Journal of Food Biosciences and Technology,
Islamic Azad University, Science and Research Branch, Vol. 6, No. 1, 13-22, 2016

A Comparative Study Concerning Linear and Nonlinear Models to Determine Sugar Content in Sugar Beet by Near Infrared Spectroscopy (NIR)

S. Minaei[a*], H. Bagherpour[b], M. Abdollahian Noghabi [c], M. E. Khorasani Fardvani [d], F. Forughimanesh [e]

[a] *Associate Professor of the Department of Biosystems Engineering, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran.*
[b] *Assistant Professor of the Department of Biosystems Engineering, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran.*
[c] *Associate Professor of Agricultural Research and Education Organization, Sugar Beet Seed Institute, Karaj, Iran.*
[d] *Assistant Professor of the Department of Agricultural Machinery Engineering, Shahid Chamran University, Ahvaz, Iran.*
[e] *Member of the Department of Agronomy and Plant Breeding, Shahed University, Tehran, Iran.*

Received: 12 July 2015 Accepted: 15 November 2015

ABSTRACT: This paper reports on the use of Artificial Neural Networks (ANN) and Partial Least Square regression (PLS) combined with NIR spectroscopy (900-1700 nm) to design calibration models for the determination of sugar content in sugar beet. In this study a total of 80 samples were used as the calibration set, whereas 40 samples were used for prediction. Three pre-processing methods, including Multiplicative Scatter Correction (MSC), first and second derivatives were applied to improve the predictive ability of the models. Models were developed using partial least squares and artificial neural networks as linear and nonlinear models, respectively. The correlation coefficient (R), sugar mean square error of prediction (RMSEP) and SDR were the factors used for comparing these models. The results showed that NIR can be utilized as a rapid method to determine soluble solid content (SSC), sugar content (SC) and the model developed by ANN gives better correlation between predictions and measured values than PLS.

Keywords: *Artificial Neural Networks, NIR Spectroscopy, Partial Least Squares, Soluble Solids Content.*

Introduction

Sucrose is the main product of sugar beet and represents about 95% of the total sugar in the crop and the price of sugar beet is completely dependent upon its sucrose content. Thus, determination of sucrose content of sugar-beet is very important. A wide variety of methods have been proposed for the determination of sucrose in beet sugar. Analytical methods for determination of sucrose in beet sugar are based on diffractometric and polarimetric measurements. However, these methods need extracting of molasses by leaching with water in different experimental conditions, therefore, they are time consuming and costly methods (Garrigues *et al*., 2000).

Recently, with the development of computer science and chemometrics, applications of NIRS technique have received more attention from food researchers. Sugar content, hardness and titratable acidity are three major parameters to determine the internal quality and characteristics of fruits and other agricultural products. For measurement of these parameters, several rapid methods are available such as, ultrasound (Patist & Bates,

*Corresponding Author: sminaee@gmail.com

2008), microwave absorption (Clerjon *et al*., 2003), Nuclear Magnetic Resonance (NMR) (Winning *et al*., 2008), and Near InfraRed Spectroscopy (NIRS) (Karoui & Baerdemaeker, 2007). Among these methods commercial producer have shown more interest in using Near-infrared technology. NIRS has been used to measure internal quality in a wide range of fruits and vegetables, such as melon (Dull *et al*., 1992), peach (Kawano & Abe, 1995) apple (McGlone *et al*., 2002; Yan-de & Yi-bin, 2004) and tomato (Shao *et al*., 2007). Lu (2001) conducted a study to predict the firmness and sugar content of apples and sweet cherries in the spectral region between 800 nm and 1700 nm with the use of an InGaAs detector. In this study, the NIR models gave excellent predictions of the sugar content of sweet cherries, with corresponding r values of 0.95 and 0.89 and SEP values of 0.71 and 0.65 Brix for Hedelfinger and Sam sweet cherries, respectively. In a study conducted by Park for prediction of soluble solids content of both Washington Delicious and Pennsylvania Gala apples, results showed good correlation with NIR diffuse reflectance data. The coefficients of determination for predicting soluble solids were 0.93 for Gala apples, and 0.966 for Delicious apples with NIR (800-1100nm) reflectance measurement (Park *et al*., 2004).

Calibration is a critical component of a NIR analysis system. Principal component regression (PCR) and partial least squares (PLS) are two calibration procedures most frequently used in a linear relationship between the target parameter and the intensity of spectral absorption bands. Regarding the high non-linearity, however, both above calibration techniques might lead to substantial errors. Hence an alternative chemometric tool must be used instead. In this study, artificial neural networks as nonlinear method were used to develop a calibration model. ANNs are widely used mathematical algorithms for overcoming non-linearity in calibration model and have been widely applied during the past several years (Dou *et al*., 2007).

The specific objectives of this study are to evaluate the potential of near-infrared spectroscopy to predict quality factors such as soluble solids and sugar contents of sugar beet as well as comparing PLS and ANN as two important methods to design calibration models for determination of these parameters in sugar beet.

Materials and Methods

- *Sample preparation*

One hundred and twenty sugar beet samples were harvested from the Sugar Beet Seed Institute's farm in Karaj, Iran. After washing, the Scalps were removed and one thin cross section of the root was cut out from the Crown part of the beet. The samples were sealed in separate plastic bags and transferred to the laboratory for spectra acquisition. Each sample was mounted on the sample holder and the signals were acquired from these layers.

- *Spectroscopic measurements and software*

Reflectance spectra of sugar beet samples were collected using a near-infrared scanning spectrophotometer (EPP2000NIR Stellar Net, Inc. Oldsmar, Fl) in the wavelength range of 900-1700 nm, with a data resolution of 2.5 nm. Each spectrum on the average consists of 15 individual readings. NIR spectra were collected in reflectance form using Spectra Wiz software reflectance form. The data were then exported from Spectra Wiz software and imported directly into The Unscrambler X 10.2 software (CAMO ASA, Oslo, Norway) for spectral processing and multivariate analysis.

- *Reference method*

After acquiring the spectra, samples were milled to make sure of tissue disintegration

and then were frozen at -20°C for 24 h. Samples were then thawed, the juice was filtered using filter paper and the sucrose content was determined using polarimetry method (Betalyser, Anton Paar Optotec). This technique establishes a correlation between rotation of the polarised light and the asymmetric centre of sucrose molecule. For the measurement of Brix and the percentage of dry matter in sugar beet juice, first the dark juice was filtered and then analyzed using a refractometer (ATAGO DR-A1). In this study, the number of calibration and test sets were 80 and 40 samples, respectively.

- *Spectral pre-processing and development of the model*

First, the spectra in reflectance(R) form were converted to absorbance (log (1/R)) values in order to obtain correlations between NIR spectra and SSC and SC. Then, the spectra were pre-processed using several tools such as first derivative, second derivative and multiplicative scatter correction (MSC) methods based on Savitsky-Golay algorithm, with five points smoothing filter. These tools were used to minimize the data noise, eliminate baseline offset and remove multiplicative interferences of scatter and particle size.

Once the preprocessing was completed, partial least squares method (PLS) was used to develop calibration models for predicting the SSC and SC. The best number of factors used by PLS was selected using leave-one-out cross-validation. Leave-one-out cross-validation consists of removing one sample from the calibration set and estimating its predicted value based on a model developed with all the other samples (da Costa Filho, 2009).

ANNs have high processing speeds, robustness, and generalization capabilities, and are able to deal with high dimensional data spaces. More particularly, ANNs incorporating supervised training algorithms such as feed-forward back-propagation networks are capable of distinguishing interesting features from voluminous and noisy data sets having distorted patterns (Zhai *et al*., 2006). In this study, multilayer feed forward neural networks with one input, one hidden layer and one output layer topology were selected for modeling the pretreated spectrum. The activation function for the hidden layer was logsig and a Linear function best suited the output layer. Algorithms such as back-propagation gradient descent and gradient descend with momentum are too slow for practical problems because they require small learning rates for stable learning, while faster algorithms such as conjugate gradient, quasi-Newton, and Levenberg–Marquardt (LM) use standard numerical optimization techniques. These algorithms eliminate some of the disadvantages mentioned above. LM algorithm uses the second-order derivatives of the cost function so that a better convergence behavior can be obtained (Ghobadian *et al*., 2009). In the ordinary gradient descent search, only the first order derivatives are evaluated and the parameter change information contains solely the direction along which the cost is minimized, whereas, the Levenberg–Marquardt technique extracts more significant parameter change vector. In this study the Levenberg–Marquardt (LM) technique was used as a faster algorithm to evaluate the results. For comparison of the models, root mean- square error of prediction (RMSEP) (eq1), correlation coefficient (R) (eq2) and SDR (eq3) were considered. In theory, SDR is a more direct indicator as compared with either R or RMSEP. The higher the SDR value, the greater the model's power (Liu *et al*., 2010).

$$\mathrm{RMSEP} = \sqrt{\frac{1}{n_p}\sum_{i=1}^{n_p}(\hat{y}_i - y_i)^2} \qquad (1)$$

$$R = \sqrt{\frac{\sum_{i=1}^{n_p}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n_p}(\hat{y}_i - y_m)^2}} \quad (2)$$

$$\text{SDR} = \frac{\text{SD}}{\text{RMSEP}} \quad (3)$$

where $\hat{y}_i$ is the predicted value of the ith observation; y_i the measured value of the ith observation; y_m the mean value of measured value in prediction set; n_p the number of observations in the prediction set and SD the standard deviation in the prediction set.

Results and Discussion

- *Spectral analysis*

Figure 1 shows the original spectra of sugar beet samples. Because of some noise in the 1550-1700 nm region, this region was not utilized to develop the calibration models.

Some spectral ranges contain more significant information: 1st overtone combinations of C-H and O-H elongation at 1400-1500 nm; 2nd overtone of C-H elongation at 1150-1200 nm; and 3rd overtone of C-H and 2nd overtone of O-H elongation at 900-1000 nm (Anonymous, 2005). There are three broad band peaks around 985 nm, 1190 nm, and 1450 nm. These absorption peaks are close to the three absorption wavelengths of pure water (958 nm, 1153 nm, and 1460 nm) (Lu, 2001), however, near 958nm and 1460nm, sugar has absorption bands that overlap with absorption peaks of water (Park *et al*., 2004; Mireei *et al*., 2010).

- *Development of model by PLS method*

After pre-processing of spectra, the calibration model was developed using Partial Least Squares Regression (PLS) method as linear regression. For developing PLS model, the number of PCs factors is very important. Using too many PCs, generates an over- fitted model with low RMSECV but perform poorly in the prediction set (Xie *et al*., 2009). The effect of the number of PCs on RMSECV is shown in Figure 2. As Figure 2 indicates, up to 6 and 7 latent variables, RMSECV decreased, but after these points increased. Thus, these values can be selected as the optimum for developing models for prediction of SSC and SC content. For fist-derivative and second- derivative preprocessing, the optimum latent variables are shown in Table 1.

For choosing the best model, first the spectra were preprocessed by three important preprocessing techniques and then their PLS models were developed. The results of PLS models are shown for both indices in Table 1.

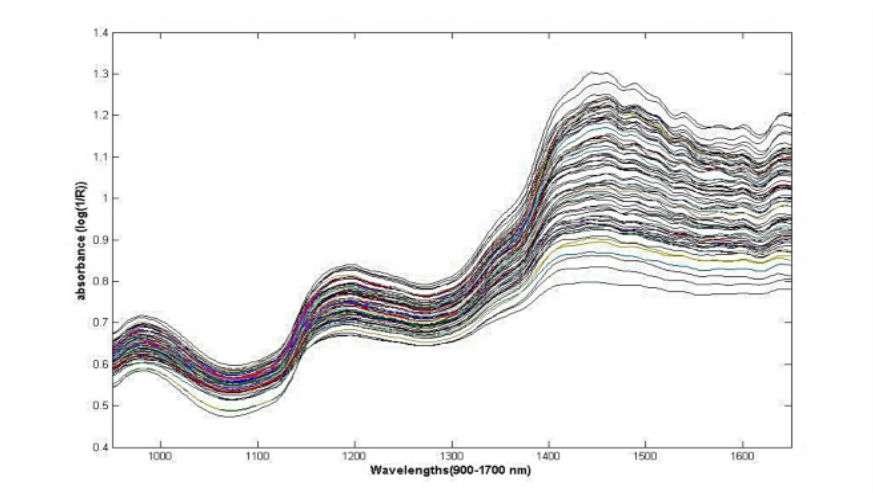

Fig. 1. Raw NIR spectra of sugar beet samples

Table 1 shows that values of SDR, R and RMSEP for MSC that are lower than those of other preprocessing techniques thus MSC can be selected as suitable preprocessing for prediction of both SSC and SC.

Figure 3 shows acceptable correlation between the actual and NIR predicted values. Also, they indicate that the PLS model could predict SSC better than SC. The results of this study are quite similar to those reported by researchers for fruits. Lu *et al* (2010). Predicted sugar content (R= 0:82, 0.78 and RMSEP = 0:56, 0.64) of Empire apple (Lu & Ariana, 2002) and determined the sugar content of Gannan navel orange using PLS regression (R= 0:88, 0:86 and RMSE= 0.46, 0.49), respectively for calibration and prediction sets (Liu *et al*., 2010). For comparing models and selection of the best one, SDR factor is better than R and RMSEP. As Table 2 shows, with MSC as pre-processing, SDR (2.66 and 2.56) is the highest for both indices.

- *Development of model using ANN method*

In the ANN analysis it is very important to work with a reduced number of input

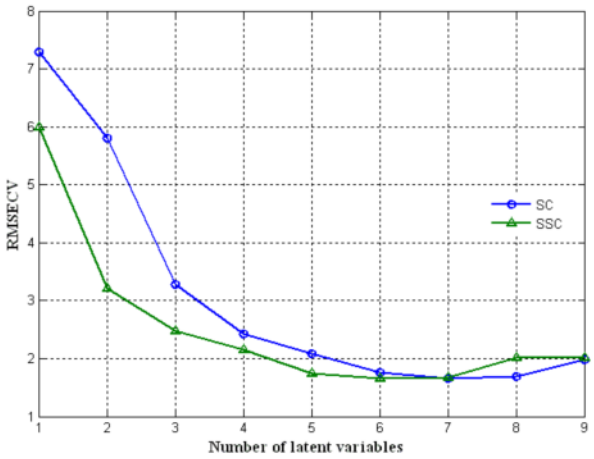

Fig. 2. Effect of the number of latent variables on RMSECV value

Table 1. Calibration and prediction results of PLS models for SSC (a) and SC (b)

(a)

Pre-processing	LVs	RMSEC	RMSEP	R	SDR
MSC	6	1.6	1.7	0.95	2.66
Fist-derivative	5	2.2	2.5	0.80	1.81
Second-derivative	4	1.7	1.9	0.84	2.38

(b)

Pre-processing	LVs	RMSEC	RMSEP	R	SDR
MSC	7	1.5	1.8	0.84	2.56
Fist-derivative	5	2.4	2.4	0.74	1.92
Second-derivative	6	1.9	2	0.80	2.30

Table 2. values of the explained variance vs number of PCs

PCs	1	2	3	4	5	6	7	8	9	10	11	12	13
Explained Variance	42	56	68.2	77.3	82.3	86.9	90.2	91.3	92.5	93.1	94	94.3	94.7

neurons to avoid memorizing effects. In this sense, it was necessary to reduce the number of data in each spectrum (Oliveira *et al*., 2006). Spectrum variables were reduced by using principal component analysis (PCA). PCs should be able to explain at least 85% of the variance in the NIR spectral matrix (He *et al*., 2006). For arriving at the best number of PCs, first, spectra were preprocessed by using known preprocessing methods such as MSC, 1st and 2nd derivatives. Then PCA analysis was conducted using Unscrambler software, In MSC preprocessing, Table 2 show that the first thirteen components represent 94.7% of the data variance and after 13 components, very little change is observed in this value. Thus by considering the analysis time and to avoid over-fitting, 13 PCs were selected as input vectors. Based on a similar analysis, the number of PCs for 1st and 2nd derivatives were obtained to be 7 and 9, respectively. In this study, MATLAB 7.2 neural network toolbox was used for ANN design, the training and testing performance MSE was chosen to be 0.00001 for all ANNs.

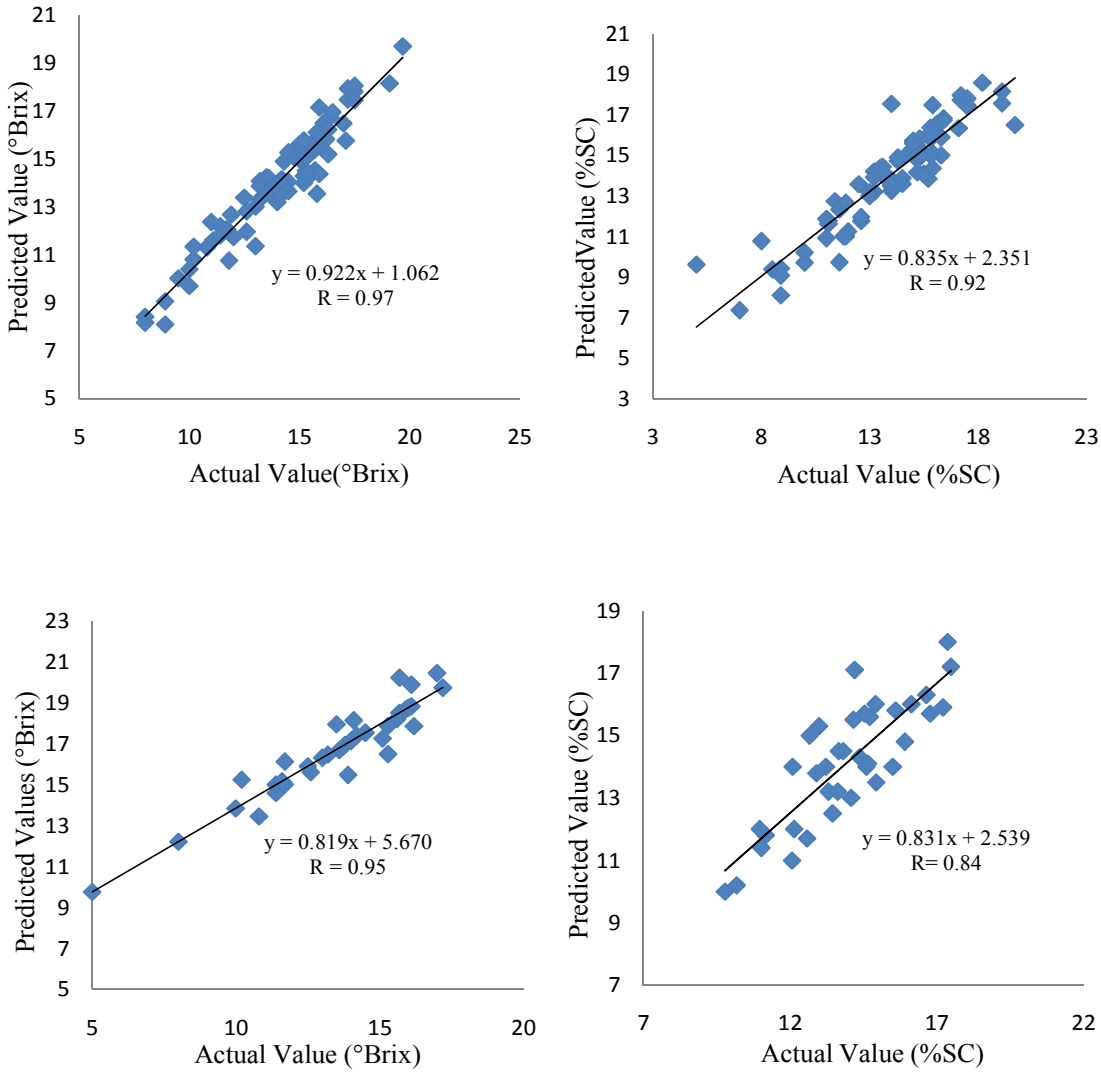

Fig. 3. Scatter plots of predicted versus actual values obtained by PLS (a, b) for calibration models and (c, d) for predicted models

Figure 4 shows the structural outline of the ANN used in this study. As stated earlier the structural outline consists of one input layer, one hidden layer and one output layer. The first selected PCs were used as input variables, the number of hidden neurons was changed from 4 to 24, and two neurons were taken for the hidden layer (desired SSC and SC). In the model, two-thirds of the data set was randomly assigned as the training set, while the remaining data were put aside for prediction and validation. For finding the optimum number of neurons in the hidden layer the R value was used as a comparison index. When the MSC was applied as preprocessing method, the R-value did not increase beyond 13 neurons in the hidden layer (Table 3). Hence, the network with 13 neurons in the hidden layer would be considered satisfactory. Similar tests were applied to data sets preprocessed using first derivative and second derivative, for which the optimum number of neurons was obtained to be 10 and 12, respectively.

For comparing PLS and ANN models, RMSE values for training and prediction data sets were computed and the results are given in Table 4. Comparison of Table 4 with Table 3, shows that RMSEP, R and SDR values in Table 4 are better than those that are obtained by PLS models. Thus the ANN is suitable for modeling the SSC and SC contents of sugar beet. Graphs a and b In Figure 5 show the correlation between the ANN model output and the actual values for the training set while graphs c and d show reasonably good correlation between the actual and predicted values.

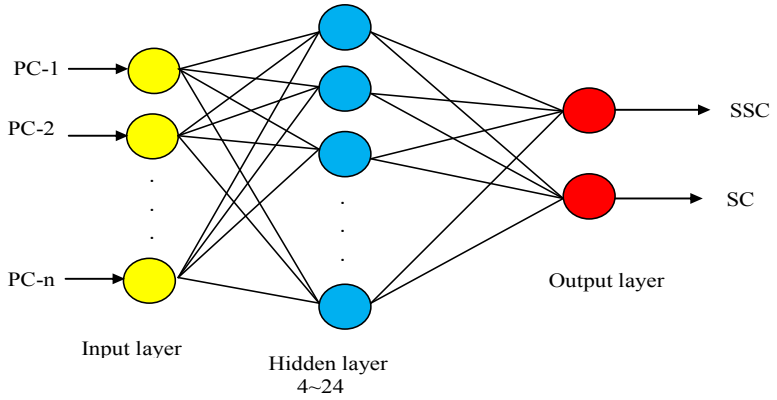

Fig. 4. Structural outline of the neural network for sugar content prediction

Table 3. The effect of the number of neurons on the network performance (MSC preprocessing)

Neurons in hidden layer	Training rule	RMSE(training)	R
7	Trainlm	1.864	0.864
8	Trainlm	1.524	0.862
9	Trainlm	1.231	0.874
10	Trainlm	1.107	0.923
11	Trainlm	1.102	0.934
12	Trainlm	1.075	0.937
13	Trainlm	1.087	0.951
14	Trainlm	1.645	0.951
15	Trainlm	1.974	0.901
16	Trainlm	1.436	0.872
12	Trainrp	1.845	0.920
12	Trainscg	1.746	0.853

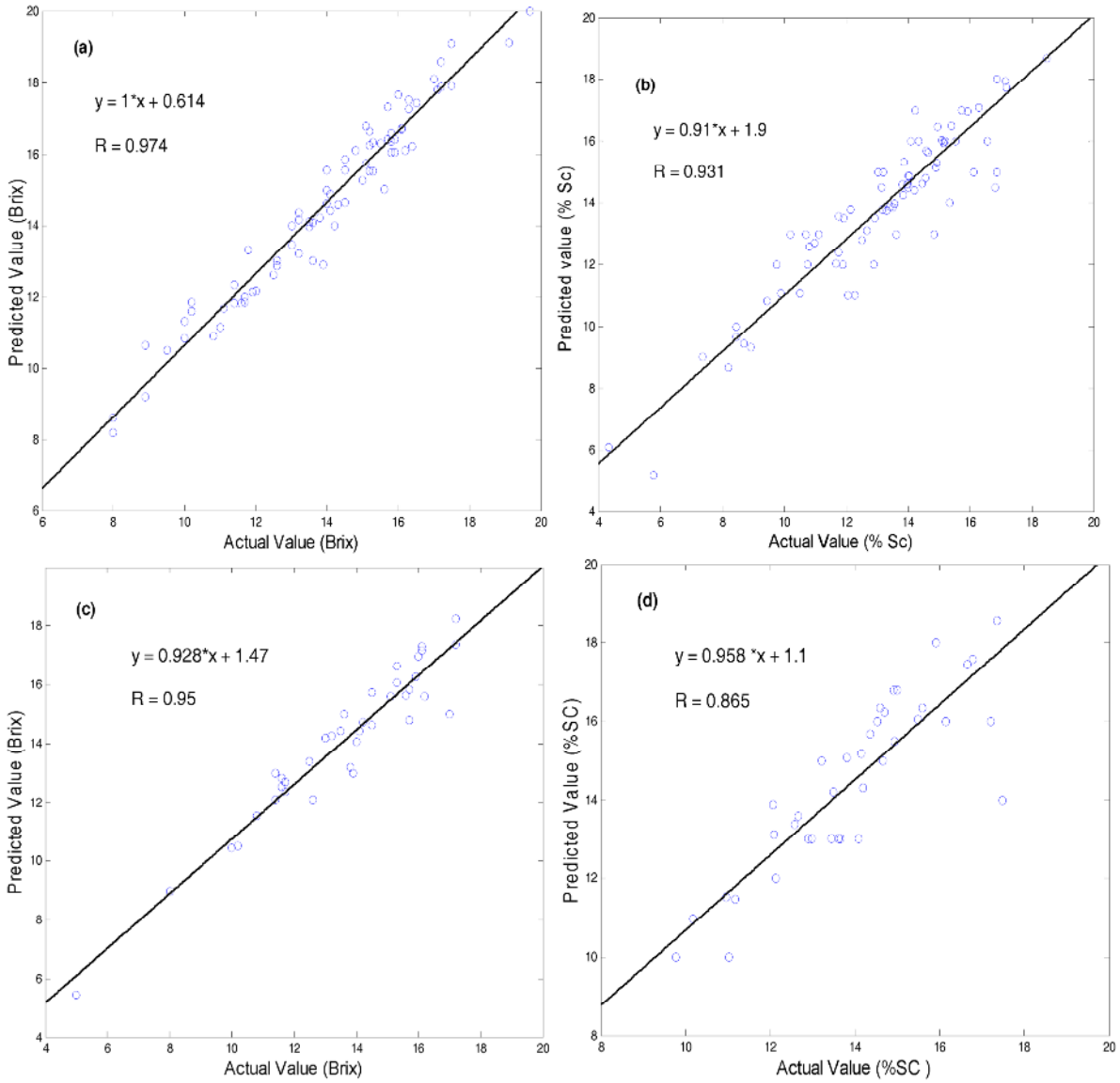

Fig. 5. Scatter plots of predicted versus actual values obtained by ANN (a, b) for calibration models and (c, d) for predicted models

Table 4. calibration and prediction results of ANN models for SSC(a) and SC(b).

(a)

Pre-processing	LVs	RMSEC	RMSEP	R	SDR
MSC	13	0.86	0.90	0.95	5.11
Fist-derivative	7	1.07	1.23	0.85	3.74
Second-derivative	9	1.32	1.38	0.81	3.33

(b)

Pre-processing	LVs	RMSEC	RMSEP	R	SDR
MSC	13	1.25	1.26	0.85	3.5
Fist-derivative	7	1.8	1.23	0.83	3.74
Second-derivative	9	1.46	1.85	0.79	2.48

Conclusion

In this study, NIR spectroscopy was utilized for assessing the sugar content of sugar beet. Also, Partial Least Square regressions (PLS) and Artificial Neural Networks (ANN) as linear and nonlinear regression analyses for estimating sugar beet soluble solid content (SSC) and sugar content (SC) were compared. Results have shown that NIR spectroscopy can satisfactorily predict the sugar and soluble solid contents of sugar beet. Both PLS and ANN models indicate that NIR spectroscopy can estimate SSC more accurately than SC. The developed models showed that the highest SDR and R values can be obtained with MSC preprocessing. In comparison with PLS models, ANN models can better predict sugar beet SSC and SC. Using MSC preprocessing with ANN, SDR values of 5.11 and 3.5 were obtained while in PLS models, these values were 2.66 and 2.56 for SSC and SC, respectively. Similar to SDR, MSCEP values were 0.9 and 1.26 in ANN while in PLS those were 1.7 and 1.8 for estimation of SSC and SC, respectively. These results indicate that ANN nonlinear models have a better potential for estimating sugar beet SSC and SC than do linear PLS models.

References

Anonymous. (2005). Near-IR Absorption Bands, Analytical Spectral Devices, www.asdi.com (accessed Sep. 2014).

Clerjon, S., Daudin, J. D. & Damez, J. L. (2003). Water activity and dielectric properties of gels in the frequency range 200 MHz–6 GHz. Food Chemistry, 82, 87-97.

Da Costa Filho, P. A. (2009). Rapid determination of sucrose in chocolate mass using near infrared spectroscopy. Analytica Chimica Acta, 631, 206–211.

Dou, Y., Zou, T., Liu T., Qu, N. & Ren, Y. (2007). Calibration in non-linear NIR spectroscopy using principal component artificial neural networks. Spectrochimica Acta Part A , 68, 1201–1206.

Dull, G. G., Leffler, R. G., Birth, G. S. & Smittle, D. A. (1992). Instrument for nondestructive measurement of soluble solids in honeydew melons. Transactions of the ASAE, 35(2), 735-737.

Garrigues, J. M., Akssira, M., Rambla, F. J., Garrigues S. & de la Guardia, M. (2000). Direct ATR-FTIR determination of sucrose in beet root. Talanta, 51, 247–255.

Ghobadian, B., Rahimi, H., Nikbakht, A. M., Najafi, G. & Yusaf, T. F. (2009). Diesel engine performance and exhaust emission analysis using waste cooking biodiesel fuel with an artificial neural network. Renewable Energy, 34, 976–982.

He, Y., Feng, S., Deng, X. & Li, X. (2006). Study on lossless discrimination of varieties of yogurt using the Visible/NIR-spectroscopy. Food Research International, 39, 645–650.

Karoui, R. & Baerdemaeker, J. (2007). A review of the analytical methods coupled with chemometric tools for the determination of the quality and identity of dairy products. Food Chem, 102, 621-640.

Kawano, S. & Abe, H. (1995). Development of a calibration equation with temperature compensation for determining the Brix value in intact peaches. Journal of Near Infrared Spectroscopy, 3, 211-218.

Liu, Y., Sun, X., Zhou, J., Zhang, H. & Yang, C. (2010). Linear and nonlinear multivariate regressions for determination of sugar content of intact Gannan navel orange by Vis_NIR diffuse reflectance spectroscopy. Mathematical and Computer Modelling, 51, 1438_1443.

Lu, R. (2001). Predicting firmness and sugar content of sweet cherries using near–infrared diffuse reflectance spectroscopy. American Society of Agricultural Engineers, 44, 1265–1271.

Lu, R. & Ariana, D. (2002). A near-infrared sensing technique for measuring internal quality of apple fruit. Applied Engineering in Agriculture, 18 (5), 585_590.

McGlone, A. V., Jordan, B. & Martinsen, P. J. (2002). Vis/NIR estimation at harvest of pre- and post-storage quality indices for 'Royal Gala' apple. *Postharvest Biology and Technology*, 25, 135-144.

Mireei, S. A., Mohtasebi, S.d., Massudi, R., Rafiee, S., Arabanian, A.S. & Berardinelli, A.

(2010). Non-destructive measurement of moisture and soluble solids content of Mazafati date fruit by NIR spectroscopy. AJCS, 4, 175-179.

Oliveira, J. S., Montalv˜ao, R., Daher, L., Suarez, P. A. Z. & Rubim, J. C. (2006). Determination of methyl ester contents in biodiesel blends by FTIR-ATR and FTNIR spectroscopies. Talanta, 69, 1278–1284.

Park, B., Abbott, J. A., Lee, K., Choi, C. & Choi, K. (2004). Near-infrared diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of delicious and gala apples. Transactions of the ASAE, *46*(6), 1721-1731.

Patist, A. & Bates, D. (2008). Ultrasonic innovations in the food industry: From the laboratory to commercial production, Food Sci. Emerg. Technology, 9, 147–154.

Shao, Y., He, Y., Gomez, A. H., Pereira, A.G., Qiu, Z. & Zhang, Y. (2007). Visible/near infrared spectrometric technique for non-destructive assessment of tomato 'Heatwave' quality characteristics. Journal of Food Engineering, 81, 672–678.

Winning, H., Larsen, F. H., Bro, R. & Engelsen, S. B. (2008). Quantitative analysis of NMR spectra with chemometrics. Journal of Magn. Reson., 190, 26-32.

Yan-de, L. & Yi-bin, Y. (2004). Measurement of sugar content in Fuji apples by FT-NIR spectroscopy. Zhejiang University SCIENCE, *5(6),651-65.*

Zhai, Y., Thomasson, J. A., BoggessIII, J. E. & Sui, R. (2006). Soil texture classification with artificial neural networks operating on remote sensing data. Computers and Electronics in Agriculture, 54, 53–68.

Xie, L., Ye, X., Liu, D. & Ying, Y. (2009). Quantification of glucose, fructose and sucrose in bayberry juice by NIR and PLS. Food Chemistry, 114, 1135–1140.