Computers and Electronics in Agriculture 118 (2015) 225–236

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Proximal hyperspectral sensing and data analysis approaches for field-based plant phenomics

K.R. Thorp [a,*], M.A. Gore [b], P. Andrade-Sanchez [c], A.E. Carmo-Silva [d], S.M. Welch [e], J.W. White [a], A.N. French [a]

[a] USDA-ARS, U.S. Arid Land Agricultural Research Center, 21881 N Cardon Ln, Maricopa, AZ 85138, United States
[b] Cornell University, Plant Breeding and Genetics Section, School of Integrative Plant Science, 310 Bradfield Hall, Ithaca, NY 14853, United States
[c] University of Arizona, Department of Agricultural and Biosystems Engineering, Maricopa Agricultural Center, 37860 W. Smith-Enke Road, Maricopa, AZ 85138, United States
[d] Rothamsted Research, Plant Biology and Crop Science Department, Hertsfordshire AL5 2JQ, United Kingdom
[e] Kansas State University, Department of Agronomy, 2104 Throckmorton Plant Sciences Center, Manhattan, KS 66506, United States

ARTICLE INFO

Article history:
Received 12 February 2015
Received in revised form 18 June 2015
Accepted 6 September 2015
Available online 22 September 2015

Keywords:
Cotton
High performance computing
Inverse modeling
Phenotyping
PROSAIL
Remote sensing

ABSTRACT

Field-based plant phenomics requires robust crop sensing platforms and data analysis tools to successfully identify cultivars that exhibit phenotypes with high agronomic and economic importance. Such efforts will lead to genetic improvements that maintain high crop yield with concomitant tolerance to environmental stresses. The objectives of this study were to investigate proximal hyperspectral sensing with a field spectroradiometer and to compare data analysis approaches for estimating four cotton phenotypes: leaf water content (C_w), specific leaf mass (C_m), leaf chlorophyll $a+b$ content (C_{ab}), and leaf area index (LAI). Field studies tested 25 Pima cotton cultivars grown under well-watered and water-limited conditions in central Arizona from 2010 to 2012. Several vegetation indices, including the normalized difference vegetation index (NDVI), the normalized difference water index (NDWI), and the physiological (or photochemical) reflectance index (PRI) were compared with partial least squares regression (PLSR) approaches to estimate the four phenotypes. Additionally, inversion of the PROSAIL plant canopy reflectance model was investigated to estimate phenotypes based on 3.68 billion PROSAIL simulations on a supercomputer. Phenotypic estimates from each approach were compared with field measurements, and hierarchical linear mixed modeling was used to identify differences in the estimates among the cultivars and water levels. The PLSR approach performed best and estimated C_w, C_m, C_{ab}, and LAI with root mean squared errors (RMSEs) between measured and modeled values of 6.8%, 10.9%, 13.1%, and 18.5%, respectively. Using linear regression with the vegetation indices, no index estimated C_w, C_m, C_{ab}, and LAI with RMSEs better than 9.6%, 16.9%, 14.2%, and 28.8%, respectively. PROSAIL model inversion could estimate C_{ab} and LAI with RMSEs of about 16% and 29%, depending on the objective function. However, the RMSEs for C_w and C_m from PROSAIL model inversion were greater than 30%. Compared to PLSR, advantages to the physically-based PROSAIL model include its ability to simulate the canopy's bidirectional reflectance distribution function (BRDF) and to estimate phenotypes from canopy spectral reflectance without a training data set. All proximal hyperspectral approaches were able to identify differences in phenotypic estimates among the cultivars and irrigation regimes tested during the field studies. Improvements to these proximal hyperspectral sensing approaches could be realized with a high-throughput phenotyping platform able to rapidly collect canopy spectral reflectance data from multiple view angles.

Published by Elsevier B.V.

1. Introduction

To improve food security, adapt to climate change, and reduce resource requirements for crop production, scientists must better understand the connection between a plant's observable characteristics (phenotype) and its genetic makeup (genotype). Unprecedented advances in DNA sequencing have unlocked the genetic code for many important food crops, including rice (*Oryza sativa* L.), sorghum (*Sorghum bicolor* L.), and maize (*Zea mays* L.) (Bolger et al., 2014). However, understanding how genes control complex plant traits, such as drought tolerance, time to anthesis, and harvestable yield, remains challenging. Field-based plant phenomics seeks to implement information technologies, including sensing and computing tools in combination with genetic mapping approaches, to rapidly characterize the physiological responses of

* Corresponding author.
E-mail address: kelly.thorp@ars.usda.gov (K.R. Thorp).

http://dx.doi.org/10.1016/j.compag.2015.09.005
0168-1699/Published by Elsevier B.V.

genetically diverse plant populations in the field and relate these responses to individual genes (Araus and Cairns, 2014; Furbank and Tester, 2011; Houle et al., 2010; Montes et al., 2007; White et al., 2012). When validated, crop improvement strategies based on targeted quantitative trait loci and genomic selection can be used for efficient development of crop cultivars that are both high yielding and resilient to environmental stresses.

A variety of electronic sensors have been deployed for field-based plant phenomics, mainly on ground-based vehicles. Andrade-Sanchez et al. (2014) developed a sensing platform on a high-clearance tractor that collected data over four Pima cotton (*Gossypium barbadense* L.) rows simultaneously. Ultrasonic sensors, infrared radiometers, and active multispectral radiometers were used to measure canopy height, temperature, and reflectance, respectively. Scotford and Miller (2004) mounted passive two-band radiometers and ultrasonic sensors on a tractor boom and used the system to estimate tiller density and leaf area index (LAI) of winter wheat (*Triticum aestivum* L.). Other sensing systems have incorporated passive hyperspectral radiometers (spectroradiometers) for measuring crop canopy spectral reflectance continuously over a range of wavelengths, typically within the visible and near-infrared spectrum. For example, the phenotyping platform of Comar et al. (2012) incorporated four spectroradiometers sensitive between 400 and 1000 nm at 3 nm spectral resolution and two RGB digital cameras. Also, Montes et al. (2011) developed a system with light curtains for canopy profiling and spectroradiometers sensitive between 320 and 1140 nm at 10 nm spectral resolution. Rundquist et al. (2004) compared machine-based versus hand-held deployment of a spectroradiometer and found reduced variability and higher reproducibility of sensor measurements when the instrument was positioned by a machine.

Following sensor platforms, the next challenge for field-based plant phenomics is the development of methodologies to extract meaningful information from the sensor data, with the ultimate goal to quantify specific crop phenotypes. However, the fundamental measurements of many sensors have little utility for crop phenotyping without additional post-processing and analysis. For simple, empirical processing of canopy spectral reflectance data, a multitude of vegetation indices have been developed (Bannari et al., 1995) and used to estimate several crop characteristics, including canopy cover, LAI, and biomass (Wanjura and Hatfield, 1987). The popular normalized difference vegetation index (NDVI) is traditionally calculated as

$$\mathrm{NDVI} = \frac{\rho_2 - \rho_1}{\rho_2 + \rho_1} \tag{1}$$

where ρ_2 is the spectral reflectance in the near-infrared waveband and ρ_1 is the spectral reflectance in the red waveband. However, with the advent of hyperspectral sensors, other narrow-band indices have been developed using the NDVI equation with reflectance data in different wavebands. For example, Gamon et al. (1992) developed the physiological (or photochemical) reflectance index (PRI), a narrow-band index using reflectance at 531 nm to track xanthophyll cycle pigments and estimate photosynthetic efficiency. Likewise, Gao (1996) developed the normalized difference water index (NDWI) to estimate vegetation water content. Many other studies have identified optimum wavebands for a given application by calculating narrow-band NDVI for all possible waveband combinations for a given hyperspectral sensor (Fu et al., 2014; Hansen and Schjoerring, 2003; Thenkabail et al., 2000; Thorp et al., 2004). Babar et al. (2006) demonstrated several narrow-band spectral reflectance indices that explained genetic variability in wheat biomass. Mistele and Schmidhalter (2008) measured spectral reflectance of maize canopies from four view angles and found the spectral reflectance indices were strongly correlated ($0.57 \leqslant r^2 \leqslant 0.91$) with total nitrogen uptake and dry biomass weight. In a study by Gutierrez et al. (2012), spectral reflectance indices explained over 87% and 93% of the variability in biomass and LAI, respectively, for three upland cotton varieties. Seelig et al. (2008) correlated shortwave infrared spectral reflectance indices with relative water content and thickness of peace lily (*Spathiphyllum lynise*) leaves ($r^2 > 0.94$).

Other spectral data analysis approaches consider all the visible, near-infrared, and shortwave infrared wavebands collectively. Statistical procedures such as principal component regression (PCR) and partial least squares regression (PLSR) reduce dimensionality by decomposing the hyperspectral data into a set of independent factors, against which crop biophysical traits are regressed. For example, Thorp et al. (2008) used PCR to estimate maize stand density from aerial hyperspectral imagery ($r^2 = 0.79$). Also, Thorp et al. (2011) used proximal spectral reflectance data with PLSR to estimate dry biomass weight, flower counts, and silique counts of lesquerella (*Lesquerella fendleri*) with root mean squared errors of prediction equal to 2.1 Mg ha^{-1}, 251 flowers, and 1018 siliques, respectively. In another study, PLSR models developed from spectral reflectance of rice canopies explained up to 71% of the variability in plant nitrogen (Bajwa, 2006). Hansen and Schjoerring (2003) compared estimates of wheat biophysical variables using (1) linear regression on narrow-band NDVI with optimal wavebands and (2) PLSR with all wavebands from 400 to 900 nm. The NDVI approach better estimated LAI and chlorophyll concentration, while the PLSR approach better estimated green biomass weight and nitrogen concentration.

Another potential solution for quantifying crop phenotypes involves combining measured spectral reflectance data with physical models of radiative transfer in the plant canopy. Input parameters for such models describe attributes (i.e., phenotypes) of the crop canopy, which are used to simulate canopy spectral reflectance. For example, with 14 input parameters that describe plant characteristics and illumination conditions, the PROSAIL model (Jacquemoud et al., 2009) can simulate plant canopy spectral reflectance from 400 to 2500 nm in 1 nm wavebands. Using model inversion techniques, spectral reflectance measurements from spectroradiometers can be used to estimate PROSAIL input parameters. These estimates represent additional crop phenotypes that could be useful in subsequent genetic analyses. By linking crop phenotypes to sensor data through the theoretical knowledge contained in the simulation model, the approach is less empirical than the vegetation index and PLSR approaches.

Literature provides examples of PROSAIL model inversion for vegetation characterization in diverse environments, but field-based plant phenomics is a novel application. Jacquemoud (1993) first investigated the practical limitations of PROSAIL model inversion using synthetic spectra. A subsequent study tested field spectroradiometer data with PROSAIL model inversion to retrieve sugar beet (*Beta vulgaris*) canopy characteristics, such as chlorophyll $a + b$ concentration, leaf water thickness, LAI, and leaf inclination angle (Jacquemoud et al., 1995). At coarser spatial and spectral scales, Zarco-Tejada et al. (2003) used data from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite to invert PROSAIL for estimation of chaparral vegetation water content in a central California shrub land. Yang and Ling (2004) estimated leaf water thickness of New Guinea impatiens (*Impatiens hawkeri*) in a controlled environment using PROSAIL model inversion from 1300 nm to 2500 nm, but spectral artifacts between 400 and 1300 nm due to artificial lighting prevented the estimation of other plant characteristics. PROSAIL model inversion also provided estimates of LAI and chlorophyll $a + b$ concentration for potato (*Solanum tuberosum* L.) and wheat managed with variable nitrogen fertilization rates (Botha et al., 2007, 2010). Others have linked