Mapping and Modeling Chlorophyll-*a* Concentrations in the Lake Manassas Reservoir Using Landsat Thematic Mapper Satellite Imagery

By Paul Bartholomew

Thesis submitted to the Faculty of Virginia Polytechnic Institute and State University in partial fulfillment for the degree of

Master of Science

In

Civil Engineering

Dr. Randy Dymond, Co chair

Dr Thomas J. Grizzard Jr.

Dr. Randolph Wynne, Co chair

May 21, 2002

Blacksburg, Virginia

Key Words: Remote Sensing, Landsat TM Imagery, Water Quality, Chlorophyll-a

Mapping and Modeling Chlorophyll-a Concentrations in the Lake Manassas Reservoir Using Landsat Thematic Mapper Satellite Imagery

Paul J. Bartholomew

ABSTRACT

Carried out in collaboration with the Occoquan Water Monitoring Lab, this thesis presents the results of research that sought to ascertain the spatial distribution of chlorophyll-a concentrations in the Lake Manassas Reservoir using a combination of Landsat TM satellite imagery and ground based field measurements. Images acquired on May 14, 1998 and March 8, 2000 were analyzed with chlorophyll-a measurements taken on 13, 1998 and March 7, 2000. A ratio of Landsat TM band 3: Landsat Band 4 was used in a regression with data collected at eight water quality monitoring stations run by the Occoquan Watershed Monitoring Lab. Correlation coefficients of 0.76 for the 1998 data and 0.73 for the 2000 data were achieved. Cross validation statistical analysis was used to check the accuracy of the two models. The standard error and error of the estimate were reasonable for the models from both years. In each instance, the ground data was retrieved approximately 24 hours before the Landsat Image acquisition and was a potential source of error. Other sources of error were the small sample size of chlorophyll-a concentration measurements, and the uncertainty involved in the location of the water quality sampling stations.

Acknowledgements

Gratitude goes to my committee members, Dr. Randy Dymond, Dr. Tom Grizzard, and Dr. Randy Wynne, whose support was vital to this research. Dr. James Campbell, head of the Department of Geography, gave support by locating grant opportunities that brought in funding for this work and was always encouraging. Special thanks goes to Craig Moore, PhD candidate in the CEE GIS/CAD Applications group, and Dr. Newland Agbenowosi, former CEE GIS/CAD Applications research assistant. Their encouragement and expert help were invaluable. Harry Post, research assistant at the Occoquan Watershed Monitoring Lab (OWML), provided me with the field data and the details of the OWML data collection system. Critical information for this thesis was obtained on forays with the OWML field measurement team, which includes Phil Spellerberg, Doug Holladay, Mark Lucas, and George Underwood. These professionals went out of their way to help me understand the way water quality data is collected and analyzed. Also appreciated are the efforts of Steven J. Miller, graduate student at the Statistical Consulting lab, who assisted me in ensuring that the results of my analysis are accurately portrayed. Diana Kapiszewski, PhD student at the University of California, Berkley, provided editorial assistance. Finally, I offer special thanks to my father, Francis E. Bartholomew, for his unwavering support and unending encouragement.

Table of Contents

Table of Figures

List of Tables

I. Introduction

Water pollution sources can be categorized into two major types, point source and non-point source pollution. Point sources are those that can be traced to one specified location, such as the outfall from an industrial plant or a failing septic system. Though point source pollution was a problem in the past, the Clean Water Act allowed the federal government to set strict guidelines controlling the levels of pollution from point sources. Most present pollution problems result mainly from non-point sources, which are diffuse or scattered sources in the environment and are difficult to measure and determine their origination. Non-point pollution sources include urban runoff, erosion from construction sites, misguided hydrologic engineering practices, agricultural runoff, landfills, stream bank erosion, and feces from domestic and wild animals.

Detrimental impacts on the environment from non-point source pollution include erosion problems that result in the reduction of the storage capacity in reservoirs, detention basins, and flood control impoundments. These sediments also transport nutrients, such as phosphorous, which can spur unrestrained algal growth. These algal blooms can adversely impact wildlife habitat, decrease the recreational potential of a water body and increase costs of water purification for drinking purposes.

To help monitor and mitigate the damages of non-point pollution on water bodies, state and local governments have funded water monitoring organizations such as the Occoquan Watershed Monitoring Lab located in Northern Virginia. In the decade of the 1960's, new urban development within the Occoquan watershed resulted in water quality degradation from increased flows of poorly treated wastewater. Due to the increase in

nutrient loading into the system, massive algal blooms, episodes of poor smelling and tasting water, fish kills, and hypolimnetic anoxic conditions were common (Grizzard, 2001). In the 1970's the State Water Control Board (SWCB) implemented the "Occoquan Policy" to improve the water quality in the reservoirs. To reduce point source pollution, the SWCB required construction of a regional high performance tertiary treatment installation, the Upper Occoquan Sewage Authority (UOSA) facility, to replace the eleven secondary treatment plants then in use. The SWCB also required the establishment of an independent agency to monitor water quality within the watershed. The Occoquan Watershed Monitoring Laboratory (OWML) was created to monitor the reservoir water quality and advise regulatory agencies, local governments, and public service authorities on the means of best preserving the water supply for its intended use(s) (Occoquan Watershed Monitoring Laboratory, 1998). The problems that still plague the Occoquan Reservoir and Lake Manassas are related to issues of non-point source pollution, including suspended sediment from erosion and urban and agricultural runoff.

As non-point source pollution is a problem that is spatial in nature, it is important to develop methods to track and analyze both the cause and effect of such pollution using ground based monitoring, remote sensing, and GIS. Presently, the OWML has a series of water quality monitoring stations where water is withdrawn and analyzed on a regular basis. The purpose of these monitoring sites is to develop a historical dataset for tracking the health of the watershed and to track the limnological dynamics of the associated reservoirs.

Problem Statement

Few procedures have been developed to ascertain the concentrations of chlorophyll-a in large water bodies using the relatively cheap and historically available datasets of the Landsat Thematic Mapper series of satellite imagery. Though ground monitoring stations track temporal changes taking place at specific points in large water bodies, they cannot detail the characteristics of those changes across the entire surface area of the water body and may miss important details regarding the spatial distribution of this important water quality parameter.

Conceptual Solution

To date, determining the spatial distribution of chlorophyll-a concentrations across open water involved the use of spectral signatures derived from hyperspectral sensors used in controlled laboratory settings or mounted on an aerial platform. The proposal being established here is to integrate the information from the hyperspectral reflectance curves of waterborne algal chlorophyll-a into the wide multispectral bands of Landsat TM data to ascertain the spatial distribution of chlorophyll-a concentrations in the Lake Manassas Reservoir. The procedure uses a combination of processed Landsat TM satellite imagery and ground based field measurements to perform a regression analysis of chlorophyll-a concentration (in milligrams per liter) vs. a function of band three and four image brightness values. The resulting regression formula, if found to be statistically sound, can be used to classify all pixels of the reservoir image to map the two-dimensional continuous surface distribution of the chlorophyll-a.

Research Objectives:

1) Based on previous work developed with hyperspectral data to determine chlorophyll-a concentrations in open waters, develop similar procedures and regression models using the more available wide band Landsat multispectral data.
2) Perform a statistical analysis on the resulting chlorophyll-a concentration models vs. reflectance to verify the adequacy of the procedures.
3) If statistically sound, use the chlorophyll-a vs. reflectance models to produce maps of chlorophyll-a concentrations that occurred at the surface of Lake Manassas on May 14, 1998, and on March 8, 2000.

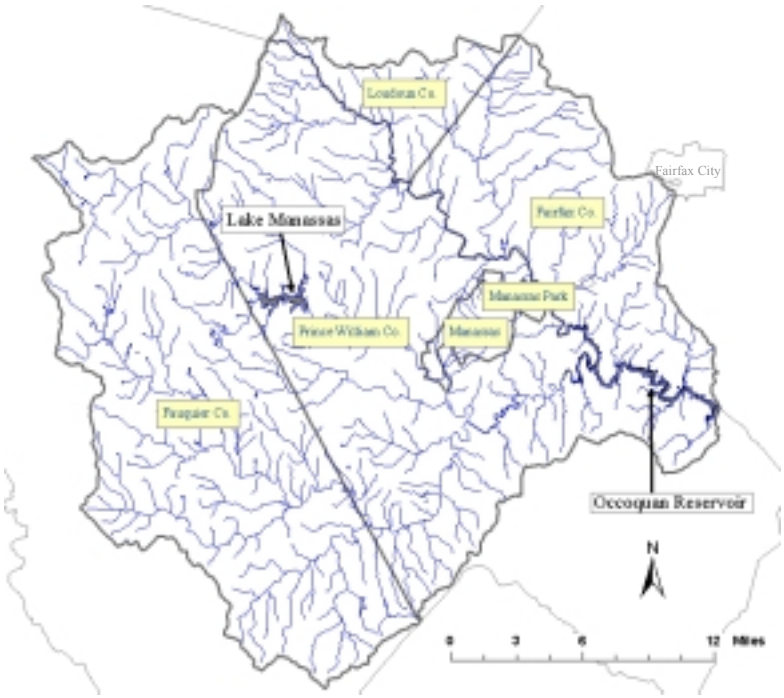


Figure 1 - Occoquan Watershed showing political divisions and reservoir locations.

Lake Manassas Reservoir Overview

Today, the principal problems that plague the Occoquan Reservoir and Lake Manassas are related to issues of non-point source pollution, including suspended sediment from erosion and high nutrient loading from urban and agricultural runoff. Approximately 46,000 acres of the upper Occoquan Watershed drain into the Lake Manassas Reservoir (figure 1). The impounded water serves as the principle source of drinking water for the City of Manassas and parts of western Prince William County. At pool full level of 285 feet mean sea level (MSL), the reservoir surface covers approximately 706 acres. The reservoir is partially surrounded by two golf courses and two residential developments. Though runoff from golf courses is traditionally blamed for high levels of nutrients and pesticides entering into water systems, the Robert Trent Jones International Golf Resort, located on the northern shore of Lake Manassas seems to

have controlled the problem of high nutrient runoff by using appropriate turfgrass management techniques (Eggink, 2001).

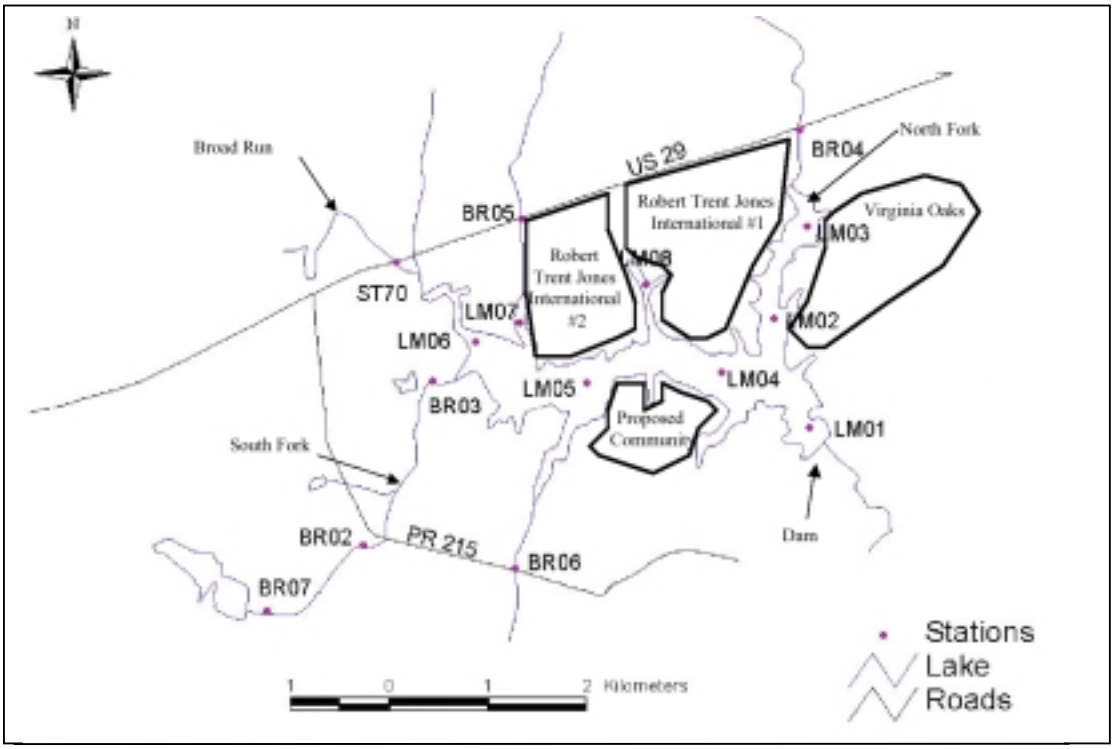


Figure 2 – Lake Manassas with eight water quality monitoring station locations and surrounding landmarks (Eggink, 2001).

However, since Lake Manassas Reservoir came online to serve the public, there have been pollution problems plaguing the water drawn from the reservoir. Some of the most significant quality issues are caused by the excess nutrients of phosphorous and, to a lesser extent, nitrogen washing in from the watershed by way of the tributaries entering the reservoir (Eggink, 2001). The excess nutrient influx causes algae to grow out of control, clogging filters used in the water treatment process and causing odor and taste problems in public tap water. These issues make the monitoring and control of nutrient loading and algal concentrations a very important part of the water management practice.

Nutrient loading problems in the Lake Manassas Reservoir are perpetuated by changing land use characteristics of the upper watershed.

Though the Occoquan Watershed remains predominantly forest, population growth trends are such that jurisdictions are continually torn between protecting the watershed by controlling growth, and allowing more development in order to form a broader tax base. As such, between 1977 and 1995 the amount of pasture and open forest decreased in area by 27%, as urban growth greatly increased (NVPDC, 1991, 1998). Over time, measures implemented to control sources of pollution from runoff have been overwhelmed by steady development, which tends to degrade the quality of the natural water system. Based on metrics such as Carlson's Trophic State Index, the reservoirs in the watershed have remained eutrophic in nature even as measures are taken to arrest their degradation.

Runoff, Chlorophyll-a Concentration, and Algae

In many instances, the degradation of surface waters is caused by sediment that is transported by runoff into streams (Clark II, 1982). Phosphorous, the nutrient that limits the growth of algae in the Lake Manassas Reservoir (Eggink, 2001), binds to these suspended sediments and can become trapped in the bottom of a lake or reservoir as these sediments settle out from runoff water. In most large lakes and reservoirs (especially those found in the southeastern United States), water becomes thermally stratified during the summer months with a warm upper layer (epilimnion) blanketing a cold lower layer (hypolimnion). Because water does not circulate among the stratified layers, oxygen from the surface is unable to reach the hypolimnion except during times of turnover or

destratification. Consequently, the hypolimnion becomes a place of anaerobic activity when oxygen levels are insufficient to meet the oxygen demand of microbes and other benthic creatures throughout the summer stratification. This anoxic condition results in the release of phosphorous, and possibly iron, magnesium, and hydrogen sulfide from the benthic sediments. During times of thermal instability, this released phosphorus-laden water can mix with oxygenated surface water resulting in algae blooms and increased turbidity (Cook, 1989).

II. Literature Review

Concentrations of various suspended substances related to water quality have been successfully mapped using remote sensing (Schalles, 1998). Parameters that have been determined using multispectral reflectance imagery include sestonic organic matter, secchi disk depth, and nephelometric turbidity. The same remote images can be investigated to determine the sources of soil erosion that feed the hydrologic system by combining image analysis with ancillary data in a GIS system (Advances in Remote Sensing, 1999). A short perspective of the history of remote sensing of water quality parameters is given below.

Some of the earliest work involving the remote sensing of water quality parameters came about in the early 1970's when a large quantity of aerial photography and other aerial imagery became available (Ritchie, et. al, 1976). In July of 1972, Landsat 1 was placed in orbit to provide multispectral imagery taken at regular time intervals over the US and many other places on Earth (Jensen, 2000). Early researchers found that the red portion of the electromagnetic spectrum was roughly proportional to the concentration of sediment in open waters (McCluney, 1975). Some of the early work done to determine the spectral response curve of water bodies of various suspended sediment concentrations was done in 1975 using portable boat mounted spectroradiometers that were placed very close to the water surface (approximately 50 cm) to record the near noon sun reflectance of various Mississippi Reservoirs (Schiebe, et al, 1975). The best linear relationship between suspended sediment at the water surface and reflectance was found to be between 7 and 8 micrometers with a correlation coefficient of 0.9. The high correlation suggested that suspended sediments in surface

water of reservoirs could be reliably estimated using remote sensing. Although very interesting at the time, this type of research in hyperspectral analysis of water quality parameters did not have much of a practical application due to the lack of wide-spread hyperspectral data taken from either aerial or satellite based platforms.

Research into more practical applications of multispectral imagery that became widely available in the late seventies were in the form of studies which used multispectral sensors onboard the Landsat series of Multispectral satellites to determine water circulation patterns by remotely sensing the distribution of suspended sediments. At the same time, studies were being conducted which mapped the turbidity of open waters by modeling field measurements with reflectance data. For example, Landsat 1 Multispectral Scanner (MSS) satellite images were used in 1975 to study ocean currents by looking at the spectral response changes due to the suspended matter off the North Carolina Coast (Welby, 1975). In 1989, work was done to track total suspended solids and water transparency (in the form of secchi disk depth) in the Green Bay part of Lake Michigan using Landsat TM imagery (Lathrop, et. al, 1990).

In Minnesota, using similar remote sensing methods like those developed in the Lake Michigan study, a lake water-clarity/trophic state database was developed for more than 500 lakes in the Minneapolis and St. Paul metropolitan areas (Olmanson, 2000). The study used a total of 14 images spanning over 10 years from 1973 to 1998. Lakes were chosen which had data containing at least 20 ground observation points distributed over a wide area on well studied lakes. The field data was taken from work done by non-profit organizations and local area schools. Models developed from these lakes were then applied to more remote lakes having similar limnological parameters in regions of

comparable geomorphology. The resulting database was used to determine spatial patterns of seasonal and temporal trends in the turbidity. The final dataset was then used in a GIS to perform a statistical analysis to determine the cumulative impacts of land development on water quality. Similar studies have been performed in other parts of the world including a eutrophic status assessment of the Te-Chi reservoir in Taiwan using French SPOT multispectral imagery (Yang, 1999), and water quality assessments done on large lakes in Northern Italy using Landsat TM (Giardino, 1998).

Most of the work done with remote sensing of chlorophyll-a, and suspended sediment when it is found in algal laden waters, has been done using either very narrow band multispectral satellite instruments, low resolution hyperspectral sensors carried aboard satellite platforms, or aerial based hyperspectral sensors. An example of a very narrow band (0.02 um) multispectral sensor is the Coastal Zone Color Scanner (CZCS) carried aboard the Nimbus-7 satellite (Mayo, 1993) used to determine suspended sediment concentrations in the Southern Mediterranean where there were very low concentrations of algae. Chlorophyll-a concentrations have been mapped in the open ocean using the low spatial resolution hyperspectral sensor known as the Moderate Resolution Imaging Spectrometer (MODIS) flown aboard NASA's Earth Observation System Terra satellite. Much of the work done to determine the concentrations of chlorophyll-a has been done with spectroradiometers either mounted on ships (Schiebe, et al, 1975) or performed in controlled lab experiments (Han, 1997)

Previous Studies on Remote Sensing Of Chlorophyll-*a* Using Landsat Imagery

There have been some relatively successful attempts in Europe to use Landsat Thematic Mapper data determine the concentration of chlorophyll-a in large lakes and reservoirs. One study was performed while mapping the water quality of Lake Erken in Sweden using CASI 2 hyperspectral data and Landsat TM data (Ostlund, 2001). Although the regression of chlorophyll-a concentration vs. Landsat band two chromaticity green produced a model with a high correlation coefficient ($R^2 = 0.93$), the ground data included very few observations and over-fitting the model to the data set was a problem. The results also showed very little variation in the spatial distribution of chlorophyll-a. Other seemingly similar studies mentioned in the research were dissertations or conference proceedings not available in English.

In an experiment done by Dr. Han at the University of Alabama (1997), large black body tanks and mounted hyperspectral radiometers were used to develop reflectance curves for suspended sediment concentrations in both clear and algal laden water. As part of the research, reflectance curves were developed for clear water, dry and wet (red loam) sediment, water with algae, water with suspended sediment, and finally algal laden water with various concentrations of suspended sediment.

The algae laden water, without sediment, was found to have a maximum reflectance peak at 550 nm green frequency and 700 nm in the infrared (IR) (See figure 3). The suspended plant also exhibits absorption troughs in the blue and red/infrared wavelength locales. The absorption areas of the reflectance curve are caused by the absorption of blue and red light by chlorophyll as it uses the light to produce energy. The

reflectance peak in the green area wavelength is due to the partial reflectance (i.e., less absorption) of the energy by chlorophyll, which causes algae to be perceived as green (Han, 1997).

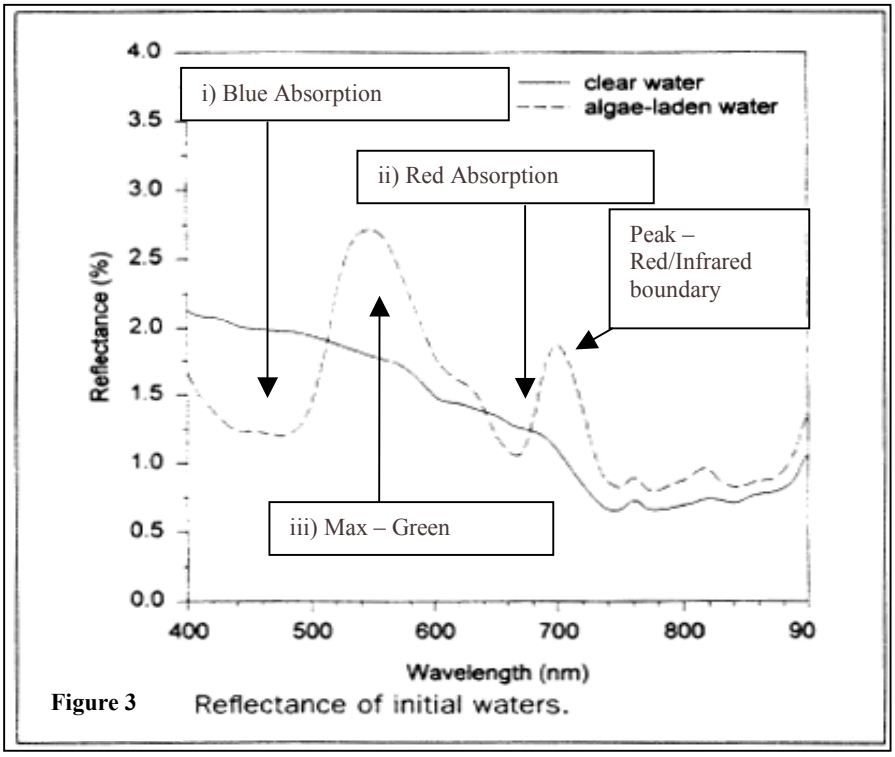


Figure 3 – Reflectance characteristics of clear water and algal laden water. (Han, 1997)

Dry sediments that are similarly analyzed show increases in reflectance between 500 and 900 nm. This reflectance curve is muted when the soil is wet, due to the absorption of light by the water, but retains the basic shape of the reflectance signature. The reflectance curve for various suspended sediment concentrations (SSC) in both clear and algal laden water are shown in figure 4 and 5. For the clear water curve, the reflectance signature graph shows an increase in the percent of reflectance with increasing SSC. The important thing to note is that the soils reflectance in the infrared portion of the spectrum increases linearly with increasing SSC. Also, the coefficient of determination is at or very close to 1.0 for the regression of spectral reflectance values in

the infrared range of approximately 750 nm and 950 nm versus SSC. Further, this area of the reflectance signature is not affected by the presence of algae in the water column (compare figures 4 and 5).

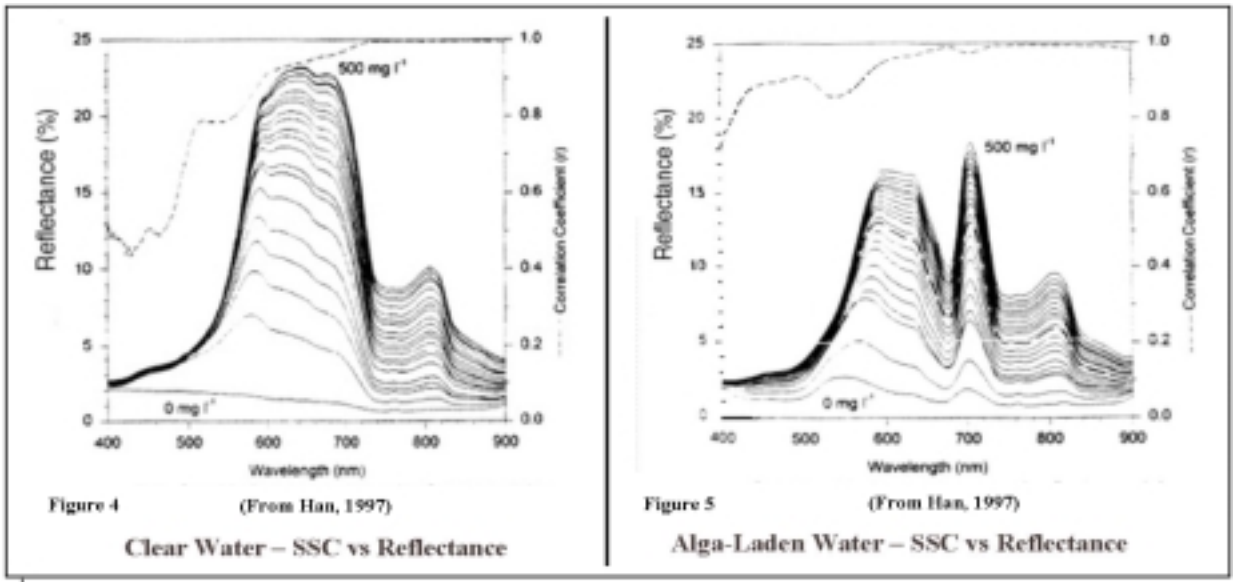


Figure 4 and 5 – Clear water and algal laden water reflectance curves (Han, 1997).

This makes the use of remote sensing a possibility for determining the concentrations of suspended sediment and algae in large water bodies using remote sensing. In theory, if the SSC is known for a location, it can be known what the magnitude of values should be for the red band portion of the spectral curve if the water were otherwise clear. Thus, by comparing the expected clear water reflectance curve with the measured reflectance curve, and calculating how much the curve was depressed by the presence of algae, one can determine the amount of algae in the water.

The use of a band ratio technique, similar to the Normalized Difference Vegetation Index (NDVI) used in land vegetation analysis, simplifies this idea by taking the reciprocal of a function that normalizes the red band suspended chlorophyll-a reflectance/absorption with the near infrared suspended sediment reflectance using a near

infrared (NIR)/RED (R) Ratio. Mittenzewy et al., using hyperspectral data, found a high coefficient of determination (0.98) using the NIR/R reflectance ratio (Mittenzewy, et al., 1992), although Han et al., found that the uncertainty for this ratio grows as the chlorophyll-a concentration dips below 10 ug/L. Another method is to use a NIR-R difference, although this method seems to work only for waters with little or no sediment (Han, et al., 1997).

III. Materials and Methods

Exploration: Method for Extracting Suspended Sediments and Chlorophyll-a Concentrations from Landsat TM Imagery

The above discussion pertains to continuous spectral signatures derived from hyperspectral sensors used in either controlled laboratory settings or aerial platforms. The objective being presented here is to integrate the hyperspectral curve information into the wide multispectral bands of Landsat TM data (see figure 6).

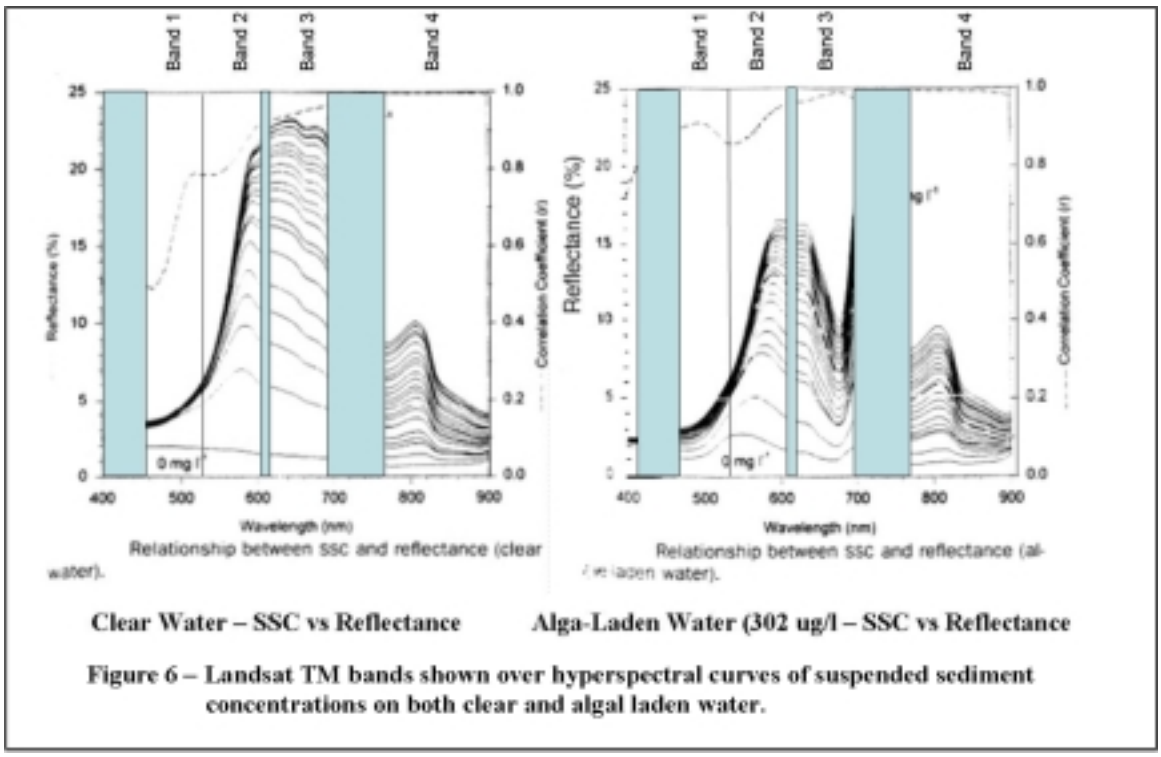


Figure 6 – Landsat TM bands superimposed over the hyperspectral curves of suspended sediment and algal laden water (adapted from Han, 1997).

The full portion of the sediment laden hyperspectral curve not affected by the presence of algae is contained in TM band 4. This portion of the spectral curve increases linearly with increasing sediment loads and theoretically may be used directly in the determination of sediment concentrations. The most significant part of the suspended sediment spectral

curve depressed by the algae is contained in TM band 3. For determining chlorophyll-a concentrations, a R/NIR ratio is proposed that theoretically produces the same 'relatively' normalized value of band-three radiance (see discussion below). The reason for using the R/NIR ratio instead of the inverse NIR/R ratio, discussed above, is to produce a positive correlation of algae concentration vs. reflectance. For the hyperspectral NIR/R ratio, the inverse was taken because the 670 nm band (pure absorption band) used involves the depressed part of the chlorophyll-a spectral curve only; however, the 630 – 690 nm range of the Landsat TM sensor includes part of the green reflectance and also the half of the red peak of the spectral curve, so that the average reflectance value across the band 3 range is assumed to increase with increasing chlorophyll-a concentrations.

As a preliminary exploration, band 4 of the May 4, 1998 Landsat TM image of the Occoquan Reservoir can be seen in figure 7. The changing brightness values, as one traces an eye downstream towards the dam (towards the right of the image), suggests the decreased distribution of suspended sediment concentrations.

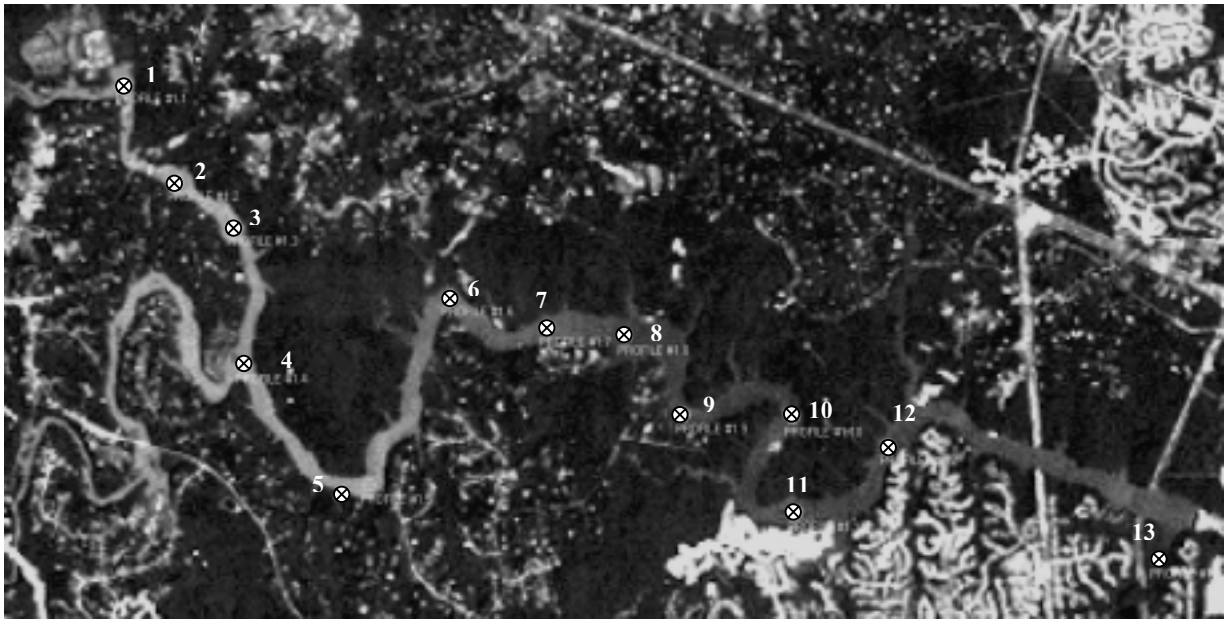


Figure 7- Landsat TM band 4 from May 4, 1998 showing possible distribution of suspended sediment in the reservoir. See figure 8 for spectral profile of points distributed along the image.

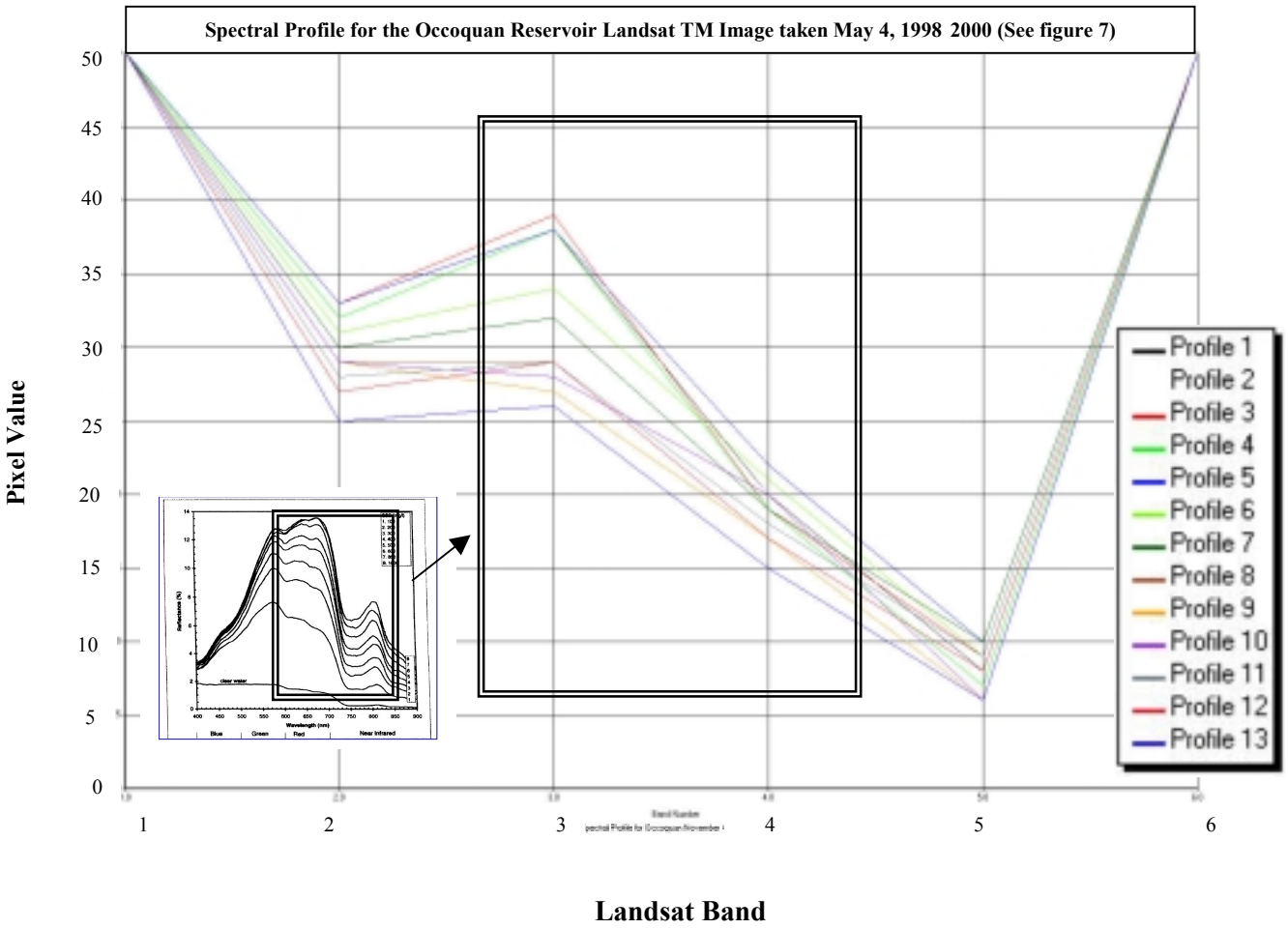


Figure 8 – Spectral Profile of points distributed along the Occoquan Reservoir image in figure 7

Figure 8 shows the 6 band TM spectral signatures (profiles) of the points randomly selected along the length of the Occoquan Reservoir image as shown in figure 7. A graph of suspended sediment reflectance curves for selected suspended sediment concentrations, similar to that shown in figure 4, is inserted for comparison. It is evident that there is separability of the data, however slight, in band four, suggesting that this band can be used in a regression analysis of suspended sediment concentrations in the reservoir.

Figure 9 shows a preliminary look at the possibility of developing an enhanced normalized band ratio that can be used in a regression analysis to determine chlorophyll-a

concentrations. The following normalized band ratio was developed through the use of theoretical reasoning and some trial and error:

((Landsat TM band 3) / (0.0001+Landsat TM band 4))^5

This ratio is meant to normalize the chlorophyll-a value while accounting for the difference in reflectance caused by an increase in suspended sediment as discussed. The 0.0001 was added to the denominator in case of '0' band 4 values, and the exponent was added to give an exponential stretch to the output for visual enhancement. This preliminary exploration is nothing more then a preliminary test of the ideas presented here and not meaningful without a comparison to field data.

Figure 9 - Possible chlorophyll-a distribution produced from an enhanced normalized ratio application.

Imagery Parameters:

Two Landsat images have been obtained for the analysis as follows:

May 14, 1998	Landsat 5 TM	World reference system path 15 - row 34
March 8, 2000	Landsat 7 ETM+	World reference system path 15 - row 34

For the purpose of this research, there is no difference between the band width and spatial resolution of the two satellite platforms and as such, they are completely comparable. Both Landsat 5 Thematic Mapper and Landsat 7 ETM+ have the spatial and spectral characteristics shown in table 1 (Jensen, 2000).

Table 1 – Spatial and spectral characteristics of Landsat 5 TM and Landsat 7 ETM+ satellites (summarized from Jensen, 2000).

Band	Spectral Resolution	Spatial Resolution	Spectral Characteristics
1	0.450 – 0.515 μm	30 x 30 meters	From just below peak transmittance of water to upper limit of blue chlorophyll absorption
2	0.525 – 0.605 μm	30 x 30 meters	Corresponds to red and blue (A.K.A. green) chlorophyll absorption for healthy vegetation
3	0.630 – 0.690 μm	30 x 30 meters	Red chlorophyll absorption region for healthy vegetation
4	0.750 – 0.900 μm	30 x 30 meters	Reflective infrared band – responsive to the amount of vegetation biomass present in scene
5	1.55 – 1.75 μm	30 x 30 meters	Sensitivity to turbidity and the water content of plants
6	10.40 – 12.50 μm	60 x 60 meters	Thermal infrared
7	2.08 – 2.35 μm	30 x 30 meters	Mid-infrared
8	0.52 – 0.90 μm	15 x 15 meters	High resolution panchromatic

Ground Based Field Data

Field data for chlorophyll-a, (by fluorometric analysis in ug/L) are gathered weekly by OWML at four strategically distributed water quality monitoring stations in the Occoquan Reservoir and eight monitoring stations located on Lake Manassas. As mentioned earlier, the field data from the Occoquan Reservoir were determined to be less suited for spectral analysis given that the date of image acquisition was taken a week before the water quality field data acquisition, and because the water moves relatively swiftly through the Occoquan Reservoir due to its riverine nature. Thus, field data from the eight Lake Manassas water quality stations were used for the final analysis.

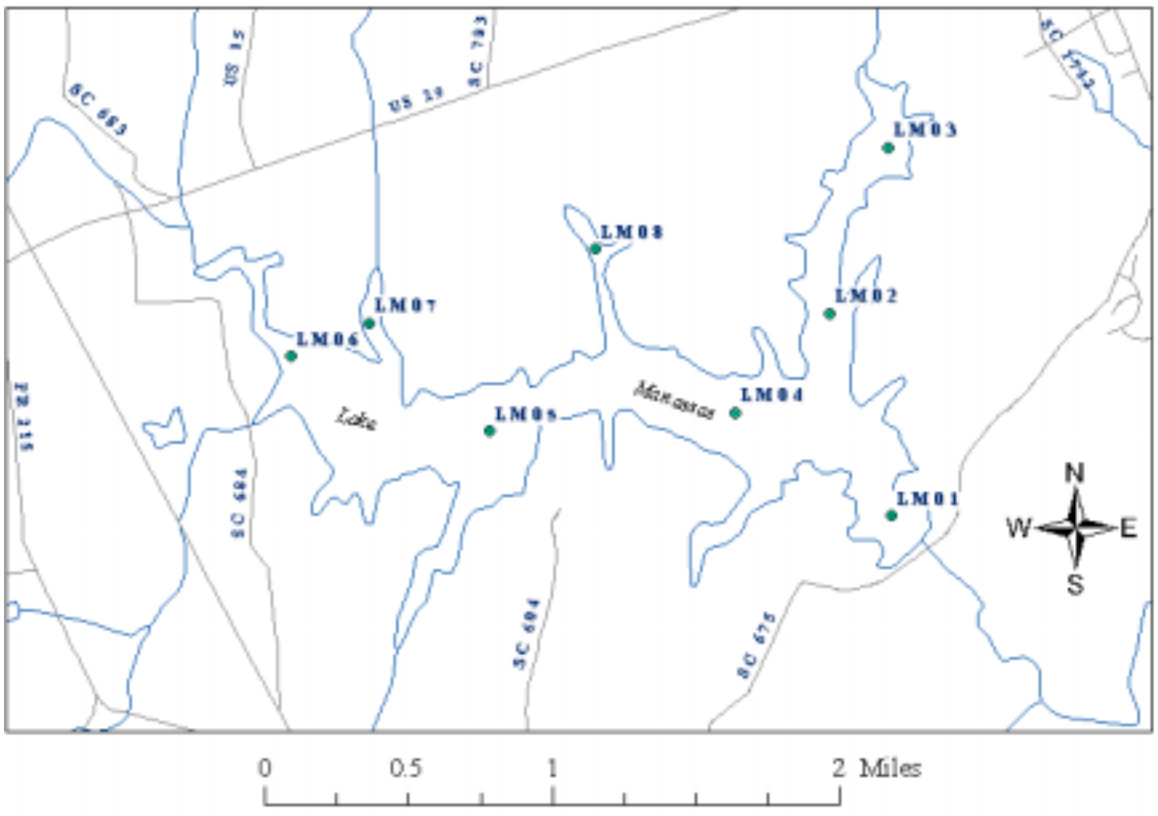


Figure 10 – Eight Lake Manassas water monitoring stations

Chlorophyll-a is a photosynthetic pigment present in all plants. It is a fluorescent molecule, and thus its concentration may be determined by fluorometry (a type of spectral analysis performed in a lab or with a field instrument). The concentration of chlorophyll-a can be related to the concentration of algae in water.

The eight Lake Manassas water monitoring stations are shown in figure 10 and listed in table 2. The location of these field stations were originally located by compass triangulation using appropriate landmarks. The coordinates of each of the stations were determined by taking GPS measurements with a Trimble GeoExplorer III GPS unit (OWML, 2002).

Table 2 – Field data collection stations and locations (OWML, 2002).			
Station	Latitude	Longitude	Field Notes
LM01	38d 45’ 51.55”	77d 37’ 24.62”	Near Lake Manassas Dam
LM02	38d 46’ 26.82”	77d 37’ 41.54”	Near Camp Glenkirk
LM03	38d 47’ 00.81”	77d 37’ 29.91”	Near Sam's Junkyard
LM04	38d 46’ 12.14”	77d 37’ 59.00”	Near RTJ Golf Course
LM05	38d 46’ 02.72”	77d 39’ 03.98”	Near L.M. Marina
LM06	38d 46’ 23.45”	77d 39’ 48.66”	Located very close to the southern shore (within 5 meters) near Broad Run
LM07	38d 46’ 26.88”	77d 39’ 27.19”	Located close to southern shore (within 10 meters) near Thoroughfare Creek
LM08	38d 46’ 40.78”	77d 38’ 35.58”	Located in narrow channel (30 to 35 meters in width) near Beaver Creek

The ranges of the chlorophyll-a data from the above ground stations for both data sets are as shown in table 3 below. The ranges of the data are small in terms of the change that would occur in the reflective response that might be perceived by a Landsat TM satellite sensor and is noted here as a possible problem. The May 1998 data is especially troublesome since it has been stated in the literature that previous remote sensing studies of chlorophyll-a concentrations were less reliable when concentrations were below 10 ug/L (Han, et al., 1997).

Table 3 – Ranges of chlorophyll-a concentrations found in the Lake Manassas field data	
Chlorophyll-a	Total range of 17 to 25 ug/L for 3/7/2000 data
	Total range of 2.4 to 7.2 ug/L for 5/13/1998 data

Preprocessing of Landsat TM data:

Registration of image to appropriate geographical coordinates

The March 8, 2000 Landsat 7 image was received projected to WGS 84 State Plane Virginia north (Lambert Conic). The 1998 Image was rectified to the 2000 image using ERDAS Imagine software using 47 control points (mostly road intersections close to the reservoir) and 15 checkpoints. The total check point error was 0.4675 (less than half a pixel). The accuracy of the projection was visually confirmed by overlaying a VDOT Roads shape file.

Reflectance Conversion of Satellite Data

As seen in table 3, the differences in the ranges of the ground data are very small in terms of the change that would occur in the reflective response. Consequently, removing the noise from the Landsat data is critical for the application proposed here. The raw digital numbers of a Landsat image are not only dependent on the reflectance characteristics of the specific scene, but also contain noise and digital number value offsets that are a result of the viewing geometry of the satellite, the angle of the sun's incoming radiation, atmospheric depth due to viewing angle, and the design characteristics of the sensor. For this reason, the data must first be converted to radiance, which removes the voltage bias and gains from the satellite sensor. The radiance values are then further converted to at-satellite reflectance. This conversion accounts for the varying sun angle due to differences in latitude, season, and time of day, and the variation in the distance between the Earth and Sun. Topographical terrain features that effect the

sun angle, atmospheric effects, or radiation scattering due to non-lambertian surfaces are not accounted for in either calculation.

Radiometric corrections of the satellite digital number values were done for each of the two Landsat images. Radiometric corrections of the 1998 data were performed using an algorithm available in Research System's ENVI software package. The output was saved as an ER Mapper file and imported into ERDAS to be rectified to the 2000 image as described above. Unfortunately, this function in ENVI was only available for Landsat 5 TM data, and not available for Landsat 7 ETM+. As such, the Landsat 7 image (2000 image) was corrected for radiometric noise using the procedure outlined in the appendix on page 65.

Atmospheric Correction

The objective of atmospheric correction is to retrieve the actual 'clear sky' surface reflectance from remotely sensed imagery by removing the specific weather related atmospheric noise from a specific scene. Atmospheric correction has been shown to significantly improve the accuracy of image classification in some instances, but decrease the accuracy in other instances (Wynne, 2002). As such, all subsequent procedures were performed on both the original data (after being converted to radiance) and the data corrected for atmospheric attenuation. The atmospheric correction was done using both the Tasseled Cap method (Jensen, 1996) and the Dark Object Subtraction (DOS) haze reduction method (Chavez, 1988).

The Tasseled Cap method is applicable only to Landsat MSS and Landsat 5 TM data sets and could only be applied to the 1998 Landsat TM 5 image. This method is a linear transformation that produces four new axes termed as follows: 1) soil brightness

index, 2) 4-5 difference, 3) moisture, and 4) haze. The fourth axis, haze, has been determined through rigorous analysis (Jensen, 2000) to be noise, most of which is associated with atmospheric haze. Removing this last axis and performing an inverse Tasseled Cap operation on the data removes this noise from the image. One of the problems with this approach is that it sometimes takes too much information out of the image (Jensen, 1996).

The dark object subtraction method, applied to the 2000 Landsat 7 ETM+ image, assumes that at least one of the pixels in the data set have a true value of zero (i.e., no reflectance). As scattering affects the shorter visible bands to a greater degree than bands of longer wavelengths, this process is generally applied only to Landsat TM bands 1, 2, and 3. Inspection of the band histograms will reveal that data in these bands generally do not begin to appear until some point to the right of zero. This value is noted and the next lowest value is subtracted from the band, thus setting the first digital number value to 1.

As a final data preparation step, the land surrounding the reservoir was masked from both images to increase the efficiency of data manipulation and so that statistics produced from the data would be more meaningful during the sequence of the analysis.

Final Data Preparation

In order to develop the final regression formulas, the following procedures where performed to enhance the accuracy of the information mined from the data.

1. Both of the final data sets, the original processed Landsat TM data and the same data further processed to reduce the effects of atmospheric haze, where analyzed in parallel.

2. New data sets were produced for the analysis of chlorophyll-a concentrations using the Landsat TM band 3 to band 4 ratio.
3. The final data sets were then taken into ArcMap and overlain with a point shape file showing the locations of each of the water quality monitoring stations.
4. The resulting map was queried to obtain the digital number brightness value at each of the monitoring stations. The values were exported into a Microsoft Excel spreadsheet containing corresponding chlorophyll-a measurements taken at the water quality monitoring stations.
5. A regression was performed on each data set.
6. Regression equations and correlation coefficients were compared to determine the best method of analysis pertaining to the following factors:
    a. Whether to use haze reduced data or original data.
    b. Whether to include all stations, or just those clearly in the pelagic zone.
    c. Whether to run the data through smoothing filters to find a stronger correlation between the satellite and field data.
7. The resulting models where checked for accuracy using the cross validation statistical method.

Location Accuracy of the Water Monitoring Stations

When OWML first set up the water monitoring stations on the open waters of the Lake Manassas Reservoir, the stations were located by using landmarks along the shoreline and compass triangulation methods. Some years later, these sites were located with GPS units in order to produce digital maps in a GIS. It was determined that these

original GPS coordinates had not been taken to an accurate enough degree for the present analysis, possibly because of problems in the early implementation of GPS[1]. As such, new GPS measurements were taken while the water monitoring team collected field data in July of 2000. These GPS measurements were taken with the highly accurate Trimble GeoExplorer III GPS unit (5 meter accuracy) along with a backup Garmin 12 GPS unit used as a rough check on the Trimble coordinates (OWML, 2002). However, even with the new coordinates, another source of error was established while taking the new coordinates.

When the monitoring team goes into the field, the crew does not use triangulation or GPS to relocate the station locations, but rather 'eyes up' the correct position and checks the location by testing for the proper depth. All members of this water quality monitoring crew have from 10 to 20 years experience working for OWML on the Occoquan and Lake Manassas Reservoirs and are highly competent in the work that they perform. However, it is not known whether the accuracy of locating the field station, which is both sufficient and efficient for OWML purposes, is accurate enough for the present analysis. A graphical view of the problem is shown in figure 11, where the pixels in the reflectance of band four are shown with the location of station LM07. Referring to figure 11, a five-meter difference in the location of the station changes the results of the data set markedly.

Due to the positional accuracy for the water quality measurement, a sensitivity analysis was performed by looking at pixel values close to the given water monitoring station position, and determining what effect replacing these values had on the regression

[1] The coordinates were taken using an original Garmin 12 GPS unit during a time when the military was still scrambling the signals from GPS satellites, which could cause errors of 30 meters or more. Also, the positions were given only to the degree-second, which is not accurate enough for this application.

results. The sensitivity analysis showed that position errors was suspect in both datasets; as such, all data were run through several types of filters. The various filters are described below in table 4. These various filters were used on the image data to see if the output produced better results in the final analysis. The idea is that, by smoothing the data, the error in location of the field stations may be less prone to extreme digital number values found between adjacent pixels. It should be noted that the data came from the EROS Data Center (EDC), which delivers most imagery preprocessed with convolution filtering already applied. Filtering the data proved to be successful in increasing the correlation coefficient of both models considerably. A discussion of the results and the related graphs can be found in the next section on chlorophyll-a/Landsat image analysis procedures.

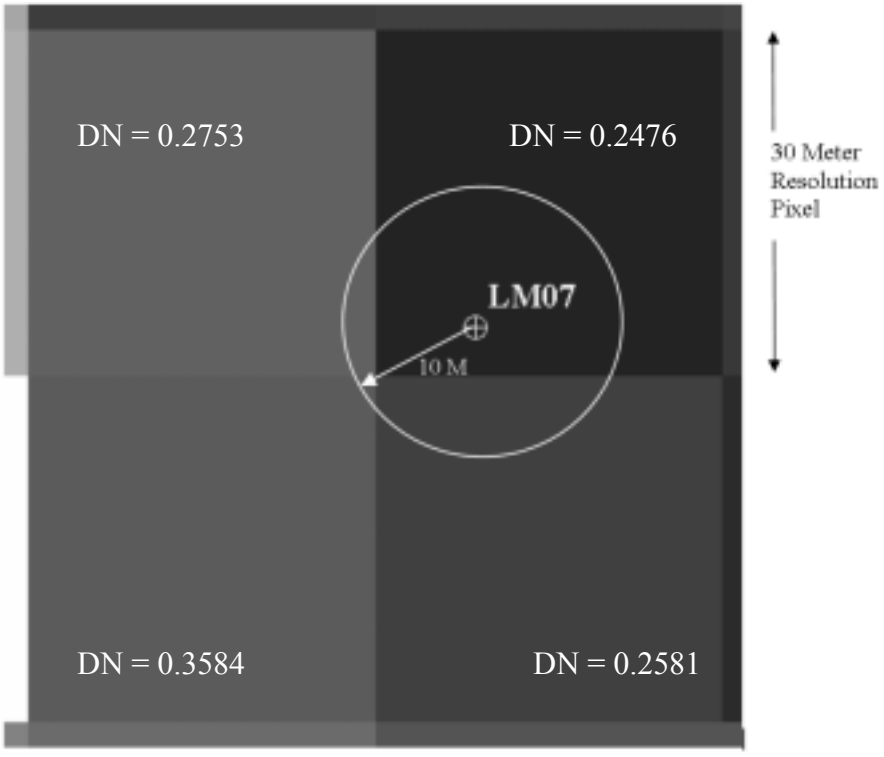


Figure 11 - Square 30 meter pixels of band 4 reflectance values are shown with the position of station LM07. It can be seen that a five-meter change in the stated field collection position can produce a noticeable change in the resulting digital number value and skew regression results.

Table 4 – Description of filtering processes and the characteristics of the resulting image (Jenson, 1996)

Filter Type	Process	Characteristics of output image
Mean (3 x 3 pixel convolution)	A neighborhood analysis technique that calculates the mean value of a 3 x 3 pixel region, placing the result into the position of the central pixel of the matrix. All computations are performed on the original image and results used to build a new output image.	This is a low pass filter that removes high frequency data which may be related to image noise. The data is significantly smoothed and data 'edges' are blurred.
Median (3 x 3 pixel convolution)	Same as Mean Filter except that the central pixel is replaced by the median value of all nine pixels in the neighborhood.	The output image data is 'smoothed', as is the case with the mean filter. The advantage is that edges are preserved in the output data.
Bilinear Interpolation	A neighborhood analysis that takes the inverse distance weighted average of the four nearest neighboring pixels, placing the result into the position of the central pixel of the matrix.	Alters original data and reduces contrast. Pixel DN's are averaged, smoothing edges and extreme values.
Cubic Convolution	Works in a similar fashion to the Bilinear Interpolation, but requires 16 surrounding samples and uses cubic splines rather than a linear interpolation.	Evaluates a block of the 16 nearest pixels (4X 4 Matrix) with a type of a non-linear (cubic) weighted averaging. The mean and variance of the output distribution matches the input distribution, however data values may be altered. The process can both sharpen images and smooth out noise.
The intent of using a filter is to compensate for the distortions introduced by a variety of factors, in this case, errors in the position of field data.		

IV. Analysis Results

As discussed on pages 21 through 25, the post-processed Landsat data were run through the ERDAS Imagine modeler function to create an output image using the following relation:

$$Digital\ Number \equiv Band\ 3\_DN \Big/ (0.000000001 + Band\ 4\_DN)$$

Where

Band3_DN = digital number values of bands three of the input image

Band4_DN = digital number values of bands four of the input image

The ratio function is meant to normalize the magnitude of chlorophyll-a reflectance from the water with the magnitude of the reflectance caused by suspended sediment. Data output from the ratio function was coupled with chlorophyll-a concentration measurements taken from eight water quality-monitoring stations distributed across the reservoir. These data couples were used in forming regression models that predict the chlorophyll-a concentration at any pixel location based on the digital values of the processed image. As described below, the regression analysis for chlorophyll-a showed good results for both the 1998 and 2000 data sets.

Chlorophyll a - 1998 Data

The results of both the haze reduced and non-haze reduced Landsat data analysis is shown in table 9, and the graphs of each exhibited regression are shown in figures 13 through 17.

Referring to table 5, note that the chlorophyll-a measurement of '-2'ug/L for water quality field station LM03 is not used in the regression analysis as the "-2" means the value was below 2 ug/L or not detectable.

The correlation between the chlorophyll-a concentration and reflectance ratio is lower for both the original and haze reduced data sets. Because of the problem of knowing the actual position of where the water quality measurement was taken, as discussed earlier, a sensitivity analysis was performed on the data by looking at pixel values close to the given water monitoring station position, and determining whether using neighboring values in the regression analysis had an effect on the resulting coefficient of determination. The greatest positive effect occurred with the haze-reduced data, and so this data was run through various smoothing filters, as described earlier, with the results shown in figures 12 through 16.

Table 5 - Lake Manassas Data for March 13, 1998 - Pool Elevation 285.62

Monitoring Station Name	Chlorophyll-a Concentration (ug/L)	Band 3/4 Ratio	Band 3/4 Ratio Haze-Reduced	Haze-Reduced B3/B4 Median Filter	Haze-Reduced B3/B4 Bilinear Convolution	Haze-Reduced B3/B4 ratio Cubic Convolution
LM01	7.2	1.2841	1.6803	1.4886	1.6803	1.1086
LM02	4.3	1.3452	1.6803	1.9409	1.6803	1.2698
LM03	-2.0	1.2841	1.6803	1.7056	1.7884	1.3127
LM04	5.7	1.1649	1.3522	1.5314	1.3522	1.1681
LM05	5.7	1.4107	1.8778	1.6819	1.7884	1.4751
LM06	4.9	1.0690	1.5722	0.9717	1.3522	1.0089
LM07	4.3	1.1784	1.3493	1.1847	0.2798	1.193
LM08	2.4	0.7105	0.6808	0.227	0.6808	0.5954

*** LM03 was not used in analysis due an misreported chlorophyll-a concentration value**

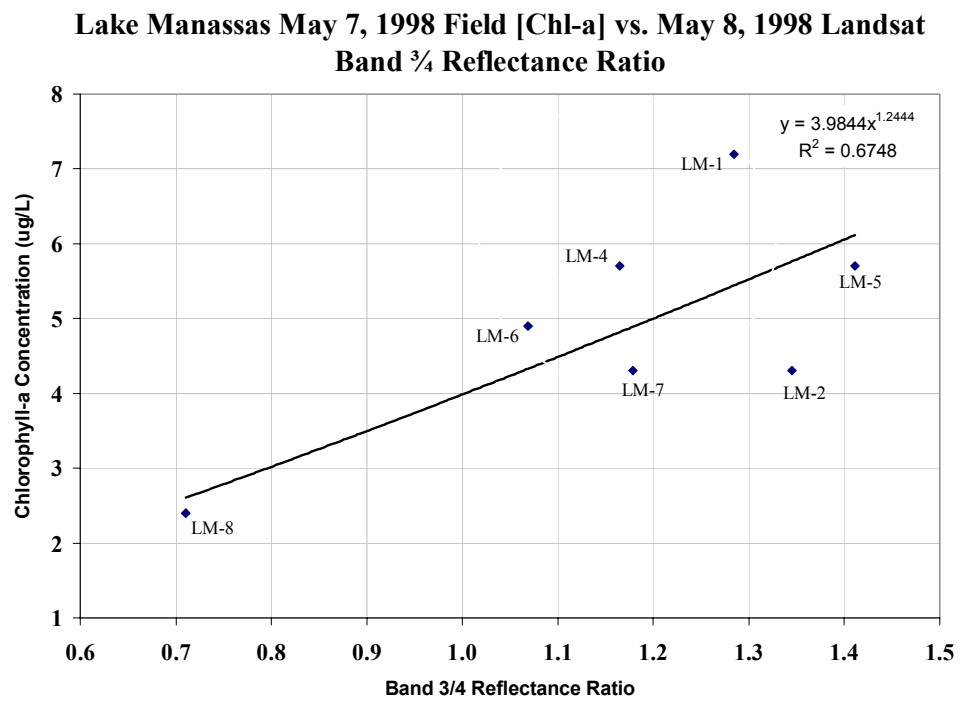


Figure 12 – Lake Manassas May 13, 1998 chlorophyll-a vs. band ¾ ratio. R2 = 0.6748

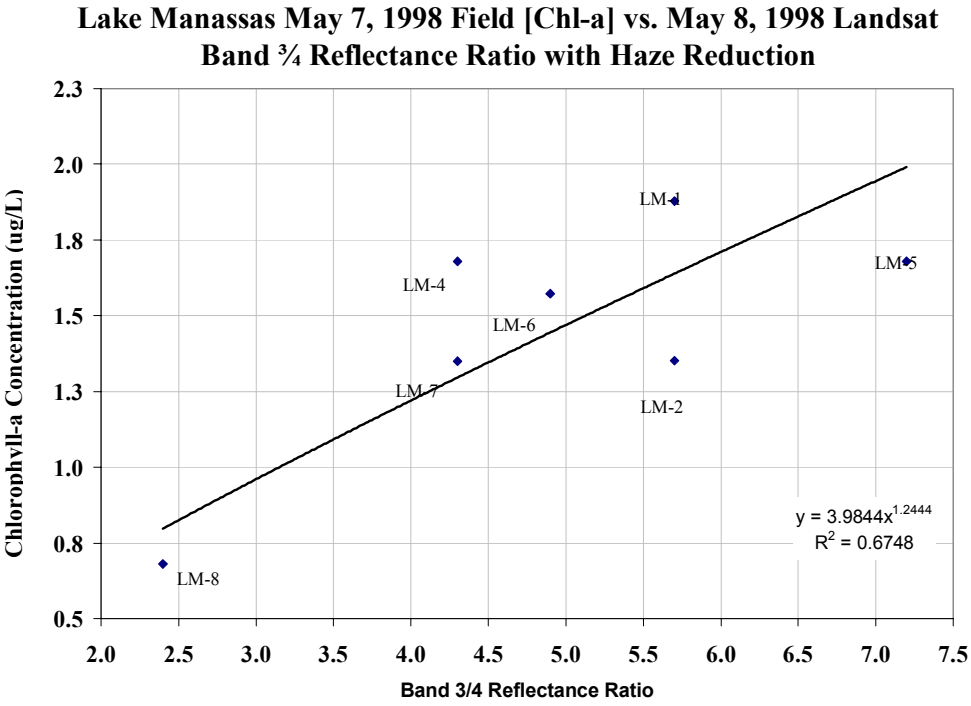


Figure 13 – Lake Manassas May 13, 1998 chlorophyll-a vs. band ¾ ratio with haze reduction. R2 = 0.6748

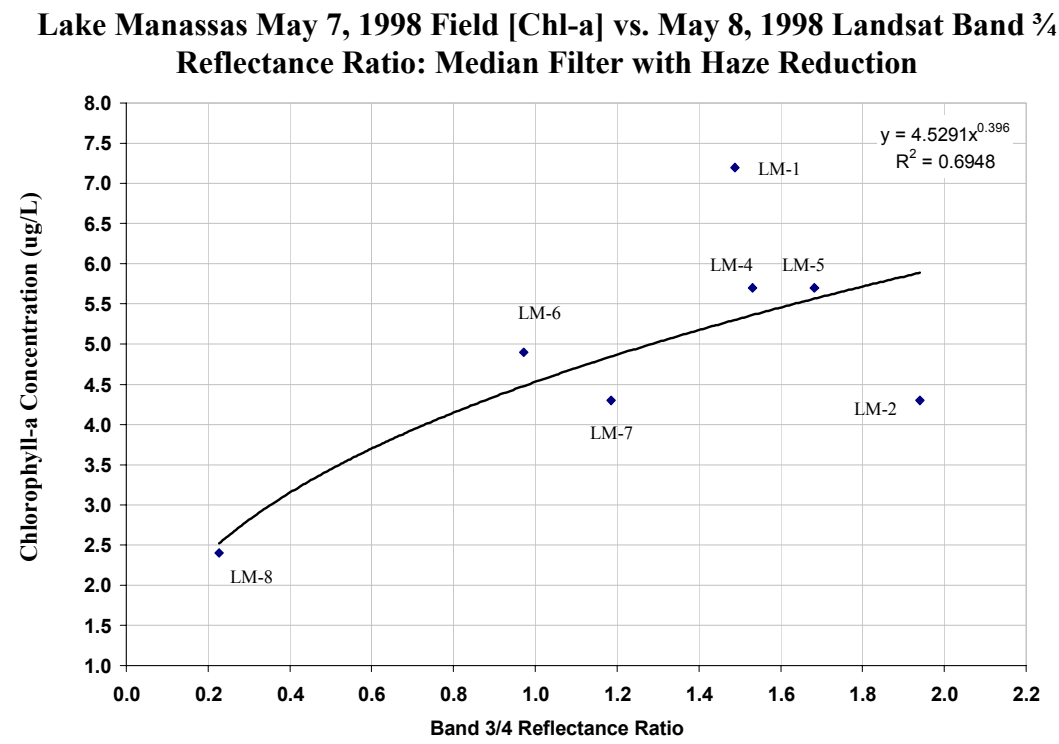


Figure 14 - Lake Manassas May 13, 1998 chlorophyll-a vs. band ¾ ratio with haze reduction and median filtering. $R^2 = 0.6948$

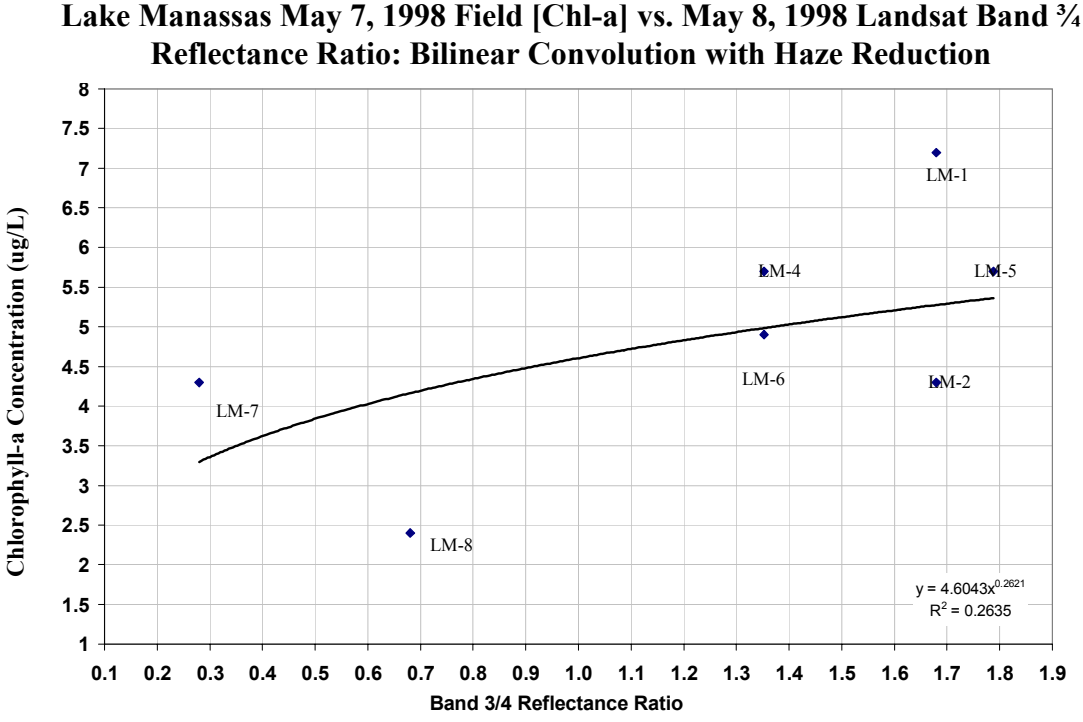


Figure 15 – Lake Manassas May 13, 1998 chlorophyll-a vs. band ¾ ratio with haze reduction and bilinear convolution filter. $R^2 = 0.2635$

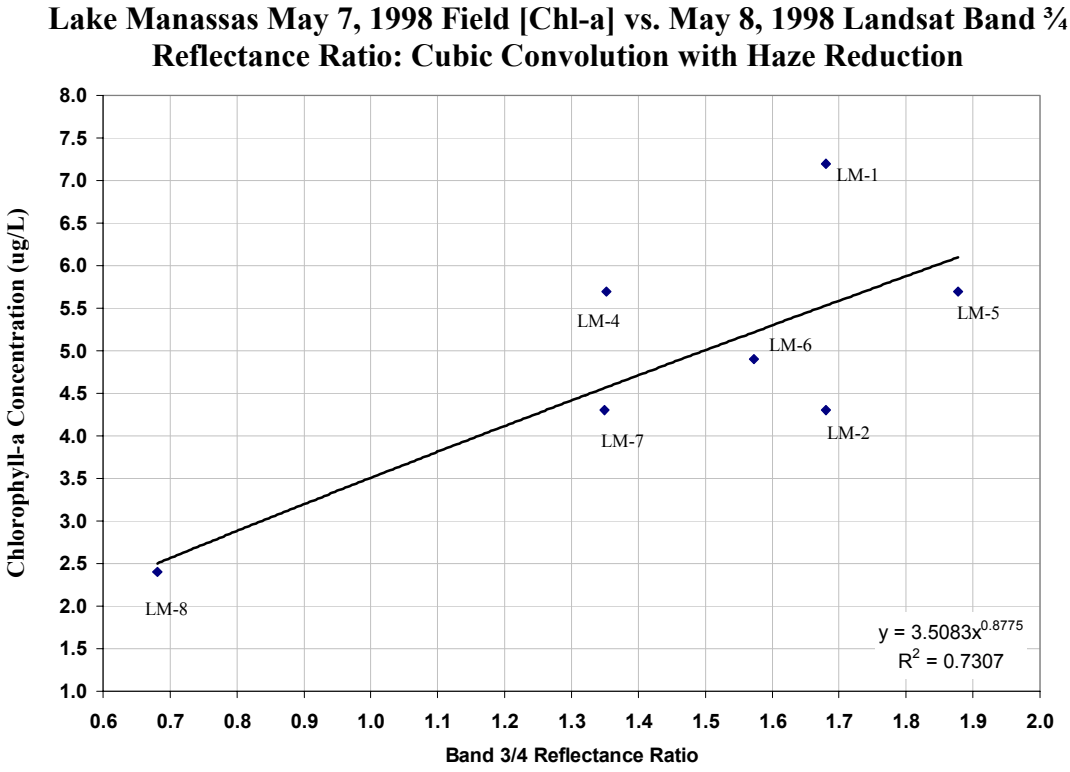


Figure 16 – Haze Reduced with cubic convolution filter: $R^2 = 0.7307$

Chlorophyll a - 2000 Data

The March 7, 2000 data were analyzed with the same procedures described for the May 13, 1998 data (See table 6 and figures 20 through 23). As with the 1998 data, the use of both haze reduction and cubic convolution filter algorithms on the 2000 data greatly increased the success of developing an appropriate model, as shown in figure 23. Station LM08 and station LM06 were taken out of the data set, as they seem to be mixed pixel outliers (See Figures 17 and 18). Station LM06 is very close to the southern shoreline and, as with LM08, may have a mixed pixel value due to shoreline or littoral zone bed reflection.

Table 6 – Water quality and reflectance parameters

Lake Manassas Data for March 7, 2000 - Pool Elevation 288.97

Water Quality Collection Station	Chlorophyll-a Concentration (ug/L)	Band 3/4 Reflectance Ratio	Band 3/4 Ratio w/ Cubic Convolution	Band 3/4 Ratio w/ Haze Reduction	Band 3/4 Ratio w/ Haze Reduction & Cubic Convolution
LM01	25	0.7871	0.8754	0.9035	1.6803
LM02	25	0.8477	0.842	0.4518	1.6803
LM03	21	0.7341	0.7389	1.3545	1.6803
LM04	21	0.7616	0.7362	1.0163	1.3522
LM05	23	0.6816	0.7394	2.7083	1.7884
LM06*	4.4	0.8271	0.8816	0.9035	0.8387
LM07	17	0.6853	0.8496	2.0322	0.2798
LM08*	21	0.9968	0.8402	0.5271	0.1219

*** LM08 & LM06 were not used in analysis due to a mixed pixel difficulty in reflectance parameters**

March 13, 2000 Field [Chl-a] vs. March 14, 2000 Landsat Band ¾ Reflectance Ratio
Residual Analysis – Outlier Detection

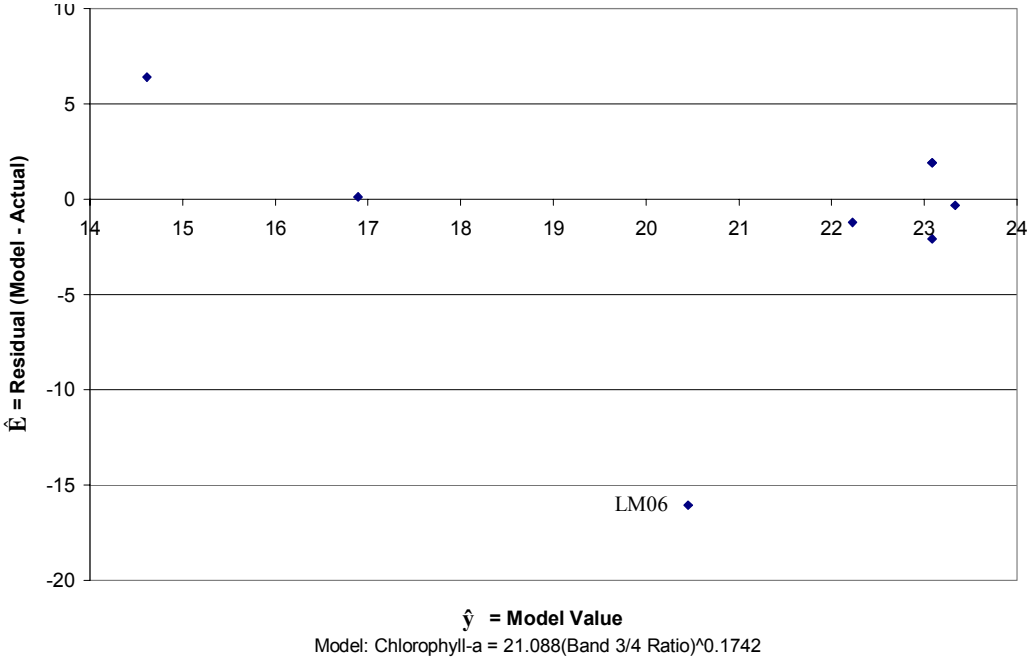


Figure 17 – Outlier detection using residual analysis. LM06 and LM08 are possibly outliers due to mixed reflectance of shore, littoral zone, and water.

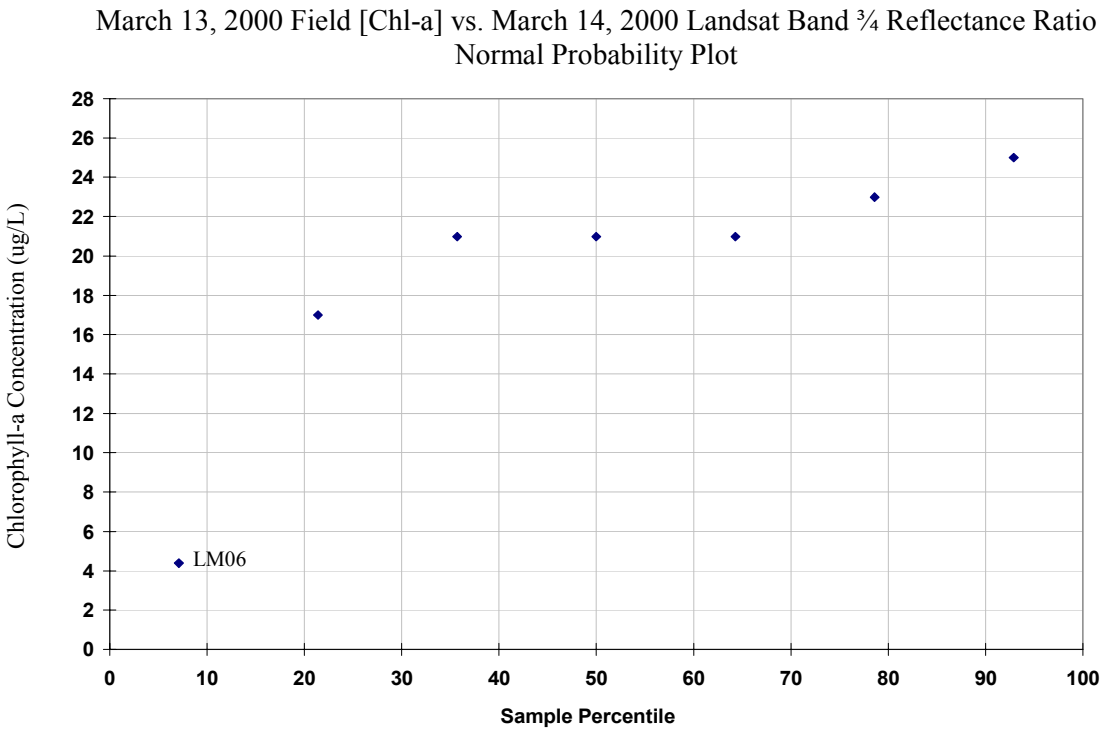


Figure 18 – Outlier detection using probability plot. LM06 is possibly and outlier due to mixed reflectance of shore, littoral zone, and water.

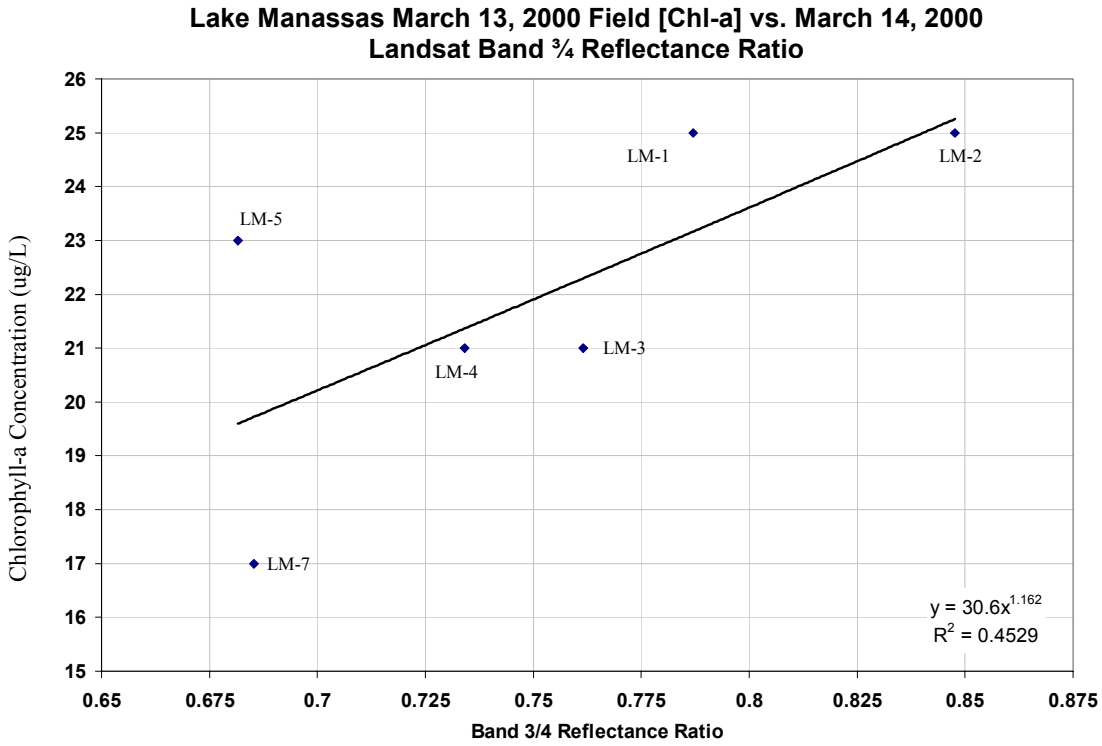
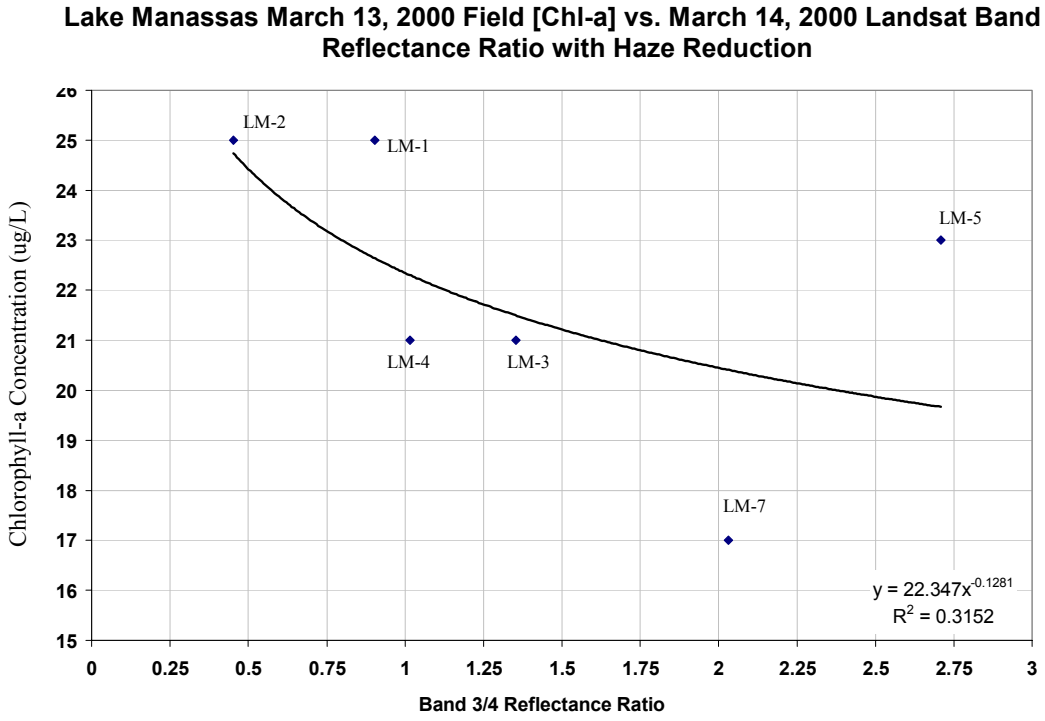


Figure 19 – Chlorophyll-a concentrations vs. band ¾ reflectance ratio: R2 = 0.4629



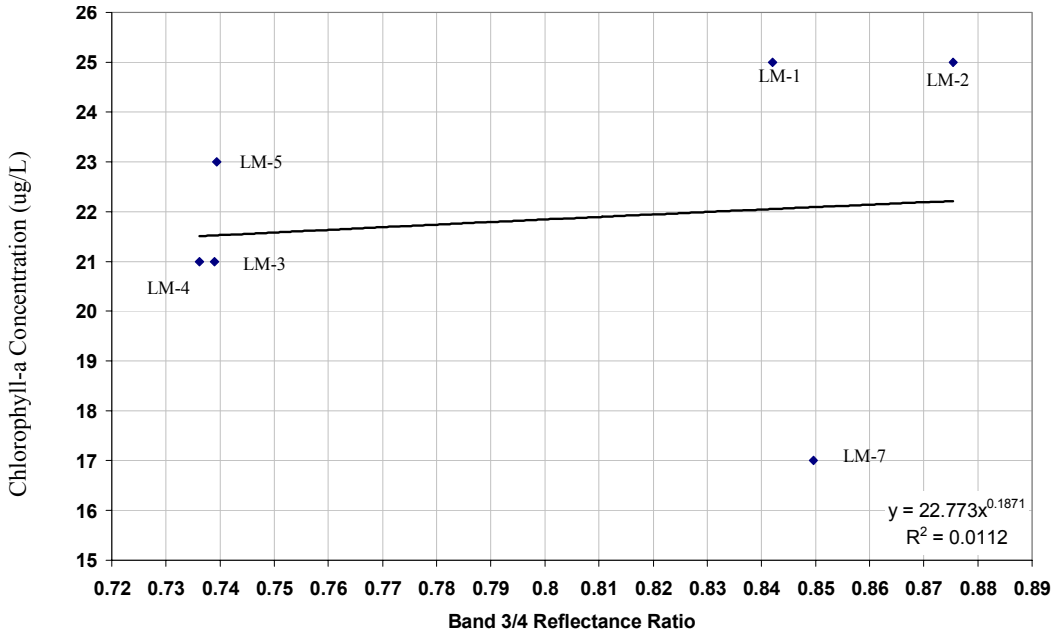


Figure 21 – Chlorophyll-a Concentrations vs. Band ¾ Reflectance Ratio with Cubic Convolution Filtering: $R^2 = 0.0112$

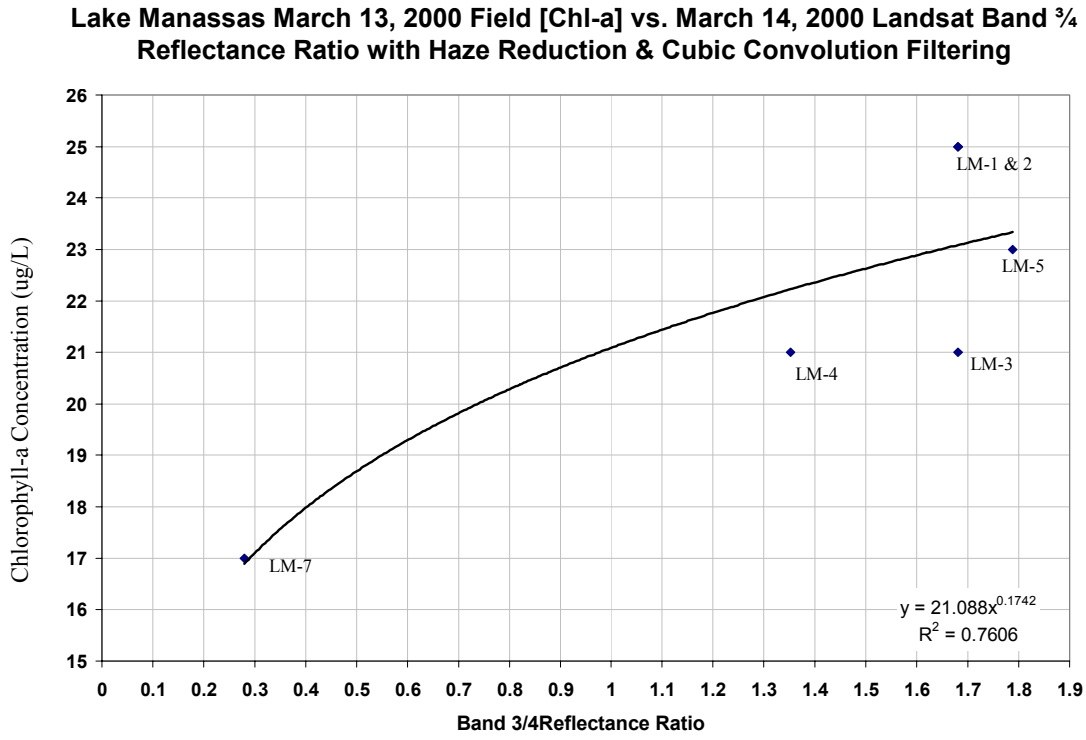


Figure 22 – Chlorophyll-a Concentrations vs. Band ¾ Reflectance Ratio with Cubic Convolution Filtering and Haze Reduction: $R^2 = 0.7606$

Statistics for Model Validation

The final regression models were evaluated using a statistical technique known as "cross validation" (Nachtsheim, 1996). This method is often used to model the replication of an experiment that cannot be repeated because of the specific nature of the study, or because it is neither practical nor feasible to perform a second set of trials. The following section is a summary of the cross validation procedure outlined in *Applied Linear Statistical Models,* 4th Edition, by Kutner Neter and Nachtsheim, Wasserman (1996, pages 436-439).

As an alternative to performing a second analysis for statistical validation of an idea, cross validation breaks the data apart into two data sets, and compares them as if the researcher had a second set of data from a second set of trials. The first set, called the

model-building set or *training sample*, is used to develop the model. The second set, called the *validation* or *prediction set* is used to evaluate the reasonableness and predictive ability of selected model. A regression is performed on the model building dataset and the resulting function is used to predict the values of the validation set. The errors are accumulated to give the mean absolute test set error, which is used to evaluate the model. This basic kind of cross validation is sometimes called the 'Holdout Method".

An enhanced form of cross validation is known as K-fold cross validation. The data set is divided into k subsets, and the holdout method is repeated k times. Each time, one of the k subsets is used as the test set and the other k-1 subsets are put together to form a training set. Then the average error across all k trials is computed. The advantage of this method is that it matters less how the data gets divided. Every data point gets to be in a test set exactly once, and gets to be in a training set k-1 times. The variance of the resulting estimate is reduced as k is increased. The disadvantage of this method is that the training algorithm has to be rerun from scratch k times, which means it takes k times as much computation to make an evaluation.

When the dataset is small, an extreme form of the k-fold validation process called 'Leave-one-out cross validation' can be used. This is simply the K-fold validation with K equal to N, the number of data points in the set. This means that N separate times, the regression analysis is performed on all the data except for one point and a prediction is made for that point. As before, the average error is computed and used to evaluate the model. Because of the small dataset being scrutinized, leave-one-out cross validation was used to analyze the accuracy of the models presented below.

Cross Validation Evaluation Procedure

The following outlines the cross validation procedure for the current analysis.

1) Perform a regression analysis on the given data set N times (N = number of data couples) leaving a different data couple out each time.
2) Compute the model predictions for each of the N regressions.
3) Compute the standard error for each of the models.
4) Compute the total standard error for the cross validation by taking the average of the individual model standard errors.

V. Statistical Results

Table 7 shows the results of the cross validation analysis performed on the reflectance ratio versus chlorophyll-a concentration regression models for the May 13, 1998 dataset. The graphical results of each regression are shown in figures 23 through 28.

Table 7 - May 13, 1998 Cross Validation Analysis

Cross Validation - Leave One Out Method Lake Manassas 1998 Data				
Station	**Chlorophyll-a Concentration (ug/L)**	**Band 3/4 Ratio**	**Model Prediction**	**(Actual-Predict)^2**
LM01	7.2	1.6803	5.5	2.78
LM02	4.3	1.6803	5.5	1.52
LM04	5.7	1.3522	4.6	1.27
LM05	5.7	1.8778	6.1	0.16
LM06	4.9	1.5722	5.2	0.10
LM07	4.3	1.3493	4.6	0.07
LM08	2.4	0.6808	2.5	0.01
Standard Error of the Estimate √Σ((Actual-Predict)^2/5) =				**2.2**

Dataset minus -LM08				
Station	**Chlorophyll-a Concentration (ug/L)**	**Band 3/4 Ratio**	**Model Prediction**	**(Actual-Predict)^2**
LM01	7.2	1.6803	5.4	3.15
LM02	4.3	1.6803	5.4	1.26
LM04	5.7	1.3522	4.9	0.64
LM05	5.7	1.8778	5.7	0.00
LM06	4.9	1.5722	5.3	0.13
LM07	4.3	1.3493	4.9	0.35
Standard Error of the Estimate √Σ((Actual-Predict)^2/5) =				**2.3**
LM08	**2.4**	**0.2798**	**2.3**	
			Error of Estimate	**0.1**

Dataset minus LM07				
Station	**Chlorophyll-a Concentration (ug/L)**	**Band 3/4 Ratio**	**Model Prediction**	**(Actual-Predict)^2**
LM01	7.2	1.6803	5.6	2.61
LM02	4.3	1.6803	5.6	1.65
LM04	5.7	1.3522	4.6	1.17
LM05	5.7	1.8778	6.2	0.21
LM06	4.9	1.5722	5.3	0.14
LM08	2.4	0.6808	2.5	0.02
Standard Error of the Estimate √Σ((Actual-Predict)^2/5) =				**2.4**
LM07	**4.3**	**1.3493**	**4.6**	
			Error of Estimate	**0.3**

Table 7 - May 13, 1998 Cross Validation Analysis continued

Dataset minus LM05				
Station	Chlorophyll-a Concentration (ug/L)	Band 3/4 Ratio	Model Prediction	(Actual-Predict)^2
LM01	7.2	1.6803	5.6	2.41
LM02	4.3	1.6803	5.6	1.81
LM04	5.7	1.3522	4.6	1.15
LM06	4.9	1.5722	5.3	0.17
LM07	4.3	1.3493	4.6	0.10
LM08	2.4	0.6808	2.465	0.004
			Standard Error of the Estimate √Σ((Actual-Predict)^2/5) =	**2.3**
LM05	**5.7**	**1.7884**	**6.0**	
			Error of Estimate	**0.3**

Dataset minus LM04				
Station	Chlorophyll-a Concentration (ug/L)	Band 3/4 Ratio	Model Prediction	(Actual-Predict)^2
LM01	7.2	1.6803	5.3	3.45
LM02	4.3	1.6803	5.3	1.09
LM05	5.7	1.8778	5.9	0.04
LM06	4.9	1.5722	5.0	0.02
LM07	4.3	1.3493	4.4	0.01
LM08	2.4	0.6808	2.4	0.00
			Standard Error of the Estimate √Σ((Actual-Predict)^2/5) =	**1.9**
LM04	**5.7**	**1.3522**	**4.4**	
		Error of Estimate		**1.3**

Dataset minus LM02				
Station	Chlorophyll-a Concentration (ug/L)	Band 3/4 Ratio	Model Prediction	(Actual-Predict)^2
LM01	7.2	1.6803	5.9	1.76
LM04	5.7	1.3522	4.8	0.87
LM05	5.7	1.8778	6.5	0.70
LM06	4.9	1.5722	5.5	0.37
LM07	4.3	1.3493	4.8	0.21
LM08	2.4	0.6808	2.465	0.004
			Standard Error of the Estimate √Σ((Actual-Predict)^2/5) =	**1.6**
LM02	**4.3**	**1.6803**	**5.9**	
			Error of Estimate	**1.6**

Table 7 - May 13, 1998 Cross Validation Analysis

Dataset minus LM01				
Station	Chlorophyll-a Concentration (ug/L)	Band 3/4 Ratio	Model Prediction	(Actual-Predict)^2
LM02	4.3	1.6803	5.2	0.80
LM04	5.7	1.3522	4.4	1.75
LM05	5.7	1.8778	5.7	0.00
LM06	4.9	1.5722	4.9	0.00
LM07	4.30	1.3493	4.37	0.005
LM08	2.4	0.6808	2.5	0.02
Standard Error of the Estimate √Σ((Actual-Predict)^2/5) =				**1.1**
LM01	**7.2**	**1.6803**	**5.2**	
			Error of Estimate	**2.0**

Cross Validation Standard Error	**0.9**
Average Standard Error of the Estimate	**1.5**

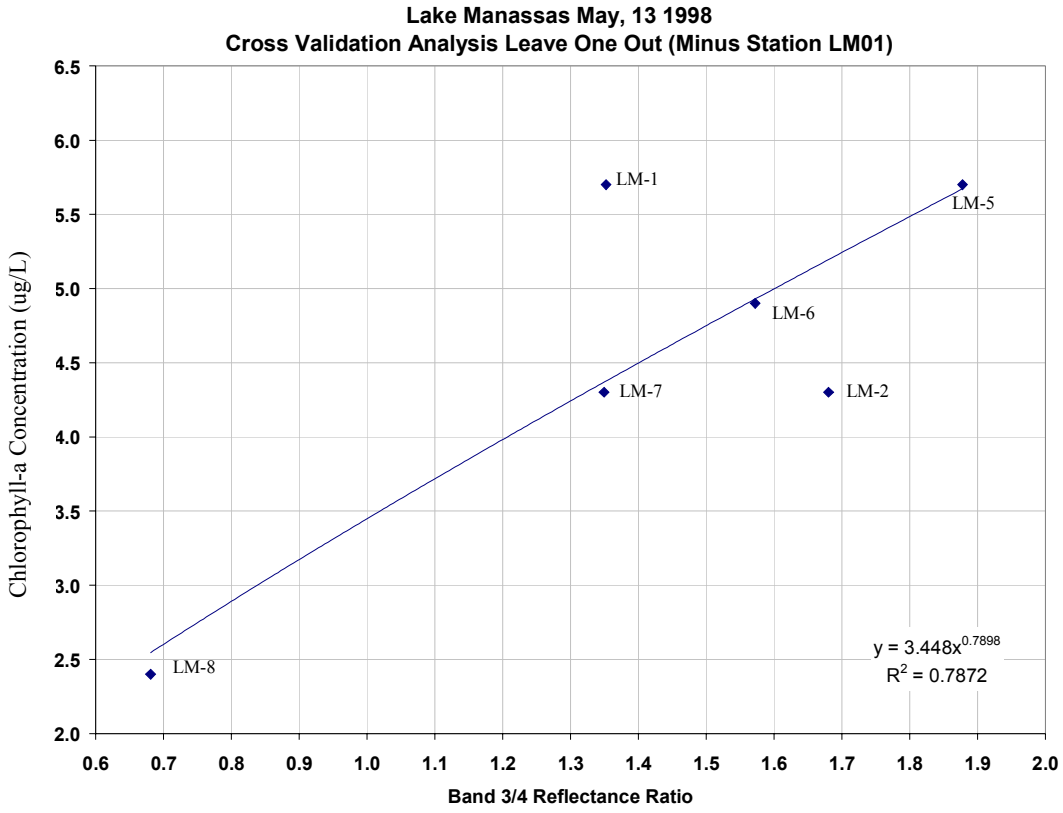


Figure 23 – 1998 Chlorophyll-a Concentrations vs. Band ¾ Reflectance Ratio with Cubic Convolution Filtering – Data minus LM01

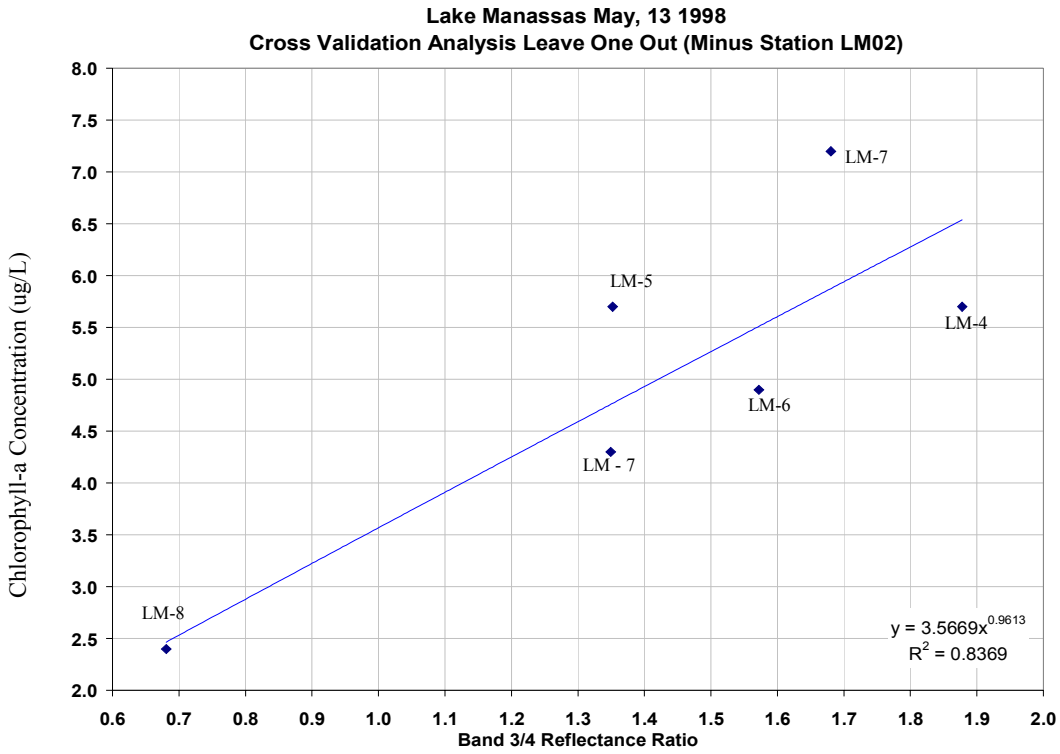


Figure 24 – 1998 Chlorophyll-a Concentrations vs. Band ¾ Reflectance Ratio with Cubic Convolution Filtering – Data minus LM02

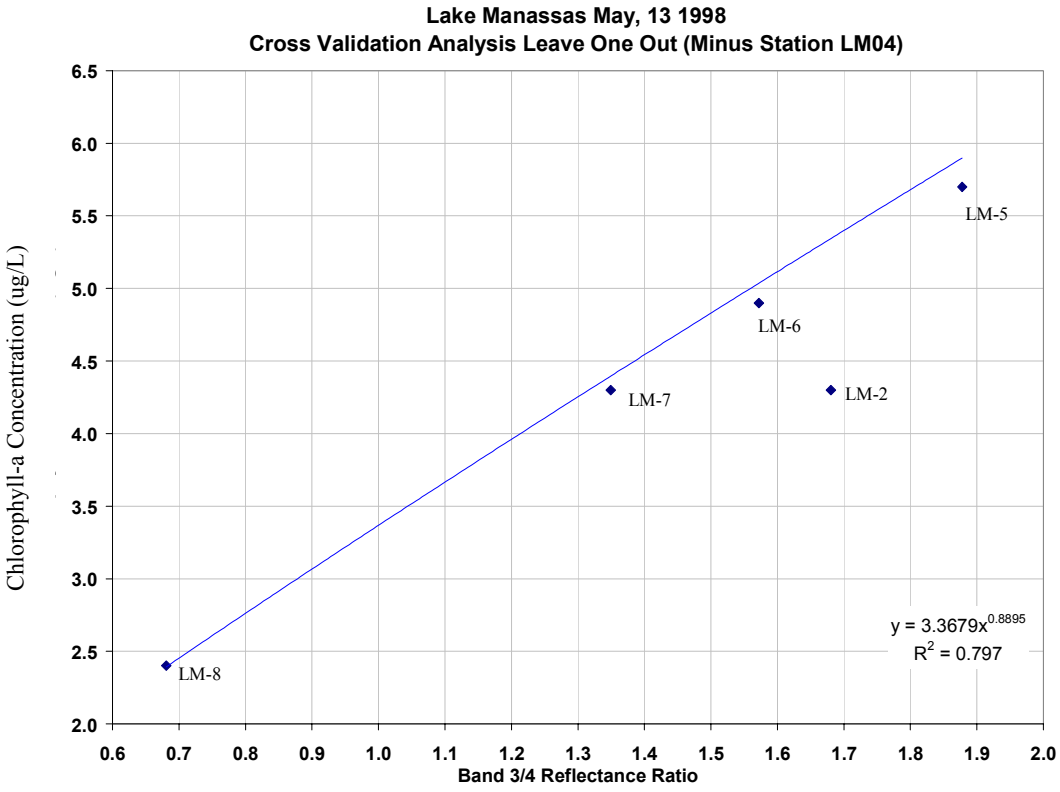


Figure 25 – 1998 Chlorophyll-a Concentrations vs. Band ¾ Reflectance Ratio with Cubic Convolution Filtering – Data minus LM04

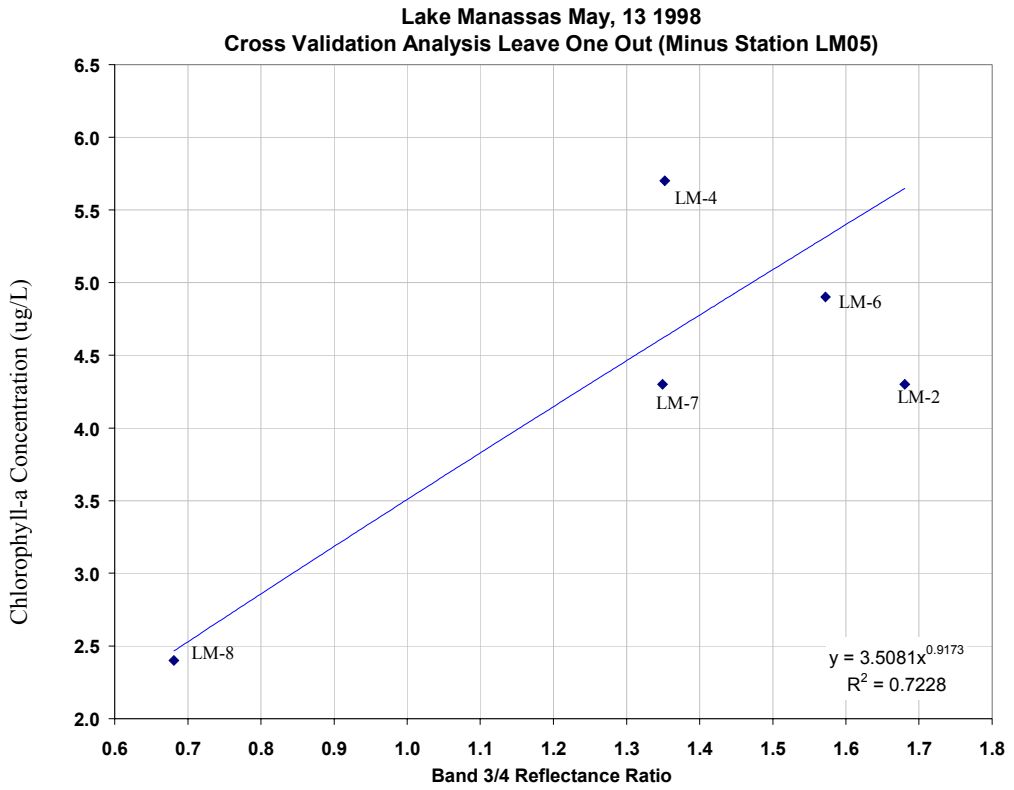


Figure 26 – 1998 Chlorophyll-a Concentrations vs. Band ¾ Reflectance Ratio with Cubic Convolution Filtering – Data minus LM05

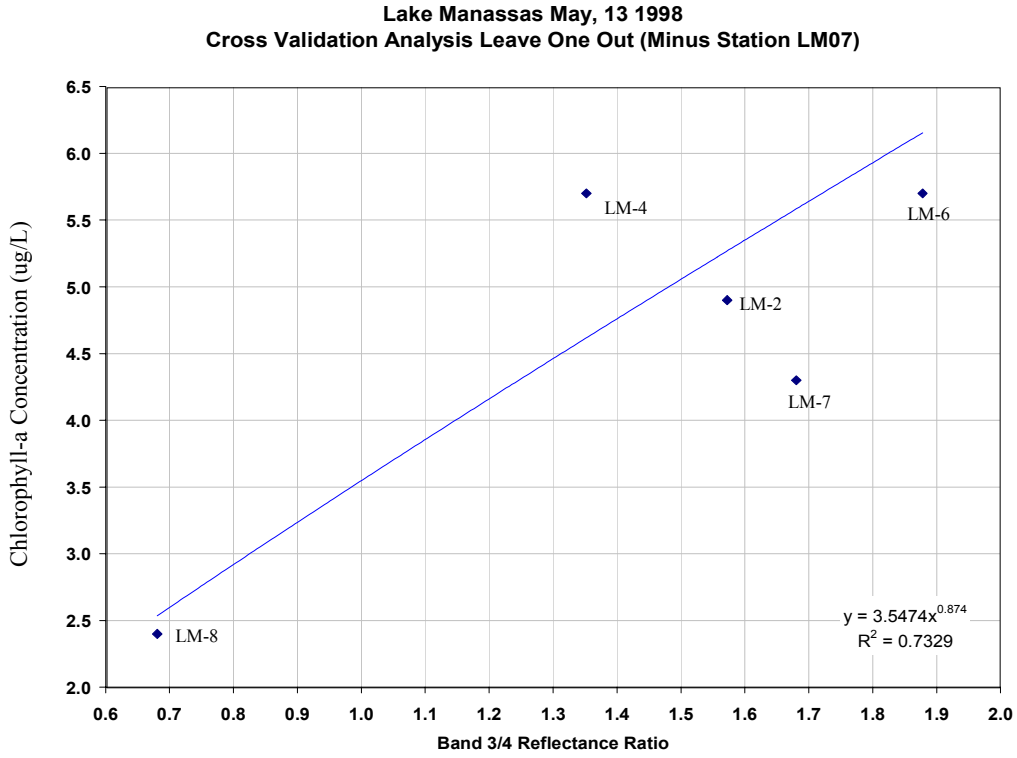


Figure 27 – 1998 Chlorophyll-a Concentrations vs. Band ¾ Reflectance Ratio with Cubic Convolution Filtering – Data minus LM07

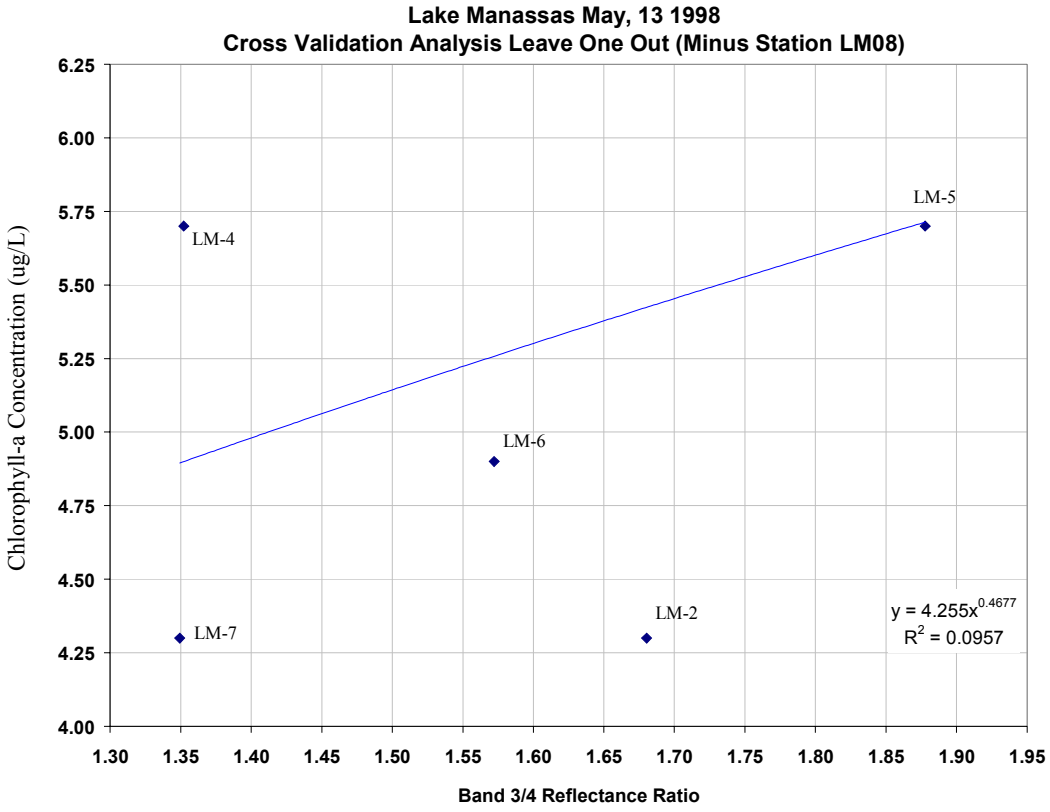


Figure 28 – 1998 Chlorophyll-a Concentrations vs. Band ¾ Reflectance Ratio with Cubic Convolution Filtering – Data minus LM08

Table 8 shows the results of the cross validation analysis performed on the reflectance ratio versus chlorophyll-a concentration regression models for the March 7, 2000 dataset.

The graphical results of each regression are shown in figures 29 through 34. Maps of the chlorophyll-a distribution, based on the present analysis, are shown in figures 35 and 36. A discussion of the results follows in the final chapter.

Table 8 - March 7, 2000 Cross Validation Analysis

Cross Validation - Leave One Out Method Lake Manassas 2000 Data				
Station	**Chlorophyll-a Concentration (ug/L)**	**Band 3/4 Ratio**	**Model Prediction**	**(Actual-Predict)^2**
LM01	25	1.6803	23.1	3.7
LM02	25	1.6803	23.1	3.7
LM03	21	1.6803	23.1	4.3
LM04	21	1.3522	22.2	1.5
LM05	23	1.7884	23.3	0.1
LM07	17	0.2798	16.9	0.0
			√Σ((Actual-Predict)^2/6)	**2.2**

Dataset minus LM07				
Station	**Chlorophyll-a Concentration (ug/L)**	**Band 3/4 Ratio**	**Model Prediction**	**(Actual-Predict)^2**
LM01	25	1.6803	23.2	3.1
LM02	25	1.6803	23.2	3.1
LM03	21	1.6803	23.2	5.0
LM04	21	1.3522	21.2	0.0
LM05	23	1.7884	23.8	0.7
		Standard Error of the	√Σ((Actual-Predict)^2/5) =	**5.4**
LM07	**17**	**0.2798**	**11.0**	
			Error of Estimate	**6.0**

Dataset minus LM05				
Station	**Chlorophyll-a Concentration (ug/L)**	**Band 3/4 Ratio**	**Model Prediction**	**(Actual-Predict)^2**
LM01	25	1.6803	23.2	3.3
LM02	25	1.6803	23.2	3.3
LM03	21	1.6803	23.2	4.7
LM04	21	1.3522	22.3	1.7
LM07	17	0.2798	16.9	0.0
		Standard Error of the	√Σ((Actual-Predict)^2/5) =	**5.9**
LM05	**23**	**1.7884**	**23.4**	
			Error of Estimate	**0.4**

Dataset minus LM04				
Station	**Chlorophyll-a Concentration (ug/L)**	**Band 3/4 Ratio**	**Model Prediction**	**(Actual-Predict)^2**
LM01	25	1.6803	23.4	2.7
LM02	25	1.6803	23.4	2.7
LM03	21	1.6803	23.4	5.6
LM05	23	1.7884	23.6	0.4
LM07	17	0.2798	17.0	0.0
		Standard Error of the	√Σ((Actual-Predict)^2/5) =	**5.1**
LM04	**21**	**1.3522**	**22.5**	
			Error of Estimate	**1.5**

Table 8 – March 7, 2000 Cross Validation Analysis continued

Dataset minus LM03				
Station	**Chlorophyll-a Concentration (ug/L)**	**Band 3/4 Ratio**	**Model Prediction**	**(Actual-Predict)^2**
LM01	25	1.6803	23.7	1.8
LM02	25	1.6803	23.7	1.8
LM04	21	1.3522	22.7	2.9
LM05	23	1.7884	23.9	0.9
LM07	17	0.2798	16.9	0.0
		Standard Error of the	√Σ((Actual-Predict)^2/5) =	**3.3**
LM03	**21**	**1.6803**	**23.7**	
			Error of Estimate	**2.7**

Dataset minus LM02				
Station	**Chlorophyll-a Concentration (ug/L)**	**Band 3/4 Ratio**	**Model Prediction**	**(Actual-Predict)^2**
LM01	25	1.6803	22.6	5.7
LM03	21	1.6803	22.6	2.6
LM04	21	1.3522	21.8	0.7
LM05	23	1.7884	22.8	0.0
LM07	17	0.2798	16.9	0.0
		Standard Error of the	√Σ((Actual-Predict)^2/5) =	**4.0**
LM02	**25**	**1.6803**	**22.6**	
			Error of Estimate	**2.4**

Dataset minus LM01				
Station	**Chlorophyll-a Concentration (ug/L)**	**Band 3/4 Ratio**	**Model Prediction**	**(Actual-Predict)^2**
LM02	25	1.6803	22.6	5.7
LM03	21	1.6803	22.6	2.6
LM04	21	1.3522	21.8	0.7
LM05	23	1.7884	22.8	0.0
LM07	17	0.2798	16.9	0.0
		Standard Error of the	√Σ((Actual-Predict)^2/5) =	**4.0**
LM01	**25**	**1.6803**	**23.1**	
			Error of Estimate	**1.9**

Average Cross Validation Standard Error	**2.2**
Average Standard Error of the Estimate	4.6

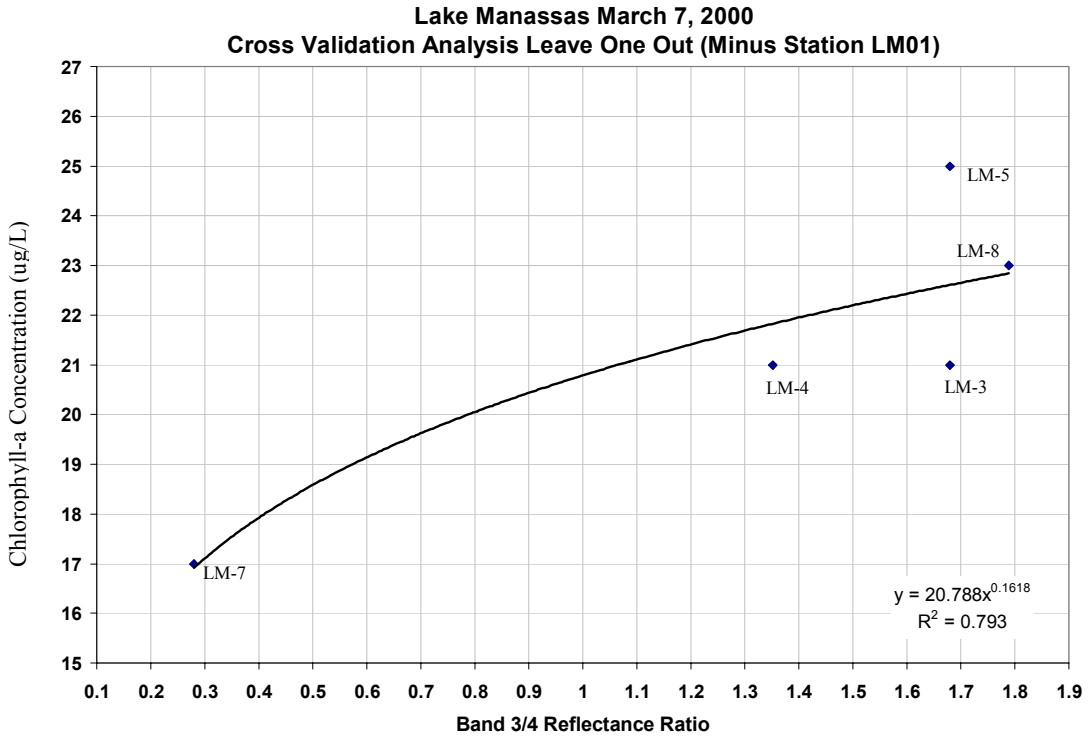


Figure 29 – 2000 Chlorophyll-a Concentrations vs. Band ¾ Reflectance Ratio with Cubic Convolution Filtering – Data minus LM01

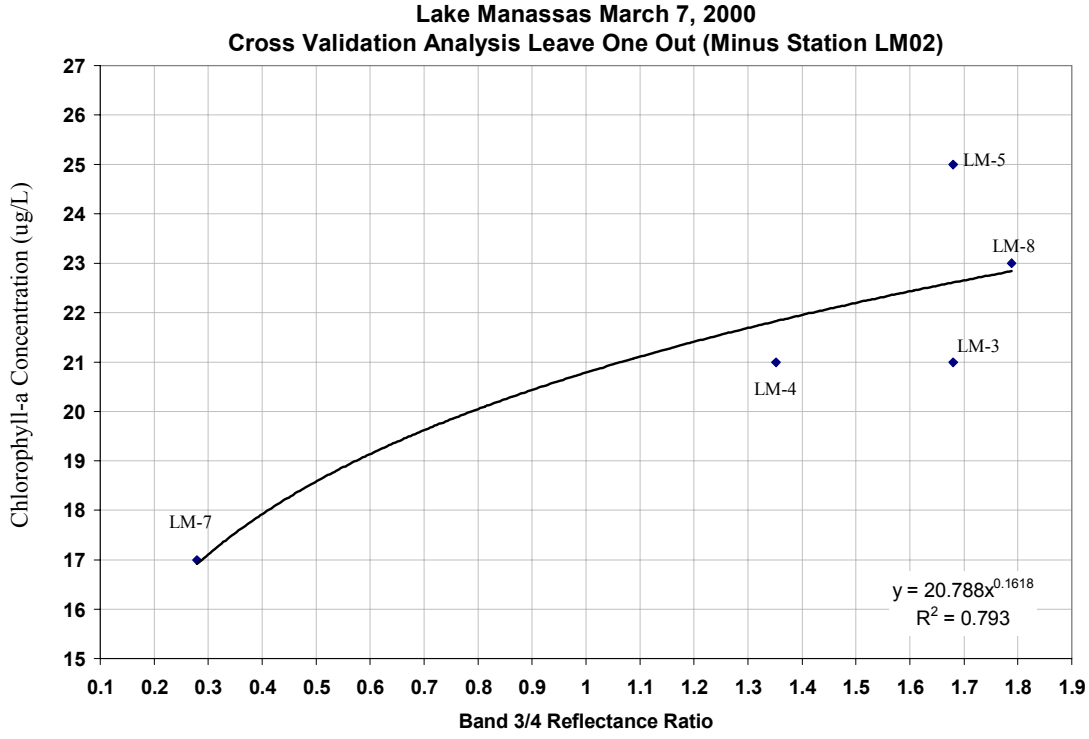


Figure 30 – 2000 Chlorophyll-a Concentrations vs. Band ¾ Reflectance Ratio with Cubic Convolution Filtering – Data minus LM02

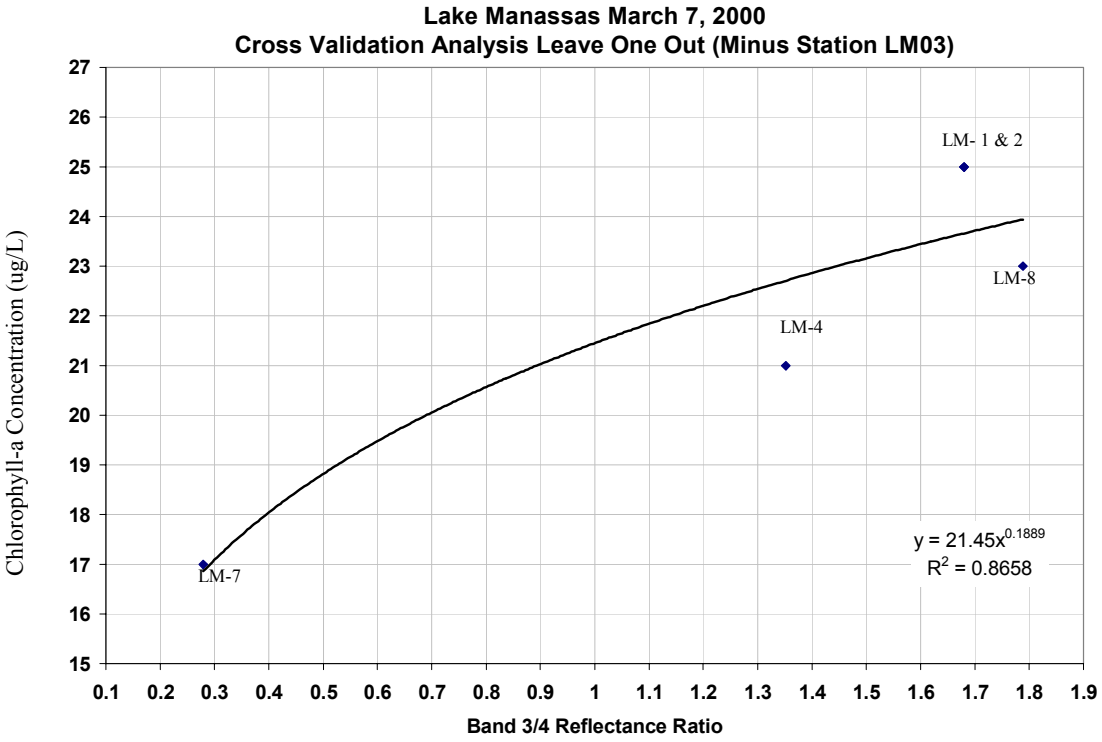


Figure 31 – 2000 Chlorophyll-a Concentrations vs. Band ¾ Reflectance Ratio with Cubic Convolution Filtering – Data minus LM03

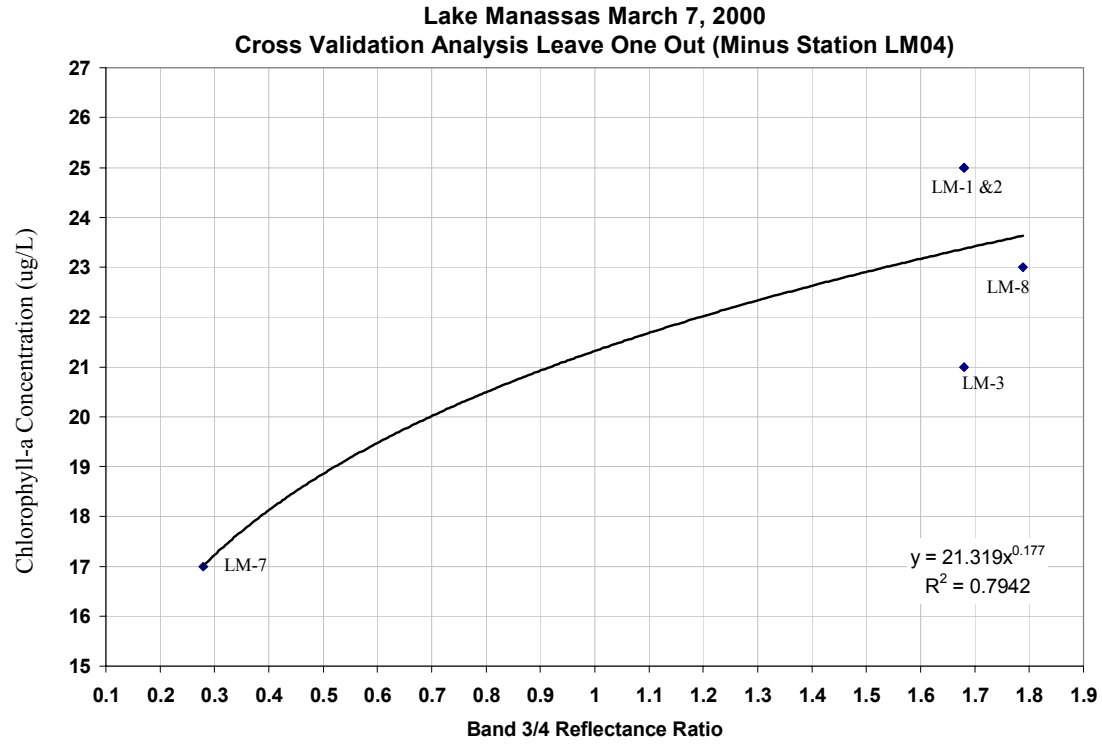


Figure 32 – 2000 Chlorophyll-a Concentrations vs. Band ¾ Reflectance Ratio with Cubic Convolution Filtering – Data minus LM04

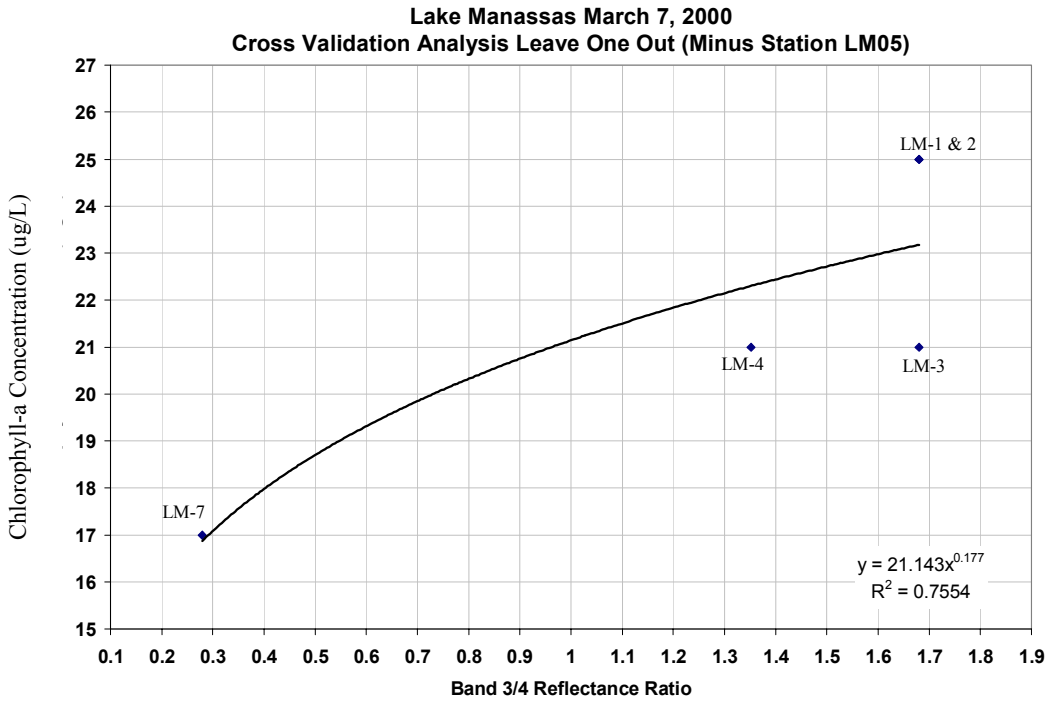


Figure 33 – 2000 Chlorophyll-a Concentrations vs. Band ¾ Reflectance Ratio with Cubic Convolution Filtering – Data minus LM05

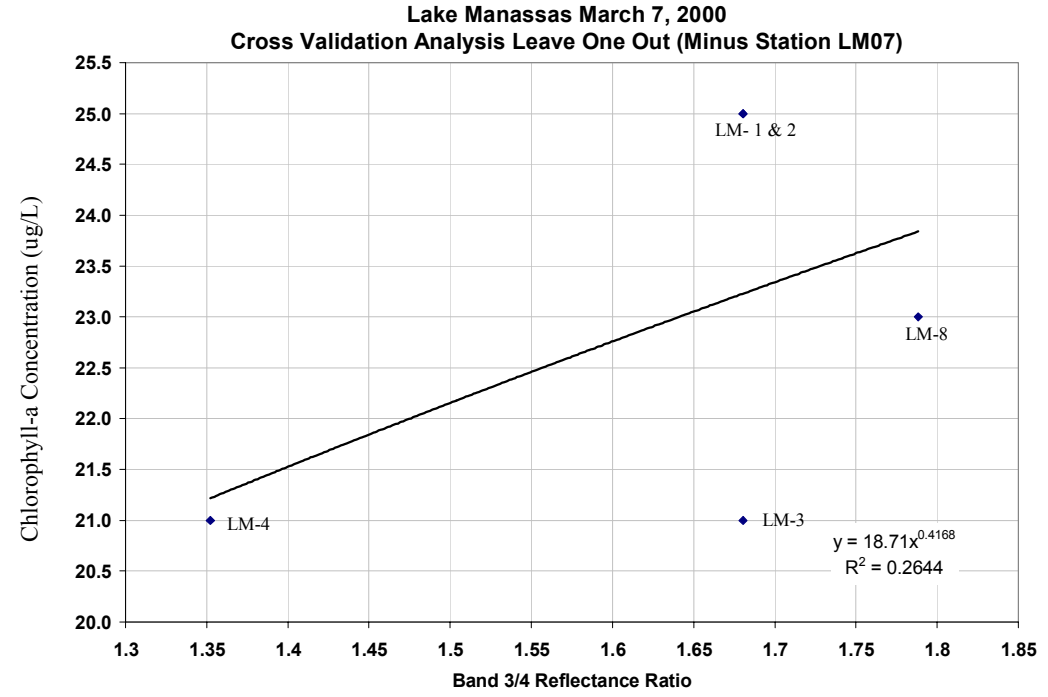


Figure 34 – 2000 Chlorophyll-a Concentrations vs. Band ¾ Reflectance Ratio with Cubic Convolution Filtering – Data minus LM07

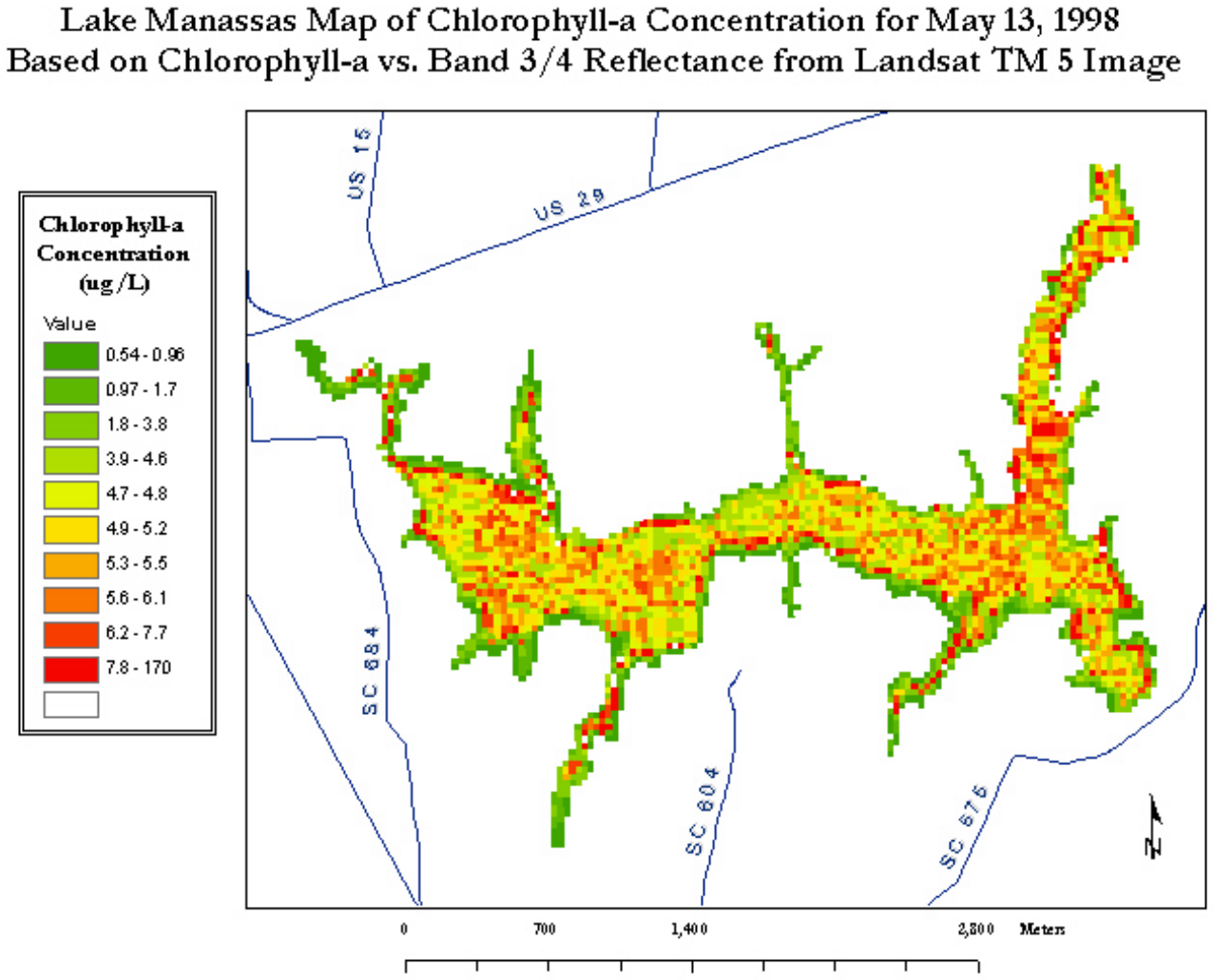


Figure 35 - Lake Manassas map of chlorophyll-a concentration for May 13, 1998 based on chlorophyll-a vs. band 3/4 reflectance from Landsat TM 5 image

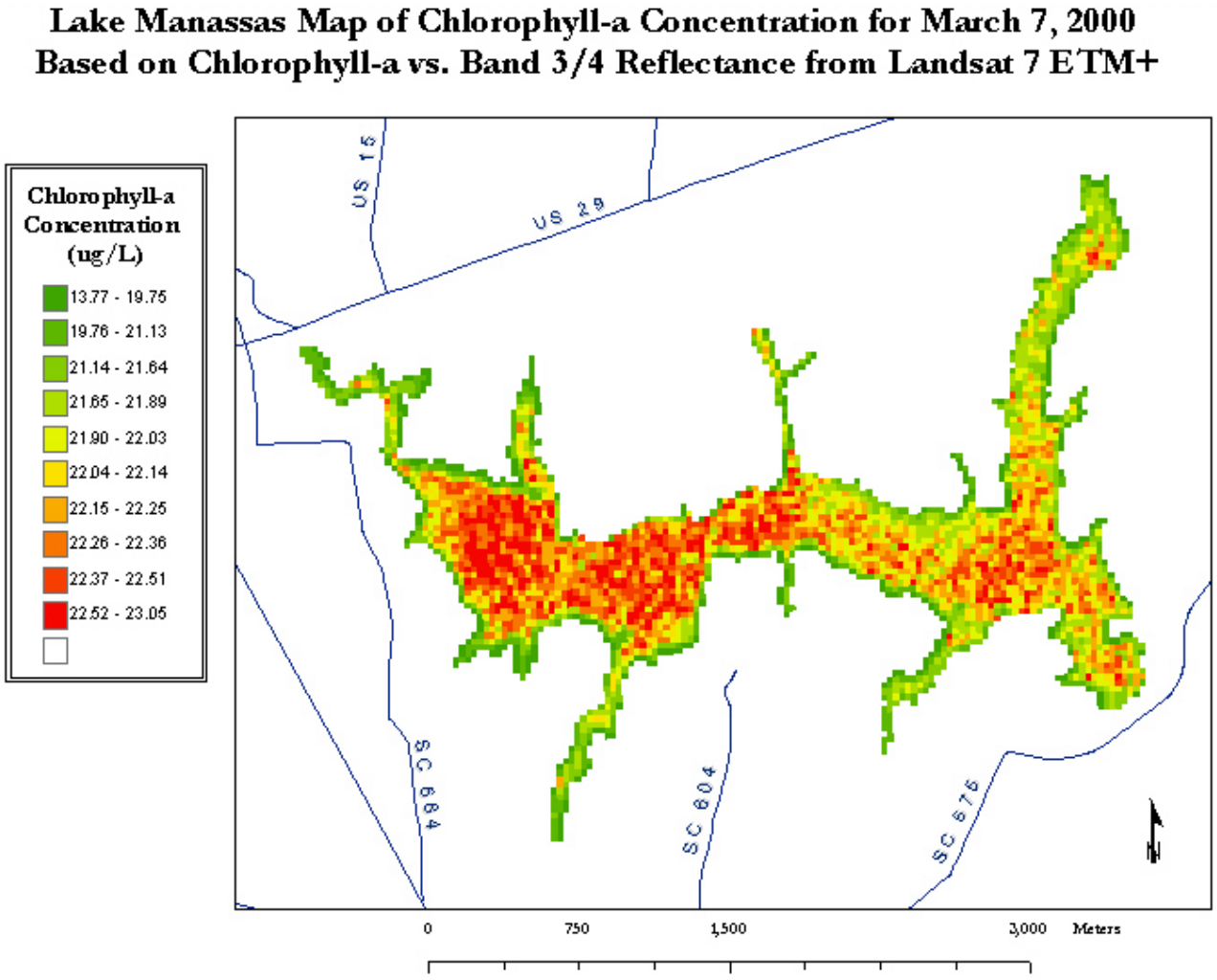


Figure 36 - Lake Manassas map of chlorophyll-a concentration for March 7, 2000 based on chlorophyll-a vs. band 3/4 reflectance from Landsat TM 7 ETM+ image

VI. Discussion

Though there are many parameters associated with the Lake Manassas water that affect the reflectance characteristics of Landsat TM band-four, this band proved to be useful in isolating the reflectance feature associated with the concentration of chlorophyll-a, though not without some preprocessing. Specifically, haze reduction was needed to reduce atmospherically induced noise, and cubic convolution filtering was needed to mitigate the uncertainty in location errors discovered in the associated field data.

Most likely the major source of error for this analysis was the 24-hour difference in the time of image acquisition (May 14, 1998 and March 8, 2000) and the time when the chlorophyll-a measurements were taken (May 13, 1998 and March 13, 2000). During this time, the algae in the reservoir may have easily been displaced by winds that mix the epilimnion. Data from the National Weather Service Dulles Airport station shows average wind speed ranging from 6 to 2.4 miles per hour with gusts ranging from 11 to 9 miles per hour for May 13 and 14 respectively with winds turning from 70 degrees WNW to 306 degrees N. For March 7 and 8, 2000 the average wind speeds ranged 6.4 to 5.1 miles per hour with gust of 20 miles per hour coming steadily out of the south. The velocity of wind driven surface water currents (depth < 1.0 cm) is approximately .02 times the wind speed. This relationship is linear up till a wind speed of about 6 meters/second or a little more than 13 miles per hour (Wetzel, 2001). Theoretically, a six mile an hour wind blowing over open waters could blow the surface water a distance of about three miles if it where unimpeded. This dispersion is greatly reduced by the confines of an enclosed reservoir. Much of the wind energy is used to pile the water on

the lee side of the fetch where it is forced downward by gravity. The downwelling water on the leeside causes water to up-well windward. There are other agents of water dispersion such as water density differentials caused by radiant energy, the coriolis effect, the water retention time of the reservoir, and so on. Though the true calculation of the dispersion rate is beyond the scope of this research, it suggests that the surface waters could have dispersed considerably over the 24-hour period between image acquisition and field data collection.

Another problematic aspect of the regression analysis for both the 1998 and 2000 data is the importance of certain stations to the measure of correlation in the model. For instance, in the cross validation analysis of the 1998 chlorophyll-a concentration model, all coefficients of determinations for the validation models fell between 0.7228 and 0.8369, except for the case were station LM08 was left out of the regression analysis, producing a coefficient of determination of 0.0957. Similarly, for the chlorophyll-a cross validation analysis of the 2000 model, all coefficients of determinations for the validation models fell between 0.7554 and 0.8658, except for the case were station LM07 was left out of the regression analysis, producing a coefficient of determination of 0.2644. The coefficient of determination is biased when there are data couples that fall far away from the main cluster of data points, as in the two cases noted above, and so the resulting correlation parameter must be referred to with some suspicion (Kleinbaum, 1988). Further, both of these water quality collection stations are located near shorelines where there is a possibility that the reflectance values are contaminated with reflectance from land along the shoreline or bottom reflectance of the near shore littoral zone. Obviously,

repeated trials of the methods shown are needed to support the validity of this type of analysis.

The standard errors of 1.5 ug/L and 4.6 ug/L associated respectively with the 1998 and 2000 data seemed to be slightly large. The standard error statistic is analogous to that of the standard deviation. As such, parallel lines constructed a vertical distance of one standard error away from the regression model would include, given a sufficiently large sample population, 68% of all sample points.

Though the standard error seems slightly large, the error of the estimate seemed to be quite low, meaning the model was able to predict the validation points with relatively fine accuracy. The average error of the estimate is 1.5 ug/L for the 1998 data and 2.2 ug/L for the year 2000 data.

Interestingly, the map of the chlorophyll-a distribution for March 7, 2000 shows a distinctive pattern that seems to suggest an algal bloom after the spring overturn. The temperature profile for both May 13, 1998 and March 7, 2000, suggest that the spring overturn was occurring close to the time of the March 8, 2000 image acquisition (See figures 37 and 38).

Appropriate future research goals would include the following:

1. Rerunning the analysis with better data: The main sources of error were in the unsure accuracy of locating the water quality ground stations. Also of concern was the time difference between Landsat image acquisition and water quality measurements. The effect of outliers on the model by data that was not evenly distributed across the data range is also problematic. An analysis on data that did not contain these sources of error would help determine the

efficacy of the procedure to remotely sense water born chlorophyll-concentrations.

2. Automate the process: If the above procedure were proven by further research, the next logical step would be to automate the procedure to make it more efficient and practical to use. Once the process becomes time tested and automated, the main goal of producing maps that show the changing dynamics of a reservoir would be more easily attained.

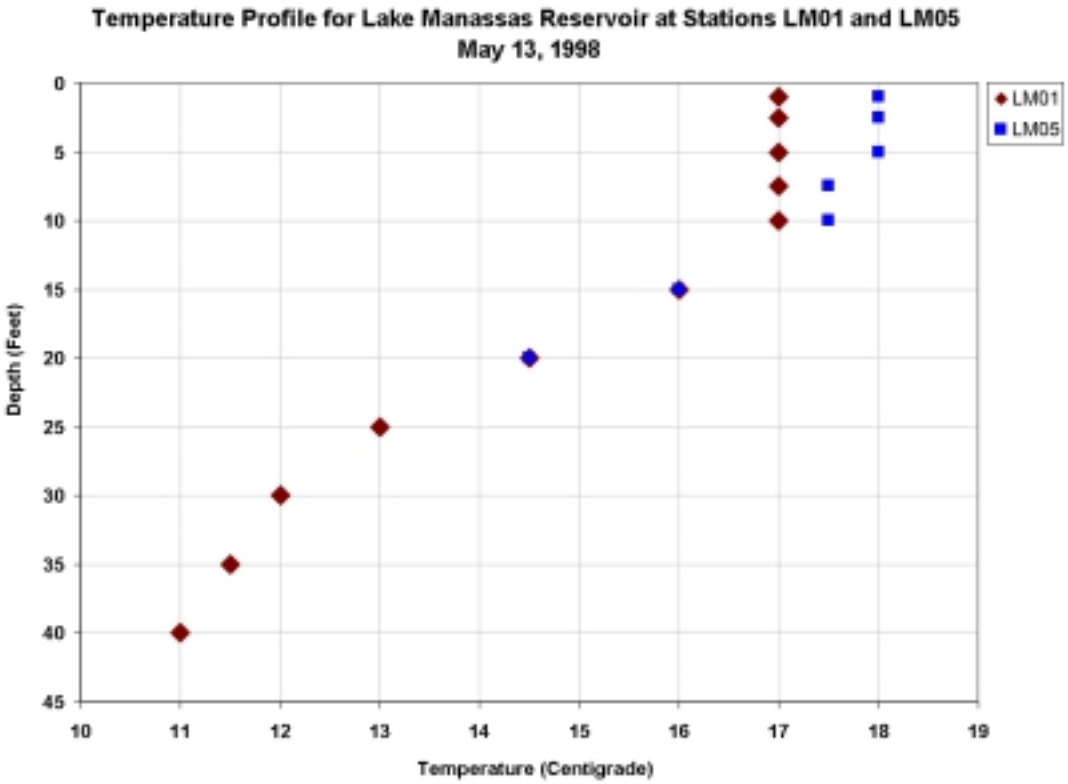


Figure 37 – Temperature profile of two pelagic zone stations, LM01 and LM05, for May 5, 1998

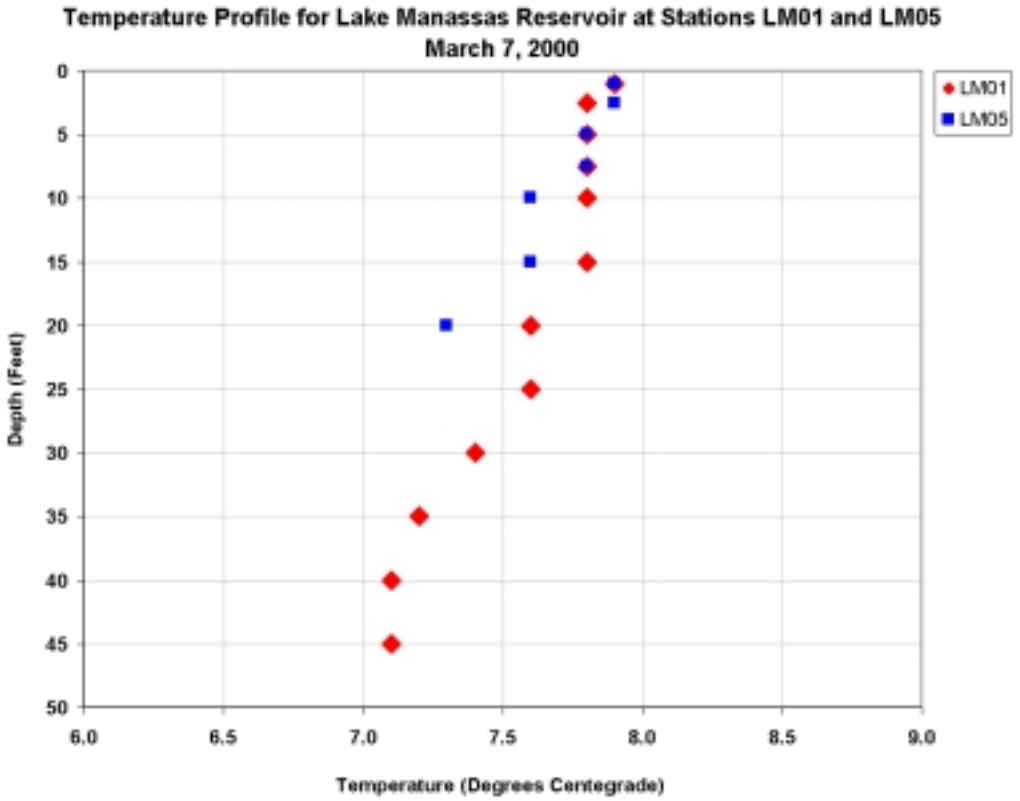


Figure 38 – Temperature profile of two pelagic zone stations, LM01 and LM05, for March 7, 2000

VII. Appendix

Summary of model results

Results for Lake Manassas May 7, 1998 Field [Chl-a] vs. May 8, 2000 Landsat Band ¾ Reflectance Ratio with Haze Reduction & Cubic Convolution Filtering			
Station	**Chlorophyll-a Concentration (ug/L)**	**Band 3/4 Ratio**	**Model Prediction**
LM01	7.2	1.6803	5.5
LM02	4.3	1.6803	5.5
LM04	5.7	1.3522	4.6
LM05	5.7	1.8778	6.1
LM06	4.9	1.5722	5.2
LM07	4.3	1.3493	4.6
LM08	2.4	0.6808	2.5

Results for Lake Manassas March 13, 2000 Field [Chl-a] vs. March 14, 2000 Landsat Band ¾ Reflectance Ratio with Haze Reduction & Cubic Convolution Filtering			
Station	**Chlorophyll-a Concentration (ug/L)**	**Band 3/4 Ratio**	**Model Prediction**
LM01	25	1.6803	23.1
LM02	25	1.6803	23.1
LM03	21	1.6803	23.1
LM04	21	1.3522	22.2
LM05	23	1.7884	23.3
LM07	17	0.2798	16.9

At Satellite Reflectance Calculations

The conversion of the 1998 Landsat 5 TM digital numbers to reflectance was done using ENVI 3.5 software; however, the function to convert the Landsat 7 ETM+ was not available at the time and was instead performed using the modeling capability of ERDAS 8.5 software. Sensor gain information in the Landsat TM leader of the year 2000 image was used to calculate radiance values in units of mW cm-2 Sr^{-1} um^{-1}. The result

was then combined in the reflectance conversion formula for the final output. The procedure is given below.

At Satellite Radiance

The radiance conversion removes bias and gain offsets introduced by the Landsat sensor. The formulas and procedures are given in a USGS paper by Chengquan Huang (Huang, 2001).

$$L_\lambda = Gain_\lambda * DN + Bias_\lambda$$

L_λ = At-satellite radiance (W m^{-2} sr^{-1})

Gain λ = Sensor gain (different for each band measured as a voltage)

DN = Digital number (unitless)

Bias λ = Sensor bias (different for each band w/(m^2 sr um))

At Satellite Reflectance

The satellite reflectance removes the effect of the viewing geometry of the satellite and atmospheric interference caused by the angle of the sun's incoming radiation. Notice that the formula includes L_λ (At-satellite radiance).

$$\rho_\lambda = \frac{\pi * L_\lambda * d^2}{ESUN_\lambda * \sin\theta}$$

ρ_λ = At-satellite reflectance (the reflectance value without bias or gain offset and sun at zenith)

L_λ = At-satellite radiance (W m^{-2} sr^{-1})

$ESUN_\lambda$ = Solar irradiance (W m^{-2} sr^{-1})

θ = Sun elevation angle given in Landsat TM header file (degrees).

d = SQRT ($1/E_0$) = the normalized Sun/Earth distance (unitless) (Iqbal, 1983)

The at-satellite reflectance with the imbedded radiance conversion formula is as follows:

$$\rho_\lambda = \frac{\pi(Gain_\lambda * DN + Bias_\lambda) * d^2}{ESUN_\lambda * \sin\theta}$$

The values for the solar energy available to each of the Landsat bands (solar spectral irradiances or $ESUN_\lambda$), as well as the Earth-sun distance table, where taken directly from The Landsat 7 Science Data Users Handbook, Chapter 11 (Landsat 7, 2002). Values are shown in tables 9 and 10 below.

Table 9 - Landsat 7 ETM+ Solar Spectral Irradiances (Landsat 7 Science Data Users Handbook)

Band	**Watts/(meter squared*µ)**
1	1969.000
2	1840.000
3	1551.000
4	1040.000
5	225.700
`7	82.07
8	1368.000

Table 10 - Earth-Sun Distance in Astronomical Units (Iqbal, 1983)

Julian Day	Distance	Julian Day	Distance	Julian Day	Distance	Julian Day	Distance	Julian Day	Distance
1	0.9832	74	0.9945	152	1.0140	227	1.0128	305	0.9925
15	0.9836	91	0.9993	166	1.0158	242	1.0092	319	0.9892
32	0.9853	106	1.0033	182	1.0167	258	1.0057	335	0.9860
46	0.9878	121	1.0076	196	1.0165	274	1.0011	349	0.9843
60	0.9909	135	1.0109	213	1.0149	288	0.9972	365	0.9833

Acquisition date = March 8, 2000; Julian day = 68 (of 2000)
d (Earth-Sun distance in astronomical units) = SQRT($1/E_0$)

And, as such,

For March 8: $E_0 = 1.0153 = (r_0/r)^2$; $d = \sqrt{(1/1.0153)} = 0.9924\ 36679288$ (Iqbal, 1983)

The gain and bias values for the May 8, 2000 data, taken from the metadata, is given below:

Gain λ

first_band_gain = 0.78627452
second_band_gain = 0.81725488
third_band_gain = 0.63960787
fourth_band_gain = 0.63529412
fifth_band_gain = 0.12847059
sixth_band_gain = 0.044243138

Bias λ **=**

first_band_bias = -6.1999998
second_band_bias = - 6
third_band_bias = - 4.5
fourth_band_bias = - 4.5
fifth_band_bias = - 1
sixth_band_bias = - 0.34999999

The solar spectral irradiances, $ESUN_\lambda$, is given by:

$\theta = 40.826799999999999$ (Form Landsat TM header file)
d = SQRT (1/1.0153) = 0.9924 36679288 or d = 0.9924 36679288

The above formulas and associated values where programmed into a reflectance model using ERDAS Imagines Model Maker (see figure 39). For Band 1, the model is as follows:

$$\rho_\lambda = ((\pi *(Gain*DN + (Bias))*d^2))/(ESUN\_band\_x*SIN (\theta_x))$$

The equations[2] that were put into the modules of the ERDAS modeler are as follows:

Reflectance Model, Band 1:

$$\rho_\lambda = (3.14159265359*(0.78627452*\$n1\_occ2000(1) + (-6.1999998))* 0.992436679288 **2)/(1969*SIN (40.826799999999999))$$

Reflectance Model, Band 2:

$$\rho_\lambda = (3.14159265359*(0.81725488*\$n1\_occ2000(2) + (- 6))* 0.992436679288 **2)/(1840*SIN (40.826799999999999))$$

Reflectance Model, Band 3:

$$\rho_\lambda = (3.14159265359*(0.63960787*\$n1\_occ2000(3) + (- 4.5))* 0.992436679288 **2)/(1551*SIN (40.826799999999999))$$

Reflectance Model, Band 4:

$$\rho_\lambda = (3.14159265359*(0.63529412*\$n1\_occ2000(4) + (- 4.5))* 0.992436679288 **2) /(1044*SIN (40.826799999999999))$$

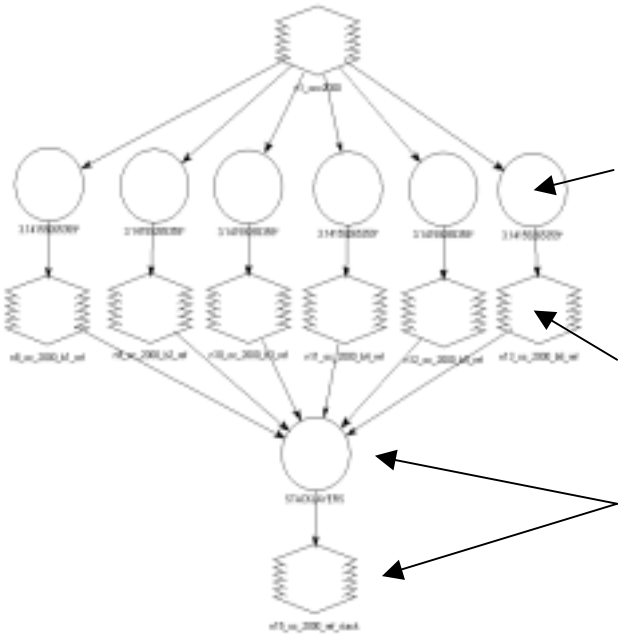
Reflectance Model, Band 5:

$$\rho_\lambda = (3.14159265359*(0.12847059*\$n1\_occ2000(5) + (- 1))* 0.992436679288 **2)/(225.7*SIN (40.826799999999999))$$

Reflectance Model, Band 6:

$$\rho_\lambda = (3.14159265359*(0.044243138*\$n1\_occ2000(6) + (- 0.34999999))* 0.992436679288 **2)/(82.07*SIN (40.826799999999999))$$

[2] *$n1_occ2000(x) refers to the Landsat ETM+ pixel (digital number) values for band x of the March 8, 2000 image.



The model takes each individual digital number from each band of the raw image and passes it through the appropriate relation as given above.

The output digital numbers are placed back in the original x-y positions creating a new file for each band.

All 6 bands, converted to reflectance, are stacked back together in one output file.

Figure 39 – Schematic of haze reduction model built in ERDAS Imagine.

Image Metadata

Metadata for May 14, 1998 Landsat 5 TM Image

```
struct Endf_Header {
   float ndf_revision = 0
   char *pixel_format[5] = "byte"
   char *pixel_order[13] = "not_inverted"
   enum data_set_type = ENDF_EDC_TM
   struct upper_left_corner {
      double lon = -775339.11
      double lat = 394718.36
      double easting = 252172.11
      double northing = 4408283.3
   } UpperLeftCorner
   struct upper_right_corner {
      double lon = -753242.69
      double lat = 392646.3
      double easting = 453088.65
      double northing = 4366435
   } UpperRightCorner
   struct lower_left_corner {
      double lon = -783322.01
      double lat = 364529.32
      double easting = 182533.39
```

```
   double northing = 4073943.2
} LowerLeftCorner
struct lower_right_corner {
   double lon = -761800.64
   double lat = 362537.16
   double easting = 383449.93
   double northing = 4032094.9
} LowerRightCorner
struct reference_position {
   double lon = -770441.3
   double lat = 380640.2
   double easting = 317811.02
   double northing = 4220189.1
   double pixelno = 3601.5
   double lineno = 5992.5
} ReferencePosition
char *product_number_string[24] = "1199062400510001.000000"
long bits_per_pixel = 8
long pixels_per_line = 7202
long lines_per_data_file = 11984
char *data_position[11] = "upper_left"
char *data_direction[6] = "right"
char *data_file_interleaving[4] = "bsq"
char *reference_point[13] = "scene_center"
char *map_projection_name[4] = "utm"
char *horizontal_datum[6] = "wgs84"
char *satellite[10] = "landsat_5"
char *satellite_instrument[3] = "tm"
char *product_size[12] = "multi_scene"
char *pixel_spacing_units[7] = "meters"
char *resampling[3] = "cc"
char *processing_software[13] = "nlaps_3_5_4e"
char *band1_name[10] = "tm_band_1"
char *band2_name[10] = "tm_band_2"
char *band3_name[10] = "tm_band_3"
char *band4_name[10] = "tm_band_4"
char *band5_name[10] = "tm_band_5"
char *band6_name[10] = "tm_band_6"
char *band7_name[10] = "tm_band_7"
double band1_wavelength_start = 0.45
double band2_wavelength_start = 0.52
double band3_wavelength_start = 0.63
double band4_wavelength_start = 0.76
double band5_wavelength_start = 1.55
double band6_wavelength_start = 10.4
double band7_wavelength_start = 2.08
```

double band1_wavelength_end = 0.52
double band2_wavelength_end = 0.6
double band3_wavelength_end = 0.69
double band4_wavelength_end = 0.9
double band5_wavelength_end = 1.75
double band6_wavelength_end = 12.5
double band7_wavelength_end = 2.35
double band1_gain = 0.6024314
double band2_gain = 1.1750981
double band3_gain = 0.8057647
double band4_gain = 0.814549
double band5_gain = 0.1080784
double band6_gain = 0.0551582
double band7_gain = 0.0569804
double band1_bias = -1.52
double band2_bias = -2.8399999
double band3_bias = -1.17
double band4_bias = -1.51
double band5_bias = -0.37
double band6_bias = 1.2377996
double band7_bias = -0.15
double sun_elevation = 61.31
double sun_azimuth = 126.04
long processing_level = 8
long usgs_projection_number = 1
long usgs_map_zone = 18
long number_of_data_files = 7
long start_line_number = 1
long start_data_file = 1
long lines_per_volume = 83888
long blocking_factor = 1
long record_size = 7202
long tape_number = 1
long tape_total = 1
double reference_offset_xpos = -333.69
double reference_offset_ypos = 2547.52
double orientation = 11.765749
double ee_semi_major_axis = 6378137
double ee_semi_minor_axis = 6356752.3
double ee_origin_offset_xpos = 0
double ee_origin_offset_ypos = 0
double ee_origin_offset_zpos = 0
double ee_rotation_offset_xpos = 0
double ee_rotation_offset_ypos = 0
double ee_rotation_offset_zpos = 0
double pixel_spacing_horizontal = 28.5

```
    double pixel_spacing_vertical = 28.5
    long wrs_path = 15
    double wrs_row = 34
    calendar_time acquisition_date_time = Thu May 14 15:24:05 1998
    char *acquisition_date_time_raw[16] = "051498/15240570"
    calendar_time process_date_time = Sun Jun 27 20:23:54 1999
    char *process_date_time_raw[16] = "062799/20235400"
    long no_bands_in_volume = 7
    double usgs_projection_parameters[15] = 6378137, 6356752.3, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0
} Endf_Header
```

Metadata for the March 8, 2000 Landsat 7 ETM+ Image

```
Band
struct Eimg_Layer {
   long width = 7202
   long height = 11984
   enum layerType = athematic
   enum pixelType = u8
   long blockWidth = 64
   long blockHeight = 64
} Eimg_Layer
struct L7A_Header {
   char product_id[19] = "0750006130162_0011"
   char first_scene_location[12] = "015/0330000"
   char first_scene_acquisition_date[9] = "2000-03-08"
   char first_scene_satellite[9] = "LANDSAT7"
   char first_scene_sensor[5] = "ETM+"
   char first_scene_mode[7] = "NORMAL"
   double first_scene_lookangle = 0
   char second_scene_location = ""
   char second_scene_acquisition_date = ""
   char second_scene_satellite = ""
   char second_scene_sensor = ""
   char second_scene_mode = ""
   double second_scene_lookangle = 0
   char third_scene_location = ""
   char third_scene_acquisition_date = ""
   char third_scene_satellite = ""
   char third_scene_sensor = ""
   char third_scene_mode = ""
   double third_scene_lookangle = 0
   char fourth_scene_location = ""
   char fourth_scene_acquisition_date = ""
   char fourth_scene_satellite = ""
   char fourth_scene_sensor = ""
   char fourth_scene_mode = ""
   double fourth_scene_lookangle = 0
   char product_type[13] = "MAP ORIENTED"
   char product_size[11] = "FULL SCENE"
   char type_of_processing[11] = "SYSTEMATIC"
   char resampling[3] = "CC"
   long tape_volume_number = 1
   long number_of_volumes = 1
   long pixels_per_line = 7851
   long lines_in_volume = 7041
```

long lines_in_outputimage = 7041
long start_line_number = 0
long blocking_factor = 0
long record_length = 55278891
double pixel_size = 30
long output_bits_per_pixel = 8
long acquired_bits_per_pixel = 8
char bands_present[7] = "123457"
char first_band_name[30] = "L71015033_03320000308_B10.FST"
char second_band_name[30] = "L71015033_03320000308_B20.FST"
char third_band_name[30] = "L71015033_03320000308_B30.FST"
char fourth_band_name[30] = "L71015033_03320000308_B40.FST"
char fifth_band_name[30] = "L71015033_03320000308_B50.FST"
char sixth_band_name[30] = "L72015033_03320000308_B70.FST"
char revision[4] = "L7A"
double first_band_bias = -6.1999998
double first_band_gain = 0.78627452
double second_band_bias = -6
double second_band_gain = 0.81725488
double third_band_bias = -4.5
double third_band_gain = 0.63960787
double fourth_band_bias = -4.5
double fourth_band_gain = 0.63529412
double fifth_band_bias = -1
double fifth_band_gain = 0.12847059
double sixth_band_bias = -0.34999999
double sixth_band_gain = 0.044243138
double seventh_band_bias = 0
double seventh_band_gain = 0
double eigth_band_bias = 0
double eigth_band_gain = 0
char map_projection_name[4] = "LCC"
char earth_ellipsoid[6] = "WGS84"
char datum_name[6] = "WGS84"
double USGS_projection_params1 = 6378137
double USGS_projection_params2 = 6356752.3
double USGS_projection_params3 = 37000000
double USGS_projection_params4 = 39030000
double USGS_projection_params5 = -79030000
double USGS_projection_params6 = 36000000
double USGS_projection_params7 = 0
double USGS_projection_params8 = 0
double USGS_projection_params9 = 0
double USGS_projection_params10 = 0
double USGS_projection_params11 = 0
double USGS_projection_params12 = 39030000

```
    double USGS_projection_params13 = 0
    double USGS_projection_params14 = 0
    double USGS_projection_params15 = 0
    long zone_number = 0
    double upperleft_lon = 78.185207
    double upperleft_lat = 39.871468
    double upperleft_mapx = 112500
    double upperleft_mapy = 430500
    double upperright_lon = 75.43525
    double upperright_lat = 39.80983
    double upperright_mapx = 348000
    double upperright_mapy = 430500
    double lowerright_lon = 75.540728
    double lowerright_lat = 37.908877
    double lowerright_mapx = 348000
    double lowerright_mapy = 219300
    double lowerleft_lon = 78.219363
    double lowerleft_lat = 37.968957
    double lowerleft_mapx = 112500
    double lowerleft_mapy = 219300
    double scenecenter_lon = 76.845025
    double scenecenter_lat = 38.89773
    double scenecenter_mapx = 230250
    double scenecenter_mapy = 324900
    long scenecenter_pixelno = 3926
    long scenecenter_lineno = 3521
    long horizontal_offset = 0
    double orientation_angle = 0
    double sun_elevation_angle = 40.826799999999999
    double sun_azimuth_angle = 146.45500000000001
} L7A_Header
```

VIII. References

1. An Updated Water Quality Assessment for the Occoquan Reservoir and Tributary Watershed 1973-1997 Prepared by the Occoquan Watershed Monitoring Laboratory, The Charles E Via, Jr. Department of Civil and Environmental Engineering, Manassas VA 20110. July 1998

2. Advances in Remote Sensing and GIS Analysis, Peter M. Atkinson and Nicholas J. Tate Editors. John Wiley & Sons, NY, New York, 1999 pp 241-261

3. Baker, E. T., & Lavelle, J. W. The Effect of Particle Size in the Light Attenuation Coefficient of Natural Suspensions. Journal of Geophysical Research, 89, 1984, pp 8197– 8203.

4. Bale, A. J., Tocher, M. D., Weaver, R., Hudson, S. J., & Aiken, J. Laboratory Measurements of The Spectral Properties Of Estuarine Suspended Particles. Netherlands Journal of Aquatic Ecology, 28, 1994, pp 237– 244.

5. Brezonic, Patrick L., Erin Day, S. Kloiber and Leif Olmanson, Lake Water Clarity of Minnesota's 10,000 Lakes: A Comprehensive Review from Space. Virginia Water Research Symposium 2000, The Charles E Via, Jr. Department of Civil and Environmental Engineering, Manassas VA. VWRRC Special Report SR-10-2000 pp 191-214

6. Chavez, P. S. (1988), "An Improved Dark-Object Subtraction Technique for Atmospheric Scattering Correction of Multispectral Data," Remote Sensing of the Environment, 24, pp 459-479.

7. Cole, G. A., Textbook of Limnology, 4th ed., Waveland Press, Inc., Prospect Heights, IL, 1994.

8. Cooke, G. D. and R. E. Carlson, Reservoir Management for Water Quality and THM Precursor Control, AWWA Research Foundation, Denver, CO, 1989.

9. Eggink, Judith, 2001, "An Exploration of the Limnological Dynamics of Lake Manassas", M.S. Thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA.

10. Yang, Ming-Dir, Yeah-Fen Yang, Eutrophic Status Assessment Using Remote Sensing Data, Department of Construction Engineering, Chaoyang University of Technology, Taichung County, Taiwan, 1999. Printed in SPIE Vol. 3868 1999

11. George, D.G. (2000) Remote Sensing Evidence For Episodic Transport On Phosphorous From The Littoral Zone Of A Thermally Stratified Lake Freshwater Biology, vol. 43, pp. 571-578

12. Grizzard, Tom. Term Project: Background Information for Occoquan Reservoir Analysis, CEE 5104 Engineering Aspects of Water Quality Class, Fall Semester, 2001, Virginia Polytechnic Institute and State University, Blacksburg, VA.

13. Grizzard, Tom, CEE 5104 Engineering Aspects of Water Quality, Class Notes, Fall Semester, 2001, Virginia Polytechnic Institute and State University, Blacksburg, VA.

14. Han, Luoheng, Spectral Reflectance with Varying Suspended Sediment Concentrations in Clear and Algal-Laden Waters, Photogrammetric Engineering and Remote Sensing, Vol. 63, No.6, June 1997, pp 701-705

15. Han, Luoheng, Rundquist, Donald C., Comparison of NIR/RED Ratio and First Derivative on Reflectance in Estimating Algal-Chlorophyll-a Concentration: A Case Study in a Turbid Reservoir, Remote Sensing Environment, Vol. 62, 1997, pp 253-261

16. Huang, Chengquan, et.al., USGS Raytheon ITSS, EROS Data Center Sioux Falls, SD 57198, USA: 2001

17. Iqbal, Muhammad, An Introduction To Solar Radiation, Academic Press, London, pp .5, 1983

18. Jensen, John R., Remote Sensing of the Environment, Prentice Hall Publications, Upper Saddle River, NJ, 2000, pp 379 - 406

19. Kiefer, Ralf W, and Lillesand, Thomas M., Remote Sensing and Image Interpretation, Fourth Edition, John Wiley & Sons Inc. 1999 pp 17-19.

20. Kleinbaum, Kupper and Muller, *Applied Regression Analysis and Other Multivariable Methods,* 2nd Edition, 1988, pp 328-33 1.

21. Landsat 7 Science Data Users Handbook, 2002 http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_toc.html

22. Lathrop R.G. Jr., Vande Castle JR ., Lillesand TM. Monitoring river plume transport and mesoscale circulation in Green Bay, Lake Michigan, through satellite remote sensing; Journal of Great Lakes Research, vol.16, no. 43, 1990, pp.471-8. USA.

23. Olmanson, Lief G., Patrick L. Bresonick, Steve M. Kloiber, Marvin E Bauer, Erin E. Day, Lake Water Clarity Assessment of Minnesota's 10,000 Lakes: A Comprehensive View from Space, Water Resources Center, University of Minnesota, 1999

24. Lodhi, Mahtab A., Rundquist, Donald C., Han, Luoheng, Kuzila, Mark S., Estimation of Suspended Sediment Concentration in Water Using Integrated Surface Reflectance; Geocarto International, Vol. 13, No. 2, June 1998 Geocarto International Centre, G.P.O. Box 4122, Hong Kong

25. McCluney, W.R. 1975, Remote Measurement of "turbidity" and other Water Quality Parameters. NASA/Goddard Space Flight Center, Greenbelt, MD

26. Mikkelsen, Ole Aarup, Variation in the Projected Surface Area of Suspended Particles: Implications for Remote Sensing Assessment of TSM, Remote Sensing and the Environment, Vol. 79, 2002, pp. 23-29

27. Mikkelsen, O. A., & Pejrup, M., In Situ Particle Size Spectra and Density of Particle Aggregates in a Dredging Plume. Marine Geology, 170, 2000, pp 443–459.

28. Mittenzewy, K. H., Gitelson, A. A., and Kondratyev, K. Y., Determination of Chlorophyll-a of Inland Waters on the Basis Of Spectral Reflectance Limnological Engineering, 37, 1992, pp 147-149

29. National Weather Service historical weather data for Dulles Airport http://www.erh.noaa.gov/er/box/dailystns.html

30. Nachtsheim, Neter, et.al., *Applied Linear Statistical Models, 4th* Edition, McGraw-Hill, NY, 1996, pp 436-439

31. National Institute of Standards and Technology/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, 2002.

32. Olmanson, Lief G., Patrick L. Bresonick, Steve M. Kloiber, Marvin E Bauer, Erin E. Day, Lake Water Clarity Assessment of Minnesota's 10,000 Lakes: A Comprehensive View from Space, Water Resources Center, University of Minnesota, 2000

33. Östlund, Catherine, Peter Flink, , Niklas Strömbeck, Don Pierson, Tommy Lindell, Mapping Of The Water Quality Of Lake Erken, Sweden, From Imaging Spectrometry And Landsat Thematic Mapper, The Science of the Total Environment, 2001 vol. 268, pp 139-154,

34. OWML - Occoquan Watershed Monitoring Lab, From field trip with the OWML field data collection team (Phil Spellerberg, Doug Holladay, Mark Lucas, and George Underwood), August, 2002

35. Ritchie, Jerry C, McHenry, J. Roger, Remote Sensing in Research, US Sedimentation Laboratory, Agricultural Research Service, United States Department of Agriculture, Oxford Mississippi, 38655

36. Schalles JF; Gitelson AA; Yacobi YZ; Kroenke AE Estimation of chlorophyll a from time series measurements of high spectral resolution reflectance in an eutrophic lake JOURNAL OF PHYCOLOGY 1998, Vol. 34, Iss 2, pp 383-390

37. Schiebe, F.R. and J. C. Ritchie, Color Measurements and Suspended Sediments in North Mississippi Reservoirs, Remote Sensing of Earth Resources, Vol. IV, The University of Tennessee Space Institute, Tullahoma, Tennessee

38. Welby, C.W. 1975, Use of ERTS Imagery in the Study of Cape Fear Plumes, Southeastern Geology 16, 1975, pp 189-203

39. Wetzel, Robert G., Limnology and River Ecosystems, Third Edition, Academic Press, NY, 2001

40. Wynne, Dr. Randolph, Class notes from *Remote Sensing of Natural Resources,* Forestry 5204, Spring, 2002, Virginia Polytechnic Institute and State University, Blacksburg, VA.

41. Yang, Min-Der, et.al., Adaptive Short Term Water Quality Forecasts Using Remote Sensing and GIS. Virginia Water Resources Research Center Symposium Proceedings, September 22 - 26, 1996, Fort Lauderdale Florida , pp 109 – 118

IX. Vita

Besides graduating from Virginia Polytechnic Institute in 2003 with a Masters in Civil and Environmental Engineering, Paul Bartholomew obtained a BA in General Science from Rutgers University in 1989, a Masters in Teaching Secondary Mathematics from The American University in 1993, and has finished 21 masters degree credits in Geological Engineering from Radford University. He received monetary awards to complete the research for this thesis from the Wastewater Policy Institute. He also received research funding from the American Water Resources Association for work on the design and implementation of a remote data logging system for the Stroubles Creek Watershed in Blacksburg VA. Currently, Paul is teaching mathematics and science in New Jersey where he and his wife Amy are expecting their first child.

Paul Bartholomew

May 21, 2003

Blacksburg, Virginia