African Journal of Biotechnology Vol. 11(10), pp. 2518-2527, 2 February, 2012
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB11.2733

Full Length Research Paper

Spectroscopic determination of leaf water content using linear regression and an artificial neural network

Xiangwei Chen[1], Wenting Han[2,3,4]* and Min Li[2]

[1]College of Food Science and Engineering, Northwest A&F University, Yangling, China 712100.
[2]Institute of Efficient Water Use for Arid Agriculture of China, Northwest A&F University, Yangling, China.
[3]Institute of Soil and Water Conservation, Chinese Academy of Sciences and Ministry of Water Resources, Yangling, China.
[4]National Engineering Laboratory for Crop Water Efficient Use, Yangling, China.

Accepted 3 January, 2011

In order to detect crop water status with fast, non-destructive monitoring based on its spectral characteristics, this study measured 33 groups of peach tree leaf reflectance spectra (350 to 1075 nm). Linear regression and backpropagation artificial neural network methods were used to establish peach tree leaf water content and perform quantitative analyses between spectral indices. The results show that a linear relationship existed between the peach tree leaf water content (relative water content and equivalent water thickness) and its leaf reflectance spectral index. The models performed satisfactorily and could be used to detect the water content of the peach tree leaves.

Key words: Spectroscopy, crop water, linear regression, artificial neural network.

INTRODUCTION

Water-saving irrigation is a fundamental measure that is performed to ease shortages of water resources and ensure sustainable agricultural development. The localized, precise, rapid, and continuous collection of information concerning crop water is an important foundation for the implementation of irrigation strategies and management; this information is also an important consideration in modern agricultural technology and precision irrigation systems.

Clevers et al. (2008) used hyperspectral imaging technology to study plant canopy water content and showed that a spectrum at 970 nm produced good test results. Cheng et al. (2011) also used a near-infrared spectrum (350 to 2500 nm) to detect the water content of plant leaves and established the reflectance, leaf dry matter content, and fresh material content using a continuous wavelet analysis method. However, the correlation coefficient obtained in this study was lower (R^2= 0.62-0.75). In addition, Sancho-Knapika et al. (2011) investigated the use of microwave (1730 MHz) technology to detect the water content of poplar leaves and compared their results with those obtained using near-infrared (1350 and 1450 nm) detection.

This study focuses on peach trees; the experiments described herein measured the leaf spectral reflectance, correlation coefficients for the leaf water content, leaf spectral reflectance and spectral index were analyzed. Linear regression and a backpropagation (BP) artificial neural network were established and the spectral index was used to determine the water content of the leaves and consequently to provide the basis for decisions regarding water-saving irrigation for the peach tree.

MATERIALS AND METHODS

Sample collection

The samples were collected on May 20, 2007. The sampling location was in the water-saving technology testing area of the Northwest Agricultural and Forestry University, Yangling, Shaanxi. The Lady Red Peach tree strain was selected and tested after fruiting. A total of 33 living peach tree leaves were randomly selected (old and new leaves were from the same tree), and good

*Corresponding author. E-mail: hanwt2000@126.com. Tel: 0086 29 8709 1325. Fax: 0086 29 8701 1354.

leaves were marked (leaf number, time, etc.) and picked after the spectral reflectance test. The collected leaves were promptly transferred into a cooler with ice to preserve their freshness. The leaf samples were then transported back to the laboratory for additional tests.

Spectral measurements

A portable visible and near-infrared FieldSpec® UV/VNIR HandHeld Spectrometer was used to conduct the spectral measurements. The range of the wavelength was 325 to 1075 nm, and the spectral resolution was 1 nm. The measurements were conducted in sunny, cloudless and windless weather between 10:00 am and 2:00 pm Beijing time to ensure that the proper solar altitude angle and light intensity were present. Before conducting measurements, a standard panel was used to calibrate the instrument. During measurements, the sensor probe was placed vertically downward, approximately 10 cm from the height of the leaves. A total of 10 dark current samples were collected, while the white reference corrected 10 collected samples. Overall, a total of 10 samples were measured for each leaf. The water content and spectral absorption values were obtained from a total of 33 groups of leaves, of which 18 constituted the control group, while 15 comprised the predictive group.

Leaf water content measurements

The oven-drying method was used to determine the relative water content. First, an electronic scale with an accuracy of 0.001 g was used to weigh the bag containing the leaf samples, and the weight (w_1) was recorded. After weighing the leaves and the bag, the bagged leaves were placed in an oven for 30 min at 85°C for fixation and re-weighed (w_2). The bagged leaves were then placed back into the oven, and the temperature was maintained at 60°C for 24 h for complete drying; the total weight of each of the dried leaves and the paper bag was then recorded (w_3). The fuel moisture content (FMC) of each of the leaves was calculated as follows:

$$FMC = \frac{w_2 - w_3}{w_2 - w_1} \times 100\% \tag{1}$$

Dong et al. (2006) have noted that another commonly used indicator of leaf water content is the equivalent water thickness (EWT). This indicator is related to the leaf surface area. In this study, the fresh and dried weight of each sample leaf was first measured, and the surface area (s) was then calculated. The formula to determine the EWT is as follows:

$$EWT = \frac{w_2 - w_3}{s} \times 100\% \tag{2}$$

Data processing

Microsoft Excel was used for initial processing of the measured data, and Matlab was used for statistical analysis. Initially, researchers used canopy spectra as the most direct statistical method for extracting water content data. However, Shen et al. (2005) found that prediction equations vary with the time and place and are affected by many factors; identifying a universal relationship is difficult because the derived relationship depends on the mechanism used by that particular method. The canopy spectrum is a mixed spectrum of vegetation and background; the use of a mixed spectrum and vegetation parameters in establishing a relationship is affected by too many random factors, and developing a theoretical explanation is thus very challenging. Duan and Meno (2007) pointed out that the internal physical mechanism of the vegetation must be considered to establish the spectral index; an effective spectral index can differentiate between vegetation information and environmental background information.

The ratio vegetation index (RVI) is more stable than is single-wave information for monitoring vegetation because it reinforces the difference between near-infrared and infrared reflectance of the vegetation. Zhang et al. (2006) reported that the water index (WI = R_{970} / R_{900}) could clearly indicate changes in water status. Rouse et al. (2009) also reported that the normalized difference vegetation index (NDVI) is the most commonly used vegetation index and that the NDVI could eliminate most of the changes due to instrument calibration, sun angle, terrain, cloud shadow, and the irradiance of the atmospheric conditions; and this effect enhances instrument response to vegetation.

In this experiment, the spectrometer wavelength range was 325 to 1075 nm; within this range, a reflection valley exists near 980 nm, and the 950 and 970 nm wavelengths are sensitive to water. Therefore, in this study, a wavelength of 900 nm was selected as the reference wavelength, and the water indices were defined as WI_{950} and WI_{970}; the reflectance outside of 460 and 810 nm was used to construct the ratio vegetation index R (810, 460), and the reflectance outside of 610 and 810 nm was used to construct a normalized index of the NDVI (810, 610).

(1) Water index (WI):

$$WI_{xxx} = R_{900} / R_{xxx} \tag{3}$$

(2) Ratio vegetation index (RVI):

$$RVI = R_{NIR} / R_{\mathrm{Re}d} \tag{4}$$

(3) Normalized difference vegetation index (NDVI):

$$NDVI = \frac{R_{NIR} - R_{\mathrm{Re}d}}{R_{NIR} + R_{\mathrm{Re}d}} = \frac{RVI - 1}{RVI + 1} \tag{5}$$

In these equations, R_{NIR} and R_{Red} represent the reflectance of crops in the near-infrared and the infrared domains, respectively.

Modeling

Simple linear regression

Simple linear regression is the most basic modeling approach; because the modeling construction is simple and the model is reliable, this approach is frequently utilized by researchers. The correlation coefficient (R) is calculated as follows:

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{S_{xy}}{S_x S_y} \tag{6}$$

In this equation, y_i and $\bar{y}$ are the measured and mean values,

Table 1. Results for leaf water content tested by the oven-drying method.

Parameter	Mean	Standard deviation	Maximum	Minimum
Calibrated leaf water content (%)	64.5956	1.7867	68.4211	61.6883
Calibrated EWT (kg/m^2)	0.1217	0.0205	0.1597	0.0863
Predicted leaf water content (%)	65.2528	1.5285	67.7686	64.1111
Predicted EWT (kg/m^2)	0.1275	0.0150	0.1665	0.1027
Total leaf water content (%)	64.8943	1.6818	68.4211	61.1111
Total EWT (kgm^2)	0.1243	0.0182	0.1665	0.0863

EWT, Equivalent water thickness.

respectively, of the objective function (soil moisture); x_i and $\bar{x}$ are the measured and mean values, respectively, of the independent variable (absorbance level); S_x and S_y are the estimated standard deviations of the variables X and Y, respectively; and S_{xy} is the covariance of X and Y.

The measured coefficient R^2, the residual standard deviation σ_p and the residual variation coefficient c_p are the main determining indicators of the model. The R^2 value represents the degree of fit: as the magnitude of R^2 increases, the ability of the independent variables to explain the dependent variable increases. As the percentage of changes that are due to the independent variables increases, the density of the points observed near the regression line also increases.

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (7)$$

$$\sigma_p = \sqrt{\frac{1}{n-2}\sum(y_i - \hat{y}_i)^2} \quad (8)$$

$$c_p = \frac{\sigma_p}{\bar{y}} \quad (9)$$

Here, $\hat{y}_i$ is the dependent variable calculated according to the model. The model used the F test and the t test to study the overall linearity of the model and the significance of the regression coefficients.

BP networks: A perceptron is a feed-forward neural network that has a typical neural network structure. A perceptron has a hierarchical structure; information is input into the network and is then passed forward layer by layer to the output layer. Multilayer perceptrons are used to solve linearly inseparable problems. The multilayer perceptrons of the error reaction propagation algorithm are by far the most widely used class of networks. In this experiment, the key is to ascertain if there is a linear model between the spectral index and the moisture conditions of the sample that does not exhibit a non-linear relationship. Theoretical analysis shows that the perceptron with a single hidden layer can be mapped to all continuous functions, so a three-layer artificial neural network was chosen for construction. Nodes were hidden using the trial and error method, based on the following empirical formula:

$$m = \sqrt{n + l} + \alpha \quad (10)$$

Where m is the hidden nodes, n is the input layer nodes, l is the output nodes, and α is a constant between one and nine.

Model accuracy evaluation parameters

To determine whether the calibration equation achieved the general requirements of high accuracy and stable computing, a set of sample predictions was used to establish that the equation was calibrated and to evaluate its accuracy. Lu et al. (2004) reported that the near-infrared predicted values, the residual standard deviation, and the average residuals of the traditional method can be used as indicators for evaluation of prediction models. The average residual bias of the model is:

$$Bias = r = \frac{1}{n}\sum r_i \quad (11)$$

The residual standard deviation (SEP) of the model is:

$$SEP = \sqrt{\sum \frac{(r_i - \bar{r})^2}{n-1}} \quad (12)$$

Where r_i is the difference between the predicted value of the pre-built near-infrared model and the analysis value of the traditional method and n is the number of samples evaluated.

RESULTS AND DISCUSSION

Statistical results for leaf water contents

Table 1 shows the results for leaf water content tested by the oven-drying method. Due to the fact that *in vivo* measurements were performed and the same types of peach tree leaves were collected at one collection time, the leaf water contents were consistent despite the presence of new and old leaves.

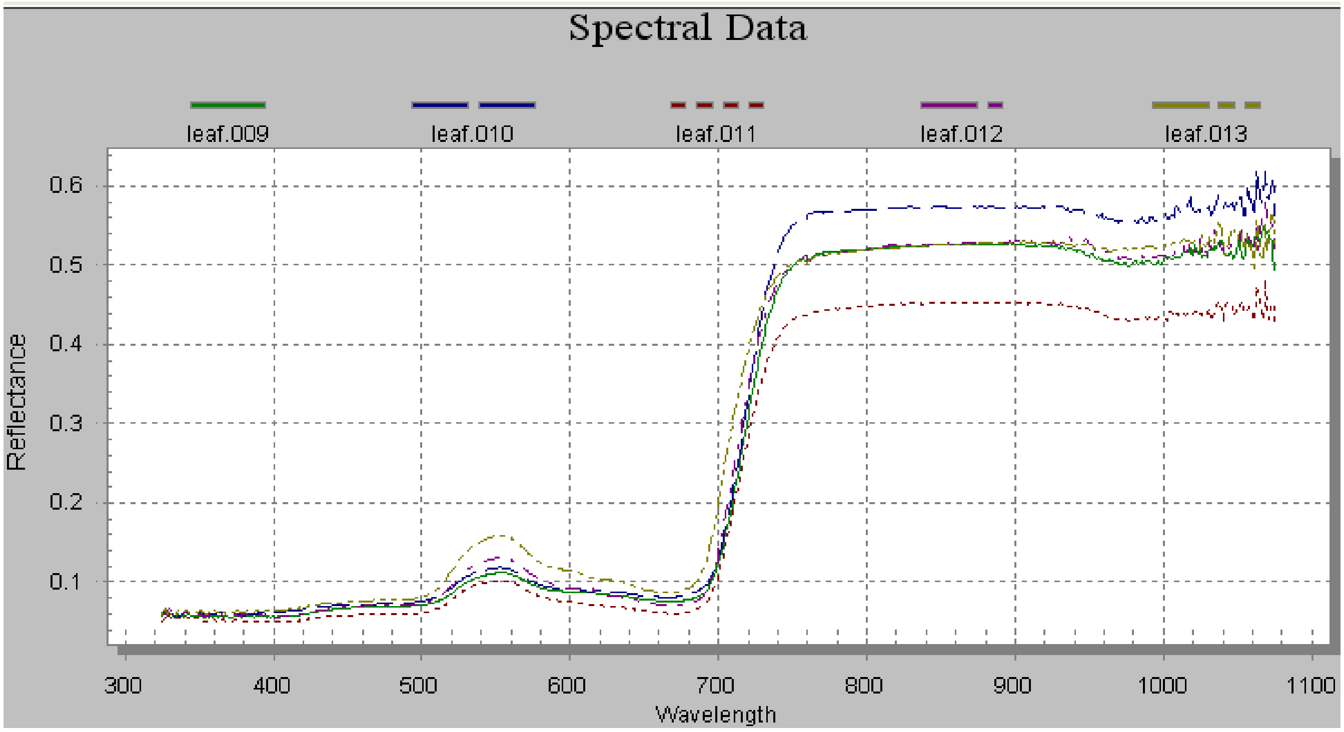

Figure 1. Spectral characteristics of leaf reflectance.

Table 2. Spectral indices and correlation coefficients for the calibrated FMC and EWT values of the leaves.

Spectral index	Relative water content	EWT
WI_{950}	0.0812	0.6172
WI_{970}	-0.0892	0.7131
R (810, 460)	-0.6611	-0.0757
NDVI (810, 610)	-0.7219	-0.0668

EWT, Equivalent water thickness; NDVI, normalized difference vegetation index; WI, water index; R, correlation coefficient.

Spectral characteristics of leaf reflectance

Figure 1 shows the measured leaf reflectance curve for a wavelength range of 350 to 1075 nm at a spectral resolution of 1 nm. Taking into account that the gradual regression of the wavelength selection lacks logical causality (Cheng et al. 2011), that is - the spectral reflectance of the leaf is susceptible to the impact of a variety of biochemical components, we used the spectral index to retrieve the leaf water content.

Regression analysis of spectral index and leaf water content

Table 2 shows the spectral index and the correlation coefficients for the calibrated FMC and EWT values of the leaves. As shown in the Table, the leaf water content (FMC) and the spectral indices R (810, 460) and NDVI (810, 610) had higher correlations. This result is similar with Dong et al. (2006). The leaf EWT values and the spectral indices WI_{950} and WI_{970} were also highly correlated. Therefore, models were constructed that used either the leaf FMC values and the spectral indices R (810, 460) and NDVI (810, 610) or the leaf EWT values and the spectral indices WI_{950} and WI_{970}. The results are shown in Table 3; the model R^2 values range from 0.38 to 0.52.

Spectral index prediction of leaf water content and model evaluation

The model was tested with a prediction set, and the results are shown in Table 4. Using the established models for the leaf EWT with either the WI_{950} or the WI_{970} indices, the correlation coefficient R and the model R^2 values were higher when the WI_{970} index was used, and the residual standard deviation (SEP) of the model and the predicted standard deviation bias were smaller, indicating that use of the WI_{970} index was better for predicting the peach tree leaf EWT. Using the models established for the leaf EWT with either the R (810, 460) or NDVI (810, 610) indices, the correlation coefficient R and model R^2 values were higher when the NDVI was used, and the predicted standard deviation bias was

Table 3. Regression model for FMC, EWT and spectral indices.

Parameter	WI_{950} and EWT	WI_{970} and EWT	R (810, 460) and leaf water content	NDVI (810, 610) and EWT
Model	y = 0.8242x - 0.6992	y = 0.5992x - 0.4810	y = -1.1540x + 72.8957	y = -39.4537x + 92.4447
R^2	0.38	0.51	0.44	0.52
Standard deviation of model residual	0.0166	0.0181	1.3819	1.2745
Variation coefficient of model residual	0.1367	0.1489	0.0214	0.0197

FMC, Fuel moisture content; EWT, equivalent water thickness; NDVI, normalized difference vegetation index; WI, water index; R, correlation coefficient.

Table 4. Spectral index prediction of leaf water content using the model.

Parameter	WI_{950}	WI_{970}	R (810, R460)	NDVI (810, 610)
Residual standard deviation (SEP)	0.0178	0.0122	1.4340	1.7035
Mean residual bias	0.0091	-0.0006	-0.5894	0.0272

NDVI, Normalized difference vegetation index; WI, water index; R, correlation coefficient.

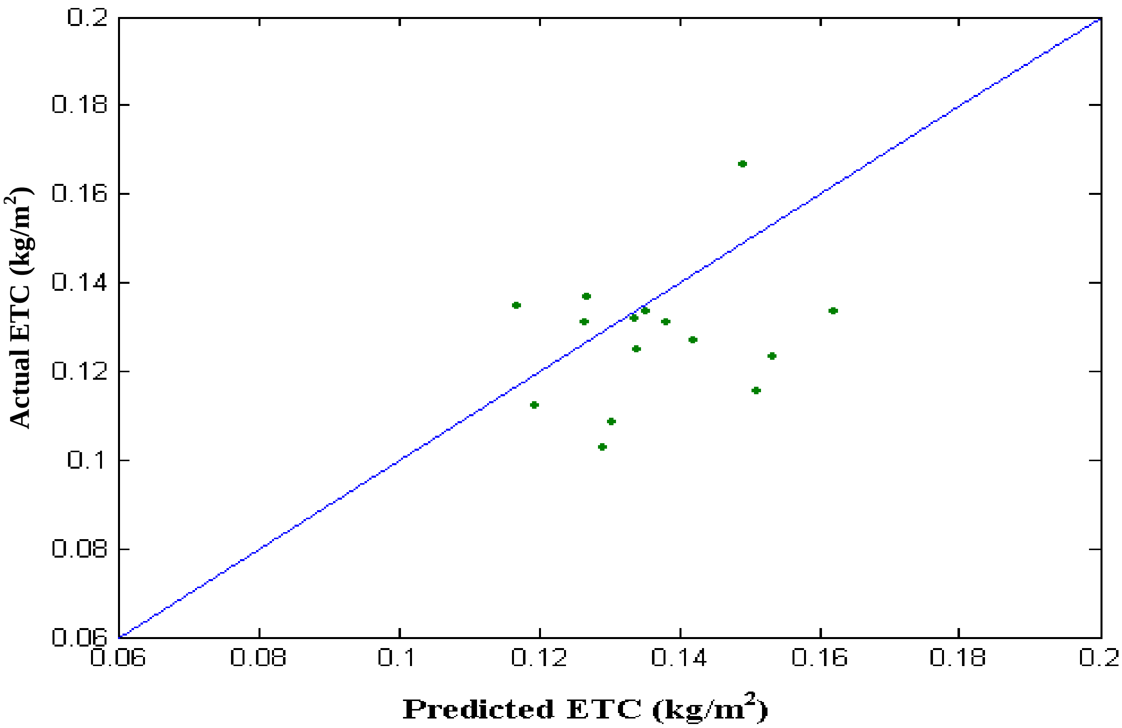

Figure 2. Equivalent water thickness prediction using WI_{950}. WI, Water index.

smaller; however, the SEP model residual standard deviation was higher. In general, the NDVI (810, 610) index performed better. The correlation coefficient for the NDVI (810, 610) and the leaf water content and the correlation coefficient for the WI_{970} and the leaf EWTwere at least 0.7, indicating that these relationships can be used to model and analyze peach tree leaf water status. Figures 2 to 5 illustrate the prediction maps of the model.

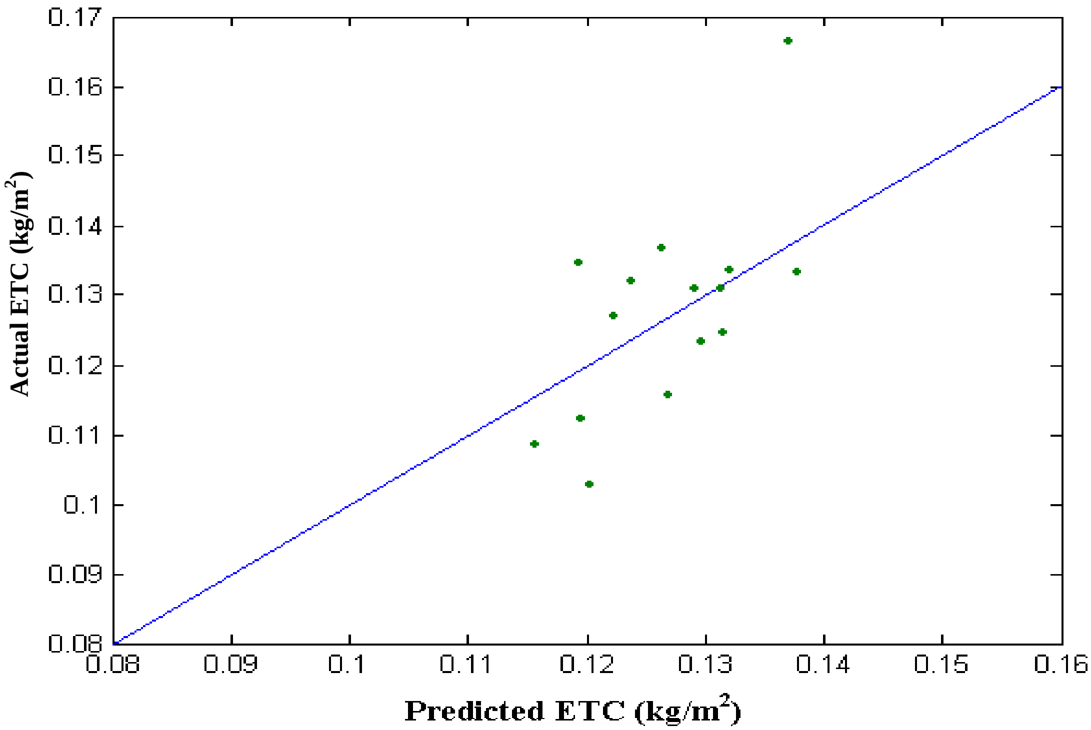

Figure 3. Equivalent water thickness prediction using WI_{970}. WI, Water index.

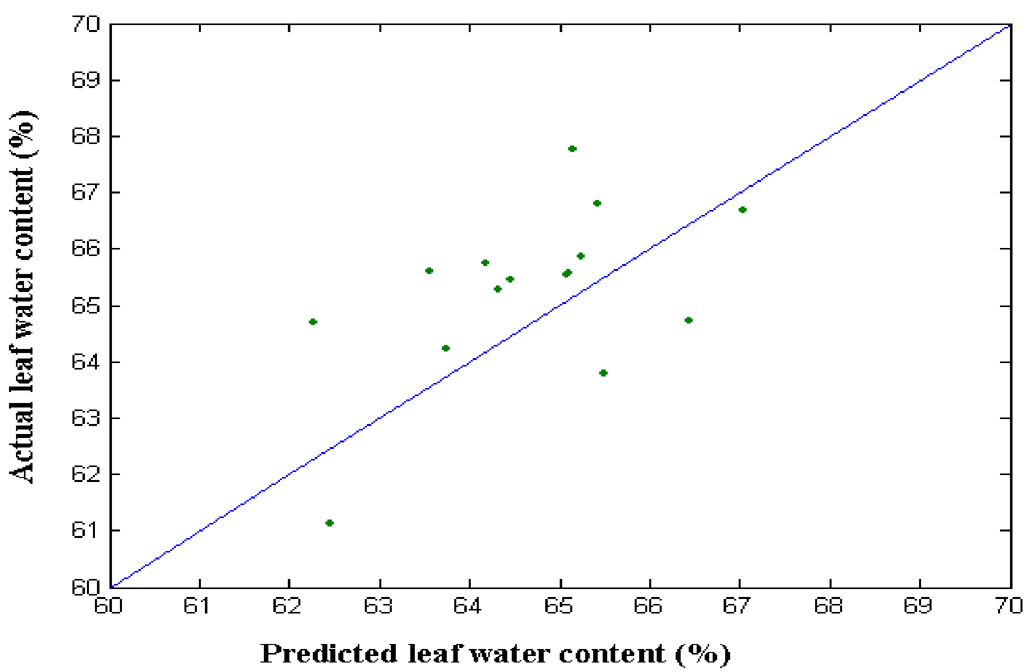

Figure 4. Leaf water content prediction using R (810, 460). R, correlation coefficient.

A multi-layer perceptron based on the BP algorithm for modeling

The regression analysis described above used the spectral reflectance at wavelengths of 460, 610, 810, 900, 950 and 970 nm as a multi-layer perceptron input to create two three-layer perceptrons and to construct models for the EWT and water content. After several

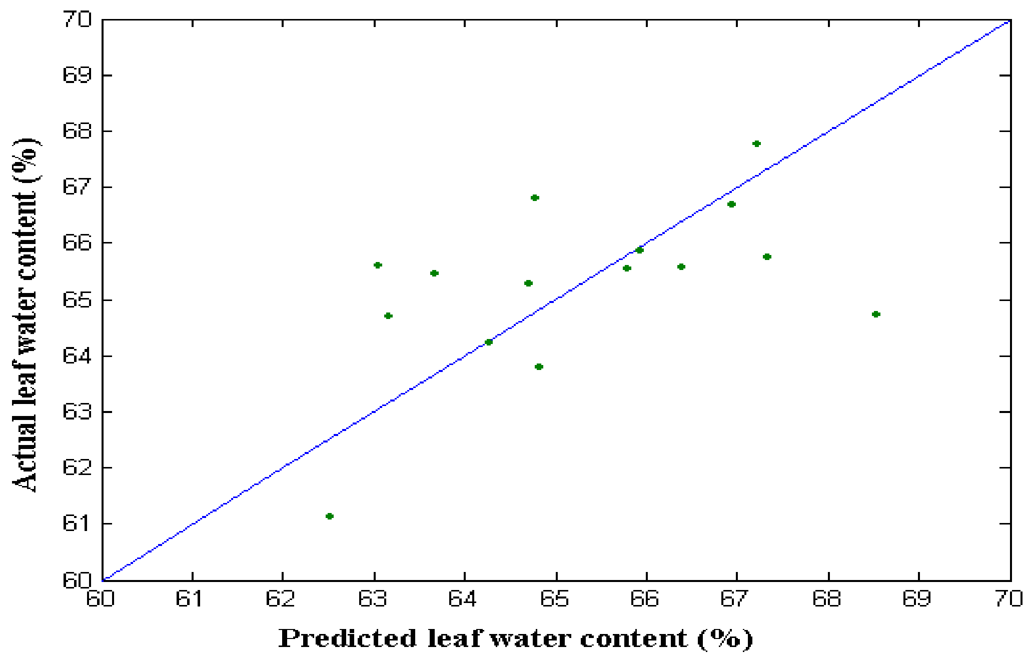

Figure 5. Leaf water content prediction using NDVI (810, 610). NDVI, Normalized difference vegetation index.

Table 5. BP network model of leaf moisture content, EWT, and spectral indices.

Parameter	EWT	Relative water content
Neural network layer	3	3
Hidden nodes	4	5
Mean square error of training set	0.00032	1.2832
Mean square error of prediction set	0.00026	2.7131
Residual standard deviation (SEP) of tested model	0.00083	0.0385
Residual mean of bias of tested model	0.0172	1.7688

EWT, Equivalent water thickness; BP, backpropagation.

tests, results were obtained and are summarized in Table 5. Figures 6 to 9 illustrate the prediction maps of the model. Similar results were achieved with the obtained model and the values derived from the regression method.

Conclusion

A linear correlation was observed between the leaf water content (relative water content and EWT) and the leaf reflectance spectral index of peach tree leaves. The correlation coefficients for the water indices and the EWT at wavelengths of 950 and 970 nm were 0.6172 and 0.7131, respectively. The correlation coefficient for the RVI and the relative water content ratio at wavelengths of 810 and 460 nm was -0.6611, and the correlation coefficient for the normalized index and relative water content was -0.7219; this latter relationship could be used to construct predictive models. The residual standard deviation (1.7035) and residual mean (0.0272) of the model for the normalized index and the relative water content at wavelengths of 810 and 610 nm were smaller, and the residual difference (0.0122) and residual mean (-0.0006) of the water index and the EWT model at a wavelength of 970 nm were smaller. The model using these parameters showed superior performance for detecting the water content of peach tree leaves.

ACKNOWLEDGEMENTS

This paper was founded by the Natural Science Research Project from the Department of Science and

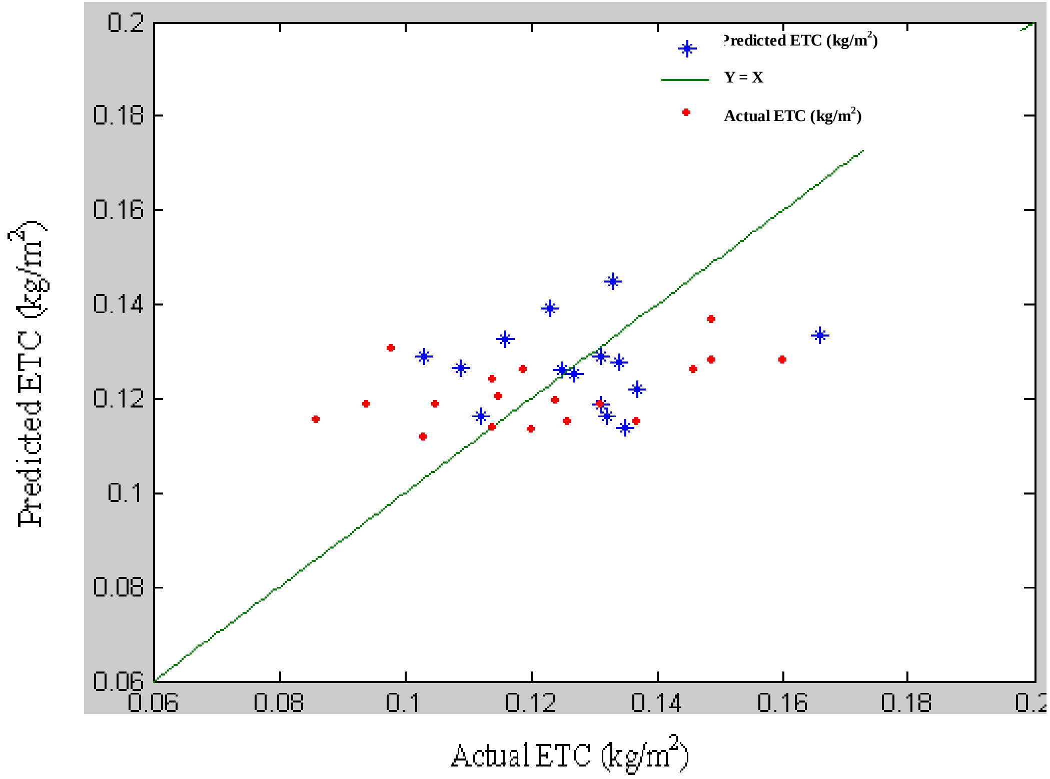

Figure 6. Predicted VS actual EWT. EWT, Equivalent water thickness.

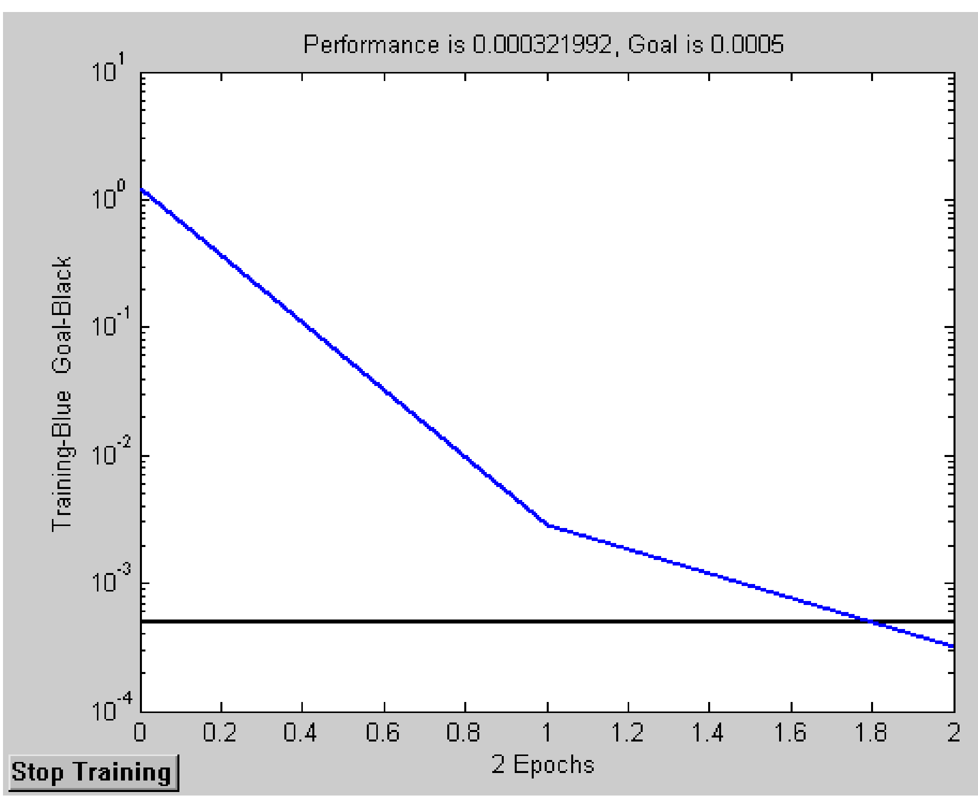

Figure 7. Decline curve of the modeling error of EWT. EWT, Equivalent water thickness.

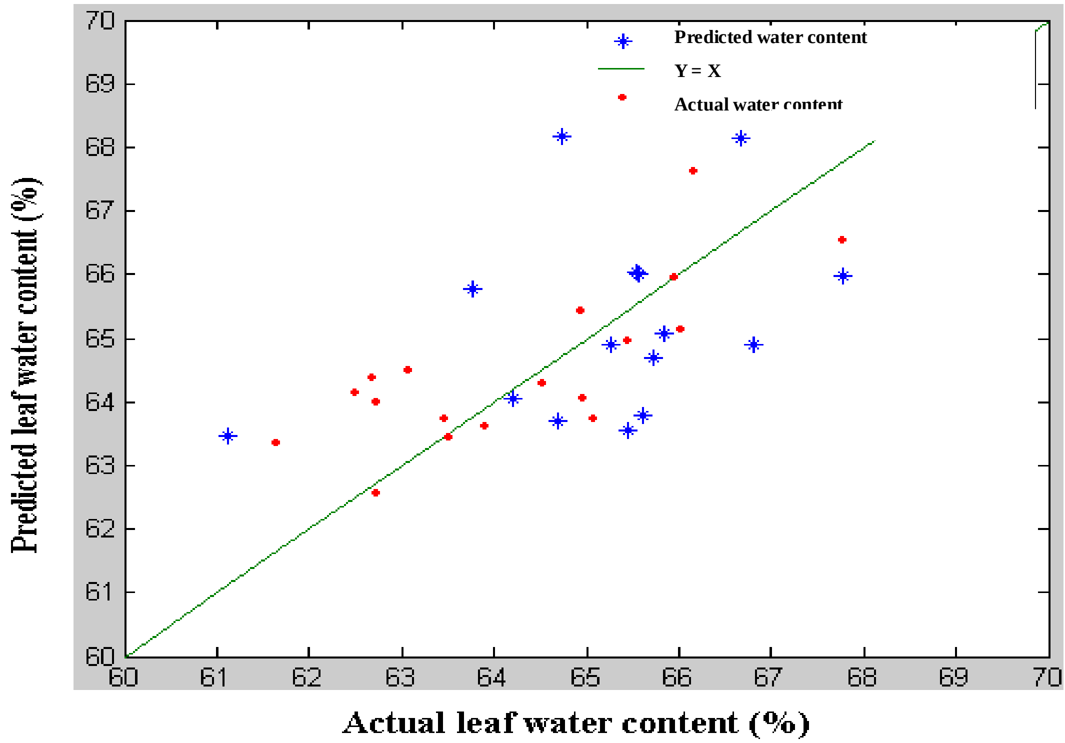

Figure 8. Predicted VS actual water content.

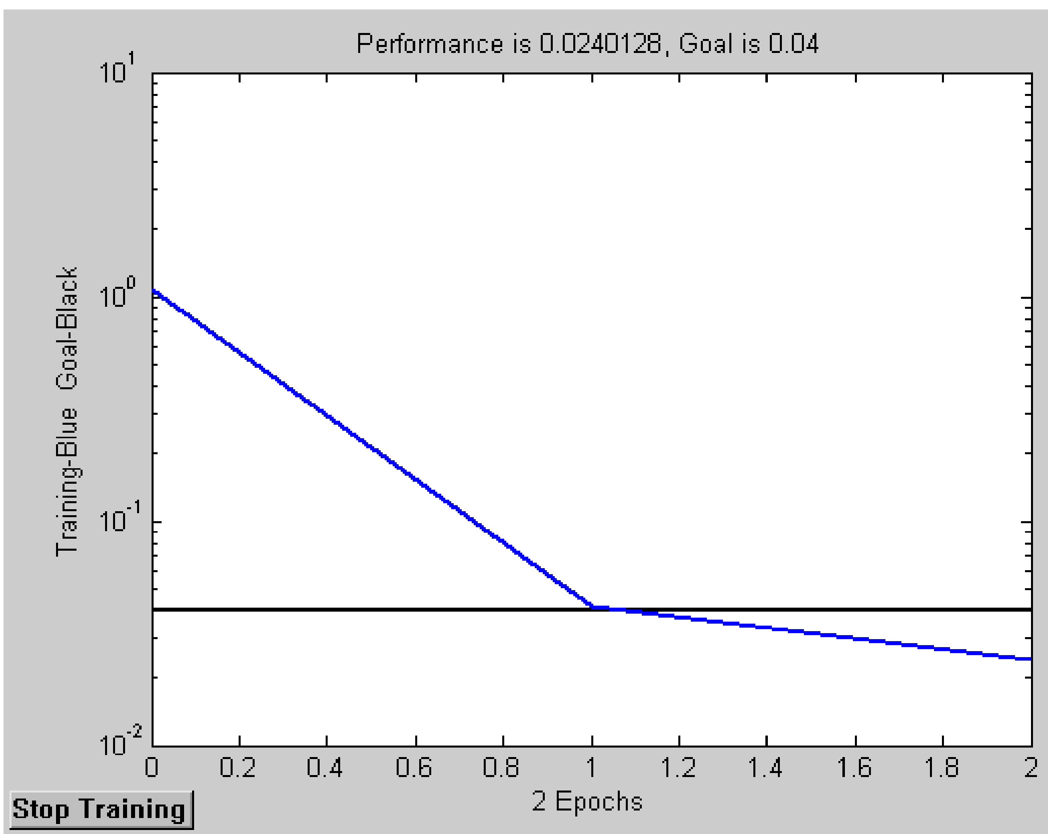

Figure 9. Decline curve of the modeling error of the relative water content.

Technology of Shaanxi Province (2011JM3005), the National Science and Technology Supporting Plan from the Ministry of Science and Technology of People's Republic of China (2011BAD29B08), the "111" Project from the Ministry of Education of People's Republic of China and the State Administration of Foreign Experts Affairs of People's Republic of China (B12007), and the Basic Research Supporting Plan from the Northwest A & F University (QN2009086).

REFERENCES

Cheng T, Rivar B, Sánchez-Azofeifa A (2011). Spectroscopic determination of leaf water content using continuous wavelet analysis. Remote Sens. Environ. 115: 659-670.

Clevers JGPW, Kooistra L, Schaepman ME (2008). Using spectral information from the NIR water absorption features for the retrieval of canopy water content. Int. J. Appl. Earth Obs. 10: 388-397.

Dong JJ, Niu Z, Shen Y, Yuan JG (2006). Comparison of the methods of obtaining leaf water content using spectroscopy reflectance data. Acta. Agric. Jiangxi. 28: 587-591.

Duan AW, Meno ZL (2007). Present situation of techniques and equipments of monitoring crop water status. Rev. China Agr. Sci. Tech. 9: 6-14.

Lu XJ, Pan JJ, Zhang JB (2004). Rice canopy spectral reflectance and leaf area index. Soils, 36: 648-653.

Rouse JW, Haas RH, Schell JA, Deering DW (2009). Monitoring vegetation systems in the Great Plains with ERTS1. Third Earth Res Technol Satellite Sym. 1: 309-317.

Sancho-Knapik D, Gismero J, Asensio A, Pequero-Pina JJ, Fernandez V, Alvarez-Arenas T, Gil-Pelegrin GE (2011). Microwave L-band (1730 MHz) accurately estimates the relative water content in poplar leaves. A comparison with a near infrared water index (R_{1300}/R_{1450}). Agr. Forest Meteorol. 151: 827-832.

Shen Y, Niu Z, Yan C (2005). Estimation models for vegetation water content at both leaf and canopy levels. Chin. J. Appl. Ecol. 16: 1218-1223.

Zhang J, Duan A, Sun J, Meng Z, Liu Z (2006). Advances in automated monitoring and diagnosis of crop water status. Transactions of the CSAE. 22: 174-178.