

# UC San Diego

## UC San Diego Electronic Theses and Dissertations

### Title

Development and Assessment of Inverse Optical Models in Support of Ocean Color Applications

### Permalink

<https://escholarship.org/uc/item/1x63f1cg>

### Author

Kehrli, Matthew Douglas

### Publication Date

2025

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Development and Assessment of Inverse Optical Models in Support of Ocean Color Applications

A Dissertation submitted in partial satisfaction of the requirements  
for the degree Doctor of Philosophy

in

Oceanography

by

Matthew Douglas Kehrl

Committee in charge:

Dariusz Stramski, Chair  
Lihini Aluwihare  
Alexander Groisman  
Rick Reynolds  
David Sandwell

2025

Copyright

Matthew Douglas Kehrl, 2025

All rights reserved.

The Dissertation of Matthew Douglas Kehrli is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2025

iii

## **DEDICATION**

*To the star throwers of the world — those who seek to make a difference, even when it seems futile.*

## TABLE OF CONTENTS

Dissertation Approval Page .....	iii
Dedication .....	iv
Table of Contents .....	v
List of Figures .....	viii
List of Tables .....	x
Acknowledgements .....	xi
Vita.....	xiv
Abstract of the Dissertation .....	xv
Introduction.....	1
References.....	10
Chapter 1 Estimation of chromophoric dissolved organic matter and non-algal particulate absorption coefficients of seawater in the ultraviolet by extrapolation from the visible spectral region .....	15
1.0 Abstract .....	16
1.1 Introduction.....	17
1.2 Methods.....	21
1.2.1 Assembly of the development dataset of absorption coefficients .....	21
1.2.2 Assembly of an independent validation dataset of absorption coefficients .....	28
1.2.3 Methods for extrapolating absorption coefficients from the VIS into the UV .....	29
1.2.4 Statistical evaluation of extrapolation models .....	32
1.3 Results and Discussion .....	33
1.3.1 Determination and assessment of the optimal extrapolation method.....	33
1.3.2 Development of correction functions and assessment of the final extrapolation .....	37
1.3.3 Application of the final extrapolation model using spectrally subsampled data in .....	43
the VIS .....	43
1.3.4 Validation of the extrapolation model with an independent dataset .....	46
1.4 Summary .....	47

1.5 Acknowledgements.....	50
1.6 Figures.....	51
1.7 Tables.....	64
1.8 References.....	69
Chapter 2 Model for partitioning the non-phytoplankton absorption coefficient of seawater in the ultraviolet and visible spectral range into the contributions of non-algal particulate and dissolved organic matter .....	76
2.0 Abstract.....	77
2.1 Introduction.....	78
2.2 Model Development and Validation Datasets .....	84
2.2.1 Assembly of the Development Dataset of Absorption Coefficients.....	84
2.2.2 Assembly of an Independent Validation Dataset of Absorption Coefficients .....	90
2.3 Statistical Methods for Model Evaluation .....	91
2.4 Description of ADG Absorption Partitioning Model.....	92
2.4.1 Model Overview .....	92
2.4.2 Libraries of Spectral Shape Functions of Constituent Absorption Coefficients.....	93
2.4.3 Speculative Solutions of the ADG Model .....	97
2.4.4 Feasible and Optimal Solutions of the ADG Model.....	99
2.5 Assessment of Model Performance .....	101
2.5.1 Evaluation with the Development Dataset.....	101
2.5.2 Evaluation with the Independent Validation Dataset.....	105
2.6 Summary and Conclusions .....	107
2.7 Acknowledgements.....	111
2.8 Figures.....	112
2.9 Tables.....	122
2.10 References.....	126
Chapter 3 Performance and uncertainty assessment of a novel multi-step semi-analytical algorithm for estimating seawater optical properties from ocean reflectance.....	134
3.0 Abstract.....	135
3.1 Introduction.....	136
3.2 Methods.....	141
3.2.1 Summary of Methodology .....	141
3.2.2 Summary of Synthetic Database.....	144
3.2.3 Description of Performance Assessment .....	146
3.2.4 Description of Uncertainty Assessment.....	147
3.2.5 Strategy to Assess the Performance and Uncertainty of 4SAA.....	149
3.3 Results and Discussion .....	151
3.3.1 Performance Assessment of Individual Component Models.....	151

3.3.2 Performance Assessment of Sequential Component Models.....	159
3.3.3 Uncertainty Assessment of Sequential Component Models .....	162
3.4 Summary and Conclusions .....	164
3.5 Acknowledgements.....	168
3.6 Figures.....	169
3.7 Tables.....	176
3.8 References.....	179

## LIST OF FIGURES

<b>Figure 1.1.</b> Geographic locations of oceanographic stations where near-surface measurements of the spectral absorption coefficients were collected for generation of the final model development dataset utilized in this study .....	51
<b>Figure 1.2.</b> Spectral absorption coefficients (a) $a_g(\lambda)$ , (b) $a_{dg}(\lambda)$ , and (c) $a_d(\lambda)$ available in the final model development dataset.....	52
<b>Figure 1.3.</b> Comparison of measured spectra of $a_g(\lambda)$ with fitted curves from exponential linear regression (ELR), exponential nonlinear regression (ENLR), stretched exponential nonlinear regression (SENLR), power linear regression (PLR), and power nonlinear regression (PNLR) .	53
<b>Figure 1.4.</b> Median ratio, $MdR$ , and median absolute percent difference, $MdAPD$ , vs. stop wavelength ( $\lambda_{stop}$ ) of the spectral regression window in the VIS range for (a, b) extrapolated $a_g(350)$ ( $N = 1294$ ) and (c, d) extrapolated $a_{dg}(350)$ ( $N = 409$ ) .....	54
<b>Figure 1.5.</b> Median ratio, $MdR$ , and median absolute percent difference, $MdAPD$ , vs. stop wavelength ( $\lambda_{stop}$ ) of (a, b) extrapolated $a_g$ ( $N = 1294$ ) and (c, d) extrapolated $a_{dg}$ ( $N = 409$ ) at five near-UV wavelengths.....	55
<b>Figure 1.6.</b> Scatter plots comparing extrapolated ( $a_g$ and $a_{dg}$ ) or calculated ( $a_d$ ) vs. measured absorption coefficients at 350 nm (a–c) and 380 nm (d–f).....	56
<b>Figure 1.7.</b> The relationship between the ratio of measured to extrapolated $a_g(\lambda)$ vs. extrapolated $a_g(\lambda)$ at 350 nm (a) and 380 nm (b) obtained from the optimal (ELR) extrapolation method.....	57
<b>Figure 1.8.</b> Median ratio, $MdR$ , of modeled to measured $a_d(\lambda)$ in the near-UV .....	58
<b>Figure 1.9.</b> Schematic describing the final extrapolation model for extending the spectral absorption coefficients $a_g(\lambda)$ , $a_d(\lambda)$ , and $a_{dg}(\lambda)$ from the VIS to the near-UV spectral region .....	59
<b>Figure 1.10.</b> The performance of the final extrapolation model applied to the development dataset and illustrated with scatter plots of the model-derived vs. measured absorption coefficients at 350 nm (a–c) and 380 nm (d–f) .....	60
<b>Figure 1.11.</b> Median ratio, $MdR$ , (a-c) and median absolute percent difference, $MdAPD$ , (d-f) in the near-UV spectral region for estimates of $a_g(\lambda)$ , $a_{dg}(\lambda)$ , and $a_d(\lambda)$ obtained from the final extrapolation model .....	61
<b>Figure 1.12.</b> Same statistics as presented in Fig. 1.11, but results determined with input data with the original 1 nm spectral sampling interval (black lines) are compared with results obtained by subsampling the input data.....	62
<b>Figure 1.13.</b> The performance of the final extrapolation model applied to the validation dataset ( $N = 149$ ) as scatter plots of model-derived versus measured absorption coefficients at 350 nm (a–c) and 380 nm (d–f).....	63
<b>Figure 2.1.</b> Global map depicting geographic locations of near-surface measurements comprising the final development dataset utilized in this study .....	112
<b>Figure 2.2.</b> Histograms of $a_d(443)$ and $a_g(443)$ from the final development dataset.....	113
<b>Figure 2.3.</b> Flowchart depicting the ADG partitioning model to partition the non-phytoplankton absorption coefficient, $a_{dg}(\lambda)$ , into its non-algal, $a_d(\lambda)$ , and color dissolved organic matter, $a_g(\lambda)$ .....	114

<b>Figure 2.4.</b> Joint probability distributions of the spectral steepness parameters, $S_d$ and $S_g$ , calculated respectively from (a) the spectral shape functions of non-algal particulate absorption coefficient, $\hat{a}_d(\lambda)$ , and (b) the spectral shape functions of CDOM absorption coefficient, $\hat{a}_g(\lambda)$	115
<b>Figure 2.5.</b> Characteristic spectral shape functions $\hat{a}_d(\lambda)$ and $\hat{a}_g(\lambda)$ comprising the final libraries used in the ADG partitioning model.....	116
<b>Figure 2.6.</b> (a–c) The median ratio, $MdR$ , and (d–f) the median symmetric accuracy, $MdSA$ , calculated from comparison of model-derived and measured values of $a_d(\lambda)$ , $a_g(\lambda)$ , and $a_{dg}(\lambda)$ using the set of concurrent measurements of $a_d(\lambda)$ and $a_g(\lambda)$ in the model development dataset .....	117
<b>Figure 2.7.</b> Assessment of the ADG_UV-VIS model applied to the development dataset as scatter plots of model-derived versus measured absorption coefficients $a_d(\lambda)$ and $a_g(\lambda)$ for light wavelengths of (a,d) 443 nm, (b,e) 555 nm, and (c,f) 670 nm .....	118
<b>Figure 2.8.</b> Similar to Fig. 2.7 but for wavelengths selected from the near-UV range: (a, c) 350 nm and (b, d) 380 nm.....	119
<b>Figure 2.9.</b> Spectral values of the (a–b) median ratio, $MdR$ , and (c–d) median symmetric accuracy, $MdSA$ , calculated from comparison of model-derived and measured values of $a_d(\lambda)$ and $a_g(\lambda)$ obtained from the ADG partitioning model applied to the development dataset ( $N = 390$ ).....	120
<b>Figure 2.10.</b> Similar to Fig. 2.9 but for the independent validation dataset ( $N = 149$ ) .....	121
<b>Figure 3.1.</b> Flowchart of the 4-step Semi-Analytical Algorithm (4SAA) .....	169
<b>Figure 3.2.</b> Synthetic database of 3320 samples of (a) remote sensing reflectance, $R_{rs}(\lambda)$ , (b) the average diffuse attenuation coefficient for downwelling irradiance over the first attenuation depth, $K_d(\lambda)$ , (c) the non-water absorption coefficient, $a_{nw}(\lambda)$ , and (d) the non-phytoplankton absorption coefficient, $a_{dg}(\lambda)$ .....	170
<b>Figure 3.3.</b> Assessment of Step 1, the KdNN model, applied to synthetic $R_{rs}(\lambda)$ as scatter plots of model-predicted versus synthetic $K_d(\lambda)$ for light wavelengths of (a) 350 nm, (b) 440 nm, (c) 490 nm, (d) 510 nm, (e) 550 nm, and (f) 650 nm .....	171
<b>Figure 3.4.</b> Spectral values of (a) median ratio, $MdR$ , and (b) median symmetric accuracy, $MdSA$ calculated from model-derived and synthetic reference values. The LS2 model is run with synthetic inputs of $R_{rs}(\lambda)$ and $K_d(\lambda)$ .....	172
<b>Figure 3.5.</b> Assessment of model-predicted non-water IOPs obtained from the LS2 model (Step 2) applied to synthetic $R_{rs}(\lambda)$ and KdNN (Step 1) model-derived $K_d(\lambda)$ .....	173
<b>Figure 3.6.</b> Assessment of model-predicted constituent absorption coefficients obtained from the ANW model (Step 3) applied to coupled KdNN and LS2 (Step 1 and Step 2) derived $a_{nw}(\lambda)$ ..	174
<b>Figure 3.7.</b> Spectral relative uncertainty obtained from the Monte Carlo analysis .....	175

## LIST OF TABLES

<b>Table 1.1.</b> Final Model Development Dataset of Absorption Coefficients, $a_g(\lambda)$ and $a_d(\lambda)$ .....	64
<b>Table 1.2.</b> Regression Models Evaluated for Extrapolation.....	65
<b>Table 1.3.</b> Spectral Values of the Parameters of the $E_g(\lambda)$ Function.....	66
<b>Table 1.4.</b> Spectral Values of $MdR$ for $a_d(\lambda)$ .....	67
<b>Table 1.5.</b> Error Statistics of the Final Extrapolation Model .....	68
<b>Table 2.1.</b> Model Development and Validation Datasets.....	122
<b>Table 2.2.</b> Statistical Metrics to Assess Model Performance .....	123
<b>Table 2.3.</b> Performance of ADG Partitioning Model with the Development Dataset .....	124
<b>Table 2.4.</b> Performance of ADG Partitioning Model with the Independent Validation Dataset	125
<b>Table 3.1.</b> Statistical Metrics to Assess Model Performance .....	176
<b>Table 3.2</b> Independent performance assessment of the ANW and ADG absorption partitioning models assessed with inputs from the synthetic database.....	177
<b>Table 3.3</b> Performance assessment of the ADG absorption partitioning model using the UV-VIS model variant assessed with the inputs obtained from the multistep sequence of the KdNN, LS2, and ANW .....	178

## ACKNOWLEDGEMENTS

This dissertation was only possible thanks to the countless people who have supported me.

First, I sincerely thank each of the members serving on my committee for their time and guidance: Dariusz Stramski, Lihini Aluwihare, Alexander Groisman, Rick Reynolds, and David Sandwell.

I am particularly grateful to Dariusz Stramski. He has been instrumental in molding both my personal character and career as a scientist. My first experience working with Dariusz was as a member of the Marine Physical Laboratory Summer Internship Program. Despite my lack of knowledge in ocean optics, he took the time to explain things to me in detail and ensured I had ample opportunities to ask questions. At the conclusion of the internship, he encouraged me to apply to the graduate program at the Scripps Institution of Oceanography. This initial patience and belief in my potential to develop into an early-career oceanographer is the source of my inspiration to pursue research in ocean optics. Dariusz has challenged my perspective of science in a positive manner with his diligent and disciplined approach to investigating problems by focusing on the physical concepts behind observations. Alongside this intensity, Dariusz is also genuine and empathetic, which has fostered my growth as a scientist. I aspire to continue to conduct high-quality research that parallels his rigorous approach and mentor others as he has compassionately mentored me in all my future scientific endeavors.

I would also like to thank Rick Reynolds. Rick fills many roles in his position as the Project Scientist for the Ocean Optics Research Laboratory and has served as an essential mentor during my graduate studies. In addition to teaching me the importance of maintaining an organized lab space, the usefulness of knowing boat knots, and the realities of working at large research universities, Rick has been an invaluable second set of eyes and ears for my research. The

trajectory of my studies would have been substantially altered if it were not for the quick feedback that Rick has provided during our many informal conversations throughout my Ph.D.

I want to thank all of the other current and former members of the lab, in particular Daniel Koestner. Daniel is a significant role model to me and brings a sense of calm thoughtfulness to research. I am truly grateful to call him a friend, colleague, and mentor. I would also like to thank Anjali Narayanan and Ishan Joshi for being the lab members that I can turn to for both constructive feedback about science and for light conversations about events outside of research.

To support a graduate student, it also takes a solid community of friends and peers. I am thankful for the friends I made in my life outside of grad school as their encouragement to continue to advance my career pushed me to pursue a Ph.D. I believe that personality and values are fundamentally determined by those you surround yourself with, and I am grateful to all of those in my past who shaped me into the person I am today. I would also like to thank my incoming SIO cohort as well as all SIO students in general. Whether we survived a global pandemic in the same house together or shared a drink at TG, you have made an essential impact on my life. I will always be here for you and please don't hesitate to reconnect. I'm confident that many of the friends I've made at SIO are lifelong - I can't wait to see what we accomplish!

I would also like to thank all of my family for their unwavering support and encouragement to follow my dreams in life and pursue my passions in my career. First, thank you to my grandparents: Wiley & Nancy Kehrli and James & Arlene Wick. I am so proud to be able to say that I'm the first doctor in the family, and without your selfless and lasting support this opportunity would not have been possible. I want to thank my dad, Steve Kehrli, for igniting my original curiosity to learn about mathematics and physics. I will always remember informally learning mathematical concepts, such as the properties of square roots or algebra, at an all too early age. I

am also grateful to my mom, Diane Kehrli, for the love and kindness she brings to my life. She has most importantly taught me to always be myself and appreciate the little things that bring joy in life, such as admiring a dog's breath on glass, cheesy movies on TV, or the simple enjoyment of a daily routine. Furthermore, I would like to thank my sister, Megan Kehrli, for listening to me when I need it most and always providing unique perspectives to appreciate the world.

Finally, a special thanks to all of my former teachers. It is through your dedication and commitment to education that any of this was possible in the first place. I am forever grateful!

Chapter 1, in full, is a reprint of the material as it appears in *Optics Express*. The dissertation author was the primary investigator and author of this paper: Kehrli, M. D., Stramski, D., Reynolds, R. A., & Joshi, I. D. (2023). Estimation of chromophoric dissolved organic matter and non-algal particulate absorption coefficients of seawater in the ultraviolet by extrapolation from the visible spectral region. *Optics Express*, 31(11), 17450–17479. <https://doi.org/10.1364/OE.486354>.

Chapter 2, in full, is a reprint of the material as it appears in *Applied Optics*. The dissertation author was the primary investigator and author of this paper: Kehrli, M. D., Stramski, D., Reynolds, R. A., & Joshi, I. D. (2024). Model for partitioning the non-phytoplankton absorption coefficient of seawater in the ultraviolet and visible spectral range into the contributions of non-algal particulate and dissolved organic matter. *Applied Optics* (2004), 63(16), 4252–4270. <https://doi.org/10.1364/AO.517706>.

Chapter 3, in part, is currently being prepared to be submitted for publication with authors Kehrli, M. D., Stramski, D., & Reynolds, R. A. The dissertation author will be the primary investigator and author of this paper.

## VITA

- 2019 Bachelor of Science in Physics, California Polytechnic State University San Luis Obispo
- 2021 Master of Science in Oceanography, University of California San Diego
- 2025 Doctor of Philosophy in Oceanography, University of California San Diego

## PUBLICATIONS

- Kehrli, M. D., Stramski, D., Reynolds, R. A., & Joshi, I. D. (2023). Estimation of chromophoric dissolved organic matter and non-algal particulate absorption coefficients of seawater in the ultraviolet by extrapolation from the visible spectral region. *Optics Express*, *31*(11), 17450–17479. <https://doi.org/10.1364/OE.486354>
- Kehrli, M. D., Stramski, D., Reynolds, R. A., & Joshi, I. D. (2024). Model for partitioning the non-phytoplankton absorption coefficient of seawater in the ultraviolet and visible spectral range into the contributions of non-algal particulate and dissolved organic matter. *Applied Optics* (2004), *63*(16), 4252–4270. <https://doi.org/10.1364/AO.517706>

## **ABSTRACT OF THE DISSERTATION**

Development and Assessment of Inverse Optical Models in Support of Ocean Color Applications

by

Matthew Douglas Kehrl

Doctor of Philosophy in Oceanography

University of California San Diego, 2025

Dariusz Stramski, Chair

Observations of ocean color from satellite imagery provide valuable opportunities to monitor biogeochemical systems of the ocean over broad spatiotemporal scales. To fully leverage the potential of these measurements, it is necessary to continue research on the development, assessment, and validation of inversion methods that estimate inherent optical properties (IOPs), which physically regulate the characteristics of ocean color. However, existing inversion models often face limitations in retrieving a complete suite of IOPs, particularly in the retrievals of absorption coefficients of seawater constituents. This limitation is relevant in the context of the

recently launched NASA PACE satellite mission which has the capability to measure hyperspectral radiances from the near-UV (350–400 nm) through the visible (400–700 nm). To address these challenges, this dissertation first focuses on advancing an absorption partitioning model that is incorporated into the broader structure of a four-step Semi-Analytical Algorithm (4SAA) designed to retrieve a full suite of seawater optical properties from satellite measurements. In the first chapter, a high-quality dataset of 1610 field samples of constituent absorption coefficients collected from diverse oceanic environments is compiled to develop a model to extrapolate the non-phytoplankton absorption coefficient,  $a_{dg}(\lambda)$ , the non-algal absorption coefficient,  $a_d(\lambda)$ , and the colored dissolved organic matter absorption coefficient,  $a_g(\lambda)$ , into the near-UV from the visible spectral region. This extrapolation model is assessed and validated using the compiled in situ dataset and provides a method to extend absorption partitioning models constrained to the visible spectral region. In the second chapter, the development of an absorption partitioning model, called ADG, is described. This model separates  $a_{dg}(\lambda)$  into its constituents,  $a_d(\lambda)$  and  $a_g(\lambda)$ , across the near-UV through visible spectral range. The ADG model is assessed and validated with the compiled dataset described in the previous chapter and demonstrates a strong capability to partition  $a_{dg}(\lambda)$ . This model consists of two variants, one of which incorporates the extrapolation model described in the first chapter and is implemented as the final component of the 4SAA model. The third chapter describes the analysis of a comprehensive performance and uncertainty assessment of the 4SAA model. This analysis utilized a recently published synthetic optical database to quantify the performance of each component model of 4SAA and investigate error propagation through the multi-step inversion sequence of 4SAA. A Monte Carlo approach was implemented to evaluate the uncertainty introduced by each component model as well as the propagation of uncertainty of the complete 4SAA model.

## INTRODUCTION

The perception of color is a fundamental human experience. When an object is exposed to white light, we observe the object's color as the light that is preferentially scattered by the object. Let's consider the green leaf of a tree as an example. We observe the leaf as green because it contains pigments, namely chlorophyll-a, that primarily absorb blue and red light. This light is absorbed for photosynthetic processes to store energy provided by sunlight as sugars for the tree. The process leaves more unabsorbed green light which is scattered by the leaf and eventually observed. The variations in the behavior of the different types of visible light as they are scattered or absorbed in the environment explain why objects appear as distinct colors.

The availability and behavior of light, namely sunlight, drive many physical and biogeochemical processes in the ocean. For example, microscopic phytoplankton absorb sunlight for photosynthesis and play a fundamental role in the global carbon cycle by forming the base of the marine food web. These organisms are responsible for carbon uptake, and without them, atmospheric carbon dioxide levels would be approximately 50% higher (Parekh et al., 2016). Phytoplankton respond to excess sunlight through photoadaptation by changing their cellular structure (Sournia 1974), and by synthesizing protective pigments to defend against the harmful effects of ultraviolet (UV) radiation (Karentz & Lutze, 1990; Karsten et al., 1999; Karentz 2001; Llewellyn & Harbour, 2003). UV radiation also affects non-phytoplankton material in the ocean, such as dissolved organic substances, and prolonged exposure can lead to their degradation through photobleaching processes (Armstrong et al., 1966; Vodacek et al., 1997; Mopper & Kieber, 2002; Helms et al., 2008). Taken together, these processes highlight the importance of studying the absorption of light in aquatic environments across a wide spectral range from the UV through the visible.

When sunlight reaches the top of Earth's atmosphere, a portion of the light propagates through the atmosphere, crosses the air-sea interface, and interacts with the seawater medium. Much of this light is lost through absorption and a fraction of it is scattered in the backward direction and exits the water column back into the atmosphere. By analyzing the ratio of the water-leaving light to the total amount of light that initially entered the water column it is possible to infer the optical properties of the seawater medium. The color of the ocean is characterized by the spectral composition of water-leaving light and typically it is quantitatively described by the remote sensing reflectance,  $R_{rs}(\lambda)$ , where  $\lambda$  represents the wavelength of light in vacuum. It is defined as the ratio of the spectral water-leaving radiance,  $L_w(\lambda)$ , to the spectral downwelling plane irradiance,  $E_d(\lambda)$ , just above the water surface (Jerlov 1976; Mobley 2022).  $R_{rs}(\lambda)$  is an apparent optical property (AOP) that can be measured in the field using near-surface radiometers or estimated from atmospherically-corrected measurement of top-of-atmosphere radiance with satellite remote sensing platforms. As an AOP,  $R_{rs}(\lambda)$  is defined as the ratio of two radiometric variables just above the water surface, making it dependent mostly on seawater inherent optical properties (IOPs) and only weakly influenced by temporally unstable other environmental factors, such as sun angle, sea surface state, and cloud cover. This temporal stability enables  $R_{rs}(\lambda)$  to be closely related to the compositional characteristics of seawater, making it an ideal AOP for investigating the seawater IOPs and associated biogeochemical constituents of seawater using satellites.

Various satellite-based sensors specially designed for remote measurements of ocean color have been developed over the past half century. The first pioneering mission began with the launch of the Nimbus 7 satellite in 1978 with its onboard Coastal Zone Color Sensor (CZCS) that measured ocean color over four channels in the visible electromagnetic spectrum (400–700nm).

The mission demonstrated the tremendous potential of utilizing ocean color measurements to acquire information about biological processes in the global ocean (Gordon et al., 1983; Yentsch & Yentsch, 1984). The success of CZCS led to the development of other satellite ocean color sensors and missions. Examples of heritage ocean color sensors designed for global ocean observations include the Sea-viewing Wide Field of View Sensor (SeaWiFS; 1997–2010), Moderate Resolution Imaging Spectroradiometer (MODIS; 1999–present), and the Visible Infrared Imaging Radiometer Suite (VIIRS; 2012–present) and together they have provided continuous spatiotemporal coverage of ocean color for over 25 years. Most recently, the Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) satellite mission was successfully launched in February 2024 and is now providing science-ready data products. The Ocean Color Instrument (OCI) is the advanced sensor aboard PACE that measures the top-of-the-atmosphere radiance across the near-ultraviolet through the near-infrared at a 5 nm hyperspectral resolution (Werdell et al., 2019).

These observations from ocean color satellite missions provide data products that are essential for understanding biogeochemical processes on large spatiotemporal scales. Among these, satellite-derived chlorophyll-a concentration has served as the flagship data product to characterize phytoplankton biomass since the CZCS era (Gordon et al., 1983; Feldman et al., 1984). Another critical product obtained from satellite observations is the concentration of particulate organic carbon, which characterizes the dynamic pool of organic carbon associated with particles in the ocean (Stramski et al., 1999). These fundamental data products have facilitated the development of many other satellite-derived data products to investigate phytoplankton community composition (Mouw et al., 2017), primary production (Antoine et al., 1996; Behrenfeld et al., 2005; Uitz et al., 2010; Siegel et al., 2013; Gregg et al., 2017; Westberry et al., 2023), carbon transport and sequestration (Brewin 2021), and water quality (Hu et al., 2004; Shaeffer et al., 2012;

Zheng et al., 2017; Turner et al., 2022). Furthermore, time series of measurements of ocean color advance the knowledge of trends in ocean color over larger timescales. By merging ocean color data from various satellites, the surface of the ocean has become greener at low-latitudes over the last twenty years (Cael et al., 2023). Numerous studies have also examined deviations of satellite-derived surface chlorophyll-a concentrations derived to investigate long-term trends of global phytoplankton distributions (Gregg & Conkright, 2002; Antoine et al., 2005).

On a more fundamental level,  $R_{rs}(\lambda)$  is directly related to the physical properties of the seawater medium itself, and optical oceanographers rely on  $R_{rs}(\lambda)$  to study the intrinsic optical characteristics of the ocean. More formally, the propagation of light through seawater is controlled by its inherent optical properties (IOPs), which are defined as the light absorption and scattering characteristics of the seawater medium that depend solely on the concentrations and composition of seawater constituents. Measurements of IOPs are independent of the ambient light field and remain the same whether determined in the field or the laboratory, provided that the temperature, pressure, and concentration of seawater constituents of the medium remain unchanged (Mobley 2022). When light interacts with seawater, it is either absorbed, elastically (i.e., no change in light wavelength) or inelastically (i.e., change in wavelength) scattered at an angle,  $\psi$ , or continues to propagate unattenuated. The two IOPs that fully describe (neglecting polarization effects) the behavior of light in seawater are the total spectral absorption coefficient,  $a(\lambda)$ , and the volume scattering function,  $\beta(\psi, \lambda)$ . Integrating  $\beta(\psi, \lambda)$  over all angles (i.e., from  $\psi$  equals  $0^\circ$  to  $180^\circ$  relative to the original direction of photon travel) yields the total scattering coefficient,  $b(\lambda)$ , and integration over all backward angles (i.e., from  $\psi$  equals  $90^\circ$  to  $180^\circ$  relative to the original direction of photon travel) yields the backscattering coefficient,  $b_b(\lambda)$ . The reciprocal of each bulk IOP corresponds to the average distance light travels before undergoing a specific attenuation event. For instance, if

the absorption coefficient  $a(650)$  is equal to  $0.5 \text{ m}^{-1}$ , light with a wavelength of 650 nm will on average travel two meters before being absorbed.

Bulk IOPs are dependent on the additive contributions of seawater constituents, which consist of molecular water, phytoplankton, non-algal particles (also referred to as detrital or depigmented particles), and chromophoric (also referred to as colored) dissolved organic matter (CDOM). The IOPs of these constituent absorption coefficients are denoted as  $a_w(\lambda)$  for the pure seawater absorption coefficient,  $a_{ph}(\lambda)$  for the phytoplankton absorption coefficient,  $a_d(\lambda)$  for the non-algal particulate absorption coefficient, and  $a_g(\lambda)$  for the CDOM absorption coefficient. Constituent IOPs vary in proportion to the concentration of each seawater constituent and investigating their relationships with AOPs can reveal underlying information about the characteristics of the seawater medium itself. For example, observations of  $R_{rs}(\lambda)$  are closely coupled to the amount of light that is absorbed and backscattered within seawater and, to first order, vary in proportion to the ratio of  $b_b(\lambda)$  to  $a(\lambda)$ . Another essential AOP is the spectral diffuse attenuation coefficient of downward plane irradiance,  $K_d(\lambda)$ , which describes the attenuation of light propagating through the water column in downward directions at various wavelengths  $\lambda$ .  $K_d(\lambda)$  approximately covaries with the sum of  $a(\lambda)$  and  $b_b(\lambda)$  since these IOPs describe the portion of light that is removed from downward propagating light by the absorption and backscattering processes (Kirk 2011, Mobley 2022). A primary goal of satellite remote sensing in optical oceanography is to use the relationships between AOPs and IOPs to gain insights about light propagation in the water column and to advance the understanding of ecological dynamics and biogeochemical processes in the upper ocean.

Measurements of ocean color obtained by PACE OCI and heritage sensors offer the potential to estimate total and constituent IOPs through inversion modeling techniques. It is well-

established through radiative transfer theory how to determine a sought-after AOP, such as  $R_{rs}(\lambda)$  or  $K_d(\lambda)$ , from a complete set of measurements of IOPs (Preisendorfer 1976). The inverse problem, which is to determine an IOP from a measurement of an AOP, is more challenging as there are no exact deterministic relationships that link IOPs and AOPs, and there are many potential values for IOPs that lead to the same AOP observation. Typically, inversion algorithms designed to solve this problem are developed by using a semi-analytical approach and rely on a combination of empirical measurements and analytical solutions of the radiative transfer equation to derive constituent IOPs from radiometric data.

Many different versions of semi-analytical algorithms (SAAs) have been constructed since the launch of the first ocean color sensors to estimate constituent IOPs from  $R_{rs}(\lambda)$  (Werdell et al., 2018). One solution method is to first determine bulk IOPs and then utilize independent partitioning models to estimate constituent IOPs. An advantage of this multi-step approach for solving the inverse problem is that it permits the development of independent mechanistically-based component models which can be optimized to accurately retrieve target variables of interest, such as constituent absorption coefficients. Typically, multi-step SAAs first estimate values of  $a(\lambda)$  and  $b_b(\lambda)$  and then subtract well-known values of pure seawater components,  $a_w(\lambda)$  (Twardowski et al., 2018) and  $b_{bw}(\lambda)$  (Zhang 2009), to obtain estimates of the non-water absorption coefficient,  $a_{nw}(\lambda)$ , and the particulate backscattering coefficient,  $b_{bp}(\lambda)$ . Further partitioning the estimated value of  $a_{nw}(\lambda)$  into constituent-specific absorption coefficients presents a challenge as the concentrations and compositions of each non-water constituent are highly variable throughout the ocean. Most inversion algorithms that partition  $a_{nw}(\lambda)$  do not separate its non-algal particulate,  $a_d(\lambda)$ , and chromophoric dissolved organic matter,  $a_g(\lambda)$ , components and instead combine them into the single non-phytoplankton absorption coefficient,  $a_{dg}(\lambda)$ , due to the similarity in the spectral

shapes of  $a_d(\lambda)$  and  $a_g(\lambda)$ . Other methods that retrieve  $a_d(\lambda)$  and  $a_g(\lambda)$  are limited as they assume a single exponential function for the spectral shapes of  $a_d(\lambda)$  and  $a_g(\lambda)$  over broad spectral ranges.

A proposed 4-step Semi-Analytical Algorithm (4SAA) has been recently developed in the Ocean Optics Research Laboratory at Scripps Institution of Oceanography with parts of this development involving a collaboration with French investigators in support of the NASA PACE satellite mission (See Figure 3.1 in Chapter 3). The main aim of 4SAA is to retrieve bulk and constituent IOPs by alleviating limitations found in previous inversion schemes. The 4SAA is comprised of four independent component models to estimate one AOP ( $K_d(\lambda)$ ) and nine IOPs ( $a(\lambda)$ ,  $a_{nw}(\lambda)$ ,  $a_p(\lambda)$ ,  $a_{dg}(\lambda)$ ,  $a_{ph}(\lambda)$ ,  $a_d(\lambda)$ ,  $a_g(\lambda)$ ,  $b_b(\lambda)$ , and  $b_{bp}(\lambda)$ ) from input  $R_{rs}(\lambda)$ . The first step of the 4SAA model employs a neural-network based model to estimate hyperspectral  $K_d(\lambda)$  in the near-UV and VIS spectral regions from input  $R_{rs}(\lambda)$  (Jamet et al., 2012, Loisel et al., 2018, Jorge et al., 2021). The model was developed with in situ and analytically-derived synthetic datasets of  $R_{rs}(\lambda)$  and  $K_d(\lambda)$ . The second step of 4SAA is an inverse AOP-IOP model, also known as LS2, that estimates  $a(\lambda)$  and  $b_b(\lambda)$  and their non-water components (Loisel & Stramski, 2000, Loisel et al., 2018). The inverse AOP-IOP model utilizes look-up tables generated from radiative transfer simulations that are independent of light wavelength and makes no assumptions about the spectral shapes of constituent IOPs to solve the inverse problem. Independently developed absorption partitioning models are then applied in the final two steps of 4SAA. In the third step, an absorption partitioning model, known as the ANW model, separates  $a_{nw}(\lambda)$  into  $a_{ph}(\lambda)$  and  $a_{dg}(\lambda)$  (Stramski & Reynolds, in preparation). The ANW partitioning model implements a spectral shape function library of  $a_{dg}(\lambda)$  to initially partition  $a_{nw}(\lambda)$  and then applies multiple inequality constraints associated mainly with spectral band ratios of  $a_{ph}(\lambda)$  (Zheng and Stramski, 2013) to generate realistic estimates of  $a_{ph}(\lambda)$  and  $a_{dg}(\lambda)$ . Finally, the 4SAA model uses an ADG partitioning model

to separate  $a_d(\lambda)$  and  $a_g(\lambda)$  from  $a_{dg}(\lambda)$  derived from in pervious step. The proposed ADG model incorporates a partitioning approach based on Stramski et al. (2019), however, improvements are needed to this model as it is constrained to the visible spectrum, its spectral shape function library contains data from sources with inconsistent measurement methodology, and additional model parameterizations could enhance its overall performance. The potential advantages of the multi-step structure of the 4SAA and promising initial performance of each component model are directly relevant to the topic of the dissertation presented herein.

The dissertation is comprised of three chapters to investigate the development and assessment of the 4SAA with a particular focus on the ADG absorption partitioning modeling methods to separate the non-phytoplankton absorption coefficient,  $a_{dg}(\lambda)$ , into its  $a_d(\lambda)$  and  $a_g(\lambda)$  components in the 4<sup>th</sup> step of the 4SAA model. The complete dissertation highlights the compilation of a novel in situ database of measurements of absorption coefficients, the development of ADG partitioning model covering both the near-UV (350–400 nm) and VIS (400–700 nm) spectral ranges, and the performance and uncertainty assessment associated with each individual component model of 4SAA. The first chapter, “Estimation of chromophoric dissolved organic matter and non-algal particulate absorption coefficients of seawater in the ultraviolet by extrapolation from the visible spectral region” addresses the challenge of obtaining spectral values of  $a_{dg}(\lambda)$  and its  $a_d(\lambda)$  and  $a_g(\lambda)$  constituents in the near-UV spectral region. The aim of the chapter is to assemble rigorously quality controlled development and validation datasets of constituent absorption coefficients from a diverse set of optical environments to formulate and evaluate the optimal extrapolation model to extend  $a_{dg}(\lambda)$ ,  $a_d(\lambda)$ , and  $a_g(\lambda)$  to the near-UV from measurements in the visible. The second chapter, “Model for partitioning the non-phytoplankton absorption coefficient of seawater in the ultraviolet and visible spectral range into the contributions of non-

algal particulate and dissolved organic matter” seeks to develop the ADG model to partition  $a_{dg}(\lambda)$  into  $a_d(\lambda)$  and  $a_g(\lambda)$  with an intent to be utilized in the fourth step of the 4SAA. The ADG model aims to complete this partitioning across the near-UV through VIS spectral range without the reliance on limiting assumptions about the single exponential spectral shapes of constituent absorption coefficients. This analysis synthesizes the investigation of a variety of spectral shape function libraries necessary for the algebraic-based solution method to partition  $a_{dg}(\lambda)$  and the assessment of model performance when applied to the datasets compiled in the previous chapter. The third chapter, “Performance and uncertainty assessment of a novel multi-step semi-analytical algorithm for estimating seawater optical properties from ocean reflectance” focuses on quantifying the performance and propagation of uncertainties in each step of the multistep algorithm. The analysis evaluates each component model of 4SAA through a performance assessment with a recently published synthetic optical database (Loisel et al., 2023). Furthermore, this analysis highlights a novel approach to evaluate the uncertainty of each component model independently as well as the propagation of model uncertainty through the complete multi-step model sequence of 4SAA.

## References

- Antoine, D., André, J., & Morel, A. (1996). Oceanic primary production: 2. Estimation at global scale from satellite (Coastal Zone Color Scanner) chlorophyll. *Global Biogeochemical Cycles*, *10*(1), 57–69. <https://doi.org/10.1029/95GB02832>
- Antoine, D., Morel, A., Gordon, H. R., Banzon, V. F., & Evans, R. H. (2005). Bridging ocean color observations of the 1980s and 2000s in search of long-term trends. *Journal of Geophysical Research. C. Oceans*, *110*(C6), C06009. <https://doi.org/10.1029/2004JC002620>
- Armstrong, F. A. J., Williams, P. M., & Strickland, J. D. H. (1966). Photo-oxidation of Organic Matter in Sea Water by Ultra-violet Radiation, Analytical and Other Applications. *Nature*, *211*(5048), 481–483. <https://doi.org/10.1038/211481a0>
- Behrenfeld, M. J., Boss, E., Siegel, D. A., & Shea, D. M. (2005). Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochemical Cycles*, *19*(1), GB1006. <https://doi.org/10.1029/2004GB002299>
- Brewin, R. J. W., Sathyendranath, S., Platt, T., Bouman, H., Ciavatta, S., Dall’Olmo, G., Dingle, J., Groom, S., Jonsson, B., Kostadinov, T. S., Kulk, G., Laine, M., Martinez-Vicente, V., Psarra, S., Raitzos, D. E., Richardson, K., Rio, M.-H., Rousseaux, C. S., Salisbury, J., ... Walker, P. (2021). Sensing the ocean biological carbon pump from space; a review of capabilities, concepts, research gaps and future developments. *Earth-Science Reviews*, *217*, 103604-. <https://doi.org/10.1016/j.earscirev.2021.103604>
- Cael, B. B., Bisson, K., Boss, E., Dutkiewicz, S., & Henson, S. (2023). Global climate-change trends detected in indicators of ocean ecology. *Nature*, *619*(7970), 551–554. <https://doi.org/10.1038/s41586-023-06321-z>
- Feldman, G., Clark, D., & Halpern, D. (1984). Satellite Color Observations of the Phytoplankton Distribution in the Eastern Equatorial Pacific During the 1982-1983. El Nino. *Science*, *226*(4678), 1069–1071. <https://doi.org/10.1126/science.226.4678.1069>
- Gordon, H., Clark, D., Brown, J., Brown, O., Evans, R., & Broenkow, W. (1983). Phytoplankton pigment concentrations in the Middle Atlantic Bight: comparison of ship determinations and CZCS estimates. *Applied Optics*, *22*(1), 20–36. <https://doi.org/10.1364/ao.22.000020>
- Gregg, W. W., & Conkright, M. E. (2002). Decadal changes in global ocean chlorophyll. *Geophysical Research Letters*, *29*(15), 20-1. <https://doi.org/10.1029/2002GL014689>
- Gregg, W. W., Rousseaux, C. S., & Franz, B. A. (2017). Global trends in ocean phytoplankton: a new assessment using revised ocean colour data. *Remote Sensing Letters*, *8*(12), 1102–1111. <https://doi.org/10.1080/2150704X.2017.1354263>
- Helms, J. R., Stubbins, A., Ritchie, J. D., Minor, E. C., Kieber, D. J., & Mopper, K. (2008). Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and

- photobleaching of chromophoric dissolved organic matter. *Limnology and Oceanography*, 53(3), 955–969. <https://doi.org/10.4319/lo.2008.53.3.0955>
- Hu, C., Chen, Z., Clayton, T. D., Swarzenski, P., Brock, J. C., & Muller–Karger, F. E. (2004). Assessment of estuarine water-quality indicators using MODIS medium-resolution bands: Initial results from Tampa Bay, FL. *Remote Sensing of Environment*, 93(3), 423–441. <https://doi.org/10.1016/j.rse.2004.08.007>
- Jamet, C., Loisel, H., & Dessailly, D. (2012). Retrieval of the spectral diffuse attenuation coefficient  $K_d(\lambda)$  in open and coastal ocean waters using a neural network inversion. *Journal of Geophysical Research*, 117(C10), C10023. <https://doi.org/10.1029/2012JC008076>
- Jerlov, N. G. (1976). *Marine optics*. Elsevier Scientific Pub. Co.
- Jorge, D. S. F., Loisel, H., Jamet, C., Dessailly, D., Demaria, J., Bricaud, A., Maritorea, S., Zhang, X., Antoine, D., Kutser, T., Bélanger, S., Brando, V. O., Werdell, J., Kwiatkowska, E., Mangin, A., & d’Andon, O. F. (2021). A three-step semi analytical algorithm (3SAA) for estimating inherent optical properties over oceanic, coastal, and inland waters from remote sensing reflectance. *Remote Sensing of Environment*, 263, 112537. <https://doi.org/10.1016/j.rse.2021.112537>
- Karentz, D., & Lutze, L. H. (1990). Evaluation of biologically harmful ultraviolet radiation in Antarctica with a biological dosimeter designed for aquatic environments. *Limnology and Oceanography*, 35(3), 549–561. <https://doi.org/10.4319/lo.1990.35.3.0549>
- Karentz, D. (2001). Chemical Defenses of Marine Organisms Against Solar Radiation Exposure: UV-Absorbing Mycosporine-Like Amino Acids and Scytonemin. In *Marine Chemical Ecology* (pp. 493–532). CRC Press. <https://doi.org/10.1201/9781420036602-22>
- Karsten, U., Bischof, K., Hanelt, D., Tüg, H., & Wiencke, C. (1999). The effect of ultraviolet radiation on photosynthesis and ultraviolet-absorbing substances in the endemic Arctic macroalga *Devaleraea ramentacea* (Rhodophyta). *Physiologia Plantarum*, 105(1), 58–66. <https://doi.org/10.1034/j.1399-3054.1999.105110.x>
- Kirk, J. T. O. (2011). *Light and photosynthesis in aquatic ecosystems* (3rd ed.). Cambridge University Press.
- Loisel, H., & Stramski, D. (2000). Estimation of the inherent optical properties of natural waters from the irradiance attenuation coefficient and reflectance in the presence of Raman scattering. *Applied Optics*, 39(18), 3001–3011. <https://doi.org/10.1364/AO.39.003001>
- Loisel, H., Stramski, D., Dessailly, D., Jamet, C., Li, L., & Reynolds, R. A. (2018). An inverse model for estimating the optical absorption and backscattering coefficients of seawater from remote-sensing reflectance over a broad range of oceanic and coastal marine environments. *Journal of Geophysical Research: Oceans*, 123(3), 2141–2171. <https://doi.org/10.1002/2017JC013632>

- Loisel, H., Jorge, D. S. F., Reynolds, R. A., & Stramski, D. (2023). A synthetic optical database generated by radiative transfer simulations in support of studies in ocean optics and optical remote sensing of the global ocean. *Earth System Science Data*, 15(8), 3711–3731. <https://doi.org/10.5194/essd-15-3711-2023>
- Llewellyn, C. A., & Harbour, D. S. (2003). A temporal study of mycosporine-like amino acids in surface water phytoplankton from the English Channel and correlation with solar irradiation. *Journal of the Marine Biological Association of the United Kingdom*, 83(1), 1–9. <https://doi.org/10.1017/S0025315403006726h>
- Mobley, C. (2022). *The oceanic optics book*.
- Mopper, K., & Kieber, D. J. (2002). Photochemistry and the cycling of carbon, sulfur, nitrogen and phosphorus. *Biogeochemistry of marine dissolved organic matter*, 455-507.
- Mouw, C. B., Hardman-Mountford, N. J., Alvain, S., Bracher, A., Brewin, R. J. W., Bricaud, A., Ciotti, A. M., Devred, E., Fujiwara, A., Hirata, T., Hirawake, T., Kostadinov, T. S., Roy, S., & Uitz, J. (2017). A Consumer’s Guide to Satellite Remote Sensing of Multiple Phytoplankton Groups in the Global Ocean. *Frontiers in Marine Science*, 4, 237601. <https://doi.org/10.3389/fmars.2017.00041>
- Parekh, P., Dutkiewicz, S., Follows, M. J., & Ito, T. (2006). Atmospheric carbon dioxide in a less dusty world. *Geophysical Research Letters*, 33(3). <https://doi.org/10.1029/2005GL025098>
- Preisendorfer, R. W. (1976). *Hydrologic optics*. U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, Environmental Research Laboratories, Pacific Marine Environmental Laboratory.
- Schaeffer, B. A., Hagy, J. D., Conmy, R. N., Lehrter, J. C., & Stumpf, R. P. (2012). An Approach to Developing Numeric Water Quality Criteria for Coastal Waters Using the SeaWiFS Satellite Data Record. *Environmental Science & Technology*, 46(2), 916–922. <https://doi.org/10.1021/es2014105>
- Siegel, D. A., Behrenfeld, M. J., Maritorena, S., McClain, C. R., Antoine, D., Bailey, S. W., Bontempi, P. S., Boss, E. S., Dierssen, H. M., Doney, S. C., Eplee, R. E., Evans, R. H., Feldman, G. C., Fields, E., Franz, B. A., Kuring, N. A., Mengelt, C., Nelson, N. B., Patt, F. S., ... Yoder, J. A. (2013). Regional to global assessments of phytoplankton dynamics from the SeaWiFS mission. *Remote Sensing of Environment*, 135, 77–91. <https://doi.org/10.1016/j.rse.2013.03.025>
- Sournia, A. (1974). Circadian Periodicities in Natural Populations of Marine Phytoplankton.. *Advances in Marine Biology*, 12, 325–389.
- Stramski, D., Reynolds, R. A., Kahru, M., & Mitchell, B. G. (1999). Estimation of Particulate Organic Carbon in the Ocean from Satellite Remote Sensing. *Science*, 285(5425), 239–242. <https://doi.org/10.1126/science.285.5425.239>

- Stramski, D., Li, L., & Reynolds, R. A. (2019). Model for separating the contributions of non-algal particles and colored dissolved organic matter to light absorption by seawater. *Applied Optics (2004)*, 58(14), 3790–3806. <https://doi.org/10.1364/AO.58.003790>
- Stramski, D., & Reynolds, R. A. (in preparation). *Model for partitioning the nonwater absorption coefficient of seawater in the ultraviolet and visible spectral range into the contributions of phytoplankton and non-phytoplankton components.*
- Turner, J. S., Fall, K. A., & Friedrichs, C. T. (2023). Clarifying water clarity: A call to use metrics best suited to corresponding research and management goals in aquatic ecosystems. *Limnology and Oceanography Letters*, 8(3), 388–397. <https://doi.org/10.1002/lol2.10301>
- Twardowski, M., Röttgers, R., and Stramski, D., “Chapter 1: The Absorption Coefficient, An Overview,” In: IOCCG Protocol Series Inherent Optical Property Measurements and Protocols: Absorption Coefficient, Neeley, A. R. and Mannino, A. (eds.), IOCCG Ocean Optics and Biogeochemistry Protocols for Satellite Ocean Colour Sensor Validation, Volume 1.0, IOCCG, Dartmouth, NS, Canada (2018).
- Uitz, J., Claustre, H., Gentili, B., & Stramski, D. (2010). Phytoplankton class-specific primary production in the world’s oceans: Seasonal and interannual variability from satellite observations. *Global Biogeochemical Cycles*, 24(3), GB3016. <https://doi.org/10.1029/2009GB003680>
- Vodacek, A., Blough, N. V., DeGrandpre, M. D., Peltzer, E. T., & Nelson, R. K. (1997). Seasonal variation of CDOM and DOC in the middle Atlantic Bight; terrestrial inputs and photooxidation. *Limnology and Oceanography*, 42(4), 674–686. <https://doi.org/10.4319/lo.1997.42.4.0674>
- Werdell, P. J., McKinna, L. I. W., Boss, E., Ackleson, S. G., Craig, S. E., Gregg, W. W., Lee, Z., Maritorena, S., Roesler, C. S., Rousseaux, C. S., Stramski, D., Sullivan, J. M., Twardowski, M. S., Tzortziou, M., & Zhang, X. (2018). An overview of approaches and challenges for retrieving marine inherent optical properties from ocean color remote sensing. *Progress in Oceanography*, 160, 186–212. <https://doi.org/10.1016/j.pocean.2018.01.001>
- Werdell, P. J., Behrenfeld, M. J., Bontempi, P. S., Boss, E., Cairns, B., Davis, G. T., Franz, B. A., Gliese, U. B., Gorman, E. T., Hasekamp, O., Knobelspiesse, K. D., Mannino, A., Martins, J. V., McClain, C. R., Meister, G., & Remer, L. A. (2019). The Plankton, Aerosol, Cloud, Ocean Ecosystem Mission: Status, Science, Advances. *Bulletin of the American Meteorological Society*, 100(9), 1775–1794. <https://doi.org/10.1175/BAMS-D-18-0056.1>
- Westberry, T. K., Silsbe, G. M., & Behrenfeld, M. J. (2023). Gross and net primary production in the global ocean: An ocean color remote sensing perspective. *Earth-Science Reviews*, 237, 104322. <https://doi.org/10.1016/j.earscirev.2023.104322>
- Yentsch, C. M., and C. S. Yentsch. 1984. Emergence of optical instrumentation for measuring biological properties. *Oceanography and Marine Biology Annual Revue*, 2, 55–98.

- Zhang, X., Hu, L., & He, M.-X. (2009). Scattering by pure seawater: Effect of salinity. *Optics Express*, 17(7), 5698–5710. <https://doi.org/10.1364/oe.17.005698>
- Zheng, G., & Stramski, D. (2013). A model based on stacked-constraints approach for partitioning the light absorption coefficient of seawater into phytoplankton and non-phytoplankton components. *Journal of Geophysical Research. Oceans*, 118(4), 2155–2174. <https://doi.org/10.1002/jgrc.20115>
- Zheng, G., & DiGiacomo, P. M. (2017). Uncertainties and applications of satellite-derived coastal water quality products. *Progress in Oceanography*, 159, 45–72. <https://doi.org/10.1016/j.pocean.2017.08.007>

## **Chapter 1**

**Estimation of chromophoric dissolved organic matter and non-algal  
particulate absorption coefficients of seawater in the ultraviolet by  
extrapolation from the visible spectral region**

## 1.0 Abstract

Extending the capabilities of optical remote sensing and inverse optical algorithms, which have been commonly focused on the visible (VIS) range of the electromagnetic spectrum, to derive the optical properties of seawater in the ultraviolet (UV) range is important to advancing the understanding of various optical, biological, and photochemical processes in the ocean. In particular, existing remote-sensing reflectance models that derive the total spectral absorption coefficient of seawater,  $a(\lambda)$ , and absorption partitioning models that partition  $a(\lambda)$  into the component absorption coefficients of phytoplankton,  $a_{\text{ph}}(\lambda)$ , non-algal (depigmented) particles,  $a_{\text{d}}(\lambda)$ , and chromophoric dissolved organic matter (CDOM),  $a_{\text{g}}(\lambda)$ , are restricted to the VIS range. We assembled a quality-controlled development dataset of hyperspectral measurements of  $a_{\text{g}}(\lambda)$  ( $N = 1294$ ) and  $a_{\text{d}}(\lambda)$  ( $N = 409$ ) spanning a wide range of values across various ocean basins, and evaluated several extrapolation methods to extend  $a_{\text{g}}(\lambda)$ ,  $a_{\text{d}}(\lambda)$ , and  $a_{\text{dg}}(\lambda) \equiv a_{\text{g}}(\lambda) + a_{\text{d}}(\lambda)$  into the near-UV spectral region by examining different sections of the VIS as a basis for extrapolation, different extrapolation functions, and different spectral sampling intervals of input data in the VIS. Our analysis determined the optimal method to estimate  $a_{\text{g}}(\lambda)$  and  $a_{\text{dg}}(\lambda)$  at near-UV wavelengths (350 to 400 nm) which relies on an exponential extrapolation of data from the 400–450 nm range. The initial  $a_{\text{d}}(\lambda)$  is obtained as a difference between the extrapolated estimates of  $a_{\text{dg}}(\lambda)$  and  $a_{\text{g}}(\lambda)$ . Additional correction functions based on the analysis of differences between the extrapolated and measured values in the near-UV were defined to obtain improved final estimates of  $a_{\text{g}}(\lambda)$  and  $a_{\text{d}}(\lambda)$  and then the final estimates of  $a_{\text{dg}}(\lambda)$  as a sum of final  $a_{\text{g}}(\lambda)$  and  $a_{\text{d}}(\lambda)$ . The extrapolation model provides very good agreement between the extrapolated and measured data in the near-UV when the input data in the blue spectral region are available at 1 or 5 nm spectral sampling intervals. There is negligible bias between the modeled and measured values of all three absorption coefficients and the median absolute percent difference (*MdAPD*) is small, e.g.,  $< 5.2\%$  for  $a_{\text{g}}(\lambda)$

and  $< 10.5\%$  for  $a_d(\lambda)$  at all near-UV wavelengths when evaluated with the development dataset. Assessment of the model on an independent dataset of concurrent  $a_g(\lambda)$  and  $a_d(\lambda)$  measurements ( $N = 149$ ) yielded similar findings with only slight reduction of performance and *MdAPD* remaining below  $6.7\%$  for  $a_g(\lambda)$  and  $11\%$  for  $a_d(\lambda)$ . These results are promising for integration of the extrapolation method with absorption partitioning models operating in the VIS.

## 1.1 Introduction

The propagation of light in natural waters is governed by their inherent optical properties (IOPs). Two basic IOPs that characterize the attenuation of a light beam through absorption and scattering processes at a particular light wavelength,  $\lambda$  (which is generally reported in vacuum in units of nm), are the total spectral absorption coefficient of seawater,  $a(\lambda)$  ( $\text{m}^{-1}$ ), and the spectral volume scattering function,  $\beta(\psi, \lambda)$  ( $\text{m}^{-1} \text{sr}^{-1}$ ) where  $\psi$  is the scattering angle [1]. The total absorption coefficient  $a(\lambda)$  is typically divided into three or four additive components, i.e., absorption by pure water,  $a_w(\lambda)$ , absorption by chromophoric dissolved organic matter (CDOM),  $a_g(\lambda)$ , and absorption by suspended particles,  $a_p(\lambda)$ , in which the latter coefficient can be further partitioned into contributions by phytoplankton,  $a_{ph}(\lambda)$ , and non-algal particles (NAP),  $a_d(\lambda)$ . Given the methodology of measuring  $a_d(\lambda)$  following extraction and removal from the sample of pigments that are contained primarily in phytoplankton [2], it is worthwhile to recall that the NAP component can be thought of as representing the absorption by depigmented particulate matter. Although the values of  $a_w(\lambda)$  are quite well quantified from the near-ultraviolet (near-UV) through the visible (VIS) and into the infrared (IR) region of the electromagnetic spectrum [3,4], the particulate and dissolved components of seawater are highly variable and drive the large variability of  $a(\lambda)$  in the ocean. The two absorption components,  $a_g(\lambda)$  and  $a_d(\lambda)$ , exhibit similarities in the spectral shapes with a general tendency to increase with decreasing wavelength, and are often summed together as  $a_{dg}(\lambda) = a_g(\lambda) + a_d(\lambda)$ . All three of these absorption coefficients have been

often approximated as an exponentially decaying function of light wavelength from the UV through the VIS as defined by:

$$a_x(\lambda) = a_x(\lambda_r) \exp[-S_x(\lambda - \lambda_r)] \quad (1.1)$$

where  $a_x(\lambda)$  is either  $a_g(\lambda)$ ,  $a_d(\lambda)$ , or  $a_{dg}(\lambda)$ ,  $\lambda_r$  is the reference wavelength,  $a_x(\lambda_r)$  is the value of the absorption coefficient at the reference wavelength, and  $S_x$  is the spectral slope parameter [1,5–10].

A result of the nearly exponential increase of the CDOM and NAP absorption coefficients with decreasing wavelength is that they contribute significantly to the non-water component of absorption,  $a_{nw}(\lambda) \equiv a_{pg}(\lambda) = a(\lambda) - a_w(\lambda)$ , in the UV and blue spectral regions [6,9,11]. The spectral characteristics of these absorption coefficients are often employed to understand physical and biogeochemical processes in the ocean. Specifically, CDOM absorption properties in the UV and blue spectral regions can be utilized to trace the transport of dissolved organic carbon [12,13], as well as monitor changes and identify the evolution of water masses in aquatic environments [14–18]. Furthermore, information about the characteristics of non-algal particulate matter, including measurements of  $a_d(\lambda)$ , has provided substantial insights about processes such as the remineralization of carbon and the vertical flux of particulate matter [19–22]. Another consequence directly resulting from the spectral characteristics of CDOM and NAP absorption is that variations in their properties, such as their concentration and chemical composition, have the potential to dramatically impact UV and VIS radiation fields in marine environments. Furthermore, exposure of CDOM itself to UV leads to its photodegradation over time and further alters the light environment [23–26]. Variations in the UV radiation field can impact various photochemical reactions in water [25,27,28], and an increase in exposure of marine organisms to UV radiation is potentially harmful [29–32]. As a result, changes in CDOM or NAP concentrations that alter the UV radiation environment directly impact the structure of vertical distributions and migration patterns of organisms in marine ecosystems [33].

A specific objective of ocean color satellite missions is the derivation of absorption coefficients of seawater constituents in support of characterization of the concentration and composition of particulate and dissolved matter as well as light propagation in the upper ocean layer. A variety of inverse ocean color models relying on measurements of ocean remote-sensing reflectance  $R_{rs}(\lambda)$  in the VIS spectral bands have been developed over recent decades to aid in pursuing this objective [34]. One important class of semi-analytical inverse models aims at deriving the total IOPs of seawater from ocean reflectance measurements [35–42]. These models can be applied in conjunction with stand-alone independently-developed models for partitioning the total absorption coefficient,  $a(\lambda)$ , or its non-water component,  $a_{nw}(\lambda)$ , into the constituent absorption coefficients. Thus, in this approach the derivation of constituent absorption coefficients is based on the sequential application of multiple algorithms operating in a stepwise fashion. There are four main types of absorption partitioning models: first, models that partition  $a_{nw}(\lambda)$  into the three constituent absorption coefficients,  $a_{ph}(\lambda)$ ,  $a_d(\lambda)$ , and  $a_g(\lambda)$  [43–46], second, models that partition  $a_{nw}(\lambda)$  into  $a_{ph}(\lambda)$  and  $a_{dg}(\lambda)$  [47,48], third, models that partition  $a_{dg}(\lambda)$  into  $a_g(\lambda)$  and  $a_d(\lambda)$  [46,49], and finally, models that partition  $a_p(\lambda)$  into  $a_{ph}(\lambda)$  and  $a_d(\lambda)$  [50–54]. Most partitioning models utilize assumptions about model outputs, most notably the spectral shapes of constituent absorption coefficients, which can substantially limit the model’s applicability and performance across diverse aquatic environments. To alleviate these limitations some absorption partitioning models have been designed with a purpose to avoid or include only weakly restrictive assumptions about the spectral shapes of output absorption coefficients [46,47,49,54]. With regard to the spectral domain of applicability, the existing absorption partitioning models have been, however, limited to the VIS range.

Recently, studies have also been conducted with the aim to estimate  $a_g(\lambda)$  in the UV either directly from measurements of  $R_{rs}(\lambda)$  in the VIS using a multivariate statistical approach [55], or from a single-wavelength estimate of  $a_{dg}(443)$  derived from a semi-analytical inverse reflectance model [35,38], which also operates in the VIS range [56]. The estimation of  $a_{dg}(\lambda)$  and  $a_{ph}(\lambda)$  in the near-UV at a single wavelength of 380 nm was also explored with a modified semi-analytical inverse reflectance model [37], which requires the input of  $R_{rs}(\lambda)$  at 380 nm rather than  $R_{rs}(\lambda)$  just from the VIS range [57]. While these studies reflect a growing interest to estimate the constituent absorption coefficients from optical measurements in the VIS range, they also reveal a need for further enhancement of these capabilities. In particular, there is a need to address the estimation of both non-phytoplankton components of absorption, i.e., the  $a_d(\lambda)$  and  $a_g(\lambda)$  coefficients, in the UV from measurements in the VIS.

Given the demonstrated potential of absorption partitioning models to separate  $a_g(\lambda)$  and  $a_d(\lambda)$  components from  $a_{dg}(\lambda)$  in the VIS [46,49], the goal of this study is to determine the optimal extrapolation method to extend these spectra from the VIS to the near-UV. The portion of near-UV that is considered in this study covers the wavelength range of 350 to 400 nm. Our focus is on this near-UV range rather than a wider range including shorter UV wavelengths primarily because of limitations and increased uncertainties with decreasing UV wavelength in the current-state-of-the-art methodology for measuring the spectral particulate absorption coefficient  $a_p(\lambda)$  including its  $a_d(\lambda)$  and  $a_{ph}(\lambda)$  components [58,59,60]. In addition, the new Ocean Color Instrument (OCI) to be deployed on the upcoming NASA Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) satellite mission will have the capabilities to make measurements of  $R_{rs}(\lambda)$  in this near-UV portion of the spectrum [61]. At present, however, the capabilities and performance of inverse optical models that utilize near-UV measurements, in particular the inverse reflectance and absorption partitioning

models, are not yet proven. Therefore, while the extrapolation model presented in this study is expected to have an immediate applicability to measurements and optical models operating in the VIS, it can also provide a valuable tool for comparison and assessment of forthcoming optical models that utilize near-UV measurements.

For formulation of the extrapolation model, we assembled a development dataset of field measurements of  $a_g(\lambda)$  and  $a_d(\lambda)$ , and hence their sum  $a_{dg}(\lambda)$ , from a diverse range of marine optical environments. We also assembled an independent dataset of field measurements of  $a_g(\lambda)$  and  $a_d(\lambda)$  for the purpose of model validation. The absorption coefficients in both these datasets were collected over the spectral range from the UV through near-IR with high spectral sampling interval and they satisfied a set of method-related and data quality criteria. One of the key method-related criteria used in this study is the inclusion of  $a_d(\lambda)$  data that were obtained solely from spectrophotometric measurements with the inside integrating-sphere filter-pad technique, which reduces the risk of biased measurements, including the biased spectral shape, compared to more traditional transmittance filter-pad technique [59,60]. Several approaches to extrapolation into the near-UV based on absorption data in the VIS were tested, and the formulation, performance, and validation of the optimal model are described.

## **1.2 Methods**

### **1.2.1 Assembly of the development dataset of absorption coefficients**

A development dataset of field measurements containing spectra of constituent absorption coefficients was assembled for the formulation and evaluation of different extrapolation methods to extend  $a_g(\lambda)$ ,  $a_d(\lambda)$ , and  $a_{dg}(\lambda)$  from the VIS into the near-UV. The outcomes of this study aim to extend existing models to the UV, which commonly work to derive absorption coefficients from satellite data, so the assembled dataset includes only near-surface water samples. Surface data are

defined as a sampling depth  $\leq 5\text{m}$  if the water depth is  $< 200\text{m}$  or a sampling depth  $\leq 15\text{m}$  if the water depth is  $\geq 200\text{ m}$ .

Multiple data sources were initially considered, but only those which contain data that pass basic methodology-related inclusion criteria were used to create a preliminary dataset for further quality control and processing. All measurements of constituent absorption coefficients considered in this study were made on discrete water samples using state-of-the-art techniques providing data at high spectral sampling intervals over a wide spectral region. First, measurements of spectral absorption coefficients  $a_g(\lambda)$  and  $a_p(\lambda)$  including the non-algal particulate component  $a_d(\lambda)$  were required to span the UV-VIS range, at a minimum from 350 to 700 nm for  $a_g(\lambda)$  and 350 to 800 nm for  $a_p(\lambda)$  and  $a_d(\lambda)$ , with a 1 nm sampling interval. Second, the spectra of  $a_g(\lambda)$  were measured either with the spectrophotometric method using a 10-cm path length cuvette or a long path length liquid waveguide capillary cell following standard sampling and measurement protocols [62,63]. Finally, a methodological prerequisite for inclusion of  $a_p(\lambda)$  and  $a_d(\lambda)$  data was that they were obtained with the spectrophotometric filter-pad method using the inside integrating-sphere configuration of measurement, which provides high-quality results owing largely to negligible or very small artifacts caused by light scattering during measurements of both sample filters and blank filters [59,60,64,65]. The  $a_d(\lambda)$  spectra were measured on the same sample filters as the  $a_p(\lambda)$  spectra after subjecting the filters to methanol treatment which extracts pigments present mostly in phytoplankton [2]. It must be noted that most historical data of  $a_p(\lambda)$  and  $a_d(\lambda)$  were measured with the filter-pad method using a transmittance configuration of measurement, which can result in significant error in both the magnitude and spectral shape of measured absorption [59,60]. Thus, measurements with the transmittance method were not included in this study.

The preliminary dataset of absorption coefficients that passed these basic methodology-related inclusion criteria consisted of 1610 unique water samples from 18 oceanographic experiments including 42 distinct cruises that collected 1493 spectra of  $a_g(\lambda)$  and 722 spectra of  $a_p(\lambda)$  and  $a_d(\lambda)$ . We note that this dataset also includes the spectra of  $a_{ph}(\lambda)$  derived as a difference between the measured  $a_p(\lambda)$  and  $a_d(\lambda)$ , however, our interest in this study is focused on the  $a_d(\lambda)$  component of particulate absorption. Additional processing was applied to the  $a_g(\lambda)$  spectra within the preliminary dataset. Most  $a_g(\lambda)$  spectra within the preliminary dataset had low signal-to-noise ratios in the green-to-red spectral region and were corrected by extrapolating the measurements with higher signal in the short-wavelength portion of the VIS spectrum to longer wavelengths. The extrapolation to the long-wavelength portion of the spectrum utilized a linear fit to  $\log_e$ -transformed  $a_g(\lambda)$  values versus light wavelength within the 400 nm to  $\lambda_{cutoff}$  range where  $\lambda_{cutoff}$  is the cutoff wavelength at which  $a_g(\lambda_{cutoff})$  drops to instrumental noise level. This level was assumed to be  $0.03 \text{ m}^{-1}$  for the 10-cm cuvette spectrophotometric method and  $0.005 \text{ m}^{-1}$  for the liquid waveguide capillary cell and is consistent with previously reported values [56,63]. The long-wavelength extrapolations were generally applied at wavelengths greater than 500 nm. If the noise level for a given sampling method occurred below 440 nm the measured spectrum of  $a_g(\lambda)$  was excluded from our final dataset.

Non-algal particulate absorption spectra available in the preliminary dataset were also subjected to additional processing. To remove instrumental noise, all  $a_d(\lambda)$  spectra were smoothed across the entire spectral range with a 7 nm moving median window and then three consecutive 7 nm moving mean windows. This method of smoothing was found adequate to smooth out the potential instrumental noise present in the measured spectra without affecting the shape of the spectra which generally does not exhibit any sharp spectral features. Additionally, it was assumed

the phytoplankton contribution to  $a_p(\lambda)$  is zero in the near-IR region. According to this assumption, a spectrally constant offset was applied to measured  $a_d(\lambda)$  to ensure that the resultant  $a_d(\lambda)$  is equal to measured  $a_p(\lambda)$  in the near-IR. Specifically,  $a_p(\lambda)$  measurements were first smoothed from 720 nm to the long-wavelength end of the spectrum by applying the same smoothing method as applied to  $a_d(\lambda)$  data. The difference between the average  $a_p(\lambda)$  and average  $a_d(\lambda)$  within the 780–820 nm range, or from 780 nm to the last available wavelength if data were not available up to 820 nm, was determined and added back to the entire  $a_d(\lambda)$  spectrum to correct for any offset between the original measurements of  $a_p(\lambda)$  and  $a_d(\lambda)$  in the near-IR. After processing  $a_d(\lambda)$ , it was evident that some spectra contained absorption features in spectral regions corresponding to phytoplankton pigments resulting from an incomplete pigment extraction process. To address this potential issue, all  $a_d(\lambda)$  spectra were analyzed to quantify the height of the residual absorption peak in the spectral region between 650 and 700 nm by applying an exponential fit to  $a_d(\lambda)$  values at two discrete wavelengths of 650 and 700 nm. This curve was treated as the expected portion of the  $a_d(\lambda)$  spectrum in the 650–700 nm range which is unaffected by potentially incomplete extraction of phytoplankton pigments. If the maximum value of measured  $a_d(\lambda)$  between 650 and 700 nm was greater than 8% above its expected fit value, the  $a_d(\lambda)$  spectrum was removed from further analysis and excluded from the final dataset. It is described below that a necessary condition of our extrapolation model is the input of an  $a_{dg}(\lambda)$  spectrum, so from the final dataset we also omit  $a_d(\lambda)$  without a matching concurrent measurement of  $a_g(\lambda)$ .

A final quality control was then applied to the remaining  $a_g(\lambda)$  and  $a_d(\lambda)$  by inspecting each individual spectrum visually to identify spectra that were characterized by the presence of clearly erroneous features in the UV-VIS range. Typical visual manifestation of erroneous features included an unrealistic spectral shape and a low signal-to-noise ratio. The spectra exhibiting such

erroneous features were not included in the final dataset. The entire data quality control process, which consisted of the application of initial basic methodology-related criteria and several additional inclusion and exclusion criteria, allowed us to greatly minimize the risk of including measurements of  $a_g(\lambda)$  and  $a_d(\lambda)$  that are subject to potentially significant errors in the final dataset.

The final model development dataset of  $a_g(\lambda)$ ,  $a_d(\lambda)$ , and  $a_{dg}(\lambda)$  consists of 1294 unique  $a_g(\lambda)$  and 409  $a_d(\lambda)$  spectra with concurrent  $a_g(\lambda)$  measurements as described in Table 1.1. The subset of data with concurrent measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$  also includes the  $a_{dg}(\lambda)$  spectra obtained as a sum of  $a_d(\lambda)$  and  $a_g(\lambda)$ . Table 1.1 also provides a list of field experiments and cruises where these data were collected.

Figure 1.1 depicts the geographic locations of all oceanographic stations where samples of absorption measurements included within the final development dataset were collected. This final dataset represents a diverse set of marine environments ranging from ultraoligotrophic waters found across the subtropical gyre within the southeastern Pacific collected on the BIOSOPE cruise to turbid waters in coastal and Arctic regions. Most samples with concurrent measurements of  $a_g(\lambda)$  and  $a_d(\lambda)$  come from polar and coastal regions. The only cruise in open ocean waters where concurrent measurements of  $a_g(\lambda)$  and  $a_d(\lambda)$  were made and met all inclusion criteria of our final dataset is the CLIVAR P16S cruise in the southern Pacific.

The final absorption spectra of  $a_g(\lambda)$ ,  $a_d(\lambda)$ , and  $a_{dg}(\lambda)$  as well as their corresponding histograms at  $\lambda = 443$  nm are displayed in Fig. 2. The shapes of  $a_g(\lambda)$  spectra are generally close to linear behavior when depicted on the semi-logarithmic scale (Fig. 1.2(a)) which is consistent with the nearly exponential shape. The average exponential slope parameter describing the decrease of  $a_g(\lambda)$  with increasing wavelength calculated over the spectral range 350–500 nm for

the entire final dataset is  $0.0157 \text{ nm}^{-1}$  (standard deviation  $\text{SD} = 0.0023 \text{ nm}^{-1}$ ) which is in general agreement with the reported values in the literature [9]. Compared to  $a_g(\lambda)$ , the spectra of  $a_d(\lambda)$  (Fig. 1.2(b)), and also  $a_{dg}(\lambda)$  (Fig. 1.2(c)), exhibit less linear behavior on the semi-logarithmic scale as their spectral slopes tend to decrease with increasing wavelength. The average spectral slope of  $a_d(\lambda)$  spectra for the entire final dataset is  $0.0076 \text{ nm}^{-1}$  ( $\text{SD} = 0.0013 \text{ nm}^{-1}$ ), which was calculated over the spectral range 380–730 nm, excluding 400–480 nm and 620–710 nm ranges to remove the potential methodological artifact associated with any remaining residual absorption by phytoplankton pigments and to maintain methodological consistency with previous work [66]. The average slope of  $a_d(\lambda)$  is smaller than values reported in other studies, for example  $0.011 \text{ nm}^{-1}$  [67,68],  $0.0123 \text{ nm}^{-1}$  [9], and  $0.0094 \text{ nm}^{-1}$  [66]. Most previous measurements of  $a_p(\lambda)$  and  $a_d(\lambda)$  utilized the spectrophotometric filter-pad method in the transmittance configuration of measurement [66–68] and some measurements were also made in the transmittance-reflectance configuration [9]. The transmittance method typically utilizes a null-point correction which assumes no particulate absorption in the near-IR. Following this assumption, the null-point correction typically consists of subtracting the average value of measured  $a_d(\lambda)$  within some portion of the near-IR from all spectral values of  $a_d(\lambda)$ . We note that similar null-point correction is applied to measurements of  $a_p(\lambda)$  although this coefficient is not of direct interest to the present study. Importantly, it can be shown that the wavelength-independent null-point correction leads to an artificially increased spectral slope of  $a_p(\lambda)$  and  $a_d(\lambda)$  compared to the  $a_p(\lambda)$  and  $a_d(\lambda)$  that exhibit some real (non-zero) absorption in the near-IR. The inside integrating-sphere spectrophotometric configuration of filter-pad method allows to measure and retain the non-zero values of  $a_p(\lambda)$  and  $a_d(\lambda)$  in the near-IR if the examined sample exhibits such absorption feature, and the null-point correction is not applied. Because the particulate absorption measurements in our dataset were all

made with the superior methodology of the inside integrating-sphere spectrophotometric configuration, the differences in the average spectral slope of  $a_d(\lambda)$  between our dataset and previous studies appear to be related to an overestimation of this parameter due to null-point correction that was commonly used in the previous studies.

The  $a_g(\lambda)$  spectra in the final dataset span approximately three orders of magnitude at 350 nm and this dynamic range remains relatively the same across the spectrum to the wavelength of 700 nm. The  $a_g(443)$  histograms (Fig. 1.2(d)) display a bimodal distribution that is associated with the large number of samples collected in either the Pacific or Atlantic Ocean basins. The samples collected from the Pacific Ocean typically have smaller magnitudes of  $a_g(443)$ , whereas  $a_g(443)$  data from the Atlantic Ocean more often originate from coastal waters and have larger  $a_g(443)$  values. Samples of  $a_d(\lambda)$  spectra span approximately four orders of magnitude at 350 nm and this dynamic range remains consistent across the spectrum. In contrast to  $a_g(443)$ , the histograms for both  $a_d(443)$  and  $a_{dg}(443)$  do not exhibit a clear bimodal shape (Fig. 1.2(e),(f)), and the majority of data come from experiments conducted in coastal Atlantic waters.

The formulation of the extrapolation model presented hereafter implements the complete development dataset of available near-surface water samples. This dataset was not partitioned into two independent subsets of data for separate purposes of model formulation and validation because we did not want to compromise on the representativeness of the model development dataset across the entire dynamic ranges of  $a_g(\lambda)$ ,  $a_d(\lambda)$ , and  $a_{dg}(\lambda)$ . In particular, partitioning of the development dataset into two subsets of data would result not only in reduction of total amount of data in the development dataset but, more importantly, would reduce the amount of data in portions of dynamic range where there are relatively few data, most notably in the ranges of relatively low and high values of  $a_d(\lambda)$  and  $a_{dg}(\lambda)$  (Fig. 1.2(e),(f)). This could adversely impact the formulation

of the extrapolation model and its performance over the entire investigated dynamic range of absorption coefficients. Therefore, for the purpose of validation of the extrapolation model with independent data we compiled a separate dataset as described below.

### **1.2.2 Assembly of an independent validation dataset of absorption coefficients**

An independent dataset of measured absorption spectra was compiled to validate the extrapolation model. The same sources of data used to create the model development dataset were employed in the assembly of the independent validation dataset. The validation dataset includes sub-surface measurements on samples collected between depths of 15 and 50 m, which satisfied the same method-related and data quality criteria as the model development dataset described in Section 2.1. This approach to validation with sub-surface (but not very deep) data is reasonable and appears as the only approach available at the present time which satisfies the method-related inclusion criteria used in the development dataset, especially in view of the overall limited availability of particulate absorption measurements with the inside integrating-sphere filter-pad technique. In the future, as more data with prerequisite methodology are collected, the validation analysis can be extended by including near-surface data. The model formulation itself can also be refined in the future with more data, if deemed appropriate.

The assembled validation dataset consists of 149 concurrent measurements of  $a_g(\lambda)$  and  $a_d(\lambda)$  from a variety of marine environments. We inspected all spectra individually within the validation dataset and found the general behavior and shapes of the spectra are consistent with those in the development dataset. However, we also observed that the validation dataset exhibits a narrower dynamic range of magnitudes of all three absorption coefficients compared to the development dataset. Nevertheless, given that the dynamic ranges overlap significantly between the validation and development datasets including the range of relatively low absorption

magnitudes, the validation dataset provides a suitable basis for assessing the performance of the extrapolation model with independent data.

### **1.2.3 Methods for extrapolating absorption coefficients from the VIS into the UV**

Although the spectral shape of non-phytoplankton constituent absorption coefficients,  $a_g(\lambda)$ ,  $a_d(\lambda)$ , and  $a_{dg}(\lambda)$ , have often been approximated with an exponential function of light wavelength, there is experimental evidence that the spectral shapes generally do not maintain a fixed exponential slope across the broad range of wavelengths from the UV through the VIS spectral region [9,69–73]. To address this challenge, functions other than a single exponential function of wavelength were examined in the past as potential descriptors of  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  spectra, such as power function, double exponential function, and stretched exponential function [69,73].

In this study, five different extrapolation functions were investigated to extend  $a_g(\lambda)$ ,  $a_d(\lambda)$ , and  $a_{dg}(\lambda)$  spectra from the VIS into the near-UV region between 350 and 400 nm: (i) an exponential function calculated by linear regression applied to  $\log_e$ -transformed absorption data vs. wavelength (ELR), (ii) an exponential function calculated by non-linear regression to absorption data vs. wavelength (ENLR), (iii) a stretched exponential function calculated by non-linear regression to absorption data vs. wavelength (SENLR), (iv) a power function calculated by linear regression to  $\log_{10}$ -transformed absorption data vs.  $\log_{10}$ -transformed wavelength (PLR), and (v) a power function calculated by non-linear regression to absorption data vs. wavelength (PNLR). Table 1.2 summarizes the different regression models.

We examined the blue spectral range from 400 to 500 nm as a basis for fitting the regression models to absorption measurements in the VIS region. This choice appears justifiable as the spectral slope of non-phytoplankton absorption coefficients in the near-UV is most similar to the spectral slope in the blue spectral region, for example for  $a_g(\lambda)$  [72]. Multiple spectral windows

with variable width from within the 400–500 nm range were tested. For each tested spectral window, we fixed the shortest (start) wavelength at 400 nm to ensure that measured data nearest the extrapolation UV region were always included in the formulation of each regression. The longest (stop) wavelength,  $\lambda_{\text{stop}}$ , of the spectral windows was allowed to vary in 1 nm increments from 405 nm (i.e., the narrowest window of 400–405 nm) to 500 nm (i.e., the widest window of 400–500 nm). We did not examine the regression windows that extend beyond 500 nm because such broader spectral windows are expected to have limited capability to predict the spectral shape in the UV. In addition, as the magnitudes of  $a_g(\lambda)$ ,  $a_d(\lambda)$ , and  $a_{dg}(\lambda)$  decrease with wavelength beyond 500 nm, the measurements at longer wavelengths are at increased risk of being affected by low signal-to-noise ratio or dropping to the instrumentation sensitivity limit, in particular for low CDOM absorption in clear waters. To maintain spectral continuity across the UV-VIS boundary, the absorption values calculated from the regression were adjusted such that the predicted value at 400 nm is equal to the measured value at 400 nm and the spectral slope obtained from the regression is preserved. Standard MATLAB fitting routines were applied to determine the regression parameters for all linear and non-linear fits, where non-linear regressions were determined using the Levenberg-Marquardt algorithm [74,75].

It is important to note that differences between the ELR and ENLR or between PLR and PNLR regression models arise from different weighting of each measurement during the determination of each regression fit. Whereas linear regressions involving the use of  $\log_e$ -transformed or  $\log_{10}$ -transformed data weigh each measurement within the regression spectral window equally, nonlinear regressions give more weight to higher-signal measurements typically found at shorter wavelengths within the regression window, so in our case the wavelengths nearest 400 nm. A linear regression model with an equal weighting of measurements can have drawbacks

if applied over a broad spectral range extending to long-wavelength portion of the VIS because it would include data with low signal-to-noise ratio, and would be subject to higher uncertainties [8,14,69]. In our analysis, however, this potential disadvantage is circumvented by using a regression window extending to at most 500 nm, which generally ensures that spectral measurements of constituent absorption coefficients maintain a high signal-to-noise ratio. Figure 1.3 compares two example measurements of  $a_g(\lambda)$  from differing marine optical environments with results from all five regression models calculated using a regression window of 400–500 nm. These examples demonstrate the differences between the fitted spectra by each type of regression model and also highlight a tendency for increased discrepancy between the fitted and measured spectral values as the wavelength decreases in the near-UV range of 350 to 400 nm, which is of particular interest to this study in the context of determining the optimal extrapolation model from the VIS into the UV.

The analysis aiming at determination of the optimal regression model and spectral window for the purpose of extrapolation from the blue to the near-UV spectral region was focused on data of  $a_g(\lambda)$  and  $a_{dg}(\lambda)$ . This approach is justified by existing experimental evidence that the absorption spectra of CDOM,  $a_g(\lambda)$ , are less likely to exhibit significant departures from a single-slope exponential-like behavior than the spectra of non-algal particles,  $a_d(\lambda)$ , especially in the blue spectral region [9]. The analysis of our dataset supported this conjecture. Thus, it can be expected that fitting the  $a_g(\lambda)$  measurements with a regression model in the blue spectral region for the purpose of subsequent extrapolation to the near-UV region will generally lead to better results compared to a similar approach applied to  $a_d(\lambda)$ . In addition, it is important to note that in the UV and blue spectral regions the magnitude of  $a_g(\lambda)$  is often larger than the magnitude of  $a_d(\lambda)$  across different marine environments, which implies that  $a_{dg}(\lambda)$  is often dominated by the contribution of

CDOM [9,11]. This tendency is also observed in our dataset. Specifically, for the dataset of 409 concurrent measurements of  $a_g(\lambda)$  and  $a_d(\lambda)$ , the average fractional contribution of  $a_g(\lambda)$  to  $a_{dg}(\lambda)$  at the example wavelength of 400 nm is 0.778 (SD = 0.165). These results suggest that, after  $a_g(\lambda)$ , the spectral measurements of  $a_{dg}(\lambda)$  appear to be the second-best candidate for the application of a regression model in the blue spectral region with a purpose of subsequent extrapolation to near-UV. Thus, the third absorption coefficient associated with non-algal particles,  $a_d(\lambda)$ , does not require fitting a regression model to measured data in the blue and extrapolation to the near-UV because it can be derived as a difference between  $a_{dg}(\lambda)$  and  $a_g(\lambda)$ . In summary, for the formulation of the optimal extrapolation model, we evaluated different extrapolation methods (i.e., different regression models and spectral windows in the blue spectral region) by assessing the degree of agreement between extrapolated and measured values of  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  in the near-UV. For  $a_d(\lambda)$ , analogous agreement in the near-UV was evaluated using the model-derived values of  $a_d(\lambda)$  obtained as a difference between the extrapolated  $a_{dg}(\lambda)$  and  $a_g(\lambda)$ .

#### 1.2.4 Statistical evaluation of extrapolation models

A suite of statistical metrics between predicted values from the model,  $P_i$ , and observed values,  $O_i$  (where the integer index  $i$  varies from 1 to  $N$  and  $N$  is the number of data points), for a given spectral value of absorption coefficient was used to assess the performance of each extrapolation method within the 350–400 nm spectral region. This analysis was made with the model development dataset. The statistical metrics include the nondimensional median ratio,  $MdR$ , defined as the median of the ratio between  $P_i$  and  $O_i$ , the median bias,  $MB$ , defined as the median difference between  $P_i$  and  $O_i$ , the median absolute percent difference,  $MdAPD = \text{median}(|P_i - O_i| / O_i) \times 100\%$ , and the root-mean-square deviation,  $RMSD = (N^{-1} \sum (P_i - O_i)^2)^{0.5}$ . These metrics provide information on bias and random error statistics. Additionally, Model II linear regression (a reduced major axis method) was applied to scatterplots of  $\log_{10}$ -transformed predicted versus

measured values of absorption coefficients in the UV range to further evaluate each extrapolation method. We report on the values of two regression parameters,  $A$  which is the slope, and  $B$  which is equal to  $10^I$  where  $I$  is the  $y$ -intercept of linear regression of  $\log_{10}$ -transformed variables. The scenario with both  $A$  and  $B$  equal to 1 is most desirable in a sense that it is indicative of no biasing effects across the dynamic range of the examined absorption coefficient. The Pearson correlation coefficient,  $r$ , between the  $\log_{10}$ -transformed predicted and measured values was also calculated. Hereafter, we typically illustrate the statistical metrics of  $MdR$  and  $MdAPD$  spectrally across the near-UV spectral region to allow for a comparison between each of these statistics when evaluating the output of different extrapolation models. Ultimately, the selection of the optimal extrapolation model was based on comparisons of multiple statistical metrics to establish the best regression model and best regression spectral window in the blue spectral region.

### 1.3 Results and Discussion

#### 1.3.1 Determination and assessment of the optimal extrapolation method

Figure 1.4 depicts the aggregate error statistics  $MdR$  and  $MdAPD$  for  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  at the example wavelength of 350 nm obtained from the extrapolation of data in the blue spectral region using five regression models. These error statistics were calculated using the model development dataset and are reported at 350 nm where the largest uncertainties are expected because this wavelength is furthest away from the regression region, which has been confirmed by our analysis shown below. These results are plotted as a function of  $\lambda_{stop}$ . We recall that  $\lambda_{stop}$  defines the regression spectral window which is set to start at 400 nm and end at  $\lambda_{stop}$  that was allowed to vary between 405 and 500 nm in these calculations. Thus, Fig. 1.4 summarizes the consequences associated with the choice of both the type of regression model and the regression spectral window on the quality of the extrapolation of  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  from the blue to the near-UV.

The spectral error statistics in Fig. 1.4 depict the characteristics in the uncertainty of UV absorption estimates obtained from exponential and power function regressions. The three exponential regression models, ELR, ENLR, and SENLR, have similar performance as indicated by similar curves of  $MdR$  and  $MdAPD$  across the range of  $\lambda_{stop}$ . The same is true for the power function regressions, PLR and PNLR, whose spectral error statistics are very similar. However, Fig. 1.4 also shows there are substantial differences in the magnitude and pattern of error statistics between the two groups of regression functions. Other statistical metrics also exhibited differences (not shown) which further supports the conjecture that the optimal regression window (i.e., the optimal  $\lambda_{stop}$ ) is different for each group of functions. Specifically, the patterns of error statistics in Fig. 1.4 suggest that for all three exponential functions the 400–450 nm range can be used as the optimal regression spectral window for the UV determinations of both  $a_g(\lambda)$  and  $a_{dg}(\lambda)$ . This is demonstrated by relatively good balance in the optimization of both statistical parameters displayed in Fig. 1.4, which is achieved in the vicinity of  $\lambda_{stop} = 450$  nm in the case of estimation of both  $a_g(350)$  and  $a_{dg}(350)$ . Specifically, if  $\lambda_{stop} = 450$  nm  $MdR$  for estimates of  $a_g(350)$  and  $a_{dg}(350)$  assumes the value closest or nearly closest to 1 (Fig. 1.4(a),(c)) and  $MdAPD$  for estimates of  $a_g(350)$  and  $a_{dg}(350)$  assumes the lowest or nearly lowest value across the range of  $\lambda_{stop}$  (Fig. 1.4(b),(d)). These  $MdR$  values are about 0.98 and the  $MdAPD$  values are about 4.5% to 6%. We note that although the  $MdR$  and  $MdAPD$  statistics in Fig. 1.4 can maintain a relatively good level at other stop wavelengths, especially for  $\lambda_{stop} > 450$  nm, the selection of  $\lambda_{stop} = 450$  nm has been made by consideration of the combination of patterns of the statistical parameters of estimates of both  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  not only at 350 nm but also at other near-UV wavelengths, which is addressed in more detail below. In addition, the selection of  $\lambda_{stop} = 450$  nm also included the consideration of statistics not shown in Fig. 1.4, such as the linear regression of predicted vs. measured

absorption coefficients at near-UV wavelengths and the correlation between the spectral slopes of measured absorption coefficients in the near-UV and the spectral slopes calculated over different spectral subranges from the 400–500 nm range.

The results in Fig. 1.4 also indicate that for the power functions, PLR and PNLR, the optimal regression window is much narrower compared to the optimal regression window for the exponential functions. Specifically, the optimal combination of  $MdR$  (Fig. 1.4(a),(c)) and  $MdAPD$  (Fig. 1.4(b),(d)) values of estimates of  $a_g(350)$  and  $a_{dg}(350)$  is observed when the PLR or PNLR regression is calculated over the spectral window from 400 nm to  $\lambda_{stop} \cong 425$  nm. In addition, the  $MdR$  and  $MdAPD$  values exhibit significant sensitivity to the choice of  $\lambda_{stop}$ , including the wavelengths in the vicinity of  $\lambda_{stop} = 425$  nm. We disregarded the power function regression models in further analysis because of such sensitivity and the relatively narrow optimal regression window of only 25 nm which implies a relatively small number of absorption data within this window. These features of the power function models can be disadvantageous for the model's predictive accuracy. Therefore, we chose to consider the exponential regression model as a means to estimate  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  in the near-UV from data in the blue spectral region.

As illustrated in Fig. 1.4, the estimation of  $a_g(\lambda)$  or  $a_{dg}(\lambda)$  in the near-UV is similar regardless of which of the three exponential regression models, ELR, ENLR, or SENLR, is used. In particular, this result holds precisely for the selected optimal regression window of 400–450 nm, i.e., when  $\lambda_{stop} = 450$  nm. Therefore, for further analysis we chose the simplest version of the exponential regression model, ELR (Table 1.2). We recall that the ELR involves the use of ordinary linear regression with loge-transformed values of spectral absorption coefficients, so each spectral value within the regression window of 400–450 nm has the same statistical weight in the calculation of the fitted function.

Figure 1.5 provides additional support for selection of the 400–450 nm spectral range as the optimal regression window. This figure depicts the error statistics,  $MdR$  and  $MdAPD$ , plotted as a function of  $\lambda_{stop}$  for estimates of  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  at five example near-UV wavelengths ranging from 350 to 390 nm in 10 nm increments. These absorption estimates were obtained by extrapolation of the exponential regression model, ELR, that was calculated using different regression windows as indicated by varying  $\lambda_{stop}$ . When  $\lambda_{stop}$  is around 450 nm, the  $MdR$  values are close to 1 for all near-UV wavelengths and assume values either slightly below 1 (i.e., slightly negative bias) for wavelengths closer to 350 nm or above 1 (slightly positive bias) for wavelengths closer to 400 nm (Fig. 1.5(a),(c)). When  $\lambda_{stop}$  decreases from 450 nm towards 400 nm there is a tendency for negative bias in the  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  estimates at all near-UV wavelengths. When  $\lambda_{stop}$  increases from the 450–460 nm range towards 500 nm, there is also a tendency for negative bias in the  $a_{dg}(\lambda)$  estimates at most near-UV wavelengths (Fig. 1.5(c)), although no such pattern is observed for  $a_g(\lambda)$  (Fig. 1.5(a)). Thus,  $\lambda_{stop} = 450$  nm appears to represent good choice from the standpoint of  $MdR$  statistic for both the  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  estimates within the entire near-UV region. In addition, when  $\lambda_{stop} = 450$  nm the  $MdAPD$  statistic remains low (below 6%) for both  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  estimates at all near-UV wavelengths, although slightly lower values of  $MdAPD$  are observed at shorter  $\lambda_{stop}$  (Fig. 1.5(b),(d)). On the other hand,  $MdAPD$  tends to increase when  $\lambda_{stop}$  continues to increase beyond 450 nm. The tradeoffs between the error statistics highlighted in Fig. 1.5 justify the selection of  $\lambda_{stop} = 450$  nm, and the complete results presented in both Fig. 1.4 and Fig. 1.5 support the conjecture that the optimal extrapolation method to estimate  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  in the near-UV from the VIS can be based on the use of the ELR model within the spectral regression window of 400–450 nm.

The assessment of the optimal extrapolation method for estimating  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  in the near-UV is provided in Fig. 1.6 which depicts the correlation plots between the extrapolated and measured values of the absorption coefficients at 350 and 380 nm, as obtained from the analysis of the model development dataset. Similar correlation plots are also illustrated for the  $a_d(\lambda)$  coefficient where the predicted values of  $a_d(\lambda)$  were calculated as a difference between the extrapolated values of  $a_{dg}(\lambda)$  and  $a_g(\lambda)$ . The correlation plots generally demonstrate a very good agreement between extrapolated and measured values of  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  as the data align well with the 1:1 line and show little scatter about this line (Fig. 1.6(a),(b),(d),(e)). This is also supported by the closeness of the 1:1 line and the best-fit lines from the Model II linear regression applied to  $\log_{10}$ -transformed extrapolated vs. measured data of  $a_g(\lambda)$  or  $a_{dg}(\lambda)$ . However, despite such overall good agreement, the extrapolation method tends to underestimate  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  at 350 nm when the magnitude of these coefficients is low (Fig. 1.6(a),(b)). This tendency is not clearly observed at 380 nm (Fig. 1.6(d),(e)) indicating the potential effects of underestimation decrease with increasing wavelength between 350 nm and 400 nm. With regard to  $a_d(\lambda)$ , the correlation plots suggest that while the extrapolation method performs generally well, there is a persistent underestimation of  $a_d(\lambda)$  at 350 nm across the entire range of  $a_d(\lambda)$  (Fig. 1.6(c)). Although the tendency for underestimation still appears to be present in the data for 380 nm, it is greatly reduced (Fig. 1.6(f)). In summary, the results presented in Fig. 1.6 suggest the need to develop a correction to minimize or eliminate the potential presence of bias in the predictions of absorption coefficients at near-UV wavelengths.

### **1.3.2 Development of correction functions and assessment of the final extrapolation model**

We evaluated several options to correct the predicted absorption coefficients in the near-UV and determined that the most effective approach involves the use of two separate correction

functions: one for correcting the extrapolated  $a_g(\lambda)$  and the other for correcting  $a_d(\lambda)$  that is initially calculated as a difference between the extrapolated  $a_{dg}(\lambda)$  and  $a_g(\lambda)$ . Note that in this correction approach the extrapolated  $a_{dg}(\lambda)$  is not subject to any direct correction. Instead, the final  $a_{dg}(\lambda)$  is obtained as a sum of corrected  $a_g(\lambda)$  and corrected  $a_d(\lambda)$  which ensures the closure between the three absorption coefficients. Details of the correction functions are described below.

The development of the correction of  $a_g(\lambda)$  was based on the analysis of the relationship between the ratio of measured to extrapolated  $a_g(\lambda)$  as a function of the magnitude of extrapolated  $a_g(\lambda)$  at each wavelength within the near-UV region, i.e., from 350 to 399 nm in 1 nm increments. This analysis provided a means to quantify the potential bias in extrapolated  $a_g(\lambda)$  as a function of  $a_g(\lambda)$  across the near-UV spectral region. Specifically, for any given near-UV wavelength, if the ratio of measured to extrapolated  $a_g(\lambda)$  shows a clear tendency to be greater than 1 for a given value of extrapolated  $a_g(\lambda)$ , it means that the extrapolated  $a_g(\lambda)$  tends to be negatively biased. On the other hand, if the ratio tends to be smaller than 1 the extrapolated  $a_g(\lambda)$  is positively biased. Values of the ratio that are evenly distributed around 1 indicate no bias. Examples of this analysis for two wavelengths, 350 nm and 380 nm, are presented in Fig. 1.7. The results for 350 nm are selected in this example illustration because at this wavelength we identified the presence of largest negative bias in extrapolated  $a_g(\lambda)$  at low values of  $a_g(\lambda)$  (Fig. 1.7(a)). Figure 1.7(a) displays all 1294 data points from our model development dataset as well as a trend in these data represented by the median values of the ratio of measured to extrapolated  $a_g(350)$  (red crosses). These median values were calculated within 19 successive bins of extrapolated  $a_g(350)$  data across the entire range of  $a_g(350)$ . As seen, the median values increase significantly above 1 when the extrapolated  $a_g(350)$  has low magnitude, specifically when  $a_g(350)$  is less than about  $0.1 \text{ m}^{-1}$ . Thus, in this range the extrapolated  $a_g(350)$  tends to be underestimated. The other important result in Fig. 1.7(a)

indicates that as the magnitude of extrapolated  $a_g(350)$  increases beyond approximately  $0.2\text{m}^{-1}$ , the bias becomes very small or negligible because the median ratio of measured to extrapolated  $a_g(350)$  remains very close to 1. Similar patterns were observed at wavelengths longer than 350 nm, however, the extent of negative bias at low values of extrapolated  $a_g(\lambda)$  was found to decrease gradually with increasing wavelength. Eventually, at a wavelength of about 380 nm and beyond towards 400 nm, no significant bias was identified in extrapolated  $a_g(\lambda)$  across the entire dynamic range of  $a_g(\lambda)$ . This result is illustrated in Fig. 1.7(b) for  $\lambda = 380$  nm where the median ratio of measured to extrapolated  $a_g(380)$  remains close to 1 across the range of  $a_g(380)$ . Therefore, we formulated a correction function that applies the correction to extrapolated  $a_g(\lambda)$  in the wavelength range between 350 nm and 380 nm, and no correction is applied at wavelengths  $\geq 380$  nm.

To establish the correction function for  $a_g(\lambda)$ , several formulas were tested as potential candidates to provide the best fit to the relationship between the individual points of the measured to extrapolated  $a_g(\lambda)$  as a function of extrapolated  $a_g(\lambda)$ . As a result of this analysis, the final best-fit function for any given wavelength  $\lambda$  is defined as:

$$E_g(\lambda) = 1 + \alpha(\lambda)^{X(\lambda) - \beta(\lambda)} \quad (1.2)$$

where  $X(\lambda)$  is the base 10 logarithm of extrapolated  $a_g(\lambda)$  at a given  $\lambda$ , and  $\alpha(\lambda)$  and  $\beta(\lambda)$  are the parameters derived from a non-linear regression analysis of data at a given  $\lambda$ . The best-fit curves of  $E_g(\lambda)$  for  $\lambda = 350$  nm and 380 nm are shown in Fig. 1.7 as blue dashed lines. The  $E_g(350)$  curve captures the patterns of individual data points and the median values which increase at low  $a_g(350)$  (Fig. 1.7(a)). In contrast, the  $E_g(380)$  curve is nearly flat with the values very close to 1 across the range of  $a_g(380)$  (Fig. 1.7(b)). A minor exception is observed at extremely low  $a_g(380)$  where  $E_g(380)$  increases above 1. However, this feature is considered insignificant given the actual patterns of individual data points and median values in the range of low  $a_g(380)$ .

To account for the weakening of the dependence of  $E_g(\lambda)$  on the magnitude of extrapolated  $a_g(\lambda)$  as the near-UV wavelength increases from 350 to 380 nm, the weighting factors,  $w_1(\lambda)$  and  $w_2(\lambda)$ , are applied to obtain the final correction function,  $CF_g(\lambda)$ , as follows:

$$CF_g(\lambda) = w_1(\lambda)E_g(\lambda) + w_2(\lambda) \quad (1.3)$$

The first weight,  $w_1(\lambda)$ , is defined as:

$$w_1(\lambda) = \frac{1}{2} \cos\left[\frac{\pi}{30}(\lambda - 350)\right] + \frac{1}{2} \quad (1.4)$$

which has a value of 1 at 350 nm and decreases to 0 at 380 nm. The  $w_1(\lambda)$  values at all near-UV wavelengths longer than 380 nm are also set to zero. The second weight,  $w_2(\lambda)$ , is defined as  $1 - w_1(\lambda)$  such that the two weights always add to a value of 1. Thus, the weighting factors act to alter the influence of  $E_g(\lambda)$  within the correction function as the wavelength increases from 350 to 380 nm. Table 1.3 specifies the final values for  $\alpha(\lambda)$ ,  $\beta(\lambda)$ ,  $w_1(\lambda)$ , and  $w_2(\lambda)$  from 350 to 380 nm in 1 nm increments. To obtain the final corrected value of  $a_g(\lambda)$  at any given wavelength  $\lambda$ , the extrapolated value of  $a_g(\lambda)$  is multiplied by the correction function  $CF_g(\lambda)$  for a given extrapolated value of  $X(\lambda) \equiv \log_{10}[a_g(\lambda)]$  and values of  $\alpha(\lambda)$ ,  $\beta(\lambda)$ ,  $w_1(\lambda)$ , and  $w_2(\lambda)$ .

The correction for bias in  $a_d(\lambda)$  estimates is based on the analysis of data such as shown in Fig. 1.6(c),(f) but for all near-UV wavelengths. This analysis demonstrated that the predicted  $a_d(\lambda)$ , which is calculated as a difference between the extrapolated  $a_{dg}(\lambda)$  and  $a_g(\lambda)$ , tends to be biased low with the extent of bias decreasing as the wavelength increases from 350 to 400 nm. In addition, it was determined that for any given  $\lambda$  there was no need to introduce a correction for bias that is dependent on the magnitude of  $a_d(\lambda)$  because the underestimation of  $a_d(\lambda)$  persisted quite consistently across the entire range of  $a_d(\lambda)$ , which is particularly evident in the data at 350 nm (Fig. 1.6(c)). The formulation of the correction for  $a_d(\lambda)$  is therefore simpler than that for  $a_g(\lambda)$ . Specifically, this correction is based on the spectral values of  $MdR$  of predicted to measured  $a_d(\lambda)$  (Fig. 1.8) and the final corrected  $a_d(\lambda)$  at any given wavelength  $\lambda$  is obtained by dividing the

uncorrected  $a_d(\lambda)$  by the value of  $MdR$  at a given  $\lambda$ . The values of  $MdR$  from 350 to 399 nm in 1 nm increments are provided in Table 1.4. The value of  $MdR$  is the smallest and differs the most from 1 at 350 nm, so the correction for underestimation is largest at this wavelength. The  $MdR$  values increase as a function of wavelength and approach the value of 1 near 390 nm and beyond towards 400 nm (Fig. 1.8). Thus, in this long-wavelength portion of the near-UV region the correction of  $a_d(\lambda)$  is minimal. It is also to be noted that as a result of such correction based on the spectral values of  $MdR$ , the median aggregate bias of corrected  $a_d(\lambda)$  is forced to zero at all near-UV wavelengths.

A diagram that depicts the sequence of steps for the complete extrapolation model is presented in Fig. 1.9. In summary, in the first step the input data of spectral  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  in the VIS are assembled. If the input data are based on field measurements of  $a_g(\lambda)$  and  $a_d(\lambda)$ , as in the present study, the values of  $a_{dg}(\lambda)$  are calculated as a sum of  $a_g(\lambda)$  and  $a_d(\lambda)$ . However, it is to be noted that  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  can also originate from models such as absorption partitioning models where prior knowledge of  $a_d(\lambda)$  is not necessarily needed. In the second step, the input data of  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  are subject to the ELR regression which provides a basis for extrapolation from the VIS into the near-UV. This is then followed by calculation of  $a_d(\lambda)$  in the near-UV as a difference between the extrapolated  $a_{dg}(\lambda)$  and  $a_g(\lambda)$ . In the third step, two separate corrections are applied to minimize or eliminate bias in the near-UV estimates of  $a_g(\lambda)$  and  $a_d(\lambda)$  obtained in the second step. In the final step, the corrected values of  $a_g(\lambda)$  and  $a_d(\lambda)$  are summed to obtain  $a_{dg}(\lambda)$  in the near-UV region. This fourth step completes the extrapolation model by providing the final estimates of  $a_g(\lambda)$ ,  $a_d(\lambda)$ , and  $a_{dg}(\lambda)$  in the near-UV, which satisfy the closure equation  $a_{dg}(\lambda) = a_g(\lambda) + a_d(\lambda)$ .

An assessment of the final extrapolation model to obtain  $a_g(\lambda)$ ,  $a_d(\lambda)$ , and  $a_{dg}(\lambda)$  in the near-UV spectral region is presented in Table 1.5 and Fig. 1.10. These results were obtained from the

analysis of the model development dataset. The aggregate error statistics of estimates of the three absorption coefficients indicate an overall good performance of the model (Table 1.5). The aggregate bias of  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  estimates is very small or negligible across the near-UV region as indicated by  $MdB$  which is close to  $0 \text{ m}^{-1}$  and  $MdR$  which is very near 1 to within about 1%. Given the formulation of the correction applied to  $a_d(\lambda)$ , the bias in this absorption coefficient was completely removed at all near-UV wavelengths, so  $MdB$  and  $MdR$  values are exactly equal to  $0 \text{ m}^{-1}$  and 1, respectively. The  $MdAPD$  and  $RMSD$  statistics indicate that the largest uncertainty in estimates of  $a_g(\lambda)$ ,  $a_d(\lambda)$ , and  $a_{dg}(\lambda)$  occurs at 350 nm. For example, the  $MdAPD$  values for the estimates of  $a_g(350)$ ,  $a_d(350)$ , and  $a_{dg}(350)$  are 5.14%, 10.45%, and 4.54%, respectively (Table 1.5). This is expected as 350 nm is the furthest away from 400 nm where the transition occurs between the near-UV extrapolation range and the VIS range containing the input data used in the extrapolation model. Importantly, the  $MdAPD$  values are small across the near-UV region, typically below 5%, which further supports the overall good performance of the extrapolation model. In addition, it is notable that the error statistics of final corrected absorption coefficients obtained in step 4 of the model are improved compared to the statistics of initial estimates of absorption coefficients obtained in step 2 of the model. Consistent with the correction design and purposes, these improvements are most significant within the short-wavelength portion of the near-UV region. It is also notable that in the process of estimating  $a_g(\lambda)$ , the correction procedure in step 4 of the model brings in especially significant improvements at low values of  $a_g(\lambda)$ . This can be demonstrated by comparing the slope ( $A$ ) and intercept-related ( $B$ ) parameters of Model II linear regression of ( $\log_{10}$ -transformed) predicted vs. measured values of  $a_g(350)$  where the predicted values are obtained in step 2 (i.e., before correction) or step 4 (i.e., after correction) of the extrapolation model and the measured values are restricted to the low-magnitude range less than

$0.1 \text{ m}^{-1}$ . In this range we have 536 data points available for this regression analysis. In the case of  $a_g(350)$  predicted in step 2 of the model, the  $A$  and  $B$  values are 1.128 and 1.350, respectively. When the corrected values of  $a_g(350)$  from step 4 of the model are used in this analysis, the  $A$  and  $B$  parameters are closer to the most desirable value of 1, specifically 0.959 and 0.886, respectively.

Figure 1.10 is analogous to Fig. 1.6 but shows the correlation plots of the final corrected estimates of  $a_g(\lambda)$ ,  $a_{dg}(\lambda)$ , and  $a_d(\lambda)$  versus measured values of these coefficients at 350 nm and 380 nm. Compared to results in Fig. 1.6, Fig. 1.10(a),(b) demonstrates improved estimates of  $a_g(350)$  and  $a_{dg}(350)$ , especially for low-magnitude data points which are now distributed nearer and more evenly around the 1:1 line. Significant improvement is also seen for  $a_d(350)$  in Fig. 1.10(c) where the biasing tendency observed in Fig. 1.6 has been essentially removed. The differences between the correlation plots presented in Fig. 1.10 and Fig. 1.6 for  $\lambda = 380 \text{ nm}$  are naturally small because the correction at this wavelength either vanishes (for  $a_g$ ) or becomes small (for  $a_d$ ). In summary, the aggregate error statistics (Table 1.5) and correlation plots (Fig. 1.10) demonstrate the final extrapolation model involving the correction functions (Fig. 1.9) provides a means to estimate with satisfactory accuracy the three non-phytoplankton components of absorption,  $a_g(\lambda)$ ,  $a_d(\lambda)$ , and  $a_{dg}(\lambda)$  in the near-UV (350 to 400 nm) spectral region.

### **1.3.3 Application of the final extrapolation model using spectrally subsampled data in the VIS**

Although the extrapolation model described in previous sections was developed and assessed with hyperspectral absorption measurements available at 1 nm intervals, it is also of interest to evaluate the performance of this model using the input absorption data at lower spectral sampling intervals within the VIS region. Testing such scenarios can be relevant in the context of absorption coefficients derived from inverse optical models at certain wavelengths in the VIS which are available on satellite ocean color sensors. The past and current satellite sensors, such as

SeaWiFS, MODIS, VIIRS, MERIS, or OLCI, are multispectral and have only a few spectral bands in the blue spectral region which is utilized as a basis in our extrapolation model. In contrast, the Ocean Color Instrument (OCI) to be launched in 2024 on NASA's Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) satellite mission will provide a capability for high spectral sampling-interval measurements at 5 nm intervals from the near-UV (350 nm) through the VIS into the near-IR (890 nm).

We have evaluated the performance of our extrapolation model for a few specific spectral scenarios in which the 1 nm hyperspectral absorption measurements in our model development dataset were subsampled in the VIS spectral range to provide input data to the ELR regression component of the extrapolation model with lower spectral sampling interval. Below we present results for three scenarios of subsampled spectral data of  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  in the VIS range which serve as input to the model. The first scenario includes the input data in the 400–450 nm range at 5 nm intervals, which is consistent with the spectral characteristics of the PACE-OCI and represents rather minor degradation of spectral resolution compared to the original 1 nm data used in the formulation of the extrapolation model. The second scenario utilizes the input data only at two wavelengths, 412 and 443 nm, which correspond to spectral bands available on MODIS. Finally, the third scenario makes use of input data at three wavelengths from the blue spectral region which are available on MODIS, 412, 443, and 488 nm. Note, however, that in this case the wavelength of 488 nm is outside the 400–450 nm range which was used in the extrapolation model based on original 1 nm data. We also note that we tested the two remaining combinations of the blue spectral bands corresponding to those available with the MODIS sensor, but the absorption estimates in the near-UV were not superior to those obtained with the aforementioned two- or three-spectral band scenarios.

Figure 1.11 compares the error statistics for two scenarios of estimation of the three absorption coefficients in the near-UV region: one scenario utilized the input data with the original 1 nm interval in the ELR regression component of the extrapolation model, and the other scenario utilized the input data with a 5 nm interval. In both cases the absorption coefficients were estimated at 1 nm intervals within the near-UV region, so the values of error parameters in Fig. 1.11 are also plotted at 1 nm intervals. These results demonstrate that compared to the use of 1 nm input data, the use of 5 nm data within the 400–450 nm regression window yields very similar or nearly identical values of  $MdR$  and  $MdAPD$  across the near-UV region for all three absorption coefficients. This is indicative of essentially no degradation in performance of the extrapolation model when the spectral sampling interval of input data within the 400–450 nm regression window is increased from the 1 nm to 5 nm interval.

Figure 1.12 depicts analogous results to Fig. 1.11, but the scenario of 1 nm input data is compared with two MODIS-like scenarios in which only two or three wavelengths from the blue spectral region are used as input to the ELR regression analysis. These results show that the exponential fit to only two or three data points in the blue region results in a decline in performance of the extrapolation model. The  $MdR$  and  $MdAPD$  statistics for estimates of the three absorption coefficients in the near-UV region are clearly inferior when the input data are only at 412 and 443 nm or 412, 443, and 488 nm, if compared to the 1 nm input data within the 400–450 nm regression window. The most significant worsening of error statistics is observed for the estimates of  $a_d(\lambda)$  when the input data at the three wavelengths, 412, 443, and 488 nm, are used. If the two MODIS-like scenarios are compared to each other, the use of two wavelengths of 412 nm and 443 nm tends to provide superior estimation of near-UV absorption coefficients from the extrapolation model than the use of three wavelengths which additionally include 488 nm. It is also notable that the

error statistics for the two-wavelength scenario are quite satisfactory for all three absorption coefficients across the near-UV region as the  $MdR$  values remain within the range between about 1.02 and less than 1.06 and the  $MdAPD$  values remain below 10% with the exception of  $a_d(\lambda)$  estimation in the short-wavelength end of near-UV region where  $MdAPD$  reaches about 24%. Thus, for the potential application of the extrapolation model when the input absorption data in the VIS are available only at limited number of spectral bands such as in multispectral satellite ocean color sensors, our results suggest the use of two bands from the 400–450 nm range, such as 412 nm and 443 nm available on MODIS, rather than the use of additional absorption data at longer wavelengths from the blue and/or green spectral region.

#### **1.3.4 Validation of the extrapolation model with an independent dataset**

The performance of the final extrapolation model for estimating  $a_g(\lambda)$ ,  $a_{dg}(\lambda)$ , and  $a_d(\lambda)$  in the near-UV when applied to the validation dataset ( $N = 149$ ) is provided in the correlation plots between extrapolated and measured values at 350 and 380 nm (Fig. 1.13). We observe a generally good overall agreement between modeled and measured values of absorption coefficients. The most notable differences between measured and modeled data are observed as underestimations of  $a_d(350)$  at low magnitudes and more scatter in data of  $a_d(350)$  about the 1:1 line compared to  $a_g(350)$  and  $a_{dg}(350)$ . Specifically, for  $a_d(350)$  the  $MdR$  is 0.975 and  $MdAPD$  is 11%. These statistics are comparable to the model performance found in Section 3.2 when applied to the development dataset. We also examined the validation dataset using the input data to the extrapolation model that were subsampled at 5 nm intervals in the VIS spectral region (not shown), and we obtained similar results to those with 1 nm input data. The validation analysis with the independent field dataset supports the good performance of the final extrapolation model and its capability to extend  $a_g(\lambda)$ ,  $a_{dg}(\lambda)$ , and  $a_d(\lambda)$  from the VIS to the near-UV.

## 1.4 Summary

In this study we developed and assessed a method for extrapolating data of constituent absorption coefficients,  $a_g(\lambda)$ ,  $a_d(\lambda)$ , and  $a_{dg}(\lambda)$ , from the VIS spectral region, specifically its blue portion, into the near-UV (350 to 400 nm) spectral region. The extrapolation model was developed for the purpose of applicability to absorption data obtained either from measurements or models in the VIS range when the data in the UV are lacking or are deemed to be subject to significantly larger uncertainty than the available data in the VIS. One example scenario of applicability of the extrapolation model is the availability of data of constituent absorption coefficients derived from absorption partitioning models that were developed specifically for the VIS spectral region. Another example is the availability of constituent absorption data at VIS spectral bands as derived from inverse reflectance models, such as those designed for operation with satellite multispectral measurements of ocean color in the VIS. In both example cases, it is desirable to have an extrapolation model providing a reliable capability to extend data of  $a_g(\lambda)$ ,  $a_d(\lambda)$ , and  $a_{dg}(\lambda)$  from the VIS into the UV with satisfactory accuracy for a wide range of oceanic conditions.

We formulated the extrapolation model using an assembled development dataset that includes field measurements of 1294  $a_g(\lambda)$  spectra and 409 concurrent spectra of  $a_d(\lambda)$  and  $a_g(\lambda)$ , and hence also 409 spectra of  $a_{dg}(\lambda)$ . These measurements are available at 1 nm intervals over a broad spectral range from the UV to near-IR, and were collected within the near-surface ocean layer in diverse environments which made this dataset well suited for the purpose of our analysis. Importantly, the model development dataset includes data that satisfy several method-related and data quality criteria, most notably the accepted hyperspectral  $a_d(\lambda)$  data were obtained solely from spectrophotometric measurements using the inside integrating-sphere filter-pad technique. We determined that the exponential function fitted to the 1 nm data of  $a_g(\lambda)$  and  $a_{dg}(\lambda)$  within the 400–450 nm spectral window provides an adequate first step of the extrapolation model. The initial

estimates of  $a_d(\lambda)$  in the near-UV region are calculated as a difference between the extrapolated values of  $a_{dg}(\lambda)$  and  $a_g(\lambda)$ . Then, to further improve the overall performance of the extrapolation model, we formulated the correction functions to minimize or eliminate the bias in the initial estimates of  $a_g(\lambda)$  and  $a_d(\lambda)$  in the near-UV. The significance of these corrections is, however, restricted mostly to the short-wavelength portion of the near-UV region, and in the case of  $a_g(\lambda)$  to low magnitudes of this coefficient. The final products of the extrapolation model include the corrected values of  $a_g(\lambda)$  and  $a_d(\lambda)$  as well as the final  $a_{dg}(\lambda)$  calculated as a sum of corrected  $a_g(\lambda)$  and  $a_d(\lambda)$ .

The comparative analysis of model-derived estimates and measured values of the absorption coefficients using the model development dataset indicates a very good performance of the extrapolation model across the near-UV region. For example, the spectral values of aggregate bias at different wavelengths in the 350–400 range are generally less than 1% and the uncertainty in terms of median absolute percent difference (*MdAPD*) is typically less than 5% with the exception of about 10% for  $a_d(\lambda)$  estimates at or near 350 nm. The extrapolation model provides equally satisfactory performance if the exponential fit in the 400–450 nm spectral window is applied to subsampled absorption data at 5 nm intervals. We also determined that if absorption data are available only at relatively few wavelengths in the VIS including its blue spectral region, the extrapolation model still performs reasonably well if the exponential function is fitted to data at just two wavelengths in the 400–450 nm range, such as 412 nm and 443 nm available on MODIS satellite ocean color sensor. In this case, when compared to the 1 or 5 nm input data in the blue, the error statistics for near-UV absorption estimates deteriorates the most for  $a_d(\lambda)$  with *MdAPD* values reaching 20–25% at the short-wavelength end of the near-UV region. The extrapolation model was also validated with an independent dataset of subsurface (depths of 15 to 50 m)

measurements which were collected at the same oceanographic stations and processed using the same protocols as our model development dataset. This validation analysis supports the good performance of the extrapolation model and its capability to extend the spectral absorption coefficients  $a_g(\lambda)$ ,  $a_{dg}(\lambda)$ , and  $a_d(\lambda)$  from the VIS to the near-UV.

Whereas, the present study utilized the field datasets of absorption measurements to formulate and assess the performance of the extrapolation model, it will be desirable to further test this model in conjunction with input absorption data in the VIS derived from absorption partitioning models. For example, the partitioning models that were developed for the VIS spectral region and would be suitable for such analysis are described in Zheng and Stramski (2013) [47], Zheng et al. (2015) [46], and Stramski et al. (2019) [49]. In addition, in the future it will be desirable to test the extrapolation model with additional independent field datasets. A prerequisite for such analysis is the acquisition of more high-quality hyperspectral data of constituent absorption coefficients which satisfy the methodology-related and data quality criteria, especially the measurements of  $a_p(\lambda)$  and  $a_d(\lambda)$  with the spectrophotometric filter-pad technique using the inside integrating-sphere configuration or another absorption technique providing equivalent or better quality of particulate absorption data, such as PSICAM technique [65].

In summary, the significance of the proposed extrapolation model stems primarily from its applicability to and enhancement of existing and potential future absorption partitioning models that are formulated for the VIS spectral region. However, because the absorption partitioning models can be used in conjunction with inverse reflectance models with an ultimate goal of estimating the constituent absorption coefficients from ocean color data, our extrapolation model can also be relevant as a potential component of a suite of inverse optical models that are applied to satellite ocean color observations. At present, to our knowledge, no validated or sufficiently

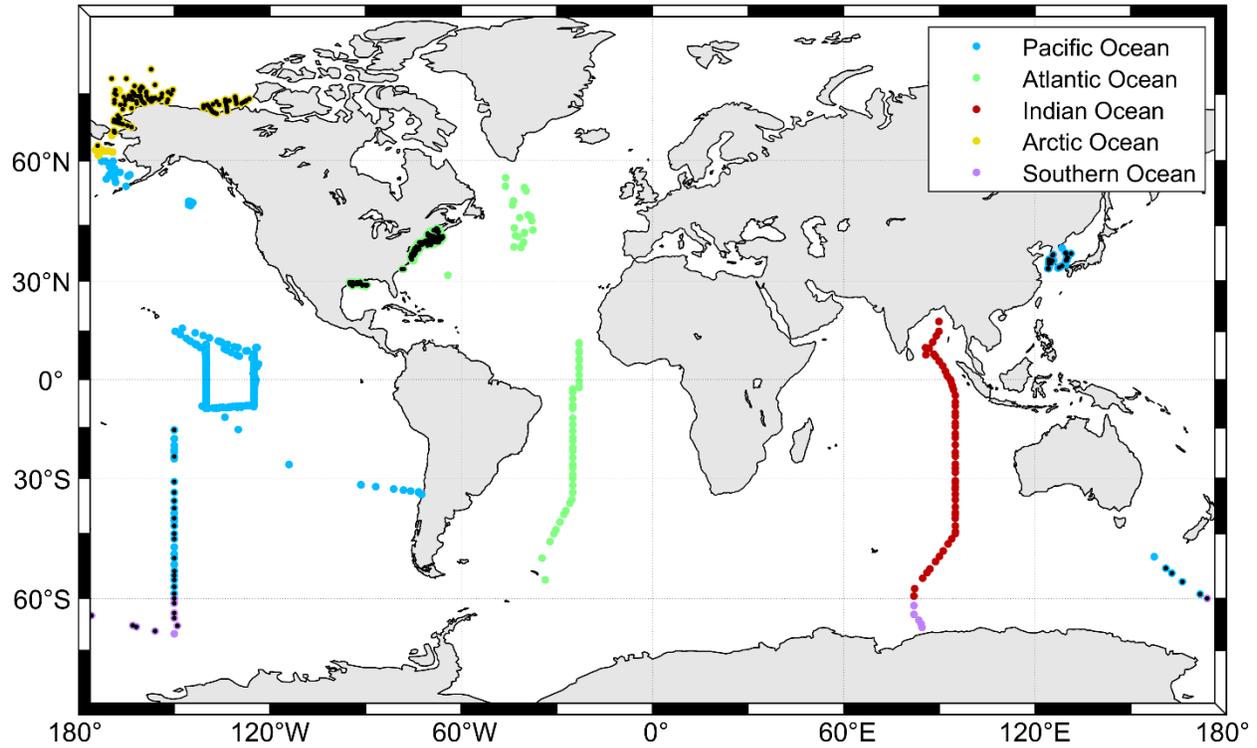
mature capabilities to estimate the constituent absorption coefficients in the UV from inverse reflectance and absorption partitioning models that utilize UV measurements are available, so our extrapolation model provides a tool for immediate applicability to existing models that operate in the VIS. Our present extrapolation model can also provide a valuable tool for comparison and assessment of future optical models that utilize UV measurements.

### **1.5 Acknowledgements**

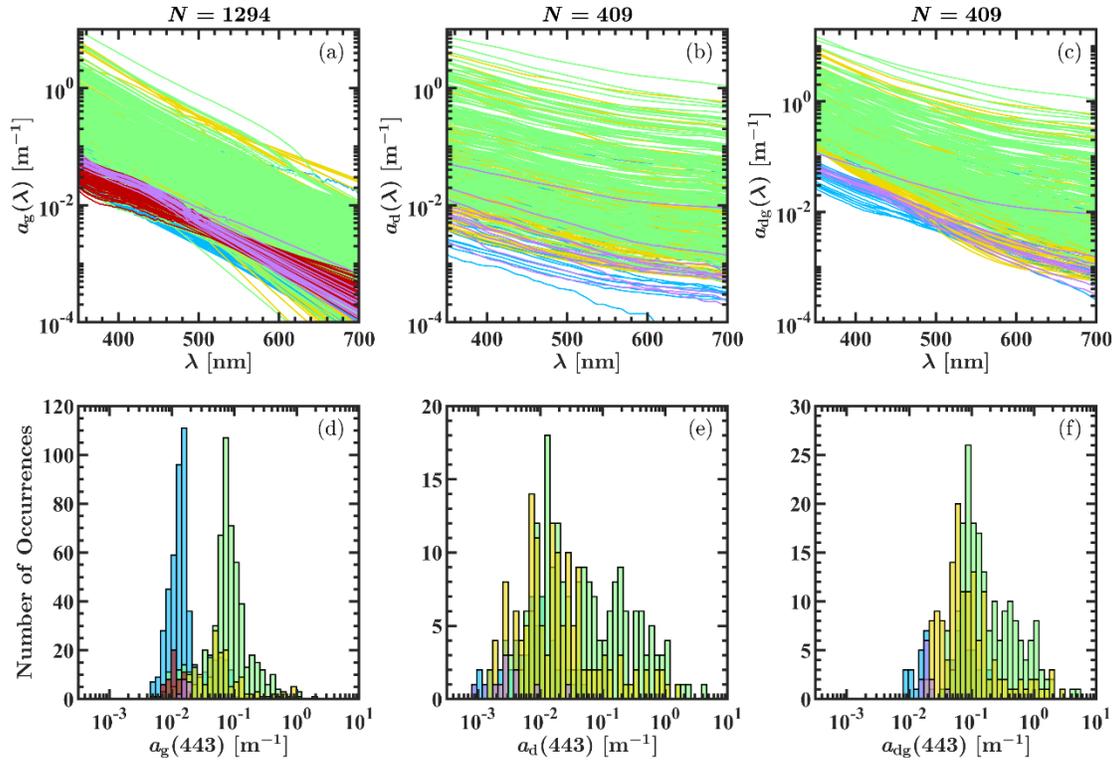
We thank all investigators who contributed to the collection, processing, and distribution through public databases of the field data of absorption coefficients used in this study. The public data sources assembled for this manuscript were available on NASA's SeaWiFS Bio-optical Archive and Storage System (SeaBASS) and from the Biogeochemistry and Optics South Pacific Experiment (BIOSOPE) dataset.

Chapter 1, in full, is a reprint of the material as it appears in *Optics Express*. The dissertation author was the primary investigator and author of this paper: Kehrli, M. D., Stramski, D., Reynolds, R. A., & Joshi, I. D. (2023). Estimation of chromophoric dissolved organic matter and non-algal particulate absorption coefficients of seawater in the ultraviolet by extrapolation from the visible spectral region. *Optics Express*, 31(11), 17450–17479. <https://doi.org/10.1364/OE.486354>.

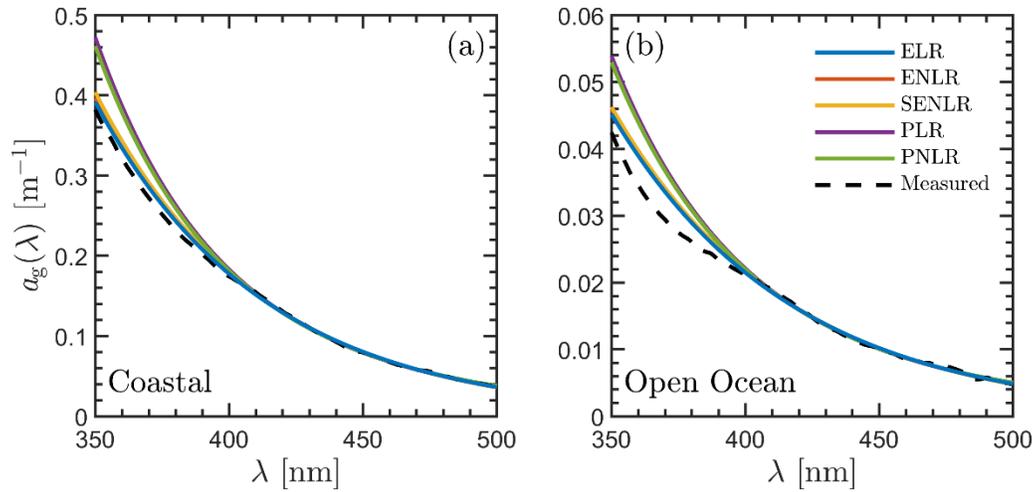
## 1.6 Figures



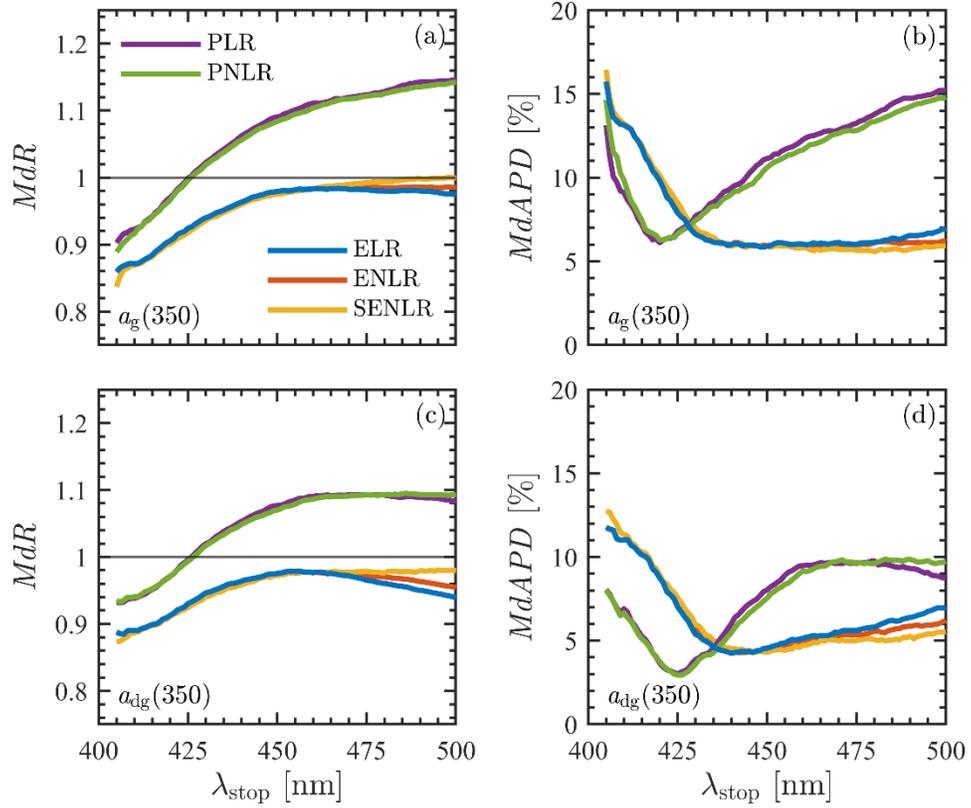
**Figure 1.1.** Geographic locations of oceanographic stations where near-surface measurements of the spectral absorption coefficients were collected for generation of the final model development dataset utilized in this study. Colored markers represent measurements of  $a_g(\lambda)$  (number of measurements  $N = 1294$ ) with the marker color corresponding to ocean basins as indicated in the legend, with the Arctic Ocean defined in this study as north of  $60^\circ\text{N}$  and the Southern Ocean defined as south of  $60^\circ\text{S}$ . Black dots in the center of markers indicate samples with concurrent measurements of  $a_g(\lambda)$  and  $a_d(\lambda)$  ( $N = 409$ ).



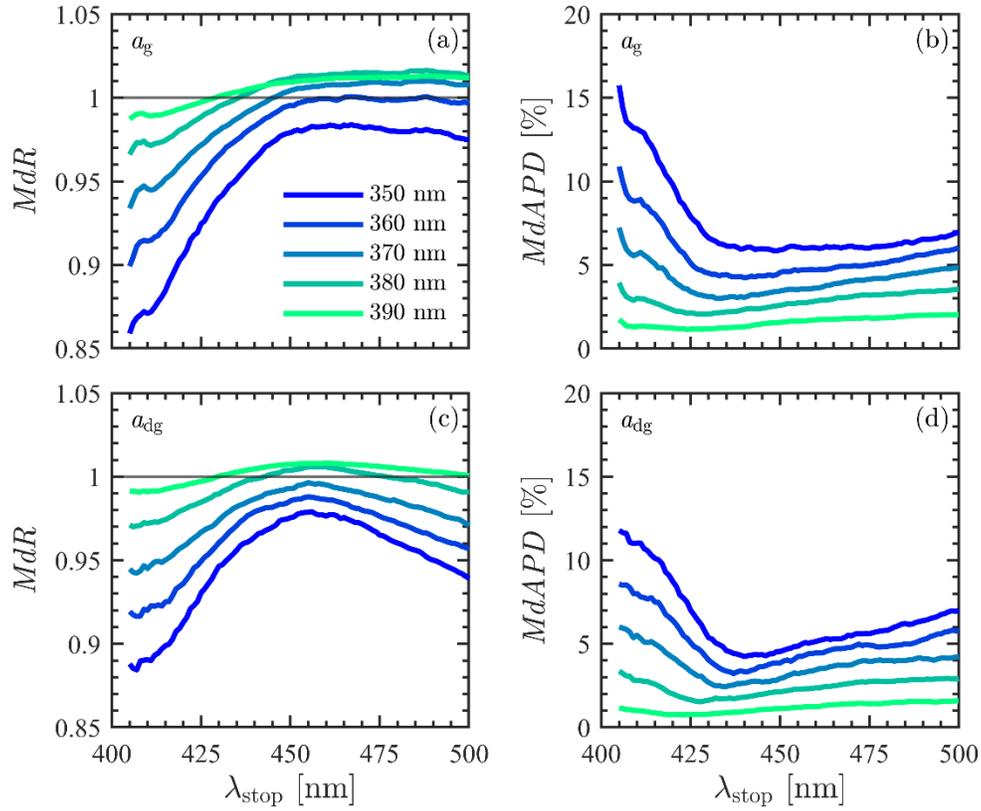
**Figure 1.2.** Spectral absorption coefficients (a)  $a_g(\lambda)$ , (b)  $a_{dg}(\lambda)$ , and (c)  $a_d(\lambda)$  available in the final model development dataset. Blue lines are data from the Pacific Ocean, green lines are data from the Atlantic Ocean, red lines are data from the Indian Ocean, yellow lines are data from the Arctic Ocean (defined in this study as north of  $60^\circ\text{N}$ ), and lavender lines are data from the Southern Ocean (defined in this study as south of  $60^\circ\text{S}$ ). (d)–(f) Histograms of each absorption coefficient at  $\lambda = 443$  nm are color coded in the same manner as (a)–(c).



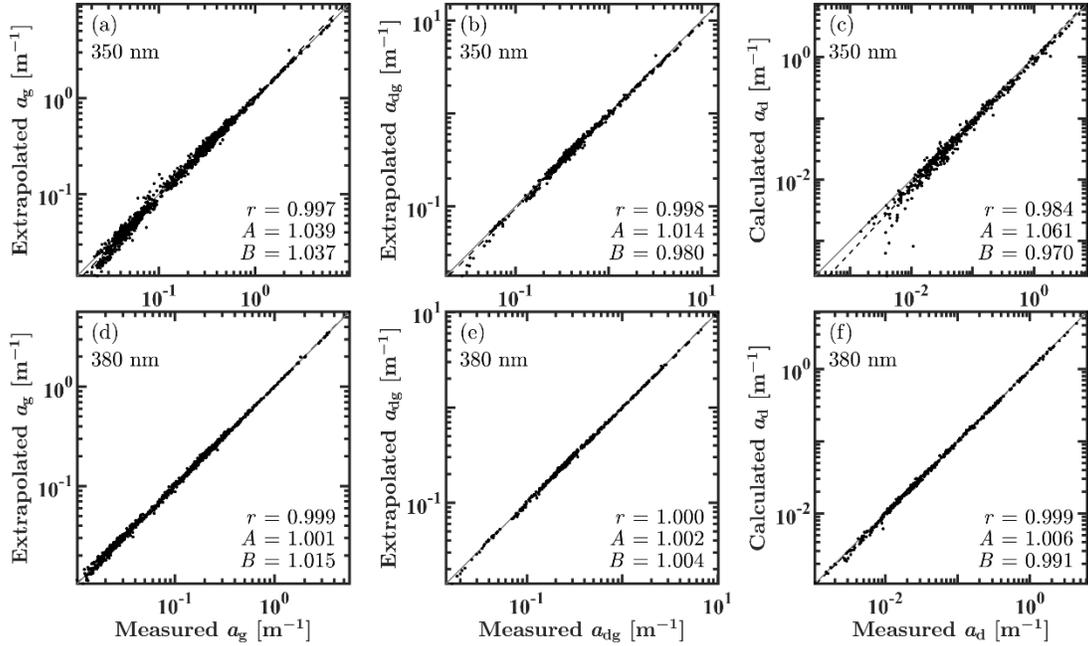
**Figure 1.3.** Comparison of measured spectra of  $a_g(\lambda)$  with fitted curves from exponential linear regression (ELR), exponential nonlinear regression (ENLR), stretched exponential nonlinear regression (SENLR), power linear regression (PLR), and power nonlinear regression (PNLR) using a regression spectral window of 400–500 nm. (a) Example data from coastal waters collected during the MR17-05C cruise in the Chukchi Sea, and (b) example data from open ocean waters collected during the CLIVAR I8SI9N cruise in the Pacific Ocean.



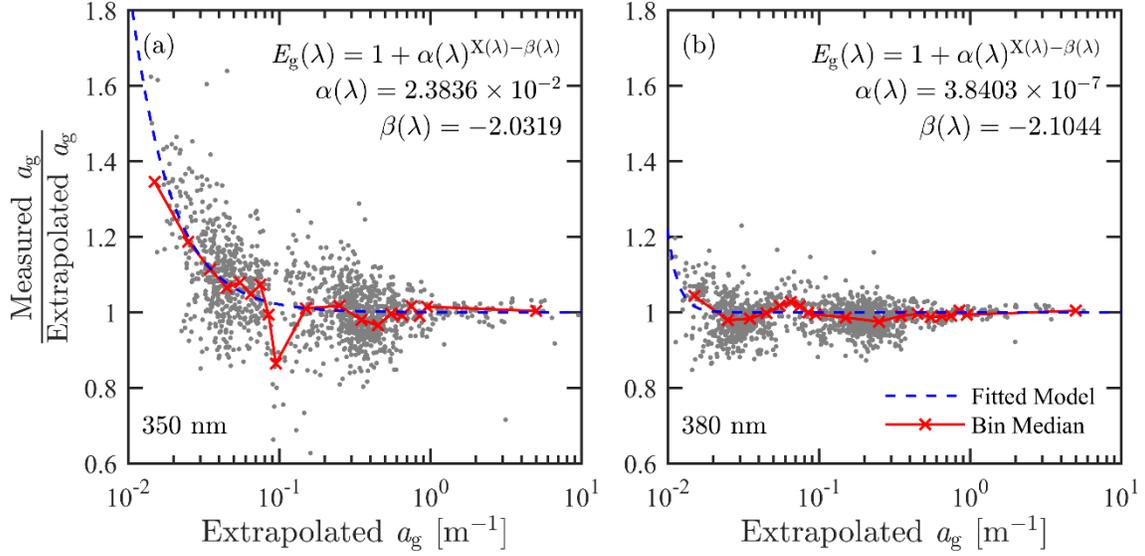
**Figure 1.4.** Median ratio,  $MdR$ , and median absolute percent difference,  $MdAPD$ , vs. stop wavelength ( $\lambda_{\text{stop}}$ ) of the spectral regression window in the VIS range for (a, b) extrapolated  $a_g(350)$  ( $N = 1294$ ) and (c, d) extrapolated  $a_{dg}(350)$  ( $N = 409$ ). The regression window always starts at 400 nm and ends at varying  $\lambda_{\text{stop}}$  from 405 to 500 nm in 1 nm increments. The statistical parameters of  $a_g(350)$  and  $a_{dg}(350)$  are shown for five regression models as indicated in panel (a) (see text for more details).



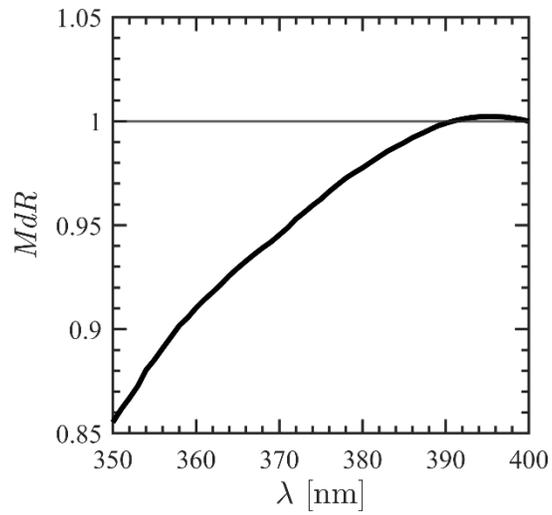
**Figure 1.5.** Median ratio,  $MdR$ , and median absolute percent difference,  $MdAPD$ , vs. stop wavelength ( $\lambda_{stop}$ ) of (a, b) extrapolated  $a_g$  ( $N = 1294$ ) and (c, d) extrapolated  $a_{dg}$  ( $N = 409$ ) at five near-UV wavelengths (350, 360, 370, 380, and 390 nm) as obtained from the ELR regression model applied to the spectral regression window starting at 400 nm and ending at varying  $\lambda_{stop}$  with ELR.



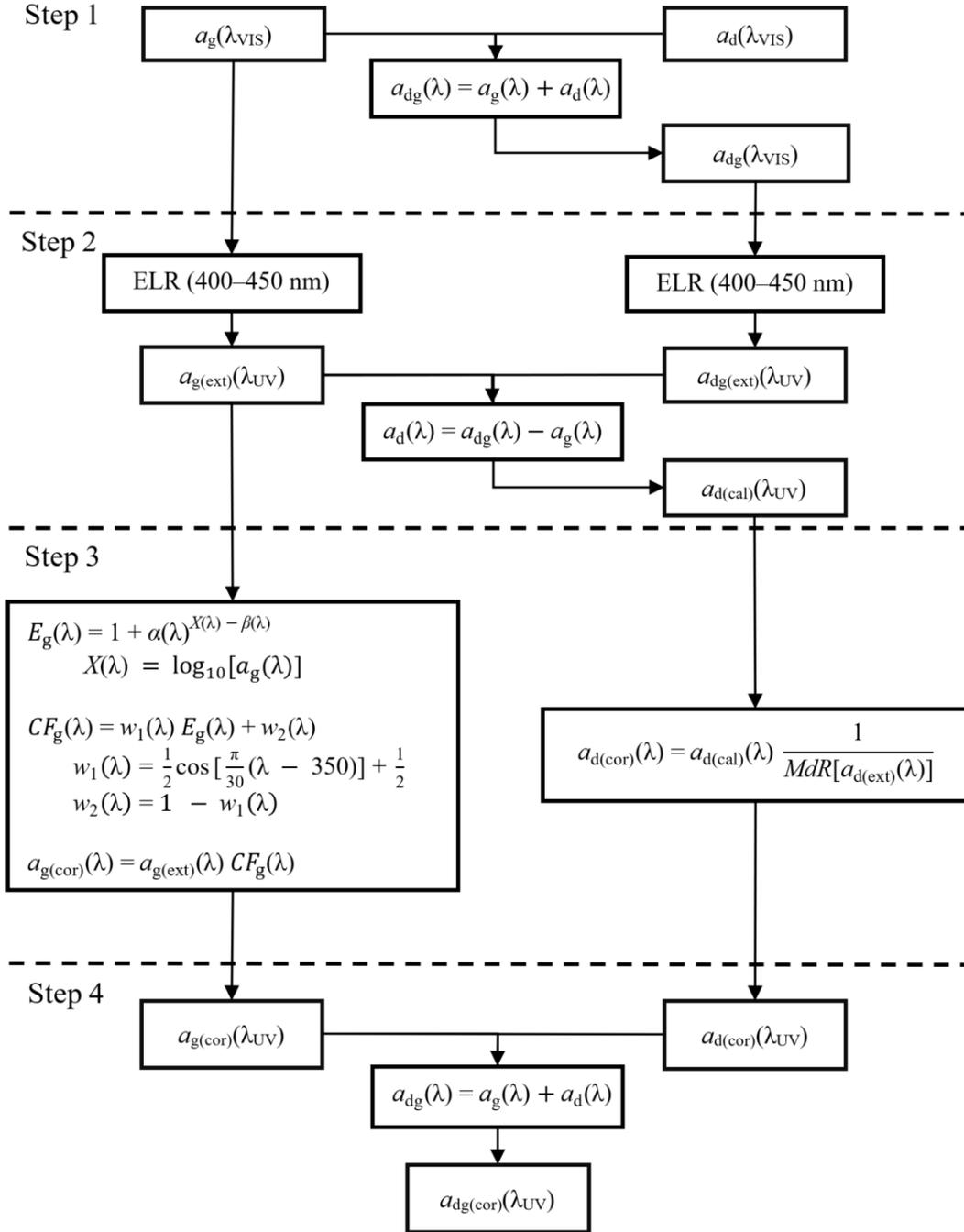
**Figure 1.6.** Scatter plots comparing extrapolated ( $a_g$  and  $a_{dg}$ ) or calculated ( $a_d$ ) vs. measured absorption coefficients at 350 nm (a–c) and 380 nm (d–f). The statistical parameters including Pearson correlation coefficient,  $r$ , between  $\log_{10}$ -transformed predicted and measured values, as well as slope,  $A$ , and coefficient  $B$ , where  $B = 10^I$  and  $I$  is the y-intercept obtained from the Model-II linear regression of  $\log_{10}$ -transformed predicted vs. measured values are depicted in each panel. The 1:1 line and best-fit line derived from the Model II linear regression to  $\log_{10}$ -transformed data are represented by the solid gray and dashed black lines, respectively.



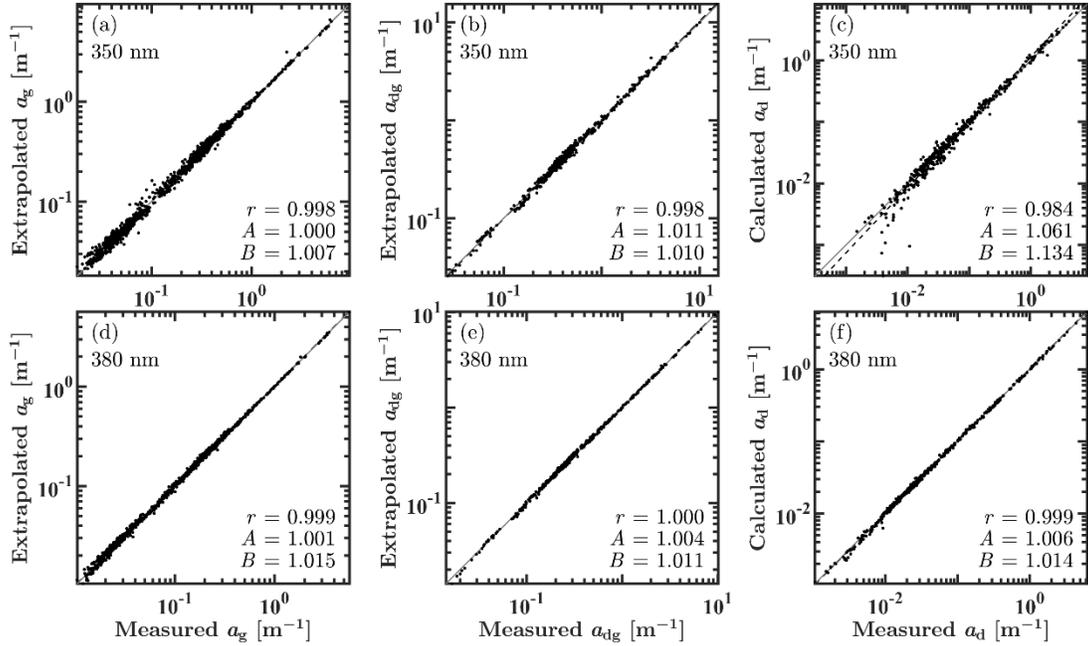
**Figure 1.7.** The relationship between the ratio of measured to extrapolated  $a_g(\lambda)$  vs. extrapolated  $a_g(\lambda)$  at 350 nm (a) and 380 nm (b) obtained from the optimal (ELR) extrapolation method (see text for more details). The red line and crosses depict the median of each bin between every  $x$ -tick mark, except above  $1\text{m}^{-1}$  where the median of all points is calculated and plotted at  $5\text{m}^{-1}$ . The blue dashed line denotes the non-linear fit to the function  $E_g(\lambda)$ , where the steepness parameter,  $\alpha(\lambda)$ , and shift parameter,  $\beta(\lambda)$ , derived from the fit are displayed in each panel.



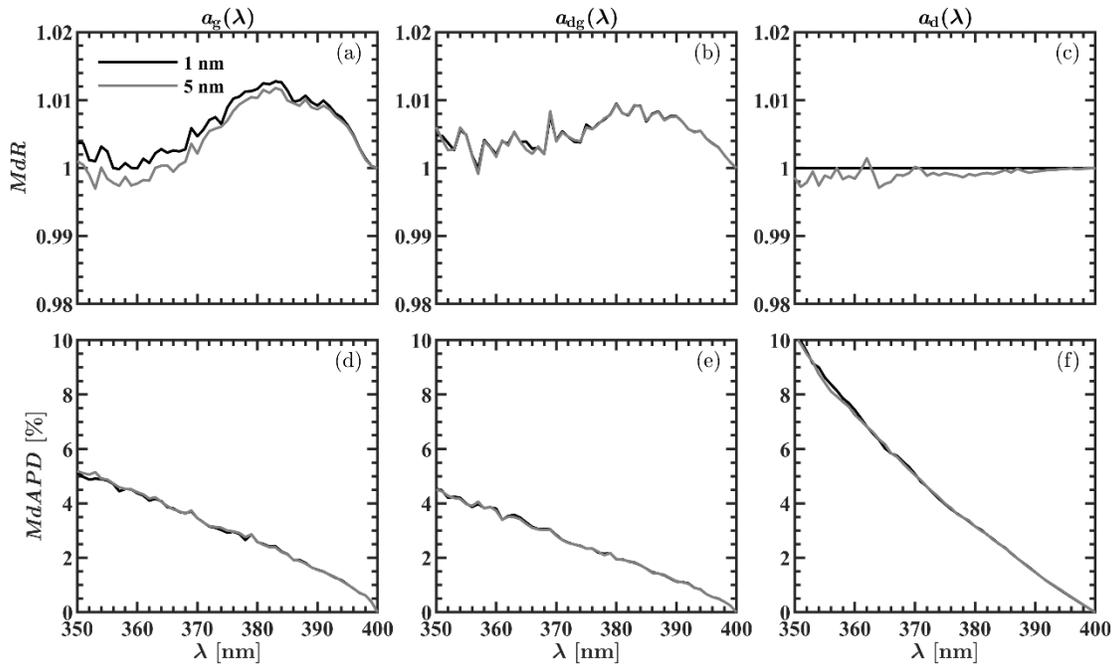
**Figure 1.8.** Median ratio,  $MdR$ , of modeled to measured  $a_d(\lambda)$  in the near-UV spectral region. Modeled values of  $a_d(\lambda)$  were calculated from the difference  $a_{dg}(\lambda) - a_g(\lambda)$ , where values of  $a_{dg}(\lambda)$  and  $a_g(\lambda)$  were determined with the optimal (ELR) extrapolation method. The numerical values of  $MdR$  are reported in Table 1.4 and utilized to correct the initial estimates of  $a_d(\lambda)$  (see text for more details).



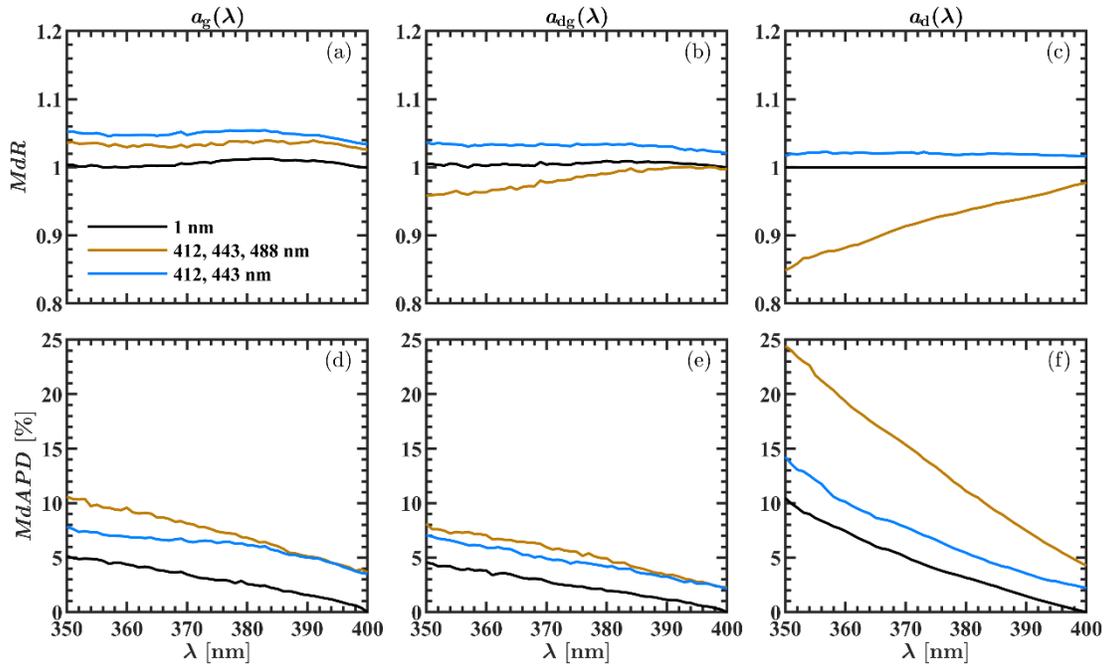
**Figure 1.9.** Schematic describing the final extrapolation model for extending the spectral absorption coefficients  $a_g(\lambda)$ ,  $a_d(\lambda)$ , and  $a_{\text{dg}}(\lambda)$  from the VIS to the near-UV spectral region. Step 1 depicts the input absorption data in the VIS. The extrapolation subsection (Step 2) outlines the initial optimal (ELR) extrapolation method to determine the three absorption coefficients in the near-UV. The correction subsection (Step 3) describes the two independent corrections applied to  $a_g(\lambda)$  and  $a_d(\lambda)$  (for model parameters see Tables 1.3, 1.4) to obtain the final output of each absorption coefficient in the near-UV (Step 4).



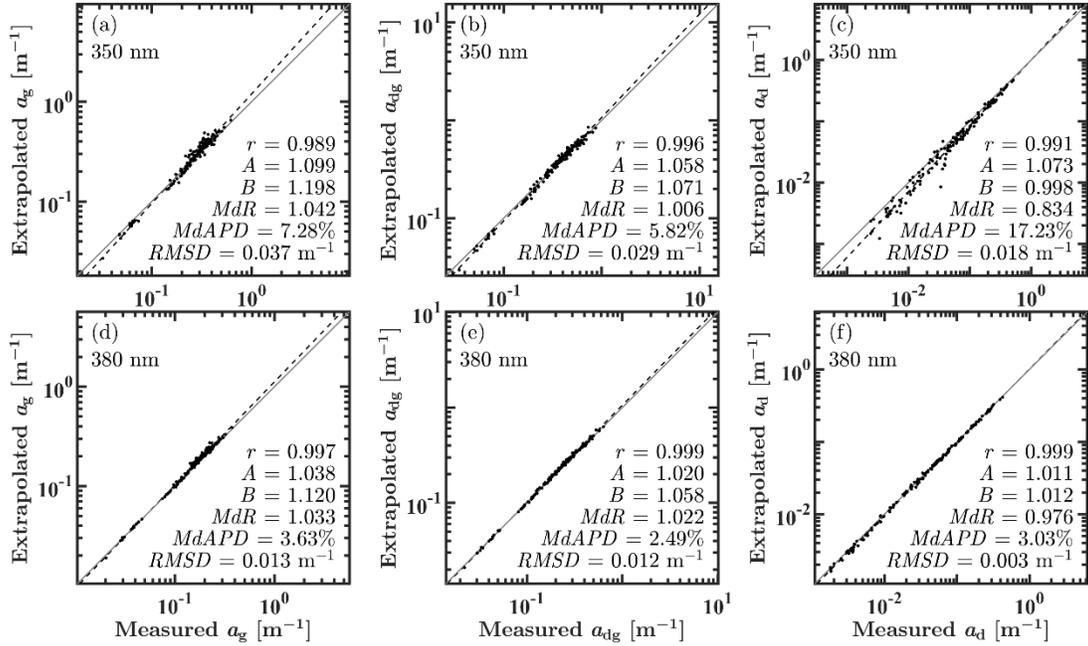
**Figure 1.10.** The performance of the final extrapolation model applied to the development dataset and illustrated with scatter plots of the model-derived vs. measured absorption coefficients at 350 nm (a–c) and 380 nm (d–f). The statistical parameters including Pearson correlation coefficient,  $r$ , between  $\log_{10}$ -transformed predicted and measured values, as well as slope,  $A$ , and coefficient  $B$ , where  $B = 10^I$  and  $I$  is the y-intercept obtained from the Model-II linear regression of  $\log_{10}$ -transformed predicted vs. measured values are depicted in each panel. The 1:1 line and best-fit line derived from the Model II linear regression to  $\log_{10}$ -transformed data are represented by the solid gray and dashed black lines, respectively.



**Figure 1.11.** Median ratio,  $MdR$ , (a-c) and median absolute percent difference,  $MdAPD$ , (d-f) in the near-UV spectral region for estimates of  $a_g(\lambda)$ ,  $a_{dg}(\lambda)$ , and  $a_d(\lambda)$  obtained from the final extrapolation model. The results are calculated with input data from the VIS spectral region with the original 1 nm spectral sampling interval (black lines) and by subsampling the input data into a 5 nm spectral interval (gray lines).



**Figure 1.12.** Same statistics as presented in Fig. 1.11, but results determined with input data with the original 1 nm spectral sampling interval (black lines) are compared with results obtained by subsampling the input data into three MODIS bands (gold lines) and two MODIS bands (light blue lines) from the blue portion of the VIS spectral region.



**Figure 1.13.** The performance of the final extrapolation model applied to the validation dataset ( $N = 149$ ) as scatter plots of model-derived versus measured absorption coefficients at 350 nm (a–c) and 380 nm (d–f). The statistical parameters of the Pearson correlation coefficient,  $r$ , between log<sub>10</sub>-transformed predicted and measured values, slope,  $A$ , coefficient  $B$ , where  $B = 10^I$  and  $I$  is the y-intercept obtained from the Model-II linear regression of log<sub>10</sub>-transformed predicted vs. measured values, median ratio,  $MdR$ , median absolute percent difference,  $MdAPD$ , and root-mean-square deviation,  $RMSD$ , are depicted in each panel. The 1:1 line and best-fit line derived from the Model II linear regression to log<sub>10</sub>-transformed data are represented by the solid gray and dashed black lines, respectively.

## 1.7 Tables

**Table 1.1.** Final Model Development Dataset of Absorption Coefficients,  $a_g(\lambda)$  and  $a_d(\lambda)$ . Description of the final development dataset of hyperspectral absorption coefficients,  $a_g(\lambda)$  and  $a_d(\lambda)$ , collected in near-surface waters in different ocean basins. The number,  $N$ , of individual  $a_g(\lambda)$  and concurrent  $a_g(\lambda)$  and  $a_d(\lambda)$  measurements are listed for each experiment.

Ocean basin	Experiment (Cruises)	$N [a_g(\lambda)]$	$N [a_g(\lambda) \& a_d(\lambda)]$	
Pacific	BIOSOPE (BIOSOPE)	31	0	
	TAO (gp1-06-ka, gp5-05-ka, gp5-06-ka)	292	0	
	CLIVAR (P16S)	44	31	
	KORUS (KR_2016)	61	40	
	EXPORTS (Process, Survey)	63	0	
Atlantic	CLIVAR (A16S)	29	0	
	AMMA (AMMA-RB-06)	12	0	
	CLiVEC (CV1, CV2, CV4, CV5, CV7)	357	93	
	ECOMON (PC1301)	25	24	
	GEO-CAPE (GOMEX_2013)	87	75	
	BBOP (bats311a, bats 312)	3	0	
	NAAMES (NA1, NA2, NA3, NA4)	37	0	
	Cyanate (Cyanate 2016)	7	4	
	Indian	CLIVAR (I8SI9N)	62	0
	Arctic	BEST (hly0803)	26	0
MALINA (MALINA)		60	53	
ICESCAPE (HLY1001, HLY1101)		76	71	
ArCS (MR17-05C)		22	18	
Total		1294	409	

**Table 1.2.** Regression Models Evaluated for Extrapolation. Summary of the five regression models considered to extrapolate the spectral constituent absorption coefficients,  $a_g(\lambda)$  and  $a_{dg}(\lambda)$ , from the VIS to the near-UV spectral region (the subscript x in the symbol  $a_x(\lambda)$  indicates g or dg).

Regression	Function	Best-fit coefficients
Exponential Linear Regression (ELR)	$\log_e [a_x(\lambda)] = -A_1\lambda + B_1$	$A_1, B_1$
Exponential Non-Linear Regression (ENLR)	$a_x(\lambda) = B_2 \exp(-A_2\lambda)$	$A_2, B_2$
Stretched Exponential Non-Linear Regression (SENLR)	$a_x(\lambda) = B_3 \exp[-A_3(\lambda)^{C_3}]$	$A_3, B_3, C_3$
Power Linear Regression (PLR)	$\log_{10}[a_x(\lambda)] = -A_4 \log_{10}(\lambda) + B_4$	$A_4, B_4$
Power Non-Linear Regression (PNLR)	$a_x(\lambda) = B_5\lambda^{-A_5}$	$A_5, B_5$

**Table 1.3.** Spectral Values of the Parameters of the  $E_g(\lambda)$  Function. Spectral values of the parameters of the  $E_g(\lambda)$  function within the spectral range of 350–380 nm (Eq. 1.2 in text), which are involved in the correction of the absorption coefficient  $a_g(\lambda)$  in the near-UV spectral region (see text and Fig. 1.9 for more details on this correction). Values of the parameters in the 381–399 nm range are not displayed because  $a_g(\lambda)$  is not corrected beyond 380 nm.

$\lambda$ [nm]	$\alpha(\lambda)$	$\beta(\lambda)$	$w_1(\lambda)$	$w_2(\lambda)$
350	$2.3836 \times 10^{-2}$	-2.0319	1.0000	0.0000
351	$1.9847 \times 10^{-2}$	-2.0376	0.9973	0.0027
352	$1.7033 \times 10^{-2}$	-2.0463	0.9891	0.0109
353	$1.3399 \times 10^{-2}$	-2.0480	0.9755	0.0245
354	$1.0390 \times 10^{-2}$	-2.0477	0.9568	0.0432
355	$7.8771 \times 10^{-3}$	-2.0477	0.9330	0.0670
356	$5.9628 \times 10^{-3}$	-2.0486	0.9045	0.0955
357	$4.2241 \times 10^{-3}$	-2.0474	0.8716	0.1284
358	$3.0184 \times 10^{-3}$	-2.0452	0.8346	0.1654
359	$2.1416 \times 10^{-3}$	-2.0434	0.7939	0.2061
360	$1.5069 \times 10^{-3}$	-2.0427	0.7500	0.2500
361	$1.0299 \times 10^{-3}$	-2.0427	0.7034	0.2966
362	$6.9047 \times 10^{-4}$	-2.0422	0.6545	0.3455
363	$4.5629 \times 10^{-4}$	-2.0416	0.6040	0.3960
364	$3.1039 \times 10^{-4}$	-2.0423	0.5523	0.4477
365	$2.4258 \times 10^{-4}$	-2.0464	0.5000	0.5000
366	$1.6812 \times 10^{-4}$	-2.0492	0.4477	0.5523
367	$9.1705 \times 10^{-5}$	-2.0495	0.3960	0.6040
368	$4.8465 \times 10^{-5}$	-2.0499	0.3455	0.6545
369	$4.9241 \times 10^{-5}$	-2.0571	0.2966	0.7034
370	$5.0495 \times 10^{-5}$	-2.0646	0.2500	0.7500
371	$2.1476 \times 10^{-5}$	-2.0639	0.2061	0.7939
372	$7.3611 \times 10^{-6}$	-2.0626	0.1654	0.8346
373	$4.7468 \times 10^{-6}$	-2.0663	0.1284	0.8716
374	$2.9035 \times 10^{-6}$	-2.0688	0.0955	0.9045
375	$1.4236 \times 10^{-6}$	-2.0728	0.0670	0.9330
376	$5.5883 \times 10^{-7}$	-2.0764	0.0432	0.9568
377	$4.0701 \times 10^{-7}$	-2.0826	0.0245	0.9755
378	$5.0469 \times 10^{-7}$	-2.0901	0.0109	0.9891
379	$5.1102 \times 10^{-7}$	-2.0975	0.0027	0.9973
380	$3.8403 \times 10^{-7}$	-2.1044	0.0000	1.0000

**Table 1.4.** Spectral Values of  $MdR$  for  $a_d(\lambda)$ . Spectral values of the median ratio,  $MdR$ , of modeled to measured  $a_d(\lambda)$  within the spectral range of 350–399 nm, which are used in the correction of  $a_d(\lambda)$  in the near-UV spectral region (see text and Fig. 1.9 for more details on this correction). The reported values of  $MdR$  are also depicted in Fig. 1.8.

$\lambda$ [nm]	$MdR$ for $a_d(\lambda)$	$\lambda$ [nm]	$MdR$ for $a_d(\lambda)$
350	0.855	375	0.963
351	0.862	376	0.966
352	0.867	377	0.969
353	0.873	378	0.973
354	0.880	379	0.975
355	0.885	380	0.978
356	0.891	381	0.980
357	0.896	382	0.983
358	0.902	383	0.986
359	0.906	384	0.988
360	0.910	385	0.990
361	0.914	386	0.992
362	0.918	387	0.994
363	0.922	388	0.996
364	0.926	389	0.998
365	0.929	390	0.999
366	0.933	391	1.000
367	0.936	392	1.001
368	0.939	393	1.002
369	0.942	394	1.002
370	0.945	395	1.002
371	0.949	396	1.002
372	0.953	397	1.002
373	0.956	398	1.001
374	0.960	399	1.001

**Table 1.5.** Error Statistics of the Final Extrapolation Model. The aggregate error statistics of the final extrapolation model applied to the development dataset and based on the comparison of absorption coefficients  $a_g(\lambda)$ ,  $a_d(\lambda)$ , and  $a_{dg}(\lambda)$  estimated from the final extrapolation model with measured values of these absorption coefficients in 10 nm intervals within the near-UV spectral range. *MdB*, *MdR*, *MdAPD*, and *RMSD* denote the median bias, median ratio, median absolute percent difference, and root-mean-square deviation, respectively.  $N$  is the number of data,  $r$  is the Pearson correlation coefficient, and  $A$  and  $B$  are the parameters obtained from Model II linear regression of  $\log_{10}$ -transformed model-derived vs. measured values (see section 2.3 for more details on statistical parameters).

	$N$	$r$	$A$	$B$	$MdB$ [ $m^{-1}$ ]	$MdR$	$MdAPD$ [%]	$RMSD$ [ $m^{-1}$ ]
$a_g(350)$	1294	0.9975	1.000	1.007	0.0005	1.0037	5.14	0.037
$a_g(360)$	1294	0.9982	1.003	1.007	0.0000	1.0000	4.37	0.024
$a_g(370)$	1294	0.9988	1.003	1.013	0.0004	1.0047	3.46	0.015
$a_g(380)$	1294	0.9993	1.001	1.015	0.0008	1.0116	2.58	0.009
$a_g(390)$	1294	0.9998	1.001	1.013	0.0006	1.0092	1.56	0.004
$a_d(350)$	409	0.9844	1.061	1.134	0.0000	1.0000	10.45	0.075
$a_d(360)$	409	0.9913	1.032	1.071	0.0000	1.0000	7.43	0.043
$a_d(370)$	409	0.9982	1.013	1.030	0.0000	1.0000	5.08	0.025
$a_d(380)$	409	0.9994	1.006	1.014	0.0000	1.0000	3.15	0.013
$a_d(390)$	409	0.9999	1.002	1.004	0.0000	1.0000	1.47	0.006
$a_{dg}(350)$	409	0.9983	1.011	1.010	0.0023	1.0049	4.54	0.092
$a_{dg}(360)$	409	0.9988	1.009	1.006	0.0007	1.0020	3.81	0.054
$a_{dg}(370)$	409	0.9991	1.007	1.008	0.0010	1.0041	2.84	0.034
$a_{dg}(380)$	409	0.9996	1.004	1.011	0.0019	1.0095	1.95	0.020
$a_{dg}(390)$	409	0.9999	1.000	1.007	0.0015	1.0076	1.13	0.009

## 1.8 References

1. C. D. Mobley, *Light and Water: Radiative Transfer in Natural Waters* (Academic Press, 1994).
2. M. Kishino, M. Takahashi, N. Okami, and S. Ichimura, “Estimation of the spectral absorption coefficients of phytoplankton in the sea,” *Bull. Mater. Sci.* **37**(2), 634–642 (1985).
3. R. M. Pope and E. S. Fry, “Absorption spectrum (380–700 nm) of pure water. II. Integrating cavity measurements,” *Appl. Opt.* **36**(33), 8710–8723 (1997).
4. M. Twardowski, R. Röttgers, and D. Stramski, “Chapter 1: The Absorption Coefficient, An Overview,” In: *IOCCG Protocol Series Inherent Optical Property Measurements and Protocols: Absorption Coefficient*, A. R., Neeley and A. Mannino eds., IOCCG Ocean Optics and Biogeochemistry Protocols for Satellite Ocean Colour Sensor Validation, Volume 1.0, IOCCG, Dartmouth, NS, Canada (2018).
5. N. G. Jerlov, *Marine Optics* (Elsevier, 1976).
6. A. Bricaud, A. Morel, and L. Prieur, “Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains,” *Limnol. Oceanogr.* **26**(1), 43–53 (1981).
7. N. B. Nelson and D. A. Siegel, *Biogeochemistry of Marine Dissolved Organic Matter: Chromophoric DOM in the Open Ocean* (Elsevier, 2002), pp. 547–578.
8. N. V. Blough and R. Del Vecchio, *Biogeochemistry of Marine Dissolved Organic Matter: Chromophoric DOM in the Coastal Environment* (Elsevier, 2002), pp. 509–546.
9. M. Babin, D. Stramski, G. M. Ferrari, H. Claustre, A. Bricaud, G. Obolensky, and N. Hoepffner, “Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe,” *J. Geophys. Res.* **108**(C7), 3211 (2003).
10. 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 U.S. coast,” *Remote Sens. Environ.* **152**, 576–602 (2014).
11. N. B. Nelson and D. A. Siegel, “The global distribution and dynamics of chromophoric dissolved organic matter,” *Annu. Rev. Mar. Sci.* **5**(1), 447–476 (2013).
12. C. E. Del Castillo and R. L. Miller, “On the use of ocean color remote sensing to measure the transport of dissolved organic carbon by the Mississippi River Plume,” *Remote Sens. Environ.* **112**(3), 836–844 (2007).
13. C. G. Fichot and R. Benner, “A novel method to estimate DOC concentrations from CDOM absorption coefficients in coastal waters,” *Geophys. Res. Lett.* **38**(3), na (2011).

14. C. A. Stedmon, S. Markager, and H. Kaas, "Optical properties and signatures of chromophoric dissolved organic matter (CDOM) in Danish coastal waters," *Estuar. Coast. Shelf Sci.* **51**(2), 267–278 (2000).
15. R. F. Chen, P. Bissett, P. Coble, R. Conmy, G. B. Gardner, M. A. Moran, X. Wang, M. L. Wells, P. Whelan, and R. G. Zepp, "Chromophoric dissolved organic matter (CDOM) source characterization in the Louisiana Bight," *Mar. Chem.* **89**(1-4), 257–272 (2004).
16. N. B. Nelson, D. A. Siegel, C. A. Carlson, C. Swan, W. M. Smethie Jr, and S. Khatiwala, "Hydrography of chromophoric dissolved organic matter in the North Atlantic," *Deep Sea Res., Part I* **54**(5), 710–731 (2007).
17. R. N. Conmy, P. G. Coble, J. P. Cannizzaro, and C. A. Heil, "Influence of extreme storm events on West Florida Shelf CDOM distributions," *J. Geophys. Res.* **114**(G4), G00F04 (2009).
18. A. Matsuoka, A. Bricaud, R. Benner, J. Para, R. Sempère, L. Prieur, S. Bélanger, and M. Babin, "Tracing the transport of colored dissolved organic matter in water masses of the Southern Beaufort Sea: relationship with hydrographic characteristics," *Biogeosciences* **9**(3), 925–940 (2012).
19. R. W. Eppley and B. J. Peterson, "Particulate organic matter flux and planktonic new production in the deep ocean," *Nature* **282**(5740), 677–680 (1979).
20. D. M. Karl, G. A. Knauer, and J. H. Martin, "Downward flux of particulate organic matter in the ocean: a particle decomposition paradox," *Nature* **332**(6163), 438–441 (1988).
21. M. Babin and D. Stramski, "Variations in the mass-specific absorption coefficient of mineral particles suspended in water," *Limnol. Oceanogr.* **49**(3), 756–767 (2004).
22. J. C. Moore, E. L. Berlow, D. C. Coleman, P. C. de Ruiter, Q. Dong, A. Hastings, N. Collins Johnson, K. S. McCann, K. Melville, P. J. Morin, K. Nadelhoffer, A. D. Rosemond, D. M. Post, J. L. Sabo, K. M. Scow, M. J. Vanni, and D. H. Wall, "Detritus, trophic dynamics and biodiversity," *Ecol. Lett.* **7**(7), 584–600 (2004).
23. F. A. J. Armstrong, P. M. Williams, and J. D. H. Strickland, "Photo-oxidation of organic matter in sea water by ultra-violet radiation, analytical and other applications," *Nature* **211**(5048), 481–483 (1966).
24. A. Vodacek, N. V. Blough, M. D. DeGrandpre, E. T. Peltzer, and R. K. Nelson, "Seasonal variation of CDOM and DOC in the Middle Atlantic Bight: terrestrial inputs and photooxidation," *Limnol. Oceanogr.* **42**(4), 674–686 (1997).
25. K. Mopper and D. J. Kieber, *Biogeochemistry of Marine Dissolved Organic Matter: Photochemistry and the cycling of carbon, sulfur, nitrogen and phosphorus* (Elsevier, 2002), pp. 455–507.

26. J. R. Helms, A. Stubbins, J. D. Ritchie, E. C. Minor, D. J. Kieber, and K. Mopper, "Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter," *Limnol. Oceanogr.* **53**(3), 955–969 (2008).
27. J. J. Cullen and P. J. Neale, "Ultraviolet radiation, ozone depletion, and marine photosynthesis," *Photosynth. Res.* **39**(3), 303–320 (1994).
28. W. P. Bissett, O. Schofield, S. Glenn, J. J. Cullen, W. L. Miler, A. J. Plueddemann, and C. D. Mobley, "Resolving the impacts and feedback of ocean optics on upper ocean ecology," *Oceanography* **14**(3), 30–53 (2001).
29. A. G. Huntsman, "No. 4: Limiting Factors for Marine Animals: I. The Lethal Effect of Sunlight," *Contrib. Can. Biol. Fish.* **2**(1), 81–88 (1924).
30. A. B. Klugh, "The effect of the ultra-violet component of the sun's radiation upon some aquatic organisms," *Can. J. Res.* **2**(5), 312–317 (1930).
31. D. Karentz and L. H. Lutze, "Evaluation of biologically harmful ultraviolet radiation in Antarctica with a biological dosimeter designed for aquatic environments," *Limnol. Oceanogr.* **35**(3), 549–561 (1990).
32. R. C. Smith, B. B. Prézelin, K. S. Baker, R. R. Bidigare, N. P. Boucher, T. Coley, D. Karentz, S. MacIntyre, H.A. Matlick, D. Menzies, M. Ondrusek, Z. Wan, and K. J. Waters, "Ozone depletion: ultraviolet radiation and phytoplankton biology in Antarctic waters," *Science* **255**(5047), 952–959 (1992).
33. D. M. Leech and C. E. Williamson, "In situ exposure to ultraviolet radiation alters the depth distribution of *Daphnia*," *Limnol. Oceanogr.* **46**(2), 416–420 (2001).
34. P. J. Werdell, L. I. W. McKinna, E. Boss, S. G. Ackleson, S. E. Craig, W. W. Gregg, Z. Lee, S. Maritorena, C. S. Roesler, C. S. Rousseaux, D. Stramski, J. M. Sullivan, M. S. Twardowski, M. Tzortziou, and X. Zhang, "An overview of approaches and challenges for retrieving marine inherent optical properties from ocean color remote sensing," *Prog. Oceanogr.* **160**, 186–212 (2018).
35. S. A. Garver and D. A. Siegel, "Inherent optical property inversion of ocean color spectra and its biogeochemical interpretation 1. Time series from the Sargasso Sea," *J. Geophys. Res.* **102**(C8), 18607–18625 (1997).
36. H. Loisel and D. Stramski, "Estimation of the inherent optical properties of natural waters from the irradiance attenuation coefficient and reflectance in the presence of Raman scattering," *Appl. Opt.* **39**(18), 3001–3011 (2000).

37. 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," *Appl. Opt.* **41**(27), 5755–5772 (2002).
38. S. Maritorena, D. A. Siegel, and A. R. Peterson, "Optimization of a semianalytical ocean color model for global-scale applications," *Appl. Opt.* **41**(15), 2705–2714 (2002).
39. M. H. Pinkerton, G. F. Moore, S. J. Lavender, M. P. Gall, K. Oubelkheir, K. M. Richardson, P. W. Boyd, and J. Aiken, "A method for estimating inherent optical properties of New Zealand continental shelf waters from satellite ocean colour measurements," *N. Z. J. Mar. Freshwater Res.* **40**(2), 227–247 (2006).
40. T. J. Smyth, G. F. Moore, T. Hirata, and J. Aiken, "Semianalytical model for the derivation of ocean color inherent optical properties: description, implementation, and performance assessment," *Appl. Opt.* **45**(31), 8116–8131 (2006).
41. P. J. Werdell, B. A. Franz, S. W. Bailey, G. C. Feldman, E. Boss, V. E. Brando, M. Dowell, T. Hirata, S. J. Lavender, Z. Lee, H. Loisel, S. Maritorena, F. Mélin, T. S. Moore, T. J. Smyth, D. Antoine, E. Devred, O. H. F. d'Andon, and A. Mangin, "Generalized ocean color inversion model for retrieving marine inherent optical properties," *Appl. Opt.* **52**(10), 2019–2037 (2013).
42. H. Loisel, D. Stramski, D. Dessailly, C. Jamet, L. Li, and R. A. Reynolds, "An inverse model for estimating the optical absorption and backscattering coefficients of seawater from remote-sensing reflectance over a broad range of oceanic and coastal marine environments," *J. Geophys. Res. Oceans* **123**(3), 2141–2171 (2018).
43. C. L. Gallegos and P. J. Neale, "Partitioning spectral absorption in case 2 waters: discrimination of dissolved and particulate components," *Appl. Opt.* **41**(21), 4220–4233 (2002).
44. O. Schofield, T. Bergmann, M. J. Oliver, A. Irwin, G. Kirkpatrick, W. P. Bissett, M. A. Moline, and C. Orrico, "Inversion of spectral absorption in the optically complex coastal waters of the Mid-Atlantic Bight," *J. Geophys. Res.* **109**(C12), C12S04 (2004).
45. J. Lin, W. Cao, G. Wang, and S. Hu, "Approach for determining the contributions of phytoplankton, colored organic material, and nonalgal particles to the total spectral absorption in marine waters," *Appl. Opt.* **52**(18), 4249–4257 (2013).
46. G. Zheng, D. Stramski, and P. M. DiGiacomo, "A model for partitioning the light absorption coefficient of natural waters into phytoplankton, nonalgal particulate, and colored dissolved organic components: a case study for the Chesapeake Bay," *J. Geophys. Res. Oceans* **120**(4), 2601–2621 (2015).

47. G. Zheng and D. Stramski, "A model based on stacked-constraints approach for partitioning the light absorption coefficient of seawater into phytoplankton and non-phytoplankton components," *J. Geophys. Res. Oceans* **118**(4), 2155–2174 (2013).
48. X. Zhang, Y. Huot, A. Bricaud, and H. Sosik, "Inversion of spectral absorption coefficients to infer phytoplankton size classes, chlorophyll concentration, and detrital matter," *Appl. Opt.* **54**(18), 5805–5816 (2015).
49. D. Stramski, L. Li, and R. A. Reynolds, "Model for separating the contributions of non-algal particles and colored dissolved organic matter to light absorption by seawater," *Appl. Opt.* **58**(14), 3790–3806 (2019).
50. A. Bricaud and D. Stramski, "Spectral absorption coefficients of living phytoplankton and nonalgal biogenous matter: a comparison between the Peru upwelling area and the Sargasso Sea," *Limnol. Oceanogr.* **35**(3), 562–582 (1990).
51. J. S. Cleveland and M. J. Perry, "A model for partitioning particulate absorption into phytoplanktonic and detrital components," *Deep Sea Res., Part I* **41**(1), 197–221 (1994).
52. K. Oubelkheir, H. Claustre, A. Bricaud, and M. Babin, "Partitioning total spectral absorption in phytoplankton and colored detrital material contributions," *Limnol. Oceanogr.: Methods* **5**(11), 384–395 (2007).
53. G. Wang, W. Cao, D. Yang, and J. Zhao, "Partitioning particulate absorption coefficient into contributions of phytoplankton and nonalgal particles: a case study in the northern South China Sea," *Estuar. Coast. Shelf Sci.* **78**(3), 513–520 (2008).
54. G. Zheng and D. Stramski, "A model for partitioning the light absorption coefficient of suspended marine particles into phytoplankton and nonalgal components," *J. Geophys. Res. Oceans* **118**(6), 2977–2991 (2013).
55. F. Cao and W. L. Miller, "A new algorithm to retrieve chromophoric dissolved organic matter (CDOM) absorption spectra in the UV from ocean color," *J. Geophys. Res. Oceans* **120**(1), 496–516 (2015).
56. C. M. Swan, N. B. Nelson, D. A. Siegel, and E. A. Fields, "A model for remote estimation of ultraviolet absorption by chromophoric dissolved organic matter based on the global distribution of spectral slope," *Remote Sens. Environ.* **136**, 277–285 (2013).
57. J. Wei and Z. Lee, "Retrieval of phytoplankton and colored detrital matter absorption coefficients with remote sensing reflectance in an ultraviolet band," *Appl. Opt.* **54**(4), 636–649 (2015).
58. I. Laurion, I. F. Blouin, and S. Roy, "The quantitative filter technique for measuring phytoplankton absorption: Interference by MAAs in the UV waveband," *Limnol. Oceanogr. Methods* **1**(1), 1–9 (2003).

59. D. Stramski, R. A. Reynolds, S. Kaczmarek, J. Uitz, and G. Zheng, “Correction of pathlength amplification in the filter-pad technique for measurements of particulate absorption coefficient in the visible spectral region,” *Appl. Opt.* **54**(22), 6763–6782 (2015).
60. C. Roesler, D. Stramski, E. J. D’Sa, R. Röttgers, and R. A. Reynolds, “Chapter 5: Spectrophotometric Measurements of Particulate Absorption Using Filter Pads,” In: *IOCCG Protocol Series Inherent Optical Property Measurements and Protocols: Absorption Coefficient*, A. R., Neeley and A. Mannino, eds., IOCCG Ocean Optics and Biogeochemistry Protocols for Satellite Ocean Colour Sensor Validation, Volume 1.0, IOCCG, Dartmouth, NS, Canada (2018).
61. P. J. Werdell, M. J. Behrenfeld, P. S. Bontempi, E. Boss, B. Cairns, G. T. Davis, B. A. Franz, U. B. Gliese, E. T. Gorman, O. Hasekamp, K. D. Knobelspiesse, A. Mannino, J. V. Martins, C. R. McClain, G. Meister, and L. A. Remer, “The Plankton, Aerosol, Cloud, Ocean Ecosystem (PACE) mission: Status, science, advances,” *Bull. Am. Meteorol. Soc.* **100**(9), 1775–1794 (2019).
62. B. G. Mitchell, A. Bricaud, and K. Carder, *et al.*, “Determination of spectral absorption coefficients of particles, dissolved material, and phytoplankton for discrete water samples,” In: *Ocean Optics Protocols for Satellite Ocean Color Sensor Validation*, Revision 2, NASA TM–2000–209966, G. S. Fargion and J. L. Mueller, eds. (NASA, 2000), pp. 125–153.
63. A. Mannino, M. G. Novak, N. B. Nelson, M. Belz, J.-F. Berthon, N. V. Blough, E. Boss, A. Bricaud, J. Chaves, C. Del Castillo, R. Del Vecchio, E. J. D’Sa, S. Freeman, A. Matsuoka, R. L. Miller, A. R. Neeley, R. Röttgers, M. Tzortziou, and P. J. Werdell, “Measurement protocol of absorption by chromophoric dissolved organic matter (CDOM) and other dissolved materials (DRAFT),” In: *IOCCG Protocol Series Inherent Optical Property Measurements and Protocols: Absorption Coefficient*, A., Mannino and M.G. Novak eds., IOCCG Ocean Optics and Biogeochemistry Protocols for Satellite Ocean Colour Sensor Validation, Volume 5.0, IOCCG, Dartmouth, NS, Canada (2019).
64. R. Röttgers and S. Gehnke, “Measurement of light absorption by aquatic particles: improvement of the quantitative filter technique by use of an integrating sphere approach,” *Appl. Opt.* **51**(9), 1336–1351 (2012).
65. I. Kostakis, M. Twardowski, C. Roesler, R. Röttgers, D. Stramski, D. McKee, A. Tonizzo, and S. Drapeau, “Hyperspectral optical absorption closure experiment in complex coastal waters,” *Limnol. Oceanogr.: Methods* **19**(9), 589–625 (2021).
66. A. Bricaud, M. Babin, H. Claustre, J. Ras, and F. Tièche, “Light absorption properties and absorption budget of Southeast Pacific waters,” *J. Geophys. Res.* **115**(C8), C08009 (2010).

67. C. S. Roesler, M. J. Perry, and K. L. Carder, "Modeling in situ phytoplankton absorption from total absorption spectra in productive inland marine waters," *Limnol. Oceanogr.* **34**(8), 1510–1523 (1989).
68. 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," *J. Geophys. Res.* **103**(C13), 31033–31044 (1998).
69. M. S. Twardowski, E. Boss, J. M. Sullivan, and P. L. Donaghay, "Modeling the spectral shape of absorption by chromophoric dissolved organic matter," *Mar. Chem.* **89**(1-4), 69–88 (2004).
70. D. G. Bowers and C. E. Binding, "The optical properties of mineral suspended particles: A review and synthesis," *Estuar. Coast. Shelf Sci.* **67**(1-2), 219–230 (2006).
71. D. Stramski, M. Babin, and S. B. Woźniak, "Variations in the optical properties of terrigenous mineral-rich particulate matter suspended in seawater," *Limnol. Oceanogr.* **52**(6), 2418–2433 (2007).
72. S. A. Loiselle, L. Bracchini, A. M. Dattilo, M. Ricci, A. Tognazzi, A. Cózar, and C. Rossi, "Optical characterization of chromophoric dissolved organic matter using wavelength distribution of absorption spectral slopes," *Limnol. Oceanogr.* **54**(2), 590–597 (2009).
73. B. B. Cael and E. Boss, "Simplified model of spectral absorption by non-algal particles and dissolved organic materials in aquatic environments," *Opt. Express* **25**(21), 25486–25491 (2017).
74. K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Q. Appl. Math.* **2**(2), 164–168 (1944).
75. D. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM J. Appl. Math.* **11**(2), 431–441 (1963).

## **Chapter 2**

**Model for partitioning the non-phytoplankton absorption coefficient of seawater in the ultraviolet and visible spectral range into the contributions of non-algal particulate and dissolved organic matter**

## 2.0 Abstract

Non-algal particles and chromophoric dissolved organic matter (CDOM) are two major classes of seawater constituents which contribute substantially to light absorption in the ocean within the ultraviolet (UV) and visible (VIS) spectral region. The similarities in the spectral shape of these two constituent absorption coefficients,  $a_d(\lambda)$  and  $a_g(\lambda)$  respectively, have led to their common estimation as a single combined non-phytoplankton absorption coefficient,  $a_{dg}(\lambda)$ , in optical remote sensing applications. Given the different biogeochemical and ecological roles of non-algal particles and CDOM in the ocean, it is important to determine and characterize the absorption coefficient of each of these constituents separately. We describe an ADG model that partitions  $a_{dg}(\lambda)$  into  $a_d(\lambda)$  and  $a_g(\lambda)$ . This model improves upon a recently published model [Appl. Opt. 58, 3790 (2019)] through implementation of a newly assembled dataset of hyperspectral measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$  from diverse oceanic environments to create the spectral shape function libraries of these coefficients, a better characterization of variability in spectral shape of  $a_d(\lambda)$  and  $a_g(\lambda)$ , and a spectral extension of model output to include the near-UV (350–400 nm) in addition to the VIS (400–700 nm) part of the spectrum. We developed and tested two variants of the ADG model; the ADG\_UV-VIS model which determines solutions over the spectral range from 350 to 700 nm, and the ADG\_VIS model which determines solutions in the VIS but can also be coupled with an independent extrapolation model to extend output to the near-UV. This specific model variant is referred to as ADG\_VIS-UV<sub>Ext</sub>. Evaluation of the model with development and independent datasets demonstrate good performance of both ADG\_UV-VIS and ADG\_VIS-UV<sub>Ext</sub>. Comparative analysis of model-derived and measured values of  $a_d(\lambda)$  and  $a_g(\lambda)$  indicates negligible or small median bias, generally within  $\pm 5\%$  over the majority of the 350–700 nm spectral range but extending to or above 10% near the ends of the spectrum, and the median percent difference generally below 20% with a maximum reaching about 30%. The presented ADG models

are suitable for implementation as a component of algorithms in support of satellite ocean color missions, especially the NASA PACE mission.

## 2.1 Introduction

The concentrations of optically significant seawater constituents such as phytoplankton, non-algal particles, and chromophoric (colored) dissolved organic matter (CDOM) play a fundamental role in controlling the propagation of light in aquatic environments [1]. An ongoing goal of optical oceanography is to assess the concentration and variability of these constituents by leveraging information obtained from optical measurements of seawater. The total spectral absorption coefficient of seawater,  $a(\lambda)$ , where  $\lambda$  is the wavelength of light in vacuum, is an inherent optical property (IOP) that quantifies the absorption of light per unit pathlength within the water column. The coefficient  $a(\lambda)$  can be decomposed into the sum of the spectral absorption coefficients of optically significant constituents, i.e.,  $a(\lambda) = a_w(\lambda) + a_{ph}(\lambda) + a_d(\lambda) + a_g(\lambda)$ , where  $a_w(\lambda)$ ,  $a_{ph}(\lambda)$ ,  $a_d(\lambda)$ , and  $a_g(\lambda)$  represent absorption coefficients due to pure water, phytoplankton, non-algal particles (also referred to as detritus or depigmented particles given the approach used in the measurement methodology), and CDOM, respectively [1–3]. The absorption coefficient of pure water is well characterized from the near-ultraviolet (near-UV) through visible (VIS) and into the near-infrared (near-IR) spectral region although the region of very low water absorption (near-UV to short visible wavelengths) still poses considerable challenges for accurate determinations of  $a_w(\lambda)$  [4,5].

The remaining non-water absorption coefficient,  $a_{nw}(\lambda) \equiv a(\lambda) - a_w(\lambda)$ , is highly variable across the world's oceans, and various methods of in situ measurements as well as optical inversion models have been employed to characterize the non-water constituent absorption coefficients  $a_{ph}(\lambda)$ ,  $a_d(\lambda)$ , and  $a_g(\lambda)$ . Pairs of absorption coefficients are often represented together either for methodological or conceptual reasons, namely the total particulate absorption coefficient,  $a_p(\lambda) \equiv$

$a_{\text{ph}}(\lambda) + a_{\text{d}}(\lambda)$ , and the non-phytoplankton absorption coefficient,  $a_{\text{dg}}(\lambda) \equiv a_{\text{d}}(\lambda) + a_{\text{g}}(\lambda)$ . From the modeling standpoint significant challenges arise in attempts to partition  $a_{\text{dg}}(\lambda)$  into  $a_{\text{d}}(\lambda)$  and  $a_{\text{g}}(\lambda)$  because these two coefficients generally exhibit similar spectral shapes with the magnitude of absorption decreasing gradually with increasing wavelength from the UV through VIS spectral region [1–3,6]. The relationships between the constituent absorption coefficients and concentrations of biogeochemically important constituents of seawater are of fundamental importance for optically-based observations. These include satellite observations over extended spatiotemporal scales of multiple oceanic properties and processes such as particulate and dissolved carbon reservoirs and cycling [7–14], primary production [15–20], phytoplankton community composition [21–29], dynamics of total suspended particulate matter [30,31], and water quality [32–35].

In the satellite remote sensing context, ocean color sensors measure upwelling radiance at the top of the atmosphere and, after applying an atmospheric correction, the spectral remote-sensing reflectance of the ocean,  $R_{\text{rs}}(\lambda)$ , is determined [36,37].  $R_{\text{rs}}(\lambda)$  is an apparent optical property that is generally weakly dependent on environmental conditions such as the geometry of the ambient underwater light field, cloud cover, solar angle, or sea surface conditions. Furthermore, measurements of  $R_{\text{rs}}(\lambda)$  contain substantial information about seawater IOPs especially  $a(\lambda)$  and the total backscattering coefficient,  $b_{\text{b}}(\lambda)$ . A variety of different inverse modeling approaches exist to estimate IOPs, including the total and constituent absorption coefficients, from measurements of  $R_{\text{rs}}(\lambda)$  [38]. Of primary relevance to the present study is a category of inverse reflectance models whose formulations and primary objectives are focused on retrieval of the total absorption coefficient  $a(\lambda)$  from input data of  $R_{\text{rs}}(\lambda)$  [39–43], rather than simultaneous retrieval of multiple constituent absorption coefficients directly from  $R_{\text{rs}}(\lambda)$  [44–50]. This is because the output of  $a(\lambda)$

from the former category of inverse reflectance models can be used as input to independently-developed absorption partitioning models to determine the non-water constituent absorption coefficients without a need to use  $R_{rs}(\lambda)$  data, and the present study is focused on one such specific type of a stand-alone absorption partitioning model which derives  $a_d(\lambda)$  and  $a_g(\lambda)$  from the sole input of  $a_{dg}(\lambda)$  data. This partitioning model is referred to as the ADG model.

The motivation for this study is associated largely with the fact that the multi-step inverse modeling approach that first retrieves the total  $a(\lambda)$ , and hence  $a_{nw}(\lambda)$ , from an inverse  $R_{rs}(\lambda)$ -based model and then the constituent absorption coefficients,  $a_{ph}(\lambda)$ ,  $a_d(\lambda)$ , and  $a_g(\lambda)$ , from absorption partitioning models has important advantages. Specifically, the development of separate inverse reflectance and absorption partitioning models enables the optimization of formulation and performance of each model independently in terms of targeting the estimation of specific output variable(s) using specific input variable(s) under the circumstances of strong mechanistic coupling between these variables. For example, there is evidence that  $a(\lambda)$  can be retrieved from reflectance models with better accuracy compared to other IOPs including the constituent absorption coefficients [42,43,51–53]. In addition, the multi-step approach consisting of inverse reflectance and absorption partitioning models provides an advantageous framework for understanding and quantifying uncertainties associated with each model separately as well as the propagation of uncertainties through the multi-step sequence of independent component models.

Multiple modeling strategies have been proposed to partition  $a(\lambda)$  or  $a_{nw}(\lambda)$  into constituent absorption coefficients. Our interest is in the strategy consisting of two separate absorption partitioning models operating in sequence, first an ANW model which partitions  $a_{nw}(\lambda)$  into  $a_{ph}(\lambda)$  and  $a_{dg}(\lambda)$  [40,54–57], and second an ADG model which partitions  $a_{dg}(\lambda)$  into  $a_d(\lambda)$  and  $a_g(\lambda)$  which is a focus of the present study. This two-step partitioning strategy represents an effective

pathway for obtaining the three main constituent absorption coefficients,  $a_{\text{ph}}(\lambda)$ ,  $a_{\text{d}}(\lambda)$ , and  $a_{\text{g}}(\lambda)$ , using just the total absorption  $a(\lambda)$  and pure water absorption  $a_{\text{w}}(\lambda)$  coefficients, and hence  $a_{\text{nw}}(\lambda)$ , as input data. It is notable that the total particulate absorption coefficient,  $a_{\text{p}}(\lambda)$ , is also obtainable as a sum of  $a_{\text{ph}}(\lambda)$  and  $a_{\text{d}}(\lambda)$ .

Historically, development of absorption partitioning models with a capability to separate  $a_{\text{d}}(\lambda)$  and  $a_{\text{g}}(\lambda)$  has been subject to distinctive challenges associated primarily with the similarity of spectral shapes of these coefficients [58–64]. These models are generally characterized by significant limitations such as restrictive assumptions about the output absorption coefficients, the use of ancillary input data (i.e., other than the input absorption data), and parameterizations reflecting the intended applicability to specific spectral range or discrete wavebands, specific geographic regions, and/or measurements taken with specific instruments. In particular, the existing models usually assume a single exponential function for the spectral shapes of  $a_{\text{d}}(\lambda)$  and  $a_{\text{g}}(\lambda)$  over a broad spectral range. However, the existing experimental data provide ample evidence that the spectra of  $a_{\text{d}}(\lambda)$  and  $a_{\text{g}}(\lambda)$ , and hence also  $a_{\text{dg}}(\lambda)$ , can depart significantly from a single exponential shape, especially when broad spectral ranges encompassing the UV and VIS are considered [10,65–72].

Recently, an ADG absorption partitioning model for estimating  $a_{\text{d}}(\lambda)$  and  $a_{\text{g}}(\lambda)$  from the sole input of  $a_{\text{dg}}(\lambda)$  data which avoids the restrictive assumption about the exponential spectral shape of non-phytoplankton constituent absorption coefficients was developed [73]. This model operates through the implementation of predefined libraries of characteristic spectral shape functions constructed from a dataset of hyperspectral measurements of  $a_{\text{d}}(\lambda)$  and  $a_{\text{g}}(\lambda)$  within the VIS spectral region. Such libraries are intended to account for the large variability of these coefficients observed across diverse oceanic environments. While the model of Stramski *et al.*

[73], was shown to be promising with good performance when tested with the dataset that was used to create the libraries of spectral shape functions, it is subject to some limitations which are worth considering for further improvements. First, although the model output is hyperspectral, it is constrained to the VIS spectral region. Second, the dataset of hyperspectral absorption coefficient  $a_d(\lambda)$  that was implemented to build the spectral shape library of  $a_d(\lambda)$  includes measurements with the spectrophotometric filter-pad method in the transmittance configuration. This measurement configuration is inferior to the inside integrating-sphere configuration which is recommended as the best option for obtaining hyperspectral data of particulate absorption with spectrophotometric filter-pad measurements [74–75]. In particular, the spectral shape of  $a_d(\lambda)$  obtained from the transmittance measurements is susceptible to substantial error as a result of the common application of the null-point correction which assumes that all types of marine particles, including non-algal particles, exhibit no absorption in the near-IR. This assumption, however, is not satisfied across diverse natural assemblages of marine particles [68,70,71,76–78]. In addition to the direct effect on the spectral values of  $a_d(\lambda)$ , the null-point correction produces a biasing effect on the spectral shape with a steeper spectral slope [79]. Third, to construct the spectral shape library the spectra of  $a_d(\lambda)$  and  $a_g(\lambda)$  were characterized by a single spectral shape parameter across the spectral region from 440 to 550 nm. This can be limiting in terms of accounting for potential variations in the spectral shape across this relatively wide blue-green region.

In this paper we present a new ADG partitioning model which builds upon the concepts used in Stramski *et al.* [73] and aims at enhancements and improvements of the model performance. One of the main enhancements is the extension of the new model from the VIS (400–700 nm) to the near-UV by including the wavelength range of 350–400 nm. This feature is particularly important in the context of the significance of  $a_d(\lambda)$  and  $a_g(\lambda)$  for absorption of UV

radiation in the ocean [1,6,80–83], and also a growing interest in extending optical remote-sensing capabilities to the near-UV, especially in view of spectral measurements to be collected with the satellite Ocean Color Instrument (OCI) as part of NASA’s Plankton Aerosol Cloud ocean Ecosystem (PACE) satellite mission [84]. Second, the development of the new ADG model and its spectral shape function libraries of  $a_d(\lambda)$  and  $a_g(\lambda)$  is based on assembling a new dataset of constituent absorption coefficients which were measured in diverse oceanic environments and cover a broad spectral range from at least 350 to 700 nm with a 1 nm interval. Importantly, several stringent inclusion and exclusion criteria related to methodology of absorption measurements and data quality were applied to reduce the risk of large uncertainties in the assembled dataset, especially biasing effects on the spectral shapes. Most notably, we disregarded data of  $a_d(\lambda)$  that were acquired with the spectrophotometric filter-pad transmittance method because of the potential errors as mentioned above. This exclusion criterion is one of the essential features of our  $a_d(\lambda)$  spectral shape library and ensures consistency with the currently recommended protocols for most accurate determinations of hyperspectral particulate absorption coefficient over a broad spectral range from the near-UV to near-IR from the inside integrating-sphere configuration of filter-pad measurements [74,75]. It is to be noted, however, that most historical data of hyperspectral particulate absorption coefficient  $a_p(\lambda)$ , including its  $a_d(\lambda)$  and  $a_{ph}(\lambda)$  components, were measured with the filter-pad transmittance method. Therefore, excluding these data from consideration naturally affects the size of the assembled dataset. An additional important enhancement of the new ADG model is the use of two spectral shape parameters in the wavelength range between 410 nm and 530 nm, which improves the ability to account for variations in the shape of  $a_d(\lambda)$  and  $a_g(\lambda)$  across the blue and blue-green spectral regions.

We present two variants of the new ADG model. The first variant, referred to as the ADG\_UV-VIS model, is based on the spectral shape libraries covering the 350–700 nm range and this model directly yields the output spectra of  $a_d(\lambda)$  and  $a_g(\lambda)$  over this entire spectral range. The second variant, referred to as the ADG\_VIS model, is based on the spectral shape libraries covering only the visible portion of the spectrum (400–700 nm). The direct output of this model in the visible range can, however, be extended to the near-UV range of 350–400 nm by implementing an independent extrapolation model described in Kehrli *et al.* [79]. The combination of the ADG\_VIS model with the extrapolation model is referred to as the ADG\_VIS-UV<sub>Ext</sub> model.

## **2.2 Model Development and Validation Datasets**

### **2.2.1 Assembly of the Development Dataset of Absorption Coefficients**

The new ADG partitioning model relies on a development dataset of hyperspectral measurements of constituent absorption coefficients  $a_d(\lambda)$  and  $a_g(\lambda)$ . The primary role of this dataset in model development is the formulation of spectral shape libraries of  $a_d(\lambda)$  and  $a_g(\lambda)$  which are used by the model to solve for  $a_d(\lambda)$  and  $a_g(\lambda)$  for a given input of  $a_{dg}(\lambda)$ , as described in detail in Section 4. This development dataset also provides a basis for initial performance evaluation of the model (Section 5.1). The ADG model is intended for applications to oceanic surface waters across diverse environments, so the assembled development dataset contains measurements made on numerous oceanographic cruises in different regions of the world's oceans. Importantly, the process of assembling the dataset involved the use of strict methodology-related and data quality criteria. This process was to ensure that the measurements of constituent absorption coefficients which passed the inclusion criteria were all collected with consistent methodology following currently recommended protocols for optimal determination of these coefficients. This aspect is important for the development of our model and its spectral shape libraries as this task requires

high-quality measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$ , including realistic detailed features that can be present in the spectral shapes of these two constituent absorption coefficients.

Recently, we assembled a dataset of  $a_d(\lambda)$  and  $a_g(\lambda)$  measurements for the purpose of developing a model to extrapolate  $a_d(\lambda)$ ,  $a_g(\lambda)$ , and  $a_{dg}(\lambda)$  from the blue part of the spectrum to the near-UV part of the spectrum [79]. For the development of the ADG partitioning model in this study, we assembled a similar dataset with some modifications. A detailed description of the approach to assembling the dataset is found in Kehrli *et al.*[79]. Here, we give a summary of the most critical aspects of this approach as applied in this study. The dataset for developing the ADG model is constrained to near-surface measurements. This is because we envision that the main application of the model will be in the context of ocean color remote sensing observations where other models provide an estimate of  $a_{dg}(\lambda)$  which serves as input to the ADG model. We defined the criteria of the near-surface measurements as a sampling depth  $\leq 5$  m if the water depth is  $< 200$  m or a sampling depth  $\leq 15$  m if the water depth is  $\geq 200$  m. The spectral criteria for inclusion in our dataset are defined as measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$  reported at a 1 nm sampling interval and spanning a broad spectral region at a minimum from the near-UV wavelength of 350 nm through the visible region up to 700 nm.

Data of  $a_g(\lambda)$  were required to be collected using either the spectrophotometric method with a 10 cm path length cuvette or a long path length liquid waveguide capillary cell where both methods follow standard measurement protocols [85,86]. Regarding the particulate absorption coefficient,  $a_p(\lambda)$ , and especially  $a_d(\lambda)$  which is of direct interest to this study, we required data to be collected with the spectrophotometric filter-pad method using the inside integrating-sphere (IS) configuration of measurement in accordance with recommended protocols [75]. Spectra of  $a_d(\lambda)$  were obtained after subjecting the  $a_p(\lambda)$  sample filters to methanol to remove phytoplankton

pigments [87]. Compared to the transmittance (T) or transmittance-reflectance (T-R) configurations of the filter-pad method, the IS configuration is highly advantageous for measuring the  $a_p(\lambda)$  and  $a_d(\lambda)$  spectra, as it minimizes different sources of error, especially those resulting from light-scattering effects on sample and blank filters [74]. Most historical data of  $a_p(\lambda)$  and  $a_d(\lambda)$  were collected using the T configuration which includes a null-point correction as a part of data processing. This correction assumes that particles exhibit no absorption at near-IR wavelengths and hence the measured signal at these wavelengths is associated solely with the unwanted effects of light scattering. However, this assumption is not necessarily satisfied by all types of marine particles and, consequently, data processing with the null-point correction can alter both the magnitudes and the spectral shapes of  $a_p(\lambda)$  and  $a_d(\lambda)$ , resulting in artificial steepening of the spectra [79]. Data prone to such errors were considered inadequate for the purpose of ADG model development. Thus, we use the  $a_d(\lambda)$  spectra collected only with the IS configuration of measurement which is the most accurate spectrophotometric configuration of the filter-pad method [74,75,88].

Historically, the studies of  $a_d(\lambda)$  and  $a_g(\lambda)$  spectra have often involved fitting a single exponential function of light wavelength to the original measured data across the entire spectral range of measurement from near-UV through visible and possibly extending into the near-IR, or within the visible range if primary research interest was limited to this range. For assembling our dataset, we considered only the sources of data which included the original measurements because preserving the presence of actual spectral features in the measured spectra of  $a_d(\lambda)$  and  $a_g(\lambda)$  is essential for development of spectral shape libraries of the ADG model and its operation. Data processing which was applied in this study to original measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$  has only a minimal effect on the final spectra included in our dataset. Specifically, many measurements of

$a_g(\lambda)$  selected for our dataset originally had a very low signal-to-noise ratio of the measured values in the long-wavelength portion of the spectrum, i.e., the green-red part of the spectrum. Therefore, the original values in this portion of the spectrum were replaced with the values of the exponential function that was fitted to the measured data in the blue portion of the spectrum where measurements had high signal-to-noise ratio. The fitted values of  $a_g(\lambda)$  in the long-wavelength portion of the spectrum are inconsequential to the development of our ADG model because the spectral shape library of  $a_g(\lambda)$  is based on the actual measured data in the blue and blue-green spectral regions. Processing of original measurements of  $a_d(\lambda)$  consisted of smoothing the spectral values to remove the potential presence of small-scale instrumental noise without affecting the actual features and variations in the spectral shape. In addition, as mentioned above we used only the  $a_d(\lambda)$  measurements collected with the IS configuration of the spectrophotometric filter-pad method. Thus, no assumption about zero absorption by particles in the near-IR was made. When the total particulate absorption measurement,  $a_p(\lambda)$ , displayed non-negligible (non-zero) signal in the near-IR, it was assumed that this signal is associated entirely with the  $a_d(\lambda)$  component, so the  $a_{ph}(\lambda)$  component in the near-IR is zero [89]. Accordingly, the originally measured spectral values of  $a_d(\lambda)$  were corrected by a spectrally constant offset to ensure that the resultant  $a_d(\lambda)$  is equal to measured  $a_p(\lambda)$  in the near-IR. It is notable that for many samples the original data of  $a_p(\lambda)$  and  $a_d(\lambda)$  measured with the IS spectrophotometric configuration in the near-IR are indistinguishable from one another within the instrumental noise. In some cases, the differences are distinguishable and interpreted as an artifactual shift in  $a_d(\lambda)$  values associated with the method of  $a_d(\lambda)$  measurement that involves the methanol treatment of sample filters. The process of data quality control of  $a_d(\lambda)$  and  $a_g(\lambda)$  measurements also included a visual inspection of every measurement which initially passed our method-related inclusion criteria. The measurements that exhibited a

high risk of gross experimental error, for example the presence of highly unrealistic spectral features, were excluded from our final dataset.

The final dataset for ADG model development contains measurements obtained on 29 research cruises as part of 18 field programs (Table 2.1). The measurements within the dataset span various geographic locations within five ocean basins (Fig. 2.1). Data of  $a_g(\lambda)$  were collected over a range of latitudes and well represent both coastal and open-ocean pelagic environments. Data of  $a_d(\lambda)$  were primarily collected in coastal areas in mid and high latitudes over or near continental shelves. A smaller proportion of  $a_d(\lambda)$  data was collected in more oligotrophic regions, for example on the Sea2Space cruise along the east-west transect from the Hawaiian Islands to the Oregon coast in the US and the north-south transect on the p16S cruise as a part of the CLIVAR program. In total, our final dataset includes 1293 measurements of  $a_g(\lambda)$  which were used for development of the spectral shape library of  $a_g(\lambda)$ , and 521 measurements of  $a_d(\lambda)$  which were used for development of the spectral shape library of  $a_d(\lambda)$  (Table 2.1). In addition, this dataset includes a subset of 390 concurrent measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$  available to obtain  $a_{dg}(\lambda)$  as a sum of  $a_d(\lambda)$  and  $a_g(\lambda)$ . This subset of concurrent  $a_d(\lambda)$  and  $a_g(\lambda)$  measurements is used to constrain parameterizations of the ADG partitioning model and assess the performance of the model.

The model development dataset covers the wide range of variability in  $a_d(\lambda)$  and  $a_g(\lambda)$  observed across diverse oceanic environments which supports the notion that the ADG model utilizing the spectral shape libraries derived from this dataset will have correspondingly wide applicability. Figure 2 displays histograms of  $a_d(443)$  and  $a_g(443)$  from the subset of the final development dataset that is used in the formation of spectral shape function libraries. In the blue spectral region, the range of  $a_d(443)$  spans ~4 orders of magnitude from about  $6 \times 10^{-4}$  to  $4 \text{ m}^{-1}$  and

the range of  $a_g(443)$  spans  $\sim 3$  orders of magnitude from 0.005 to 1.9  $\text{m}^{-1}$ . The distribution of observations of  $a_d(443)$  approximates a normal distribution in log-space and contains measurements primarily from coastal environments (Fig. 2.2a). In contrast, the distribution of  $a_g(443)$  is clearly bimodal and reflects that many of the observations were collected during expeditions in either oligotrophic waters of the Pacific Ocean or coastal waters in the Atlantic Ocean (Fig. 2.2b). A similar distribution pattern is observed in the red spectral region although absorption values are naturally smaller, for example at  $\lambda = 650$  nm the range of  $a_d(\lambda)$  is  $1.29 \times 10^{-4}$  to 1.32  $\text{m}^{-1}$  and the range of  $a_g(\lambda)$  is  $3.25 \times 10^{-5}$  to 0.05  $\text{m}^{-1}$ . The spectral slope of  $a_d(\lambda)$  ranges from 0.003 to 0.012  $\text{nm}^{-1}$  (calculated using the exponential function fit over the spectral range 380–730 nm, excluding 400–480 nm and 620–710 nm) with a mean value of 0.0077  $\text{nm}^{-1}$  (standard deviation (SD) = 0.0013  $\text{nm}^{-1}$ ) and the spectral slope of CDOM absorption ranges from 0.008 to 0.024  $\text{nm}^{-1}$  (calculated over the spectral range 350–500 nm) with a mean value of 0.016  $\text{nm}^{-1}$  (SD = 0.002  $\text{nm}^{-1}$ ). The spectral slopes for both absorption coefficients were calculated just for illustrative purposes and to show their consistency with previous literature [67], which suggests the dataset is representative of diverse oceanic environments and adequate for our purpose to develop the spectral shape libraries of ADG partitioning model. When considering the 390 concurrent measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$  within the dataset,  $a_g(\lambda)$  typically dominates the total non-phytoplankton absorption coefficient,  $a_{dg}(\lambda)$ , from the near-UV through the blue spectral region. For example, the average contribution of  $a_g$  to  $a_{dg}$  at  $\lambda = 443$  nm is 71.2%. In contrast, in the red  $a_d(\lambda)$  typically dominates the contribution to  $a_{dg}(\lambda)$  with the average contribution of 69.9% at  $\lambda = 650$  nm.

### 2.2.2 Assembly of an Independent Validation Dataset of Absorption Coefficients

A thorough assessment of model performance typically requires the use of validation data that are independent of data utilized in model development. Thus, such validation analysis of the ADG model requires an independent validation dataset of concurrent measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$ . In addition, to make the validation results most meaningful the validation dataset should be consistent with the intended applicability of the model and ideally satisfy similar method-related and data quality criteria as those used in the model development dataset. To our knowledge, the 390 concurrent measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$  included in our model development dataset, specifically the development of spectral shape function libraries, represent all presently available concurrent near-surface data of these two absorption coefficients which meet our strict method-related and data quality criteria, especially the use of the IS spectrophotometric method to measure  $a_d(\lambda)$ . Therefore, it appears that the most reasonable approach at this time to create an independent validation dataset for evaluating the ADG model is to use concurrent measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$  collected at depths below the near-surface layer on the same cruises as those used in the model development dataset. We constrained the sampling depth of measurements within the validation dataset to a range between 15 and 50 m. This dataset contains 149 samples of concurrent measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$  from a variety of geographic locations (Table 2.1). These measurements are restricted to a narrower dynamic range of magnitudes compared to the development dataset. Importantly, however, the measurements within the validation dataset are characterized by large spectral variability as well as a significant overlap with the development dataset in terms of the magnitude and spectral shape of the constituent absorption coefficients which is important for assessment of model performance. We also note that further details about this validation dataset can be found in Kehrli *et al.* [79], where the same data were used to validate the extrapolation model presented in that study.

### 2.3 Statistical Methods for Model Evaluation

The statistical metrics employed to assess the quality and performance of the ADG partitioning model through comparison of values predicted by the model,  $P_i$ , with measured (observed) values,  $O_i$ , are described in Table 2.2. Each statistical metric described in this section is determined at individual wavelengths from 350 nm to 700 nm to compare the degree of agreement between the model-derived and measured spectral values in the development and validation datasets, unless otherwise stated. To assess bias in model output compared to its associated measurement we calculated the Median Ratio,  $MdR$ , defined as the median of the ratio of predicted to observed variables and the Median Bias,  $MdB$ , defined as the median of the difference between predicted and observed variables. The ideal values (e.g., best agreement between modeled and measured values) for  $MdR$  and  $MdB$  are equal to 1 and  $0 \text{ m}^{-1}$ , respectively. To characterize random deviation between model-derived and measured values we calculated the Median Absolute Percent Difference,  $MdAPD$ , and the Root-Mean-Square Deviation,  $RMSD$ . The ideal values for  $MdAPD$  and  $RMSD$  are equal to 0% and  $0 \text{ m}^{-1}$ , respectively.  $MdAPD$  and  $RMSD$  are common uncertainty metrics that are implemented to assess model performance and can facilitate comparison of our ADG partitioning models to heritage models; however, it is notable that these metrics penalize the model overestimations to a greater extent than the model underestimations. To remove this penalizing imbalance, we also calculated the Median Symmetric Accuracy,  $MdSA$  (expressed in percent) [90,91].

A Model II linear regression based on the reduced major axis method [92] was also used to evaluate the relationship between the log<sub>10</sub>-transformed values of model-derived vs. measured absorption coefficients. The Model II analysis computes the linear regression slope,  $A$ , and intercept,  $I$ , which supports the assessment of the general performance of the ADG model and also

provides insight into potential bias in specific regions across the dynamic range of absorption coefficients. The value  $I$  was converted to a parameter  $B$  as described in Table 2.2, so the ideal values for  $A$  and  $B$  are both 1 which corresponds to the overall best agreement between the model-derived and measured values. The deviations of  $A$  and  $B$  from 1 provide information about the performance of the model across the range of measured values of the examined absorption coefficient. For example, when  $A$  is greater than 1, it indicates the tendency of the model to be negatively biased toward low-magnitude measurements and to be positively biased toward high-magnitude measurements.

## **2.4 Description of ADG Absorption Partitioning Model**

### **2.4.1 Model Overview**

The schematic of the ADG model is depicted in Fig. 2.3. The model relies on the use of libraries of spectral shape functions for the non-algal particulate absorption coefficient,  $\hat{a}_d(\lambda)$ , and the CDOM absorption coefficient,  $\hat{a}_g(\lambda)$ . The libraries of  $\hat{a}_d(\lambda)$  and  $\hat{a}_g(\lambda)$  were created using the  $a_d(\lambda)$  and  $a_g(\lambda)$  measurements included in our model development dataset described in Section 2.1, and the complete approach to create the libraries is described in Section 4.2. For each constituent absorption coefficient, the spectral shape libraries were created for two spectral regions: in the UV-VIS from 350 to 700 nm and in the VIS from 400 to 700 nm. As a result, there are two variants of the ADG model depending on whether the model operates with the UV-VIS library or VIS library. Conceptually, the two variants of the model operate in the same way except for how the solutions for  $a_d(\lambda)$  and  $a_g(\lambda)$  are determined in the near-UV region from 350 to 400 nm. When the UV-VIS library is used, the solutions are determined directly from the ADG model over the UV-VIS spectral region from 350 to 700 nm. This variant of the ADG model will be referenced as the ADG\_UV-VIS model. When the VIS library is used, solutions are obtained directly over the visible region from 400 to 700 nm, and this variant of the ADG model will be referenced as the

ADG\_VIS model. However, in this case the solutions in the near-UV can also be obtained through extrapolation from the blue part of the spectrum into the 350 to 400 nm wavelength range using a separate extrapolation model described in Kehrli *et al.* [79]. This combination of ADG\_VIS and extrapolation model will be referred to as the ADG\_VIS-UV<sub>Ext</sub> model. The distinction between ADG\_UV-VIS and ADG\_VIS-UV<sub>Ext</sub> is not shown explicitly in Fig. 2.3 as this schematic is focused on highlighting the common components of the model.

In the first step of model operation, the input values of spectral absorption coefficient,  $a_{dg}(\lambda)$ , and the libraries of spectral shape functions  $\hat{a}_d(\lambda)$  and  $\hat{a}_g(\lambda)$  are used to determine solutions of  $a_d(\lambda)$  and  $a_g(\lambda)$  that are physically possible and are referred to as speculative solutions. In the next step, the pool of speculative solutions is reduced to a smaller range of feasible solutions. This range has a very high probability to include the actual desired (“true”) spectra of  $a_d(\lambda)$  and  $a_g(\lambda)$  associated with the input  $a_{dg}(\lambda)$ . Finally, the optimal solution for  $a_d(\lambda)$  and  $a_g(\lambda)$  is selected from the pool of feasible solutions. This step aims at selecting one pair of  $a_d(\lambda)$  and  $a_g(\lambda)$  solutions which represents the closest agreement in a statistical sense with the true spectra of  $a_d(\lambda)$  and  $a_g(\lambda)$ .

We note that the overall approach of the ADG model depicted in Fig. 2.3 is similar to the approach utilized in Stramski *et al.* [73]. There are, however, several differences associated with the development dataset used to create the spectral shape function libraries, the approach to create these libraries, and some details related to the determination of feasible and optimal solutions.

#### 2.4.2 Libraries of Spectral Shape Functions of Constituent Absorption Coefficients

The first step in the formation of spectral shape function libraries,  $\hat{a}_d(\lambda)$  and  $\hat{a}_g(\lambda)$ , is the normalization of each spectrum of  $a_d(\lambda)$  and  $a_g(\lambda)$  in the development dataset by the spectral mean value of each relevant absorption coefficient:

$$\hat{a}_x(\lambda) = N_\lambda a_x(\lambda) \left[ \sum_{\lambda=\lambda_{\min}}^{700} a_x(\lambda) \right]^{-1} \quad (2.1)$$

where  $x$  is the subscript for either the non-algal particulate (d) or CDOM (g) absorption coefficient,  $N_\lambda$  is the number of discrete wavelengths included in the summation, and  $\lambda_{\min}$  is either 350 nm for the UV-VIS library or 400 nm for the VIS library. Because the development dataset consists of absorption spectra with a 1 nm spectral interval,  $N_\lambda$  is 351 for the UV-VIS library and 301 for the VIS library.

In the next step, for each spectral shape function of  $\hat{a}_d(\lambda)$  and  $\hat{a}_g(\lambda)$ , the spectral steepness parameters,  $S_x$ , are calculated within two non-overlapping spectral regions as:

$$S_x = \frac{\ln \hat{a}_x(\lambda_1) - \ln \hat{a}_x(\lambda_2)}{\lambda_2 - \lambda_1} \quad (2.2)$$

where  $\ln$  is the natural logarithm function, the first spectral region (blue) is delimited by  $\lambda_1 = 410$  nm and  $\lambda_2 = 460$  nm and the second spectral region (blue-green) by  $\lambda_1 = 480$  nm and  $\lambda_2 = 530$  nm. The use of these two non-overlapping wavelength ranges, 410–460 nm and 480–530 nm, for calculating the two spectral steepness parameters provides the flexibility to account for potential changes in the spectral shapes of  $a_d(\lambda)$  and  $a_g(\lambda)$  across a relatively broad blue-green spectral region between 410 nm and 530 nm where both  $a_d(\lambda)$  and  $a_g(\lambda)$  typically exhibit significant magnitudes and a general trend to decrease in magnitude with increasing wavelength. These choices are well-suited for library development in terms of spectral scales that capture the variations in spectral shape, especially the spectral slope, of  $a_d(\lambda)$  and  $a_g(\lambda)$  across the blue-green region. The steepness parameters drive the formation of the library of spectral shape functions  $\hat{a}_d(\lambda)$  and  $\hat{a}_g(\lambda)$ . Specifically, the blue and blue-green steepness parameters  $S_d$  and  $S_g$  serve as relatively simple metrics of the degree of similarity between the spectral shapes of  $a_d(\lambda)$  and  $a_g(\lambda)$  within the blue-green spectral region measurements in the development dataset. The values of  $S_d$  and  $S_g$  are used to classify all measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$  available in the development dataset into multiple spectral-shape classes with each class constrained by the similarity of the steepness parameters.

This classification is made using the joint probability distributions of the spectral steepness parameters which are depicted in Fig. 2.4. Each individual bin of the probability distributions includes measurements that are characterized by a similar pair of blue and blue-green steepness parameters. The characterization implies that the variation in the spectral shape of constituent absorption coefficients within any given single bin is sufficiently small, so that a single characteristic spectral shape function can be determined from data contained in that bin. Accordingly, for each bin we determined such characteristic spectral shape functions,  $\hat{a}_d(\lambda)$  and  $\hat{a}_g(\lambda)$ , by averaging all  $\hat{a}_d(\lambda)$  and  $\hat{a}_g(\lambda)$  spectra contained in the bin. The final libraries of spectral shapes are comprised of such characteristic spectral shape functions of  $\hat{a}_d(\lambda)$  and  $\hat{a}_g(\lambda)$  determined for each bin of joint probability distributions. Thus, the number of characteristic spectral shape functions of  $\hat{a}_d(\lambda)$  and  $\hat{a}_g(\lambda)$  is defined by the number of bins in the probability distributions shown in Fig. 2.4a and 2.4b, respectively. Specifically, the library of  $\hat{a}_d(\lambda)$  consists of 86 characteristic spectral shape functions which were created from 521 measurements of  $a_d(\lambda)$  available in the development dataset. The library of  $\hat{a}_g(\lambda)$  consists of 227 characteristic spectral shape functions which were created from 1293 measurements of  $a_g(\lambda)$ . It is notable that the data in Fig. 2.4 indicate the spectral steepness parameters are on average larger for  $S_g$  than for  $S_d$ . For example, in the 410–460 nm spectral range, the median values of  $S_g$  and  $S_d$  are  $0.0165 \text{ nm}^{-1}$  and  $0.0094 \text{ nm}^{-1}$ , respectively. For the 480–530 nm range, the corresponding median values are  $0.0158 \text{ nm}^{-1}$  and  $0.0069 \text{ nm}^{-1}$ .

Figure 2.5 depicts the spectra of  $\hat{a}_d(\lambda)$  and  $\hat{a}_g(\lambda)$  in both the UV-VIS and VIS libraries. We recall that the steepness parameters are defined within the visible spectral range (Eq. 2.2). Therefore, the joint probability distributions (Fig. 2.4) used to create the spectral shape libraries for a given absorption coefficient are the same for the UV-VIS and VIS libraries. The only

differences between the UV-VIS and VIS libraries for a given absorption coefficient is the overall spectral range covered by the characteristic spectral shape functions and the magnitude of these functions at any common wavelength because the normalization in Eq. (2.1) is associated with a different spectral range. Note also that the final number of characteristic shape functions in the  $\hat{a}_d(\lambda)$  and  $\hat{a}_g(\lambda)$  libraries represents approximately the same percentage ( $\sim 17\%$ ) of the total number of available spectra of  $a_d(\lambda)$  and  $a_g(\lambda)$  within the development dataset. This result essentially implies the postulation that the entire variability of spectral shape functions of  $\hat{a}_d(\lambda)$  and  $\hat{a}_g(\lambda)$  observed within the development dataset can be adequately represented by reducing the size of spectral shape dataset to 17%.

The binning method used to define the joint probability distributions shown in Fig. 2.4 naturally played a key role in determining the size of the spectral shape libraries. We adopted a discretization technique following the rule of Freedman and Diaconis [93], which determines the bin width in the middle 50% of the distribution based on the interquartile range of variables involved in the distribution. Within this range of the distribution the bin size is referred to as a 1x1 Freedman and Diaconis (FD) bin. We then increased the bin size of joint probability distribution as it extends away from its median value along the  $x$  or  $y$  axis. Specifically, the bin size was increased to 2x2 FD once the value on the  $x$  or  $y$  axis is more than one median absolute deviation from the distribution's median, and 4x4 FD once the value on the  $x$  or  $y$  axis is more than two median absolute deviations from the distribution's median. The purpose of such increase in bin size with increasing distance from the median of the distribution is to account for the fact that fewer spectra exist in the peripheral space of the distribution, and it is generally preferred to incorporate more than a single spectrum into a single bin. The average coefficient of variation of  $\hat{a}_d(\lambda)$  and  $\hat{a}_g(\lambda)$  within individual bins for both libraries was found to be consistently below 6% in

the spectral range from 400 to 550 nm which provides support that the adopted binning scheme is appropriate.

### 2.4.3 Speculative Solutions of the ADG Model

In the first step of the ADG model operation, a large pool of speculative solutions of output  $a_d(\lambda)$  and  $a_g(\lambda)$  is generated from given input of  $a_{dg}(\lambda)$  by implementing the spectral shape function libraries of  $\hat{a}_d(\lambda)$  and  $\hat{a}_g(\lambda)$  (Fig. 2.3). The ADG\_UV-VIS model uses UV-VIS libraries and generates speculative solutions over the spectral range from 350 to 700 nm. The VIS-UV<sub>Ext</sub> model uses VIS libraries and generates speculative solutions over the spectral range from 400 to 700 nm. The generation of speculative solutions involves the use of weighting factors defined as the ratio of the spectral mean of either  $a_d(\lambda)$  or  $a_g(\lambda)$  to the spectral mean of  $a_{dg}(\lambda)$ :

$$w = \left( \sum_{\lambda=\lambda_{\min}}^{700} a_g(\lambda) \right) \left( \sum_{\lambda=\lambda_{\min}}^{700} a_{dg}(\lambda) \right)^{-1} \quad (2.3)$$

$$w' = \left( \sum_{\lambda=\lambda_{\min}}^{700} a_d(\lambda) \right) \left( \sum_{\lambda=\lambda_{\min}}^{700} a_{dg}(\lambda) \right)^{-1} \quad (2.4)$$

where  $w$  and  $w'$  are the weighting factors for  $a_g(\lambda)$  and  $a_d(\lambda)$  respectively and range from 0 to 1 while satisfying the condition  $w + w' = 1$ , and  $\lambda_{\min}$  is either 350 nm for ADG\_UV-VIS or 400 nm for ADG\_VIS-UV<sub>Ext</sub>. From Equations (2.1), (2.3), and (2.4) we obtain:

$$\hat{a}_{dg}(\lambda) = (1 - w) \hat{a}_d(\lambda) + w \hat{a}_g(\lambda) \quad (2.5)$$

where  $\hat{a}_{dg}(\lambda)$  is the spectral shape function of the non-phytoplankton absorption coefficient,  $a_{dg}(\lambda)$ , which is defined analogously to Equation (2.1):

$$\hat{a}_{dg}(\lambda) = N_{\lambda} a_{dg}(\lambda) \left( \sum_{\lambda=\lambda_{\min}}^{700} a_{dg}(\lambda) \right)^{-1} \quad (2.6)$$

Equation (2.5) can be rewritten as a system of linear equations:

$$\mathbf{A} \times w = \mathbf{B} \quad (2.7)$$

where  $\mathbf{A}$  and  $\mathbf{B}$  are vectors containing 351 data points of  $\hat{a}_g(\lambda) - \hat{a}_d(\lambda)$  and  $\hat{a}_{dg}(\lambda) - \hat{a}_d(\lambda)$  respectively for the UV-VIS model or 301 analogous data points for the VIS-UV<sub>Ext</sub> model.

Speculative solutions for a given input  $a_{dg}(\lambda)$  are computed by first calculating the spectral shape function  $\hat{a}_{dg}(\lambda)$  from Equation (2.6). The weighting factor  $w$  is then calculated for every pair of spectral shape functions of  $\hat{a}_d(\lambda)$  and  $\hat{a}_g(\lambda)$  from a given library by solving Equation (2.7) for  $w$  using the \ (mldivide) operator in MATLAB. Only solutions for  $w > 0$  and  $< 1$  are selected as speculative solutions. The speculative solutions with valid  $w$  and  $w' (\equiv 1 - w)$  associated with each pair of spectral shape functions are then utilized to calculate a pair of speculative solutions for non-algal particulate absorption coefficient,  $a'_d(\lambda)$ , and CDOM absorption coefficient,  $a'_g(\lambda)$ :

$$a'_d(\lambda) = N_\lambda^{-1} (1 - w) \hat{a}_d(\lambda) \sum_{\lambda=\lambda_{\min}}^{700} a_{dg}(\lambda) \quad (2.8)$$

$$a'_g(\lambda) = N_\lambda^{-1} w \hat{a}_g(\lambda) \sum_{\lambda=\lambda_{\min}}^{700} a_{dg}(\lambda) \quad (2.9)$$

The maximum possible number of speculative solutions for a given input of  $a_{dg}(\lambda)$  is 19522 which corresponds to a computational scenario when all weighting factors obtained by solving Equation (2.7) are bounded within the physically realistic range between 0 and 1. This maximum number of speculative solutions is equivalent to all possible combinations of the 86 spectral shape functions of  $\hat{a}_d(\lambda)$  and the 227 shape functions of  $\hat{a}_g(\lambda)$ . Our analysis of both model variants applied to the subset of the development dataset containing 390 concurrent measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$  found the distribution of the number of speculative solutions to be left-skewed. The median and mean number of speculative solutions for ADG\_UV-VIS is 15170 and 13420 (SD = 5280) respectively. The median and mean number of speculative solutions for ADG\_VIS-UV<sub>Ext</sub> is 16500 and 15120 (SD = 4041) respectively.

#### 2.4.4 Feasible and Optimal Solutions of the ADG Model

The subsequent step of the ADG model involves the selection of solutions, referred to as feasible solutions, from the pool of speculative solutions  $a'_d(\lambda)$  and  $a'_g(\lambda)$  (Fig. 2.3). We recall that the main purpose of this step is to create a pool of feasible solutions which is significantly smaller than the pool of speculative solutions while ensuring that the desired (“true”) spectra of  $a_d(\lambda)$  and  $a_g(\lambda)$  fall within a range of feasible solutions with a very high probability. Feasible solutions are determined by comparing the sum of all pairs of speculative solutions  $a'_d(\lambda) + a'_g(\lambda)$  to input  $a_{dg}(\lambda)$ , which means that all speculative solutions  $a'_{dg}(\lambda)$  are compared with input  $a_{dg}(\lambda)$ .

The best possible outcome for  $a'_{dg}(\lambda)$  is one where the summed pair of speculative solutions  $a'_d(\lambda) + a'_g(\lambda)$  is equal to input  $a_{dg}(\lambda)$  at all wavelengths. This perfect scenario is, however, highly unlikely. Thus, the method for identifying feasible solutions involves analyzing the pool of speculative solutions and selecting those that fulfill the criterion that minimizes the differences between  $a'_{dg}(\lambda)$  and  $a_{dg}(\lambda)$  at multiple wavelengths.

Specifically, we used the statistical metric of the normalized sum of squared residuals (NSSR) defined as:

$$\text{NSSR} = \sum_{\lambda=\lambda_{\min}}^{600} \left( \frac{a'_d(\lambda) + a'_g(\lambda) - a_{dg}(\lambda)}{a_{dg}(\lambda)} \right)^2 \quad (2.10)$$

We chose to exclude wavelengths longer than 600 nm because the magnitude of  $a_{dg}(\lambda)$  in the red portion of the spectrum is typically small and subject to relatively low values of signal-to-noise. Like in the previous considerations,  $\lambda_{\min}$  in Equation (2.10) is either 350 nm for the ADG\_UV-VIS variant of the ADG model or 400 nm for the ADG\_VIS-UV<sub>Ext</sub> variant of the ADG model. To determine the pool of feasible solutions we calculated NSSR for every pair of speculative solutions and identified the pairs of speculative solutions which have the smallest NSSR. For a given input spectrum of  $a_{dg}(\lambda)$ , our analysis demonstrated the distribution of NSSR is consistently skewed to

the right independent of the input spectrum (not shown). This result indicates that multiple speculative solutions producing relatively low values of NSSR show relatively high degree of similarity between one another, and thus this subset of speculative solutions is well suited for the determination of feasible solutions. Specifically, the pairs of speculative solutions of  $a'_d(\lambda)$  and  $a'_g(\lambda)$  which fall within the lowest 10<sup>th</sup> percentile of the NSSR distribution were selected as feasible solutions. Given the maximum possible number of speculative solutions is 19522, the corresponding maximum number of feasible solutions is 1952. From our analysis of ADG model solutions with the subset of the development dataset containing 390 concurrent measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$ , we found the number of feasible solutions ranges from 78 to 1933. The distribution of the total number of feasible solutions is analogous to the distribution of speculative solutions (since they are the lowest 10<sup>th</sup> percentile of speculative solutions) and is left-skewed. As a consequence, the statistics which characterize the distribution of feasible solutions are one-tenth the values reported for speculative solutions, so the median and mean number of feasible solutions of 1517 and 1342 for the ADG\_UV-VIS model, respectively. For the ADG\_VIS-UV<sub>Ext</sub> model, these median and mean values are 1650 and 1512, respectively. Our analysis of both the development and validation datasets indicated that the range of feasible solutions for  $a_d(\lambda)$  and  $a_g(\lambda)$  derived from both model variants has a very high probability (>95%) of containing the measured values.

The final step of the ADG model is to select the pair of feasible solutions that have the highest likelihood of matching the true values of  $a_d(\lambda)$  and  $a_g(\lambda)$  as optimal solutions for model output. A reasonable selection for the pair of feasible solutions to output as optimal solutions are those with the minimum NSSR produced through the evaluation of speculative solutions in Equation (2.10). However, it is unlikely that this pair of feasible solutions will perfectly match the

actual values of  $a_d(\lambda)$  and  $a_g(\lambda)$ , so we selected an approach that incorporates the entire distribution of feasible solutions we obtained. We chose the spectral median (i.e., the 50<sup>th</sup> percentile) of feasible solutions as output for the optimal solution for the ADG model and this selection of median values was made at each wavelength independently. Figure 2.6 displays the differences in the aggregate error metrics of  $MdR$  and  $MdSA$  compared to measured values of  $a_d(\lambda)$  and  $a_g(\lambda)$  as well as the resultant  $a_{dg}(\lambda)$  when selecting different percentiles from the pool of feasible solutions for both ADG\_UV-VIS (solid lines) and ADG\_VIS-UV<sub>Ext</sub> (dashed lines) at 350, 440, 550, and 660 nm. The results presented in Fig. 2.6 show  $MdR$  increases for all four wavelengths with increasing percentile that approach the optimal  $MdR$  value of 1 near the 50<sup>th</sup> percentile of feasible solutions for  $a_d(\lambda)$  and  $a_g(\lambda)$  (Fig. 2.6a–c). Furthermore,  $MdSA$  values reach a minimal value near the 50<sup>th</sup> percentile of feasible solutions for all constituent absorption coefficients for all wavelengths. These results provide justification in support of the selection of the spectral median of feasible solutions as the optimal solution for model output.

## 2.5 Assessment of Model Performance

### 2.5.1 Evaluation with the Development Dataset

The fundamental difference between the ADG\_UV-VIS and ADG\_VIS-UV<sub>Ext</sub> variants of the ADG partitioning model is the spectral range of the libraries used to determine the speculative, feasible, and optimal solutions. It is of interest to examine the performance of both model variants because they may exhibit different advantages or disadvantages depending on the specific scenario of model application. For example, the ADG\_UV-VIS model has the benefit of directly providing estimates of  $a_d(\lambda)$  and  $a_g(\lambda)$  in the spectral range 350–700 nm, thus including the near-UV. However, this model may have a disadvantage in the context of remote sensing applications when the ADG model is intended to be applied as the last component of a multi-step algorithm that begins with an inverse reflectance model to first derive the total absorption coefficient  $a(\lambda)$ , which

is then followed by ANW and ADG absorption partitioning models. It is conceivable that in this scenario, the ADG\_UV-VIS model may be impacted by larger uncertainties of input  $a_{dg}(\lambda)$  in the UV than VIS due to larger uncertainties in the satellite-derived reflectance in the UV than VIS, which then propagate downstream through the multi-step algorithm. The ADG\_VIS-UV<sub>Ext</sub> model, which is optimized for the VIS but still enables the estimation in the near-UV through extrapolation from the blue part of VIS, may circumvent this potential issue as the ADG model performance will not be directly impacted by these potentially larger uncertainties in the near-UV satellite-derived  $R_{rs}(\lambda)$  and the downstream near-UV products of  $a(\lambda)$  and  $a_{dg}(\lambda)$ .

The subset of our development dataset containing concurrent measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$ , and hence  $a_{dg}(\lambda)$  ( $N = 390$ ) can be used to evaluate the general performance of each ADG model. It is important to note that although these data were utilized in the construction of the spectral shape function libraries, the libraries were determined from a much larger dataset than these 390 pairs of spectra. The comparison of model-derived optimal solutions and measured values of absorption coefficients for the ADG\_UV-VIS model is presented for three example wavelengths from the VIS range (443, 555, and 670 nm) in Fig. 2.7. The model demonstrates a strong performance at all three wavelengths with relatively even distribution of data points about the 1:1 line suggesting no tendency for bias across the dynamic range of  $a_d(\lambda)$  and  $a_g(\lambda)$ . However, the scatter of data points around the 1:1 line for  $a_d(\lambda)$  increases as wavelength decreases, especially in the range of relatively low values of  $a_d(\lambda)$  (Fig. 2.7 a-c). In contrast, the scatter of data points of  $a_g(\lambda)$  increases as wavelength increases, and the spread of points is less dependent on the magnitude of  $a_g(\lambda)$  (Fig. 2.7d-f). These findings suggest that the model performs better at predicting either  $a_d(\lambda)$  or  $a_g(\lambda)$  in spectral regions where a given constituent absorption coefficient tends to dominate  $a_{dg}(\lambda)$ . This is particularly well-pronounced when comparing the data points for

$a_g(\lambda)$  (Fig. 2.7d) with data points for  $a_d(\lambda)$  (Fig. 2.7a) in the blue spectral region where  $a_g(\lambda)$  has generally a dominant contribution to  $a_{dg}(\lambda)$ .

Figure 2.8 depicts the performance of the ADG\_UV-VIS model at two wavelengths from the near-UV spectral region, 350 and 380 nm. The scatter plots of model-derived optimal solutions vs. measured values of constituent absorption coefficients support the notion that the model continues to yield good results in the near-UV. It is notable, however, that while the model estimates of  $a_g(\lambda)$  at the near-UV wavelengths align closely with the 1:1 line across the entire dynamic range (Fig. 2.8c,d), the model-derived values of  $a_d(\lambda)$  exhibit significantly larger scatter around the 1:1 line albeit with no tendency for bias (Fig. 2.8a,b). An analogous analysis of the ADG\_VIS-UV<sub>Ext</sub> model in the VIS and near-UV provided generally similar patterns in the scatter plots of model-derived vs. measured values of constituent absorption coefficients (not shown).

Figure 2.9 depicts two statistical metrics,  $MdR$  and  $MdSA$ , which characterize the performance of ADG\_UV-VIS and ADG\_VIS-UV<sub>Ext</sub> in terms of estimating  $a_d(\lambda)$  and  $a_g(\lambda)$  across the entire spectral range of model output. These results were calculated for the subset of development dataset containing concurrent measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$  and they generally demonstrate good performance of both ADG models. The spectral values of  $MdR$  are close to 1 (generally within a few percent) indicating negligible or small aggregate bias of model-derived absorption coefficients (Fig. 2.9a,b). It is notable, however, that ADG\_UV-VIS yields the values of  $MdR$  generally somewhat closer to 1 than ADG\_VIS-UV<sub>Ext</sub>, except for the  $a_d(\lambda)$  estimation in the red part of the spectrum. The difference in  $MdR$  values between the two model variants is most pronounced for model-derived  $a_d(\lambda)$  in the blue part of the spectrum where ADG\_VIS-UV<sub>Ext</sub> produces the highest departure of  $MdR$  from 1 with a positive bias of up to about 5% (Fig. 2.9a).

For the estimation of  $a_g(\lambda)$ , the differences between the model variants in terms of *MdR* values are small across the entire spectral range, < 2% at most wavelengths (Fig. 2.9b).

The *MdSA* statistics support the generally good performance of both ADG model variants in terms of aggregate percent uncertainty (Fig. 2.9c,d). For the most part the spectral values of *MdSA* are below 20%. The highest value of *MdSA* is about 31% in the case of  $a_d(\lambda)$  estimation by ADG\_UV-VIS at the short-wavelength end (350 nm) of the examined spectral range. The lowest values of *MdSA* are below 10% in the case of  $a_g(\lambda)$  estimation by both models within the near-UV and blue portions of the spectrum and in the case of  $a_d(\lambda)$  estimation by ADG\_VIS-UV<sub>Ext</sub> in the red part of the spectrum. There exist clear spectral trends and differences in *MdSA* values between ADG\_UV-VIS and ADG\_VIS-UV<sub>Ext</sub> (Fig. 2.9c,d). Specifically, *MdSA* of  $a_d(\lambda)$  decreases with increasing wavelength and the opposite spectral trend is observed for  $a_g(\lambda)$ . There is one minor exception in the spectral trend of  $a_g(\lambda)$  estimation by ADG\_VIS-UV<sub>Ext</sub> in the near-UV but in this spectral region the *MdSA* values remain low below 10% regardless of the model. The general spectral trends of *MdSA* depicted in Fig. 2.9c,d can be attributed to typical patterns of wavelength dependence of relative contributions of  $a_d(\lambda)$  and  $a_g(\lambda)$  to  $a_{dg}(\lambda)$ ; namely, while  $a_g(\lambda)$  tends to be more significant than  $a_d(\lambda)$  at short wavelengths, the opposite situation is typical for the long-wavelength portion of visible spectrum. With regards to differences in *MdSA* between the two model variants, ADG\_VIS-UV<sub>Ext</sub> generally yields lower *MdSA* (by about 5% ± a few percent) than ADG\_UV-VIS. An exception is observed for the  $a_g(\lambda)$  estimation in the near-UV, but again it is of minor significance because *MdSA* in this spectral region remains below 10% for both models (Fig. 2.9d).

Table 2.3 provides statistical metrics of performance of ADG\_UV-VIS and ADG\_VIS-UV<sub>Ext</sub> for several wavelengths across the examined spectral range, namely 350, 380, 443, 555, and

670 nm. These statistics include *MdR* and *MdSA* as well as other metrics that characterize the degree of agreement between the model-derived optimal solutions and measured values of  $a_d(\lambda)$  and  $a_g(\lambda)$ , which are described in Section 3 and Table 2.2. We do not provide the statistics based on comparison of the input  $a_{dg}(\lambda)$  with the sum of model outputs of  $a_d(\lambda)$  and  $a_g(\lambda)$  because the agreement for  $a_{dg}(\lambda)$  is very good by virtue of the ADG model formulation. Overall, the results in Fig. 2.9 and Table 2.3 provide no clear indication that ADG\_UV-VIS can be considered superior to ADG\_VIS-UV<sub>Ext</sub> or vice versa. Each of these two models may have slight advantage over the other model depending on the constituent absorption coefficient, statistical metrics, and/or the spectral region being considered. Both models show the best predictive capability for  $a_g(\lambda)$  in the near-UV and blue spectral regions although ADG\_VIS-UV<sub>Ext</sub> may be generally preferable given consistently lower *MdSA*, *MdAPD*, and *RMSD* over the visible spectral range compared to ADG\_UV-VIS. For predicting  $a_d(\lambda)$ , ADG\_VIS-UV<sub>Ext</sub> also appears to provide some advantage in terms of *MdSA*, *MdAPD* and *RMSD* statistics; however, this model has somewhat inferior measures of bias in the blue part of the spectrum compared with ADG\_UV-VIS.

Because this assessment of model performance has been made with the subset of field dataset that was used to create the libraries of the spectral shape functions,  $\hat{a}_d(\lambda)$  and  $\hat{a}_g(\lambda)$ , there is a need for further evaluation of ADG models with datasets that are independent of those used to develop the spectral shape libraries. Such an initial effort is described in Section 5.2.

### **2.5.2 Evaluation with the Independent Validation Dataset**

Figure 2.10 and Table 2.4 present the results of performance assessment of ADG\_UV-VIS and ADG\_VIS-UV<sub>Ext</sub>, which was conducted using the independent validation dataset consisting of 149 subsurface measurements (i.e., samples obtained at depths between 15 m and 50 m) of  $a_d(\lambda)$  and  $a_g(\lambda)$  coefficients (see Section 2.2 for the description of this dataset). These results indicate

that the statistical parameters of model performance are either very similar or slightly degraded compared to the analysis with the subset of development dataset presented in Section 5.1. The most pronounced manifestation of this degradation is observed for the *MdR* statistic, especially in the case of  $a_d(\lambda)$  estimation in the short-wavelength portion of the spectrum where *MdR* values exceed 1.1 (Fig. 2.10a). This implies an aggregate positive bias with a median value >10% which is markedly higher compared to the analysis with the development dataset displayed in Fig. 2.9a where *MdR* reaches a value of about 1.05 (5% bias) in the worst case of  $a_d(\lambda)$  estimation in the blue spectral range from ADG\_VIS-UV<sub>Ext</sub>. The *MdR* values for the  $a_g(\lambda)$  estimation based on the analysis of independent validation dataset (Fig. 2.10b) are generally below 1 and exhibit somewhat stronger tendency for model underestimation compared with analogous results obtained with the development dataset (Fig. 2.9b). The differences in the *MdSA* statistic between the analyses of independent dataset (Fig. 2.10c,d) and development dataset (Fig. 2.9c,d) are very small. It is also notable that these small differences do not necessarily indicate a tendency of *MdSA* to increase when the analysis is made with the independent dataset compared with the development dataset. In fact, the model estimates of  $a_d(\lambda)$  tend to have slightly improved (i.e., lower) *MdSA* when tested with the independent dataset (Fig. 2.10c) compared with the results for the development dataset (Fig. 2.9c).

Comparison of ADG\_UV-VIS and ADG\_VIS-UV<sub>Ext</sub> in terms of their performance statistics obtained from the analysis of independent validation dataset (Fig. 2.10, Table 2.4) suggests that ADG\_VIS-UV<sub>Ext</sub> tends to generally provide slightly better estimates of both  $a_d(\lambda)$  and  $a_g(\lambda)$  for most of the examined spectral region. It is noteworthy that the spectral values of *MdR* associated with ADG\_VIS-UV<sub>Ext</sub> are generally closer to 1 than those associated with ADG\_UV-VIS (Fig. 10a,b). We recall that the opposite tendency was observed in the analysis of development

dataset (Fig. 2.9a,b). The *MdSA* statistic is also generally slightly better for ADG\_VIS-UV<sub>Ext</sub> compared with ADG\_UV-VIS (Fig. 2.10c,d) and this tendency is similar to analogous results obtained with the development dataset (Fig. 2.9c,d). Only at near-UV wavelengths close to the short-wavelength end of the spectrum, *MdSA* and a few other statistics shown in Tables 2.3 and 2.4 are somewhat better for ADG\_UV-VIS than ADG\_VIS-UV<sub>Ext</sub>. Thus, if one of the two model variants were to be selected for use across the entire examined spectral range from 350 to 700 nm, the results of model assessment presented in Figs. 2.9, 2.10 and Tables 2.3, 2.4 appear to be in favor of ADG\_VIS-UV<sub>Ext</sub>. However, because the differences in the performance statistics between the two variants of the ADG model are mostly minor, our results may also support the appropriateness of ADG\_UV-VIS for applications, especially under the circumstances when the uncertainty in the input data of  $a_{dg}(\lambda)$  exhibits no significant increase as the light wavelength decreases from the VIS into the UV spectral region.

## 2.6 Summary and Conclusions

This study describes the ADG partitioning model that separates the spectral absorption coefficients of non-algal (depigmented) particulate matter,  $a_d(\lambda)$ , and chromophoric dissolved organic matter (CDOM),  $a_g(\lambda)$ , from the spectral non-phytoplankton absorption coefficient,  $a_{dg}(\lambda)$ , which represents the sole input to the model. Historically, this absorption partitioning problem has proven highly challenging, mainly because of similarity in the spectral shapes of  $a_d(\lambda)$  and  $a_g(\lambda)$ . Previous modeling approaches have been commonly limited by restrictive assumptions and generalizations about the spectral shape of these constituent absorption coefficients. In particular, it has been commonly assumed that the spectral behavior of  $a_d(\lambda)$  and  $a_g(\lambda)$  can be described by a single exponential function of light wavelength over a broad spectral range. However, there is ample experimental evidence that the actual spectra of  $a_d(\lambda)$  and  $a_g(\lambda)$  often deviate considerably

from such exponential approximation. Recently, Stramski *et al.* [73] developed an ADG model that aims to circumvent these limitations by relaxing the restrictive assumptions about the spectral shapes of  $a_d(\lambda)$  and  $a_g(\lambda)$ . Specifically, this model was designed to use the spectral shape function libraries representing the variability in the measured spectral shapes of  $a_d(\lambda)$  and  $a_g(\lambda)$  with no additional assumptions about these coefficients.

In this study, we presented a new ADG partitioning model that builds upon and enhances the concepts and model formulation used in Stramski *et al.* [73]. The foundation of the ADG model is provided by the spectral shape function libraries of  $a_d(\lambda)$  and  $a_g(\lambda)$  which are created from near-surface measurements of these coefficients collected in diverse oceanic environments. For this purpose, the new ADG model implements a newly assembled field dataset consisting of hyperspectral measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$  covering the near-UV and VIS spectral range. These measurements were subject to stringent method-related and data quality control inclusion criteria. This process was to ensure consistent quality of data; for example, we excluded hyperspectral measurements of  $a_d(\lambda)$  with a spectrophotometric filter-pad method in transmittance configuration, which has long been used but is known to involve the biasing effects on  $a_d(\lambda)$  determinations. Instead, we used only the  $a_d(\lambda)$  data obtained with a superior filter-pad method that is based on the inside integrating-sphere configuration of measurement [74,75].

We created two versions of spectral shape function libraries of  $a_d(\lambda)$  and  $a_g(\lambda)$ , one covering the spectral range from 350 to 700 nm (UV-VIS libraries) and the other covering the range from 400 to 700 nm (VIS libraries). Accordingly, we developed two variants of the ADG model. The variant referred to as ADG\_UV-VIS determines solutions directly within the near-UV and VIS by utilizing the UV-VIS libraries. The variant referred to as ADG\_VIS determines solutions in the VIS by utilizing the VIS libraries. The ADG\_VIS model can, however, be coupled with an

independent extrapolation model described in Kehrli *et al.* [79] to extend output from the VIS to the near-UV, and this specific model variant is referred to as ADG\_VIS-UV<sub>Ext</sub>. In addition to the use of a newly assembled field dataset, the key enhancements of the new ADG model compared to the version described in Stramski *et al.* [73] include an improved formulation of spectral shape functions of constituent absorption coefficients which better accounts for actual variability observed within the blue-green spectral region as well as spectral extension of model output to include the near-UV (i.e., the 350–400 nm range) in addition to the VIS portion of the spectrum. In this study, we also evaluated the performance of both ADG\_UV-VIS and ADG\_VIS-UV<sub>Ext</sub> model variants using two datasets. The first contains concurrent near-surface measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$  extracted from the model development dataset that was used to create the spectral shape function libraries. The second is an independent validation dataset of concurrent sub-surface (depth range 15 – 50 m) measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$ . These analyses demonstrated an overall good performance of both variants of the ADG model. The optimal model solutions of  $a_d(\lambda)$  and  $a_g(\lambda)$  were generally characterized by an aggregate percent difference relative to measurements of less than 20% and a negligible or small aggregate bias within a few percent over the majority of examined spectrum. The new ADG model also provides a range of feasible solutions which encompass the reference measurements with a very high probability that exceeds 95%.

We anticipate the ADG partitioning model developed in this study can play a unique role as a component model of a multi-step inverse optical algorithm applicable in the context of ocean color remote sensing observations. The envisioned multi-step algorithm will begin with an inverse reflectance model that derives the total absorption coefficient of seawater from ocean reflectance such as satellite-derived reflectance [43], which is then followed by the first absorption partitioning model that derives the phytoplankton,  $a_{ph}(\lambda)$ , and non-phytoplankton,  $a_{dg}(\lambda)$ , absorption

coefficients (e.g., see [56]), and finally the ADG absorption partitioning model presented in this study. This multi-step approach can produce a full suite of main constituent absorption coefficients with hyperspectral coverage over a broad spectral range, which offers an important capability for NASA's PACE mission that will, for the first time, provide satellite ocean color measurements with hyperspectral coverage from the near-UV through VIS into the near-IR [84]. We conclude that the ADG model described in this study has reached the level of readiness for implementation into such multi-step algorithm approach or other potential types of ocean color algorithms that are commensurate with the purpose of ADG model.

It is also noteworthy that the two variants of ADG model, ADG\_UV-VIS and ADG\_VIS-UV<sub>Ext</sub>, provide an added value of flexibility for use in various application scenarios. Although the analyses in this study suggest that both model variants can be appropriate for applications to estimate  $a_d(\lambda)$  and  $a_g(\lambda)$  in the near-UV and VIS spectral range from the sole input of  $a_{dg}(\lambda)$ , the choice of specific model variant may depend on specific application scenario. For example, if the uncertainty in the spectral values of input  $a_{dg}(\lambda)$  in the UV is known to be substantially higher than the uncertainty in the VIS, then ADG\_VIS-UV<sub>Ext</sub> would be a preferred option to provide the model output from the near-UV through VIS. Consideration of such scenario appears to be particularly important in the context of near-future applications of the multi-step algorithm to satellite ocean color measurements obtained from Ocean Color Instrument (OCI) during the PACE mission. This mission will provide the data product of ocean reflectance extending to the near-UV for the first time, so the quantification of uncertainties in the near-UV reflectance is yet to be established.

In closing, we call attention to a need for further validation studies of the proposed ADG model, especially as more high-quality hyperspectral data of constituent absorption coefficients are collected in diverse oceanic environments with the best currently available and recommended

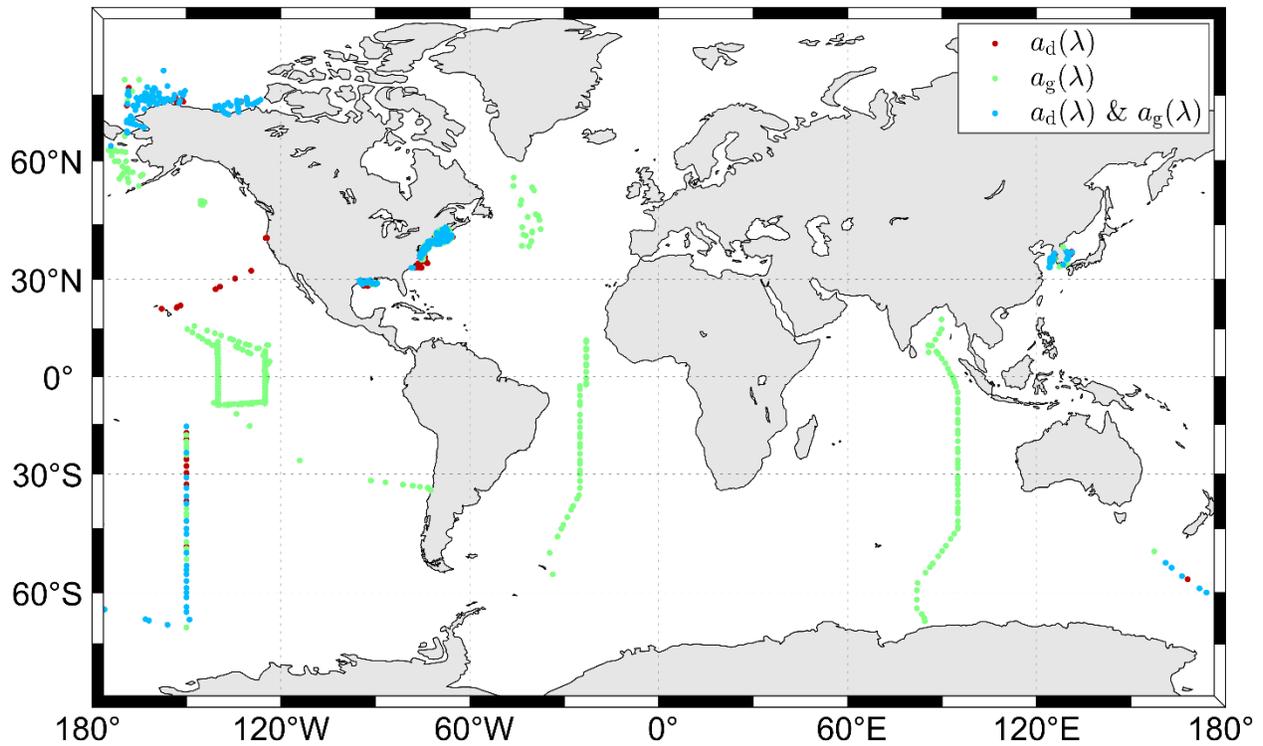
methods. Likewise, the implementation of the ADG model as a component model of comprehensive ocean color algorithms, such as the multi-step algorithm mentioned in this study, will require further dedicated work to validate and assess the algorithm performance.

## **2.7 Acknowledgements**

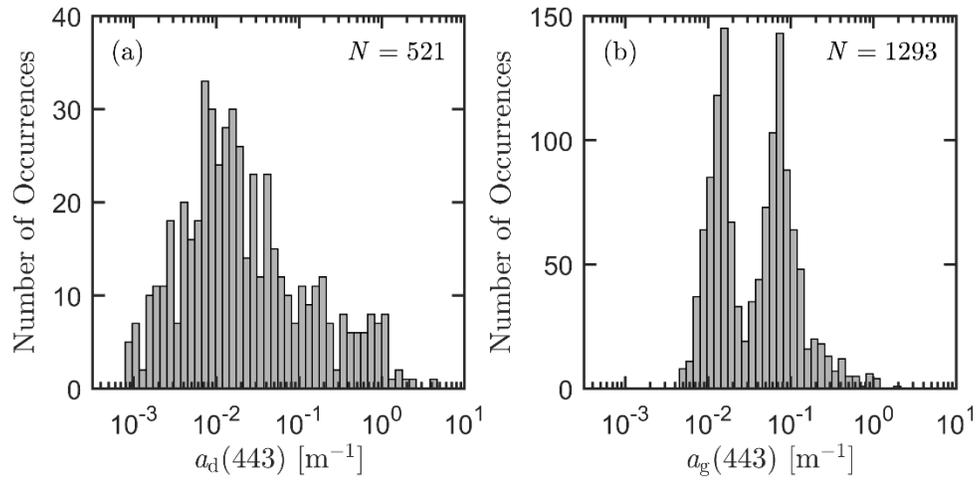
We thank all investigators who contributed to the collection, processing, and distribution through public databases of the field data of absorption coefficients used in this study. The data sources used in this study are available on NASA's SeaWiFS Bio-optical Archive and Storage System (SeaBASS) and the Biogeochemistry and Optics South Pacific Experiment (BIOSOPE) database. We also thank Michael Novak for comments on the manuscript.

Chapter 2, in full, is a reprint of the material as it appears in *Applied Optics*. The dissertation author was the primary investigator and author of this paper: Kehrli, M. D., Stramski, D., Reynolds, R. A., & Joshi, I. D. (2024). Model for partitioning the non-phytoplankton absorption coefficient of seawater in the ultraviolet and visible spectral range into the contributions of non-algal particulate and dissolved organic matter. *Applied Optics* (2004), 63(16), 4252–4270. <https://doi.org/10.1364/AO.517706>.

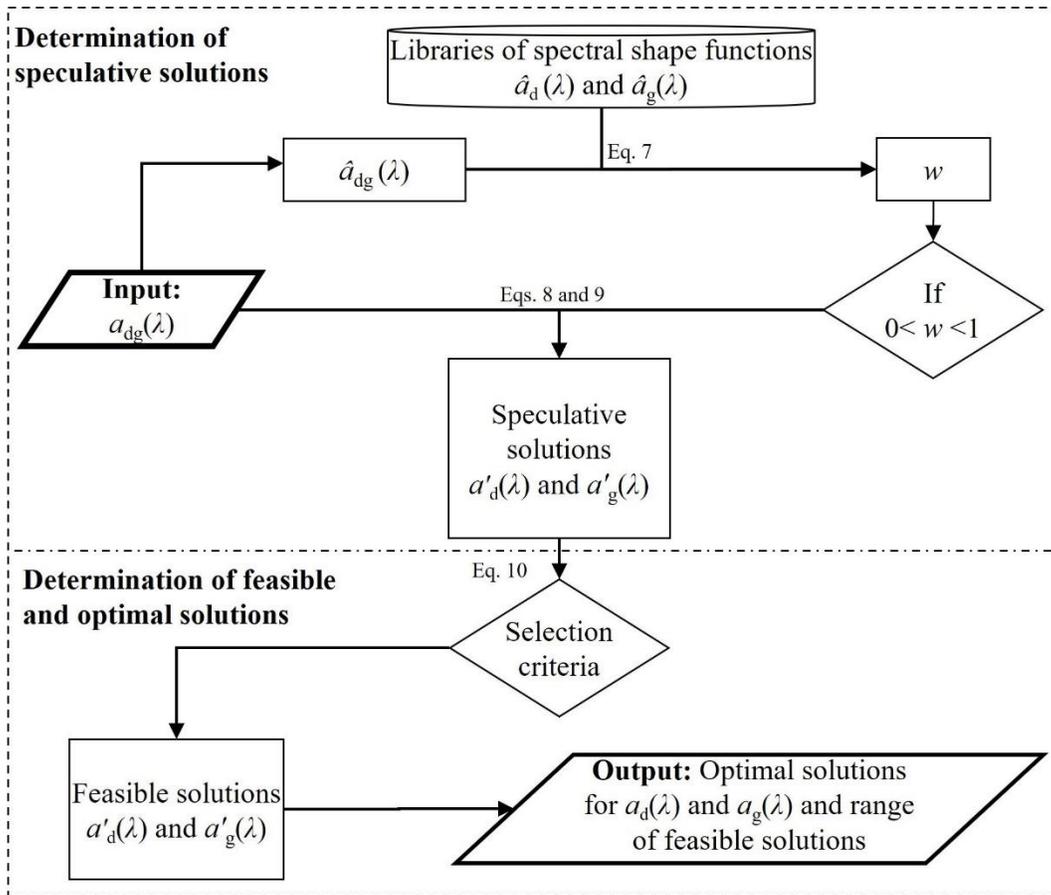
## 2.8 Figures



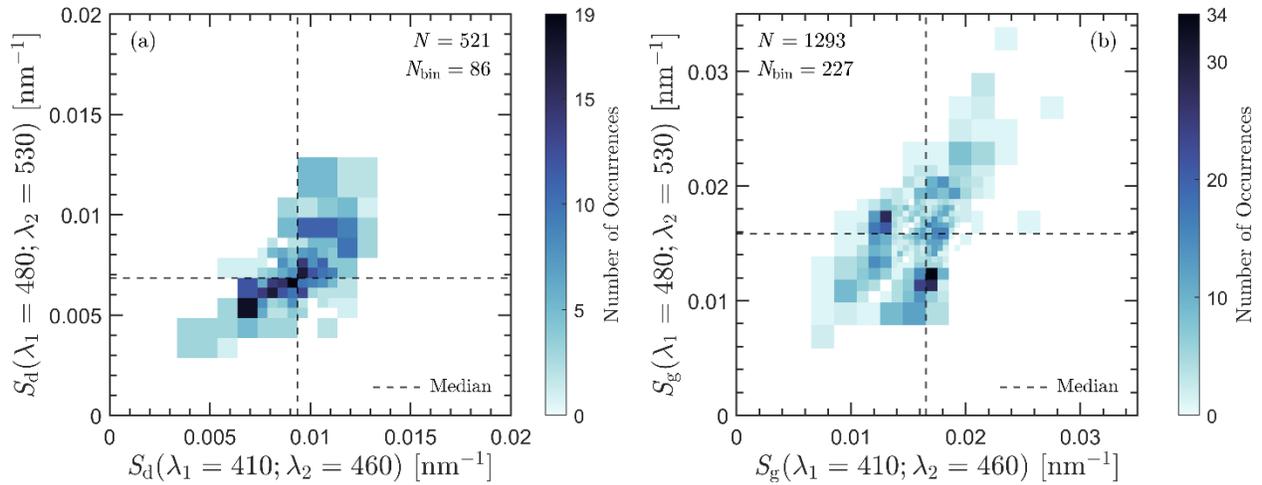
**Figure 2.1.** Global map depicting geographic locations of near-surface measurements comprising the final development dataset utilized in this study. Red markers are stations where only  $a_d(\lambda)$  measurements were collected ( $N = 131$ ), green markers are stations where only  $a_g(\lambda)$  measurements were collected ( $N = 903$ ), and blue markers are stations with concurrent measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$  ( $N = 390$ ).



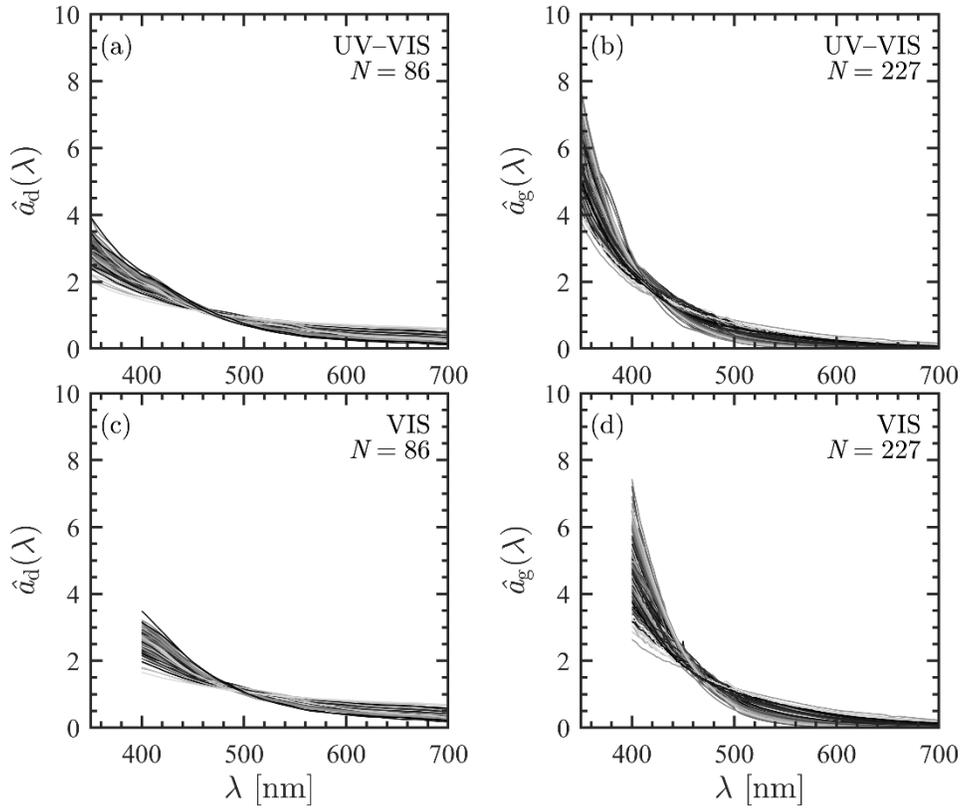
**Figure 2.2.** Histograms of  $a_d(443)$  and  $a_g(443)$  from the final development dataset.



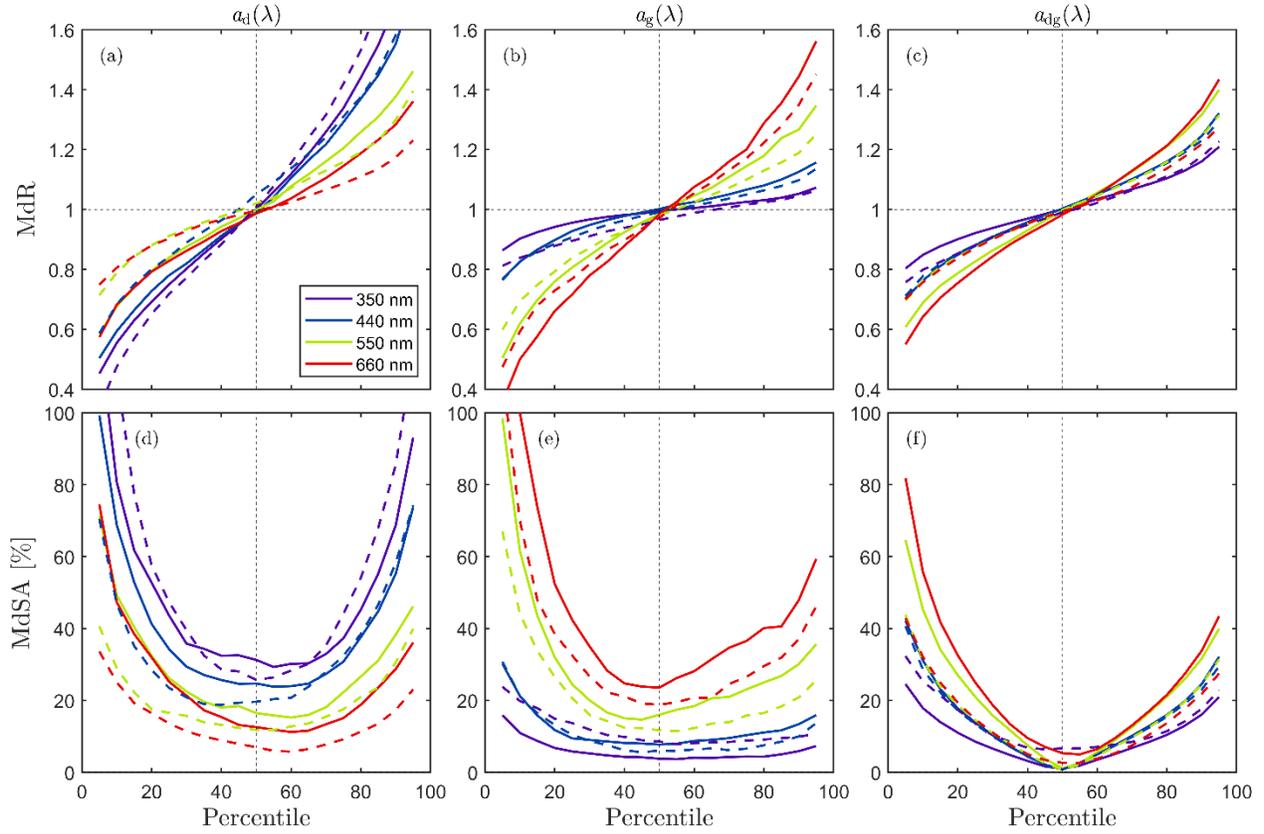
**Figure 2.3.** Flowchart depicting the ADG partitioning model to partition the non-phytoplankton absorption coefficient,  $a_{dg}(\lambda)$ , into its non-algal,  $a_d(\lambda)$ , and color dissolved organic matter,  $a_g(\lambda)$ .



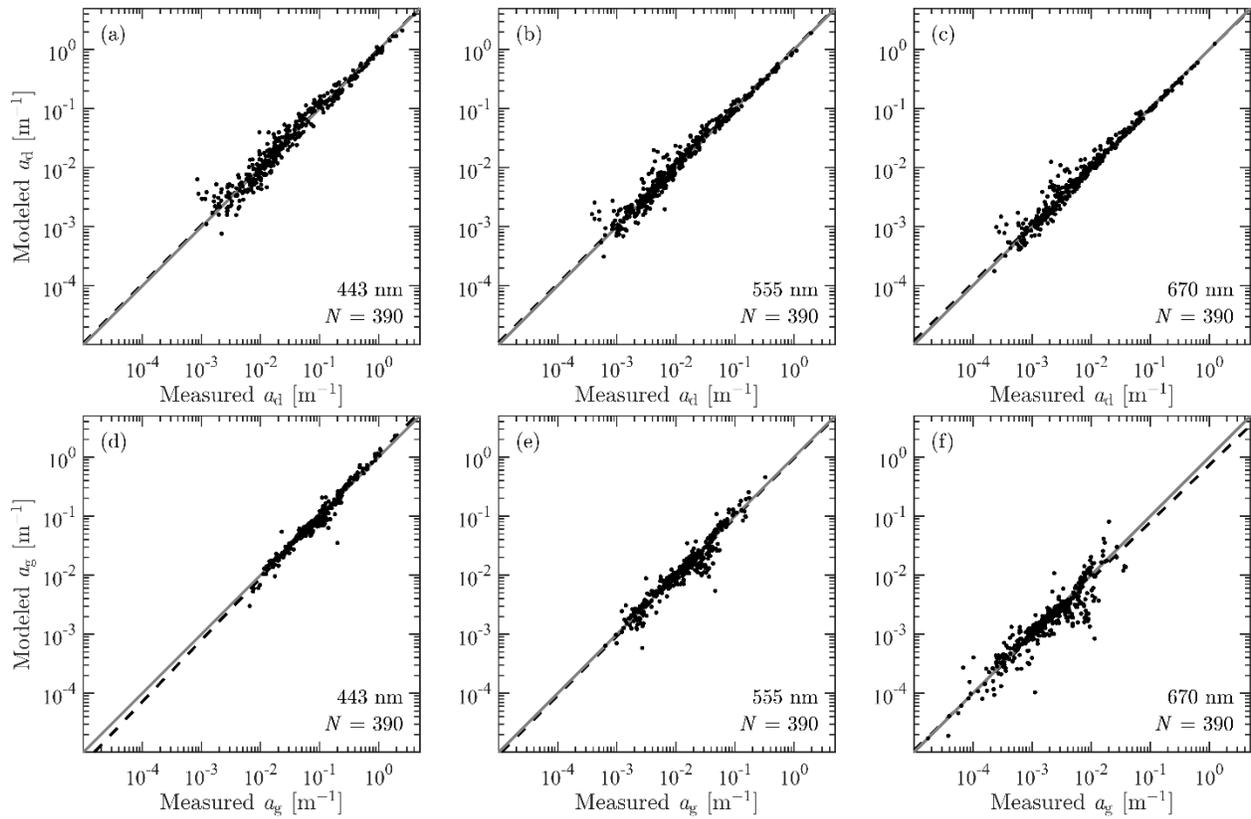
**Figure 2.4.** Joint probability distributions of the spectral steepness parameters,  $S_d$  and  $S_g$ , calculated respectively from (a) the spectral shape functions of non-algal particulate absorption coefficient,  $\hat{a}_d(\lambda)$ , and (b) the spectral shape functions of CDOM absorption coefficient,  $\hat{a}_g(\lambda)$ . In each panel, the steepness parameters are shown for two spectral regions representing the blue (from  $\lambda_1 = 410$  nm to  $\lambda_2 = 460$  nm) and blue-green (from  $\lambda_1 = 480$  nm to  $\lambda_2 = 530$  nm) portions of the spectrum. The black dashed lines indicate the median of the distribution for each axis. The number of spectra within each distribution,  $N$ , and the number of bins containing at least one spectrum,  $N_{\text{bin}}$ , are denoted in each panel.



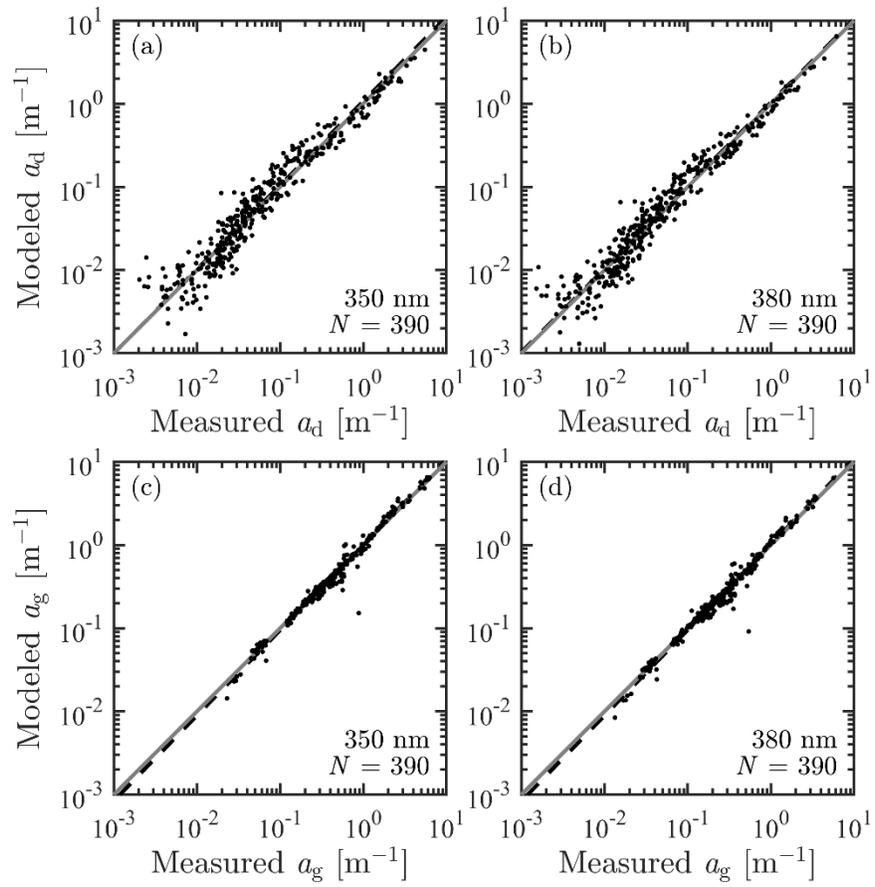
**Figure 2.5.** Characteristic spectral shape functions  $\hat{a}_d(\lambda)$  and  $\hat{a}_g(\lambda)$  comprising the final libraries used in the ADG partitioning model. (a-b) The UV–VIS library for the spectral range 350–700 nm. (c-d) The VIS library for the spectral range 400–700 nm. The number of shape functions in each library,  $N$ , is denoted in each panel. Note that the grayscale lines in all panels are for the illustrative purpose to delineate spectral shape functions in each library.



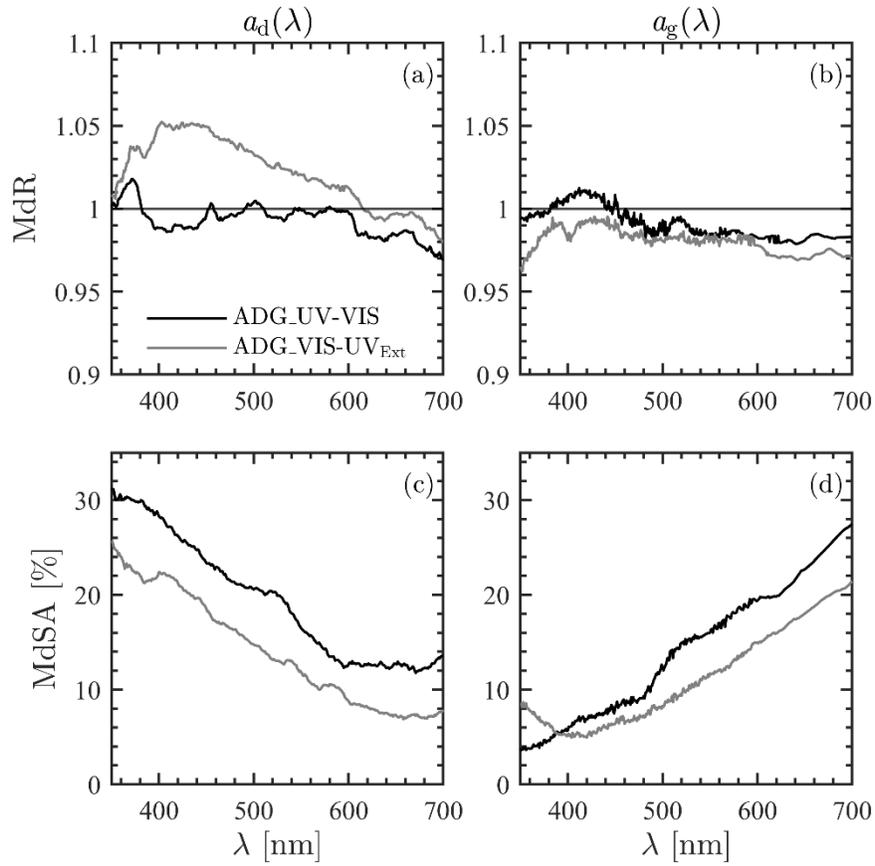
**Figure 2.6.** (a–c) The median ratio,  $MdR$ , and (d–f) the median symmetric accuracy,  $MdSA$ , calculated from comparison of model-derived and measured values of  $a_d(\lambda)$ ,  $a_g(\lambda)$ , and  $a_{dg}(\lambda)$  using the set of concurrent measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$  in the model development dataset ( $N = 390$ ). All values are shown as functions of different percentiles chosen from the pool of feasible solutions obtained from the ADG\_UV-VIS (solid lines) and ADG\_VIS-UV<sub>Ext</sub> (dashed lines) models. Results are depicted for light wavelengths of 350 nm (purple), 440 nm (blue), 550 nm (green), and 660 nm (red). The vertical dashed black line denotes the 50<sup>th</sup> percentile of feasible solutions for  $a_d(\lambda)$  and  $a_g(\lambda)$ . The horizontal dashed black line in plots (a–c) indicates a  $MdR$  value equal to 1.



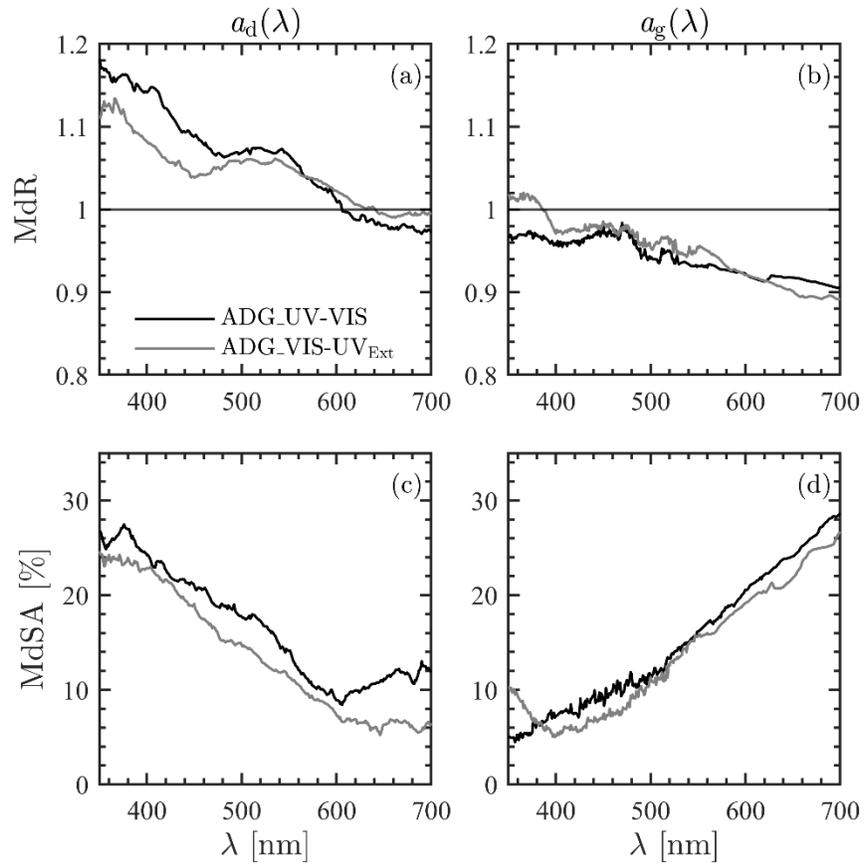
**Figure 2.7.** Assessment of the ADG\_UV-VIS model applied to the development dataset as scatter plots of model-derived versus measured absorption coefficients  $a_d(\lambda)$  and  $a_g(\lambda)$  for light wavelengths of (a,d) 443 nm, (b,e) 555 nm, and (c,f) 670 nm. The 1:1 line and best-fit line derived from the Model II linear regression to  $\log_{10}$ -transformed data are represented by the solid gray and dashed black lines, respectively. Additional statistical metrics are provided in Table 2.3.



**Figure 2.8.** Similar to Fig. 2.7 but for wavelengths selected from the near-UV range: (a, c) 350 nm and (b, d) 380 nm.



**Figure 2.9.** Spectral values of the (a–b) median ratio,  $MdR$ , and (c–d) median symmetric accuracy,  $MdSA$ , calculated from comparison of model-derived and measured values of  $a_d(\lambda)$  and  $a_g(\lambda)$  obtained from the ADG partitioning model applied to the development dataset ( $N = 390$ ). Results are depicted for the two model variants ADG\_UV-VIS (black lines) and ADG\_VIS-UV<sub>Ext</sub> (gray lines). The horizontal solid black line in panels (a, b) indicates a  $MdR$  value equal to 1.



**Figure 2.10.** Similar to Fig. 2.9 but for the independent validation dataset ( $N = 149$ ).

## 2.9 Tables

**Table 2.1.** Model Development and Validation Datasets. The number,  $N$ , of individual  $a_d(\lambda)$  and  $a_g(\lambda)$  and concurrent  $a_d(\lambda)$  and  $a_g(\lambda)$  measurements included in model development and validation datasets are listed for each experiment.

Ocean basin	Experiment (Cruises)	Development			Validation
		$N [a_d(\lambda)]$	$N [a_g(\lambda)]$	$N [a_d(\lambda) \& a_g(\lambda)]$	$N [a_d(\lambda) \& a_g(\lambda)]$
Pacific	BIO SOPE (BIO SOPE)	0	31	0	0
	TAO (gp1-06-ka, gp5-05-ka, gp5-06-ka)	0	292	0	0
	BEST (hly0803)	0	18	0	0
	CLIVAR (P16S)	31	32	19	12
	KORUS (KR_2016)	52	61	38	5
	EXPORTS (Process, Survey)	0	63	0	0
	Sea2Space (FK170124)	10	0	0	0
Atlantic	CLIVAR (A16S)	0	29	0	0
	AMMA (AMMA-RB-06)	0	12	0	0
	CLiVEC (CV1, CV2, CV4, CV5, CV7)	106	357	84	10
	ECOMON (PC1301)	35	25	24	18
	GEO-CAPE (GOMEX_2013)	100	87	74	20
	NAAMES (NA1, NA2, NA3, NA4)	0	39	0	0
	Cyanate (Cyanate 2016)	18	7	3	4
Indian	CLIVAR (I8SI9N)	0	57	0	0
Arctic	BEST (hly0803)	0	8	0	0
	MALINA (MALINA)	65	60	53	24
	ICESCAPE (HLY1001, HLY1101)	75	76	71	43
Southern	ArCS (MR17-05C)	18	22	13	12
	CLIVAR (I8SI9N)	0	5	0	0
	CLIVAR (P16S)	11	12	11	1
Total		521	1293	390	149

**Table 2.2.** Statistical Metrics to Assess Model Performance.

Symbol	Description
$N$	Number of samples
$O_i$ or $P_i$	Observed or model-predicted value for sample $i$ of $N$
$MdR$	Median Ratio; $MdR = \text{Median}\left(\frac{P_i}{O_i}\right)$
$MdB$	Median Bias; $MdB = \text{Median}(P_i - O_i)$
$MdAPD$	Median Absolute Percent Difference; $MdAPD = \text{Median}\left(\left \frac{P_i - O_i}{O_i}\right \right) \times 100\%$
$RMSD$	Root-mean-square deviation; $RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2}$
$MdSA$	Median Symmetric Accuracy; $MdSA = (10^{Md \log_{10}(\frac{P_i}{O_i}) } - 1) \times 100\%$
$r$	Pearson correlation coefficient of log-transformed $P_i$ vs. $O_i$
$A$	Slope of model II linear regression of log-transformed $P_i$ vs. $O_i$
$B$	Ten to the power of the y-intercept of Model II linear regression of log-transformed $P_i$ vs. $O_i$

**Table 2.3.** Performance of ADG Partitioning Model with the Development Dataset. The aggregate error statistics of ADG\_UV-VIS and ADG\_VIS-UV<sub>Ext</sub> applied to the development dataset ( $N = 390$ ) at the light wavelengths of 350, 380, 443, 555, and 670 nm (see Section 3 and Table 2.2 for more details on statistics).

Model Variant	Variable	$r$	$A$	$B$	$MdB$ [ $m^{-1}$ ]	$MdR$	$MdAPD$ [%]	$MdSA$ [%]	$RMSD$ [ $m^{-1}$ ]
ADG_UV-VIS	$a_d(350)$	0.9688	1.010	1.073	0.00001	1.001	25.95	31.26	0.1393
	$a_d(380)$	0.9723	1.000	1.036	0.00025	1.005	24.81	29.78	0.0958
	$a_d(443)$	0.9770	0.996	1.023	-0.00016	0.991	21.15	24.39	0.0410
	$a_d(555)$	0.9815	0.996	1.045	-0.00001	0.997	14.61	16.05	0.0136
	$a_d(670)$	0.9846	0.987	0.991	-0.00010	0.982	11.45	11.95	0.0065
	$a_g(350)$	0.9903	1.026	1.003	-0.00191	0.994	3.77	3.79	0.1494
	$a_g(380)$	0.9887	1.029	1.018	-0.00021	0.999	4.62	4.65	0.0898
	$a_g(443)$	0.9817	1.047	1.103	0.00025	1.005	7.76	7.99	0.0449
	$a_g(555)$	0.9550	1.004	0.944	-0.00008	0.982	15.77	16.77	0.0132
	$a_g(670)$	0.9199	0.967	0.729	-0.00001	0.983	21.90	24.60	0.0046
ADG_VIS-UV <sub>Ext</sub>	$a_d(350)$	0.9686	1.060	1.218	0.00051	1.008	22.07	25.70	0.1360
	$a_d(380)$	0.9752	1.017	1.105	0.00084	1.036	19.65	22.00	0.0711
	$a_d(443)$	0.9802	1.007	1.079	0.00059	1.050	17.27	19.62	0.0322
	$a_d(555)$	0.9852	1.003	1.078	0.00013	1.018	10.64	11.48	0.0114
	$a_d(670)$	0.9895	0.993	1.018	-0.00002	0.996	6.91	7.38	0.0043
	$a_g(350)$	0.9904	0.998	0.954	-0.01140	0.965	8.07	8.56	0.1632
	$a_g(380)$	0.9902	1.015	0.984	-0.00182	0.986	6.10	6.43	0.0835
	$a_g(443)$	0.9866	1.030	1.041	-0.00030	0.994	6.23	6.37	0.0307
	$a_g(555)$	0.9639	0.994	0.894	-0.00016	0.985	11.13	11.90	0.0111
	$a_g(670)$	0.9338	0.956	0.677	-0.00002	0.974	17.84	19.57	0.0044

**Table 2.4.** Performance of ADG Partitioning Model with the Independent Validation Dataset. Similar to Table 2.3 but for the independent validation dataset ( $N = 149$ ).

Model Variant	Variable	$r$	$A$	$B$	$MdB$ [ $m^{-1}$ ]	$MdR$	$MdAPD$ [%]	$MdSA$ [%]	$RMSD$ [ $m^{-1}$ ]	
ADG_UV-VIS	$a_d(350)$	0.9659	1.023	1.257	0.00843	1.182	25.32	26.58	0.0394	
	$a_d(380)$	0.9682	1.007	1.175	0.00456	1.161	25.64	26.99	0.0281	
	$a_d(443)$	0.9720	0.991	1.056	0.00128	1.095	20.23	21.29	0.0129	
	$a_d(555)$	0.9807	0.966	0.930	0.00035	1.058	12.48	12.93	0.0034	
	$a_d(670)$	0.9838	0.930	0.742	-0.00013	0.982	11.41	11.84	0.0023	
	$a_g(350)$	0.9663	1.012	0.949	-0.00687	0.968	4.86	5.09	0.0386	
	$a_g(380)$	0.9547	1.002	0.926	-0.00476	0.970	6.14	6.44	0.0278	
	$a_g(443)$	0.9293	1.015	0.973	-0.00097	0.972	8.76	8.94	0.0121	
	$a_g(555)$	0.8890	0.968	0.785	-0.00036	0.932	14.21	16.57	0.0033	
	$a_g(670)$	0.8459	0.913	0.507	-0.00007	0.913	21.96	26.04	0.0008	
	ADG_VIS-UV <sub>Ext</sub>	$a_d(350)$	0.9647	1.063	1.309	0.00266	1.112	22.73	24.63	0.0426
		$a_d(380)$	0.9670	1.020	1.140	0.00210	1.104	22.46	23.93	0.0211
		$a_d(443)$	0.9713	1.004	1.039	0.00078	1.047	18.34	19.20	0.0090
		$a_d(555)$	0.9855	1.005	1.060	0.00035	1.048	10.24	11.03	0.0027
$a_d(670)$		0.9925	0.974	0.897	-0.00005	0.993	6.22	6.58	0.0015	
$a_g(350)$		0.9495	1.078	1.121	0.00391	1.016	9.56	10.33	0.0540	
$a_g(380)$		0.9622	1.060	1.121	0.00187	1.011	6.28	6.53	0.0262	
$a_g(443)$		0.9585	1.025	1.049	-0.00053	0.982	6.81	6.87	0.0089	
$a_g(555)$		0.9213	0.955	0.747	-0.00030	0.953	13.64	15.75	0.0027	
$a_g(670)$		0.8764	0.903	0.471	-0.00009	0.895	21.18	24.71	0.0007	

## 2.10 References

1. J. T. O. Kirk, *Light and Photosynthesis in Aquatic Ecosystems*, 2nd ed. (Cambridge University Press, 1994).
2. N. G. Jerlov, *Marine Optics* (Elsevier, 1976).
3. C. D. Mobley, *The Ocean Optics Book*. (IOCCG, 2022).
4. R. M. Pope and E. S. Fry, “Absorption spectrum (380–700 nm) of pure water II Integrating cavity measurements,” *Appl. Opt.* **36**(33), 8710–8723 (1997).
5. M. Twardowski, R. Röttgers, and D. Stramski, “Chapter 1: The Absorption Coefficient, An Overview,” In: *IOCCG Protocol Series Inherent Optical Property Measurements and Protocols: Absorption Coefficient*, A. R., Neeley and A. Mannino eds., IOCCG Ocean Optics and Biogeochemistry Protocols for Satellite Ocean Colour Sensor Validation, Volume 1.0, IOCCG, Dartmouth, NS, Canada (2018).
6. A. Bricaud, A. Morel, and L. Prieur, “Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains,” *Limnol. Oceanogr.* **26**(1), 43–53 (1981).
7. A. Vodacek, N. V. Blough, M. D. DeGrandpre, *et al.*, “Seasonal variation of CDOM and DOC in the Middle Atlantic Bight: Terrestrial inputs and photooxidation,” *Limnol. Oceanogr.* **42**(4), 674–686 (1997).
8. D. Stramski, R. A. Reynolds, M. Kahru, *et al.*, “Estimation of Particulate Organic Carbon in the Ocean from Satellite Remote Sensing,” *Science* **285**(5425), 239–242 (1999).
9. M. Stramska and D. Stramski, “Variability of particulate organic carbon concentration in the north polar Atlantic based on ocean color observations with Sea-viewing Wide Field-of-view Sensor (SeaWiFS),” *J. Geophys. Res.* **110**(C10), C10018 (2005).
10. A. Mannino, M. E. Russ, and S. B. Hooker, “Algorithm development and validation for satellite-derived distributions of DOC and CDOM in the U.S. Middle Atlantic Bight,” *J. Geophys. Res.* **113**(C7), C07051 (2008).
11. A. Matsuoka, A. Bricaud, R. Benner, *et al.*, “Tracing the transport of colored dissolved organic matter in water masses of the Southern Beaufort Sea: relationship with hydrographic characteristics,” *Biogeosciences* **9**(3), 925–940 (2012).
12. I. D. Joshi, E. J. D’Sa, C. L. Osburn, *et al.*, “Assessing chromophoric dissolved organic matter (CDOM) distribution, stocks, and fluxes in Apalachicola Bay using combined field, VIIRS ocean color, and model observations,” *Remote Sens. Environ.* **191**, 359–372 (2017).
13. D. Stramski, I. Joshi, and R. A. Reynolds, “Ocean color algorithms to estimate the concentration of particulate organic carbon in surface waters of the global ocean in support

- of a long-term data record from multiple satellite missions,” *Remote Sens. Environ.* **269**, 112776 (2022).
14. R. J. W. Brewin, S. Sathyendranath, G. Kulk, *et al.*, “Ocean carbon from space: Current status and priorities for the next decade,” *Earth-Sci. Rev.* **240**, 104386 (2023).
  15. A. Morel, “Available, usable, and stored radiant energy in relation to marine photosynthesis,” *Deep-Sea Res.* **25**(8), 673–688 (1978).
  16. S. Sathyendranath and T. Platt, “Computation of aquatic primary production: Extended formalism to include effect of angular and spectral distribution of light,” *Limnol. Oceanogr.* **34**(1), 188–198 (1989).
  17. D. Antoine, J. André, and A. Morel, “Oceanic primary production: 2. Estimation at global scale from satellite (Coastal Zone Color Scanner) chlorophyll,” *Global Biogeochem. Cy.* **10**(1), 57–69 (1996).
  18. J. Marra, C. C. Trees, and J. E. O’Reilly, “Phytoplankton pigment absorption: A strong predictor of primary productivity in the surface ocean,” *Deep-Sea Res. I: Oceanogr. Res. Pap.* **54**(2), 155–163 (2007).
  19. J. Uitz, H. Claustre, B. Gentili, *et al.*, “Phytoplankton class-specific primary production in the world’s oceans: Seasonal and interannual variability from satellite observations,” *Global Biogeochem. Cy.* **24**(3), GB3016 (2010).
  20. T. K. Westberry, G. M. Silsbe, and M. J. Behrenfeld, “Gross and net primary production in the global ocean: An ocean color remote sensing perspective,” *Earth Sci. Rev.* **237**, 104322 (2023).
  21. 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,” *Limnol. Oceanogr.* **47**(2), 404–417 (2002).
  22. S. Sathyendranath, V. Stuart, T. Platt, *et al.*, “Remote sensing of ocean colour: Towards algorithms for retrieval of pigment composition,” *Indian J. Mar. Sci.* **34**(4), 333–340 (2005).
  23. A. Bricaud, C. Mejia, D. Blondeau-Patissier, *et al.*, “Retrieval of pigment concentrations and size structure of algal populations from their absorption spectra using multilayered perceptrons,” *Appl. Opt.* **46**(8), 1251–1260 (2007).
  24. T. Hirata, J. Aiken, N. Hardman-Mountford, *et al.*, “An absorption model to determine phytoplankton size classes from satellite ocean colour,” *Remote Sens. Environ.* **112**(6), 3153–3159 (2008).

25. R. J. W. Brewin, E. Devred, S. Sathyendranath, *et al.*, “Model of phytoplankton absorption based on three size classes,” *Appl. Opt.* **50**(22), 4535–4549 (2011).
26. E. Devred, S. Sathyendranath, V. Stuart, *et al.*, “A three component classification of phytoplankton absorption spectra: Application to ocean-color data,” *Remote Sens. Environ.* **115**(9), 2255–2266 (2011).
27. E. Organelli, A. Bricaud, D. Antoine, *et al.*, “Multivariate approach for the retrieval of phytoplankton size structure from measured light absorption spectra in the Mediterranean Sea (BOUSSOLE site),” *Appl. Opt.* **52**(11), 2257–2273 (2013).
28. J. Uitz, D. Stramski, R. A. Reynolds, *et al.*, “Assessing phytoplankton community composition from hyperspectral measurements of phytoplankton absorption coefficient and remote-sensing reflectance in open-ocean environments,” *Remote Sens. Environ.* **171**, 58–74 (2015).
29. R. A. Reynolds and D. Stramski, “Optical characterization of marine phytoplankton assemblages within surface waters of the western Arctic Ocean,” *Limnol. Oceanogr.* **64**(6), 2478–2496 (2019).
30. S. B. Woźniak, D. Stramski, M. Stramska, *et al.*, “Optical variability of seawater in relation to particle concentration, composition, and size distribution in the nearshore marine environment at Imperial Beach, California,” *J. Geophys. Res.* **115**(C8), (2010).
31. D. Stramski, S. Constantin, and R. A. Reynolds, “Adaptive optical algorithms with differentiation of water bodies based on varying composition of suspended particulate matter: A case study for estimating the particulate organic carbon concentration in the western Arctic seas,” *Remote Sens. Environ.* **286**, 113360 (2023).
32. C. Hu, Z. Chen, T. D. Clayton, *et al.*, “Assessment of estuarine water-quality indicators using MODIS medium-resolution bands: Initial results from Tampa Bay, FL,” *Remote Sens. Environ.* **93**(3), 423–441 (2004).
33. B. A. Schaeffer, J. D. Hagy, R. N. Conmy, *et al.*, “An Approach to Developing Numeric Water Quality Criteria for Coastal Waters Using the SeaWiFS Satellite Data Record,” *Environ. Sci. Technol.* **46**(2), 916–922 (2012).
34. G. Zheng and P. M. DiGiacomo, “Uncertainties and applications of satellite-derived coastal water quality products,” *Prog. Oceanogr.* **159**, 45–72 (2017).
35. J. S. Turner, K. A. Fall, and C. T. Friedrichs, “Clarifying water clarity: A call to use metrics best suited to corresponding research and management goals in aquatic ecosystems,” *Limnol. Oceanogr. Lett.* **8**(3), 388–397 (2022).
36. H. R. Gordon and D. K. Clark, “Remote sensing optical properties of a stratified ocean: an improved interpretation,” *Appl. Opt.* **19**(20), 3428–3430 (1980).

37. H. R. Gordon and M. Wang, "Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm," *Appl. Opt.* **33**(3), 443–452 (1994).
38. P. J. Werdell, L. I. W. McKinna, E. Boss, *et al.*, "An overview of approaches and challenges for retrieving marine inherent optical properties from ocean color remote sensing," *Prog. Oceanogr.* **160**, 186–212 (2018).
39. H. Loisel and D. Stramski, "Estimation of the inherent optical properties of natural waters from the irradiance attenuation coefficient and reflectance in the presence of Raman scattering," *Appl. Opt.* **39**(18), 3001–3011 (2000).
40. 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," *Appl. Opt.* **41**(27), 5755–5772 (2002).
41. M. H. Pinkerton, G. F. Moore, S. J. Lavender, *et al.*, "A method for estimating inherent optical properties of New Zealand continental shelf waters from satellite ocean colour measurements," *New Zeal. J. Mar. Fresh.* **40**(2), 227–247 (2006).
42. T. J. Smyth, G. F. Moore, T. Hirata, *et al.*, "Semianalytical model for the derivation of ocean color inherent optical properties: description, implementation, and performance assessment," *Appl. Opt.* **45**(31), 8116–8131 (2006).
43. H. Loisel, D. Stramski, D. Dessailly, *et al.*, "An Inverse Model for Estimating the Optical Absorption and Backscattering Coefficients of Seawater From Remote-Sensing Reflectance Over a Broad Range of Oceanic and Coastal Marine Environments," *J. Geophys. Res-Oceans* **123**(3), 2141–2171 (2018).
44. C. S. Roesler and M. J. Perry, "In situ phytoplankton absorption, fluorescence emission, and particulate backscattering spectra determined from reflectance," *J. Geophys. Res.* **100**(C7), 13279–13294 (1995).
45. F. E. Hoge and P. E. Lyon, "Satellite retrieval of inherent optical properties by linear matrix inversion of oceanic radiance models: An analysis of model and radiance measurement errors," *J. Geophys. Res.* **101**(C7), 16631–16648 (1996).
46. S. A. Garver and D. A. Siegel, "Inherent optical property inversion of ocean color spectra and its biogeochemical interpretation: 1. Time series from the Sargasso Sea," *J. Geophys. Res.* **102**(C8), 18607–18625 (1997).
47. S. Maritorena, D. A. Siegel, and A. R. Peterson, "Optimization of a semianalytical ocean color model for global-scale applications," *Appl. Opt.* **41**(15), 2705–2714 (2002).

48. P. Wang, E. S. Boss, and C. Roesler, “Uncertainties of inherent optical properties obtained from semianalytical inversions of ocean color,” *Appl. Opt.* **44**(19), 4074–4085 (2005).
49. E. Devred, S. Sathyendranath, and T. Platt, “Inversion based on semi-analytical reflectance model,” in *Remote Sensing of Inherent Optical Properties: Fundamentals, Tests of Algorithms and Applications* (IOCCG, 2006), pp.87–94.
50. V. E. Brando, A. G. Dekker, Y. J. Park, *et al.*, “Adaptive semianalytical inversion of ocean color radiometry in optically complex waters,” *Appl. Opt.* **51**(15), 2808–2833 (2012).
51. H. Loisel, D. Stramski, B. G. Mitchell, *et al.*, “Comparison of the ocean inherent optical properties obtained from measurements and inverse modeling,” *Appl. Opt.* **40**(15), 2384–2397 (2001).
52. Z. Lee, R. Arnone, C. Hu, *et al.*, “Uncertainties of optical parameters and their propagations in an analytical ocean color inversion algorithm,” *Appl. Opt.* **49**(3), 369–381 (2010).
53. G. Zheng, D. Stramski, and R. A. Reynolds, “Evaluation of the Quasi-Analytical Algorithm for estimating the inherent optical properties of seawater from ocean color: Comparison of Arctic and lower-latitude waters,” *Remote Sens. Environ.* **155**, 194–209 (2014).
54. C. S. Roesler, M. J. Perry, and K. L. Carder, “Modeling in situ phytoplankton absorption from total absorption spectra in productive inland marine waters,” *Limnol. Oceanogr.* **34**(8), 1510–1523 (1989).
55. A. M. Ciotti and A. Bricaud, “Retrievals of a size parameter for phytoplankton and spectral light absorption by colored detrital matter from water-leaving radiances at SeaWiFS channels in a continental shelf region off Brazil,” *Limnol. Oceanogr-Meth.* **4**(7), 237–253 (2006).
56. G. Zheng and D. Stramski, “A model based on stacked-constraints approach for partitioning the light absorption coefficient of seawater into phytoplankton and non-phytoplankton components,” *J. Geophys. Res-Oceans* **118**(4), 2155–2174 (2013).
57. X. Zhang, Y. Huot, A. Bricaud, *et al.*, “Inversion of spectral absorption coefficients to infer phytoplankton size classes, chlorophyll concentration, and detrital matter,” *Appl. Opt.* **54**(18), 5805–5816 (2015).
58. G. C. Chang and T. D. Dickey, “Partitioning in situ total spectral absorption by use of moored spectral absorption–attenuation meters,” *Appl. Opt.* **38**(18), 3876–3887 (1999).
59. H. Claustre, F. Fell, K. Oubelkheir, *et al.*, “Continuous monitoring of surface optical properties across a geostrophic front: Biogeochemical inferences,” *Limnol. Oceanogr.* **45**(2), 309–321 (2000).

60. C. L. Gallegos and P. J. Neale, "Partitioning spectral absorption in case 2 waters: discrimination of dissolved and particulate components," *Appl. Opt.* **41**(21), 4220–4233 (2002).
61. O. Schofield, T. Bergmann, M. J. Oliver, *et al.*, "Inversion of spectral absorption in the optically complex coastal waters of the Mid-Atlantic Bight," *J. Geophys. Res.* **109**(C12), C12S04 (2004).
62. Q. Dong, S. Shang, and Z. Lee, "An algorithm to retrieve absorption coefficient of chromophoric dissolved organic matter from ocean color," *Remote Sens. Environ.* **128**, 259–267 (2013).
63. J. Lin, W. Cao, G. Wang, *et al.*, "Approach for determining the contributions of phytoplankton, colored organic material, and nonalgal particles to the total spectral absorption in marine waters," *Appl. Opt.* **52**(18), 4249–4257 (2013).
64. G. Zheng, D. Stramski, and P. M. DiGiacomo, "A model for partitioning the light absorption coefficient of natural waters into phytoplankton, nonalgal particulate, and colored dissolved organic components: A case study for the Chesapeake Bay," *J. Geophys. Res-Oceans* **120**(4), 2601–2621 (2015).
65. R. Iturriaga and D. A. Siegel, "Microphotometric characterization of phytoplankton and detrital absorption properties in the Sargasso Sea," *Limnol. Oceanogr.* **34**(8), 1706–1726 (1989).
66. N. V. Blough and R. Del Vecchio, *Biogeochemistry of Marine Dissolved Organic Matter: Chromophoric DOM in the Coastal Environment* (Elsevier, 2002), pp. 509–546.
67. M. Babin, D. Stramski, G. M. Ferrari, *et al.*, "Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe," *J. Geophys. Res.* **108**(C7), 3211 (2003).
68. M. Babin and D. Stramski, "Variations in the mass-specific absorption coefficient of mineral particles suspended in water," *Limnol. Oceanogr.* **49**(3), 756–767 (2004).
69. M. S. Twardowski, E. Boss, J. M. Sullivan, *et al.*, "Modeling the spectral shape of absorption by chromophoric dissolved organic matter," *Mar. Chem.* **89**(1–4), 69–88 (2004).
70. D. G. Bowers and C. E. Binding, "The optical properties of mineral suspended particles: A review and synthesis," *Estuar. Coast. Shelf Sci.* **67**(1–2), 219–230 (2006).
71. D. Stramski, M. Babin, and S. B. Woźniak, "Variations in the optical properties of terrigenous mineral-rich particulate matter suspended in seawater," *Limnol. Oceanogr.* **52**(6), 2418–2433 (2007).

72. S. A. Loisel, L. Bracchini, A. M. Dattilo, *et al.*, “The optical characterization of chromophoric dissolved organic matter using wavelength distribution of absorption spectral slopes,” *Limnol. Oceanogr.* **54**(2), 590–597 (2009).
73. D. Stramski, L. Li, and R. A. Reynolds, “Model for separating the contributions of non-algal particles and colored dissolved organic matter to light absorption by seawater,” *Appl. Opt.* **58**(14), 3790–3806 (2019).
74. D. Stramski, R. A. Reynolds, S. Kaczmarek, *et al.*, “Correction of pathlength amplification in the filter-pad technique for measurements of particulate absorption coefficient in the visible spectral region,” *Appl. Opt.* **54**(22), 6763–6782 (2015).
75. C. Roesler, D. Stramski, E. J. D’Sa, *et al.*, “Chapter 5: Spectrophotometric Measurements of Particulate Absorption Using Filter Pads,” In: IOCCG Protocol Series Inherent Optical Property Measurements and Protocols: Absorption Coefficient, A. R., Neeley and A. Mannino, eds., IOCCG Ocean Optics and Biogeochemistry Protocols for Satellite Ocean Colour Sensor Validation, Volume 1.0, IOCCG, Dartmouth, NS, Canada (2018).
76. S. Tassan and G. M. Ferrari, “Variability of light absorption by aquatic particles in the near-infrared spectral region,” *Appl. Opt.* **42**(24), 4802–4810 (2003).
77. D. Stramski, S. B. Woźniak, and P. J. Flatau, “Optical properties of Asian mineral dust suspended in seawater,” *Limnol. Oceanogr.* **49**(3), 749–755 (2004).
78. R. Röttgers, C. Dupouy, B. B. Taylor, *et al.*, “Mass-specific light absorption coefficients of natural aquatic particles in the near-infrared spectral region,” *Limnol. Oceanogr.* **59**(5), 1449–1460 (2014).
79. M. D. Kehrl, D. Stramski, R. A. Reynolds, *et al.*, “Estimation of chromophoric dissolved organic matter and non-algal particulate absorption coefficients of seawater in the ultraviolet by extrapolation from the visible spectral region,” *Opt. Express* **31**(11), 17450–17479 (2023).
80. F. A. J. Armstrong, P. M. Williams, and J. D. H. Strickland, “Photo-oxidation of Organic Matter in Sea Water by Ultra-violet Radiation, Analytical and Other Applications,” *Nature* **211**(5048), 481–483 (1966).
81. A. Bricaud and D. Stramski, “Spectral absorption coefficients of living phytoplankton and nonalgal biogenous matter: A comparison between the Peru upwelling area and the Sargasso Sea,” *Limnol. Oceanogr.* **35**(3), 562–582 (1990).
82. K. Mopper and D. J. Kieber, *Biogeochemistry of Marine Dissolved Organic Matter: Photochemistry and the cycling of carbon, sulfur, nitrogen and phosphorus* (Elsevier, 2002), pp. 455–507.

83. J. R. Helms, A. Stubbins, J. D. Ritchie, *et al.*, “Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter,” *Limnol. Oceanogr.* **53**(3), 955–969 (2008).
84. P. J. Werdell, M. J. Behrenfeld, P. S. Bontempi, *et al.*, “The Plankton, Aerosol, Cloud, Ocean Ecosystem Mission: Status, Science, Advances,” *Bull. Am. Meteorol. Soc.* **100**(9), 1775–1794 (2019).
85. B. G. Mitchell, A. Bricaud, K. Carder, *et al.*, “Determination of spectral absorption coefficients of particles, dissolved material, and phytoplankton for discrete water samples,” In: *Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 2*, NASA TM–2000–209966, G. S. Fargion and J. L. Mueller, eds. (NASA, 2000), pp. 125–153.
86. A. Mannino, M. G. Novak, N. B. Nelson, *et al.*, “Measurement protocol of absorption by chromophoric dissolved organic matter (CDOM) and other dissolved materials (DRAFT),” In: *IOCCG Protocol Series Inherent Optical Property Measurements and Protocols: Absorption Coefficient*, A. Mannino and M.G. Novak eds., IOCCG Ocean Optics and Biogeochemistry Protocols for Satellite Ocean Colour Sensor Validation, Volume 5.0, IOCCG, Dartmouth, NS, Canada (2019).
87. M. Kishino, M. Takahashi, N. Okami, and S. Ichimura, “Estimation of the spectral absorption coefficients of phytoplankton in the sea,” *Bull. Mater. Sci.* **37**(2), 634–642 (1985).
88. R. Röttgers and S. Gehnke, “Measurement of light absorption by aquatic particles: improvement of the quantitative filter technique by use of an integrating sphere approach,” *Appl. Opt.* **51**(9), 1336–1351 (2012).
89. M. Babin and D. Stramski, “Light absorption by aquatic particles in the near-infrared spectral region,” *Limnol. Oceanogr.* **47**(3), 911–915 (2002).
90. S. K. Morley, T. V. Brito, and D. T. Welling, “Measures of model performance based on the log accuracy ratio,” *Space Weather* **16**, 69–88 (2018).
91. D. Koestner, D. Stramski, and R. A. Reynolds, “Characterization of suspended particulate matter in contrasting coastal marine environments with angle-resolved polarized light scattering measurements,” *Appl. Opt.* **60**(36), 11161–11179 (2021).
92. W. E. Ricker, “Linear Regressions in Fishery Research,” *J. Fish. Res. Bd. Can.* **30**(3), 409–434 (1973).
93. D. Freedman and P. Diaconis, “On the histogram as a density estimator:  $L_2$  theory,” *Z. Wahrscheinlichkeit.* **57**(4), 453–476 (1981).

## **Chapter 3**

### **Performance and uncertainty assessment of a novel multi-step semi-analytical algorithm for estimating seawater optical properties from ocean reflectance**

### 3.0 Abstract

Inversion methods applied to satellite-derived ocean color data provide a valuable tool for investigating the optical and biogeochemical properties of the ocean over broad spatiotemporal scales. A current challenge of inversion models designed to retrieve the inherent optical properties (IOPs) of seawater from radiometric measurements is that they retrieve a limited subset of IOPs or impose restrictive spectral assumptions on output variables. To address these challenges, a 4-step Semi-Analytical Algorithm (4SAA), which implements four independent component models, has been proposed to retrieve the hyperspectral attenuation coefficient of downward irradiance,  $\langle K_d(\lambda) \rangle$ , and a suite of total and constituent hyperspectral IOPs from input remote-sensing reflectance,  $R_{rs}(\lambda)$ . We develop a systematic approach to assess the performance and uncertainty of the 4SAA model across the 350–700 nm spectral range using a recently published synthetic optical database covering a wide range of oceanic environments. The performance assessment first quantifies the error introduced by each component model of 4SAA and then evaluates the propagation of error through the multi-step inversion sequence. The uncertainty assessment utilizes a Monte Carlo method to perturb input parameters of each component model when operated in a multi-step sequence. Results indicate that retrievals of  $\langle K_d(\lambda) \rangle$ , the total absorption coefficient  $a(\lambda)$ , and the total backscattering coefficient  $b_b(\lambda)$  exhibit a strong performance with errors below 15% across the entire spectral range. Constituent IOPs are retrieved well in specific spectral regions, particularly in the blue and blue-green portions of the spectrum, but often exhibit substantial errors in the near-UV and red which is typically associated with a reduced contribution of non-water constituents compared to the pure seawater contribution to the total IOPs. This study provides insight into the quantification and propagation of error and uncertainty in the multi-step model and highlights potential limitations of model evaluation with datasets that assume fixed

spectral shapes for constituent IOPs. These findings also offer guidance for refining component models within 4SAA to improve retrievals for future ocean color applications.

### **3.1 Introduction**

In optical oceanography, inversion algorithms aim to estimate the optical properties of the ocean that govern light propagation through the seawater medium. These models derive optical properties by inverting the problem of radiative transfer theory, which defines the effects of seawater inherent optical properties (IOPs) on radiometric quantities and apparent optical properties (AOPs) of a given water body. The conceptual development of inversion modeling techniques to estimate IOPs from water-leaving upwelling light field measurements began over half a century ago (Gordon & Brown, 1973; Gordon 1973), following the discovery of the relationship between measurements of chlorophyll-a concentrations and ocean color (Clarke et al., 1970). Over the next several decades, inversion algorithms continued to develop (e.g., Morel & Prieur, 1977; Kirk 1984; Aas 1987; Roesler & Perry, 1995; Loisel & Stramski, 2000; Lee et al., 2002; Maritorena et al., 2002; Pinkerton et al., 2006; Smyth et al., 2006; Werdell et al., 2013; Loisel et al., 2018), driven by their demonstrated potential to monitor IOPs over extensive spatial and temporal scales with airborne and spaceborne sensors. Given the relationships between the concentrations of seawater constituents and IOPs, inversion models provide a valuable tool to study biogeochemical parameters on a global scale. For example, satellite-based inversion algorithms have yielded substantial information about distribution and variability of chlorophyll-a concentration (Yoder et al., 1993; Yoder & Kennelly, 2003; Antoine et al., 2005; O'Reilly & Werdell, 2019; McClain et al., 2022), dissolved organic carbon (Mannino et al., 2008; Matsuoka et al., 2012; Joshi et al., 2017; Aurin et al., 2018), particulate organic carbon (Stramski et al., 1999; Stramska & Stramski, 2005; Stramski et al., 2008; Allison et al., 2010; Evers-King et al., 2017; Stramski et al., 2022), phytoplankton community composition (Hirata et al., 2008; Devred et al.,

2011; Li et al., 2013; Mouw et al., 2017), primary production (Feldman et al., 1984; Antoine et al., 1996; Behrenfeld et al., 2005; Uitz et al., 2010; Siegel et al., 2013; Gregg et al., 2017; Westberry et al., 2023), water quality (Hu et al., 2004; Shaeffer et al., 2012; Zheng et al., 2017; Turner et al., 2022), suspended particulate matter (Woźniak et al., 2010; Stramski et al., 2023), and climate change (Gregg & Conkright, 2002; Antoine et al., 2005; Kahru & Mitchell, 2008; Cael et al., 2023). Such studies are invaluable to the general oceanographic community and justify further advancements to inversion models for studying ocean optics and biogeochemistry.

A preferred approach to solving the inverse problem is to utilize values of remote sensing reflectance,  $R_{rs}(\lambda)$ , where  $\lambda$  is the wavelength of light in vacuum, to derive total IOPs. This approach is advantageous because total IOPs are the physical parameters most directly linked to observations of radiometric variables and ocean apparent optical properties and form the fundamental basis of the radiative transfer theory (Preisendorfer 1976). Of particular interest are the total spectral absorption coefficient,  $a(\lambda)$ , which quantifies the attenuation of light beam due to absorption processes per unit pathlength, and the total backscattering coefficient,  $b_b(\lambda)$ , which quantifies the attenuation of light beam due to backscattering per unit pathlength. The total absorption and backscattering coefficients are equal to the sum of contributions by optically significant seawater constituents. For the total absorption coefficient, the sum includes the additive contributions from pure seawater,  $a_w(\lambda)$ , phytoplankton,  $a_{ph}(\lambda)$ , non-algal particulates (also referred to as detritus or depigmented particles owing to the measurement methodology),  $a_d(\lambda)$ , and chromophoric (colored) dissolved organic matter (CDOM),  $a_g(\lambda)$ . Thus,  $a(\lambda) = a_w(\lambda) + a_{ph}(\lambda) + a_d(\lambda) + a_g(\lambda)$ . Often, these constituent absorption coefficients are grouped together for conceptual or operational reasons such as to express the non-water contribution,  $a_{nw}(\lambda) \equiv a(\lambda) - a_w(\lambda) = a_{ph}(\lambda) + a_d(\lambda) + a_g(\lambda)$ , the non-phytoplankton absorption coefficient,  $a_{d\&g}(\lambda) \equiv a_d(\lambda) + a_g(\lambda)$ , or the

particulate absorption coefficient,  $a_p(\lambda) \equiv a_{ph}(\lambda) + a_d(\lambda)$ . For the total backscattering coefficient, the sum includes the additive contributions from pure seawater,  $b_{bw}(\lambda)$ , and particulates,  $b_{bp}(\lambda)$ . Scattering by dissolved substances is assumed to be negligible, thus  $b_b(\lambda) = b_{bw}(\lambda) + b_{bp}(\lambda)$  (Jerlov 1976; Kirk 2011; Mobley 2022).

Given the scientific value of inversion models, the capabilities and limitations associated with models must be quantified through a thorough assessment of their performance and uncertainty. These assessments should be well-defined and adhere to the widely accepted guidelines of metrology. The objective of a measurement, or strictly speaking in this case an estimation, obtained from ocean color inversion algorithms is to determine the value of a well-defined physical quantity subject to measurement (estimation), called the measurand, through a measurement (estimation) procedure defined by the steps outlined in a particular algorithm (McKinna et al., 2019). The performance of an ocean color inversion algorithm is assessed by evaluating the difference between the estimated (algorithm-derived) value of the measurand and the true value of the measurand, this is known as error. In general, the ocean color community employs a suite of statistical metrics to quantify error and evaluate inversion models (Brewin et al., 2015; Seegers et al., 2018). Another key concept is that a model-derived measurement is an estimate of the measurand and is only complete when accompanied by an uncertainty value (Melin 2019). The uncertainty of a measurement is distinct from error and defines the range of values within which the true value of the measurand can be said to lie within a specified level of confidence (GUM 2008; VIM 2012). The ocean color community implements various approaches to assess model uncertainty, including validation using concurrent and coincident satellite-derived and in situ measurements (Antoine et al., 2008; Mélin et al., 2016), first-order first-moment methods (Lee et al., 2010; McKinna et al., 2019), and Monte Carlo simulations (Wang et al., 2005).

These performance and uncertainty assessments are essential for evaluating ocean color inversion models. Following these concepts we assess a newly proposed ocean color inversion algorithm.

The inversion algorithm of interest for this study is the 4-step Semi-Analytical Algorithm (4SAA) that consists of four component models which are implemented in a multi-step framework (Figure 3.1). This multi-component structure offers an advantage over other inversion methods by providing additional flexibility to interchangeably implement new or updated mechanistically-based component models, allowing for continual improvements as measurement methodology data availability and models change over time. The 4SAA model first solves for one specific AOP, then derives total absorption and backscattering coefficients, and finally partitions the non-water absorption coefficient into its constituent absorption coefficients. The 4SAA is designed to yield hyperspectral optical properties from the near-ultraviolet (near-UV) through visible (VIS) spectral region. Step 1 of the 4SAA model is a neural network, referred to as KdNN, which uses input  $R_{rs}(\lambda)$  to estimate the average diffuse attenuation coefficient of downwelling irradiance over the first attenuation depth,  $\langle K_d(\lambda) \rangle$ .  $K_d(\lambda)$  is an AOP, and, for brevity, we omit the average bracket notation for the remainder of this chapter. This neural network approach was first developed over a decade ago (Jamet et al., 2012) and has undergone multiple updates to improve model output. These updates include the utilization of more comprehensive training datasets, consideration of radiometric effects associated with solar zenith angle and Raman scattering, and the optimization of the model to consider different optical water types by adjusting the number of input parameters (Loisel et al., 2018; Jorge et al., 2021). In the present study, we implement a version of the KdNN that uses the input of  $R_{rs}(\lambda)$  at 12 spectral bands within the visible spectrum to estimate hyperspectral  $K_d(\lambda)$  from the near-UV (350–400 nm) through the VIS (400–700 nm). Step 2 of the 4SAA model is an inverse AOP-IOP model referred to as LS2 (Loisel et al., 2018), which

introduces multiple improvements to the similar approach described previously by Loisel and Stramski (2000). The LS2 model utilizes a look-up table approach to estimate spectrally independent solutions of total  $a(\lambda)$  and  $b_b(\lambda)$  and non-water  $a_{nw}(\lambda)$  and  $b_{bp}(\lambda)$  from input  $R_{rs}(\lambda)$  and  $K_d(\lambda)$  while avoiding assumptions about the spectral shapes of constituent IOPs. Step 3 of the 4SAA model utilizes an absorption partitioning model, called ANW, to separate  $a_{nw}(\lambda)$  into its  $a_{ph}(\lambda)$  and  $a_{dg}(\lambda)$  constituents (Stramski & Reynolds, in preparation). The model combines the stacked inequality constraints approach associated mostly with band ratios of  $a_{ph}(\lambda)$  (Zheng & Stramski, 2013) with a spectral shape function library of  $a_{dg}(\lambda)$  developed from a high-quality dataset containing measurements of constituent absorption coefficients. The model requires input of  $a_{nw}(\lambda)$  at a minimum of seven spectral bands (determined by values necessary to implement the pre-defined inequality constraints) to estimate  $a_{dg}(\lambda)$  from 350 to 700 nm with a 1 nm spectral interval as well as  $a_{ph}(\lambda)$  at light wavelengths that correspond to the input  $a_{nw}(\lambda)$ . In the present study, quantities of  $a_{nw}(\lambda)$  are generated at a 1 nm spectral interval, so both  $a_{dg}(\lambda)$  and  $a_{ph}(\lambda)$  retrievals are hyperspectral and span the complete near-UV through visible range. Step 4 of the 4SAA model is the ADG partitioning model (Kehrli et al., 2024, see also Chapter 2). The ADG model used in this study implements the ADG\_UV-VIS model variant and operates with input of  $a_{dg}(\lambda)$  to estimate  $a_d(\lambda)$  and  $a_g(\lambda)$  without assuming an exponential spectral shape for these coefficients. This model is the first of its kind to separate  $a_{dg}(\lambda)$  across the near-UV through the visible range while avoiding assumptions about the spectral shapes for its constituent absorption coefficients.

Here, we establish a methodology for examining the performance and uncertainty of the 4SAA algorithm. Given the initially promising results that were obtained from the evaluation of the individual component models of the 4SAA model (Loisel et al., 2018; Kehrli et al., 2024), here

we construct an advanced method to evaluate the 4SAA model using a recently published synthetic optical database (Loisel et al., 2023). This synthetic database is comprehensive and aligns with the high spectral resolution capabilities of NASA’s Plankton, Aerosol, Cloud and ocean Ecosystem (PACE) Ocean Color Instrument (OCI) (Werdell et al., 2019). We aim to determine a suite of error metrics to evaluate agreement between model estimates and reference values for each component model of the 4SAA model separately, as well as implement a Monte Carlo (MC) approach to approximate uncertainty and track its propagation through the multi-step sequence of the 4SAA component models. Specifically, we first quantify model performance at the individual component model level by assessing model retrievals for each step of the 4SAA model. This evaluation of model performance using only inputs of synthetic data quantifies the error produced from each component model under ideal conditions with minimal input error. Next, we assess error propagation as it passes downstream through each component model by using inputs from the preceding model. For example, we use  $a_{nw}(\lambda)$  estimated from Steps 1 and 2 as input into Step 3 of 4SAA and evaluate the error metrics of the resulting  $a_{ph}(\lambda)$  and  $a_{dg}(\lambda)$  model outputs. This assessment quantifies the impact of the multi-step 4SAA structure by determining how the upstream model errors propagate through subsequent steps. Finally, we evaluate the uncertainty and its propagation at each step within the multi-step framework of 4SAA. We define and apply a MC approach beginning with uncorrelated input  $R_{rs}(\lambda)$  uncertainties of 20, 5, and 10% from 350–399 nm, 400–600 nm, and 600–700 nm ranges, respectively, to estimate the uncertainty of all model outputs.

## **3.2 Methods**

### **3.2.1 Summary of Methodology**

The scope of the assessment of the 4SAA model depends on both the characteristics of the database used for evaluation and the methods applied to assess its performance and uncertainty. A

variety of data sources were considered for the initial assessment, each of which has its own advantages and disadvantages. Synthetic databases derived from radiative transfer simulations contain minimal error, as forward-modeled AOPs depend solely on errors associated with input IOPs. However, a potential disadvantage of this approach is that the database is constrained to the IOP dataset from which it was developed. If the IOP database is not globally distributed or contains some assumptions or errors, these limitations will propagate through the model and reduce the applicability of the synthetic database for the purpose of evaluating the inverse modeling results. There are also simplifications and assumptions associated with forward radiative transfer models which can lead to differences compared to observations from the natural environment. Another approach is to implement an in situ field database of measurements that contains a complete set of AOPs and IOPs. Although this method provides real-world optical property data, it has limitations which include measurement error and uncertainty, a limited number of observations with a full-suite of concurrent optical measurements, and a measurement distribution that is unlikely to represent the global ocean. Satellite measurements paired with ground-truth observations of optical properties were also considered. While global satellite observations provide globally distributed AOPs, satellite-in situ match-up measurements that contain both AOPs and IOPs are rare. Considering these factors and the main objective of this study, we employ a synthetic optical database (Loisel et al., 2023). This database is described in further detail in Section 3.2.2.

A key aspect of this study concerns the distinction between the methods used to assess the performance and uncertainty of the 4SAA model. Our study employs aggregate statistical metrics to quantify model error as an indicator of model performance. Aggregate statistical metrics compare model predictions against true or reference values to provide information about the performance of a model in terms of error. Model error consists of both systematic error, when the

model consistently over- or underestimates true values, and random error, which results in variations in repeated observations of a measurand. The complete suite of statistical metrics utilized are described in Section 3.2.3.

Although error metrics provide a basis to describe model performance, they are constrained because they only compare predictions to known outcomes. Alone, these metrics provide an indication of model performance but are independent of model uncertainty. Model uncertainty refers to the degree of the dispersion of values that can be reasonably attributed to a measurand. One method to assess the uncertainty of a model is the Monte Carlo approach, which randomly samples the distribution of model inputs through an iterative process to generate a distribution of model outputs that can be evaluated to quantify uncertainty. Additionally, MC analysis can be compared to model output to assess precision by determining how perturbations in model input influence the consistency of model predictions. The specific approach implemented in this study randomly perturbs model input variables from established estimates and examines the distribution of model outputs to form an uncertainty assessment of 4SAA. The complete method to quantify model uncertainty is described in Section 3.2.4.

Finally, the strategy to assess model performance and uncertainty can become impractical if every possible model implementation is considered. For example, one could evaluate each component model individually as well as all combinations of component models (e.g., using output from Step 1 as input to Step 2, using the output generated from the chained Step 1, 2 and 3 model as input for Step 4, etc.) to complete the full assessment of 4SAA. In total there are six combinations of coupled models that one can evaluate; however, this assessment is not necessary or efficient. Our assessment is structured to be both concise and comprehensive by aiming to

quantify the performance and uncertainty of 4SAA while reducing complexity as much as possible. The structure of the analysis is described in Section 3.2.5.

### 3.2.2 Summary of Synthetic Database

The performance and uncertainty assessment of the 4SAA model is based on a synthetic optical database developed by Loisel et al. (2023), which contains a full suite of spectral AOPs and IOPs. The optical properties from this database include measurements of  $a_{\text{ph}}(\lambda)$ , estimates of  $a_{\text{d}}(\lambda)$ ,  $a_{\text{g}}(\lambda)$ , and  $b_{\text{bp}}(\lambda)$ , as well as derived values of  $R_{\text{rs}}(\lambda)$  and  $K_{\text{d}}(\lambda)$  from radiative transfer simulations. This synthetic database serves as the basis for the assessment of model error and uncertainty propagation of each step of the complete 4SAA model. Additionally, the database provides a method to evaluate the model with a database that contains minimal measurement errors. Although the database contains near-surface measurements of  $a_{\text{ph}}(\lambda)$  obtained from various oceanic environments, the remaining IOPs within the dataset are derived from these measurements, and AOPs, including  $R_{\text{rs}}(\lambda)$ , are computed through forward radiative modeling using the IOPs as input to these simulations. Thus, the only source of measurement error within the synthetic database originates from sources associated with the collection of  $a_{\text{ph}}(\lambda)$  data. This characteristic of the synthetic database is essential for evaluating the 4SAA model as this task aims to only assess errors attributable to the techniques implemented in the model itself, requiring a database with minimal errors from other sources such as in situ measurements.

The original study that generated the synthetic optical database describes its characteristics and provides an in-depth comparison with in situ measurements of optical properties. Here, we highlight specific aspects of a subset of the database utilized for our analysis. The recently assembled optical synthetic database used in this study was developed using radiative transfer simulations that aim to address previous limitations of other similar synthetically generated datasets from past studies. The new synthetic dataset includes inelastic Raman scattering by water

molecules and chlorophyll-a fluorescence which have the potential to significantly contribute to light propagation in the upper ocean and impact AOPs (Marshall & Smith, 1990; Westberry et al., 2013; Li et al., 2016). Furthermore, the IOPs reported in the new synthetic database are consistent with observed global distributions of in situ measurements and satellite-based estimates of IOPs. The complete synthetic database contains 29880 simulations that rely on 3320 combinations of input IOPs, three different assumptions regarding inelastic radiative processes, three different solar zenith angles, and the assumption of an infinitely deep ocean with vertical homogeneity of IOPs. The radiative transfer-simulated AOPs of interest to this study represent the near-surface ocean layer. All optical properties reported in the synthetic dataset span from 350 to 700 nm with a 5 nm spectral interval. The subset of the synthetic database employed in this study uses radiative transfer scenarios that consider both Raman scattering and chlorophyll-a fluorescence and a solar zenith angle of  $0^\circ$ , i.e., the sun is directly overhead at its zenith.

In total, there are 3320 synthetic samples of IOPs and concurrent AOPs from the database used to assess the performance and uncertainty of 4SAA. Figure 3.2 displays the available  $R_{rs}(\lambda)$ ,  $K_d(\lambda)$ ,  $a_{nw}(\lambda)$ , and  $a_{dg}(\lambda)$  spectra utilized for the performance and uncertainty assessment. The compilation of  $R_{rs}(\lambda)$  spectra (Fig. 3.2a) suggests the collection of samples within the dataset is consistent with the global distributions and is representative of eutrophic, mesotrophic, and oligotrophic waters. The median and interquartile range of  $R_{rs}(\lambda)$  indicate the majority of spectra within the database are representative of open-ocean waters. This is further supported by the calculation of the median apparent visible wavelength (Vandermeulen et al., 2020) of the database, which is equal to 464 nm.  $K_d(\lambda)$  is most variable in the blue spectral region where it spans approximately 2 orders of magnitude with a median value of  $0.034 \text{ m}^{-1}$  at 440 nm. As wavelength increases into the red, the irradiance attenuation within the surface layer is increasingly dominated

by pure seawater absorption, and  $K_d(\lambda)$  begins to exhibit lower variability across most database samples (Fig. 3.2b). Importantly,  $a_{nw}(\lambda)$  and  $a_{dg}(\lambda)$  exhibit a wide range of variability, both of which span approximately 2 orders of magnitude across all wavelengths (Fig. 3.2c,d). The representative distribution of all optical properties and spectral variability enables us to evaluate the performance and uncertainty of the 4SAA model in the context of potential global applications.

### 3.2.3 Description of Performance Assessment

The performance assessment of 4SAA is completed by evaluating a suite of statistical metrics described in Table 3.1. These statistical measures compare model predictions,  $P_i$ , with observations,  $O_i$ , from the synthetic optical database. We evaluated the random error of the model by calculating the root-mean-square deviation, RMSD, the median absolute percent difference, MdAPD, and the median symmetric accuracy, MdSA, of model-derived values compared to values reported in the synthetic database. RMSD measures the average deviation between  $P_i$  and  $O_i$  to provide an assessment of random error in physical units of the variable being investigated, but it is excessively sensitive to outliers within model output. MdAPD is a more robust measurement and provides the normalized deviation between  $P_i$  and  $O_i$ , expressed as a percentage. However, in the calculation of MdAPD overestimations are penalized more heavily than underestimations of the same magnitude. Thus, we also calculated MdSA, which is equally affected by over- and underestimations of model output, to assess random deviations of model predictions (Morley et al., 2018). Additionally, we quantified the systematic error of model output by calculating the median ratio, MdR (dimensionless), of the predicted to observed variables, as well as the median bias (MdB), which is calculated as the residual between predicted and observed variables and therefore expressed in physical units of the variable in question. In this analysis, it is also important to recognize that not all original input samples or combinations of perturbed values from the synthetic database yielded valid model outputs. Because the number of valid retrievals varied

across models, model output parameters, and wavelengths, we explicitly report the number of valid model outputs, denoted as  $N_{\text{out}}$ .

We also used a model-II linear regression analysis based on the reduced major axis method to assess log10-transformed values of model predictions versus reference values from the synthetic optical database (Ricker 1973; Bellacicco et al., 2019). The best-fit coefficients which describe the regression include the slope,  $A$ , y-intercept,  $I$ , and Pearson correlation coefficient,  $r$ . The value of  $I$  is converted to the parameter,  $B$ , as described in Table 3.1, to express the y-intercept in logarithmic space for consistency with log10-transformed data. These statistical metrics provide additional context in the assessment of error for each model within study. For example, the ideal value for  $A$  is equal to 1, and deviations from this value suggest the potential of model bias near the upper or lower ends of the measurement dynamic range, provided the value of  $B$  remains close to 1. This type of systematic error appears in the correlation plots between model predictions and observations as over- or underestimations of observations at the extreme end of measurements on the  $x$ -axis.

### **3.2.4 Description of Uncertainty Assessment**

We use a Monte Carlo (MC) method for the assessment of model uncertainty and its propagation for the outputs produced in each step of 4SAA. The method is a well-established technique for estimating model output uncertainty in optical oceanography and the broader Earth science community (Anderson 1976; Wang et al., 2005; Refsgaard et al., 2007). MC simulations calculate an estimate of model uncertainty by repeatedly perturbing model inputs according to a predefined uncertainty distribution. At each iteration, input values are randomly sampled from this distribution, and new model outputs are determined. After a sufficient number of iterations, a distribution of model output across all simulations is generated, which is analyzed to quantify model uncertainty (Anderson 1976). In regard to this study, MC techniques provide a means to

estimate the uncertainty resulting from each step of the complete 4SAA model and track the uncertainty propagation through the multi-step sequence of 4SAA component models.

The MC method used to assess uncertainties of 4SAA follows a similar approach as described by the uncertainty assessment of other ocean color algorithms (McKinna et al., 2019). Recall that in Step 1 of 4SAA,  $K_d(\lambda)$  is retrieved from synthetic  $R_{rs}(\lambda)$  input using the KdNN model. The estimates of uncertainty for  $K_d(\lambda)$  are determined by defining the characteristics of the  $R_{rs}(\lambda)$  input uncertainty distribution. For our analysis, we consider the allowable uncertainties of the required science data products for the NASA PACE mission. These uncertainties are defined as 20%, 5%, and 10% for  $R_{rs}(\lambda)$  in the spectral ranges from 350 to 400 nm, 400 to 600 nm, and 600 to 700 nm, respectively (Werdell et al., 2019). Each  $R_{rs}(\lambda)$  input is perturbed independently using a normal distribution centered on its reference value from the synthetic database, with a standard deviation defined by the relative uncertainty. For example, consider a hypothetical model where input  $R_{rs}(440) = 0.01 \text{ sr}^{-1}$  and  $R_{rs}(610) = 0.003 \text{ sr}^{-1}$ . Given the defined uncertainty distribution of  $R_{rs}(\lambda)$ , input values are randomly selected from the normal distribution defined by  $R_{rs}(440) = 0.01 \pm 0.0005 \text{ sr}^{-1}$  and  $R_{rs}(610) = 0.003 \pm 0.00015 \text{ sr}^{-1}$ . We also considered absolute uncertainties for PACE OCI-derived  $R_{rs}(\lambda)$ , however, we found that applying these uncertainties across broad spectral ranges produced unrealistically low and occasionally negative  $R_{rs}(\lambda)$  which limited the analyses. Next, we ran 100 model simulations for all sample inputs where at each iteration the input to the model (e.g., the input  $R_{rs}(\lambda)$  to the KdNN model) is randomly perturbed using the predefined probability distribution. Each iteration produces a unique model output for every sample, and the relative model uncertainty is quantified by calculating the median coefficient of variation (CV) across all samples, expressed as a percentage. For example, the relative uncertainty of Step 1 of 4SAA (i.e., KdNN model) is determined by running the model 100 times for all 3320

samples within the synthetic database. The CV for each of the 3320 resulting  $K_d(\lambda)$  distributions is then calculated. The median of these values is determined and quantifies the relative uncertainty for this component model of 4SAA. The formulation of relative uncertainty is expressed in Table 3.1. The same methodology is repeated for downstream models of 4SAA, in which the input uncertainty distribution of each component model is defined with the relative uncertainty from the preceding upstream model.

### **3.2.5 Strategy to Assess the Performance and Uncertainty of 4SAA**

We employ a strategy to assess the performance of 4SAA over two stages, first evaluating the component models individually and then combining component models in a sequential manner. By evaluating each component model individually, we assess their ability to estimate output variables under optimal input conditions (i.e., the actual input values from the synthetic database without MC perturbations) to complete an independent assessment of each model's performance. AOPs and IOPs from the synthetic optical database were interpolated to a 1 nm spectral interval and used as input for each component model to conduct the performance assessment. In the first stage of our assessment, Step 1 of 4SAA (KdNN model) is evaluated using synthetic  $R_{rs}(\lambda)$  as input, and model-predicted values of  $K_d(\lambda)$  are compared to synthetic  $K_d(\lambda)$  to assess model error. The performance of the remaining component models (LS2, ADG, and ANW) are assessed in a similar manner. The LS2 model takes synthetic  $R_{rs}(\lambda)$  and  $K_d(\lambda)$  as input, and the resulting model output of  $a(\lambda)$ ,  $b_b(\lambda)$ ,  $a_{nw}(\lambda)$ , and  $b_{bp}(\lambda)$  is compared with the corresponding values in the synthetic database. Next, we use synthetic  $a_{nw}(\lambda)$  as input for the ANW model and compare model predictions of  $a_{ph}(\lambda)$  and  $a_{dg}(\lambda)$  with the corresponding values in the synthetic database. Finally, the ADG model uses input of synthetic  $a_{dg}(\lambda)$  and the resulting estimates of  $a_d(\lambda)$  and  $a_g(\lambda)$  are compared with corresponding values in the synthetic database. If available, we compare the

behavior and spectral trends of the performance statistics with previous studies of each component model, which were evaluated with different synthetic or in situ datasets.

We then completed a performance assessment of three of the six possible combinations of the coupled component models that are required to evaluate the complete performance of the 4SAA model when operated in a step-wise manner. The three scenarios included in our analysis are: (i) the coupled Steps 1 and 2 of 4SAA, in which synthetic  $R_{rs}(\lambda)$  and KdNN-derived  $K_d(\lambda)$  are used as input for the LS2 model to obtain values of  $a(\lambda)$ ,  $b_b(\lambda)$ ,  $a_{nw}(\lambda)$ , and  $b_{bp}(\lambda)$ ; (ii) the coupled Steps 1, 2, and 3 of 4SAA, in which  $a_{nw}(\lambda)$  acquired from scenario (i) is used as input for ANW model to obtain values of  $a_{ph}(\lambda)$  and  $a_{dg}(\lambda)$ ; and (iii) the coupled Steps 1, 2, 3, and 4 of 4SAA, in which  $a_{dg}(\lambda)$  acquired from scenario (ii) is used as input for ADG model to obtain values of  $a_d(\lambda)$  and  $a_g(\lambda)$ . We note that comparisons of model predictions to synthetic values of  $a_{dg}(\lambda)$ ,  $a_d(\lambda)$ , and  $a_g(\lambda)$  may be limited because these IOPs are generated in the synthetic database by using a fixed exponential function to represent the whole spectra of  $a_d(\lambda)$  and  $a_g(\lambda)$ . Nevertheless, these synthetic IOPs are still useful for an initial assessment of the ANW and ADG models beyond the field datasets that were used for their development (see Chapter 2), and furthermore, this method is consistent with the other evaluations of performance utilized in this study.

The uncertainty assessment strategy based on MC simulations first estimates the uncertainty in Step 1 (KdNN model) and then follows the multi-step structure of 4SAA to estimate the uncertainty produced in the remaining steps of 4SAA. We first assessed the uncertainty in Step 1 by using synthetic  $R_{rs}(\lambda)$  and conducting a MC-based assessment by sampling the input uncertainty distribution of  $R_{rs}(\lambda)$  as quantified by the relative uncertainty described in Werdell et al. (2018). We then evaluated the uncertainty in the multi-step scenarios (i), (ii), and (iii) that are described in the paragraph above by using the same model inputs and conducting MC assessments

with input uncertainty distributions propagated from the previous step. For scenario (i), the input uncertainty distribution is defined by the relative uncertainty of  $R_{rs}(\lambda)$  described in Werdell et al. (2018) and the relative uncertainty of  $K_d(\lambda)$  propagated from the KdNN model. For scenario (ii), the input uncertainty distribution is defined by the relative uncertainty of  $a_{nw}(\lambda)$  propagated from scenario (i). Finally, for scenario (iii), the input uncertainty distribution is defined by the relative uncertainty of  $a_{dg}(\lambda)$  propagated from scenario (ii). This step-wise uncertainty assessment provides a well-defined approach to quantify the propagation of uncertainty throughout the complete 4SAA model by accounting for uncertainties produced by each upstream component model and provides the foundation for a comprehensive evaluation of 4SAA.

### **3.3 Results and Discussion**

#### **3.3.1 Performance Assessment of Individual Component Models**

An important distinction between the performance assessment of individual component models of 4SAA and the assessment of coupled-component models within the multi-step structure of 4SAA is that the individual component model analysis isolates each model for a detailed examination of its output. The individual component model analysis is valuable for identifying and attributing model error from each step of 4SAA under ideal conditions, where model input is free of both measurement error and model-propagated error to compare predictions with reference values from the synthetic optical database. For example, the independent performance assessment of Step 2 (LS2 model) with synthetic values of  $R_{rs}(\lambda)$  and  $K_d(\lambda)$  quantifies the retrieval error for the total and non-water IOPs,  $a(\lambda)$ ,  $b_b(\lambda)$ ,  $a_{nw}(\lambda)$ , and  $b_{bp}(\lambda)$ , produced from the LS2 model alone, without interference of potential errors generated by Step 1 (KdNN model). Furthermore, the independent evaluation of the absorption partitioning models, ANW and ADG, in Steps 3 and 4 of 4SAA compares model output to synthetic reference values. We recall that the synthetic database includes the assumption of fixed exponential spectral shapes of  $a_d(\lambda)$  and  $a_g(\lambda)$ . This assessment

will reveal the compatibility of each model with this type of synthetic dataset and provide insight into the potential model performance as well as guidance on the potential range of applications (e.g., spectral range or range of optical water types).

We begin the individual model performance assessment by evaluating the KdNN model using synthetic  $R_{rs}(\lambda)$  input. Figure 3.3 presents a comparison of KdNN-derived and synthetic reference values of  $K_d(\lambda)$  at six example wavelengths: 350, 440, 490, 510, 550, and 650 nm. The scatter plots include one UV wavelength (350 nm) to directly evaluate the model results at the shortest wavelength of output, where the largest errors are expected. This expectation is supported by this analysis. Figure 3.3 also highlights the correlations between model predictions and synthetic values in the blue and green portions of the spectrum (440, 490, 510, and 555 nm) to focus on the spectral region where the magnitude of input  $R_{rs}(\lambda)$  is typically the largest. These wavelengths are also near-analogous to measurements provided by heritage ocean color sensors. The red spectral region is assessed at 650 nm, which is where pure water absorption usually dominates the attenuation of light. Thus, the results presented in Fig. 3.3 summarize the capability of KdNN model to estimate  $K_d(\lambda)$  across a broad spectral range that is consistent with the performance goals of Step 1 of 4SAA.

Figure 3.3 shows relatively low scatter of data points which are evenly distributed along the 1:1 line between model predictions and synthetic data at all six wavelengths, indicating strong model performance across the near-UV through visible spectral range. This finding is further supported by the small random error observed as MdSA is consistently below 5% in the visible. Random error tends to increase with decreasing wavelength as model predictions extend to the near-UV, but remain within acceptable limits, as values of MdSA stay below 10%. Additionally, these results indicate minimal systematic error in  $K_d(\lambda)$  retrievals since points on the scatter plot

generally fall along the 1:1 line across the dynamic range of model output, and MdR values remain close to a value of 1. In general, this performance assessment of Step 1 of 4SAA demonstrates that the KdNN model provides reliable estimates of  $K_d(\lambda)$  using synthetic inputs of  $R_{rs}(\lambda)$  across the near-UV through visible spectrum.

The next stage of the individual model performance assessment evaluates the LS2 model using synthetic input of  $K_d(\lambda)$  and  $R_{rs}(\lambda)$  to generate estimates of  $a(\lambda)$ ,  $b_b(\lambda)$ ,  $a_{nw}(\lambda)$ , and  $b_{bp}(\lambda)$ . We reiterate that in this scenario, model inputs are free of measurement error and any influence from outputs generated by KdNN model. Figure 3.4 presents the spectral behavior of two error metrics, MdR and MdSA, to characterize the performance of the LS2 model in retrieving total and non-water IOPs. A strong correlation is observed between model predictions and synthetic reference values across the dynamic range of model output. Specifically, the results indicate a strong model performance in retrievals of the total IOPs,  $a(\lambda)$  and  $b_b(\lambda)$ , across most of the spectral range of interest. The MdR values indicate nearly negligible positive systematic error and high accuracy in estimates of  $a(\lambda)$ , with MdSA consistently below 4%. The performance in retrievals of  $b_b(\lambda)$  is slightly diminished but remains acceptable for most wavelengths below 660 nm where MdR and MdSA are consistently below 1.15 and 15%, respectively. However, the model performance weakens at longer wavelengths as indicated by the sharp increase in MdR and MdSA for  $b_b(\lambda)$  beyond 660 nm. This deterioration is attributable to the effect of chlorophyll-a fluorescence on  $R_{rs}(\lambda)$  which was included in the generation of the synthetic database but was omitted from the process of the LS2 model development. Consequently, the presented retrievals of  $b_b(\lambda)$  from LS2 model at red wavelengths are enhanced because of the fluorescence-induced enhancement of input  $R_{rs}(\lambda)$  in the red spectral band.

Figure 3.4 also demonstrates that retrievals of non-water IOPs exhibit a strong overall model performance, displaying analogous behavior to their total IOP counterparts across specific portions of the near-UV and visible spectral region. Retrievals of  $a_{nw}(\lambda)$  and  $a(\lambda)$  perform similarly in the short-wavelength portion of the spectrum. Estimates of  $a_{nw}(\lambda)$  agree well in comparison to synthetic values in blue spectral range between 350 and 500 nm, where MdR and MdSA are consistently below 1.03 and 3%, respectively. However, beyond 500 nm, the spectral error statistics show a significant decline in the ability of LS2 to accurately predict  $a_{nw}(\lambda)$ . In this spectral region, pure seawater absorption begins to increase and contributes an increasing fraction of the total magnitude of  $a(\lambda)$  relative to the non-water components, with the pure water contribution becoming usually highly dominant in the red portion of the spectrum. Thus, even if the total  $a(\lambda)$  in the red is retrieved with a small error, the error in  $a_{nw}(\lambda)$  can become large. A similar behavior in retrievals of  $a_{nw}(\lambda)$  was observed when LS2 is applied to a different synthetic database of optical properties in Loisel et al. (2018) (see Fig. 7 and Table 1 in their publication). Given our understanding of the contribution of pure seawater to total spectral absorption of seawater, along with the similarities in the findings of this study and previously published findings, we conclude that the LS2 model does not adequately retrieve the non-water absorption coefficient,  $a_{nw}(\lambda)$ , beyond approximately 500 nm. This limitation suggests the need for further improvements or modifications to the model to better account for the significant or dominant contribution of pure seawater absorption to total absorption in the green and red spectral regions. Such improvements will be particularly important to obtain adequate estimates of  $a_{nw}(\lambda)$  across the complete near-UV through visible spectral range, which is also required for reliable retrievals of constituent absorption coefficients in downstream absorption partitioning models of 4SAA.

Compared to retrievals of  $b_b(\lambda)$ , the spectral error statistics presented in Figure 3.4 for  $b_{bp}(\lambda)$  suggest a similar spectral trend but slight degradation in the performance. In the visible portion of the spectrum there is a positive bias and MdSA remains generally below 30%. At wavelengths extending into the near-UV MdSA significantly increases, which can be attributed largely to the increasing contribution of pure seawater backscattering to total backscattering as wavelength decreases into the near-UV. This effect is expected to be particularly well-pronounced in very clear waters with low levels of backscattering. The deterioration of  $b_{bp}(\lambda)$  at wavelengths greater than 650 nm is similar to results of  $b_b(\lambda)$  over the same portion of the spectrum and is again attributable to the increased values of input reflectance due to chlorophyll-a fluorescence.

In the next stage of analysis, we investigate the performance of the Step 3 ANW absorption partitioning model with synthetic inputs of  $a_{nw}(\lambda)$  to generate spectral estimates of  $a_{ph}(\lambda)$  and  $a_{dg}(\lambda)$ . The ANW model is designed to produce output spectra only if they satisfy specific built-in spectral inequality constraints that are consistent with a large range of variability in the measured spectral data across diverse oceanic environments. In this analysis not all synthetic inputs of  $a_{nw}(\lambda)$  lead to a solution of partitioned  $a_{ph}(\lambda)$  and  $a_{dg}(\lambda)$  at all wavelengths as indicated by  $N_{out}$  in Table 3.2. Table 3.2 also presents the errors statistics for five select wavelengths to provide a quantitative assessment of ANW model performance across a broad spectral range spanning from the near-UV to the red (350, 440, 500, 550, and 650 nm). The ANW model performs well in the visible spectral region for retrievals of both  $a_{ph}(\lambda)$  and  $a_{dg}(\lambda)$ . Model error metrics for derived  $a_{ph}(\lambda)$  in the visible (440, 500, 550, and 650 nm) indicate that MdAPD remains consistently below 30%. MdAPD increases in the long-wavelength portion of the spectrum, a trend that also appears in MdSA. Notably, MdSA shows a more pronounced increase at 550 and 650 nm equal to 37% and 42%, respectively. The MdR values range between 0.71 at 650 nm and 0.95 at 440 nm which indicates

that the ANW model tends to underestimate  $a_{\text{ph}}(\lambda)$  across the visible spectral range of interest. Regarding the spectral error metrics in model predictions of  $a_{\text{dg}}(\lambda)$ , the lowest retrieval errors for  $a_{\text{dg}}(\lambda)$  occur in the blue to blue-green spectral range (440–550 nm) where MdAPD and MdSA values remain below 15% and model bias indicates that MdR ranges from 1.114 at 440 nm to 1.055 at 550 nm. Beyond 550 nm, ANW retrievals of  $a_{\text{dg}}(\lambda)$  also begin to deteriorate and overestimations increase substantially as MdR is equal to 1.356 and random error rises to 35.99% and 36.78% in MdAPD and MdSA, respectively. This deteriorated performance of both  $a_{\text{ph}}(\lambda)$  and  $a_{\text{dg}}(\lambda)$  in the red (650 nm) can be attributed to the relatively low values of  $a_{\text{nw}}(\lambda)$  in this spectral region.

In the near-UV, the performance of the ANW model deteriorates. At the shortest wavelength of interest in this study (350 nm), the results suggest reasonable retrievals of  $a_{\text{dg}}(\lambda)$  with MdR and MdSA values increasing to 1.228 and 22.76%, respectively. These error statistics indicate an increasing positive bias and greater random error in the short-wavelength portion of the spectrum. However, there are much fewer retrievals of  $a_{\text{ph}}(\lambda)$  in the near-UV that are notably worse in comparison to the performance of  $a_{\text{ph}}(\lambda)$  estimates in the visible. The best-fit regression line of model versus synthetic  $a_{\text{ph}}(350)$  shows  $r = 0.5701$  which is a significant decrease from values reported at visible wavelengths (Table 3.2). This weaker correlation is also observed in the scatter plot comparing model predictions and reference values of  $a_{\text{ph}}(350)$  (not shown). Additionally, the MdR value for retrievals of  $a_{\text{ph}}(350)$  is equal to 0.269 indicating that estimates of  $a_{\text{ph}}(\lambda)$  in the near UV are greatly underestimated.

These observations suggest that the ANW model can reasonably partition  $a_{\text{nw}}(\lambda)$  into its constituent absorption spectra in the blue and blue-green portions of the spectrum. Furthermore, the error statistics describing the performance of  $a_{\text{dg}}(\lambda)$  retrievals in the near-UV suggest that the model provides reasonable estimates of this absorption coefficient at the short-wavelength portion

of the spectrum of interest in this study. However, the retrievals of  $a_{\text{ph}}(\lambda)$  in both the short- and long-wavelength spectral regions (i.e., 350–400 nm and 550–700 nm) as well as  $a_{\text{dg}}(\lambda)$  at wavelengths longer than 550 nm, should be considered with great caution.

We provide error statistics to assess the performance of the ADG absorption partitioning model at four wavelengths (350, 440, 500, and 550 nm) with a special focus on retrievals at 440 nm. The emphasis on the ADG model performance at 440 nm is motivated by the characteristics of the synthetic database and the design of the model. First, the available spectra of  $a_{\text{d}}(\lambda)$  and  $a_{\text{g}}(\lambda)$ , and consequently  $a_{\text{dg}}(\lambda)$ , in the synthetic database are calculated based on assumptions about their magnitude and spectral behavior. The magnitudes of synthetic  $a_{\text{d}}(440)$  and  $a_{\text{g}}(440)$  are estimated using empirical relationships established from measurements of  $a_{\text{ph}}(\lambda)$  included in the database. The spectral values of both absorption coefficients are then obtained by assuming they follow an exponential function with the spectral slope parameter selected randomly from a predetermined range of values (Loisel et al., 2023). Second, the ADG partitioning model is developed using a spectral shape function library that is free of exponential spectral shape assumptions and optimized to work with real-world inputs of  $a_{\text{dg}}(\lambda)$  which deviate from a simple exponential function (Kehrli et al., 2024, see also Chapter 2). Given the empirical relationships used to calculate  $a_{\text{d}}(\lambda)$  and  $a_{\text{g}}(\lambda)$  within the synthetic database and the design of the ADG partitioning model, 440 nm is an appropriate choice of wavelength to assess model performance. We also expand our analysis to three additional wavelengths to evaluate changes in model performance as we move away from the assessment at 440 nm. We exclude analysis in the red portion of the spectrum as the magnitudes of these constituent absorption coefficients are typically very small or negligible in this spectral domain.

Table 3.2 presents the results from the performance assessment of the ADG model using synthetic inputs of  $a_{dg}(\lambda)$ . Focusing on the performance at 440 nm, the results show a substantial negative bias in retrievals of  $a_d(440)$  with MdB equal to  $-0.00161 \text{ m}^{-1}$  and MdR equal to 0.543 which indicates the model underestimates  $a_d(440)$ . In contrast, the bias assessment shows that  $a_g(440)$  is overestimated with MdB and MdR values equal to  $0.00158 \text{ m}^{-1}$  and 1.249, respectively. The best-fit regression analysis of model-predicted versus synthetic reference data demonstrates that the data are highly scattered around the 1:1 line with  $r$  equal to 0.7219 for  $a_d(440)$  and equal to 0.894 for  $a_g(440)$ . The error metrics reveal that random error exceeds 50% for  $a_d(440)$  and 30% for  $a_g(440)$ , which is further exemplified by MdSA values of 105% and 35% for  $a_d(440)$  and  $a_g(440)$ , respectively.

Compared to a previous analysis presented in Kehrli et al. (2024) (see also Chapter 2), the model performance results in this study are notably weaker across all error metrics. However, we postulate that these weaker results do not indicate that the model is unreliable but rather highlight the need for caution when selecting input data for the model. As previously mentioned, non-phytoplankton absorption values of  $a_d(\lambda)$  and  $a_g(\lambda)$  available in the synthetic database are derived from empirical relationships with  $a_{ph}(440)$  and constrained by the assumption of a single exponential spectral shape. The ADG model is optimized to generate solutions using the spectral shape function libraries containing the actual spectral measurements with no exponential fits to the data, so the model is particularly sensitive to the spectral shape of the input data by design, especially when a single exponential function is assumed to represent the entire spectrum of  $a_d(\lambda)$  or  $a_g(\lambda)$ . In this scenario, the fixed shape of input data affects the ability of the model to accurately estimate non-phytoplankton constituent absorption coefficients, even at a single wavelength, as demonstrated in this analysis. Similar remarks can be made for the performance of the ANW

model, which also utilizes a spectral library of  $a_{dg}(\lambda)$  shapes determined from field measurements of  $a_d(\lambda)$  and  $a_g(\lambda)$  that do not utilize any assumption of fixed exponential shapes.

### 3.3.2 Performance Assessment of Sequential Component Models

The evaluation of component models combined in sequential steps of 4SAA aims to quantify the propagation of model error in model predictions at each step of the 4SAA model. The assessment of model performance after every step of the 4SAA model is essential because each step introduces additional errors that vary in magnitude and impact the performance of downstream model outputs. This assessment will reveal the sensitivity of the 4SAA model to errors from individual component models when used in sequence to obtain a complete suite of IOPs and AOPs. The analysis begins by utilizing synthetic  $R_{rs}(\lambda)$  and KdNN-derived  $K_d(\lambda)$  as input for the LS2 model to examine the performance of the combined Step 1 and Step 2 models. Next, estimates of  $a_{nw}(\lambda)$  derived from the coupled KdNN-LS2 model are used to evaluate predictions generated in the ANW model when combined with the previous two steps of the 4SAA model. The final analysis evaluates the influence of model-derived  $a_{dg}(\lambda)$  from the previous three steps on retrievals in the ADG model. These three scenarios provide key insights into how each individual step of the 4SAA model contributes to the total error in model predictions.

We focus our analysis of the coupled Steps 1 and 2 models on comparisons between model predictions of non-water IOPs with synthetic reference values. The analysis is presented for three example wavelengths in the near-UV and blue portions of the spectrum (350, 440, and 500 nm) in Fig. 3.5. We observe strong agreement between model predictions and synthetic values of  $a_{nw}(\lambda)$  in this spectral range with a high correlation ( $r > 0.98$ ) and limited scatter between points depicted in the plot. Error statistics indicate that  $a_{nw}(\lambda)$  is typically overestimated as MdR values exceed 1.05 (5% positive bias), and this positive bias increases at longer wavelengths (MdR = 1.164 at 500 nm). An analysis of the model performance at longer wavelengths, which is omitted for

brevity, revealed a deterioration in model performance above 500 nm. The spectral behavior of the error statistics describing the performance of model retrievals of  $a_{nw}(\lambda)$  is consistent with the individual component model assessment of LS2 and demonstrates a subtle degradation when using model-predicted  $K_d(\lambda)$  values. In particular, MdSA increases by 9.43%, 2.74%, and 13.81% at 350, 440, and 500 nm, respectively, when using KdNN-estimated values of  $K_d(\lambda)$  instead of synthetic values, suggesting that Step 1 introduces approximately 3–14% additional error to LS2 model retrievals in this spectral region (Fig. 3.5a–c). There is substantial positive bias in estimates of  $b_b(\lambda)$  as MdR exceeds 1.25 at all analyzed wavelengths. Furthermore, estimates of  $b_b(\lambda)$  in the UV show weak correlation with synthetic values at 350 nm ( $r = 0.457$ ) and model error deteriorates (MdSA = 72.94%). These spectral error trends are similar to those observed in the individual component model analysis using synthetic inputs of  $R_{rs}(\lambda)$  and  $K_d(\lambda)$  and we can quantify the additional error introduced by coupling Steps 1 and 2. In this case, MdSA increases by 44.42%, 3.66%, and 12.56% at the three wavelengths presented in the analysis, indicating that Step 1 contributes an additional 4–13% error in the blue and greater than 40% error in the near-UV (Fig. 3.5d–f).

Figure 3.6 displays the performance of model-predicted versus synthetic values of  $a_{ph}(\lambda)$  and  $a_{dg}(\lambda)$  obtained from the ANW model using the  $a_{nw}(\lambda)$  inputs retrieved from the coupled Step 1 and Step 2 models. Results are presented at 440, 500, and 550 nm which represent the blue and green portion of the spectrum. Notably, we omit presenting results in the near-UV in this scenario as they were similar to individual component model analysis of ANW described in the previous section and from a qualitative perspective that  $a_{ph}(\lambda)$  retrievals were poor, and that retrievals of  $a_{dg}(\lambda)$  were reasonable. MdR values for  $a_{ph}(\lambda)$  retrievals are equal to 1.210 and 1.589 at 440 and 500 nm, respectively, and correlation plots indicate that the model produces stronger overestimates

for smaller values of  $a_{\text{ph}}(\lambda)$  in this spectral region. This behavior subsides at 550 nm where MdR values indicate that model predictions of  $a_{\text{ph}}(550)$  are typically underestimated, albeit with greater scatter compared to 440 and 500 nm. Compared to the individual model assessment (Table 3.2), MdSA values increase by 9.93%, 47.83%, and 10.03% at 440, 500, and 550 nm, respectively, suggesting that model error of  $a_{\text{ph}}(\lambda)$  retrievals increases by approximately 10–50% in the blue and green portions of the spectrum when used in conjunction with the KdNN and LS2 models. In contrast to the  $a_{\text{ph}}(\lambda)$ , the coupled ANW, LS2, and KdNN models demonstrate a stronger performance in retrievals of  $a_{\text{dg}}(\lambda)$  at all three wavelengths as MdSA is less than 30% and data points are distributed more evenly about the 1:1 line. Retrievals of  $a_{\text{dg}}(\lambda)$  are typically underestimated as MdR ranges from 0.842 at 500 nm to 0.926 at 440 nm. In comparison to the individual model assessment (Table 3.2), MdSA values increase by 6.77%, 17.07%, and 12.95% at 440, 500, and 550 nm, respectively, suggesting that the model error in  $a_{\text{dg}}(\lambda)$  retrievals increases by approximately 7–18% from 440 to 550 nm when used in conjunction with the KdNN and LS2 models.

The error metrics describing the performance of retrievals of  $a_{\text{d}}(\lambda)$  and  $a_{\text{g}}(\lambda)$  at four wavelengths (350, 440, 500, and 550 nm) from the complete multistep sequence of the 4SAA model are available in Table 3.3. Most notably, the performance assessment shows an improvement in  $a_{\text{d}}(\lambda)$  retrievals as MdSA decreases by 44.78%, 38.67%, 34.99%, and 21.7% for 350, 440, 500, and 550 nm, respectively. We attribute the improved performance of the multistep algorithm to the rejection of potentially problematic  $a_{\text{dg}}(\lambda)$  values derived from Steps 1 through 3 as well as to the additional variability introduced by the multistep sequence which allowed inputs to deviate from a single exponential function. In contrast, the error assessment of  $a_{\text{g}}(\lambda)$  retrievals shows MdSA increasing by 27.33%, 0.87%, and 9.68% for 350, 400, and 500 nm while decreasing

by 3.76% for 550nm. Furthermore, and more importantly, correlations between model-predicted and synthetic values of both coefficients are generally weaker in comparison to the evaluation of the standalone ADG model (Table 3.2). These findings suggest that while some aggregate error statistics indicate improved model performance in this scenario, the use of synthetic data with fixed exponential spectral shapes of  $a_d(\lambda)$  and  $a_g(\lambda)$  is incompatible for providing a sufficiently meaningful assessment of all outputs produced by Step 4 of 4SAA.

### 3.3.3 Uncertainty Assessment of Sequential Component Models

The uncertainty assessment of each step in the complete 4SAA model is conducted in a stepwise manner to calculate the relative uncertainty introduced by each component model. This evaluation is performed using MC simulations on each component model with assumptions about the uncertainty associated with all model inputs. We reiterate that input uncertainties used for the assessment of each model are obtained from the prior MC simulation performed on the respective upstream model. The first step of the multistep sequence assumes  $R_{rs}(\lambda)$  input uncertainties as described in Table 1 of Werdell et al. (2019). The subsequent relative uncertainties for downstream inputs are calculated as the median coefficient of variation, expressed as a percentage.

Figure 3.7 depicts the results of the MC uncertainty analysis of outputs for each component model of 4SAA at 1 nm increments from 350 to 700 nm. We find the relative uncertainty of  $K_d(\lambda)$  from Step 1 KdNN model has a maximum value of 12.5% at 350 nm and decreases in a smooth and gradual manner as it reaches a minimum value of 1% at 628 nm (Fig. 3.7a). The relative uncertainty of  $a(\lambda)$  obtained from the Step 2 LS2 model closely follows the previous analysis for  $K_d(\lambda)$ , suggesting that the LS2 introduces relatively little additional uncertainty in  $a(\lambda)$  retrievals when applied on the synthetic database (Fig. 3.7b). This result is consistent with the fact that  $a(\lambda)$  is the most significant contributor to downward irradiance attenuation within the synthetic dataset. The relative uncertainty of  $a_{nw}(\lambda)$  is slightly larger (~15% at 350 nm, gradually decreasing to ~13%

at 440 nm) and follows a similar trend to  $a(\lambda)$  in the blue portion of the spectrum. However, a marked change in this trend occurs at  $\sim 440$  nm, where the relative uncertainty of  $a_{\text{nw}}(\lambda)$  begins to diverge from the relative uncertainty of  $a(\lambda)$ , becoming more pronounced above 500 nm. The relative uncertainty spectrum of  $a_{\text{nw}}(\lambda)$  becomes noisier and reaches a mean value of 26% for wavelengths greater than 500 nm. The behavior of the uncertainties of  $b_{\text{b}}(\lambda)$  and  $b_{\text{bp}}(\lambda)$  are similar and achieve their largest values in the near-UV of  $\sim 20\%$  for  $b_{\text{b}}(\lambda)$  and  $\sim 40\%$  for  $b_{\text{bp}}(\lambda)$ . Both scattering uncertainties also experience discontinuities at 400 nm and 600 nm, which can be attributed to uncertainty discontinuities of  $R_{\text{rs}}(\lambda)$  inputs at these wavelengths.

Results from the Step 3 ANW model reveal contrasting behaviors in the calculated uncertainties of  $a_{\text{ph}}(\lambda)$  and  $a_{\text{dg}}(\lambda)$  (Fig. 3.7c). The relative uncertainty of  $a_{\text{dg}}(\lambda)$  is continuous across the spectrum and ranges from a minimum value of 25% at 430 nm to a maximum value of 50% at 700 nm with a mean value of 36%. Alternatively, the calculated relative uncertainty for  $a_{\text{ph}}(\lambda)$  exceeds 100% in the near-UV and displays a noisy behavior in the visible portion of the spectrum that ranges from 21% at 412 nm to 84% at 552 nm with a mean value of 37%. Finally, the uncertainty assessment of the ADG model shows continuous functions for  $a_{\text{d}}(\lambda)$  and  $a_{\text{g}}(\lambda)$  (Fig. 3.7d). The relative uncertainty  $a_{\text{g}}(\lambda)$  ranges from 7% at 368 nm to 35% at 700 nm with a mean value of 21%. In comparison, the relative uncertainty of  $a_{\text{d}}(\lambda)$  is approximately 30–55% larger than  $a_{\text{g}}(\lambda)$  and spans from 63% at 461 nm to 66% at 700 nm. This analysis demonstrates the lowest estimated relative uncertainties from the MC analysis occur in spectral regions where a particular constituent absorption coefficient is a dominant contributor to the absorption coefficient from which it is derived. For example, relative uncertainties are the smallest for  $a_{\text{nw}}(\lambda)$ ,  $a_{\text{dg}}(\lambda)$ , and  $a_{\text{g}}(\lambda)$  from approximately 350 to 500 nm, and for  $a_{\text{ph}}(\lambda)$  in the blue and red portions of the spectrum.

### 3.4 Summary and Conclusions

This study assessed the performance and uncertainty of a 4-step Semi-Analytical Algorithm, 4SAA, using a synthetic database of optical properties. The complete 4SAA model utilizes input of  $R_{rs}(\lambda)$  to estimate values of one AOP ( $K_d(\lambda)$ ) and nine IOPs ( $a(\lambda)$ ,  $a_{nw}(\lambda)$ ,  $a_p(\lambda)$ ,  $a_{ph}(\lambda)$ ,  $a_{dg}(\lambda)$ ,  $a_d(\lambda)$ ,  $a_g(\lambda)$ ,  $b_b(\lambda)$ , and  $b_{bp}(\lambda)$ ) through a multistep inversion scheme. The 4SAA model consists of four individually developed component models, KdNN, LS2, ANW, and ADG, which have been evaluated using independent synthetic and in situ datasets. This multistep sequential structure of the entire inversion scheme is advantageous as it provides flexibility to modify its current configuration with the mechanistically-based component models as they are further enhanced or developed in the future. The 4SAA model also has significant potential for the bio-optical community, with its capability to retrieve a full suite of total and constituent IOPs at global and regional scales with hyperspectral coverage, as part of the science data products incorporated into the NASA PACE satellite mission. The developmental advantages and wide-ranging potential applicability of 4SAA demonstrates the need for a comprehensive error and uncertainty assessment of the model.

We established a methodology to conduct a comprehensive performance and uncertainty assessment of the 4SAA model at both individual and coupled component model levels using a synthetic optical database. The previously published synthetic database described in Loisel et al. (2023) contains a full suite of total and constituent IOPs as well as their corresponding AOPs obtained from forward radiative transfer simulations. This database provided reference data that are free of measurement error, which allowed us to evaluate model-generated error in a state where errors arose solely from the model itself. The individual component model evaluation was designed to enable the quantification of error produced by each step of 4SAA. Importantly, this assessment configuration isolated each component model to identify the steps responsible for the

largest sources of error. By coupling the component models, we examined the accumulation and propagation of error across sequential model steps of 4SAA, which is essential for understanding the sensitivity of each model with additional error contributed from upstream component models. We also evaluated the uncertainty and its propagation throughout the complete 4SAA model by conducting a Monte Carlo uncertainty assessment. These MC simulations provided insights into model uncertainty with anticipated input uncertainty distributions for the OCI sensor aboard NASA PACE. The final uncertainty products provided key details about the current status of predictive capabilities of the 4SAA model and guidance for potential future model improvements.

In the individual component model analysis, where model-derived estimates were compared to synthetic values of optical properties, we found a strong performance for  $K_d(\lambda)$ ,  $a(\lambda)$ , and  $b_b(\lambda)$  retrievals across the entire spectral domain of model solutions. Aggregate percent differences were generally below 15%, with the exception of  $b_b(\lambda)$  retrievals beyond 650 nm. This apparent deterioration in the performance of the LS2 model at these longer wavelengths is attributable to the inclusion of chlorophyll-a fluorescence in the synthetic database and the absence of this inelastic radiative process in the development of the LS2 model. Retrievals of  $a_{nw}(\lambda)$  were reasonably accurate up to 500 nm as aggregate percent differences were below 5% before considerably increasing at longer wavelengths. These general trends align with findings from Loisel et al. (2018) and reemphasize the challenge to accurately retrieve  $a_{nw}(\lambda)$  in the spectral region where pure water absorption has large or highly dominant contribution to total seawater absorption.

Regarding the performance of the individual absorption partitioning models, retrievals of  $a_{ph}(\lambda)$  were less reliable but the most accurate in the blue through green spectral regions where MdSA values generally remained below 40% with the performance deteriorating below 400 nm.

Retrievals of non-phytoplankton absorption and its constituent absorption coefficients were highly variable and exhibited the best performance in the near-UV through green (350–500 nm) where MdSA remained generally below 50%. However,  $a_d(\lambda)$  retrievals exceed 100% error in the blue and near-UV and highlighted a key challenge to accurately estimate this absorption constituent. These findings provided the first indication to be cautious when applying or evaluating these absorption partitioning models using spectral data of  $a_d(\lambda)$  and/or  $a_g(\lambda)$  with fixed exponential spectral shapes, as is the case with the synthetic optical database used in this study. In reality, the spectral shapes of  $a_d(\lambda)$  and  $a_g(\lambda)$  can deviate significantly from a single exponential function (see Chapter 1 and 2).

In the coupled component model analysis, we focused on the retrievals of non-water IOPs and observed a similar but slightly degraded performance in Step 2 LS2 model due to the error propagation from Step 1 KdNN model. MdAPD values increased by 3–14% and for both  $a_{nw}(\lambda)$  and  $b_{bp}(\lambda)$  across the blue and green spectral regions (440–550 nm). As error propagated into Steps 3 and 4, the ANW model rejected more potential output spectra for  $a_{ph}(\lambda)$  as a result of the realistic constraints defined within the model, and the retrievals of constituent absorption coefficients deteriorated by 10–50% with the exception of  $a_d(\lambda)$  obtained from the ADG model, which demonstrated a subtle improvement. This analysis provided further confirmation of the incompatibility of this synthetic database for the fully meaningful performance assessment of 4SAA, as the exponential spectral shape assumptions pertaining to the non-phytoplankton absorption spectra within the database can alter model performance.

Finally, the uncertainty analysis involving perturbations of model input variables via Monte Carlo computations revealed small relative uncertainties of less than 15% from 350 to 700 nm for  $K_d(\lambda)$  and  $a(\lambda)$  retrievals. Notably, we found the relative uncertainty of  $a_{nw}(\lambda)$  remained

below 15% at wavelengths shorter than 500 nm and generally below 30% at longer wavelengths. Relative uncertainties were also the smallest for  $a_{\text{ph}}(\lambda)$  in the blue and red portions of the spectrum and for  $a_{\text{dg}}(\lambda)$  and  $a_{\text{g}}(\lambda)$  in the near-UV through blue portion of the spectrum. These spectral regions are particularly important as they correspond to wavelengths where these coefficients contribute most substantially to non-water or non-phytoplankton absorption. These results indicate that model uncertainties are the lowest in the spectral regions where constituent absorption coefficients are the dominant contributor to non-water or non-phytoplankton absorption.

While this study relied on a synthetic database of optical properties to assess 4SAA, it highlighted the need for further evaluation of the model using a dataset that is not constrained by assumptions about the exponential spectral shapes of  $a_{\text{d}}(\lambda)$ ,  $a_{\text{g}}(\lambda)$  and the resulting  $a_{\text{dg}}(\lambda)$ . Ideally, such a dataset should include a full-suite of hyperspectral in situ measurements from diverse oceanic environments that are rigorously quality controlled (similar to the field datasets presented in Chapters 1 and 2) and contain optical properties analogous to those in the synthetic database used in this study. This envisioned study would utilize the new field dataset of actual high-quality measurements (with no assumptions or approximations of spectral shapes) and use our analysis presented in this chapter as a template to comprehensively assess the error and uncertainty of model predictions from 4SAA. Furthermore, the conclusions of this study highlight the potential avenues for future improvements to the LS2 model. These enhancements should focus on improvements in retrievals of  $a_{\text{nw}}(\lambda)$  through a detailed analysis of sources of error at wavelengths beyond 500 nm, where pure seawater absorption begins to contribute more substantially to total absorption of seawater. Another example of potential improvement of LS2 model is related to retrievals of  $b_{\text{b}}(\lambda)$  which can be pursued by updating the look-up tables used in the model through incorporation of the effects of chlorophyll-a fluorescence in the red spectral region.

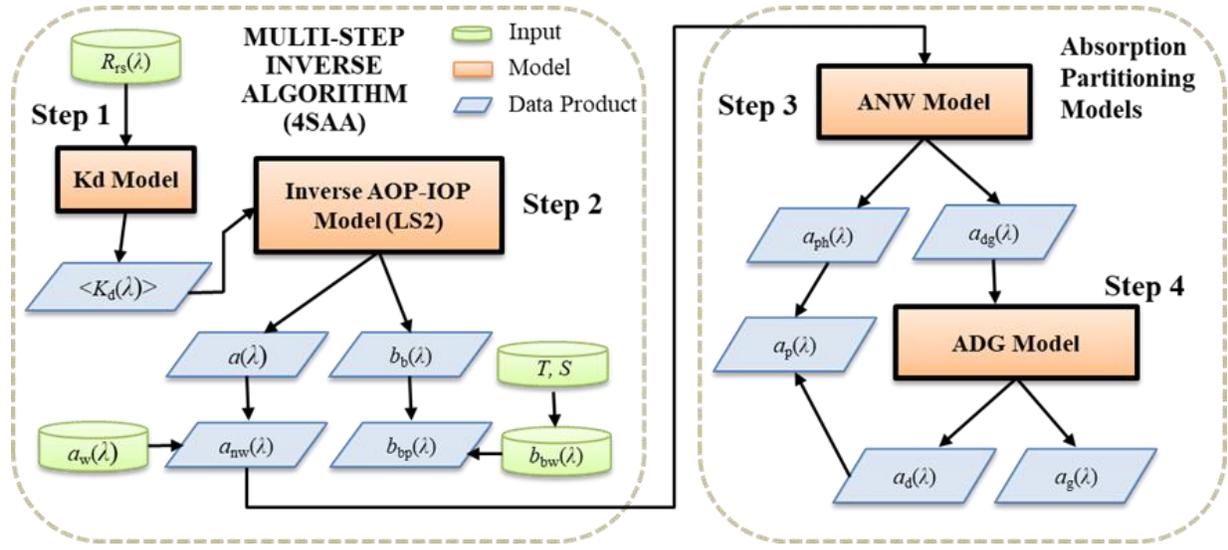
The significance of the presented statistical analysis of the 4SAA model lies in its ability to characterize and identify the sources and propagation of error and uncertainty at each step of the inversion scheme. These considerations are often overlooked or inadequately addressed in the development of satellite-derived ocean color data products, which can have significant consequences in the assessment and interpretation of satellite-derived bio-optically significant properties of the ocean. Our evaluation methodology provides a well-defined framework for assessing inversion modeling methods to better understand and characterize retrieval error and uncertainty. Additionally, we emphasize the importance of further investigation of the 4SAA model presented herein with new high-quality datasets of optical properties as they become available, especially given the already demonstrated strong performance of the individual component models of 4SAA and the unique capabilities and potential value of the complete 4SAA model for broad applications in the field of satellite ocean color observations.

### **3.5 Acknowledgements**

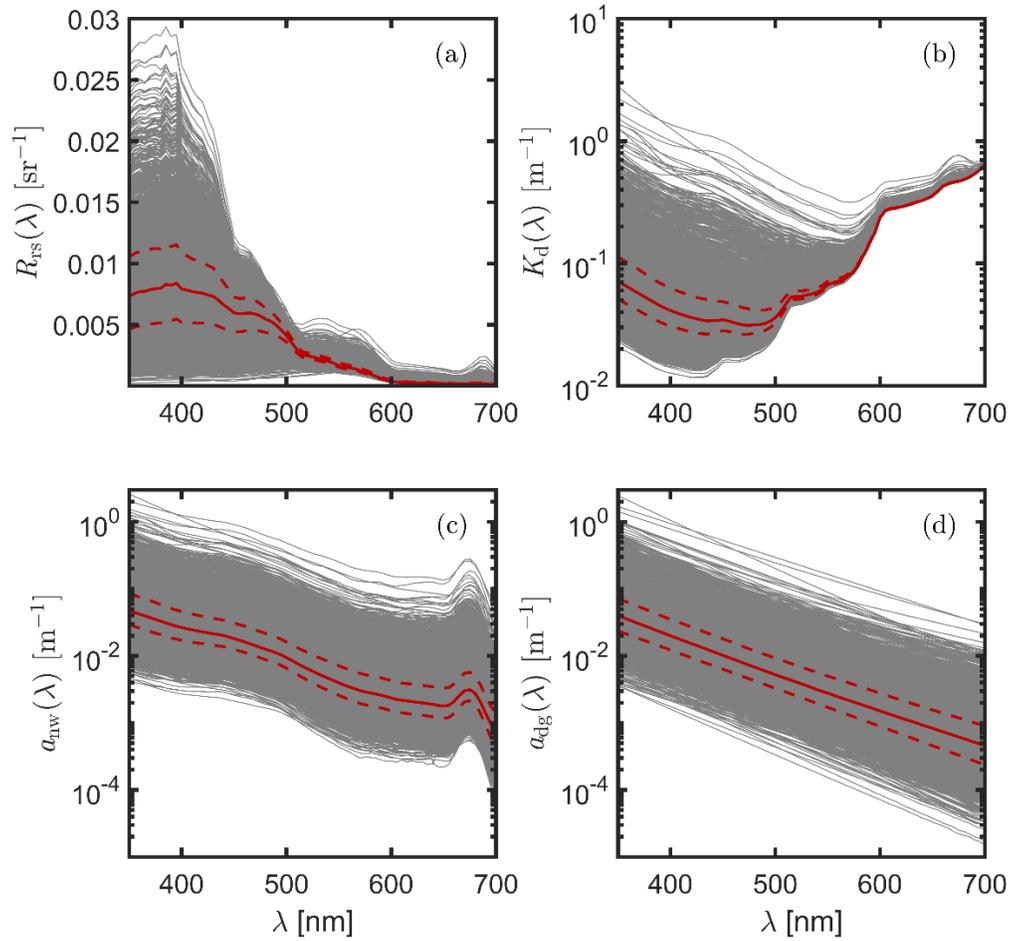
This work was supported by NASA Grant 80NSSC20M0252. I am thankful to Dariusz Stramski and Rick Reynolds for providing constructive feedback during the research process that has led to the ideas and content presented in this manuscript.

Chapter 3, in part, is currently being prepared to be submitted for publication with authors Kehrli, M. D., Stramski, D., & Reynolds, R. A. The dissertation author will be the primary investigator and author of this paper.

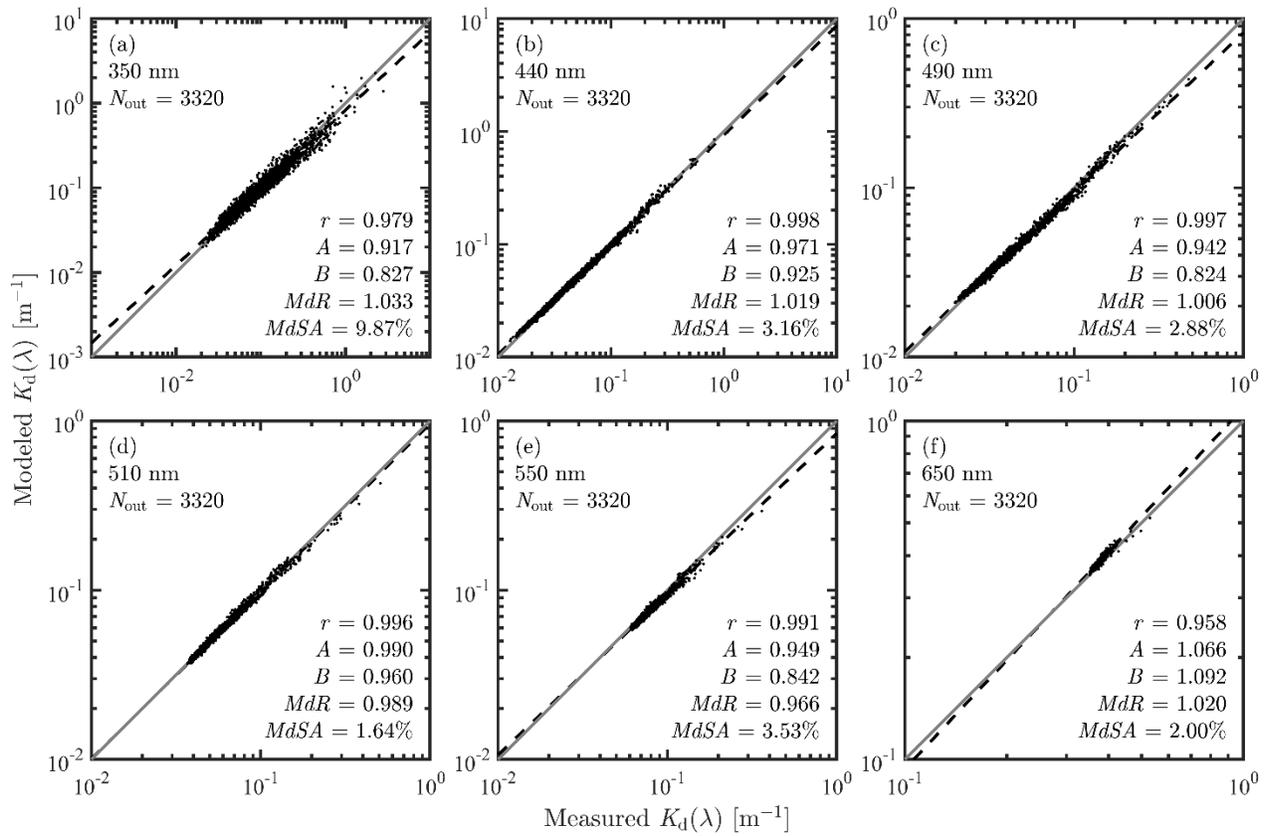
### 3.6 Figures



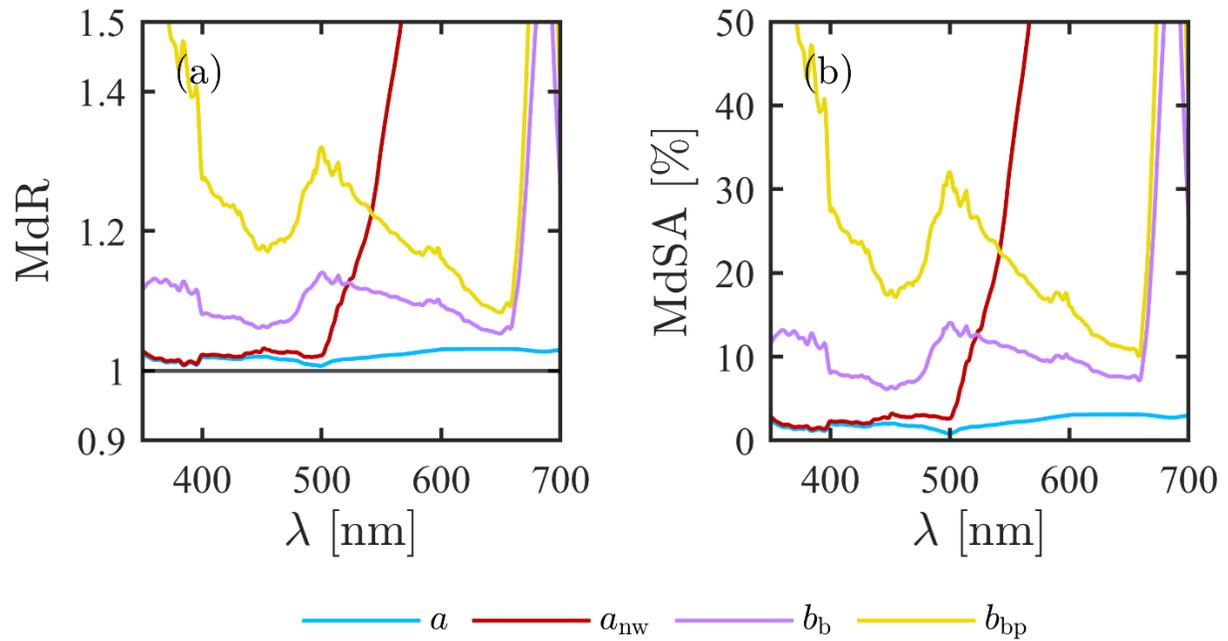
**Figure 3.1.** Flowchart of the 4-step Semi-Analytical Algorithm (4SAA).



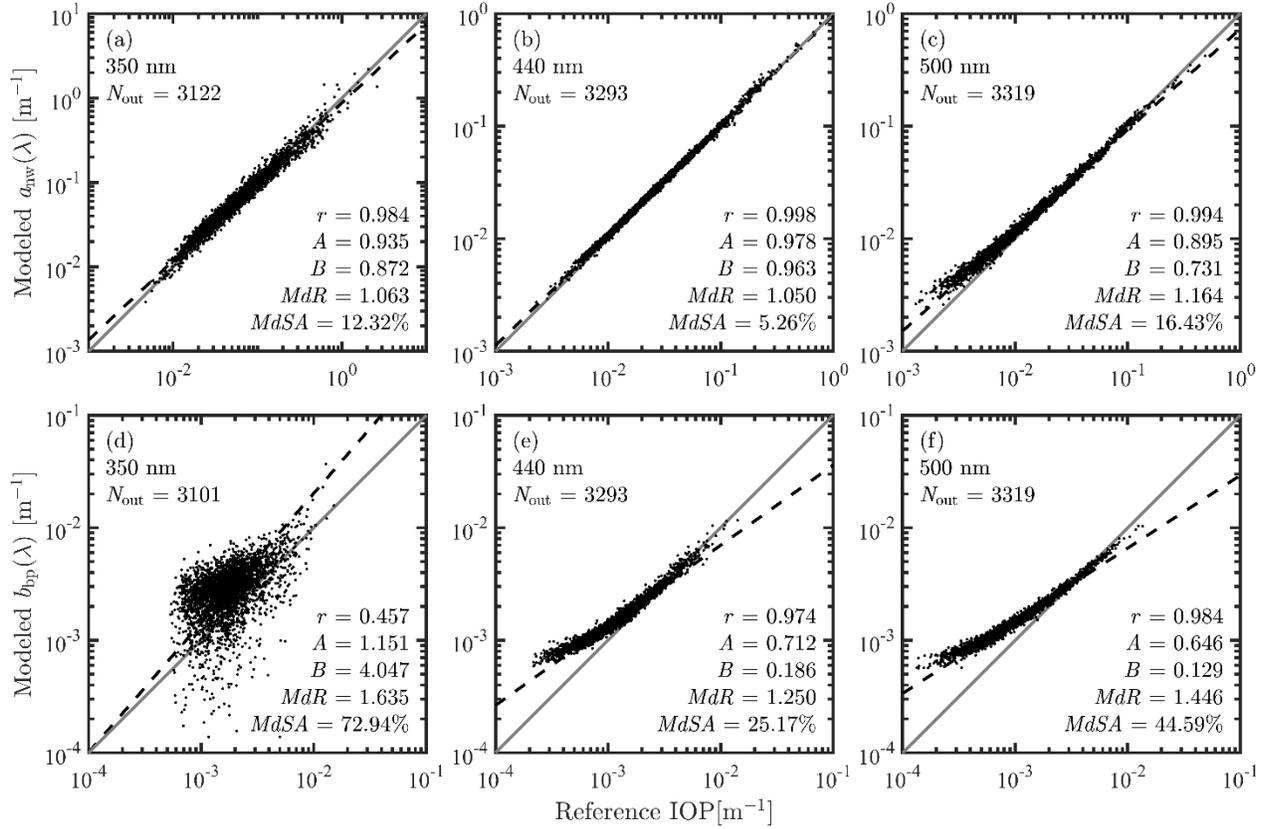
**Figure 3.2.** Synthetic database of 3320 samples of (a) remote sensing reflectance,  $R_{rs}(\lambda)$ , (b) the average diffuse attenuation coefficient for downwelling irradiance over the first attenuation depth,  $K_d(\lambda)$ , (c) the non-water absorption coefficient,  $a_{nw}(\lambda)$ , and (d) the non-phytoplankton absorption coefficient,  $a_{dg}(\lambda)$ . Solid red lines denote the median and dashed red lines depict the interquartile range for a given optical property from the database.



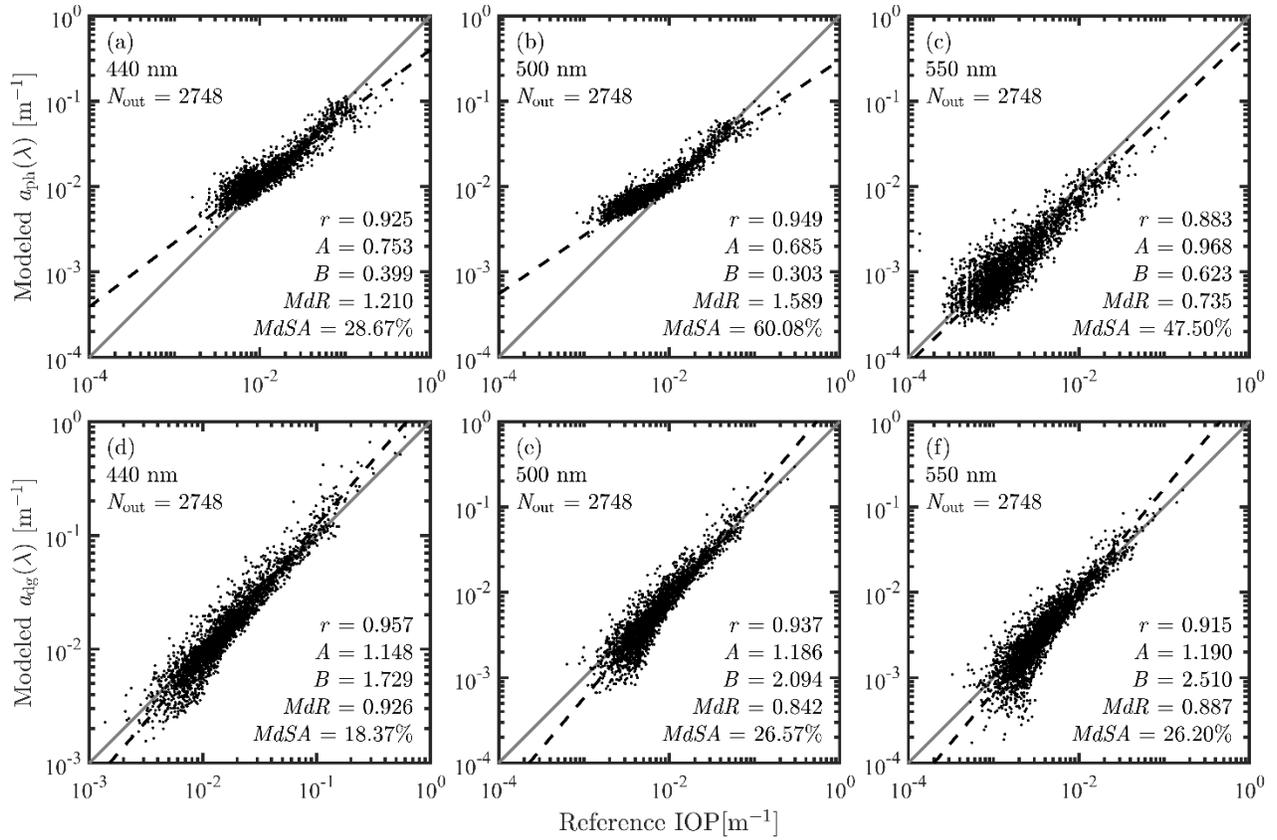
**Figure 3.3.** Assessment of Step 1, the KdNN model, applied to synthetic  $R_{rs}(\lambda)$  as scatter plots of model-predicted versus synthetic  $K_d(\lambda)$  for light wavelengths of (a) 350 nm, (b) 440 nm, (c) 490 nm, (d) 510 nm, (e) 550 nm, and (f) 650 nm. The 1:1 line and best-fit line derived from the Model II linear regression to log-transformed data are represented by the solid gray and dashed black lines, respectively.



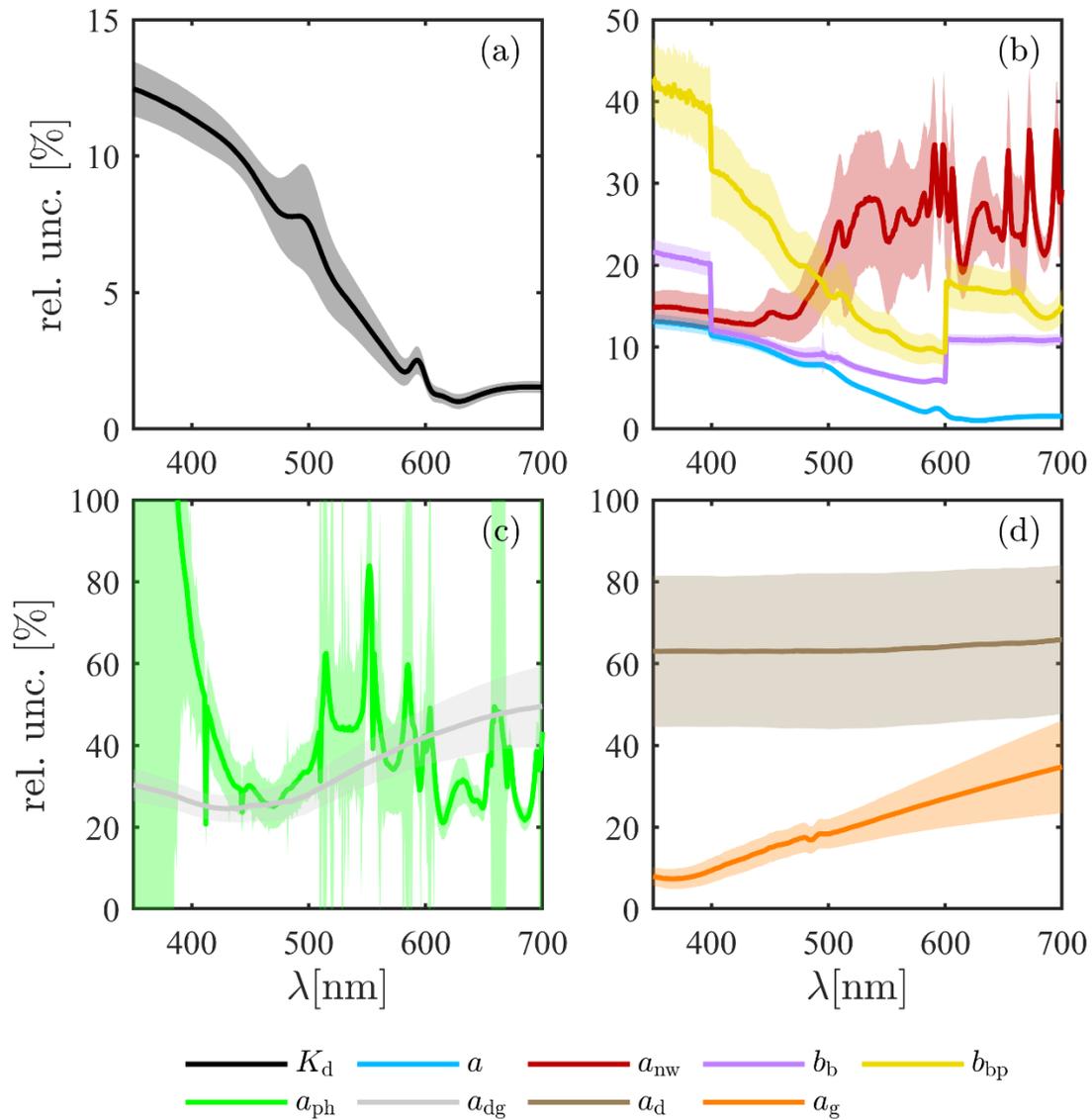
**Figure 3.4.** Spectral values of (a) median ratio, MdR, and (b) median symmetric accuracy, MdSA calculated from model-derived and synthetic reference values. The LS2 model is run with synthetic inputs of  $R_{rs}(\lambda)$  and  $K_d(\lambda)$ . Results are depicted for each model output of  $a(\lambda)$  (blue),  $a_{nw}(\lambda)$  (red),  $b_b(\lambda)$  (purple), and  $b_{bp}(\lambda)$  (yellow). The horizontal black line in panel (a) signifies a MdR value equal to 1.



**Figure 3.5.** Assessment of model-predicted non-water IOPs obtained from the LS2 model (Step 2) applied to synthetic  $R_{rs}(\lambda)$  and KdNN (Step 1) model-derived  $K_d(\lambda)$ . The assessment implements scatter plots of model-predicted versus synthetic (a–c)  $a_{nw}(\lambda)$  and (d–f)  $b_{bp}(\lambda)$  for light wavelengths of (a,d) 350 nm, (b,e) 440 nm, and (c,f) 500 nm. The 1:1 line and best-fit line derived from the Model II linear regression to log-transformed data are represented by the solid gray and dashed black lines, respectively.



**Figure 3.6.** Assessment of model-predicted constituent absorption coefficients obtained from the ANW model (Step 3) applied to coupled KdNN and LS2 (Step 1 and Step 2) derived  $a_{nw}(\lambda)$ . The assessment implements scatter plots of model-predicted versus synthetic (a–c)  $a_{ph}(\lambda)$  and (d–f)  $a_{dg}(\lambda)$  for light wavelengths of (a,d) 440 nm, (b,e) 500 nm, and (c,f) 550 nm. The 1:1: line and best-fit line derived from the Model II linear regression to log-transformed data are represented by the solid gray and dashed black lines, respectively.



**Figure 3.7.** Spectral relative uncertainty obtained from the Monte Carlo analysis of (a)  $K_d(\lambda)$  retrievals (black) from the KdNN model, (b)  $a(\lambda)$  (blue),  $a_{nw}(\lambda)$  (red),  $b_b(\lambda)$  (purple), and  $b_{bp}(\lambda)$  (yellow) retrievals from the LS2 model, (c)  $a_{ph}(\lambda)$  (green) and  $a_{dg}(\lambda)$  (gray) retrievals from the ANW model, and (d)  $a_d(\lambda)$  (brown) and  $a_g(\lambda)$  (orange) retrievals from the ADG model. Shaded regions denote the median absolute deviation around each respective relative uncertainty curve.

### 3.7 Tables

**Table 3.1.** Statistical Metrics to Assess Model Performance.

Symbol	Description
$N$	Number of samples
$O_i$ or $P_i$	Observed or model-predicted value for sample $i$ of $N$
$MdR$	Median Ratio; $MdR = \text{Median}\left(\frac{P_i}{O_i}\right)$
$MdB$	Median Bias; $MdB = \text{Median}(P_i - O_i)$
$MdAPD$	Median Absolute Percent Difference; $MdAPD = \text{Median}\left(\left \frac{P_i - O_i}{O_i}\right \right) \times 100\%$
$RMSD$	Root-mean-square deviation; $RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2}$
$MdSA$	Median Symmetric Accuracy; $MdSA = (10^{\text{Median} \log_{10}(\frac{P_i}{O_i}) } - 1) \times 100\%$
$r$	Pearson correlation coefficient of log-transformed $P_i$ vs. $O_i$
$A$	Slope of model II linear regression of log-transformed $P_i$ vs. $O_i$
$B$	Ten to the power of the y-intercept of Model II linear regression of log-transformed $P_i$ vs. $O_i$
$CV$	Coefficient of Variation: $CV = \left(\frac{\text{standard deviation}(P_i)}{\text{mean}(P_i)}\right)$
$rel\ unc.$	Relative Uncertainty: $rel\ unc. = \text{Median}(CV(P_i)) \times 100\%$

**Table 3.2** Independent performance assessment of the ANW and ADG absorption partitioning models assessed with inputs from the synthetic database. The aggregate error statistics of ANW are evaluated at light wavelengths equal to 350, 440, 500, 550, and 650 nm and the same error statistics of ADG are evaluated at 350, 440, 500, and 550 nm.

Model	Variable	$N_{out}$	$r$	$A$	$B$	$MdB [m^{-1}]$	$MdR$	$MdAPD [\%]$	$MdSA [\%]$	$RMSD [m^{-1}]$	
ANW	$a_{ph}(350)$	1271	0.5701	1.490	2.113	-0.00565	0.269	74.11	275.42	0.0279	
	$a_{ph}(440)$	3303	0.9846	1.008	0.873	-0.00132	0.845	15.84	18.74	0.0056	
	$a_{ph}(500)$	3303	0.9811	0.985	0.891	-0.00019	0.948	11.26	12.25	0.0026	
	$a_{ph}(550)$	3303	0.9035	0.950	0.671	-0.00013	0.869	29.30	37.47	0.0020	
	$a_{ph}(650)$	3303	0.9610	0.954	0.557	-0.00024	0.714	29.88	41.90	0.0017	
	$a_{dg}(350)$	3303	0.9897	0.976	1.152	0.00871	1.228	22.76	22.77	0.0319	
	$a_{dg}(440)$	3303	0.9918	0.994	1.093	0.00132	1.114	11.54	11.60	0.0056	
	$a_{dg}(500)$	3303	0.9845	0.986	0.979	0.00019	1.041	9.13	9.50	0.0026	
	$a_{dg}(550)$	3303	0.9727	0.977	0.935	0.00013	1.055	12.56	13.25	0.0020	
	$a_{dg}(650)$	3303	0.9368	0.950	0.990	0.00024	1.356	35.99	36.78	0.0017	
	ADG	$a_d(350)$	3320	0.6263	1.166	0.848	-0.00544	0.446	62.64	141.60	0.0339
		$a_d(440)$	3320	0.7219	1.185	1.403	-0.00161	0.543	54.58	105.49	0.0099
		$a_d(500)$	3320	0.7701	1.174	1.736	-0.00069	0.608	50.61	86.77	0.0045
		$a_d(550)$	3320	0.8023	1.140	1.863	-0.00029	0.703	44.52	68.56	0.0024
$a_g(350)$		3320	0.9290	0.984	1.260	0.00663	1.253	28.47	29.60	0.0367	
$a_g(440)$		3320	0.8941	0.982	1.181	0.00158	1.249	33.02	35.31	0.0099	
$a_g(500)$		3320	0.8465	0.955	0.955	0.00057	1.244	37.64	42.40	0.0044	
$a_g(550)$		3320	0.8002	0.913	0.696	0.00026	1.249	43.08	51.42	0.0024	

**Table 3.3** Performance assessment of the ADG absorption partitioning model using the UV-VIS model variant assessed with the inputs obtained from the multistep sequence of the KdNN, LS2, and ANW. The aggregate error statistics of ADG are evaluated at light wavelengths equal to 350, 440, 500, and 550 nm.

Model	Variable	$N_{out}$	$r$	$A$	$B$	$MdB [m^{-1}]$	$MdR$	$MdAPD [\%]$	$MdSA [\%]$	$RMSD [m^{-1}]$
ADG	$a_d(350)$	2748	0.7163	1.059	0.690	-0.00629	0.556	51.94	96.82	0.0322
	$a_d(440)$	2748	0.7528	1.072	1.004	-0.00169	0.679	42.68	66.82	0.0091
	$a_d(500)$	2748	0.7562	1.043	1.083	-0.00046	0.813	37.32	51.78	0.0043
	$a_d(550)$	2748	0.7478	1.003	1.096	0.00004	1.027	36.37	46.86	0.0026
	$a_g(350)$	2748	0.8664	0.910	1.245	0.01433	1.564	56.47	56.93	0.1065
	$a_g(440)$	2748	0.8439	1.013	1.288	0.00079	1.132	30.12	36.18	0.0200
	$a_g(500)$	2748	0.7952	1.073	1.447	-0.00029	0.888	38.85	52.08	0.0070
	$a_g(550)$	2748	0.7416	1.116	1.662	-0.00032	0.737	47.66	74.62	0.0033

### 3.8 References

- Aas, E. (1987). Two-stream irradiance model for deep waters. *Applied Optics*, 26(11), 2095–2101. <https://doi.org/10.1364/AO.26.002095>
- Allison, D. B., Stramski, D., & Mitchell, B. G. (2010). Empirical ocean color algorithms for estimating particulate organic carbon in the Southern Ocean. *Journal of Geophysical Research: Oceans*, 115(C10), C10044. <https://doi.org/10.1029/2009JC006040>
- Anderson, G. M. (1976). Error propagation by the Monte Carlo method in geochemical calculations. *Geochimica et Cosmochimica Acta*, 40(12), 1533–1538. [https://doi.org/10.1016/0016-7037\(76\)90092-2](https://doi.org/10.1016/0016-7037(76)90092-2)
- Antoine, D., André, J., & Morel, A. (1996). Oceanic primary production: 2. Estimation at global scale from satellite (Coastal Zone Color Scanner) chlorophyll. *Global Biogeochemical Cycles*, 10(1), 57–69. <https://doi.org/10.1029/95GB02832>
- Antoine, D., Morel, A., Gordon, H. R., Banzon, V. F., & Evans, R. H. (2005). Bridging ocean color observations of the 1980s and 2000s in search of long-term trends. *Journal of Geophysical Research. C. Oceans*, 110(C6), C06009. <https://doi.org/10.1029/2004JC002620>
- Antoine, D., D’Ortenzio, F., Hooker, S. B., Becu, G., Gentili, B., Tailliez, D., & Scott, A. J. (2008). Assessment of uncertainty in the ocean reflectance determined by three satellite ocean color sensors (MERIS, SeaWiFS and MODIS-A) at an offshore site in the Mediterranean Sea (BOUSSOLE project). *Journal of Geophysical Research*, 113(C7), C07013. <https://doi.org/10.1029/2007JC004472>
- Aurin, D., Mannino, A., & Lary, D. J. (2018). Remote Sensing of CDOM, CDOM Spectral Slope, and Dissolved Organic Carbon in the Global Ocean. *Applied Sciences*, 8(12), 2687. <https://doi.org/10.3390/app8122687>
- Behrenfeld, M. J., Boss, E., Siegel, D. A., & Shea, D. M. (2005). Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochemical Cycles*, 19(1), GB1006. <https://doi.org/10.1029/2004GB002299>
- Bellacicco, M., Vellucci, V., Scardi, M., Barbieux, M., Marullo, S., & D’Ortenzio, F. (2019). Quantifying the Impact of Linear Regression Model in Deriving Bio-Optical Relationships: The Implications on Ocean Carbon Estimations. *Sensors*, 19(13), 3032. <https://doi.org/10.3390/s19133032>
- Brewin, R. J. W., Sathyendranath, S., Müller, D., Brockmann, C., Deschamps, P.-Y., Devred, E., Doerffer, R., Fomferra, N., Franz, B., Grant, M., Groom, S., Horseman, A., Hu, C., Krasemann, H., Lee, Z., Maritorena, S., Mélin, F., Peters, M., Platt, T., ... White, G. N. (2015). The Ocean Colour Climate Change Initiative: III. A round-robin comparison on in-water bio-optical algorithms. *Remote Sensing of Environment*, 162, 271–294. <https://doi.org/10.1016/j.rse.2013.09.016>

- Cael, B. B., Bisson, K., Boss, E., Dutkiewicz, S., & Henson, S. (2023). Global climate-change trends detected in indicators of ocean ecology. *Nature*, *619*(7970), 551–554. <https://doi.org/10.1038/s41586-023-06321-z>
- Clarke, G. L., Ewing, G. C., & Lorenzen, C. J. (1970). Spectra of Backscattered Light from the Sea Obtained from Aircraft as a Measure of Chlorophyll Concentration. *Science*, *167*(3921), 1119–1121. <https://doi.org/10.1126/science.167.3921.1119>
- Devred, E., Sathyendranath, S., Stuart, V., & Platt, T. (2011). A three component classification of phytoplankton absorption spectra: Application to ocean-color data. *Remote Sensing of Environment*, *115*, 2255–2266. <https://doi.org/10.1016/j.rse.2011.04.025>
- Evers-King, H., Martinez-Vicente, V., Brewin, R. J. W., Dall’Olmo, G., Hickman, A. E., Jackson, T., Kostadinov, T. S., Krasemann, H., Loisel, H., Röttgers, R., Roy, S., Stramski, D., Thomalla, S., Platt, T., & Sathyendranath, S. (2017). Validation and Intercomparison of Ocean Color Algorithms for Estimating Particulate Organic Carbon in the Oceans. *Frontiers in Marine Science*, *4*, 251. <https://doi.org/10.3389/fmars.2017.00251>
- Feldman, G., Clark, D., & Halpern, D. (1984). Satellite Color Observations of the Phytoplankton Distribution in the Eastern Equatorial Pacific During the 1982-1983. El Nino. *Science*, *226*(4678), 1069–1071. <https://doi.org/10.1126/science.226.4678.1069>
- Gordon, H. R., & Brown, O. B. (1973). Irradiance Reflectivity of a Flat Ocean as a Function of Its Optical Properties. *Applied Optics*, *12*(7), 1549–1551. <https://doi.org/10.1364/ao.12.001549>
- Gordon, H. R. (1973). Simple Calculation of the Diffuse Reflectance of the Ocean. *Applied Optics*, *12*(12), 2803–2804. <https://doi.org/10.1364/AO.12.002803>
- Gregg, W. W., & Conkright, M. E. (2002). Decadal changes in global ocean chlorophyll. *Geophysical Research Letters*, *29*(15), 20-1. <https://doi.org/10.1029/2002GL014689>
- Gregg, W. W., Rousseaux, C. S., & Franz, B. A. (2017). Global trends in ocean phytoplankton: a new assessment using revised ocean colour data. *Remote Sensing Letters*, *8*(12), 1102–1111. <https://doi.org/10.1080/2150704X.2017.1354263>
- GUM (2008). Evaluation of measurement data – Guide to the expression of uncertainty in measurements. JCGM 100. Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures, 134pp
- Hirata, T., Aiken, J., Hardman-Mountford, N., Smyth, T. J., & Barlow, R. G. (2008). An absorption model to determine phytoplankton size classes from satellite ocean colour. *Remote Sensing of Environment*, *112*(6), 3153–3159. <https://doi.org/10.1016/j.rse.2008.03.011>
- Hu, C., Chen, Z., Clayton, T. D., Swarzenski, P., Brock, J. C., & Muller–Karger, F. E. (2004). Assessment of estuarine water-quality indicators using MODIS medium-resolution bands:

- Initial results from Tampa Bay, FL. *Remote Sensing of Environment*, 93(3), 423–441. <https://doi.org/10.1016/j.rse.2004.08.007>
- Jamet, C., Loisel, H., & Dessailly, D. (2012). Retrieval of the spectral diffuse attenuation coefficient  $K_d(\lambda)$  in open and coastal ocean waters using a neural network inversion. *Journal of Geophysical Research*, 117(C10), C10023. <https://doi.org/10.1029/2012JC008076>
- Jerlov, N. G. (1976). *Marine optics*. Elsevier Scientific Pub. Co.
- Jorge, D. S. F., Loisel, H., Jamet, C., Dessailly, D., Demaria, J., Bricaud, A., Maritorena, S., Zhang, X., Antoine, D., Kutser, T., Bélanger, S., Brando, V. O., Werdell, J., Kwiatkowska, E., Mangin, A., & d'Andon, O. F. (2021). A three-step semi analytical algorithm (3SAA) for estimating inherent optical properties over oceanic, coastal, and inland waters from remote sensing reflectance. *Remote Sensing of Environment*, 263, 112537. <https://doi.org/10.1016/j.rse.2021.112537>
- Joshi, I. D., D'Sa, E. J., Osburn, C. L., Bianchi, T. S., Ko, D. S., Oviedo-Vargas, D., Arellano, A. R., & Ward, N. D. (2017). Assessing chromophoric dissolved organic matter (CDOM) distribution, stocks, and fluxes in Apalachicola Bay using combined field, VIIRS ocean color, and model observations. *Remote Sensing of Environment*, 191, 359–372. <https://doi.org/10.1016/j.rse.2017.01.039>
- Kahru, M., & Mitchell, B. G. (2008). Ocean Color Reveals Increased Blooms in Various Parts of the World. In *EOS*, (Vol. 89, Number 18, pp. 170–170). Blackwell Publishing Ltd. <https://doi.org/10.1029/2008EO180002>
- Kehrli, M. D., Stramski, D., Reynolds, R. A., & Joshi, I. D. (2023). Estimation of chromophoric dissolved organic matter and non-algal particulate absorption coefficients of seawater in the ultraviolet by extrapolation from the visible spectral region. *Optics Express*, 31(11), 17450–17479. <https://doi.org/10.1364/OE.486354>
- Kehrli, M. D., Stramski, D., Reynolds, R. A., & Joshi, I. D. (2024). Model for partitioning the non-phytoplankton absorption coefficient of seawater in the ultraviolet and visible spectral range into the contributions of non-algal particulate and dissolved organic matter. *Applied Optics (2004)*, 63(16), 4252–4270. <https://doi.org/10.1364/AO.517706>
- Kirk, J. T. O. (1984). Dependence of relationship between inherent and apparent optical properties of water on solar altitude. *Limnology and Oceanography*, 29(2), 350–356. <https://doi.org/10.4319/lo.1984.29.2.0350>
- Kirk, J. T. O. (2011). *Light and photosynthesis in aquatic ecosystems* (3rd ed.). Cambridge University Press.
- Lee, Z., Carder, K. L., & Arnone, R. A. (2002). Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. *Applied Optics*, 41(27), 5755–5772. <https://doi.org/10.1364/ao.41.005755>

- Lee, Z., Arnone, R., Hu, C., Werdell, P. J., & Lubac, B. (2010). Uncertainties of optical parameters and their propagations in an analytical ocean color inversion algorithm. *Applied Optics*, 49(3), 369–381. <https://doi.org/10.1364/AO.49.000369>
- Li, Z., Li, L., Song, K., & Cassar, N. (2013). Estimation of phytoplankton size fractions based on spectral features of remote sensing ocean color data. *Journal of Geophysical Research: Oceans*, 118(3), 1445–1458. <https://doi.org/10.1002/jgrc.20137>
- Li, L., Stramski, D., & Reynolds, R. A. (2016). Effects of inelastic radiative processes on the determination of water-leaving spectral radiance from extrapolation of underwater near-surface measurements. *Applied Optics*, 55(25), 7050–7067. <https://doi.org/10.1364/AO.55.007050>
- Loisel, H., & Stramski, D. (2000). Estimation of the inherent optical properties of natural waters from the irradiance attenuation coefficient and reflectance in the presence of Raman scattering. *Applied Optics*, 39(18), 3001–3011. <https://doi.org/10.1364/AO.39.003001>
- Loisel, H., Stramski, D., Dessailly, D., Jamet, C., Li, L., & Reynolds, R. A. (2018). An inverse model for estimating the optical absorption and backscattering coefficients of seawater from remote-sensing reflectance over a broad range of oceanic and coastal marine environments. *Journal of Geophysical Research: Oceans*, 123(3), 2141–2171. <https://doi.org/10.1002/2017JC013632>
- Loisel, H., Jorge, D. S. F., Reynolds, R. A., & Stramski, D. (2023). A synthetic optical database generated by radiative transfer simulations in support of studies in ocean optics and optical remote sensing of the global ocean. *Earth System Science Data*, 15(8), 3711–3731. <https://doi.org/10.5194/essd-15-3711-2023>
- Mannino, A., Russ, M. E., & Hooker, S. B. (2008). Algorithm development and validation for satellite-derived distributions of DOC and CDOM in the U.S. Middle Atlantic Bight. *Journal of Geophysical Research. C. Oceans*, 113(C7), C07051. <https://doi.org/10.1029/2007JC004493>
- Maritorena, S., Siegel, D. A., & Peterson, A. R. (2002). Optimization of a semianalytical ocean color model for global-scale applications. *Applied Optics*, 41(15), 2705–2714. <https://doi.org/10.1364/ao.41.002705>
- Marshall, B. R., & Smith, R. C. (1990). Raman scattering and in-water ocean optical properties. *Applied Optics*, 29(1), 71–84. <https://doi.org/10.1364/AO.29.000071>
- Matsuoka, A., Bricaud, A., Benner, R., Para, J., Sempéré, R., Prieur, L., Bélanger, S., & Babin, M. (2012). Tracing the transport of colored dissolved organic matter in water masses of the Southern Beaufort Sea: relationship with hydrographic characteristics. *Biogeosciences*, 9(3), 925–940. <https://doi.org/10.5194/bg-9-925-2012>
- McClain, C. R., Franz, B. A., & Werdell, P. J. (2022). Genesis and Evolution of NASA’s Satellite Ocean Color Program. *Frontiers in Remote Sensing*, 3, 938006. <https://doi.org/10.3389/frsen.2022.938006>

- McKinna, L. I. W., Cetinic, I., Chase, A. P., & Werdell, P. J. (2019). Approach for Propagating Radiometric Data Uncertainties Through NASA Ocean Color Algorithms. *Frontiers in Earth Science*, 7, 440217. <https://doi.org/10.3389/feart.2019.00176>
- Mélin, F. (2019). Uncertainties in Ocean Colour Remote Sensing. *IOCCG Report*, 18, 1–164. <https://doi.org/10.25607/OBP-696>
- Mélin, F., Sclep, G., Jackson, T., & Sathyendranath, S. (2016). Uncertainty estimates of remote sensing reflectance derived from comparison of ocean color satellite data sets. *Remote Sensing of Environment*, 177, 107–124. <https://doi.org/10.1016/j.rse.2016.02.014>
- Mobley, C. (2022). The oceanic optics book.
- Morel, A., & Prieur, L. (1977). Analysis of Variations in Ocean Color. *Limnology and Oceanography*, 22(4), 709–722. <https://doi.org/10.4319/lo.1977.22.4.0709>
- Morley, S. K., Brito, T. V., & Welling, D. T. (2018). Measures of Model Performance Based on the Log Accuracy Ratio. *Space Weather*, 16(1), 69–88. <https://doi.org/10.1002/2017sw001669>
- Mouw, C. B., Hardman-Mountford, N. J., Alvain, S., Bracher, A., Brewin, R. J. W., Bricaud, A., Ciotti, A. M., Devred, E., Fujiwara, A., Hirata, T., Hirawake, T., Kostadinov, T. S., Roy, S., & Uitz, J. (2017). A Consumer’s Guide to Satellite Remote Sensing of Multiple Phytoplankton Groups in the Global Ocean. *Frontiers in Marine Science*, 4, 237601. <https://doi.org/10.3389/fmars.2017.00041>
- O’Reilly, J. E., & Werdell, P. J. (2019). Chlorophyll algorithms for ocean color sensors - OC4, OC5 & OC6. *Remote Sensing of Environment*, 229, 32–47. <https://doi.org/10.1016/j.rse.2019.04.021>
- Pinkerton, M. H., Moore, G. F., Lavender, S. J., Gall, M. P., Oubelkheir, K., Richardson, K. M., Boyd, P. W., & Aiken, J. (2006). A method for estimating inherent optical properties of New Zealand continental shelf waters from satellite ocean colour measurements. *New Zealand Journal of Marine and Freshwater Research*, 40(2), 227–247. <https://doi.org/10.1080/00288330.2006.9517417>
- Preisendorfer, R. W. (1976). *Hydrologic optics*. U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, Environmental Research Laboratories, Pacific Marine Environmental Laboratory.
- Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L., & Vanrolleghem, P. A. (2007). Uncertainty in the environmental modelling process – A framework and guidance. *Environmental Modelling & Software: With Environment Data News*, 22(11), 1543–1556. <https://doi.org/10.1016/j.envsoft.2007.02.004>
- Ricker, W. E. (1973). Linear Regressions in Fishery Research. *Canadian Journal of Fisheries and Aquatic Sciences*, 30(3), 409–434. <https://doi.org/10.1139/f73-072>

- Roesler, C. S., & Pery, M. J. (1995). In situ phytoplankton absorption, fluorescence emission, and particulate backscattering spectra determined from reflectance. *Journal of Geophysical Research*, *100*(C7), 13279–13294. <https://doi.org/10.1029/95JC00455>
- Schaeffer, B. A., Hagy, J. D., Conmy, R. N., Lehrter, J. C., & Stumpf, R. P. (2012). An Approach to Developing Numeric Water Quality Criteria for Coastal Waters Using the SeaWiFS Satellite Data Record. *Environmental Science & Technology*, *46*(2), 916–922. <https://doi.org/10.1021/es2014105>
- Seegers, B. N., Stumpf, R. P., Schaeffer, B. A., Loftin, K. A., & Werdell, P. J. (2018). Performance metrics for the assessment of satellite data products: an ocean color case study. *Optics Express*, *26*(6), 7404–7422. <https://doi.org/10.1364/oe.26.007404>
- Siegel, D. A., Behrenfeld, M. J., Maritorena, S., McClain, C. R., Antoine, D., Bailey, S. W., Bontempi, P. S., Boss, E. S., Dierssen, H. M., Doney, S. C., Eplee, R. E., Evans, R. H., Feldman, G. C., Fields, E., Franz, B. A., Kuring, N. A., Mengelt, C., Nelson, N. B., Patt, F. S., ... Yoder, J. A. (2013). Regional to global assessments of phytoplankton dynamics from the SeaWiFS mission. *Remote Sensing of Environment*, *135*, 77–91. <https://doi.org/10.1016/j.rse.2013.03.025>
- Smyth, T. J., Moore, G. F., Hirata, T., & Aiken, J. (2006). Semianalytical model for the derivation of ocean color inherent optical properties: description, implementation, and performance assessment. *Applied Optics*, *45*(31), 8116–8131. <https://doi.org/10.1364/ao.45.008116>
- Stramska, M., & Stramski, D. (2005). Variability of particulate organic carbon concentration in the north polar Atlantic based on ocean color observations with Sea-viewing Wide Field-of-view Sensor (SeaWiFS). *Journal of Geophysical Research: Oceans*, *110*(C10), C10018 <https://doi.org/10.1029/2004JC002762>
- Stramski, D., Constantin, S., & Reynolds, R. A. (2023). Adaptive optical algorithms with differentiation of water bodies based on varying composition of suspended particulate matter: A case study for estimating the particulate organic carbon concentration in the western Arctic seas. *Remote Sensing of Environment*, *286*, 113360. <https://doi.org/10.1016/j.rse.2022.113360>
- Stramski, D., Joshi, I., & Reynolds, R. A. (2022). Ocean color algorithms to estimate the concentration of particulate organic carbon in surface waters of the global ocean in support of a long-term data record from multiple satellite missions. *Remote Sensing of Environment*, *269*, 112776. <https://doi.org/10.1016/j.rse.2021.112776>
- Stramski, D., Reynolds, R. A., Kahru, M., & Mitchell, B. G. (1999). Estimation of Particulate Organic Carbon in the Ocean from Satellite Remote Sensing. *Science*, *285*(5425), 239–242. <https://doi.org/10.1126/science.285.5425.239>
- Stramski, D., Reynolds, R. A., Babin, M., Kaczmarek, S., Lewis, M. R., Röttgers, R., Sciandra, A., Stramska, M., Twardowski, M. S., Franz, B. A., & Claustre, H. (2008). Relationships between the surface concentration of particulate organic carbon and optical properties in

- the eastern South Pacific and eastern Atlantic Oceans. *Biogeosciences*, 5(1), 171–201. <https://doi.org/10.5194/bg-5-171-2008>
- Stramski, D., & Reynolds, R. A. (in preparation). *Model for partitioning the nonwater absorption coefficient of seawater in the ultraviolet and visible spectral range into the contributions of phytoplankton and non-phytoplankton components.*
- Turner, J. S., Fall, K. A., & Friedrichs, C. T. (2023). Clarifying water clarity: A call to use metrics best suited to corresponding research and management goals in aquatic ecosystems. *Limnology and Oceanography Letters*, 8(3), 388–397. <https://doi.org/10.1002/lol2.10301>
- Uitz, J., Claustre, H., Gentili, B., & Stramski, D. (2010). Phytoplankton class-specific primary production in the world's oceans: Seasonal and interannual variability from satellite observations. *Global Biogeochemical Cycles*, 24(3), GB3016. <https://doi.org/10.1029/2009GB003680>
- Vandermeulen, R. A., Mannino, A., Craig, S. E., & Werdell, P. J. (2020). 150 shades of green: Using the full spectrum of remote sensing reflectance to elucidate color shifts in the ocean. *Remote Sensing of Environment*, 247, 111900. <https://doi.org/10.1016/j.rse.2020.111900>
- VIM (2012). International vocabulary of metrology - Basic and general concepts and associated terms. JCGM 200. Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures, 108pp
- Wang, P., Boss, E., & Roesler, C. (2005). Uncertainties of inherent optical properties obtained from semianalytical inversions of ocean color. *Applied Optics*, 44(19), 4074–4085. <https://doi.org/10.1364/ao.44.004074>
- Werdell, P. J., Franz, B. A., Bailey, S. W., Feldman, G. C., Boss, E., Brando, V. E., Dowell, M., Hirata, T., Lavender, S. J., Lee, Z., Loisel, H., Maritorena, S., Mélin, F., Moore, T. S., Smyth, T. J., Antoine, D., Devred, E., d'Andon, O. H. F., & Mangin, A. (2013). Generalized ocean color inversion model for retrieving marine inherent optical properties. *Applied Optics (2004)*, 52(10), 2019–2037. <https://doi.org/10.1364/AO.52.002019>
- Werdell, P. J., Behrenfeld, M. J., Bontempi, P. S., Boss, E., Cairns, B., Davis, G. T., Franz, B. A., Gliese, U. B., Gorman, E. T., Hasekamp, O., Knobelspiesse, K. D., Mannino, A., Martins, J. V., McClain, C. R., Meister, G., & Remer, L. A. (2019). The Plankton, Aerosol, Cloud, Ocean Ecosystem Mission: Status, Science, Advances. *Bulletin of the American Meteorological Society*, 100(9), 1775–1794. <https://doi.org/10.1175/bams-d-18-0056.1>
- Westberry, T. K., Boss, E., & Lee, Z. (2013). Influence of Raman scattering on ocean color inversion models. *Applied Optics (2004)*, 52(22), 5552–5561. <https://doi.org/10.1364/AO.52.005552>
- Westberry, T. K., Silsbe, G. M., & Behrenfeld, M. J. (2023). Gross and net primary production in the global ocean: An ocean color remote sensing perspective. *Earth-Science Reviews*, 237, 104322. <https://doi.org/10.1016/j.earscirev.2023.104322>

- Woźniak, S. B., Stramski, D., Stramska, M., Reynolds, R. A., Wright, V. M., Miksic, E. Y., Cichocka, M., & Cieplak, A. M. (2010). Optical variability of seawater in relation to particle concentration, composition, and size distribution in the nearshore marine environment at Imperial Beach, California. *Journal of Geophysical Research: Oceans*, *115*(C8). <https://doi.org/10.1029/2009JC005554>
- Yoder, J. A., McClain, C. R., Feldman, G. C., & Esaias, W. E. (1993). Annual cycles of phytoplankton chlorophyll concentrations in the global ocean: a satellite view. *Global Biogeochemical Cycles*, *7*(1), 181–193. <https://doi.org/10.1029/93GB02358>
- Yoder, J. A., & Kennelly, M. A. (2003). Seasonal and ENSO variability in global ocean phytoplankton chlorophyll derived from 4 years of SeaWiFS measurements. *Global Biogeochemical Cycles*, *17*(4), 1112. <https://doi.org/10.1029/2002GB001942>
- Zheng, G., & Stramski, D. (2013). A model based on stacked-constraints approach for partitioning the light absorption coefficient of seawater into phytoplankton and non-phytoplankton components. *Journal of Geophysical Research. Oceans*, *118*(4), 2155–2174. <https://doi.org/10.1002/jgrc.20115>
- Zheng, G., & DiGiacomo, P. M. (2017). Uncertainties and applications of satellite-derived coastal water quality products. *Progress in Oceanography*, *159*, 45–72. <https://doi.org/10.1016/j.pocean.2017.08.007>