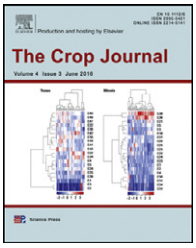
THE CROP JOURNAL 4 (2016) 212–219

HOSTED BY

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect



Estimation of biomass in wheat using random forest regression algorithm and remote sensing data

Li'ai Wang[a], *Xudong Zhou*[b], *Xinkai Zhu*[a], *Zhaodi Dong*[a], *Wenshan Guo*[a,*]

[a]*Key Laboratory of Crop Genetics and Physiology of Jiangsu Province, Yangzhou University, Yangzhou 225009, China*
[b]*Information Engineering College of Yangzhou University, Yangzhou 225009, China*

ARTICLE INFO

Article history:
Received 15 October 2015
Received in revised form
29 January 2016
Accepted 15 March 2016
Available online 30 March 2016

Keywords:
Above-ground dry biomass
Triticum aestivum
Vegetation indices
Wheat

ABSTRACT

Wheat biomass can be estimated using appropriate spectral vegetation indices. However, the accuracy of estimation should be further improved for on-farm crop management. Previous studies focused on developing vegetation indices, however limited research exists on modeling algorithms. The emerging Random Forest (RF) machine-learning algorithm is regarded as one of the most precise prediction methods for regression modeling. The objectives of this study were to (1) investigate the applicability of the RF regression algorithm for remotely estimating wheat biomass, (2) test the performance of the RF regression model, and (3) compare the performance of the RF algorithm with support vector regression (SVR) and artificial neural network (ANN) machine-learning algorithms for wheat biomass estimation. Single HJ-CCD images of wheat from test sites in Jiangsu province were obtained during the jointing, booting, and anthesis stages of growth. Fifteen vegetation indices were calculated based on these images. In-situ wheat above-ground dry biomass was measured during the HJ-CCD data acquisition. The results showed that the RF model produced more accurate estimates of wheat biomass than the SVR and ANN models at each stage, and its robustness is as good as SVR but better than ANN. The RF algorithm provides a useful exploratory and predictive tool for estimating wheat biomass on a large scale in Southern China.

1. Introduction

Biomass is one of the most useful indicators of crops vegetation development and health. Measuring biomass directly is a destructive and expensive procedure. More recent estimates are based on remotely sensed data, such as vegetation indices (VIs) [1–4]. Kross et al. [1] established relationships between corn biomass and VIs such as the NDVI (Normalized Difference Vegetation Index), Green-NDVI, RVI (Ratio Vegetation Index), and MTVI2 (Modified Triangular Vegetation Index 2) computed from the SPOT and Landsat images. Gnyp et al. [3] found that SAVI (Soil-Adjusted Vegetation Index), OSAVI (Optimized Soil-Adjusted Vegetation Index), and MTVI2 had stronger relationships with rice biomass at the jointing stage than that at booting. Gao et al. [4] proposed that maize biomass could be estimated by VIs

* *Corresponding author.*
E-mail addresses: wla001@163.com (L. Wang), guows@yzu.edu.cn (W. Guo).
Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

http://dx.doi.org/10.1016/j.cj.2016.01.008

calculated using Chinese environmental satellite (HJ) images [e.g. NDVI, RVI, and the enhanced vegetation index (EVI)]. Jin et al. [5] reported that the estimation accuracy of wheat biomass was better using a combination of VIs and radar polarimetric parameters (RPPs) than using VIs or RPPs alone.

Remote estimation of biomass requires application of diverse methods and techniques. In recent years machine-learning algorithms were trialed for ability to perform flexible input–output nonlinear mappings between remotely sensed data and biomass [6–8]. Typically, artificial neural networks (ANNs) and support vector regressions (SVRs) were employed to couple with VIs to build monitoring models with improved prediction accuracy of remote estimation of biomass in crops. For instance, Wang et al. [9] provided an effective model for assessing the biomass of wheat with ANNs and VIs (i.e. RVI, NDVI, GNDVI, SAVI, OSAVI, RDVI) calculated based on ASD FieldSpec data. Clevers et al. [10] estimated grassland biomass using SVRs and VIs such as the RVI, NDVI, WDVI, SAVI, GEMI (Global Environmental Monitoring Index), and EVI (Enhanced Vegetation Index) calculated based on ASD FieldSpec data.

Among various machine-learning algorithms, the emerging Random Forest (RF) algorithm proposed by Leo Breiman and Cutler Adele in 2001 has been regarded as one of the most precise prediction methods for classification and regression, as it can model complex interactions among input variables and is relatively robust in regard to outliers. The RF algorithm presents several advantages; it runs efficiently on large datasets, it is not sensitive to noise or over-fitting [11], it can handle thousands of input variables without variable deletion, and it has fewer parameters compared with that of other machine-learning algorithms (e.g. ANN or SVR). The RF classification algorithm has been applied to many remote sensing domains such as land cover classification [12–14] and other fields related to the environment and water resources [15–16]. To our knowledge, only a few studies have reported the use of the RF regression algorithm in remote sensing applications, including monitoring of forest growth, wetland vegetation, and water resources [6,17–18]. Furthermore, few studies have employed the RF regression algorithm based on VIs for estimating the biomass of winter wheat.

The major objectives of this study were to: (i) investigate the applicability of the RF regression algorithm in combination with VIs to remotely estimate wheat biomass, (ii) test the performance of RF regression for estimating biomass, and (iii) compare the performance of RF with that of other machine-learning algorithms for the estimation of wheat biomass. Specifically, based on VIs calculated from China's environmental satellite (HJ) charge-coupled device (CCD) images, we employed the RF algorithm to construct models to estimate wheat biomass, and then, the RF algorithm was compared with the SVR and ANN machine-learning algorithms in terms of accuracy, goodness of fit, and robustness for estimating wheat biomass.

2. Data source

2.1. Experimental design and data acquisition

Experiments were carried out in four counties (YiZheng, JiangYan, GaoYou and TaiXing) of Jiangsu province during the winter wheat growing seasons of 2010, 2011, 2012 and 2014. The local wheat cultivars were Yangmai 13, Yangmai 15, Yangmai 16, and Yangfumai 2. For each year's experiment, fifteen sample sites were established in each county and a plot of 30 × 30 m was randomly demarcated at each site. Within each plot, five subplots of 0.4 m × 0.4 m were established at least 10 m from each other. During three growth stages (jointing, booting and anthesis) wheat plants from each subplot (positions determined with a Global Positioning System GPS, Trimble GeoExplorer 2008 Series GeoXH, Trimble Navigation Limited, USA) at each site were collected, sealed in plastic bags, and sent to a laboratory for analysis. In the laboratory, the wheat plants from each subplot were dried in an oven at 80 °C for 48 h, after which the dry weight was determined. The dry weight was divided by the surface area of the subplot, and then the weight was converted to kg ha^{-1}. The biomass values of plants from the five subplots within each plot were averaged to represent the biomass of the entire plot.

For each stage, the pooled data from 2010, 2011, 2012 and 2014 were randomly divided into a training dataset and an independent test dataset (75% and 25% of the pooled data, respectively). For the training dataset, the number of samples was 174 at jointing, 174 at booting, and 147 at anthesis. For the test dataset, the number of samples was 58 at jointing, 58 at booting, and 49 at anthesis. The training dataset was used to establish models to predict biomass during each growth stage, and the test dataset was used to test the quality and reliability of each prediction model.

2.2. Remote sensing data and preprocessing

Remotely sensed data (HJ satellite charge-coupled device) of wheat from the three stages were retrieved online from the China Centre for Resources Satellite Data and Application (CRESDA). The HJ satellite charge-coupled device (HJ-CCD) satellite system is China's environmental disaster and environmental monitoring satellite system. It includes two optical satellites, HJ-1A and HJ-1B, which are symmetrically equipped with two CCD cameras. They comprise four multispectral bands with a 30-m resolution and a 720-km swath. The spectral ranges of the four bands are 430–520 nm (B_1-blue), 520–600 nm (B_2-green), 630–690 nm (B_3-red) and 760–900 nm (B_4-near infrared).

All HJ-CCD image data used in this study were completely corrected using ENVI4.7 remote sensing image processing software. Ground control points were located with a differential GPS unit during the field experiments. The map projection used a geographic coordinate system (Lat/Lon) as the projection type (WGS84) and a pixel size of 30 m × 30 m. A radiometric calibration was conducted using the HJ satellite calibration coefficients (e.g. gains and offsets). Atmospheric corrections were conducted using the MOTRAN 4 model embedded in the ENVI/FLAASH module of ENVI 4.7 software, and the input parameters were set based on the location, sensor type and ground weather conditions observed on the day each image was acquired. To improve the accuracy of pixel registration to within one pixel, coarse geometric corrections were made based on the 1:10,000 digitized raster map, after which, precise geometric corrections were made based on the GPS ground control points.

2.3. Vegetation indices

Vegetation indices (VIs) are usually used to quantify crop biomass. This study examined 15 VIs (Table 1) reported in literature to be well correlated with biomass. These VIs were calculated based on the four HJ-CCD bands.

3. Models and statistics

Based on the vegetation indices (VIs) in Table 1, RF, SVR and ANN were respectively used to remotely estimate wheat biomass during each growth stage. In each model, the vegetation indices were considered to be independent variables and biomass was the dependent variable.

3.1. Random forest regression algorithm (RF)

The RF regression algorithm is an ensemble-learning algorithm that combines a large set of regression trees. A regression tree represents a set of conditions or restrictions that are hierarchically organized and successively applied from a root to a leaf of the tree [34–36]. The RF begins with many bootstrap samples that are drawn randomly with replacement from the original training dataset. A regression tree is fitted to each of the bootstrap samples. For each node per tree, a small set of input variables selected from the total set is randomly considered for binary partitioning. The regression tree splitting criterion is based on choosing the input variable with the lowest Gini Index, i.e. $I_G(t_{X(x_i)}) = 1 - \sum_{j=1}^{m} f(t_{X(x_i)}, j)^2$, where $f(t_{X(x_i)}, j)$ is the proportion of samples with the value x_i belonging to leave j as node t [36]. The predicted value of an observation is calculated by averaging over all the trees. Two parameters need to be optimized in the RF: the number of regression trees (*ntree*; default value is 500 trees) and the number of input variables per node (*mtry*; default value is 1/3 of the total number of variables).

To model the relationship between VIs and wheat biomass in this study, given the set of training input–output (i.e. VIs–biomass) pairs, the RF regression model was conducted as follows:

1) *ntree* bootstrap sample sets, i.e. X_i (i = bootstrap iteration, and its value was limited to the range of [1, *ntree*]), were randomly drawn with replacement from the original training dataset. The elements not included in X_i are referred to as out-of-bag data (OOB) for that bootstrap sample set.
2) At each node per tree, *mtry* vegetation indices were randomly selected from all 15 vegetation indices and the best split from among those indices was chosen according the lowest Gini Index.
3) For each tree, the data splitting process in each internal node of a rule was repeated from the root node until a previously specified stop condition was reached.

For the three stages, the parameter values (*ntree* and *mtry*) were optimized using the training dataset and RMSE to find the values that could best predict the wheat biomass. For each stage, *ntree* values from 1000 to 9000 with intervals of length 1000 were tested [37–41], and *mtry* was tested from 3 to 10 (Fig. 1). The *ntree* and *mtry* values that yielded the lowest RMSE were selected. According to Fig. 1, the values of *ntree* and *mtry* were 1000 and 3 at jointing and booting, respectively, and 3000 and 9 at anthesis.

3.2. Support vector regression (SVR)

The Support Vector Machine (SVM) was originally used for classification problems, i.e. support vector classification (SVC) and was then extended for use with regression problems, i.e. namely support vector regression (SVR) [42]. The quality of the SVR models depends on a proper setting of the SVR meta-parameters, the loss function ε and the error penalty factor C. In addition, selection of the kernel function has an important

Table 1 – Formulas of remote sensing vegetation indices.

Acronym	Index	Formula	Reference
NDVI	Normalized Difference Vegetation Index	$(R_{NIR}–R_R)/(R_{NIR} + R_R)$	[19]
SAVI	Soil-Adjusted Vegetation Index	$(R_{NIR}–R_R) / (R_{NIR} + R_R + 0.5) \times 1.5$	[20]
OSAVI	Optimized Soil-Adjusted Vegetation Index	$(R_{NIR}–R_R) / (R_{NIR} + R_R + 1.6) \times 1.16$	[21]
NRI	Nitrogen Reflectance Index	$(R_G–R_R) / (R_G + R_R)$	[22]
GNDVI	Green-NDVI	$(R_{NIR}–R_G) / (R_{NIR} + R_G)$	[23]
SIPI	Structure Insensitive Pigment Index	$(R_{NIR}–R_B) / (R_{NIR} + R_B)$	[24]
PSRI	Plant Senescence Reflectance Index	$(R_R–R_B) / R_{NIR}$	[25]
RVI	Ratio Vegetation Index	R_{NIR} / R_R	[26]
CRI	*Carotenoid* Reflectance Index	$1/R_G + 1/R_{NIR}$	[27]
EVI	Enhanced Vegetation Index	$2.5 \times (R_{NIR} - R_R) / (1 + R_{NIR} + 6 \times R_R - 7.5 \times R_B)$	[28]
MSR	Modified Simple Ratio Index	$((R_{NIR} / R_R) - 1) / \sqrt{R_{NIR}/R_R + 1}$	[29]
NLI	Nonlinear Vegetation Index	$(R_{NIR} \times R_{NIR} - R_R) / (R_{NIR} \times R_{NIR} + R_R)$	[30]
RDVI	Re-normalized Difference Vegetation Index	$(R_{NIR} - R_R) / \sqrt{R_{NIR} + R_R}$	[31]
TVI	Transformational Vegetation Index	$\sqrt{NDVI + 0.5}$	[32]
MTVI2	Modified Triangular Vegetation Index 2	$1.5 \times [1.2 \times (R_{NIR} - R_G) - 2.5 \times (R_R - R_G)] / \left[\sqrt{2 \times (R_{NIR} + 1)^2 – 6 \times R_{NIR} + 5 \times \sqrt{R_R}} – 0.5\right]$	[33]

R_i denotes reflectance at band i (nanometer); R_B represents reflectance of the blue band of HJ-CCD; R_G represents reflectance of the green band of HJ-CCD; R_R represents reflectance of the red band of HJ-CCD; R_{NIR} represents reflectance of near infrared band of HJ-CCD.

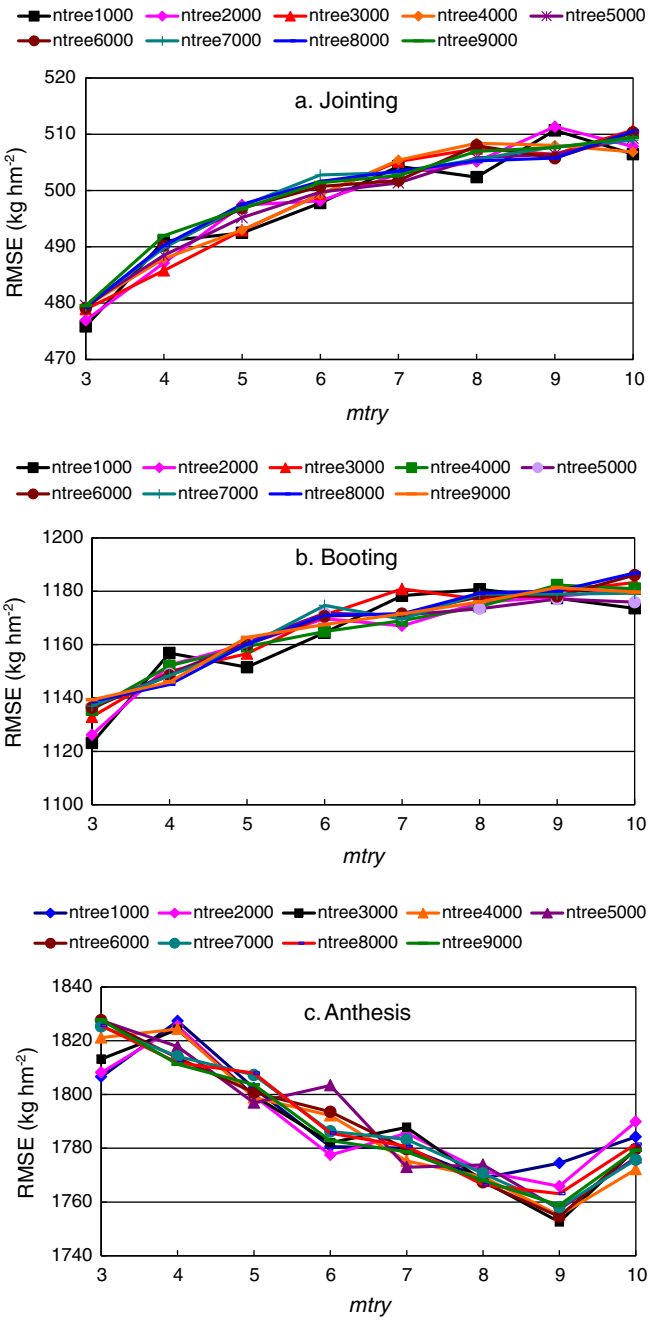


Fig. 1 – Optimization of random forest parameters (*ntree* and *mtry*) using RMSE.

impact on the final models. The commonly used radial basis kernel function (RBF), i.e. $K(x,x') = \exp(-||x-x'2/\sigma^2)$ was applied in this study. Finally, we employed a cross-validation procedure to optimize these parameters including C, ε, and the RBF kernel parameter σ, yielding values of 30, 470 and 2.5 at jointing, 5, 400 and 1.1 at booting, and 5, 850 and 8 at anthesis, respectively.

3.3. Artificial neural network (ANN)

Among various machine-learning algorithms, artificial neural networks (ANNs) are the most common approaches to develop nonlinear regression [43]. Training an ANN needs selections including the network structure (i.e. the number of hidden layers and nodes per layer), proper initialization of the weights, learning rate, and training algorithm. In this work, the input layer was vegetation indices, and the output layer was wheat biomass. We optimized a two-layer back propagation neural network (BPNN) with tan-sigmoid (i.e. $f(x) = \frac{2}{1+e^{-2x}} - 1$) hidden neurons and log-sigmoid (i.e. $g(x) = \frac{1}{1+e^{-x}}$) output neurons using the Levenberg–Marquardt algorithm. The ANN weights were initialized randomly according to the Nguyen–Widrow method [44]. Meanwhile, a cross-validation procedure was employed to set the number nodes per layer (i.e. 67 at jointing and booting, and 49 at anthesis, respectively).

3.4. Statistical analysis

Regarding model performances in this study, we used the coefficient of determination (R^2) to account for goodness-of-fit, and the root mean square error (RMSE) and relative RMSE (%) to assess accuracy. The relative RMSE was used to compare performances across different machine-learning algorithms [44]. Generally, the performance of the model was estimated by comparing the differences in R^2 and RMSE of the estimated-versus-measured value plots. Higher R^2 and lower RMSE values, respectively, corresponded to higher precision and accuracy of a model for predicting wheat biomass.

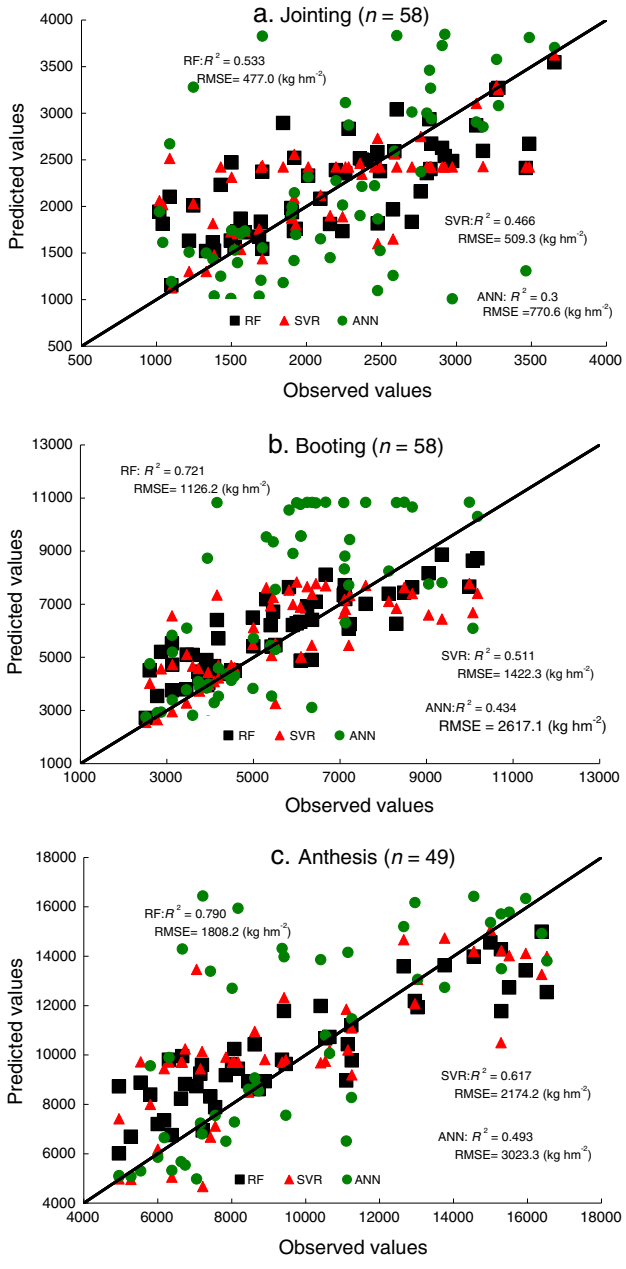


Fig. 2 – One-to-one relationships between predicted and observed biomass values.

4. Results

4.1. Evaluating model accuracy

Using the R^2 and RMSE values as metrics the performances of all models at each stage were evaluated with the test data from the corresponding stages and compared to identify the best model. For each stage, the R^2 and RMSE values between estimated (using the RF, SVR or ANN model) and measured biomass values were compared by means of scatter plots (Fig. 2). The performance of the RF model shows an overall improvement compared to that of the SVR and ANN models. Compared with SVR, the RMSE of the RF model decreased to 32.3 kg ha^{-1} at jointing, 296.1 kg ha^{-1} at booting, and 366.0 kg ha^{-1} at anthesis, and the corresponding R^2 values increased to 0.067, 0.210 and 0.173; compared with ANN, the RMSE decreased to 293.6, 1490.9 and 1215.1 kg ha^{-1} at each stage, and R^2 increased by 0.233, 0.287 and 0.297, respectively.

4.2. Evaluating model robustness

Relative RMSE results of the three regression methods for biomass at different growth stages are presented in Fig. 3. The error bars provide an idea of model robustness with respect to the input data. Different stages hardly impact the RF and SVR models performance in training or testing datasets. For each stage, the relative RMSE for the RF and SVR models respectively stabilize around 8% in the training and 20% in testing datasets. Regarding ANN, the performance of the training dataset was also robust at all three stages with the relative RMSE about 4%, but it performed unstably when applied to testing dataset. Specifically, the relative RMSE is about 35% at jointing, about 45% at booting, and about 30% at anthesis. For each model at each stage, the performance in testing is poorer than in the training dataset. ANN, in particular, showed a much better performance in training than in testing. Hence, in further analysis it will be important to determine how accurately a trained model performs when tested against ground reference measurements rather than the training data [44].

5. Discussion

The objective of this study was to employ accurate and robust random forest (RF) machine-learning algorithms to accurately estimate wheat biomass. Previous studies already used machine-learning algorithms such as SVR or ANN for remote estimations of biomass [6–10]. It remains however to be questioned whether these are the most adequate algorithms to fulfill the requirement. This study compared RF with SVR and ANN for accuracy and robustness.

By analyzing the estimated-versus-measured values (Fig. 2) the RF model had higher R^2 and lower RMSE values than the SVR and ANN models for biomass estimates at each growth stage, indicating that RF models can provide accurate biomass estimations. Each node of the standard regression tree is created using the best split among all variables. Unlike this strategy, RF splits each node using the best among a

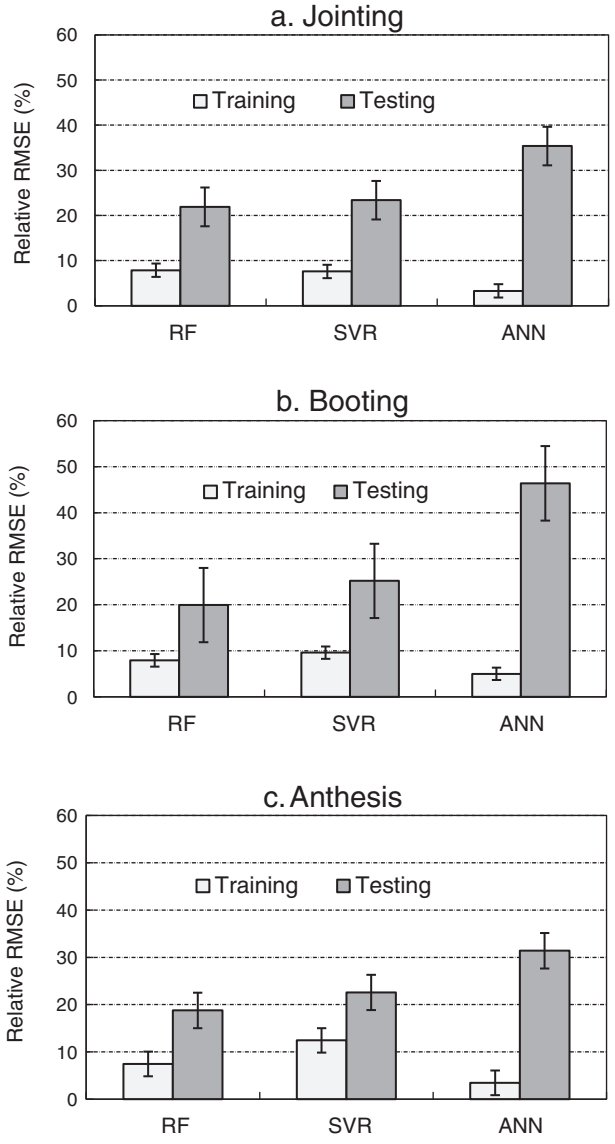


Fig. 3 – Relative RMSE (%) results for biomass estimation using RF, SVR, and ANN at different growth stages.

subset of variables chosen randomly at the node. The specific size of the subset is the parameter *mtry*. Although this method seems to be contradictory, it performs relatively well compared to SVR and ANN (Fig. 2).

RF rendered similar robustness with SVR at different growth stages in both the training and testing datasets (Fig. 3), and shows better robustness than ANN at each stage. Meanwhile, the RF model for each stage has a little better generalization capability than the ANN model, which behaves relatively unpredictable when used with independent input data that deviate from what was presented during the training stage [44,45]. Compared with the RF and SVR results for each stage, ANN shows much poorer performance in testing than in training. This is due to the fact that ANN is often applied to large amounts of sampling data, but SVR and RF are suitable for small amounts of sampling data. Another reason for this is possibly that the learning ability is too strong during the ANN process training, and thus the model obtained cannot reflect the hidden rules of samples that ultimately weaken prediction ability.

Most of the 15 vegetation indices in this study are correlated. However, as demonstrated by Cutler et al. [46] RF is not sensitive to collinearity. This is very valuable in modeling, especially for a complex, nonlinear system because it is commonly difficult to decide which variable to remove when two (or more) variables correlate with each other [47].

For estimation models of vegetation biochemical and biophysical variables to be useful in guiding on-farm crop management, they must perform well in farmers' fields. Therefore, data that fully represents real farm conditions should be included in model training and testing. Data in many previous studies were based on designated experimental sites rather than farmers' fields [48–50]. In the present study, we pooled data from farmers' fields in 2010, 2011, 2012 and 2014, and then randomly divided it into a training dataset and an independent testing dataset (75% and 25%, respectively).

A single vegetation index was usually selected in previous studies, to remotely estimate biomass in crops [51–52]. However, a single vegetation index is influenced by different degrees of saturability or soil background, and is consequently affected by regional specificity and timeliness [53]. This study shows that use of a combination of 15 vegetation indices and the RF regression algorithm improved the accuracy of prediction of wheat biomass. We propose for the first time use of RF regressions for remote monitoring of biomass, but the prediction accuracy of the method should be further investigated by optimizing the modeling algorithms.

Previous studies of crop growth monitoring based on remotely sensed data have often used a single algorithm to monitor different growing parameters at different growth stages [54–55]. In this work, we used RF to estimate wheat biomass on a much larger scale, assuming that it would help to improve wheat growth monitoring in the study areas. It would be interesting to apply the method to monitor other crop growth parameters with different features to verify reproducibility. This research contributes to the establishment of management strategies for non-destructive monitoring and precise modeling methods.

6. Conclusion

Biomass is an important indicator of crop growth. To estimate biomass in wheat rapidly and non-destructively, an improved method that combines vegetation indices based on HJ-CCD and random forest (RF) regression method is proposed. Estimation accuracy and robustness of the RF model were verified for each stage (i.e. jointing, booting, and anthesis). Furthermore, the RF model results were compared with support vector regression (SVR) and artificial neural network (ANN) models. The estimation accuracy of RF outperformed that of SVR and ANN at each stage. For RF models, the R^2 values for the estimated-versus-measured biomass regression for the three stages were 0.533, 0.721 and 0.79, respectively, and the corresponding RMSE values were 477, 1126.2 and 1808.2 kg ha^{-1}. The RF model was as robust as SVR and more robust than ANN. The relative RMSE values obtained

from the RF and SVR models were about 8% in training and 20% in testing for each stage, respectively. The relative RMSE of ANN was about 4% in training at each stage, whereas the RMSE values in testing were about 35% at jointing, 45% at booting, and 30% at anthesis.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 31271642), the Natural Science Foundation of Education Department of Jiangsu Province (No. 09KJB20013, No. 12KJB520018), the Six Talent Summit Project of Jiangsu Province (No. 2011-NY039), and the Science and Technology Innovation Development Foundation of Yangzhou University (No. 2015CXJ022).

REFERENCES

[1] A. Kross, H. McNairn, D. Lapen, M. Sunohara, C. Champagne, Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops, Int. J. Appl. Earth Obs. Geoinf. 34 (2015) 235–248.
[2] Y.Y. Fu, G.J. Yang, J.H. Wang, X.Y. Song, H.K. Feng, Winter wheat biomass estimation based on spectral indices, band depth analysis and partial least squares regression using hyperspectral measurements, Comput. Electron. Argic. 100 (2014) 51–59.
[3] M.L. Gnyp, Y.X. Miao, F. Yuan, S.L. Ustin, K. Yu, Y.K. Yao, S.Y. Huang, G. Bareth, Hyperspectral canopy sensing of paddy rice aboveground biomass at different growth stages, Field Crops Res. 155 (2014) 42–55.
[4] S. Gao, Z. Niu, N. Huang, X.H. Hou, Estimating the leaf area index, height and biomass of maize using HJ-1 and RADARSAT-2, Int. J. Appl. Earth Obs. Geoinf. 24 (2013) 1–8.
[5] X.L. Jin, G.J. Yang, X.G. Xu, H. Yang, H.K. Feng, Z.H. Li, J.X. Shen, C.J. Zhao, Y.B. Lan, Combined multi-temporal optical and radar parameters for estimating LAI and biomass in winter wheat using HJ and RADARSAR-2 data, Remote Sens. 7 (2015) 13251–13272.
[6] C.J. Gleason, J. Im, Forest biomass estimation from airborne LiDAR data using machine learning approaches, Remote Sens. Environ. 125 (2012) 80–91.
[7] J.M. Montes, F. Technow, B.S. Dhillon, F. Mauch, A.E. Melchinger, High-throughput non-destructive biomass determination during early plant development in maize under field conditions, Field Crops Res. 121 (2011) 268–273.
[8] R. Prasad, A. Pandey, K.P. Singh, V.P. Singh, R.K. Mishra, D. Singh, Retrieval of spinach crop parameters by microwave remote sensing with back propagation artificial neural networks: a comparison of different transfer functions, Adv. Space Res. 50 (2012) 363–370.
[9] D.C. Wang, J.H. Wang, N. Jin, Q. Wang, C.J. Li, J.F. Huang, Y. Wang, F. Huang, ANN-based wheat biomass estimation using canopy hyperspectral vegetation indices, Trans. CSAE 24 (2008) 196–201 (in Chinese with English abstract).
[10] J.G.P.W. Clevers, G.W.A.M. van der Heijden, S. Verzakov, M.E. Schaepman, Estimating grassland biomass using SVM band shaving of hyperspectral data, Photogramm. Eng. Remote. Sens. 73 (2007) 1141–1148.
[11] X.L. Jin, W.Y. Diao, C.H. Xiao, F.Y. Wang, B. Chen, K.R. Wang, S.K. Li, Estimation of wheat agronomic parameters using new spectral indices, PLoS One 8 (2013), e72736.
[12] R. Jhonnerie, V.P. Siregar, B. Nababan, L.B. Prasetyo, S. Wouthuyzen, Random forest classification for mangrove land cover mapping using Landsat5 TM and ALOS PALSAR imageries, Procedia Environ. Sci. 24 (2015) 215–221.
[13] I. Nitze, B. Barrett, F. Cawkwell, Temporal optimization of image acquisition for land cover classification with random forest and MODIS time-series, Int. J. Appl. Earth Obs. Geoinf. 34 (2015) 136–146.
[14] P.O. Gislason, J.A. Benediktsson, J.R. Sveinsson, Random forests for land cover classification, Pattern Recogn. Lett. 27 (2006) 294–300.
[15] A. Puissant, S. Rougier, A. Stumpf, Object-oriented mapping of urban trees using random forest classifiers, Int. J. Appl. Earth Obs. Geoinf. 26 (2014) 235–245.
[16] D.J. Booker, T.H. Snelder, Comparing methods for estimating flow duration curves at ungauged sites, J. Hydrol. 434 (2012) 78–84.
[17] M.L. Liu, X.N. Liu, D. Liu, C. Ding, J.L. Jiang, Multivariable integration method for estimating sea surface salinity in coastal waters from in situ data and remotely sensed data using random forest algorithm, Comput. Geosci. 75 (2015) 44–56.
[18] O. Mutanga, E. Adam, M. Azong Cho, High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm, Int. J. Appl. Earth Obs. Geoinf. 18 (2014) 399–406.
[19] J.W. Rouse, R.H. Haas, J.A. Schell, Monitoring the Vernal Advancement of Retrogradation (Green Wave Effect) of Natural Vegetation, Remote Sensing Center, Texas A&M University College Station, USA, 1974.
[20] A.R. Huete, A soil vegetation adjusted index (SAVI), Remote Sens. Environ. 25 (1998) 295–309.
[21] G. Rondeaux, M. Steven, F. Baret, Optimization of soil-adjusted vegetation indices, Remote Sens. Environ. 55 (1996) 95–107.
[22] T.D. Schleicher, W.C. Bausch, J.A. Delgado, P.D. Ayers, Evaluation and refinement of the nitrogen reflectance index (NRI) for site-specific fertilizer management, in: 2001 ASABE Annual International Meeting, St-Joseph, MI, USA. ASABE Paper No. 01–11151, 2001.
[23] A.A. Gitelson, Y. Kaufman, M.N. Merzlyak, Use of a green channel in remote sensing of global vegetation from EOS-MODIS, Remote Sens. Environ. 58 (1996) 289–298.
[24] J. Penuelas, I. Filella, J.A. Gamon, Assessment of photosynthetic radiation-use efficiency with spectral reflectance, New Phytol. 131 (1995) 291–296.
[25] M.N. Merzlyak, A.A. Gitelson, O.B. Chivkunova, Y.R. Rakitin, Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening, Physiol. Plant. 106 (1999) 135–141.
[26] F. Baret, G. Guyot, Potentials and limits of vegetation indices for LAI and APAR assessment, Remote Sens. Environ. 35 (1991) 161–173.
[27] A.A. Gitelson, Y. Zur, O.B. Chivkunova, M.N. Merzlyak, Assessing carotenoid content in plant leaves with reflectance spectroscopy, Photochem. Photobiol. 75 (2002) 272–281.
[28] H.Q. Liu, A.R. Huete, A feedback based modification of the NDVI to minimize canopy background and atmospheric noise, IEEE Trans. Geosci. Remote Sens. 33 (1995) 457–465.
[29] J.M. Chen, Evaluation of vegetation indices and modified simple ratio for boreal applications, Can. J. Remote. Sens. 22 (1996) 229–242.
[30] N.S. Goel, W.H. Qin, Influences of canopy architecture on relationships between various vegetation indexes and LAI and FPAR: a computer simulation, Remote Sens. Environ. 10 (1994) 309–347.
[31] K. Wang, Z.Q. Shen, R.C. Wang, Effects of nitrogen nutrition on the spectral reflectance characteristics of rice leaf and canopy, J. Zhejiang Agric. Univ. 24 (1998) 93–97.
[32] N.H. Broge, E. Leblanc, Comparing prediction power and stability of broad band and hyperspectral vegetation indices

for estimation of green leaf area index and canopy chlorophyll density, Remote Sens. Environ. 76 (2000) 156–172.
[33] D. Haboudane, J.R. Miller, E. Pattey, P.J. Zarco-Tejada, I.B. Strachan, Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: modeling and validation in the context of precision agriculture, Remote Sens. Environ. 90 (2004) 337–352.
[34] L. Breiman, J. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, The Wadsworth Statistics/Probability Series, Wadsworth, Belmont, CA, 1984.
[35] J.R. Quinlan, C4.5 Programs for Machine Learning, Morgan Kaurmann, San Mateo, CA, 1993.
[36] V. Rodriguez-Galiano, M.P. Mendes, M.J. Garcia-Soldado, M. Chica-Olmo, L. Ribeiro, Predictive modeling of groundwater nitrate pollution using random forest and multisource variables related to intrinsic and specific vulnerability: a case study in an agricultural setting (southern Spain), Sci. Total Environ. 476–477 (2014) 189–206.
[37] A. Prasad, L. Iverson, A. Liaw, Newer classification and regression tree techniques bagging and random forests for ecological prediction, Ecosystems 9 (2006) 181–199.
[38] A. Huete, K. Didan, T. Miura, E.P. Rodriguez, X. Gao, L.G. Ferreira, Overview of the radiometric and biophysical performance of the MODIS vegetation indices, Remote Sens. Environ. 83 (2002) 195–213.
[39] F. Li, Y.X. Miao, S.D. Hennig, M.L. Gnyp, X.P. Chen, L.L. Jia, G. Bareth, Evaluating hyperspectral vegetation indices for estimating nitrogen concentration of winter wheat at different growth stages, Precis. Agric. 11 (2010) 335–357.
[40] F. Li, B. Mistele, Y.C. Hu, X.L. Yue, Y.X. Miao, X.P. Chen, Z.L. Cui, Q.F. Meng, U. Schmidthalter, Remotely estimating aerial N status of phenologically differing winter wheat cultivars grown in contrasting climatic and geographic zones in China and Germany, Field Crops Res. 138 (2012) 21–32.
[41] J.L. Hatfield, J.H. Prueger, Value of using different vegetative indices to quantify agricultural crop characteristics at different growth stages under varying management practices, Remote Sens. 2 (2010) 562–578.
[42] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.
[43] S.O. Haykin, Neural Networks: A Comprehensive Foundation, second ed. Prentice Hall, 1999.
[44] J. Verrelst, J. Muñoz, L. Alonso, J. Delegido, J.P. Rivera, G. Camps-Valls, J. Moreno, Machine learning regression algorithms for biophysical parameter retrieval: opportunities for Sentinel-2 and -3, Remote Sens. Environ. 118 (2012) 127–139.
[45] F. Baret, S. Buis, Estimating canopy characteristics from remote sensing observations: review of methods and associated problems, in: S. Liang (Ed.), Advances in Land Remote Sensing: System, Modeling, Inversion and Application, Springer, The Netherlands 2008, pp. 173–201.
[46] R.D. Cutler, T.C. Edwards, K.H. Beard, K.T. Cutler, H.J. Gibson, J.J. Lawler, Random forests for classification in ecology, Ecology 88 (2007) 2783–2792.
[47] S. Fukuda, E. Yasunaga, M. Nagle, K. Yuge, V. Sardsud, W. Spreer, J. Müller, Modelling the relationship between peel colour and the quality of fresh mango fruit using random forests, J. Food Eng. 131 (2014) 7–17.
[48] J. Bendig, K. Yu, H. Aasen, A. Bolten, S. Bennertz, J. Broscheit, M.L. Gnyp, G. Bareth, Combining UAV-based plant height from crop surface models, visible and near infrared vegetation indices for biomass monitoring in barley, Int. J. Appl. Earth Obs. Geoinf. 39 (2015) 79–87.
[49] K. Prabhakara, W.D. Hively, G.W. McCarty, Evaluating the relationship between biomass, percent groundcover and remote sensing indices across six winter cover crop fields in maryland, United States, Int. J. Appl. Earth Obs. Geoinf. 39 (2015) 88–102.
[50] X.L. Jin, K.R. Wang, C.H. Xiao, W.Y. Diao, F.Y. Wang, B. Chen, S.K. Li, Comparison of two methods for estimation of leaf total chlorophyll content using remote sensing in wheat, Field Crops Res. 135 (2012) 24–29.
[51] P.F.C. Monteiro, R.A. Filho, A.C. Xavier, R.O.C. Monteiro, Assessing biophysical variable parameters of bean crops with hyperspectral measurements, Sci. Agric. 69 (2012) 87–94.
[52] P.M. Hansen, J.K. Schjoerring, 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 (2003) 542–553.
[53] X.C. Li, X.G. Xu, Y.S. Bao, W.J. Huang, J.H. Luo, Y.Y. Dong, X.Y. Song, J.H. Wang, Retrieving LAI of winter wheat based on sensitive vegetation index by the segmentation method, Sci. Agric. Sin. 45 (2012) 3486–3496 (in Chinese with English abstract).
[54] E. Boegh, R. Houborg, J. Bienkowski, C.F. Braban, T. Dalgaard, N. van Dijk, U. Dragosits, E. Holmes, Remote sensing of LAI, chlorophyll and leaf nitrogen pools of crop and grasslands in five European landscapes, Biogeosciences 10 (2013) 6279–6307.
[55] J.U.H. Eitel, T.S. Magney, L.A. Vierling, T.T. Brown, D.R. Huggins, LiDAR based biomass and crop nitrogen estimates for rapid, non-destructive assessment of wheat nitrogen status, Field Crops Res. 159 (2014) 21–32.