APPROPRIATE WAVELENGTHS FOR WINTER WHEAT GROWTH STATUSE BASED ON MULTI SPECTRAL REFLECTANCE DATA

Vali Rasooli Sharabian, Noboru Noguchi, Issei Han-ya, Kazunobu Ishii

Graduate School of Agriculture
Hokkaido University
Sapporo, Japan
E-mail: vrasooli@bpe.agr.hokudai.ac.jp

ABSTRACT

One of the applications of remote sensing in agriculture is to obtain crop status for estimation and management of variable rate of inputs in the crop production. In order to select the appropriate wavelengths related for crop characteristics, the field experiments were conducted in June-July in the farming area of Hokkaido University (43° 4' N 141° 20' E), Japan in two years. To make difference in growth conditions, the field was divided into 8 areas and four level of fertilizer (Ammonium Nitrate) 0, 30, 60, and 90 kg ha^{-1} with two repetitions were applied at the reviving stage (GS 26). Hyper spectral reflectance data using a portable field spectroradiometer (351 to 2500 nm) were recorded from 10 am to 2 pm, under cloudless conditions at four different growth stages of winter wheat in first year and at eight different growth stages in second year. The partial least squares (PLS) regression was used to determine important wavelengths. Some wavelengths in green, amber, red, red edge, near infrared and short wave bands were identified by PLS as significant wavelengths for measured crop reference data. All two-band combinations of the several vegetation indices were subsequently calculated in an algorithm for determining linear regression analysis against SPAD value, protein content, and grain yield. R square matrix used in ArcMap to make contour plot. Using overlaying in analysis tools between first and second year results, a number of grouped wavebands with high correlation were revealed. The results from the calibration models built by PLS showed strong relationships between predicted and measured SPAD value, protein concentration and Yield.

Keywords: remote sensing; multi spectral reflectance; winter wheat; PLS;

INTRODUCTION

Measurement of various crop canopy variables during the growing season provides an opportunity for improving grain yields and quality by site-specific application of fertilizers (Hansen et al, 2003). Plant reflectance is affected by leaf surface properties, internal structure, plant stress, and the concentration and distribution of biochemical components; therefore, analysis of remote reflected light may be used to assess plant biomass and the physiological status of a plant (Penuelas and Filella 1998). Wavelengths in the red and near infrared (NIR) wavebands are frequently used for indirect measurements of plant characteristics (Wood et al. 2003). In the study by Wainjura and Hatfield (1987), a vegetation index ratio (NIR/red) was most sensitive to high crop biomass production in corn (Zea Mays L.) and soybean (Glycine max L.), but during early vegetative growth (e.g., Zadoks growth stage (GS) 25 in winter wheat), (Zadoks et al. 1974) a normalized difference vegetation index (NDVI), Eq. 1, has been shown to provide more accurate estimates of biomass (Raun et al. 2001). Spectral vegetation indices such as NDVI are useful for obtaining crop information indirectly, such as photosynthetic efficiency, productivity potential, and potential yield (Raun et al. 2001).

$$\mathrm{NDVI} = (\mathrm{NIR} - \mathrm{Red}) / (\mathrm{NIR} + \mathrm{Red}) \qquad (1)$$

NDVI is a broadband index that is well correlated to leaf area index and green biomass, and thus it is sensitive to photosynthetic efficiency (Aparicio et al. 2002). Remote sensing, using passive sensor systems (imagery or spectral radiometers), has long been advocated as a way to characterize spatial variability in fields (Bhatti et al. 1991). Passive sensors detect the canopy radiance (reflected radiation) of natural sunlight with a down-facing sensor. Using a four-band passive sensor system (blue, green, red and NIR), Shanahan et al (2003) were able to show that sensor-determined green NDVI values could potentially be used to direct in-season N applications. However, because passive sensor systems rely on natural sunlight, their effectiveness for assessing canopy N status is limited by numerous factors including intermittent cloud cover, narrow time window for operation (around solar noon), and bidirectional reflection issues associated with solar angle (Solari et al. 2008). The majority of agricultural studies use measurements in the visible (400–700nm wavelength) and near infra red (700–2500nm wavelength) region of the electromagnetic spectrum. The principle is that the majority of the red light is absorbed by the chlorophyll in the canopy and therefore little is reflected, in contrast a high proportion of the near infra red light is reflected. As canopy green area increases, either due to increasing crop density or chlorophyll content, the percentage of red reflectance decreases whilst the near infra red reflectance increases. Depending on the canopy and soil type the position of the red edge can also change, this spectral shift is exploited in some research (Boochs et al., 1990). The objective of the present investigation was to compare the predictive power of a) models based on predefined short and broad band for a normalized difference type of index, b) the best combination of wavelengths for a normalized difference type of vegetation index, and c) partial least squares regression (PLS) using all available wavebands.

MATERIAL AND METHODS

The test winter wheat fields were conducted for two years in the farming area of Hokkaido University (43° 4' N 141° 20' E), Sapporo, Japan with annual average precipitation of 1106.5 mm and minimum temperature (-7 c^{o}) in January and maximum temperature (26.4 c^{o}) in August. The fields dimension was 40m × 120m divided into 8 areas and four levels of fertilizer (Ammonium Nitrate) 0, 30, 60, and 90 kg ha^{-1} with two repetitions, were applied at the reviving stage (GS 26), (Zadoks et al. 1974), so that the difference of the growth conditions could be grasped. In the 2010 after the flag leaf stage (GS 37) and in the 2011 after the stem elongation (GS 30), growth investigation in the four growth stages in the first year from 20 points of the reference area and in the eight growth stages in the second year from 56 points, which were set in the fields, was done. Reflectance data by using a Spectroradiometer and The Soil Plant Analysis Development (SPAD) value by a SPAD meter were collected during of growth season in two years. The protein content and yield of grain were measured after harvesting and threshing 1m × 3m area in each reference point in both years. The SPAD meter (MINOLTA Co. LTD.) determines the relative amount of chlorophyll presence by measuring the absorbance of the leaf in two wavelength regions of red and near-infrared. It can provide an indication of chlorophyll content present in plant leaves. According to the catalogue of SPAD 502 (www.konicaminolta.eu 2011) there is high relationship ($R^2 > 0.9$) between SPAD value and leaf nitrogen concentration, therefore it has been widely used in detecting crop chlorophyll and nitrogen content and the guidance of plant healthy and topdressing (Zhang et al. 2003).

Reflectance measurement

Canopy spectral reflectance was measured using a portable spectroradiometer, FieldSpec®3 (FS3), (Analytical Spectral Devices, Inc., USA) from 10 am to 2 pm, under cloudless conditions. The FS3 is a hyper spectral sensor which can measure spectral reflectivity within the range from 350nm to 2500nm with a sampling interval of 1.4nm and 10nm of 350-1050nm and 1000-2500nm respectively that is designed to collect solar radiance, irradiance, and reflectance measurements. The PC was used to acquire data for the control of the FS3 in which special software is installed. The PC and The FS3 communicate through wireless connection that the measured data is transmitted to the PC and saved. The viewing angle was set at 25 degree and height was 150cm from the ground. The calibration was done using a standard white board immediately before measuring reflectance value. Thus, all reflectance data which was more than 100 % in amount of reflection was removed because of noises or absorption by the atmosphere.

Pre-processing for reflectance data

A large amount of spectral data is usually obtained from spectral instruments and yields useful analytical information (Blanco & Villarroya, 2002; Osborne et al, 1993). However, the data acquired from spectrometer contains back- ground information and noise besides sample information. In order to obtain reliable, accurate and stable calibration models, it is very necessary to pre-process spectral data before modelling with Partial Least Square (PLS) (Cen and He, 2007). Spectral pre-processing techniques are required to remove any irrelevant information including noise, uncertainties, variability, interactions and unrecognised features. A lot of pre-processing techniques for spectral data have been developed recently. In this study several pre-processing methods were used such as smoothing (median filter, SavitzkyeGolay and wavelet), differentiation (first derivative) and Max-Min normalisation (standard normal variate transformation (SNV)), and a combination of this three pre-processing.

Partial least squares regression

PLS is a bilinear calibration method using data compression by reducing the large number of measured collinear spectral variables to a few non-correlated principal components (PCs). The PCs represents the relevant structural information, which is present in the reflectance measurements to predict the dependent variable. The spectral data were mean-centred before analysis and the optimal number of principal components was determined by the guidelines described in Esbensen (2000). In principle, PLS regression uses component projection successively to find latent structures. Visual inspection of score-plots and validation residual variance plots was used to find the optimal number of PCs, so that over-fitting was prevented. In most cases, this procedure can reduce the number of spectral variables to a few independent PCs. The final model predicting $\hat{y}$ had the following form Eq. 2:

$$\hat{y}_i = b_0 + b_1 t_{1i} + b_2 t_{2i} + \cdots + b_n t_{ni} \qquad (2)$$

Where t_{1i} to t_{ni} were the scores from principal component (PC) 1 to n for variable i. The scores were calculated on the basis of mean-centred data. By linear regression of t versus y in the calibration iteration process, the regression coefficient b_n was obtained. Due to the initial centring of y, the centred mean b_0 was added in order to obtain $\hat{y}_i$. The performance of the model was evaluated by comparing the differences in the coefficient of determination (R^2) and root mean squares error (RMSE) of prediction. The larger R^2 and the smaller RMSE are the greater precision and the accuracy of the model, to predict crop variables, (Li et al. 2010). The RMSE was calculated using Eq. 3:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (3)$$

Where y_i and $\hat{y}_i$ and n were the measured, predicted and the number of samples, respectively. All data handling, pre-processing and the PLS analysis were performed using Essential FTIR software (www.essentialftir.com 2012).

Crop variable waveband selection

The basis of crop variable waveband selection was the two-band and three-band vegetation indices which listed in table 1. An algorithm was developed in the MATLAB 7.1 SP3 software (The MathWorks, Inc) to calculate all possible two and three wavelengths combinations which presented in table 1. Then linear regression was performed in order to determine the coefficient of determination (R^2). Linear regression was preferred in the initial analysis in order to operate with linear relationships. All the R^2 values derivative from three-band indices were plotted in a three dimension matrix and the plot revealed a characteristic pattern with a number of ‘‘hot spots’’ with relatively high correlation coefficients. These spots were selected by choosing the wavelength combinations that showed a highest R^2 between. On the other hand, the R2 values which calculated with two-band indices were transferred to the Arc map 9.3 (ESRI, CA, USA. 2008) and inverse distance weighted (IDW) interpolated.

Using surface analysis, contour plots for datasets of each year as layers with the same coordinate were made separately. Then intersect overlaying was applied in the analysis tools and common areas with highest elevation (R^2) were calculated. The center wavelength and bandwidth for each of the common spots were determined by fitting a coordinate that could hold the spot of interest inside its limits. At the end, linear ($y= ax + b$), quadratic ($y=ax^2+bx+c$), and exponential ($y= ae^{bx}$) fitting procedures were tested using these new indices and their performance compared.

Table 1. Seven different equations that mentioned in prior literatures were used as vegetation indices to calculate all combination of 2 or 3 bands of reflectance data

RESULTS AND DISCUSSION

The experimental treatments, including different years, nitrogen application, and strategies together with the temporal timing of plant sampling, caused a wide range of variation within the investigated crop variables (Table 2). This wide

Table 2. Selected properties of the investigated crop variables

No. of Indices	**Vegetation Index**	**Equation**	**Reference**
1	Difference Vegetation Index (VDI)	$(\lambda_2/ \lambda_1) -1$	Gitelson et al. (2005)
2	Normalized Difference Vegetation Index (NDVI)	$(\lambda_2- \lambda_1)/ (\lambda_2+ \lambda_1)$	Rouse et al. (1974)
3	Root Difference Vegetation Index (RDVI)	$\sqrt{NDVI*DVI}$	Roujean & Breon (2001)
4	Optimal Soil Adjusted Vegetation Index (OSAVI)	$(1+0.16) (\lambda_2- \lambda_1)/ (\lambda_2+ \lambda_1+ 0.16)$	Rondeax et al. (1996)
5	Plant Senescence Reflectance Index (PSRI)	$(\lambda_2- \lambda_1)/ \lambda_3$	Sime & Gamon (2002)
6	Modified Simple Rate Index (MSRI)	$(\lambda_2- \lambda_1)/ (\lambda_2- \lambda_3)$	Dash and Curan (2004)
7	Simple Rate (SR)	$\lambda_2/ (\lambda_1* \lambda_3)$	Datt (1998)

range in the investigated crop variables (Table 2) was planned in order to make

Year	Crop Variables	Unit	Mean	S.D.	Min	Max	Range
	SPAD	-	41.92	3.23	34.5	45.9	11.4
2010	Yield	Kg/ha	7510.3	836.8	5589	8786	3197
	Protein	%	9.95	1.2	7.9	11.5	3.6
	SPAD	-	43	1.75	38.8	46.3	7.5
2011	Yield	Kg/ha	6948.7	1532.5	4323.3	10440	6116.7
	Protein	%	11.34	1.02	9.45	13.45	4

the relationship between plant performance and reflectance measurements (Fig. 1) as realistic and universal as possible. The reflectance was at all wavelengths on average higher in 2010 in visible (VIS) and short wavebands (SW) regions compared to 2011 (Fig. 1a), but lower in near infrared (NIR) regions. Increasing nitrogen supply caused on average lower reflectance in the VIS spectral range (400~700nm), while the reflectance was higher in the NIR and SW spectral range (700 ~900nm and 900~1350nm, Fig. 1b). The canopy reflection in the near infrared spectral range increased with growth stage, but there was exception in the GS 60 stage of 2010 due to effect of crop lodging (Fig. 1c and d).

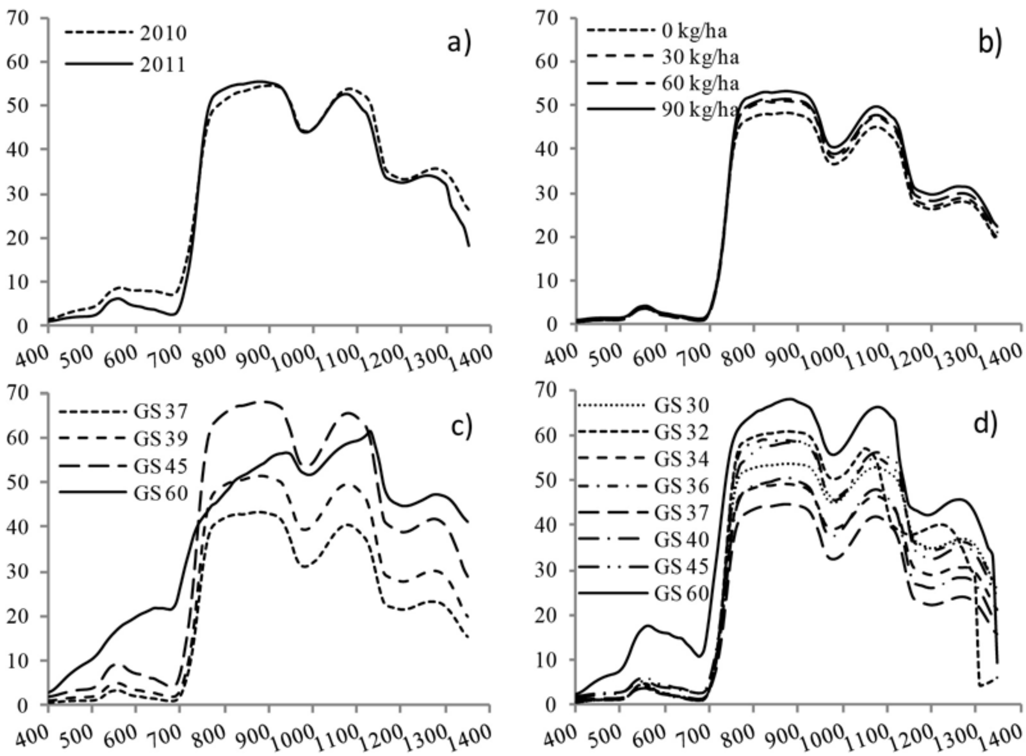

Fig. 1. Average reflectance spectrum of the different experimental treatments; years
(n = 20+56), nitrogen application (n = 20+40), and date of measurement (n = 20 and 56).

Various calibration models were developed by using different pre-processing techniques on the spectral data. Each calibration model was used to predict SPAD value, grain yield and protein content of prediction data set in order to verify the improved ability of models based on different pre-processing techniques. A proper model should have a low root mean squares error (RMSE) and a high coefficient of determination (R^2) between the predicted and measured value of each property. Moreover, a low number of PLS factors are desirable. Models have been developed using different number of PLS factors and different combination of pre-processing techniques. However, only the most accurate models were presented in Table 3 with their R^2 and RMSE. If no pre-processing was applied, a minimum R^2 was observed in the 5 factors of PLS model for prediction of crop variables. However, if pre-processing was applied, R^2 increases and the RMSE were reduced. In the meantime it was found that the number of PLS factors could be reduced by the use of data pre-processing. Max-Min normalisation (SNV) was the best pre-processing method. This pre-processing method with using 8 selected samples of 40 samples as a test set and 32 selected samples of 40 samples as a calibration set increased R^2 from 0.67 to 0.77, from 0.66 to 0.76 and from 0.66 to 0.77 for prediction of SPAD value, yield and protein, respectively whilst the RMSE decreased from 1.98 to 1.28 for SPAD value, from 789 to 677 for yield and from 0.75 to 0.62 for protein in the 5 number of PLS factors.

Table 3. The results of bilinear PLS regression with Max-Min normalization pre-processing method including cumulative coefficient of determination (R^2), root mean squares error (RMSE) with the 5 number of factors

Crop variables	Per-processing	Prediction with 5 Factors		The most important Factors	Efficacy of important factor %	Ranges of important wavelengths	
		R^2	RMSE				
SPAD	Max-Min normalization	0.774	1.28	2	45.6	Range 1	400~500
						Range 2	520~750
						Range 3	1300~1350
Yield	Max-Min normalization	0.763	677	1 , 3	21.1, 29.9	Range 1	405~480
						Range 2	520~750
						Range 3	1010~1350
Protein	Max-Min normalization	0.767	0.62	1	41.5	Range 1	470~530
						Range 2	650~770
						Range 3	840~900

The different components can be defined by their respective scores and PLS regression coefficients. The scores are related to the single samples, while the coefficients quantify the contribution of different wavelengths to the model. The coefficients allow the optimal fit to be achieved for the specific crop variable of interest. The coefficients related to the three investigated crop variables are shown in Fig. 2. According to the figures there are some ranges of important reflecting (table 3), the relationship between the spectral data and the canopy. Three zones at approximately 400~500, 520~750 and 1300~1350 nm for SPAD value, 405~480, 520~750 and 1010~1350 nm for yield and 470~530, 650~770 and 840~900 nm for protein of major importance for the PLS models could be identified. These zones showed either a shift or significant peaks.

A number of ‘‘hot spots’’ with high R^2 were revealed in a linear regression analysis of the individual crop variables against different vegetation indices that calculated according to presented equations in table 1 for all possible two and three combinations of the reflectance measured at the 400 to 1350nm wavelengths which centred with 5nm. Analysis of the centre wavelength and bandwidth in both directions revealed from one to seven indices depending on crop variable (Table 4). Bands with centre wavelengths in the red and NIR spectral regions from 600 to 750nm were represented in almost 50% of all selected bands for all crop variables in total. The bands (λ_1, λ_2 and λ_3) were often paired so that these bands were closely spaced in the steep linear shift between red, NIR and SW reflection (Table 4). The most effective two-band combination estimating SPAD was provided by λ_1 at 730~735 nm combined with λ_2 in the short wave area (SW) 1315~1330nm. For

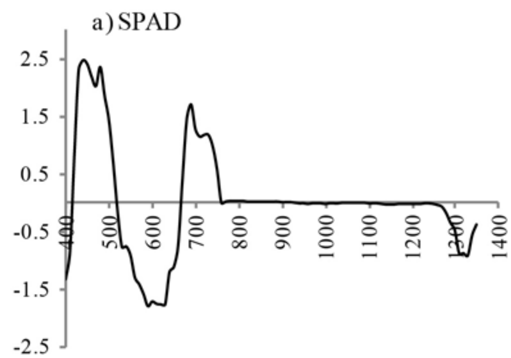

Fig. 2. PLS regression coefficients related to the three investigated crop variables, a) SPAD value, b) grain yield and c) protein content

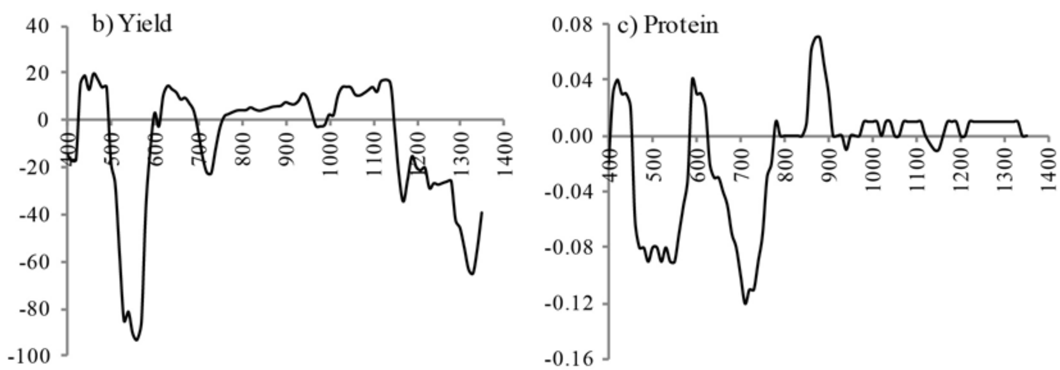

Table 4. Band centers derivative from wavelengths combination in the different vegetation indices according to highest R^2 (‘‘hot spots’’)

indices with three-band combination estimating SPAD was provided by 05~425nm at blue region, 605nm at red area, 705~725nm at red edge and 1325nm at SW region. A relatively broadband (λ_1) in the green region at 525~530nm with a narrow band (λ_2) at the SW shift (1325 nm) was efficient (indices 1, 2 and 4 in Table 2) for yield estimation. Furthermore the blue region with 405~425nm and SW region with 1245~1265nm was appeared in indices 5, 6 and 7 to estimate yield by three-band combination. The red and red edge regions were the most effective bands in the indices for protein prediction, however there were other bands such as green aria at 490nm in the index 3 and NIR region in the indices 5 and 6 for estimation of protein. The relationship between broadband and short-band vegetation indices for the crop physiological variables was investigated and the performance compared to the selected wavelengths. The best coefficients of determination (R^2) using linear, quadratic and exponential regression were obtained for estimates of SPAD value, yield and protein. There was a high degree of coincidence between the selected broad and narrow bands for the best vegetation index and the size of the numerical PLS regression coefficients (comparison of Fig. 2, Table 3, and Table 4). This means that the same wavelengths were important in both methods. The table 5 shows a summarizing of comparison between the results of PLS models and the best two-band and three band vegetation indices.

Table 5. The results of R^2 and RMSE were compared to PLS modeling and the best broad and narrow-band vegetation indices with linear and quadratic fit, respectively for investigated crop biophysical variables; SPAD value, yield,

Crop Variables	Band centre and width (nm)	Band centres (λ_1, λ_2 and λ_3) and band widths ($\Delta\ \lambda_1$, $\Delta\ \lambda_2$ and $\Delta\ \lambda_3$) for two and three vegetation indices						
		Index 1	Index 2	Index 3	Index 4	Index 5	Index 6	Index 7
SPAD	λ_1	735	730	730	730	705	405	705
	$\Delta\ \lambda_1$	8	6	7	5	15	15	15
	λ_2	1325	1330	1315	1330	725	605	425
	$\Delta\ \lambda_2$	10	5	9	5	15	15	15
	λ_3	-	-	-	-	1325	425	1325
	$\Delta\ \lambda_3$	-	-	-	-	15	15	15
Yield	λ_1	530	525	1030	530	1245	405	1325
	$\Delta\ \lambda_1$	5	7	5	5	15	15	15
	λ_2	1325	1325	1120	1325	1265	1265	405
	$\Delta\ \lambda_2$	14	12	10	9	15	15	15
	λ_3	-	-	-	-	405	1245	425
	$\Delta\ \lambda_3$	-	-	-	-	15	15	15
Protein	λ_1	635	635	490	635	745	505	665
	$\Delta\ \lambda_1$	7	7	5	5	15	15	15
	λ_2	700	700	670	700	665	665	885
	$\Delta\ \lambda_2$	12	10	8	8	15	15	15
	λ_3	-	-	-	-	765	685	765
	$\Delta\ \lambda_3$	-	-	-	-	15	15	15

protein.

CONCLUSIONSC

The vegetation indices are a potentially useful for early estimation of crop physiological variables. However, selection of the correct wavelengths and bandwidths are important. The difference vegetation index (VDI) with two wavelengths and plant senescence reflectance index (PSRI) with three wavelengths are appropriate to estimate SPAD value. The root difference vegetation index (RDVI) was high linear fit for early prediction of both yield and protein. Selection of the optimal waveband combination in several kinds of vegetation indices improves the relation to the investigated crop using a linear regression method. The PLS models can further improve his relation to specify important wave bands which was related to crop variables.

REFERENCE

Aparicio, N., Villegas, D., Araus, et all. 2002. Relationship between growth traits and spectral vegetation indices in Durum Wheat. Crop Science, 42, 1547–1555.

Crop Variables	PLS			The best Index linear fit			The best Index Quadratic fit		
	factors	R^2	RMSE	Index	R^2	RMSE	Index	R^2	RMSE
SPAD	5	0.774	1.28	1	0.709	1.45	5	0.841	1.08
Yield	5	0.763	677	3	0.631	831.87	7	0.678	777.09
Protein	5	0.767	0.62	3	0.730	0.667	6	0.779	0.604

Cen, H., & He, Y. 2007. Theory and application of near infrared reflectance spectroscopy in determination of food quality. Trends in Food Science & Technology, 18, 72e83.

Gitelson, A. A, Vin˜a, A, Ciganda, V, et all . 2005. Remote estimation of canopy chlorophyll content in crops. Geophysical Research Letters, 32, L08403. doi:10.1029/ 2005GL022688.

Li Fei., Yuxin Miao., Simon D. Hennig., Martin L., et al. 2010. Evaluating hyperspectral vegetation indices for estimating nitrogen concentration of winter wheat at different growth stages. Precision Agric.11:335–357.

Penuelas, J., & Filella, I. 1998. Visible and near-infrared reflectance techniques for diagnosing plant physiological status. Trends in Plant Science, 3(4), 151–156.

P.M. Hansena, J.K. Schjoerringb. 2003. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. Remote Sensing of Environment 86: 542–553.

Raun, W. R., Solie, J. B., Johnson, G. V., Stone, M. L., Lukina, E. V., et al. (2001). In-season prediction of potential grain yield in winter wheat using canopy reflectance. Agronomy Journal, 93, 131-138.

Solari Fernando, John Shanahan., Richard Ferguson., et al. 2008. Active Sensor Refl ectance Measurements of Corn Nitrogen Status and Yield Potential. Agron. J. 100:571–579.

Wood, G. A., Taylor, J. C., & Godwin, R. J. 2003. Calibration methodology for mapping within-field crop variability using remote sensing. Biosystems Engineering, 84(4), 409–423.

Zadoks, J. C., Chang, T. T., & Konzak, D. F. 1974. A decimal code for the growth stages of cereals. Weed Research, 14, 415–421.

Zhang, J. H., K. Wang, R. C. Wang. 2003. Application of chlorophyll meter SPAD in diagnosis nitrogen status and nitrogenous fertilizer in rice. Journal of Northwest A&F University (Natural Science Edition). 31(2): 177-180.