Taxonomic Classification of Phytoplankton with Multivariate Optical Computing, Part III: Demonstration

Megan R. Pearl,[a] Joseph A. Swanstrom,[a] Laura S. Bruckman,[a] Tammi L. Richardson,[b] Timothy J. Shaw,[a] Heidi M. Sosik,[c] Michael L. Myrick[a,*]

[a] *Department of Chemistry and Biochemistry, University of South Carolina, Columbia, SC 29208 USA*
[b] *Department of Biological Sciences, University of South Carolina, Columbia, SC 29208 USA*
[c] *Department of Biology, Woods Hole Oceanographic Institution, Woods Hole, MA 02543 USA*

We describe the automatic analysis of fluorescence tracks of phytoplankton recorded with a fluorescence imaging photometer. The optical components and construction of the photometer were described in Part I and Part II of this series in this issue. An algorithm first isolates tracks corresponding to a single phytoplankter transit in the nominal focal plane of a flow cell. Then, the fluorescence streaks in the track that correspond to individual optical elements on the filter wheel are identified. The fluorescence intensity of each streak is integrated and used to calculate ratios. This approach was tested using 853 fluorescence measurements of the coccolithophore *Emiliania huxleyi* and the diatom *Thalassiosira pseudonana*. Average intensity ratios for the two classes closely follow those predicted in Part I of this series, with a distribution of ratios in each class that is consistent with the signal-to-noise ratio calculations in Part II for single cells. No overlap of the two class ratios was observed, yielding perfect classification.

Index Headings: **Phytoplankton; Fluorescence; Multivariate optical computing; Photometer; Classification.**

INTRODUCTION

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 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.[1] show that experts, who are routinely involved in classification, have accuracies in the range 84–95%. Unfortunately, this method lacks the high-frequency sampling capacity needed for community structure monitoring.[2,3]

Flow cytometry alone, although 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.[4] were the first to report automated classification of phytoplankton cells from video images, but no statistical summary was provided. Since then, advancements in morphological classification methods have combined microscopy with flow cytometry and have resulted in great progress.[1,5–8] Using the Video Plankton Recorder, Davis et al.[9] achieved accuracies between 45 and 91% in identifying individual taxa. Culverhouse et al.[10] developed the Harmful Algal Bloom Buoy for both zooplankton and phytoplankton identification and report identification rates of 80% for phytoplankton species. Sosik et al.[5] have developed perhaps the most innovative method using cytometry, fluorescence, and image analysis with the FlowCytobot, where 88% accuracy between 22 categories is reported.

CHEMTAX, a high-performance liquid chromatography method, has been used to identify the relative concentrations of taxonomic species in bulk monocultures and mixed cultures.[11] CHEMTAX is useful in determining pigment concentration in bulk samples for calibration or validation, but it is not suited to in situ measurements.[12]

Remote sensing methods such as satellite imagery of chlorophyll *a* and phycoerythrin also have been used for monitoring phytoplankton.[13,14] 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 they are limited or unable to discriminate the speciation of the source of the fluorescence emission.[15,16]

Despite the advancements described above, a rugged and deployable method suitable for open ocean monitoring is still desired.[9,17–21] In situ fluorescence excitation spectroscopy provides an alternative approach. Fluorescence excitation spectroscopy uses spectral characteristics of a phytoplankton cell that is independent of morphology.[22] Beutler et al.[23] developed an in situ method using light-emitting diodes to selectively excite a bulk sample at five wavelength bands and recorded the chlorophyll *a* emission for each, but they provided no statistical analysis. The potential of this instrument for bulk in situ fluorometric measurement of phytoplankton community structure has been described previously.[12]

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 report, we focus on the automatic analysis of spectroscopic content in images from a fluorescence imaging photometer.

In a previous report, we showed that full-spectrum fluorescence excitation spectroscopy could be used as the basis for distinguishing at least limited classes of phytoplankton in cultures.[24] In Part I[25] of this series, we showed that the full-spectrum information of single phytoplankton cells could be used to develop special optical elements, enabling rapid measurements based on fluorescence excitation spectroscopy.

Received 28 January 2013; accepted 1 February 2013.
* Author to whom correspondence should be sent. E-mail: myrick@mailbox.sc.edu.
DOI: 10.1366/12-06785

0003-7028/13/6706-0640/0

In Part II,[26] we described an instrument capable of supporting measurements using these special optical elements, called multivariate optical elements (MOEs). The output of this instrument consists of images of phytoplankton fluorescence tracks, with individual streaks whose intensity is related to fluorescence excitation through particular MOEs, forming a sort of "bar code" for an individual phytoplankter. Analysis of the data produced from phytoplankton using the fluorescence imaging photometer instrument presents a unique situation, since image analysis is required, but it serves as a proxy for spectral analysis. This report completes the series by describing the function and performance of the software developed for automatic interpretation of phytoplankton data from a fluorescence imaging photometer, Streak Integrator for Multivariate Optical Computing, version 1.0 (SIMOC).

Although the end goal for this instrument lies in field measurements and studies involving variability of laboratory cultures, the goal of this study is to evaluate the true performance of the instrument versus theoretical expectation, for which known monocultures of phytoplankton are the closest available samples to instrument standards. We validate the MOE/fluorescence imaging photometer concept by measuring experimental fluorescence ratios and comparing them to those calculated from the optical model given in Part I,[25] as well as the classification accuracy and ratio variability for each class using two similarly pigmented phytoplankton species: the coccolithophore *Emiliania huxleyi* and the diatom *Thalassiosira pseudonana*. We found that the measured MOE ratio for *E. huxleyi* differed from the theoretical MOE ratio 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. The distribution of individual cell ratios was well explained by the signal-to-noise ratio (S/N) of the instrument reported in Part II,[26] and no misclassifications were seen for 853 cells of two species analyzed.

EXPERIMENTAL

Data Collection. Details of the phytoplankton culture conditions are found in Part I,[25] and instrument details are found in Part II[26] of this series. The fluorescence imaging photometer consists of a filter wheel with six openings that rotates counterclockwise as viewed from the lamp. One of the six positions was blocked with a 2.54 cm (1 in.) diameter black painted substrate. Using this as a reference, the order in which filters intersected the light path was 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 three-species separation described in Part I[25] of this series, *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 S/N 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 an 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+, whereas 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 s were acquired. In total, 20 such files for each organism constituted the complete sample data set. Additional files characterizing the background, dark count, and flat field also were acquired.

Data Preprocessing. All algorithms were written in the MatLab® 7.7 (Mathworks, Inc., Natick, MA) programming environment on an Apple iMac computer running OS X, version 10.7.7. Preprocessing was performed as follows. Data sets of 16-bit binary image files were read using a MatLab routine. The first sets of data imported included three 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 of Part II.[26]

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 frame was then processed according to the formula

$$s_c = \frac{s_i - \langle b \rangle}{\langle f \rangle} \quad (1)$$

where s_c is a single corrected image, $\langle f \rangle$ is the normalized corrected flat field, s_i is a single uncorrected image, and $\langle b \rangle$ is the average sample background frame. Eq. 1 was applied pixel by pixel for each image.

Data Analysis Algorithm. In the following section, tracks of fluorescence streaks from a phytoplankter are assumed to be largely parallel to the column direction of the charge-coupled device array. After the preprocessing described above, each corrected 16-bit sample image is initially evaluated for the presence of measurable tracks that show good modulation by the spinning filter wheel. In Fig. 1a, for example, although several tracks are visible, some are clear and sharp and others are blurred and indistinct.

The average and standard deviation of fluorescence intensities in each column of the image is first calculated. Then, a baseline-corrected standard deviation is calculated with the algorithm described in Part II.[26] Figure 1b shows the column standard deviations plotted along with the baseline-corrected standard deviations. Next, the ratio of the baseline-corrected standard deviation to the average intensity along the central pixel column of each region is calculated.

Potential tracks are then identified. First, tracks centered in columns 1–10 and 247–256 of the 256 column image are ignored because optical aberrations can cause these tracks to curve outside the field of view of the camera at points. Then, a threshold value in the array of baseline-corrected column standard deviations is established based on a histogram of the array.

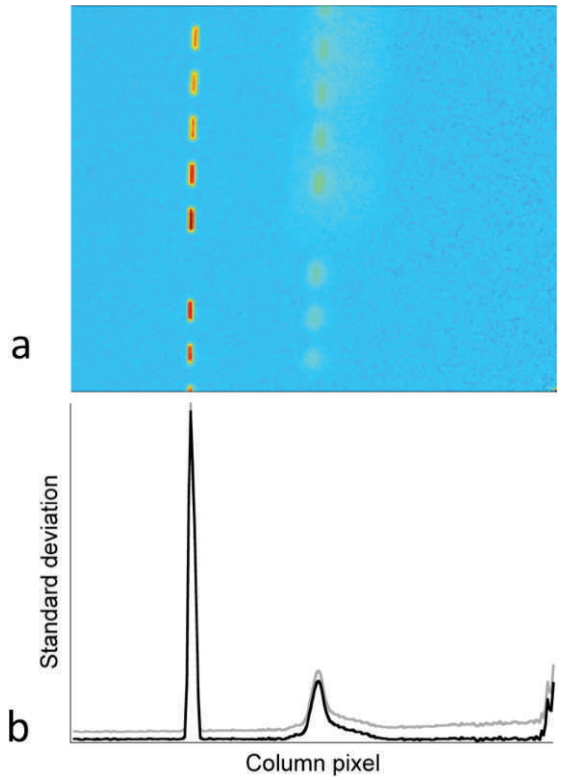

FIG. 1. (**a**) Preprocessed data image. There are two tracks of the coccolithophore *Emiliania huxleyi* visible in this image; however, only the track near column 64 is well modulated and therefore considered "good." (**b**) Plot of the standard deviation and corrected standard deviation along the rows for each column of the image in panel a. The gray dashed line corresponds to the uncorrected column standard deviation, and the solid black line corresponds to the baseline-corrected column standard deviation.

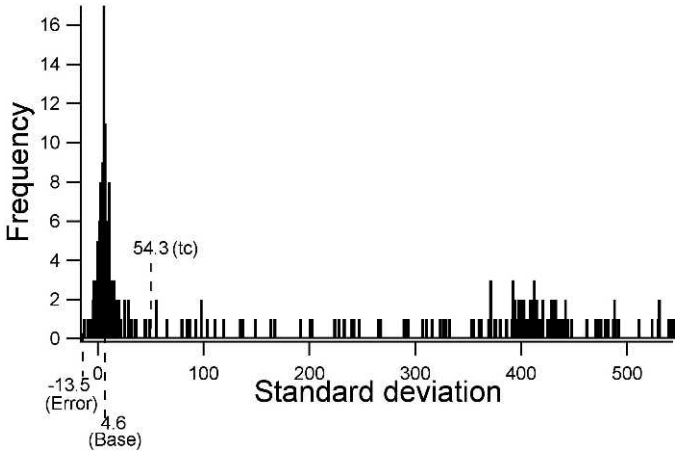

FIG. 2. Histogram of the column standard deviations shows the highest frequency standard deviation to be 4.6, a value that corresponds to the baseline value. The lowest value measured (error = −13.5) is considered 1 standard deviation from the baseline. The threshold for the image is 3*(base-error) or 54.3, labeled as "tc".

The most common value in the histogram is taken to be the best estimate of the corrected baseline (nominally at zero). The most negative value in the histogram is taken as the maximum deviation expected from the baseline that is not related to the signal (i.e., from noise and baseline correction errors alone). Three times the absolute value of the difference between the most negative and most common values, added to the most common value, was found to serve well as a threshold for identifying potential tracks. This process is illustrated in Fig. 2. From the histogram, the most common value in the baseline-corrected standard deviation array was 4.6. This value occurred 17 times. The lowest measured value was −13.5. The absolute value of the difference between these values was 18.1. Three times this value, added to the most common value, gave a threshold for identifying potential tracks of 54.3 (Fig. 2).

As a first test, each column value of the standard deviation array was scored against the threshold, 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, whereas the right edge is indicated by the change from 1 to 0. Each contiguous set of 1's identifies columns containing a potential track. This identifies all potentially usable tracks.

Tracks that are a few micrometers in diameter near the plane of focus of the imaging system show discrete streaks that are fully modulated (returning to baseline or near baseline between different filter wheel elements). In profile, these tracks approximate a square-wave pattern in which the base is near 0 and the top is at some finite positive value. For a perfect square wave with its base at 0, the ratio of the standard deviation to the average fluorescence intensity is exactly unity. In practice, it was found that a ratio of standard deviation to average intensity above 0.6 discriminates well between streaks that would be visually judged "good" versus those that appear indistinct or overlapped to the eye.

This ratio threshold is applied to each track identified by the first test. The column at the center of each potential phytoplankton track is identified, and the ratio of the standard deviation for that column to the average intensity in that column is calculated. Potential tracks with ratios above 0.6 are retained, whereas those with ratios below 0.6 are deleted, forming a second test. Take, for example, Fig. 1b. Although there are two phytoplankton tracks present, only one track has a ratio above 0.6. The track located near column 64 would be retained for further analysis, and the other track near column 125 would be ignored.

Information about phytoplankton tracks in each image is compiled in the corresponding cell of a cell array in MATLAB. One such array, t_e, is created to hold the left/beginning and right/ending boundaries, c_b and c_e respectively, of each track. This array has two columns, with c_b in the first column and c_e in the second column. The number of rows in the array is equal to the number of tracks identified in a given image frame.

For each track identified in an image, the rows within the track are summed to form a first representation of the track. These are placed in a single column of a separate cell array. The array in a given cell has a size determined by the number of rows in the image (256 in all cases reported here) and the number of tracks passing the first two tests. This array contains the average intensities of the width of each streak and is automatically baseline corrected. A row threshold is determined for the 256 row sums of each track using the histogram method described above to help locate the individual streaks in each track.

The pixel intensity in each row i of this track array (Fig. 3, solid black line) is compared to its row threshold (solid gray line) using the same approach described above to assign scores

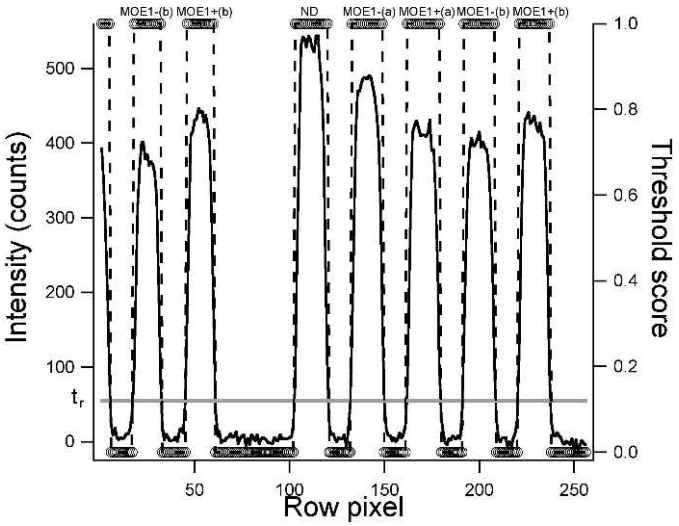

FIG. 3. Example analysis of fluorescence intensities of streaks with in a phytoplankter cell. The solid black line is a plot of the sums of the column intensities along each row in the image. The threshold determined by the histogram method for this streak is $t_r = 54.8$ counts (solid gray line). Rows with pixel values above t_r are scored with a 1, whereas pixel values below t_r are scored with a 0. The rows that contain 1 in the score array (as indicated by the dashed line and open circles) are row numbers that are identified to contain streaks. The names of the filters corresponding to each streak are listed above for a reference, the labels a and b indicate replicates of the same MOE located in two positions in the filter wheel. Rows 1–8 contain a portion of a streak, but because there is no transition in the score array from a 0 to a 1, the streak is not used.

to each row. Working from bottom to top in the image, a change from 0 to 1 indicates the bottom edge, r_b, of a fluorescence streak, whereas the change from 1 to 0 indicates the top edge, r_e, of the same streak. Figure 3 shows the fluorescence intensity of the streaks in the track (dashed rectangle regions).

The length of each streak produced by a single MOE is determined by the filter wheel rotation frequency and the pump speed and is therefore nominally the same for all streaks in all image frames. For this study, streaks occupy at least 8 pixels. This length does vary slightly due to the position of a phytoplankter relative to the level of focus in the image. Lengths less than 8 pixels, however, represent partial filter elements, extremes of noise, or serious baseline errors. For the purpose of this analysis, only complete filter elements and full filter wheel rotations have been used. Therefore, if $r_e - r_b$ is less than 8 or the total number of filter elements detected is less than 5 (the number of filters in the filter wheel), all data referring to a given track is deleted from the corresponding cell arrays, representing a third test used to reject potential tracks.

Another cell array, f_e, is created to contain the vertical (row) boundaries of each streak in a given track. This array, for each track, contains two columns of numbers, the first column holding r_b for each streak and the second column holding r_e. The number of rows in the f_e cell arrays corresponds to the number of filter elements detected in tracks passing the first three tests, but is not less than 5.

We then refine the horizontal (column) boundaries of each streak in each track. These boundaries were originally determined by testing all columns above a threshold, as described above. Because of track slope and curvature, the horizontal boundaries of the individual streaks can be refined as follows.

A cell array is created, hb, with one data cell per image in the data set. Inside each cell is placed another cell array with one data cell per track. Inside the cells of the track cell array, an array is placed in which each column contains a cross section for each streak in the track. These cross sections are sums of the rows between the vertical boundaries of the streak (r_b and r_e, in the fe array), computed from 15 columns to the left of the horizontal boundary c_b to 15 columns to the right of the boundary c_e. The extensions to the sides of the original track boundaries were used so more of the baseline could be considered: baseline correction for each streak cross section was the first step in refinement.

Two new thresholds are then calculated based on the individual streak cross section. The lower threshold, t_l, is set at 2.5 times the standard deviation of the baseline residuals in the region outside the streak. This approach is possible because of the generally flatter baseline for single streaks, and it includes more of the actual streak intensity than the histogram method, the latter of which performs better when the baseline is not very flat. An upper threshold, t_u, was then selected to be the average of the two highest histogram levels with multiple observations. This unusual construction for the upper threshold gives a level that was near the top of the highest peak in the cross section. It is used to specify the dominant fluorescence peak in a cross section, a specification that generally does not matter, but becomes important when a neighboring track starts to bleed fluorescence into the edges of the track being quantified.

The cross section of each filter element is first compared to t_u. Rows in hb that contain pixel intensities above t_u are scored with a 1 in a corresponding score array, whose size is equal to hb. The cross section is then compared to t_l and rows in hb that contain intensities above t_l have their entries in the corresponding score array incremented by 1. The score array now contains values of 0 for pixels below both t_u and t_l, 1 for pixels between t_l and t_u, and 2 for pixels above t_u.

Because the original horizontal boundaries chosen were extended by 15 pixels in each direction, some cases arise where the cross section impinges on a streak from a nearby track. In these cases, 2's in the score array indicate the dominant peak. All rows holding a score of 1 and that are immediately preceding or following rows containing a value of 2 are then incremented to 2, whereas those that are not are decremented to 0. The score array for a streak cross section is then divided by two to change the 2's to 1's. Transitions from 0 to 1 and 1 to 0 then precisely define the horizontal boundaries of the proper filter element streaks. Because hb contains the sums within the row boundaries, the scalar product of hb with the score array yields a single value representing the integrated intensity of each filter element streak. The cross section of a single streak and score results are shown in Fig. 4. This streak has an integrated intensity of 6.207×10^4 counts.

A new cell array, rec, is used as a record of these integrated intensities for each streak in each track. Intensities are not entered in this array in the order they occur in the track. Instead, the distances between streaks are used to identify the blocked aperture in the filter wheel. This position represents row 1 in the intensity array, where a defined 0 is entered. Streak identities are defined by their positions relative to a blocked position in the filter wheel rotation in a track. Five non-zero values are entered for the integrated intensities of the five open

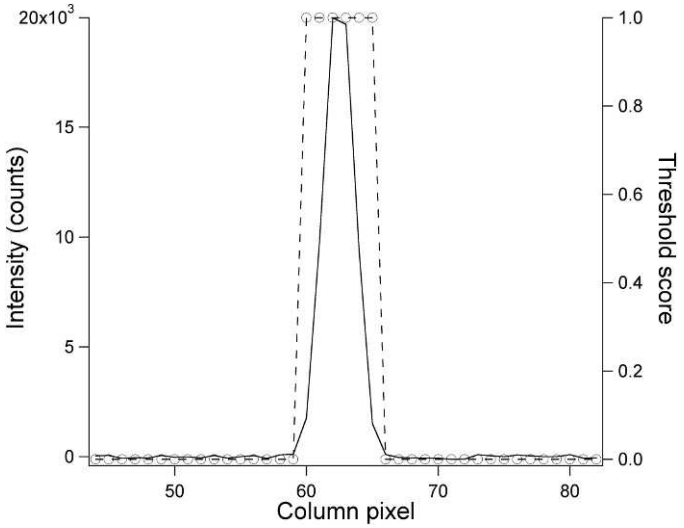

FIG. 4. The solid black line is the column profile of a single fluorescence streak corresponding to MOE1+. Circles indicate the threshold score, and the dashed lines are added as a guide. The product of the threshold scores and intensity values for the streak is 6.207×10^4 counts.

TABLE I. Example showing calculation of the experimental ratio for a single cell of the coccolithophore *Emiliania huxleyi* from a single track that is shown in Fig. 1.

MOE[a]	Integrated intensity ($\times 10^4$)[b]	Average intensity ($\times 10^4$)[c]	Ratio[d]
$1+^{i}$	5.549	5.370	0.9112
$1+^{ii}$	5.190		
$1-^{i}$	6.207	5.893	
$1-^{ii}$	5.578		

[a] Column 1 shows the identity of the MOEs, with (i) indicating the first measurement of a given MOE type, and (ii) indicating measurement of the second MOE of the given type. Two copies each of the two MOE types (1+ and 1−) mimicking the first linear discriminant function described in Part I were sampled in each rotation of the filter wheel. + and − symbols indicate the sign of the relationship expected between the (MOE/ND) ratio and the score of a single phytoplankton fluorescence excitation spectrum on the first linear discriminant function. ND in this case stands for a neutral density filter (i.e., a filter with a flat spectral function). The streaks associated with each MOE were identified and integrated.
[b] Column 2 shows integrated intensity for the streaks associated with the MOEs in column 1.
[c] Column 3 shows the averaged intensities for the repeat measurements, and
[d] Column 4 shows the ratio of 1+ to 1−.

positions in the filter wheel, and they are assigned to rows 2–6 in order of their positions around the filter wheel. The cell array is thus converted into a conventional array for ease of analysis.

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. In the example image shown in Fig. 1a, the integrated intensities for each streak (as in Fig. 3) are shown in Table I.

RESULTS AND DISCUSSION

Species Classification. Images containing fluorescence streaks of *E. huxleyi* and *T. pseudonana* were recorded and analyzed using the approach described above. This automated analysis resulted in 565 *E. huxleyi* ratios and 382 *T. pseudonana* ratios. The mean ratio of fluorescence intensities for excitation through (MOE1+/MOE1–) for *E. huxleyi* was 0.9081 ± 0.0072, whereas *T. pseudonana* had a mean ratio of 1.1788 ± 0.0040 (Table II). Figure 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.

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 whether the calculated ratios were accurately determined by the program. There were 19 *T. pseudonana* and 28 *E. huxleyi* ratios tested as outliers, and 15 *T. pseudonana* and 27 *E. huxleyi* tracks showed obvious defects that were visible in tour opinion but passed the tests by the program. The ratios resulting from the 42 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, resulting in a fraction of the integrated intensity of the streak; a phytoplankter track that flowed in and out of the plane of focus caused the later streaks in the track to be in a different state of focus than the earlier streaks in the track, resulting 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 II shows the number of ratios, calculated means, standard deviation of the means, and standard deviations for the distributions after the outliers were removed.

The inset of Fig. 5 shows the ratios of *T. pseudonana* and *E. huxleyi*, with the defective ratios removed. The separation of mean ratios for the two classes was 0.278. The sample standard deviations for the *E. huxleyi* and *T. pseudonana* ratio distributions shown in Table II were 10 and 12% of the difference between the class mean ratios, respectively.

The program in its current version miscalculates ratios for approximately 4.4% of tracks for a variety of minor thresholding and timing factors, but once these errors were manually identified and removed, there were no remaining misclassifications observed in these data. It still remains to automate the process for removing residual bad tracks or to improve the algorithm to correctly calculate the ratios for marginal tracks. If the "bad" tracks are not removed, the rate of correct

TABLE II. Experimental results for ratios of the coccolithophore *Emiliania huxleyi* and the diatom *Thalassiosira pseudonana* recorded on a fluorescence imaging photometer.

Phytoplankton species	N[a]	$\bar{X}$[b]	$S_{\bar{X}}$[c]	S_X[d]
E. huxleyi	538	0.900	0.001	0.029
T. pseudonana	367	1.179	0.002	0.036

[a] N is the number of phytoplankton tracks analyzed for each distribution.
[b] $\bar{X}$ is the mean ratio.
[c] $S\bar{X}$ is the standard deviation of the mean.
[d] SX is the sample standard deviation for each distribution.

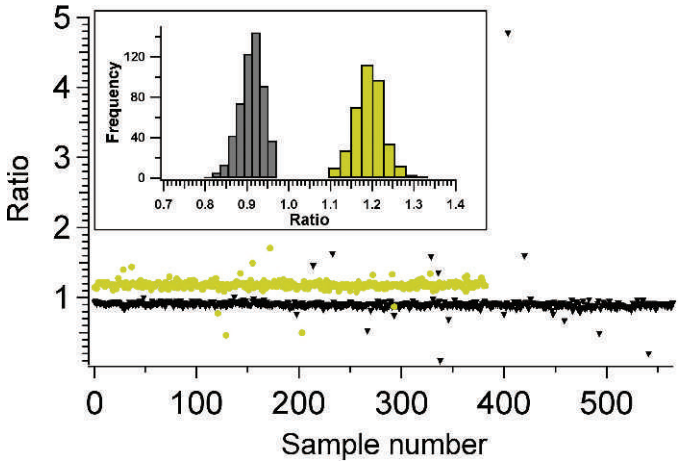

FIG. 5. Distribution of measured ratios for the coccolithophore *Emiliania huxleyi* (black, 565 ratios) and the diatom *Thalassiosira pseudonana* (yellow, 382 ratios). The inset is the frequency distribution of the measured ratios where 47 outlier ratios were visually inspected and 42 were removed due to defects in the image that were missed by the algorithm.

classification is reduced to about 98%, when only two classes are considered. Thus, there is still room for improvement in SIMOC.

Comparisons to Theory. The performance of the fluorescence imaging photometer with the MOEs can be compared to the modeled performance of single-cell fluorescence excitation spectra in Part I.[25] To do this, we return to the calibration spectra originally used in the design of the optical elements and shown in Fig. 1 of Part I.[25] The transmission spectrum of a particular MOE can be used to estimate the fluorescence intensity expected when a phytoplankter is represented by each of the original fluorescence excitation spectra and is excited by light passing through it using the equation

$$I_{MOE,\,j} = \sum_{\lambda=550}^{610} T_{sys}(\lambda)\, E_{plk,\,i}(\lambda)\, T_{MOE}(\lambda) \qquad (2)$$

In this Equation, λ is wavelength (nm). T_{sys} is the measured spectral radiance of the photometer system on the excitation side and is equal to the product of the profile of the excitation source and the transmission or reflection of the other optics between the lamp and sample except for the filter wheel elements. $E_{plk,i}$ represent the fluorescence excitation spectrum of the *i*th species of phytoplankton in the original calibration set. $T_{MOE,j}$ represents the transmission spectrum of *j*th MOE on the filter wheel. Since the optics of the system are chosen to isolate the region between 550 and 610 nm, the sum of all terms over this region is calculated. When this is repeated for a single phytoplankter spectrum using both the MOE1+ and MOE1− spectra, the ratio between these values can be computed for the original calibration phytoplankton spectra.

The theoretical ratio averages and sample standard deviations of single phytoplankter cells of *E. huxleyi* and *T. pseudonana* were found by this method to be 0.926 ± 0.012 and 1.177 ± 0.014, respectively. The experimental averages acquired with the fluorescence imaging photometer, shown in Table II, deviate from these theoretical values by −2.9% for *E. huxleyi* and +0.2% for *T. pseudonana*. The major difference between the theoretical and experimental result is that the experimental sample standard deviations are about three-fold higher than those in theory.

Minor differences in the mean ratios for each species can be attributed to a variety of sources. First, the theoretical response was calculated using the transmission spectrum of the actual MOE as measured through the center of the filter using a well-collimated beam. The filters we fabricate usually show minor spectral shifting across the substrate due to slight spatial variations in deposition rates. Also, the light passing through the optical elements in the photometer is not as well collimated as in the research grade ultraviolet-visible spectrometry used to measure the transmission spectrum of the elements. This was intentionally done to increase the etendue of the photometer, but results in additional spectral blurring and shifting. It is also likely that the spectral character of coatings on the curved surfaces of lenses have not been adequately accounted for in the model and that the original calibration spectra are themselves not perfectly corrected for the original recording instrument response, despite our best experimental effort. Overall, we expected a mean separation in class ratios of 0.251 based on an optical model of the system, and we observed a separation of 0.278 experimentally.

The differences in the sample standard deviations between the model and experiment can likewise be attributed to several factors. Among these factors are effects arising from the sources mentioned above, plus a couple of others. The first is that variability in the original calibration data may be artificially narrow due to their selection process. Potential calibration spectra were rejected for the technical reasons described in Part I,[25] such as escape of the cell from the optical trap. These technical rejections may have introduced a bias toward cells of a particularly uniform type. However, the most likely difference between the model and experiment lies in the S/N of the fluorescence imaging photometer, described in Part II.[26] In that article, we showed that the photometer meets, but does not significantly exceed, the minimum criteria for separating cells with the stated theoretical ratios with 95% confidence. Doubling up on the optical elements and taking an average was done in these tests specifically to improve the S/N of the measurement. Using Eq. 8 of Part I,[25] coupled with the S/N measured for the instrument in Part II,[26] we estimate the theoretical sample standard deviation due to variability in the single phytoplankton measurements to be near 0.06 ratio units. By doubling the MOEs, we expect this value to drop by about $\sqrt{2}$ to around 0.04 ratio units. The average experimental sample standard deviation was found to be around 0.033 units, so that the measured uncertainty is likely to be dominated by instrument effects.

Distribution of Ratios. The distribution of ratios for phytoplankton cells around the mean ratio for each species is of interest when considering how well species are truly separated. In the case of *E. huxleyi*, the distribution of ratios shown in Fig. 5 has a standardized second moment, a test for skewness,[27] of −0.360. This value suggests the data are not skewed compared with a normal distribution. The standardized third moment provides a test statistic for kurtosis;[27] the value for *E. huxleyi* is 3.22 in these data. This value is consistent with a normal distribution.

The same statistics for the *T. pseudonana* ratios are 0.36 and 4.87. Again, these values indicate a distribution not significantly skewed to the left or right, but the distribution is somewhat short-tailed compared with a normal distribution.

CONCLUSIONS

An imaging multivariate optical computing system was constructed for the classification of phytoplankton based on fluorescence excitation. The use of MOEs for classification implies that the system can be easily adapted to classify species based on any spectral property as long as the species or groupings of interest classify using linear discriminant analysis, or any other method amenable to analysis using spectral patterns.

We have shown that even two similarly pigmented species grown in culture, *E. huxleyi* and *T. pseudonana*, can be classified with a low error rate using this system via the algorithms described herein to process the data.

Several of improvements to and extensions of this approach are suggested by the work reported in this three-part series. First, the approach to designing optical elements based on discriminant analysis results is rather ad hoc; no native approaches to the design of elements have yet been implemented. To do so would require new code implementing new figures of merit in the optical design algorithms, and there are a wide variety of such calculations that could be done mimicking discriminants analysis natively, as well as hierarchical cluster analysis and other methods of classification.

Second, the implication from Part II that phytoplankton exhibit curiously variable fluorescence intensities at the single-cell level has yet to be investigated further. This has significant implications for the potential and speed of the analysis by our fluorescence imaging photometer.

Finally, improvements in the consistency of the filter wheel rotation frequency, and the pump stability, are desirable. Depth of focus for the instrument is likewise a concern because it limits the number of observed tracks that can be suitably measured. Extraction of size and shape information from the observed tracks is also desirable, using either image processing approaches or via measurements extracted from the track cross-sections, and spaces between streaks in a track.

Understanding how natural phytoplankton cells might appear in this measurement, and extension to studies in the field (e.g., shipboard) is of great interest. On a related note, extension of the design of MOEs from discrete species to classes of organisms, such as distinguishing haptophytes from diatoms from cyanobacteria, is also of great interest.

Another area for development is the analysis of multiple MOE streaks in a track by more flexible methods. In the present work, MOEs were measuring a single spectral pattern (linear discriminant vector 1), but in general there might be any number of MOEs measuring on many vectors. Simple ratios are not the most flexible means of combining the discrete measurements. An alternative approach would be a multivariate analysis applied directly to the MOE signals. This approach has the advantage of enabling post-processing or tuning of the response of the system long after the hardware has been placed in service, and allows it to be applied to tasks for which is was not originally designed.

Furthermore, we are interested in pushing the sensitivity of the method down to detection of picophytoplankton (nominally 0.2–2 μm in diameter) by a combination of improved S/N and automated control over the filter wheel rotation frequency and pump speed so that different regimes can be explored.

To a greater or lesser degree, work in all these areas is currently underway in our laboratories, although experience suggests that the kind reader should not hold his or her breath between subsequent installments of these reports.

ACKNOWLEDGMENTS

The authors acknowledge helpful conversations with Rob Olson of Woods Hole Oceanographic Institution. Funding for this study was provided by the National Science Foundation (grants OCE0623400 and OCE0958831 to TLR, MLM, and TJS).

1. P.F. Culverhouse, R. Williams, B. Reguera, V. Herry, S. Gonzalez-Gil. "Do Experts Make Mistakes? A Comparison of Human and Machine Identification of Dinoflagellates". Mar. Ecol. Prog. Ser. 2003. 247: 17-25.
2. A. Verikas, A. Gelzinis, M. Bacauskiene, I. Olenina, S. Olenin, E. Vaiciukynas. "Phase Congruency-Based Detection of Circular Objects Applied to Analysis of Phytoplankton Images". Pattern Recogn. 2012. 45(4): 1659-1670.
3. 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, G. Gorsky. "RAPID Research on Automated Plankton Identification". Oceanography. 2007. 20(2): 172-187.
4. D. Uhlmann, O. Schlimpert, W. Uhulmann. "Automated Phytoplankton Analysis by a Pattern Recognition Method". Int. Rev. Hydrobiol. 1978. 63(4): 575-583.
5. H.M. Sosik, R.J. Olson. "Automated Taxonomic Classification of Phytoplankton Sampled with Imaging-In-Flow Cytometry". Limnol. Oceanogr. Methods. 2007. 5: 204-216.
6. K.V. Embleton, C.E. Gibson, S.I. Heaney. "Automated Counting of Phytoplankton by Pattern Recognition: A Comparison with a Manual Counting Method". J. Plankton Res. 2003. 25(6): 669-681.
7. K. Rodenacker, B. Hense, U. Jutting, P. Gais. "Automatic Analysis of Aqueous Specimens for Phytoplankton Structure Recognition and Population Estimation". Microsc. Res. Tech. 2006. 69(9): 708-720.
8. T.P.A. Rutten, B. Sandee, A.R.T. Hofman. "Phytoplankton Monitoring by High Performance Flow Cytometry: A Successful Approach?". Cytometry A. 2005. 64A(1): 16-26.
9. C.S. Davis, Q. Hu, S.M. Gallager, X. Tang, C.J. Ashjian. "Real-Time Observation of Taxa-Specific Plankton Distributions: An Optical Sampling Method". Mar. Ecol. Prog. Ser. 2004. 284: 77-96.
10. 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, G. del Punta. "HAB Buoy: A New Instrument for In Situ Monitoring and Early Warning of Harmful Algal Bloom Events". Afr. J. Mar. Sci. 2006. 28(2): 245-250.
11. M.D. Mackey, D.J. Mackey, H.W. Higgins, S.W. Wright. "CHEMTAX - A Program for Estimating Class Abundances from Chemical Markers: Application to HPLC Measurements of Phytoplankton". Mar. Ecol. Prog. Ser. 1996. 144(1–3): 265-283.
12. T.L. Richardson, E. Lawrenz, J.L. Pinckney, R.C. Guajardo, E.A. Walker, H.W. Paerl, H.L. MacIntyre. "Spectral Fluorometric Characterization of Phytoplankton Community Composition Using the Algae Online Analyser (R)". Water Res. 2010. 44(8): 2461-2472.
13. A. Longhurst, S. Sathyendranath, T. Platt, C. Caverhill. "An Estimate of Global Primary Production in the Ocean from Satellite Radiometer Data". J. Plankton Res. 1995. 17(6): 1245-1271.
14. S.W. Jeffrey, M. Vesk. "Introduction of Marine Phytoplankton and Their Pigment Signatures". In: S.W. Jeffrey, R.F.C. Mantoura, S.W. Wright, editors. Phytoplankton Pigments in Oceanography: Guidelines to Modern Methods. Paris, France: UNESCO, 2005. Pp. 37-84.
15. S. Sathyendranath, G. Cota, V. Stuart, H. Maass, T. Platt. "Remote Sensing of Phytoplankton Pigments: A Comparison of Empirical and Theoretical Approaches". Int. J. Remote Sens. 2001. 22(2–3): 249-273.
16. J. Vepsalainen, T. Pyhalahti, E. Rantajarvi, K. Kallio, S. Pertola, T. Stipa, M. Kiirikki, J. Pulliainen, J. Seppala. "The Combined Use of Optical Remote Sensing Data and Unattended Flow-Through Fluorometer Measurements in the Baltic Sea". Int. J. Remote Sens. 2005. 26(2): 261-282.
17. C.S. Davis, F.T. Thwaites, S.M. Gallager, Q. Hu. "A Three-Axis Fast-Tow Digital Video Plankton Recorder for Rapid Surveys of Plankton Taxa and Hydrography". Limnol. Oceanogr. Methods. 2005. 3: 59-74.
18. G. Gorsky, P. Guilbert, E. Valenta. "The Autonomous Image Analyzer - Enumeration, Measurement and Identification of Marine-Phytoplankton". Mar. Ecol. Prog. Ser. 1989. 58(1–2): 133-142.
19. N. Hashemi, J.S. Erickson, J.P. Golden, K.M. Jackson, F.S. Ligler.

"Microflow Cytometer for Optical Analysis of Phytoplankton". Biosens. Bioelectron. 2011. 26(11): 4263-4269.
20. R.J. Olson, A. Shalapyonok, H.M. Sosik. "An Automated Submersible Flow Cytometer for Analyzing Pico- and Nanophytoplankton: FlowCytobot". Deep Sea Res. Part 1: Oceanogr. Res. Pap. 2003. 50(2): 301-315.
21. J.W. Hofstraat, W.J.M. Vanzeijl, M.E.J. Devreeze, J.C.H. Peeters, L. Peperzak, F. Colijn, T.W.M. Rademaker. "Phytoplankton Monitoring by Flow-Cytometry". J. Plankton Res. 1994. 16(9): 1197-1224.
22. C.S. Yentsch, D.A. Phinney. "Spectral Fluorescence - An Ataxonomic Tool for Studying the Structure of Phytoplankton Populations". J. Plankton Res. 1985. 7(5): 617-632.
23. M. Beutler, K.H. Wiltshire, B. Meyer, C. Moldaenke, C. Luring, M. Meyerhofer, U.P. Hansen, H. Dau. "A Fluorometric Method for the Differentiation of Algal Populations In Vivo and In Situ". Photosynth. Res. 2002. 72(1): 39-53.
24. L.S. Bruckman, T.L. Richardson, J.A. Swanstrom, K.A. Donaldson, M. Allora, T.J. Shaw, M.L. Myrick. "Linear Discriminant Analysis of Single-Cell Fluorescence Excitation Spectra of Five Phytoplankton Species". Appl. Spectrosc. 2012. 66(1): 60-65.
25. J.A. Swanstrom, L.S. Bruckman, M.R. Pearl, M.N. Simcock, K.A. Donaldson, T.L. Richardson, T.J. Shaw, M.L. Myrick. "Taxonomic Classification of Phytoplankton with Multivariate Optical Computing, Part I: Design and Theoretical Performance of Multivariate Optical Elements". Appl. Spectrosc. 2013. 67(6): 620-629. doi: 10.1366/12-06783.
26. J.A. Swanstrom, L.S. Bruckman, M.R. Pearl, E. Abernathy, T.L. Richardson, T.J. Shaw, M.L. Myrick. "Taxonomic Classification of Phytoplankton with Multivariate Optical Computing, Part II: Design and Experimental Protocol of a Shipboard Fluorescence Imaging Photometer". Appl. Spectrosc. 2013. 67(6): 630-639. doi: 10.1366/12-06784.
27. S.M. Kendall, A. Stuart. The Advanced Theory of Statistics: Distribution Theory. New York: Macmillan Publishing Co., Inc., 1977. Vol 1.