Remote sensing models using Landsat satellite data to monitor algal blooms in Lake Champlain
A. Trescott and M.-H. Park

ABSTRACT

Lake Champlain is significantly impaired by excess phosphorus loading, requiring frequent lake-wide monitoring for eutrophic conditions and algal blooms. Satellite remote sensing provides regular, synoptic coverage of algal production over large areas with better spatial and temporal resolution compared with in situ monitoring. This study developed two algal production models using Landsat Enhanced Thematic Mapper Plus (ETM⁺) satellite imagery: a single band model and a band ratio model. The models predicted chlorophyll a concentrations to estimate algal cell densities throughout Lake Champlain. Each model was calibrated with in situ data compiled from summer 2006 (July 24 to September 10), and then validated with data for individual days in August 2007 and 2008. Validation results for the final single band and band ratio models produced Nash–Sutcliffe efficiency (NSE) coefficients of 0.65 and 0.66, respectively, confirming satisfactory model performance for both models. Because these models have been validated over multiple days and years, they can be applied for continuous monitoring of the lake.

Key words | algal blooms, chlorophyll a, eutrophication, remote sensing

INTRODUCTION

Many lakes in the USA have been impaired by eutrophication and algal blooms. For example, Lake Champlain has been impaired by excess phosphorus loading associated with agricultural, residential and urban runoff, resulting in severe eutrophication in many areas (USEPA 2011). Nutrient loading to the lake is primarily associated with terrestrial runoff from seasonal snowmelt and rainfall, but real-time drivers that spur algal blooms and toxins in many regions are poorly understood. The timing and intensity of bloom propagation are difficult to predict. Frequent lake-wide monitoring is required to understand the dynamics of algal blooms in Lake Champlain.

Traditional lake monitoring programs rely on shipboard field sampling to determine in situ concentrations of chlorophyll a (Chl-a), a photosynthetic pigment used as a proxy to measure algal biomass (Carlson 1977; Coskun et al. 2008), among other parameters. Chl-a concentrations are often determined with established in vitro laboratory analyses where a solvent is used to extract chlorophyll pigments from phytoplankton samples. The optical signals of the extracted pigments are then measured using a fluorometer (Arar & Collins 1997) or spectrophotometer (Arar 1997). Although recent advancement of fluorometer technology provides real-time measurements (Oxborough 2004; Izydorczyk et al. 2009; Richardson et al. 2010), in situ monitoring still has limitations on spatial coverage that cannot provide the distribution of algal blooms. Due to the inherent patchiness of blooms, measured algal concentrations taken from sampling stations may not be representative of levels throughout the study area and important seasonal fluctuations of a bloom’s size and duration may be missed (Shafique et al. 2001).

Satellite remote sensing can provide regular, synoptic coverage of algal blooms over large areas for regional monitoring programs at resolutions unattainable by field measurements (Richardson 1996; Park & Ruddick 2007). Satellite measurements are especially useful for algae detection because of the unique spectral absorbance/reflectance characteristics of Chl-a. Chl-a in algae-laden water exhibits four distinct reflectance/absorption features, including: (1) high absorbance of blue light between 400 and 500 nm; (2) a reflectance peak in the green region at approximately 550 nm; (3) strong absorption of red light near 675 nm; and (4) a pronounced reflectance peak between 690 and
700 nm (Jensen 2007). The relationship of light reflected or absorbed at specific wavelengths (λ) measured by a satellite sensor can be used to estimate Chl-a concentrations with bio-optical algorithms (O’Reilly et al. 1998; Brown et al. 2008).

Scientists have been using satellites to estimate water quality parameters since the early 1970s (Ekstrand 1992). Satellite optical images at varying degrees of spatial, spectral and temporal resolution have been used to provide the distribution of Chl-a to estimate phytoplankton: Coastal Zone Color Scanner (Gordon et al. 1983; Kutser et al. 1995), the Sea-viewing Wide Field-of-View Sensor (SeaWiFS) (Vos et al. 2003; Budd & Warrington 2004; Shuchman, et al. 2006; Werdell et al. 2009), the Moderate Resolution Imaging Spectroradiometer (MODIS) (Bergamino, et al. 2009; Werdell et al. 2009; Wang et al. 2011; Binding et al. 2012; Chang et al. 2012), and the Medium Resolution Imaging Spectrometer (MERIS) (Park & Ruddick 2007; Gitelson et al. 2008; Moses et al. 2009; Binding et al. 2011).

Satellite data from the Landsat series are useful for studying inland lakes. In previous studies, empirical Chl-a algorithms for Landsat imagery are mostly developed using the reflectance of bands 1 (blue), 3 (red), and sometimes band 2 (green) (Hellweger et al. 2004; Sass et al. 2007). Though Landsat visible bands are relatively broad (Band 1: 450–515 nm; Band 2: 525–605 nm; Band 3: 630–690 nm) compared to other multi-spectral sensors (e.g. SeaWiFS, MODIS and MERIS), these bands are capable of capturing Chl-a’s optical signature. High spatial resolution of Landsat data (28.5 m for reflective bands) allows for detailed analysis of optical features in large open waters as well as smaller lakes and embayments (>5 ha) (Sass et al. 2007). Landsat is good for Lake Champlain because small embayments such as Missisquoi Bay and St Albans Bay have the most problems associated with eutrophication. These areas are too small for meaningful observations by other available sensors with coarse resolution (e.g. 1 km for MODIS or SeaWiFS and 300 m for MERIS). Moreover, Landsat provides the longest continuous data archives since 1972, which can be used for retrospective analyses of water quality.

The objective of this study was to develop an algal production model using remote sensing to predict Chl-a throughout Lake Champlain. The algal production models were developed using images from the multispectral Landsat Enhanced Thematic Mapper Plus (ETM+). The satellite models were calibrated and validated using a 3-year composite of in situ water quality monitoring data.

METHODS

Lake Champlain borders the States of Vermont, New York, USA and the Province of Quebec, Canada. With a surface area of 1,127 km² and a maximum depth of 122 m, the lake provides a drinking water source to about 200,000 people living in the surrounding basin (LCBP 2006; USEPA 2011). In Lake Champlain, the Vermont Department of Environmental Conservation (DEC) and the New York State DEC have performed regular monitoring of water quality and biological parameters at 15 sampling stations to detect long-term environmental change and effects of management actions (Vermont DEC and New York State DEC 2010). After a dog-poisoning caused by cyanobacteria toxins occurred in 1999, the Lake Champlain Basin Program (LCBP) began a study to monitor the occurrence and extent of potential toxin-producing cyanobacteria in 2000. Monitoring is performed based on weekly ship-based sampling of the lake-wide stations in the Vermont DEC’s long-term monitoring program. Additional lake sampling stations, collected and analyzed by the University of Vermont, were added to regions with high bloom-frequencies in 2003 to further assess the occurrence of cyanobacteria toxins. Though the existing monitoring program in Lake Champlain provides invaluable in situ measurements, much of the data are not available rapidly enough to drive an instantaneous public alert system for potentially harmful bloom levels (LCBP 2006).

The models developed in this study are based on in situ water quality data for Lake Champlain, obtained from the Vermont DEC and the LCBP (Figure 1). The 3-year water quality monitoring dataset was obtained for the 2006, 2007 and 2008 sampling seasons. They sampled lake water from the photic zone, defined as twice the Secchi disk depth at the time of sampling, at established long term monitoring stations that were located by GPS latitude and longitude coordinates. They analyzed samples for Chl-a concentrations and total density of net phytoplankton according to established methods described by the US Environmental Protection Agency (USEPA 1997) and the American Public Health Association (APHA et al. 2005).

Acquisition and pre-processing of Landsat data

Landsat ETM+ images (Path 14, Row 29) were obtained from the US Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center
Previous remote sensing studies recommend a time gap between satellite overpass and field measurement of ±1 – 2 days for good predictive chlorophyll values (Stadelmann et al. 2000). However, the shape and composition of algal blooms in Lake Champlain are highly changeable due to the effects of wind and precipitation. Therefore, a satellite overpass time window of ±1 day from in situ sampling was used to establish coincident pairs. Any replicate in situ samples collected for a specific station within the overpass time window were averaged to give one Chl-a value per station per overpass.

For the purposes of this study, all satellite images were processed with ENVI (Ver. 4.7, ITT®) and ArcGIS (Ver. 9.3, ESRI®) software. To minimize atmospheric interference, all satellite images were prescreened for cloud cover and haze by inspection under different color enhancements. The reflective bands in each of the prescreened images were further converted to exoatmospheric (or ‘top-of-atmosphere’) reflectance before analysis using ENVI, which allows for standardized comparison of data within one scene or between multiple images from different days. The process involves converting the radiometrically calibrated digital numbers to at-sensor spectral radiance ($L_\lambda$), followed by conversion of $L_\lambda$ values to exoatmospheric reflectance (Chander et al. 2009).

Development of chlorophyll-a models

For each coincident in situ Chl-a measurement, the mean exoatmospheric reflectance values ($R_\lambda$) from the six multispectral ETM+ bands were extracted from a 5 by 5 pixel window, centered on the nearest sampling station coordinates. The window size, used in previous studies (Ekstrand 1992), is considered to minimize the effects of noise in the satellite data and account for patchiness of algae.
In addition to single band information, spectral reflectance ratios ($R_{i}/R_{j}$, where $R_{i}$ is reflectance of the $i$th band and $R_{j}$ is reflectance of the $j$th band) were calculated for each of the non-reciprocal pairs of the six reflectance bands, following the method outlined in Vincent et al. (2004).

Step-wise multiple linear regression was used to develop the remote sensing Chl-a algorithms using all combinations of single-band and non-reciprocal reflectance band-ratios for coincident pairs. The final regression models were calibrated using coincident pairs obtained in summer 2006 (July 24 to September 10, including three images with a total of 16 coincident pairs), which is the time period associated with the most severe algal blooms in Lake Champlain. The single-band and band-ratio Chl-a models were validated by applying them to ETM+ images in August 2007 and 2008. Because 2007 is a known low-bloom year, whereas 2008 is a known high-bloom year, basing the validation on days from each of these years provided a good indication of the strength of each model across a range of bloom conditions.

To assess the performance of each model, the Nash–Sutcliffe efficiency (NSE) was applied to the prediction results (Nash & Sutcliffe 1970):

$$
\text{NSE} = 1 - \frac{\sum_{i=1}^{n} (X_{i,\text{obs}} - X_{i,\text{prd}})^2}{\sum_{i=1}^{n} (X_{i,\text{obs}} - \bar{X}_{i,\text{obs}})^2}
$$

(1)

where $n$ is the number of values, $X_{i,\text{obs}}$ and $X_{i,\text{prd}}$ are measured and predicted values, respectively.

NSE is widely used in water resources and environmental modeling assessments. It represents a normalized statistic that determines the relative magnitude of the residual variance (noise) to observed variance, indicating how well the plot of observed versus predicted data fits the 1:1 line (Nash & Sutcliffe 1970). NSE ranges from $-\infty$ to 1.0, but model performance can be judged as satisfactory for NSE $>0.5$ (Moriasi et al. 2007).

The best single band and band-ratio models that are based on the highest coefficients of determination and acceptable $p$-values for each parameter coefficient were then applied to entire satellite scenes to obtain Chl-a estimates for all of Lake Champlain.

**RESULTS AND DISCUSSION**

The following shows the best single band and band ratio models to detect Chl-a in Lake Champlain. Both models showed strong correlations between predicted and actual Chl-a (Figure 2):

$$
\text{Chl-a(µg/L)} = -666.18 \times (R_{B1}) + 840.39 \times (R_{B2}) + 16.14
$$

(2)

$$
\text{Chl-a(µg/L)} = 85.34 \times (R_{B2}/R_{B1}) - 26.54 \times (R_{B3}/R_{B1}) - 39.10
$$

(3)

where $R_{B1}$, $R_{B2}$ and $R_{B3}$ are the normalized exoatmospheric reflectance values for Landsat ETM+ bands 1, 2 and 3, respectively. $R_{Bi}/R_{Bj}$ stands for the reflectance ratio of the $i$th band over the $j$th band.

The single band model (Equation (2)) used the Chl-a reflectance minimum in the sensor’s blue-sensitive band (B1) and the reflectance maximum in the green-sensitive band (B2). Previous studies had found Landsat’s red-sensitive band (B3) useful in single band Chl-a models because it is

![Figure 2](https://iwaponline.com/wst/article-pdf/67/5/1113/441759/1113.pdf)
least sensitive to atmospheric effects, yet it has the lowest penetration distance in the water column (Richardson 1996; Sass et al. 2007). The statistical tests performed during this step-wise multiple linear regression showed B1 and B2 to be the most significant Chl-a predictors in single-band combinations. This is a key difference from other studies, likely due to the collection method of Chl-a water samples in Lake Champlain. All Chl-a samples were collected as representative composites of the photic zone. The higher water column penetrations achieved by Landsat ETM+ bands 1 and 2 may give better predictions of phytoplankton present in the entire photic zone rather than surficial concentrations collected via grab sampling. This is a significant finding with a better prediction of algal production through the entire photic zone as opposed to the surface layer of the lake. It should be noted, however, that ETM+ bands 1 and 2 are more susceptible to atmospheric interference, therefore limiting this single band model to clear sky coincident pairs.

The band ratio model (Equation (3)) only slightly outperformed the single band model by using the ratio of the Chl-a green reflectance peak to the blue reflectance minimum, as well as the ratio of the red reflectance trough to the blue reflectance minimum. These two ratios maximize the effects of three distinctive optical features of Chl-a. The improved performance of the band ratio model over the single band model agrees with the findings of past studies (Vincent et al. 2004) that found more robust results using band ratios.

The validation results for both models were similar and followed general in situ trends (Figure 3). Most of the observation concentration falls within the prediction interval range except one (Station 40, 2008). The band ratio model had a root mean square error (RMSE) of 2.48 μg/L while the single band model had a RMSE of 2.52 μg/L. The NSE coefficients for single band and band ratio models were 0.65 and 0.66, respectively. The NSE results imply satisfactory performance for both models.

Based on the relationship of Chl-a to net phytoplankton cell densities determined by the in situ dataset, the validated Landsat ETM+ band ratio Chl-a model results were converted to approximate phytoplankton cell densities to give a clearer picture of algal bloom distributions in Lake Champlain. Figure 4 shows the linear model of phytoplankton cell density using measured Chl-a and time series maps of Chl-a concentration and algal bloom distribution in 2006. The figure focuses on Missisquoi Bay because the bay is one of the most eutrophic areas of lake. The results of bloom patterns from July 24, August 9 and August 24 show significant concentrations forming in the northeastern side of the bay, in close proximity to a major tributary (Pike River), while the result on September 10 reveals the effects of wind-driven advection, spreading algae in a southern direction. The July and September images show the most intense blooms compared to August.

CONCLUSION

Eutrophication remains a significant problem in many regions of Lake Champlain. Algal blooms often shift and
change composition dramatically in a short period of time, highlighting the need for continued lake-wide monitoring efforts. The remote sensing models developed in this study successfully predicted Chl-α concentrations and algal bloom distribution throughout Lake Champlain over multiple years. Inclusion of Landsat remote sensing models developed in this study would benefit Lake Champlain’s monitoring program by providing regular synoptic coverage of algal productivity throughout the lake at a very high spatial resolution. Both the band ratio model and single band model provided satisfactory performance to estimate Chl-α concentrations in the lake. Because these models have been validated over multiple days and years, they can be applied to continuous monitoring efforts. The cumulative data availability of Landsat series can be used for retrospective study. New Landsat satellite planned to launch in early 2013 will provide continuing monitoring with enhanced resolution. Although the number of bands and bandwidth of the new Landsat sensor will be slightly different from those of Landsat 7, the bandwidth of the visible will be very similar. Therefore our approach will be applicable to detect algal blooms for continuous monitoring, although the model might need to be calibrated. Time series analysis using Landsat model results could also provide resource managers with a clearer understanding of bloom dynamics in eutrophic areas and point to the various land use, environmental, and climate-based drivers of eutrophication.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Mary Watzin at the University of Vermont’s Rubenstein Lab and the Vermont Department of Environmental Conservation for providing the in situ water quality dataset used for the models developed in this study.

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

REFERENCES


LCBP (Lake Champlain Basin Program) 2006 Monitoring and evaluation of cyanobacteria in Lake Champlain: Summer 2006. LCBP (Lake Champlain Basin Program) 2012 Lake Champlain Long-Term Water Quality and Biological Monitoring Program.


US Environmental Protection Agency 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.

Vermont Department of Environmental Conservation (DEC) and New York State DEC 2010 Lake Champlain long-term water quality and biological monitoring program: program description.


First received 1 May 2012; accepted in revised form 15 October 2012