radiometric concepts and optical properties," tech. rep., Scripps Institution Of Oceanography La Jolla, CA Visibility Lab, 1958.

[10] C. S. Roesler and A. H. Barnard, "Optical proxy for phytoplankton biomass in the absence of photophysiology: Rethinking the absorption line height," *Methods in Oceanography*, vol. 7, pp. 79–94, 2013.

[11] J. C. Kitchen, J. R. V. Zaneveld, and H. Pak, "Effect of particle size distribution and chlorophyll content on beam attenuation spectra," *Applied Optics*, vol. 21, no. 21, pp. 3913–3918, 1982.

[12] H. Van de Hulst, *Light scattering by small particles.* Wiley, New York, 1957.

Chapter 1

Atmospheric correction effects on coastal ocean satellite datasets

Overview: Atmospheric correction of satellite imagery is more difficult in coastal (versus oceanic) waters due in part to the presence of non-maritime aerosols, as well as increases in the natural ranges of various oceanic constituents, such as algal and non-algal particles. Higher suspended particle loads elevate water-leaving reflectances at near-infrared wavelengths, in which the ocean is assumed to be a dark target. The increased near-infrared signals can cause atmospheric correction failures during image processing and result in the removal of bright pixels. This chapter compares biomass measurements for satellite pixels that satisfy versus fail standard atmospheric correction processing. Many users of satellite imagery rely on temporal and spatial composites, and this work provides suggestions for determining whether composite products in coastal waters are strongly influenced by atmospheric correction failures.

The following contains material that was published in *Remote Sensing* in September 2019:

Ocean color quality control masks contain the high-phytoplankton fraction of coastal ocean observations

Henry F. Houskeeper & Raphael M. Kudela

Abstract: Satellite estimation of oceanic chlorophyll *a* content has enabled characterization of global phytoplankton stocks, but the quality of retrievals for many ocean color products (including chlorophyll *a*) degrades with increasing particle concentrations. Quality control of ocean color products is achieved primarily through the application of masks based on standard thresholds designed to identify suspect or low-quality retrievals. This study compares the masked and unmasked fractions of ocean color datasets from two Eastern Boundary Current upwelling ecosystems (the California and Benguela Current Systems) using satellite proxies for phytoplankton biomass that are applicable to satellite imagery without correction for atmospheric aerosols. Evaluation of the differences between the masked and unmasked fractions indicates that high biomass observations are preferentially masked in National Aeronautics and Space Administration (NASA) ocean color datasets as a result of decreased retrieval quality for waters with high concentrations of phytoplankton. This study tests whether dataset modification persists into the default composite data tier commonly disseminated to science end-users. Further, this study suggests that statistics describing a dataset's masked fraction can be helpful in assessing the quality of a composite dataset and in determining the extent to which retrieval quality is linked to biological processes in a given study region.

1.1 Introduction

Ocean color remote sensing has greatly improved our ability to monitor global scale biological processes of ocean systems [1,2], but the potential for conventional satellite ocean color tools to characterize coastal ecosystems is limited by the assumptions used in various algorithms, for example that diverse phytoplankton communities match global bio-optical relationships [3] or that backscattered light from particles does not strongly elevate signals at near-infrared (NIR) wavelengths [4]. Although these assumptions are often invalid for coastal waters [5], satellite assessment of coastal marine ecosystems is an area of intense focus in part because of reported increases in the frequency of coastal phytoplankton blooms that are harmful to humans and wildlife [6–10].

Fundamental challenges for ocean color remote sensing of coastal marine ecosystems arise from the increased complexity of constituents, as well as the difficulty in separating

the atmospheric and oceanic signals. Next-generation ocean color satellite missions aim to overcome these difficulties in part by increasing the sensor spectral resolutions to resolve variability in phytoplankton pigmentation [11] and by increasing the spectral range to discern the effects of absorbing and non-absorbing aerosols [12]. Nonetheless, data from existing satellite platforms is presently required to assess coastal marine ecosystems. When next-generation sensing platforms become operational, interpreting legacy measurements of coastal waters and detecting regional biases will still be necessary in order to construct and interpret climate-quality data records [13–15]. For coastal targets, constructing climate-quality datasets will require an approach to maintain atmospheric correction efficacy across variable phytoplankton concentrations.

Conventional approaches for atmospheric correction of ocean color satellite imagery take advantage of the strong light absorption by water at NIR wavelengths to estimate that the water-leaving radiance, $L_W(\lambda)$, is negligible in the NIR [16]. After removal of glint and white capping effects, the derived top-of-atmosphere NIR radiance is attributed to the atmospheric contributions by Rayleigh scattering, aerosol scattering and multiple interactions between Rayleigh and aerosol scattering [17], which enables the estimation of aerosol thickness. The approximation that $L_W(\lambda)$ is zero at NIR wavelengths, or the black-pixel assumption [4], is often incorrect, frequently so in coastal waters where high near-surface particle loads (organic or inorganic) can strongly backscatter light. Because top-of-atmosphere signals at NIR wavelengths are attributed to atmospheric constituents, aquatic contributions cause overestimation of atmospheric aerosols and thus result in incorrect (often negative) derivation of $L_W(\lambda)$ in the visible domain, particularly in the blue bands used for, among others, chlorophyll *a* (Chl *a*) derivation [18]. As a result, ocean color satellite retrievals are more problematic for water masses with high particle loads (including phytoplankton cells), and high biomass pixels are frequently removed during quality control processing of satellite datasets (Fig. 1.1).

Alternate atmospheric correction methods have been developed with the goal of improving $L_W(\lambda)$ retrieval in coastal waters, for example by estimating aerosol contributions from longer (i.e., short-wave infrared) wavebands [19], by assuming stable NIR reflectance ratios within a scene [20], or by neural network determination of atmospheric contributions [21]. Alternate methods for atmospheric correction of coastal imagery have generally improved performance of nearshore ocean color retrievals compared with the conventional (NIR-based) methods [22], although users must still decide when these methods are prefer-

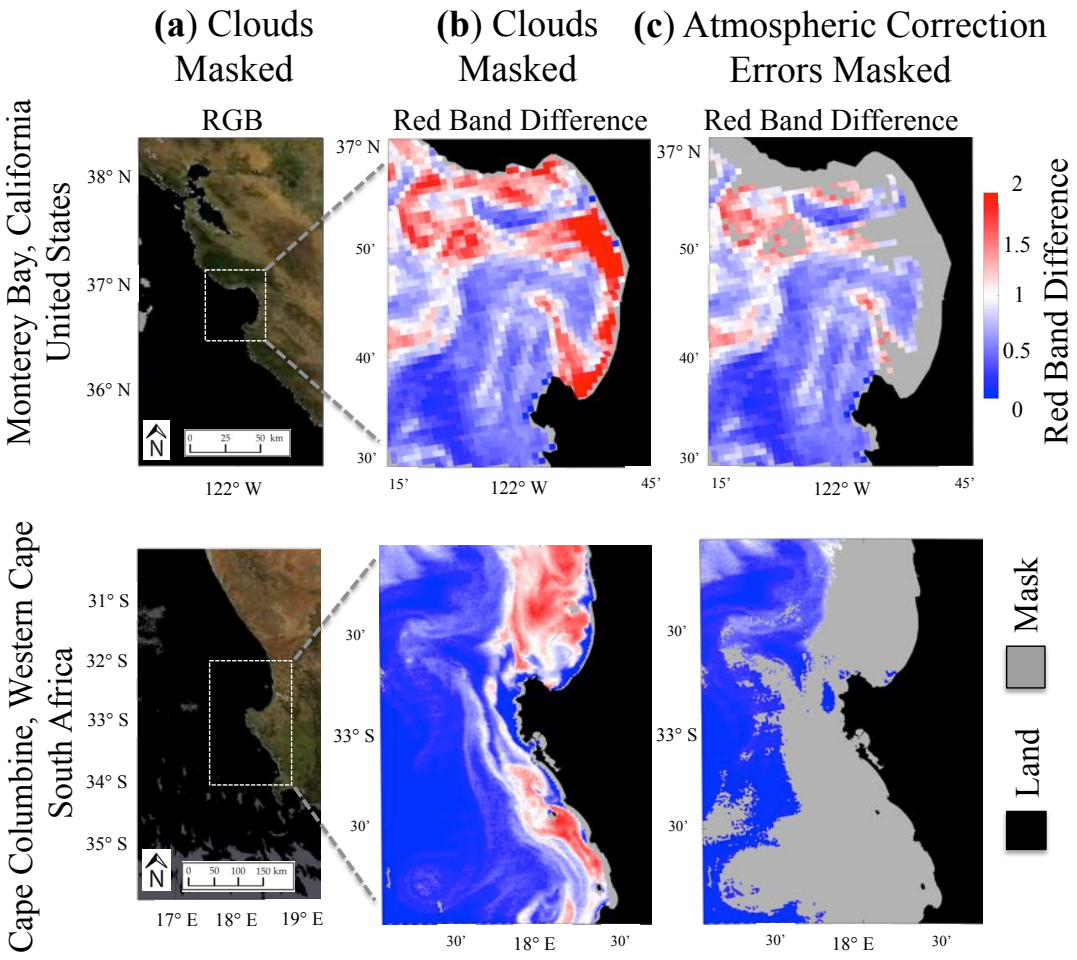

Figure 1.1: *MODIS Aqua imagery of phytoplankton blooms obtained on 26 October 2016 in Monterey Bay, California (upper) and 27 September 2011 near Cape Columbine, Western Cape (lower).* ***(a)*** *Pseudo-true color images with clouds masked in grey;* ***(b)*** *Red Band Difference algorithm (a proxy for phytoplankton biomass) with clouds masked in grey; and* ***(c)*** *Red Band Difference algorithm with clouds and suspect atmospheric correction (defined as maximum aerosol iterations reached and low water-leaving radiances) masked in grey.*

able given tradeoffs of noisy wavebands, non-analytical solutions, and increased difficulty in obtaining alternate processing for National Aeronautics and Space Administration (NASA) imagery. Another explanation for widespread use of default NIR-corrected imagery, evaluated in detail within this work, is that the degradation of satellite datasets by atmospheric correction errors is obscured by the default quality control masks.

Quality control of NASA ocean color imagery is achieved in part through flag assignments that trigger masking (i.e., removal) of individual pixels that do not satisfy predefined thresholds. Increasingly rigorous flag criteria remove observations from sequential data tiers based on the quality requirements of each tier's expected end-users. Pixels within atmospherically-corrected imagery (Level 2 data tier, L2) are masked by default when the derivation of meaningful products is severely inhibited, for example when the sensor saturates or is viewing land or clouds. L2 datasets contain shifting pixel coordinates, frequent data gaps (i.e., from clouds), and large file sizes inconvenient for users requiring continuous or less computationally expensive products [23]. To satisfy these user needs, statistical composites of geophysical variables binned in space and time (Level 3 data tier, L3) are provided by the NASA Ocean Biology Processing Group (OBPG) and are valuable to users beyond the ocean color community, for example as inputs into biogeochemical models. In order to provide higher quality composites for a larger end-user community, the default masks applied during L3 processing are more rigorous than during L2 processing, for example removing observations flagged for likely or known errors in atmospheric correction. Spatial distortions may also arise during compositing, and although not evaluated here, are likewise relevant to L3 end-users [24].

In this study, we compare estimates of phytoplankton biomass – obtained without applying aerosol corrections – between observations that satisfy (versus fail) standard quality control thresholds for two Eastern Boundary Current (EBC) ecosystems. We assess whether observations that satisfy quality control methods, hereafter the masked fraction, provide an unbiased perspective of phytoplankton biomass in coastal ecosystems. Finally, we provide examples for L3 end-users to consider when determining whether to apply standard composite products for a specific study region.

1.2 Methods

1.2.1 Site selection

The mid-latitude eastern margins (or EBCs) of the world's oceans are regions of heightened biological primary production due to coastal upwelling, or the wind-driven transport of nutrient-rich subsurface waters to the illuminated surface layer. Heightened nutrient availability, coupled with the persistence of seed stocks supplied from the shelf and retained

in the lee of headlands, supports high phytoplankton concentrations that periodically form blooms, some of which are harmful to humans and wildlife [7, 25]. Here we consider EBC ecosystems of Monterey Bay (MB), California, United States and St. Helena Bay (SHB), Western Cape, South Africa. MB is within a marine sanctuary and is partially sheltered from the predominant alongshore winds of the central California Current System (CCS) by coastline geometry. Phytoplankton in the region follow a distinct climatology, with spring onset of upwelling-favorable winds supporting diatom-rich phytoplankton blooms, followed by a mid-summer reduction in phytoplankton, possibly associated with rapid advection to offshore waters [26]. Fall relaxation of upwelling-favorable winds and increasing vertical stratification of the surface layer facilitate a community shift towards dinoflagellates, which periodically form dense red tides with concentrations that may reach or exceed those of spring diatom blooms [25].

SHB is an upwelling ecosystem in the lee of Cape Columbine within the southern Benguela Current System (BCS). The region's proximity to the Cape Peninsula upwelling cell, its shelter from the lee, and its widened shelf support persistently high phytoplankton production [27]. In addition, sea surface temperature is relatively high within SHB compared with other EBCs, allowing elevated phytoplankton populations to persist throughout all seasons [28]. As with MB, phytoplankton succession in the southern BCS, including within SHB, is primarily controlled by the intensification and relaxation of alongshore winds, with characteristic diatom and dinoflagellate regimes dominating in the spring and fall, respectively [29].

1.2.2 Atmospheric dataset

Climatological datasets for Ångstrom exponent and aerosol optical depth (500 nm) were obtained from the Aerosol Robotic Network (AERONET; `aeronet.gsfc.nasa.gov`) for Monterey, California, United States (36.59°N, 121.85°W) and Simonstown, Western Cape, South Africa (34.18°S, 18.43°E), shown in Fig. 1.2. The Monterey AERONET site is located to the southeast of MB and separated from a nearby agricultural region by a coastal mountain range, although diurnal sea breeze north of this range may increase mixing between terrestrial and marine air masses. The Simonstown AERONET site is located on the eastern slope of the Cape Peninsula, roughly 150 km south of Cape Columbine. Predominant windstress is equatorward (towards SHB) with summer intensification [30].

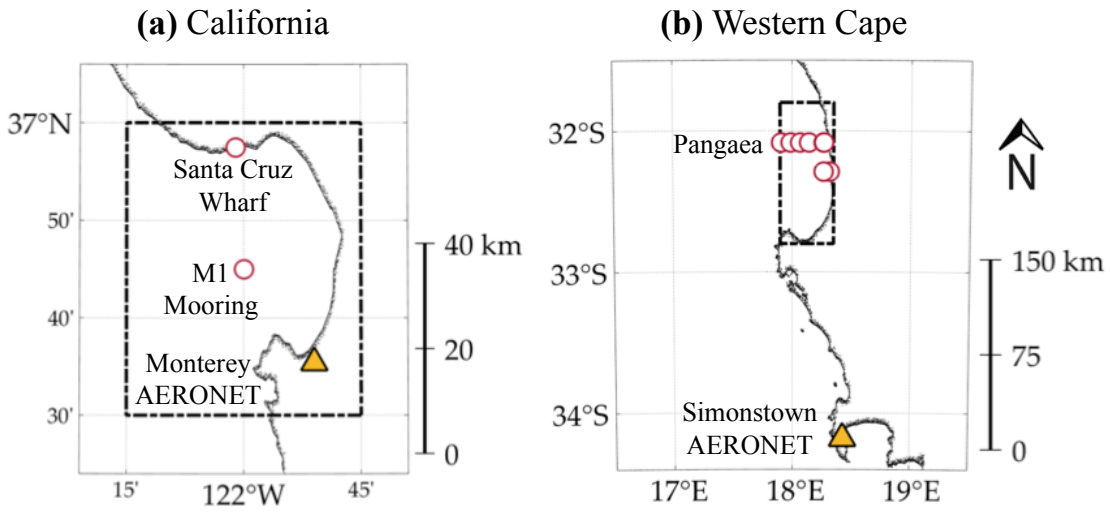

Figure 1.2: *Location of biological and atmospheric measurements used in this study for* ***(a)*** *Monterey Bay, California and* ***(b)*** *St. Helena Bay, Western Cape. Chl a and fluorometer measurements were obtained at sites denoted with red circles, and atmospheric measurements at sites denoted with orange triangles. Regions of satellite L2 datasets used in this analysis are indicated by dashed black lines.*

1.2.3 Biological field data

Weekly fluorometric Chl *a* measurements were obtained from the Southern California Coastal Ocean Observing System portal (`sccoos.org/data/habs/`) for the Santa Cruz Wharf (SCW; 36.958°N, 122.017°W) in northern MB. Daily mean *in situ* surface (1 m depth) fluorescence measurements were obtained from a HydroScat-2 fluorometer (HOBI Labs) mounted on an oceanographic mooring in central MB (M1; 36.750°N, 122.000°W), maintained by the Monterey Bay Aquarium Research Institute (MBARI; `mbari.org`). Measurements of Chl *a* within SHB were obtained from the European Space Agency (ESA) Ocean Colour Climate Change Initiative (OC-CCI) dataset, available on the Pangaea portal (`doi.pangaea.de`). Only fluorometric measurements (not pigment analyses) were used for matchups with satellite products for SHB since the majority of Chl *a* estimates in the database were based on that methodology. The locations of the sampling sites are indicated in Fig. 1.2.

1.2.4 Satellite data

MODIS Aqua (MODISA) calibrated, non-atmospherically corrected imagery (L1A) were obtained from the NASA Ocean Color website (`oceancolor.gsfc.nasa.gov`) for dates spanning July 2002 to September 2018 within MB (36.50 – 37.00°N; 121.75 – 122.25°W) and SHB (31.80 – 32.80°S; 17.90 – 18.35°E) for matchup validation and for analysis of L2 datasets (Fig. 1.2). Surface reflectances, $\rho_s(\lambda)$, were obtained for both regions from geo-referenced and atmospherically corrected imagery produced using the NASA OBPG software SeaDAS (version 7.5) with observations removed if viewing land or clouds or for non-physical retrievals, such as $\rho_s(\lambda)$ outside of the range 0–1. Spatial subsets used for comparison with AERONET results were selected from the MB and SHB domains based on the AERONET site locations and the local topography (south of 36.65°N and east of 121.92°W for MB; south of 32.10°S and east of 18.10°E for SHB). Datasets for L3 analyses were obtained for the same timeframe and for similar regions in MB (36.50 – 37.00°N; 121.75 – 122.25°W) and SHB (32.20 – 32.80°S; 17.90 – 18.35°E), as well as for two nearby transects placed in regions with relatively north-south coastlines of the CCS (37.10 – 37.50°N, 122.40 – 123.40°W) and BCS (31.80 – 32.20°S, 17.30 – 18.35°E). Processing for the L3 products applied additional SeaDAS software `l2bin` and `l2mapgen` to form daily, 4 km standard map grid composites.

Quality control flags were assigned to all pixels during the L2 processing chain according to standard OBPG L2 flag thresholds. Masks were applied (i.e., the data was removed) for one identical set of the L2 data if flag assignments indicated likely errors in atmospheric correction. Flags chosen included warnings for low $L_W(\lambda)$, maximum iterations exceeded during atmospheric correction processing, and out-of-range spectral slope of derived aerosol radiances. This combination of flags will be hereafter referred to as AC flags. More detailed information on thresholds and applications of default flags can be found in the SeaWiFS postlaunch documentation [31].

Neural network Chl *a* estimates were included for evaluating satellite products using processing tools provided by the Coast Colour project (`coastcolour.org`). In short, MODISA images were georeferenced and calibrated using SeaDAS. Atmospheric correction and derivation of Chl *a* was performed for each image using the Sentinel Application Platform (SNAP; `step.esa.int`) with the Case-2 Regional Coast Colour (C2RCC; `brockmann-consult.de`) plugin.

1.2.5 Remote estimation of phytoplankton biomass

Remote measurements of the spectral radiance anomalies generated by sun-induced fluorescence of the Chl *a* molecule have been applied as proxies for phytoplankton biomass for over four decades [32,33]. The fluorescence line height approach (FLH), which subtracts a red to NIR baseline from the Chl *a* fluorescence peak to correct for brightness effects, is the most widely used of these satellite tools. FLH has been proposed to complement traditional Chl *a* satellite algorithms in high biomass coastal waters [34, 35] and a normalized version (nFLH) is disseminated in standard L2 and L3 OBPG data. Another FLH-type method, the Red Band Difference algorithm (RBD), subtracts the signal derived at the nearest shorter wavelength from the signal measured at the Chl *a* fluorescence peak and was chosen for this work because of its relative robustness to sediment effects [36]. Fluorescence line height products, including RBD, may be derived from top-of-atmosphere (TOA) measurements, thus bypassing the potential errors arising during the atmospheric correction procedure. Here we derive RBD from $\rho_s(\lambda)$, a partial atmospheric correction product that accounts for Rayleigh but not aerosol effects, defined as follows:

$$\rho_s(\lambda) \;=\; \left(\frac{\pi}{F_0\,\mu_0}\right)\left(\frac{L_{TOA}}{t_{solar}\,t_{sensor}}\right)\left(t'_{solar}\,t'_{sensor}\,t'_{O_2}\,t'_{H_2O}\right)^{-1}\,, \tag{1.1}$$

where F_0 is the solar downward irradiance, μ_0 is the cosine of the solar zenith angle, t and t' are the direct and diffuse atmospheric transmittances, respectively, for the sun-to-surface and surface-to-sensor path lengths and for the atmospheric effects of oxygen and water vapor. RBD was derived as the difference between MODISA surface reflectances, as follows:

$$\mathrm{RBD} \;=\; \rho_\mathrm{s}(678\,\mathrm{nm}) \;-\; \rho_\mathrm{s}(667\,\mathrm{nm})\,, \tag{1.2}$$

where $\rho_s(678\,\mathrm{nm})$ corresponds to the Chl *a* fluorescence maximum (nominally 683 nm) and $\rho_s(667\,\mathrm{nm})$ provides an adjacent baseline to account for overall spectral brightness.

1.2.6 Matchups, climatological averages, and error statistics

Validation statistics for all satellite products were derived using only same-day, 3×3 pixel median matchups centered on the in-water samples due to high spatial and temporal variability at the matchup sites. Climatologies were derived as the mean monthly values for each region. Composite datasets were compared using standardized bias (SB), defined

as follows:

$$\mathrm{SB} \;=\; \left(\frac{\overline{X} \;-\; \overline{Y}}{\sigma_{_Y}}\right), \tag{1.3}$$

where $\overline{X}$ and $\overline{Y}$ correspond to the mean RBD composite values for datasets with AC flagged pixels masked or not masked, respectively. The difference in composite means were normalized by the standard deviation of the unmasked (AC flagged pixels retained) dataset.

1.3 Results

1.3.1 Association between RBD and phytoplankton biomass

Satellite matchups at both MB and SHB indicate that the RBD algorithm associated more strongly (based on Pearson correlation) with *in situ* proxies for phytoplankton concentrations than either a standard NASA blue-green Chl *a* algorithm (OC3M) or a neural-network-based Chl *a* algorithm (Coast Colour), shown in (Table 1.1). An attempt to model surface Chl *a* from satellite measurements (using a linear, least-squares approach) resulted in higher accuracy of the RBD method versus the other remote products in most instances. More valid matchups were possible at the M1 buoy (MB) location because of increased distance from land and because of the greater number of *in situ* records. More valid matchups were also possible for RBD versus OC3M because the use of the $\rho_s(\lambda)$ product prevented retrieval failures related to aerosol overestimation. Visual inspection of Coast Colour matchup scenes suggested that common culprits for the reduction of valid matchups were both the out-of-range inputs to the atmospheric neural network as well as incorrect cloud mask assignments.

The lowest Pearson correlation coefficients for all remote products occurred at the Santa Cruz Wharf (in northern MB), where matchups were more difficult because of the increased spatial and temporal heterogeneity of the near-shore environment, and because fewer adjacent oceanic pixels were available due to obstruction by shore. The Pearson correlation coefficient for the RBD product was highest relative to the other products at this site, suggesting that the nearshore matchups were also strongly affected by resuspended sediment, riverine discharge or terrestrial aerosols given the relative robustness of RBD to signal brightening effects and to absorption by riverine constituents, such as colored dissolved organic material. Error between modeled and *in situ* Chl *a* for the OC3M algorithm was lowest at the M1 buoy (in central MB), the site with the greatest prevalence of optically

Table 1.1: *Matchup statistics for three MODIS Aqua phytoplankton biomass proxies in Monterey Bay, California, USA and St. Helena Bay, Western Cape, South Africa where* n *is the number of valid matchups, r is the Pearson correlation coefficient and* nRMSE *is the root mean square error of the linear fit of the satellite data to the in situ data, normalized by the in situ data range.*

Location	**Product**	**n**	***r***	**nRMSE**
Monterey Bay,	Red Band Difference	1012	0.4190	14.4%
California	Blue-Green Chl *a* Algorithm[1]	773	0.3344	14.8%
M1 Buoy	Neural Network Chl *a* Algorithm[2]	840	0.2179	17.0%
Monterey Bay,	Red Band Difference	361	0.2055	14.1%
California	Blue-Green Chl *a* Algorithm[1]	8	0.0359	28.8%
Santa Cruz Wharf	Neural Network Chl *a* Algorithm[2]	132	0.0488	17.3%
St. Helena Bay,	Red Band Difference	90	0.5493	19.2%
South Africa	Blue-Green Chl *a* Algorithm[1]	74	0.4283	20.8%
Various Locations	Neural Network Chl *a* Algorithm[2]	75	0.3973	19.9%

[1]NASA OBPG standard Chl *a* product (OC3M algorithm). [2]Coast Colour standard Chl *a* product.

simple (case-1) water types among our validation sets. The highest Pearson correlation coefficient for each product was derived from SHB matchups, with RBD showing the strongest association with *in situ* Chl *a* among the evaluated remote products. The SHB matchups were unique from the two MB sites in that the *in situ* measurements were obtained by ship at various distances from shore, allowing the SHB matchup dataset to include a wider diversity of water types than either the wharf or mooring datasets in MB.

1.3.2 Climatology of atmospheric correction flags

Comparison of AERONET and satellite climatologies did not reveal similarities between atmospheric constituents and satellite atmospheric correction flags. In particular, Pearson correlation coefficients were non-significant between AC flags and either aerosol optical depths or Ångstrom exponents for the Monterey or Simonstown sites, respectively. AC flag assignments indicated positive, significant correlation with RBD in MB ($P(r) < 0.01$), and positive although non-significant correlation in SHB. AERONET sites revealed local maxima of both aerosol optical depths and Ångstrom exponents during summer that did

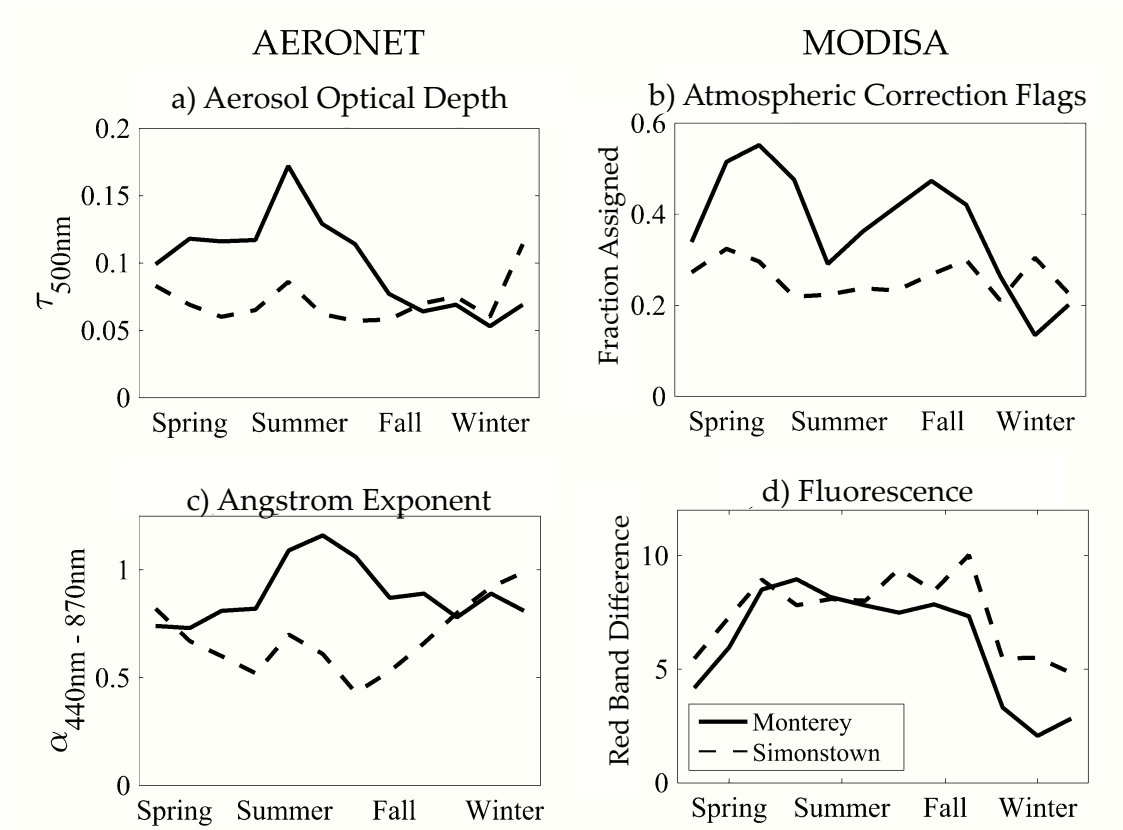

Figure 1.3: *Climatology of atmospheric and satellite products near Monterey Bay and St. Helena Bay, shifted for the phase timing of northern and southern hemispheres (Monterey: solid black line, March-February; Simonstown: dashed black line, September-August). Panels indicate:* ***(a)*** *Aerosol optical depth;* ***(b)*** *assignment frequency of AC flags;* ***(c)*** *Ångstrom exponent; and* ***(d)*** *Red Band Difference.*

not correspond to spikes in AC flag assignments during the same month. The Monterey site summertime peak in atmospheric complexity coincided with a local minimum in AC flag assignments, which suggests that pollution and aerosol characteristic of summer months are not dominant mechanisms for low atmospheric correction efficacy in this sample. The seasonality of AC flags in the Simonstown region was more uniform than in the Monterey region but was similarly incongruous with the AERONET results (Fig. 1.3).

Although not elucidated by this analysis, the relatively higher Ångstrom exponent and aerosol optical depths measured from spring through fall at the Monterey AERONET station may in part be responsible for the decreased performance in the OC3M matchups

within MB compared with SHB. Other differences between the Monterey and Simonstown results may be due to the AERONET locations, with the Monterey site nearer the sheltered retentive zone in southern MB, and the Simonstown site located near a headland with more exposure to wind and further from the SHB study site.

1.3.3 Impact of atmospheric correction masks on RBD datasets

Satellite retrievals with higher RBD values were more frequently assigned AC flags, with 33.8% and 33.1% of observations assigned AC flags within the upper quartile of RBD data, compared with only 5.3% and 8.2% within the lower quartile at MB and SHB, respectively. As a result, masking of AC flagged retrievals decreased the right-hand tails of the RBD dataset distributions at both sites (Fig. 1.4). The resulting masked fractions describe lower average RBD values (mean: -18.1% and -11.0%; median: -14.7% and -8.0%) with less variability (standard deviation: -13.3% and -13.0%) compared with the initial (AC flagged pixels not masked) datasets for MB and SHB, respectively.

L3 spatial composites (4 km, 1 day) were compared between the masked and unmasked RBD datasets using SB. The greatest magnitude in SB occurred in the lee of the retentive features that outline MB and SHB, regions prone to frequent phytoplankton blooms due to recirculation of water masses and protection from offshore advection during upwelling pulses (Fig. 1.5). SB was more negative in near-shore composite grids within the BCS compared to the CCS, with near-shore RBD composites approximately a fifth of a standard deviation lower in the masked versus unmasked dataset. Similar differences were observed in the SHB and MB regions.

Composites within the CCS and BCS transects were partitioned by longitude to derive SB as a function of distance from shore. Transect regions were chosen to be adjacent to relatively north-south coastlines and were each greater than 50 km equatorward of the largest nearby coastline points (e.g., Point Reyes and Cape Columbine). The BCS transect showed more rapid improvement of data quality as a function of distance from shore, with the magnitude in SB reduced to within a tenth of one standard deviation beyond approximately 30 km from shore, compared to within 15 km from shore in the CCS transect (Fig. 1.6). The differences in the offshore persistence of SB may be due in part to regional circulation, for example due to differences in coastline geometries and proximities to upwelling hot spots. Beyond approximately 50 km from shore, both transects indicate

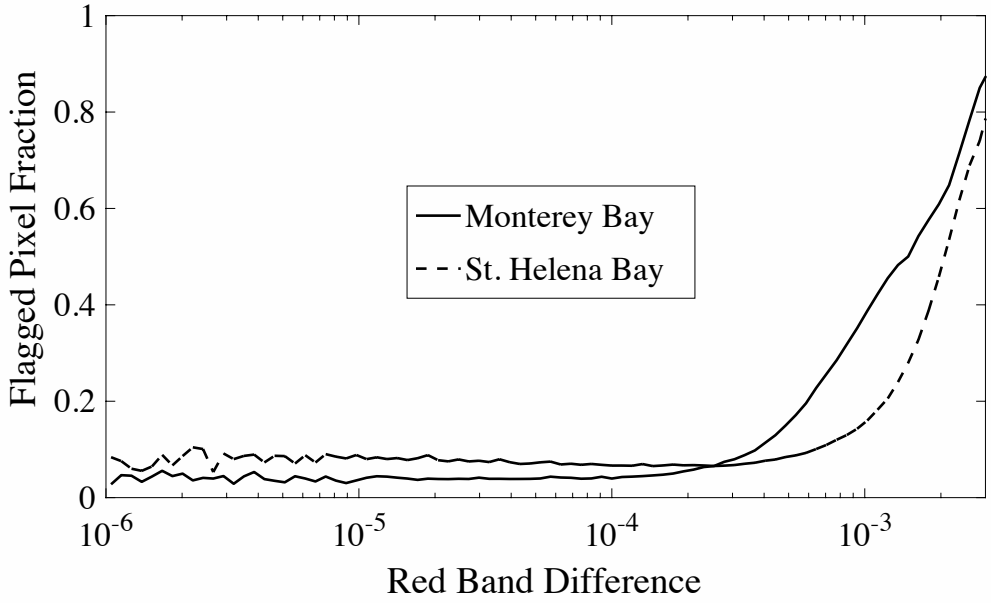

Figure 1.4: *Fraction of pixels assigned atmospheric correction (AC) flags as a function of Red Band Difference, with Monterey Bay and St. Helena Bay shown with solid and dashed black lines, respectively. The range shown encompasses greater than 98% of the data for both regions.*

low magnitude of SB, although the signs remain negative across both datasets.

From each of the four masked and unmasked L3 datasets, composites were also partitioned by the number of underlying L2 pixels that were assigned AC flags. Fig. 1.7 shows a steady increase in the magnitude of SB corresponding with more L2 pixels assigned AC flags, with the masked datasets indicating, on average, about four tenths of one standard deviation less fluorescence than the unmasked datasets for the most frequently masked composites. The similarity in the slopes shown in Fig. 1.7 implies that despite the atmospheric and ecological differences between the regions considered, the relationship between fluorescence and the likelihood of atmospheric correction failure was consistent. About 4% of the L3 transect composites were generated from at least 25% flagged L2 pixels. For the MB and SHB sites, which didn't extend offshore, 8% and 10% of the L3 composites were generated from at least 25% flagged L2 pixels.

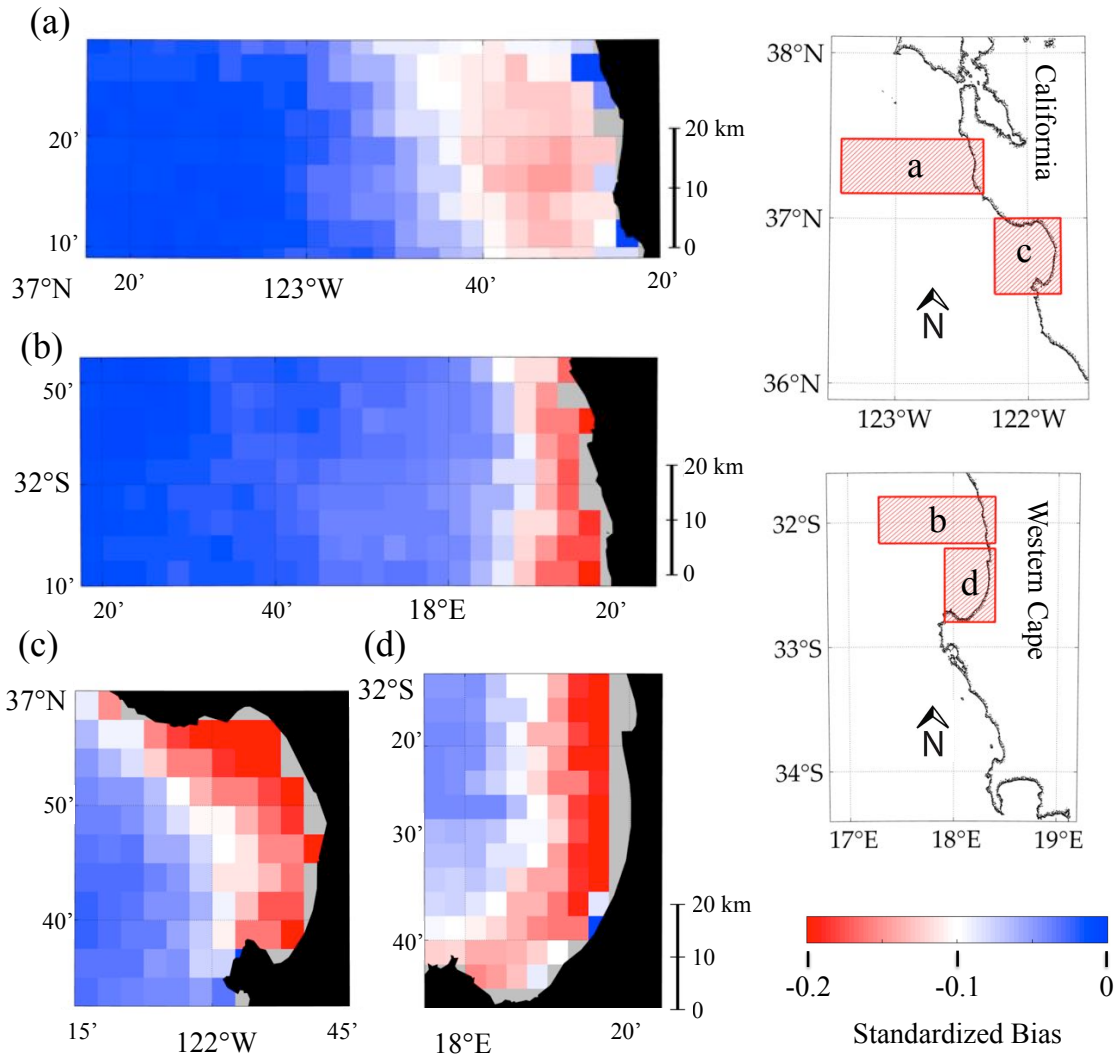

Figure 1.5: *Standardized bias in Red Band Difference from MODISA composites (L3) due to the removal of underlying (L2) pixels, shown for standard map grids within:* ***(a)*** *a transect within the central California Current System;* ***(b)*** *a transect within the southern Benguela Current System;* ***(c)*** *within Monterey Bay; and* ***(d)*** *within St. Helena Bay.*

1.4 Discussion

1.4.1 Performance of satellite products at matchup sites

Based on the matchup results, RBD is a reasonable proxy to describe relative changes in Chl *a* within the study regions, although this comparison is not intended as a validation

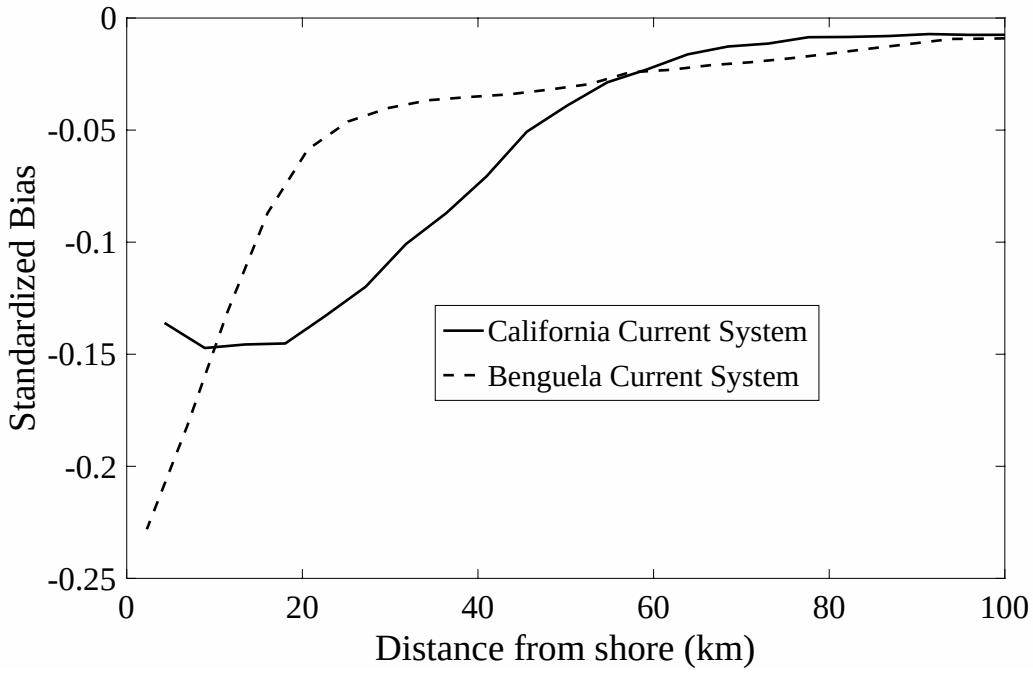

Figure 1.6: *Standardized bias in Red Band Difference composites (L3) across longitudinal transects due to the removal of underlying (L2) pixels assigned AC flags. The California Current System transect is indicated by a solid line and the Benguela Current System transect is indicated by a dashed line.*

activity to assess OC3M, C2RCC or other alternate processing methods. For the OC3M products, atmospheric correction errors were common in the matchup regions, but rigorous quality screening of the matchups would have been counter to the goals targeted by this study, namely, to characterize the observations that fail such screenings. C2RCC performed marginally worse than RBD in the comparisons, but the development of C2RCC was not primarily aimed at use with NASA products. In addition, a recent inter-comparison of atmospheric correction methods for coastal waters found strong improvements for C2RCC after the updates from an expanded training set [22].

FLH algorithms such as RBD are useful for scene comparisons and for simplistic Chl *a* proxies, but fluorescence approaches are not a satisfactory solution to remote sensing challenges in coastal waters. Although Chl *a* fluorescence and concentration generally covary, their relationship is inconsistent. Factors that may alter the relationship between Chl *a* fluorescence and concentration include phytoplankton species composition, pigment pack-

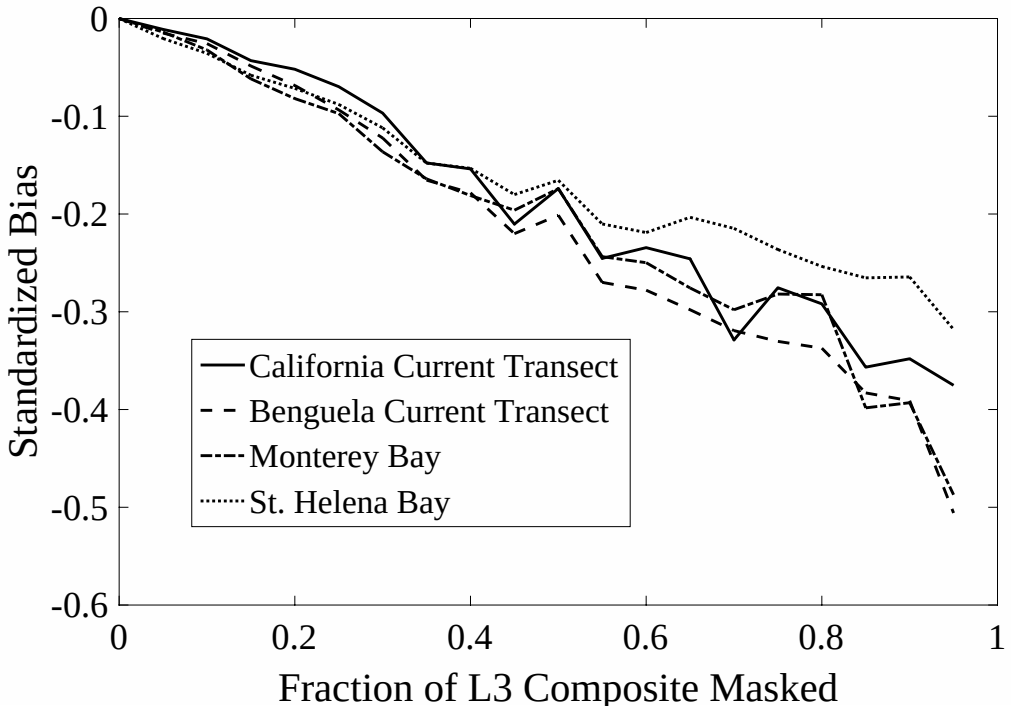

Figure 1.7: *Standardized bias in Red Band Difference for L3 composites, shown as a function of L2 AC flag assignments for data within the California Current System (solid line) and Benguela Current System (dashed line) transects and within Monterey (dashed-dotted) and St Helena (dotted) Bays.*

aging effects, physiology, limitation by nutrients or light, and solar-induced fluorescence quenching [37]. The ability to measure fluorescence is also strongly affected by attenuation from water and its constituents, particularly by non-algal particles [38] and from sensor-specific response functions, for example, if the fluorescence peak shifts between response bands [38, 39]. However, despite the inherent difficulties in quantifying phytoplankton concentrations with Chl *a* fluorescence, such proxies are reasonable for the analysis shown here because of their robustness to atmospheric correction errors and because the RBD biomass comparison is not used across large spatial domains. Comparisons here are assumed to be relative within the region, and this work did not focus on modeling RBD onto Chl *a*.

Products that use inputs from blue wavelengths (such as OC3M) were considered the most sensitive to poor atmospheric correction and therefore were not reasonable options for comparing the masked and unmasked observations. Similarly, products such as OC3M

often cannot be derived for AC flagged pixels, for example, when overestimation of aerosol thickness causes derivation of negative radiances at blue wavelengths. Finally, while RBD was preferred here over the OBPG default nFLH product because of robustness to sediment effects, nFLH is expected to be a reasonable alternative for users who intend to perform a similar analysis of their study region but who require direct downloads of default L2 products.

1.4.2 Variability of atmospheric constituents and ocean color AC flags

Mitigating the atmospheric correction errors that arise from elevated $L_W(\lambda)$ at NIR wavelengths has been a major challenge to remote sensing of sedimented or phytoplankton-rich waters. The associations shown here between AC flags and phytoplankton concentrations within MB and SHB are intended to demonstrate that satellite retrieval quality is linked to biology. In our study areas, the dataset fractions that satisfy default quality control criteria (i.e., are not assigned AC flags) describe ecosystems which are generally reduced in biomass and have lower variability than described by the parent dataset. The results of this study, however, do not suggest that users relax AC flag criteria, because the flag assignments are in most cases reasonable indicators of degraded data quality, particularly in the portion of the spectrum relevant to blue-green band-ratio Chl a algorithms (i.e., OC3M).

The removal of high biomass observations from the satellite record additionally screens out important biological processes, such as the formation of phytoplankton-dense fronts, and removes regions that may be disproportionately important to an ecosystem's dynamics and species succession. For example, a northern MB retentive zone, which maintains dinoflagellate stocks that play an important role in species succession by seeding the surrounding waters [40], was most frequently masked from the MB satellite record.

1.4.3 Potential for user evaluation of Level 2 and 3 datasets

Defining the transition zone between regions where default NIR-based atmospheric correction methods can and cannot be used is challenging, as evidenced by the different composite transect results for the two EBC ecosystems. The effects of the atmospheric correction errors may extend farther offshore than anticipated given the physics of the region (e.g., advection offshore by mesoscale eddies or jets). For regions where the removal

of ephemeral high-biomass pixels may be more infrequent, research targeting ecosystem processes may still suffer from the loss of rare but important events.

How can L2 and L3 end-users test whether satellite datasets contain a strong bias from the removal of high phytoplankton observations? In regions where the biology is well-studied, L2 users can compare flag assignment climatologies with expected phytoplankton dynamics to determine whether flag assignments covary with phytoplankton concentrations. For some flags, seasonality due to the Earth-sun geometry or regional cloud dynamics may resemble biological parameters without tracking phytoplankton biomass within an individual image. Interpretation of the flag assignment frequency should be considered cautiously, because infrequent flag assignments can have outsized effects in regions with high environmental variability. In the absence of *a priori* knowledge of a region, fluorescence products are anticipated to provide useful comparisons of the masked and unmasked data fractions, as shown herein. L3 end-users could make use of compositing statistics in order to assess the representativeness of spatial or temporal, quality-controlled averages. The similarity between all regions shown in Fig. 1.7 suggests that L2 flag assignments are useful parameters for interpreting L3 composites. As such, metadata that includes flag assignment statistics may be a beneficial addition for disseminated L3 products, particularly for users requiring inputs for coastal ocean biogeochemical models.

1.5 Conclusion

Decreased performance of satellite ocean color products in coastal waters has encouraged the development of alternative atmospheric correction methods. However, satellite imagery processed using standard oceanic (i.e., black NIR) approaches are still commonly used in coastal ocean research. This work compares non-aerosol-corrected satellite biomass proxies to test whether the high-quality portion of satellite retrievals are representative of the complete set of retrievals in coastal regions. Key findings shown are that, for productive ecosystems like MB and SHB, the masked satellite record underreports phytoplankton biomass due to the difficulty of removing the atmospheric signal over phytoplankton-rich waters. The changes to recorded biomass for L2 datasets due to atmospheric correction failures persist into L3 composites, and the distances that the changes extend offshore are variable even among broadly similar systems (i.e., EBCs). Finally, users may assess the sensitivity of a satellite product to atmospheric correction failures for a specific study site

by comparing FLH products between the masked and unmasked fractions of their dataset or by retaining flag assignment statistics when generating L3 products. In cases where the fluorescence of the masked fraction is dissimilar to the unmasked fraction, users may prefer alternative atmospheric correction methods regardless of the performance of matchup tests, which are generally obtained from the masked fraction only. Research directed towards coastal ocean ecosystems should evaluate whether quality-controlled satellite estimates of phytoplankton concentrations are representative compared to the statistics of the parent dataset. When possible, TOA proxies are useful tools for such comparisons.

Acknowledgements

We are grateful to Christopher Edwards (University of California, Santa Cruz), Stan Hooker (NASA Goddard Space Flight Center), John P. Ryan (Monterey Bay Aquarium Research Institute; MBARI) and David Draper (University of California, Santa Cruz) for advising on this work. The Pangaea dataset used is available through (`doi.pangaea.de`), the Santa Cruz Wharf Chl *a* data through the Southern California Coastal Ocean Observing System portal (`sccoos.org/data/habs/`), M1 mooring fluorometry data through the MBARI repository (`dods.mbari.org`), the atmospheric climatologies through the AERONET repository (`aeronet.gsfc.nasa.gov`), the MODIS Aqua data and SeaDAS processing software through the NASA OceanColor Web (`oceancolor.gsfc.nasa.gov`), the SNAP software through the ESA Science Toolbox Exploitation Platform (`step.esa.int`), and the C2RCC plugin from Brockmann consulting (`brockmann-consult.de`). The version of the manuscript which was published in *Remote Sensing* was improved by feedback from three anonymous reviewers.

Funding: Support for H.F.H. was partially provided by NASA C-HARRIER Campaign grant number NNX17AK89G, and the NOAA Central and Northern California Ocean Observing System (CeNCOOS) grant number NA16NOS0120021.

References

[1] W. W. Gregg and M. E. Conkright, "Decadal changes in global ocean chlorophyll," *Geophysical Research Letters*, vol. 29, no. 15, pp. 20–1, 2002.

[2] C. R. McClain, "A decade of satellite ocean color observations," *Annual Review of Marine Science*, vol. 1, pp. 19–42, 2009.

[3] H. M. Dierssen, "Perspectives on empirical approaches for ocean color remote sensing of chlorophyll in a changing climate," *Proceedings of the National Academy of Sciences*, vol. 107, no. 40, pp. 17073–17078, 2010.

[4] D. A. Siegel, M. Wang, S. Maritorena, and W. Robinson, "Atmospheric correction of satellite ocean color imagery: the black pixel assumption," *Applied Optics*, vol. 39, no. 21, pp. 3582–3591, 2000.

[5] S. Sathyendranath, R. P. Bukata, R. Arnone, M. Dowell, C. O. Davis, M. Babin, J. F. Berthon, O. V. Kopelevich, and J. W. Campbell, "Colour of case 2 waters," in *Remote sensing of ocean colour in coastal, and other optically-complex, waters* (S. Sathyendranath, ed.), pp. 23 – 46, Dartmouth, Nova Scotia, Canada: International Ocean-Colour Coordinating Group, 2000.

[6] M. Kahru and B. G. Mitchell, "Ocean color reveals increased blooms in various parts of the world," *Eos, Transactions American Geophysical Union*, vol. 89, no. 18, pp. 170–170, 2008.

[7] D. A. Jessup, M. A. Miller, J. P. Ryan, H. M. Nevins, H. A. Kerkering, A. Mekebri, D. B. Crane, T. A. Johnson, and R. M. Kudela, "Mass stranding of marine birds caused by a surfactant-producing red tide," *PLoS One*, vol. 4, no. 2, p. e4550, 2009.

[8] A. J. Lewitus, R. A. Horner, D. A. Caron, E. Garcia-Mendoza, B. M. Hickey, M. Hunter, D. D. Huppert, R. M. Kudela, G. W. Langlois, J. L. Largier, *et al.*, "Harmful algal blooms along the north american west coast region: History, trends, causes, and impacts," *Harmful Algae*, vol. 19, pp. 133–159, 2012.

[9] R. M. McCabe, B. M. Hickey, R. M. Kudela, K. A. Lefebvre, N. G. Adams, B. D. Bill, F. M. D. Gulland, R. E. Thomson, W. P. Cochlan, and V. L. Trainer, "An unprecedented coastwide toxic algal bloom linked to anomalous ocean conditions," *Geophysical Research Letters*, vol. 43, no. 19, pp. 10–366, 2016.

[10] G. C. Pitcher, F. G. Figueiras, R. M. Kudela, T. Moita, B. Reguera, and M. Ruiz-Villareal, "Key questions and recent research advances on harmful algal blooms in

eastern boundary upwelling systems," in *Global Ecology and Oceanography of Harmful Algal Blooms*, pp. 205–227, Springer, 2018.

[11] A. Bracher, H. A. Bouman, R. J. Brewin, A. Bricaud, V. Brotas, A. M. Ciotti, L. Clementson, E. Devred, A. Di Cicco, S. Dutkiewicz, *et al.*, "Obtaining phytoplankton diversity from ocean color: a scientific roadmap for future development," *Frontiers in Marine Science*, vol. 4, p. 55, 2017.

[12] R. J. Frouin, B. A. Franz, A. Ibrahim, K. D. Knobelspiesse, Z. Ahmad, B. Cairns, J. Chowdhary, H. M. Dierssen, J. Tan, O. Dubovik, *et al.*, "Atmospheric correction of satellite ocean-color imagery during the pace era," *UMBC Faculty Collection*, 2019.

[13] R. A. Barnes, D. K. Clark, W. E. Esaias, G. S. Fargion, G. C. Feldman, and C. R. McClain, "Development of a consistent multi-sensor global ocean colour time series," *International Journal of Remote Sensing*, vol. 24, no. 20, pp. 4047–4064, 2003.

[14] W. W. Gregg and N. W. Casey, "Improving the consistency of ocean color data: A step toward climate data records," *Geophysical Research Letters*, vol. 37, no. 4, 2010.

[15] M. Kahru, R. M. Kudela, M. Manzano-Sarabia, and B. G. Mitchell, "Trends in the surface chlorophyll of the california current: Merging data from multiple ocean color satellites," *Deep Sea Research Part II: Topical Studies in Oceanography*, vol. 77, pp. 89–98, 2012.

[16] H. R. Gordon and M. Wang, "Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with seawifs: a preliminary algorithm," *Applied Optics*, vol. 33, no. 3, pp. 443–452, 1994.

[17] C. D. Mobley, P. J. Werdell, B. Franz, Z. Ahmad, and S. Bailey, "Atmospheric correction for satellite ocean color radiometry," 2016.

[18] H. R. Gordon, D. Antoine, H. Fukushima, M. Wang, R. Frouin, A. Morel, P. Y. Deschampe, and J. M. Nicolas, "Atmospheric correction algorithm description," in *Atmospheric correction for remotely-sensed ocean-colour products* (M. Wang, ed.), vol. 10, pp. 5 – 14, Dartmouth, Nova Scotia, Canada: International Ocean-Colour Coordinating Group, 2010.

[19] M. Wang, "Remote sensing of the ocean contributions from ultraviolet to near-infrared using the shortwave infrared bands: simulations," *Applied Optics*, vol. 46, pp. 1535–1547, Mar 2007.

[20] K. G. Ruddick, F. Ovidio, and M. Rijkeboer, "Atmospheric correction of seawifs imagery for turbid coastal and inland waters," *Applied Optics*, vol. 39, no. 6, pp. 897–912, 2000.

[21] C. Brockmann, R. Doerffer, M. Peters, S. Kerstin, S. Embacher, and A. Ruescas, "Evolution of the c2rcc neural network for sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters," in *Living Planet Symposium*, vol. 740, p. 54, 2016.

[22] M. A. Mograne, C. Jamet, H. Loisel, V. Vantrepotte, X. Mériaux, and A. Cauvin, "Evaluation of five atmospheric correction algorithms over french optically-complex waters for the sentinel-3a olci ocean color sensor," *Remote Sensing*, vol. 11, no. 6, p. 668, 2019.

[23] J. W. Campbell, J. M. Blaisdell, and M. Darzi, "Level-3 seawifs data products: spatial and temporal binning algorithms," *Oceanographic Literature Review*, vol. 9, no. 43, p. 952, 1996.

[24] J. P. Scott and P. J. Werdell, "Comparing level-2 and level-3 satellite ocean color retrieval validation methodologies," *Optics Express*, vol. 27, no. 21, pp. 30140–30157, 2019.

[25] J. P. Ryan, A. M. Fischer, R. M. Kudela, J. F. Gower, S. A. King, R. Marin III, and F. P. Chavez, "Influences of upwelling and downwelling winds on red tide bloom dynamics in monterey bay, california," *Continental Shelf Research*, vol. 29, no. 5-6, pp. 785–795, 2009.

[26] J. T. Pennington and F. P. Chavez, "Seasonal fluctuations of temperature, salinity, nitrate, chlorophyll and primary production at station h3/m1 over 1989–1996 in monterey bay, california," *Deep Sea Research Part II: Topical Studies in Oceanography*, vol. 47, no. 5-6, pp. 947–973, 2000.

[27] G. C. Pitcher, P. C. Brown, and B. A. Mitchell-Innes, "Spatio-temporal variability of phytoplankton in the southern benguela upwelling system," *South African Journal of Marine Science*, vol. 12, no. 1, pp. 439–456, 1992.

[28] R. Barlow, H. Sessions, M. Balarin, S. J. Weeks, C. Whittle, and L. Hutchings, "Seasonal variation in phytoplankton in the southern benguela: pigment indices and ocean colour," *African Journal of Marine Science*, vol. 27, no. 1, pp. 275–287, 2005.

[29] A. Fawcett, G. C. Pitcher, S. Bernard, A. D. Cembella, and R. M. Kudela, "Contrasting wind patterns and toxigenic phytoplankton in the southern benguela upwelling system," *Marine Ecology Progress Series*, vol. 348, pp. 19–31, 2007.

[30] E. H. Schumann and J. A. Martin, "Climatological aspects of the coastal wind field at cape town, port elizabeth and durban," *South African Geographical Journal*, vol. 73, no. 2, pp. 48–51, 1991.

[31] S. B. Hooker, E. R. Firestone, F. S. Patt, R. A. Barnes, R. E. Eplee Jr, B. A. Franz, W. D. Robinson, G. C. Feldman, S. W. Bailey, *et al.*, "Algorithm updates for the fourth seawifs data reprocessing," Tech. Rep. TM-2003-206892, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, 2003.

[32] R. A. Neville and J. F. R. Gower, "Passive remote sensing of phytoplankton via chlorophyll α fluorescence," *Journal of Geophysical Research*, vol. 82, no. 24, pp. 3487–3493, 1977.

[33] H. R. Gordon, "Diffuse reflectance of the ocean: the theory of its augmentation by chlorophyll a fluorescence at 685 nm," *Applied Optics*, vol. 18, no. 8, pp. 1161–1166, 1979.

[34] R. M. Letelier and M. R. Abbott, "An analysis of chlorophyll fluorescence algorithms for the moderate resolution imaging spectrometer (modis)," *Remote Sensing of Environment*, vol. 58, no. 2, pp. 215–223, 1996.

[35] J. F. Gower, "On the use of satellite-measured chlorophyll fluorescence for monitoring coastal waters," *International Journal of Remote Sensing*, vol. 37, no. 9, pp. 2077–2086, 2016.

[36] R. Amin, J. Zhou, A. Gilerson, B. Gross, F. Moshary, and S. Ahmed, "Novel optical techniques for detecting and classifying toxic dinoflagellate karenia brevis blooms using satellite imagery," *Optics Express*, vol. 17, no. 11, pp. 9126–9144, 2009.

[37] C. S. Roesler and M. J. Perry, "In situ phytoplankton absorption, fluorescence emission, and particulate backscattering spectra determined from reflectance," *Journal of Geophysical Research: Oceans*, vol. 100, no. C7, pp. 13279–13294, 1995.

[38] J. F. Gower and G. A. Borstad, "On the potential of modis and meris for imaging chlorophyll fluorescence from space," *International Journal of Remote Sensing*, vol. 25, no. 7-8, pp. 1459–1464, 2004.

[39] J. P. Ryan, C. O. Davis, N. B. Tufillaro, R. M. Kudela, and B.-C. Gao, "Application of the hyperspectral imager for the coastal ocean to phytoplankton ecology studies in monterey bay, ca, usa," *Remote Sensing*, vol. 6, no. 2, pp. 1007–1025, 2014.

[40] J. P. Ryan, J. F. Gower, S. A. King, W. P. Bissett, A. M. Fischer, R. M. Kudela, Z. Kolber, F. Mazzillo, E. V. Rienecker, and F. P. Chavez, "A coastal ocean extreme bloom incubator," *Geophysical Research Letters*, vol. 35, no. 12, 2008.

Chapter 2

Remote sensing of colored dissolved organic matter for oceanic, coastal, and inland waters

Overview: Next-generation satellite and airborne missions aim to measure key biogeochemical constituents, including $a_{\mathrm{CDOM}}(440)$, within smaller and more optically complex inland water bodies. This will require new algorithms that quantify $a_{\mathrm{CDOM}}(440)$ without relying on the case-1 assumption of covariance between various constituents and Chl a. The following work evaluates bio-optical parameterizations, *in situ* data, and airborne matchups to test the importance of spectral range for two-channel $a_{\mathrm{CDOM}}(440)$ remote sensing algorithms. The findings indicate that the dynamic range of band-ratio algorithms and their correlation to $a_{\mathrm{CDOM}}(440)$ increase with greater spectral separation of the waveband ratio toward ultraviolet and near-infrared wavelengths. The approach does not require regional tunings to various inland water bodies, and it confers additional advantages to sensing of inland waters, including mitigation of bottom reflectance (as ultraviolet and near-infrared wavelengths are strongly attenuated by water) and decreased spectral sensing requirements.

The following contains material that was submitted to *Remote Sensing of Environment* on 10 February 2020:

Spectral range within global $a_{\mathrm{CDOM}}(440)$ algorithms for oceanic, coastal, and inland waters with application to airborne measurements

Henry F. Houskeeper, Stanford B. Hooker, & Raphael M. Kudela

Abstract: The optically active component of dissolved organic material in aquatic ecosystems, or colored dissolved organic matter (CDOM), is represented by the coefficient of absorption due to the dissolved aquatic constituents at 440 nm, $a_{\mathrm{CDOM}}(440)$. Remote sensing of $a_{\mathrm{CDOM}}(440)$ enables characterization of ecosystem processes and aids in retrieval of chlorophyll a, a proxy for phytoplankton biomass, which overlaps in spectral dependencies with $a_{\mathrm{CDOM}}(440)$. Spectrally adjacent band-ratios, e.g., blue to green, have previously been applied for remote sensing of $a_{\mathrm{CDOM}}(440)$ in coastal and oceanic waters with similar results compared to more complex semi-analytical algorithms. Estimation of $a_{\mathrm{CDOM}}(440)$ from ratios of the most spectrally separate ocean color bands (end-members), e.g., ultraviolet to near-infrared, termed end-member analysis (EMA), has been shown to increase the accuracy of $a_{\mathrm{CDOM}}(440)$ retrievals from in-water measurements and to enable a unified algorithmic perspective without requiring regional adjustment of internal bio-optical parameters. EMA of above-water measurements is evaluated herein, with a focus on coastal and inland waters in which increasing optical complexity and higher incidence of bottom reflectance challenge the oceanic algorithms developed for deep and optically simple (case-1) waters. Analysis herein of three independent bio-optical datasets indicates strong association between $a_{\mathrm{CDOM}}(440)$ and end-member band-ratios (320 and 780 nm or 412 and 670 nm) with a coefficient of determination, R^2, of 0.82 or higher based on a global waters dataset. For applicable wavelengths, EMA algorithms are shown to agree with case-1 relationships and to produce consistent log-scale uncertainties across nearly three orders of magnitude in $a_{\mathrm{CDOM}}(440)$. EMA is applied to airborne measurements and satisfies a 25% log-scale uncertainty criteria within each of three dissimilar matchup sites ranging in $a_{\mathrm{CDOM}}(440)$ from $0.02-0.57\,\mathrm{m}^{-1}$. Results suggest that EMA is a useful and robust approach for remote sensing of $a_{\mathrm{CDOM}}(440)$ in coastal and inland waters, which generally are more optically diverse and span a greater range in $a_{\mathrm{CDOM}}(440)$ than oceanic waters.

2.1 Introduction

Aquatic dissolved organic matter (DOM) modifies the solar heat flux to surface waters, provides carbon storage on ephemeral to geologic timescales, and fuels the microbial pump that recycles nutrients to maintain global phytoplankton stocks [1–3]. Oceanic sources of

DOM include microbial degradation of phytoplankton [4, 5] and phytoplankton secretions for photo-protection [6], e.g., mycosporine-like amino acids, or MAAs [7, 8]. Weathering of terrestrial material injects DOM into aquatic ecosystems, and rivers transport DOM from inland to estuarine and coastal waters [9]. The optically active component of DOM, termed colored dissolved organic matter (CDOM), has been used as a DOM tracer [10, 11] and enables remote sensing (space or airborne) of ecosystem processes, e.g., for analysis of large runoff events [12] or for measurement of oceanic circulation rates [13].

Decadal analysis of satellite imagery reveals global fluctuations in ocean color related to basin-scale oscillations in oceanic CDOM content [14]. Improving remote measurement of CDOM has been a key objective for ocean color remote sensing, because the accuracy of chlorophyll *a* concentration (Chl *a*; a proxy for phytoplankton biomass) derived by satellites degrades in waters with elevated CDOM loads [15–20]. Traditional satellite Chl *a* algorithms, i.e., blue-green band-ratios [21], target high Chl *a*-specific absorption of blue light, a spectral characteristic that also describes CDOM absorption, in which spectral dependencies are modeled as follows [22]:

$$a_{\mathrm{CDOM}}(\lambda) \;=\; a_{\mathrm{CDOM}}(440)\; e^{-S(\lambda\,-\,440)}\,, \tag{2.1}$$

where λ is wavelength, $a_{\mathrm{CDOM}}(440)$ is the absorption coefficient of CDOM at 440 nm, and S is the spectral slope of CDOM absorption. Natural variability in S corresponds to differences in lability (humic versus fulvic) or source (marine versus riverine) of DOM [15]. S values ranging from $0.015 - 0.03\,\mathrm{nm}^{-1}$ are common among oceanic waters [14], although wide discrepancies exist in the literature due, in part, to differences in wavelength intervals and fitting [23]. An expanded range of $0.0095 - 0.0410\,\mathrm{nm}^{-1}$ for oceanic, coastal, and inland waters was reported in [24].

Remote measurement of $a_{\mathrm{CDM}}(440)$, or the combination of $a_{\mathrm{CDOM}}(440)$ and detrital absorption at 440 nm, has been most successful in the optically simple oligotrophic ocean using inversion, e.g., the Garver-Siegel-Maritorena algorithm (GSM), or semi-analytical, e.g., the Quasi-Analytical Algorithm (QAA), methods [25, 26]. Regional optimizations of existing quasi- or semi-analytical algorithms have enabled ocean color observations of specific coastal ecosystems [27–29], and method updates to the GSM and QAA have been proposed for separating the dissolved and detrital signals [30, 31]. Band-ratio approaches for estimating $a_{\mathrm{CDOM}}(440)$ using spectrally adjacent wavelengths (e.g., 412 – 443 nm or 443 –

555 nm) have produced similar results as algorithms with greater spectral and computational requirements in oceanic environments [32] and have been applied to coastal [33, 34] and lacustrine [35, 36] environments. The performance of band-ratio algorithms is anticipated to improve by using more spectrally separate ocean color bands because the dynamic range of ocean color measurements as a function of $a_{\mathrm{CDOM}}(440)$ increases with greater separation between waveband ratios, e.g., ultraviolet (UV) to near-infrared (NIR), as shown by [37]. This algorithmic approach, hereafter referred to as end-member analysis (EMA), also has the advantage of minimizing the effects of photosynthetic pigments, which have strong spectral dependencies across visible (VIS) wavelengths [38]. For *in situ* measurements of global water masses (e.g., oceanic, estuarine, lacustrine, riverine), EMA of diffuse attenuation coefficients for spectral downward irradiance, $K_d(\lambda)$, enables estimation of $a_{\mathrm{CDOM}}(440)$ within a 2% root mean square error (RMSE), with consistent $\log_{10}$-scale accuracy for $a_{\mathrm{CDOM}}(440)$ values spanning over three decades in range [24]. The efficacy of an EMA algorithmic approach derived from above-water measurements, i.e., the normalized water-leaving radiance, $\left[L_W(\lambda)\right]_{\mathrm{N}}$, is evaluated herein using theoretical, *in situ*, and airborne matchup results.

2.2 Methods

2.2.1 Derivation of synthetic case-1 bio-optical datasets

The radiance distribution and constituents of natural, optically simple waters may be estimated first-order based solely on the Chl *a* content, because the optically active constituents (including algal and non-algal particles, CDOM, etc.) covary; a condition defined in optical oceanography as case-1 [39], and all other conditions described herein as case-2. Under this scenario, biogeochemical relationships between $a_{\mathrm{CDOM}}(440)$ and Chl *a*, e.g., [40], enable derivation of a case-1 model relating $a_{\mathrm{CDOM}}(440)$ to bio-optical constituents and therefore to the theoretical light field of a natural water mass. A synthetic, case-1 $a_{\mathrm{CDOM}}(440)$ dataset is derived using the parameterizations of [41] as follows:

$$a_{\mathrm{CDOM}}(440) \;=\; 0.032\,[\mathrm{Chl}\,a]^{0.63}\,. \tag{2.2}$$

Although the case-1 scenario is not upheld in optically complex waters, algorithmic agreement with case-1 relationships may be considered as a baseline for evaluating whether an algorithm is internally consistent with the predefined bio-optical relationships of case-1 wa-

ters. The parameterizations used in this study are intended to provide a general framework for considering the basis of an EMA algorithm within the parameterized ranges in Chl a and are not intended as a thorough review of bio-optical models. The derivations for two $K_d(\lambda)$ and one $\left[L_W(\lambda)\right]_{\mathrm{N}}$ synthetic case-1 datasets are briefly described below, with more detail on the case-1 parameterizations provided in this chapter's appendix.

The scattering and absorption coefficients of seawater are obtained from [42] for red and NIR wavelengths and from [43] for blue and UV wavelengths, with the backscattering contribution of seawater approximated as one-half of the total scattering coefficient. The absorption by particles (including algal and non-algal) are parameterized with spectral dependencies according to [44], with coefficients selected from the HydroLight Mid-range UV option of the *New Case-1 Model* [45]. The scattering contributions of particles are parameterized according to [41], with the effects of backscattering efficiencies considered by sensitivity testing.

2.2.2 Biogeochemical parameterizations of $K_d(\lambda)$

$K_d(\lambda)$ has been parameterized as a power-law function of Chl a within the VIS wavelength range for waters spanning Chl a of $0.035-5.45\ \mathrm{mg\,m^{-3}}$ [46] and $0.02-30\ \mathrm{mg\,m^{-3}}$ [47], as:

$$K_d(\lambda) \;=\; K_w(\lambda) \;+\; \chi(\lambda)\,[\text{Chl } a]^{\mathrm{e}(\lambda)}\,, \tag{2.3}$$

where $K_w(\lambda)$ is the diffuse attenuation coefficient of pure water, and $\chi(\lambda)$ and $\mathrm{e}(\lambda)$ are fitting coefficients. The χ and e coefficients are extended to NIR wavelengths [48] by assuming no biological contributions above 750 nm (i.e., χ decreases linearly between 700–750 nm) and to UV wavelengths by assuming a chlorophyll-specific $K_d(\lambda)$ that resembles the specific absorption of non-algal particles (in clear waters) or of a pure algal suspension (in eutrophic waters). A synthetic $K_d(\lambda)$ dataset derived from Eq. 2.3 following [46] and [48] is hereafter referred to as MSyn.

$K_d(\lambda)$ may be treated as an inherent optical property (IOP), wherein it is dependent on the constituents of a water mass and not on the ambient light field, if $K_d(\lambda)$ is normalized by the downwelling distribution function, $D_o(\lambda)$, which relates the subsurface downwelling irradiance to the subsurface scalar downwelling irradiance, following [49]. The correction allows for the partitioning of $K_d(\lambda)$ among additive constituents (e.g., water, particles, and CDOM). Least-squares fitting relates $K_d(\lambda)$ to absorption $a(\lambda)$ and back-scattering $b_b(\lambda)$

coefficients [49], with the spectral dependencies dropped for brevity such that:

$$\frac{K_d}{D_o} \;=\; 1.0395 \left[\sum_{i=1}^{m} (a)_i + \sum_{i=1}^{m} (b_b)_i \right], \tag{2.4}$$

where a and b_b are partitioned as i of m components (e.g., water, particles, and CDOM). Parameterizations of a and b_b (see Appendix) enable a case-1 solution relating $K_d(\lambda)$ band-ratios to $a_{\mathrm{CDOM}}(440)$ within the variability arising from the spectral dependencies of D_o (due to scattering of the atmosphere). The ratios $\frac{D_o(670)}{D_o(412)}$ and $\frac{D_o(780)}{D_o(320)}$ were considered using [49] as well as by approximating the spectral fraction of downward irradiance contained within the solar beam at relevant wavelengths following [50], and the variability conferred to an EMA algorithmic approach was not significant given other sources of uncertainty in the case-1 parameterizations, e.g., natural variability in S. A synthetic $K_d(\lambda)$ dataset derived from Eq. 2.4 following [49] is hereafter referred to as GSyn.

2.2.3 Biogeochemical parameterizations of $\left[L_W(\lambda)\right]_{\mathrm{N}}$

The case-1 model was used to derive a synthetic $\left[L_W(\lambda)\right]_{\mathrm{N}}$ dataset based on the approach of [51], with updates to the methodology, as well as applicable terms and coefficients, provided in [46], [52], [53], and [54]. The case-1 model $\left[L_W(\lambda)\right]_{\mathrm{N}}$ dataset was compared with a HydroLight model [55] using the relative percent difference (RPD), with HydroLight as the reference in the difference calculations. The mean RPD value as a function of the spectral domains and total Chl a spanning $0.02-20$ mg m^{-3} was 1.1%, and the overall absolute percent difference (APD) was 13.9%. The comparison of spectral shape was significantly better with values of the Pearson correlation coefficient, ρ, for eutrophic, mesotrophic, and oligotrophic waters of 0.991, 0.966, and 0.987, respectively. The synthetic $\left[L_W(\lambda)\right]_{\mathrm{N}}$ dataset was processed using the Processing of Radiometric Observations of Seawater using Information Technologies (PROSIT) software [56], and is hereafter referred to as PSyn.

2.2.4 Description of above- and in-water data products

Ratios of $\left[L_W(\lambda_1)\right]_{\mathrm{N}} / \left[L_W(\lambda_2)\right]_{\mathrm{N}}$, hereafter $\Lambda_{\lambda_2}^{\lambda_1}$, were selected for comparison to future capabilities of next-generation sensors (Λ_{780}^{320}), to current capabilities of existing ocean color satellites (Λ_{670}^{412}), to proposed band-ratio approaches (Λ_{555}^{443}), and to capabili-

ties of legacy platforms or those not primarily dedicated to ocean color missions (Λ_{625}^{465}). The $\Lambda_{\lambda_2}^{\lambda_1}$ pairs were obtained from *in situ* datasets containing coincident measurements of $a_{\mathrm{CDOM}}(440)$ and $\left[L_W(\lambda)\right]_{\mathrm{N}}$ spanning up to 320 – 780 nm, described below.

An oceanic dataset was created by reprocessing data obtained with legacy instruments, hereafter referred to as the OCEAN dataset. The reprocessing was performed for 16 wavebands spanning 320 – 780 nm with 10 nm central bandwidths. The OCEAN dataset was obtained from both rocket-shaped profiler designs as well as the first Compact-Optical Profiling System (C-OPS), which used a novel kite-shaped backplane [57]. Consequently, vertical sampling resolution (VSR) was never less than 1 cm, and the first upwelling radiance observations, $L_u(z, \lambda)$, were typically obtained deeper than 0.3 m (the nominal first measurement depth of the C-OPS downward-facing radiometer). The geographical areas sampled ranged from the Southern Ocean to tropical waters, plus northern mid-latitudes. The majority of the measurements were made in deep oceanic waters (63%) with the remainder in shallower coastal waters (37%); no inland waters were sampled, and the depth of the profiles were to the 1% light level or more.

A global dataset of oceanic, coastal, and inland waters was obtained using a C-OPS with an average VSR of 6.0 mm (0.9 mm within very shallow or turbid waters) and with the first $L_u(z, \lambda)$ observations routinely obtained at approximately 0.3 m [24]. The C-OPS spectral configuration contained 19 wavebands with 10 nm bandwidths, and was obtained with the Compact-Propulsion Option for Profiling Systems (C-PrOPS) thruster accessory [58], which enabled sampling of shallow, non-navigable waters, and which mitigated sampling difficulties related to adjacency, shading, and physical perturbations. The global dataset was designed with the intention of enabling algorithm development and validation in optically complex water bodies, and sampling was uniformly spread across oceanic (31.2%), coastal (36.6%), and inland (32.2%) ecosystems. The uniform distribution achieved is important for unbiased curve fitting of global waters.

Factual observations regarding essential water mass attributes were recorded prior to collection of the global dataset to establish a subjective classification scheme [24]. The rationale for prior subjective classification was not to exclude optically complex water bodies, but rather to identify waters with non-conservative perturbations (i.e., modifications not consistent with a linear mixing of parent water masses), which are not expected to satisfy a global algorithmic approach. Examples of relevant categorical information include the persistence of a severe drought, which cuts off inflow to lakes and rivers and can increase

resuspension due to shallowing of the water level, or flooding of lakes and rivers beyond nominal fill levels, which can introduce scoured terrigenous materials. The subjective classification scheme does not signify that these conditions are ignored in the evaluation of algorithm efficacy. Rather, the classifications provide extra information to enable testing of whether particular environments (e.g., a drought-stricken lake) require an alternate or regionalized algorithmic approach. Measurements from the global dataset with no subjective classification assignments (conservative water bodies relevant to a global algorithmic perspective) are hereafter referred to as the GLOBC dataset, while measurements with non-conservative classifications are referred to as the GLOBN dataset, which is considered separately in Section 2.3.4.

A dataset containing 308 above-water EMA pairs for legacy (VIS) wavelengths (Λ_{670}^{412}, Λ_{555}^{443}, and Λ_{625}^{465}) was obtained from the NASA bio-Optical Marine Algorithm Dataset (NOMAD), using version 2a (`https://seabass.gsfc.nasa.gov/wiki/NOMAD`), which is part of a long-standing repository of bio-optical measurements produced for ocean optics algorithm modeling and validation [59–61]. This dataset is hereafter referred to as the NOMAD dataset. In-water products from the NOMAD repository were previously applied to evaluate EMA based on $K_d(\lambda)$ in [24], which found that the legacy products confirmed the in-water algorithm after quality control of the dataset, for example by removing coefficients clearer than pure water. Here, a synthetic $K_d(412)/K_d(670)$ dataset is produced from NOMAD above-water products using [62], which estimates in-water products from above-water measurements using a neural network approach. This dataset, referred to in Section 2.4.1 as NNet, is included to investigate whether remote sensing methods to predict $a_{\mathrm{CDOM}}(440)$ would benefit from working through $K_d(\lambda)$ intermediates rather than directly estimating $a_{\mathrm{CDOM}}(440)$ from above-water measurements.

2.2.5 Quality control and partitioning of datasets

For the OCEAN dataset, in which prior subjective classification information was not available, an objective approach for partitioning $K_d(\lambda)$ spectra based on fuzzy c-means analysis [24] was applied following [63]. Median spectra that satisfied criteria corresponding with conservative water bodies in [24], and which did not contain $K_d(\lambda)$ below that of natural water, were retained.

Quality control was also performed for the OCEAN, GLOBC, and GLOBN datasets

due to the inherent difficulty of measuring $L_u(z,\lambda)$ from an in-water profiler within extremely turbid waters, i.e., $K_d(\lambda) > 10\,\mathrm{m}^{-1}$. Although the VSR was optimized in the GLOBC and GLOBN datasets by technological improvements designed to increase surface loitering and stabilize aperture planar geometries [58], the shallowest possible $L_u(z,\lambda)$ measurement is limited by the length of the downward-facing radiometer. Exclusion of $K_d(\lambda)$ values above $10\,\mathrm{m}^{-1}$ removed roughly 1% of the OCEAN dataset, 8% of the GLOBC dataset, and 30% of the GLOBN dataset. More measurements were removed from the GLOBN dataset because the non-conservative water bodies contained a greater range in turbidity. For validation comparisons herein, all spectra were required to satisfy the $K_d(\lambda)$ quality control threshold for all wavelengths.

2.2.6 Description of airborne datasets

The Compact-Airborne Environmental Radiometers for Oceanography (C-AERO) instrument suite uses above-water (airborne) radiometers, which typically have 19 wavelengths (nominally with 10 nm bandwidths), spanning the UV–NIR with 16 wavelengths plus the short-wave infrared (SWIR) with three wavelengths [58]. The instruments are used to derive normalized data products by simultaneously measuring the total radiance $L_T(\lambda)$ from the water surface at a specified angle with respect to nadir (typically 40°), the indirect (sky) radiance $L_i(\lambda)$ measured in the same plane and at a complementary zenith angle as $L_T(\lambda)$, and the global solar irradiance $E_s(\lambda)$. The radiance radiometers have a narrow field of view (2.5° full view angle) and are fitted with a shroud to reduce long-wavelength scattering at the aperture. The $L_W(\lambda)$ term is derived from the $L_T(\lambda)$ observations by filtering out sun glint in the $L_T(\lambda)$ data and removing the sky reflection based on a reflectance model depending on the viewing geometry (i.e., pointing angle of the radiometers) and wind speed [64, 65].

The C-AERO instrument suite was flown aboard a Twin Otter aircraft operated by the Center for Interdisciplinary Remotely Piloted Aircraft Studies (CIRPAS) at the lowest safe altitude (LSA), which was as low as 100 ft (30.5 m) in order to exclude atmospheric effects, and measurements were screened for improper pointing relative to the solar azimuth. C-AERO observations were obtained in the San Francisco Bay Delta (SFBD), Lake Tahoe (LT), and Monterey Bay (MB). All flights coincided with water sampling for $a_{\mathrm{CDOM}}(440)$, in which samples were filtered through a Whatman GD/X 0.2 μm syringe filter, and then measured with a Cary Varian 50 spectrophotometer with a 10 cm path-length. The locations

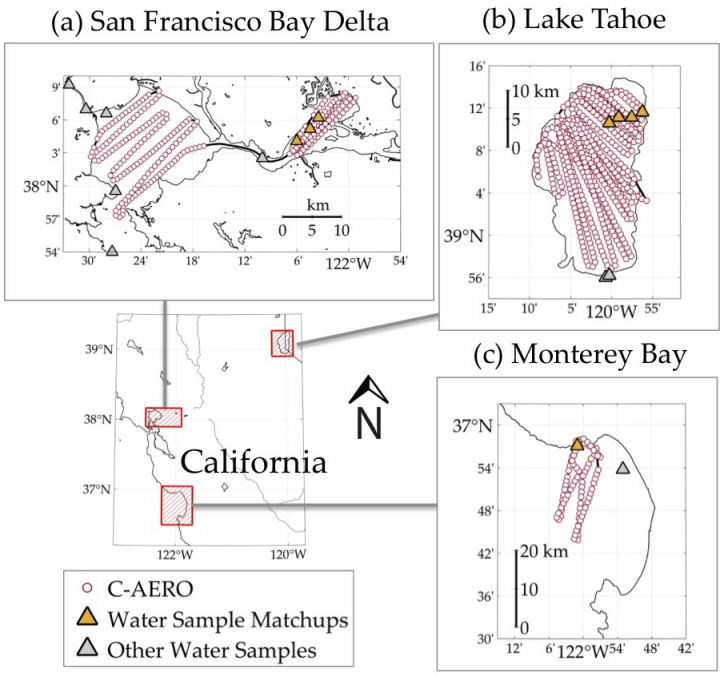

Figure 2.1: *Locations of C-AERO flight lines and data products after glint correction:* ***(a)*** *SFBD in California;* ***(b)*** *LT in California and Nevada; and* ***(c)*** *MB in California. Twin Otter flight lines are indicated in solid black, with C-AERO data products shown as red circles, and water sampling matchup sites indicated as orange triangles. Grey triangles indicate water sampling sites that were not considered for matchups due to spatial or temporal separation.*

of the C-AERO data products and water sampling sites are shown in Fig. 2.1.

Signal-to-noise ratios for the C-AERO measurements exceed those of other relevant remote platforms within SFBD and LT [66], and coincident in-water C-OPS data products demonstrated radiometric $\left[L_W(\lambda)\right]_{\mathrm{N}}$ agreement of $\rho = 0.989$ and 0.999 at LT and SFBD, respectively. Geospatial imagery (50 m) was produced from the C-AERO measurements (approximately 3.4 m surface spot size at LSA for native resolution) using natural neighbor interpolation [67] without extrapolation beyond the boundary of the airborne measurements, and with the closest nearshore retrievals masked.

2.2.7 Derivation of algorithm fits for $\left[L_W(\lambda)\right]_{\mathrm{N}}$ EMA

The $\Lambda_{\lambda_2}^{\lambda_1}$ pairs indicate log-linear association with coincident $a_{\mathrm{CDOM}}(440)$ measurements, and a power-law relationship was applied to model $a_{\mathrm{CDOM}}(440)$ from $\left[L_W(\lambda)\right]_{\mathrm{N}}$ as

follows:

$$a_{\mathrm{CDOM}}(440) \;=\; A\left[\Lambda_{\lambda_2}^{\lambda_1}\right]^{B}\,, \tag{2.5}$$

where the coefficients A and B were derived by minimizing an absolute deviation cost function, and their associated uncertainties were estimated by non-parametric bootstrapping. Because the distribution of $a_{\mathrm{CDOM}}(440)$ measurements within the OCEAN dataset emphasizes a narrow range in $a_{\mathrm{CDOM}}(440)$, an iterative random thinning of the oversampled range was performed to obtain a median curve fit that was representative of the full dataset range in $a_{\mathrm{CDOM}}(440)$. The combined absorption by non-algal particles and CDOM, $a_{\mathrm{CDM}}(440)$, was derived using the GSM [25] and the updated version (v6) of the QAA, originally developed by [26], which require wavelengths spanning 412 – 555 nm and 412 – 670 nm, respectively, including the central VIS wavelengths provided by legacy ocean color sensors. The effects of separating the dissolved and detrital contributions, i.e., $a_{\mathrm{CDM}}(440)$ versus $a_{\mathrm{CDOM}}(440)$, are considered further in the discussion section of this chapter.

2.2.8 Summary of algorithm validation statistics

Algorithm performance was assessed using RMSE statistics, as follows:

$$\mathrm{RMSE} \;=\; \left[\frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{n}\right]^{1/2}\,, \tag{2.6}$$

where X and Y correspond to the algorithm and measured values of $a_{\mathrm{CDOM}}(440)$, respectively. When reported as a percentage, RMSE was normalized by the range in the measured values. For comparison of matchup statistics and evaluation of subjective subcategories, algorithm uncertainty was assessed using root mean square log-error (RMSLE) and log-scale bias (LB) which have been used for evaluating the success of the Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) mission, e.g., [16], and were derived as follows:

$$\mathrm{RMSLE} \;=\; 100\left[\frac{\sum_{i=1}^{n}\big(\log_{10}(X_i) - \log_{10}(Y_i)\big)^2}{n}\right]^{1/2}\,, \tag{2.7}$$

and

$$\mathrm{LB} \;=\; 100\,\frac{\sum_{i=1}^{n}\big(\log_{10}(X_i) - \log_{10}(Y_i)\big)}{n}\,. \tag{2.8}$$

Uncertainty requirements of RMSLE to within 35% were described in [68] for operational ocean color products (i.e., Chl a). Allotting half of the total uncertainty budget to *in situ* activities (which includes radiometry and algorithm development), and assuming quadrature summing of uncertainties, the target uncertainty for a near-surface activity (C-AERO operated at LSA) would be an RMSLE to within 25% in order to satisfy calibration and validation requirements for legacy oceanic measurements [69]. For future NASA missions, e.g., the Phytoplankton, Aerosol, Cloud, ocean Ecosystem mission (PACE), calibration and validation activities within open ocean waters should aim to satisfy RMSLE to within 15% [69]. In smaller and more challenging inland water bodies, which were not targeted by legacy ocean color missions, next generation activities may aim to satisfy the 25% criteria.

An uncertainty budget was also estimated for C-AERO EMA measurements following the first-order-first-moment (FOFM) uncertainty propagation methods described in [70]. Briefly, the FOFM uncertainty in the EMA product is the quadrature sum of the radiometric and model contributions, each derived through partial differentiation. Spectrally dependent radiometric uncertainties are estimated from [56,65], and algorithm coefficient uncertainties are estimated by non-parametric bootstrapping.

2.3 Results

2.3.1 Case-1 evaluation of a $K_d(\lambda)$ end-member approach

A straightforward algorithmic approach for estimating $a_{\mathrm{CDOM}}(440)$ was evaluated in [24], in which ratios of spectrally separate $K_d(\lambda)$ pairs were predictors of $a_{\mathrm{CDOM}}(440)$. The case-1 relationships between $K_d(\lambda)$ band-ratios and $a_{\mathrm{CDOM}}(440)$ were evaluated herein by comparing the GSyn and MSyn datasets with $K_d(\lambda)$ EMA algorithms using the [24] dataset. EMA algorithms based on linear [24] and power-law fits were considered for $K_d(320)/K_d(780)$ and $K_d(412)/K_d(670)$, respectively, which produced normally distributed residuals between algorithm and data pairs, shown in Fig. 2.2. GSyn and MSyn datasets were not extended beyond the Chl a value of $25\,\mathrm{mg\,m^{-3}}$, consistent with the range considered in [38,44], and were compared to the EMA algorithms using unbiased percent difference (UPD), or the APD with the mean of the paired values taken as the reference. The GSyn a_{CDOM} slope coefficients were varied between 0.01 and $0.03\,\mathrm{nm^{-1}}$ for sensitivity testing, shown in the Fig. 2.2 panel inlays.

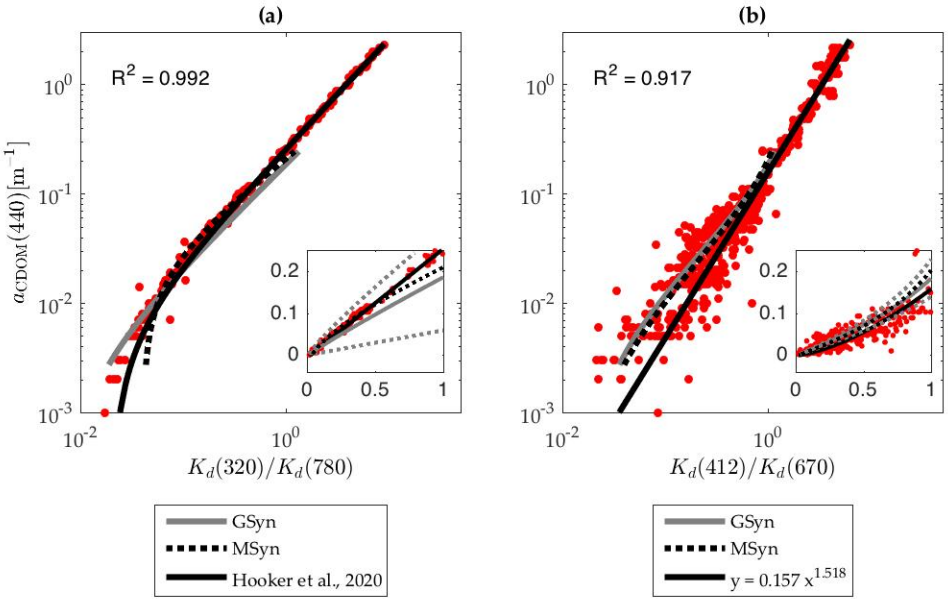

Figure 2.2: $K_d(\lambda_1)/K_d(\lambda_2)$ *relationships to* $a_{\mathrm{CDOM}}(440)$*, with the dataset of Hooker et al., 2020 shown in red and overlaid, respectively, as follows:* ***(a)*** *GSyn with* $S = 0.018\,\mathrm{nm}^{-1}$ *in solid grey, MSyn in dashed black, and EMA from Hooker et al., 2020 in solid black; and* ***(b)*** *GSyn with* $S = 0.018\,\mathrm{nm}^{-1}$ *in solid grey, MSyn in dashed black, and a mean absolute deviation power-law fit in solid black.* R^2 *is reported for* $\log_{10}$*-transformed products. Panel inlays show relationships in linear space, with sensitivity testing for GSyn with* $S = 0.01, 0.03\,\mathrm{nm}^{-1}$ *included in dashed grey.*

For the UV-NIR $K_d(320)/K_d(780)$ pair, the EMA relationship proposed in [24] is within 25% UPD of MSyn for $a_{\mathrm{CDOM}}(440)$ above $0.006\,\mathrm{m}^{-1}$ ($\mathrm{Chl}\,a > 0.07\,\mathrm{mg\,m}^{-3}$) and also to within 25% of GSyn for $a_{\mathrm{CDOM}}(440)$ between $0.009 - 0.098\,\mathrm{m}^{-1}$ ($0.13 < \mathrm{Chl}\,a < 5.91\,\mathrm{mg\,m}^{-3}$). For the VIS $K_d(412)/K_d(670)$ pair, the power-law EMA relationship is within 25% UPD of MSyn for $a_{\mathrm{CDOM}}(440)$ between $0.062 - 0.201\,\mathrm{m}^{-1}$ ($2.88 < \mathrm{Chl}\,a < 18.52\,\mathrm{mg\,m}^{-3}$), and to within 25% of GSyn for $a_{\mathrm{CDOM}}(440)$ above $0.076\,\mathrm{m}^{-1}$ ($\mathrm{Chl}\,a > 3.96\,\mathrm{mg\,m}^{-3}$).

The $K_d(412)/K_d(670)$ EMA power-law algorithm produces lower $a_{\mathrm{CDOM}}(440)$ values than the synthetic case-1 datasets, although both remain within the spread of the data due to the increased variability of the $K_d(412)/K_d(670)$ relationship compared to $K_d(320)/K_d(780)$. For GSyn, an expanded range in S produces greater variability in $K_d(320)/K_d(780)$ than is apparent in the *in situ* data products. For the legacy wavelength pair, the effect of S on the GSyn data products is less than the variability of the *in situ* data products for the clearer fraction of waters, e.g., $a_{\mathrm{CDOM}}(440)$ below $0.1\,\mathrm{m}^{-1}$.

2.3.2 EMA algorithms derived for $\left[L_W(\lambda)\right]_{\mathrm{N}}$ pairs

EMA algorithms for UV-NIR (Λ_{780}^{320}) and VIS (Λ_{670}^{412}, Λ_{555}^{443}, and Λ_{625}^{465}) $\left[L_W(\lambda)\right]_{\mathrm{N}}$ pairs were generated from the OCEAN, GLOBC, and NOMAD datasets. For all datasets, the dynamic range in $\Lambda_{\lambda_2}^{\lambda_1}$ increases with spectral separation between wavelength pairs, with the greatest ranges observed in Λ_{780}^{320} for the OCEAN and GLOBC datasets and in Λ_{670}^{412} for the NOMAD dataset, which lacks the relevant UV and NIR wavelengths. The $\Lambda_{\lambda_2}^{\lambda_1}$ EMA algorithm fits are overlaid on the OCEAN and GLOBC datasets in Figs. 2.3 and 2.4, respectively, with the PSyn synthetic case-1 dataset included for relevant legacy wavelengths.

Compared to the coefficients derived from the GLOBC dataset, the OCEAN dataset normalizing coefficients (A) were 8.6% greater (Λ_{780}^{320}) and 0.3% less (Λ_{670}^{412}), and the exponential coefficients (B) were 3.0% (Λ_{780}^{320}) and 18.1% (Λ_{670}^{412}) greater. NOMAD A and B coefficients for the Λ_{670}^{412} algorithm were elevated by 17.7% and 33.6%, respectively, compared with the algorithm fit to the GLOBC dataset. The coefficient differences between datasets for the VIS $\Lambda_{\lambda_2}^{\lambda_1}$ pairs are in some instances greater than the standard errors of the coefficients, indicating differences in the technology, distribution, or range of the datasets. For example, if a higher-order relationship were appropriate, differences in the range of the data would alter the coefficients of a power-law fit. If the GLOBC dataset is confined to the $a_{\mathrm{CDOM}}(440)$ range in the OCEAN dataset ($0.004-0.613\,\mathrm{m}^{-1}$), B increases by about 11% and the uncertainties for the coefficients overlap. Higher-order algorithms were not pursued herein given the size and variability of the *in situ* datasets, and because the GLOBC dataset suggests log-linearity for the Λ_{780}^{320} and Λ_{670}^{412} algorithms within higher $a_{\mathrm{CDOM}}(440)$ waters.

The relevant $\Lambda_{\lambda_2}^{\lambda_1}$ comparisons are considered using the GLOBC dataset, which contains the greatest range in $a_{\mathrm{CDOM}}(440)$ with the most uniform distribution, and which was obtained with extrapolation intervals nearer to the surface and with finer VSR to improve measurement at end-member wavelengths. EMA algorithms fit to the GLOBC dataset indicate that the dynamic range in $\Lambda_{\lambda_2}^{\lambda_1}$ decreases by about two orders of magnitude between Λ_{780}^{320} and Λ_{555}^{443}, in agreement with the perspective of [37]. Due to decreasing sensitivity for the Λ_{555}^{443} and Λ_{625}^{465} algorithms within the global waters dataset, the subsequent analysis herein focuses on validation for the Λ_{780}^{320} and Λ_{670}^{412} algorithms.

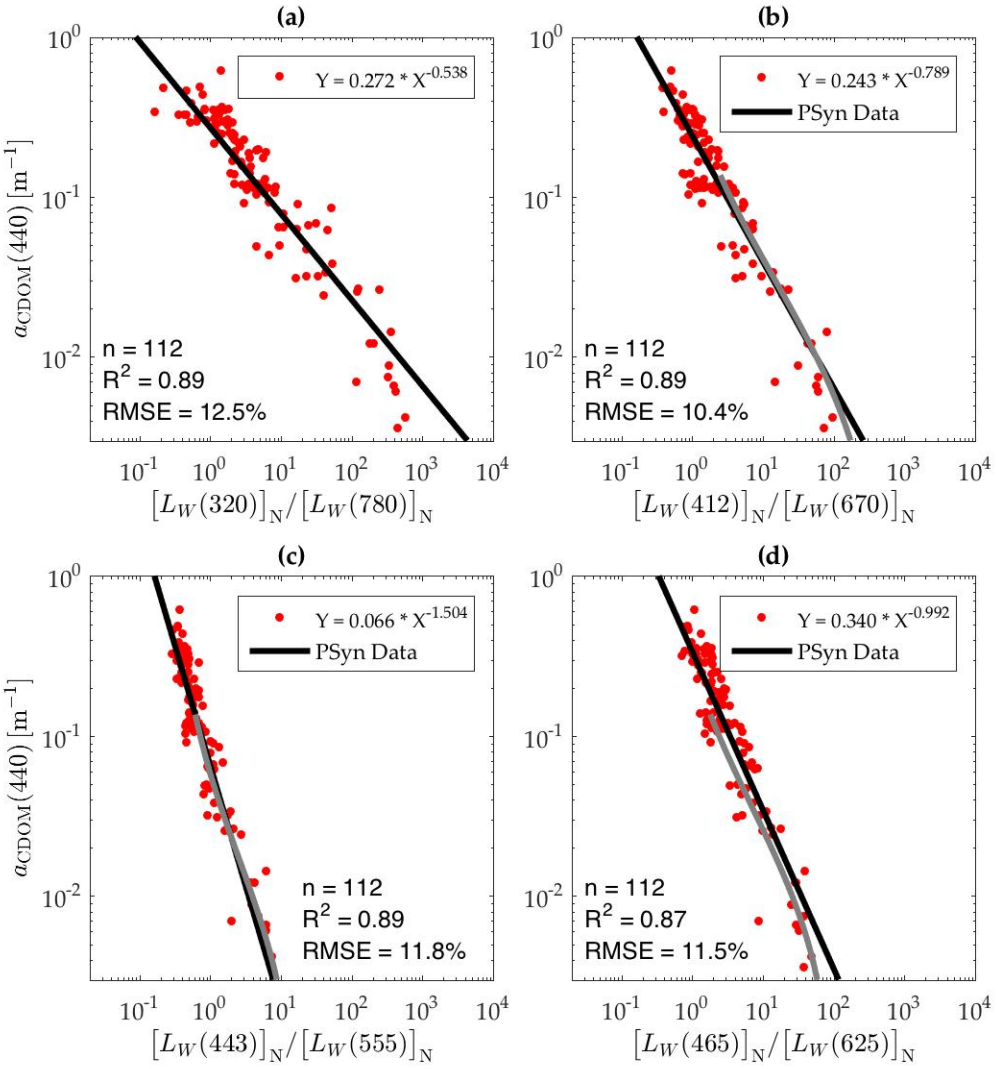

Figure 2.3: *OCEAN dataset (red dots), EMA algorithm (solid black line), and PSyn case-1 dataset (solid grey line) for:* ***(a)*** Λ_{780}^{320}*;* ***(b)*** Λ_{670}^{412}*;* ***(c)*** Λ_{555}^{443}*; and* ***(d)*** Λ_{625}^{465}*.* R^2 *is reported for* $\log_{10}$*-transformed products.*

2.3.3 Validation of $a_{\mathrm{CDOM}}(440)$ algorithms

The performance of the EMA (Λ_{780}^{320} and Λ_{670}^{412}), GSM, and QAA algorithms was evaluated using the GLOBC dataset, which includes oceanic, coastal, and inland water bodies and spans nearly three decades ($0.001 - 0.861\,\mathrm{m}^{-1}$) of $a_{\mathrm{CDOM}}(440)$. The EMA algorithms considered were those fit to the OCEAN dataset in order to maintain independence from the GLOBC validation dataset. Results for each algorithm are shown in Fig. 2.5.

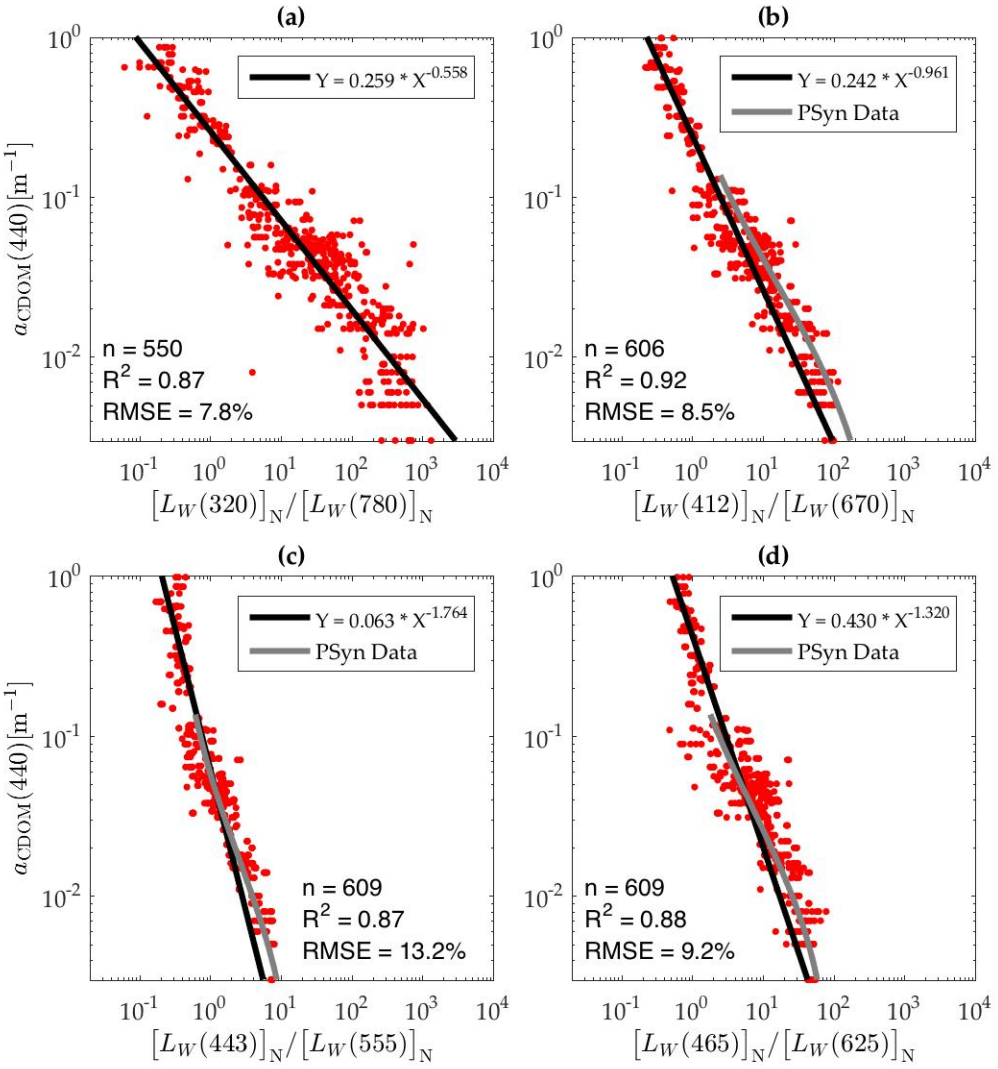

Figure 2.4: *GLOBC dataset (red dots), EMA algorithm (solid black line), and PSyn case-1 dataset (solid grey line) for:* ***(a)*** Λ_{780}^{320}*;* ***(b)*** Λ_{670}^{412}*;* ***(c)*** Λ_{555}^{443}*; and* ***(d)*** Λ_{625}^{465}*.* R^2 *is reported for* $\log_{10}$*-transformed products.*

The GSM and QAA algorithms produced the lowest RMSLE within the clearer waters of the GLOBC dataset (generally oceanic, although with some inland targets, such as LT), while the EMA algorithms produced the lowest RMSLE within the higher $a_{\mathrm{CDOM}}(440)$ portion of the dataset (generally coastal and inland water bodies), resulting in lower RMSE for the EMA algorithms across the full range. Applying the $a_{\mathrm{CDOM}}(440)$ threshold of $0.1\,\mathrm{m}^{-1}$ as a partition separating primarily oceanic from primarily coastal or inland waters,

Table 2.1: *The* $\left[L_W(\lambda)\right]_{\mathrm{N}}$ *EMA power-law coefficients derived from in situ datasets.*

$\boldsymbol{\lambda}$ **pair**	**Fit**	$\boldsymbol{u_m(A)}$ **[m^{-1}]**	$\boldsymbol{u_m(B)}$	**Dataset**
$\left[L_W(320)\right]_{\mathrm{N}}/\left[L_W(780)\right]_{\mathrm{N}}$	$y = 0.2814x^{-0.5420}$ †	0.0413	0.0662	OCEAN
$(x = \Lambda_{780}^{320})$	$y = 0.2589x^{-0.5583}$	0.0085	0.0197	GLOBC
$\left[L_W(412)\right]_{\mathrm{N}}/\left[L_W(670)\right]_{\mathrm{N}}$	$y = 0.2416x^{-0.7874}$	0.0225	0.0583	OCEAN
$(x = \Lambda_{670}^{412})$	$y = 0.2423x^{-0.9614}$	0.0100	0.0359	GLOBC
	$y = 0.2852x^{-0.6379}$	0.0102	0.0393	NOMAD
$\left[L_W(443)\right]_{\mathrm{N}}/\left[L_W(555)\right]_{\mathrm{N}}$	$y = 0.0660x^{-1.5227}$	0.0059	0.1922	OCEAN
$(x = \Lambda_{555}^{443})$	$y = 0.0630x^{-1.7640}$	0.0076	0.1293	GLOBC
	$y = 0.0649x^{-1.3992}$	0.0029	0.0964	NOMAD
$\left[L_W(465)\right]_{\mathrm{N}}/\left[L_W(625)\right]_{\mathrm{N}}$	$y = 0.3491x^{-0.9960}$	0.0405	0.0690	OCEAN
$(x = \Lambda_{625}^{465})$	$y = 0.4297x^{-1.3204}$	0.0226	0.0796	GLOBC
	$y = 0.1278x^{-0.5641}$	0.0433	0.1001	NOMAD

†Algorithm applied to airborne imagery in Section 2.3.5.

the GSM and QAA algorithms failed or produced non-physical values (i.e., negative or above $4\,\mathrm{m}^{-1}$) for 11 (5%) and 2 (1%) of the high $a_{\mathrm{CDOM}}(440)$ measurements, respectively. Not considering the non-physical retrievals for the GSM and QAA algorithms (but including them for the EMA approaches), RMSLE within the high ($> 0.1\,\mathrm{m}^{-1}$) $a_{\mathrm{CDOM}}(440)$ range was 12.8%, 13.2%, 21.5%, and 30.1% for the Λ_{780}^{320}, Λ_{670}^{412}, GSM, and QAA algorithms, respectively. Within this range the QAA produced positive LB (over-prediction) of 22.6%, although the method indicated high R^2, perhaps indicating its potential for regional optimization and in part driven by increased detrital contributions, described more in the discussion section. For a higher $a_{\mathrm{CDOM}}(440)$ range of conservative waters ($> 0.5\,\mathrm{m}^{-1}$; made possible by the prior subjective classifications), the EMA algorithms approximately satisfy the next-generation 15% RMSLE criteria with 15.0% and 14.7% for Λ_{780}^{320} and Λ_{670}^{412} algorithms, respectively, while RMSLE increased for the GSM and QAA algorithms to 31.5% and 30.0%, respectively.

Within clear waters, or $a_{\mathrm{CDOM}}(440) \leq 0.1\,\mathrm{m}^{-1}$, the QAA algorithm resulted in one negative retrieval, while the EMA and GSM algorithms always produced physical retrievals. Ignoring the negative retrieval, the GSM and QAA algorithms performed best in the clear-water partition, with RMSLE of 24.3%, 22.6%, 19.6%, and 18.2% for the Λ_{780}^{320}, Λ_{670}^{412}, GSM,

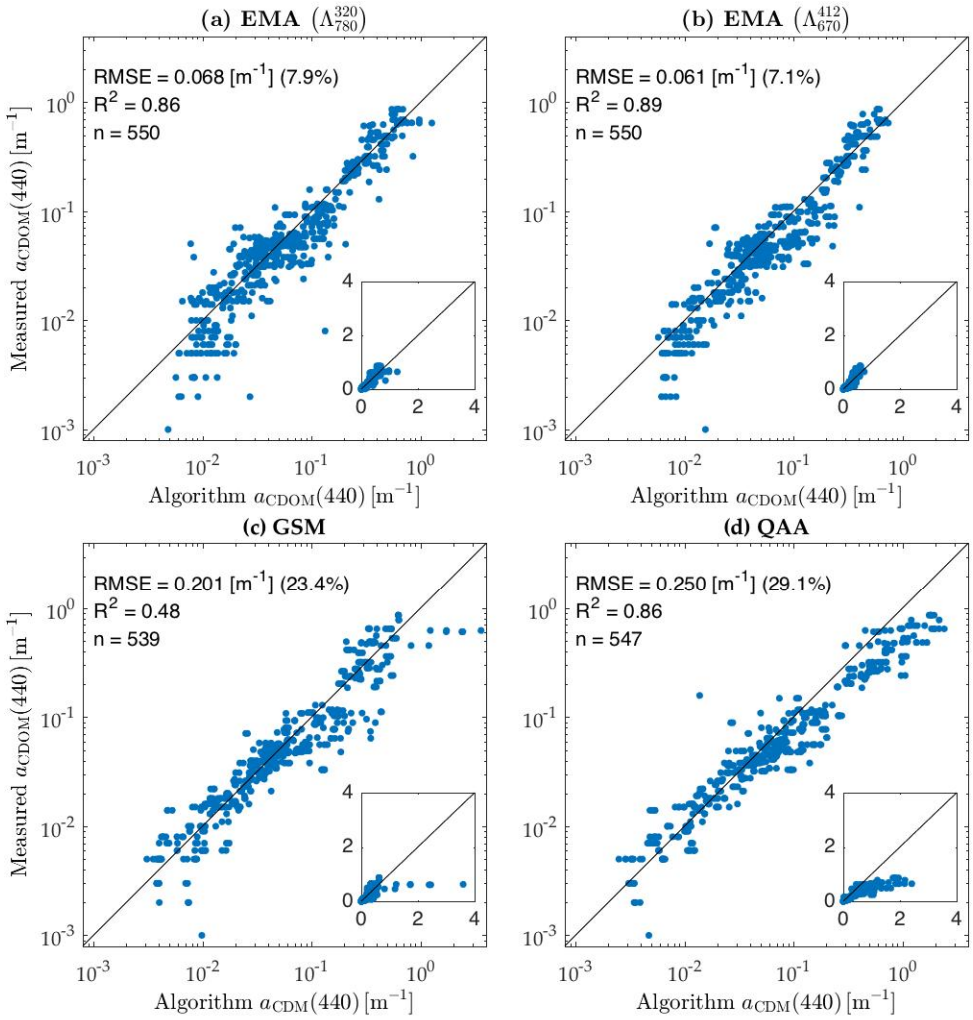

Figure 2.5: *Algorithm versus measured* $a_{\mathrm{CDOM}}(440)$ *from the GLOBC dataset for:* ***(a)*** Λ_{780}^{320}*;* ***(b)*** Λ_{670}^{412}*;* ***(c)*** *GSM; and* ***(d)*** *QAA. Linear axes are included in sub-panels, with a one-to-one line shown in solid black for all panels.*

and QAA algorithms, respectively. Performance of the Λ_{780}^{320} EMA algorithm was degraded relative to the Λ_{670}^{412} algorithm within this validation exercise in part by the difficulty of retrieving $\left[L_W(780)\right]_{\mathrm{N}}$ from an in-water profiler within clear waters, and in part by the fitting of the coefficients to the legacy OCEAN dataset, which produced a less negative exponential term than the corresponding fit to the GLOBC dataset. For example, if an iterative cross-validation exercise (80% – 20% modeling-validation split) is considered within the GLOBC dataset, which included technological improvements for retrieving UV and NIR wavelengths, the performance of the Λ_{780}^{320} algorithm improves in clear waters, with median

RMSLE values for the Λ_{780}^{320} and Λ_{670}^{412} EMA algorithms of 23.0% and 23.5%, respectively.

Although the $a_{\mathrm{CDOM}}(440) > 0.1\,\mathrm{m}^{-1}$ partition represents a small fraction of global waters by area, it constitutes a large portion of the total dynamic range in $a_{\mathrm{CDOM}}(440)$ relevant to a global algorithmic perspective. For example, over 33% or 95% (in $\log_{10}$- or linear-space, respectively) of the $a_{\mathrm{CDOM}}(440)$ range for conservative waters exceeds an $a_{\mathrm{CDOM}}(440)$ value of $0.1\,\mathrm{m}^{-1}$ based on the complete *in situ* range in the GLOBC dataset. Decreased performance of the GSM and QAA algorithms in a global dataset was anticipated given that both were tuned for oceanic ecosystems, the versions shown in Fig. 2.5 did not partition detrital and dissolved signals, and the GLOBC dataset is disproportionately non-oceanic. The GLOBC dataset contains roughly uniform density of $a_{\mathrm{CDOM}}(440)$ across the full range in $\log_{10}$-space, which was not in itself an objective of the sampling, but which resulted from the effort to represent a wide diversity of water bodies (e.g., lakes, rivers, bays, and marshes).

2.3.4 Efficacy of $a_{\mathrm{CDOM}}(440)$ retrievals in non-conservative water bodies

Measurements of non-conservative waters from the GLOBC dataset were assessed within classifications defined in [24]. The classes do not necessarily indicate that optical or ecological conditions invalidate a global algorithmic approach, but they apply available prior information in order to quantify algorithm performance in conditions where algorithms might otherwise be expected to degrade. The classifications are briefly summarized as follows: a) Resuspension: shallow regions with strong mixing (e.g., from tides or winds) which resuspends bottom material, b) Refilled or Flooded: lakes or rivers above their average mean high-water line or which have recently been replenished, c) Drought-Stricken: a water body significantly below normal elevation (water line), d) Harbor: a sheltered area for docking vessels, e) HAB: harmful algal blooms or elevated concentrations of phytoplankton, which may produce toxic compounds, f) Wetland or Marsh: shallow, tidally influenced estuarine areas, often with brackish properties, g) Polluted: waters containing anthropogenic sources that alter the natural water properties, h) Alkaline Lake: lacustrine ecosystems containing significantly elevated loads of salt compounds and generally high pH, i) River Mouth: regions where riverine ecosystems mix with larger water bodies, including lakes and bays, and j) Invasive Species: ecosystems perturbed by anthropogenic introduction of noxious macroalgae. The classifications are consistent with the presentation of [24], in which a more

detailed discussion is provided on the optically relevant water mass modifications that might be anticipated for each condition. The EMA performance results for the non-conservative (GLOBN) dataset subcategories were assessed using EMA algorithms fit to the conservative (GLOBC) dataset, and are shown in Table 2.2. The statistics shown indicate whether particular environments are expected to be consistent with the EMA algorithms derived herein, or whether a regionalized approach is supported (e.g., if RMSLE approximates LB, R^2 is significant, and n is large).

Despite decreases in algorithm efficacy for the non-conservative waters, some subjective categories satisfy the 25% RMSLE target for inland water bodies, including (for the Λ_{780}^{320} algorithm) harbors, HABs, river mouths, and refilled or flooded rivers and lakes. Drought-stricken and alkaline lakes indicate the potential for the development of a regionalized algorithmic approach based on high correlation (R^2) between $a_{\mathrm{CDOM}}(440)$ and Λ_{780}^{320}, RMSLE near 25%, and RMSLEcp (RMSLE with LB removed in quadrature) below 25%, with a relatively large number (n = 66 and 34) of observations considered for an inland waters ecosystem-specific dataset.

For drought-stricken lakes, harbors, and river mouths, classification is possible from a remote perspective. For example, maps of harbors and river mouths are accessible, and lake levels can be assessed through satellite imagery or by applying external knowledge of drought conditions. For other environments, classification is more difficult. For example, tidal estuaries are strongly affected by tidal cycles, which may alternate source water masses or lead to changes in depth and resuspension processes. An operator in a boat can better avoid mud flats or vegetation, although the validation dataset primarily sampled high tide conditions in these environments because access by boat was sometimes not possible during low tides.

2.3.5 Analysis of airborne imagery

The C-AERO EMA data products were evaluated with respect to contemporaneous water sampling, with EMA algorithms at all sites satisfying the next-generation RMSLE criteria for inland (25%) and oceanic (15%) waters, although water sample matchups were sparse (n = 3, 4, and 1 unique replicates at SFBD, LT, and MB, respectively). RMSLE values from the matchup validation were 19.4%, 13.6%, and 6.7% for the Λ_{780}^{320} algorithm at SFBD, LT, and MB, respectively, and 10.5%, 12.5%, and 11.2% for the Λ_{670}^{412} algorithm

Table 2.2: *Performance of* $\left[L_W(\lambda)\right]_{\mathrm{N}}$ *end-member algorithms within various water body classifications. Bold* R^2 *values indicate satisfaction of* $P \leq 0.01$*, and bold RMSLE values indicate satisfaction of the 25% algorithm uncertainty target for next generation calibration and validation activities in inland waters.*

Algorithm	**Water Type**	**n**	$\mathbf{R^2}$	**RMSE**	**RMSLE**	**LB**
				$[\mathrm{m}^{-1}]$	(%)	(%)
$\left[L_W(320)\right]_{\mathrm{N}}/\left[L_W(780)\right]_{\mathrm{N}}$	Resuspension	66	0.15	0.031	30.8	-11.5
(Λ_{780}^{320})	Refilled or Flooded	15	**0.43**	0.219	**18.3**	-10.2
$\left(y = 0.2589x^{-0.5583}\right)$	Drought-Stricken	60	**0.86**	0.194	26.9	19.7
	Harbor	48	**0.70**	0.120	**16.2**	2.1
	HAB	11	**0.71**	0.150	**25.0**	-14.2
	Wetland or Marsh	28	0.18	0.146	39.9	26.9
	Polluted	38	0.04	0.142	40.7	20.3
	Alkaline Lake	34	**0.49**	0.165	37.4	32.1
	River Mouth	16	**0.58**	0.113	**10.3**	-2.3
	Invasive Species	10	**0.75**	0.165	50.4	48.0
$\left[L_W(412)\right]_{\mathrm{N}}/\left[L_W(670)\right]_{\mathrm{N}}$	Resuspension	66	0.07	0.053	36.3	-15.6
(Λ_{670}^{412})	Refilled or Flooded	15	0.11	0.281	27.3	-12.8
$\left(y = 0.2423x^{-0.9614}\right)$	Drought-Stricken	60	**0.76**	0.169	26.8	18.0
	Harbor	48	**0.82**	0.095	**17.7**	8.7
	HAB	11	**0.97**	0.095	**13.7**	-8.9
	Wetland or Marsh	28	0.00	0.213	48.7	33.9
	Polluted	38	**0.43**	0.065	26.3	9.9
	Alkaline Lake	34	**0.39**	0.174	38.4	31.4
	River Mouth	16	**0.79**	0.100	**13.0**	-7.3
	Invasive Species	10	**0.99**	0.105	28.8	28.7

at SFBD, LT, and MB, respectively. The Λ_{670}^{412} matchup results were in overall qualitative agreement with the FOFM approach of [70], in that they indicated lower percent uncertainties for the EMA approach within higher $a_{\mathrm{CDOM}}(440)$ water masses. FOFM results were sensitive to decisions of radiometric and algorithm coefficient uncertainties, and cor-

responded to RMSLE for the Λ_{780}^{320} and Λ_{670}^{412} algorithms below the 15% criteria at all sites using the comprehensive and spectrally dependent radiometric uncertainty budgets outlined in [56] and [65]. FOFM analysis of the algorithms fit to the legacy OCEAN dataset indicate lower uncertainty for the Λ_{670}^{412} EMA algorithm, while analysis of the algorithms fit to the next-generation GLOBC dataset indicate the lowest uncertainty for the Λ_{780}^{320} EMA algorithm. FOFM results for the OCEAN and GLOBC coefficient uncertainties reveal that the end-member perspective requires high radiometric data quality in order to achieve algorithmic improvements from increased spectral separation of waveband pairs.

Coincident water samples for the SFBD C-AERO flights were obtained in Grizzly Bay and not in San Pablo Bay, although the observed structure in $a_{\mathrm{CDOM}}(440)$ within San Pablo Bay is consistent with sampling in a prior year, as well as information on tidal cycles and bathymetry. In particular, a transition between roughly $0.25\,\mathrm{m}^{-1}$ (yellow) and $0.35\,\mathrm{m}^{-1}$ (red) is apparent in the San Pablo image, which occurs adjacent to the 2 m bathymetric contour (the bottom depth was greater than the 0.1% light level within the regions discussed herein). A similar $a_{\mathrm{CDOM}}(440)$ gradient was observed *in situ* during prior sampling, with $a_{\mathrm{CDOM}}(440)$ of $0.554\,\mathrm{m}^{-1}$ measured in the northern edge of San Pablo Bay, compared to $0.337\,\mathrm{m}^{-1}$ measured roughly 13 km due south. Measurements at four sites farther south into San Francisco Bay recorded a mean value of $0.207\,\mathrm{m}^{-1}$. The longitudinal gradient is also supported by *in situ* sampling within the Carquinez Strait joining San Pablo Bay and Grizzly Bay, with a mean $a_{\mathrm{CDOM}}(440)$ value of $0.553\,\mathrm{m}^{-1}$ and decreased salinity (9 ppt) relative to the San Francisco Bay water (28 – 30 ppt) recorded on the same day as the C-AERO flights. The C-AERO observations of SFBD were obtained on a flood tide, with an eastward-flowing current at the in-water sampling locations. Although the Grizzly Bay SFBD matchup site was in a tidally-mixed system denoting the non-conservative Resuspension classification, application of the C-AERO EMA algorithm satisfied the 25% RMSLE criteria.

2.4 Discussion

2.4.1 Consistency of EMA with case-1 parameterizations

A band-ratio algorithmic approach in which spectrally separate $K_d(\lambda)$ or $\left[L_W(\lambda)\right]_{\mathrm{N}}$ end-member pairs are related to $a_{\mathrm{CDOM}}(440)$ was evaluated using synthetic bio-optical

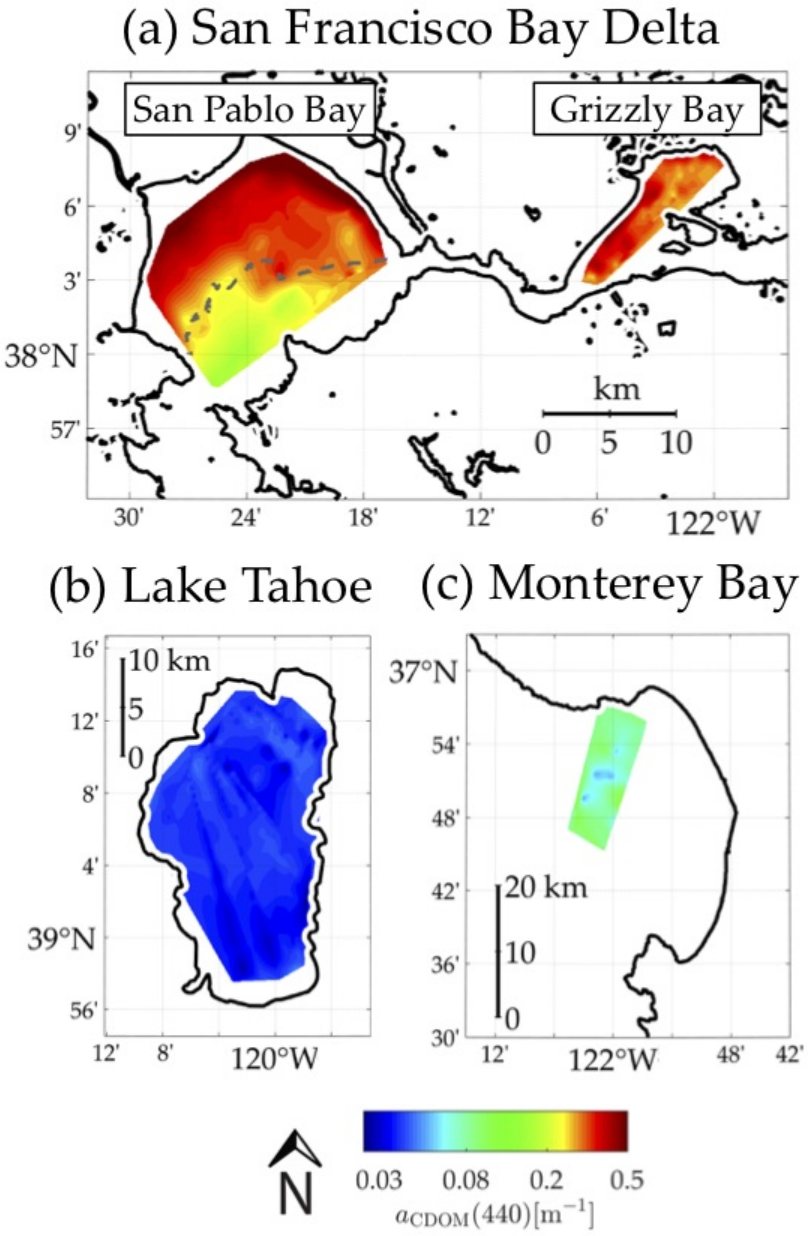

Figure 2.6: *The Λ_{780}^{320} EMA Algorithm applied to C-AERO measurements obtained at LSA and under clear-sky conditions at the following sites:* ***(a)*** *SFBD with larger San Pablo Bay on the left and smaller Grizzly Bay on the right (the northwest 2 m bathymetric contour is overlaid for San Pablo Bay with a dashed grey line);* ***(b)*** *LT; and* ***(c)*** *MB.*

datasets derived from case-1 parameterizations. In-water EMA algorithms for UV-NIR $K_d(320)/K_d(780)$ and VIS $K_d(412)/K_d(670)$ pairs indicated that case-1 parameterizations produce similar algorithmic relationships as those derived using a global *in situ* dataset of in-water measurements within a Chl a range relevant to case-1 parameterizations. $K_d(\lambda)$ EMA algorithm residuals were within the variability due to a global range in S for UV-NIR but not VIS band-ratios. Approximations for $\frac{D_o(\lambda_1)}{D_o(\lambda_2)}$ following [50] suggest that variability due to solar geometry is less than other natural sources considered herein.

If the derivation of the synthetic GSyn dataset is considered (i.e., a band-ratio of Eq. 2.4) the relative contributions of the optical constituents within each of the UV, VIS, and NIR spectral domains demonstrate the basis for the EMA approach. For example, if a case-1 condition is considered in which $K_d(\lambda)$ is approximately proportional to the total absorption coefficient (e.g., sun at zenith with $a \gg b_b$), the greatest spectral dependency in a UV-NIR $K_d(\lambda)$ band-ratio is due to CDOM, i.e., $a_{\mathrm{CDOM}}(780) \ll a_{\mathrm{CDOM}}(320)$, and an approximately linear EMA relationship emerges. If the same logic is applied to the VIS EMA, the absorption contribution by phytoplankton increases (relative to that of CDOM) and contributes non-linear spectral dependencies [38], thereby supporting a non-linear EMA relationship. These results are supported by the linear and power-law algorithms shown in Fig. 2.2.

Similar logic (partitioning constituent contributions across the UV, VIS, and NIR domains) is relevant to the $\left[L_W(\lambda)\right]_{\mathrm{N}}$ EMA approach, although $\left[L_W(\lambda)\right]_{\mathrm{N}}$ EMA performance is degraded by sensitivity to brightness, as well as to technological limitations of in-water measurements, i.e., no $L_u(z,\lambda)$ within the upper 0.3 m of the water column. Variability in target brightness (e.g., due to particle loading) increases uncertainty of the $\left[L_W(\lambda)\right]_{\mathrm{N}}$ EMA relative to $K_d(\lambda)$ EMA, but an $\left[L_W(\lambda)\right]_{\mathrm{N}}$ algorithmic approach is useful because $\left[L_W(\lambda)\right]_{\mathrm{N}}$ is derived directly from ocean color measurements, whereas $K_d(\lambda)$ derivations are spectrally incomplete [71], despite ongoing improvements, e.g., [72]. An alternate approach of estimating $K_d(\lambda)$ from above-water ocean color measurements and then deriving $a_{\mathrm{CDOM}}(440)$ was considered, wherein a neural network dataset of synthetic $K_d(\lambda)$ products (NNet) was derived from NOMAD above-water measurements [62]. The performance of an NNet EMA algorithm is compared with an $\left[L_W(\lambda)\right]_{\mathrm{N}}$ EMA algorithm in Fig. 2.7. The neural network solution indicates close agreement between NNet and MSyn, but does not show improvement in correlation between NNet $K_d(\lambda)$ ratios and $a_{\mathrm{CDOM}}(440)$. EMA from

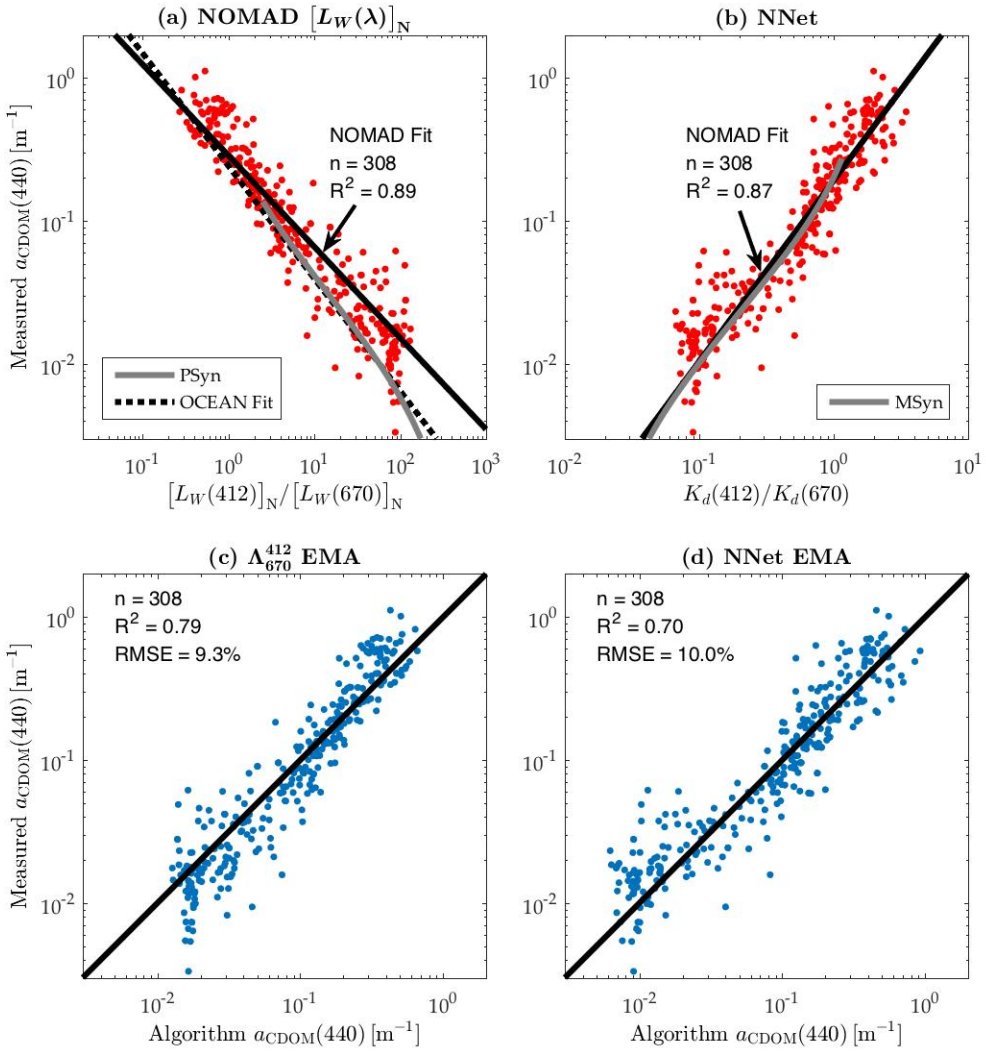

Figure 2.7: ***(a)*** *NOMAD dataset shown in red for the* Λ_{670}^{412} *products, with the PSyn synthetic case-1 dataset overlaid in solid grey, the* Λ_{670}^{412} *EMA algorithm fit to the OCEAN dataset overlaid in dashed black, and a least absolute deviation power-law fit of the NOMAD data in solid black;* ***(b)*** *NNet dataset shown in red for synthetic* $K_d(412)/K_d(670)$ *products, with the MSyn synthetic case-1 dataset overlaid in solid grey and a least absolute deviation power-law fit of the NNet data in solid black;* ***(c)*** *Validation results from the NOMAD* Λ_{670}^{412} *EMA with a one-to-one line indicated in solid black; and* ***(d)*** *Validation results from the NNet* $K_d(412)/K_d(670)$ *EMA with a one-to-one line indicated in solid black.* R^2 *values for panels* ***(a)*** *and* ***(b)*** *are derived from* $\log_{10}$*-transformed variables.*

NNet data results in degraded performance relative to EMA using the original $\left[L_W(\lambda)\right]_N$ measurements.

2.4.2 Performance of EMA algorithms

Three sets of $\left[L_W(\lambda)\right]_{\mathrm{N}}$ EMA algorithms were derived using the OCEAN, GLOBC, and NOMAD datasets, which contain coincident $a_{\mathrm{CDOM}}(440)$ and $\left[L_W(\lambda)\right]_{\mathrm{N}}$ *in situ* measurements. For all datasets, the dynamic range in $\Lambda_{\lambda_2}^{\lambda_1}$ increases with increasing spectral separation between wavelength pairs, which is anticipated to increase the sensitivity and robustness of the algorithmic approach. Scalar coefficients (A) were similar between the OCEAN and GLOBC datasets for the Λ_{780}^{320}, Λ_{670}^{412}, and Λ_{555}^{443} algorithm pairs. Exponential coefficients (B) obtained for each wavelength were more variable between datasets, with the GLOBC dataset consistently deriving more negative (steeper in log-log space) B values than the OCEAN dataset, which in turn produced more negative B values than the NOMAD dataset. The differences in coefficients are in some instances greater than the coefficient uncertainties derived by bootstrapping, indicating that model uncertainties are underestimated. The in-water datasets may contain measurement or processing differences, or differences in $a_{\mathrm{CDOM}}(440)$ range and distribution may cause differences in curve fitting. It is also possible that the $\left[L_W(\lambda)\right]_{\mathrm{N}}$ EMA relationship is more accurately modeled with a higher-order function, which is not evaluated herein.

The $\left[L_W(\lambda)\right]_{\mathrm{N}}$ EMA algorithms fit to the OCEAN dataset were compared with existing $a_{\mathrm{CDOM}}(440)$ methods using a global dataset containing similar representation of inland, coastal, and oceanic waters. Within the clear water fraction, $a_{\mathrm{CDOM}}(440) \leq 0.1\,\mathrm{m}^{-1}$, the lowest RMSLE was achieved by QAA, followed by GSM, Λ_{670}^{412} EMA, and Λ_{780}^{320} EMA, which all satisfied the legacy 25% oceanic criteria, but not the next-generation 15% criteria. RMSLE may have been elevated for the Λ_{780}^{320} algorithm by the difficulty of deriving $\left[L_W(780)\right]_{\mathrm{N}}$ from an in-water profiler in clear waters due to the strong attenuation scale for longer wavelengths, and by differences between the *in situ* datasets, supported by the results of cross-validation using the GLOBC dataset. For higher $a_{\mathrm{CDOM}}(440)$ waters ($> 0.1\,\mathrm{m}^{-1}$), RMSLE for the Λ_{780}^{320} and Λ_{670}^{412} EMA algorithms satisfied the 15% next-generation RMSLE criteria and produced physical retrievals for all spectra. RMSLE for GSM and QAA increased in the higher $a_{\mathrm{CDOM}}(440)$ fraction of the dataset, and resulted in 11 (6.3%) and 2 (1.1%) failed retrievals, respectively. QAA did not satisfy the 25% threshold for the high $a_{\mathrm{CDOM}}(440)$ fraction.

Validation results understate the usefulness of GSM and QAA, which derive other relevant bio-optical parameters, such as phytoplankton absorption and particle backscat-

tering, for which EMA has not been applied. The decreasing performance of the GSM and QAA algorithms for higher $a_{\mathrm{CDOM}}(440)$ was anticipated given that these methods were not tuned to inland waters, which constitute a significant portion of the GLOBC dataset. Inland waters are optically more variable than oceanic waters, and are less likely to match the internal bio-optical parameterizations of oceanic algorithms. For example, tuning the a_{CDOM} spectral slope parameter (S) could improve performance for GSM (QAA derives S). However, no relationship between algorithm uncertainty and S is observed in the GLOBC dataset, as shown for the western USA portion in Fig. 2.8.

The GSM and QAA methods generated large positive residuals (in addition to producing failed and non-physical retrievals) indicating that internal algorithm parameterizations, which were selected for oceanic waters, do not resemble the bio-optical properties of coastal and inland water masses and generally lead to overestimation of $a_{\mathrm{CDOM}}(440)$. EMA does not assume a slope parameter, although global average differences in S as a function of $a_{\mathrm{CDOM}}(440)$ might be relevant to the coefficients derived for the EMA algorithms across different datasets. The residuals generated for all algorithms indicate low correlation to S, suggesting that resolving natural S variability would not in itself be likely to significantly improve algorithm performance. Various regionalized improvements to the GSM and QAA algorithms have been proposed but are not the focus of this work given the goal of evaluating a global, rather than regional, algorithmic perspective. Positive biases by the GSM and QAA methods are also due to contributions from detrital matter, which is not separated out in the versions of the GSM and QAA models shown in Fig. 2.5. Improved methods for separating the dissolved and detrital contributions have been provided for the GSM [30] and QAA [31], and were also evaluated in this work. Following the update described by [31], the performance of the QAA method in the high $a_{\mathrm{CDOM}}(440)$ fraction ($> 0.1\,\mathrm{m}^{-1}$) of the GLOBC dataset improves, with RMSLE reduced from 30.1% to 25.8%, in agreement with the findings of [73]. A GSM update based on [30] results in an increase in RMSLE for the high $a_{\mathrm{CDOM}}(440)$ partition, perhaps because the updated version was tuned within arctic waters.

Consideration of non-conservative waters suggests that the global Λ_{780}^{320} EMA algorithmic approach is applicable to harbors, HABs, river mouths, and refilled or flooded lakes and rivers. Drought-stricken and alkaline lakes are potential candidates for developing regionally tuned Λ_{780}^{320} algorithmic approaches. Waters with high resuspension showed decreased correlation between $a_{\mathrm{CDOM}}(440)$ and Λ_{780}^{320} in agreement with the sensitivity of above-water

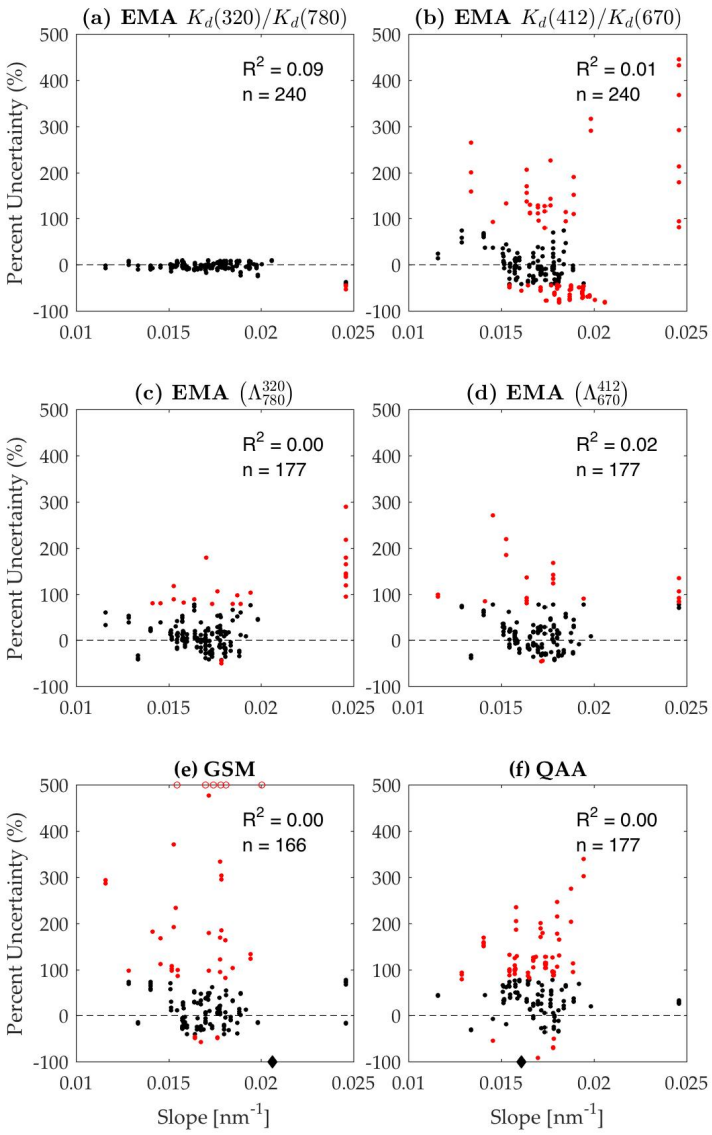

Figure 2.8: *The relative percent uncertainty from the western USA portion of the GLOBC dataset (positive indicates overestimation) as a function of* a_{CDOM} *spectral slope as follows:* (***a***) $K_d(320)/K_d(780)$ *EMA using a linear fit described in Hooker et al., 2020 ;* (***b***) $K_d(412)/K_d(670)$ *EMA using an absolute deviation power-law fit;* (***c***) Λ_{780}^{320} *EMA fit to the OCEAN dataset;* (***d***) Λ_{670}^{412} *EMA fit to the OCEAN dataset;* (***e***) *GSM with failed retrievals shown as open circles at the top of the y-axis and the defined* a_{CDOM} *slope shown as a black diamond on the x-axis; and* (***f***) *QAA with the median derived* a_{CDOM} *slope shown as a black diamond on the x-axis. Data points which satisfy the 25% RMSLE criteria are shown in black, and those which exceed the criteria are shown in red.*

methods to differences in brightness. Congruous with the case-1 versus case-2 logic for oceanic ecosystems, the conservative versus non-conservative classification provides a logical partition for algorithm development within inland waters, in which potentially problematic water bodies may be individually assessed for agreement with a global approach or else evaluated for regional tuning.

The Λ_{780}^{320} and Λ_{670}^{412} algorithms were applied to airborne $\left[L_W(\lambda)\right]_{\mathrm{N}}$ measurements of three environments, which spanned an order of magnitude in $a_{\mathrm{CDOM}}(440)$ and were independent from the OCEAN dataset used to fit the airborne EMA coefficients. Retrievals at all sites satisfied the RMSLE criteria for inland (25%) and oceanic waters (15%), although caution is warranted due to the limited number of available matchups. In the Λ_{780}^{320} comparison, LT produced the highest RMSLE matchup, which was anticipated given that EMA performance was worse for the clear-water partition in the algorithm validation intercomparison. LT is most similar to a case-1 oceanic water-type, for which the GSM and QAA methods produced the best retrievals. Λ_{780}^{320} also produced lower RMSLE in MB than in SFBD, wherein the latter is most similar to a Resuspension classification due to shallow depths plus strong tidal and wind-driven mixing, i.e., an ecosystem classification which produced elevated RMSLE and lower R^2 in the non-conservative validation comparison. The relative performance of the C-AERO products between matchup sites was consistent with the FOFM uncertainty derived from spectrally dependent radiometric uncertainties. FOFM with the GLOBC coefficient uncertainties indicated the lowest above-water algorithm uncertainty for the more spectrally separate Λ_{780}^{320} pair.

Previous work has indicated that inclusion of longer wavelengths (i.e., above 600 nm) for $a_{\mathrm{CDOM}}(440)$ algorithms improves $a_{\mathrm{CDOM}}(440)$ measurements in complex inland or freshwater environments by separating effects from detritus particles [73], which is in agreement with the findings herein. Other studies have suggested that $a_{\mathrm{CDOM}}(440)$ algorithm performance may decrease with the addition of information from short wavelengths (i.e., 412 and 443 nm) due to phytoplankton, decreasing radiometric sensitivity and increasing difficulty of atmospheric correction, e.g., [35, 36, 74]. Results discussed herein are not in opposition with these findings because the extension to UV wavelengths is anticipated to mitigate the relative importance of phytoplankton pigmentation, and because satellite radiometric capabilities and accuracy of atmospheric correction are not evaluated herein. However, this work indicates that based on an algorithmic perspective, obtaining accurate radiometry of end-members improves the robustness of remote $a_{\mathrm{CDOM}}(440)$ measurements within complex

waters. Already, airborne capabilities enable high signal-to-noise retrievals at relevant UV and NIR wavelengths [66], and various advances in atmospheric correction methodologies have increased the robustness of satellite retrievals in complex waters [75].

2.5 Conclusion

A band-ratio EMA algorithmic approach in which spectrally separate end-member pairs, either $\left[L_W(\lambda)\right]_{\mathrm{N}}$ or $K_d(\lambda)$, are related to $a_{\mathrm{CDOM}}(440)$ is shown to agree with case-1 relationships and to predict $a_{\mathrm{CDOM}}(440)$ with comparable log-scale statistics across nearly three decades of $a_{\mathrm{CDOM}}(440)$ range, including within oceanic, coastal, and inland conservative water bodies. The correlation to $a_{\mathrm{CDOM}}(440)$ and dynamic range expressed by Λ_{555}^{443} and Λ_{625}^{465} are less than that of Λ_{780}^{320} and Λ_{670}^{412}, indicating that increased spectral separation is useful from an algorithmic perspective. The performance of Λ_{780}^{320} and Λ_{670}^{412} EMA algorithms was greatest (RMSLE $< 15\%$) in the higher $a_{\mathrm{CDOM}}(440)$ partition ($> 0.1\,\mathrm{m}^{-1}$) of the global conservative waters dataset, a range generally associated with case-2 or inland waters. For the non-conservative dataset, the Λ_{780}^{320} EMA algorithm satisfied the RMSLE inland waters uncertainty criteria when applied to harbors, HABs, river mouths, and refilled or flooded lakes and rivers.

Satisfaction of the next-generation oceanic (15%) and inland (25%) RMSLE criteria for the Λ_{780}^{320} and Λ_{670}^{412} algorithms when applied to airborne imagery supports further research into EMA applications for remote measurement of $a_{\mathrm{CDOM}}(440)$ in inland waters, in which EMA confers multiple advantages. First, inland waters are overall more optically diverse than oceanic waters, and EMA does not require internal bio-optical parameterizations to specific water bodies. Second, inland waters contain a greater range in $a_{\mathrm{CDOM}}(440)$ than oceanic waters (LT to SFBD spans an order of magnitude), and EMA maintained stable log-scale uncertainty across a global range in $a_{\mathrm{CDOM}}(440)$ for the waters considered in this analysis. Third, many inland waters require increased spatial resolution of remote sensors because of their smaller areas and greater adjacency effects. Sensors with high spatial resolution often sacrifice spectral resolution, which is needed for inversion methods but not for two-band algorithms such as EMA. Fourth, inland waters are frequently shallow, and bottom effects are problematic for central (VIS) wavelengths. Λ_{780}^{320} is relatively insensitive to bottom effects (as is Λ_{670}^{412} within moderately turbid waters) due to the high attenuation of light by water at UV and NIR wavelengths.

Finally, the robustness of EMA across nearly three decades of $a_{\mathrm{CDOM}}(440)$ suggests that algorithms based on spectrally distant wavelengths, or end-members, are less sensitive to variations in optical complexity than algorithms which utilize the central, VIS wavelengths. Future work to improve an understanding of EMA may influence the technological goals for remote sensing of optically complex inland systems, for example by emphasizing the quality of retrievals at end-member wavebands. This will require high accuracy of atmospheric corrections, without preconceived brightness approximations (e.g., no black-pixel-assumptions), so that the relevant UV and NIR information is preserved.

Acknowledgements

We are grateful to contributions to this work from (in alphabetical order) Jim Brown (University of Miami), Liane Guild (NASA Ames), Randall Lind (Biospherical Instruments), Atsushi Matsuoka (University of New Hampshire), John Morrow (Biospherical Instruments), Kendra Negrey (University of California, Santa Cruz), Koji Suzuki (Hokkaido University) and Youhei Yamashita (Hokkaido University). Useful feedback on this manuscript was provided by Meredith McPherson (University of California, Santa Cruz), John P. Ryan (Monterey Bay Aquarium Research Institute), and three anonymous reviewers.

Funding: This work was supported by the NASA C-HARRIER campaign (grant number NNX17AK89G), as well as by the California State Water Resources Control Board (grant number A17-0553).

References

[1] W. S. Pegau, "Inherent optical properties of the central arctic surface waters," *Journal of Geophysical Research: Oceans*, vol. 107, no. C10, pp. SHE–16, 2002.

[2] G. C. Chang and T. D. Dickey, "Coastal ocean optical influences on solar transmission and radiant heating rate," *Journal of Geophysical Research: Oceans*, vol. 109, no. C1, 2004.

[3] N. Jiao, G. J. Herndl, D. A. Hansell, R. Benner, G. Kattner, S. W. Wilhelm, D. L. Kirchman, M. G. Weinbauer, T. Luo, F. Chen, *et al.*, "Microbial production of recalci-

trant dissolved organic matter: long-term carbon storage in the global ocean," *Nature Reviews Microbiology*, vol. 8, no. 8, p. 593, 2010.

[4] E. Rochelle-Newall and T. Fisher, "Production of chromophoric dissolved organic matter fluorescence in marine and estuarine environments: an investigation into the role of phytoplankton," *Marine Chemistry*, vol. 77, no. 1, pp. 7–21, 2002.

[5] N. B. Nelson, C. A. Carlson, and D. K. Steinberg, "Production of chromophoric dissolved organic matter by sargasso sea microbes," *Marine Chemistry*, vol. 89, no. 1-4, pp. 273–287, 2004.

[6] C. R. Castillo, H. Sarmento, X. A. Alvarez-Salgado, J. M. Gasol, and C. Marraséa, "Production of chromophoric dissolved organic matter by marine phytoplankton," *Limnology and Oceanography*, vol. 55, no. 1, pp. 446–454, 2010.

[7] J. R. Morrison and N. B. Nelson, "Seasonal cycle of phytoplankton uv absorption at the bermuda atlantic time-series study (bats) site," *Limnology and Oceanography*, vol. 49, no. 1, pp. 215–224, 2004.

[8] D. A. Jessup, M. A. Miller, J. P. Ryan, H. M. Nevins, H. A. Kerkering, A. Mekebri, D. B. Crane, T. A. Johnson, and R. M. Kudela, "Mass stranding of marine birds caused by a surfactant-producing red tide," *PLoS One*, vol. 4, no. 2, 2009.

[9] C. E. Del Castillo, F. Gilbes, P. G. Coble, and F. E. Müller-Karger, "On the dispersal of riverine colored dissolved organic matter over the west florida shelf," *Limnology and Oceanography*, vol. 45, no. 6, pp. 1425–1432, 2000.

[10] R. G. Spencer, K. D. Butler, and G. R. Aiken, "Dissolved organic carbon and chromophoric dissolved organic matter properties of rivers in the usa," *Journal of Geophysical Research: Biogeosciences*, vol. 117, no. G3, 2012.

[11] V. Vantrepotte, F.-P. Danhiez, H. Loisel, S. Ouillon, X. Mériaux, A. Cauvin, and D. Dessailly, "Cdom-doc relationship in contrasted coastal waters: implication for doc retrieval from ocean color remote sensing observation.," *Optics Express*, vol. 23, no. 1, pp. 33–54, 2015.

[12] R. M. Kudela and F. P. Chavez, "The impact of coastal runoff on ocean color during an el nino year in central california," *Deep Sea Research Part II: Topical Studies in Oceanography*, vol. 51, no. 10-11, pp. 1173–1185, 2004.

[13] N. B. Nelson, D. A. Siegel, C. A. Carlson, and C. M. Swan, "Tracing global biogeochemical cycles and meridional overturning circulation using chromophoric dissolved organic matter," *Geophysical Research Letters*, vol. 37, no. 3, 2010.

[14] N. B. Nelson and D. A. Siegel, "The global distribution and dynamics of chromophoric dissolved organic matter," *Annual Review of Marine Science*, vol. 5, pp. 447–476, 2013.

[15] K. L. Carder, R. G. Steward, G. R. Harvey, and P. B. Ortner, "Marine humic and fulvic acids: Their effects on remote sensing of ocean chlorophyll," *Limnology and Oceanography*, vol. 34, no. 1, pp. 68–81, 1989.

[16] W. W. Gregg and N. W. Casey, "Global and regional evaluation of the seawifs chlorophyll data set," *Remote Sensing of Environment*, vol. 93, no. 4, pp. 463–479, 2004.

[17] D. A. Siegel, S. Maritorena, N. B. Nelson, M. J. Behrenfeld, and C. R. McClain, "Colored dissolved organic matter and its influence on the satellite-based characterization of the ocean biosphere," *Geophysical Research Letters*, vol. 32, no. 20, 2005.

[18] C. A. Brown, Y. Huot, P. J. Werdell, B. Gentili, and H. Claustre, "The origin and global distribution of second order variability in satellite ocean color and its potential applications to algorithm development," *Remote Sensing of Environment*, vol. 112, no. 12, pp. 4186–4203, 2008.

[19] M. J. Sauer, C. S. Roesler, P. J. Werdell, and A. H. Barnard, "Under the hood of satellite empirical chlorophyll a algorithms: revealing the dependencies of maximum band ratio algorithms on inherent optical properties," *Optics Express*, vol. 20, no. 19, pp. 20920–20933, 2012.

[20] D. A. Siegel, M. J. Behrenfeld, S. Maritorena, C. R. McClain, D. Antoine, S. W. Bailey, P. S. Bontempi, E. Boss, H. M. Dierssen, S. C. Doney, *et al.*, "Regional to global assessments of phytoplankton dynamics from the seawifs mission," *Remote Sensing of Environment*, vol. 135, pp. 77–91, 2013.

[21] J. E. O'Reilly, S. Maritorena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. R. McClain, "Ocean color chlorophyll algorithms for seawifs," *Journal of Geophysical Research: Oceans*, vol. 103, no. C11, pp. 24937–24953, 1998.

[22] A. Bricaud, A. Morel, and L. Prieur, "Absorption by dissolved organic matter of the sea (yellow substance) in the uv and visible domains," *Limnology and Oceanography*, vol. 26, no. 1, pp. 43–53, 1981.

[23] M. S. Twardowski, E. Boss, J. M. Sullivan, and P. L. Donaghay, "Modeling the spectral shape of absorption by chromophoric dissolved organic matter," *Marine Chemistry*, vol. 89, no. 1-4, pp. 69–88, 2004.

[24] S. B. Hooker, A. Matsuoka, R. M. Kudela, Y. Yamashita, K. Suzuki, and H. F. Houskeeper, "A global end-member approach to derive $a_{\mathrm{CDOM}}(440)$ from near-surface optical measurements," *Biogeosciences*, vol. 17, no. 2, pp. 475–497, 2020.

[25] S. Maritorena, D. A. Siegel, and A. R. Peterson, "Optimization of a semianalytical ocean color model for global-scale applications," *Applied Optics*, vol. 41, no. 15, pp. 2705–2714, 2002.

[26] Z. Lee, K. L. Carder, and R. A. Arnone, "Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters," *Applied Optics*, vol. 41, no. 27, pp. 5755–5772, 2002.

[27] D. A. Aurin and H. M. Dierssen, "Advantages and limitations of ocean color remote sensing in cdom-dominated, mineral-rich coastal and estuarine waters," *Remote Sensing of Environment*, vol. 125, pp. 181–197, 2012.

[28] Y. Wang, F. Shen, L. Sokoletsky, and X. Sun, "Validation and calibration of qaa algorithm for cdom absorption retrieval in the changjiang (yangtze) estuarine and coastal waters," *Remote Sensing*, vol. 9, no. 11, p. 1192, 2017.

[29] I. D. Joshi and E. J. D'Sa, "An estuarine-tuned quasi-analytical algorithm (qaa-v): assessment and application to satellite estimates of spm in galveston bay following hurricane harvey," *Biogeosciences*, vol. 15, no. 13, pp. 4065–4086, 2018.

[30] A. Matsuoka, S. B. Hooker, A. Bricaud, B. Gentili, and M. Babin, "Estimating absorption coefficients of colored dissolved organic matter (cdom) using a semi-analytical

algorithm for southern beaufort sea waters: application to deriving concentrations of dissolved organic carbon from space," *Remote Sensing of Environment*, vol. 10, pp. 917–927, 2013.

[31] W. Zhu, Q. Yu, Y. Q. Tian, R. F. Chen, and G. B. Gardner, "Estimation of chromophoric dissolved organic matter in the mississippi and atchafalaya river plume regions using above-surface hyperspectral remote sensing," *Journal of Geophysical Research: Oceans*, vol. 116, no. C2, 2011.

[32] A. Morel and B. Gentili, "A simple band ratio technique to quantify the colored dissolved and detrital organic material from ocean color remotely sensed data," *Remote Sensing of Environment*, vol. 113, no. 5, pp. 998–1011, 2009.

[33] M. Kahru and B. G. Mitchell, "Seasonal and nonseasonal variability of satellite-derived chlorophyll and colored dissolved organic matter concentration in the california current," *Journal of Geophysical Research: Oceans*, vol. 106, no. C2, pp. 2517–2529, 2001.

[34] A. Mannino, M. G. Novak, S. B. Hooker, K. Hyde, and D. Aurin, "Algorithm development and validation of cdom properties for estuarine and continental shelf waters along the northeastern us coast," *Remote Sensing of Environment*, vol. 152, pp. 576–602, 2014.

[35] T. Kutser, D. C. Pierson, K. Y. Kallio, A. Reinart, and S. Sobek, "Mapping lake cdom by satellite remote sensing," *Remote Sensing of Environment*, vol. 94, no. 4, pp. 535–540, 2005.

[36] P. L. Brezonik, L. G. Olmanson, J. C. Finlay, and M. E. Bauer, "Factors affecting the measurement of cdom by remote sensing of optically complex inland waters," *Remote Sensing of Environment*, vol. 157, pp. 199–215, 2015.

[37] S. B. Hooker, J. H. Morrow, and A. Matsuoka, "Apparent optical properties of the canadian beaufort sea–part 2: The 1% and 1 cm perspective in deriving and validating aop data products," *Biogeosciences*, vol. 10, no. 7, pp. 4511–4527, 2013.

[38] A. Bricaud, M. Babin, A. Morel, and H. Claustre, "Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: Analysis and parameteriza-

tion," *Journal of Geophysical Research: Oceans*, vol. 100, no. C7, pp. 13321–13332, 1995.

[39] A. Morel and L. Prieur, "Analysis of variations in ocean color," *Limnology and Oceanography*, vol. 22, no. 4, pp. 709–722, 1977.

[40] L. Prieur and S. Sathyendranath, "An optical classification of coastal and oceanic waters based on the specific spectral absorption curves of phytoplankton pigments, dissolved organic matter, and other particulate materials 1," *Limnology and Oceanography*, vol. 26, no. 4, pp. 671–689, 1981.

[41] A. Morel, "Are the empirical relationships describing the bio-optical properties of case 1 waters consistent and internally compatible?," *Journal of Geophysical Research: Oceans*, vol. 114, no. C1, 2009.

[42] R. C. Smith and K. S. Baker, "Optical properties of the clearest natural waters (200–800 nm)," *Applied Optics*, vol. 20, no. 2, pp. 177–184, 1981.

[43] A. Morel, B. Gentili, H. Claustre, M. Babin, A. Bricaud, J. Ras, and F. Tieche, "Optical properties of the "clearest" natural waters," *Limnology and Oceanography*, vol. 52, no. 1, pp. 217–229, 2007.

[44] A. Bricaud, A. Morel, M. Babin, K. Allali, and H. Claustre, "Variations of light absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters: Analysis and implications for bio-optical models," *Journal of Geophysical Research: Oceans*, vol. 103, no. C13, pp. 31033–31044, 1998.

[45] C. Mobley and L. Sundman, "Hydrolight 5 ecolight 5 users' guide," *Sequoia Scientific Inc., US*, 2008.

[46] A. Morel and S. Maritorena, "Bio-optical properties of oceanic waters: A reappraisal," *Journal of Geophysical Research: Oceans*, vol. 106, no. C4, pp. 7163–7180, 2001.

[47] A. Morel, "Optical modeling of the upper ocean in relation to its biogenous matter content (case i waters)," *Journal of Geophysical Research: Oceans*, vol. 93, no. C9, pp. 10749–10768, 1988.

[48] A. Morel and D. Antoine, "Heating rate within the upper ocean in relation to its bio–optical state," *Journal of Physical Oceanography*, vol. 24, no. 7, pp. 1652–1665, 1994.

[49] H. R. Gordon, "Can the lambert-beer law be applied to the diffuse attenuation coefficient of ocean water?," *Limnology and Oceanography*, vol. 34, no. 8, pp. 1389–1409, 1989.

[50] X. Pan and R. C. Zimmerman, "Modeling the vertical distributions of downwelling plane irradiance and diffuse attenuation coefficient in optically deep waters," *Journal of Geophysical Research: Oceans*, vol. 115, no. C8, 2010.

[51] H. R. Gordon, O. B. Brown, R. H. Evans, J. W. Brown, R. C. Smith, K. S. Baker, and D. K. Clark, "A semianalytic radiance model of ocean color," *Journal of Geophysical Research: Atmospheres*, vol. 93, no. D9, pp. 10909–10924, 1988.

[52] A. N. Bashkatov and E. A. Genina, "Water refractive index in dependence on temperature and wavelength: a simple approximation," in *Saratov Fall Meeting 2002: Optical Technologies in Biophysics and Medicine IV*, vol. 5068, pp. 393–395, International Society for Optics and Photonics, 2003.

[53] M. S. Twardowski, H. Claustre, S. A. Freeman, D. Stramski, and Y. Huot, "Optical backscattering properties of the "clearest" natural waters," *Biogeosciences Discussions*, vol. 4, pp. 2441–2491, July 2007.

[54] Y. Huot, A. Morel, M. S. Twardowski, D. Stramski, and R. A. Reynolds, "Particle optical backscattering along a chlorophyll gradient in the upper layer of the eastern south pacific ocean," *Biogeosciences Discussions*, vol. 4, no. 6, pp. 4571–4604, 2007.

[55] C. C. Liu, J. D. Woods, and C. D. Mobley, "Optical model for use in oceanic ecosystem models," *Applied Optics*, vol. 38, no. 21, pp. 4475–4485, 1999.

[56] S. B. Hooker, R. N. Lind, J. H. Morrow, J. W. Brown, R. M. Kudela, H. F. Housekeeper, and K. Suzuki, "Advances in above-and in-water radiometry, volume 2: Autonomous atmospheric and oceanic observing systems," Tech. Rep. TP-2018-219033, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, 2018.

[57] J. Morrow, C. Booth, R. Lind, and S. Hooker, "The compact-optical profiling system (c-ops)," *Advances in Measuring the Apparent Optical Properties (AOPs) of Optically Complex Waters, NASA Tech. Memo*, vol. 215856, pp. 42–50, 2010.

[58] S. B. Hooker, R. N. Lind, J. H. Morrow, J. W. Brown, K. Suzuki, H. F. Houskeeper, T. Hirawake, and E. d. R. Maúre, "Advances in above-and in-water radiometry, volume 1: Enhanced legacy and state-of-the-art instrument suites," Tech. Rep. TP-2018-219033, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, 2018.

[59] S. B. Hooker, C. R. McClain, J. K. Firestone, T. L. Westphal, E.-N. Yeh, Y. Ge, and E. R. Firestone, "The seawifs bio-optical archive and storage system (seabass), part 1.," Tech. Rep. TM-104566, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, 1994.

[60] P. J. Werdell and S. W. Bailey, "The seawifs bio-optical archive and storage system (seabass): Current architecture and implementation," Tech. Rep. TM-2002-211617, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, 2002.

[61] S. Bailey, "An improved bio-optical data set for ocean color algorithm development and satellite data product variation," *Remote Sensing of Environment*, vol. 98, pp. 122–140, 2005.

[62] C. Jamet, H. Loisel, and D. Dessailly, "Retrieval of the spectral diffuse attenuation coefficient kd (λ) in open and coastal ocean waters using a neural network inversion," *Journal of Geophysical Research: Oceans*, vol. 117, no. C10, 2012.

[63] T. S. Moore, J. W. Campbell, and H. Feng, "A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms," *IEEE Transactions on Geoscience and Remote sensing*, vol. 39, no. 8, pp. 1764–1776, 2001.

[64] S. B. Hooker, G. Lazin, G. Zibordi, and S. McLean, "An evaluation of above-and in-water methods for determining water-leaving radiances," *Journal of Atmospheric and Oceanic Technology*, vol. 19, no. 4, pp. 486–515, 2002.

[65] S. B. Hooker, G. Zibordi, J. F. Berthon, and J. W. Brown, "Above-water radiometry in shallow coastal waters," *Applied Optics*, vol. 43, no. 21, pp. 4254–4268, 2004.

[66] R. M. Kudela, S. B. Hooker, H. F. Houskeeper, and M. McPherson, "The influence of signal to noise ratio of legacy airborne and satellite sensors for simulating next-generation coastal and inland water products," *Remote Sensing*, vol. 11, no. 18, p. 2071, 2019.

[67] R. Sibson, *Interpreting multivariate data*. John Wiley & Sons, 1981.

[68] S. B. Hooker and C. R. McClain, "The calibration and validation of seawifs data," *Progress in Oceanography*, vol. 45, no. 3-4, pp. 427–465, 2000.

[69] S. B. Hooker, C. R. McClain, and A. Mannino, "Nasa strategic planning document: A comprehensive plan for the long-term calibration and validation of oceanic biogeochemical satellite data," Tech. Rep. SP-2007-214152, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, 2007.

[70] L. McKinna, I. Cetinić, A. Chase, and P. J. Werdell, "Approach for propagating radiometric data uncertainties through nasa ocean color algorithms," *Frontiers in Earth Science*, vol. 7, p. 176, 2019.

[71] Z. Lee, C. Hu, S. Shang, K. Du, M. Lewis, R. Arnone, and R. Brewin, "Penetration of uv-visible solar radiation in the global oceans: Insights from ocean color remote sensing," *Journal of Geophysical Research: Oceans*, vol. 118, no. 9, pp. 4241–4255, 2013.

[72] F. Cao, C. G. Fichot, S. B. Hooker, and W. L. Miller, "Improved algorithms for accurate retrieval of uv/visible diffuse attenuation coefficients in optically complex, inshore waters," *Remote Sensing of Environment*, vol. 144, pp. 11–27, 2014.

[73] W. Zhu, Q. Yu, Y. Q. Tian, B. L. Becker, T. Zheng, and H. J. Carrick, "An assessment of remote sensing algorithms for colored dissolved organic matter in complex freshwater environments," *Remote Sensing of Environment*, vol. 140, pp. 766–778, 2014.

[74] K. Menken, P. Brezonik, and M. Bauer, "Influence of chlorophyll and humic color on reflectance spectra of lakes: implications for measurement of lake-water properties by remote sensing," *Lake Reservoir Management*, vol. 22, no. 3, pp. 179–190, 2006.

[75] M. A. Mograne, C. Jamet, H. Loisel, V. Vantrepotte, X. Mériaux, and A. Cauvin, "Evaluation of five atmospheric correction algorithms over french optically-complex waters for the sentinel-3a olci ocean color sensor," *Remote Sensing*, vol. 11, no. 6, p. 668, 2019.

[76] A. P. Vasilkov, J. R. Herman, Z. Ahmad, M. Kahru, and B. G. Mitchell, "Assessment of the ultraviolet radiation field in ocean waters from space-based measurements and full radiative-transfer calculations," *Applied Optics*, vol. 44, no. 14, pp. 2863–2869, 2005.

Appendix: Case-1 parameterizations

Scattering and absorption coefficients of seawater were obtained from [42] for red and NIR wavelengths and from [43] for blue and UV wavelengths. The backwards scattering contribution of seawater was approximated as one-half of the total scattering coefficient. Although scattering by seawater is a small contribution to total scattering, the contribution of seawater in the backwards direction is not negligible, for example exceeding the particulate contribution for the clearer fraction of the synthetic case-1 dataset.

The absorption by particles (including algal and non-algal) was parameterized as a power-law function of Chl a with spectral dependencies [44] as follows:

$$a_p(\lambda) \;=\; A_p(\lambda)\,[\text{Chl } a]^{E(\lambda)}\,. \tag{2.9}$$

Coefficients A_p and E were selected as the default coefficients for the HydroLight *New Case-1 Model* option (Mid-range UV), in which coefficients from various sources [7, 44, 76] are combined in order to extend the $a_p(\lambda)$ spectral domain to UV and NIR wavelengths [45].

The scattering by particles was parameterized in oligotrophic waters with spectral dependencies, which flatten with increasing Chl a [41] as follows:

$$b_p(\lambda) \;=\; \left(\frac{\lambda}{550}\right)^{v} 0.300\,[\text{Chl } a]^{0.62}\,, \tag{2.10}$$

where the spectral dependency is modified through $v = \max(0, (\log_{10}[\text{Chl } a] - 0.3)/2)$. The backwards component of b_p is obtained by multiplying b_p by the fraction of scattering which occurs in the backwards direction (B_{bp}). For case-1 waters, B_{bp} is thought to range between 0.5 and 1.5% [41]. B_{bp} of 1% is assumed herein, although the case-1 range is considered in sensitivity testing of this work's results.

Glossary

$a_{\mathrm{CDOM}}(440)$	Absorption coefficient of CDOM at 440 nm
$a(\lambda)$	Total spectral absorption coefficient
$a_p(\lambda)$	Spectral absorption coefficient of particles
$a_w(\lambda)$	Spectral absorption coefficient of water
$b_b(\lambda)$	Total spectral backscattering coefficient
$b_{bp}(\lambda)$	Spectral backscattering coefficient of particles
B_{bp}	Backscattering efficiency of particles
$b_{bw}(\lambda)$	Spectral backscattering coefficient of water
$b_p(\lambda)$	Spectral scattering coefficient of particles
$b_w(\lambda)$	Spectral scattering coefficient of water
C-AERO	Compact Airborne Environmental Radiometers for Oceanography
C-OPS	Compact-Optical Profiling System
C-PrOPS	Compact-Propulsion Option for Profiling Systems
CDOM	Colored dissolved organic matter
Chl *a*	Chlorophyll *a*
CIRPAS	Center for Interdisciplinary Remotely Piloted Aircraft Studies
$D_o(\lambda)$	Downwelling Distribution Function
$E_s(\lambda)$	Global solar irradiance
End-members	The most spectrally separate ocean color wavebands
EMA	End-member analysis
FOFM	First Order First Moment
GSM	Garver-Siegel-Maritorena algorithm [25]
IOP	Inherent Optical Property
$K_d(\lambda)$	Diffuse attenuation coefficient for spectral downward irradiance
$L_i(\lambda)$	Indirect (sky) radiance
$L_T(\lambda)$	Total radiance from the water surface
$L_u(z,\lambda)$	Upwelling radiance
$L_W(\lambda)$	Water-leaving radiance
$\left[L_W(\lambda)\right]_{\mathrm{N}}$	Normalized water-leaving radiance
$\Lambda_{\lambda_2}^{\lambda_1}$	The ratio $\left[L_W(\lambda_1)\right]_{\mathrm{N}}/\left[L_W(\lambda_2)\right]_{\mathrm{N}}$
LSA	Lowest Safe Altitude
LT	Lake Tahoe, California, USA
MB	Monterey Bay, California, USA
NIR	Near-infrared
NOMAD	NASA bio-Optical Marine Algorithm Dataset
S	Spectral slope of CDOM
SFBD	San Francisco Bay Delta, California, USA
SWIR	Short-wave infrared
QAA	Quasi-Analytical-Algorithm [26]
UV	Ultra-violet
VIS	Visible
VSR	Vertical Sampling Resolution

Chapter 3

Estimating chlorophyll *a* from hyperspectral particulate beam-attenuation in oceanic and coastal ecosystems

Overview: Electromagnetic theory predicts spectral dependencies in particle extinction efficiency that arise in the spectral vicinity of a narrow absorption band, for example produced by the primary phytoplankton photopigment, Chl *a*, at red wavelengths. The shapes of the predicted spectral anomalies depend on the size and refractivity of the particles. This chapter demonstrates that Chl *a* absorption can be accurately predicted from hyperspectral particulate beam-attenuation measurements. In addition, the relationships observed between the particle size distribution and the beam-attenuation spectral anomalies near the red absorption maxima are consistent with the predictions of electromagnetic theory.

The following was submitted to *Applied Optics* on 2 May 2020:

Chlorophyll absorption and phytoplankton size information inferred from hyperspectral particulate beam-attenuation

Henry F. Houskeeper, David Draper, Raphael M. Kudela, & Emmanuel Boss

Abstract: Electromagnetic theory predicts spectral dependencies in extinction efficiency near a narrow absorption band for a particle with an index of refraction close to that of the medium in which it is immersed. These absorption band effects are anticipated in oceanographic beam-attenuation (beam-c) spectra primarily due to the narrow red peak in absorption produced by the phytoplankton photopigment, chlorophyll *a* (Chl *a*). Here we present a method to obtain Chl *a* absorption and size information by analyzing an eigendecomposition of hyperspectral beam-c anomalies measured in oceanic surface waters by an automatic underway system. We find that three principal modes correspond to over 99% of the variance in beam-c anomalies at wavelengths near the Chl *a* red absorption peak. The spectral shapes of the eigenvectors resemble extinction efficiency anomalies attributed to absorption band effects. Projection of the eigenvectors onto the beam-c anomalies produces a time series of amplitude functions with absolute values that are strongly correlated to concurrent Chl *a* absorption line height (a_{LH}) measurements (r values of 0.59 to 0.83) and hence provide a method for estimating Chl *a* absorption. Multiple linear regression of a_{LH} on the amplitude functions enables an independent estimate of a_{LH}, with RMSE of $3.19 \cdot 10^{-3}\,\mathrm{m}^{-1}$ (3.3%) or $\log_{10}$-RMSE of 18.6%, and a raw-scale R^2 value of 0.894. Relationships between the amplitude functions and the beam-c exponential slopes are in agreement with theory relating beam-c to the particle size distribution. Compared with multispectral analysis of beam-c slope, hyperspectral analysis of absorption band effects is anticipated to be relatively insensitive to the addition of non-pigmented particles and to monodispersion.

3.1 Introduction

Bio-optical characterization of the marine environment supports global ocean monitoring by enabling the use of sensing infrastructure on platforms that range from autonomous floats to satellite imagers. In-water sensors lack the coverage and re-sample rates of satellite platforms but enable the direct measurement of light absorption and scattering processes. Multispectral absorption measurements have useful applications for measuring marine ecosystems, for example by enabling the estimation of chlorophyll *a* (Chl *a*) concentration [1, 2]. Multispectral measurements of scattering and attenuation are strongly

correlated [3], and their spectral slopes enable prediction of the particle size distribution (PSD) but cannot resolve phytoplankton Chl *a* content [4]. Recent advances in hyperspectral instrumentation have enabled the decomposition of *in situ* particulate absorption spectra to discern accessory pigmentation relevant to describing phytoplankton community composition [5]. For suspended particles with an index of refraction near that of the medium in which they are immersed, electromagnetic theory predicts wavelength (λ) dependencies in scattering or attenuation for spectral regions adjacent to narrow absorption bands, described as *anomalous dispersion* or *absorption band effects* [6–12]. The λ dependencies related to absorption band effects provide information about phytoplankton size, pigmentation, and refractive index at fine spectral scales measured by current hyperspectral sensors. In this paper, we present observations relevant to absorption band effects using a global and methodologically consistent dataset of particulate hyperspectral beam-c anomalies, and we consider potential applications for characterizing oceanic ecosystems. In particular, we show that these anomalies enable estimation of Chl *a* absorption and provide size information on phytoplankton.

3.1.1 Overview of absorption band effects on beam-attenuation spectra

The modification of the internal oceanic light field occurs through scattering and absorption processes, which are defined as the inherent optical properties (IOPs) of the oceanic medium [13]. The beam-attenuation (beam-c) coefficient $c(\lambda)$ describes the decay or directional change in a beam of collimated light and is the sum of the total absorption $a(\lambda)$ and scattering $b(\lambda)$ coefficients. In practice, $c(\lambda)$ and $a(\lambda)$ are more readily measured, and $b(\lambda)$ is obtained through subtraction. Although often treated as such, $a(\lambda)$ and $b(\lambda)$ are not independent properties. The coefficients may be further specified to represent the particulate (algal and non-algal) contributions by subtracting the properties of the dissolved (filtered) materials from those of the whole water [14], expressed as $a_{\mathrm{p}}(\lambda)$, $b_{\mathrm{p}}(\lambda)$, and $c_{\mathrm{p}}(\lambda)$ for the particulate absorption, scattering, and beam-c coefficients, respectively.

For individual particles within a medium, the ratios of the optical cross-sections to the geometric cross-sections define the absorption $Q_{\mathrm{a}}(\lambda)$, scattering $Q_{\mathrm{b}}(\lambda)$, and extinction $Q_{\mathrm{ext}}(\lambda)$ efficiency factors, which contribute to the bulk IOPs of a water mass, i.e., to $a_{\mathrm{p}}(\lambda)$, $b_{\mathrm{p}}(\lambda)$, and $c_{\mathrm{p}}(\lambda)$, respectively. Anomalous diffraction theory approximates $Q_{\mathrm{ext}}(\lambda)$ for large, non-absorbing, and homogeneous spheres using the phase-lag parameter, ρ, defined as fol-

lows [6]:

$$\rho = 2\,x\,(n - 1)\ , \tag{3.1}$$

in which n is the real index of refraction and x is the ratio of the particle's circumference to the wavelength of light in the medium, ranging from 0 to ∞. Within this paper, the real index of refraction n is defined relative to seawater, and $(n - 1)$ is assumed to be positive. The spectral dependency in x provides a theoretical basis for estimating the PSD from the spectral slope of $c_{\mathrm{p}}(\lambda)$ that is sensitive to pigmented and non-pigmented constituents [15].

For absorbing particles, including phytoplankton, the imaginary component n' of the complex index of refraction m (defined as $m = n + i\,n'$) corresponds to a particle's pigmentation [16] and is included in the anomalous diffraction approximation of the optical efficiency factors [6]. In the case of phytoplankton cells, Chl *a* and various accessory pigments elevate $Q_{\mathrm{a}}(\lambda)$ and suppress $Q_{\mathrm{b}}(\lambda)$ across a relatively broad range of blue wavelengths [17]. At red wavelengths, a special situation arises from the specific absorption spectra of Chl *a*, which produces a narrow red absorption band. Changes to n and n' in the vicinity of the absorption band (nominally centered at 676 nm) modify $Q_{\mathrm{ext}}(\lambda)$ in a way that depends on the value of ρ. Panel **(a)** of Fig. 3.1 illustrates characteristic $Q_{\mathrm{ext}}(\lambda)$ spectra in the vicinity of an absorption band using various ρ values and is based on the anomalous diffraction approximation of [6].

The predicted changes in $Q_{\mathrm{ext}}(\lambda)$ are described as a function of ρ as follows: for lower ρ values, $Q_{\mathrm{ext}}(\lambda)$ is elevated at the absorption band, e.g., resembling an increase in $Q_{\mathrm{a}}(\lambda)$; for higher ρ values, $Q_{\mathrm{ext}}(\lambda)$ is reduced at the absorption band; and for moderate ρ values, an anomalous dispersion curve emerges with $Q_{\mathrm{ext}}(\lambda)$ reduced at shorter wavelengths and elevated at longer wavelengths, relative to the center of the absorption band. We use the more general term *absorption band effects*, following [11], to describe the spectral features in $Q_{\mathrm{ext}}(\lambda)$ or $c_{\mathrm{p}}(\lambda)$ that are observed near absorption bands.

3.1.2 Relevance of absorption band effects to phytoplankton composition

The approximation that the real index of refraction is near that of seawater is valid for many types of phytoplankton, although natural variability exists due to differences in cellular composition [12]. For example, calcification generally corresponds to higher refractivity of coccolithophores [18]. Cell size is relevant to the phase lag parameterization in such a way that, for constant cellular composition, smaller cells are associated with lower

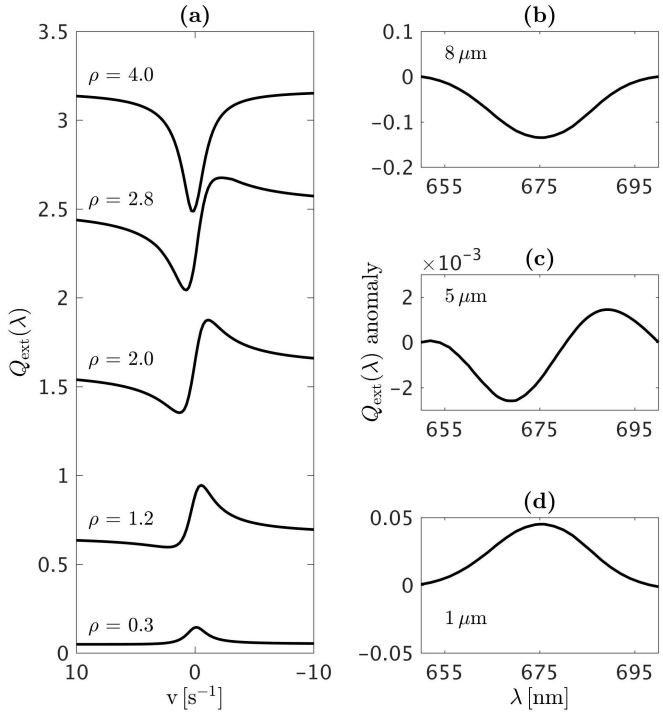

Figure 3.1: ***(a)*** *Anomalous diffraction approximation for $Q_{\mathrm{ext}}(\lambda)$ at a narrow absorption band as a function of light frequency for various ρ values, recreated from van de Hulst, 1957 using a lookup table; and* ***(b-d)*** *illustrative examples of $Q_{\mathrm{ext}}(\lambda)$ anomalies for various small-sized phytoplankton (diameters 8, 5, and 1 μm, respectively), with a fixed n of 1.0344 and spectral n' with a maximum value of 0.0024 at the Chl a red absorption peak. The sizes presented in panels* ***(b-d)*** *are sensitive to the selection of real and imaginary refractive index. For example, Aas, 1996 illustrates an anomalous dispersion curve for $Q_{\mathrm{b}}(\lambda)$ using a 1 μm absorbing sphere and spectral dependencies in both n and n'.*

ρ values and larger cells with higher ρ values. The combined effects of cell size and index of refraction therefore yield the result that $Q_{\mathrm{ext}}(\lambda)$ anomalies in the spectral vicinity of a narrow absorption band can be positive for small phytoplankton with n near seawater, or negative for larger or more refractive phytoplankton. Within a narrow, intermediate range in size and refractivity, predicted $Q_{\mathrm{ext}}(\lambda)$ anomalies resemble an anomalous dispersion curve. Thresholds for ρ have been approximated (e.g., $\rho < 3$ can correspond to an anomalous dispersion curve), but these limits are not particularly useful due to intracellular variability in refractivity, pigmentation, and cellular shape, as well as uncertainties in the actual size distribution when representing ρ for a theoretical mean equivalent particle ($\tilde{\rho}$) [8]. Illustrative $Q_{\mathrm{ext}}(\lambda)$ anomalies for various sizes of phytoplankton are shown in panels

(b-d) of Fig. 3.1 using idealized parameterizations based on [6] and [9].

Considering IOPs of polydisperse systems rather than single particle efficiencies, the $c_{\mathrm{p}}(\lambda)$ and $b_{\mathrm{p}}(\lambda)$ spectra are anticipated to be smoothed by the diversity of refractivity and cell sizes, i.e., polydispersion, in natural marine phytoplankton communities [8]. However, communities dominated by small cell sizes, characteristic of many oligotrophic oceanic environments, are anticipated to produce peaks and anomalous dispersion curves in $c_{\mathrm{p}}(\lambda)$ measurements. The objective of this paper is to advance understanding of how absorption band effects influence $c_{\mathrm{p}}(\lambda)$ spectra at red wavelengths, in particular by partitioning the contributions from polydisperse systems of phytoplankton to enable inference of biomass or community information. We do not focus here on $b_{\mathrm{p}}(\lambda)$ or on its backward component, $b_{\mathrm{bp}}(\lambda)$, as hyperspectral $c_{\mathrm{p}}(\lambda)$ is more routinely measured *in situ*. However, we note that $b_{\mathrm{bp}}(\lambda)$ is most strongly affected by absorption band effects [11] and is likewise most relevant to a remote sensing perspective [19].

3.2 Methods

3.2.1 Description of bio-optical dataset

Bio-optical oceanographic data was collected during the Tara Oceans Expedition, in which an aluminum-hulled schooner sailed through the Pacific, Atlantic, and Indian ocean basins, as well as the Caribbean, Mediterranean, and Red seas, while continually sampling surface waters using a flow-through system [20]. Briefly, seawater was routed to a WET Labs ac-s meter, which measures $c(\lambda)$ and $a(\lambda)$ by passing water through separate columns illuminated by collimated and diffuse light sources, respectively. The ac-s instrument is hyperspectral and measures approximately 80 wavelengths spanning 400 to 730 nm.

The relative calibration of the ac-s meter in the underway configuration was achieved during the Tara Oceans Expedition by periodically filtering the flow-through samples (every 30 or 60 min) and subtracting the measurements of the dissolved samples from the total, as described in [21]. The difference corresponds to the particulate contributions, $c_{\mathrm{p}}(\lambda)$ and $a_{\mathrm{p}}(\lambda)$, with scattering corrections performed following [22]. The ac-s instrument has a non-negligible acceptance angle of 0.93 degrees for beam-c measurements, which decreases the sensitivity to scattering by large particles. This can introduce a bias by decreasing the contribution of large particles in measured $c_{\mathrm{p}}(\lambda)$ relative to theoretical $c_{\mathrm{p}}(\lambda)$ [23].

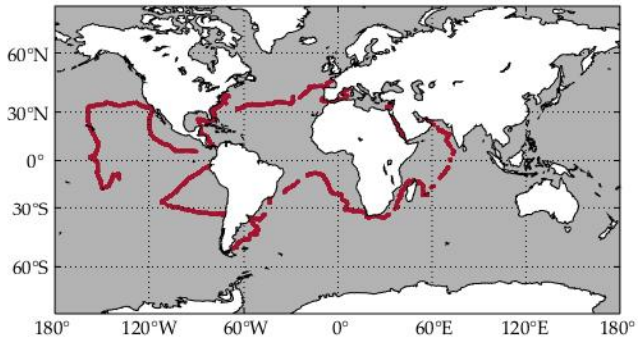

Figure 3.2: *Sampling locations for the bio-optical measurements obtained from the Tara Oceans Expedition archive in SeaBASS.*

We accessed Tara Oceans Expedition $c_{\mathrm{p}}(\lambda)$ and $a_{\mathrm{p}}(\lambda)$ measurements at one-minute temporal resolution through the NASA SeaWiFS Bio-Optical Archive and Storage System (SeaBASS; `seabass.gsfc.nasa.gov`) data repository [24–26]. The global sampling locations of the concurrent $a_{\mathrm{p}}(\lambda)$ and $c_{\mathrm{p}}(\lambda)$ measurements used in this study (245,277 in total) are shown in Fig. 3.2.

The exact spectral locations of the ac-s wavebands differ, both between instruments and within the same instrument following factory recalibration; in response to this, we linearly interpolated all $c_{\mathrm{p}}(\lambda)$ spectra onto a consistent waveband set. Following [5], we removed filter artifacts that resulted from default smoothing of the ac-s spectra across $c_{\mathrm{p}}(\lambda)$ channels. We modeled broad spectral dependencies in $c_{\mathrm{p}}(\lambda)$ that result from the relationship between particle size and wavelength as a power-law function, following [4], and subtracted the power-law model from the $c_{\mathrm{p}}(\lambda)$ spectra to obtain residuals, denoted $c'_{\mathrm{p}}(\lambda)$, as follows:

$$c'_{\mathrm{p}}(\lambda) \;=\; c_{\mathrm{p}}(\lambda) \;-\; \left[A \cdot \lambda^{-\gamma}\right] \,. \tag{3.2}$$

The PSD of natural oceanic particles is also represented by a power-law function [14], with exponential slope, ξ, that may be predicted from the $c_{\mathrm{p}}(\lambda)$ exponential slope, γ [4]. This relationship is valid within an appropriate range of particle sizes [27], with the result that increases in γ have been shown to correspond to increases in ξ, e.g., steeper (more negative) $c_{\mathrm{p}}(\lambda)$ slopes indicate greater relative proportions of smaller particles.

We obtained an independent proxy for Chl *a* absorption from coincident $a_{\mathrm{p}}(\lambda)$ measurements based on the height of the red peak in the Chl *a*-specific absorption spectrum.

Briefly, we linearly interpolated an absorption baseline between shoulder wavebands and subtracted this from the region of maximum Chl a red absorption, as follows:

$$a_{\mathrm{LH}} \;=\; a_{\mathrm{p}}(\lambda_2) - a_{\mathrm{p}}(\lambda_1) - \left(\frac{\lambda_2 - \lambda_1}{\lambda_3 - \lambda_1}\right)\left[a_{\mathrm{p}}(\lambda_3) - a_{\mathrm{p}}(\lambda_1)\right] \,, \tag{3.3}$$

in which λ_2 is the red peak (nominally 676 nm), and λ_1 and λ_3 are adjacent legacy wavelengths (e.g., 650 and 715 nm). Estimation of Chl a from a_{LH} has been shown to be relatively insensitive to pigment packaging effects [28, 29] compared with blue wavelength algorithms, as well as to changes in phytoplankton physiology and the ambient light field compared with Chl a fluorescence-based techniques [30] (see [31] for field data).

3.2.2 Reduction of $c'_{\mathrm{p}}(\lambda)$ spectra using empirical orthogonal functions

We reduced the dimensionality of the beam-c residuals dataset using an eigenanalysis of the $c'_{\mathrm{p}}(\lambda)$ dataset within a narrow spectral subset (13 wavebands spanning approximately 50 nm) centered on the Chl a red absorption peak (nominally 676 nm). Briefly, we performed an eigendecomposition of the $c'_{\mathrm{p}}(\lambda)$ covariance matrix of the form:

$$\boldsymbol{C}\,\boldsymbol{\psi} \;=\; \boldsymbol{\Lambda}\,\boldsymbol{\psi}\,, \tag{3.4}$$

in which $\boldsymbol{C}$ is the covariance matrix of the $c'_{\mathrm{p}}(\lambda)$ dataset and $\boldsymbol{\psi}$ is the eigenfunction matrix, with columns $\boldsymbol{\psi}_i(\lambda)$ describing modes of variability across the wavelength domain of the $c'_{\mathrm{p}}(\lambda)$ data. In Sect. 3.3.2 below we compare the spectral shape of the $\boldsymbol{\psi}_i(\lambda)$ eigenvectors with the $Q_{\mathrm{ext}}(\lambda)$ anomalies predicted for absorption band effects. The diagonal matrix $\boldsymbol{\Lambda}$ contains eigenvalues relating scalar information for each eigenvector, with the sum of the eigenvalues equal to the sum of the wavelength-specific variances in the $c'_{\mathrm{p}}(\lambda)$ dataset (the diagonal elements of the covariance matrix $\boldsymbol{C}$), expressed as follows:

$$\sum_{i=1}^{k=13} \Lambda_{i,i} \;=\; \sum_{i=1}^{k=13} \sigma^2_{(\lambda_i,\lambda_i)}\,. \tag{3.5}$$

Comparison of the eigenvalues $\Lambda_{i,i}$ enables consideration of the variance captured by each eigenfunction. We reduced the spectra of the $c'_{\mathrm{p}}(\lambda)$ dataset to scalar amplitudes by project-

ing the $c_{\mathrm{p}}'(\lambda)$ data onto the eigenfunction matrix, as follows:

$$\boldsymbol{S} \;=\; c_{\mathrm{p}}' \, \boldsymbol{\psi} \,, \tag{3.6}$$

in which the 13 columns of the $\boldsymbol{S}$ matrix are time series of amplitude functions that quantify the stretching and compressing necessary to represent the $c_{\mathrm{p}}'(\lambda)$ dataset in the new coordinates defined by the eigenvector basis functions. Positive and negative signs in $\boldsymbol{S}_i$ indicate corresponding phase shifts in $\boldsymbol{\psi}_i(\lambda)$, which may resemble the different idealized shapes of $Q_{\mathrm{ext}}(\lambda)$ anomalies (i.e., maxima or minima) shown in Fig. 3.1.

We assessed the relationships between the components of the $\boldsymbol{S}$ time series and the a_{LH} and γ data products through univariate and multivariate regression, with the positive and negative phases of each $\boldsymbol{S}$ component i treated as separate $\boldsymbol{S}$ matrix predictors, $\boldsymbol{P}_i^{(+)}$ and $\boldsymbol{P}_i^{(-)}$, defined as follows:

$$\begin{aligned} \boldsymbol{P}_i^{(+)} \;&=\; \boldsymbol{S}_i \,; \quad & \boldsymbol{P}_i^{(-)} \;&=\; 0 \qquad & (\text{if } \boldsymbol{S}_i \;>\; 0) \,; \\ \boldsymbol{P}_i^{(+)} \;&=\; 0 \,; \quad & \boldsymbol{P}_i^{(-)} \;&=\; |\boldsymbol{S}_i| \qquad & (\text{if } \boldsymbol{S}_i \;\leq\; 0) \,. \end{aligned} \tag{3.7}$$

We evaluated regressions using root mean squared error of prediction (RMSE) which, when represented as a percentage for raw-scale values, was normalized by the range in the a_{LH} dataset. We analyzed the $\boldsymbol{S}$ matrix using a thinned dataset to reduce auto-correlation related to the relatively long spatial decorrelation scales of oceanic waters compared with the average speed of the Tara vessel. We sub-sampled the dataset across approximate length scales of 11.1 km and 33.3 km for coastal (within 200 km from the shore) and oceanic (over 200 km from the shore) water masses, respectively.

3.3 Results

3.3.1 Effects of autocorrelation

Autocorrelation is a persistently challenging topic in oceanography because spatial and temporal decorrelation scales are variable among regions and seasons and because large differences between oceanic provinces (e.g., coastal zones, oligotrophic gyres, upwelling regions) often overshadow smaller-scale variability within each region. Our spatial-scale thinning of the Tara $c_{\mathrm{p}}'(\lambda)$ dataset decreased the total number of observations by over

98%. Before thinning, the coastal zones were measured less frequently than oceanic waters because the Tara vessel generally maintained a trans-oceanic course. Our spatial thinning, which used different length scales for coastal and oceanic measurements, increased the relative contribution of coastal waters (i.e., within 200 km of shore) from 27% to 50% of the dataset. Despite the changes in size and regional representation due to thinning, the spectral shapes and ordering of the eigenfunctions were not significantly altered by sub-sampling, and the eigenvalues corresponding to the first three modes of the sub-sampled dataset were each within 2% of those derived from the full (not sub-sampled) dataset.

3.3.2 Interpretation of the eigenfunctions

The bulk optical properties of a water mass integrate contributions from various constituents, as well as from the medium. An idealized equation relating $c_{\mathrm{p}}(\lambda)$ to the size-dependent $Q_{\mathrm{ext}}(\lambda, x)$ contributions is of the form [6]:

$$c_{\mathrm{p}}(\lambda) \;=\; \int_0^{\infty} Q_{\mathrm{ext}}(\lambda, x)\, N(x)\, \pi\, x^2\, dx\,, \tag{3.8}$$

in which $N(x)$ is the number of x-sized particles. The $c_{\mathrm{p}}(\lambda)$ spectra is a bulk property that arises from the addition of the underlying particle properties, making the spectral dependencies in $Q_{\mathrm{ext}}(\lambda, x)$ directly related to the spectral shape of $c_{\mathrm{p}}(\lambda)$. Our eigenanalysis of the $c'_{\mathrm{p}}(\lambda)$ covariance matrix quantified the primary modes of variability in the $c'_{\mathrm{p}}(\lambda)$ spectra, which are shown in Fig. 3.3. Although both positive and negative phases of the eigenfunctions may resemble the spectral anomalies associated with absorption band effects, only one representative phase is shown for each eigenvector.

The first three components of the eigenanalysis captured more than 99% of the variance, with the first, second, and third eigenvectors comprising 61.1%, 37.0%, and 1.2% of the total, respectively. We did not examine the additional components, which individually corresponded to 0.25% or less of the variance. We propose the following interpretations of the spectral shapes of the relevant positive (+) and negative (−) eigenfunctions, with the local maxima for Chl *a* absorption at red wavelengths defined as the *absorption band* for brevity:

- Eigenfunction 1 (+): a negative anomaly shifted ∼10 nm shorter than the absorption band.

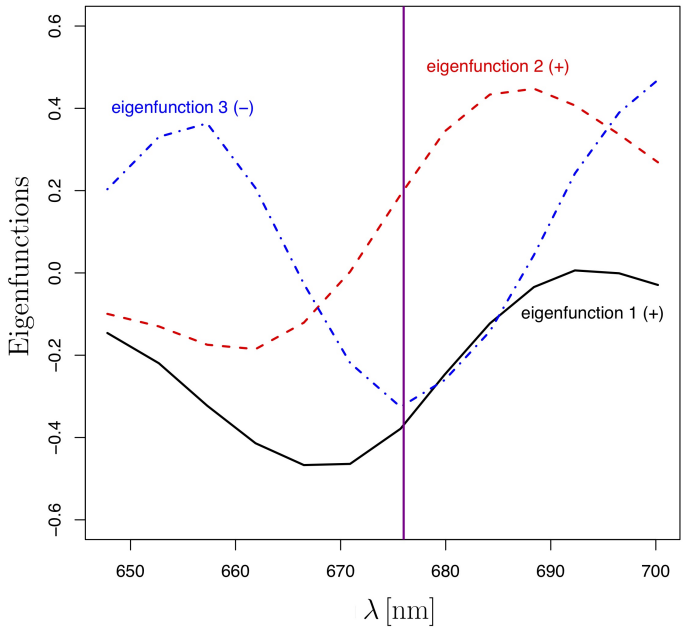

Figure 3.3: *Principal components of the* $c'_{\mathrm{p}}(\lambda)$ *covariance matrix, with the nominal location of the Chl a red absorption maximum indicated by the vertical line.*

- Eigenfunction 1 (−): a positive anomaly shifted ∼10 nm shorter than the absorption band.
- Eigenfunction 2 (+): a positive anomaly shifted (∼12 nm) longer than the absorption band, a negative anomaly shifted (∼15 nm) shorter than the absorption band, and an inflection near the absorption band.
- Eigenfunction 2 (−): a negative anomaly shifted (∼12 nm) longer than the absorption band, a positive anomaly shifted (∼15 nm) shorter than the absorption band, and an inflection near the absorption band.
- Eigenfunction 3 (+): a positive anomaly centered at the absorption band.
- Eigenfunction 3 (−): a negative anomaly centered at the absorption band.

Considering the theoretical parameterizations governing absorption band effects on $Q_{\mathrm{ext}}(\lambda)$, the eigenvector basis functions may also be interpreted through a transition in ρ values, as follows: the lowest ρ (an increase at the absorption band) corresponds to

eigenfunctions 1 (−) and 3 (+); intermediate ρ (the anomalous dispersion curve) corresponds to eigenfunction 2 (+); and the highest ρ (a decrease at the absorption band) corresponds to eigenfunctions 1 (+) and 3 (−). The spectral shape of eigenfunction 2 (−) was not in agreement with the $Q_{\text{ext}}(\lambda)$ anomalies predicted for absorption band effects, as shown in Fig. 3.1, and we therefore regard it as potentially non-physical within the scope of this work. This is supported by the low expression of eigenfunction 2 (−) in the $\boldsymbol{S}$ matrix time series, with less than 1% of the $\boldsymbol{S}_2$ amplitudes negative. The same is true, but to a lesser extent, for eigenfunction 1 (−), for which the maxima was less spectrally separate from the absorption band compared with eigenfunction 2 (−). Similarly, less than 11% of the $\boldsymbol{S}_1$ amplitudes were negative.

3.3.3 Interpretation of the predictors

The relationships between the $\boldsymbol{S}$ matrix predictors and the a_{LH} dataset are shown individually in Fig. 3.4, with the negative amplitudes from the time series for component 2 (26 out of 2827 observations) omitted. Each $\boldsymbol{S}$ matrix predictor had highly significant positive correlation with a_{LH} (posterior probability < 0.001 of non-positive correlation in the population to which our results generalize), except for the negative amplitudes in the time series for component 2, which indicated a positive but non-significant relationship.

We found that, for each eigenanalysis component, slope coefficients derived by linear univariate regression of a_{LH} and the $\boldsymbol{S}$ matrix predictors were greater for the phases that corresponded with higher ρ values based on the interpretations described in Sect. 3.3.2. For example, the negative phase of component 1, which corresponds to a negative anomaly near the absorption band (higher ρ), produced a significantly steeper slope than the positive phase of component 1, which corresponds with a positive anomaly near the absorption band (lower ρ). Similarly, the positive phase of component 2 (anomalous dispersion curve; intermediate ρ) produced a significant positive slope, while the slope of the negative phase (non-physical) was flattened and not significant. The negative phase of component 3 (absorption band minima; higher ρ) produced a greater slope than the positive phase (absorption band maxima; lower ρ).

We evaluated the accuracy of the combined $\boldsymbol{S}$ matrix predictors for estimating a_{LH} using multiple linear regression over 10,000 cross-validation replications. In each replication, we randomly partitioned the dataset into modeling and validation subsets using

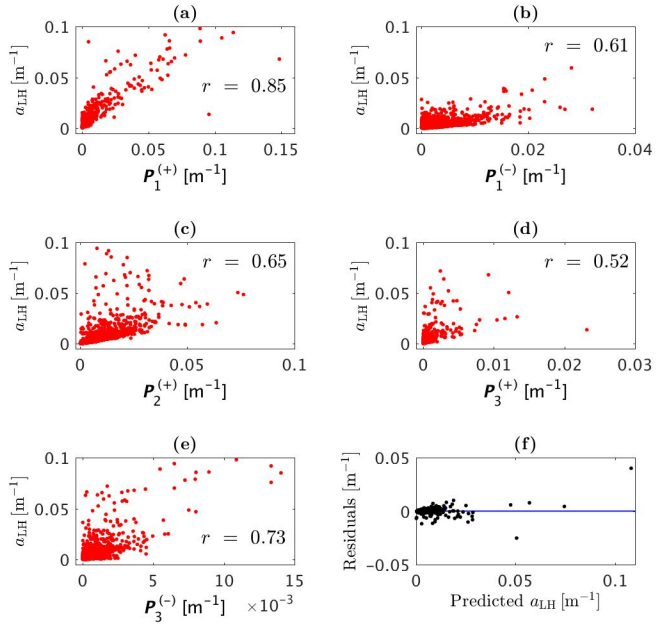

Figure 3.4: *Relationships between* a_{LH} *and the* $\boldsymbol{S}$ *matrix predictors* ***(a)*** $\boldsymbol{P}_1{}^{(+)}$*;* ***(b)*** $\boldsymbol{P}_1{}^{(-)}$*;* ***(c)*** $\boldsymbol{P}_2{}^{(+)}$*;* ***(d)*** $\boldsymbol{P}_3{}^{(+)}$*; and* ***(e)*** $\boldsymbol{P}_3{}^{(-)}$*.* ***(f)*** *Residuals against predicted* a_{LH} *values from a multivariate linear regression of the* $\boldsymbol{S}$ *matrix predictors.*

an $80\% - 20\%$ split, which corresponded to 2262 and 565 data points, respectively. The repeated cross-validation produced a median RMSE of $3.19 \cdot 10^{-3} \pm 0.55 \cdot 10^{-3}\,\mathrm{m}^{-1}$, corresponding to 3.3% of the range in a_{LH}. RMSE derived from $\log_{10}$-transformed variables ($\log_{10}$-RMSE) indicated uncertainty of $18.6\% \pm 1.7\%$. The median R^2 value in the validation datasets was 0.894. The log-log scatter plot of measured and predicted a_{LH} values in the median validation subset is shown in Fig. 3.5.

The relationships for the $\boldsymbol{S}$ matrix predictors and the exponential slopes of the $c_{\mathrm{p}}(\lambda)$ dataset, γ, were evaluated using the non-parametric scatterplot smoother `lowess` [32], shown in Fig. 3.6 with $\log_{10}$ horizontal scales. As in Fig. 3.4, we omitted the predictor corresponding to the negative phase of $\boldsymbol{S}_2$ due to the low number of observations and lack of physical interpretability. The non-parametric smoothers indicate that positive spikes in component 1 and negative spikes in component 3 of the $\boldsymbol{S}$ matrix correspond with decreasing γ, and negative spikes in component 1 correspond with increasing γ. Based on our Sect. 3.3.2 interpretation of the eigenvector spectra, the relationships in Fig. 3.6 could also be expressed in terms of ρ, with panels **(a)**, **(b)**, and **(e)** indicating a negative association

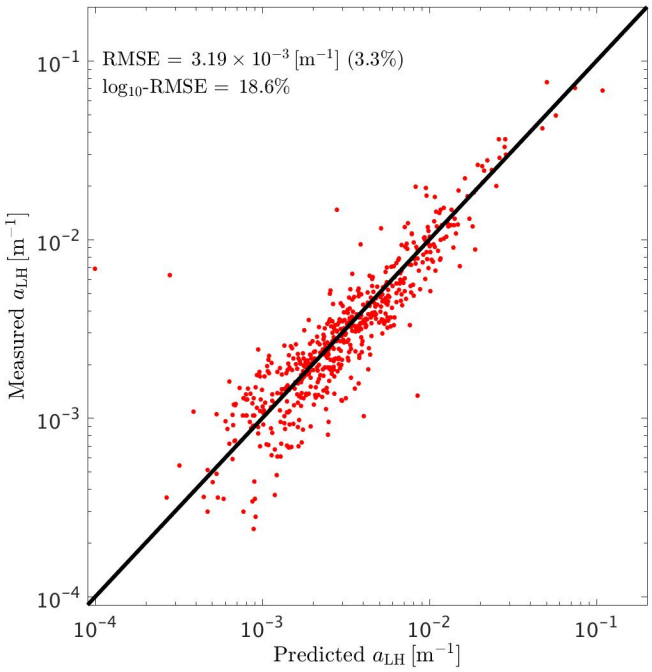

Figure 3.5: *Median validation scatterplot relating measured (vertical) and predicted (horizontal)* a_{LH} *values from the* $\boldsymbol{S}$ *matrix predictors, with both axes on the* $\log_{10}$ *scale.*

between γ and ρ. Ignoring variability in m (i.e., $\rho \propto x$), the results are in agreement with theory relating γ to the PSD. For example, an increase in the concentration of larger cells corresponds to lower γ and greater expression of the high-ρ amplitude functions, $\boldsymbol{P}_1^{(+)}$ and $\boldsymbol{P}_3^{(-)}$. However, the theoretical relationship between PSD and γ corresponds to the full particle population, while the relationship with *absorption band effects* corresponds to the pigmented particle fraction.

3.4 Discussion

3.4.1 Relevance of absorption band effects to phytoplankton dynamics

The development of hyperspectral IOP sensors and their deployment in a continuous, underway configuration provided us with a large surface ocean IOP dataset with 13 wavebands within about 25 nm of the Chl *a* red absorption peak. Our eigendecomposition found that three principal modes captured more than 99% of the variance in the $c_{\mathrm{p}}'(\lambda)$ dataset.

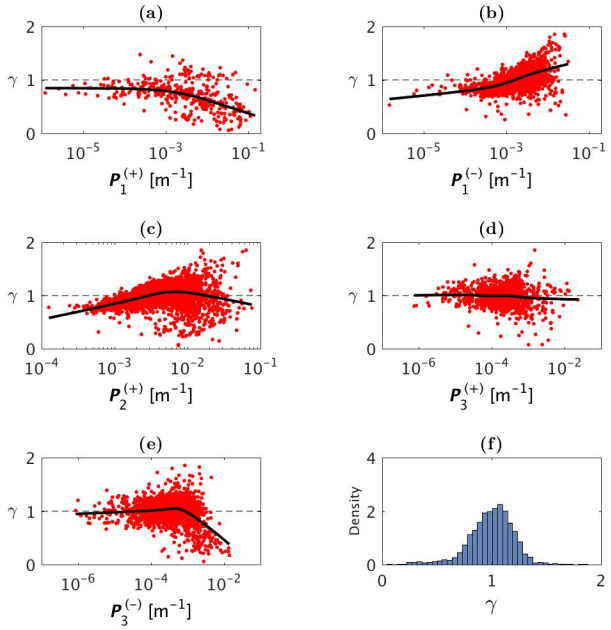

Figure 3.6: *Relationships between* γ *and the* $\boldsymbol{S}$ *matrix predictors* **(a)** $\boldsymbol{P}_1{}^{(+)}$; **(b)** $\boldsymbol{P}_1{}^{(-)}$; **(c)** $\boldsymbol{P}_2{}^{(+)}$; **(d)** $\boldsymbol{P}_3{}^{(+)}$; *and* **(e)** $\boldsymbol{P}_3{}^{(-)}$, *with locally-weighted scatterplot smoothing (*`lowess`*) functions overlaid in solid black; the horizontal scales are* $\log_{10}$. **(f)** *Histogram estimate of the probability density function of the* γ *values derived from the* $c_\mathrm{p}(\lambda)$ *dataset.*

The shapes of these principal modes resembled the $Q_\mathrm{ext}(\lambda)$ spectral anomalies predicted by electromagnetic theory to arise in the vicinity of narrow absorption bands.

The second principal mode, which captured about one third of the $c_\mathrm{p}'(\lambda)$ variance, most closely resembled the anomalous dispersion curve illustrated in panel **(c)** of Fig. 3.1. Absorption band effects are not anticipated to produce spectral shapes resembling the reflection of this curve, and likewise the negative phase of the second principal component was rarely present in the $\boldsymbol{S}$ matrix time series (less than 1% of the $\boldsymbol{S}_2$ values were negative). The variance captured by the second principal mode is interesting because previous analyses have indicated that the anomalous dispersion result is only relevant to very small phytoplankton, e.g., [12]. Although caution is warranted in relating ρ to phytoplankton size because of uncertainty in cellular shape, refractivity, and pigmentation [8], the results of our eigenanalysis interpretation are in general agreement with the ecological perspectives that picoplankton (diameters generally less than 1 or 2 μm) are ubiquitous across oceanic gyre ecosystems [33] and that increases in biomass often result from the addition of larger

cells over a more stable background of smaller cells [34].

The eigenvector projections onto the $c'_{\mathrm{p}}(\lambda)$ dataset indicate that the separate phases of the $\boldsymbol{S}$ matrix predictors, i.e., $\boldsymbol{P}_i^{(+)}$ and $\boldsymbol{P}_i^{(-)}$ (Eq. 3.7), are each significantly and positively correlated with a_{LH}, with the exception of $\boldsymbol{P}_2^{(-)}$, which corresponded to an eigenvector spectral shape that was not predicted from absorption band effects. Predictors associated with higher ρ domains (based on interpretations in Sect. 3.3.2) indicate steeper relationships with a_{LH} compared to those associated with lower ρ domains. The results suggest that additions of larger or more refractive phytoplankton correspond to greater increases in a_{LH} than additions of smaller or less refractive phytoplankton. In general, this perspective is in agreement with the phytoplankton ecological paradigm that larger phytoplankton predominantly occupy more productive water masses, while smaller phytoplankton are ubiquitous across less productive, more oligotrophic regimes [35, 36]. Considering the relationships between γ and the $\boldsymbol{S}$ matrix predictors, our findings are consistent with electromagnetic theory, although we could not separate the effects of refractivity and size with the approach taken here.

Our theoretical description of absorption band effects is based on the treatment of phytoplankton as simple, homogeneous spheres. Phytoplankton are often non-spherical, but the $Q_{\mathrm{ext}}(\lambda)$ approximations discussed here can be generalized to describe the average efficiency factors for non-spherical particles, provided that the particles are randomly oriented within the medium [37]. Phytoplankton also contain various internal cellular structures that produce large variability in the refractive index, particularly for structures bound by lipid membranes, containing gases, or housing pigmented molecules, as well as layering by plates and frustules that encase coccolithophore and diatom cells, respectively. As a conceptualization, our theoretical description provides a basis for interpreting the eigenanalysis and is consistent with previous approaches that have advanced fundamental concepts in marine optics using simple representations of phytoplankton cells as homogeneous [38] or layered [39] spheres.

Despite the terminology *anomalous dispersion*, the theory described in this paper is in agreement with basic principles, e.g., ρ decreases with increasing wavelength (all else being equal), consistent with normal dispersion. However, key differences between absorption band effects and γ analysis are relevant to measuring marine systems. First, while $c_{\mathrm{p}}(\lambda)$ slope methods are sensitive to the sizes of pigmented and non-pigmented particles [4], absorption band effects are only anticipated for pigmented particles, as spectral dependencies

in particle n' values arise primarily through absorption by photopigments. Second, the relationship between the $c_{\mathrm{p}}(\lambda)$ and PSD slope is robust in polydisperse environments within an appropriate PSD size range but is challenged in monodisperse systems [27]. Absorption band effects are confounded by polydispersion [8]; thus hyperspectral resolution is needed for decomposition of overlapping signals.

3.4.2 Potential for application of absorption band effects

The methods presented here enabled us to assess spectral variability in $c_{\mathrm{p}}(\lambda)$ near a photopigment absorption band without requiring *a priori* decisions about the shapes of the extracted signals. Optimization of the spectral shapes of extracted signals, e.g., using theoretical response functions, may improve estimates of a_{LH} from $c_{\mathrm{p}}(\lambda)$ datasets. We compared our analysis with an eigendecomposition using $c'_{\mathrm{p}}(\lambda)$ spectra that were peak-normalized, and we found that the accuracy for estimating Chl *a* absorption decreased, although the relationships between the $\boldsymbol{S}$ matrix predictors and γ were similar. Advancing the capability to estimate Chl *a* absorption from beam-c would provide useful redundancies in instances in which both beam-c and absorption are measured concurrently (e.g., by an ac-s) and would be useful in turbid waterbodies, where absorption meters are more prone to fouling than beam-c meters. In general, beam-c is more easily measured than absorption.

PSD can be derived from the $c_{\mathrm{p}}(\lambda)$ exponential slope using legacy multispectral instrumentation because the approach only requires measurement of two wavelengths [27]. Decomposition methods require greater spectral resolution because the targeted signals are often more complex and spectrally overlapping. Because the signals associated with absorption band effects are most apparent within a narrow range in λ, even the hyperspectral ac-s instrument only provided up to 13 relevant wavebands in this study. Improvement in spectral resolution generally coincides with a trade-off in radiometric accuracy, which is problematic for decomposition of low amplitude signals. For example, the signals associated with absorption band effects are relatively low compared to the ranges in $c_{\mathrm{p}}(\lambda)$ that result from variability in the refractivity and size distributions of pigmented or non-pigmented particles in natural marine systems.

Low signal difficulties may be partially mitigated by the measurement of backscattering, rather than beam-attenuation or total scattering, because the backscattering spectra are more sensitive to absorption band effects [11]. Recently, a commercial hyperspectral

$b_{\mathrm{b}}(\lambda)$ instrument has been been developed, which could potentially be applied to advance this topic [40]. For water bodies dominated by large refractive phytoplankton, $c_{\mathrm{p}}(\lambda)$ and $b_{\mathrm{bp}}(\lambda)$ are anticipated to produce local minima at the Chl *a* red absorption band, and estimation of phytoplankton biomass could be performed using a line height approach to reduce spectral resolution requirements. Research to develop multispectral backscatter instruments that target absorption band effects in order to predict Chl *a* is underway [41]. In the work summarized here, we did not investigate the importance of accessory pigmentation on absorption band effects, because most phytoplankton photopigments are not active in the vicinity of the Chl *a* red peak. Chlorophyll *b*, which can form an absorption plateau with Chl *a* at high concentrations, is one notable exception.

3.5 Conclusions

We reduced the dimensionality of a surface ocean hyperspectral beam-c dataset with minimal loss of information by identifying three principal modes, which were similar in shape to the spectra of theoretical particle extinction anomalies associated with absorption band effects. The results indicate that at wavelengths adjacent to the Chl *a* red absorption peak, absorption band effects are a primary source of variability in beam-c spectra, due in part to the absence of other strong spectral dependencies within the region. Challenges to our approach include low signal and high spectral requirements, polydispersion of natural marine ecosystems, and variability in the pigmentation, refractivity, and shape of marine phytoplankton.

The positive and negative amplitudes of the major principal components we found provided useful predictors for a_{LH} in our study, indicating that analysis of absorption band effects in $c_{\mathrm{p}}(\lambda)$ spectra can enable an alternative estimate of Chl *a* absorption. The relationships between the principal components and the $c_{\mathrm{p}}(\lambda)$ exponential slopes are in agreement with electromagnetic theory and suggest that useful size parameters could be estimated from decomposition of hyperspectral beam-c measurements. Future improvements in measuring the index of refraction of phytoplankton will be useful in advancing these topics, and the upper and lower limits for interpreting absorption band effects should be explored. Culture work in particular could help better elucidate the bio-optical relationships described by this analysis.

Finally, in this work we considered absorption band effects using an observational

approach that was made possible by advances in hyperspectral IOP instrumentation. Basic, rather than applied, scientific research was first necessary to develop an understanding of the optics of narrow absorption bands, and our work relies on advances achieved through electromagnetic theory, e.g., [7–11]. Our principal conclusion is that, with recent advances in IOP instrumentation, current hyperspectral beam-c datasets enable accurate estimation of Chl *a* absorption based on information captured from absorption band effects.

Acknowledgments

We are grateful to Kenneth Voss (University of Miami), Collin Roesler (Bowdoin College), Curtis Mobley (Sequoia Scientific), and Ivona Cetinić (NASA GSFC) for useful discussions that helped clarify topics related to this work. Numerous student participants in the NASA Ocean Optics summer course also provided valuable feedback for this project.

Funding: Support for H.F.H. was partially provided by NASA C-HARRIER Campaign grant Number NNX17AK89G. NASA grant 80NSSC19K0812 supported the Ocean Optics class where this study began.

References

[1] R. F. Davis, C. C. Moore, J. R. V. Zaneveld, and J. M. Napp, "Reducing the effects of fouling on chlorophyll estimates derived from long-term deployments of optical instruments," *Journal of Geophysical Research: Oceans*, vol. 102, no. C3, pp. 5851–5855, 1997.

[2] E. S. Boss, R. Collier, G. Larson, K. Fennel, and W. Pegau, "Measurements of spectral optical properties and their relation to biogeochemical variables and processes in crater lake, crater lake national park, or," *Hydrobiologia*, vol. 574, pp. 149–159, 2007.

[3] G. Dall'Olmo, T. K. Westberry, M. J. Behrenfeld, E. Boss, and W. H. Slade, "Significant contribution of large particles to optical backscattering in the open ocean," *Biogeosciences*, vol. 6, no. 6, p. 947, 2009.

[4] J. C. Kitchen, J. R. V. Zaneveld, and H. Pak, "Effect of particle size distribution

and chlorophyll content on beam attenuation spectra," *Applied Optics*, vol. 21, no. 21, pp. 3913–3918, 1982.

[5] A. Chase, E. Boss, R. Zaneveld, A. Bricaud, H. Claustre, J. Ras, G. Dall'Olmo, and T. K. Westberry, "Decomposition of in situ particulate absorption spectra," *Methods in Oceanography*, vol. 7, pp. 110–124, 2013.

[6] H. Van de Hulst, *Light scattering by small particles*. Wiley, New York, 1957.

[7] A. Morel and A. Bricaud, "Theoretical results concerning light absorption in a discrete medium, and application to specific absorption of phytoplankton," *Deep Sea Research Part A. Oceanographic Research Papers*, vol. 28, no. 11, pp. 1375–1393, 1981.

[8] A. Bricaud, A. Morel, and L. Prieur, "Optical efficiency factors of some phytoplankters," *Limnology and Oceanography*, vol. 28, no. 5, pp. 816–832, 1983.

[9] A. Morel and A. Bricaud, "Inherent optical properties of algal cells including picoplankton: theoretical and experimental results," *Canadian Bulletin of Fisheries and Aquatic Sciences*, vol. 214, pp. 521–559, 1986.

[10] D. Stramski, A. Morel, and A. Bricaud, "Modeling the light attenuation and scattering by spherical phytoplanktonic cells: a retrieval of the bulk refractive index," *Applied Optics*, vol. 27, no. 19, pp. 3954–3956, 1988.

[11] J. R. V. Zaneveld and J. C. Kitchen, "The variation in the inherent optical properties of phytoplankton near an absorption peak as determined by various models of cell structure," *Journal of Geophysical Research: Oceans*, vol. 100, no. C7, pp. 13309–13320, 1995.

[12] E. Aas, "Refractive index of phytoplankton derived from its metabolite composition," *Journal of Plankton Research*, vol. 18, no. 12, pp. 2223–2249, 1996.

[13] R. W. Preisendorfer and J. E. Tyler, "The measurement of light in natural waters; radiometric concepts and optical properties," tech. rep., Scripps Institution Of Oceanography La Jolla, CA Visibility Lab, 1958.

[14] N. G. Jerlov, *Marine Optics*, vol. 14. Elsevier, 1976.

[15] P. Diehl and H. Haardt, "Measurement of the spectral attenuation to support biological-research in a plankton tube experiment," *Oceanologica Acta*, vol. 3, no. 1, pp. 89–96, 1980.

[16] D. Stramski, "Refractive index of planktonic cells as a measure of cellular carbon and chlorophyll a content," *Deep Sea Research Part I: Oceanographic Research Papers*, vol. 46, no. 2, pp. 335–351, 1999.

[17] J. L. Mueller, *The influence of phytoplankton on ocean color spectra*. PhD thesis, Oregon State University, Corvallis, OR, 1973.

[18] A. Bricaud, J. R. V. Zaneveld, and J. C. Kitchen, "Backscattering efficiency of coccolithophorids: use of a three-layered sphere model," in *Ocean Optics XI*, vol. 1750, pp. 27–33, International Society for Optics and Photonics, 1992.

[19] H. R. Gordon, O. B. Brown, R. H. Evans, J. W. Brown, R. C. Smith, K. S. Baker, and D. K. Clark, "A semianalytic radiance model of ocean color," *Journal of Geophysical Research: Atmospheres*, vol. 93, no. D9, pp. 10909–10924, 1988.

[20] E. Boss, M. Picheral, T. Leeuw, A. Chase, E. Karsenti, G. Gorsky, L. Taylor, W. Slade, J. Ras, and H. Claustre, "The characteristics of particulate absorption, scattering and attenuation coefficients in the surface ocean; contribution of the tara oceans expedition," *Methods in Oceanography*, vol. 7, pp. 52–62, 2013.

[21] W. H. Slade, E. Boss, G. Dall'Olmo, M. R. Langner, J. Loftin, M. J. Behrenfeld, C. S. Roesler, and T. K. Westberry, "Underway and moored methods for improving accuracy in measurement of spectral particulate absorption and attenuation," *Journal of Atmospheric and Oceanic Technology*, vol. 27, no. 10, pp. 1733–1746, 2010.

[22] J. R. V. Zaneveld, J. C. Kitchen, and C. C. Moore, "Scattering error correction of reflection-tube absorption meters," in *Ocean Optics XII*, vol. 2258, pp. 44–55, International Society for Optics and Photonics, 1994.

[23] E. Boss, W. H. Slade, M. J. Behrenfeld, and G. Dall'Olmo, "Acceptance angle effects on the beam attenuation in the ocean," *Optics Express*, vol. 17, no. 3, pp. 1535–1550, 2009.

[24] S. B. Hooker, C. R. McClain, J. K. Firestone, T. L. Westphal, E.-N. Yeh, Y. Ge, and E. R. Firestone, "The seawifs bio-optical archive and storage system (seabass), part 1.," Tech. Rep. TM-104566, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, 1994.

[25] P. J. Werdell and S. W. Bailey, "The seawifs bio-optical archive and storage system (seabass): Current architecture and implementation," Tech. Rep. Tech. Memo 2002-211617, NASA Goddard Space Flight Center, Greenbelt, Maryland, 2002.

[26] P. J. Werdell, S. Bailey, G. Fargion, C. Pietras, K. Knobelspiesse, G. Feldman, and C. R. McClain, "Unique data repository facilitates ocean color satellite validation," *Eos, Transactions American Geophysical Union*, vol. 84, no. 38, pp. 377–387, 2003.

[27] E. Boss, M. S. Twardowski, and S. Herring, "Shape of the particulate beam attenuation spectrum and its inversion to obtain the shape of the particulate size distribution," *Applied Optics*, vol. 40, no. 27, pp. 4885–4893, 2001.

[28] L. Duysens, "The flattening of the absorption spectrum of suspensions, as compared to that of solutions," *Biochimica et biophysica acta*, vol. 19, pp. 1–12, 1956.

[29] J. T. O. Kirk, "A theoretical analysis of the contribution of algal cells to the attenuation of light within natural waters i. general treatment of suspensions of pigmented cells," *New Phytologist*, vol. 75, no. 1, pp. 11–20, 1975.

[30] C. S. Roesler and A. H. Barnard, "Optical proxy for phytoplankton biomass in the absence of photophysiology: Rethinking the absorption line height," *Methods in Oceanography*, vol. 7, pp. 79–94, 2013.

[31] R. J. Brewin, G. Dall'Olmo, S. Pardo, V. van Dongen-Vogels, and E. S. Boss, "Underway spectrophotometry along the atlantic meridional transect reveals high performance in satellite chlorophyll retrievals," *Remote Sensing of Environment*, vol. 183, pp. 82–97, 2016.

[32] W. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association*, vol. 74, pp. 829–836, 1979.

[33] G. E. Fogg, "Some comments on picoplankton and its importance in the pelagic ecosystem," *Aquatic Microbial Ecology*, vol. 9, no. 1, pp. 33–39, 1995.

[34] A. M. Ciotti, M. R. Lewis, and J. J. Cullen, "Assessment of the relationships between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient," *Limnology and Oceanography*, vol. 47, no. 2, pp. 404–417, 2002.

[35] R. Margalef, "Life-forms of phytoplankton as survival alternatives in an unstable environment," *Oceanol. Acta*, vol. 1, pp. 493–509, 1978.

[36] A. Bricaud, H. Claustre, J. Ras, and K. Oubelkheir, "Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size structure of algal populations," *Journal of Geophysical Research: Oceans*, vol. 109, no. C11, 2004.

[37] W. Clavano, E. Boss, and L. Karp-Boss, "Inherent optical properties of non-spherical marine-like particles – from theory to observation," *Oceanography and Marine Biology: An Annual Review*, 2007.

[38] A. Bricaud and A. Morel, "Light attenuation and scattering by phytoplanktonic cells: a theoretical modeling," *Applied Optics*, vol. 25, no. 4, pp. 571–580, 1986.

[39] J. C. Kitchen and J. R. V. Zaneveld, "A three-layered sphere model of the optical properties of phytoplankton," *Limnology and Oceanography*, vol. 37, no. 8, pp. 1680–1690, 1992.

[40] O. A. Mikkelsen, W. H. Slade, T. Leeuw, D. R. Dana, and C. Pottsmith, "A new submersible instrument for measuring hyperspectral backscattering in natural waters (hyper-bb)," in *Ocean Sciences Meeting 2020*, AGU, 2020.

[41] S. Graban, S. Goult, R. Sauzede, and G. Dall'Olmo, "Accurate deep-learning estimation of chlorophyll-a concentration from the spectral particulate beam-attenuation coefficient," *Optics Express*, In prep.

Appendix: Comparison of absorption band effects with pigments and satellite imagery

Matchups of Tara measurements and satellite imagery are included to demonstrate that increases in phytoplankton biomass can correspond with increases in the amplitudes of large or small ρ components of the $\boldsymbol{S}$ matrix predictors. Fig. 3.7 depicts MODIS images of Chl a obtained off the Yucatan Peninsula of Mexico (left panels) and off the Western Cape of South Africa (right panels). As the Tara vessel sails into higher Chl a waters off the Yucatan Peninsula, the $\boldsymbol{P}_2^{(+)}$ predictor, which is associated with intermediate size and refractivity, is elevated by about three standard deviations. Meanwhile, the $\boldsymbol{P}_3^{(-)}$ predictor, associated with large size and high refractivity, remains near zero. Pigment analysis from water sampled near the high Chl a region indicates that micro, nano, and pico phytoplankton size fractions are similar in abundance.

When the Tara vessel sails into a phytoplankton bloom off the coast of the Western Cape, both the $\boldsymbol{P}_2^{(+)}$ and $\boldsymbol{P}_3^{(-)}$ predictors are elevated. A water sampling station within the bloom and locally surrounded by high $\boldsymbol{P}_3^{(-)}$ and low $\boldsymbol{P}_2^{(+)}$ measurements indicates that the phytoplankton community is dominated by the micro size fraction.

Similar scenarios, in which elevated satellite Chl a coincide with changes in a subset of the $\boldsymbol{S}$ matrix predictors, were found in the Tara dataset at sites including the Bay of Biscay off of the French coastline and the New York/New Jersey Bight off the Eastern United States. Most of these examples lack coincident pigment measurements to verify that the $\boldsymbol{S}$ matrix predictors are consistent with phytoplankton community shifts. Future data collection of hyperspectral beam-c measurements with coincident pigment analysis would be helpful for advancing the capability to discern phytoplankton community from beam-c spectra, which was explored within this dissertation.

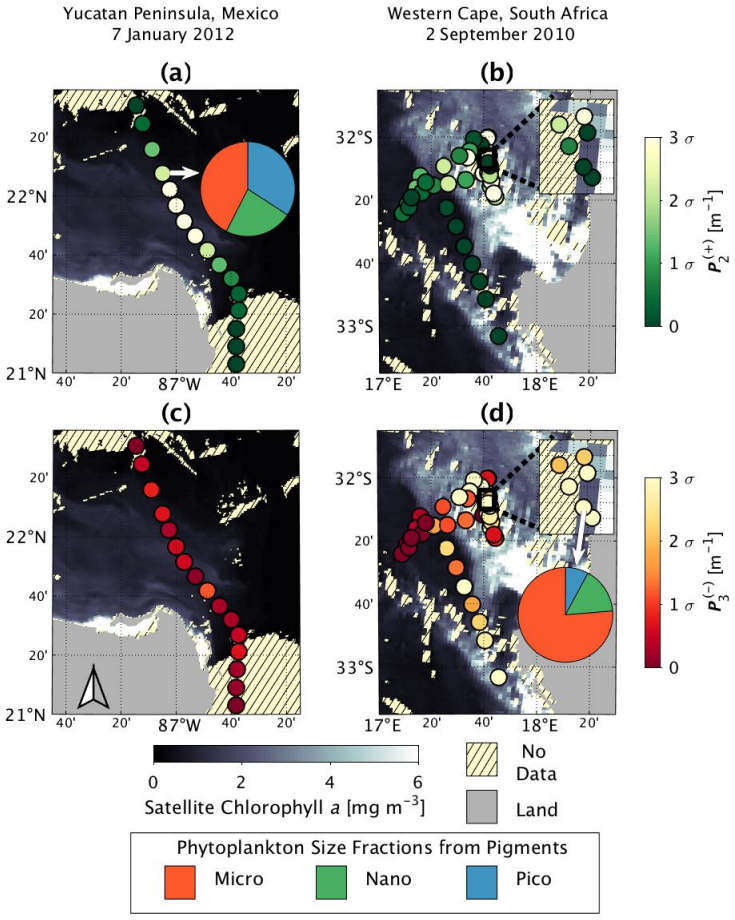

Figure 3.7: ***S*** *matrix predictors derived from the Tara Oceans Expedition data (red and green dots) overlaid onto satellite Chl a imagery (represented by blue shading). The coast of the Yucatan Peninsula (MODIS Aqua) is shown in panels* ***(a)*** *and* ***(c)****, and the Western Cape (MODIS Terra) in panels* ***(a)*** *and* ***(c)****. Pie charts indicate phytoplankton size fractions determined using pigment analysis. The* ***S*** *matrix predictor* $\boldsymbol{P}_2{}^{(+)}$ *is shown in panels* ***(a)*** *and* ***(b)****, and* $\boldsymbol{P}_3{}^{(-)}$ *is shown in panels* ***(c)*** *and* ***(d)****. The satellite matchup shown in panels* ***(b)*** *and* ***(d)*** *precedes the Tara measurements by an average of 5 days due to cloud cover.*

Conclusions

The achievements of legacy satellite and *in situ* activities have greatly increased our ability to detect daily changes in phytoplankton distributions at global scales for the open ocean environment. However, ocean color tools are generally not reliable in optically complex coastal and inland water bodies, where remote sensing capabilities are needed to measure the spatially heterogeneous ecosystems most directly affected by human activities. For example, major human population centers pollute the adjacent coastal and inland water bodies through a complex web of activities including land-use change, eutrophication, diversion or damming of natural waterways, release of persistent chemical compounds, and poor waste management. These practices manifest as the proliferation of toxins in estuaries [1], increased frequency and intensity of harmful algal blooms and fish kills [2], the accumulation of marine debris on beaches, suspended in seawater and within wildlife digestive tracts [3], and many other examples.

Future improvements in optical oceanography are needed to reliably measure complex coastal and inland water systems, and atmospheric correction will continue to be a persistent obstacle for satellite remote sensing. Technological improvements that are anticipated to improve atmospheric correction during upcoming NASA missions (e.g., PACE) include the expansion of sensor spectral range into the ultraviolet (UV), improvement of the spectral resolution at wavelengths spanning the UV to the near-infrared (NIR), and the addition of multi-angle polarimeters [4]. Results shown in this dissertation are relevant to assessing the algorithms that have been proposed to utilize anticipated sensor spectral improvements.

First, the likelihood of atmospheric correction failure should not be strongly correlated to the environmental parameters of interest, or else removal of the uncorrected imagery will alter the perspective obtained by remote sensing. Second, methods that retrieve aerosol content by assuming negligible UV water reflectances, e.g., [5], would prevent the use of

the spectral region that provides the greatest dynamic range and correlation to colored dissolved organic matter. Atmospheric correction schemes that assume negligible short-wave infrared (SWIR) signals [6] are more robust than similar NIR-based schemes, but the environmental signals at the legacy SWIR wavelengths are non-negligible in turbid inland water bodies or are not yet verified by field measurements [7]. High variability in the spectral shapes of coastal and inland waters supports the development of atmospheric correction approaches that do not make *a priori* assumptions of the water's signal, particularly at singular wavelengths.

Improvements to in-water and airborne methods will also be required to supplement satellite observations, particularly in the optically complex water bodies where atmospheric correction difficulties will persist. Chapter 2 demonstrated an airborne application for surveying colored dissolved organic matter in complex water bodies, and Chapter 3 described an approach for deriving chlorophyll *a* absorption using a beam-attenuation meter, which would reduce instrument redundancies and mitigate vulnerability to fouling in turbid systems. These projects made use of increased spectral range (Chapter 2) and resolution (Chapter 3), consistent with the improvements that are anticipated for next-generation satellite sensors. Finally, increased use of neural network methods are currently being tested for deriving atmospheric and aquatic constituents in optically complex water types, e.g., [8]. These new methods will likely become important tools in oceanographic remote sensing, but simpler approaches, for example based on fundamental bio-optical relationships (Chapter 2) or electromagnetic theory (Chapter 3), should not be discounted in future work on the measurement of optically complex water bodies.

References

[1] A. E. Hilton, J. T. Bausell, and R. M. Kudela, "Quantification of polychlorinated biphenyl (pcb) concentration in san francisco bay using satellite imagery," *Remote Sensing*, vol. 10, no. 7, p. 1110, 2018.

[2] A. J. Lewitus, R. A. Horner, D. A. Caron, E. Garcia-Mendoza, B. M. Hickey, M. Hunter, D. D. Huppert, R. M. Kudela, G. W. Langlois, J. L. Largier, *et al.*, "Harmful algal blooms along the north american west coast region: History, trends, causes, and impacts," *Harmful Algae*, vol. 19, pp. 133–159, 2012.

[3] S. C. Gall and R. C. Thompson, "The impact of debris on marine life," *Marine Pollution Bulletin*, vol. 92, no. 1-2, pp. 170–179, 2015.

[4] R. J. Frouin, B. A. Franz, A. Ibrahim, K. D. Knobelspiesse, Z. Ahmad, B. Cairns, J. Chowdhary, H. M. Dierssen, J. Tan, O. Dubovik, *et al.*, "Atmospheric correction of satellite ocean-color imagery during the pace era," *UMBC Faculty Collection*, 2019.

[5] X. He, Y. Bai, D. Pan, J. Tang, and D. Wang, "Atmospheric correction of satellite ocean color imagery using the ultraviolet wavelength for highly turbid waters," *Optics Express*, vol. 20, no. 18, pp. 20754–20770, 2012.

[6] M. Wang, "Remote sensing of the ocean contributions from ultraviolet to near-infrared using the shortwave infrared bands: simulations," *Applied Optics*, vol. 46, pp. 1535–1547, Mar 2007.

[7] N. Pahlevan, J.-C. Roger, and Z. Ahmad, "Revisiting short-wave-infrared (swir) bands for atmospheric correction in coastal waters," *Optics Express*, vol. 25, no. 6, pp. 6015–6035, 2017.

[8] N. Pahlevan, B. Smith, J. Schalles, C. Binding, Z. Cao, R. Ma, K. Alikas, K. Kangro, D. Gurlin, N. Hà, *et al.*, "Seamless retrievals of chlorophyll-a from sentinel-2 (msi) and sentinel-3 (olci) in inland and coastal waters: A machine-learning approach," *Remote Sensing of Environment*, p. 111604, 2020.