ASSESSING THE EFFECTIVENESS OF LANDSAT 8 CHLOROPHYLL-A RETRIEVAL ALGORITHMS FOR REGIONAL FRESHWATER MANAGEMENT

JONAH BOUCHER
Hamilton College, Clinton, NY 13323 USA

MENTOR SCIENTISTS: DRS. KATHLEEN WEATHERS[1] AND HAMID NOROUZI[2]
[1]*Cary Institute of Ecosystem Studies, Millbrook, NY 12545 USA*
[2]*The City University of New York, Brooklyn, NY 11201 USA*

Abstract. Predicting algal blooms has become a priority for municipalities, businesses, and citizens. Remote sensing offers solutions to the spatial and temporal challenges facing existing lake monitoring programs that rely primarily on high-investment *in situ* measurements. Techniques to remotely measure chlorophyll-a (chl-a) as a proxy for algal biomass have been limited to large water bodies in particular seasons and chl-a ranges. Thus, a first step toward prediction of algal blooms is generating regionally robust algorithms using in-situ and remote sensing data. This study explores the relationship between in-lake measured chl-a data in Maine and New Hampshire and remotely-sensed chl-a retrieval algorithm outputs. Landsat 8 images were obtained and then required atmospheric and radiometric corrections. Six existing algorithms were tested on a regional scale on nine scenes from 2013-2015 covering 169 lakes. Landsat 8's Bands 2-4 proved most useful for correlation with chl-a, and for late-summer scenes, existing algorithms accounted for nearly 90% of the variation in in-situ measurements. Imposing a chl-a floor and talking only samples from within a day of the satellite improved this relationship so that over 98% of variation was explained. A significant effect of the time of year on several indices was demonstrated. A sensitivity analysis revealed that a longer time difference between *in situ* measurements and the satellite image increased noise in the models. The quantification of these confounding influences points to potential solutions such as incorporating remotely sensed water temperature into models as a proxy for seasonal effects. These results suggest that remote sensing could be an effective and accessible tool for monitoring programs at the regional scale.

INTRODUCTION

Scientists and citizens alike are becoming increasing interested in monitoring and predicting changes in global water quality. Freshwater systems in particular provide a wide range of ecosystem services, from habitats to recreation to irrigation. Environmental stressors related to climate and land-use change are threatening many of these services, leading public and private organizations to rejuvenate or begin monitoring strategies. Many such efforts have begun in response to anthropogenic eutrophication from increasing nutrient inputs and higher water temperatures that are changing freshwater biota community dynamics (Dörnhöfer and Oppelt 2016). The effects of eutrophication can range from aesthetic annoyances like odor and color to toxic blooms that decimate wildlife populations and make the water unsuitable to drink (Shen et al. 2012). Monitoring programs have thus aimed to predict phytoplankton population change with the hope of preventing these problematic situations.

While *in situ* and primarily summer-time measurements have long been the basis of these monitoring programs, these collection methods are limited because of the high demands for time and labor and the cost of data collection. An alternative, or complement, to this cost-intensive approach uses remote sensing technology, which aims to use satellite imagery to derive water quality parameters. Remote sensing has shown promise for supplementing or replacing field data, but limited spatial resolution and optically complex inland waters have posed challenges to progress (Palmer et al. 2015).

Recent advances in the field of remote sensing have the potential to overcome some of these challenges. Landsat 8 (L8) is the most recent addition to the Landsat series of satellites, which have provided a publically accessible record of over four million images dating back to 1972. L8 features eight times better signal to noise ratio than previous iterations, a faster 12-bit quantization processing speed, and two new wavelengths (Roy et al. 2014). Although the 16-day repeat cycle of the satellite results in temporal limitations, the 30m x 30m spatial resolution has made L8 one of the most promising versions of remote sensing technology yet for inland water monitoring (Gerace et al. 2013, Feyisa et al. 2014, Concha and Schott 2015a, Andrzej Urbanski et al. 2016, Beck et al. 2016). The real practicality of Landsat 8 satellite data, though, remains to be fully tested.

The application of remote sensing to monitoring water quality is limited to measuring spectrally active water constituents since these satellites can only produce data about absorbance and reflectance of surfaces. Luckily, the issue of eutrophication of lakes has potential to be investigated with remote sensing due to the fact that chlorophyll a (chl-a) is a spectrally active compound in phytoplankton. Chl-a can be used as a proxy for phytoplankton biomass, therefore serving as an indicator of lake productivity (Beck et al. 2016). Many different algorithms have been developed to estimate chl-a values from reflectance outputs of specific bands from satellite data. Robust relationships have been developed in many cases (Duan et al. 2007, Bresciani et al. 2011, Keith et al. 2012, Tebbs et al. 2013) between measured and satellite-retrieved chl-a values, but their application is not necessarily accurate outside of the original area of study, often just a single water body.

Lakes in the New England region have long been monitored using traditional *in situ* methods. The Volunteer Lake Assessment Program (VLAP) has used volunteer-driven lake sampling sponsored by the Department of Environmental Services to monitor New Hampshire lakes since 1985. The organization is currently collecting monthly water samples and compiling annual reports on 176 different lakes in the state (New Hampshire DES 2015). Likewise, the Volunteer Lake Monitoring Program in Maine, with the help of over 1,200 active volunteers, has been monitoring more than 500 lakes since 1971. Of primary concern to the managers, citizens, and scientists involved is eutrophication, which is expected to be a growing problem in the region due to nutrient loading from land-use and climate change (Moore et al. 2014). Studies on lakes in the region have largely been limited to *in situ* case studies on specific lakes (Davis et al. 2006, Carey et al. 2009), making remote sensing technology for monitoring lakes a very promising too.

Monitoring small oligotrophic lakes remotely was suggested as early as 1989 (Vertucci and Likens 1989), but limitations on the resolution and accuracy of remote sensing technology precluded any further action at the time. In New Hampshire, early Landsat data were used to examine the relationship between satellite reflectance and lake transparency, but were limited by the band availability of Landsat (Schloss et al. 2002). More recent work has identified remote sensing of chl-a as a priority, but has used hyperspectral imaging from boats or aircraft rather than available satellite imagery (Bradt 2010, Keith et al. 2012). Urbanski et al. (2016) provided an initial assessment of L8's suitability for regional lake quality assessment, finding that remote sensing for trophic state assessment of lakes in Poland was possible, but often required a specific formula for each specific satellite scene.

This investigation seeks to utilize the currently available *in situ* and satellite data in the region to assess the strength of existing chl-a retrieval algorithms and their potential as predictive tools in over 150 small lakes. Such a study will contribute to improving the existing algorithms for water quality predictions and identify the existing benefits of using this kind of technology. We primarily hope to suggest to monitoring programs and remote sensing specialists ways to improve the applicability of remote sensing to the field of freshwater ecology.

METHODS

1. Satellite Data

Five Landsat images over Maine (Path 12 Row 29) and seven over New Hampshire (Path 13 Row 30) were acquired from USGS EarthExplorer (https://earthexplorer.usgs.gov/) (Table 1, Figure 1). The Level 1 GeoTIFF Data Products from the Landsat 8 OLI_TIRS sensor were downloaded for each image, except for the August 26, 2000 image which instead came from Landsat 7's Level 1 Product to provide a comparison between Landsat iterations. The images were processed using ENVI software. For each of Landsat 8's eleven different bandwidths (Appendix) Reflective Radiometric Calibration was performed to convert radiance – the surface brightness measured directly by the satellite – to the unit-less surface reflectance using metadata about the acquisition time and sun elevation when the image was taken (Harris Geospatial). A simple Dark Object Subtraction (DOS) was performed to minimize the effects of atmospheric haze (Giardino et al. 2001, Andrzej Urbanski et al. 2016).

2. In Situ Data

In situ chl-a data were obtained for New Hampshire from the Department of Environmental Service's Volunteer Lake Assessment Program (VLAP) database and for Maine from the Volunteer Lake Monitoring Program (VLMP) (Figure 1).

The 265 different chl-a samples in Maine were taken from 145 unique sampling points from 133 different lakes. Two scenes had on average 15 associated sampling points in common, and no two images shared more than 23 sampling points. In New Hampshire, the 90 chl-a samples were taken from 61 unique lakes that each had just one sampling spot. The scenes had on average below three sample sites in common and no two images shared more than six corresponding sampling points.

Chl-a measurements taken from lakes within the satellite image and sampled within four days for New Hampshire images and five days for Maine images were recorded (Figure 1). There was not a standardized procedure for chl-a measurement within either of the two programs, and samples were taken from depths ranging from 0.5m-10m depending on the lake (mainevlmp.org and des.nh.gov/organization/divisions/water/wmb/vlap/). Of the 265 *in situ* samples from the five Maine images, 87% were below 10ppb. Five of the six chl-a values above 30ppb were sampled in the two late August samples. These points were excluded from the regression analyses in order to minimize the effect of just a few points on the strength of the correlations. In New Hampshire, only four of the 90 samples had chl-a values above 10ppb, and none were above 13ppb (Table 1).

3. Assessing Algorithm Performance

Six existing chl-a retrieval algorithms were chosen to incorporate several different potentially ecologically-significant bandwidths (Table 2). The Surface Algal Bloom Index (SABI) and Normalized Difference Vegetation Index (NDVI) were developed for satellites with narrower bandwidths and were adapted to fit the wider bands of Landsat 8. The 3BDA (KIVU) and 2BDA algorithms were developed with Landsat 7TM bands, so the band math presented here substituted the corresponding Landsat 8 bands (Appendix). Kab1 and Kab 2 were best-fit algorithms built specifically for low-chl-a coastal settings, and were include to assess how well a highly specific algorithm might translate to other regions of study.

The VLAP and VLMP provided exact coordinates for where each *in situ* sample was taken, and ENVI Geospatial Software was used to gather the reflectance output values for each Landsat 8 band at each site. The "Compute Spatial Statistics" function was used to average the reflectance values over a 3x3 pixel grid surrounding the sampling point to account for any variability from the local mean that might exist

within just one 30m x 30m pixel. The individual band reflectances could then be combined to find the output in remote-sensing reflectance (*Rrs*) of each algorithm at each sampling location.

4. Sensitivity Analyses

Time Window. The nine satellite images from Table 1 were used for an analysis of the importance of the time window between sample acquisition date and satellite image date. Measured *in situ* chl-a values were plotted against *Rrs* output for each algorithm and the plots were fitted with quadratic regression curves. The strength of the correlation was judged using Pearson's R^2 and a p value. The few chl-a values above 30ppb were excluded from the analyzed data sets to examine relationships without the influence of outliers. For each image, the time window was progressively decreased down to one day, wherein only points sampled within one day of the satellite image were included.

Minimum chl-a. The same time window sensitivity analysis was conducted with the dataset restricted to only samples above 5ppb to investigate the sensitivity of algorithms to highly oligotrophic waters.

Season. In addition to the nine images used in the previous analyses, three more images from New Hampshire were used to expand the temporal resolution of the seasonal analysis (Table 1). All of the *in situ* measurements corresponding to these images were sorted by Julian Day. Each algorithm output was plotted against Julian Day while the lake chl-a level was held relatively constant by restricting the data used to only points within a specific chl-a concentration range. A "Relative Slope" metric was designed to judge how much of an influence water quality parameters besides chl-a that changed throughout the year had on each algorithm.

5. Model Validation

The quadratic best-fit models with the one-day time frame from the Time-Window and Minimum chl-a sensitivity analyses for the August 25th, 2014 satellite image of Maine were validated on the data sets from August 26, 2000, August 22, 2013, and August 9, 2014 to assess the possibility of using the models predictively. The time-window used for the validation set was chosen to ensure a high number and range of chl-a values. A perfectly predictive validation would produce a line of best fit with a slope of one, y-intercept of 0, and RMSE of zero. The image from the year 2000 was a Landsat 7 image to provide insight on how well models generated in the future from Landsat 8 data could be used to retroactively study archived Landsat images from before 2013. Chl-a levels predicted by each model based on each algorithm's output were plotted against the corresponding measured chl-a values. A linear regression and RMSE were used to assess the model's predictive performance.

RESULTS

Using data from both states and from all different times of year, relationships between each algorithm and the *in situ* chl-a measurements were very weak (Figure 2). There was some evidence of trends – like the highest *in situ* chl-a measurements all having relatively low KIVU outputs – that became more apparent when specific images were studied instead of using all the data points across different regions and seasons.

1. Time Window and Minimum chl-a Level: Maine

Limiting the time window between sample acquisition date and satellite image date produced clear and consistent results for August 26, 2000 and August 25, 2014, the two images with the most chl-a values above 10ppb. For the August 26, 2000 image, the correlation improved slightly only for KIVU and Kab1 from a five to a three day window, but with a two day window each of the five algorithms showed the

strongest relationship to the *in situ* measurements. Limiting the sample set to only points above 5ppb had a marginal and inconsistent effect on the correlations (Table 3). For all four data sets considered, KIVU and Kab1 displayed the strongest correlations, explaining over 95% of the variation in the most specific sample (within two days and over 5ppb) (Figure 3).

For August 25, 2014, reducing the time window removed noise and led to a stronger correlation between the *in situ* measurement and the output for all six algorithms (Figure 4). The algorithms were seemingly less sensitive to chl-a values below 5ppb, so the correlations were significantly stronger when only points above this floor were considered. For the most specific data set (within one day and above 5ppb), each algorithm explained over 75% of the variation in the eight data points (Table 4).

The strength of the correlations for the August 22, 2013 image was highly dependent on the algorithm. There were no significant correlations for SABI, NDVI, or 2BDA, but there was evidence that a larger sample size would have resulted in similar correlations to the previous two images because the algorithm outputs for the four samples with chl-a levels above 10ppb were grouped together at one end of the output range. The KIVU, Kab1, and Kab2 algorithms were moderately correlated with the measured chl-a levels. The correlations improved with a decreasing time window and – except for the one day sample that only had a seven data points – improved greatly by imposing a 5ppb floor. For the one day window with no chl-a floor, KIVU, Kab1, and Kab2 explained 48.77%, 37.91%, and 46.64% of the variation in the *in situ* data, respectively.

For June 19, 2013, there were no significant correlations between *in situ* chl-a and algorithm output for any of the six algorithms. There were not enough samples with chl-a above 5ppb to impose the minimum value limitation, and the outputs for each algorithm were extremely variable for the low chl-a value sample points. Likewise, there were no significant correlations for August 9, 2014, and all algorithms showed minimal sensitivity for chl-a values below 10ppb. In both cases, neither imposing a time window limitation nor a minimum chl-a limitation helped to eliminate the noise in the correlations between the small range of chl-a values and the associated algorithm outputs.

2. Time Window and Minimum chl-a Level: New Hampshire

The sensitivity analyses on New Hampshire data revealed significant relationships only for the August 3, 2015 image (Figure 5), drawn in part by the two samples with chl-a values above 10ppb. These relationships did improve from five days to two days for each algorithm, except SABI, which showed no significant correlations under any time window (Table 5). July 12, 2013 also had two points that were slightly above 10ppb that caused very slight trends for the NDVI and 2BDA algorithms, but all other points were scattered. For June 29, 2014 and August 19, 2015, none of the six algorithms could differentiate between the samples of chl-a values that were all below 10ppb.

3. Effect of Season

The amount of variation in algorithm output for a given chl-a range explained by the date varied from explaining almost none of the variation in the output data to explaining over 30%. The low relative slopes for the NDVI and 2BDA algorithms indicate that these algorithms changed the least over time for relatively constant chl-a values (Table 6). The SABI algorithm only displayed a slight effect of season for chl-a values in the 6-10ppb range. The Kab1 algorithm was consistently negatively correlated to the day of the year, and for the last two chlorophyll ranges, the time of year significantly explained 23.53% and 11.79%, respectively, of the variation in algorithm output. The Kab2 algorithm had a consistently positive relationship with the day of year that was again most significant for the two higher chlorophyll ranges, where the time of year significantly explained 30.76% and 25.91%, respectively, of the variation in algorithm output. The KIVU algorithm output showed the greatest effect of season, with a consistently

positive relationship. For the 4-6ppb chl-a range, time of year explained 39.53% of the variation in the positive trend of algorithm outputs (Figure 6).

4. Model Validation

The models used for validation were the quadratic models of best fit for the August 25th, 2015 Maine data points sampled within one day of the satellite image. The models created using all of the data points (M1), as well as the models created using only data points with chl-a values above 5ppb (M2) were validated.

The models were validated on the August 9th, 2014 Maine data set with a time-window of three days. Each M1 model except for SABI had positive slopes, but they were all below 0.55 and the intercepts were also all above five, suggesting that the M1 models overestimated low chl-a values and underestimated higher chl-a values. The models were most accurate for the KIVU algorithm. The same pattern was true for the M2 models, but the slopes were slightly lower and the intercepts slightly higher, indicating even more extreme over- and under-estimation. There were SABI, NDVI, and 2BDA values in the 8/22/13 set that fell well outside the algorithm output domain on which the 8/25/14 model was built, so these were excluded from the validation.

The models were validated on the 8/22/13 data with a time-window of two days. The M1 models had slopes close to 0.5 and intercepts ranging from 3.4 (SABI) to 9.2 (KIVU), again indicating the same pattern of prediction inaccuracy. The NDVI and 2BDA models had the highest RMSE. There were SABI, NDVI, and 2BDA values in the 8/22/13 set that fell well outside the algorithm output domain on which the 8/25/14 model was built, so these were excluded from the validation. The M2 models for KIVU, NDVI, Kab1, and Kab2 had slopes above seven and intercepts below one, suggesting that these models were overestimating chl-a values as the actual values increased. In general, the performance of the models on 8/25/14 and on 8/22/13 samples was poor and inconsistent. No model was able to predict known chl-a values with any significant accuracy.

For both M1 and M2 models validated on the Landsat 7 data from August 26, 2000 with a time window of two days, all algorithms had positive correlations between predicted and measured chl-a (Table 7). The models had slopes much closer to one and intercepts closer to zero compared to the other two validations, and RMSE values were consistently smaller. There was no clear distinction between any of the algorithms. The M1 models proved relatively accurate for the many smaller chl-a values as well as the three chl-a values that were above 15ppb. The M2 models all overestimated the smaller chl-a values and underestimated the higher ones. The models could not be validated for Kab2 since it uses a band that is not available from Landsat 7 data.

DISCUSSION

This study aimed to identify the applicability of Landsat 8 remote sensing technology for monitoring many small water bodies simultaneously primarily by comparing chl-a retrieval algorithm predictions to nearly coincident *in situ* observations in New Hampshire and Maine. The relationships between the two data sources were extremely robust in some cases, but minimal in others. The primary factors affecting the strength of the correlations were the range of *in situ* chl-a values sampled for a particular satellite image, the specific algorithm employed, the percentage of cloud cover for the satellite scene, and the time window between *in situ* samples and the satellite image.

CHL-A RANGE: The stark differences between the robust correlations in Maine and the unenlightening relationships in New Hampshire is attributable to the fact that the Maine lakes had many more chl-a concentrations outside of the 0-10ppb range. The plots of Maine algorithm outputs, particularly for 8/26/00 and 8/25/24 - which had 12 and 11 lakes with chl-a above 10ppb (Table 1) - show clearly that

most of the retrieval algorithms produce varied results for lakes with low chl-a values. Since almost all of the lakes sampled in NH fell into this category, the retrieval algorithms did not appear to be closely related to the *in situ* measurements. When there were enough sampled lakes with higher chl-a values, imposing a chl-a floor consistently improved correlations. An arbitrary floor of 5ppb was generally sufficient to achieve this result, but in some cases a slightly higher floor was more suitable (Figure 7). This is slightly higher than the 3ppb sensitivity limit described by Brivio et al. (2001) using the KIVU algorithm, but Kabbara et al. (2008) found that the Kab1 and Kab2 algorithms explained above 70% of the variation in 34 chl-a samples below 4ppb in coastal Tripoli. It is likely the heterogeneity of the lakes in the study region that introduced much of the variation in the oligotrophic, low chl-a lakes, suggesting that studying a wide variety of different water bodies comes with a tradeoff of sensitivity.

ALGORITHMS: The specific chl-a retrieval algorithm itself also proved to be a strong indicator of how robust the relationships were. The KIVU, Kab1, and Kab2 algorithms were consistently more related to the *in situ* samples than the 2BDA, SABI, and NDVI algorithms. These distinguishing characteristic of these first three algorithms compared to the other three is that they used L8 bands 1-4 and not band 5.The 2BDA, SABI, and NDVI algorithms have proved effective retrieval algorithms using satellites other than L8 because of the narrower near-infrared band (NIR). Landsat's NIR Band 5, however, is the closest comparable band, with wavelength of 850-880nm. This is well over the 706nm wavelength absorbance peak of chl-a that is more accurately represented in NDVI and NDCI (Normalized Difference Chlorophyll Index) remote sensing indices using other satellites. This is made evident by the fact that the NDVI retrieval algorithm presented here produces the reverse trend as would be expected for a measure of greenness, but despite the specific chl-a absorbance wavelength relevance it still proved a decent indicator of measure chl-a. Beck et al. (2016) noted too that Landsat 8's NIR band distance from the chl-a reflectance peak made algorithms using Band 5 less reliable. They found that the algorithms that compared the visible green peak to red or blue minima using Bands 1-4 performed the best. Derived combinations of these bands like those used by Kabbara et al. (2008) could be used at least for prediction within a specific region and season, but the success of Kab1 and Kab2 in this study suggest that these complex algorithms could still be useful outside of the original area of study too. Odermatt et al. (2012) found that green/blue ratios were most applicable to retrieve 0-10ppb chl-a concentrations, but red-NIR ratios became more useful as the measured concentrations increased, suggesting that a combination of algorithms might be necessary to study a wide range of lakes.

CLOUD COVER: Although remote sensing is often useful for its ability to overcome limitations of *in situ* sampling, there was evidence that environmental conditions and weather affected the retrieval algorithms just like poor weather would interfere with sampling efforts in the field. In addition to the fact that the 8/26/00 and 8/25/24 Maine images had a wider range of chl-a samples, these two scenes also had much lower cloud cover in the region. The same was true in New Hampshire, where the only image with significant correlations was the 8/3/15 image that had only 0.15% cloud cover, an order of magnitude less than all other New Hampshire images. Mishra et al. (2012) noted that MERIS satellite images that were not cloud-free required cloud masking before effectively interpreting chl-a retrieval algorithm data. Urbanski et al.'s (2016) study in Poland relied on uncloudy images too, and concluded that at least in that region, Landsat 8 and the European Space Agency's Sentinal 2 satellite would together provide at least one uncloudy image each month. Investigations of atmospheric correction techniques for L8 are aiming to reduce the weather limitations of remote sensing so that some cloud presence in an image can still produce robust chl-a prediction results (Concha and Schott 2015b).

TEMPORAL LIMITATIONS: Decreasing the time between the measured chl-a concentration and the satellite image consistently improved algorithm performance. The inconsistent distribution of samples around each scene made a rigorous analysis about the minimum acceptable time window difficult, but it was evident that more time introduces more noise. This is likely the result of either increased particulate matter in the water from rain or from actual changes in phytoplankton density and distribution that occur

between the *in situ* sample and the satellite image (Toming et al. 2016). Studying fewer sampling locations makes coincident surface measurement much more feasible (Beck et al. 2016), but studies involving dozens of different lakes used time windows of as little as one day (Kabbara et al. 2008, Keith et al. 2012) and as many as six days (Urbanski et al. 2016) with relative success.

On a wider temporal scale, the time of year also had a clear effect on the algorithm outputs, with the three algorithms that seemed best tuned to chl-a in the water column being the most affected. As turbidity and bottom vegetation likely increased throughout the year along with chl-a, these indices that in any given date did the best at characterizing the "greenness" were inconsistent throughout the year. This seasonal specificity poses a challenge for predictive remote sensing, which would ideally be tuned specifically to chl-a so that seasonal variation in other water quality parameters does not affect the output. Despite the limitations imposed by L8's relatively narrow bands, a preliminary investigation suggested that L8's thermal bands 10 and 11 are relatively well correlated with surface water temperature, which could potentially be incorporated into chl-a retrieval algorithms as a seasonal correction term.

POSSIBILITY OF PREDICTION: Although even correlations between chl-a retrieval algorithms and *in situ* samples across many small lakes is an advancement, remote sensing technology would ideally be used to develop models built on one or several satellite scenes to accurately predict chl-a from other scenes acquired in the future. The validation results from this study clearly indicate that this type of prediction is possible even when many different lakes are in question. The models were built using the August 25, 2014 scene with the most robust correlations and widest range of algorithm outputs to maximize the applicability to scenes from other times of the year and other years. It is not surprising based on the results of the correlational study that the validations on the other two L8 scenes were not very accurate.

The validation on the L7 8/22/00 image, however, proved most successful since this was the second of the two scenes with the most robust correlations to begin with. This result is particularly important because it demonstrates how models built using the newest Landsat satellite iterations could not only be useful for monitoring chl-a in the future, but could also be used retroactively to investigate chl-a concentrations from archived satellite images. Even using a validation set with a time window of five days – which is known from the correlational analysis to introduce some noise – the model is very clearly able to distinguish between "high" and "low" chl-a concentrations. When the time window of samples used was narrowed to within three days of the image date, some of the outlying points are removed as expected (Figure 8). Even though this model does not predict within a narrow margin of error the actual chl-a concentration, its clear ability to differentiate between classes of lakes is still useful, especially on a regional scale. Recent studies for large-scale remote sensing have been satisfied with simply classifying the tropic status of lakes into broad categories (Bresciani et al. 2011, Watanabe et al. 2015, Andrzej Urbanski et al. 2016). In a region like the Northeastern United States, remotely sensed classification of the hundreds of lakes in the area would be a powerful tool for water quality management.

CONCLUSIONS AND SUGGESTIONS: This study compared *in situ* chl-a concentration measurements in to chl-a retrieval algorithm predictions from satellite observations over Maine and New Hampshire. The KIVU, Kab1, and Kab2 algorithms in particular – which used Landsat 8 bands 1-4 rather than band 5 – explained large amounts of the variation in measured vs. predicted chl-a. The correlations were highest on cloud-free days, and improved dramatically with increased temporal coincidence of the sampling effort. Additionally, the algorithms explained over 95% of variation in the data after imposing a chl-a minimum of 5ppb, suggesting that the chl-a models have a difficulty capturing low chl-a levels, but are much more successful with higher concentrations. Significant seasonal variation in the algorithm outputs can be explained by the wide Landsat 8 bands that do not record only the narrow band of chl-a reflectance, and can potentially be mitigated with remotely sensed water temperature in lieu of narrower satellite bands. The validation results showed that models built using a wide enough range of sampled concentrations can

in some cases reliably classify the chl-a concentration in any given 30m pixel in past or future satellite images that is cloud-free.

This investigation points to improvements that both remote sensing specialists and ecologists could make to advance large scale remote sensing of freshwater resources. First, the coordination of sampling efforts with satellite overpasses would provide much more data for model development. Samples nearly coincident with the satellite image are clearly more useful for investigating correlations and building predictive models. Organizing VLMP or VLAP volunteers to conduct their water quality analysis on days when a satellite like L8 will be acquiring an image of the region will provide the data required for refining retrieval algorithms. Second, the application of remote sensing to solving ecological problems will require intentional band design to maximize the ecological relevance of reflectance outputs. This study confirms Beck et al.'s (2016) conclusion that high spatial resolution satellites like Sentinal-2A and Landsat-8 are most relevant for monitoring small-inland water bodies even though they sacrifice the spectral resolution that other satellites offer. As engineers continue to weigh the tradeoffs between temporal resolution, spatial resolution, and spectral resolution, it will be important to take into account the needs of end users like ecologists and policy makers. Ultimately, this investigation demonstrates how further communication and collaboration between scientists, engineers, and policy makers can help to solve large-scale problems like protecting valuable freshwater resources.

ACKNOWLEDGEMENTS

I would like to thank Dr. Kathleen Weathers and Dr. Hamid Norouzi for their guidance and Dr. Satya Prakash, Saba Saberi, and Jaelyn Bos for technical assistance.

LITERATURE CITED

Alawadi, F. 2010. Detection of surface algal blooms using the newly developed algorithm surface algal bloom index (SABI). Proceedings of the International Society for Optics and Photonics 7825:1–14.

Andrzej Urbanski, J., A. Wochna, I. Bubak, W. Grzybowski, K. Lukawska-Matuszewska, M. Łącka, S. Śliwińska, B. Wojtasiewicz, and M. Zajączkowski. 2016. Application of Landsat 8 imagery to regional-scale assessment of lake water quality. International Journal of Applied Earth Observation and Geoinformation 51:28–36.

Beck, R., S. Zhan, H. Liu, S. Tong, B. Yang, M. Xu, Z. Ye, Y. Huang, S. Shu, Q. Wu, S. Wang, K. Berling, A. Murray, E. Emery, M. Reif, J. Harwood, J. Young, C. Nietch, D. Macke, M. Martin, G. Stillings, R. Stump, and H. Su. 2016. Comparison of satellite reflectance algorithms for estimating chlorophyll-a in a temperate reservoir using coincident hyperspectral aircraft imagery and dense coincident surface observations. Remote Sensing of Environment 178:15–30.

Bradt, S. 2010. Using remote sensing to monitor water quality in New England lakes. Gauging the Health of New England's Lakes and Ponds: A Survey Report and Decision-Making Resource October:35–38.

Bresciani, M., D. Stroppiana, D. Odermatt, G. Morabito, and C. Giardino. 2011. Assessing remotely sensed chlorophyll-a for the implementation of the Water Framework Directive in European perialpine lakes. Science of the Total Environment 409:3083–3091.

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.

Carey, C. C., K. C. Weathers, and K. L. Cottingham. 2009. Increases in phosphorus at the sediment-water interface may influence the initiation of cyanobacterial blooms in an oligotrophic lake. International Association of Theoretical and Applied Limnology, Vol 30, Pt 8, Proceedings: Now: International Society of Limnology 30:1185–1188.

Concha, J. A., and J. R. Schott. 2015a. Retrieval of color producing agents in Case 2 waters using Landsat

8. Remote Sensing of Environment.
Concha, J. A., and J. R. Schott. 2015b. Atmospheric correction for Landsat 8 over case 2 waters. Earth Observing Systems XX 9607:96070R.
Dall'Olmo, G., and A. a Gitelson. 2006. Effect of bio-optical parameter variability and uncertainties in reflectance measurements on the remote estimation of chlorophyll-a concentration in turbid productive waters: modeling results. Applied optics 45:3577–3592.
Davis, R. B., D. S. Anderson, S. S. Dixit, P. G. Appleby, and M. Schauffler. 2006. Responses of two New Hampshire (USA) lakes to human impacts in recent centuries. Journal of Paleolimnology 35:669–697.
Dörnhöfer, K., and N. Oppelt. 2016. Remote sensing for lake research and monitoring - Recent advances. Ecological Indicators 64:105–122.
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.
Feyisa, G. L., H. Meilby, R. Fensholt, and S. R. Proud. 2014. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. Remote Sensing of Environment 140:23–35.
Gerace, A. D., J. R. Schott, and R. Nevins. 2013. Increased potential to monitor water quality in the near-shore environment with Landsat's next-generation satellite. Journal of Applied Remote Sensing 7:73558.
Kabbara, N., J. Benkhelil, M. Awad, and V. Barale. 2008. Monitoring water quality in the coastal area of Tripoli (Lebanon) using high-resolution satellite data. ISPRS Journal of Photogrammetry and Remote Sensing 63:488–495.
Keith, D. J., B. Milstead, H. Walker, H. Snoo, J. Szykman, M. Wusk, L. Kagey, C. Howell, C. Mellanson, and C. Drueke. 2012. Trophic status, ecological condition, and cyanobacteria risk of New England lakes and ponds based on aircraft remote sensing. Journal of Applied Remote Sensing 6:063577.
Mishra, S., and D. R. Mishra. 2012. Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. Remote Sensing of Environment 117:394–406.
Moore, T. S., M. D. Dowell, S. Bradt, and A. Ruiz Verdu. 2014. An optical water type framework for selecting and blending retrievals from bio-optical algorithms in lakes and coastal waters. Remote Sensing of Environment 143:97–111.
Palmer, S. C. J., T. Kutser, and P. D. Hunter. 2015. Remote sensing of inland waters: Challenges, progress and future directions. Remote Sensing of Environment 157:1–8.
Roy, D. P., M. A. Wulder, T. R. Loveland, W. C.E., R. G. Allen, M. C. Anderson, D. Helder, J. R. Irons, D. M. Johnson, R. Kennedy, T. A. Scambos, C. B. Schaaf, J. R. Schott, Y. Sheng, E. F. Vermote, A. S. Belward, R. Bindschadler, W. B. Cohen, F. Gao, J. D. Hipple, P. Hostert, J. Huntington, C. O. Justice, A. Kilic, V. Kovalskyy, Z. P. Lee, L. Lymburner, J. G. Masek, J. McCorkel, Y. Shuai, R. Trezza, J. Vogelmann, R. H. Wynne, and Z. Zhu. 2014. Landsat-8: Science and product vision for terrestrial global change research. Remote Sensing of Environment 145:154–172.
Schloss, A. L., S. Spencer, J. A. Schloss, J. Haney, S. Bradt, and J. Nowak. 2002. Using Landsat TM data to aid the assessment of long-term trends in lake water quality in New Hampshire lakes. Geoscience and Remote Sensing Symposium 00:3095–3098.
Shen, L., H. Xu, and X. Guo. 2012. Satellite remote sensing of harmful algal blooms (HABs) and a potential synthesized framework. Sensors (Switzerland) 12:7778–7803.
Tebbs, E. J., J. J. Remedios, and D. M. Harper. 2013. Remote sensing of chlorophyll-a as a measure of cyanobacterial biomass in Lake Bogoria, a hypertrophic, saline-alkaline, flamingo lake, using Landsat ETM+. Remote Sensing of Environment 135:92–106.
Toming, K., T. Kutser, A. Laas, M. Sepp, B. Paavel, and T. Nõges. 2016. First Experiences in Mapping Lake Water Quality Parameters with Sentinel-2 MSI Imagery. Remote Sensing 8:640.
Vertucci, F. a., and G. E. Likens. 1989. Spectral reflectance and water quality of Adirondack mountain

region lakes. Limnology and Oceanography 34:1656–1672.
Watanabe, F. S. Y., E. Alcântara, T. W. P. Rodrigues, N. N. Imai, C. C. F. Barbosa, and L. H. da S. Rotta. 2015. Estimation of chlorophyll-a concentration and the trophic state of the barra bonita hydroelectric reservoir using OLI/landsat-8 images. International Journal of Environmental Research and Public Health 12:10391–10417.

APPENDIX

TABLE 1. Landsat 8 images (*besides August 26, 2000, from Landsat 7) acquired over Maine (P12R29) and over New Hampshire (P13R30) and the available corresponding *in situ* chl-a data. The NH images from September and October were used for seasonal analysis, but not assessing algorithm performance.

	Date	*Cloud Cover*	*Lakes Sampled*	*Chl-a Min*	*Chl-a Max*	*Chl-a Median*	*Lakes with Chl-a above 10ppb*
NH	July 12, 2013	11.44%	19	1.410	11.050	4.750	2
	September 30, 2013	2.56%	2	5.65	5.73	5.69	0
	June 29, 2014	12.15%	13	1.29	6.74	3.76	0
	September 17, 2014	11.91%	5	1.790	6.550	2.590	0
	October 3, 2014	0.15%	6	1.320	2.540	9.510	0
	August 3, 2015	3.14%	17	1.550	12.100	3.280	2
	August 19, 2015	7.36%	28	1.070	9.770	2.530	0
ME	August 26, 2000	0.31%	70	1.200	48.400	4.400	12
	June 19, 2013	4.69%	19	1.300	48.000	3.000	1
	August 22, 2013	12.78%	65	1.300	56.000	3.300	5
	August 9, 2014	16.21%	54	1.200	17.000	4.050	3
	August 25, 2014	0.47%	57	1.200	65.000	4.400	11

TABLE 2. Descriptions of each of the six algorithms assessed. The original algorithm band math was converted to the closest comparable Landsat 8 band math. "Original Use" describes the type of water and the satellite for which the algorithm was designed.

Algorithm	Landsat 8 Band Math	Original Use	Source
Surface Algal Bloom Index (SABI)	(B5-B4)/(B2+B3)	Ocean, designed to minimize variations in cloud shadow and atmospheric conditions, using MODIS satellite.	Alawadi 2010
3BDA-like (KIVU)	(B2-B4)/B3	Large freshwater lake, above 3ppb, Landsat TM.	Brivio et al. 2001
Normalized Difference Vegetation Index (NDVI)	(B5-B4)/(B5+B4)	Estuarine and coastal waters 1-60ppb, using MERIS satellite.	Mishra and Mishra 2012
2BDA	B5/B4	Simulated turbid productive freshwater, using Landsat TM.	Dall'Olmo and Gitelson 2006
Kab1	e^(1.67-3.94*ln(B2)+3.78*ln(B3))	Coastal, best-fit algorithm, chl-a below 4ppb, using Landsat 7.	Kabbara et al. 2008
Kab2	e^(6.92274-5.7581*(ln(B1)/ln(B3)))	Coastal, best-fit algorithm, chl-a below 4ppb, using Landsat 7.	Kabbara et al. 2008

TABLE 3. Descriptive statistics for correlations between *in situ* chl-a values for Maine lakes and algorithm outputs from August 26, 2000 image. Data sets used were limited by days between the image and the sample as well as by imposing a chl-a minimum of 5ppb. The Kab2 algorithm was not used because it requires a bandwidth unique to Landsat 8.

	5 days				*3 days*				*2 days*			
	All points (n = 68)		> 5 ppb (n = 27)		All points (n = 44)		> 5 ppb (n = 19)		All points (n = 25)		> 5 ppb (n = 13)	
	R^2	p	R^2	p	R^2	p	R^2	p	R^2	p	R^2	p
SABI	0.4165	<0.001	0.4659	<0.001	0.3163	<0.001	0.2929	0.0244	0.807	<0.001	0.8157	<0.001
KIVU	0.8732	<0.001	0.8542	<0.001	0.8942	<0.001	0.8846	<0.001	0.9245	<0.001	0.954	<0.001
NDVI	0.6326	<0.001	0.6717	<0.001	0.578	<0.001	0.5718	<0.001	0.7975	<0.001	0.7894	<0.001
2BDA	0.3395	<0.001	0.3897	<0.001	0.2589	<0.001	0.2342	0.0461	0.7533	<0.001	0.7581	<0.001
Kab1	0.8402	<0.001	0.8204	<0.001	0.9029	<0.001	0.9102	<0.001	0.9343	<0.001	0.962	<0.001
Kab2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

TABLE 4. Descriptive statistics for correlations between *in situ* chl-a values for Maine lakes and algorithm outputs from August 25, 2014 image. Data sets used were limited by days between the image and the sample as well as by imposing a chl-a minimum of 5ppb.

	5 days				*3 days*				*1 day*			
	All points (n = 55)		> 5 ppb (n = 23)		All points (n = 38)		> 5 ppb (n = 16)		All points (n = 16)		> 5 ppb (n = 8)	
	R^2	p	R^2	p	R^2	p	R^2	p	R^2	p	R^2	p
SABI	0.1976	<0.001	0.4392	0.002	0.2754	<0.001	0.547	0.003	0.5785	<0.001	0.9752	<0.001
KIVU	0.5686	<0.001	0.4402	0.003	0.6664	<0.001	0.6446	<0.001	0.8004	<0.001	0.7469	0.032
NDVI	0.4042	<0.001	0.6042	<0.001	0.3702	<0.001	0.8504	<0.001	0.8504	<0.001	0.9859	<0.001
2BDA	0.2066	<0.001	0.5514	<0.001	0.2118	0.004	0.7134	<0.001	0.5071	0.002	0.9806	<0.001
Kab1	0.5874	<0.001	0.4891	0.001	0.6565	<0.001	0.6644	<0.001	0.7795	<0.001	0.7828	0.022
Kab2	0.5496	<0.001	0.4375	0.003	0.6465	<0.001	0.6589	<0.001	0.7864	<0.001	0.7688	0.026

TABLE 5. Descriptive statistics for correlations between *in situ* chl-a values for New Hampshire lakes and algorithm outputs from four Landsat 8 images. Data sets used were limited by days between the image and the sample, with the smaller time window chosen to ensure an adequate sample size for each date. Meaningful relationships (bolded) were only present for the August 3, 2015 image.

New Hampshire Lakes		**July 12, 2013**		**June 29, 2014**		**August 3, 2015**		**August 19, 2015**	
		5 days (n = 19)	3 days (n = 14)	5 days (n = 13)	3 days (n = 9)	5 days (n = 17)	2 days (n = 13)	5 days (n = 28)	1 days (n = 16)
SABI	R^2	-0.1117	-0.1234	-0.1918	-0.2017	-0.1364	-0.1583	-0.04217	-0.1074
	p	0.9094	0.7566	0.9665	0.732	0.9609	0.8379	0.6404	0.7657
KIVU	R^2	-0.00695	0.05568	0.01494	0.2425	**0.3543**	**0.5781**	-0.07218	0.02713
	p	0.4119	0.2912	0.3727	0.1833	0.01839	0.005374	0.9132	0.3299
NDVI	R^2	0.1694	0.1861	0.1422	0.1041	**0.2299**	**0.4878**	-0.01767	0.1448
	p	0.08832	0.1286	0.1866	0.3034	0.06309	0.01417	0.4757	0.1427
2BDA	R^2	0.1806	0.1906	0.1879	0.07866	**0.2498**	**0.5068**	-0.03746	0.1704
	p	0.07924	0.1247	0.1419	0.3299	0.05254	0.01173	0.6051	0.1171
Kab1	R^2	-0.01449	-0.0226	0.00177	0.1476	**0.3548**	**0.5848**	-0.06049	-0.1411
	p	0.4373	0.4512	0.3983	0.2612	0.01828	0.004959	0.7962	0.9305
Kab2	R^2	-0.02982	-0.0220	-0.1327	-0.0721	**0.3627**	**0.6062**	-0.0556	-0.0995
	p	0.493	0.4497	0.7494	0.5199	0.01676	0.003806	0.7515	0.7309

TABLE 6. Comparison of relationships between algorithm output and Julian Day of year. The relative slope is a metric to compare the slopes for different algorithms calculated by normalizing each slope by the range of the algorithm and then dividing each by the smallest corrected slope. Negative values represent decreasing algorithm outputs as the year went on. Each * represents an outlier that was removed from the data set to ensure representative regression lines.

CHl-A Ranges:		**0 - 2.5 ppb (n = 81)**	**2.5 - 4 ppb (n = 104)**	**4 - 6 ppb (n = 79)**	**6 - 10 ppb (n = 52)**
SABI	Relative Slope	4.88*	1.49	-1.97***	19.83*
KIVU	Relative Slope	15.57	22.17*	42.34*	38.92
NDVI	Relative Slope	-5.63	-2.36	-11.74	-5.56
2BDA	Relative Slope	1.10	1.00	-14.29	-4.41
KAB1	Relative Slope	-7.31	-16.58*	-27.81**	-31.24*
Kab2		*n = 75*	*n = 85*	*n = 56*	*n = 43*
	Relative Slope	14.12	14.84*	29.05	38.07*

TABLE 7. Linear validation (ax+b) of 8/25/14 models built with all data points on 8/26/00 data sampled within two days of image acquisition (n = 25). M1 is the model built using all 8/26/00 data points, while M2 is that built using just those with chl-a > 5ppb.

		SABI	***KIVU***	***NDVI***	***2BDA***	***Kab1***	***Kab2***
M1	**a**	0.7879	0.86705	1.0237	1.1806	1.34812	NA
	b	1.0804	0.4001	0.8425	0.8151	0.3126	NA
	RMSE	3.0372	2.5308	3.5808	4.9112	2.2568	NA
M2	**a**	0.6677	0.53172	0.52237	0.7151	0.53831	NA
	b	5.3607	6.36092	5.60774	5.7730	6.71173	NA
	RMSE	2.0121	0.9446	1.6677	2.9022	0.9031	NA

TABLE 8 (not referenced in text). Band descriptions and wavelengths for USGS Landsat 7 (L7) and Landsat 8 (L8). * Denotes that those bands are acquired at 100m resolution, but are resampled to 30m in the data product.

L8 Band	**Wavelength (um)**	**Resolution**	**Comparable L7 Band**
Band 1 – Coastal aerosol	0.43 – 0.45	30	None
Band 2 – Blue	0.45 – 0.51	30	Band 1
Band 3 – Green	0.53 – 0.59	30	Band 2
Band 4 – Red	0.64 – 0.67	30	Band 3
Band 5 – Near IR	0.85 – 0.88	30	Band 4
Band 6 – Short-wave IR 1	1.57 – 1.65	30	Band 5
Band 7 – Short-wave IR 2	2.11 – 2.29	30	Band 7
Band 8 - Panchromatic	0.50 – 0.68	15	Bands 2, 3, 4
Band 9 – Cirrus	1.36 – 1.38	30	None
Band 10 – Thermal IR 1	10.60 – 11.19	30*	Band 6
Band 11 – Thermal IR 2	11.50 – 12.51	30*	Band 6

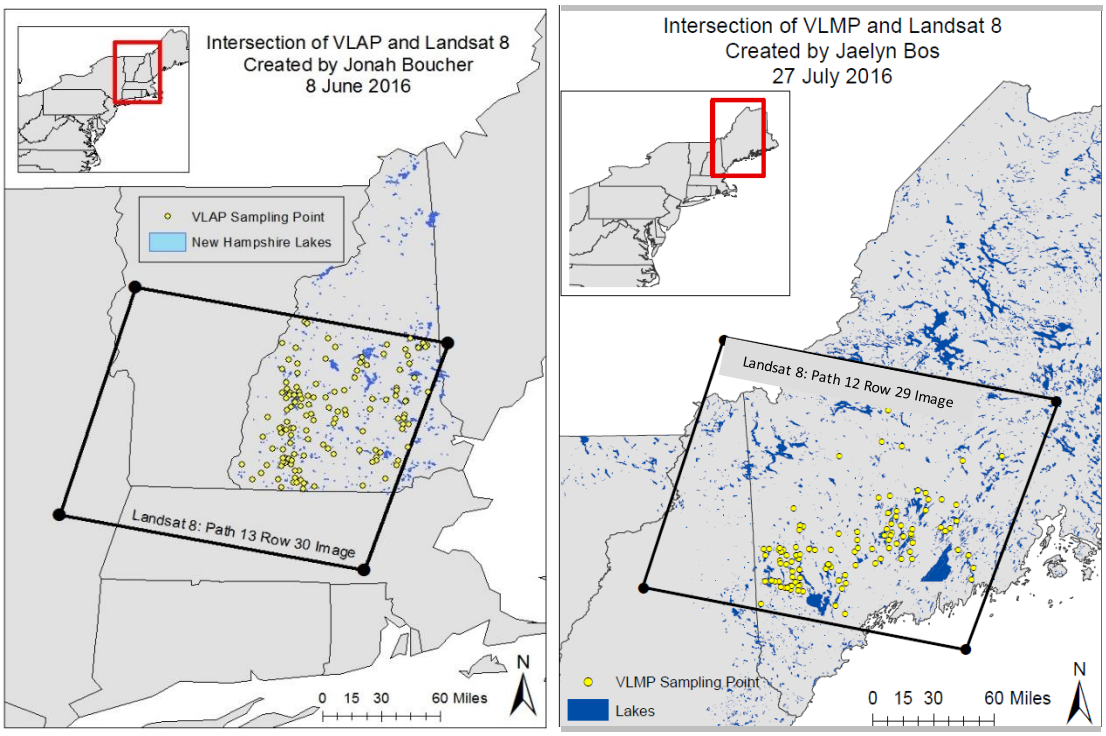

FIGURE 1. Lakes sampled by New Hampshire's Volunteer Lake Assessment Program (left) and Maine's Volunteer Lake Monitoring Program (right) that fall within Landsat satellite Path 13 Row 30 and Path 12 Row 29, respectively.

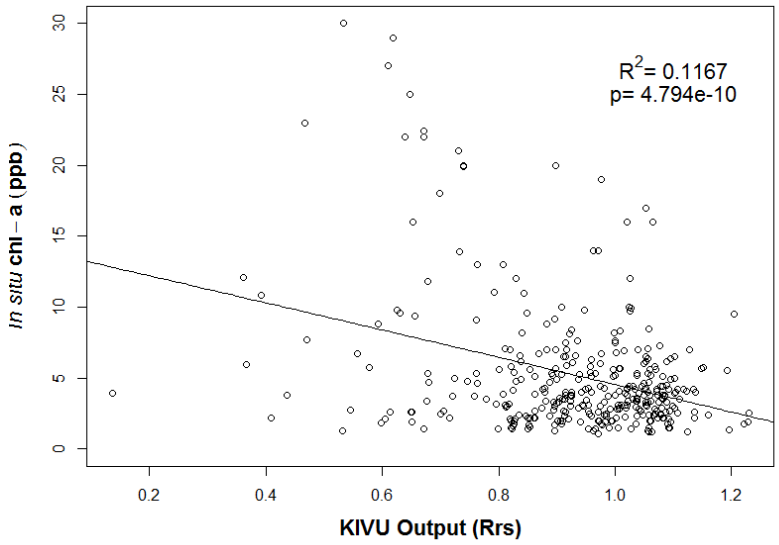

FIGURE 2. Correlation between the KIVU chl-a retrieval algorithm output and the *in situ* measured chl-a for 349 sampling points in Maine and New Hampshire lakes. Each of the six algorithms examined had similarly weak correlations.

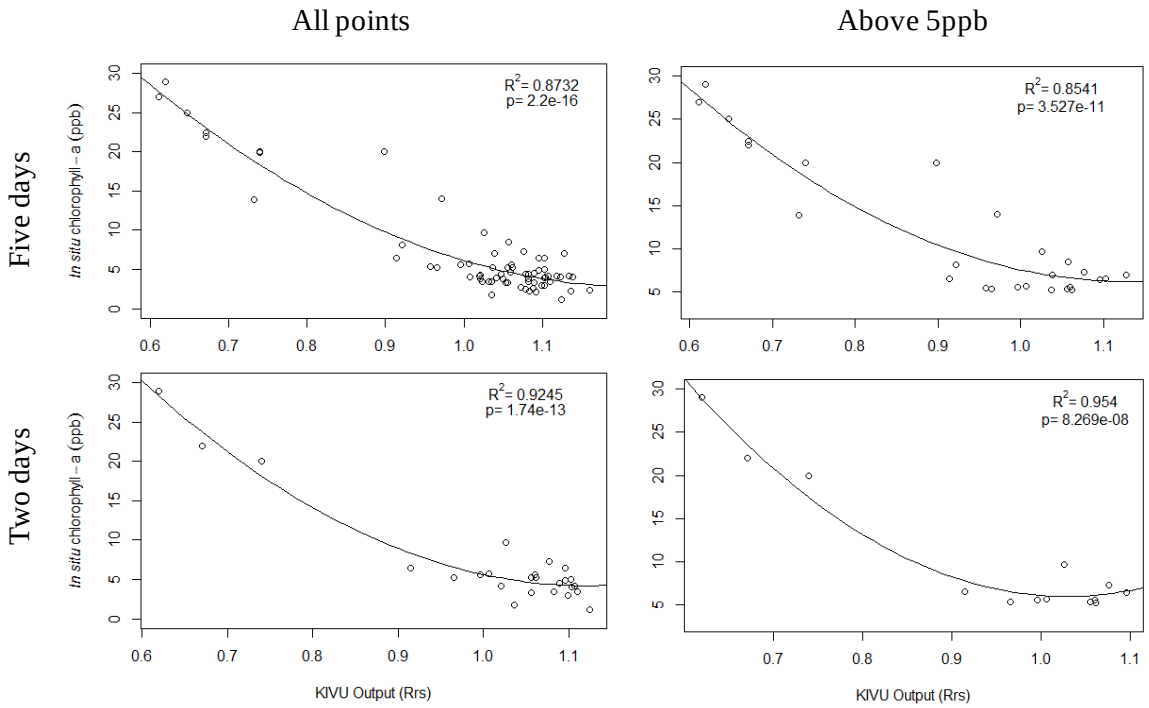

FIGURE 3. *In-situ* chl-a values for Maine lakes plotted against KIVU output from August 26, 2000 satellite image. Samples used are limited by time window (within five days or two days from the image date) and by imposing a minimum chl-a level (5ppb). These graphs are representative of the correlation trends of the other four algorithms described in Table 3.

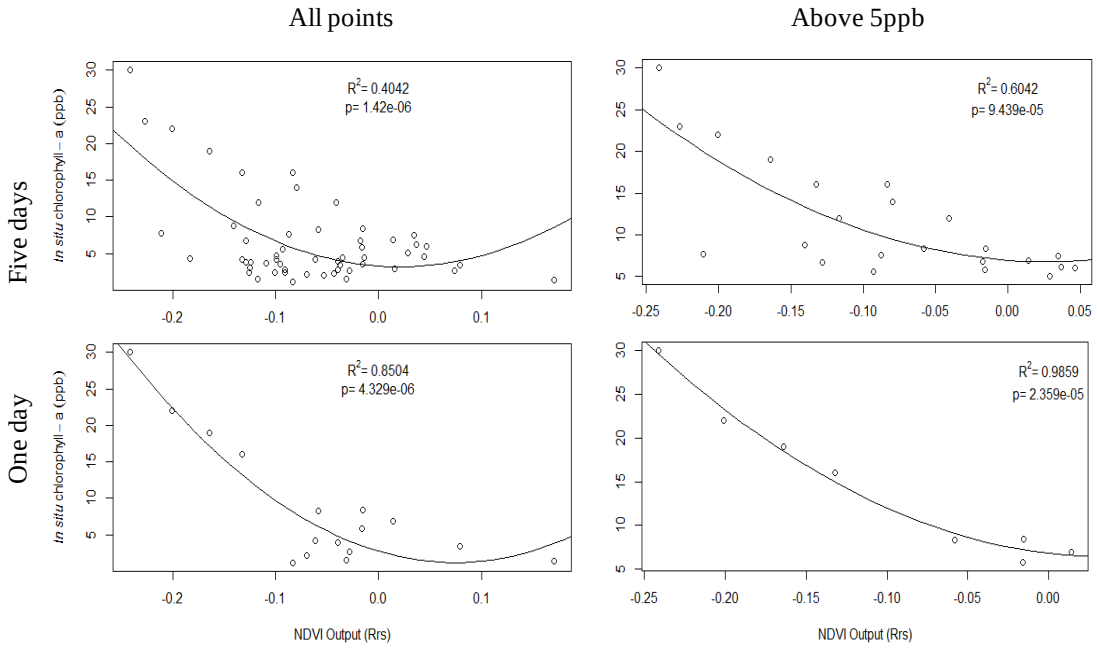

FIGURE 4. *In-situ* chl-a values for Maine lakes plotted against NDVI output from August 25, 2014 satellite image. Samples used are limited by time window (within five days or one days from the image date) and by imposing a minimum chl-a level (5ppb). These graphs are representative of the correlation trends of the other five algorithms described in Table 4.

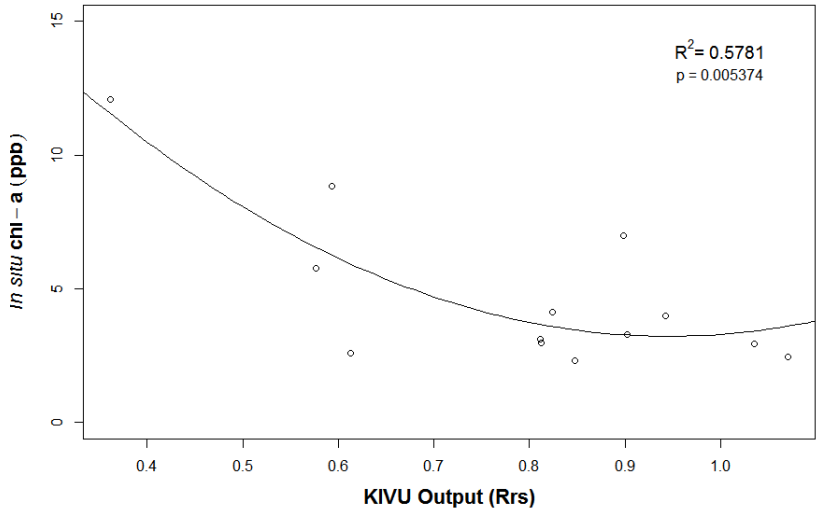

FIGURE 5. Correlation between New Hampshire lakes *in situ* chl-a measurements and KIVU algorithm outputs for August 3, 2015 Landsat 8 image with a two day time window. This graph is representative of the relationship using NDVI, 2BDA, Kab1, and Kab1.

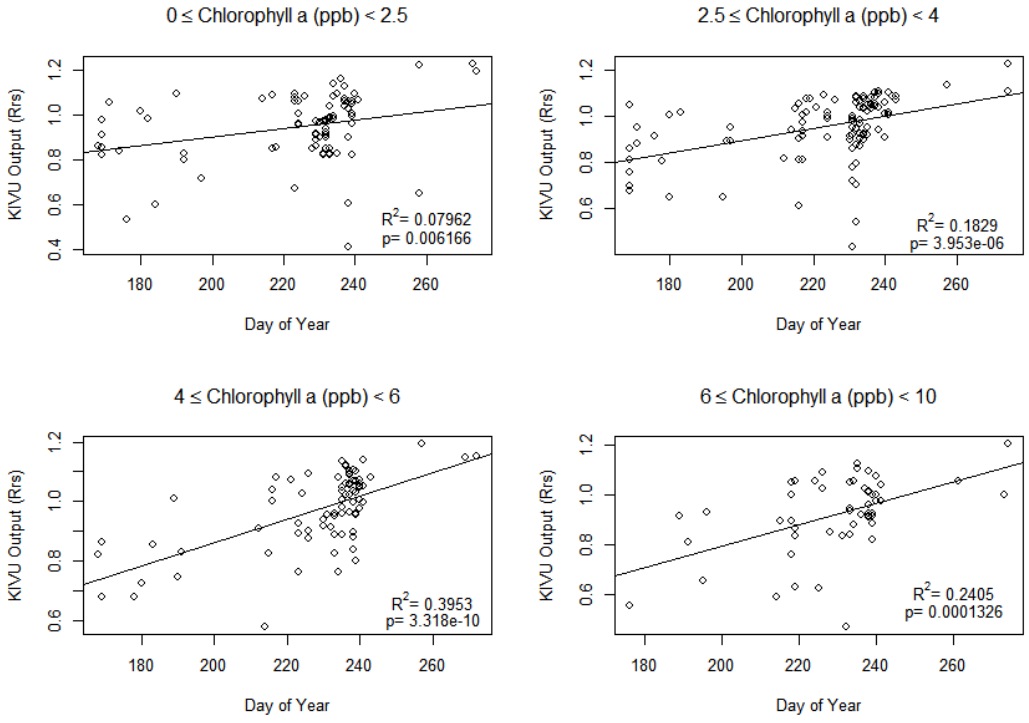

FIGURE 6. The effect of Julian Day of Year on KIVU output for all NH and ME samples within specific chl-a ranges.

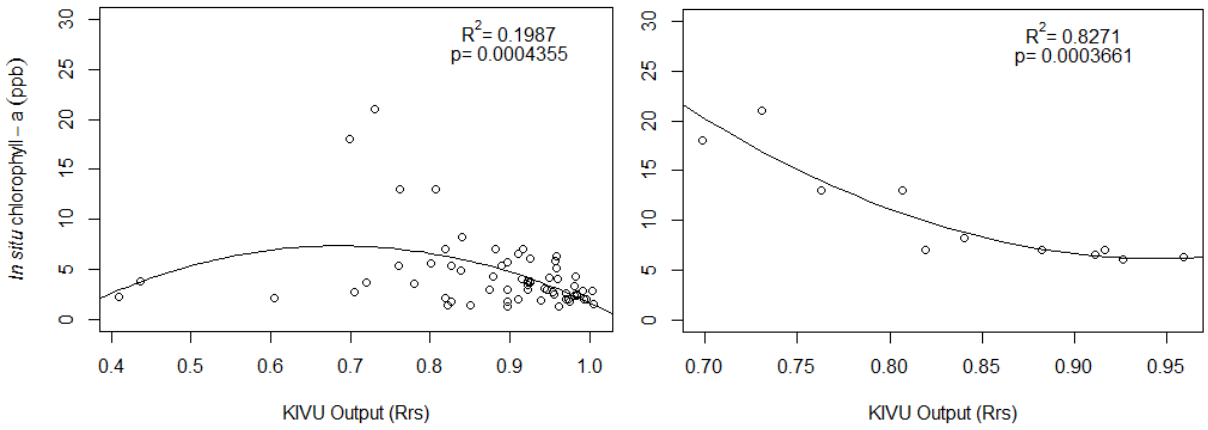

FIGURE 7. *In-situ* chl-a values for Maine lakes sampled within five days of satellite image plotted against KIVU output from August 22, 2013. A 6ppb floor was used on the right to demonstrate how the algorithm's sensitivity to low chl-a can affect apparent relationships.

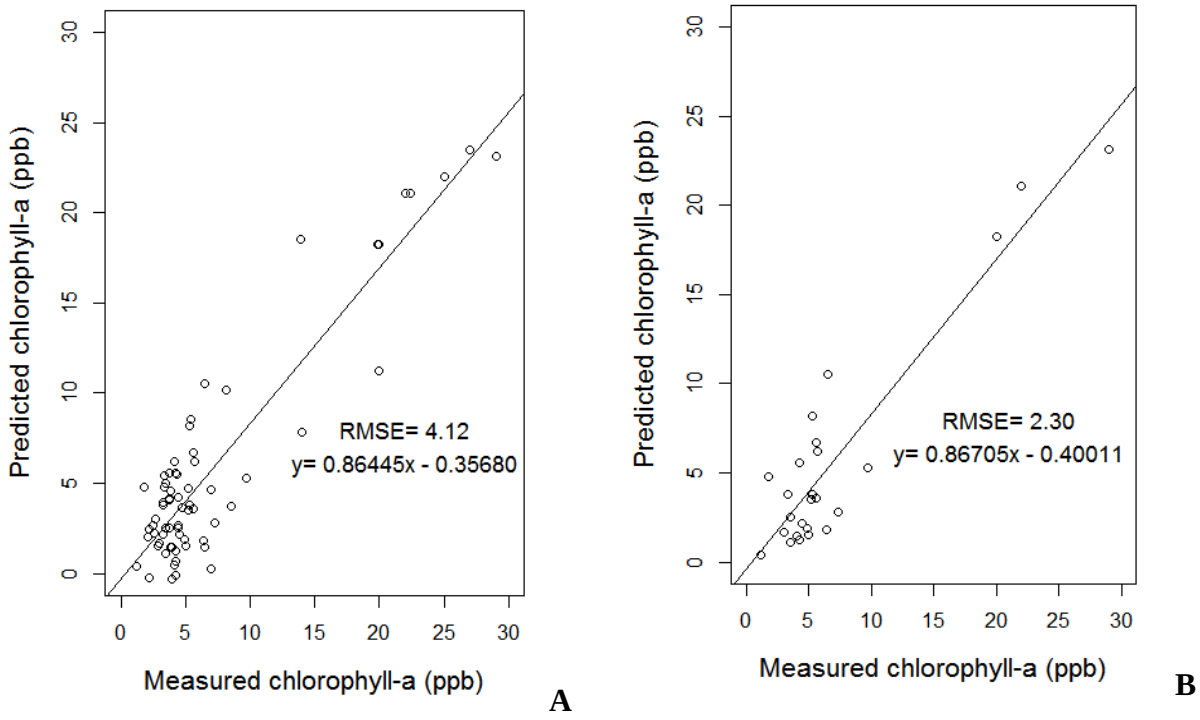

FIGURE 8. Validation of 8/25/2014 Maine model built using all samples on 8/22/00 image and coincident samples with a time window of five days (A) and a time window of three days (B).