ISPRS Journal of Photogrammetry and Remote Sensing 95 (2014) 68–80

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Remote quantification of phycocyanin in potable water sources through an adaptive model

Kaishan Song [a,b,*], Lin Li [b], Lenore P Tedesco [c], Shuai Li [d], Bob E Hall [b], Jia Du [a]

[a] Northeast Institute of Geography and Agroecology, CAS, Changchun, JL 130102, China
[b] Department of Earth Science, Indiana University-Purdue University, Indianapolis, IN 46202, USA
[c] The Wetlands Institute, 1075 Stone Harbor Blvd, Stone Harbor, NJ 08247, USA
[d] Department of Geological Sciences, Brown University, Providence, RI 02912, USA

ARTICLE INFO

Article history:
Received 3 November 2013
Received in revised form 6 April 2014
Accepted 9 June 2014

Keywords:
Cyanobacteria
Hyperspectral
PLS–ANN
Productive waters
Drinking water sources
Phycocyanin
Three-band model

ABSTRACT

Cyanobacterial blooms in water supply sources in both central Indiana USA (CIN) and South Australia (SA) are a cause of great concerns for toxin production and water quality deterioration. Remote sensing provides an effective approach for quick assessment of cyanobacteria through quantification of phycocyanin (PC) concentration. In total, 363 samples spanning a large variation of optically active constituents (OACs) in CIN and SA waters were collected during 24 field surveys. Concurrently, remote sensing reflectance spectra (R_{rs}) were measured. A partial least squares–artificial neural network (PLS–ANN) model, artificial neural network (ANN) and three-band model (TBM) were developed or tuned by relating the R_{rs} with PC concentration. Our results indicate that the PLS–ANN model outperformed the ANN and TBM with both the original spectra and simulated ESA/Sentinel-3/Ocean and Land Color Instrument (OLCI) and EO-1/Hyperion spectra. The PLS–ANN model resulted in a high coefficient of determination (R^2) for CIN dataset ($R^2 = 0.92$, R: 0.3–220.7 μg/L) and SA ($R^2 = 0.98$, R: 0.2–13.2 μg/L). In comparison, the TBM model yielded an $R^2 = 0.77$ and 0.94 for the CIN and SA datasets, respectively; while the ANN obtained an intermediate modeling accuracy (CIN: $R^2 = 0.86$; SA: $R^2 = 0.95$). Applying the simulated OLCI and Hyperion aggregated datasets, the PLS–ANN model still achieved good performance (OLCI: $R^2 = 0.84$; Hyperion: $R^2 = 0.90$); the TBM also presented acceptable performance for PC estimations (OLCI: $R^2 = 0.65$, Hyperion: $R^2 = 0.70$). Based on the results, the PLS–ANN is an effective modeling approach for the quantification of PC in productive water supplies based on its effectiveness in solving the non-linearity of PC with other OACs. Furthermore, our investigation indicates that the ratio of inorganic suspended matter (ISM) to PC concentration has close relationship to modeling relative errors (CIN: $R^2 = 0.81$; SA: $R^2 = 0.92$), indicating that ISM concentration exert significant impact on PC estimation accuracy.

1. Introduction

As population growth continues to place increasing demand on drinking water supplies, resource managers urgently need better assessment tools to protect and maintain the quantity and quality of our drinking water sources (Li et al., 2010; Song et al., 2012a). Unpleasant and sometimes harmful cyanobacterial blooms degrade water quality due to the production of surface scums, toxin, and earthy compounds affecting the water taste and odor (Codd et al., 2005; Paerl and Paul, 2012). Monitoring programs to track cyanobacterial blooms and the conditions conducive to bloom formation are often limited to a small number of stations that are sampled infrequently (Hunter et al., 2009; Guanter et al., 2010). Although water supply sources have been studied intensively (Randolph et al., 2008; Gitelson et al., 2008; Song et al., 2012a), water resource managers lack a powerful assessment tool capable of providing timely information on the spatial distribution and concentration of algal communities (Hunter et al., 2009; Li et al., 2010). This study will address this practical need by demonstrating that remote sensing techniques can provide a fast and efficient method for determining the intensity of cyanobacterial blooms (Randolph et al., 2008; Hunter et al., 2009, 2010).

Cyanobacterial growth is dependent on temperature, light, and nutrient availability and is often associated with eutrophication (Codd et al., 2005; Guo, 2007; Paerl and Paul, 2012), which is a

* Corresponding author.
E-mail address: songks@neigae.ac.cn (K. Song).

http://dx.doi.org/10.1016/j.isprsjprs.2014.06.008

natural process hastened by anthropogenic activity (Chorus and Bartram, 1999; Paerl and Paul, 2012). As nitrogen and phosphorus levels increase in water bodies, the conditions become more conducive for cyanobacterial blooms (Codd et al., 2005; Tedesco and Clercin, 2011, Paerl and Paul, 2012). Current monitoring practices often involve widely dispersed station sampling and laboratory analysis (Simis et al., 2005; Paerl and Paul, 2012). The ephemeral nature of algal blooms makes effective monitoring in this manner difficult (Gons et al., 2008; Hunter et al., 2010). Remote sensing offers an alternative to minimal station monitoring by providing a synoptic view of the target feature (Gons et al., 2008; Hunter et al., 2010). It is widely accepted that remote sensing can be used as a powerful tool for monitoring the spatiotemporal dynamics of chlorophyll-a (Chl-a) (Gons et al., 2008; Yang et al., 2011; Moses et al., 2012). Phycocyanin (PC), a pigment unique to cyanobacteria, demonstrates a diagnostic spectral absorption in freshwater systems based on its diagnostic spectral band located around 620 nm, including the maximum absorption of a number of modifications of PC (Dekker 1993; Ruiz-Verdú et al., 2007), which makes the remote detection of cyanobacteria possible (Dekker, 1993; Schalles and Yacobi, 2000; Simis et al., 2005; Song et al., 2012b).

Multiple algorithms have been developed to estimate PC concentrations using remotely sensed data. These include the models created by Dekker (1993), Schalles and Yacobi (2000) based on empirical algorithms and the semi-analytical algorithm developed by Simis et al. (2005, 2007). A better understanding of water quality constituents that impact the PC concentration estimation derived from algorithms is necessary to improve their predicative capabilities and utility to water resource managers. Currently, only a few successful studies have used multi- and hyper-spectral remote sensing to map PC concentrations in inland waters (Vincent et al., 2004; Hunter et al., 2010; Guanter et al., 2010; Song et al., 2012b). Simis et al. (2005) developed an optical model for determining this ancillary pigment abundance using the optical properties of PC along with the attenuation and backscattering of other optically active constituents (OACs) presented in turbid inland waters. This algorithm was developed using a portable spectroradiometer (Photoresearch, PR-650) to accommodate other remote sensing platforms (e.g., the MEdium Resolution Imaging Spectrometer, MERIS). Investigations have been made to apply MERIS satellite data to monitor PC concentrations in inland waters (Guanter et al., 2010; Matthews et al., 2010).

Investigations have proved that the accuracy of the remote estimation of Chl-a for inland turbid productive water is strongly influenced by nonalgal particles and colored dissolved organic matter (CDOM) (Schalles and Yacobi, 2000; Gitelson et al., 2008; Gilerson et al., 2010; Yang et al., 2011). The package effects and impact from ancillary pigments are also factors that influence Chl-a estimates (Babin et al., 2003). Compared to Chl-a, PC demonstrates even weaker absorption, which makes it even more challenging for remote estimation (Simis et al., 2005). Nonlinearity is the major issue for PC estimation using remotely sensed data (Ruiz-Verdú et al., 2007; Randolph et al., 2008). Water quality remote-sensing algorithms basically employ upwelling radiance or remote sensing reflectance (input) to retrieve water quality parameters (output). Algorithm development can be regarded as a regression problem where once the functional form of the algorithm is determined, the parameters of the function are derived from a set of input–output pairs (D'Alimonte et al., 2003; Bricaud et al., 2007; Song et al., 2013). The pairs are affected by the training dataset range and representative values for model performance.

The interactions between PC, non-algal particles (dominated by suspended mineral particles), and other ancillary pigments have restricted the development of empirical algorithms to specific regions, due to the nonlinearity of various water constituents (Ruiz-Verdú et al., 2007; Hunter et al., 2010; Yang et al., 2011; Song et al., 2014). An adaptive model based upon spectral variables derived from *in situ* reflectance was developed with a partial least squares-artificial neural network (PLS–ANN) based on the PC diagnostic spectral band or band ratios (Song et al., 2014), which shows potential for total suspended matter (TSM) and total phosphorus estimation with both *in situ* and airborne imaging data (Song et al., 2012a, b). Our research will address this technical needs for effective remote sensing of PC concentration by demonstrating the PLS–ANN performance and testing the applicability of the semi-empirical algorithms developed by Dall'Olmo and Gitelson (2005) for Chl-a estimates in turbid inland waters, which has a potential applicability to PC inversion (Guanter et al., 2010; Song et al., 2012b) for inland productive waters. The objectives are twofold: (1) to examine the PLS–ANN model for PC estimates using *in situ* collected spectral data and satellite-borne sensor simulated spectra, e.g., Sentinel-3/OLCI scheduled launching in 2015 and EO-1/Hyperion, in comparison with an ANN and a semi-empirical model adapted from three-band model (TBM); and (2) to examine the major optically active constituents (OACs) that confound the PC estimates for productive drinking water sources.

2. Materials and methods

2.1. Study sites

The Eagle Creek (ECR: 86°18′13.07″ W, 39°51′09.84″N; surface area (A) = 5.0 km^2; depth (Z) = 4.2 m), Morse (MR: 86°2′17.22″W, 40°6′16.84″N; A = 6.0 km^2; Z = 4.7 m) and Geist reservoirs (GR: 85°57′47.22″W, 35°56′16.84″N; A = 7.5 km^2; Z = 3.2 m) in central Indiana (CIN) are the major drinking water sources for over 900,000 residents of the Indianapolis metropolitan region. The long water residence time and a high percentage of agricultural land use (ECR: 60.1%, MR: 76.9%, and GR: 60.5%) in the watersheds (Li et al., 2010) contribute to the high nitrogen and phosphorus loading of these reservoirs with a mean total phosphorus and total nitrogen concentration of 94 μg P/L and 1.47 mg N/L, respectively. The occurrence of nuisance algal blooms impairs the water quality in three reservoirs every summer (Tedesco and Clercin, 2011). From June 2007 to November 2010, five, six, and seven field surveys were conducted across the MR, GR, and ECR, respectively, and overall, 18 field surveys were conducted (see supplementary Table 1).

The dataset for South Australia (SA) was collected from three study sites. Myponga Reservoir (MPR) is an inland water body (138°26′13.29″E, 35°24′10.02″S; A = 2.8 km^2; Z = 21.5 m) located approximately 60 km south of Adelaide, which provides approximately 5% of the potable water for Adelaide. Murray River at Mannum (MRM) is situated on the broad reaches of the Lower Murray channel (139°19′33.22″E, 34°44′55.11″S; Z = 5.6 m), where a main pumping station is sited to supply water to Adelaide. The Murray River at Wellington (MRW) is connected with Alexandrina Lake (139°27′33.11″E, 35°23′55.19″S; Z = 5.3 m), serving as a major water supply source for Adelaide through pumping station. Six field surveys were conducted over three potable water sources from February to March 2009 (see supplementary Table 1).

2.2. In situ data collection

The Secchi disk depth (SDD) was collected at each site to determine water transparency. The following physical parameters were measured at each sampling station using YSI 600XLM-SV multi-parameter probes (YSI, Yellow Springs, Ohio, USA) positioned 0.5 m below the water surface: temperature (°C), turbidity (NTU) and pH. The coordinates were recorded at each station using a GPS unit. Surface water grab samples were collected at each station at approximately 0.5 m below the water surface for laboratory analysis.