Bulgarian Journal of Agricultural Science, 25 (No 5) 2019, 819–832

Estimation of biophysical and biochemical variables of winter wheat through Sentinel-2 vegetation indices

Petar Dimitrov[1*], Ilina Kamenova[1], Eugenia Roumenina[1], Lachezar Filchev[1], Iliana Ilieva[1], Georgi Jelev[1], Alexander Gikov[1], Martin Banov[2], Veneta Krasteva[2], Viktor Kolchakov[2], Milena Kercheva[2], Emil Dimitrov[2], Nevena Miteva[2]

[1]*Bulgarian Academy of Sciences, Space Research and Technology Institute, Department of Remote Sensing and GIS, 1113 Sofia, Bulgaria*
[2]*Agricultural Academy, Institute of Soil Science, Agrotechnologies and Plant Protection "Nikola Poushkarov", 1331 Sofia, Bulgaria*
[*]*Corresponding author:* petar.dimitrov@mail.space.bas.bg

Abstract

Dimitrov, P., Kamenova, I., Roumenina, E., Filchev, L., Ilieva, I., Jelev, G., Gikov, A., Banov, M., Krasteva, V., Kolchakov, V., Kercheva, M., Dimitrov. E., & Miteva. N. (2019) Estimation of biophysical and biochemical variables of winter wheat through Sentinel-2 vegetation indices. *Bulgarian Journal of Agricultural Science, 25*(5), 819–832

Traditionally, the growth and physiological status of winter wheat (*Triticum aestivum* L.) is monitored in the field by measuring different biophysical and biochemical variables such as Above Ground Biomass (AGB), Nitrogen content (N), N uptake, Leaf Area Index (LAI), Fraction of vegetation Cover (fCover), Canopy Chlorophyll Content (CCC), and fraction of Absorbed Photosynthetically Active Radiation (fAPAR). The objective of this study was to investigate the possibility of estimating these crop variables through statistical regression modelling and spectral vegetation indices derived by the Sentinel-2 satellites. Field data were collected over two growing seasons, 2016/2017 and 2017/2018, in test fields around Knezha, northern Bulgaria. A combination of spectral data from Sentinel-2 images and field spectroscopy obtained through the first growing season was used for model calibration and cross-validation. The models were further validated with Sentinel-2 image data from the second growing season. The accuracy of the models varied widely across crop variables. According to the cross-validation, the relative RMSE was below 25% for fAPAR, fCover, and fresh AGB, with particularly good result for fAPAR (13%). For N content and dry AGB the error was between 25% and 30%. The accuracy was low for CCC, LAI, and N uptake (error between 30% and 43%). The models' performance was worse when they were applied to the data from the second growing season, resulting in relative RMSE which were 3-8% higher in the general case. The cross-validation results suggested that the variety-specific models are more accurate than the generally calibrated models for most crop variables. The accuracy obtained in this study for the prediction of fAPAR, fCover and AGBf through VIs is promising. Future studies and incorporation of new field data will be needed to better account for variety, season, and site variations in the modelled relationships and to improve their generalisation potential.

Keywords: biomass; canopy chlorophyll; leaf area; nitrogen content; satellite imagery

Introduction

Winter wheat (*Triticum aestivum* L.) is the most widespread crop in Bulgaria covering one third of the total arable land of the country (MAFF, 2016). As a main food crop it has strategic importance at national level and large economic effect for the individual agricultural producers. The monitoring of growth and the timely assessment of status of winter wheat along the season are thus essential for food security and the economic results of farmers.

Different biophysical and biochemical variables (hereafter called crop variables for brevity) can be used to describe and monitor winter wheat growth and status. For instance, the Above Ground Biomass (AGB), Leaf Area Index (LAI) and Fraction of vegetation Cover (fCover) are directly related to the amount of vegetation and are thus indicators of crop growth. The nitrogen (N) content in leaves is an important abiotic stress indicator (Banerjee et al., 2015). Other variables used to characterize the nitrogen status of crops are N uptake (nitrogen content per unit ground area) and Canopy Chlorophyll Content (CCC). The information about the fraction of Absorbed Photosynthetically Active Radiation (fAPAR) may be used to calculate dry matter and grain yield (Casanova et al., 1998; Lobell et al., 2003). These and other crop variables are traditionally measured by destructive sampling or other labor-intensive field methods. However, remote sensing methods have been demonstrated to be a feasible cost-effective alternative (Feng et al., 2015).

Relationships between different crop variables and remotely sensed spectral vegetation indices (VIs) have been established based on field spectroscopy (Hinzman et al., 1986; Wiegand et al., 1992; Viña et al., 2011; Feng et al., 2015) and satellite data (Delegido et al., 2013; Kross et al., 2015; Shang et al., 2015; Clevers et al., 2017). These relations are the basis for the regression technique of crop variable retrieval - an empirical approach using ground measurements of the crop variable to calibrate a statistical model. A vegetation index map and the calibrated model can then be used to map the crop variable over an area of interest and thus characterise its spatial variability.

Sentinel-2 is an optical high-resolution mission of the European Space Agency with two satellites currently in orbit – Sentinel-2A and Sentinel-2B. The mission is characterised by a unique combination of high spatial resolution (up to 10 m), wide field of view (290 km) and spectral coverage (13 spectral bands spanning from the visible and the near infrared to the short wave infrared) (Drusch et al., 2012).

The capabilities of Sentinel-2 vegetation indices for quantitative estimation of LAI, CCC, and leaf chlorophyll concentration have been evaluated before (Delegido et al., 2011; Frampton et al., 2013; Verrelst et al., 2015). Most previous studies, however, used hyperspectral airborne or satellite sensors to simulate Sentinel-2 data. There is still little experience with deriving crop variables by using actual Sentinel-2 images (Söderström et al., 2017; Chemura et al., 2018; Delloye et al., 2018; Zheng et al., 2017; Pan et al., 2018).

The objective of this study was to investigate the capability of Sentinel-2 vegetation indices to estimate AGB, N, N uptake, LAI, CCC, fAPAR and fCover of winter wheat.

Material and Methods

Field campaigns

The field data used in this study was collected on *Zlatia* test site (Figure 1) as part of the TS2AgroBG project. The test site is situated in northern Bulgaria, municipality of Knezha. Four field campaigns were carried out during the 2016/2017 growing season and two in 2017/2018 growing season (Table 1). In each season, six winter wheat fields were sampled, three sown with winter wheat Enola variety and three sown with winter wheat Annapurna variety. Thirty Elementary Sampling Units (ESUs) were distributed over these fields and measured during every field campaign. The size of ESU was ~20×20 m^2.

Samples to determine AGB, N content and N uptake were collected by cutting all aboveground parts of the crop at ground level from one square meter in each ESU. The fresh weight of the samples were measured immediately after the field work at the end of the day and used to calculate the fresh AGB (AGBf) in grams per m^2. Dry biomass (AGBd) was determined after drying representative sub-sample at 60°C to constant weight. N content of the above-ground organs (leaf and stem) is determined by the Kjeldahl method and expressed as per-cent of the dry weight. N uptake (g m^{-2}) is determined as the product of N content and AGBd. Therefore, N uptake increases with the canopy development (Mistele & Schmidhalter, 2008). AccuPAR LP-80 instrument was used to measure LAI, fAPAR and fCover in three locations along the diagonal of the ESU. In the same locations, the chlorophyll content of leaves was measured with hand-held chlorophyll content meter CCM-300. CCM-300 is a modulated fluorometer that uses the emission ratio of red chlorophyll fluorescence at 700 nm to the far red fluorescence at 735 nm (Gitelson et al., 1999). The instrument provides direct read out of chlorophyll content at leaf level in mg m^{-2}. The CCC (g m^{-2}) was derived as the product of the LAI (m^2m^{-2}) and the CCM-300 readings converted in g m^{-2}.

When the weather conditions permitted, crop reflectance spectra were collected using field spectroradiometer ASD FieldSpec4. Three measurements in each ESU were collected from the same locations where LAI, fAPAR, fCover and chlorophyll content were measured. The crop reflectance

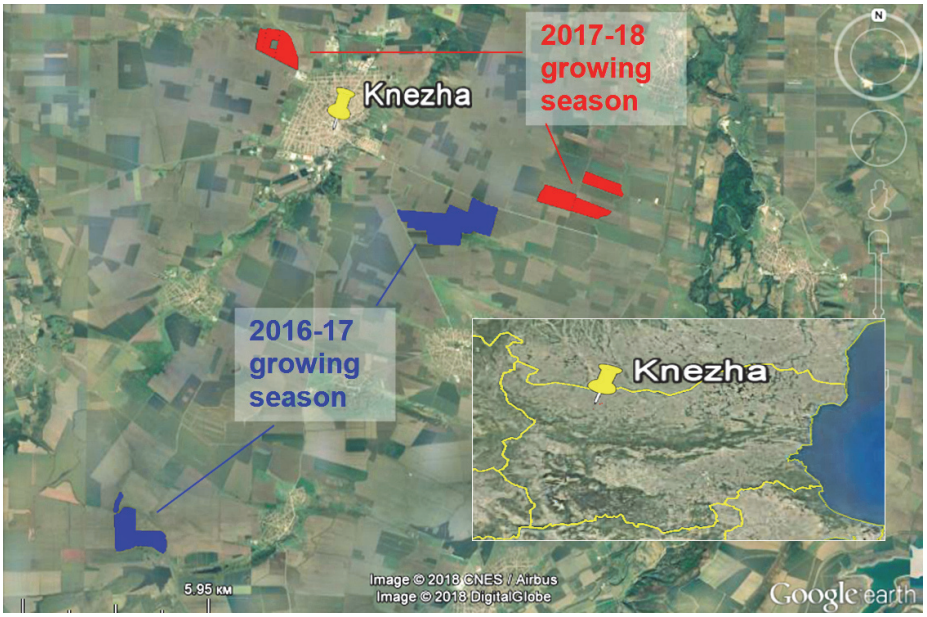

Fig. 1. Location of the test site and the winter wheat fields sampled in 2016/2017 and 2017/2018 growing seasons

Table 1. Field campaigns and corresponding Sentinel-2 images used in the study

Growing season	Field campaign dates	Development stage (after Zadoks)	Date of used Sentinel-2 image
2016/2017	07–11.11.2016	Tillering (Z20)	31.10.2016
	20–24.03.2017	Tillering (Z21-Z26)	–
	24–28.04.2017	Stem elongation (Z31-Z34)	19.04.2017
	15–19.05.2017	Anthesis (Z65 – Z69)	–
2017/2018	06–10.11.2017	Tillering (Z20-Z23)	31.10.2017
	02–05.04.2018	Stem elongation (Z30-Z31)	30.03.2018

spectrum for the ESU was then derived by averaging the three measurements. Auxiliary data were also collected including development stage and height of the crop and qualitative assessment of weed status. For this study, ESUs with significant weeds were excluded from the calibration dataset of 2016/2017 growing season but retained in the validation dataset of 2017/2018 growing season. The geographic coordinates of ESUs were determined on the field with GPS. Table 2 shows descriptive statistics of the final dataset used in the further analysis.

Table 2. Descriptive statistics on the field measurements from 2016/2017 and 2017/2018 growing seasons

	LAI	fAPAR	fCover	CCC (g m^{-2})	AGBf (g m^{-2})	AGBd (g m^{-2})	N (%)	N uptake (g m^{-2})
	2016/2017 growing season							
min	0.04	0.12	0.02	0.02	53	9	0.57	0.27
max	7.88	0.95	0.97	3.22	4629	1194	3.23	18.66
mean	3.17	0.69	0.64	1.46	2149	525	1.53	5.49
std.dev.	2.19	0.28	0.34	0.98	1461	371	0.86	4.08
# ESUs	83	83	83	83	87	87	87	87
	2017/2018 growing season							
min	0.06	0.11	0.02	0.02	80	15	0.98	0.31
max	3.75	0.90	0.79	1.72	2740	518	2.64	9.35
mean	1.14	0.45	0.32	0.55	786	134	1.77	1.95
std. dev.	1.05	0.24	0.26	0.50	710	124	0.52	1.71
# ESUs	58	58	58	55	60	60	60	60

Sentinel-2 data

Sentinel-2 images acquired close to the date of the field campaigns were sought and downloaded from the Copernicus Open Access Hub (https://scihub.copernicus.eu/). There were, in total, four cloud-free images for the two growing seasons (Table 1), which was suitable to use as their acquisition dates were not further than 10 days from the field measurements. The data were downloaded as Level 1C product and pre-processed to Level 2A (i.e. top of canopy reflectance) in the Sen2Cor software (http://step.esa.int/main/third-party-plugins-2/sen2cor). The spectral data from the 10 m and 20 m bands were used in this study (Table 3). Sen2Cor provides the corrected bands B02, B03, B04, and B08 in two versions – at their native resolution of 10 m and at 20 m resolution. The reflectance values in the individual spectral bands were extracted for each ESU using a GIS tool and stored in spreadsheet for further processing.

The missing Sentinel-2 image data for two of the field campaigns (Table 1) were substituted with simulated data. Reflectance values for the Sentinel-2 spectral bandwidths were simulated at ESU level using data from the ASD FieldSpec4. The weighted average method was used for that purpose, where the published spectral responses of the Sentinel-2A bands were used as weights.

The spectral dataset, including the extracted Sentinel-2 image data and complemented with the simulated data for two of the field campaigns was used for calculation of VIs (Table 4). These indices were selected based on a previous literature review (Kamenova et al., 2017).

The aim of the regression modelling is to derive equation which quantitatively describes the relation between a variable of interest, y, (in this case, the crop variable) and a predictor variable, x, (in this case, VI or reflectance in a given spectral band). In this study, models with one predictor were only used. All potential predictors, i.e. the selected VIs (Table 4) and the Sentinel-2 bands were systematically tested for their predicting capabilities using linear and exponential regression models. The general form of the models is shown in (1):

$$y = a + b * x$$
$$y = a * \exp(x * b), \quad (1)$$

where a and b are parameters to be estimated. The regression parameters a and b were estimated by the least-squares method within the software package R (R Core Team, 2017).

We used the data from 2016/2017 growing season to calibrate the models. As mentioned previously, these data include only ESUs with little or no weeds, thus permitting relation to be established, which is valid for pure winter wheat crop and is not contaminated by the spectral effect of other vegetation. The data from 2017/2018 growing season were spared for independent validation of the models. In addition, a cross-validation following the leave-one-out method was performed using the 2016/2017 dataset. In both cases predictions were compared with observed data and Root Mean Square Error (RMSE) calculated using (2):

$$RMSE = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n}}, \quad (2)$$

where y_i is the measured value and $\hat{y}_i$ is the predicted value for the i-th observation. For the cross-validation, $\hat{y}_i$ is the predicted value derived from a model calibrated with all observations except i. The two statistics, corresponding to the cross-validation and the independent validation are denoted as $RMSE_{cv}$ and $RMSE_{iv}$ respectively. Their relative counterparts ($rRMSE_{cv}$ and $rRMSE_{iv}$) were calculated as percentage of the average of ground measurements from 2016/2017 and 2017/2018 growing season respectively.

Note that compared to the 2016/2017 growing season the number of observations in 2017/2018 growing season is smaller and later development stages (e.g. Anthesis) are not represented. Thus, in evaluating the relative performance of different VIs we did not put the emphasis on $RMSE_{iv}$ but used also $RMSE_{cv}$. Models were sought, which fit well the calibration data (2016/2017) and are stable when applied to new data (2017/2018).

Results and Discussions

The regression analysis was conducted, first, irrespective of the winter wheat variety, and then, separately for each of the two varieties. The results are presented in turn below.

Combined analysis

To facilitate the comparison of the models, $rRMSE_{cv}$ and $rRMSE_{iv}$ obtained using each VI were plotted together. Figure 2 shows example plots for the linear and the exponential models of fAPAR containing the ten best performing VIs in terms of $rRMSE_{cv}$. The VIs are sorted in increasing order of

Table 3. Spectral bands of Sentinel-2 used in this study (Drusch et al., 2012)

Band number	B02	B03	B04	B05	B06	B07	B08	B8A	B11	B12
Central Wavelength (nm)	490	560	665	705	740	783	842	865	1610	2190
Spatial Resolution (m)	10	10	10	20	20	20	10	20	20	20

Table 4. List of spectral vegetation indices used in the study and formulas for their calculation with Sentinel-2 bands

Vegetation Index	Reference
REP = 705 + 35 * (((B04 + B07) / 2 – B05) / (B06 – B05))	Clevers & Gitelson, 2013
MTCI = (B06 – B05) / (B05 – B04)	Dash & Curran, 2004
CI red edge B07 = (B07 / B05) – 1	Gitelson et al., 2003, 2006
CI green B07 = (B07 / B03) – 1	Gitelson et al., 2003, 2006
CI red edge B8A = (B8A / B05) – 1	
CI green B08 = (B08 / B03) – 1	
NDRE = (B06 – B05) / (B06 + B05)	Gitelson & Merzlyak, 1994, Sims & Gamon, 2002
NDRE 1 = (B07 – B05) / (B07 + B05)	Barnes et al., 2000
CCCI = ((B07 – B05) / (B07 + B05)) / ((B07 – B04) / (B07 + B04))	Barnes et al., 2000
$NDVI_{b7}$ = (B07 – B04) / (B07 + B04)	Rouse et al., 1973
$NDVI_{b8}$ = (B08 – B04) / (B08 + B04)	Rouse et al., 1973
SR = B08 / B04	Jordan, 1969
SR 1 = B06 / B04	
SR 2 = B07 / B04	
SR 3 = B07 / B06	Gitelson & Merzlyak, 1994
SR 4 = B07 / B05	
VARI = (B03 – B04) / (B03 + B04 – B02)	Gitelson et al., 2002
Red edge NDVI = (B8A –B06) / (B8A + B06)	Gitelson & Merzlyak, 1994
WDRVI = (0.3 * B08 – B04) / (0.3 * B08 + B04)	Gitelson, 2004
VIgreen = (B03 – B04) / (B03 + B04)	Gitelson et al., 2002
NDI = (B05 – B04) / (B05 + B04)	Delegido et al., 2011
GBM_Index2 = (B05 – B03) / (B05 + B03)	Hansen & Schjoerring, 2003
TCARI = 3 * [(B05 – B04) – 0.2 * (B05 – B03) * (B05 / B04)]	Haboudane et al., 2002
$TCARI/OSAVIre = \frac{3 * [(B06 – B05) – 0.2 * (B06 – B03) * (B06 / B05)]}{(1 + 0.16) * (B06 – B05)/(B06 + B05 + 0.16)}$	Wu et al., 2008
$MCARI/OSAVIre = \frac{[(B06 – B05) – 0.2 * (B06 – B03)] * (B06 / B05)}{(1 + 0.16) * (B06 – B05)/(B06 + B05 + 0.16)}$	Wu et al., 2008; Franceschini et al., 2017
MCARI = ((B05 – B04) – 0.2 * (B05 – B03)) * (B05 / B04)	Daughtry et al., 2000
$TCI = 1.2 * (B05 – B03) – 1.5 * (B04 – B03) * (B05 / B04)^{1/2}$	Haboudane et al., 2008
gNDVI = (B08 – B03) / (B08 + B03)	Gitelson et al., 1996
gNDVI 1 = (B06 – B03) / (B06 + B03)	
$MTVI2 = \frac{1.5 * [1.2 * (B8A – B03) – 2.5 * (B04 – B03)}{\sqrt{(2 * B8A + 1)^2 – (6 * B8A – 5\sqrt{B04}) – 0.5}}$	Haboudane et al., 2004
GIPVI = B08/ (B08 + B03)	
OSAVI = (1+0.16)*(B08 - B04) / (B08 + B04 + 0.16)	Rondeaux et al., 1996
DVI = B08 – B04	Jordan, 1969
EVI = 2.5(B08 − B04)/(B08 + 6 B04 – 7.5 B02 + 1)	Huete et al., 1997
NPCI = (B04 − B02)/(B04 + B02)	Peñuelas et al., 1994
MCARI/OSAVI = MCARI/OSAVI	Daughtry et al., 2000
TCARI/OSAVI = TCARI/OSAVI	Haboudane et al., 2002

$rRMSE_{cv}$ for better readability and ease of interpretation. This graphical representation of the data permitted to identify the models which have lowest $rRMSE_{cv}$ and in the same time to check if they are stable, i.e. if they have comparably low $rRMSE_{iv}$ as well. As can be seen, lowest $rRMSE_{cv}$ was obtained for the models utilizing GIPVI, both for the linear and the exponential regression. However, these models had much higher $rRMSE_{iv}$; this is true especially for the linear model (Figure 2a) where the error with the independent data is more than twice as high as that obtained with the cross-

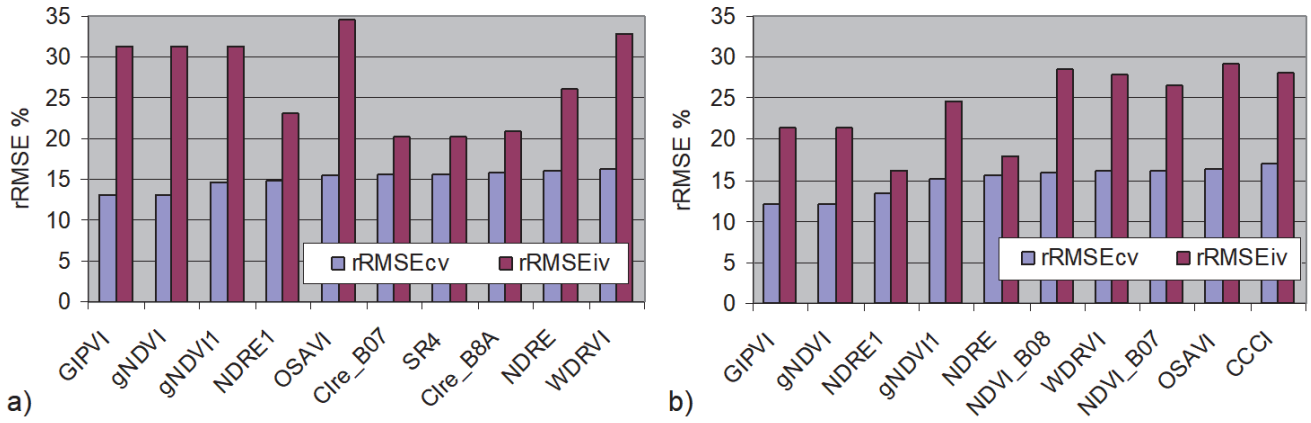

Fig. 2. Part of the validation results for the linear (*a*) and exponential (*b*) fAPAR models
Note: the ten VIs with lowest rRMSEcv (sorted in increasing order from left to right) are shown in each plot

validation. From the other hand, the exponential model with NDRE1 has slightly higher $rRMSE_{cv}$ than GIPVI, but this is compensated by low $rRMSE_{iv}$. Thus, the exponential model with NDRE1 is more stable and this comes at the expense of only slightly higher $rRMSE_{cv}$.

The plots for the other crop variable are not presented here due to the limited space. However, following the same reasoning as presented above, an "optimal" model was selected for each crop variable which both, fits the calibration data well and is stable with the independent data. The results are summarized in Table 5. It was possible to select regression models based on the criteria outlined above for all crop variables except for the AGBd. The most accurately predicted crop variable was fAPAR for which the relative errors were 13% and 16% (here, and in the following, the first figure represents $rRMSE_{cv}$ whereas the second represents $rRMSE_{iv}$). Relatively good were the results for fCover (20% and 28%), N content (20% and 27%), and AGBf (24% and 32%). Worst were the results for LAI (41% and 42%), CCC (33% and 41%), and N uptake (43% and 47%) (Table 5). Figure 3 shows the scatter plots of measured *versus* estimated crop variables. The correspondence between the measured and the estimated values is very good for fAPAR, fCover, and AGBf, with R^2 of up to ~0.90 (Figure 3h, Figure 3g, and Figure 3a). Also, the regression line was close to the 1:1 line, i.e. slope = 1 and intercept = 0. The correspondence of measured and estimated values of N content was weak (Figure 3d) with R^2 of only 0.59 for the independent validation. This is in contrast with the RMSE, which was relatively low.

For AGBd none of the developed models was found stable. The model which performed best during the cross-validation ($rRMSE_{cv}$ = 26%) resulted in extremely high error when applied on the independent data ($rRMSE_{iv}$ = 80%) (Table 5). As can be seen on Figure 3b this error is due to overestimation.

Figure 4 shows the relationships between the crop variables and the corresponding VIs used as predictors in the "optimal" models (Table 5). The regression lines represent the models

Table 5. Selected optimal regression models of the crop variables derived by combined analysis of both winter wheat varieties

Crop variable	Model type	Vegetation Index	Cross-validation		Independent validation	
			$RMSE_{cv}$	$rRMSE_{cv}$ (%)	$RMSE_{iv}$	$rRMSE_{iv}$ (%)
AGBf (g/m^2)	Linear	SR3	526	24.5	252	32.1
AGBd (g/m^2)	Linear	SR3	136	26.0	108	80.0
N uptake (g/m^2)	Exponential	CCCI	2.35	42.9	0.91	46.9
N content (%)	Linear	gNDVI	0.41	27.0	0.36	20.2
CCC (g/m^2)	Linear	SR3	0.48	33.1	0.22	40.6
LAI (m^2/m^2)	Linear	SR3	1.29	40.7	0.48	42.5
fCover	Exponential	NDRE1	0.13	20.0	0.09	28.0
fAPAR	Exponential	NDRE1	0.09	13.4	0.07	16.1

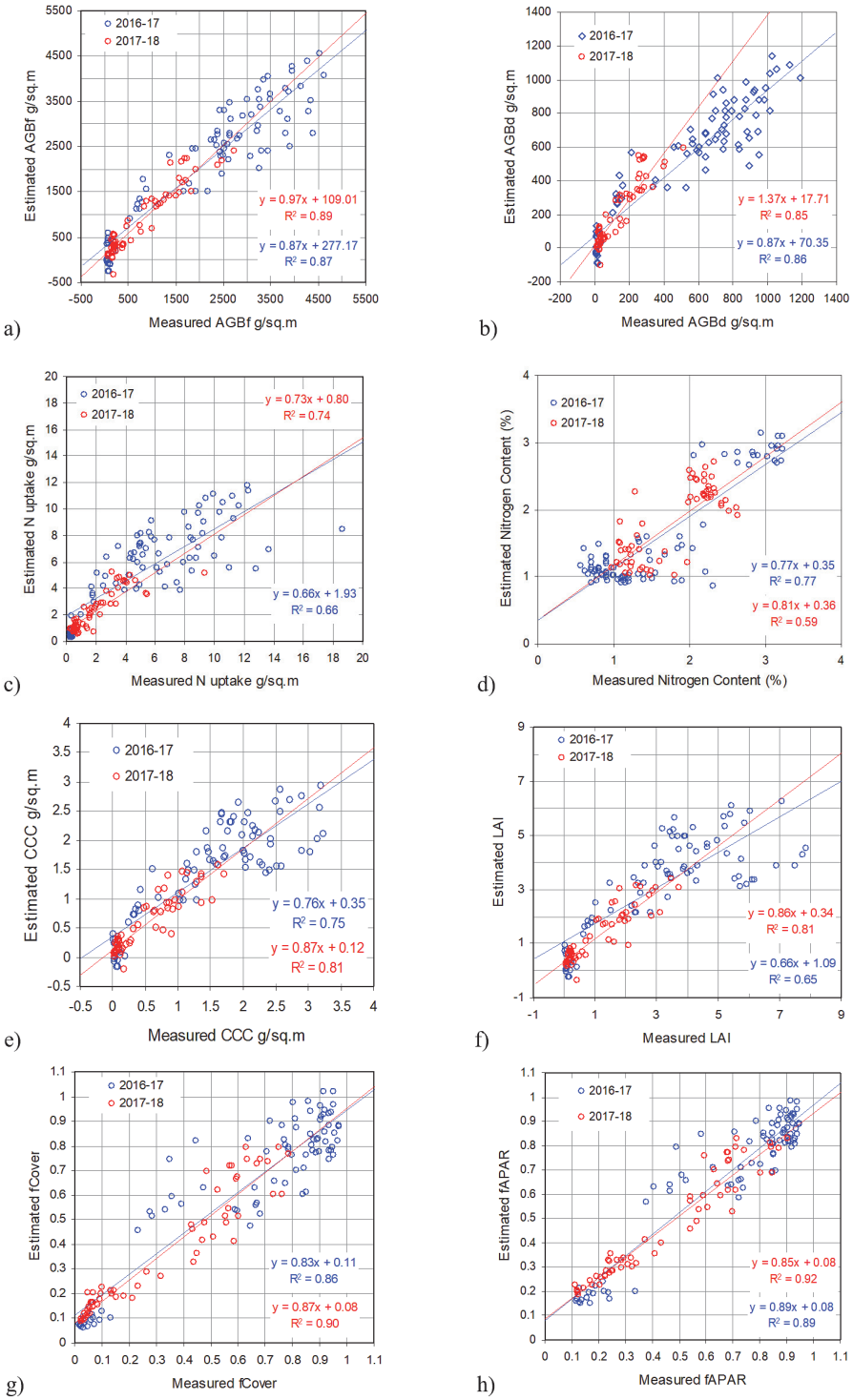

Fig. 3. Scatter plots of measured *versus* estimated values of crop variables

Note: the models used for estimation are listed in Table 5; blue dots show estimated values derived from the cross-validation procedure based on data from the first growing season (2016/2017); the red dots show the validation with the independent data from the second growing season (2017/2018)

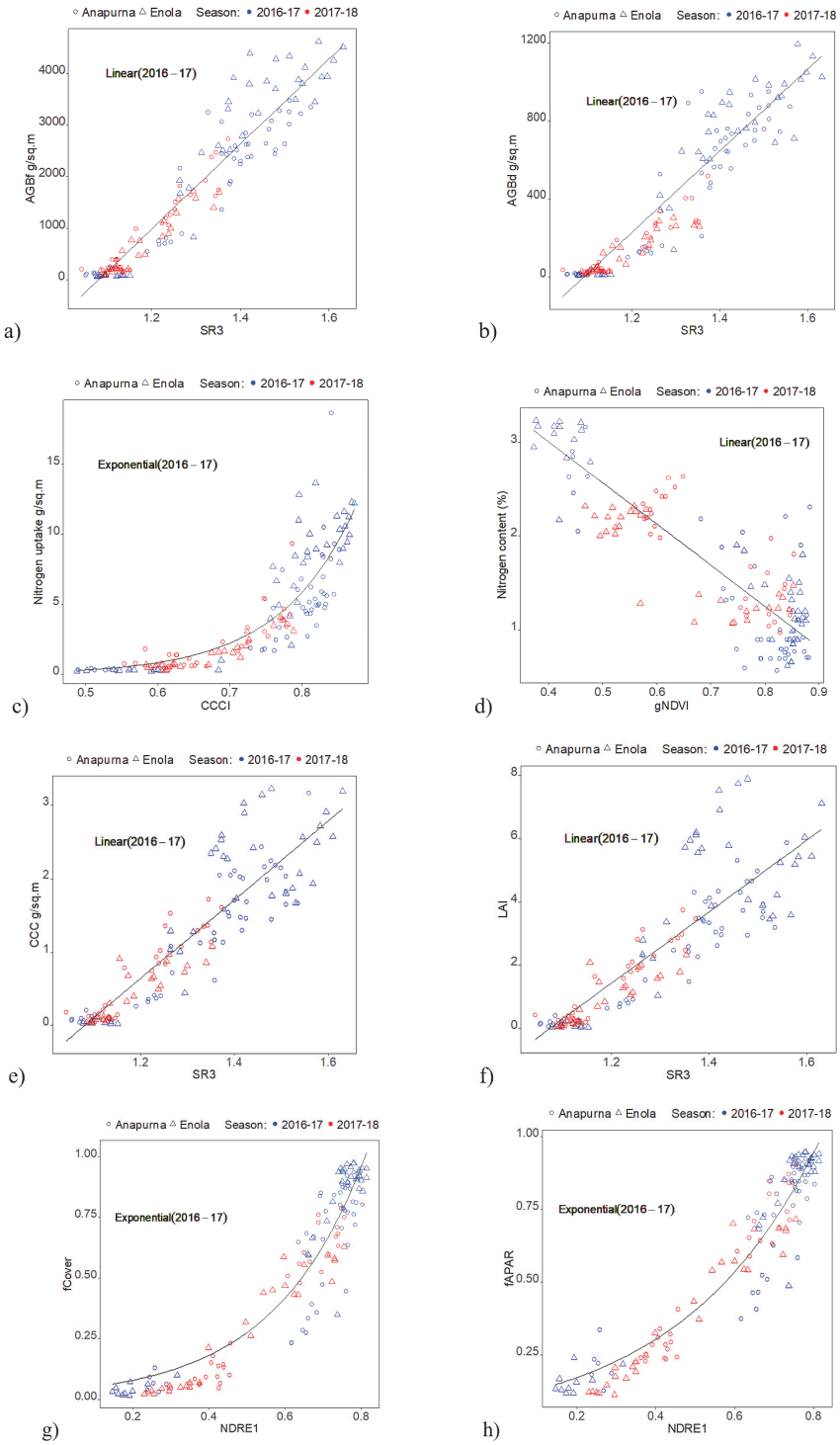

Fig. 4. Relationships between crop variables and the corresponding vegetation indices used as predictors for the models in Table 5

Note: the regression line represents the model which had been fit to the calibration data from 2016/2017 growing season

which had been fit to the calibration data of 2016/2017 growing season. The data from the 2017/2018 growing season are plotted in different colour to access how they are spread around the regression line and indicate possible instability of the models. Both, AGBf and AGBd are linearly related with the SR3 vegetation index (Figure 4a and Figure 4b). Note that, unlike the AGBf model, the AGBd model did not fit the 2017/2018 data well enough and most observations systematically deviate from the regression line. This led to the overestimation observed in Figure 3b and the instability of the model. The relationship of N uptake and CCCI is not linear and, thus, the index saturates at N uptake values over 5 g m^{-2} (Figure 4c). The relationship of N content and gNDVI (Figure 4d) is linear but is characterized with large scatter of data points. A clustered pattern is observed, which appears related with the individual field campaigns. More specifically, the first field campaign in each growing season (November 2017 and 2018, Tillering development stage; Table 1) is characterized with high N content (above ~2%) while in the later field campaigns the measured values were mostly below ~2%. The relationships of LAI and CCC with SR3 are linear (Figure 4e and Figure 4f), but the scatter around the regression line became very important with high values of the two variables. SR3 is not sensitive to LAI over ~3 m^2 m^{-2} and CCC over ~1.5 g m^{-2}. Note that the large scatter of data points observed above these values is due to the Enola variety (Figure 4e and Figure 4f). The relationships between NDRE1 and fAPAR and fCover are exponential (Figure 4g and Figure 4h). Both models were not very good in fitting the low fAPAR and fCover observations in the validation dataset (2017/2018) but performed well for higher values.

Analysis at variety level

The results discussed so far were based on regression models calibrated with data collected from both winter wheat varieties. These can be referred to as "combined models". However, evidence from the literature suggested that relationships between VIs and crop variables may be variety-specific. To test this, regression models were developed separately for each variety – Annapurna and Enola, and their errors compared with that of the combined models. The comparison based on the cross-validation errors is shown in Figure 5.

Figure 5 shows that for all but the N content and N uptake the accuracy of models increased when varieties are analyzed separately. The most substantial is the improvement with LAI where the combined model had $rRMSE_{cv}$ of 39% whereas the variety-specific models have $rRMSE_{cv}$ of 27%. Note that, the lowest $rRMSE_{cv}$ for the variety-specific models may have been obtained using different VI in each variety or using different VI compared to the combined model. This situation probably arise because, as observed in this study, several VIs may have very similar prediction capabilities and chance sampling variations may account for incident superiority of one or the other. For example, SR3 is the best predictor of AGBf in the combined model and in the Annapurna model. For Enola however, $rRMSE_{cv}$ of the model using SR3 is only slightly higher than that of another index - CI red edge B8A. Nevertheless, Figure 5 suggests that, for all but the N content and N uptake, it is feasible to use variety-specific regression models because they produced lower errors than the combined model. The results for N content and N

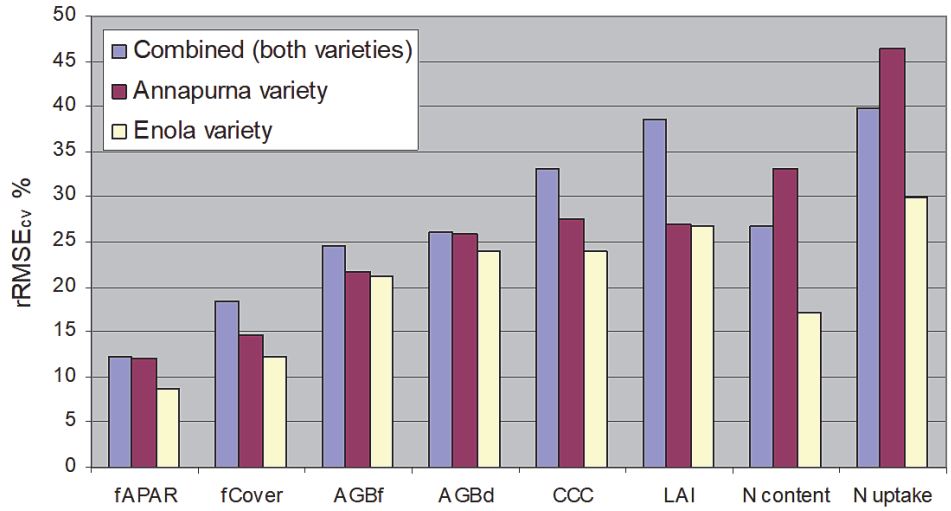

Fig. 5. Lowest $rRMSE_{cv}$ achieved for the two varieties, regardless of model type and predictor, compared with the lowest $rRMSE_{cv}$ achieved for the combined models

Table 6. Selected optimal regression models of the crop variables derived by analysis of winter wheat Annapurna variety

Crop variable	Model type	Vegetation Index	Cross-validation		Independent validation	
			$RMSE_{cv}$	$rRMSE_{cv}$ (%)	$RMSE_{iv}$	$rRMSE_{iv}$ (%)
AGBf (g/m^2)	Linear	SR3	424	21.6	263	27.6
AGBd	Exponential	CCCI	174	34.4	52	32.8
N uptake (g/m^2)	Linear	SR3	2.24	46.5	1.12	46.8
N content (%)	Linear	B12	0.44	33.1	0.37	20.2
CCC (g/m^2)	Linear	SR3	0.36	27.6	0.24	33.9
LAI	Linear	SR3	0.72	26.9	0.50	35.0
fCover	Exponential	NDRE1	0.14	22.9	0.08	22.0
fAPAR	Exponential	NDRE1	0.10	15.3	0.07	13.5

uptake are ambiguous because the error is lower for Enola variety but higher for Annapurna variety as compared with the corresponding combined model.

The above conclusion was drawn based on the cross-validation. However, as we saw, models calibrated with data from the 2016/2017 growing season may not perform well for the 2017/2018 growing season. Thus, we tested also the stability of the variety-specific models. Stable models were found for Annapurna variety and are shown in Table 6. These models were indeed more accurate than the "optimal" combined models (Table 5).

However, it was not possible to find stable models for Enola variety. The errors were very high when the models were applied to the 2017/2018 dataset ($rRMSE_{iv}$ > 50% for LAI, CCC, AGBd, and AGBf). Even for fAPAR, $rRMSE_{iv}$ is twice as high as the error from the cross-validation (22% and 11% respectively). These high errors are mostly related with overestimation. The scatter plots in Figure 6 demonstrate the overestimation problem in the Enola models of some of the crop variables. For example, AGBd and CCC from April 2018 are all spread below the regression line (Figure 6a and Figure 6b) and the model estimations are thus higher than the true values. For CCC, the linear model is inadequate to describe the relationship formed by the combined data of the two seasons. This is because instead of linear the relationship is curvilinear; this however was not evident in the calibration dataset (2016/2017). This is clear indication that the data used for model calibration were not sufficient to review the true nature of the relation. This is mostly due to the small number of observations from March 2017 for Enola variety. LAI shows clear nonlinear relation with gNDVI1 in the calibration dataset and this is confirmed by the validation dataset as well (Figure 6c). However, the validation data reviewed that the saturation starts earlier (at lower LAI), something which was not clear from the calibration dataset. For fAPAR the overestimation problem affects only the lower values (November 2017) (Figure 6d).

The general objective of this study was to evaluate different VIs for the remote estimation of several crop variables of winter wheat. With Sentinel-2 data available every 5 days (not considering clouds) opportunity exists now for monitoring winter wheat growth and status over large regions and with great detail. The challenge still remains how to extract relevant information from the satellite images, which to be meaningful to farmers and decision makers. The regression analysis based on VIs was used in literature for deriving quantitative crop characteristics from remotely sensed data because this method is simple, comprehensive and fast.

Few studies have reported results for crop variable estimation in winter wheat based on Sentinel-2 data. For example Delloye et al. (2018) used Radiative Transfer Model with an Artificial Neural Network and calculated CCC with RMSE of 0.35 g m^{-2} and relative RMSE of 26%. Our models (combined and Annapurna) had RMSE of 0.22-0.48 g m^{-2} (28-41%) depending on the validation strategy. The same authors derived Green Area Index (GAI), a parameter closely related with LAI with RMSE of 0.70 m^2 m^{-2} and relative RMSE of 26%. In this study, the combined model of LAI was of low accuracy (relative RMSE as high as 41-42%). The results were better for the Annapurna variety-specific model with RMSE of 0.50-0.72 m^2 m^{-2} (27-35%). Pan et al. (2018) evaluated four VIs for estimation of LAI through exponential regression and found that the models based on Red edge NDVI (defined as (B08-B06)/(B08+B06)) provided the most accurate predictions with RMSE of 0.85 m^2 m^{-2} (22%). Söderström et al. (2017) used combination of Sentinel-2 and DMC satellite data to predict N uptake as estimated by the handheld Yara NSensor. They used linear regression on MSAVI2 index and obtained mean absolute error (MAE) of 10-15 kgN ha^{-2} (i.e. 1.0-1.5 g m^{-2}), depending on the model and validation strategy. In our study the RMSE for the combined and the Annapurna model is between 0.91 and 2.35 g m^{-2}, depending on the validation strategy. Note however that MAE and RMSE are not comparable and, in the gen-

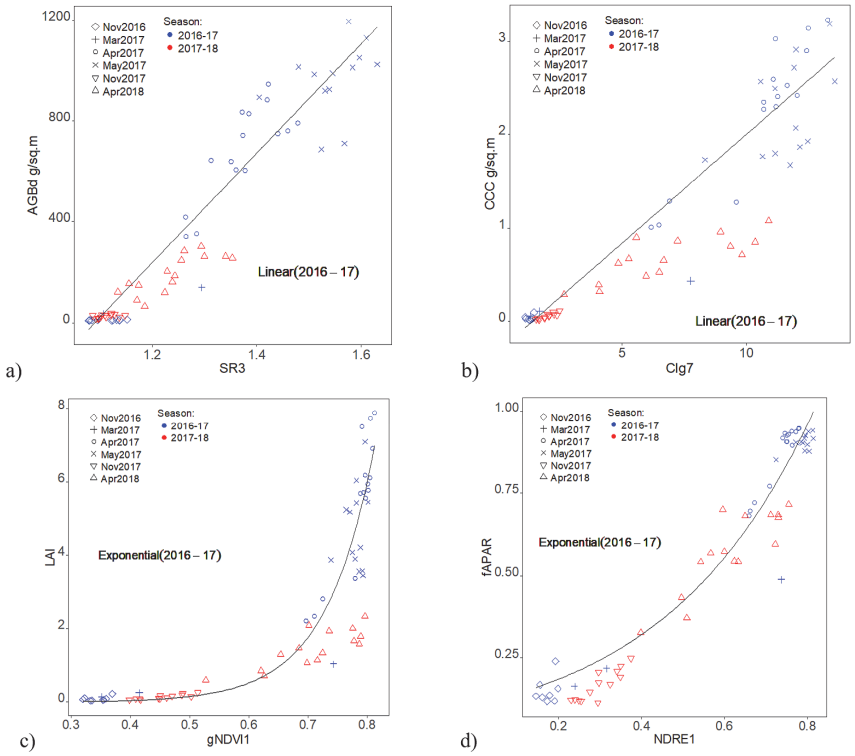

Fig. 6. Relationships between selected vegetation indices and crop variables for the winter wheat Enola variety: *a*) AGBd, *b*) CCC, *c*) LAI, and *d*) fAPAR

Note: the poor alignment of the validation data points (red colour) to the regression lines

eral case, MAE is lower than RMSE for any given dataset (Willmott & Matsuura, 2005). The potential of winter wheat biomass estimation based on Sentinel-2 data was studied by Zheng et al. (2017). Among 17 VIs they recommend the Red-edge Simple Ratio (NIR/Rededge) as a stable index for AGB estimation covering the entire growing season. This result is in line with our study where AGBf and AGBd were best predicted by SR3, which is the ratio of Sentinel-2 bands B07 and B06 (Table 4).

The possibility for estimation of N content by Sentinel-2 data is of particular interest for implementation of site-specific N fertilization systems. The errors for the proposed regression models (combined and Annapurna) were 0.36-0.44% of dry matter (20-33% of dataset mean). The errors are thus lower than those we had for LAI, CCC, and N uptake. The models, however, should be considered with care because of the high level of noise as indicated in Figure 3d and Figure 4d. As pointed out before, the pattern in these plots is partly explained with different N content measured in different field campaigns. Consistently higher N content was measured in the early season, before the winter dormancy period. This observation is in line with Mistele and Schmidhalter (2008) who point out that the N content of plants is highest at early growth stages and decreases continually up to the stage of senescence. This change is related with the combined effect of the differential tissue N content (higher content in some tissues and lower in other tissues) and the change in the proportion of different tissues with growth (increasing the proportion of N poor tissues). Thus, N content depends on the plant biomass (Mistele & Schmidhalter, 2008). The observed relation between VI and N content is therefore affected by biomass as well; an effect potentially introducing noise in the relation. Also, this study was conducted on intensively managed farmers' fields with N fertilizers applied several times during the season which is also a possible source of noise.

Part of the field data collected during the 2016/2017 field campaigns were not used in the regression analysis. For example ESUs with high percent cover of weeds were removed as well as those ESUs for which reliable field spectrometry data were not available due to bad weather conditions during the measurements. This reduced the number of observations available for model calibration. The problem was the most

severe for the March 2017 field campaign when data from 11 out of 30 ESUs (of which only 3 in Enola variety) were available. The effect of the insufficient data for Enola variety at the time of the March 2017 field campaign was significant and resulted in poorly calibrated models for that variety (Figure 6). As a consequence the models were not stable and performed poorly on the new data from the 2017/2018 growing season.

It should be noted also that the field campaigns were made within up to 10 days of the Sentinel-2 image acquisitions. Some errors in the models may have been introduced due to this temporal shift. During the 2016/2017 growing season only Sentinel-2A was operational and the frequency of image acquisitions was 10 days.

In this study, measurements from a single season were used for models calibration. It was shown that some models may not predict well data from other season. The empirical relations may differ not only due to crop variety and study site characteristics, but also from one year to the next (Mistele & Schmidhalter, 2008). Data collected across multiple years, sites and varieties may be needed to better describe relations between crop variables and VIs and to study the possibility for calibration of generally applicable regional models (Feng et al., 2015).

Conclusions

In this study we tested different VIs for their capability to predict several crop variables in winter wheat fields through linear and exponential regression. A combination of spectral data from Sentinel-2 images and field spectroscopy obtained through one growing season was used for model calibration and cross-validation. The models were further validated with Sentinel-2 image data from a second growing season. The main conclusions from this study can be summarised as follow:

The accuracy of the models varied widely across crop variables. According to the cross-validation, the relative RMSE was below 25% for fAPAR, fCover, and AGBf, with particularly good result for fAPAR ($rRMSE_{cv}$ = 13%). For N content and AGBd the error was between 25 and 30%. The accuracy was low for CCC, LAI, and N uptake ($rRMSE_{cv}$ between 30 and 43%).

In general, the models' performance was worse when they were applied to the data from the second growing season, resulting in $rRMSE_{iv}$ which were 3-8% higher than the $rRMSE_{cv}$ reported above. The model for AGBd was an extreme case where $rRMSE_{iv}$ reached 80%. It should be noted, however, that the validation dataset from the second season included only measurements from two field campaigns/development stages and thus, is not representative of the possible range of values of the crop variables.

The optimally performing models for fAPAR and fCover were the exponential models with NDRE1 while linear models using SR3 were best for predicting AGBf, AGBd, LAI, and CCC. The exponential model with CCCI was found the most useful for predicting N uptake. The N content was best predicted by linear model using gNDVI. Note however, that other VIs such as NDRE, CI red edge B07, CI red edge B8A, GIPVI, and gNDVI1 may also provide good results.

The cross-validation results suggested that the variety-specific models are more accurate than the generally calibrated models for most crop variables. This was confirmed, for Annapurna variety, also by the independent validation. The models for Enola variety were not successful in predicting the measurements from the second growing season but this is probably due to the insufficient number of observations available for calibration.

The results presented in this paper are preliminary. Though the models for fAPAR, fCover and AGBf, seem suitable for practical application due to their low error, further studies and more data will be needed to better account for variety, season, and site variations in the modelled relationships. This is necessary before models applicable at regional level are proposed, thoroughly validated and implemented in the practice. The first step will be to combine the data from the two growing seasons and re-calibrate the models. This is part of an ongoing research.

Acknowledgments

This study was conducted as part of the Testing Sentinel-2 vegetation indices for the assessment of the state of winter crops in Bulgaria (TS2AgroBg) project funded by the government of the Republic of Bulgaria through an ESA Contract (4000117474/16/NL/NDe) under the Plan for European Cooperating States.

References

Banerjee, V., Krishnan, P., Das, B., Verma, A. P. S., & Varghese, E. (2015). Crop Status Index as an indicator of wheat crop growth condition under abiotic stress situations. *Field Crops Research*, *181*, 16–31.

Barnes, E. M., Clarke, T. R., Richards, S. E., Colaizzi, P. D., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., Thompson, T., Lascano, R. J., Li, H., & Moran, M. S. (2000). Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data. In: *Proceedings of the Fifth International Conference on Precision Agriculture, Bloomington, MN, USA*. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.463.8007

Casanova, D., Epema, G. F., & Goudriaan, J. (1998). Monitoring rice reflectance at field level for estimating biomass and LAI. *Field Crops Research*, *55*, 83–92.

Chemura, A., Mutanga, O., Odindi, J., & Kutywayo, D. (2018). Mapping spatial variability of foliar nitrogen in coffee (Coffea arabica L.) plantations with multispectral Sentinel-2 MSI data. *ISPRS Journal of Photogrammetry and Remote Sensing, 138*, 1–11.

Clevers, J. G. P. W., & Gitelson, A. A. (2013). Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. *International Journal of Applied Earth Observation and Geoinformation, 23,* 344–351.

Clevers, J. G. P. W., Kooistra, L., & Van den Brande, M. M. M. (2017). Using Sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop. *Remote Sensing, 9,* 405. doi:10.3390/rs9050405.

Dash, J., & Curran, P. J. (2004). The MERIS terrestrial chlorophyll index. *International Journal of Remote Sensing, 25(23)*, 5403–5413.

Daughtry, C. S. T., Walthall, C. K., Kim, M. S., Brown de Colstoun, E., & McMurtrey, J. E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment, 74*, 229–239.

Delegido, J., Verrelst, J., Alonso, L., & Moreno, J. (2011). Evaluation of Sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. *Sensors, 11*(7), 7063–7081.

Delegido, J., Verrelst. J., Meza, C. M., Rivera, J. P., Alonso, L., & Moreno, J. (2013). A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems. *Europ. J. Agronomy, 46*, 42– 52.

Delloye, C., Weiss, M., & Defourny, P. (2018). Retrieval of the canopy chlorophyll content from Sentinel-2 spectral bands to estimate nitrogen uptake in intensive winter wheat cropping systems. *Remote Sensing of Environment, 216*, 245–261.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment, 120*, 25–36.

Feng, W., Guo, B., Zhang, H., He, L., Zhang, Y., Wang, Y., Zhu, Y., & Guo, T. (2015). Remote estimation of above ground nitrogen uptake during vegetative growth in winter wheat using hyperspectral red-edge ratio data. *Field Crops Research, 180,* 197–206.

Frampton, W. J., Dash, J., Watmough, G., & Milton, E. J. (2013). Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing, 82,* 83–92.

Franceschini, M. H. D., Bartholomeus, H., Van Apeldoorn, D., Suomalainen, J., & Lammert, K. (2017). Inter-comparison of unmanned aerial vehicle and ground-based narrow band spectrometers applied to crop trait monitoring in organic potato production. *Sensors, 17,* 1428. doi:10.3390/s17061428

Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *J. Plant Physiol., 161,* 165–173.

Gitelson, A. A., & Merzlyak, M. N. (1994). Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves - spectral features and relation to chlorophyll estimation. *J. Plant Physiol. 143*, 286–292.

Gitelson, A. A., Buschmann, C., & Lichtenthaler, H. K. (1999). The chlorophyll fluorescence ratio F735/F700 as an accurate measure of chlorophyll content in plants. *Remote Sensing of Environment, 69,* 296–302.

Gitelson, A. A., Gritz, Y., & Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology, 160*, 271–282.

Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ., 58*, 289–298.

Gitelson, A. A., Keydan, G. P., & Merzlyak, M. N. (2006). Three-band model for non-invasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters, 33*(11).

Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment, 80,* 76–87.

Haboudane, D., Tremblay, N., Miller, J. R., & Vigneault, P. (2008). Remote estimation of crop chlorophyll content using spectral indices derived from hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing, 46*, 423–437.

Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., & Strachan, I. B. (2004). Hyper-spectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: modelling and validation in the context of precision agriculture. *Remote Sens. Environ. 90,* 337–352.

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment, 81*, 416–426.

Hansen, P. M., & Schjoerring, J. K. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sens. Environ., 86*(4), 542-553.

Hinzman, L. D., Bauer, M. E., & Daughtry, C. S. T. (1986). Effects of Nitrogen Fertilization on Growth and Reflectance Characteristics of Winter Wheat. *Remote Sensing of Environment, 19*, 47–61.

Huete, A. R., Liu, H. Q., Batchily, K., & Yan-Leeuwen, W. (1997). A comparison of vegetation indices global set of TM images for EOS-MODIS. *Remote Sensing of Environment, 59*, 440–451.

Jordan, C. F. (1969). Derivation of leaf-area index from quality of light on the forest floor. *Ecology, 50*, 663–666.

Kamenova, I., Filchev, L., & Ilieva, I. (2017). Review of spectral vegetation indices and methods for estimation of crop bio physical variables. *Aerospace Research in Bulgaria, 29*, 72–82.

Kross, A., McNairn, H., Lapen, D., Sunohara, M., & Champagne, C. (2015). Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soy-

bean crops. *International Journal of Applied Earth Observation and Geoinformation, 34,* 235–248.

Lobell, D. B., Asner, G. P., Ortiz-Monasterio, J. I., & Benning, T. L. (2003). Remote sensing of regional crop production in the Yaqui Valley, Mexico: estimates and uncertainties. *Agriculture, Ecosystems and Environment, 94,* 205–220.

MAFF, Ministry of Agriculture, Food and Forestry. (2016). *Survey of land cover and land use in Bulgaria.* Ministry of Agriculture, Food and Forestry, Agrostatistics Department, Sofia.

Mistele, B., & Schmidhalter, U. (2008). Estimating the nitrogen nutrition index using spectral canopy reflectance measurements. *Europ. J. Agronomy, 29*(4)*,* 184–190.

Pan, H., Chen, Z., Ren, J., Li, H., & Wu, S. (2018). Modeling winter wheat leaf area index and canopy water content with three different approaches using Sentinel-2 multispectral instrument data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(2), 482-492.

Peñuelas, J., Gamon, J. A., Fredeen, A. L., Merino, J., & Field, C. B. (1994). Reflectance indices associated with physiological changes in nitrogen and water limited sunflower leaves. *Remote Sensing of Environment, 48*(2), 135–146.

R Core Team. (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment, 55(2)*, 95-107.

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. In: *Third ERTS Symposium.* NASA SP-351, 1, NASA, Washington, DC, 309–317.

Shang, J., Liu, J., Ma, B., Zhao, T., Jiao, X., Geng, X, Huffman, T., Kovacs, J. M., & Walters, D. (2015). Mapping spatial variability of crop growth conditions using RapidEye data in Northern Ontario, Canada. *Remote Sensing of Environment, 168*, 113–125.

Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment 81*, 337–354.

Söderström, M., Piikki, K., Stenberg, M., Stadig, H., & Martinsson, J. (2017). Producing nitrogen (N) uptake maps in winter wheat by combining proximal crop measurements with Sentinel-2 and DMC satellite images in a decision support system for farmers. *Acta Agriculturae Scandinavica, Section B - Soil & Plant Science*, *67*(7), 637-650.

Verrelst, J., Rivera, J. P., Veroustraete, F., Muñoz-Marí, J., Clevers, J. G. P. W., Camps-Valls, G., & Moreno, J. (2015). Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods - a comparison. *ISPRS Journal of Photogrammetry and Remote Sensing, 108*, 260–272.

Viña, A., Gitelson, A. A., Nguy-Robertson, A. L., & Peng, Y. (2011). Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sensing of Environment*, *115*(12), 3468-3478.

Wiegand, C. L., Maas, S. J., Aase, J. K., Hatfield, J. L., Pinter Jr., P. J., Jackson, R. D., Kanemasu, E. T., & Lapitan, R. L. (1992). Multisite analyses of spectral-biophysical data for wheat. *Remote Sensing of Environment, 42,* 1–21.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research, 30,* 79–82.

Wu, C., Niu, Z., Tang, Q., & Huang, W. (2008). Estimating chlorophyll content from hyperspectral vegetation indices: modelling and validation. *Agricultural and forest meteorology, 148*, 1230-1241

Zheng, Y., Wu, B. & Zhang, M. (2017). Estimating the above ground biomass of winter wheat using the Sentinel-2 data. *Yaogan Xuebao/Journal of Remote Sensing, 21*(2), 318–328 (Ch).

Received: June, 14, 2019; *Accepted:* July, 23, 2019; *Published:* October, 31, 2019