Computers and Electronics in Agriculture 152 (2018) 186–197

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Prediction of cotton lint yield from phenology of crop indices using artificial neural networks

Amir Haghverdi[a,*], Robert A. Washington-Allen[b], Brian G. Leib[c]

[a] Environmental Sciences Department, University of California, Riverside, 900 University Ave., Riverside, CA 92521, United States
[b] Department of Agriculture, Nutrition, and Veterinary Science (ANVS), University of Nevada, Reno, Mail Stop 202, Reno, NV 89557, United States
[c] Biosystems Engineering and Soil Science Department, University of Tennessee-Knoxville, 2506 E.J. Chapman Drive, Knoxville, TN 37996, United States

ARTICLE INFO

Keywords:
Machine learning
Phenology
Soil variability
Yield map

ABSTRACT

A primary utility of satellite remote sensing technology is monitoring and assessment of agricultural lands for determining the area, amount, type, and quality of crop production. Since the mid-1970s agricultural scientists have sought to advance this utility through development of precision agriculture (PA) methods and technologies. Consequently, PA has taken advantage of freely available medium-spatial resolution remote sensing technology and instrumented fields to monitor crop biomass, phenology, and yield of crops at the sub-field to larger scales. The main goal of this study was to determine cotton lint yield in a 73-ha irrigated field in western Tennessee using remote sensing technology. We used two growing seasons (2013 and 2014) of Landsat 8 transformed to 8 input predictors including *Red*, near infra-red (*NIR*), the simple ratio (*SR*), normalized difference vegetation index (*NDVI*), green NDVI (*GNDVI*), and the tasselled cap transformation's greenness, wetness, and soil brightness indices: *GI*, *WI,* and *SBI,* respectively, as proxies for cotton lint yield and crop phenology (in this study all input predictors are being referred to as crop indices, CIs). We used artificial neural network (ANN) approach to generate 61,200 models relating individual CIs and CI phenology to field estimates of lint yield to predict and map the field's cotton lint yield in two cropping seasons. The correlation between cotton lint yield and CIs ranged from -0.20 to 0.60 in 2013 and from −0.79 to 0.84 in 2014. The best ANN models were in 2013 (r = 0.68 and the normalized MAE = 11%) and 2014 (r = 0.86 and the normalized MAE = 8%) growing seasons. The *WI* and *GI* were the best CI predictors of cotton lint yield, and overall for the early to mid-season prediction, CI phenologies had better performance than single date CI models. Consequently, we recommend the use of Landsat 8 derived *WI* or *GI* phenology to predict crop yields.

1. Introduction

Precision agriculture (PA) has progressed as a result of emerging innovations in instrumentation and measurements that allow consideration of agricultural practices at multiple spatial (and temporal) scales from farm to subfield to individual plants. The main promise of PA is the development of an understanding of crop growth and yield dynamics in response to spatiotemporal variabilities in climate and the physical environment. This knowledge will help farmers optimize site-specific management decisions. *In situ* application of mid- to high spatial resolution, airborne and/or satellite remote sensing technology has been of particular interest to the PA community, particularly for the investigation of the variability of crop growth and yield (Leon et al., 2003, Guo et al., 2012; Heermann et al., 2002; Vellidis et al., 2004; Wang and Shen, 2015; Yang & Everitt, 2002).

The Landsat series of satellites have been collecting global data since 1972, which means that long term historical archives of multispectral data are freely available to conduct annual and interannual time series analyses of the status of crop phenology over a number of growing seasons (Wulder et al., 2012). The reflectance of crop canopies and the bare soil background changes throughout a cropping season as plants go through different growth stages and soil water status. Proxy indicators of soil properties as well as crop phenology and yield dynamics can be derived from moderate pixel resolution Landsat data.

A variety of vegetation or crop indices (CIs) have been developed from the portions of the electromagnetic spectrum detected by particular sensors on various platforms (e.g., the Operational Land Imager (OLI) on Landsat 8) including the simple ratio (SR, Birth and McVey 1968), the normalized difference vegetation index (NDVI, Rouse et al., 1973), the green NDVI (GNDVI, Gitelson et al., 1996), and the

* Corresponding author.
E-mail address: amirh@ucr.edu (A. Haghverdi).

https://doi.org/10.1016/j.compag.2018.07.021
Received 24 May 2018; Received in revised form 9 July 2018; Accepted 10 July 2018
0168-1699/

"Tasselled Cap" transformation (TCAP, Kauth & Thomas 1976). These CIs are calculated as:

$$SR = \frac{NIR}{R} \tag{1}$$

$$NDVI = \frac{NIR - R}{NIR + R} \tag{2}$$

$$GNDVI = \frac{NIR - G}{NIR + G} \tag{3}$$

where *NIR* is near-infrared reflectance (e.g., Landsat 8's band 5 at 0.851–0.879 μm), *R* is red reflectance (e.g., Landsat 8's band 4 at 0.636–0.673 μm), and *G* is green reflectance (e.g., Landsat 8's band 3 at 0.533–0.590 μm).

Among CIs in general, NDVI is arguably the most widely used index in the world, much less in PA. NDVI and other crop indices have been shown to be effective measurement proxies for various crop characteristics including biomass, net primary productivity, density, plant canopy cover, leaf area, and plant water content (Sellers, 1985; Washington-Allen et al., 2006; West et al., 2010). Most of these indices were developed to take advantage of the consistent difference in interactions (i.e., absorption, emittance, transmission, and reflectance) with crop canopies between red and near-infrared radiation (Clevers, 2014). Other parts of the electromagnetic spectrum have also proven useful for agricultural applications. For instance, leaf emittance in the thermal region of the electromagnetic spectrum has been widely used to quantify evapotranspiration and thus crop response to water stress (Hatfield et al., 2008).

Kauth & Thomas (1976) developed the TCAP to reduce the correlated radiation components of the 4-dimensional structure of Landsat 1 Multispectral Scanner data when tracking the phenology of yellow corn crops. The TCAP is a weighted form of a principle component analysis that was used to reduce the dimensionality or degree of correlation between the bands or channels of a multispectral sensor. This is done through the consolidation of a sensor's highly correlated bands into a set of new bands that are uncorrelated with each other (Baig et al., 2014). For example, Landsat 5's Thematic Mapper sensor detects 7 parts of the electromagnetic spectrum or individual bands or channels with differing levels of correlation between each band. Application of the TCAP produced three new bands and/or indices of interest to this study called the soil brightness index (*SBI*), that detects bare ground; a wetness index (*WI*), that detects water bodies, soil and vegetation moisture content; and a greenness index (*GI*) that detects vegetation cover or biomass (Crist 1985; Crist & Cicone 1984). When these 2–3 bands are visually compared in Cartesian space using either a two-dimensional (2-D) or 3-D scatter plot a "tassell cap" relationship is observed (Kauth & Thomas, 1976). Plotting a 2-D or 3-D time series of the TCAP tracks different crop phenologies and soil moisture dynamics within a growing season (Baig et al., 2014; Kauth & Thomas, 1976).

New sets of sensor specific TCAP weighted coefficients have been developed for each of the sensors in the Landsat series as well as other sensors including RapidEye (Crist, 1985; Schönert et al., 2015). Schönert et al. (2015) reported promising preliminary results for the retrieval of crop biophysical attributes (e.g. leaf area index, plant chlorophyll and nitrogen) at field-level using RapidEye sensor-derived TCAP.

Consequently, these past studies suggested that remote sensing, particularly mid-resolution sensors like Landsat 8, could be used as effective PA tools for tracking spatially and temporally heterogeneous plant and soil moisture dynamics over the course of a few growing seasons, i.e., crop phenology. Thus, these studies demonstrated that an annual time series of remotely-sensed CIs may correlate with variable cotton lint yield (due to the heterogeneous spatial distribution of soil moisture) (Hypothesis 1). Further, a time series of annual CIs will characterize cotton lint phenology and may predict cotton lint yield at the subfield to farm spatial scales of observation (Hypothesis 2).

We examined two key research questions:

(I) Which of the Landsat 8 derived CIs provides the best prediction of cotton lint yield at the subfield level?
(II) Which approach provides the best growing season prediction of cotton lint yield, the use of multiple CI's at a single date or CI crop phenology produced from an annual time series?

The objectives of this study will be implemented over the 2013 and 2014 growing seasons and are to:

(I) Use an annual time series of Landsat 8 derived CIs to characterize cotton phenology.
(II) Determine the utility of these remotely-sensed CIs as input predictors for in-season prediction of cotton lint yield using Artificial Neural Network (ANN) models.

2. Material and methods

2.1. Study area & data collection

We conducted a two-year (2013 and 2014) on-farm irrigation experiment on a 73-ha property in the semi-humid region of west Tennessee to investigate cotton's response to supplemental irrigation management (Fig. 1). Climate records for the area indicate a mean monthly precipitation of 97 mm and a mean temperature of 21 °C throughout the May to November growing season (National Climate Data Center, 2015). The field's elevation ranges from 77 to 80-m and it contains Robinsonville loam and fine sandy loam, Commerce silty clay loam, and Crevasse sandy loam soils that were formed in the loamy and sandy alluvium of Mississippi river terraces (Soil Survey Staff, 2015).

Haghverdi et al. (2015b) described the soil data collection where a total of 400 undisturbed core soil samples were collected from 100 locations at 4 depths on March 21 and 22, 2014. Measures of soil texture, soil water content, and bulk density within the crop effective root zone (i.e. 0–100 cm) were determined from these samples in a laboratory. Additionally, soil apparent electrical conductivity (ECa) was collected in situ by means of a Veris 3100 (Veris Technologies, Salina, KS) instrument. As much as a four-fold variability in plant available water (PAW: the difference between water contents at field capacity and

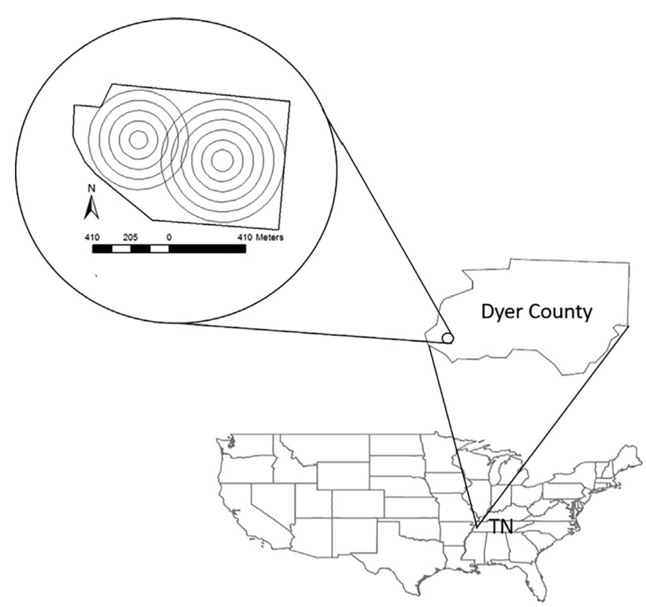

Fig. 1. The 73-ha study field and its 2 center pivots in the Dyer County humid region of west Tennessee.
Adapted from Haghverdi et al. (2016)