WHEAT BIOMASS ESTIMATION USING VISIBLE AERIAL IMAGES AND ARTIFICIAL NEURAL NETWORK

Mirayr Raul Quadros de Souza[1], Adriane Parraga[2], Marcelo Negreiros[3], Thiago Bertani[3], Carolina Trentin[4], Christian Bredemeier[4], Dionísio Doering[2], Altamiro Susin[5]

[1]Graduate Program in Electrical Engineering, UFRGS, Porto Alegre, Brazil

[2]Computer Engineering Department, UERGS, Guaiba, Brazil

[3]Department of Environment and Sustainability, UERGS, São Francisco de Paula, Brazil

[4]Department of Crop Science, UFRGS, Porto Alegre, Brazil.

[5]Department of Electrical Engineering, UFRGS, Porto Alegre, Brazil

{mirayr.raul, ddoering, marcelo.negreiros.rs}@gmail.com, adriane-parraga@uergs.edu.br, carolinatrentin@live.com, {bredemeier, altamiro.susin}@ufrgs.br

A paper from the Proceedings of the
14th International Conference on Precision Agriculture
June 24 – June 27, 2018
Montreal, Quebec, Canada

Abstract.

In this study, visible RGB-based vegetation indices (VIs) from UAV high spatial resolution (1.9 cm) remote sensing images were used for modeling shoot biomass of two Brazilian wheat varieties (TBIO Toruk and BRS Parrudo). The approach consists of a combination of Artificial Neural Network (ANN) with several Vegetation Indices to model the measured crop biomass at different growth stages. Several vegetation indices were implemented: NGRDI (Normalized Green-Red Difference Index), CIVE (Color Index of Vegetation Extraction), ExG (Excess green) SCOM (Simplified Combined Index) and a new index that we called ExRM (Excess red modified). An experiment containing 120 test plots was designed to assess wheat growth and test the vegetation indices (VIs) performance by correlating them with measured shoot dry biomass. Variability in crop growth was created for all test areas by varying nitrogen availability. For determining shoot biomass, plants were sampled at two different crop growth stages: V6 (stage of six fully developed leaves) and flowering. These measures were considered as the golden standard for the biomass model estimators. The images of the test areas were captured using an UAV flying at 50 meters above ground, a mosaic was created, and then the regions of interest were segmented. An ANN was trained to predict Biomass using Vegetation Indices as the features. We also compared the results with a linear regression to estimate shoot biomass from the VIs. The accuracy of the estimated model was evaluated based on the coefficient of determination (R^2). The best result for the cultivar BRS Parrudo was R^2=0.81 obtained using ANN and all VI as features versus biomass. For the cultivar TBIO Toruk the best result was R^2=0.86, modeling the biomass with four selected VIs. Our research results indicate that the proposed estimation model, based on RGB images and ANN, can be used in precision agriculture for predicting the spatial variability of shoot biomass, considering the two wheat cultivars tested.

Keywords.
Biomass estimation, wheat, UAV, Vegetation Index.

Introduction

Crop biomass estimation has been widely studied from different ways and cultivars in the last decade (Ballesteros 2018, Olschofsky 2016, Wang 2016). Considering the importance of constantly assessing crop biomass accumulation to predict grain yield and to delineate crop management practices, remote sensing images acquired with Unmanned Aerial Vehicles (UAVs) bring new opportunities to estimate photosynthetic activity and potential gross primary productivity systematically throughout the crop cycle.

UAVs are more flexible and inexpensive compared to its common alternatives, namely field measurement and satellite imagery. With respect to the latter option, it includes the possibilities of having higher image resolution and temporal resolution compared to satellite. These UAV imaging advantages can improve precision agriculture by measuring crop biomass as demonstrated by several publications. In (Bendig 2015) and (Brocks and Bareth 2018) a combination of selected vegetation indices (VIs) and plant height information was used to estimate biomass of summer barley. In (Ballesteros 2018) the estimation of onion crop biomass from high-resolution imaging obtained with UAV was based on green canopy cover, crop height and canopy volume as the predictor variables. The authors found strong correlation between canopy volume versus dry leaf biomass and canopy volume versus dry bulb biomass, with adjusted coefficient of determination values of 0.76 and 0.95, respectively. Messinger (2016) investigated the use of UAVs to measure aboveground carbon density in forests to replace expensive or labor-intensive inventory methods. The authors used a structure-from-motion (SFM) software approach to create a 3D model of the forest canopy and they found that the measured aboveground carbon density was highly correlated to previous LiDAR estimates, demonstrating that UAVs could be used as an alternative technique for local measurement.

Many works found in the literature use vegetation indices in an attempt to model shoot biomass, such as the normalized difference vegetation index (NDVI), which is computed from the red and near-infrared (NIR) bands (Guo 2018, Deng 2014). In (Ge 2016) it was demonstrated the characterization of temporal dynamics of plant growth and water use as well as leaf water content of two maize genotypes under two different water treatments. Hyperspectral images of plants were processed to extract plant leaf reflectance and correlated with leaf water content. Strong correlations were found between projected plant area and all three measured plant parameters (R2 > 0.95) at early growth stages

However, in order to evaluate near-infrared bands it is necessary to have multispectral images acquired with special infrared cameras which are more expensive than conventional RGB cameras. In this work, the potential of RGB imaging captured by an UAV and vegetation indices to model wheat shoot biomass at two different wheat-growing stages is analyzed. This work focuses on two different Brazilians wheat varieties, namely TBIO Toruk and BRS Parrudo, where the images from these crops were studied in conjunction with several vegetation indices evaluation. The challenge was to model shoot biomass of two kinds of Brazillians wheat cultivars throughout the crop cycle with variation of nitrogen fertilizer at the two different crop growth stages, i.e., at six fully developed leaves (V6) and at flowering. The goals of the present study were the following:

1. To evaluate several vegetation indices for wheat shoot biomass estimation;
2. To determine if a new index can increase correlation of the models;
3. To assess the potential of Visible RGB images for shoot biomass estimation;
4. To evaluate the use of an artificial neural network (ANN) to model shoot biomass for each of two Brazilian wheat varieties used in the present study.

This paper is structured as follows. The experimental setup is described in the section "Material and Methods" along with details of data acquisition and processing, including the vegetation indices, followed by a section were the experimental results are presented. The analysis and discussion of the results are presented in the "Discussion" section and the paper finishes with conclusions and further work.

Material and Methods

In this section, the experimental setup is described along with the associated techniques used for image acquisition and processing, as well as the vegetation indices used.Study Area. In this work an experimental wheat crop field with circa 100 m in length and 60 m in width was used as shown in Figure 1. The experiments were conducted at the Agriculture Experimental Station (51°40'32,788"W 30°6'40,678"S) of the Federal University of Rio Grande do Sul (EEA/UFRGS), with an altitude of 46 m, located in Eldorado do Sul (RS/Brazil).

The study area contains several 2.5m x 1.8m plots that hold three wheat varieties (Sossego, Toruk and Parrudo). Different nitrogen fertilizer rates were applied to the test areas during the experiment allowing for a large variability in crop growth. Biomass measurements were only available for test areas inside the marked zones in Figure 1 for the genotype TBIO Toruk (lines 8-10) and BRS Parrudo (lines1-3) wheat varieties.

The field experiment was conducted in 2017 in a typical dystrophic Red Argisol exhibiting the following physical and chemical characteristics in the 0-20cm layer: clay=250g dm^{-3}; pHwater=5.2; P=42mg dm^{-3}; K=202mg dm^{-3}; and organic matter=18g dm^{-3}. The climate of the region is classified as Cfa (subtropical with wet and hot summer). Treatments consisted of nitrogen rates applied at plant emergence (0, 15, 30, and 45kg ha^{-1}) and as topdressing (0, 20, 40, 60, and 80kg ha^{-1}) in the form of urea at the growth stage of six fully expanded leaves on the main stem. Different N rates were chosen to generate crop growth variability, in order to evaluate the response of vegetation indices and grain yield to N availability. The genotypes used were TBIO Toruk and BRS Parrudo. Sowing on maize straw was performed on July 10, 2017 at a density of 330 seeds m^{-2}.The experiment was conducted in a randomized block experimental design with split-plots and three replicates, being the main plots constituted by the N doses applied at plant emergence and the subplots by the topdressing N doses. Each plot comprised 10 rows of 2.5m in length, with row spacing of 0.18m, constituting 4.5m^2. Fertilizer rates at sowing were 60kg ha^{-1} of P_2O_5 and 90kg ha^{-1} of K_2O. Other crop management practices were performed according to recommendations for wheat in southern Brazil.

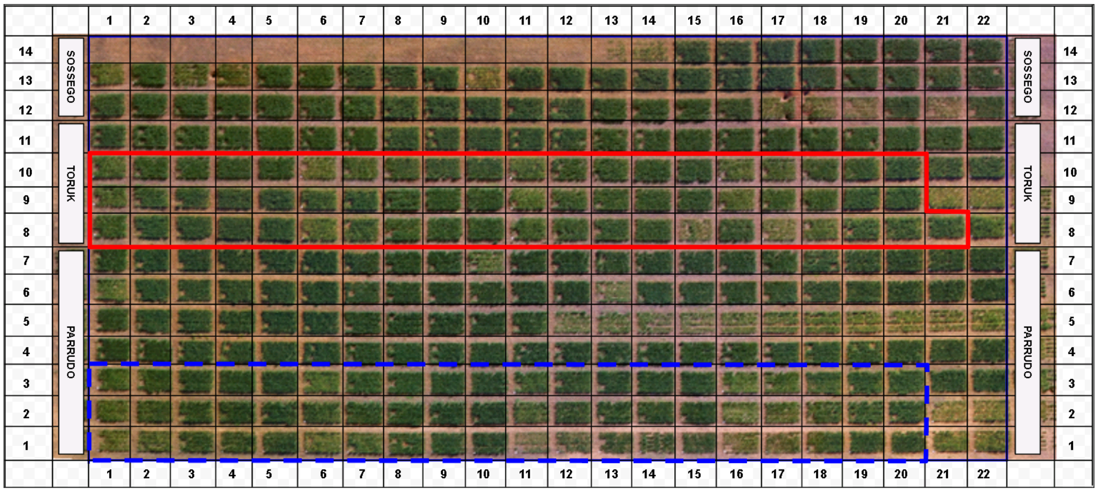


Fig 1. Experimental wheat crop field with annotated test plots used in this work. The TBIO Toruk wheat variety test area is marked in lines 8-10 and the BRS Parrudo wheat variety test area is marked in lines 1 to 3.

Biomass Measurement

Shoot dry biomass was determined at the stage of six fully expanded leaves (V6 growth stage) and at flowering by the collecting of plants in an area of 0.27 m². The plants collected were oven dried at 65°C until constant weight and weighed in a semi analytical balance.

Biomass measurements are presented in Table 1 along with number of plots sampled and average values obtained. The large biomass deviation value in the flowering stage is also related to different levels of fertilizer used in the experiment.

Table 1. Biomass measured dates and average values.

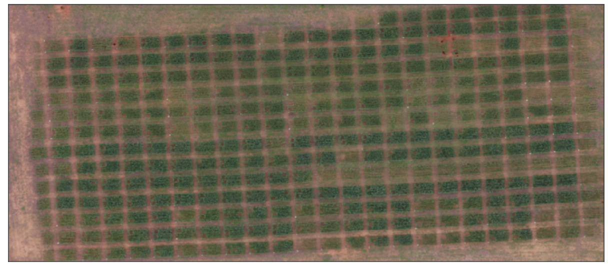
Biomass measured			Shoot biomass (kg/ha)			
			TBIO Toruk		BRS Parrudo	
Growth stage	Date	Sample size per wheat variety	Average	Standard Deviation	Average	Deviation
Six fully expanded leaves (V6 stage)	July 28, 2017	36	1579.7	189.6	1497.5	177.1
Flowering	September 12, 2017	60	5428.4	1122.6	4148.1	1133.0

UAV-based Data Collection

The images used in this work were acquired at a height of 50m aboveground using a camera coupled to a DJI Matrice 100 Quadcopter. This camera is a single channel DJI Zenmuse X3 Visible (RGB), with 12MB resolution and 8-bit pixel depth. The resulting pixel size for the image was 1.9 cm.

Post-processing of the acquired images included georeferencing and mosaicking using photogrammetry softwares, where a large set of overlapping images are post-processed to produce a single global orthoimage. An example of a single image of the study area can be seen in Figure 2 (a), and the resulting orthophoto after processing is presented in Figure 2(b).

(a)



(b)

Fig 2. Experimental test plots in September 13, 2017, one day after biomass sampling: (a) aerial image captured by UAV camera and (b) orthophoto obtained after image processing.

UAV-based image acquisition was performed in the dates shown in Table 2, within one-day interval from the biomass sampling date. Weather conditions were similar (sunny day and clear skies).

Table 2. UAV-based image acquisition.

Crop growth stage	UAV-based acquisition Date
Six fully expanded leaves (V6 stage)	July 27, 2017
Flowering	September 13, 2017

Image Pre-Processing

As described in the section UAV-based Data Collection, after the mosaicking of the images acquired by the UAV an orthophoto was generated. The ortophoto can then be used to derive parameters for each test plot. Since two different dates were used for biomass sampling, at least two orthophotos must be processed for each test plot.

In this work, the georeferenced orthophotos were co-registered, so that a region of interest (ROI) for every test plot used in the experiment can be defined. In this way, the ROI defined for images of different dates correspond to the same area of interest in the test plots.

From the co-registrated orthophotos, one can extract the mean value of the RGB channel from each test plot of the experiment. A region of interest was manually defined for each test plot in order to avoid areas with low plant density.

In this way, every test plot is linked to a single value of RGB data corresponding to the mean value of the test plot. The average data for each channel (red, green and blue) of each test plot is then used to calculate the corresponding vegetation indices as described in the following section.

Vegetation indices

In this section, we present a review of color based vegetation indices used in the literature and propose a new vegetation index.

According to Huete (2002), vegetation indices (VIs) are spectral transformations of two or more bands designed to enhance the contribution of vegetation properties and allow reliable spatial and temporal inter-comparisons of terrestrial photosynthetic activity and canopy structural variations. Vegetation indices can be considered as the basis of remote sensing in the analysis of natural and agricultural vegetation. An accurate vegetation index is required to identify plant biomass versus soil and residue backgrounds for automated remote sensing and machine vision applications, plant ecological assessments, precision crop management, and weed control (Meyer, 2008).

In this work, visible RGB-based vegetation indices (VIs) were evaluated from the orthophotos of the experimental site. For every test plot a zone was selected and the average value of the vegetation indices were calculated from the digital numbers of each visible channel. The vegetation indices used in this work are presented next.

The chromatic coordinates are a normalization introduced in the evaluation of color vegetation indices in Woebbecke (1993):

$$r = \frac{R^*}{R^*+G^*+B^*}\ , \quad g = \frac{G^*}{R^*+G^*+B^*}\ , \qquad b = \frac{B^*}{R^*+G^*+B^*} \tag{1}$$

Where

$$R^* = R/R_{max}, G^* = G/G_{max}, B^* = B/B_{max} \tag{2}$$

and R,G,B are the digital numbers from the RGB image for the channels red, green and blue, respectively. R_{max}, G_{max} and B_{max} are the maximum value for each channel.

When using 24-bit color images (8 bits per channel), the maximum value for all channels is the same so that Rmax=Gmax=Bmax=255 and then (1) can be rewritten as

$$r = \frac{R}{R+G+B}\ , \quad g = \frac{G}{R+G+B}\ , \quad b = \frac{B}{R+G+B} \tag{3}$$

Some authors have not applied any normalization and used the digital numbers from the image directly when evaluating vegetation indices, like Meyer (2008) and Kataoka(2003). In the present study, vegetation indices evaluated without normalization of the input channels are indicated by the suffix “Raw”.

The Excess Green Index or ExG (Woebbecke 1993; Hamuda 2016) is defined as in (4) and was used to separate plants from bare soil.

$$ExG = 2*g - r - b = \frac{2*G-R-B}{R+G+B} \tag{4}$$

The non-normalized version of the index is given by (5).

$$ExGRaw = 2*G - R - B \tag{5}$$

The Excess Red Index or ExR (Meyer 1999) was also used to separate plant pixels from background pixels, although not as accurate as ExG (Hamuda 2016). In the original paper a coefficient of value 1.3 was used (Meyer 1999; Hamuda 2016), however recent publications have used the value of 1.4 (Meyer 2004, Meyer 2008). The equation used in this work is presented in

(6).

$$ExR = 1.4 * r - g = \frac{1.4*R-G}{R+G+B} \quad (6)$$

The non-normalized version of the index is given by (7).

$$ExRRaw = 1.4 * R - G \quad (7)$$

The Excess Green minus Excess Red Index or ExGR (Meyer 2004, Hamuda 2016) was also used for separating plants from soil background, and it is given by (8).

$$ExGR = ExG - ExR = \frac{3*G-2.4*R-B}{R+G+B} \quad (8)$$

The non-normalized version of the index is given by (9).

$$ExGRRaw = ExGRaw - ExRaw = 3 * G - 2.4 * R - B \quad (9)$$

The Colour Index of Vegetation Extraction or CIVE was proposed by Kataoka (2003) in order to separate green plants from soil background. A normalized version is used in Guijarro (2011) as shown in (10).

$$CIVE = 0.441 * r - 0.811 * g + 0.385 * b + 18.78745 = \frac{0.441*R-0.811*G+0.385*B}{R+G+B} + 18.78745 \quad (10)$$

The non-normalized version of the index is given by (11).

$$CIVERaw = 0.441 * R - 0.811 * G + 0.385 * B + 18.78745 \quad (11)$$

The Normalized Green-Red Difference Index or NGRDI (Hunt 2005; Hamada 2016) was used in Hunt (2005) for estimating nutrient status of corn and crop biomass of corn, alfalfa, and soybeans. This index is sometimes referred as NDI (Meyer 2008).

$$NGRDI = \frac{G-R}{G+R} \quad (12)$$

In this paper, a simplified version of the combined index used in Guijarro (2011) is used, as described by (13)

$$SCOM = 0.25 * ExG + 0.30 * ExGR + 0.33 * CIVE \quad (13)$$

The non-normalized version of the simplified combined index is shown in (14).

$$SCOMRaw = 0.25 * ExGRaw + 0.30 * ExGRRaw + 0.33 * CIVERaw \quad (14)$$

In our work, a new index is proposed, called Excess Red Modified Index or ExRM, and its equation is presented in (15).

$$ExRM = 2 * r - g - b = \frac{2*R-G-B}{R+G+B} \quad (15)$$

The color based vegetation indices defined in this section will be used in order to model the biomass of the test plots in the following sections of this paper using linear regression and artificial neural networks.

Metodology

In this study, the use of RGB vegetation indices evaluated from images captured by an UAV in order to model wheat shoot biomass at two different growth stages is analyzed. Two different wheat varieties (TBIO Toruk and BRS Parrudo) were used in an experimental wheat crop field described in the section "Study Area" and the biomass of several test plots were sampled according to the procedures described in the section "Biomass Measurement". The sampling of biomass occurred at two different growing stages: at the stage of six fully developed leaves (V6) and at flowering. The measured biomass was then used as the desired output of the model, and the vegetation indices described in the section "Vegetation Indices" were used as inputs to the model. The block diagram of the modeling methodology is presented in Figure 3.

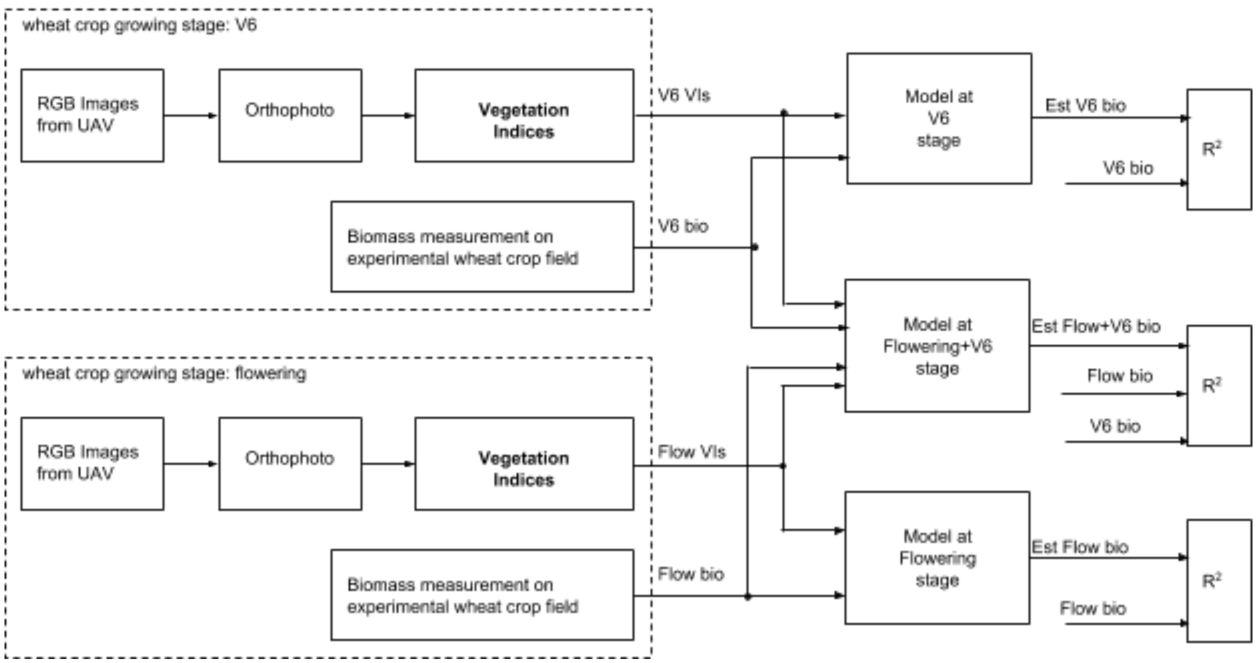


Fig. 3 - Block diagram of the modeling methodology

A model was constructed for data from each crop growing stage separately and for the whole data, so a model for V6, Flowering and V6+Flowering could be obtained. The models used linear interpolation and artificial neural networks, as will be described in the next sections.

In order to evaluate model quality, the coefficient of determination (R^2) was evaluated from the output of each model against the measured shoot biomass, for the corresponding vegetation indices. This way, both single input and multiple input modeling results were compared.

The coefficient of determination (R^2) was used to evaluate the quality and reliability of the estimate models for wheat biomass prediction. The coefficient of determination is a well-known statistic parameter and is often used to judge the adequacy of a regression model. According to Montgomery (2003), R^2 is often referring to the amount of variability in the data explained or accounted for by the regression model. The coefficient of determination R^2 is the square of the Pearson correlation coefficient between two variables. The Pearson correlation coefficient (R) is a measure of the strength and direction of the linear relationship between two variables.

Linear Regression Modeling

The first attempt to predict biomass using vegetation index data was to model the relationship between the measured biomass for each individual vegetation index through a linear model (y=ax+b), where "y" is the estimated or predicted biomass and "x" is the vegetation index. A simple linear regression was performed in MATLAB™ R2014a (MathWorks, Natick, MA, US) using the curve fitting tool and data from the different wheat genotypes (BRS Parrudo and TBIO Toruk) at all growing stages (V6, Flowering and combined V6+Flowering).

The linear regression modeling did not used any kind of data normalization. The approach evaluated only a single vegetation index and no multiple input modeling was attempted at this time. The linear regression results are presented in the Results section

Artifical Neural Network

Artificial Neural Networks (ANN) are computing systems vaguely inspired by the biological neural networks that constitute animal brains. The constituent units of this architecture are based on nodes or neurons weighted from their synaptic weights. The basic architecture consists of an input layer that is connected to the input data, an intermediate layer and an output layer. In the case of an ANN designed for function approximation, there will be only one neuron in the output layer. The input-output relationship can be mapped through a Euclidean space of dimension equal to the size of the input characteristics of the network in a Euclidean space of dimensions equal to the output layer (Haykin 2010).

In the supervised case, the ANN is trained with separate data set for training, usually with most of the data set available. Consequently, the ANN test and accuracy is performed with the separate test data that were not subjected to training in order to minimize the error between the desired and predicted output. It is expected that ANN will be able to predict as accurately as possible the values of the function in which it is approaching.

Machine learning and data mining algorithms are widely used in agricultural applications (Gandhi, 2016a). ANN was used in Gandhi (2016b) to estimate rice production and a vector with 7 dimensions was used at ANN input for prediction. The ANN was designed with the help of the WEKA tool (Waikato Environment for Knowledge Analysis).

A system to prevent drift of plant protection products with ANN algorithm was implemented in (Junior, 2017), so that ANN is intended to predict the appropriate amount of liquid product to be applied as a function of vegetation size. Prediction of ANN biomass were used for the sorghum crop in Zhang (2017), where the coefficient of determination was 0.62.

In Stas (2016), a comparison between machine learning algorithms was performed for wheat prediction. Using as input 76 characteristics related to chemical fertilizers, an ANN was used to choose which of these characteristics best represents the production of wheat. The output consists of NDVI data with a resolution of 1km in the evaluated period. In Kadir (2014) a data set such as sun, frost, rain and temperature was evaluated to predict wheat production.

As described in the "Image Pre-Processing" section, the mean value of the RGB channel for every test plot was used to calculate the vegetation indices. Theses vegetation indices were then used to model shoot biomass as described in the subsection "Metodology".

The modeling approach used an Artificial Neural Network of FeedForward Multilayer Perceptron. The python programming language was used to manipulate the RGB data, to calculate the vegetation indices and, with the help of the Keras (Chollet,2015) and sciKit libraries (Pedregosa, 2011), to use the Machine Learning algorithms. The vegetation indices evaluated were the following: exRM, exGRaw, exGRRaw, exGR, CIVERaw, CIVE, SCOM, SCOMRaw, exG e NGRDI, as described in the section "Vegetation Indices".

Biomass modeling was performed taking into account vegetation indices as described in the section "Metodology". The ANN architecture used an input layer with 10 neurons and three other layers with 15 neurons and an output layer with a single neuron.

The biomass measurement methodology consisted of collecting samples of 36 test plots of each cultivar at growing stage V6 and samples of all test plots of the two cultivars at flowering, that is 120 samples, as described in Table 1 in the section “Biomass Sampling”. Several biomass prediction models were developed and are described in the results and discussion section.

Results

In this section, the results obtained from the modeling approaches using linear regression and artificial neural networks are presented.

Linear Regression

Table 3 presents the linear regression results obtained for each different wheat cultivar (TBIO Toruk and BRS Parrudo) at different growing stages (V6 or Flowering). A total of 40 linear regression models were implemented and their coefficient of determination was evaluated and presented in Table 3. Table 4 presents the linear regression results obtained for each different wheat cultivar (TBIO Toruk or BRS Parrudo) but considering data from all growing stages (V6 + Flowering). Figure 4 presents the linear regression for the best result of Table 4 for Toruk cultivar.

Table 3. Coefficient of Determination (R^2) for linear model considering different Vegetation Indices for each wheat cultivar.

Cultivar	Stage	ExRM	ExGRaw	ExGRRaw	SCOMRaw	CIVERaw	CIVE	SCOM	ExGR	ExG	NGRDI
Toruk	V6	0.17	0.21	0.22	0.22	0.21	0.21	0.21	0.21	0.20	0.21
Toruk	Flow	0.26	0.12	0.26	0.24	0.16	0.23	0.24	0.25	0.22	0.26
Parrudo	V6	0.59	0.63	0.67	0.67	0.64	0.69	0.68	0.68	0.69	0.66
Parrudo	Flow	0.33	0.30	0.36	0.36	0.33	0.33	0.36	0.36	0.32	0.35

Table 4. Coefficient of Determination (R^2) for linear model

Cultivar	Stage	ExRM	ExGRaw	ExGRRaw	SCOMRaw	CIVERaw	CIVE	SCOM	ExGR	ExG	NGRDI
Toruk	All	0.76	0.70	0.21	0.37	0.66	0.46	0.24	0.14	0.50	0.16
Parrudo	All	0.73	0.31	0.14	0.03	0.21	0.00	0.11	0.20	0.02	0.55

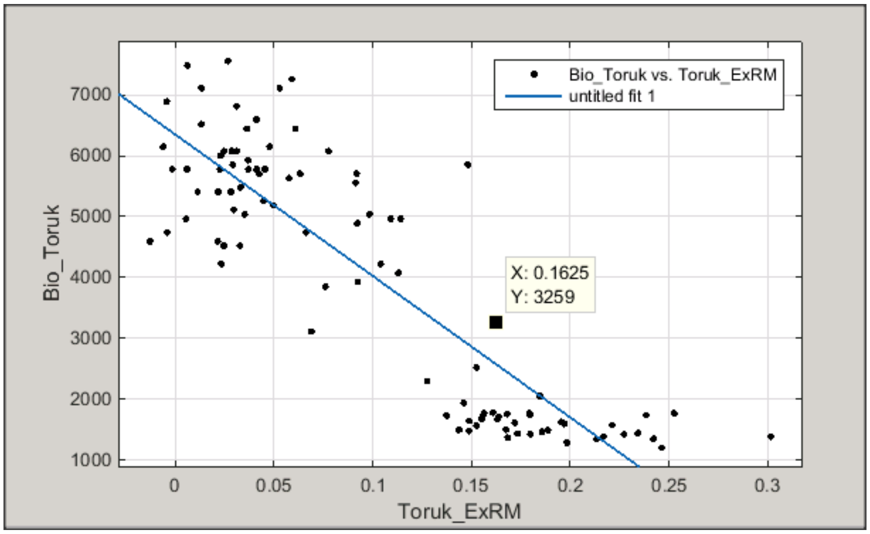


Fig. 4 – Scatterplot and Linear Model for shoot biomass of Toruk cultivar (axis y) versus Excess Red Modified Index – ExRM (axis x)

ANN Biomass Prediction results

The biomass prediction model developed using an Artificial Neural Network consists primarily in using the vegetation indices extracted at each growth stage for each cultivar as inputs to the ANN, as already described in Figure 3 in the Metodology section. Models were also developed using vegetation indices extracted in the two stages for the same cultivar. To evaluate the models, we use the determination coefficient R^2.

For each model, the total set of available samples was divided into 33% for testing and 67% for training. The coefficient of determination was calculated between the biomass data predicted by the model using ANN and the biomass measurements evaluated for each test plot. To calculate the coefficient of determination, we squared the Pearson correlation coefficient (R) generated by the Pearson function available in the scipy.stats Python library.

Table 5. Coefficient of Determination (R^2) for ANN models: (a) modeling using ANN for each cultivar on a specific growth stage and (b) modeling using ANN for each cultivar using data from both growth stages.

Cultivar	Stage	Number of Samples	R^2
Parrudo	V6	36	0.77
Toruk		36	0.12
Parrudo	Flowering	60	0.28
Toruk		60	0.26

(a)

Cultivar	Stage	Number of Samples	R^2
Parrudo	V6 + Flowering	96	0.81
Toruk	V6 + Flowering	96	0.86

(b)

The first model was constructed for predicting shoot biomass using all vegetation indices (exRM,exGRaw,exGRRaw,SCOMRaw,CIVERaw,CIVE,SCOM,exGR,exG,NGRDI) of a crop in the same growth stage. In Table 5(a), we can see that the best result $R^2 = 0.77$ is to use the vegetation indices for the BRS Parrudo cultivar in stage V6. On the other hand, the second model developed in the flowering stage for the same cultivar resulted in a very small R^2. The results of the coefficient of determination achieved for TBIO Toruk in this configuration is not acceptable because the value of R^2 in this case is very small and the model does not explain the observed values.

The third and fourth models developed took into account the vegetation indices of growth stage V6 and Flowering. Table 5(b) shows that the coefficient of determination increased significantly, i.e., $R^2 = 0.81$ and $R^2 = 0.86$ for BRS Parrudo and TBIO Toruk, respectively. The results obtained using the model in Table 5(b) for TBIO Toruk are presented in Figure 5.

From the results observed in Table 5, we can reason that the prediction model of biomass is more suitable when the indices of vegetation and biomass are modeled taking into account the two growth stages. Thus, the next test is performed with vegetation and biomass indices using data from both growth stages. Taking into account that the vegetation indices have different maximum, minimum and mean values, the standard normalization was performed for ANN to perform adequately (Haykin, 2010).

Table 6. Coefficient of Determination (R^2) for ANN models using a single vegetation index as input data.

Cultivar	exRM	exGRaw	exGRRaw	SCOMRaw	CIVERaw	CIVE	SCOM	exGR	exG	NGRDI
Parrudo	0.73	0.57	0.14	0.75	0.53	0.32	0.05	0.04	0.34	0.49
Toruk	0.86	0.84	0.56	0.82	0.86	0.73	0.51	0.50	0.73	0.12

Table 7. Coefficient of Determination (R^2) for ANN models using all vegetation indices as input data.

Cultivar	exRM, exGRaw, exGRRaw, SCOMRaw, CIVERaw, CIVE, SCOM, exGR, exG, NGRDI
Parrudo	0.81
Toruk	0.82

Table 8. Coefficient of Determination (R^2) for ANN models using normalized and non-normalized versions of only three vegetation indices.

Cultivar	Non-normalized vegetation indices (exGRaw, SCOMRaw, exGRRaw)	Normalized vegetation indices (exG, SCOM, exGR)
Parrudo	0.77	0.75
Toruk	0.86	0.82

In this second ANN modeling approach, data were first normalized according to (Nayak, 2014), so that the mean value of each class is zero and the standard deviation is unitary. This normalization consists of subtracting each sample from the mean and dividing it by the standard deviation of the corresponding class data set. This normalization prevents the ANN from being biased or from modulating the output data as a function of an input data class that has large numeric values in comparison to others.

The results generated by the ANN are shown in Tables 6, 7 and 8. First, prediction models were designed with a single vegetation index to characterize the biomass. The results of this test can be summarized in Table 6. The prediction model that had the highest value for BRS Parrudo was $R^2 = 0.73$ with the exRM vegetation index. For TBIO Toruk, R^2 was 0.86 when vegetation indices exRM or CIVERaw were used as input to the model.

As presented in Table 7, when all vegetation indices were used in the ANN input, the coefficient of determination was $R^2 = 0.81$ and $R^2 = 0.82$ for BRS Parrudo and TBIO Toruk.

The last test consists of developing two models using normalized and non-normalized vegetation indices as ANN inputs. For both BRS Parrudo and TBIO Toruk, the best value of R^2 was observed for non-normalized (raw) vegetation indices as presented in Table 8. This test shows that the three vegetation indices used to model biomass for TBIO Toruk have the same coefficient of determination when using only the exRM or CIVERaw vegetation indices used in the tests shown in Table 6.

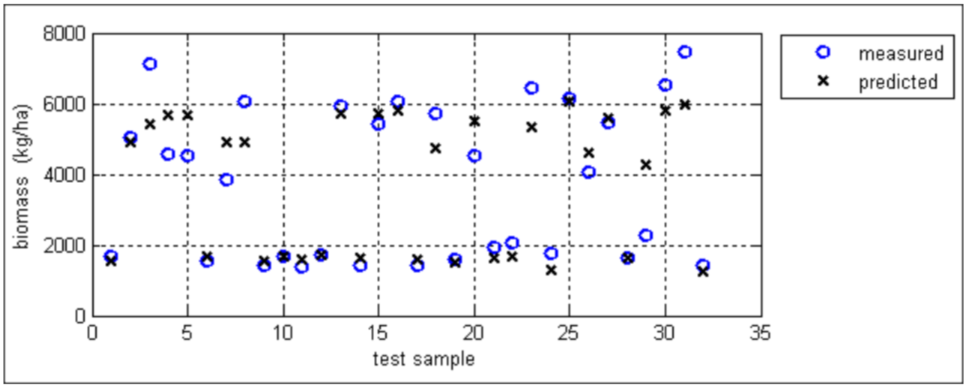


Fig. 5 Predicted and measured biomass for TBIO Toruk corresponding to the model presented in Table 5(b). Biomass values around 2000 kg/ha correspond to the growth stage V6.

Discussion

The first attempt to predict biomass using vegetation indices was to search for a linear relationship between biomass and each individual vegetation index. Table 3 presents the values of R^2 for both cultivars and for each growth stage. The results show that separating growth stage for each cultivar did produce a good fit between VI versus shoot biomass. However, when combining data for both growth stages as presented in Table 4, it is clear that there is a relationship between some vegetation indices and shoot biomass. Also, we found that the Excess Red Modified Index was the best index in this linear model. In Figure 4 the shoot biomass versus ExRM was presented. The negative slope in the plot is because the Excess Red Modified Index is based on the red channel. This could be explained by the fact that higher levels of biomass will have higher chlorophyll content per area and lower reflectance (greater absorption) in the red band.

In order to improve prediction accuracy it is proposed here to use machine learning for this task, since it should be able to find the best model (linear or non-linear) based on the data. Previous studies already used machine-learning algorithms, such as ANN, for remote estimations (Wang 2016).

As can be observed in Table 7, when using ANN to model biomass, it was possible to reach a coefficient of determination $R^2 = 0.81$ for the cultivar BRS Parrudo when all vegetation indices were used to train the model. On the other hand, the model constructed for the cultivar TBIO Toruk had a $R^2 = 0.86$ with only a single vegetation index of exRM or CIVERaw, as shown in Table 6. The same coefficient of determination value was reached when the features used in the training were exGRaw, exGRRaw, SCOMRaw (Table 8).

The value of coefficient of determination obtained by the ANN algorithm performed better when compared to a linear regression model with only one vegetation index. This is because the ANN is a more complex estimation method.

Conclusion

RGB image acquisition sensors are more commonly commercialized and less expensive, so predicting biomass using visible indices and vegetation is a challenge that can be useful for precision agriculture applications. The results presented here suggest that estimating wheat shoot biomass using aerial images are a promising alternative to other sensors and platforms, as they have high spatial and temporal resolution to perform biomass monitoring.

A non-linear model was implemented using artificial neural networks and several RBG vegetation indices at the input layer. The obtained results demonstrate the feasibility of the proposed approach to model shoot biomass for each of two Brazilian wheat varieties. The best results for an individual vegetation index was ExRM. Considering a combination of vegetation indices, the best results for TBIO Toruk was R^2=0.86 using the vegetation indices without normalization, and for BRS Parrudo was R^2=0.81 using all vegetation indices.

UAVs are able to capture data with high spatial as well as temporal resolutions by using visible and multispectral sensors. By doing the acquisition of high-resolution imagery, it is possible to model the gross primary productivity, perform site-specific fertilizer application, identify the presence of insects, weeds and others invasive plants. The information obtained from these images can then be used for crop management decisions.

Future research includes increasing the database to predict grain yield variability on the data based on aerial visible images and also improving the modelling approach developed in this paper.

Acknowledgements Authors would like to thank Fapergs (the Brazilian Foundation for Science and Technology) for funding the project under grant 16/2551-0000524-9.

References

Ballesteros R., Ortega, J. F., Hernandez D., Moreno M. A., (2018). Onion biomass monitoring using UAV-based RGB Imaging. *Journal of Precision Agriculture* https://doi.org/10.1007/s11119-018-9560-y.

Bendig, J., et. al. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Journal of International Journal of Applied Earth Observation and Geoinformation* 39 (2015) 79–87.

Brocks S, Bareth G. (2018). Estimating Barley Biomass with Crop Surface Models from Oblique RGB Imagery. *Journal of Remote Sensing*. 2018; 10(2):268.

Deng S., Katoh, M., Guan O., Yin N., Li, M. (2014). Estimating Forest Aboveground Biomass by Combining ALOS PALSAR and WorldView-2 Data: A Case Study at Purple Mountain National Park, Nanjing, China. *Journal of Remote Sens. 2014, 6, 7878-7910*; doi:10.3390/rs6097878.

Gandhi, N., & Armstrong, L. J. (2016, December). A review of the application of data mining techniques for decision making in agriculture. In *Contemporary Computing and Informatics (IC3I), 2016 2nd International Conference on* (pp. 1-6). IEEE, doi:10.1109/IC3I.2016.7917925

Gandhi, N., Petkar, O., & Armstrong, L. J. (2016, July). Rice crop yield prediction using artificial neural networks. In Technological Innovations in ICT for Agriculture and Rural Development (TIAR), 2016 IEEE (pp. 105-110). IEEE, doi:10.1109/TIAR.2016.7801222

Ge, Y, Bai,. G, Stoerger V., Schnable, J. C., (2016). Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput RGB and hyperspectral imaging. *Journal of Computers and Electronics in Agriculture 127 (2016) 6.*

Guo, C., Zhang, L., Zhou, X. et al. (2018). Integrating remote sensing information with crop model to monitor wheat growth and yield based on simulation zone partitionin. *Precision Agric (2018) 19: 55.* https://doi.org/10.1007/s11119-017-9498-5

Haykin, S. (2010). Neural networks: a comprehensive foundation, 1999. Mc Millan, New Jersey.ISBN 81-7808-300-0

Junior, C. R., Gomes, P. H., Mano, L. Y., de Oliveira, R. B., de Carvalho, A. C. D. L., & Faiçal, B. S. (2017, October). A Machine Learning-Based Approach for Prediction of Plant Protection Product Deposition. In *Intelligent Systems (BRACIS), 2017 Brazilian Conference on* (pp. 234-239). IEEE, doi: 10.1109/BRACIS.2017.26

Kadir, M. K. A., Ayob, M. Z., & Miniappan, N. (2014, August). Wheat yield prediction: Artificial neural network based approach. In Engineering Technology and Technopreneuship (ICE2T), 2014 4th International Conference on (pp. 161-165). IEEE, doi: 10.1109/ICE2T.2014.7006239

Li W., Yuan H., Song, L.(2016). Prediction of Wheat Gains with Imagery from Four-rotor UAV. . In Proceedings of the 2nd IEEE International Conference on Computer and Communications.

Messinger M., Asner G. P, and Silman M. (2016). Rapid Assessments of Amazon Forest Structure and Biomass Using Small Unmanned Aerial Systems. *Journal of Remote Sensing. 2016, 8, 615*; doi:10.3390/rs8080615*.*

Montgomery, D.C., Runger, G. C. (2003). Applied Statistics and Probability for Engineers pdf download Introduction to Probability and Statistics for Engineers and Scientists, Third Edition. New York: Wiley.

Natividade, J., Prado J. and Marques L. (2017), Low-cost Multi-spectral Vegetation Classification using an Unmanned Aerial Vehicle. In Proceedings of the IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC) April 26-28, Coimbra, Portugal.

Nayak, S. C., Misra, B. B., & Behera, H. S. (2014). Impact of data normalization on stock index forecasting. *Int. J. Comp. Inf. Syst. Ind. Manag. Appl*, *6*, 357-369.

Panda S. S., Ames, D. P., and Panigrahi S. (2010). Application of Vegetation Indices for Agricultural Crop Yield Prediction Using Neural Network Techniques. *Journal of Remote Sensing, 2, 673-696;* doi:10.3390/rs2030673*.*

Pedregosa, F. et al. (2011). “Scikit-learn: Machine Learning in Python”. In: Journal of Machine Learning Research 12, pp. 2825–2830.

Olschofsky, K., Mues, V., Köhl,M. (2016). Operational assessment of aboveground tree volume and biomass by terrestrial laser scanning. Computers and Electronics in Agriculture. *Journal of Computers and Electronics in Agriculture 127 (2016) 699–707.*

Stas, M., Van Orshoven, J., Dong, Q., Heremans, S., & Zhang, B. (2016, July). A comparison of machine learning algorithms for regional wheat yield prediction using NDVI time series of SPOT-VGT. In Agro-Geoinformatics (Agro-Geoinformatics), 2016 Fifth International Conference on (pp. 1-5). IEEE, doi:10.1109/Agro-Geoinformatics.2016.7577625

Zhang, Z., Masjedi, A., Zhao, J., & Crawford, M. M. (2017, July). Prediction of sorghum biomass based on image based features derived from time series of UAV images. In *Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International* (pp. 6154-6157). IEEE, doi: 10.1109/IGARSS.2017.8128413

Zermas D., Teng D, Stanitsas P., Bazakos M., Kaiser D. Morellas V. Mulla D. and Papanikolopoulos N. Automation Solutions for the Evaluation of Plant Health in Corn Fields (2015). In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) Congress Center Hamburg Sept 28 - Oct 2, 2015. Hamburg, Germany.

Wang L., Zhou, X., Zhu, X. Dong Z., Guo, W., (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *Journal of The crop journal 4 (2016) 212–219.*