# EVALUATING LAKE PHYTOPLANTON RESPONSE TO HUMAN DISTURBANCE AND CLIMATE CHANGE USING SATELLITE IMAGERY

By

Linda Nicole Novitski

### A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Zoology – Doctor of Philosophy

### ABSTRACT

### EVALUATING LAKE PHYTOPLANTON RESPONSE TO HUMAN DISTURBANCE AND CLIMATE CHANGE USING SATELLITE IMAGERY

By

### Linda Nicole Novitski

Accurate and cost-effective assessment of water quality is necessary for proper management and restoration of inland water bodies susceptible to algal bloom conditions. Landsat and MODIS satellite images were used to create chlorophyll and Secchi depth predictive models for algal assessment of Great Lakes and other lakes of the United States. Boosted regression tree (BRT) models using satellite imagery are both easy to use and can have high predictive performance. BRT models inferred chlorophyll and Secchi depth more accurately than linear regression models for all study locations. Inferred chlorophyll of inner Saginaw Bay was subsequently used in ecological models to help understand the ecological drivers of algal blooms in this ecosystem.

For small lakes (non-Great Lakes), the best national Landsat model for lntransformed chlorophyll was the BRT model and had a cross-validation R<sup>2</sup> of 0.44 and a 0.76 ln-transformed  $\mu$ g/L RMSE. The best national Landsat model for Secchi depth was also a BRT model that had an adjusted R<sup>2</sup> of 0.52 and a 0.80 m RMSE. We assessed the applicability of the national chlorophyll model for ecological analysis by comparing the total phosphorus- chlorophyll relationship with chlorophyll determined from sampling or remote sensing, which showed the total phosphorus-chlorophyll relationship had an adjusted R<sup>2</sup> = 0.58 and 1.02 ln-transformed  $\mu$ g/L RMSE with sampled chlorophyll versus an adjusted R<sup>2</sup> = 0.56 and 1.04 ln-transformed  $\mu$ g/L RMSE with chlorophyll determined by the boosted regression tree remote sensing model. For Great Lakes models, the MODIS BRT model predicted chlorophyll most accurately of the three BRT models and compared well to other models in the literature. BRT models for Landsat ETM+ and TM more accurately predicted chlorophyll than the MSS model and all Landsat models had favorable results when compared to the literature. BRT chlorophyll predictive models are useful in helping to understand historical, long-term chlorophyll trends and to inform us of how climate change may alter ecosystems in the future. In inner Saginaw Bay, annual average and upper guartile Landsat-derived chlorophyll decreased from 7.44 to 6.62 and 8.38 to 7.38 µg/L between 1973-1982, and annual upper quartile of 8-day phosphorus loads increased from 5.29 to 6.79 kg between 1973-2012. Simple linear and multiple regression models and Wilcoxon rank test results for MODIS and Landsat-derived chlorophyll indicate that distance from the Saginaw River mouth influences chlorophyll concentration in Saginaw Bay; Landsat-derived surface water temperature and phosphorus loads to a lesser extent. Mixed-effect models for MODIS and Landsat-derived chlorophyll were related to chlorophyll better than simple linear or multiple regressions, with random effects of pixel and sample date contributing substantially to predictive power (NSE=0.35-70), though phosphorus loads, distance to Saginaw River mouth, and water were significant fixed effects in most models. Water quality changes in Saginaw Bay between 1972-2012 were influenced by phosphorus loading and distance to the Saginaw River's mouth.

Landsat and MODIS imagery are complementary platforms because of the long history of Landsat operation and the finer spectral resolution and image frequency of MODIS. Remote sensing water quality assessment tools can be valuable for limnological study, ecological assessment, and water resource management. Copyright by LINDA NICOLE NOVITSKI 2014

### ACKNOWLEDGEMENTS

Many thanks are due to the research group in which this research was created; my main advisor, Dr. R. Jan Stevenson, Dr. Jiaguo Qi, Dr. Dave Hyndman, Dr. Peter Esselman, Dr. Anthony Kendall, Dr. Sherry Martin, Dr. Siam Lawawirojwong, Dr. Tanita Suepa, and Dr. Shengpan Lin. Thanks are also due to committee members Dr. Elena Litchman, Dr. Orlando Sarnelle, and Dr. Mantha Phanikumar. Thanks to Dr. Narumon Wiangwang and Zhen Zhang for their assistance. Finally, I am deeply appreciative of the continuous support of my loving parents, sisters, and husband through this process.

LIST OF TABLES	viii
LIST OF FIGURES	ix
KEY TO ABBREVIATIONS	xi
CHAPTER 1	
REMOTE SENSING OF CHLOROPHYLL-A IN THE GREAT LAKES USING LANDSAT AND	)
MODIS SATELLITE IMAGERY; A HISTORY	1
	1 5
	5
CHAPTER 2 NATIONAL CHLOROPHYLL AND SECCHI DETECTION BY LANDSAT 7 ETM+ SATELLI' IMACEBY FOR WATER OLIALITY ASSESSMENT	<b>ГЕ</b> 11
Abstract	11
Introduction	11
Mathads	15
Fcological data satellite imagery and nivel extraction	10
Remote sensing images and data processing	10
l inpar rearpssion analyses	17
Boosted rearession tree analyses	20
Model diagnostics	21
Results	22
National and lake type models to predict chl	22
National and lake type models to predict Secchi denth	
Discussion	
Evaluation of national chl model	25
Evaluation of national Secchi depth model	25
Possible improvement of models	
Robustness and application of national models	
Summary and concluding remarks	
APPENDIX	
LITERATURE CITED	
CHAPTER 3 modis and landsat boosted decression thee chi ododuvi i ddedictive	
MODIS AND LANDSAT BOOSTED REGRESSION TREE CHLOROPHTLL PREDICTIVE MODELS FOR THE GREAT LAKES	52

AUDELS FOR THE GREAT LAKES	
Abstract	
Introduction	
Methods	

Study locations and ecological data	55
MODIS satellite image processing and pixel extraction for BRT model training .	55
Landsat image processing and pixel extraction for BRT model training	56
Data preparation and statistics for MODIS and Landsat data	58
MODIS image processing for model application	59
Landsat image processing for model application	60
Model evaluation	61
Results	61
BRT models and partial correlation plots	61
Discussion	63
APPENDIX	67
LITERATURE CITED	80

### **CHAPTER 4**

# IMPACTS OF CLIMATE CHANGE AND LAND-USE/LAND-COVER CHANGE ON PHYTOPLANKTON IN INNER SAGINAW BAY USING LANDSAT AND MODIS SATELLITE IMAGERY, 1973-2012 84 Abstract 84 Introduction 85 Research objectives 88 Methods 88 Data 88 Statistics for MODIS and Landsat 90 Results 92 Temporal trends 92 Relationships between algal blooms and global change variables 93 Discussion 94 APPENDIX 98 LITERATURE CITED 115

### **CHAPTER 5**

SUMMARY AND SYNTHESIS: USE OF REMOTE SENSING IN ECOLOGIAL ASSESSMENT12	21
National Lakes Assessment chlorophyll and Secchi depth predictive models12	21
Landsat and MODIS boosted regression tree models for chlorophyll prediction	
in the Great Lakes12	22
Using Landsat- and MODIS-derived chlorophyll to assess water quality in	
Saginaw Bay from 1972-201212	23
LITERATURE CITED12	25

## LIST OF TABLES

Table 2-1: Landsat 7 ETM+ wavelength band features (USGS 2011a).	32
Table 2-2: US EPA National Lakes Assessment lake types	33
Table 2-3: National boosted regression tree chl and Secchi depth models.	34
Table 2-4: National and lake type linear regression chl detection models	35
Table 2-5: National and lake type linear regression chl and Secchi model coefficients	. 36
Table 2-6: National and lake type linear regression Secchi depth detection models	37
Table 2-7: Chlorophyll and Secchi depth models from the literature	38
Table 3-1: MODIS wavelength band features (NASA 2013).	68
Table 3-2: Landsat 7 ETM+ wavelength band features (USGS 2011)	69
Table 3-3: Landsat 1-3 MSS wavelength band features (USGS 2011).	70
Table 3-4: Boosted regression tree results for MODIS, Landsat ETM+, TM, and MSS	71
Table 3-5: Comparison of BRT chl predictive models to models in the literature	72
Table 4-1: Linear regression results for temporal trends in chlorophyll, water temperature, and phosphorus loads in inner Saginaw Bay	99
Table 4-2: Wilcoxon Rank test results using near and far groupings of MODIS or Landsat chl or water temperature data.	100
Table 4-3. MODIS (2000-2012) linear regression results.	101
Table 4-4: MODIS linear mixed-effect model results.	102
Table 4-5: Landsat (1973-2012 or 1984-2012) linear regression results.	103
Table 4-6: Landsat linear mixed-effect model results.	104
Table 4-7: Comparison of MODIS- and Landsat-derived inner Saginaw Bay         chlorophyll values and values from the literature	105

# **LIST OF FIGURES**

Figure 2-1: Histograms of untransformed and In-transformed chlorophyll and Secchi depth4	2
Figure 2-2: Remotely sensed (linear regression predicted) ln-transformed chlorophyll a using band ratio 1/3 versus measured ln-transformed chlorophyll a (R <sup>2</sup> adj=0.19, 1.38 ln-transformed $\mu$ g/L)	:3
Figure 2-3: Remotely sensed (boosted regression tree predicted) natural ln- transformed chlorophyll versus measured ln-transformed chlorophyll a (R <sup>2</sup> adj=0.44, 0.76 ln-transformed $\mu$ g/L)	.4
Figure 2-4: Remotely sensed (linear regression predicted) natural ln-transformed Secchi depth using band ratio 1/3 versus measured ln-transformed Secchi depth. (R <sup>2</sup> adj=0.49, 0.77 ln-transformed m RMSE)	-5
Figure 2-5: Remotely sensed (boosted regression tree predicted) Secchi depth versus measured untransformed Secchi depth (R <sup>2</sup> adj=0.52, 0.80 m RMSE)	:6
Figure 3-1: Measured chl versus MODIS BRT predicted chl for 2007-2009 GLNPO training data with 1:1 line	'3
Figure 3-2: Partial dependence plots for the MODIS BRT7	4
Figure 3-3: Measured chl versus Landsat ETM+ and TM BRT predicted chl for 2007- 2009 GLNPO training data with 1:1 line	6
Figure 3-4: Partial dependence plots for the Landsat ETM+ and TM BRT7	7
Figure 3-5: Measured chl versus Landsat MSS predicted chl for 2007-2009 GLNPO training data with 1:1 line	'8
Figure 3-6: Partial dependence plots for the Landsat MSS BRT7	′9
Figure 4-1: Saginaw Bay's inner bay with depth ≥ 5 m1	.07
Figure 4-2: Inner Saginaw Bay study area and river mouth inputs into Saginaw Bay1	.08
Figure 4-3: Annual average and upper quartile MODIS and Landsat-derived chlorophyll concentration ( $\mu$ g/L) in inner Saginaw Bay, 2000-2012	.09
Figure 4-4: Annual average and upper quartile Landsat-derived chlorophyll concentration ( $\mu$ g/L) in inner Saginaw Bay, 1973-20121	.10

Figure 4-5: Annual average and upper quartile 8-day phosphorus load (kg) in inner Saginaw Bay, 1973-2012	.111
Figure 4-6: Annual average and upper quartile Landsat-derived surface water temperature (°C) in inner Saginaw Bay, 1984-2012	.112
Figure 4-7: MODIS-derived chlorophyll vs. mixed-effect model chlorophyll ( $\mu$ g/L) in inner Saginaw Bay, 2000-2012 with 1:1 line	.113
Figure 4-8: Landsat-derived chlorophyll vs. mixed-effect model chlorophyll ( $\mu$ g/L) in inner Saginaw Bay, 1984-2012 with 1:1 line	.114

### **KEY TO ABBREVIATIONS**

### A=August

- AIC = Akaike information criterion
- avg. = average
- BIC = Bayesian information criterion
- BRT = boosted regression tree
- chl = chlorophyll a
- D = distance between satellite image pixel and Saginaw River mouth (km)
- diff. days = difference between in-lake sample date and date of satellite overpass
- DOC = dissolved organic carbon

DOY = day of year

- ETM+ = enhanced thematic mapper
- GLNPO = Great Lakes National Program Office
- JL=July

JN=June

- L1G = Landsat image processing level
- L1T = Landsat image processing level
- LOESS = locally weighted scatterplot smoothing
- max = maximum
- MCKT = MODIS conversion toolkit
- MERIS = Medium Resolution Imaging Spectrometer

min = minimum

MOD02 = MODIS satellite data product

- MOD09 = MODIS satellite data product
- MODIS = Moderate Resolution Imaging Spectroradiometer
- Month=sample month
- MRT = MODIS reprojection tool
- MSS = multispectral scanner
- NASA = National Aeronautics and Space Administration
- NLA = United States Environmental Protection Agency National Lakes Assessment
- NOAA = National Oceanic and Atmospheric Administration
- NS = not statistically significant
- NSE = Nash-Sutcliffe efficiency
- OT=one-time
- P = 8-day phosphorus load (kg)
- R=range
- RMSE = root mean square error
- S=September
- SA=seasonal average
- SeaWiFS = Sea-Viewing Wide Field-of-View Sensor
- TM = thematic mapper
- TOA = top of atmosphere
- TOTS=two one-time samples
- TP = total phosphorus
- Type=sample type

UQ = upper quartile

# USGS = United States Geological Survey

# WT = Landsat-derived water temperature (°C)

### CHAPTER 1 REMOTE SENSING OF CHLOROPHYLL-A IN THE GREAT LAKES USING LANDSAT AND MODIS SATELLITE IMAGERY; A HISTORY

### **Literature Review**

Since the 1970s, improving water quality has been a national priority, however, in the Great Lakes, concerns, such as fish population declines (Burkholder 1930) and eutrophication (Beeton 1965), were noted much earlier. One retrospective study of diatom records indicates nutrient enrichment as early as European settlement and forest clearing in some part of the Great Lakes (Schelske et al. 1983). In 1972, guidelines were established by the Clean Water Act that mandated regulation of point source pollution into water bodies, including municipal and industrial sources of nutrients (Copeland 2010). Amendments to the Clean Water Act in 1987 added regulations for nonpoint nutrient sources, including agriculture (Copeland 2010). In addition to the protection provided to the Great Lakes by the Clean Water Act, independent remediation action plans have been created and carried out for numerous areas of concern, including Saginaw Bay (USEPA 2013). In the years following remediation, nutrients and chemicals from municipal, industrial, and agricultural sources decreased significantly and led to improvements in water quality in some areas (Auer et al. 1976, Bierman et al. 1984). Meanwhile, other perturbations, including invasive species introductions and climate change, have caused additional ecological challenges (Fahnenstiel et al. 1995, Lofgren and Gronewold 2012).

To enable effective assessment, remediation, and management of affected water bodies, it is necessary to understand how the water quality varies in space and time and how it is related to sources of pollution. Chlorophyll a (chl) concentration, a proxy of algal

biomass, is often used to help quantify ecological health of water bodies, as it can be indicative of high nutrient levels and water temperatures (Carlson 1977, Paerl and Huisman 2008). In some cases, high algal biomass may have deleterious effects on fish and other aquatic organisms, as well as humans who touch or ingest water with algae (Paerl et al. 2001). Numerous studies indicate that algal blooms have the potential to increase (Hauer et al. 1997) and, in some areas, are already increasing in frequency, which could be caused by increased nutrient input, higher water temperatures and increasing atmospheric CO<sub>2</sub> (Shapiro 1997, Hunter 2003).

Water quality assessments are usually done by physically taking water samples and analyzing them in the lab. However, advancements in remote sensing technology and development of modeling tools in recent decades have allowed ecological assessment from a distance, enabling evaluation with larger spatial and temporal scales and providing a more comprehensive view at, potentially, less cost. Landsat satellite imagery has often been used to detect chl in inland lakes (Lillesand et al. 1983, Mayo et al. 1995) and coastal ocean waters (Erkkila and Kalliola 2004, Han and Jordan 2005). Landsat also has a long history in water quality detection in the Great Lakes, from observation of colors in Lake Erie (Strong 1974), calcium carbonate precipitation in Lake Michigan and Lake Ontario (Strong 1978), chl detection in central Lake Michigan and Green Bay (Lathrop and Lillesand 1986), total suspended solids or Secchi depth in Green Bay (Lathrop et al. 1991), or phycocyanin detection in western Lake Erie (Vincent et al. 2004). While some Landsat studies have looked back at water quality over time (Olmanson et al. 2008), none have looked back at chl trends over the lifetime of Landsat.

Since its inception, chl algorithms have also been created, tested, and used for the

Moderate Resolution Imaging Spectroradiometer (MODIS) satellite in both ocean waters (O'Reilly et al. 2000, Carder et al. 2004) and in the Great Lakes (Shuchman et al. 2006, Kerfoot et al. 2008, Weghorst 2008, Becker et al. 2009, Kerfoot et al. 2010, Binding et al. 2012, Lesht et al. 2013). One of the early MODIS chl predictive models, the OC3 model that was created for ocean waters (O'Reilly et al. 2000) was applied with some success in freshwater lakes with little non-algal color producing agents (CPAs), or when chl was the dominant CPA at the time (Kerfoot et al. 2008), but not as well when other CPAs are present. More recent MODIS algorithms for freshwater lakes have tried to take account of the other CPAs, such as colored dissolved organic matter (CDOM) and suspended particular matter (Shuchman et al. 2006, Binding et al. 2012).

Whether using chl models for Landsat or MODIS, complications arise when detecting chl in shallow areas, where reflectance from the bottom of the lake could interfere with the chl signal detected by the satellite. Numerous studies have acknowledged and tried to address this issue (Højerslev et al. 1977, Lyzenga 1981, Rundquist et al. 1995, Lee et al. 1998, Tolk et al. 2000, Lodhi and Rundquist 2001, Cannizzaro and Carder 2006), however no standard approach has been widely accepted and applied.

Historically, two general approaches have been use to develop models for predicting chl with satellite images: 1) deriving empirical relationships between chl and satellite bands or band ratios, alone or separately from other CPAs using linear or polynomial regression analyses (Carpenter and Carpenter 1983, O'Reilly et al. 1998, Giardino et al. 2001) or 2) deriving relationships between two or more CPAs (including chl) at once (Carder et al. 2004, Pozdnyakov et al. 2005). Many of the current models are either not widely applicable, have not been tested for wide use, are challenging to implement without

expert statistical and programming knowledge, or require costly computer programs.

Other algorithms have been created and tested for the Great Lakes using other satellites, including the medium resolution imaging spectrometer (MERIS)(Gons et al. 2008, Wynne et al. 2008), CZCS and SeaWiFS (Bukata et al. 1981, Shuchman et al. 2006, Binding et al. 2007), or hyperspectral devices (Ali 2011). For additional information, there are comprehensive summaries of remote sensing of cyanobacteria (Kutser 2009), of case II waters (Bukata 2005, Odermatt et al. 2012, Lesht et al. 2013), or ocean waters (Stumpf and Tomlinson 2005).

In my dissertation I used field data in conjunction with satellite imagery to test models for chl and Secchi depth using existing remote sensing models and using a new approach with boosted regression tree (BRT) model-building techniques. In the first project, US Environmental Protection Agency field measurements and Landsat satellite imagery were used to create chl concentration and Secchi depth predictive models for use in small inland lakes across the United States. The second chapter explorers the use of BRT models for chl prediction in the Great Lakes, using both Landsat and MODIS satellite imagery. Landsat and MODIS complement each other in that Landsat allows for long-term analysis of algal biomass at a finer spatial scale, and MODIS allows for chl detection with finer spectral wavelengths at more frequent intervals. The third chapter uses Landsat and MODIS-derived chl concentrations in mixed-effects models to better understand the influence of river discharge and nutrient loading on chl concentration in Saginaw Bay. I expect that developing remote sensing tools for water quality assessment can be useful for ecological analysis to not only understand the natural dynamics of the ecosystem, but also help predict future trends caused by climate change.

LITERATURE CITED

### LITERATURE CITED

- Ali, K. A. 2011. Prediction of water qualitly parameters from VIS-NIR radiometry: Using Lake Erie as a natural laboratory for anallysis of case 2 waters. Kent State University.
- Auer, M. T., R. P. Canale, and P. L. Freedman. 1976. The limnology of Grand Traverse Bay, Lake Michigan. Michigan Sea Grant Program.
- Becker, R. H., M. I. Sultan, G. L. Boyer, M. R. Twiss, and E. Konopko. 2009. Mapping cyanobacterial blooms in the Great Lakes using MODIS. Journal of Great Lakes Research **35**:447-453.
- Beeton, A. M. 1965. Eutrophication of the St. Lawrence Great Lakes. Limnology and Oceanography **10**:240-254.
- Bierman, V. J., Jr., D. M. Dolan, and R. Kasprzyk. 1984. Retrospective analysis of the response of Saginaw Bay, Lake Huron, to reductions in phosphorus loadings. Environmental Science & Technology 18:23-31.
- Binding, C. E., T. A. Greenberg, and R. P. Bukata. 2012. An analysis of MODIS-derived algal and mineral turbidity in Lake Erie. Journal of Great Lakes Research **38**:107-116.
- Binding, C. E., J. H. Jerome, R. P. Bukata, and W. G. Booty. 2007. Trends in water clarity of the lower Great Lakes from remotely sensed aquatic color. Journal of Great Lakes Research **33**:828-841.
- Bukata, R. P. 2005. Satellite monitoring of inland and coastal water quality: Retrospection, inspection, and future directions. Taylor & Francis, Boca Raton, FL.
- Bukata, R. P., J. E. Bruton, J. H. Jerome, S. C. Jain, and H. H. Zwick. 1981. Optical water quality model of Lake Ontario. 2: Determination of chlorophyll a and suspended mineral concentrations of natural waters from submersible and low altitude optical sensors. Applied Optics 20:1704-1714.

Burkholder, P. R. 1930. A biological survey of Lake Erie. Science **71**:288-289.

Cannizzaro, J. P. and K. L. Carder. 2006. Estimating chlorophyll a concentrations from remote-sensing reflectance in optically shallow waters. Remote Sensing of Environment **101**:13-24.

- Carder, K. L., F. R. Chen, J. P. Cannizzaro, J. W. Campbell, and B. G. Mitchell. 2004. Performance of the MODIS semi-analytical ocean color algorithm for chlorophyll-a. Advances in Space Research **33**:1152-1159.
- Carlson, R. E. 1977. A trophic state index for lakes. Limnology and Oceanography **22**:361-369.
- Carpenter, D. J. and S. M. Carpenter. 1983. Modeling inland water quality using Landsat data. Remote Sensing of Environment **13**:345-352.
- Copeland, C. 2010. Clean Water Act: A Summary of the Law. Congressional Researc Service.
- Erkkila, A. and R. Kalliola. 2004. Patterns and dynamics of coastal waters in multi-temporal satellite images: support to water quality monitoring in the Archipelago Sea, Finland. Estuarine, Coastal, and Shelf Sciences **60**:165-177.
- Fahnenstiel, G. L., T. B. Bridgeman, G. A. Lang, M. J. McCormick, and T. F. Nalepa. 1995. Phytoplankton productivity in Saginaw Bay, Lake Huron: Effects of zebra mussel (*Dreissena polymorpha*) colonization. Journal of Great Lakes Research 21:465-475.
- Giardino, C., M. Pepe, P. A. Brivio, P.Ghezzi, and E.Zilioli. 2001. Detecting chlorophyll, Seechi disk depth and surface tempterature in a sub-alpine lake using Landsat imagery. The Science of the Total Environment **268**:19-29.
- Gons, H. J., M. T. Auer, and S. W. Effler. 2008. MERIS satellite chlorophyll mapping of oligotrophic and eutrophic waters in the Laurentian Great Lakes. Remote Sensing of Environment **112**:4098-4106.
- Han, L. and K. J. Jordan. 2005. Estimating and mapping chlorophyll-a concentration in Pensacola Bay, Florida using Landsat ETM+ data. International Journal of Remote Sensing **26**:5245-5254.
- Hauer, F. R., J. S. Baron, D. H. Campbell, K. D. Fausch, S. W. Hostetler, G. H. Leavesley, P. R. Leavitt, D. M. McKnight, and J. A. Stanford. 1997. Assessment of climate change and freshwater ecosystems of the Rocky Mountains, USA and Canada. Hydrological Processes 11:903-924.
- Højerslev, N., N. G. Jerlov, and G. Kullenberg. 1977. Colour of the ocean as an indicator in phtosynthetic studies. Journal du Conseil international pour l'Exploration de la Mer **37**:313-316.
- Hunter, P. R. 2003. Climate change and waterborne and vector-borne disease. Journal of Applied Microbiology, Journal of Applied Microbiology Symposium Supplement **94**:37S-46S.

- Kerfoot, W. C., J. W. Budd, S. A. Green, J. B. Cotner, B. A. Biddanda, D. J. Schwab, and H. A. Vanderploeg. 2008. Doughnut in the desert: Late-winter production pulse in southern Lake Michigan. Limnology and Oceanography 53:589-604.
- Kerfoot, W. C., F. Yousef, S. A. Green, J. W. Budd, D. J. Schwab, and H. A. Vanderploeg. 2010. Approaching storm: Disappearing winter bloom in Lake Michigan. Journal of Great Lakes Research 36:30-41.
- Kutser, T. 2009. Passive optical remote sensing of cyanobacteria and other intense phytoplankton blooms in coastal and inland waters. International Journal of Remote Sensing **30**:4401-4425.
- Lathrop, R. G., Jr. and T. M. Lillesand. 1986. Use of Thematic Mapper data to assess water quality in Green Bay and central Lake Michigan. Photogrammetric Engineering and Remote Sensing **52**:671-680.
- Lathrop, R. G., T. M. Lillesand, and B. S. Yandell. 1991. Testing the utility of simple multidate Thematic Mapper calibration algorithms for monitoring turbid inland waters. International Journal of Remote Sensing **12**:2045-2063.
- Lee, Z., K. L. Carder, C. D. Mobley, R. G. Steward, and J. S. Patch. 1998. Hyperspectral remote sensing for shallow waters. I. A semianalytical model. Applied Optics **37**:6329-6338.
- Lesht, B. M., R. P. Barbiero, and G. J. Warren. 2013. A band-ratio algorithm for retrieving open-lake chlorophyll values from satellite observations of the Great Lakes. Journal of Great Lakes Research **39**:138-152.
- Lillesand, T. M., W. L. Johnson, R. L. Deuell, O. M. Lindstrom, and D. E. Meisner. 1983. Use of Landsat data to predict the trophic stat of Minnesota lakes. Photogrammetric Engineering and Remote Sensing **49**:219-229.
- Lodhi, M. A. and D. C. Rundquist. 2001. A spectral analysis of bottom-induced variation in the colour of Sand Hills lakes, Nebraska, USA. International Journal of Remote Sensing **22**:1665-1682.
- Lofgren, B. and A. Gronewold. 2012. Water Resources. Great Lakes Integrated Sciences and Assessments Center.
- Lyzenga, D. R. 1981. Remote sensing of bottom reflectance and water attenuation parameters in shallow water using aircraft and Landsat data. International Journal of Remote Sensing **2**:71-82.
- Mayo, M., A. Gitelson, Y. Z. Yacobi, and Z. Ben-Avraham. 1995. Chlorophyll distribution in Lake Kinneret determined from Landsat Thematic Mapper data. International Journal of Remote Sensing **16**:175-182.

- O'Reilly, J. E., S. Maritorena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. McClain. 1998. Ocean color chlorophyll algorithms for SeaWiFS. Journal of Geophysical Research **103**:24,937-924,953.
- O'Reilly, J. E., S. Maritorena, M. C. O'Brien, D. A. Siegel, D. Toole, D. Menzies, R. C. Smith, J. L. Mueller, B. G. Mitchell, M. Kahru, F. P. Chavez, P. Strutton, G. F. Cota, S. B. Hooker, C. R. McClain, K. Carder, F. Müller-Karger, L. Harding, A. Magnuson, D. Phinney, G. F. Moore, J. Aiken, K. R. Arrigo, R. Letelier, and M. Culver. 2000. SeaWiFS postlaunch calibration and validation analyses, Part 3.*in* S. B. a. E. R. F. Hooker, editor. SeaWiFS Postlaunch Techinical Report Series. NASA.
- Odermatt, D., A. Gitelson, V. E. Brando, and M. Schaepman. 2012. Review of constituent retrieval in opticaly deep and complex waters from satellite imagery. Remote Sensing of Environment **118**:116-126.
- Olmanson, L. G., M. E. Bauer, and P. L. Brezonik. 2008. A 20-year Landsat water clarity census of Minnesota's 10,000 lakes. Remote Sensing of Environment **112**:4086-4097.
- Paerl, H. W. and J. Huisman. 2008. Blooms like it hot. Science 320:57-58.
- Paerl, H. W., I. R.S. Fulton, P. H. Moisander, and J. Dyble. 2001. Harmful freshwater algal blooms, with and emphasis on cyanobacteria. The Scientific World **1**:76-113.
- Pozdnyakov, D., R. Shuchman, A. Korosov, and C. Hatt. 2005. Operational algorithm for the retrieval of water qualilty in the Great Lakes. Remote Sensing of Environment **97**:352-370.
- Rundquist, D. C., J. F. Schalles, and J. P. Peake. 1995. The response of volume reflectance to manipulated algal concentrations above bright and dark bottoms at various depth in an experimental pool. Geocarto International **10**:5-14.
- Schelske, C. L., E. F. Stoermer, D. J. Conley, J. A. Robbins, and R. M. Glover. 1983. Early eutrophication in the lower Great Lakes: New evidence from biogenic silica in sediments. Science **222**:320-322.
- Shapiro, J. 1997. The role of carbon dioxide in the initiation and mantenance of blue-green dominance in lakes. Freshwater Biology **37**:307-323.
- Shuchman, R., A. Korosov, C. Hatt, D. Pozdnyakov, J. Means, and G. Meadows. 2006. Verification and application of a bio-optical algorithm for Lake Michigan using SeaWiFS: A 7-year inter-annual analysis. Journal of Great Lakes Research 32:258-279.
- Strong, A. E. 1974. Remote sensing of algal blooms by aircraft and satellite in Lake Erie and Utah Lake. Remote Sensing of Environment **3**:99-107.

- Strong, A. E. 1978. Chemical whitings and chlorophyll distributions in the Great Lakes as viewed by Landsat. Remote Sensing of Environment **7**:61-72.
- Stumpf, R. P. and M. C. Tomlinson. 2005. Remote sensing of harmful algal blooms. Pages 277-296 in R. L. Miller, C.E. Del Castillo, and B.A. McKee, editor. Remote Sensing of Coastal Aquatic Environments. Springer, Ah Dordrecht, The Netherlands.
- Tolk, B. L., L. Han, and D. C. Rundquist. 2000. The impact of bottom brightness on spectral reflectance of suspended sediments. International Journal of Remote Sensing **21**:2259-2268.
- USEPA. 2013. Great Lakes Areas of Concern. United States Environmental Protection Agency.
- Vincent, R. K., X. Qin, R. M. McKay, J. Miner, K. Czajkowski, J. Savino, and T. Bridgeman. 2004. Phycocyanin detection from Landsat TM data for mapping cyanobacterial blooms in Lake Erie. Remote Sensing of Environment 89:381-392.
- Weghorst, P. L. 2008. MODIS algorithm assessment and principal component analysis of chlorophyll concentration in Lake Erie. Kent State University.
- Wynne, T. T., R. P. Stumpf, M. C. Tomlinson, R. A. Warner, P. Tester, J. Dyble, and G. L. Fahnenstiel. 2008. Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. International Journal of Remote Sensing **29**:3665-3672.

### CHAPTER 2 NATIONAL CHLOROPHYLL AND SECCHI DETECTION BY LANDSAT 7 ETM+ SATELLITE IMAGERY FOR WATER QUALITY ASSESSMENT

### Abstract

We developed remote sensing models for chlorophyll a (chl) and Secchi depth for use at national and regional scales with US EPA's National Lake Assessment data (NLA) and Landsat 7 ETM+ satellite images. Using linear regression, separate Landsat models for chl and Secchi depth were developed at the national scale and for each of seven NLA lake classes to determine if climatic as well as naturally varying physical and chemical attributes could reduce variation in Landsat models. National boosted regression tree (BRT) models were also made and compared to national models made with linear regression. For all models, Landsat bands and band ratios were used to infer chl or Secchi depth, and adjusted  $R^2$  and RMSE were used to select the best models. For the linear regression models, the national chl model performed better than the lake type models, except for lake type B, and the national Secchi depth model performed better than all lake types except for lake types F and G. Overall, the best national Landsat model for In-transformed chl was the boosted regression tree model and had a cross-validation  $R^2$  of 0.44 and a 0.76 ln-transformed  $\mu g/L$ RMSE. The best national Landsat model for Secchi depth was also a boosted regression tree model that had an adjusted R<sup>2</sup> of 0.52 and a 0.80 m RMSE. Potential sources of error in the national chl boosted regression tree model were indicated by residuals being positively correlated with DOC (adjusted R<sup>2</sup>=0.14) and cyanobacterial density (adjusted R<sup>2</sup>=0.18), and negatively correlated with lake depth (adjusted R<sup>2</sup>=0.25). Lake area, difference between image and sample date, and percent cloud cover were not correlated with national chl boosted regression tree model residuals. National Secchi boosted regression tree model

residuals were weakly, negatively correlated with DOC (adjusted  $R^2=0.08$ ) and cyanobacterial density (adjusted  $R^2=0.06$ ) and weakly, positively correlated with difference between image and sample dates (adjusted  $R^2=0.02$ ) and lake depth (adjusted  $R^2=0.17$ ). Percent cloud cover of Landsat image and lake area were not correlated with national Secchi boosted regression tree model residuals. We assessed the applicability of the national chl model for ecological analysis by comparing the total phosphorus-chl relationship with chl determined from sampling or remote sensing, which showed the total phosphorus-chl relationship had an adjusted  $R^2=0.58$  and 1.02 ln-transformed  $\mu$ g/L RMSE with sampled chl versus an adjusted  $R^2=0.56$  and 1.04 ln-transformed  $\mu$ g/L RMSE with chl determined by the boosted regression tree remote sensing model. Remote sensing water quality assessment tools can be valuable for limnological study, ecological assessment, and water resource management.

### Introduction

Lakes are important sources of water for drinking, agriculture, industrial processes, recreation, and aesthetic well being. Water quality is highly threatened by withdrawals, pollution, and climate change-related alterations in precipitation and temperature (Parry et al. 2007). Toxic or nuisance algal growth in lakes can result from human activities generating nitrogen and phosphorus pollution (Carpenter et al. 1998) and may be exacerbated by increasing temperature with climate change (Paerl and Huisman 2008). Excessive algal biomass causes a variety of problems ranging from decreasing oxygen as the algae decompose (Wetzel 2001), harboring pathogens (Ishii et al. 2006, Byappanahalli et al. 2009), and generating toxins that harm aquatic organisms (Carmichael 1996) and cause sickness or death in terrestrial vertebrates (Briand et al. 2003) and humans (Watson et al. 2008, Backer et al. 2010). Therefore, developing and refining tools to assess algae at spatial and temporal scales that will advance ecological understanding for resource management is imperative as human needs and climate change stress lake ecosystems.

Satellite imagery has been used to assess water quality of freshwater ecosystems for several decades (Strong 1974, Scarpace et al. 1979, Lillesand et al. 1983, Gitelson et al. 1993, Kloiber et al. 2002a, Bustamante et al. 2009) Remote water quality assessment has the potential to be more time and cost-effective than traditional sampling techniques because the cost of acquiring some satellite imagery and time for processing are relatively low. Remotely sensed imagery also has advantages over more traditional sampling techniques by incorporating spatial variation over the size of a pixel and providing spatial detail throughout a lake. In addition, many historic satellite images are available, which enables temporal study of climate and land use change.

Landsat has greater potential for historical analyses than other satellites with nearly four decades of available data. Perhaps most importantly, unlike other satellites (e.g., MODIS, SeaWiFS), Landsat's small pixel size allows for the assessment of small lakes and fine resolution of spatial patterns within bigger lakes (Chipman et al. 2004, Han and Jordan 2005, Olmanson et al. 2008, Torbick et al. 2008). The disadvantages of Landsat are the relatively low number of bands compared to other satellites and relatively wide bandwidths, which produce some limitations on spectral resolution. However, Landsat imagery was used as early as Strong (1974) to detect the presence of algae in freshwater lakes. Since Strong's study, many models using spectral band measurements from Landsat images have been developed for inference of chl and Secchi depth using data from relatively small regions, such as states, and using only a subset of the available Landsat bands (Kloiber et al. 2002b, Nelson et al. 2003, Chipman et al. 2004).

While Landsat's visible wavelengths capture peak absorption and reflectance wavelengths for chl, they also capture changes in reflectance due to the increased scattering of light by particles near the water's surface (Lathrop et al. 1991, Han et al. 1994). Therefore, we might expect the visible wavelength bands to be the most useful for inferring chl and Secchi depth. However, infrared bands can also capture the scattering caused by either particles in the surface water (Lavery et al. 1993, Kloiber et al. 2002a), or by floating biomass, which reflects more light than clear water (Strong 1974). Additionally, the thermal infrared band can detect increases in surface water temperature that can both cause increased algal biomass and indicate warming by energy absorption and heat dissipation by organic and inorganic matter (Wetzel 2001). Furthermore, using bands in ratios can help minimize the relative influence of atmosphere between images (Jensen

2005). Despite this demonstrated utility, many studies have excluded all or most of the infrared bands in Landsat models of water quality parameters (Brivio et al. 2001, Kloiber et al. 2002b, Nelson et al. 2003, Chipman et al. 2004, Olmanson et al. 2008).

Most Landsat remote sensing studies have been limited in geographic range, or in the ranges of chl or Secchi depth they use to develop the model, for example, not covering both the lower and upper range of chl concentrations found in nature (Baban 1993, Cox et al. 1998, Allee and Johnson 1999, Giardino et al. 2001, Duan et al. 2007). Often the geographic ranges of studies are single lakes or states (Carpenter and Carpenter 1983, Mayo et al. 1995, Brivio et al. 2001, Nelson et al. 2003, Mishra and Mishra 2010). Having one or a strategically designed set of remote sensing models that could be used throughout the country would enhance lake studies by providing data at spatial and temporal scales that practically cannot be achieved by water sampling and also provide a standard method.

The goal of our study was to develop and test remote sensing models of chl and Secchi depth using Landsat 7 ETM+ imagery for application across large spatial and temporal scales by using field measurements from the National Lakes Assessment (NLA) of the United States Environmental Protection Agency (USEPA). We developed national chl and national Secchi models and compared them to a set of models for different types of lakes found across the US, using linear regression. National chl and Secchi models were also made with boosted regression tree statistical techniques and compared to the national models made with linear regression. In this effort, all Landsat bands and band ratios, including infrared and thermal, were used in model development. Sources of error, such as confounding environmental factors and non-linearities of relationships were evaluated. Finally, to determine if the national remote sensing model for chl was accurate enough to

be useful for ecological and management studies, we compared relationships between total phosphorus (TP) concentrations in lakes and either chl measured with water samples or chl measured with Landsat imagery.

### Methods

### Ecological data, satellite imagery, and pixel extraction

Algal and water chemistry data were collected during the USEPA's NLA from more than 1000 lakes in the 48 contiguous states (USEPA 2009). Lakes were sampled between May and October of 2007 (USEPA 2009). Landsat 7 ETM+ images with level 1G processing were downloaded from the US Geological Survey's Global Visualization Viewer website (USGS 2011b) for dates as close as possible to when lakes were sampled. Images were reprojected to a WGS-84 projection system in Erdas Imagine 9.3. Exact location of lakes was determined by overlaying a shapefile (made in Arc Map 9.3) of NLA sampling sites on each Landsat image. Landsat 7 ETM+ images are comprised of seven layers, each layer representing a different range of wavelengths, called "bands" (Table 2-1). Each layer is comprised of 30 m<sup>2</sup> pixels except for the thermal infrared, which has 60 m<sup>2</sup> pixels. Each of the pixel's seven layers has an associated, unitless numeric value that represents the amount of reflectance emitted by those Landsat band wavelengths. Pixel values were extracted from Landsat images using Erdas Imagine 9.3 software.

Lakes were included in our study if the sample location could be found in an image without being covered by clouds and images were taken within eight days of the NLA sample date. We picked eight days because Landsat images are taken every sixteen days. Thus, the maximum time between sample collection and a Landsat image would be eight days. Due to a mechanical malfunction, some Landsat 7 ETM+ images contain black stripes

in areas where information has been lost. If a sample location coincided with a black stripe, a nearby pixel (when possible in the same row or column as the sample pixel) was selected. The same method was used when sample locations were located in a mixed pixel that had both water and land signatures. We considered using all 1028 lakes in the 2007 NLA, however, clear Landsat images could not be found for some lakes within a month of the sample date and so were not used in the analysis. Also, lakes with maximum depth less than three times Secchi depth were removed due to the possible influence of lake bottom detection on Landsat images. This criterion allowed for retention of lakes in the dataset with a range of depth, from shallow to deep. The final dataset had 447 lakes. Chl ranged from  $0.14-684 \mu g/L$ . TP ranged from  $1-2147 \mu g/L$  in these lakes. Secchi depth ranged from 0.04 meters to 13.7 m.

### Remote sensing images and data processing

Radiometric calibration, provided by conversion to top of the atmosphere (TOA) reflectance values, is necessary to account for image differences in sun angle, in Earth-Sun distance, and in solar irradiance for different wavelengths that occur above the earth's atmosphere (Chander et al. 2009). Therefore, after pixel extraction, raw pixel values were converted to spectral radiance values with the equation:

Radiance = 
$$\frac{(LMAX - LMIN)}{255}$$
 (pixel digital number) + LMIN

Where pixel digital number is the raw pixel value, and LMAX and LMIN correspond to known spectral radiances for each of the Landsat 7 ETM+ spectral bands (Chander 2009). Radiance values were further converted to TOA reflectance values with this equation:

$$P_p = \frac{(\Pi)(L_{\lambda})(d^2)}{(E_{\lambda})(\cos\theta_s)}$$

Here  $P_p$  is unitless planetary reflectance,  $L_\lambda$  is spectral radiance at the sensor's aperture, d is the earth-sun distance in astronomical units based on the image acquisition date (Chander et al. 2009), and  $E_\lambda$  is mean solar exoatmospheric irradiance. Solar zenith (( $\theta_s$ ) values were obtained from NOAA's Solar Position Calculator (NOAA 2011). High gain thermal band radiance values were converted directly to temperature (°C) with this equation:

$$T = \frac{K2}{\left(\frac{K1}{L\lambda} + 1\right)} - 273.15$$

Where, T is the effective at-satellite temperature in Kelvin, K2 is calibration constant, 1282.71, K1 is calibration constant, 666.09, and  $L_{\lambda}$  is the spectral radiance in watts/((m<sup>2</sup>)(steradians)(µm)) (Chander 2009).

If the TOA value was 0.3 or greater, pixel values for that lake were excluded from further analyses due to the likelihood of nearby cloud interference. No additional atmospheric correction was done due to lack of information on atmospheric conditions for the sampling date and location.

### Linear regression analyses

Natural log transformations were formed on both chl and Secchi depth data to obtain normal distributions before use in linear regression analyses (Figure 2-1). Linear regression analyses were run using Landsat's visible bands, infrared bands (including thermal infrared), and all possible two-band ratios to determine the best Landsat 7 ETM+ model. It should also be noted that, although including inverse ratios (e.g., 1/3 and 3/1) in the same model would seem to cancel effects of each other, inverse ratios have different rates of change along the same band ranges, and therefore have different weights in

different band ranges. Linear regression analyses were run, first, with each of the 49 possible band or band ratios, individually, to elucidate the relationship between each band or ratio and chl or Secchi depth. Testing all bands/ratios also formed a base for comparison to previous studies that often did not test the same subset of bands or ratios for their respective predictive power. Bootstrapping was run on each model to assess the robustness of the model. The bands/ratios that yielded the highest R<sup>2</sup> and lowest RMSE were chosen for multiple linear regression analysis. Multiple linear regression was done starting with the best predictive bands from the first round of linear regression and adding all other bands or ratios one at a time. The model was retained if it improved the linear regression model R<sup>2</sup> by at least 0.05. Next, bootstrapping was done on the multiple linear regression model to determine if R<sup>2</sup> value of the multiple linear regression retained the 0.05 increase in predictive power over the linear regression model. None of the bootstrapped multiple linear regression models improved the R<sup>2</sup> linear regression model by at least 0.05, so no further work was done on multiple linear regression models. We checked for non-linearities by running a LOESS analysis and visually comparing model fits. To assess model robustness, we looked for consistencies in band and band ratios used in the national models, the lake type models, and models in the literature.

We also developed chl and Secchi depth linear regression models by lake type to test whether grouping lakes by natural features allows for better predictive power. Lake types were determined by a cluster analysis of variables that are relatively unaffected by human activities (Herlihy et al. 2013), but do influence the lakes' ecology (Table 2-2). These variables included: surface area, maximum depth, morphology, elevation, maximum annual air temperature in the watershed, annual maximum of monthly precipitation, calcium

concentration, latitude, and longitude. Variables were transformed to reduce skewness, and they were standardized to provide similar weights. The cluster analysis was constrained by level three ecoregions, which were: Eastern Highlands; Plains and Lowlands; and Xeric and Mountain West. Seven lake types resulted from the cluster analysis, labeled A through G. As with national models, all bands and ratios were considered in the lake type models. We compared national model and lake type models by: 1) comparing the adjusted R<sup>2</sup> and RMSE of relationships between chl and Secchi depths measured by sampling and Landsat imagery using national and lake type specific models, and 2) comparing the resulting R<sup>2</sup> and RMSE values to remote sensing water quality models in the literature.

### Boosted regression tree analyses

To test whether simultaneously incorporating multiple Landsat spectral bands/ratios and accounting for non-linearities could improve predictive power, national boosted regression tree models were created using the same Landsat ETM+ pixel data as the linear regression models and all 49 band and band ratios that can be created from Landsat ETM+ bands 1-7, rather than previous analyses with one or two bands or ratios (as with linear or multiple linear regression models). Although response variables should, generally, not require transformation for use in boosted regression tree analysis, we found it necessary to ln-transform chl for with this dataset. This may be due to the extreme skewness of the chl data (Figure 2-1). Secchi depth was left untransformed for boosted regression tree analysis. Initial boosted regression tree models were simplified using the gbm.simplify program in R that determines the number of variables that can be removed without increasing predictive deviance (Elith et al. 2008). Root mean square error (RMSE)

values, calculated using predicted and measured chl, were compared for final model selection. Since boosted regression tree models were already cross-validated with data internal to the dataset we did cross-validate them a second time with bootstrapping. All statistics were run in gbm package 2.9 in R. R 2.15.1.

### Model diagnostics

An internal 10-fold cross-validation procedure, incorporated into the boosted regression tree program code, is designed to avoid overfitting the data. The crossvalidation process began with a division of the data into 10 subsets. Ten boosted regression trees were run, simultaneously. Model training was done on 90% of the data and testing it on the remaining 10% until each of the 10 subsets has been used as a test dataset. Next, a new set of 10 boosted regression trees was run and tested with a larger number of trees. This process was continued until, at a given number of trees, the average predictive power of the first five boosted regression trees was surpassed in predictive power by the second set of five boosted regression trees. At the point, the minimum (optimal) number of trees for the model had been passed. From this, we determined optimal tree number and continued to refine other model parameters, including learning rate, bag fraction, and tree complexity. Learning rate, or relative contribution of each tree to the final model, was determined to be 0.005 for the chl model and 0.011 for the Secchi model; optimal bag fraction, or amount of data chosen at random from the data set for tree construction, was determined to be 0.50 for both chl and Secchi models; optimal tree complexity, or approximate order of the model, was determined to be three for both chl and Secchi models; number of trees in final boosted regression tree models was 3050 for the chl model and 1800 for the Secchi model (Table 2-3).

To determine if the national boosted regression tree models could be improved with available information, we analyzed national model residuals to determine if other ecological variables were confounding the inference of chl or Secchi depth. Non-remotely sensed predictor variables used in the residuals analysis included dissolved organic carbon (DOC), cyanobacterial density (cells/cm<sup>2</sup>), "diff. days" (the number of days between sampling and Landsat image acquisition), percent cloud cover of the Landsat image, lake area, and maximum measured lake depth. Any variables that were correlated to national model residuals were further analyzed to elucidate trends. Partial correlation coefficients were calculated to determine the relative influence of band ratios vs. depth on chl or Secchi depth detection. Finally, to determine if the national boosted regression tree models are useful for ecological and management studies, we used linear regression to relate TP and actual, measured chl and Landsat-inferred chl. All statistics were run with R, version 2.15.0.

### Results

### National and lake type models to predict chl

For the linear regression national model, band ratios 1/2 and 1/3 were selected for the national model (Tables 2-4 and 2-5, Figure 2-2). Bootstrap adjusted R<sup>2</sup> values for all linear regression chl models were about the same as the original model adjusted R<sup>2</sup> (Table 2-4), perhaps, due to the simplicity of the model and the size of the training data set. No two-band/ratio multiple linear regression models had bootstrap R<sup>2</sup> values that improved the R<sup>2</sup> of the original linear regression model by more than 0.05. Band ratio 1/2 was also the best predictor of chl for lake types B, C, and D, while other bands or ratios were best

predictors for other lake types. Lake type chl models had lower adjusted R<sup>2</sup>, but also lower RMSE values than the national model in all cases, except for lake types B (Table 2-4).

Lake type B yielded the best model, while lake type D yielded the weakest (Table 2-4). Bands ratio 1/2 was most frequently the best predictor of chl in the lake type ratios and band 1 was used in at least one of the best predictor ratios in all lake types except for F. Band 3 was the next most common band used in best predictor band ratios. Band 4, near infrared, was used in a best predictor ratio in lake type E; band ratio 7/5 was just as good a predictor of chl as ratio 1/2 in lake type D; band 6, thermal infrared, was the best predictor in lake type F. Using ratio 1/2 on lake types for which it was not the best predictor, did not yield substantially poorer results, except for lake type A (Table 2-4). Some of the lake type models with smaller sample size also had some of the highest R<sup>2</sup> and lowest RMSE values (Table 2-4). Upon visual inspection, the national linear regression chl model did not display substantial non-linearities.

The national chl boosted regression tree model had the highest R<sup>2</sup> and lowest RMSE of all models (0.44 cross-validation R<sup>2</sup> and 0.76 ln-transformed  $\mu$ g/L RMSE; Table 2-3, Figure 2-3). National chl boosted regression tree model residuals were positively correlated with DOC (adjusted R<sup>2</sup> = 0.14) and cyanobacterial density (adjusted R<sup>2</sup> = 0.18), and negatively correlated with lake depth (adjusted R<sup>2</sup> = 0.25). Lake area, difference between image and sample date, and percent cloud cover were not correlated with national chl boosted regression tree model residuals. The relationship between TP and chl using NLA data had an adjusted R<sup>2</sup> = 0.58 and 4.90 µg/L RMSE, whereas the relationship between TP and Landsat inferred chl had an R<sup>2</sup> of 0.40 and 5.64 µg/L RMSE.
#### National and lake type models to predict Secchi depth

For the national linear regression Secchi model, band ratio 1/3 produced the best model, while ratios 3/1, 1/2, and 2/1 produced similarly predictive models (Table 2-5 and 2-6 and, Figure 2-4). Bootstrap adjusted R<sup>2</sup> values for all linear regression Secchi depth models were about the same as the adjusted R<sup>2</sup> of the original model (Table 2-6). As with the chl models, this similarity may be due to the simplicity of the model and the size of the training data set. As with chl models, no two-band/ratio multiple linear regression models had bootstrap  $R^2$  values that improved the original linear regression model  $\leq 0.05$ . The best Secchi depth models were for lake types F and G and the poorest in lake types D, E, and A. Band ratio 1/3 was also the best predictor of chl for lake types C, D, and E, while ratio 3/1 was equally predictive for lake type E. Ratio 2/1 yielded the best model for lake types B and F. Band 1 was used in all best predictor models for Secchi depth. No infrared bands were found to be useful in a best predictive model. Lake type models generally had lower adjusted R<sup>2</sup>, but higher RMSE values than the national model in all cases, except for lake types F and G (Table 2-6). Using ratio 1/3 on lake types for which it was not the best predictor, yielded similar results for best predictive models using band ratio 3/1 and more poorly for best predictive models using band ratio 2/1 (Table 2-6). All results for national and lake type models were statistically significant at  $p \le 0.05$ . As with the lake type chl models, some of the lake type Secchi depth models with smaller sample size also had some of the highest R<sup>2</sup> and lowest RMSE values (Table 2-6). Upon visual inspection, the national linear regression Secchi depth model did not display any non-linearities.

The national Secchi boosted regression tree model had the highest R<sup>2</sup> and lowest RMSE (0.52 cross-validation R<sup>2</sup> and 80 m RMSE; Table 2-6, Figure 2-5) of all models.

National Secchi boosted regression tree model residuals were weakly, negatively correlated with DOC (adjusted R<sup>2</sup> = 0.08) and cyanobacterial density (adjusted R<sup>2</sup> = 0.06) and positively correlated with lake area (adjusted R<sup>2</sup> = 0.02) and lake depth (adjusted R<sup>2</sup> = 0.17). Percent cloud cover of Landsat image and diff. days were not correlated with national Secchi boosted regression tree model residuals.

## Discussion

#### Evaluation of national chl model

The national chl boosted regression tree model made using a combination of Landsat bands/ratios performed well based on our evaluation criteria. The adjusted R<sup>2</sup> of our national and lake type models was lower than many other studies, and those studies were done at smaller scales or with satellite imagery with narrower bandwidths (Table 2-7). The linear regression model for lake type B also had good performance compared to others from smaller geographic areas. Therefore, grouping eastern highlands lakes and reservoirs may be useful for improvement of chl prediction. Alternatively, small RMSE values and correspondingly high adjusted R<sup>2</sup> values of some lake type models may be a result of a smaller sample size and/or smaller chl range. These results may also indicate that the sampling structure of the NLA, with one chl sample per lake, cannot adequately capture the spatial variability needed for model chl development. Having a gradient of chl within one lake could help to calibrate the model in relation to other lakes attributes, such as DOC. Chl inferred from Landsat imagery was a good predictor of TP when compared to measured chl.

#### Evaluation of national Secchi depth model

The national boosted regression tree Secchi depth model made using a combination of Landsat bands/ratios performed adequately, while the lake type models performed well based on our criteria, falling well within the performance range seen in the literature for smaller regions, based on R<sup>2</sup> values (Table 2-7). As with chl model results, smaller RMSE values with corresponding higher adjusted R<sup>2</sup> values for Secchi depth lake type models may, in some cases, be a result of a smaller Secchi depth range. Overall, our national boosted regression tree model performs well compared to the lake type linear regression models, and by incorporating spatial variation across the United States, it can be a valuable tool for larger-scale ecological studies. Interestingly, at least one each of the best predictor band ratios for national and lake type chl and Secchi models overlapped in best predictor band ratios, except for lake types C and D. This may be due to the overlapping information in chl and Secchi depth measurements, as Secchi depth includes algal-related turbidity. *Possible improvement of models* 

Investigation of possible non-linear trends in the relationship between measured chl in samples and Landsat inferred chl or measured Secchi depth and Landsat-inferred Secchi depth were not evident upon visual comparison of linear and LOESS models. However, testing a variety of boosted regression tree models indicated a third order trend, which agrees with previous chl models made for MODIS (O'Reilly et al. 2000). Analysis of national boosted regression tree chl and Secchi model residuals elucidated additional trends. The correlation between maximum depth and both chl and Secchi depth, and the partial correlation coefficients suggested that Landsat may detect differences in color intensity based on the volume of water. For example, it is possible that more chl is detected in deeper water bodies due to the increased volume of water for algae to inhabit. The fact

that depth explained so much variation in chl agrees with the idea that shallower lakes are more susceptible to algal blooms, due to lower phosphorus processing capacity and higher water temperatures. A similar explanation can be made for the correlation between national Secchi model residuals and lake depth. With increasing depth of a lake, there is more cumulative suspended matter throughout the water column that Landsat may be able to detect, but that is unaccounted for in field measurements that only represent a small portion of the water column.

The weak, positive and negative correlation of cyanobacterial density and boosted regression tree model residuals of the Landsat chl and Secchi depth models, respectively, was likely driven by differences in the taxonomic composition of algal communities, which affects algal color because of accessory pigments, such as phycocyanin. Alternatively, this positive correlation could be due to different cell mass/volume relationships of taxa (Morel and Bricaud 1981, Gitelson et al. 2000, Wetzel 2001). The fact that there was a correlation with cyanobacterial density indicates that Landsat's utility for cyanobacteria density prediction should be tested.

The weak, positive and negative correlation of DOC to national boosted regression tree chl and Secchi models, respectively, indicates that the varying levels of reddish-brown color reflected by DOC may be introducing noise into the signals reaching Landsat. For the chl model, the positive correlation suggests that the reflectance in the red band that DOC creates may overwhelm the subtle absorption changes we expect to see detecting chl.

A weak, positive relationship between Secchi boosted regression tree model residuals and lake area may indicate that surface area-related ecological attributes were not all account for in the model. This is also suggested by the lack of significance in the

relationship between chl boosted regression tree model residuals and lake area. Lack of correlation between diff. days and boosted regression tree model residuals indicates that number of days between sampling and image acquisition was not a factor in ability to predict chl and Secchi. We conclude that 8-day differences in sampling and imagery had little effect on model development.

Percent cloud cover was not related to either boosted regression tree chl or Secchi model residuals. Non-normality of percent cloud data, even after transformation, could have contributed to this result, however, precautions in choosing pixels for data extraction could have prevented any major cloud interference in chl and Secchi prediction. *Robustness and application of national models* 

Since each Landsat 7 ETM+ band provides different biological or physical information about the ecosystem, it is not surprising that the best boosted regression tree models included all of the visible bands in various combinations as well as infrared bands. The visible bands represent pigments or colors of matter suspended near the water's surface, infrared bands detecting biological structure, and thermal infrared detecting differences in temperature. Although there are differences in the bands and ratios in national models compared to those included in the lake type models (Tables 2-4 and 2-5), the similarities in the linear regression models agree with previous studies (Table 2-7) that visible bands and sometimes infrared bands alone or in ratios are important for chl and Secchi depth inference.

The boosted regression tree models, but especially the linear regression national and lake type models emphasize the importance of the visible bands as part of the chl and Secchi depth models. As there is variation in the bands used in the literature, there is also

variation in exactly which bands used in our national and lake type models. Such variation probably reflects the variety of algal species or inorganic suspended sediments in the water in those areas. For example, Lake type D includes relatively shallow lakes, which might become stratified more quickly. Lake type A, whose lakes are not particularly shallow, but whose temperature (similarly to lake type D) is significantly higher than other lakes types might also stratify more quickly, allowing high temperature tolerant phytoplankton species to thrive, and change overall pigment content or amount of floating biomass on the water's surface (Gitelson et al. 2000). Lake type E has a much larger range of DOC than other lake types which may flood the more subtle changes in red wavelength absorptance and reflectance of chl or colors of various sediment types. Lake depth, itself, may affect which bands/ratios are most useful in detecting chl due to change in color at different depths (Provost et al. 2004). In short, differences in algal pigments, DOC, and sediment types may explain variation in bands and ratios for each lake type.

Another reason for the variety of models is that the spectral signatures of certain pigments or particular matter may lie at the edge of Landsat wavelength bands. For example, Gitelson et al. (2000) found that chl has a subtle absorption feature starting around 690 nm, which coincides with the upper edge of the red Landsat 7 ETM+ detection band (Gitelson et al. 2000). In addition, algae have many accessory pigments with absorption peaks varying across the visible spectrum. Slight differences in environmental conditions can alter algal physiology, accessory pigment concentration, and color of the algae (Reynolds 2006).

Summary and concluding remarks

The national boosted regression tree models to infer chl and Secchi depth using Landsat imagery performed well and shows substantial potential for use in ecological studies. Confounding environmental variables limited the precision of national models, but future studies could incorporate natural landscape feature that can help predict the confounding variables. Refinement of national or regional-scale remotely sensed water quality might include accounting for different algal types, or wind speed near the lake. Other refinements may also be possible with the launch of the Landsat 8 (Landsat Data Continuity Mission) that occurred in February, 2013. Landsat 8 includes a second blue wavelength band, a new shortwave infrared band, and a narrower near infrared band that better cover the absorption features of chl (USGS 2011a). Continued refinement of remote sensing tools for characterizing water quality will be valuable for water resource management. APPENDIX

Landsat bands	Band wavelength (nm)	Resolution (m)
Blue	450-520	30
Green	520-600	30
Red	630-690	30
Near Infrared	770-900	30
Short Wave Infrared	1550-1750	30
Thermal Infrared	10400-12500	60
Short Wave Infrared	2090-2350	30

 Table 2-1: Landsat 7 ETM+ wavelength band features (USGS 2011a).

**Table 2-2: US EPA National Lakes Assessment lake types.**Lake types, based on clusteranalysis of US EPA National Lakes Assessment water chemistry.

Lake type	Lake type description
A	Eastern highlands: Low to moderate-elevation,
	warm reservoir
В	Eastern highlands: Low to moderate-elevation,
	northern lakes and reservoirs
С	Northern central plains and lowlands: Low-
	elevation lakes and reservoirs with small surface
	area
D	Southern and coastal plains and lowlands: Low-
	elevation, shallow, warm lakes and reservoirs with
	large surface area
Е	Northern central plains and lowlands: Low to
	moderate-elevation, shallow lakes and reservoirs
F	North-western mountains: Moderate to high-
	elevation, cold, deep lakes and reservoirs
G	South-western mountains: Moderate to high-
	elevation, cold lakes and reservoirs

**Table 2-3: National boosted regression tree chl and Secchi depth models.** National boosted regression tree models for ln-transformed chl and untransformed Secchi depth made with Landsat 7 ETM+ bands and ratios. Units of RMSE are ln-transformed  $\mu$ g/L for chl and m for Secchi, CV=cross-validation, NT=number of trees in final BRT model, LR=BRT learning rate, BF=bag fraction, TC=tree complexity.

		Training						
Model	Ν	R <sup>2</sup>	RMSE	CV R <sup>2</sup>	NT	LR	BF	ТС
Ln Chl	447	0.78	0.76	0.44	3050	0.005	0.50	3
Secchi	447	0.87	0.80	0.52	1800	0.008	0.50	3

Table 2-4: National and lake type linear regression chl detection models. National and lake type chl detection models made with Landsat 7 ETM+ bands and ratios. All results for national and lake type models were statistically significant at p $\leq$  0.05. Both national models using band ratio 1/2 and 1/3 were tested on each lake type, consecutively. Units of RMSE are ln-transformed µg/L, LT = lake type, NS = not statistically significant, NA = not applicable.

Model	N	Best model	Adj R <sup>2</sup>	RMSE	Adj R² Bootstrap	Nat. Model on LT RMSE
Nat.	447	1/2 1/3	0.18 0.19	1.39 1.38	0.19 0.20	NA
А	44	2/3 3/1 3/2	0.14 0.14 0.15	1.38 1.38 1.37	0.15 0.15 0.16	1.46, 1.40
В	48	1/2 2/1	0.31 0.31	1.17 1.16	0.33 0.33	1.17, 1.26
С	102	1/2	0.11	1.00	0.12	1.01, 1.04
D	44	1/2 7/5	0.09 0.09	0.94 0.94	0.10 0.09	1.60 1.52
Е	102	4/1	0.24	1.31	0.25	1.54
F	45	6	0.14	1.04	0.15	1.55, 1.60
G	62	1/3	0.17	1.17	0.18	1.47, 1.40

Model	Chl BR	Chl coeff.	Chl int.	Secchi BR	Secchi coeff.	Secchi int.
National	1/2	-3.64	7.21	1/3	1.58	-3.04
	1/3	-1.42	5.18			
А	2/3	-3.28	7.19	3/1	-5.62	3.15
	3/1	6.54	-1.06			
	3/2	7.50	-2.80			
В	1/2	-4.31	8.12	2/1	-6.61	5.31
	2/1	9.35	-4.68			
С	1/2	-2.51	5.65	1/3	1.38	-2.56
D	1/2	-2.20	6.53	1/3	0.90	-2.30
	7/5	-1.62	4.59			
Е	4/1	3.26	1.36	1/3	1.53	-3.29
				3/1	-4.28	1.95
F	6	0.14	-1.78	2/1	-8.04	6.77
G	1/3	-1.12	3.92	3/1	-4.56	2.90

**Table 2-5: National and lake type linear regression chl and Secchi model coefficients.**BR=band ratios., coeff. = coefficients, int.= intercept.

**Table 2-6: National and lake type linear regression Secchi depth detection models.** National and lake type Secchi depth detection models made with Landsat 7 ETM+ band ratios. All results for national and lake type models were statistically significant at  $p \le 0.05$ . Units of RMSE are ln-transformed meters. LT = lake type, NA = not applicable.

		<b>D</b>				Nat. Model
		Best	Adj		Adj R <sup>2</sup>	on LT
Model	N	model	R <sup>2</sup>	RMSE	Bootstrap	RMSE
Nat.	447	1/3	0.49	0.77	0.49	NA
А	44	3/1	0.36	0.66	0.36	5.19
В	48	2/1	0.40	0.69	0.41	6.44
С	102	1/3	0.47	0.55	0.47	4.91
D	44	1/3	0.34	0.41	0.34	4.12
E	102	1/3 3/1	0.35 "	0.77 "	0.35 "	3.38
F	45	2/1	0.65	0.53	0.63	7.26
G	62	3/1	0.61	0.75	0.60	4.38

**Table 2-7: Chlorophyll and Secchi depth models from the literature.** Previous studies using various satellites and bands or band ratios to infer chl and Secchi depth. Dashes indicate duplicate information from the line above.

Year	Authors	Satellite	Geographic extent	Bands used	Chl/Secchi	R <sup>2</sup>
1983	Lillesand et al.	Landsat MSS	Twenty-eight Minnesota lakes	Green, red, near- infrared	Chl	0.84
				Green, red, near- infrared	Secchi	0.88- 0.94
1983	Carpenter and Carpenter	Landsat MSS	Lake Burley Griffin, Australia	Green, red, near- infrared	Chl	0.50- 0.85
1989	Lathrop and Lillesand	Spot-1	Lake Michigan, USA	Near- infrared	Secchi	0.83
1991	Khorram et al.	Landsat TM	August Bay, Italy	Blue, green	Chl	0.84
				Blue, green	Secchi	0.83
1992	Mittenzwey et al.	Ship-based spectrometer	Several lakes and rivers in Germany	Two red bands	Chl	0.98
1994	Pattiaratchi et al.	Landsat TM	Cockburn Sound, Australia	Blue, green, red	Chl	0.73- 0.77

Year	Authors	Satellite	Geographic extent	Bands used	Chl/Secchi	R <sup>2</sup>
				Red	Secchi	0.72- 0.78
1995	Gitelson et al.	Landsat TM and ship-based radiometer	Haifa Bay, Israel	Blue, green, red	Chl	0.37- 0.93
1995	Mayo et al.	Landsat TM	Lake Kinneret, Israel	Blue, green, red	Chl	0.49
2000	Gitelson et al.	LI-1800, Ocean Optics ST1000, ASD spectrometers	Several water bodies in Israel and several lakes in Iowa, USA	Two red bands	Chl	083- 0.96
2001	Brivio et al.	Landsat TM	Lake Garda, Italy	Blue, green, red	Chl	0.68- 0.82
2001	Giardino et al.	Landsat TM	Lake Iseo, Italy	Blue, green	Chl, Secchi	
2001	Pulliainen et al.	Airborne imaging spectrometer	Eleven lakes in Finland	Two red bands	Chl	0.41- 0.97
2002b	Kloiber et al.	Landsat MSS	Five hundred Minnesota lakes	Blue, green	Secchi	0.60- 0.79

# Table 2-7 (cont'd)

Year	Authors	Satellite	Geographic extent	Bands used	Chl/Secchi	R <sup>2</sup>
		Landsat TM	Five hundred Minnesota lake	Blue, red	Secchi	0.72- 0.93
2003	Nelson et al.	Landsat ETM+	State of Michigan	Blue, red	Secchi	0.43- 0.82
2004	Chipman et al.	Landsat TM and ETM+	State of Wisconsin	Blue, red	Secchi	0.42- 0.88
2005	Brezonik et al.	Landsat TM	Fifteen Minnesota lakes	Blue, green, red, near infrared	Chl	0.88
2008	Kabbara et al.	Landsat ETM+	Coast of Tripoli	Blue, green, red	Chl	0.72
				Blue, green	Secchi	0.54
2008	Olmanson et al.	Landsat MSS, TM, and ETM+	State of Minnesota	Blue, red	Secchi	0.71- 0.95
2009	Nas et al.	ASTER	Lake Beysehir, Turkey	Green, red, near- infrared	Chl	0.86

# Table 2-7 (cont'd)

Table 2-7 (cont'd)							
2010	Mishra and Mishra	MODIS	Lake Pontchatrain, USA	Green, red	Chl	0.65	



Figure 2-1: Histograms of untransformed and In-transformed chlorophyll and Secchi depth.



Measured Chlorophyll In(µg/L)

Figure 2-2: Remotely sensed (linear regression predicted) ln-transformed chlorophyll a using band ratio 1/3 versus measured ln-transformed chlorophyll a (R<sup>2</sup>adj=0.19, 1.38 ln-transformed μg/L). Line indicates model's fit.



Measured Chlorophyll In(µg/L)

Figure 2-3: Remotely sensed (boosted regression tree predicted) natural lntransformed chlorophyll versus measured ln-transformed chlorophyll a (R<sup>2</sup>adj=0.44, 0.76 ln-transformed μg/L). Line indicates model fit.



Figure 2-4: Remotely sensed (linear regression predicted) natural ln-transformed Secchi depth using band ratio 1/3 versus measured ln-transformed Secchi depth. (R<sup>2</sup>adj=0.49, 0.77 ln-transformed m RMSE). Line indicates model's fit.



**Figure 2-5: Remotely sensed (boosted regression tree predicted) Secchi depth versus measured untransformed Secchi depth (R<sup>2</sup>adj=0.52, 0.80 m RMSE).** Line indicates model's fit.

LITERATURE CITED

# LITERATURE CITED

- Allee, R. J. and J. E. Johnson. 1999. Use of satellite imagery to estimate surface chlorophyll a and Secchi disc depth of Bull Shoals Reservoir, Arkansas, USA. International Journal of Remote Sensing **20**:1057-1072.
- Baban, S. M. J. 1993. Detecting water quality parameters in the Norfolk Broads, U.K., using Landsat imagery. International Journal of Remote Sensing **14**:1247-1267.
- Backer, L. C., S. V. McNeel, T. Barber, B. Kirkpatrick, C. Williams, M. Irvin, Y. Zhou, T. B. Johnson, K. Nierenberg, M. Aubel, R. LePrell, A. Chapman, A. Foss, S. Corum, V. R. Hill, S. M. Kieszak, and Y.-S. Cheng. 2010. Recreational exposure to microcystins during algal blooms in two California lakes. Toxicon 55:909-921.
- Briand, J.-F., S. Jacquet, C. Bernard, and J.-F. Humbert. 2003. Health hazards for terrestrial vertebrates from toxic cyanobacteria in surface water ecosystems. Veterinary Research **34**:361-377.
- Brivio, P. A., C. Giardino, and E. Zilioli. 2001. Determination of chlorophyll concentration changes in Lake Garda using an image-based radiative transfer code for Landsat TM images. International Journal of Remote Sensing **22**:487-502.
- Bustamante, J., F. Pacios, R. Diaz-Delgado, and D. Aragonés. 2009. Predictive models of turbidity and water depth in the Doñana marshes using Landsat TM and ETM+ images. Journal of Environmental Management **90**:2219-2225.
- Byappanahalli, M. N., R. Sawdey, S. Ishii, D. A. Shively, J. A. Ferguson, R. L. Whitman, and M. J. Sadowsky. 2009. Seasonal stability of *Cladophora*-associated *Salmonella* in Lake Michigan watersheds. Water Research **43**:806-814.
- Carmichael, W. W. 1996. Toxic *Microcystis* in the environment. Pages 1-11 *in* M. F. Wantabe, K. H. Harada, W. W. Carmichael, and H. Fujiki, editors. Toxic Microcystis. CRC Press, Boca Raton, FL.
- Carpenter, D. J. and S. M. Carpenter. 1983. Modeling inland water quality using Landsat data. Remote Sensing of Environment **13**:345-352.
- Carpenter, S. R., N. F. Caraco, D. L. Correll, R. W. Howarth, A. N. Sharpley, and V. H. Smith. 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecological Applications **8**:559-568.

- Chander, G., B. L. Markham, and D. L. Helder. 2009. Summary of current radiometeric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. Remote Sensing of Environment **113**:893-903.
- Chipman, J. W., T. M. Lillesand, J. E. Schmaltz, J. E. Leale, and M. J. Nordheim. 2004. Mapping lake water clarity with Landsat images in Wisconsin, U.S.A. Canadian Journal of Remote Sensing **30**:1-7.
- Cox, R. M., R. D. Forsythe, G. E. Vaughan, and L. L. Olmsted. 1998. Assessing water quality in Catawba River reservoirs using Landsat Thematic Mapper satellite data. Lake and Reservoir Management **14**:405-416.
- Duan, H., Y. Zhang, B. Zhang, K. Song, and Z. Wang. 2007. Assessment of chlorophyll-a concentration and trophic state for Lake Chagan using Landsat TM and field spectral data. Environmental Monitoring and Assessment **129**:295-308.
- Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. Journal of Animal Ecology **77**:802-813.
- Giardino, C., M. Pepe, P. A. Brivio, P. Ghezzi, and E. Zilioli. 2001. Detecting chlorophyll, Secchi disk depth and surface temperature in a sub-alpine lake using Landsat imagery. The Science of the Total Environment **268**:19-29.
- Gitelson, A., G. Garbuzov, F. Szilagyi, K.-H. Mittenzwey, A. Karnieli, and A. Kaiser. 1993. Quantitative remoe sensing methods for real-time monitoring of inland waters quality. International Journal of Remote Sensing **14**:1269-1295.
- Gitelson, A., Y. Z. Yacobi, J. F. Schalles, D. C. Rundquist, L. Han, R. Stark, and D. Etzion. 2000. Remote estimation of phytoplankton density in productive waters. Archiv für Hydrobiologie Special Issues in Advanced Limnology **55**:121-136.
- Han, L. and K. J. Jordan. 2005. Estimating and mapping chlorophyll-a concentration in Pensacola Bay, Florida using Landsat ETM+ data. International Journal of Remote Sensing **26**:5245-5254.
- Han, L., D. C. Rundquist, L. L. Liu, R. N. Fraser, and J. F. Schalles. 1994. The spectral response of algal chlorophyll in water with varying levels of suspended sediment. International Journal of Remote Sensing 15:3707-3718.
- Herlihy, A. T., N. C. Kamman, J. C. Sifneos, D. Charles, M. D. Enache, and R. J. Stevenson. 2013. Using multiple approaches to develop nutrient criteria for lakes in the conterminous USA. Freshwater Science **32**:367-384.
- Ishii, S., T. Yan, D. A. Shively, M. N. Byappanahalli, R. L. Whitman, and M. J. Sadowsky. 2006. *Cladorphora* (Chlorophyta) spp. harbor human bacterial pathogens in nearshore water of Lake Michgian. Applied and Environmental Microbiology **72**:4545-4553.

- Jensen, J. R. 2005. Introductory Digital Image Processing. A Remote Sensing Perspective. Prentice Hall, Upper Saddle River, New Jersey.
- Kloiber, S. M., P. L. Brezonik, and M. E. Bauer. 2002a. Application of Landsat imagery to regional-scale assessments of lake clarity. Water Research **36**:4330-4340.
- Kloiber, S. M., P. L. Brezonik, L. G. Olmanson, and M. E. Bauer. 2002b. A procedure for regional lake water clarity assessment using Landsat multispectral data. Remote Sensing of Environment **82**:38-47.
- Lathrop, R. G., T. M. Lillesand, and B. S. Yandell. 1991. Testing the utility of simple multidate Thematic Mapper calibration algorithms for monitoring turbid inland waters. International Journal of Remote Sensing **12**:2045-2063.
- Lavery, P., C. Pattiaratchi, A. Wyllie, and P. Hick. 1993. Water quality monitoring in estuarine waters using the Landsat Thematic Mapper. Remote Sensing of Environment **46**:268-280.
- Lillesand, T. M., W. L. Johnson, R. L. Deuell, O. M. Lindstrom, and D. E. Meisner. 1983. Use of Landsat data to predict the trophic state of Minnesota Lakes. Photogrammetric Engineering and Remote Sensing **49**:219-229.
- Mayo, M., A. Gitelson, Y. Z. Yacobi, and Z. Ben-Avraham. 1995. Chlorophyll distribution in Lake Kinneret determined from Landsat Thematic Mapper data. International Journal of Remote Sensing **16**:175-182.
- Mishra, D. R. and S. Mishra. 2010. Plume and bloom: Effect of the Mississippi River diversion on the water quality of Lake Pontchartrain. Geocarto International **25**:555-568.
- Nelson, S. A. C., P. A. Soranno, K. S. Cheruvelil, S. A. Batzli, and D. L. Skole. 2003. Regional assessment of lake water clarity using satellite remote sensing. Journal of Limnology **62**:27-32.
- NOAA. 2011. Solar Position Calculator National Oceanic and Atmospheric Administration.
- O'Reilly, J. E., S. Maritorena, M. C. O'Brien, D. A. Siegel, D. Toole, D. Menzies, R. C. Smith, J. L. Mueller, B. G. Mitchell, M. Kahru, F. P. Chavez, P. Strutton, G. F. Cota, S. B. Hooker, C. R. McClain, K. Carder, F. Müller-Karger, L. Harding, A. Magnuson, D. Phinney, G. F. Moore, J. Aiken, K. R. Arrigo, R. Letelier, and M. Culver. 2000. SeaWiFS postlaunch calibration and validation analyses, Part 3.*in* S. B. a. E. R. F. Hooker, editor. SeaWiFS Postlaunch Techinical Report Series. NASA.

- Olmanson, L. G., M. E. Bauer, and P. L. Brezonik. 2008. A 20-year Landsat water clarity census of Minnesota's 10,000 lakes. Remote Sensing of Environment **112**:4086-4097.
- Paerl, H. W. and J. Huisman. 2008. Blooms like it hot. Science **320**:57-58.
- Provost, J.-N., C. Collet, P. Rostaing, P. Pérez, and P. Bouthemy. 2004. Hierarchial Markovian segmentation of multispectral images for the reconstruction of water depth maps. Computer Vision and Image Understanding **93**:155-174.
- Reynolds, C. S. 2006. Ecology of Phytoplankton. Cambridge University Press, Cambridge, United Kingdom.
- Scarpace, F. L., K. W. Holmquist, and L. T. Fisher. 1979. Landsat analysis of lake quality. Photogrammetric Engineering and Remote Sensing **45**:623-633.
- Strong, A. E. 1974. Remote sensing of algal blooms by aircraft and satellite in Lake Erie and Utah Lake. Remote Sensing of Environment **3**:99-107.
- Torbick, N., F. Hu, J. Zhang, J. Qi, H. Zhang, and B. Becker. 2008. Mapping chlorophyll-a concentrations in West Lake, China using Landsat 7 ETM+. Journal of Great Lakes Research **34**:559-565.
- USEPA. 2009. National Lakes Assessment: A collaborative survey of the nation's lakes.*in* O. o. W. a. O. o. R. a. D. Environmental Protection Agency, editor., Washington, D.C.
- USGS. 2011a. Landsat Missions. United States Geological Survey.
- USGS. 2011b. US Geological Survey Gloval Visualization Viewer United States Geological Survey.
- Watson, S. B., J. Ridal, and G. L. Boyer. 2008. Taste and odour and cyanobacterial toxins: impairment, prediction, and management in the Great Lakes. Canadian Journal of Fisheries and Aquatic Sciences **65**:1779-1796.
- Wetzel, R. G. 2001. Limnology: Lake and Reservoir Ecosystems. Third edition. Academic Press, New York, New York.

# CHAPTER 3 MODIS AND LANDSAT BOOSTED REGRESSION TREE CHLOROPHYLL PREDICTIVE MODELS FOR THE GREAT LAKES

#### Abstract

Numerous models for predicting chlorophyll (chl) based on satellite images have been tested on the Great Lakes, and some have been built specifically for the Great Lakes. However many of these models either lack the ability to predict chl or are technically complicated or expensive to employ. Also, few studies have taken advantage of archived satellite images to assess spatial and temporal dynamics over chlorophyll during past decades. Boosted regression tree (BRT) models of chl using satellite imagery are both easy to use and can have high predictive performance. Landsat and MODIS imagery are complementary platforms because of the long history of Landsat operation and the finer spectral resolution and image frequency of MODIS. Three BRT models were created: 1) a MODIS model; 2) a Landsat ETM+ and TM model; and 3) a Landsat MSS model. MODIS BRT model predicted chl most accurately of the three models and compared well to other models in the literature. BRT models for Landsat ETM+ and TM more accurately predicted chl than the MSS model and all Landsat models had favorable results when compared to the literature. BRT chl predictive models are useful in helping to understand historical, longterm chl trends and to inform us of how climate change may alter ecosystems in the future.

# Introduction

Early water quality assessments were done by physically taking samples. However, advancements in remote sensing technology and development of modeling tools in recent decades have allowed ecological assessment from a distance, enabling evaluation from larger spatial and temporal scales and providing a more comprehensive view at potentially

less cost. Historically, the two main approaches to remote sensing chl predictive model development in the Great Lakes have been: 1) deriving empirical relationships between chl and satellite bands or band ratios, alone or separately from other color producing agents using linear or polynomial regression analyses (Carpenter and Carpenter 1983, O'Reilly et al. 1998, Giardino et al. 2001) or 2) deriving relationships between two or more color producing agents (including chl) at once (Carder et al. 2004, Pozdnyakov et al. 2005). Many of the current chl models are either not widely applicable, have not been tested for wide use, or are challenging to implement without expert statistical and programming knowledge, or without costly computer programs. One model-building approach that has potential to alleviate these challenges is boosted regression tree (BRT) analyses, which is relatively easy and fast to run with free software, the R program. Code for models can be easily shared and applied with models to output from standard remote sensing software.

BRT analysis is a machine learning approach for building models (De'ath 2007, Elith et al. 2008) with advantages of both regression trees and stochastic boosting. In regression tree analysis, models are fit by recursive partitioning of the response variable, such as chl, into the two most homogenous groups possible at each split in a way that maximally reduces error in chl prediction. Regression trees are able to simultaneously handle predictor variables of different types and scales, they can function in the presence of outliers and missing data, they can model non-linear relationships and interactions between predictors, and they do not require that the response variable is normally distributed (Elith et al. 2008). Boosting is the process of iteratively fitting the regression tree models to explain the largest error in residuals, in contrast to common modeling approaches where only one best model is created and chosen (Friedman 2002). Stochastic

boosting can decrease overfitting the data and increase predictive power in the final model by randomly selecting subsets of training data to use in each stage of the tree fitting process. The combination of regression tree analyses and boosting further increases predictive power (Elith et al. 2008). In addition, BRT analysis is relatively easy to perform and apply to satellite imagery. Results of BRT models can be used alone or in conjunction with other models depending on the study goals. While BRT models often have increased predictive performance, they can be harder to interpret due to their complexity, causing them to be more "black box" in nature when compared to simple linear or polynomial regression models. For both increased predictive performance and interpretability, a multimodel approach could be employed, combining BRTs and more simple regression models.

Our research objectives were to develop and test boosted regression tree (BRT) predictive models for chl in the Laurentian Great Lakes using Landsat and MODIS imagery and chl measured in water samples from 2007-2009 by the Great Lakes National Program Office. Landsat and MODIS imagery are complementary platforms due to the fine spatial resolution and long history of Landsat operation and the finer spectral resolution and image frequency of MODIS. To take advantage of the strengths of both satellites, we developed separate models for Landsat and MODIS. A total of two models were made for Landsat due to significant differences in number and location of spectral bands of Landsat ETM+ and TM compared to the spectral bands of Landsat MSS. The three models developed include: 1) a MODIS model; 2) a Landsat ETM+ and TM model; and 3) a Landsat MSS model. In this effort, all MODIS and Landsat bands and band ratios, including infrared and thermal, were used in model development. Models were validated by k-fold cross validation. Root

mean square error (RMSE) was the final determinant of model function. BRT models were created and compared to algorithms from the literature. BRT models were applied to historical Landsat (1972-2012) and MODIS (2000-2012) images for long-term analysis of chl concentration in the Great Lakes. Historical trend data can also help elucidate how algal biomass may change in the Great Lakes due to climate change.

#### Methods

#### Study locations and ecological data

Study locations include the five Great Lakes, with a focus on three targeted areas: Saginaw Bay (43°55'N 83°35'W), Grand Traverse Bay (43°3'53.24"N, 83°29'9.56"W), and the eastern shore of Lake Michigan, near the Grand River mouth (43°3'47"N 86°13'42"). Great Lakes chl data were collected by the Great Lakes National Program Office (GLNPO) in August and September of 2007-2009 across the Great Lakes. Data for the MODIS model ranged from 7.2-306.1 m in sample station depth and 0.31-12.64  $\mu$ g/L chl from 226 chl observations. Data for the Landsat ETM+ and TM model and the Landsat MSS model ranged from 7.9-306.1 m in sample station depth and 0.29-12.64  $\mu$ g/L chl from 177 chl observations.

# MODIS satellite image processing and pixel extraction for BRT model training

MOD021km (1 km resolution) images, available free of charge, were downloaded from http://ladsweb.nascom.nasa.gov (Masuoka and Horrocks 2010) and converted to spectral reflectance using the MODIS Conversion Toolkit (MCTK). Images were chosen based on proximity of image date to GLNPO sample date. When sampling stations were located on pixels covered by clouds or cloud shadows, an image from a previous or subsequent day would be assessed. This process continued until a cloud-free pixel was

found. Number of days between image capture and sample collection ranged from 0-6 days. A shapefile containing GLNPO chl sample stations was created in ArcMap 9.3 and used in conjunction with MODIS images in Erdas Imagine 9.3 to extract pixels values closest to the sample station from bands 1-16 (for a total of 18 bands, with bands 13 and 14 each having a low and high gain bands, Table 3-1).

#### Landsat image processing and pixel extraction for BRT model training

Landsat 7 ETM+ (Table 3-2) images with level L1T processing were downloaded from the US Geological Survey (USGS) Global Visualization Viewer website: http://glovis.usgs.gov/ (USGS 2013c) for dates as close as possible to when lakes were sampled. Exact location of lakes was determined by overlaying a shapefile (made in Arc Map 9.3) of GLNPO sampling sites on each Landsat image. When sampling stations were located on pixels covered by clouds or cloud shadows, an image from a previous or subsequent image would be assessed. This process continued until a cloud-free pixel was found. Number of days between image date and sample date ranged from 0-19 days for Landsat training data. Due to mechanical malfunction, some ETM+ images contain black stripes in areas causing loss of information. If a sample location coincided with a black stripe, a nearby pixel (when possible in the same row or column as the sample pixel) was selected. The same method was used when sample locations were located in a mixed pixel that had both water and land signatures.

Radiometric calibration, provided by conversion to top of the atmosphere (TOA) reflectance values, is necessary to account for image differences in sun angle, in Earth-Sun distance, and in solar irradiance for different wavelengths that occur above the earth's atmosphere (Chander et al. 2009). Therefore, after pixel extraction, raw pixel values were

converted to spectral radiance values with the following equation:

Radiance = 
$$\frac{(LMAX - LMIN)}{255}$$
 (pixel digital number) + LMIN

Where pixel digital number is the raw pixel value, and LMAX and LMIN correspond to known spectral radiances for each of the Landsat 7 ETM+ spectral bands (Chander et al. 2009). Radiance values were further converted to TOA reflectance values with this equation:

$$P_p = \frac{(\Pi)(L_{\lambda})(d^2)}{(E_{\lambda})(\cos\theta_s)}$$

Here  $P_p$  is unitless planetary reflectance, L is spectral radiance at the sensor's aperture, d is the earth-sun distance in astronomical units based on the image acquisition date (Chander et al. 2009), and  $E_{\lambda}$  is mean solar exoatmospheric irradiance. Solar zenith ((s) values were obtained from NOAA's Solar Position Calculator at:

http://www.esrl.noaa.gov/gmd/grad/solcalc/ (NOAA 2013). High gain thermal band radiance values were converted directly to temperature (°C) with this equation:

$$T = \frac{K2}{\left(\frac{K1}{L\lambda} + 1\right)} - 273.15$$

Where, T is the effective at-satellite temperature in Kelvin, K2 is calibration constant, 1282.71, K1 is calibration constant, 666.09, and L is the spectral radiance in watts/(m<sup>2</sup> µm steridians)(Chander et al. 2009).

If the TOA value was 0.3 or greater, pixel values for that lake were excluded from further analyses due to the likelihood of nearby cloud interference. No additional atmospheric correction was done due to lack of information on atmospheric conditions for the sampling date and location.

#### Data preparation and statistics for MODIS and Landsat data

To avoid the effects of lake bottom reflectance, stations for which depth was less than twice their Secchi depth were removed from the training data set before performing statistical analyses. For stations where Secchi was not available, it was estimated using Carlson's (1977) Secchi-chl relationship:

$$Ln(Secchi) = -0.68 * Ln(chl) + 2.04$$

Pixel value data were analyzed in separate BRT analyses for MODIS, for Landsat ETM+ and TM, and for Landsat MSS. For the Landsat MSS model, three of the ETM+ bands from the training data were used that corresponded most closely with MSS bands (Table 3-3). All statistics were run in gbm package 2.9 in R 2.15.1. For each BRT, all available visible and infrared bands, including thermal infrared were used, along with all possible band ratios. A total of 49 bands and ratios were used in BRT analyses for Landsat ETM+ and TM, ten bands and ratios were used for Landsat MSS, and 324 were used for MODIS. Although including inverse ratios (e.g., 1/3 and 3/1) in the same model would seem to introduce redundancy into the model, inverse ratios have different rates of change along the same band ranges, and therefore have different weights in high and low ranges of band ratios. Also, since each waveband covers a slightly different part of the electromagnetic spectrum and because chl detection is influenced by algal community structure and time of day sampled (Owens et al. 1980, Lebouteiller and Herbland 1982, Reynolds 2006), even bands in the same general category, such as "red bands," can give us additional information about the presence of algae.

Bag fractions (the percent of data taken from the training data set, at random) for

every BRT model were set to 0.5 and, learning rate, which controls the contribution of each tree to the model (Elith et al. 2008), was set to 0.01 for every model. Initial BRT models were simplified by looking at the partial correlation plots of each band or ratio and measured chl. Final BRT models were chosen using 10-fold internal cross validation procedure, incorporated into the gbm package. Top bands or ratios were kept, the number of which was based on the total number of data rows, so that the ratio of observations (rows of data) to predictors (bands/ratios) was no lower than 10:1 (Peng et al. 2002, Babyak 2004), which reduced the chance of overfitting the data. Root mean square error (RMSE) values, calculated between predicted and measured chl, were compared for final model selection. For the final MSS BRT model, measured (not remotely sensed) water depth was used as a predictor variable in conjunction with Landsat bands and ratios for better chl predictive performance.

## MODIS image processing for model application

MODIS MOD02L2 images (1km resolution) were downloaded

http://ladsweb.nascom.nasa.gov (Masuoka and Horrocks 2010), covering July 20 -September 6, 2000 to 2012 across the Great Lakes. Level 1B of MOD02L2 images were selected because they contain calibrated and geolocated at-aperture radiances for bands 1-16, generated from MODIS Level 1A sensor counts (MOD01). Spectral radiance, in units of watts/(m<sup>2</sup> μm steridians), was calculated for MOD02L2 bands 1-16 using the MODIS Conversion Toolkit (MCTK) and the images were mosaicked in the same date to generate the daily images at 1 km spatial resolution.

MODIS MOD09A1 images (500 m resolution) were also downloaded http://e4ftl01.cr.usgs.gov/MOLT (USGS 2013b), covering July 20 - September 6, 2000 to
2012 across the Great Lakes. After processing, MOD09A1 images were reprojected, and converted to the GeoTIFF format using the MODIS Reprojection Tool (MRT) to provide the day of year (DOY) and cloud images. Since MOD09A1 images only include bands 1-7, we did not use the MOD09A1 8-day composite image product for our study. Instead, metadata from MOD09A1 images were used to create 8-day composite images from MOD02L2 images. We assumed that the best possible TOA reflectance of MOD02L2 bands 1-16 during 8-day composite periods was the same as for the MOD09A1 product. For each pixel, the best possible observation of spectral radiance obtained during the 8-day composite period was assumed to be the DOY provided by MOD09A1. The criteria for selecting the DOY of MOD09A1 is based on high observation coverage, low view angle, absence of clouds, shadow, and aerosol loading, and minimum value of blue band (Toller et al. 2006). This compositing method minimized cloud effect as well as view angle, bi-directional effects, and other atmospheric effects. DOY and corresponding name of the final MOD02L2 8-day composite images, therefore, represent the first day of the 8-day period when the best observation is recorded. Cloud masking information, also found in the MOD09A1 product, was used to perform cloud masking of 8-day composite MOD02L2 images before applying the BRT chl model.

## Landsat image processing for model application

Landsat 7 ETM+, Landsat 4 and 5 TM, and Landsat 1-3 MSS images (Tables 3-2 and 3-3) with level L1T or L1G processing were downloaded from the US Geological Survey (USGS) Global Visualization Viewer website: <u>http://glovis.usgs.gov/</u> (USGS 2013c) or the USGS Earth Explorer website: <u>http://earthexplorer.usgs.gov/</u> (USGS 2013a). Landsat images were processed from digital number to TOA reflectance (Chander et al. 2009) using

Python 2.6. Images downloaded at L1G processing were georectified using four control points with RMSE of less than 30 m. The reference layer was Landsat TM image with 0% cloud cover and quality of 9. Cloud mask algorithm (Oreopoulos et al. 2011) was performed for all Landsat TM and ETM+ images. This method was modified from a cloud/clear masking algorithm initially developed for MODIS clear-image compositing. This cloud masking technique applies a simple threshold scheme by selection four bands (band 1, 3, 4, and 5) in Landsat TM and ETM+ to remove cloud pixels.

#### Model evaluation

When possible, we tested models in the literature with our data to standardize the evaluation process (Lathrop and Lillesand 1986, O'Reilly et al. 2000, Lesht et al. 2013). When either the full algorithm was not available to test with our data (Shuchman et al. 2006, Becker et al. 2009) or we did not have measurements of other color producing agents to use with dual retrieval approaches (Binding et al. 2012), we based the performance of our model on reported R<sup>2</sup> or RMSE values from those models.

#### Results

#### BRT models and partial correlation plots

The final MODIS BRT model had 27 predictors, 0.10 µg/L RMSE and 0.85 crossvalidation R<sup>2</sup> (Table 3-4, Figure 3-1). Chl was accurately predicted through the range of measured chl values (Figure 3-1). As previously shown (Table 3-1), MODIS has multiple bands that fall into each general band category. Many of the bands were used in the final model, and some of those bands are from the same general band category. For example, 1/13H is a band ratio comprised of two red bands. MODIS band ratios with the most predictive power included blue/green, green/blue, red/red and red/blue, using bands 1,10, 11, and 13H (Figure 3-2). The MODIS BRT model had a lower RMSE and higher R<sup>2</sup> than

other MODIS Great Lakes chl models (Table 3-5).

The final Landsat ETM+ and TM BRT model had 17 predictors, 0.55 µg/L RMSE and 0.69 cross-validation R<sup>2</sup> (Table 3-4, Figure 3-3). Chl was accurately predicted between 0-8 µg/L chl, above which higher measured chl are inaccurately estimated to be approximately 8 µg/L. Bands and ratios with the most predictive power included short wave infrared (SWIR), green/blue, blue/green, thermal, and blue, using bands 1, 2, 6, and 7 (Figure 3-4). The Landsat ETM+ and TM BRT model had a lower RMSE and higher R<sup>2</sup> than other Landsat Great Lakes chl models (Table 3-5).

The final Landsat MSS BRT model had 10 predictors, 0.73  $\mu$ g/L RMSE and 0.66 cross-validation R<sup>2</sup> (Table 3-4, Figure 3-5). As with the ETM+ and TM model, chl is fairly accurately predicted up until about 8  $\mu$ g/L chl, above which measured chl are inaccurately predicted around 8  $\mu$ g/L. Variables that had the most predictive power in this model included depth, green, and red; bands 2 and 3 (Figure 3-6). This is the first Landsat MSS chl predictive model created for the Great Lakes.

Overall, MODIS was able to detect chl more accurately than Landsat, and Landsat ETM+ and TM models outperformed MSS models. Green/blue and blue/green band ratios were the most common predictors of chl, however the most powerful predictor of chl was different for each of the three models, including blue/green, SWIR, and depth (Figures 3-2, 3-4, 3-6, 3-8, 3-10). The blue/green and green/blue ratios were commonly found among the top predictors, except for the MSS models. MSS does not contain a blue band. Interestingly, the red band was a common variable in MODIS and MSS models, but not in the Landsat ETM+ and TM models.

#### Discussion

To date, our BRT models are the some of the most accurate chl predictive models for the Great Lakes (Table 3-5). Both MODIS and Landsat BRT models were able to predict chl well compared to Great Lakes models in the literature as indicated by R<sup>2</sup> and RMSE values (Table 3-5). Bands and ratios that are often used by models in the literature were also found among the top predictors in the BRT models, including blue-green ratios (O'Reilly et al. 2000, Lesht et al. 2013), red-green ratios (Lillesand et al. 1983), and red-blue ratios (Kabbara et al. 2008, Torbick et al. 2008). This was expected, since chl reflects light in the green wavelengths and has strong absorption features in the blue and red wavelengths. Infrared bands were also found in all BRT models, which may indicate that incorporation of physical features, such as algal scums on the surface of the water, is a relevant component to algal biomass detection. Predictive power of physical variables such as the thermal band and measured water depth in the Landsat models indicate that detection of chl can be improved by accounting for physical features of the sample location. Though in this study station depth was obtained by in-lake measurement, future values could be obtained by bathymetric map.

One notable difference between BRT models is the incorporation of multiple band of the same general category (for example, red bands), in the MODIS model. This is compatible with the idea of satellites with more numerous and finer bands being able to detect different amount of chl or chl that comingle with other algal pigments more accurately (Gower and Borstad 2004). Great Lakes algal communities span over six divisions that include variety of non-chl a pigments, including chlorophylls b, c, and d, as well as carotenoids, xanthophylls, and phycobilins (Vollenweider et al. 1974, Reynolds

2006). Finally, pixel size of the remote sensing image influences the spatial scale at which chl is being summarized and possibly which bands would be most useful in chl prediction. In areas of the Great Lakes where chl values are fairly consistent, a larger pixel size may be adequate. A smaller pixel size becomes more important in areas with highly variable chl concentrations throughout the summer.

Reasons that MODIS outperformed Landsat may include that MODIS is able to capture more subtle spectral features with greater temporal frequency (NASA 2013). Also, MODIS has more wavelength bands with which to build BRT models that have the potential to explain more color-related or physical attributes of water. As with MODIS to Landsat comparisons, the favorable performance of ETM+ and TM to MSS models may have to do, in part, with more numerous spectral bands, slightly finer spectral and temporal resolution.

Application of BRT models to the 40-year Landsat image archive can help elucidate changes in algal biomass across the Great Lakes for as long as the Clean Water Act has been in place, with high spatial resolution. These patterns can help discern regular, seasonal fluctuations from climate change, invasive species, and habitat alteration effects on algal biomass. With its high chl prediction accuracy, application of the MODIS BRT model to the 12-year MODIS image archive can give us additional information on the seasonal dynamics of algal biomass within each year. The notably high accuracy of the MODIS BRT model throughout the range of measured chl values indicates its applicability across the Great Lakes in areas with both somewhat high and low chl values. The Landsat models are less accurate at detecting higher chl values than MODIS, but can still capture algal conditions above 5  $\mu$ g/L. Therefore, application of the MODIS BRT models would be appropriate across the Great Lakes and at times when chl values range from 0-12  $\mu$ g/L. Landsat models

would be most appropriate for use in areas or during times of year where chl does not reach go much beyond 8  $\mu$ g/L. The advantages of Landsat and MODIS complement each other and enable future research and management programs to be focused on parts of the Great Lakes that are most vulnerable to stress, including climate change related impacts.

BRT predictive models could also be used conjunction with other chl models. As previously mentioned, a main advantage of BRT models is that they have high predictive accuracy. However, they are less easily interpretable than simple linear or polynomial algorithms, leading BRTs to be more "black box" in nature (Elith et al. 2008). If predictive accuracy and not interpretability is the objective, then the black box nature of the models is not as much of a drawback. However, if the study goal is both predictive accuracy and interpretability, then using a multimodel approach that incorporates results from BRT and other models may be best. With a multimodel approach one may gain benefits of both, for results that may be more accurate and more widely applicable than either model alone. Yet another approach to chl prediction might use different models in different parts of the Great Lakes. Where the dominant color producing agent is chl and concentrations are relatively low and constant, the ocean models may work well enough. In areas with higher or more variable chl, or with more reflective influence by the bottom in nearshore zones, other models may be more appropriate.

Chl BRT predictive models, used alone or in conjunction with other models, could be used to establish a standard method in chl prediction in the Great Lakes. Advantages to a well-tested standard method include being able to compare chl concentrations across broader spatial and temporal scales and efficient use of resources for the implementation and improvement of one model rather than production of additional models. Such efforts

can advance our understanding of ecosystem changes and allow managers to reliably assess and target regions of the Great Lakes that experience the largest algae-related stresses. Improving water quality in these parts of the Great Lakes would allow for recovery of habitat for aquatic species that are negatively impacted by algal bloom conditions that often include, low oxygen levels, presence of toxins, and low light levels (Wetzel 2001). People, too, would benefit from more reliable chl prediction in Great Lakes waters, to ensure safe recreation, water and fish consumption, and aesthetic wellbeing.

Our study sites may have been variably affected by color producing agents, such as dissolved organic carbon. Since we did not have total inorganic carbon or dissolved organic carbon data, we were not able to separate the chl signal from these color producing agents. Also, the range of chl provided by the training data is not representative of the upper range of chl experienced during bloom events (Freedman 1974). A larger data set, with more frequent sampling over the summer might allow for more bloom events to be sampled. Possible problems with developing or applying Landsat and MODIS model across lakes, time periods include turbulence caused by wind on the lake surface (Singh 1994). Such weather patterns can change within a short period, and can be hard to determine with accuracy after the sampling period in the vicinity of the sample station. Time delays in image and sample collection date in the training data could have introduced some additional error, particularly with Landsat images. For the Landsat models, better results might have been somewhat dependent on sample size.

APPENDIX

Band Number	Band Band wavelength (nm)		Resolution (m)
1	Red	620-670	250
2	Near Infrared	841-876	250
3	Blue	459-479	500
4	Green	545-565	500
5	Near Infrared	1230-1250	500
6	Short Wave Infrared	1628-1652	500
7	Short Wave Infrared	2105-2155	500
8	Blue	405-420	1000
9	Blue	438-448	1000
10	Blue	483-493	1000
11	Green	526-536	1000
12	Green	546-556	1000
13 L/H	Red	662-672	1000
14 L/H	Red	673-683	1000
15	Near Infrared	743-753	1000
16	Short Wave Infrared	1550-1750	1000

Table 3-1: MODIS wavelength band features (NASA 2013).

Band number	Landsat bands	Band wavelength (nm)	Resolution (m)
1	Blue	450-520	30
2	Green	520-600	30
3	Red	630-690	30
4	Near Infrared	770*-900	30
5	Short Wave Infrared	1550-1750	30
6	Thermal Infrared	10400-12500	60**
7	Short Wave Infrared	2090***-2350	30

Table 3-2: Landsat 7 ETM+ wavelength band features (USGS 2011).Values are thesame for Landsat 4-5 TM, unless indicated by and asterisk. \* = 760 for TM, \*\* = 120 m forTM, and \*\*\* = 2080 for TM .

Band	Corresponding	Landsat bands	Band wavelength	Resolution
number	ETM+/TM band		(nm)	(m)
4	2	Green	500-600	60
5	3	Red	600-700	60
6	4	Near Infrared	700-800	60
7		Near Infrared	800-1100	60

Satellite	RMSE	R <sup>2</sup>	Cross-Val	Number of
	(µg/L)		R <sup>2</sup>	BRT Trees
MODIS	0.10	0.98	0.85	8300
ETM+ & TM	0.55	0.93	0.69	1800
MSS	0.73	0.88	0.66	1350

# Table 3-4: Boosted regression tree results for MODIS, Landsat ETM+, TM, and MSS.

Satemite	Source	Model	K²	RMSE
				(µg/L)
TM & ETM+	Novitski et al. in	BRT	0.69	0.55
	prep.			
TM & ETM+	Lathrop &	Band ratio	0.16	5.55
	Lillesand 1986	(Green bay)		
TM & ETM+	Lathrop &	Band ratio	0.16	4.39
	Lillesand 1986	(Lake Michigan)		
MSS	Novitski et al. in	BRT	0.66	0.73
	prep.			
MODIS	Novitski et al. in	BRT	0.85	0.10
	prep.			
MODIS	O'Reilly et al.	OC3	0.75	0.47
	2000			
MODIS	Lesht et al.	OC3-refine	0.75	0.49
	2013			

# Table 3-5: Comparison of BRT chl predictive models to models in the literature.SatelliteSourceModelR2RMSE



Great Lakes Measured Chl versus MODIS Predicted Chl

Figure 3-1: Measured chl versus MODIS BRT predicted chl for 2007-2009 GLNPO training data with 1:1 line.







# Figure 3-2 (cont'd)



Figure 3-3: Measured chl versus Landsat ETM+ and TM BRT predicted chl for 2007-2009 GLNPO training data with 1:1 line.



**Figure 3-4: Partial dependence plots for the Landsat ETM+ and TM BRT.** Each plot shows the relationship of chl (response variable) to the individual band or ratio (predictor variables) after accounting for the effects of all other predictor variables in the final BRT model.



Great Lakes Measured Chl versus Landsat MSS Predicted Chl

Figure 3-5: Measured chl versus Landsat MSS predicted chl for 2007-2009 GLNPO training data with 1:1 line.



**Figure 3-6: Partial dependence plots for the Landsat MSS BRT.** Each plot shows the relationship of chl (response variable) to the individual band or ratio (predictor variables) after accounting for the effects of all other predictor variables in the final BRT model.

# LITERATURE CITED

# LITERATURE CITED

- Babyak, M. A. 2004. What you see may not be what you get: A brief, nonethical introduction to overfitting in regression-type models. Psychosomatic Medicine **66**:411-421.
- Becker, R. H., M. I. Sultan, G. L. Boyer, M. R. Twiss, and E. Konopko. 2009. Mapping cyanobacterial blooms in the Great Lakes using MODIS. Journal of Great Lakes Research **35**:447-453.
- Binding, C. E., T. A. Greenberg, and R. P. Bukata. 2012. An analysis of MODIS-derived algal and mineral turbidity in Lake Erie. Journal of Great Lakes Research **38**:107-116.
- Carder, K. L., F. R. Chen, J. P. Cannizzaro, J. W. Campbell, and B. G. Mitchell. 2004. Performance of the MODIS semi-analytical ocean color algorithm for chlorophyll-a. Advances in Space Research **33**:1152-1159.
- Carpenter, D. J. and S. M. Carpenter. 1983. Modeling inland water quality using Landsat data. Remote Sensing of Environment **13**:345-352.
- Chander, G., B. L. Markham, and D. L. Helder. 2009. Summary of current radiometric calibration coefficients for Landsat MSS, TM, and ETM+, and EO-1 ALI sensors. Remote Sensing of Environment **113**:893-903.
- De'ath, G. 2007. Boosted trees for ecological modeling and prediction. Ecology 88:243-251.
- Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. Journal of Animal Ecology **77**:802-813.
- Freedman, P. L. 1974. Saginaw Bay: An evaluation of existing and historical conditions. Pages 1-137 in U. S. E. P. Agency, editor. Environmental Protection Agency, University of Michigan.
- Friedman, J. H. 2002. Stochastic gradient boosting. Computational Statistics & Data Analysis **38**:367-378.
- Giardino, C., M. Pepe, P. A. Brivio, P.Ghezzi, and E.Zilioli. 2001. Detecting chlorophyll, Secchi disk depth and surface temperature in a sub-alpine lake using Landsat imagery. The Science of the Total Environment **268**:19-29.
- Gower, J. F. R. and G. A. Borstad. 2004. On the potential of MODIS and MERIS for imaging chlorophyll fluorescence from space. International Journal of Remote Sensing **25**:1459-1464.

- Kabbara, N., J. Benkhelil, M. Awad, and V. Barale. 2008. Monitoring water qualitly in the coastal area of Tripoli (Lebanon) using high-resolution satellite data. ISPRS Journal of Photogrammetry & Remote Sensing **63**:488-495.
- Lathrop, R. G., Jr. and T. M. Lillesand. 1986. Use of Thematic Mapper data to assess water quality in Green Bay and central Lake Michigan. Photogrammetric Engineering and Remote Sensing **52**:671-680.
- Lebouteiller, A. and A. Herbland. 1982. Diel variation of chlorophyll-a as evidenced from a 13-day station in the equatorial atlantic-ocean. Oceanologica Acta **5**:433-441.
- Lesht, B. M., R. P. Barbiero, and G. J. Warren. 2013. A band-ratio algorithm for retrieving open-lake chlorophyll values from satellite observations of the Great Lakes. Journal of Great Lakes Research **39**:138-152.
- Lillesand, T. M., W. L. Johnson, R. L. Deuell, O. M. Lindstrom, and D. E. Meisner. 1983. Use of Landsat data to predict the trophic state of Minnesota lakes. Photogrammetric Engineering and Remote Sensing **49**:219-229.
- Masuoka, E. and K. Horrocks. 2010. Level 1 and atmospheric archive and distribution system. NASA Goddard Space Flight Center.
- NASA. 2013. MODIS Specifications.*in* N. A. a. S. Administration, editor., National Aeronautics and Space Administration.
- NOAA. 2013. NOAA solar calcuator. U.S. Department of Commerce, National Oceanic and Atmospheric Adminitration.
- O'Reilly, J. E., S. Maritorena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. McClain. 1998. Ocean color chlorophyll algorithms for SeaWiFS. Journal of Geophysical Research **103**:24,937-924,953.
- O'Reilly, J. E., S. Maritorena, M. C. O'Brien, D. A. Siegel, D. Toole, D. Menzies, R. C. Smith, J. L. Mueller, B. G. Mitchell, M. Kahru, F. P. Chavez, P. Strutton, G. F. Cota, S. B. Hooker, C. R. McClain, K. Carder, F. Müller-Karger, L. Harding, A. Magnuson, D. Phinney, G. F. Moore, J. Aiken, K. R. Arrigo, R. Letelier, and M. Culver. 2000. SeaWiFS postlaunch calibration and validation analyses, Part 3.*in* S. B. a. E. R. F. Hooker, editor. SeaWiFS Postlaunch Techinical Report Series. NASA.
- Oreopoulos, L., M. J. Wilson, and T. Varnai. 2011. Implementation on Landsat data of a simple cloud mask algorithm developed for MODIS land bands. . IEEE Geoscience and Remote Sensing Letters **8**:597-601.
- Owens, T. G., P. G. Falkowski, and T. E. Whitledge. 1980. Diel periodicity in cellular chlorophyll content in marine diatoms. Marine Biology **59**:71-77.

- Peng, C.-Y. J., K. L. Lee, and G. M. Ingersoll. 2002. An introduction to logistic regression analysis and reporting. The Journal of Educational Reserach **96**:3-14.
- Pozdnyakov, D., R. Shuchman, A. Korosov, and C. Hatt. 2005. Operational algorithm for the retrieval of water qualilty in the Great Lakes. Remote Sensing of Environment **97**:352-370.

Reynolds, C. 2006. Ecology of Phytoplankton. Cambridge University Press, New York.

- Shuchman, R., A. Korosov, C. Hatt, D. Pozdnyakov, J. Means, and G. Meadows. 2006. Verification and application of a bio-optical algorithm for Lake Michigan using SeaWiFS: A 7-year inter-annual analysis. Journal of Great Lakes Research 32:258-279.
- Toller, G. N., A. Isaacman, J. Kuyper, and V. Salomonson. 2006. MODIS Level 1B Product User's Guide. NASA/Goddard Space Flight Center.
- Torbick, N., F. Hu, J. Zhang, J. Qi, H. Zhang, and B. Becker. 2008. Mapping chlorophyll-a concentrations in West Lake, China using Landsat 7 ETM+. Journal of Great Lakes Research **34**:559-565.
- USGS. 2011. Landsat Missions. United States Geological Survey.
- USGS. 2013a. EarthExplorer. United States Geological Survey.
- USGS. 2013b. USGS FTP server parent directory. USGS.
- USGS. 2013c. USGS Global Visualization Viewer. United States Geological Survey.
- Vollenweider, R. A., M. Munawar, and P. Stadelmann. 1974. A comparative review of phytoplankton and primary production in the Laurentian Great Lakes. Journal of Fisheries Research Board of Canada **31**:739-762.
- Wetzel, R. G. 2001. Limnology: Lake and Reservoir Ecosystems. Adacemic Press, New York, New York.

# CHAPTER 4 IMPACTS OF CLIMATE CHANGE AND LAND-USE/LAND-COVER CHANGE ON PHYTOPLANKTON IN INNER SAGINAW BAY USING LANDSAT AND MODIS SATELLITE IMAGERY, 1973-2012

## Abstract

Saginaw Bay has experienced significant fluctuations in water quality over the last four decades. Remote sensing of chlorophyll (chl) can help to track water quality over large spatial and temporal scales. Our research objectives were to determine: 1) if water temperature and nutrient inputs have changed in inner Saginaw Bay from 1973-2012; and 2) how any changes in water temperature and nutrient inputs as well as distance from Saginaw River inputs affect algal blooms in Saginaw Bay. Chl concentrations were predicted from Landsat 1973-2012 and MODIS 2000-2012 maps with boosted regression tree analysis. Phosphorus loading was estimated with a mixed-effect model based existing phosphorus, precipitation, discharge, and land use/land cover data from various sources. Surface water temperature was derived from the thermal band aboard Landsat's ETM and TM satellites.

In inner Saginaw Bay, annual average and upper quartile Landsat-derived chlorophyll decreased from 7.44 to 6.62 and 8.38 to 7.38 µg/L between 1973-1982, and annual upper quartile of 8-day phosphorus loads increased from 5.29 to 6.79 kg between 1973-2012. Wilcoxon rank tests comparing chl values in groups, "near" and "far" from the Saginaw River mouth showed increases in minimum, average, and upper quartile chl values for MODIS (2000-2012) and Landsat- (1984-2012) derived chl near the Saginaw River mouth. Simple linear and multiple regression models for MODIS-derived chl indicate that distance from the Saginaw River mouth may influence chl concentration in Saginaw Bay; Landsat-derived surface water temperature and phosphorus loads to a lesser extent. Mixed-effect models for MODIS and Landsat-derived chl were related to chl better than simple linear or multiple

regressions, with random effects of pixel and sample date contributing substantially to predictive power (NSE=0.35-70), though phosphorus loads, distance to Saginaw River mouth, and water were significant fixed effects in most models. Water quality changes in Saginaw Bay between 1972-2012 may be influenced by phosphorus loading and distance to the Saginaw River's mouth. Other factors, such as the zebra mussel invasion may have had an impact on chl concentration.

## Introduction

Saginaw Bay has been the subject of great ecological interest for decades. This shallow, elongated bay in southwestern Lake Huron, with surface area of approximately 2,960 km<sup>2</sup>, shows marked differences in water chemistry, chlorophyll (chl) concentration, and bacterial activity (Schelske et al. 1974, Moll et al. 1980) than the rest of Lake Huron. Even the inner and outer portions of Saginaw Bay that are separated by a broad shoal (Freedman 1974) differ in circulation patterns (Danek and Saylor 1977) and algal assemblages (Stoermer and Theriot 1985). This may be due, in part, to differences in mean depth; 4.6 and 14.6 m for inner and outer bays, respectively (Freedman 1974).

Degradation of water quality in Saginaw Bay over the years has been influenced not only by natural basin characteristics, such as water depth, but also by direct municipal, agricultural, and industrial inputs, and invasive species introductions. In the 1970s, approximately 50% of the land in the watershed was agricultural (Freedman 1974), and 55% of the phosphorus in Saginaw Bay was coming from agricultural runoff (MDNR 1988). However, large quantities of nutrient and chemical inputs were also coming from industrial and municipal sources from the Saginaw River, which provides over 90% of Saginaw Bay's drainage inputs (Bierman et al. 1984). Around this time, phytoplankton studies indicated that cyanobacteria, including potential toxinforming *Anabaena* and *Microcystis*, were the dominant algal taxonomic group present in

Saginaw Bay between July and September (Munawar and Munawar 1982). High cyanobacterial abundance even caused taste and odor problems in municipal drinking water supplies until the late 1970s (MDNR 1988).

After Saginaw Bay earned an area of concern designation (MDNR 1988), improvements in water quality followed reductions in nutrient inputs, including from a phosphorus ban in laundry and other detergents in the Great Lakes (Hartig and Horvath 1982, Bierman et al. 1984) and additional targeted remediation efforts. However, by 1991 other water quality changes occurred after zebra mussels (*Dreissena polymorpha*) became established (Nalepa and Fahenestiel 1995), namely a 59% decrease in phytoplankton biomass as chlorophyll a (Fahnenstiel et al. 1995b) allowing planktonic cyanobacterial species to thrive relative to diatoms and chlorophytes (Bridgeman et al. 1995, Heath et al. 1995, Lavrentyev et al. 1995, Vanderploeg et al. 2001, Bierman et al. 2005). As planktonic algae decreased, benthic algal production increased (Fahnenstiel et al. 1995a) along with macrophytes (Skubinna et al. 1995). Benthic community composition shifted from diatoms to filamentous green algae (Lowe and Pillsbury 1995).

Ecosystem changes that involve algal abundance and community composition have a number of possible causes and implications, including changes in nutrient cycling and presence of toxins. Some studies indicate that zebra mussels control algal abundance and community composition directly by selective feeding (Vanderploeg et al. 2001) or indirectly by controlling the amount of phosphorus in the water column through assimilation, or sedimentation and resuspension of feces (Johengen et al. 1995, Vanderploeg et al. 2002). Nicholls et al. (1999) found that chl:phosphorus ratios after zebra mussel invasion were significantly different in different parts of the Great Lakes, having implications for the efficacy of phosphorus reduction

efforts (Nicholls et al. 1999). Zebra mussels may even indirectly affect their own filtering rate by selective feeding on certain phytoplankton types, leaving the less edible taxa (Pillsbury et al. 2002). Along with the increased presence of *Microcystis*, is increased chance for toxins. One study showed that the gene for microcystin production was more prevalent in Saginaw Bay than in Lake Erie (Dyble et al. 2008). Another study found that phosphorus has a large influence on microcystin based on its influence on *Microcystis* growth rates and abundance (Fahnenstiel et al. 2008).

Remote sensing is a useful tool for broad-scale, long-term assessment of water quality indicators, such as chl concentration. Early remote sensing studies of Saginaw Bay, using Landsat MSS satellite imagery, were done for the purpose of creating maps of water features, including chlorophyll, Secchi depth, and temperature (Rogers 1975, Rogers et al. 1976, McKeon et al. 1977). More recently, the AVHRR satellite has been used to evaluate differences in Secchi depth and total suspended solids before and after zebra mussel invasion in Saginaw Bay (Budd et al. 2001). Other satellites that have been used to study various parameters of Saginaw Bay, such as, chl concentration or concentration of cyanobacterial species, include the SeaWiFS (Pozdnyakov et al. 1999), and MERIS (Wynne et al. 2008). Little work has been done in Saginaw Bay with Landsat and MODIS satellites, both which are currently functional and have images available to the public free of charge. In addition, the constellation of Landsat satellites has an image archive dating back to 1972, the year the Clean Water Act was enacted. MODIS has an image archive that goes back to 2000. Therefore, these satellites have the potential to be extremely valuable for water quality assessment of water bodies, such as Saginaw Bay, that have variable ecological history.

#### Research objectives

Using inferred-chl from MODIS and Landsat images and satellite boosted regression tree models, USGS river discharge data, and distance-to-river mouth data, our research objectives were to determine: 1) if water temperature and nutrient inputs have changed in inner Saginaw Bay from 1973-2012; and 2) how changes in water temperature and nutrient inputs as well as distance from Saginaw River inputs affect algal blooms in Saginaw Bay using satellite-derived chl concentrations.

#### Methods

#### Data

Chl concentration was inferred from boosted regression tree (BRT) models (Novitski et al. in prep) applied to MODIS and Landsat MSS, TM, and ETM+ images using Python with GDAL and RPy2 (Lawawirojwong et al. in prep). Chl data were extracted from images with pixels that corresponded to a portion of Saginaw Bay's inner bay (Figure 4-1) deemed deep enough to avoid the influence of reflection from the lake bottom, using Python 2.5 software (Lawawirojwong et al. in prep). This subset of pixels from the inner bay was chosen for the analysis based on a depth to chl relationship, derived from Carlson's Secchi to chl relationship (Carlson 1977).

$$Ln(Secchi) = -0.68 * Ln(chl) + 2.04$$

If depth was at least as deep as Secchi disk depth, we assumed that reflectance from the bottom of the bay would not interfere with detection of chl in the water column. Most of the predicted chl concentrations for Saginaw Bay were > 2  $\mu$ g/L, which agrees with historical data of the inner bay (CCIW 1972, Schelske and Roth 1973, Schelske et al. 1974, Smith et al. 1977, Fahnenstiel et al. 1995b, Nalepa et al. 1995, Suzuki et al. 1995, Fanslow et al. 2001, Vanderploeg et al. 2001,

Lehman et al. 2004). Based on Carlson's equation, with at least 2  $\mu$ g/L, the minimum depth we used in our analyses is 4.8 m. For practicality, we rounded this number to 5 m. MODIS-inferred chl data consists of pixel values from 8-day composite MODIS images, starting from the last week of July through the first week of September, for a total of six images per year, except for the year 2000 in which only 5 composite images were created because images between 08/04/2000-08/11/2000 were not available. Therefore, a total of 77 MODIS images were used in these analyses. The Landsat-inferred chl data consist of 209 images, taken from July through September of 1972-2012. Due to cloud interference or low image quality for some images, the number of images ranges from 0-11 per year, 1972 and 1983 having no images. To reduce the Landsat data to a more manageable size for statistical analyses, we found the latitude and longitude of each MODIS pixel centroid using ArcGIS 9.3 and extracted the Landsat-inferred chl value from the Landsat pixel at that same location using Python 2.5 software. Therefore, the resulting number of chl values per Landsat image, 1887, was the same for the MODIS data.

Locations of river mouth in Saginaw Bay data were determined using hydrological information from NHDPlus (USEPA and USGS 2005). Once MODIS pixel centroids were determined in ArcGIS 10.1, distance-to-river mouth data were created in Microsoft Excel 2007 by applying the haversine formula. For the purposes of this study, we selected the distance to the Saginaw River mouth for each pixel.

Phosphorus loads at river mouths emptying into Saginaw Bay (Figure 4-2) were estimated for 1972-2012 using linear mixed-effects models (Esselman et al. in prep). Land use/land cover information used to inform phosphorus load models included 2001 and 2006 National Land Cover Database data, 1978 Michigan Resource Inventory System data, and 1996 Coastal Change Analysis Program data. Discharge data used in phosphorus load model

development were, themselves, created using precipitation data from NOAA North American Regional Reanalyis and National Centers for Environmental Prediction and from a watershed characteristics database of Dr. Dave Hyndman's lab. Measured phosphorus data used in the phosphorus load model building came from three general sources: EPA STORET, Kendall et al. (unpublished), and Dr. R.J. Stevenson. To obtain phosphorus load values at a level that was spatially relevant to the chl data, phosphorus loads were weighted by the distance data using two weighting approaches (Van Sickle and Johnson 2008) inverse distance weighting f(d) = d<sup>-α</sup> and exponential weighting f(d) = exp(-<sup>α</sup>d) where d= distance between the pixel centroid and the river mouth. In early analyses, inverse distance weighting resulted in better model outcomes, so it was used for the remainder of analyses.

Surface water temperature data were obtained from Landsat ETM+ and TM images by extracting thermal band pixel values corresponding to the location of MODIS pixel centroids and converting the digital number to degrees Celsius (Chander et al. 2009) using Python 2.5. *Statistics for MODIS and Landsat* 

Eighteen MODIS composite images and seven Landsat images had missing pixel values due to clouds. For one of the MODIS images (8/4/2000), no pixel values were available, and for Landsat, images from 1972 and 1983 missing due to cloud coverage or image quality problems. For the images with missing pixel values, temporal analyses of these pixels or sample dates could not be carried out.

Simple linear regressions were done to determine temporal trends of chl, water temperature, and phosphorus loads in inner Saginaw Bay over the lifetime of each satellite. However, since Landsat MSS did not have a thermal band with which to detect water temperature, Landsat-derived water temperature only goes back to 1984. Landsat-derived chl over the 1973-2012 had a non-normal, bimodal distribution. A category and regression tree analysis was run to determine if there were two statistically distinct chl groups (1973-1982 and 1984-2012). Based on the category and regression tree results and the distribution of chl data in the resulting groups, linear regressions were run on the 1973-1982 and 1984-2012 data, individually. Wilcoxon rank test was done on "near" and "far" pixels (below and above 43.8°N, respectively) to further discern the influence of the Saginaw Bay river mouth.

Simple linear regression analyses were also run to model effects of water temperature  $(X_1)$ , phosphorus loads  $(X_2)$  on chl concentration (Y). The effects on chl of distance to the Saginaw River mouth was tested, but never in conjunction with phosphorus loads due to the redundancy of the distance-weighted phosphorus loads.

Finally, mixed-effects models were calculated to determine if the predictive power can be improved by adding the unexplained or random variation of time or space using the following equation:

$$Y_{ij} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2j} + \alpha_i + \gamma_j + \varepsilon_{ij}$$

where  $\beta$  are the coefficients for the *X* predictors (fixed effects), the site index is represented by i = 1, 2, ..., N, the time index is represented by j = 1, 2, ..., T. Also, the residual term is  $\varepsilon_{ij} \sim N(0, \sigma^2)$ , the random effect on sample site is  $\alpha_i \sim N(0, \tau^2)$ , and the random effect on sample date is  $\gamma_j \sim N(0, \delta^2)$ , where  $\sigma^2, \tau^2$  and  $\delta^2$  represent variability parameters.

Criteria used to evaluate the mixed-effects models included Bayesian information criterion (BIC) and Akaike information criterion (AIC). Both BIC and AIC evaluate the performance of the model while penalizing for increasing complexity. Root mean square error (RMSE) was calculated to help understand the error associated with the final predicted values of the model, and Nash-Sutcliffe efficiency (NSE) was used determine the proportion of variation in the chl data that were explained, much like R<sup>2</sup> values in linear regression models. Smaller BIC, AIC, and RMSE values and larger NSE values are preferable. All statistics were run in R version 2.15.1. Linear regression models were run in R the package, stats, and mixed-effect models were run in the lme4 package. Satellite-inferred chl was then compared to values in the literature from inner Saginaw Bay.

## Results

#### *Temporal trends*

Although there was a slight increasing trend in annual average and maximum MODISderived chl in the 2000-2012 time period during which MODIS has been operational, the results were not statistically significant (Table 4-1, Figure. 4-3). Annual average and upper quartile Landsat-derived chl decreased between 1972-2012, with a notable drop in the early 1980s, decreasing in one year from 6.5 to 4.5 µg chl/L and 8.5 to 6.5 µg chl/L in average and upper quartile values, respectively (Figure 4-4). Regression tree analysis showed a significant changepoint producing two distinct periods for Landsat-derived chl: 1973-1983 and 1984-2012 (complexity parameter=0.95). It seemed likely that this changepoint was due to differences in differences in band ranges and widths for sensors on Landsat satellites, so the two periods were analyzed separately. Linear regressions run on each group of Landsat-derived chl showed a decrease in average annual chl between 1973-1982 from 7.44 to 6.62 µg/L, but no change from 1984-2012 (Table 4-1, Figure 4-4). For both MODIS and Landsat (1894-2012), Wilcoxon test of "near" and "far" derived chl, average, upper quartile, and minimum values were statistically significantly higher near the Saginaw River mouth.

Between 2000-2012, MODIS-derived chl concentrations were consistently higher than Landsat concentrations, although trends over time had some similarities with relatively low concentrations in 2000 and 2003, and higher concentrations in 2001, 2005, and 2007 (Figure 4-3). Annual upper quartile 8-day phosphorus load increased between 1973-2012 from about from 5.29 to 6.79 kg/8 days, while temporal trends in annual average 8-day phosphorus were not significant (4-1, Figure 4-5). Upper quartile values of 8-day phosphorus loading are notably smaller than average values (Figure 4-5). Between 1984-2012, annual upper quartile water temperature increased from about 19°C to 20°C, however trends in annual average water temperature were not statistically significant (Table 4-1, Figure 4-6). *Relationships between algal blooms and global change variables* 

MODIS-derived chl was significantly (p<0.05) related to phosphorus loading, distance to the mouth of the Saginaw River. Landsat-derived surface water temperature and interactions among these variables, but phosphorus loading and water temperature or other combinations of those predictor variables had almost no predictive power (Table 4-3) Interestingly, the best mixed-effect models for MODIS-derived chl were any that included random effects of sample site and sample date, explaining 56% (RMSE=1.10  $\mu$ g/L) of the variation in the satellite-inferred chl data (Table 4-4; Figure 4-7). Water temperature was the fixed effect that was most commonly not statistically significant in mixed effect models, while phosphorus loads and distance were always significantly (p<0.05) related to MODIS-derived chl (Table 4-4).

For 1973-2012, phosphorus loads and distance were negatively correlated to Landsatderived chl (p<0.05, Table 4-5). For 1984-2013, and phosphorus loads, water temperature were positively correlated to Landsat-derived chl, while distance was negatively correlated with chl (p<0.05, Table 4-5). Similarly to the MODIS mixed-effect model results, the best 1973-1982 Landsat mixed-effect model results were with models that included either sample date or both random effects of pixel and sample date, accounting for 70% (RMSE=0.80  $\mu$ g/L) of the variation in the chl data (Table 4-6). Models that included sample date as the only random effect were almost as strong as those with both random effects (Table 4-6).

The 1984-2012 mixed-effect models (some of which included water temperature as a fixed effect) showed similar trends to other Landsat and MODIS mixed-effect models with those that included both random effects being the strongest predictors of chl, explaining about 35% (RMSE=0.77  $\mu$ g/L) of the variation in the chl data, and those with just sample date as explaining 29-34% (RMSE=0.78-0.81  $\mu$ g/L) of the variation (Table 4- 6). Only two Landsat mixed-effect models with non-significant fixed effect of phosphorus loads (Table 4-6). As with MODIS mixed effects models, the strongest of all Landsat mixed models was one that included distance as the only fixed effect, as well as pixel and sample date as random effects (Table 4-6; Figure 4-8).

# Discussion

Long-term trends in satellite-derived chl, phosphorus loading, and surface water temperature, as well as and relationships among these variables and to distance to the Saginaw River mouth, indicate that land use change, climate change, and invasive species may be influencing water quality in inner Saginaw Bay. Although, it appears that chl concentration has decreased, overall, since the Clean Water Act was enacted in 1972, a possible weakening of that trend in recent years indicates a continued need for water quality assessment and management. Increasing upper quartile values of 8-day phosphorus load in the last 40 years, and substantially larger average vs. upper quartile phosphorus load values indicate that sizable phosphorus inputs, though possibly infrequent, could be one factor in the lack of chl decrease in recent years (Paerl and Huisman 2008). This occurred despite efforts to decrease phosphorus loads into Saginaw

Bay (Fahnenstiel et al. 1995b) since the mid-1970s. Possible explanations for increasing phosphorus could be a changing urban or agricultural landscape in the Saginaw River Watershed, or more frequent precipitation events in the Midwest (Lofgren and Gronewold 2012) that wash phosphorus directly into Saginaw Bay, or through nearby rivers. Also, though trends in surface water temperature were not apparent using Landsat-derived values, recent studies of warming in the Great Lakes indicate that warming may still be a factor in Saginaw Bay (Austin and Colman 2007).

Indirect or interactive effects of phosphorus, water temperature and chl may be important relationships to take into account. For example, the indirect effects of increasing water temperature on chl may be mediated by physical factors, such as length of water column stratification, and may effect phosphorus cycling and accessibility of phosphorus to biological organisms (Scavia 1979). We did not see an overall effect of zebra mussels on chl. This contradicts other findings of decreased phytoplanktonic chl after the zebra mussel invasion (Fahnenstiel et al. 1995b), which may be accounted for by the inclusion of shallow sampling stations. Zebra mussels, themselves, could be mediating the effects of phosphorus or water temperature on phytoplankton (Sarnelle et al. 2005) or they could be moderating effects of phosphorus or surface water temperature on chl by removing algae from the water column (Fahnenstiel et al. 1995b). Algal communities that have been grazed by zebra mussel tend to have more buoyant cyanobacterial species, such as *Microcystis*, which are thought to be harder for zebra mussels to ingest (Vanderploeg et al. 2001). Shifts toward more toxin-producing algal taxa is a major concern for managers and restoration efforts.

Both MODIS and Landsat-derived chl were consistent with chl values observed in Saginaw Bay (Table 4-7). The July-September seasonal average of MODIS-inferred chl fit
within a range of average and one-time chl values from the literature (Table 4-7). Seasonal average Landsat-inferred chl for 1991 and 1994 were within 1  $\mu$ g/L chl of average values from the literature for these years (Table 4-7). Overall, MODIS-inferred chl shows a more dynamic range than Landsat-inferred chl. and is possibly more accurate due to narrower band widths and larger numbers of spectral bands. MODIS also has finer temporal resolution to capture the chl temporal signature. However, lakes that are too small for MODIS will still require the use of a Landsat model or a model that hybridizes MODIS and Landsat-derived chl values. In such a hybrid model, the bias between MODIS and Landsat models could be determined and used to recalibrated Landsat models to reflect a more accurate surface chl concentration.

Mixed-effects models were able to capture relationships between predictor variables and satellite-derived chl much more effectively than simple linear or multiple regression models. Continued refinement of mixed-effect models with existing data can enable us to understand which ecosystem attributes are most influential determinants of algal abundance over space and through time. Similar models can be applied to other water bodies for improved management. Future models could take into account the lag in the response of chl concentration to nonpoint source pollution, internal nutrient cycling, or water circulation effects. We would expect this connection to be particularly important in extreme weather events, including heavy storms or drought. Currently, the data used for chl model calibration do not necessarily include a spring or summer bloom event for each year of data collected. Having more chl data as well as information on the extent and duration of blooms events would help us better capture bloom timing and extent in our models, and it would also help us understand the timing of these events in relation to other ecological factors, such as storm events. Having satellite inferred-chl data for shallower depths could help to elucidate any relationships due to the proximity of river water and

96

nutrient inputs. In addition, the Landsat-inferred water temperature may have noise associated with it due to atmospheric conditions that could not be accounted for. Modeled phosphorus loads were extrapolated from numerous data sources. Refining both of these data sets could potentially allow for more accurate chl prediction. Finally, accounting for interactions not just within each pixel or sample date, but also between pixels and sample dates may help to improve model outcomes. Additional years of data may help to elucidate trends in MODIS-derived chl, which has the potential to be a powerful ecological assessment tool with its daily overpass and relatively fine spectral bands (NASA 2013).

Our results show the utility of using satellite-derived chl data to monitor and understand water quality trends in ecological models. Though changes in the data over the 40 year study period are modest, they show a distinct trend that could become more prominent in the future with possible tipping points after which management or restoration of Saginaw Bay is not practical or feasible.

APPENDIX

Table 4-1. Linear regression results for temporal trends in chlorophyll, water temperature, and phosphorus loads in inner Saginaw Bay. The 2000-2012 results correspond to MODIS-derived chl. Landsat-derived chl results are broken into two year groups (1973-1982 and 1984-2012) based on regression tree groupings. Landsat-derived water temperature values span 1984-2012. Chl = chlorophyll concentration ( $\mu$ g/L), P load = 8-day phosphorus load (kg), water temp. = surface water temperature (°C), Avg = annual mean value. IIO = annual upper quartile value. NS = not statistically significant.

mean value, og – annual upper quartile value, NS – not statistically significant.											
Satellite	Years	Variable	Annual	Increase/	Coefficients	Adjuste	p-				
			Avg/Max	Decrease		$d R^2$	value				
MODIS	2000-2012	Chl	Avg			-0.06	NS				
MODIS	2000-2012	Chl	UQ			-0.08	NS				
Landsat	1973-1982	Chl	Avg	Decrease	-0.09	0.33	0.05				
Landsat	1973-1982	Chl	UQ			0.15	NS				
Landsat	1984-2012	Chl	Avg			-0.03	NS				
Landsat	1984-2012	Chl	UQ			-0.04	NS				
Landsat	1973-2012	Р	Avg.			0.04	NS				
Landsat	1973-2012	Р	UQ	Increase	0.02	0.26	< 0.01				
Landsat	1984-2012	WT	Avg			0.00	NS				
Landsat	1984-2012	WT	UQ			0.07	NS				

**Table 4-2: Wilcoxon Rank test results using near and far groupings of MODIS or Landsat chl or water temperature data.** Chl=chlorophyll, WT=water temperature, Max=annual maximum, Min=annual minimum, Avg=annual mean, UQ=annual upper quartile. \* indicates ties in ranking for which an exact p-value cannot be calculated.

Satellite	Chl/WT	Index	W	p-value	
MODIS 2000-2012	Chl	Max	80	NS	
		Min	27	< 0.01	
		Avg	24	< 0.01	
		UQ	24	< 0.01	
	WT	Max	76	NS*	
		Min	51	NS*	
		Avg	65	NS	
		UQ	61	NS*	
Landsat 1973-1982	Chl	Max	53	NS*	
		Min	36.5	NS*	
		Avg	55	NS	
		UQ	49	NS	
Landsat 1984-2012	Chl	Max	529.5	NS*	
		Min	117.5	< 0.01	
		Avg	119	< 0.01	
		UQ	177	< 0.01	
	WT	Max	393.5	NS *	
		Min	339.5	NS *	
		Avg	367	NS *	
		UQ	337.5	NS *	

**Table 4-3. MODIS (2000-2012) linear regression results.** Linear regression results using combinations of variables (water temperature, phosphorus load, and distance-to-river mouth to predict chlorophyll concentration in inner Saginaw Bay. P=8-day phosphorus load (kg), D=distance to Saginaw River mouth from MODIS pixel centroid (km), WT=Landsat-derived surface water temperature (°C), pixel=sample site, date=date of satellite-derived chl. Phosphorus load and distance were never used in the same model because phosphorus load is inverse-distance weighted for regression analyses.

Secure prosprior as road is my		9
Predictor Variables	Adjusted R <sup>2</sup>	p-value
Р	0.01	< 2x10 <sup>-16</sup>
WT	0.01	< 2x10 <sup>-16</sup>
D	0.12	< 2x10 <sup>-16</sup>
P + WT	0.01	< 2x10 <sup>-16</sup>
WT + D	0.13	< 2x10 <sup>-16</sup>

**Table 4-4: MODIS linear mixed-effect model results**. P=8-day phosphorus load (kg), D=distance to Saginaw River mouth from MODIS pixel centroid (km), WT=Landsat-derived surface water temperature (°C), pixel=sample site, date=date of satellite-derived chl, RMSE=root mean square error, NSE=Nash-Sutcliffe efficiency, AIC=Akaike information criterion, BIC=Bayesian information criterion. Statistical significance is at p≤0.05 level.

Fixed-effects variables	Significant fixed	RMSE	NSE	AIC	BIC
	effect variables				
P + pixel	Р	1.52	0.17	466231	466270
P + date	Р	1.21	0.47	403940	403979
P + pixel + date	WT	1.10	0.56	387856	387905
WT + pixel	WT	1.53	0.15	467454	467493
WT + date	WT	1.28	0.41	417343	417382
WT + pixel + date		1.10	0.56	387859	387908
D + pixel	D	1.54	0.15	464877	467916
D + date	D	1.14	0.53	388388	388427
D + pixel + date	D	1.11	0.56	384771	384820
P + WT + pixel	Р	1.52	0.17	465758	465807
P + WT + date	Р	1.21	0.47	403757	403805
P + WT + pixel + date	Р	1.10	0.56	387867	387926
D + WT + pixel	D,WT	1.53	0.15	464435	464483
D + WT + date	D, WT	1.14	0.53	388398	388447
D + WT + pixel + date	D	1.10	0.56	384782	384840

**Table 4-5: Landsat (1973-2012 or 1984-2012) linear regression results.** Linear regression results using combinations of variables (water temperature, phosphorus load, and distance-to-river mouth to predict chlorophyll concentration in inner Saginaw Bay. P=8-day phosphorus load (kg), D=distance to Saginaw River mouth from Landsat pixel centroid (km), WT=Landsat-derived surface water temperature (°C), pixel=sample site, date=date of satellite-derived chl. Phosphorus load and distance were never used in the same model because phosphorus load is inverse-distance weighted for regression analyses.

Year Range	Predictor Variables	Adjusted	p-value
		R <sup>2</sup>	
1973-2012	Р	0.05	< 0.01
1973-2012	D	0.01	< 0.01
1984-2012	Р	0.00	< 0.01
1984-2012	D	0.05	< 0.01
1984-2012	WT	0.00	< 0.01
1984-2012	P + WT	0.00	< 0.01
1984-2012	D + WT	0.06	< 0.01

**Table 4-6: Landsat linear mixed-effect model results.** P=8-day phosphorus load (kg), D=distance to Saginaw River mouth from Landsat pixel centroid (km), WT=Landsat-derived surface water temperature (°C), pixel=sample site, date=date of satellite-derived chl, RMSE=root mean square error, NSE=Nash-Sutcliffe efficiency, AIC=Akaike information criterion, BIC=Bayesian information criterion. Statistical significance is at p≤0.05 level.

Year Range	Fixed-effects	Signif.	RMSE	NSE	AIC	BIC
	variables	fixed				
		effect				
		variables				
1973-2102	P + pixel	Р	1.39	0.09	1291662	1291705
1973-2102	P + date	Р	0.81	0.69	890991	891035
1973-2102	P + pixel + date	Р	0.80	0.70	885052	885106
1973-2102	D + pixel	D	1.45	0.14	1314953	1314996
1973-2102	D + date	D	0.80	0.70	883763	883807
1973-2102	D + pixel + date	D	0.80	0.70	882928	882982
1984-2012	P + pixel	Р	0.92	0.07	777789	777831
1984-2012	P + date	Р	0.80	0.31	687537	687579
1984-2012	P + pixel + date		0.80	0.35	676255	676308
1984-2012	WT + pixel	WT	0.93	0.07	778085	778127
1984-2012	WT + date	WT	0.81	0.29	697233	697276
1984-2012	WT + pixel + date	WT	0.77	0.35	675714	675767
1984-2012	D + pixel	D	0.93	0.06	776769	776811
1984-2012	D + date	D	0.78	0.34	674787	674829
1984-2012	D + pixel + date	D	0.77	0.35	673112	673165
1984-2012	P + WT + pixel	P,WT	0.92	0.08	776176	776229
1984-2012	P + WT + date	P,WT	0.80	0.31	687340	687393
1984-2012	P + WT + pixel + date	WT	0.77	0.35	675733	675797
1984-2012	D + WT + pixel	D,WT	0.93	0.07	774963	775016
1984-2012	D + WT + date	D, WT	0.78	0.34	674239	674292
1984-2012	D + WT + pixel + date	D, WT	0.77	0.35	672553	672617

Table 4-7: Comparison of MODIS- and Landsat-derived inner Saginaw Bay chlorophyll values and values from
the literature. Pub=publication,1=Schelske & Roth 1973, 2=CCIW 1872*, 3=Schelske et al. 1974, 4=Smith et al. 1977,
Fahnenstiel et al. 1995, 6=Fanslow et al. 2001, 7=Nalepa et al. 1995, 8=Lehman et al. 2004, 9=Suzuki et al. 1995,
10=Vanderploeg et al. 2001, 11=Landsat-derived chlorophyll, 12=MODIS-derived chlorophyll. *Cited in Freedman et al.
1974, **samples taken from wetland area, †some values estimated from graph and some exact numbers taken from
Fanslow et al. 2001. Month=sample month, JN=June, JL=July, A=August, S=September. Type=sample type, Avg.=average,
R=range, Max=maximum, SA=seasonal average, OT=one-time, TOTS=two one-time samples. Last two digits used to
indicate sample year (e.g., 70=1970, 00=2000).

Pub	Month	Туре	70	72	74	89	90	91	92	93	94	95	96	99	00
1	JL	Avg.	7.26												
				4.0-											
		R;		12;											
2*	А	Max		20.0											
3	S	Avg.		15.0											
					16.26;										
		Avg.;			2.33-										
4		R			45.7										
5	IN-S	SA					11.0	7.5	6.0	3.0					
6	, JL											16.1	3.49		
6	Â	ОТ										9.01	13.45		
7	S	ОТ										6.28	16.79		
7	JL	ОТ						9.1	7.8	1.1					
7	А	ОТ						2.6	11.9	2.6					
7	S	ОТ						2.6	16.4	2.9					
8	А	ОТ												10.22	
															17.4
8	S	TOTS													17.5

Tal	ble 4-7 (	conťd)													
Pub	Month	Туре	70	72	74	89	90	91	92	93	94	95	96	99	00
9**	А	R				0.5- 42.0									
10† 10†	JL A	OT OT						9.0 2.5	8.0 12.0	1.0 2.5	8.0 15.5	13.0 7.64	1.85 16.79		
10†	S	ОТ					14.0	2.5	16.5	3.0	11.0	3.74	7.68		
11 12	JL-S JL-S	SA SA			7.03	4.58	4.69	4.32	4.43	4.52	4.40	4.56	4.40	4.45	4.47 6.68



**Figure 4-1: Saginaw Bay's inner bay with depth**  $\geq$  **5.** The gray dots indicate the portion Saginaw Bay's inner bay with depth  $\geq$  5 m that was used for statistical analyses. In the lower left is the Saginaw River.



**Figure 4-2: Inner Saginaw Bay study area and river mouth inputs into Saginaw Bay.** Inner Saginaw Bay study area (gray dots) and river mouth inputs into Saginaw Bay (black dots).



Figure 4-3. Annual average and upper quartile MODIS and Landsat-derived chlorophyll concentration ( $\mu$ g/L) in inner Saginaw Bay, 2000-2012. Annual average MODIS and Landsat-derived chl concentration ( $\mu$ g/L) in inner Saginaw Bay, 2000-2012 (left). Upper quartile MODIS and Landsat-derived chl concentration in inner Saginaw Bay, 2000-2012 (right).



**Figure 4-4: Annual average and upper quartile Landsat-derived chlorophyll concentration (µg/L) in inner Saginaw Bay, 1973-2012.** Annual average Landsat-derived chl concentration (µg/L) in inner Saginaw Bay, 1973-2012 (left). Upper quartile Landsat-derived chl concentration in inner Saginaw Bay, 1973-2012 (right).



**Figure 4-5: Annual average and upper quartile 8-day phosphorus load (kg) in inner Saginaw Bay, 1973-2012.** Annual average 8-day phosphorus load (kg) in inner Saginaw Bay, 1973-2012 (left). Upper quartile 8-day phosphorus load (kg) in inner Saginaw Bay, 1973-2012 (right).



**Figure 4-6: Annual average and upper quartile Landsat-derived surface water temperature (°C) in inner Saginaw Bay, 1984-2012.** Annual average Landsat-derived surface water temperature (°C) in inner Saginaw Bay, 1984-2012 (left). Upper quartile Landsat-derived surface water temperature (°C) in inner Saginaw Bay, 1984-2012 (right).



MODIS-derived chlorophyll (µg/L)

Figure 4-7: MODIS-derived chl vs. mixed-effect model chl (µg/L) in inner Saginaw Bay, 2000-2012 with 1:1 line.



Landsat-derived chlorophyll (µg/L)

Figure 4-8: Landsat-derived chl vs. mixed-effect model chl ( $\mu$ g/L) in inner Saginaw Bay, 1984-2012 with 1:1 line.

- Austin, J. A. and S. M. Colman. 2007. Lake Superior summer water temperatures are increasing more rapidly than regional air temperature: A positive ice-albedo feedback. Geophysical Research Letters **34**.
- Bierman, V. J., Jr., D. M. Dolan, and R. Kasprzyk. 1984. Retrospective analysis of the response of Saginaw Bay, Lake Huron, to reductions in phosphorus loadings. Environmental Science and Technology 18:23-31.
- Bierman, V. J., Jr., J. Kaur, J. V. DePinto, T. J. Feist, and D. W. Dilks. 2005. Modeling the role of zebra mussels in the proliferation of blue-green algae in Saginaw Bay, Lake Huron. Journal of Great Lakes Research **31**:32-55.
- Bridgeman, T. B., G. L. Fahnenstiel, G. A. Lang, and T. F. Nalepa. 1995. Zooplankton grazing during the zebra mussel (*Dreissena polymorpha*) colonization of Saginaw Bay, Lake Huron. Journal of Great Lakes Research **21**:567-573.
- Budd, J. W., T. D. Drummer, T. F. Nalepa, and G. L. Fahnenstiel. 2001. Remote sensing of biotic effects: Zebra mussels (*Dreissena polymorpha*) influence on water clarity in Saginaw Bay, Lake Huron. Limnology and Oceanography 46:213-223.
- Carlson, R. E. 1977. A trophic state index for Lakes. Limnology and Oceanography **22**:361-369.
- CCIW. 1972. Canadian Center for Inland Waters: Biological and chemical water quality dat available from their various Great Lakes cruises.
- Chander, G., B. L. Markham, and D. L. Helder. 2009. Summary of current radiometeric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. Remote Sensing of Environment **113**:893-903.
- Danek, L. J. and J. H. Saylor. 1977. Measurements of the summer currents in Saginaw Bay, Michigan. Journal of Great Lakes Research **3**:65-71.
- Dyble, J., G. L. Fahnenstiel, R. W. Litaker, D. F. Millie, and P. A. Tester. 2008. Microcystin concentrations and genetic diversity of *Microcystis* in the lower Great Lakes. Environmental Toxicology **23**:507-516.
- Fahnenstiel, G. L., T. B. Bridgeman, G. A. Lang, M. J. McCormick, and T. F. Nalepa. 1995a.
  Phytoplankton productivity in Saginaw Bay, Lake Huron: Effects of zebra mussel (Dreissena polymorpha) colonization. Journal of Great Lakes Research 21:465-475.

- Fahnenstiel, G. L., G. A. Lang, T. F. Nalepa, and T. H. Johengen. 1995b. Effects of zebra mussel (Dreissena polymorpha) colonization on water quality parameters in Sagainaw Bay, Lake Huron. Journal of Great Lakes Research **21**:435-448.
- Fahnenstiel, G. L., D. F. Millie, J. Dyble, R. W. Litaker, P. A. Tester, M. J. McCormick, R.Rediske, and D. Klarer. 2008. Microcystin concentrations and cell quotas in Saginaw Bay, Lake Huron. Aquatic Ecosystem Health & Management 11:190-195.
- Fanslow, D. L., T. F. Nalepa, and T. H. Johengen. 2001. Seasonal changes in the respiratory electron transport system (ETS) and respiration of the zebra mussel, *Dreissena polymorpha* in Saginaw Bay, Lake Huron. Hydrobiologia **448**:61-70.
- Freedman, P. L. 1974. Saginaw Bay: An evaluation of existing and historical conditions. Pages 1-137 *in* U. S. E. P. Agency, editor. Environmental Protection Agency, University of Michigan.
- Hartig, J. H. and J. Horvath. 1982. A preliminary assessment of Michigan's phosphorus detergent ban. Journal (Water Pollution Control Federation) **54**:193-197.
- Heath, R. T., G. L. Fahnenstiel, W. S. Gardner, J. F. Cavaletto, and S.-J. Hwang. 1995.
  Ecosystem-level effects of zebra mussels (*Dreissena polymorpha*): An enclosure experiement in Saginaw Bay, Lake Huron. Journal of Great Lakes Research 21.
- Johengen, T. H., T. F. Nalepa, G. L. Fahnenstiel, and G. Goudy. 1995. Nutrient changes in Saginaw Bay, Lake Huron, after the establishment of zebra mussel (*Dreissena polymorpha*). Journal of Great Lakes Research **21**:449-464.
- Lavrentyev, P. J., W. S. Gardner, J. F. Cavaletto, and J. R. Beaver. 1995. Effects of the zebra mussel (Dreissena polymorpha Pallas) on protozoa and phytoplankton from Saginaw Bay, Lake Huron. Journal of Great Lakes Research **21**:545-557.
- Lehman, J. T., A. Bazzi, T. Nosher, and J. O. Nriagu. 2004. Copper inhibition of phytoplankton in Saginaw Bay, Lake Huron. Canadian Journal of Fisheries and Aquatic Sciences **61**:1871-1880.
- Lofgren, B. L. and A. Gronewold. 2012. Water Resources.
- Lowe, R. L. and R. W. Pillsbury. 1995. Shifts in benthic algal community structure and function following the appearance of zebra mussels (*Dreissena polymorpha*) in Saginaw Bay, Lake Huron. Journal of Great Lakes Research **21**:558-566.
- McKeon, J. B., R. H. Rogers, and V. E. Smith. 1977. Production of a water quality map of Saginaw Bay by computer processing of Landsat-2 data.1045-1054.
- MDNR. 1988. Saginaw River and Saginaw Bay. Page 588 *in* M. D. o. N. Resources, editor. MDNR, Lansing, Michigan.

- Moll, R. A., C. O. Davis, and C. L. Schelske. 1980. Phytoplankton productivity and standing crop in the vicinity of the Lake Huron-Saginaw Bay front. Journal of Great Lakes Research **6**:232-246.
- Munawar, M. and I. F. Munawar. 1982. Phycological studies in lakes Ontario, Erie, Huron, and Superior. Canadian Journal of Botany **60**:1837-1858.
- Nalepa, T. F. and G. L. Fahenestiel. 1995. *Driessina polymorpha* in Saginaw Bay, Lake Huron ecosystem: Overiew and perspective. Journal of Great Lakes Research **21**:411-416.
- Nalepa, T. F., J. A. Wojcik, D. L. Fanslow, and G. A. Lang. 1995. Initial colonization of the zebra mussel (*Dreissena polymorpha*) in Saginaw Bay, Lake Huron: Population recruitment, density, and size structure. Journal of Great Lakes Research 21:417-434.
- NASA. 2013. MODIS Specifications.*in* N. A. a. S. Administration, editor., National Aeronautics and Space Administration.
- Nicholls, K. H., G. J. Hopkins, and S. J. Standke. 1999. Reduced chlorophyll to phosphorus ratios in nearshore Great Lakes waters coincide with the establishment of dreissenid mussels. Canadian Journal of Fisheries and Aquatic Sciences **56**:153-161.
- Paerl, H. W. and J. Huisman. 2008. Blooms like it hot. Science **320**:57-58.
- Pillsbury, R. W., R. L. Lowe, Y. D. Pan, and J. L. Greenwood. 2002. Changes in the benthic algal community and nutrient limitation in Saginaw Bay, Lake Huron, during the invasion of the zebra mussel (*Dreissena polymorpha*). Journal of the North American Benthological Society **21**:238-252.
- Pozdnyakov, D. V., A. V. Lyaskovsky, F. J. Tannis, and D. R. Lyzenga. 1999. Modeling of apparent hydro-optical properties and retrievals of water quality in the Great Lakes for SeaWiFS: A comparison with in situ measurements. Pages 2742-2744 Geoscience and Remote Sensing Symposium, 1999. IGARSS'99 Proceedings. IEEE 1999 International.
- Rogers, R. H. 1975. Application of Landsat to the surveillance and control of eutrophication in Saginaw Bay. National Aeronautics and Space Administration, Ann Arbor, Michigan.
- Rogers, R. H., N. J. Shah, J. B. McKeon, and V. E. Smith. 1976. Computer mapping of water quality in Saginaw Bay with Landsat digital data. Proceedings of the American Society of Photogrammetry, Annual Meeting:584-596.

- Sarnelle, O., A. E. Wilson, S. K. Hamilton, L. B. Knoll, and D. F. Raikow. 2005. Complex interactions between zebral mussel, *Dreissena polylmorpha*, and the harmful phytoplankter, *Microcystis aeruginosa*. Limnology and Oceanography **50**:896-904.
- Scavia, D. 1979. Examination of phosphorus cycling and control of phytoplankton dynamics in Lake Ontario with an ecological model. journal of the Fisheries Research Board of Canada **36**:1336-1346.
- Schelske, C. L., L. E. Feldt, M. S. Simmons, and E. F. Stoermer. 1974. Storm induced relationships among chemical conditions and phytoplankton in Saginaw Bay, Lake Huron. Proceedings of the 17th Conference on Great Lakes Research:78-91.
- Schelske, C. L. and J. C. Roth. 1973. Limnological survey of Lakes Michigan, Superior, Huron, and Erie. University of Michigan, Ann Arbor, MI.
- Skubinna, J. P., T. G. Coon, and T. R. Batterson. 1995. Increased abundance and depth of submersed macrophytes in response to decreased turbidity in Saginaw Bay, Lake Huron. Journal of Great Lakes Research **21**:476-488.
- Smith, V. E., K. W. Lee, J. C. Filkins, K. W. Hartwell, K. R. Rygwelski, and J. M. Townsend. 1977. Survey of chemical and biological factors in Saginaw Bay (Lake Huron). EPA Contract Number R802685, Cranbrook Institute of Science; Environmental Research Laboratory
- Stoermer, E. F. and E. Theriot. 1985. Phytoplankton distribution in Saginaw Bay. Journal of Great Lakes Research **11**:132-142.
- Suzuki, N. N., S. Endoh, M. Kawashima, Y. Itakura, C. D. McNabb, F. M. D'Itri, and T. R. Batterson. 1995. Discontinuity bar in a wetland on Lake Huron's Saginaw Bay. Journal of Freshwater Ecology **10**:111-123.
- USEPA and USGS. 2005. National Hydrography Dataset Plus NHDPlus Version 1.0.
- Van Sickle, J. and C. B. Johnson. 2008. Parametric distance weighting of landscape influence on streams. Landscape Ecology **23**:427-438.
- Vanderploeg, H. A., J. R. Liebig, W. W. Carmichael, M. A. Agy, T. H. Johengen, G. L. Fahnenstiel, and T. F. Nalepa. 2001. Zebra mussel (*Dreissna polymorpha*) selective filtration promoted toxic *Microcystis* blooms in Saginaw Bay (Lake Huron) and Lake Erie. Canadian Journal of Fisheries and Aquatic Sciences 58:1208-1221.
- Vanderploeg, H. A., T. F. Nalepa, D. J. Jude, E. L. Mills, K. T. Holeck, J. R. Liebig, I. A. Grigorovich, and H. Ojaveer. 2002. Dispersal and emerging ecological impacts of Ponto-Caspian species in the Laurentian Great Lakes. Canadian Journal of Fisheries and Aquatic Sciences 59.

Wynne, T. T., R. P. Stumpf, M. C. Tomlinson, R. A. Warner, P. A. Tester, J. Dyble, and G. L. Fahnenstiel. 2008. Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. International Journal of Remote Sensing **29**:3665-3672.

### CHAPTER 5 SUMMARY AND SYNTHESIS: USE OF REMOTE ESNSING IN ECOLOGICAL ASSESSMENT National Lakes Assessment chlorophyll and Secchi depth predictive models

National Landsat boosted regression tree (BRT) models of chlorophyll (chl) concentration and Secchi depth are able to predict chl better than linear regression models for inland lakes in the United States (chl:  $R^2$  of 0.44 and a 0.76 ln-transformed  $\mu g/L$  RMSE; Secchi depth: R<sup>2</sup> of 0.52 and a 0.80 m RMSE). Applicability of the national BRT chl model for ecological analysis was demonstrated with a comparable total phosphorus-chl relationship whether using measured chl ( $R^2 = 0.58$  and 1.02 ln-transformed µg/L RMSE) or remotely sensed chl ( $R^2 = 0.56$  and 1.04 ln-transformed  $\mu g/L$  RMSE). Using multiple bands and ratios in the model increased predictive power. There is, however, much improvement to be made in broad-scale chl and Secchi depth models for lakes with widely variable physical, chemical, and biological attributes, including differing depth, dissolved organic carbon levels, and algal community composition. Correlations of national BRT model residuals with algal community composition, lake depth, and dissolved organic carbon levels, as well as time between sampling and image capture indicate that other factors may need to be addressed to further model development. Overall, subsetting the lakes into groups by similar regional characteristics did not improve chl or Secchi model performance.

Next steps in models for satellite measures of chl in inland lakes using BRT should consider incorporation of natural landscape variables, such as land use, into the model. Also, sampling each lake at multiple locations and over several times may help understand fluctuations algal biomass in each individual lake more clearly. Matching sample date as closely as possible with Landsat passover date may also help to reduce the noise in the relationship between predicted vs. measured chl. Finally, including field measurements of

121

water quality attributes, such as inorganic carbon, can help distinguish the chl signal from other substances in the water. Remote sensing water quality assessment tools can be valuable for limnological study, ecological assessment, and water resource management.

# Landsat and MODIS boosted regression tree models for chlorophyll prediction in the Great Lakes

Landsat and MODIS satellites are complementary platforms because of the long history of Landsat operation and the finer spectral resolution and image frequency of MODIS. Landsat and MODIS BRT models are better predictors of chl in the Great Lakes than other, readily available chl models. BRT models are also relatively easy to produce and to share. Of the three chl BRT predictive models that were made for the Great Lakes, the MODIS model was most accurate (0.85 cross-validation R<sup>2</sup> and 0.10 µg/L RMSE) and compared well to other models in the literature. BRT models for Landsat ETM+ and TM more accurately predicted chl (0.69 cross-validation R<sup>2</sup> and 0.55 µg/L RMSE) than the MSS model (0.66 cross-validation R<sup>2</sup> and 0.73 µg/L RMSE). All our Landsat models had favorable results when compared to models in the literature.

Though several of the current Great Lakes remote sensing models involving MODIS and Landsat were not publicly available for a direct comparison with our models, a comparison of this kind would be informative. Next steps should also include validation with external data and tests to see how widely the models can be applied. Model improvement may involve accounting for biological factors, such as position of the algae in the water column (Kutser et al. 2008). More consistent time of day for sampling may allow for more accurate detection with each image, and better means for comparison of longterm ecological trends. Also, application of BRT models could assist with discerning

122

different algal groups in the Great Lakes, whether they be used with Landsat, MODIS, or with satellite products with higher spectral or spatial resolution (Bergmann et al. 2004, Hunter et al. 2008, Wynne et al. 2008, Bracher et al. 2009) various algal types and simulated responses from Landsat and other less-spectrally sensitive satellites. Accounting for other color producing agents will elucidate the benefits of using all bands and ratios in model creation. Also, further tests should be done to determine if models that use all bands and ratios are less sensitive to shallow water due to incorporation of more spectral information. Application of BRT chl predictive models to the lifetime of both Landsat and MODIS could be useful understanding historical, long-term chl trends and to inform us of how climate change may alter ecosystems in the future.

# Using Landsat- and MODIS-derived chlorophyll to assess water quality in Saginaw Bay from 1972-2012

Deriving remotely-sensed chl concentrations from Landsat and MODIS satellite imagery enabled a long-term historical analysis of water quality trends in Saginaw Bay, which has a history of water quality problems and continues to experience regular, summertime algal blooms. Chl concentrations were predicted from Landsat 1973-2012 and MODIS 2000-2012 with boosted regression tree analysis. In a mixed-effect model, remotely-sensed chl concentration can be used with other physical, chemical, or biological parameters to test specific hypotheses about what drives algal biomass. Linear regression analyses show that Landsat-inferred chl decreased between 1973-1982 and that 8-day phosphorus load increased between 1972-2012. Simple linear regression showed influences of 8-day phosphorus load, distance to the Saginaw River mouth, and water temperature on satellite-inferred chl. Mixed effects models also show significant effects of 8-day phosphorus load, distance to the Saginaw River mouth, and Landsat-inferred water temperature on MODIS and Landsat-inferred chl, however, the results suggest that a large proportion of the random spatial effects are unaccounted for.

In future studies, variables such as water circulation patterns or internal nutrient cycling could be added to the mixed-effects model to help elucidate the cause for spatial correlations in chl. Future models could also take into account the lag in the response of chl concentration to precipitation events and nutrient exposure. We would expect this connection to be particularly important in extreme weather events, including heavy storms or drought. Having more chl data for spring blooms might also be informative in seeing long-term trends. Finally, having satellite inferred-chl data for shallower depths could help elucidate any relationships due to the proximity of external river water and nutrient inputs. Using remotely-sensed chl in ecological analyses can help determine which factors have, historically, been the largest drivers of algal biomass production in Saginaw Bay, and, therefore, which may be the most critical for management in the face of climate change.

- Bergmann, T., G. Fahnentiel, S. Lorhenz, D. Millie, and O. Schofield. 2004. Impacts of a recurrent resuspension event and variable phytoplankton community composition on remote sensing reflectance. Journal of Geophysical Research **109**:1-12.
- Bracher, A., M. Vountas, T. Dinter, J. P. Burrows, B. Schmitt, R. Röttgers, and I. Peeken. 2009. Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data. Biogeosciences **6**:751-764.
- Hunter, P. D., A. N. Tyler, M. Présing, A. W. Kovács, and T. Preston. 2008. Spectral discrimination of phytoplankton colour groups: The effect of suspended particulate matter and sensor spectral resolution. Remote Sensing of Environment **112**:1527-1544.
- Kutser, T., L. Metsamaa, and A. G. Dekker. 2008. Influence of the vertical distribution of cyanobacteria in the water column on the remote sensing signal. Estuarine, Coastal, and Shelf Sciences **78**:649-654.
- Wynne, T. T., R. P. Stumpf, M. C. Tomlinson, R. A. Warner, P. Tester, J. Dyble, and G. L. Fahnenstiel. 2008. Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. International Journal of Remote Sensing 29:3665-3672.