Canadian Journal of Fisheries and Aquatic Sciences
 Journal canadien des sciences halieutiques et aquatiques

Improved estimates of phytoplankton community composition based on in situ spectral fluorescence: use of ordination and field-derived norm spectra for the bbe FluoroProbe

Journal:	<i>Canadian Journal of Fisheries and Aquatic Sciences</i>
Manuscript ID	cjfas-2015-0360.R3
Manuscript Type:	Article
Date Submitted by the Author:	16-Feb-2016
Complete List of Authors:	Harrison, Joel; University of Waterloo, Biology Howell, Todd; Ontario Ministry of the Environment Watson, Susan; Environment Canada, National Water Research Institute Smith, Ralph; University of Waterloo,
Keyword:	FluoroProbe, fluorescence, PHYTOPLANKTON < Organisms, TAXONOMY < General, ordination

SCHOLARONE™
 Manuscripts

Improved estimates of phytoplankton community composition based on *in situ* spectral fluorescence: use of ordination and field-derived norm spectra for the bbe FluoroProbe

*Joel W. Harrison¹, E. Todd Howell², Susan B. Watson³, and Ralph E. H. Smith¹

* Current Address: North Bay-Mattawa Conservation Authority, 15 Janey Avenue, North Bay, ON, Canada, P1C 1N1.

¹Department of Biology, University of Waterloo, 200 University Avenue West, Waterloo, ON, Canada, N2L 3G1.

²Environmental Monitoring & Reporting Branch, Ontario Ministry of the Environment and Climate Change, 125 Resources Road, Toronto, ON, Canada, M9P 3V6.

³Watershed Hydrology and Ecology Division, Canada Centre for Inland Waters, Canada Centre for Inland Waters, Environment Canada, 867 Lakeshore Road, Burlington, ON, Canada, L7S 1A1.

ABSTRACT

The use of spectral fluorometers for assessing phytoplankton concentrations and taxonomic composition in aquatic environments is increasingly common. However, the accuracy of such assessments suffers because the necessary norm spectra (spectral fingerprints) are derived using selected taxa and laboratory conditions that may not adequately represent the taxa and environmental conditions in the study area. Ordination analysis of raw fluorescence data has been proposed as a better means of interpreting spectral fluorescence data. We applied nonmetric multidimensional scaling and cluster analysis to raw *in situ* fluorescence data from Sturgeon Bay, a small, mesotrophic embayment of Georgian Bay (Lake Huron) in order to obtain system-specific norm spectra for the bbe FluoroProbe. The revised spectra gave improved estimates of phytoplankton taxonomy (RMSE of 10% vs. 14%) and of dissolved organic carbon and chlorophyll *a* concentrations. While promising, this method should be further explored in other systems with different and/or weaker gradients in phytoplankton biomass and taxonomic composition.

INTRODUCTION

Major phytoplankton groups differ in photosynthetic pigmentation and therefore can be distinguished by their fluorescence excitation and emission spectra (Yentsch and Yentsch, 1979; Yentsch and Phinney, 1985). A number of commercially-available spectral fluorometers allow phytoplankton community composition to be estimated through space and time at scales for which manual sample collection and identification via microscopic examination would be impractical. The need for such monitoring tools is acute, given the importance of phytoplankton to global biogeochemical cycling (Field et al. 1998), their role in aquatic foodwebs (Kalff 2002), and the potential for bloom-forming taxa to degrade water quality, especially as cultural eutrophication and climate change progress (Paerl & Otten 2013). The FluoroProbe (FP; bbe Moldaenke GmbH, Kiel, Germany) is a commonly-used profiling instrument that measures fluorescence emission at ~680 nm in response to excitation by light emitting diodes (LEDs) centered at approximately 370, 470, 525, 570, 590, and 610 nm. The FP software provides estimates of

chromophoric dissolved organic matter (CDOM; referred to as ‘yellow substances’ by the software) and chlorophyll *a* (chl *a*) concentrations of 4 phytoplankton groups based on their fluorescence excitation spectra: chlorophytes, chromophytes, cryptophytes, and cyanobacteria. In lakes, these groups chiefly comprise the Chlorophyta and Euglenophyta (chlorophylls *a* & *b*), Heterokontophyta, Haptophyta, and Dinophyta (chlorophylls *a* & *c*, fucoxanthin/peridinin), Cryptophyta (chlorophylls *a* & *c*, phycobilins), and Cyanophyta (chl *a*, phycobilins), respectively. The FP estimates are made by linear unmixing, i.e., by solving linear equations to obtain group-specific concentrations from the observed fluorescence excitation spectrum and group-specific norm spectra (Beutler et al. 2002). Alternative statistical approaches for estimating group-specific concentrations from spectral data also exist (MacIntyre et al. 2010; Neveux et al. 2011). The default norm spectra for each FP unit provided by the manufacturer are determined by measuring fluorescence excitation spectra of laboratory cultures (of known chl *a* concentration) of phytoplankton species representative of the major pigment groups (Beutler et al. 2002).

Despite increasingly-common use, assessments of FP performance under field conditions have noted inaccuracy in chl *a* determination (Ghadouani and Smith 2005; Gregor and Maršálek 2004), and misclassification of phytoplankton taxa (Catherine et al. 2012). The linear unmixing approach, as described above, has a number of limitations (reviewed by MacIntyre et al. 2010). In contrast to lab cultures, the irradiance and nutrient conditions experienced by phytoplankton in nature are dynamic, with effects on both light absorption and fluorescence. The fluorescence yield of chl *a* is not constant *in situ* but declines at high light due to non-photochemical (photoprotective) quenching (NPQ) of absorbed excitation energy (Falkowski and Raven 2007). Changes in nutrient availability and irradiance (photoacclimation, nonphotochemical quenching) influence the pigmentation and optical properties (including the chl *a*:fluorescence ratio) of different phytoplankton taxa, potentially affecting the accuracy of taxonomic identification based on spectral fluorescence (MacIntyre et al. 2010 and references therein; Escoffier et al. 2015). Furthermore, fluorescence excitation spectra can vary considerably among

different taxa within the same major pigment groups (Johnsen and Sakshaug 2007) and the representative species from which norm spectra are obtained will likely differ from the dominant taxa *in situ*.

Ordination, for example Principal Components Analysis (PCA), allows patterns of similarity to be discerned and quantified in multivariate data, such as measurements of fluorescence at multiple wavelengths at a range of sites (Alexander et al. 2012). It has been used to derive spectral signatures ('base spectra') and reveal putative patterns in CDOM and phytoplankton community composition without reliance on pre-defined norm spectra for each fluorescent group (Alexander et al. 2012, Alexander & Imberger 2013), but only in a single, highly-eutrophic lake. It is desirable to explore the applicability of the ordination approach in additional freshwater systems and further validate its utility with more direct comparison to independent measurements than in previous studies. We applied the ordination method to FP data collected in Sturgeon Bay, a small mesotrophic embayment of Georgian Bay (Lake Huron) characterized by strong spatial-temporal gradients in phytoplankton biomass, community composition, and dissolved organic carbon (DOC) concentration. In a step forward from Alexander et al. (2012), we conducted microscopic analysis and made quantitative comparisons against the fluorescence measurements. Our aims in doing so were to (1) assess the accuracy of the FP when used with its default norm spectra, (2) determine whether ordination analysis could identify base spectra, and (3) to determine whether revised norm spectra, calculated from base spectra, could produce more accurate estimates of chl *a*, phytoplankton community composition and DOC concentration than those obtained using the instrument-default norm spectra.

METHODS

Field sampling

Sturgeon Bay is a small (5.56 km²) embayment located on the eastern coast of Georgian Bay, Lake Huron, Canada (Fig. 1). Distinctive spatial variability in water quality is a feature of this narrow and irregularly sub-divided embayment (Schiefel 2004). The northern portion is deeper, more productive, and

further removed from exchanges with Georgian Bay than the southern portion. The bay was sampled at an approximately bi-weekly interval from early-June to mid-October in 2006 and 2007. Four sites (686, 687, 688, 689) were sampled regularly in 2006 (13 times) and 2007 (10 times) for the parameters described below; in 2006, two additional sites (Central 2, South 2) were sampled for FP data only (Fig. 1; Table 1). Vertical profiles of water column fluorescence (from surface to 1-m-above bottom) were measured with a FP (serial number 1401; bbe Moldaenke GmbH, Kiel, Germany) at each site on each sampling date. Integrated epilimnetic samples (limited to the depth within the photic zone) were collected from sites 686, 687, 688, and 689 using a thick-walled plastic tube lowered through the water column and the water stored on ice. All samples were shipped to the Ontario Ministry of the Environment Laboratory in Toronto within 24 hours and analyzed using standard Ontario Ministry of Environment and Climate Change methods (Chow et al. 2010) for a suite of water chemistry parameters, including total phosphorus (TP), DOC, and chl *a*. The term FP chl *a* is used to distinguish FP-derived estimates of chl *a* made *in situ* from those based on laboratory analysis of filtered samples (i.e., acetone-extracted chl *a* determined by spectrophotometry). Subsamples (100 mL) were preserved with Lugol's iodine for later identification and enumeration of the phytoplankton community.

Determination of phytoplankton biomass & taxonomy

Preserved phytoplankton samples for 45 of the 92 date-sites (Table 1) were analysed by a professional taxonomist for species composition and biomass using the standard Utermöhl inverted microscope technique. Depending on sample density, subsamples of 2-5 mL were settled over 24 hours and counted across transects at 100×, 200× or 400× using an inverted phase microscope. Cell counts were converted to biomass (wet weight; mg/m³) using biovolumes (µm³) calculated from average measured cell dimensions and geometric approximations of shape, assuming a specific density of 1.0 g/cm³. Taxa were identified to genus or species levels; colonial and filamentous forms were measured individually and biomass calculated as a function of average cell density per biovolume (Findlay and Kling 1988). Xanthophyte

algae, which contain chlorophylls *a* & *c* but lack fucoxanthin, were assigned to the chlorophyte pigment group (see MacIntyre et al. 2010).

Data analysis

All statistics were performed using R (R Core Team 2013).

Estimation of CDOM and phytoplankton chl a from FluoroProbe fluorescence

CDOM and group-specific FP chl *a* concentrations were estimated from norm spectra and the *in situ* fluorescence measured at each date-site-depth based on the following relationship:

$$sfs = Kc + e \quad (1)$$

where *sfs* is a 6×1 matrix (column vector) of detected fluorescence (from excitation at 370, 470, 525, 570, 590, 610 nm) for the date-site-depth, *c* is a 5×1 matrix of concentrations (5 groups = CDOM + 4 taxa) for the date-site-depth, *K* is a 6×5 matrix comprising the norm spectra (6 wavelengths $\times$ 5 groups), and *e* is a 6×1 error matrix (residuals) (see Neveux et al. 2011 or Seppälä and Olli 2008). The Lawson-Hanson least squares algorithm (Lawson and Hanson, 1995) was used to find a non-negative solution for *c* (R package: *nnls*) for each date-site-depth. A randomly-chosen subset (site 689 on 16 Oct 2006) of the estimates of CDOM and the total and group-specific FP chl *a* were checked against those generated by the FP software (v. 2.2.4) using Pearson correlation analysis to ensure computational accuracy (all $r > 0.998$).

Derivation of base spectra through ordination of spectral fluorescence data

Ordination of FP fluorescence data was performed according to Alexander et al. (2012) with some modifications. Prior to analysis, fluorescence data were averaged into 0.5-m bins. Shapiro-Wilk tests (performed separately on the data from each excitation wavelength), inspection of histograms and quantile-quantile plots, and the Doornik-Hansen omnibus test for multivariate normality (performed on the entire dataset) all indicated that the fluorescence dataset lacked multivariate normality. Non-metric

multidimensional scaling (NMDS), a non-parametric ordination method, was therefore performed instead of PCA. For use in the NMDS analysis, the (binned) fluorescence data for each date-site-depth were normalized by dividing the value for each excitation wavelength by the average value for all 6 excitation wavelengths. NMDS was performed on the Bray-Curtis dissimilarity matrix to produce a 2-dimensional ordination of the fluorescence data. To aid in identifying the termini of the ordination, which identify samples that represent base spectra (Alexander *et al.* 2012), k-means clustering (a common agglomerative clustering technique based on Euclidean distance) was used to assign the points to 100 groups. Terminal clusters of the ordination were then chosen based on visual inspection.

Calculation of new norm spectra from base spectra to improve predictions by linear unmixing

New norm spectra were calculated from base spectra via several steps (Fig. 2). The raw fluorescence excitation spectra for the date-site-depths comprising two of the termini of the NMDS plot were averaged, then corrected by subtracting the fluorescence signal of deionized water. The resultant spectra were assumed to correspond to phytoplankton groups based on visual inspection of their shapes and knowledge of the default phytoplankton norm spectra; these were corrected for dissolved fluorescence by subtracting the CDOM contribution (putative CDOM excitation spectrum $\times$ DOC concentration) to the total fluorescence. The putative CDOM spectrum was obtained from the terminus of the NMDS corresponding to data from hypolimnetic waters characterized by relatively high CDOM and no FP chl *a* (based on the default spectra). The only date-site-depth (9 Jul 2007-site 688-11 m) of its cluster for which DOC concentration had been determined (only a limited number of hypolimnetic grab samples were obtained for water chemistry analysis) was used to produce a fluorescence excitation spectrum for 1 mg/L DOC (after subtracting the fluorescence signal of deionized water). The putative phytoplankton spectra were normalized to the respective (extracted, epilimnetic) chl *a* concentrations to obtain excitation spectra for 1 μ g/L chl *a* of each pigment group. CDOM and group specific FP chl *a* concentrations were then recalculated for the entire dataset using these new spectra, as described above (Eq. 1). Both the default and new norm spectra were tested for linear independence using multiple regression analysis; the spectra were

considered to be linearly independent if no one spectrum could be significantly ($p < 0.05$) expressed as a weighted combination of the others.

Assessment of accuracy of default vs. revised norm spectra

Simple linear regression analysis was applied to the 88 date-sites for which paired FP and water quality data were obtained (see *Field sampling*) to quantify the proportion of variation in DOC and chl *a* that could be explained by FP-based estimates of CDOM and total FP chl *a*. The accuracy of taxonomic discrimination was quantified as the root-mean-squared-error (RMSE) of prediction for each phytoplankton group.

RESULTS

Water Clarity and Chemistry

TP, DOC, and chl *a* generally increased (Fig. 3) with distance from the mouth of the bay (Fig. 1). Likewise, water clarity was higher at sites 686 and 689 than at 687 and 688 (Fig. 3).

Phytoplankton Biomass and Taxonomy

The total and group-specific phytoplankton biomass (wet weight) varied among date-sites by more than an order of magnitude (Table 2). Chromophytes and cyanobacteria were most prominent, comprising up to 95% and 89% of the total biomass, respectively, at times, whereas chlorophytes and cryptophytes generally made a minor contribution to the biomass (Table 2). The (identified) taxa that most commonly dominated or codominated the phytoplankton biomass (Table 3) were the cyanobacterial genus *Anabaena* (29 times), the dinoflagellate genus *Peridinium* (8 times), and the diatom genera *Aulacoseira* and *Tabellaria* (4 times each).

Ordination of Fluorescence Excitation Spectra

166 The NMDS plot (2-D stress = 0.032) revealed 3 main spectral gradients and associated termini (Fig. 4).
167 The FP data comprising terminus 1 (11, 25 Jul 2006, 9 Jul 2007; site 688; 10-12.5 m) originated from
168 hypolimnetic waters, as previously noted (see Methods). The FP data comprising terminus 2 were from
169 1-7.5 m at site 688 on 17 Sep 2007. The FP data comprising terminus 3 were from 1.5-2.5 m at site 686
170 on 27 Jun 2006. Epilimnetic phytoplankton biomass, as estimated by microscopy, was dominated by
171 cyanobacteria (73%) for the samples associated with terminus 2. Phytoplankton biomass and taxonomy
172 were not analyzed for the date-site associated with terminus 3; however, the FP data which clustered
173 closest to terminus 3 (1-3.5 m at site 689 on 4 Sep 2007; red inverted triangles in Fig. 4) were associated
174 with an epilimnetic sample for which the biomass was dominated by chromophytes (81%). Interestingly,
175 FP data (obtained using the default norm spectra) corresponding to the date-sites when and where the
176 community biomass was most strongly dominated by cyanobacteria (89%; 20 Sep 2006 at 688) and
177 chromophytes (95%; 27 Jun 2006 at 689) plotted near, but not at, termini 2 and 3, respectively, of the
178 NMDS plot (Fig. 4).

179 *New Norm Spectra*

180 The new norm spectra differed strongly from the default spectra in magnitude (new spectra ~50% of
181 default spectra) but more subtly in shape (Fig. 5). In particular, the ratio between fluorescence at 590 nm
182 and 610 nm (F590:610) of the default cyanobacteria spectrum was less than unity, whereas F590:610 of
183 the new cyanobacteria spectrum was greater than unity. The F470:525 ratio was much higher for the
184 default CDOM spectrum than for the new spectrum. The shapes of the default and new chromophyte
185 spectra were very similar (Fig. 5). Both the default (all $p > 0.32$) and new (all $p > 0.34$) norm spectra
186 were found to be linearly independent based on multiple linear regression analyses.

187 *Fluorescence-Based Estimates of DOC, chl a, and Phytoplankton Taxonomy*

188 FP-based estimates of CDOM are reported by the instrument software in relative units, an
189 acknowledgement of the potentially-large variability in the optical properties of DOC. Here, a linear

relationship was observed between FP-based estimates of CDOM and DOC concentration, when CDOM was estimated using the default norm spectra (slope = 2.70; intercept = 1.46). Use of the revised CDOM spectrum resulted in an almost 1:1 relationship between FP-based estimates of CDOM and DOC (slope = 0.94, intercept = 1.90; Fig. 6). There was a marginal decrease in the proportion of variance in DOC explained by CDOM upon application of the new CDOM norm spectrum ($R^2 = 0.62$ vs. $R^2 = 0.61$). FP chl *a* underestimated the chl *a* concentration by a factor of >2 when estimated using the default norm spectra (slope = 1.91; intercept = 1.04); the relationship was substantially improved with use of the new norm spectra (slope = 1.09; intercept = -0.51; Fig. 7). There was a marginal decrease in the proportion of variance in chl *a* explained by FP chl *a* upon application of the new norm spectra ($R^2 = 0.94$ vs. $R^2 = 0.91$).

Using the default norm spectra, FP chl *a* (as % total) usually underestimated the chromophyte contribution to the total biomass and overestimated the cryptophyte contribution (Figs. 8, 9). Application of the new norm spectra (Fig. 5) improved the overall relationship between taxonomic composition as biomass and as FP chl *a*, particularly for the chromophyte and cryptophyte groups (Figs. 8, 9). The RMSE of prediction for percent composition using the new and default spectra, respectively, was 13% vs. 17% for chromophytes, 6% vs. 20% for cryptophytes, no change (8%) for chlorophytes, and 12% vs. 11% for cyanobacteria; on average the RMSE was 10% using the new spectra vs. 14% using the default spectra.

DISCUSSION

Ordination (NMDS) of FP observations from our temperate, mesotrophic system showed the pattern theoretically expected (Alexander et al. 2012) if samples represent variable mixtures of a finite number of spectral groups and include mixtures that are heavily dominated by just one group. This is, to our knowledge, the first demonstration outside of Lake Victoria that such a pattern can be observed in nature. K-means clustering successfully identified putative base spectra that allowed us to calculate new norm

spectra from which we made revised estimates of DOC and phytoplankton group-specific chl *a*. Overall, these estimates showed improved agreement with traditional measures of phytoplankton biomass (extracted chl *a*) and taxonomy (microscopy), with some exceptions, despite the apparent base spectra not solely comprising pure phytoplankton pigment groups (see *Prediction of Phytoplankton Taxonomy*).

FluoroProbe vs. Extracted Chlorophyll a

Use of the default spectra gave strong correlations between chl *a* and FP chl *a*, but the magnitude of chl *a* was substantially underestimated by the FP. This result is consistent with that of several other field studies (Catherine et al. 2012; Ghadouani and Smith 2005; Gregor and Maršálek 2004; Twiss 2011) which found FP to underestimate chl *a*. NPQ induced by solar radiation may cause the FP to underestimate the true chl *a* concentration (J.W. Harrison, unpublished data; Leboulanger et al. 2002; Serra et al. 2009); however, 28% of our FP profiles (with complementary chl *a* data) were collected before 10:00 h or after 16:00 h, when NPQ would be expected to be minimal. Even during periods of high insolation, NPQ would not be expected at depth in Sturgeon Bay due to the relatively low water clarity (Fig. 3). Furthermore, work with cultures has shown that the chl *a* concentration of different taxa within the same pigment group can be strongly either over- or under-estimated using the same instrument and norm spectra under laboratory conditions (Escoffier et al. 2015; Kring et al. 2011), which may reflect differences between the chl *a*-specific fluorescence yields of the taxa used for calibration by the manufacturer and those under study. Whatever the underlying cause, use of the new norm spectra resolved this issue within our dataset. We note that if only total chl *a* is of interest then correction factors could be obtained more easily from linear regression of FP chl *a* vs. chl *a* (provided that a strong linear correlation between chl *a* and the FP chl *a* exists for the phytoplankton community of interest) without the need to calculate revised norm spectra from field data.

Prediction of Phytoplankton Taxonomy

There are a number of factors that could explain the more accurate correspondence between group-specific FP chl *a* and group-specific phytoplankton biomass we observed when using the field data-derived norm spectra in place of the default norm spectra for chromophytes, cyanobacteria, and CDOM. If linear unmixing is to be successfully employed, norm spectra must be (1) linearly independent, (2) determined with sufficient accuracy, and (3) of invariable shape (Beutler et al. 2002). The first condition can be validated arithmetically for a given set of norm spectra (Beutler et al. 2002; Escoffier et al. 2015) as was done for the default and new spectra used in this study. Given that the default norm spectra were in fact linearly independent, one possible explanation for the improvement we observed is that criterion 2 was not met, i.e., that the FP used in our study was not accurately calibrated by the manufacturer. We have no reason to suspect that this is the case, nor can we directly test such a suspicion, but the results of an inter-instrument comparison (Twiss 2011) suggest that the quality of calibration may vary among FPs. What is known with certainty is that condition 3 is not valid. The chl *a*:fluorescence ratio of phytoplankton is not constant, but varies between species within a pigment-group, and with irradiance conditions (NPQ, photoacclimation), nutrient status, and cell size; these factors are given comprehensive consideration by MacIntyre et al. (2010). Recently, Escoffier et al. (2015) found the effects of both nitrogen (N) availability (i.e., N-replete vs. N-depleted media) and recent light exposure on the accuracy of the FP to be species-specific for a number of laboratory cultures of freshwater phytoplankton. With respect to inter-specific differences in excitation spectra, Beutler et al. (2002) explicitly stated that ‘a determination of norm curves with algae obtained from the actual place of interest is required for high-precision determinations.’ Given that isolating and culturing representative species to obtain site- or system-specific norm spectra is not practical, except perhaps for where a single species completely dominates the biomass of a pigment group (e.g., Houliez et al. 2012), the ordination approach may represent a more viable alternative.

Use of the revised norm spectra led to an overall improvement in the predictive accuracy of the FP, but in general the instrument continued to underestimate the contribution of chromophytes and overestimate the

contribution of cryptophytes (though by less than with the default norm spectra), and considerable scatter remained in the relationships between group-specific FP chl *a* and group-specific phytoplankton biomass. A number of factors likely contributed to the remaining error. First, only 3 of 5 norm spectra were revised, as neither cryptophytes nor chlorophytes ever dominated the community biomass. Second, the NMDS termini fluorescence data used to calculate the new norm spectra were partially representative of phytoplankton pigment groups other than those that the new spectra were used to estimate (i.e., the date-site-depths corresponding to the phytoplankton base spectra data were only 73% cyanobacteria and 81% chromophytes by biomass; see Results, and *General Applicability of the Ordination Method*). Third, the error around microscope estimates of phytoplankton biomass is typically $\pm 10\text{--}20\%$, reflecting misidentification, subsampling and counting error, and biovolume calculations, and estimates based on microscopy do not include picoplankton. Direct observations of autotrophic picoplankton have not been reported in our study area, but elsewhere in Lake Huron they have been reported to range from 0.5 to 50% of phytoplankton biomass, averaging 10% (Fahnenstiel and Carrick 1992). Picoplankton in other eutrophic areas of the Laurentian Great Lakes (e.g., Bay of Quinte, Lake Erie) typically comprise <6% of the total phytoplankton biomass (S.B. Watson, unpublished data). Fourth, it is well established that the biomass (carbon) to chl *a* ratio of phytoplankton cells can vary considerably with temperature, photoacclimation status, and/or nutrient availability (Falkowski and Raven 2007), so perfect correlation between FP chl *a* and phytoplankton biovolume would not be expected *a priori*. Lastly, as discussed below, fluorescence from CDOM may have played a role. We note that if both spectral fluorescence and phytoplankton biomass (or chemotaxonomic) data are obtained for a given system, site-specific norm spectra can be obtained using least squares estimation, as described in Seppälä and Olli (2008).

CDOM and Prediction of DOC Concentration

The fluorescence excitation spectrum of DOC is a function of the pool of molecular constituents it comprises, and is therefore a function of its source (Hudson et al. 2007), the degree of microbial processing (Cammack et al. 2004), and photochemical modification by ultraviolet radiation (Morris and

Hargreaves 1997). Using the default spectrum, the FP generates CDOM estimates in relative units. Here, we derived a new norm spectrum with the intention of predicting DOC concentration in absolute units (i.e., mg/L). However, upon application of this revised spectrum, we underestimated DOC concentration by ~2 mg/L. This is likely because the hypolimnetic DOC from which the new norm spectrum was obtained was of a more chromophoric nature than the epilimnetic DOC, due to being less photobleached (i.e., the intercept of 1.90 mg/L (Fig. 6) likely represents the concentration of colorless DOC in the total DOC pool, for which fluorescence emission, and therefore CDOM, is zero). This inference is consistent with the study of Twiss (2011) who recommended the use of stratum-specific CDOM norm spectra for the FP, based on vertical differences in CDOM observed in the Laurentian Great Lakes. The applicability of a CDOM norm spectrum obtained from a single date/site (especially if obtained from hypolimnetic waters) to a broader spatial or temporal scale may represent an improvement on the FP's default CDOM spectrum, but due to the potential for considerable spatial variation in the optical properties (spectral absorption) of DOC within a single system (Smith et al. 2004), it is still likely to be of rather limited utility. Furthermore, the ordination approach used here cannot be applied to datasets for which dissolved fluorescence was not measured in the absence of appreciable phytoplankton fluorescence *in situ*. That is, data from a continuous subsurface water intake on survey cruises, or from vertical profiles during periods of complete water column mixing will not include hypolimnetic data from which to isolate a CDOM norm spectrum. The measurement of fluorescence from filtered (e.g., <0.2–0.8 μm) water samples to complement all FP data is therefore recommended where logistically practical, especially in systems characterized by strong temporal and/or spatial variability in DOC concentration and/or spectral properties.

General Applicability of the Ordination Method

The apparent effectiveness of the ordination method, reported here, and espoused (though not rigorously validated) by Alexander et al. (2012), reflects its application to relatively large datasets, obtained from systems characterized by strong spatial-temporal gradients in phytoplankton taxa. Despite the strong

gradients in phytoplankton community composition observed in Sturgeon Bay, the base spectra identified by NMDS were of partly mixed composition. This underscores the inherent limitation of this approach: it is not possible to isolate 'pure' spectral signatures (Alexander et al. 2012) from within a dataset that does not include them. Furthermore, the groups dominant in both Sturgeon Bay (this study) and Lake Victoria (Alexander and Imberger, 2013) were chromophytes and cyanobacteria, the fluorescence excitation spectra of which differ markedly in shape, at least when the cyanobacteria are dominated by phycocyanin-rich genera, as they were in Sturgeon Bay. Phycoerythrin-rich picocyanobacteria typically constitute less than 20% of the total picocyanobacterial biomass in eutrophic embayments of the Laurentian Great Lakes (S.B. Watson, unpublished data). Distinguishing chromophytes from chlorophytes (both with high light absorption in the blue region of the spectrum due to chlorophylls) or cyanobacteria from cryptophytes (both with high light absorption in the red region of the spectrum due to phycobilins) via ordination could be more challenging, though chlorophyte and cryptophyte algae are rarely dominant contributors to the phytoplankton biomass in most lakes (Watson et al. 1999; Kalff 2002). Finally, we note that while both our study and that of Alexander et al. (2012) analyzed spectral fluorescence data from the bbe FP, the same ordination approach could be more broadly applied (i.e., to any type of spectral data).

Summary

Norm spectra for chromophytes, cyanobacteria, and CDOM were obtained from ordination of *in situ* spectral fluorescence data from FP and used to make improved, though still imperfect, estimates of DOC, chl *a*, and phytoplankton taxonomy in Sturgeon Bay, a small, hydrologically-isolated embayment of southeastern Georgian Bay, Lake Huron. While further research is needed, the utility of this approach is likely to be restricted to aquatic systems characterized by strong spatial or temporal gradients in phytoplankton community composition.

ACKNOWLEDGEMENTS

Funding for this study was obtained from the Lake Simcoe / South-eastern Georgian Bay Clean-Up Fund, the Ontario Ministry of the Environment and Climate Change, and Environment Canada. Microscopy-based estimates of phytoplankton taxonomy and biomass were provided by Hedy Kling of Algal Taxonomy and Ecology Inc., Winnipeg, Manitoba. We thank Myles M. Harrison for providing useful statistical advice.

REFERENCES

- Alexander, R., Gikuma-Njuru, P., and Imberger, J. 2012. Identifying spatial structure in phytoplankton communities using multi-wavelength fluorescence spectral data and principal component analysis. *Limnology and Oceanography: Methods*, 10: 402-415.
- Alexander, R. and Imberger, J. 2013. Phytoplankton patchiness in Winam Gulf, Lake Victoria: a study using principal component analysis of in situ fluorescent excitation spectra. *Freshwater Biology*, 58: 275–291.
- Beutler, M., Wiltshire, K.H., Meyer, B., Moldaenke, C., Lürling, C., Meyerhöfer, M., Hansen, U.-P., Dau, H. 2002 A fluorometric method for the differentiation of algal populations *in vivo* and *in situ*. *Photosynthesis Research*, 72:39–53
- Cammack, W.K.L., Kalff, J., Prairie, Y.T., and Smith, E.M. 2004. Fluorescent dissolved organic matter in lakes: relationships with heterotrophic metabolism. *Limnology and Oceanography*, 49: 2034–2045.
- Catherine, A., Escoffier, N., Belhocine, A., Nasri, A.B., Hamlaoui, S., Yéprémian, C., Bernard, C., and Troussellier, M. 2012. On the use of the FluoroProbe[®], a phytoplankton quantification method based on

- 360 fluorescence excitation spectra for large-scale surveys of lakes and reservoirs. *Water Research*, 46:
361 1771–1784.
- 362 Chow, J., Abbey, A.I., Khan, Z., Dermicheva, S., Jennings, W., and Wilson, P. 2010. 2009 Performance
363 Report: General chemistry and microbiology analyses section. Ontario Ministry of the Environment
364 Report, Queen's Printer for Ontario, Ontario, Canada.
365
366
- 367 Escoffier, N., Bernard, C., Hamlaoui, S., Groleau, A., and Catherine, A. 2015. Quantifying
368 phytoplankton communities using spectral fluorescence: the effects of species composition and
369 physiological state. *Journal of Plankton Research*, 37: 233–247.
- 370
- 371 Fahnenstiel, G.L. and Carrick, H. J. 1992. Phototrophic picoplankton in Lakes Huron and Michigan:
372 Abundance, distribution, composition and contribution to biomass and production. *Canadian Journal of*
373 *Fisheries and Aquatic Sciences*, 49: 379-388.
- 374
- 375 Falkowski, P. G. and Raven J. A. 2007. *Aquatic Photosynthesis*. 2nd edn. Princeton University Press,
376 Princeton, New Jersey, USA.
- 377
- 378 Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkowski, P.G. 1998. Primary production of the
379 biosphere: integrating terrestrial and oceanic components. *Science*, 281: 237–240.
- 380
- 381 Findlay, D. L. and Kling, H. J. 1988. Protocols for Measuring Biodiversity: Phytoplankton in Freshwater.
382 [www.eman-rese.ca/350 eman/ecotools/protocols/freshwater/phytoplankton/intro.html](http://www.eman-rese.ca/350%20eman/ecotools/protocols/freshwater/phytoplankton/intro.html)
383

- 384 Ghadouani, A. and Smith, R.E.H. 2005. Phytoplankton distribution in Lake Erie as assessed by a new *in*
385 *situ* spectrofluorometric technique. *Journal of Great Lakes Research*, 31 (Suppl. 2): 154–167.
386
- 387 Gregor, J. and Maršálek, B. 2004. Freshwater phytoplankton quantification by chlorophyll
388 *a*, a comparative study of *in vitro*, *in vivo* and *in situ* methods. *Water Research*, 38: 517–552.
389
- 390 Houliez, E., Lizon, F., Thyssen, M., Artigas, L.F., and Schmitt, F.G. 2012. Spectral fluorometric
391 characterization of Haptophyte dynamics using the FluoroProbe: an application in the eastern English
392 Channel for monitoring *Phaeocystis globosa*. *Journal of Plankton Research*, 34, 136–151.
393
- 394 Hudson, N., Baker, A., and Reynolds, D. 2007. Fluorescence analysis of dissolved organic matter in
395 natural, waste and polluted waters—a review. *River Research and Applications*, 23: 631–649.
396
- 397 Johnsen, G. and Sakshaug, E. 2007. Biooptical characteristics of PSII and PSI in 33 species (13 pigment
398 groups) of marine phytoplankton, and the relevance for pulse-amplitude-modulated and fast-repetition-
399 rate fluorometry. *Journal of Phycology*, 43: 1236–1251.
400
- 401 Kalff, J. 2002. *Limnology: inland water ecosystems*. Prentice-Hall Inc., Upper Saddle River, NJ, USA.
402
- 403 Kring, S.A., Figary, S.E., Boyer, G.L., Watson, S.B., and Twiss, M.R. 2014. Rapid *in situ* measures of
404 phytoplankton communities using the bbe FluoroProbe: evaluation of spectral calibration, instrument
405 intercompatibility, and performance range. *Canadian Journal of Fisheries and Aquatic Sciences*, 71:
406 1087–1095.
407

- 408 Leboulanger, C., Dorigo, U., Jacquet, S., Le Berre, B., Paolini, G., Humbert, J.-F. 2002. Application of a
409 submersible spectrofluorometer for rapid monitoring of freshwater cyanobacterial blooms: a case study.
410 *Aquatic Microbial Ecology*, 30: 83–89.
411
- 412 Lawson, C.L. and Hanson, R.J. 1995. *Solving Least Squares Problems. Classics in Applied Mathematics.*
413 SIAM, Philadelphia, USA.
414
- 415 MacIntyre, H.L., Lawrenz, E., and Richardson, T.L. 2010. Taxonomic discrimination of phytoplankton
416 by spectral fluorescence. In: D.J. Suggett et al. (eds.), *Chlorophyll a Fluorescence in Aquatic Sciences:*
417 *Methods and Applications*, Developments in Applied Phycology 4, DOI 10.1007/978-90-481-9268-7_7
418
- 419 Morris, D.P. and Hargreaves, B.R. 1997. The role of photochemical degradation of dissolved organic
420 carbon in regulating the UV transparency of three lakes on the Pocono Plateau. *Limnology and*
421 *Oceanography*, 42: 239–249.
422
- 423 Neveux, J., Seppälä, J., and Dandonneau, Y. 2011. Multivariate analysis of extracted pigments using
424 spectrophotometric and spectrofluorometric methods. In: Roy et al. (eds.), *Phytoplankton Pigments:*
425 *Characterization, Chemotaxonomy and Applications in Oceanography*, Cambridge University Press.
426
- 427 Paerl, H.W. and Otten, T.G. 2013. Harmful cyanobacteria blooms: Causes, consequences, and controls.
428 *Environmental Microbiology*, DOI 10.1007/s00248-012-0159-y
429
- 430 R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for
431 Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
432

Schiefer, K. 2004. Water Quality Study of Sturgeon Bay: Results of Sampling on 20 September 2004.
A report prepared for the Township of The Archipelago.

Seppälä, J. and Olli, K. 2008. Multivariate analysis of phytoplankton spectral *in vivo* fluorescence:
estimation of phytoplankton biomass during a mesocosm study in the Baltic Sea. *Marine Ecology
Progress Series*, 370: 69-85.

Serra, T., Borrego, C., Quintana, X., Calderer, L., López, R., and Colomer, J. 2009. Quantification of the
effect of nonphotochemical quenching on the determination of *in vivo* chl *a* from phytoplankton along the
water column of a freshwater reservoir. *Photochemistry and Photobiology*, 85, 321–331.

Smith, R.E.H., Allen, C.D., Charlton, M.N. 2004. Dissolved organic matter and ultraviolet radiation
penetration in the Laurentian Great Lakes and tributary waters. *Journal of Great Lakes Research*, 30:
367–380.

Twiss, M.R. 2011. Variations in chromophoric dissolved organic matter and its influence on the use of
pigment-specific fluorimeters in the Great Lakes. *Journal of Great Lakes Research*, 37: 124–131.

Watson, S., McCauley, E. and Downing, J.A. 1997. Patterns in phytoplankton taxonomic composition
across temperate lakes of differing nutrient status. *Limnology and Oceanography*, 42: 487-495.

Yentsch, C.S. and Phinney, D.A. 1985 Spectral fluorescence – an ataxonomic tool for studying the
structure of phytoplankton populations. *Journal of Plankton Research*, 7:617–632.

Yentsch ,C.S. and Yentsch, C.M. 1979 Fluorescence spectral signatures – characterization of
phytoplankton populations by the use of excitation and emission-spectra. *Journal of Marine Research*,
37:471–483.

Draft

TABLES

Table 1. Sturgeon Bay site names, UTM coordinates, approximate maximum depths, and the number of samples (dates) for which phytoplankton biomass was analysed by microscopy.

Station	Northing (m)	Easting (m)	Zone	Depth (m)	# samples
Central 2	5050433	545358	17	7.0	0
South 2	5049147	546338	17	10.4	0
686	5048485	546280	17	10.9	3
687	5050395	546038	17	6.5	3
688	5051240	544244	17	13.8	23
689	5049170	547071	17	6.4	16

Table 2. Summary statistics for phytoplankton biomass, as estimated by microscopy, for the 45 date-sites for which counts were completed.

Group	Biomass (mg m ⁻³)			Contribution to Biomass (%)		
	Min.	Median	Max.	Min.	Median	Max.
total	150	1089	4373	—	—	—
Chlorophytes	16	66	478	0	7	27
Chromophytes	64	501	3062	8	48	95
Cryptophytes	5	69	267	1	5	40
Cyanobacteria	18	254	3909	1	30	89

Draft

Table 3. Major contributors to the phytoplankton biomass in Sturgeon Bay for each date-site for which microscope counts were completed.

Taxonomic Group	Genus	Times dominant or codominant	Contribution to total biomass where dominant (%)		
			Min.	Avg.	Max.
Chlorophytes	<i>Botrococcus</i>	1	9	9	9
Chrysophytes	not identified	7	12	41	66
	<i>Uroglena</i>	3	17	26	36
	<i>Chrysosphaerella</i>	1	23	23	23
Cryptophytes	<i>Cryptomonas</i>	1	26	26	26
Cyanobacteria	<i>Anabaena</i>	29	8	28	68
	<i>Woronichinia</i>	3	17	20	25
	<i>Microcystis</i>	2	8	15	22
	<i>Aphanocapsa</i>	1	9	9	9
	<i>Lyngbya</i>	1	8	8	8
Diatoms	<i>Aulacoseira</i>	4	9	15	18
	<i>Tabellaria</i>	4	14	16	17
	<i>Cyclotella</i>	3	15	21	25
	<i>Fragilaria</i>	3	8	9	11
	<i>Asterionella</i>	2	20	24	27
Dinoflagellates	<i>Peridinium</i>	8	6	12	19
	<i>Ceratium</i>	3	6	8	12
	<i>Gymnodinium</i>	2	11	13	15
	<i>Glenodinium</i>	1	9	9	9

FIGURE CAPTIONS

Figure 1. The study sites on Sturgeon Bay; the bay's position on the eastern coast of Georgian Bay, Lake Huron, is indicated by the dot in the inset panel. Coordinates correspond to Zone 17T of the UTM conformal projection. Map base data drawn from the Ontario Ministry of Natural Resources and Forestry Ontario Hydrographic Network Database.

Figure 2. Step-by-step depiction of the procedure used to calculate new phytoplankton norm spectra (chromophyte data shown). (a) Base spectrum is corrected for deionized water signal. (b) CDOM fluorescence for the date-site-depth corresponding to the phytoplankton base spectrum is calculated from the CDOM norm spectrum and DOC concentration, a proxy for CDOM concentration and (c) subtracted from the DIW-corrected base spectrum. (d) The CDOM-corrected spectrum is then normalized by dividing by the acetone-extracted, epilimnetic chl *a* concentration that corresponds to the date-site from which the base spectrum was isolated. This procedure was repeated for the data from each date-site-depth of each terminus of the NMDS plot and the resultant spectra averaged to obtain the new norm spectra.

Figure 3. Estimated photic depths (2 x Secchi depth), and epilimnetic concentrations of TP, DOC, and extracted chl *a* at the 4 main study sites in Sturgeon Bay (from outermost to innermost) during 2006-07 ($n = 23$).

Figure 4. NMDS ordination plot produced from FP fluorescence data collected during Jun-Oct 2006-07 at 6 sites on Sturgeon Bay. The points are shaded in proportion to chl *a* concentration. Termini data are depicted by yellow triangles. One (hypolimnetic) outlying datum was excluded from the cluster chosen as terminus 3. Colored squares denote epilimnetic FP data (obtained using the default reference spectra) for the date-sites when/where cyanobacteria (89.4%; 20 Sep 2006 at 688; cyan squares) and chromophytes (94.7%; 27 Jun 2006 at 689; orange squares) made the highest relative contributions to the phytoplankton biomass. The inverted red triangles denote the cluster closest to terminus 3 for which phytoplankton biomass and taxonomy were analysed.

Figure 5. Default norm spectra (open symbols) and new norm spectra (closed symbols) for CDOM (triangles), chromophytes (squares), and cyanobacteria (circles). Error bars represent standard deviations.

Figure 6. FP-based CDOM estimates as a predictor of DOC for all date-sites ($n=86$) based on default spectra (grey circles) and new spectra (black circles). The solid lines are the respective trend lines for each data series.

Figure 7. FP chl a as a predictor of extracted chl a for all date-sites ($n=86$) based on default spectra (grey circles) and new spectra (black circles). Triangles represent outliers (based on visual inspection) excluded from each regression analysis (site 688 on 11 Sep and 3 Oct 2006). The solid lines are the respective trend lines for each data series; the dashed line is the 1:1 line.

Figure 8. Differences between relative community composition based on biomass (microscopy) and FP-inferred chl a equivalents calculated using the default (left panel) and new reference spectra (right panel). Whiskers represent the max and min values (exclusive of outliers, defined as data beyond 1.5 x interquartile range from 1st and 3rd quartiles), boxes span the 1st to 3rd quartiles, and black lines within boxes represent medians.

Figure 9. Percent FP chl a by pigment group as a predictor of percent biomass by pigment group using the instrument-default spectra (top) and the NMDS-derived spectra for CDOM, cyanobacteria, and chromophytes with the default spectra for chlorophytes and cryptophytes (bottom). The solid lines are the trendlines from simple linear regression analyses; the dashed lines are the 1:1 lines.

FIGURES

Figure 1. The study sites on Sturgeon Bay; the bay’s position on the eastern coast of Georgian Bay, Lake Huron, is indicated by the dot in the inset panel. Coordinates correspond to Zone 17T of the UTM conformal projection. Map base data drawn from the Ontario Ministry of Natural Resources and Forestry Ontario Hydrographic Network Database.

599

Figure 2. Step-by-step depiction of the procedure used to calculate new phytoplankton norm spectra (chromophyte data shown). (a) Base spectrum is corrected for deionized water signal. (b) CDOM fluorescence for the date-site-depth corresponding to the phytoplankton base spectrum is calculated from the CDOM norm spectrum and DOC concentration, a proxy for CDOM concentration and (c) subtracted from the DIW-corrected base spectrum. (d) The CDOM-corrected spectrum is then normalized by dividing by the acetone-extracted, epilimnetic chl *a* concentration that corresponds to the date-site from which the base spectrum was isolated. This procedure was repeated for the data from each date-site-depth of each terminus of the NMDS plot and the resultant spectra averaged to obtain the new norm spectra.

Figure 3. Estimated photic depths (2 x Secchi depth), and epilimnetic concentrations of TP, DOC, and extracted chl *a* at the 4 main study sites in Sturgeon Bay (from outermost to innermost) during 2006-07 ($n = 23$).

Figure 4. NMDS ordination plot produced from FP fluorescence data collected during Jun-Oct 2006-07 at 6 sites on Sturgeon Bay. The points are shaded in proportion to chl *a* concentration. Termini data are depicted by yellow triangles. One (hypolimnetic) outlying datum was excluded from the cluster chosen as terminus 3. Colored squares denote epilimnetic FP data (obtained using the default reference spectra) for the date-sites when/where cyanobacteria (89.4%; 20 Sep 2006 at 688; cyan squares) and chromophytes (94.7%; 27 Jun 2006 at 689; orange squares) made the highest relative contributions to the phytoplankton biomass. The inverted red triangles denote the cluster closest to terminus 3 for which phytoplankton biomass and taxonomy were analysed.

Figure 5. Default norm spectra (open symbols) and new norm spectra (closed symbols) for CDOM (triangles), chromophytes (squares), and cyanobacteria (circles). Error bars represent standard deviations.

Figure 6. FP-based CDOM estimates as a predictor of DOC for all date-sites ($n=86$) based on default spectra (grey circles) and new spectra (black circles). The solid lines are the respective trend lines for each data series.

Figure 7. FP chl *a* as a predictor of extracted chl *a* for all date-sites ($n = 86$) based on default spectra (grey circles) and new spectra (black circles). Triangles represent outliers (based on visual inspection) excluded from each regression analysis (site 688 on 11 Sep and 3 Oct 2006). The solid lines are the respective trend lines for each data series; the dashed line is the 1:1 line.

Figure 8. Differences between relative community composition based on biomass (microscopy) and FP-inferred chl *a* calculated using the default (left panel) and new norm spectra (right panel). Whiskers represent the max and min values (exclusive of outliers, defined as data beyond 1.5 x interquartile range from 1st and 3rd quartiles), boxes span the 1st to 3rd quartiles, and black lines within boxes represent medians.

Figure 9. Percent FP chl *a* by pigment group as a predictor of percent biomass by pigment group using the default norm spectra (top) and the new norm spectra (bottom). The solid lines are the trendlines from simple linear regression analyses; the dashed lines are the 1:1 lines.