Harvest chronological planning using a method based on satellite-derived vegetation indices and artificial neural networks

Sepideh Taghizadeh (Taghizadeh, S)¹, Hossein Navid (Navid, H)¹, Reza Adiban (Adiban, R)² and Yasser Maghsodi (Maghsodi, Y)²

¹University of Tabriz, Biosystems Engineering Dept., Tabriz 5166614776, Iran. ²Agriculture Engineering Research Department, Ardabil Agricultural and Natural Resources Research and Education Center, AREEO, Ardabil (Moghan), Iran. ³K.N. Toosi University of Technology, Photogrammetry and Remote Sensing Dept., Tehran, Iran.

Abstract

Aim of study: Wheat appropriate harvest date (WAHD) is an important factor in farm monitoring and harvest campaign schedule. Satellite remote sensing provides the possibility of continuous monitoring of large areas. In this study, we aimed to investigate the strength of vegetation indices (VIs) derived from Landsat-8 for generating the harvest schedule regional (HSR) map using Artificial Neural Network (ANN), a robust prediction tool in the agriculture sector.

Area of study: Qorveh plain, Iran.

Material and methods: During 2015 and 2016, a total of 100 plots was selected. WAHD was determined by sampling of plots and specifying wheat maximum yield for each plot. The strength of eight Landsat-8 derived spectral VIs (NDVI, SAVI, GreenNDVI, NDWI, EVI, EVI2, CVI and CI\textsubscript{green}) was investigated during wheat growth stages using correlation coefficients between these VIs and observed WAHD. The derived VIs from the required images were used as inputs of ANNs and WAHD was considered as output. Several ANN models were designed by combining various VIs data.

Main results: The temporal stage in agreement with dough development stage had the highest correlation with WAHD. The optimum model for predicting WAHD was a Multi-Layer Perceptron model including one hidden layer with ten neurons in it when the inputs were NDVI, NDWI, and EVI2. To evaluate the difference between measured and predicted values of ANNs, MAE, RMSE, and \(R^2\) were calculated. For the 3-10-1 topology, the value of \(R^2\) was estimated 0.925. A HSR map was generated with RMSE of 0.86 days.

Research highlights: Integrated satellite-derived VIs and ANNs is a novel and remarkable methodology to predict WAHD, optimize harvest campaign scheduling and farm management.

Additional keywords: harvest date; Landsat-8 satellite; remote sensing; wheat.

Abbreviations used: ANN (artificial neural network); CI\textsubscript{green} (green chlorophyll index); CVI (chlorophyll vegetation index); EVI (enhanced vegetation index); EVI2 (2 bands EVI); FLAASH (fast line-of-sight atmospheric analysis of spectral hypercubes); GDM (gradient descent with momentum); GreenNDVI (green normalized difference vegetation index); HSR (harvest schedule regional); MAE\textsubscript{p} (mean absolute error for prediction); MLP (multi-layer perceptron); N (number of samples); NDVI (normalized difference vegetation index); NDWI (normalized difference water index); NIR (near-infrared); OLI (operational land imager); R (correlation coefficient); \(R^2\)\textsubscript{p} (coefficient of determination for prediction); RMSE\textsubscript{p} (root mean square error for prediction); SAVI (soil adjusted vegetation index); SWIR (short wave); STICS (simulator multidisciplinaire pour les cultures standard); TANH (hyperbolic tangent); TIRS (thermal infrared sensor); VI (vegetation index); w (weight); WAHD (wheat appropriate harvest date); WOFOST (Worlds Food Study); \(\bar{Y}\)\textsubscript{m} (mean of estimated values of WAHD); \(\bar{Y}\)\textsubscript{m} (mean of measured values of WAHD); \(\bar{Y}\)\textsubscript{est} (measured WAHD); \(\bar{Y}\)\textsubscript{est} (estimated WAHD); \(\alpha\) (momentum term); \(\eta\) (learning rate); \(o\) (output); \(\delta\) (error).

Authors’ contributions: Conceived and research design: ST, RA, and YM. Acquisition data and statistical analysis and writing: ST, RA. Critical revision of the manuscript for important intellectual content: ST, RA, NA. Coordinating and supervising the research work: HN, YM. All authors read and approved the final manuscript.


Received: 03 Dec 2018. Accepted: 14 Oct 2019.

Copyright © 2019 INIA. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC-by 4.0) License.

Funding: University of Tabriz (Biosystems Engineering Dept.), Tabriz, Iran.

Competing interests: The authors have declared that no competing interests exist.

Correspondence should be addressed to Hossein Navid: navid@tabrizu.ac.ir
Introduction

Wheat (*Triticum aestivum* L.) is one of the most important cereals grown in the world. More humans consume wheat as their main food more than other cereal grains (Pimentel & Pimentel, 2007). So, wheat plays a very important role in the world’s supply chain and food security. Therefore, maximum yield achievement at minimum expense is the main goal of wheat production. Awareness of crop growing dates, such as planting and harvesting dates, helps farm managers to reach the objectives and farm management. Harvest date is an important factor in field management. Planning the crop appropriate harvest date helps to schedule for harvest operation with less cost and maximize the profit (Abawi, 1993; Suwannachatkul et al., 2014). In these studies, wheat appropriate harvest date (WAHD) is the date in which the wheat has reached maturity, and its moisture content is convenient for harvest by a combine harvester. In Iran, the appropriate moisture content is in the range of 12-14 percent (Mansouri-Rad, 2000). If wheat is harvested earlier than WAHD, there will be yield loss and its moisture content will be inappropriate for harvesting and storage (Burnett & Bakke, 1930; Philips & O'Callaghan, 1974; Sabir et al., 2005). In contrast, harvesting after this date causes dry matter yield diminution and grain quality degradation. Moreover, yield loss due to weather conditions (such as temperature, rain, and wind), birds and insects damages to the matured product are disadvantages of delay in the harvest (BoIIand, 1984; Abawi, 1993; Farrer et al., 2006; Sun et al., 2007). In some areas of the world, the majority of farmers are smallholders, and the size of most agricultural farms is small. Consequently, it is not affordable for them to own their combine harvesters, leading to a shortage of harvesters in the harvest time (Bougaris et al., 2013). Lack of necessary machinery during harvest campaign could further delay the harvest date. Therefore, the wheat producers require prior planning for providing the combine harvester rental and labors which requires former notice of WAHD for each field. Field surveying operations are costly, time-consuming and infeasible in large areas to determine harvest date (Moran & Pearce, 1997; Pinter et al., 2003). Some researchers (Porter & Gawith, 1999; McMaster & Wilhelm, 2003; Streck et al., 2003; Evers et al., 2010) have improved wheat phenological stages prediction models using photoperiod, water and vernalization to predict wheat maturity date. Other models such as STICS (Simulateur multidiplinaire pour les Cultures Standard) and WOFOST (Worlds Food Study) have been developed by taking more factors into account such as temperature, nutrient and water stress for crop growth stage prediction (Boogaard et al., 1998; Brisson et al., 1998). Most of these models are based on meteorological parameters, making it difficult to discriminate within and between field differences in a dense configuration of the spatial grid (Meng et al., 2015). Moreover, the main limitation of crop models is the challenges in preparing reliable input data. The uncertainty about the spatial repartition of soil properties and micro meteorological variables at farm scale limits these model output’s assurance (Therond et al., 2011).

Remote sensing provides appropriate and timely images of the agricultural farms. The higher revisit frequency capability is the merit of remote sensing (Atzberger, 2013) for collecting the farm information. In recent decades, satellite remote sensing has been applied for agricultural field management operations such as yield and biomass estimation (Panda et al., 2010; Ren et al., 2008; Xie et al., 2009), phenological date prediction (Sakamoto et al., 2010; De Bernardis et al., 2016), and mapping of land use (Galford et al., 2008; Atzberger & Rembold, 2013). Vegetation indices (VIs), which are computational combinations of different spectral bands of the electromagnetic spectrum, simplify the analysis and processing of big data obtained by satellites (Govarts et al., 1999; Viña et al., 2011). The strong contrast of absorption and scattering of the red and near-infrared bands can be combined into different quantitative indices to explain the vegetation conditions. VIs are the semi-analytical measurements of plant vegetation activity which explain vegetation conditions during the growth stages. The benefit of VIs utilization is spectral reflectance data enhancement, by considering the variability of vegetation and minimizing of the atmospheric effect, sun-target-sensor geometry and soil (Moulin, 1999; Viña et al., 2011).

The relationship between remote sensing indices and plant bio-physical variables is nonlinear (Haboudane et al., 2004). Therefore, the usage of methods creating a non-linear relation between independent variables (VIs) and dependent variables could help predict better the plant bio-physical behavior changes. The artificial neural network (ANN) is the technique, widely used in the field of geo/bio-physical variable detections (Beale et al., 2008). Application of ANN techniques using the VIs and also visible blue, green, red, near-Infrared (NIR) and short wave (SWIR) regions of the electromagnetic spectrum have led to successful results for crop monitoring, crop cover, crop growth, crop nitrogen, crop yield and biomass estimation (Chen & McNairn, 2006; Karimi et al., 2006; Li et al., 2007; Xie et al., 2009; Prasad et al., 2012). The aim of this study was to generate a harvest schedule regional (HSR) map for predicting WAHD by using two years (2015...
& 2016) VIs data derived by operational land imager (OLI) sensor (Landsat-8) and ANN techniques.

Material and methods

Study site

The study was carried out in Qorveh plain in southeast of Kordestan province, West of Iran (35°15´N, 47°80´E), altitude 1900 m, with cold snowy winters and temperate summers (Fig. 1). The mean annual temperature of the site is 10.6 °C, and the average annual precipitation is 439 mm. The rainfed winter wheat is the major crop in this region, and it is sown in mid-October and harvested in mid-July of next year. The wheat-growing season is about 250 days.

Data acquisition

Field data

Field data for rainfed wheat yield measurement was collected during the wheat maturing stage of 2015 and 2016 years (from June 19 to July 11). Fifty sampling plots were selected from fifty wheat farms, in each year. They were flat and homogenous with areas bigger than 175×175 m². The size of plots chosen was 60×60 m² and plots were located at the center of wheat farms. A GPS receiver (Garmin GPSMAP 62s, Taiwan, with a spatial accuracy 3 m) was used for recording the geo-coordinates of each plot. The plant density in each plot was obtained in early June using a 1 m × 1 m quadrate. The plant density measurements were done by ten times random throwing of the quadrate through the zigzag sweep path and counting the number of plants in each throw and averaging of plant numbers of the quadrates. According to the field observations, plant densities often were constant on each farm.

The yield sampling was carried out by five random throws of the quadrate and choosing three plants in each throw. The plants were randomly chosen in the quadrate. At each sampling, fifteen plants were sampled for each plot, in total. By measuring the average grain weight of sampled plants in each plot and knowing plant density, the yield amount was acquired for each plot. The grains were used to measure the yield of fully matured plants. The sampling operation was conducted at a frequency of 2 days, from June 19 to July 11. Generally, eleven yield sampling observations were carried out for each plot in each year. The interpolation between yield observations was done by fitting a spline curve. The WAHD was considered the day of wheat highest yield.

Satellite image acquisition and pre-processing

Landsat-8 is the latest in a series of Landsat satellites, and it was launched on February 11, 2013 (http://science.nasa.gov/missions/ldcm). The operational land imager (OLI) and the thermal infrared sensor (TIRS) are two sensors which are carried by the Landsat-8 (Roy et al., 2014). We used OLI sensor data in this study. The appropriate spatial resolution of OLI in comparison with the common size of agricultural farms is its advantage for using in agricultural studies. OLI consist of nine spectral bands. The spatial resolution for bands 1 to 7 (coastal, blue, green, red, NIR, SWIR1, SWIR2) and band 9 (cirrus), is 30 m and for panchromatic band (band 8) is 15 m. A series of OLI images of the

![Figure 1. Study site and sampling farm locations. Background: Landsat-8 OLI at 17 May 2016 (Row: 36, Path: 166).](image-url)
study area (path: 166, 167 & row: 35, 36) was acquired after wheat dormancy stage with no cloud or less than 10% amount of cloud in each year. The images were acquired in 2015 on April 4, April 20, April 29, May 6, May 31, June 7, June 16 and June 23 which there were coincided with tilling, stem extension (20th & 29th of April), booting, heading, dough development, physiological maturity and harvest ripe stages of wheat, respectively. Also, in 2016 the images on April 6, April 22, May 1, May 8, May 17, June 9, June 18 and June 25 were acquired which there were coincided with tilling, stem extension (22nd of April & 1st of May), booting, heading, dough development, physiological maturity and harvest ripe stages of wheat, respectively.

The pre-processing of images was performed in ENVI 5.1 software. Since the OLI images had been corrected geometrically, we just did the second geo-referencing operation to ensure the correct location of terrain. The ground control points derived from 1:25000 topographic maps were applied for geometric correction. The geo-correction error less than 0.5 pixel (15 m) was attained. The nearest neighbor method and linear polynomial geometric model were used. Atmospheric corrections were performed using MODTRAN 4 model in FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes) module in the ENVI 5.1. The parameters used in the FLAASH were adjusted based on scene center location, sensor type, sensor altitude, ground elevation, pixel size, information about sensor flight date and weather conditions on the image acquisition date. The output of the FLAASH package was the surface reflectance of OLI images.

Extraction of vegetation indices

There are some studies which have suggested double or multi-band spectral indices to estimate bio-physical changes of crops (Garroutte et al., 2016; Li et al., 2016). This study included eight widely used VIs to investigate the strength of VIs during wheat growing stages. These indices included: NDVI (normalized difference vegetation index), SAVI (soil adjusted vegetation index), EVI (enhanced vegetation index), EVI2 (2 bands enhanced vegetation index), NDWI (normalized difference water index), GreenNDVI (green normalized difference vegetation index), CVI (chlorophyll vegetation index) and CIGreen (green chlorophyll index). The VIs mentioned above can be classified in three groups based on their sensitivity to green biomass (NDVI, SAVI, EVI, EVI2), the liquid water content of vegetation (NDWI) and leaf chlorophyll (CVI, GreenNDVI, and CIGreen). Table 1 represents the formula of these indices.

Determining the best temporal stage for WAHD prediction

In each year, to determine the best temporal and phenological stage for predicting WAHD, correlation coefficients (R) between VIs and WAHD were calculated for different images acquired during the wheat growth phases. The date with the highest R values was recognized as the best temporal stage for WAHD prediction.

The artificial neural network model

The ANNs are flexible mathematical models that accomplish a computational simulation based on the behavior of human brain neurons. The ANN is a non-linear machine learning algorithm with a high potential for modeling and prediction (Foody, 2004). It is composed of artificial neuron groups which are interconnected with weighted links and could create a strong relation between inputs and outputs using a learning approach (Omer et al., 2016). The network architecture is the first step for developing an ANN model that is determined by artificial neurons and layers.

### Table 1. Formula for several spectral vegetation indices.

<table>
<thead>
<tr>
<th>Vegetation index</th>
<th>Formula</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDVI</td>
<td>( \text{NDVI} = (\rho_{\text{NIR}} - \rho_{\text{Red}})/(\rho_{\text{NIR}} + \rho_{\text{Red}}) )</td>
<td>Rouse et al., 1974</td>
</tr>
<tr>
<td>SAVI</td>
<td>( \text{SAVI} = (\rho_{\text{NIR}} - \rho_{\text{Red}})(1 + L)/(\rho_{\text{NIR}} + \rho_{\text{Red}} + L) )</td>
<td>Huete, 1988</td>
</tr>
<tr>
<td>EVI</td>
<td>( \text{EVI} = 2.5 (\rho_{\text{NIR}} - \rho_{\text{Red}})(\rho_{\text{NIR}} + 6\rho_{\text{Red}} - 7.5\rho_{\text{Blue}} + 1) )</td>
<td>Huete et al., 1997</td>
</tr>
<tr>
<td>EVI2</td>
<td>( \text{EVI2} = 2.5 (\rho_{\text{NIR}} - \rho_{\text{Red}})(\rho_{\text{NIR}} + 2.4\rho_{\text{Red}} + 1) )</td>
<td>Jiang et al., 2008</td>
</tr>
<tr>
<td>NDWI</td>
<td>( \text{NDWI} = (\rho_{\text{NIR}} - \rho_{\text{SWIR}})(\rho_{\text{NIR}} + \rho_{\text{SWIR}}) )</td>
<td>Gao, 1996</td>
</tr>
<tr>
<td>CVI</td>
<td>( \text{CVI} = (\rho_{\text{NIR}}/\rho_{\text{Green}}) * (\rho_{\text{Red}}/\rho_{\text{Green}}) )</td>
<td>Vinci et al., 2007</td>
</tr>
<tr>
<td>GreenNDVI</td>
<td>( \text{GreenNDVI} = (\rho_{\text{NIR}} - \rho_{\text{Green}})(\rho_{\text{NIR}} + \rho_{\text{Green}}) )</td>
<td>Gitelson &amp; Merzlyak, 1998</td>
</tr>
<tr>
<td>CI(_{\text{Green}})</td>
<td>( \text{CI}<em>{\text{Green}} = (\rho</em>{\text{Red}}/\rho_{\text{Green}}) - 1 )</td>
<td>Gitelson et al., 2003</td>
</tr>
</tbody>
</table>

\( \rho_{\text{NIR}}, \rho_{\text{Green}}, \rho_{\text{Red}}, \) and \( \rho_{\text{SWIR}} \) are spectral reflectance of red, green, blue, near infrared and short wave infrared bands which are in accordance with bands 2, 3, 4, 5 and 6 for OLI sensor, respectively.
A usual ANN consists of an input layer, an output layer, and several hidden layers. Also, each layer contains some neurons and activation functions. In this study, we used a Multi-layer perceptron (MLP) network which is extensively used type of ANNs in the community of remote sensing. The gradient descent with momentum (GDM) algorithm was used to train the network. For this algorithm, the weights are updated during the \( n \)th training iteration as follows (Omid et al., 2009):

\[
w_{ij}^{(n)} = w_{ij}^{(n-1)} + \Delta w_{ij}^{(n)}
\]

Additionally, the set of weights are given by:

\[
\Delta w_{ij}^{(n)} = \eta \delta_j^{(n)} o_i^{(n)} + \alpha \Delta w_{ij}^{(n-1)}
\]

where \( w_{ij}^{(n)} \) is the weight between the \( j \)th neuron of the upper layer and the \( i \)th neuron of the lower layer, \( \delta_j \) is the error of the \( j \)th neuron, \( o_i \) is the output value of the \( i \)th neuron of the previous layer, \( \eta \) is the learning rate and \( \alpha \) is the momentum term. \( \Delta w_{ij}^{(n)} \) is the gradient vector affiliated with the weights.

**ANNs structure**

To consider the efficacy of inputs (Green biomass sensitive, chlorophyll sensitive and water content sensitive VIs) on WAHD prediction, eight MLP type models were designed. Fig. 2 shows a typical scheme of an MLP model which comprises the input layer, the hidden layer, and the output layer. Different models were designed using various combination of VIs as the neurons in the input layer (ANN-1 to ANN-8). The input variables for ANN-1 to ANN-3 were green biomass sensitive VIs, chlorophyll sensitive VIs and water content sensitive VIs, respectively. The input variables for ANN-4 were EVI2 (green biomass sensitive), NDWI (water content sensitive) and GreenNDVI (chlorophyll sensitive) that each of them had the highest correlation with the observed WAHD in their groups. The input variables for ANN-5 was a combination of the NDVI and EVI2 (green biomass sensitive) and NDWI (water content sensitive). ANN-6 was a combination of green biomass sensitive VIs (NDVI, SAVI, EVI2) and chlorophyll sensitive VI (GreenNDVI). The input variables for ANN-7 were the combination of chlorophyll sensitive and water content sensitive VIs. The input variables for ANN-8 were the combination of green biomass sensitive and water content sensitive VIs (Table 2).

For determining the optimum number of hidden layer neurons, the minimum value of RMSE was considered. In this study, the used activation function was hyperbolic tangent (TANH) for hidden and output layers. Also, the values of \( \alpha=0.7 \) and \( \eta=0.1 \) were used.

The data set on 100 VIs for predicting WAHD was split into three parts: 70% of VIs for training, 15% for cross-validation and 15% for testing data. After sufficient training, the weights of the network were adapted and applied for validation to determine the model overall performance.

**ANN model performance and validation**

The performance of eight ANN constructed model using a different number of VIs was evaluated by statistical parameters including coefficient of determination for prediction (test data) \( (R^2_p) \), mean absolute error for prediction (MAE_p), and the root mean square error for...
prediction (RMSEp). The values of $R^2$, MAEp, and RMSEp were calculated using equations 3, 4 and 5, respectively.

$$R^2_p = \frac{\sum_{i=1}^{N}(Y_{est,i} - \bar{Y}_{mea,i})(Y_{mea,i} - \bar{Y}_{mea,i})^2}{\sum_{i=1}^{N}(Y_{est,i} - \bar{Y}_{est,i})^2 \sum_{i=1}^{M}(Y_{mea,i} - \bar{Y}_{mea,i})^2}$$ (3)

$$MAEp = \frac{\sum_{i=1}^{N}|Y_{est,i} - Y_{mea,i}|}{N}$$ (4)

$$RMSEp = \sqrt{\frac{\sum_{i=1}^{N}(Y_{est,i} - Y_{mea,i})^2}{N}}$$ (5)

where $Y_{mea,i}$ is the measured WAHD, $Y_{est,i}$ is the estimated WAHD, $\bar{Y}_{mea,i}$ is the mean of measured values of WAHD, $\bar{Y}_{est,i}$ is the mean of estimated values of WAHD, and $N$ is the number of samples. The higher values for $R^2_p$ and lower values of MAEp and RMSEp represent further precision and accuracy of the model. All calculations for ANN models were implemented using custom-written scripts of MATLAB R2015a software.

**Harvest schedule regional (HSR) map**

The predicted data of the best MLP model (in terms of precision and accuracy) was used for generating the HSR map using ArcGIS 10 software. Furthermore, the WAHD map was provided using the data of the test set. Moreover, both of the generated maps were compared with maps of measured data.

**Results**

### The best temporal stage for WAHD prediction

Table 3 and Table 4 show the variation of $R$-values between VIs and observed WAHD in 2015 and 2016, respectively. The $R$-values were calculated to specify which temporal stage and VIs are appropriate for WAHD prediction. Also, for better comprehend the results in Tables 3 and 4, the temporal variations of $R$-values were presented in Fig. 3. Most of $R$-values had an ascendant trend before Jun 7, 2015 (Fig. 3a), and Jun 9, 2016 (Fig. 3b). However, a decline was observed in $R$-values after these dates. Based on the results the days of early of June were the best temporal phase for WAHD prediction which was coincided with dough development stage of wheat.

In 2015, the maximum $R$-values for NDVI, SAVI, EVI, EVI2, GreenNDVI, CIgreen, and NDWI were 0.729, 0.698, 0.744, 0.594, 0.558 and 0.753 respectively.

**Table 2. Combination of various VIs data as inputs of estimated ANN models.**

<table>
<thead>
<tr>
<th>Prediction model</th>
<th>NDVI</th>
<th>SAVI</th>
<th>EVI</th>
<th>EVI2</th>
<th>GreenNDVI</th>
<th>CVI</th>
<th>CIgreen</th>
<th>NDWI</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN-1</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td></td>
<td>□</td>
<td>■</td>
</tr>
<tr>
<td>ANN-2</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td></td>
<td>□</td>
<td>■</td>
</tr>
<tr>
<td>ANN-3</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td></td>
<td>□</td>
<td>■</td>
</tr>
<tr>
<td>ANN-4</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td></td>
<td>□</td>
<td>■</td>
</tr>
<tr>
<td>ANN-5</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td></td>
<td>□</td>
<td>■</td>
</tr>
<tr>
<td>ANN-6</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td></td>
<td>□</td>
<td>■</td>
</tr>
<tr>
<td>ANN-7</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td></td>
<td>□</td>
<td>■</td>
</tr>
<tr>
<td>ANN-8</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td></td>
<td>□</td>
<td>■</td>
</tr>
</tbody>
</table>

**Table 3. The $R$ values between WAHD and various VIs at different dates in 2015. The maximum $R$ values are in bold type.**

<table>
<thead>
<tr>
<th>Vis 2015</th>
<th>Apr 4</th>
<th>Apr 20</th>
<th>Apr 29</th>
<th>May 6</th>
<th>May 31</th>
<th>Jun 7</th>
<th>Jun 16</th>
<th>Jun 23</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDVI</td>
<td>0.341</td>
<td>0.376</td>
<td>0.377</td>
<td>0.511</td>
<td>0.625</td>
<td>0.729</td>
<td>0.564</td>
<td>0.412</td>
</tr>
<tr>
<td>SAVI</td>
<td>0.298</td>
<td>0.401</td>
<td>0.392</td>
<td>0.456</td>
<td>0.607</td>
<td>0.698</td>
<td>0.618</td>
<td>0.441</td>
</tr>
<tr>
<td>EVI</td>
<td>0.376</td>
<td>0.369</td>
<td>0.395</td>
<td>0.524</td>
<td>0.664</td>
<td>0.744</td>
<td>0.628</td>
<td>0.514</td>
</tr>
<tr>
<td>EVI2</td>
<td>0.345</td>
<td>0.384</td>
<td>0.389</td>
<td>0.557</td>
<td>0.668</td>
<td>0.758</td>
<td>0.632</td>
<td>0.513</td>
</tr>
<tr>
<td>GreenNDVI</td>
<td>0.179</td>
<td>0.241</td>
<td>0.298</td>
<td>0.439</td>
<td>0.459</td>
<td>0.594</td>
<td>0.466</td>
<td>0.404</td>
</tr>
<tr>
<td>CVI</td>
<td>0.197</td>
<td>0.310</td>
<td>0.309</td>
<td>0.329</td>
<td>0.435</td>
<td>0.423</td>
<td>0.563</td>
<td>0.437</td>
</tr>
<tr>
<td>CIgreen</td>
<td>0.067</td>
<td>0.183</td>
<td>0.138</td>
<td>0.236</td>
<td>0.401</td>
<td>0.558</td>
<td>0.444</td>
<td>0.351</td>
</tr>
<tr>
<td>NDWI</td>
<td>0.175</td>
<td>0.208</td>
<td>0.257</td>
<td>0.379</td>
<td>0.627</td>
<td>0.753</td>
<td>0.608</td>
<td>0.356</td>
</tr>
</tbody>
</table>
Table 4. The $R$ values between WAHD and various VIs at different dates in 2016. The maximum $R$ values are in bold type.

<table>
<thead>
<tr>
<th>VIs 2016</th>
<th>Apr 6</th>
<th>Apr 22</th>
<th>May 1</th>
<th>May 8</th>
<th>May 17</th>
<th>Jun 9</th>
<th>Jun 18</th>
<th>Jun 25</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDVI</td>
<td>0.334</td>
<td>0.412</td>
<td>0.468</td>
<td>0.514</td>
<td>0.592</td>
<td><strong>0.698</strong></td>
<td>0.501</td>
<td>0.357</td>
</tr>
<tr>
<td>SAVI</td>
<td>0.330</td>
<td>0.361</td>
<td>0.410</td>
<td>0.478</td>
<td>0.517</td>
<td><strong>0.686</strong></td>
<td>0.551</td>
<td>0.473</td>
</tr>
<tr>
<td>EVI</td>
<td>0.409</td>
<td>0.326</td>
<td>0.319</td>
<td>0.480</td>
<td>0.597</td>
<td><strong>0.709</strong></td>
<td>0.506</td>
<td>0.467</td>
</tr>
<tr>
<td>EVI2</td>
<td>0.360</td>
<td>0.371</td>
<td>0.410</td>
<td>0.532</td>
<td>0.605</td>
<td><strong>0.763</strong></td>
<td>0.607</td>
<td>0.522</td>
</tr>
<tr>
<td>GreenNDVI</td>
<td>0.216</td>
<td>0.221</td>
<td>0.249</td>
<td>0.419</td>
<td>0.425</td>
<td>0.560</td>
<td><strong>0.582</strong></td>
<td>0.455</td>
</tr>
<tr>
<td>CVI</td>
<td>0.165</td>
<td>0.257</td>
<td>0.314</td>
<td>0.380</td>
<td>0.444</td>
<td>0.530</td>
<td><strong>0.554</strong></td>
<td>0.541</td>
</tr>
<tr>
<td>CI_green</td>
<td>0.114</td>
<td>0.169</td>
<td>0.278</td>
<td>0.312</td>
<td>0.335</td>
<td>0.482</td>
<td><strong>0.541</strong></td>
<td>0.480</td>
</tr>
<tr>
<td>NDWI</td>
<td>0.203</td>
<td>0.286</td>
<td>0.301</td>
<td>0.418</td>
<td>0.601</td>
<td><strong>0.747</strong></td>
<td>0.568</td>
<td>0.455</td>
</tr>
</tbody>
</table>

Figure 3. Temporal variations of $R$ values between WAHD and VIs (a) in 2015 and (b) in 2016.
that were obtained on Jun 7, but the R-value of CVI (0.563) was maximum on Jun 16 (Table 3).

As shown in Table 4, the maximum R-values were achieved on Jun 9 for NDVI (0.698), SAVI (0.686), EVI (0.709), EVI2 (0.763), and NDWI (0.747) in 2016. Additionally, the maximum of R-values for GreenNDVI, CVI, and CI green were 0.582, 0.554 and 0.541 on Jun 18. Considering the R-value as an indicator to assess the ability of VIs to estimate WAHD at wheat different growth stages, the June 7, 2015, and June 9, 2016, was selected for prediction.

**Evaluation of ANN models**

The results of designed models and the $R_p^2$, $\text{MAE}_p$, and $\text{RMSE}_p$ values for these models are summarized in Table 5. For these models, $R_p^2$, $\text{MAE}_p$ and $\text{RMSE}_p$ values ranged from 0.611 to 0.925, 0.70 to 1.51, and 0.87 to 2.00, respectively. From the designed and trained networks, the ANN-5 presented better results than other networks. Therefore, the ANN-5 was selected as the best estimation model for WAHD (eq. 6):

$$WAHD_{ANN-5} = \sum_{j=1}^{10} w_j^{(0)} f \left( \sum_{i=1}^{3} w_{ji}^{(b)} f(\text{net}_i) + w_0^{(b)} \right) + w_0^{(o)}$$

(6)

where $WAHD_{ANN-5}$ is the predicted value of the WAHD using the ANN-5, and $f(\text{net})$ is tanh ($\text{net}$). Hence, the (1- 10-3) - MLP network, consists of three VIs as input variables (EVI2, NDWI, NDVI), ten neurons in the hidden layer and a single output variable (WAHD) was selected as the optimum network. The values of $R_p^2$, $\text{MAE}_p$ and $\text{RMSE}_p$ for this topology were obtained 0.925 (Fig. 4b), 0.70 and 0.87, respectively. Fig. 4a shows the comparison between predicted WAHD values (test data) with the measured data. It could be observed that the trends were similar and superimposed over the others in some parts. The second model that had good performance for WAHD estimation was ANN-8 ($R_p^2$ =0.882; $\text{MAE}_p$=1.05; $\text{RMSE}_p$=1.11). Three VIs (NDVI, EVI2, and NDWI) were common parameters for both models ANN-5 and ANN-8. In other words, NDVI, EVI2, and NDWI can predict WAHD better than other indices.

ANN-2 ($R_p^2$=0.706; $\text{MAE}_p$=1.51; $\text{RMSE}_p$=2.00) which included GreenNDVI, CVI, and CI green had low accuracy for WAHD modeling. Therefore, it can be concluded that GreenNDVI, CVI, and CI green are not good indicators for WAHD prediction. In addition, ANN-3 model with individual input (NDWI) did not have high accuracy estimation for WAHD ($R_p^2$=0.611; $\text{MAE}_p$=1.41; $\text{RMSE}_p$=1.86). But, it can be seen in ANN-7 model that adding NDWI to GreenNDVI, CVI, CI green VIs (ANN-2) led to better estimations for WAHD prediction ($R_p^2$=0.759; $\text{MAE}_p$=1.22; $\text{RMSE}_p$=1.49). Besides, by adding the NDWI to vegetation indices of ANN-1, the accuracy of model improved ANN-8 (Table 5). In total, it is concluded that NDWI composition with other indicators could improve the forecast of WAHD.

**HSR map**

The HSR map using all data is shown in Fig. 5b, where ANN-5 model has been used for estimation as respects it had the best prediction accuracy (as explained in detail earlier). Also, the comparison between measured and estimated HSR maps is presented in Fig. 5, where the ANN-5 shows to have made good estimations of WAHD. Also, the trend of harvesting dates for different farms is the same in the two maps (Fig. 5).

As shown in estimated and measured maps (Fig. 5), the WAHD of northern regions is after the southern regions. Therefore, the harvest operation should start from the farms in southeast regions in the maps and completed in northwest farms. The differences in sowing dates, topographic variations, and the angle of sunlight, temperature, precipitation and soil moisture

<table>
<thead>
<tr>
<th>Prediction model</th>
<th>Best structure of MLP</th>
<th>$R_p^2$</th>
<th>$\text{MAE}_p$</th>
<th>$\text{RMSE}_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN-1 (NDVI, SAVI, EVI, EVI2)</td>
<td>1-10-4</td>
<td>0.857</td>
<td>1.06</td>
<td>1.12</td>
</tr>
<tr>
<td>ANN-2 (CVI, GreenNDVI, CI green)</td>
<td>1-13-3</td>
<td>0.706</td>
<td>1.51</td>
<td>2.00</td>
</tr>
<tr>
<td>ANN-3 (NDWI)</td>
<td>1-8-1</td>
<td>0.611</td>
<td>1.48</td>
<td>1.86</td>
</tr>
<tr>
<td>ANN-4 (EVI2, NDWI, GreenNDVI)</td>
<td>1-12-3</td>
<td>0.844</td>
<td>0.99</td>
<td>1.08</td>
</tr>
<tr>
<td>ANN-5 (EVI2, NDWI, NDVI)</td>
<td>1-10-3</td>
<td><strong>0.925</strong></td>
<td><strong>0.70</strong></td>
<td><strong>0.87</strong></td>
</tr>
<tr>
<td>ANN-6 (NDVI, EVI2, SAVI, GreenNDVI, NDWI)</td>
<td>1-16-5</td>
<td>0.843</td>
<td>1.13</td>
<td>1.38</td>
</tr>
<tr>
<td>ANN-7 (GreenNDVI, CVI, CI green, NDWI)</td>
<td>1-14-4</td>
<td>0.759</td>
<td>1.22</td>
<td>1.49</td>
</tr>
<tr>
<td>ANN-8 (NDVI, SAVI, EVI, EVI2, NDWI)</td>
<td>1-13-5</td>
<td>0.882</td>
<td>1.05</td>
<td>1.11</td>
</tr>
</tbody>
</table>
amount are the reasons of WAHD variation among these farms.

The estimated (using test data of ANN-5) and measured WAHD maps are shown in Fig. 6a and 6b, respectively. As seen in Fig. 6, the ANN-5 could provide very good predictions of WAHD. However, there were few differences between measured and estimated WAHD, and one disagreement was labeled with a black circle in Fig. 6. Therefore, it can be concluded that the MLP model with 1-10-3 topology had very good performance for harvest date predictions.

Discussion

In this study, we present a method for predicting WAHD, based on Landsat-8 derived VIs and ANNs to optimize harvest campaign scheduling and farm management. The utilized approach is based on the spectral variation of wheat canopy during phenological stages. The VIs, using the combination of different spectral bands, could explain the vegetation conditions during wheat growth stages. The VIs used in this study were classified in three groups based on their sensitivity to green biomass (NDVI, SAVI, EVI, EVI2), the liquid water content of vegetation (NDWI) and leaf chlorophyll (CVI, GreenNDVI, and CI

In the present study, application of ANNs, as non-linear modeling techniques, had excellent performance for estimating WAHD. The performance and accuracy of MLP models to WAHD estimation was due to the key features of these type models: (i) the intrinsic abilities of ANNs, such as learning, cross-validating and flexible processing; (ii) generating non-linear patterns between inputs and outputs which lead to accurate estimations of complex and dynamic data; and (iii) the VIs included intensive data which correlate non-linearly with spatial based WAHD. Therefore, considering the nature of VIs data, the ANN modeling techniques are a very good substitute for linear statistical methods. As reported

Figure 4. (a) Estimated (ANN-5) and measured WAHD on testing data. (b) Measured vs. estimated WAHD values (ANN5).

Figure 5. Comparison between measured (a) and estimated (using ANN-5) (b) HSR maps for All data.
by Xie et al. (2009), the ANN models could provide an accurate estimation ($R^2=0.817$ and $\text{RMSE}=0.4236$) for predicting above ground grassland biomass based on Landsat ETM+. Also, other researches which have been done using ANN techniques and VIs, suggest that utilizing of ANNs in remote sensing applications, has been very successful in the agriculture sector (Li et al., 2007; Fortin et al., 2010; Pantazi et al., 2016).

The ANN-5 had the best estimation of WAHD. The input variables of this model were EVI2, NDWI, and NDVI. EVI2 could be recognized as the best index for WAHD prediction because of its high $R$ value with WAHD. EVI2 consists of NIR and red bands; strong contrast of leaf scattering in the NIR wavelength and chlorophyll absorption in the red wavelength, makes EVI2 very robust to explain the vegetation conditions. EVI2 not only has an improved sensitivity over high biomass, in comparison with the SAVI, but also minimizes soil influences (Jin et al., 2016). EVI2 can capture subtle changes in vegetation structure and condition, especially to discriminate between leaf area index surface and greenness for vegetation with various soil background reflectances (Rocha & Shaver, 2009). Therefore, EVI2 is a suitable indicator for WAHD prediction. Results achieved by Bolton & Friedl (2013) as well as by Wang et al. (2015), proved that EVI2 has a high ability to predict maize yield and estimate rice phenology, respectively. NDWI is sensitive to the total amounts of the liquid water content of vegetation canopies (Gao, 1996). The water content of wheat canopy varies during wheat growth stages, and NDWI could present these water content variations. As reported by Liu et al. (2006) and Bao et al. (2008), NDWI had a good correlation with wheat biophysical parameters and yield. Although in the present study NDWI did not have a good ability to WAHD estimation alone (ANN-3 model; $R^2=0.611; \text{MAE}=1.41; \text{RMSE}=1.86$), adding NDWI to other indices improved the estimation power of models (ANN-2 vs. ANN-7 and ANN1 vs. ANN8). NDVI is the most practical indicator which is widely used in vegetation monitoring. According to Suwannachatkul et al. (2014), the use of NDVI to estimate rice harvest date resulted in acceptable accuracy (about eight days). NDVI was the only indicator used for predicting. The results of the current study showed that the combination of VIs for WAHD prediction improved the estimation accuracy (about one day). This was also confirmed in the present study, where green biomass sensitive VIs (NDVI, SAVI, EVI, EVI2) could estimate WAHD better than chlorophyll sensitive VIs (CVI, GreenNDVI, CI green) by comparing models ANN-1 and ANN-2. Moreover, ANN-6 and ANN-7 represent this issue. Therefore, it is suggested that biomass-sensitive VIs be used in the harvesting date prediction researches.

The methodology applied in this paper is straightforward and just needs satellite imagery bands to predict WAHD. The substantial advantage of the used method is its relative simplicity which makes it suitable for regional scale applications. The present method can be extended to other regions of the world. Also, other satellite sensors can be applied with this method. Moreover, the current model (ANN-5) may be helpful to regions with similar climate, where the sowing and harvesting dates are similar to the study region. The main limitation of this method is the accessibility of valid satellite imagery. The cloudy weather disrupts access to satellite images. A solution to this problem is the simultaneous use of several satellite images (Shang et al., 2014; Wang et al., 2015).

The current study presents a preliminary investigation of WAHD using satellite imagery. This method can be applied to other crops considering the crop conditions. Also, the combination of satellite remote sensing...
models, crop models, and meteorological statistical models will improve predictions.

References


