Combined approach based on principal component analysis and canonical discriminant analysis for investigating hyperspectral plant response

Anna Maria Stellacci,[1] Annamaria Castrignanò,[2] Mariangela Diacono,[1] Antonio Troccoli,[3] Adelaide Ciccarese,[1] Elena Armenise,[1] Antonio Gallo,[3] Pasquale De Vita,[3] Antonio Lonigro,[1] Mario Alberto Mastro,[1] Pietro Rubino[1]

[1]Dipartimento di Scienze Agro-Ambientali e Territoriali, Università di Bari;
[2]CRA – SCA, Unità di Ricerca per i Sistemi Colturali degli Ambienti Caldo-Aridi, Bari;
[3]CRA – CER, Centro di Ricerca per la Cerealicoltura, Foggia, Italy

Abstract

Hyperspectral (HS) data represents an extremely powerful means for rapidly detecting crop stress and then aiding in the rational management of natural resources in agriculture. However, large volume of data poses a challenge for data processing and extracting crucial information. Multivariate statistical techniques can play a key role in the analysis of HS data, as they may allow to both eliminate redundant information and identify synthetic indices which maximize differences among levels of stress. In this paper we propose an integrated approach, based on the combined use of Principal Component Analysis (PCA) and Canonical Discriminant Analysis (CDA), to investigate HS plant response and discriminate plant status. The approach was preliminary evaluated on a data set collected on durum wheat plants grown under different nitrogen (N) stress levels. Hyperspectral measurements were performed at anthesis through a high resolution field spectroradiometer, ASD FieldSpec HandHeld, covering the 325-1075 nm region. Reflectance data were first restricted to the interval 510-1000 nm and then divided into five bands of the electromagnetic spectrum [green: 510-580 nm; yellow: 581-630 nm; red: 631-690 nm; red-edge: 705-770 nm; near-infrared (NIR): 771-1000 nm]. PCA was applied to each spectral interval. CDA was performed on the extracted components to identify the factors maximizing the differences among plants fertilised with increasing N rates. Within the intervals of green, yellow and red only the first principal component (PC) had an eigenvalue greater than 1 and explained more than 95% of total variance; within the ranges of red-edge and NIR, the first two PCs had an eigenvalue higher than 1. Two canonical variables explained cumulatively more than 81% of total variance and the first was able to discriminate wheat plants differently fertilised, as confirmed also by the significant correlation with aboveground biomass and grain yield parameters.

The combined approach proved to be effective, being able to synthesise the redundant radiometric information in a reduced number of indicators of plant nutritional status, which could be utilized to delineate homogeneous within-field areas to be submitted to site-specific fertilization.

Introduction

The rapid and precise detection of crop stress represents the basis for a rationale and sustainable management of natural resources in agriculture. The assessment of health status of agricultural crops or natural vegetation resources is considered of crucial and equal importance for biodiversity preservation, farmland management and precise agriculture running (Kancheva and Borisova, 2008). For characterising plant health status, the analysis of vegetation spectral response may be of great support because stress, due to a variety of causes, induces changes in leaf optical properties (Carter and Knapp, 2001). Spectral reflectance measures are able to provide information on several biophysical and biochemical parameters, such as canopy structure, quantity of biomass, chlorophyll concentration, water content and overall vegetative health (Goetz, 2009).

In the last years, the availability of hyperspectral sensors has widened the potential for plant status characterization. Hyperspectral (HS) narrowbands provide, in comparison to broadbands, a significant improvement in information content and detection accuracy for discriminating land cover types, identifying small differences in green vegetation cover and crop moisture (Jain et al., 2007). In addition, they have proven to be of great utility in detecting plant stress (Carter and Knapp, 2001; Ray et al., 2010).

Although HS data furnish a wealth of detailed information, large volume of data and redundancy (Thenkabail et al., 2002; Broge and Leblanc, 2000) pose a challenge for data processing and for the extraction of crucial information embedded (Delalieux et al., 2007;

Hogervorst and Schwering, 2011; Zhang *et al.,* 2011). Specifically, in the processing of HS data researchers face two, apparently contrasting, needs: on one hand to comprehensively analyse the whole vegetation spectral response, in order to not lose information, and, on the other one, to individuate synthetic indices (optimum set of narrowbands and/or new HS vegetation indices, *etc.*) able to best characterize crop status and different levels of stress (Thenkabail, 2002; Thenkabail *et al.*, 2004; Jain *et al.*, 2007).

Many approaches have been so far proposed for discriminating crops under different conditions, such as: use of reflectance from individual narrowbands; computation of difference and ratio indices (structural indices, chlorophyll indices, red edge indices); calculation of combined indices, which compare an index that has a relatively greater sensitivity to chlorophyll content with one that has greater sensitivity to LAI and canopy cover (*e.g.* CCCI, TCARI/OSAVI, MCARI/OSAVI, MCARI/MTVI2) (Eitel *et al.*, 2008; Li *et al.,* 2010; Basso *et al.*, 2011); computation of derivatives of reflectance spectra; use of multivariate statistical analysis techniques (Thenkabail *et al.*, 2004; Jain *et al.*, 2007; Ray *et al.*, 2010). Among them, the last methods can play a crucial role in analysing HS data set. Indeed, they may allow both to eliminate the redundant information, reducing the data set to a limited number of components, and to identify synthetic indices which maximize differences among levels of stress (Broge and Leblanc, 2000).

Principal Component Analysis (PCA) and Discriminant Analysis (DA) are two major techniques for dimensionality reduction in HS data and crop characterization. PCA aims to derive a new set of orthogonal factors which explain the pattern of correlations and capture most of the variance of the original data set. In this way the numerous initial variables are reduced to a few new explicative components. DA is aimed to identify the variables within the data set which maximize between group variability while minimizing within group variability. Specifically, Canonical Discriminant Analysis (CDA) is a dimension-reduction technique which derives canonical variables, *i.e.* linear combinations of the original variables, that summarize between-class variation in much the same way principal components summarize total variation (SAS, 1999). The methods may then provide complementary information.

In previous research, discriminant analysis furnished positive results, being able to differentiate fertilizer and irrigation managements and identify areas of the canopy under eco-physiological stress; the method has been also successfully used at both leaf and canopy level (Filella *et al.*, 1995; Ray *et al.*, 2010). Apan *et al.* (2004) applied canonical discriminant analysis with a stepwise selection method for discriminating sugarcane disease. Schmidt and Skidmore (2003) also found canonical variate analysis to significantly increase classification accuracy when applied to HS data. The authors stated that it should be considered as the preferred data reduction technique in the analysis of such data set.

Thenkabail *et al.* (2004) adopted a comprehensive analysis to detect optimal wavebands that best described vegetation characteristics, of several spontaneous and cultivated plant species, using: i) principal component analysis (PCA); ii) lambda-lambda R^2 models (LL R^2M); iii) stepwise discriminant analysis (SDA); iv) derivative greenness vegetation indices (DGVI). A similar approach (lambda-lambda plots, PCA and SDA) was also adopted by Jain *et al.* (2007) and Ray *et al.* (2010) to characterise potato spectral response to several stressors [nitrogen (N) and water deficiency; diseases]. The last authors initially performed a reduction of data dimensionality, averaging the narrowbands of the range 395-1075 nm over 10 nm and thus obtaining 69 variables; then, after applying the statistical methods separately, pooled together the wavebands that provided the best results in the three methods and quantified their frequency of occurrence. As concerns N stress, the wavelengths most frequently associated with plant response to the examined N levels were 560, 650, 730 and 760 nm (Ray *et al.*, 2010).

From the previous considerations, it emerges the importance of analysing the whole HS data set in order to clearly interpret plant response to stress, as well as the complementary role of PCA and CDA as a data dimensionality reduction method and a data classification technique, respectively. For these reasons, we propose an integrated approach, based on the combined use of PCA and CDA, to investigate the vegetation spectral response and individuate synthetic indices that best discriminate different N stress levels and could be utilized to delineate homogeneous within-field areas. The approach was preliminary evaluated on a data set collected on durum wheat plants grown under rainfed conditions in southern Italy.

Materials and methods

Site description and hyperspectral data collection

Hyperspectral data were collected within a field trial started in the autumn of 2009 at the Cereal Research Centre of the Italian Agricultural Research Council (CRA-CER, Foggia, 41° 27' N, 15° 36' E, 96 asl, Apulia region, southern Italy).

The area is characterized by climatic conditions typical of the Mediterranean environment, with a dry season between May and September and a cold and rainy season from October-November to March-April (Troccoli *et al.*, 2007). The soil, typical of the Apulian Tavoliere, is clay-loam of alluvial origin classified as *chromic calcixerert* according to USDA Soil Taxonomy, and is characterized by a high average chemical fertility, as indicated by the soil organic matter content (27.4 g kg^{-1}). Mean values of the main chemical and physical soil properties are reported in Table 1.

The experiment was aimed to compare the effect of ten N rates, corresponding to the supply of N amounts ranging between 0 and 180 kg ha^{-1} (0, 60, 80, 90, 100, 110, 120, 140, 160, 180 kg N ha^{-1}), on durum wheat (*Triticum durum* Desf., cv PR22D89) grown continuously. The experimental design consisted of ten unreplicated plots (10 m × 80 m size), each treated with a different N dose.

Wheat was sown at the rate of 350 seeds m^{-2} on 24 December 2010 and harvested at the end of July 2011. Nitrogen was supplied using ammonium nitrate, which was split in two applications, at the end of tillering (2/3, on 22 March 2011) and at booting (1/3, on 29 April 2011) stages. Other cultivation practices were typical of the growing area.

For the collection of spectral signatures, inside each plot (N treatment), two sub-plots were selected (13 m × 10 m size), within which

Table 1. Main soil physical and chemical properties of the experimental site.

Properties	Units	Average values
Physical		
Sand	g kg^{-1}	238
Silt	g kg^{-1}	459
Clay	g kg^{-1}	303
Texture (USDA)		Clay-loam
Chemical		
SOM	g kg^{-1}	27.4
TN	g kg^{-1}	1.11
P	mg kg^{-1}	23.8
CEC	meq 100 g^{-1}	22.4
CaCO$_3$	g 100 g^{-1}	11.4
pH		8.8
EC	ds m^{-1}	0.2

SOM, soil organic matter (Walkley and Black); TN, total nitrogen (Kjeldahl); P, extractable phosphorus (Olsen); CEC, BaCl$_2$ cation exchange capacity; pH in 1:2.5 soil/water extracts; EC in 1:2 soil/water extracts.

the spectral readings were carried out for a total of forty measures (10 plots × 2 sub-plots × 2 readings per sub-plot). Reflectance data were gathered on 24 May 2011 with a high resolution hyperspectral radiometer, ASD FieldSpec HandHeld (ASD Inc., Boulder, CO, USA), covering the wavelength range from 325 nm to 1075 nm. The instrument, which relies on a 512-element photodiode array, acquires hyperspectral data with a spectral resolution (Full Width at Half Maximum - FWHM) of 3.5 nm at 700 nm and a sampling interval (spacing between sample points in the measured spectrum) of 1.5 nm. The FOV (field of view) of the bare fiber-optic probe is 25°. The reflectance of the target is calculated with the calibration measurements of dark current and a white reference panel with known reflectance properties. The sensor was held about 40 cm above the canopy in nadir orientation, and this allowed a spot size of approximately 250 cm^2. The measurements were carried out under clear and cloudless sky conditions between 11:00 and 14:00 at local time. Examples of average experimental reflectance signatures of wheat canopy are reported in Figure 1. On the same sampling date, aboveground biomass was measured in each sub-plot replicate on 1 m by 0.17 m areas. Fresh and dry biomass of the collected samples were quantified.

At harvesting, grain yield was determined by harvesting ten sampling areas of 21.08 m^2 (15.5 m × 1.36 m size) per plot using a plot combine harvester. Yield components and quality parameters (hectoliter weight, grain protein and gluten content) were determined on collected samples. Two sampling areas fell within each previously selected sub-plot replicate; thus, the corresponding yield data were averaged to be used in the following correlation analysis.

Data analysis

Collected reflectance data were initially restricted to 600 spectral wavelengths from 400 to 1000 nm, as they were considered noise-free. In addition, in this study it was decided to concentrate the attention on the 510-1000 nm range because of both the particular importance and suitability of this range for crop stress detection (Carter and Knapp, 2001; Ray *et al*., 2010) and the strong and stable relationship between red and blue reflectances (Jiang *et al*., 2008).

The selected hyperspectral data (510-1000 nm) were afterwards divided into five bands of the electromagnetic spectrum, chosen on the basis of their capabilities to highlight specific vegetation features (Carter, 1993; Sims *et al.,* 2002): i) Green band (510-580 nm, G) is able to focus on the peak reflectance of healthy vegetation; ii) Yellow band (581-630 nm, Y) detects *yellowness*, *i.e.* the degree of chlorosis and senescence of vegetation; iii) Red band (631-690 nm, R) is better focused on the absorption of red light by chlorophyll in healthy plants; iv) Red-Edge band (705-770 nm, RE) is very effective in measuring plant health as strongly related to the plant response to several stressors (Carter and Knapp, 2001); v) near-infrared band (771-1000 nm, NIR) is strongly correlated to canopy cover and leaf area index.

Hyperspectral data were then processed through Principal Component Analysis (PCA) in order to remove the redundant information given by neighbouring bands, by linearly transforming the original set of reflectances into a new set of orthogonal factors. PCA was applied to the correlation matrix and performed separately for each spectral interval. Within each interval only the principal components with eigenvalue greater than 1 (Kaiser criterion) were retained for further analysis. The component loadings were used to interpret the meaning of the new variables.

Canonical Discriminant Analysis (CDA) was applied to the retained PCs for investigating the radiometric differences among N treatments and identifying some synthetic indices able to discriminate treatments effectively. Discriminant Analysis (DA) uses multiple quantitative attributes to discriminate a single classification variable. A discriminant model, also known as classification criterion, is determined by a measure of generalised squared distance (SAS, 1999). The classifica-

tion criterion may be based on either the individual within-group covariance matrices or the pooled covariance matrix, depending on the homogeneity of the within-group matrices. To verify the homogeneity of the within-group covariance matrices, the Bartlett's modification of the likelihood ratio test (Morrison, 1976; Anderson, 1984) was used.

CDA was performed to extract linear combinations of the quantitative variables (in our case, the principal components) that best revealed the differences between the groups (N treatments). The first extracted canonical variables (or components) have the highest possible multiple correlation with the groups. The standardised canonical coefficients indicate the partial contribution of each variable to the canonical component and these structure loadings are commonly used to interpret the meaning of the canonical variable. A scatter plot of the sample data on the plane of the first two canonical variables was computed to visually evaluate the efficacy of CDA in segregating different N fertilizer treatments (stress levels). In order to evaluate the agronomic potential of the extracted canonical variables, a correlation analysis was computed between canonical variables and biophysical parameters, specifically fresh and dry aboveground biomass and grain yield.

Finally, being the red edge region very important in characterizing plant response to stress, a red edge position index (REIP, defined by Guyot and Baret, 1988) was calculated using a combination of four wavebands that approximate the red edge inflection point. REIP index is based on a linear four-point interpolation of the wavelengths 670, 700, 740 and 780 nm (Sticksel *et al.,* 2004; Li *et al*., 2009), and is computed using the following equation:

$$REIP = 700 + 40\ \frac{[R_{670} + R_{780})/2 - R_{700}]}{(R_{740} - R_{700})}$$

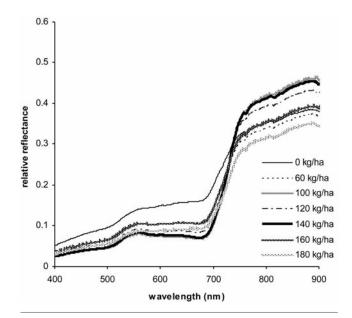PCA and CDA were implemented using PRINCOMP and DISCRIM procedures of SAS/STAT software package (SAS, 2010).

Figure 1. Examples of average experimental reflectance signatures of wheat canopy.

Results and discussion

Weather conditions and yield response

The monthly minimum and maximum air temperatures and the rainfall recorded across the wheat growing cycle, in comparison with the long-term trends (1955-2009), are reported in Figure 2. The 2010-2011 autumn-spring period showed higher rainfall than the 55-year averages (627.4 mm *vs* 436.2 mm). Particularly abundant rainfalls were recorded in the October-December period (+123.6 mm than long-term values), with +104.2 mm in October and +45.8 mm in November. This allowed the constitution of a significant water reserve in the soil for the following wheat growing cycle. Consistent rainfall events were also recorded in January (+14.6 mm), March (+43.9 mm) and in April-May (+28.5 mm). Slightly lower rainfall (-13.6 mm) was finally recorded in June. As concerns the thermal regime, the 2010-2011 growing season was characterized by a cold winter and a fresh spring. Monthly air temperatures were indeed lower than long-term averages, with minimum air temperatures of 1.65 °C lower in the December-February period and maximum temperatures lower in March and May (-1.6 and -1.9°C).

Grain yield was strongly affected by different N supply. Average grain yield was 4.28 t ha^{-1} and ranged from 2.34 (0 kg N ha^{-1}) to 5.62 t ha^{-1} (120 kg N ha^{-1}). As a consequence of the high average soil fertility, highest yields were recorded under intermediate N rates, whereas a yield decrement was observed in correspondence of the highest N level (Table 2). Protein content showed a linear increase up to the rate of 100 kg N ha^{-1}; further increases in N supply did not result in higher values. Hectoliter weight and seeds weight were less affected by varying N supply.

Hyperspectral plant response

According to the Kaiser criterion and to the percentage of variance explained, only the first principal component (PC) was retained within the spectral intervals of green, yellow and red, as it had an eigenvalue greater than 1 (Table 3) and explained more than 95% of the total variance (PC_G1, PC_Y1, PC_R1). These results point out the high correlation existing between neighbouring wavelengths and underline the importance of adopting multivariate statistical techniques in order to reduce data dimensionality and synthesise redundant information.

Within the ranges of red-edge (RE) and NIR, the first two PCs showed an eigenvalue greater than 1. In the NIR region, the first PC (PC_NIR1) was able to synthesize the major part of the total variance and the general behaviour of this spectral interval, as its component loadings were all equally weighted. The second component (PC_NIR2), which however explained only a low percentage of the total variance,

was instead more related to the wavelengths of the 952-970 nm range. Within the RE interval, the first two components (PC_RE1, PC_RE2) explained respectively 72.8% and 27% of the total variance. The first component was more correlated to the wavelengths in the central part of the spectral interval (730-745 nm). The second component was instead characterised by positive loadings up to 735 nm while negative up to the end of the interval. In particular, it was observed a positive high correlation to the wavelengths of the range 705-717 nm, whereas a negative correlation to those of the range 754-770 nm. The existence of two significant components in the RE interval can be attributed to the wide variability of plant radiometric response in this spectral region.

The characteristic red edge reflectance pattern of vegetation has been deeply investigated by many researches, all of which reported that the observed blue-shift and red-shift of the red edge inflection point (REIP) can be related to plant growth conditions (Broge and Leblanc, 2000). REIP can be located at around 720 nm on average; shifts toward shorter wavelengths (blue-shift) can be associated with a decrease in green vegetation density; whereas shifts toward longer wavelengths (red-shift) with an increase in green plant material (Broge and Leblanc, 2000; Carter and Knapp, 2001). The position of REIP is there-
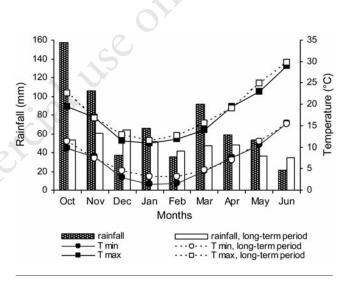

Figure 2. Monthly mean minimum and maximum air temperature (T_{min} and T_{max}) and total rainfall recorded over the experimental period in comparison with long-term trends (1955-2009).

Table 2. Quantitative and qualitative yield response of the 10 compared nitrogen treatments.

Treatments (kg N ha^{-1})	Grain yield (t ha^{-1})	Protein (% DM)	Gluten (% DM)	Hectoliter weight (kg hL^{-1})	1000 seeds weight (g)
0	2.34±0.15	9.90±0.07	5.88±0.06	80.83±0.33	45.93±0.38
60	3.73±0.14	12.53±0.29	8.85±0.35	80.45±1.14	45.60±2.38
80	3.07±0.10	13.98±0.39	10.15±0.33	78.68±1.23	43.35±1.04
90	4.96±0.05	13.93±0.29	10.08±0.31	81.33±0.36	46.78±1.44
100	4.89±0.08	15.05±0.24	11.43±0.32	79.95±0.80	46.28±0.73
110	4.23±0.28	14.88±0.37	11.05±0.47	79.28±0.93	45.13±1.23
120	5.62±0.25	13.65±0.33	9.60±0.30	81.93±0.90	50.00±1.50
140	5.05±0.26	14.98±0.44	10.98±0.44	79.55±1.92	46.55±2.06
160	4.95±0.09	14.83±0.02	10.90±0.11	79.85±0.96	45.93±1.98
180	3.94±0.32	15.35±0.12	11.43±0.11	77.33±0.96	45.33±1.42

N, nitrogen; DM, dry matter. Data are expressed as mean ± standard error.

Table 3. Eigenvalue and percentage of variance explained by the first three principal components of each examined spectral interval.

Spectral interval	Spectral range (nm)	n of variables	PCs	Eigenvalue	% of variance explained
Green	510-580	71	*PC_G1*	*70.65*	*99.51*
			PC_G2	0.27	0.38
			PC_G3	0.06	0.08
Yellow	581-630	50	*PC_Y1*	*49.89*	*99.79*
			PC_Y2	0.06	0.12
			PC_Y3	0.04	0.07
Red	631-690	60	*PC_R1*	*59.89*	*99.82*
			PC_R2	0.09	0.15
			PC_R3	0.01	0.02
Red-Edge	705-770	66	*PC_RE1*	*48.03*	*72.78*
			PC_RE2	*17.82*	*27.00*
			PC_RE3	0.09	0.13
NIR	771-1000	230	*PC_NIR1*	*226.43*	*98.45*
			PC_NIR2	*2.06*	*0.89*
			PC_NIR3	0.61	0.27

PCs, principal components; NIR, near-infrared; the selected PCs within each interval are indicated in italics.

fore a strong indicator of chlorophyll content: the higher the number, the greater the amount.

Although explaining only 27% of total variance, the second RE component showed a close indirect association with the REIP index (r=-0.954, P<0.0001), computed according to Guyot and Baret (1988). The REIP index increased passing from the not fertilised treatment (719.6 nm) to the fertilised plants (*i.e.* 725.9 nm for the treatment which received 100 kg N ha^{-1}, Table 4). Although this index is derived by a simplified formula, its efficacy has been pointed out in other studies. Sticksel *et al.* (2004), among several hyperspectral indices, observed that REIP together with IRI (InfraRed Index) were the most sensitive to the N induced heterogeneities in winter wheat crops. In addition, Li *et al.* (2009), comparing different red edge vegetation indices for estimating winter wheat N status, noticed that red-edge position indices performed better than normalised and simple ratio red edge indices. Further statistical analysis was carried out on the seven new variables extracted by means of Principal Component Analysis (PC_G1, PC_Y1, PC_R1, PC_RE1, PC_RE2, PC_NIR1, PC_NIR2). The bands more affecting wheat spectral response were the green, yellow, red, the second component of RE and the first component of NIR (Table 5). This proved that reflectance in these intervals is able to discriminate crop nutritional status at anthesis and that canopy optical properties are highly affected by nutritional stress (Carter and Knapp, 2001). Specifically, passing from the unfertilised control to the fertilised plants, the recorded radiometric reflectances showed a trend to decrease in the visible region, whereas to increase in the NIR region. These results were in agreement with findings by Zhu *et al.* (2006), which analysed the canopy reflectance response spectra to N rates on wheat. The authors observed that the spectral reflectance was negatively correlated to N rate in the visible wavebands (460-710 nm) and near infrared (NIR) long wavebands (1480-1650 nm), whereas in the NIR short wavebands (760-1220 nm) the reflectance tended to increase with increasing N rate (Zhu *et al.,* 2006). Liew *et al.* (2008) also reported that many nutrient deficiencies induced a decrease in chlorophyll content, a concomitant increase in reflectance in the visible range (400-700 nm) and a blue-shift in the red edge inflection point. The Bartlett's modification of the likelihood ratio test, carried out to verify the homogeneity of the within-group covariance matrices, was not significant at a probability level P<0.05; therefore, the discriminant analysis was carried out on the pooled covariance matrix. Two canonical variables showed an eigenvalue higher than one, explaining cumulatively more than 81% of the total variance (Tables 6 and 7). The performance of the discriminant function in classifying future observa-

Table 4. Average values of the Red Edge Inflection Point index, computed according to Guyot and Baret (1988), for the 10 compared nitrogen treatments.

Treatments(kg N ha^{-1})	REIP (nm)
0	719.6 (±0.35)
60	721.9 (±0.19)
80	723.0 (±0.64)
90	724.4 (±0.46)
100	725.9 (±0.36)
110	723.7 (±0.49)
120	724.4 (±0.51)
140	724.7 (±0.65)
160	722.1 (±0.72)
180	723.7 (±0.44)

REIP, Red Edge Inflection Point index; N, nitrogen. Data are expressed as mean ± standard error.

Table 5. R^2 performed on the extracted principal components.

	R^2
PC_G1	0.7366
PC_Y1	0.7586
PC_R1	0.7399
PC_RE1	0.3192
PC_RE2	0.7477
PC_NIR1	0.4071
PC_NIR2	0.1709
Average R^2	0.5543

PC, principal component.

Table 6. Eigenvalue, percentage of variance explained and significance of the seven canonical variables extracted.

Canonical variables	Eigenvalue	% of variance explained	Likelihood ratio	Pr>F
Can 1	5.15	59.6	0.0166	<0.0001
Can 2	1.88	21.7	0.1021	0.0257
Can 3	0.99	11.5	0.2939	0.3717
Can 4	0.45	5.2	0.5876	0.8822
Can 5	0.13	1.5	0.8518	0.9931
Can 6	0.03	0.3	0.9635	0.9972
Can 7	0.01	0.1	0.9899	0.9585

tions was assessed through the analysis of the confusion matrix. In table 8, the single treatment classification accuracy, expressed as proportion of correctly classified observations in each group, and the overall accuracy are reported. Unfertilised plants were clearly classified and no confusion was observed with other treatments. Treatments ranging from 60 to 110 kg N ha^{-1} showed also a high accuracy. Lower treatment accuracy was observed for higher N levels with particular regard to the rate of 120 kg N ha^{-1}. The overall discrimination may be considered satisfactory since the overall accuracy is 70%; this indicates that about 3/4 of the data was correctly classified.

The observation of the standardised canonical coefficients (Table 7) showed that on the first canonical variable, which accounted for 59.6% of total variance, the yellow component and the first of RE weighted more and positively and were indirectly correlated to the green, the first NIR and the second RE components. On the second canonical variable, which accounted for 21.7% of the total variance, both red edge components weighted more and positively and were negatively correlated with yellow, red and NIR1 components (Table 7). From the observation of the first canonical variable coefficients, it emerged the role of the green-yellow and far-red (red-edge) intervals, as well as of the NIR region (that at canopy level is highly related to canopy cover and LAI), in characterising the vegetation spectral response. Carter and Knapp (2001) observed that the major differences in optical plant response to stress occur in the far-red and green-yellow regions of the spectrum. Reflectance differences between stressed and healthy leaves in the 400-500 and 670-680 nm tend instead to be low. The reason is that, in the 670-680 nm range, the absorption may saturate due to the strong chlorophyll absorptivity and then, only with the lost of large amounts of chlorophyll from the leaves, significant optical differences occur (Carter and Knapp, 2001). To evaluate the discriminating capability of the extracted canonical variables, the scatter plot of the sample data on the plane of the first two canonical variables (Figure 3) was inspected. The plot showed that the first canonical component was able to clearly distinguish the unfertilised from all the fertilised plants. Moreover, among the fertilised treatments, the plants receiving 100 kg N ha^{-1} (which is the N dose commonly used by the farmers of the study area) were differently classified from those which received the lowest (60 and 80 kg N ha^{-1}) and the highest (160 and 180 kg N ha^{-1}) N rates (Figure 3). The other N supplies resulted in an intermediate behaviour and appeared close to the 100 kg N ha^{-1} treatment. The agronomic potential of the first canonical variable extracted was finally evaluated through a correlation analysis with biophysical (plant biomass) and quantitative and qualitative yield parameters. The analysis highlighted the high degree of association between the canonical variable and the examined plant response parameters with particular regard to grain yield and protein content (Table 9). This finding was indicative of the ability of this canonical component to characterise agronomic plant response to N treatments.

The results of this preliminary study pointed out that canonical variables may represent synthetic indices able to characterize main spectral features of plant response to N supply and plant nutritional status. Therefore these indices could play a key role in delineating homogeneous within-field areas to be submitted to site-specific management.

Conclusions

The analysis of the spectral response of vegetation through remote and proximal sensing could be a valid means to rapidly investigate and define the overall plant health and nutritional status. The development of efficient procedures for data analysis is still one of the most essential issues and its importance is directly related to the ever-increasing amount of data provided by numerous sensors, such as hyperspectral
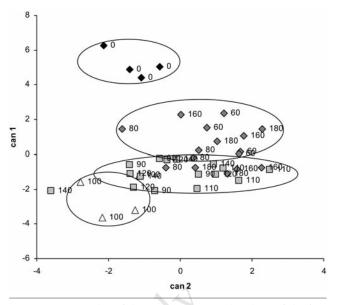

Figure 3. Scatterplot of the sample data on the plane of the first two canonical variables. N rates supplied (kg ha^{-1}) are reported near to the sample points.

Table 7. The total-sample Standardised Canonical Coefficients of the first two canonical variables.

	Can 1	Can 2
PC_G1	-19.3647	1.6752
PC_Y1	23.3357	-4.4118
PC_R1	-0.7744	-3.9016
PC_RE1	16.1896	3.1499
PC_RE2	-11.0119	4.9738
PC_NIR1	-16.9738	-3.2185
PC_NIR2	1.6229	0.6919
Eigenvalue	5.15	1.88
% variance explained	59.6	21.7

PC, principal component.

Table 8. Treatment accuracy and overall accuracy from the error matrix.

				Treatment accuracy						Overall accuracy
0	60	80	90	100	110	120	140	160	180	
1.00	1.00	0.75	0.75	0.75	1.00	0.25	0.50	0.50	0.50	0.70

Table 9. Correlation matrix of the first canonical variable, biophysical (dry and fresh plant biomass) and yield parameters (grain yield, grain protein and gluten content).

	Can 1	Grain yield	Grain protein	Gluten	Fresh biomass
Grain yield	-0.736***	-			
Grain protein	-0.805***	0.591**	-		
Gluten	-0.816***	0.573**	0.995***	-	
Fresh biomass	-0.582**	0.659**	0.680**	0.671**	-
Dry biomass	-0.556*	0.644**	0.652**	0.646**	0.991***

***$P<0.001$; **$P<0.01$; *$P<0.05$.

instruments (Kancheva and Borisova, 2008). Multivariate statistical techniques may play a crucial role in the analysis of hyperspectral data sets. The combined approach adopted in this study, based on the use of PCA and CDA, proved to be effective, being able to summarize the redundant radiometric information in a reduced number of indicators of plant nutritional status, which could be utilized to delineate homogeneous within-field areas to be submitted to site-specific fertilization. In particular, the first canonical variable extracted discriminated most differently N fertilised durum wheat plants and was significantly correlated to biophysical and yield parameters. Further research is needed to evaluate the efficacy of this approach across the crop cycle, *i.e.* at different phenological stages of wheat and on different crops and to compare the results to those obtained with other methodological approaches. Since the correlation between individual bands and biophysical parameters varies during the crop cycle, the use of only one band may not be able to discriminate the different stress levels at any phenological stage, whereas a weighted combination of all the monitored bands may lead to a better discrimination of differently fertilised areas. No single best approach is so far available to define optimal indices or band sub-sets which provide the best estimates of vegetation characteristics (Thenkabail *et al.,* 2004). Different approaches may be followed, with different combination of statistical techniques, but the adoption of multivariate methods represents undoubtedly one of the most promising way.

References

Anderson TW, 1984. An introduction to multivariate statistical analysis. 2nd ed. John Wiley & Sons, New York, NY, USA.

Apan A, Held A, Phinn S, Markley J, 2004. Detecting sugarcane 'orange rust' disease using EO-1 Hyperion hyperspectral imagery. Int. J. Remote Sens. 25:489-498.

Basso B, Cammarano D, Cafiero G, Marino S, Alvino A, 2011. Cultivar discrimination at different site elevations with remotely sensed vegetation indices. Ital. J. Agron. 6:e1.

Broge NH, Leblanc E, 2000. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. Remote Sens. Environ. 76:156-172.

Carter GA, 1993. Responses of leaf spectral reflectance to plant stress. Am. J. Bot. 80:239-243.

Carter GA, Knapp AK, 2001. Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. Am. J. Bot. 88:677-684.

Delalieux S, van Aardt J, Keulemans W, Schrevens E, Coppin P, 2007. Detection of biotic stress (Venturia inaequalis) in apple trees using hyperspectral data: non-parametric statistical approaches and physiological implications. Eur. J. Agron. 27:130-143.

Eitel JUH, Long DS, Gessler PE, Hunt ER, 2008. Combined spectral index to improve ground-based estimates of nitrogen status in dryland wheat. Agron. J. 100:1694-1702.

Filella I, Serrano L, Serra J, Peñuelas J, 1995. Evaluating wheat nitrogen status with canopy reflectance indices and discriminant analysis. Crop Sci. 35:1400-1405.

Goetz AFH, 2009. Three decades of hyperspectral remote sensing of the Earth: a personal view. Remote Sens. Environ. 113:S5-S16.

Guyot G, Baret F, 1988. Utilisation de la haute resolution spectrale pour suivre l'etat des couverts vegetaux. pp 279-286 in Proc. 4th Int. Colloquium on Spectral Signatures of Objects in Remote Sensing, Aussois, France.

Hogervorst MA, Schwering PBW, 2011. Hyperspectral data analysis and visualization. In: K. Funatsu and K. Hasegawa (ed.) Knowledge-oriented applications in data mining. In Tech Publ., pp 183-205.

Jain N, Ray SS, Singh JP, Panigraphy S, 2007. Use of hyperspectral data to assess the effects of different nitrogen applications on a potato crop. Precis. Agric. 8:225-239.

Jiang Z, Heute AR, Didan K, Miura T, 2008. Development of a two-band enhanced vegetation index without a blue band. Remote Sens. Environ. 112:3833-3845.

Kancheva RH, Borisova DS, 2008. Vegetation spectral response to stress conditions. Remote Sensing of the Earth and Planet. Fundamental Space Research, Bulgaria. Available from: http://www.stil.bas.bg/FSR/PDF/TOP2Kancheva_Rumiana22508.pdf

Li F, Miao Y, Gnyp ML, Hennig SD, Chen X, Jia L, Bareth G, 2009. Evaluating red edge vegetation indices for estimating winter wheat N status under high canopy coverage conditions. Proc. 16th Int. Plant Nutrition Colloquium, Davis, CA, USA. Available from: http://escholarship.org/uc/item/0f46f63r

Li F, Miao Y, Hennig SD, Gnyp ML, Chen X, Jia L, Bareth G, 2010. Evaluating hyperspectral vegetation indices for estimating nitrogen concentration of winter wheat at different growth stages. Precis. Agric. 11:335-357.

Liew OW, Chong PCJ, Li B, Asundi AK, 2008. Signature optical cues: emerging technologies for monitoring plant health. Sensors 8:3205-3239.

Morrison DF, 1976. Multivariate statistical methods. McGraw-Hill Ed., New York, NY, USA.

Ray SS, Singh JP, Panigraphy S, 2010. Use of hyperspectral remote sensing data for crop stress detection: ground-based studies. International Archives of Photogrammetry, Remote Sensing and Spatial Information Science, volume 38, Part 8. Kyoto, Japan.

SAS, 1999. The CANDISC Procedure and The DISCRIM Procedure. SAS/STAT User's Guide, Ver. 8. SAS Institute Inc., Cary, NC, USA.

SAS, 2010. SAS/STAT Software, Rel. 9.2. SAS Institute Inc., Cary, NC, USA.

Schmidt KS, Skidmore AK, 2003. Canonical variate analysis increases vegetation classification accuracy with hyperspectral field data. In "Hyperspectral Remote Sensing of Vegetation Species Distribution in a Saltmarsh". Degree Diss., International Institute for Geo-Information Science and Earth Observation, Entsched, The Netherlands.

Sims DA, Gamon JA, 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. Remote Sens. Environ. 81:337-354.

Sticksel E, Schächtl J, Huber G, Liebler J, Maidl FX, 2004. Diurnal variation in hyperspectral vegetation indices related to winter wheat biomass formation. Precis. Agric. 5:509-520.

Thenkabail PS, 2002. Optimal hyperspectral narrowbands for discriminating agricultural crops. Rem. Sens. Rev. 20:257-291.

Thenkabail PS, Enclona EA, Ashton MS, Van Der Meer B, 2004. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. Remote Sens. Environ. 91:354-376.

Thenkabail PS, Smith RB, De Pauw E, 2002. Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization. Photogramm. Eng. Rem. S. 68:607-621.

Troccoli A, Colecchia SA, Cattivelli L, Gallo A, 2007. Caratterizzazione agroclimatica del capoluogo dauno - Analisi della serie storica delle temperature e delle precipitazioni rilevate a Foggia dal 1955 al 2006. Typography Digital Print, Orta Nova (FG), Italy.

Zhang ZY, Ma XM, Liu GS, Jia FF, Qiao HB, Zhang YW, Lin S, Song WF, 2011. A study of hyperspectral estimating models of tobacco leaf area index. Afr. J. Agric. Res. 6:289-295.

Zhu Y, Li Y, Feng W, Tian Y, Yao X, Cao W, 2006. Monitoring leaf nitrogen in wheat using canopy reflectance spectra. Can. J. Plant Sci. 86:1037-1046.