SPECTRAL PROCESSING FOR ALGAE MONITORING AND MAPPING (SPAMM):

REMOTE SENSING METHODOLOGIES FOR RIVER ECOLOGY

by

Riley Donovan Logan

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

of

Doctor of Philosophy

in

Electrical Engineering

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2024

©COPYRIGHT

by

Riley Donovan Logan

2024

All Rights Reserved

DEDICATION

To my parents, Jill and Jeff Logan, for showing me what it means to give unquestioning support, setting an incredibly high bar for what it means to be good parents, always answering the phone, and encouraging me to get the keys to open as many doors as possible.

ACKNOWLEDGEMENTS

I would like to thank everyone who has provided support, open ears, advice, enthusiasm, mentorship, and hands for scrubbing rocks throughout this project. In particular, I would like to thank my advisor, Dr. Joseph Shaw, for always finding a way to squeeze a 25th hour out of the day for me, for being an exceptional role model in research, life, and patience, and showing us all how to properly rock the long hair (though I did learn through experimentation that not *all* of us can pull off the look).

To Dr. Wataru Nakagawa for somehow managing to make teaching and research lighthearted and fun, but also rigorous and terrifying. My gratitude for the years of patience, advice, and jokes at my expense is only matched by the number of working hours you fit into a week.

To Dr. Maury Valett, Dr. Wyatt Cross, Dr. Rafa Feijó-Lima, and all of my colleagues at The University of Montana for patiently teaching me how to speak Ecology and the proper way to scrub a rock.

To Dr. Charles Kankelborg, for spreading your joy for science, always being curious, and sharing your second office at the tea house.

To all of the members of the Optical Remote Sensor Laboratory for your continual support, kind words, clever suggestions, and the many (many, many, many) hours spent helping me in the field over the years. I consider myself very fortunate to be surrounded by such kind and talented people.

To my friends and family, for acting like you aren't tired of hearing about my research

and always being there in support. I'm lucky to have you all, even if you are a bunch of liars.

The work completed and proposed within this document was part of a larger collaborative project between several Montana universities called the Consortium for Research on Environmental Water Systems (CREWS), a five-year National Science Foundation project. The CREWS project began in late 2018 and work on the UAV-based portion was completed between 2019 and 2024. This work would never have been completed without my many collaborators and mentors. Thank you all for your advice, support, brilliance, and patience throughout the last five years. Or was it six?

This material is sponsored by the National Science Foundation's Established Program to Stimulate Competitive Research (EPSCoR) Cooperative Agreement OIA-1757351. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

TABLE OF CONTENTS

1. INTRODUCTION 1

Basis of the Remote Sensing Problem 4
What's the Deal with Chlorophyll? 4
Satellite-Based Remote Sensing of Water Quality Parameters 5
Aircraft-Based Remote Sensing of Water Quality Parameters 11
Lakes and Large Inland Water Bodies. 11
Rivers and Small Inland Water Bodies. 13
UAV-Based Remote Sensing of River-Water Quality 15
Objectives 19
Develop Models for Estimating Benthic Algal Pigments 19
Test Models for Estimating Benthic Algal Pigments 20
Develop Classification Machine Learning Algorithms 20
A Combination of Algorithms 20
In Support of Developing a Low-Cost Multispectral Imager 21

2. THEORETICAL BASIS FOR ISOLATING ALGAE REFLECTANCE 23

Deep Water 23
Shallow Water 26

3. UAV-BASED HYPERSPECTRAL IMAGING FOR RIVER ALGAE PIGMENT ESTIMATION 29

Contribution of Authors and Co-Authors 29
Manuscript Information 30
Abstract 31
Introduction 31
Background 31
Existing Methods 33
Ecological Context 34
Objectives 35
Materials and Methods 36
Study Sites 36
RABs and Algal Phenology 36
Imaging System 37
Imager 37
UAV. 38
Flight Overview 38
Ground-Truth Data Collection 39

TABLE OF CONTENTS – CONTINUED

Image Preprocessing 41
Image Calibration. 41
Pixel Selection. 42
Image Processing 42
Theoretical Basis for Isolating Stream Bed Reflectance. 43
Development of Spectral Indices and Application to Algal Standing Crops. 45
Uncertainty Analysis 46
Results 47
RAB Characteristics 47
Spectral Models 49
Separation by Growth Form 50
Filamentous Plus Epiphytic. 51
Epilithic. 51
Separation by Phenology 53
Bear Gulch. 53
Gold Creek. 54
Results Summary. 55
Generalizability of Band Ratios 59
Comparison of Brute-Force Method with Existing Indices 60
Discussion 62
Conclusions 64

4. RIVER ALGAE MAPPING USING HYPERSPECTRAL IMAGERY FROM UAVS 67

Contribution of Authors and Co-Authors 67
Manuscript Information 68
Abstract 69
Introduction 69
Materials and Methods 72
Study Sites 72
Data Collection 72
Image Preprocessing 73
Classification Algorithm 75
Training. 75
Percent Cover Estimates. 77
Results and Discussion 79
Conclusion 85
Parts of Manuscript 87

TABLE OF CONTENTS – CONTINUED

Acknowledgments 87

5. THE DEVELOPMENT AND USE OF SPAMM 88

Contribution of Authors and Co-Authors 88
Manuscript Information 89
Introduction 90
Materials and Methods 91
An Overview of SPAMM 91
Study Sites 92
Ground Truth Field Sites. 93
Analysis Field Sites. 94
Ground Truth Measurements 95
Hyperspectral Image Collection 96
Imaging System. 96
UAV. 97
Flight Overview. 97
Image Preprocessing 97
Pigment Abundance Estimation 98
Overview. 98
Training and Testing. 99
Results and Discussion 100
Pigment Estimation 100
Ground Truth Data. 100
Regression and Classification Models. 100
Case Study: SPAMM on the UCFR 105
2020 Field Sites. 105
2021 Field Sites. 109
2022 Field Sites. 113
Field Site Evolution 113
Water Quality Standards. 117
Discussion 119
Conclusion 122

6. CONCLUSION 124

REFERENCES CITED 130

APPENDIX: Supplementary Figures 147

LIST OF TABLES

Table Page

1.1 Example spaceborne sensors used for water quality remote sensing, all of which have spatial resolution in relevant spectral bands that is too coarse to measure water quality in rivers and streams that are often below 30 m in width. 10

3.1 Standing crop of chlorophyll a (chl *a*), benthic organic matter (BOM), and phycocyanin at Gold Creek and Bear Gulch. Data are means ± standard error; n=6 at Gold Creek on 17 August, n=7 at Bear Gulch on 17 August, and n=20 at Gold Creek on 9 September. 48

3.2 Model performance and optimal band ratio selected for chl *a* abundance after separating data sets by growth form and phenological state. The ΔR^2 value describes the change in R^2 after separating the data set, relative to the performance of the combined data set estimating total pigment abundance. 56

3.3 Model performance and optimal band ratio selected for phycocyanin abundance after separating data sets by growth form and phenological state. The ΔR^2 value describes the change in R^2 after separating the data set, relative to the performance of the combined data set estimating total pigment abundance. 57

3.4 Optimal band ratio, mean R^2 value, and standard deviation identified after running the uncertainty analysis on total pigment abundance and after separating by growth form. 60

3.5 Comparison of selected spectral band ratios for estimating total chl *a* abundance in the UCFR from all field sites (n = 33). Listed wavelengths represent band centers for each spectral ratio while regression equations are presented as linear equations in the form of y = mx + b. 63

4.1 Optimal random forest classifier hyper parameters selected after randomized search with 5-fold cross-validation across 50 iterations. 77

LIST OF TABLES – CONTINUED

Table Page

4.2 Accuracy metrics of the random forest classification algorithms after testing on unseen validation images, reported as precision, recall, F1-Score, and the total number of samples in each class. Validation Image 1 contained no examples of regions below the visually detectable threshold for benthic growth so no metrics are included. 79

4.3 Dates, locations, and percent cover estimates of UAV flights during the 2021 data collection period. The average percent cover (bottom) indicates that non-*Cladophora* benthic growth generally dominated across the 2021 growing season, with the exception of Gold Creek, where percent cover was approximately equal. *The percent cover estimates from 17 August at Gold Creek are based on a 80-m reach at Gold Creek due to limited data collection during flight. 81

5.1 Standing crop of chlorophyll a (chl *a*) Deer Lodge, Garrison, Gold Creek, and Bear Gulch. Data are means ± standard error, after five outlying points were removed. 101

5.2 Regression results from the six selected regression algorithms, reported as R^2 and MSE after prediction of the test set. MSE is reported in transform space across all fits. 102

5.3 Regression results from the six selected regression algorithms, reported as R^2 and MSE after prediction of the test set. MSE is reported in transform space across all fits. 103

5.4 Estimates of the percent *Cladophora* and diatoms/SAV cover of the three 2020 field sites. Values are reported as the average across the full reach analyzed. 108

5.5 Estimated chl *a* abundance for the 2020 field sites. Mean chl *a* is reported as the average value across the river reach analyzed, "Low" values are abundance values between 1-50 mg/m^2, "Med" values are abundance values between 50-150 mg/m^2, and "High" values are abundance values greater than 150 mg/m^2. 109

LIST OF TABLES – CONTINUED

Table Page

5.6 Estimates of the percent *Cladophora* and diatoms/SAV cover of the three 2021 field sites. Values are reported as the average across the full reach analyzed. 111

5.7 Estimated chl *a* abundance for the 2021 field sites. Mean chl *a* is reported as the average value across the river reach analyzed, "Low" values are abundance values between 1-50 mg/m^2, "Med" values are abundance values between 50-150 mg/m^2, and "High" values are abundance values greater than 150 mg/m^2. 112

5.8 Estimates of the percent *Cladophora* and diatoms/SAV cover of the three 2022 field sites. Values are reported as the average across the full reach analyzed. 114

5.9 Estimated chl *a* abundance for the 2022 field sites. Mean chl *a* is reported as the average value across the river reach analyzed, "Low" values are abundance values between 1-50 mg/m^2, "Med" values are abundance values between 50-150 mg/m^2, and "High" values are abundance values greater than 150 mg/m^2. 115

A.1 Measurements of filamentous chl *a* and phycocyanin for data collected on 17 August 2021 at the Gold Creek and Bear Gulch field sites. 148

A.2 Measurements of filamentous chl *a* and phycocyanin for data collected on 9 September 2021 at the Gold Creek field site. 149

LIST OF FIGURES

Figure Page

1.1 The Upper Clark Fork River Basin, flowing from Butte to Missoula in southwestern Montana. 2

1.2 Example spectral absorbance of chl *a* (solid green), with decrease in absorption for lower concentrations (dashed green), and absorption features (vertical dashed lines). 6

2.1 Depiction of the radiance terms detected by an airborne imaging system, including radiance sources from sunlight such as radiance from atmospheric scattering (L_{atm}), radiance scattered from suspended particles within the water column (L_{ws}), and radiance reflected from the river bottom (L_b) as well as radiance from surface reflections at the air-water interface from diffuse sky radiance (L_{ws}). 24

3.1 The Upper Clark Fork River originates at Warms Spring and flows approximately 200 km downstream to its confluence with the Blackfoot River. Stars indicate sample sites located on the main stem at Gold Creek and Bear Gulch. 37

3.2 Sample plots and transects at the Gold Creek site on 9 September 2021. Five hoops (small white circles) were placed along four transects (red ovals, two of five shown) to establish plots moving from areas of low to high algal growth. 40

3.3 RGB image created from one of the hyperspectral image cubes gathered using the UAV system at Gold Creek. Examples of the complete spectra contained at each pixel are shown for bare river substrate, Cladophora algae, and bank vegetation for the spectral range of 400–850 nm. Note the varying reflectance axes and strong water absorption beyond approximately 725 nm that suppresses the strong near-infrared reflectance of the submerged algae. 50

3.4 Analytics from the regression analysis of total chl *a* standing crops (sum of filamentous, epiphytic, and epilithic sources) from the combined (all field sites) data set. 52

LIST OF FIGURES – CONTINUED

Figure Page

3.5 Linear regressions, correlation maps, and histograms for the Gold Creek data set. Results gathered via the brute-force analysis for estimating fila/epip chl *a* abundance (left column) and epilithic phycocyanin abundance (right column)..... 58

3.6 Estimated versus measured fila/epip chl *a* abundance at the Gold Creek field site using the regression analysis generated by the brute-force method. Points falling along the 1:1 line (dashed) represent perfect performance. 59

3.7 Histograms of R^2 values generated by running linear regressions on each iteration of pigment subsampling against the optimal band ratios identified for estimating pigment abundance for each growth form.....61

4.1 Study sites along the Upper Clark Fork River, numbered from upstream to downstream at Deer Lodge, Gold Creek, and Bear Gulch. Data collection sites were selected to capture representative algal growth along the full 120-mile reach. 73

4.2 Example spectra for the three labeled classes, including pixels containing benthic growth visually identified as *Cladophora* (upper left), non-*Cladophora* (bottom right), and below a detectable threshold (upper right). 76

4.3 Training and validation accuracy was very high (>99%) assessed across 5-fold cross-validation (a); the confusion matrix shows little confusion between classes (b); and top spectral features hovered around green wavelengths and surrounded a known chl *a* absorption line. 78

4.4 Confusion matrices for Validation Images. 80

4.5 Hyperspectral data cube before and after classification using the trained random forest model. The data cube covers approximately 80 m and shows a large bloom of *Cladophora* in the center of the river, with percent cover estimations of approximately 55% and 29% for *Cladophora* and non-*Cladophora*, respectively. 80

LIST OF FIGURES – CONTINUED

Figure Page

4.6 Decimal percent cover estimates for Deer Lodge (a), Gold Creek (b), and Bear Gulch (c) across the summer growing season of 2021. *Cladophora* shown in green and non-*Cladophora* shown in dashed blue. 83

4.7 RBG and classified image of the full river reach at Bear Gulch on 2 September 2021 (a) and connected scatter plot of percent cover as a function of downstream distance along the reach (b). 84

5.1 Overview of the SPAMM method, starting with a georectified hyperspectral imager converted to units of reflectance, producing a spatial distribution map and chl *a* abundance map, as well as coverage of areal density of chl *a*, mean chl *a* abundance, and area above Montana water quality thresholds. 93

5.2 The Upper Clark Fork River flowing from its headwaters near Butte, MT to Missoula, MT. Ground truth field sites are shown in orange and ground truth plus analysis field sites are shown in green. 94

5.3 Ground truth samples were gathered from circular plots (white circles) intentionally placed throughout each field site to capture a range of ground cover conditions and algal abundances. 95

5.4 Ground truth data measurements before (a) and after (b) applying a power transform to improve normality. 96

5.5 Stacked regression model performance of actual versus predicted values (red dots) shown with a 1:1 line (solid line), where any point falling on this line is a perfect prediction, in transform space (a) and abundance space (b). 102

5.6 Metrics from chl *a* abundance classifier, including the confusion matrix (a) and top ten wavelengths for classifying abundance (b). 104

LIST OF FIGURES – CONTINUED

Figure Page

5.7 Example outputs of SPAMM, shown at Garrison on 3 August 2020. Spectral products are created from input image, shown in RGB (a), spatial distribution maps show *Cladophora* cover (green) and diatom/SAV cover (brown) from which percent cover estimates are made (b), algal abundance maps show heat map of chl *a* values (c), and histograms show spatial coverage of chl *a* abundance, including areas in excess of Montana water quality standards (red bars)(d). 106

5.8 Large bloom of *Cladophora* observed at Deer Lodge in August of 2022, causing percent cover estimates to increase to over 80% with significant portions estimated in excess of water quality standards. 114

5.9 Deer Lodge decimal percent cover estimates measured from 0 to 1 for 11 June through 11 September of 2020 (top), 2021 (middle), and 2022 (bottom) for *Cladophora* (a) and diatoms/SAV (b). Mean value for each flight is shown with solid dot and standard deviation shown with fill. 116

5.10 Garrison decimal percent cover estimates measured from 0 to 1 for 11 June through 11 September of 2020 (top), 2021 (middle), and 2022 (bottom) for *Cladophora* (a) and diatoms/SAV (b). Mean value for each flight is shown with solid dot and standard deviation shown with fill. 117

5.11 Bear Gulch decimal percent cover estimates measured from 0 to 1 for 11 June through 11 September of 2020 (top), 2021 (middle), and 2022 (bottom) for *Cladophora* (a) and diatoms/SAV (b). Mean value for each flight is shown with solid dot and standard deviation shown with fill. 118

5.12 Estimated areal extent of *Cladophora* in excess of Montana water quality standards (150 mg/m^2) measured from 0 to 17 m^2 for 11 June through 11 September of 2020 (top), 2021 (middle), and 2022 (bottom) for Deer Lodge (a), Garrison (b), and Bear Gulch (c). Mean value for each flight is shown with solid dot and standard deviation shown with fill. 118

LIST OF FIGURES – CONTINUED

Figure Page

5.13 Gage height measurements at Garrison from 2020 (blue), 2021 (orange), and 2022 (green) showing the low water year experiences in 2021, possibly contributing to early accrual of *Cladophora* and higher levels of diatoms. 120

5.14 RGB image, spatial distribution map, and chl *a* abundance heat map (a) and spatial variation (b) for Deer Lodge on 10 September 2020....121

A.1 Analytics from the regression analysis of total phycocyanin standing crops from the combined (all field sites) data set. 150

A.2 Analytics from the regression analysis of fila/epip chl *a* standing crops from the combined (all field sites) data set. 151

A.3 Analytics from the regression analysis of epiphytic phycocyanin standing crops from the combined (all field sites) data set. 152

A.4 Analytics from the regression analysis of epilithic chl *a* standing crops from the combined (all field sites) data set. 153

A.5 Analytics from the regression analysis of epilithic phycocyanin standing crops from the combined (all field sites) data set. 154

A.6 Analytics from the regression analysis of total chl *a* standing crops from the Bear Gulch data set. 155

A.7 Analytics from the regression analysis of total phycocyanin standing crops from the Bear Gulch data set....156

A.8 Analytics from the regression analysis of fila/epip chl *a* standing crops from the Bear Gulch data set....157

A.9 Analytics from the regression analysis of epiphytic phycocyanin standing crops from the Bear Gulch data set. 158

A.10 Analytics from the regression analysis of epilithic chl *a* standing crops from the Bear Gulch data set....159

A.11 Analytics from the regression analysis of epilithic phycocyanin standing crops from the Bear Gulch data set. 160

LIST OF FIGURES – CONTINUED

Figure Page

A.12Analytics from the regression analysis of total chl *a* standing crops from the Gold Creek data set. 161

A.13Analytics from the regression analysis of total phycocyanin standing crops from the Gold Creek data set..... 162

A.14Analytics from the regression analysis of epiphytic phycocyanin standing crops from the Gold Creek data set. 163

A.15Analytics from the regression analysis of epilithic chl *a* standing crops from the Gold Creek data set..... 164

ABSTRACT

Inland water quality is a growing concern to public health, riparian ecosystems, and recreational uses of our waterways. Many modern water quality programs include measures of the presence and abundance of harmful and nuisance algae. In southwestern Montana, large blooms of the nuisance algae, *Cladophora glomerata*, have become common in the Upper Clark Fork River due to a combination of warming water temperatures, naturally high phosphorus levels, and an influx of contaminants through wastewater and anthropogenic activity along its banks. To improve understanding of bloom dynamics, such as algal biomass and percent algae cover, and their effects on water quality, a UAV-based hyperspectral imaging system was used to monitor several locations along the Upper Clark Fork River. Image data were collected across the spectral range of 400 - 1000 nm with 2.1 nm spectral resolution during field sampling campaigns across the entirety of the project, beginning in 2019 and ending in 2023. In this dissertation, methodologies for monitoring water quality were developed. These methods include estimating benthic algal pigment abundance using spectral band ratios achieving R^2 values of up to 0.62 for chlorophyll *a* and 0.96 for phycocyanin; creating spatial algae distribution maps and estimating percent algae cover using machine learning classification algorithms with accuracies greater than 99%; combining spatial algae distribution maps and improved pigment estimation using machine learning regression algorithms for creating chlorophyll *a* abundance maps, achieving an R^2 of 0.873, while also comparing abundance values to Montana water quality thresholds; and identifying salient wavelengths for monitoring and mapping algae to inform the design of a low-cost and compact multispectral imager. Throughout all field campaigns, significant spatial variations in algal growth within each river reach and frequent violations of current water quality standards were observed, demonstrating the need for high-spatial resolution monitoring techniques to be incorporated in current water quality monitoring programs.

1

INTRODUCTION

In 1983, one of the largest Superfund sites in the United States was created in southwestern Montana. The Environmental Protection Agency (EPA) designated a 777-square-kilometer site located in the Deer Lodge Valley after nearly a century of mining and smelting activity in the area, the byproducts from which drained into the Upper Clark Fork River (UCFR). The UCFR is the segment of the Clark Fork of the Columbia River that flows from its headwaters near Butte, Montana to Missoula, Montana (Figure 1.1). Though the UCFR was declared as an impaired river with the Superfund designation, the characteristically clear and shallow river had a long history of eutrophication and subsequent water quality monitoring. Between warming water temperatures, naturally high phosphorus levels, and the introduction of wastewater through anthropogenic activity along its banks, large blooms of the rapidly growing filamentous algae, *Cladophora glomerata*, have been reported since the 1970s [1, 2].

Thirty-six years after the UCFR Basin was designated as a Superfund site, the Montana University System, in partnership with Tribal College partners, was awarded a five-year, National Science Foundation project to "address water-quality issues in Montana and connect those efforts to the national research priorities." The Consortium for Research on Environmental Water Systems (CREWS) was split between five academic institutions and focused on three study areas spread across Montana, including the UCFR [3]. The CREWS project had four broad goals, two of which were focused on furthering research on ground and surface water quality and the development of field-deployable sensors and methodologies to monitor water quality. Due to its size and scope, the project was broken into Research Thrusts that contained unanswered questions with success metrics. The CREWS project

provided the motivation and starting point for the work contained in this dissertation, with particular focus paid to furthering remote sensing research on surface water quality through field-deployable sensors and the development of water quality monitoring methodologies based on algae monitoring and mapping.

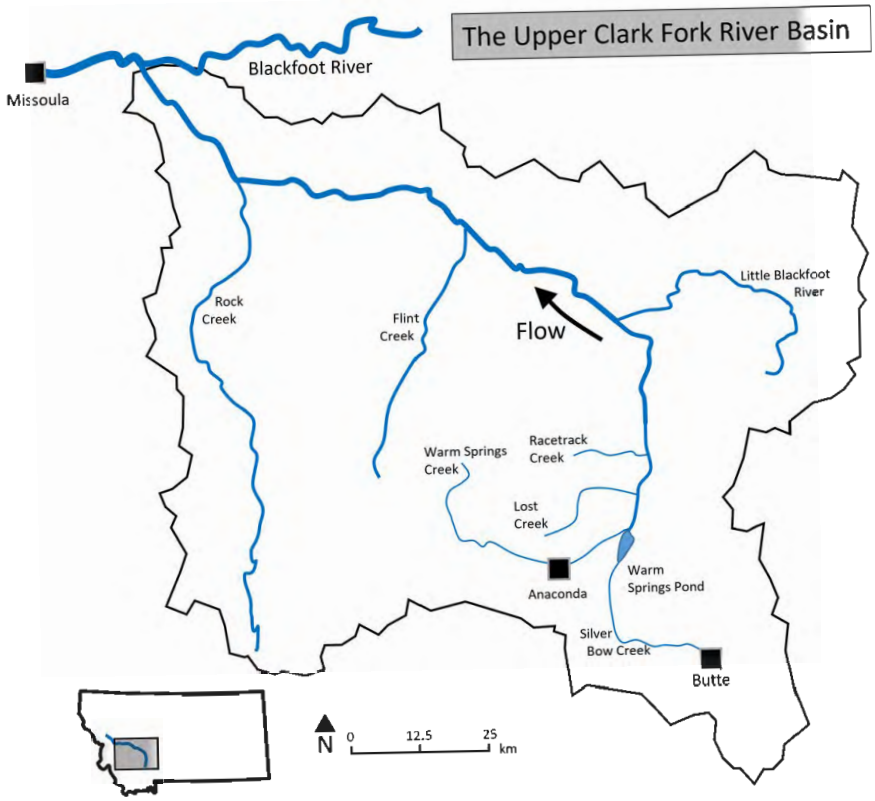


Figure 1.1: The Upper Clark Fork River Basin, flowing from Butte to Missoula in southwestern Montana (adapted from [4]).

Dense blooms of *Cladophora* commonly appear on the UCFR in the early summer and typically peak in August, though this seasonal pattern is variable [5]. As these blooms grow, they can disrupt dissolved oxygen and pH levels [6, 7], reduce benthic biodiversity [8], deteriorate water quality [6, 9, 10, 11], and decrease recreational opportunities [6, 9, 11]. To understand the ecological status, composition, and health of the water body in which algal blooms are growing, the concentrations of various algal pigments and spatial extent of the blooms are commonly measured.

Photosynthesis in algae is supported by three major classes of pigments: chlorophyll, carotenoids, and biliproteins. Among these major classes, chlorophyll *a* and phycocyanin are commonly used to assess algal bloom characteristics [12, 13]. Conventional methods to determine pigment concentration require gathering organic material *in situ*, then processing the samples in a laboratory setting in a process known as spot sampling. Though effective, this measurement method has several shortcomings, including time-intensive data collection, high costs, limited spatial sampling distribution, depth limitations (i.e. the river must be wadeable), measurement errors, and long delays between measurements and results [1, 14].

The spatial extent of algal blooms is measured in terms of percent cover, or the amount of a river's natural substrate that is covered by algal growth, ranging from 0-100%. Percent cover is often visually estimated by volunteers or field scientists moving in transects across segments of the river [15]. Though volunteers and scientists are trained to identify various types of algae, this method suffers from problems with repeatability and objectivity and relies on limited spatial measurements to estimate the total cover of a river reach.

Many of the limitations of conventional sampling methods can be overcome through the use of remote sensing systems. Remote sensing platforms, whether based on satellites, airplanes, or unoccupied aerial vehicles (UAVs), provide a synoptic view of a water body, allowing for improved spatial and temporal generality, repeatable methods and flight paths, and faster data processing times [14].

In this dissertation, methods for remotely sensing the areal density of algal pigments and the spatial distribution and extent of algal blooms are presented. All methods were developed using a UAV-based hyperspectral imaging system flown over river reaches along the length of the UCFR. These methods relate hyperspectral data to in-situ ground-truth pigment data. Additionally, techniques for classifying and mapping the spatial extent of algal blooms using hyperspectral imagery and machine leaning algorithms are presented. These two methods are then combined to create a process that identifies algal blooms, quantifies

their percent cover, and estimates the areal density of pigments.

Basis of the Remote Sensing Problem

Fifteen years prior to the creation of the Superfund site in the Deer Lodge Valley in southwestern Montana, a meeting was held in Washington, D.C. at the Federal Water Pollution Control Administration. Following this meeting, a list of twenty-two water pollution problems was created. Among these problems, eight referenced the need for better understanding of algal blooms and water quality, with five calling for the development of remote sensing techniques to address water pollution problems [16]. Remote sensing for water quality parameters was still a relatively new field in the late 1960s and early 1970s, with most missions performed by aircraft carrying imaging systems with few spectral channels [17, 18]. However, nearly all of these systems included a targeted measurement of the concentration of one pigment: chlorophyll.

What's the Deal with Chlorophyll?

One of the most common water quality metrics measured is chlorophyll *a* (chl *a*) concentration [14]. Chl *a* is a pigment essential for supporting oxygenic photosynthesis and is found in most plants, algae, and cyanobacteria. The concentration of chl *a* is indicative of the trophic state (i.e., sustainable biological productivity) of a water body, as it is directly tied to nutrient cycling and algal production; therefore, water quality is often measured in terms of volumetric chl *a* concentration (e.g., mg/L or μg/L) in oceans, lakes, and wetlands; however, in wadeable streams and rivers, chl *a* is often measured on an areal basis (e.g., mg/m^2 or μg/cm^2). Remote measurement of chl *a* concentration or areal abundance is based on related changes in the spectral absorbance of a substance that contains chl *a*.

Many bio-optical algorithms have been developed to estimate chl *a*, largely based on the prominent spectral features found in the spectral reflectance of chl *a* near 670 nm [19, 20],

a feature partially responsible for the "red edge" used to identify vegetation [21]. The example spectra in Figure 1.2 show the change in spectral reflectance as a function of chl *a* concentration (solid and dashed green lines) as well as approximate locations of spectral absorption features (vertical dashed lines) on which these algorithms are based.

Most remote sensing methods attempt to exploit the concentration-dependent change in spectral reflectance to estimate chl *a* concentration. However, to quantitatively interpret the reflectance change for suspended (or benthic) algae, the measured signal must be converted from raw digital counts to physical units, such as radiance or reflectance. This is done through radiometric calibration of the sensor system. Furthermore, the light passes through an optically lossy water path on its way to the algae and back to the sensor, which introduces a depth-dependent spectral variation that may require compensation.

Though work in the field of remote sensing of water quality has been used across the past 50 years and beyond, significant advancement began in the era of spaceborne remote sensing systems.

Satellite-Based Remote Sensing of Water Quality Parameters

The first major leap forward for remotely sensing water quality parameters occurred in the late 1970s, with the launch of a trio of satellites, including *Seasat*, *Tiros-N*, and *Nimbus-7* [22]. The Coastal Zone Color Scanner carried by the *Nimbus-7* provided image data that could be related to ocean chlorophyll concentration, with increased capability after the launch of the *NOAA-6* satellite carrying the Advanced Very-High Resolution Radiometer [23, 24]. Together, these two systems could capture simultaneous image data of the same region, allowing for more precise calculation of ocean temperature and volumetric phytoplankton chl *a* concentration [22, 25]. Though these systems provided an unprecedented tool set for scientists to study water quality in oceanic waters, they suffered from several limitations, including poor spatial resolution, significant delays due to slow data transfer rates, and data

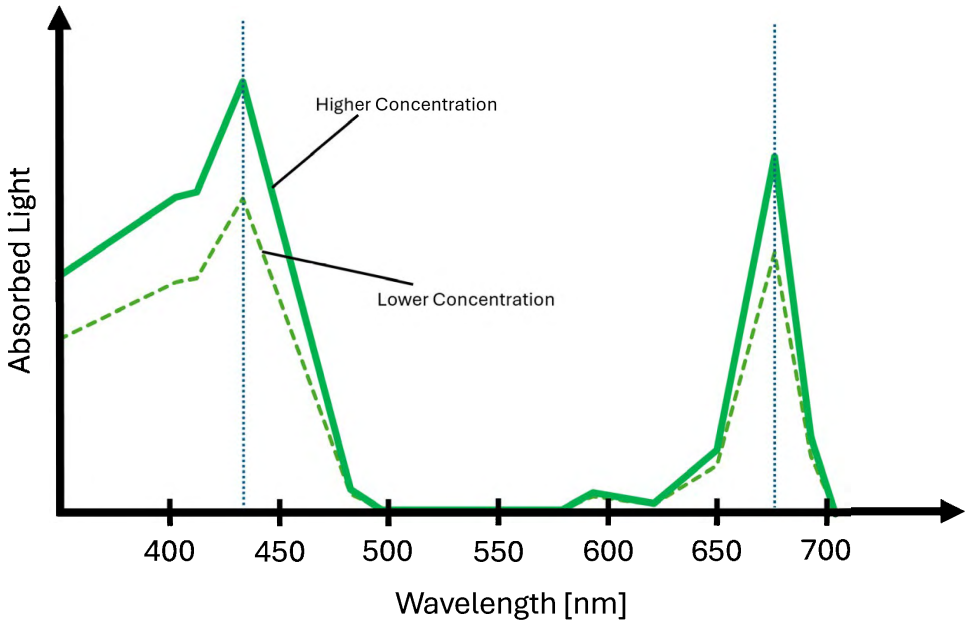


Figure 1.2: Example spectral absorbance of chl *a* (solid green), with decrease in absorption for lower concentrations (dashed green), and absorption features (vertical dashed lines).

access issues.

Since the late 1970s, sensor systems have continually improved and remote sensing of water quality parameters has been dominated by satellite-based observations of coastal waters, with the launch of several series of meteorological and Earth-observing sensors such as the Sea-Viewing Wide Field-of-View Sensor (SeaWiFS), MODerate Resolution Imaging Spectroradiometer (MODIS), and MEdium Resolution IMaging Spectrometer (MERIS).

SeaWiFS was launched aboard the SeaStar (OrbView-2) satellite with the mission of obtaining worldwide ocean color data, processing the data to produce meaningful biological parameters, and making the data available to scientists and researchers for further analysis [26]. Specifically, SeaWiFS sought to continue the data collection started by the CZCS by assessing the spatial and temporal distributions of blooms of phytoplankton, while also monitoring their primary production and volumetric chl *a* concentration; determine relationships between ocean physics and patterns of primary productivity; and further scientific applications

of ocean-color data through improvement of technical capabilities, among others. These objectives were achieved using a multispectral imager with eight spectral bands, typically 20 nm wide, spanning the spectral range of approximately 400 to 885 nm with a spatial resolution of up to 1.13 km [27].

SeaWiFs has been successful in many studies on ocean water quality, with many focusing on estimating surface chl *a* concentration to assess algal biomass and identify algal blooms [28, 29, 30]. Many of the data processing techniques used to analyze SeaWiFS data used the concept of spectral band ratios, or mathematical functions that rely on manipulating ratios of radiometrically calibrated image data. The ocean chlorophyll 2 (OC2) algorithm was chosen as the at-launch method for analyzing image data received from the SeaWiFS sensor [28]. The OC2 algorithm is based on the quotient of the remote sensing reflectance (i.e., reflectance per steradian, R_{rs}) at 490 nm and 555 nm with a cubic polynomial fit that had been tuned using an iterative approach to minimize root mean squared errors (RMSE) and maximize the R^2 values between the model and chl *a* concentration. The OC2 algorithm provided impressive performance, with an $R^2 = 0.918$ and RMSE = 0.172.

The OC2 algorithm showed excellent performance when estimating volumetric chl *a* concentration, but further improvements were made in the ocean chlorophyll 4 (OC4) algorithm. The OC4 algorithm exploited the tendency for spectral band ratios to shift toward longer wavelengths with increasing chlorophyll concentration. The algorithm had a tuned cubic polynomial fit at its core, but used three band ratios instead of the one used in OC2. The OC4 algorithm measured three band ratios, reported as wavelength in nanometers: $R_{rs}443/R_{rs}555$, $R_{rs}490/R_{rs}555$, and $R_{rs}510/R_{rs}555$, where R_{rs} is the remote sensing reflectance. The ratio that generated the highest value was then used to estimate volumetric chlorophyll concentration in what was known as a maximum band ratio (MBR) method. The OC4 method showed greater performance than that of the OC2 algorithm with $R^2 = 0.932$ and RMSE = 0.156 after tuning. Though both the OC2 and OC4 algorithms

performed well during modeling, the fit between *in-situ* and SeaWiFS data was later estimated to have $R^2 = 0.76$ and RMS log error of 31% globally, $R^2 = 0.72$ and RMS log error of 27.7% for the open ocean, and $R^2 = 0.60$ and RMS log error of 33% in coastal regions [31].

Following the launch of SeaWiFS in 1997, two additional sensors, MODIS and MERIS, were launched in 2002 aboard a series of Earth-observing satellites. Two MODIS sensors were launched by the National Aeronautics and Space Administration (NASA) aboard the satellites Terra (1999) and Aqua (2002), with each carrying a single MODIS sensor. Between the two satellites, complete imagery of the Earth could be acquired every day across 36 spectral bands. The MODIS sensor was designed as a multispectral imager, capturing spectral information in bands designated for primary uses, ranging from cloud and aerosol properties to land cover analysis. Of the 36 spectral bands on MODIS, nine channels were designated for the purpose of ocean color and water quality analysis ranging from 405 - 877 nm [32, 33]. The spectral bands designated for oceanic analysis were designed with a spatial resolution of approximately 1 km; however, several bands on MODIS had a higher spatial resolution of 250 m, which would allow for observation of smaller terrestrial features.

Data from MODIS have been used in many water quality analyses throughout its design lifetime, with many focusing on ocean-water quality assessment through volumetric chl *a* estimation [34, 35, 36]. Following the algorithmic development from SeaWiFS, data collected from MODIS were often analyzed using modified algorithms from SeaWiFS, such as the ocean color (OC) series of algorithms [28]. These modified algorithms were also based on spectral band ratios of wavelengths found to be sensitive to pigment concentration, but included modifications to correct for atmospheric effects and changes in the spectral bands of SeaWiFS and MODIS. The algorithms often relied on a log-transform of the pigment data to account for large differences in concentrations, such as the OC3 M algorithm, defined as:

$$R = log_{10}\left[\frac{max(R_rs445, R_rs489)}{R_rs555}\right], \tag{1.1}$$

where *max* denotes the greater of the two R_{rs} values [28]. The performance of the OC3 M algorithm has been compared to two other algorithms, OC5 and a local algorithm, in remote monitoring of coastal waters using MODIS data [34]. The OC5 algorithm is a modified version of the OC4 algorithm, with improvements in atmospheric corrections and adjustments for more turbid coastal waters, which has been shown to produce better results than previous OC algorithms [37]. The local algorithm was developed as a linear regression between log-transformed *in-situ* chl *a* data and *in-situ* reflectance data, constrained to a minimum and maximum reflectance to remove potential artifacts from atmospheric particles and coastal haze [34]. The resulting algorithm was adapted to MODIS data, defined as:

$$\text{Chl-a} = 10^{\left[-0.18-(5.85\times log_{10}(R_{rs}(\frac{531}{547})))-(5.56\times log_{10}(R_{rs}(\frac{678}{667})))+(0.442\times log_{10}(R_{rs}(\frac{748}{678})))\right].} \tag{1.2}$$

Comparison of the OC3 M, OC5, and local algorithm for predicting oceanic chl *a* concentrations per unit volume showed moderate performance from all three algorithms, with the local algorithm and OC5 achieving the best results. The local algorithm produced fits with an $R^2 = 0.48$ and RMSE = 96.5% while the OC5 algorithm achieved an $R^2 = 0.51$ and RMSE = 94.4%. The performance of each algorithm achieved better results than the previous state-of-the-art, OC3 M, which produced an $R^2 = 0.25$ and RMSE = 300%.

With the launch of the MODIS sensors, techniques for oceanic-water quality took a step forward; however, satellite-based inland water quality analysis was still limited by the relatively poor spatial resolution found on the spectral channels designed for water-quality analysis. Using the medium-resolution spectral bands outside those designed for analysis of water quality, chl *a* prediction has been attempted with good results [36]. These methods rely on spectral band ratios of uncorrected reflectance regressions and scattering-corrected radiance regressions fitted against *in-situ* volumetric chl *a* concentration. Though these band ratios were found to show promising results for monitoring coastal water quality, several

limitations of satellite-based systems were noted. These limitations included problems with highly turbid water, poor calibration or characterization and atmospheric correction of imagery leading to inaccurate measures of reflectance, and application limits introduced by poor spatial resolution (i.e., use for analyzing ocean areas smaller than 1 km $\times$ 1 km).

Satellites carrying sensors with higher spatial resolution, such as the Landsat Thematic Mapper (TM) with 30-m spatial resolution, have been used to estimate volumetric chl *a* concentration in large lakes [38, 39]. The models developed to monitor large lakes often rely on building regressions between *in-situ* data and spectral band ratios, and have been useful for mapping spatial distributions of phytoplankton; however, many studies have noted difficulties with the TM sensors, such as poor sensor sensitivity, poor temporal resolution, and need for data preprocessing techniques [36, 40]. Though the medium-resolution spectral channels of MODIS have been used to solve some of the limitations introduced by TM with good results for monitoring lake-chlorophyll concentration [40] and total suspended sediments [41], the limited spatial resolution of satellite-based sensors is still problematic for remotely monitoring small-to-medium inland water bodies, such as lakes and rivers where widths are often under 30 m (Table 1.1). To overcome the limitations of spaceborne platforms, remote sensors can be flown aboard aircraft.

Sensor	**Launch Year**	**Spatial Resolution (m)**
Coastal Zone Color Scanner	1978	825
Advanced Very-High Resolution Radiometer	1998	1100
SeaWiFS	1997	1100
MODIS	1999 and 2002	1000
Landsat Thematic Mapper	1984	30

Table 1.1: Example spaceborne sensors used for water quality remote sensing, all of which have spatial resolution in relevant spectral bands that is too coarse to measure water quality in rivers and streams that are often below 30 m in width.

Aircraft-Based Remote Sensing of Water Quality Parameters

Lakes and Large Inland Water Bodies. Airborne remote sensing systems provide advantages over satellite-based systems for monitoring inland water bodies, including improved configuration and mission-planning flexibility, higher spectral and spatial resolution, and more options for imaging systems, such as hyperspectral imagers [14]. Though hyperspectral imagers, such as Hyperion, have been successfully used as a part of satellite-based remote sensing systems [42], they are often limited by coarse spatial resolution, restricting their application to oceanic regions or large lakes. The spatial resolution problem in satellite-based hyperspectral imagers can be solved when using airborne systems, where sub-meter spatial resolution is frequently achieved [14]. High-spatial-resolution hyperspectral imaging systems provide advantages over traditional multispectral systems, giving the user contiguous spectral information across a broad range of wavelengths while maintaining spatial detail. Detailed spectral information allows for more careful analysis of absorption features, such as those associated with chl *a*.

Many airborne hyperspectral imaging systems have been developed for water quality analysis, with some of the most commonly used being the Airborne Imaging Spectrometer for Applications (AISA), Compact Airborne Spectrographic Imager (CASI), and Airborne Visible / Infrared Imaging Spectrometer (AVIRIS). AISA has been used in many water quality studies in lakes, commonly assessing parameters such as volumetric chl *a* concentration, turbidity, and colored dissolved organic matter (CDOM) [14, 43, 44, 45]. The AISA imaging system is based on a pushbroom-style sensor with 286 spectral channels across the spectral band of 450 - 900 nm, achieving a spatial resolution of 1 m when flown at 1000 m [45]. Similar to satellite-based remote sensing systems, analysis of airborne image data often rely on calculating band ratios to correct for atmospheric effects and measurement geometry. In one case, AISA data were used to classify lakes based on their turbidity and Secchi depth; however, chl *a* was found to be more representative of the state of the lake [45]. To estimate

the water quality parameters of the lake, several radiance band ratios were used. To estimate volumetric chl *a* concentration, a relationship between *in situ* chl *a* concentration and the radiance at 662, 700, and 781 nm was found as

$$\text{Chl-a} = a_0 + a_1 \frac{L_{700} - L_{781}}{L_{662} - L_{781}} \tag{1.3}$$

where L_λ designates the radiance at wavelength λ. The resulting fit from Equation 1.3 led to good predictive performance, with an $R^2 = 0.937$.

Though difference ratios at varying wavelengths are common in spectral analysis of chl *a* concentration, three-band models have also been used to assess lake-water quality [43]. Traditional two-band models often come in the form of

$$\text{Chl-a} \propto \frac{R_{\lambda_1}}{R_{\lambda_2}} \tag{1.4}$$

where R_λ designates the reflectance at wavelength λ [19]. Three-band models add an extra term to this relationship, with successful models taking the form

$$\text{Chl-a} \propto \frac{R_{\lambda_3}}{R_{\lambda_1} - R_{\lambda_2}}, \tag{1.5}$$

where again, R_λ designates the reflectance at wavelength λ [46]. Interestingly, both the two-band and three-band models have been shown to have similar performance when estimating volumetric chl *a* concentration, with fits from each achieving an $R^2 = 0.93$ after atmospheric correction [43]. The wavelengths found to be most sensitive to volumetric chl *a* concentration were 666, 704, and 723 nm; however, the 723 nm channel was originally modeled at 730 nm, but due to excessive instrument noise the wavelength was shortened. Though the majority of studies of water quality are focused on large lakes and oceanic waters, with the rise of airborne imaging systems, studies of smaller inland waters, such as rivers and streams have

become more common.

Rivers and Small Inland Water Bodies. Improvements in the spatial resolution of airborne hyperspectral imaging systems has expanded the applications of water quality remote sensing to smaller bodies of water, namely large-to-medium rivers. Rivers pose a greater challenge than lakes and oceans when remotely assessing water quality due to several factors, including temporal variability and heterogeneity; lack of depth, where the optical signal received by an imaging system becomes a function of substrate reflectance and water parameters; and need for higher spatial resolution to handle spatial heterogeneity [47].

Though river remote sensing poses a greater challenge than assessing large bodies of water, hyperspectral imaging systems have been successfully deployed over large rivers to monitor water quality. Assessing water quality in rivers follows a similar methodology to that applied to larger bodies of water, such as oceans and lakes. That is, image data are collected using an airborne system then related to *in-situ* measurements of water quality, such as volumetric chl *a* concentration or total suspended solids (TSS). This methodology has been applied to large riverine systems, such as the Mississippi River, several tributaries in Minnesota, and the Great Miami River in the Ohio River Basin.

To assess water quality in the Mississippi River in Minnesota, ground truth water quality data, including volumetric chl *a* concentration, TSS, and turbidity were gathered across a three-year period [47]. The ground-truth data were then related to hyperspectral image data collected by both AISA-Classic and AISA-Eagle imaging systems, with the latter having slight improvements in spectral resolution obtained between 435 - 730 nm. Linear regressions were used to relate simple reflectance ratios and ground truth data, while also experimenting with performing log-transformations on the data sets. Several relationships were developed, with different band ratios achieving optimal results for each year of data collected. The highest-performing ratios for estimating volumetric chl *a* concentration were found to be R_{705}/R_{670},

R_{705}/R_{620}, and R_{705}/R_{592}. Though all three band ratios had impressive performance, with R^2 values of 0.75-0.93, the band ratio of R_{705}/R_{670} was determined to be the most broadly applicable, as it depended on well-known chl *a* absorption features near 670 nm and 700 nm. The other two ratios were noted to likely depend on other accessory pigments, such as phycocyanin, with an absorption feature near 620 nm, and B-phycoerythrin, with an absorption feature near 590 nm.

Monitoring water quality in the Ohio River Basin relied on airborne hyperspectral data collected by the CASI instrument [48]. The CASI instrument is similar to AISA in that it is a pushbroom-style imager that collects spectral data between 400 - 1000 nm with user-programmable configurations [49]. Hyperspectral image data collected along the Ohio River Basin used 19 spectral bands with 5-nm spectral resolution that were chosen based on analysis of preliminary field spectrometer data. The selected spectral bands were then tested against ground truth data to see which band ratios had the highest correlation with the target water quality parameter. The three water quality parameters in this Ohio River Basin study were volumetric chl *a* concentration, total phosphorus, and turbidity. After relating ground-truth data to image data, high correlations were found using the relationships R_{705}/R_{675}, R_{554}/R_{675}, $R_{710} - R_{740}$ for chl *a* concentration, total phosphorus, and turbidity, respectively, with all relationships providing high R^2 values between 0.72 - 0.77. The model for predicting chl *a* concentration was found to work equally well for both rivers in the study, whereas the other two relationships required slight modification. Additionally, the best fit between reflectance data and chl *a* concentration was found after including the concentration of pheophytin, suggesting that these two biological parameters are tied together and are best estimated as a sum.

Remote sensing systems aboard aircraft solve the spatial resolution limits introduced by satellite-based systems, but still have drawbacks, including time-intensive data collection, high cost, and limited flexibility and repeatability in flight path and planning. An emerging

method of airborne imaging using UAVs addresses these drawbacks, while maintaining many of the advantages of aircraft-based remote sensing.

UAV-Based Remote Sensing of River-Water Quality

The use of UAV-based remote sensing systems is a relatively new field, but has already gathered interest due to their ability to address the problems introduced by spaceborne and airborne systems. Water quality studies based on multispectral and hyperspectral UAV-based systems have been have become popular in small reservoirs [50, 51], deep reservoirs [52], and coastal regions [53]; however, little work has been published analyzing small-to-mid-size rivers. Recently, multispectral and hyperspectral imaging systems have been used for ground-based water quality monitoring in rivers [54], but have seen limited use onboard UAVs [55, 56, 57, 58]. UAV-based hyperspectral data are often analyzed in ways similar to airborne and spaceborne systems, relying on spectral band ratios to estimate water quality parameters [51, 52, 58]; however, the use of machine learning (ML) models is becoming more widespread [54, 55, 56, 57].

With the growing popularity of ML methods, myriad models have been developed with each method optimized for a particular application and data type. Various ML algorithms have been compared to assess performance in monitoring water quality [55]. Among the algorithms compared were several regression models, including AdaBoost regression (ABR) [59], gradient boost regression tree (GBRT) [60], extreme gradient boosting regression (XGBR) [61], and CatBoost regression (CBR) [62]; as well as many traditional ML methods, such as random forest (RF) [63], extremely randomized trees (ERT) [64], support vector machine (SVM) [65], multi-layer perceptron regression (MLPR) [66], and elastic net (EN) [67]. Each method was tasked with estimating typical water quality parameters, such as volumetric chl a concentration and TSS, and assessed based on the R^2, RMSE, and mean absolute error (MAE) of the resulting fit. Data were split into training and testing sets in a typical

80/20 split (i.e. 80% of the data randomly chosen for testing, 20% selected for testing), with accuracy assessed based on performance on the test data. Of the methods tested, CBR, a customized gradient boosting method which has shown to outperform methods of similar size and complexity [62], generated the best predictions of both chl *a* and TSS, producing an R^2 = 0.96, RMSE = 0.96 mg/m^3, and MAE = 0.53 mg/m^3 when estimating chl *a* concentration.

In addition to the ML algorithms listed above, neural networks have been shown to be a promising tool for predicting water quality parameters in rivers [54]. A back propagation (BP) neural network was used to predict eight water quality parameters in the Haihe River in northern China, including volumetric chl *a* and phycocyanin concentrations. The analysis was completed using data collected by a ground-based hyperspectral imaging system mounted on the bank of the river. Prior to training the network, the data were preprocessed to remove outliers using a series of reflectance-based filters. To begin, any spectra with a reflectance above 0.2 at 420 nm was excluded to remove spectra with an anomalously high water reflectance. Next, spectra with a normalized difference vegetation index (NDVI) greater than 0.2 were removed to account for any vegetation on the surface of the water. Finally, spectra with a reflectance value higher than 0.15 at 880 nm were removed to account for pixels which likely contained objects floating on the surface of the water. Once preprocessed, the remaining 1,221 samples were used as inputs to train the BP network. The data set was split into training and validation data sets, using 70% of the data for training and 30% for validation. The model performed quite well, producing $R^2 > 0.72$ in the validation set for all water quality parameters. The model performed particularly well when estimating volumetric chl *a* concentration, producing an $R^2 = 0.88$ and RMSE = 4.77 ug/L.

Though spectral band ratios, regression techniques, and ML methods have been shown to be promising paths for estimating water quality parameters, nearly all estimates of chl *a* concentration rely on volumetric measurements (i.e., pigment mass per unit volume of water). These estimates are the variable of interest in large bodies of water, such as oceans and lakes,

where the bottom is not visible; however, for clear and shallow rivers, benthic algae become the object of interest.

When monitoring benthic algae, two parameters are often of interest: the percent algal cover and algal biomass [5, 68]. Percent algal cover is a simple measurement of the amount of stream bed which is covered by benthic algal growth, calculated as the quotient of covered to total substrate. Total algal biomass is a measure of the amount of biologically active material, often approximated using the concentration of photosynthetically active pigment, such as chl *a*. Percent cover has been approximated using UAV systems with success [69]; however, no known estimates of the areal density of chl *a* (i.e., pigment mass per unit area) have been attempted.

To monitor the chl *a* areal density of *Cladophora*, the algae that commonly reaches nuisance levels in the UCFR, the body of research on estimating chlorophyll concentration in leaves provides a starting point. Estimating leaf chl *a* concentration is a fairly mature field due to its many applications in precision agriculture and forest management, with many methods using hyperspectral imaging systems to collect image data. Numerous techniques have been used to analyze leaf chl *a* concentration, but many focus on the use of spectral band ratios and ML techniques to relate image data to *in-situ* data. A form of spectral band ratio in which a difference between two wavelengths is divided by their sum, known as a normalized difference (ND) index, was found to accurately estimate leaf chlorophyll content [70]. To find the wavelengths for the ND index, a radiative transfer model was used to determine optimal ratios, then compared to experimental data. The model determined an optimal ND index of

$$ND_{Chla} = \frac{R_{925} - R_{710}}{R_{925} + R_{710}} \tag{1.6}$$

where R_λ represents the reflectance at wavelength λ. The optimal ND index performed well,

generating a low RMSE of 8.2 ug/cm^2.

In a similar study, the efficacy of several existing band ratios used to assess vegetation health and chlorophyll concentration, such as the NDVI, modified simple ratio (MSR), modified chlorophyll absorption ratio index (MCARI,TCARI), and the integrated forms of MCARI/TCARI, as well as several custom indices, were tested using a radiative transfer model to assess their performance in estimating chlorophyll concentration of winter wheat [71]. After comparing ten vegetation indices for estimating leaf chlorophyll concentration, a quotient of two existing indices, MCARI [72] and optimal soil-adjusted vegetation index (OSAVI) [73], was found to perform best after modifying the wavelengths used. The original MCARI and OSAVI indices both used 550, 670, 700, and 800 nm; however, optimization of these indices modified the wavelengths to 550, 705, and 750 nm. The best performing index, MCARI/OSAVI was calculated as

$$\frac{MCARI}{OSAVI} = \frac{[(R_{750} - R_{705}) - 0.2 \times (R_{750} - R_{550})](R_{750}/R_{705})}{(1 + 0.16)(R_{750} - R_{705})/(R_{750} + R_{705} + 0.16)} \tag{1.7}$$

where R_λ represents the reflectance at wavelength λ. The MCARI/OSAVI index produced an excellent fit, achieving an $R^2 = 0.9406$ against chl a concentration measured across four days.

Spectral indices have been shown to be good predictors of leaf chlorophyll concentration, however, ML algorithms have also been shown to be effective when applied to the task. To estimate the chlorophyll concentration of winter wheat, a combination of a 2D enhanced correlation spectrum (2DCOS) and transfer learning was used [74]. The technique was applied to a simulated data set, from which nine wavelengths were selected, including 455, 545, 571, 615, 641, 662, 706, 728, and 756 nm. These wavelengths were then incorporated into the transfer learning scheme, from which a deep neural network was trained and tested on field measurements of leaf chlorophyll concentration. When tested against field data collected across two years, the model performed well, achieving R^2 values ranging from 0.61 to 0.85.

Though many methods exist for estimating leaf chlorophyll concentration and volumetric chl *a* concentration in water bodies, no known methods exist for combining these methods to monitor benthic algal biomass. The work presented here is aimed at filling this gap and addresses the need for improved methods for monitoring water quality in clear and shallow rivers, where benthic algae is of particular interest.

Objectives

The objectives of this work are to 1) develop models for estimating the areal density [mg/m^2] of pigments present in benthic algal growth, including spectral band ratio methods and machine learning techniques; 2) test the performance of the developed models against established methods for predicting leaf chlorophyll concentration and volumetric chlorophyll concentration; 3) develop machine learning classification methodology for producing spatial distribution maps of algal cover; 4) combine the techniques for estimating pigment areal density and spatial distribution maps into a single algorithm, capable of classifying algal pixels and estimating the pigment abundance at each location; and 5) identify salient wavelengths in support of developing a low-cost multispectral imager capable of monitoring water quality.

Develop Models for Estimating Benthic Algal Pigments

Monitoring water quality is often centered around measuring photosynthetically active pigments in benthic vegetation and the surrounding water column. Many methods have been developed for estimating chlorophyll concentration in oceans, lakes, reservoirs, and in limited cases, rivers. However, these methods measure volumetric chl *a* concentration, not the areal density associated with rooted algae coverage. Spectral indices and machine learning methods have been shown to be effective for estimating both volumetric chl *a* concentration and terrestrial leaf chlorophyll concentration as well as other accessory algal pigments, such as phycocyanin and pheophytin. Here, both spectral indices and machine learning algorithms are

used, combining concepts from volumetric and terrestrial estimates of pigment concentration to estimate benthic chl *a* areal density.

Test Models for Estimating Benthic Algal Pigments

To determine accuracy, models are compared to *in-situ* field data collected at field sites along the UCFR. Performance of each method use ground truth data to determine overall accuracy by splitting *in-situ* data into training and testing sets, where models are trained using a portion of the data and tested on the remaining unseen data.

Develop Classification Machine Learning Algorithms

Along with the demand for methods to remotely estimate benthic chl *a* areal density in clear and shallow rivers, a technique to accurately determine percent algal cover is needed. Building off the work of Flynn et al. [69], classification algorithms based on machine learning methods are developed to distinguish between *Cladophora*, diatoms/submerged aquatic vegetation (SAV), and bare substrate. The maps generated through this process are georectified and placed in a user-friendly format such that the end user can quickly and easily determine the location and extent of the spatial distribution of algae along the surveyed reach.

A Combination of Algorithms

With the completion of Objectives 1-3, the methods are combined into a single program, addressing the need for a self-contained tool for analyzing water quality from hyperspectral imagery. The program has the same capabilities listed above, but stacks output products, creating a software functional for end users without deep knowledge on the topics contained within. The software, Spectral Processing for Algae Monitoring and Mapping (SPAMM), will classify georectified hyperspectral images converted to reflectance, identify regions of

algal coverage, estimate the chl *a* areal density of the identified pixels, produce percent cover estimates, and spatial histograms of area covered by binned chl *a* abundance.

In Support of Developing a Low-Cost Multispectral Imager

Analysis and monitoring of water quality relies on information gathered by spectral imaging systems. With the increasing popularity of hyperspectral imaging systems, coupled with their decreasing cost, remote monitoring of water quality is often performed using a hyperspectral imager. Though powerful, these systems have drawbacks, including high costs, large data density, and complex image analysis. As many spectral indices have been shown to accurately estimate chl *a* abundance using limited spectral data, a low-cost multispectral system has been developed to utilize the spectral band ratios identified as effective in the UCFR. The system is still in testing, but would allow for continuous monitoring rivers, identifying algal bloom extent and estimating algal biomass. The salient wavelengths identified in this work will be used to refine the spectral channels of the existing system to improve performance.

The following chapters in this dissertation are largely based on published or submitted work addressing the objectives listed above. In Chapter 2, the theoretical background for the radiometric analysis and isolation of algae reflectance is developed. Chapter 3 is a reproduction of published work, where a method for estimating the areal density of benthic pigments using spectral band ratios along the UCFR is discussed. Chapter 4 is also a reproduction of submitted work, in which machine learning algorithms are used to produce spatial distribution maps, distinguishing between benthic growth identified as *Cladophora*, benthic growth (non-*Cladophora*), and areas below a detectable threshold of benthic growth (bare substrate) along with generating percent cover estimates. The methodology is then used to analyze three field sites along the UCFR in 2021. Finally, Chapter 5 is dedicated to the development of SPAMM, where the spatial distribution maps from Chapter 4 are

combined with machine learning regression algorithms aimed at estimating total chl *a* areal density. The method produces both algae spatial distribution maps and chl *a* areal density maps, along with associated statistics, such as percent algal cover and regions where chl *a* levels are in excess of state standards. To begin, our first step in monitoring water quality along the UCFR was to explore simple methods, such as spectral band ratios, for estimating pigment areal density.

THEORETICAL BASIS FOR ISOLATING ALGAE REFLECTANCE

The goal of many remote sensing image processing methods is to relate the measured radiance or reflectance to the pigment concentration within algae suspended in the water column (volumetric concentration) or rooted on the bottom substrate of a body of water (benthic). In an approach similar to Legleiter, et al. [75, 76], the total upwelling radiance [W/m^{2}sr] measured by a water-viewing imaging system can be written as

$$L_{tot}(\lambda) = L_b(\lambda) + L_{wc}(\lambda) + L_{ws}(\lambda) + L_{atm}(\lambda), \tag{2.1}$$

where L_{tot} is the total upwelling radiance, L_b is the radiance reflected from the benthos (i.e, bottom of the river), L_{wc} is the radiance scattered to the imager from the water column, L_{ws} is the radiance reflected from the water's surface, L_{atm} is the radiance scattered from the atmosphere and all terms are a function of wavelength, λ (omitted here on) (Figure 2.1). For deep water in lakes and oceans, the approach is generally to estimate L_{wc}, while in shallow river water it is to estimate L_b.

Deep Water

To relate image data to a change in suspended algae pigments, often the primary parameter of interest in lakes and oceans, the primary variable of interest is L_{wc}, or the optical signal scattered by the water column. To isolate this signal, several assumptions can be made. First, if measuring chlorophyll a concentration suspended in a water column, it is unlikely that the bottom substrate is visible. With no light reflected from the bed of the water body, L_b goes to zero. When imaging larger bodies of water, data are often collected from spaceborne platforms, meaning the remaining terms all play a role in the signal received by the imager. However, the signal generated from surface reflections has been shown to

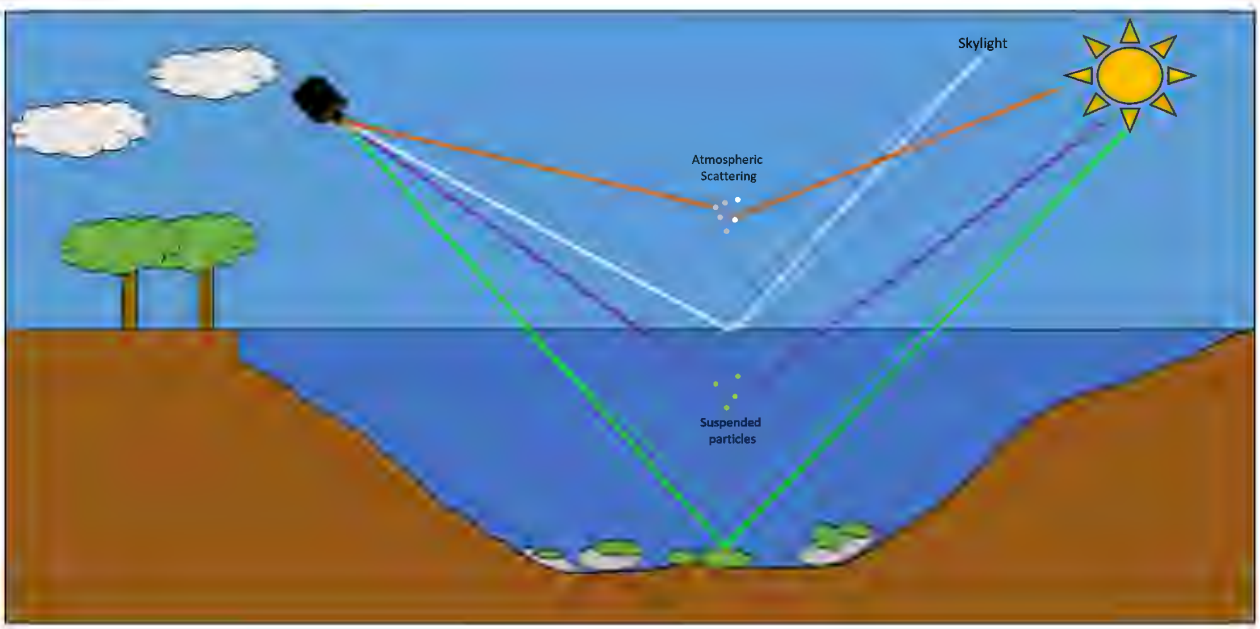


Figure 2.1: Depiction of the radiance terms detected by an airborne imaging system, including radiance sources from sunlight such as radiance from atmospheric scattering (L_{atm}), radiance scattered from suspended particles within the water column (L_{ws}), and radiance reflected from the river bottom (L_b) as well as radiance from surface reflections at the air-water interface from diffuse sky radiance (L_{ws}).

be a function of surface roughness, with reflectance values approaching 0 for calm surfaces [75]. Therefore, when imaging near nadir under calm conditions, L_{ws} may be negligible compared to the signal generated from the remaining terms. Otherwise, a wavelength- and polarization-dependent term may be added to account for transmission through the air-water interface. Finally, throughout this section we assume Lambertian reflectors and drop the notation of remote sensing reflectance (reflectance per steradian) used in the previous chapter.

With these assumptions in place, the total radiance received by the imaging system can be written in terms of the radiance generated in the water body as

$$L_{tot} = \frac{E_d T_{aw}^2 T_{atm} R_{wc}}{\pi} + T_{atm} R_{ws} L_{sky} + L_{atm}, \tag{2.2}$$

where E_d is the downwelling irradiance [W/m^2] at the surface of the water, T_{aw} is the transmittance across the air-water interface, T_{atm} is the atmospheric transmittance, R_{wc}

is the reflectance of the water column, the factor of π accounts for assumed Lambertian reflectance into a hemisphere, R_{ws} is the reflectance of the water's surface, L_{sky} is the diffuse sky radiance [W/m^{2}sr], and L_{atm} is the radiance captured by the imager through atmospheric scattering [W/m^{2}sr].

When considering the radiance observed from a body of water, the optical extinction coefficient, denoted κ in the exponential form of the Beer-Lambert Law (Equation 2.3), is often used to quantify the amount of light lost to absorption and scattering. The Beer-Lambert Law is written as

$$L_d = L_0 e^{-\kappa d}, \tag{2.3}$$

where L_d is the radiance at depth d and L_0 is the at-surface radiance. This exponential decay can be useful when performing optical measurements as a function of depth, such as observing the benthic zone (i.e., deepest ecological zone in a deep water body or river bottom). Extinction loss is not explicitly considered in Equation 2.2 due to the inclusion of R_{wc} and because the depth is assumed to be large enough that the benthic-reflected light never emerges from the water. The Beer-Lambert Law captures the amount of optical power (radiance) lost to absorption and scattering in a uniform water body as a simplified path integral to depth d. In deeper water bodies where no light reaches the bottom substrate, the amount of light reflected from the water column, R_{wc}, contains the sum of the information captured in the path integral of the Beer-Lambert Law. That is, R_{wc} represents the sum of the optical signal that is scattered back to the imaging system divided by the light incident at each depth in the path integral and inherently accounts for the signal lost to absorption or scattering outside the field of view of the imager.

The goal of imaging systems measuring volumetric chlorophyll a concentration is to isolate the signal of primary interest, R_{wc}. Rewriting Equation 2.2 in terms of reflectance

accomplished by dividing each term by the downwelling irradiance, multiplying by π steradians ($L\pi/E_d$), and solving for R_{wc} leads to

$$R_{wc} = \frac{R_{tot}}{T_{aw}^2 T_{atm}} - \frac{R_{ws} R_{sky}}{T_{aw}^2} - \frac{R_{atm}}{T_{aw}^2 T_{atm}} \tag{2.4}$$

Equation 2.4 is a complex interaction of several variables that vary with wavelength. However, many of the complications can be removed through the use of spectral band ratios, or the quotient of two or more wavelengths. In this case, many of the terms cancel, simplifying the calculation of spectral reflectance.

Shallow Water

In smaller water bodies, such as rivers and streams, in which benthic algae biomass is a water quality parameter of interest, the calculation of spectral reflectance can be further simplified. To relate image data to a change in algal pigments present in benthic algae, the signal of primary interest is the river bed radiance, L_b. Due to the nature of smaller water bodies and airborne data collection, several assumptions can be made to simplify Equation 2.1 and isolate the signal of interest.

When imaging shallow and clear waters, the majority of the downwelling solar irradiance will pass through the water column and be reflected from the stream bed, making L_b the dominant signal reaching the imager. The upwelling radiance from optical scattering within the water body prior to reaching the bottom substrate, L_{wc}, largely depends on the water's inherent optical properties (IOPs). In general, the fundamental IOPs of natural waters are the absorption coefficient and volume scattering function, which are functions of both depth and dissolved and suspended matter present in the stream [77, 78]. Due to the clarity and shallow depth common to many rivers and streams, the radiance generated through L_{wc} can be considered negligible when compared to L_b. The signal generated by L_{ws} largely

depends on the reflectance of the water's surface, which is a function of surface conditions. The reflectance of a pure-water surface is approximately 2% at normal incidence between 400 - 800 nm and are often negligible for calm surfaces compared to L_b [75]. Finally, airborne image data are commonly captured at heights of tens to hundreds of meters, meaning the optical signal takes a relatively short path through the atmosphere, making the effects of atmospheric scattering and absorption negligible compared to L_b under clear-sky conditions.

Taking these assumptions and simplifications into account, the total upwelling radiance reaching the airborne imaging system can be considered to be dominated by the benthic upwelling radiance, L_b. Equation 2.1 can be rewritten only in terms of L_b as

$$L_{tot} \approx L_b = \frac{E_d T_{aw}^2 R_b e^{-2kd}}{\pi}, \tag{2.5}$$

where E_d is the downwelling irradiance [W/m^2], π accounts for an assumed Lambertian reflectance, T_{aw} is the air-water interface transmittance, R_b is the reflectance of the stream bed, k is the optical extinction coefficient [1/m], and d is the depth of the water body [m].

As the primary signal of interest for identifying algal pigment abundance is the stream bed reflectance, Equation 2.5 can be rewritten as

$$R_b = \frac{R_{tot}}{T_{aw}^2 e^{-2kd}}, \tag{2.6}$$

by dividing all terms by the downwelling irradiance and solving for R_b. R_{tot} is the reflectance measured at the imager, or $L_{tot}\pi/E_d$, with the assumption that the imaged surface is Lambertian. Equation 2.6 provides a basis for isolating the reflectance of the river bed to accurately assess changes in pigment abundance.

A water-viewing imaging system measures a signal that is a combination of atmospheric scattering, water surface reflection, water column scattering, and river bottom reflection terms. The signal can be simplified by considering deep-water or shallow-water limits where

certain terms can be neglected. When analyzing organic matter in a water column (e.g., volumetric chl *a*) the deep-water limit is used and the bottom reflection term can be ignored. In the shallow-water limit, the measured signal is dominated by bottom reflections and many of the other terms can be considered negligible. Depth correction using the exponential decay and extinction coefficient characteristic of water can be helpful, but requires information that is not generally collected without a water-penetrating lidar or another method of determining depth. In the following chapter, the shallow-water limit outlined here is used to develop a method for estimating benthic algal pigments using spectral band ratios.

29

UAV-BASED HYPERSPECTRAL IMAGING FOR RIVER ALGAE PIGMENT ESTIMATION

Contribution of Authors and Co-Authors

Manuscript in following chapter

Author: Riley D. Logan

Contributions: Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing—original draft

Author: Madison A. Torrey

Contributions: Data curation

Author: Rafael Feijó-Lima

Contributions: Data curation, Formal analysis, Methodology, Writing—review and editing

Author: Benjamin P. Colman

Contributions: Formal analysis, Resources, Supervision, Validation, Writing—review and editing

Author: H. Maurice Valett

Contributions: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing—review and editing

Author: Joseph A. Shaw

Contributions: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing—review and editing

Manuscript Information

Riley D. Logan, Madison A. Torrey, Rafael Feijó-Lima, Benjamin P. Colman, H. Maurice Valett, Joseph A. Shaw

Remote Sensing

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
_X_ Published in a peer-reviewed journal

MDPI
Submitted: 2 May 2023
Published: 16 June 2023
Volume 15, Issue 12
https://doi.org/10.3390/rs15123148

Abstract

Harmful and nuisance algal blooms are becoming a greater concern to public health, riverine ecosystems, and recreational uses of inland waterways. Algal bloom proliferation has increased in the Upper Clark Fork River due to a combination of warming water temperatures, naturally high phosphorus levels, and an influx of nitrogen from various sources. To improve understanding of bloom dynamics and how they affect water quality, often measured as algal biomass measured through pigment standing crops, a UAV-based hyperspectral imaging system was deployed to monitor several locations along the Upper Clark Fork River in western Montana. Image data were collected across the spectral range of 400–1000 nm with 2.1 nm spectral resolution during two field sampling campaigns in 2021. Included are methods to estimate chl *a* and phycocyanin standing crops using regression analysis of salient wavelength bands, before and after separating the pigments according to their growth form. Estimates of chl *a* and phycocyanin standing crops generated through a linear regression analysis are compared to in situ data, resulting in a maximum R^2 of 0.96 for estimating filamentous plus epiphytic chl-a and 0.94 when estimating epiphytic phycocyanin. Estimates of pigment standing crops from total abundance, epiphytic, and the sum of filamentous and epiphytic sources are also included, resulting in a promising method for remotely estimating algal standing crops. This method addresses the shortcomings of current monitoring techniques, which are limited in spatial and temporal scale, by proposing a method for rapid collection of high-spatial-resolution pigment abundance estimates.

Introduction

Background

In 1983, the largest collection of US Environmental Protection Agency Superfund sites was established to include the Upper Clark Fork River (UCFR), a snowmelt-fed river located

in western Montana, USA. The river's headwaters were the locus of over 100 years of mining and smelting activity, the byproducts of which were repeatedly deposited along the river and its adjacent floodplain via recurrent flooding. Along with heavy metal contamination, the UCFR suffers from anthropogenic nitrogen enrichment that, coupled with naturally high levels of phosphorous availability and insolation, leads to annual nuisance blooms of filamentous algae (*Cladophora glomerata*) during periods of enhanced net primary productivity [1, 79, 80]. Furthermore, as the growing season progresses, *Cladophora* is replaced by bluegreen algal growth (Valett et al., in review). As these dense mats of algae grow, they can disrupt dissolved oxygen and pH [6, 7], reduce benthic biodiversity [8], deteriorate water quality [6, 9, 10, 11], and decrease recreational opportunities [6, 9, 11]. To understand the ecological status, composition, and health of water bodies (i.e., water quality) in which riverine algal blooms (RABs) occur, the standing crops of various algal pigments are commonly measured, rendering information on the extent and duration of the bloom. Monitoring RAB development is particularly important in Montana, which was one of the first states to adopt water quality standards based in part on numeric algal biomass, measured as standing crops of benthic algal chl *a* [81].

Photosynthesis in algae is supported by three major classes of pigments: chlorophyll, carotenoids, and biliproteins. Among the major classes of pigments, chlorophyll *a* (chl *a*) is commonly used to assess bloom characteristics for an array of algal types, while phycocyanin is used for cyanobacteria [12, 13]. Methods for measurement generally require gathering organic material in situ and processing in the laboratory. While biomass abundance can be determined by burning the sample to obtain measures of ash-free dry mass (AFDM), pigment content is determined by extraction using various solvents [82]. Current methods of assessing algal abundance via pigment and organic matter standing crops include several shortcomings: time-intensive data collection, limited spatial distribution, and often long delays between sample collection and results. In response, remote sensing has been introduced as a promising method to overcome many of the shortcomings of these traditional techniques [83].

Existing Methods

Remotely sensed imagery, from either satellite or airborne platforms, allow for rapid data collection over large areas with relatively short data processing times [14]. Satellite remote sensing has been used extensively to monitor water quality and assess algal bloom dynamics in oceans [84, 85, 86, 87, 88, 89, 90] and to study algal pigments in large lakes [89, 90, 91, 92]. However, satellite sensing of rivers and small lakes is prevented or impaired by disadvantages that include coarse spatial resolution, long revisit times, and obscuration by clouds. Unoccupied aerial vehicles (UAVs) have provided a means of capturing high-spatial-resolution image data while also maintaining precise control over the sampling location and time, with improved repeatability and data collection times [93].

UAV-based remote sensing systems have gained popularity in recent years for monitoring algal pigment concentrations in oceans [94], lakes and reservoirs [52, 95, 96], and rivers [58, 69, 97]. In smaller running water systems, UAV-based multispectral imagers have been used to estimate pigment concentrations using spectral indices such as the normalized difference vegetation index (NDVI), which was shown to be strongly correlated with chl *a* concentration [58]. Remote sensing systems have also been successfully deployed along larger water systems, such as the UCFR, where a small UAV-based color imager with red, green, and blue (RGB) channels was used to assess percent algal cover [69]. UAV-based multispectral imagers also have been used with machine learning techniques to measure chl *a* concentration and total suspended solids in reservoirs and artificial lakes [97].

Spectral indices are commonly used to estimate biological markers in vegetation, ranging from assessing forest health and stress [98] to estimating leaf chlorophyll [99]. These indices vary from simple ratios of two spectral bands to more complex formulations involving several wavelengths. Though spectral indices have been developed for estimating chl *a* concentration in coastal waters [88], inland lakes [100], and streams [58], they have been developed to estimate concentrations within a water body (i.e., per unit volume of water) rather than

per unit area. In addition, several established spectral indices have been shown to have decreased performance when generalized across systems with differing water properties [101]. Furthermore, no known studies have assessed water quality through estimation of algal standing crops (i.e., the mass of algal pigment in a given area) using UAV-based hyperspectral imaging systems over shallow rivers.

Ecological Context

While many rivers and streams are heavily canopied and their food webs are subsidized by inputs of terrestrial organic matter [102, 103, 104, 105, 106, 107], others, such as streams in semi-arid systems [102] or grasslands [108], and mid-order rivers with open, well-lit channels [109], rely on autochthonous organic matter production and processing [110, 111, 112]. In these types of streams, autotrophic organisms typically colonize and grow on the stream bottom (i.e., benthos) and proliferate when warm temperatures, ample insolation, and abundant nutrients support growth. The proliferation of benthic algae influences rates of primary production and respiration, trophic relationships among consumers such as insects and fish, and alters water quality. As systems dominated by the flow of water, measures of ecosystem metabolism [113, 114] and nutrient cycling in streams and rivers reflect transport and processing that typically occur over kilometers of river distance upstream from points of assessment. Accordingly, the ability to assess the spatial and temporal distributions of benthic algae is critical for understanding in-stream influences over riverine process. Current methods for quantifying algal standing crops are constrained to scales smaller (cm^2–m^2) than are needed to couple structure to functional attributes like metabolism or nutrient uptake that extend over larger scales (i.e., 100–1000 m) reflecting the influences of flow [115]. Such constraint makes typical measures susceptible to site-specific conditions (i.e., local variation in growth conditions), rendering extrapolation to larger scales inappropriate.

Objectives

Here, this methodological knowledge gap is addressed by presenting an approach for quantifying algal abundance from image data collected by a UAV-based hyperspectral imaging system to determine standing crops of chl *a* and phycocyanin. Data were collected from the UCFR during the annual algal bloom dominated by the nuisance filamentous green algae, *Cladophora*, but include colonists that live on algal filaments (i.e., epiphytic forms) and more adnate forms associated with rock surfaces (i.e., epilithic algae), and bluegreen cyanobacteria that contribute to periphyton composition as the bloom progresses (Valett et al., in revision). Concurrently, ground-truth data were collected from designated plots to superimpose in situ sampling locations and corresponding image pixels. Using these data sets, the following research objectives were pursued:

1. present a framework for collecting high-spatial and -spectral image data using a UAV-based hyperspectral imaging system along a clear and shallow river;

2. determine optimal linear relationships between spectral band ratios (i.e., the ratio of reflectance values at two wavelengths) and the algal pigments chl *a* and phycocyanin present among filamentous, epiphytic, and epilithic forms of benthic algae;

3. compare the performance of the derived optimal linear relationships to select existing spectral indices developed for monitoring chl *a*.

This paper expands on a conference paper that provided a preliminary summary of this work [4], by including expanded discussions of the ecology of the Upper Clark Fork River; algae phenology; ground-truth data collection; imager calibration; the theoretical basis for the employed method; algorithmic development; results of ground-truth data collection; expanded modeling by separating by algal growth form, including another sampling location, and estimating phycocyanin abundance; an uncertainty analysis to aid in assessments of generalizability; and performance comparisons with several established spectral band ratios.

Materials and Methods

Study Sites

The UCFR is a clear and shallow cobble-bed river that forms at the confluence of Warm Springs Creek and Silver Bow Creek, with a long history of eutrophication and heavy algal growth [1, 79, 80]. Data were collected at two sites on the main stem of the river (Figure 3.1). The field sites, Gold Creek (46.59°N, 112.93°W) and Bear Gulch (46.70°N, 113.43°W), feature channels of comparable width (30 and 28 m) and depth (35 cm) during the time of field sampling and both have been shown to contain elevated nutrients, likely from wastewater and other anthropogenic activities in the area [116]. Coupled to little shading from the surrounding riparian canopy, such enrichment creates conditions conducive to RAB proliferation [79].

RABs and Algal Phenology

Growth of *Cladophora* in the UCFR is controlled by many factors, but is typically initiated after spring runoff events in May or June [117]. Peak runoff scours benthic organic matter through shear forces created by elevated flow. Algal biomass then accrues as flow declines (Valett et al., in press) [118], temperatures increase, and *Cladophora* standing crops become visually evident [68]. The UCFR typically approaches base flow by early July, after which *Cladophora* blooms reach peak biomass [118]. As blooms progress, diatoms begin to accumulate on *Cladophora* filaments, as seen in other river systems experiencing RABs dominated by *Cladophora* [119, 120]. Self-shading, grazing pressures, and thermal stress can lead to rapid extirpation and *Cladophora* decline. During this process, *Cladophora* blooms experience a shift in their surface appearance, transitioning from light green (early phases) to yellow (later stages with epiphytic coloration) to a darker brown (heavily colonized or senescent) [121]. As *Cladophora* blooms senesce in late summer and early fall, they give way

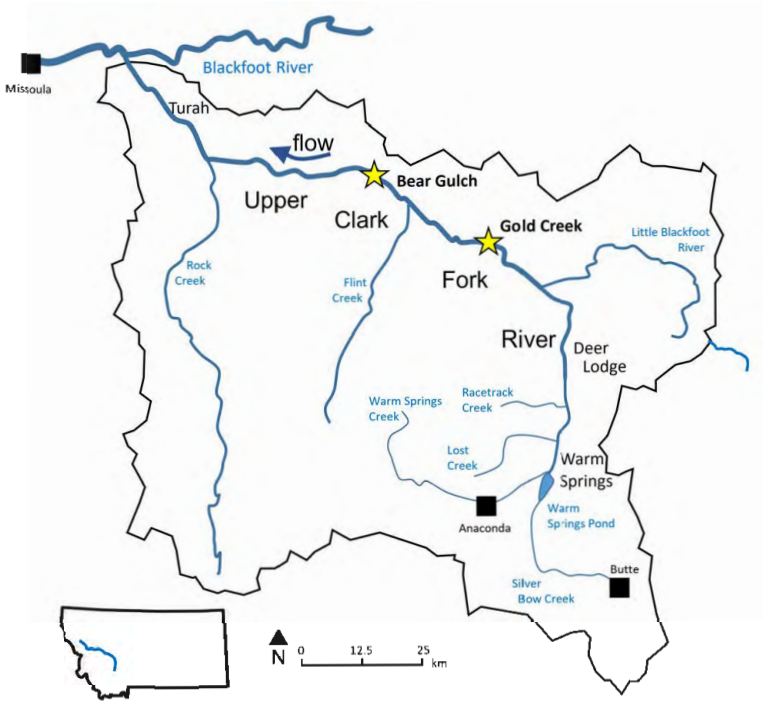


Figure 3.1: The Upper Clark Fork River originates at Warms Spring and flows approximately 200 km downstream to its confluence with the Blackfoot River. Stars indicate sample sites located on the main stem at Gold Creek and Bear Gulch.

to mats of bluegreen algae (Valett et al., in press), which uniquely employ the accessory pigment phycocyanin. The progression of RABs in the UCFR is spatially variable with stages of the bloom differing among locations. The visible change in *Cladophora* blooms, including accrual of epiphytic algae and change in physiological status, along with the shift to bluegreen bacteria and its exclusive pigment, phycocyanin, form the basis for remotely sensing algal standing crops of different composition and character.

Imaging System

Imager. Hyperspectral images were gathered using a Pika L hyperspectral imaging system (Resonon Inc., Bozeman, MT, USA), with a spectral range of approximately 387–1023 nm and nominal band spectral resolution of 2.1 nm, resulting in 300 spectral channels.

During data collection, the Pika L was fitted with a 17 mm objective lens, resulting in a 17.6° across-track full-angle field of view. This pushbroom imager was combined with an airborne imaging system that included a system control unit with onboard storage, GPS and inertial measurement unit, and downwelling irradiance sensor, forming the payload for the UAV.

UAV. The airborne hyperspectral system described above was mounted to a DJI Matrice 600 Pro hexacopter (DJI, Shenzhen, China) via a DJI Ronin-MX gimbal. The Matrice 600 Pro measures 1668 × 1518 × 727 mm, with all frame arms, propellers, and GPS antennae unfolded, and a total weight of approximately 9.5 kg. With a total takeoff weight of 15 kg, the UAV was able to carry a payload of approximately 5.5 kg. Operation of the UAV was handled via a remote controller with a 2.4 GHz radio link and maximum transmission distance of approximately 5 km coupled with a tablet computer, allowing for autonomous flight using DJI Ground Station Pro software. In addition to the GPS system incorporated into the imaging system, location information is supplied through the onboard flight control system on the UAV via three additional GPS antennae.

Flight Overview

All UAV flights were conducted either by, or under the supervision of, a pilot certified through the Federal Aviation Administration (FAA) Part 107 licensure during each data collection campaign on 17 August and 9 September 2021. The airspace above each sampling site was verified as Class D using the FAA-certified AirMap software (AirMap, Santa Monica, CA, USA). To minimize changes in lighting conditions, flights were scheduled within two hours of solar noon, or between approximately 11:00–15:00 MDT on mostly clear days with calm wind to minimize image blurring and errors in calibration techniques caused by tilting of the UAV. Prior to flying, both the UAV and airborne imaging system were inspected and calibrated according to manufacturer specifications [122, 123] to ensure proper operation and precise heading and location information. After inspection and calibration, reflectance tarps

were laid out along the river bank within the field of view of the imaging system.

Flight plans were created prior to each mission using DJI Ground Station Pro software. Each flight was conducted in 300–500 m straight-line segments over the center of the river channel at an above ground level of 120 m, the maximum allowed by the FAA, and a flight speed of approximately 2.5 m/s. Flight parameters were chosen to maximize the amount of river channel captured in each image while minimizing image distortion caused by increased flying speeds and frequent aerial movements.

The airborne hyperspectral imaging system and downwelling irradiance sensor were configured prior to each flight using Resonon Ground Station software. Exposure settings on the imager were held constant throughout each flight after being set through an auto exposure routine using the 11% reflectance target illuminated by direct sunlight. Frame rate was calculated by inputting the flight height and speed into the Ground Station software. The pushbroom hyperspectral imager was set to capture a new data cube after 2000 image lines were captured, at which time the spectrometer would capture a downwelling spectral irradiance measurement.

Ground-Truth Data Collection

To build a relationship between image data collected by the hyperspectral imager and algal pigment abundance, ground-truth data were collected immediately prior to UAV flights on 17 August and 9 September 2021. On 17 August, prior to UAV flights, seven plots were established randomly on the stream bottom at each sample site. Circular plots were delineated with 112 cm sections of white hose filled with sand to ensure that spectral and pigment data were derived from the same locations. Sampling on 9 September was conducted at the Gold Creek site with 20 plots placed in four sets of five organized transects, purposefully moving from areas of low to high growth (Figure 3.2).

After image data were collected, water depth was measured and all benthic substrata



Figure 3.2: Sample plots and transects at the Gold Creek site on 9 September 2021. Five hoops (small white circles) were placed along four transects (red ovals, two of five shown) to establish plots moving from areas of low to high algal growth.

(rocks, etc.) were collected within each plot. Pigment abundance was determined as areal standing crops for three growth forms: filamentous algae, epiphytes, and epilithon. Filamentous algae were removed from the surface of stones within each plot, rinsed and squeezed, and retained independently. Rinse from cleaning (i.e., squeezing and rinsing) filamentous algae was retained separately to quantify epiphytic pigments. Epilithic algae were removed using steel brushes and the resultant slurry was processed following Steinmann et al. [82]. All samples were bagged and placed in a cooler to avoid direct sunlight and keep samples chilled.

Samples were analyzed for chl *a* using 90% buffered acetone for pigment extraction and spectrophotometric assessment with acidification following the Ritchie method [124]. Phycocyanin standing crops extracted from epilithic and epiphytic samples were obtained through the use of a handheld fluorimeter (Aquafluor 8000, Turner Designs, Sunnyvale, CA, USA) following Moulton et al. [125] and Cremella et al. [126]. Calculations for final standing crops included corrections for sub-sampling and were normalized to sampling area.

For each plot, benthic organic matter (BOM) standing stocks were measured for each growth form as AFDM (g/m^2) following standard protocols [82]; dry mass was measured following 24 h at 60 °C, and ash mass after sample ignition for one hour at 550 °C, cooling and re-weighing. For slurries of epilithon, subsamples were filtered through pre-weighed Whatman GF/F (0.7 μm pore size) filters. For filamentous algae, subsamples of wet weight were used for drying and combustion. One corrupted data point was removed from the 17 August Gold Creek data set.

Image Preprocessing

Image Calibration. Hyperspectral data captured by the imaging system have a native pixel format of raw digital counts. Calibrating the pixel values to a common physical scale is key to achieving accurate image processing and comparison among imaging systems. Image calibration was performed using two methods: reflectance targets laid out along the river bank and a downwelling irradiance sensor mounted on the UAV.

The first of these methods used two G8T portable fabric reflectance targets (Group 8 Technology Inc., Provo, UT, USA). The reflectance standards had spectral reflectances of approximately 6% and 11% from 400–1000 nm, with sizes of 1.2 × 1.2 m and 2 × 2 m, respectively. Both tarps were laid out in flat areas along the river bank in locations that would be captured by the imager during UAV flights. As the tarps were contained in only a portion of a single image, the pixels containing the tarps were used as a reflectance standard from which all other image pixels were calibrated throughout the flight using Equation (3.1):

$$\rho(\lambda)_{scene} = \frac{L(\lambda)_{scene} - L(\lambda)_{dark}}{L(\lambda)_{tarp} - L(\lambda)_{dark}} \rho(\lambda)_{tarp}, \tag{3.1}$$

where $\rho(\lambda)_{scene}$ is the spectral reflectance of the scene, $L(\lambda)_{scene}$ is the spectral radiance of the imaged scene, $L(\lambda)_{tarp}$ is the spectral radiance of the reflectance standard, $L(\lambda)_{dark}$ is the dark current signal on the image sensor, and $\rho(\lambda)_{tarp}$ is the spectral reflectance of

the standard. Though this method is an accepted calibration technique, it relies on several assumptions, such as constant lighting conditions throughout the flight and pixel-to-pixel response consistency.

The second method removed the assumptions implicit in the first method by using an Ocean Insight Flame UV-VIS spectrometer (Ocean Insight, Orlando, FL, USA) fitted with a cosine corrector and calibrated to measure downwelling spectral irradiance from 400–1000 nm with nominal spectral resolution of 1.34 nm. The spectrometer was configured to synchronously measure spectral downwelling irradiance with each hyperspectral data cube captured by the imager. This method allowed each hyperspectral data cube to be converted to reflectance with precise illumination information throughout the flight and thereby removed the assumption of unchanging illumination, using

$$\rho(\lambda)_{scene} = \frac{L(\lambda)_{scene} - L(\lambda)_{dark}}{\frac{E(\lambda)_{downwelling}}{\pi} - L(\lambda)_{dark}}, \tag{3.2}$$

where $E(\lambda)_{downwelling}$ is the downwelling spectral irradiance and π is the hemispheric projected solid angle, with radiometric calibration of both the Pika L and Flame spectrometer performed by Resonon.

Pixel Selection. After the hyperspectral images were converted from digital counts to reflectance, spatial pixels within each plot that were not saturated by sun glint were averaged using Resonon Spectronon software. For a 120 m flight altitude with a 17 mm focal length lens, the across-track image contained 900 pixels in a 17.6-degree field of view, such that each plot contained an average of approximately 40 pixels after flying over each plot. The spectra contained in each of these spatial pixels was averaged to create a single spectrum for each plot.

Image Processing

Hyperspectral image data were processed by isolating the signal reflected from the stream bed harboring algal growth, converting image data to physical units of reflectance,

and developing and applying spectral indices following the steps outlined below.

Theoretical Basis for Isolating Stream Bed Reflectance. The goal of the image processing method presented was to relate the upwelling radiance measured by the hyperspectral imaging system to the pigment abundance contained within benthic algal blooms. Following a development similar to that presented by Legleiter, et al. [75, 76], total upwelling spectral radiance [W / (m^2 sr nm)] measured by the UAV-based imaging system can be written as

$$L_{tot}(\lambda) = L_b(\lambda) + L_{wb}(\lambda) + L_{ws}(\lambda) + L_{atm}(\lambda), \tag{3.3}$$

where L_{tot} is the total upwelling spectral radiance, L_b is the spectral radiance reflected from the benthos, L_{wb} is the spectral radiance scattered to the imager from the water body, L_{ws} is the spectral radiance reflected from the water's surface, L_{atm} is the spectral radiance scattered from the atmosphere, and all terms are a function of wavelength, λ. Each term can also have variable polarization, which creates radiometric errors if the polarization-sensitive instrument response is ignored [127, 128]; however, this becomes negligible for the small incidence angles (less than 8°) employed in this work. Of the terms addressed in Equation (4.1), L_b is of primary interest when relating image data to a difference in benthic algal pigment contents. Owing to the nature of the UCFR and UAV-based data collection, several assumptions can be made to simplify Equation (4.1) and isolate the signal of interest.

With the exception of stormflow and snowmelt conditions, the UCFR is clear and shallow, meaning the majority of the downwelling solar irradiance will be reflected from the stream bed itself, making L_b the dominant signal reaching the imager. Upwelling radiance from optical scattering within the water body prior to reaching the bottom substrate, L_{wb}, largely depends on the water's inherent optical properties. These generally include the absorption coefficient, scattering coefficient, and volume scattering function, which are functions of both depth and dissolved and suspended matter present in the stream [77, 78]. Owing to the clarity

and shallowness of the UCFR, radiance generated through L_{wb} can be considered negligible when compared to the relatively large signal of L_b. The signal generated by L_{ws} largely depends on the reflectance of the water's surface, which is a function of surface conditions and stream morphology. The reflectance of the water's surface has been measured as low as 0.03 between 400–800 nm, with reflectance values approaching 0 for calm surfaces [75], making L_{ws} small compared to L_b. The final term, L_{atm}, typically becomes negligible when optical signals take relatively short paths through a clear atmosphere, such as UAV operation at 120 m on a clear day. However, under smoky conditions, which often occur during the summer in Montana, atmospheric radiance may increase. Radiative transfer simulations show that elevated atmospheric scattering driven by smoke generates a spectrally-flat reduction in contrast from 400–1000 nm; however, the increased L_{atm} is still much smaller than the signal generated by L_b.

Taking these assumptions and simplifications into account, total upwelling radiance reaching the imager can be considered to be dominated by upwelling radiance from the benthos and Equation (4.1) can be rewritten only in terms of L_b as

$$L_{tot} \approx L_b = \frac{E_d T_{aw} T_{wa} R_b e^{-2\kappa d}}{\pi}, \tag{3.4}$$

where E_d is the downwelling irradiance [W / m^2]; T_{aw} is the air-water interface transmittance; T_{wa} is the water-air interface transmittance; R_b is the reflectance of the stream bed; κ is the optical extinction coefficient; d is the depth of the water body, which must be considered on both its downwelling and upwelling path through the water; and the factor of pi represents the bidirectional reflectance distribution function for a Lambertian surface. Transmission across the air-water interface occurs twice, once when the optical signal enters the water (T_{aw}) and again when the scattered sunlight leaves the water (T_{wa}). For small incidence angles, the transmittance across this interface is equal when moving between the media; however, for

large incidence angles these terms may need to be considered separately.

As the primary signal of interest for identifying algal pigment standing crops is the stream bed reflectance, Equation (3.4) can be rewritten as

$$R_b = \frac{R_{tot}}{T_{aw} T_{wa} e^{-2\kappa d}}, \tag{3.5}$$

where R_{tot} is the reflectance measured at the imager, or L_{tot}/E_d. Equation (3.5) provides a basis for isolating the reflectance of the stream bed to accurately assess changes in pigment abundance.

Calculating or measuring the extinction coefficient, κ, enables correction for the effects of optical absorption and scattering within the water as a function of wavelength and depth. Accounting for these losses may allow for the use of wavelengths that are strongly absorbed by water, which contain important pigment information. However, correcting for the effects of water extinction requires continuous measurement of the depth of the studied water body, making the method largely inaccessible for most remote sensing applications. The results presented here do not include an extinction correction because it is less important for shallow water and it is preferred to present a method that is more readily implemented by others using the equipment presented herein.

Development of Spectral Indices and Application to Algal Standing Crops. Using methods similar to those presented in earlier works [76, 101], a brute-force analysis was performed to relate hyperspectral reflectance data to in situ samples of chl *a* and phycocyanin. Reflectance data across the spectral range of 400–850 nm from each measurement plot were extracted to form all possible band ratio combinations of the form $R_{\lambda_1}/R_{\lambda_2}$, where R_λ represents the reflectance at a single 2.1 nm-wide spectral channel. Data were omitted for wavelengths longer than 850 nm, where the signal-to-noise ratio declines rapidly because of increasing water absorption and decreasing detector responsivity. The resulting values were

matched to corresponding in situ pigment measurements and linear regressions performed to relate pigment standing crops to each band ratio's spectral signal. The goodness of fit for each model was assessed via the coefficient of determination (R^2) for each data set. The approach generated a large number (45,796) of band ratios and R^2 values for each data set analyzed. Data sets containing the band ratios and the resultant R^2 values were analyzed with frequency distributions to characterize the spread of potentially useful band ratios. A matrix of R^2 values for each band ratio combination was developed to depict correlation intensity across all band ratio combinations. The matrix was used to highlight areas sensitive to the target pigment, from which the band ratio with the highest R^2 value was saved as the optimal band ratio for the data set.

With this approach, two major factors influencing the identity of the appropriate spectral signal (i.e., band ratio) and the quality of the linear models employed to estimate pigment standing crops were addressed. First, independent models were generated for different growth forms including: (i) filamentous plus epiphytic (fila/epip), (ii) epilithic, and (iii) total (filamentous, epiphytic, and epilithic). Models estimating phycocyanin abundance were restricted to epiphytic and epilithic growth, as no bluegreen bacteria were associated with *Cladophora* filaments. Second, the influence of RAB development over time was addressed by comparing combined and independent models derived from multiple dates (17 August, 9 September). Finally, the combined influence of form and stage was analyzed by generating models specific to individual growth forms for different sampling dates.

Uncertainty Analysis

The optimal band ratios selected by the brute-force method were based on correlating a small number of pigment observations (n = 33) to spectral models with a large number of potential explanatory variables (n = 45,796), creating conditions that could generate spurious relationships. To test the generalizability and uncertainty around all band ratios, including

those identified as optimal by the brute-force method, a resampling approach was used. In this approach, the 33 pigment observations were subsampled into groups of 26 (representing 80% of the observations) and these subsamples were repeated 1000 times without replacement. This resampling approach was chosen as it would allow comparison between consistently sized subsamples of the data. For each subsample, a linear regression was performed between each of the 45,796 possible band ratios and the subsample of pigments, generating 45,796,000 linear models and R^2 values (1000 for each band ratio). The average R^2 value was calculated for each band ratio across all 1000 resamples. With this method, the performance of each of the optimal band ratios identified by the brute-force analysis could be assessed against each pigment subsample, giving a distribution of R^2 values and a standard deviation around the mean R^2 value for each pigment growth form. The subsampling analysis was run for total standing crops of chl *a* and phycocyanin, as well as after separating by growth form.

Results

RAB Characteristics

When plots were randomly established at both Gold Creek and Bear Gulch, total chl *a* standing crop ranged from a high of 316.5 mg/m^2 to a low of 69.3 mg/m^2 and averaged 191.4 ± 33.1 mg/m^2 and 183.5 ± 31.8 mg/m^2 at Gold Creek and Bear Gulch, respectively (Table 3.1). When samples at Gold Creek were purposefully distributed to include transects into the primary RAB, total chl *a* standing crops ranged from a high of 377.8 mg/m^2 to a low of 46.1 mg/m^2 and averaged 186.0 ± 30.3 mg/m^2. The majority of chl *a* was found in standing crops of filamentous growth across all field sites and sampling dates. Epilithic chl *a* standing crops tended to be greater than those of epiphytic chl *a* at Gold Creek, whereas Bear Gulch contained similar levels of each. Total standing crop of BOM was highest on 17 August at Gold Creek (97.4 g AFDM/m^2), including a large proportion (88%) of filamentous growth. Between 17 August and 9 September, total BOM decreased at Gold Creek, largely

due to a decrease in filamentous forms. Bear Gulch showed the lowest standing stocks of filamentous BOM, but larger standing stocks of epiphytic and epilithic BOM compared to the Gold Creek data sets. Total phycocyanin standing crops were similar between field sites; however, abundance was distributed almost evenly between epiphytic and epilithic growth forms at Bear Gulch whereas the majority of phycocyanin standing crops were contained in epilithon at Gold Creek.

Table 3.1: Standing crop of chlorophyll a (chl *a*), benthic organic matter (BOM), and phycocyanin at Gold Creek and Bear Gulch. Data are means ± standard error; n=6 at Gold Creek on 17 August, n=7 at Bear Gulch on 17 August, and n=20 at Gold Creek on 9 September.

Field Site	**Field Date**	**Growth Form**	**Chl *a* (mg/m^2)**	**BOM (g FDM/m^2)**	**Phycocyanin ($\mu g/m^2$)**
Gold Creek	17 August	Filamentous	158.9 ± 30.3	85.8 ± 13.3	NA
		Epiphytic	7.3 ± 2.8	4.2 ± 0.4	6.5 ± 2.2
		Epilithic	25.2 ± 2.2	8.8 ± 13.9	31.3 ± 4.6
		Total	191.4 ± 33.1	97.4 ± 13.9	37.8 ± 4.6
	9 September	Filamentous	149.5 ± 24.0	67.4 ± 11.0	NA
		Epiphytic	10.0 ± 2.5	5.0 ± 1.7	7.2 ± 2.4
		Epilithic	26.6 ± 3.8	9.9 ± 1.3	27.2 ± 4.6
		Total	186.0 ± 30.3	82.4 ± 12.1	34.4 ± 5.5
Bear Gulch	17 August	Filamentous	124.8 ± 25.8	55.5 ± 14.8	NA
		Epiphytic	26.8 ± 9.9	17.2 ± 3.7	20.8 ± 5.2
		Epilithic	31.9 ± 5.2	15.4 ± 2.5	21.7 ± 2.6
		Total	183.5 ± 31.8	88.1 ± 18.1	42.5 ± 5.3

Spectral Models

Estimative modeling began by running the brute-force analysis using the combined data set, composed of pigment data collected across all field sites and dates and averaged spectra from corresponding plots. Each plot consisted of a single spectrum obtained by averaging the spectra contained in all hyperspectral pixels within the plot (Figure 3.3). The model was run using total chl *a* standing crops (i.e., sum of all growth forms) and total phycocyanin abundance, generating three output products (Figure 3.4): a heat map of coefficient of determination (R^2) values for every possible band ratio (Figure 3.4a), a histogram showing the counts of R^2 values (Figure 3.4b), and a linear regression between the optimal band ratio and the targeted pigment (Figure 3.4c).

The heat maps for each regression represent the resulting R^2 value generated for every possible band ratio, created from a particular wavelength as denominator (y-axis) and numerator (x-axis). The heat maps allow for a quantifiable means of selecting the optimal band ratio (i.e., The band ratio with the highest R^2 value) and determining spectral regions which are sensitive to the targeted pigment. Next, the histograms show the salience of the regressions generated by the brute-force analysis by displaying the number of band ratios which generated the R^2 values shown in the heat maps. Finally, the linear regressions for each data set are a graphical representation of the fit between the best-performing band ratio, selected by the brute-force algorithm, and the pigment under analysis. For each linear regression, the field location and date from which the ground-truth pigment standing crop was measured is designated with a unique marker and color.

A band ratio of 684/674 nm was selected as the optimal ratio for estimating chl *a* standing crops, which is likely based on the known chl *a* absorption line near 665 nm [129, 130]. Of the 45,796 possible band ratios, the optimal band ratio was one of eight which produced an R^2 above 0.40 (Figure 3.4b). Analysis of the combined data set showed promising results when estimating chl *a* standing crops, with an $R^2 = 0.57$ and root-mean-square

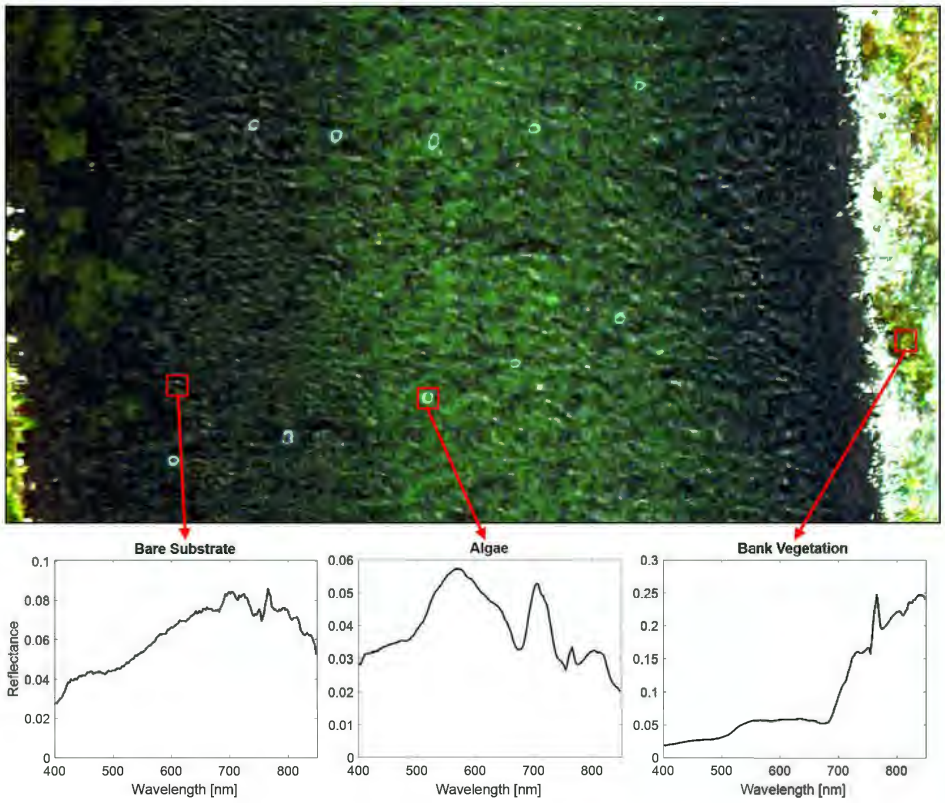


Figure 3.3: RGB image created from one of the hyperspectral image cubes gathered using the UAV system at Gold Creek. Examples of the complete spectra contained at each pixel are shown for bare river substrate, Cladophora algae, and bank vegetation for the spectral range of 400–850 nm. Note the varying reflectance axes and strong water absorption beyond approximately 725 nm that suppresses the strong near-infrared reflectance of the submerged algae.

error (RMSE) = 66.29 mg/m^2. Regression analysis of phycocyanin standing crops resulted in worse estimation performance, with an $R^2 = 0.41$ and RMSE = 15.54 mg/m^2. An optimal band ratio of 753/824 nm was selected, highest among 10 band ratios which resulted in an R^2 above 0.35 (See Supplementary Figures).

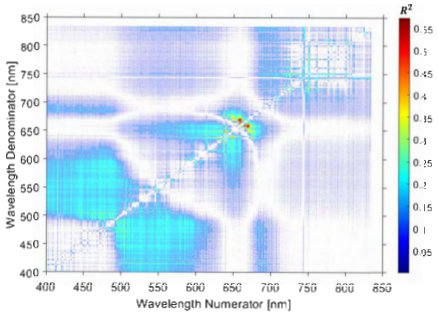
Separation by Growth Form. Though the combined data set showed promising results, the optical signal received by the imaging system was likely dominated by either large mats of filamentous algae obscuring the epilithic growth or areas of solely epilithic growth with minimal filamentous algae. To explore this possibility, the pigment abundances related to the filamentous algae and epiphytes were separated from the abundances generated from

epilithic sources, and each data set was used as input to the brute-force analysis using the same imagery, with zero-valued points removed. As epiphytes tend to collect and grow on the surface of filamentous algae, the optical signals for these growth forms were likely conflated; therefore, the chl *a* standing crops associated with these two forms were combined into a single group.

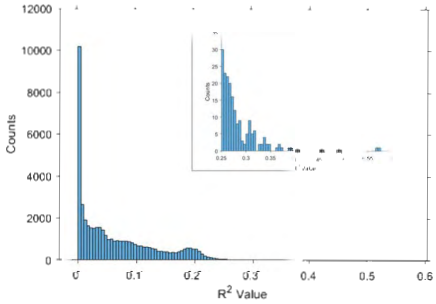
Filamentous Plus Epiphytic. After separating each data set, the strength of relationships between target pigment abundance and image data were compared to the combined data set's performance when estimating total pigment standing crops. The difference in model performance is described as ΔR^2, which is positive for increases in R^2 and negative for decreases. The performance of the regression analysis modestly improved when estimating fila/epip chl *a* standing crops individually, with a ΔR^2 of 0.05, increasing from 0.57 to 0.62, while maintaining an optimal band ratio of 684/674 nm. With improved performance, the salience of this ratio also increased; the optimal band ratio was one of six generating an R^2 above 0.50 (See Supplementary Figures).

Estimation of phycocyanin standing crops was limited to epiphytic and epilithic growth forms as phycocyanin was not associated with the filamentous growth forms present in the UCFR. Estimative performance of epiphytic phycocyanin standing crops decreased compared to the combined data set estimates, with a ΔR^2 of -0.08. The decreased performance was associated with an optimal band ratio similar to those selected for chl *a* estimation, with a band ratio of 688/655 nm (See Supplementary Figures).

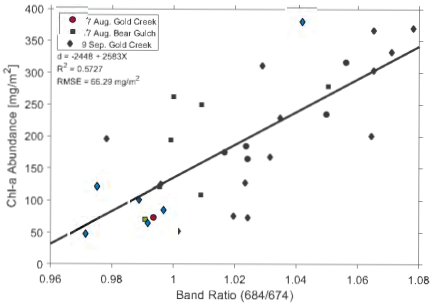
Epilithic. Estimation of epilithic chl *a* standing crops decreased compared to the combined data set, with a loss in R^2 resulting in a ΔR^2 of -0.13, decreasing from 0.57 to 0.44 while the optimal band ratio shifted to longer wavelengths of 729/809 nm (See Supplementary Figures). Performance for estimating epilithic phycocyanin standing crops improved compared to the combined data set, with R^2 values increasing from 0.41 to 0.50,



(a) Heat map of R^2 values generated by fitting each band combination against total chl *a* standing crops. Colors within the heat map represent the R^2 value, with wavelength numerators and denominators shown on the x- and y-axes, respectively.



(b) Histogram of the R^2 values generated using the brute-force analysis. The insert shows counts beyond an $R^2 = 0.25$, showing that relatively few band ratios generate fits in this region.



(c) Linear regression analysis between chl *a* abundance (y-axis) and optimal band ratio (x-axis). Data collected on 17 August 2021 at the Gold Creek and Bear Gulch site represented by red circles and green squares, respectively, while data collected on 9 September 2021 are shown with blue diamonds.

Figure 3.4: Analytics from the regression analysis of total chl *a* standing crops (sum of filamentous, epiphytic, and epilithic sources) from the combined (all field sites) data set.

resulting in a ΔR^2 of 0.09 while using an optimal band ratio of 738/804 nm (See Supplementary Figures).

The improved strength of the relationship for estimating chl *a* in fila/epip growth forms separately and decrease in performance for epilithic chl *a* suggests that the optical signal from filamentous and epiphytic algal growth dominates the imaged scene during a large bloom. The increased performance for epilithic phycocyanin is likely due to the larger abundance of phycocyanin commonly found in epilithic algal growth. Those pixels which contain epilithic growth, which are not obscured by filamentous growth, are likely better estimated when separated by growth form.

Separation by Phenology. Sources of algal biomass change throughout the summer growing season, with the ratio of biomass sourced from epilithon, epiphytes, and filamentous algae shifting as a function of stream conditions. This change has been observed to depend on sampling location along the UCFR, with field sites farther downstream progressing through their life cycle prior to those closer to the headwaters. Compared to both Gold Creek sites, the Bear Gulch site visually appeared to contain elevated epiphyte colonization relative to filamentous organic matter. To explore the role phenology had in band ratio selection, the data collected from the two field sites were separated, forming a Bear Gulch data set (n = 7) and Gold Creek data set (n = 26). These two data sets were used as inputs to the brute-force analysis to estimate standing crops from all pigment sources, epiphytic and filamentous sources, and epilithon.

Bear Gulch.

Total. Separation based on phenology led to improvements in the overall strength of the regression relationship, with performance increasing for all pigments and sources. When estimating total chl *a*, the brute-force analysis selected a band ratio of 554/536 nm, resulting in a ΔR^2 of 0.38, increasing from the combined model's R^2 value of 0.57 to 0.95

(See Supplementary Figures). Estimation of total phycocyanin at Bear Gulch shifted the optimal band ratio from 753/824 nm in the combined model to 625/619 nm (See Supplementary Figures). The associated ΔR^2 was 0.51, with R^2 values increasing from 0.41 to 0.92.

Filamentous plus Epiphytic. Improvements in the estimation of the fila/epip and epilithic groups were also seen after separating the data by location. When estimating fila/epip chl *a* at Bear Gulch, the brute-force analysis selected a shorter wavelength ratio of 554/536 nm with a ΔR^2 of 0.39 associated with an R^2 of 0.96 (See Supplementary Figures). Similarly, the estimation of epiphytic phycocyanin improved, with a ΔR^2 of 0.53, after R^2 values increased from 0.41 to 0.94 and selecting an optimal band ratio of 748/740 nm (SSee Supplementary Figures).

Epilithic. Epilithic pigment estimation accuracy also increased for both chl *a* and phycocyanin, with R^2 values improving from 0.57 to 0.86 and 0.41 to 0.89, respectively, leading to ΔR^2 values of 0.29 and 0.48 compared to the combined model (See Supplementary Figures). Estimation of epilithic chl *a* from the Bear Gulch data set relied on shorter-wavelength ratios as compared to the combined data set, selecting a band ratio of 573/589 nm. The ratios selected for epilithic phycocyanin also shifted to shorter wavelengths, using a band ratio of 525/519 nm (See Supplementary Figures).

Gold Creek.

Total. Analyzing the Gold Creek data separately resulted in improved estimates compared to the combined data set across all growth forms. The strength of the relationship between total chl *a* and image data at Gold Creek improved from an R^2 of 0.57 to 0.63 ($\Delta R^2 = 0.06$) while for total phycocyanin, R^2 values increased from 0.41 to 0.46, resulting in a ΔR^2 of 0.05. The optimal band ratio for estimating total chl *a* from the Gold Creek data set remained the same between the data sets, maintaining a band ratio of 684/674 nm (See

Supplementary Figures). The optimal band ratio for total phycocyanin showed a near-inverse relationship compared to the combined data set, shifting from 753/824 nm to 804/748 nm (See Supplementary Figures).

Filamentous plus Epiphytic. The strength of the linear regression was highest at Gold Creek when estimating fila/epip chl *a*, with a ΔR^2 of 0.11, after R^2 values increased from 0.57 to 0.68 (Figure 3.5). Notably, the optimal band ratio remained the same for the combined data set and the Gold Creek data set. This band ratio is the vicinity of known chl *a* absorption near 665 nm, suggesting the brute-force analysis relied on a change in chl *a* abundance when estimating total and fila/epip chl *a* standing crops. Estimation of epiphytic phycocyanin also showed modest improvements, increasing from an R^2 of 0.41 to 0.43, an improvement associated with a ΔR^2 of 0.02 and the selection of a band ratio of 778/774 nm (See Supplementary Figures).

Epilithic. The regression analysis for estimating epilithic pigmentation also improved for chl *a* (See Supplementary Figures) and phycocyanin (Figure 3.5), with R^2 values increasing from 0.57 to 0.62 and 0.41 to 0.59, respectively, leading to ΔR^2 values of 0.05 and 0.18. The optimal band ratio for estimating epilithic phycocyanin shifted to longer wavelengths, selecting an optimal band ratio of 804/740 nm.

The correlation maps for the Gold Creek data set show that wavelengths surrounding chl *a* absorption lines are favored when estimating total chl *a* abundance, fila/epip chl *a*, and epiphytic phycocyanin. However, longer-wavelength band ratios are preferred when estimating the remaining pigments and forms, often selecting wavelengths over 700 nm.

Results Summary. Separating the combined data set by growth form (pigment source) and phenological state (field site) generally resulted in improved performance for estimating both chl *a* (Table 3.2) and phycoycanin (Table 3.3) standing crops, with the exception of epilithic chl *a* for which performance dropped after growth form separation. The strongest

relationship between band ratio and pigment abundance was found after separating by phenological state and estimating pigment standing crops at the Bear Gulch. Although the linear regression models at the Bear Gulch site produced R^2 values above 0.86, this field site was limited to seven data points, which may introduce problems such as spurious fits. Optimal band ratio selection was fairly robust when estimating chl *a* abundance at Bear Gulch, converging on a band ratio of 554/536 nm for total and fila/epip chl *a*; however, optimal band ratio selection for phycocyanin abundance was more variable.

Table 3.2: Model performance and optimal band ratio selected for chl *a* abundance after separating data sets by growth form and phenological state. The ΔR^2 value describes the change in R^2 after separating the data set, relative to the performance of the combined data set estimating total pigment abundance.

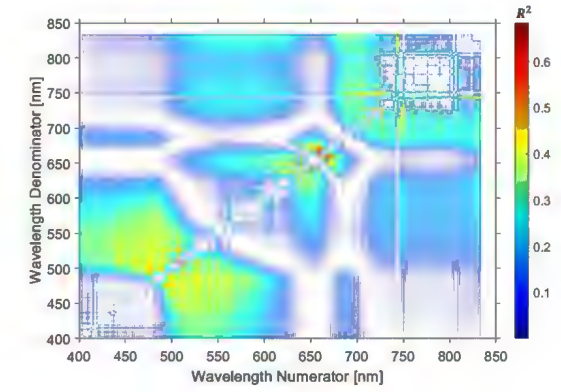
Data Set	**Growth Form**	**Optimal Band Ratio**	$\boldsymbol{R^2}$	$\boldsymbol{\Delta R^2}$
Combined	Total	684/674	0.57	-
	Fila/epip	684/674	0.62	0.05
	Epilithic	729/809	0.44	−0.13
Bear Gulch	Total	554/536	0.95	0.38
	Fila/epip	554/536	0.96	0.39
	Epilithic	573/589	0.86	0.29
Gold Creek	Total	684/674	0.63	0.06
	Fila/epip	684/674	0.68	0.11
	Epilithic	804/738	0.62	0.05

The strongest relationship using the Gold Creek data set was found after separating the data sets by growth form and analyzing epilithic phycocyanin and fila/epip chl *a* standing crops (Figure 3.5). Estimation of epilithic phycocyanin abundance produced an R^2 of 0.59 and used a band ratio of 804/740 nm (Figure 3.5b,d,f). Analysis of fila/epip chl *a* produced

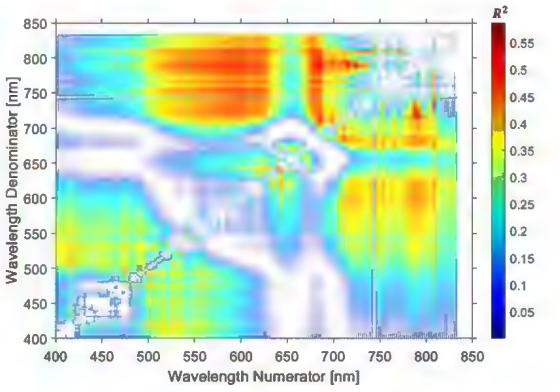
Table 3.3: Model performance and optimal band ratio selected for phycocyanin abundance after separating data sets by growth form and phenological state. The ΔR^2 value describes the change in R^2 after separating the data set, relative to the performance of the combined data set estimating total pigment abundance.

Data Set	**Growth Form**	**Optimal Band Ratio**	$\boldsymbol{R^2}$	$\boldsymbol{\Delta R^2}$
Combined	Total	753/824	0.41	-
	Epiphytic	688/655	0.33	−0.08
	Epilithic	738/804	0.50	0.09
Bear Gulch	Total	625/619	0.92	0.51
	Epiphytic	748/740	0.94	0.53
	Epilithic	525/519	0.89	0.48
Gold Creek	Total	804/748	0.46	0.05
	Epiphytic	778/774	0.43	0.02
	Epilithic	804/740	0.59	0.18

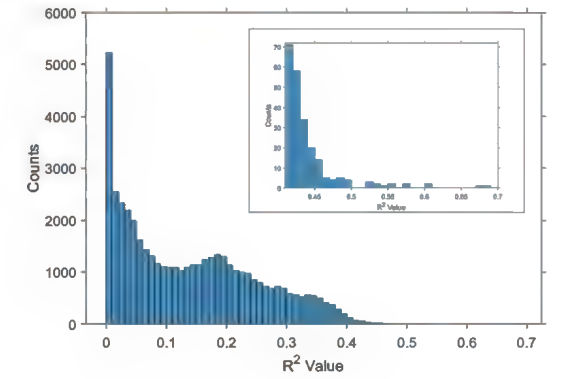
an R^2 of 0.68 and converged on a band ratio of 684/674 nm, found in several chl *a* abundance measurements for the Gold Creek and combined data (Figure 3.5a,c,e). When applied at Bear Gulch, the optimal band ratio for the combined and Gold Creek data produced reasonable fits when estimating chl *a* standing crops. Fitting to total and fila/epip chl *a* standing crops, the band ratio of 684/674 nm produced R^2 values of 0.37 and 0.36, respectively, meaning this band ratio may carry meaningful information across field sites (i.e., phenological states).



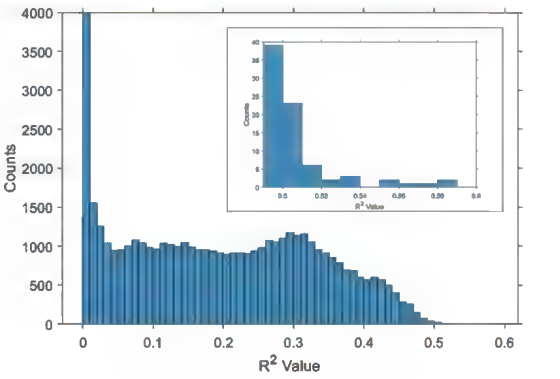
(a) Gold Creek fila/epip chl *a* correlation map



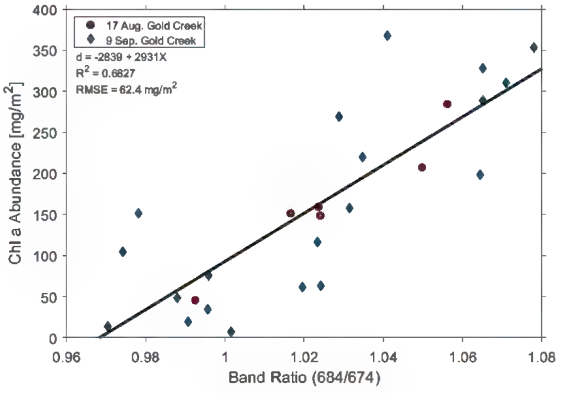
(b) Gold Creek epilithic phycocyanin correlation map



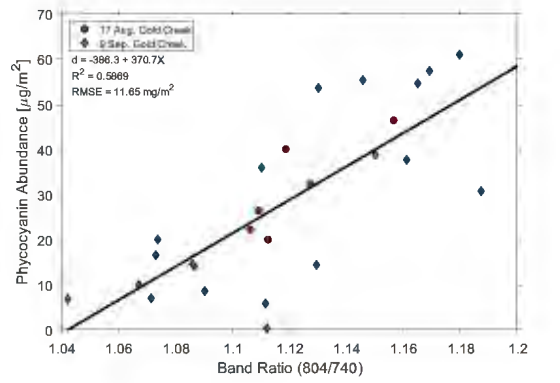
(c) Gold Creek fila/epip chl *a* histogram



(d) Gold Creek epilithic phycocyanin histogram



(e) Gold Creek fila/epip chl *a* linear regression



(f) Gold Creek epilithic phycocyanin linear regression

Figure 3.5: Linear regressions, correlation maps, and histograms for the Gold Creek data set. Results gathered via the brute-force analysis for estimating fila/epip chl *a* abundance (left column) and epilithic phycocyanin abundance (right column).

Due to its strong performance and frequent selection as optimal, the linear equation obtained from the band ratio of 684/674 was used to estimate pigment abundances for the top-performing data set, found for fila/epip chl *a* measurements at the Gold Creek field site, and compared to the measured abundances (Figure 3.6). In general, the model showed promising results, but tended to overestimate low chl *a* abundances (below ≈ 100 mg/m^2) and underestimate elevated abundances (above ≈ 250 mg/m^2).

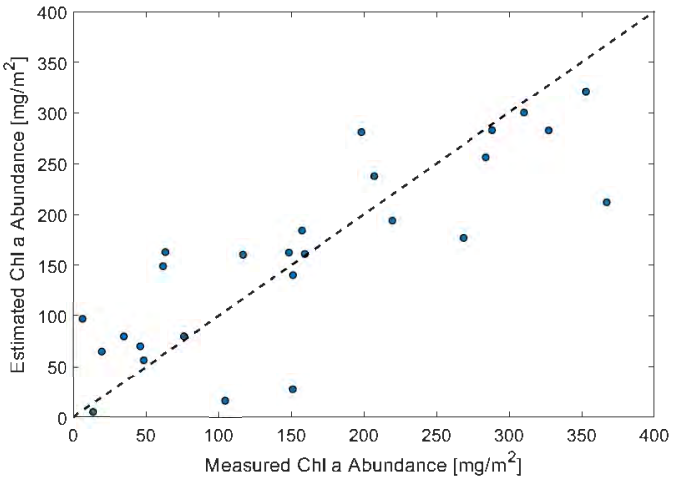


Figure 3.6: Estimated versus measured fila/epip chl *a* abundance at the Gold Creek field site using the regression analysis generated by the brute-force method. Points falling along the 1:1 line (dashed) represent perfect performance.

Generalizability of Band Ratios

The resampling approach identified the same optimal band ratios and similar average R^2 values as the brute-force method for four out of the six combinations of pigment and growth form, and identified the inverse band ratio for one. For all of the chl *a* growth forms, the same optimal band ratios were identified by both the brute-force method and resampling 80% of the observations 1000 times, and the mean R^2 value from the resampling approach was identical to that identified by the brute-force method on all of the samples (Table 3.4). Additionally, the resulting data distribution of the R^2 values across all 1000 resamples for

the optimal band ratios were tightly constrained (Figure 3.7), with standard deviations between 0.05 and 0.06. For phycocyanin, the resampling approach selected the same optimal band ratio for total abundance, the inverse band ratio for epilithic (i.e., brute force, 738/804; resampling, 804/738), while both band ratios had R^2 values that were nearly identical between the resampling and brute-force approaches and small standard deviations. In contrast, the resampling approach identified a different optimal band ratio for the epiphytic growth form for phycocyanin (i.e., brute force, 688/655; resampling, 684/674) and the R^2 value was higher for the resampling approach for this alternative band ratio. If the R^2 is examined from the resampling approach for the optimal band ratio identified by the brute-force method, a mean R^2 of 0.18 $\pm$ 0.19 is generated.

Table 3.4: Optimal band ratio, mean R^2 value, and standard deviation identified after running the uncertainty analysis on total pigment abundance and after separating by growth form.

Pigment	**Growth Form**	**Optimal Band Ratio**		$\boldsymbol{R^2}$	
		Brute Force	**Resampling**	**Brute Force**	**Resampling**
Chl *a*	Total	684/674	684/674	0.57	0.57 $\pm$ 0.06
	Fila/epip	684/674	684/674	0.62	0.62 $\pm$ 0.05
	Epilithic	729/809	729/809	0.44	0.44 $\pm$ 0.06
Phycocyanin	Total	753/824	753/824	0.41	0.42 $\pm$ 0.09
	Epiphytic	688/655	684/674	0.33	0.59 $\pm$ 0.22
	Epilithic	738/804	804/738	0.50	0.51 $\pm$ 0.06

Comparison of Brute-Force Method with Existing Indices

To provide context for the performance of the presented linear relationships, the derived optimal band ratio for estimating total chl *a* abundance from all field data (n = 33) was compared to several existing band ratio methods for monitoring chl *a*, including the Ha

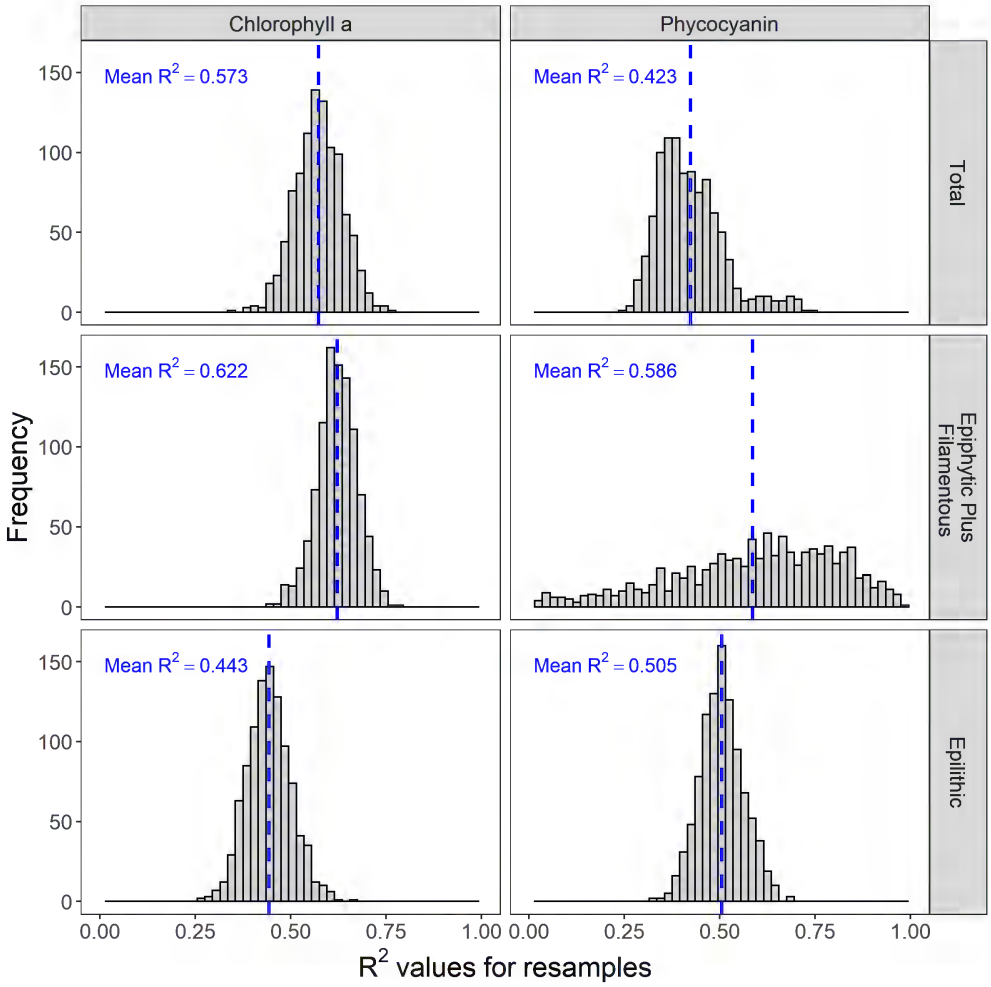


Figure 3.7: Histograms of R^2 values generated by running linear regressions on each iteration of pigment subsampling against the optimal band ratios identified for estimating pigment abundance for each growth form.

Index [131], Shafique Index [48], and NDVI [132]. All three methods employ the use of band ratios for estimating chl *a* concentration in systems such as tropical lakes using Sentinel 2A imagery [131], large rivers using airborne hyperspectral imagery [48], and general vegetated systems [132]. The selected band ratios were generated using UCFR hyperspectral data, matching the spectral resolution reported in the original work, where NDVI was calculated by integrating across the reported spectral width of Landsat 8 bands [133].

When compared to the optimal band ratio identified by the brute-force method, the

selected spectral band ratios performed poorly when tasked with estimating total chl *a* abundance in the UCFR, with a maximum R^2 value of 0.13 (Table 3.5). Surprisingly, although many of the band ratios used similar wavelengths to those selected by the brute-force method, their performance was notably lower, with NDVI having almost no correlation with total chl *a* abundance. The poor performance associated with NDVI is likely attributable to the near-infrared band centered at 865 nm, a spectral region that is strongly absorbed by water. Previous research has shown that the efficacy of NDVI is a function of water depth, with NDVI becoming ineffective for detecting submerged vegetation at depth greater than 30 cm [134].

To explore the effects of spectral resolution on pigment abundance estimations, the spectral channels used to generate NDVI were reduced to 10 nm and 2.1 nm then fit to pigment abundance. When reduced to 10 nm, the R^2 modestly improved to 0.05 (p-value = 0.23) while a spectral resolution of 2.1 nm generated similar results ($R^2 = 0.05$, p-value = 0.20), suggesting spectral bandwidth did not play a major role in performance in this case. Finally, the brute-force method was set to search for the optimal normalized difference index across the spectral region of 650–850 nm, selecting the same wavelengths used in the simple ratio identified by the brute-force method ($\frac{684 - 674}{684 + 674}$), with nearly identical performance ($R^2 = 0.57$, p-value < 0.001).

Discussion

The optimal band ratio for estimating total and fila/epip chl *a* standing crops in the combined and Gold Creek data sets is likely explained by a change in the reflectance spectra near a chl *a* absorption region driven by variations in pigment abundance. The denominator selected by the brute-force method is near the spectral trough of this region whereas the numerator is along the red edge, a wavelength pair which is likely sensitive to changes in chl *a* abundance. The shift to shorter wavelengths at Bear Gulch is likely driven by a larger

Table 3.5: Comparison of selected spectral band ratios for estimating total chl *a* abundance in the UCFR from all field sites (n = 33). Listed wavelengths represent band centers for each spectral ratio while regression equations are presented as linear equations in the form of y = mx + b.

Band Ratio	**Center Wavelengths (nm)**	**Spectral Resolution (nm)**	**Regression Equation**	R^2	***p*-Value**
Ha Index	$\frac{559.8}{664.6}$	30, 35	y = 177.4X − 61.3	0.12	0.048
Shafique Index	$\frac{705}{675}$	5	y = 154.5X − 47.4	0.13	0.041
NDVI	$\frac{865-655}{865+665}$	30	y = -235.5X + 125.8	0.04	0.25
Brute-Force Method	$\frac{684}{674}$	2.1	y = 2583X − 2448	0.57	< 0.001
Brute-Force Method, ND	$\frac{684-674}{684+674}$	2.1	y = 5276X + 135.6	0.57	< 0.001

proportion of total BOM derived from epiphytic and epilithic growth, representing later stages of bloom progression. The results presented here suggest that late-stage algae are characterized by a decrease in NIR reflectance, shortening of the red edge, and increased reflectance across visible wavelengths, all of which likely reduce the information contained in the vicinity of the 665 nm chl *a* absorption line. During this progression, spectral information near the visible reflectance peak between approximately 530–600 nm carries more information, leading the brute-force method to this region.

The optimal band ratios selected by the brute-force method show promising generalizability across pigment growth forms assessed in the resampling analysis. The results of this analysis showed elevated maximum mean R^2 for the optimal band ratio identified for each growth form, as well as low standard deviation. Though this analysis is still data-limited, these early results show that the optimal band ratios may be a promising tool for assessing pigment abundance in clear and shallow rivers similar to the UCFR.

When compared to existing spectral indices developed for estimating chl *a* concentration, the band ratio selected by the brute-force method showed superior performance to all others, with NDVI performing quite poorly. The poor performance of the existing band ratios may be attributed to their general development for measuring volumetric chl *a* concentration, broader spectral resolution, which tends to smear small spectral features [48, 131], or, in the case of NDVI, the use of longer wavelengths that are strongly absorbed by water [132, 134]. Overall, the derived band ratios suggest that the spatially averaged spectra (average of all pixels contained in each plot) contain unique and useful information for estimating the abundance of each growth form present.

Conclusions

A new method using a UAV-based hyperspectral imager for quantitatively characterizing algal blooms in narrow, shallow rivers has been presented. The results from this analysis suggest that spectral band ratios are a promising method for estimating chl *a* and phycocyanin standing crops contained within, near, and on the surface of blooms of the filamentous nuisance algae, *Cladophora glomerata*, growing in the UCFR. The brute-force analysis used to generate spectral band ratios for estimating chl *a* and phycocyanin standing crops from filamentous, epiphytic, and epilithic growth forms often selected wavelengths near known absorption bands of the pigments [129, 130]. The absorption bands represent the spectral nature of primary production carried out both by eukaryotic algae and cyanobacteria and reflects their

structural and functional character derived from evolutionary selection for light harvesting and carbon fixation. These features are robust and consistent over the time frames applicable to ecological investigations addressing algal abundance; however, the data presented here are limited to the species and conditions present along the UCFR in 2021. Additionally, the scope of this work is limited to the exploration of linear relationships between spectral band ratios and pigment abundance. In cases where the linear fit is not highly correlated (i.e., low R^2 values), a nonlinear fit may explain more of the observed variance in pigment abundance, but also raises the risk of over-fitting. Questions still remain about the generalizability of the selected optimal band ratios, meaning that more data must be collected and analyzed, though preliminary analyses show strong agreement across data sets.

The brute-force method converged on two optimal band ratios when estimating fila/epip chl *a* standing crops across data sets, selecting a ratio of 684/674 nm for both the combined and Gold Creek data sets and a ratio of 554/536 nm at the Bear Gulch site. The brute-force analysis appeared to identify other important wavelengths for estimating chl *a* for algae in different phenological states, selecting shorter wavelengths (554, 536 nm) when analyzing *Cladophora* blooms with significant diatom coverage, such as those seen at the Bear Gulch site. The method also illuminates regions sensitive to phycocyanin standing crops, often selecting wavelengths over 700 nm (Table 3.3). Interestingly, the brute-force analysis selected similar band ratios for epilithic chl *a* and phycocyanin abundance, indicating that these signals may be conflated. That is, the presence of one pigment may be related to the presence (or absence) of the other.

The brute-force analysis outlined here represents the early framework for developing a network of low-cost sensors to estimate pigment standing crops over large stretches of riverine systems affected by algal bloom activity. The adoption of a sensor network based on the proposed method would allow for more frequent data collection, while also providing information on a spatial scale large enough to understand ecological factors such as algal

metabolism and nutrient uptake. While current measurement methods are constrained to small spatial scales (cm^2–m^2), the proposed method allows for much broader spatial coverage (100–1000 m), allowing for greater understanding of the relationship between RAB development and functional attributes related to water quality. Expanding access to RAB detection and monitoring techniques would help maintain water quality standards, especially in areas where standards are based on algal biomass or nutrient standards, or lead to remediation in contaminated waterways.

In future work, a larger data set collected in the UCFR in 2022 will be analyzed and new data will be collected in the Gallatin River in Montana to begin assessing how well the results presented here apply to other times, conditions, and rivers. These additional data also will aid in the design of multispectral imaging systems capable of RAB detection, along with the development of algal classification maps leading to percent cover estimates.

To strengthen the applicability of the methods outlined above, maps of the spatial distribution of algae growth had to be developed. These maps could not only provide additional information, such as percent algal cover, but could also be used to augment pigment abundance estimation, where only pixels identified as algae would be analyzed. The following chapter addresses the second objective of this work by outlining the development and use of such spatial distribution maps.

RIVER ALGAE MAPPING USING HYPERSPECTRAL IMAGERY FROM UAVS

Contribution of Authors and Co-Authors

Manuscript in following chapter

Author: Riley D. Logan

Contributions: Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing—original draft

Author: Joseph A. Shaw

Contributions: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing—review and editing

Manuscript Information

Riley D. Logan and Joseph A. Shaw

Journal of Applied Remote Sensing

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
_X_ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

SPIE
Submitted: 19 January 2024
Published: Type date the manuscript will appear here
[Issue and Volume]
[Type DOI, if available]

Abstract

The increasing prevalence of harmful and nuisance algal blooms in both oceanic and inland waters has led to water quality monitoring programs based on the presence and abundance of algae. Large blooms of the nuisance filamentous algae, *Cladophora glomerata*, have become common in the waters of the Upper Clark Fork River in western Montana. To aid in the understanding of algal growth dynamics, UAV-based hyperspectral images were gathered at three field sites along the length of the river throughout the growing season of 2021. Select regions within images covering the spectral range of 400-1000 nm were labeled based on a combination of professional judgment and spectral profiles and used to train a random forest classifier to identify benthic algal growth across several classes, including *Cladophora*, benthic growth other than *Cladophora*, and areas below a visually detectable threshold of benthic growth. After classification, images were stitched together to produce spatial distribution maps of each river reach while also calculating the average percent cover for each reach, with an accuracy of approximately 99%. Results of this analysis showed strong variability across each reach, both temporally (up to 40%) and spatially (up to 46%), indicating that traditional measurement techniques, such as spot sampling, may create misleading or inaccurate results that may be avoided by using recurrent UAV-based imaging.

Introduction

The presence of harmful and nuisance algal blooms is becoming an increasing concern in oceans, lakes, and rivers, where widespread growth of algae has been shown to be indicative of deteriorated water quality and ecosystem health [135, 136]. Though submerged aquatic vegetation (SAV) and macroalgae, such as the nuisance filamentous algae *Cladophora glomerata*, are commonly found in healthy aquatic ecosystems, excessive growth is indicative of eutrophication [15, 137, 138]. *Cladophora glomerata* is one of over 400 species of *Cladophora*,

a genus that is considered one of the most abundant alga in alkaline rivers worldwide and one of the most important harmful filamentous alga in inland waters [139]. The Upper Clark Fork River (UCFR) in western Montana has a long history of abundant *Cladophora glomerata* (*Cladophora*) growth, where naturally high levels of phosphorus and an influx of wastewater due to anthropogenic activity along its banks have led to nutrient enrichment, creating advantageous conditions for large blooms [1, 79, 80, 140].

Monitoring of water quality can be based on chemical or biological analyses, both of which provide valuable insight into nutrient enrichment and stress in fluvial systems [141]. Historically, water quality standards in Montana have been based on either visual standards or numeric levels of chemical properties, such as benthic algal standing crops, measured as the mass of the pigment chlorophyll *a* (chl *a*) per square meter [142, 143], however, these assessments methods have limitations. Visual standards lack quantitative biological analysis while numeric standards are based on *in-situ* sampling that is limited in the spatial distribution and coverage required to accurately assess large-scale algal growth [144, 145].

Water quality monitoring along the UCFR has a long history, with major impairment recorded as early as 1908 [2]. Beginning in the 1970s, large blooms of *Cladophora* became a regular occurrence between the UCFR headwaters near Butte, MT and approximately 193 river kilometers downstream, near Missoula, MT [1, 2]. *Cladophora* growth in the UCFR is spatially patchy and dependent on stream morphology and flow [69], indicating that water quality measurements that include assessment of algal standing crops must take this spatial variability into account; however, laboratory-based analysis of algal samples does not capture information on the spatial extent of algal growth [15]. Current methods for determining algal coverage rely on direct visual assessment based on qualitative estimates of percent algal cover, predominant algal color, and growth condition, measured at several points along a transect [146]. Transect methodologies require little training and time to perform, but only capture small regions, are not often quantitatively rigorous, and are commonly site-specific. Some

sampling-based methods are designed to be quantitative, such as the USGS Richest Targeted Habitat Method [147, 148], but are resource intensive and have limited spatial representation.

Optical remote sensing systems have been used to assess the spatial coverage and biological metrics of algal blooms in large bodies of water for decades and is ongoing [96, 149, 150]; however, most rely on satellite imagery with coarse spatial resolution (10 m - km), inhibiting their use on anything other than the largest rivers or lakes [151, 152, 153, 154]. More recently, the use of unoccupied aerial vehicle (UAV) systems has grown in popularity due to their flexibility in revisit times and flight planning, ability to carry a variety of imaging systems, and higher spatial resolution [155, 156]. UAV-based systems have captured imagery using uncalibrated red-green-blue (RGB) imagers [69, 157] and multispectral systems with near-infrared sensitivity [158, 159, 160] with promising results for identifying algal blooms. Methods to classify imagery and estimate algal coverage typically rely on spectral information that is analyzed using a variety of methods, such as spectral angle mapping [69], statistical approaches [158], and machine learning techniques [157, 160, 161]. Though hyperspectral satellite imagery has shown promise in differentiating between cyanobacteria genera in large bodies of inland water using spectral analysis [153], there is little work exploring the efficacy of UAV-based hyperspectral imagery to map the spatial distribution of benthic macroalgae and SAV in mid-sized inland waterways.

This paper builds on a recently published method in which a UAV-mounted hyperspectral imager was used to estimate river algae pigment abundance [4]. A random forest classification method was adopted to explore all available wavelengths, but we also report elsewhere progress toward a low-cost multispectral imager that will rely on more traditional spectral classification methods [162]. Here, the spatial distribution of river algae is mapped using UAV-based hyperspectral imagery collected across the 2021 growing season, with the following objectives:

1. Differentiate between *Cladophora*, other forms of benthic growth (non-*Cladophora*), and

areas below a visually detectable threshold of benthic growth using a custom-trained random forest classification model.

2. Generate georectified classification maps from UAV-based hyperspectral imagery.

3. Estimate average percent cover at three field sites along the UCFR sampled between June and September 2021.

Materials and Methods

Study Sites

The UCFR is formed at the junction of Warm Springs and Silver Bow Creeks in southwestern Montana and flows nearly 193 kilometers northwest to its confluence with the Blackfoot River. Three field sites were selected along the UCFR to capture algal growth characteristics near the headwaters (Deer Lodge, 46.38°N, 112.74°W), midway through its length (Gold Creek, 46.59°N, 112.93°W), and near its confluence with the Blackfoot River (Bear Gulch, 46.71°N, 113.33°W) to capture possible influence of nutrient transport and effects of input drainages (Fig. 4.1). Algal bloom development along the UCFR follows a complex seasonal pattern, influenced by many factors such as net primary productivity, flow rates, temperature, and canopy shading and is typically initiated after snowmelt-fed runoff hydraulically scours benthic material in May or June [118, 140], informing data collection periods.

Data Collection

Image data were collected using a Resonon Pika L Airborne Hyperspectral Imaging System (Resonon Inc., Bozeman, MT, USA) mounted on a DJI Matrice 600 Pro hexacopter (DJI, Shenzhen, China) using a DJI Ronin-MX gimbal between 30 June 2021 and 16 September 2021 on an approximate biweekly schedule. The Pika L Hyperspectral Imager has a spectral

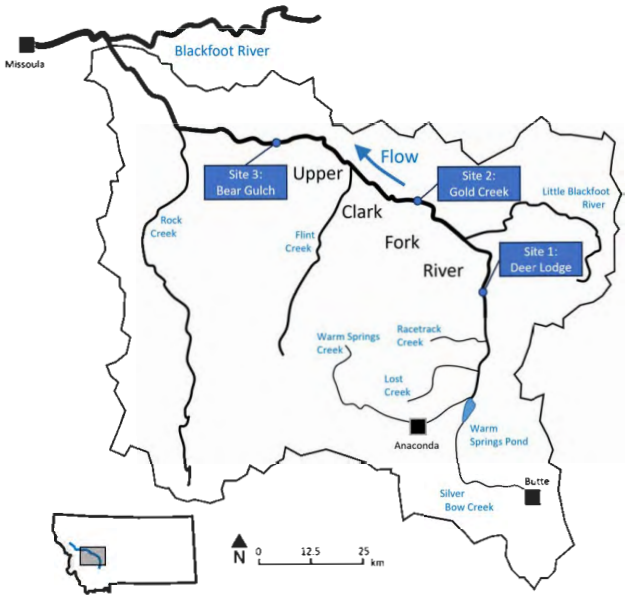


Figure 4.1: Study sites along the Upper Clark Fork River, numbered from upstream to downstream at Deer Lodge, Gold Creek, and Bear Gulch. Data collection sites were selected to capture representative algal growth along the full 120-mile reach (adapted from [4]).

range of approximately 387–1023 nm and nominal spectral resolution of 2.1 nm, producing 300 spectral channels per pixel. Images were captured using a 17-mm objective lens with a 17.6° across-track full-angle field of view directed nadir, resulting in an across-track ground swath of approximately 37 m when flown at a typical height of 120 m. Flight paths were established to follow the midline of each river segment for approximately 1 km and held constant throughout the data collection period to ensure the same reach was evaluated throughout the study. All UAV flights were planned and conducted by, or under the supervision of, a pilot certified through the Federal Aviation Administration (FAA) Part 107 licensure program.

Image Preprocessing

Image data were converted from raw digital number to reflectance using a cosine-corrected Ocean Insight Flame VIS-NIR spectrometer (Ocean Insight, Orlando, FL, USA) mounted to the top chassis of the UAV and calibrated to measure downwelling spectral irradiance

from 400–1000 nm with a nominal spectral resolution of 1.34 nm. Downwelling irradiance spectra were measured throughout each UAV flight, with measurements taken whenever a new hyperspectral image was recorded by the imager, then interpolated to match the spectral channels on the Pika L. Images were converted to reflectance using

$$R_{image}(\lambda) = \frac{L_{upwelling}(\lambda)\pi}{E_{downwelling}(\lambda)}, \tag{4.1}$$

where all terms are a function of wavelength (λ) and R_{image} is the reflectance of the image, $L_{upwelling}$ is the upwelling radiance measured by the Pika L, π accounts for assumed Lambertian reflectors in the scene, and $E_{downwelling}$ is the downwelling irradiance measured by the spectrometer. Once converted to reflectance, images were smoothed using a Savitzky-Golay filter with a window length of 13 and polynomial order of 3 [163].

Due to strong water absorption in the near-infrared, wavelengths beyond 850 nm were removed, leaving 221 spectral channels in each image; however, the reflectance of the 850 nm channel was used as a method to identify and remove both sun glint and bank pixels from the image. In order to simplify the classification task, pixels containing material other than water (e.g., floating debris, bank vegetation), or pixels containing shadows or sun glint, needed to be removed from images prior to classification. Using spectral profiles and professional judgement, a reflectance greater than 5% at 850 nm was found to be an effective means of separating pixels containing submerged objects from pixels containing sun glint or non-submerged material, while a reflectance less than 1.5% at 550 nm was found to isolate shadowed water pixels for removal.

Hyperspectral image data were spatially corrected using Resonon Spectronon Pro software to align the imagery with cardinal directions and latitude and longitude, while also accounting for spatial deviations in imagery caused by following bends in the river. This spatial correction was performed using an automated georectification tool that used data

from the onboard GPS and inertial measurement unit data, along with ground elevation, flight height, and focal length of the imaging system.

Classification Algorithm

Training. Image classification was based on a supervised random forest classification model implemented using Scikit-Learn. [164] The model was trained and validated on 2,157,996 labeled pixels from hyperspectral imagery collected along the UCFR during 2021-2022. Images were selected between June and September to capture algal growth under a variety of conditions throughout the summer growing season, including changes in physical characteristics (e.g., flow rates, water temperature) and biological characteristics (e.g., phenological state, biomass accrual). Representative pixels from 13 hyperspectral images were labeled into three classes: visually identified as *Cladophora* (n = 759,873 pixels), benthic growth above a visibly detectable threshold identified as non-*Cladophora* (n = 705,483), and benthic growth below a visually detectable threshold (n = 692,640). Pixels were labeled manually using professional judgment based on RGB composite color and spectral profiles (Figure 4.2). Pixels that appeared to contain *Cladophora* were identified by their strong reflectance in green and near-infrared wavelengths (modulated by water depth) giving them a bright green appearance while healthy, with yellow-shifting during senescence. Pixels with benthic growth visually identified as non-*Cladophora* were identified by decreased and broadened reflectance at green wavelengths, while still maintaining elevated reflectance at near-infrared wavelengths giving them a dark brown appearance. Benthic growth below a visually detectable threshold was identified by a relatively featureless spectrum and light gray or tan appearance. Benthic growth identified as *Cladophora* was often highly separable from the other classes; however, non-*Cladophora* and pixels below a visible threshold of benthic growth were spectrally similar at times, likely due to trace amounts of algal growth across the benthos, which may have introduced error in the labeling process. Labeled pixels

were randomly split into 30% testing and 70% training subsets using the *train_test_split()* function in Scikit-Learn [164]. Hyper parameters were tuned using a randomized search across the training data subset with 5-fold cross-validation across 50 iterations, leading to the parameters shown in Table 4.1.

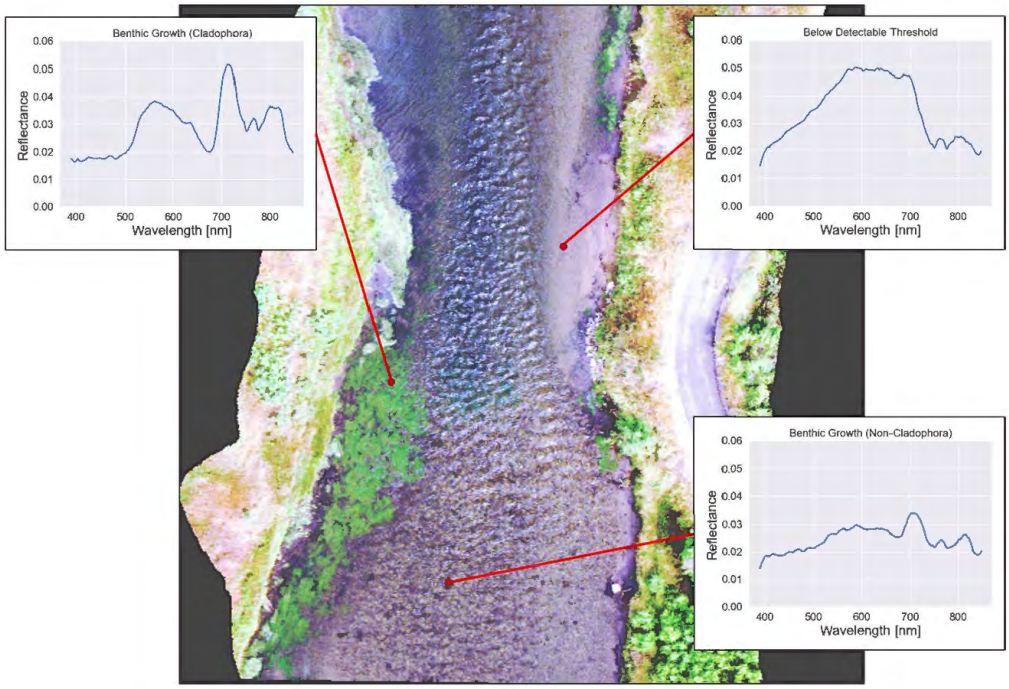


Figure 4.2: Example spectra for the three labeled classes, including pixels containing benthic growth visually identified as *Cladophora* (upper left), non-*Cladophora* (bottom right), and below a detectable threshold (upper right).

The 5-fold cross-validation on the tuned random forest model produced a mean training accuracy of 99.9% and a mean validation accuracy of 99.6% across all folds. The tuned classifier showed little confusion when predicting classes, with the largest error attributed to incorrect classification between non-*Cladophora* and areas below a visually detectable threshold of benthic growth (Fig. 4.3a). As full-spectrum images were used during classification (i.e., without dimensionality reduction), the importance of the ten spectral bands with the highest predictive power were analyzed using the mean decrease in impurity (Fig. 4.3b).

Two of the top ten bands matched spectral channels used to predict chl *a* abundance

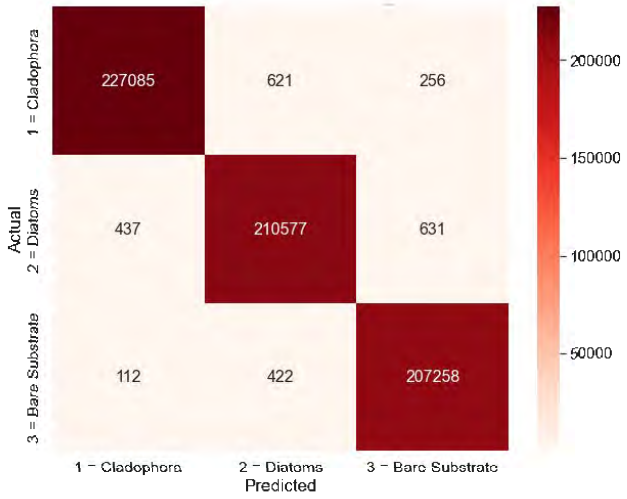
Parameter	Value
Number of Trees	75
Loss Function	Gini Impurity
Minimum Samples to Split a Node	10
Minimum Samples for Leaf Node	2
Maximum Number of Features	Square Root
Maximum Depth of Tree	50

Table 4.1: Optimal random forest classifier hyper parameters selected after randomized search with 5-fold cross-validation across 50 iterations.

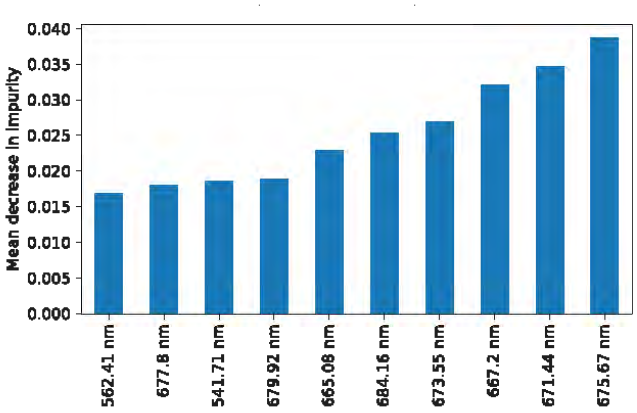
in the UCFR [4] and surrounded a known chl *a* absorption line near 675 nm [129, 130]. In addition, two green wavelengths were identified, likely because of their utility in differentiating between the typical bright green *Cladophora* and areas of benthic growth other than *Cladophora* which are generally dark green or brown.

Though the classifier showed promising early results, the random train/test split was based on pixel-by-pixel assignment, therefore, the accuracy metrics may have been biased high due to spatial autocorrelation (i.e., adjacent pixels that are spectrally related present in both training and validation sets). To further assess the accuracy of the method, two labeled hyperspectral images were withheld from the training and validation process and used only to test accuracy. The trained algorithm was run on these unseen images and the accuracy of the labeled regions was reported per class in terms of precision, recall, and F1-score (Table 4.2) along with confusion matrices (Figure 4.4). The classification algorithm showed good performance, with an average F1-score of 99% across the 122,075 labeled pixels.

Percent Cover Estimates. After classification and preprocessing, including identification and removal of pixels containing riverbank material, all labeled pixels contained only



(a) Confusion matrix of the tuned random forest model, showing a summary of the predictions made by the classifier. The largest classification error was between the non-*Cladophora* and areas below a visually detectable threshold of benthic growth (bare substrate) classes (though this error was less than 1%).



(b) Top ten wavelengths for making class predictions extracted using mean decrease in impurity. The most important features (right) surround known chl *a* absorption lines [129, 130] and match salient wavelengths identified in previous work [4].

Figure 4.3: Training and validation accuracy was very high (>99%) assessed across 5-fold cross-validation (a); the confusion matrix shows little confusion between classes (b); and top spectral features hovered around green wavelengths and surrounded a known chl *a* absorption line.

submerged material, from which percent cover estimates could be obtained. Percent cover was calculated for both *Cladophora* and non-*Cladophora* by dividing the number of pixels

	Class	Precision	Recall	F1-Score	Number of Samples
Validation Image 1	*Cladophora*	1.00	1.00	1.00	37,501
	Non-*Cladophora*	1.00	0.97	0.98	35,276
	Below Threshold	0.00	0.00	0.00	0
Validation Image 2	*Cladophora*	1.00	0.99	1.00	16,862
	Non-*Cladophora*	0.98	0.97	0.98	7,846
	Below Threshold	0.99	0.99	0.99	24,590

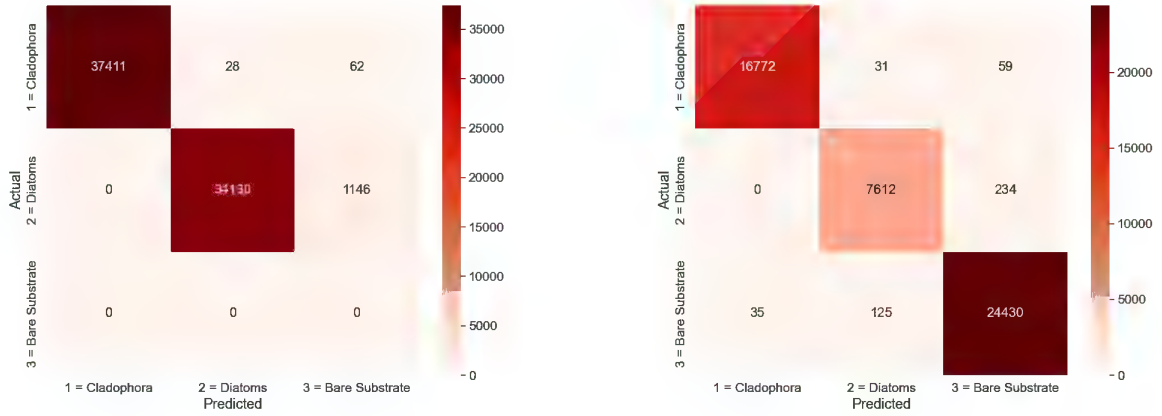
Table 4.2: Accuracy metrics of the random forest classification algorithms after testing on unseen validation images, reported as precision, recall, F1-Score, and the total number of samples in each class. Validation Image 1 contained no examples of regions below the visually detectable threshold for benthic growth so no metrics are included.

classified as either class by the total number of pixels identified as being within the water (i.e., pixels with a reflectance less than 5% at 850 nm). The resulting value gave a percent cover estimate for each classified data cube, spanning roughly 80 m. The average percent cover for each UAV flight was calculated as the mean of all classified data cubes collected along each 1 km river reach.

Results and Discussion

UAV flights were conducted between June and September of 2021 (Tab. 4.3) at the three field sites discussed in Section 4. Individual, preprocessed hyperspectral data cubes captured during each flight were used as inputs to the trained random forest classifier. Each classified image contained predictions for every pixel in the data cube, where green pixels represent *Cladophora*, blue represent benthic growth identified as non-*Cladophora*, gray represent areas below a visually detectable threshold of benthic growth, and black represent non-submerged or shadowed pixels removed during preprocessing (Fig. 4.5).

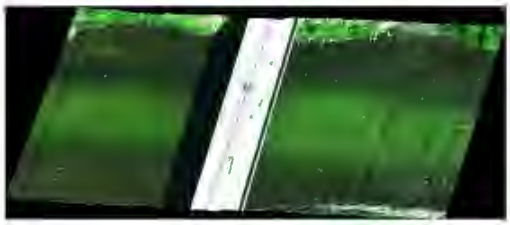
After classification, the spatially corrected images gathered at each site were manually



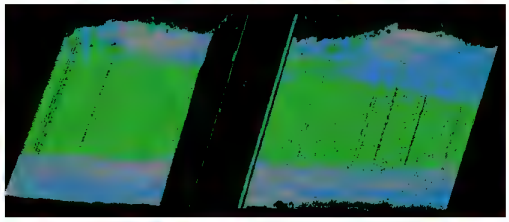
(a) Confusion matrix of the tuned random forest model tested on Validation Image 1. Confusion between *Cladophora* and non-*Cladophora* pixels was quite low, though there was confusion between non-*Cladophora* and areas below the detectable threshold.

(b) Confusion matrix of the tuned random forest model tested on Validation Image 2. As in Validation Image 1, confusion between *Cladophora* and non-*Cladophora* pixels was quite low, though there was confusion between non-*Cladophora* and areas below the detectable threshold.

Figure 4.4: Confusion matrices for Validation Images.



(a) RGB image of a single data cube at Gold Creek on 1 July 2021 showing large central bloom of *Cladophora*, regions of non-*Cladophora* near the banks, bridge across the center of the image, and associated shadow.



(b) Classified image of Gold Creek 1 July 2021 showing large central bloom of *Cladophora* in green, regions of non-*Cladophora* near the banks in red, areas below a visually detectable threshold of benthic growth in gray, and non-submerged objects in black. Note the removal of the bridge and associated shadow.

Figure 4.5: Hyperspectral data cube before and after classification using the trained random forest model. The data cube covers approximately 80 m and shows a large bloom of *Cladophora* in the center of the river, with percent cover estimations of approximately 55% and 29% for *Cladophora* and non-*Cladophora*, respectively.

Date (2021)	**Field Site**					
	Deer Lodge		Gold Creek		Bear Gulch	
	Clado	Non-Clado	Clado	Non-Clado	Clado	Non-Clado
30 June	41%	42%	–	–	–	–
1 July	–	–	47%	30%	61%	21%
8 July	19%	70%	–	–	–	–
9 July	–	–	43%	38%	35%	48%
14 July	20%	72%	–	–	–	–
15 July	–	–	61%	28%	30%	61%
21 July	36%	41%	–	–	–	–
22 July	–	–	21%	64%	3%	27%
28 July	–	–	26%	61%	16%	29%
12 August	17%	72%	–	–	–	–
13 August	–	–	–	–	21%	49%
17 August*	–	–	41%	45%	–	–
25 August	15%	69%	–	–	–	–
1 September	31%	44%	–	–	–	–
2 September	–	–	44%	34%	40%	27%
16 September	–	–	53%	35%	–	–
Season Average:	**26%**	**58%**	**42%**	**42%**	**29%**	**37%**

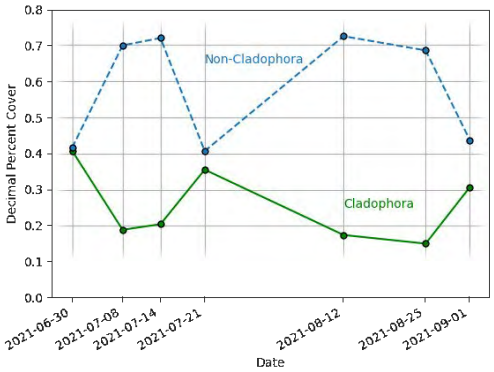
Table 4.3: Dates, locations, and percent cover estimates of UAV flights during the 2021 data collection period. The average percent cover (bottom) indicates that non-*Cladophora* benthic growth generally dominated across the 2021 growing season, with the exception of Gold Creek, where percent cover was approximately equal. *The percent cover estimates from 17 August at Gold Creek are based on a 80-m reach at Gold Creek due to limited data collection during flight.

stitched together and the average percent cover was calculated across the entirety of the reach. Percent cover estimates generally showed higher levels of non-*Cladophora* across the growing season, with an average percent *Cladophora* cover of 26%, 42%, and 29% and average percent non-*Cladophora* cover of 58%, 42%, and 37% for Deer Lodge, Gold Creek, and Bear Gulch, respectively (Tab. 4.3). *Cladophora* coverage estimates generally declined at all sites until late July, with the exception of significant growth at Gold Creek on 15 July, indicating that maximum algal cover occurred prior to UAV flights (Fig. 4.6). *Cladophora* growth showed signs of resurgence in early August at Gold Creek and Bear Gulch, with percent cover estimates gradually increasing until the end of the sampling period. Deer Lodge also showed signs of a resurgence event, with percent cover increasing during the final UAV flight in early September. Non-*Cladophora* benthic growth showed an inverse relationship with *Cladophora* at Deer Lodge and Gold Creek, with more variability observed at Bear Gulch.

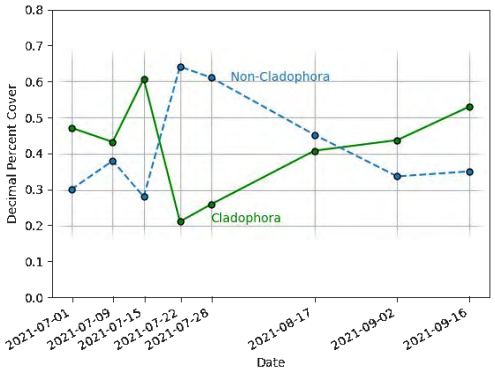
Surprisingly, the seasonal average of non-*Cladophora* coverage was higher than *Cladophora* coverage at Deer Lodge and Bear Gulch, while being equal at Gold Creek. Non-*Cladophora* coverage tended to decrease moving farther downstream from the headwaters of the UCFR, dropping from a seasonal average of 58% at the most upstream site to 37% at the most downstream site.

As noted in previous work on the UCFR, algal growth was spatially variable within each reach [69], ranging from nearly total coverage to areas with little to no algal growth (Fig. 4.7a). Estimating percent cover in 80-m steps along the full 1-km reach the Bear Gulch field site on 2 September 2021 revealed that *Cladophora* and other forms of benthic growth varied by up to 42% and 46%, respectively (Fig. 4.7b). Given this spatial variability, existing measurement methods that rely on spot sampling along a small section of a reach may not capture the true character of algal activity in the river.

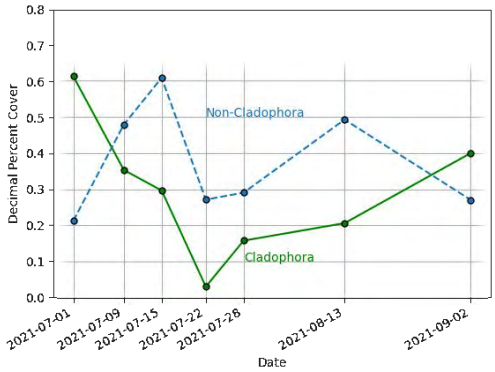
Many factors influence *Cladophora* growth rates; however, growth has been shown to be related to stream velocity [165], which may explain the tendency of blooms to form along



(a) Deer Lodge



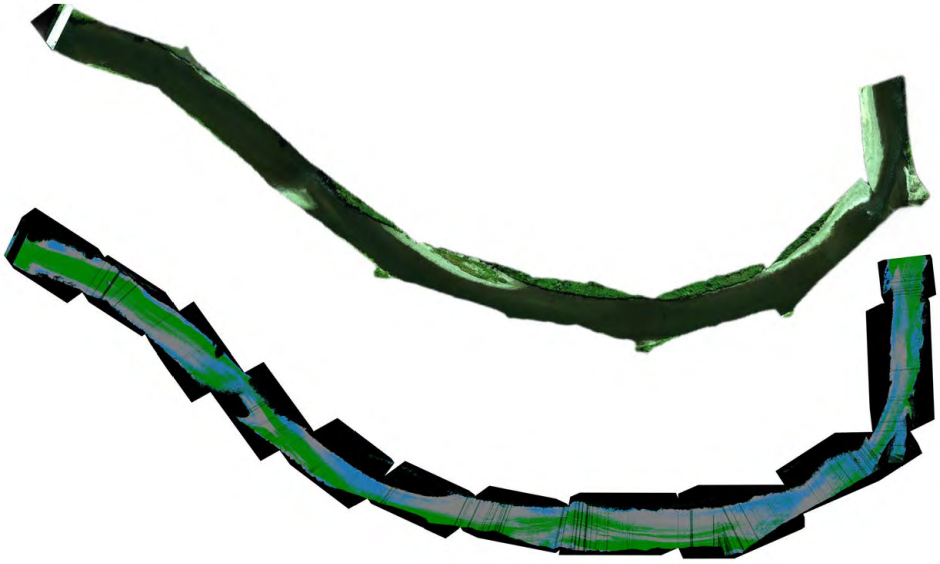
(b) Gold Creek



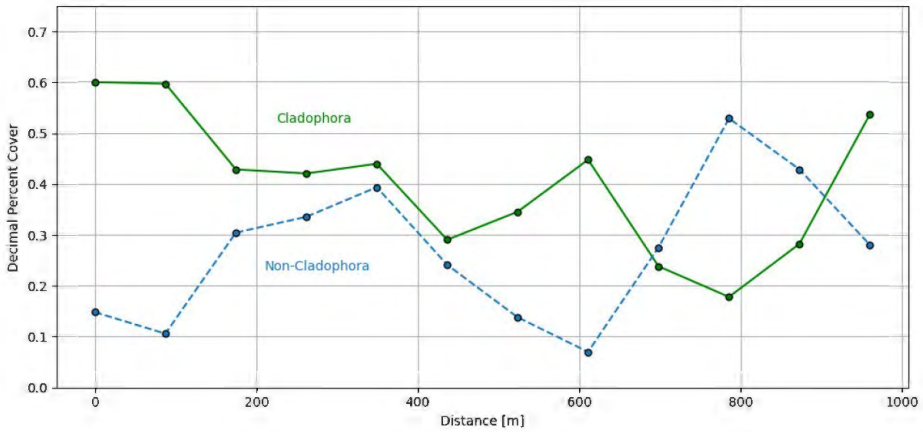
(c) Bear Gulch

Figure 4.6: Decimal percent cover estimates for Deer Lodge (a), Gold Creek (b), and Bear Gulch (c) across the summer growing season of 2021. *Cladophora* shown in green and non-*Cladophora* shown in dashed blue.

outer bends in the river, though more work is required to understand the exact relationship between stream morphology and algal growth [137]. In addition to spatially variable growth, rapid temporal shifts were also noted across all sites and sampling periods, with *Cladophora* cover fluctuating by 40% between 15 and 22 July at Gold Creek and non-*Cladophora* coverage following similar patterns.



(a) Example of a full-reach image at Bear Gulch on 2 September 2021, shown in RGB (top), and after classification (bottom). The reach spans roughly 1 km with an average percent *Cladophora* cover (green) of 40% and percent non-*Cladophora* (blue) cover of 27% with no detectable benthic growth shown in grey.



(b) Percent cover as a function of distance along river. Values were averaged across each hyperspectral data cube or approximately 80 m. Percent cover varies by approximately 42% for Cladophora and 46% for non-*Cladophora* along the 1 km reach, with the rest of the reach classified as not containing growth above a visually identifiable threshold, suggesting that spot measurement methods may create misleading results.

Figure 4.7: RBG and classified image of the full river reach at Bear Gulch on 2 September 2021 (a) and connected scatter plot of percent cover as a function of downstream distance along the reach (b).

Conclusion

The significant spatial and temporal variability of *Cladophora* and other forms of benthic growth in the UCFR suggests that traditional methods of measuring percent cover may provide misleading results when applied to short reaches. Though multispectral imaging systems, such as RGB imagers, have been used with promising results [69], the method presented here suggest that UAV-based hyperspectral imaging coupled with machine learning classification methods provides improved accuracy for capturing the spatial and temporal variability through continuous, high-spatial-resolution spatial distribution maps while also easily isolating water pixels and removing shadowed pixels. Using a custom-trained random forest classifier, georectified spatial distribution maps were created for three field sites along the UCFR between June and September of 2021, highlighting the spatial and temporal variation of algal growth in the UCFR and an inverse relationship between *Cladophora* and other types of benthic growth. Though the method showed promising results, with training and validation accuracies greater than 99% and an average F1-score of 99%, there are limitations. First, ground control points were not used during the georectification process, introducing possible errors in spatial accuracy. Second, the removal of bank and shadow pixels was not field-verified, again possibly introducing small errors in the estimation of percent cover. Finally, though the presence of *Cladophora* and other forms of benthic growth were verified in the field, ground truth samples were not collected to assess the accuracy of the classification method versus *in-situ* samples and the role of depth in identifying benthic growth was not explored.

The ability to classify algal growth along continuous river reaches allows for better assessment of benthic algal cover, but is also a step forward in species identification by differentiating between *Cladophora* and benthic growth other than *Cladophora*. Separation by growth form allows for further studies of habitat preference, seasonal shifts in dominant

growth forms, and a better understanding of eutrophication over larger scales. In future work, the classification methods presented here will be combined with techniques for estimating algal biomass, as chl *a* abundance, to provide georectified classification and biomass maps and the salient spectral bands are being used to develop a low-cost, compact multispectral algae imager [162].

Parts of Manuscript

Acknowledgments

This material is based upon work supported in part by the National Science Foundation EPSCoR Cooperative Agreement OIA-1757351. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Code, Data, and Materials Availability

The software and data presented in this article are publicly available at: Logan, R. (2024). River Algae Mapping Using Hyperspectral Imagery from UAVs, HydroShare, http://www.hydroshare.org/resource/840938d8c4004e22999fe8a2eb89523d

Acknowledgments

The authors would like to extend their thanks to H. M. Valett and Rafael Feijó-Lima for their extensive help with understanding and interpreting the ecology of the Upper Clark Fork River and its aquatic ecosystem.

THE DEVELOPMENT AND USE OF SPAMM

Contribution of Authors and Co-Authors

Manuscript in following chapter

Author: Riley D. Logan

Contributions: Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing—original draft

Author: Joseph A. Shaw

Contributions: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing—review and editing

Manuscript Information

Riley D. Logan and Joseph A. Shaw

Status of Manuscript:
__X__ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

Submitted: Published: [Issue and Volume] [Type DOI, if available]

Introduction

Nearly 25 years before developing national water quality standards as part of the Clean Water Act (CWA) of 1972, the United States passed the Federal Water Pollution Control Act (FWPCA) to confront the growing degradation of inland waters caused by urban development and major industrial growth [166, 167]. The FWPCA was intended to address impairment of "all rivers, lakes and other waters that flow across, or form a part of, State boundaries"[168]; however, over 75 years later, our nation's waters still suffer from clear signs of biological stressors, such as eutrophication, which has led to large algal blooms in coastal waters [169, 170] and freshwater systems [171, 172, 173]. Even in less populated areas, algal blooms have become a growing issue. For example, the Upper Clark Fork River (UCFR) in southwestern Montana experiences large and recurrent blooms of the nuisance filamentous algae, *Cladophora glomerata*, that have been documented since the mid 1970s [1].

Water quality standards in Montana have a convoluted history, with a combination of descriptive, or narrative, standards and more robust numeric standards [142, 143]. Numeric water quality standards for nutrient concentrations affecting algal growth (i.e., total nitrogen and phosphorus levels) and the area density of algal biomass were proposed along the UCFR no later than 1996 [5] and inform the current state standards of a summer mean area density of 100 mg/m^2 with an upper limit of 150 mg/m^2 [143]. Although numeric nutrient and biomass standards are objective and testable, visual surveys of percent algal cover (i.e., the percentage of bottom substrate covered by algal growth) have also been identified as an important factor in water quality monitoring [68]. When algal levels grow in excess (i.e., above regulatory thresholds), indigenous flora and fauna become stressed [8, 137, 174], dissolved oxygen and pH levels are altered [6, 7], general water quality is reduced [6, 9, 10, 11], and recreational use is disrupted [6, 9, 11]. Current methods for assessing algal biomass and coverage in rivers typically involve transect measurements and spot sampling, though these

methods are time intensive, costly, and can be difficult to extrapolate to full-reach conditions [5].

Here we introduce methods for creating spatial distributions maps of algae biomass and estimating algal biomass based on UAV-based hyperspectral image data collected along the UCFR. The methods are packaged in a Python software, which we call Spectral Processing for Algal Monitoring and Mapping (SPAMM), that builds on our previous work in estimating river algae pigment abundance [4] and creating spatial distribution maps in the UCFR [175]. The objectives of this work are as follows:

1. Present a method for improving our algal biomass estimation based on machine learning regression algorithms and a larger ground-truth data set collected along the UCFR throughout 2022.

2. Introduce a single software capable of creating spatial algae distribution maps locating and distinguishing between *Cladophora*, diatoms and submerged aquatic vegetation (SAV), and bare substrate, while also producing algal biomass abundance maps based on identified algae locations.

3. Use SPAMM to assess average percent cover and algal biomass along the UCFR at select field sites across the 2020, 2021, and 2022 summer growing seasons.

4. Assess the results of SPAMM as a tool for monitoring water quality in accordance with Montana standards.

Materials and Methods

An Overview of SPAMM

SPAMM was developed using geolocated UAV-based hyperspectral reflectance imagery and *in-situ* ground truth measurements collected across the summer algal growing seasons of

2020-2022 to make percent cover and algal abundance predictions. Therefore, SPAMM requires georeferenced hyperspectral images across the spectral range of 400-1000 nm converted to physical units of reflectance as inputs. The software then builds a spatial algae distribution map by identifying pixels that contain *Cladophora*, diatoms or SAV, and bare substrate, then generating percent cover estimates for *Cladophora* and diatoms/SAV, the details of which are discussed elsewhere [175]. At all located algae pixel locations, a trained machine learning regression algorithm estimates the algal biomass and creates an abundance map as the areal density of chlorophyll *a* (chl *a*) based on ground truth data collected from 2021-2022. The abundance maps calculate mean chl *a* across the imaged reach, report the percentage of algae pixels that fall into numeric abundance categories (see Overview), and the spatial coverage of algal biomass, including regions above Montana water quality thresholds, shown with a histogram. A graphical overview of this process is shown in Figure 5.1. Here, we discuss the development of modified pigment abundance estimation methods, then use SPAMM to assess water quality at three field sites on the UCFR from 2020-2022.

Study Sites

All field sites were selected along the UCFR in southwestern Montana, a cobble-bed river characterized by clear and shallow waters, flowing between the cities of Butte and Missoula (Figure 5.2). Each field site was selected due to a long history of nutrient enrichment and physical conditions, such as direct sun exposure, creating conditions conducive to significant algal growth. Field sites were further broken down into ground truth collection sites and field analysis sites. At ground truth collection sites, *in-situ* measurements of algal abundance were taken as described in the Ground Truth Measurements section, with the measurements used to train regression algorithms to estimate chl *a* abundance. Analysis field sites were used to study particular reaches of the UCFR by estimating their percent cover and algal abundance. Sites were approximately 320 - 800 m in 2020, 1 km in 2021, and 240 m in 2022,

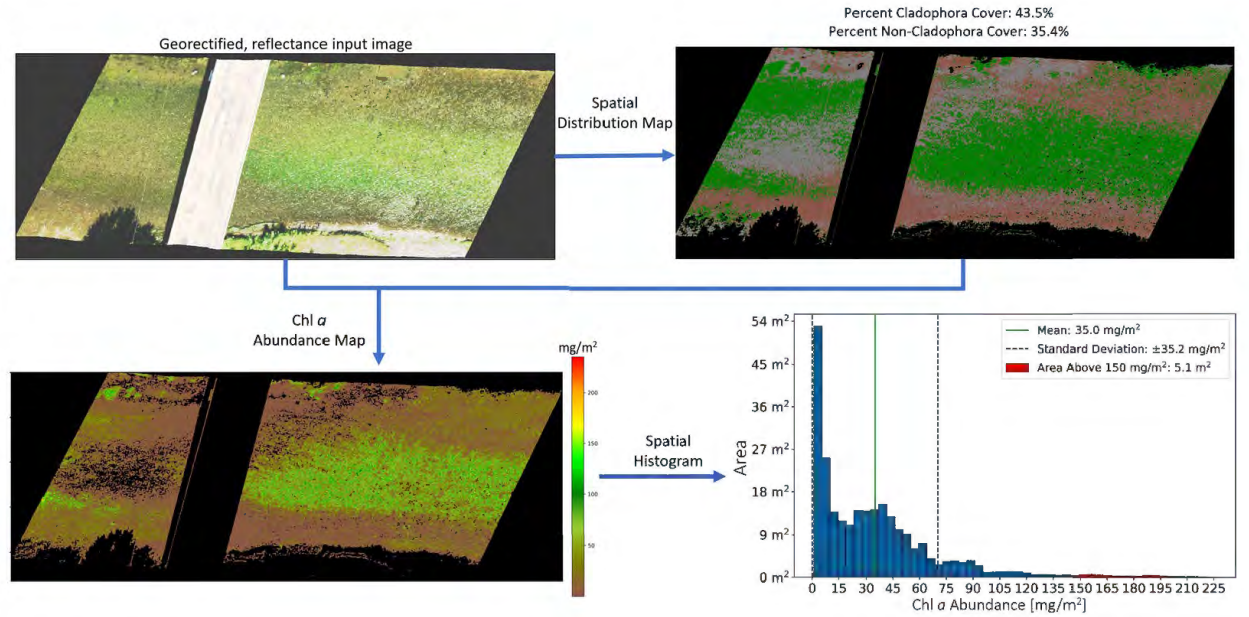


Figure 5.1: Overview of the SPAMM method, starting with a georectified hyperspectral imager converted to units of reflectance, producing a spatial distribution map and chl *a* abundance map, as well as coverage of areal density of chl *a*, mean chl *a* abundance, and area above Montana water quality thresholds.

with portions of ground truth field sites also used as analysis field sites in 2021 and 2022.

Ground Truth Field Sites. Ground truth data collection was split between the end of the 2021 summer algae growing season and the full 2022 summer growing season. In 2021, *in-situ* algae samples were collected at Gold Creek (46.59°N, 112.93°W) and Bear Gulch (46.70°N, 113.43°W), both of which feature similar channels widths of 30 and 28 m, respectively, and had approximate average depths of 35 cm during the sampling period. In 2022, ground truth sampling was expanded to include Deer Lodge (46.38°N, 112.74°W) and Garrison (46.50°N, 112.74°W), while also sampling at Gold Creek and Bear Gulch. The Garrison field site had a channel width of approximately 19 m and average depth of 34 cm during sampling, while the Deer Lodge field site had an approximate channel width of 18 m and average depth of 26.3 cm.

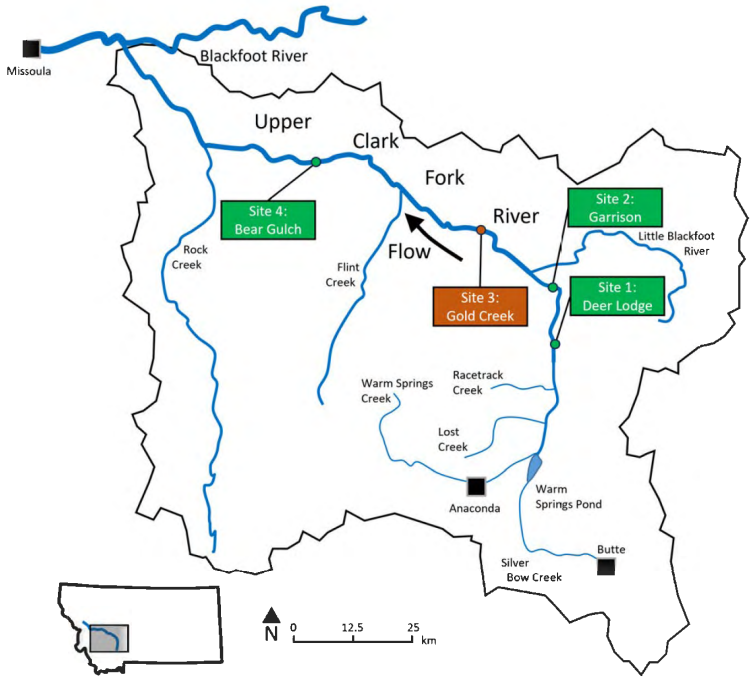


Figure 5.2: The Upper Clark Fork River flowing from its headwaters near Butte, MT to Missoula, MT. Ground truth field sites are shown in orange and ground truth plus analysis field sites are shown in green (adapted from [4]).

Analysis Field Sites. Data collection at our analysis field sites spanned the summer algal growing seasons of 2020-2022 at Deer Lodge, Garrison, and Bear Gulch. The field sites were analyzed using data collected between June and September of each year, with specific flight schedules varying by site and year. In 2020, image data were collected from 12 June through 11 September in one-to-two week periods, resulting in eight flights at Garrison and nine flights at Deer Lodge and Bear Gulch. In 2021, image data were collected from 30 June through 16 September in one-to-two-week periods, providing seven flights at all three sites. In 2022, data collection was limited to ground truth field sites and captured sporadically between 28 June and 15 August, resulting in one flight at Bear Gulch and two flights at Deer Lodge and Garrison. For each site, we used SPAMM to generate percent *Cladophora* cover and percent diatom/SAV cover estimates along with pixel-level estimates of chl *a* abundance.

Ground Truth Measurements

Ground truth data collection methodology followed the process outlined in our previous work on the UCFR [4]. Prior to UAV flights, circular plots built from 112-cm sections of white hose were used to mark *in-situ* measurement locations. Plots were purposefully placed to represent a variety of conditions ranging from bare substrate to full algal cover at each field site (Figure 5.3).

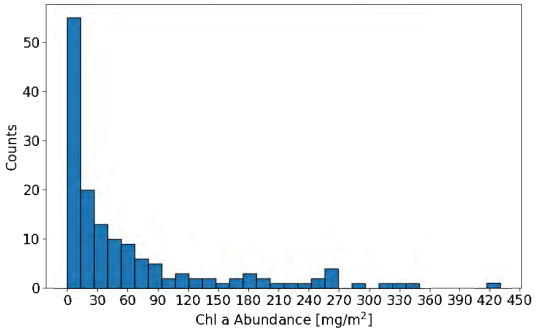


Figure 5.3: Ground truth samples were gathered from circular plots (white circles) intentionally placed throughout each field site to capture a range of ground cover conditions and algal abundances.

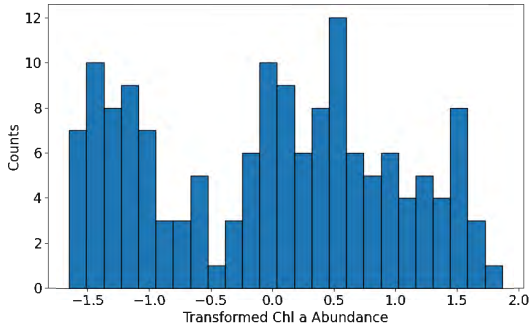
After initial placement, plot locations were adjusted slightly to achieve near-homogeneous cover within each plot to account for spatial averaging of hyperspectral image pixels (see Image Preprocessing). Following UAV flights, water depth at each plot was measured and benthic subtrata contained within each plot were collected. Pigment abundance was measured as an areal density, or standing crop, for three algal growth forms found within each hoop, including filamentous algae, epiphytes, and epilithon. Chl *a* abundance was measured for each growth form in the laboratory following the Ritchie [124] and Hauer[176] methodologies, and then summed across growth forms to obtain the total chl *a* standing crop. In total, 154 *in-situ* samples were collected across the 2021-2022 growing season with the majority (n=120)

gathered in 2022. Five outlying data points were removed, resulting in a ground truth data set of 149 samples.

The resulting ground truth data set distribution was highly skewed, with the majority of samples falling beneath a total chl *a* abundance of approximately 30 mg/m^2 (Figure 5.4a). To account for this high skewness, we transformed the ground-truth data using a Yeo-Johnson power transform [177] to improve the normality of the distribution (Figure 5.4b).



(a) Histogram of ground truth measurements showing a highly skewed data set, with the majority of values falling below approximately 30 mg/m^2.



(b) Histogram of ground truth measurement after Yeo-Johnson power transform, showing an improvement in the normality of the distribution.

Figure 5.4: Ground truth data measurements before (a) and after (b) applying a power transform to improve normality.

Hyperspectral Image Collection

Imaging System. Our image data were collected using a Resonon Pika L Airborne Hyperspectral Imaging System (Resonon Inc., Bozeman, MT, USA) across a spectral range of 387–1023 nm with a nominal spectral resolution of 2.1 nm. Fitted with a 17-mm objective lens, the Pika L had a 17.6° across-track full-angle field of view, resulting in an approximate spatial resolution of 4.2 cm when flown at a typical height of 120 m.

The imaging system included an onboard flight computer that controlled image capture and stored imagery, along with an inertial measurement unit, GPS, and calibrated Ocean

Insight downwelling irradiance sensor (Ocean Insight, Orlando, FL, USA) fitted with a cosine corrector.

UAV. The Pika L imaging system was carried by a DJI Matrice 600 Pro hexacopter (DJI, Shenzhen, China) and mounted on a three-axis DJI Ronin-MX gimbal. The UAV was controlled autonomously via an onboard navigation system coupled with flight plans created using DJI Ground Station Pro software run on a tablet computer, and manually using a 2.4 GHz radio remote controller.

Flight Overview. UAV flights generally followed the same protocol as our previous data collection campaigns on the UCFR [4]. Flights were performed within two hours of solar noon on calm days with little wind to minimize changes in lighting and image distortions introduced by unplanned movement of the UAV, such as tipping or tilting. When coupled with ground-truth measurements, UAV flights were conducted in 160–240 m segments centered over the middle channel of the river at a height of 120 m and ground speed of 2.5 m/s. When collecting imagery at analysis field sites in 2021, the speed and height of flights remained the same, but flight length was extended to approximately 1 km over the center of the reach.

Image Preprocessing

The hyperspectral image data used to train and test the SPAMM algorithms were first converted from digital counts to physical units of reflectance using the downwelling irradiance sensor and a radiometric calibration for the imager. The sensor, calibrated to spectral radiance across the range of 400-100 nm with a spectral resolution of 1.34 nm, was configured to record a downwelling irradiance spectrum whenever a hyperspectral data cube was recorded by the Pika L imager. By recording the downwelling irradiance at this rate, changes in lighting could be captured throughout the flight.

To account for the significant decrease in signal-to-noise ratio introduced by water absorption in the near-infrared, we removed all spectral information beyond 850 nm from our imagery before use in algorithm development and analysis. Next, we smoothed each image using a Savitzky-Golay filter [163] with a 3rd-order polynomial and window length of 13. Once smoothed, we took the first derivative of each spectrum to enhance spectral absorption features. Finally, we normalized each spectrum by removing the mean and scaling the variance to one.

Images were georeferenced using a built-in package in the Resonon Spectronon Pro software that relied on flight parameters (flight height and ground elevation), imager parameters (focal length), and location information (GPS and inertial measurement unit). The spatial correction aligned each image to cardinal directions and latitude and longitude, but also corrected spatial distortions introduced by both planned and unplanned movement, such as turns to follow the river or tilting caused by wind.

Pigment Abundance Estimation

Overview. Building on our previous work using spectral band ratios to estimate pigment abundance in the UCFR [4], one of our primary objectives for SPAMM was to refine our methodology by using a much larger ground truth data set (n=149) to train and test machine learning regression models; however, the underlying principle of relating the average spectrum from each plot to pigment abundance remained the same. That is, all spatial pixels contained within a plot were spectrally averaged to produce a single spectrum for each plot. The averaged spectrum was coupled with biomass measurements from each plot and used as training or testing pairs for our regressions methods. Several common regression algorithms were trained and tested on our data set, with the top-performing method selected as the main regression method for SPAMM. The regression algorithms were implemented in Python 3.11 using Scikit-learn [164] and included Extra Trees Regression (ETR) [64], Histogram-based

Gradient Boosting Regression Tree (HGBR) [178], k-nearest neighbors regression (KNN) [179], support vector machine regression (SVR) [180], Lasso regression [181], and a stacked regression [182] using KNN, SVR, and ET as base estimators and a Bayesian Ridge Regression [183] as the final regression algorithm.

Due to the relative lack of elevated chl *a* abundance values, the output from our models generally underestimated high abundance values. To account for this, a second-stage classifier was added to the regression algorithms above. A random forest classifier was trained on four labeled categories of chl *a* abundance: Trace (0-1 mg/m^2), Low (1-50 mg/m^2), Medium (50-150 mg/m^2), and High (>150 mg/m^2). The outputs of the regression algorithm and the classifier were then compared. If a disagreement occurred where the classifier predicted a higher algal abundance category than the numeric level predicted by the regression model and the classifier had a probability greater than 30%, the mean squared error (MSE) of the model was added to its prediction.

Training and Testing.

Regression. Prior to training the machine learning regression algorithms, the data set was randomly divided into train and test sets using a 70/30 split ratio. The training set (n=104) was used to fit each regression model while the test set (n=45) was used as unseen data to validate each model. Hyperparameters were tuned using a combination of randomized grid and manual searches to optimize the performance on these two data sets, measured by the coefficient of determination (R^2) and MSE. The algorithm that achieved the best results on the test set was selected as the primary regression model for SPAMM.

Classifier. The classifier required categorical labels rather than the numeric values used to train the regression algorithms. The ground truth data set was broken into the categories described above (i.e., Trace, Low, Medium, High) based on numeric value. The labeled data were then randomly divided into train and test sets using a stratified 70/30 split ratio to

evenly represent each class. Hyperparameters were tuned using a randomized grid search and validated using a 5-fold cross-validation. To assess classifier performance in our unbalanced data set, the F1 score was used in place of typical accuracy assessments based on number of correct predictions alone.

Results and Discussion

We begin by discussing the results of pigment estimation techniques, including statistics on the ground truth data collection across 2021 and 2022, the performance of our regression algorithms, and the selection of the primary algorithms for SPAMM. Finally, we use SPAMM to assess water quality conditions in the UCFR at three field sites from 2020 to 2022.

Pigment Estimation

Ground Truth Data. The total chl *a* abundance measured from the plots placed at Deer Lodge, Garrison, Gold Creek, and Bear Gulch ranged from a maximum value of 348.9 mg/m^2 to a minimum of 34.8 mg/m^2 in 2021 and a maximum of 430.1 mg/m^2 and minimum of 0.06 mg/m^2 in 2022, with average values of 171.1 $\pm$ 91.9 mg/m^2 and 36.9 $\pm$ 64.7 mg/m^2, respectively (Table 5.1). Average total chl *a* levels were higher during all 2021 field collection campaigns, indicating that *Cladophora* growth was less significant in 2022 or had not yet reached peak growth in the earlier summer months of 2022.

Regression and Classification Models. The six regression models we selected for analysis were assessed according to their R^2 and MSE values after comparing predictions made by the algorithm and the test set (Table 5.2). All MSE values were calculated as errors around predictions made in transform space, thus units of mg/m^2 are not reported. All regression algorithms showed strong performance, ranging from R^2 values of 0.787 to 0.873, with the stacked regression performing the best. The stacked model was built using the tuned KNN,

Table 5.1: Standing crop of chlorophyll a (chl *a*) Deer Lodge, Garrison, Gold Creek, and Bear Gulch. Data are means ± standard error, after five outlying points were removed.

Field Site	**Field Date**	**Chl *a* (mg/m^2)**	**Number of Samples**
Deer Lodge	29 June 2022	4.9 ± 7.4	19
Garrison	28 June 2022	1.0 ± 1.2	16
	12 July 2022	64.2 ± 97.9	21
Gold Creek	17 August 2021	191.4 ± 74.0	6
	9 September 2021	150.7 ± 92.0	19
	13 July 2022	27.4 ± 20.1	21
	26 July 2022	65.6 ± 86.7	21
Bear Gulch	17 August 2021	167.6 ± 72.8	6
	27 July 2022	37.2 ± 31.5	20

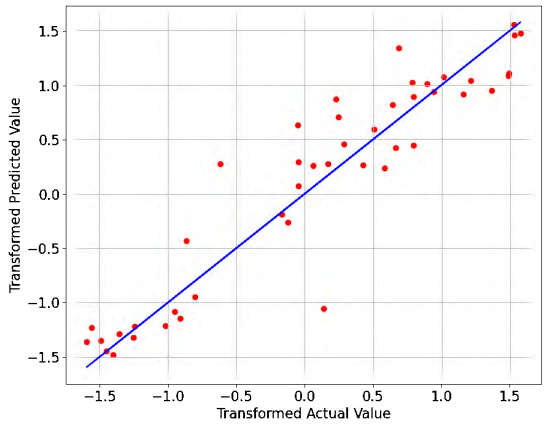
SVR, and ETR models as base estimators and a Bayesian Regression model tuned using a 5-fold cross validation on the base estimator outputs.

The stacked regression was selected as the primary algorithm for SPAMM; however, due to the lack of high chl *a* abundance values in our 2022 ground truth data, the model tended to perform poorly for elevated chl *a* values, particularly after reverse transforming the predictions into areal density values (Figure 5.5). To explore how the error grew with increasing chl *a* abundance, we calculated the MSE of the regression model for the Trace, Low, Medium, and High categories described in the Overview section, resulting in MSE values of 0.62, 22.30, 44.18, and 71.33 mg/m^2, respectively. To help account for the growing error in high values caused by underestimation, we added a second stage to our abundance estimation model.

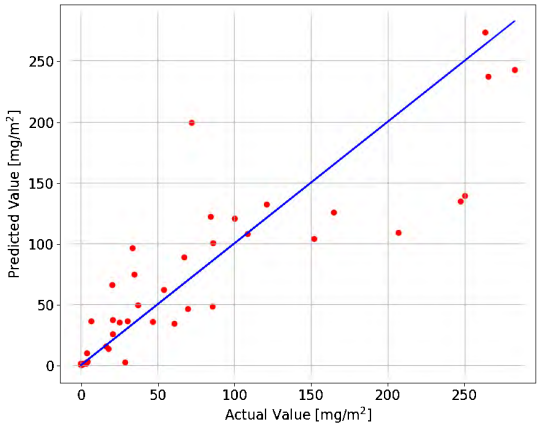
Pigment Abundance Classification. The abundance classifier showed strong performance for the Low, Medium, and High abundance categories, with degraded performance for the

Table 5.2: Regression results from the six selected regression algorithms, reported as R^2 and MSE after prediction of the test set. MSE is reported in transform space across all fits.

Model	**R^2**	**MSE**
Extra Trees Regression	0.825	0.170
Histogram-based Gradient Boosting Regression	0.788	0.206
K-Nearest Neighbors Regression	0.871	0.125
Support Vector Regression	0.861	0.135
Lasso Regression	0.787	0.206
Stacked Regression	**0.873**	**0.123**



(a) Predicted versus actual values of the stacked regressor algorithm in transform space, showing a tight fit for most values.



(b) Predicted versus actual values of the stacked regression algorithm in abundance space, showing the large error in higher abundance values and trend of underpredicting high areal biomass values.

Figure 5.5: Stacked regression model performance of actual versus predicted values (red dots) shown with a 1:1 line (solid line), where any point falling on this line is a perfect prediction, in transform space (a) and abundance space (b).

Trace category (Table 5.3). This pattern was also seen in the confusion matrix, where Trace values are often confused with the Low category (Figure 5.6a); however, the overall absolute accuracy was still strong at 82.6%. As previously described, with an unbalanced data set such as ours, the absolute accuracy may not be the best method for assessing model performance. Therefore, for tuning hyperparameters, we used the F1 score to assess model performance, with our top model averaging an F1 score of 0.80 across all abundance classes. When used with the stacked regression model described above, the classifier added the MSE for each abundance category to the output of the regression model when the classifier disagreed with the numeric prediction of the regression and had a probability greater than 30%. With the second stage in place, R^2 improved by approximately 0.024. Finally, to aid in the design of future multispectral imaging systems, we used the mean decrease in impurity to determine the top 10 wavelengths for determining algal abundance category (Figure 5.6b). Unsurprisingly, the identified wavelengths were very similar to those used to discriminate between *Cladophora*, diatoms/SAV, and bare substrate in our previous work [175], though wavelengths tended to be closer to a known chl *a* absorption feature [129, 130] and included more wavelengths beyond 730 nm.

Table 5.3: Regression results from the six selected regression algorithms, reported as R^2 and MSE after prediction of the test set. MSE is reported in transform space across all fits.

Category	**Precision**	**Recall**	**F1 Score**	**Number of Samples**
Trace	0.67	0.57	0.62	7
Low	0.84	0.91	0.87	23
Medium	0.80	0.80	0.80	10
High	1.0	0.83	0.91	6
Average	0.83	0.78	0.80	46



(a) Confusion matrix for the algal abundance classifier showing little confusion between classes, other than the Trace category, which was confused for both Low and Medium across the testing set.

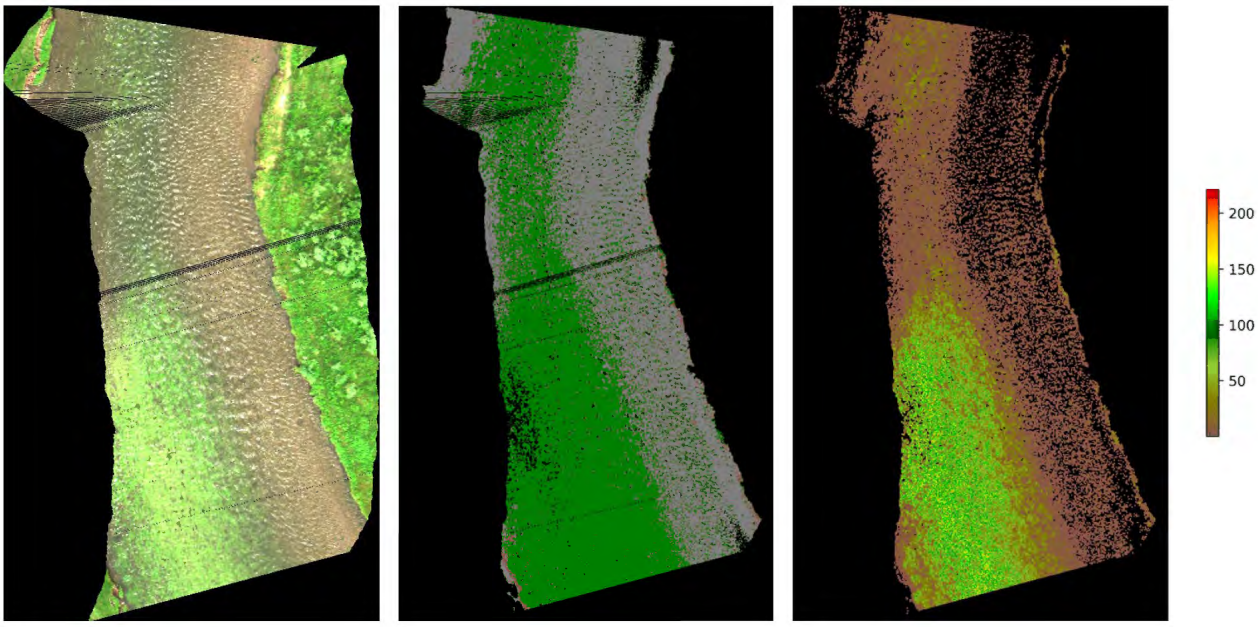
(b) Top ten wavelengths identified using mean decrease in impurity, showing a collection of wavelengths around a known chl *a* absorption line and several wavelengths beyond 700 nm.

Figure 5.6: Metrics from chl *a* abundance classifier, including the confusion matrix (a) and top ten wavelengths for classifying abundance (b).

Case Study: SPAMM on the UCFR

Here we explore the use and efficacy of SPAMM as a water quality monitoring tool by analyzing imagery collected from 2020, 2021, and 2022 at our analysis field sites along the UCFR. Using a georectified, reflectance hyperspectral image (Figure 5.7a), SPAMM creates a spatial distribution map of algal growth from which percent cover estimates are made (Figure 5.7b), a chl *a* abundance map from which mean standing crop and algal abundance category are estimated (Figure 5.7c), and a spatial histogram of abundance values from which areal coverage of algal abundance categories are estimated, including area in excess of Montana water quality standards (Figure 5.7d). The following sections use these estimated metrics averaged across each full reach (i.e., field site) to give a single value for each flight, broken down by year of collection. Finally, we synthesize all results and discuss the use of SPAMM as a water quality monitoring tool.

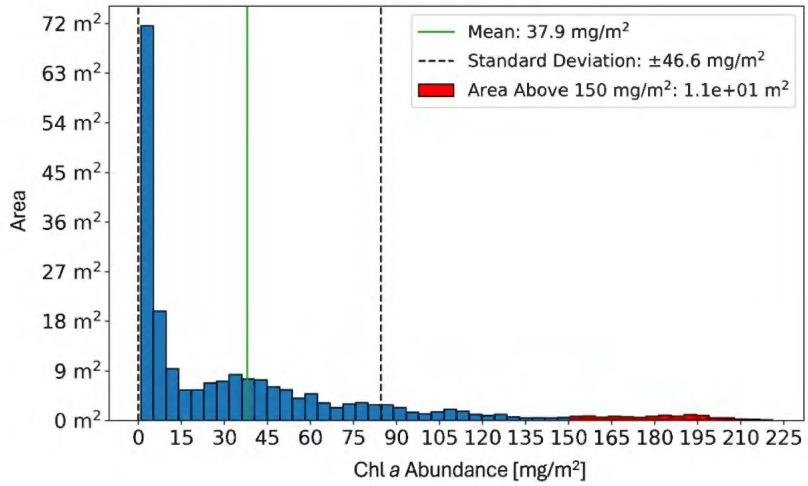
2020 Field Sites. We began by using SPAMM to analyze field imagery collected at Deer Lodge, Garrison, and Bear Gulch in 2020, producing estimates of average spatial coverage of *Cladophora* and diatoms/SAV for each flight and seasonal averages (Table 5.4). In addition, we estimated chl *a* abundance as a mean value for each flight, percentage of abundance that fell into the Low, Medium, and High categories described in the Overview section, and season averages (Table 5.5). Though measurements began in early June, no algae was detected until 13 July, when blooms of *Cladophora* began to emerge at Deer Lodge and Garrison. Little algal growth was detected at Bear Gulch on 14 July; however, a large bloom occupying nearly 75% of the reach was observed on 28 July. Maximum percent cover occurred in late July at Deer Lodge, early August at Garrison, and between late July and early August at Bear Gulch. Interestingly, all field sites showed signs of a resurgence event at the beginning of September, with percent *Cladophora* cover increasing by approximately 37% between 28 August and 11 September at Deer Lodge and Garrison and approximately 32% at Bear Gulch.



(a) Input hyperspectral image shown in RGB.

(b) Spatial distribution map with bank, saturated, and shadowed pixels removed.

(c) Chl *a* abundance map, with estimates only calculated for pixels identified as algae from (b).



(d) Spatial histogram of areal coverage for algal abundance categories, with mean (solid green), standard deviation (dashed), and areas in excess of water quality standards (red bars) shown.

Figure 5.7: Example outputs of SPAMM, shown at Garrison on 3 August 2020. Spectral products are created from input image, shown in RGB (a), spatial distribution maps show *Cladophora* cover (green) and diatom/SAV cover (brown) from which percent cover estimates are made (b), algal abundance maps show heat map of chl *a* values (c), and histograms show spatial coverage of chl *a* abundance, including areas in excess of Montana water quality standards (red bars)(d).

Across the season, Bear Gulch averaged the highest percent cover for both *Cladophora* and diatoms/SAV, with means of 40.9% ± 20.3% and 24.6% ± 17.5%, respectively.

Estimates of total chl *a* told a similar story, with a few subtle differences. Maximum chl *a* abundance did not occur until early August at Deer Lodge, compared to late July as reported by algal cover estimates. Garrison had a local maximum in early August, but did not hit its peak chl *a* abundance until the beginning of September. No chl *a* was detected until the end of July at Bear Gulch, with a local maximum observed at the beginning of August and a peak value at the beginning of September. As observed in our percent cover estimates, chl *a* abundance increased at the end of September at all of our field sites, again indicating a resurgence event. Contrary to the percent cover maps, Deer Lodge averaged the highest average chl *a* abundance across the season, with an average value of 33.0 ± 1.2 mg/m^2. Deer Lodge showed particularly high chl *a* abundance values at the end of the sampling period, with an average of 2.1% and 1.8% pixels categorized as falling above Montana water quality standards on 27 August and 10 September, respectively.

Table 5.4: Estimates of the percent *Cladophora* and diatoms/SAV cover of the three 2020 field sites. Values are reported as the average across the full reach analyzed.

2020 Flight Date	Deer Lodge		Garrison		Bear Gulch	
	Cladophora	**Diatoms/SAV**	**Cladophora**	**Diatoms/SAV**	**Cladophora**	**Diatoms/SAV**
11 June	0.2%	0%	–	–	–	–
12 June	–	–	–	–	0.2%	0%
26 June	–	–	0%	0%	0%	0%
13 July	37.1%	0.1%	33.6%	4.6%	–	–
14 July	–	–	–	–	6.7%	0.2%
22 July	46.9%	14.2%	32.2%	2.7%	–	–
27 July	44.2%	14.3%	39.7%	2.0%	–	–
28 July	–	–	–	–	74.9%	13.2%
3 August	29.0%	0%	47.5%	4.4%	–	–
4 August	–	–	–	–	43.2%	5.2%
10 August	9.4%	6.5%	47.4%	12.4%	–	–
11 August	–	–	–	–	47.4%	25.4%
20 August	39.3%	14.3%	–	–	–	–
21 August	–	–	–	–	37.5%	46.3%
27 August	2.7%	20.6%	1.8%	67.2%	–	–
28 August	–	–	–	–	22.8%	44.6%
10 September	39.1%	10.1%	38.8%	26.7%	–	–
11 September	–	–	–	–	54.0%	37.6%
Season Average	31.1% ± 15.0%	10.1% ± 6.8%	34.4% ± 14.4%	17.1% ± 22.0%	40.9% ± 20.3%	24.6% ± 17.5%

Table 5.5: Estimated chl *a* abundance for the 2020 field sites. Mean chl *a* is reported as the average value across the river reach analyzed, "Low" values are abundance values between 1-50 mg/m^2, "Med" values are abundance values between 50-150 mg/m^2, and "High" values are abundance values greater than 150 mg/m^2.

2020 Flight Date	Deer Lodge				Garrison				Bear Gulch			
	Mean [mg/m^2]	**"Low"**	**"Med"**	**"High"**	**Mean [mg/m^2]**	**"Low"**	**"Med"**	**"High"**	**Mean [mg/m^2]**	**"Low"**	**"Med"**	**"High"**
11 Jun	0	0%	0%	0%	–	–	–	–	–	–	–	–
12 Jun	–	–	–	–	–	–	–	–	0	0%	0%	0%
26 Jun	–	–	–	–	0	0%	0%	0%	–	–	–	–
13 Jul	31.5	92%	7.6%	0.4%	27.9	94.6%	5.3%	0%	–	–	–	–
14 Jul	–	–	–	–	–	–	–	–	0	0%	0%	0%
22 Jul	33.2	86.1%	13.4%	0.4%	29.04	84.5%	14.7%	0.8%	–	–	–	–
27 Jul	33.3	83.9%	15.2%	0.9%	31.9	78.7%	19.3%	2.0%	–	–	–	–
28 Jul	–	–	–	–	–	–	–	–	31.1	83.9%	14.0%	2.1%
3 Aug	35.2	82.3%	16.6%	1.1%	32.0	82.2%	16.1%	1.7%	–	–	–	–
4 Aug	–	–	–	–	–	–	–	–	31.8	81.0%	17.3%	1.7%
10 Aug	32.6	84.5%	13.1%	2.5%	32.1	82.5%	16.1%	1.5%	–	–	–	–
11 Aug	–	–	–	–	–	–	–	–	32.9	90.6%	9.0%	0.4%
20 Aug	31.4	90.3%	8.8%	0.9%	–	–	–	–	–	–	–	–
21 Aug	–	–	–	–	–	–	–	–	30.7	91.8%	7.9%	0.3%
27 Aug	33.6	86.9%	11.1%	2.1%	31.69	89.7%	9.9%	0.4%	–	–	–	–
28 Aug	–	–	–	–	–	–	–	–	32.0	89.7%	9.7%	0.6%
10 Sep	33.6	87.5%	10.7%	1.8%	32.70	85.6%	13.6%	0.8%	–	–	–	–
11 Sep	–	–	–	–	–	–	–	–	34.8	87.5%	11.8%	0.7%
Season Average	33.0 ± 1.2	86.7% ± 3.1%	12.1% ± 2.9%	1.3% ± 0.7%	31.0 ± 1.7	85.4% ± 4.9%	13.6% ± 4.3%	1.0% ± 0.7%	32.2 ± 1.4	87.4% ± 3.8%	11.6% ± 3.2%	1.0% ± 0.7%

2021 Field Sites. Growing conditions during the summer of 2021 followed a different pattern than we observed in 2020, with peak *Cladophora* growth occurring on, or prior to, data collection beginning on 30 June. Estimates of percent cover indicated that maximum algal cover occurred earlier than the typical July or August peak at all field sites, with the

greatest cover at Deer Lodge and Bear Gulch measured on 30 June and 1 July, respectively (Table 5.6). *Cladophora* coverage at Garrison remained relatively low throughout the sampling period, with maximal coverage occurring on 27 July and 1 September, though both values were relatively low compared to our other field sites. As in 2020, Bear Gulch had the highest seasonal *Cladophora* coverage, though it also showed the greatest variability, with a mean value of 29.1% and standard deviation of 17.3%. Garrison had the highest seasonal average of diatom/SAV growth, with an average value of 63.1% $\pm$ 17.8%. As observed in 2020, all field sites showed signs of a resurgence event at the end of the sampling period, with increased *Cladophora* growth and decreased diatom/SAV growth at all sites.

Estimation of chl *a* abundance at our 2021 field sites revealed a different pattern than percent cover values, with peak chl *a* abundance observed at the end of the sampling period for all field sites (Table 5.7). Bear Gulch again showed the greatest average chl *a* abundance, though Deer Lodge had a similar average value. Our final sampling event at each site showed significant increases in the average chl *a* abundance, particularly exemplified by the percent of pixels identified as "High," or a value greater than 150 mg/m^2. More than 1% of pixels at all sites fell into this category, with 1.7% of pixels at Deer Lodge over Montana water quality standards on 1 September.

Table 5.6: Estimates of the percent *Cladophora* and diatoms/SAV cover of the three 2021 field sites. Values are reported as the average across the full reach analyzed.

2021 Flight Date	Deer Lodge		Garrison		Bear Gulch	
	Cladophora	**Diatoms/SAV**	**Cladophora**	**Diatoms/SAV**	**Cladophora**	**Diatoms/SAV**
30 June	40.8%	41.6%	13.2%	26.7%	–	–
1 July	–	–	–	–	61.2%	21.3%
8 July	18.9%	69.9%	–	–	–	–
9 July	–	–	13.6%	49.2%	33.7%	47.7%
14 July	20.5%	71.9%	9.1%	71%	–	–
15 July	–	–	–	–	29.4%	61.0%
21 July	35.8%	42.4%	–	–	–	–
22 July	–	–	10.9%	78.1%	3.1%	27.7%
27 July	–	–	18.5%	73.0%	–	–
28 July	–	–	–	–	15.4%	30.2%
12 August	17.5%	72.5%	8.5%	76.1%	–	–
13 August	–	–	–	–	20.6%	49.2%
25 August	15.1%	68.5%	11.7%	73.4%	–	–
1 September	31.0%	43.7%	18.6%	43.6%	–	–
2 September	–	–	–	–	40.0%	27.1%
Season Average	25.7% ± 9.3%	58.7% ± 14.0%	12.9% ± 3.6%	63.1% ± 17.8%	29.1% ± 17.3%	37.7% ± 13.7%

Table 5.7: Estimated chl *a* abundance for the 2021 field sites. Mean chl *a* is reported as the average value across the river reach analyzed, "Low" values are abundance values between 1-50 mg/m^2, "Med" values are abundance values between 50-150 mg/m^2, and "High" values are abundance values greater than 150 mg/m^2.

2021 Flight Date	Deer Lodge				Garrison				Bear Gulch			
	Mean [mg/m^2]	**"Low"**	**"Med"**	**"High"**	**Mean [mg/m^2]**	**"Low"**	**"Med"**	**"High"**	**Mean [mg/m^2]**	**"Low"**	**"Med"**	**"High"**
30 Jun	29.4	88.0%	11.5%	0.5%	28.2	89.6%	9.8%	0.7%	–	–	–	–
1 Jul	–	–	–	–	–	–	–	–	31.1	81.2%	18.0%	0.8%
8 Jul	29.8	88.6%	10.9%	0.4%	–	–	–	–	–	–	–	–
9 Jul	–	–	–	–	29.4	88.9%	10.4%	0.7%	32.8	86.5%	12.7%	0.8%
14 Jul	31.9	87.9%	11.2%	0.9%	–	–	–	–	–	–	–	–
15 Jul	–	–	–	–	–	–	–	–	32.0	90.2%	9.3%	0.5%
21 Jul	29.7	91.5%	8.3%	0.2%	–	–	–	–	–	–	–	–
22 Jul	–	–	–	–	29.8	88.9%	10.6%	0.5%	30.1	91.7%	8.0%	0.2%
27 Jul	–	–	–	–	29.5	90.8%	8.9%	0.3%	–	–	–	–
28 Jul	–	–	–	–	–	–	–	–	30.3	90.5%	9.2%	0.2%
12 Aug	30.7	87.0%	12.5%	0.5%	29.8	87.1%	12.4%	0.6%	–	–	–	–
13 Aug	–	–	–	–	–	–	–	–	31.2	87.5%	12.0%	0.5%
25 Aug	33.1	85.1%	13.9%	0.9%	32.8	83.7%	15.0%	1.3%	–	–	–	–
1 Sep	35.0	83.2%	15.1%	1.7%	33.5	81.8%	16.7%	1.6%	–	–	–	–
2 Sep	–	–	–	–	–	–	–	–	34.2	85.9%	13.0%	1.1%
Season Average	31.4 ± 1.9	87.3% ± 2.4%	11.9% ± 2.0%	0.7% ± 0.5%	30.4 ± 1.8	87.2% ± 3.1%	12.0% ± 2.7%	0.8% ± 0.4%	31.7 ± 1.3	87.7% ± 3.3%	11.7% ± 3.1%	0.6% ± 0.3%

2022 Field Sites. Our data collection in 2022 was focused on ground truth measurements of algal abundance, leading to fewer analysis sites; however, observations of general trends can be extracted. *Cladophora* growth in 2022 indicated a more typical pattern, with the greatest percent cover values observed in late July and early August across all sites (Table 5.8). Algal growth was still low when measured on 29 June at Deer Lodge, with both *Cladophora* and diatom/SAV covering only 16.2% of the reach; however, Garrison had greater algal growth, with over 40% of the reach covered by either algal growth type on 28 June. Deer Lodge averaged the highest *Cladophora* cover in 2022, driven by the presence of a massive bloom observed in August (Figure 5.8).

Our estimations of algal abundance matched the conclusions drawn from percent cover measurements, with maximum chl *a* standing crops measured at the end of the sampling period at all sites (Table 5.9). Deer Lodge showed the highest average chl *a* abundance, likely driven by the presence of the large *Cladophora* bloom at the sampling reach leading to between 2-3% of pixels identified as above water quality standards. As previously noted, the regression algorithm may be underpredicting for the large chl *a* abundance likely present in the bloom, meaning a greater area is likely over the water quality standards.

Field Site Evolution

The progression of *Cladophora* and diatom growth from 2020 to 2022 at Deer Lodge indicated that the 2021 growing season was shifted earlier than in 2020 and 2022 (Figure 5.9a). *Cladophora* growth in 2020 had three peaks in percent cover, occurring on 22 July, 20 August, and 10 September, whereas growth in 2021 appeared to peak prior to the first sampling period on 30 June, though a local peak was observed on 21 July. Though limited to two data points, the 2022 growing season appeared to peak around the middle of August, with over 80% coverage of *Cladophora*. We found growth across all three years to be extremely variable along the length of all reaches, with percent cover estimates varying by over 50%, reiterating the



Figure 5.8: Large bloom of *Cladophora* observed at Deer Lodge in August of 2022, causing percent cover estimates to increase to over 80% with significant portions estimated in excess of water quality standards.

Table 5.8: Estimates of the percent *Cladophora* and diatoms/SAV cover of the three 2022 field sites. Values are reported as the average across the full reach analyzed.

2022 Flight Date	Deer Lodge		Garrison		Bear Gulch	
	Cladophora	**Diatoms/SAV**	**Cladophora**	**Diatoms/SAV**	**Cladophora**	**Diatoms/SAV**
28 June	–	–	19.9%	19.4%	–	–
29 June	8.4%	7.8%	–	–	–	–
12 July	–	–	41.8%	7.6%	–	–
21 July	–	–	45.8%	11.6%	–	–
27 July	–	–	–	–	42.4%	32.5%
15 August	80.3%	5.0%	–	–	–	–
Season Average	44.3% ± 35.9%	6.4% ± 1.4%	35.8% ± 11.4%	12.9% ± 4.9%	42.4%	32.5%

need for high-resolution sampling to augment or replace traditional spot sampling. Diatom coverage was much higher in 2021 than in either 2020 or 2022, with percent cover regularly measured above 60% (Figure 5.9b). As observed in our previous work [4][Logan and Shaw, in

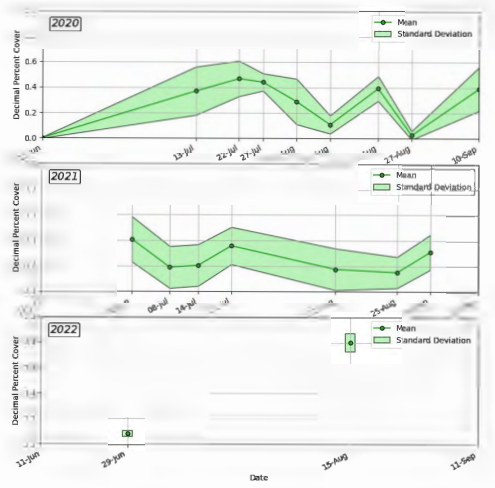
Table 5.9: Estimated chl *a* abundance for the 2022 field sites. Mean chl *a* is reported as the average value across the river reach analyzed, "Low" values are abundance values between 1-50 mg/m^2, "Med" values are abundance values between 50-150 mg/m^2, and "High" values are abundance values greater than 150 mg/m^2.

2022 Flight Date	Deer Lodge				Garrison				Bear Gulch			
	Mean [mg/m^2]	**"Low"**	**"Med"**	**"High"**	**Mean [mg/m^2]**	**"Low"**	**"Med"**	**"High"**	**Mean [mg/m^2]**	**"Low"**	**"Med"**	**"High"**
28 Jun	–	–	–	–	29.8	87.4%	11.5%	1.1%	–	–	–	–
29 Jun	27.9	89.6%	9.9%	0.5%	–	–	–	–	–	–	–	–
12 Jul	–	–	–	–	31.2	84.2%	14.7%	1.2%	–	–	–	–
21 Jul	–	–	–	–	34.8	80.7%	17.3%	2.0%	–	–	–	–
27 Jul	–	–	–	–	–	–	–	–	29.9	89.5%	10.1%	0.3%
15 Aug	36.8	78.0%	19.2%	2.9%	–	–	–	–	–	–	–	–
Season Average	32.3 ± 4.5	83.8% ± 5.8%	14.5% ± 4.6%	1.7% ± 1.2%	31.9 ± 2.1	84.1% ± 2.7%	14.5% ± 2.4%	1.4% ± 0.4%	29.9	89.5%	10.1%	0.3%

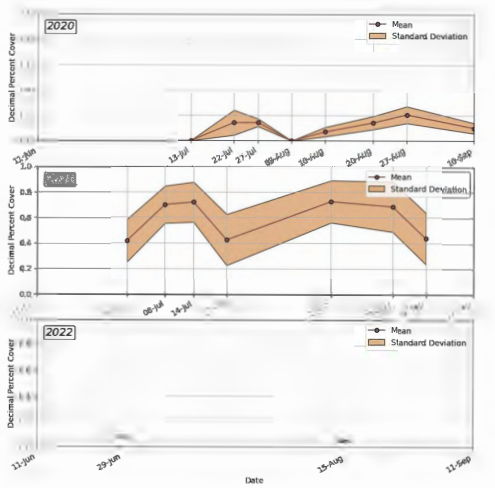
submission], *Cladophora* and diatom growth appeared to have an inverse relationship.

Cladophora growth at Garrison showed less variability within each year, with peak growth occurring in late July or early August in all three years measured (Figure 5.10a). Coverage estimates were significantly lower throughout 2021 compared to 2020 and 2022, with all mean coverage estimates below 20% in 2021. *Cladophora* growth in 2020 and 2022 had similar patterns, with an average percent cover of approximately 40% throughout the summer season once algae had been established. As previously noted, *Cladophora* growth in both 2020 and 2021 showed signs of a late-summer resurgence event, with percent cover estimates increasing on the final sampling period in early September.

Diatom growth at Garrison peaked in late August, with over 60% coverage estimated along the reach (Figure 5.10b). As we observed at Deer Lodge in 2021, diatom coverage at Garrison in 2021 was elevated throughout the season, with most percent cover estimates over 70%. Diatom growth in 2020 and 2022 indicated a similar pattern, with peak growth not



(a) Decimal percent *Cladophora* cover.



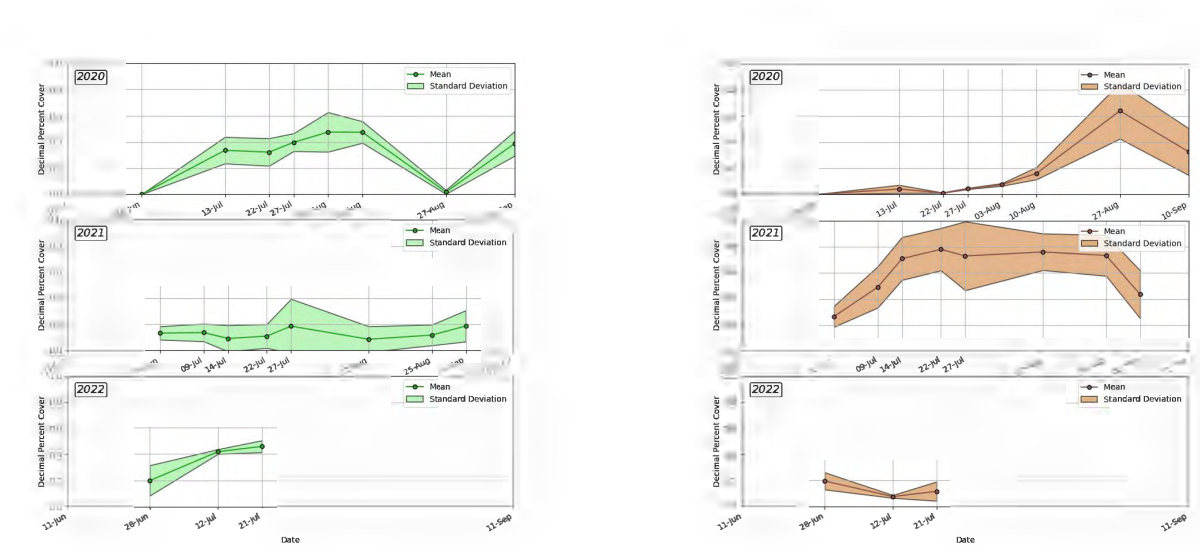
(b) Decimal percent diatom/SAV cover.

Figure 5.9: Deer Lodge decimal percent cover estimates measured from 0 to 1 for 11 June through 11 September of 2020 (top), 2021 (middle), and 2022 (bottom) for *Cladophora* (a) and diatoms/SAV (b). Mean value for each flight is shown with solid dot and standard deviation shown with fill.

occurring until late summer.

Bear Gulch coverage peaked relatively early in 2020, with maximum *Cladophora* cover measured on 28 July; however, coverage did not drop to the late-August lows we observed at Deer Lodge and Garrison, with coverage remaining above 20% once *Cladophora* was established (Figure 5.11a). *Cladophora* coverage again showed a negative trend from the beginning of our sampling in 2021, reaching coverage estimates of less than 5% by 22 July. Interestingly, the lowest coverage at Bear Gulch in 2021 nearly coincided with peak algal growth in 2020, and occurred when Deer Lodge and Garrison had local maxima in 2021. As in other sites in 2021, after reaching minimal coverage, *Cladophora* recovered and percent cover increased until the end of the sampling period. Though limited to a single measurement, *Cladophora* coverage in 2022 fell closer to values measured in 2020, with a percent cover of slightly over 40% at the end of July.

Diatom coverage at Bear Gulch followed similar trends to Deer Lodge and Garrison



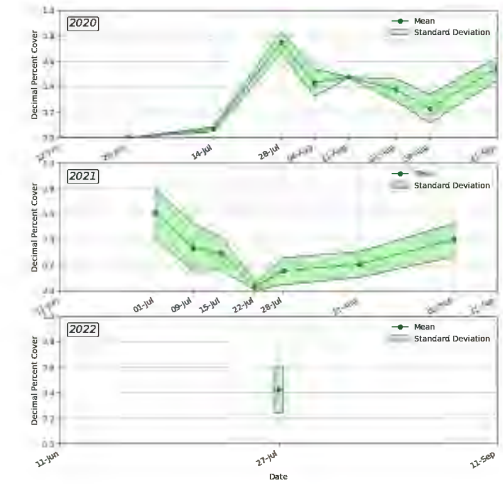
(a) Decimal percent *Cladophora* cover.

(b) Decimal percent diatom/SAV cover.

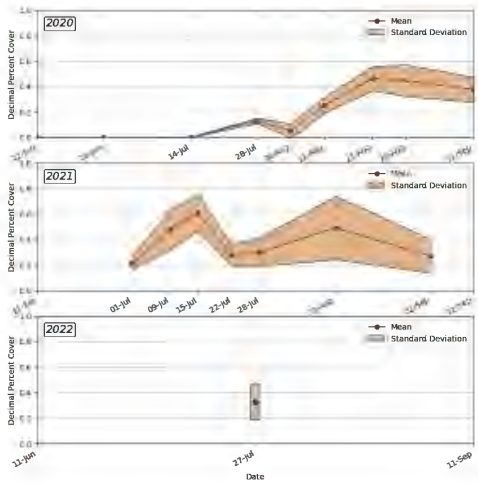
Figure 5.10: Garrison decimal percent cover estimates measured from 0 to 1 for 11 June through 11 September of 2020 (top), 2021 (middle), and 2022 (bottom) for *Cladophora* (a) and diatoms/SAV (b). Mean value for each flight is shown with solid dot and standard deviation shown with fill.

in 2020, with growth starting in late July and peaking in late August (Figure 5.11b). We observed peak growth at Bear Gulch in 2021 in mid-July, similar to our estimations at Deer Lodge and Garrison, with values remaining relatively high throughout the growing season. Our only measurement in 2022 was similar to percent cover estimates in 2021, though limited data make it difficult to draw firm conclusions.

Water Quality Standards. The spatial histograms discussed in the Case Study: SPAMM on the UCFR section provided estimates of the areal extent of blooms above the Montana water quality standard of 150 mg/m^2. Using these histograms, we calculated the average area above the regulatory threshold for each measurement date and reach (Figure 5.12). In general, the areal extent in excess of standards followed the patterns we observed in our percent cover estimates, with greater percent cover typically increasing the area above the threshold. Once algal biomass began to accrue, State water quality standards were exceeded



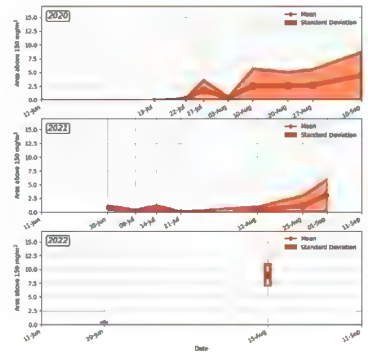
(a) Decimal percent *Cladophora* cover.



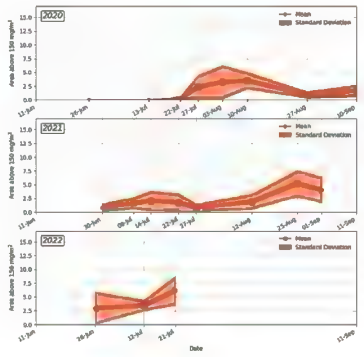
(b) Decimal percent diatom/SAV cover.

Figure 5.11: Bear Gulch decimal percent cover estimates measured from 0 to 1 for 11 June through 11 September of 2020 (top), 2021 (middle), and 2022 (bottom) for *Cladophora* (a) and diatoms/SAV (b). Mean value for each flight is shown with solid dot and standard deviation shown with fill.

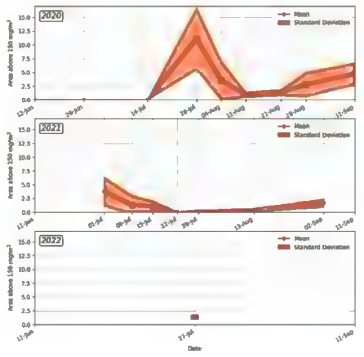
at every site during every sampling period across the three years we analyzed.



(a) Deer Lodge



(b) Garrison



(c) Bear Gulch

Figure 5.12: Estimated areal extent of *Cladophora* in excess of Montana water quality standards (150 mg/m^2) measured from 0 to 17 m^2 for 11 June through 11 September of 2020 (top), 2021 (middle), and 2022 (bottom) for Deer Lodge (a), Garrison (b), and Bear Gulch (c). Mean value for each flight is shown with solid dot and standard deviation shown with fill.

Discussion

Uncovering generalized annual trends of *Cladophora* and diatom/SAV growth from our 2020-2022 data proved elusive due to the significant difference in growth patterns we observed between 2020 and 2021 and relative lack of measurements in 2022. Though *Cladophora* growth is affected by many factors, including temperature, nutrient levels, and insolation [137], the large difference we observed between 2020 and 2021 may have been partially caused by river depth and temperature. Hydrograph measurements provided by the United States Geological Survey (USGS) show that 2020 and 2021 were substantially different water years, with peak discharges at Garrison measured at approximately 1770 ft^3/s on 30 June 2020, 178 ft^3/s on 22 August 2021, and 412 ft^3/s on 4 July 2022 (Figure 5.13). Low discharge in 2021 helps to explain the early onset of *Cladophora*, as algal biomass was allowed to accrue much earlier than in 2021 and 2022, while the resurgence of algal biomass in 2021 may also be explained by the introduction of additional flow from reservoirs released at the beginning of August. The renewed growth at the end of 2020 may be partially explained by water temperatures, but remains unknown.

Once we detected *Cladophora* at a site, some portion of the reach always exceeded the current water quality standards. The largest average area covered by algae above the regulatory threshold was at Bear Gulch on 28 July 2020, where we estimated an average of over 14 m^2 along the reach.

Whether estimating percent cover, chl *a* abundance, or the area in excess of water quality standards, we observed significant spatial variation within each reach. Using a portion of our 10 September 2020 flight at Deer Lodge as an example, we created a mosaic from four hyperspectral images to give us an RGB image of the flight (Figure 5.14a, top); a spatial distribution map of algal growth, including *Cladophora* (green), diatoms/SAV (brown), and bare substrate (grey) (Figure 5.14a, middle); and a chl *a* abundance heat map (Figure 5.14a, bottom). Next, we plotted decimal percent *Cladophora* cover (Figure 5.14b, top), chl *a*

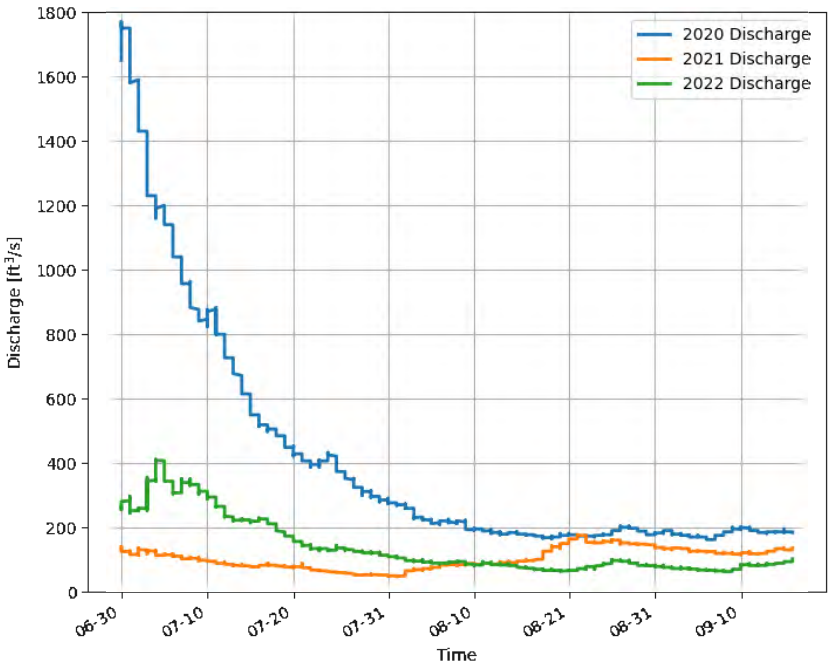
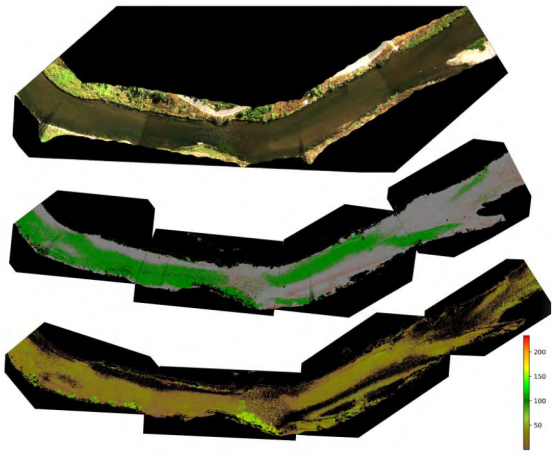


Figure 5.13: Gage height measurements provided by the USGS [184] at Garrison from 2020 (blue), 2021 (orange), and 2022 (green) showing the low water year experiences in 2021, possibly contributing to early accrual of *Cladophora* and higher levels of diatoms.

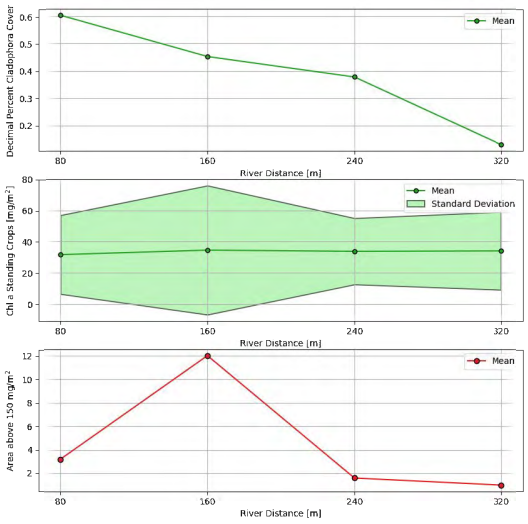
abundance (Figure 5.14b, middle), and area above 150 mg/m^2 (Figure 5.14b, bottom) as a function of downstream distance, using the average value from the hyperspectral images, each of which covered approximately 80 m of the reach, giving a total length of 320 m, broken into four points (Figure 5.14).

Visual inspection of the RGB image paints an enigmatic picture of the spatial variability of *Cladophora* growth and abundance; however, the spatial distribution and chl *a* abundance maps show the significant variation in not only growth patterns, but also chl *a* abundance. The variability is also evident in the line plots as a function of reach distance, where percent *Cladophora* cover varies by nearly 50% between the first and last portions of the reach. Though the average chl *a* abundance appears consistent across the 320-m reach, it belies the significant chl *a* standing crops found along the bank, giving rise to a standard deviation of over 41 mg/m^2 across the second 80-m reach and a 12-m^2 bloom of *Cladophora* above the

regulatory threshold.



(a) 320-m portion of the Deer Lodge site created from a four-image mosaic, shown in RGB (top); as spatial distribution map with *Cladophora* (green), diatoms/SAV (brown), and bare substrate (grey) classes (middle); and chl *a* abundance heat map (bottom).



(b) Metrics as a function of river distance for the 320-m reach at Deer Lodge, including decimal percent *Cladophora* cover (top), chl *a* abundance (middle), and area above 150 mg/m^2 (bottom).

Figure 5.14: RGB image, spatial distribution map, and chl *a* abundance heat map (a) and spatial variation (b) for Deer Lodge on 10 September 2020.

Conclusion

In this work, we have presented a new remote sensing methodology for monitoring water quality in small, flowing freshwater systems. The method is based on monitoring protocols focused on measuring the spatial extent of algal growth (i.e., percent cover) and numeric standards of the spatial density of algal biomass (i.e., chl *a* abundance [mg/m^2]). We used our software, SPAMM, to create spatial algae distribution maps, chl *a* abundance heat maps, and spatial chl *a* abundance histograms for three field sites along the UCFR between 2020 and 2022 in southwestern Montana as a case study for its use in water quality monitoring.

We created spatial distribution maps using a custom-trained random forest classification algorithm, trained on over 2,000,000 labeled pixels from imagery collected on the UCFR. The distribution maps highlighted pixels identified as either *Cladophora*, diatoms/SAV, or bare substrate, then calculated the percent cover for *Cladophora* and diatoms/SAV. After identification, we used the distribution map as a filter to estimate chl *a* abundance only for pixels containing algae. From the abundance maps, we found the mean chl *a* abundance and standard deviation across each reach, comprised of several hyperspectral images covering roughly 80-m per image on average. In addition, we used spatial histograms to assess the area covered by chl *a* abundance values, including the area above the Montana water quality standard of 150 mg/m^2.

We had several findings from our analysis of the UCFR between 2020 and 2022. First, the low water year in 2021 resulted in much earlier accrual of algal biomass, with our estimates indicating that peak growth occurred prior to our first sampling event on 30 June, nearly two months prior to typical maximum growth on the UCFR [1, 69]. In addition to earlier *Cladophora* accrual, we also observed early and elevated diatom growth across the summer growing season compared to 2020 and 2022.

Importantly, we found significant spatial and temporal variation in percent cover and

chl *a* abundance estimates at all field sites across all three years. Percent cover variations of over 50% were recorded during a single flight, and the reach-average percent cover varied by nearly 75% between sampling events approximately two weeks apart, with chl *a* abundance following similar patterns. This variation indicates that current spot sampling techniques may not adequately capture water quality conditions, suggesting that higher spatial and temporal resolution monitoring is needed. Finally, we found that once *Cladophora* had been established at a field site, portions of the reach always exceeded Montana water quality standards, with some reaches estimated at having an average of over 15 m^2 above the threshold.

In future work, we plan to expand the capabilities of SPAMM by collecting more ground truth data with higher chl *a* abundance to improve accuracy at elevated levels. In addition, we plan to use SPAMM to analyze other rivers across Montana that have been identified as impaired to test its generalizability to other freshwater systems.

CONCLUSION

The work I have presented in this dissertation was aimed at addressing the research questions originally proposed in the CREWS project, sometimes modifying them, comparing results with success criteria, and filling established knowledge gaps. In this concluding chapter, I summarize the work that I have completed since the CREWS project was awarded by providing insight into how I have answered these research questions and filled knowledge gaps.

At the time of this writing, water quality sampling is still based on *in-situ* measurement techniques, largely based on transect measurements or spot sampling. These sampling techniques suffer from several problems, including high costs, long collection times, and poor temporal and spatial resolution, yet they remain one of the only reliable ways to accurately measure water quality parameters. These parameters are commonly based on the measurement of the photosynthetically active pigment, chlorophyll *a*, which can be used as a proxy for algal biomass. Water quality standards in Montana are a blend of narrative standards (i.e., visual description and assessment) and the more robust and testable numeric standards. The current numeric standards include a measurement of the spatial density of chl *a*, which can be used as an indicator of river impairment.

Remote sensing techniques have been applied to monitor water quality since the late 1970s as a means of addressing the shortcomings of *in-situ* methods, though many of the established methods have been created for satellite-based imaging systems. Spaceborne remote sensing has had great success in monitoring coastal waters and large freshwater systems, though the spatial resolution of current satellite-based systems is simply too large to monitor the smaller rivers and streams found across the world. Airborne systems, such as those flown on crewed aircraft, solve some of the issues of spaceborne systems; however, still suffer from long sampling times, high costs, and poor repeatability. UAV-based systems have

grown in popularity due to their low cost, highly repeatable flight patterns, fast sampling times, and high-spatial resolution. UAV imaging systems have been used successfully in freshwater lakes and large rivers to measure volumetric chl *a*, though no known work has been completed on assessing the area density of chl *a* in rivers and streams. This brings us to the primary research question presented in my portion of the CREWS project: can hyperspectral optical sensors be adapted to UAV systems to reduce data costs and thus provide an unparalleled perspective on biological controls of surface water quality?

My hope is that by now, the reader agrees that the answer to this question is "yes," but, if not, let me attempt to convince you. With extensive help from many friends and colleagues, I began by creating a methodology for combining benthic algal sampling and UAV-based hyperspectral imaging toward the goal of relating spectral data to chl *a* abundance. I applied this methodology across three field sampling campaigns at two field sites along the UCFR in 2021, where we collected *in-situ* benthic algal samples at known locations after collecting hyperspectral image data of the area. Once the ground truth samples were processed into pigment abundances by colleagues at The University of Montana, I created a brute-force algorithm to determine the best spectral band ratio for estimating chl *a* standing crops and phycocyanin abundance. The algorithm was configured to find the best band ratio for several growth forms of algae, including filamentous, epiphytic, epilithic, and their sum. On its search for the optimal band ratio, the algorithm assessed all possible ratios and recorded their performance, meaning spectral locations with sensitivity to the target pigment could be determined. Though limited to 33 data points, the method showed promising results, with R^2 values of 0.57, 0.62, and 0.44 for estimating total, filamentous plus epiphytic, and epilthic chl *a* abundance, respectively. The algorithm also worked fairly well for estimating phycocyanin, with R^2 values of 0.41, 0.33, and 0.50 for estimating total, filamentous plus epiphytic, and epilithic phycocyanin abundance, respectively. The method was also found to be generalizable, assessed through a resampling approach, and outperformed several common band ratios

found in the literature. With a promising method for estimating pigment abundance present in algal blooms, the next step was to create a method for estimating the spatial extent of algal growth.

The UCFR is commonly home to several types of submerged aquatic vegetation, including a variety of aquatic grasses and mosses, single-cell algae called diatoms, and our dissertation-favorite algae, *Cladophora*. Determining the spatial extent of blooms of *Cladophora* is important for assessing impairment and biological stress in river systems. Chapter 3 of this dissertation is focused on the methodology I developed for assessing the spatial distribution of *Cladophora*, diatoms/SAV, and bare substrate. The methodology is built from ground truth image data generated from labeling over 2,000,000 pixels of river data collected on the UCFR between 2020 and 2022. I used the labeled pixels to train and test a random forest classification algorithm with the goal of identifying the three groups discussed above. The final model showed promising results, with training and validation accuracies greater than 99% and little class confusion. Once trained, I used the algorithm to generate spatial distribution maps of three field sites along the UCFR across the growing season of 2021, while also estimating the percent *Cladophora* cover and percent diatom/SAV cover of each image. I averaged the percent cover estimates for each image, then took the mean of all images to get an average value across the full reach, then compared them throughout the growing season to uncover trends in algal growth in the UCFR in 2021. I found algal growth to have significant spatial and temporal variability within each reach and across sampling dates, varying by over 40% within certain reaches. The spatial and temporal variation I observed was another indicator that traditional sampling methods may be insufficient and could lead to misleading conclusions about the water quality state in our rivers. With the classification method performing well, the last step was to combine the spatial algae distribution maps with improved regression techniques to create a single software capable of building distribution maps and accurately estimating chl *a* abundance at identified algae locations.

Using a larger *in-situ* ground truth data set collected and processed in 2022 with help from our colleagues at The University of Montana, I updated and improved the chl *a* abundance methodology we previously published. The resulting software I developed, Spectral Processing for Algal Monitoring and Mapping (SPAMM), generates spatial distribution maps using the methods outlined in Chapter 3, but uses the larger ground truth data set to train a machine learning regression algorithm to make chl *a* estimates. I tested several regression algorithms and data processing techniques then selected the combination that provided the best results of the testing data set. The top-performing algorithm had promising results, with an R^2 of 0.87 on the test set. Due to our unbalanced data set, with few ground truth measurements with elevated chl *a* abundance, the regression algorithm often underestimated high chl *a* values. To partially correct for this effect, I added a second stage classifier to the prediction pipeline to boost underpredicted values when the classifier and regression model had conflicting predictions, leading to modest improvements in R^2. The final methodology of SPAMM was based on identifying algal locations using the classification algorithm, then estimating chl *a* abundance at all pixels labeled as algae. The final outputs were a spatial distribution map of algae locations and groups along with chl *a* abundance heat maps. From these outputs, I calculated percent algae cover and created spatial histograms, showing the areal coverage of chl *a* abundance groups including areas in excess of Montana water quality standards. Once functioning and performing well on training and testing data sets, I used SPAMM to analyze three years' worth of data at three field sites along the UCFR to assess its use as a tool for monitoring water quality. The results were promising, revealing trends in algal growth between the field sites and years. The most prominent findings were the significant spatial and temporal variation we observed at all field sites throughout all three years, similar to the findings presented at the end of Chapter 3, again signalling the need for improved water quality monitoring methods. Additionally, once *Cladophora* growth was established at a field site, a portion of every reach during every year sampled exceeded the

Montana water quality standard for chl *a* abundance.

I now hope that the reader agrees that UAV-based hyperspectral imaging can be adapted to provide "unparalleled perspectives on measurements of surface water quality" by considering the following findings:

1. Prior to this work, water quality sampling based on chl *a* abundance and percent algal cover in rivers and streams was limited to spot sampling and subjective visual descriptions and assessments, with no established remote sensing methods available. Using the methodologies described here, centimeter-scale spatial resolution data can be collected at 1-km river reach lengths. Given the significant variation we observed within each reach at our field sites, current water quality monitoring techniques may be providing misleading information about the state of our waters, suggesting the need for improved methods.

2. With growing impairment in our waterways, water quality monitoring needs to become easier and more accurate. The methodologies presented here provide means of rapidly quantifying percent algal cover while also estimating algal biomass, allowing comparison to, or updating of, state water quality standards. Across the three years we analyzed field sites along the UCFR, once *Cladophora* had been established at a site, a portion of every reach always exceeded the current numeric standards in Montana, again signalling the need for improved monitoring and remediation.

3. The methods for estimating percent algal cover and algal biomass largely presented here rely on the use of machine learning algorithms, but we have identified salient wavelengths that may be used in the design of low-cost and compact multispectral imagers capable of performing the tasks currently only tested with hyperspectral imaging.

My hope is that the remote sensing methodologies presented here lay the groundwork for improving water quality monitoring across the state of Montana and beyond. However,

for the methodologies described here to be used for regulatory work, I recommend several next steps. First, more *in-situ* data at elevated chl *a* abundances should be collected to refine the pigment abundance regression algorithm output. Second, SPAMM should be tested on other shallow, clear, and cobble-bottomed rivers with a history of *Cladophora* growth as an assessment of generalizability. Finally, the wavelengths identified throughout this work should be used in the design of a multispectral algae imager that can be used to augment current water quality monitoring stations.

REFERENCES CITED

[1] V Watson, P Berlind, and L Bahls. Control of algal standing crop by p and n in the clark fork river. *Clark Fork River Symposium*, page 47–62, 1990.

[2] Montana Department of Environmental Quality. Clark fork river operable unit (ou 3) mlt reservoir/clark fork river superfund site cerlis mtd980717565 explanation of significant differences (esd), 2015.

[3] National Science Foundation. Award # 1757351 - rii track-1 consortium for research on environmental water systems, 2018.

[4] Riley D Logan, Madison A Torrey, Rafael Feijó-Lima, Benjamin P Colman, H Maurice Valett, and Joseph A Shaw. Uav-based hyperspectral imaging for river algae pigment estimation. *Remote Sensing*, 15(12):3148, 2023.

[5] Vicki Watson and Bonnie Gestring. Monitoring algae levels in the clark fork river. *Intermountain Journal of Sciences*, 2:17–26, 1996.

[6] B. A. Whitton. Biology of cladophora in freshwaters. *Water Research*, 4:457–476, 1970.

[7] Barry J.F. Biggs and Geoff M. Price. A survey of filamentous algal proliferations in new zealand rivers. *New Zealand Journal of Marine and Freshwater Research*, 21:175–191, 1987.

[8] O. A. Timoshkin, D. P. Samsonov, M. Yamamuro, M. V. Moore, O. I. Belykh, V. V. Malnik, M. V. Sakirko, A. A. Shirokaya, N. A. Bondarenko, V. M. Domysheva, G. A. Fedorova, A. I. Kochetkov, A. V. Kuzmin, A. G. Lukhnev, O. V. Medvezhonkova, A. V. Nepokrytykh, E. M. Pasynkova, A. E. Poberezhnaya, N. V. Potapskaya, N. A. Rozhkova, N. G. Sheveleva, I. V. Tikhonova, E. M. Timoshkina, I. V. Tomberg, E. A. Volkova, E. P. Zaitseva, Yu M. Zvereva, A. B. Kupchinsky, and N. A. Bukshuk. Rapid ecological change in the coastal zone of lake baikal (east siberia): Is the site of the world's greatest freshwater biodiversity in danger? *Journal of Great Lakes Research*, 42:487–497, 6 2016.

[9] Richard L. Whitman, Dawn A. Shively, Heather Pawlik, Meredith B. Nevers, and Muruleedhara N. Byappanahalli. Occurrence of escherichia coli and enterococci in cladophora (chlorophyta) in nearshore water and beach sand of lake michigan. *Applied and Environmental Microbiology*, 69:4714–4719, 8 2003.

[10] M. P. Verhougstraete, M. N. Byappanahalli, J. B. Rose, and R. L. Whitman. Cladophora in the great lakes: Impacts on beach water quality and human health, 2010.

[11] Amy Vanden Heuvel, Colleen McDermott, Robert Pillsbury, Todd Sandrin, Julie Kinzelman, John Ferguson, Michael Sadowsky, Muruleedhara Byappanahalli, Richard Whitman, and Gregory T. Kleinheinz. The green alga, cladophora , promotes escherichia coli growth and contamination of recreational waters in lake michigan. *Journal of Environmental Quality*, 39:333–344, 1 2010.

[12] Laurie L Richardson. Remote sensing of algal bloom dynamics. *BioScience*, 46:492–501, 1996.

[13] Lin Li, Rebecca E. Sengpiel, Denise L. Pascual, Lenore P. Tedesco, Jeffrey S. Wilson, and Andemmanuel Soyeux. Using hyperspectral remote sensing to estimate chlorophyll-a and phycocyanin in a mesotrophic reservoir. *International Journal of Remote Sensing*, 31:4147–4162, 2010.

[14] Mohammad Haji Gholizadeh, Assefa M. Melesse, and Lakshmi Reddi. A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors (Switzerland)*, 16, 2016.

[15] Adam N Griggs, Gordon Michael Selckmann, James Cummins, and Claire Buchanan. Methods for estimating filamentous algae cover in streams and rivers of the shenandoah river basin. *Final Report. US EPA Region*, 3:33, 2015.

[16] David Bendersky, Andrew J Winfrey, and Edward T Fago. Application of aerospace-generated technology to water pollution and other public sector problems, 1969.

[17] Richard R Anderson, National Aeronatics, and Space Administrtation. Hydrology spectral reflectance characteristics and automated data reduction techniques which identify wetland and water quality conditions in the chesapeake bay. 1970.

[18] Fabian Casimer Polcyn and R. A. Rollin. Remote sensing techniques for the location and measurement of shallow-water features. 1969.

[19] A. Gitelson. The peak near 700 nm on radiance spectra of algae and water: Relationships of its magnitude and position with chlorophyll. *International Journal of Remote Sensing*, 13:3367–3373, 1992.

[20] A. Gitelson, G. Garbuzov, F. Szilagyi, K. H. Mittenzwey, A. Karnieli, and A. Kaiser. Quantitative remote sensing methods for real-time monitoring of inland waters quality. *International Journal of Remote Sensing*, 14:1269–1295, 1993.

[21] DNH Horler, Mo Dockray, and J Barber. The red edge of plant leaf reflectance. *International journal of remote sensing*, 4(2):273–288, 1983.

[22] W. Stanley Wilson, Jean-Louis Fellous, Kawamura H., and Leonid Mitnik. *A History of Oceanography from Space*, pages 1–31. 01 2006.

[23] W. A. Hovis, D. K. Clark, F. Anderson, R. W. Austin, W. H. Wilson, E. T. Baker, D. Ball, H. R. Gordon, J. L. Mueller, S. Z. El-Sayed, B. Sturm, R. C. Wrigley, and C. S. Yentsch. Nimbus-7 coastal zone color scanner: System description and initial imagery. *Science*, 210(4465):60–63, 1980.

[24] R. W. Johnson and R. C. Harriss. Remote sensing for water quality and biological measurements in coastal waters. *Photogrammetric Engineering and Remote Sensing*, 46:77–85, 1980.

[25] G J PRANGSMA and J N ROOZEKRANS. Using noaa avhrr imagery in assessing water quality parameters. *International Journal of Remote Sensing*, 10:811–818, 1989.

[26] Stanford B. Hooker and Wayne E. Esaias. An overview of the seawifs project. *Eos, Transactions American Geophysical Union*, 74:241–246, 1993.

[27] S.B. Hooker, Goddard Space Flight Center, United States. National Aeronautics, and Space Administration. *SeaWiFS Technical Report Series: An overview of SeaWiFS and ocean color*. NASA technical memorandum. National Aeronautics and Space Administration, 1992.

[28] John E. O'Reilly, Stphane Maritorena, B. Greg Mitchell, David A. Siegel, Kendall L. Carder, Sara A. Garver, Mati Kahru, and Charles McClain. Ocean color chlorophyll algorithms for seawifs. *Journal of Geophysical Research: Oceans*, 103:24937–24953, 1998.

[29] D. A. Siegel, M. J. Behrenfeld, S. Maritorena, C. R. McClain, D. Antoine, S. W. Bailey, P. S. Bontempi, E. S. Boss, H. M. Dierssen, S. C. Doney, R. E. Eplee, R. H. Evans, G. C. Feldman, E. Fields, B. A. Franz, N. A. Kuring, C. Mengelt, N. B. Nelson, F. S. Patt, W. D. Robinson, J. L. Sarmiento, C. M. Swan, P. J. Werdell, T. K. Westberry, J. G. Wilding, and J. A. Yoder. Regional to global assessments of phytoplankton dynamics from the seawifs mission. *Remote Sensing of Environment*, 135:77–91, 2013.

[30] M. J. Behrenfeld, J. T. Randerson, C. R. McClain, G. C. Feldman, S. O. Los, C. J. Tucker, P. G. Falkowski, C. B. Field, R. Frouin, W. E. Esaias, D. D. Kolber, and N. H. Pollack. Biospheric primary production during an enso transition. *Science*, 291:2594–2597, 2001.

[31] Watson W. Gregg and Nancy W. Casey. Global and regional evaluation of the seawifs chlorophyll data set. *Remote Sensing of Environment*, 93:463–479, 2004.

[32] John R.G. Townshend and Christopher O. Justice. Towards operational monitoring of terrestrial systems by moderate-resolution remote sensing. *Remote Sensing of Environment*, 83:351–359, 2002.

[33] C. O. Justice, J. R.G. Townshend, E. F. Vermote, E. Masuoka, R. E. Wolfe, N. Saleous, D. P. Roy, and J. T. Morisette. An overview of modis land data processing and product status. *Remote Sensing of Environment*, 83:3–15, 2002.

[34] S. Novoa, G. Chust, Y. Sagarminaga, M. Revilla, A. Borja, and J. Franco. Water quality assessment using satellite-derived chlorophyll-a within the european directives, in the southeastern bay of biscay. *Marine Pollution Bulletin*, 64:739–750, 2012.

[35] Hae Cheol Kim, Seunghyun Son, Yong Hoon Kim, Jong Seong Khim, Jungho Nam, Won Keun Chang, Jung Ho Lee, Chang Hee Lee, and Jongseong Ryu. Remote sensing and water quality indicators in the korean west coast: Spatio-temporal structures of modis-derived chlorophyll-a and total suspended solids. *Marine Pollution Bulletin*, 121:425–434, 2017.

[36] Chuanmin Hu, Zhiqiang Chen, Tonya D. Clayton, Peter Swarzenski, John C. Brock, and Frank E. Muller-Karger. Assessment of estuarine water-quality indicators using modis medium-resolution bands: Initial results from tampa bay, fl. *Remote Sensing of Environment*, 93:423–441, 2004.

[37] F. Gohin, J. N. Druon, and L. Lampert. A five channel chlorophyll concentration algorithm applied to sea wifs data processed by seadas in coastal waters. *International Journal of Remote Sensing*, 23:1639–1661, 2002.

[38] P. A. Brivio, C. Giardino, and E. Zilioli. Determination of chlorophyll concentration changes in lake garda using an image-based radiative transfer code for landsat tm images. *International Journal of Remote Sensing*, 22:487–502, 2001.

[39] Yosef Z. Yacobi, Anatoly Gitelson, and Meir Mayo. Remote sensing of chlorophyll in lake kinneret using highspectral-resolution radiometer and landsat tm: Spectral features of reflectance and algorithm development. *Journal of Plankton Research*, 17:2155–2173, 1995.

[40] Ian M. McCullough, Cynthia S. Loftin, and Steven A. Sader. High-frequency remote monitoring of large lakes with modis 500m imagery. *Remote Sensing of Environment*, 124:234–241, 2012.

[41] Essayas K. Ayana, Abeyou W. Worqlul, and Tammo S. Steenhuis. Evaluation of stream water quality data generated from modis images in modeling total suspended solid emission to a freshwater lake. *Science of the Total Environment*, 523:170–177, 2015.

[42] Africa I. Flores-Anderson, Robert Griffin, Margaret Dix, Claudia S. Romero-Oliva, Gerson Ochaeta, Juan Skinner-Alvarado, Maria Violeta Ramirez Moran, Betzy Hernandez, Emil Cherrington, Benjamin Page, and Flor Barreno. Hyperspectral satellite remote sensing of water quality in lake atitlán, guatemala. *Frontiers in Environmental Science*, 8, 2020.

[43] Wesley J. Moses, Anatoly A. Gitelson, Richard L. Perk, Daniela Gurlin, Donald C. Rundquist, Bryan C. Leavitt, Tadd M. Barrow, and Paul Brakhage. Estimation of chlorophyll-a concentration in turbid productive waters using airborne hyperspectral data. *Water Research*, 46:993–1004, 2012.

[44] K. Kallio, S. Koponen, and J. Pulliainen. Feasibility of airborne imaging spectrometry for lake monitoring - a case study of spatial chlorophyll a distribution in two meso-eutrophic lakes. *International Journal of Remote Sensing*, 24:3771–3790, 2003.

[45] Sampsa Koponen, Jouni Pulliainen, Kari Kallio, and Martti Hallikainen. Lake water quality classification with airborne hyperspectral spectrometer and simulated meris data. *Remote Sensing of Environment*, 79:51–59, 2002.

[46] Giorgio Dall'Olmo, Anatoly A. Gitelson, and Donald C. Rundquist. Towards a unified approach for remote estimation of chlorophyll-a in both terrestrial vegetation and turbid productive waters. *Geophysical Research Letters*, 30:8–11, 2003.

[47] Leif G. Olmanson, Patrick L. Brezonik, and Marvin E. Bauer. Airborne hyperspectral remote sensing to assess spatial distribution of water quality characteristics in large rivers: The mississippi river and its tributaries in minnesota. *Remote Sensing of Environment*, 130:254–265, 2013.

[48] Naseer A Shafique, Florence Fulk, Bradley C Autrey, and Joseph Flotemersch. Hyperspectral remote sensing of water quality parameters for large rivers in the ohio river basin. pages 216–221, 2003.

[49] L.W. Harding, W.D. Miller, R.N. Swift, and C.W. Wright. Aircraft remote sensing. In John H. Steele, editor, *Encyclopedia of Ocean Sciences (Second Edition)*, pages 138–146. Academic Press, Oxford, second edition edition, 2001.

[50] Carmen Cillero Castro, Jose Antonio Domínguez Gómez, Jordi Delgado Martín, Boris Alejandro Hinojo Sánchez, Jose Luis Cereijo Arango, Federico Andrés Cheda Tuya, and Ramon Díaz-Varela. An uav and satellite multispectral data approach to monitor water quality in small reservoirs. *Remote Sensing*, 12, 5 2020.

[51] Min Xu, Hongxing Liu, Richard Beck, John Lekki, Bo Yang, Song Shu, Yang Liu, Teresa Benko, Robert Anderson, Roger Tokars, Richard Johansen, Erich Emery, and Molly Reif. Regionally and locally adaptive models for retrieving chlorophyll-a concentration in inland waters from remotely sensed multispectral and hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 57(7):4758–4774, 2019.

[52] Yong Sung Kwon, Jong Cheol Pyo, Yong Hwan Kwon, Hongtao Duan, Kyung Hwa Cho, and Yongeun Park. Drone-based hyperspectral remote sensing of cyanobacteria using vertical cumulative pigment concentration in a deep reservoir. *Remote Sensing of Environment*, 236, 1 2020.

[53] Yingying Gai, Dingfeng Yu, Yan Zhou, Lei Yang, Chao Chen, and Jun Chen. An improved model for chlorophyll-a concentration retrieval in coastal waters based on uav-borne hyperspectral imagery: A case study in qingdao, china. *Water*, 12(10), 2020.

[54] Qi Cao, Gongliang Yu, Shengjie Sun, Yong Dou, Hua Li, and Zhiyi Qiao. Monitoring water quality of the haihe river based on ground-based hyperspectral remote sensing. *Water (Switzerland)*, 14, 1 2022.

[55] Qikai Lu, Wei Si, Lifei Wei, Zhongqiang Li, Zhihong Xia, Song Ye, and Yu Xia. Retrieval of water quality from uav-borne hyperspectral imagery: A comparative study of machine learning algorithms. *Remote Sensing*, 13, 10 2021.

[56] Hong Liu, Tao Yu, Bingliang Hu, Xingsong Hou, Zhoufeng Zhang, Xiao Liu, Jiacheng Liu, Xueji Wang, Jingjing Zhong, Zhengxuan Tan, Shaoxia Xia, and Bao Qian. Uav-borne hyperspectral imaging remote sensing system based on acousto-optic tunable filter for water quality monitoring. *Remote Sensing*, 13, 10 2021.

[57] Yi Xiao, Yahui Guo, Guodong Yin, Xuan Zhang, Yu Shi, Fanghua Hao, and Yongshuo Fu. Uav multispectral image-based urban river water quality monitoring using stacked ensemble machine learning algorithms—a case study of the zhanghe river, china. *Remote Sensing*, 14, 7 2022.

[58] Eun Ju Kim, Sook Hyun Nam, Jae Wuk Koo, and Tae Mun Hwang. Hybrid approach of unmanned aerial vehicle and unmanned surface vehicle for assessment of chlorophyll-a imagery using spectral indices in stream, south korea. *Water (Switzerland)*, 13, 2021.

[59] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

[60] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[61] Tianqi Chen, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, Kailong Chen, Rory Mitchell, Ignacio Cano, Tianyi Zhou, et al. Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4):1–4, 2015.

[62] Anna Veronika Dorogush, Vasily Ershov, and Andrey Gulin. Catboost: gradient boosting with categorical features support. *arXiv preprint arXiv:1810.11363*, 2018.

[63] Leo Breiman. Random forests. *Machine Learning*, 45:5–32, 10 2001.

[64] Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine learning*, 63:3–42, 2006.

[65] Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.

[66] Geoffrey E Hinton. Connectionist learning procedures. In *Machine learning*, pages 555–610. Elsevier, 1990.

[67] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2):301–320, 2005.

[68] Michael W Suplee, Vicki Watson, Mark Teply, and Heather McKee. How green is too green? public opinion of what constitutes undesirable algae levels in streams 1. *JAWRA Journal of the American Water Resources Association*, 45(1):123–140, 2009.

[69] Kyle F. Flynn and Steven C. Chapra. Remote sensing of submerged aquatic vegetation in a shallow non-turbid river using an unmanned aerial vehicle. *Remote Sensing*, 6:12815–12836, 2014.

[70] Guerric le Maire, Christophe François, Kamel Soudani, Daniel Berveiller, Jean Yves Pontailler, Nathalie Bréda, Hélène Genet, Hendrik Davi, and Eric Dufrêne. Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. *Remote Sensing of Environment*, 112:3846–3864, 2008.

[71] Chaoyang Wu, Zheng Niu, Quan Tang, and Wenjiang Huang. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agricultural and Forest Meteorology*, 148:1230–1241, 2008.

[72] C.S.T Daughtry, C.L Walthall, M.S Kim, E.Brown de Colstoun, and J.E McMurtrey. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment*, 74(2):229–239, 2000.

[73] Geneviève Rondeaux, Michael Steven, and Frédéric Baret. Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 55(2):95–107, 1996.

[74] Yao Zhang, Jian Hui, Qiming Qin, Yuanheng Sun, Tianyuan Zhang, Hong Sun, and Minzan Li. Transfer-learning-based approach for leaf chlorophyll content estimation of winter wheat from hyperspectral data. *Remote Sensing of Environment*, 267(July):112724, 2021.

[75] Carl J. Legleiter, Dar A. Roberts, W. Andrew Marcus, and Mark A. Fonstad. Passive optical remote sensing of river channel morphology and in-stream habitat: Physical basis and feasibility. *Remote Sensing of Environment*, 93:493–510, 2004.

[76] Carl J. Legleiter, Dar A. Roberts, and Rick L. Lawrence. Spectrally based remote sensing of river bathymetry. *Earth Surface Processes and Landforms*, 34:613–628, 2009.

[77] Vladimir I. Haltrin. *Absorption and scattering of light in natural waters*, pages 445–486. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[78] H. Buiteveld, J.H.M. Hakvoort, and M. Donze. Optical properties of pure water. volume 2258, pages 9–93, 1994.

[79] W. K. Dodds, V. H. Smith, and B. Zander. Developing nutrient targets to control benthic chlorophyll levels in streams: A case study of the clark fork river. *Water Research*, 31:1738–1750, 1997.

[80] Michael W Suplee, Vicki Watson, Walter K Dodds, and Chris Shirley. Response of algal biomass to large-scale nutrient controls in the clark fork river, montana, united states. *Resources Association (JAWRA)*, 48:1008–1021, 10 2012.

[81] Montana Department of Environmantal Quality. 17.30.631 numeric algal biomass and nutrient standards. `https://rules.mt.gov/gateway/ruleno.asp?RN=17%2E30%2E631`, 2002. Online; accessed 31 May 2023.

[82] Alan D Steinman, Gary A Lamberti, Peter R Leavitt, and Donald G Uzarski. Biomass and pigments of benthic algae. In *Methods in Stream Ecology, Volume 1*, pages 223–241. Elsevier, 2017.

[83] Simon N Topp, Tamlin M Pavelsky, Daniel Jensen, Marc Simard, and Matthew RV Ross. Research trends in the use of remote sensing for inland water quality science: Moving towards multidisciplinary applications. *Water*, 12(1):169, 2020.

[84] Gui Feng Wei, Dan Ling Tang, and Sufen Wang. Distribution of chlorophyll and harmful algal blooms (habs): A review on space based studies in the coastal environments of chinese marginal seas. *Advances in Space Research*, 41:12–19, 2008.

[85] S. Sathyendranath, G. Cota, V. Stuart, H. Maass, and T. Platt. Remote sensing of phytoplankton pigments: A comparison of empirical and theoretical approaches. *International Journal of Remote Sensing*, 22:249–273, 2001.

[86] S. Sathyendranath, L. Prieur, and A. Morel. A three-component model of ocean colour and its application to remote sensing of phytoplankton pigments in coastal waters. *International Journal of Remote Sensing*, 10:1373–1394, 1989.

[87] Vittorio E. Brando and Arnold G. Dekker. Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality. *IEEE Transactions on Geoscience and Remote Sensing*, 41:1378–1387, 6 2003.

[88] S. Ekstrand. Landsat tm based quantification of chlorophyll-a during algae blooms in coastal waters. *International Journal of Remote Sensing*, 13:1913–1926, 1992.

[89] Mark William Matthews, Stewart Bernard, and Lisl Robertson. An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. *Remote Sensing of Environment*, 124:637–652, 9 2012.

[90] Nima Pahlevan, Brandon Smith, Caren Binding, Daniela Gurlin, Lin Li, Mariano Bresciani, and Claudia Giardino. Hyperspectral retrievals of phytoplankton absorption and chlorophyll-a in inland and nearshore coastal waters. *Remote Sensing of Environment*, 253:112200, 2021.

[91] Michael Papenfus, Blake Schaeffer, Amina I. Pollard, and Keith Loftin. Exploring the potential value of satellite remote sensing to monitor chlorophyll-a for us lakes and reservoirs. *Environmental Monitoring and Assessment*, 192, 12 2020.

[92] Stefan G H Simis, Steef W M Peters, and Herman J Gons. Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. *Limnol. Oceanogr*, 50:237–245, 2005.

[93] Karen Anderson and Kevin J. Gaston. Lightweight unmanned aerial vehicles will revolutionize spatial ecology, 4 2013.

[94] K. H. Cheng, S. N. Chan, and Joseph H.W. Lee. Remote sensing of coastal algal blooms using unmanned aerial vehicles (uavs). *Marine Pollution Bulletin*, 152, 3 2020.

[95] Di Wu, Ruopu Li, Feiyang Zhang, and Jia Liu. A review on drone-based harmful algae blooms monitoring. *Environmental Monitoring and Assessment*, 191, 4 2019.

[96] Tyler King, Stephen Hundt, Konrad Hafen, Victoria Stengel, and Scott Ducar. Mapping the probability of freshwater algal blooms with various spectral indices and sources of training data. *Journal of Applied Remote Sensing*, 16, 2022.

[97] Lucas Silveira Kupssinskü, Tainá Thomassim Guimarães, Eniuce Menezes De Souza, Daniel C. Zanotta, Mauricio Roberto Veronez, Luiz Gonzaga, and Frederico Fábio Mauad. A method for chlorophyll-a and suspended solids prediction through remote sensing and machine learning. *Sensors (Switzerland)*, 20, 2020.

[98] Neha Kureel, Jyoti Sarup, Shafique Matin, Suresh Goswami, and Kapil Kureel. Modelling vegetation health and stress using hyperspectral remote sensing data. *Modeling Earth Systems and Environment*, 8:733–748, 2022.

[99] Russell Main, Moses Azong Cho, Renaud Mathieu, Martha M. O'Kennedy, Abel Ramoelo, and Susan Koch. An investigation into robust spectral indices for leaf chlorophyll estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66:751–761, 2011.

[100] Caren Binding, Tracie Greenberg, and Robert Bukata. The meris maximum chlorophyll index; its merits and limitations for inland water algal bloom monitoring. *Journal of Great Lakes Research*, 39:100–107, 12 2013.

[101] Yaohuan Huang, Dong Jiang, Dafang Zhuang, and Jingying Fu. Evaluation of hyperspectral indices for chlorophyll-a concentration estimation in tangxun lake (wuhan, china). *International Journal of Environmental Research and Public Health*, 7:2437–2451, 2010.

[102] Stuart G. Fisher, Lawrence J. Gray, Nancy B. Grimm, and David E. Busch. Temporal succession in a desert stream ecosystem following flash flooding. *Ecological Monographs*, 52(1):93–110, 1982.

[103] G Wayne Minshall, Robert C Petersen, Kenneth W Cummins, Thomas L Bott, James R Sedell, Colbert E Cushing, and Robin L Vannote. Interbiome comparison of stream ecosystem dynamics. *Ecological monographs*, 53(1):1–25, 1983.

[104] Michael E McClain, Jeffrey E Richey, Jay A Brandes, and Tania P Pimentel. Dissolved organic matter and terrestrial-lotic linkages in the central amazon basin of brazil. *Global Biogeochemical Cycles*, 11(3):295–311, 1997.

[105] Roberto D Garcia, Mariana Reissig, Claudia P Queimaliños, Patricia E Garcia, and Maria C Dieguez. Climate-driven terrestrial inputs in ultraoligotrophic mountain streams of andean patagonia revealed through chromophoric and fluorescent dissolved organic matter. *Science of the total environment*, 521:280–292, 2015.

[106] Robin L. Vannote, G. Wayne Minshall, Kenneth W. Cummins, James R. Sedell, and Colbert E. Cushing. The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences*, 37(1):130–137, 1980.

[107] James H. Thorp and Michael D. Delong. The riverine productivity model: An heuristic view of carbon sources and organic processing in large river ecosystems. *Oikos*, 70(2):305–308, 1994.

[108] Cathy M Tate. Patterns and controls of nitrogen in tallgrass prairie streams. *Ecology*, 71(5):2007–2018, 1990.

[109] H Maurice Valett, Marc Peipoch, and Geoffrey C Poole. Nutrient processing domains: Spatial and temporal patterns of material retention in running waters. *Freshwater Science*, 41(2):000–000, 2022.

[110] G Wayne Minshall. Autotrophy in stream ecosystems. *BioScience*, 28(12):767–771, 1978.

[111] Robert O Hall Jr, Charles B Yackulic, Theodore A Kennedy, Michael D Yard, Emma J Rosi-Marshall, Nicholas Voichick, and Kathrine E Behn. Turbidity, light, temperature, and hydropeaking control primary productivity in the c olorado river, g rand c anyon. *Limnology and Oceanography*, 60(2):512–526, 2015.

[112] Jacques C. Finlay. Patterns and controls of lotic algal stable carbon isotope ratios. *Limnology and Oceanography*, 49:850–861, 2004.

[113] ES Bernhardt, JB Heffernan, NB Grimm, EH Stanley, JW Harvey, Maite Arroita, Alison P Appling, MJ Cohen, WH McDowell, RO Hall Jr, et al. The metabolic regimes of flowing waters. *Limnology and Oceanography*, 63(S1):S99–S118, 2018.

[114] Tom J Battin, Ronny Lauerwald, Emily S Bernhardt, Enrico Bertuzzo, Lluís Gómez Gener, Robert O Hall Jr, Erin R Hotchkiss, Taylor Maavara, Tamlin M Pavelsky, Lishan Ran, et al. River ecosystem metabolism and carbon biogeochemistry in a changing world. *Nature*, 613(7944):449–459, 2023.

[115] Joanna R Blaszczak, Charles B Yackulic, Robert K Shriver, and Robert O Hall Jr. Models of underlying autotrophic biomass dynamics fit to daily river ecosystem productivity estimates improve understanding of ecosystem disturbance and resilience. *Ecology Letters*, 26(9):1510–1522, 2023.

[116] Gary L Ingman and Mark A Kerr. Nutrient sources in the clark fork river basin. 1990.

[117] Kirk Lohman and John C. Priscu. Physiological indicators of nutrient deficiency in cladophora chlorophyta. *Journal of Phycology*, 28:443–448, 1992.

[118] Nicholas J Banish. Factors influencing cladophora biomass abundance factors influencing cladophora biomass abundance in the upper clark fork river, montana. Master's thesis, 2017.

[119] M Power, R Lowe, P Furey, J Welter, M Limm, J Finlay, C Bode, S Chang, Maria Goodrich, and J Sculley. Algal mats and insect emergence in rivers under mediterranean climates: towards photogrammetric surveillance. *Freshwater Biology*, 54(10):2101–2115, 2009.

[120] Mary E Power, Keith Bouma-Gregson, Patrick Higgins, and Stephanie M Carlson. The thirsty eel: summer and winter flow thresholds that tilt the eel river of northwestern california from salmon-supporting to cyanobacterially degraded states. *Copeia*, 103(1):200–211, 2015.

[121] Paula C Furey, Rex L Lowe, Mary E Power, and Alexis M Campbell-Craven. Midges, cladophora, and epiphytes: shifting interactions through succession. *Freshwater Science*, 31(1):93–107, 2012.

[122] DJI. Matrice 600 pro - user manual, 2022.

[123] Resonon. Airborne user manual, 2022.

[124] R J Ritchie. Universal chlorophyll equations for estimating chlorophylls a, b, c, and d and total chlorophylls in natural assemblages of photosynthetic organisms using acetone, methanol, or ethanol solvents. *PHOTOSYNTHETICA*, 46:115–126, 2008.

[125] Timothy P Moulton, Marcelo L Souza, Taldi L Walter, and Felipe AM Krsulović. Patterns of periphyton chlorophyll and dry mass in a neotropical stream: a cheap and rapid analysis using a hand-held fluorometer. *Marine and Freshwater Research*, 60(3):224–233, 2009.

[126] Bruno Cremella, Yannick Huot, and Sylvia Bonilla. Interpretation of total phytoplankton and cyanobacteria fluorescence from cross-calibrated fluorometers, including sensitivity to turbidity and colored dissolved organic matter. *Limnology and Oceanography: Methods*, 16(12):881–894, 2018.

[127] Joseph A Shaw. The effect of instrument polarization sensitivity on sea surface remote sensing with infrared spectroradiometers. *Journal of Atmospheric and Oceanic Technology*, 19(5):820–827, 2002.

[128] Riley D Logan, Erica Venkatesulu, and Joseph A Shaw. Comparing the polarimetric response of hyperspectral imagers. In *Polarization Science and Remote Sensing X*, volume 11833, pages 90–97. SPIE, 2021.

[129] Pasi Ylöstalo, Kari Kallio, and Jukka Seppälä. Absorption properties of in-water constituents and their variation among various lake types in the boreal region. *Remote Sensing of Environment*, 148:190–205, 2014.

[130] Dariusz Karcz, Bozena Boroń, Arkadiusz Matwijczuk, Justyna Furso, Jakub Staroń, Alicja Ratuszna, and Leszek Fiedor. Lessons from chlorophylls: Modifications of porphyrinoids towards optimized solar energy conversion. *Molecules*, 19:15938–15954, 2014.

[131] Nguyen Thi Thu Ha, Nguyen Thien Phuong Thao, Katsuaki Koike, and Mai Trong Nhuan. Selecting the best band ratio to estimate chlorophyll-a concentration in a tropical freshwater lake using sentinel 2a images from a case study of lake ba be (northern vietnam). *ISPRS International Journal of Geo-Information*, 6, 9 2017.

[132] Compton J Tucker. Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, 8(2):127–150, 1979.

[133] Eric Vermote, Chris Justice, Martin Claverie, and Belen Franch. Preliminary analysis of the performance of the landsat 8/oli land surface reflectance product. *Remote Sensing of Environment*, 185:46–56, 2016.

[134] Hyun Jung Cho, Philemon Kirui, and Harene Natarajan. Test of multi-spectral vegetation index for floating and canopy-forming submerged vegetation. *International Journal of Environmental Research and Public Health*, 5(5):477–483, 2008.

[135] Alexandria B. Boehm, Niveen S. Ismail, Lauren M. Sassoubre, and Elizabeth A. Andruszkiewicz. Oceans in peril: Grand challenges in applied water quality research for the 21st century. *Environmental Engineering Science*, 34:3–15, 1 2017.

[136] Didem Gökçe. Algae as an indicator of water quality. *Algae-Organisms for Imminent Biotechnology*, pages 81–101, 2016.

[137] Walter K. Dodds and Dolly A. Gudder. The ecology of cladophora. *Journal of Phycology*, 28:415–427, 8 1992.

[138] Marta Pikosz, Beata Messyasz, and Maciej Gabka. Functional structure of algal mat (cladophora glomerata) in a freshwater in western poland. *Ecological Indicators*, 74:1–9, 3 2017.

[139] J.M. Burkholder. Harmful algal blooms. In Gene E. Likens, editor, *Encyclopedia of Inland Waters*, pages 264–285. Academic Press, Oxford, 2009.

[140] H. Maurice Valett, Rafael Feijó de Lima, Marc Peipoch, and Royce C. Engstrom. Bloom succession and nitrogen dynamics during snowmelt in a mid-order montane river. *Biogeochemistry*, pages 1–20, 9 2023.

[141] F E Round. Diatoms in river water-monitoring studies. *Journal of Applied Phycology*, 3:129–145, 1991.

[142] Montana Department of Environmental Quality. Base numeric nutrient standards implementation guidance version 3.0, 2019.

[143] Montana Department of Environmental Quality. Deq-7 montana numeric water quality standards, 2019.

[144] Cameron Murray, Albert Larson, Joseph Goodwill, Yeqiao Wang, Dawn Cardace, and Ali S Akanda. Water quality observations from space: A review of critical issues and challenges. *Environments*, 9(10):125, 2022.

[145] Mohammad Haji Gholizadeh, Assefa M Melesse, and Lakshmi Reddi. Spaceborne and airborne sensors in water quality assessment. *International Journal of Remote Sensing*, 37(14):3143–3180, 2016.

[146] M Suplee, R Sada, K Makarowski, and G Steffens. Sample collection and laboratory analysis of chlorophyll-a standard operation procedure. *Montana Department of Environmental Quality. WQPBWQM-011, Version 8.0*, 2021.

[147] Donald F Charles, Candia Knowles, and Robin S Davis. Protocols for the analysis of algal samples collected as part of the us geological survey national water-quality assessment program. *Patrick Center for Environmental Research Report*, (02-06), 2002.

[148] SD Porter, TF Cuffney, ME Gurtz, and MR Meador. Methods for collecting algal samples as part of the national water-quality assessment program. open-file report 93–409. *US Geological Survey, Raleigh, NC*, 1993.

[149] Victoria G. Stengel, Jessica M. Trevino, Tyler V. King, Scott D. Ducar, Stephen A. Hundt, Konrad C. Hafen, and Christopher J. Churchill. Near real-time satellite detection and monitoring of aquatic algae and cyanobacteria: how a combination of chlorophyll-a indices and water-quality sampling was applied to north texas reservoirs. *Journal of Applied Remote Sensing*, 17, 11 2023.

[150] Bruno Munhoz Araújo, Rogério Galante Negri, Pedro Henrique Moraes Ananias, Adriano Bressane, and Suzelei Rodgher. Remotely sensed-based analysis about climatic and landscape change effects on phytoplankton bloom in barra bonita reservoir (são paulo state, brazil). *Journal of Applied Remote Sensing*, 17, 2 2023.

[151] Jayesh Ganpat Ghatkar, Rakesh Kumar Singh, and Palanisamy Shanmugam. Classification of algal bloom species from remote sensing data using an extreme gradient boosted decision tree model. *International Journal of Remote Sensing*, 40:9412–9438, 12 2019.

[152] Yu Hwan Ahn, Palanisamy Shanmugam, Joo Hyung Ryu, and Jong Chul Jeong. Satellite detection of harmful algal bloom occurrences in korean waters. *Harmful Algae*, 5:213–231, 3 2006.

[153] Carl J. Legleiter, Tyler V. King, Kurt D. Carpenter, Natalie C. Hall, Adam C. Mumford, Terry Slonecker, Jennifer L. Graham, Victoria G. Stengel, Nancy Simon, and Barry H. Rosen. Spectral mixture analysis for surveillance of harmful algal blooms (smash): A field-, laboratory-, and satellite-based approach to identifying cyanobacteria genera from remotely sensed data. *Remote Sensing of Environment*, 279:113089, 9 2022.

[154] R. Dwivedi, M. Rafeeq, B. R. Smitha, K. B. Padmakumar, Lathika Cicily Thomas, V. N. Sanjeevan, Prince Prakash, and Mini Raman. Species identification of mixed algal bloom in the northern arabian sea using remote sensing techniques. *Environmental Monitoring and Assessment*, 187:1–11, 2 2015.

[155] Chippie Kislik, Iryna Dronova, and Maggi Kelly. Uavs in support of algal bloom research: A review of current applications and future opportunities. *Drones*, 2:1–14, 12 2018.

[156] Maciej M. Nowak, Katarzyna Dziób, and Paweł Bogawski. Unmanned aerial vehicles (uavs) in environmental biology: A review. *European Journal of Ecology*, 4:56–74, 12 2018.

[157] Chippie Kislik, Laurel Genzoli, Andy Lyons, and Maggi Kelly. Application of uav imagery to detect and quantify submerged filamentous algae and rooted macrophytes in a non-wadeable river. *Remote Sensing*, 12:1–24, 10 2020.

[158] Fleur Visser, Caroline Wallis, and Anne M. Sinnott. Optical remote sensing of submerged aquatic vegetation: Opportunities for shallow clearwater streams. *Limnologica*, 43:388–398, 9 2013.

[159] Kevin Stanfield. Developing methods to differentiate species and estimate coverage of benthic autotrophs in the potomac using digital imaging. Master's thesis, 2018.

[160] Carl J. Legleiter and Shawn W. Hodges. Mapping benthic algae and cyanobacteria in river channels from aerial photographs and satellite images: A proof-of-concept investigation on the buffalo national river, ar, usa. *Remote Sensing 2022, Vol. 14, Page 953*, 14:953, 2 2022.

[161] Andrew Lensen, Harith Al-Sahaf, Mengjie Zhang, and Brijesh Verma. Genetic programming for algae detection in river images. *2015 IEEE Congress on Evolutionary Computation, CEC 2015 - Proceedings*, pages 2468–2475, 9 2015.

[162] Shannon M Hamp, Riley D Logan, and Joseph A Shaw. Developing a low-cost multispectral imager for detecting algal blooms in rivers. In *SPIE Future Sensing Technologies 2023*, volume 12327, pages 100–109. SPIE, 2023.

[163] Abraham Savitzky and Marcel JE Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964.

[164] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[165] Kyle F Flynn and Steven C Chapra. Evaluating hydraulic habitat suitability of filamentous algae using an unmanned aerial vehicle and acoustic doppler current profiler. *Journal of Environmental Engineering*, 146(3):04019126, 2020.

[166] William L Andreen. *The Evolution of Water Pollution Control in the United States-State, Local, and Federal Efforts, 1789-1972: Part I*, volume 22, pages 146–200. 2003.

[167] Claudia Copeland. Clean water act: A summary of the law. Congressional Research Service, Library of Congress Washington, DC, 2016.

[168] USC. Federal water pollution control act - federal-water-pollution-control-act-508full, 2002.

[169] Donald M. Anderson, Joann M. Burkholder, William P. Cochlan, Patricia M. Glibert, Christopher J. Gobler, Cynthia A. Heil, Raphael M. Kudela, Michael L. Parsons, J. E. Jack Rensel, David W. Townsend, Vera L. Trainer, and Gabriel A. Vargo. Harmful algal blooms and eutrophication: Examining linkages from selected coastal regions of the united states. *Harmful Algae*, 8(1):39–53, 2008. HABs and Eutrophication.

[170] Kevin G Sellner, Gregory J Doucette, and Gary J Kirkpatrick. Harmful algal blooms: causes, impacts and detection. *Journal of Industrial Microbiology and Biotechnology*, 30:383–406, 2003.

[171] Hans W Paerl, Rolland S Fulton, Pia H Moisander, Julianne Dyble, et al. Harmful freshwater algal blooms, with an emphasis on cyanobacteria. *The Scientific World Journal*, 1:76–113, 2001.

[172] H Kenneth Hudnell. The state of us freshwater harmful algal blooms assessments, policy and legislation. *Toxicon*, 55(5):1024–1034, 2010.

[173] Wayne A Wurtsbaugh, Hans W Paerl, and Walter K Dodds. Nutrients, eutrophication and harmful algal blooms along the freshwater to marine continuum. *Wiley Interdisciplinary Reviews: Water*, 6(5):e1373, 2019.

[174] Chris Kronberg. Biological effects of water pollution in the clark fork river. 1983.

[175] Riley D Logan and Joseph A Shaw. River algae mapping using hyperspectral imagery from uavs. *Journal of Applied Remote Sensing*, TBA(TBA):3148, 2024.

[176] F Richard Hauer and Gary Lamberti. *Methods in stream ecology*. Academic Press, 2011.

[177] In-Kwon Yeo and Richard A Johnson. A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4):954–959, 2000.

[178] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.

[179] Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.

[180] Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14:199–222, 2004.

[181] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996.

[182] David H Wolpert. Stacked generalization. *Neural networks*, 5(2):241–259, 1992.

[183] Michael E Tipping. Sparse bayesian learning and the relevance vector machine. *Journal of machine learning research*, 1(Jun):211–244, 2001.

[184] U.S. Geological Survey. National water information system, February 2024.

APPENDIX

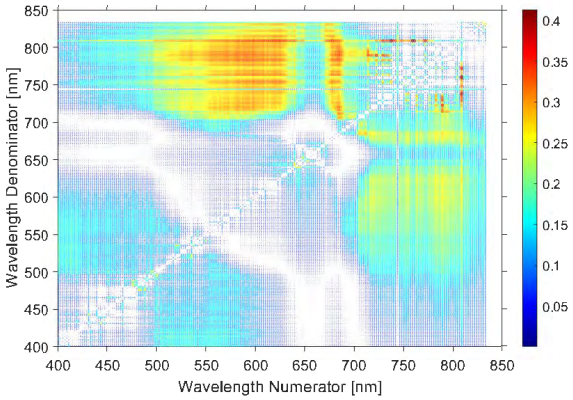
SUPPLEMENTARY FIGURES

Table A.1: Measurements of filamentous chl *a* and phycocyanin for data collected on 17 August 2021 at the Gold Creek and Bear Gulch field sites.

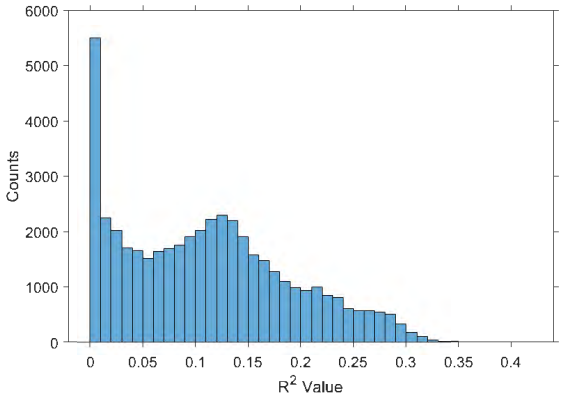
Field Site	**Plot Number**	**Filamentous chl *a* Abundance** $[mg/m^2]$	**Epiphytic chl *a* Abundance** $[mg/m^2]$	**Epilithic chl *a* Abundance** $[mg/m^2]$	**Epiphytic Phycocyanin Abundance** $[mg/m^2]$	**Epilithic Phycocyanin Abundance** $[mg/m^2]$
Gold Creek	1	0	0	18.46	0	52.01
	2	45.82	0	26.14	0	46.52
	3	148.51	10.80	25.56	7.87	40.11
	4	140.62	10.75	23.78	7.98	22.27
	5	207.34	0	27.68	0	20.06
	6	143.72	4.95	16.08	9.17	26.41
	7	267.12	17.26	32.12	13.95	32.40
Bear Gulch	1	163.55	9.60	21.20	43.49	19.66
	2	31.04	12.53	25.72	9.84	23.72
	3	170.84	73.35	34.57	27.52	19.41
	4	174.94	34.53	52.85	20.31	36.70
	5	202.15	47.91	0	29.67	0
	6	67.98	6.46	34.31	4.56	21.58
	7	63.09	3.23	54.72	10.01	30.98

Table A.2: Measurements of filamentous chl *a* and phycocyanin for data collected on 9 September 2021 at the Gold Creek field site.

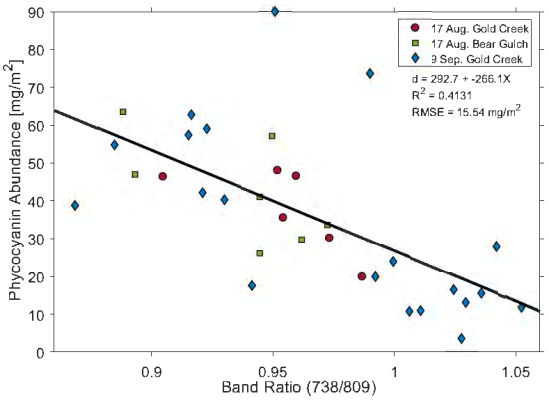
Plot Number	**Filamentous chl *a* Abundance** $[mg/m^2]$	**Epiphytic chl *a* Abundance** $[mg/m^2]$	**Epilithic chl *a* Abundance** $[mg/m^2]$	**Epiphytic Phyco-cyanin Abundance** $[mg/m^2]$	**Epilithic Phyco-cyanin Abundance** $[mg/m^2]$
1	0	13.54	32.54	7.92	30.73
2	28.48	6.21	48.42	1.77	55.43
3	234.45	34.35	42.48	36.16	53.67
4	299.27	11.26	21.99	7.84	20.12
5	145.55	11.88	10.63	5.43	9.98
6	0	6.64	45.20	4.19	37.77
7	44.67	3.50	50.47	1.57	60.94
8	282.72	44.69	39.41	37.41	36.06
9	60.58	2.98	9.61	1.91	8.66
10	105.81	10.88	10.62	4.55	7.07
11	76.16	0	48.27	0	54.66
12	149.62	1.65	44.78	1.35	38.94
13	192.23	6.29	2.14	3.11	0.32
14	345.15	7.98	16.73	5.79	14.16
15	211.01	8.73	9.25	4.88	5.91
16	17.00	2.69	44.05	1.43	57.38
17	100.92	3.75	14.79	3.05	14.37
18	276.19	12.52	14.79	7.26	16.66
19	359.82	7.27	10.67	6.07	6.89
20	59.39	2.57	14.15	1.65	14.62



(a) Heat map of R^2 values generated by fitting each band combination against total phycocyanin standing crops. Colors within the heat map represent the R^2 value, with wavelength numerators and denominators shown on the x- and y-axes, respectively.

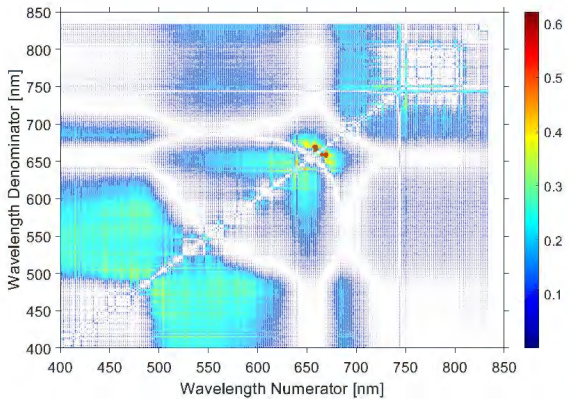


(b) Histogram of the R^2 values generated using the brute-force analysis.

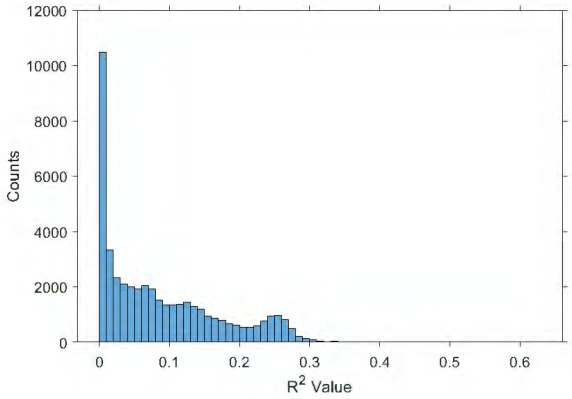


(c) Linear regression analysis between total phycocyanin abundance (y-axis) and optimal band ratio (x-axis). Data collected on 17 August 2021 at the Gold Creek and Bear Gulch site represented by red circles and green squares, respectively, while data collected on 9 September 2021 are shown with blue diamonds.

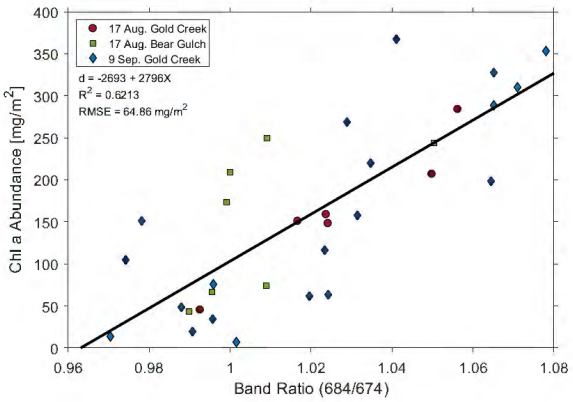
Figure A.1: Analytics from the regression analysis of total phycocyanin standing crops from the combined (all field sites) data set.



(a) Heat map of R^2 values generated by fitting each band combination against fila/epip chl *a* standing crops. Colors within the heat map represent the R^2 value, with wavelength numerators and denominators shown on the x- and y-axes, respectively.

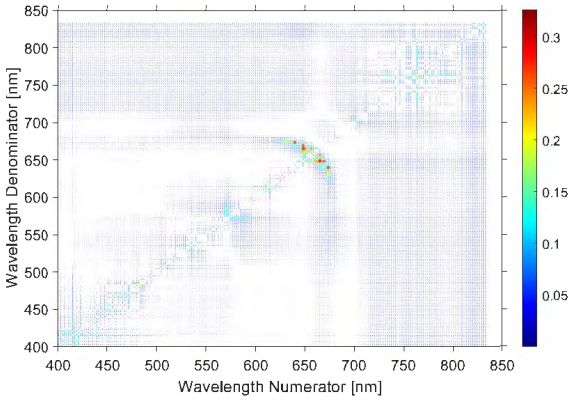


(b) Histogram of the R^2 values generated using the brute-force analysis.

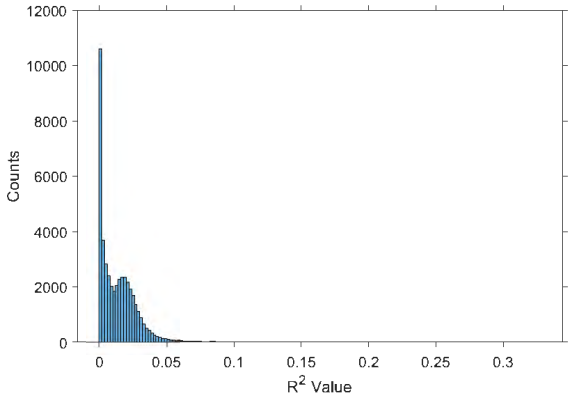


(c) Linear regression analysis between chl *a* abundance (y-axis) and optimal band ratio (x-axis). Data collected on 17 August 2021 at the Gold Creek and Bear Gulch site represented by red circles and green squares, respectively, while data collected on 9 September 2021 are shown with blue diamonds.

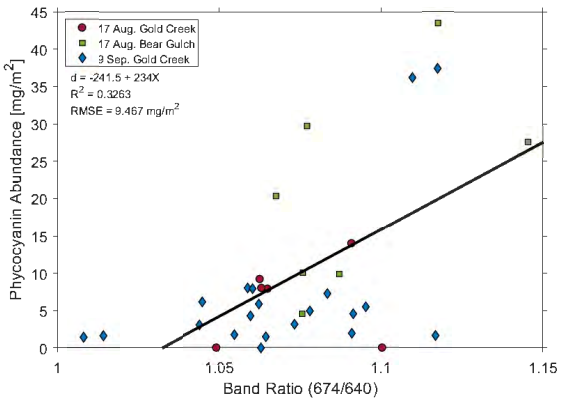
Figure A.2: Analytics from the regression analysis of fila/epip chl *a* standing crops from the combined (all field sites) data set.



(a) Heat map of R^2 values generated by fitting each band combination against epiphytic phycocyanin standing crops. Colors within the heat map represent the R^2 value, with wavelength numerators and denominators shown on the x- and y-axes, respectively.

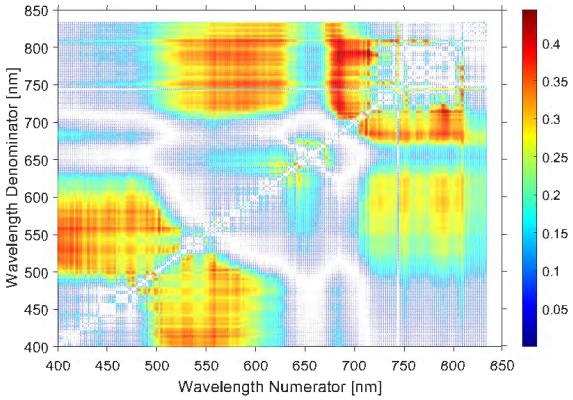


(b) Histogram of the R^2 values generated using the brute-force analysis.

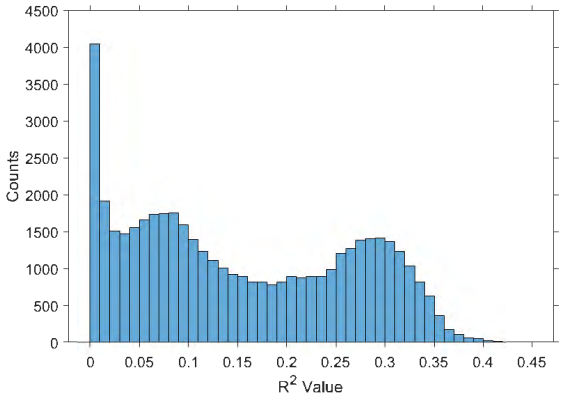


(c) Linear regression analysis between phycocyanin abundance (y-axis) and optimal band ratio (x-axis). Data collected on 17 August 2021 at the Gold Creek and Bear Gulch site represented by red circles and green squares, respectively, while data collected on 9 September 2021 are shown with blue diamonds.

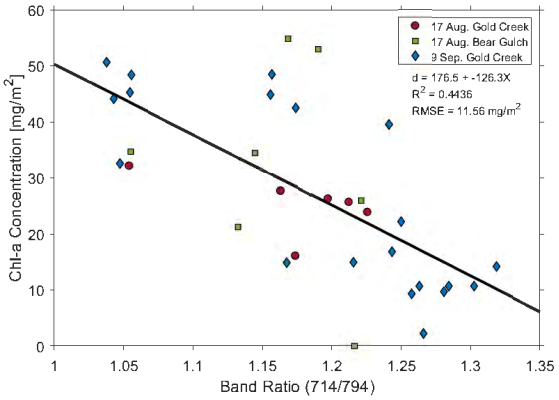
Figure A.3: Analytics from the regression analysis of epiphytic phycocyanin standing crops from the combined (all field sites) data set.



(a) Heat map of R^2 values generated by fitting each band combination against epilithic chl *a* standing crops. Colors within the heat map represent the R^2 value, with wavelength numerators and denominators shown on the x- and y-axes, respectively.

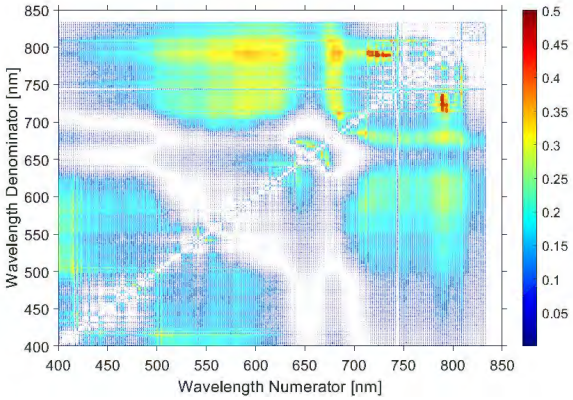


(b) Histogram of the R^2 values generated using the brute-force analysis.

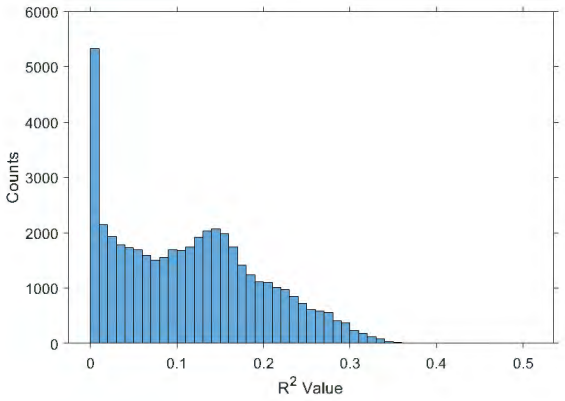


(c) Linear regression analysis between epilithic chl *a* abundance (y-axis) and optimal band ratio (x-axis). Data collected on 17 August 2021 at the Gold Creek and Bear Gulch site represented by red circles and green squares, respectively, while data collected on 9 September 2021 are shown with blue diamonds.

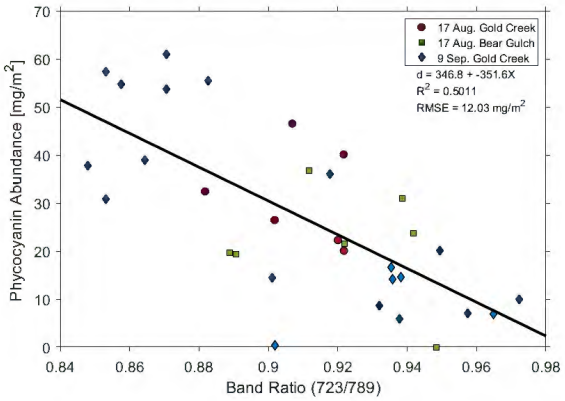
Figure A.4: Analytics from the regression analysis of epilithic chl *a* standing crops from the combined (all field sites) data set.



(a) Heat map of R^2 values generated by fitting each band combination against epilithic phycocyanin standing crops. Colors within the heat map represent the R^2 value, with wavelength numerators and denominators shown on the x- and y-axes, respectively.

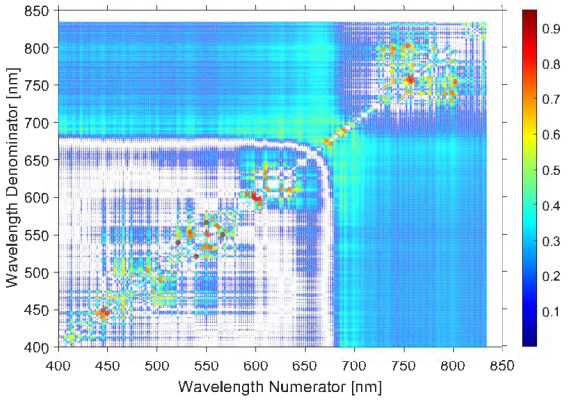


(b) Histogram of the R^2 values generated using the brute-force analysis.

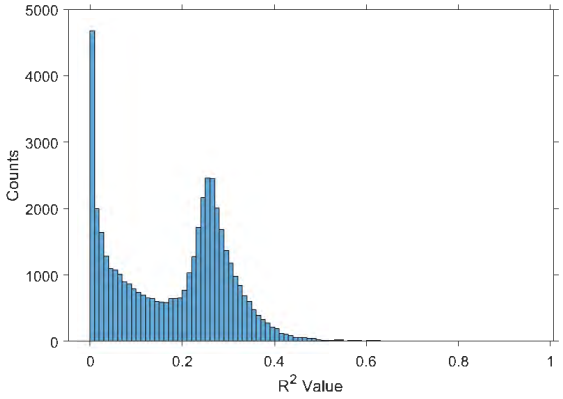


(c) Linear regression analysis between epilithic phycocyanin abundance (y-axis) and optimal band ratio (x-axis). Data collected on 17 August 2021 at the Gold Creek and Bear Gulch site represented by red circles and green squares, respectively, while data collected on 9 September 2021 are shown with blue diamonds.

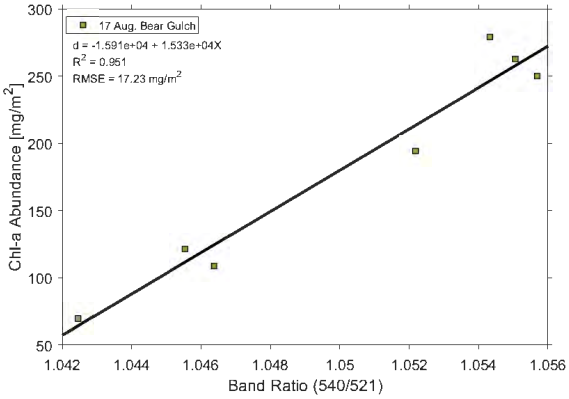
Figure A.5: Analytics from the regression analysis of epilithic phycocyanin standing crops from the combined (all field sites) data set.



(a) Heat map of R^2 values generated by fitting each band combination against total chl *a* standing crops. Colors within the heat map represent the R^2 value, with wavelength numerators and denominators shown on the x- and y-axes, respectively.

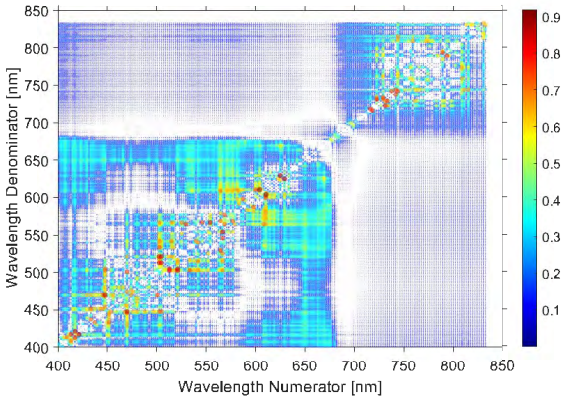


(b) Histogram of the R^2 values generated using the brute-force analysis.

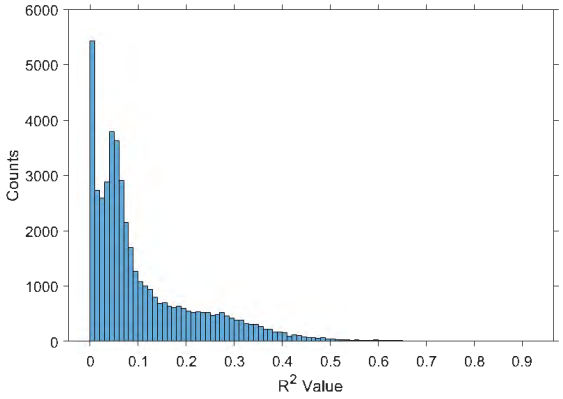


(c) Linear regression analysis between total chl *a* abundance (y-axis) and optimal band ratio (x-axis). Data collected on 17 August 2021 at the Bear Gulch site represented by green squares.

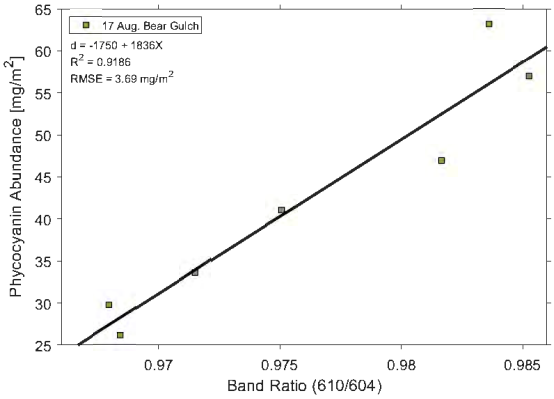
Figure A.6: Analytics from the regression analysis of total chl *a* standing crops from the Bear Gulch data set.



(a) Heat map of R^2 values generated by fitting each band combination against total phycocyanin standing crops. Colors within the heat map represent the R^2 value, with wavelength numerators and denominators shown on the x- and y-axes, respectively.

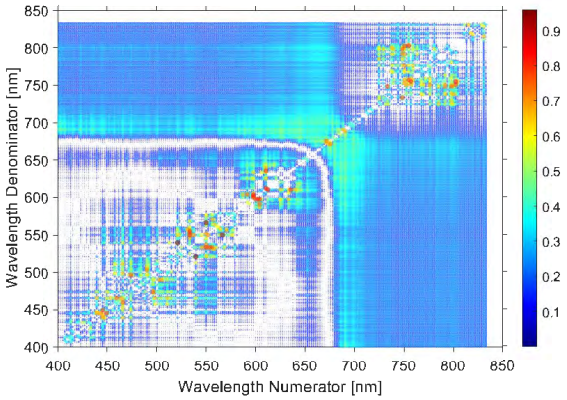


(b) Histogram of the R^2 values generated using the brute-force analysis.

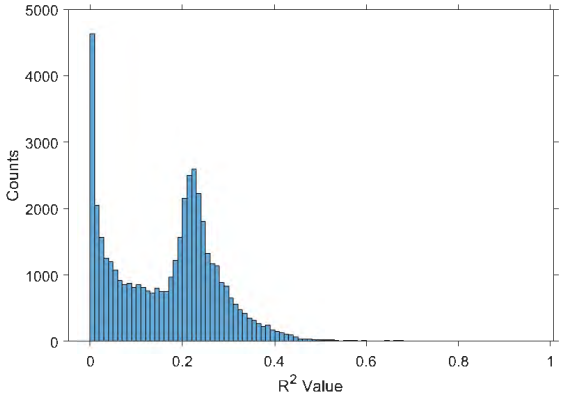


(c) Linear regression analysis between total phycocyanin abundance (y-axis) and optimal band ratio (x-axis). Data collected on 17 August 2021 at the Bear Gulch site represented by green squares.

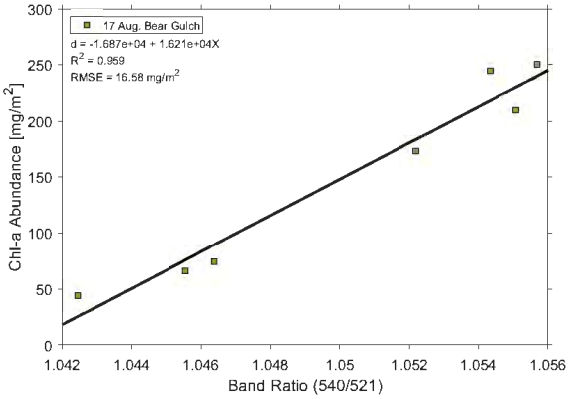
Figure A.7: Analytics from the regression analysis of total phycocyanin standing crops from the Bear Gulch data set.



(a) Heat map of R^2 values generated by fitting each band combination against fila/epip chl *a* standing crops. Colors within the heat map represent the R^2 value, with wavelength numerators and denominators shown on the x- and y-axes, respectively.

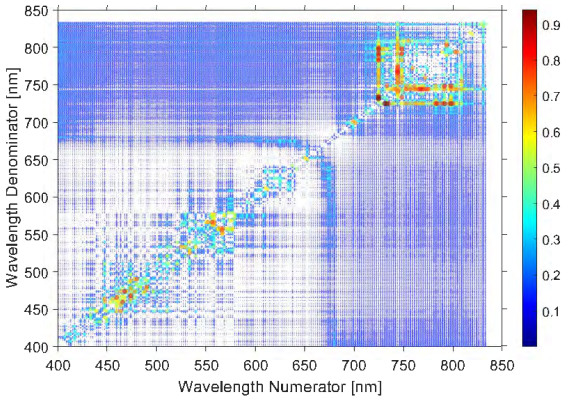


(b) Histogram of the R^2 values generated using the brute-force analysis.

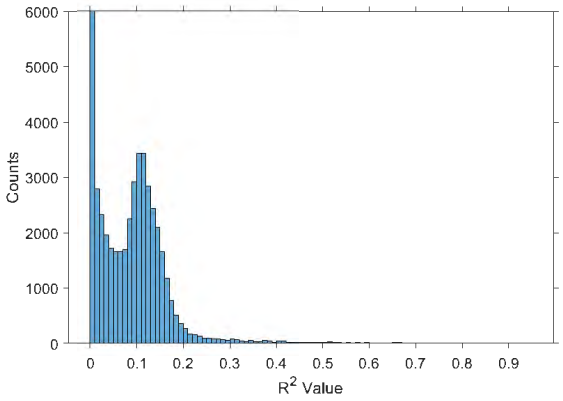


(c) Linear regression analysis between fila/epip chl *a* abundance (y-axis) and optimal band ratio (x-axis). Data collected on 17 August 2021 at the Bear Gulch site represented by green squares.

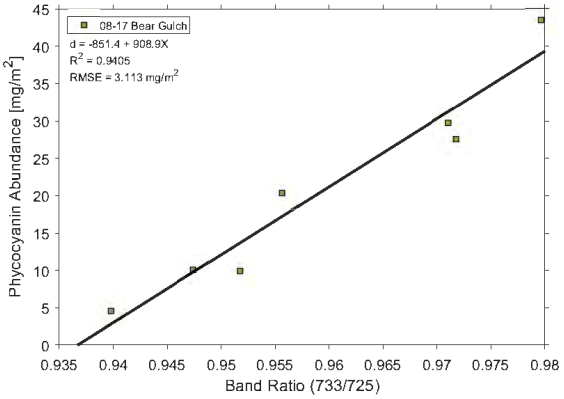
Figure A.8: Analytics from the regression analysis of fila/epip chl *a* standing crops from the Bear Gulch data set.



(a) Heat map of R^2 values generated by fitting each band combination against epiphytic phycocyanin standing crops. Colors within the heat map represent the R^2 value, with wavelength numerators and denominators shown on the x- and y-axes, respectively.

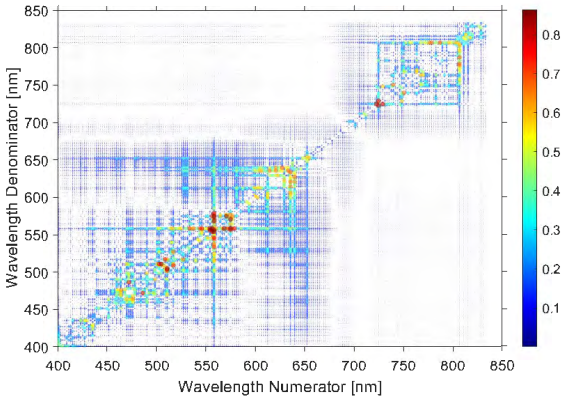


(b) Histogram of the R^2 values generated using the brute-force analysis.

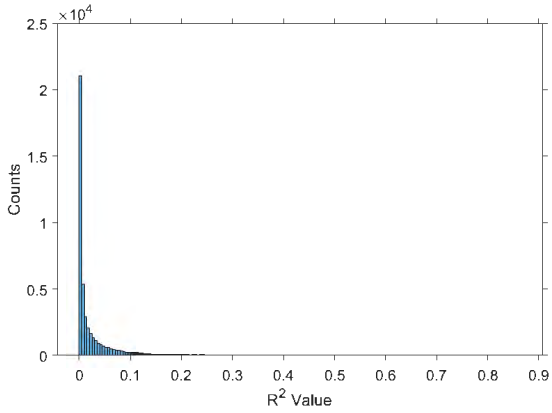


(c) Linear regression analysis between epiphytic phycocyanin abundance (y-axis) and optimal band ratio (x-axis). Data collected on 17 August 2021 at the Bear Gulch site represented by green squares.

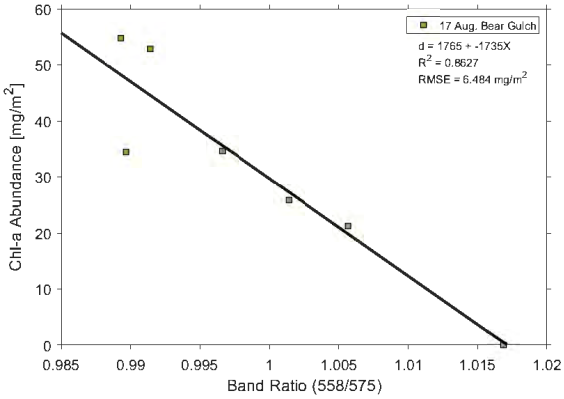
Figure A.9: Analytics from the regression analysis of epiphytic phycocyanin standing crops from the Bear Gulch data set.



(a) Heat map of R^2 values generated by fitting each band combination against epilithic chl *a* standing crops. Colors within the heat map represent the R^2 value, with wavelength numerators and denominators shown on the x- and y-axes, respectively.

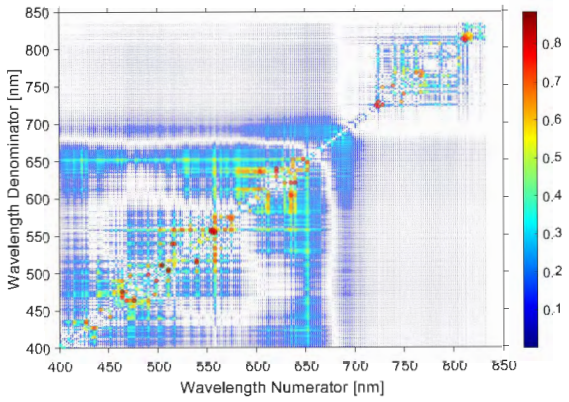


(b) Histogram of the R^2 values generated using the brute-force analysis.

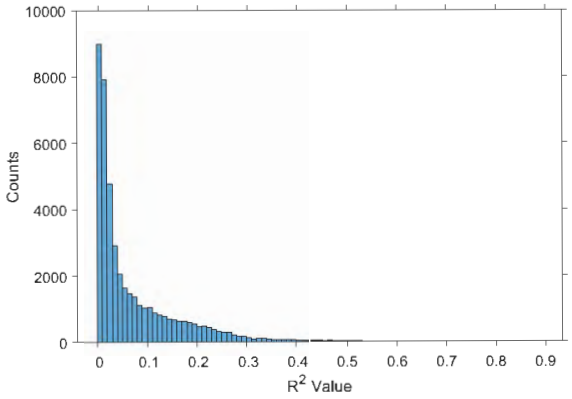


(c) Linear regression analysis between epilithic chl *a* abundance (y-axis) and optimal band ratio (x-axis). Data collected on 17 August 2021 at the Bear Gulch site represented by green squares.

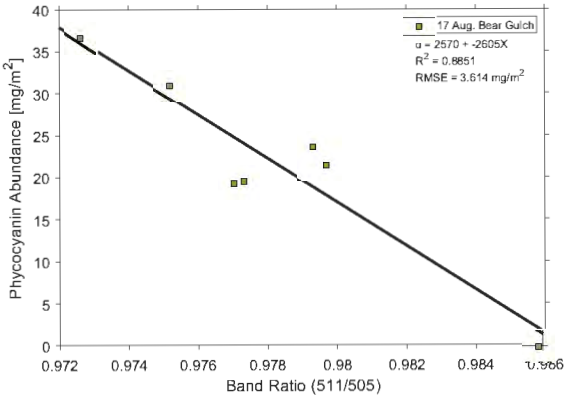
Figure A.10: Analytics from the regression analysis of epilithic chl *a* standing crops from the Bear Gulch data set.



(a) Heat map of R^2 values generated by fitting each band combination against epilithic phycocyanin standing crops. Colors within the heat map represent the R^2 value, with wavelength numerators and denominators shown on the x- and y-axes, respectively.

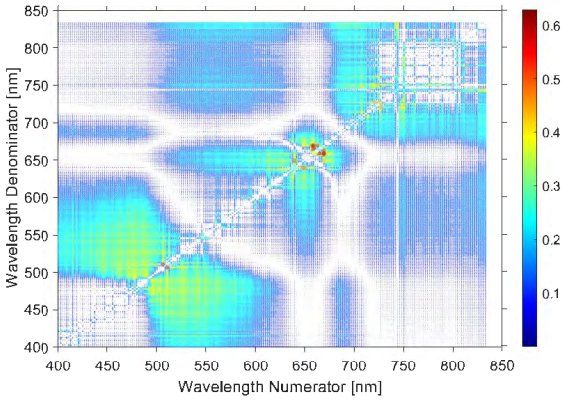


(b) Histogram of the R^2 values generated using the brute-force analysis.

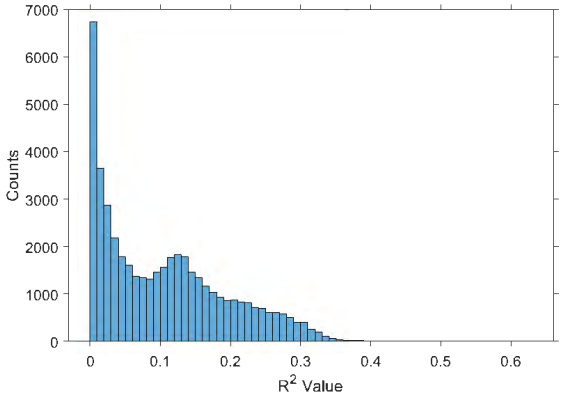


(c) Linear regression analysis between epilithic phycocyanin abundance (y-axis) and optimal band ratio (x-axis). Data collected on 17 August 2021 at the Bear Gulch site represented by green squares.

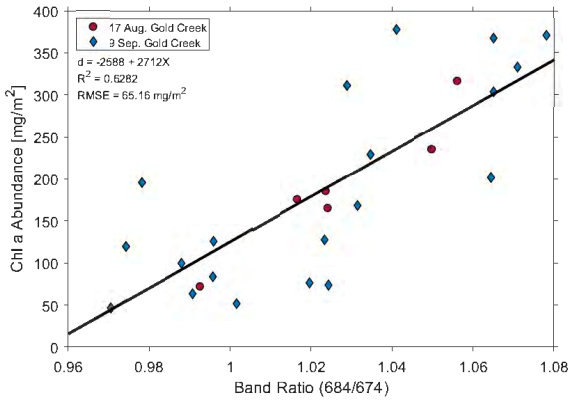
Figure A.11: Analytics from the regression analysis of epilithic phycocyanin standing crops from the Bear Gulch data set.



(a) Heat map of R^2 values generated by fitting each band combination against total chl *a* standing crops. Colors within the heat map represent the R^2 value, with wavelength numerators and denominators shown on the x- and y-axes, respectively.

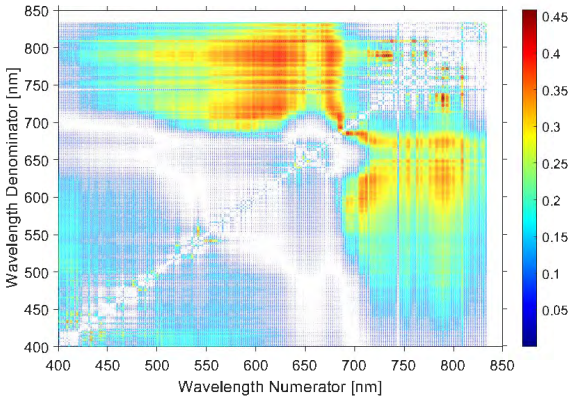


(b) Histogram of the R^2 values generated using the brute-force analysis.

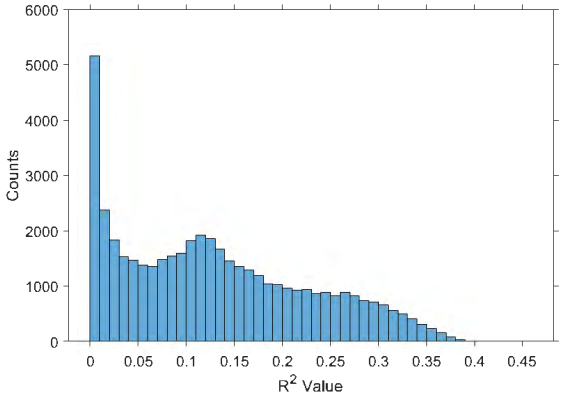


(c) Linear regression analysis between total chl *a* abundance (y-axis) and optimal band ratio (x-axis). Data collected on 17 August 2021 at the Gold Creek site represented by red circles, while data collected on 9 September 2021 are shown with blue diamonds.

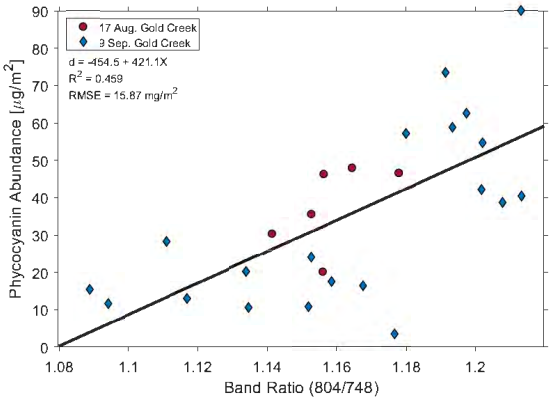
Figure A.12: Analytics from the regression analysis of total chl *a* standing crops from the Gold Creek data set.



(a) Heat map of R^2 values generated by fitting each band combination against total phycocyanin standing crops. Colors within the heat map represent the R^2 value, with wavelength numerators and denominators shown on the x- and y-axes, respectively.

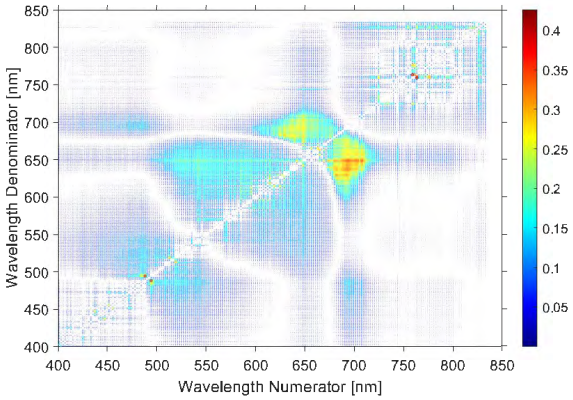


(b) Histogram of the R^2 values generated using the brute-force analysis.

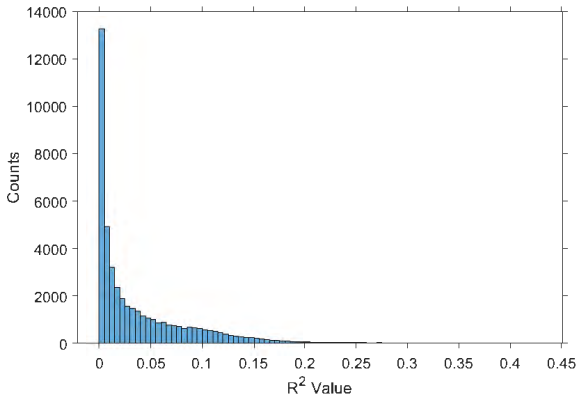


(c) Linear regression analysis between total phycocyanin abundance (y-axis) and optimal band ratio (x-axis). Data collected on 17 August 2021 at the Gold Creek site represented by red circles, while data collected on 9 September 2021 are shown with blue diamonds.

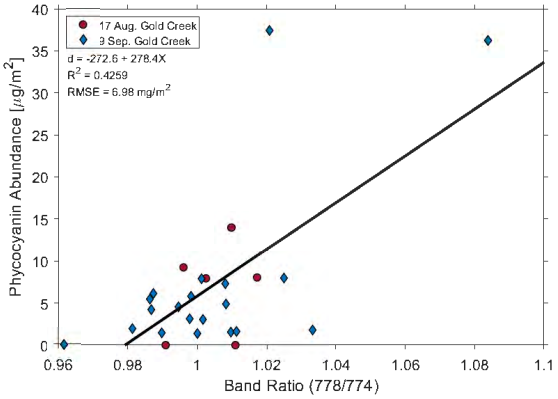
Figure A.13: Analytics from the regression analysis of total phycocyanin standing crops from the Gold Creek data set.



(a) Heat map of R^2 values generated by fitting each band combination against epiphytic phycocyanin standing crops. Colors within the heat map represent the R^2 value, with wavelength numerators and denominators shown on the x- and y-axes, respectively.

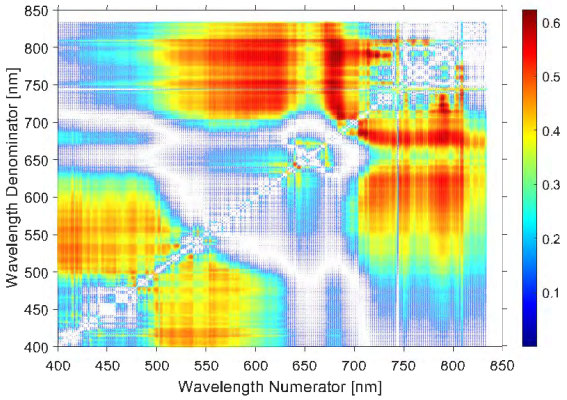


(b) Histogram of the R^2 values generated using the brute-force analysis.

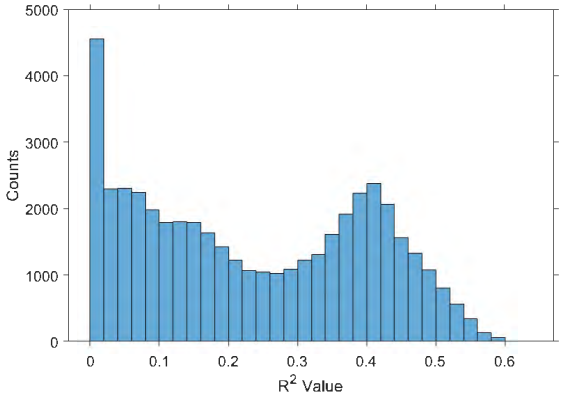


(c) Linear regression analysis between epiphytic phycocyanin abundance (y-axis) and optimal band ratio (x-axis). Data collected on 17 August 2021 at the Gold Creek site represented by red circles, while data collected on 9 September 2021 are shown with blue diamonds.

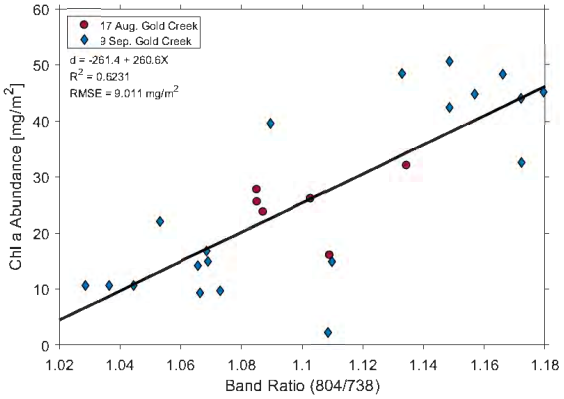
Figure A.14: Analytics from the regression analysis of epiphytic phycocyanin standing crops from the Gold Creek data set.



(a) Heat map of R^2 values generated by fitting each band combination against epilithic chl *a* standing crops. Colors within the heat map represent the R^2 value, with wavelength numerators and denominators shown on the x- and y-axes, respectively.



(b) Histogram of the R^2 values generated using the brute-force analysis.



(c) Linear regression analysis between epilithic chl *a* abundance (y-axis) and optimal band ratio (x-axis). Data collected on 17 August 2021 at the Gold Creek site represented by red circles, while data collected on 9 September 2021 are shown with blue diamonds.

Figure A.15: Analytics from the regression analysis of epilithic chl *a* standing crops from the Gold Creek data set.