Europ. J. Agronomy 78 (2016) 44–59

Contents lists available at ScienceDirect

European Journal of Agronomy

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

Data fusion of spectral, thermal and canopy height parameters for improved yield prediction of drought stressed spring barley

Pablo Rischbeck[a,*], Salah Elsayed[b], Bodo Mistele[c], Gero Barmeier[a], Kurt Heil[a], Urs Schmidhalter[a]

[a] Chair of Plant Nutrition, Technical University of Munich, Emil-Ramann-Straße 2, D-85350 Freising-Weihenstephan, Germany
[b] Evaluation of Natural Resources Department, Environmental Studies and Research Institute, Sadat City University, Egypt
[c] Department of Agricultural Engineering, Fachhochschule Südwestfalen, Soest, Germany

ARTICLE INFO

Article history:
Received 30 October 2015
Received in revised form 5 April 2016
Accepted 27 April 2016
Available online 10 May 2016

Keywords:
Abiotic stress
Data fusion
High throughput phenotyping
Multi-annual
Multiple linear regression
Partial least square
Phenomics
Yield modelling

ABSTRACT

Yield modelling based on visible and near infrared spectral information is extensively used in proximal and remote sensing for yield prediction of crops. Distance and thermal information contain independent information on canopy growth, plant structure and the physiological status. In a four-years′ study hyperspectral, distance and thermal high-throughput measurements were obtained from different sets of drought stressed spring barley cultivars. All possible binary, normalized spectral indices as well as thirteen spectral indices found by others to be related to biomass, tissue chlorophyll content, water status or chlorophyll fluorescence were calculated from hyperspectral data and tested for their correlation with grain yield. Data were analysed by multiple linear regression and partial least square regression models, that were calibrated and cross-validated for yield prediction. Overall partial least square models improved yield prediction ($R^2 = 0.57$; RMSEC = 0.63) compared to multiple linear regression models ($R^2 = 0.46$; RMSEC = 0.74) in the model calibration. In cross-validation, both methods yielded similar results (PLSR: $R^2 = 0.41$, RMSEV = 0.74; MLR: $R^2 = 0.40$, RMSEV = 0.78). The spectral indices R_{780}/R_{550}, R_{760}/R_{730}, R_{780}/R_{700}, the spectral water index R_{900}/R_{970} and laser and ultrasonic distance parameters contributed favourably to grain yield prediction, whereas the thermal based crop water stress index and the red edge inflection point contributed little to the improvement of yield models. Using only more uniform modern cultivars decreased the model performance compared to calibrations done with a set of more diverse cultivars. The partial least square models based on data fusion improved yield prediction ($R^2 = 0.62$; RMSEC = 0.59) compared to the partial least square models based only on hyperspectral data ($R^2 = 0.48$; RMSEC = 0.69) in the model calibration. This improvement was confirmed by cross-validation (data fusion: $R^2 = 0.39$, RMSEV = 0.76; hyperspectral data only: $R^2 = 0.32$, RMSEV = 0.79). Thus, a combination of spectral multiband and distance sensing improved the performance in yield prediction compared to using only hyperspectral sensing.

1. Introduction

In agricultural remote and proximal sensing the visible and near infrared spectral range is used for modelling crop yield (Hansen et al., 2002; Li et al., 2015). A number of yield related plant traits like green biomass (Mistele et al., 2012), canopy water mass (Winterhalter et al., 2011), leaf senescence and chlorophyll content (Kipp et al., 2014) can be assessed spectrally. The spectral reflectance in the water band at 970 nm (Peñuelas et al., 1993) and 1450 nm in addition to various normalised difference indices combining NIR and VIS wavelengths (Rischbeck et al., 2014) can be used to assess plant tissue dehydration. Mistele et al. (2004) tested several spectral indices that were obtained from an oligo view optic device connected to a spectrometer in winter wheat field trials. The NDVI $(R_{780} - R_{670})/(R_{780} + R_{670})$ and NIR/Red (R_{780}/R_{670}) indices were correlated with biomass. The indices Red Edge/NIR (R_{740}/R_{780}), NIR/Green (R_{780}/R_{550}) and NIR/Red (R_{780}/R_{700}) are useful for assessing biomass and the nitrogen status. Unlike the simple ratio or normalised indices that are sensitive to the pigment absorption in the visible range, the red edge inflection point (REIP) indicates the shift of the slope connecting the reflectance in the red and NIR spectral regions (Herrmann et al., 2010). The REIP is

* Corresponding author.
E-mail address: Pablo.rischbeck@wzw.tum.de (P. Rischbeck).

http://dx.doi.org/10.1016/j.eja.2016.04.013
1161-0301/

Fig. 1. PhenoTrac 4, measuring platform equipped with hyperspectral, thermal and laser distance sensors developed by the Chair of Plant Nutrition of the Technical University of Munich.

an indicator of the amount of chlorophyll in the field of view of the sensor. Mistele et al. (2004) found the REIP is useful for assessing both the biomass and the nitrogen status, because both parameters are related to the amount of absorbing pigments.

However, the information that can be obtained from spectrometry is limited to the interactions of light with plant or soil material. Transpiration is not directly related to spectral reflectance. Canopy height is related to passive spectral reflectance by the leaf number and leaf architecture. Sensors using other physical principles might add information to spectral assessments. The use of information from different sensors might also increase the flexibility of yield prediction under different weather conditions and at different growth stages of the crop.

For example, thermometry detects thermal infrared emission in the range of 8–20 μm and is used to measure surface temperature (Jones, 2004). Idso et al. (1981) developed a crop water stress index based on thermal measurements that ranges between 0 (full transpiration) and 1 (minimum stomatal conductance).

Laser distance sensors measure the traveling time of light to and from a reflecting surface. Ultrasonic distance sensors measure the time interval between sending a sound signal and receiving the echo to measure the distance to an object. Both sensors enable the measurement of canopy height and structure (Zhao et al., 2010).

In recent years unmanned aerial vehicles have been equipped with spectrometers (Agüera Vega et al., 2015) and stereo camera systems are used for distance measurements (Honkavaara et al., 2013) for applications in precision agriculture. These systems allow rapid image and data acquisition of neighbouring fields. A number of workgroups have used carrier vehicles equipped with spectral and other sensors to increase the throughput of trait assessments in breeding nurseries and field trials (Schmidhalter et al., 2001; Montes et al., 2011; White et al., 2012; Comar et al., 2012; Rebetzke et al., 2013). The Chair of Plant Nutrition from the Technical University of Munich developed a carrier vehicle (PhenoTrac 4, Fig. 1), equipped with hyperspectral, thermal and distance sensors (Mistele and Schmidhalter, 2010; Erdle et al., 2011; Winterhalter et al., 2013; Kipp et al., 2014). In this study statistical data fusion of sensor measurements was used and tested for improving yield modelling under different levels of drought stress and at different phenological stages.

Under temperate European climatic conditions the interannual precipitation can vary strongly, ranging from sufficient to deficient. In Germany since 2000 the yearly precipitation was below the longterm average (1881–2014) in the years 2003, 2006, 2011, 2012 and 2014 (Deutscher Wetterdienst, 2015). Drought stress due to low water availability in the soil and high potential evapotranspiration affects cereal crops by decreasing their growth and yield. The direct impacts of drought stress are reduced water uptake by roots, decreasing water status, reduced stomatal conductance, reduced CO_2 uptake, reduced transpiration and increased surface temperature (Bradford and Hsiao, 1982; Schmidhalter et al., 1998; Siddique et al., 2000; Wall et al., 2011). Reduced water uptake from dry soils also decreases nitrogen uptake and affects the nutritional status of the crop (Rostamza et al., 2011). For this study rain-out shelters were used to impose mild to severe drought stress on spring barley grown in field trials. Variable growth conditions in cereal production can be tested by including artificially created drought conditions into model calibrations.

Yield models can be adapted to a certain species, genotype and local soil and growth conditions (Laurent et al., 2015). However, statistical yield modelling based on remote sensing is most commonly used on a regional scale using sensors mounted on satellites (Zhao et al., 2015). Regional and multi-field applications require models, that can cope with different genotypes and soils and the spatial and interannual variability of weather and abiotic stresses that affect plant growth. Combining data across phenological stages, field trials and years in broad calibrations might lead to such broadly adapted yield models or improve yield prediction.

There has been progress in developing sophisticated statistical approaches for linking data gathered from sensors with agronomic or physiological traits. Thus, yield modelling approaches can be based on single spectral indices (Gutierrez-Rodriguez et al., 2004). Both linear regression models and multiple linear regression models can also be derived from spectral indices. Royo et al. (2003) found that 17.3% to 65.2% of durum wheat yield could be accounted for with multiple linear regression models based on varying spectral indices. However, multiple linear regression analysis is limited to a few predictor variables in cases of collinearity of the measured variables. Additionally, multiple linear regression analysis also requires homogeneity of variance (homoscedasticity) (Heil and Schmidhalter, 2012).

An alternative approach is to use partial least square regression (PLSR). In PLSR orthogonal components are unaffected by collinearity and are derived from all variables. Partial least square regression models of hyperspectral reflectance were used by Weber et al. (2012) at anthesis and during the milk-grain stage of maize (*Zea mays* L.). The models explained between 49% to 69% of the variation in grain yield in calibration studies and 23% to 40% after model validation. Hyperspectral data has also been used for yield modelling in partial least square regressions (Sharabian et al., 2014). Fu et al. (2013) combined optimal binary, normalized indices retrieved in a spectral range of 350 nm to 2500 nm with band depth ratios in partial least square regression models for assessing winter wheat biomass. Biomass could be predicted with a maximum R^2 of 0.84 and a minimum RMSE of 0.177 kg m^{-2}.

The purpose of this work was: (i) to compare the performance of multiple linear regression and partial least square models for assessing grain yield in multi-annual experiments, (ii) to investigate which traits should be included in data fusion, (iii) to investigate the advantages of data combinations, (iv) to assess improvements in yield prediction by data fusion models, and (v) to test the methods by using more and less variable cultivars.

2. Materials and methods

2.1. Field trials

The field experiments were conducted at the Dürnast research station of the Technical University of Munich in southwestern