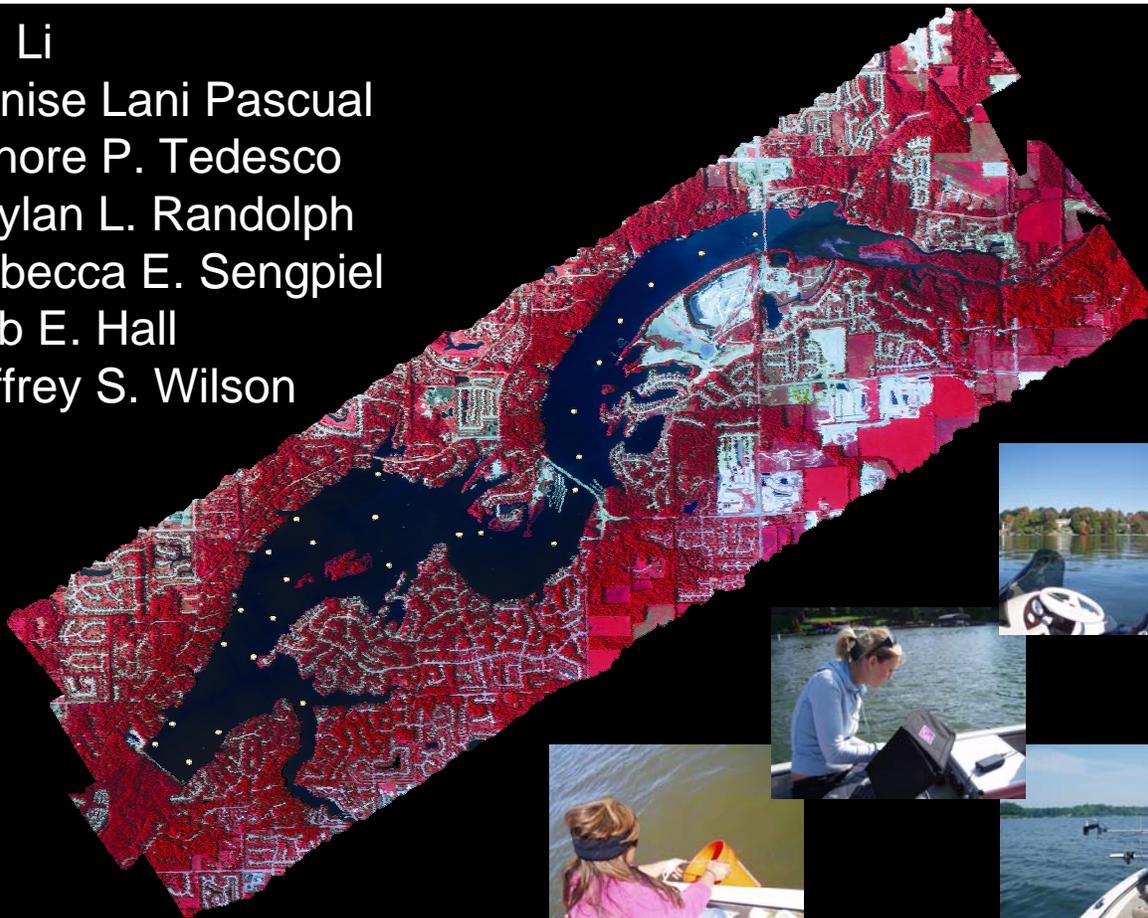


## Developing a Survey Tool for the Rapid Assessment of Blue-Green Algae in Central Indiana's Reservoirs

Lin Li  
Denise Lani Pascual  
Lenore P. Tedesco  
Kaylan L. Randolph  
Rebecca E. Sengpiel  
Bob E. Hall  
Jeffrey S. Wilson



DEVELOPING A SURVEY TOOL FOR THE RAPID ASSESSMENT OF BLUE-GREEN  
ALGAE IN CENTRAL INDIANA'S RESERVOIRS

Preliminary Report of 2005 Research

by

Lin Li  
Denise Lani Pascual  
Lenore P. Tedesco  
Kaylan L. Randolph  
Rebecca E. Sengpiel  
Bob E. Hall  
Jeffrey S. Wilson

CEES Publication 2006-01



Lin Li

Assistant Professor, Department of Earth Sciences  
Affiliated Faculty, Center for Earth and Environmental Science  
Indiana University – Purdue University, Indianapolis  
723 W. Michigan St. SL 118  
Indianapolis, IN 46202  
317-274-7484

Denise Lani Pascual

Research Scientist, Center for Earth and Environmental Science

Lenore Tedesco

Associate Professor, Department of Earth Sciences  
Director, Center for Earth and Environmental Science

Kaylan Lee Randolph

Master's Student, Department of Geography, IUPUI

Rebecca Elizabeth Sengpiel

Master's Student, Department of Earth Sciences, IUPUI

Bob E. Hall

Systems Engineer, Department of Earth Sciences  
Research Scientist, Center for Earth and Environmental Science

Jeffrey S. Wilson

Associate Professor and Chairman, Department of Geography  
Affiliated Faculty, Center for Earth and Environmental Science

Acknowledgements:

The authors wish to thank all of the graduate students and faculty in the Department of Earth Sciences, IUPUI who participated in the September 6 and 7, 2005 sampling excursion on the three reservoirs.

The authors also wish to thank the Veolia Water Indianapolis, LLC White River Laboratories for all of their assistance with sample analyses; the University of Nebraska CALMIT laboratory for acquiring the AISA imagery; IndyParks and Eagle Creek Park for dock space and launch access; and Kent Duckwell for launch access to Geist and Morse Reservoirs.

---

TABLE OF CONTENTS

---

1.0	Executive Summary .....	1
2.0	Introduction .....	3
2.1	Problem to be Addressed .....	3
2.2	Background on Pigment Spectroscopy .....	4
2.2.1	Developing a new mapping tool.....	5
2.2.2	Mapping the spatial distribution of phycocyanin .....	6
2.3	Study Sites .....	7
3.0	Methodology .....	10
3.1	Study Design .....	10
3.2	Field Spectroscopy .....	11
3.3	AISA (Airborne) Spectroscopy.....	11
3.3	GPS Data .....	12
3.4	Water Samples and Sample Analysis .....	12
3.5	Data Analysis.....	14
3.5.1	GA-PLS .....	14
3.5.2	Empirical, Semi-Empirical, and Bio-optical Models .....	17
4.0	Results and Discussion .....	20
4.1	Water Quality Data .....	20
4.2	GA-PLS Algorithms .....	22
4.2.1	Band selection results .....	22
4.2.2	Regression models and results .....	23
4.3	Performance of Empirical and Semi-Empirical Algorithms .....	25
4.3.1	Field Spectra .....	25
4.3.2	Airborne Spectra.....	26
4.4	Relationship between Phycocyanin and Blue-green Abundance .....	28
4.5	Derivation of spatial maps of phycocyanin abundances.....	29
4.5.1	Eagle Creek Reservoir.....	29
4.5.2	Geist Reservoir .....	30
4.5.3	Morse Reservoir.....	30
4.6	Assessment of robustness .....	39
4.6.1	Robustness of model built with data from Geist Reservoir.....	39
4.6.2	Robustness of model built with data from Morse Reservoir .....	39
4.6.3	Robustness of model built with data from two reservoirs .....	40
3.0	Summary .....	41
4.0	On-going and Future Work.....	43
7.0	Academic Scholarship.....	44
8.0	Work Cited .....	45
	APPENDIX A: Chlorophyll <i>a</i> Maps .....	50
	APPENDIX A: Chlorophyll <i>a</i> Maps .....	51
	APPENDIX A: Chlorophyll <i>a</i> Maps .....	52
	APPENDIX B: Analytical Methods Descriptions .....	53
	APPENDIX B: Analytical Methods Descriptions .....	54
	APPENDIX C: Genetic Algorithm Components .....	55

## LIST OF FIGURES

Figure 1: Field spectradiometer spectra of Eagle Creek Reservoir water surface.....	5
Figure 2: Geo-referenced location of sampling sites on Morse, Geist, and Eagle Creek Reservoirs. ....	10
Figure 3: Framework for study design.....	11
Figure 4: Surface water reflectance measurements. ....	12
Figure 5: A diagram showing the outer and inner relations in PLS regression .....	15
Figure 6: Linear regressions of measured phycocyanin and estimated phycocyanin using developed GA-PLS algorithms for Geist Reservoir .....	24
Figure 7: Linear regressions of measured phycocyanin an estimated phycocyanin using developed GA-PLS algorithms for Morse Reservoir.....	25
Figure 8: Linear regressions of measured phycocyanin and estimated phycocyanin using developed semi-empirical model.....	27
Figure 9: Linear regression of measured phycocyanin and estimated phycocyanin using developed the Gitelson <i>et al.</i> algorithm on Geist Reservoir: .....	28
Figure 10: Relationship between phycocyanin concentrations and measures of blue-green algal abundance .....	29
Figure 11: Interpolated map of ground truth phycocyanin concentration in Eagle Creek Reservoir on September 6, 2005.....	31
Figure 12: Interpolated map of phycocyanin concentration in Geist Reservoir on September 6, 2005 estimated from GA-PLS derived algorithm .....	32
Figure 13: Interpolated map of ground truth phycocyanin concentration in Geist Reservoir on September 6, 2005.....	33
Figure 14: Interpolated map of phycocyanin concentration in Geist Reservoir on September 6, 2005 estimated from the modified Simis <i>et al.</i> algorithm.....	34
Figure 15: Map of phycocyanin concentrations in Geist Reservoir on September 6, 2005 estimated from the modified Gitelson <i>et al.</i> algorithm .....	35
Figure 16: Interpolated map of phycocyanin concentration in Morse Reservoir on September 6, 2005 estimated from GA-PLS derived algorithm .....	36
Figure 17: Interpolated map of ground truth phycocyanin concentrations in Morse Reservoir on September 6, 2005).....	37
Figure 18: Interpolated map of phycocyanin concentration in Morse Reservoir on September 6, 2005 estimated from the modified Simis <i>et al.</i> algorithm .....	38
Figure 19: Linear regressions of measured phycocyanin and estimated phycocyanin using GA-PLS algorithms developed from Geist Reservoir datasets and applied to Morse Reservoir. ....	39
Figure 20: Linear regressions of measured phycocyanin and estimated phycocyanin using GA-PLS algorithms developed from Morse Reservoir datasets and applied to Geist Reservoir.....	40
Figure 21: Linear regressions of measured phycocyanin and estimated phycocyanin using GA-PLS algorithms developed from Geist and Morse Reservoir combined datasets.....	41
Figure 22: Summary of study results. ....	43

## LIST OF TABLES

Table 1: Reservoir and Watershed Descriptions.....	9
Table 2: Summary of empirical and semi-empirical algorithms tested. $R_{xxx}$ refers to the reflectance measurement a wavelength xxx.....	19
Table 3: Summary statistics of water quality parameters.....	21
Table 4: Spectral bands selected by GA-PLS modeling. ....	22

## 1.0 Executive Summary

The three reservoirs which supply the Indianapolis drinking water system, Eagle Creek, Morse, and Geist Reservoir, experience nuisance blue-green algal blooms. These blooms can cause water quality degradation, reducing the aesthetic quality of water by producing scums, foul odors, and off tastes; affecting the ecological structure of the reservoir by creating an anoxic hypolimnion which results in a shift from cold water to warm water fisheries; and reducing the function of the reservoir as a drinking water resource by the formation of taste and odor and/or toxin-producing blooms. Because these blooms can have such detrimental effects, determining the location and intensity of blue-green algal populations (e.g., *Anabaena*, *Aphanizomenon*, and *Cylindrospermopsis*) is critical to developing prevention, mitigation, and response strategies to minimize the negative impacts of blooms. This report demonstrates the applicability of remote sensing techniques to meet this mapping need.

Remote sensing techniques have been used to map productivity throughout the world's oceans since the 1970s: Landsat Multispectral Scanner (MSS); Landsat Thematic Mapper (TM), SPOT satellite HRV data, Coastal Zone Color Scanner; and Advanced Very High Resolution Radiometer (AVHRR). Using imagery acquired by these satellite sensors, researchers developed empirical models to relate spectral bands with measured concentrations of chlorophyll. These techniques have been further applied using newer satellite sensors with greater spectral band and spatial resolution, enabling more detailed mapping of ocean productivity and resulting in a growing interest in mapping inland lakes. Recent research efforts in the 1990s have described the applicability of remote sensing techniques to mapping algal pigments in inland systems where turbidity often confounds spectral response.

Therefore, the objective of this study was to demonstrate that remote sensing techniques can be applied to Eagle Creek, Geist, and Morse Reservoirs as a means to map the spatial distribution and concentration of blue-green algae, providing a rapid blue-green bloom assessment tool for water resource managers. This study tested the applicability of using both field sensor and airborne sensor acquired reflectance spectra to predict phycocyanin concentration, an accessory photosynthetic pigment unique to blue-green algae in freshwater ecosystems.

To accomplish this, field sensor and airborne sensor spectral data was collected for all three reservoirs on September 6 and 7, 2005, and additional field reflectance spectra were collected from June – October 2006. Spectral data were measured concurrently with surface water sample collection. Samples were extracted and fluorometrically analyzed for chlorophyll *a* and phycocyanin. Whole water samples were analyzed for phytoplankton biovolume and analyzed by Veolia Water Indianapolis, LCC for water quality constituents (e.g., nitrate/nitrite, total phosphorous, and MIB/Geosmin).

Research in 2005 and 2006 resulted in the following major findings:

- (1) Empirical and semi-empirical models relating field sensor acquired spectral reflectance to phycocyanin yielded high squared correlation coefficients between measured and estimated phytoplankton pigment concentrations.

- a. Models developed using genetic algorithms coupled with partial least squares (GA-PLS) to empirically relate spectral reflectance and phycocyanin resulted in a high correlation between measured and estimated phycocyanin for Geist Reservoir ( $r^2=0.77$ ; R: 10 – 160 ppb of phycocyanin) and Morse Reservoir ( $r^2=0.92$  R: 0 – 140 ppb of phycocyanin).
  - b. Empirical models by Gitelson *et al.* (1995) relating spectral reflectance to phycocyanin resulted in a high correlation between measured phycocyanin and estimated phycocyanin from airborne acquired spectra collected from Geist Reservoir:  $r^2 = 0.66$ , R: 15 – 255 ppb of phycocyanin.
  - c. Semi-empirical algorithms developed by Simis *et al.* (2005), modified using a backscattering model and absorption coefficients, and used to relate field spectral reflectance with phycocyanin concentration performed well. Analysis resulted in a high correlation between measured and estimated phycocyanin for Geist and Morse Reservoirs individually:  $r^2 = 0.78$  and  $r^2 = 0.91$ , respectively, for a range of 0 – 140 ppb of phycocyanin. An aggregated dataset of Geist and Morse Reservoirs resulted in an  $r^2 = 0.85$  for the same phycocyanin range.
- (2) The relationship between measures of blue-green biomass and phycocyanin resulted in high correlations between phycocyanin and blue-green counts and biovolume ( $r^2 = 0.91$  and  $r^2 = 0.95$ , respectively).

Based on these results, field and airborne hyperspectral remote sensors can provide both quantitative and qualitative data on the distribution and concentration of blue-green algae in Eagle Creek, Geist, and Morse Reservoirs<sup>1</sup>. On-going work is focused on refining the predictive algorithms for application across a range of inland water bodies that commonly experience blue-green algal blooms, which pose challenges to water resource managers.

---

<sup>1</sup> Eagle Creek Reservoir data was not included in this study due to an error in sensor calibration.

## 2.0 Introduction

As population growth continues to increasingly stress our drinking water resources, resource managers continually need better assessment tools to protect and maintain the quantity and quality of our drinking water sources. Of concern in the US and abroad (e.g., India, China, Europe, and Australia) is the composition and concentration of blue-green algae (e.g., *Anabaena*, *Aphanizomenon* and *Cylindrospermopsis*). These nuisance and sometimes harmful phytoplankton blooms can result in aesthetic degradation of lakes and reservoirs due to the production of surface scums and musty, earthy smells from taste and odor causing metabolites such as MIB (2-methylisoborneol) and geosmin ((E)-1,10-dimethyl-9-decalol); recreational degradation due to hypolimnetic anoxia and fish kills of desirable sport fish (such as bass and walleye); as well as human health impairment due to the production of toxins such as anatoxins (neurotoxin), microcystins (hepatotoxin), and cylindrospermopsin (hepatotoxin). Long-term low level exposures to microcystin-LR has been suspected to contribute to high rates of liver cancer in certain parts of China (Charmichael, 1994), and a short-term acute exposure to anatoxin-a was the likely cause of death for a Wisconsin teenager in July 2002 (Behm, 2003).

Currently, Eagle Creek Reservoir, Morse Reservoir, and Geist Reservoir, are impaired due to high nutrient concentrations and the occurrence of nuisance algae: the Indiana Department of Environmental Management (IDEM)'s 303(d) list of impaired water bodies (2002, 2004, and 2006) characterizes the three reservoirs as having persistent impairment due to algal blooms and taste and odor. These three reservoirs are a part of the Indianapolis drinking water system which supplies over 800,000 Indianapolis residents with potable drinking water. Therefore, developing tools to that can quickly assess and track the spatial and temporal occurrence of blue-green algal blooms is critical to protecting and improving drinking and recreational use water quality for the people of Indianapolis and around the world.

Funding for this research comes from both the Indiana Department of Natural Resources (IDNR) Lake and River Enhancement (LARE) program and the Central Indiana Water Resources Partnership (CIWRP). The purpose of this research is to apply hyperspectral remote sensing techniques to quantitatively map the concentration and spatial distribution of blue-green algae in Eagle Creek, Geist, and Morse Reservoir on September 6 and 7, 2005 and to extend this work to develop a survey tool to rapidly determine the concentrations and spatial distribution of blue-green algae in other lake and reservoir systems. LARE funding was used to evaluate the feasibility of using hyperspectral data from a hand-held sensor as a blue-green algae rapid assessment tool, while CIWRP funding was used to evaluate the feasibility of using an airborne hyperspectral sensor as a means to map blue-green algal distribution using a high spatial resolution remote sensor. These combined projects will provide the framework for our long term goal of developing a hyperspectral airborne/satellite remote sensing monitoring system for predicting and tracking blue-green algae in the over 600 hundred public reservoirs and lakes across the state of Indiana and, potentially, elsewhere in the world.

### 2.1 Problem to be Addressed

Monitoring programs for blue-green algal blooms and the conditions conducive to bloom formation are often limited to a small number of stations that are sampled relatively infrequently. While several reservoirs such as the Indianapolis reservoirs and Lake

Monroe have been studied intensively, water resource managers in Indiana lack a powerful assessment tool capable of providing timely information on the spatial distribution and concentration of algal communities. As field sample collection and phytoplankton identification and enumeration are time and labor intensive, this research will address this practical need by demonstrating that remote sensing techniques can provide a fast and efficient method for determining the location and intensity of blue-green algal blooms, improving water resource manager understanding and response time.

This research will concurrently address the need to further understand how remote sensing techniques can be applied to mapping phycocyanin in inland water bodies. Currently, only a few successful studies have used multi- and hyper-spectral remote sensing to map phycocyanin concentrations in inland lakes. Our research will address this theoretical need by demonstrating a new method for creating empirical algorithms by using genetic algorithms with partial least squares analyses (GA-PLS), as well as testing the applicability of existing empirical algorithms in Eagle Creek, Geist, and Morse Reservoirs.

## 2.2 Background on Pigment Spectroscopy

Algae contain colored pigments, which give them characteristic spectral features. For example, chlorophyll *a* absorbs light in the blue and red bands<sup>2</sup> of the visible light spectrum and reflects light in the green. Blue-green algae such as those nuisance taxa found in the Indianapolis reservoirs all contain two major photosynthetic pigments: chlorophyll *a* and phycocyanin. While chlorophyll *a* is common to many phytoplankton taxa, phycocyanin is unique to only blue-green algae in freshwater environments. Therefore, the spectral features of phycocyanin can be used to determine the occurrence and concentration of blue-green algae in inland lakes and reservoirs such as Eagle Creek, Geist and Morse Reservoirs.

**Figure 1** shows *in situ* reflectance spectra of Eagle Creek reservoir measured in August of 2004. The spectral signature signifying the presence of algal pigments in the water include (A) low reflectance between 400 and 500 nm (**Figure 1**) resulting from algal chlorophyll (Rundquist *et al.*, 1996; Han and Rundquist, 1997); (B) maximum reflectance between 560 and 570 nm (**Figure 1**) due to the lack of absorption by algal chlorophyll, thus giving algae a green color to our eyes (Dekker *et al.*, 1991); (C) a strong phycocyanin absorption at 620 nm (**Figure 1**) unique to the blue-green (and oceanic red) algae due to phycocyanin absorbing primarily green and red light (Dekker *et al.*, 1991; Gitelson, 1992; Jensen, 2000); (D) a weak reflection at 640 nm (**Figure 1**) ascribed to backscattering from dissolved organic matter or accessory pigment (Gitelson, 1992); (E) strong chlorophyll *a* absorption of red light at about 675 nm (**Figure 1**) (Rundquist *et al.*, 1996; Han and Rundquist, 1997); (F) strong reflectance peak around 690- 700 nm (**Figure 1**) caused by an interaction of algal-cell scattering and a combined effect of pigment and water absorption (Dekker *et al.*, 1991; Gitelson, 1992; Rundquist *et al.*, 1996; Han and Rundquist, 1997); and (G) a weak reflectance peak at about 810 nm

---

<sup>2</sup> In remote sensing jargon, a “band” is range of wavelengths in the electromagnetic spectrum that are grouped together based on the wavelengths’ characteristic. For instance, light between wavelength ranges of 492 – 577 nanometers (nm) is referred to as the “green” band since our eyes perceive this light as green, and the wavelengths between 700 and 1500 nm are referred to as the “near infrared” (NIR) band.

(Figure 1) likely due to backscattering from algal cells combined with the general absorption of the near infrared band by clear water (Rundquist *et al.*, 1996; Han and Rundquist, 1997). Spectral characteristics of phycocyanin pigments provide a physical basis for quantifying the concentrations of phycocyanin and, thus, blue-green algal concentrations using remote sensing techniques.

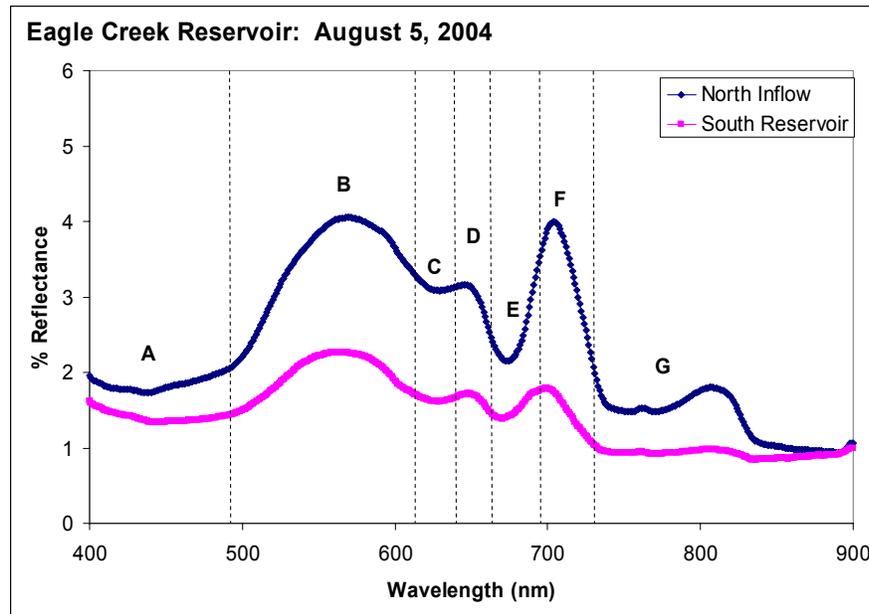


Figure 1: Field spectroradiometer spectra of Eagle Creek Reservoir water surface.

### 2.2.1 Developing a new mapping tool

Multispectral approach: The use of remote sensing as a tool to map algal blooms of inland water has been used since the early 1970s (Strong, 1974). Numerous studies have been done to map algal abundance with satellite multispectral remotely sensed data such as Landsat Multispectral Scanner (MSS) (Strong, 1974; Carpenter and Carpenter, 1983; Lillesand *et al.*, 1983; Richie *et al.*, 1990), Landsat Thematic Mapper (TM) (Verdin, 1985; Lathrop and Lillesand, 1986; Richie *et al.*, 1990; Lathrop *et al.*, 1991; Lavery *et al.*, 1993), SPOT satellite HRV data (Lathrop and Lillesand, 1989; Cairns *et al.*, 1997) and Advanced Very High Resolution Radiometer (AVHRR) (Prangma and Roozkrans, 1989). The retrieval of the algal abundance from these multispectral data was accomplished through the regressed relationship between *in situ* measurement of algal abundance and their reflectance values (Liu *et al.*, 2003). While successful cases of using multispectral satellite data to map chlorophyll *a* exist, Dekker *et al.* (1991) noted that TM and SPOT data both have the disadvantage of missing the spectral bands for the detection of chlorophyll *a* in Case II waters<sup>3</sup>; Danson *et al.* (1991) suggested that airborne TM could not be directly related to chlorophyll *a* concentration; Cairns *et al.* (1997) found the correlation of SPOT data to chlorophyll *a* abundance to be so weak that they could not be used as a practical tool for monitoring water quality. All

<sup>3</sup> Case II waters are defined as those waters in which light scattering is due a combination of non-algal turbidity, algal pigments, and water itself (e.g., turbid estuaries and eutrophic inland lakes). This is juxtaposed to Case I waters in which light scattering is due to algal pigments and water itself (e.g., ocean water and oligotrophic inland lakes).

these limitations to the effectiveness of TM and SPOT are very likely due to the broad band width of these multispectral data.

Hyperspectral approach: The past decade has seen a rapid advance in remote sensing technologies, as signified by the presence and deployment of hyperspectral imaging spectrometers. These sensors provide high spectral resolution<sup>4</sup> composed of a large number of narrow, contiguous bands and better spatial resolution<sup>5</sup>. Zhang *et al.* (2003) demonstrated that the narrow band data acquired from Airborne Imaging Spectrometer for Application (AISA) and Moderate Resolution Imaging Spectroradiometer (MODIS) resulted in more accurate estimates of chlorophyll *a* than those from satellite multispectral data (TM and AVHRR). However, MODIS spatial resolution is 250 m, while TM spatial resolution is 30 m. Several studies have argued that the spectral resolution is more important than spatial resolution for reliable retrieval of water quality parameters (Koponen *et al.*, 2001; Pulliainen *et al.*, 2001; Koponen *et al.*, 2002; Zhang *et al.*, 2003). However, for small lakes (Area <10 km<sup>2</sup>), high spatial resolution is a necessity as pixel sizes larger than the lakes will obscure the lake's spectral signatures. Therefore, while the use of new generation satellite and airborne imaging spectrometers with greater spectral and spatial resolution has improved capabilities for monitoring and mapping phytoplankton distribution in inland lakes and reservoirs.

In recent years, a number of investigators have obtained success in mapping algal abundance in lakes and reservoirs for water quality retrieval using airborne hyperspectral data (Dekker, 1993; Mille *et al.*, 1992; Jupp *et al.*, 1993; Fraser, 1998) with high spatial resolution. Data from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), which as a spatial resolution of 20 m, were used to estimate chlorophyll *a* content in Lake Tahoe (Hamilton *et al.*, 1993); AISA data were used to retrieve chlorophyll *a* and other parameters of water quality of various lake types in different seasons (Kallio *et al.*, 2001; Koponen *et al.*, 2002). With the increasing spatial resolution and spectral resolution of sensors, these studies have shown that hyperspectral remote sensing is a highly successful technique for mapping chlorophyll *a* from space, and can be operationally used on a routine basis to produce accurate estimations of algal pigment spatial distribution (Pulliainen, *et al.*, 2001).

### **2.2.2 Mapping the spatial distribution of phycocyanin**

As described above, numerous applications of multi- and hyper- spectral remote sensing in mapping algal pigments have been focused on chlorophyll *a*, the pigment common to all phytoplankton. To date, little research has been done to map phycocyanin using multi- and hyper-spectral remote sensing. There are few successful cases reported in the published literature. The first example used multiple linear regressions between different combinations of TM band ratios and measured concentration to map spatial distribution of phycocyanin in Lake Erie (Vincent *et al.*, 2004), but this algorithm is saturated when phycocyanin abundance is greater than 14 µg-L<sup>-1</sup> (ppb). The second case mapped high concentrations of phycocyanin in a pond near southern San

---

<sup>4</sup> Spectral resolution refers to the ability of a sensor to differentiate the intensity of electromagnetic radiation between wavelengths. The greater the spectral resolution, the more bands can be differentiated by the sensor.

<sup>5</sup> Spatial resolution refers to the size of the smallest unit that can be identified by a sensor. This is important when trying to delineate between adjacent features. With many remote sensors, spatial resolution is often sacrificed for spectral resolution.

Francisco Bay though the use of spectral mixture analysis (SMA) of AVIRIS data (Richardson, 1996 and Kruse *et al.*, 1997). SMA is a simple additive linear model used to estimate the abundances of the materials measured by the imaging spectrometer. Each mixed spectrum is a linear combination of the "pure" spectra, each weighted by their fractional abundance within the pixel, a simple averaging technique (Boardman, 1989). While this approach has been widely used in mapping land use and land cover change as well as geological mapping, it requires a spectrally "pure" endmember which is completely impossible from field water samples. Schalles *et al.* (1998) proposed an algorithm based on the band ratio R648/R624 to determine phycocyanin concentration. Application of the Schalles *et al.* algorithm to field spectra collected with a high resolution spectroradiometer on Carter Lake in Nebraska, yielded an  $r^2$  value of 0.61. Simis *et al.* (2005) created an optical model for determining phycocyanin pigment abundance using the optical properties of phycocyanin and the attenuation and backscattering of other optically active constituents present in turbid inland water. This algorithm was developed using a portable spectroradiometer (Photoresearch, PR-650) to accommodate other remote sensing platforms (*i.e.*, Medium Resolution Imaging Spectrometer, MERIS). Application of the Simis *et al.* (2005) algorithm to field spectra collected on two lakes in the Netherlands yielded an  $r^2$  value of 0.94,  $n=34$ ; however, data showed a 20% error in estimated phycocyanin concentrations to measured phycocyanin concentrations up to 80 ppb of phycocyanin.

### **2.3 Study Sites**

Eagle Creek, Geist, and Morse Reservoirs were created between 1943 and 1968 to provide flood control and/or drinking water for the City of Indianapolis. As the City's population grew, all three reservoirs became critical resources for maintaining surface drinking water supplies. In recent years, as the reservoirs have aged, drinking water managers have documented blooms of nuisance and possibly harmful algae in all three reservoirs. IDEM has classified the three reservoir's trophic status in the Mesotrophic to Eutrophic range (IDEM 2002, 2004, and 2006). With mean Secchi disk measurements of less than 1.0 m and TSS concentrations greater than  $10 \text{ mg-L}^{-1}$  (ppm), all three reservoirs can be functionally defined as Case II waters, or waters in which light transmittance, reflectance, and absorbance are affected by suspended sediments and/or colored dissolved organic matter as well as phytoplankton pigments and water itself.

#### *Eagle Creek Reservoir*

Eagle Creek Reservoir was constructed in 1967 in the northwest corner of Marion County. Eagle Creek Reservoir is a small (area of  $5.0 \text{ km}^2$ ), shallow reservoir (mean depth of 4.2 m) with an estimated reservoir volume of  $20,900,000 \text{ m}^3$  (Tedesco *et al.*, 2005). Originally constructed for flood control, the reservoir became a direct drinking water source in 1976 when the T.W. Moses Drinking Water Plant came on-line. The reservoir is fed by Eagle Creek Watershed (HUC 05120201120), which encompasses a  $420 \text{ km}^2$  ( $162 \text{ mi}^2$ ) area above the Eagle Creek Dam. Three streams contribute the majority of inflow into Eagle Creek Reservoir with the largest flow contribution coming from the trunk stream, Eagle Creek (Tedesco *et al.*, 2003). According to USGS Stream Gage (#03353200) data (1957-2003), estimated median daily instantaneous flow into Eagle Creek Reservoir was  $0.9 \text{ m}^3\text{-s}^{-1}$  (cms). Water balance estimates of Eagle Creek Reservoir resulted in a calculated residence time of 56 days (Pascual *et al.*, 2006). An average watershed slope of <5% and the presence of productive soils allowed for crop production in the watershed, resulting in 60.1% (2003) of Eagle Creek Watershed land to be in agricultural land use (Tedesco *et al.*, 2005). High rates of nutrient input from

watershed sources such as agricultural and suburban fertilizer run-off, tile drainage, septic system leaching, waste water treatment plants, feed lot run-off, and natural sources has led to high phosphorous loads into the reservoir. CIWRP bi-weekly monitoring during the growing season from 2003 – 2005 showed a reservoir mean total phosphorus concentration of  $94 \mu\text{g P-L}^{-1}$  and a mean total N (TKN + nitrate) of  $2.6 \text{ mg N-L}^{-1}$  (**Table 1**).

#### *Geist Reservoir*

Geist Reservoir was constructed in 1944 to regulate the flow into Fall Creek, which is the surface water source to the Fall Creek Water Treatment Facility. Geist Reservoir is located northeast of the City in Marion, Hamilton, and Hancock Counties. The reservoir is small (area of  $7.5 \text{ km}^2$ ) and shallow (mean depth  $3.2 \text{ m}$ ) with an estimated volume of  $23,810,000 \text{ m}^3$  (Wilson *et al.*, 1996). Geist Reservoir is fed by Fall Creek Watershed, which encompasses a  $420 \text{ km}^2$  ( $227 \text{ mi}^2$ ) area above Geist Dam. Four streams contribute the majority of flow into Geist Reservoir with Fall Creek, the trunk stream, contributing the largest flow volume (Tedesco *et al.*, 2003). USGS Stream Gage (#03351500) data from Fall Creek (1941-2003) showed a median daily instantaneous flow of  $2.6 \text{ m}^3/\text{s}$  into the reservoir. Residence time based on this inflow rate was estimated to be 55 days. In 2000, the Fall Creek Watershed land use was dominated by agriculture (58.3%; Tedesco *et al.*, 2003). CIWRP sampling dates in 2004 and 2005 showed a reservoir mean total phosphorus concentration of  $100 \mu\text{g P-L}^{-1}$  and a mean total nitrogen concentration of  $2.0 \text{ mg N-L}^{-1}$  (**Table 1**).

#### *Morse Reservoir*

Morse Reservoir was constructed in 1956 to regulate flow to Cicero Creek at its confluence with the White River, a surface water source to the White River North Water Treatment Facility. Morse Reservoir is located north of Indianapolis in Hamilton County. Morse is a small (area of  $6 \text{ km}^2$ ), shallow (mean depth of  $4.7 \text{ m}$ ) reservoir with an estimated reservoir volume of  $28,012,000 \text{ m}^3$  (Wilson *et al.*, 1996). Morse Reservoir is fed by the Cicero Creek Watershed, which encompasses a  $590 \text{ km}^2$  ( $215 \text{ mi}^2$ ) area above the Morse Dam. Four streams contribute the majority of flow into Morse with Cicero Creek, the trunk stream, contributing the largest inflow volume to the reservoir (Tedesco *et al.*, 2003). According to USGS Stream Gage (#03349510) data from Cicero Creek (2004-2006), estimated median daily instantaneous flow into Morse Reservoir is  $1.0 \text{ m}^3/\text{s}$ . Residence time based on this inflow rate was estimated to be 70 days. Long water residence time in Morse Reservoir and a high percentage of agricultural land use in Cicero Creek Watershed, 76.9% in 2000 (Tedesco *et al.*, 2003), contribute high phosphorus loading into Morse Reservoir. CIWRP sampling dates in 2003 and 2005 showed a reservoir mean total phosphorous concentration of  $94 \mu\text{g P-L}^{-1}$  and a mean total nitrogen of  $4.1 \text{ mg N-L}^{-1}$  (**Table 1**).

DEVELOPING A SURVEY TOOL FOR THE RAPID ASSESSMENT OF BLUE-GREEN ALGAE IN  
CENTRAL INDIANA'S RESERVOIRS

**Table 1: Reservoir and Watershed Descriptions.**

Reservoir	Eagle Creek	Geist	Morse	Units
Original Purpose	Flood Control	Water Supply	Water Supply	
Date of Service	1968	1943	1956	
Surface Area	1.9	2.9	2.3	mi <sup>2</sup>
	5.0	7.5	6.0	km <sup>2</sup>
Reservoir Volume	5,500	6,300	7,400	million gallons
	21.0	23.8	28.0	million m <sup>3</sup>
Maximum Depth	54	48	42	ft
	16.7	14.7	12.9	m
Mean Depth	13' 9"	10' 6"	15' 5"	ft inches
	4.2	3.2	4.7	m
Residence Time	56	55	70	days
Watershed Area above Dam	162	215	227	mi <sup>2</sup>
	420	560	590	km <sup>2</sup>
Trophic Status	Mesotrophic- Eutrophic*	Mesotrophic*	Eutrophic*	
Mean Total P	94 <sup>†</sup>	100 <sup>†</sup>	94 <sup>†</sup>	µg P-L <sup>-1</sup> (ppb)
Mean Total N	2.6 <sup>†</sup>	2.0 <sup>†</sup>	4.1 <sup>†</sup>	mg N-L <sup>-1</sup> (ppm)
% Agriculture in Watershed	60.1%‡	58.3%‡	76.9%‡	
Trunk Stream (median flow)	Eagle Creek	Fall Creek	Cicero Creek	m <sup>3</sup> -s <sup>-1</sup> (cms) cfs
	(0.9) (31.7)	(2.6) (91.8)	(1.0) (35.3)	
Other Inflow Streams	Fishback Creek and School Branch	Thorpe Creek, Bee Camp, and Dry Branch	Little Cicero Creek, Bear Slide Creek, and Hinkle Creek	

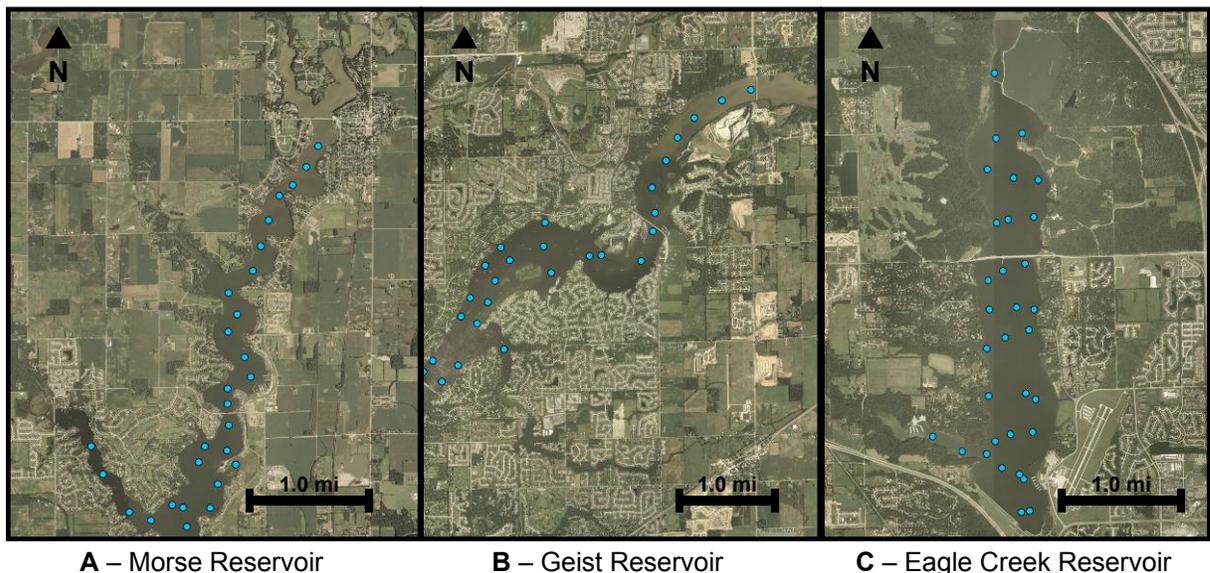
\* (IDEM 2002, 2004, and 2006)

† Various CIWRP studies from 2003 – 2005

‡ Based on 2003 land use/land cover assessments for Eagle Creek Watershed (Tedesco *et al.*, 2005) and 2000 land use/land cover assessments for Fall Creek and Cicero Creek Watersheds (Tedesco *et al.*, 2003)

### 3.0 Methodology

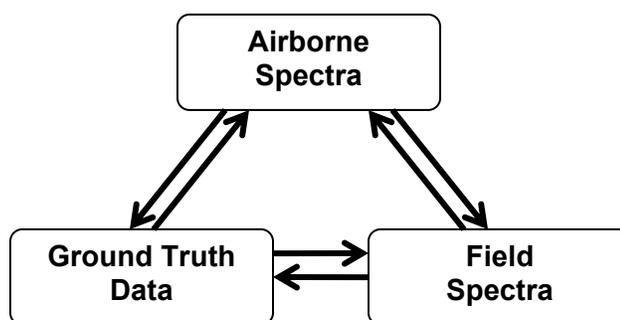
On September 6 and 7, 2005, field water surface spectroradiometer reflectance spectra, *in situ* physical water quality measurements, and ground truth water samples were collected at 32 sites on Eagle Creek Reservoir, 27 sites on Geist Reservoir, and 29 sites on Morse Reservoir (**Figure 2**) as a part of the Indiana Department of Natural Resources (DNR) Lake and River Enhancement (LARE) program grant. As a part of the companion CIWRP study, airborne hyperspectral data was also collected using an AISA-Eagle (Spectral Imaging Ltd. Oulu, Finland) sensor on board the University of Nebraska, Lincoln (UNL) Center for Advanced Land Management Information Technologies (CALMIT) Piper Saratoga airplane.



**Figure 2: Geo-referenced locations of sampling sites on Morse, Geist, and Eagle Creek Reservoirs.**

#### 3.1 Study Design

A training set of field water surface spectroradiometer reflectance spectra and extracted phycocyanin concentrations were used to develop an algorithm for predicting phycocyanin concentrations from field and airborne reflectance spectra. Algorithms developed using GA-PLS (Genetic Algorithm – Partial Least Squares) or based on published models were then tested against a validation set. Phycocyanin concentrations were then related to blue-green algal counts and biovolume in order to demonstrate a strong relationship between phycocyanin concentrations and these two commonly used measures of blue-green algal abundance (**Figure 3**).



**Figure 3: Framework for study design.** Data from airborne spectra will be related to ground truth data and field spectra. Field spectra will be related to airborne spectra and ground truth data. Ground truth data will be related to both field spectra and airborne spectra.

### 3.2 Field Spectroscopy

*In situ* irradiance spectra of water were collected at each sampling site in the three reservoirs using an ASD FieldSpec ultraviolet/visible and near-infrared (UV/VNIR) (Analytical Devices, Inc., Boulder, CO, USA) or Ocean Optics USB2000 visible and near infrared (V/NIR) spectroradiometer (Ocean Optics, Inc., Dunedin, FL, USA). The ASD spectrometer recorded a continuous irradiance spectrum over 512 bands from 348 to 1074 nm, while the Ocean Optics USB2000 recorded a continuous irradiance spectrum over 1,700+ bands from 350 to 1000 nm. For both sensors, the fiber-optical head was held at a height of 1.2 m (h) above the water surface and positioned at nadir viewing direction, the diameter of the instantaneous field of view (IFOV)<sup>6</sup> (d) on the water surface was 0.2 m, as given by  $d=h\beta$ .

Downwelling irradiance was measured using a Spectralon reference panel (Labsphere, Inc., North Sutton, NH, USA). Therefore, water surface reflectance was calculated by dividing upwelling radiance by downwelling irradiance (**Figure 4**). To reach optimal signal-to-noise ratio, *in situ* water surface reflectance spectrums for each site were calculated by averaging 15 upwelling radiance measurements (**Figure 4-B**) then dividing by the average of 15 downwelling measurements (**Figure 4-A**) captured at each site.

### 3.3 AISA (Airborne) Spectroscopy

Airborne imaging spectrometer for applications AISA-Eagle (Spectral Imaging Ltd., Oulu, Finland) was used to acquire hyperspectral imagery of the Eagle Creek, Geist, and Morse reservoirs. The AISA-Eagle sensor is a pushbroom imager with a charge coupled device CCD sensor. AISA-Eagle was set to collect data from 40 bands in the spectral region of approximately 436-800 nm with a band width of 7 – 8 nm. The IFOV across the track is 1 mrad, resulting in 1 m wide pixels and a 1000 m wide swath from an altitude of 1000 m. The sensor was installed aboard a Piper-Saratoga aircraft, the research aircraft of the UNL-CALMIT laboratory. The image data (AISA data) were processed (normalized, registered, mosaicked, and calibrated) using ENVI 4.2 software (Research Systems, Inc.). Field spectroradiometer measurements from the sample locations were

<sup>6</sup> In remote sensing jargon, the IFOV, which stands for instantaneous field of view, is the solid angle through which a detector is sensitive to radiation. Instantaneous field of view is commonly expressed in milliradians. Here, the IFOV also refers to the ground area covered by this solid angle.

used for image calibration. These were matched by using the spatial coordinates for each sample site and associating each field measurement with a 3x3 m area of the AISA image. The spectra from the nine 1x1 m pixels were averaged to obtain a single spectra matched to the field spectra taken at the same location.

### 3.3 GPS Data

The spatial coordinates of each measured site was recorded using a Trimble Pro-XRS (Trimble Navigation, Inc., Sunnyvale, CA) global positional system (GPS). The GPS coordinates were used to project the site locations onto the generated spatial maps.

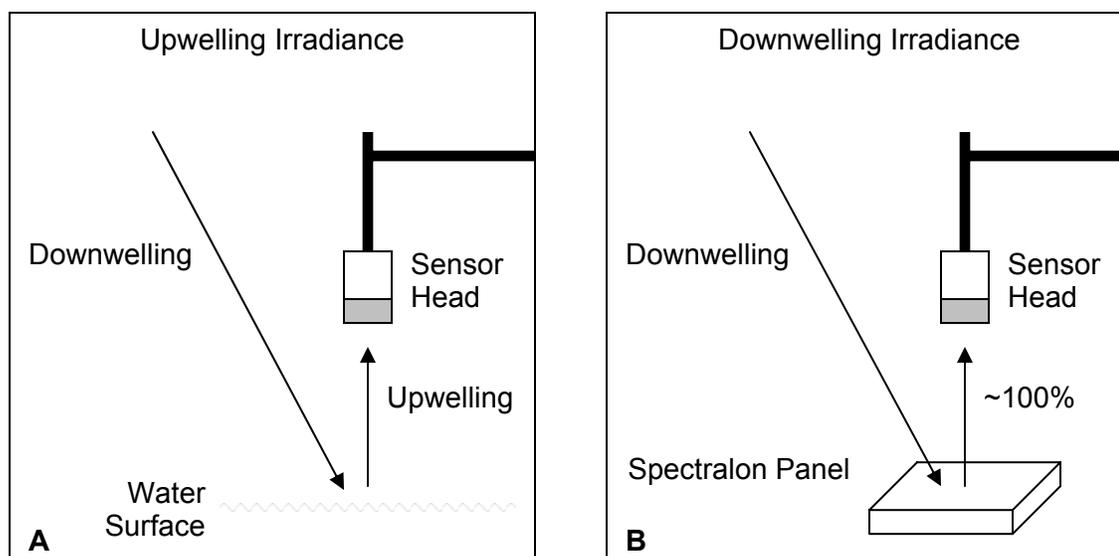


Figure 4: Surface water field reflectance spectra from the hand held sensor were calculated by dividing the upwelling irradiance measurement (A) by the downwelling irradiance measurement (B).

### 3.4 Water Samples and Sample Analysis

Ground truth, surface water grab samples (upper 30 cm) were collected at each station, concomitant with field spectra acquisition, for biological and chemical analysis (Figure 3).

#### *Pigment Analysis*

Samples analyzed for chlorophyll *a*, phycocyanin, and phytoplankton identification and enumeration were collected in 1 L amber HDPE bottles.

Chlorophyll *a* – Sample pretreatment consisted of filtering 150 – 200 mL onto a 47 mm, 0.45 micron pore size acetate filters using a filtration manifold. Filters were then placed into a 15 mL falcon tube then frozen and kept in the dark freezer (-9 °C) until analysis. Filtration and freezing occurred within 8 hours of sample collection. Samples were frozen for no longer than 3 months. Extracted chlorophyll *a* was analyzed according to EPA Method 445.0 (1997). Prior to analysis, filters were dissolved in 10 mL of 90% buffered acetone and allowed to extract in a dark freezer (-9 °C) for at least 24 hours and no longer than 48 hours. After a 1:5 or 1:10 dilution, pheophytin corrected chlorophyll *a* was measured fluorometrically using a TD-700

Fluorometer (Turner Designs, Inc.) equipped with a Daylight White Lamp and Chlorophyll Optical Kit (340-500 nm excitation filter and emission filter > 665 nm) and calibrated with chlorophyll *a* from Spinach standard (Sigma-Aldrich 10865). All steps in the chlorophyll *a* extraction process were performed under subdued light conditions. All samples were run in duplicate.

While not discussed in this report, interpolated ground truth maps of Chlorophyll *a* distribution in the reservoirs can be found in APPENDIX A.

Phycocyanin – Sample pretreatment consisted of filtering 150 to 200 mL onto a 47 mm, 0.7 micron pore size glass fiber filters using a filtration manifold. Filters were placed into a 50 mL centrifuge tube then frozen and kept in a dark freezer (-9 °C) until analysis. Filtration and freezing occurred within 8 hours of sample collection. Samples were frozen for no longer than 3 months. Extracted phycocyanin was analyzed according to Sarada *et al.* (1999). Prior to analysis, filters were transferred to a 50 mL polycarbonate centrifuge tube and suspended in 15 mL of 50 mM phosphate buffer. Filters were broken up using a stainless steel spatula. The spatula was rinsed with 5 mL of 50 mM phosphate buffer with the rinse collected in the centrifuge tube. The filter and 20 mL of buffer were homogenized for 2 minutes using a Teflon coated pestle. Pestles were rinsed into the sample with an additional 5 mL of buffer with the rinse collected in the centrifuge tube. Samples (now with 25 mL of buffer) were centrifuged at 5°C, 27,200 x g for 25 minutes using a Beckman J2-21M centrifuge. Filters were homogenized again and the pestle was rinsed using 5 mL buffer. As before, rinse buffer was collected in the centrifuge tube. The sample was then centrifuged again using the same settings. Supernatant was collected, diluted using a 1:5 or 1:10 dilution. Extracted samples were analyzed for phycocyanin concentrations fluorometrically using a TD-700 Fluorometer (Turner Designs, Inc.) equipped with a Cool White Mercury Vapor Lamp and a Phycocyanin Optical Kit (630 nm excitation and 660 nm emission filters) and calibrated using C-phycocyanin from *Spirulina sp.* (Sigma-Aldrich P6161). All steps in the phycocyanin extraction process were performed under subdued light conditions. All samples were run in duplicate.

#### *Blue-green Algal Biovolume*

Samples to be analyzed for phytoplankton identification and enumeration were split from the 1 L amber HDPE bottle taken for pigment analyses. A 50 mL aliquot was poured into a centrifuge tube and preserved with 0.5 mL of Lugol's solution. Samples were stored in the dark at 5 °C until concentrating through sedimentation and counting. Samples were stored for no longer than 9 months before analysis. Twenty-five of the 90 samples were analyzed for blue-green algal biovolume. Phytoplankton were identified to species. A minimum of 400 natural units of blue-green algae were counted thus giving a counting precision of ±10% within 95% C.L. (Lund *et al.*, 1958), and each measured for biovolume. Cell volumes were estimated by approximation to the nearest simple geometric solid (e.g., sphere, ovoid, or rod).

#### *Total Suspended Solids (TSS)*

Samples to be analyzed for TSS were taken in 1 L HDPE bottles. 150 – 700 mL of sample was filtered onto a pre-ashed, pre-weighed 47 mm, 0.7 micron pore size glass fiber filters using a filtration manifold. Filters were dried in a 60 °C oven for at least 1

hour before cooling in a dessicator and weighing (SM 2540D). All samples were run in duplicate.

#### *In-situ Water Quality Parameters*

Physical and chemical surface water data (temperature, conductivity, pH, DO, total dissolved solids (TDS), and salinity) were measured at each sample site using an YSI 600 XLM Multi-Parameter Sonde. Secchi disk depth was also measured.

#### *Other Water Quality Parameters*

Collected water samples were analyzed for nutrients and other water quality constituents. Total Kjeldahl Nitrogen (TKN) was analyzed by ESG Laboratories (Indianapolis, IN). Ortho-phosphate was analyzed by IUPUI's CEES Laboratories. The following analytes were analyzed by the Veolia Water Indianapolis, LCC Laboratories. All parameters were analyzed using the EPA and American Public Health Association Standard Methods (APPENDIX B):

- ♦ alkalinity (mg CaCO<sub>3</sub>-L<sup>-1</sup> as)
- ♦ total hardness (mg CaCO<sub>3</sub>-L<sup>-1</sup>)
- ♦ dissolved organic carbon (DOC; mg C-L<sup>-1</sup>)
- ♦ total organic carbon (TOC; mg C-L<sup>-1</sup>)
- ♦ chloride (mg-L<sup>-1</sup>)
- ♦ sulfate (mg-L<sup>-1</sup>)
- ♦ total phosphorus (µg-L<sup>-1</sup>)
- ♦ NH<sub>4</sub>-N (mg-L<sup>-1</sup>)
- ♦ nitrate (mg-L<sup>-1</sup>)
- ♦ nitrite (mg-L<sup>-1</sup>)
- ♦ total and dissolved silica (mg-L<sup>-1</sup>)
- ♦ calcium (mg-L<sup>-1</sup>)
- ♦ magnesium (mg-L<sup>-1</sup>)
- ♦ potassium (mg-L<sup>-1</sup>)
- ♦ sodium (mg-L<sup>-1</sup>)
- ♦ MIB/Geosmin (µg-L<sup>-1</sup>)
- ♦ Turbidity (NTU)

### **3.5 Data Analysis**

#### **3.5.1 GA-PLS**

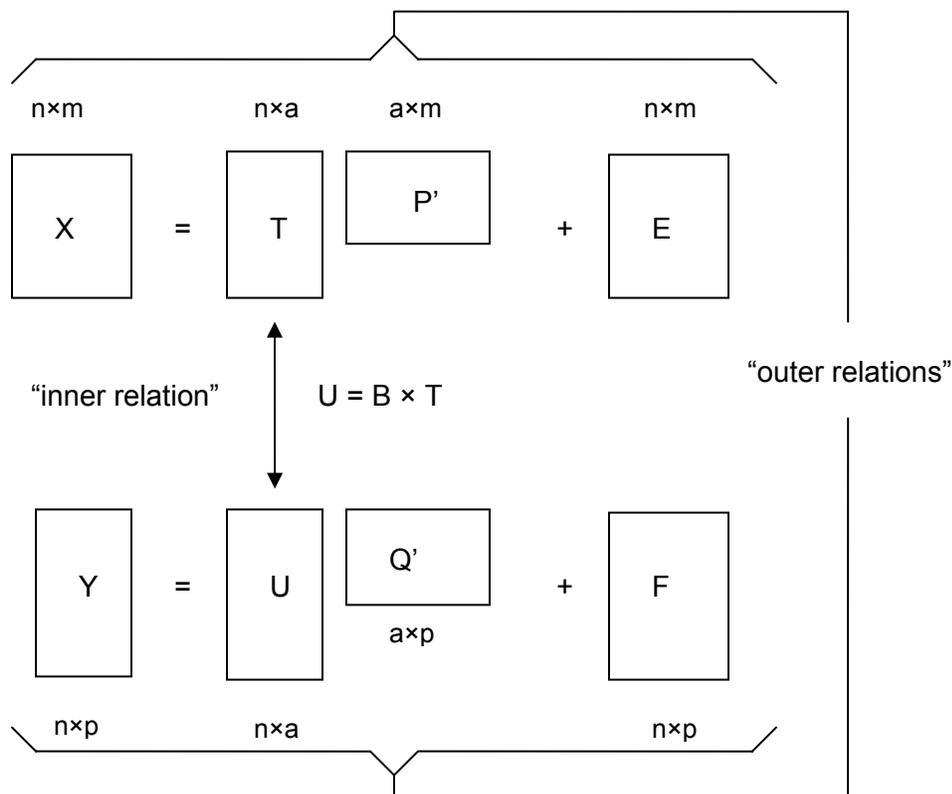
Both Genetic Algorithms (GA) and Partial Least Square (PLS) regression were used in this project for estimation of phycocyanin from *in situ* reflectance of water samples. GA-PLS combines genetic algorithms with partial least squares regression for spectroscopic analysis of material composition in which genetic algorithms are used for the band selection and PLS links the selected bands to the compositional parameters.

#### *PLS regression*

PLS is a standard multivariate regression method developed by Herman Wold (1966a, 1966b). Assuming that the system of interest is driven by a few factors or components, PLS determines a few eigenvectors of the explanatory variables such that the corresponding scores not only explain the variance of the explanatory variables but also have a high correlation to the response variables. A simplified PLS model consists of two outer relations resulting from the eigenstructure decomposition of both the matrix containing explanatory variables (*i.e.*, spectral bands) and the matrix containing response variables (*i.e.*, soil chemistry and mineralogy), and an inner relation that links the resultant score matrices from these two eigenstructure decompositions (Geladi and Kowalski, 1986). The diagram shown in **Figure 5** shows these three relations, where both  $X$  ( $n \times m$ ) and  $Y$  ( $n \times p$ ) represent the explanatory and response variable matrices, respectively. The first outer relation is derived by applying principle component analysis

(PCA) to  $X$ , resulting in the score matrix  $T (n \times a)$  and the loading matrix  $P (m \times a)$  plus an error matrix  $E (n \times m)$ . In the similar way, the second outer relation is derived by decomposing  $Y$  into the score matrix  $U (n \times a)$  and the loading matrix  $Q (p \times a)$  and the error term  $F (n \times p)$ . The prime represents matrix transpose<sup>7</sup>.

The goal of PLS modeling is to minimize the norm of  $F$  while maximizing the covariance between  $X$  and  $Y$  by the inner relation. This inner relation is a multiple linear regression between the score matrices  $U$  and  $T$  in which  $B$  is an  $n \times n$  regression coefficient matrix determined via least square minimization.



**Figure 5: A diagram showing the outer and inner relations in PLS regression**

The selection of the optimal number of PLS components is a key step to obtain a model with good predictive power. The optimum number of components is determined via cross validation. The cross validation is often performed on the calibration samples that are one portion of all available samples and separated from the validation samples, the other portion of the total samples. During the cross validation, the model is increased one PLS component until the prediction on the calibration samples shows that further PLS components do not improve predictive ability for concentration. In the authors' previous work (Li *et al.*, 2006), the cross-validation was performed by leaving out one sample at a time. In this leave-one-out cross validation, a calibration PLS model was built using  $N-1$  samples, and the abundance of the sample left out was then predicted. Given a set of  $N$  calibration samples, this process was repeated  $N$  times so that each

<sup>7</sup> Transpose of a matrix is obtained by exchanging rows and columns of that matrix such that the rows (columns) of the matrix transpose match the columns (rows) of that matrix and vice versa

sample had been left out once. The prediction error sum of squares (PRESS) was used to derive the root-mean-square error of cross-validation (RMSECV) and RMSECV is used as a measure to select the optimal number of PLS components because RMSECV indicates how well the model predicts new samples.

$$PRESS = \sum_{i=1}^N \left( \hat{y}_{(i)} - y_i \right)^2 \quad \text{Equation 1}$$

$$RMSECV = \sqrt{\frac{PRESS_k}{N}} \quad \text{Equation 2}$$

Where  $\hat{y}_{(i)}$  is the predicted value for the sample  $i$ ,  $y_i$  is known concentration of the sample  $i$ , and  $k$  is the number of components used in a PLS model. The selection of the optimal number of PLS components is based on the minimum RMSECV or  $F$ -test on RMSECV (Haadland and Thomas, 1988a, 1988b; Lestander *et al.*, 2003). In addition to leave-one-out, there are several other cross validation methods such as Venetian blinds, contiguous blocks and random sets (Wise *et al.*, 2004).

#### *Genetic algorithm*

Genetic algorithms (GA) are a computer model that simulates natural selection (Forrest, 1993). In a genetic algorithm, a population of individuals is created automatically and typically stored as binary strings in a computer memory. Over time, the functionality of an individual is evolved by computer operations with use of the mechanisms of natural variation, selection and inheritance. Because the ability of genetic algorithms to simulate a natural evolution of an individual, GA is well suitable for solving variable subset selection problems (Jouan-Rimbaud *et al.*, 1995; Banglore *et al.*, 1996; Schaffer *et al.*, 1996; Ding *et al.*, 1998; Gourvéneec *et al.*, 2004).

A simple genetic algorithm includes at least five components: encoding, population initialization, individual selection, crossover and mutation. These GA components have been described in many references (e.g., Forrest, 1993; Mitchell, 1998; Coley, 2003). A brief summary of each is in APPENDIX C.

#### *GA-PLS Application*

The software used in this study is a group of Matlab programs (Leardi and González, 1998). The value for GA-PLS parameters is empirically determined by the developer. The major risk of using GA is over fitting because of too many variables (*i.e.*, spectral bands). To minimize this risk, the program was set using the following features: (1) the parameters were set with the highest elitism, meaning a very limited population size and high mutation rate were used to assure a fast response increase and to find a good solution at an early stage of the process (Leardi, 2000); (2) the model was set to be determined after 100 independent, short GA runs; and (3) each run's frequency of selection of the variables was set to be a weighted average between the frequency of selection of the variables in the starting run and the previous run. In this way, each run is able to "learn" information from the previous runs; a moving average (window size 3) is applied to the frequency of selection of the variables to take into account of high spectral correlation and to assure that the highly correlated spectral bands are selected together.

The fitness function with which the individuals are subject to evaluation is the percentage of predicted variance of a constituent abundance, defined as:

$$100 - \left\{ \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / n}{\sum_{i=1}^n (y_i - \bar{y}_i)^2 / k} \right\} \times 100 \quad \text{Equation 3}$$

where  $\hat{y}_i$  is the predicted value of a sample  $i$ ,  $n$  is the number of samples to be considered,  $k = n-1$  in the case of cross-validation. The fitness of the individuals indicates the prediction power of the selected combination of spectral variables. The final model is picked via a stepwise regression, and the variables are selected in terms of their frequency. The plot of the response (% cross-validated explained variance) vs. the number of variables in the model is referenced to determine the final model.

The result of GA-PLS modeling is evaluated based on the root mean square error in the dataset (RMSE). RMSEP (root-mean-square error of prediction) is written as:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}} \quad \text{Equation 4}$$

where  $N$  is the number of samples in the dataset set, and  $y$  denotes a constituent variable.

#### *Algorithm Validation*

The retrieved algorithm predicting phycocyanin from surface water reflectance spectra was validated using a second subset (*i.e.*, a validation set comprised of 10% of the total spectra collected): the retrieved algorithm was applied to the validation set's surface water reflectance spectra to obtain a predicted phycocyanin concentration value, and the predicted phycocyanin concentration values were plotted against actual measured phycocyanin concentrations to obtain a coefficient of determination and a root-mean-square error (RMSE).

#### *Algorithm Transferability*

The algorithm's transferability was tested by using one reservoir as a training set and another reservoir as a validation set.

### **3.5.2 Empirical, Semi-Empirical, and Bio-optical Models**

Studies by Dekker (1993), Schalles and Yacobi (2000), and Simis *et al.* (2005) have resulted in the development of empirical models for the estimation of phycocyanin from multi- and hyper-spectral reflectance data. The applicability of these algorithms to data outside of their calibration data sets can be improved upon by applying a backscattering model and absorption coefficients that are specific to certain water quality constituents (*e.g.*, colored organic matter, chlorophyll *a*, and total suspended solids). The combination of both an empirical relationship and the inherent optical properties of the water itself and the water's constituents results in the creation of semi-empirical models. Unlike empirical and semi-empirical models, bio-optical models utilize only the inherent properties of the water and water constituents. As each constituent will attenuate light

differently, these spectral signatures of the pure end-member can be used to estimate concentrations of each constituent from reflectance data.

#### *Algorithm Selection*

The previously developed algorithms chosen for evaluation represent a range of model types from empirical to semi-empirical models (**Table 2**). As recommended by previous research, regression analysis was used to calibrate these algorithms using the spectral and field data collected from our study. This analysis resulted in the calculation of coefficients for each empirical model for each reservoir.

#### *Algorithm Validation*

The calibrated algorithms for predicting phycocyanin from surface water reflectance spectra will be validated using a second subset (*i.e.*, a validation set comprised of field spectra data taken in 2006): the calibrated algorithms will be applied to the validation set's surface water reflectance spectra to obtain a predicted phycocyanin concentration value, and the predicted phycocyanin concentration values will be plotted against actual measured phycocyanin concentrations to obtain a coefficient of determination and an RMSE.

#### *Algorithm Transferability*

The calibrated algorithms' transferability will be tested by using two methods: (1) one reservoir will be used as a training set and another reservoir as a validation set; and (2) an aggregated dataset will be used as a training set and individual reservoirs will be used as a validation set.

**Table 2: Summary of empirical and semi-empirical algorithms tested.  $R_{xxx}$  refers to the reflectance measurement a wavelength xxx.**

Phycocyanin Algorithm	Reference
<i>Empirical Models</i>	
1. [PC] = 0.5( $R_{600} + R_{648}$ ) - $R_{624}$	Dekker 1993
2. [PC] = 0.5( $R_{591-609} + R_{647-656}$ )- $R_{619-638}$	Dekker 1993
3. [PC] =0.5( $R_{600} + R_{647}$ ) - $R_{628}$	Dekker 1993
4. [PC] = $R_{650} / R_{625}$	Schalles and Yacobi, 2000
5. [PC] = $R_{647} / R_{628}$	Schalles and Yacobi 2000
6. [PC] = $R_{628}$	Modified; Gitelson <i>et al.</i> 1995
7. [PC] = $R_{704} / R_{628}$	Modified; Simis <i>et al.</i> 2005

*Semi-Empirical Model*

$$[PC] = a_{pc}(620) = (\{[R_{709} / R_{620}] \times [a_w(709) + b_b]\} - b_b - a_w(620)) \times \delta^{-1} - [\varepsilon \times a_{chl}(665)]$$

Where: Adapted from Simis *et al.* 2005

$a_{pc}(620)$  = absorption of phycocyanin at 620 nm which can be related to phycocyanin concentrations

$R(\lambda)$  = reflectance value at a specified wavelength

$a_w(\lambda)$  = pure water absorption coefficients at specified locations

$$a_w(709) = 0.70 \text{ m}^{-1} \text{ (Buiteveld et al., 1994)}$$

$$a_w(620) = 0.30 \text{ m}^{-1} \text{ (Pope and Fry, 1997)}$$

$b_b$  = backscattering coefficient (as obtained for chlorophyll *a*)

$\delta$  = 0.89; a constant derived from the linear least-squares fit of measured versus predicted phycocyanin absorption (Simis *et al.*, 2005)

$a_{chl}(665)$  = absorption of chlorophyll *a* at 665 nm, determined using the equation  $(\{[R_{709} / R_{665}] \times [a_w(709) + b_b]\} - b_b - a_w(665)) \times \gamma$

$\varepsilon$  = 0.24; correction factor to define absorption of chlorophyll *a* at 620 nm relative to 665 nm, derived from Lake Loosdrecht data (Simis *et al.*, 2005)

## 4.0 Results and Discussion

Results show the progress of work as of October 2006, focusing on data collected on Geist and Morse Reservoirs in 2005. Unfortunately, due to a problem with instrument calibration, 2005 field reflectance data for Eagle Creek Reservoir were not able to be used for algorithm development, calibration, or validation.

However, work is on-going and, as a result of 2006 data collections of over 300 field spectra and ground truth samples, we anticipate that algorithm development, calibration, and validation results will continue to improve and be extended to Eagle Creek Reservoir.

### 4.1 Water Quality Data

*Water Clarity*– For all reservoirs, Secchi disk depths averaged 70 cm with the greatest depths seen in Morse Reservoir. This corresponded with the lowest TSS measurement also being recorded in Morse Reservoir; however, the lowest turbidity measurement was recorded in Eagle Creek Reservoir (**Table 3**).

*Dissolved Constituents*– For all reservoirs, the mean TDS and DOC were 0.277 g-L<sup>-1</sup> and 4.3 mg C-L<sup>-1</sup>, respectively. The highest TDS was recorded in Geist Reservoir and the highest DOC was recorded in Morse Reservoir. Eagle Creek Reservoir had the lowest concentrations of both TDS and DOC (**Table 3**).

*Pigments*– The mean concentrations of chlorophyll *a* and phycocyanin for all reservoirs were 59.7 ppb and 73.6 ppb, respectively. The highest concentration of chlorophyll *a* was recorded in Morse Reservoir and the highest phycocyanin concentration was measured in Geist Reservoir. Morse Reservoir showed the highest chlorophyll *a*-to-phycocyanin ratio, showing that more chlorophyll *a* than phycocyanin was present in the water column (**Table 3**).

*Nutrients*– For all reservoirs the mean Total P and Total N (TKN + nitrate) concentrations were 79.6 µg P-L<sup>-1</sup> and 1.7 mg N-L<sup>-1</sup>, respectively (**Table 3**).

DEVELOPING A SURVEY TOOL FOR THE RAPID ASSESSMENT OF BLUE-GREEN ALGAE IN  
CENTRAL INDIANA'S RESERVOIRS

**Table 3: Summary statistics of water quality parameters.**

Parameter	Mean	Median	Minimum	Maximum	$\sigma$	N
<i>ALL Reservoirs</i>						
Secchi Depth (cm)	70	68	30	135	28	96
Turbidity (NTU)	7.4	5.6	2.3	30.0	4.5	96
TSS (mg-L <sup>-1</sup> )	15.3	12.0	4.4	54.4	9.8	96
TDS (g-L <sup>-1</sup> )	0.277	0.267	0.242	0.369	0.035	96
DOC (mg C-L <sup>-1</sup> )	4.3	4.4	0.5	5.2	0.8	91
TOC (mg C-L <sup>-1</sup> )	6.1	6.1	4.4	10.3	1.0	96
Chlorophyll a (µg-L <sup>-1</sup> ppb)	59.7	50.9	18.0	168.6	32.2	92
Phycocyanin (µg-L <sup>-1</sup> ppb)	73.6	76.6	2.0	185.1	43.6	92
Chl a-to-PC Ratio	1.8	0.9	0.3	12.7	2.6	89
Total P (µg P-L <sup>-1</sup> )	79.6	70.0	23.0	204.0	42.7	95
Total N (mg N-L <sup>-1</sup> )	1.7	1.7	0.6	2.8	0.4	98
<i>Eagle Creek Reservoir</i>						
Secchi Depth (cm)	69	73	35	86	12	34
Turbidity (NTU)	5.6	4.8	0.5	5.1	0.7	34
TSS (mg-L <sup>-1</sup> )	11.2	10.4	7.2	24.4	3.5	34
TDS (g-L <sup>-1</sup> )	0.248	0.245	0.242	0.280	0.008	34
DOC (mg C-L <sup>-1</sup> )	4.5	4.6	0.5	5.11	0.7	34
TOC (mg C-L <sup>-1</sup> )	6.4	6.3	5.4	9.4	0.8	34
Chlorophyll a (µg-L <sup>-1</sup> ppb)	52.1	46.0	26.4	107.1	23.8	33
Phycocyanin (µg-L <sup>-1</sup> ppb)	80.1	83.3	24.6	130.5	25.9	33
Chl a-to-PC Ratio	0.7	0.6	0.3	1.7	0.3	32
Total P (µg P-L <sup>-1</sup> )	66.0	59.0	28.0	152.0	27.7	33
Total N (mg N-L <sup>-1</sup> )	1.7	1.7	1.5	2.4	0.2	35
<i>Geist Reservoir</i>						
Secchi Depth (cm)	49	45	30	75	12	30
Turbidity (NTU)	10.3	9.8	7.0	18.0	2.0	30
TSS (mg-L <sup>-1</sup> )	20.4	19.4	13.2	29.2	4.2	30
TDS (g-L <sup>-1</sup> )	0.322	0.306	0.300	0.369	0.028	30
DOC (mg C-L <sup>-1</sup> )	3.9	4.2	0.5	4.86	1.0	30
TOC (mg C-L <sup>-1</sup> )	6.2	5.9	4.4	10.3	1.2	30
Chlorophyll a (µg-L <sup>-1</sup> ppb)	71.3	64.4	34.7	118.9	26.0	29
Phycocyanin (µg-L <sup>-1</sup> ppb)	96.2	100.4	25.2	185.1	43.8	30
Chl a-to-PC Ratio	0.9	0.8	0.4	2.9	0.5	29
Total P (µg P-L <sup>-1</sup> )	111.2	113.0	38.0	191.0	39.1	30
Total N (mg N-L <sup>-1</sup> )	2.0	2.0	0.6	2.8	0.4	31
<i>Morse Reservoir</i>						
Secchi Depth (cm)	92	90	35	135	36	31
Turbidity (NTU)	6.7	4.6	2.3	30.0	6.4	31
TSS (mg-L <sup>-1</sup> )	15.1	8.4	4.4	54.4	15.0	31
TDS (g-L <sup>-1</sup> )	0.267	0.268	0.260	0.281	0.005	31
DOC (mg C-L <sup>-1</sup> )	4.4	4.2	3.9	5.2	0.4	26
TOC (mg C-L <sup>-1</sup> )	5.6	5.1	4.6	7.7	0.9	31
Chlorophyll a (µg-L <sup>-1</sup> ppb)	57.2	35.6	18.0	168.6	42.9	29
Phycocyanin (µg-L <sup>-1</sup> ppb)	41.8	28.6	2.0	135.1	43.4	28
Chl a-to-PC Ratio	4.0	1.6	0.7	12.7	3.8	27
Total P (µg P-L <sup>-1</sup> )	64.0	50.0	23.0	204.0	44.3	31
Total N (mg N-L <sup>-1</sup> )	1.5	1.5	0.6	2.3	0.3	32

## 4.2 GA-PLS Algorithms

### 4.2.1 Band selection results

Data were normalized by two methods: mean centering in which the sample mean was subtracted from the data and autoscaling which includes a second step in which mean centered data are divided by the standard deviations. The former method can remove the effect of path radiance, and the latter can remove or compress the effect of path radiance and illumination. The results from GA modeling of *in situ* reflectance from Geist and Morse Reservoirs indicated that spectral bands selected for individual reservoirs differ (Table 4).

**Table 4: Spectral bands selected by GA-PLS modeling.**

Reservoir	Mean Centered	Autoscaled
Geist	628.8	563.2
	694.6	628.8
	704.0	933.5
Morse	694.6	647.6
	704.0	704.0

For Geist Reservoir, 3 spectral bands were selected with mean centered data and autoscaled data, respectively. The band at 628.8 nm was selected because it is located in the region where phycocyanin has strong absorption. Other bands shorter than or longer than 628.8 nm were selected though they are not diagnostic to the presence of phycocyanin. We interpret that the selection of the non-causal bands such as 563.2 nm, 694.6, 704.0, and 933.5 nm was due to the interaction of chlorophyll's spectral features with those related to phycocyanin. Although the selected bands from mean centered data differ from those resulting from autoscaled data in the number and spectral locations, the difference is insignificant.

For Morse Reservoir, two spectral bands were selected with mean centered and autoscaled data, respectively. Compared with the selected bands for Geist Reservoir, fewer spectral bands were selected and none of the spectral bands diagnostic to phycocyanin were selected. The selected bands are all relevant to the presence of chlorophyll *a*. The selection of the band at 674.6 nm is due to the strong chlorophyll *a* absorption of red light at about 675.0 nm, while the selection of the band 704.0 nm is due to its location in the reflectance peak around 690-700 nm, caused by algal-cell scattering and a combined effect of pigment and water absorption. We interpret that this band selection for the phycocyanin estimation is due to a high correlation between chlorophyll *a* and phycocyanin. Examining a correlation matrix of water constituent abundances indicates a high correlation between chlorophyll *a* and phycocyanin. Another reason is the dominance of chlorophyll *a* over phycocyanin in driving spectral variations. This phenomenon is most notable in Morse Reservoir samples where the mean chlorophyll *a*-to-phycocyanin ratio was 4.0, showing that samples had 3x the amount of chlorophyll *a* as phycocyanin. Comparatively, Geist Reservoir's mean chlorophyll *a*-to-phycocyanin ratio was less than 1.0, showing that samples had more phycocyanin than chlorophyll *a*.

## 4.2.2 Regression models and results

The regression models for the phycocyanin estimation from reflectance are shown in **Equation 5** with mean centered data and **Equation 6** with autoscaled data for Geist Reservoir, and **Equation 7** with mean centered data and **Equation 8** with autoscaled data for Morse Reservoir. The estimated and measured phycocyanin concentrations are plotted in **Figure 6** for Geist and **Figure 7** for Morse with the corresponding coefficient of determinations and RMSE. Note: Due to a sensor calibration error, spectra collected on Eagle Creek Reservoir were not used for GA-PLS algorithm development.

$$[PC_{\text{Geist}}] \text{ (ppb)} = 1.5797R_{628.8} - 0.3967R_{694.6} - 0.0458R_{704.0} \quad \text{Mean Centered} \quad \text{Equation 5}$$

$$[PC_{\text{Geist}}] \text{ (ppb)} = 1.1383R_{628.8} - 1.0555R_{694.6} - 0.0138R_{704.0} \quad \text{Autoscaled} \quad \text{Equation 6}$$

$$[PC_{\text{Morse}}] \text{ (ppb)} = -1.3812R_{694.6} + 0.2230R_{704.0} \quad \text{Mean Centered} \quad \text{Equation 7}$$

$$[PC_{\text{Morse}}] \text{ (ppb)} = -1.6889R_{647.6} + 0.3726R_{704.0} \quad \text{Autoscaled} \quad \text{Equation 8}$$

Where PC is phycocyanin and  $R_{xxx.x}$  denotes the reflectance at the wavelength xxx.x.

For Geist Reservoir, the GA-PLS modeling resulted in a coefficient of determination 0.6637 and a RMSE of 21.46 ppb when using the mean centered data (**Figure 6(A)**), and a coefficient of determination 0.5704 and a RMSE of 23.50 ppb when using autoscaled data (**Figure 6(B)**). The estimation for the sample with the most extreme phycocyanin abundance is poor. The reason for this is either the spectral reflectance was saturated, thus not sensitive to the change in phycocyanin abundance, or an error exists in laboratory extraction and analysis of this sample. Elimination of this extreme sample improves coefficients of determination significantly, resulting in an increase to 0.7655 when using mean centered data (**Figure 6(C)**) and to 0.7039 when using autoscaled data (**Figure 6(D)**).

For Morse Reservoir, a coefficient of determination 0.8119 and a RMSE of 17.74 ppb were obtained with the GA-PLS modeling of mean centered data (**Figure 7(A)**), and the corresponding values are 0.8018 and 18.30 ppb when using autoscaled data (**Figure 7(B)**). The removal of a sample with large estimated error improved the coefficients of determination, which increased to 0.9151 using mean centered data (**Figure 7(C)**) and 0.8958 using autoscaled data (**Figure 7(D)**).

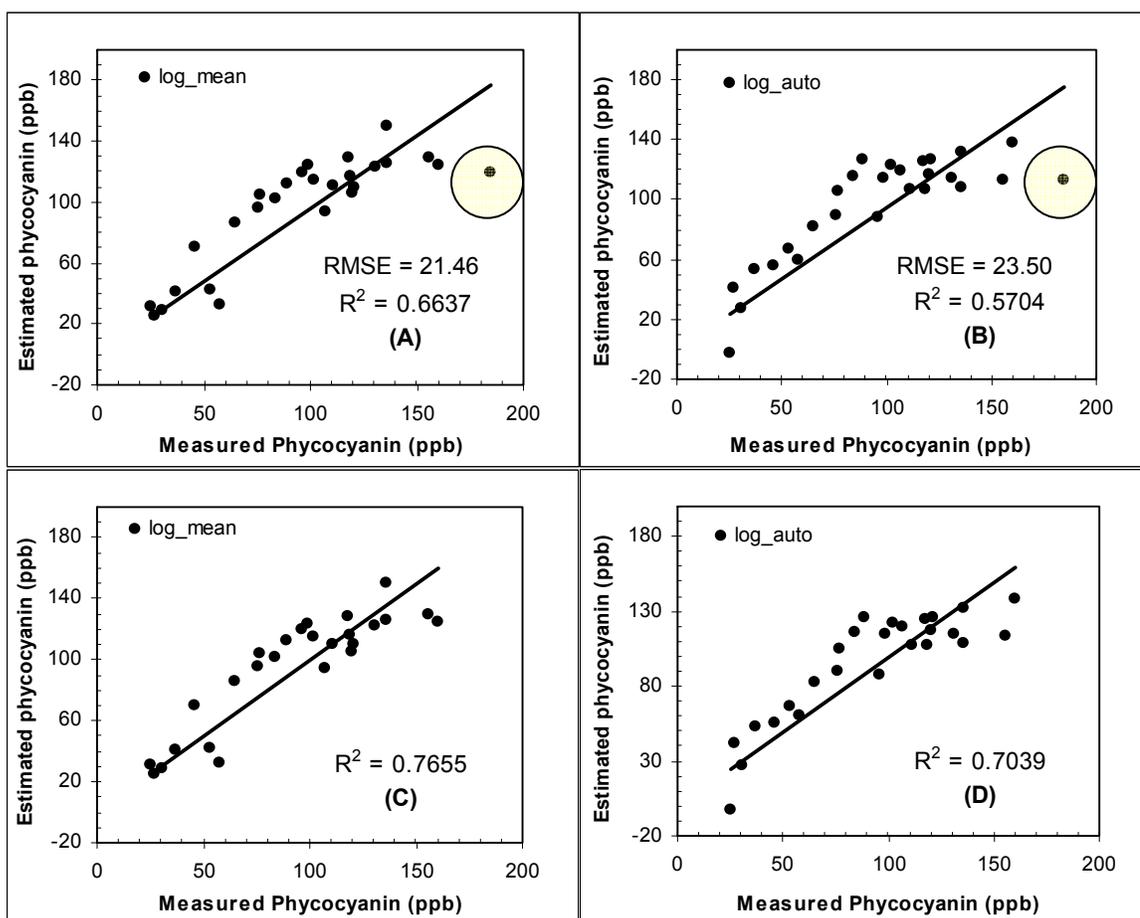


Figure 6: Linear regressions (black line) and root mean square error (RMSE) of measured phycocyanin and estimated phycocyanin using developed GA-PLS algorithms for Geist Reservoir: (A) Equation 5; (B) Equation 6; (C) Equation 5 after elimination of high data point highlighted by the yellow circle; (D) Equation 6 after elimination of high data point highlighted by the yellow circle yielded high squared correlation coefficients ( $r^2 \geq 0.57$ ).

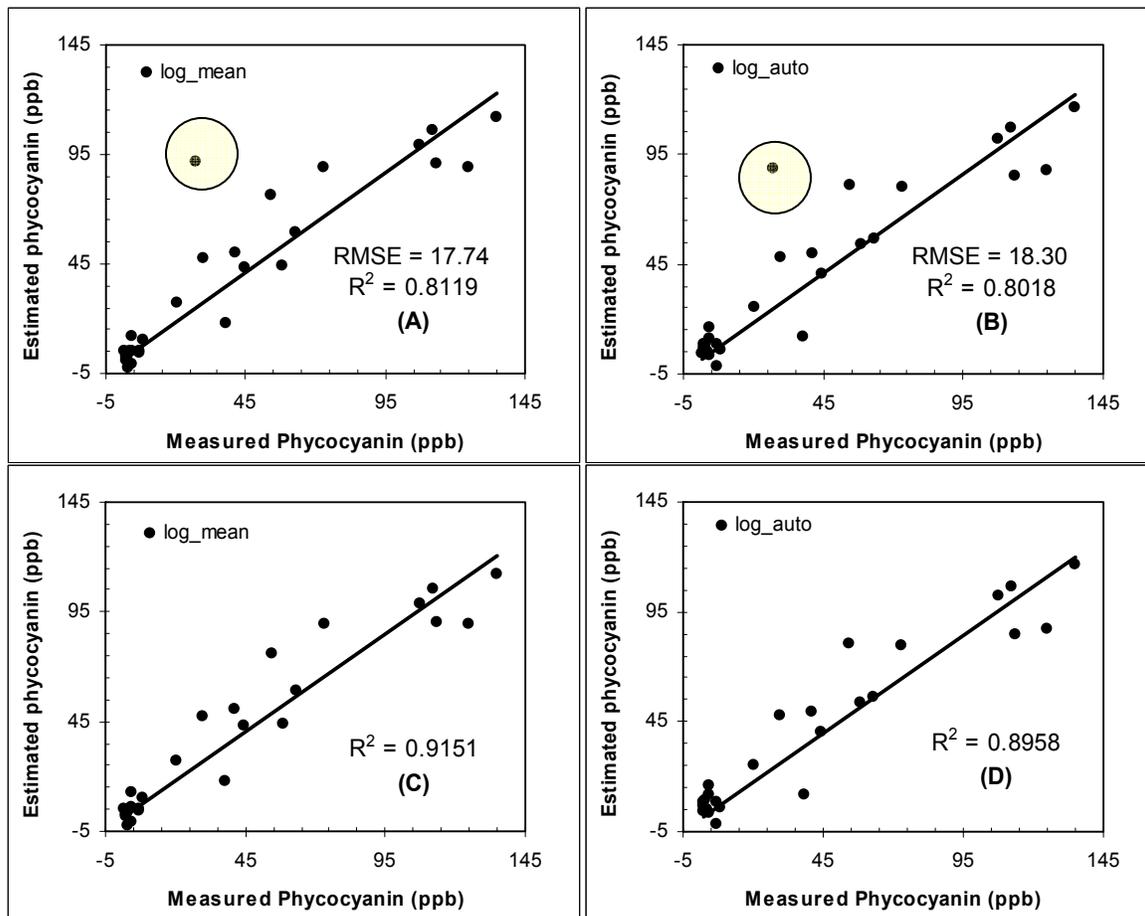


Figure 7: Linear regressions (black line) and root mean square error (RMSE) of measured phycocyanin and estimated phycocyanin using developed GA-PLS algorithms for Morse Reservoir: (A) Equation 7; (B) Equation 8; (C) Equation 7 after elimination of data point with high estimated error highlighted by the yellow circle; (D) Equation 8 after elimination of data point with high estimated error highlighted by the yellow circle yielded high squared correlation coefficients ( $r^2 \geq 0.80$ ).

### 4.3 Performance of Empirical and Semi-Empirical Algorithms

Preliminary results from these studies will focus on the best performing algorithms and will not include a discussion on the algorithm derivation. (Information on the derivation of these algorithms can be found in the referenced literature.) This aspect of our research is the focus of two Masters Theses by Kaylan L. Randolph and Rebecca E. Sengpiel. The full results of their research will be reported at the conclusion of their program.

#### 4.3.1 Field Spectra

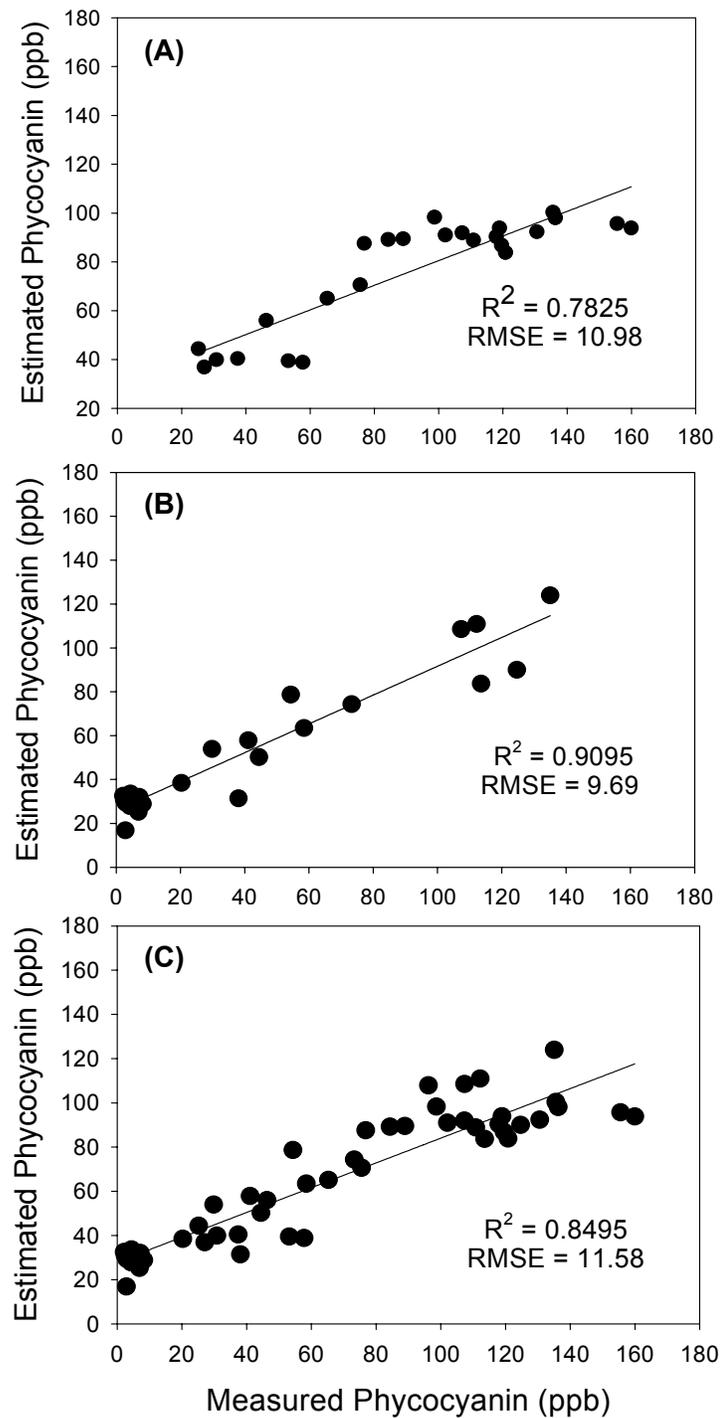
Simis *et al.* (2005) increased the predictive power of the simple ratio  $R_{709}/R_{620}$  for phycocyanin estimation in Case II waters by incorporating the measured specific absorption coefficient of phycocyanin at 620 nm ( $a_{pc}(620)$ ) (Simis *et al.*, 2005), pure water coefficients at specified wavelengths  $a_w(\lambda)$  (Buiteveld, *et al.*), backscattering coefficients retrieved from reflectance at 778 nm ( $b_b$ ) (Gons, 1999), and the chlorophyll  $a$

absorption coefficient at 620 nm ( $a_{chl}(620)$ ) (**Table 2: Semi-Empirical Model**). The result of the semi-empirical algorithm was the absorption coefficient for phycocyanin at wavelength location 620 nm ( $a_{PC}(620)$ ). The division of the absorption coefficient  $a_{PC}(620)$  by the specific absorption coefficient ( $a_{PC}^*(620)$ ) of phycocyanin yielded an estimation of phycocyanin concentration.

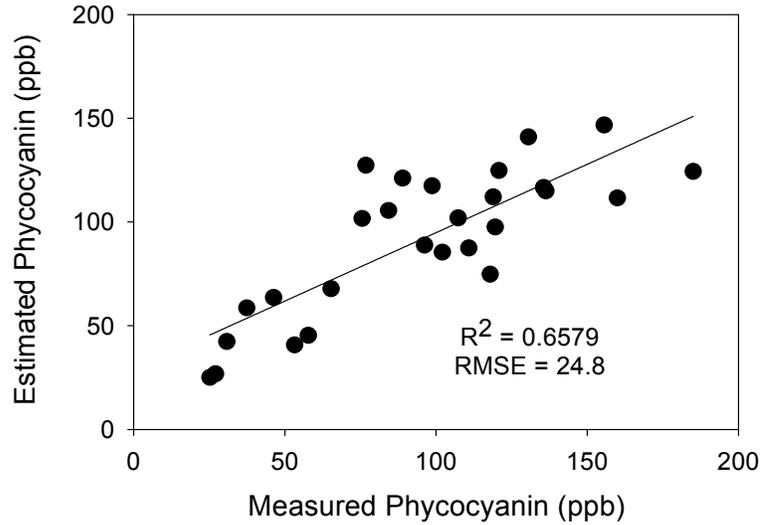
Results showed that prediction of phycocyanin concentrations, within the range of 0 to 140 ppb, for Geist and Morse Reservoirs resulted in  $r^2$  values of 0.7825 (n=25), RMSE = 10.98 ppb and 0.9095, RMSE = 9.69, (n=25), respectively (**Figure 8A and B**). Application of the Simis *et al.* (2005) algorithm to an aggregated dataset resulted in  $r^2$  values of 0.8495, RMSE = 11.58 (**Figure 8C**). The success of this technique demonstrates the need to further explore the use of semi-empirical and bio-optical models for estimating phycocyanin from field spectra collections. Currently, data are being further analyzed to optimize their applicability to Indianapolis' reservoirs. This continued analysis includes algorithm validation and a test of algorithm robustness.

#### 4.3.2 Airborne Spectra

Estimated pigment concentrations were retrieved from the airborne reflectance spectra by calibrating phycocyanin algorithms (**Table 2: Empirical Models**) to Geist Reservoir AISA data and ground truth phycocyanin concentrations. However, these algorithms were developed from data collected using field spectroradiometers. As band resolution on the AISA sensor is coarser than sensors from which these algorithms were developed, algorithms were modified to account for the band resolution of the AISA sensor (**Table 2: Empirical Models 6 and 7**). Of the algorithms tested, a simple model of reflectance at 628 nm to phycocyanin performed the best (Gitelson *et al.*, 1995; **Table 2: Empirical Model 6**). Results showed a strong correlation between measured phycocyanin and estimated phycocyanin from airborne acquired spectra collected from Geist Reservoir over a phycocyanin range of 15-225 ppb:  $r^2 = 0.6579$  (n=27), RMSE = 24.8 ppb (**Figure 9**). Application of backscattering and absorption coefficients introduced by Simis *et al.* (2005) did not significantly improve the algorithms' performance. This demonstrates that even though algorithms were developed on different sensors and different platforms (i.e., hand-held field sensor versus an airborne sensor), techniques for estimating phycocyanin from reflectance spectra are robust. Currently, AISA imagery for Morse Reservoir is being processed. Data are being further analyzed to optimize algorithms.



**Figure 8: Linear regressions (black line) and root mean square errors (RMSE) of measured phycocyanin and estimated phycocyanin using the Simis *et al*, (2005) semi-empirical model (Table 2) for (A) Geist Reservoir and (B) Morse Reservoir, and (C) the aggregated dataset of both Geist and Morse Reservoirs yielded  $r^2 \geq 0.78$  and low RMSE ( $RMSE \leq 11.58$ ).**



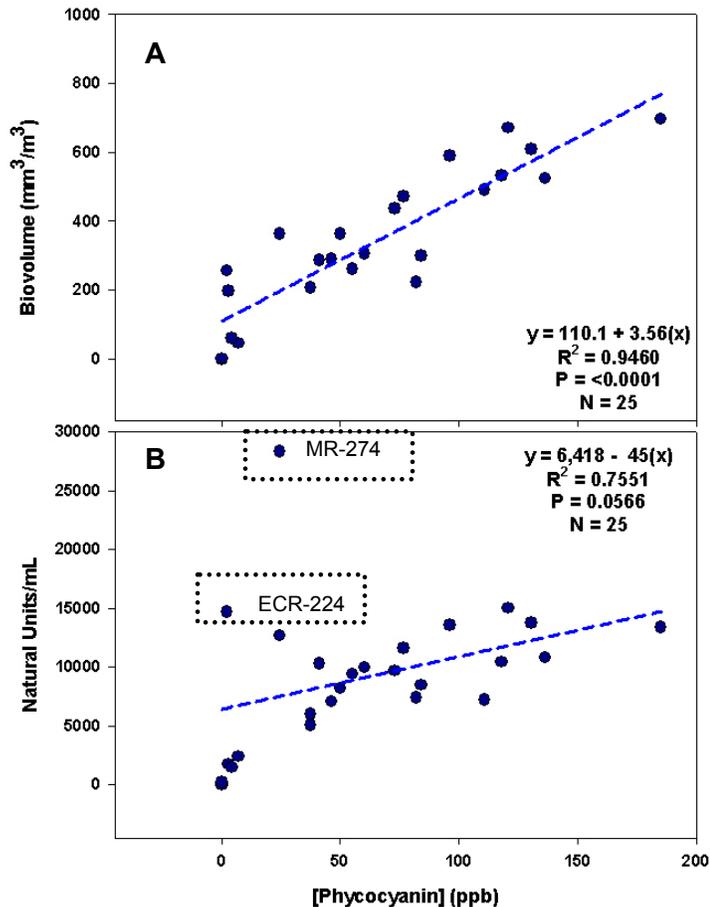
**Figure 9: Linear regression (black line) and root mean square error (RMSE) of measured phycocyanin and estimated phycocyanin using the Gitelson *et al.* (1995; Table 2: Empirical Model 6) algorithm on Geist Reservoir yielded an  $r^2 = 0.66$  and a low RMSE (RMSE = 24.8).**

#### **4.4 Relationship between Phycocyanin and Blue-green Abundance**

To determine the relationship between phycocyanin pigment concentrations and measures of blue-green algal abundance, a subset of 25 samples across all three reservoirs underwent phytoplankton identification, enumeration, and biovolume calculations. These data show a strong correlation between ground truth phycocyanin concentrations and biovolume measurements ( $r^2 = 0.9460$ ,  $p < 0.0001$ ) and a moderately strong relationship between phycocyanin concentrations and phytoplankton counts as natural units- $\text{mL}^{-1}$  ( $r^2 = 0.7551$ ,  $p = 0.0566$ ; **Figure 10**). While most management decisions are based on the number of cells or natural units, biovolume is a better measure of the actual biomass of blue-green cells as blue-green algae (and most algae) are not uniform in size and shape. This variation in algal size and shape is evident in samples with high concentrations of natural units but relatively low concentrations of phycocyanin. This disparity was caused by the prevalence of small blue-green taxa such as *Merismopedia minima* and *Pseudanabaena limnetica*, which contribute little to overall biovolume and phycocyanin concentrations but can dominate counts. In sample ECR-224, small taxa accounted for 68% of the natural units but only 16% of the biovolume; and in sample MR-274 small taxa accounted for 77% of the natural units but only 21% of the biovolume. Removing these two data points resulted in a stronger relationship between phycocyanin and natural units ( $r^2 = 0.9104$ ,  $p < 0.0001$ ). As such, both biovolume and natural unit measures for blue-green abundance can be reasonably estimated from phycocyanin concentrations (**Equation 9**, **Equation 10**), allowing managers a method in which to compare phycocyanin concentrations to blue-green algal abundance.

Blue-green Biovolume ( $\text{mm}^3/\text{m}^3$ ) =  $110.1 + (3.56 \times [\text{PC}] (\text{ppb}))$  Equation 9

Blue-green Natural Units / mL =  $3,629 + (67 \times [\text{PC}] (\text{ppb}))$  Equation 10



**Figure 10: Relationship between phycocyanin concentrations and measures of blue-green algal abundance. (A) Shows a strong correlation between phycocyanin and biovolume, while (B) shows a poor relationship between phycocyanin and blue-green algal counts. The data points in the boxes show samples (ECR-224 and MR-274) with high algal counts but low phycocyanin concentrations. This deviation from the trend is caused by an abundance of small blue-green taxa that do not contribute much to pigment concentrations but dominate counts.**

#### 4.5 Derivation of spatial maps of phycocyanin abundances

Applying the algorithms developed using empirical (GA-PLS and Gitelson *et al.*, 1995) and semi-empirical (modified Simis *et al.* 2005) modeling to field spectra, spatial maps were created for the reservoirs.

##### 4.5.1 Eagle Creek Reservoir

Due to an error in sensor calibration, spectra collected on Eagle Creek Reservoir were not utilized for mapping. However, actual measured field data (ground truth data) were

used to create an interpolated map of phycocyanin in the reservoir (**Figure 11**) and chlorophyll *a* (APPENDIX A).

#### 4.5.2 Geist Reservoir

##### *GA-PLS Algorithms*

Applying the best fit GA-PLS algorithm, Equation 5 (mean centered), after elimination of the high data point (**Figure 6(C)**) to the field spectra collected at each reservoir site resulted in an interpolated map of estimated phycocyanin concentration and distribution. This can be compared to ground truth measurements of phycocyanin (**Figure 12 and Figure 13**). An interpolated map of ground truth chlorophyll *a* data can be found in APPENDIX A.

##### *Empirical and Semi-Empirical Algorithms*

Field Spectra– Applying the modified Simis *et al.* (2005; **Table 2: Semi-Empirical Model**) algorithm to field spectra data from Geist Reservoir resulted in an interpolated map of estimated phycocyanin concentration and distribution (**Figure 14**).

Airborne Spectra– Applying the modified Gitelson *et al.*, (1995; **Table 2: Empirical Model 6**) to the over 5 million airborne spectra that make up the Geist Reservoir image resulted in a map of estimated phycocyanin concentration and distribution over the surface of the reservoir based on a pixel size of 1 m<sup>2</sup> (**Figure 15**).

#### 4.5.3 Morse Reservoir

Applying the best fit algorithm, **Equation 7** (mean centered), after elimination of the data point with high standard error (**Figure 7(C)**) to the field spectra collected at each reservoir site resulted in an interpolated map of phycocyanin concentration and distribution. This can be compared to ground truth measurements of phycocyanin (**Figure 16 and Figure 17**). An interpolated map of ground truth chlorophyll *a* data can be found in APPENDIX A.

##### *Empirical and Semi-Empirical Algorithms*

Field Spectra– Applying the modified Simis *et al.* (2005; **Table 2: Semi-Empirical Model**) algorithm to field spectra data from Morse Reservoir resulted in an interpolated map of estimated phycocyanin concentration and distribution (**Figure 18**).

Airborne Spectra– AISA imagery for Morse Reservoir are still being processed; therefore, no map is available at this time.

While similar maps of phycocyanin concentration and distribution can be created using actual, extracted and fluorometrically measured phycocyanin concentrations, the map created using field spectra and a calibrated algorithm can be made in a few hours after collecting field spectra (**Figure 12 and Figure 16**) while a map created from extracted and fluorometrically measured phycocyanin or blue-green algal counts will take at least 48 hours (usually a week) to generate after collecting water samples (**Figure 13 and Figure 17**).

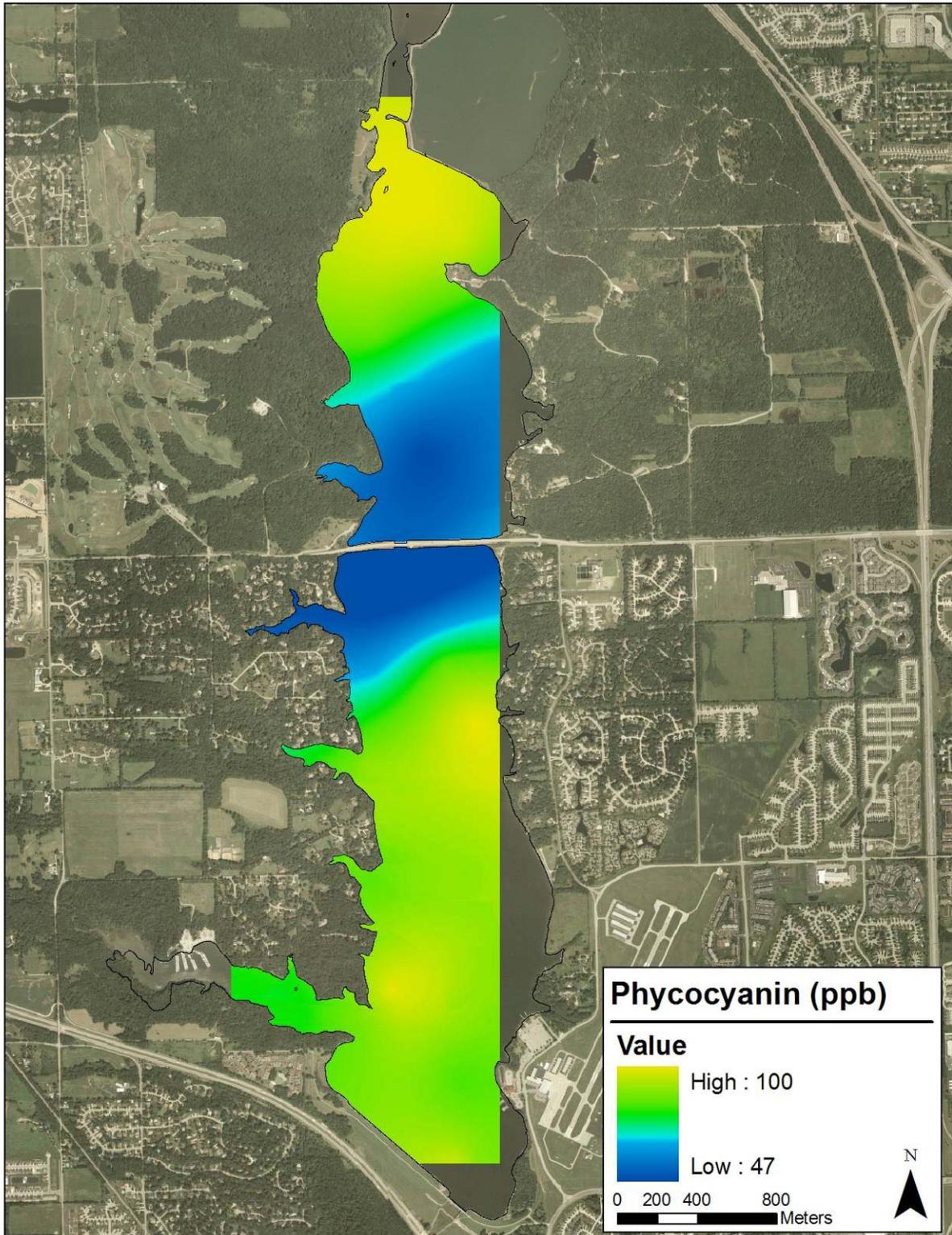


Figure 11: Interpolated map of ground truth phycocyanin concentration in Eagle Creek Reservoir on September 6, 2005.

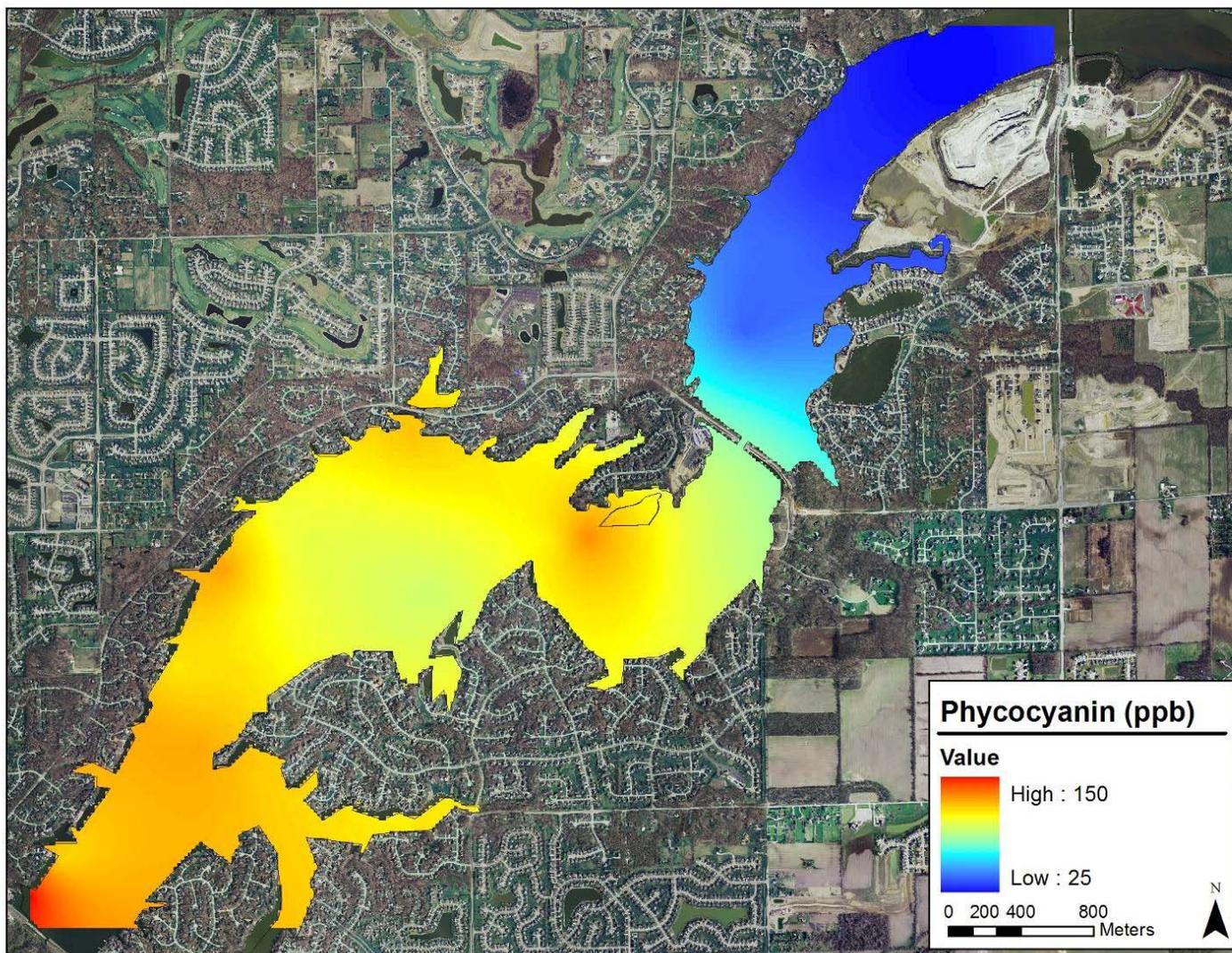


Figure 12: Interpolated map of phycocyanin concentration in Geist Reservoir on September 6, 2005 estimated from GA-PLS derived algorithm ( $r^2 = 0.7655$ ; Equation 5) applied to field spectra.

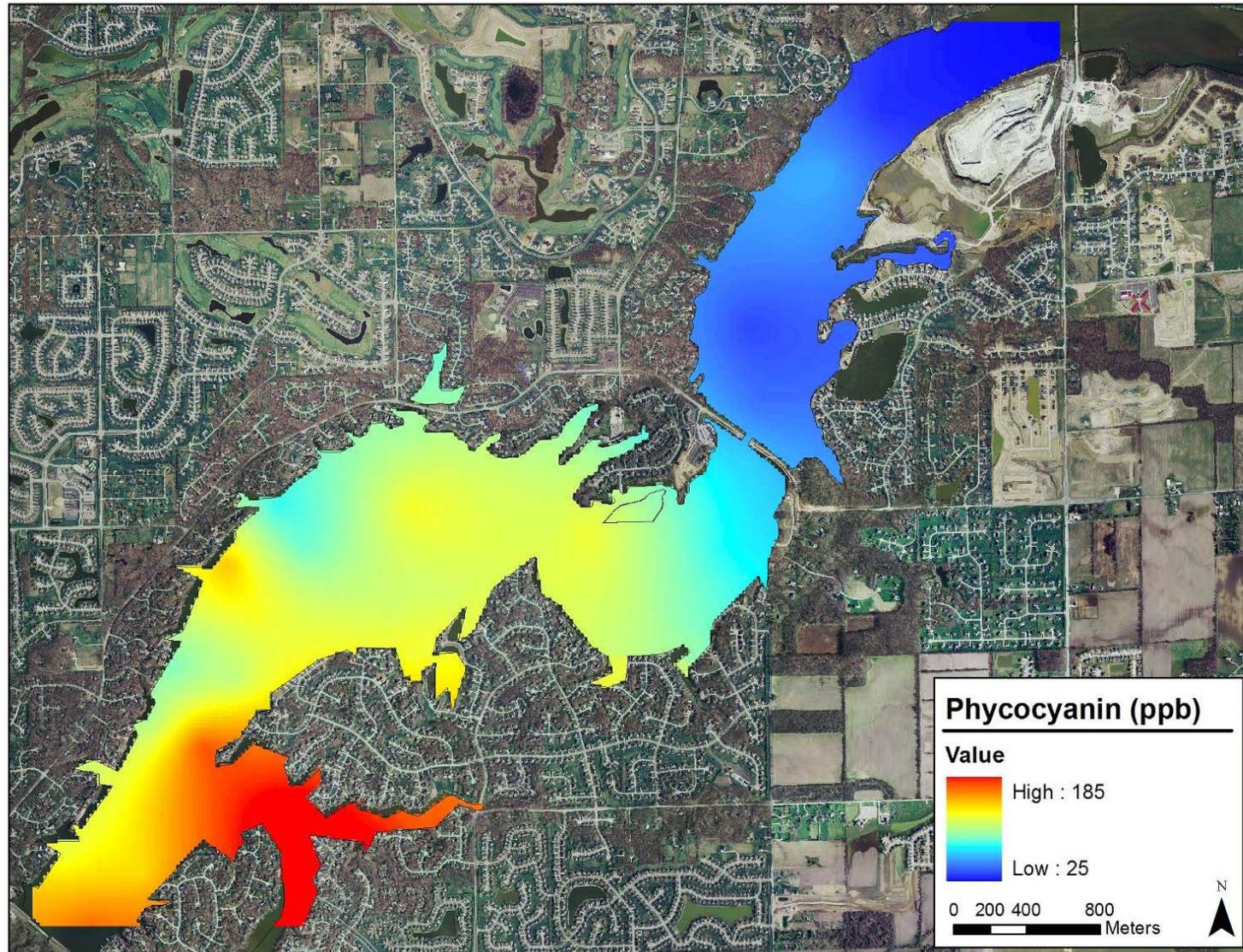


Figure 13: Interpolated map of ground truth phycocyanin concentration in Geist Reservoir on September 6, 2005.

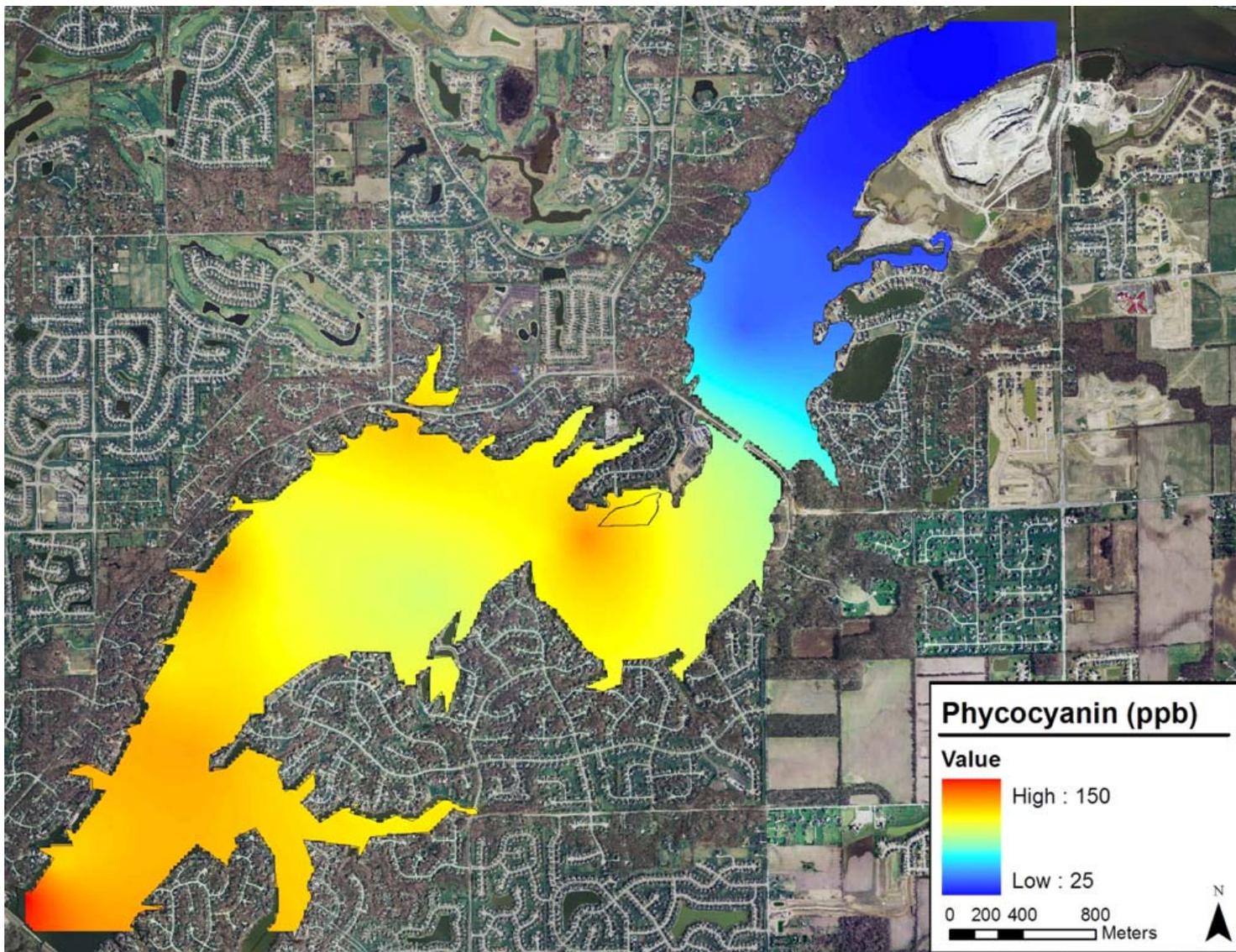


Figure 14: Interpolated map of phycocyanin concentration in Geist Reservoir on September 6, 2005 estimated from the modified Simis *et al.* (2005) algorithm ( $r^2 = 0.7825$ ; Table 2: Semi-Empirical Model) applied to field spectra.

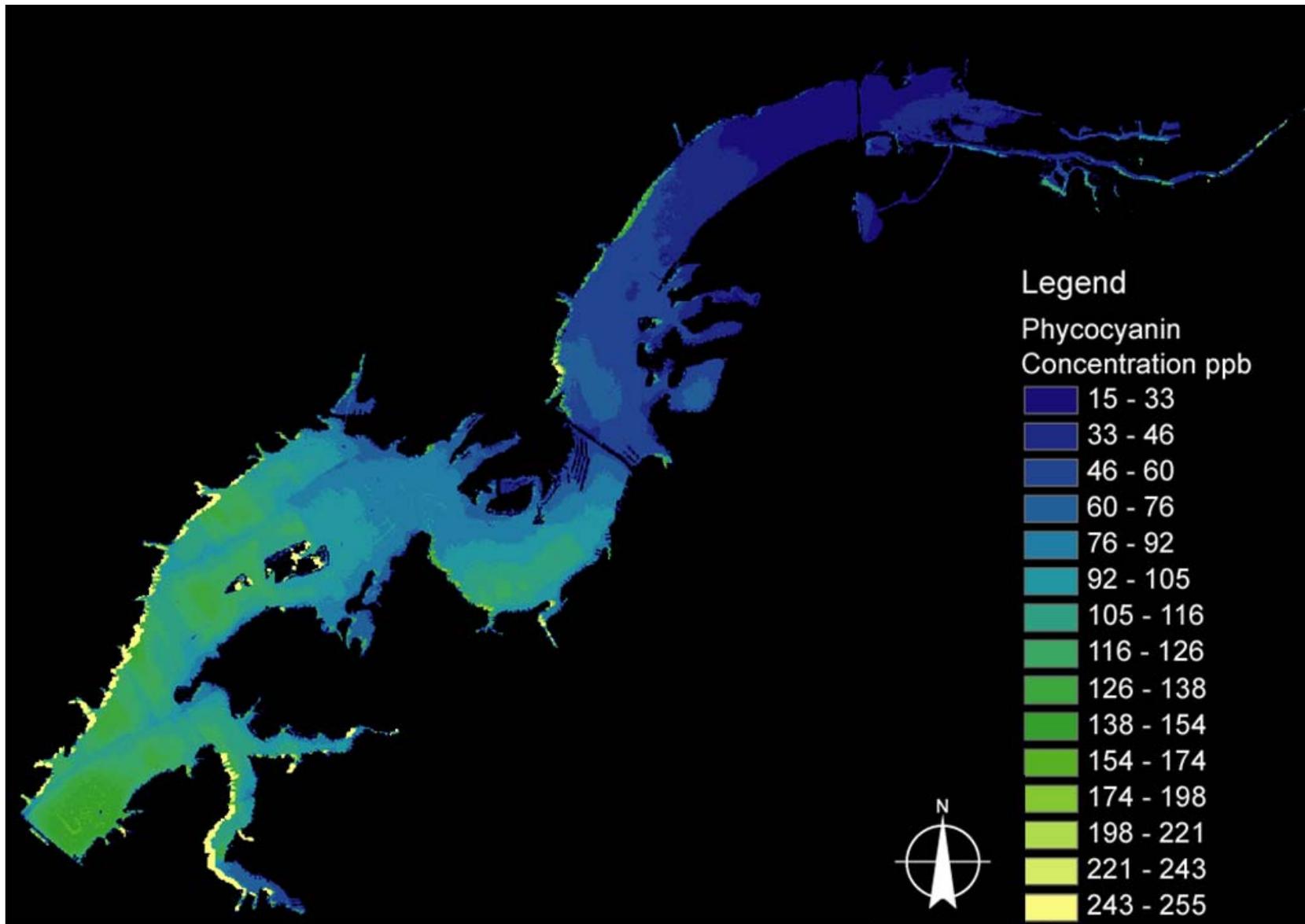


Figure 15: Map of phycocyanin concentrations in Geist Reservoir on September 6, 2005 estimated from the modified Gitelson *et al.* (1995) algorithm ( $r^2 = 0.6579$ ; Table 2: Empirical Model 6) applied to over 5 million airborne spectra that make up the image.

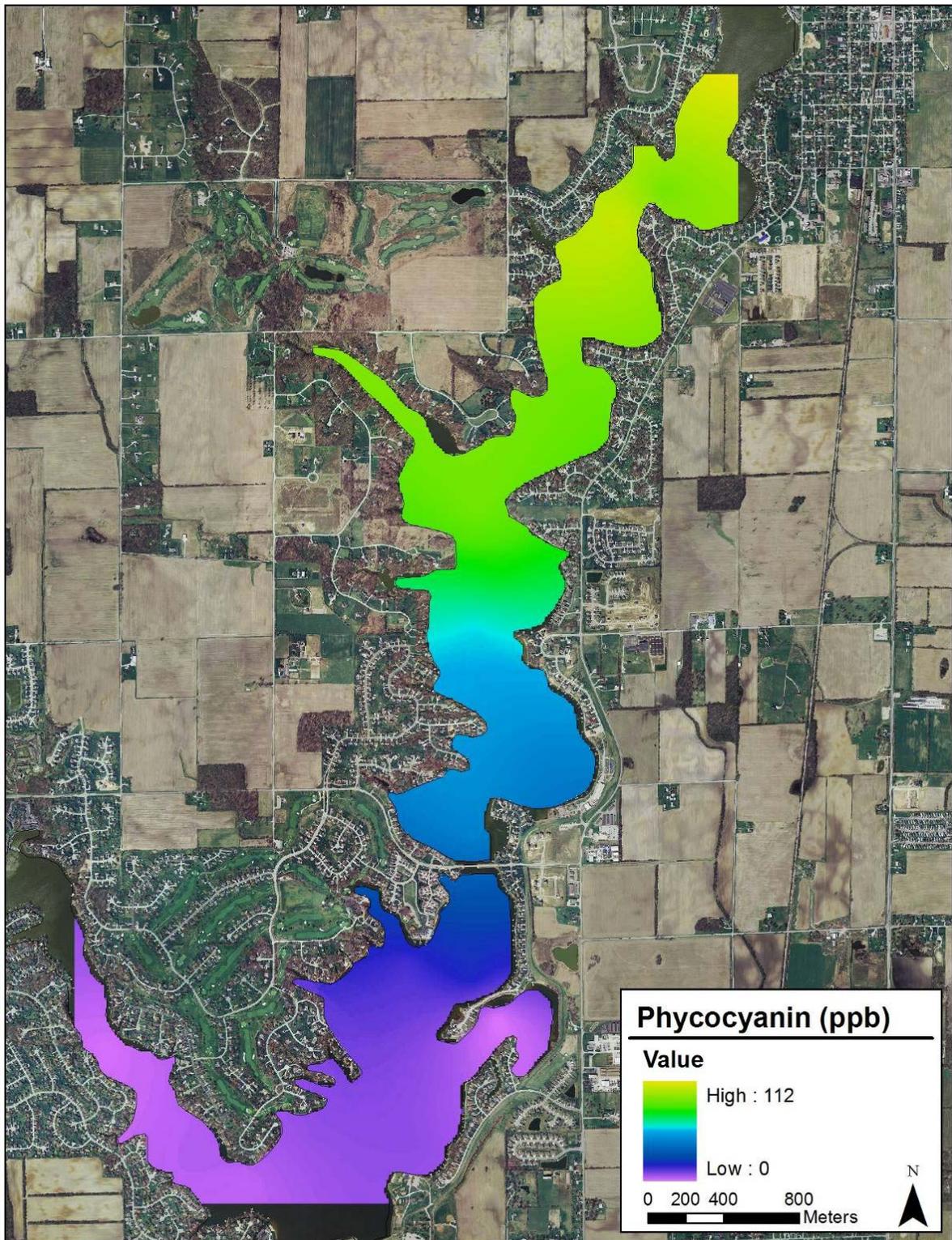


Figure 16: Interpolated map of phycocyanin concentration in Morse Reservoir on September 6, 2005 estimated from GA-PLS derived algorithm ( $r^2 = 0.9151$ ; Equation 7) applied to field spectra.

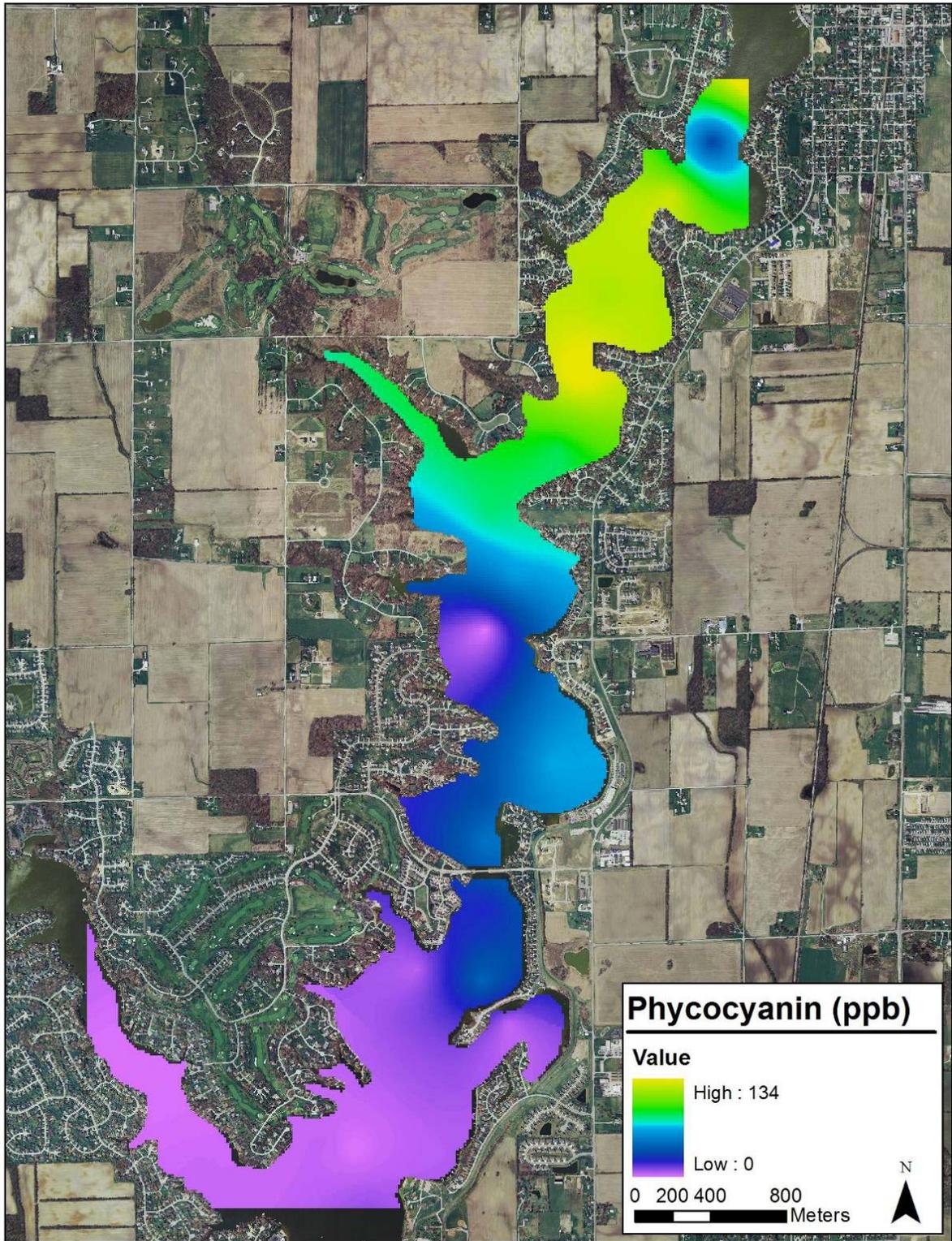


Figure 17: Interpolated map of ground truth phycocyanin concentrations in Morse Reservoir on September 6, 2005.

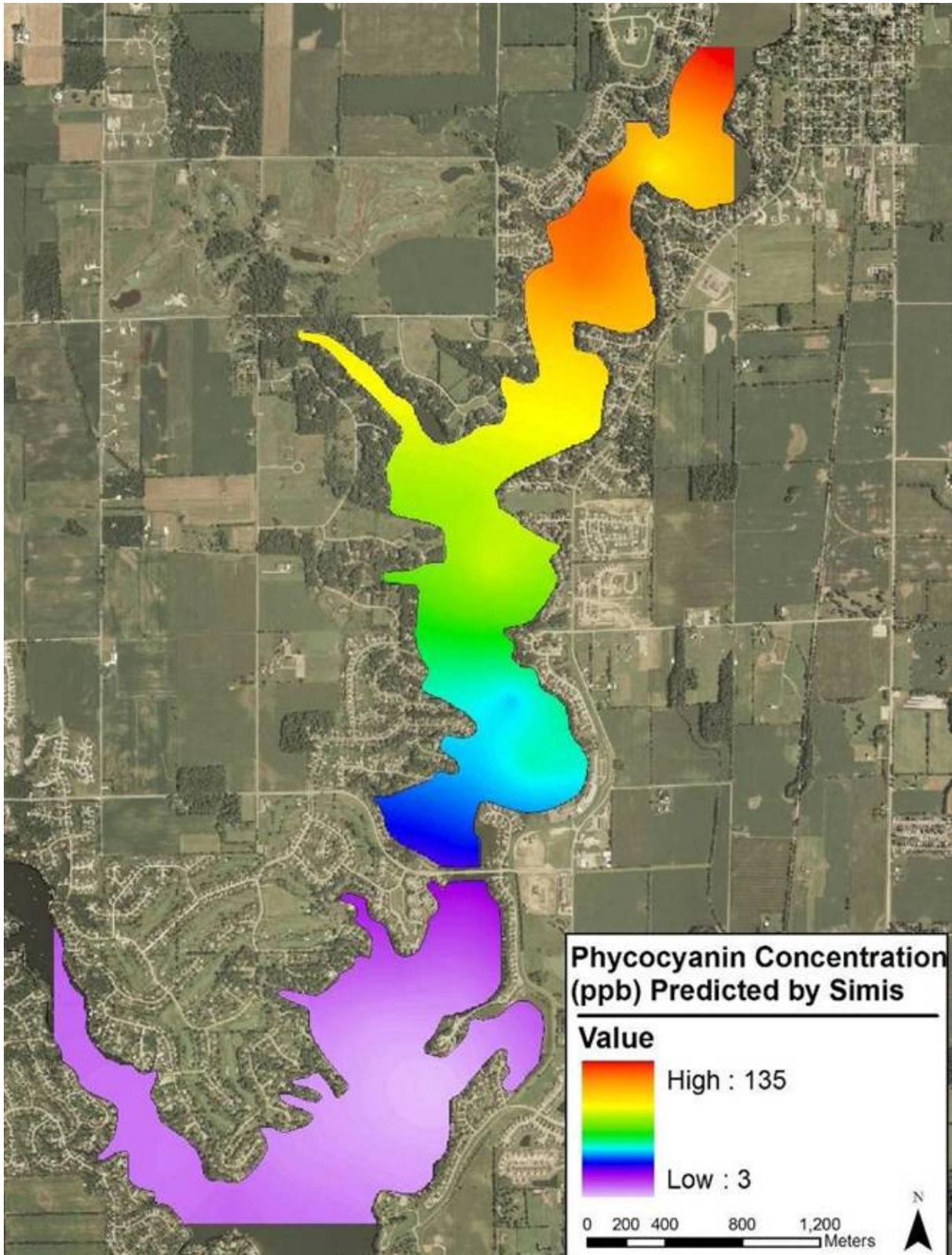


Figure 18: Interpolated map of phycocyanin concentration in Morse Reservoir on September 6, 2005 estimated from the modified Simis *et al.* (2005) algorithm ( $r^2 = 0.9095$ ; Table 2: Semi-Empirical Model) applied to field spectra.

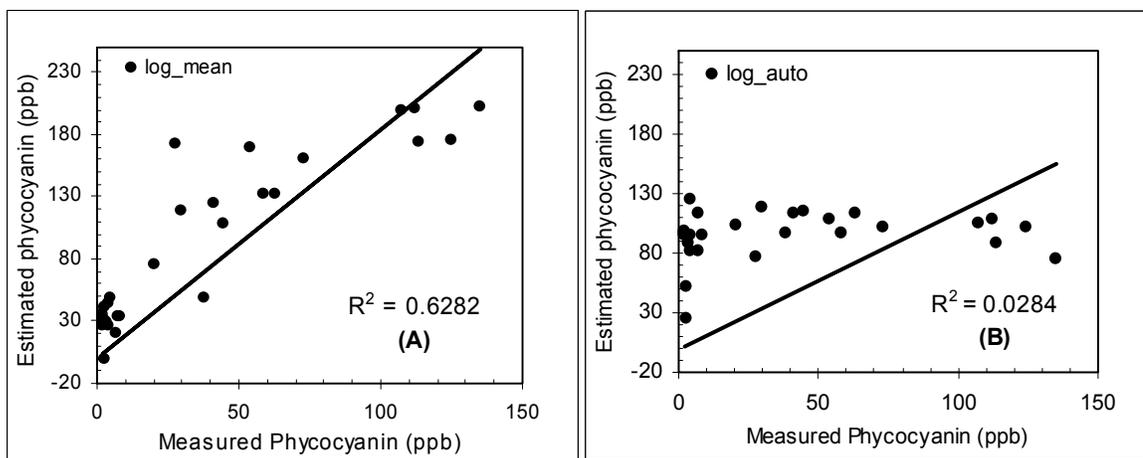
## 4.6 Assessment of robustness

This assessment has been completed for GA-PLS models applied to field spectra. Ongoing research will evaluate the robustness of empirical and semi-empirical models applied to both field spectra and airborne spectra.

PLS regression models are empirical relationships between *in situ* ASD reflectance and phycocyanin abundance. As with any other empirical model, the PLS algorithms might be site-specific, *i.e.*, a model developed for one reservoir may need to undergo site-specific calibration before being extrapolated to another lake or reservoir. To examine the algorithm's transferability, we have done several experiments with the datasets available for Geist and Morse Reservoirs. First, we used the Geist Reservoir dataset to develop an independent model for Geist Reservoir, and then examined the predictive power of the Geist-derived model on the Morse Reservoir dataset. Second the similar process was repeated for the model developed from the Morse Reservoir dataset, whereby the Morse-derived model was tested on the Geist Reservoir dataset. Lastly, PLS models were developed with the datasets from both Geist and Morse Reservoirs.

### 4.6.1 Robustness of model built with data from Geist Reservoir

Applying **Equation 5** and **Equation 6**, the PLS regression models derived from the Geist Reservoir dataset to the reflectance spectra data from Morse Reservoir generated poor predictions in which the coefficient of determination was 0.6282 and the RMSE 64.28 ppb using mean centered data (**Figure 19(A)**) and the corresponding values were 0.0284 and 70.187 ppb for autoscaled data (**Figure 19(B)**).

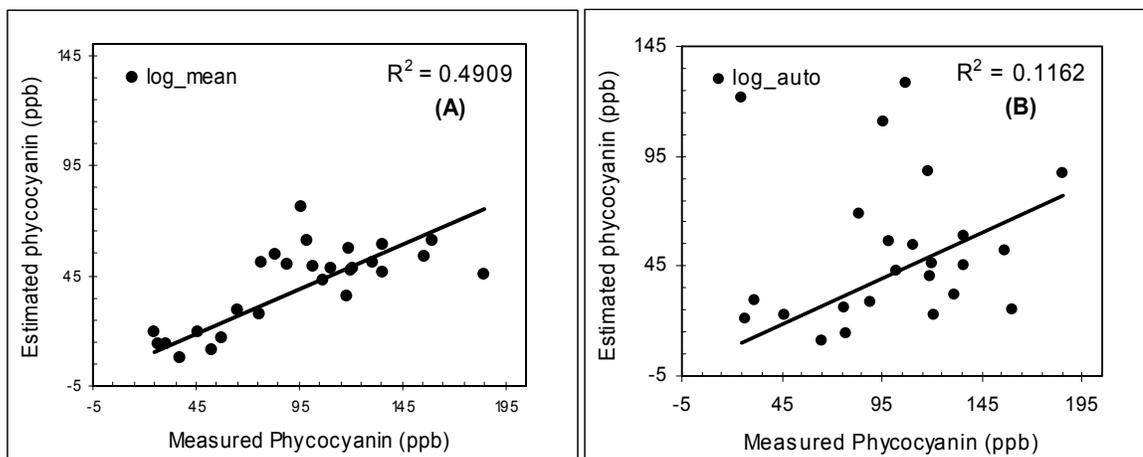


**Figure 19:** Linear regressions (black line) of measured phycocyanin and estimated phycocyanin using GA-PLS algorithms developed from Geist Reservoir datasets and applied to Morse Reservoir field reflectance spectra: (A) Equation 5 ( $R^2=0.63$ ), and (B) Equation 6 ( $R^2=0.03$ ), suggest that the model built from Geist Reservoir data is a poor predictor of Morse Reservoir phycocyanin.

### 4.6.2 Robustness of model built with data from Morse Reservoir

Application of **Equation 7** and **Equation 8**, the PLS regression models derived from the Morse Reservoir dataset to the reflectance spectra data from Geist reservoir generated poor predictions resulting in a coefficient of determination of 0.4909 and a RMSE 63.80

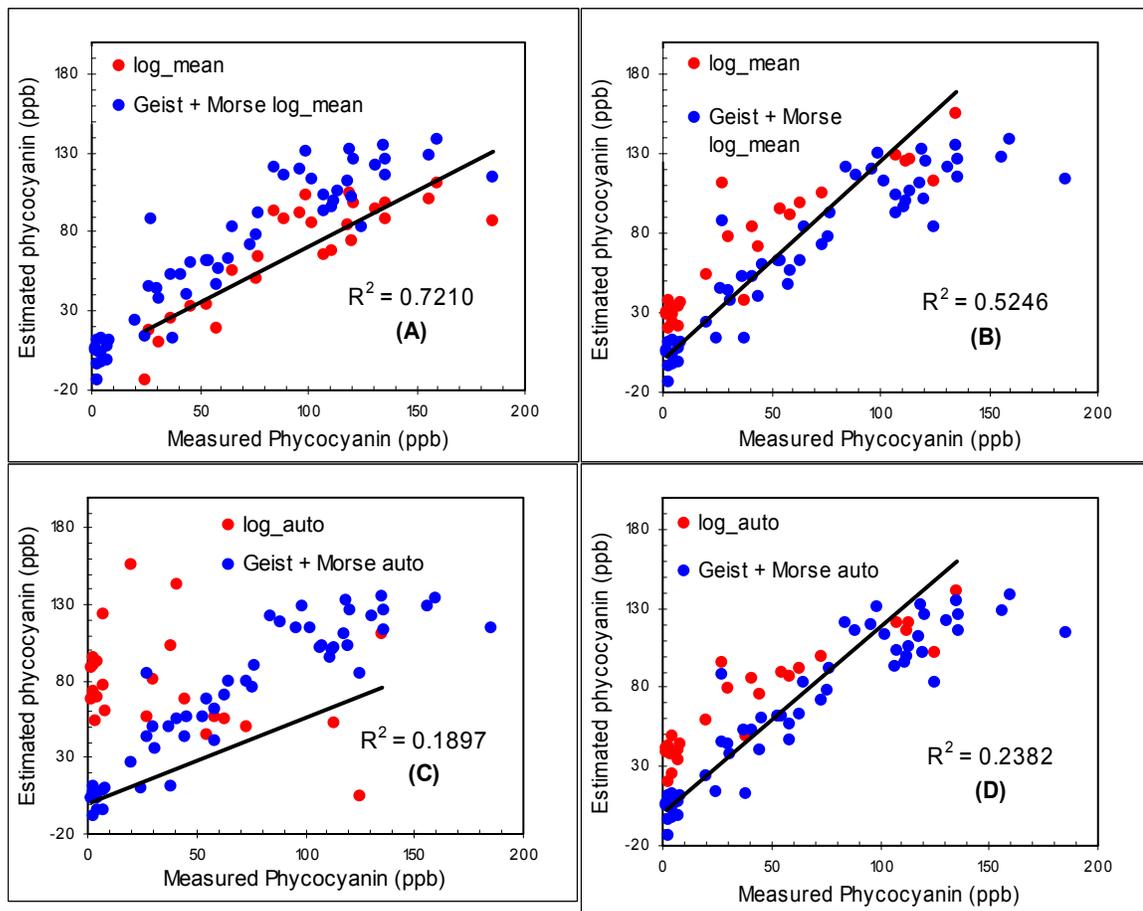
ppb with mean centered data, respectively and 0.1161 and 73.43 ppb for autoscaled data, respectively (**Figure 20**).



**Figure 20:** Linear regressions (black line) of measured phycocyanin and estimated phycocyanin using GA-PLS algorithms developed from Morse Reservoir datasets and applied to Geist Reservoir field reflectance spectra data, (A) Equation 7 ( $R^2=0.49$ ), and (B) Equation 8 ( $R^2=0.12$ ), suggest that the model built from Morse Reservoir data is a poor predictor of Geist Reservoir phycocyanin.

#### 4.6.3 Robustness of model built with data from two reservoirs

PLS models based on an aggregated dataset of both Geist and Morse Reservoir datasets were derived and applied to Geist and Morse Reservoir datasets separately. The PLS model derived from mean centered data from both reservoirs resulted in relatively good prediction with a coefficient of determination and RMSE of 0.7210 and 35.19 ppb respectively for Geist Reservoir (**Figure 21(A)**), and 0.5246 and 31.917 ppb for Morse Reservoir (**Figure 21(B)**). The PLS model derived with autoscaled data gave a poor prediction resulting in a coefficient of determination and RMSE of 0.1897 and 58.81 ppb respectively for Geist Reservoir (**Figure 21(C)**), and 0.2382 and 32.59 ppb for Morse Reservoir (**Figure 21(D)**).



**Figure 21: Linear regressions (black line) of measured phycocyanin and estimated phycocyanin using GA-PLS algorithms developed from Geist and Morse Reservoir combined datasets: (A) PLS log<sub>mean</sub> equation applied to Geist Reservoir data yielded an  $r^2 = 0.72$ ; (B) PLS log<sub>mean</sub> equation applied to Morse Reservoir data yielded an  $r^2 = 0.52$ ; (C) PLS autoscaled equation applied to Geist Reservoir data yielded an  $r^2 = 0.19$ ; and (D) PLS autoscaled equation applied to Morse Reservoir data yielded an  $r^2 = 0.24$ .**

### 3.0 Summary

This study shows the feasibility of using hyperspectral remote sensing techniques as a rapid assessment tool for determining the concentration and spatial distribution of blue-green algae in the three reservoirs which supply Indianapolis's drinking water system.

#### GA-PLS Algorithms

- GA modeling resulted in the selection of spectral bands most related to chlorophyll *a* and phycocyanin variability in the Indianapolis reservoirs. These GA selected bands are determinant of chlorophyll *a* and phycocyanin: the spectral features identified by GA modeling are those spectral features caused by the optical properties of the pigments as reported in the literature.
- GA-PLS derived algorithms for predicting phycocyanin using mean centered data performed better than using autoscaled data, showing that additional normalization was not necessary to improve algorithm performance.

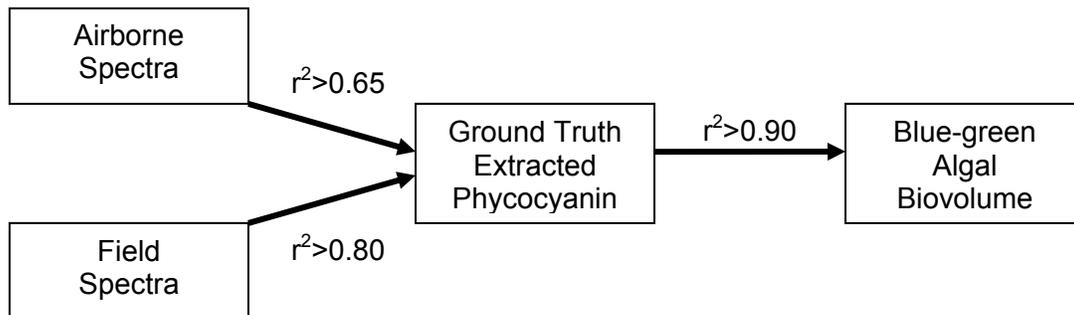
- GA-PLS derived algorithms performed well when applied to the reflectance spectra from the same reservoir from which the algorithm was developed.
- Developed algorithms from a training set of a specific reservoir did not perform well at predicting phycocyanin on a different reservoir. Differences in phytoplankton community structure between reservoirs could explain the low performance when transferring algorithms. At the time of sampling, Geist Reservoir had a mean chlorophyll *a*-to-phycocyanin ratio of less than 1.0, while Morse Reservoir's ratio was 4.0. This difference in photosynthetic pigment dominance caused the GA modeling to select different spectral bands for each reservoir's algorithm. Because Geist Reservoir had more phycocyanin than chlorophyll *a*, GA selected bands specific to phycocyanin (628.8 nm), chlorophyll *a* (694.6 nm and 563.2 nm), and phytoplankton biomass (704.0 nm). However, Morse Reservoir GA selected bands specific only to chlorophyll *a* (694.6 nm), dissolved organic matter (647.6 nm), and phytoplankton biomass (704.0 nm). The dominance of chlorophyll *a* in Morse Reservoir water samples was evident in the reflectance spectra, whereby chlorophyll *a* features dominated. As such, GA modeling selected bands specific to chlorophyll *a* for Morse Reservoir while selecting bands specific to both chlorophyll *a* and phycocyanin for Geist Reservoir.
- Algorithms developed from an aggregated training set of both Geist and Morse Reservoir performed moderately well on the reservoirs individually. This shows that despite the differences in chlorophyll *a*-to-phycocyanin ratios, the algorithms were able to account for those differences. However, to develop a more robust algorithm that will accurately predict phycocyanin concentration in all three reservoirs, a larger dataset which encompasses a larger range of chlorophyll *a*-to-phycocyanin ratio variability will likely be necessary.

#### *Empirical and Semi-Empirical Algorithms*

- Algorithms developed from semi-empirical models and applied to Geist and Morse Reservoir field reflectance data performed well and demonstrate the utility of incorporating the inherent optical properties of water quality constituents in future algorithm development.
- Algorithms developed from empirical models and applied to Geist Reservoir airborne spectra performed reasonably well, showing that simple models can result in adequate estimation of phycocyanin from AISA data and demonstrating that these techniques can be applied across sensors and sensor platforms (e.g., field, airborne, and satellite sensors).

#### *Phycocyanin and Blue-green Algal Measures*

- The relationship between phycocyanin and blue-green algal measures used by water resource managers (*i.e.*, natural units- $\text{mL}^{-1}$  and biovolume) resulted in strong relationships ( $r^2 = 0.9104$  and  $0.9460$ , respectively); thus, these measures can be accurately estimated from predicted phycocyanin concentrations.



**Figure 22: Summary of study results.**

Overall, the data show remote sensing techniques were successful at predicting blue-green algal abundance in Geist and Morse Reservoirs (**Figure 22**). Therefore, hyperspectral remotely sensed data from field and airborne sensors can be used to successfully and rapidly map the relative concentration and distribution of blue-green algae in Eagle Creek, Geist, and Morse Reservoirs. However, a larger spectral library of different water conditions (e.g., non-algal turbidity, non-organic seston, and chlorophyll *a*-to-phycocyanin ratios) will need to be collected to improve GA-PLS band selection and algorithm development and empirical and semi-empirical model performance across a range of water quality conditions.

#### 4.0 On-going and Future Work

In June 2006, data collection began for the 2006 CIWRP project “Quantifying Blue-Green Algae of Central Indiana Reservoirs Using Hyperspectral Reflectance.” This study was initiated to further improve the performance of GA-PLS derived algorithms at retrieving phycocyanin concentrations from field spectra. Data will also be used to validate and test the robustness of empirical and semi-empirical algorithms used to estimate phycocyanin concentrations from field spectra. This dataset includes field reflectance spectra from both ASD and Ocean Optics spectroradiometers, measurements of optically active constituents (e.g., *in-vitro* (extracted) measurements of chlorophyll *a* and phycocyanin, TSS, and loss-on-ignition carbon), measurements of water quality parameters (e.g., nutrients, MIB/Geosmin, and major anions and cations), and *in situ* physical water conditions (e.g., Secchi depth, temperature, DO, SPC, and pH). Between June and October 2006, 122 samples have been collected on Eagle Creek Reservoir over seven sampling days; 112 samples have been collected on Geist Reservoir over seven sampling days; and 102 samples have been collected on Morse Reservoir over six sampling days. Data were collected to obtain a larger range of non-algal turbidity to improve the robustness of algorithms over a range of suspended sediment concentrations and to further optimize the algorithms for use in Case II waters.

Hyperspectral imagery from the AISA sensor for Eagle Creek and Morse Reservoirs are currently being processed (normalized, registered, mosaicked, and calibrated) to obtain reflectance data to which the algorithms can be applied to estimate phycocyanin distribution and concentration in these reservoirs. While the spatial resolution of 1 m<sup>2</sup> pixels allows for the creation of high resolution maps, the data file size (over a terabyte per image) requires a powerful computer and time to process the data. AISA imagery of Eagle Creek and Morse Reservoirs should be processed by January 2007. Phycocyanin algorithms will be applied to these data and maps will be generated by Spring 2007.

Based on these studies preliminary findings, future research will focus on improving our techniques for predicting water quality from hyperspectral data. This will include (1) increasing the number of water quality parameters retrieved from remotely sensed data; (2) expanding our approach to developing algorithms from empirical based models to bio-optical models; and (3) scaling up our use of spectroradiometer platforms to include satellite imagery such as NASA's MODIS sensor in larger water bodies. Currently, research is focused on using empirically or semi-empirically derived algorithms to map the phytoplankton pigments chlorophyll *a* and phycocyanin. Expanding the research focus to include bio-optical models, which are based upon the inherent optical properties of water quality parameters, will allow for the simultaneous determination of chlorophyll *a*, total suspended solids, and colored dissolved organic matter. These three water quality constituents are the most optically active components of natural waters. We hope to expand the use of bio-optical models to include the determination of phycocyanin. In addition to increasing the number of water quality parameters retrieved from remotely sensed data, bio-optical models are more robust than empirical models when applied across a range of sensors and locations: the same bio-optical models could be applied to the data collected by hand held, airborne, or satellite platforms to retrieve water quality parameters from many locations without the need for extensive sensor or site calibration. This robustness of bio-optical models will allow for the transferability of models developed using hand-held sensors (e.g. ASD Fieldspec (UV/VNIR) and Ocean Optics USB2000 (V/NIR) spectroradiometers) to airborne and satellite sensors.

Long term interests include integrating watershed-scale processes through remote sensing and GIS tools. Remotely sensed data and GIS tools can be used to monitor temporal changes in watershed processes such as land use and land cover change, soil erosion and transport, nutrient load export, and nuisance algal bloom occurrence. These processes can be integrated into a system-wide approach to predicting water quality changes over time. This tool could then be displayed as a web-based forecasting and alarm system for educating the public about current and predicted water quality conditions.

## **7.0 Academic Scholarship**

As of October 2006, data obtained from this research study have been presented at the 2006 Indiana Water Resources Association's Spring 2006 meeting, West Lafayette, Indiana. The two graduate students who presented this research, Kaylan L. Randolph and Rebecca E. Sengpiel both received awards for Outstanding Student Presentation. Additionally, both graduate students and Dr. Lin Li presented aspects of this research at the 2006 North American Lake Management Society Annual Meeting in Indianapolis, IN (November 8 – 10, 2006).

## 8.0 Work Cited

- Bangalore, A. S., R. E. Schaffer, G. W. Small, and M. A. Arnold, 1996. Genetic algorithm-based method for selecting wavelengths and model size for use with partial least squares regression: Application to near-infrared spectroscopy. *Analytical Chemistry*, 68: 4200-4212.
- Behm, D. 2003. Coroner cites algae in teen's death. *The Milwaukee Journal Sentinel*, September 6, 2003.
- Boardman, J. W. 1989. Inversion of imaging spectrometry data using singular value decomposition, in *Proceedings IGARSS '89, 12th Canadian Symposium on Remote Sensing*, 4: 2069 – 2072.
- Buiteveld H, J. Hakvoort, and M. Donze. 1994. The optical properties of pure water. *Ocean Optics XII*, J. Jaffe, ed. Proc. SPIE 2258. 174-183.
- Cairns, S.H., K.L. Dickson, and S.f. Atkinson. 1997. An examination of measuring selected water quality trophic indicators with SPOT satellite HRV data. *Photogrammetric Engineering and Remote Sensing*, 63: 263–65.
- Coley, D. A. 2003. *An Introduction to Genetic Algorithms for Scientists and Engineers*, World Scientific, Singapore, 227 pp.
- Carmichael, W.W. 1994. "The Toxins of Cyanobacteria." *Scientific America*, January: 78-86.
- Carpenter, D. J., and S.M. Carpenter. 1983. Modeling inland water quality using Landsat data, *Remote Sensing of Environment*, 13: 345-352.
- Danson, F.M., A.L. Heathwaite and S.T. Trudgill. 1991. Multi-temporal monitoring of lake water quality, In Hunt, J.J., editor, *Physical measurements and signatures in remote sensing*, Proceedings 5th International Colloquium, Courchevel. Noordwijk: European Space Agency, Special Publication 319, 1: 169–72.
- Dekker, A. G., T.J. Malthus, and E. Seyhan. 1991. Quantitative modeling of inland water quality for high resolution MSS system. *IEEE Transactions Geoscience and Remote Sensing*, 29: 89-95.
- Dekker, A. G. 1993. Detection of the Optical Water Quality Parameters for Eutrophic Waters by High Resolution Remote Sensing. Ph. D. Thesis, Free University, Amsterdam, The Netherlands.
- Ding, Q., G.W. Small, and M.A. Arnold. 1998. Genetic algorithms-based wavelength selection for the near-infrared determination of glucose in biological matrixes: Initialization strategies and effects of spectral resolution. *Analytical Chemistry*, 70: 4472-4479
- Forrest. S. 1993. Genetic algorithms: Principles of natural selection applied to computation. *Science*, 261. 872-878.
- Fraser, R.N. 1998. Hyperspectral remote sensing of turbidity and chlorophyll *a* among Nebraska Sand Hills lakes, *International Journal of Remote Sensing*, 19: 1579-1589.
- Geladi, P. and B. Kowalski. 1986. Partial least squares regression: A tutorial. *Analytica Chimica Acta*, 185: 1-17.
- Gitelson, A. 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: 3367-3373.

DEVELOPING A SURVEY TOOL FOR THE RAPID ASSESSMENT OF BLUE-GREEN ALGAE IN  
CENTRAL INDIANA'S RESERVOIRS

---

- Gitelson, A. A., S. Laorawat, G.P Keydan, and A. Vonshak. 1995. Optical Properties of dense algal cultures outdoors and their application to remote estimation of Biomass and pigment concentration in *Spirulina Platensis* (CYANOBACTERIA). *Journal of Phycology*, 31: 828-834.
- Gons, H. (1999). Optical teledetection of chlorophyll a in turbid inland waters. *Environmental Science Technology*. 33: 1127-1132.
- Gordon, H.R. D.K. Clark, J.W. Brown, O.B. Brown, R.H. Evans, and W.W. Broenkow. 1983. Phytoplankton pigment concentrations in the middle Atlantic Bight – comparison of ship determinations and CZCS estimates. *Applied Optics*, 22(1): 20-36.
- Gourvéneç, S., X. Capron, and D.L. Massart, 2004. Genetic algorithms (GA) applied to the orthogonal projection approach (OPA) for variable selection. *Analytica Chimica Acta*, 519: 11-21.
- Haadland, D. M. and E.V. Thomas. 1988a. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Analytical Chemistry*, 60: 1193-1202.
- Haadland, D. M. and E.V. Thomas. 1988b. Partial least-squares methods for spectral analyses. 2. Application to simulated and glass spectral data. *Analytical Chemistry*, 60: 1202-1208.
- Hamilton, M. K., V.O. Davis, W.J. Rhea, S.H. Pilorz, and K.L Carder. 1993. Estimating chlorophyll content and bathymetry of Lake Tahoe using AVIRIS data. *Remote Sensing of Environment*, 44: 217-230.
- Han, L., and D.C Rundquist. 1997. Comparison of nir/red ratio and first derivative of reflectance in estimating algal-chlorophyll concentration: A case study in a turbid reservoir. *Remote Sensing of Environment*, 62: 253-261.
- Hovis, W.A., D.K. Clark, F. Anderson, R.W. Austin, W.H. Wilson, E.T. Baker, D. Ball, H.R. Gordon, J.L. Mueller, S.Z. El-Sayed, B. Strum, R.C. Wrigley, and C.S. Yentsch. 1980. Nimbus-7 coastal zone color scanner – system description and early imagery. *Science*, 210: 60-63.
- Indiana Department of Environmental Management. 2002. Indiana Integrated Water Quality Monitoring and Assessment Report. IDEM/34/02/004/2002.
- Indiana Department of Environmental Management. 2004. Indiana Integrated Water Quality Monitoring and Assessment Report. <http://www.in.gov/idem/water/planbr/wqs/quality.html>.
- Indiana Department of Environmental Management. 2006. Indiana Integrated Water Quality Monitoring and Assessment Report. <http://www.in.gov/idem/water/planbr/wqs/quality.html>.
- Jensen, J.R. 2000. *Remote Sensing of the Environment: An Earth Resources Perspective*. Prentice-Hall, Inc., Upper Saddle River, NJ.
- Jouan-Rimbaud, D., D-L. Massart, R. Leardi, and O.E. De Noord. 1995. Genetic algorithms as a tool for wavelength selection in multivariate calibration, *Analytical Chemistry*, 67: 4295-4301.
- Jupp, D. L.B., A. Held, G. Byrne, P. Hutton, E. McDonald, and D. Parkin. 1993. Potential use of airborne scanning for monitoring algal dynamics in Australian inland waters, Office of Space Science and Applications, Publication No. 30. 141. pp.

- Kallio, K., T. Kuster, T. Hannonen, S. Koponen, J. Pullianinen, J. Vepsalainen, and T. Pyhalhti. 2001. Feasibility of airborne imaging spectrometry for lake monitoring—a case study of spatial chlorophyll *a* distribution in two meso-eutrophic lakes. *International Journal of Remote Sensing*, 24: 3771-3790.
- Koponen, S., J. Pullianinen, H. Servomaa, Y. Zhang, M. Hallikainen, K. Kallio, K. Eloheimo, and T. Hannonen. 2001. Analysis on the feasibility of multisource remote sensing observations for chl-*a* monitoring in Finnish lakes, *The Science of the Total Environment*, 268: 95-106.
- Koponen, S., J. Pullianinen, K. Kallio, and M. Hallikainen. 2002. Lake water quality classification with airborne hyperspectral spectrometer and simulated MERIS data. *Remote Sensing of Environment*, 79: 51-59.
- Kruse, F. A., LL. Richardson, and V.G. Ambrosia. 1997. Techniques Developed for Geologic Analysis of Hyperspectral Data Applied to Near-Shore Hyperspectral Ocean Data, *Presented at the Fourth International Conference on Remote Sensing for Marine and Coastal Environments*, Orlando, Florida. March 17 – 19, 1997.
- Lathrop, R.G., Jr. and T.M Lillesand. 1986. Use of Thematic Mapper data to assess water quality in Green Bay and central Lake Michigan, *Photogrammetric Engineering and Remote Sensing*, 52, 671–80.
- Lathrop, R.G., Jr. and T.M Lillesand. 1989. Monitoring water quality and river plume transport in Green Bay, Lake Michigan with SPOT-1 imagery, *Photogrammetric Engineering and Remote Sensing*, 55: 349–54.
- Lathrop, R.G., Jr., T.M. Lillesand and B.S. Yandell. 1991. Testing the utility of simple multivariate Thematic Mapper calibration algorithms for monitoring turbid inland waters, *International Journal of Remote Sensing*, 12: 2045–63.
- Lavery, P., C. Pattiaratchi, A. Wyllie and P. Hick. 1993. Water quality monitoring in estuarine waters using the Landsat Thematic Mapper, *Remote Sensing of Environment*, 46: 268–80.
- Leardi, R. 2000. Application of genetic algorithm–PLS for feature selection in spectral data sets, *Journal of Chemometrics*, 14: 643–655.
- Leardi, R. and A.L. González. 1998. Genetic algorithms applied to feature selection in PLS regression: How and when to use them, *Chemometrics and Intelligent Laboratory System*, 41: 195-207.
- Lestander, T., R. Leardi, and P. Geladi. 2003. Selection of near infrared wavelengths using genetic algorithms for the determination of seed moisture content, *Journal of Near Infrared Spectroscopy*, 13: 433-446.
- Li L., S. L. Ustin, and D. Riano. 2006. Retrieval of Fresh Leaf Fuel Moisture Content Using Genetic Algorithm – Partial Least Squares Modeling (GA-PLS), *IEEE Trans. Geosci. Remote Sens. Lett.*, in press.
- Lillesand, T. M., W.L. Johnson, R.L. Deuell, O.M. Lindstrom, and D.E. Miesner. 1983. Use of Landsat data to predict trophic status of Minnesota lakes, *Photogrammetric Engineering and Remote Sensing*, 49: 219-229.
- Liu, Y., M.A. Islam, and J. Gao. 2003. Quantification of shallow water quality parameters by means of remote sensing. *Progress in Physical Geography*, 27: 24-43.

DEVELOPING A SURVEY TOOL FOR THE RAPID ASSESSMENT OF BLUE-GREEN ALGAE IN  
CENTRAL INDIANA'S RESERVOIRS

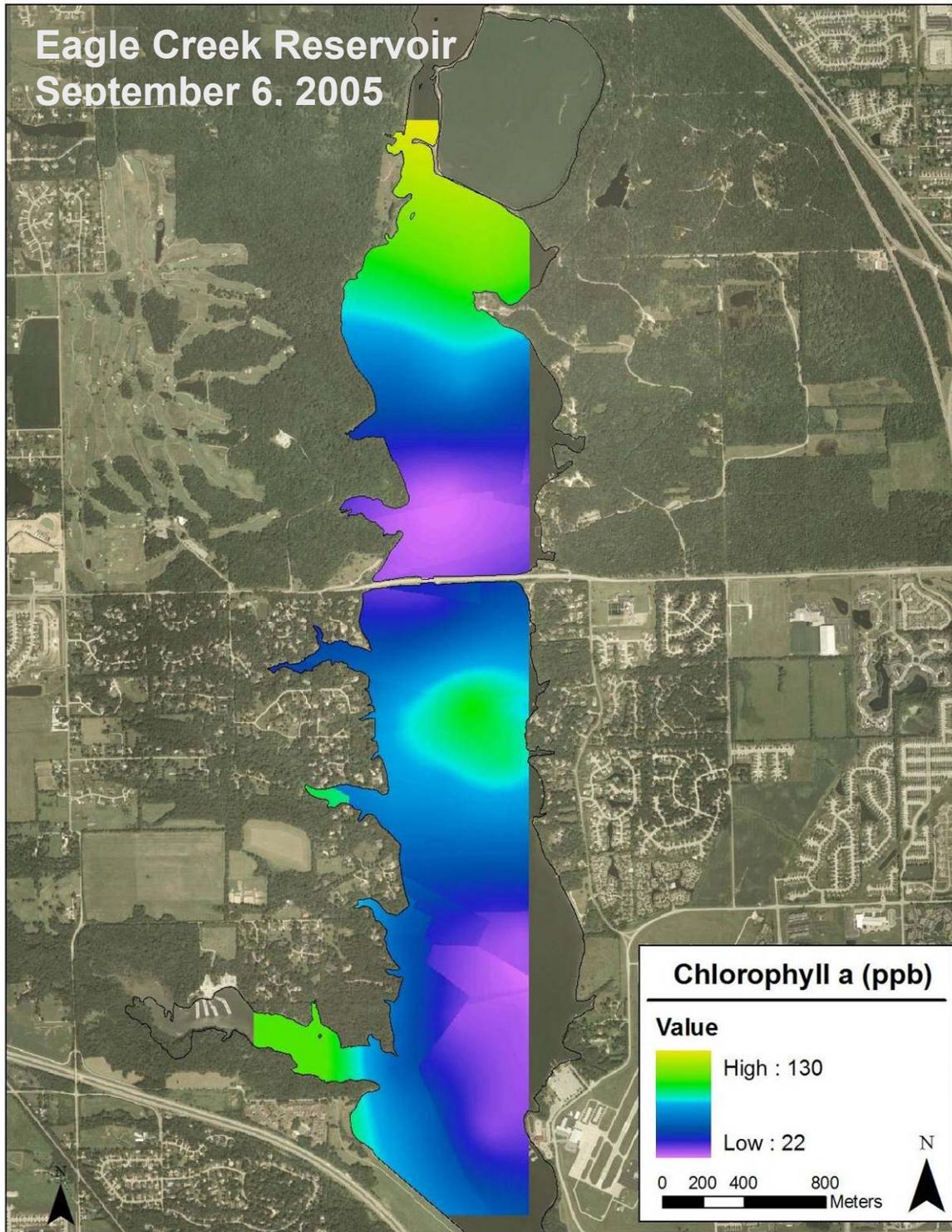
---

- Lund, J.W.G., C. Kipling, and E.D. LeCren. 1958. The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. *Hydrobiologia*, 11:143 – 170.
- Mille, D.F., M.C. Baker, C.D. Tucker, B.T. Vinyard, and C.P. Dionigi. 1992. High resolution airborne remote sensing of bloom-forming phytoplankton, *Journal of Phycology*, 28: 281-290.
- Mitchell, M. 1998. *In Introduction to Genetic Algorithms*, The MIT Press, Massachusetts Institute of Technology, Cambridge, Massachusetts, pp. 209.
- Pascual, D.L., R. Raftis, G. Filippelli, and L.P. Tedesco. 2006. Run-off and tile drainage versus internal recycling: three year mass balance approach to understand phosphorous loading and productivity in a small, urban reservoir, Eagle Creek Reservoir, IN, USA. *Presented at the North American Lake Managers Association, 2006 International Symposium*. Indianapolis, Indiana. November 8 – 10, 2006.
- Polliainen, J., K. Kallio, K. Eloheimo, S. Koponen, H. Servomaa, T. Hannonen, S. Tauriainen, and M. Hallikainen. 2001. A semi-operational approach to water quality retrieval from remote sensing data, *The Science of the Total Environment*, 268: 79–93.
- Pope, R. and E. Fry. 1997. Absorption spectrum (380-700 nm) of pure water. II. Integrating cavity measurements. *App. Opt.*, 36:8710-8723.
- Prangma, G.J. and J.N. Roozkrans. 1989. Using NOAA AVHRR imagery in assessing water quality parameters, *International Journal of Remote Sensing*, 10: 811–818.
- Richardson, L. L. 1996. Remote sensing of algal bloom dynamics. *BioScience*, 46: 492-501.
- Ritchie, J.C., C.M. Cooper, and F.R. Schiebe. 1990. The relationship of MSS and TM digital data with suspended sediments, chlorophyll, and temperature in Moon lake, Mississippi. *Remote Sensing Environment*, 33: 137-148.
- Rundquist, D., L. Han, J.F. Schalles, and J.S. Peake. 1996. Remote measurement of algal chlorophyll in surface water: The case for the first derivative of reflectance near 690 nm. *Photogrammetric Engineering and Remote Sensing*, 62: 195-200.
- Sarada, R. M.G. Pillai, and G.A. Ravishanker. 1999. Phycocyanin from *Spirulina* sp: influence of processing of biomass on phycocyanin yield, analysis of efficacy of extraction methods and stability studies on phycocyanin. *Process Biochemistry*, 34: 795-801.
- Shaffer, R., G. W. Small, and M. A. Arnold. 1996. Genetic algorithm-based protocol for coupling digital filtering and partial least-squares regression: Application to the near-infrared analysis of glucose in biological matrices. *Analytical Chemistry*, 68: 2663-2675.
- Schalles, J, A. Gitelson, Y. Yacobi, and A. Kroenke. 1998. Estimation of Chlorophyll *a* from time series measurements of high spectral resolution reflectance in a eutrophic lake. *Journal of Phycology*, 34: 383-390.
- Schalles, J.F., and Y.Z. Yacobi. 2000. Remote detection and seasonal patterns of phycocyanin, carotenoid, and chlorophyll pigments in eutrophic waters. *Archives fur Hydrobiologia - Special Issues Advancements in Limnology*, 55: 153-168.
- Simis, S, S. Peters, and S. Gons. 2005. Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. *Limnology and Oceanography*, 50(11): 237-245.

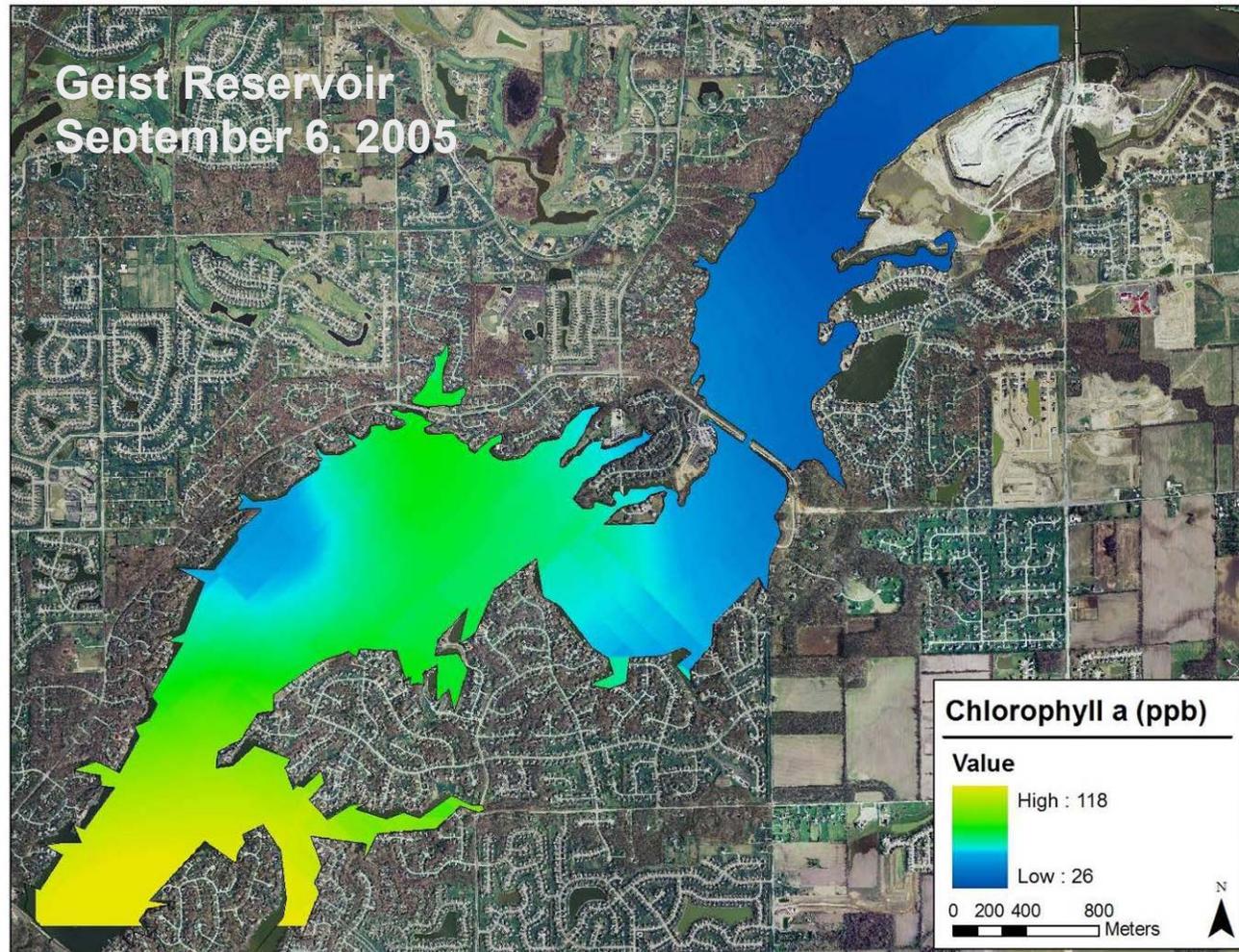
- Strong, A. E. 1974. Remote sensing of algal blooms by aircraft and satellite in Lake Erie and Utah lake, *Remote Sensing Environment*, 3: 99-107.
- Tedesco, L.P., E.A. Atekwana, G. Filippelli, K. Licht, L.K. Shrake, B.E. Hall, D.L. Pascual, J. Latimer, , R. Raftis, D. Sapp, G. Lindsey, R. Maness, D. Pershing, D. Peterson, K. Ozekin, C. Mysore, and M. Prevost 2003. Water Quality and Nutrient Cycling in Three Indiana Watersheds and Their Reservoirs: Eagle Creek/Eagle Creek Reservoir, Fall Creek/Geist Reservoir and Cicero Creek/Morse Reservoir. Central Indiana Water Resources Partnership, CEES Publication 2003-01, IUPUI, Indianapolis, IN, 163 pp.
- Tedesco, L.P., D.L. Pascual, L.K. Shrake, L.R. Casey, B.E. Hall, P.G.F. Vidon, F.V. Hernly, R.C. Barr, J. Ulmer, and D. Pershing. 2005. Eagle Creek Watershed Management Plan: An Integrated Approach to Improved Water Quality. Eagle Creek Watershed Alliance, CEES Publication 2005-07, IUPUI, Indianapolis, 182 pp.
- Verdin, J.P. 1985. Monitoring water quality conditions in a large western reservoir with Landsat imagery (USA), *Photogrammetric Engineering and Remote Sensing*, 51, 343-53.
- Vincent, R. K., X. Qin, R.M.L. McKay, J. Miner, K. Czajkowski, J. Savino, and T. Bridgemanet. 2004. Phycocyanin detection from LANDSAT TM data for mapping cyanobacterial blooms in Lake Erie, *Remote Sensing of Environment*, 89: 381- 392.
- Wilson, J.T., S.E. Morlock, and N.T. Baker. 1996. Bathymetric Surveys of Morse and Geist Reservoirs in Central Indiana Made with Acoustic Doppler Current Profiler and Global Positioning System Technology. U.S. Geological Survey Water Resources Division. <http://in.water.usgs.gov/bathymetry.web/> from December 12, 2002.
- Wise, B. M., J. M. Shaver, N. B. Gallagher, W. Windig, R. Bro, and R. S. Koch 2005. *PLS\_Toolbox 3.5*, Eigenvector Research Incorporation.
- Wold, H. 1966a. Nonlinear estimation by iterative least squares procedure, *In Research papers in statistics* (ed. David F., Wiley & Sons, New York), 441-444.
- Wold, H. 1966b. Estimation of principal components and related models by iterative least squares, *In Multivariate Analysis* (ed. Krishnaiah P. R., Academic Press, New York), 391-420.
- Yao, H. and L. Tian 2003. A genetic-algorithm-based selective principal component analysis (GA-SPCA) method for high-dimensional data feature extraction, *IEEE Transaction on Geoscience and Remote Sensing*, 41(6): 1469-1478.
- Zhang,Y., S.S. Koponen, J.T. Pulliainen, and M.T. Hallikainen. 2003. Application of Empirical Neural Networks to Chlorophyll-a Estimation in Coastal Waters Using Remote Optosensors, *IEEE Sensors Journal*, 3: 376-382.

## APPENDIX A: Chlorophyll a Maps

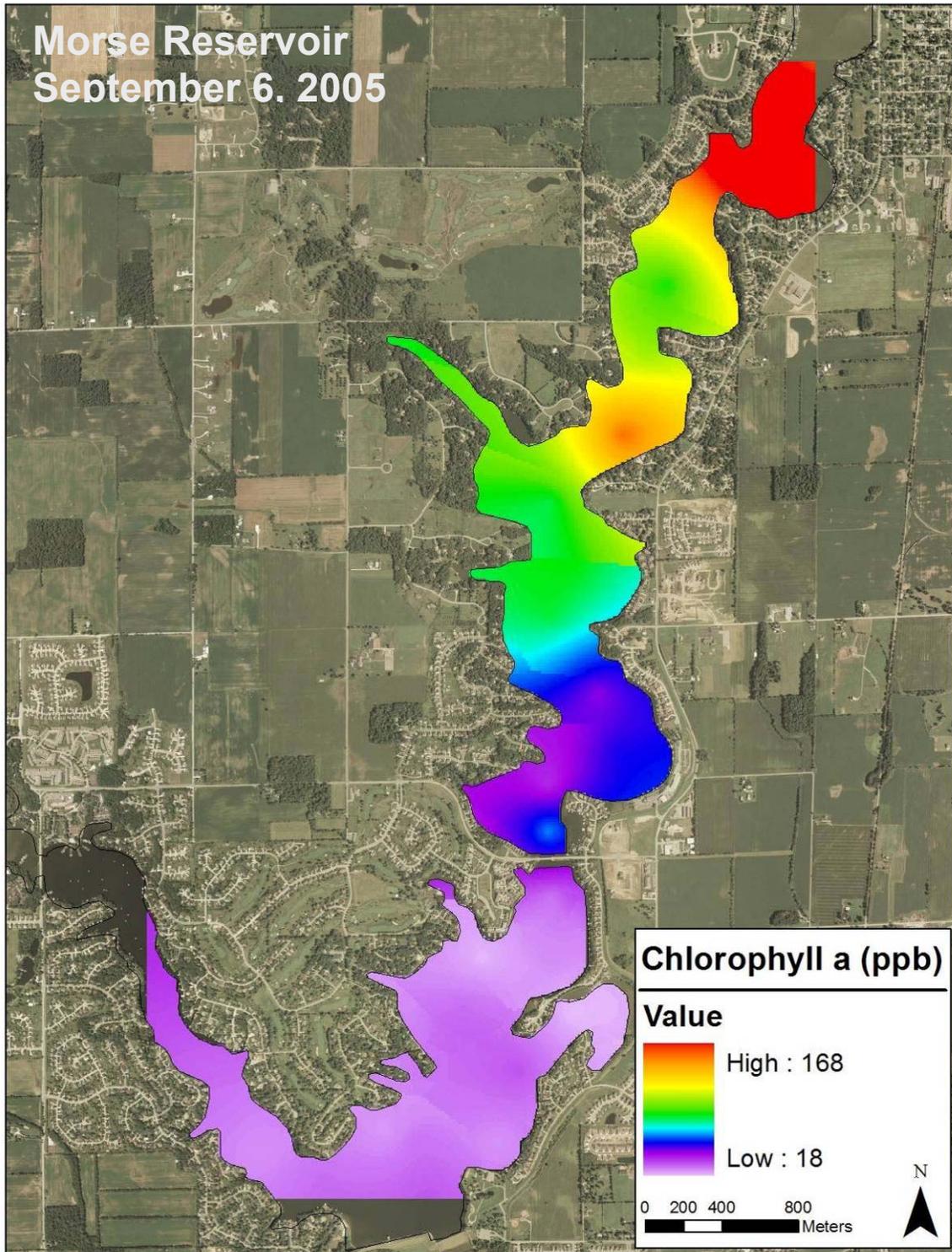
Interpolated maps of extracted and flourometrically measured chlorophyll a concentrations.



## APPENDIX A: Chlorophyll a Maps



## APPENDIX A: Chlorophyll a Maps



## APPENDIX B: Analytical Methods Descriptions

Sample	Analytical Method	Detection Limit	Method Description	Description
Alkalinity (mg/L as CaCO <sub>3</sub> )	EPA (310.1)	2.0	Titrametric	Alkalinity by titration to pH 4.5. Acid neutralizing capacity (sum of all titratable bases). Primarily a function of carbonate, bicarbonate, and hydroxide content (usually an indicator of the concentration of these constituents)
DOC (mgC/L)	SM (5310C)	0.5	Persulfate	Oxidation-Amount of TOC that passes through a 0.45µm-pore-diam filter
TOC (mgC/L)	SM (5310C)	0.5	Persulfate	Oxidation-Persulfate and Ultraviolet Oxidation with IR detection. Carbon atoms covalently bonded in organic molecules are broken down to be measured quantitatively. Organic Carbon is oxidized into CO <sub>2</sub> by persulfate using UV light. CO <sub>2</sub> is removed from the sample, dried, and transferred with a carrier gas to an IR analyzer.
Chloride (mg/L)	EPA (300.0)	8.0	Ion	Method 300.0 is an ion chromatograph method using a Dionex DX-600 system with a conductivity detector. Sample is added to an ion chromatograph. Anions are separated and measured using a guard column, analytical column, suppressor device, and conductivity detector.
Sulfate (mg/L)	EPA (300.0)	8.0	Ion	Method 300.0 is an ion chromatograph method using a Dionex DX-600 system with a conductivity detector. Sample is added to an ion chromatograph. Anions are separated and measured using a guard column, analytical column, suppressor device, and conductivity detector.
O-Phos (mg/L)	EPA (300.0)	0.05	Ion	Method 300.0 is an ion chromatograph method using a Dionex DX-600 system with a conductivity detector. Sample is added to an ion chromatograph. Anions are separated and measured using a guard column, analytical column, suppressor device, and conductivity detector.
Total P (mg/L)	SM (4500-P E.)	0.010	Colorimetric	Ascorbic Acid Colorimetric method-Ammonium molybdate and potassium antimonyl tartrate react in acid with orthophosphate to form phosphomolybdic acid that is reduced to colored molybdenum blue by ascorbic acid.
Nitrite (mg/L)	EPA (300.0)	0.0	IC	Method 300.0 is an ion chromatograph method using a Dionex DX-600 system with a conductivity detector. Sample is added to an ion chromatograph. Anions are separated and measured using a guard column, analytical column, suppressor device, and conductivity detector.
Nitrate (mg/L)	EPA (300.0)	0.10	IC	Method 300.0 is an ion chromatograph method using a Dionex DX-600 system with a conductivity detector. Sample is added to an ion chromatograph. Anions are separated and measured using a guard column, analytical column, suppressor device, and conductivity detector.
NH <sub>4</sub> -N (mg/L)	SM (4110)	0.02	IC	Sample injected into carbonate-bicarbonate and passed through series of ion exchangers. Anions are separated by their relative affinities for for a strongly basic anion exchanger. Anions passed through a fiber suppressor coated with a strong acid solution to convert anions to highly conductive acid form, conductivity is measured. Concentration is determined from measurement of peak height or area.
TKN (mg/L)	EPA (351.4)	0.30	Contracted Out	Determined by digestion, followed by ammonia determination by ion selective Electrode by a contract lab.

## APPENDIX B: Analytical Methods Descriptions

Sample	Analytical Method	Detection Limit	Method Description	Description
Silica (mg/L) unfiltered	EPA (370.1)	0.10	Colorimetric	Heteropoly acids are produced by the addition of Ammonium molybdate (at pH 1.3) to sample containing silica and phosphates. Molybdosilicic acid is preserved and molybdphosphoric destroyed with the addition of oxalic acid. Intensity of yellow color is indicative of concentration of molybdate-reactive silica.
Ca (mg/L)	EPA (300.7)	3.0	IC	Method 300.0 is an ion chromatograph method using a Dionex DX-600 system with a conductivity detector. Sample is added to an ion chromatograph. Anions are separated and measured using a guard column, analytical column, suppressor device, and conductivity detector.
Mg (mg/L)	EPA (300.7)	1.0	IC	Method 300.0 is an ion chromatograph method using a Dionex DX-600 system with a conductivity detector. Sample is added to an ion chromatograph. Anions are separated and measured using a guard column, analytical column, suppressor device, and conductivity detector.
K (mg/L)	EPA (300.7)	0.05	IC	Method 300.0 is an ion chromatograph method using a Dionex DX-600 system with a conductivity detector. Sample is added to an ion chromatograph. Anions are separated and measured using a guard column, analytical column, suppressor device, and conductivity detector.
Na (mg/L)	EPA (300.7)	1.0	C	Method 300.0 is an ion chromatograph method using a Dionex DX-600 system with a conductivity detector. Sample is added to an ion chromatograph. Anions are separated and measured using a guard column, analytical column, suppressor device, and conductivity detector.
Total Hardness (mg)	SM (2340 B)	12.0	Calculation	Total hardness is calculated from the sum of Calcium and Magnesium Concentrations (mg CaCO <sub>3</sub> /L)
MIB (ng/L)	SM (6040)	3.0	Mass Spectrometric	Organics are extracted from water by closed-loop stripping. Extracted organics are injected into a gas chromatograph/mass spectrometer for identification based on retention time and spectrum comparison. Single-ion current integration is used to quantify MIB.
Geosmin (ng/L)	SM (6040)	3.0	Mass Spectrometric	Organics are extracted from water by closed-loop stripping. Extracted organics are injected into a gas chromatograph/mass spectrometer for identification based on retention time and spectrum comparison. Single-ion current integration is used to quantify Geosmin.

## APPENDIX C: Genetic Algorithm Components

A genetic algorithm consists of five steps including encoding, population initialization, individual selection, cross-over and mutation. Each of these steps is briefly described below:

**Encoding:** For spectral variable selection, a genetic algorithm works with binary strings (chromosomes). Chromosomes are generated by mapping spectral variable combinations into binary strings. A gene or bit of a chromosome, randomly set to be "0" or "1," represents a spectral band. If the bit is set to one, this band is selected in PLS modeling. Otherwise, the band is not used. The encoding of band combination into the chromosome follows the order of band number to allow for the neighboring bits representing neighboring bands. The number of integers in a coded string equals to the number of the total spectral bands.

**Population initialization:** GA needs a number of possible candidate solutions to start with. This initial solution pool is generally randomly generated by setting each bit of a chromosome to either zero or one. The initial population should be large.

**Individual selection:** Selection attempts to apply a selection rule upon the population in a manner similar to that of natural selection applied to biological systems. Each chromosome or individual is evaluated based on a predefined fit function that the genetic algorithm intends to optimize. An individual is selected to survive to next generation according to the probability of selection that is an increasing function of the fitness. The selection is done with replacement such that the same chromosome can be selected as parent for multiple times (Forrest, 1993).

**Crossover:** Crossover is one process of reproducing new offspring to simulate gene exchange between natural organisms. During the crossover, a chromosome from each parent is segmented at randomly selected one or several points and recombined with the segments from the other parent based on a crossover rate. The crossover rate is the probability that parents will crossover at a single point. If no crossover takes place, two offspring are formed by copying their respective parents.

**Mutation:** As the other process of reproducing new offspring, mutation operates on the offspring following the crossover to simulate the change of the chromosome due to the random disturbance. The mutation is performed by randomly flipping the value of single bits in a chromosome. The mutation rate is usually kept low to avoid destroying good chromosomes, and this parameter is determined by the population size, selection methods and elitism policy (Yao and Tian, 2003).