Water Quality and Algal Bloom Sensing from Multiple Imaging Platforms

Ryan Ford
rmf3259@rit.edu

Follow this and additional works at: https://scholarworks.rit.edu/theses

Recommended Citation

This Dissertation is brought to you for free and open access by RIT Scholar Works. It has been accepted for inclusion in Theses by an authorized administrator of RIT Scholar Works. For more information, please contact ritscholarworks@rit.edu.
Water Quality and Algal Bloom Sensing from Multiple Imaging Platforms

by

Ryan Ford

B.A. Physics, State University of New York at Geneseo, 2013

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Chester F. Carlson Center for Imaging Science College of Science Rochester Institute of Technology

August 8, 2019

Signature of the Author

Accepted by

Dr. Charles Bachmann, Coordinator, Ph.D. Degree Program

Date
The Ph.D. Degree Dissertation of Ryan Ford has been examined and approved by the dissertation committee as satisfactory for the dissertation required for the Ph.D. degree in Imaging Science.

Dr. Anthony Vodacek, Dissertation Advisor

Dr. Christy Tyler, External Chair

Dr. John Kerekes

Dr. Charles Bachmann
Water Quality and Algal Bloom Sensing from Multiple Imaging Platforms

Abstract

Harmful cyanobacteria blooms have been increasing in frequency throughout the world resulting in a greater need for water quality monitoring. Traditional methods of monitoring water quality, such as point sampling, are often resource expensive and time consuming in comparison to remote sensing approaches, however the spatial resolution of established water remote sensing satellites is often too coarse (300 m) to resolve smaller inland waterbodies. The fine scale spatial resolution and improved radiometric sensitivity of Landsat satellites (30 m) can resolve these smaller waterbodies, enabling their capability for cyanobacteria bloom monitoring.

In this work, the utility of Landsat to retrieve concentrations of two cyanobacteria bloom pigments, chlorophyll-a and phycocyanin, is assessed. Concentrations of these pigments are retrieved using a spectral Look-Up-Table (LUT) matching process, where an exploration of the effects of LUT design on retrieval accuracy is performed. Potential augmentations to the spectral sampling of Landsat are also tested to determine how it can be improved for waterbody constituent concentration retrieval.

Applying the LUT matching process to Landsat 8 imagery determined that concentrations of chlorophyll-a, total suspended solids, and color dissolved organic matter were retrieved with a satisfactory accuracy through appropriate choice of atmospheric compensation and LUT design, in agreement with previously reported implementations of the LUT matching process. Phycocyanin proved to be a greater challenge to this process due to its weak effect on waterbody spectrum, the lack of Landsat spectral sampling over its predominant spectral feature, and error from atmospheric compensation. From testing potential enhancements to Landsat spectral sampling, we determine that additional spectral sampling in the yellow and red edge regions of the visible/near-infrared (VNIR) spectrum can lead to improved concentration retrievals. This performance further improves when sampling is added to both regions, and when Landsat is transitioned to a VNIR imaging spectrometer, though this is dependent on band position and spacing. These results imply
that Landsat can be used to monitor cyanobacteria blooms through retrieval of chlorophyll-a, and this retrieval performance can be improved in future Landsat systems, even with minor changes to spectral sampling. This includes improvement in retrieval of phycocyanin when implementing a VNIR imaging spectrometer.
Acknowledgements

I would like to thank my advisor, Dr. Anthony Vodacek, for all the support, guidance, and opportunities he has provided me throughout my graduate school career. Similarly I would like to thank my committee for all of the ways in which they shaped and assisted my research. In addition, I am grateful to Dr. Aaron Gerace and Dr. John Schott for starting me on this project. I also greatly appreciate the rest of CIS faculty and staff for all the help, knowledge, and advice they have given me through this process.

I would also like to thank all the people who helped me perform my field collections. This includes the late Bob Brower and the Owasco Watershed Lake Association as well as Terry Gronwall of the Honeoye Lake Watershed Task Force. Most prominently I would like to thank Nina Raqueno, who chauffeured me around while I napped, rode in the boat with me despite the feeling of tipping over at any moment, and performed all other manner of assistance during collections.

In consideration of the little bit of life outside of work, I am so grateful to all the amazing friends that I have gained by virtue of being colleagues. I hope the best for them in their continued studies and subsequent lives. Last, and probably mostly, I am grateful for my family and all their support over these years. Their assistance in taking care of the minutia of life has allowed me to focus on and complete this work.
Dedicated to my family, who was willing to give their full support through this process, whether or not they understood my work.
## Contents

1 Introduction 1

2 Objectives 4
   2.1 Problem Statement ........................................... 4
   2.2 Project Objectives ........................................... 5
   2.3 Contributions to Field ........................................ 6

3 Background and Theory 7
   3.1 Concepts of Radiometry ....................................... 7
      3.1.1 Radiometric Quantities .................................. 8
      3.1.2 Exoatmospheric Radiance and Atmospheric Effects ....... 10
      3.1.3 Solar Energy Paths ...................................... 13
      3.1.4 Governing Equation ..................................... 15
   3.2 Imaging Systems and Platforms ............................... 17
      3.2.1 Imaging Systems ......................................... 17
      3.2.2 Sensor Performance Specifications ....................... 18
      3.2.3 Platforms ................................................ 20
   3.3 Radiative Transfer in Water .................................. 24
      3.3.1 Inherent Optical Properties ............................. 24
      3.3.2 Water Constituents ...................................... 27
      3.3.3 Energy Paths in Water .................................. 37
      3.3.4 Glint ..................................................... 38
3.3.5 Hydrolight & Ecolight ........................................ 40
3.4 Past Work in Remote Sensing of Cyanobacteria Blooms ................. 41
  3.4.1 Empirical Methods ........................................ 41
  3.4.2 Semi-Empirical Methods .................................... 44
3.5 Past Implementations of Hydrolight LUTs ................................ 46

4 Methods ................................................................. 49
  4.1 Imagery Acquisition and Processing .................................. 49
    4.1.1 UAS Imagery ............................................... 50
    4.1.2 Satellite Imagery .......................................... 53
  4.2 Field Collections .................................................. 54
    4.2.1 Water Sample Collection .................................. 55
    4.2.2 Water Spectra Collection .................................. 55
  4.3 Laboratory Analysis ................................................ 56
    4.3.1 Absorption Measurements ................................... 56
    4.3.2 Concentration Measurements ................................ 58
  4.4 Collection Images and Data Sets .................................... 61
    4.4.1 Landsat 8 .................................................. 61
    4.4.2 Sentinel-2 .................................................. 61
    4.4.3 UAS ......................................................... 63
  4.5 Atmospheric Compensation ........................................... 65
    4.5.1 Empirical Line Method ...................................... 65
    4.5.2 Model-Based Empirical Line Method ......................... 67
    4.5.3 SeaDAS MUMM .............................................. 67
    4.5.4 Acolite .................................................... 68
    4.5.5 Sen2Cor .................................................... 69
  4.6 Model Based Spectral Look-Up-Table Matching Process .................. 70
    4.6.1 LUT Generation ............................................. 70
    4.6.2 LUT Spectra Matching ....................................... 71
    4.6.3 Design and Implementation of IOP Variability in LUTs ........... 74
    4.6.4 Potential Sources of Error in the LUT Retrieval Method ........ 81
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Geometric visualization of projected irradiance</td>
<td>9</td>
</tr>
<tr>
<td>3.2</td>
<td>Exoatmospheric spectral irradiance from Mobley (1994)</td>
<td>11</td>
</tr>
<tr>
<td>3.3</td>
<td>MODTRAN simulated atmospheric transmission and upwelled radiance</td>
<td>12</td>
</tr>
<tr>
<td>3.4</td>
<td>Scattering phase function profiles of both Mie and Rayleigh scattering</td>
<td>14</td>
</tr>
<tr>
<td>3.5</td>
<td>Optical paths of visible light interacting with the target and surrounding region</td>
<td>15</td>
</tr>
<tr>
<td>3.6</td>
<td>Relative Spectral Response of Landsat 8</td>
<td>21</td>
</tr>
<tr>
<td>3.7</td>
<td>Relative Spectral Response of the MSI bands aboard Sentinel-2</td>
<td>22</td>
</tr>
<tr>
<td>3.8</td>
<td>Spectral filter transmissions and sensor response of the Tetracam imaging system</td>
<td>23</td>
</tr>
<tr>
<td>3.9</td>
<td>Scattering phase function from Petzold (1972) compared to those of Fournier and Forand (1994)</td>
<td>26</td>
</tr>
<tr>
<td>3.10</td>
<td>Absorption and scattering of pure water</td>
<td>28</td>
</tr>
<tr>
<td>3.11</td>
<td>Absorption and scattering of Chl from measurements in the Finger Lakes region of New York and the literature</td>
<td>29</td>
</tr>
<tr>
<td>3.12</td>
<td>Relation between Chl mass specific absorption and concentration</td>
<td>31</td>
</tr>
<tr>
<td>3.13</td>
<td>Normalized absorption of CDOM from measurements in the Finger Lakes region of New York</td>
<td>32</td>
</tr>
<tr>
<td>3.14</td>
<td>Absorption and scattering of Total Suspended Solids from measurements in the Finger Lakes region of New York</td>
<td>33</td>
</tr>
<tr>
<td>3.15</td>
<td>Relation between TSS absorption and concentration from Tilstone et al. (2012)</td>
<td>34</td>
</tr>
<tr>
<td>3.16</td>
<td>Absorption and estimated scattering of phycocyanin</td>
<td>35</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3.17</td>
<td>Reported covariations of Chl, TSS, and CDOM concentration from Babin et al. (2003)</td>
<td>36</td>
</tr>
<tr>
<td>3.18</td>
<td>Various optical paths through waterbodies</td>
<td>39</td>
</tr>
<tr>
<td>4.1</td>
<td>Landsat 8 imagery of Honeoye Lake with marked sampling points</td>
<td>62</td>
</tr>
<tr>
<td>4.2</td>
<td>Landsat 8 imagery of Owasco Lake with marked sampling points</td>
<td>63</td>
</tr>
<tr>
<td>4.3</td>
<td>Sentinel-2 imagery of Lake Erie with marked sampling points</td>
<td>64</td>
</tr>
<tr>
<td>4.4</td>
<td>Example linear relation determined from the Empirical Line Method of atmospheric compensation</td>
<td>66</td>
</tr>
<tr>
<td>4.5</td>
<td>Example LUT generated in Ecolight</td>
<td>71</td>
</tr>
<tr>
<td>4.6</td>
<td>MODTRAN output used in simulating imagery for testing potential improvements to future systems</td>
<td>85</td>
</tr>
<tr>
<td>4.7</td>
<td>Spectral response of the first 5 Landsat 8 bands with a modeled Super-Gaussian NIR band</td>
<td>87</td>
</tr>
<tr>
<td>4.8</td>
<td>Landsat band system and quantization noise for bands 1 through 5 from Morfitt et al. (2015)</td>
<td>89</td>
</tr>
<tr>
<td>5.1</td>
<td>Effect of concentration steps within the LUT on retrieval error</td>
<td>93</td>
</tr>
<tr>
<td>5.2</td>
<td>Effect of variability in scattering phase function with the scene/LUT on retrieval</td>
<td>95</td>
</tr>
<tr>
<td>5.3</td>
<td>Comparison of atmospheric compensated data with in-situ data from Owasco Lake on September 20, 2016</td>
<td>96</td>
</tr>
<tr>
<td>5.4</td>
<td>Error of various atmospheric compensation methods when compared to re-sampled in-situ spectra</td>
<td>97</td>
</tr>
<tr>
<td>5.5</td>
<td>Results of unstructured Mean IOP and structured Split IOP LUTs to Finger Lakes Imagery</td>
<td>98</td>
</tr>
<tr>
<td>5.6</td>
<td>Results of applying individualized structured LUTs to Owasco Lake imagery</td>
<td>100</td>
</tr>
<tr>
<td>5.7</td>
<td>Results of applying individualized structured LUTs to MoB-ELM and MUMM imagery with 5 or 3 band fitting</td>
<td>101</td>
</tr>
<tr>
<td>5.8</td>
<td>Effect of considering PC as a scattering vs non-scattering component on retrieval</td>
<td>104</td>
</tr>
</tbody>
</table>
5.9 Mean water spectra Rrs for varying component concentrations without PC scattering compared to atmospheric compensation MAE .......................... 105
5.10 Retrieved versus measured concentrations generated using global LUTs .................. 107
5.11 NRMSE comparison for all global LUTs .................................................. 108
5.12 Retrieved versus measured concentrations from Sentinel-2 Imagery generated using the Literature LUT .......................................................... 110
5.13 Results of LUT applied to a sub-sectioned Tetracam image of Long Pond ........ 112
5.14 Results of Literature LUT applied to a Landsat 8 image of Long Pond .............. 113
5.15 Results of applying the Literature LUT to UAS imagery over a bloom in Honeoye Lake ................................................................. 114
5.16 Comparison of Headwall Nano Spectra to LUT entries .............................. 115
5.17 Comparison of TOA Landsat 8 imagery over cyanobacteria bloom waters to modeled TOA imagery .............................................. 117
5.18 Retrieval NRMSE of the modeled Landsat baseline system generated using 1000 modeled test spectra ......................................................... 118
5.19 Retrieval NRMSE for potential future Landsat systems with an added yellow band .......................................................... 119
5.20 Retrieval NRMSE for potential future Landsat systems with an added red edge band .......................................................... 121
5.21 Retrieval NRMSE for potential future Landsat systems with both an added yellow and red edge band ......................................................... 123
5.22 Retrieval NRMSE for modeled potential imaging spectrometers varying in bandwidths .......................................................... 124
5.23 Retrieval error of imaging spectrometers with band centers shifted during testing ........................................................................ 125
5.24 Retrieval error of a 20 nm imaging spectrometer resampled to Landsat 8 RSRs 126
5.25 Histogram of wavelengths chosen for best UAS band results ....................... 127

6.1 Comparison of all tested potential future Landsat systems .......................... 133
Chapter 1

Introduction

Water constitutes 71% of the Earth’s surface. Earth’s waters provide a habitat for flora and fauna, with oceans containing over one million species of animals and plants alone. These waters and the photosynthetic life within them supply almost 50% to 80% of the Earth’s oxygen. Furthermore, water served as the origin point of all life on Earth. On a more pragmatic level, water acts as a recreational and economic resource for mankind. It supplies food, allows for transportation of people and goods, spurs tourism, and freshwater (a mere 2.5% of all of Earth’s waters, a fraction of which is surface water) supplies potable water to all people and communities. As water is such a necessary resource, it is imperative that its quality be monitored so it can be preserved for mankind and other species.

Accurately monitoring Earth’s waters is a challenge due to the variation in scale. Areas within oceans and seas may be resource taxing to monitor, due to their remoteness. Smaller ponds and lakes may be considered to be too small of a concern, meaning that monitoring resources may be hard to acquire. Furthermore, collections from boat or buoy occur at a single point in a waterbody, and field samples can be taxing on time and resources for analysis. Remote sensing is a powerful tool for overcoming these obstacles, as it is able to provide comprehensive measurements over large areas in a short time period. Imagery can be easily accessible for interested groups with limited resources as well, such as Landsat imagery which is freely available.

The use of remote sensing for water quality monitoring truly began with the launch of
the proof of concept Coastal Zone Color Scanner (CZCS) satellite in 1978. This led to the
development of future passive imaging systems, for example the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) in 1997, the Moderate-resolution Imaging Spectroradiometer (MODIS) in 1999, and various systems within the European Space Agency (ESA) Copernicus Program (e.g. Sentinel-3). Along with the implementation of the SeaDAS imagery processing software, these systems have lead to great success in the study of relatively clear ocean waters. These systems and their algorithms, however, still need further development for monitoring turbid coastal and inland waters.

The difficulty of using remote sensing in coastal and inland waters comes from the optical overlap of their constituents. Total suspended solids, color dissolved organic matter, phytoplankton, and water itself all play a role in shaping the water spectra. Inland waterbodies are also variable in size, meaning they may require a finer spatial resolution than is necessary for ocean waters. The importance of monitoring these inland waters is increased by the fact that they are sometimes used to supply potable water to surrounding communities. One possible method of overcoming the optical complexity of these waters is to simultaneous retrieve information about all of its constituents. This can be done through the use of spectral modeled Look-Up-Tables (LUTs). The viability of this technique has already been proven from satellite and plane based sensing systems by Raqueno (2003), Gerace (2010), and Concha (2015).

Cyanobacteria blooms in inland waterbodies are an area of growing concern, given their potential to degrade water quality and release toxins into otherwise potable water. Past attempts to use remote sensing to monitor these types of Harmful Algal Blooms (HABs) has mostly lead to the creation of sensor specific algorithms calibrated for the optical properties of the waters they were developed around. Furthermore, sensors with the appropriate spectral sampling to implement these algorithms often have coarse spatial resolutions, which fail to resolve smaller inland waterbodies.

The issues presented here motivate the objectives of this work. The first objective is to determine if the spectral LUT matching process, a sensor non-specific process, can be extended to retrieve concentrations to pigments related to cyanobacteria blooms and to explore how LUT design affects this retrieval. Specific focus is given to imagery from Landsat 8 as its 30 m spatial resolution enables capture of smaller inland waters. The
second objective is to assess the potential of using the LUT process on imagery from Unmanned Aerial Systems (UAS). The final objective is to identify how future imaging systems could be improved to perform retrieval.
Chapter 2

Objectives

2.1 Problem Statement

Cyanobacteria blooms, just one type of harmful algal blooms, present danger to the ecology of waterbodies as well as the communities that depend on them through their effect on water quality and their potential to release toxins. These blooms require monitoring and observation to inform communities of potential risks and to better understand their source and dynamics. Remote sensing is a powerful potential tool for monitoring blooms through its ability to perform comprehensive observations of waterbodies and the recent implementation of UAS in remote sensing has made collections at fine spatial resolutions viable. The spectral LUT matching process is a proven process for monitoring water quality from remotely sensed imagery, and has potential to be extended for monitoring cyanobacteria blooms. This study proposes to answer three questions based on this premise: Can the spectral LUT matching process be used to extract concentrations of cyanobacteria pigments from remote sensing data? Can these methods be implemented on imagery with fine spatial resolution? Finally in what ways can imaging systems and the retrieval process be improved to perform this retrieval?
2.2 Project Objectives

Based on the questions raised in the problem statement, this project is constructed with 3 main objectives.

1) **Determine the accuracy of using LUTs to retrieve concentrations of pigments associated with cyanobacteria blooms**

   This objective requires generating spectral LUTs that include pigments associated with cyanobacteria blooms. These LUTs will then be applied to satellite imagery (e.g. Landsat 8, Sentinel-2) of areas known to be undergoing cyanobacteria blooms. The imagery will have georeferenced data collected as close to the overpass as possible. In addition to testing on imagery, the effects of elements of the retrieval process (e.g. LUT design) will be analyzed to determine potential areas of improvement.

2) **Assess the potential of using UAS imagery in the spectral LUT matching process**

   This objective is motivated by the fine spatial resolution of UAS imagery and how it may affect the retrieval process. Similar to the first objective, this will require imagery collections along with collections of reference data. This objective will also require determination of appropriate calibration methods and processing of imagery to minimize issues such as glint.

3) **Determine system specifications that increase pigment retrieval accuracy**

   The final objective of this work is to determine system specifications improve performance of the retrieval task. A specific focus is placed on potential spectral sampling of future Landsat systems. Modeled water spectra will be propagated through the atmosphere to a simulated system with noise, quantization, and its own spectral sampling. These spectra will be compensated for atmospheric effects and used in the spectral LUT matching process to extract waterbody component concentrations. Suggestions for future systems will be determined based on comparison of retrieval accuracy. This objective will focus on both examining future Landsat systems as multispectral and visible/near-infrared (VNIR) imaging spectrometers. These same
methods will be used to determine a unique set of bands for UAS systems that optimize performance of this retrieval task.

2.3 Contributions to Field

The novel contributions of this work pertain most directly to remote sensing of inland and coastal waters. This work will be the first to extend the LUT matching process to more than three waterbody components through also attempting retrieval of the accessory pigment phycocyanin. Previous work using this method has only focused on chlorophyll, total suspended solids, and color dissolved organic matter. Furthermore, as phycocyanin weakly affects the water signal, this work also tests the ability of the LUT matching process to retrieve concentrations of components with weak optical features. This is also the first time the retrieval process has been implemented using both Sentinel-2 satellites and UAS imagery.

This work is also the first to study how design of the LUT affects retrieval. It examines how variability of Inherent Optical Properties (IOPs) (e.g. absorption and scattering), both in the scene and in the LUT, affect the accuracy of the retrieval. This work is the first to examine the effects of varying the population of component concentrations within the LUT, specifically looking at the distance between adjacent concentrations.

Finally, the modeling work used to test system designs improves on the modeling done by Gerace (2010) through inclusion of variability within the water inherent optical properties of the modeled scene. Furthermore, this model is not used to determine performance of an upcoming system but to discover how spectral coverage of future systems can be improved to optimize performance of the LUT retrieval process.
Chapter 3

Background and Theory

This chapter describes the scientific theory and basis which underpins remote sensing, specifically of water quality. First, in Section 3.1, the basic concepts of radiometry and how light interacts with the Earth and its atmosphere is covered. Section 3.2 describes the design of imaging systems and platforms on which they are installed, with a focus on those that were used in this work. Following that, Section 3.3 covers radiative transfer in water, including a discussion of the various optical components of a waterbody and the Hydrolight radiative transfer model. Finally in Sections 3.4 and 3.5, previous work in remote sensing of algal blooms and use of spectral LUTs generated using Hydrolight is discussed.

3.1 Concepts of Radiometry

This section will discuss basic concepts of radiometry, and by extension remote sensing. Radiometry is defined as the science of measuring energy from electromagnetic radiation at a location in space in a certain direction (Schott, 2007). First the definition of various radiometric quantities will be covered to remove any ambiguity in terminology. Following this is a discussion of the Sun’s radiative energy and how it interacts with the Earth. This will include discussing how the atmosphere affects light, and the paths light can take when it reaches the Earth. Finally this information will be summarized in mathematical notation in the Governing Equation.
3.1.1 Radiometric Quantities

The most basic unit of light is the photon. The photon is considered to be analogous to a single particle of light, however due to the particle-wave duality of light we can describe its energy as:

$$q = \frac{hc}{\lambda} = h\nu \ [J] \quad (3.1)$$

Where $h$ is Planck’s Constant ($6.626 \times 10^{-34} J \cdot s$), $c$ is the speed of light ($3 \times 10^8 m/s$), $\lambda$ is the wavelength of the particle, and $\nu$ is the frequency. The equality on the right hand side holds due to the constant speed of light in a vacuum. A light source generally creates numerous photons, therefore it is useful to describe the source by the rate at which energy is delivered or its flux. Flux is defined as (Schott, 2007):

$$\Phi = \frac{dq}{dt} \left[ \frac{J}{s} \text{ or } W \right] \quad (3.2)$$

Here $q$ still represents the photon energy, and $t$ is the amount of time that the incoming energy is observed. As indicated by equation 3.1, the wavelength (or frequency) of the photon dictates its energy, thus the amount of flux can also vary depending on the wavelength of incoming light as well. As such, the spectral nature of flux can be defined with units of $W/nm$. This spectral quality is applicable to all other equations described in this section, however has been omitted for brevity.

We can further describe the nature of a light field by considering its spatial and angular distributions. A useful quantity for determining the rate photons reach a certain area in space is irradiance, $E$ (Schott, 2007):

$$E = \frac{d\Phi}{dA} \left[ \frac{W}{m^2} \right] \quad (3.3)$$

Where $A$ is the area of space over which light is being measured. Irradiance may vary depending on the angle between its surface and the incident light, as seen in Figure 3.1.
In this case the irradiance can be accounted for by determining the projected irradiance:

\[ E_\theta = E \cos \theta \left( \frac{W}{m^2} \right) \]  
(3.4)

Here theta is the angle between the incident light and the normal of the surface \( dA \). The cosine nature of this projection arises out of simple geometric arguments.

![Figure 3.1: Geometric visualization of projected irradiance](image)

A need for a new term which is capable of describing both the spatial and angular distribution of light arises from irradiance only describing the distribution spatially. This is accomplished through the measurement of radiance, \( L \) (Schott, 2007):

\[ L = \frac{d^2\Phi}{dA \cos \theta \, d\Omega} \left( \frac{W}{m^2 \, sr} \right) \]  
(3.5)

The only term in this section not yet defined is \( \Omega \), which is the solid angle. The solid angle is considered to be the conic angle which carves out an area of a spherical surface \( (d\Omega = dA/r^2) \). The units of solid angles are steradians. It should also be noted that the cosine arises again from the projection of the area element.

As radiance is a term that describe light flux per direction and spatial area, and irradiance is a term that describes light flux per spatial area, it infers that there is another term which only describes the light flux per direction only. This term is called intensity, \( I \)
Chapter 3. Background and Theory

(Schott, 2007):

\[ I = \frac{d \Phi}{d \Omega} \left[ \frac{W}{sr} \right] \]  \hspace{1cm} (3.6)

This term is included for completeness, but is not necessary for understanding this work.

The final two terms to be discussed in this section are built from what has already been defined. The first is irradiance reflectance, which is the ratio of upwelled irradiance to downwelled irradiance off the surface of a sample. It is defined as (Mobley, 1994):

\[ r = \frac{E_u}{E_d} \]  \hspace{1cm} (3.7)

Where \( E_u \) and \( E_d \) are the upwelled and downwelled irradiance respectively. In this case upwelled can be thought of irradiance passing through the lower hemisphere of a sphere to the upper, and downwelled to be the reverse case. The upwelled irradiance is the irradiance leaving the illuminated surface, which is referred to as exittance, but retains the same units as irradiance. Related to this term, but more useful in working with water is the Remote Sensing Reflectance (Mobley, 1994):

\[ R_{rs} = \frac{L_w}{E_d} \left[ \frac{1}{sr} \right] \]  \hspace{1cm} (3.8)

\( E_d \) similarly represents the downwelled irradiance, where as \( L_w \) represents the upwelled water leaving radiance. This term will be used frequently within this work.

3.1.2 Exoatmospheric Radiance and Atmospheric Effects

The first step in the imaging chain is the light source (Fiete, 2010) and the light source most often used for passive imaging is the Sun. The Sun can be closely approximated to be a blackbody of effective temperature 5778 K (Lide, 2004). Using Wein’s displacement law it can be determined that the peak wavelength is in the visible spectrum, around 500 nm. Furthermore the Stefan-Boltzmann law can be used in conjunction with the mean Sun-Earth distance to determine the total irradiance incident on the Earth, \( 1367 \frac{W}{m^2} \) (Mobley, 1994). This irradiance is know as the exoatmospheric irradiance, as it is incident on Earth’s
atmosphere. Exoatmospheric irradiance has been well characterized, and an example of it can be seen in Figure 3.2.

Figure 3.2: Exoatmospheric spectral irradiance at the mean Earth-Sun distance
(Source: Mobley (1994))

This solar irradiance must then pass through the atmosphere in order to reach and interact with the Earth. The atmospheric plays a large role in affecting the amount of light that reaches the ground, and these atmospheric effects are not consistent spatially, temporally, or spectrally. Understanding the how the atmosphere affects solar irradiance is especially important when trying to determine radiometric quantities in low signal situations, such as over water. As this work primarily deals with observing water in the VNIR, the follow discussion focuses on atmospheric effects within this regime.

The atmosphere affects solar irradiance through two processes fundamental to radiative transfer, absorption and scattering. The first process, absorption, is a process by which photon energy is removed from incident light (Hapke, 2012). Light passing through the
atmosphere interacts with its various constituents (e.g. molecules such as water vapor, ozone, carbon monoxide and dioxide, and various aerosols) and is then absorbed. The amount and wavelength of light absorbed depends on the constituent and its concentration, which varies spatially and temporally. The effect of this absorption can be seen in a spectral irradiance profile as a series of valleys, known as absorption lines, indicating absorption of photons at that wavelength (and therefore energy). An example of absorption lines can be seen in Figure 3.3 which shows a plot of atmospheric transmission generated using the MODerate resolution atmospheric TRANsmission software (MODTRAN) (Berk et al., 1989). The Sun also exhibits this behavior as the gases surrounding it absorb some of its energy. The absorption lines that this causes are called Fraunhofer Lines (Schott, 2007).

Figure 3.3: A plot of spectral transmission of a mid-latitude summer atmosphere with rural aerosols made using MODTRAN. The dips seen in transmission are caused by atmospheric absorption.
The second process, scattering, is much more complicated than the former. Scattering is when a photon is deflected from its original path of travel due to interactions of electromagnetic fields. Scattering caused by the atmosphere generally takes two forms, Rayleigh scattering and Mie scattering (Schott, 2007).

Rayleigh scattering occurs when a photon interacts with a particle that is much smaller than its wavelength. In the case of the atmosphere, this means interactions with various molecules. This type of scattering is best defined by the fact that the amount of energy scattered is inversely proportional to the wavelength of the photon to the 4th power \( \left( \frac{1}{\lambda^4} \right) \) (Hapke, 2012). This means that higher energy photons, such as blue, are more scattered by the atmosphere as opposed to lower ones, such as red.

Mie scattering can describe the interaction between photons and particles that are on the same order as their wavelength. It should be noted that Mie scattering is only a model for scattering caused by particles that are perfectly spherical (Hapke, 2012), but it adequately describes the effect of atmospheric aerosols. Mie scattering has much less of a dependence on wavelength than Rayleigh scattering and forward scatters much more highly (Schott, 2007). An example comparing the scattering profiles of Rayleigh and Mie scattering can be seen in Figure 3.4.

### 3.1.3 Solar Energy Paths

The path that light takes when interacting with the Earth and its atmosphere and returning to the sensor is of great importance to understanding the contributions of the cumulative measured signal. The paths most relevant to this work are illustrated in Figure 3.5.

The primary path taken by light is path A. In path A, light from the Sun passes through the atmosphere undisturbed, is reflected by the target and reaches the sensor without any interaction with the atmosphere. The next most important path is path B. Photons traveling in path B enter into the atmosphere and are scattered into the sensor before interacting with the target at all. This is called upwellled radiance (Schott, 2007), and is generally compensated for when performing analysis on the imagery. Path C occurs when photons pass through the atmosphere, interact with an area other than the target and then
are scattered by the atmosphere into the path of the sensor. This phenomenon is referred to as the adjacency effect, as it is most relevant when dealing with two areas of different material that are adjacent to each other, such as a small waterbody and surrounding woodland.

The remaining paths in Figure 3.5 (D, E, and F) contribute to the overall signal, but to a lesser extent than the paths A, B, and C. They are included for completeness. Path D occurs when light is scattered onto the target by the atmosphere and is reflected into the sensor. Similarly path E consists of light scattered by a nearby object onto the target and into the sensor. Finally the photons in path F reach the target undisturbed by the atmosphere, but they are scattered back onto the target by the atmosphere, and then reflected again by the target finally reaching the sensor. This is referred to as multiple scattering, and is not limited to interaction with the target. Multiple scattering can also occur between the various constituents of the atmosphere, contributing to the upwelled radiance seen from path B.
Figure 3.5: Various optical paths, labeled A through F, taken by light as it enters the atmosphere and is directed to the sensor.

### 3.1.4 Governing Equation

This qualitative understanding of the various paths of that light can take between sensor and target allows us to build a quantitative description. This description is called the governing equation, and in this work it will be simplified to contain only the relevant contributions within the reflective domain. We consider first the initial exoatmospheric irradiance which approaches the Earth at some zenith angle \( \phi' \):

\[
E_a \cos \phi'
\]

(3.9)

This irradiance then passes through the atmosphere to a target on the ground. Energy is then lost due to attenuating processes. The amount of energy transmitted to the surface
is described by $\tau_1$.
\begin{equation}
E_s \tau_1 \cos \phi' 
\end{equation}

The light that reaches the target is then reflected toward the sensor. The target’s reflection which is presented here as the target’s bi-directional reflectance factor ($r$) divided by $\pi$.
\begin{equation}
L_{Total} = E_s \tau_1 \frac{r}{\pi} \cos \phi' 
\end{equation}

Downwelled radiance, scattered by the atmosphere, also reaches the target and is reflected in the direction of the sensor. The diffuse downwelled sky irradiance is represented by the term $E_{sd}$. This term can represent light that reflects off of a nearby object onto the target as well.
\begin{equation}
L_{Total} = E_s \tau_1 \frac{r}{\pi} \cos \phi' + E_{sd} \frac{r}{\pi} 
\end{equation}

After reflection of the target the incident energy must travel back through the atmosphere towards the sensor. We can describe the energy loss during this section as $\tau_2$.
\begin{equation}
L_{Total} = \left( E_s \tau_1 \frac{r}{\pi} \cos \phi' + E_{sd} \frac{r}{\pi} \right) \tau_2 
\end{equation}

Finally we must also consider the contribution of signal from upwelled radiance due to atmospheric scattering or light reflected off of another target and into the sensor. This contribution is described as $L_u$.
\begin{equation}
L_{Total} = \left( E_s \tau_1 \frac{r}{\pi} \cos \phi' + E_{sd} \frac{r}{\pi} \right) \tau_2 + L_u 
\end{equation}

This simplified form of the governing equation, while omitting effects such as shape factor, contains all of the factors that determine what is observed by an imaging system. This includes issues such as upwelled radiance which contain no useful information about the target under observation. The approach to address these issues (i.e. Atmospheric Compensation) is described in Chapter 4, Methods.
3.2 Imaging Systems and Platforms

This section discusses imaging systems types and specifications that are relevant to this work. The first focus is on the different ways an imaging system can collect imagery, followed by descriptions of imaging system characteristics that effect their ability to be used for scientific analysis. Finally this section covers specific platforms and the imaging sensors they carry, along with their characteristics.

3.2.1 Imaging Systems

In this work, an imaging system is defined as a system which is capable of measuring radiant energy in discrete spatial samples. These systems have a wide variation of operation, spectral coverage, and performance, which influences its ability to perform certain tasks. These characteristics are discussed here as it is important in understanding how imagery is formed.

3.2.1.1 Framing Arrays

Framing Arrays are the most conceptually simple type of imaging system, as it is the dominant type of system used in consumer grade cameras (e.g. cell phones, DSLRs). These systems consist of a 2-D sensor array which is occluded by a shutter. This shutter opens and closes to expose the sensor to signal. The exposure time of these types of systems is often short as they are susceptible to motion blur (Schott, 2007). This type of sensor is ideal for platforms such as aircraft or drones because its nearly instant exposures allow it to be unaffected by movement.

3.2.1.2 Push-Broom Systems

Push-Broom systems use linear arrays of detectors to capture a single row of pixels at any time. This allows large swaths to be captured without the use of dynamic parts, only movement in the along track direction. Furthermore this increases the dwell time and therefore amount of received signal signal (Schott, 2007) which can be further increased by use of time delay and integration stages (Fiete, 2010). The constant forward track
movement of satellites make them ideal platform for push-broom sensors. Airplanes and UAS often change in orientation and speed, which poses potential error in data collection.

### 3.2.1.3 Other Imaging System Designs

Beyond framing arrays and push-brooms, the other most used imaging systems in remote sensing are line-scanners and whisk-broom systems. Line-scanners use a spinning mirror to scan across the scene of interest, and project the image of the scene onto the sensor (Schott, 2007). Line-scanners have been used in both aircraft and satellite systems, however they can suffer from changes in platform movement and generally have a short dwell time.

Whisk-broom systems use an oscillating scan mirror to sweep over multiple lines of data at a time. This sweeping is timed appropriately so all ground information is covered (Schott, 2007). Whisk-brooms are mostly used on satellite sensors as they also have issues with unstable platform movement. They offer more dwell time than a line-scanning system but do not allow for the spectral data to be inherently registered. Furthermore the amount of dynamic parts in a whisk-broom system makes them more susceptible to mechanism malfunctions, such as the scan line correction error that occurred with Landsat 7.

### 3.2.2 Sensor Performance Specifications

Specifications of imaging sensors can greatly affect their ability to be used in scientific research. The main characteristics important in this study, especially the modeling work, are spectral coverage, ground sample distance, signal-to-noise ratio (SNR), and radiometric resolution. The first, spectral coverage, is defined to be a system’s ability to collect light across the electro-magnetic spectrum. Spectral coverage is generally described by the spectral bands of a system. These bands are described by their responsivity as a function of wavelength, which is averaged into one single data point. This is often referred to as the system’s Relative Spectral Response (RSR). Systems can have few or many spectral bands, known as multi-spectral and hyper-spectral systems respectively (Schott, 2007).

Ground sample distance, or ground instantaneous field of view, is considered to be the spatial footprint captured by one pixel of the imaging system. It is considered to be
the projection of an array element onto the ground from the sensor optics. Fiete (2010) describes ground sample distance using the following equation:

\[ GSD = \frac{H \times p}{f} \]  

(3.15)

Where \( H \) is the sensor altitude, \( p \) is the sensor pixel pitch, and \( f \) is the focal length of the sensor. Ground sample distance can greatly affect how the scene being imaged can be resolved, and how much area gets averaged into a single pixel. It should be noted that this equation for GSD is a simplification of a more complicated set of parameters that defines the system’s full spatial resolution. This includes the resolving power of the optics as well as the pixel sampling interval.

Signal-to-noise ratio, or SNR, is simply the ratio of signal photons received to that of noise photons as described by equation 3.16.

\[ SNR = \frac{S}{\sigma} \]  

(3.16)

Here \( S \) is the signal, described as the portion of measured energy consisting of photons that have interacted with the target object or substance under study and arrived at the sensor. Noise, denoted by \( \sigma \), is considered to be measured energy that arises from other parts of the imaging chain, such as the propagation from the source or to the sensor, or the characteristics of the sensor. Noise can arise from both signal dependent sources, such as photons that interact with the atmosphere and shot noise, or signal independent sources, such as dark noise from current in the sensor, quantization noise from digitizing the signal, or read noise (Fiete, 2010). As noise is inevitable in one form or another, it is generally important to make sure that enough signal is being collected to allow for the highest SNR possible. This can be done in many ways including increasing exposure or dwell time, choosing bands with larger bandwidths, or spatial binning.

The final specification that will be covered is the radiometric resolution. Electro-optical sensors are only capable of recording radiometric energy discrete levels, a process commonly referred to as quantization. Therefore each level of quantization represents a discrete step in radiance, with lower quantization leading to larger steps. The number of quantization
levels can be described by:

\[
\text{Number of Quantization Levels} = 2^n
\]  \hspace{1cm} (3.17)

Where \(n\) is the number of bits (Schott, 2007). A consumer digital camera typically has 8 bits. The name radiometric resolution comes from the fact that the number of bits determines how well the amount of radiant energy coming from a target can be resolved.

3.2.3 Platforms

3.2.3.1 Landsat 8

Landsat 8 is the 8th system in a series of space based Earth imaging sensors that began in 1972 (Landgrebe, 1997). Landsat 8 was equipped with two instruments, the Operational Land Imager (OLI), which captured data in the visible and infrared/near-infrared regimes, and the Thermal InfraRed Sensor (TIRS), which captured two bands of information in the longwave infrared (Irons et al., 2012). This project mainly focuses on the OLI.

The OLI captures imagery in seven bands, four in the visible, one in near-infrared, and two in the shortwave infrared each of which has a GSD of 30 meters. The relative spectral response of these bands can be seen in Figure 3.6. Also included is a panchromatic band (500 nm to 680 nm) and a band for detecting cirrus clouds (1360 nm to 1380 nm) with 15 meter and 30 meter GSD respectively. OLI increased its capability for imaging over the Enhance Thematic Mapper Plus (ETM+) on the Landsat 7 system through having an added coastal band, an improved signal to noise ratio, and a greater radiometric resolution (8 bit to 12 bit) (Gerace et al., 2013). The OLI was also designed as a push-broom imaging system, as opposed to whisk-broom designs of previous Landsat systems, after testing the design using the Advanced Land Imager (ALI) on the Earth Observer-1 satellite (Lencioni et al., 1999). Landsat 8 has a temporal revisit time of 16 days, however areas in overlapping collection paths are captured nearly weekly.
3.2.3.2 Other Satellite Systems

This work is mainly focused on the use of Landsat 8, however the methods tested are applicable to other systems as well. In this work we also examine the use of Sentinel-2, and potential designs for future Landsat missions. The design of Landsat 9 has already been finalized as a copy of OLI with an improved bit depth (Wulder et al., 2019), but designs for following missions have yet to be settled on. This work intends to determine potential designs of future systems that improve Landsat’s ability to monitor water quality.

Sentinel-2 is a satellite imaging system launched by the European Space Agency. On-board is Sentinel-2 is the Multispectral Instrument (MSI) which is a multi-spectral push-broom system with 4 visible, 6 infrared, and 3 near infrared bands. The RSR of all bands in the 400 nm to 900 nm range is shown in Figure 3.7. The GSD of these bands range from 10 m, 20 m, and 60 m. Sentinel-2’s radiometric resolution is 12 bits. As Sentinel-2 is a constellation of 2 satellites its temporal revisit rate is on the order of 5 days (Drusch
et al., 2012).

Figure 3.7: Relative Spectral Response of Sentinel-2’s visible and near-infrared spectral bands in the 400 nm to 900 nm region.

3.2.3.3 Unmanned Aerial Systems

UAS are capable of having numerous imaging systems attached. One of the two imaging systems used for this study was the Tetracam Micro-MCA 6 Snap. The Micro-MCA 6 Snap is a multi-spectral framing array system that has six CMOS cameras on which spectral filters can be installed interchangeably. The filters installed in the camera were (including band center/width): Blue (490 nm/10 nm), Green (550 nm/10 nm), Red (680 nm/10 nm), Near-Infrared 1 (720 nm/10 nm), Near-Infrared 2 (800 nm/10 nm), Near-Infrared 3 (900 nm/20 nm). Each camera has a 1280 x 1024 pixel detector array, pixel pitch of 5.2 m, and a focal length of 9.6 mm. The average collection altitude of this system was about 200 meters, meaning that it had a GSD of about 10 cm. The exposure time of the Tetracam
was used both in fixed and variable modes. This system also captured GPS latitude and longitude with each image. The platform used for this system was an octocopter, the DJI Spreading Wings 1000. The spectral band transmission and sensor response of the Tetracam can be seen in Figure 3.8.

![Figure 3.8: Spectral filter transmissions (colored) and sensor response (Black) of the Tetracam imaging system.](image)

The second system used in this work is the Headwall Nano-Hyperspec, a hyperspectral push-broom system. The Nano-Hyperspec, occasionally referred to as just Nano, has 270 bands throughout the 400 nm to 1000 nm range. As it is a push-broom system, the light captured by this system is dispersed over 670 pixels with a pixel size of 7.4 µm. The sensor of the Nano-Hyperspec is a CMOS sensor, and has a 12 bit quantization. The Nano was implemented as part of the MX-1 sensor array, and flown on a DJI Matrice 600 Pro UAV (Kaputa et al., 2019).
3.3 Radiative Transfer in Water

This section focuses on how light interacts with a body of water and its constituents. First the major optically active components found in a waterbody will be discussed as well as their inherent optical properties. This is followed by the major energy paths that light can take in water, similar to the solar energy paths seen in Section 3.1.3 and a section on how glint effects the signal received by an imaging sensor. This section ends with a discussion of Hydrolight, a radiative transfer code that models the light field in a waterbody using specified properties.

3.3.1 Inherent Optical Properties

In Section 3.1.2, absorption and scattering were defined in the context of how light interacts with the atmosphere. These two phenomenon were described as being optical properties of the particles found within it. Furthermore these are considered to be inherent optical properties (IOPs) which are optical properties that do not depend on the ambient light field (Mobley, 1994). Absorption and scattering are the predominate drivers of water color and therefore should be more formally defined, as well as a few other terms, for this work.

Absorption can be described first as the amount of energy that is lost as light passes through a medium:

\[ A = \frac{\Phi_a}{\Phi_i} \]  

(3.18)

Where \( \Phi_a \) is the flux that is absorbed by the medium, and \( \Phi_i \) is the total flux that enters the system. It should be noted that both fluxes can vary with wavelength, making absorption a spectral quantity. This is more usefully defined as the spectral absorption coefficient (Mobley, 1994):

\[ a = \lim_{r \to \infty} \frac{A}{r} \left[ \frac{1}{m} \right] \]  

(3.19)

In this case \( r \) is the path length that the light travels within a medium. A related term, the specific spectral absorption coefficient (\( a^* \), units: \( m^2/mg \)), can be determined by dividing this term by the concentration of the component within the medium.
Scattering unlike absorption is not a loss of energy, but only a redirection of energy from its path of travel. The notation for scattering is similar to that of absorption. In this case scattering can be defined as:

\[ B = \frac{\Phi_s}{\Phi_i} \]  

(3.20)

Where the numerator is the amount of light that is scattered \((\Phi_s)\) instead of the amount of light that is absorbed. Similarly the spectral scattering coefficient \((b)\) and the specific specific spectral scattering coefficient \((b^*)\) can be defined by applying the same operations that were applied to the absorption. The spectral scattering and absorption coefficient, as well as their specific counterparts, can be summed together to make the total attenuation coefficient of the water.

\[ c(\lambda) = a(\lambda) + b(\lambda) \]  

(3.21)

The \(\lambda\) dependency is shown in this equation to infer that the attenuation coefficient is spectral as well.

The definition given for the scattering coefficient only gives information to how much is scattered away from the direction of travel spectrally, but does not provide directional information. The volume scattering function describes both the spectral and directional nature of scattering. The volume scattering function is defined as:

\[ \beta(\psi, \lambda) = \lim_{\Delta r \to \infty} \lim_{\Delta \Omega \to \infty} \frac{\Phi(\psi, \lambda) \Delta r \Delta \Omega}{\Phi_i(\lambda) \Delta \Omega} \]  

(3.22)

Where \(\Delta \Omega\) is the solid angle containing the angular information of the scattered light. This term can be related back to the spectral scattering coefficient purely by integration of all solid angles.

\[ b = \int_0^{2\pi} \int_0^\pi \beta(\psi, \lambda) \sin \psi d\psi d\phi = 2\pi \int_0^\pi \beta(\psi, \lambda) \sin \psi d\psi \]  

(3.23)

In this case \(\psi\) is the angle between the initial and scattered travel direction while \(\phi\) is the azimuthal angle of the incoming light. The volume scattering coefficient can similarly be related for the forward \((b_f)\) and backward \((b_b)\) scattering coefficient by setting the limits of \(\psi\) as 0 to \(\frac{\pi}{2}\) or \(\frac{\pi}{2}\) to \(\pi\), respectively.

The volume scattering function can also be used to derive a term that solely describes
the directional nature of the scattered light. This term is called the scattering phase function and is defined as the volume scattering function divided by the spectral scattering coefficient:

\[ \tilde{\beta} = \frac{\beta(\psi, \lambda)}{b(\lambda)} \]  

(3.24)

The shape of a scattering phase function is dependent on the water constituents. The most notable measurement of scattering phase function in water was done by Petzold (1972). This phase function was defined by the ratio of back scattering coefficient to scattering coefficient. Petzold found this value to be about 0.018, or 1.8%. This work assumes that a Fournier-Forand scattering phase function (Fournier and Forand, 1994), whose shape is dependent on the ratio of back scattering to scattering, approximately represents that of water constituents. Figure 3.9 shows a few different Fournier-Forand scattering phase functions, as well as the Petzold phase scattering function.

Figure 3.9: Fournier-Forand scattering phase functions of different back scatter ratios in red. The black line is the Petzold scattering phase function.

(Source: Mobley and Sundman (2013))
The nature of IOPs is variable, especially when applied to the constituents of waterbodies. An example of this is shown in Figure 3.11, where a series of measurements of Chlorophyll-a for a single body of water (Owasco Lake) across different areas and different days is shown. This natural variability must be accounted for in any remote sensing water monitoring procedure to be robust.

3.3.2 Water Constituents

The optical signature of waterbodies are defined by the IOPs of water as well as its constituents. Gordon and Morel (1983) developed a classification system for waters of varying constituents. First is Case 1 waters, where the optical signal is heavily dominated by photosynthetic entities and color dissolved organic matter. Case 2 waters have the same composition as Case 1, but also contain suspended sediment/materials. This work focuses on monitoring of Case 2 waters. The following subsection discusses each of these constituents, as well as pure water and phycocyanin, a pigment found in cyanobacteria.

3.3.2.1 Pure Water

It is important to understand that pure water is only one of the constituents that make up a waterbody. Pure water is very weakly absorbing in the visible regime, while highly absorbing outside of it, thus the reason that ocean color algorithms are based in the visible. Algorithms for atmospheric compensation and glint removal often take advantage of the fact pure water’s high near-infrared absorption in Case I waters, as it can be assumed that in there is no signal from water. Generally in small quantities pure water appears to be colorless, however in large quantities it appears blue due to the molecular scattering much like the atmosphere. Water scattering is better approximated by Einstein Smoluchowski scattering rather than rayleigh scattering, however both are similar in their dependency on wavelength ($\lambda^{-4}$) (Mobley, 1994). The absorption spectrum of pure water as measured by Smith and Baker (1981) and Pope and Fry (1997) as well as scattering spectra of both sea water and fresh water from Morel (1974) can be seen in Figure 3.10. The phase scattering function can be seen in green in Figure 3.9.
3.3.2.2 Chlorophyll (Chl)

Chlorophyll is a pigment found in phytoplankton, photosynthetic bacteria (e.g. cyanobacteria), and plants/macrophytes. Chlorophyll is a photosynthetic pigment used to absorb radiant energy (Avila, 1995). There are multiple forms of chlorophyll such as a, b, and c but this work focuses on specifically on chlorophyll-a. In the rest of this work it will often be referred to as just chlorophyll or Chl. Chlorophyll-b and c are considered to be accessory pigments. Chl has a strong fluorescence signal with an excitation wavelength around 430 nm and emission around 665 nm (Yentsch and Menzel, 1963). Absorption spectra and scattering spectra of Chl can be seen in Figure 3.11.

The absorption of Chl can be highly variable between waterbodies or even within one. This is largely a result of the way various photosynthetic entities structure their cells
and store Chl. This phenomenon is known as the packaging effect (Kirk, 1984). Cell size, irregularities, physiological state, and accessory pigment presence are also sources of variability in measured Chl absorption spectra. This is exacerbated by the fact that these properties also depend on the nature of the environment in which the entity develops.

The absorption spectrum of Chl is signified by its high absorption in the 400 nm to 500 nm range with peaks near 440 nm and 680 nm. Scattering spectra are also sensitive to the variations in the same cell properties as absorption (Bukata et al., 1995). As Chl is organic, it is generally considered to contribute less to waterbody scattering than other components such as suspended solids (Stramski et al., 2004). It should also be noted that cyanobacteria cells generally show higher levels of scattering than some phytoplankton due to gas vacuoles within the cell (Ganf et al., 1989; Ahn et al., 1992). The scattering profile of Chl is considered to generally decrease with increasing wavelength.

Figure 3.11: Specific absorption and scattering spectra of Chl. The absorption spectra were measured in western New York waterbodies. The scattering spectrum is taken from the literature (Bukata et al., 1995)
Examining the relation between Chl mass specific absorption and concentration, Bricaud et al. (1995) determined that Chl mass specific absorption tends to decrease with increasing concentration, as shown in Figure 3.12. This was considered to be a result of two factors, changes in the packaging effect and the increased presence of accessory pigments. Changes in the packaging effect were assumed to be due to the differences of dominant species cell size and structure between oligotrophic and eutrophic waterbodies, as well as differences in cell pigment concentration with depth. Accessory pigments were assumed to have an impact due to reported higher concentrations in oligotrophic waterbodies than eutrophic.

### 3.3.2.3 Color Dissolved Organic Matter (CDOM)

Color/Chromophoric Dissolved Organic Matter (CDOM) is the name for decayed organic matter within the waterbody, often less that 0.4\( \mu \)m in size (Mobley, 1994). In the literature it can also go by the name gelbstoff, yellow matter, or gilvin. CDOM can occur from terrestrial runoff as well as decaying of phytoplankton and other aquatic organic life and is often found in greater concentration in inland and coastal waters than open ocean (Kirk, 1984; Bukata et al., 1995). CDOM absorption increases exponentially towards shorter wavelengths (Bricaud et al., 1981) leading to a yellowish color in waterbodies where it is present. CDOM is also highly variable in its absorption, which is generally represented by the slope of decrease in absorption, referred to in exponential models as the slope factor (Equation 3.25).

\[
a_{CDOM}(\lambda) = a_{CDOM}(443nm)e^{-S(\lambda-443nm)}
\]

(Equation 3.25)

Here \( a_{CDOM} \) is the spectral absorption of CDOM, \( S \) is the slope factor, and \( \lambda \) is the wavelength in nm. One cause of this variability is composition, for example CDOM dominated by humic acid generally has a much larger absorption that dominated by fulvic acid (Bukata et al., 1995). CDOM variability can also be affected by solar photobleaching (Vodacek et al., 1997). A range of CDOM absorption measurements taken in New York state can be seen in Figure 3.13 normalized at 440 nm. No scattering data is included as CDOM scattering is considered to be negligible (Dall’Olmo et al., 2009).
3.3.2.4 Total Suspended Solids (TSS)

Total Suspended Solids (TSS) represents suspended particulate matter within the water column (Kirk, 1984). This term is broadly inclusive, encapsulating organic and inorganic matter. The absorption spectrum of TSS is very similar to that of CDOM, increasing with decreasing wavelength. They are differentiated in that TSS also scatters light, and generally has the greatest scattering of all waterbody components. The scattering spectra
Figure 3.13: Normalized absorption spectra of CDOM, each fit to an exponential. The original absorption spectra were measured in western New York waterbodies.

of TSS also decreases monotonically with increasing wavelength, though at a much slower rate than absorption, with notable scattering in the NIR at high concentrations. This causes high TSS waters to appear bright with a brownish color and also is the reason that standard water remote sensing algorithms that make the black pixel assumption (i.e. no signal in the NIR) are often invalid for turbid waterbodies. The absorption and scattering spectra for TSS can be seen in Figure 3.14. The absorption and scattering IOPs show large natural variation due to their diversity in size, shape composition, and internal structure. Particle sizes of TSS can range anywhere from 0.4 \( \mu m \) to 250 \( \mu m \) or greater (Bukata et al.,
1995), though particle concentration has been reported to typically decrease with particle size (Stramski et al., 2004). The particle assemblages which drive the variability vary greatly due to geographic and limnological properties.

Figure 3.14: Specific absorption and scattering spectra of TSS. The absorption spectra were measured in Western New York waterbodies.

Tilstone et al. (2012) reported a linear relation between non-algal particle absorption and Total Suspended Material (TSM, i.e. TSS) as measured in the North Sea and English Channel. This relation is shown in Figure 3.15. This linear relation infers that mass specific absorption is also linear, though there is some variance to its exact value due to the aforementioned reasons (e.g. composition, particle size). As expected the mass specific absorption reported in the literature varies between datasets (Kirk, 1984; Babin et al.,
3.3.2.5 Phycocyanin (PC)

Phycocyanin is a biliprotein pigment found almost exclusively in cyanobacteria (Kirk, 1984). This pigment is used to increase the amount of light harvested for photosynthesis. The visual appearance that this pigment creates is the reason that cyanobacteria is sometimes referred to erroneously as "blue-green algae". The absorption spectrum of phycocyanin is characterized by an absorption maximum centered near 620 nm. Scattering of phycocyanin, to the best of our knowledge, has never been directly measured, however for this work it has been estimated by subtracting the specific scattering of chlorophyll from Bukata et al. (1995) from the specific scattering of cyanobacteria measured by Ahn et al. (1992). The absorption and scattering spectra for PC can be seen in Figure 3.16.
3.3.2.6 Covariation of Waterbody Component Concentrations

Concentrations of waterbody components have been reported to covary within the literature. This would seem fairly intuitive as phytoplankton cells contain both the pigment Chl as well as other organic matter that falls within the categorization of TSS. Furthermore, as phytoplankton die off they can decompose into small particles that can be categorized as CDOM. Babin et al. (2003) reported the relations between Chl, TSS (labeled as suspended particulate matter, SPM), and CDOM for multiple European waterbodies, shown in Figure 3.17. The relation between these three components is shown to generally increase together. This may not necessarily be true for waterbodies where TSS is predominately composed of inorganic materials (e.g. sediments) which may attenuate enough light to limit the growth of phytoplankton.
As phycocyanin and chlorophyll are both light harvesting pigments found in cyanobac-
teria, they also show a general relation in concentration. The reported mean ratio of phycocyanin to chlorophyll ranges from 0.27 to 2.14, most often being reported as $\approx 1$. (Simis et al., 2005; Randolph et al., 2008; Ruiz-Verdú et al., 2008; Mishra and Mishra, 2014; Li et al., 2015). It has also been reported as high as 5.19 by Li et al. (2015). This variation is stated to be dependent on the conditions of the waterbody and availability of light. It should be noted that while high concentrations of phycocyanin and chlorophyll are generally found together, this is not true for the reverse case (i.e. high concentrations of phycocyanin and low concentrations of chlorophyll).

### 3.3.2.7 Spatial and Temporal Variability of Cyanobacteria Blooms

Cyanobacteria blooms often show high spatial and temporal variability, leading to difficulty in monitoring. In the attempt to monitor cyanobacteria blooms using Hyperion and the Advanced Land Imager, Kutser (2004) noted that retrieved concentrations of chlorophyll ranged two orders of magnitude within tens of meters of the sampling area. Kutser (2004) noted that this may indicate that even fine scale spatial resolution system like Landsat may be inadequate for resolving the full structure of cyanobacteria blooms. Furthermore, Kutser et al. (2008) performed a modeling study looking at the impacts of vertically distributed cyanobacteria concentration and found that the distribution of chlorophyll had a notable effect on the remote sensing signal. Spatial and temporal variability have been reported both on diurnal (Qi et al., 2018) and seasonal scales (Wynne and Stumpf, 2015; Vander Woude et al., 2019). As such the use of point sampling within the top meter of the waterbody, as is done in this work, should be noted as a potential source of error for retrieval.

### 3.3.3 Energy Paths in Water

Light’s interaction with water can be summarized by the use of energy paths, much like the discussion of solar energy paths in Section 3.1.3. This can be considered an extension of the solar energy paths, as it describes how paths that reach the water target interact with it. These water paths are described by the Roman numerals I-VII, and are shown in Figure 3.18.
The first path, path I, describes light which is scattered by pure water into the sensor. Path II is light that makes it through the entire length of the water column to end up interacting with the bottom of the waterbody and being scattered back towards the sensor. This signal reaching the sensor from the bottom is sometimes called the bottoming effect. In path III, the light is absorbed by CDOM, with no scattering. In path IV, the light reflects off of the surface of the air-water boundary and is reflected into the sensor, usually as a specular reflection. This is known as glint and is covered in more detail in Section 3.3.4.

The final paths, V through VII describe how different constituents of the water affect the energy paths. Path V describes the interaction of light with chlorophyll which as previously stated is found in phytoplankton and cyanobacteria within the water. Energy from this path is partially absorbed and partially scattered. Similarly, in Path VI, light interactions with phycocyanin which is also a pigment found in cyanobacteria. In this case light is also partially scattered and absorbed. Finally, Path VII describes the interaction of suspended materials (SM, i.e. TSS) and light, which as with the previous two constituents, scatters and absorbs light.

Each of these processes have an affect on the light energy that reaches the sensor. Although these energy paths are described separately, photons may interact with multiple constituents multiple times on its path to the sensor. Any model needs to take into account any possibility of interactions in order to produce accurate results as does Hydrolight which is discussed more in Section 3.3.5.

3.3.4 Glint

As mentioned in Section 3.3.3, glint is the reflection of light off of the surface of the water. Glint manifests in two forms, Sun glint and sky glint, the former being the reflection of light from the Sun and the latter the reflection of sunlight scattered by the atmosphere. Fresnel equations describe the amount of energy that is reflected by the water’s surface.

Although glint may carry information about the waterbody, such as the shape of the surface or wave height (Cox and Munk, 1956) it is undesirable in water color remote sensing. Effects of glint can be minimized both through collection and post processing means.
The amount of glint collected by an imaging system is highly dependent on the Sun-water-sensor angle. It is possible to set the sensor viewing angle to minimize glint when collecting imagery. This technique has been implemented on water remote sensing focused missions in the past (e.g. SeaWiFs) (Gregg and Patt, 1994). Fougnie et al. (1999) suggest keeping the sensor facing a 40° zenith angle, and 135° azimuthal angle from the Sun. Furthermore the work by Fink (2014) developed code to determine high glint areas based on the flight plan and solar position allowing sensor viewing angles to be set accordingly.

Although it is favorable to minimize glint effects during collection, it is also possible to remove them using post processing algorithms. Kay et al. (2009) produced a survey of post processing methods for correcting glint. This work organized methods into two

Figure 3.18: Various optical paths, labeled I through VII, taken by light within a waterbody.
groups, statistical removal methods and near-infrared methods. The statistical methods were focused on modeling how glint should appear, and required systems with coarse spatial resolutions (e.g. satellites). The near-infrared methods can be implemented on systems with fine spatial resolutions and appropriate spectral coverage, however these methods assume that glint is the major contributor of the NIR region of the spectrum, which is an invalid assumption in turbid water. It is also possible to incorporate glint removal as part of the atmospheric compensation process, assuming in-situ measurements of the waterbody are available (Ford and Vodacek, 2019).

3.3.5 Hydrolight & Ecolight

Hydrolight is a radiative transfer model focused on the transfer of light through bodies of water. It is based on the work of Dr. Rudolph PresENDORF and Dr. Curtis Mobley (Mobley, 1994). Hydrolight numerically solves a one dimensional time independent radiative transfer equation to determine factors such as the water leaving radiance, irradiance, and remote sensing reflectance over a discretized hemisphere. This makes the output applicable to sensors at various orientations. Hydrolight also includes Ecolight which solves the radiative transfer equation only for the orientation nadir to the water, allowing for decreased computation time (Mobley and Sundman, 2013), but is otherwise identical in operation. This work primarily uses the Ecolight model. Hydrolight is also capable of modeling in-water light fields at any depth, although computation time increases linearly with depth. All solutions include the effects of wave shape as well as internal sources and inelastic scattering (e.g. Raman scattering, chlorophyll fluorescence, and bioluminescence).

Hydrolight uses inputs of IOPs (e.g. absorption and scattering) from any number of waterbody constituents to define parameters within the radiative transfer equation. IOPs can be entered from both field measurements or models. The radiative transfer equation is solved numerically using the invariant imbedding technique, which allows for efficient calculation of all orders of multiple scattering between components. This technique determines transmittance and reflectance values as a function of optical depth from the water surface which can be imbedded between conditions of the air-water and water-bottom boundaries. The input sky irradiance can be calculated using analytical and semi-empirical...
models from the Hydrolight software or determined purely by user input. The boundary conditions of a non-flat air-water surface are determined by Monte Carlo simulations based off Cox and Munk wave slope statistics, while the bottom boundary is defined in terms of its reflectance. Combining the transmittance, reflectance, and boundary conditions allows a determination of the water leaving radiance based on the full path that light travels.

There are limitations to the Hydrolight model. Hydrolight is only capable of producing spectra between 300 to 1000 nm, though it is possible to extended this range through a rigorous number of changes to built-in spectra as well as user input. Hydrolight only uses a linear interpolation to fill missing gaps in data (e.g. varying concentration with depth) which may lead to error in the modeled spectrum. Any neccessary extrapolation only uses the last known value as a constant, which is often not physically accurate. As Hydrolight solves the radiative transfer equation in one dimension, it cannot simultaneously model waters with constituent properties that vary spatial or temporally, though it can be applied to each set of constituent properties in turn. White caps and wave shadowing are not considered when Hydrolight models the effects of wave shape which can affect the accuracy of modeling certain conditions. Hydrolight also assumes a parallel plane waterbody in its calculations. Finally polarization is not currently a feature in Hydrolight, although this functionality is currently in development (Mobley, 2012).

3.4 Past Work in Remote Sensing of Cyanobacteria Blooms

This section discusses previous work done using remote sensing for extracting information about cyanobacteria blooms. These works generally fall into two categories, empirical methods (indexes, band ratios, or regressions) and semi-empirical/semi-physical methods (algorithms partially based on physical properties of the water).

3.4.1 Empirical Methods

Landsat sensors were the main focus of a bloom retrieval attempt by Vincent et al. (2004). Concentrations of chlorophyll-a and phycocyanin were determined via fluorometric analysis performed in the field. These collections were concurrent with overpasses of both Landsat 7 and Landsat 5, in July and September respectively. Atmospheric compensation
was performed via dark object subtraction. Multiple single band and spectral ratio multiple regression models were tested, however only the best performing models were reported. In the single band model case, the best performing model retrieved phycocyanin with an $R^2$ of 0.74 from L7 imagery while the spectral ratio model had an $R^2$ of 0.77. The $R^2$ values were not reported for the Landsat 5 retrievals due to high error. This method was not tested for years other than 2000, and not applied to other waterbodies.

Kahru et al. (2007) used satellite imagery from SeaWiFS, CZCS, and MODIS in order to observe the interannual and spatial variability of cyanobacteria blooms in the Baltic Sea. Pixels with accumulations of cyanobacteria were determined by looking at the radiance level of each system’s green band, typically around 550 nm band, after the imagery had been atmospherically compensated in SeaDAS. A radiance level above $0.8 \text{ mW/(cm}^2 \text{ um sr)}$ implied the pixel had an accumulation of cyanobacteria, unless it was flagged as being turbid (by a threshold radiance at 670 nm), cloudy, or showing bottom/coast (determined by SeaDAS processing). The frequency of cyanobacteria accumulations was determined by finding the ratio of cyanobacteria accumulations to non-flagged scenes from July-August. Though concentration retrievals were not attempted, it was reported that there was a positive correlation between residual phosphate in May-June and frequency of accumulations. This relation was only valid for the entire sea, and could not infer spatial distribution for the bloom.

One of the most notable empirical methods, the Cyanobacteria Index (CI), was first written about in a paper by Wynne et al. (2008). This method is notable as it is still used by the National Oceanic and Atmospheric Administration (NOAA) to distribute information about blooms. The theoretical basis for this method is comes from Ganf et al. (1989), where it was noted that cyanobacteria scattering near 700 nm produced a signal greater than chlorophyll fluorescence. It was also noted that cyanobacteria had a lower fluorescence than is generally seen in algae (Seppala et al., 2007) and that it has a greater scattering effect due to its gas vacuoles. This led to the development of a spectral shape algorithm based on the second derivative of the spectrum in the red/near-infrared, analogous to fluorescent line height. There was no attempt at extraction of information in this 2008 paper, however a paper by Wynne et al. (2010) would continue the work on the original algorithm. In this paper, there was an attempt to relate CI statistically to cell counts,
which produced an $R^2$ value of 0.48. Lunetta et al. (2015) would attempt to build on the use of the CI. This study looked at MERIS bloom imagery in New England, Ohio, and Florida which had associated cell counts and again attempted to find a statistical fit to cell count with an improved $R^2$ of 0.87. The majority of error in this work came from misidentifying high and moderate concentration levels. Tomlinson et al. (2015) also attempted to build further on this work. A radiometer was used to collect field spectra of ongoing blooms. Associated chlorophyll concentrations were also measured. The measured spectra was sampled to MERIS bands. A linear relation between chlorophyll concentration and CI was determined, and this relation was implemented on MERIS data from 2008-2009. A root-mean squared error in retrieval of 27% was reported.

Matthews et al. (2010) attempted to retrieve concentrations of water quality components from Lake Zeekoevlei in South Africa. Empirical algorithms were determined for MERIS imagery in order to retrieve concentrations of Chlorophyll-a, Total Suspended Solids, Secchi Disk depth, and absorption by CDOM but not phycocyanin. A MERIS Neural Network was also created to attempt to retrieve these same values. Results of the Neural Network were not reported due to high errors. This was attributed to artifacts from atmospheric compensation and the simulated training data generated using of IOPs that did not properly represent the waterbody being studied. The empirical algorithms developed performed notably better with Chlorophyll-a having an $R^2$ of 0.964 and an RMSE of 9.8%. Similarly TSS and CDOM had an RMSE of 14.1% and 13% respectively. These algorithms were not applied to other waterbodies for testing. It was stated that future work may implemented to create empirical algorithms for retrieval of phycocyanin.

Kudela et al. (2015) used water samples collected in three California water bodies to test how well various indices could be empirical related to cyanobacteria blooms. The water samples measured cell abundance, surface chlorophyll concentration, temperature, toxin concentration, and phycocyanin concentration. The indices used were the CI (Wynne et al., 2008, 2010), the Scattering Line Height index (an index with the same theoretical basis as the Cyanobacteria Index but adapted to the MODIS/ASTER Airborne sensor), and the Aphanizomenon-Microcystis Index (which focused on determining toxic from non-toxic genera of cyanobacteria). Kudela reported an $R^2$ for the relationship between the first two indices and phycocyanin concentrations based on both imagery from HICO and radiome-
ter measurements. The Cyanobacteria Index had an $R^2$ of 0.14 and the Scattering Line Height Index had an $R^2$ of 0.4. Although no quantified assessment of the Aphanizomenon-Microcystis Index was published, it was stated that it seemed to show the general trend of the waterbodies switching between the dominance of each genera.

Shi et al. (2015) used a MODIS index to retrieve concentrations of toxin in waters undergoing a cyanobacteria bloom. Shi et al. used two data sets, one ranging from 2003 to 2013 and another from 2009/2010, both from Lake Taihu. The former data set contained chlorophyll concentrations, and the latter had both chlorophyll and microcystin concentrations. Shi’s methodology required determining an index to relate MODIS imagery to chlorophyll concentrations, and then used a linear regression to relate chlorophyll concentration to microcystin concentration. Multiple indexes were tested to find the one with the best regression with chlorophyll concentration. The index chosen was reported to have a $R^2$ relation of 0.75. The $R^2$ regression between chlorophyll and microcystin was determined to be 0.83. This algorithm was then used to retrieve microcystsins and study its trends from historical MODIS imagery.

3.4.2 Semi-Empirical Methods

Kutser (2004) published a paper on retrieving chlorophyll concentrations of cyanobacteria blooms from satellite imagery. Kutser used imagery from Hyperion and ALI, corrected using FLAASH. The algorithm presented in this paper is based on the semi-empirical model of Gordon and Wang (1994), with values taken from Kirk (1984). This model is augmented by a factor proposed by Austin (1980) to take into account the air-water interface. This model was used to build a spectral library where the only varying factor was chlorophyll absorption based on concentration. Each pixel in imagery was compared to this library using the Spectral Angle Mapper (Kruse et al., 1993), where the mapping determined concentration. Although a research vessel was measuring chlorophyll concentrations using a flow-through system, no quantitative measures of error are reported. Kutser states that this is due to concerns of the ship disrupting the bloom structure, as well as high variability of retrieved concentrations between adjacent pixels. Kutser finishes the paper by stressing concern for chlorophyll retrieval given the coarse spatial resolution of imaging satellites.
Metsamaa et al. (2006) used modeling work to test the effectiveness of this method and found that chlorophyll concentrations needed to be close to 8 \( \text{mg/m}^3 \) in order to clearly show absorption peaks.

Simis et al. (2005) focused on retrieving concentrations of phycocyanin. Spectral data of two lakes in the Netherlands was captured by using a radiometer and water samples were collected to measure phycocyanin concentration. A semi-empirical model was developed that related the spectral ratio of two wavelengths to the inherent optical properties of the waterbody based on the work of Gordon et al. (1975). This was used to extract the absorption of phycocyanin at 620 nm which was then divided by the specific absorption the same wavelength to determine concentration. When the specific absorption of each water sample was determined and applied this retrieval had an \( R^2 \) of 0.94. It performed significantly worse when a single specific absorption value was used, having only a \( R^2 \) of 0.77 and was highly skewed for low concentrations. This paper was followed by another by Simis et al. (2007), focused on the effects of the accessory pigments chlorophyll-b and c on phycocyanin retrieval. This time water samples were taken from Norwegian and Spanish lakes. Only a single specific absorption value was used and the best \( R^2 \), 0.9, was found when the effects of chlorophyll-b were removed.

Wang et al. (2016) used a data set of remote sensing reflectances, absorption coefficients, and phycocyanin concentrations from a Mississippi pond (taken of Mishra et al. (2013)) to build a retrieval algorithm for hyperspectral systems. This algorithm used the semi-empirical model of Gordon et al. (1988) and the air-sea interface effect from Lee et al. (2002) to model remote sensing reflectance curves. The absorption and backscattering coefficients were taken from measurements and models throughout literature. The absorption coefficient for phytoplankton was modeled as a series of Gaussian curves centered at the absorption peaks of their various pigments, including phycocyanin, following Hoepffner and Sathyendranath (1993). The peak value at 617.5 nm was considered to be the peak absorption coefficient for phycocyanin. The phycocyanin concentrations from the Mississippi data set and the retrieved absorption coefficient values were used to find a power relation which was implemented on a set of radiometer measurements and phycocyanin concentration from Lake Taihu. Wang reports an \( R^2 \) regression of 0.96 between the measured and retrieved values.
3.5 Past Implementations of Hydrolight LUTs

Though by no means the first attempt at physical model LUT inversion, one of the first documented uses of Hydrolight to invert waterbody properties was by Raqueno (2003). This work serves as the initial basis for the retrieval process used in this document. Hydrolight LUTs generated using only a singular set of IOPs (taken from Bukata et al. (1995)) were implemented in retrieval of waterbody components from the Rochester Embayment in Western New York. A retrieval process, including spectral matching and interpolation across LUT entries, was applied to hyperspectral Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data to retrieve concentrations of Chl, TSS, and CDOM. Interestingly, this is the first documented use of the Model-Based Empirical Line method, which will be described in more detail in 4.5.2. The accuracy of retrieval through this method was determined to be satisfactory with a root-mean-square error for Chl, TSS, and CDOM being 12 $mg/m^3$, 4 $g/m^3$, and 2.2 $m^{-1}$, respectively.

In an approach to perform bathymetry instead of monitoring of water quality, Mobley et al. (2005) developed a spectral matching process using Hydrolight LUTs. The LUTs used 11 sets of absorption and scattering measurements (based on potential chlorophyll concentrations and in-situ measurements), 63 potential bottom reflectances, and 60 depths for a total number of 41,591 entries. This LUT was applied to hyperspectral imagery from the Ocean Portable Hyperspectral Imager for Low-Light Spectroscopy over relatively clear waters near Lee Stocking Island in the Bahamas. The retrieval process purely consisted of spectral matching, no interpolation between spectra was performed. Retrievals where spectral bands were and were not weighted in the matching process based on importance, referred to as the constrained and unconstrained test respectively, were performed. In the bathymetry task, both the constrained and unconstrained LUTs performed satisfactory, though the constrained retrieval showed better performance which increased further when IOPs deemed less likely to be present where removed from the LUT. Mobley et al. (2005) attributed this success of this method mainly to the accurate calibration of the imagery data (i.e. atmospheric compensation), as well as the accurate waterbody properties (e.g. IOPs, bottom reflectances) used in the LUT generation. Lesser and Mobley (2007) would revisit this process for benthic classification where similarly the constrained approach out-
performed the unconstrained approach by a small margin. It was also determined that the constrained approach could result in increased error if inappropriate limitations of the LUT generation and search were applied. Mobley would continue to develop the LUT approach, examining aspects of the retrieval process such as the spectral matching algorithm.

Liu and Miller (2008) attempted to use Hydrolight LUTs in a process to correct for sources of inelastic scatter within waterbodies. While this work was not a direct process of performing model inversion using spectral matching, it did provide a sensitivity study to determine which inputs to Hydrolight most affected the simulated spectrum. It was determined that the IOPs and inelastic scatter played much larger roles than the ambient optical environment and surface wind speed in shaping the water spectrum, which would be expected through the use of Rrs. This also showed agreement to a similar study performed by Gerace (2010).

The LUT retrieval process of Raqueno (2003) was revisited by Gerace (2010) in order to estimate how improvements in the upcoming Landsat 8 imaging satellite would impact its potential for water quality monitoring. This work included both simulating the end-to-end retrieval process using modeled water spectra, as well as AVIRIS imagery resampled to the Landsat 8 RSRs. It should be noted that like the LUT of Raqueno (2003), Gerace (2010) only use one set of IOPs in generating their LUTs which is determined to be a source of error when applied to spectra whose IOPs deviate from these values. The results of both the modeled and real tests indicated a notable increase in performance of the upcoming Landsat system, mostly due to the decrease in system noise and the change from 8 bit to 12 bit quantization. Furthermore, in testing multiple potential atmospheric compensations, the best performing compensation approach for the retrieval process is determined to be the MoB-ELM, as used by Raqueno (2003). In a portion of the work by Pahlevan (2012), this same process was applied to Hyperion imagery, again resampled to the Landsat 8 spectral bands, and agreed with the improvement indicated by Gerace (2010).

Actual application of Hydrolight LUTs to Landsat 8 imagery would be performed by Concha (2015). Three Landsat 8 images were used, collected through 2013 to 2015 over the Rochester Embyment study area used by Raqueno (2003) and compensated using the MoB-ELM. The retrieval process of Concha (2015) deviates from that of Raqueno (2003) as it is the first to implement more than one set of IOPs into the retrieval process. The set
of IOPs used to model spectra were determined based on component concentrations, where higher concentrations used one set of IOPs and lower concentrations had another. The retrieval accuracy determined by this process was considered to be satisfactory with RMSEs for Chl, TSS, and CDOM being $15.51 \text{ mg/m}^3$, $4.98 \text{ g/m}^3$, and $0.08 \text{ m}^{-1}$, respectively.
Chapter 4

Methods

This chapter is focused on the methods used to accomplish the objectives of this work. The first section discusses how imagery is collected and processed, both for satellites and UAS. This is followed by sections describing the reference data that is collected in the field, how it is analyzed in the laboratory, and the collections and datasets used in this work. The methods used to remove atmospheric effects are then described, as well as the model based retrieval method used in analysis. Finally, this chapter covers the methods used to simulate water scenes for determining improvements in spectral sampling for future imaging systems.

4.1 Imagery Acquisition and Processing

Imagery of the subject of interest must first be acquired and processed before any analysis is carried out. The acquisition and processing of this imagery can be simple or complex depending on the system type. Satellite imagery is generally shared between many users, and therefore most of the processing is completed before its distribution. Preparation of UAS imagery is often a more involved, as it is the task of the user to take the imagery through all processing steps. This is further complicated due to the numerous sensors with different properties that can be implemented on a UAS. The acquisition and processing of both systems types are considered here.
4.1.1 UAS Imagery

This section focuses on practices used to collect and process drone imagery. The sensors used in these collects were the Tetracam Micro-MCA Snap sensor as well as the Headwall Nano-Hyperspec.

4.1.1.1 UAS Imagery Collection

In order to collect high resolution imagery of waterbodies, both a Tetracam Micro-MCA Snap and a Headwall Nano-Hyperspec were mounted on a UAS. Software was used in conjunction with the on-board GPS system of the UAS to create a flight path that carved out a swath of area over the waterbody, as well as control the altitude. Flight plans were designed so that overlap occurred between each flight line, therefore covering the entire spatial extent. The altitude and spatial coverage of the imagery varied between collections.

For dark targets, such as most water bodies, sensor integration time, the amount of time the camera captures light, is important as the amount of signal produced by water is low. This may make longer exposure times seem desirable, however they may cause saturation, where the sensor receives more than the maximum amount of signal that can be recorded, making information unrecoverable. High signal targets, such as calibration panels or areas of glint, are the most susceptible to saturation. Further, longer exposure times can lead to motion degradation from the continuous UAS movement (Fiete, 2010). Conversely, if exposure is too low, system noise, such as dark current or read noise, may dominate low signal water pixels. Ideally a setting between too high or low exposure should be used, although midflight changes in illumination can cause issues for fixed exposure times. For this reason a variable exposure was used (i.e. auto-exposure) and each band was normalized by its exposure time before processing.

4.1.1.2 UAS Imagery Processing

Imagery from the Tetracam sensor required processing from raw format. The imagery data was transferred to a computer in a proprietary .RWS format where of the 6 bands records its own RWS file. This is a format that stores imagery data as 16 bit values,
despite the Tetracam having a quantization of 10 bits. The imagery data starts at byte 10 (indexed from 0). The RWS format differs from a RAW format in that it does not rearrange the image columns from the order they were read out from the sensor. This reduces the Input/Output time of collecting imagery while in-flight allowing sub-second capture of imagery. The proper remapping of the columns in the image can be seen in Table 4.1.

<table>
<thead>
<tr>
<th>Original Column</th>
<th>New Column</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>15</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 4.1: Column remapping required to assemble RWS format data.

Once all of the Tetracam band images are reordered out of RWS format, they must be registered. In order to do this the imagery must be offset, rotated, and scaled. This is because of the sensor’s multiple camera setup, each camera must be aligned with the other. Tetracam alignments change between systems, and each individual system is provided a set of specifications in a data file with the extension ”.mca”. The .mca file inputs can be seen in Table 4.2. As each camera is slightly offset, they capture data where others do not. Any pixel not captured by all cameras is trimmed from the imagery. Each image is
then stacked into an array in wavelength order. Due to the Nano-Hyperspec’s push-broom architecture, such an involved registration process is not required.

<table>
<thead>
<tr>
<th>X offset</th>
<th>Y offset</th>
<th>Rotation</th>
<th>Scaling</th>
<th>Center Wave Length-Bandwidth</th>
<th>Vignette</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>800-10</td>
<td>0</td>
</tr>
<tr>
<td>-5</td>
<td>-2</td>
<td>0.0109</td>
<td>1.00634</td>
<td>490-10</td>
<td>0</td>
</tr>
<tr>
<td>-3</td>
<td>-4</td>
<td>0.0419</td>
<td>1.00271</td>
<td>550-10</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0.3124</td>
<td>0.99463</td>
<td>680-10</td>
<td>0</td>
</tr>
<tr>
<td>-2</td>
<td>1</td>
<td>0.1191</td>
<td>1.00180</td>
<td>720-10</td>
<td>0</td>
</tr>
<tr>
<td>-1</td>
<td>5</td>
<td>0.1811</td>
<td>0.99820</td>
<td>900-20</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.2: Values stored in the Tetracam .mca file for registering imagery.

At this point it is possible to mosaic imagery into one total swath. Mosaicing is generally completed by determining features shared between adjacent images and using these features to align the imagery. This method is difficult to perform over water due to the fact that the waterbody surface is constantly changing with time. GPS data may be used to create an image swath, placing each image in their relative locations, although the overlap may not fit visibly due to the changes in water surface. As the sensor viewing angle can change midflight it would be important to have an Inertial Measurement Unit (IMU) to record changes in viewing angle (roll, pitch, yaw) so that the true location coordinates can be determined.

The effect of lens fall-off for the Tetracam was characterized by taking a full frame image of the output of an integrating sphere. The ratio of signal received from the center pixel to the all other pixels was determined and then used to correct each pixel. The fall-off of the imagery was corrected using the equation:

\[
I_{\text{corr}} = I_{(x,y)} \frac{C_{\text{center}}}{C_{(x,y)}}
\]  

(4.1)

Here, \( I_{(x,y)} \) is the Digital Count (DC) of the pixel, \( C_{\text{center}} \) is the DC of the center pixel in the calibration image, \( C_{(x,y)} \) is the DC of the pixel in the calibration, and \( I_{\text{corr}} \) is the corrected DC.
As previously stated, auto-exposure was used during imagery collections. Removing the effects of auto-exposures is required in processing of the imagery. Uncorrected auto-exposure leads to issues such as bright targets in one image having lower DCs than dark targets in another (e.g. land appearing darker than water). The DCs of each band image were normalized by their exposure time stored in the metadata. This was sufficient to set the relative DCs of each image to appropriate levels.

After the collection with the Headwall Nano-Hyperspec, it was noted that the spectrum contained significant noise. In order to decrease this noise a spatial binning of the imagery was performed. Pixel values in an 8x8 pixel window were averaged and re-formatted into imagery. An 8x8 pixel window was chosen as it was determined to be the most efficient trade-off between noise reduction and spatial resolution. Similarly, all Tetracam retrievals were performed on the scene averaged spectrum.

4.1.2 Satellite Imagery

4.1.2.1 Satellite Imagery Acquisition

Although the effort required to design and implement a satellite imaging system is much greater than that of UAS, distribution of its imagery is made simple for its wide variety of users. Landsat imagery is distributed through the USGS’ Earth Explorer site ([earthexplorer.usgs.gov](http://earthexplorer.usgs.gov)). In order to find imagery of a select region, one needs to access the site, enter the coordinates of a location of interest and a date range of overpass. The user can then select from a variety of archives, including Landsat. Similarly, Sentinel-2 imagery can be accessed from the ESA’s Scihub website ([scihub.copernicus.eu](http://scihub.copernicus.eu)). Imagery from other ESA satellite missions can be acquired from this source as well. Another example, specific to Ocean Color imagery, is acquisition from NASA’s Ocean Color Web ([oceancolor.gsfc.nasa.gov](http://oceancolor.gsfc.nasa.gov)). The data browser can be accessed from the main website. This data hub allows acquisition of imagery from numerous sensors (MODIS, SeaWiFS, etc.) from a date and location of the user’s choosing.
4.1.2.2 Satellite Imagery Processing

Similar to the acquisition of satellite imagery, generally minimal processing is required once it has been acquired. In most instances, the distributing agency will have performed processing of data from raw formats, image registration, and other similar processes. Imagery from Landsat 8 is available as scaled DCs, which can be simply radiometrically calibrated using Equation 4.2.

\[ L_{\text{cal}} = L_{\text{gain}} \times DC + L_{\text{bias}} \]  

(4.2)

Where \( L_{\text{gain}} \) and \( L_{\text{bias}} \) are the radiance gain and bias provided in the metadata, and \( L_{\text{cal}} \) is the calibrated radiance value.

Systems can offer imagery that has been further processed. The best example of this is the imagery from the NASA Ocean Color Web database. NASA provides imagery that has been processed using their standardized processing techniques. This includes the radiometrically calibrated imagery, but can also include imagery compensated for atmospheric effects or maps of water parameters (e.g. Chlorophyll-a Concentration, Sea Surface Temperature) generated using standard algorithms. It should be noted that these products can be generated for oceanic or inland/coastal waters, but often the algorithms used are validated only for the former. These same products can be generated using the SeaDAS software provided by NASA. This may be preferable as it allows users to have more control over the processing.

4.2 Field Collections

As this work focuses on retrieving physical properties of waterbodies, reference data is necessary to measure accuracy of the retrieval. Collection of reference data should preferably occur as close to the time of imagery collection as possible to minimize the effect of temporal variability. Error is implicit in any measurement, therefore reference data can never be considered completely truthful, however methodology can be put in place to minimize error in the analysis. This section focuses on the first step of acquiring reference data, collection of samples in the field.
4.2.1 Water Sample Collection

Water samples must be collected in the field so that they can be analyzed in the laboratory to determine properties of interest. Although collecting a water sample can be a trivial task, there are a few steps that can be implemented to better preserve the sample until analysis. First, when collecting a water sample the vessel must be anchored as much as possible. This is to avoid drift and to guarantee that imagery pixel containing the sampling location can be determined. Next the container for the sample should be cleaned. This is done by filling and emptying the container with sample water 3 times. The procedure for this requires that the container lid be capped until submerged a few inches beneath the surface, uncapped so it can fill, and then recapped without surfacing. This prevents collection of any surface floating objects. Finally the water sample should be stored on ice to prevent degradation. After the collection is over the water sample should be stored at $4^\circ - 8^\circ$ Celsius.

4.2.2 Water Spectra Collection

A measurement of the water’s spectrum should be collected at each water sampling site. This spectrum can be used to implement or verify the atmospheric compensation applied in analysis, or to verify accurate modeling of spectra in Hydrolight. This work used the SVC HR-1024i spectroradiometer for measurement. Spectral measurements were taken above the surface of the water and therefore were susceptible to sun and sky glint. The methods of Mobley (1999) were used to minimize their effects, but other methods exist in the literature (e.g. Kutser et al. (2013)). This required sky and water spectra, collected at angles $+40^\circ$ and $-40^\circ$ from the horizontal both at an azimuthal angle of $135^\circ$ from the sun. These spectra were captured in triplicate and averaged. The same white Spectralon reference measurement was used for all 6 spectra. The sky glint is removed from the water spectra through the equation:

$$R_{rs} = \frac{L_w - \rho L_{sky}}{\pi L_g}$$

Where $L_w$ is the water radiance, $L_{sky}$ is the sky radiance, $L_g$ is the reference radiance, $\rho$ is the proportionality factor given by Mobley (1999), and $R$ is the reflectance of the
reference (1.0 for Spectralon). The value of the proportionality factor is about 0.028 for the conditions of a clear sky and less than 5 m/s wind speeds. This output units of this correction is Remote Sensing Reflectance (Rrs) as defined in Section 3.1.1.

4.3 Laboratory Analysis

This section describes how the two properties of interest, absorption and concentration, are measured for each component of the water.

4.3.1 Absorption Measurements

In order to determine the IOPs of water, absorption measurements are made in the laboratory. Ideally scattering measurements would be collected as well, however these measurements are difficult to collect and require specialized equipment. These values are taken from literature instead. This section describes the protocol to obtain the absorption of each individual component. These protocol are mainly based on the Ocean Optics Protocols for Satellite Ocean Color Sensors by Mitchell et al. (2002).

4.3.1.1 CDOM Absorption

Absorption of CDOM is determined through the methods given in Mitchell et al. (2002). Water samples are first filtered through a 0.2 \( \mu \text{m} \) filter membrane which has been pre-washed using purified water. It should be noted that other works may use a 0.4 \( \mu \text{m} \) membrane for this process. The filtered water is allowed to sit in a graduated cylinder for 4-5 hours to reach room temperature, and allow the settling of any filter particulate that may not have been removed by the pre-wash. The water sample is then measured in a dual-beam spectrophotometer using a 10 cm cuvette along with the same size reference cell filled with purified water. The output absorbance spectrum output is converted to absorption (units \( \text{m}^{-1} \)) using equation 4.4.

\[
a_{\text{cdom}} = \ln(10) \frac{OD_{\text{cdom}}}{l}
\]  

(4.4)
Here $OD_{\text{cdom}}$ is the measured absorbance, $ln$ is the natural log, and $l$ is the cuvette path length. It is also possible to remove any bias in the absorbance curve by subtracting any absorbance in the near infrared as it is assumed to be zero. It should be noted that no concentration measurements are made for CDOM, as the retrieval process is focused on determining the absorption value at 440 nm.

4.3.1.2 TSS and Chl Absorption

Measurements of TSS and Chl absorption spectra are determined by the methods given in Mitchell et al. (2002). A recorded volume of a water sample is filtered through an 0.7 $\mu$m GF/F filter, diameter 2.5 cm. The absorption of the filter is then measured using a dual-beam spectrophotometer with a filter wetted with purified water as a reference. It should be noted that this spectrum contains both the absorption of TSS and Chl. The filter is then soaked in warmed methanol for 10 minutes to depigment chlorophyll, after which the filter is measured again. This second absorbance measurement is considered to be a measurement of TSS only.

As both pigments being measured are assumed to have zero absorbance in the near-infrared, any bias in that region is subtracted off for both spectra. The multiple scattering effects in both spectra of glass fiber filters are corrected using equation 4.5, taken from Cleveland and Weidemann (1993).

$$OD_{\text{corr}} = 0.379 \times OD_{\text{org}} + 0.523 \times OD_{\text{org}}^2$$

(4.5)

Here $OD_{\text{corr}}$ is the corrected absorbance and $OD_{\text{org}}$ is the original absorbance. The TSS only spectrum is subtracted from the first spectrum measured, leaving only the absorbance from Chl.

As with the CDOM measurements, these absorbance spectra need to be converted to absorption spectra. As the absorbance spectra of both components are now separated, they are converted absorption through equation 4.6.

$$a_{\text{TSS/Chl}} = ln(10) \frac{OD_{\text{TSS/Chl}}}{V/A}$$

(4.6)
Where $a_{TSS/Chl}$ is the absorption of TSS or Chl, $OD_{TSS/Chl}$ is the absorbance of the component, $V$ is the volume of water measured, and $A$ is the area of the filter. Both absorption spectra are divided by their respective component concentrations to determine the mass specific absorption.

4.3.1.3 Phycocyanin Absorption

The final pigment absorption that is measured is that of phycocyanin, extracted by the methods of Sarada et al. (1999). Water samples are filtered through glass fiber filters the same pore size, $0.7 \mu m$, diameter 4.7 cm. These filters are then stored in a 15 ml centrifuge tube with 9 ml of 10 mM sodium phosphate buffer, pH = 6.8. In order to lyse the cyanobacteria cells, the buffer is frozen and thawed 3 times at 0° and 8° Celsius respectively.

The thawed tubes are then centrifuged, and the supernatant is then collected and measured in the spectrophotometer using 1 cm cuvettes with the sodium phosphate buffer as a reference. The absorbance spectrum is measured and then converted to absorption using equation 4.6.

4.3.2 Concentration Measurements

This section describes how concentration is measured for each of the components. As noted in Section 4.3.1.1, no concentration measurements of CDOM are made, only measurements of absorption at 440 nm.

4.3.2.1 Chlorophyll-a Concentration

The process extracting chlorophyll-a concentration is taken from Concha (2015), based on Lorenzen (1967) and Ritchie (2007). A measured volume of water is filtered using a glass fiber filter with pore size, $0.7 \mu m$ and a 4.7 cm diameter as soon after collection as possible (to avoid degradation). The filters are stored at −80° C until full analysis is possible. Once ready for analysis, the filters are ground to lyses the cells and allow the release of Chl. The ground filters are transferred centrifuge tubes containing 6 ml of 90% acetone solution, with a further 3 ml of acetone that has been used to rinse out any remaining filter.
particulate in the grinding tube. Once all samples have been processed, the centrifuge tubes are refrigerated overnight (at 4° - 8° C).

After this resting period, the tubes are centrifuged for 20 minutes in order to force the filter particulate to the bottom. The supernatant is then collected and absorbance is measured in a spectrophotometer using a 1 cm cuvette with a reference cell of 90% acetone. The absorbance at 630, 647, 664, 665, 691, and 750 nm is recorded. Three drops of a 5% hydrochloric acid solution is then added to the cuvette and allowed to sit for 1 minute. The absorbance is measured again at the same wavelengths.

Once both sets of measurements are taken, the concentration is determined using two different calculations. The first, taken from Lorenzen (1967), is shown by equation 4.7.

\[
C_a = \frac{26.7 \times [(665_o - 750_o) - (665_a - 750_a)] \times v}{V \times l} \tag{4.7}
\]

In this equation \(C_a\) is the chlorophyll concentration, the o/a subscript represent measurements before or after acidification respectively, \(v\) is the extraction volume of acetone in ml, \(V\) is the amount of water filtered in liters, and \(l\) is the pathlength of the cuvette in cm.

The second calculation performed, equation 4.8, is from Ritchie (2007). It was determined empirically and uses 4 wavelengths to determine concentration.

\[
C_a = \frac{-0.3319 \times 630_o - 17.485 \times 647_o + 11.9442 \times 664_o - 1.41306 \times 691_o \times v}{V} \tag{4.8}
\]

All variables in this equation have the same definition as they do for equation 4.7, however \(V\) and \(v\) are entered with units of \(m^3\) and \(L\), respectively.

4.3.2.2 **TSS Concentration**

The protocol for measuring TSS concentration is taken from Concha (2015). The first step in this process is to measure the mass of the filters to be used. These are the same filter type as those used for the Chl concentration measurement. Water samples are then filtered and the volume of water filtered is recorded. These filters are then stored in a drying oven for at least 12 hours at 75° C. After the filters are completely dry, the dried
filters are re-measured. The concentration is then determined using equation 4.9.

\[ C_{TSS} = \frac{M_{after} - M_{before}}{V} \]  

(4.9)

In this equation, \( C_{TSS} \) is the concentration, \( M_{after/before} \) is the filter mass after/before filtering the sample, and \( V \) is the volume of water filtered.

### 4.3.2.3 Phycocyanin Concentration

Phycocyanin concentration is measured using fluorometric analysis based on the methods of Dr. Gregory Boyer at the State University of New York College of Environmental Science and Forestry (Pers. Comm.). Phycocyanin concentration is the only measurement that is made concurrently with absorption. It is also possible to measure concentrations using absorbance measurements through use of equation 4.10, taken from Bennett and Bogorad (1973).

\[ C_{pc} = \frac{A_{615} - 0.474 \times A_{652}}{5.34} \]  

(4.10)

Here \( A_{615/652} \) are the absorbance values at 615 and 652 nm respectively, and the concentration is output in mg/ml. This method only works for very high concentrations (> 100 \( \mu \)g/L) and therefore is often not useful for field samples.

A spectrofluorometer was calibrated via a serial dilution process using powdered phycocyanin (acquired from Sigma-Aldrich) mixed with sodium phosphate buffer (described in section 4.3.1.3). The initial concentration of the phycocyanin buffer mix was quantified using equation 4.10. The spectrofluorometer was set to have excitation and emission wavelengths of 577 nm and 660 nm with bandwidths of 20 nm and 40 nm, respectively. The response time was set to 6 seconds, and the integration time was set to 10 seconds. This calibration method output a linear regression relating relative fluorescence units to concentration. The same process as listed in section 4.3.1.3 was used for phycocyanin extraction from water samples. The supernatant was then measured in the spectrofluorometer, and the response was used to determine the concentration of the filtered sample. This filtered concentration was then corrected to determine the concentration in the field using equation...
4.11.

\[ C_{corr} = \frac{C_{meas} \cdot \upsilon}{V} \]  

In this equation \( C_{meas} \) is the concentration determined from calibration, \( \upsilon \) is the volume of extraction buffer used, \( V \) is the volume of water initially filtered, and \( C_{corr} \) is the corrected concentration.

### 4.4 Collection Images and Data Sets

#### 4.4.1 Landsat 8

Landsat 8 retrievals focused predominately on imagery over two lakes in the Finger Lakes region, Owasco and Honeoye. These lakes were focused on due to their consistent annual off-shore cyanobacteria blooms. Honeoye, the shallower of the two, tended to have higher bloom concentrations than Owasco. Two field collections were performed in Honeoye Lake, and three in Owasco Lake both near-concurrent with Landsat 8 overpasses. The two Landsat 8 images of Honeoye Lake were captured on September 27, 2016 (LC80160302016271LGN01) and August 29, 2017 (LC80160302017241LGN00). It should be noted that for the August 2017 image, the sky was mostly cloudy except a significant break around Honeoye lake. The three Landsat 8 images of Owasco Lake were captured on September 20, 2016 (LC80150302016264LGN01), September 27, 2016 (LC80160302016271LGN01), and September 23, 2017 (LC80150302017266LGN00). The images along with sampling areas are shown in Figure 4.1 for Honeoye Lake and Figure 4.2 for Owasco Lake. Multiple atmospheric compensation methods were performed on imagery, however the majority of analysis was performed using imagery compensated using the MoB-ELM (Section 4.5.2).

#### 4.4.2 Sentinel-2

As there was no overlap between the Landsat 8 collections and Sentinel-2 overpasses, concentration retrieval from Sentinel-2 focused on Lake Erie instead. Lake Erie was chosen as concentrations of Chl and PC were available based on weekly reporting by the National Oceanic and Atmospheric Administration (NOAA) Great Lakes Environmental Research
Laboratory (GLERL) (Vander Woude et al., 2019). These concentrations are solely focused on western Lake Erie. As no CDOM or TSS measurements were reported, this work does not perform any analysis for these components. Only Sentinel-2 imagery within one day of measurement was used in analysis resulting in four images captured on August 4, 2015 (S2A_MSIL1C_20150804T163126_N0204_R040_T17TLG_20150804T163120), October 2, 2017 (S2A_MSIL1C_20171002T162111_N0205_R040_T17TLG_20171002T162618), July 9, 2018 (S2A_MSIL2A_20180709T161901_N9999_R040_T17TLG_20190616T170610), and August 13, 2018 (S2B_MSIL2A_20180813T161859_N9999_R040_T17TLG_20190616T160615). Sentinel-2 imagery and the reported sampling sites are shown in Figure 4.3. All Sentinel-2 images were compensated for atmospheric effects using the Sen2Cor algorithm (Section 4.5.5).

Figure 4.1: Landsat 8 imagery of Honeoye Lake with marked sampling points (Red). A) Honeoye Lake on September 27, 2016 B) Honeoye Lake on August 29, 2017
Figure 4.2: Landsat 8 imagery of Owasco Lake with marked sampling points (Red). A) Owasco Lake on September 20, 2016 B) Owasco Lake on September 27, 2016 C) Owasco Lake on September 23, 2017

4.4.3 UAS

Three UAS collections were performed, two using the Tetracam system and one with the Nano-Hyperspec system. The first of the UAS collections focused on Long Pond in the Rochester Embayment in western New York on November 3, 2015. This was the first of the two Tetracam based collections. Imagery collection was performed to be near-concurrent with a Landsat 8 overpass (LC80170302015307LGN01). Two field samples were collected within 24 hours of both image collections. As there was no cyanobacteria bloom reported in Long Pond at this time, only concentrations of Chl, TSS, and CDOM were measured and therefore retrieved in analysis. Imagery from both the UAS and Landsat collections can be seen in Figure 5.13 and Figure 5.14, respectively. Atmospheric compensation for
both the UAS and Landsat imagery was performed using the MoB-ELM (Section 4.5.2). In the case of the UAS imagery, calibration panels were placed in-scene.

Two UAS collections were performed on the north end of Honeoye Lake, near Sandy Bottom Beach. The first, performed on September 12, 2016, used the Tetracam imaging system and the second, performed September 17, 2017, used the Nano-Hyperspec. Both collections occurred during on-going cyanobacteria blooms, and bloom conditions were greater for the 2016 collection. In both collections, three water samples were taken and analyzed for all four components. Sampling points were selected to be within distinct regions of the collected imagery. The MoB-ELM (Section 4.5.2) was used to correct the 2016 Tetracam imagery. The 2017 Nano-Hyperspec imagery was afflicted by noise which
often caused the MoB-ELM to over correct the pixels (i.e. negative Rrs). As such, the 2017 imagery was corrected using the regular ELM with in-scene calibration panels (Section 4.5.1).

4.5 Atmospheric Compensation

Atmospheric effects can be a major contributor to the amount of signal captured over dark targets, such as water. It is imperative that these effects be removed from the imagery so that analysis can take place, even for UAS despite their low flight altitudes. As atmospheric compensation is often imperfect, multiple compensation techniques were tested and implemented in this work. This section describes each of these compensation techniques.

4.5.1 Empirical Line Method

The simplest atmospheric compensation implemented for this work was the Empirical Line Method (ELM) (Smith and Milton, 1999). The ELM requires two calibration panels, one highly reflective and one with minimal reflection. The bright and dark panels are placed in the field to be imaged at the same flight altitude as the rest of the collection. The spectral reflectance of the panels is characterized in the laboratory or in the field. These measurements are used to create a linear relation between DCs in imagery and physical values of spectral reflectance at each wavelength or band of the sensor. The slope and offset of this wavelength dependent linear relation allows for the removal of atmospheric effects. This method assumes that the atmosphere is uniform over the entire collection area and the slope and offsets are applied to the entire image to convert to spectral reflectance.

The reflectance panels used for this study, specifically during the UAS collections, were large enough to fill a 1 m² area. One panel was painted matte black and had a spectral reflectance of 4-5% across the VNIR spectrum, and the other was gray with a spectral reflectance of 50-65% across the VNIR spectrum. These panels were chosen to bracket the dynamic range of the collection area. Characterization of these panels took place after collection in the laboratory. Imagery was collected under the same illumination as that of the panel overpass.
The spectral characterization of the two panels used in the collection was converted to Rrs by dividing their spectral reflection by $\pi$, as they were assumed to be approximately Lambertian. Using the transmission values for the spectral filters implemented for this study, as well as the CMOS camera response function, the panel Rrs values were sampled to each Tetracam spectral band using equation 4.12.

$$Rrs'_i = \frac{\int t_i \ast S \ast Rrs \, d\lambda}{\int t_i \ast S \, d\lambda}$$  \hspace{1cm} (4.12)

Where $t_i$ is the filter transmission of the ith band, $S$ is the responsivity of the sensor, and $Rrs$ is the measured panel Rrs. A linear relation of band DCs (x-axis) and Rrs values (y-axis) was then determined for each band, which was used to convert all pixels to Rrs. An example of this linear relation is shown in Figure 4.4.

![Figure 4.4](image-url)  

Figure 4.4: Example ELM linear relation based on imagery collected over Long Pond on November 3, 2015. The inset graph better shows the clustered dark panel points.
4.5.2 Model-Based Empirical Line Method

A variation of the ELM was suggested by Concha and Schott (2016), initially conceived and implemented by (Raqueno, 2003). This method is mainly focused on satellite systems as their spatial resolution makes it difficult to implement the traditional ELM. The Model-Based Empirical Line Method (MoB-ELM) uses pseudo-invariant features and models to make up the light and dark pixels within the scene.

The bright pixels used in the MoB-ELM are usually pseudo-invariant urban features. Non-urban features are masked by combining a vegetation and water mask. This mask is applied to the original Landsat 8 image and the Landsat 8 Surface Reflection Product, which gives both the Top of Atmosphere radiance and a reflectance value for those pixels. The reflectance pixels are assumed lambertian and divided by $\pi$ to convert to $R_{rs}$.

The dark pixel reflectance cannot be determined using the same method as the bright pixels as the Landsat 8 Surface Reflectance product is not calibrated to work over water. Instead, an Ecolight run is generated to represent the $R_{rs}$ of a region of interest within a waterbody. The inputs to Ecolight are determined by analysis of a collected water sample. The sampling location is selected as the region of interest. Once this "dark panel" data and the data for the "bright panel" are collected, they can be used to implement an MoB-ELM correction using the same equations for the ELM stated in Section 4.5.1.

4.5.3 SeaDAS MUMM

Another atmospheric compensation algorithm implemented on the imagery is the SeaDAS MUMM. The name MUMM is an acronym for the Management Unit of North Sea Mathematical Models, and it is based on the work of Ruddick et al. (2000). This atmospheric compensation scheme builds on the method developed by Gordon and Wang (1994). Gordon & Wang’s method assumes that the water does not reflect light of any wavelength greater than or equal to 700 nm, therefore all measured radiance is caused by the atmosphere. This is also known as the black pixel assumption. As the total signal in the near-infrared is assumed to be from aerosols, the ratio between two bands in this region are used to fit an aerosol model. This is done by searching a database of aerosol types, based on relative humidity, and then finding aerosol types whose ratios bracket the ratio.
from the imagery. The proportional difference between the bracketing ratios and the image ratio is assumed to be constant for all wavelengths. These values are used to interpolate the atmospheric contribution to the signal for each band. This can be used to remove the atmospheric effects from the imagery.

It should be noted that the black pixel assumptions fails over case 2 waters, the type of waters studied by this work. Thus, the method develop by Ruddick et al. (2000) does not make this assumption. Instead they make the assumption that the multiple-scattering aerosols and aerosol-Rayleigh reflectances are spatially homogeneous in the near-infrared and that the water-leaving reflectances normalized by the sun-sea atmospheric transmittance in the same region is spatially homogeneous. These assumptions are applied to determine the aerosols despite the water leaving radiance in the near-infrared, and then implement the compensation.

The MUMM correction is implemented in SeaDAS. Landsat 8 bands 4 and 5 are chosen for the calibration process. A calibration parameter of $\alpha = 8.7$ is used as suggested by Concha (2015).

4.5.4 Acolite

Acolite (atmospheric correction for OLI lite) is compensation method designed and described by Vanhellemont and Ruddick (2015) and later extended to Sentinel-2 by Vanhellemont and Ruddick (2016).

The first step in the compensation technique is to calculate the OLI band averaged extraterrestrial solar irradiance, pure-water absorption, Rayleigh thickness for a standard atmosphere, and ozone optical thickness for 300 DU of atmospheric ozone as they relate to each of Landsat 8’s spectral bands. These values allow correction of TOA radiance to TOA reflectance. Furthermore they are used along with the sun zenith angle and viewing zenith angle to determine the sun-sea and sea-sensor transmission.

The next step is to calculate the Rayleigh reflectance. The Rayleigh reflectance is estimated by the equations of Gordon et al. (1988). This equation takes into consideration the Rayleigh scattering phase function, the Fresnel reflectance, the sun and sensor azimuthal angles, and the Rayleigh optical thickness. Once this is completed, the effect the Rayleigh
reflectance can be removed from the overall signal, leaving only the aerosol reflectance and the water reflectance.

The final step in this process is to determine the aerosol correction. This is done using the same assumptions described in Section 4.5.3. Once these assumptions are made, the arguments determined can be used to solve for the water reflectance. This reflectance is divided by $\pi$ to turn it into $R_{rs}$.

Software to implement the Acolite atmospheric compensation method was developed by the Royal Belgian Institute of Natural Sciences, and is made available by them. It is capable of producing output in reflectance or $R_{rs}$. Images were processed using this software, and geoTIFFs of $R_{rs}$ were produced. The images were processed using the default settings, which used the shortwave infrared spectral bands. The gains used for this correction were the ”Franz Ave [Average]” gains, as calculated by Franz et al. (2015) and implemented by Concha (2015).

## 4.5.5 Sen2Cor

Sen2Cor is the primary atmospheric compensation method for Sentinel satellites, provided by the European Space Agency (ESA). Sen2Cor provides bottom-of-atmosphere reflectance, as well as maps of atmospheric properties (e.g. Aerosol Optical Thickness (AOT)) determined as part of the correction process. Due to the varying spatial resolutions of Sentinel-2 bands, output can be in resolutions of 60 m, 20 m, or 10 m. This work used corrections to 20 m resolution as a compromise of spatial and spectral resolution. Furthermore this work used version 2.8 of the Sen2Core code, distributed from the ESA website for Python. The following paragraphs describe the process of Sen2Cor as described by Main-Knorn et al. (2017).

The first step of the Sen2Cor process is performing a TOA scene classification, the results of which are used in the atmospheric compensation process to determine clear, cloudy, or water pixels. The AOT of the scene is determined by using the Dense Dark Vegetation Algorithm with a 550 nm band, specifically using pixels focused on dark vegetation, soil, or water bodies. If none of these are present, a constant AOT is chosen to based on the starting visibility set during configuration. The water vapor content is estimated using the
Atmospheric Pre-corrected Differential Absorption algorithm using bands 8a (864 nm) and 9 (945 nm) as reference and measurement bands, respectively.

Once the AOT and water vapor content is known, a correction LUT is generated using the libRadtran radiative transfer software package. The choice of stratospheric aerosols, ozone content, and other atmospheric properties are set based on the scene geographic location as well as climatology. The generated LUT is then used to invert TOA reflectance to bottom of atmosphere. In this work, the bottom of atmosphere reflectance is further divided by \( \pi \) to convert to \( \text{Rrs} \).

4.6 Model Based Spectral Look-Up-Table Matching Process

This section describes how the LUTs are generated and applied to imagery in order to retrieve concentrations. First, the generation of the LUTs will be discussed. Following that, the matching process is discussed. Finally, the methods of testing how variability affects the LUT are described.

4.6.1 LUT Generation

In this study LUTs are generated by repeatedly running Ecolight and varying the input file each time. The desired variability of component concentration, absorption spectra, scattering phase function, and scattering spectra are listed in separate files. Concentration files are listed with numeric values and spectra files contain names of specially formatted data files containing their respective values. The user can enter any number of inputs for these properties.

Once the user has set all property files, a bash script is implemented that creates an Ecolight input file for every possible combination of component properties. The number of input files created (and LUT size) can be determined by multiplying the number possible values for each property of each component. Each of these input files are entered into their own directories and submitted in turn to Ecolight. Ecolight then generates the water leaving radiance and Rrs spectra. Once it is confirmed that all spectra have been generated, they are assembled into the LUT. Also generated is a file containing a list of the property
inputs associated with each LUT entry. An example of an output LUT is shown in Figure 4.5.

![Figure 4.5: Example LUT generated using the process detailed in Section 4.6.1. This LUT contains about 22,000 entries.](image)

This method of LUT generation makes all combinations of the input variability given, including non-physical combinations. An example of this would be having a low chlorophyll content but a high phycocyanin content. In this work, alternate designs for the LUT that limit these outputs are also tested. These alternate forms are described in Section 4.6.3.2.

### 4.6.2 LUT Spectra Matching

Once the LUTs are generated and the imagery has been atmospherically compensated, the spectra matching algorithm can be implemented. This algorithm is the point in which the LUTs are used to form concentration maps from the imagery. The first step in the matching algorithm is to set the import path for the LUT, the LUT concentrations file, and the image. It should also be verified that the spectral band files that the LUT will be sampled to are the same as the system that captured the imagery. As the LUTs are
produced spectrally, this algorithm can be implemented with any system as long as its relative spectral response is known.

Once the settings listed are entered, the algorithm begins. The LUT is sampled to the bands of the system using equation 4.12, however satellite systems use their bands relative spectral response in lieu of transmission and responsivity. The RMSE between each individual pixel and every entry in the sampled LUT is determined using equation 4.13.

\[
RMSE = \sqrt{\frac{\sum (Rrs_{image} - Rrs_{LUT})^2}{N_{bands}}} \tag{4.13}
\]

Where RMSE is the root mean square error, \(Rrs_{image}\) is the Rrs of a single pixel from the imagery, \(Rrs_{LUT}\) is one LUT entry, and \(N_{bands}\) is the number of bands used. The concentrations of the LUT entry that produces the minimum RMSE are then chosen as the initial guess.

Once the initial guess concentrations are chosen an interpolation and minimization process are used to determine concentrations that fall between the discretely spaced values in the LUT. This requires that the non-concentration component properties be fixed. The absorption/scattering spectrum and scattering phase functions of the best fit spectrum are determined and a smaller LUT is built from the original using only entries with these specific IOPs. Once this smaller LUT is generated, the interpolation process begins.

The interpolation function is a linear interpolation based on the four component concentrations. The interpolation begins by first finding spectra with component concentrations above and below the initial guess concentrations of each component. During the initial run, as the minimization process repeats the interpolation multiple times, the spectra with the concentrations below the initial guess is set to be that of the initial guess. If there is no value above the initial guess in the LUT, the initial guess spectrum becomes the upper concentration spectrum and the spectrum with the second highest concentration is set as the lower spectrum.

Once the upper and lower concentration spectra are set, the location of the initial guess spectrum between these two bounds is determined for each component. This is done using
equation 4.14.

\[
m_{\text{comp}} = \frac{C_{\text{guess}} - C_{\text{lower}}}{C_{\text{upper}} - C_{\text{lower}}} \quad (4.14)
\]

Where \( m_{\text{comp}} \) is the location, \( C_{\text{guess}} \) is the test concentration, \( C_{\text{lower}} \) is the lower concentration bound, and \( C_{\text{upper}} \) is the upper concentration bound. Once this is done, 16 spectra from the abbreviated LUT are found, one for every combination of the 4 components at the higher or lower concentration. This can be considered to be spectrum 0000 through spectrum 1111, where each digit represents a different component and 0,1 indicate a high or low concentration. The number of spectra needed increases as \( 2^n \) where \( n \) is the number of components in the model. The contribution of each of these spectra is weighted by the guess concentration’s distance between them. An abbreviated example of this is shown in equation 4.15.

\[
S_{\text{interp}} = S_{0000} \ast (1 - m_{\text{chl}}) \ast (1 - m_{\text{sm}}) \ast (1 - m_{\text{cdom}}) \ast (1 - m_{\text{pc}}) \\
+ S_{0001} \ast (1 - m_{\text{chl}}) \ast (1 - m_{\text{sm}}) \ast (1 - m_{\text{cdom}}) \ast (m_{\text{pc}}) \\
+ \ldots \\
+ S_{1111} \ast (m_{\text{chl}}) \ast (m_{\text{sm}}) \ast (m_{\text{cdom}}) \ast (m_{\text{pc}}) 
\]

(4.15)

Where \( S \) represent the interpolated spectra and \( m_{\text{comp}} \) is the guess location determined by equation 4.14.

Once the interpolated spectrum is determined, the difference is found between it and the pixel spectrum. This difference is returned by the interpolation process. A least squares non-linear minimization process is then used to minimize this returned difference by augmenting the guess concentrations. The range of concentrations is bounded by that of the LUT, so component concentrations cannot be extrapolated past the minimum and maximum values used in the LUT generation. The concentrations determined to minimize the difference of this interpolation spectra and the pixel spectra are reported as the final results.
4.6.3 Design and Implementation of IOP Variability in LUTs

One key aspect of this work was the testing of how the design of an LUT, including the implementation of the natural variability of water’s optical properties, affects retrieval. This was approached in two ways, first through modeling studies looking at how variability in the LUT and waterbody affect retrieval, and through applying LUTs with different designs/structures to real imagery. The following sections discuss the methodology for each of these tests.

4.6.3.1 Modeled LUT Design Tests

Two modeling studies were performed to determine how LUT design and IOP variability affect retrieval. The first study focused the number of steps within component concentration ranges and the second focused on variability within IOPs and within the waterbody. For both tests random test spectra were generated using the methods described in Section 4.7.1. The concentration ranges used for both tests were 0.1 to 200 mg/m$^3$ for Chl and PC, 0.1 to 100 g/m$^3$ for TSS, and 0.1 to 4 m$^{-1}$ for CDOM. This range was chosen based on reported concentrations from various studies performed on cyanobacteria blooms (Simis et al., 2005; Randolph et al., 2008; Nguy-Robertson et al., 2013; Mishra et al., 2013). The mass-specific scattering for both Chl and TSS were taken from Bukata et al. (1995) and phycocyanin was assumed to be a scattering component with its mass-specific estimated through the method discussed in Section 3.3.2.5. The mass-specific absorption for Chl and TSS were taken from Nguy-Robertson et al. (2013), and Mishra et al. (2013) for phycocyanin. All mass-specific absorptions were chosen to represent the median of the reported range. CDOM absorption, normalized at 440 nm, was also taken from Nguy-Robertson et al. (2013) chosen again to represent the median of the reported range. The backscatter ratio used to described the Fournier Forand scattering phase function (Fournier and Forand, 1994) was chosen to be 1.8% based on measurement by Petzold (1972) for the concentration tests, and was allowed to vary for IOP variability tests. It should be noted that no noise, atmospheric effects, or band sampling were used for these tests (i.e. the spectra were matched at full resolution).

For tests focused on the number of concentrations in the LUT, LUTs with 3 through 14
equally sized steps throughout each component’s range of concentrations were generated and applied to the same test set. This test set consisted of 10,000 spectra for which concentrations were randomly chosen for generation. The retrieval error of applying each LUT was then determined, which is reported in Section 5.1.1. In addition to testing equally spaced concentrations, two tests were performed that biased concentration steps towards the upper and lower halves of the component concentration ranges. Two LUTs with 8 concentration steps where generated where the first was 'Bottom-Heavy' (i.e. 5 values in the first half of the range, 3 in the second) and the second was 'Top-Heavy' (i.e. 3 values in the first half of the range, 5 in the second). These LUTs were both applied to the same test set as the others, and the retrieval error was reported as well. It should be noted that these tests assume that the IOP set in the LUT perfectly matches that of the test spectra, that the IOPs are consistent with increasing concentration, and that there is no noise from an imaging system or atmospheric compensation. The effects of changing these assumptions remains an area for future study.

For tests focused on the variability of IOPs within the LUT and the scene, multiple sets of paired LUTs and test spectra were generated, each with a different number of IOPs. Each set of test spectra contained 10,000 entries. The IOPs used were chosen to span the range of expected values. In the case of scattering phase function, which in this work is defined by the ratio of backscattering to total scattering, an appropriate range was determined from literature to be 0.5% to 4.0% (Ahn et al., 1992; Twardowski et al., 2001; Aas et al., 2005; Loisel et al., 2007; Snyder et al., 2008). LUT and test spectra data sets were generated with 2 to 13 scattering phase functions varying in backscatter ratio across this range. As an example, a set that had 3 entries would have scattering phase functions with backscatter ratios of 0.5%, 2.0%, and 4.0% whereas one with only two entries would only have the 0.5% and 4.0% ratios. Each different LUT (containing varying numbers of IOP entries) was used to retrieve concentrations from each set of test spectra (containing spectra generated from randomly chosing IOP entries from the possible range), and the error for each retrieval was reported in Section 5.1.2.
4.6.3.2 Real Imagery LUT Design Tests

The effect of LUT design on retrievals from real imagery was tested as part of this work. LUT design is here defined as the selected combinations of inputs used to generate the spectra that populate the LUT. The LUT generation described in Section 4.6.1 represents what this work refers to as an unstructured LUT. In this case no combination of LUT inputs is prohibited. This is the generation method used by Raqueno (2003) and Gerace (2010). Two types of unstructured LUTs were generated for this work, those generated individually for each image, and a single unstructured LUT generated to be globally applicable to all imagery.

For the case of the individual unstructured LUTs, only one set of IOPs was used in the generation process. The absorption spectrum used was the average of all measured absorption spectra within the waterbody and scene to which it was applied. As not all scenes had measured mass-specific phycocyanin absorption, two spectra were used, one measured in Honeoye Lake and another representing the median measured value by Mishra et al. (2013). The mass-specific scattering spectra used was chosen based on the IOPs used in the MoB-ELM atmospheric compensation process, which varied between images. As it was indeterminate as to whether phycocyanin was appropriately considered a scattering component, two mass-specific scattering IOPs were used, one that used the estimated scattering from Section 3.3.2.5 and one that assumed no scattering. The backscatter ratio of the scattering phase function was similarly determined from the MoB-ELM correction spectrum, but allowed to range with one entry above and below the MoB-ELM spectrum value. The range of component concentrations are shown in Table 4.3, chosen based on measured concentrations waterbody components in the Finger Lakes and Lake Erie regions. With each of the number component concentrations, 2 phycocyanin mass-specific absorption and scattering spectra, 1 set of all other IOPs, and 3 scattering phase functions the total number of entries in each of the individualized Mean IOP LUTs was 19200.

For the case of the global unstructured LUT, the same phycocyanin mass-specific absorption spectra and concentration values were used as the individualized LUTs. It should be noted that at this point it was determined that phycocyanin was more accurately assumed scattering negligible, so only one mass-specific scattering was used. Three phase
Table 4.3: Component concentrations for Unstructured LUTs applied to imagery

<table>
<thead>
<tr>
<th>Chl mg/m³</th>
<th>TSS g/m³</th>
<th>CDOM 1/m</th>
<th>PC mg/m³</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>5.0</td>
<td>10.0</td>
<td>0.33</td>
<td>5</td>
</tr>
<tr>
<td>10.0</td>
<td>20.0</td>
<td>0.66</td>
<td>10.0</td>
</tr>
<tr>
<td>25.0</td>
<td>35.0</td>
<td>1.0</td>
<td>25.0</td>
</tr>
<tr>
<td>50.0</td>
<td>50.0</td>
<td>1.66</td>
<td>50.0</td>
</tr>
<tr>
<td>100.0</td>
<td></td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>150.0</td>
<td></td>
<td>150.0</td>
<td></td>
</tr>
<tr>
<td>225.0</td>
<td></td>
<td>215.0</td>
<td></td>
</tr>
</tbody>
</table>

scattering functions with backscatter ratios ranging over the values determined for all scenes were used (i.e. 1.0%, 1.4%, and 1.8%). The mass-specific scattering spectra for Chl and TSS were taken from (Ahn et al., 1992) and (Bukata et al., 1995), respectively, as they determined were most frequently used in performing the MoB-ELM compensation of all scenes. Three mass-specific absorption spectra were used for each Chl and TSS, while only two for CDOM. In each case, the chosen absorption spectra were selected to represent the entire range of absorption within the imagery while also limiting the computation time for generating the LUT. This LUT had 172,800 entries.

A series of structured LUTs, akin to that used by (Concha, 2015), was also generated for both the individual and global case. These structured LUTs were split into high and low concentrations sections according to the distributions shown in Table 4.4. Note that phycocyanin is not split up between the higher and lower portions of the table due to both it being an accessory pigment not necessarily found in all waterbodies and its reported relation to Chl reported in Section 3.3.2.6. For the individual case, IOPs were chosen similarly to that of the unstructured LUTs. The same two mass-specific absorption and scattering phycocyanin spectra were used in both portions of the structured LUTs. The same mass-specific scattering and scattering phase functions chosen in the individualized unstructured LUTs were used in both portions of the structured LUTs as well. The mass-specific absorption spectra for Chl and TSS were determined individually for each scene. The absorption spectra measured in each scene corresponding to the highest concentrations was used in
the high concentration portion of the LUT, while the lowest concentration sampling point absorption was used in the low concentration portion. Overall the individualized structured tables had 2880 entries in their lower portion and 4608 entries in their upper portion, a total size of 7488 entries.

<table>
<thead>
<tr>
<th>Chl $mg/m^3$</th>
<th>TSS $g/m^3$</th>
<th>CDOM $1/m$</th>
<th>PC $mg/m^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>5.0</td>
<td>10.0</td>
<td>0.33</td>
<td>5</td>
</tr>
<tr>
<td>10.0</td>
<td>0.66</td>
<td>10.0</td>
<td></td>
</tr>
</tbody>
</table>

Low Concentrations

| 25.0         | 25.0        |
| 50.0         |
| 100.0        |
| 150.0        |
| 215.0        |

| 51.0         | 11.0        | 0.67        | 0.1         |
| 100.0        | 20.0        | 1.00        | 5           |
| 150.0        | 35.0        | 1.66        | 10.0        |

High Concentrations

| 215.0        | 50.0        | 25.0        |
| 50.0         |
| 100.0        |
| 150.0        |
| 215.0        |

Table 4.4: Component concentrations for Structured LUTs applied to imagery

For the global structured LUT, IOPs were chosen that were similar to the global unstructured LUT and the concentrations stated for the individualized structured LUTs. This LUT used the same Phycocyanin mass-specific absorption spectra as in the Global Unstructured LUT, as well as the same mass-specific scattering spectra and scattering phase functions for all components in both portions of the LUT. In the higher portion of the LUT, four mass specific absorption spectra were chosen for Chl and TSS, while three were chosen for CDOM. These were also chosen to be representative of all the sampled points that fell within the upper portion of the LUT. The same number of spectra were chosen for the lower portion, except for the TSS which only had three entries due to the
lower amount of variability at this end of the range. The lower portion of this LUT had 51,840 entries while the higher portion had 110,592 entries, for a grand total of 162,432 entries.

The final LUT design tested was one that was based on noted relations between component concentrations and their IOPs, both in the literature and from measured values. This LUT is referred to as the Literature LUT. The Literature LUT is broken up into six stages, each dependent on increasing concentrations of Chl. The concentrations of TSS and CDOM in each stage of the LUT were determined based on the relation reported by Babin et al. (2003), with some adjustment based on the concentrations measured in the scene. The concentrations of phycocyanin in each stage were determined based on the reported ratios of PC/Chl in Section 3.3.2.5. The concentrations for each component in each stage are shown in Table 4.5. The same set of mass-specific scattering IOPs as well as phase functions used in the other Global LUTs were used in all stages of the Literature LUT.

In order to set the CDOM absorption spectra, all CDOM measurements were fit to the model shown in equation 3.25, and the slope factor of each of these fits was determined. The mean and standard deviation in slope factor were calculated. Three CDOM absorption spectra were selected from the entire set of measured values, based on the closest value to the mean, and the mean plus/minus the standard deviation. The three CDOM absorption spectra were used in all stages. As the change in TSS mass-specific absorption was shown to be fairly linear with concentration by Tilstone et al. (2012), only three mass-specific absorption spectra were used in all stages of the LUT as well. These were chosen based on range of TSS mass-specific absorption at 442 nm reported by Tilstone et al. (2012). Each stage of the LUT had two mass-specific Chl absorption spectra, excluding stages 5 and 6 which only had one each. These spectra bracketed the range of absorption values at 442 nm reported by Tilstone et al. (2012). Generally, as Chl concentration increased the peak value of the higher absorption spectra would decrease, mimicking the relation shown by Bricaud et al. (1995). All entries with a Chl Concentration greater than 25 mg/m$^3$ were only represented by the single lowest value absorption spectrum. The Literature LUT was used for retrievals from all Sentinel-2 imagery and UAS imagery.
Table 4.5: Component concentrations in each stage of the Literature based LUT

<table>
<thead>
<tr>
<th>Stage</th>
<th>Chl $mg/m^3$</th>
<th>TSS $g/m^3$</th>
<th>CDOM $1/m^3$</th>
<th>PC $mg/m^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.01</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.5</td>
<td>0.05</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>1.0</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.0</td>
<td>1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>2</td>
<td>0.4</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>3</td>
<td>0.7</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25.0</td>
</tr>
<tr>
<td>3</td>
<td>5.0</td>
<td>1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>7.5</td>
<td>4</td>
<td>0.4</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>10.0</td>
<td>8</td>
<td>0.7</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50.0</td>
</tr>
<tr>
<td>4</td>
<td>10.0</td>
<td>1</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>17.5</td>
<td>5</td>
<td>0.7</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>10.0</td>
<td>1</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100.0</td>
</tr>
<tr>
<td>5</td>
<td>25.0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>37.5</td>
<td>5</td>
<td>0.8</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>50.0</td>
<td>10.0</td>
<td>1.25</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>150.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>215.0</td>
</tr>
<tr>
<td>6</td>
<td>50.0</td>
<td>10.0</td>
<td>0.66</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>100.0</td>
<td>20.0</td>
<td>1.00</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>150.0</td>
<td>35.0</td>
<td>1.66</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td>225.0</td>
<td>50.0</td>
<td></td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>150.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>215.0</td>
</tr>
</tbody>
</table>
4.6.4 Potential Sources of Error in the LUT Retrieval Method

No retrieval process is immune to potential sources of error, and this includes the spectral LUT matching process. As such, this section outlines potential sources of error from all parts of retrieval. It should be noted that often these sources of error are linked, influencing other parts of the retrieval process and leading to erroneous results.

4.6.4.1 Atmospheric Compensation

As atmospheric compensation often relies on assumptions about the scene, such as the black pixel assumption when performing the atmospheric compensation by Gordon and Wang (1994), it is a potential source of error to the retrieval process. As such, these assumptions are not always perfectly valid, and lead to discrepancy between the retrieved and actual BOA signal. The spectral matching process determines waterbody concentrations based on the quality of the match between the modeled spectra, without atmospheric effects, and the BOA spectrum. Residual atmospheric effects in the BOA spectrum could lead to an improper matching and therefore error. In this work, multiple atmospheric compensations are tested to determine which is most appropriate for the LUT matching process.

4.6.4.2 Imaging System Specifications

Imaging systems used in the spectral LUT matching process also present a potential source of error, mainly through its spectral and radiometric resolution. In terms of spectral resolution, the spectral sampling of an imaging system limits how well resolved the waterbody spectrum is resolved, including all of its spectral features. This limits the amount of information that the matching process has about the waterbody, and therefore the quality of the match. Furthermore, spectral resolution can also limit the use of glint correction and atmospheric compensation algorithms. In terms of radiometric resolution, noise or coarse quantization of the signal can affect the spectrum and therefore cause error in the matching process, similar to atmospheric compensation. Future generations of imaging systems are expected to improve in both spectral and radiometric resolution, and this work also explores how changes in spectral resolution can lead to improved performance by the imaging
system. Due to the spatial and temporal variability in cyanobacteria blooms, the spatial and temporal resolution of a system plays a role in HAB monitoring as well, though not directly affecting the LUT matching process.

4.6.4.3 Insufficient and Unrepresentative LUT Population

Insufficient and unrepresentative population of spectra in the LUT is a major source of error for all spectral matching tasks. The error from this mainly arises from the spectral matching and interpolation processes. If the spectra contained within the LUT are unrepresentative of that which is in the scene, then a poor match is chosen by default. Furthermore, poorly representative spectra can exaggerate the effects of error from a system’s radiometric performance or atmospheric compensation. In the interpolation process, poorly representative IOPs and error from atmospheric compensation and noise can exacerbate retrieval error of components with weak influence on the spectrum. The interpolation process will often freely increase or decrease the concentrations of these components to force a better output match, especially if the component only affects one or two bands of the imaging system.

Insufficient spectra within an LUT can lead to error in the spectral matching process, for the reasons mentioned above, but also in the interpolation process. If the interpolation method does not have the information to adequately describe the change in Rrs with change in concentration, this portion of the spectrum matching cannot be properly performed. It should also be noted that this method does not allow extrapolation to concentrations outside those modeled in the LUT, which can also limit performance if the LUT is not generated correctly.

4.7 Improved System Modeling

This section discusses the simulation work performed to determine potential improvements in spectral resolution for both future Landsat systems, and for a system that can be implemented on a six band UAS imager (e.g. Tetracam).
4.7.1 Scene and Retrieval LUT Generation

4.7.1.1 Test Pixel and Look-Up-Table Generation

In this work, both the randomly generated test pixels as well as LUTs were generated using the EcoLight radiative transfer model. It should be noted that in this and subsequent system improvement tests, LUTs and test spectra of water leaving radiance were used instead of spectra of 

\[ \text{Rrs}. \]

EcoLight uses the IOPs of each of the optically active components in the water to output a spectral water leaving radiance between 400 nm and 900 nm at 5 nm increments. This LUT used the freshwater absorption spectrum of Pope and Fry (1997) and the freshwater scattering spectrum of Morel (1974), both of which are built into EcoLight. The mass-specific scattering spectrum for Chl and for TSS were taken from Bukata et al. (1995) while the PC mass-specific scattering spectrum was estimated using the methods described in Section 3.3.2.5. Although a single scattering spectrum was used, two Fournier-Forand phase scattering functions (Fournier and Forand, 1994), whose shape depends on the ratio of backscattering to total scattering, were used to describe the angular dependence of scattering for the non-water components. The ratio values chosen were 1.8% and 2.0%, measured in harbor waters by Petzold (1972) and the mean ratio reported in the Oslo Fjord by Aas et al. (2005), respectively. Each component is represented by three values of specific absorption, or absorption in the case of CDOM. This was to represent the variability of specific absorption spectra reported in the literature (Fujiki and Taguchi, 2002; Stramski et al., 2007; Vodacek et al., 1997; Mishra et al., 2013). Two PC absorption spectra were taken from samples of a 2017 cyanobacteria bloom in Honeoye lake, located in the Finger Lakes region of New York state. PC was extracted using the freeze-thaw method of Sarada et al. (1999). The third PC absorption spectrum with a value of 0.005 \( m^2/mg \) at 620 nm was taken from measurements by Mishra et al. (2013) of Mississippi ponds. The three absorption spectra for Chl, TSS, and CDOM were obtained from the measurements of Nguy-Robertson et al. (2013) from lakes and reservoirs in Indiana bracketing the range of variability they reported. The set of random test pixels and the structured retrieval LUT were built using these IOP spectra. EcoLight uses RADTRAN to determine the scene irradiance, thus requiring date (September 16), time (16 GMT), and location (Lat/Long 42.83 N/77.7 W). These values corresponded to the cyanobacteria bloom season and time...
of day for a Landsat 8 overpass corresponding to western New York.

Concentrations for each of the test pixels were chosen randomly from preset ranges. 1000 test pixels were generated in all. These ranges where 10 to 150 $mg/m^3$ for Chl, 25 to 120 $mg/m^3$ for PC, 3 to 20 $g/m^3$ for TSS, and 0.1 to 2.0 $m^{-1}$ for CDOM, chosen based on reported bloom conditions (Randolph et al., 2008; Mishra et al., 2013; Yacobi et al., 2015). The water column is assumed to be optically thick with no bottom effects. One of the three absorption spectra and one the of two scattering phase functions were randomly chosen for each run. The same scattering, absorption, and phase scattering data were used for generation of the LUT in a structured fashion. Unlike the random concentration levels for the test pixels, a structured set of predefined concentration levels for each component were used in the LUT. A full list of concentration levels is shown in Table 4.6. The highest component concentration levels in the LUT were chosen to be greater than those for the test pixels. Based on the concentrations listed in Table 4.6, three possible specific absorption spectra for each of the four absorbing components, a single specific scattering spectrum for each component, and two scattering phase functions, the LUT contains 48600 entries.

<table>
<thead>
<tr>
<th>Chl $mg/m^3$</th>
<th>TSS $g/m^3$</th>
<th>CDOM $m^{-1}$</th>
<th>PC $mg/m^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0</td>
<td>3.0</td>
<td>0.1</td>
<td>25</td>
</tr>
<tr>
<td>50.0</td>
<td>10.0</td>
<td>1.0</td>
<td>50</td>
</tr>
<tr>
<td>75.0</td>
<td>17.0</td>
<td>2.5</td>
<td>75</td>
</tr>
<tr>
<td>125.0</td>
<td>25.0</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>175.0</td>
<td></td>
<td>140</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.6: Component concentrations for future Landsat modeling study

4.7.1.2 Addition of Atmospheric Effects

In order to make the scene more realistic, atmospheric effects are added to each of the generated water spectra. Atmospheric transmission and upwelled radiance are modeled using MODTRAN. The inputs used are for a nadir looking sensor at sun synchronous orbit, rural aerosols (23 km visibility), and as mid-latitude summer. The MODTRAN output transmission and solar scattering can be scene in Figure 4.6. This output is combined with
the water spectra to form the sensor reaching radiance using equation 4.16.

\[
L_{\text{sens}} = L_{\text{water}} \ast t_{\text{atmo}} + L_{\text{upwell}}
\] (4.16)

Where \(L_{\text{water}}\) is the water leaving radiance, \(L_{\text{upwell}}\) is the upwelled radiance, \(t_{\text{atmo}}\) is the atmospheric transmission, and \(L_{\text{sens}}\) is the sensor reaching radiance.

![Figure 4.6: MODTRAN output upwelled radiance and atmospheric transmissions for a mid-latitude summer with rural aerosols. The upwelled radiance has been peak normalized to fit on the graph with transmission.](image)

After this sensor reaching radiance is modeled through the sensor, as described in Section 4.7.2, it must then be compensated for atmospheric effects. The compensation is done through use of equation 4.17.

\[
L_{\text{ret}} = \frac{L_{\text{sensor}} - L'_{\text{upwell}}}{t'_{\text{atmo}}}
\] (4.17)

Where \(L_{\text{ret}}\) is the retrieved radiance, \(L_{\text{sensor}}\) is the radiance recorded by the sensor, and \(L'_{\text{upwell}}/t'_{\text{atmo}}\) are the band sampled atmospheric upwelled radiance and transmission. It
should be noted that this compensation assumes perfect knowledge of the atmosphere, however it does not achieve perfect compensation as the effects of the sensor retain noise artifacts (Section 4.7.2.2). The compensated test pixels were then entered in the spectral LUT matching retrieval process.

4.7.2 Sensor Modeling

Sensor modeling takes place in two steps, first the sampling to the spectral coverage of the sensor, and then adding the noise effects present in the sensor. The following subsections describe how this is accomplished in the model.

4.7.2.1 Spectral Coverage

Spectral coverage was varied in the modeled sensors in order to understand its effect on retrieval. As this modeling work was mainly focused on determining improvements for future Landsat missions, usually this meant adding bands to the existing Landsat 8 spectral coverage. Although it is possible to use existing relative spectral responses from bands on existing imaging systems, new band responses where chosen to be created. This was so a band response could be created with any center and bandwidth. Bands were created using equation 4.18, a function type known as a Super-Gaussian.

\[
RSR_{modeled} = e^{-\left(\frac{\lambda - \lambda_{center}}{w}\right)^{10}}
\]  

(4.18)

Where \( RSR_{modeled} \) is the modeled relative spectral response, \( \lambda \) is the wavelength, \( \lambda_{center} \) is the band center wavelength, and \( w \) is the approximate Half-Width Half-Max (HWHM), which is half of the Full-Width Half-Max (FWHM). This shape was chosen because it has similar features to the other Landsat 8 spectral responses. It has a high spectral response at the center wavelength which falls off quickly at the edges. An example of this added band can be seen in Figure 4.7.

The testing of additional spectral bands mainly focuses on gaps in Landsat’s coverage, specifically in the yellow (\( \approx 620 \) nm) and red edge (\( \approx 700 \) nm) regions, as well as the addition of a band in those two regions combined. Band centers tested range from 590 nm to 630 nm for the yellow and 675 nm to 725 nm for the red edge both in 1 nm increments.
Figure 4.7: Landsat 8 relative spectral response with an added modeled spectra band. The modeled band is shown in pink. It is a Super-Gaussian function centered around 710 nm.

The approximate HWHMs of all bands ranged from 15 to 30 nm wide (FWHMs 30 nm to 60 nm) which are similar to the existing Landsat 8 bands. For the simultaneous test, band centers range from 590 nm to 630 nm for the yellow and 680 nm to 725 nm both in 5 nm increments. The HWHMs are 15 nm, 18 nm, 20 nm, 25 nm, and 30 nm for both bands.

For a future imaging spectrometer, continuous coverage systems were tested with 5 nm FWHM bands as well as band FWHMs between 10 nm to 100 nm at 10 nm increments in the range of 400 to 900 nm. In each case, the center of the shortest wavelength band is 400 nm. The approximate HWHM of the bands were chosen to be the point where an adjacent bands approximate HWHM overlaps (e.g. a system with 20 nm FWHM would have bands with an approximate HWHM of 10 nm). In a second test of band center locations, spectrometers with 20 nm, 40 nm, 60 nm, and 80 nm FWHM were incrementally shifted to determine how changing placement would affect retrieval. For each system, band centers were shifted by 5 nm increments until the wavelength repeated. For example, a system that had 20 nm spacing would be tested with the location of its shortest wavelength band center at
400, 395, 390, 385 nm (shifts of 0, 5, 10, and 15 nm). If the band centers where shifted another 5 nm it would return the sampling to that of the original system. The 20 nm system was also used to test how well a spectrometer could be resampled to the historic Landsat product for data continuity. This was done by modeling a 20 nm spectrometer, resampling its retrieved radiances to Landsat 8 RSRs, and using the resampled spectra in the constituent retrieval process. In this data continuity test, the band center locations were also shifted by 5, 10, and 15 nm.

4.7.2.2 System Noise and Quantization

System noise for the existing Landsat 8 bands was implemented in the modeling chain using the noise model from Morfitt et al. (2015). As shown in Equation 4.19, this model determines the total noise radiance ($N$), with units $W/(m^2sr\mu m)$, using the quantization noise coefficient ($q$), the signal-independent noise coefficient ($a$), the signal dependent noise coefficient ($b$), the scene radiance ($S$), and a factor of 0.8 to account for error reduction from the cubic convolution resampling of the Level-1 Product. The values for $a$, $b$, and $q$ are band specific and were generated using on-orbit Landsat 8 noise measurements. This noise is added to the signal of each band as the standard deviation of a zero mean normally distributed value. The noise radiance per signal radiance is shown in Figure 4.8.

$$N = \sqrt{0.8(b \ast S + a)^2 + q^2}$$ (4.19)

To determine the most effective spectral sampling and spectral resolution, a process was needed to estimate the effect of band position and HWHM on the noise of the modeled spectral bands for both the multispectral and imaging spectrometer cases. To model the noise of a given band center we interpolate from the Landsat 8 bands and then adjust based on the modeled band HWHM. For example, for a modeled band at 620 nm with a HWHM of 20 nm an interpolation is performed of the noise and HWHM of Landsat bands 3 (center 561 nm, HWHM 28.5 nm) and 4 (center 654 nm, HWHM 18.5 nm) at 620 nm, producing the noise for a band with a HWHM of 22.1 nm. The noise of the modeled band is then multiplied by the ratio of the interpolated width to its desired width, thus narrower bands have more noise and wider bands have less. With this process the signal
dependent and independent noise are implicitly included. The sampled radiance spectra are then entered in the component retrieval process.

![Graph](image)

**Figure 4.8:** The output noise radiance versus the input signal radiance of the first 5 Landsat 8 bands based on the model by Morfitt et al. (2015).

### 4.7.3 Original Multi-Spectral Setup Determination

The scene and sensor modeling technique discussed in Sections 4.7.1 and 4.7.2 were used to determine a best performing set of six bands that could be implemented on the Tetracam system. In comparison to the future Landsat study, there is an infinitely large number of spectral band centers and widths for testing. In order to narrow down the potential testing space, band properties were limited to only those available through the Andover Corporation catalogue [www.andovercorp.com/products/bandpass-filters/standard/](http://www.andovercorp.com/products/bandpass-filters/standard/). This included band centers that ranged from 400 nm to 900 nm, and band FWHMs ranging...
from 1 nm to 100 nm.

The best band selection process began by randomly selecting six all of the potential bands. The selection process was based on a uniform random distribution. Once the six bands where chosen, the bands would be used in the same modeling process used for testing potential future Landsat systems, and the retrieval NRMSE (Section 4.8) would be determined. After this process one of the six bands, chosen at random, would be replaced by a new band. The band chosen to be replaced and the new band were both selected based on a uniform distribution. The end-to-end retrieval process would then be performed with this new band. The retrieval NRMSE of the system with the new band was then compared to that of the previous system and the system with the lowest NRMSE would be recorded as the best performing system. If the best performing system was the previous system, a counter would be incremented and the single band replacement would occur again using the previous system’s bands. If the best performing system was the new band system, the counter would be reset and the single band replacement would occur using this new system. If no new band outperforms the previous system after 350 tests (i.e. 350 band replacements), the test ends and the band properties of the best reporting system are recorded. The value of 350 iterations was chosen to allow robust results that minimized computation time. This overall process was repeated 20 times to further avoid the potential of a returning a band combination that was trapped in a local error minimum.

4.8 Error Reporting

Error reporting in this work is predominately done using two metrics, Root-Mean-Square-Error (RMSE) and Normalized Root Mean Square Error (NRMSE). RMSE is similarly defined as in equation 4.13, however instead the error in concentration is determined instead of the difference in Rrs. NRMSE is very similar to RMSE, but normalized over the highest and lowest concentration present within the scene. This is done to scale the reported error for each component, making them more comparable. NRMSE is described in equation 4.20.

\[
NRMSE = \sqrt{\frac{\sum_{i=0}^{N}(C_{r,\text{est}}-C_{\text{truth}})^2}{N \frac{C_{\text{max}} - C_{\text{min}}}{}}}
\]  

(4.20)
Here $N$ is the number of test spectra, the $ret$ and $truth$ subscripts indicate the retrieved concentration and truth concentration respectively, and $max/min$ represent the maximum and minimum concentrations in the scene. This equation has the disadvantage of seemingly high errors when retrieval is applied to components with small concentration ranges such as CDOM.
Chapter 5

Results

This chapter contains the results of the various experiments carried out in this study. The first section describes the research into how LUT design affects retrieval. Section 5.2 includes all results from satellite imagery retrievals. Following this is Section 5.3, which focuses on retrieval from UAS systems including a comparison of overlapping satellite and UAS imagery collections. The final section focuses on the modeling studies looking at the designs of future Landsat missions and an original multispectral band combination that could be implemented on a UAS.

5.1 LUT Design Study

5.1.1 Concentration Step Range

The results of applying LUTs with varying numbers of concentration steps to a single data set of modeled random test spectra indicated that more steps lead to improved retrieval performance, as shown in Figure 5.1. The retrieval error of each component was shown to decrease for each additional step added to the concentration range, nearly exponentially. The decrease in retrieval error tended to slow after the addition of 7 or 8 steps in the given range, indicating this would be the ideal number of steps to minimize retrieval error and computation time simultaneously. This test also indicates that the interpolation scheme of the LUT can be fairly accurate, assuming that spectra with appropriate IOPs are
used in the interpolation process. The application of an eight entry Top-Heavy and Bottom-
Heavy LUT (i.e. half of steps in the top half and bottom of half of the concentration range, 
respectively) indicated that additional concentration steps are more impactful at lower 
concentrations than higher concentrations. The Bottom-Heavy LUT showed performance 
akin to a LUT with 11-12 equally spaced steps, while the Top-Heavy LUT showed higher 
error akin to a LUT with 5 equally spaced steps. This shows agreement with Gerace (2010), 
who noted that changes in concentration at high component concentrations had a lesser 
effect on the water spectrum.

Figure 5.1: Effect of the number of concentration steps within the LUT on retrieval error. 
The x and o markers represent LUTs where more concentration steps were placed in the 
top half and bottom half of the concentration range, respectively.

5.1.2 Scattering Phase Function Variability

The results of testing scattering phase function variability in the LUT and scene deter-
mined that LUTs perform best when they include all variability within the scene, as shown 
in Figure 5.2. This figure shows best performing retrievals generally lie on the diagonal, 
an intuitive result as this is when the LUTs and random sets have the same amount of
variability in their IOPs. The single best retrieval occurs when the LUT with two scattering phase functions is applied to the set with the same scattering phase functions, as this combination has the lowest potential to incidentally match spectra with the wrong phase function. This can occur easily when applying a LUT with high variability to a data set with low variability (e.g. 13 potential scattering phase functions in the LUT and 2 in the random set), which leads to an increase in retrieval error. The converse of this, high variability in the scene and low variability in the LUT, leads to the highest retrieval error as the LUT entries are not appropriately representative of the scene. This overall this test shows that scattering phase function plays an important part in the retrieval process. This presents a possible problem, as measuring the phase scattering of a waterbody is a difficult task, therefore it may be difficult to determine how to implement it in the model. A future work of substantial interest to attempt to measure the phase scattering of waters of varying turbidity, possibly by using the same methods as Petzold (1972). As a closing note, a similar process to this test was performed for the specific absorption and scattering IOPs which showed identical results to what is reported here. As such, these results are not presented in this section for brevity.

5.2 Satellite Based Retrievals

5.2.1 Landsat 8

The following section shows the results of using Landsat 8 imagery with modeled LUTs to retrieve concentrations of cyanobacteria bloom related pigments. The first section deals covers performance of the atmospheric compensation methods described earlier. This is followed by concentration retrievals from the spectral LUT matching process.

5.2.1.1 Atmospheric Compensation

Atmospherically compensated spectra from Landsat 8 imagery were compared to measured in-situ spectrometer measurements (Section 4.2.2). Three atmospheric compensation methods were used in this comparison, the MoB-ELM, SeaDAS MUMM, and Acolite (Sections 4.5.2, 4.5.3, and 4.5.4 respectively). In-situ data used from three different Owasco
Chapter 5. Results

Figure 5.2: Tables showing the overall normalized root mean square error (NRMSE) for the retrieval of each component. The y-axis represents the number of phase functions in the modeled scene and the x-axis represents the number of phase functions in the LUT.

Lake collections were used in comparison (September 20 & 27 2016, September 23, 2017). A visual comparison of the resampled in-situ and atmospherically compensated spectra for the September 20, 2016 collection is shown in Figure 5.3. Note that the in-situ spectra is represented by the mean and standard deviation of the multiple measurements taken.

This visual comparison shows that all compensation methods generally agree with the in-situ data, however each method shows some error. All three methods tend to preserve the spectral shape of the in-situ spectra, but often overestimate Rrs. This is most notable for the Acolite compensation which overestimates Rrs at all locations and in all bands. The MUMM showed better agreement for the Coastal, Blue, and Green bands (bands 1, 2, and 3) however often overestimated the Red and NIR bands (bands 4 and 5). The MoB-ELM shows better retrieval of the Coastal, Red, and NIR bands compared to the other methods with occasional error in the Blue and Green bands. The relative performance of
Chapter 5. Results

Figure 5.3: Comparison of atmospheric compensated data with in-situ data from Owasco Lake on September 20, 2016. The three compensation methods tested were Acolite, SeaDAS MUMM, and MoB-ELM. These comparisons are shown for 4 different sites throughout the waterbody.

the atmospherically compensated spectra to the in-situ spectra also shows agreement with that reported by Concha (2015).

The Mean Absolute Error (MAE) between the in-situ spectra of three Owasco lake collects was calculated for each atmospheric compensation method, shown in Figure 5.4. The reported Rs error shows good agreement with Figure 5.3. The MoB-ELM shows the best performance across all bands with the MUMM showing the second lowest amount of error and Acolite the largest. Acolite and MUMM both show higher error in the Coastal and NIR bands that the MoB-ELM, as was noted in Figure 5.3. As the LUT retrieval process depends on accurately matching scene spectra with that in the LUT, the varying accuracy of these atmospheric compensation methods infers that atmospheric compensation choice can greatly impact concentration retrieval accuracy. The rest of this work uses the MoB-
Chapter 5. Results

5.2.1.2 Individualized LUT Retrievals

The results of applying individualized LUTs to cyanobacteria bloom imagery indicated that the structured Split LUTs, akin to the method of Concha (2015), performed significantly better than the unstructured Mean LUTs as indicated in Figure 5.5. In Figure 5.5, the structured LUT results are presented in red while the mean LUT results are presented in blue. The retrieval RMSE of each component is shown in Table 5.1 for both the structured and mean IOP LUTS as well.

The retrievals for the unstructured Mean IOP LUT tended to overestimate Chl concentrations and CDOM absorption while underestimating TSS. This is primarily a result of the mean IOPs poorly representing the range of IOPs within the waterbody. In a similar test of applying LUTs generated with the mean in-scene measured IOPs, Gerace (2010) noted that retrieval accuracy decreases the further the actual waterbody IOPs deviate from the

ELM as it shows the best retrieval when compared to in-situ data and has previously shown good performance when implemented in the LUT retrieval process (Raqueno, 2003; Gerace, 2010; Concha, 2015).

Figure 5.4: Mean Absolute Error of various atmospheric compensation methods when compared to resampled in-situ spectra.
Further, the presented results indicate that the retrieval algorithm is attributing the TSS induced scattering to higher concentrations of Chl and the TSS induced absorption to CDOM considering the similarity of both of these IOPs spectral shapes.

The retrievals for the structured IOP LUT showed more consistent performance for Chl, TSS, and CDOM retrievals. The retrieval RMSE also agrees with that reported by Concha (2015), who also structured their LUTs in this manner. Chl retrieval shows fair accuracy at low concentrations, and a slight underestimate at higher concentrations. This is likely caused by atmospheric compensation of the two Honeoye Lake images, as Honeoye consistently had higher concentrations of bloom pigments than Owasco. One of
Chapter 5. Results

the Honeoye images (captured September 27, 2016) also contained Owasco Lake during a cyanobacteria bloom for which in-situ data was also collected. As Owasco generally contained lower concentrations, it was used as the dark target for the MoB-ELM. It is likely that variation of the atmospheric properties between the two lakes lead to a poor compensation, considering the assumptions of homogeneity required by ELM compensation and its derivatives. This is likely also the reason for the two outlying CDOM adsorptions. The second image (captured August 29, 2017) was captured while Honeoye was visible through a break in the otherwise overcast sky. LUT entries were generated with the assumption of a clear sky atmosphere which would impact retrieval, though this effect would be partially mitigated through the use of Rrs in retrieval.

<table>
<thead>
<tr>
<th></th>
<th>Chl mg/m³</th>
<th>TSS g/m³</th>
<th>CDOM 1/m</th>
<th>PC mg/m³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Means LUT</td>
<td>26.33</td>
<td>3.42</td>
<td>0.34</td>
<td>62.25</td>
</tr>
<tr>
<td>Split LUT</td>
<td>13.87</td>
<td>2.32</td>
<td>0.15</td>
<td>71.3</td>
</tr>
</tbody>
</table>

Table 5.1: RMSE of Individualized Split and Mean LUTs to Finger Lakes Imagery

Retrieval of phycoyanin for both sets of individualized LUTs was generally poor with high error. The retrieved PC concentration shows overestimation at nearly all points. Retrieval of pixels with low PC concentration are still fairly concentrated to the lower end of retrieval but distributed across a large range, ranging from 0 to 20 mg/m³, despite the actual values being near zero. Pixels with PC concentrations ranging from 10 to 50 mg/m³ are overestimated in some place by a factor of nearly 10, ranging in retrieved value from 50 to 200 mg/m³. Finally the highest concentrations show a massive underestimate, excluding one point for the unstructured Mean IOP LUTs near 100 mg/m³. This over/underestimation results in a considerably high RMSE for PC retrieval and indicates that retrieval of PC using this method and Landsat 8 imagery is challenging. This is further explored in Section 5.2.1.3.

The performance reported for the structured IOP LUTs is heavily dependent on atmospheric compensation with the lower error MoB-ELM outperforming the SeaDAS MUMM, as shown in Figure 5.7. Applying the same individualized LUTs to the same pixels from imagery corrected by the MUMM lead to a significant increase an error, as indicated by the
Figure 5.6: Results of applying individualized structured LUTs to Owasco Lake imagery captured September 20, 2016.

5 band MUMM results. In Figures 5.4 and 5.3 the MoB-ELM shows similar performance to the MUMM in the Red, Green, and Blue bands but often outperforms the MUMM in the Coastal and NIR. The poor correction of these bands by the MUMM likely cause the retrieval algorithm to misestimate concentrations. This is supported by the fact that the components with the highest retrieval error, Chl and TSS, are the predominant sources of scattering in the NIR band which is consistently overestimated by the MUMM.

Removing the noisiest MUMM bands (Coastal and NIR) from the initial fitting of the retrieval process shows that the increased retrieval error was primarily driven by atmospheric compensation error, as shown in the 3 band MUMM retrieval error of Figure 5.7. The MUMM retrieval error from not using these bands becomes similar to the retrieval of
Figure 5.7: Results of applying individualized structured LUTs to MoB-ELM and MUMM imagery with 5 or 3 band fitting.

the MoB-ELM with 5 bands. Performing retrieval using only the same three bands in the initial fitting with the MoB-ELM compensation shows little to no effect on retrieval error. Retrieval error for the 3 band MUMM showed a little more error than the MoB-ELM, likely driven by the fact that although only 3 bands were used for the initial fitting, all 5 bands were used for the interpolation inducing concentration error. This demonstrates the importance of the initial fitting in the retrieval process, as fitting spectra with appropriate concentrations and IOPs can help mitigate error in the interpolation process. The decreased error after removing the two bands also indicate that it is not only the magnitude of atmospheric compensation error that influences the retrieval process, but also which bands are most afflicted by it, an intuitive result considering the spectral nature of IOPs.
5.2.1.3 Retrieval of PC

The retrieval error of phycoyanin for both sets of individualized LUTs was consistently higher than those of the other components. This is likely caused by a variety of reasons, but is predominately a result of its weak effect on the spectrum. The various sources of error affecting retrieval of PC are discussed in this section.

PC’s IOPs have a very small effect on the overall Rrs spectrum compared to other components. First, it is reported to only have one major absorption feature within the visible, located near 620 nm (Sarada et al., 1999). The mass specific absorption of this peak is reported to vary between 0.003 to 0.007 m$^2$ mg$^{-1}$ by Mishra et al. (2013). Similarly, Simis et al. (2005) reported a mean and standard deviation mass-specific absorption for PC of 0.0095 ± 0.0033 m$^2$ mg$^{-1}$, which was later recalculated to be 0.007 m$^2$ mg$^{-1}$ by Simis et al. (2007), showing agreement with Mishra et al. (2013). This is notably lower than features of other components, for example Chl’s two absorption peaks ($\lambda \approx 440$ nm and $\lambda \approx 680$ nm) reported average values of 0.0313 m$^2$ mg$^{-1}$ and 0.0142 m$^2$ mg$^{-1}$, respectively, for all water samples collected for this study. No other component has a notable absorption feature at 620 nm, however according to the mean reported phycocanin and total absorption values by Mishra et al. (2013) (1.94 m$^{-1}$ and 5.29 m$^{-1}$, respectively) phycocanin accounts for only about third of the absorption at this wavelength. Furthermore, in the context of this work, Landsat 8 has a coverage gap over this specific feature meaning it is not part of the measured spectrum.

In addition to its small effect on the waterbody total absorption, PC likely has little to no effect on total scattering. This is inferred from performing retrieval using individualized structured LUTs where PC is considered either a scattering (red) or non-scattering (blue) component, shown in Figure 5.8. The retrieval RMSE for each set of LUTs is also shown in Table 5.2. Note that the LUTs used in Section 5.2.1.2 had entries for both cases. The retrieval error for both LUT sets is generally similar for all components, excluding Chl and PC. The scattering LUTs show significantly more error than the non-scattering LUTs for Chl. The overestimation in Chl concentration is likely explained by the method of estimating PC scattering discussed in section 3.3.2.5, where measurements of Chl scattering measured by Bukata et al. (1995) were subtracted from measurements of scattering cyanobacteria cells by Ahn et al. (1992). A significant portion of the estimated scattering
is likely directly related to cyanobacteria cells and Chl as opposed to PC. The retrieval algorithm attempts to account for this residual scattering in the retrieval process by fitting spectra with higher Chl concentrations. Retrieval error of PC was lower when using the scattering LUTs, but not low enough to imply that it was being meaningfully retrieved by the LUT process. The increased retrieval error from applying the non-scattering LUTs is likely due to the retrieval process having more difficulty determining PC’s contribution to the waterbody spectrum when it is treated as purely absorbing component. PC was considered a non-scattering component for all LUTs in Section 5.2.1.4. It should be noted that the majority of imagery pixels fit to structured LUT entries without PC scattering based on the similarity of retrieval to the LUTs with no PC scattering (Table 5.1 and 5.2, respectively).

<table>
<thead>
<tr>
<th></th>
<th>Chl mg/m³</th>
<th>TSS g/m³</th>
<th>CDOM 1/m</th>
<th>PC mg/m³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scattering LUT</td>
<td>35.18</td>
<td>2.68</td>
<td>0.12</td>
<td>47.27</td>
</tr>
<tr>
<td>Non-Scattering LUT</td>
<td>13.33</td>
<td>2.28</td>
<td>0.15</td>
<td>71.3</td>
</tr>
</tbody>
</table>

Table 5.2: RMSE of Scattering PC and Non-Scattering PC LUTs to Finger Lakes Imagery

Atmospheric compensation plays a role in the poor retrieval of PC, also as a result of PC’s small influence on the spectrum. Figure 5.9 compares the MAE of the MoB-ELM atmospheric compensation (from Figure 5.4) with the average Rrs of waterbody spectra per component per concentration, determined from spectra within the unstructured mean LUTs. This average was calculated for each concentration of each component and includes all variation in component IOPs and other component concentrations. Note that PC was considered to be non-scattering in Figure 5.9. Comparing the average Rrs per component per concentration and atmospheric compensation error shows that the concentration change necessary to have an Rrs change greater than atmospheric compensation error is larger for PC than other components, close to 100 mg/m³. This infers that atmospheric compensation could be a large source of error in retrieving PC. It should also be noted that the Landsat 8 band in which the PC has the most impact (Green, which covers a portion of the slope of the absorption peak) is also the band which contains the most atmospheric
compensation error. Interestingly the concentration change that surpasses the error for Chl (slightly less than 25 mg/m$^3$) is similar to the retrieval RMSE from applying the structured LUTs, indicating that atmospheric compensation error may be limiting retrieval performance. It should be noted that for retrieval of CDOM the change in absorption required to cause an Rrs change greater than the reported atmospheric compensation error infers that CDOM should be poorly retrieved as well, which is not supported by the retrieval results. This would imply that atmospheric compensation is only a partial contributor to error, however not the only source.

The initial fitting of the LUT process may also be a source of error in the retrieval of
Figure 5.9: Mean water spectra Rrs for varying component concentrations compared to atmospheric compensation MAE. In this case PC is considered non-scattering.

PC. As stated in Section 4.6.2, the initial fitting process weights all bands equally when determining which LUT spectrum is least different than the pixel. This means that a component with a strong influence on the water’s Rrs spectrum (e.g. Chl) could cause the retrieval algorithm to select a LUT entry that poorly represents components with weaker influence (e.g. PC) on the basis of better fitting bands that the weaker influence component has little to no effect on. As an example with PC and Chl, a LUT entry that poorly represents PC may be chosen due to it matching the effects of Chl in the Coastal and NIR bands, where PC has little influence. This improper matching would allow the error to propagate through the interpolation process to the retrieved results, leading to the high error reported. It should be noted that this misfitting is not unique to PC, however PC’s weak influence increases its susceptibility.
5.2.1.4 Global LUT Retrievals

The results of implementing a single global LUT in the retrieval process shows LUTs with more structure tend to have less retrieval error. Figure 5.10 shows the comparison of retrieved to measured component concentrations for each of the three global LUTs. The retrieval RMSEs are shown in Table 5.3 and a comparison of NRMSEs is shown in Figure 5.11. As indicated by these figures, the Literature LUT had the best retrieval, followed by the structured LUT and lastly the unstructured LUT. The structured and unstructured LUTs perform fairly similarly across components. For Chl, both of these LUTs generally match the measured concentrations at low values \((\approx 0 - 20 \text{ mg/m}^3)\) with a few high overestimates. At mid to high values \((\approx 20 - 50 \text{ mg/m}^3)\), the Unstructured LUT generally outperforms the Structured which over both and underestimates concentrations. At very high concentrations, both LUTs highly underestimate the Chl concentrations. For TSS, the Unstructured LUT performs slight better for lower concentrations while the Structured LUT performs slightly better at higher concentrations, however both frequently underestimate compared to the measured value. For CDOM, both LUTs have similar retrievals which often overestimate low concentrations. Both LUTs show similarly poor PC retrievals, likely for the same reasons as the individualized LUTs (Section 5.2.1.3).

<table>
<thead>
<tr>
<th></th>
<th>Chl mg/m^3</th>
<th>TSS g/m^3</th>
<th>CDOM 1/m</th>
<th>PC mg/m^3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstructured LUT</td>
<td>21.08</td>
<td>4.10</td>
<td>0.413</td>
<td>79.93</td>
</tr>
<tr>
<td>Structured LUT</td>
<td>19.48</td>
<td>4.89</td>
<td>0.39</td>
<td>82.67</td>
</tr>
<tr>
<td>Literature LUT</td>
<td>15.44</td>
<td>3.16</td>
<td>0.15</td>
<td>63.43</td>
</tr>
</tbody>
</table>

Table 5.3: RMSE of Global LUTs applied to Finger Lakes imagery.

The Literature LUT outperformed the other global LUTs in retrieving all components. The largest relative drop in retrieval error occurred with the retrieval of CDOM, especially at lower concentrations. Chl also saw a notable drop in error, mostly from concentrations on the lower end of the range. In contrast the Literature LUT tended to underestimate higher concentrations, possibly as a result of keeping a constant mass-specific scattering spectrum and scattering phase functions for all stages of the LUT. TSS retrieval was similar to that of Chl, accurate at lower concentrations but underestimated at higher ones.
Finally PC performed poorly, but better than the other LUTs, likely to limiting the range of phycocyanin according to its relation with Chl (Section 3.3.2.5).

The higher errors of the global LUTs can be partially attributed the choice of IOPs. Unlike the individualized LUTs, the global LUTs only use a subset of the measured IOPs to limit computation time. This is especially true for the Unstructured LUT which consists of 172,800 entries despite only containing three absorption IOPs for Chl & TSS and two for CDOM & PC. As with applying the Individualized Means LUTs, the more the scene IOPs deviate from the limited selections made, the more error is incurred within the retrieval.

Figure 5.10: Retrieved versus measured concentrations generated using global LUTs.
It is likely that the variability that each table does contain assists with improving retrieval compared to the individualized unstructured Mean IOP LUTs.

The varying structures of the global LUTs also contribute to the higher retrieval errors. The Unstructured LUT is representative of numerous different waterbody conditions as a result of it containing every combination of its concentrations and IOPs, increasing the likelihood of misfitting in the initial matching process and leading to higher retrieval error. The Structured LUT lessens this problem through its division between high and low concentration IOPs as the components that have a greater effect on the water spectrum are more likely to influence the fitting process and choose the appropriate portion of the LUT (high concentration or low concentration), leading to a retrieval that mimics the relations noted between components (Section 3.3.2.6). If the split between the high and low concentration portions is not appropriate, or if the concentrations of the waterbody fall...
near the area of the split, error can be incurred. It should be noted that none of the higher concentration pixels were fit with entries in the higher concentration LUT, hence why the structured LUT generally underestimated higher concentrations of all components. This is a problem for the Literature LUT as well, considering as many high concentration bright pixels ended up selecting a stage of the table that had middling concentrations of Chl and low concentrations of TSS and CDOM. This combination minimizes absorption and increases scattering to make bright LUT entries that the retrieval process chooses over the appropriate fits.

Given that the structured LUT is split based on high/low concentrations and the Literature LUT is set into stages based on Chl concentration, they are limited in the scope of waters they represent. While appropriate for lakes or ponds where the components will generally be found in high concentrations together, they would be unable to retrieve appropriately from waters where single components drive the color such as sediment heavy waters. As such their true ability to be "global" is further limited. This suggests that future work should focus on methods of how the retrieval process can better choose from the LUT rather than to change the design of the LUT itself in order to better handle IOP variability between waters.

5.2.2 Sentinel-2

The application of the Literature based LUT to Sentinel-2 imagery of Lake Erie showed a strong relation of measured to retrieved concentrations of Chl, as shown in Figure 5.12. This result is promising considering that the regression consistent over the entire range \((1 \text{ mg m}^{-3} \text{ to } 130 \text{ mg m}^{-3})\), which was greater than that of the Landsat 8 data set. Unlike the application to the Landsat 8 data set, Chl concentrations were not underestimated at higher concentrations. This is likely due to the differences in band sampling between the two systems. Sentinel-2, unlike Landsat 8, has 3 bands in the NIR (705 nm, 741 nm, and 782 nm) which have been indicated to improve Chl retrieval (Section 5.4.1.3). The retrieved concentrations have a larger RMSE than Landsat 8 retrievals, as noted in Table 5.4, likely caused by the IOPs in the Literature LUT being measured from the Finger Lakes as opposed to Lake Erie.
Chapter 5. Results

Figure 5.12: Retrieved versus measured concentrations from Sentinel-2 Imagery generated using the Literature LUT.

<table>
<thead>
<tr>
<th>RMSE</th>
<th>Chl $mg/m^3$</th>
<th>PC $mg/m^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentinel-2</td>
<td>28.66 $mg/m^3$</td>
<td>54.14 $mg/m^3$</td>
</tr>
</tbody>
</table>

Table 5.4: RMSE of Sentinel-2 Retrievals.

The retrieval of PC from Sentinel-2 imagery showed similar levels of accuracy as the retrievals of Landsat 8, with a fairly weak regression. The low accuracy of retrieval can easily be explained to be a result of all of the reason listed in Section 5.2.1.3, though the atmospheric compensation error cannot be fully determined due to lack of reference in-situ spectra. Sentinel-2 also has a larger coverage gap near the 620 nm PC absorption feature than Landsat 8. The relation between the measured and retrieved values of PC likely arises from the structure built into the Literature LUT, and the retrieval error would likely be greater if that structure were to be removed.
Chapter 5. Results

5.3 UAS Based Retrievals

5.3.1 Matching UAS and Satellite Retrievals

Applying the Literature LUT to UAS and satellite imagery of Long Pond showed similar levels in retrieval accuracy, as shown in Table 5.5. The UAS and satellite retrievals were similarly accurate retrieving Chl and CDOM while both consistently underestimating the concentrations of TSS. There is no retrieval of PC as this was no active bloom in Long Pond at this time. The Chl concentrations retrieved from the drone imagery were generally more accurate than those from satellites. This is a result of the Tetracam system having better spectral coverage in the NIR (Section 3.2.3.3), similar to the results of the Sentinel-2 testing.

<table>
<thead>
<tr>
<th></th>
<th>Long Pond West</th>
<th></th>
<th>Long Pond East</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chl ($mg/m^3$)</td>
<td>59.3</td>
<td>61.1</td>
<td>1.8</td>
<td>71.5</td>
</tr>
<tr>
<td>TSS ($g/m^3$)</td>
<td>21.1</td>
<td>10.0</td>
<td>-11.2</td>
<td>10.0</td>
</tr>
<tr>
<td>CDOM ($1/m$)</td>
<td>1.40</td>
<td>1.43</td>
<td>0.03</td>
<td>1.65</td>
</tr>
</tbody>
</table>

Table 5.5: Results of applying Literature LUT to near-simultaneous UAS and Satellite Imagery of Long Pond

The concentration maps, along with original imagery, for both the UAS and satellite are shown in Figures 5.13 and 5.14, respectively. These figures show that glint has a much greater affect on UAS imagery, than satellite imagery. This is in part due to the different times of day of the collect, the higher winds/wavy water surface at the time of the UAS collection, and the relative sun-target-sensor geometry of the UAS including unintended roll, pitch, and yaw changes caused by wind. Through scene averaging and use of the MoB-ELM correction most of this glint can be accounted for, allowing the accurate retrieval.

5.3.2 Bloom Retrievals

The reported RMSE (Table 5.6) for both multispectral Tetracam and hyperspectral Nano-Hyperspec were similar to those reported from the Landsat 8 retrievals. It should be reiterated that these retrievals were also performed using the Literature LUT. The
Figure 5.13: Results of LUT applied to a sub-sectioned Tetracam image of Long Pond.

retrieved values versus the measured values for the multispectral (MSI) and hyperspectral (HSI) systems are shown in Figure 5.15. These comparison show that both systems perform well for Chl and TSS, though the hyperspectral system outperforms the MSI for the latter. This is likely due to the differences in spectral sampling between the two systems, with the MSI system leaving considerable gaps in the 400 to 900 nm region.

For the MSI system, the retrieval error of CDOM and PC is fairly similar to that of Landsat 8. The high error in PC retrieval is likely due to the reasons listed in Section 5.2.1.3 as this system also lacks a spectral band over the PC absorption feature. Similar to Sentinel-2, the higher PC concentrations retrieved are likely a result of the structure of the Literature LUT. The lower retrieval error in CDOM, especially compared to the HSI system, is likely due to the lower noise of the MSI system. This lower noise also enabled a MoB-ELM atmospheric compensation as opposed to a panel based ELM correction, which
Figure 5.14: Results of Literature LUT applied to a Landsat 8 image of Long Pond.

Improves retrieval results as shown in Section 5.2.1.2.

<table>
<thead>
<tr>
<th></th>
<th>Chl</th>
<th>TSS</th>
<th>CDOM</th>
<th>PC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multispectral</td>
<td>11.35 mg/m³</td>
<td>4.27 g/m³</td>
<td>0.15 m⁻¹</td>
<td>60.87 mg/m³</td>
</tr>
<tr>
<td>Hyperspectral</td>
<td>9.29 mg/m³</td>
<td>2.67 g/m³</td>
<td>0.78 m⁻¹</td>
<td>25.39 mg/m³</td>
</tr>
</tbody>
</table>

Table 5.6: RMSE of Literature LUT applied to UAS imagery of cyanobacteria blooms.

For the HSI system, CDOM and PC retrieval error was notably different than that of the Landsat 8 retrievals. This is likely a cause of both the differences in spectral coverage as well as noise. The measured HSI spectra were generally noisy, even after spatial binning, as shown in 5.16. As the interpolation process of the retrieval method will often increase or decrease the concentrations of components with weaker influence to better fit the over
all spectra, which contributes to the poor retrieval error of both components. This is mitigated in the case of PC due to the HSI system having coverage over the PC absorption feature, and thus retrieving PC with the lowest error of all systems. It should be reiterated that this is likely assisted by the structured of the Literature LUT, and error may increase if tested with a less structured LUT. It can also be seen in 5.16 that the HSI spectra is brighter in the 400 to 500 nm range than the best fit LUT entry. This indicates that potentially the absorption spectra used in the LUT for CDOM was likely a poor fit for the

Figure 5.15: Results of applying the Literature LUT to UAS imagery over a bloom in Honeoye Lake.
actual absorption of CDOM during this period, leading to a worse retrieval.

Figure 5.16: Comparison of Headwall Nano Spectra (Black) to the best fit (Blue) and most appropriate (Red) LUT entries.

5.4 Improved Spectral Coverage Modeling Studies

Modeling studies were implemented to determine how systems could be improved for water quality remote sensing, the results of which are presented in this section. The first portion presents results for improving future Landsat missions, both considering multispectral systems and imaging spectrometers. The second portion presents the results of determining an ideal multispectral setup that could be implemented on a 6 band UAS system.
5.4.1 Future Landsat Designs

5.4.1.1 Modeled Data Validation

As a simple test to verify our modeling process we compare the TOA radiance spectra of our test pixels to that of radiometrically calibrated Landsat 8 scenes containing cyanobacteria blooms. Two Landsat 8 images, collected in July and September 2015 (LC8019032015209LGN00 and LC8019032015257LGN00) over Lake Erie, were selected for comparison. A region of interest containing bloom and non-bloom waters north of Port Clinton, Ohio was selected by visual inspection. The region of interest mean radiance and its standard deviation for each band were calculated and compared against the mean and standard deviation of the 1000 test pixels which also represent bloom and non-bloom conditions, shown in Figure 5.17.

In each band, the standard deviation radiance of the modeled spectra overlaps with the standard deviation radiance of at least one of the two images. It is expected that these spectra should be similar considering the signal is dominated by atmospheric radiance. Spectral differences may arise from IOP variability, random aspect of the test pixels, and the MODTRAN inputs used to generate the atmosphere. These results indicate the modeling process produces reasonably realistic spectra.

5.4.1.2 Baseline Data

The Landsat 8 system was modeled, using its spectral sampling and noise characteristics, to set a baseline of retrieval error against which other systems could be compared. Water component concentrations were retrieved from all 1000 modeled test pixels. Error was reported as NRMSE in Figure 5.18.

5.4.1.3 Additional Multispectral Coverage

The results of modeled water constituent retrieval are provided in Figure 5.19 for the addition of a yellow band, in Figure 5.20 for the addition of a red edge band, and in Figure 5.21 for the addition of both bands. For these three figures the NRMSE of the retrieved component varies as a function of band center and bandwidth allowing critical assessment
of tradeoffs for these variables.

The results in Figure 5.19 indicate a yellow band with a HWHM in the range of 15 nm and 20 nm with a center around 625 nm leads to the greatest improvement in retrieval. The retrieval of Chl, TSS, and CDOM all outperform the baseline Landsat 8 system when a band is added with said properties. CDOM outperforms the base system at all locations regardless of band placement, an intuitive result as adding more spectral information should allow for better fitting of the observed spectra. Chl and TSS outperform the baseline system for the majority of tested bands except those closer to 590 nm. This degraded retrieval accuracy is a result of redundant information from the overlap of the modeled yellow band and the existing Landsat 8 green band which leads to confusion in the LUT retrieval process. This is most clearly illustrated by Chl, which shows retrieval error decrease faster for thinner bands as band centers increase from 590 nm.

Unlike the other water components, for PC the added spectral sampling beats the base Landsat 8 system only sporadically. Furthermore, the band centers that show more
improvement (605 nm to 615 nm) are slightly removed from the PC absorption feature. This offset may allow the algorithm to better parse out the IOPs of PC from the other components in this region. Conversely this offset may help the retrieval algorithm better determine the IOPs of the other components and subsequently PC however this is not supported by the reported retrieval errors of the other components in said region.

The retrieval results of the added red edge band show comparable retrieval improvement as that of the added yellow band driven primarily by Chl and PC. These components show an improvement over the baseline system regardless of band placement and width. The retrieval error of Chl decreased significantly with the addition of the red edge band, as was expected based on its absorption minimum and fluorescence emission peak in this region. This improvement outperforms the yellow band while remaining fairly insensitive to location, an encouraging result considering that the placement of a red edge band would most likely be determined based on terrestrial needs. The best Chl retrieval tends to favor band centers around 695 nm. It is most consistent with centers ranging from 690 nm to 705 nm, with an approximate HWHM around 15 nm to 20 nm.
Figure 5.19: Retrieval NRMSE for each component using modeled future Landsat systems with Landsat 8 spectral coverage and an added yellow band. Each graph represents a different component with the y-axis being the center of the added yellow band, the x-axis being the approximate HWHM, and the color axis being the NRMSE. Carefully note that the NRMSE scales for each component are variable to allow for better visualization. The retrieval NRMSE from the baseline Landsat 8 model is labeled on the colorbar with an asterisk.
PC shows improvement with all possible bands with the best retrievals taking place with band centers at the end of the tested range. This is most likely caused by the mass specific scattering spectrum used for PC, which may have retained scattering properties of the cell not related to Chl, but still related to cyanobacteria. This includes an increased scattering at longer wavelengths, specifically in this NIR region, which has been documented in the literature (Ganf et al., 1989).

TSS and CDOM showed a high sensitivity to band center with the best performing bands corresponding to those closest to the visible regime. Bands with centers greater than 685 nm tended to underperform the baseline at retrieving these components, most likely due to the retrieval algorithm misattributing the impact of TSS and CDOM in this region. This is an intuitive result TSS and CDOM absorption both decrease with increasing wavelength leading to minimal information in the NIR region of the spectrum. TSS shows more potential improvement and less error than CDOM due to its contribution to scattering in the NIR.

The simultaneous addition of a yellow band and red edge band, shown in Figure 5.21, outperformed the addition of either band individually. In Figure 5.21, each of the 30 outlined boxes shown represent a different band combination, where each individual colored square represents a different combination of widths. Wider bandwidths are placed lower in each box for the yellow band (y-axis), and further right for the red edge band (x-axis). This means that the results generated using the two most narrow bands is always in the top-left of the box and those from the two widest bands are always in the bottom-right.

The two-band system retrieval error for all components was lower than that of the baseline for nearly all configurations. The only exception to this is the retrieval of CDOM by systems with yellow bands at 590 nm and red edge bands at longer wavelengths. These bands have previously been noted to lead to under performance in CDOM retrieval when added individually. The best configuration for each individual component vary though they generally agree with what has been observed from the bands individually with the exception of Chl and PC. A 590 nm band paired with already well performing red edge led to the lowest Chl retrieval error despite the 590 nm band causing under performance when added individually. Chl also saw the greatest decrease in retrieval error, outperforming the sole addition of yellow band retrieval regardless of band position or width. PC was
Figure 5.20: Retrieval NRMSE for each component using modeled future Landsat systems with Landsat 8 spectral coverage and an added red edge band. Each graph represents a different component with the y-axis being the center of the added red edge band, the x-axis being the approximate HWHM, and the color axis being the NRMSE. Carefully note that the NRMSE scales for each component are variable to allow for better visualization. The retrieval NRMSE from the baseline Landsat 8 model is labeled on the colorbar with an asterisk.
retrieved equally well by a two-band system with yellow bands centered from 590 nm to 630 nm despite yellow bands at longer wavelengths showing worse performance when added alone. The change in Chl and PC retrieval performance for these noted bands is a result of the spectral nature of IOPs and the LUT retrieval process. The addition of a single band can lead to an improvement in retrieval, but as the LUT retrieval process is essentially a spectral unmixing problem there is uncertainty in determining the extent that each component affects the newly sampled region. Changes in the spectrum by each component covary due to the spectral of IOPs, meaning that this uncertainty can be decreased by sampling additional regions of the spectrum, leading to better a fit from the LUT and hence decreased retrieval error.

To summarize, the addition of a yellow and red edge band to the existing Landsat spectral coverage improves retrieval in most cases. Both the addition of only a yellow band or only a red edge band showed improvement in retrieval with some dependency on position and width. The addition of both bands simultaneously showed improvement nearly regardless of band position and width meaning that band configurations would be amenable to the needs of other application areas.

5.4.1.4 Imaging Spectrometer

The results of testing imaging spectrometers tended to show increasing retrieval error with increasing band spacing, as shown in Figure 5.22. This is an intuitive result as wider bands average more spectral features when sampling the spectrum. Across all components, each modeled spectrometer had a retrieval error that was less than the baseline Landsat 8 model. This result may seem unexpected considering that a system with 100 nm increment spacing has only six bands in the 400 nm to 900 nm range, compared to Landsat 8’s five, but the spectrometer blankets this entire range while Landsat leaves considerable gaps (Figure 4.7). Systems with a 60 nm or lower band spacing showed consistent retrieval error, with only small variations. A small upturn of noise is noticeable for the 5 nm and 10 nm spectrometer, especially in the case of CDOM, which is most likely from increased noise due to smaller bandwidths. For spectrometers with a spacing between 100 nm and 70 nm retrieval error did not consistently decrease, but instead changed sporadically. This
Figure 5.21: Retrieval NRMSE for each component using modeled future Landsat systems with Landsat 8 spectral coverage and both an added red edge and yellow band. Each graph represents a different component with the y-axis being the center of the added yellow band, the x-axis the center of the added red edge band, and the color axis being the NRMSE. Each of the 30 large boxes represent a different band combination, with the sub-elements representing different HWHMs. The sub-elements represented HWHMs of 15, 18, 20, 25, and 30 nm from top to bottom for the yellow band and left to right for the red edge band. Carefully note that the NRMSE scales for each component are variable to allow for better visualization. The retrieval NRMSE from the baseline Landsat 8 model is labeled on the colorbar with an asterisk.
Figure 5.22: NRMSE of retrieval for modeled imaging spectrometers with varying incrementing bandwidths. Retrieval error of each component is represented by a different line, where all lines show increasing error with increasing bandwidth.

is likely due to the different band spacings changing the sampling of spectrum features leading to better or worse retrieval. This sporadic change in retrieval error suggests that a more optimal alignment of band sampling locations and features exists for wider spaced systems and was in part motivation for examining how shifting the spectrometer band center locations affected the performance of imaging spectrometers.

Imaging spectrometer band centers showed sensitivity to band shifting that increased with bandwidth, as shown in Figure 5.23. Figure 5.23 also confirms that the sporadic error in the wider spaced system was due to the relative position of the band RSRs and spectrum features. The two systems most resilient to shifting were the ones with lower spacing, i.e. 20 nm and 40 nm, with the 20 nm system outperforming the 40 nm system. The two systems with larger band spacing showed a larger increase in error as centers were shifted. These wide band systems have larger areas of decreased spectral response which would average multiple spectral features depending on placement. The 60 nm and 80 nm spectrometer show a notable decrease in retrieval error when shifted 40 nm and 60 nm, respectively, caused by both systems adding an additional band to maintain full coverage of the entire 400 nm to 900 nm range. These results indicate positions of the band centers
Figure 5.23: Retrieval error of imaging spectrometers with band centers shifted until their positions repeated. Four spectrometer spacings were tested (20, 40, 60, and 80 nm). Each plot in this section is relates to the retrieval error of a different component.

need to be considered if moving to an imaging spectrometer with wider band spacing. A move to a system with narrower band spacing may allow for a better performance across application areas.

The retrieval error of the resampled spectra with a 20 nm FWHM spectrometer, shown in Figure 5.24, matches the retrieval of the base Landsat 8 system well, with little sensitivity to band shifting. The NRMSE difference between the baseline and resampled systems was less than 0.01 for almost all components except CDOM which was slightly higher around 0.01 to 0.02. Reconfiguring the spectrometer sampling to maximize the number of band centers that fall within the Landsat 8 band RSRs eliminates this difference. Similarly, performing the same test using a spectrometer with 10 nm spacing and no noise adjustment for bandwidth eliminates this difference as well. This indicates that retrieval ability of the
resampled system is dependent on the relative placement of spectrometer band centers to the Landsat RSRs, inferring that spectrometers with narrower spacing are better suited for this task. Overall, these results show potential that spectrometer data can be resampled to maintain data continuity with Landsat 8 for water applications.

### 5.4.2 Original Multi-Spectral Setup

The results of randomly testing combinations of six multispectral bands for implementation on a UAS system indicated that the most desirable area for spectral coverage is
between 400 to 420 nm, 690 to 700 nm, and 790 to 810 nm, as shown in Figure 5.25. Figure 5.25 displays a histogram that counts each time a wavelength in the 400 to 900 nm range fell within a chosen set of best bands from the random process. Secondary peaks appeared around 460 nm, 555 nm, and 665 nm as well. The band selection process also heavily favored narrower bands (10 or 20 nm FWHMs) over wider bands even though the noise scaling with bandwidth was still included in this test.

Figure 5.25: Histogram of each time each wavelength in the 400 to 900 nm range was chosen from the random best UAS band selection process.

Four different band combinations were determined based on the results in Figure 5.25. The band centers and HWHMs for each combination are reported in Table 5.7. The first combination was the best performing set of bands from the random test (Best Random Result). The second set was determined based on inspection of Figure 5.1 (Results Based Combination). In addition to these two sets, a different combination capable of imple-
menting multiple algorithms from literature was also determined (Multi-Algorithm). This combination is capable of implementing blue/green band ratios akin to the Ocean Color Chl algorithms (O’Reilly and Werdell, 2019), red/NIR band ratios which allow retrieval of Chl in turbid waterbodies (Gitelson et al., 2007, 2008), NDVI which is useful for detecting surface scums, the Cyanobacteria Index (Wynne et al., 2008), various algorithms for retrieval of PC (Lee et al., 2002; Mishra et al., 2013), and true color imagery. This combination was considered as the other two sets optimize band placement solely for the LUT matching process, and these optimized sets do not have sufficient spectral coverage for most of these approaches, excluding the red/NIR ratios and NDVI. The final combination chosen was a compromise between the results of the random testing and the coverage which allowed for multiple band algorithms. This compromise retains the ability to do blue/green ratios, red/NIR ratios, NDVI, and the PC retrieval algorithm by Mishra et al. (2013). Each of these band sets were used to in the modeled end-to-end retrieval process, and the NRMSE is reported in Table 5.8 along with the performance of the base Tetracam system.

<table>
<thead>
<tr>
<th>Band Centers/HWHMs</th>
<th>Band 1</th>
<th>Band 2</th>
<th>Band 3</th>
<th>Band 4</th>
<th>Band 5</th>
<th>Band 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tetracam</td>
<td>490/5</td>
<td>550/5</td>
<td>680/5</td>
<td>720/5</td>
<td>800/5</td>
<td>900/10</td>
</tr>
<tr>
<td>Best Random Result</td>
<td>400/5</td>
<td>656/2.5</td>
<td>700/10</td>
<td>740/2.5</td>
<td>810/5</td>
<td>830/5</td>
</tr>
<tr>
<td>Multi-Algorithm</td>
<td>440/10</td>
<td>555/5</td>
<td>620/5</td>
<td>665/5</td>
<td>680/5</td>
<td>710/5</td>
</tr>
<tr>
<td>Compromise Combination</td>
<td>440/10</td>
<td>555/5</td>
<td>620/5</td>
<td>665/5</td>
<td>710/5</td>
<td>800/10</td>
</tr>
<tr>
<td>Results Based Combination</td>
<td>410/10</td>
<td>460/10</td>
<td>620/5</td>
<td>680/5</td>
<td>710/5</td>
<td>800/10</td>
</tr>
</tbody>
</table>

Table 5.7: Band centers/HWHMs for various possible UAS band combinations.

The reported NRMSEs for each of the combinations indicate that the best performing band combinations were those that were optimized for the LUT method. The top band combination was the Results Based Combination followed by the Best Random Result. This is intuitive as the Best Random Result is dependent on only one iteration of testing while the Results Based Combination is based on the aggregated results of all tests. The majority of improvement over the base Tetracam system occurred through the retrieval of CDOM, followed by Chl. PC and TSS showed little change. The Multi-Algorithm showed improvement only in the retrieval of CDOM, with most other components showing more
error than the baseline system. This combination should not be expected to perform as well as the others as it has not been optimized for this process. The Compromise Combination performs Chl retrieval better than the Multi-Algorithm and Tetracam bands, however incurs higher error for the other components. These results indicate that spectral bands can be selected that optimize the LUT retrieval process, however attempting to implement bands that allow for an ensemble application of algorithms will degrade performance.

<table>
<thead>
<tr>
<th></th>
<th>Chl</th>
<th>TSS</th>
<th>CDOM</th>
<th>PC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tetracam</td>
<td>0.16</td>
<td>0.28</td>
<td>0.44</td>
<td>0.17</td>
</tr>
<tr>
<td>Best Random Result</td>
<td>0.13</td>
<td>0.30</td>
<td>0.23</td>
<td>0.18</td>
</tr>
<tr>
<td>Multi-Algorithm</td>
<td>0.17</td>
<td>0.29</td>
<td>0.26</td>
<td>0.25</td>
</tr>
<tr>
<td>Compromise Combination</td>
<td>0.13</td>
<td>0.31</td>
<td>0.31</td>
<td>0.21</td>
</tr>
<tr>
<td>Results Based Combination</td>
<td>0.11</td>
<td>0.27</td>
<td>0.26</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Table 5.8: NRMSE of various possible UAS band combinations.
Chapter 6

Conclusions

6.1 Conclusions

6.1.1 LUT Design

The results of the LUT design study indicate that population of the LUT used in the retrieval process has a significant affect on retrieval accuracy. Performance of the interpolation portion of the retrieval process showed sensitivity to the number of steps within the range of concentrations modeled for each component, with more steps across the given range leading to better results for all components. Furthermore, these steps lead to better performance at the lower end of the concentration range than the upper end. Testing the effect of phase function variability within the LUT and within the scene showed that the best performance occurs when the IOPs of the scene are known and used to generate the LUT. Lacking that, LUTs with a comprehensive number of IOP sets performs second best as the appropriate set of IOPs exists within the LUT, though pixels can sometime fit to entries with improper IOPs leading to error. The largest error occurs when the LUT does not contain an appropriate IOP set. Though this was only demonstrated through testing variability within the scattering phase function, results not shown performing similar tests with specific absorption and scattering indicate similar results.
6.1.2 Satellite Imagery Retrievals

The results of applying the spectral LUT matching process to Landsat 8 imagery indicated that accurate concentrations of chlorophyll, TSS, and CDOM could be retrieved, though this is dependent on method of atmospheric compensation and LUT design. The MoB-ELM atmospheric compensation showed the lowest error in retrieving waterbody Rrs spectra and lead to the most accurate concentration retrievals. The best performing LUTs were those that were individualized to each waterbody, and that were split into two portions: high and low concentrations. In comparison, individualized LUTs that used the mean IOPs for each scene performed worse, which shows agreement with the LUT design study as the mean IOPs are not similar to any one IOP measured in the scene and therefore not an appropriate set for any pixel used in testing. This also showed agreement with results from Gerace (2010). Retrieval of phycocyanin from Landsat imagery was often inaccurate, attributed to its low impact on the waterbody signal, the lack of spectral sampling over the phycocyanin absorption feature, and error in the waterbody spectrum caused by atmospheric compensation.

In testing global LUTs, it was determined that a LUT structured around component relations reported in waterbodies could perform retrieval with levels of accuracy similar to that of the individualized LUTs. While the Literature LUT outperformed the unstructured LUT and structured LUT, it was noted that the design built into the LUT limited its applicability to water types (e.g. rivers with high concentrations of TSS and low concentrations of chlorophyll). Furthermore, the Literature LUT tended to cause underestimates of high concentration pixels from the Landsat imagery most notably in Chlorophyll and TSS. This underestimation disappeared when applying the same LUT to Sentinel-2 imagery, which showed a fairly strong relation between measured and retrieved chlorophyll. The difference in performance of each system with the same LUT indicates the importance of spectral sampling in the retrieval process.

6.1.3 UAS Imagery Retrievals

The retrieval accuracy determined from applying the spectral LUT matching process to UAS imagery showed similar levels accuracy to that as the satellite systems. This is most
notable in the matched UAS and satellite retrievals where both systems performed fairly similarly. The differences that arose were suggested to be a result of each system’s spectral sampling but may also arise from the increased susceptibility of glint for UAS systems, though in this case the glint was able to be accounted for through use of the MoB-ELM compensation. Retrieval from multispectral Tetracam and hyperspectral Headwall Nano-Hyperspec imagery of bloom water also showed similar levels of retrieval error as those of satellite retrievals. Though the hyperspectral system showed significant noise in its spectral measurements, it had the best reported retrieval of PC. It should be noted that this is the only system tested that had spectral coverage of PC’s spectral features, indicating the need for additional spectral coverage for other systems. This also indicates that the retrieval of phycocyanin could be improved if the amount of noise in the hyperspectral measurements was decreased.

6.1.4 Future System Improvements

Modeling studies to determine how future Landsat systems could be improved for water quality monitoring showed that adding a single band to the existing Landsat 8 system could decrease error in retrieving all components (Figure 6.1), both in the case of the Red Edge and Yellow band. This was dependent on the bands placement and width. The best performing bands were generally those with smaller bandwidths. Simultaneously adding both bands not only lessened this dependency but allowed for a greater decrease in error. Retrieval accuracy was indicated to relate better to the increased sampling over the visible/NIR region. This is reinforced by the modeled imaging spectrometer with 100 nm band spacing (i.e. 5 bands) that outperformed the Landsat 8 baseline model.

This testing also demonstrates that transition to an imaging spectrometer for the VNIR portion of the spectrum for a future Landsat system will outperform the Landsat 8 baseline for water applications. The lowest retrieval error was achieved when the bands had an incremental spacing of about 30 nm. This spacing can be inferred to be the best trade-off between spectral resolution and the increased noise caused by the narrow bands while also being insensitive to band center position. Our results also imply that resampling to a historic Landsat product is feasible with this spacing. The transition to an imaging spec-
Figure 6.1: Comparison of spectral coverages tested along with the baseline system. Multispectral systems are represented by the mean retrieval error over all centers and widths. The imaging spectrometers reported range in FWHM from 60 nm to 5 nm with centers shifted to the location of optimal performance.

Modeling studies to determine a best set of six bands to implement on a UAS Tetracam sensor indicated that a few distinct regions of the VNIR spectrum were ideal for sampling. This included regions around 400, 680, 710, and 800 nm, as determined by examining 20 implementations of the best band search. Selecting bands within the highlighted regions allowed for an improvement over the default Tetracam spectral coverage, mostly through retrieval of Chl and TSS. Attempting to adapt these selected bands to be amenable for implementing other algorithms from literature showed a decrease in retrieval accuracy.
indicating that a user’s desired algorithm should be considered before band selection.

6.2 Future Work

6.2.1 Satellite Retrievals

6.2.1.1 Examination of Spectral Matching Functions

In this work, spectra from imagery are matched to the LUT by examining the least squares distance between spectra (i.e. RMSE). This process has a significant impact on retrieval, as it determines the set of IOPs that will be used in the interpolation and subsequent reporting of the final determined concentrations. As such, it may be possible to determine a function other than RMSE that improves this initial matching. Raqueno (2003) suggested the possibility of weighting spectral bands based on each component’s effect in the region of the spectrum that they measure. It may also be possible to randomly select bands used for fitting, or some different metric to help minimize the issues from noise or atmospheric compensation in the spectra. It may be also of interest to determine an interpolation process that is more efficient (i.e. requires less population of the LUT, and therefore less computation time).

6.2.1.2 Examination of Error due to Unrepresentative IOPs

While this work attempted to examine how variability within the LUT affected retrieval of waterbody components, this could be examined further in future work. Examinations into how different types of error (shifting, scaling, white noise) in the IOPs used for LUT generation affect retrieval could assist in determining better methods for incorporating IOP variability into LUTs. This could be examined through both models and real imagery.

Furthermore this work focused mainly on variability in mass-specific absorption, although variability in scattering phase function and mass-specific scattering has been reported in the literature as well. Future work could focus on examining the effects of scattering variability and how to best incorporate it in the LUT matching process. This would include examining how the scattering phase function of Chl and TSS differ, adding another dimension of analysis for the LUT.
6.2.1.3 Examination of Bloom Composition with Depth

The LUTs modeled for this work assumed component concentrations were homogeneous with depth, and that the waterbodies were optically thick (i.e. infinite depth). As noted by Kutser et al. (2008), this is often not the case and the variation of cyanobacteria concentration with depth can have a large impact on the water spectrum and therefore retrieved concentrations. As such, future implementations of the LUT process should consider examining concentrations with depth, as this may potentially improve retrieval. Furthermore, examinations into including bottom reflectance should allow retrieval in shallow, non-optically deep areas and coastlines.

6.2.1.4 Increased Exploration of High Concentration Component Covariance

A portion of this work was focused on developing and testing a LUT based on reported relations between component concentrations from literature. These reported relations were determined based on sampling that was more frequent in the lower end of the concentration range in regard to blooms (i.e. less than \(50 \text{ mg/m}^3 \) of Chl). As such, it may be useful to better determine how these relations perform at higher concentrations, which may help to inform the design of LUTs.

6.2.1.5 Phytoplankton Function Type Sensing of Cyanobacteria Blooms

Although it was determined that Landsat did not have appropriate spectral coverage to retrieve concentrations of phycocyanin, it may still be possible to identify and monitor cyanobacteria blooms by observing other properties of cyanobacteria cells. It may be possible to differentiate blooms of cyanobacteria from green algae via the increased scattering discussed by Ganf et al. (1989). As such a potential future work would be to examine Landsat’s potential for determining phytoplankton functional types.

6.2.2 UAS Retrievals

While the UAS retrievals shown in this work indicated satisfactory performance, until analysis is performed with more data points it should be considered a proof of study.
Performing this research with a larger focus on the UAS portion could be a fruitful area of study due to the rising interest in the use of UAS in environmental monitoring and the adaptability of UAS spectral coverage. This would require more data points to be collected, and a better standardization for the calibration of data. Furthermore, future efforts should focus on determining methods of image collection and processing to reduce the effects of glint.

6.2.3 Future System Improvements

This study laid out potential spectral coverages for both Landsat and UAS systems that could improve its performance in the spectral LUT matching task. Future work should focus on validating these results through real world data, potentially through resampled in-situ spectrometer measurements, or multispectral and hyperspectral imaging systems on UAS.

Furthermore, in the case of future Landsat missions, examinations considering system noise and quantization should be performed to determine if better radiometric performance would improve retrieval. As this study focus primarily on the use of Landsat for water quality monitoring, similar modeling studies should be performed for other domains such as agriculture or forestry. These studies can allow the USGS and NASA to create meaningful design requirements for a system which performs optimally across applications.
References


References


