Lake Clarity Assessment in Minnesota’s Northern Superior Forest Ecoregion using MODIS Satellite Imagery

M. Taylor Long
Peter Marinello
FR 3262/5262 Final Project
December 10, 2012
INTRODUCTION

Monitoring lake clarity represents an important part of assessing the overall water quality of lakes. In a state like Minnesota, where lakes are numerous, monitoring lake clarity across entire regions requires a more economically feasible approach than on-the-ground observation. Researchers at the University of Minnesota and elsewhere have demonstrated that lake clarity can be monitored successfully through the analysis of remote sensing satellite imagery (Kloiber, 2002; Sawaya, 2003; Olmanson, 2008; Knight, 2012). Analyzing satellite imagery offers a reasonable means of assessing lake clarity across large areas to conduct comparisons on a broad spatial scale. Furthermore, the availability of frequent, recurring satellite imagery allows for temporal comparisons to detect changes over many years.

Olmanson (2008) demonstrated how lake clarity could be assessed using Landsat imagery for the entire state of Minnesota over a 20 year time span at five year intervals. Knight and Voth (2012) demonstrated how a similar analysis could be done using imagery captured by the MODIS sensor aboard the Terra (EOS AM) satellite. Knight’s study looked at a number of different images within one year, highlighting the benefits of the MODIS sensor’s higher temporal resolution. Based on the foundations established by these two papers, we decided to conduct a lake clarity analysis limited to the Northern Superior Forest Ecoregion of Minnesota and compare 3 years of August imagery from 2001-2011 (see Study Area image below).

Purpose of Study

By looking at lake clarity within one common ecoregion we hope to shift the focus of our analysis away from ecological differences to more subtle spatial differences. Within the Northern Superior Forest Ecoregion, we can compare lake water clarity on a finer scale than Knight and Voth’s 2012 analysis. Our hope is that this regional analysis might prove useful for future studies, possibly comparing lakes located inside of state and federal protected lands with those located outside of these boundaries in unprotected or suburban areas. By conducting a change analysis on images taken from three separate years (2001, 2006, 2011) over a ten year period, we can look for changes occurring over time and explore whether or not those temporal changes have a spatial dimension as well.

The initial identification of lake clarity and change for these Northern Minnesota lakes, both inside and outside of protective boundaries, has the potential to play a key role in current and future studies such as rural development impact studies or the proposed copper mining within and near the Boundary Waters Canoe Area Wilderness (BWCAW) and the St. Louis River watershed. Our study will focus on the process of creating a baseline change analysis for lake water clarity within our study area. Future studies could be conducted, for instance: to determine if there is detectable lake water clarity changes within the BWCAW on lakes with close proximity to future copper mines, in order to detect any mining water pollution. Whether or not these changes will be detectable or statistically relevant remains unknown. Assessing error and confidence will play an important part in our accuracy assessment and the discussion of our results.

We decided to follow the techniques outlined by Knight and Voth (2012) to assess lake clarity using a combination of secchi disk transparency (SDT) data and MODIS imagery. SDT measurements are commonly used to infer lake clarity. By comparing SDT measurements with MODIS sensor values for the same lake location, it is possible to create an equation to predict SDT readings for all other lakes. With these values, one can roughly assess the average lake clarity for each lake throughout a broad region.

We took advantage of this process to conduct our analysis of lake clarity. We chose to analyze imagery from the same summer period (August 13th-20th) for 2001, 2006 and 2012. We also chose to examine lake clarity differences between four analysis regions: the Boundary Waters Canoe Area Wilderness, National Forest Lands, State Forest Lands, and Other (remaining study area that does not fall within any of the first three boundaries).

Hypothesis

We hypothesize that our analysis of MODIS imagery may detect an overall difference in lake clarity between the three years of study, and that this difference may have a general downward trend, indicating a decrease in lake clarity, with consideration to increasing human impact on these water features over time. We also hypothesize that our analysis will detect lake clarity differences related to the management boundaries within which a lake is located.
Figure 1. These maps show the different boundaries used in our analysis. The top left map shows the Northern Superior Forest Ecoregion — our selected study site. The remaining maps show our analysis regions of the Boundary Waters Canoe Area Wilderness (light green), the National Forest Lands (orange), and the State Forest Lands (dark tan).
METHODS

Data Selection
To compare lake clarity within each analysis region, we downloaded boundary data from the Minnesota Department of Natural Resources (DNR) Data Deli website. These shapefiles included boundaries for our study area (The Northern Superior Forest Ecoregion), the Boundary Waters Canoe Area Wilderness, National Forests, and State Forests. Additionally, we downloaded the DNR 100K Hydrography dataset for Minnesota, including boundaries of all rivers, streams, wetlands and lakes. We clipped all these shapefiles to our study area.

Following the methods set forth by Knight and Voth (2012), we chose to use Moderate Resolution Imaging Spectroradiometric (MODIS) imagery which is collected by the National Aeronautics and Space Administration's (NASA) Terra and Aqua satellites. The MODIS sensor records spectral responses in 36 bands, at 250m, 500m and 1km spatial resolution. MODIS bands are narrower, with a higher radiometric resolution (12 bit rather than the 8-bit of Landsat imagery). MODIS imagery is available at no cost and is provided in single and multi-day composite products, of varying spatial and spectral resolutions. Knight and Voth mention that previous studies have used higher resolution Landsat-7 imagery to monitor lake clarity (Kloiber, 2002), but MODIS imagery's lower spatial resolution (500 meters) offers higher temporal resolution (daily images of the same area). Daily images make it more feasible to collect clear (cloudless) repeat imagery at the same time of year. The large image swath also eliminates the need to stitch together multiple images of our study area, which was fully contained in one MODIS image.

One of the greatest challenges posed by satellite image analysis is obtaining cloudless imagery, especially during wet summer months in Northern Minnesota. Because of the MODIS satellite's daily revisit capabilities, NASA offers 8-day composite images which eliminate as many clouds as possible, through a series of pre-processing algorithms provided by NASA. This allowed us to acquire nearly cloud-free imagery for our study area, over the same calendar days from three separate years, and avoiding lots of painstaking pre-processing. Each pixel in these images represents the most cloudless capture taken within an 8-day period (NASA, 2012). Given the scope of our project, we chose to use this product to save time, acknowledging that these images could still contain some cloud cover. The benefits of using MODIS imagery (large swath, high temporal resolution/daily revisit time) outweighed the costs (lower spatial resolution and timely pre-processing) for our purposes.

Data Preparation
We downloaded the MOD09A1 - MODIS Surface Reflectance 8-Day L3 Global 500m from NASA's Reverb Echo site, choosing an 8-day temporal window from August 13th-20th for 2001, 2006, and 2011. We then drew a bounding box around our study area, which extended roughly from the Northwest corner of Lake of the Woods (the northernmost lake in Minnesota), South to the just below perpendicular with Lake Mille Lacs, and east to include the arrowhead of Minnesota. This gave us a rough-enough definiton of our study area to find our MODIS granules. The results show two separate granules for our area (h11 and h12) of which we used h11, which contained the entirety of our study area, quite conveniently.

After downloading our MODIS images we reprojected them from their original sinusoidal projection into NAD83 UTM Zone 15N using ArcMap 10.1. Continuing to use ArcMap, we then simplified the layer stack in each image leaving only bands 1-7, the bands typically used in MODIS imagery classification (See: Discussion Section - Clouds and Haze). Reviewing the imagery, we encountered the same concerns with Band 5 and Band 7 as Knight and Voth (2012): both of these bands had considerable artifacts throughout the scene that seemed to be related to a sensor issue. Following Knight and Voth's methods (2012), we removed both bands from our layer stack to keep this artifacting from impacting the results of our classification. Finally, we clipped each year's image to our study area.

Because our MODIS imagery only offered 500 m spatial resolution, we anticipated removing some smaller lakes from our analysis to minimize the presence of too many mixed land and water pixels. As recommended by Knight and Voth (2012) we chose to eliminate all lakes with an area of less than 400 acres from our analysis. To do this, we altered the Minnesota hydrology data by removing all water features without a lake designation. Then we calculated the area of each remaining lake polygon and removed all features of less than 400 acres, leaving only lakes greater than 400 acres. We used this layer later to select acceptable secchi disk transparency data points and remove clear water pixels from smaller lakes from our analysis.
Unsupervised Classification

We used ERDAS Imagine to conduct an unsupervised classification on each year’s image. These stacked images contained only MODIS pixels within our study area for Bands 1, 2, 3, 4, and 6. We performed an ISODATA unsupervised classification set to create 20 clusters with a maximum of 50 iterations. We then examined the resulting classified image and determined that Class 2 represented clear water pixels (the same class was found to represent clear water in Knight and Voth, 2012). Based on this classification, we manually re-classified the image, setting all the Class 2 water pixels to a value of 1 and all other pixels to a value of 0. I repeated this process for each image of our three study years.

We then returned to ArcMap to create a mask for each image's water pixels. Using the classified images, we created a polygon layer for each image representing all of the clear water pixels with a value of 1. We used these clipping layers to return to each original MODIS image and clip out the clear water pixels from Band 1 and Band 3. As demonstrated by Knight and Voth (2012), Band 1 (red) and Band 3 (blue) offer useful information for correlating MODIS imagery with secchi disk transparency data. Our clipping resulted in two clipped images for each year, representing only Band 1 and Band 3 values over clear water pixels.

Secchi Disk Transparency Data

Secchi disks are used as a simple means of collecting data on lake clarity. These black and white disks are lowered into the water slowly until they disappear from view – the depth at which they disappear is recorded as the Secchi Disk Transparency (SDT) and used to make inferences about an entire lake’s water quality. We requested SDT data from the Minnesota Pollution Control Agency by email, which was collected through the MPCA’s Citizen Lake-Monitoring Program (CLMP). This program utilizes citizen volunteers to collect SDT readings to assess water clarity in lakes across Minnesota. Measurements for over 1,000 lakes exist in this volunteer-created data source, and this is the only source of water clarity data for many of the lakes. Volunteer-based programs that collect SDT data throughout the summer have been shown to produce high-quality, trustworthy Secchi data for many lakes that would not otherwise be monitored. Each CLMP database entry includes the sample date, GPS location, and SDT(depth) in meters (MPCA CLMP, 2012).

Using this data, we could find SDT readings taken close to our image capture dates and match actual SDTs with values from Band 1 and Band 3 of our MODIS images to predict SDT values for all other pixels in each image. For each year of SDT data, we identified all points collected during the same date range as our 8-day composite images (August 13th-20th, 2001, 2006, 2011). We then converted this data to point locations in ArcMap used both our Lakes >400 Acres layer and our

Figure 2. Altered 100K hydrography shapefile showing only lakes with an area greater than 400 acres.
clear water pixel mask for each image to select points that did not fall over an acceptable clear water pixel captured during their corresponding image year. We did this for all three years, giving us 206 SDT points for our 2001 lakes, 254 points on the 2006 image, and 177 on the 2011 image. The number of points varies for each year due to the variation of water-classification for each year’s image as well as variations in secchi disk data collection year to year (See: Discussion Section - SDT Data Selection).

In the interest of time, we chose to use all of our filtered SDT points in our regression analysis. Knight and Voth (2012) used bathymetry contours and aerial imagery to hand-pick acceptable SDT points for analysis, but we decided to bypass this step in the interest of maintaining the scope of this project. The impact this may have on our results is discussed below.

**Extracting Band Values for Regression**

Before extracting band values from each SDT location, we followed the Knight and Voth (2012) paper by using the Focal Statistics tool in ArcMap to smooth out the variation for each pixel value in each red and blue image. This tool uses a 3 x 3 kernel to calculate an average value for each pixel based on the 8 surrounding pixels’ values. Because these red and blue images were already clipped to contain clear lake pixels, pixels on the edge of a lake would only take into account the value of the lake pixels around them, ignoring the “no data” values in their calculation of a surrounding average. This process helped lessen the effect that edges or small islands might have on creating mixed pixel values.

With red band and blue band clear-water pixels prepared from each date’s image, we used the Extract Multi Values to Points tool in ArcMap to retrieve DN values from each SDT location for 2001, 2006, and 2011. Because satellite imagery can vary from one image to the next, we conducted a separate regression analysis for each image. We loaded the extracted values and their corresponding SDT values into SPSS and conducted a bivariate linear regression using SDT (meters) as our dependent variable and the Band 1 values and Band 3 values as our independent variables. The results of this regression analysis are presented below.

![Figure 3. Regression analysis results of SDT and corresponding Band 1 and Band 3 values.](image)

<table>
<thead>
<tr>
<th>Image Date</th>
<th>Regression Equation</th>
<th>$R^2$</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001 (n=206)</td>
<td>$Y = (-0.025 \times \text{Band 1}) + (0.28 \times \text{Band 3}) + 3.822$</td>
<td>0.284</td>
<td>1.50</td>
</tr>
<tr>
<td>2006 (n=254)</td>
<td>$Y = (-0.023 \times \text{Band 1}) + (0.14 \times \text{Band 3}) + 5.190$</td>
<td>0.434</td>
<td>1.10</td>
</tr>
<tr>
<td>2011 (n=177)</td>
<td>$Y = (-0.026 \times \text{Band 1}) + (0.26 \times \text{Band 3}) + 4.662$</td>
<td>0.344</td>
<td>1.28</td>
</tr>
</tbody>
</table>

We used the Raster Calculator tool in ArcMap to insert the above regression equation and calculate a predicted SDT values for the entirety of our image pixels. Our newly processed images now contained clear water pixels where each pixel value represented a predicted SDT depth (m) for that location.

**Trophic State Index**

SDT depth measurements gathered by MPCA are typically converted into a number on Carlson’s Trophic State Index (TSI) to describe the quality and productivity of a lake. While collecting water samples for chemical analysis offers the most accurate calculation of TSI, SDT data can be substituted for reasonably accurate results using the following equation presented by Carlson (1977):

$$\text{TSI(SDT)} = 60 - (14.41 \times \ln(\text{SDT m}))$$

TSI values range from 0 to 100 with low values representing clear, high water quality lakes and higher values representing low visibility, low water quality lakes. The table below summarizes the meaning of each range of TSI values.
Figure 4. Relationship between Carlson’s Trophic State Index (TSI) and other lake qualities. Copied from Knight and Voth (2012), derived from Minnesota Pollution Control Agency (2011) and Lake Access (2012).

<table>
<thead>
<tr>
<th>Lake Trophic State</th>
<th>Oligotrophic Lakes</th>
<th>Mesotrophic Lakes</th>
<th>Eutrophic Lakes</th>
<th>Hypereutrophic Lakes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water Quality</td>
<td>Extremely High</td>
<td>Moderate</td>
<td>Poor</td>
<td>Extremely Poor</td>
</tr>
<tr>
<td>Photosynthetic Productivity (TSI)</td>
<td>Low (&lt;30-40)</td>
<td>Intermediate (40-50)</td>
<td>High (50-70)</td>
<td>Extremely High (&gt;70)</td>
</tr>
<tr>
<td>Nutrient Levels</td>
<td>Low</td>
<td>Intermediate</td>
<td>High</td>
<td>Extremely High</td>
</tr>
<tr>
<td>Typical Lake</td>
<td>Very clear, deep lakes</td>
<td>Seasonal algae blooms, various lake depths</td>
<td>Green water, shallow lakes</td>
<td>Green water, shallow lakes, summer blooms of blue-green algae (toxic) and surface scum.</td>
</tr>
<tr>
<td>Secchi Transparency</td>
<td>High</td>
<td>Intermediate</td>
<td>Low</td>
<td>Very Low</td>
</tr>
</tbody>
</table>

With our three SDT-calculated images, we again used the Raster Calculator tool to convert each SDT pixel value into a TSI pixel value. The resulting image for each year represented TSI values for all clear water pixels throughout our study area. We reclassified this image into five colored categories based on Knight and Voth’s color classification scheme.

Figure 5. TSI Color Classification Scheme.

<table>
<thead>
<tr>
<th>Color</th>
<th>TSI Category</th>
<th>Lake Clarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grey</td>
<td>&gt;100</td>
<td>Considered invalid. No quality inferences made.</td>
</tr>
<tr>
<td>Red</td>
<td>70-100</td>
<td>Very Low</td>
</tr>
<tr>
<td>Yellow</td>
<td>60-70</td>
<td>Low</td>
</tr>
<tr>
<td>Green</td>
<td>50-60</td>
<td>Low</td>
</tr>
<tr>
<td>Blue</td>
<td>40-50</td>
<td>Moderate</td>
</tr>
<tr>
<td>Dark Blue</td>
<td>0-40</td>
<td>High</td>
</tr>
</tbody>
</table>

Classifying each image by the above color scheme revealed a number differences in TSI pixel values within larger lakes. Given that single SDT readings are typically used to develop a TSI value for an entire lake, we followed the methods of Knight and Voth (2012) in calculating an average TSI value for each lake based on all the individual pixel values within it. Because of the high potential for error in any one pixel’s TSI value, lake averages represent a better general assessment of probable lake clarity. For the purposes of our change analysis, we avoided any per pixel statistics, focusing results based only on the number of lakes falling into each of the five TSI categories listed above.

Processing of Results

With average TSI values calculated for each lake (or lake pixel cluster) we could begin our temporal and regional analysis. We used the Lakes >400 Acres polygon layer to extract each lake’s average TSI value from the underlying classified pixels. We also used the same Lakes >400 Acres layer to extract data on whether or not any given lake’s centroid fell within one of our analysis areas. For example, lakes with center points inside of the National Forest boundaries were designated as such by adding a “Yes” value to the National Forest field in the Lakes >400 acres table. Collecting all of this data allowed us to assess the number of lakes in each region that fell within each of the five TSI categories. Because the number of lakes in each region vary considerably, we found it more useful to calculate percentages. This way, we could ask “What percentage of lakes in the Boundary Waters Canoe Area Wilderness were classified as Yellow (60-70 TSI) in 2001?” and compare this answer to other regions as well as to other years. Our results are highlighted below.
RESULTS

Large versions of results maps and graphs are included in the Appendix at the end of this document. Highlighted results are included below.

Figure 6. The 2006 Per Pixel TSI classification of lakes within our study area from 2006 based on our SDT regression and TSI conversion.

Figure 7. The 2006 Lake Average TSI classification, represented by the average TSI value for each lake, symbolized by class.
Figure 8. Percentage of Lakes in each TSI category for each year, accompanied by a table below with actual percentage values listed for each TSI category.

The above graph and table summarizes the percentage of lakes in each TSI category for each year. For each lake in the study area, an average TSI value was calculated so each lake could be assigned to one category. The total number of lakes in each category were added up so that category percentages could be determined for each year. These results show a sharp decline in the percentage of clearest lakes (0-40 TSI) as well as a rise in the percentage of poorer clarity lakes (50-60 TSI). Also notable is the rise and fall of lakes in the 70-100 TSI category. Our data suggest a general downward trend in regional lake clarity over the ten year period.
Figure 9. Percentage of Lake in each TSI category in 2006, summarized separately by analysis region.

Our results also revealed interesting regional differences in the number of lakes that fell in different TSI categories. The figure above shows our results for 2006, a year in which this regional difference was most pronounced. Lakes that fall within the Boundary Waters Canoe Area Wilderness were mostly classified in the clearer water categories (0-40, 40-50 TSI), but the portion of clearer lakes declines in National Forests, further in State Forests, and further still in the “Other Lakes” category (lakes not falling within any of the analysis region boundaries). Our data also suggest a higher portion of lakes in the poor clarity categories (50-60, 60-70, 70-100 TSI) in the “Other Lakes” category. These data suggest that the regional differences we considered may be connected to differences in management rules governing the areas surrounding any given lake.
**Figure 10.** This map shows the results of our per lake change analysis for 2001-2006. Lakes are colored based on whether or not they rose or fell by a full TSI classes (based on the five-class scheme established above).

![Changes in TSI Category 2001-2006](image)

**Figure 11.** This map shows the results of our per lake change analysis for 2006-2011. Lakes are colored based on whether or not they rose or fell by a full TSI classes (based on the five-class scheme established above).

![Changes in TSI Category 2006-2011](image)
The above change analysis maps offer a sense for the kinds of change taking place over our study period for each lake. Comparing the map for 2001-2006 with 2006-2011, we could see that more lakes increased by one TSI category during the first change period than in the second change period. Also visible in the second map is the greater presence of lakes that dropped one TSI category from 2006-2011 (See: Discussion Section - Results - Temporal Trends, and Results - Change Analysis). These trends are highlighted more clearly in the graphs below.

Figure 12. Change analysis summary of TSI class changes across the entire study area for each period of change.

This graph shows the quantitative results of our change analysis. Between each study date, we recorded whether or not each lake’s TSI category changed positively or negatively. Our analysis suggests that for both periods of change, more lakes increased in TSI category than decreased. However, from 2006-2011 there were more lakes that decreased in TSI category than there had been in the previous period. Similarly, 2006-2011 showed fewer increases in TSI category than in 2001-2006.
The two graphs above show the results of our change analysis by region. The results of the 2006-2011 period show the most increase in TSI category in the “Other Lakes” analysis region. However, this same region shows the least increase in TSI category during the 2006-2011 period. Comparing Figures 13 and 14 — looking specifically at State Forest Lakes and Other Lakes — suggests that some of the increases in TSI category during 2001-2006 may have been reversed in 2006-2011 (See: Discussion Section - Results - Analysis Region Trends).
ACCURACY ASSESSMENT

Following the guidelines laid out in Knight and Voth (2012), we conducted an accuracy assessment on one of our three images, selecting 2006 because of its interesting regional differences (See: Discussion Section - Accuracy Assessment Results). We randomly divided up our 177 original SDT points for 2006 into two subgroups: A and B. We used the points from subgroup A to extract Band 1 and 3 values, and conduct a linear regression to establish a new formula for predicting SDT, then TSI for all lake pixels in the 2006 image. Then, we used the SDT points from subgroup B to compare their actual TSI with their predicted TSI based on the equation generated from the subgroup A points. This process was repeated by switching subgroups A and B so all point values were compared to predicted values based on the other set. From this analysis we were able to compare actual and predicted TSI values to see the percentage of correct classifications.

Figure 15. Accuracy Assessment regression analysis results.

<table>
<thead>
<tr>
<th>2006 Image (n=254)</th>
<th>R²</th>
<th>Standard Error</th>
<th>Range SDT (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subset A (n=127)</td>
<td>.401</td>
<td>1.21</td>
<td>0.46 - 10.21</td>
</tr>
<tr>
<td>Subset B (n=127)</td>
<td>.475</td>
<td>.099</td>
<td>0.46 - 7.32</td>
</tr>
</tbody>
</table>

Figure 16. Accuracy Assessment table for 2006 image.

<table>
<thead>
<tr>
<th>2006 Image (n=254)</th>
<th>Trophic Class Accuracy</th>
<th>Within 1 TSI</th>
<th>Within 2 TSI</th>
<th>Within 5 TSI</th>
<th>Within 10 TSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subset A (n=127)</td>
<td>55%</td>
<td>12%</td>
<td>25%</td>
<td>64%</td>
<td>92%</td>
</tr>
<tr>
<td>Subset B (n=127)</td>
<td>54%</td>
<td>21%</td>
<td>34%</td>
<td>69%</td>
<td>94%</td>
</tr>
<tr>
<td>2006 Averages</td>
<td>54.5%</td>
<td>16.5%</td>
<td>29.5%</td>
<td>66.5%</td>
<td>93%</td>
</tr>
</tbody>
</table>

The accuracy assessment shows that while Trophic Class accuracy was somewhat low (54.5%), the actual TSI and predicted TSI value for any given point was within 10 in 93% of cases.

DISCUSSION

Our data do support the hypothesis that lake clarity varies within the Northern Superior Forest ecoregion both temporally (between 2001, 2006 and 2011) and spatially (between BWCAW, National Forest, State Forest, and Other). These data suggest that the percentage of lakes classified in the low clarity range (50-100) has increased from 2001 to 2011 and that in 2006 there was a clear difference in the percentage of high-moderate clarity (0-50) lakes between our analysis regions, with BWCAW containing the highest percentage of clear lakes. However, because of our low R² values in our regression analysis, and our 54.5% Trophic Class Accuracy, these findings can only be used to recommend exploring these trends further through more robust, accurate studies.

Clouds and Haze

Clouds and haze represent one serious potential for error in our data. In the interest of time, we chose to use 8-day composite imagery and “assume” that these images contained mostly clear, cloudless pixels. To create this MODIS L3 Global 8-Day composite, a series of algorithms are applied to capture the “best” observation during an eight-day period, for every cell in the image, which helps to reduce or remove clouds from the scene (Yale University, 2012). The 8-day composites have the same seven spectral data bands as the regular single day imagery as well as 6 additional bands of quality control, solar zenith, view zenith, relative azimuth, surface reflectance 500m state flags, and surface reflectance day of the year (Yale University, 2012).

It is likely that these L3 images still contained pixels for which there was no clear measurement recorded during the 8 days between August 13th and 20th for each year. Examining these images, we found that because of their composite nature, visually detecting cloud effects was actually more difficult because cloud interference was more pixelated than in a natural
image. Given this difficulty, and the limitations of our time and expertise, we accepted the error that some images may still contain cloud interference, and performed our unsupervised classification on the images without any spectral corrections. With time to conduct a more robust study, we would choose single day images carefully by their absence of cloud cover and we would manually inspect each image to eliminate all lakes with cloud interference from our study. This process would require great effort in identifying acceptable images within a small time frame for each year, but would likely yield more accurate TSI classification results.

SDT Data Selection

Our method of selecting SDT data represents another potential source of error in our study. We filtered SDT data by time and location, but not by lake depth as Knight and Voth (2012) did for their study. For our study, we first selected all SDT points collected within the time range of our images (August 13th-20th), then filtered these to leave only those points that coincided with a clear water image pixel for that year. At this point in Knight and Voth’s 2012 study, they compared SDT locations to the DNR’s bathymetry contour data to manually selected up to 50 points for each image that fill over water with a depth greater than 2m. This reduced the influence of mixed land/water pixels near lake edges, islands and shallow areas, providing more accurate Band 1 and Band 3 values to compare with SDT values for regression. In the interest of time, we chose to use all SDT points that fit the time and location criteria, without manually selecting an ideal subset. We believe this choice negatively impacted our $R^2$ value for each image’s regression analysis by likely including some Band 1 and Band 3 values that represented mixed pixels. Therefore, this accepted error resulted in less accurate equations for predicting SDT in all other lake pixels.

After performing our first regression analysis of the 2011 image which yielded an $R^2$ value of .344, we did attempt to improve our $R^2$ results by performing a quick manual selection of 30 SDT points. Without consulting bathymetry data, we simply picked SDT points that seemed to fall well within lake pixels and tried to select a range of points from all over the image. We thought this might improve our $R^2$ value slightly, but this second regression analysis actually yielded an even lower $R^2$ value, so we continued on, applying our first selection method to the remaining images. This accepted error certainly impacted the accuracy of each lake pixel’s predicted SDT value, but we feel this degree of error was lessened by our use of lake average pixel values in our analysis. Still, this impacts the accuracy of our assessment and is a likely contributor to the 54.5% Trophic Class Accuracy we found in our 2006 image.

Results - Temporal Trends

Our data suggest that the number of clear and moderate clarity lakes may have decreased from 2001-2011. This temporal change is most clearly represented in Figure 8. In 2001, 95% of lakes throughout our entire study region were classified as either high or moderate clarity (0-40 or 40-50 TSI). In 2006, this percentage dropped to 84%, and in 2011 it dropped to 74%. While these data show a clear downward trend in the percentage of clear lakes across the study area, more robust studies would be needed to confirm these findings.

Knight and Voth (2012) found that inter-annual lake clarity can vary widely. Despite the fact that we chose the same 8 day period of each year (August 13th-20th), hoping to capture peak algal blooms, we acknowledge that intra-annual and inter-annual variation could skew the results of our study. To conduct a stronger comparison of year-to-year variation, it would be better to calculate average TSI values for each lake using 3 or more images from July, August, and September of each year. This would help minimize the impact that extreme weather events like floods and fires might have on any one image.

Another way to improve the investigation of temporal differences in regional lake clarity would involve collecting images from each year over a ten year period. Three collection dates over ten years represents the bare minimum for detecting any kind of trend, and even then, more images would give a more complete impression of the range of changes possible between one year and the next. Given more time, we would replicate our analysis on images during our gap years: 2002-2005 and 2007-2010. Based on the trend we observed in our analysis, we would recommend further study to replicate these results with more sample images for better statistical validity.
Results - Analysis Region Trends

Our data also suggest that the rules governing the land surrounding lakes may influence their likelihood of clarity. Figure 9 shows this pronounced difference in the 2006 data, but the same general trends were seen in the 2001 and 2011 data. In 2006, 98% of lakes within the Boundary Waters Canoe Area were classified as high or moderate clarity (0-40 or 40-50 TSI). For National Forest lakes, this number was 92% and for State Forests, 82%. Among all other lakes not included in any analysis region, only 78% were classified as having high or moderate clarity. We found it particularly interesting that there was a noticeable difference between lakes located in National Forest boundaries versus lakes located within State Forest boundaries. This trend suggests that summaries of lake clarity may differ based on the specific land management policies surrounding those lakes. Still, due to our low percentage of TSI class accuracy, even a 10% difference in the percentage of clear and moderate lakes should be treated with suspicion. Given these findings, we would recommend conducting more statistically valid analyses of lake clarity in State Forests and National Forests to see how different management practices may have impacted overall clarity trends.

Results - Change Analysis

The results of our change analysis (Figure 12) show an interesting overall trend of more lakes rising in TSI categories during the period from 2001-2006 than from 2006-2011. However, these results are particularly questionable considering the troubling 54.4% rate of TSI class accuracy found in the 2006 image. While our change analysis looked at the percentage of lakes that changed TSI categories (based on our five, color-coded categories), it is likely that many of the +1 or -1 category changes are due to random noise introduced by clouds and haze or by the inaccuracy of our regression. With this in mind, it’s best to focus on changes of more than one TSI category, and our data reveal relatively few lakes that changed by this degree. From 2001-2006 only 2.4% of lakes changed by more than 1 TSI category and from 2006-2011, only 1.3% of lakes changed by this degree.

We did notice one interesting difference between the two change periods for State Forest lakes. In the 2001-2006 period, 3.5% of State Forest lakes increased by 2 or more TSI categories (none decreased by 2 or more TSI categories). Then, from 2006-2011, 1.4% of the same State Forest lakes decreased by 3 TSI categories, while only 1.4% increased by 2 TSI categories. The difference is best viewed by comparing Figures 13 and 14 above. While small changes are difficult to statistically justify, this finding deals with much more dramatic shifts in TSI that may actually reflect reality to some degree. Still, the effects of clouds and haze from one image year to the next could account for even some of these higher level changes, especially on smaller lakes. While our change analysis helped remind us that the results of our study should be treated with caution, it at least offered a different perspective on some of the trends observed in the previous figures.

Accuracy Assessment Results

We chose to perform our accuracy assessment on the 2006 imagery because it was used in both change analyses, and because it showed the clearest differences between analysis regions. Given our interest in these results, and the fact that this image’s regression analysis yielded the highest R² value (0.434), we wanted to assess the accuracy of at least this one image to determine the degree with which we could trust our results. We found that 93% of the SDT points were classified within 10 TSI of their actual TSI based on actual SDT measurements. However, only 66.5% were classified within 5 TSI of their actual TSI.

A separate measure determined the accuracy of the points’ placement within a given TSI class (one of the five we used for color classification). This measure is particularly important to our results because we based our analysis on any given lake's inclusion in one class or another. Our accuracy assessment revealed that only 54.5% of actual TSI values were classified into the same class through our equation. This number is predictably lower than the 62% Trophic Class Accuracy found by Knight and Voth (2012) and suggests that the trends we witnessed in our data cannot be considered conclusive.

It is important to remember that this accuracy assessment measures the rate of any one point being correctly classified based on its underlying pixel values. Our data, on the other hand, represent TSI averages for entire lakes, so any one pixel’s error is mitigated by being averaged with its surrounding TSI values. Still, this accuracy assessment was conducted on only the 2006 image – potentially our most accurately classified image based on R² value. It’s reasonable to assume that accuracy assessments
for the other two images could yield even lower percentages of accuracy. Given more time, we would complete accuracy assessments on each image and experiment with different corrections on our original imagery and SDT points to see if we could improve the accuracy of our results. Taking all this into account, our results still may contain some degree of truth, and certainly reflect patterns that make sense intuitively, but ultimately cannot be used to draw any real conclusions.

CONCLUSIONS

Using MODIS 500m resolution 8-day composites to assess lake water clarity can offer interesting insights into temporal and spatial patterns of lake clarity. Given our limited experience analysing MODIS imagery, we found that following the general approach established by Knight and Voth (2012) provided a feasible set of methods for conducting water clarity analysis. Because of our limited time frame, we chose to omit some steps, which introduced error and ultimately weakened the validity of our results. Nevertheless, our data offered interesting clues into trends that could warrant confirmation of our findings by further study, and gave us experience conducting a full remote sensing project.

Monitoring lake clarity in the Northern Superior Forest ecoregion is particularly important because it is a region of Minnesota that is known for its abundance of clear lakes. This study sought to identify how management boundaries might influence lake clarity, and we found some evidence that these man-made boundaries may play some role in lake clarity. Our data also led us to believe that there may be a general trend of reduction in lake clarity across the region during the last decade, but no real conclusions can be drawn without further study.

We feel that with the factors inherent to this project (16 week timeline, user inexperience, uncertain outcome and results) have generated some valuable lessons learned in regard to our project intent. Our level of experience has risen through application of lab skills and research and literature review, as well as trial and error. If future students wish to build upon the results we've found here, we recommend consulting with Maggie Voth early in the process to get a better sense of the technical tools needed to conduct this analysis and the time expected for each step. Mastering this process earlier in the semester may have allowed us to analyse more images, and certainly might have given us more time to improve our regression analysis. Overall, this project succeeded in providing interesting results that should pique the curiosity of anyone interested in how lake clarity varies throughout Northern Minnesota.

Lessons Learned

Our initial intent was to generate significant and insightful results with our project as stated in our hypothesis, hoping to see a significant difference between lake water clarity inside of protective boundaries versus outside of those boundaries. We have been humbled a bit, and are now acutely aware of the complexities and difficulties inherent with a remote sensing project. It is our feeling that the initial scope of this project and our projected results could be better attained with a higher level of accuracy if we had been more familiar with the complexities of the MODIS 8-day composite images. After conducting our analysis and exploring the reasons behind our low $R^2$ value, we returned to the MODIS imagery to explore potentials for error. The detailed description of these images mentions the importance of the quality assessment layers and validation layers included with the composite images. Given our lack of experience using MODIS data, our study ignored these quality assessment layers, focusing only on the spectral band layers of the image during data collecting and processing.

We failed to grasp the full impact of ignoring this information until near the end of our project with insufficient time to start over and learn how to process these images more accurately. Furthermore, we were unsure how much of an impact this would have on improving our accuracy, so we chose to acknowledge the source of error and move on. After reviewing the imagery later-on in the project, it seemed that this 8-day composite imagery may still contain the faint or residual presence of cloud reflections on some water bodies. This simple, yet significant issue may have influenced our low $R^2$ in our regression analysis, as well as our accuracy assessment. Here, we offer two possible solutions to our assumption that the MODIS imagery contains reflectance residuals on water bodies from clouds and haze:

The first way to more accurately prepare MODIS 8-day composite images would be to filter out pixels that still contain some effect from clouds and haze. Assuming the extra accuracy assessment layers could be used to identify any pixel’s amount of cloud interference, some tolerance level could be set to remove all pixels below a certain threshold of clarity. This kind
of preprocessing would likely yield better results by removing the residual cloud presence and improving the accuracy of regression and classification.

The second way we could have addressed the effect of clouds would be to spend more time searching for individual calendar days that provide clearer images. This method would involve careful review and manual processing to remove clouds and their reflection from all lakes in your area of interest. This would require a great deal of time to search and process many individual images in order to stitch together a perfect composite to classify and analyze. Thus, we are satisfied with our choice to use the MODIS 500m 8-day composite to conduct our project analysis, even though we have discovered that it may not have been the ideal image format for this type of project, or that further pre-processing of the images and more time to research and attempt new methods would most likely have yielded better results with these images.

In the end, this project left us with a deep appreciation for the challenges posed by remote sensing analysis, and for the skills and time required to execute a project of this nature. It also offered a glimpse into the great potential that this kind of analysis can offer if performed correctly.

**WORKS CITED**


METADATA

MODIS Imagery Metadata


Shapefile Metadata


DNR 100K Hydrography Layer, Shapefile, Accessed: 05 October 2012, MnDRN Data Deli: http://deli.dnr.state.mn.us/metadata.html?id=L390003370201


Secchi Disk Transparency Metadata

2001, 2006 and 2011 SDT data provided by Nancy Flandrick, Minnesota Pollution Control Agency. (Received electronically 29 November 2012).

Questions?

E-mail Taylor Long

longx598@umn.edu
Per Pixel TSI - 2006

<table>
<thead>
<tr>
<th>TSI (SDT)</th>
<th>Lake Clarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 100</td>
<td>Very Low</td>
</tr>
<tr>
<td>70 - 100</td>
<td>Low</td>
</tr>
<tr>
<td>60 - 70</td>
<td>Intermediate</td>
</tr>
<tr>
<td>50 - 60</td>
<td>High</td>
</tr>
<tr>
<td>40 - 50</td>
<td></td>
</tr>
<tr>
<td>0 - 40</td>
<td></td>
</tr>
</tbody>
</table>

NORTH

0 50 Miles 100
Lake Average TSI - 2001

<table>
<thead>
<tr>
<th>TSI (SDT)</th>
<th>Lake Clarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 100</td>
<td>Very Low</td>
</tr>
<tr>
<td>70 - 100</td>
<td>Low</td>
</tr>
<tr>
<td>60 - 70</td>
<td>Intermediate</td>
</tr>
<tr>
<td>50 - 60</td>
<td>High</td>
</tr>
<tr>
<td>40 - 50</td>
<td></td>
</tr>
<tr>
<td>0 - 40</td>
<td></td>
</tr>
</tbody>
</table>

NORTH

0  50 Miles  100
APPENDIX

Changes in TSI Category 2001-2006

TSI Index Category Changes

+3
+2
+1
Same TSI Category
-1
-2
-3
Insufficient Data

NORTH

0 50 Miles 100
Changes in TSI Category 2006-2011

TSI Index Category Changes

-3
-2
-1
Same TSI Category
+1
+2
+3
Insufficient Data
APPENDIX

Percentage of Lakes in Each TSI Category by Region - 2001

Analysis Region

Percentage of Lakes in Each TSI Category by Region - 2006

Analysis Region

Percentage of Lakes in Each TSI Category by Region - 2011

Analysis Region
APPENDIX

TSI Category Change Summary by Region - 2001-2006

Percentage of Lakes

Analysis Region

TSI Category Change Summary by Region - 2006-2011

Percentage of Lakes

Analysis Region