

1-1-2013

Instrument and Method Development For Single-Cell Classification Using Fluorescence Imaging Multivariate Optical Computing

Joseph Swanstrom
University of South Carolina

Follow this and additional works at: <https://scholarcommons.sc.edu/etd>

 Part of the [Chemistry Commons](#)

Recommended Citation

Swanstrom, J.(2013). *Instrument and Method Development For Single-Cell Classification Using Fluorescence Imaging Multivariate Optical Computing*. (Doctoral dissertation). Retrieved from <https://scholarcommons.sc.edu/etd/2398>

This Open Access Dissertation is brought to you by Scholar Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact digres@mailbox.sc.edu.

INSTRUMENT AND METHOD DEVELOPMENT FOR SINGLE-CELL CLASSIFICATION
USING FLUORESCENCE IMAGING MULTIVARIATE OPTICAL COMPUTING

by

Joseph Andrew Swanstrom

Bachelor of Science
South Dakota State University, 2007

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Chemistry and Biochemistry

College of Arts and Sciences

University of South Carolina

2013

Accepted by:

Michael Myrick, Major Professor

S. Michael Angle, Committee Chair

Stephen Morgan, Committee Member

Tammi Richardson, Committee Member

Lacy Ford, Vice Provost and Dean of Graduate Studies

© Copyright by Joseph Andrew Swanstrom, 2013
All Rights Reserved.

ACKNOWLEDGMENTS

There are many people I would like to thank for their contributions both directly and indirectly to my career as a graduate student. I would like to acknowledge first and foremost my advisor, Dr. Michael Myrick. He has shared vast amounts of knowledge with all his students, and has been the single most influential figure in my academic life. I also wish to thank my Ph.D committee members Dr. S. Michael Angel, Dr. Stephen Morgan and Dr. Tammi Richardson for their advice and guidance over the past four years.

A great many thanks to the current and former members of the Myrick group; Dr. Michael Simcock for teaching me the art of thin film deposition; Dr. Ryan Priori for his willingness to advise and answer questions regarding designs of multivariate optical elements; Dr. Laura Bruckman and Dr. Megan Pearl for teaching a newbie graduate student the ways of the world; and Shawna Tazik, Stephanie DeJong, Elizabeth Abernathy, Wayne O'Brien, Nick Viole and Kathleen Donaldson for their help and encouragement inside and outside the lab.

The love and support of my wife Jennah has been a solid foundation for my life, and for that I am immensely thankful. Her willingness to leave her family, friends and career in South Dakota to join me in this chapter of my life in South Carolina is a testament to her love and support as my partner in life. I wish to thank my beautiful daughter Lily Jo for bringing so much joy and perspective to my life. I want to thank my parents Dale and Nancy Swanstrom for their guidance, love, support, discipline, work ethic, faith and the many more attributes they have instilled on me over the past 28 years. To everyone that has been so instrumental in my life, thank you all.

ABSTRACT

Multivariate optical computing (MOC) is an all-optical approach of predictive spectroscopy that utilizes multivariate calibration and spectral pattern recognition techniques while operating in a simple filter photometer instrument, removing the need for expensive instrumentation and post-processing of spectral data. This is accomplished with specially designed interference filters called multivariate optical elements (MOEs).

MOC can provide analytical solutions for applications requiring low cost, rugged, and simple to operate instrumentation for use in remote and hazardous environments such as open ocean waters. These instrument specifications are central for developing a method for classifying phytoplankton in their natural environment. Phytoplankton are photosynthetic single cell algae and cyanobacteria that inhabit nearly all natural bodies of water. The size and taxonomic composition of the phytoplankton community structure has global implications on carbon transport.

This dissertation describes the development of a single streak imaging multivariate optical computing (SSIMOC) method for single-cell classification of phytoplankton. The discussion of the SSIMOC method is broken into four main categories: MOE design and fabrication, design and protocol of filter photometer instrumentation, analysis of SSIMOC data images of known cultured phytoplankton, and testing in coastal ocean water at Martha's Vineyard Coastal Observatory. The designing and fabrication of MOEs for traditional predictive spectroscopy is reviewed as well a discussion on modifying these models for applying predictive spectroscopy to classification of phytoplankton. A theoretical model is presented for classifying three species

of phytoplankton and the requirements of a filter photometer needed for making MOC measurements is discussed. The design and experimental protocol of the SSIMOC filter photometer details the components and design of the filter photometer constructed for phytoplankton MOC measurements. The discussion will analyze the sources of variance in the filter photometers measurements of the coccolithophore *Emiliana huxleyi*. Streak deblurring and second generation streak integration algorithms for analyzing the images collected by the SSIMOC filter photometer are discussed. These algorithms are then applied on cultured *Emiliana huxleyi* and *Thalassiosira pseudonana* data to validate the theoretical models the MOEs were designed from. Finally, the application of the SSIMOC method tested on "real world" coastal ocean waters at MVCO will be presented. A comparison of data collected from a cytometer deployed at MVCO, called the FlowCytobot, will be used to determine the size range and phytoplankton detection efficiency of the SSIMOC.

CONTENTS

| | |
|---|-----|
| ACKNOWLEDGMENTS | iii |
| ABSTRACT | iv |
| LIST OF TABLES | ix |
| LIST OF FIGURES | xii |
| CHAPTER 1 INTRODUCTION | 1 |
| 1.1 Global Impact of Phytoplankton | 3 |
| 1.2 Photosynthesis and Pigments | 4 |
| 1.3 Method for Phytoplankton Classification | 4 |
| 1.4 Multivariate Analysis and Optical Computing | 7 |
| BIBLIOGRAPHY | 11 |
| CHAPTER 2 DESIGN AND THEORETICAL PERFORMANCE OF MULTIVARI- ATE OPTICAL ELEMENTS FOR TAXONOMIC CLASSIFICA- TION OF PHYTOPLANKTON | 20 |
| 2.1 Introduction | 20 |
| 2.2 Experimental | 21 |
| 2.3 Results and Discussion | 31 |
| 2.4 Conclusions | 48 |
| BIBLIOGRAPHY | 49 |

| | | |
|-----------|---|-----|
| CHAPTER 3 | DESIGN AND EXPERIMENTAL PROTOCOL OF A SHIPBOARD | |
| | STREAK IMAGING MULTIVARIATE OPTICAL COMPUTING | |
| | PHOTOMETER (SSIMOC) | 52 |
| 3.1 | Introduction | 52 |
| 3.2 | Experimental | 55 |
| 3.3 | Results and Discussion | 67 |
| 3.4 | Conclusions | 81 |
| | BIBLIOGRAPHY | 82 |
| CHAPTER 4 | SIMOC IMAGE ANALYSIS AND DEBLURRING ALGORITHM . . | 86 |
| 4.1 | Introduction | 86 |
| 4.2 | Experimental | 87 |
| 4.3 | Results and Discussion | 96 |
| 4.4 | Conclusion | 99 |
| | BIBLIOGRAPHY | 100 |
| CHAPTER 5 | DETERMINING PHYTOPLANKTON SIZE FROM FLUORESCENCE | |
| | IMAGES | 103 |
| 5.1 | Introduction | 103 |
| 5.2 | Experimental | 104 |
| 5.3 | Results and Discussion | 106 |
| 5.4 | Conclusion | 110 |
| | BIBLIOGRAPHY | 113 |
| CHAPTER 6 | SSIMOC FIELD TESTING | 115 |
| 6.1 | Introduction | 115 |
| 6.2 | Experimental | 117 |

| | | |
|-----|--|-----|
| 6.3 | Results and Discussion | 120 |
| 6.4 | Conclusion | 123 |
| | BIBLIOGRAPHY | 125 |
| | BIBLIOGRAPHY | 129 |
| | APPENDIX A SNR MATLAB CODE | 143 |
| | APPENDIX B SIMOCv2.0 MATLAB CODE | 174 |

LIST OF TABLES

| | | |
|-----------|---|----|
| Table 2.1 | <p>(a) Design suite results. The theoretical characteristics of the designs selected for fabrication are shown. SEC is the standard error of calibration; sensitivity is given in % change of the ratio of fluorescence excited through the MOE to that excited through a neutral density for each unit off LD1 score; Gain is a measure of the slope of the relationship between the ratio and score. Note that one is chosen with positive gain, and the other with negative gain. (b) Calculated results after final layer deposited. This shows the same characteristics as part (a) except with changes to the design following fabrication of the optical elements.</p> | 38 |
| Table 2.2 | <p>(a) Layer thicknesses for Design 212. This shows the designed thicknesses of the first 5 layers that composed Design 212 after design (initial) and after fabrication was completed (final). (b) Layer thicknesses for Design 23551. This shows the designed thicknesses of 5 layers that compose design 23551 after the design stage (initial) and after fabrication was completed (final).</p> | 39 |

| | | |
|-----------|--|----|
| Table 3.1 | Streak statistics. For each filter wheel rotation frequency, the average integrated area of 90 streaks (9 streaks each for 10 <i>E. huxleyi</i> tracks) is given. s.d. (inter) is the standard deviation of the average streak intensities for 10 phytoplankton tracks. s.d. (intra) is the average of the standard deviations of 9 streaks in each track. SNR (intra) is the average signal to noise ratio. More details are found in the text. | 68 |
| Table 3.2 | Background statistics. For each filter wheel frequency, the average integrated area of 90 regions of equivalent size to those in Table I, but chosen from a nearby region of background near each phytoplankton track, is provided in column 2. Column 3 gives the standard deviation of the average streak-area intensities for ten background "tracks". Column 4 gives the average of the standard deviations of 9 streak areas in 10 background "tracks". | 68 |
| Table 4.1 | Mean (\bar{x}), standard deviation of the mean ($\sigma_{\bar{x}}$), standard deviation (σ) and the number of samples (N) for <i>E. huxleyi</i> and <i>T. pseudonana</i> | 98 |

| | | |
|-----------|--|-----|
| Table 6.1 | <p>The rows in column 1 are the month and years of the cruises that the SSIMOC participated in to MVCO. Column 2 is the MOE filter sets that were used on each cruise described in the text. D is a 0.3 optical density filters used a reference and the subscripts refers to the Linear Discriminant vector each MOE is designed to mimic as described in chapter 2. Column 3 is the total number of data images collected on each cruise. Column 4 is the number of fluorescent particles detected by the SSIMOC. Column 5 is the number of tracks detected that have a high enough SNR for MOC calculations. Column 6 is the calculated cell density based on the total number of fluorescent particles detected and the total number of mL analyzed.</p> | 121 |
|-----------|--|-----|

LIST OF FIGURES

| | | |
|------------|--|----|
| Figure 2.1 | <p>Fluorescence excitation spectra of individual cells of three species.</p> <p>The area-normalized fluorescence excitation spectra of 77 <i>E. huxleyi</i> (black), 46 <i>T. pseudonana</i> (yellow) and 76 <i>Synechococcus sp.</i> (pink) cells excited in 2 nm increments over the range of 350 to 650 nm with a spectral resolution of ~ 10 nm. The inset shows the same spectra renormalized in the wavelength range for best classification, 550 to 610 nm.</p> | 23 |
| Figure 2.2 | <p>Schematic representation of an MOC instrument. A white light source (a) is filtered (b) to give a range of wavelengths suited to classifying phytoplankton. An MOE designed for that purpose is placed into the beam, alternating with a reference filter, (REF). A dichroic beamsplitter reflects the excitation toward a microscope objective (d) and then to a phytoplankter (e). Fluorescence from the phytoplankter cell is returned to a detector system (f) which integrates the signals and determines the filter type (MOE vs. REF) for each signal.</p> | 26 |
| Figure 2.3 | <p>Optical functions of the components discussed in figure 2.2 used for designing MOEs in this report. a) Irradiance of Xe arc lamp, b) percent transmission of broadband bandpass filter, c) percent transmission of 60X 0.7 NA plan-fluor objective, d) percent reflectance of the beamsplitter at 45 degrees. Details on components are given in the text.</p> | 28 |

| | | |
|------------|--|----|
| Figure 2.4 | Linear discriminant analysis (LDA) results based on inset data in Figure 2.1. (a) Plot of <i>E. huxleyi</i> (circles), <i>T. pseudonana</i> (squares) and <i>Synechococcus sp.</i> (triangles) scores of the individual calibration spectra along linear discriminant 1 (LD1) and 2 (LD2). (b) Spectral profile of the loadings of LD1 in the wavelength range specified by the results of LDA, 550 to 610 nm. | 33 |
| Figure 2.5 | Sensitivity vs. standard error of calibration (SEC) for 823 unique designs meeting selection criteria. Sensitivity is defined as the percentage change in MOE/REF ratio per unit change in LD1 score. SEC is the standard error of calibration in units of LD1 score. | 36 |
| Figure 2.6 | Transmission curves of MOE 212 (Dotted line) and MOE 23551 (solid line) over the wavelength range of 540-620 nm. | 37 |
| Figure 2.7 | Distribution (circles) and multi-Poisson distribution model (line) of design duplicates of 823 unique MOE designs observed in a run of 120,000 cycles. Four populations of designs were modeled with 19.1 ± 1.0 , 9.5 ± 0.5 , 3.4 ± 0.4 , and 0.44 ± 0.41 repeats on average. The four populations have 64 ± 13 , 242 ± 26 , 400 ± 50 , and 360 ± 150 members, respectively, for a total accessible design population of 1060 ± 160 designs. The top axis shows the residuals between observation and model, with error bars representing the square root of the modeled number of occurrences. The fitting procedure is described in the text. Reduced χ^2 for the fit was 1.16. | 42 |
| Figure 2.8 | Theoretical results of classification performance of designed MOE applied to calibration data. a) Ratio of a single MOE to a 0.3 OD neutral density filter b) Ratio of MOE 23551 to MOE 212. | 45 |

Figure 3.1 The rays travel from left (component 1) to right (component 7), originating at the source and pass through the optical components to a Nikon 60x objective. Component 1 is a 75 W Xe lamp (Sci 200, Sciencetech, Ontario, Canada); component 2 is a 1-inch f/1 biconvex lens (KBX046, Newport, Irvine, CA) ; component 3 is a 1-inch f/6 planoconvex lens (KPX100, Newport, Irvine, CA) ; component 4 is a 2-inch f/3 planoconvex field lens (KPX199, Newport, Irvine, CA) ; component 5 is another f/3 planoconvex lens (KPX199, Newport, Irvine, CA); component 6 is a 2-inch diameter beam splitter (p0643dr1p, Omega Optical, Brattleboro, VT) positioned at a 45° angle to the incident light to reflect it onto component 7 which is a 60X 0.70 NA Nikon infinity-corrected objective. 56

Figure 3.2 Schematic of fluorescence imaging photometer. Ex, the excitation source is a Sciencetech 500-200 75W Xe Arc lamp and K is a ~ 2.5 mm aperture. The lenses, in order of excitation to emission, are as follows: L1 is a 1 in. diameter biconvex lens with a focal length of 1 in., L2 and L3 are 2 in. diameter planoconvex lenses with 6 in. focal lengths, and L4 is a 2 in. diameter biconvex lens with a 4 in. focal length. The filters are as follows: F1 is a 550-610nm Chroma bandpass (HQ580/60), F2 is a 681 ± 5 nm Omega Optical (HBP10) bandpass, F3 is an OG 530 Schott long pass, and F4 is an OG 590 Schott long pass. BS is an Omega Optical 2 in. diameter dichroic beam-splitter (640drlp). The filter wheel, W, is a Thorlabs FW103, powered by a Faulhaber 2057B brushless motor. The objective, M, is a Nikon, Plan Fluor, 60x magnification, 0.70 NA objective. The sample is contained in a beaker, S, and is pulled the flow cell, C, by a Cole-Parmer 75211-10 gear pump, G, to a waste beaker, R. The fluorescence emitted by the sample is passed back through the objective, M, and beamsplitter, BS, and imaged onto a back-illuminated Princeton PIXIS 1024B CCD. 57

Figure 3.3 Measurement scheme for collecting data using fluorescence imaging photometer. (A) The x-axis represents the exposure time of the camera in seconds, where at $x=0$ the shutter opens and $x=1$ the shutter closes. The top graphic ("CCD Integration") illustrates the CCD integration time extends well before and after the timeline shown. The center graphic ("Transit time") illustrates the time a phytoplankter takes to pass through the region being imaged by the CCD. The lower graphic ("FW") illustrates the position of the filter wheel in respect to the excitation beam as a function of time, with each filter opening being shown for filters 1-5 and a blocked opening marked "B". This timing diagram is for a phytoplankter that enters the image volume 186 milliseconds after the camera shutter opens, at a moment when filter 3 of the filter wheel is starting to line up with the excitation beam. During the approximately 75 milliseconds required for the phytoplankton to traverse the image volume, the filter wheel completes a little over 1.5 rotations. (B) An example of the tracks that two individual phytoplankton cells would produce in the CCD image. The track surrounded by a dashed box is aligned and scaled relative to the graphics in part A to emphasize the relation between the observed track and the timing diagram. The track below the dashed box is an example of a phytoplankter that has entered the image area at a different time and is unrelated to the track in the dashed box.

Figure 3.4 Example image from the fluorescence imaging photometer. (A) The image contains a track of *E. huxleyi*'s streaks where there are no filters placed in the filter wheel. Regions of interest for each streak in the track are illustrated (black outline), along with corresponding regions of the same size just below the phytoplankton track (gray outline) that were used for image background measurements. A second *E. huxleyi* track appears in the middle of the image but is unrelated to the *E. huxleyi* track of interest. (B) Plot of summed rows from example phytoplankton track and reference track areas shown in A. This plot shows the summed fluorescence intensity of the rows between the upper and lower row boundaries of the given track. 66

Figure 3.5 Average streak integrated fluorescence intensity as function of inverse filter wheel frequency. At each frequency, 90 fluorescence streaks were averaged (9 each for 10 phytoplankter tracks). Error bars are 95% confidence intervals of the mean. The fitted line has a slope of $(1.91 \pm 0.21) \times 10^5$ counts/sec and an intercept of $(2.6 \pm 1.0) \times 10^3$ counts. When forced through zero, the slope of the best fit line is $(2.09 \pm 0.03) \times 10^5$ counts/sec, where all error limits are estimated standard errors. The normalized value of χ^2 for the plot shown is approximately 0.1. 70

| | | |
|------------|---|----|
| Figure 3.6 | (A) Single-streak measurement variability averaged for 10 <i>E. huxleyi</i> and 10 equivalent background regions. Open circles represent <i>E. huxleyi</i> streaks and solid circles are equivalent background streaks. Error bars are 95% confidence intervals for the single-streak standard deviations. (B) Variance of single- <i>E. huxleyi</i> fluorescence signals, correcting for background variance, based on top figure. The fit is of the form $y=ax$, where $a = 45.2$ counts. | 73 |
| Figure 4.1 | A) Preprocessed data image. There are 3 tracks of the coccolithophore <i>Emiliania huxleyi</i> visible in this image. The solid black rectangle indicates the region of the image defined by the algorithm as a track and is used as an example in the data analysis below. B) A plot of the corrected standard deviation along the rows for each column of the image in Fig. 1A. The solid black line corresponds to the baseline corrected column standard deviation and the dotted line is the calculated threshold. | 90 |
| Figure 4.2 | The image area of a track presented in Fig 1A. 256 rows by 11 columns. A) Red-Green-Blue (RGB) image of the track area. B) Gray scale image. C) Enhanced gray scale image D) Binary image of all the objects found in the image E) Binary image with the perimeter of each object filled out F) Binary image with the interior of each object filled in G) Binary image with where the object of non-streaks removed | 93 |

| | | |
|------------|--|-----|
| Figure 4.3 | Diagram of of the filter wheel with MOE placements and their assignments on a theoretical track mask. The wheel is turning in a counter clock wise direction and vertical the flow direction in the image would result in the first streak after the blank to be assigned to MOE1-. | 94 |
| Figure 4.4 | A) Corrected data image of a 6 micron microsphere that is out of focus B) Deblurred image formed by deconvolving the image in A. C) Corrected data image of a 6 micron microsphere that is in focus. D) Deblurred image formed by deconvolving the image in C. | 95 |
| Figure 4.5 | Distribution of measured ratios for the coccolithophore <i>Emil- iania huxleyi</i> (black, 703 ratios) and the diatom <i>Thalassiosira pseudonana</i> (yellow, 858 ratios). The inset is the frequency dis- tribution of the measured ratios where 81 outlier ratios were visually inspected and 65 were removed due to defects in the image that were missed by the algorithm. | 97 |
| Figure 5.1 | A) A picture of the SSIMOC filter photometer on a lab bench on the RV Tioga. B) A picture of Martha's Vineyard Coastal Observatory tower, positioned about one mile off the eastern coast of Martha's Vinyard, MA. | 105 |

| | | |
|------------|--|-----|
| Figure 5.2 | A) The histogram plot in red of the log of the cell density (cell counts/mL) vs log of the chlorophyll a fluorescence (volts) collected during a July 2012 cruise from the FlowCytobot. The histogram plot outlined in black is of the log of the cell density (cell counts/mL) vs log of the chlorophyll a fluorescence (photon counts) collected during a July 2012 cruise from the SSIMOC filter photometer. B) Same histograms as in (A) except that the SSIMOC plot is shifted to the left 4.95 log units and the FlowCytobot plot is shifted up 0.1191 log units. This is done to visually determine if the plots have overlapping features, which are visible in (B) between -0.6 and 0.9 volt log units. | 108 |
| Figure 5.3 | The bottom plot is of the SSIMOC counts vs the FlowCytobot counts/mL for each bin in the histograms from Figure 2, and the the red line represents the best fit line with a χ^2 of 0.0252. The error bars represent 95% confidence intervals for the SSIMOC. The top plot represents the distance between the fit and each point. | 110 |
| Figure 5.4 | A) Histogram plots for the corrected SSIMOC volume (red) and converted FlowCytobot chlorophyll a intensity to photon counts (blue) B) Plot of the cell counts per mL efficiency of the SSIMOC method compared to the FlowCytobot where a value of one indicates a perfect efficiency. | 111 |
| Figure 6.1 | Left: Example of data image collected on SSIMOC photometer used at MVCO. Right: Example of MOEs in the filter | 117 |

| | | |
|------------|--|-----|
| Figure 6.2 | Classification distribution of 200 phytoplankton tracks collected from MVCO in March of 2012. The dashed regions represent 95% confidence intervals calculated from measurements of cultured dinoflagellate (red), diatom (green), and haptophyte (black) when interrogated through the 3 class MOEs. The solid circles represent the MVCO natural population recorded using the 3 class set MOEs, the size of the circle represents it's relative size (largest circle represents 44 microns and smallest circle represents less than 3 microns) and color gradient represents average fluorescence intensity over all 5 streaks. Unknown samples that fall within the 95% confidence limits would be classified as either a dinoflagellate, diatom or haptophyte. Unknown samples that fall outside these bounds would be classified as not belonging to any of these three class. | 122 |
| Figure 6.3 | | 123 |

CHAPTER 1

INTRODUCTION

The development of new technologies and methodologies is a difficult and risky endeavor. Many environmental and ecological projects are essential for understanding and planning future and remediation of impacts of anthropologic pollution. With no immediate financial gains needed for industrial investment, the support for these projects often come from private foundations and federal and state agencies. The scale and magnitude of the work presented in this dissertation details the advancements made in the area of optical computing for classification and is part of a larger project aimed toward creating a method for long term characterizing the phytoplankton community in open ocean waters.

The first chapter reviews the important role phytoplankton play in the global carbon cycle and the impact of mapping the size and taxonomic composition of the phytoplankton community structure has on future of the environment. Here we will examine current techniques for phytoplankton classification and highlight the difficulties associated with such a complicated task. This first chapter also attempts to give a brief overview of predictive spectroscopy and how multivariate optical computing is utilized as an all optical method for applying predictive spectroscopy both in situ and real time by use of specially designed interference filters called multivariate optical elements (MOEs).

Chapter two will show that the full spectrum information of single phytoplankton cells can be used to develop special optical elements, enabling rapid measurements based on fluorescence excitation spectroscopy. The results show the theoretical

classification power of the MOEs as well as setting the Signal-to-Noise ratio (SNR) specification requirements of an imaging photometer needed in order to realize their application on real phytoplankton samples.

Chapter three characterizes the performance of a fluorescence imaging photometer designed and constructed to use the MOEs from chapter two for making classification measurements of real phytoplankton. The SNR of the photometer meets the specifications described in chapter two but is much lower than expected if limited by photon counting noise. A thorough evaluation of possible sources from mechanical and electrical instrumental sources is described and the hypothesis that the sources of noise lies with the inherent variability of the test species *E. huxleyi* is presented.

Chapter four evaluates an algorithm developed in the MATLAB coding environment for analyzing the fluorescent images produced by the imaging photometer described in chapter three and validates the theoretical model in chapter two. The results show that the shipboard streak imaging multivariate optical computing (SSIMOC) method can successfully classify two species of similarly pigmented phytoplankton.

Chapter five evaluates the SSIMOC efficiency at identifying Chlorophyll a containing fluorescence particles in coastal ocean waters at the Martha's Vineyard Coastal Observatory (MVCO). A standardized flow cytometry, the FlowCytobot, is used as a validation method for comparison and correction.

Chapter six evaluates the operation and results of the SSIMOC instrument and method on board the RV Tioga on five separate research cruises at the Martha's Vineyard Coastal Observatory (MVCO) from March, 2011 to July, 2012. A description of the operation and results show that the SSIMOC is able to operate on a research vessel and collect and analyze real ocean samples containing natural phytoplankton.

1.1 GLOBAL IMPACT OF PHYTOPLANKTON

Phytoplankton are single cell autotrophs that inhabit virtually all natural water bodies on earth.[35, 20] These microscopic single cell organisms have an enormous impact on the global carbon cycle and oxygen production through their ability to fix atmospheric CO_2 through photosynthesis.[31, 35, 20, 5, 42] The extent of the fixed carbon that is effectively removed from the carbon cycle by sinking to the ocean bottom is tied to species that make up the community structure.[22, 9, 5]

Marine biologists are inherently interested in the dynamics of the phytoplankton community structure, but due to the increasing levels of atmospheric carbon dioxide from anthropologic sources and its effects of global warming, the greater scientific community is growing more interested as well.[18, 33, 50, 48] Characterizing the phytoplankton community structure is important to obtain a better understanding of how the oceanic ecosystem is responding to global climate changes.[5, 18, 33, 50, 55, 12]

Phytoplankton, being the oceans net primary producers, impact more than the carbon cycle. They become problematic when their ability to grow rapidly causes harmful blooms that are sometimes toxic or cause oxygen depletion.[1, 39] This is a major concern along coastal regions where river run off can bring an influx of nutrients that spark these blooms.[43, 39] The blooms have the potential to cause kill zones over large regions, wrecking havoc on regional economies.[32] In addition to kill zones, a more sinister problem occurs when certain blooms do not kill shell fish but rather cause increase levels of toxins to build in the shell fish and harm humans, *i.e.* diarrhetic shellfish poisoning (DSP) caused by blooms of *Prorocentrum lima*. [2]

To monitor and better understand the phytoplankton community structure, it is advantageous to have a deployable, continuous and rugged method capable of discriminating species or classes of species in open and costal waters.[3, 8, 56].

1.2 PHOTOSYNTHESIS AND PIGMENTS

Photosynthesis is the process by which autotrophic organisms, primarily plants, use solar radiation, carbon dioxide and water to synthesize organic carbon and oxygen.[48, 42] The most important molecule in this photosynthetic process is Chlorophyll *a*, a pigment at the center of the photosynthetic reaction center with a fluorescence emission maxima at 680 nm.[60, 63] Chlorophyll *a* is not only critical to photosynthesis in the reaction center, but also has a main absorption peak at 425 nm for light harvesting applications as part of the antenna pigment system.[46] While chlorophyll *a* is the most important pigment for photosynthesis, phytoplankton also have accessory pigments with absorption maxima in different regions to maximize the solar spectrum. These include chlorophyll *b*, *c*, *d*, carotenes and phycobilins. [46] Almost every species of phytoplankton contains chlorophyll *a*, but the compositions of the accessory antenna pigments vary between species. The varying concentrations of pigments is an ideal discriminating characteristic that can be probed using excitation fluorescence spectroscopy.[64, 63, 11]

1.3 METHOD FOR PHYTOPLANKTON CLASSIFICATION

Historically, it has been challenging to understand the temporal and spatial variations of oceanic phytoplankton community structure because the available techniques for phytoplankton counting and classification have not been suited to high frequency open ocean measurements.

The gold standard for community structure measurement is traditional microscopy of fixed and stained samples by a skilled phytoplankton taxonomist who manually counts and classifies the phytoplankton. This task is both difficult and time consuming, as the morphological differences between some taxonomic classes are minute. Culverhouse et al. show that experts, who are routinely involved in classification,

have accuracies in the range 84-95%.[14] Unfortunately this method lacks the high frequency sampling capacity needed for community structure monitoring.[59, 6]

Flow cytometry alone, while suited to high frequency measurement, has not been reported as an acceptable method for phytoplankton community structure. Likewise, classification methods using solely morphological information are often confounded by similarities in morphology between phytoplankton species. Uhlmann *et al.* were the first to report automated classification of phytoplankton cells from video images, but no statistical summary was provided.[57] Since then, advancements in morphological classification methods have combined microscopy with flow cytometry and resulted in great progress.[14, 53, 19, 45, 47] Using the Video Plankton Recorder (VPR) Davis *et al.* achieved accuracies between 45 and 91% in identifying individual taxa.[16] Culverhouse *et al.* developed the Harmful Algal Bloom (HAB) Buoy for both zooplankton and phytoplankton identification and report identification rates of 80% for phytoplankton species.[15] Sosik *et al.* have developed perhaps the most innovative method using cytometry, fluorescence and image analysis with the FlowCytobot, where 88% accuracy between 22 categories is reported.[53]

CHEMTAX, a high performance liquid chromatography method, has been used to identify the relative concentrations of taxonomic species in bulk monocultures and mixed cultures.[36] CHEMTAX is useful in determining pigment concentration in bulk samples for calibration or validation, but is not suited to in-situ measurements.[44]

Remote sensing methods such as satellite imagery of chlorophyll *a* and phycoerythrin have also been used for monitoring phytoplankton.[35, 29] Clarke *et al.* first demonstrated that backscattered light measurements from aircraft could be used to estimate total chlorophyll concentrations in the water.[13] Satellite images that isolate chlorophyll *a* fluorescence at a band around 680 nm are useful in selectively targeting photosynthetic organisms and estimating bulk chlorophyll *a* concentrations over large areas, but are limited or unable to discriminate the speciation of the source of the

fluorescence emission.[49, 58]

Despite the advancements described above, a rugged and deployable method suitable for open ocean monitoring is still desired.[17, 24, 26, 40, 27] In-situ fluorescence excitation spectroscopy provides an alternative approach. Fluorescence excitation spectroscopy uses spectral characteristics of a phytoplankton cell that is independent of morphology.[63] Beutler et al. developed an in situ method using light emitting diodes (LEDs) to selectively excite a bulk sample at 5 wavelength bands and recorded the chlorophyll a emission for each, but provided no statistical analysis.[7] The potential of this instrument for bulk in situ fluorometric measurement of phytoplankton community structure has recently been described.[44]

The Myrick group is exploring an approach to phytoplankton classification that combines some of the power of imaging with fluorescence excitation spectroscopy to classify phytoplankton. This thesis will focus on the automatic analysis of spectroscopic content in images from a fluorescence imaging photometer and show that full-spectrum fluorescence excitation spectroscopy can be used as a basis for distinguishing at least limited classes of phytoplankton in cultures. The Myrick group has, through publications and content in this thesis, shown that optical computing combined with a fluorescence imaging is able to differentiate phytoplankton by measuring experimental fluorescence ratios and comparing them to those calculated from an optical model from chapter 2 as well as the classification accuracy and ratio variability for each class using two similarly pigmented phytoplankton species: the coccolithophore *Emiliana huxleyi* (*E. huxleyi*) and the diatom *Thalassiosira pseudonana* (*T. pseudonana*).[41] We found that the measured MOE ratio for *E. huxleyi* differed from the theoretical MOE ratio (chapter 1) by -3%, and that of *T. pseudonana* differed by +0.1%, with a measured ratio difference of 0.281 versus a theoretical ratio difference of 0.251.[41] The distribution of individual cell ratios was well explained by the signal to noise ratio (SNR) of the instrument reported in chapter 3 and no misclassifications were

seen for 853 cells of two species analyzed.

1.4 MULTIVARIATE ANALYSIS AND OPTICAL COMPUTING

Multivariate Calibration and Discriminant Analysis

Principal Component Analysis

The measured spectral variables of a complex data set can be transformed using linear multivariate models such as principal component analysis (PCA) into orthogonal basis vectors called principle components (PCs).[23, 34] These PCs model the statistical significant variations in the data along with noise, the first PC describes the most significance variations in the data with each successive PC containing less until eventually describing only noise. The PCs containing the significant variations can then be used for data analysis while the PCs containing mostly noise are not, attempting to remove the noise in further analysis of the data. Or, reduce the dimensionality of the data set while retaining the significant variations within the data set.[11]

A common and popular method for calculating PCs of a spectral data set is with singular value decomposition (SVD), a predefined algorithm in the Matlab programming environment. The fluorescence excitation spectra described in the next two chapters, and in a recent Myrick group publication[11], were decomposed using equation 1.1:

$$A = USV^T \quad (1.1)$$

where the columns of U contain the columns PC scores of A , S contains the square root of the PCs of $A^T A$ and V^T contains the row PC loadings of A . [10, 4] The greatest variation of the data set A is described by the first row vector in V^T , with each following row describing less. The PC scores are computed by multiplying the U matrix by the S matrix.

Linear Discriminant Analysis

Linear discriminant analysis (LDA) was first developed by Ronald Fisher in 1936 in his work with taxonomic classification in genetic research.[21] LDA is a form of supervised classification, meaning that a known set of known parameters are used in the model, whereas unsupervised methods do not and are useful in discovering patterns in data sets.[23] LDA is used to classify samples into separate classes based on known variables of each sample, such as their fluorescence excitation spectrum. LDA is a common technique for classification that has been used in image pattern methods for facial recognition, microarray gene expression, and pattern recognition.[37, 61, 30, 28, 51, 65, 52]

Unlike PCA, where calculated vectors describe the data, LDA calculates vectors that describe the features in the data matrix that show the greatest difference between classes. The independent features of the data set are used to calculate linear combinations of these features that maximize the variance between classes while minimizing the variance within each class.[62, 23] A requirement of LDA is that the number of samples must be larger than the number of variables. When a data set has a large number of independent variables, such as a fluorescence excitation spectrum, PCA can be used to reduce the number of variables to the number of significant PCs in the data set. [10, 62]

The Fisher ratio maximized in LDA is given by 1.2:

$$F = \frac{\sigma_{between}^2}{\sigma_{within}^2} \quad (1.2)$$

where σ^2 is the variance, shows Fisher Ratio that LDA algorithms maximize. The greater this ratio, the more significant the discrimination function. [21]

Optical Computing

Traditional multivariate calibration techniques such as principle component regression (PCR), partial least squares (PLS) and linear discriminant regression (LDR) extract spectral patterns from a digitized spectrum obtained using a wavelength selection spectrometer or other method to predict a pure component spectral variations that are correlated to a digitized spectral vector. A regression vector is then calculated from a data set of known chemical or physical properties (i.e. concentration, particle size, ect.) and correlated with the magnitude of the dot product between the calculated vector and the spectrum of the known sample. Using the regression vector, a predictive model is generated for determining the properties of interest of unknown samples.

Multivariate optical computing (MOC) is a predictive spectroscopic technique that combines the data collection and processing steps of a traditional multivariate analysis into a single and instantaneous step.[54] It achieves this by applying an all optical technique to masking or filtering part of the optical path with a specially designed interference filter called a multivariate optical element (MOE). This approach drastically lessens the complexity of the instrumentation required for making the measurements from a powerful desecrate wavelength spectrometer to a simple filter photometer instrument with little to no moving components.

Multivariate Optical Elements

MOEs are specially designed and fabricated interference filters that perform the equivalent dot product operation, as described in PCA and LDA above, by filtering the optical train of a measurement with an optical filter with a specific transmission function.[38] Each MOE filter design is analyze specific and intended to replace the need for a conventional spectroscopic instrumentation and computer processing used for multivariate calibration and prediction. It allows for an optical prediction of a

chemical or physical prediction of an unknown sample to be made without the collection of a spectrum, but rather a single optical response.

The spectral function of an MOE is generated by randomly seeding (i.e., initial guesses at the structure of a MOE) the thickness of thin film stacks that are then refined.[25] In operation, the random seed points for the optimization are constrained in several ways. The materials from which the layers will be designed are predefined as a binary pair of high and low-index materials (e.g. Nb_2O_5 and SiO_2). From this point, the algorithm seeks to optimize the thicknesses of the layers in the film stack of the interference filter so as to minimize the standard error of calibration (SEC) for prediction. The design process is repeated many times to form a population of optimized MOE designs with different characteristics.

BIBLIOGRAPHY

- [1] P. Ajanik, B. Steve, K. Martin, S. Peter, G. Webster, and L Armand. The risk of harmful algal blooms (habs) in the oyster-growing estuaries of new south wales, australia. *Environmental monitoring and assessment*, 185:5295–526, 2013.
- [2] D. Baden, L.E. Fleming, and J.A Bean. *Chapter: Marine Toxins. in: Handbook of Clinical Neurology: Intoxications of the Nervous System Part II. Natural Toxins and Drugs*. Elsevier Press, 1995.
- [3] C. Barnes, X. Irigoien, J. A. A. De Oliveira, D. Maxwell, and S. Jennings. Predicting marine phytoplankton community size structure from empirical relationships with remotely sensed variables. *Journal of Plankton Research*, 33(1): 13–24, 2011.
- [4] K. R. Beebe, R. J. Pell, and M.B. Seasholtz. *Chemometrics: A Practical Guide*. John Wiley and Sons, New York, 1998.
- [5] M. J. Behrenfeld, R. T. O’Malley, D. A. Siegel, C. R. McClain, J. L. Sarmiento, G. C. Feldman, A. J. Milligan, P. G. Falkowski, R. M. Letelier, and E. S. Boss. Climate-driven trends in contemporary ocean productivity. *Nature*, 444(7120): 752–755, 2006.
- [6] M. C. Benfield, P. Grosjean, P. F. Culverhouse, X. Irigoien, M. E. Sieracki, A. Lopez-Urrutia, H. G. Dam, Q. Hu, C. S. Davis, A. Hansen, C. H. Pilskaln, E. M. Riseman, H. Schultz, P. E. Utgoff, and G. Gorsky. Rapid research on automated plankton identification. *Oceanography*, 20(2):172–187, 2007.

- [7] M. Beutler, K. H. Wiltshire, B. Meyer, C. Moldaenke, C. Luring, M. Meyerhofer, U. P. Hansen, and H. Dau. A fluorometric method for the differentiation of algal populations in vivo and in situ. *Photosynthesis Research*, 72(1):39–53, 2002.
- [8] R. R. Bidigare and M. E. Ondrusek. Spatial and temporal variability of phytoplankton pigment distributions in the central equatorial pacific ocean. *Deep-Sea Research Part Ii-Topical Studies in Oceanography*, 43(4-6):809–833, 1996.
- [9] P. W. Boyd, S. C. Doney, R. Strzepek, J. Dusenberry, K. Lindsay, and I. Fung. Climate-mediated changes to mixed-layer properties in the southern ocean: assessing the phytoplankton response. *Biogeosciences*, 5(3):847–864, 2008. 321KK Times Cited:16 Cited References Count:106.
- [10] R. G. Brereton. *Chemometrics for Pattern Recognition: A Practical Guide*. John Wiley and Sons, Chichester, UK, 2009.
- [11] L. S. Bruckman, T. L. Richardson, J. A. Swanstrom, K. A. Donaldson, M. Allora, T. J. Shaw, and M. L. Myrick. Linear discriminant analysis of single-cell fluorescence excitation spectra of five phytoplankton species. *Applied Spectroscopy*, 66(1):60–65, 2012.
- [12] F.P. Chavez, K.R. Buck, and R.T. Barber. Phytoplankton taxa in relation to primary production in the equatorial pacific. *Deep-Sea Research Part A-Oceanographic Research Papers*, 37:1733–1752, 1990.
- [13] G.L. Clarke, G.C. Ewing, and C.J. Lorenzen. Spectra of backscattered light from sea obtained from aircraft as a measure of chlorophyll concentration. *SCIENCE*, 167:1119, 1970.
- [14] P. F. Culverhouse, R. Williams, B. Reguera, V. Herry, and S. Gonzalez-Gil. Do experts make mistakes? a comparison of human and machine identification of dinoflagellates. *Marine Ecology-Progress Series*, 247:17–25, 2003.

- [15] P. F. Culverhouse, R. Williams, B. Simpson, C. Gallienne, B. Reguera, M. Cabrini, S. Fonda-Umani, T. Parisini, F. A. Pellegrino, Y. Pazos, H. Wang, L. Escalera, A. Morono, M. Hensey, J. Silke, A. Pellegrini, D. Thomas, D. James, M. A. Longa, S. Kennedy, and G. del Punta. Hab buoy: a new instrument for in situ monitoring and early warning of harmful algal bloom events. *African Journal of Marine Science*, 28(2):245–250, 2006.
- [16] C. S. Davis, Q. Hu, S. M. Gallager, X. Tang, and C. J. Ashjian. Real-time observation of taxa-specific plankton distributions: an optical sampling method. *Marine Ecology-Progress Series*, 284:77–96, 2004.
- [17] C. S. Davis, F. T. Thwaites, S. M. Gallager, and Q. Hu. A three-axis fast-tow digital video plankton recorder for rapid surveys of plankton taxa and hydrography. *Limnology and Oceanography-Methods*, 3:59–74, 2005.
- [18] S. J. Davis, K. Caldeira, and J. D. Mathews. Future co2 emissions and climate change from existing energy infrastructure. *Science*, 329:1330–1333, 2010.
- [19] K. V. Embleton, C. E. Gibson, and S. I. Heaney. Automated counting of phytoplankton by pattern recognition: a comparison with a manual counting method. *Journal of Plankton Research*, 25(6):669–681, 2003.
- [20] P. G. Falkowski and J. A. Raven. *Aquatic Photosynthesis*. Princeton University Press, Princeton, 2nd edition, 2007.
- [21] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Ann. Eugenics*, 7:179–188, 1936.
- [22] M. H. Forget, C. Fuentes-Yaco, S. Sathyendranath, T. Platt, J. Pommier, and E. Devred. Computation of primary production from remote sensing of ocean colour at the northwestern atlantic c-solas lagrangian site. *Marine*

Ecology-Progress Series, 352:27–38, 2007. 257JQ Times Cited:2 Cited References Count:35.

- [23] P. Gemperline. *Practical Guide to Chemometrics*. CRC Press, Boca Raton, 2006.
- [24] G. Gorsky, P. Guilbert, and E. Valenta. The autonomous image analyzer - enumeration, measurement and identification of marine-phytoplankton. *Marine Ecology-Progress Series*, 58(1-2):133–142, 1989.
- [25] F. G. Haibach, A. E. Greer, M. V. Schiza, R. J. Priore, O. O. Soyemi, and M. L. Myrick. On-line reoptimization of filter designs for multivariate optical elements. *Applied Optics*, 42(10):1833–1838, 2003.
- [26] N. Hashemi, J. S. Erickson, J. P. Golden, K. M. Jackson, and F. S. Ligler. Microflow cytometer for optical analysis of phytoplankton. *Biosensors and Bioelectronics*, 26(11):4263–4269, 2011.
- [27] J. W. Hofstraat, W. J. M. Vanzeijl, M. E. J. Devreeze, J. C. H. Peeters, L. Peperzak, F. Colijn, and T. W. M. Rademaker. Phytoplankton monitoring by flow-cytometry. *Journal of Plankton Research*, 16(9):1197–1224, 1994.
- [28] E. B. Huerta, B. Duval, and J. K. Hao. A hybrid lda and genetic algorithm for gene selection and classification of microarray data. *Neurocomputing*, 73:2375–2383, 2010.
- [29] S. W. Jeffrey and M. Vesk. *Introduction of marine phytoplankton and their pigment signitures*. Phytoplankton Pigments in Oceanography: Guidlines to Modern Methods. UNESCO, Paris, 2nd edition, 2005.
- [30] T. K. Kim, H. Kim, W. Hwang, and J. Kittler. Component-based lda face description for image retrieval and mpeg-7 standardization. *Image Vis. Comput.*, 23(7):631–642, 2005.

- [31] C. M. Lalli and T. R. Parsons. *Biological Oceanography: An Introduction*. Butterworth-Heinemann, second edition, 1997.
- [32] A.J. Lewitus, R.A. Horner, D.A. Caron, E. Garcia-Mendoza, B.M. Hickey, M. Hunter, D.D. Huppert, R.M. Kudela, G.W. Langlois, J.L. Largier, E.J. Lessard, R. RaLonde, J.E.J. Rensel, P.G. Strutton, V.L. Trainer, and J.F. Tweddle. Harmful algal blooms along the north american west coast region: History, trends, causes, and impacts. *HARMFUL ALGAE*, 19:133–159, 2012.
- [33] L. H. Liska, P. R. Epstein, and W. H. Schesinger. Rising co₂, climate change, and public health: Exploring the links to plant biology. *Environmental Health Perspective*, 117:155–158, 2009.
- [34] D Livingstone. *A Practical Guide to Scientific Data Analysis*. John Wiley and Sons, Chichester, UK, 2009.
- [35] A. Longhurst, S. Sathyendranath, T. Platt, and C. Caverhill. An estimate of global primary production in the ocean from satellite radiometer data. *Journal of Plankton Research*, 17(6):1245–1271, 1995.
- [36] M. D. Mackey, D. J. Mackey, H. W. Higgins, and S. W. Wright. Chemtax - a program for estimating class abundances from chemical markers: Application to hplc measurements of phytoplankton. *Marine Ecology-Progress Series*, 144(1-3): 265–283, 1996.
- [37] J. Mazanec, M. Melisek, M. Oravec, and J. Pavlovicova. Support vector machines, pca and lda in face recognition. *J. Electr. Eng.*, 59:203–209, 2008.
- [38] M. P. Nelson, J. F. Aust, J. A. Dobrowolski, P. G Verly, and M. L. Myrick. Multivariate optical computing for predictive spectroscopy. *Analytical Chemistry*, 70:73–82, 1997.

- [39] T. Okaishi. *Red Tides: Biology*. Kluwer Academic Pub, Boston, 1994.
- [40] R. J. Olson, A. Shalapyonok, and H. M. Sosik. An automated submersible flow cytometer for analyzing pico- and nanophytoplankton: Flowcytobot. *Deep-Sea Research Part I-Oceanographic Research Papers*, 50(2):301–315, 2003.
- [41] M. R. Pearl, J. A. Swanstrom, L. S. Bruckman, T. L. Richardson, T. J. Shaw, H. Sosik, and M.L. Myrick. Taxonomic classification of phytoplankton with multivariate optical computing, part iii: Demonstration. *Applied Spectroscopy*, 67(6), 2013.
- [42] Colin Reynolds. *Ecology of Phytoplankton*. Number 53. Cambridge University Press, Cambridge, 2006.
- [43] P. J. Rice and B. P. Horgan. Evaluation of nitrogen and phosphorus transport with runoff from fairway turf managed with hollow tine core cultivation and verticutting. *Science of The Total Environment*, 456-457:61–68, 2013.
- [44] T. L. Richardson, E. Lawrenz, J. L. Pinckney, R. C. Guajardo, E. A. Walker, H. W. Paerl, and H. L. MacIntyre. Spectral fluorometric characterization of phytoplankton community composition using the algae online analyser (r). *Water Research*, 44(8):2461–2472, 2010.
- [45] K. Rodenacker, B. Hense, U. Jutting, and P. Gais. Automatic analysis of aqueous specimens for phytoplankton structure recognition and population estimation. *Microscopy Research and Technique*, 69(9):708–720, 2006.
- [46] K.S. Rowan. *Photosynthetic Pigments of Algae*. Cambridge University Press, 1989.
- [47] T. P. A. Rutten, B. Sandee, and A. R. T. Hofman. Phytoplankton monitoring by high performance flow cytometry: A successful approach? *Cytometry Part A*, 64A(1):16–26, 2005.

- [48] J. H. Ryther. Photosynthesis and fish production in sea. *Science*, 166(3901):72–78, 1969.
- [49] S. Sathyendranath, G. Cota, V. Stuart, H. Maass, and T. Platt. Remote sensing of phytoplankton pigments: a comparison of empirical and theoretical approaches. *International Journal of Remote Sensing*, 22(2-3):249–273, 2001.
- [50] A. Schmittner, A. Oschlies, H. D. Mathews, and E. D. Galbraith. Future changes in climate, ocean circulation, ecosystems, and biogeochemical cycling simulated for a business-as-usual co2 emission scenario until year 4000 ad. *Global Biogeochemical Cycles*, 22:1013–1033, 2008.
- [51] A. Sharma and K. K. Paliwal. Cancer classification by gradient lda technique using microrray gene expression data. *Data Knowl. Eng.*, 66:338–347, 2008.
- [52] A. Sharma, K. K. Paliwal, and G. C. Onwubolu. Class-dependent pca, mdc and lda: A combined classifier for pattern classification. *Pattern Recognition*, 39:1215–1229, 2006.
- [53] H. M. Sosik and R. J. Olson. Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. *Limnology and Oceanography-Methods*, 5:204–216, 2007. ISI Document Delivery No.: 200DB Times Cited: 7 Cited Reference Count: 38 Sosik, Heidi M. Olson, Robert J. AMER SOC LIMNOLOGY OCEANOGRAPHY.
- [54] O. Soyemi, D. Eastwood, L. Zhang, H. Li, J. Karunamuni, P. Gemperline, R. A. Synowicki, and M. L. Myrick. Design and testing of a multivariate optical element: The first demonstration of multivariate optical computing for predictive spectroscopy. *Analytical Chemistry*, 73(17):4393–4393, 2001.

- [55] M.A. Srokosz, A.P. Martin, and M.J.R. Fasham. On the role of biological dynamics in plankton patchiness at the mesoscale: An example from the eastern north atlantic ocean. *Journal of Marine Research*, 61:517–537, 2003.
- [56] M. Thyssen, G. Tarran, M.V. Zubkov, R.J. Holland, and G. Gregori. The emergence of automated high-frequency flow cytometry: revealing temporal and spatial phytoplankton variability. *Journal of Plankton Research*, 30(3):333–343, 2008.
- [57] D. Uhlmann, O. Schlimpert, and W. Uhlmann. Automated phytoplankton analysis by a pattern recognition method. *Internationale Revue der Gesamten Hydrobiologie*, 63(4):575–583, 1978.
- [58] J. Vepsäläinen, T. Pyhalahti, E. Rantajarvi, K. Kallio, S. Pertola, T. Stipa, M. Kiirikki, J. Pulliainen, and J. Seppala. The combined use of optical remote sensing data and unattended flow-through fluorometer measurements in the baltic sea. *International Journal of Remote Sensing*, 26(2):261–282, 2005.
- [59] A. Verikas, A. Gelzinis, M. Bacauskiene, I. Olenina, S. Olenin, and E. Vaiciukynas. Phase congruency-based detection of circular objects applied to analysis of phytoplankton images. *Pattern Recognition*, 45(4):1659–1670, 2012.
- [60] D Voet and J.G Voet. *Biochemistry: Biomolecules, Mechanisms of Enzyme Action, and Metabolism*, volume 1. John Wiley and Sons, 3rd edition, 2004.
- [61] H. Y. Wang, Z. F. Wang, Y. Leng, X. J. Wu, and Q Li. Pca plus f-lda: A new approach to face recognition. *Int. J. Pattern Recognition*, 21:1059–1068, 2007.
- [62] H. Yang and J. Y. Yang. Why can lda be performed in pca transformed space? *Pattern Recognition*, 36:563–566, 2003.

- [63] C. S. Yentsch and D. A. Phinney. Spectral fluorescence - an ataxonomic tool for studying the structure of phytoplankton populations. *Journal of Plankton Research*, 7(5):617–632, 1985.
- [64] C. S. Yentsch and C. M. Yentsch. Fluorescence spectral signatures - characterization of phytoplankton populations by the use of excitation and emission-spectra. *Journal of Marine Research*, 37(3):471–483, 1979.
- [65] Y. Zhao, J. Wang, Q. J. Lu, and R. Jiang. Pattern recognition of eggshell cracks using pca and lda. *Innov. Food Sci. Emerg. Technol*, 11:520–525, 2011.

CHAPTER 2

DESIGN AND THEORETICAL PERFORMANCE OF MULTIVARIATE OPTICAL ELEMENTS FOR TAXONOMIC CLASSIFICATION OF PHYTOPLANKTON

2.1 INTRODUCTION

Phytoplankton are single-celled, photosynthetic algae and cyanobacteria found in all aquatic environments. Phytoplankton photosynthesis accounts for at least 90% of global ocean primary productivity and is the nutritional foundation for virtually all life in the sea.[3] The size and taxonomic composition of phytoplankton communities determine how carbon is transferred through ocean food webs, and the extent to which it is recycled vs. transported to the deep ocean.[16] Characterization of phytoplankton communities over a wide range of temporal and spatial scales is critical to developing models of carbon cycling, and to predicting the ocean response to global climate change.

The potential use of spectral fluorescence for characterizing phytoplankton communities was first recognized by Yentsch and Phinney.[22] The general approach is based on selective excitation of differing antenna and accessory pigments between taxonomic groups of algae, with quantification of the resulting red fluorescence of chlorophyll *a*. [7] A number of similar approaches have recently been developed that rely on the use of fluorescence excitation spectra, emission spectra, or both, for phytoplankton quantification and/or characterization.[21, 1, 15]

Our work uses multivariate optical computing (MOC) to emulate linear discriminant (LD) vectors of phytoplankton fluorescence excitation spectra using a simple filter-photometer arrangement. Our previous studies have shown that optical measurements based on spectral pattern analysis can be performed using MOC.[18, 14] At the center of MOC is the design of interference filters that we call multivariate optical elements (MOEs). These interference filters have complex spectral transmission functions that make possible a variety of multivariate measurements. In the past, the main application of MOC was the measurement of chemical concentrations in complex samples based on near-infrared, UV-Visible, and Raman spectroscopies of the samples as an optical, in situ alternative to the post-sampling, multivariate techniques of partial least squares and principal components regression analysis.[10, 9, 19]

In the present work we present a new application of MOC to the classification of phytoplankton based on linear discriminants analysis (LDA) of fluorescence excitation spectroscopy. We first use LDA to confirm that three differently pigmented species of phytoplankton are differentiable from one another using single-cell fluorescence excitation spectroscopy. Thousands of possible MOEs were then designed to make the same measurement in a more robust filter-photometer system. Finally, we show how the theoretical characteristics of the designs were used in determining which optical elements were selected for fabrication. Two subsequent manuscripts in this series describe how the MOE designs were then used to construct and evaluate an instrument for the in situ classification of phytoplankton.

2.2 EXPERIMENTAL

Calibration Spectra

Single cell fluorescence excitation spectra were collected with a custom-built single cell fluorometer described by Hill et al.[6] Fluorescence was excited between 350 nm

and 650 nm at 2 nm increments using a spectral bandwidth of approximately 10 nm over the excitation wavelength region. Fluorescence emission was detected at 680 ± 5 nm defined with a bandpass filter.

Unialgal cultures of three phytoplankton species were obtained from the Center for the Culture of Marine Phytoplankton (CCMP strain; recently re-named the National Center for Marine Algae and Microbiota) at the Bigelow Laboratory for Ocean Sciences, East Boothbay Harbor, Maine. These cultures were: *Emiliania huxleyi* (CCMP strain 375, *E. huxleyi*), *Thalassiosira pseudonana* (CCMP strain 1335, *T. pseudonana*) and *Synechococcus* sp. (CCMP 833, *Synechococcus*). Cells were grown in batch culture at 23°C and 80 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$ with illumination from the top and the sides on a 12:12 light:dark cycle in filter-sterilized f/2 culture medium.[4] Spectra were collected from nutrient-replete cells in exponential growth phase. A total of 200 single cell fluorescence excitation spectra were used as a calibration set: 77 for *E. huxleyi* spectra, 46 for *T. pseudonana* and 76 for *Synechococcus*. The 200 spectra used for calibration were those that did not exhibit any of the errors described in reference [Hill et al.], which include high apparent noise due to cell motion, collisions with other phytoplankton, and escape of the phytoplankter from the trap during study.

The critical measurement for using MOC for classification is the ratio of the recorded fluorescence from a single phytoplankter cell when excitation first passes through a MOE to that recorded when excitation passes through another filter, either a neutral density filter or another MOE. Acquiring the MOE measurement referenced to a neutral density filter is optically equivalent to normalizing the MOE measurement to the total integrated intensity in a fluorescence excitation spectrum over the excitation band.[17] For the purposes of LDA, the equivalent procedure is normalization of the calibration spectra to unit area. Thus, the fluorescence excitation spectra in this study were normalized to unit integrated area to correct for variations in total

emission prior to analysis by LDA to verify if classification was practical under this normalization condition. Figure 2.1 shows the full spectral data set normalized to unit area.[6]

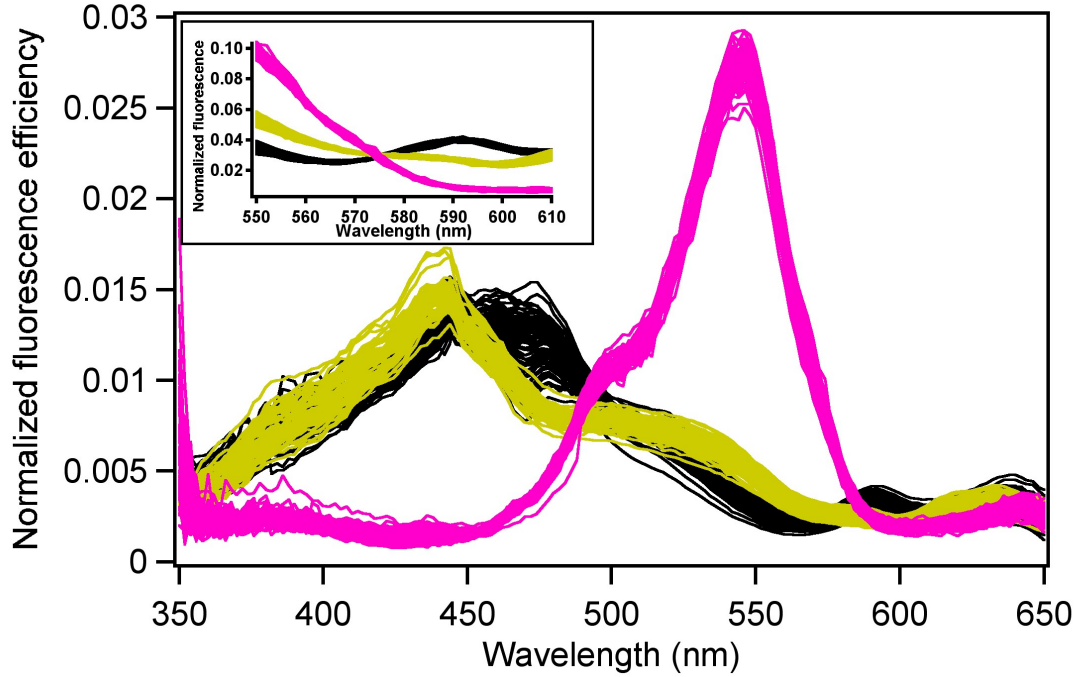


Figure 2.1 Fluorescence excitation spectra of individual cells of three species. The area-normalized fluorescence excitation spectra of 77 *E. huxleyi* (black), 46 *T. pseudonana* (yellow) and 76 *Synechococcus sp.* (pink) cells excited in 2 nm increments over the range of 350 to 650 nm with a spectral resolution of ~ 10 nm. The inset shows the same spectra renormalized in the wavelength range for best classification, 550 to 610 nm.

LDA was conducted using the approach described by Bruckman et al., which uses LDA to compare different continuous wavelength regions for their performance.[2] The result of this LDA-based analysis is an estimate of the best achievable discrimination among species based on the available calibration spectra as well as the best continuous wavelength region for performing that discrimination for the three classes of phytoplankton tested. As a result of this brute-force automated approach to LDA, the 550-610 nm wavelength region was found to have the most favorable characteristics for distinguishing the three phytoplankton species.

Multivariate Optical Element Design

In the absence of a native MOC algorithm for approximating LDA, MOEs were designed using the existing calibration code. This was accomplished by treating the scores of each calibration spectrum on the linear discriminant vector from LDA as if they were quantities we would like the MOEs to predict. The existing calibration code is based on a nonlinear optimization algorithm described previously by Soyemi et al.[19] that has been modified to run on a computer cluster.[13] This algorithm is coded in MATLAB; the cluster itself and the software that implements the algorithm are referred to below as the design suite because it is capable of MOE design using a variety of different criteria and for a variety of different optical computing styles.[13] The computer cluster that runs the design suite software consists of three Dell PCs with a total of 12 processors running in parallel.

The design suite algorithm operates by generating random seed points (i.e., initial guesses at the structure of a MOE) that are then refined. In operation, the random seed points for the design suite optimization are constrained in several ways. First, the materials from which the layers will be designed are predefined as a binary pair of high and low-index materials. For the work here, this pair was Nb_2O_5/SiO_2 . The second constraint is the total thickness of all seed points, expressed as an approximate spectral resolution in wavenumbers. In this study, we constrained the stack to start at 256 cm^{-1} , which for a quarter-wave stack of Nb_2O_5 and SiO_2 is equivalent to a physical thickness of 7.28 micrometers.[13] The third constraint is in the number of layers that form the seed stack. In this study, seed stacks contained 11 layers, 6 of which were high index, and 5 of which were low index.

From this point, the algorithm seeks to optimize the thicknesses of the layers in the film stack of the interference filter so as to minimize the standard error of calibration (SEC) for prediction of the phytoplankton's score on a linear discriminant vector. Optimization continues until no changes in the film stack yield an improved SEC.

The design process is repeated many times to form a population of optimized MOE designs with different characteristics.

The MOE ratio is used to determine the score of the optical spectrum on the particular spectroscopic vector embodied in the MOE by a simple linear equation with a gain and offset. MOEs can be designed so that the ratio increases with the score (a positive gain) or so that the ratio decreases with the score (a negative gain). In fluorescence excitation spectroscopy, these filters are introduced into the excitation light path, and fluorescence intensities are the quantities used to construct ratios. The reference filter in MOC may be an open window, a neutral density filter, or another MOE. If the latter is chosen, the two MOEs are generally chosen to have opposite signs of their gain so that sensitivity is increased.

Each MOE is designed to make predictions only under the particular set of design conditions used for the design suite calculation⁸. Specifically, the MOEs are designed to operate in real instruments and measure real (*i.e.*, not ideal) intensities. Real instruments do not respond equally to all wavelengths of light, so anything that affects the spectral response of the instrument must be included in the design process. The choices of optical components in the eventual photometer system are important in MOC; they should have stable, measurable and reproducible spectral profiles, although the absolute response is of lesser importance. These components should be known or selected and their spectral characteristics determined before an MOE can be designed and manufactured. Figure 2.2 is a schematic of an imaging photometer that would be capable of collecting the measurements required for applying optical computing to classifying phytoplankton, including a description of the optical components.

The optimal excitation wavelength region defined by LDA for discriminating the three species was 550-610 nm as shown in the inset of Fig. 2.1. MOC with MOEs provides no intrinsic method of restricting the wavelength region exactly. Instead, the

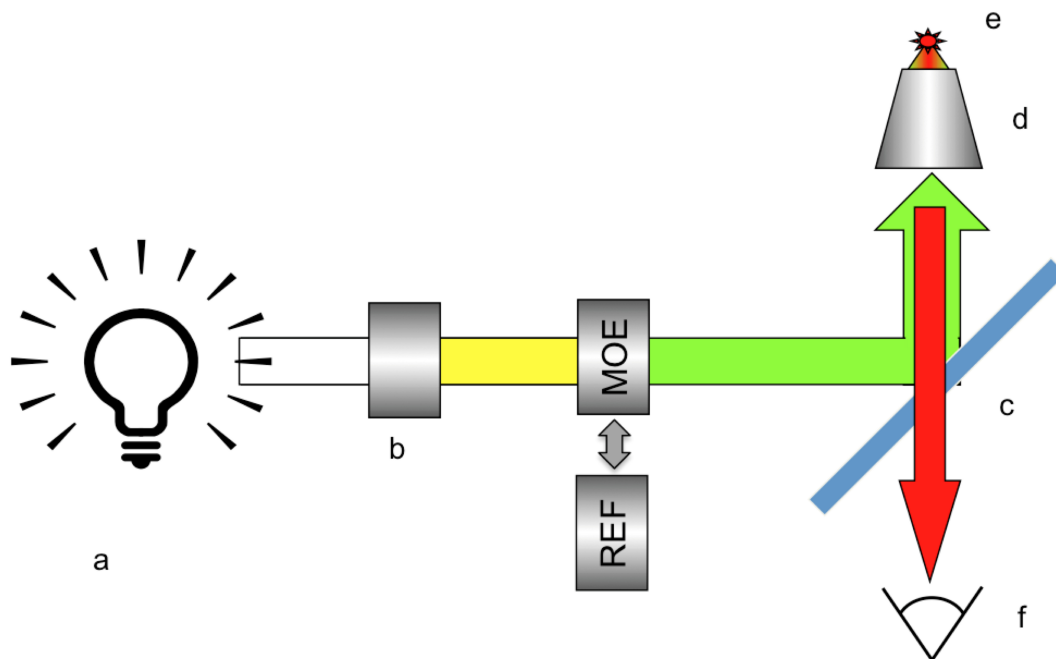


Figure 2.2 Schematic representation of an MOC instrument. A white light source (a) is filtered (b) to give a range of wavelengths suited to classifying phytoplankton. An MOE designed for that purpose is placed into the beam, alternating with a reference filter, (REF). A dichroic beamsplitter reflects the excitation toward a microscope objective (d) and then to a phytoplankter (e). Fluorescence from the phytoplankter cell is returned to a detector system (f) which integrates the signals and determines the filter type (MOE vs. REF) for each signal.

wavelength region is defined by selecting an off-the-shelf bandpass interference filter that approximates the required selection. In this case, a bandpass filter (element b in Fig. 2.2) produced by Chroma Technology, Inc. (HQ580/60M) was selected for this purpose.

E. huxleyi, *T. pseudonana* and *Synechococcus sp.* are reported by the CCMP to be 1-10 micrometers in size, requiring a microscope objective to image them individually (element d in Fig. 2.2). The microscope objective selected was a Nikon 60X 0.7NA plan-fluor objective. Nikon provided transmission values for this objective of 77, 74, 68 and 72% at 550nm, 600 nm, 650 nm and 700 nm, respectively. These transmission values were fit to a quadratic function, which was then used to estimate the transmission of the objective at all the discrete wavelengths used in the design

suite.

The lamp selected to provide excitation (element a in Fig. 2.2) was a 75W Xe arc lamp (Sciencetech, London, Ontario)(201-100 arc lamp housing, SCI 201 power supply, OSRAM XBO 75W/2 OFR short-arc lamp). The Xe arc lamp spectral irradiance profile was measured on an automated Spectroradiometric Measurement System (Optronic Laboratories, Inc. Orlando, FL; OL Series 750) over a 280-1100 nm range. The standard lamp used was an OL 200IR Standard of Spectral Irradiance (Optronic Laboratories S-1041). The detector was a silicon detector (Optronic Laboratories 750-HSD-301EC). An OL 750-M-S monochromator was used for calibration with 1.25 mm slit and LG-05-600 grating with 600 grooves/mm density and blaze wavelength of 500 nm and resolution of 10 nm. The distance from the Xe arc lamp and the standard lamp was 50 cm to the monochromator entrance for calibration. The Xe lamp has a relatively flat spectral irradiance profile and maintains a stable power output, attributes required for MOC.

The excitation and emission light in Fig. 2.2 follow the same path through the objective so a dichroic beamsplitter (element c in Fig. 2.2) was required to reflect the selected excitation range from 550-610 nm and transmit chlorophyll a emission at 680 nm. In this case, a dichroic beamsplitter produced by Omega Optical, Inc. (640DCLP) was selected for this purpose. Figure 2.3 shows plots of the spectral characteristics of the optical components assumed in designing the MOEs. Inputs to the design suite for this study included: normalized and corrected calibration spectra; the spectral characteristics shown in Fig. 2.3; and scores of each of the calibration spectra on the linear discriminant vectors previously determined by LDA.

MOE designs consist of alternating layers of materials with different refractive indices whose optical interference yields transmission spectra mimicking the target linear discriminant vector determined in LDA. Often this mimicry is functional rather than exact. The final spectrum of a given MOE design may perform well at repro-

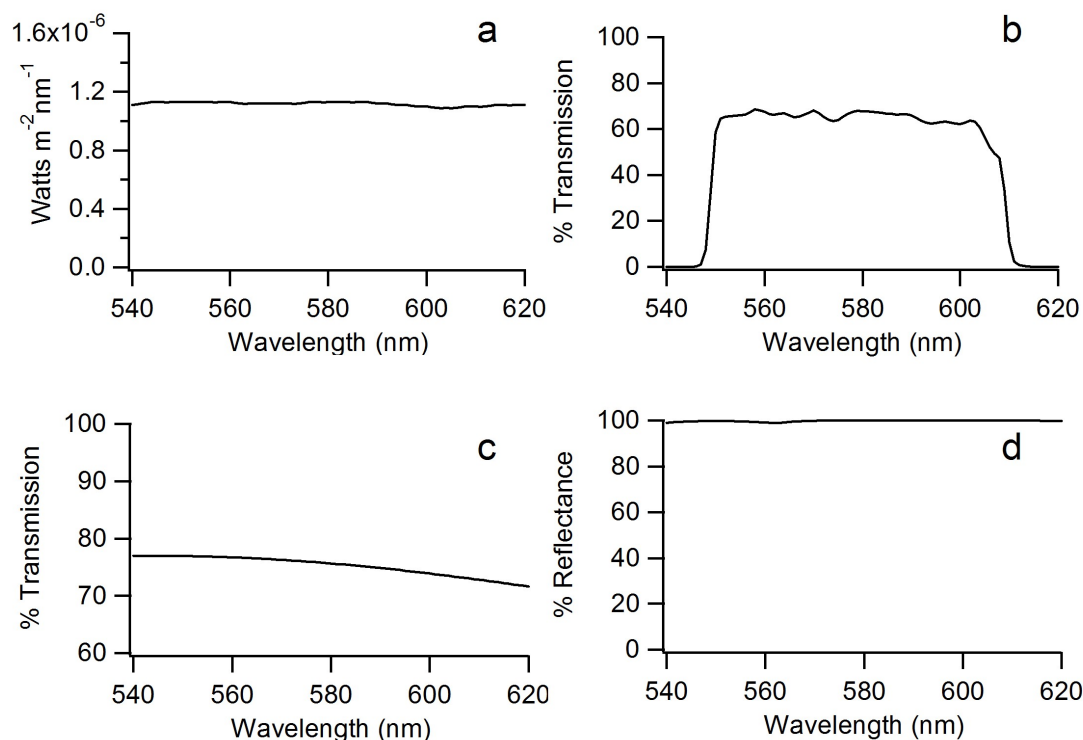


Figure 2.3 Optical functions of the components discussed in figure 2.2 used for designing MOEs in this report. a) Irradiance of Xe arc lamp, b) percent transmission of broadband bandpass filter, c) percent transmission of 60X 0.7 NA plan-fluor objective, d) percent reflectance of the beamsplitter at 45 degrees. Details on components are given in the text.

ducing the loadings of the linear discriminant while looking similar to, inverted from, or only sharing some of the features of the wavelength loadings of the target linear discriminant function.

The design space exhibits numerous local minima of the SEC merit function, some of which correspond to designs that would be satisfactory, but many of which are not. To find sufficient satisfactory designs, the design process is repeated many times in an effort to find minima corresponding to designs that satisfy a number of performance criteria at the same time. For this study, 121,982 designs were generated that predict the scores of spectra on LD1. Generation and optimization of these designs required approximately 6 hours.

Statistics that were reported for each design include sensitivity, standard error of calibration, gain and offset, % transmittance, total thickness, design time, and the number of layers in the design as well as the thickness of each layer. An analysis of these designs is given in a following section. The result of the analysis is the selection for fabrication of one or two target designs.

Curve Fitting

In an analysis below it was necessary to curve fit a distribution using a function representing a sum of Poisson distributions. Most automated fitting routines assume the variability for underlying data is the same at all points by default. This was not appropriate here, but nor was it clear what the underlying variability should be. To explore this question, the fit was obtained by iteration. In the first iteration, the default variability was assumed. The result of the fit was used in a simulation of 50,000 distributions that could have resulted from the model if it were correct, and at each point in the distribution the variability was calculated. It was observed that high points in the distribution had a standard deviation that was nearly the square root of the frequency of occurrence and simultaneously near 1/4th of the total range. Low points in the distribution had standard deviations that were much less than 1/4th of their ranges and that produced too high a weighting effect in the conventional curve fitting process. Thus we chose weighting factors that were 1/4th of the range found at each point in the simulated distributions. After this decision, we repeated the curve fit with these new weights to each point. The process was repeated until the results began iterating around the same values. This couldn't be avoided because the fit coefficients were not constrained to integer values, but some had to be truncated to integer values for the simulation. No convergence could be obtained for this reason, but a narrow range of possible values for each coefficient was determined.

After the curve fitting was complete, we compared these weighting factors to the

square root of the average distribution number at each point in the model. The latter approach gave a slightly lower weight to low points in the histogram and a slightly higher weight to high points than we would have estimated by simulation. We then conducted another iterative solution in which the variability was chosen as the square root of the model value at each point, the advantage of this approach being that it does not require any simulation and truncation to integer values. This approach converged after a few iterations. The values found in this approach are not the best "true" values for the fit we have performed because the weights at each point are not ideal. Nevertheless, they are a significant improvement from the assumption of uniform errors at all points, and each of the coefficients we found by this approach was within the range of the coefficients found in the simulation approach.

The approach reported in the section on characterization of unique designs that follows (*vide infra*) is the simpler square-root approach because it is far more rapid (minutes instead of days) but also reasonably consistent with the more rigorous simulation approach. It is also worth noting that the more rigorous approach based on simulation is still not completely accurate because the fitting process assumes the points exhibit a symmetric distribution around their mean value, while in a Poisson distribution points with low "counts" exhibit skew.

Fabrication

The fabrication of the MOE is done in our laboratory by reactive magnetron sputtering (RMS) using a chamber from Corona Vacuum Coaters Inc. RMS produces films that can be uniformly and reproducibly deposited onto a substrate.[11] The MOEs are fabricated by depositing thin layers of materials with different refractive indexes onto a glass substrate (Corning BK-7). Silicon dioxide, SiO₂ ($n \approx 1.45$ at 540 nm) is used as a low refractive index material and niobium pentoxide, Nb₂O₅ ($n \approx 2.26$ at 540 nm) is used as a high refractive index material⁸. In situ spectroscopy of the

MOE during fabrication is used to determine the actual thickness and other properties of the films being deposited in real time. Minor errors in thickness are then compensated as described previously by Haibach et al.[5] The average time to deposit a single layer of each material in our laboratory is about one day, depending on the thickness of the layer. Unlike typical off-the-shelf interference filters, MOEs are not generally required to exhibit high optical density, which makes it possible to choose designs with a minimum number of layers.

2.3 RESULTS AND DISCUSSION

Linear Discriminant Analysis

Fluorescence excitation spectra for individual cells of the three phytoplankton species (77 *E. huxleyi* cells, 46 *T. pseudonana* cells and 76 *Synechococcus sp.* cells grown as described above) are shown in Fig. 2.1; the spectra for each cell are normalized to unit area over the wavelength range 350-650 nm. Linear discriminant analysis (LDA) is a method for optimally separating classes by maximizing Fisher ratios. The combinatorial approach described in reference [Bruckman et al.] showed that the spectral window 550-610 nm was best for separating the three phytoplankton classes in this report. The inset graph in Fig. 2.1 shows this sub-region of the full spectrum with each individual spectrum normalized to unit area over just this spectral window.

The LDA approach described in reference 16 uses principal components analysis as a preprocessing step before discriminant analysis. The first two principal components of the spectra normalized over the 550-610 nm region provided perfect classification accuracy based on leave-one-out cross-validation, and yielded Fisher ratios no less than 132 for distinguishing any two classes from one another.

Figure 2.4 a shows how the individual cell spectra score on the two linear discriminant vectors that result from LDA using the first two principal components.

These two discriminant axes are the result of a rotation in the plane of the first two principal components, but with the rotation done in such a way that the first linear discriminant provides most of the classifying power of the first two principal components. Figure 2.4 b shows the spectral profile of this first linear discriminant vector, and since this vector carries most of the spectral information necessary for classification, it will be the sole focus of the remainder of this manuscript. Likewise, the lowest Fisher ratio of 132 was found for distinguishing *E. huxleyi* cells from *T. pseudonana* cells, *i.e.*, these cells are the most difficult of the pairs to distinguish from one another. Since this classification is the most difficult, the separation of *E. huxleyi* from *T. pseudonana* will be the sole focus of the remainder of this manuscript.

Multivariate Optical Elements

Multivariate optical elements were designed to mimic the function of the linear discriminant vector shown in Fig. 2.4. Most applications of MOC have been directed at quantitative chemical measurement, so the design suite software that designs MOEs does so by refinement of initial interference filter structures to minimize SEC.[17] Ideally, design of MOEs for classification of phytoplankton would be driven using the Fisher ratio as a figure of merit. But since analytical measurement software already exists, we chose the simple expedient of designing MOEs to "predict" the scores of the phytoplankton spectra on the first linear discriminant vector revealed by LDA.

Each design calculated by the design suite has a set of properties that define the physical properties of the filter and also how well that particular design will function in predicting the LD score of a phytoplankton. These properties include: optimized SEC, sensitivity, overall filter thickness, number of layers, individual layer thicknesses, average transmission, and gain. The SEC establishes how well the MOE is capable of predicting the score of a phytoplankton on LD1 in an ideal, noiseless measurement (other than the noise present in the original calibration spectra). Sensitivity describes

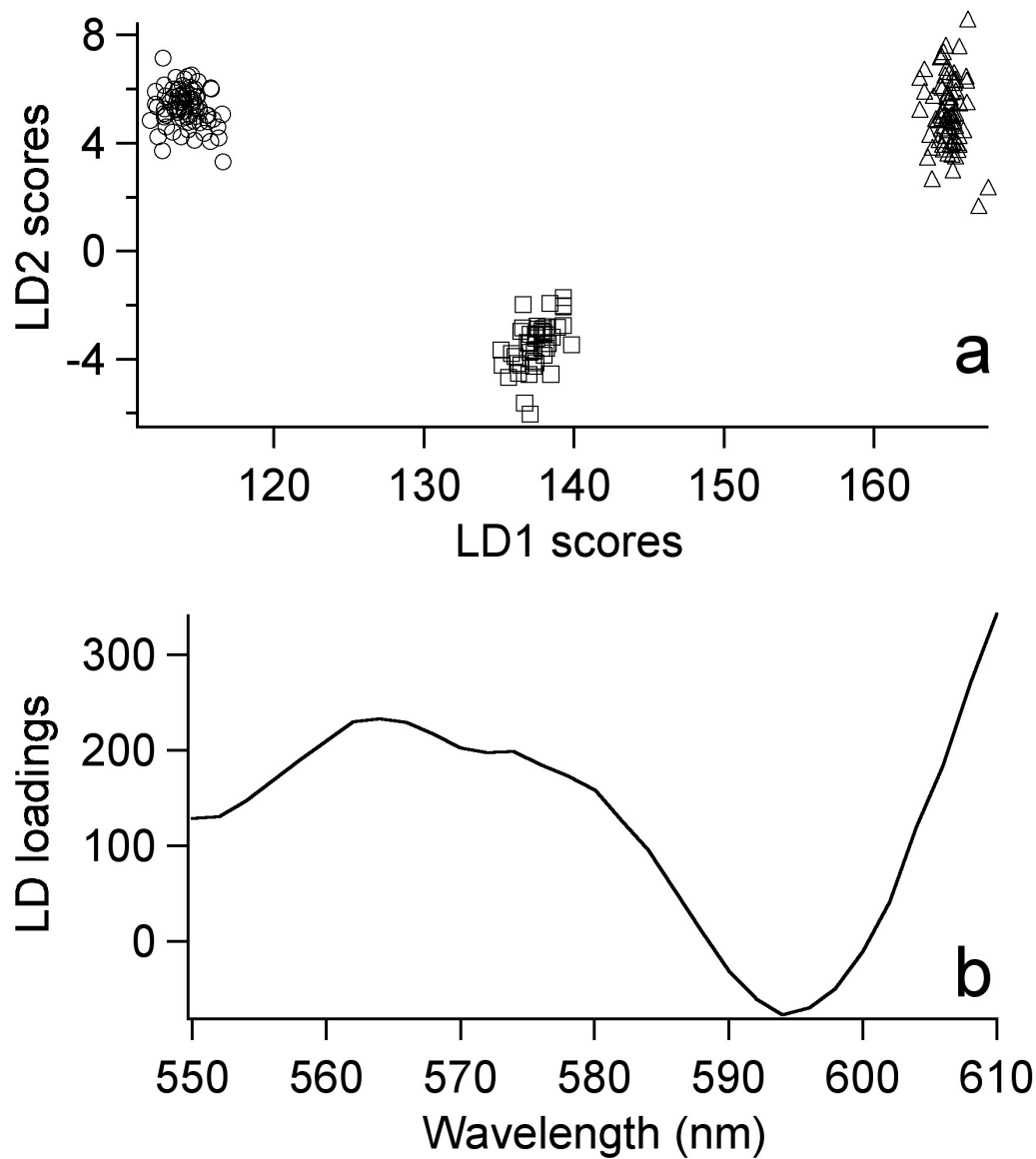


Figure 2.4 Linear discriminant analysis (LDA) results based on inset data in Figure 2.1. (a) Plot of *E. huxleyi* (circles), *T. pseudonana* (squares) and *Synechococcus sp.* (triangles) scores of the individual calibration spectra along linear discriminant 1 (LD1) and 2 (LD2). (b) Spectral profile of the loadings of LD1 in the wavelength range specified by the results of LDA, 550 to 610 nm.

how much the intensity ratio for phytoplankton fluorescence changes per unit of score on LD1, and is related to how easily the measurement can be made in the presence of noise or variability. Total filter thickness is related to the length of time necessary to fabricate a given MOE, while the number of filter layers is related to the number of total steps required to complete a given MOE. Average transmission for MOEs is related to the expected SNR in the measurement, with low transmission yielding small signals with low signal to noise ratio (SNR). Gain is the slope of the predicted calibration of the MOE. Unlike the other properties, the maximum number of layers is established as an environment within which the MOEs are designed, so no criterion needs to be applied to this property later.

The first step in evaluating the results of a design cycle is to apply criteria to each of the properties of MOEs. A non-predictive MOE will always have an SEC that is identical to the standard deviation of the dependent variable in a design run — in this case, the standard deviation of the vector of LD1 scores used in the design process which was found to be 22.3 units. This sets the upper limit to the SEC that could be determined for an MOE; the lower limit is zero.

We generally find that peak sensitivity increases monotonically along with SEC, so it is useful to define the maximum value of SEC that would represent an acceptable MOE. In this case, we decided that an SEC equal to the standard deviation of the LD1 scores within a given class of phytoplankton would make a reasonable cutoff for acceptable MOEs. LD1 scores, as illustrated in Fig. 2.4, varied over a range of approximately 51 units. On the same scale, the standard deviation within the tightest single phytoplankton class cluster was 0.793 units, so this was chosen as a maximum value criterion for SEC. A spreadsheet of designs was sorted by SEC and then all designs with SEC greater than 0.793 units were deleted. 87,229 of the initial 121,982 designs passed this threshold.

Setting the minimum acceptable sensitivity is the second criterion we applied to

the selection of MOE designs. In the present case, *E. huxleyi* and *T. pseudonana* differ in their average scores on LD1 by 23.37 score units on LD1. A MOE with a nominal sensitivity of 1% per unit would vary in its response by 23.37% over this range. Sample variability is already included in the calibration set, but the future instrument's variability and noise are not. The criterion for minimum sensitivity is then set by the operating conditions of the future instrument in which the MOE would be installed. As an example, if a 24% change in response were needed to be sure samples were not misclassified due to instrument variability, then a 1% per unit lower limit to MOE sensitivity would need to be set. In our case we set a lower limit of 0.5% per unit based on little more than an educated guess of what the future instrument's performance might be like - it had not been constructed at that point - and deleted the designs that failed to meet this criterion. Only 5607 designs remained after this threshold was set.

The maximum thickness criterion was set at 6 micrometers to keep fabrication time down to one week or less in our laboratory deposition chamber. The minimum average transmittance was set at 6% from previous experience. After these criteria were applied, 5,279 designs remained.

Among these potentially useful designs are many that are duplicates of one another. The rate at which duplicates of designs are found contains useful information regarding the number of designs and the likely manufacturability of the design, as described in a section below. But duplicates must be removed before a true picture of the number of usable unique designs can emerge. In this case, removal of duplicates left 823 unique designs to consider manufacturing that met all the minimum criteria. Figure 2.5 shows the dependence of sensitivity of the remaining MOEs on their SEC. One fact that becomes apparent from Fig. 2.5 is that while the highest sensitivity tends to decrease along with the SEC, many low sensitivity designs are found at all values of SEC. All other things being equal, a high level of sensitivity is

always desirable, so the cream of the crop of designs always lies along the top right corner of the sensitivity-vs-SEC distribution.

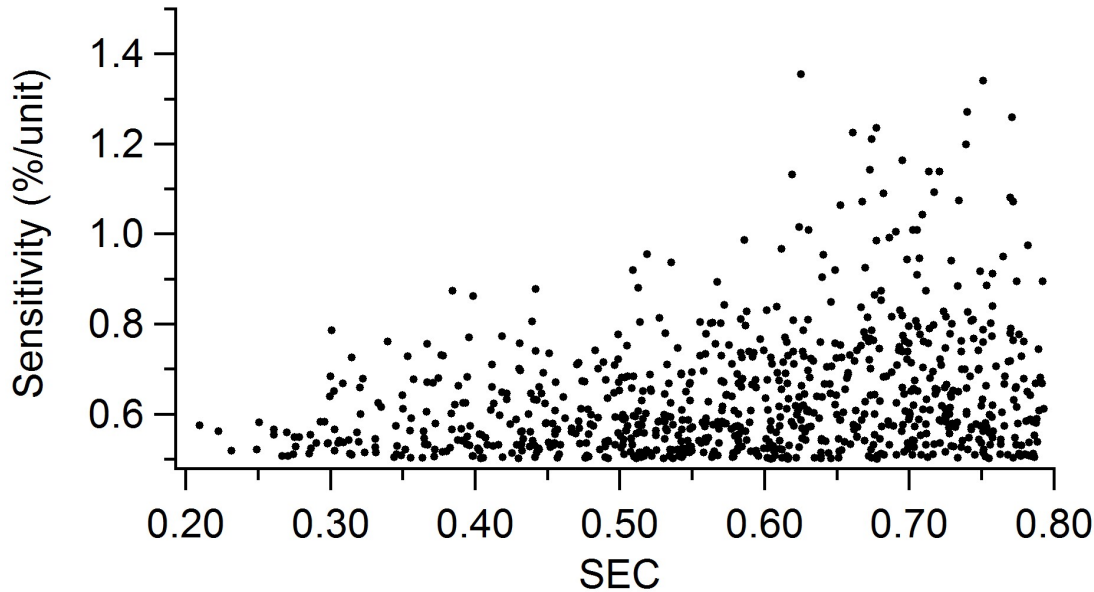


Figure 2.5 Sensitivity vs. standard error of calibration (SEC) for 823 unique designs meeting selection criteria. Sensitivity is defined as the percentage change in MOE/REF ratio per unit change in LD1 score. SEC is the standard error of calibration in units of LD1 score.

A final stage in selecting a MOE for manufacturing is to examine the remaining designs for their likelihood to be manufactured correctly. This determination is made by a person skilled in thin film operations and relies on both experience in the art as well as additional calculations. One of the calculations that can be performed on prospective MOE designs is to determine what degradation of performance would result if layers of the resulting design were deposited inaccurately. Such a calculation can use randomly generated errors within a fixed limit added to the thickness of each layer, with a recalculation of the SEC for each possible set of errors. After a number of error vectors have been tested, the RMS SEC for the set can be calculated and used to compare different designs for manufacturability. On the other hand, art comes into play when the specialist makes determinations about minimum or maximum layer

thicknesses that he/she believes make the deposition process particularly difficult.

Following a detailed examination of the available designs, two different MOEs with opposite gains were selected for manufacturing. These were designs #212 and #23551, whose initial characteristics are shown in Table 2.1. The re-optimized designs of the two MOEs are shown in Table 2.1 using the most recent optical constants for our deposition chamber, with the layers shown from closest to the substrate to furthest away, and the MOE designed for use with the film stack side of the substrate facing the incident radiation. Figure 2.6 shows the transmission curves of the two MOEs.

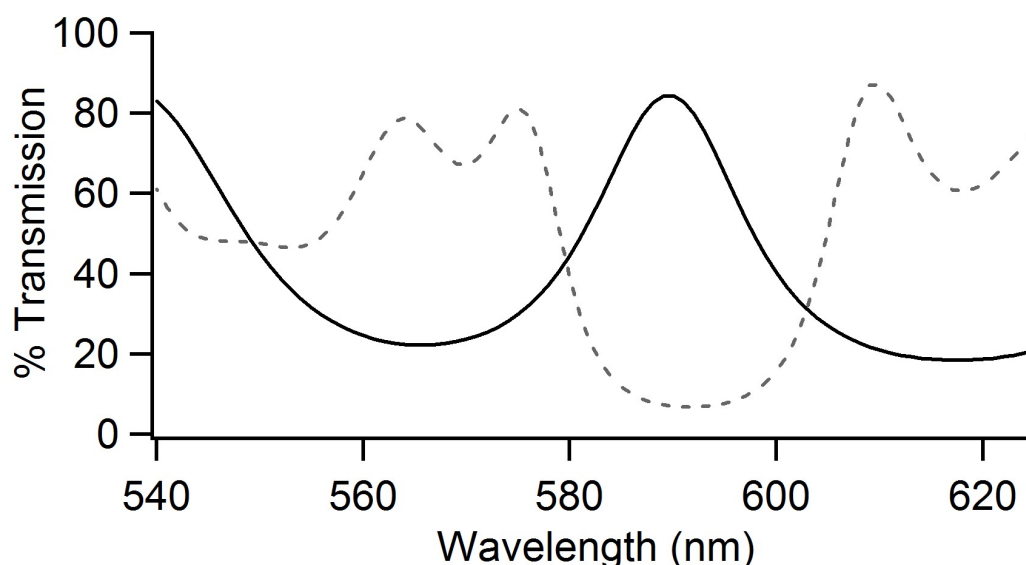


Figure 2.6 Transmission curves of MOE 212 (Dotted line) and MOE 23551 (solid line) over the wavelength range of 540-620 nm.

The two MOEs, based on the optical constants recorded in our laboratory (nominally 1.46 for SiO_2 and 2.24 for Nb_2O_5 over the 550-610 nm spectral range), were designed to have the initial structures given in Table 2.2. During deposition, an optical model of the evolving structure is used to update the effective optical constants of each film material, and the "recipe" (list of layer thicknesses) of the MOE is reoptimized to take advantage of this improved information. As a result, the final layer thicknesses of each optical element may differ from the initial recipe, and the

spectrum of the MOE may differ slightly from the original spectrum expected of it. However, the target of deposition and reoptimization is not to maintain a particular spectrum or particular layer thicknesses, but to maintain performance as measured by the SEC figure of merit. A comparison of the initial and estimated final layer thicknesses shows that reoptimization changed some layers by substantial amounts, yet the resulting theoretical sensitivities and SECs have remained relatively stable through the process of converting a hypothetical optical component into a real device. The estimated final layer thicknesses were the result of the modeling process during deposition; attempts to confirm these values through variable-angle spectroscopic ellipsometry failed due to the complexity of the film stack and the fact that the films have slight birefringence.

Table 2.1 (a)Design suite results. The theoretical characteristics of the designs selected for fabrication are shown. SEC is the standard error of calibration; sensitivity is given in % change of the ratio of fluorescence excited through the MOE to that excited through a neutral density for each unit off LD1 score; Gain is a measure of the slope of the relationship between the ratio and score. Note that one is chosen with positive gain, and the other with negative gain. (b) Calculated results after final layer deposited. This shows the same characteristics as part (a) except with changes to the design following fabrication of the optical elements.

Table 1(a)

| Design no. | Design Time (sec) | SEC | Sensitivity | Gain | No. of Layers |
|------------|-------------------|-------|-------------|--------|---------------|
| 212 | 0.0297 | 0.182 | 0.557 | -437.7 | 8 |
| 23551 | 0.00598 | 0.210 | 0.575 | 366.5 | 9 |

Table 1 (b)

| Design no. | Fab time (days) | SEC | Sensitivity | Gain | No. of Layers |
|------------|-----------------|-------|-------------|--------|---------------|
| 212 | 6 | 0.173 | 0.533 | -463.3 | 8 |
| 23551 | 8 | 0.216 | 0.567 | 370.2 | 9 |

Table 2.2 (a) Layer thicknesses for Design 212. This shows the designed thicknesses of the first 5 layers that composed Design 212 after design (initial) and after fabrication was completed (final). (b) Layer thicknesses for Design 23551. This shows the designed thicknesses of 5 layers that compose design 23551 after the design stage (initial) and after fabrication was completed (final).

Table 2 (a)

| Stage | High (nm) | Low (nm) | High (nm) | Low (nm) | High (nm) |
|---------|-----------|----------|-----------|----------|-----------|
| Initial | 70.276 | 112.625 | 214.794 | 462.724 | 537.092 |
| Final | 69.215 | 84.804 | 197.068 | 531.806 | 520.585 |

Table 2 (b)

| Stage | High (nm) | Low (nm) | High (nm) | Low (nm) | High (nm) |
|---------|-----------|----------|-----------|----------|-----------|
| Initial | 728.282 | 514.714 | 478.465 | 917.282 | 435.529 |
| Final | 718.964 | 513.638 | 471.666 | 918.782 | 427.213 |

Duplicate Designs

Whenever large numbers of designs are generated, duplicate designs are found in the output. Of the 5,279 original designs that met all our minimum criteria, only 823 were unique.

The fact that duplicate designs occur when we seed the design process with random starting points indicates that there are not an infinite number of designs for MOEs possible under the constraints applied during the design step (e.g., maximum number of layers; starting thickness of the film stack; type of optical computing, etc).

It is qualitatively observed that the number of unique designs increases with additional design cycles but with a rate that decreases over time. One question that often arises in the design of MOEs is: how many unique designs are there? A study of the statistical distribution of MOE designs allows us to make an estimate of this quantity. The statistics of rare events, such as the finding of even the most readily-found design, are governed by Poisson statistics.[8] If each design has the same probability of being found, then the entire ensemble corresponds to a single population governed

by a single distribution. A less severe approximation is to model the ensemble of designs as the summation of a small set of populations, within each of which all the designs have the same probability of being found in a search. We model the distribution of the repetitive designs as such a sum where each population is represented by a Poisson distribution in the form of Eq. 2.1:

$$F_{r,\langle r \rangle} \approx \frac{\langle r_i \rangle^r e^{-\langle r_i \rangle}}{r!} \quad (2.1)$$

In this version of the Poisson distribution equation, r is interpreted as the number of times a given design is repeated in the results and $\langle r_i \rangle$ is the average number of repeats for designs in the i^{th} population. $\langle r_i \rangle$ itself is given by:

$$\langle r_i \rangle = p_i * T \quad (2.2)$$

where p_i is the probability of a given design in population i being found in each design trial and T is the total number of trials in a particular design run. Populations of readily-found designs advance in the r dimension more rapidly with T than those that are less readily found, leading to a separation along the r coordinate.

Using Eq. 2.1 to evaluate the design space for a system modeled to have P populations of designs, the number of designs repeated r times, represented by $n(r)$, is:

$$n(r) \approx \sum_{i=1}^P N_i \frac{\langle r_i \rangle^r e^{-\langle r_i \rangle}}{r!} \quad (2.3)$$

where N_i is the total number of designs in the i^{th} design population. The number of designs that have not yet been found is given by substituting $r=0$:

$$n(0) \approx \sum_{i=1}^P N_i e^{-\langle r_i \rangle} \quad (2.4)$$

Using this approximation and the definition of $\langle r_i \rangle$ given in Eq. 2.2, the number of new, unique designs that would be found in a new run of X design cycles is found

by determining the number of designs that would not yet be found even after an additional X design cycles:

$$-n\Delta(0) \approx \sum_{i=1}^P N_i e^{-\langle r_i \rangle} \left(1 - e^{-\langle r_i \rangle \frac{X}{T}}\right) \quad (2.5)$$

In Eq. 2.5, T is the initial number of trials that were run (e.g., 120,000), X is the number of additional trials, represents the average repetition number for the i^{th} population after the initial T trials, and $-\Delta n(0)$ is a measure of the new designs found in the next X design trials.

After many design cycles have already been carried out, the average number of repetitions of designs becomes larger, and the first exponential begins to slow the rate of new discovery of easily-found designs. The designs that are left to be found tend to become weighted toward the designs that are least likely to be found, so that in any system that fits a multiple population model, the rate of discovery decreases faster than exponentially at first. Eventually, it reaches the exponentially-decreasing rate of discovery expected of the most refractory designs.

To date, there is nothing known about populations of designs that would lead one to believe that refractory designs have more desirable characteristics than easily-found designs, so there is no rationale for carrying out extensive design trials once the majority of designs is already found.

Figure 2.7 is a plot showing the distribution of the repeats of the 823 unique designs found above. The form of the plot is of the number of occurrences of designs (y axis) that were repeated a given number of times (x axis). The sum of all the values on the plot is 823, while the sum of the product of the x and y values for each point is 5,279, the total number of design trials that met criteria. A fit to this graph assuming one or two populations yielded large residuals with clear trending. A four-population fit using Eq. reffr:

$$f(r) = a_2 \frac{a_1^r e^{-a_1}}{r!} + a_4 \frac{a_3^r e^{-a_3}}{r!} + a_6 \frac{a_5^r e^{-a_5}}{r!} + a_8 \frac{a_7^r e^{-a_7}}{r!} \quad (2.6)$$

gave the fit and residuals shown in the caption of Fig. 2.7.

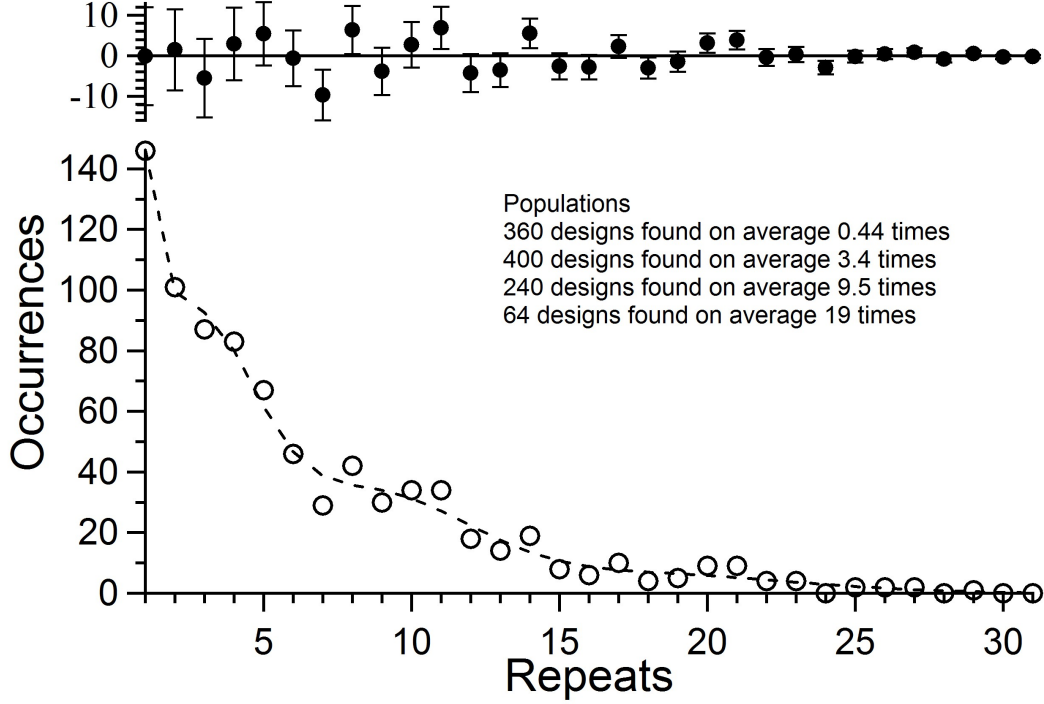


Figure 2.7 Distribution (circles) and multi-Poisson distribution model (line) of design duplicates of 823 unique MOE designs observed in a run of 120,000 cycles. Four populations of designs were modeled with 19.1 ± 1.0 , 9.5 ± 0.5 , 3.4 ± 0.4 , and 0.44 ± 0.41 repeats on average. The four populations have 64 ± 13 , 242 ± 26 , 400 ± 50 , and 360 ± 150 members, respectively, for a total accessible design population of 1060 ± 160 designs. The top axis shows the residuals between observation and model, with error bars representing the square root of the modeled number of occurrences. The fitting procedure is described in the text. Reduced χ^2 for the fit was 1.16.

The sum of the factors a_2, a_4, a_6 and a_8 gives an estimate of the total design population, including designs not yet found, of 1060 ± 160 designs. If this is the case, then our trial of $\sim 20,000$ design trials identified almost 80% of all designs that could be obtained within the constraints of our starting points. Eq. 2.4 tells us that if we ran an additional 120,000 design trials, we could expect to find another 95 unique

designs, increasing our yield only to 87%. The final 13% would be nearly all in the most refractory population.

This observation tells us that, if we feel more design work is necessary before selection preliminary to manufacturing, it would be more fruitful to vary the constraints on seeding the designs than to continue working in the constraints behind Fig. 2.7.

Further work along these lines is currently under way in our laboratories in an effort to explore the population model of MOE design and to understand whether the ease of discovery for different populations relates to their manufacturability.

Theoretical Performance

This manuscript describes the design and fabrication of MOEs for discriminating three classes of phytoplankton from one another. Each MOE is designed to transmit a spectral function of the excitation light whose detected intensity relative to a neutral density can be used to predict the score of phytoplankton's spectra on the most important linear discriminant function found using LDA.

The measurement of how a phytoplankton responds to the MOE is made by measuring its fluorescence intensity in a 10-nm band centered at 680 nm, mimicking the detection of the calibration instrument. This fluorescence intensity is usually measured relative to the fluorescence intensity detected when the same phytoplankton is excited through a neutral density filter. By convention, the neutral density is chosen to have approximately the same overall throughput as the MOE to avoid as much detector nonlinearity as possible.

An alternate approach is to measure the MOE-excited intensity relative to that of another MOE of opposite sign. For example, MOE #23551 above has a positive gain, meaning that the MOE/ND ratios it produces increase with increasing score of the phytoplankton fluorescence excitation spectrum on LD1. MOE #212, on the other hand, produces MOE/ND ratios that decrease with the score on LD1. Since

the sign of a linear discriminant function is arbitrary, the definitions of which MOE has a positive gain and which has a negative are likewise arbitrary. The fact that they are of differing signs, however, is not.

To increase the sensitivity of measurement to the phytoplankton fluorescence excitation spectrum, we choose to measure the ratio of opposite-sign MOEs. Take for example the ratios expected of *E. huxleyi* and *T. pseudonana* in Fig. 2.8 under different conditions of measurement. A 0.3 optical density filter gives a relatively flat excitation spectrum over the 550-610 nm wavelength range, and has an average transmittance similar to the MOEs described above. Figure 8 shows the difference in ratio responses using the positive-gain MOE #23551 relative to both a 0.3 OD neutral density (ND) reference filter and also to the negative-gain MOE #212. The mean values of the MOE/ND ratio are 0.865 ± 0.007 , 0.994 ± 0.007 and 1.167 ± 0.005 for *E. huxleyi*, *T. pseudonana* and *Synechococcus sp.* respectively, and the mean values of the MOE/MOE are 0.926 ± 0.012 , 1.177 ± 0.014 , and 1.620 ± 0.013 for *E. huxleyi*, *T. pseudonana* and *Synechococcus sp.* respectively. These ratios were calculated by finding the direct product of MOE #23551 transmission spectrum and the ND or MOE #212 transmission spectrum with the computed system response to each of the spectra in the calibration set, where the system response is given by the product of the phytoplankton calibration spectrum with the spectra shown in Fig. 2.1. The greater difference in ratios for the two most similar species with a two-MOE measurement versus a single MOE and ND measurement leads to reduced demands on the performance of the MOC instrument.

Signal to Noise Requirements

How the MOEs will perform in a real instrument is determined by how much noise is introduced by the measurement. Variability in the ratios determined for phytoplankton comes from three sources: (a) the inherent variability of the time-averaged

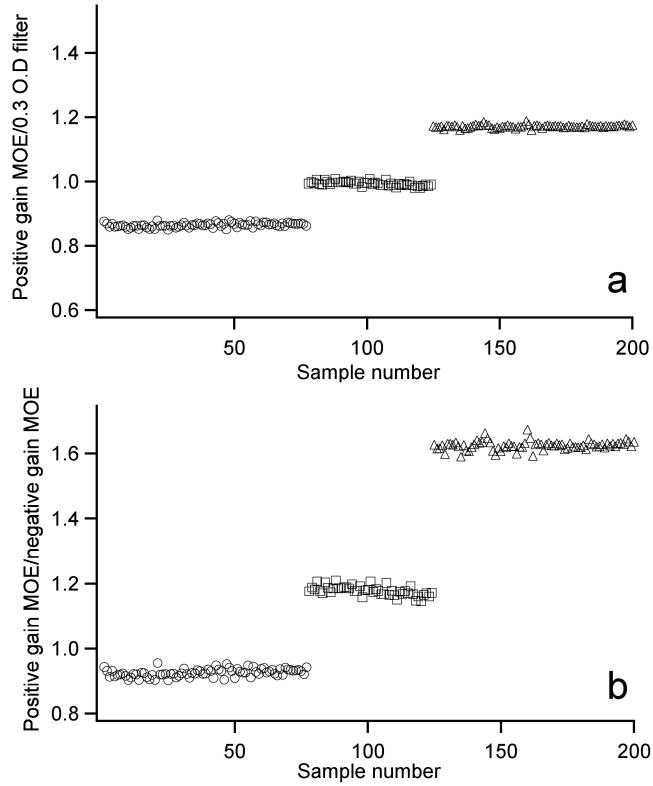


Figure 2.8 Theoretical results of classification performance of designed MOE applied to calibration data. a) Ratio of a single MOE to a 0.3 OD neutral density filter b) Ratio of MOE 23551 to MOE 212.

fluorescence excitation spectrum of individual organisms, (b) variability in the excitation cross-section and apparent fluorescence quantum yield of an individual organism with time, and (c) noise due to instrumental factors. The variability of the ratios calculated from the calibration spectra places an upper limit on the first of these factors for monocultures that are grown in nutrient replete conditions. The second factor originates from the fact that the individual measurements from which the ratio is constructed are recorded at different times if the MOEs are sampled sequentially. It is well known that phytoplankton fluorescence efficiency responds to an increase in light level, increasing rapidly as the photoactive centers are flooded and then decreasing gradually as photoprotective pigments are brought into play.[3] The third factor consists of numerous contributors such as dark noise of detectors, variability of illumi-

nation, spatial dependence of response, noise due to background signals, nonlinearity of detector response, and photon shot noise.

The variability in calculated ratios using the calibration spectra for *E. huxleyi* and *T. pseudonana* is similar: approximately 0.01 within each species in Fig. 2.8. The ratio separation between the averages of the two species is approximately 0.25 as illustrated in Fig. 2.8. Since the second and third sources of variability listed above should contribute nearly equally to each MOE measurement, it is possible to estimate the performance required to give reasonable classification.

We make the estimate by considering only discrimination between two classes (1 and 2) using a single MOE ratio, where the two classes cluster around different average ratios, one that is lower (R_1) and one that is higher (R_2), respectively. We then assume that somewhere between these two ratios lies a point of discrimination, R_d , and that any phytoplankton with a measured ratio above R_d is assigned to class 2, else it is assigned to class 1. We further assume that the distribution of measurements in each class around their average values is described by a normal distribution, although we are simply using this as a guide and do not require the distribution to be rigorously Gaussian.

The value of R_d for optimum separation is the point at which the two probability distributions for the two classes cross one another. For ideal Gaussian distributions, the distance between the class means and this discrimination point is the same when normalized to the sample standard deviation of the class in question. To be certain that at least 95% of plankton belonging to class 1 are correctly assigned by this method, we require that the sample standard deviation times the one-sided 95% confidence value of Student's t , $t_{0.95}$, be less than or equal to the difference between $R_d - R_1$. Because of the way R_d is selected, this assures that at least 95% of samples belonging to class 2 will also be correctly classified.

This can be simplified greatly if we assume both classes have the same sample

standard deviation. In this case, R_d lies exactly between R_1 and R_2 , and we require a sample standard deviation, s , such that

$$2 * t_{0.95} * s \leq (R_2 - R_1) \equiv \Delta R_{21} \quad (2.7)$$

The necessary sample standard deviation is found by solving for s ,

$$s \leq \frac{\Delta R_{21}}{2 * t_{0.95}} \quad (2.8)$$

The sample standard deviation here is that of a ratio between two separate measurements, an MOE and its reference. Each of these measurements can contribute to the variability of s , and propagation of errors allows us to obtain an analytical expression in terms of each separate measurement. However, it is likely in most experiments that ratios never depart greatly from unity because the reference is chosen in practice to have broadly the same order of overall transmission to avoid detector range and nonlinearity problems. Since we are simply seeking a guide to the necessary SNR in a measurement system, it is convenient to further simplify the result by assuming both the MOE and its reference have nearly the same signals and nearly the same sample standard deviations. In this case, the standard deviation of the ratio is $\sqrt{2}$ times the relative standard deviation of one of the measured signals. The relative standard deviation of the measured signal is simply the reciprocal of the instrument SNR, so substitution into Eq. 2.8 gives:

$$\text{SNR} \geq \frac{2 * \sqrt{2} * t_{0.95}}{\Delta R_{21}} \approx \frac{4.7}{\Delta R_{21}} \quad (2.9)$$

This final form assumes we have large samples from which a reasonable sample standard deviation can be computed. For the case in hand, $\Delta R_{21} = 0.25$, and we therefore require a SNR of 19 to distinguish *textit{E. huxleyi}* from *T. pseudonana* with around 95% accuracy. In the next manuscript of this series we describe an instrument capable of measuring fluorescence from an individual phytoplankter in a

few milliseconds with a SNR ratio meeting this requirement. Two more manuscripts further describe taxonomic classification of phytoplankton with MOC by using the MOEs designed and fabricated here, places them in the photometer,[20] and applies them to live cultures of *E. huxleyi* and *T. pseudonana*. [12] The mean ratios reported there for *E. huxleyi* and *T. pseudonana* were 0.900 ± 0.029 and 1.179 ± 0.036 respectively, versus the theoretical ratios of 0.926 ± 0.012 and 1.177 ± 0.014 reported here. This gives a class separation of 0.278 ratio units compared to the optical model here of 0.251 ratio units, and standard deviations for the classes that are limited more by the photometer SNR than by the inherent variability of the phytoplankter cells, while still meeting the minimum criteria developed above for successful classification.

2.4 CONCLUSIONS

In this chapter I have shown how the results of LDA can be used as a basis for designing optical elements whose sole job is to score fluorescence excitation spectra against linear discriminant vectors. This can potentially be used to discriminate between two classes of phytoplankton in a fluorescence-based instrument. In the following chapters, I report the design, construction and characterization of a CCD-based instrument capable of analyzing moderate volumes of water and classifying phytoplankton found therein. The SNR of the instrument is found to be adequate for the purpose outlined in this chapter using measurements from a single cycle of a rotating filter wheel in which the positive and negative MOEs described in Table Ib are cycled at a rate of ~ 17 Hz, for a total measurement time of about 60 msec. A final published manuscript shows that measurement of *E. huxleyi* and *T. pseudonana* cells under these conditions results in measured fluorescence ratios very close to those predicted here for each cell type, and with high efficiency for classification.

BIBLIOGRAPHY

- [1] M. Beutler, K. H. Wiltshire, B. Meyer, C. Moldaenke, C. Luring, M. Meyerhofer, U. P. Hansen, and H. Dau. A fluorometric method for the differentiation of algal populations in vivo and in situ. *Photosynthesis Research*, 72(1):39–53, 2002.
- [2] L. S. Bruckman, T. L. Richardson, J. A. Swanstrom, K. A. Donaldson, M. Allora, T. J. Shaw, and M. L. Myrick. Linear discriminant analysis of single-cell fluorescence excitation spectra of five phytoplankton species. *Applied Spectroscopy*, 66(1):60–65, 2012.
- [3] P. G. Falkowski and J. A. Raven. *Aquatic Photosynthesis*. Princeton University Press, Princeton, 2nd edition, 2007.
- [4] R. R. Guillard. *Culture of phytoplankton for feeding marine invertebrates*. Culture of Marine Invertebrate Animals. Plenum Press, New York, 1975.
- [5] F. G. Haibach, A. E. Greer, M. V. Schiza, R. J. Priore, O. O. Soyemi, and M. L. Myrick. On-line reoptimization of filter designs for multivariate optical elements. *Applied Optics*, 42(10):1833–1838, 2003.
- [6] L. S. Hill, T. L. Richardson, L. T. M. Profeta, T. J. Shaw, C. J. Hintz, B. S. Twining, E. Lawrenz, and M. L. Myrick. Construction, figures of merit, and testing of a single-cell fluorescence excitation spectroscopy system. *Review of Scientific Instruments*, 81(1):013103, 2010.
- [7] S. W. Jeffrey and M. Veski. *Introduction of marine phytoplankton and their pigment signatures*. Phytoplankton Pigments in Oceanography: Guidelines to Modern Methods. UNESCO, Paris, 2nd edition, 2005.

- [8] S. M. Kendall and A. Stuart. *The Advanced Theory of Statistics: Volume 1 Distribution Theory*. Macmillan Publishing Co., Inc, Cambridge, 1977.
- [9] M. L. Myrick, O. Soyemi, J. Karunamuni, D. Eastwood, H. Li, L. Zhang, A. E. Greer, and P. Gemperline. A single-element all-optical approach to chemometric prediction. *Vibrational Spectroscopy*, 28(1):73–81, 2002.
- [10] Micheal L. Myrick, Olusola. Soyemi, Fred Haibach, Lixia Zhang, Ashley Greer, Hongli Li, Ryan Priore, Maria V. Schiza, and J. R. Farr. Application of multivariate optical computing to near-infrared imaging. *Proceedings of SPIE*, 4577 (Vibrational Spectroscopy-Based Sensor Systems):148–157, 2002.
- [11] R. Parsons. *Sputter Deposition Process*. Thin Film Processes II. Academic Press, San Diego, 1991.
- [12] M. R. Pearl, J. A. Swanstrom, L. S. Bruckman, T. L. Richardson, T. J. Shaw, H. Sosik, and M.L. Myrick. Taxonomic classification of phytoplankton with multivariate optical computing, part iii: Demonstration. *Applied Spectroscopy*, 67(6), 2013.
- [13] R. J. Priore. *The design, fabrication and implementation of point detection and chemical imaging sensors utilizing optical and digital regression techniques*. PhD thesis, 2005.
- [14] R. J. Priore, F. G. Haibach, M. V. Schiza, A. E. Greer, D. L. Perkins, and M. L. Myrick. Miniature stereo spectral imaging system for multivariate optical computing. *Applied Spectroscopy*, 58(7):870–873, 2004.
- [15] T. L. Richardson, E. Lawrenz, J. L. Pinckney, R. C. Guajardo, E. A. Walker, H. W. Paerl, and H. L. MacIntyre. Spectral fluorometric characterization of phytoplankton community composition using the algae online analyser (r). *Water Research*, 44(8):2461–2472, 2010.

- [16] J. L. Sarmiento and N. Gruber. *Ocean Biogeochemical Dynamics*. Princeton University Press, Princeton, 2006.
- [17] M. N. Simcock and M. L. Myrick. Precision in imaging multivariate optical computing. *Applied Optics*, 46(7):1066–1080, 2007.
- [18] O. Soyemi, D. Eastwood, L. Zhang, H. Li, J. Karunamuni, P. Gemperline, R. A. Synowicki, and M. L. Myrick. Design and testing of a multivariate optical element: The first demonstration of multivariate optical computing for predictive spectroscopy. *Analytical Chemistry*, 73(17):4393–4393, 2001.
- [19] O. Soyemi, F. G. Haibach, P. J. Gemperline, and M. L. Myrick. Nonlinear optimization algorithm for multivariate optical element design. *Applied Spectroscopy*, 56(4):477–487, 2002.
- [20] J. A. Swanstrom, L. S. Bruckman, M. R. Pearl, E. Abernathy, T. L. Richardson, T. J. Shaw, and M. L. Myrick. Taxonomic classification of phytoplankton with multivariate optical computing, part ii: Design and experimental protocol of a shipboard fluorescence imaging photometer. *Applied Spectroscopy*, 67(6), 2013.
- [21] Z. Uysal. Chroococcoid cyanobacteria synechococcus spp. in the black sea: pigments, size, distribution, growth and diurnal variability. *Journal of Plankton Research*, 23(2):175–189, 2001.
- [22] C. S. Yentsch and D. A. Phinney. Spectral fluorescence - an ataxonomic tool for studying the structure of phytoplankton populations. *Journal of Plankton Research*, 7(5):617–632, 1985.

CHAPTER 3

DESIGN AND EXPERIMENTAL PROTOCOL OF A SHIPBOARD STREAK IMAGING MULTIVARIATE OPTICAL COMPUTING PHOTOMETER (SSIMOC)

3.1 INTRODUCTION

Characterization of phytoplankton size and taxonomic composition is critical to understanding the biogeochemical functioning of marine ecosystems.[10, 21] Fluorometric measurement of chlorophyll a is a common approach to quantifying phytoplankton biomass, but while it is useful for estimating the concentration of chlorophyll a, it lacks the ability to discriminate phytoplankton taxa or discern cell size.[6, 15, 16] Currently, the most common approach to determining phytoplankton community structure is by visual examination (and counting) of species using light microscopy of fixed samples. Unfortunately, this technique is laborious and time-consuming and is impractical for systems that vary greatly in space and time.[22, 24, 28] In situ instruments that monitor continuously offer a better chance of following phytoplankton community composition in real or near-real time.[2, 3, 5, 20]

In 1978, Uhlmann et al.[26, 27] discussed the importance of frequent sampling to track the behavior of a dynamic phytoplankton community. They reported construction and evaluation of an imaging system that used pattern recognition specifically for this purpose.[27] Later developments in flow cytometry and imaging, in particular the development of the charged couple device (CCD) and the personal computer, led

to more automated, continuous flow instruments.[13, 25, 7, 11] The Imaging FlowCytobot developed by Olsen and Sosik represents the state-of-the-art in automated classification of phytoplankton by image analysis, eliminating discrete sampling and providing continuous monitoring of a coastal ocean site near Marthass Vineyard.[18] For the many organisms that are distinguishable by their size and shape, the Imaging FlowCytobot provides rapid and excellent classification. For those that are not readily distinguished in this way, additional means of classification are still sought.

All phytoplankton contain chlorophyll *a* or a derivative thereof, with a fluorescence emission maximum near 680 nm.[12] As a result, the emission wavelength of 680 nm is used in most fluorometric applications precisely because it is not unique or taxa-dependent. This wavelength is also used in cytometers to distinguish phytoplankton cells from other living or non-living particles.

While chlorophyll *a* is relatively ubiquitous, phytoplankton taxa vary in their complement of antenna and accessory pigments. These differences are used for taxonomic classification by identifying biomarker pigments and also by fluorescence excitation spectroscopy.[20, 4, 29, 17] Conventional fluorescence excitation spectroscopy records the fluorescence at a particular emission wavelength or band of wavelengths as a function of the excitation wavelength. Ideally, it is corrected for the source excitation intensity. Unlike fluorescence emission spectroscopy, in which intensity at every emission wavelength can easily be recorded in parallel, fluorescence excitation spectroscopy generally requires scanning, either of a discrete excitation wavelength or combinations of excitation wavelengths (e.g., via an interferometer). This restricts the amount of light that can be used for a measurement and slows the process down. Filter-based or LED-based excitation sources trade information content for speed and sensitivity by reducing the number of discrete channels that are sampled.

As an alternative to conventional fluorescence excitation spectroscopy, we have developed a method based on multivariate optical computing (MOC) that uses flu-

orescence excitation spectral information as a basis for classification via discriminant analysis. MOC uses multivariate optical elements (MOEs) designed to optically mimic linear discriminant functions optimized for distinguishing phytoplankton taxa.[23] These MOEs are special optical filters that provide broad spectral coverage without sacrificing spectral content, enabling high quality spectral information to be obtained in real time. Part I of this series describes how these MOEs are designed and fabricated for a generic instrument.[23]

To make use of MOEs, we have developed a dynamic flow fluorescence imaging filter photometer. This photometer differs from most in that it uses an imaging CCD array to record the fluorescence emitted from a phytoplankter, e.g. *Emiliania huxleyi* (*E. huxleyi*), when excited from light filtered through an MOE. Phytoplankton are drawn into a flow cell and passed through an area of excitation that is modulated by a rotating 6-position filter wheel. The filter wheel contains up to 5 MOEs, plus an opaque glass element in the 6th aperture to create a reference position in every rotation of the wheel.

The aim of this article is to describe the dynamic imaging photometer above, to characterize its performance and to detail all the mechanical, optical, electrical, physical, and biological contributions to the variability in its measured signals. In the sections below, we detail the optical layout, flow cell design, magnification, depth of field, constraints on filter wheel and flow velocities, procedures for blank subtraction and flat-field correction, the measurement scheme of the photometer, and its signal to noise ratio (SNR) as a function of filter wheel frequency. A semi-automatic approach to extracting fluorescence intensities from the imaging photometer data is presented and used to determine the sensitivity of the instrument. After that we perform a quantitative analysis of contributions to noise from 1) arc lamp ripple, 2) aperture-to-aperture variability, 3) variability in filter wheel frequency, 4) detector read and dark noise, 5) errors in the flat field correction, and 6) variability due to background

fluorescence and baseline subtraction error, and 7) photon counting error.

In some cases we find that the SNR of this photometer is not limited by any property of the photometer itself but by variance in the individual phytoplankter fluorescence that increases linearly with signal and integration time. We then consider possible sources of linearly dependent variance or apparent variance in individual cell fluorescence through a qualitative analysis of 1) flow direction artifacts, 2) interference effects in the flow cell, 3) photoacclimation effects in the phytoplankton, 4) photochemical decay, 5) detector non-linearity and 6) Brownian rotation. Based on this analysis we hypothesize a dynamically fluctuating fluorescence quantum efficiency in the phytoplankton to explain this unanticipated fluorescence variability. Despite this, we find the photometer can be operated to meet the minimum SNR requirements set forth in part I of this series.

3.2 EXPERIMENTAL

Instrument Design

A ray-tracing program (OSLO[®], Lambda Research, Littleton, MA) was used to aid in designing the optical configuration of the instrument to efficiently transfer the excitation radiation from the Xe lamp source, through the MOEs and to the sample plane without distorting the radiation flux over the image plane. A diagram of the excitation path is shown in Fig. 3.1.

The photometer is designed to be easily transported and operated within the space restrictions of a small oceanographic research vessel Fig. 3.2. A 75 W Xe lamp (Sci 200, Sciencetech, Ontario, Canada) with a f/2 condenser is aligned to project the beam onto a ~ 2.5 mm aperture formed by an adjustable iris. The excitation beam is collimated by a 1-inch f/1 biconvex lens (KBX046, Newport, Irvine, CA) and passed through a band pass filter (HQ580/60, Chroma, Bellows Falls, VT) to restrict the

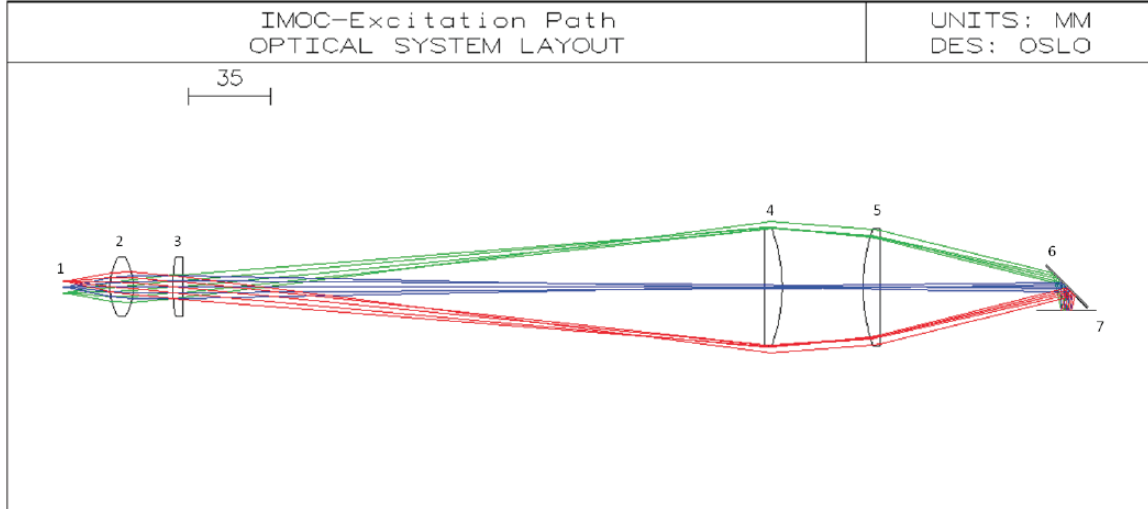


Figure 3.1 The rays travel from left (component 1) to right (component 7), originating at the source and pass through the optical components to a Nikon 60x objective. Component 1 is a 75 W Xe lamp (Sci 200, Sciencetech, Ontario, Canada); component 2 is a 1-inch f/1 biconvex lens (KBX046, Newport, Irvine, CA) ; component 3 is a 1-inch f/6 planoconvex lens (KPX100, Newport, Irvine, CA) ; component 4 is a 2-inch f/3 planoconvex field lens (KPX199, Newport, Irvine, CA) ; component 5 is another f/3 planoconvex lens (KPX199, Newport, Irvine, CA); component 6 is a 2-inch diameter beam splitter (p0643dr1p, Omega Optical, Brattleboro, VT) positioned at a 45° angle to the incident light to reflect it onto component 7 which is a 60X 0.70 NA Nikon infinity-corrected objective.

wavelength band of the excitation beam to 550-610 nm. This is the same filter whose characteristic spectrum was provided in Chapter 2, Fig. 3. A 1-inch f/6 planoconvex lens (KPX100, Newport, Irvine, CA) transfers the beam through a 6-position filter wheel (FW103, Thor labs, Newton, NJ) that houses the MOEs. The filter wheel is rotated by a brushless DC-servomotor (2057B, Faulhaber, Schonaich, Germany) and the frequency is set by a PC configurable speed controller (SC 2804 S, Faulhaber, Schonaich, Germany).

The filter wheel modulates the excitation beam with each "on" cycle passing through a different MOE. The excitation is then collected by a 2-inch f/3 planoconvex field lens (KPX199, Newport, Irvine, CA) after the chopper wheel and then further focused by another f/3 planoconvex lens (KPX199, Newport, Irvine, CA) onto a 2-

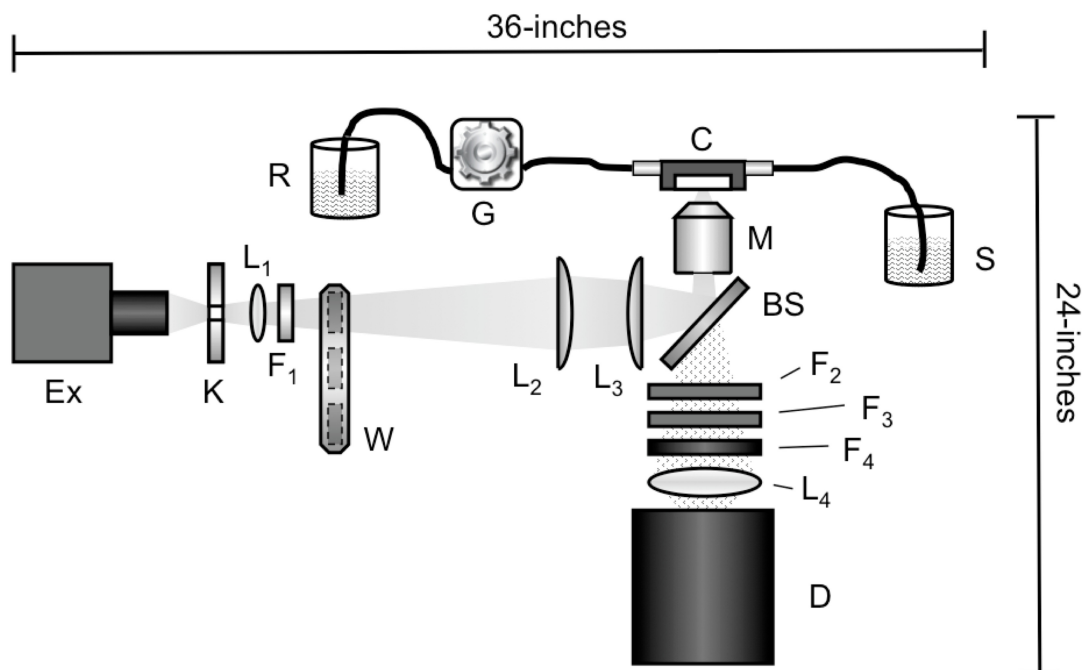


Figure 3.2 Schematic of fluorescence imaging photometer. Ex, the excitation source is a Sciencetech 500-200 75W Xe Arc lamp and K is a ~ 2.5 mm aperture. The lenses, in order of excitation to emission, are as follows: L₁ is a 1 in. diameter biconvex lens with a focal length of 1 in., L₂ and L₃ are 2 in. diameter planoconvex lenses with 6 in. focal lengths, and L₄ is a 2 in. diameter biconvex lens with a 4 in. focal length. The filters are as follows: F₁ is a 550-610nm Chroma bandpass (HQ580/60), F₂ is a 681 ± 5 nm Omega Optical (HBP10) bandpass, F₃ is an OG 530 Schott long pass, and F₄ is an OG 590 Schott long pass. BS is an Omega Optical 2 in. diameter dichroic beamsplitter (640drlp). The filter wheel, W, is a Thorlabs FW103, powered by a Faulhaber 2057B brushless motor. The objective, M, is a Nikon, Plan Fluor, 60x magnification, 0.70 NA objective. The sample is contained in a beaker, S, and is pulled the flow cell, C, by a Cole-Parmer 75211-10 gear pump, G, to a waste beaker, R. The fluorescence emitted by the sample is passed back through the objective, M, and beamsplitter, BS, and imaged onto a back-illuminated Princeton PIXIS 1024B CCD.

inch diameter beam splitter (p0643drlp, Omega Optical, Brattleboro, VT) positioned at a 45° angle to the incident light reflecting 550-610 nm onto the back entrance of a 60X 0.70 NA Nikon infinity-corrected objective and thence onto the flow cell.

The flow cell is made from black Delrin®(DuPont, Wilmington, DE), a material used for fluorescence measurements because it has low reflectance and fluorescence characteristics.[1] The flow cell was designed and fabricated in house. The window of the flow cell is a 22 x 60 mm, 0.17 mm thick glass cover slip (12-544-G, Fisher Scientific, Pittsburg, PA) and is affixed to the cell with epoxy (Torr Seal, Varian, Lexington, MA). The cell depth is approximately 200 μ m and the focus of the objective is placed approximately at the middle of the cell by focusing through the depth of the cell from the front to the rear and then reversing to the mid-point.

A gear pump (75211-10, Cole-Parmer, Vernon Hills, IL) with pump head (07002-25, Micropump, Vancouver, WA) draws the sample through the flow cell at a flow rate of \sim 3 mL/min. The final flow rate for each filter wheel frequency is adjusted slightly to obtain nine to ten streaks of an individual phytoplankter per image. Phytoplankton are pulled through the sample cell before passing through the gear pump, this is to ensure that the pump does not damage the phytoplankton prior to entering the flow cell.

The fluorescence emission from a phytoplankter at 680 nm returns through the objective, through the beam splitter, through a 2-inch diameter 695 ± 35 nm band pass filter (695HBP70, Omega Optical, Brattleboro, VT), and finally through 2-inch square glass long pass filters (OG530 and OG590, Schott, Elmsford, NY). The fluorescence is then focused by a 2-inch f/1 BK7 antireflection-coated biconvex lens (KBX142AR.14, Newport, Irvine, CA) onto a thermoelectrically cooled back-illuminated 1024 x 1024 CCD camera with a 13.3 x 13.3 mm CCD (Pixis 1024B, Princeton Instrument, Trenton, NJ). Hardware binning is performed on the CCD to produce a 256 x 256 pixel image, where each binned pixel is a sum of 16 camera pixels. Binning is per-

formed to increase the signal strength per picture element and to reduce the readout time of the image. The digitization rate is set to 100 kHz, the lower-speed/lower-noise selection available via software. The read noise measured by the manufacturer for the camera is 4.19 e- rms (about 4.28 photons rms, vide infra). The gain on the camera is set to its highest setting of 0.98 e-/AD unit for low light conditions. The magnification of the fluorescence image projected onto the camera was determined with a 1951 USAF glass slide resolution positive target (Edmond Optics, NT36-275, Barrington, NJ) and was found to be 25.90 ± 0.16 .

The sample flow cell is oriented so that fresh sample enters at the bottom and flows upward to the outlet. This allows for air bubbles to pass easily through the flow cell. The area of the flow cell that is imaged onto the camera is given by the area of the detector (177 mm^2) divided by the square of the magnification, or 0.264 mm^2 . The effective depth of field is determined by observing whether a streak returns to baseline between each filter element during the rotation or not. If a phytoplankter is too far outside the plane of focus, then the circle of confusion of the organism's image extends into the dark spaces between the filter elements and makes the integration of fluorescence intensity for a specific streak difficult. Thus, the effective depth of field is related to the size of the organism and the flow rate, with an approximate geometric optics value of 40-80 micrometers for small particles and a sample volume of 0.0105 - 0.0211 ($0.264 \text{ mm}^2 \times (40\text{-}80 \text{ um})$) mm^3 .

The rate at which this volume passes through the image plane of the detector is determined from the number of rotations of the filter wheel observed during passage of a phytoplankter through the flow cell. For instance, at 500 rpm or 8.33 Hz, a full cycle of the filter wheel from one dark reference mark to the next takes about 120 milliseconds. In most cases, the filter wheel frequency and flow rate are set so that a phytoplankter crosses about 60% of the field of view during filter wheel rotation. Thus, 60% of a sample volume is displaced in 120 milliseconds in this example, for a

volume sampling rate of $0.0211 \text{ mm}^3 \times 0.60/0.120 \text{ sec} = 0.106 \text{ mm}^3/\text{sec}$, or $1.06 \times 10^{-4} \text{ mL/sec}$. Increasing the filter wheel frequency permits an increase in volumetric flow rate, but decreases the signal for a single organism due to reduced residence time in the illuminated volume.

Blank/Flat-Field Correction

All CCD images collected use an integration time of 1 second. A "blank" sample of filtered seawater is collected for subtracting the instrument and filtered seawater fluorescence from the raw data sets. Raw data images are corrected for the instrument and filtered seawater fluorescence background by averaging 100 images of filtered seawater acquired with the flow rate and filter wheel frequency matching that of the raw data set.

A flat-field correction is applied to a raw data set to compensate for spatial heterogeneity of the excitation irradiance profile in the sample cell. A flat-field data set is acquired using a dense culture of a small ($\sim 5\text{-}10 \text{ }\mu\text{m}$) phytoplankton (e.g. *E. huxleyi*) that is passed through the flow cell at a constant flow rate and filter wheel frequency (e.g. 0.1 mL/sec and 500 rpm). The flat-field image is acquired with illumination modulated by the filter wheel containing no filters. Each flat-field image contains hundreds of phytoplankton cells resulting in thousands of streaks, making it necessary to average 1500 images to produce a suitably smooth flat-field profile for correcting the raw data. Dividing each pixel in the averaged flat-field image by the average pixel value across both dimensions normalizes the flat-field array.

Measurement scheme

The measurement scheme for the instrument is illustrated in Fig. 3.3, where a partial timing diagram is shown for observation of a single phytoplankter, as well as an illustrated image containing hypothetical data for that organism. In Fig. 3.3A, the

timing diagram portion of Fig. 3.3, a time axis is provided that assumes the opening of the camera shutter occurred at time $t=0$, and that it will close at time $t=1$ second. The portion illustrated in Fig. 3.3A covers a small part of that complete window, as illustrated by the top panel labeled "CCD Integration" that extends both to the left and right of the portion illustrated.

During the 1-second integration of the camera, under most conditions, few particles are observed. Single particles enter the view of the camera randomly, but pass through it with a fixed velocity. In this illustration, a phytoplankton enters the field of view of the camera at about time $t=0.186$ seconds and exits about 0.078 seconds later. During this time it first appears at the entrance of the visible region and crosses the entire field of view of the camera, finally exiting on the opposite side of the visible region. The camera shutter remains open and integrating throughout this transit of the flow cell by the phytoplankton. During the transit time, the filter wheel is continuously turning with a rotation period of about 0.049 seconds in this illustration. During each rotation, five filters are brought into the excitation path in sequence, labeled with numbers 1-5 in the illustration. Between filter 5 and filter 1 is a blocked filter opening labeled B in the illustration, during which the phytoplankton is not illuminated. The timing diagram illustrates that there is no synchronization between the camera shutter operation, the transit of a phytoplankton, or the rotation of the filter wheel with the exception that (a) the CCD shutter is open at least as long as the transit time of a phytoplankton, and (b) the rotation period of the filter wheel should be somewhat less than the transit time of the phytoplankton so that at least a full revolution of the filter wheel occurs during its transit through the field of view of the camera.

Figure 2B illustrates how the "streaks" of fluorescence might appear in an image. The image shown here illustrates the tracks of two individual phytoplankton cells that were captured during 1 second of integration by the CCD. The track represented in

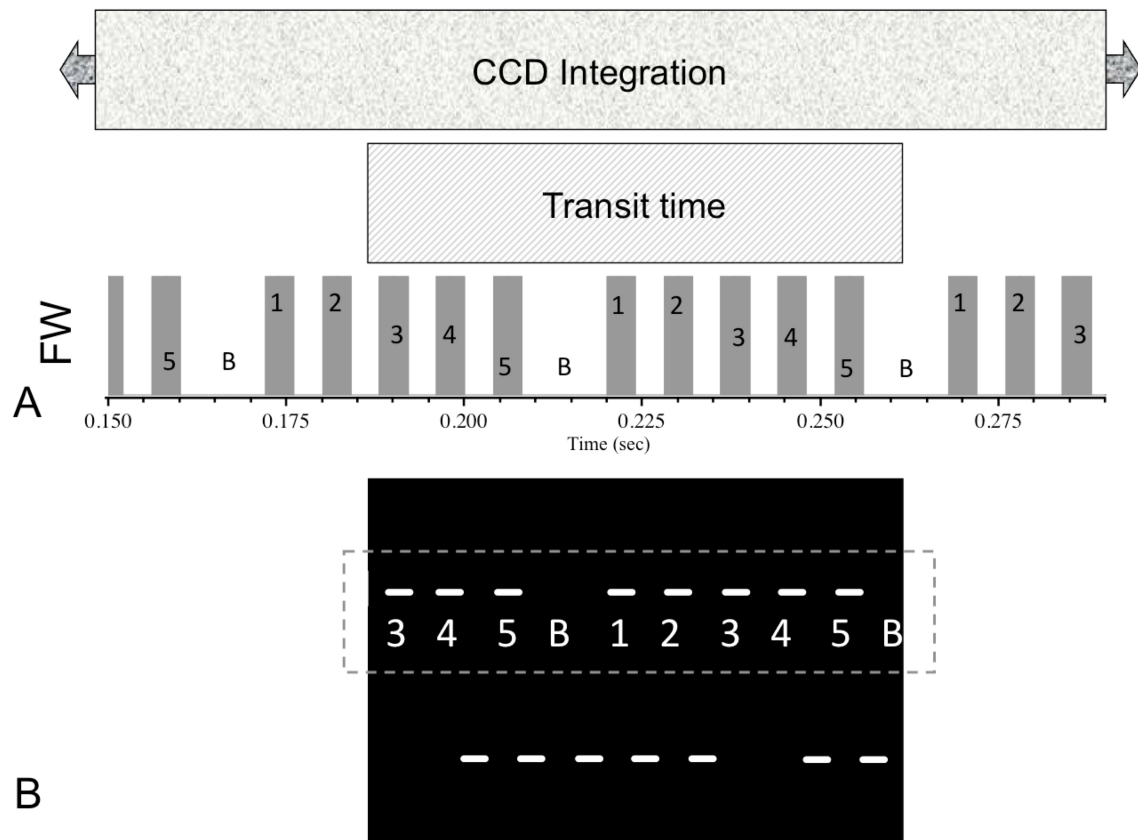


Figure 3.3 Measurement scheme for collecting data using fluorescence imaging photometer. (A) The x-axis represents the exposure time of the camera in seconds, where at $x=0$ the shutter opens and $x=1$ the shutter closes. The top graphic ("CCD Integration") illustrates the CCD integration time extends well before and after the timeline shown. The center graphic ("Transit time") illustrates the time a phytoplankter takes to pass through the region being imaged by the CCD. The lower graphic ("FW") illustrates the position of the filter wheel in respect to the excitation beam as a function of time, with each filter opening being shown for filters 1-5 and a blocked opening marked "B". This timing diagram is for a phytoplankter that enters the image volume 186 milliseconds after the camera shutter opens, at a moment when filter 3 of the filter wheel is starting to line up with the excitation beam. During the approximately 75 milliseconds required for the phytoplankton to traverse the image volume, the filter wheel completes a little over 1.5 rotations. (B) An example of the tracks that two individual phytoplankton cells would produce in the CCD image. The track surrounded by a dashed box is aligned and scaled relative to the graphics in part A to emphasize the relation between the observed track and the timing diagram. The track below the dashed box is an example of a phytoplankter that has entered the image area at a different time and is unrelated to the track in the dashed box.

Fig. 3.3A is surrounded by a dashed box. The image is sized and aligned with Fig. 3.3A so that it is easier to see how the filter wheel cycling is related to the fluorescence that is emitted by each organism. Regardless of the position of the filter wheel when a phytoplankter enters the view, the position of the blank in the wheel is readily found, either by eye or by an automatic algorithm. Once this reference position is found, then a filter is easily assigned to each streak by counting in either direction from the blank position in the image. Streaks that are completely within the area of the image are examined, some filters may produce more than a single streak in an image (in the example, filters 3, 4 and 5 appear twice in the example).

In summary, images captured by the CCD record 1 second worth of events in the mode of operation described here and in Pearl et. al.[19] Typical transit times are near 150 milliseconds under normal flow conditions corresponding to filter rotation frequency near 7 Hz (about 400 rpm). Multiple tracks can be recorded during the 1-second integration time and the total number of tracks is related to the concentration of phytoplankton cells in the water sample. There is nothing synchronous about the measurement, so there is nothing preventing a track from being in progress when the integration time begins, or being incomplete when the integration time ends, except that the length of an individual track relative to the integration time makes it statistically unlikely.

Sample Preparation

For these experiments, we used unialgal cultures of *E. huxleyi* (strain 375) obtained from the Center for the Culture of Marine Phytoplankton (CCMP; recently re-named National Center for Marine Algae and Microbiota) at the Bigelow Laboratory for Ocean Sciences, East Boothbay Harbor, Maine. The phytoplankton cultures were grown in 0.2 μm filter-sterilized f/2 culture medium, salinity ~ 35 , at 23°C . [9, 8] Illumination was provided from the top and sides at an irradiance of approximately

80 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$ on a 12:12 light:dark cycle.

The fluorescence imaging photometer was set up as described above except that no MOEs or blocks were placed in the filter wheel. The culture of *E. huxleyi* was used as a test organism to determine the uniformity of fluorescence measured as a single phytoplankter flows past the image area. In order to maintain uniform excitation irradiance similar to that anticipated with use of MOEs, a 0.3 OD neutral density filter (Newport) was added to the excitation path filter set. For a given filter wheel frequency, an aliquot of the culture was diluted and passed through the flow cell at a flow rate that allowed a single cell to be modulated into 9-10 individual streaks, forming a track. Filter wheel frequencies used in this study were 4.17, 8.33, 16.7, 25.0 and 33.3 Hz. A total of 2000 images were collected at each filter wheel frequency.

Streak and Track Delineation

An in-house Matlab TM(R2008b, Mathworks, Natick, MA) program was used to identify "good" tracks produced by a single *E. huxleyi* cell and to sum the streak intensities. The Matlab code for this program is in Appendix A. A "good" track was defined as one containing 9-10 complete streaks with distinct boundaries to each streak. The track highlighted by a black box in Fig. 3.4A is an example of a "good" track. The program functions by importing the raw images and using a flat-field correction and blank subtraction (detailed above) to correct the images for small irregularities in the excitation plane as well as background fluorescence. The corrected images are then processed one at a time by the user who selects images that appear to be in focus. This initial screening of the data resulted in about 100 images out of the 2000 that passed visual inspection. From these roughly 100 images for each filter wheel frequency, 10 representative frames were selected that were similar to one another in average streak intensity and focus to use for calculating the variation of integrated streak intensities. This process may introduce bias in cases of low average

signal strength or high frequency filter wheel settings, which we explain in further detail in the discussion.

For each image, the integrated streak intensities were found by manually selecting the top and bottom rows of a track in the 256 x 256 pixel image. Typically, the fluorescence images of *E. huxleyi* streaks are about 10-12 pixels wide in the row dimension. The rows in each column were then summed along the columns to produce a 256 row vector. This row vector is a cross-section of the track, with each peak representing a streak. A baseline correction is applied to correct for drift of the baseline (detailed below), then the column boundaries are identified for each peak and the values between the boundaries are summed to find the integrated intensity for each streak. The column boundaries of each peak are determined by identifying columns in which the value exceeds 3 standard deviations of the baseline, and then the average distance between adjacent peaks is found and used as the peak boundary. Continuous columns bracketed by the peak boundaries are summed and stored as the streak intensities for the track selected in each image.

The baseline subtraction routine used for correcting for slope and offset in the baseline operates as follows. First, only the y-values enter the routine. An x-axis is created that is centered so that $x=0$ corresponds to the point in the y-array with the highest intensity. The top 50% of y intensities and their corresponding x values are removed from the raw data in Fig. 3.4B before proceeding to the fitting process. The remaining data array enters a loop. In this loop, a 3rd order polynomial fit to the remaining data is performed, and an array of deviations from the fit is created. If the maximum positive error is greater than 2.2 times the average absolute value of all the errors, then the point containing the highest absolute error is removed from the remaining data. This loop repeats until there are no points that error more than 2.2 times the average absolute error based on the most recent polynomial fit. Then the loop ends and the final curve fit parameters are used with the original x-axis values

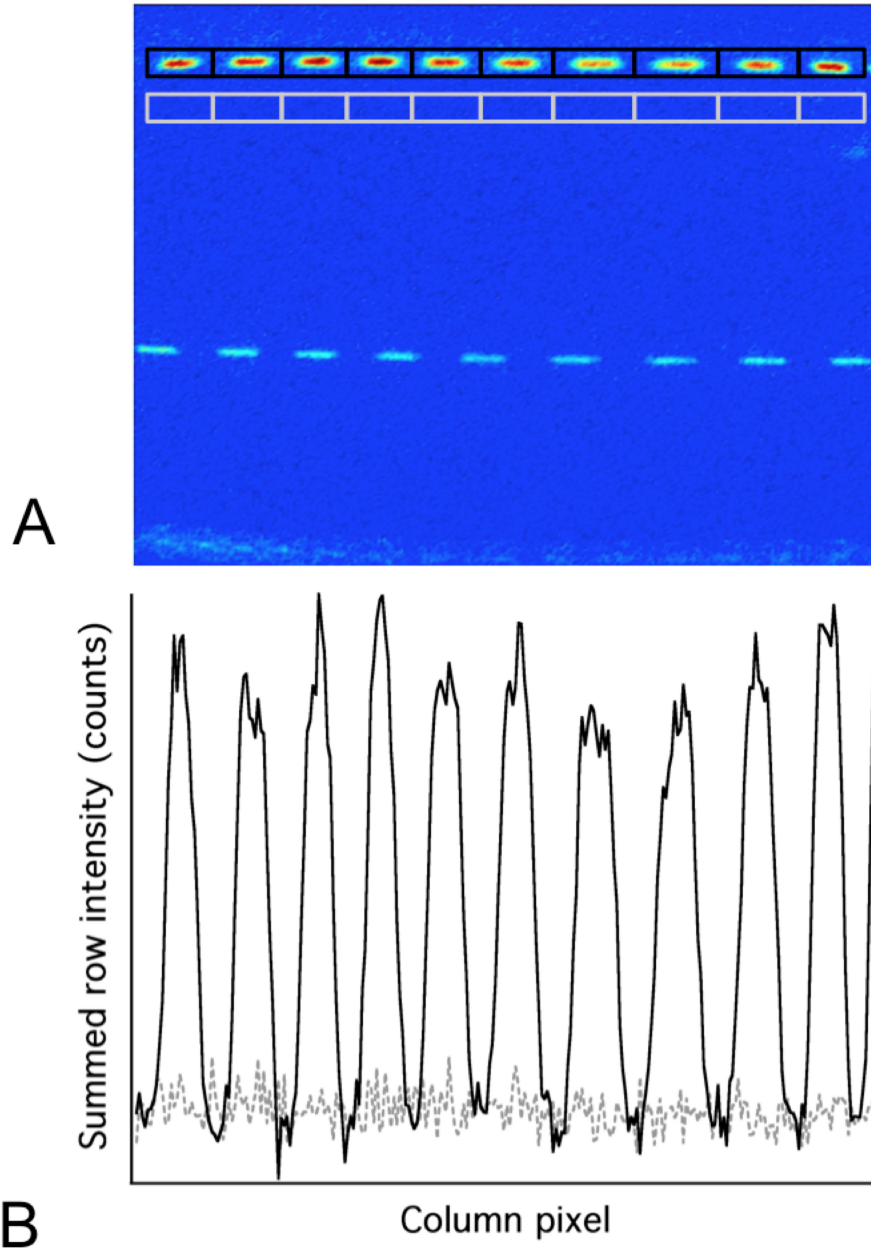


Figure 3.4 Example image from the fluorescence imaging photometer. (A) The image contains a track of *E. huxleyi*'s streaks where there are no filters placed in the filter wheel. Regions of interest for each streak in the track are illustrated (black outline), along with corresponding regions of the same size just below the phytoplankton track (gray outline) that were used for image background measurements. A second *E. huxleyi* track appears in the middle of the image but is unrelated to the *E. huxleyi* track of interest. (B) Plot of summed rows from example phytoplankton track and reference track areas shown in A. This plot shows the summed fluorescence intensity of the rows between the upper and lower row boundaries of the given track.

to create a baseline for the original y-axis values. The standard deviation of the last error vector is calculated and multiplied by 3 to serve as a threshold value that can be used in finding the edge of the peaks in Fig. 3.4B. The baseline subtracted y-values and this threshold value are then returned to the calling program.

3.3 RESULTS AND DISCUSSION

Experimental Results

Figure 3.4 A shows an example image with a filter wheel rotation frequency of 8.33 Hz. The uppermost horizontal row of highlighted regions in Fig. 3.4 A (black) illustrates the regions of interest for a particular track made by a single *E. huxleyi* cell. The lower horizontal row of highlighted regions in Fig. 3.4 A (gray) illustrates regions of the image that are used to estimate the variability in the "image background" with the same pixel area and number of streaks as the *E. huxleyi* track. Figure 3.4 B shows the plot of the row sums of each column in both of the regions of interest in Fig. 3.4 A. Both traces in Fig 3.4 B have been baseline subtracted.

At each filter wheel frequency, integrated intensities were calculated for a total of 90 streaks for 10 *E. huxleyi* cells selected manually from the images as representative. A total of 90 background signals were likewise extracted from the 10 images by offsetting the regions of interest from the phytoplankton streaks as described in the discussion above and for Fig. 3.4 B. The statistics shown in Tables 3.1 and 3.2 were calculated from these 180 values at each filter wheel frequency.

The second column of Table 3.1 shows the average intensity recorded for the 90 *E. huxleyi* streaks at each frequency. Values in the third and fourth columns were obtained by taking the integrated intensities of 9 streaks for a single phytoplankton track and calculating the single-organism average and standard deviation. The third column is the standard deviation of the 10 single organism track averages at each fil-

Table 3.1 Streak statistics. For each filter wheel rotation frequency, the average integrated area of 90 streaks (9 streaks each for 10 *E. huxleyi* tracks) is given. s.d. (inter) is the standard deviation of the average streak intensities for 10 phytoplankton tracks. s.d. (intra) is the average of the standard deviations of 9 streaks in each track. SNR (intra) is the average signal to noise ratio. More details are found in the text.

| Hz. | I_{avg} (counts) | s.d. (inter) | s.d. (intra) | SNR (intra) |
|-------|--------------------|--------------|--------------|-------------|
| 4.67 | 43.4 | 14.1 | 1.67 | 31.0 |
| 8.33 | 25.2 | 13.3 | 1.37 | 19.0 |
| 17.00 | 14.0 | 5.64 | 1.17 | 12.5 |
| 25.33 | 11.2 | 6.47 | 1.23 | 8.91 |
| 33.33 | 8.25 | 1.99 | 0.959 | 9.97 |

Table 3.2 Background statistics. For each filter wheel frequency, the average integrated area of 90 regions of equivalent size to those in Table I, but chosen from a nearby region of background near each phytoplankton track, is provided in column 2. Column 3 gives the standard deviation of the average streak-area intensities for ten background "tracks". Column 4 gives the average of the standard deviations of 9 streak areas in 10 background "tracks".

| Hz. | I_{avg} (counts) | s.d. (inter) | s.d. (intra) |
|-------|--------------------|--------------|--------------|
| 4.67 | 61.9 | 403 | 937 |
| 8.33 | 72.4 | 266 | 815 |
| 17.00 | -25.1 | 466 | 819 |
| 25.33 | 73.6 | 418 | 1030 |
| 33.33 | 55.5 | 246 | 734 |

ter wheel frequency. It is labeled "inter" because it describes the variability between different *E. huxleyi* cells. The fourth column is the average of the 10 single organism standard deviations at each filter wheel frequency. It is labeled "intra" because it describes the variability when measuring an individual *E. huxleyi*. The fifth column, SNR, is calculated by dividing the average signal for a single *E. huxleyi* by the standard deviation of the integrated intensities of the 9 streaks for that *E. huxleyi* cell. Values for all 10 *E. huxleyi* at the same filter wheel frequency are then averaged.

Table 3.2 shows values for the background regions of interest in the images where the *E. huxleyi* data presented in Table 3.1 were collected. The second column of this Table shows the average values of 90 equivalent streak windows in the background for comparison to the nearby *E. huxleyi* streaks. The third column is the standard deviation of the average values in the 10 background tracks; the fourth column is the average of the standard deviations of the 10 background tracks.

Instrument Signal and Noise Characteristics

Baseline-Subtracted Background Intensity

The first observation from the data in Tables 3.1 and 3.2 is that the average baseline-subtracted background intensity is statistically indistinguishable from zero. For each filter wheel frequency, the average background signal in Table 3.2 column 2 is well within the 95% confidence interval of zero calculated from the standard deviation in column 3. This is significant because the signals in these background regions are subject to the same data treatment as the signals of the phytoplankton, and show no obvious bias from errors in the algorithm used for processing them.

Single-*E. Huxleyi* Fluorescence Rate

The second observation from the data in Tables 3.1 and 3.2 is that the average intensity of the fluorescence in a streak scales directly with the illumination time.

Figure 3.5 shows the average integrated intensity of a streak as a function of the reciprocal of the filter wheel frequency. The error bars shown in Fig. 3.5 are 95% confidence intervals of the mean based on the standard deviations in column three of Table 3.1 . Calculating the normalized value of χ^2 is a common test for goodness of fit; it is calculated to be 0.1 for these data when the standard deviation of the mean is used in the calculation.[14] This suggests that the apparent uncertainty in each of the data points is being overestimated, perhaps because the distribution of intensities is not normal (vide infra), or because of uncertainty in the baseline that is unrelated to the true intensity.

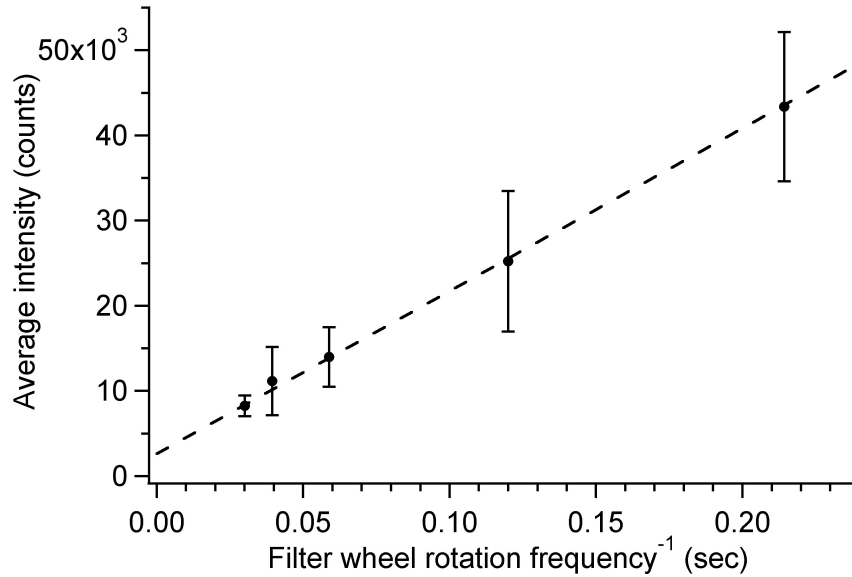


Figure 3.5 Average streak integrated fluorescence intensity as function of inverse filter wheel frequency. At each frequency, 90 fluorescence streaks were averaged (9 each for 10 phytoplankter tracks). Error bars are 95% confidence intervals of the mean. The fitted line has a slope of $(1.91 \pm 0.21) \times 10^5$ counts/sec and an intercept of $(2.6 \pm 1.0) \times 10^3$ counts. When forced through zero, the slope of the best fit line is $(2.09 \pm 0.03) \times 10^5$ counts/sec, where all error limits are estimated standard errors. The normalized value of χ^2 for the plot shown is approximately 0.1.

The intercept of the fit in Fig. 3.5 is distinctly non-zero, which could be due to defects in the baseline fitting and integration routines. In that case one would expect the values for "background" measurements to be subject to the same error, but there

is no evidence of such an error. A second and more likely source for the offset in the y-intercept is bias in selecting *E. huxleyi* tracks. In general, the visual approach to selecting *E. huxleyi* tracks works well when the streaks are easy to pick out in the images - i.e., when their average SNR is high. But when the average SNR is low, both manual and automated track selection tends to under-represent *E. huxleyi* tracks of low average intensity, leading to a bias toward higher intensity. In fig. 3.5, this affects the points to the left of the graph much more than those to the right, leading to a positive y-intercept.

When the intercept is forced through zero, the slope of the data in Fig. 3.5 is $(2.09 \pm 0.03) \times 10^5$ counts/sec. The actual individual streaks are a fraction of the full wheel revolution, since each open window occupies about 8.2% of the circumference of the filter wheel at the radius of the windows. Thus the detected fluorescence intensity of the average *E. huxleyi* in these studies is $\sim 2.5 \times 10^6$ counts per second of illumination. The value expected in a real experiment would depend on whether the optical filters placed into the filter wheel have an average transmission that is more or less than the 50% value of the neutral density used to simulate them for this study.

Intra - vs. Inter- *E. huxleyi* Cell Variability

The s.d. (inter) is up to an order of magnitude larger for measurements made in *E. huxleyi* streaks than s.d. (intra). Generally both values decrease with increasing filter wheel frequency and decreasing average integrated intensity, and they diverge most strongly for the lowest frequency/highest intensity. The difference between s.d. (inter) and s.d. (intra) occurs because the tracks represent different *E. huxleyi* cells that often have very different average intensities. The first chapter reported the variation in integrated fluorescence intensity for 77 *E. huxleyi* cells from a monoculture of the type used here. In that work, we found the integrated intensity ranged over a factor of 7.7. Here the factor ranges from 2.5 to 5 in the sets of 10 *E. huxleyi* cells

at each filter wheel frequency, depending on the dataset, the range perhaps being reduced somewhat by intentional selection of "typical" *E. huxleyi* tracks. The large spread of individual *E. huxleyi* cell intensities accounts for the remarkably large value of s.d. (inter), but has no effect on s.d. (intra) since all the measurements compared with that statistic are made on a single *E. huxleyi* cell.

Streak vs. Background Variability

The top graph in Fig. 3.6 shows a comparison of s.d. (intra) for background and *E. huxleyi* tracks at our five different filter wheel frequencies, plotted against the inverse of the filter wheel frequency. Each point represents the average s.d.(intra) measurement for 10 tracks, with error bars representing its 95% confidence interval. A small horizontal offset has been applied to the background values (solid circles) to make it easier to see which error bars go with which points. In all cases the background values are lower than the *E. huxleyi* values, but they begin to converge toward the left in the figure. This is expected because as the filter wheel frequency increases to infinity (on the left axis of Fig. 3.5), the illumination time for a single streak approaches zero and the *E. huxleyi* streak becomes infinitesimally different from a background measurement. As expected, the background measurements are not significantly different from one another at different filter wheel frequencies.

Because the experiments at different filter wheel frequencies were done separately - with new background images, dark frames, and long periods of flushing the flow cell between sets - the apparent phytoplankton variability and the background variability are not perfectly consistent across the plot on top of Fig. 3.6. However, the background measurements were made from data in the same CCD image as the phytoplankton measurements, so they share some of the same variability. The result is that the *E. huxleyi* streak variability moves with the background variability to an extent.

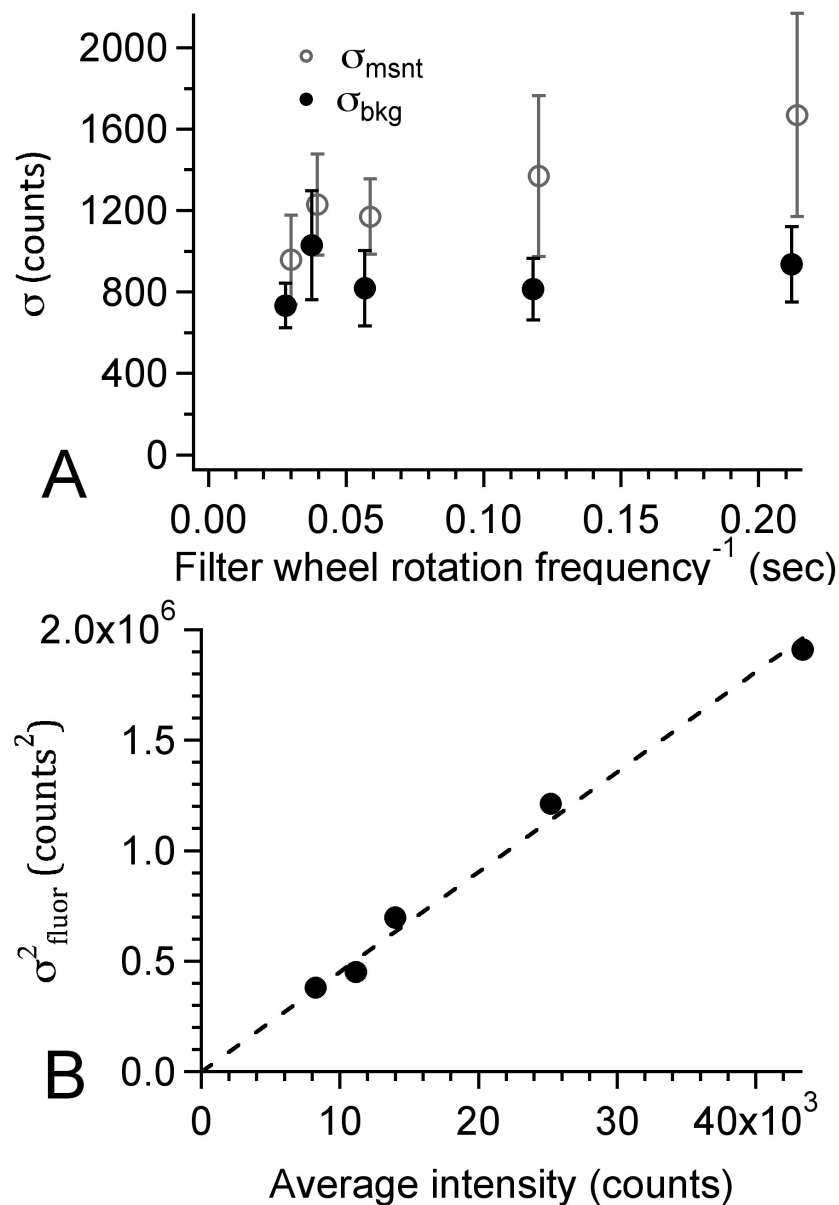


Figure 3.6 (A) Single-streak measurement variability averaged for 10 *E. huxleyi* and 10 equivalent background regions. Open circles represent *E. huxleyi* streaks and solid circles are equivalent background streaks. Error bars are 95% confidence intervals for the single-streak standard deviations. (B) Variance of single-*E. huxleyi* fluorescence signals, correcting for background variance, based on top figure. The fit is of the form $y=ax$, where $a = 45.2$ counts.

The primary goal for this imaging photometer is to meet the minimum SNR requirements set forth in chapter I. In that chapter we show that a minimum SNR of 19 is theoretically necessary to distinguish *E. huxleyi* from *T. pseudonana*. Table 3.1 shows that this criterion is met for the two lowest filter wheel frequencies tested. The highest SNR reported in Table 3.1 is 31.0 which corresponds to a RMS noise level of 3.2%. To understand the contribution of mechanical and electrical defects in the instrument itself to the intra-track SNR reported in Table 3.1, we measured several relevant parameters.

Quantitative Analysis of Mechanical and Electrical Sources of Noise

Xe arc lamp ripple

The Xe arc lamp intensity was tested for ripple using a silicon photodiode measuring the intensity after passing through the optics of the instrument, but without modulation by the filter wheel. The ripple was estimated at 0.843% RMS with the dominant frequency being approximately the 60 Hz line frequency. For measurement times longer than cycle of the line frequency the RMS noise is expected to be lower. Xe arc lamp ripple may become a limiting factor if the SNR exceeds 100.

Aperture to Aperture variability

The once-around aperture variability would be defined as the variability in streaks due to the same aperture on subsequent cycles of the wheel, separated from lamp flicker/ripple and drift in the filter wheel speed. These could not be easily decoupled in our experiments. If they could be, the once-around variability for a constant-speed, perfectly stable light source would be determined by the mechanical stability of the filter wheel, which is relatively high. We therefore believe the once-around variability

to be a negligible contribution to the track SNR.

Different apertures on the filter wheel, however, might be in slightly different physical positions, with different diameters and edge shapes. This variability does not matter in the intended application because each MOE is viewed through a single aperture. But in the data reported here, all apertures were treated as if they were identical. An estimate of the aperture variability, assuming constant speed and a constant lamp, was made by placing a detector in the position of the flow cell and recording data with an oscilloscope, averaging over 128 cycles to average out speed and lamp variability. In each experiment, all apertures were covered except the one being tested. Experiments were repeated for all six apertures sequentially, and the integrated intensity detected by a silicon photodiode for each aperture was calculated. All apertures were found to be very similar, with a relative standard deviation of 0.68% or better between apertures. Aperture to aperture variability is always a negligible contribution in normal usage, but could limit the apparent SNR in these studies to approximately 150.

Variability in filter wheel frequency

Drift in the filter wheel speed can also give rise to variability in the total illumination during a streak. The drift was generally found to be larger in relative terms for lower speeds. The worst was 3.46% peak-to-peak for a filter wheel frequency of 4.66 Hz, with a sinusoidal oscillation in frequency having a period of 2.6 seconds. At this filter wheel frequency, an *E. huxleyi* cell will typically pass through the observation window in about 0.36 seconds. Under the most unfavorable conditions (an *E. huxleyi* cell passing through at the moment of greatest rate of change in the filter wheel frequency), the frequency would be expected to change by 2.9% across the image, with the average measurement showing a relative standard deviation of 0.98%. The deviations from the average for the filter wheel frequencies of 8.33, 17.00, 25.33 and 33.33 Hz did not

vary as a simple sinusoid and had a P-P relative deviation of approximately 0.68%, 0.17%, 0.19% and 0.17%, respectively. Assuming a rms deviation from the average that is 2-1/2 of the peak deviation, the relative standard deviations at these higher frequencies are estimated to be 0.24%, 0.06%, 0.07% and 0.06%. Varying filter wheel frequency could limit the SNR to approximately 100 at the lowest speed, or 400 at the second lowest.

Detector Read and Dark Noise

Dark measurements made over a second of integration time were used to determine the contribution of the dark or thermal noise of the camera to our measurements. The standard deviation of the signal on an individual pixel in our images due to the combination of dark and read noise was found to be 5.2 counts, a negligible contributor to s.d. (intra). Since this is a source of absolute noise, it places no fundamental limit on SNR. For the lowest filter wheel frequency, read and dark noise would limit the SNR to no more than 800 based on the signal recorded for that condition and estimating 110 pixels for a streak area.

Flat-Field correction error

Because of varying illumination intensity across the field of view, raw fluorescence intensities are compensated with a flat-field correction, which is itself determined by measurement of cultured phytoplankton fluorescence. If the flat-field correction applied deviates from the "true" flat-field, identical signals measured in different regions of the image will appear to have different intensities which we will interpret as noise. We estimate the uncertainty in the integrated intensity of a single streak due to flat-field correction error to be 0.6% based on variability in repeated measurements of the flat field correction. This would limit SNR to no more than 170.

Background Fluorescence and baseline subtraction

Background fluorescence and other interference can also lead to noise, so 100 background images were collected and used to estimate the single-pixel variance from all sources therein. The average single-pixel variance of the background was found to be 1530 counts squared (a single pixel s.d. of 39 counts). The average background fluorescence signal on a single pixel was found to be 1390 counts, suggesting that most of this background fluorescence noise simply comes from photon counting error.

For a streak area of 110 pixels, and assuming background subtracted streak images, background fluorescence by itself would account for background intra-track variance of 2.4×10^5 counts squared, or a contribution to s.d.(intra) in Table 3.2 of about 490 counts. This represents about one-third of the observed variance between background streak areas for the same track. Much, if not all, of the remaining background intra-track variability can be attributed to baseline subtraction error since the baseline is found from the pixel intensities between the streak areas that are themselves noisy.

It is likely that the precision of the baseline subtraction can be improved by reducing the background fluorescence observed in the instrument. The apparent intra-track phytoplankton fluorescence noise can also be improved this way whenever it is dominated by background fluorescence noise. As we show below, this occurs at high filter wheel frequencies, but is not the limiting factor at low frequencies.

If the background fluorescence cannot be reduced it will limit the SNR at the lowest filter wheel frequency (highest signal strength) to no more than 75. Background fluorescence is the second most important source of error for MOE measurements with this instrument. Another factor becomes more important at low filter wheel frequencies as we show below.

Photon Counting Error

Figure 3.6B allows us to evaluate the contribution of Poisson photon counting noise, which sets the lower limit to noise in a photon measurement. Our noise is considerably higher than this limit. For example, at a 4.67 Hz filter wheel frequency, the average streak intensity is 43,400 counts. At the gain setting of the camera, this is approximately the number of photons detected in an average streak. The square root of this value would be the expected fluorescence noise due to photon counting - 208 counts. The actual value is 1380 counts, nearly a factor of 7 greater. The observed dependence of our fluorescence noise on intensity does, however, follow the trend expected of photon counting noise, with the variance proportional to the average intensity (*vide infra*).

Variance of Fluorescence in Measurement

The variability observed in s.d. (intra) for phytoplankton streaks always exceeds the variability of the background, and appears to diverge as the illumination time and the fluorescence intensity increase. If we assume the fluorescence variability of the phytoplankton is independent of the noise in the background, then we can assert that:

$$\sigma_{msmt}^2 = \sigma_{fluor}^2 + \sigma_{bkg}^2 \quad (3.1)$$

where σ represents the intra-track standard deviation, and the subscripts msmt, fluor and bkg represent the standard deviation of a streak measurement (msmt), in the phytoplankton fluorescence signal itself (fluor), and in the background (bkg).

Figure 3.6 A gives s.d.(intra) for both the phytoplankton and background tracks, so we can compute the variance of the fluorescence from Eq. 3.1 by rearrangement. Figure 3.6 B shows the result for the square of the fluorescence standard deviation - the fluorescence variance - as a function of the average intensity at each filter wheel frequency. The data give a linear fit whose y-intercept is indistinguishable from zero

when performed with no constraints. The best-fit line shown in Fig. 3.6 B is the fit with the y-intercept constrained to exactly zero. This appears to show that the variance is linear with fluorescence intensity, but it is equally true that the variance is linear with the length of time a phytoplankter is excited in each streak because of the way the data are measured.

We have shown above that the background variance is independent of the filter wheel frequency. Since the phytoplankton fluorescence variance increases with decreasing filter wheel frequency, background noise will dominate SNR at high filter wheel frequencies, with phytoplankton fluorescence noise dominant at low filter wheel frequencies. The average variance in the background calculated from Fig. 3.6 A is 0.76×10^6 counts squared. The slope of the best-fit line shown in Fig. 3.6 B is 45.2 counts. From the latter slope, the fluorescence variance will equal the average background variance when the fluorescence intensity equals 17,000 counts. Using the slope of Fig. 4 when the intercept is forced through zero, we find that a fluorescence intensity of 17,000 counts corresponds to a filter wheel frequency of 13 Hz. This represents the crossover frequency below which fluorescence noise dominates SNR. Of the five filter wheel frequencies tested, the two with the highest SNR are therefore limited by the variance of the phytoplankton fluorescence. These are also the only two conditions tested that meet the SNR requirement in chapter 1. Further improvement of the photometer's performance requires some better understanding of the source of the phytoplankton fluorescence noise.

Possible sources of noise and apparent noise in the phytoplankton fluorescence signals include: inadequate flat fielding; the *E. huxleyi* cells moving in and out of focus during transit; interference effects in the flow cell; photoacclimation of the *E. huxleyi*; unstable filter wheel rotation frequency; lamp ripple; inter-aperture throughput variability; photochemical decay of the pigments in the *E. huxleyi* cell; non-linearity of the detector; Brownian rotation of the *E. huxleyi* cell during transit of the image

region; and/or some other process.

Qualitative analysis of phytoplankton fluorescence noise

Any source of phytoplankton fluorescence noise or apparent noise would have to explain the magnitude of the observed noise and also the linear dependence of variance on signal shown in Fig. 3.6B. The first 8 of the sources in the preceding paragraph can be rejected because they would lead to a standard deviation of fluorescence intensity that varies directly with the fluorescence intensity.

Nonlinearity of the detector is not a likely explanation because (a) CCD arrays are quite linear, (b) intensities are not particularly high, and (c) saturation effects, if they were observed, would tend to make the variability at high intensity lower than at low intensity - the opposite of what is observed.

Brownian rotation, which could present different aspects of an *E. huxleyi* cell to measurement and change its apparent fluorescence intensity through orientation dependence of pigment packaging or inner filtering effects, can be rejected as an explanation because both calculation and experimental observation of *E. huxleyi* show the average time for rotation to be more than an order of magnitude longer than the transit time observed here for even our lowest filter wheel rotation frequency.[30] In addition the observed variation in fluorescence intensity for a single phytoplankter given in a previous report from this laboratory is lower than that observed here, even though the observation time was long enough to allow for full rotation.

This leaves us with the somewhat unfulfilling conclusion that some other process is responsible for the fluorescence variability. The main clue that we are given is that this unknown process has the characteristic of having a variance proportional to average intensity or average illumination time. A dynamically varying fluorescence quantum efficiency in the phytoplankton is one possible source of a "noise" of this type. When we think of fluorescence from a phytoplankton, we often implicitly assume that

all the emitting photocenters are equivalent. But if instead most fluorescence comes from a small sub-population of the photocenters, and if that sub-population were in a dynamic equilibrium of some type, the fluorescence variance would scale with measurement time as observed here. Most chlorophyll *a* fluorescence observed from phytoplankton comes from the photosystem II complex, which is only an intermediate step in photosynthesis. If the photosynthetic site (photosystem I) is unavailable for some reason, this could have a large effect on the apparent fluorescence efficiency of nearby photosystem II complexes. We may be able to test this hypothesis by temperature or chemical perturbations to phytoplankton. A study is underway in our laboratory to investigate this question.

3.4 CONCLUSIONS

In chapter 1 we concluded with the required specification of an instrument SNR to enable multivariate optical computing to distinguish *E. huxleyi* from *T. pseudonana*. The necessary value was SNR $\hat{\geq}$ 19. This manuscript has described an instrument suitable for the optical elements designed and fabricated in the previous manuscript that demonstrates SNR of the requisite level. In Pearl et. al[19], we show results for the instrument in real measurements attempting to classify individual *E. huxleyi* and *T. pseudonana*.

BIBLIOGRAPHY

- [1] N. Agnihotri, W. S. Kisaalita, and C. H. Keith. Micro-perfusion flow cell for imaging cultured cells. *Biotechniques*, 27(4):722–726, 1999.
- [2] C. Barnes, X. Irigoien, J. A. A. De Oliveira, D. Maxwell, and S. Jennings. Predicting marine phytoplankton community size structure from empirical relationships with remotely sensed variables. *Journal of Plankton Research*, 33(1):13–24, 2011.
- [3] R. R. Bidigare and M. E. Ondrusek. Spatial and temporal variability of phytoplankton pigment distributions in the central equatorial pacific ocean. *Deep-Sea Research Part Ii-Topical Studies in Oceanography*, 43(4-6):809–833, 1996.
- [4] L. S. Bruckman, T. L. Richardson, J. A. Swanstrom, K. A. Donaldson, M. Allora, T. J. Shaw, and M. L. Myrick. Linear discriminant analysis of single-cell fluorescence excitation spectra of five phytoplankton species. *Applied Spectroscopy*, 66(1):60–65, 2012.
- [5] T. D. Dickey. The emergence of concurrent high-resolution physical and biooptical measurements in the upper ocean and their applications. *Reviews of Geophysics*, 29(3):383–413, 1991.
- [6] T. Fujiki, K. Matsumoto, S. Watanabe, T. Hosaka, and T. Saino. Phytoplankton productivity in the western subarctic gyre of the north pacific in early summer 2006. *Journal of Oceanography*, 67(3):295–303, 2011.

- [7] G. Gorsky, P. Guilbert, and E. Valenta. The autonomous image analyzer - enumeration, measurement and identification of marine-phytoplankton. *Marine Ecology-Progress Series*, 58(1-2):133–142, 1989.
- [8] R. R. Guillard. *Culture of phytoplankton for feeding marine invertebrates*. Culture of Marine Invertebrate Animals. Plenum Press, New York, 1975.
- [9] R. R. Guillard and J. H. Ryther. Studies of marine planktonic diatoms .1. cyclotella nana hustedt, and detonula confervacea (cleve) gran. *Canadian Journal of Microbiology*, 8(2):229–239, 1962.
- [10] B. Hansen, P. K. Bjornsen, and P. J. Hansen. The size ratio between planktonic predators and their prey. *Limnology and Oceanography*, 39(2):395–403, 1994.
- [11] N. Hashemi, J. S. Erickson, J. P. Golden, K. M. Jackson, and F. S. Ligler. Microflow cytometer for optical analysis of phytoplankton. *Biosensors and Bioelectronics*, 26(11):4263–4269, 2011.
- [12] S. W. Jeffrey and M. Veski. *Introduction of marine phytoplankton and their pigment signatures*. Phytoplankton Pigments in Oceanography: Guidelines to Modern Methods. UNESCO, Paris, 2nd edition, 2005.
- [13] H. P. Jeffries, M. S. Berman, A. D. Poularikas, C. Katsinis, I. Melas, K. Sherman, and L. Bivins. Automated sizing, counting and identification of zooplankton by pattern-recognition. *Marine Biology*, 78(3):329–334, 1984.
- [14] S. M. Kendall and A. Stuart. *The Advanced Theory of Statistics: Volume 1 Distribution Theory*. Macmillan Publishing Co., Inc, Cambridge, 1977.
- [15] C. S. Kinkade, J. Marra, T. D. Dickey, and R. Weller. An annual cycle of phytoplankton biomass in the arabian sea, 1994-1995, as determined by moored optical sensors. *Deep-Sea Research Part II-Topical Studies in Oceanography*, 48 (6-7):1285–1301, 2001.

- [16] Z. Kolber and P. G. Falkowski. Use of active fluorescence to estimate phytoplankton photosynthesis in-situ. *Limnology and Oceanography*, 38(8):1646–1665, 1993.
- [17] M. D. Mackey, D. J. Mackey, H. W. Higgins, and S. W. Wright. Chemtax - a program for estimating class abundances from chemical markers: Application to hplc measurements of phytoplankton. *Marine Ecology-Progress Series*, 144(1-3): 265–283, 1996.
- [18] R. J. Olson and H. M. Sosik. A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging flowcytobot. *Limnology and Oceanography-Methods*, 5:195–203, 2007.
- [19] M. R. Pearl, J. A. Swanstrom, L. S. Bruckman, T. L. Richardson, T. J. Shaw, H. Sosik, and M.L. Myrick. Taxonomic classification of phytoplankton with multivariate optical computing, part iii: Demonstration. *Applied Spectroscopy*, 67(6), 2013.
- [20] T. L. Richardson, E. Lawrenz, J. L. Pinckney, R. C. Guajardo, E. A. Walker, H. W. Paerl, and H. L. MacIntyre. Spectral fluorometric characterization of phytoplankton community composition using the algae online analyser (r). *Water Research*, 44(8):2461–2472, 2010.
- [21] R. B. Rivkin and L. Legendre. Biogenic carbon cycling in the upper ocean: Effects of microbial respiration. *Science*, 291(5512):2398–2400, 2001.
- [22] T. P. A. Rutten, B. Sandee, and A. R. T. Hofman. Phytoplankton monitoring by high performance flow cytometry: A successful approach? *Cytometry Part A*, 64A(1):16–26, 2005.
- [23] Joseph A. Swanstrom, Laura S. Bruckman, Megan R. Pearl, Michael N. Simcock, Kathleen A. Donaldson, Tammi L. Richardson, Timothy J. Shaw, and Michael L.

Myrick. Taxonomic classification of phytoplankton with multivariate optical computing, part i: Design and theoretical performance of multivariate optical elements. *Applied Spectroscopy*, 67(6), 2013.

- [24] C. M. Thomas. *Identifying Marine Phytoplankton*. Academic Press, New York, 1997.
- [25] Takashi Tsuji and Takashi Nishikawa. Automated identification of red tide phytoplankton *prorocentrum triestinum* in costal areas by image analysis. *Journal of the Oceanographical Society of Japan*, 40:425–431, 1984.
- [26] D. Uhlmann, H. Mihan, and A. Gnauck. Variations of oxygen-content and of biological structure of extremely eutrophied waters under constant environmental-conditions (model experiments). *Acta Hydrochimica Et Hydrobiologica*, 6(5): 421–444, 1978.
- [27] D. Uhlmann, O. Schlimpert, and W. Uhlmann. Automated phytoplankton analysis by a pattern recognition method. *Internationale Revue der Gesamten Hydrobiologie*, 63(4):575–583, 1978.
- [28] M. L. Yallop. Distribution patterns and biomass estimates of diatoms and autotrophic dinoflagellates in the ne atlantic during june and july 1996. *Deep-Sea Research Part Ii-Topical Studies in Oceanography*, 48(4-5):825–844, 2001.
- [29] C. S. Yentsch and C. M. Yentsch. Fluorescence spectral signatures - characterization of phytoplankton populations by the use of excitation and emission-spectra. *Journal of Marine Research*, 37(3):471–483, 1979.

CHAPTER 4

SIMOC IMAGE ANALYSIS AND DEBLURRING

ALGORITHM

4.1 INTRODUCTION

Many instruments for phytoplankton classification are based on either image analysis or fluorescence spectroscopy.[3, 4, 6, 7, 8, 10, 11, 12]. Automated interpretation of data from these instruments involves development of algorithms for image analysis or spectral analysis.[3, 4, 6, 7, 11, 6, 13, 14, 15, 16, 2] Research in imaging-based classification is often confounded by similarities in morphology between phytoplankton species.[3] Wide ranges of methods and approaches have been applied to automatic analysis with varying levels of success at overcoming this and other problems.[16]

Fluorescence excitation spectroscopy provides an alternative approach. Because of its sensitivity to accessory pigments, fluorescence excitation spectroscopy provides classification using variables that differ from those used in image analysis and provide information about taxonomic status of a phytoplankton cell that is independent of morphology. We are exploring an approach to phytoplankton classification that combines some of the power of imaging with fluorescence excitation spectroscopy to classify phytoplankton. In this chapter, we focus on the automatic analysis of spectroscopic content in images from a fluorescence imaging photometer described in chapter 3.

The main features from the fluorescent images are the integrated intensities of each streak within a selected track and the width of the streaks. Using the blank position in

the filter wheel as a reference, each streak can be assigned to a filter in the filter wheel. The ratios of the streak intensities are then used to assign the phytoplankter track a species or class as described in chapter 2. Along with the spectral characteristics, we can also extract physical characteristics of the phytoplankter in a given track. The physical size of the chloroplast, region of the cell containing the chlorophyll. This added physical feature can then be used as an additional measure for classification as the relative size of each species is well understood.[5, 16] This is not a trivial task in the measurement scheme of the fluorescence imaging photometer described in chapter 3 due to there being no way to guarantee that the phytoplankton pass through the flow cell in the focal plane of the objective. Here, a deblurring algorithm is described that attempts to deconvolve the image of an ‘in focus’ streaks from the out of focus streaks in an image.

4.2 EXPERIMENTAL

Phytoplankton cultures and Data Collections

The details of the phytoplankton culture conditions can be found in chapter 2. The details of the imaging photometer can be found in chapter 3. The filter wheel of the imaging photometer has 6 openings that rotated counter-clockwise from the prospective of the lamp. One of the filter wheel positions contained a 1-inch substrate painted black to block the light. This blacked out substrate is used as a reference that can be seen in the image to correlate the fluorescent streaks to each of the MOEs. The order of the MOEs as they rotated into the light path is MOE1-, MOE1+, MOE1+, MOE1-, ND. The labels on each MOE, 1+ and 1-, represent MOEs built to mimic the operation of the first linear discriminant vector in the 3-species separation described in chapter 2, *E. huxleyi*, *T. pseudonana* and *Synechococcus sp.* ND represents a 0.3 Newport neutral density filter. Duplicate MOEs were loaded into the instrument to

improve the SNR of the measurement by repetitive sampling.

In the design of MOEs, performance is characterized by the ratio of fluorescence intensity when a phytoplankter is excited through a MOE to fluorescence intensity when excited through a ND. The difference between MOEs labeled 1+ and 1- is that the predicted ratio values increase with the scores of each calibration spectrum on the first linear discriminant for 1+, while the ratio values decrease with the scores for 1-. Since the sign of a linear discriminant is arbitrary, the absolute sign of the MOE has no particular significance. Regardless of the sign of the linear discriminant, however, MOEs with different signs have opposite responses.

For these measurements, the filter wheel was rotated at 6.67 Hz. The pump speed was adjusted to give approximately 9-10 streaks during the transit of a phytoplankter across the field of view. During a set of measurements, files containing 500 16-bit image frames with integration times of 1 second were acquired. A total of 20 such files for each organism constituted the complete sample data set. Additional files characterizing the background, dark count and flat field were also acquired.

Image Preprocessing

All algorithms were written in the MatLab®R2011b 7.13 (Mathworks, Inc., Natick, MA) programming environment and utilized predefined functions in Image Analysis and on a Dell PC running Windows 7.

Preprocessing was performed as follows. Data sets of 16-bit binary image files of 256x256 pixels matrix were read using a MatLab®routine. The first sets of data imported included 3 sets of 500 image frames of a dense culture of *E. huxleyi* to use for a flat-field correction. Details on how the flat field images are acquired may be found in the experimental section in chapter 3.

A corrected and normalized flat field frame, $\langle f \rangle$, was then obtained by subtracting the average background frame from the average flat field frame and dividing the result

by its average pixel value. The normalized flat field measures the distribution of excitation radiation in the image plane.

Sample measurements were typically loaded in 500-frame image data files, along with a 100-frame file of sample background images acquired directly before or after each sample measurement. An average sample background image, $\langle b \rangle$, was calculated from the 100 sample background images.

Each sample image was then processed pixel by pixel according to 4.1 to generate a corrected image.

$$S_c = \frac{S_i - \langle d \rangle}{\langle f \rangle} \quad (4.1)$$

where S_c is a single corrected image and S_i is a single uncorrected sample image.

Streak Integration and Deblurring Algorithm

This section outlines the approach and steps of the algorithm we termed the Streak Integrator for Multivariate Optical Computing (SIMOC, version 2.0), the Mabl@code is shown in Appendix A. After the raw data files are imported and preprocessed as described above. Each image is first analyzed for fluorescent tracks that are well modulated. 4.1a is an example of an image containing three visible track that are nearly overlapping to the left of the image.

A well defined track is determined by finding the average and standard deviation of the 256 rows for each of the 256 columns in an image. Then, a baseline-corrected standard deviation and threshold are calculated as described in chapter 3. 4.1B shows the column standard deviations of the three tracks on the left side of the image in 4.1A.

Track Identification and Streak Integration

The track boundaries are identified by the baseline-corrected standard deviation crossing the threshold. Each column value of the standard deviation (solid line 4.1B) was

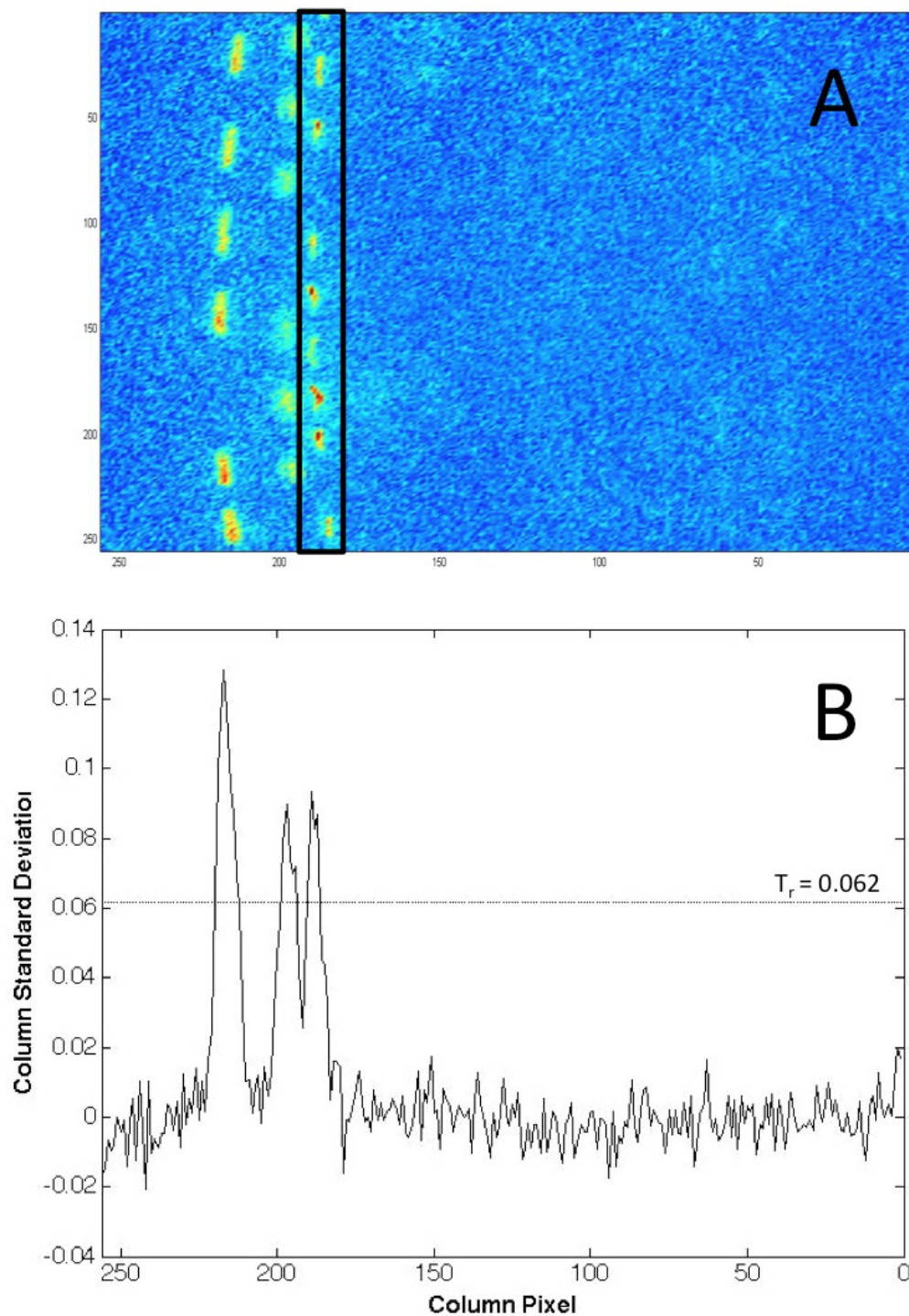


Figure 4.1 A) Preprocessed data image. There are 3 tracks of the coccolithophore *Emiliana huxleyi* visible in this image. The solid black rectangle indicates the region of the image defined by the algorithm as a track and is used as an example in the data analysis below. B) A plot of the corrected standard deviation along the rows for each column of the image in Fig. 1A. The solid black line corresponds to the baseline corrected column standard deviation and the dotted line is the calculated threshold.

scored against the threshold (dotted line 4.1B), assigning a score of 0 for standard deviations below the threshold and 1 for standard deviations equal to or greater than the threshold. Within a single image, moving from left to right, the left edge of a track is indicated by the change from 0 to 1 in the score array while the right edge is indicated by the change from 1 to 0. Each contiguous set of 1s identifies columns containing a potential track. This identifies all potentially usable tracks. The column boundaries of each track are compiled into a cell array where each cell corresponds to the image number from the data set. In each cell the row number corresponds to the track number in the image and the first column represents the left boundary value and the second column represents the right boundary value for the track.

From this point in the algorithm, each track is viewed and analyzed individually. For clarity, a track is a matrix, $\langle T \rangle$, extracted from the corrected image and is 256 rows long and the column width defined by the track selection described in the previous paragraph. Next, a track is analyzed for streaks. 4.1A is an example an image with three tracks that nearly overlap one another, a scenario which makes analyzing each individually a difficult task. For instance, in 4.1B the column standard deviations between the tracks does not fall to the baseline, indicating that streaks within each track share a common column pixel. This can be seen in the image by the track in the black box region . The track is slightly curved, resulting in it sharing part of column 191 with the track to the left of it in the image. This is problematic for calculating integrated streak intensities by summing along the columns of a track as we described in an earlier version of the algorithm, SIMOC, version 1.0, described in Pearl et. al.[9] To resolve this, each streak is found in the image by utilized the image processing toolbox in MATLAB. The streaks are treated as objects located in an image with a fairly uniform background. A matrix mask of 0's and 1's is created, where the 1's correspond to pixels of a streak and 0's to the background so that when multiplied with the track matrix results with only the streak pixels having a value

not 0.

The mask is created by duplicating the track area and converting it to gray scale, redefining the matrix with values ranging from 0 (black) to 1 (white) with the lowest pixel value in the original data array assigned 0 and highest assigned 1, shown in 4.2B. This will preserve the pixel to pixel differences in the image but allow removes negative pixel values from the image. Next, the contrast between the top 1% and bottom 1% of the pixels is increased, resulting in the background being reduced and the pixels with signal being increased as shown in 4.2C. A derivative is then calculated across both dimension of the contrasted image using the ‘canny’ method, which finds objects by the identifying local maxima in the image derivative.[1] This method defines the majority of the perimeters of the streaks in the track as well as the perimeters of parts of the streaks in the adjacent track, shown in the binary image 4.2D where 1’s indicate the perimeter edge of an object. The perimeter of each object is filled out, shown in figure 4.2F. This is done by looking at the 3x3 array around each pixel of 4.2D and if the central pixel in the 3x3 array is a 0 bordered by pixel values of 1, it is replaced with a 1. Figure 4.2E shows the ‘filled’ in objects of 4.2D. From here, the objects that have a fully contained perimeter are ‘filled’ in by making the pixel values within the perimeter equal to 1 as shown in 4.2F. The streaks are then identified from the objects as having a total pixel area greater than 60% of the average area of the 5 largest objects. Those with smaller areas are removed and the final mask is generated, as shown in 4.2G. Then the mask (4.2G) is multiplied with original data image (4.2A) leaving only streaks 4.2H.

The data array from 4.2H is then used to identify the order of the streaks in reference to the blank filter and integrate their signals. This is done by generating an array $\langle T_c \rangle$ by summing track image along the column dimension and identifying the row values in $\langle T_c \rangle$ where a streak begins and ends. By multiplying the original image array by the mask, the background and blank spot pixel values in $\langle T_c \rangle$ are all

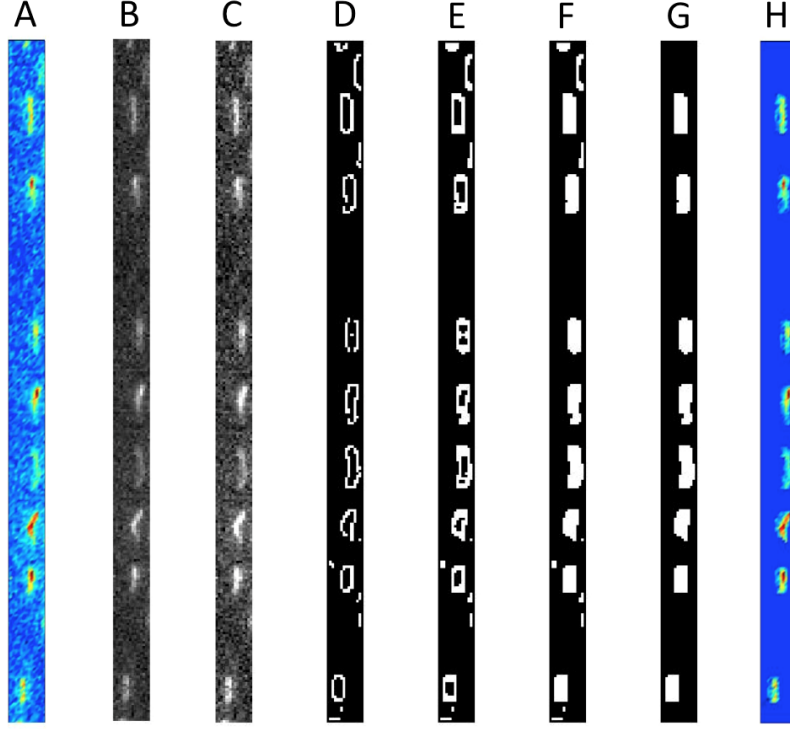


Figure 4.2 The image area of a track presented in Fig 1A. 256 rows by 11 columns. A) Red-Green-Blue (RGB) image of the track area. B) Gray scale image. C) Enhanced gray scale image D) Binary image of all the objects found in the image E) Binary image with the perimeter of each object filled out F) Binary image with the interior of each object filled in G) Binary image with where the object of non-streaks removed

equal to zero so the transition from the background to the beginning of a streak is simply identifying the rows where there is a transition from a zero value to a non-zero value. Equally, a transition from a non-zero value to zero is indicative the end of a streak. A row vector, $\langle T_i \rangle$ is then generated where values of 1 indicate a streak and 0 indicate the background values in $\langle T_c \rangle$. The blank spots and streaks in the track image are identified by finding the lengths of the streaks and background areas $\langle T_i \rangle$. The continuous background areas in $\langle T_i \rangle$ that are longer than 1.25 times the length of the average streaks are identified as blank positions. 4.3 illustrates the MOEs in the filter wheel and their assignment to streaks in the track.

The boundaries of each streak defined in $\langle T_i \rangle$ are then used to sum each streak in

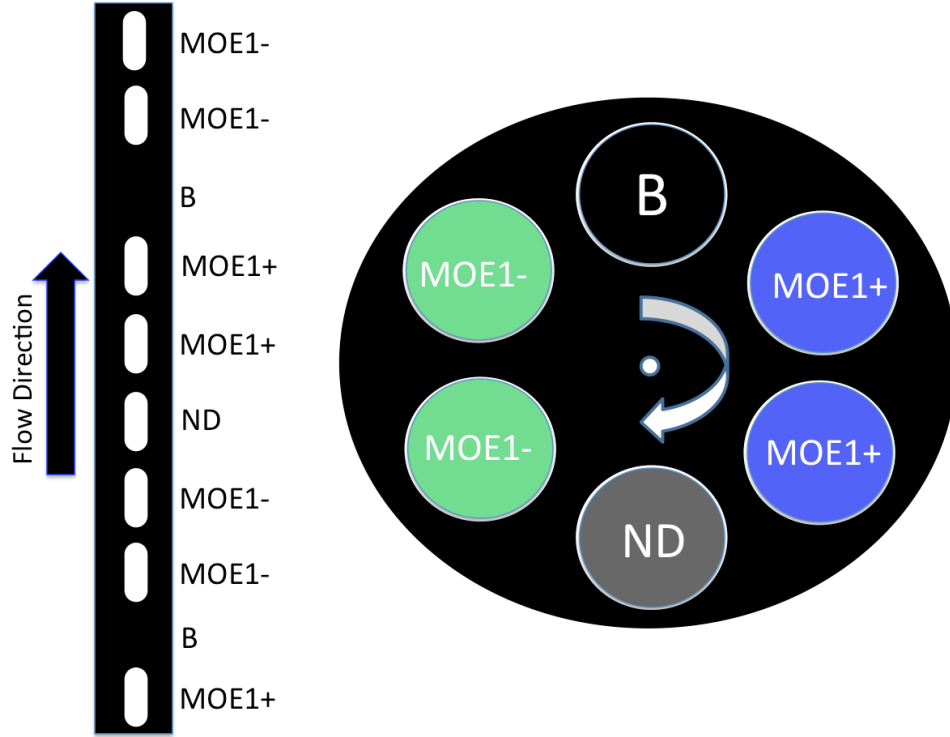


Figure 4.3 Diagram of of the filter wheel with MOE placements and their assignments on a theoretical track mask. The wheel is turning in a counter clock wise direction and vertical the flow direction in the image would result in the first streak after the blank to be assigned to MOE1-.

$\langle T_c \rangle$ and ordered according to their distance from the blank positions. The integrated streaks are then stored in a cell array, $\langle S_i \rangle$. The critical measurement for classification of *E. huxleyi* and *T. pseudonana* is the ratio of fluorescence intensity recorded when the excitation beam passes through MOE 1+ to that recorded when the beam passes through MOE 1-. Two of each of these MOEs are loaded into the filter wheel. The fluorescence responses for each MOE of the same type are averaged and then the ratio is calculated.

Deblurring

As described earlier, the out of focus streaks in an image are due to the photyoplankter passing through the flow cell out of the focal plan of the objective. The portion of the

algorithm discussed below tries to recreate the image with ‘in focus’ streaks. Images of fluorescent microspheres (Fuoresbrite, Polysciences Inc., Warrington, PA) are used replacements for phytoplankton because their diameter is known and thus the pixel widths of ‘in focus’ streaks is also known.

4.4A is an image of a 6 micron microsphere. The average width of the streaks in this image, measured with the algorithm described in the previous section, is 9 pixels or 18 microns. In chapter 3 it was shown that image pixels are equal to 2 microns. 4.4B is a deblurred image with an average streak width of 4 pixels or 8 microns.

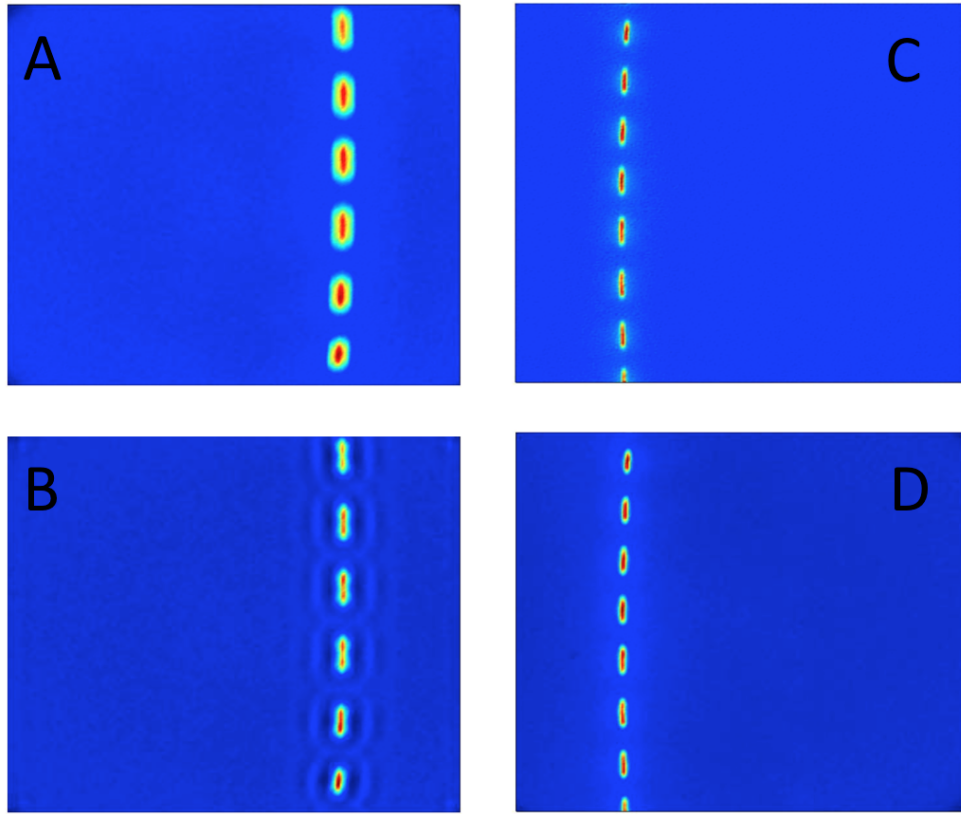


Figure 4.4 A) Corrected data image of a 6 micron microsphere that is out of focus B) Deblurred image formed by deconvolving the image in A. C) Corrected data image of a 6 micron microsphere that is in focus. D) Deblurred image formed by deconvolving the image in C.

The deblurring algorithm deblurs the image using the Lucy-Richardson deconvolution method. The assumption made here is that the image is blurred by a disk shaped

degradation pattern, so a point spread function that most resembles this degradation pattern can be used to recreate the original, unblurred image. The degree in which the beads (or phytoplankton) are out of focus has a tremendous impact on the size of the point spread function. Since the degree of blurriness is not known or consistent between tracks, disk point spread functions with diameters ranging from 2 to 10 in 0.5 increments are applied to all tracks. This approach results in a minimum streak diameter for the optimal point spread function because the image begins to blur after this optimal diameter is reached.

4.3 RESULTS AND DISCUSSION

MOE Ratios

Images containing fluorescence streaks of *E. huxley* and *T. pseudonana* were recorded and analyzed using the approach described above. This automated analysis resulted in 858 *E. huxley* ratios and 703 *T. pseudonana* ratios. The mean ratio of fluorescence intensities for excitation through (MOE1+/MOE1-) for *E. huxley* was 0.8992 ± 0.0049 , while *T. pseudonana* had a mean of 1.1956 ± 0.0052 (Table II). 4.5 shows a plot of the measured ratios for each species. An examination of this plot shows the ratios for each prediction are centered on their respective means, but with outliers. With few exceptions, these outliers were found to result from defects in analysis by SIMOC v2.0. The theoretical ratios from chapter 2 for *E. huxley* was 0.926 ± 0.012 , while *T. pseudonana* had a mean of 1.177 ± 0.014 .

An outlier was defined as having a ratio outside two standard deviations of the sample population. The images that contained tracks with ratios outside this range were found and examined to determine if the calculated ratios were accurately determined by the program. There were 37 *T. pseudonana* and 44 *E. huxley* ratios tested as outliers, and 29 *T. pseudonana* and 36 *E. huxley* tracks showed obvious defects

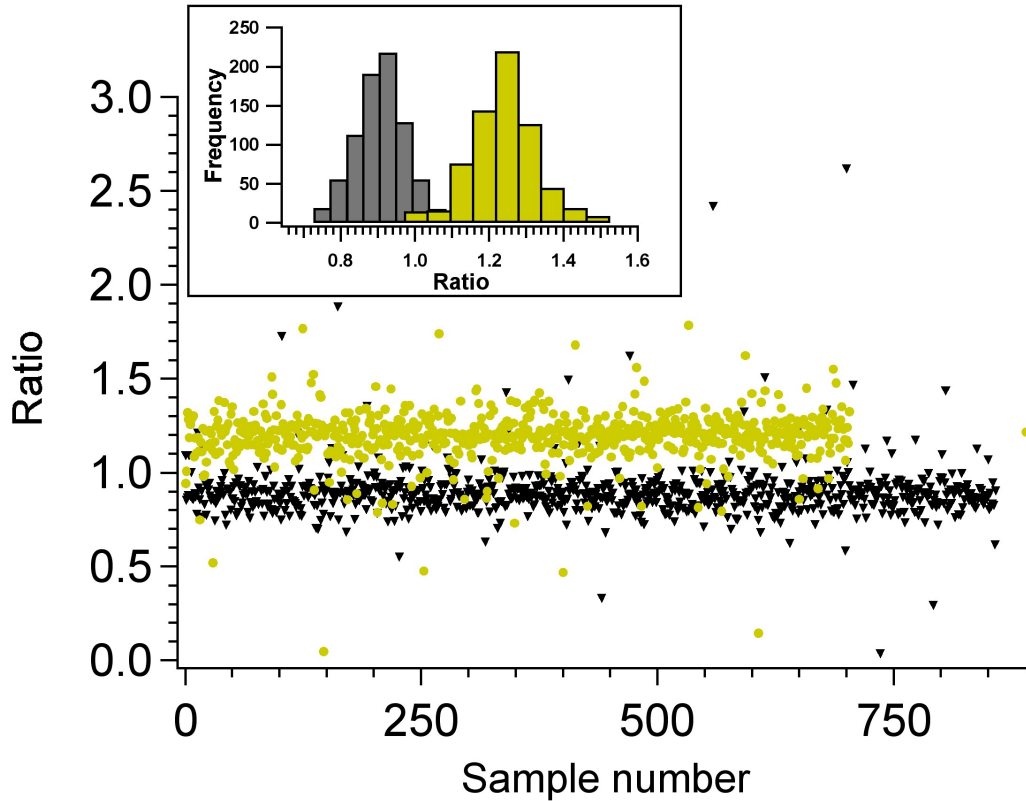


Figure 4.5 Distribution of measured ratios for the coccolithophore *Emiliana huxleyi* (black, 703 ratios) and the diatom *Thalassiosira pseudonana* (yellow, 858 ratios). The inset is the frequency distribution of the measured ratios where 81 outlier ratios were visually inspected and 65 were removed due to defects in the image that were missed by the algorithm.

that were visible in the author's opinion but passed the tests by the program. The ratios resulting from the 65 tracks that showed obvious defects were removed. The defects in the tracks include: the camera integration terminating during the acquisition of the final streak in the track which resulted in a fraction of the integrated intensity of the streak; a phytoplankter track that flowed in and out of the plane of focus which caused the later streaks in the track to be in a different state of focus than the earlier streaks in the track which results in abnormally shaped streaks with inconsistent integrated areas; and streaks that overlap the same image area as a track from a phytoplankter that passed through the flows cell at a different time in the integration. Table 4.1 shows the number of ratios, calculated means, standard

deviation of the means and standard deviations for the distributions after the outliers were removed.

Table 4.1 Mean (\bar{x}), standard deviation of the mean ($\sigma_{\bar{x}}$), standard deviation (σ) and the number of samples (N) for *E. huxleyi* and *T. pseudonana*.

| | \bar{x} | $\sigma_{\bar{x}}$ | σ | N |
|----------------------|-----------|--------------------|----------|-----|
| <i>E. huxleyi</i> | 0.8992 | 0.0045 | 0.07 | 858 |
| <i>T. pseudonana</i> | 1.1956 | 0.0052 | 0.09 | 703 |

The inset of Fig. 5 shows the ratios of *T. pseudonana* and *E. huxley*, with the defective ratios removed. The separation of mean ratios for the two classes was 0.296. The sample standard deviations for the *E. huxley* and *T. pseudonana* ratio distributions shown in Table 4.1 were 24% and 30% of the difference between the class mean ratios, respectively. While the results showed that the predicted ratios from chapter 2 were within $\pm 3\%$ of the measured mean ratios here, the standard deviation of each distribution resulted in 137 misclassifications. The cause of the high standard deviations is most likely due to the SNR of the photometer and the inability of the algorithm to distinguish poor SNR tracks from good SNR tracks. In chapter 2, a SNR of 19 was required for classification. Here, it is uncertain the value of each track's SNR. Advancements in the photometer and algorithm are underway to improve the SNR determination of each track.

The algorithm miscalculates ratios for approximately 4.2% of tracks for a variety of image processing and timing factors, but improvements are still underway and during the writing of this thesis, a subsequent SIMOCv3.0 coded in the Myrick lab has improved the image analysis algorithm and thus improving the selection of streak areas. The newer version has shown a decrease in the number of outliers and an increase in identifying tracks with very weak fluorescence signals while at the same time rejecting them if their SNR is below the theoretical SNR threshold. The description and results of SIMOCv3.0 are currently in preparation for publication where the

application of SIMOCv3.0 to data shown in Chapters 5 and 6 will be presented.

4.4 CONCLUSION

An algorithm was developed to identify phytoplankton tracks and extract size and integrated intensities from the streaks of tracks in images collected on the SSIMOC photometer. Using this algorithm to analyze images collected from cultures of *E. huxley* and *T. pseudonana*, we verified the theoretical model presented in chapter 2 and successfully classified *E. huxley* fluorescent tracks from *T. pseudonana* tracks when excitation light was filtered using the MOEs fabricated in chapter 2.

BIBLIOGRAPHY

- [1] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986. URL <http://www.mathworks.com/help/images/ref/edge.html?searchHighlight=edge>.
- [2] C. S. Davis, Q. Hu, S. M. Gallager, X. Tang, and C. J. Ashjian. Real-time observation of taxa-specific plankton distributions: an optical sampling method. *Marine Ecology-Progress Series*, 284:77–96, 2004.
- [3] C. S. Davis, F. T. Thwaites, S. M. Gallager, and Q. Hu. A three-axis fast-tow digital video plankton recorder for rapid surveys of plankton taxa and hydrography. *Limnology and Oceanography-Methods*, 3:59–74, 2005.
- [4] K. V. Embleton, C. E. Gibson, and S. I. Heaney. Automated counting of phytoplankton by pattern recognition: a comparison with a manual counting method. *Journal of Plankton Research*, 25(6):669–681, 2003.
- [5] P. G. Falkowski and J. A. Raven. *Aquatic Photosynthesis*. Princeton University Press, Princeton, 2nd edition, 2007.
- [6] G. Gorsky, P. Guilbert, and E. Valenta. The autonomous image analyzer - enumeration, measurement and identification of marine-phytoplankton. *Marine Ecology-Progress Series*, 58(1-2):133–142, 1989.
- [7] Q. Hu and C. Davis. Automatic plankton image recognition with co-occurrence matrices and support vector machine. *Marine Ecology-Progress Series*, 295:21–31, 2005. 947UC Times Cited:16 Cited References Count:49.

- [8] R. J. Olson, A. Shalapyonok, and H. M. Sosik. An automated submersible flow cytometer for analyzing pico- and nanophytoplankton: Flowcytobot. *Deep-Sea Research Part I-Oceanographic Research Papers*, 50(2):301–315, 2003.
- [9] M. R. Pearl, J. A. Swanstrom, L. S. Bruckman, T. L. Richardson, T. J. Shaw, H. Sosik, and M.L. Myrick. Taxonomic classification of phytoplankton with multivariate optical computing, part iii: Demonstration. *Applied Spectroscopy*, 67(6), 2013.
- [10] M. Picheral, L. Guidi, L. Stemann, D. M. Karl, G. Iddaoud, and G. Gorsky. The underwater vision profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. *Limnology and Oceanography-Methods*, 8:462–473, 2010. 678UM Times Cited:1 Cited References Count:39.
- [11] K. Rodenacker, B. Hense, U. Jutting, and P. Gais. Automatic analysis of aqueous specimens for phytoplankton structure recognition and population estimation. *Microscopy Research and Technique*, 69(9):708–720, 2006.
- [12] T. P. A. Rutten, B. Sandee, and A. R. T. Hofman. Phytoplankton monitoring by high performance flow cytometry: A successful approach? *Cytometry Part A*, 64A(1):16–26, 2005.
- [13] H. M. Sosik and R. J. Olson. Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. *Limnology and Oceanography-Methods*, 5:204–216, 2007. ISI Document Delivery No.: 200DB Times Cited: 7 Cited Reference Count: 38 Sosik, Heidi M. Olson, Robert J. AMER SOC LIMNOLOGY OCEANOGRAPHY.
- [14] Takashi Tsuji and Takashi Nishikawa. Automated identification of red tide phy-

- toplankton prorocentrum triestinum in costal areas by image analysis. *Journal of the Oceanographical Society of Japan*, 40:425–431, 1984.
- [15] D. Uhlmann, O. Schlimpert, and W. Uhulmann. Automated phytoplankton analysis by a pattern recognition method. *Internationale Revue der Gesamten Hydrobiologie*, 63(4):575–583, 1978.
- [16] A. Verikas, A. Gelzinis, M. Bacauskiene, I. Olenina, S. Olenin, and E. Vaiciukynas. Phase congruency-based detection of circular objects applied to analysis of phytoplankton images. *Pattern Recognition*, 45(4):1659–1670, 2012.

CHAPTER 5

DETERMINING PHYTOPLANKTON SIZE FROM FLUORESCENCE IMAGES

5.1 INTRODUCTION

A key determination for characterizing the phytoplankton community structure is accurately identifying the number of cells per mL, or mg of chlorophyll a per mL, are in a sample of ocean water. The field of cytometry is built around accurately making this measurements, often calibrated with standard particle per cells.[12, 5, 9, 8, 1, 3] The FlowCytobot developed by Sosik and Olsen, is represents one of the most technically advanced flow cytometers to date for phytoplankton analysis.[6, 4] The basic idea of a flow cytometer is to count the number of particles in a sample using a selective feature of the particles of interest to trigger, such as a fluorescent marker.[10]

The FlowCytobot uses a 532 nm solid state laser for a light source and directs the light through a sheath designed flow cell that passes sample at a high rate, one particle at a time. As a phytoplankton particle passes through the sheath flow, the phytoplankton absorbs the 532 light and emits 680 nm fluorescence. The fluorescence is detected by a photomultiplier tube and when there is sufficient fluorescence intensity from a phytoplankton particle, it is "counted", if a non phytoplankton particle passes through the system, there will be no chlorophyll a fluorescence and it will not be "counted". This manner of triggering ensures the selectivity of phytoplankton over other particles in the sample.

One of the goals of this project is to tie the information gathered by the SSIMOC method to a high resolution, high confident method. The FlowCytobot currently operating at MVCO location is ideal for this purpose. The question is how to compare a fluorescence measurement collected on the SSIMOC to the FlowCytobot? Both methods are detecting particles in the size range of tens of micrometers and collecting fluorescence intensities. This would allow the SSIMOC and FlowCytobot to record the same fluorescence measurements and for the SSIMOC to calibrate the range of fluorescent particles to the FlowCytobot.

The SSIMOC filter photometer described in chapter 3 can be treated as a flow cytometer with a few adjustments to the filter placements. The difficulty associated with the analysis for determining cell counts per mL is accurately estimating the flow volume. The FlowCytobot uses a sheath flow designed flow cell which ensures accurate counting of the entire sample. The SSIMOC flow cell in chapter 3 only views a fraction of the sample passed through the flow cell, making it a challenge to true volume sampled. In addition, the depth of field is a function of the size of the sample, meaning that smaller particles are at a lower volume than larger particles. In chapter 3 it was estimated that the depth of field is between 40 - 80 micrometers. In a data set of 500 images, this estimate puts the range of water sampled between 42 - 85 μ L. This range of sample volume makes determining an accurate cells/mL of the water sampled difficult. This chapter describes a procedure for using cell counts determined from the FlowCytobot on the same water and using the same excitation band as the SSIMOC in July 2012.

5.2 EXPERIMENTAL

The data was collected with the imaging photometer described in chapter 3. The goal for the first generation instrument is to operate on a small research vessel in Open Ocean or coastal waters. The data collected for this report was collected while

on board a Woods Hole Oceanographic research vessel, RV Tioga, at the Martha's Vineyard Coastal Observatory near Woods Hole, MA on July 2012. A total of 3000 images from 0.5 mL of coastal water was analyzed with the SSIMOC using 532 nm excitation, and 8.12 mL of sample was analyzed with the FlowCytobot.

The photometer was operated as described in chapter 3 except that the excitation band pass filter used in that study was replaced with a 532 nm bandpass filter (HQ532/10, Chroma, Bellows Falls, VT). The purpose of the 532 nm filter for excitation is to replicate the 532 nm solid state laser used for excitation as the FlowCytobot.[6] With this SSIMOC filter photometer setup, the same basic parameters are used for measuring the chlorophyll *a* fluorescence as the FlowCytobot, making it possible to use the FlowCytobot as standardized method for cell counts/mL to determine the correct cells per mL sampled by the SSIMOC filter photometer. Figure 5.1 shows the footprint of the SSIMOC filter photometer on the RV Tioga and the MVCO tower located about 1 mile east of Martha's Vineyard.

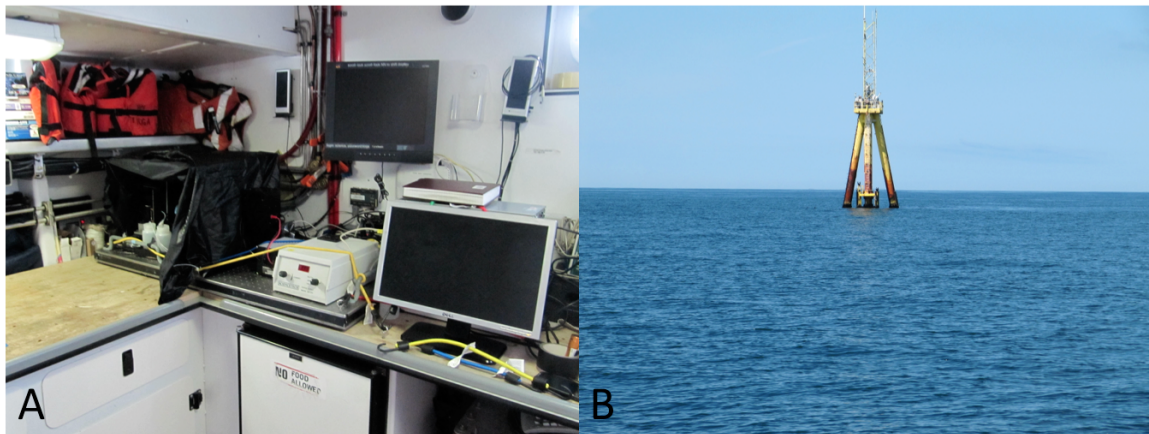


Figure 5.1 A) A picture of the SSIMOC filter photometer on a lab bench on the RV Tioga. B) A picture of Martha's Vineyard Coastal Observatory tower, positioned about one mile off the eastern coast of Martha's Vineyard, MA.

Image Analysis

The fluorescence images were analyzed using a modified version of the algorithm described in chapter 4 and in Pearl et. al.[7] The algorithm takes advantage of the distinct frequencies in the profile of the streaks in a track produced by the modulation of the filter wheel. As phytoplankton flow through the image area, phytoplankton that are in the image plane of the objective are clearly defined and produce a strong amplitude in the frequency domain. When phytoplankton are slightly above or below the image plane, the streaks begin to merge but maintain a frequency pattern that is identifiable. Phytoplankton that pass through the image area far outside the image plane will have a track with no identifiable modulated streaks and will not be identified. A full description of this algorithm will be presented in a manuscript under preparation for submission.

5.3 RESULTS AND DISCUSSION

Cell counts

Samples from the July 2012 cruise to MVCO analyzed simultaneously by both the FlowCytobot and SSIMOC contained particle fluorescence intensity results. The results of these data can be used to measure the efficiency of the SSIMOC, assuming that the FlowCytobot has 100% efficiency.

Figure 5.2A is a histogram plot of the log base 10 of the chlorophyll *a* fluorescence (measured in Volts for the FlowCytobot and photon counts for the SSIMOC) vs the log base 10 of the cells per mL for both. Log-log plots are often used when identifying relationships between two sets of empirical data.[2] Here, the goal is to identify where there is overlap between the SSIMOC and Flowcytobot in chlorophyll *a* fluorescence intensity. Simply, what fraction of the total population observed by the FlowCytobot is the SSIMOC viewing. We are assuming that the FlowCytobot detects all the

phytoplankton in a given volume, which is a safe assumption from the years of field operation and studies conducted by the FlowCytobot.[6, 11] Setting the bin widths of both histograms to 0.05 log units, we can overlay the two and estimate a region of log space where the two populations overlap as shown in Figure 5.2B. Figure 5.2B is made by manually shifting the SSIMOC distribution along the x-dimension by 4.95 log units and the FlowCytobot along the y-dimension by 0.1191 log units and used as a guide for the starting parameters for a fit described below.

Equation 5.1 and 5.2 show a relationship that can be drawn from the shifting the two histograms.

$$I_{FCB} = C_1 * I_{SSIMOC} \quad (5.1)$$

Equation 5.1 shows the relationship between the fluorescence intensity measured by the FlowCytobot (I_{FCB}), measured in volts, and SSIMOC (I_{SSIMOC}), measured in photon counts, are offset by a constant C_1 .

$$N_{SSIMOC} = C_2 * N_{FCB} * \eta_{(I_{FCB})} \quad (5.2)$$

Equation 5.2 shows that a relationship in the number of cells detected by the FlowCytobot and the the SSIMOC photometer and can be used to correct the volume sampled estimate of the SSIMOC photometer discussed earlier due to the depth of field issue. Where N_{SSIMOC} is the number of cells detected by the SSIMOC photometer, C_2 is a constant volume offset, N_{FCB} is the number of cells detected by the FlowCytobot and $\eta_{(I_{FCB})}$ is the efficiency of the SSIMOC as a function of FlowCytobot chlorophyll a fluorescence intensity.

Taking the log of each term in equation 5.1 results in 5.3:

$$\log(I_{FCB}) = \log(C_1) + \log(I_{SSIMOC}) \quad (5.3)$$

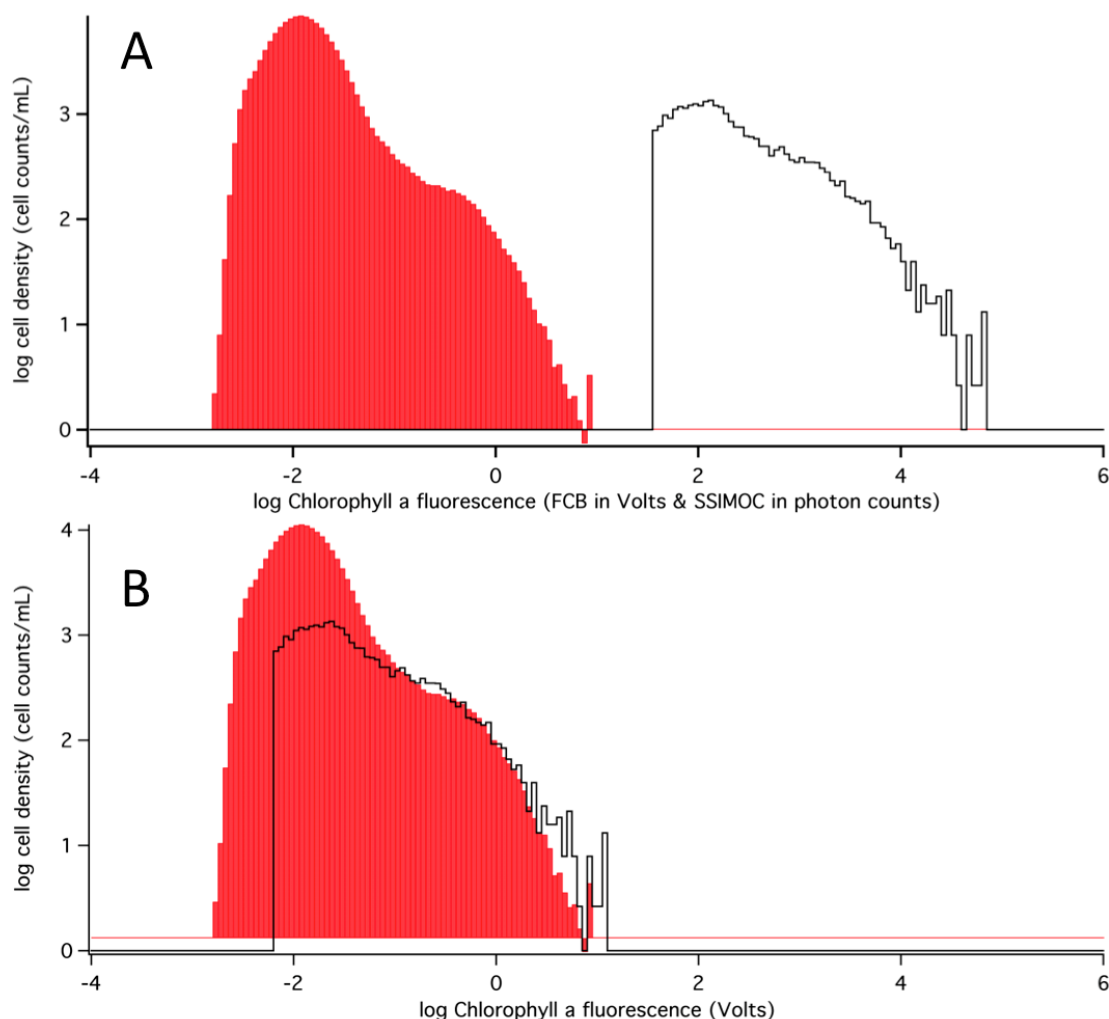


Figure 5.2 A) The histogram plot in red of the log of the cell density (cell counts/mL) vs log of the chlorophyll a fluorescence (volts) collected during a July 2012 cruise from the FlowCytobot. The histogram plot outlined in black is of the log of the cell density (cell counts/mL) vs log of the chlorophyll a fluorescence (photon counts) collected during a July 2012 cruise from the SSIMOC filter photometer. B) Same histograms as in (A) except that the SSIMOC plot is shifted to the left 4.95 log units and the FlowCytobot plot is shifted up 0.1191 log units. This is done to visually determine if the plots have overlapping features, which are visible in (B) between -0.6 and 0.9 volt log units.

The shift in the x-dimension shown in Fig. 5.2 is equal to C_1 . Estimating the correction for cell counts in the SSIMOC distribution is not as straight forward as converting the FlowCytobot's volts to photon counts. This is due to the uncertainty in the number of cells detected. The total volume measured by the FlowCytobot was 8.12 mL, giving a much greater certainty to it's counts/mL measured at each chlorophyll a fluorescence intensity than the roughly 0.3 - 0.6 mL estimated by the SSIMOC. To give a better certainty, the y-dimension of fig. 5.2 was converted back to a linear scale and 95% confidence intervals were calculated for each SSIMOC. Then, a best-fit was calculated for the plot shown in fig. 5.3 using the 95% confidence intervals as weighting factors giving more influence to the points with higher certainty. Figure 5.3 is the counts (y-dimension) for each histograms plotted against each other, SSIMOC in counts and FlowCytobot in counts/mL. If this plot were a straight line with a slope of one running through the origin, then the two histograms would perfectly match and the volume for the SSIMOC would be 1 mL. With the y-intercept restricted to run through the origin, the slope of the best fit is the volume correction for the SSIMOC. The region of the histograms used in the fit was determined by eq. 5.2, where the offset, C_2 , is the slope of the line if fig 5.3 when the efficiency, $\eta_{(ICB)}$, is unity. Using the region visually identified in manual shifting of the log/log plot described earlier, the region of overlap were bins 62-92 (FlowCytobot bins -0.9 to 0.6 volts). The slope of the line is 0.534 mL. This volume estimate puts the depth of field imaged in the flow cell at 60 microns, the center of the estimate used in chapter 3.

Figure 5.4A is the histogram plots for the corrected SSIMOC volume (red) and converted FlowCytobot chlorophyll a intensity to photon counts (blue). Figure 5.4B is the ratio of these two histograms at each bin. This shows that at chlorophyll a fluorescence counts above 10^4 , the two methods are nearly detecting the same number of particles within the 95% confidence level, but below 10^4 counts the SSIMOC's efficiency drops linearly.

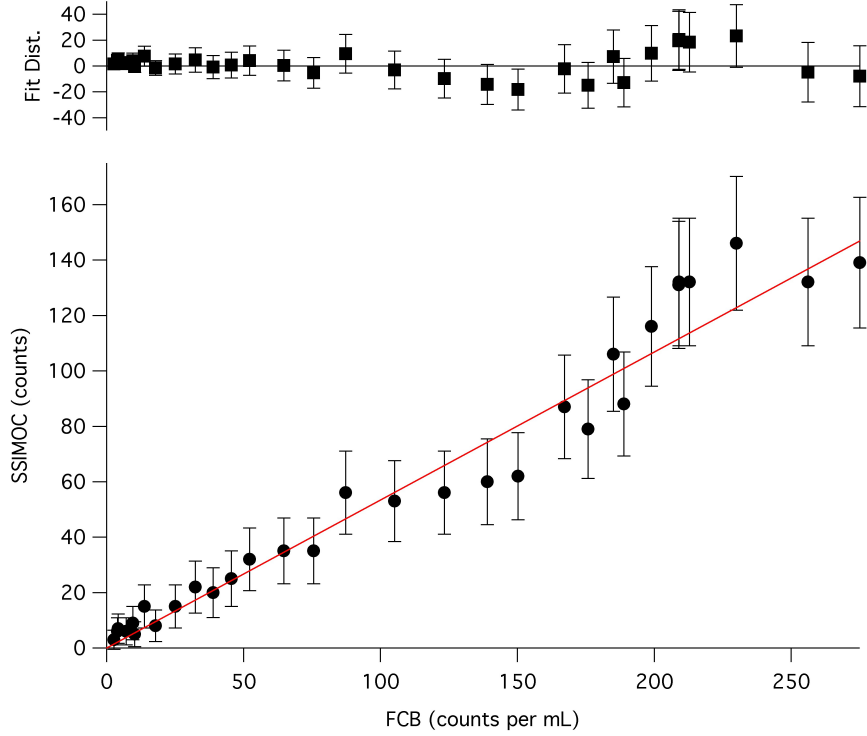


Figure 5.3 The bottom plot is of the SSIMOC counts vs the FlowCytobot counts/mL for each bin in the histograms from Figure 2, and the the red line represents the best fit line with a χ^2 of 0.0252. The error bars represent 95% confidence intervals for the SSIMOC. The top plot represents the distance between the fit and each point.

5.4 CONCLUSION

These results show that the SSIMOC, with corrected sample volumes, is detecting the same number of particles as the FlowCytobot. This is important for a couple of reasons. It shows that the classification results from the SSIMOC, when using MOE filter sets, is sampling a complete subset of the phytoplankton population within a given intensity range. The second reason is that we have shown that the measurements made from the SSIMOC can be directly tied to another classification method.

We have demonstrated that the volume sampled by the SSIMOC can be calculated from the relationship between the phytoplankton cell counts from SSIMOC images and the cell density results determined from the FlowCytobot. The corrected volume

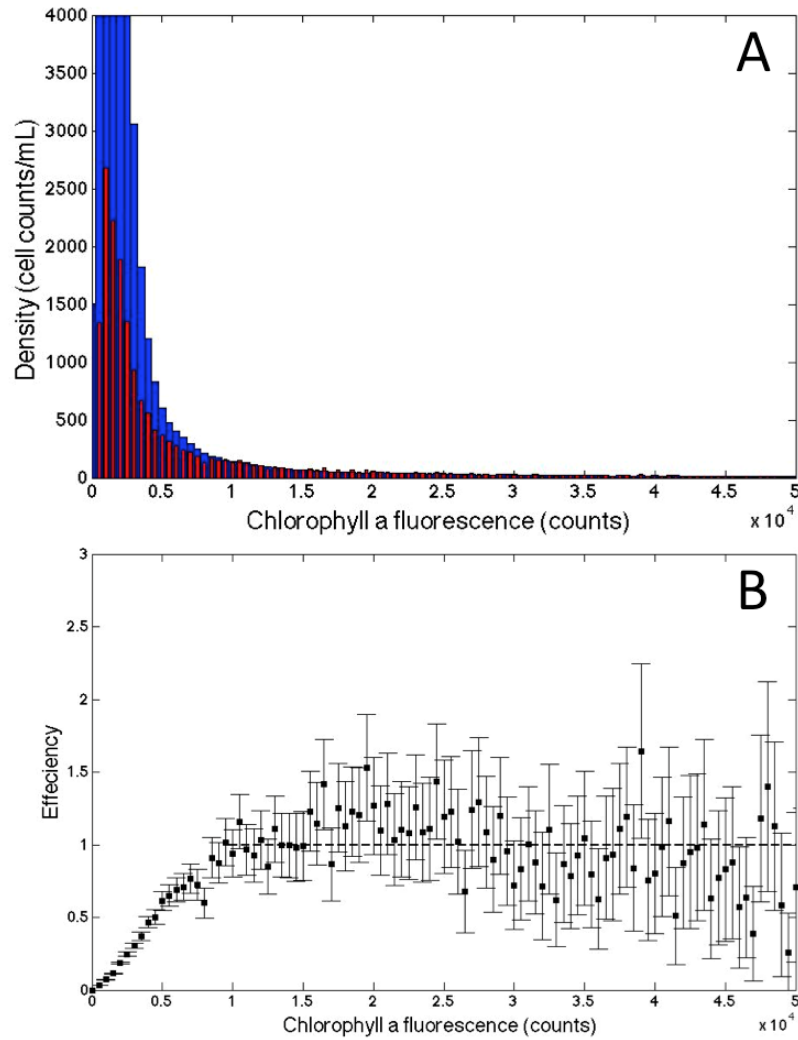


Figure 5.4 A) Histogram plots for the corrected SSIMOC volume (red) and converted FlowCytobot chlorophyll a intensity to photon counts (blue) B) Plot of the cell counts per mL efficiency of the SSIMOC method compared to the FlowCytobot where a value of one indicates a perfect efficiency.

estimates revealed that our original depth of field estimate of 40 - 80 microns in chapter 3 included the calculated depth of field of 60 microns.

BIBLIOGRAPHY

- [1] Paa A.S., J. Oro, and J.R. Cowles. Application of flow micro-fluorometry to study of algal cells and isolated-chloroplasts. *JOURNAL OF EXPERIMENTAL BOTANY*, 29:1011–1020, 1978.
- [2] A. Clauset, C.R. Shalizi, and M.E.J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.
- [3] N. Hashemi, J. S. Erickson, J. P. Golden, K. M. Jackson, and F. S. Ligler. Microflow cytometer for optical analysis of phytoplankton. *Biosensors and Bioelectronics*, 26(11):4263–4269, 2011.
- [4] R. J. Olson and H. M. Sosik. A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging flowcytobot. *Limnology and Oceanography-Methods*, 5:195–203, 2007.
- [5] R. J. Olson, D. Vault, and S.W Chisholm. Marine-phytoplankton distributions measured using shipboard flow-cytometry. *DEEP-SEA RESEARCH PART A-OCEANOGRAPHIC RESEARCH PAPERS*, 32:1273–1280, 1985.
- [6] R. J. Olson, A. Shalapyonok, and H. M. Sosik. An automated submersible flow cytometer for analyzing pico- and nanophytoplankton: Flowcytobot. *Deep-Sea Research Part I-Oceanographic Research Papers*, 50(2):301–315, 2003.
- [7] M. R. Pearl, J. A. Swanstrom, L. S. Bruckman, T. L. Richardson, T. J. Shaw, H. Sosik, and M.L. Myrick. Taxonomic classification of phytoplankton with multivariate optical computing, part iii: Demonstration. *Applied Spectroscopy*, 67(6), 2013.

- [8] T. P. A. Rutten, B. Sandee, and A. R. T. Hofman. Phytoplankton monitoring by high performance flow cytometry: A successful approach? *Cytometry Part A*, 64A(1):16–26, 2005.
- [9] M.R. Schiaffino, J.M. Gaslo, I. Izaguirre, and F. Unrein. Picoplankton abundance and cytometric group diversity along a trophic and latitudinal lake gradient. *AQUATIC MICROBIAL ECOLOGY*, 68:231–250, 2013.
- [10] Howard M. Shapiro. *Practical Flow Cytometry*. John Wiley and Sons, Hoboken, New Jersey, 4th edition, 2003.
- [11] H. M. Sosik, R. J. Olson, M. G. Neubert, A. Shalapyonok, and A. R. Solow. Growth rates of coastal phytoplankton from time-series measurements with a submersible flow cytometer. *Limnology and Oceanography*, 48(5):1756–1765, 2003. 723LK Times Cited:23 Cited References Count:37.
- [12] B.J. Trask, G.J Vandenengh, and J.H.B.W Elgershuizen. Analysis of phytoplankton by flow-cytometry. *Cytometry*, 2:258–264, 1982.

CHAPTER 6

SSIMOC FIELD TESTING

6.1 INTRODUCTION

The community composition of the phytoplankton is highly variable in both space and time. Monitoring the variability within the phytoplankton community structure is a key component to understanding the carbon flux in the ocean.[3]

The traditional method for phytoplankton classification is by microscopy, either by human identification or automated methods.[8, 5] Recent development of automated techniques such as the Video plankton Recorder by Davis et al.[7], the Harmful Algal Bloom Buoy by Culverhouse et al.[6] and the FlowCytobot by Sosik et al.[15, 18] have shown great promise and innovation in the area of phytoplankton community structure. These methods perform classification largely on the basis of morphological differences between groups of phytoplankton and give remarkably high quality results, upwards of 90% identification accuracies.[7, 6, 15, 18] They represent the leading effort in this area but require a significant amount of power and communication from a ship or coastal station.

As described in the previous chapters of this dissertation, our lab has demonstrated a spectroscopic method using fluorescence excitation spectroscopy that looks to highlight the genetic differences between phytoplankton expressed by the varying composition of photosynthetic pigments.[21, 20, 16] We call the first generation of the method the shipboard streak imaging multivariate optical computing instrument, or SSIMOC. Multivariate optical computing (MOC) is an optical method that aims

to correlate chemical or physical properties of a sample to spectral characteristics spanning multiple wavelengths.[19] The working components of MOC are multivariate optical elements (MOEs), which are specially designed interference filters with transmission functions that mimic the functionality of linear discriminant analysis (LDA).[21, 14] The MOEs operate as filters in a filter photometer system. Examples of filter photometers used for phytoplankton analysis use mainly a filter wheel, so that each filter can be brought into the optical path fixed filters, or a system of narrow band LED for excitation.[17, 2, 1, 9, 13] Unlike typical filter photometers, the MOEs do not select single wavelength bands for excitation but perform distinctive vector manipulations on the same wavelength band. The advantage of full spectral or multiwavelength excitation allows for both higher throughput and coverage of multiple pigment excitation bands. The wavelength region for excitation is identified by LDA as the optimal region for classification, (i.e. 570-650 nm).[4]

Like most spectroscopic techniques, the goal of classification using this method is not to surpass or compete with the cytometry or microscopy methods which are currently in use or in development, but to add a low cost and highly flexible spectroscopic technique to compliment the developed network of tools in the area of phytoplankton monitoring. The complexity of phytoplankton community structure in terms of organism size and special variability makes it difficult to imagine a single approach that will cover the entire population range. There is an inverse relationship between selectivity and sensitivity with sample throughput (meaning either you do a good job of identifying the components of a very small sample, milliliters, or sample large volumes with low distinguishability between components in the sample.[cite] The goal of the SSIMOC method is to make a remotely deployable instrumentation that promote and add value to existing stationary methods such as the FlowCytobot. This chapter entails the next step in the development of the method in it's application to natural water samples at Martha's Vineyard Costal Observatory (MVCO).

6.2 EXPERIMENTAL

The photometer was operated as described in chapter 4, but with different MOE filter sets. Below we will discuss the design of each MOE filter set and break down the samples analyzed with each set. Briefly, the MOEs are housed in a filter wheel that modulates a light beam, which is then reflected off a dichroic beamsplitter and focused through a 60x objective onto a flow cell. The fluorescence emission of chlorophyll a at 680 nm is then passed back through the objective and the dichroic beamsplitter and imaged onto a CCD array. An example data image and orientation of the MOEs in the filter wheel is shown in figure 1.

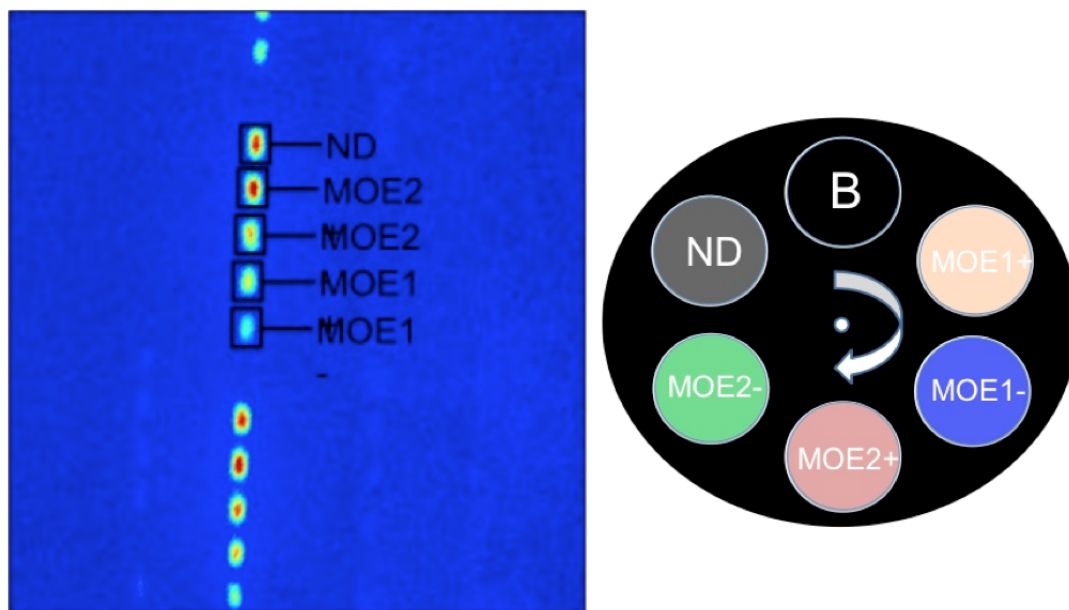


Figure 6.1 Left: Example of data image collected on SSIMOC photometer used at MVCO. Right: Example of MOEs in the filter

Multivariate Optical Elements

The process for the design and fabrication of the MOEs is described in chapters 1 and 2. For the cruises described below, three different sets of MOEs were designed

and fabricated and shown in 6.1. (A) is a set designed to classify the three species of *Emiliania huxleyi* (*E. huxleyi*), *Thalassiosira pseudonana* (*T. pseudonana*) and *Synechococcus sp.*, (B) is a set designed for the six species of *E. huxleyi*, *T. pseudonana*, *Synechococcus sp.*, *Rhodomonas salina* (*R. salina*), *Amphidinium carterae* (*A. carterae*) and *Dunaliella tertiolecta* (*D. tertiolecta*), (C) is a single MOE designed to differentiate *R. salina* in either a nitrogen replete state of growth or nitrogen deplete state of growth, and (E) is a set designed for the three classes of phytoplankton diatoms, dinoflagellates and haptophytes. The diatoms were represented by *Phaeodactylum tricornutum* (*P. tricornutum*), *Skeletonema costatum* (*S. costatum*), *Thalassiosira weissflogii* (*T. weissflogii*), *T. pseudonana*, the dinoflagellates by *A. carterae*, *Prorocentrum minimum* (*P. minimum*), *Lingulodinium polyedrum* (*L. polyedrum*), and the haptophyte *E. huxleyi*.

Sets A and B were designed using single cell fluorescence excitation spectra collected as described in Hill et. al.[12] The use of single cell spectra in the first two sets of MOEs was desired because the filter photometer is designed to collect images on a single cell basis instead of a bulk mixture. While this ensures that each spectrum collected is of a single cell, it is very labor and time expensive to collect hundreds of spectra needed for MOE designs. For this reason, it was desirable to determine if there are significant differences in the spectrum of single cells to the spectrum collected on a monoculture.

Bulk monoculture spectra were collected on a Hitachi Fluorescence Spectrophotometer (F-4500, Hitachi, Tokyo Japan). For these experiments, we used unialgal cultures of each species obtained from the Center for the Culture of Marine Phytoplankton (National Center for Marine Algae and Microbiota) at the Bigelow Laboratory for Ocean Sciences, East Boothbay Harbor, Maine. The phytoplankton cultures were grown in 0.2 μm filter-sterilized f/2 culture medium, salinity ≈ 35 , at 23° C.[11, 10] Illumination was provided from the top and sides at an irradiance of approximately

80 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$ on a 12:12 light:dark cycle.

Research Cruises

Five sampling cruises were undertaken to test SSIMOC at the Woods Hole Oceanographic Institution's Martha's Vineyard Coastal Observatory (MVCO) near South Beach in Edgartown, Massachusetts on March 23, 2011, July 6, 2011, November 22, 2011, March 6 2012 and July 7, 2012. The imaging photometer was placed on a lab bench on board the RV Tioga. Samples were collected from the waters near the MVCO at a depth of 4m. Water samples were obtained with a CTD rosette lowered to a four-meter depth. The water was sampled with Niskin Bottles and returned to the surface where the water was transferred to the instrument and data collected within minutes.

Particle size

Fluorescence imaging allows us to obtain physical as well as spectral information from phytoplankton. While we are not able to collect absolute cell size from the images, we can say something about the size of the chloroplast within the cells. The size of the cells that are visible by the SSIMOC are primarily defined by the magnification of the objective and size of the CCD chip. Selecting the magnification of the of the objective is a balance of desired magnification and numerical aperture. The 60x/0.7 N.A objective used for this study was selected because of its relatively high throughput and magnification allowed for imaging phytoplankters in the size range of 3-20 microns. If the phytoplankter passes through the cell at the focal plan, then the size of the chloroplast is determined by magnification and pixel width, but if it passes slightly out of the focal plan then the pixel width alone would overestimate the size. A manuscript detailing the algorithm is in preparation at the time of this thesis. Briefly, the algorithm calculates the intensity profile of a streak in the row

dimension and uses the shape of the streak profile to estimate the width in pixels at half max intensity. A great deal of effort and trials with multivariate calibration techniques was vetted, but the simple model of half width at full max resulted as the best model for predicting size.

6.3 RESULTS AND DISCUSSION

The SSIMOC was tested on a total of five cruises to MVCO. The first cruises were trial runs to test the instrumentation and determine the parameters of the SSIMOC, as well as gain insight from the proven experience of the scientist at Woods Hole Oceanographic Institution. The MOEs used in the March 2011, July 2011 and November 2011 were first generation MOEs. During the first cruises, we came to the conclusion that models that included three and six species were inadequate in gathering community structure information. The MOE set used in the final cruises, March 2012 and July 2012, were designed to cover a broader range. Instead of attempting to look species specific, we group species together into similar classes: Diatoms, Dinoflagellates and Haptophytes.

The five cruises, even though different MOEs were used, allowed for quantitative evaluation of the MVCO water with the SSIMOC method shown in 6.1. The total number of images collected from each cruise day varied from 9,000 to 14,000 and was influenced by the weather and amount of time the ship could spend at the MVCO. The total number of tracks detected and number of apposite tracks from each cruise both were greatly influenced by the SNR of the imaging photometer due to the set up as well as the overall photosynthetic properties of the phytoplankton community structure. The pigment concentration, and thus photosynthetic efficiency and fluorescence intensity, decrease at a function of solar radiation intensity and daily duration.

As described previously, the approach selecting which phytoplankton species the

Table 6.1 The rows in column 1 are the month and years of the cruises that the SSIMOC participated in to MVCO. Column 2 is the MOE filter sets that were used on each cruise described in the text. D is a 0.3 optical density filters used a reference and the subscripts refers to the Linear Discriminant vector each MOE is designed to mimic as described in chapter 2. Column 3 is the total number of data images collected on each cruise. Column 4 is the number of fluorescent particles detected by the SSIMOC. Column 5 is the number of tracks detected that have a high enough SNR for MOC calculations. Column 6 is the calculated cell density based on the total number of fluorescent particles detected and the total number of mL analyzed.

| Table 1 | | | | | |
|-------------|-----------------|------------------|-------------|----------------------|--------------------|
| Date | MOE Filter Set | # of Data Images | # of tracks | # of apposite tracks | Estimated Cells/mL |
| March, 2011 | A_1, B_1, C_1 | 10,600 | 9,659 | 1,034 | 2,700 |
| July, 2011 | B_1, B_2, A_1 | 14,000 | 19,690 | 1,966 | 20,000 |
| Nov, 2011 | A_1, A_2, D | 9,500 | 10,506 | 462 | 14,200 |
| March, 2012 | E_1, E_2, D | 11,600 | 4,736 | 212 | 6100 |
| July, 2012 | E_1, E_2, D | 9,000 | 7,946 | 106 | 15,900 |

MOEs were designed to classify changed over time. In 2011, the goal was to classify similarly pigments phytoplankton species using filter set A. The next set of MOEs were designed to classify six species of phytoplankton over a wider pigmentation range in an attempt to characterize a larger portion of the community. But, this approach was not suitable for the diversity of the phytoplankton community structure found in the waters near the MVCO. The phytoplankton community structure of MVCO consists of 100's of species of phytoplankton over a large range of physical sizes, 1 - 100 microns in size. Also, the presence of chain forming diatoms further complicates the community structure. Discussions with collaborators Dr. Heidi Sosik of Woods Hole Oceanographic Institution and Dr. Tammi Richardson, identified that the greatest effect for the SSIMOC is to classify the community structure at the level of phytoplankton class instead of phytoplankton species. The March 2012 cruise saw the first implementation of a set of filters designed to classify on the level of phytoplankton class instead of phytoplankton species, diatoms, dinoflagellates, and haptophytes.

Three Class MOE Results

Fig. 6.2 shows a plot of the MOE results of the March 2012 cruise with the 212 apposite tracks.

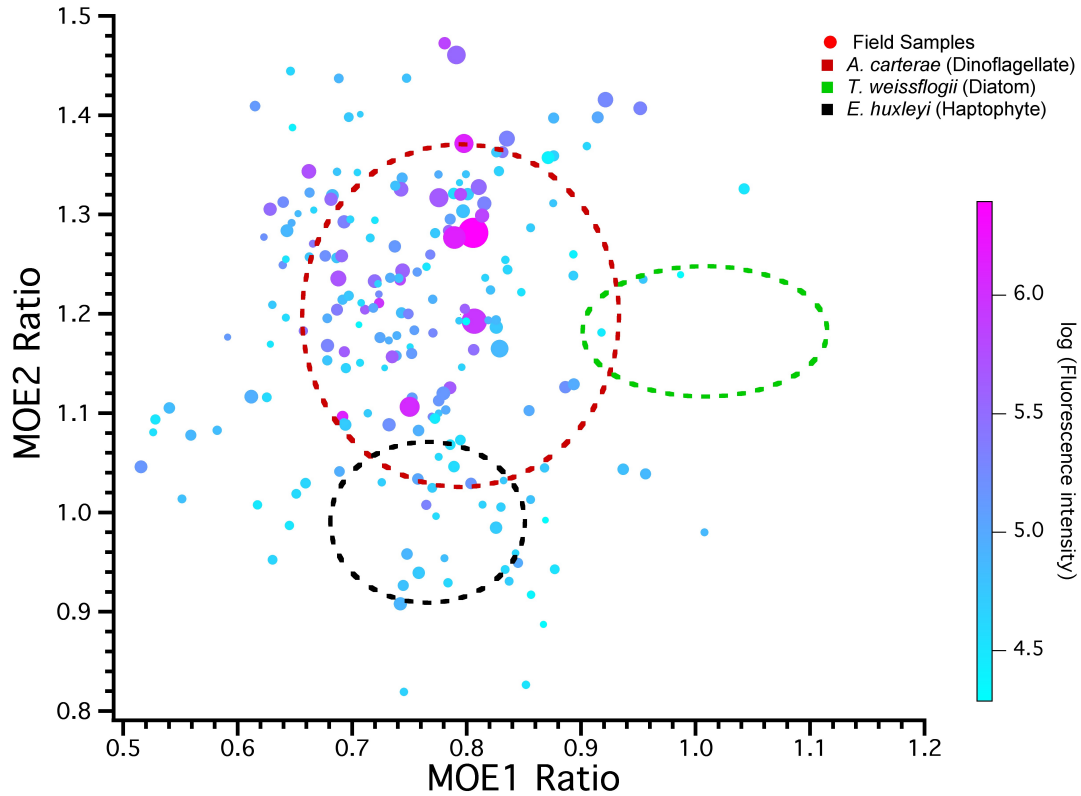


Figure 6.2 Classification distribution of 200 phytoplankton tracks collected from MVCO in March of 2012. The dashed regions represent 95% confidence intervals calculated from measurements of cultured dinoflagellate (red), diatom (green), and haptophyte (black) when interrogated through the 3 class MOEs. The solid circles represent the MVCO natural population recorded using the 3 class set MOEs, the size of the circle represents it's relative size (largest circle represents 44 microns and smallest circle represents less than 3 microns) and color gradient represents average fluorescence intensity over all 5 streaks. Unknown samples that fall within the 95% confidence limits would be classified as either a dinoflagellate, diatom or haptophyte. Unknown samples that fall outside these bounds would be classified as not belonging to any of these three class.

Fig. 6.3 shows a plot of the MOE results of the July 2012 cruise with the 212 apposite tracks.

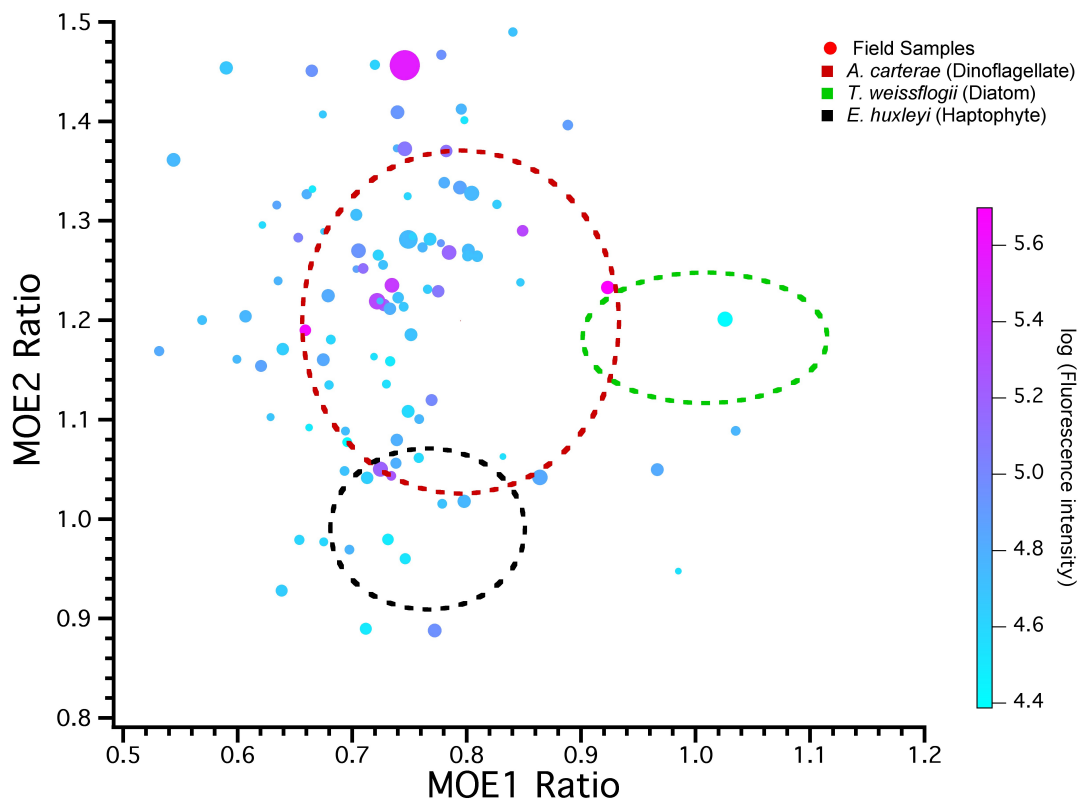


Figure 6.3

6.4 CONCLUSION

The goal of the SSIMOC research cruises was to identify where improvements to the SSIMOC filter photometer needed to be made and for the group of chemists working on the project to gain a better understanding of the environment in which the SSIMOC would be operating. The five cruises resulted in significant improvement to the robustness of the filter photometer. The development of designing MOE sets for classifying phytoplankton on the class level was an important result from this study.

We have shown the capability of the SSIMOC photometer to operate on a research vessel and collect data images of coastal ocean waters containing natural phytoplankton. This demonstrates the usefulness and ability for optical computing to operate in this environment for the purpose of classification. Further development of the SSI-

MOC photometer is underway in the Myrick group to move the instrumentation from the lab of a research vessel into a smaller and more robust enclosure with the goal to operate in a more remote setting similar to the FlowCytobot and other operating ocean monitoring phytoplankton instrumentation.

BIBLIOGRAPHY

- [1] G. Becker, H. Holfeld, A. T. Hasselrot, D. M. Fiebig, and D. A. Menzler. Use of a microscope photometer to analyze in vivo fluorescence intensity of epilithic microalgae grown on artificial substrata. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 63(4):1318–1325, 1997.
- [2] M. Beutler, K. H. Wiltshire, B. Meyer, C. Moldaenke, C. Luring, M. Meyerhofer, U. P. Hansen, and H. Dau. A fluorometric method for the differentiation of algal populations in vivo and in situ. *Photosynthesis Research*, 72(1):39–53, 2002.
- [3] P. W. Boyd and P.P. Newton. Does planktonic community structure determine downward particulate organic carbon flux in different oceanic provinces? *Deep-Sea Research Part I-Oceanographic Research Papers*, 46(1):63–91, 1999.
- [4] L. S. Bruckman, T. L. Richardson, J. A. Swanstrom, K. A. Donaldson, M. Allora, T. J. Shaw, and M. L. Myrick. Linear discriminant analysis of single-cell fluorescence excitation spectra of five phytoplankton species. *Applied Spectroscopy*, 66(1):60–65, 2012.
- [5] P. F. Culverhouse, R. Williams, B. Reguera, V. Herry, and S. Gonzalez-Gil. Do experts make mistakes? a comparison of human and machine identification of dinoflagellates. *Marine Ecology-Progress Series*, 247:17–25, 2003.
- [6] P. F. Culverhouse, R. Williams, B. Simpson, C. Gallienne, B. Reguera, M. Cabrini, S. Fonda-Umani, T. Parisini, F. A. Pellegrino, Y. Pazos, H. Wang, L. Escalera, A. Morono, M. Hensey, J. Silke, A. Pellegrini, D. Thomas, D. James, M. A. Longa, S. Kennedy, and G. del Punta. Hab buoy: a new instrument for

- in situ monitoring and early warning of harmful algal bloom events. *African Journal of Marine Science*, 28(2):245–250, 2006.
- [7] C. S. Davis, F. T. Thwaites, S. M. Gallager, and Q. Hu. A three-axis fast-tow digital video plankton recorder for rapid surveys of plankton taxa and hydrography. *Limnology and Oceanography-Methods*, 3:59–74, 2005.
- [8] K. V. Embleton, C. E. Gibson, and S. I. Heaney. Automated counting of phytoplankton by pattern recognition: a comparison with a manual counting method. *Journal of Plankton Research*, 25(6):669–681, 2003.
- [9] G. Friedrich, V. Gerhardt, U. Bodemer, and M. Pohlmann. Phytoplankton composition and chlorophyll concentration in freshwaters: comparison of delayed fluorescence excitation spectroscopy, extractive spectrophotometric method, and utermohl- method. *Limnologica*, 28:323–328, 1998.
- [10] R. R. Guillard. *Culture of phytoplankton for feeding marine invertebrates*. Culture of Marine Invertebrate Animals. Plenum Press, New York, 1975.
- [11] R. R. Guillard and J. H. Ryther. Studies of marine planktonic diatoms .1. cyclotella nana hustedt, and detonula confervacea (cleve) gran. *Canadian Journal of Microbiology*, 8(2):229–239, 1962.
- [12] L. S. Hill, T. L. Richardson, L. T. M. Profeta, T. J. Shaw, C. J. Hintz, B. S. Twining, E. Lawrenz, and M. L. Myrick. Construction, figures of merit, and testing of a single-cell fluorescence excitation spectroscopy system. *Review of Scientific Instruments*, 81(1):013103, 2010.
- [13] S. W. Jeffrey and G. F. Humphrey. New spectrophotometric equations for determining chlorophylls a, b, c1 and c2 in higher-plants, algae and natural phytoplankton. *BIOCHEMIE UND PHYSIOLOGIE DER PFLANZEN*, 167:191–194, 1975.

- [14] M. L. Myrick, O. Soyemi, J. Karunamuni, D. Eastwood, H. Li, L. Zhang, A. E. Greer, and P. Gemperline. A single-element all-optical approach to chemometric prediction. *Vibrational Spectroscopy*, 28(1):73–81, 2002.
- [15] R. J. Olson and H. M. Sosik. A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging flowcytobot. *Limnology and Oceanography-Methods*, 5:195–203, 2007.
- [16] M. R. Pearl, J. A. Swanstrom, L. S. Bruckman, T. L. Richardson, T. J. Shaw, H. Sosik, and M.L. Myrick. Taxonomic classification of phytoplankton with multivariate optical computing, part iii: Demonstration. *Applied Spectroscopy*, 67(6), 2013.
- [17] V. V. Povazhnyi. Determination of the chlorophyll "a" concentration using a combined method based on measurements with a modified photometer. *Oceanography*, 66(1):60–65, 2012.
- [18] H. M. Sosik, R. J. Olson, M. G. Neubert, A. Shalapyonok, and A. R. Solow. Growth rates of coastal phytoplankton from time-series measurements with a submersible flow cytometer. *Limnology and Oceanography*, 48(5):1756–1765, 2003. 723LK Times Cited:23 Cited References Count:37.
- [19] O. Soyemi, D. Eastwood, L. Zhang, H. Li, J. Karunamuni, P. Gemperline, R. A. Synowicki, and M. L. Myrick. Design and testing of a multivariate optical element: The first demonstration of multivariate optical computing for predictive spectroscopy. *Analytical Chemistry*, 73(17):4393–4393, 2001.
- [20] J. A. Swanstrom, L. S. Bruckman, M. R. Pearl, E. Abernathy, T. L. Richardson, T. J. Shaw, and M. L. Myrick. Taxonomic classification of phytoplankton with multivariate optical computing, part ii: Design and experimental protocol of a shipboard fluorescence imaging photometer. *Applied Spectroscopy*, 67(6), 2013.

- [21] Joseph A. Swanstrom, Laura S. Bruckman, Megan R. Pearl, Michael N. Simcock, Kathleen A. Donaldson, Tammi L. Richardson, Timothy J. Shaw, and Michael L. Myrick. Taxonomic classification of phytoplankton with multivariate optical computing, part i: Design and theoretical performance of multivariate optical elements. *Applied Spectroscopy*, 67(6), 2013.

BIBLIOGRAPHY

- [1] N. Agnihotri, W. S. Kisaalita, and C. H. Keith, *Micro-perfusion flow cell for imaging cultured cells*, Biotechniques **27** (1999), no. 4, 722–726.
- [2] P. Ajanik, B. Steve, K. Martin, S. Peter, G. Webster, and L Armand, *The risk of harmful algal blooms (habs) in the oyster-growing estuaries of new south wales, australia*, Environmental monitoring and assessment **185** (2013), 5295–526.
- [3] Paau A.S., J. Oro, and J.R. Cowles, *Application of flow micro-fluorometry to study of algal cells and isolated-chloroplasts*, JOURNAL OF EXPERIMENTAL BOTANY **29** (1978), 1011–1020.
- [4] D. Baden, L.E. Fleming, and J.A Bean, *Chapter: Marine toxins. in: Handbook of clinical neurology: Intoxications of the nervous system part ii. natural toxins and drugs*, Elsevier Press, 1995.
- [5] C. Barnes, X. Irigoien, J. A. A. De Oliveira, D. Maxwell, and S. Jennings, *Predicting marine phytoplankton community size structure from empirical relationships with remotely sensed variables*, Journal of Plankton Research **33** (2011), no. 1, 13–24.
- [6] G. Becker, H. Holfeld, A. T. Hasselrot, D. M. Fiebig, and D. A. Menzler, *Use of a microscope photometer to analyze in vivo fluorescence intensity of epilithic microalgae grown on artificial substrata*, APPLIED AND ENVIRONMENTAL MICROBIOLOGY **63** (1997), no. 4, 1318–1325.
- [7] K. R. Beebe, R. J. Pell, and M.B. Seasholtz, *Chemometrics: A practical guide*, John Wiley and Sons, New York, 1998.

- [8] M. J. Behrenfeld, R. T. O'Malley, D. A. Siegel, C. R. McClain, J. L. Sarmiento, G. C. Feldman, A. J. Milligan, P. G. Falkowski, R. M. Letelier, and E. S. Boss, *Climate-driven trends in contemporary ocean productivity*, Nature **444** (2006), no. 7120, 752–755.
- [9] M. C. Benfield, P. Grosjean, P. F. Culverhouse, X. Irigoien, M. E. Sieracki, A. Lopez-Urrutia, H. G. Dam, Q. Hu, C. S. Davis, A. Hansen, C. H. Pilskaln, E. M. Riseman, H. Schultz, P. E. Utgoff, and G. Gorsky, *Rapid research on automated plankton identification*, Oceanography **20** (2007), no. 2, 172–187.
- [10] M. Beutler, K. H. Wiltshire, B. Meyer, C. Moldaenke, C. Luring, M. Meyerhofer, U. P. Hansen, and H. Dau, *A fluorometric method for the differentiation of algal populations in vivo and in situ*, Photosynthesis Research **72** (2002), no. 1, 39–53.
- [11] R. R. Bidigare and M. E. Ondrusek, *Spatial and temporal variability of phytoplankton pigment distributions in the central equatorial pacific ocean*, Deep-Sea Research Part II-Topical Studies in Oceanography **43** (1996), no. 4-6, 809–833.
- [12] P. W. Boyd, S. C. Doney, R. Strzepek, J. Dusenberry, K. Lindsay, and I. Fung, *Climate-mediated changes to mixed-layer properties in the southern ocean: assessing the phytoplankton response*, Biogeosciences **5** (2008), no. 3, 847–864, 321KK Times Cited:16 Cited References Count:106.
- [13] P. W. Boyd and P.P. Newton, *Does planktonic community structure determine downward particulate organic carbon flux in different oceanic provinces?*, Deep-Sea Research Part I-Oceanographic Research Papers **46** (1999), no. 1, 63–91.
- [14] R. G. Brereton, *Chemometrics for pattern recognition: A practical guide*, John Wiley and Sons, Chichester, UK, 2009.

- [15] L. S. Bruckman, T. L. Richardson, J. A. Swanstrom, K. A. Donaldson, M. Al-lora, T. J. Shaw, and M. L. Myrick, *Linear discriminant analysis of single-cell fluorescence excitation spectra of five phytoplankton species*, Applied Spectroscopy **66** (2012), no. 1, 60–65.
- [16] J. Canny, *A computational approach to edge detection*, IEEE Transactions on Pattern Analysis and Machine Intelligence **8** (1986), no. 6, 679–698.
- [17] F.P. Chavez, K.R. Buck, and R.T. Barber, *Phytoplankton taxa in relation to primary production in the equatorial pacific*, Deep-Sea Research Part A-Oceanographic Research Papers **37** (1990), 1733–1752.
- [18] G.L. Clarke, G.C. Ewing, and C.J. Lorenzen, *Spectra of backscattered light from sea obtained from aircraft as a measure of chlorophyll concentration*, SCIENCE **167** (1970), 1119.
- [19] A. Clauset, C.R. Shalizi, and M.E.J. Newman, *Power-law distributions in empirical data*, SIAM Review **51** (2009), no. 4, 661–703.
- [20] P. F. Culverhouse, R. Williams, B. Reguera, V. Herry, and S. Gonzalez-Gil, *Do experts make mistakes? a comparison of human and machine identification of dinoflagellates*, Marine Ecology-Progress Series **247** (2003), 17–25.
- [21] P. F. Culverhouse, R. Williams, B. Simpson, C. Gallienne, B. Reguera, M. Cabrini, S. Fonda-Umani, T. Parisini, F. A. Pellegrino, Y. Pazos, H. Wang, L. Escalera, A. Morono, M. Hensey, J. Silke, A. Pellegrini, D. Thomas, D. James, M. A. Longa, S. Kennedy, and G. del Punta, *Hab buoy: a new instrument for in situ monitoring and early warning of harmful algal bloom events*, African Journal of Marine Science **28** (2006), no. 2, 245–250.

- [22] C. S. Davis, Q. Hu, S. M. Gallager, X. Tang, and C. J. Ashjian, *Real-time observation of taxa-specific plankton distributions: an optical sampling method*, Marine Ecology-Progress Series **284** (2004), 77–96.
- [23] C. S. Davis, F. T. Thwaites, S. M. Gallager, and Q. Hu, *A three-axis fast-tow digital video plankton recorder for rapid surveys of plankton taxa and hydrography*, Limnology and Oceanography-Methods **3** (2005), 59–74.
- [24] S. J. Davis, K. Caldeira, and J. D. Mathews, *Future co2 emissions and climate change from existing energy infrastructure*, Science **329** (2010), 1330–1333.
- [25] T. D. Dickey, *The emergence of concurrent high-resolution physical and biooptical measurements in the upper ocean and their applications*, Reviews of Geophysics **29** (1991), no. 3, 383–413.
- [26] K. V. Embleton, C. E. Gibson, and S. I. Heaney, *Automated counting of phytoplankton by pattern recognition: a comparison with a manual counting method*, Journal of Plankton Research **25** (2003), no. 6, 669–681.
- [27] P. G. Falkowski and J. A. Raven, *Aquatic photosynthesis*, 2nd ed., Princeton University Press, Princeton, 2007.
- [28] R. A. Fisher, *The use of multiple measurements in taxonomic problems*, Ann. Eugenics **7** (1936), 179–188.
- [29] M. H. Forget, C. Fuentes-Yaco, S. Sathyendranath, T. Platt, J. Pommier, and E. Devred, *Computation of primary production from remote sensing of ocean colour at the northwestern atlantic c-solas lagrangian site*, Marine Ecology-Progress Series **352** (2007), 27–38, 257JQ Times Cited:2 Cited References Count:35.
- [30] G. Friedrich, V. Gerhardt, U. Bodemer, and M. Pohlmann, *Phytoplankton composition and chlorophyll concentration in freshwaters: comparison of delayed*

- fluorescence excitation spectroscopy, extractive spectrophotometric method, and utermohl- method*, *Limnologica* **28** (1998), 323–328.
- [31] T. Fujiki, K. Matsumoto, S. Watanabe, T. Hosaka, and T. Saino, *Phytoplankton productivity in the western subarctic gyre of the north pacific in early summer 2006*, *Journal of Oceanography* **67** (2011), no. 3, 295–303.
 - [32] P. Gemperline, *Practical guide to chemometrics*, CRC Press, Boca Raton, 2006.
 - [33] G. Gorsky, P. Guilbert, and E. Valenta, *The autonomous image analyzer - enumeration, measurement and identification of marine-phytoplankton*, *Marine Ecology-Progress Series* **58** (1989), no. 1-2, 133–142.
 - [34] R. R. Guillard, *Culture of phytoplankton for feeding marine invertebrates*, *Culture of Marine Invertebrate Animals*, Plenum Press, New York, 1975.
 - [35] R. R. Guillard and J. H. Ryther, *Studies of marine planktonic diatoms .1. cyclotella nana hustedt, and detonula confervacea (cleve) gran*, *Canadian Journal of Microbiology* **8** (1962), no. 2, 229–239.
 - [36] F. G. Haibach, A. E. Greer, M. V. Schiza, R. J. Priore, O. O. Soyemi, and M. L. Myrick, *On-line reoptimization of filter designs for multivariate optical elements*, *Applied Optics* **42** (2003), no. 10, 1833–1838.
 - [37] B. Hansen, P. K. Bjørnsen, and P. J. Hansen, *The size ratio between planktonic predators and their prey*, *Limnology and Oceanography* **39** (1994), no. 2, 395–403.
 - [38] N. Hashemi, J. S. Erickson, J. P. Golden, K. M. Jackson, and F. S. Ligler, *Microflow cytometer for optical analysis of phytoplankton*, *Biosensors and Bioelectronics* **26** (2011), no. 11, 4263–4269.

- [39] L. S. Hill, T. L. Richardson, L. T. M. Profeta, T. J. Shaw, C. J. Hintz, B. S. Twining, E. Lawrenz, and M. L. Myrick, *Construction, figures of merit, and testing of a single-cell fluorescence excitation spectroscopy system*, Review of Scientific Instruments **81** (2010), no. 1, 013103.
- [40] J. W. Hofstraat, W. J. M. Vanzeijl, M. E. J. Devreeze, J. C. H. Peeters, L. Peperzak, F. Colijn, and T. W. M. Rademaker, *Phytoplankton monitoring by flow-cytometry*, Journal of Plankton Research **16** (1994), no. 9, 1197–1224.
- [41] Q. Hu and C. Davis, *Automatic plankton image recognition with co-occurrence matrices and support vector machine*, Marine Ecology-Progress Series **295** (2005), 21–31, 947UC Times Cited:16 Cited References Count:49.
- [42] E. B. Huerta, B. Duval, and J. K. Hao, *A hybrid lda and genetic algorithm for gene selection and classification of microarray data*, Neurocomputing **73** (2010), 2375–2383.
- [43] S. W. Jeffrey and G. F. Humphrey, *New spectrophotometric equations for determining chlorophylls a, b, c1 and c2 in higher-plants, algae and natural phytoplankton*, BIOCHEMIE UND PHYSIOLOGIE DER PFLANZEN **167** (1975), 191–194.
- [44] S. W. Jeffrey and M. Vesk, *Introduction of marine phytoplankton and their pigment signatures*, 2nd ed., Phytoplankton Pigments in Oceanography: Guidelines to Modern Methods, UNESCO, Paris, 2005.
- [45] H. P. Jeffries, M. S. Berman, A. D. Poularikas, C. Katsinis, I. Melas, K. Sherman, and L. Bivins, *Automated sizing, counting and identification of zooplankton by pattern-recognition*, Marine Biology **78** (1984), no. 3, 329–334.
- [46] S. M. Kendall and A. Stuart, *The advanced theory of statistics: Volume 1 distribution theory*, Macmillan Publishing Co., Inc, Cambridge, 1977.

- [47] T. K. Kim, H. Kim, W. Hwang, and J. Kittler, *Component-based lda face description for image retrieval and mpeg-7 standardization*, Image Vis. Comput. **23** (2005), no. 7, 631–642.
- [48] C. S. Kinkade, J. Marra, T. D. Dickey, and R. Weller, *An annual cycle of phytoplankton biomass in the arabian sea, 1994-1995, as determined by moored optical sensors*, Deep-Sea Research Part II-Topical Studies in Oceanography **48** (2001), no. 6-7, 1285–1301.
- [49] Z. Kolber and P. G. Falkowski, *Use of active fluorescence to estimate phytoplankton photosynthesis in-situ*, Limnology and Oceanography **38** (1993), no. 8, 1646–1665.
- [50] C. M. Lalli and T. R. Parsons, *Biological oceanography: An introduction*, second ed., Butterworth-Heinemann, 1997.
- [51] A.J. Lewitus, R.A. Horner, D.A. Caron, E. Garcia-Mendoza, B.M. Hickey, M. Hunter, D.D. Huppert, R.M. Kudela, G.W. Langlois, J.L. Largier, E.J. Lessard, R. RaLonde, J.E.J. Rensel, P.G. Strutton, V.L. Trainer, and J.F. Tweddle, *Harmful algal blooms along the north american west coast region: History, trends, causes, and impacts*, HARMFUL ALGAE **19** (2012), 133–159.
- [52] L. H. Liska, P. R. Epstein, and W. H. Schesinger, *Rising co₂, climate change, and public health: Exploring the links to plant biology*, Environmental Health Perspective **117** (2009), 155–158.
- [53] D Livingstone, *A practical guide to scientific data analysis*, John Wiley and Sons, Chichester, UK, 2009.
- [54] A. Longhurst, S. Sathyendranath, T. Platt, and C. Caverhill, *An estimate of global primary production in the ocean from satellite radiometer data*, Journal of Plankton Research **17** (1995), no. 6, 1245–1271.

- [55] M. D. Mackey, D. J. Mackey, H. W. Higgins, and S. W. Wright, *Chemtax - a program for estimating class abundances from chemical markers: Application to hplc measurements of phytoplankton*, Marine Ecology-Progress Series **144** (1996), no. 1-3, 265–283.
- [56] J. Mazanec, M. Melisek, M. Oravec, and J. Pavlovicova, *Support vector machines, pca and lda in face recognition*, J. Electr. Eng. **59** (2008), 203–209.
- [57] M. L. Myrick, O. Soyemi, J. Karunamuni, D. Eastwood, H. Li, L. Zhang, A. E. Greer, and P. Gemperline, *A single-element all-optical approach to chemometric prediction*, Vibrational Spectroscopy **28** (2002), no. 1, 73–81.
- [58] Micheal L. Myrick, Olusola. Soyemi, Fred Haibach, Lixia Zhang, Ashley Greer, Hongli Li, Ryan Priore, Maria V. Schiza, and J. R. Farr, *Application of multivariate optical computing to near-infrared imaging*, Proceedings of SPIE **4577** (2002), no. Vibrational Spectroscopy-Based Sensor Systems, 148–157.
- [59] M. P. Nelson, J. F. Aust, J. A. Dobrowolski, P. G Verly, and M. L. Myrick, *Multivariate optical computing for predictive spectroscopy*, Analytical Chemistry **70** (1997), 73–82.
- [60] T. Okaishi, *Red tides: Biology*, Kluwer Academic Pub, Boston, 1994.
- [61] R. J. Olson, A. Shalapyonok, and H. M. Sosik, *An automated submersible flow cytometer for analyzing pico- and nanophytoplankton: Flowcytobot*, Deep-Sea Research Part I-Oceanographic Research Papers **50** (2003), no. 2, 301–315.
- [62] R. J. Olson and H. M. Sosik, *A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging flowcytobot*, Limnology and Oceanography-Methods **5** (2007), 195–203.

- [63] ———, *A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging flowcytobot*, *Limnology and Oceanography-Methods* **5** (2007), 195–203.
- [64] R. J. Olson, D. Vault, and S.W Chisholm, *Marine-phytoplankton distributions measured using shipboard flow-cytometry*, *DEEP-SEA RESEARCH PART A-OCEANOGRAPHIC RESEARCH PAPERS* **32** (1985), 1273–1280.
- [65] R. Parsons, *Sputter deposition process*, *Thin Film Processes II*, Academic Press, San Diego, 1991.
- [66] M. R. Pearl, J. A. Swanstrom, L. S. Bruckman, T. L. Richardson, T. J. Shaw, H. Sosik, and M.L. Myrick, *Taxonomic classification of phytoplankton with multivariate optical computing, part iii: Demonstration*, *Applied Spectroscopy* **67** (2013), no. 6.
- [67] M. Picheral, L. Guidi, L. Stemann, D. M. Karl, G. Iddaoud, and G. Gorsky, *The underwater vision profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton*, *Limnology and Oceanography-Methods* **8** (2010), 462–473, 678UM Times Cited:1 Cited References Count:39.
- [68] V. V. Povazhnyi, *Determination of the chlorophyll "a" concentration using a combined method based on measurements with a modified photometer*, *Oceanography* **66** (2012), no. 1, 60–65.
- [69] R. J. Priore, *The design, fabrication and implementation of point detection and chemical imaging sensors utilizing optical and digital regression techniques.*, Ph.D. thesis, 2005.

- [70] R. J. Priore, F. G. Haibach, M. V. Schiza, A. E. Greer, D. L. Perkins, and M. L. Myrick, *Miniature stereo spectral imaging system for multivariate optical computing*, Applied Spectroscopy **58** (2004), no. 7, 870–873.
- [71] Colin Reynolds, *Ecology of phytoplankton*, no. 53, Cambridge University Press, Cambridge, 2006.
- [72] P. J. Rice and B. P. Horgan, *Evaluation of nitrogen and phosphorus transport with runoff from fairway turf managed with hollow tine core cultivation and verticutting*, Science of The Total Environment **456-457** (2013), 61–68.
- [73] T. L. Richardson, E. Lawrenz, J. L. Pinckney, R. C. Guajardo, E. A. Walker, H. W. Paerl, and H. L. MacIntyre, *Spectral fluorometric characterization of phytoplankton community composition using the algae online analyser (r)*, Water Research **44** (2010), no. 8, 2461–2472.
- [74] R. B. Rivkin and L. Legendre, *Biogenic carbon cycling in the upper ocean: Effects of microbial respiration*, Science **291** (2001), no. 5512, 2398–2400.
- [75] K. Rodenacker, B. Hense, U. Jutting, and P. Gais, *Automatic analysis of aqueous specimens for phytoplankton structure recognition and population estimation*, Microscopy Research and Technique **69** (2006), no. 9, 708–720.
- [76] K.S. Rowan, *Photosynthetic pigments of algae*, Cambridge University Press, 1989.
- [77] T. P. A. Rutten, B. Sandee, and A. R. T. Hofman, *Phytoplankton monitoring by high performance flow cytometry: A successful approach?*, Cytometry Part A **64A** (2005), no. 1, 16–26.
- [78] J. H. Ryther, *Photosynthesis and fish production in sea*, Science **166** (1969), no. 3901, 72–78.

- [79] J. L. Sarmiento and N. Gruber, *Ocean biogeochemical dynamics*, Princeton University Press, Princeton, 2006.
- [80] S. Sathyendranath, G. Cota, V. Stuart, H. Maass, and T. Platt, *Remote sensing of phytoplankton pigments: a comparison of empirical and theoretical approaches*, International Journal of Remote Sensing **22** (2001), no. 2-3, 249–273.
- [81] M.R. Schiaffino, J.M. Gaslo, I. Izaguirre, and F. Unrein, *Picoplankton abundance and cytometric group diversity along a trophic and latitudinal lake gradient*, AQUATIC MICROBIAL ECOLOGY **68** (2013), 231–250.
- [82] A. Schmittner, A. Oschlies, H. D. Mathews, and E. D. Galbraith, *Future changes in climate, ocean circulation, ecosystems, and biogeochemical cycling simulated for a business-as-usual co2 emission scenario until year 4000 ad*, Global Biogeochemical Cycles **22** (2008), 1013–1033.
- [83] Howard M. Shapiro, *Practical flow cytometry*, 4th ed., John Wiley and Sons, Hoboken, New Jersey, 2003.
- [84] A. Sharma and K. K. Paliwal, *Cancer classification by gradient lda technique using microrray gene expression data*, Data Knowl. Eng. **66** (2008), 338–347.
- [85] A. Sharma, K. K. Paliwal, and G. C. Onwubolu, *Class-dependent pca, mdc and lda: A combined classifier for pattern classification*, Pattern Recognition **39** (2006), 1215–1229.
- [86] M. N. Simcock and M. L. Myrick, *Precision in imaging multivariate optical computing*, Applied Optics **46** (2007), no. 7, 1066–1080.
- [87] H. M. Sosik and R. J. Olson, *Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry*, Limnology and Oceanography-Methods **5** (2007), 204–216, ISI Document Delivery No.: 200DB Times Cited:

7 Cited Reference Count: 38 Sosik, Heidi M. Olson, Robert J. AMER SOC LIMNOLOGY OCEANOGRAPHY.

- [88] H. M. Sosik, R. J. Olson, M. G. Neubert, A. Shalapyonok, and A. R. Solow, *Growth rates of coastal phytoplankton from time-series measurements with a submersible flow cytometer*, Limnology and Oceanography **48** (2003), no. 5, 1756–1765, 723LK Times Cited:23 Cited References Count:37.
- [89] O. Soyemi, D. Eastwood, L. Zhang, H. Li, J. Karunamuni, P. Gemperline, R. A. Synowicki, and M. L. Myrick, *Design and testing of a multivariate optical element: The first demonstration of multivariate optical computing for predictive spectroscopy*, Analytical Chemistry **73** (2001), no. 17, 4393–4393.
- [90] O. Soyemi, F. G. Haibach, P. J. Gemperline, and M. L. Myrick, *Nonlinear optimization algorithm for multivariate optical element design*, Applied Spectroscopy **56** (2002), no. 4, 477–487.
- [91] M.A. Srokosz, A.P. Martin, and M.J.R. Fasham, *On the role of biological dynamics in plankton patchiness at the mesoscale: An example from the eastern north atlantic ocean*, Journal of Marine Research **61** (2003), 517–537.
- [92] J. A. Swanstrom, L. S. Bruckman, M. R. Pearl, E. Abernathy, T. L. Richardson, T. J. Shaw, and M. L. Myrick, *Taxonomic classification of phytoplankton with multivariate optical computing, part ii: Design and experimental protocol of a shipboard fluorescence imaging photometer*, Applied Spectroscopy **67** (2013), no. 6.
- [93] Joseph A. Swanstrom, Laura S. Bruckman, Megan R. Pearl, Michael N. Simcock, Kathleen A. Donaldson, Tammi L. Richardson, Timothy J. Shaw, and Michael L. Myrick, *Taxonomic classification of phytoplankton with multivariate*

- optical computing, part i: Design and theoretical performance of multivariate optical elements*, Applied Spectroscopy **67** (2013), no. 6.
- [94] C. M. Thomas, *Identifying marine phytoplankton*, Academic Press, New York, 1997.
 - [95] M. Thyssen, G. Tarran, M.V. Zubkov, R.J. Holland, and G. Gregori, *The emergence of automated high-frequency flow cytometry: revealing temporal and spatial phytoplankton variability*, Journal of Plankton Research **30** (2008), no. 3, 333–343.
 - [96] B.J. Trask, G.J. Vandenengh, and J.H.B.W. Elgershuizen, *Analysis of phytoplankton by flow-cytometry*, Cytometry **2** (1982), 258–264.
 - [97] Takashi Tsuji and Takashi Nishikawa, *Automated identification of red tide phytoplankton *prorocentrum triestinum* in coastal areas by image analysis*, Journal of the Oceanographical Society of Japan **40** (1984), 425–431.
 - [98] D. Uhlmann, H. Mihan, and A. Gnauck, *Variations of oxygen-content and of biological structure of extremely eutrophied waters under constant environmental conditions (model experiments)*, Acta Hydrochimica Et Hydrobiologica **6** (1978), no. 5, 421–444.
 - [99] D. Uhlmann, O. Schlimpert, and W. Uhlmann, *Automated phytoplankton analysis by a pattern recognition method*, Internationale Revue der Gesamten Hydrobiologie **63** (1978), no. 4, 575–583.
 - [100] Z. Uysal, *Chroococcoid cyanobacteria *synechococcus* spp. in the black sea: pigments, size, distribution, growth and diurnal variability*, Journal of Plankton Research **23** (2001), no. 2, 175–189.

- [101] J. Vepsäläinen, T. Pyhälähti, E. Rantajarvi, K. Kallio, S. Pertola, T. Stipa, M. Kiirikki, J. Pulliainen, and J. Seppälä, *The combined use of optical remote sensing data and unattended flow-through fluorometer measurements in the baltic sea*, International Journal of Remote Sensing **26** (2005), no. 2, 261–282.
- [102] A. Verikas, A. Gelzinis, M. Bacauskiene, I. Olenina, S. Olenin, and E. Vaiciukynas, *Phase congruency-based detection of circular objects applied to analysis of phytoplankton images*, Pattern Recognition **45** (2012), no. 4, 1659–1670.
- [103] D Voet and J.G Voet, *Biochemistry: Biomolecules, mechanisms of enzyme action, and metabolism*, 3rd ed., vol. 1, John Wiley and Sons, 2004.
- [104] H. Y. Wang, Z. F. Wang, Y. Leng, X. J. Wu, and Q Li, *Pca plus f-lda: A new approach to face recognition*, Int. J. Pattern Recognition **21** (2007), 1059–1068.
- [105] M. L. Yallop, *Distribution patterns and biomass estimates of diatoms and autotrophic dinoflagellates in the ne atlantic during june and july 1996*, Deep-Sea Research Part Ii-Topical Studies in Oceanography **48** (2001), no. 4-5, 825–844.
- [106] H. Yang and J. Y. Yang, *Why can lda be performed in pca transformed space?*, Pattern Recognition **36** (2003), 563–566.
- [107] C. S. Yentsch and D. A. Phinney, *Spectral fluorescence - an ataxonomic tool for studying the structure of phytoplankton populations*, Journal of Plankton Research **7** (1985), no. 5, 617–632.
- [108] C. S. Yentsch and C. M. Yentsch, *Fluorescence spectral signatures - characterization of phytoplankton populations by the use of excitation and emission-spectra*, Journal of Marine Research **37** (1979), no. 3, 471–483.
- [109] Y. Zhao, J. Wang, Q. J. Lu, and R. Jiang, *Pattern recognition of eggshell cracks using pca and lda*, Innov. Food Sci. Emerg. Technol **11** (2011), 520–525.

APPENDIX A

SNR MATLAB CODE

```
1 function [goodstrk,bestframes,prd,prdback,intspot,intspotback,br,  
    goodback,gdstrk,temp0,temp1,pbegin,pend,varargout] =  
    manstrkanalysis_back2(varargin)  
2 %Combination of IMOC Streak RSD functions. The function combine  
    manual and  
3 %automatic techniques for integrating streaks collected from the IMOC  
4 %instrurment for measuring the RSD of a background and flat field  
    corrected  
5 %image with no filters in any of the filterwheel positions.  
6  
7 %Inputs:  
8 % order of varargin=(normarray,bkg); varargout=[bkg,normarray];  
9 %  
10 % varargin- This is used to call in the number of  
11 % variables required, which can change depending on the new flat  
    field or new background.  
12 % If a new ff and bkg are needed, then varargin = [].  
13 % If a ff and/or bkg exist for a data set,  
14 % then "normarray" and "bkg" will be in the workspace and will need  
    to be  
15 % added in the order specified in line 2.  
16  
17 %Outputs:  
18 % goodstrk = selected tracks used for SNR calculations
```

```

19 % bestframes = frames containing tracks used for SNR calculations
20 % prd = array of 1's and 0's defining the streaks in the tracks.
21 % prdback = array of 1's and 0's defining the areas in the background
22 %     offset from the streaks in the tracks.
23 % intspot = array of integrated streaks.
24 % intspotback = array of integrated "background equivalent" streaks
25 % br = row vector representation of integrated columns in a track
26 % goodback = selected background tracks used for SNR calculations
27 % gdstrk = output track values for a "good" track.
28 % temp0 = location of blank or non-streak areas in the tracks
29 % temp1 = location of streaks in the tracks.
30 % pbegin = beginning columns of a streak.
31 % pend = ending columns of a streak.
32 % varargout = bkg and normarrays used for correcting raw images.
33
34 if nargin==0
35     [normarray,bkg,datap,data]=plankread();
36 elseif nargin==1
37     [datap,data,bkg]=plankread3(varargin{1}); %varargin{1}=normarray;
38 elseif nargin==2
39     [datap,data]=plankread2(varargin{1},varargin{2}); %1=normarray, 2=
        bkg
40 end
41
42
43 [goodframes] = framesort(allframes);
44 [goodback,goodstrk,bestframes,gdstrk] = strkarea_back2(goodframes);
45 [br] = peakthreshold2(goodstrk);
46 [prd,prdback,intspot,intspotback,temp0,temp1,pbegin,pend]=
    product_back2(br,goodstrk,goodback);
47
48 nout = max(nargout,1)-13;
49 if nout==1

```

```

50     varargout{1}=bkg;
51 elseif nout==2
52     varargout{1}=bkg;
53     varargout{2}=normarray;
54 else
55     varargout=[];
56 end
57
58 end

```

```

1 function [goodframes] = framesort(allframes)
2 %The function "framesort" is used to select frames that contain
3 %streaks
4 %that are in focus, straight and bright from a background subtraced,
5 %flate field corrected image. Typically 'allframes' should be a 500
6 %frame
7 %array that was produced after background and flatfield correcting a
8 %500 image .SPE file collected on the IMOC instrument. The resulting
9 %'goodframes' array is a preliminary sorted variable that will be
10 %further
11 %wittled down by "strkarea".
12 % JS 8/16/11
13 % for i=1:500
14 % allframes(:, :, i)=allframes1{i};
15 % end
16 a=size(allframes);
17 j=0;
18 for i=1:a(3)
19     pcolor(allframes(:, :, i));
20     shading interp;
21     drawnow;

```

```

19     title(num2str(i));
20     b=input('keep (1) or move on (space)? ','s');
21     if b=='1'
22         j=j+1;
23         goodframes(:, :, j)=allframes(:, :, i);
24     end
25 end
26
27 end

```

```

1 function [goodback,goodstrk,bestframes,gdstrk] = strkarea_back2(
    goodframes)
2 %The function 'strkarea' is used to extract only the rows of interest
    around a
3 %streak. This function is performed on the variable resulting from
4 %"framesort", 'goodframes'. It requires the user to identify with
    the mouse
5 %where the top and bottom of the streak is in the image by clicking
    on the
6 %uppermost part of the streak and dragging the mouse to the lowest
    part of
7 %the streak. The resulting arrays are 'goodstreak'
8 %and 'bestframes'. 'goodstrk' is an array where each row
9 %contains the sum of the columns selected from a single frame in
10 %'goodframes'. The rows in 'goodstrk' are corrected by "autobase4"
    to
11 % normalize the peaks to a common baseline and remove the artificial
    intensities
12 % below the baseline. 'bestframes' is an array where is frame
    coincides with the
13 %same row number in 'goodstrk' as a reference. 'goodstrk' is used in

```

```

14
15 % Autobase7 is used so the background baseline is fitted correctly
    without
16 % bias (JS 10/14/11)
17
18 figure(1);
19 figure(2);
20 a=size(goodframes);
21 gdstrk(1:a(3),1:256)=0;
22 gdback(1:a(3),1:256)=0;
23 goodstrk(1:a(3),1:256)=0;
24 goodback(1:a(3),1:256)=0;
25 gthold(a,1)=0;
26 for i=1:a(3)
27     mx=max(max(goodframes(:, :, i), [], 1));
28     mn=min(min(goodframes(:, :, i), [], 1));
29     temp=(goodframes(:, :, i)-mn)*64/(mx-mn);
30     figure(1)
31     image(temp)
32     title(num2str(i));
33     r=getrect;
34     r;
35     ybot=floor(r(2));
36     ybot;
37     ytop=ybot+ceil(r(4));
38     ytop;
39     %backtop=ytop+(ytop-ybot)+5;    %take background values from
        pixels above
40     %backbot=ytop+5;                %streak
41
42     backtop=ybot-5;                %take background values from
        pixels below streak
43     backbot=ybot-5-(ytop-ybot);

```

```

44     backtop;
45     backbot;
46     gdstrk(i,:)=sum(goodframes(ybot:ytot, :,i));
47     gdback(i,:)=sum(goodframes(backbot:backtop, :,i));
48 end
49
50 %added to autobase the streaks before selecting
51
52
53 for i=1:size(gdstrk,1)
54     [outy,thold]=autobase5(gdstrk(i,[1:256]));
55     [outyy,tholdyb]=autobase7(gdback(i,[1:256]));
56     goodstrk(i,:)=outy(1,:);
57     goodback(i,:)=outyy(1,:);
58     %goodback=gdback;
59     gthold(i,1)=thold(1,1);
60     gtholdb(i,1)=tholdyb(1,1);
61 end
62
63 % for i=1:size(gdstrk,1)
64 %     base=min(gdstrk(i,:));
65 %     gdstrkbase(i,:)=gdstrk(i,:)-base;
66 % end
67
68
69
70
71 bestframes=goodframes;    %added bestframes to coincide with
72 for i=a(3):-1:1           %image # in bestframes ref with row # in
    goodstreaks
73     figure(2)
74     image(temp)
75     title(num2str(i));

```

```

76     plot(goodstrk(i,:));
77     b=input('keep (1) or throw (space)?','s');
78     if b=='1'
79     else
80         goodstrk(i,:)=[]; %added bestframes to coincide with
81         goodback(i,:)=[];
82         gdstrk(i,:)=[];
83         bestframes(:, :, i)=[]; %image # in bestframes ref with row #
            in goodstreaks
84         gthold(i,:)=[];
85     end
86 end
87
88
89 for i=size(goodstrk,1):-1:1 %image # in bestframes ref with
    row # in goodstreaks
90     figure(2)
91     image(temp)
92     title(num2str(i));
93     plot(goodback(i,:));
94     b=input('keep (1) or throw (space)?','s');
95     if b=='1'
96     else
97         goodback(i,:)=[]; %added bestframes to coincide with
98         goodstrk(i,:)=[];
99         gdstrk(i,:)=[];
100        bestframes(:, :, i)=[]; %image # in bestframes ref with row #
            in goodstreaks
101        gthold(i,:)=[];
102    end
103 end
104
105 for i=1:size(goodstrk)

```

```

106 goodstrk(i,257)=gthold(i,1);
107 end
108
109 for i=1:size(goodback)
110 goodback(i,257)=gtholdb(i,1);
111 end
112
113 end

```

```

1 function [br] = peakthreshold2(goodstrk)
2
3 % Purpose: Compare each pixel to a threshold and write the score (0=
    below
4 % and 1=above) to 'br'. 'br' is used in "product" to get only the
    values
5 % of the peaks and remove the values of the baseline from the
    variable
6 % 'goodstrk'.
7
8 k=size(goodstrk);
9 for i=1:k
10     for j=1:256
11         br(i,j)=goodstrk(i,j)-goodstrk(i,257);
12         if (br(i,j)≤0)
13             br(i,j)=0;
14         else br(i,j)=1;
15         end
16     end
17     for k=2:255 % Looks to either side of a 0 to see if
        there is a 1, if so change it to a one, if not leave it.
18         if br(i,k)==0

```



```

19         if br(i,k-1)==1
20             if br(i,k+1)==1
21                 br(i,k)=1;
22             else
23                 end
24         else
25             end
26     else
27         end
28 end
29 for k=2:255           % Looks to either side of a 0 to see if
                        there is a 1, if so change it to a one, if not leave it.
30     if br(i,k)==1
31         if br(i,k-1)==0
32             if br(i,k+1)==0
33                 br(i,k)=0;
34             else
35                 end
36         else
37             end
38     else
39         end
40 end
41 end

```

```

1 function [prd,prdback,intspot,intspotback,temp0,temp1,pbegin,pend]=
    product_back2(br,goodstrk,goodback)
2 %The purpose of "product" is to make a variable to hold
3 %The ouput values are:
4 %prd - Each row is from an individual streak represented in 'goodstrk
    ' and

```

```

5 %it is multiplied by the value in the same row and column br, so the
6 %resulting row has column values with the baseline values of '0' and
   peak
7 %values with the values in 'goodsrk'.
8 %strkstat—the first column is the average intensity of each spot in a
9 %streak for the given row in 'intspot' and 'goodstrk', the second
   column is
10 %the standard deviation, third column is the value for the spot with
   the
11 %highest intensity and the fourth is of the spot with the lowest
   intensity.
12 %intspot—is an array with the values of the individual spot values in
   each
13 %column before they are summed and stats are collected.
14
15 %goodstrk(:,256)=0; used before row 18–20 commented code was added.
16 prd = br .* goodstrk(:,(1:256));
17
18 for m=1:size(prd)
19     for s=1:256
20         if prd(m,s)<0
21             prd(m,s)=1;
22         else
23             end
24         end
25 end
26
27 %prd needs to remove any remnants of a streak where column 1 had a
   value to
28 %where a zero starts, and any remnant of a streak where column 256
   is a
29 %number not a zero.
30 for g=1:size(prd) %removes values from a partial first streak

```

```

31     g;
32     if prd(g,1)==0
33         h=2;
34     else
35         h=1;
36         f=1;
37         prd(g,f)=0;
38         while h==1
39             if prd(g,f+1)>0
40                 prd(g,f+1)=0;
41                 f=f+1;
42                 f;
43                 h=1;
44             else
45                 h=0;
46             end
47         end
48     end
49 end
50
51 for g=1:size(prd) %removes values from a partial last streak
52     g;
53     if prd(g,256)==0
54         h=2;
55     else
56         h=1;
57         f=256;
58         prd(g,f)=0;
59         while h==1
60             if prd(g,f-1)>0
61                 prd(g,f-1)=0;
62                 f=f-1;
63                 f;

```

```

64             h=1;
65         else
66             h=0;
67         end
68     end
69 end
70 end
71
72
73     for j=1:size(prd,1)
74         for k=1:size(prd,2)
75             if prd(j,k)==0
76                 prdback(j,k)=prd(j,k);
77             else
78                 prdback(j,k)=goodback(j,k);
79             end
80         end
81     end
82
83     %"streaktest" finds is a second integration function written to find
      the
84     %integrated intensity of a streak by summing the illumination event "
      spot"
85     %for all pixel values contained in the region half the distance
      between
86     %the peaks on either side of it.  prd column 2 needs to be zero,
      otherwise
87     %the first peak will not be recognized.  This affects only streaks
      where
88     %the first illumination event is only one pixel from the start, about
      one
89     %out of every 100 frames.
90

```

```

91 for i=1:size(prd)
92     prd(i,2)=0;
93 end
94
95 [temp0,temp1,pbegin,pend,intspot,intspotback] = streaktest (goodback,
    goodstrk,prd);
96
97
98 end

```

```

1 function [outy,thold]=autobase5(iny)
2
3 % Purpose: to autosubtract baseline from a profile
4 % outy is the baseline-subtracted profile;  sdbase is the standard
    deviation
5 % of the errors from the baseline for regions outside of peaks.  inx
    and
6 % iny are the x and y values for the curve that you want to perform a
7 % baseline fit to.  Order is either 1, 2 or 3 (higher is possible,
    but
8 % probably not too wise) – it represents the type of baseline you
    want to
9 % fit to.  1 would be a simple linear baseline;  2 would be a
    quadratic;  3
10 % would be a cubic.  Lower numbers are more conservative and less
    likely to
11 % give weird output shapes;  higher numbers make it possible to fit
    curvier
12 % baselines.  The output outy is the baseline corrected curve that
    can be
13 % plotted against inx (which is why we don't provide an outx array).

```

```

    The
14 % function is based on an autocurvefitting routine developed by
    Thermo
15 % Nicolet for the FTIR in our back laboratory – we got details from
    them of
16 % that it functions, which were published in a paper by Maria Schiza
    from
17 % our laboratory in a study of FTIR microscopy of bacterial spores.
    We
18 % modified the original to make it work better for the purpose of
    plankton
19 % streaks in the IMOC instrument.  MLM (8/17/11)
20
21 %a 3rd order fit gives a better fit to IMOC SNR data (sdbase=115)
    than
22 %1st(162.6)or 2nd(162.5) order and 4(109.7) gets a little better.  So
    use
23 %third order.  JS (8/17/11)
24 inx = [1:256];  %removed inx and order from inputs of function and
    set
25 order = 3;      %there values here for framesort.m JS 8/19/11
26
27 if inx(1)==1;
28     [y,yi]=max(iny);
29     inx=inx-yi;
30     % we're centering the spectrum around zero to make low orders
        easier to
31     % work with.  This is only done if the input x array is a
        sequence of
32     % numbers beginning with 1, i.e., not actual wavelengths for the
33     % spectrum.  In that case, the output outy should be plotted
        against
34     % the wavelengths if you want to see what the spectrum looks like

```

```

        - the
35     % inx array is just a bunch of numbers made up to make the job of
36     % fitting easier MLM
37     bot=min(iny);
38     iny2=iny;
39     inx2=inx;
40     for i=length(iny):-1:1
41         if iny(i)> (bot+(y-bot)*0.5))
42             iny2(i)=[];
43             inx2(i)=[];
44             % here we are eliminating all parts of the input curve
              that are
45             % within 20% of the highest point. The baseline function
              tries
46             % to fit the bottom of the curves, and keeping those
              highest
47             % points will sometimes cause crazy results, so we just
              get
48             % rid of them right off the bat MLM
49         end
50     end
51 end
52 outx=inx;
53 outy=iny;
54 %plot(inx,iny,'o');
55 %drawnow;
56 j=1;
57
58 while j≠0
59     j=j+1;
60     P = polyfit(inx2,iny2,order);
61     % here is where the curvefit is actually done, in an iterative
        loop

```

```

62     Y2 = polyval(P,inx2);
63     err=inY2-Y2;
64     abserr=abs(err);
65     % abserr is an array of absolute value of the errors between the
        fit
66     % and the original curve
67     j;
68     if (max(err) > 2.2*mean(abserr))
69         % if the maximum error in the array of errors is
            substantially
70         % larger than the average error, then we remove the point of
71         % maximum abs error from the arrays and repeat the curvefit,
            which
72         % gradually looks more and more just at the bottom of the
            curve.
73         % Note that err is the array of signed errors, abserr is
            unsigned.
74         % Thus we don't pay attention to errors that are negative -
            where
75         % the fit is well above the original data. Only where the
            original
76         % data are substantially higher than the fit. MLM
77     ymax=max(abserr);
78     abserr=abserr/ymax;
79     for i=length(abserr):-1:1
80         % in this function, we're often doing negative indexing so
            we
81         % can delete an element of the array without affecting our
82         % counting MLM
83         if abserr(i)==1
84             inx2(i)=[];
85             inY2(i)=[];
86             err(i)=[];

```



```

87         i;
88     end
89     i;
90 end
91 %plot(inx2,err,'o');
92 %drawnow;
93 %pause(0.2)
94 else
95     % if the error conditions don't apply, we set j=0, which ends
        the
96     % loop and we return. MLM
97     j=0;
98 end
99
100 end
101 % I don't recall why the stuff in green below was every here, or why
        it was
102 % commented out. Obviously it is designed to count how long some
103 % particular action took MLM
104
105 %tic
106 %m=1;
107 %while m~=0
108 %     y3=toc;
109 %     if y3>0.3
110 %         m=0;
111 %     else
112 %         m=m+1;
113 %     end
114 % end
115 Y2=polyval(P,outx);
116 outy=outy-Y2;
117 %sdbase=std(err);

```

```

118 thold=std(err)*3;
119 % above, we take the results of the curve fit and build a baseline
    for all
120 % the x values, then subtract that baseline from the original data
    and
121 % return it. Note that sdbase is the standard deviation of the
    baseline
122 % error - it doesn't include the points that were removed from the
    fit, so
123 % it really does mean something like "baseline error" f MLM
124
125 end %function autobase4

```

```

1
2 function [temp0,temp1,pbegin,pend,intspot,intspotback] = streaktest (
    goodback,goodstrk,prd)
3 %If the first peak is only one column away from the edge, the peak
    will
4 %not be counted.
5
6 for i=1:size(prd)
7     p=1;
8     a=1;
9     k=0;
10    m=0;
11    j=1;
12    q=1;
13    z=1;
14    %Look down the entire row, checking each column value to
        determine
15    %where a streak starts in the prd array. Once a non-zero value

```

```

        is
16    %found, the column number of the first non-zero number is saved
        in
17    %pbegin and the total number of columns before the start of the
        peak is
18    %saved in the temp0 array.
19
20    %Each column in temp0, temp1, pend, and pbegin correlates to the
        peak
21    %number in prd. For example, the value in column 1 in pbegin
22    %is the column number in prd where the first peak starts, column
        1 in
23    %pend is the column number in prd where the first peak ends,
        column 1
24    %in temp0 is the number of columns in prd before the first peak
        starts
25    %and column 1 in temp1 is the number of columns in prd that make
        up the
26    %first peak.
27    while z==1
28        while a==1
29            if j==256
30                a=0;
31                z=0;
32            elseif prd(i,j)==0
33                k=k+1;
34                j=j+1;
35                a=1;
36            else
37                pbegin_temp(i,p)=j;
38                a=2;
39                temp0(i,p)=k;
40                k=0;

```

```

41         p=p+1;
42     end
43 end
44     %After a peak is determined to have started, the row is
        followed down
45     %each column until a zero is found again. Then the last
        column of a
46     %zero is saved in pendtemp and the total number of columns in
        the peak is
47     %saved in temp1.
48     while a==2
49         if prd(i,j)>0.001
50             m=m+1;
51             j=j+1;
52             a=2;
53         else
54             pendtemp(i,q)=j-1;
55             a=1;
56             temp1(i,q)=m;
57             m=0;
58             q=q+1;
59             a=1;
60         end
61     end
62 end
63 end
64 %Below looks at values in temp1 and if the peak is less than 3 pixels
        wide,
65 %it is replaced with a zero and the value is stored in pbegintemp and
66 %pendtemp.
67 for i=1:size(temp1)
68     k=1;
69     for j=1:size(temp1,2)

```

```

70         if temp1(i,j)<3
71             peaksize(i,k)=temp1(i,j);
72             pbeginTemp(i,j)=0;
73             pendTemp(i,j)=0;
74             k=k+1;
75         else
76             k=k;
77         end
78     end
79 end
80
81 %Below looks at values in temp0 and if the gap between peaks is
    larger than
82 %2 pixels, if it is, then the values in pbeginTemp and pendTemp are
    kept.
83 %If it is not, then the pend for that column and the pbegin are made
    0,
84 %effectvily joining the two peeks. The next for loop removes the 0's
    and
85 %creates temp2 arrays.
86 for i=1:size(temp0)
87     k=1;
88     for j=1:size(temp0,2)
89         if temp0(i,j)<3
90             pbeginTemp(i,j)=0;
91             pendTemp(i,j)=0;
92         end
93     end
94 end
95
96 %Now we remove any 0's in temp arrays and store the non-zero values
    in
97 %temp2 arrays.

```

```

98
99 for i=1:size(pbegintemp)
100     k=1;
101     for j=1:size(pbegintemp,2)
102         if pbegintemp(i,j)>0
103             pbegintemp2(i,k)=pbegintemp(i,j);
104             pendtemp2(i,k)=pendtemp(i,j);
105             k=k+1;
106         else
107             k=k;
108         end
109     end
110 end
111
112
113
114 %The half way point between two peaks is found and tempt2 files are
115 %expanded in the begining and the end of a peak to include these
116 %values.
117 for i=1:size(pbegintemp2)
118     d=0;
119     for k=1:size(pbegintemp2,2) %counts how many non-zero values are
120         in a row
121             if pbegintemp2(i,k)>0
122                 d=d+1;
123             end
124         end
125     a=(pbegintemp2(i,2)-pendtemp2(i,1)-1)*.5; %using distance from
126         first and second peak as reference for start of peak one
127     a=floor(a);
128     pbegin(i,1)=pbegintemp2(i,1)-a;
129     pend(i,1)=pendtemp2(i,1)+a;
130     if pbegin(i,1)<1

```

```

129         pbegin(i,1)=1;
130     end
131     for j=2:d-1
132         a=(pbeginTemp2(i,j)-pendTemp2(i,j-1)-1)*.5; %a is distance
            from previous peak to current peak *0.5
133         a=floor(a);
134         b=(pbeginTemp2(i,j+1)-pendTemp2(i,j)-1)*.5; %b is distance
            from next peak to current peak *0.5
135         b=floor(b);
136         pbegin(i,j)=pbeginTemp2(i,j)-a;
137         pend(i,j)=pendTemp2(i,j)+b;
138     end
139
140     a=(pbeginTemp2(i,d)-pendTemp2(i,d-1)-1)*.5; %using distance from
            last and
141     a=floor(a); %second to last peak
            as reference for end of last peak
142     pbegin(i,d)=pbeginTemp2(i,d)-a;
143     pend(i,d)=pendTemp2(i,d)+a;
144     if pend(i,d)>255
145         pend(i,d)=255;
146     end
147 end
148
149 %The new peak widths stored in pbegin and pend are used to find
150 %the integrated intensities for each peak.
151 for i=1:size(pbegin)
152     a=pbegin(i,1); %start of first peak
153     b=pend(i,1); %end of first streak
154     inttemp=[];
155     inttemp(1,:)=goodstrk(i,a:b);
156     inttempback=goodback(i,a:b);
157     intspot(i,1)=sum(inttemp);

```

```

158     intspotback(i,1)=sum(inttempback);
159     clear inttemp;
160     clear inttempback;
161     d=0;
162     for k=1:size(pbegin,2) %counts how many non-zero values are in a
        row
163         if pbegin(i,k)>0
164             d=d+1;
165         end
166     end
167     for j=2:d-1
168         a=pbegin(i,j);
169         b=pbegin(i,j);
170         inttemp=goodstrk(i,a:b);
171         intspot(i,j)=sum(inttemp);
172         inttempback=goodback(i,a:b);
173         intspotback(i,j)=sum(inttempback);
174         clear inttemp;
175         clear inttempback;
176     end
177     a=pbegin(i,d); %last streak
178     b=pbegin(i,d);
179     inttemp=goodstrk(i,a:b);
180     intspot(i,d)=sum(inttemp);
181     inttempback=goodback(i,a:b);
182     intspotback(i,d)=sum(inttempback);
183     clear inttemp;
184     clear inttempback;
185 end
186
187
188 end

```



```

1
2 function [temp0,temp1,pbegin,pend,intspot,intspotback] = streaktest (
    goodback,goodstrk,prd)
3 %If the first peak is only one column away from the edge, the peak
    will
4 %not be counted.
5
6 for i=1:size(prd)
7     p=1;
8     a=1;
9     k=0;
10    m=0;
11    j=1;
12    q=1;
13    z=1;
14    %Look down the entire row, checking each column value to
        determine
15    %where a streak starts in the prd array. Once a non-zero value
        is
16    %found, the column number of the first non-zero number is saved
        in
17    %pbegin and the total number of columns before the start of the
        peak is
18    %saved in the temp0 array.
19
20    %Each column in temp0, temp1, pend, and pbegin correlates to the
        peak
21    %number in prd. For example, the value in column 1 in pbegin
22    %is the column number in prd where the first peak starts, column
        1 in
23    %pend is the column number in prd where the first peak ends,
        column 1

```

```

24      %in temp0 is the number of columns in prd before the first peak
        starts
25      %and column 1 in temp1 is the number of columns in prd that make
        up the
26      %first peak.
27      while z==1
28          while a==1
29              if j==256
30                  a=0;
31                  z=0;
32              elseif prd(i,j)==0
33                  k=k+1;
34                  j=j+1;
35                  a=1;
36              else
37                  pbegintemp(i,p)=j;
38                  a=2;
39                  temp0(i,p)=k;
40                  k=0;
41                  p=p+1;
42              end
43          end
44      %After a peak is determined to have started, the row is
        followed down
45      %each column until a zero is found again. Then the last
        column of a
46      %zero is saved in pendtemp and the total number of columns in
        the peak is
47      %saved in temp1.
48      while a==2
49          if prd(i,j)>0.001
50              m=m+1;
51              j=j+1;

```

```

52             a=2;
53         else
54             pendtemp(i,q)=j-1;
55             a=1;
56             temp1(i,q)=m;
57             m=0;
58             q=q+1;
59             a=1;
60         end
61     end
62 end
63 end
64 %Below looks at values in temp1 and if the peak is less than 3 pixels
    wide,
65 %it is replaced with a zero and the value is stored in pbegin temp and
66 %pendtemp.
67 for i=1:size(temp1)
68     k=1;
69     for j=1:size(temp1,2)
70         if temp1(i,j)<3
71             peaksize(i,k)=temp1(i,j);
72             pbegin temp(i,j)=0;
73             pendtemp(i,j)=0;
74             k=k+1;
75         else
76             k=k;
77         end
78     end
79 end
80
81 %Below looks at values in temp0 and if the gap between peaks is
    larger than
82 %2 pixels, if it is, then the values in pbegin temp and pendtemp are

```

```

        kept.
83 %If it is not, then the pend for that column and the pbegin are made
        0,
84 %effectvily joining the two peeks. The next for loop removes the 0's
        and
85 %creates temp2 arrays.
86 for i=1:size(temp0)
87     k=1;
88     for j=1:size(temp0,2)
89         if temp0(i,j)<3
90             pbegintemp(i,j)=0;
91             pendtemp(i,j)=0;
92         end
93     end
94 end
95
96 %Now we remove any 0's in temp arrays and store the non-zero values
        in
97 %temp2 arrays.
98
99 for i=1:size(pbegintemp)
100     k=1;
101     for j=1:size(pbegintemp,2)
102         if pbegintemp(i,j)>0
103             pbegintemp2(i,k)=pbegintemp(i,j);
104             pendtemp2(i,k)=pendtemp(i,j);
105             k=k+1;
106         else
107             k=k;
108         end
109     end
110 end
111

```

```

112
113
114 %The half way point between two peaks is found and tempt2 files are
115 %expanded in the begining and the end of a peak to include these
116 %values.
117 for i=1:size(pbegintemp2)
118     d=0;
119     for k=1:size(pbegintemp2,2) %counts how many non-zero values are
120         in a row
121         if pbegintemp2(i,k)>0
122             d=d+1;
123         end
124     end
125     a=(pbegintemp2(i,2)-pendtemp2(i,1)-1)*.5; %using distance from
126         first and second peak as reference for start of peak one
127     a=floor(a);
128     pbegin(i,1)=pbegintemp2(i,1)-a;
129     pend(i,1)=pendtemp2(i,1)+a;
130     if pbegin(i,1)<1
131         pbegin(i,1)=1;
132     end
133     for j=2:d-1
134         a=(pbegintemp2(i,j)-pendtemp2(i,j-1)-1)*.5; %a is distance
135             from previous peak to current peak *0.5
136         a=floor(a);
137         b=(pbegintemp2(i,j+1)-pendtemp2(i,j)-1)*.5; %b is distance
138             from next peak to current peak *0.5
139         b=floor(b);
140         pbegin(i,j)=pbegintemp2(i,j)-a;
141         pend(i,j)=pendtemp2(i,j)+b;
142     end
143
144     a=(pbegintemp2(i,d)-pendtemp2(i,d-1)-1)*.5; %using distance from

```

```

        last and
141     a=floor(a); %second to last peak
        as reference for end of last peak
142     pbegin(i,d)=pbegin2(i,d)-a;
143     pend(i,d)=pend2(i,d)+a;
144     if pend(i,d)>255
145         pend(i,d)=255;
146     end
147 end
148
149 %The new peak widths stored in pbegin and pend are used to find
150 %the integrated intensities for each peak.
151 for i=1:size(pbegin)
152     a=pbegin(i,1); %start of first peak
153     b=pend(i,1); %end of first streak
154     inttemp=[];
155     inttemp(1,:)=goodstrk(i,a:b);
156     inttempback=goodback(i,a:b);
157     intspot(i,1)=sum(inttemp);
158     intspotback(i,1)=sum(inttempback);
159     clear inttemp;
160     clear inttempback;
161     d=0;
162     for k=1:size(pbegin,2) %counts how many non-zero values are in a
        row
163         if pbegin(i,k)>0
164             d=d+1;
165         end
166     end
167     for j=2:d-1
168         a=pbegin(i,j);
169         b=pend(i,j);
170         inttemp=goodstrk(i,a:b);

```

```

171         intspot(i,j)=sum(inttemp);
172         inttempback=goodback(i,a:b);
173         intspotback(i,j)=sum(inttempback);
174         clear intemp;
175         clear intempback;
176     end
177     a=pbegin(i,d); %last streak
178     b=pending(i,d);
179     inttemp=goodstrk(i,a:b);
180     intspot(i,d)=sum(inttemp);
181     inttempback=goodback(i,a:b);
182     intspotback(i,d)=sum(inttempback);
183     clear intemp;
184     clear intempback;
185 end
186
187
188 end

```

APPENDIX B

SIMOCv2.0 MATLAB CODE

```
1 function [filtelec,deb_track,strk_hightbound,strk_widthbound,wid_x,
    deb_x,int_x,initial_wid_x,initial_width,final_width,x,
    initial_strk_hightbound,initial_strk_widthbound,varargout]=
    plankton_field_deblur(wn,varargin)
2 % Purpose: This function compiles all phytoplankton deblurring
    analysis functions to
3 % calculate the integrated intensity and size of each filter element.
4 %
5 %order of varargin=(normarray,bkg); varargout=[bkg,normarray]
6 %plankton_field_5 is used for phytoplankton.
7 %The use of varargin allows the user to determine if a new bkg and/or
8 %normarray are required for the data he/she wishes to correct.
9
10 %Inputs:
11 %
12 % wn-The total number of filter wheel positions in the filter wheel.
    i.e,
13 % the 7 position filter wheel with 6 filters and a blacked out blank
14 % position, wn=6.
15 %
16 % varargin- This is used to call in the number of
17 % variables required, which can change depending on the new flat
    field or new background.
18 % If a new ff and bkg are needed, then varargin = [].
```



```

19 % If a ff and/or bkg exist for a data set,
20 % then "normarray" and "bkg" will be in the workspace and will need
    to be
21 % added in the order specified in line 2.
22
23 %Outputs:
24 %
25 %filtelec = cell array with track boundaries.
26 % deb_track = a cell array with the image for each deblured track.
27 % strk_hightbound = a cell array with the row boundries of each
    streak.
28 % strk_widthbound = a cell array with the column boundries of each
    streak
29 % wid_x = a cell array with the final deblured widths in pixes of
    each
30 %   streak.
31 % deb_x = a cell array with the entire image deblured.
32 % int_x = the integrated values of each streak in a track
33 % initial_wid_x =a cell array with the starting widths in pixes of
    each
34 % streak.
35 % initial_width = array with starting streak widths.
36 % final_width = array with final streak widths.
37 % x = cell array with corrected images.
38 % initial_strk_hightbound = pre-deblured row boundries of each
    streak.
39 % initial_strk_widthbound = pre-deblured column boundries of each
    streak.
40 % varargout = depending on inpute values, varargout holds the
    background
41 % variable and flat field variables for correcting the raw images.
42
43

```

```

44
45 if nargin==1
46     [normarray,bkg,datap,data]=plankread();
47 elseif nargin==2
48     [datap,data,bkg]=plankread3(varargin{1}); %varargin{1}=normarray;
49 elseif nargin==3
50     [datap,data]=plankread2(varargin{1},varargin{2}); %1=normarray, 2=
        bkg
51 end
52
53
54 [r,p,x,y,th1]=planka_deblur(datap);
55 disp('a complete')
56
57 [filtelec]=plankb2(p,y,th1,x,r);
58 disp('b complete')
59
60 [ deb_track,strk_hightbound,strk_widthbound,wid_x,deb_x,int_x,
        initial_wid_x ] = plank_deblur6(wn,int,filtelec,x);
61 disp('deblur complete')
62
63
64 final_width=cell2mat(wid_x);
65 final_width= final_width';
66 final_avg_wid= mean(final_width(:,1));
67 final_std_wid=std(final_width(:,1));
68
69 initial_width=cell2mat(initial_wid_x);
70 initial_width= initial_width';
71 initial_avg_wid= mean(initial_width(:,1));
72 initial_std_wid=std(initial_width(:,1));
73
74

```

```

75 nout = max(nargout,1)-13;
76 if nout==1
77     varargout{1}=bkg;
78 elseif nout==2
79     varargout{1}=bkg;
80     varargout{2}=normarray;
81 else
82     varargout=[];
83 end
84
85 end %function plankton_field

```

```

1 function [normarray,bkg,datap,data]=plankread()
2     %need normarray, bkg and datap.
3 % Purpose: This function will read in .spe files that you have chosen
4     from a
5     % browser menu and perform all preprocessing.
6 %Choose dark, bkg and flat field data to create a corrected flat
7     field
8     %array.
9 [normarray,m]=normdat5;
10
11 % Upload a single bkg file that will be subtracted from the data
12     file. The
13 % bkg file will be divided into 256x256 images that will be averaged
14     (bkg) then
15 % dark count will be removed. This bkg-dark array, bd, will be
16     sutracted
17 % from the average data array.

```

```

15
16 [filename,pathname] = uigetfile('*.','Select the Background File for
    Data Analysis','MultiSelect','on');
17 fullname = strcat(pathname,filename);
18 fid=fopen(fullname);
19 fseek(fid,4100,'bof');
20 A=fread(fid,256*inf,'uint16');
21 [r c]=size(A);
22 count=r/(256^2);
23
24
25 bkg=reshape(A,256,256,count);
26 bkg=mean(bkg,3);
27
28
29 % bd=bkg;
30 % bd=bkg-dark; The bkg image already contains the dark noise, so no
    need to
31 % subtract it out of the bkg image. when subtracting background from
    sample image, the dark
32 % noise will also be subtracted.
33
34 % Upload a single data file and determine how many images are in the
    file,
35 % count. Reshape the data into a 3D array whose size is 256 x 256 x
    count.
36 %fclose(fid);
37 [filename,pathname] = uigetfile('*.','Select the Data File for
    Analysis','MultiSelect','on');
38 fullname = strcat(pathname,filename);
39 fid=fopen(fullname);
40 fseek(fid,4100,'bof');
41 A=fread(fid,256*inf,'uint16');

```

```

42 [r c]=size(A);
43 count=r/(256^2);
44
45 data=reshape(A,256,256,count);
46
47
48 % Preprocessing the data: subtract bd, divide by normarray
49 % (corrected flat field), multiply by m (average flat field pixel
50 % intensity).
51 [r,c,p]=size(data);
52 for i=1:p
53     datap(:, :, i)=((data(:, :, i)-bkg))./normarray);
54 end
55
56 end

```

```

1 function [datap,data]=plankread2(normarray,bkg)
2     %Already calculated normarray and bkg, need new datap
3 % Purpose: This function will read in .spe files that you choose from
4 % browser menu and perform a background fluorescence subtraction and a
5 % flatfield normalization correction. The difference between
6 % plankread and
7 % plankread2 is that plankread asks to read in a the files necessary
8 % to
9 % generate the 'dark', normlized flatfield 'normarray', and the
10 % background
11 % 'back', but plankread2 assumes that these files are already
12 % generated and
13 % can be applied to another set of data. If different correction
14 % files are

```

```

10 % needed, then plankread or plankread3 need to be used.
11
12 % JS (8-12-11)
13
14
15 [filename,pathname] = uigetfile('*.','Select the Data File for
    Analysis','MultiSelect','on');
16 fullname = strcat(pathname,filename);
17 fid=fopen(fullname);
18 fseek(fid,4100,'bof');
19 A=fread(fid,256*inf,'uint16');
20 [r c]=size(A);
21 count=r/(256^2);
22
23 for i=1:count
24     data=reshape(A,256,256,count);
25 end
26
27 % Preprocessing the data: subtract dark, subtract bd, divide by
    normarray
28 % (corrected flat field), multiply by m (average flat field pixel
29 % intensity).
30 [r,c,p]=size(data);
31 for i=1:p
32     datap(:,:,i)=(((data(:,:,i)-bkg))./normarray);
33 end
34
35
36 end

```

```

1 function [datap,data,bkg]=plankread3(normarray)

```

```

2      %Already calculated normarray, need new bkg and datap
3  % Purpose: This function will read in .spe files that you choose from
      a
4  % browser menu and perform a background fluorescence subtraction and a
5  % flatfield normalization correction. The difference between
      plankread and
6  % plankread2 is that plankread asks to read in a the files necessary
      to
7  % generate the 'dark', normlized flatfield 'normarray', and the
      background
8  % 'back', but plankread2 assumes that these files are already
      generated and
9  % can be applied to another set of data. If different correction
      files are
10 % needed, then plankread or plankread3 need to be used.
11
12 % JS (8-12-11)
13
14 [filename,pathname] = uigetfile('*.','Select the Background File for
      Data Analysis','MultiSelect','on');
15 fullname = strcat(pathname,filename);
16 fid=fopen(fullname);
17 fseek(fid,4100,'bof');
18 A=fread(fid,256*inf,'uint16');
19 [r c]=size(A);
20 count=r/(256^2);
21 bkg=reshape(A,256,256,count);
22 bkg=mean(bkg,3);
23
24
25 [filename,pathname] = uigetfile('*.','Select the Data File for
      Analysis','MultiSelect','on');
26 fullname = strcat(pathname,filename);

```

```

27 fid=fopen(fullname);
28 fseek(fid,4100,'bof');
29 A=fread(fid,256*inf,'uint16');
30 [r c]=size(A);
31 count=r/(256^2);
32
33 for i=1:count
34     data=reshape(A,256,256,count);
35 end
36
37 % Preprocessing the data: subtract dark, subtract bd, divide by
    normarray
38 % (corrected flat field), multiply by m (average flat field pixel
39 % intensity).
40 [r,c,p]=size(data);
41 for i=1:p
42     datap(:,:,i)=((data(:,:,i)-bkg))./normarray);
43 end
44
45
46 end

```

```

1 function [r,p,x,y,th1]=planka_deblur(datap)
2
3 % Purpose: Find the average & std. dev of the columns of an image.
4 % Baseline correct the std dev using autobase2. Calculate a
    threshold from
5 % the baseline corrected std dev using a histogram to determine the
    average
6 % baseline value. This average baseline value is used to determine
    the

```



```

7 % threshold value for selecting a track in an image.
8
9 % For use in plankton_field.m
10
11 %autobase_track is an adaptation of the autobase function Dr. Myrick
    and
12 %Puala used to baseline correct IR spectra. I modified it so that
    the
13 %outputs are y(3,:), the baseline corrected spectrum, and sdbase, is
    the sd of the baseline are the
14 %same. But, the th1 variable is modified from the original in that
    after the
15 %intensities within 50% of the maximum value are removed, the max
    value and
16 %min value of the corrected baseline are used to approximate the
    noise of
17 %the baseline. The threshold of a track is then equal the 2x the
    noise of
18 %the baseline. JS (4-9-12)
19
20 [r,c,p]=size(datap);
21
22 for i=1:p
23     x{i}=datap(:, :, i)';
24     x_gray{i}=mat2gray(x{i});
25     x_adj{i}=imadjust(x_gray{i});
26     y{i}(1,:)=mean(x_adj{i});
27     y{i}(2,:)=std(x_adj{i});
28     inx{i}(1,:)=1:r;
29     [y{i}(3,:),sdbase,th1(i)]=autobase_track(inx{i},y{i}(2,:),2);
30 %     [y{i}(3,:)=autobase2(inx{i},y{i}(2,:),2);
31 %     rat{i}=y{i}(3,:)./y{i}(1,:);
32 %     mins(i)=min(y{i}(2,:));

```

```

33 %      [freq1{i},xout1{i}]=hist(y{i}(3,:),r);
34 %      [vfq1(i),rfq1(i)]=max(freq1{i});
35 %      th1(i)=(xout1{i}(rfq1(i)+1)-xout1{i}(rfq1(i)))*3;
36 end
37
38 end %function

```

```

1 function [filtelec]=plankb2(p,y,th1,x,r)
2
3 % Purpose: Compare the baseline corrected std dev to the threshold
   value.
4 % Score the pixel with a 0 for below th1 and 1 for above th1. Look 10
5 % pixels in from each image edge to find the beginning (left edge)
   and end
6 % (right edge) of an illumination event (phytoplankton streak). The
7 % boundaries are written to filtelec with the beginning in column 1
   and the
8 % end in column 2.
9 % p = size of image set (often 500)
10 % y = cell array, each cell containing a 3Xr array; row 1 = mean(
   image);
11 % row 2 = std(image); row 3 = baseline subt'd std(image)
12 % th1d = threshold of baseline corrected standard deviations
13 % x = cell array of raw frames
14 % r = number of rows in each image (e.g., 256)
15 %
16 % filtelec is a cell array, each cell either being empty if nothing
   is
17 % above threshold. Otherwise it contains a jX2 array, where j is the
18 % number of tracks detected, and column 1 contains start row index,
   column

```

```

19 % 2 contains end row index
20
21 % For use in plankton_field.m
22
23 filtelec=cell(1,p);
24 b=cell(1,p);
25 for i=1:p
26     b{i}=(sign(y{i}(3,:)-th1(i))+1)/2;
27     j=0;
28     for k=2:r-1
29         b{i}(k)=sum(b{i}(k-1:k+1))/3;
30     end
31     for k=1:r
32         b{i}(k)=(sign(b{i}(k)-0.8)+1)/2;
33     end
34 %     t(i)=sum(b{i});
35 %     if (t(i)<3) || (t(i)>75)
36 %         x{i}=[];
37 %     end
38 if (isempty(x{i}))
39     filtelec{i}=[];
40 elseif (isnumeric(x{i}))
41     for k=10:r-10
42         if (b{i}(k) == 0) && (b{i}(k+1) == 1)
43             j=j+1;
44             filtelec{i}(j,1:2)=0;
45             filtelec{i}(j,1)=k;
46             %filtelec{i}(j,2)=k+1;
47         end
48         if (b{i}(k) == 1) && (b{i}(k+1) == 0) && (j>0)
49             filtelec{i}(j,2)=k+1;
50         end
51         if (b{i}(:) == 0)

```

```

52         filtelec{i}=[];
53     end
54     %i
55     %filtelec{i}
56 end
57 end
58 %i
59 %filtelec{i}
60 end
61
62 end %function plankb

```

```

1 function [ deb_track, strk_hightbound, strk_widthbound, wid_x, deb_x,
           int_x, initial_wid_x ] = plank_deblur5( x, filtelec, F, E)
2 % Version 1.3.0
3 %UNTITLED2 Summary of this function goes here
4 %   The purpose of plank_deblure2 is to deblure all "tracks" in x.
   Each
5 %   image in x will go through the debluring minimization for every "
   track"
6 %   defined in filtelec. This is done by a for loop from 1 to 50,
7 %   adjusting the disk psf diameter by 0.2, and finding the diameter
   value
8 %   that gives the largest max column vector value.
9
10 %deb_x is a cell array with each cell containing a cell array. The
   cell
11 %array in each cell of deb_x is a deblured track. There are as many
   cells
12 %in each cell of deb_x as there are tracks in each image of x.
13

```

```

14 %   x: cell array containing the "blurry" images.
15 %   filtelec: cell array containing the column boundries of each
    track
16 %       contained in x.
17 %   F: fraction of a disk diameter that is increased for each
    iteration.
18 %   E: Largest disk diameter to use.
19 compstart=tic;
20 a=size(x,2);
21 deb_track=cell(1,a);
22 strk_hightbound=cell(1,a);
23 strk_widthbound=cell(1,a);
24 initial_wid_x=cell(1,a);
25 wid_x=cell(1,a);
26 deb_x=cell(1,a);
27 int_x=cell(1,a);
28 %strk_int_x=cell(1,a);
29 for i=1:size(x,2)
30     i
31     if isempty(filtelec{i})==0
32         b=size(filtelec{i},1); %Make sure to change this once it is
            working!!!
33         tstart=tic;
34         for j=1:b
35             jj=1;
36             j;
37             % Increase the size of filtelec by 2 rows on both sides.
                This
38             % is needed to allow 'edge' to have enough pixels on the
                sides
39             % of the streaks.
40             filtwid=ceil(abs(filtelec{i}(j,2)-filtelec{i}(j,1))*0.25)
                ;

```

```

41         if filtelec{i}(j,1)>1
42             filtelec{i}(j,1)=filtelec{i}(j,1)-filtwid;
43         end
44         if filtelec{i}(j,2)<size(x{i},2)-1
45             filtelec{i}(j,2)=filtelec{i}(j,2)+filtwid;
46         end
47         if filtelec{i}(j,1)<1
48             filtelec{i}(j,1)=1;
49         end
50         if filtelec{i}(j,2)>256
51             filtelec{i}(j,2)=256;
52         end
53         % Test to determine if the track is more than 2 pixels
54         % wide
55         % before deblurring. If the track is less than 3 pixels
56         % wide,
57         % the deblurring step will not work properly.
58         % BWo=edge(x{i}(:,filtelec{i}(j,1):filtelec{i}(j,2))
59         % , 'prewitt', [], 'both');
60         track=x{i}(:,filtelec{i}(j,1):filtelec{i}(j,2));
61
62         % The perimeter of the streaks in the track are found and
63         % stored in DBWo_filt, expressed as 1's surrounded by the
64         % background which is expressed as 0's.
65         [ DBWo_filt ] = plank_edge3( track );
66
67         % The average width of the streaks in the track are
68         % output at

```

```

68         % streako_width. temp_strk_widthbound and
           temp_strk_hightbound
69         % contain the row and column boundries of the streaks as
           well
70         % as their lenght and widths.
71         [ streako_width,temp_strko_widthbound,
           temp_strko_hightbound ] = streak_dimensions (
           DBWo_filt );
72
73         % if the width of the track is less than 3, then the
           debluring
74         % algorithms do not work.
75         if streako_width>2
76             jj=1;
77         else
78             jj=2;
79         end
80
81         % If the average width of the streaks in the track are
           less
82         % than 3 pixels, then the deblur_iteration function is
           called.
83         if jj==1
84             tstart=tic;
85             [ test_deb_optval ] = deblur_iteration( streako_width
               ,filtelec{i},x{i},j,F,E );
86             deblur_iteration_time=toc(tstart)
87             [opt_val,opt_d]=min(test_deb_optval,[],1);
88             diameter=test_deb_optval(opt_d(1,1),2);
89
90             psf=fspecial('disk',diameter);
91             % [tempdeb,fpsf]=deconvblind(x{i},psf)
               ; %deconvblind may be

```

```

92         %                used to implement the psf found in
           deblur_iteration as a
93         %                starting value and then runs it's
           own iteration
94         %                internally. But there is no check
           to determin if this is
95         %                a resonable psf like there is in the
           deblur_iteration
96         %                function.
97
98     tempdeb=deconvlucy(x{i},psf);
99     deb_track{i}{j}=tempdeb(:,filtelec{i}(j,1):filtelec{i}
       )(j,2));
100     deb_x{i}{j}=tempdeb;
101     tempdeb2=deb_track{i}{j};
102
103     % The perimeter of the deblured streaks in the track
       are found and
104     % stored in DBWf_filt, expressed as 1's surrounded by
       the
105     % background which is expressed as 0's.
106     [ DBWf_filt ] = plank_edge3( tempdeb2 );
107     imshow(DBWf_filt);
108     pause(0.5);
109     % The average width of the streaks in the track are
       output at
110     % streako_width. temp_strk_widthboundf and
       temp_strk_hightboundf
111     % contain the row and column boundries of the streaks
       as well
112     % as their lengths and widths.
113     [ streakf_width,temp_strk_widthboundf,
       temp_strk_hightboundf ] = streak_dimensions (

```



```

114         DBWf_filt );
115
116         % strk_mask is a mask where only the area of the
117         image where
118         % a streak is located has an intensity value and all
119         other
120         % elements are 0.
121         strk_mask=DBWf_filt.*track;
122
123         % Finds the blank gaps in the track where the blocked
124         out
125         % filter element is located and integrates the
126         streaks.
127         % The order of the rows in int is the order of the
128         streaks
129         % as they appear in the image from row 1 to 256, or
130         % inverted in datap from column 1 to column 256.
131         [ int ] = integrate_streaks( temp_strk_hightboundf,
132         strk_mask,DBWf_filt );
133
134         %g;
135         int_x{i}{j}=int;
136         wid_x{i}(1,j)=streakf_width;
137         wid_x{i}(2,j)=i;
138         initial_wid_x{i}(1,j)=streako_width;
139         initial_wid_x{i}(2,j)=i;
140         strk_hightbound{i}{j}=temp_strk_hightboundf;
141         strk_widthbound{i}{j}=temp_strk_widthboundf;
142     else
143         wid_x{i}{j}=streako_width;
144     end
145 end
146 end

```

```

140         time=toc(tstart);
141     end
142 end
143 complete=toc(compstart)
144 end

```

```

1 function [ DBW_filt ] = plank_edge3( track )
2 %Version 1.3 used for beads
3 %   The input array, track, contains a set of streaks in a track. The
   track
4 %   is defined by the filtelec. Here, the streaks are enhanced and
   their
5 %   exact location within the track array are found. Their location
   is
6 %   stored in DBWo_filt and marked by a 1, where the background area
   is
7 %   marked by a 0. v1.3 uses the region found by edge and then
8 %   interpolates the region and determines width by selecting the
   edges of
9 %   the summed row vector profile.
10
11 gray_track=mat2gray(track);
12 %imresize interpolates the image by 10 pixels. This allows the image
   to
13 %expand by 10 times. Later it is contracted by using the same
   imadjust
14 %function, but with 0.1, to bring the image back to its original
   size.
15
16 % gray_track_interp=imresize(gray_track,2,'box');
17 %nearest, bilinear, bicubic, box, lanczos2, lanczos3

```

```

18
19 %imadjust increases the contrast between streaks and the bkg.
20 gray_track_con=imadjust(gray_track);
21
22 %edge outputs a logical array of 1's and 0's, where 1's
23 %represent an element in the array where the derivative along
24 %one dimension specifies the edge of an object. BWo is the
25 %trace of the original image.
26 BW=edge(gray_track_con,'canny'); %sobel, prewitt, roberts, log,
    zerocross, canny
27 % DBWo is converted from logical to double.
28 DBW=double(BW);
29
30 % perimeter_draw is a function that fills in the perimeter of
31 % an object from the input where the gap in the perimeter is
32 % less than 3 elements.
33 [ DBW_perim ] = perimeter_draw( DBW );
34 % imfill fills in the interior of an object drawn from
35 % perimeter_draw.
36 DBW_fill=imfill(DBW_perim,'holes');
37 % object_locate separates the streaks from noise objects
38 % detected in edge.
39 [ DBW_filt ] = object_locate( DBW_fill );
40 % imshow(DBW_filt);
41 % pause(0.2);
42 % figure:imshow(gray_track_con);title('nearest-gray track con');
43 % figure:imshow(gray_track_interp);title('nearest-gray track interp')
    ;
44 % figure:imshow(BW);title('nearest-sobel-BW');
45 % figure:imshow(DBW_fill);title('bicubic-sobel-DBWfill');
46 end

```

```

1 function [test_deb_optval] = deblur_iteration2(filtelec,x,j)
2 %deblur_iteration: This function is used in plank_deblur6. It finds
   the
3 %optimum disk radius used for debluring an image of a phytoplankton
   track.
4
5 %   Inputs—
6 %   F: fraction of a disk diameter that is increased for each
   iteration.
7 %   E: Largest disk diameter to use.
8 %   strk_widthboundo: the width of the track in an image before any
9 %       deblur processing is performed.
10 %   filtelec: predefined edges of the track found in plank a & b.
11 %   x: 256x256 array of original image
12 %   j: track number in filtelec. This is retrieved from
   plank_deblur3.
13 %   strk_lengthboundt: the section of the image that is used to
   determine
14 %       the "original" streak length in plank_deblure3. This number
   is used
15 %       only if the debluring step finds the streak width to be less
   than 2
16 %       pixels. Usually, this occurs because the streak is "
   deblured" into
17 %       nothing and only one or two pixels are left.
18
19 %   Outputs—
20 %   test_deb_optval: Is an array where each row is the result of one
   step
21 %       in the for loop. Column 1 contains the width of the deblured
   streak.
22 %       Column 2 contains the diameter of the psf used in 'deconvlucy

```

```

    ' to
23 %         deblur the image.
24 E=2;
25 F=10;
26 test_deb_optval(1:(E-F),1:2)=zeros;
27 EE=E*F;
28 Fd=F-1;
29 for d =F:EE
30     dd=d/F; %disk psf diameter is dd
31
32     % deblurring function for the entire image i
33     psf=fspecial('disk',dd);
34     tempdeb=deconvlucy(x,psf);
35
36     % put row boundries in for either a single streak or all
37     % streaks in a track, if this is know.
38     test_deb2=tempdeb(:,filtelec(j,1):filtelec(j,2));
39
40 %     pcolor(tempdeb(:,40:75));shading interp
41 %     pause(0.5);
42 %     d;
43
44     [r,p,x_t,y,th1]=planka_3(test_deb2);
45
46     [ streakt_width,temp_strk_widthbound,temp_strk_hightbound ] =
        streak_dimensions (DBWt_filt );
47
48     if streakt_width<2
49         streakt_width=strk_width;
50     end
51
52     test_deb_optval(d-Fd,1)=streakt_width;
53     test_deb_optval(d-Fd,2)=dd;

```

```

54
55 end
56
57 end

```

```

1 function [ streak_width,temp_strk_widthbound,temp_strk_hightbound ] =
    streak_dimensions2 (DBW_filt,track )
2 %UNTITLED6 Summary of this function goes here
3 % Detailed explanation goes here
4
5     % The row dimension of DBWo_filt is summed so that the
        each
6     % continouse row that contains a number defines the
        length of
7     % the streak.
8     strk_bound=sum(DBW_filt,2);
9
10    % temp_strk_hightbound is a variable that contains the
        hight or
11    % length of a streak. Rows represent the streak number
        in the
12    % track, column 1 is the start of the streak, 2 is the
        end and
13    % 3 is the length.
14    [ temp_strk_hightbound ] = streak_height_boundries(
        strk_bound );
15
16    % temp_strk_widthbound is a variable tghat contains the
        width
17    % of a streak. Rows represent the streak number in the
18    % track, column 1 is the start of the streak, 2 is the

```

```

        end and
19         % 3 is the width.
20         [ temp_strk_widthbound] = strk_width_boundries2(
            temp_strk_hightbound,DBW_filt,track );
21
22         streak_width=mean(temp_strk_widthbound(:,3));
23 end

```

```

1 function [ int,tot_int ] = integrate_streaks2( strk_heightbound,track
    ,track_outline,strk_widthbound,wn,strk_area )
2 %UNTITLED7 Summary of this function goes here
3 % This function uses the inputs from strk_dimensions and plank_edge
    and
4 % sums the pixel values found in each streak in a track.
5
6 %Inputs:
7 % wn=number of MOE positions in the filter wheel.  example: 7
    position
8 % filter wheel will have wn=6.
9 % temp_strk_heightbound = each row is a streak in a track, col 1 is
10 % beginning row value and col 2 is ending row value.
11 % strk_mask = track area image containing only the pixel values for
    each streak in a
12 % track.
13 % DBWf_filt = logical array of track image where 1's are pixels in
    the
14 % image containing a streak and 0's are pixels containing background.
15
16 %Outputs:
17 % int = array with integrated streak values and ordered from 1:end in
18 % increasing row values from the "blank" streak in the track.

```

```

19 % tot_int = cell array the same size as int where each cell is
    another cell array.
20     %The first row of the inner cell array contains the integrated
        intensity
21     %for each streak in the track and it's filter wheel position.
        The
22     %second and third rows in the inner cell are the image and track
23     %numbers.
24 wn=wn+1;
25 int=cell(1,size(track,2));
26 tot_int=cell(1,size(track,2));
27 for i=1:size(track,2)
28     for j=1:size(track{1,i},2)
29         gs=size(strk_heightbound{i}{j},1);
30         gap=zeros(gs+1,1); %make an array with the number of streaks
            plus 1.
31         % The for-loop fills in "gap" with each row representing
            the distance between streaks in the track.
32         % the first number will be between row 1 (beginning of
            image) to the first streak.
33         for g=1:gs+1
34             if g==1
35                 gap(g,1)=strk_heightbound{i}{j}(g,1);
36             elseif g==size(strk_heightbound{i}{j},1)+1
37                 gap(g,1)=size(track,1)-strk_heightbound{i}{j}(g
                    -1,2);
38             else
39                 gap(g,1)=abs(strk_heightbound{i}{j}(g-1,2)-
                    strk_heightbound{i}{j}(g,1));
40             end
41         end
42         %Below the average length of a streak in the track is
            found.

```



```

43         %From this average length, a gap threshold is found.
44         %This is used for determining where a "blank" is located
           in the
45         %track, variable 'sg_blank'.
46         streak_gap=gap(1:gs,:);
47         sg_blank=ceil((mean(strk_heightbound{i}{j}(:,3)))*1.5);
48         k=1;
49         temp_int=zeros(gs,1);
50         %Below populated the "integrated" intensities held in
           column 2
51         %of temp_int. If the length of the gap between streaks
           is
52         %greater than the previously determined thershold, then a
           0 is
53         %used in col 2 to identify it as the blank position.
           This
54         %blank will later be used to order the streaks.
55         %streak_outline=track{i}{j}.*track_outline{i}{j};
56         for q=1:gs
57             if q==1
58                 temp_int(k,1)=sum(sum(track{i}{j}(
                           strk_heightbound{i}{j}(q,1):strk_heightbound{
                           i}{j}(q,2),strk_widthbound{i}{j}(q,1):
                           strk_widthbound{i}{j}(q,2))),2);
59                 temp_int(k,4)=strk_area{1,i}{1,j}(1,q);
60                 %temp_int(k,1)=sum(sum(streak_outline(
                           strk_heightbound{i}{j}(q,1):strk_heightbound{
                           i}{j}(q,2),:)));
61                 k=k+1;
62             elseif streak_gap(q,1)< sg_blank
63                 temp_int(k,1)=sum(sum(track{i}{j}(
                           strk_heightbound{i}{j}(q,1):strk_heightbound{
                           i}{j}(q,2),strk_widthbound{i}{j}(q,1):

```

```

        strk_widthbound{i}{j}(q,2)),2);
64     temp_int(k,4)=strk_area{1,i}{1,j}(1,q);
65     %temp_int(k,1)=sum(sum(streak_outline(
        strk_heightbound{i}{j}(q,1):strk_heightbound{
        i}{j}(q,2),:)));
66     k=k+1;
67     else
68     temp_int(k,1)=0;
69     k=k+1;
70     temp_int(k,1)=sum(sum(track{i}{j}(
        strk_heightbound{i}{j}(q,1):strk_heightbound{
        i}{j}(q,2),strk_widthbound{i}{j}(q,1):
        strk_widthbound{i}{j}(q,2))),2);
71     temp_int(k,4)=strk_area{1,i}{1,j}(1,q);
72     %temp_int(k,1)=sum(sum(streak_outline(
        strk_heightbound{i}{j}(q,1):strk_heightbound{
        i}{j}(q,2),:)));
73     k=k+1;
74     end
75 end
76
77 %The third column in temp_int with have a 101 used to
    identify
78 %the blank spot, 201 is placed in all other cells in col
    3.
79 temp_int(:,3)=201;
80 temp_int(:,2)=(1:1:size(temp_int,1));
81 gs=size(temp_int,1);
82 % Look to see if there is a 0 in column 2 and change the
    row to
83 % 101 from 201, indicating it is a blank.
84 %pq is used to make sure only the first blank is labeled
    101.

```

```

85         %If there are multiple blank spots with 101, the for loop
           at 79
86         %renumbers all streaks below the second blank spot.
87         pq=1;
88         for q=1:gs
89             if temp_int(q,1)==0 && pq==1
90                 temp_int(q,3)=101;
91                 pq=2;
92             end
93         end
94         l=1;
95         for q=1:gs
96             % we only want to identify the first blank spot.
97             if temp_int(q,3)==101
98                 temp_int(q,2)=0;
99                 for m=q-1:-1:1
100                     temp_int(m,2)=wn-l; %ordering the streaks
                                           from p:1 in decending order of row pixel
                                           values. example: p=7, row 1 in
                                           height_bound
101                     l=l+1;             %corresponds to the streak from
                                           row pixels 6:35 and so on. If the blank
                                           is after say the 3 streak in the image,
                                           then streak
102                     end                %2 is labeled filter 6 and
                                           streak 1 is labled filter 5 and streak 3 is
                                           labeled streak 1 and so on.
103                     l=1;
104                     for m=q+1:size(temp_int)
105                         temp_int(m,2)=1;
106                         l=l+1;
107                     if l==wn+1           %This if statment and
                                           for loop are for when a scond blank spot

```

```

                                is found in the track.  The ordering
108                                temp_int(m,2)=0;
109                                l=1;
110                                for n=m+1:size(temp_int)
111                                %                                temp_int(n,2)=c;
112                                %                                c=c+1;
113                                %                                end
114                                end
115                                end
116                                end
117                                end
118                                temp_int(:,3)=[];
119                                % reorder int so that zero is first and the order goes
                                from 1:p.
120                                temp_int2=zeros(wn-1,2);
121                                temp_avg=zeros(1,1);
122                                for jj=1:wn-1 %jj is filt position
123                                kj=1; %kj is the number of filt positions found
124                                for ii=1:size(temp_int,1) %ii is the row number of
                                temp_int
125                                if temp_int(ii,2)==jj
126                                temp_avg(kj,1)=temp_int(ii,1);
127                                kj=kj+1;
128                                end
129                                end
130                                temp_int2(jj,1)=jj;
131                                temp_int2(jj,2)=mean(temp_avg);
132                                temp_avg=zeros(1,1);
133                                end
134                                int{1,i}(1:wn-1,j)=temp_int2(:,2);
135                                int{1,i}(wn,j)=i;
136                                int{1,i}(wn+1,j)=j;
137                                tot_int{1,i}{1,j}(:,1)=temp_int(:,1); % integrated

```

```

        intensity
138         tot_int{1,i}{1,j}(:,2)=temp_int(:,2); % MOE number in
            filter wheel
139         tot_int{1,i}{1,j}(:,3)=temp_int(:,3); % area of the
            streak in pixels
140         for mm=1:size(temp_int)
141             tot_int{1,i}{1,j}(mm,4)=sqrt((86^2.*temp_int(mm,3))
                +(45.2.*temp_int(mm,1))); % stdev in the streak
                due to background fluorescence.
142             tot_int{1,i}{1,j}(mm,5)=temp_int(mm,1)/tot_int{1,i
                }{1,j}(mm,4); % upper limit on SNR due to
                background fluorescence
143         end
144         tot_int{1,i}{2,j}=i;
145         tot_int{1,i}{3,j}=j;
146     end
147 end
148 end

```

```

1 function [strk_hightbound] = streak_height_boundries( strk_bound )
2 %   Find the row boundries of each streak in a track and remove any
   streaks
3 %   that are less than the average streak length.
4
5 p=1; % keeps track of what peak number is being counted
6 a=1; % is used to switch between looking for start and end of a
   streak
7 k=0; % keeps count of number of rows between streaks
8 m=0; % keeps count of number of rows in a streak
9 j=1; % keeps track of what row is being investigated
10 q=1; % keeps track of what row in pend was filled last

```

```

11 z=1; % remains 1 until the last row is investigated
12
13 %Look down the entire column, checking each row value to determine
14 %where a streak starts in the strk_bound array. Once a non-zero
    value is
15 %found, the row number of the first non-zero number is saved in
16 %pbegintemp and the total number of rows before the start of the peak
    is
17 %saved in the temp0 array.
18
19 % If the first streak in an image is a partial streak, then it causes
20 % problems in strk_width_boundries function becuae the width of a
    partial
21 % streak is found and can be inconsistant with the widths of full
    streaks.
22
23 % looks at the first row to see if there is a partial streak. If
    there is
24 % a partial streak, it removes it by changing all consecutive 1's to
    0's.
25
26 if strk_bound(1,1)==1 || strk_bound(2,1)==1
27     strk_bound(1,1)=0;
28     jj=2;
29     pp=1;
30     while pp==1
31         if strk_bound(jj,1)==1
32             strk_bound(jj,1)=0;
33             jj=jj+1;
34         else
35             pp=2;
36         end
37     end

```

```

38         clear jj;
39         clear pp;
40     end
41     clear bb;
42
43     % looks at the last row to see if there is a partial streak.  If
44     % there is
45     % a partial streak, it removes it by changing all consecutive 1's to
46     % 0's.
47
48     if strk_bound(size(strk_bound,1),1)==1 || strk_bound(size(strk_bound
49         ,1)-1,1)==1
50         strk_bound(size(strk_bound),1)=0;
51         jj=size(strk_bound,1)-1;
52         pp=2;
53         while pp==2
54             if strk_bound(jj,1)==1
55                 strk_bound(jj,1)=0;
56                 jj=jj-1;
57             else
58                 pp=3;
59             end
60         end
61         clear jj;
62         clear pp;
63     end
64
65     %Each row in temp0, temp1, pend, and pbegin correlates to the peak
66     %number in strk_bound.  For example, the value in row 1 in pbegin
67     %is the row number in strk_bound where the first peak starts, row 1
68     %in
69
70     %pend is the row number in strk_bound is where the first peak ends,
71     %row 1

```

```

66 %in temp0 is the number of rows in strk_bound before the first peak
    starts
67 %and row 1 in temp1 is the number of rows in strk_bound that make up
    the
68 %first peak.
69
70 while z==1
71     while a==1
72         if j>size(strk_bound,1)
73             a=0;
74             z=0;
75         elseif strk_bound(j,1)==0
76             k=k+1;
77             j=j+1;
78             a=1;
79             if j>size(strk_bound,1)
80                 a=2;
81                 j=size(strk_bound,1);
82             end
83         else
84             pbegintemp(p,1)=j;
85             a=2;
86             temp0(p,1)=k;
87             k=0;
88             p=p+1;
89         end
90     end
91     %After a peak is determined to have started, the row is followed
        down
92     %each column until a zero is found again. Then the last column
        of a
93     %zero is saved in pendtemp and the total number of columns in the
        peak is

```



```

94     %saved in temp1.
95     while a==2
96         if strk_bound(j,1)>0.001
97             m=m+1;
98             j=j+1;
99             a=2;
100         if j>size(strk_bound,1)
101             a=1;
102             pendtemp(q,1)=j-1;;
103             temp1(q,1)=m;
104             m=0;
105             q=q+1;
106         end
107     else
108         pendtemp(q,1)=j-1;;
109         temp1(q,1)=m;
110         m=0;
111         q=q+1;
112         a=1;
113     end
114 end
115 end
116
117 hight=pendtemp-pbegintemp;
118 avlength=mean(hight);
119 for i=size(hight):-1:1
120     if hight(i,1)<avlength*0.6
121         hight(i,:)=[];
122         pbegintemp(i,:)=[];
123         pendtemp(i,:)=[];
124     end
125 end
126

```

```

127 strk_hightbound(:,1)=pbeginntemp;
128 strk_hightbound(:,2)=pendntemp;
129 strk_hightbound(:,3)=hight;
130 end

```

```

1 function [ strk_widthbound ] = strk_width_boundries2(
    strk_lengthboundt,DBW,track )
2 %strk_width_boundries: After strk_high_boundries have found the row
3 %boundries(streak hights), we want to take the DBW variable and sum
    along the columns to find
4 %the column number for each row that has the greatest value. The two
5 %columns numbers with the highest value will be the edges of streak.
6
7 %used for the 'canny' method, which produces the outline of an image
    based
8 %on two thresholds, lower and upper. This function eliminates the
    value
9 %for "inner" object found with the lower threshold, if one is found.
10 a=size(strk_lengthboundt,1);
11 strk_widthbound=zeros(a,3);
12 for i=1:a
13     temp=sum(track(strk_lengthboundt(i,1):strk_lengthboundt(i,2),:),
        ,1);
14     [r,c]=size(temp);
15     c2=c*10;
16     temp2=interp1(1:1:c,temp,1:1:c); %interpolate the width of a
        streak to 10 pixels. This allow for higher resolution of
        streak widths beyond 1 pixel(2um).
17     [r2,c2]=size(temp2);
18     [outy,sdbase,thold]=autobase_track_beads2(1:c2,temp2,1);
19     streak_wid=outy;

```

```

20     streak_wid(1,end+1)=thold;
21     [br] = peakthreshold3(streak_wid);
22     % find start of streak.
23
24
25
26     j=1;
27     jj=1;
28     while j==1
29         if jj>=size(br,2)
30             j=2;
31         elseif br(1,jj)==1
32             strk_widthbound(i,1)=jj;
33             j=2;
34         else
35             jj=jj+1;
36         end
37     end
38     jj=size(br,2);
39     while j==2
40         if jj==1
41             j=3;
42         elseif br(1,jj)==1
43             strk_widthbound(i,2)=jj;
44             j=3;
45         else
46             jj=jj-1;
47         end
48     end
49     if strk_widthbound(i,1)<10
50         strk_widthbound(i,1)=10;
51     end
52     strk_widthbound(i,3)=abs(strk_widthbound(i,2)-strk_widthbound(i

```

```

        ,1));
53     clear streak_wid;
54 end
55 strk_widthbound=strk_widthbound.*0.1; % divide by 10 to bring width
    back to pixels after the interpolation done above.
56 strk_widthbound(:,1)=floor(strk_widthbound(:,1));
57 strk_widthbound(:,2)=ceil(strk_widthbound(:,2)); %ceil and floor are
    used to give integer values for the width dimensions to be used
    for PLS width calibration.
58
59 end

```