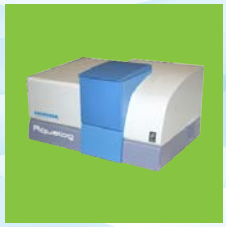
HORIBA Scientific

Harmful Algal Bloom Species

FLSS-36

ELEMENTAL ANALYSIS
FLUORESCENCE
GRATINGS & OEM SPECTROMETERS
OPTICAL COMPONENTS
FORENSICS
PARTICLE CHARACTERIZATION
RAMAN
SPECTROSCOPIC ELLIPSOMETRY
SPR IMAGING



Identification Strategies with the Aqualog® and Eigenvector, Inc. *Solo* Software

Summary

This study describes the application of simultaneous absorbance and fluorescence excitation-emission matrix (EEM) analysis for the purpose of identification and classification of freshwater planktonic algal species. The main foci were two major potentially toxic cyanobacterial species associated with algal bloom events in the Great Lakes region of the United States. The survey also included two genera and species of diatoms and one species of green algae. The study analyzed the precision and accuracy of the technique's ability to identify algal cultures as well as resolve and quantify mixtures of the different cultures. Described and compared are the results from both 2-way and 3-way multivariate EEM analysis techniques using the Eigenvector, Inc. *Solo* program.

It is shown that the 3-way Parallel Factor Analysis (PARAFAC) method can differentiate the major algal divisions and even provide limited culture level recognition based on component scores and ratios. Importantly, however, PARAFAC could not be effectively calibrated with pure algal cultures to yield unique culture-dependent spectral component loadings. In contrast, the 2-way Classical Least Squares (CLS) methodology can be calibrated with pure algal culture spectra and taxonomic enumeration data to yield concentrations, cell densities and biovolume estimates, thus providing the potential capacity for rapid and precise algal recognition and quantification. In fact, the CLS method can effectively separate different cyanobacterial mixtures with over 2 orders of magnitude variation in relative cell density and biovolume with practical predicted limits of detection in a few hundreds of cells/ml.

Introduction

Cyanobacterial species associated with algal blooms can create health and safety issues, as well as a financial impact for drinking water treatment plants. These blooms are a particular issue in the Great Lakes region of the United States in the late summer months. Several species of cyanobacteria (also known as blue-green algae) can produce a variety of toxins including hepatotoxins and neurotoxins. In addition, some species can produce so-called taste and odor compounds that, though not toxic, can lead to drinking water customer complaints, and thus represent a considerable treatment objective. The two major cyano species in this study, *Microcystis aeruginosa* and *Anabaena flos-aquae*, are commonly associated with such notorious bloom events in Lake Erie. The bloom events can be triggered by warm weather and are believed to be exacerbated by agricultural runoff of fertilizers, particularly dissolved reactive phosphorous, and wastewater discharge.

The traditional methods involved in monitoring and prediction of bloom events center on time- and labor-intensive laboratory procedures. These primarily include: 1) Microscopy for species identification and quantification (taxonomic enumeration), 2) Enzyme Linked Immunoassay (ELISA) for quantification of the toxins, and 3) GC-MS for taste and odor compounds. Hence the main purpose of this study was to investigate the potential for rapid, reagent-free identification and quantification with the Aqualog's patented simultaneous absorbance and EEM data acquisition coupled with multi-way analysis using Eigenvector, Inc.'s *Solo* package.

JOBIN YVON Technology

Simultaneous absorption and EEM data collection affords an opportunity for either2-wayand or 3-way multivariate analyses. Two-way analysis involves the unfolding of the 3-way EEM data into individual excitation-emission-intensity pairs to generate a 2-dimensional spectral fingerprint. For this study, the 2-way model type employed the Classical Least Squares (CLS) method which is calibrated with pure component spectra and concentration data to form a library. CLS functions to evaluate by least squares fitting the linear contributions of the library components in unknown samples. Parallel Factor Analysis (PARAFAC) method is the 3-way analysis which can be calibrated with a set of sample EEMs to generate a tri-linear model of correlated excitation and emission spectral loading and concentration dependent scores based on an assumed number of components in the calibration EEM set. PARAFAC calibration cannot constrain model components to individual EEM samples or files but instead treats the calibration EEM set as a whole. PARAFAC is arguably most useful for extracting spectral component information without prior knowledge of the calibration system's components especially when individual EEM samples in the calibration set contain multiple excitation-emission spectral components.

The models were tested using two *Cyanophytes, M. aeruginosa, A. Flos-aquae*, two Chrysophyte diatom species (*Nitzschia sp. and Mayamaea sp.*) and one Chlorophyte algal species (*Selanastrum sp.*). The report below describes the significance of results obtained by attempting to calibrate the 2- and 3-way models using the same pure algal culture data followed by attempts to validate the models with test culture samples and mixtures in comparison to independent taxonomic enumeration.

Materials and Methods

Pure algal cultures were provided in log-growth phase by the USEPA (Cincinnati, OH). Suspensions of each species were diluted with high-quality water (QW: > 18 MΩ, TOC < 0.1 ppm) such that the desired peak chlorophyll absorbance (1 cm path length) at 680 nm (versus a blank of QW) was obtained. Diluted suspensions were constantly stirred in a 1-cm path quartz fluorescence cell (at 25°C).

The algal taxonomic enumeration was determined using 400X magnification according to the 2012 APHA Utermöhl method (APHA, 2012. Standard Methods for Examination of Water and Wastewater, 22 ed. American Public Health Association (APHA), American Water Works Association (AWWA), Water Environment Federation (WEF), USA.).

The Aqualog conditions included a 5 nm excitation interval, CCD binning of 2 pixels (approximately 1 nm), and medium gain with an integration time range between 0.5 and 1 second. All EEMs were normalized based on the instrument settings to a 1 µm NIST SRM936 quinine sulfate unit (QSU) in perchloric acid. For model calibration EEMs, collected in replicates of 5 per pure algal culture sample, were corrected for instrument spectral bias, Rayleigh scatter, and inner-filter effects using the Aqualog software. To maximize the signal to noise ratio for the model calibrations, the five replicate EEMs for each culture were averaged. EEMs were analyzed using Eigenvector, Inc. Solo as described below.

Results and Discussion

Typical Algal EEMs

Table 1 describes the taxonomy of the five algal species analyzed identified using the taxonomic enumeration method. Figure 1 shows the algal EEMs used for model calibrations organized to compare the three main types of algae including: the two (Fig. 1 A and B) different orders of Cyanophytes (*M. aeruginosa and A. flos-aquae*), the two (Fig. 1C and D) diatoms (*Mayamaea sp. and Nitzschia sp.*), representing two different families of Chrysophytes, and the (Fig. 1E) Chlorophyte (*Selenastrum sp.*). What is clear from these EEMs is the striking similarity of the Cyanophyte EEMs signified by the major strong excitation/emission feature around 615 nm/655 nm attributed to the phycobilin chromophores. Likewise the two diatoms also exhibited very similar EEM features with excitation bands in the blue region attributed to the fucoxantin based light-harvesting complexes and the chlorophyll emission bands in the red region (around 680 nm). Lastly the green algal specimen exhibited typical chlorophyll-carotenoid based light harvesting excitation in the blue and chlorophyll emission band in the red (around 685 nm).

Table 1. Taxonomic Description of the Five Algal Cultures Analyzed

Genus	Family	Order	Class	Division	Taxon
Anabaena	Nostocaceae	Nostocales	Myxophyceae	Cyanophyta	bluegreen
Microcystis	Microcystaceae	Chroococcales	Myxophyceae	Cyanophyta	bluegreen
Selenastrum	Oocystaceae	Chlorococcales	Chlorophyceae	Chlorophyta	green
Nitzschia	Bacillariaceae	Pennales	Bacillariophyceae	Chrysophyta	diatom
Mayamaea	Naviculaceae	Pennales	Bacillariophyceae	Chrysophyta	diatom

PARAFAC

The PARAFAC calibration model (Fig. 1) included the five separate cultures, each measured as the sum of five replicates. The calibration resolved a maximum of three spectral loadings based on the core-consistency (93%) and model redundancy notifications and accounted for >95% of the variance. The EEM contours of the spectral component loadings are shown in Figure 2. Component 1 (C1) in Fig. 2A shows one major contour mode closely resembling the cyanophyte emission in Figs. 1A and B with a major excitation/emission (Ex/Em) coordinate peak around 615 nm/655 nm. Component 2 (C2) in Fig. 2B shows several contour modes closely resembling the Chrysophyte emission overall in Figs. 1C and D with broad excitation/emission coordinate peaks around 460 nm/680 nm and minor Ex/Em peaks around 640/680 nm. Component 3 (C3) in Fig. 2C shows several contour modes closely resembling the Chlorophyte emission overall in Fig. 1E with broad excitation/emission coordinate peaks around 440 nm/680 nm and minor Ex/Em peaks around 650/680 nm.

Figure 3 shows the PARAFAC component scores corresponding to the spectral loading contours in Fig. 2. Component 1 had the highest score in both Cyanophytes, M. *aeruginosa* (MC) and *A. Flos-aquae* (A). Component 2 dominated both the Chrysophytes, *Myamaea* (MD) and *Nitzschia* (N). Component 3 dominated the Chlorophyte *Selenastrum*(S). Importantly, the results indicated that the PARAFAC model, when calibrated with five pure unique algal cultures under these conditions, did not yield five unique species-specific spectral component loadings. Clearly, however, the component scores especially the dominant ones did distinguish each of the three major algal Orders, namely, Cyanophyta, Chrysophyta, and Chlorophyta, respectively. The mean±SD component scores were graphically different comparing A and MC, with relatively higher scores for component 3 and lower scores for component 1 in A. A student's T-test confirmed that the means of both C1 and C3 for A and MC were significantly different with less than a $2x10^{-4}$ probability of either of them coming from the same mean and population. However, as indicated graphically by the overlapping mean±SD values the PARAFAC scores C2 and C3 could not be used to significantly distinguish the two diatoms MD and N; the student's t-test indicated >39% probability they shared the same mean population.

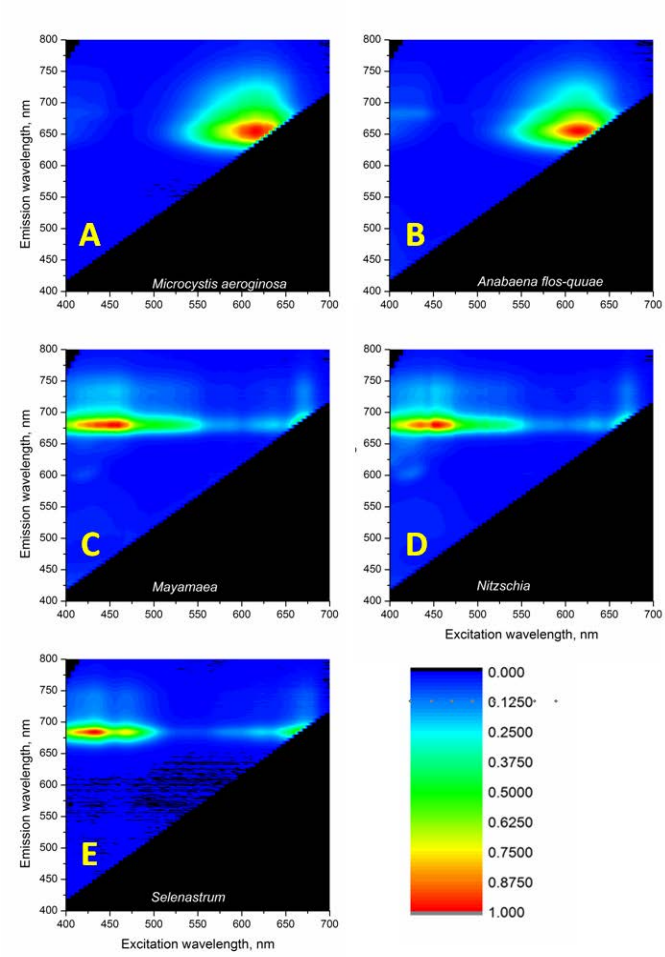


Figure 1. Excitation emission matrices representing the sum of five individual sample measurements of each species used for the 2-way and 3-way model calibrations.(A) *M. aeruginosa;* (B) *A. flos-aquae;* (C) *Mayamaea sp.;* (D) *Nitzschia sp.;* (E) *Selenastrum sp.*

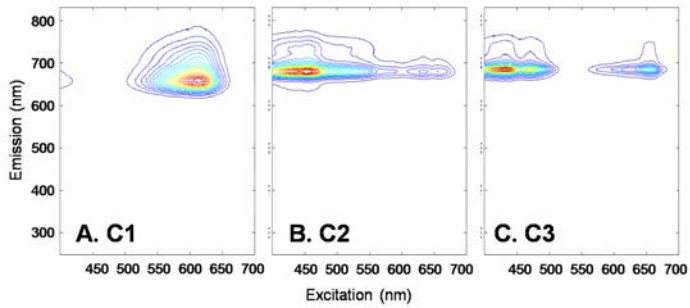


Figure 2. PARAFAC spectral loading contours for the 3 component (C1-C3) calibration model corresponding to the five species of algae and EEMs in Figure 1.

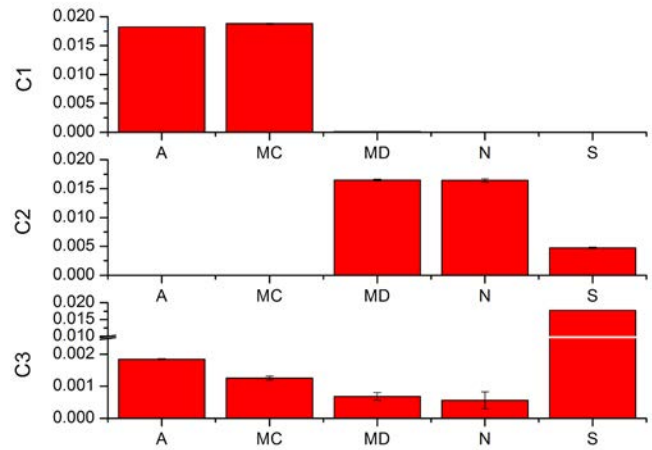


Figure 3. PARAFAC component scores (±SD) for components (C1-C3) for five replicate individual sample EEMs of each species validated with the calibration model from Figure 2. Note scale break on y-axis for C3.

The fact that the PARAFAC calibration model did not yield five unique spectral loadings is consistent with the mathematical principle of PARAFAC, which correlates orthogonal excitation and emission component spectral shapes among the entire set of EEMs in the calibration data set. It is not possible to constrain PARAFAC to generate a unique component for a given unique sample (in this case a singular algal species) in a given mixed calibration set of EEMs. Further, it is well-known that PARAFAC components do not necessarily reflect chemically distinct components and it is also possible for similar spectral shapes to be correlated with different chemical components. For example, in Figure 3 it is likely that similarities between the spectral shapes of the Fucoxanthin light harvesting components in the Chrysophytes and the chlorophyll –carotenoid light-harvesting components of the Chlorophyte spectra are purely coincidental to explain the observation of component 2 in the latter.

Classical Least Squares

In contrast to the orthogonal shape- and calibration-sample-set-dependent model calibration principles of PARAFAC, the Classical Least Square (CLS) method provides an obligatory means for calibration. In this study CLS used 2-way EEM data from the individual algal species to generate unique corresponding spectral library components. CLS also has the additional benefit of the capacity to assign quantitative concentration dependent information for each EEM in the calibration set. Thus CLS significantly enhances the potential to make positive identification in an unknown mixture at greater phylogenic resolution (e.g., order, family, genus, or perhaps species level). Figure 4 shows the results of validation of five replicate samples of each of the five algal cultures. Despite the apparently similar spectral features of the Cyanophytes and Chrysophytes the data confirm that there was a remarkable positive identification (ID) fraction on average of greater than 97% for each culture. In fact, the only evidence of minor crosstalk was observed between algae of a similar Division such as in sample N3 or A1 where a minor fraction of the MD or MC components, respectively, were included in the model.

The phylogenetic resolution of the CLS method was further tested using reciprocal mixtures of A and MC mixed to a total *A680* cm^{-1} of 0.16 to generate mixtures of a predicted linear ratio of the two key Cyanophytes. Figure 5 shows the CLS method was able to distinguish, quantify and identify the two species with a very high linear correlation over a wide range of A:MC absorbance ratios varying from 1 to 0.005 (or 100 to 0.5%).

Cell Density and Biovolume Calibration

The relationship between the cell density (cells/ml) and biovolume (µm3/ml), measured independently by taxonomic enumeration, was evaluated by linear regression for the same A and MC mixtures in Figure 5. The measured *A680* and QSU normalized fluorescence data from the CLS analysis were evaluated. Table 2 shows the linear regression statistics, described with an intercept fixed at 0, with standard errors in the slopes of around CV=13% for A and CV=9.67% for MC with adjusted r^2 values of 0.88 and 0.93, respectively. Table 2 also shows that A exhibited for the same *A680* much higher cell densities (around 34-fold) and biovolumes (around 26-fold) than M. Likewise, the fluorescence conversion factor exhibited around 16 and 12 fold higher cell densities and biovolumes, respectively, for A and MC. The ratio of *A680* and fluorescence conversion factors for cell density (or biovolumes) was around 123 for A and 58 for MC indicating the fluorescence quantum yields, based on the normalized *A680* values, were significantly different for the two cultures analyzed under these conditions.

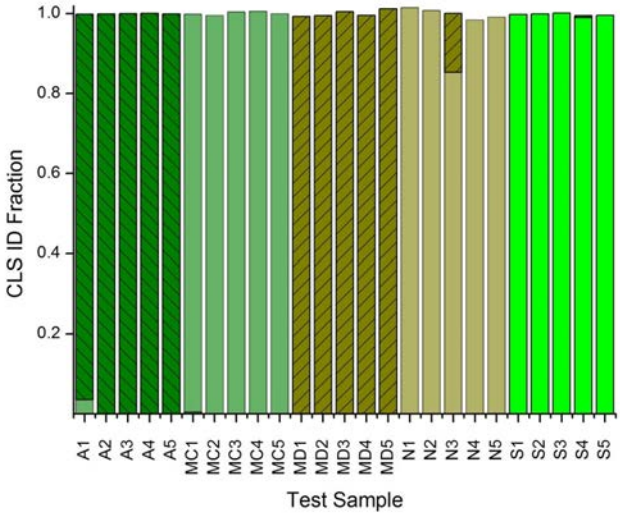


Figure 4. CLS component positive ID Fractions of five individual validation replicates for each algal sample based on the model calibrated using the summed EEMs from Figure 1.

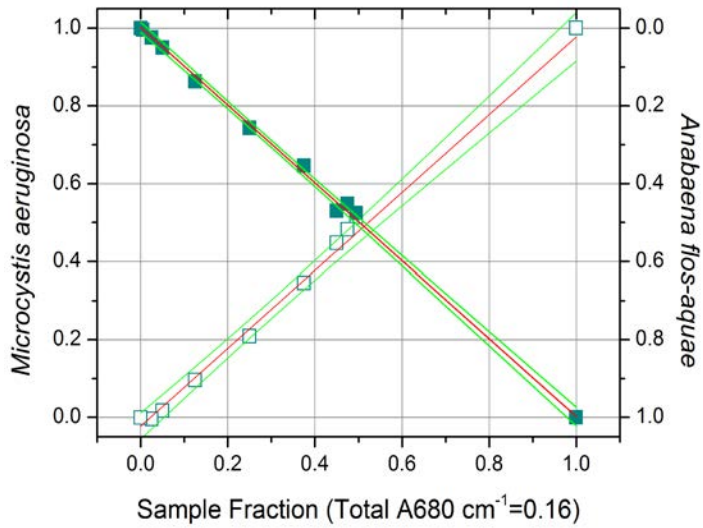


Figure 5. Linear regression plots of the actual and predicted relative absorbance values from CLS analysis for a reciprocal mixture of *Anabaena flos-quae* (closed) and *Microcystis aeruginosa* (open) corresponding to the same *A680* cm^{-1} for each corresponding sample. The slope was fixed at 1 and the red lines represent the model fit and the green lines the 95% confidence intervals. The adjusted r^2 values were 0.99797 for A and 0.9882 for MC.

Table 2. Linear regression statistics for the relationships between cell density and biovolume and the absorbance at 680 nm and QSU Fluorescence intensity for *A. flos-aquae* and *M. aeruginosa.*

	A680 (cm-1)				F-QSU					
Species	Cells/ml	SE	Biovolume (µm³/ml)	SE	Cells/ml	SE	Biovolume (µm³/ml)	SE	CV%	Adjusted r^2
A. Flos-aquae	1.88E+07	2.45E+06	9.45E+08	1.23E+08	1.53E+05	1.99E+04	7.68E+06	1.00E+06	13.01	0.879
M. aeruginosa	5.54E+05	5.36E+04	3.62E+07	3.50E+06	9.52E+03	9.20E+02	6.22E+05	6.01E+04	9.67	0.930
Ratio	33.9		26.1		16.1		12.3			

Figure 6 further explores the significance of the cell density and biovolume calibrations by plotting the values predicted using the conversion factors from Table 2 as a function of values predicted by CLS in the reciprocal mixtures of A and MC described in Figure 5. Most notably, in the presence of roughly 1.35E+06 cells/ml of A it was possible to detect and identify MC at around 1.5E+03 cells/ml (i.e., close to a 900-fold cell density ratio); this corresponded to roughly a 688-fold difference in biovolume.

In Figure 7 the relative limits of detection for the pure culture *A. flos-aquae* were estimated by linear regression for a serial dilution with three replicates for each sample of as a function of the *A680* cm^{-1} to predict the cell density and biovolume based on the enumeration calibration data in Table 2. The lowest concentration (see inset Fig. 7) used in the dilution corresponded to an *A680* cm^{-1} value of 7.4E-5 which corresponded to a predicted cell density of 2428±377 cells/ml and a biovolume of 1.22E+05±1.89E+04µm3/ml. Notably the integration time used for all samples was 0.5 s thus it is reasonable to assume that as a function of integration time the sensitivity and precision could easily be increased linearly for a predicted limit of detection of around a few hundred cells/ml.

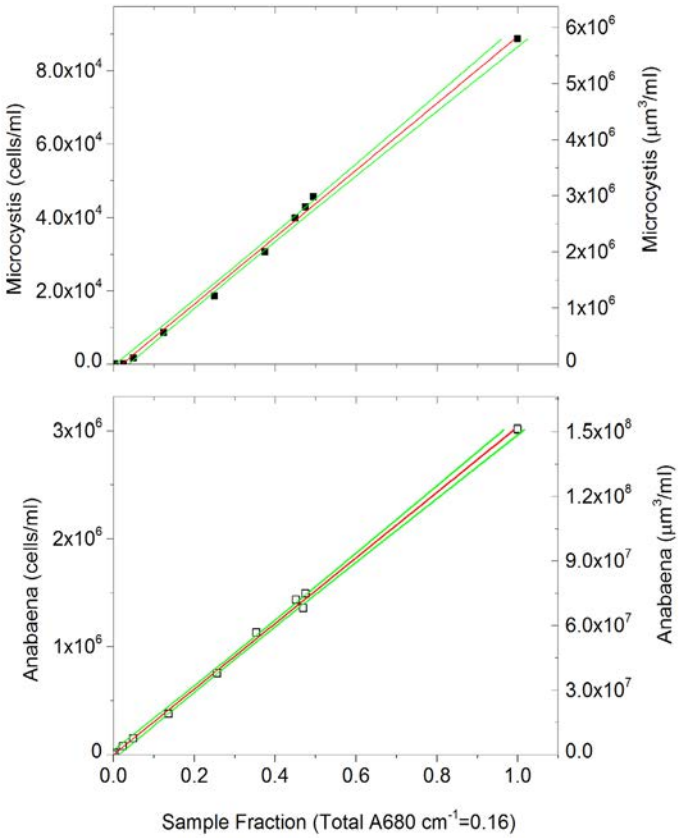


Figure 6. Comparison of the cell density (cells/ml) and biovolume (µm³/ml) model estimates for the *Anabaena* and *Microcystis* samples from Figure 5. The red line defines the fit and the green lines represent the 95% confidence bands.

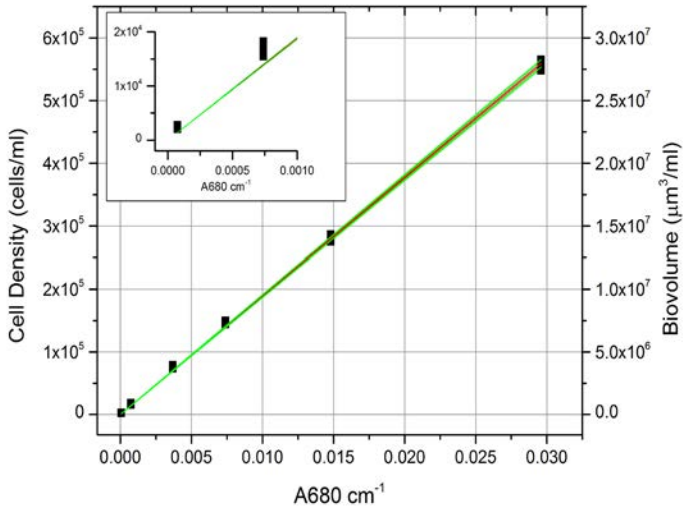


Figure 7. Estimated limit of detection for *A. flos-aquae* determined by linear regression of a serial dilution as a function of *A680* cm^{-1} and predicted cell density and biovolume. The EEM and absorbance conditions included a 0.5 s integration time at medium gain. Values for *A680* cm^{-1} were predicted by CLS as described previously. The linear regression statistics for cell density and biovolume included an intercept fixed at 0, an adjusted r^2=0.999 and predicted slopes of 1.88E7±9.32E5 and 9.45E8±4.68E6, respectively. The red line defines the fit and the green lines represent the 95% confidence bands. Note the expanded scale in the inset.

Conclusions

The conclusions of this study center on the flow chart in Figure 8 which explains how the CLS and PARAFAC methods in *Solo* can be compared as a basis for establishing rapid, reliable analysis methods for differentiating different algal types. The PARAFAC method, when calibrated with pure culture spectra, affords the capacity to rapidly and effectively separate and identify major algal orders (Cyanophyta, Chrysophyta and Chlorophyta). PARAFAC can also possibly identify different types of cyanobacteria more specifically based on evaluation component scores and score ratios. PARAFAC calibration may not yield unique spectral loadings for each algal culture nor necessarily allow unambiguous assignments of spectral components to identifiable physical components in the calibration data set. CLS on the other hand requires calibration with pure culture samples and exhibits the capacity to generate unique, culture-specific spectral component libraries. Further, based on independent cell density and biovolume calculations it was clear the CLS method can potentially provide rapid, reliable estimates of these parameters with high-precision. In conclusion, it is clear that the Aqualog EEM and *Solo* model analyses show potential for rapid evaluation of algal types and could prove useful in rapid field and laboratory level cyano harmful algal bloom identification, especially when carefully calibrated with pure algal culture samples. Due caution is noted in terms of dealing with the morphological and physiological properties of the samples in terms of spectroscopy and the requirement for due diligence in terms of establishing detection concentration limits in the ranges needed for positive identification. Importantly, it is well known that algal spectral properties are significantly affected by the light-environment, temperature and nutrient conditions among other factors so algal culture methods used for calibration would ideally take these factors into consideration and independent methods of qualification and quantification including taxonomic enumeration are advised.

Acknowledgments

HORIBA Instruments, Inc. thanks EPA Cincinnati for providing the algal cultures, and Mrs. Sakiko Akaji for assistance with analysis and data acquisition.

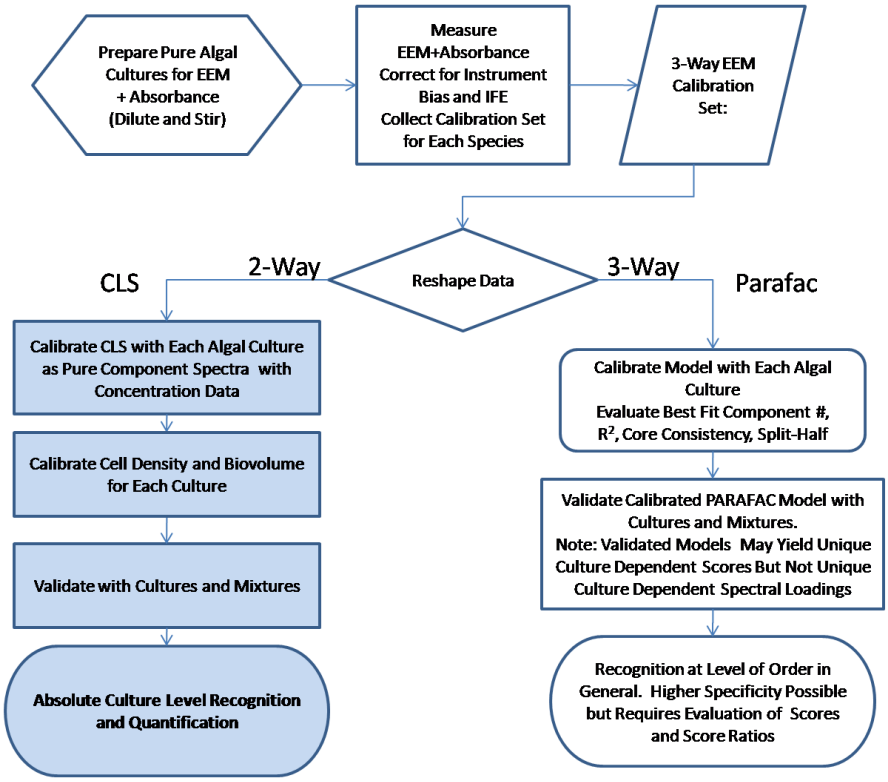


Figure 8. Flow chart of experimental design for EEM acquisition, multivariate analysis and cell density/biovolume calibration methods.