Geocarto International
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tgei20

Multivariate approach to estimate colour producing agents in Case 2 waters using first-derivative spectrophotometer data

Khalid A. Ali [a b], Donna L. Witter [b] & Joseph D. Ortiz [b]

[a] College of Charleston, Department of Geology and Environmental Sciences, Charleston, SC, 29424

[b] Kent State University, Department of Geology, Kent, OH, 44242

Accepted author version posted online: 30 Oct 2012.Published online: 07 Feb 2013.

To cite this article: Khalid A. Ali , Donna L. Witter & Joseph D. Ortiz , Geocarto International (2013): Multivariate approach to estimate colour producing agents in Case 2 waters using first-derivative spectrophotometer data, Geocarto International, DOI: 10.1080/10106049.2012.743601

To link to this article: http://dx.doi.org/10.1080/10106049.2012.743601

Geocarto International, 2013
http://dx.doi.org/10.1080/10106049.2012.743601

Taylor & Francis
Taylor & Francis Group

Multivariate approach to estimate colour producing agents in Case 2 waters using first-derivative spectrophotometer data

Khalid A. Ali[a,b]*, Donna L. Witter[b], and Joseph D. Ortiz[b]

[a]*College of Charleston, Department of Geology and Environmental Sciences, Charleston, SC 29424;* [b]*Kent State University, Department of Geology, Kent, OH 44242*

(*Received 23 May 2012; final version received 22 October 2012*)

The complex composition and distribution of colour producing agents (CPAs) in turbid aquatic environments such as the Western Basin of Lake Erie (WBLE) presents a challenge to the application of remote sensing data for differentiating among in-water constituents and estimating their concentrations independently. In this study, multivariate procedures are applied to lab-based spectrophotometer data to estimate the concentration of chlorophyll-a and suspended matters in the WBLE. Principal Component Analysis of first-derivative transformed hyperspectral data from the spectrophotometer extracted three significant spectral components for each cruise, explaining up to 88% of the spectral variability. Spectral matching using reference spectra indicated that two of the extracted patterns represent signatures of in-water constituents that govern the optical properties of the WBLE, namely, cyanobacteria and diatoms associated with green algae. The spectrophotometer data clearly revealed known spectral features associated with phytoplankton, such as the absorption minima near 550 and 700 nm, which can be attributed to the minimum of absorption and fluorescence of chlorophyll-a, respectively. The method also extracted the absorption peaks due to chlorophyll-a, near 670 nm, and due to phycocyanin, near 620 nm. Principal component regression of chlorophyll-a on the PC scores indicated that 63.4% of variation of chlorophyll-a in the WBLE can be explained by two components. Factors 2 and 3 explain 60% of the joint spatiotemporal variability of suspended matters in the WBLE. The results illustrate the potential of multivariate technique applied to remote sensing data in isolating the patterns that represent constituents in turbid Case 2 waters.

Keywords: chlorophyll-a; remote sensing; Lake Erie; PCA

Introduction

The Great Lakes in the United States are an invaluable natural resource, containing about 20% of the Earth's potable freshwater (Reynolds 1996). Although the smallest by volume, Lake Erie is the most productive of the Great Lakes (Munawar and Weisse 1989). The high productivity of Lake Erie, large population density (*ca.* 16 million people) in its watershed, and large percentage of agricultural land contained within its drainage basin (*ca.* 63%) (USEPA 2006) have resulted in a significant dependence by the local population on Lake Erie for recreational, industrial and potable water needs. One of the most striking issues is the expanded seasonal

*Corresponding author. Email: alika@cofc.edu

depletion of oxygen in the hypolimnion of the central basin that has been termed '*The Dead Zone*' (USEPA, 2009, 2010). An additional issue of concern is the reemergence of toxic cyanobacteria blooms in the Western Basin of Lake Erie (WBLE). These problems are of great concern due to embedded implications for altered biodiversity, ecological services as well as for overall ecosystem health.

Distributions of phytoplankton in space and time have major implications for water quality and ecosystem function. Water resource managers and policy makers require effective water quality monitoring that is critical to address the question of how various natural and anthropogenic factors affect the health of these environments, and to characterize the onset and the temporal and spatial scale of hypoxia and toxic cyanobacteria blooms. Obtaining these observations through in situ methods is challenging for large water systems, such as Lake Erie, where anthropogenic nutrient loads, water levels and global surface temperatures change. Water quality assessments for Lake Erie are largely based on conventional in situ measurements with limited spatial and temporal resolution. This makes it difficult to understand the dynamics of colour producing agents (CPAs), such as phytoplankton, total suspended matter (TSM) and coloured dissolved organic matter (CDOM).

Satellite-based measurements using multi-spectral sensors (MERIS, MODIS and SeaWiFS) have better spatial coverage and temporal resolution to characterize dynamic water quality properties, but require the development of algorithms that relate the spectral reflectance measured by the sensor to the concentrations of CPAs (Doerffer and Fischer 1994, IOCCG 2000, McClain 2009). Ocean satellites have radiometric sensitivities optimized for measurements in turbid waters that normally have low reflectance due to the high absorption by water and its constituents such as CDOM. A potential limiting factor for satellite-based estimates of the concentrations of in-water constituents is the depth of penetration at which the recorded reflectance signal originates. According to the work by Ortiz *et al.* (unpublished results) in the Western Lake Basin of Lake Erie, surface water conditions are representative of the bio-optical properties over a broader range of depth. This makes satellite applications to estimate concentrations of CPAs feasible.

CPAs change the optical properties of water by absorbing and scattering incident light (Gordon and Morel 1983, Bukata *et al.* 1995). Chlorophyll-a concentrations have been considered an important index in assessing water quality for coastal waters in surveillance programs for harmful algal blooms (Glasgow *et al.* 2004). Morel and Prieur (1977) have classified marine waters as Cases 1 and 2 based on optical properties. In Case 1 waters, such as epipelagic ocean environments, phytoplankton pigments govern the optical properties of the water. Many coastal and most inland waters are classified as Case 2 waters, because in addition to phytoplankton, constituents such as dissolved organic matter and suspended matter occur in abundance. The optical properties of such waters are thus governed by multiple components.

Application of satellite remote sensing methodologies to estimate concentrations of individual CPAs requires the development of empirical or semi-empirical algorithms. These algorithms are developed by relating the reflectance measured remotely with the in situ observations of individual CPA concentrations. Ground truthing is achieved by comparing spatially and temporally co-located satellite and in situ observations. Spectral variations in the backscattered flux of Case 1 waters are primarily related to concentration of chlorophyll-a, which varies as a function of the phytoplankton population. In marine environments satisfactory relationships have

been described for chlorophyll-a concentration ranging between 0.02 and 78 mg/m^3 (Gordon *et al*. 1988, Gons 1999, O'Reilly *et al*. 2000). In many coastal waters, the water-leaving irradiance is affected by multiple optically active components. As shown by Witter *et al*. (2009), optical detection of chlorophyll-a in the Case 2 waters of the Western Basin of Lake Erie (WBLE) using Case 1-derived algorithms is problematic. The challenge is attributed to optical heterogeneity of Case 2 waters due to interference from higher concentrations of suspended matter and dissolved organic matter.

One strategy for differentiating the overlapping and correlated in-water constituents is utilizing both the visible and the infrared portions of the spectrum (Gitelson 1992, Gons 1999). The combined effects of absorption due to phytoplankton, suspended matter, dissolved organic matter and pure water produces distinct spectral features in the visible and NIR (near-infrared). The main spectral features of Case 2 waters include absorption maxima at the blue and red wavelengths, and absorption minima at green wavelengths and in the NIR. Several models have been developed to estimate the in-water constituents such as CDOM independently (Binding *et al*. 2008, Becker *et al*. 2009). However, the models do not account for the variability in the characteristics of the CPA with time and space.

The colour of Case 2 waters is influenced by multiple constituents with distinct spectral properties that combine non-linearly. Hence, remote sensing of Case 2 waters is a non-linear multivariate problem, requiring a multivariate approach to discriminate among, and estimate concentrations of, the various in-water constituents. The spectral signature of the WBLE (Figure 2) has similar spectral pattern with the line labelled as 'a' in Figure 1. The spectral discrepancies observed

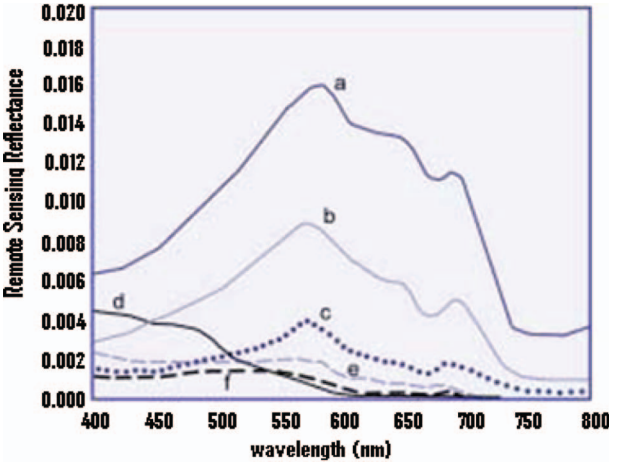

Figure 1. Some examples of remote-sensing reflectance spectra from different types of waters, including waters with (a) very high sediment and CDOM concentrations, (b) high sediment and CDOM concentrations, (c) moderate sediment and CDOM with some phytoplankton, (d) clear water, (e) waters with moderate chlorophyll and sediment concentrations, (f) waters with moderate chlorophyll concentration. Lines labelled as a, b, c and e represent Case 2 waters, while d and f are signatures of Case 1 waters (IOCCG, 20032000).

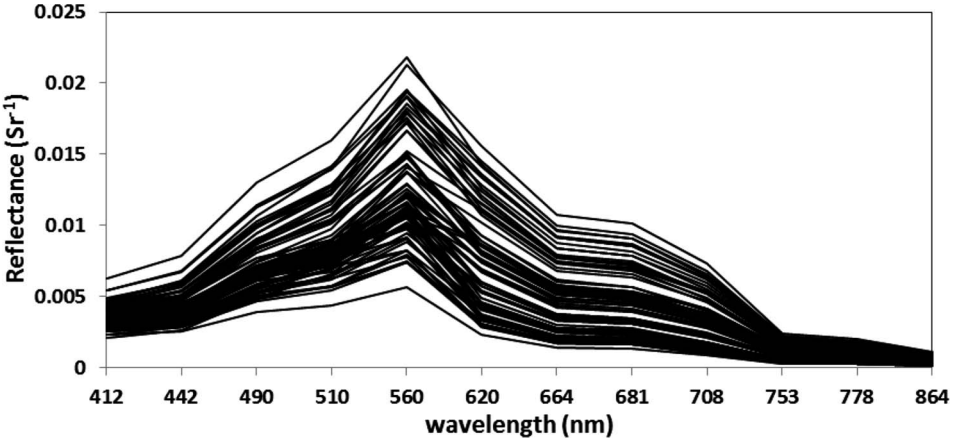

Figure 2. Reflectance spectra (400–860 nm) recorded by the Medium Resolution Imaging Spectrometer (MERIS) in the WBLE at pixels corresponding to 18 sampling stations. Pixel resolution is 290 by 290 m. The spectra shows typical signatures of Case 2 waters with the high absorption of CDOM in the blue region (440 nm), chlorophyll *a* in the red region (665 nm, red absorption) and phycocyanin in the yellow region (620 nm). The peak centred at 550 nm represents effect of backscattering from suspended matter.

from the reflectance spectra of the Medium Resolution Imaging Spectrometer (MERIS) clearly indicate the presence of multiple biotic and inorganic components in the WBLE (Figure 2). Based on this comparison, reflectance patterns from the WBLE are consistent with those measured in Case 2 water.

A powerful multivariate statistical tool that can be applied in water quality studies that have hyper-spectral data is Principal Component Analysis (PCA). PCA seeks to decompose a data matrix into uncorrelated components by finding the eigenvalues and principal components of the matrix, which is identical to solving a system of linear equations. The principal components represent linear combinations of the original variables that are uncorrelated to each other. This method maximizes the variances of leading factors and reduces invalid factors in the matrix data, thus reducing the complexity of multidimensionality of the matrix systems (Kaiser 1958). Because PCA amounts to a series of matrix operations, it can be applied to extract uncorrelated variance between variables, or in this case to extract relevant information that explains the optical complexity of Case 2 waters. Decomposition of the hyper-spectral remote sensing data measured on the Lake Erie filter samples can be used to identify important water quality indicators such as chlorophyll-a concentration, and estimate spatial and temporal variations across the WBLE (Ortiz *et al*., unpublished results). The multivariate approach is also used to estimate TSM in the WBLE.

In this study, spectral signatures of various in-water constituents that contribute to the optical complexity of the WBLE are extracted using multivariate statistics. The main objectives of the study are to (a) analyse the efficiency of remote sensing data to detect the various CPAs in the WBLE and (b) examine the effectiveness of multivariate regression techniques for predicting concentrations of chlorophyll-a and TSM using remote sensing reflectance data.

Data and methods

Study area

The Western Basin of Lake Erie (83 to 82.5 W and 41.2 to 41.7 N, Figure 3), with an average depth of 7 m, is shallow enough to have dramatic wind- and wave-driven turbidity. Its relatively warm temperature makes it conducive to high biologic productivity. In our study area, a number of rivers serve as conduits for fluxes of nutrients and sediments into the WBLE, influencing water clarity, particularly near the mouths of the Sandusky and Portage rivers (Figure 3). In recent years, oxygen depletion and the extent of both harmful and nuisance algal blooms have increased in Lake Erie, despite a general decline in eutrophication (Budd *et al.* 2002, Rinta-Kanto *et al.* 2005). Because Lake Erie has dynamic nutrient and trophic interactions, satellites can enhance monitoring efforts. The current generation of imaging satellite sensors can synoptically monitor large areas at relatively high spatial resolution with repeat sampling intervals ranging from days to weeks, depending on the cloud cover.

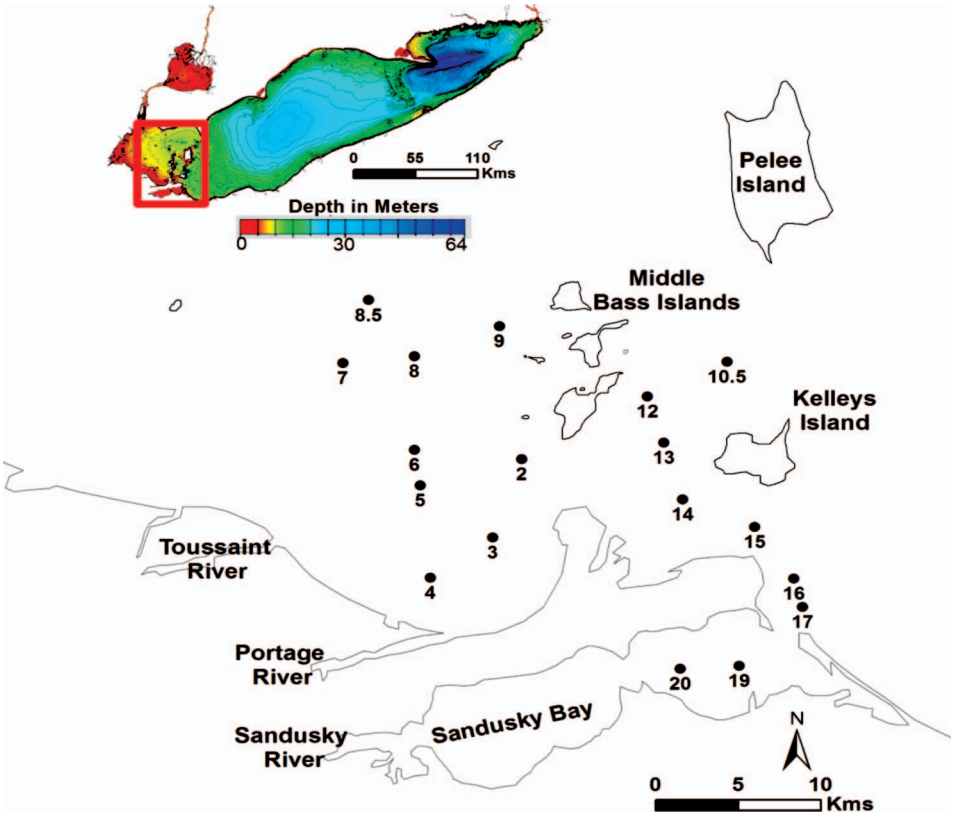

Figure 3. The Western Basin of Lake Erie and the locations of 18 sampling stations visited during multiple cruises of Research Vessels Gibraltar and Erie. The lake interacts with terrestrial environment through fluxes of the Sandusky, Portage, Toussaint and other rivers such as the Maumee and Detroit, which drain into the far western side of the basin (data not shown). The inset shows, the Lake Erie bathymetry.

Data acquisition

In the summers of 2009 and 2010, in situ data and water samples were collected from the Western Basin of Lake Erie aboard the Research Vessels Gibraltar and Lake Erie Monitor during cruises in June, July, August and September. A total of 108 stations were visited between the two summer periods. Stations were located between Sandusky Bay and Middle Bass Islands (Figure 3).

Water quality parameters such as chlorophyll-a, oxygen, pH, total dissolved solids, temperature and electrical conductivity were measured using a HACH Hydrolab along a vertical profile at 0.1 m resolution between 0 and 1 m; measurements were taken at 0.5 m resolution below 1 m depth. For this study, the in situ chlorophyll-a data were integrated to 0.5 m depth. Secchi depth was also recorded at each station using a standard Secchi disk with alternating black and white quadrants. A depth integrated 1-litre water samples were collected at each station and filtered through pre-weighed 0.47 μm Glass Fibre Filters (GF/F) to collect the lake water residue. The samples were wrapped with aluminium foil to inhibit radiation interaction and stored in a cooler with ice in the dark. Dry particulate mass was measured gravimetrically with accuracy of 0.1 mg to estimate the TSM. The filtered samples were placed below a high intensity 20 mm diameter Spectralon integrating sphere and reflectance spectra were recorded between 250 and 2500 nm using ASD lab-based Pro Full-Resolution ultraviolet/visible/near-infrared (UV/VIS/NIR) spectroradiometer. Spectra were measured at 2 nm intervals in the UV-VIS range and 4–10 nm intervals in the NIR range. A Spectralon reference panel was used for white referencing. Percent reflectance was transformed to Log (1/Reflectance) units. To enhance signal to noise ratios, the records for each sample were obtained by averaging 800 individual measurements of the spectrum. The absorption values were corrected for the GF/F filter background using averaged measurements of the blank filters.

Methodology

Principal component analysis is a multivariate procedure that involves the transformation of a number of possibly correlated variables into a smaller number of uncorrelated indices (eigenvectors). In this remote sensing application, PCA accounts for the correlation between the signals in different spectral channels and enhances the potential discrimination and reconstruction accuracy of retrieved constituents. For this study, the data matrix composed of m rows, where m is the number of stations sampled (18 stations per cruise, total of five cruises) and n columns, where n is the number of spectral bands (64 bands ranging from 360 to 1000 nm). The Principal Components (PCs) returned by PCA represent combinations of co-varying data. These are ordered based on the percent of total variance that they explain. In most applications, a small number of PCs explain the bulk of the total variance, with remaining PCs representing noise. Each PC is a linear combination of the original variables, potentially describing a different source of variation. The largest or 1st PC is oriented in the direction of the largest variation of the original variables and passes through the centre of the data. PCA is commonly used in environmental applications, including surface and ground water quality studies (Yu *et al*. 1998, Helena *et al*. 2000, Tauler *et al*. 2000, Petersen *et al*. 2001, Parinet *et al*. 2004, Ying Ouyang 2005). Works by Balsam and Deaton (1991), Deaton and Balsam (1991), Mix *et al*. (1995) and (1999) and Harris and Mix (1999)

have demonstrated the utility of this method for extracting information regarding variations in sediment composition both spatially and temporally. The method works equally well with CPAs, allowing isolation of spectral patterns that are correlated in space and time across each cruise track. In many cases, the spectral patterns identified by PCA can be interpreted physically based on known spectral signatures (Ortiz *et al*. 1999, 2004, 2009).

The reflectance spectra are sensitive to background noise that may alter the shape of reflectance spectra. These effects can be minimized by calculating the first derivative of the reflectance spectra (Ortiz *et al*. 2009, Ortiz 2011). We conducted PCA on the correlation matrix obtained from a data matrix of first-derivative-transformed Visible-Infrared (VIR) spectra. The columns of each data matrix explicitly define the 1st derivative transform represent wavelength in 10 nm increments and the rows represent a station in one of the cruise tracks. Data from all of the cruises were combined into a single matrix to allow evaluation of the joint variation in space and time within the dataset, the data matrix contained a total of 89 rows. Data from one station for a particular cruise were removed due to instrument errors. PCA was also conducted for each cruise date individually to evaluate the temporal dynamics of the leading optical components.

A varimax rotation was applied to the component matrix extracted from the correlation matrix (Kaiser 1960). This allows sufficient orthogonality between the derived component axes to enable spectrally distinguishing among several independent in-water constituents. Use of the correlation matrix weights each band equally within the analysis since the correlation coefficient is the cross product of the z-scores of the two bands. In this study, only factors exhibiting an eigenvalue of over 1 were retained (Kaiser 1960). Prior to conducting the PCA calculation, a standard normal variate transformation was applied to the data to remove a bias due to concentration differences between stations. The trend due to concentration is removed across the wavelength (λ) using the following mathematical expression (Barnes *et al*. 1989):

$$SNV_{(\lambda)} = \frac{(y_{(\lambda)} - \bar{y})}{\sqrt{\frac{\sum (y_{(\lambda)} - \bar{y})^2}{n-1}}} \tag{1}$$

In remote sensing application, the ability of PCA to discriminate among the multiple in-water constituents is attributed to its capability to resolve for the satellite signal variability caused by the constituents including interference from the atmosphere. By truncating the noise factors, the atmospheric interference and stochastic errors, PCA reduces the dimensionality of the dataset to that of the significant or leading factors that represent only signals caused by in-water constituents. Theoretically, some components of atmosphere interference may come out as a leading mode of PCA. However, in practice, this does not happen because the atmospheric correction procedures applied during pre-processing of the data remove much of the systematic atmospheric interference (Doerffer and Schiller 2008, Witter *et al*. 2009). The communality of the PCA identifies the variance in each input variable explained by the varimax-rotated factors extracted from the data set.

To identify the origin of the factors, the resulting factor loading patterns were compared with the centre-weighted derivatives of plant pigment reflectance spectra for algal groups using spectra from Moberg *et al*. (2002) and Toepel *et al*. (2005) and mineral diffuse spectral reflectance signatures from known standards measured in

our laboratory or available from version 5 of the USGS Digital Spectral Library (Clark *et al*. 2003). This is analogous to X-ray diffraction methods that use peak or whole spectrum matching methods to compare samples against known standards (Will 2006).

Results and discussions

Figure 4 represents the reflectance spectra recorded between 400 and 2500 nm using a laboratory-based spectrophotometer. The reflectance values are highly variable in the visible and NIR range. The blue-green spectral range shows reflectance patterns typically observed in turbid waters (Gitelson *et al*. 2000, Schalles *et al*. 2002, Dall'Olmo and Gitelson 2005). These patterns reflect the effects of multiple factors including absorption by phytoplankton, CDOM, and scattering due to SM. Reflectance peaks near the red/NIR edge and absorption peaks near 620 and 670 nm are apparent; these peaks confirm the presence of phytoplankton biomass. The red/NIR peaks were much higher than the green peaks for chlorophyll-a concentrations >1.5 μg/l. The absorption minima near 550 nm and the maxima

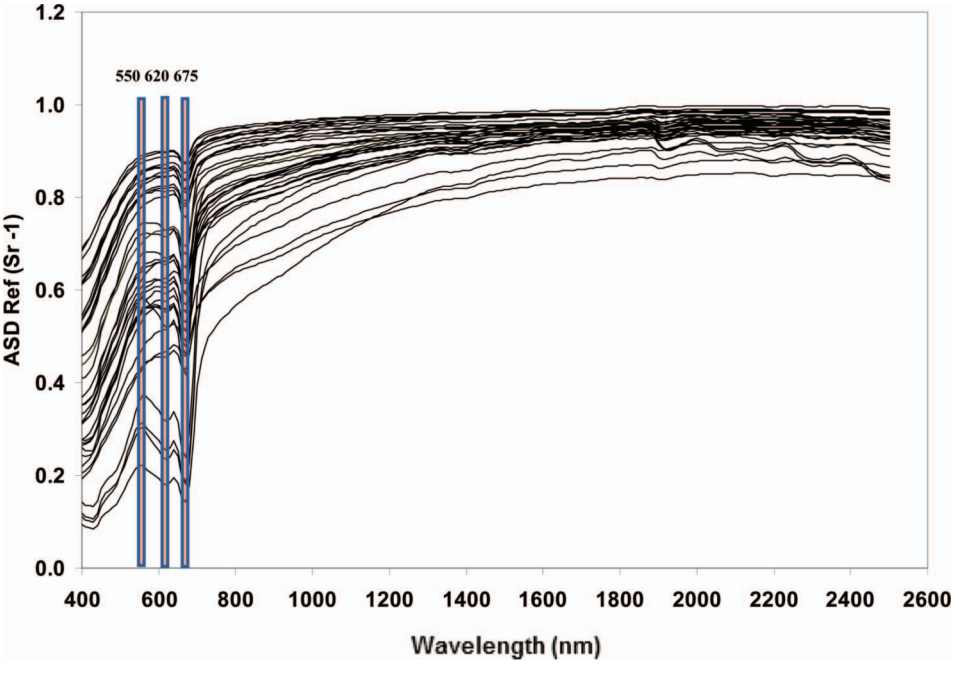

Figure 4. Reflectance spectra (400–2500 nm) generated using laboratory-based ASD spectrophotometer for samples collected during the four research expeditions. The absorption effects of CDOM, chlorophyll *a* and phycocyanin (PC) are observed in the visible range (400–700 nm). The trough near 440 nm is due to high absorption due to CDOM and phytoplankton reflects strong absorption due to CDOM and phytoplankton, toughs at 620 and 675 nm represent the presence of PC and chlorophyll *a* pigment. High reflectance values near 550 nm and beyond 720 nm occur due to the combined effects of backscattering from suspended material and the minimum absorption by plant pigments. The three line markers indicate the green peak near 550 nm, the 620 nm absorption due to PC, and the red absorption near 675 nm due to chlorophyll *a*.

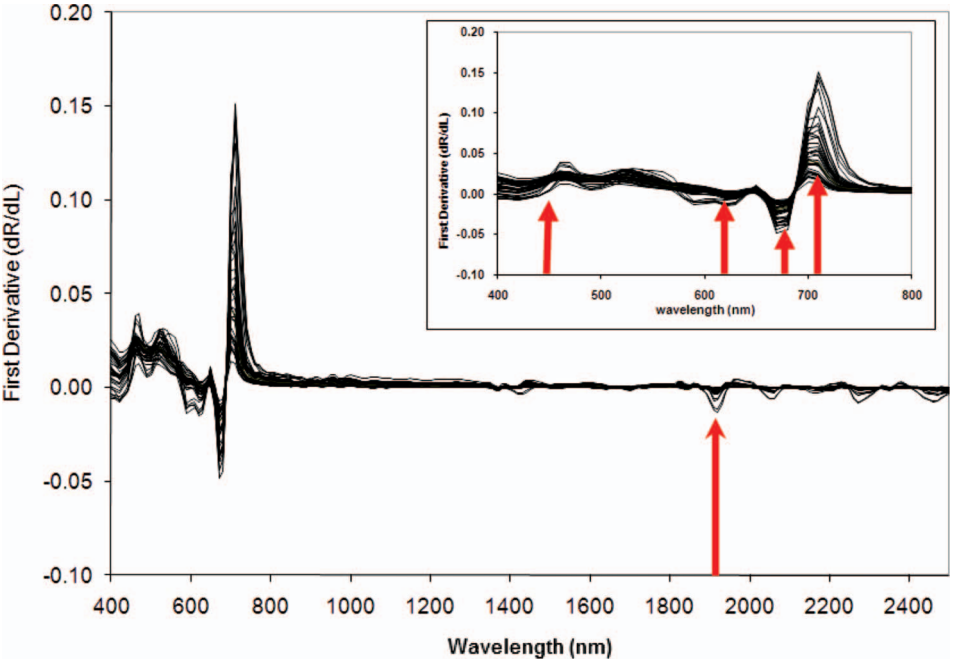

Figure 5. Centre-weighted first derivative of the reflectance spectra in Figure 2.2, the derivative removes and accentuates absorption features in the visible and NIR. The NIR absorption features between 1400 and 2500 nm correlate well with suspended sediment. The most prominent of these is at 1910 nm. The inset to Figure 3.5 illustrates prominent features in the visible range, including absorption troughs at 440, 620 and 675 nm, and a fluorescence peak near 710 nm.

near 670 nm were more distinct for samples representing stations closer to Sandusky Bay.

The centre-weighted first derivative of the reflectance spectra minimizes background noise and accentuates absorption features in the visible and NIR. This reveals troughs at 410 nm, peak at 460 nm, a broad peak centred at 550 nm, a trough at 672 nm and a peak at 710 nm in the visible range (Figure 5, inset). Hydroxyl absorption is present at 1400 and 1910 nm indicating the presence of inorganic suspended sediments (Figure 5) (Ortiz *et al.* unpublished results, Ortiz *et al.* 2009).

The ability of chlorophyll-a to fluoresce is the basis for the spectrophotometer capable of measuring the analyte in situ. In the WBLE, the average in situ concentration of chlorophyll-a was 5.68 μg/l in June 2009 and concentrations during September period ranged between 4.96 and 9.64 μg/l, with an average concentration of 8.07 μg/l. Stations 19 and 20 contributed to the high variability during the September 2009 cruise and appeared as outliers in the dataset. Average concentrations on July 8th and 27th of 2010 were 6.05 and 7.45 μg/l, respectively. In September of 2010, average chlorophyll-a concentration was 7.23 μg/l (Figure 6(a)).

The standard deviation of the chlorophyll-a concentration decreases from early summer period to late summer period for both 2009 and 2010 (Table 1). Relatively high chlorophyll-a concentrations were recorded at stations, 17, 19 and 20. These stations represent the waters of Sandusky Bay which is heavily influenced by

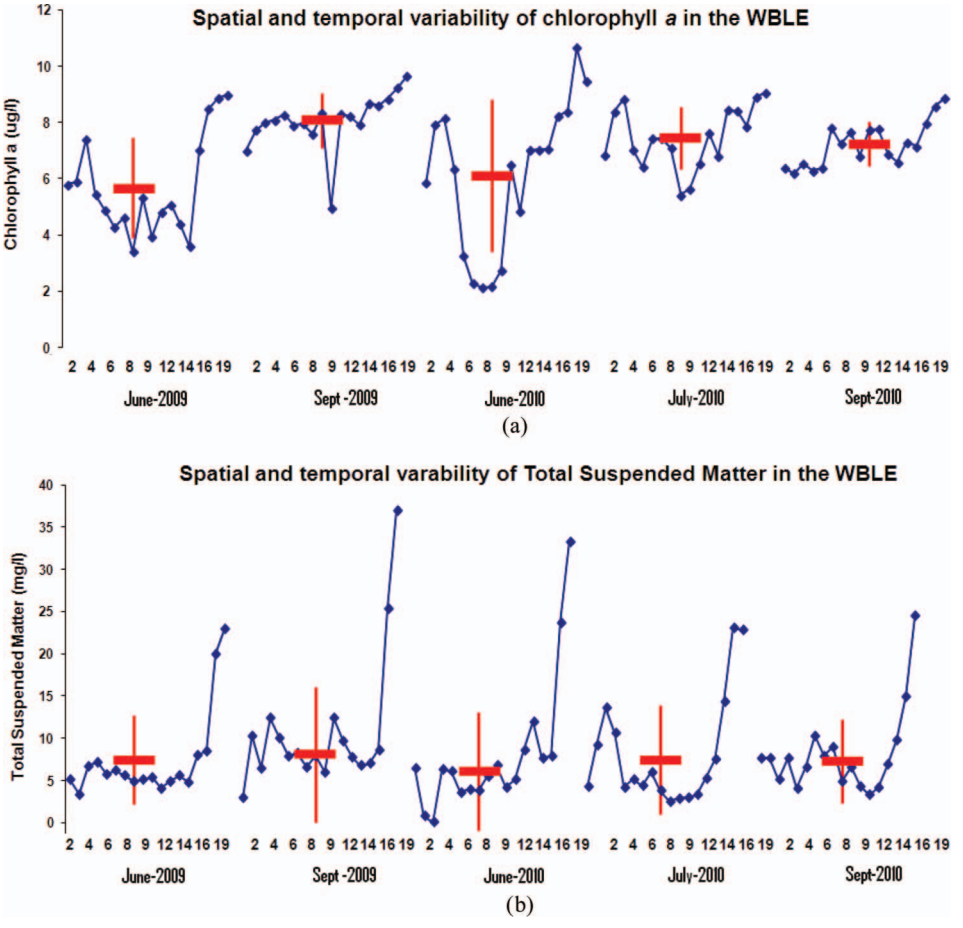

Figure 6. Plots of concentrations of water constituents along the cruise track: (a) chlorophyll *a* concentrations in μg/l at each station and (b) gravimetrically estimated TSM in mg/l at each sampling point. Relatively higher concentrations of the CPAs are recorded in the Sandusky Bay. Red horizontal bars represent average concentrations and the red vertical lines are standard variation of the concentrations.

Table 1. Statistics of chlorophyll *a* and TSM in the WBLE.

Cruise dates	Average chlorophyll *a* (μg/l)	Standard deviation	Average s TSM (mg/l)	Standard deviation
24 June 2009	5.68	1.76	7.46	5.28
2 September 2009	8.07	0.98	10.75	8.02
8 July 2010	6.05	2.716	8.1	8.07
28 July 2010	7.45	1.08	8.12	6.45
13 September 2010	7.23	0.79	8.08	4.96

Note: A low concentration of chlorophyll *a* is obtained during early summer and the concentration increases throughout summer period. Similar pattern is observed for TSM.

terrestrial influx via the Sandusky River that loads significant nutrients. The high standard deviation in chlorophyll-a during the early summer period reflects the optical difference between the waters of Sandusky Bay and the central WBLE. During the summer, prolonged periods of strong sunlight and the continuous input of nutrients allows expansion of algal blooms into the central WBLE (Becker *et al.* 2009). Moreover, turbulent mixing between the terrestrially influenced Sandusky Bay and the central WBLE water, coupled with wind-induced mixing homogenizes the water resulting in lower standard variation in the concentrations of the in-water constituents across the basin. The spatial and temporal variability of the concentrations of TSM in the WBLE is similar to that of chlorophyll-a. Higher than average values of TSM are recorded in early summer at the stations corresponding to Sandusky Bay relative to the stations representing the central WBLE, indicating difference in the optical properties between the WBLE and Sandusky Bay. The standard deviation of TSM among the sampling stations decreases over the summer period (Figure 6(b)). This signals lake circulation that involved material advecting and mixing throughout the WBLE.

PCA analysis of the laboratory-based VIR derivative spectra

The centre-weighed first-derivative spectra highlight fine-scale differences in the reflectance spectra, particularly between 600 and 700 nm (Figure 5). Application of

Table 2. Descriptive statistics of selected principal components.

	Percentage of variance after Varimax rotation									
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Variability (%)	42.4	31.1	14.3	7.5	2.8	0.7	0.4	0.2	0.2	0.1
Cumulative (%)	42.4	73.5	87.8	95.3	98.1	98.8	99.2	99.4	99.6	99.8

Note: This represents the joint spatial and temporal optically variability across the WBLE during summer of 2009 and 2010. The first four principal components represent approximately 95% of the optical variability observed during the two summer periods.

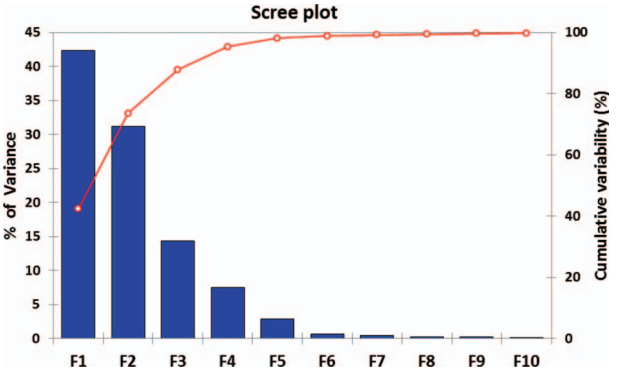

Figure 7. The 'scree plot' represents eigenvalues scaled as the percentage of variance explained by each factors. The first three factors (F1, F2 and F3) explain about 88% of optical variability. Factors beyond the F3 represent background noise and instrumental error, and therefore can be discarded reducing the dimensionality of the dataset.

varimax-rotated PCA to the derivative-transformed reflectance spectra for the 2009 and 2010 cruises produced three leading principal components with eigenvalues greater than one. The leading factors are associated with their empirical orthogonal factors that explain most of the joint spatiotemporal variability observed in the WBLE (42.4, 31.1 and 14.3%, respectively) (Table 2). The number of significant factors can also be inferred from the 'scree plot' (Figure 7). PCs with eigenvalues less than one did not give interpretable factors and therefore were not considered in this study. The factor loading values sharply decrease within the first three principal vectors and then slowly stabilize for the remaining ones which may contain a great deal of interference stochastic error, and therefore are discarded. This procedure contributes to dimensionality reduction in the data matrix while preserving relationships that exists in the data. The presence of multiple factors with high percentage of signal variability suggests that the optical variability observed in the

Table 3. Shows the PCAs with leading PC patterns, the corresponding factor loadings, which represent the largest amount of variance extracted by each factor and the variance percentages for each principal component.

PCA results on first-derivative hyperspectral data			
Factors with significant loads within each Principal components PCA with 35 centercentre weighed first derivative spectrophotometer bands Rotation sum of square loadings			
PCs	Bands (nm)	Factor Loadings	Total Variance (%)
20.26	400	0.759	42.4
	410	0.830	
	420	0.862	
	430	0.893	
	440	0.614	
	570	0.797	
	580	0.814	
	590	0.842	
	600	0.838	
	610	0.662	
	620	0.517	
	640	0.521	
	660	0.421	
	670	0.511	
	690	0.580	
	700	0.631	
6.28	450	0.889	31.1
	460	0.819	
	470	0.935	
	480	0.901	
	490	0.892	
	500	0.726	
	510	0.574	
	520	0.459	
	680	0.418	
1.96	530	0.338	14.3
	540	0.464	
	550	0.557	
	560	0.725	

Western Basin of Lake Erie is attributed to multiple, independent constituents described by the various orthogonal vectors, further confirming the categorical classification the WBLE, as Case 2 water type.

Table 3 shows the results of the general and specific PCs with the leading factor loadings, the corresponding eigenvalues, which are the amount of variance, extracted by each factor, and the variance percentages (accounted for) corresponding to the principal components. The factor loadings indicate the significant bands contributing to each PC. The loadings express the correlation between the hyper-spectral bands and the newly formed PCs. The communality, which is the proportion of variation in each hyper-spectral band explained by the PCs, is greater than 0.85 for all the bands, indicating that most of the optical variability detected by each band is explained by the PC model. The total communality is 27.2 of 30 and the proportion of the total optical variation explained by the three factors is 88%.

Plots of the first three factor loadings (PCs) as a function of wavelength indicate the spectral patterns of the major in-water constituents that governed the optical variability across the WBLE during the summer of 2009 and 2010. The plot clearly indicates the critical bands contributing for each factor (Figure 8). Factor 1 is heavily influenced by the blue (400–460 nm), green (560–640 nm) and red region (670–680 nm). The second factor represents the variance contribution from the signals in the region between 450 and 560 nm, high loading to this factor also comes from 680 nm. Maximum loading for the third factor mainly comes from the green region (520–560 nm).

Spectral matching of the derived eigenspectra as a function of wavelength with derivative-transformed reflectance spectra for known classes of in-water constituents provides a means of identifying the eigenvectors. This approach indicates that the first principal vector represents the cyanobacteria population (Figure 9(a)). The second principal vector, which describes 31.1% of the optical variability observed in the WBLE, corresponds to diatoms associated with cyanobacteria signalling the absorption features due to phycobilins (Figure 9(b)). The third factor represents a more complex mixture of phytoplankton biomass and oxy hydroxides (Figure 9(c)). The importance between the first two principal vectors flipped when performing

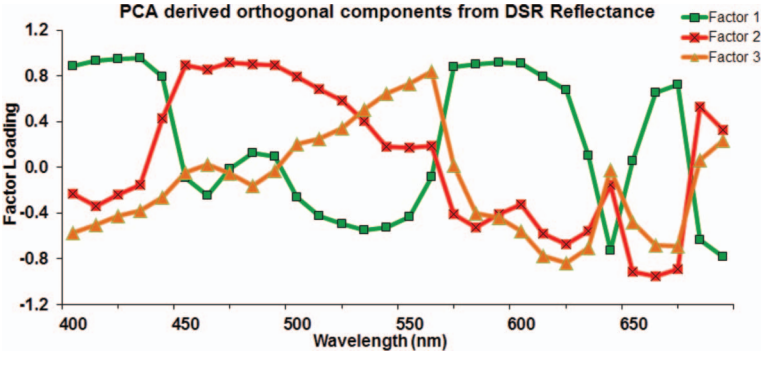

Figure 8. Plots of principal factors as function of wavelength. The three eigenvectors account for approximately 87% of the optical variability observed in the WBLE during the summer season of 2009 and 2010. These plots represent the first-derivative spectral signature of the various water constituents.

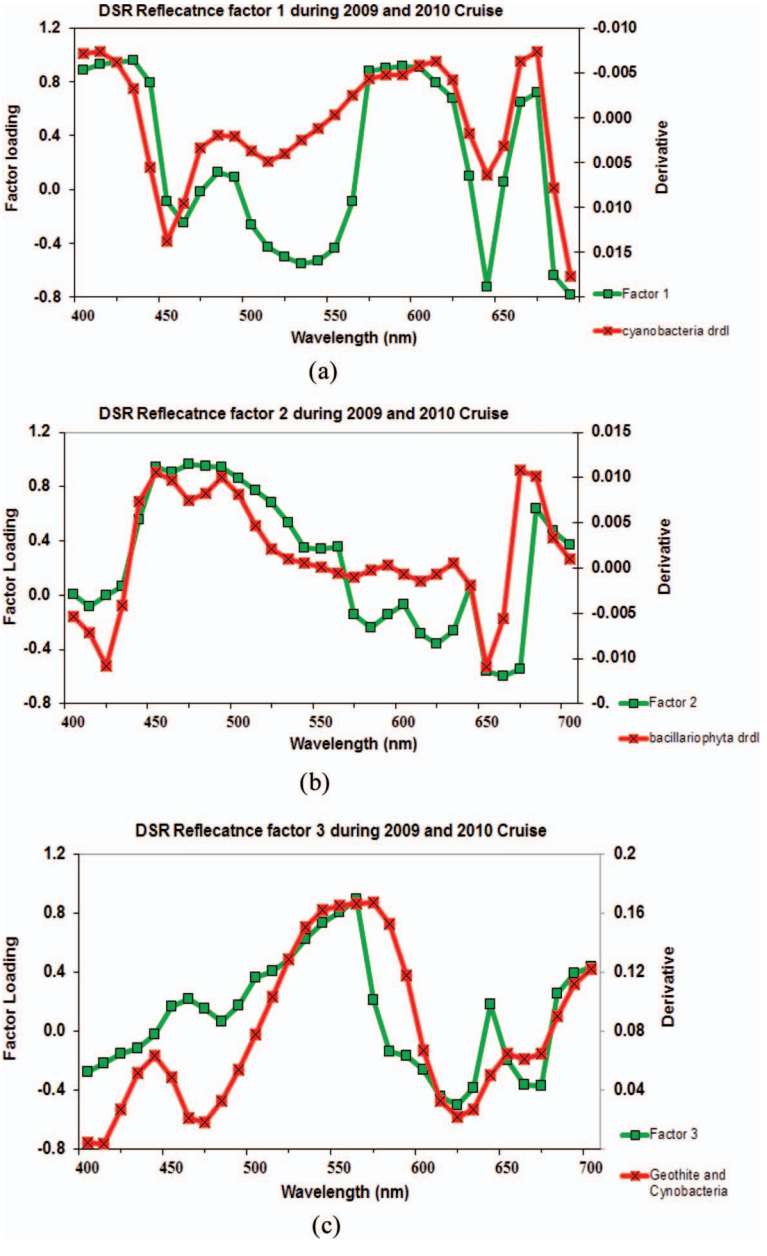

Figure 9. Factor loadings for the PCA and selected reference derivative spectra: (a) the first factor relates to cyanobacteria, (b) the second factor relates to diatoms and cyanobacteria, (c) the third factor relates to a mixture of cyanobacteria and goethite, an iron oxy-hydroxide component of the suspended sediment.

factor analysis excluding samples from Sandusky Bay. This indicates that the Sandusky Bay is optically different from the open water of the WBLE. These results are consistent with those obtained from previous studies in the WBLE (Ortiz *et al.*, unpublished results).

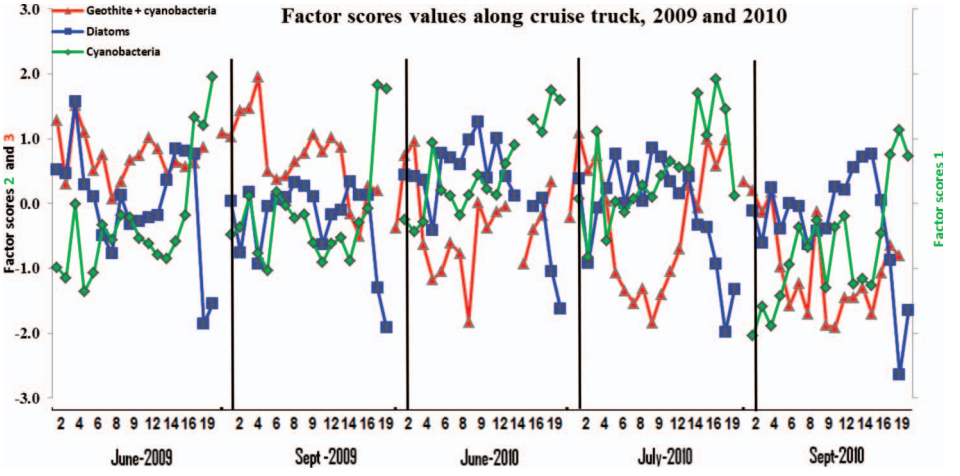

Figure 10. Factors scores from the PCA model. The importance of each component at each station is proportional to its magnitude; the sample locations are as defined in Figure 2.3. The data are presented in chronologic order with the earliest cruise on the left.

Spatial variations of the CPAs

A plot of the factor scores, representing each vector space, against station label indicates the station's contribution towards each principal factor. This in turn indicates the dominant type of CPAs along the cruise track (Figure 10).

The first component, cyanobacteria, is dominant in Sandusky Bay, and its significance decreases towards the stations in the central WBLE. This suggests a heavy presence of riverine algae discharged into the Sandusky Bay. Additionally, nutrient loading from the rivers supplies the required nutrients for cyanobacteria, increasing lake productivity and causing algal blooms (Becker *et al*. 2009, Budd *et al*. 2002). In the central WBLE, relatively higher factor 1 scores were computed for Stations 3 and 4, both are stations closer to the outlets of the Toussaint and Portage Rivers, further suggesting the influence of terrestrially derived algal components in the lake water quality. The concentration of the cyanobacteria increases between June and July. Factor score values for the cyanobacteria are low during late summer or early fall period. The relatively lower September temperatures coupled with reduced amount of nutrients in the water inhibit algal growth leading to shrinkage of the cyanobacteria community. Water temperatures in July were approximately 25°C and by late summer decreased to 15°C (NOAA, http://www.erh.noaa.gov/buf/laketemps/laketemps.php). In 2009, during summer, input concentrations of phosphorus and nitrate via the Maumee river plume dropped from an average of 0.092 and 10.15 mg/l, respectively, to 0.01 and 0.01 mg/l in late summer, respectively (NCWQR, http://www.heidelberg.edu/academiclife/distinctive/ncwqr). The second factor represents the variations of the diatoms associated with blue-green algae. Factor 2 scores are generally lower in Sandusky Bay than in the central WBLE, suggesting that these are more dominant members of the community in the central WBLE.

The third component represents a complex mixture of iron-bearing oxide minerals and blue-green algae and shows a general increase in its importance from the central WBLE to the Sandusky Bay. Sandusky Bay and stations 3 and 4

represent aquatic environments that are heavily influenced by terrestrial influx as a result of their close proximity to river discharge zones (Figure 3). These environments showed higher values in the factor 3 scores which suggest that the third PCA component represents optical variability attributed to the terrestrially derived suspended matters. Factor 3 shows a general decreasing trend over the summer period which may be associated with the decrease in discharge rates of the Sandusky River during the same period. Sandusky River discharge decreased from approximately 28.3 m^3/sec in early spring of 2010 to 1.1 m^3/sec in late summer of 2010 (USGS, http://waterdata.usgs.gov). The relatively higher factor scores in the Sandusky Bay for July 2010 cruise data may be associated with high discharge event in the days preceding the cruise.

Temporal variability of the CPAs

In order to assess the temporal dynamics of CPAs, the spectral reflectance matrix data were divided into specific cruise periods representing early and late summer. In 2009, cruises were conducted in late June and early September. In 2010, cruises were conducted between mid-July and early September. Varimax-principal component analyses were then performed for each of the four cruise periods. Factor loadings are plotted as a function of wavelength, indicating the spectral pattern of the principal in-water constituents that governed the optical variability across the WBLE during these specific periods. As discussed earlier, spectral matching was determined based on the reference library spectra (Figures 11 and 12).

PCA of the 2009 first-derivative transformed reflectance data indicates that maximum variance occurs due to the presence of blue green algae (Figure 11(a)). Factor 2 represents the diatoms which form the phytoplankton community in the central WBLE (Figure 11(b)). The third factor indicates optically complex signal that relates mixture of terrigenous oxide minerals and riverine cyanobacteria (Figure 11(c)). The third PC component for September of 2009 lack signatures of inorganic sediments (Figure 11(f)). This is possibly the result of the low discharge rate of the Sandusky River during September of 2009 (<2.8 m^3/sec, http://waterdata.usgs.gov) leading to low volume of material influx, and the settlement of the suspended clays to the benthic environment.

Close observation of Figures 11(a) and (d) shows that optical properties of the WBLE changed throughout summer of 2009. In early summer, the optical property of the WBLE was primarily controlled by cyanobacteria; however, by late summer (September) goethite became equally important and showed up in PC one. The magnification of clay mineral signal is primarily due to the significantly lower concentration of phytoplankton in the water column during the late summer or early fall period of 2009. Varimax-PC analysis of the 2010 optical data gave eigenvectors similar to the 2009 data. Spectral patterns for 2010 (Figures 12(a) and (d)) show that cyanobacteria remained as the primary optical constituent throughout the summer period. Records of 2010 Sandusky River discharge show that the discharge was significantly high compared to 2009 and fluxes of high discharge rates remained until August of 2010. Higher and persistent river flux allows for continuous loading of riverine algae community and nutrients, increasing the productivity of the lake and hence phytoplankton play major role in defining the optical properties of the WBLE.

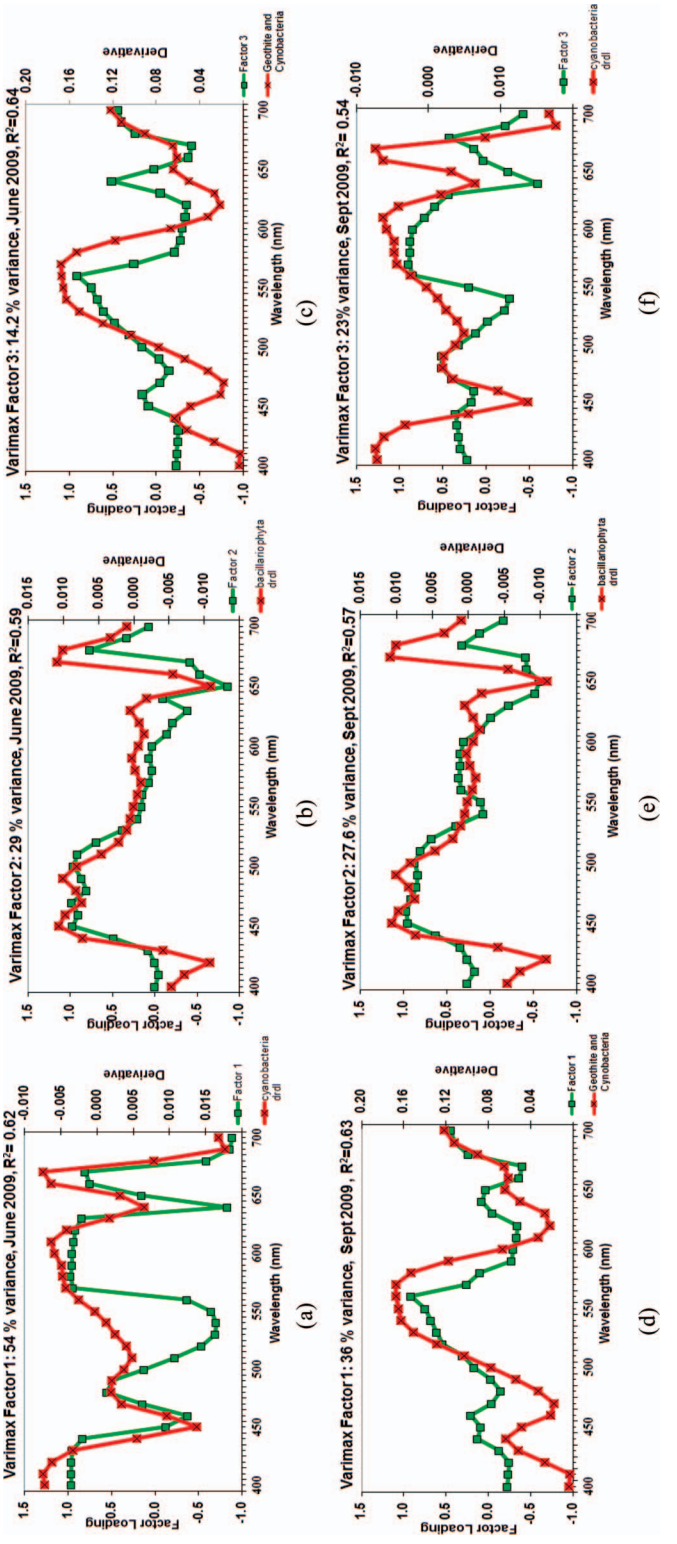

Figure 11. Factor loadings and selected reference derivative spectra for 2009. The first factor relates mainly to cyanobacteria (a, d), the second factor relates to diatoms (b, e) and the third factor relates to a mixture of cyanobacteria and goethite (c), an iron oxy-hydroxide component of the suspended sediment. For September of 2009 period, the first factor represents a mixture of cyanobacteria and goethite (d) and the third factor represents predominantly blue-green algae (f).

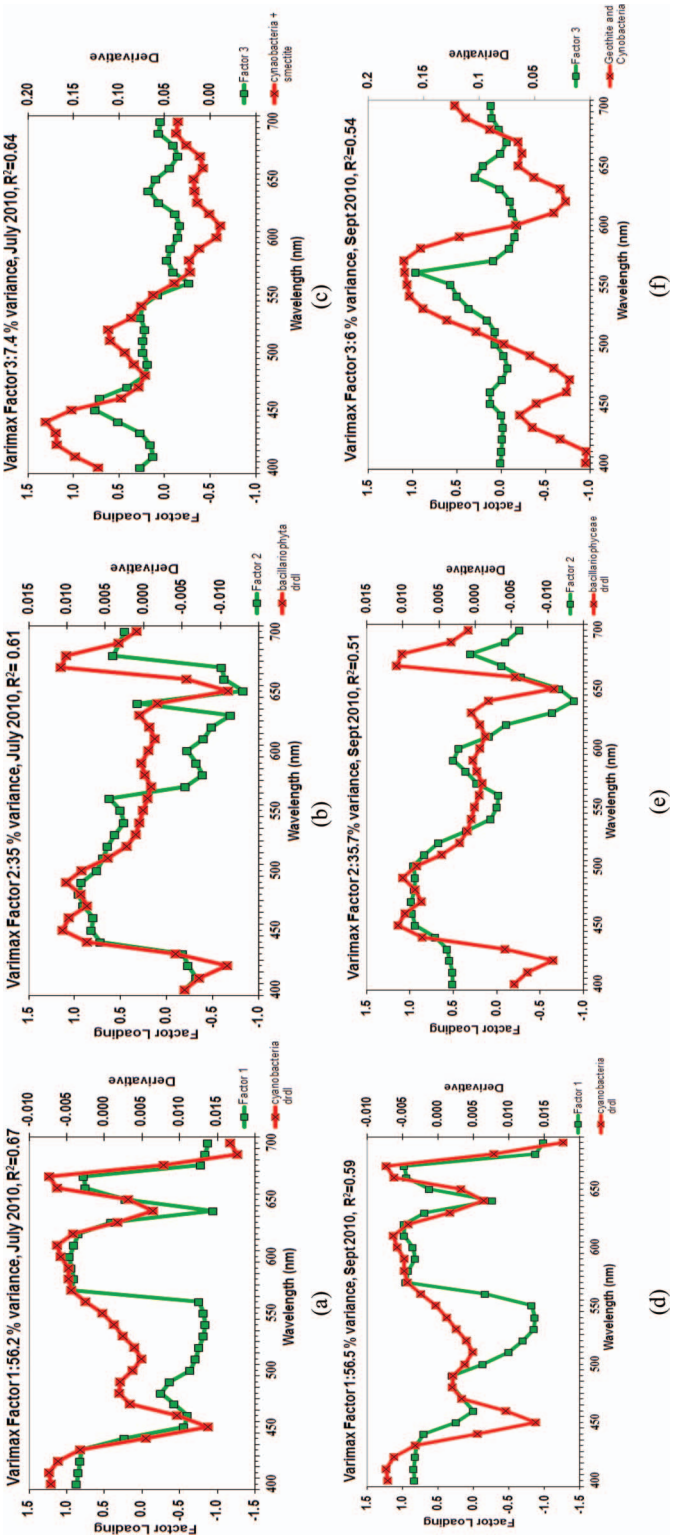

Figure 12. Factor loadings and selected reference derivative spectra for 2010. The first factor relates to cyanobacteria (a, d), the second factor relates to diatoms (b, e) and the third factor relates to a mixture of cyanobacteria and inorganic matter (c), an iron oxy-hydroxide component of the suspended sediment. The third factor for the September period represents predominantly algal components but signatures of iron bearing minerals are observed from the peak near the 550 nm (f).

Prediction of CPAs

One of the major strengths of performing principal component regression (PCR) as oppose to using least-squares methods is it removes the effects of collinearity that may exist between the spectral bands. An important index to justify the removal of collinearity effects between the PCA-based uncorrelated variables, i.e., the factor scores, is their correlation. Table 4 shows that the correlation coefficients between all

Table 4. Correlation analysis between factor scores.

	Correlations coefficients		
	Factor 1	Factor 2	Factor 3
Factor 1	1.00		
Factor 2	0.00	1.00	
Factor 3	0.00	0.00	1.00

Note: Dependent variable: chlorophyll a.

Table 5. Collinearity indices of three standardized principal component regression models for chlorophyll *a* prediction.

				Variance Proportions			
Model	Dimension	Eigenvalue	Condition Index	(Constant)	Factor 3	Factor 1	Factor 2
1	1	1.05	1.00	0.47	0.47		
	2	0.95	1.05	0.53	0.53		
2	1	1.19	1.00	0.00	0.41	0.39	
	2	1.02	1.08	0.89	0.03	0.05	
	3	0.78	1.24	0.11	0.56	0.55	
3	1	1.57	1.00	0.00	0.10	0.04	0.16
	2	1.19	1.15	0.00	0.16	0.31	0.00
	3	1.02	1.24	0.89	0.01	0.01	0.00

Note: Dependent variable: TSM.

Table 6. Collinearity indices of three standardized principal component regression models for TSM prediction.

				Variance proportions			
Model	Dimension	Eigenvalue	Condition index	(Constant)	Factor 3	Factor 2	Factor 1
1	1	1.04	1.02	0.50	0.50		
	2	1.02	0.95	0.50	0.50		
2	1	1.20	1.00	0.14	0.40	0.47	
	2	1.01	1.10	0.81	0.18	0.01	
	3	0.86	1.23	0.06	0.42	0.52	
3	1	1.14	1.00	0.14	0.39	0.46	0.00
	2	1.24	1.12	0.21	0.05	0.02	0.72
	3	1.02	1.21	0.59	0.15	0.00	0.27

Note: Dependent variable: chlorophyll a.

PCs are 0. PC regression models used for the prediction of chlorophyll-a and TSM have eigenvalues and condition indices close to 1 (Tables 5 and 6). All of these collinearity statistics suggest that all principal components are uncorrelated of each other. A general equation that shows the form of the regression model that is used is given by:

$$\widehat{Y}_i = \alpha \widehat{Y}_i = \sum_i + \beta_i F_i \tag{2}$$

where, $\hat{Y}_i$ is predicted value, α is a constant, β_i is the regression coefficients and F_i is factors.

PCR against chlorophyll-a concentrations

The score values obtained from PCA are used as uncorrelated variables in the stepwise multiple linear regression analysis to determine the most significant PCs for predicting the concentrations of chlorophyll-a. Table 7 indicates the zero collinearity of the uncorrelated variables using statistical measures such as the tolerances and Variance Inflation Factors (VIFs), both of which are equal to 1.

Factors that significantly increase the coefficient of determination are included in the model. Accordingly, scores 1, 2 and 3 were found to have significant linear relationship with chlorophyll-a (Table 7). The regression statistics shows that 70% of the chlorophyll-a variation is explained by the linear combination of factor scores 1, 2 and 3. The addition of 4th PC scores in the model raised the coefficient of determination to 72%. However, this difference is not statistically significant, $P > 0.05$. Figure 13 shows the regression between predicted chlorophyll-a values from PC regression model and the chlorophyll-a concentrations. The model provides R^2 of 0.70 for chlorophyll-a prediction and RMSE = 11%.

A total increase in chlorophyll-a level would lead to a decrease in significant variables of score 1, namely, derivatives at 400, 410, 420, 440, 580, 610, 660 and 670 nm. On the other hand, chlorophyll-a abundance would cause an increase in derivative values at 500, 510, 530, 540, 550, 560 and 680 nm. This is consistent with

Table 7. Stepwise regression analysis of PC scores against chlorophyll *a* (n = 89).

Model		Regression coefficients				Collinearity statistics		
		B	SE	*t*	*P*	Tolerance	VIF	R^2 (%)
1	(Constant)	6.98	0.14	50.35	0.00			48.8
	Factor 1	1.41	0.15	9.09	0.00	1.00	1.00	
2	(Constant)	7.01	0.12	60.14	0.00			63.4
	Factor 1	1.57	0.13	11.82	0.00	1.00	1.00	
	Factor 3	−0.79	0.13	–6.09	0.00	1.00	1.00	
3	(Constant)	7.12	0.11	64.72	0.00			70.3
	Factor 1	2.03	0.17	12.32	0.00	1.00	1.00	
	Factor 3	−1.16	0.15	−7.80	0.00	1.00	1.00	
	Factor 2	0.95	0.23	4.17	0.00	1.00	1.00	

Note: The first three principal components account for 87.8% of the optical variability in the WBLE during the summer period of 2009 and 2010. The regression model explains about 70% of the chlorophyll a variation. Dependent variable: TSM.

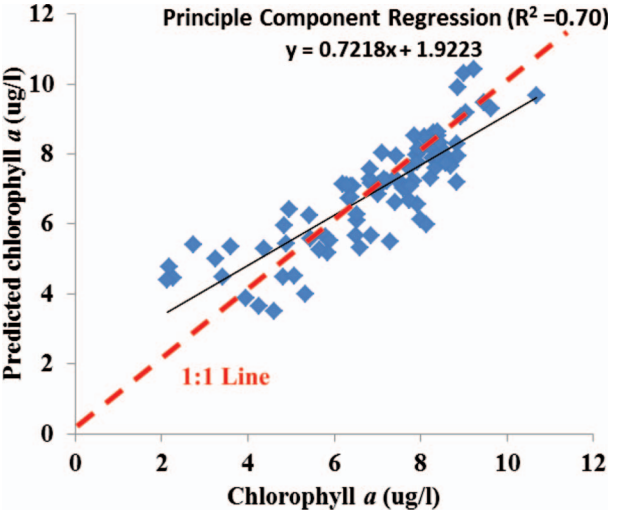

Figure 13. Regression analysis between predicted chlorophyll chlorophyll-a levels based on model 3 and in-situ chlorophyll *a* values.

chlorophyll-a's spectral features, i.e., absorption in the Soret bands and the red absorption near 670 nm. The positive relation between chlorophyll-a abundance and the back scattering recorded in the range of 500 to 560 nm is due to minimum absorption capacity of the pigment and backscattering effects from cell walls, the maxima observed at 680 nm is the fluorescence effect due to chlorophyll-a.

PCR against TSM concentrations

Principal component regression analysis is performed between the uncorrelated variables, PC scores, and dependant variable, TSM. The first three factors are found to have significant correlation with the TSM, $P < 0.05$ (Table 8). Approximately, 75% of the TSM variability is accounted for by the linear multivariate regression analysis using the first three principal factors (Figure 14), $R^2 = 0.75$ and RMSE = 30%.

The TSM variability in the WBLE is mainly explained by the third principal factor and the coefficient of determination increased with the addition of factors 2 and 1. The PCs contribute significantly to the performance of the model with all three factors having correlation coefficients statistically different from zero. TSM has a positive impact on factor 3, while it possesses a negative correlation with PC Scores 2 and 1. In other words, TSM levels would be expected to increase the values of Score 3, which refers to the reflectance level recorded at 540, 550 and 560 nm. The increase in reflectance in the green region due to suspended matter is corroborated in many previous works (Ortiz *et al.* unpublished results, Gordon and Morel 1983, Schalles 2006). Regression analysis between PCs and TSM shows that factors 2 and 1 have negative coefficients (Table 8). Factor 1 represents the Soret bands and the red region. Factor 2 represents mainly the spectral regions 440–510 nm. Figure 14 illustrates that the regression is strongly controlled by the outliers. These high value data represent concentrations of the TSM in the turbid waters of Sandusky Bay, and therefore are considered to be valid measurements that should be incorporated in the model.

Table 8. Stepwise regression analysis of PC scores against TSM ($n = 89$).

Model		Regression coefficients				Collinearity statistics		
		B	Std. Error	*t*	*P*	Tolerance	VIF	R^2 (%)
1	(Const)	8.50	0.58	14.79	0.00			33.4
	Factor 1	3.84	0.58	6.64	0.00	1.00	1.00	
2	(Const)	8.50	0.45	19.04	0.00			60.2
	Factor 1	3.84	0.45	8.54	0.00	1.00	1.00	
	Factor 2	−3.44	0.46	−7.66	0.00	1.00	1.00	
3	(Const)	8.50	0.35	24.07	0.00			75.4
	Factor 1	3.84	0.35	10.80	0.00	1.00	1.00	
	Factor 2	−3.44	0.36	−9.68	0.00	1.00	1.00	
	Factor 3	−2.58	0.35	−7.28	0.00	1.00	1.00	

Note: The first three principal components explain about 75% of the TSM variation.

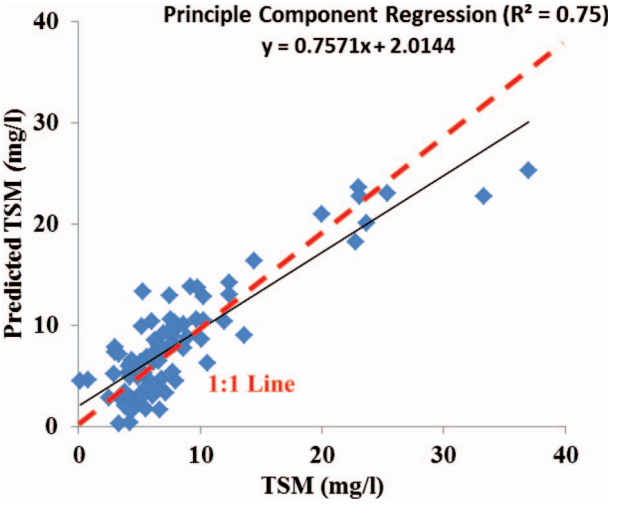

Figure 14. Regression analyses between predicted TSM levels and gravimetrically measured TSM in the WBLE.

The presence of multiple constituents in the TSM in the WBLE causes various levels of absorption and scattering across the spectrum. The inverse relationship between the reflectance near the Soret bands and TSM signals the presence of multiple constituents including CDOM which may be bound to filtered particles as suggested in the works of Binding *et al.* (2008). TSM also showed inverse relation with bands near 670 nm, signalling the presence of a phytoplankton component.

Calibration–validation

Calibration–validation procedures were applied to the PCR models for chlorophyll-a and TSM prediction. The full dataset was divided randomly into subsets of 60% for

Table 9. Summary of the validation parameters computed using the calibration and validation dataset for the PCs and two CPAs (chlorophyll *a* and TSM).

	PCA				
	N	MRES	R2	RMSE	Beta Values
Dataset – chlorophyll *a*					
Training	6261	0.02	0.70	0.98	0.72
Validation	28	0.86	0.71	2.041	0.64
Data set – TSM					
Training	6261	0.09	0.77	2.89	0.81
Validation	28	–0.39	0.69	4.17	0.63

model calibration and 30% for model validation purposes. Table 9 shows that the validation parameters using 30% of the data (mean residuals (MRES), R^2 and RMSE) for both chlorophyll-a and TSM are only slightly worse than the calibration statistical parameters. Parameters from validation datasets (28 dataset) were expected to be worse than those attained from the calibration dataset (61 dataset) because the validation data were not included in the development of the model. The beta coefficients in the calibration and validation models are within the error margin for the chlorophyll-a. The fact that all of the validation statistics have similar values to that of calibration dataset demonstrated that the model is stable and it is not affected by dataset size making the model more robust.

Conclusion

Multivariate decomposition of reflectance measured from filtered sample using a laboratory-based spectrophotometer clearly indicated the presence of multiple factors that affect the optical characteristics of the WBLE, namely, cyanobacteria, diatoms and inorganic minerals. This approach effectively (a) reduces number of the variables in multiple regression models, (b) removes scattering effect of variables and (c) eliminates multicollinearity problems. The consistency and reproducibility of the extracted Lake Erie water constituents from the PCA analysis, both in the case of the joint spatiotemporal data and temporally specific data, indicates that the multivariate approach is stable method that may be applied efficiently in Case 2 waters.

Various models have previously been applied to explain the biogeophysical characteristics of Case 2 waters from remote sensing data (Moore 1980, Gitelson 1992, Dekker 1993, Han *et al*. 1994, O'Reilly *et al*. 1998, Gower *et al*. 1999, 2005, Håkansson 2000, Gons *et al*. 2002, Ruddick *et al*. 2003, Parinet *et al*. 2004, Simis *et al*. 2005, Schalles 2006, Gitelson *et al*. 2008, 2009, Witter *et al*. 2009, Moses *et al*. 2009). The successes of bio-optical models have been commonly described using simple statistical criteria such as the coefficient of determination and the root mean square error of the predictions. A number of studies have defined the relationship of reflectance with biogeophysical factors as univariate and others have adapted a multivariate linear or nonlinear approach. Recent studies show that multivariate models are capable of assessing large number of variables and interrelations and therefore, more successful in defining and predicting complex biogeophysical processes.

In this study, the relationships between the first derivative of hyper-spectral bands with chlorophyll-a and TSM s have been investigated using a linear multivariate approach. This model has been able to predict concentration of chlorophyll-a and TSM in the Western Basin of Lake Erie with a maximum predictive success of 70% and 75%, respectively, using approximately 88% of the variation in the optical data. Calibration–validation procedures clearly demonstrated that the PCA-based models are stable and the results are not affected by dataset size.

Acknowledgements

The authors thank M. Thomas, Captain of the Stone Laboratory research vessels, Stone Laboratories, and Ohio SeaGrant for access and funds required to collect the samples.

References

Balsam, W.L., and Deaton, B.C., 1991. Sediment dispersal in the Atlantic Ocean: evaluation by visible light spectra. *Reviews in Aquatic Sciences*, 4 (4), 411–477.

Barnes, R.J., Dhanoa, M.S., and Lister, S.J., 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, 43 (5), 772–777.

Becker, R.H., *et al*., 2009. Mapping cyanobacterial blooms in the great lakes using MODIS. *Journal of Great Lakes Research*, 35 (3), 447–453.

Binding, C.E., *et al*., 2008. Spectral absorption properties of dissolved and particulate matter in Lake Erie. *Remote Sensing of Environment*, 112 (4), 1702–1711.

Budd, J.W., *et al*., 2002. Satellite observations of microcystis blooms in western Lake Erie. *Limnology*, 27, 3787–3793.

Bukata, R.P., *et al*., 1995. *Optical properties and remote sensing of inland and coastal waters*. Boca Raton, FL: CRC Press.

Clark, R.N., *et al*., 2003. *USGS digital spectral library splib05a*. US Geological Survey Open-File Report 03-395.

Dall'Olmo, G., and Gitelson, A.A., 2005. Effect of bio-optical parameter variability on the remote estimation of chlorophyll-a concentration in turbid productive waters: experimental results. *Applied Optics*, 44 (3), 412–422.

Deaton, B.C., and Balsam, W.L., 1991. Visible spectroscopy; a rapid method for determining hematite and goethite concentration in geological materials. *Journal of Sedimentary Research*, 61 (4), 628.

Dekker, A.G., 1993. *Detection of optical water quality parameters for eutrophic waters by high resolution remote sensing*. Amsterdam: Vrije Universiteit.

Doerffer, R., and Fischer, J., 1994. Concentrations of chlorophyll, suspended matter, and gelbstoff in case II waters derived from satellite coastal zone color scanner data with inverse modeling methods. *Journal of Geophysical Research*, 99 (C4), 7457–7466.

Doerffer, R. and Schiller, H., 2008. MERIS regional coastal and lake case 2 water project atmospheric correction ATBD. Geesthacht, Germany: Inst. Coastal Res., GKSS Res. Center, Rep. GKSS-KOF-MERIS-ATBD01, 1.

Gitelson, A., 1992. The peak near 700 nm on radiance spectra of algae and water: relationships of its magnitude and position with chlorophyll concentration. *International Journal of Remote Sensing*, 13 (17), 3367–3373.

Gitelson, A.A., *et al*., 2009. A bio-optical algorithm for the remote estimation of the chlorophyll-a concentration in case 2 waters. *Environmental Research Letters*, 4, 045003.

Gitelson, A.A., *et al*., 2008. A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: validation. *Remote Sensing of Environment*, 112 (9), 3582–3593.

Gitelson, A., *et al*., 2000. Remote estimation of phytoplankton density in productive waters. *Advances in Limnology.Stuttgart*, 121–136.

Glasgow, H.B., *et al*., 2004. Real-time remote monitoring of water quality: a review of current applications, and advancements in sensor, telemetry, and computing technologies. *Journal of Experimental Marine Biology and Ecology*, 300 (1–2), 409–448.

Gons, H.J., 1999. Optical teledetection of chlorophyll a in turbid inland waters. *Environmental Science & Technology*, 33 (7), 1127–1132.
Gons, H.J., Rijkeboer, M., and Ruddick, K.G., 2002. A chlorophyll-retrieval algorithm for satellite imagery (medium resolution imaging spectrometer) of inland and coastal waters. *Journal of Plankton Research*, 24 (9), 947.
Gordon, H.R., *et al*., 1988. A semi-analytic radiance model of ocean color. *Journal of Geophysical Research*, 93 (D9), 10909–10,924.
Gordon, H.R., and Morel, A.Y., 1983. *Remote assessment of ocean color for interpretation of satellite visible imagery: a review*. New York: Spring-Verlag.
Gower, J., *et al*., 2005. Detection of intense plankton blooms using the 709 nm band of the MERIS imaging spectrometer. *International Journal of Remote Sensing*, 26 (9), 2005–2012.
Gower, J.F.R., Doerffer, R., and Borstad, G.A., 1999. Interpretation of the 685 nm peak in water-leaving radiance spectra in terms of fluorescence, absorption and scattering, and its observation by MERIS. *International Journal of Remote Sensing*, 20 (9), 1771–1786.
Håkansson, B., 2000. Satellite remote sensing of the coastal ocean: water quality and algae blooms. *In*: C.R.C. Sheppard, ed. *Seas at the millennium: an environmental evaluation: 3. Global issues and processes*. Pergamon: Amsterdam.
Han, L., *et al*., 1994. The spectral responses of algal chlorophyll in water with varying levels of suspended sediment. *International Journal of Remote Sensing*, 15 (18), 3707–3718.
Harris, S.E. and Mix, A.C., 1999. Pleistocene precipitation balance in the Amazon basin recorded in deep sea sediments 1. *Quaternary Research*, 51 (1), 14–26.
Helena, B., *et al*., 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Research*, 34 (3), 807–816.
IOCCG, 2000. *Remote sensing of ocean colour in coastal, and other optically-complex, waters, International ocean color group* [online]. Dartmouth, Canada. Available from: http://www.ioccg.org/reports/report3.pdf [Accessed 22 November 2012].
Kaiser, H.F., 1958. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23 (3), 187–200.
Kaiser, H.F., 1960. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141–151.
McClain, C.R., 2009. A decade of satellite ocean color observations*. *Annual Review of Marine Science*, 1, 19–42.
Mix, A.C., Harris, S.E., & Janecek, T.R., 1995. Estimating lithology from nonintrusive reflectance spectra: ODP Leg 138. *In*: N.G. Pisias, et al., eds. *Proceedings of the ocean drilling program*, Scientific Results 138, College Station, TX (Ocean Drilling Program), 413–427.
Mix, A.C., *et al*., 1999. Rapid climate oscillations in the northeast pacific during the last deglaciation reflect northern and southern hemisphere sources. *Mechanisms of Global Climate Change at Millennial Time Scales*, 112, 127–148.
Moberg, L., *et al*., 2002. Assessment of phytoplankton class abundance using absorption spectra and chemometrics. *Talanta*, 56 (1), 153–160.
Moore, G.K., 1980. Satellite remote sensing of water turbidity. *Hydrological Sciences Bulletin*, 25 (4), 12.
Morel, A. and Prieur, L., 1977. Analysis of variations in ocean color. *Limnology and Oceanography*, 22 (4), 709–722.
Moses, W.J., *et al*., 2009. Estimation of chlorophyll-a concentration in case II waters using MODIS and MERIS data – successes and challenges. *Environmental Research Letters*, 4, 045005.
Munawar, M. and Weisse, T., 1989. Is the 'microbial loop' an early warning indicator of anthropogenic stress? *Hydrobiologia*, 188–189, 163–174.
O'Reilly, J.E., *et al*., 2000. Ocean color chlorophyll a algorithms for SeaWiFS, OC2, and OC4: version 4. *SeaWiFS Postlaunch Calibration and Validation Analyses*, Part 3, 9–23.
O'Reilly, J.E., *et al*., 1998. Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research*, 103 (C11), 24937–24953.
Ortiz, J.D., O'Connell, S., and Mix, A.C., 1999. *Data report: spectral reflectance observations from recovered sediments*. College Station, TX: Ocean Drilling Program, Texas A & M University.
Ortiz, J.D., *et al*., 2004. Enhanced marine productivity off western north America during warm climate intervals of the past 52 ky. *Geology*, 32 (6), 521.

Ortiz, J.D., *et al*., 2009. Provenance of holocene sediment on the chukchi-alaskan margin based on combined diffuse spectral reflectance and quantitative X-ray diffraction analysis. *Global and Planetary Change*, 68 (1–2), 73–84.

Ortiz, J.D., *et al*., Unpublished results. Evaluation of multiple color producing agents in the case II waters of Lake Erie using VNIR derivative spectroscopy and varimax-rotated Principal component analysis.

Ortiz, J.D., 2011. Application of visible/near infrared derivative spectroscopy to Arctic paleoceanography. *IOP Conference Series: Earth and Environmental Science*, 14. ISSN 1755-1355.

Parinet, B., Lhote, A., & Legube, B., 2004. Principal component analysis: an appropriate tool for water quality evaluation and management – application to a tropical lake system. *Ecological Modelling*, 178 (3–4), 295–311.

Petersen, W., *et al*., 2001. Process identification by principal component analysis of river water-quality data. *Ecological Modelling*, 138 (1–3), 193–213.

Reynolds, C.S., 1996. Foreword. *In*: M. Munawar and I.F. Munawar, eds. Phytoplankton dynamics in the north American Great Lakes. Vol. 1. Lakes Ontario, Erie and St. Clair: SPB Academic Publishing.

Rinta-Kanto, J., *et al*., 2005. Quantification of toxic microcystis spp. during the 2003 and 2004 blooms in western Lake Erie using quantitative real-time PCR. *Environmental Science & Technology*, 39 (11), 4198–4205.

Ruddick, K., Park, Y., & Nechad, B., 2003. MERIS imagery of Belgian coastal waters: mapping of suspended particulate matter and chlorophyll-a. *ESA Special Publications*, 549, 1–10.

Schalles, J.F., 2006. *Optical remote sensing techniques to estimate phytoplankton chlorophyll a concentrations in coastal waters with varying suspended matter and CDOM concentrations*. New York: Springer.

Schalles, J.F., Rundquist, D.C., & Schiebe, F.R., 2002. The influence of suspended clays on phytoplankton reflectance signatures and the remote estimation of chlorophyll. *Vehrein Internationale Verein Limnologie Vehr. Internat. Verein. Limnol*. 27: 3619–3625.

Simis, S.G.H., Peters, S.W.M., & Gons, H.J., 2005. Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. *Limnology and Oceanography*, 50 (1), 237–245.

Tauler, R., Barcelo, D., & Thurman, E.M., 2000. Multivariate correlation between concentrations of selected herbicides and derivatives in outflows from selected US midwestern reservoirs. *Environmental Science & Technology*, 34 (16), 3307–3314.

Toepel, J., Langner, U., & Wilhelm C., 2005. Combination of flow cytometry and single cell absorption spectroscopy to study phytoplankton structure and to calculate the chl a specific absorption coefficients at the taxon level 1. *Journal of Phycology*, 41 (6), 1099–1109.

USEPA, 2010. *Lake Erie lake wide management plan*. Annual report. US EPA and Environment Canada.

USEPA, 2009. *Lake Erie lake wide management plan*. Status of nutrients in the Lake Erie basin. US EPA and Environment Canada.

USEPA, 2006. *Lake Erie lake wide management plan*. US EPA and Environment Canada.

Will, G., 2006. *Powder diffraction: the Rietveld method and the two-stage method to determine and refine crystal structures from powder diffraction data*. New York: Springer Verlag.

Witter, D.L., *et al*., 2009. Assessing the application of SeaWiFS ocean color algorithms to Lake Erie. *Journal of Great Lakes Research*, 35 (3), 361–370.

Ying Ouyang, 2005. Evaluation of river water quality monitoring stations by principal component analysis. *Journal of Water Research*, 39, 2621–2635.

Yu, C.-C., *et al*., 1998. Effective dimensionality of environmental indicators: a principal component analysis with bootstrap confidence intervals. *Journal of Environmental Management*, 53 (1), 101–119.