International Journal of Applied Earth Observation and Geoinformation xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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

Estimating chlorophyll with thermal and broadband multispectral high resolution imagery from an unmanned aerial system using relevance vector machines for precision agriculture

Manal Elarab[a,b,*], Andres M Ticlavilca[b], Alfonso F. Torres-Rua[b], Inga Maslova[c], Mac McKee[a,b]

[a] Civil and Environmental Engineering Department, Utah State University, USA
[b] Utah Water Research Laboratory, Utah State University, USA
[c] American University, Washington D.C., USA

ARTICLE INFO

Article history:
Received 31 October 2014
Accepted 31 March 2015
Available online xxx

Keywords:
Remote sensing
High spatial resolution imagery
Relevance vector machine
Precision agriculture
Chlorophyll concentration

ABSTRACT

Precision agriculture requires high-resolution information to enable greater precision in the management of inputs to production. Actionable information about crop and field status must be acquired at high spatial resolution and at a temporal frequency appropriate for timely responses. In this study, high spatial resolution imagery was obtained through the use of a small, unmanned aerial system called AggieAir™. Simultaneously with the AggieAir flights, intensive ground sampling for plant chlorophyll was conducted at precisely determined locations. This study reports the application of a relevance vector machine coupled with cross validation and backward elimination to a dataset composed of reflectance from high-resolution multi-spectral imagery (VIS–NIR), thermal infrared imagery, and vegetative indices, in conjunction with in situ SPAD measurements from which chlorophyll concentrations were derived, to estimate chlorophyll concentration from remotely sensed data at 15-cm resolution. The results indicate that a relevance vector machine with a thin plate spline kernel type and kernel width of 5.4, having LAI, NDVI, thermal and red bands as the selected set of inputs, can be used to spatially estimate chlorophyll concentration with a root-mean-squared-error of 5.31 $\mu g\,cm^{-2}$, efficiency of 0.76, and 9 relevance vectors.

1. Introduction

Increasing world population levels will bring increased demand for food, water, and agricultural inputs. Various agricultural farming strategies are being reevaluated to determine how to improve food production, minimize environmental impact, and reduce costs. Among many, Precision Agriculture (PA) has evolved as a viable system to improve profitability and productivity (Swinton and Lowenberg-DeBoer, 1998; Lambert and Lowenberg-De Boer, 2000; Daberkow et al., 2000). PA is a process of finely adjusting agricultural inputs (e.g., water, nutrients) and in-field practices (e.g., irrigation, fertilization), through the use of site-specific information and spatial imagery, to improve measures of agricultural productivity (e.g., yield, net farm income) (Pierce and Nowak, 1999).

Use of spatial imagery in agriculture has been the focus of many studies for the past five decades (MacDonald and Hall, 1980; Bauer, 1985; Idso et al., 1977; Benedetti and Rossini, 1993; Shanahan et al., 2001; Stone et al., 1996; Mathur and Foody, 2008; Franke and Menz, 2007), requiring increased investments in relevant research and technologies (Schellberg et al., 2008) that indicate that remote sensing can be a valuable tool to enhance precision agriculture (Lamb and Brown, 2001; Haboudane et al., 2002; Seelan et al., 2003). However, remote sensing has yet to reach its full capability in PA applications. Lack of fine spatial resolution and near real-time data, compounded by high costs, has hindered remote sensing applications at the field scale (Brisco et al., 1998; Liaghat and Balasundram, 2010; Moran et al., 1997; Kalluri et al., 2002). Thirty years ago, Jackson (Jackson, 1984) envisioned an autonomous remote sensing platform that could overcome most of the limitations; this is becoming a reality with the introduction of affordable unmanned aerial systems (UAS). UAS, a potential substitute for satellite-based remote sensing, are gaining attention and recognition in the scientific community as a potential technology that can generate

* Corresponding author at: Utah Water Research Laboratory, Utah State University, USA. Tel.: +1 435 754 9299; fax: +1 435 797 3663.
E-mail address: alarabmanal@gmail.com (M. Elarab).

http://dx.doi.org/10.1016/j.jag.2015.03.017

high spatial resolution imagery (<1 m) and at a temporal frequency appropriate for timely responses in the production of actionable information about crop and field status. One such UAS, named AggieAir^{TM}, was developed by the Utah Water Research Laboratory (UWRL) at Utah State University. AggieAir is designed to carry camera payloads to acquire high resolution, georeferenced aerial imagery to be used in various water, natural resources, and agricultural applications, including PA. AggieAir holds three sensors: sensors one and two are consumer-grade cameras (personal point-and-click cameras) that capture imagery, depending on flight elevation above ground, of 6–25 cm resolution in the visible (red, green, blue spectrum) and near – infrared spectrum, respectively; sensor three is a microbolometer thermal camera that captures images of 30–150 cm resolution in the thermal infrared spectrum. The three sensors are ideal because of their small size, light weight, low-cost, and high resolution. The use of high-resolution imagery (<1 m) can potentially improve the ability to evaluate the spatial dynamics of chlorophyll and detect its temporal variation. In this study, the use of multispectral VIS-NIR-thermal high-resolution imagery is investigated as a tool to estimate plant chlorophyll concentration to provide time-critical information for PA.

Chlorophyll concentration, measured in mass per unit leaf area ($\mu g\ cm^{-2}$), is an important biophysical parameters retrievable from reflectance data. Chlorophyll is a vital pigment primarily responsible for harvesting light energy used in photosynthesis (Sims and Gamon, 2002; Evans, 1989; Niinemets and Tenhunen, 1997) and is therefore an excellent indicator of a crop's overall physiological status (Evans, 1989; Yoder and Pettigrew-Crosby, 1995) stress or disease (Zarco-Tejada et al., 2004; Peñuelas and Filella, 1998; Chaerle and Van Der Straeten, 2000), and yield predictions (Dawson et al., 2003; Gitelson et al., 2006). Chlorophyll can potentially provide an assessment of leaf nitrogen, an essential plant nutrient, due to the close relationship between leaf chlorophyll and leaf nitrogen (Daughtry et al., 2000; Moran et al., 2000; Wood et al., 1992). Chlorophyll concentration varies with vegetation growth, thus estimating chlorophyll across the field at different growth stages could offer the farmer time- and location-specific critical information ideal for assisting decision makers in monitoring their crops and managing farming activities to achieve maximum production.

Several leaf scale studies have focused on estimating chlorophyll concentration from VIS–NIR reflectance data. These studies indicate that the green and far-red regions of the visible spectrum are sensitive to variations in chlorophyll concentrations (Kim, 1994; Datt, 1999; Gitelson and Merzlyak, 1994; Zarco-Tejada et al., 2001; Demarez and Gastellu-Etchegorry, 2000). Various successful indices have been formulated to estimate chlorophyll concentration (Bonge and Leblanc, 2001; Le Maire et al., 2004; Haboudane et al., 2002). Some of these indices are ratios of reflectance in individual narrow visible wavebands (Blackburn, 1998; Carter and Spiering, 2002) or ratios of reflectance in VIS and NIR (Gitelson et al., 1996), while others are red edge reflectance ratio indices (Gitelson and Merzlyak, 1994; Zarco-Tejada and Miller, 1999; Kim et al., 1994) or first and second derivatives of reflectance spectra (Miller et al., 1990). Composites of indices have been developed (Haboudane et al., 2002) in an attempt to correct for distortions in the reflectance data caused by soil background effect and canopy architecture. Detailed discussions and thorough reviews concerning appropriate optimal wavelengths and various chlorophyll indices can be found in the literature (Haboudane et al., 2004; Bonge and Leblanc, 2001). However, most of the studies have had low spatial and coarse spectral resolution characteristics; therefore, the applicability of those indices to high spatial resolution airborne data cannot be evaluated. Regarding thermal imagery, it was mainly explored when information on plant water status was in question, for example when screening drought tolerance genotypes (Blum et al., 1982), detecting crop water stress levels (Bernie et al., 2009), estimating soil moisture and evapotranspiration (Jackson et al., 1981; Wallace et al., 2012; Hassan Esfahani et al., 2014a). However, TIR data haven't been investigated in estimating chlorophyll yet. Exploring thermal data in this study is rationalized by the close relationship between heat stress and the photosynthetic capacity of the leaves (Raison et al., 1982; Sharkey, 2005) and consequently the chlorophyll concentration. The mechanism by which moderate heat stress reduces photosynthetic capacity has been debated since the eighties where researchers attributed the photosynthesis inhibition to different factors such as the impairment of electron transport activity or the inactivation of Rubisco (Berry and Bjorkman, 1980; Murakami et al., 2000; Weis, 1981; Salvucci and Crafts-Brandner, 2004).

Estimating chlorophyll at a canopy level from optical remotely sensed data can generally be carried out by several methodologies. The simplest methodology that is widely accepted is the empirical method, such as those based on vegetation indices (Johnson et al., 1994). Nevertheless, indices generated in this context are inclined to unstable performance when applied to images that differ from the designed method (Verrelst et al., 2010). Physical behavior based methods are another approach to formulating estimates from remotely sensed data. This method is based on physical laws that describe the transfer and interaction of radiation within the atmospheric column and canopy, such as radiative transfer models (RT) (Myneni et al., 1995). This approach has become more promising with advances in atmospheric radiative transfer modeling. The biggest drawback for such a model is that it requires site-specific information for proper model parameterization, which is not always available. As a result, methods based on vegetation indices or physical models may be either too simple or too complex to deliver accurate estimates (Baret and Buis, 2008). Several books and published papers have reviewed these methodologies and highlighted the advantages and disadvantages associated with the complexity of the modeling approach selected, and the degree of general or local applicability of the methodology in remote sensing (Baret and Buis, 2008; Zarco-Tejada et al., 2001).

Considerable research has been carried out to explore advanced computational methods that are both accurate and robust. Machine learning regression algorithms present a potential approach for generating adaptive, robust, and, once trained, fast estimates (Hastie et al., 2009; Knudby et al., 2010). Recent studies have demonstrated successful performance of a very well-known machine learning algorithm in estimating biophysical parameters using neural network models (Cipollini et al., 2001; De Martino et al., 2002; Verrelst et al., 2012; González Vilas et al., 2011; Hassan Esfahani et al., 2014b). In recent studies, neural networks are being replaced by more advanced regression-based methods that are simpler to calibrate, like support vector machines (SVM) (Moser and Serpico, 2009; Camps-Valls et al., 2006; Pal and Mather, 2005) and relevance vector machines (RVM) (Camps-Valls et al., 2006b). SVMs have been widely used in various remote sensing applications; nevertheless, their large computational complexity is a major drawback. This complexity of SVM models is due to their liberal use of basis functions that typically grow linearly with the size of the training set (Tipping, 2001). Studies have shown that the behavior of relevance vector machines (RVM) is often superior to that of SVMs (Demir and Erturk, 2007). The results given by Tipping (2001) demonstrated that the RVM has a comparable generalization performance to the SVM, while requiring dramatically fewer kernel functions or model terms. RVM, is a statistical learning method proposed by Tipping in 2001 (Tipping, 2001), constitutes a Bayesian approximation for solving nonlinear regression models and is often used for classification and pattern recognition. RVMs offer excellent sparseness characteristics, are robust, and can produce probabilistic outputs that permit the capture of uncertainty in the predictions (Gómez-Chova et al., 2011; Thayananthan et al., 2008).