Ecological Indicators 67 (2016) 12–20

Contents lists available at ScienceDirect

Ecological Indicators

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

Successive projections algorithm-based three-band vegetation index for foliar phosphorus estimation

Junjie Wang[a,b], Tiezhu Shi[a,b], Huizeng Liu[a], Guofeng Wu[a,b,*]

[a] Key Laboratory for Geo-Environmental Monitoring of Coastal Zone of the National Administration of Surveying, Mapping and Geo-information & Shenzhen Key Laboratory of Spatial-temporal Smart Sensing and Services, Shenzhen University, 518060 Shenzhen, China
[b] College of Life and Marine Sciences, Shenzhen University, 518060 Shenzhen, China

ARTICLE INFO

Article history:
Received 2 October 2015
Received in revised form 6 February 2016
Accepted 10 February 2016
Available online 25 April 2016

Keywords:
Genetic algorithm
Hyperspectral remote sensing
Foliar phosphorus
Successive projections algorithm
Three-band vegetation index

ABSTRACT

Phosphorus (P) is essential for plant growth and development. Very few studies have reported the use of hyperspectral three-band vegetation indices (TBVIs) in foliar P estimation. Further, the optimal TBVI is generally chosen from millions of all possible band combinations. This study aimed to investigate resampling and two wavelength selection methods (genetic algorithm (GA) and successive projections algorithm (SPA)) in deriving TBVIs for foliar P estimation and further to compare the performances of the newly developed TBVIs and published VIs. A total of 137 field-based canopy hyperspectral reflectance (350–2500 nm) of *Carex* (*C. cinerascens*) were obtained and reduced to 1603 wavelengths due to spectral noises. Considering both the original and first derivative reflectance spectra, their resampled wavelengths and selected wavelengths by GA and SPA were employed to derive TBVIs. A total of 24 selected TBVI models were calibrated for foliar P estimation with the training dataset, and they were independently validated with the test dataset. The root mean square error of validation ($RMSE_{Val}$), determination coefficient of validation (R^2_{Val}) and residual prediction deviation (RPD) values were calculated to evaluate the performance of each model. The results demonstrated that 5474, 1972 and 1.2 s in average was taken in calculating all possible TBVIs using resampling, GA and SPA, respectively. Two SPA-based TBVIs, i.e. $(\rho_{760}-\rho_{2387})/(\rho_{723}-\rho_{2387})$ (ρ_λ, original reflectance) and $(\rho'_{729}-\rho'_{1319}+2\rho'_{714})/(\rho'_{729}+\rho'_{1319}-2\rho'_{714})$ (ρ'_λ, first derivative reflectance), had the best model performances ($R^2_{Val}=0.680$, $RMSE_{Val}=0.040\%$, $RPD=1.75$; $R^2_{Val}=0.692$, $RMSE_{Val}=0.039\%$, $RPD=1.80$) in foliar P estimation among the 24 TBVIs. Compared with 15 published VIs ($R^2_{Val}<0.64$, $RPD<1.64$), the two SPA-based TBVIs exhibited better validation performances. We concluded that SPA has the great potential for TBVI derivation due to the reduction of computation time, and the use of SPA in TBVI derivation is recommended for NDVI derivation or other biochemical parameter estimation.

1. Introduction

Phosphorus (P) is a key nutrient element for plant growth and development. In cropland ecosystem, P-deficient crops generally exhibit weak or stunted growth (Li et al., 2006; Plénet et al., 2000), thereby decreasing crop yield and economic benefit (Rodríguez et al., 1999); while in grass ecosystem, foliar P content closely associates with the distribution and feeding patterns of herbivores (Bailey et al., 1996; McNaughton, 1988; Mutanga and Kumar, 2007). Therefore, the regional spatial monitoring of P content in plant leaves is crucial for precision agriculture and wildlife conservation.

Hyperspectral remote sensing and imaging spectroscopy techniques have been proven to be promising in capturing the regional spatial variation of biochemical components in plant leaves, e.g. chlorophyll, nitrogen (N) and P (Axelsson et al., 2013; Ferwerda et al., 2005; Malenovsky et al., 2006; Mutanga et al., 2004). This feasibility is attributed to the relationship between the absorption of electromagnetic radiation and chemical compositions (Curran, 1989). Despite the fact that N and P play an equally important role in plant growth and development, much less attention has been paid to remotely estimation of P compared to that of N (Axelsson et al., 2013; Mutanga and Kumar, 2007). For green plants, foliar N content can be successfully estimated with visible and near-infrared reflectance (VNIR) through its close correlation with optically active compounds, i.e. chlorophyll and protein (Abdel-Rahman et al.,

* Corresponding author at: College of Life and Marine Sciences, Shenzhen University, 518060 Shenzhen, China. Tel.: +86 755 26933149.
E-mail address: guofeng.wu@szu.edu.cn (G. Wu).

http://dx.doi.org/10.1016/j.ecolind.2016.02.033
1470-160X/

2010; Axelsson et al., 2013; Ferwerda et al., 2005; Mitchell et al., 2012). A relationship between P and VNIR is theoretically expected, because P plays an important role in plant photosynthesis and interacts with organic compounds (e.g. starch, sugar, protein and chlorophyll) that absorb electromagnetic radiation at fundamental stretching frequencies in the VNIR domain (Alabbas et al., 1974; Sanches et al., 2013). However, the low concentration in the foliage and subtle physical absorption features still pose a challenge for the accurate estimation of P by VNIR. Given this limitation, there is a need to develop or improve techniques that can accurately estimate foliar P content in grass or cropland ecosystems.

Multivariate modeling techniques, such as partial least squares regression (PLSR) (Bogrekci and Lee, 2005; Sanches et al., 2013), stepwise multiple linear regression (SMLR) (Mutanga et al., 2005; Ramoelo et al., 2011), successive projections algorithm coupled with multiple linear regression (SPA-MLR) (Cui et al., 2013; Wang et al., 2015), artificial neural network (ANN) (Knox et al., 2011; Mutanga and Kumar, 2007) and support vector regression (SVR) (Axelsson et al., 2013; Zhai et al., 2013), have been employed to extract foliar P content information from hyperspectral reflectance at leaf, canopy or landscape scale. These techniques bring multiple wavebands into calibration models, rendering model interpretation more complicated than hyperspectral vegetation index (VI) models with only one predictor. Hyperspectral VIs are developed to enhance their sensitivities to green vegetation signals (Kawamura et al., 2011) and to minimize the variability caused by external factors such as soil background, canopy geometry, leaf optical properties and atmospheric conditions (Darvishzadeh et al., 2009; Jackson and Huete, 1991; Mutanga and Skidmore, 2004). Hyperspectral VIs have been widely used for estimating and mapping foliar N, water and chlorophyll contents (Abdel-Rahman et al., 2010; Ferwerda et al., 2005; Li et al., 2015; Malenovsky et al., 2006; Pacheco-Labrador et al., 2014; Ramoelo et al., 2012). However, very little experience has been gained regarding foliar P estimation with hyperspectral VIs (Kawamura et al., 2011; Pimstein et al., 2011). Hence, it is required to further investigate the relationship between foliar P content and hyperspectral VIs.

The most common hyperspectral VIs are ratio-based indices derived from two narrow wavelengths, i.e. simple ratio index (SRI) and normalized difference vegetation index (NDVI). These VIs are generally developed and optimized through all possible two-paired band combinations involving hundreds or thousands of wavelengths over 400–2400 nm (Darvishzadeh et al., 2009; Ferwerda et al., 2005). Recent studies have demonstrated that three-band VIs (TBVIs) might be better estimators of foliar N and chlorophyll contents compared with traditional two-band VIs (Li et al., 2015; Pacheco-Labrador et al., 2014; Tian et al., 2011; Wang et al., 2012). In general, a total of n wavelengths can generate $n \times (n-1)$ and $n \times (n-1) \times (n-2)$ combinations of all possible two-band VIs and TBVIs, respectively. Therefore, the process of selecting an optimal TBVI is often time-consuming. To overcome this limitation, some studies turned to employ spectral resampling method with an interval of 5 nm to reduce time cost (Pacheco-Labrador et al., 2014), but this method may still take long time in selecting an optimal TBVI from millions of all possible combinations. Further, due to the discard of potential wavelengths related to biochemical concentration and the multicollinearity of neighboring wavelengths, the resampling method may bring irrelevant and redundant spectral information in TBVI derivation. Hence, there is a need to select informative wavelengths related to target biochemical component and reduce time cost for TBVI derivation.

Wavelength selection methods, such as genetic algorithm (GA) (Broadhurst et al., 1997; Jarvis and Goodacre, 2005) and successive projections algorithm (SPA) (Araújo et al., 2001), have been widely used for informative wavelength extraction (Cui et al., 2013; Li et al., 2007). Compared with GA, SPA has weaker performance in the calibration process, whereas SPA is simpler and time-saving in the process of wavelength selection (Shi et al., 2014). The selected wavelengths are often combined with multivariate modeling techniques (e.g. PLSR and MLR) for estimating biochemical concentrations (Cui et al., 2013; Li et al., 2007). However, few efforts have been made for further TBVI derivation with the selected wavelengths.

With the canopy hyperspectral reflectance of *Carex* (*C. cinerascens*) in Poyang Lake, China, this study aimed to (1) investigate the resampling method and wavelength selection methods (GA and SPA) in deriving TBVIs for foliar P estimation and (2) compare the performances of the newly developed TBVIs and some published VIs.

2. Materials and methods

2.1. Field sampling and canopy reflectance measurement

The study area was located in Poyang Lake (28°52′21″–29°06′46″ N, 116°10′24″–116°23′50″ E), Jiangxi Province, China. As the largest freshwater lake in China, Poyang Lake is an important wetland in the world. *Carex* (*C. cinerascens*) is a wetland grass species widely distributed in Poyang Lake, and it is the main food of some over-wintering birds such as swan goose (*Anser cygnoides*) and white-fronted goose (*A. albifrons albifrons*) (Zhang and Lu, 1999).

In order to widen the range of foliar P contents, two filed surveys were carried out in December 2012 (vegetative stage, $n = 66$) and April 2013 (heading stage, $n = 71$), respectively. In each field survey, nine sites (150 m × 150 m) were randomly selected within the large areas of *Carex* biomes. At each site, four to eight plots (1 m × 1 m) were randomly laid out to keep at least 30 m apart between any two plots, and the canopy reflectance spectra and leaf samples were measured and collected as follows: the longitude and latitude coordinates at each plot were first obtained with a global position system receiver (Garmin Ltd., Lenexa, KS, USA); before each spectral measurement of canopy, the spectral measurement of a white Spectralon panel was then carried out for instrument calibration to minimize the effect of changes of solar irradiance and atmospheric conditions on canopy reflectance; ten successive spectra (350–2500 nm) were measured 1 m above the canopy at nadir position using an ASD FieldSpec® 3 portable spectroradiometer (Analytical Spectral Devices, Inc., Boulder, CO, USA) with a field of view of 10°, and their average value was calculated as the final spectrum for each sample; finally, for each plot, five subplots (0.25 m × 0.25 m) in the center and four corners were harvested by clipping leaves 5 cm above the ground and pooled, and one third of the pooled fresh leaves were immediately put into a labeled sample bag for their chemical analysis in laboratory.

2.2. Chemical analysis

The collected leaf samples were dried at 70 °C for 24 h in an oven, ground with an agate mortar and passed through a 65-mesh sieve (0.25 mm). The dried and ground samples were initially pre-processed by $HCLO_4$–H_2SO_4 digestion. Following digestion, the P content (% DM, dry matter) was determined using the Mo–Sb (molybdenum–antimony) colorimetric method (Yuan and Lavkulich, 1995). To ensure measurement accuracy, certified house reference materials and reagent blanks were used during chemical analysis.

2.3. Data pre-processing

Due to the large noises at both spectrum edges (<400 nm and >2450 nm) and over the water-dominated spectral regions