Remote sensing of cyanobacterial blooms in Lake Champlain, USA
Adam Trescott, Elizabeth Isenstein and Mi-Hyun Park

ABSTRACT

The objective of this study was to develop cyanobacteria remote sensing models using Landsat 7 Enhanced Thematic Mapper Plus (ETM+) as an alternative to shipboard monitoring efforts in Lake Champlain. The approach allowed for estimation of cyanobacteria directly from satellite images, calibrated and validated with 4 years of in situ monitoring data from Lake Champlain’s Long-Term Water Quality and Biological Monitoring Program (LTMP). The resulting stepwise regression model was applied to entire satellite images to provide distribution of cyanobacteria throughout the surface waters of Lake Champlain. The results demonstrate the utility of remote sensing for estimating the distribution of cyanobacteria in inland waters, which can be further used for presenting the initiation and propagation of cyanobacterial blooms in Lake Champlain.

INTRODUCTION

Cyanobacteria, also called blue-green algae, have been a serious concern in many inland freshwaters because the blooms impact drinking water supplies, recreational use, and fisheries; and pose a threat to human health and the aquatic ecosystem (Paerl et al. 2011; O’Neil et al. 2012). Some cyanobacteria species produce cyanotoxins, which can be hazardous to animals and humans. Cyanobacterial blooms pose a variety of water treatment problems (NRA 1990; Ferguson et al. 1996; Makarewicz, et al. 2006). A famous case of cyanotoxin poisoning occurred in Brazil, where 60 people were poisoned after drinking water contaminated by microcystin, a toxin produced by the cyanobacterial species, Microcystis (Pouria et al. 1998).

Many factors are known to influence cyanobacterial bloom propagation, including light, temperature and available phosphorus (De Nobel et al. 1998, Yamamoto & Nakahara 2005), but the real-time drivers that spur cyanobacterial blooms are poorly understood (Watzin et al. 2006; Ndong et al. 2010). The occurrence, intensity and duration of blooms are often difficult to predict, requiring frequent monitoring to inform public health alerts for potential risks associated with recreational uses and drinking water treatment.

Satellite remote sensing has been used for detecting water quality parameters since the early 1970s (Ekstrand 1992) and can provide regular measurement of cyanobacterial blooms over large coverage areas. Satellite imagery has been used to delineate pigment concentrations in identifying algal species (Stumpf et al. 2009). Chlorophyll a in algae-laden water shows unique reflectance/absorption characteristics, which can be used to develop bio-optical algorithms (O’Reilly et al. 1998; Brown et al. 2008). Phycocyanin, an accessory pigment to chlorophyll, has been used to detect cyanobacterial species (Vincent et al. 2004; Simis et al. 2005; Makarewicz et al. 2006; Randolph et al. 2008). Phycocyanin exhibits unique spectral characteristics of absorption spectra with peaks near 620 nm (Vincent et al. 2004; Gons et al. 2005; Metsamaa et al. 2006, Becker et al. 2006, 2009). Vincent et al. (2004) successfully used Landsat data to develop phycocyanin detection algorithms for assessing cyanobacteria blooms in Lake Erie although Landsat data does not provide the peak spectra of phycocyanin. Becker et al. (2009) derived cyanobacteria...
abundances from Moderate Resolution Imaging Spectroradiometer (MODIS) data in the lower Great Lakes using published absorption spectra for chlorophyll $a$, phycocyanin, colored dissolved organic matter and suspended sediments. Wheeler et al. (2002) retrieved Chlorophyll $a$ and phycocyanin using Envisat’s Medium Resolution Imaging Spectrometer (MERIS) and QuickBird data. These studies demonstrated that satellite remote sensing could be applied to a rapid detection and warning system for cyanobacteria blooms. Evolving technologies, new satellite programs and unprecedented access to satellite data archives have heightened interest in the development of sophisticated remote sensing models for regional and global monitoring programs (Morel & Prieur 1977; Mortimer 1988; Ekstrand 1992; O’Reilly et al. 1998).

The objective of this study is to develop a remote sensing model to detect algal and cyanobacterial bloom distributions in Lake Champlain using remote sensing: Enhanced Thematic Mapper Plus (ETM+) data on Landsat 7. The ETM+ data are well suited for monitoring algal blooms over Lake Champlain because of its relatively high spatial resolution (30 m), allowing for measurements of small embayments which tend to have the most problems associated with eutrophication and algal blooms. In theory, phycocyanin pigments are proportional to cyanobacteria concentrations, and the ETM+ sensor should be able to detect the unique spectral properties of cyanobacterial blooms. Therefore, we attempted to develop a remote sensing model based directly on the net cyanobacteria measurements. This study will demonstrate the utility of remote sensing models in lake monitoring programs. The approach will provide effective monitoring of cyanobacteria in lakes and presents new insights into the initiation and propagation of cyanobacterial blooms.

**MATERIALS AND METHODS**

**Study area**

Lake Champlain is one of the largest lakes in the USA and an important resource for the States of Vermont and New York, USA, and the Province of Quebec, Canada, providing drinking water to approximately 200,000 people living in the surrounding basin (Watzin et al. 2006; USEPA 2011). Situated between the Green Mountains of Vermont and the Adirondack Mountains of New York State, the lake covers a surface area of approximately 1,127 km$^2$, reaching a maximum depth of 122 m. On average, the lake holds 25.8 km$^3$ of water as it flows northward from Whitehall, New York to its northern outlet in the Richelieu River, Quebec. The Lake Champlain watershed drains a land area of approximately 21,326 km$^2$ (Figure 1) (Watzin et al. 2006).

Lake Champlain has been impaired by excess phosphorus loading mainly attributed to non-point source pollution, which has significantly increased in the last two decades (Smeltzer et al. 2009). As a result, the lake experiences increased eutrophic conditions and cyanobacterial blooms. This problem is most prevalent in Missisquoi Bay, where the mean phosphorus loading was 188 tons per year in the years 2002–2005. This loading exceeds Missisquoi Bay’s target of 97 tons per year, which was established by the Lake Champlain Phosphorus TMDL (Total Maximum Daily Load) in 2002 (VT/NY DEC 2002). Cyanobacterial blooms in the lake have received increased concern since the poisoning of two dogs attributed to direct ingestion of lake water contaminated with cyanotoxins in 1999 and 2000 (Shambaugh & Boccuzzo, 2011). Shortly after, the Lake Champlain Basin Program (LCBP) initiated sampling to investigate the occurrence and nature of cyanobacteria (Watzin et al., 2006). As part of the Lake Champlain Long-Term Water Quality and Biological Monitoring Program (LTMP), ship-based water quality and biological samples have been collected on a bi-weekly basis at 15 lake sampling stations since 1992 (as shown in Figure 1) and net phytoplankton, cyanobacteria and zooplankton tows have been carried out since 2006. The species composition is reported in terms of cell density and biovolume (VT/NY DEC 2012).

**In situ field measurements**

The field sampling data were obtained from the Vermont DEC LTMP (http://www.anr.state.vt.us/dec/waterq/lakes/html/lp_longterm.htm). Four years of measurements between 2006 and 2009 were used including chlorophyll $a$, net phytoplankton and specific counts of cyanobacteria biovolume and cell densities from the 15 long-term monitoring stations spread throughout Lake Champlain. All algae and cyanobacteria samples were collected from the photic zone of the lake (defined as twice the Secchi depth at the
time of sampling) (VT/NY DEC 2012). Net phytoplankton and cyanobacteria biovolumes were sampled by a vertical tow of the upper three meters of the water column (VT/NY DEC 2012). All algae and cyanobacteria samples were analyzed at the Vermont DEC's laboratory according to established methods described by the US Environmental Protection Agency (USEPA 1997), the American Public Health Association (APHA 2005) and accepted limnological texts (Wetzel & Likens 2000).

**Landsat ETM+ data processing**

In all, nine ETM+ images of Lake Champlain (Path 14, Row 29) were obtained from the US Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center (http://earthexplorer.usgs.gov). Previous studies recommended a time window of ±1–2 days between satellite overpass and field measurements to ensure accurate algal detection (Stadelmann et al. 2001). The composition of algal blooms are known to be highly changeable in Lake Champlain, due to the effects of wind and precipitation; therefore, the satellite overpass time window was limited to ±1 day from the date of field sampling to establish coincident pairs.

All satellite images were processed with ENVI (ITT®, Ver. 4.7) and ArcGIS (ESRI®, Ver. 9.3) software. Based on literature, atmospheric correction over similar lakes can result in errors as high as 15% (Pozdnyakov et al. 2005). The images selected for this study met the requirement of having little to no cloud cover, in order to ensure minimal atmospheric interference during model development. Each of the ETM+ images were prescreened by inspection under different color enhancements, and only clear-sky coincident pairs were used for the image processing. Although this approach limited the number of usable Landsat 7 flyover days, it was appropriate to the scope of this project and it minimized the uncertainty and complexity involved in atmospheric correction.
The digital numbers in each band of the prescreened images were converted to exoatmospheric reflectance before model development. This conversion step allows for standardized comparison of data between multiple images from different days (Chander et al. 2009).

**Development of remote sensing cyanobacteria models**

Coincident pairs were extracted from the six multispectral ETM+ bands ($R_i$), excluding thermal and panchromatic bands. Previous studies suggested using the mean value of Landsat reflectance from a window size of $5 \times 5$ pixels to minimize the effects of satellite signal noise and account for patchiness in algae (Ekstrand 1992). This approach was used in this study to extract coincident exoatmospheric reflectance values.

Multi-linear regression was used to develop the cyanobacteria models using the ETM+'s single band and band ratio model. The correlation between exoatmospheric reflectance for each band and cyanobacteria was calculated to identify the most significant band information. Stepwise multi-linear regression was developed using the band information with the highest correlation coefficient first. The final cyanobacteria regression models were calibrated and cross-validated using coincident pairs from the late summer bloom period (July 20 to September 10) between 2006 and 2009. One-third of the coincident pairs were randomly selected for cross-validation, leaving the rest as training data for model development. In total, 20 clear-sky coincident pairs were collected and used to develop the final cyanobacteria models in the 4-year study period.

The best regression model was selected based on four criteria: (1) acceptable $p$-values (<0.05) for each of the model’s predictor coefficients; (2) the highest $R^2$ achieved with training data; (3) performance of the model using the cross-validation dataset; and (4) overall performance of the model using the entire 4-year dataset.

**RESULTS AND DISCUSSION**

The correlation results show that the ETM+ bands 2 ($r > 0.70$), 5 ($r > 0.58$), and 1 ($r > 0.46$) ($R_{B2}, R_{B5}$ and $R_{B1}$ respectively) were highly correlated to cyanobacteria in the lake (Table 1). Band ratios of 2/1 ($r > 0.80$) and 3/1 ($r > 0.58$) among others show the highest correlation with cyanobacteria.

The regression analysis was developed using ETM+ bands 2, 3, and 1, which provided the most significant information for detecting cyanobacteria. The final ETM+ models for cyanobacteria were presented by the following equations and the regression coefficients are provided in **Table 2**:

$$
cyano = \beta_0 + \beta_1 R_{B1} + \beta_2 R_{B2}$$

**Table 3** shows the results of model performance. For cyanobacterial biovolume models, the single band model with bands 2 and 3 achieved the highest coefficients of determination, $R^2$, with the training dataset with acceptable coefficient $p$-values. However, these single band models with bands 2 and 3 also showed high root-mean-square errors (RMSEs) with the validation datasets, which were more than twice as high as the RMSEs for the other models. On the other hand, band ratio models showed low RMSEs. Based on the $R^2$ values and validation results, the band ratio model was selected as the best model to predict cyanobacteria across Lake Champlain. This could be due to the band ratio model’s ability to maximize the effects of three distinctive optical features of chlorophyll $a$ in cyanobacteria. Past studies using Landsat data have similarly found improved performance with band ratio models over single band models in detecting algal chlorophyll $a$ (Hellweger et al. 2004; Vincent et al. 2004). In general, the ETM+ band ratio model performed well at detecting cyanobacteria biovolume above 0.05 mm$^3$ mL$^{-1}$, but it did poorly at lower concentrations, sometimes predicting negative concentrations when in situ levels fell below 0.01 mm$^3$ mL$^{-1}$.

However, the regression models for cyanobacteria density did not perform well. Although their $R^2$ values for both the simple band model and the band ratio model

<table>
<thead>
<tr>
<th>Cyanobacteria</th>
<th>$R_{B1}$</th>
<th>$R_{B2}$</th>
<th>$R_{B3}$</th>
<th>$R_{B4}$</th>
<th>$R_{B5}$</th>
<th>$R_{B7}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biovolume</td>
<td>0.58</td>
<td>0.78</td>
<td>0.61</td>
<td>0.38</td>
<td>-0.22</td>
<td>-0.18</td>
</tr>
<tr>
<td>Density</td>
<td>0.46</td>
<td>0.72</td>
<td>0.59</td>
<td>0.23</td>
<td>-0.16</td>
<td>-0.28</td>
</tr>
</tbody>
</table>
were high, their RMSEs were not acceptable (above 40%). Therefore, the remote sensing is applicable to detect cyanobacteria biovolume.

The final ETM+ models with band ratio were applied to nine relatively cloudless ETM+ images of Lake Champlain between 2006 and 2009. Figure 2 shows the time series results using band ratio models for Missisquoi bay in 2006. The images were given linear color scales to show the concentration and extent of cyanobacterial blooms, allowing for time-series analysis of bloom propagation, and spatial analysis of the model results in the context of adjacent watershed features.

The resulting time series images show the utility of the models for the purposes of bloom detection and monitoring. The ETM+ models captured the extent of a documented cyanobacterial bloom in July 2006 and August 2008. Cyanobacteria levels in the bay were some of the highest measured by the LTMP in the entire 4-year monitoring dataset. In addition to displaying the full spatial extent of this bloom, the imaged model results provide a striking illustration of cyanobacterial bloom during late summer.

The ETM+ model results were also used to assess how precipitation and river discharge impact bloom formation in Missisquoi Bay. Figure 3 shows daily precipitation data.
(NOAA NCDC) at Philipsburg station in Canada, located in the vicinity of Missisquoi Bay, from July 10 to August 10, 2006, plotted. The data show the impact of precipitation events on cyanobacterial blooms, which are inferred by the model results near the outlets of two major tributaries that discharge to Missisquoi Bay. On July 24, high cyanobacteria blooms were detected near the outlet of the Missisquoi and Pike Rivers whereas on August 9, the areas around the outlets of the Missisquoi and Pike Rivers are very low in cyanobacteria, with higher concentrations forming along the central and eastern portion of the bay (Figure 4). A significant precipitation event (11.7 mm) occurred on July 22, just 2 days before the flyover on July 24. The introduction of nutrients in the stormwater runoff on July 22 likely caused the blooms near the outlets. Between July 31 and August 1, severe precipitation (11.4 and 22.8 mm, respectively) occurred about a week before the Landsat flyover on August 9. After the large storm, the distribution of the blooms could shift and subsequent dry

**Figure 3** | Daily precipitation data recorded at Philipsburg weather station.

**Figure 4** | Cyanobacteria distribution over Missisquoi Bay on July 24 and August 9, 2006.
days and a small precipitation event (1.7 mm) on the day before the August 9 flyover likely led to low levels of blooms at the mouths of the two rivers. The bloom initiation and propagation in Lake Champlain could shift within a week (personal communication with Chandra Madramootoo at McGill University). More research should be conducted to confirm this theory, but it demonstrates the power of synoptic coverage of cyanobacteria in determining the drivers of algal blooms.

CONCLUSIONS

Public access to the full archive of Landsat data and the availability of environmental datasets provide many opportunities for environmental monitoring and modeling. The remote sensing models presented in this study were developed from coincident field sampling and Landsat ETM+ flyovers. The results confirm that ETM+ data can be used to provide lakewide coverage of cyanobacterial blooms. Our ETM+ models captured documented cyanobacterial blooms providing a full snapshot of cyanobacteria estimations across the entire lake. Time series and watershed analysis also demonstrates the utility of our remote sensing models for a range of resource management issues.

Even though the ETM+ sensors have high spatial resolutions that allow for detailed synoptic coverage of algal blooms, Landsat 7’s 16-day revisit period places limitations on the monitoring capabilities of the models. With further research, similar models could be developed for the Landsat 5 Thematic Mapper (TM) sensor, and the two satellites could be used in combination to provide results every 8 days, given clear-sky conditions.

Our approach demonstrates how remote sensing models can be developed and support lake monitoring programs at relatively low costs. Our approach can be applied to other waters and serve as a monitoring tool for water resource managers.

ACKNOWLEDGEMENTS

The project described in this publication was supported by Grant/Cooperative Agreement Number G11AP20085 from the United States Geological Survey and University of Massachusetts Faculty Research/Healey Endowment Grant. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the USGS.

REFERENCES

American Public Health Association (APHA) 2005 Standard Methods for the Examination of Water and Wastewater. 21st edn, Washington, DC.


National Oceanic and Atmospheric Administration (NOAA)
National Rivers Authority (NRA)
National Climatic Data Center (NCDC) http://www.ncdc.noaa.gov/cdo-web/.


US Environmental Protection Agency (USEPA) 1997 *In Vitro Determination of Chlorophyll a and Phycocyanin in Marine and Freshwater Algae by Fluorescence (Revision 1.2)*. Washington, DC.

USEPA 2011 Reconsideration of EPA’s Approval of Vermont's 2002 Lake Champlain Phosphorus Total Maximum Daily Load (TMDL) and Determination to Disapprove the TMDL. Washington, DC.


