International Journal of Applied Earth Observation and Geoinformation 18 (2012) 368–385

Contents lists available at SciVerse ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Hyperspectral retrieval of phycocyanin in potable water sources using genetic algorithm–partial least squares (GA–PLS) modeling

Kaishan Song[a,b,*], Lin Li[a], Shuai Li[a], Lenore Tedesco[a], Bob Hall[a], Zuchuan Li[a]

[a] Department of Earth Sciences, Indiana University-Purdue University, Indianapolis, IN, USA
[b] Northeast Institute of Geography and Agricultural Ecology, CAS, Changchun, PR China

ARTICLE INFO

Article history:
Received 23 August 2011
Received in revised form 20 March 2012
Accepted 26 March 2012

Keywords:
AISA
Cyanobacteria
GA–PLS
Hyperspectral
Phycocyanin
Semi-empirical model

ABSTRACT

Eagle Creek, Morse and Geist reservoirs, drinking water supply sources for the Indianapolis, Indiana, USA metropolitan region, are experiencing nuisance cyanobacterial blooms. Hyperspectral remote sensing has been proven to be an effective tool for phycocyanin (C-PC) concentration retrieval, a proxy pigment unique to cyanobacteria in freshwater ecosystems. An adaptive model based on genetic algorithm and partial least squares (GA–PLS), together with three-band algorithm (TBA) and other band ratio algorithms were applied to hyperspectral data acquired from in situ (ASD spectrometer) and airborne (AISA sensor) platforms. The results indicated that GA–PLS achieved high correlation between measured and estimated C-PC for GR (RMSE = 16.3 μg/L, RMSE% = 18.2; range (*R*): 2.6–185.1 μg/L), MR (RMSE = 8.7 μg/L, RMSE% = 15.6; *R*: 3.3–371.0 μg/L) and ECR (RMSE = 19.3 μg/L, RMSE% = 26.4; *R*: 0.7–245.0 μg/L) for the in situ datasets. TBA also performed well compared to other band ratio algorithms due to its optimal band tuning process and the reduction of backscattering effects through the third band. GA–PLS (GR: RMSE = 24.1 μg/L, RMSE% = 25.2, *R*: 25.2–185.1 μg/L; MR: RMSE = 15.7 μg/L, RMSE% = 37.4, *R*: 2.0–135.1 μg/L) and TBA (GR: RMSE = 28.3 μg/L, RMSE% = 30.1; MR: RMSE = 17.7 μg/L, RMSE% = 41.9) methods results in somewhat lower accuracy using AISA imagery data, which is likely due to atmospheric correction or radiometric resolution. GA–PLS (TBA) obtained an RMSE of 24.82 μg/L (35.8 μg/L), and RMSE% of 31.24 (43.5) between measured and estimated C-PC for aggregated datasets. C-PC maps were generated through GA–PLS using AISA imagery data. The C-PC concentration had an average value of 67.31 ± 44.23 μg/L in MR with a large range of concentration, while the GR had a higher average value 103.17 ± 33.45 μg/L.

1. Introduction

Notorious for their negative impact on water quality, cyanobacterial blooms have been increasingly the subject of water management and scientific studies (Dekker, 1993; Simis et al., 2005; Paerl and Huisman, 2008). These nuisance and sometimes harmful phytoplankton (e.g., *Anabaena*, *Aphanizomenon, Planktothrix* and *Cylindrospermopsis*) blooms can result in both aesthetic degradation and resource use limitations of lakes and reservoirs due to the production of surface scums and musty, earthy smell taste and odor metabolites (Chorus and Bartram, 1999). Blooms also cause recreational degradation due to ecosystem degradation and human and animal health risks, including fatal human liver, neurological and skin diseases, caused by the production of toxins such as anatoxins, microcystins, and cylindrospermopsin (Codd et al., 2005; Huisman et al., 2005). Long-term low level exposures to microcystin have been suspected to contribute to high rates of liver cancer (Guo, 2007), and a short-term exposure to anatoxin has likely caused death of wild animals (Behm, 2003).

Cyanobacteria growth is dependent on temperature, light, and nutrient concentrations and is often associated with eutrophication (Codd et al., 2005; Paerl and Huisman, 2008), which is a natural process hastened by anthropogenic activity (Paerl and Huisman, 2008). As nitrogen and phosphorous levels rise in water bodies, conditions become more conducive for cyanobacterial blooms (Codd et al., 2005). Current monitoring practices often involve widely dispersed station sampling and laboratory analysis. The ephemeral nature of algal blooms makes this traditional approach ill-suited for monitoring inland waters at large scale due to the patchy distribution of algae blooms (Dekker et al., 1991; Simis et al., 2007; Hunter et al., 2010). Remote sensing offers an alternative to in situ field based monitoring by providing a synoptic view of target water quality parameters (Gitelson, 1992; Schalles and Yacobi, 2000; Simis et al., 2005; Voutilainen

* Corresponding author at: Department of Earth Sciences, Indiana University-Purdue University, Indianapolis, IN, USA. Tel.: +1 317 274 8383.
E-mail addresses: kaissong@iupui.edu, songks@neigae.ac.cn (K. Song).

0303-2434/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jag.2012.03.013

et al., 2007; Hunter et al., 2010). It is widely accepted that remote sensing can be used as a powerful tool for monitoring chlorophyll-*a* (Chl-*a*) spatiotemporal dynamics due to its diagnostic absorption spectral band around 443 and 675 nn (Gitelson, 1992; Gons, 1999; Gons et al., 2008; Igamberdiev et al., 2011). C-phycocyanin (C-PC), a pigment mainly specific to cyanobacteria and some cryptophytes as a minor pigment, demonstrates a diagnostic spectral absorption in freshwater systems at 620 nm (Dekker, 1993; Gitelson et al., 1999; Randolph et al., 2008), which makes the remote detection of cyanobacteria possible (Dekker, 1993; Schalles and Yacobi, 2000; Simis et al., 2005; Hunter et al., 2010).

To date, little research has been done to map C-PC using multi- and hyper-spectral remote sensing (Mille et al., 1992; Vincent et al., 2004; Hunter et al., 2010; Guanter et al., 2010). The first example used multiple linear regressions between different combinations of TM band ratios and measured concentration to map spatial distribution of C-PC in Lake Erie (Vincent et al., 2004), but this algorithm is saturated when C-PC abundance is greater than 14 μg/L. The second case mapped high concentrations of C-PC in a pond near southern San Francisco Bay though the use of spectral mixture analysis (SMA) of AVIRIS data (Richardson, 1996; Kruse et al., 1997). SMA is a simple additive linear model used to estimate the abundances of the materials measured by the imaging spectrometer. Although this approach has been widely used in mapping land use and land cover change as well as geological mapping, it requires a spectrally "pure" endmember which is completely impossible with in situ water samples. Schalles and Yacobi (2000) proposed an algorithm based on the band ratio R_{648}/R_{624} to determine C-PC concentration.

Simis et al. (2005) created a semi-empirical model for determining C-PC abundance using the optical properties of C-PC and the attenuation and backscattering of other optically active constituents (OACs) present in turbid inland water, particularly specific absorption (C-PC spectral absorption per unit concentration). This algorithm was developed using a portable spectroradiometer (Photoresearch, PR-650) to accommodate other remote sensing platforms (i.e., MEdium Resolution Imaging Spectrometer, MERIS). Application of the Simis et al. (2005) algorithm to field spectra collected on two lakes in the Netherlands yielded an R^2 value of 0.94, with limited samples (n = 34). Investigations have been made to apply MERIS satellite data to monitor C-PC for inland waters (Guanter et al., 2010; Matthews et al., 2010). Due to the weak absorption feature for C-PC, inherent optical properties (IOPs) determination for C-PC is a challenge to date, which makes bio-optical determination of C-PC unsuitable.

The spatial transferability of developed models is a challenge for the remote sensing community (D'Alimonte et al., 2003; Vincent et al., 2004; Gitelson et al., 2008). Since the three-band algorithm (TBA) has been tested more universally and has been shown to be stable for Chl-*a* concentration (Gitelson et al., 2008; Sun et al., 2009), the spatial transferability of models will be compared with both GA–PLS and TBA. This research will address a new method for creating empirical algorithms by using genetic algorithms with partial least squares analyses (GA–PLS). As a comparison, three-band model (Dall'Olmo and Gitelson, 2005) will be explored for its potential applicability to C-PC inversion (Hunter et al., 2008; Guanter et al., 2010) for inland productive potable waters. Specifically, the objective of this analysis is threefold: (1) to determine the optimal band ratios and TAB for C-PC retrieval with higher accuracy; (2) to assess the GA–PLS model for C-PC concentration estimation with in situ collected spectral data with comparison to TAB; and (3) to map C-PC concentration with airborne imaging spectrometer for application (AISA) data for inland water bodies which are confounded with other optically active constituents.

2. Materials and methods

2.1. Study sites

Eagle Creek Reservoir (ECR: W 86°18′13.07″, N 39°51′09.84″; surface area (A) = 5.0 km^2; mean water depth (Z) = 4.2 m; volume (V) = 21.0 million m^3), Morse Reservoir (MR: W 86°2′17.22″, N 40°6′16.84″; A = 6.0 km^2; Z = 4.7 m; V = 28.0 million m^3) and Geist Reservoir (GR: W 85°57′47.22″, N 39°55′16.84″; A = 5.9 km^2; Z = 3.2 m; V = 23.8 million m^3) are major components of the drinking water system and used for recreation for over 900,000 residents of the Indianapolis, Indiana (Fig. 1). The major features of the three reservoirs are summarized in Table 1. GR has undergone dredging for the acquisition of sand and gravel since 2001. Drinking water managers have documented blooms of nuisance and harmful algae in all three reservoirs (Tedesco et al., 2005). Indiana Department of Environmental Management (IDEM) has classified the trophic status of the three reservoirs from the mesotrophic to eutrophic range (IDEM, 2006).

2.2. In situ data collection

Field campaigns were carried out in 2005 and 2006. Samples were taken under a diverse range of algal bloom conditions. In situ water measurements were collected with YSI 6600 V-2 multi-parameter probe (YSI Inc., Yellow Springs, OH), including electrical conductivity (mS), turbidity (NTU), and pH value. The station coordinates were recorded using global positioning system (GPS) and water clarity was estimated using a Secchi disk (SDD). Surface water grab samples were collected at each location at approximately 0.25 meters below the water surface. Samples were held on ice until preprocessed for analyzing C-PC, Chl-*a*, and total suspended solids (TSS). Two field surveys for 2005 and 17 field surveys for 2006 were conducted (see Tables 2 and 3 for detail).

2.3. Spectroscopic measurements

In situ remote sensing reflectance was collected using an ASD ultraviolet/visible and near-infrared (UV/VNIR) spectrophotometer (ASD Inc., Boulder CO). The detail measurement procedures can be found in Randolph et al. (2008). Remote sensing reflectance (R_{rs}; sr^{-1}) was obtained using the ratio of upwelling water-leaving radiance (L_w; W m^{-2} sr^{-1}) at a nadir viewing angle to the downwelling irradiance (L_d; W m^{-2}):

$$R_{rs} = \frac{L_w(0^+, \lambda)}{E_d(0^+, \lambda)} \quad (1)$$

where L_w is derived from substracting total upwelling radiance (L_{up}) at 900 nm from L_{up} for each wavelength from 350 to 900 nm; E_d denotes downwelling irradiance measured at each sample site using a white reference panel (99% Lambertian reflector).

2.4. Airborne hyperspectral images

2.4.1. Image acquisition

Airborne hyperspectral data were collected using an AISA-Eagle (Spectral Imaging Ltd. Oulu, Finland) sensor on board a Piper Saratoga airplane owned by the University of Nebraska, Lincoln (UNL) Center for Advanced Land Management Information Technologies (CALMIT). This airborne sensor has a programmable set-up, allowing the collection of data in up to 512 discrete channels through the spectral range of 400–1000 nm. Detailed information on image acquisition can be found in Li et al. (2010). Images with 62 bands in the spectral region of approximately 392–982 nm with a bandwidth of 7–8 nm was acquired with 1 m spatial resolution.