Europ. J. Agronomy 84 (2017) 58–66

Contents lists available at ScienceDirect

European Journal of Agronomy

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

Hyper-spectral estimation of wheat biomass after alleviating of soil effects on spectra by non-negative matrix factorization

Yanli Li[a,b], Ya Liu[a], Shiwen Wu[a,b], Changkun Wang[a], Aiai Xu[a,b], Xianzhang Pan[a,*]

[a] *State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing, Jiangsu 210008, PR China*
[b] *University of Chinese Academy of Science, Beijing 100049, PR China*

ARTICLE INFO

Article history:
Received 22 March 2016
Received in revised form 15 October 2016
Accepted 2 December 2016

Keywords:
Non-negative matrix factorization
Spectral separation
Biomass
Partial least squares regression
Winter wheat

ABSTRACT

Hyper-spectral technology has been proven to be an effective method for the fast and non-destructive monitoring of crop biomass. However, the biomass estimation accuracy of this method is limited due to the effects of background factors, such as soils and water. In this study, a spectral separation method, non-negative matrix factorization (NMF), was proposed to alleviate the effects of soil on spectra. With the application of the NMF method, pure vegetation spectra were extracted from the field-observed spectra of wheat canopy, which were collected in four growing seasons from the tillering to the booting stages of wheat. Then, prediction models of wheat biomass (WB) were established and validated using the extracted spectra with the partial least squares regression (PLSR) method. The results showed that the NMF method could effectively separate the vegetation spectra from the mixed canopy spectra. Based on the extracted vegetation spectra, the WB prediction accuracy could be greatly improved with an increase of 31.7% for the $R^2{}_p$ and an increase of 46.6% for the ratio of performance to deviation (RPD) as compared to the original spectra, indicating that the NMF method could significantly improve the performance of the WB prediction model. This method has potential application in the estimation of biomass using remote sensing technology.

1. Introduction

Biomass is an important indicator of crop conditions and soil fertility, and is the primary factor used in crop monitoring and yield estimation (Bai et al., 2007; Clevers et al., 2007; Feng et al., 2015). However, the traditional field-based destructive means of measuring biomass are painstaking and time consuming. With the emergence of hyperspectral technology in the past few decades, crop biomass can be estimated by a portable, airborne or satellite spectrometer non-destructively (Li et al., 2015; Ramoelo et al., 2015; Prabhakara et al., 2015; Rasmussen et al., 2016).

However, the application of hyperspectral techniques in monitoring crop biomass is relatively restricted due to influence factors, such as soil, water, vegetal fragments, leaf angle, view angle, sun zenith angle and shade (Gamon et al., 1992; Verhoef and Bach, 2007; Wang et al., 2012; Wang et al., 2013a; Schittenhelm et al., 2014; Stadler et al., 2015; Van Beek et al., 2015). Among those influence factors, soil was identified as one of the most important ones when using spectral data for biomass estimation (Huete, 1988; Wessman, 1991; Guerif and Duke, 2000), especially soil surface exposure before canopy closure. A number of approaches that address the effect of the soil factor on biomass prediction have been demonstrated. Generally, the approaches for biomass prediction by hyperspectral techniques can be summarized as two types: the physically based models and the empirical-statistical models (Thenkabail et al., 2016).

The important physically based models for estimating vegetation parameters were the vegetation radiative transfer models, such as SAIL (Verhoef, 1984), PROSPECT (Jacquemoud and Baret, 1990), and GeoSAIL (Verhoef and Bach, 2003), which were established using the fundamental theories of biology, mathematics and physics, and could obtain biochemical parameters, such as biomass and LAI (leaf area index), by simulating the radiation transfer and interaction of light in vegetation. Based on those models, some more methods were proposed by considering the effects of soil factors. Baret et al. (1992) first proposed the PROSAIL model. Next, Verhoef and Bach (2007) proposed the 4SAIL2 model, which

Abbreviations: BRDF, bidirectional reflectance distribution function; BSS, blind source separation; GeoSAIL, SAIL canopy model adding second layer to mimic vertical leaf color gradient; ICA, independent component analysis; LSU, linear spectral unmixing; NMF, non-negative matrix factorization; PLSR, partial least squares regression; PROSPECT, (leaf optical) properties spectra (leaf model); RSU, residual spectral unmixing; SAIL, scattering by arbitrarily inclined leaves; SAILH, SAIL canopy model including Hot spot; PROSAIL, coupling of PROSPECT and SAILH; WB, wheat biomass.

* Corresponding author at: No. 71 East Beijing Road, Institute of Soil Science, Chinese Academy of Sciences, PR China.
E-mail address: panxz@issas.ac.cn (X. Pan).

http://dx.doi.org/10.1016/j.eja.2016.12.003
1161-0301/

was an extension of GeoSAIL, and was additionally combined with PROSPECT and a soil BRDF model (Hapke, 1981; Hapke and Wells, 1981). These approaches could effectively estimate the vegetation parameters with better stability and portability while considering the soil perturbation. However, when applying these models, it was necessary to know the various parameters of the radiative transfer model, which varied greatly from place to place over a large domain (Guerif and Duke, 2000), and were difficult to obtain in practice (Jacquemoud et al., 2009).

The empirical-statistical approach used regression analysis to calculate and predict biomass, with easily measured factors as the independent variables and the biomass as the dependent variable. This approach was widely applied for its appealing simplicity, the easily obtained parameters, and the ability of monitoring vegetation biomass continuously in regions (Li et al., 2015). Recently, researchers had made some progress in avoiding soil effects. Thenkabail et al. (2000) found that three wavelength bands, 500–550 nm, 650–700 nm and 900–940 nm, were the most sensitive bands for biomass estimation. A high correlation was observed between the absorption in the red band (550–750 nm) and biomass, which could be used for biomass estimation (Fu et al., 2014). In addition, the vegetation indices, such as the ratio vegetation index (RVI) (Pearson and Miller, 1972), the normalized difference vegetation index (NDVI) (Rouse et al., 1974) and the soil-adjusted vegetation index (SAVI) (Huete, 1988), were established by performing mathematical transformations of the relevant spectra with the purposes of enhancing vegetation information and diminishing non-vegetation information.

Moreover, the technique of spectral unmixing has been proposed and has been used for the separation of soil-vegetation spectra. Huete (1986) used a factor analysis for the separation of spectral mixtures with soil and vegetation spectra and applied the method for the prediction of biomass. Roberts et al. (1998) applied a simple linear spectral unmixing (LSU) model for the separation of soil and vegetation. Additionally, a new method of RSU was proposed to extract soil spectra from the mixed spectra by Bartholomeus et al. (2011). However, the application of the RSU method was subjected to certain restrictions in use, because the proportions and the spectral information of soil and vegetation should be known in advance, both of which are reality difficult to obtain. Ouerghemmi et al. (2011) applied an ICA algorithm for the separation of soil-vegetation spectra. However, the biggest disadvantages of the ICA were that the sources must be independent of each other and the results might be negative. Compared with the ICA, NMF was another separation algorithm of BSS, which did not necessarily require independent sources and the results were all non-negative. Liu et al. (2015) applied the NMF method to extract soil spectra from field-observed mixed spectra, and compared the prediction accuracy of NMF with ICA and RSU. The NMF method was found to be the best among the three methods. However, whether it is applicable for extracting the spectra of vegetation from the mixed spectra to improve the prediction accuracy of biomass should be further studied.

Therefore, the objectives of this study were 1) to apply the NMF method to alleviate the effects of soil on vegetation spectral extraction; and 2) to determine whether it could improve the accuracy of the WB estimation based on the extracted spectra.

2. Materials and methods

2.1. Experimental site description

The experiment was carried out in an experimental field of Dongtai Institute of Tidal Flat Research, Nanjing Branch of the Chinese Academy of Sciences (32°38′–32°40′ N, 120°52′–120°54′ E), located in the southeast of Dongtai City, Jiangsu Province, China. The area is located in the northern subtropical marine monsoon climate, with four distinct seasons, abundant sunshine and mean annual temperature of 14.6 °C. The soil type for the field experiment site is a silt loam according to the USDA texture classification, with a parent material of marine sediment, containing 4.01 g kg^{-1} of organic carbon, 0.4 g kg^{-1} of total nitrogen, 1.5 g kg^{-1} of total phosphorus and 30 g kg^{-1} of total potassium content.

2.2. Experimental design and data collection

The experimental area was divided into 30 plots, each with size of 1.5 m × 1.5 m and intervals of 30 cm. We added a different amount of salt to each plot, to obtain different soil salt contents for acquiring varying degrees of vegetation coverage. Next, winter wheat was seeded in October 2014 using conventional tillage methods. The field management followed the practice of local farmlands. We obtained the canopy spectra, pictures and aboveground biomass of each plot at the growth stages of tillering, reviving, jointing and booting. We collected data from 20 plots during the first stage, and from 30 plots during the other three stages for a total of 110 groups of data. The descriptive statistics of the biomass and vegetation coverage at the four growth stages were given in Table 1.

The spectral reflectance of each plot was measured under clear skies between 10:00 and 14:00 with an ASD spectrometer (Fieldspec 3 Hi-Res, PANalytical, Boulder, CO, USA), which covered the wavelengths of 350–2500 nm with a 1.4 nm sampling interval between 350 and 1000 nm, and a 2 nm sampling interval between 1000 and 2500 nm. The spectrometer was placed vertically approximately 1 m above wheat canopy, with a 25° field of view (FOV). A white spectral reference panel (1 × 1 m; Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences) was used to convert the spectral radiance measurements to reflectance before each measurement, and ten spectra were taken for each plot. Then, digital photos of each plot were taken by a digital camera (Canon PowerShot G10, Canon Inc., Tokyo, Japan) positioned 1 m above the wheat canopy for the vegetation coverage calculation.

After the collection of spectra and photos, the aboveground biomass was sampled by cutting the biomass in a 20 cm × 20 cm area corresponding to the position of spectrometry of each plot, and put it in a sealed plastic bag. All fresh plant samples were fixed at 105 °C in the laboratory, and oven dried at 65 °C to a constant weight of dry matter. Finally, the aboveground biomass per unit area of winter wheat was calculated.

2.3. Spectral data preparation

The average spectrum of the ten spectra taken in each plot was identified as the spectrum of the plot. In this study, we took the spectra in the wavelength region of 400–1800 nm as the research object because a high correlation between the reflectance and biomass has been found in this range (Gnyp et al., 2014). Because the reflectance data in the region of 1351–1450 nm was easily affected by moisture in the air, these data were not used in the subsequent analysis. To eliminate the noise in the spectrum, the Savitzky-Golay smoothing convolution method was used, with a moving window width of 9.

2.4. Vegetation coverage estimation

The vegetation coverage was the percentage of the vertical projection area in the unit area occupied by aboveground vegetation, which was extracted from digital photos using the G-R thresholding method in this study (Rorie et al., 2011; Wang et al., 2013b). Setting a threshold, pixels with the green channel minus red channel values higher than the threshold were classified as vegetation, with the