Blended Spectral Classification Techniques for Mapping Water Surface Transparency and Chlorophyll Concentration

Perry LaPotin, Robert Kennedy, Timothy Pangburn, and Robert Bolus

Abstract

An innovative technique for estimating Secchi Disk Transparency and Chlorophyll a concentration is examined using in situ samples and coincidental satellite imagery for West Point Lake, Georgia. The technique is divided into two main components: (1) unsupervised classification to organize and reduce spectral variance, and (2) linear logarithmic modeling to transfer class structure onto primary water quality measurements. In component 1, clusters are derived using a non-parametric approach that is computationally unique from the traditional ISODATA algorithm. The method includes focused stratified sampling, non-parametric estimation, and blending of class structure using first-order principal components. In component 2, the class structure is tied to water quality estimation using primary band ratios for visible, near infrared, and middle infrared as independent variables. The results indicate a strong association between the Landsat TM middle infrared band and observed measurements for Secchi Disk Transparency and Chlorophyll a concentration. Logarithmic ratios for the visible green to the visible red are shown to be the second most significant covariates. The resultant models are shown to explain 98 percent of the variance in Secchi Disk Transparency, and 93 percent of the variance in Chlorophyll a concentration using pooled data from 59 sampling stations acquired during two distinct periods: 08 June and 28 September 1991.

Introduction

Sampling strategies for water quality management require intensive data collection as a means to describe ambient conditions across a range of spatial scales. Focused data collection strategies are employed for individual waterbodies, especially if they exhibit water quality heterogeneities. For example, reservoirs that are large and morphologically complex frequently exhibit marked gradients or spatial patterns in water quality conditions, thus requiring data collection at numerous positions (Johnson and Gage, 1997). However, such strategies are logistically difficult, costly, and frequently do not provide data with spatial detail to adequately address management issues (EPA, 1998).

Remote sensing provides an alternative method for assessing patterns in water quality across a range of spatial scales (Zilioli and Brivio, 1997). Observations in the visible and thermal infrared portions of the electromagnetic spectrum have been commonly applied for this purpose (Schalles *et al.*, 1998), for example, for the detection of turbidity in surface waters due to the presence of suspended particles. Cox *et al.* (1998) applied non-linear regression of spectral radiance data, with *in situ* water quality measurements, to assess water quality of reservoirs on the Catawba River in North and South Carolina. Kennedy *et al.* (1994) described patterns in the distribution of water temperature, turbidity, and algal pigment at West Point Lake using empirical models derived from radiance measurements for Landsat TM bands 1–4 and 6. A similar approach was used for Lake Kinneret as a means to differentiate periodic blooms of the dinoflagellate *Peridinium* (Schalles *et al.*, 1998). However, these approaches rely on the coincidental collection of ground truth data as a means to interpret image data. Because no universal algorithms are available for estimating optically sensible water quality parameters (Cox *et al.*, 1998), coincidental sampling remains a potentially costly requirement that can limit the use of remotely sensed images, particularly when attempting to assess historical trends. Described here are results of efforts to develop and evaluate an unsupervised methodology for conducting water quality assessments. The approach is shown to be highly predictive for estimating certain water quality parameters within two distinct periods.

Water Quality Data

Water quality surveys were conducted at West Point Lake coincident with Landsat TM overflights on 08 June and 28 September 1991.[1] West Point Lake is a large (104.8 km^2) U.S. Army Corps of Engineer hydroelectric reservoir located on the Chattahoochee River 120 km downstream from Atlanta, Georgia. The waterbody has a complex morphology, and receives relatively high nutrient and sediment loads (Emmert and Bayne, 1996). High nutrient concentrations result in excessive algal growth, as evidenced by high chlorophyll concentrations in

[1]Image Specifications:

Acquisition Date (1991)	Image Identification Number	Local Standard Time (hr:min:sec)	Sun Angle (degrees)	Sun Azimuth (degrees)
08 Jun	LT5019037009115910	10:35:33.75	62.04	103.91
28 Sep	LT5019037009127110	10:36:42.72	46.82	137.82

P. LaPotin, T. Pangburn, and R. Bolus are with the Remote Sensing/GIS Center, U.S. Army Corps of Engineers, Cold Regions Research and Engineering Laboratory, 72 Lyme Road, Hanover, NH 03755 (p@sover.net).

R. Kennedy is with the U.S. Army Corps of Engineers, Waterways Experiment Station, 3909 Halls Ferry Road, Vicksburg, MS 39180.

Photogrammetric Engineering & Remote Sensing
Vol. 67, No. 9, September 2001, pp. 1059–1065.

0099-1112/01/6709–1059$3.00/0

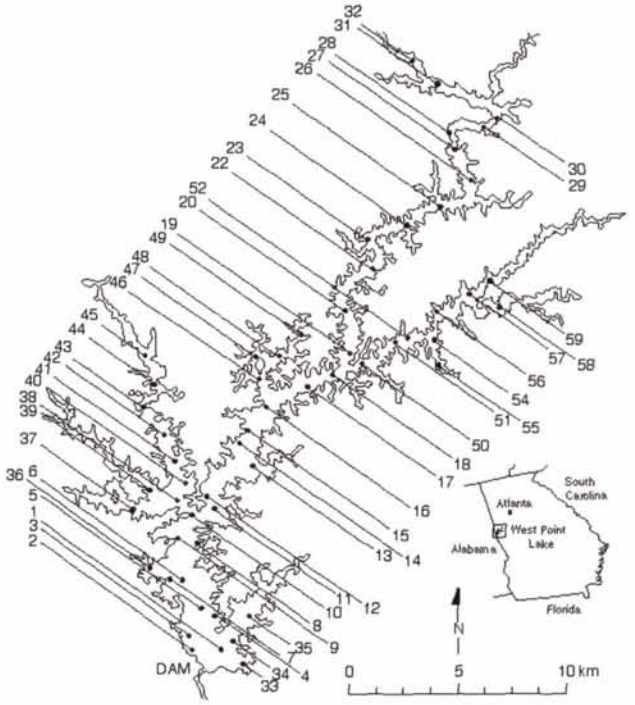


Figure 1. Water quality sample positions within West Point Lake, Georgia.

surface waters. However, turbid inflows can reduce light and inhibit algal growth downstream from the river confluence. As a result, marked gradients in water quality are apparent along the length of the reservoir and in major embayments. In general, upper reaches of the reservoir are low in chlorophyll concentration but high in turbidity due to river-borne suspended sediment concentrations. By contrast, downstream reaches are less influenced by suspended sediments and exhibit elevated chlorophyll concentrations. However, highest chlorophyll concentrations are frequently observed at mid-reservoir where nutrient concentrations are moderately high (due to the influences of in-flowing river water), and suspended sediment concentrations are relatively low.

Sample stations were distributed throughout the main body and associated tributaries in a nearly uniform manner (Figure 1). An onboard Global Positioning System (GPS) receiver was used to reference each water quality measurement. Differential correction was employed to improve positional accuracy using shoreline monuments and navigational (buoy) reference (Kennedy *et al.*, 1998). Field observations were geographically referenced to the Universal Transverse Mercator (UTM) projection to allow comparison with projected TM data. *In situ* measurements and samples of surface water (depth = 0.1 m), were collected within 2 to 3 hours of each satellite overflight. Water transparency (S_d) was determined using a Secchi Disk.[2] Water samples for determination of algal pigment were stored on ice in the dark prior to analysis. Chlorophyll *a* (C_a) concentration was determined colorimetrically following extraction from material retained on 0.45-μ glass filters (APHA 1992). The principal water quality parameters, units, and liminological rationale are provided within Table 1.

TABLE 1. DESCRIPTION OF SELECTED WATER QUALITY VARIABLES

Measure	Variable	Units	Rationale
Secchi Disk Transparency (The depth above which a standard 20-cm disk is visible at the surface)	S_d	m	General technique for measuring water transparency and assessing water quality.
Chlorophyll a (Primary algal pigment; measure of algal biomass in surface waters)	C_a	mg/m^3	Chlorophyll *a* is the primary photosynthetic plant pigment present in all algae. Algae reduces water clarity and diminishes aesthetic value.

Methodology

In water quality estimation, sharp variations in pattern and texture are apparent in multispectral data (Han *et al.*, 1994). The *spatial heterogeneity* and *spatial dependence* of these features is often pronounced due to variations in limnological features related to turbidity and chlorophyll. Standard discriminant theory (supervised classification) and cluster or factor analytical methods (unsupervised classification) neglect the spatial heterogeneity and dependence. In particular, these algorithms uniformly sample across the image, and they assume linearity, stationarity, and Independently and Identically Distributed (IID) variates (Dutilleuil and Legendre, 1993). Hence, an alternative non-parametric analysis is examined that employs focused sampling for pattern identification without IID restrictions.

The focused sampling is used to select spectral values (spectral vector $\vec{s}_m$) from the within-lake region of the TM image. The goal is to extract the maximum information concerning the distribution of the lake features and class structure. Not all pixels in the image fulfill this purpose because noise and subtle patterns are significantly under-represented using a uniform sample.[3] The sampling procedure is *adaptively* biased to select pixels that contain one type of attribute ("homogeneous" information), with a low degree of noise. The TM image is partitioned into an N-square grid (variable size). Within each grid cell, a small patch of size l is randomly selected. Within that patch, a single pixel is selected with the largest value of the local/global density ratio: i.e.,

$$R^{(l,g)}(\vec{s}_m) = \frac{\hat{f}^{(l)}(\vec{s}_m)}{\hat{f}^{(g)}(\vec{s}_m)}. \qquad (1)$$

The procedure is repeated for each pixel until no new pixels are located within the patch. The iterative process is used to define the modal *local* density function $\hat{f}^{(l)}(\vec{s}_m)$. The modal *global* density $\hat{f}^{(g)}(\vec{s}_m)$ is computed from within the neighborhood of $\vec{s}_m$. The sampling methodology is designed to minimize the influence of dominant classes, and to provide adequate representation for small, relatively scarce spectral data (e.g., oil in water, localized sedimentation patterns, tidal patterns, and limnological variations).

Hierarchical cluster analysis is used to locate spectral classes. The approach is an adaptive multivariate density estimation algorithm referred to as the *k-NN density estimator*

[2]Transparency determined as the depth above which a standard 20-cm disk is visible at the surface. Because light cannot penetrate through a water column with low transparency, remote sensing of the body is generally limited to 20 to 50 percent of the secchi disk transparency (Han *et al.*, 1994).

[3]The traditional random sample would select far too many pixels from dominant environmental strata (e.g., open water), and would likely miss the small near-shore features of water quality and bathymetry.

(Kemenade *et al.*, 1999). The algorithm estimates the density of the k^{th} nearest neighborhood that encloses each pixel. To illustrate this procedure using an *m*-dimensional sphere: (1) the pixel is the center of the sphere, and (2) the (Euclidean) distance to its k^{th} nearest neighbor is the radius. Pixels are assigned to a class if they are located within the *m*-dimensional sphere. If the spectral signature is not sufficient (i.e., the pixel is located outside the sphere), no class assignment is made. The *k*-NN density estimator for a sample pixel *x* is

$$\hat{f}(x) = \frac{k}{N d_k^m x V_m} \tag{2}$$

where d_k^m is the k^{th}-NN distance of *x*, and V_m is the volume of the unit *m*-dimensional ball.

Class structure is computed using hierarchical cluster analysis. Pixels are linked to spectral classes based upon proximity measures. A pixel is merged with a class when the union of its *k*-NN neighborhood with that of the nearest pixel (in that class) is sufficiently dense. A criterion for the separation of two clusters is

$$S_p = \frac{V}{\min\{p_k, p_l\}} \tag{3}$$

where p_l denotes the maximum density in cluster c_i and V is the density for the cylindrical envelope for the *k*-NN neighborhood.

The separation parameter S_p may be used as a threshold to control the number of classes and the level of detail in the classified image.[4]

Classification Results

Blended classification results are provided in Plates 1a and 1b using the parameters shown in Table 2. For this analysis, a cluster separation ($S_p = 0.8$) was selected to produce 16 unique classes within each of the two TM scenes. A value ($S_n = 10$) was selected based upon (1) excellent separation for near shore features and (2) significant discrimination for circulation and bathymetric effects within the main channel.

Within Plates 1a and 1b, the spectral blend is computed using the end points **a** and **b** from the first Principal Component (PC). Pixels, with projections **y** (between **a** and **b**) portrayed as a linear mixture $\mathbf{y} = \mathbf{a} + k\mathbf{b}$, where *a* is the initial point of the first PC segment, and **b** is the direction vector parallel to the first eigenvector. Therefore, each image shows the class structure (PC) weighted by the most significant spectral combination of bands.

The blended classification results (from Plates 1a and 1b) are combined with the independent (radiometric) measures (described in Table 3) to estimate Secchi Disk Transparency (S_d) and Chlorophyll *a* (C_a). The four respective logarithmic measures are calculated for *each* GPS sampling position where a corresponding water quality parameter was measured.

To compensate for possible drift during data acquisition (e.g., boat movement and related GPS error), a 3 by 3 kernel is employed. In Table 3, an *x,y* subscript is added to each measure to indicate the kernel position.[5] Because the kernel includes a 30-m pixel on either side of the GPS location, a spatial deviation of ±30 m is considered within this analysis. The minimum footprint in any direction is 90 m. An abbreviated syntax is used to simplify the presentation of the linear logarithmic model. In this notation, $GR_{x,y} = \ln\left(\frac{G}{R}\right)_{x,y}$, $NR_{x,y} = \ln\left(\frac{N}{R}\right)_{x,y}$, $IR_{x,y} = \ln\left(\frac{I}{R}\right)_{x,y}$, and $C_{x,y} = \ln(C)_{x,y}$. Hence, it is understood that the independent measure $IR_{x,y}$ is the logarithm (base e) of the ratio middle infrared to visible red at kernel position *x,y*. If no subscript is shown, then an average measure is used. Hence, $IR = \frac{1}{9}\sum_{x=1}^{3}\sum_{y=1}^{3}\ln\left(\frac{I}{R}\right)_{x,y}$ is the kernel average value for the 3 by 3 footprint using the middle infrared ratio as a case example.

The dependent water quality measures are summarized within Table 4 for the two field acquisition periods: 08 June (summer) and 28 September 1991 (fall). During the summer period, 57 *in situ* samples were acquired for Secchi and Chlorophyll *a* concentration. During the fall period, 52 samples were acquired for S_d. The remaining Chlorophyll *a* measure shows 50 cases (with two cases omitted due to measurement error). Mean, minimum, and maximum levels are indicated for each variable. Confidence bounds are shown to indicate the relative distribution of the data with resultant standard error. An initial contrast indicates a decline in Secchi Disk Transparency from 08 June to 28 September. The water clarity declines with the increase in average Chlorophyll *a* concentration from $12.64 \pm .57\ mg\ m^{-3}$ to $17.03 \pm 1.10\ mg\ m^{-3}$.

Water Quality Estimation

General linear models were constructed to estimate Secchi Disk Transparency (S_d) and Chlorophyll *a* (C_a) concentration. Summaries are provided in Tables 5a and 5b for both dependent water quality parameters. The tables are organized by significance level. The inclusion order shows the most significant model parameter. For example, in Tables 5a and 5b the kernel average infrared measure (IR) is the most significant parameter for estimating both S_d and C_a. The measure $GR_{2,1}$ is shown to be the second most significant parameter for estimating S_d, and the measure $GR_{1,3}$ is the second most significant parameter for estimating C_a. Conversely, the parameters $NR_{3,1}$ and $IR_{3,1}$ show the least significance within each respective model (adjusted for *all prior* measures). Hence, IR shows the largest partial correlation with S_d and C_a whereas $NR_{3,1}$ and $IR_{3,1}$ exhibit the least (model dependent) partial correlation in each respective model.[6] As described in Table 3, the measures are logarithmic to ensure a purely additive model for ease of interpretation.

Within Table 5a, the blended class structure is shown in position 4 and position 5. The form is a reciprocal of the logarithm represented as C^{-1}. All model coefficients $\hat{\beta}$ are shown to be significant at the 0.06 level and below, with strongly significant weights (below 0.001) in position 1, 2, 3, and 7. Tolerance measures for each included variable are displayed within the final column. Small tolerance levels close to zero indicate redundancy. As anticipated, the blended class parameters $C_{1,2}^{-1}$ and $C_{3,1}^{-1}$ exhibit fine tolerance levels because they are optimal cluster representations of the other band compositions including the visible red, near infrared, and middle infrared. The spectral model is shown to explain 98 percent of the variance for the *in situ* Secchi Disk Transparency (S_d) data acquired 08 June *and* 28 September (pooled observations). The standard error for this model is 0.147 meters of Secchi Disk Transparency.

[4] An analogous procedure is used in traditional unsupervised classification, where the number of resultant classes is declared *a priori*.

[5] The *x,y* subscript: 1,1 indicates the upper left corner pixel; 3,3 indicates the lower right pixel; the precise GPS position is in the centroid of the kernel at position 2,2.

[6] Within each table, the following statistical measures are shown: $\hat{\beta}$, the weight (regression estimate) for the independent TM measure; se($\hat{\beta}$), the standard error for the estimate $\hat{\beta}$; the magnitude of the t-statistic for the estimate $\hat{\beta}$; the p-value for the t-statistic; and the related tolerance for the estimate $\hat{\beta}$.

(a) (b)

Plate 1. (a) Blended spectral classification for West Point Lake 08 June 1991. Brightness (−33 percent) and contrast (+78 percent) enhancement applied to highlight class structure. (b) Blended spectral classification for West Point Lake 28 September 1991. Brightness (−47 percent) and contrast (+85 percent) enhancement applied to highlight class structure.

TABLE 2. CLASSIFICATION PARAMETERS

Parameter		Level	Purpose	Rationale
Cluster Separation	S_p	0.8	Merging Threshold (controls the number of classes displayed in the final classification).	Yields 16 unique classes with excellent orthogonal separation. See Discussion in Equation 3.
Spectral Neighbors	S_n	10	Determines the size of the square pixel window for the k-NN density estimation.	Excellent stochastic separation. See Discussion in Equation 2.

TABLE 3. INDEPENDENT MEASUREMENTS AT KERNEL POSITION x,y

Measure	Abbr.	Description	Rationale
$\ln\left(\frac{G}{R}\right)_{x,y}$	$GR_{x,y}$	Logarithmic ratio of the visible green (band 2 @ 0.52–0.60 μm) to the visible red (band 3 @ 0.63–0.69 μm).	Positive discrimination of Secchi Disk Transparency (S_d) in the visible green as an inverse relationship to the visible red.
$\ln\left(\frac{N}{R}\right)_{x,y}$	$NR_{x,y}$	Logarithmic ratio of the near infrared (band 4 @ 0.76–0.90 μm), to the visible red.	Positive reflectivity of Chlorophyll (C_a) in the near infrared *relative* to the visible red (Schalles *et al.*, 1998).
$\ln\left(\frac{I}{R}\right)_{x,y}$	$IR_{x,y}$	Logarithmic ratio of the middle infrared (band 5 @ 1.55–1.75 μm) to the visible red.	Secondary covariate for estimating turbidity and transparency (Allee and Johnson, 1999).
$\ln(C)_{x,y}$	$C_{x,y}$	Blended Spectral Class Member	Non-parametric class structure for seven TM bands.

TABLE 4. SUMMARY STATISTICS FOR SECCHI DISK TRANSPARENCY AND CHLOROPHYLL A CONCENTRATION

Measure	Date (1991)	Sample Size	Mean	Min	Max	95% CI	Std. Error
S_d (*m*)	08 June	57	1.400	.400	2.200	(1.26, 1.53)	.065
$C_a\left(\frac{mg}{m^3}\right)$	08 June	57	12.644	0	21.700	(11.49, 13.79)	.573
S_d (*m*)	28 September	52	1.013	.300	1.600	(.90, 1.12)	.053
$C_a\left(\frac{mg}{m^3}\right)$	28 September	50	17.031	2.270	30.180	(14.76, 19.29)	1.127

TABLE 5a. MODEL SPECIFICATIONS: SECCHI DISK TRANSPARENCY (S_d) POOLED DATA

Inclusion Order	TM Measure	$\hat{\beta}$	se($\hat{\beta}$)	t	p	Tolerance
1	IR	0.364	0.03	9.96	*	0.05
2	$GR_{2,1}$	3.510	0.26	13.09	*	0.14
3	$GR_{3,2}$	1.012	0.30	3.36	*	0.12
4	$C^{-1}_{1,2}$	−1.134	0.49	−2.27	0.02	0.004
5	$C^{-1}_{3,1}$	0.946	0.49	1.90	0.06	0.004
6	$NR_{1,1}$	−0.208	0.09	−2.14	0.03	0.07
7	$IR_{1,3}$	0.267	0.06	4.13	*	0.03
8	$NR_{3,1}$	−0.533	0.16	−3.20	0.002	0.02

General Linear Model and Summary Statistics:
$S_d = 0.364(IR) + 3.510(GR_{2,1}) + 1.012(GR_{3,2}) - 1.134(C^{-1}_{1,2}) + 0.946(C^{-1}_{3,1}) - 0.208(NR_{1,1}) + 0.267(IR_{1,3}) - 0.533(NR_{3,1})$
$R^2 = 0.98$; Adj $R^2 = 0.98$; $F_{8,64} = 781.35$; $p < 0.0001$; $\hat{\sigma} = 0.147$; N = 72; * ≤ 0.001.

TABLE 5b. MODEL SPECIFICATIONS: CHLOROPHYLL *a* (C_a) POOLED DATA

Inclusion Order	TM Measure	$\hat{\beta}$	se($\hat{\beta}$)	t	p	Tolerance
1	IR	5.153	0.88	5.84	*	0.05
2	$GR_{1,3}$	33.211	7.52	4.41	*	0.10
3	$C_{3,1}$	3.450	1.08	3.17	0.002	0.01
4	$C_{1,2}$	−2.805	1.07	−2.60	0.011	0.01
5	$NR_{3,1}$	19.311	4.38	4.40	*	0.02
6	$NR_{3,3}$	−6.414	3.98	−1.61	0.112	0.02
7	$GR_{2,1}$	−17.284	6.40	−2.69	0.008	0.14
8	$IR_{3,1}$	−4.867	1.93	−2.51	0.014	0.02

General Linear Model and Summary Statistics:
$C_a = 5.153(IR) + 33.211(GR_{1,3}) + 3.450(C_{3,1}) - 2.805(C_{1,2}) + 19.311(NR_{3,1}) - 6.414(NR_{3,3}) - 17.284(GR_{2,1}) - 4.867(IR_{3,1})$
$R^2 = 0.94$; Adj $R^2 = 0.93$; $F_{8,64} = 137.71$; $p < .0001$; $\hat{\sigma} = 3.484$; N = 72; * ≤ 0.001.

With respect to Chlorophyll *a* (C_a), the model formulation also includes the average kernel measure for the infrared (IR), and the logarithmic ratio of the visible green to the visible red GR as highly significant parameters. The blended class structure is shown in position 3 and position 4 (non-reciprocal form). The contribution of the near infrared is shown in position 5 and position 6. The spectral model is shown to explain 94 percent of the variance in the observed Chlorophyll *a* (C_a) concentration with a respective standard error of 3.484 mg m^{-3}.

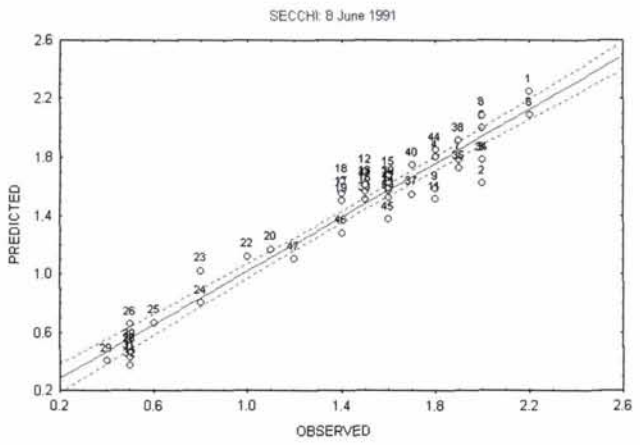
Accuracy Assessment

The models in Tables 5a and 5b are derived using pooled observations from two independent field acquisitions: 08 June and 28 September. The applicability of these models for predicting water quality characteristics on either date is described in Figures 2a and 2b and Figures 3a and 3b. In this discussion, the *pooled* model is used to predict observed water quality parameters on *each* date. Hence, although the model is based upon pooled data, the predictive efficiency of the result is assessed by position *and* by acquisition period. In Figure 2a, a plot of observed Secchi Disk Transparency (S_d) is shown for 08 June 1991. The model described in Table 5a is used as a transfer function to map the eight spectral measures onto T_n using the log-linear additive equation: $S_d = 0.364(IR) + 3.510(GR_{2,1}) + 1.012(GR_{3,2}) - 1.134(C^{-1}_{1,2}) + 0.946(C^{-1}_{3,1}) - 0.208(NR_{1,1}) + 0.267(IR_{1,3}) - 0.533(NR_{3,1})$. The predicted level for S_d is shown along the *y*-axis. The *x*-axis displays the *in situ* field observations for the 08 June 1991 period. The diagonal line indicates direct correspondence between the observed and the model predicted levels. The plot also includes a 95 percent Confidence Interval (CI) for the transfer function (linear-logarithmic model shown in Table 5a). This envelope is calculated from the se($\hat{\beta}$) levels shown within the table. As in standard regression theory, the CI is elliptical—gradually increasing at the tails of the plot (near 0.2 m and above 2.6 m). Model estimates generally fit within the CI boundary, and significant residuals are shown at positions away from the envelope. For the 08 June period, the observed and predicted values are closely aligned along the diagonal of the plot with minor residual error. For the September period, larger residuals are shown at positions 38, 40, 46–47 (within the western tributaries), and at location 3 (western edge near the dam).

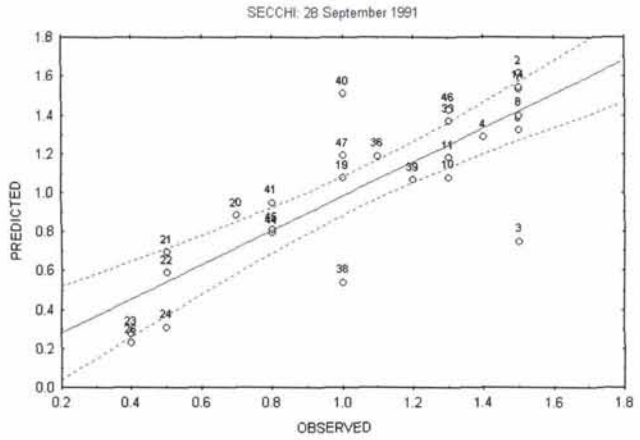
An analysis of Chlorophyll *a* (C_a), is shown in Figures 3a and 3b. For this plot, the model described in Table 5b is used as a transfer function to map the eight spectral measures onto C_a using the log-linear additive equation: $C_a = 5.153(IR) + 33.211(GR_{1,3}) + 3.450(C_{3,1}) - 2.805(C_{1,2}) + 19.311(NR_{3,1}) - 6.414(NR_{3,3}) - 17.284(GR_{2,1}) - 4.867(IR_{3,1})$. Once again, the predicted level for C_a is shown along the *y*-axis, with observed Chlorophyll *a* shown along the *x*-axis. During the June acquisition, model predicted levels are scattered between the 6- and 18-mg m^{-3} levels. The model performs well in southern locations near the main dam (positions 1–5) and within the extreme northern reaches (positions 30–32). The model is shown to systematically underestimate the observed Chlorophyll *a* within the upper tributaries at positions 25–29, and within the middle lake region at positions 13–15 and 19. Conversely, the model systematically overestimates the observed Chlorophyll *a* within the western tributary (at position 43 and 45), and within select eastern locations near the main dam (positions 33–35). The contribution of the blended class structure (east-west and north-south) is visibly apparent within Plate 1a, where the upper reaches are shown in a light intensity and the eastern edges are highlighted in a darker tones. Within the fall period (Figure 3b), the model captures the vast majority of the observed C_a within the 95 percent CI. However, the model overestimates the observed Chlorophyll *a* within station 36, 38, and 41 along the western edge of the main body near the western tributary, and within stations 3 and 6 along the western edge of the main body near the dam. Smaller residual error is shown at position 4 and 11 within the main body, and at position 48 within the mid-western tributary.

Conclusions

An innovative algorithm is presented for unsupervised classification using focused sampling and spectral blending. The

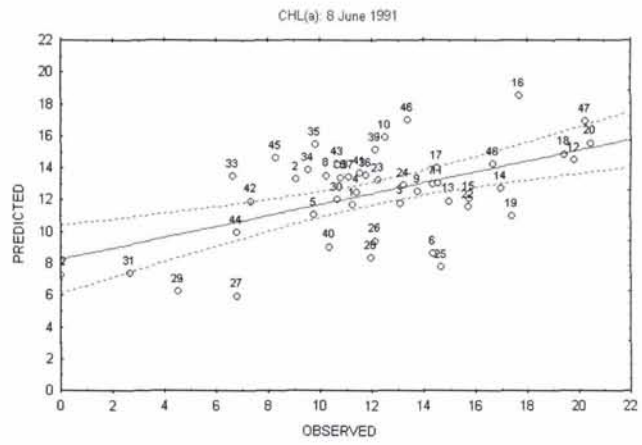


(a)

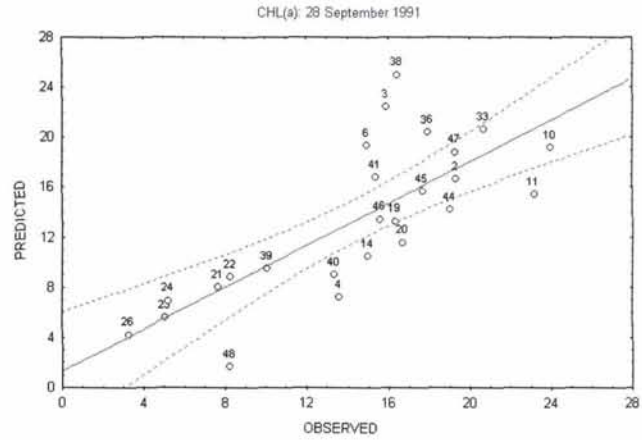


(b)

Figure 2. (a) Observed versus predicted S_d for 08 June 1991. (b) Observed versus predicted S_d for 28 September 1991.



(a)



(b)

Figure 3. (a) Observed versus predicted C_a for 08 June 1991. (b) Observed versus predicted C_a for 28 September 1991.

method uses non-parametric estimation as a technique to form unique hierarchical clusters. The clusters are used as covariates to develop linear predictive models for Secchi Disk Transparency (S_d) and Chlorophyll *a* (C_a) using Landsat TM data from 08 June and 28 September 1991.

For the estimation of S_d and C_a, the average kernel measure for the infrared

$$IR = \frac{1}{9}\sum_{x=1}^{3}\sum_{y=1}^{3}\ln\left(\frac{I}{R}\right)_{x,y}$$

is shown to be the most significant independent measure. The logarithmic ratio of the visible green (band 2 at 0.52 to 0.60 μm) to the visible red (band 3 at 0.63 to 0.69 μm) is shown as the second most significant indicator. Class structure contributes to each model as a significant covariate. For the estimation of Sd, the logarithmic classification is shown as the fourth and fifth most significant factor (i.e., fourth and fifth largest partial correlation for estimating the pooled behavior). The class structure is also shown to be highly significant for the estimation of Ca (position 3 and 4 within the hierarchy). Model results are presented using subscripts to compensate for potential GPS error and boat drift during field acquisition. The residual error is shown for each period by geographic position within the lake ecosystem. This serves two main functions:

- The statistical strength for each model is shown by period (June and September). Outliers are clearly visible, by sample position, within each residual plot.
- The statistical strength for each model is shown by position and general proximity to the dam (main channel, and in upper tributary embayments).

The model results are consistent with prior research. The primary influence of the middle infrared and the specific inclusion of the visible red band as a main indicator is well known (Rundquist *et al.*, 1996; Harding *et al.*, 1995). The linear association with suspended sediment in the visible green and visible red was identified by Roberts *et al.* (1995) and the physics of particle back-scatter (within the visible green and the visible red) was described by Topliss *et al.* (1990). Conversely, chlorophyll and dissolved organic matter act primarily as a differential absorber, causing a decline in the spectral response within the visible blue (Zilioli and Brivio, 1997). An increase of radiance (or reflectance) at the long wave-length intervals of the visible spectrum infers a decreased transparency in the water column. This relationship is closely linked to eutrophic conditions (Tassan, 1993). For this study, the inclusion of parameters in the visible green, near infrared, and middle infrared (contrasted against the visible red) provide excellent separation of the spectral response for estimating both Secchi Disk Transparency and Chlorophyll *a*.

Acknowledgments

The authors wish to acknowledge the contributions of Professor Robert Mokken, Dr. Cees van Kemenade, and Professor Han LaPoutre of the University of Amsterdam for their pioneering work in mathematical statistics and classification analysis. We are grateful to the editorial board of *PE&RS* for their excellent technical review of this manuscript. This research was supported by the U.S. Army Corps Engineers Remote Sensing Research Program.

References

Allee, R.J., and J.E. Johnson, 1999. Use of Satellite Imagery to Estimate Surface Chlorophyll *a* and Secchi Disc Depth of Bull Shoals Reservoir, Arkansas, USA, *Int. J. Remote Sensing*, 20:1057–1072.

APHA, 1992. *Standard Methods for the Examination of Water and Wastewater*, American Public Health Association, Washington, D.C., 122 p.

Cox, R.M., R.D. Forsythe, G.E. Vaughan, and L.L. Olmstead, 1998. Assessing Water Quality in Catawba River Reservoirs Using Landsat Thematic Mapper Satellite Data, *J. Lake and Res. Manage.*, 14(4):405–416.

Dutilleuil, P., and P. Legendre, 1993. Spatial Heterogeneity and Heteroscedasticity: An Ecological Paradigm versus a Statistical Concept, *Oikos*, 66:152–171.

EPA (U.S. Environmental Protection Agency), 1998. *National Strategy for the Development of Regional Nutrient Criteria*, EPA-8221-R-98-002, Washington, D.C., 286 p.

Emmert, P.P., and D.R. Bayne, 1996. Urban Phosphorus Influences on Phosphorus and Sediment Loading to West Point Lake, Georgia, *Wat. Res. Bull.*, 32:145–154.

Harding, L.W., E.C. Itsweire, and W.E. Esalas, 1995. Algorithm Development for Recovering Chlorophyll Concentrations in the Chesapeake Bay Using Aircraft Remote Sensing, *Photogrammetric Engineering & Remote Sensing*, 61:177–185.

Han, L., D.C. Rundquist, L.L. Liu, R.N. Fraser, and J.F. Schalles, 1994. The Spectral Responses of Algal Chlorophyll in Water with Varying Levels of Suspended Sediment, *Int. J. Remote Sensing*, 15:3707–3718.

Johnson, L.B., and S.H. Gage, 1997. Landscape Approaches to the Analysis of Aquatic Ecosystems, *Freshwat. Biol.*, 37:113–132.

Kemenade, C.H.M. van, J.A. La Poutré, and R.J. Mokken, 1999. Density-Based Unsupervised Classification for Multispectral Imagery, *Spatial Statistics and Remote Sensing* (A. Stein and F. Van der Meer, editors), Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 71–87.

Kennedy, R.H., J.J. Hains, S.L. Ashby, W. Jabour, B. Naugle, and B. Speziale, 1994. *Limnological Assessment of West Point Lake, Georgia*, Technical Report EL-94-6, U.S. Army Engineer Waterways Experiment Station, Vicksburg, Mississippi, 62 p.

Roberts A., C. Kirman, and L. Lesack, 1995. Suspended Sediment Concentration Estimation from Multispectral Video Imagery, *Int. J. Remote Sensing*, 16:2439–2455.

Rundquist, D.C., L. Han, J.F. Schalles, and J.S. Peake, 1996. Remote Measurement of Algal Chlorophyll in Surface Waters: The Case for the First Derivative of Reflectance Near 690 nm, *Photogrammetric Engineering & Remote Sensing*, 62:195–200.

Schalles, J., A.A. Gitelson, Y.Z. Yacobi, and A.E. Kroenke, 1998. Estimation of Chlorophyll *a* from Time Series Measurements of High Spectral Resolution Reflectance in a Eutrophic Lake, *J. Phycol.*, 34:383–390.

Tassan, S., 1993. An Improved In-Water Algorithm for the Determination of Chlorophyll and Suspended Sediment Concentration from Thematic Mapper Data in Coastal Waters, *Int. J. Remote Sensing*, 14:1221–29.

Topliss, B.J., C.L. Almos, and P.R. Hill, 1990. Algorithms for Remote Sensing of High Concentration in Organic Suspended Sediment, *Int. J. Remote Sensing*, 11:947–966.

Zilioli, E., and P.A. Brivio, 1997. The Satellite Derived Optical Information for the Comparative Assessment of Lacustrine Water Quality, *Science of the Total Environment*, 196:229–245.

(Received 25 May 2000; accepted 12 February 2001; revised 21 March 2001)