Int. Agrophysics, 2010, 24, 351-356

INTERNATIONAL
Agrophysics
www.international-agrophysics.org

Feasibility of near infrared spectroscopy for analysis of date fruits

S.A. Mireei[1]*, *S.S. Mohtasebi*[1], *R. Massudi*[2], *S. Rafiee*[1], *and A.S. Arabanian*[2]

[1]Department of Agricultural Machinery Engineering, Faculty of Agricultural Engineering and Technology, University of Tehran, P.O. Box 4111, Karaj 31587-77871, Iran
[2]Laser and Plasma Research Institute, Shahid Beheshti University, G.C., Evin, Tehran 1983963113, Iran

Received February 28, 2010; accepted June 18, 2010

A b s t r a c t. This paper deals with a research to study a near infrared spectroscopy as a nondestructive method for discrimination of date fruits according to four main ripening stages. For this purpose, a near infrared spectra of dates were acquired in the range of 900-1700 nm. Principal component analysis was performed to reduce the dimensionality of the spectral data and then the first five principal components were used as the inputs to a back propagation artificial neural network (BP-ANN) to classify the dates based on the ripening stages. The principal component analysis – artificial neural network (PCA-ANN) model provided satisfactory discrimination results with an accuracy of 95.5% for test sample set. Moreover, partial least square method with different data preprocessing methods was applied to predict the moisture content and soluble solids content of date fruits. The best predictive models showed the coefficient of determination values of 0.98 and 0. 0.96 with residual predictive deviation values of 6.22 and 5.03 for the moisture and the soluble solids content, respectively. A near infrared spectroscopy appeared to be a good method for both classification of Shahani date fruits according to ripening stages and also determination of their maturity indices.

K e y w o r d s: NIR spectroscopy, date fruit, artificial neural network, discrimination, moisture content

INTRODUCTION

Date fruit (*Pheonix dactylifera* L.) is one of the most important commercial fruit in the Middle Eastern countries. Date fruit is a good source of energy, vitamins, and a group of mineral substances like phosphorus, iron, potassium, and also a significant amount of magnesium and calcium (Biglari *et al*., 2009; Sahari e*t al*., 2007). With production of 10^9 kg of date fruit and exportation of around 143×10^6 kg in 2006, Iran has had an important role in production and exportation of different varieties of date throughout the world (FAOSTAT, 2006).

As well known, there are four distinct ripening stages almost for all varieties of date fruits which are termed as Kimri, Khalal , Rutab and Tamr, respectively (Sahari *et al*., 2007; Imad and Abdul Wahab, 1995; Al-Shahib and Marshall, 2003). However, depending on cultivar type and market demand, dates are presented in market at three final ripening stages *ie* Khalal, Rutab and Tamr (Sahari *et al*., 2007).

Shahani date cultivar, considered in the current study, is one of the most famous and delicious date which is produced commonly in the south region of Fars province (Jahrom) in Iran (Keramat Jahromi *et al*., 2008). Due to single-pass and non-selective harvesting method of date fruits, the harvesting has to include the date fruits in all ripening stages, so design and development of a rapid nondestructive method which can help in recognizing any ripening stage should be studied.

Date fruit is a climacteric fruit which is ripen even after harvesting. The main factors which will affect the future ripening of dates are moisture and sugar content (Schmilovitch *et al*., 1999). Besides of being an effective parameter for maturity determination, moisture content (MC) is also an important factor in food processing industry of date fruits. Most instrumental techniques which are used to measure MC and soluble solids content (SSC) are destructive and involve a considerable amount of manual and time-consuming work.

A near infrared (NIR) spectroscopy has been documented by a great number of researchers to be a non-destructive, rapid and accurate method for both classification and regression purposes of different fruits (Andrejko *et al.*, 2008; Baranowski and Mazurek, 2009; Kim *et al*., 2000; Nicolaï *et al.*, 2007). Classification analyses were carried out using several pattern recognition approaches such as linear discriminate analysis (LDA), partial least square

*Corresponding author's e-mail: samireei@ut.ac.ir

discriminate analysis (PLS-DA), soft independent modeling of class analogy (SIMCA) and combination of principal component analysis (PCA) with artificial neural networks (ANN) (Berrueta *et al*., 2007). On the other hand, principal component regression (PCR) and partial least squares (PLS) are the common methods in regression analysis which were used in several studies (Nicolaï *et al*., 2007).

This research was aimed to evaluate the potential of NIR spectroscopy in discriminating of date fruits according to four maturity stages, and establish calibration models to predict MC and SSC of Shahani date fruits based on NIR spectral information.

MATERIALS AND METHODS

Some bunches of Shahani date fruits were harvested in August 20, 2008 and September 20, 2008 from two different orchards in Jahrom (Fars province, Iran). After each harvest, fruits were stored at 2°C and for maximum 10 days. Totally 200 samples of Shahani dates were selected from all bunches which were generally characterized by the following ripening stages: Kimri (17%), Khalal (41%), Rutab (27%) and Tamr (15%). Before spectra acquisition, fruits were equilibrated for about 6 h to room temperature (about 22°C).

NIR spectra were acquired in interactance mode using an EPP2000 NIR non-scanning spectrometer (StellarNet, Inc. Oldsmar, Florida, USA). The indium gallium arsenide (InGaAs) detector of spectrometer could acquire the spectra in the range of 900-1700 nm with 2.5 nm resolution. Using a bifurcated cable (R400-7-VISNIR, StellarNet, Inc.) in the interactance mode, light due to specular reflection could not directly enter the detector and only diffuse reflection of the date fruits was acquired (Nicolaï *et al*., 2007). A schematic of NIR measuring system is shown in Fig. 1. Each spectrum was the average of 10 scans and for each sample three spectra were acquired at about 120° intervals around the date equatorial axis. These three spectra were then averaged and the mean spectrum was considered in the calculations.

Immediately after spectral acquisitions, ripening stage of each sample was determined. Then, each date fruit was cut from the length in two halves and the inside seed removed. One half was used to measure MC and the other half was exploited for determining the SSC. MC value was determined by use of a method described by Elleuch *et al*., (2008) and Keramat Jahromi *et al*. (2008). SSC was determined by using a digital refractometer (Abbe refractometer, Sun Instruments Corp., Torrance, CA, USA) as described in AOAC (1998).

For classification analysis, multiplicative scatter correction (MSC) was used as a data pretreatment. MSC is a linear transformation which is able to remove the effects of light scattering on spectra (Lu, 2001). In this study, all statistical and chemometric analyses were done in Unscrambler software V9.7 (CAMO ASA, Oslo, Norway).

After data pretreatment, principal components analysis (PCA) was performed to reduce the dimensionality of the spectral data. Besides of reserving the major spectral information, PCA is an useful data mining technique which can define the fewer, new and uncorrelated variables or principal components (PCs) instead of original variables. The effective PCs, which could explain the most common variations in all data, were then selected and used as the inputs in an ANN model to classify the date fruits in different stages of ripening (He *et al*., 2006; He *et al*., 2007).

ANN is a powerful computational modeling tools, consisting of a large amount of interconnected simple processing elements (artificial neurons) that is able to model complex, parallelism and non-linear systems. In particular, back propagation (BP) networks are the most commonly and appropriate networks for spectroscopic applications. A BP network is a feed-forward multilayer perceptron network consisting one input layer with the neurons as independent variables (PCs), one or more hide layers and one output layer with the neurons as dependent variable, namely ripening stages (Omid *et al*., 2009). In the feed-forward networks,

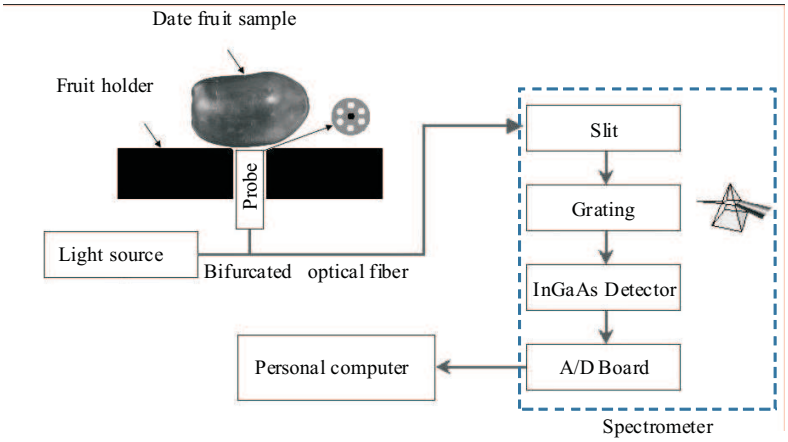

Fig. 1. Schematic of NIR measuring system.

error minimization can be achieved by a number of methods such as gradient descent (GD), Levenberg-Marquardt (LM) and conjugate gradient (CG). The standard BP uses the GD technique which is very stable when a small learning rate is used, however it has slow convergence properties (Omid *et al.*, 2010). Several methods for speeding up BP algorithm have been used such as GD with momentum (GDM) and a variable learning rate. In this paper, GDM learning rule which is an improvement of the straight GD rule is applied to avoid local minima, speed up learning and stabilizing convergence.

For classification of dates according to their ripening stages, the NIR spectra of 200 samples were divided into the training (50%), validation (25%) and test (25%) subsets. The neural networks classification was performed by using STATISTICA neural networks software (StatSoft, Inc., USA).

In order to establish the regression models for predicting MC and SSC, the relative reflectance spectra (R) were expressed as absorbance using the equation of log (1/R). Different data preprocessing: multiplicative scatter correction (MSC), first derivative and the second derivative were then performed on spectra and their effects on the results compared to raw spectra. The gap size in the first and the second derivatives was set to 9 after trials of several different gap sizes.

In this study, the PLS method was used to develop calibration models for predicting MC and SSC. Randomly 150 samples were selected for calibration and leave-one-out cross validation (75%) as well as 50 samples for test set or external validation (25%) were considered. The optimum number of latent variables in PLS models was determined by finding the minimum plot of the root mean square error of cross validation (RMSECV) against the number of latent variables. The performance of the PLS models were evaluated with regard to such parameters as determination coefficient (R^2), root mean square error of leave-one-out cross validation and prediction (RMSEP) (test set validation). Finally, residual predictive deviation values (RPD) which is the best index to evaluate the regression models were computed by dividing the standard deviation (SD) of the reference values by the standard error of cross validation or prediction (Williams and Norris, 2001).

RESULTS AND DISCUSSION

The average reflectance spectra from 900 to 1700 nm are shown in Fig. 2 for four randomly selected samples representing Kimri, Khalal, Rutab and Tamr stages of ripening. The dates having advanced ripening stages, have a lower total percentage of light reflectance in the spectra due to the flesh softening. The flesh softening leads to a lower opacity, increasing light penetration depth and decreasing the level of diffuse reflectance (Cavaco *et al*., 2009). However, this scattering effect may be obscured by absorption effects which can be removed by preprocessing methods. In all spectra, there are three absorption bands around 980, 1 180, and 1 430 nm. The absorption around 980 nm is closed to absorption bands of -CH and -OH functional groups which can be due to absorption by water and carbohydrates in date (Park *et al*., 2003). Another absorption around 1180 nm is related to -CH 2nd overtone and finally absorption around 1430 nm can be related to absorption of -CH and -OH functional groups (Miller, 2001).

After preprocessing of spectra with MSC, PCA was performed on 200 sample spectra. The first principal component (PC1) could explain 88.7% of the original variables while with five PCs, the explained variance is 99.2%. Generally, when the PCs have 85% explained variance of the original spectra, they can be used as the input to ANN instead of original data (He *et al*., 2006). The number of five PCs was selected as the input for ANN by trial and error procedure. ANN models with lower number of PCs had a lower discrimination power while by using higher number of PCs, models had more complexity without meaningful discrimination power.

With these PCs as the inputs, a BP-ANN model with three layers was developed. One input layer consisting five neurons for five PCs, one hidden layer including five neurons and an output layer with four neurons for each ripening stage. The optimum number of 5 neurons in the hidden layer was selected by using an empirical equation as $b=\sqrt{m+n}+a$, where m and n are the number of neurons in the input and output layers, respectively, and a is a constant value between 1 and 20 (Li and He, 2008). According to this equation, the number of neurons in the hidden layer was within 4-23. Several networks with different number of neurons in hidden layer were established and finally the number of five neurons was chosen. The transfer function for each neuron was considered as differentiable sigmoid function. The output vector of four output neurons was assumed to be a four-byte binary code. So the output vectors (1000), (0100), (0010) and (0001) were denoted as the Kimri, Khalal, Rutab, and Tamr stages of ripening,

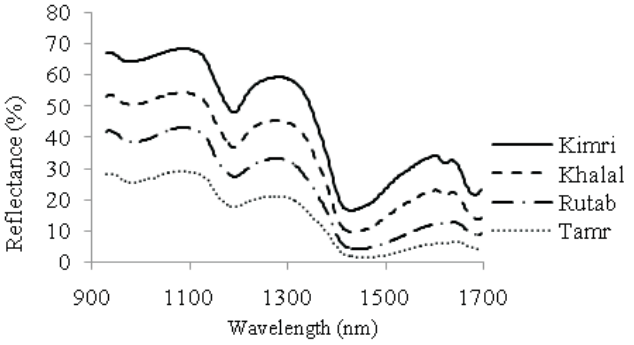

Fig. 2. Relative reflectance spectra obtained for Shahani date fruits at Kimri, Khalal, Rutab and Tamr stages of ripening.

T a b l e 1. Results of classification in training, verification and test sets of PCA-ANN model

Subsets	Ripening stages	Discrimination results					Accuracy (%)
		No.	Kimri	Khalal	Rutab	Tamr	
Training	Kimri	17	17	0	0	0	100
	Khalal	41	0	39	2	0	95
	Rutab	27	0	2	25	0	93
	Tamr	15	0	0	0	15	100
Verification	Kimri	8	8	0	0	0	100
	Khalal	21	0	20	1	0	95
	Rutab	14	0	1	13	0	93
	Tamr	7	0	0	0	7	100
Test	Kimri	9	9	0	0	0	100
	Khalal	20	0	18	2	0	90
	Rutab	13	0	1	12	0	92
	Tamr	8	0	0	0	8	100

respectively. The learning rate and momentum in the network were as 0.1 and 0.3, respectively and the number of epochs was equal to 1000 times. After the training, the root mean square errors of training, verification and test were found to be 0.15, 0.16 and 0.20, respectively.

Table 1 shows the results of classification in training, verification and test sample sets in each ripening stage. The discrimination accuracy for both training and verification sample sets was 97%. Whilst the model was used for prediction, an accuracy of 95.5% was obtained.

The MC and SSC values of Shahani date fruits obtained by the standard destructive methods and used for calibration (leave-one-out cross validation) and external validation data sets are shown in Table 2. The standard deviations of the MC and SSC were 12.16 (CV=35.0%) and 9.8°Brix (CV= 19.10%), respectively, for all date fruit samples. The small differences between standard deviations in MC and SSC at each calibration and test set samples with those in total samples indicate a good division of samples in each subset.

T a b l e 2. MC and SSC of date fruits obtained by the destructive methods for calibration and validation sample sets

Parameter	Calibration or leave one out CV		Test set validation	
	MC (%)	SSC (°Brix)	MC (%)	SSC (°Brix)
Maximum	85.90	45.6	85.71	44.4
Minimum	41.17	10.0	45.37	10.4
Mean value	63.98	27.6	62.67	28.9
SD	12.37	9.8	11.69	9.7
CV(%)	19.33	35.6	18.66	33.5

Table 3 shows the results of PLS models for determining MC using different preprocessing methods. Different preprocessing methods resulted in different prediction performances. Among different preprocessing methods, MSC resulted in the best prediction model with R^2 of leave one out cross validation and test set validation equal to 0.98 and 0.97, respectively, RMSECV of 1.82 and RMSEP of 1.88. In this model, RMSECV decreases as the number of latent variable increases until it exceeds 8 latent variables. After the 8th latent variable, the RMSECV increased slightly with increasing the number of latent variables caused by overfitting of the model. The obtained RPD values for this model were 6.80 and 6.22 for leave-one-out cross validation and test set validation, respectively. According to Williams and Norris (2001), RPD values within 6.5-8.0 indicate a very good performance of the model, while the values varying from 5.0 to 6.4 show a good performance of model which can be used for quality control purposes. After MSC, models with 1st derivative, 2nd derivative and without any preprocessing had a better results in calibration, leave-one-out cross validation and test set validation, respectively. MC values predicted by the best PLS model and the values observed by the reference method for the leave-one-out cross validation and test set validation are shown in Fig. 3a.

The results of the PLS models for determining SSC using different preprocessing methods is shown in Table 3. The best PLS model with MSC as preprocessing, yielded the R^2 of leave one out cross validation and test set validation equal to 0.97 and 0.96, respectively, RMSECV of 1.82°Brix

T a b l e 3. The results of PLS models for predicting MC and SSC using different preprocessing methods (spectrum range: 900-1700 nm)

Preprocessing	Calibration			Leave one out cross validation			Test set validation		
	LV*	R^2	RMSEC	R^2	RMSECV	RPDCV	R^2	RMSEP	RPD
					MC				
MSC	8	0.98	1.70	0.98	1.82	6.80	0.97	1.88	6.22
Log (1/R)	9	0.96	2.48	0.95	2.71	4.56	0.94	2.81	4.16
D^1log (1/R)	9	0.97	2.12	0.96	2.34	5.83	0.96	2.39	4.89
D^2log (1/R)	9	0.97	2.29	0.95	2.66	4.65	0.95	2.77	4.22
					SSC				
MSC	8	0.97	1.69	0.97	1.82	5.40	0.96	1.93	5.03
Log (1/R)	8	0.95	2.31	0.94	2.44	4.03	0.92	2.82	3.44
D^1log (1/R)	8	0.96	2.02	0.94	2.33	4.22	0.94	2.36	4.11
D^2log (1/R)	10	0.97	1.78	0.95	2.19	4.49	0.94	2.44	3.98

*LV – latent variable.

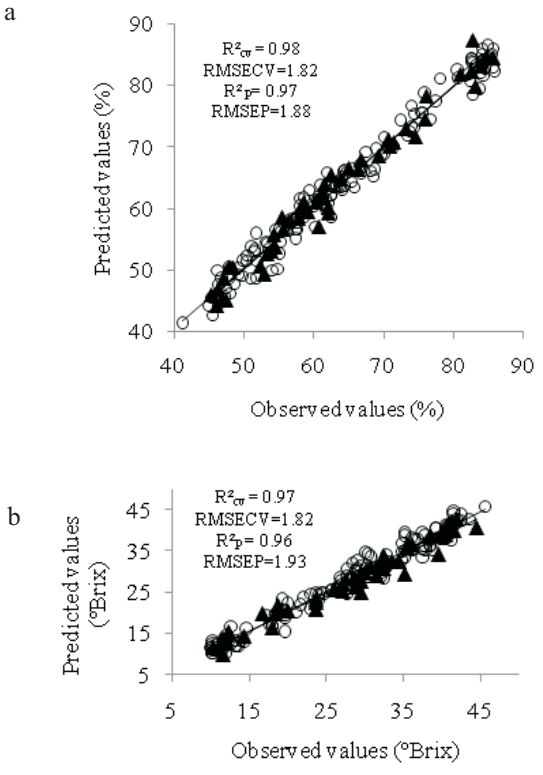

Fig. 3. Predicted and observed values of MC and SSC resulted from leave one out cross validation (○) and test set validation (▲) using PLS model with 8 latent variables and the MSC as data preprocessing.

and RMSEP of 1.93°Brix. The RMSECV plot for this model displayed a minimum at the 8th latent variable, so the optimum number of latent variables was selected as 8. The resulted RPD values for this model were found to be 5.40 and 5.03 for leave-one-out cross validation and test set validation, respectively, which according to Williams and Norris (2001), these RPD values show a good performance of model. Similar to MC, after MSC, models with 1st derivative, 2nd derivative and without any preprocessing resulted to a better performance, in calibration, leave-one-out cross validation and test set validation, respectively. The values predicted by the best PLS model versus the observed values of SSC obtained by the leave-one-out cross validation and the test set validation are depicted in Fig. 3b.

CONCLUSIONS

1. The obtained results indicate that NIR spectroscopy has a high potential to recognize the ripening stages of Shahani date fruits and to determine simultaneously some maturity indices of dates.

2. Partial components analysis combined with BP-ANN was assessed as a powerful pattern recognition technique to establish the relation between the interactance spectra and four main ripening stages of dates (Kimri, Khalal, Rutab and Tamr). The discrimination accuracy of this model was found as 97% for both training and verification sample sets and 95.5% for test sample set.

3. The best PLS models can predict the moisture content and the soluble solids content of date fruits with root mean square error of prediction 1.88% (R^2= 0.98) and 1.93°Brix (R^2=0.96), respectively. The relative high residual predictive deviation values of these models suggest that NIR spectroscopy might be an application for date fruits quality control in the industry.

REFERENCES

Al-Shahib W. and Marshall R.J., 2003. The fruit of the date palm: its possible use as the best food for the future? Int. J. Food Sci. Nutr., 54(4), 247-259.

Andrejko D., Rydzak L., Ślaska-Grzywna B., Goździewska M., and Kobus Z., 2008. Influence of preliminary thermal processing applying infra-red radiation on pea seeds cooking process. Int. Agrophysics, 22, 17-20.

AOAC, **1998.** Official Methods of Analysis. AOAC Press, Washington, DC, USA.

Baranowski P. and Mazurek W., 2009. Detection of physiological disorders and mechanical defects in apples using thermography. Int. Agrophysics, 9-17.

Berrueta L.A., Alonso-Salces R.M., and Héberger K., 2007. Supervised pattern recognition in food analysis. J. Chromatography, 1158, 196-214.

Biglari F., AlKarkhi A.F.M., and Easa A.M., 2009. Cluster analysis of antioxidant compounds in dates (*Phoenix dactylifera*): Effect of long-term cold storage. J. Food Chem., 112, 998-1001.

Cavaco A.M., Pinto P., Antunes M.D., da Silva J.M., and Guerra R., 2009. 'Rocha' pear firmness predicted by a Vis/NIR segmented model. J. Postharvest Biol. Technol., 51, 311-319.

Elleuch M., Besbes S., Roiseux O., Blecker C., Deroanne C., Drira N., and Attia H., 2008. Date flesh: Chemical composition and characteristics of the dietary fibre. J. Food Chemistry, 111, 676-682.

FAOSTAT, **2006.** Statistical Year Book of FAO. Available: http:// faostat.fao.org.

He Y., Feng S., Deng X., and Li X., 2006. Study on lossless discrimination of varieties of yogurt using the Visible/ NIR-spectroscopy. Food Res. Int., 39, 645-650.

He Y., Li X., and Deng X., 2007. Discrimination of varieties of tea using near infrared spectroscopy by principal component analysis and BP model. J. Food Eng., 79, 1238-1242.

Imad A.A. and Abdul Wahab K.A., 1995. Chemical composition of date varieties as influenced by the stage of ripening. J. Food Chemistry, 54, 305-309.

Keramat Jahromi M., Jafari A., Rafiee Sh., Mirasheh R., and Mohtasebi S.S., 2008. Changes in physical properties of date fruit (cv. Shahani) during three edible stages of ripening. Am.-Eur. J. Agric. Environ. Sci., 3 (1), 132-136.

Kim J., Mowat A., Poole P., and Kasabov N., 2000. Linear and non-linear pattern recognition models for classification of fruit from visible-near infrared spectra. Chemometrics and Intelligent Laboratory Systems, 51, 201-216.

Li X. and He Y., 2008. Discriminating varieties of tea plant based on Vis/NIR spectral characteristics and using artificial neural networks. J. Biosys. Eng., 99, 313-321.

Lu R., 2001. Predicting firmness and sugar content of sweet cherries using near-infrared diffuse reflectance spectroscopy. Transactions of the ASAE, 44(5), 1265-1271.

Miller C.E., 2001. Near-Infrared Technology in the Agricultural and Food Industry. AACC Press, St. Paul, MN, USA.

Nicolaï B.M., Beullens K., Bobelyn E., Peirs A., Saeys W., Theron K.I., and Lammertyn J., 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. J. Postharvest Biol. Technol., 46, 99-118.

Omid M., Mahmoudi A., and Omid M.H., 2009. An intelligent system for sorting pistachio nut varieties. Expert Systems with Applications, 36, 11528-11535.

Omid M., Mahmoudi A., and Omid M.H., 2010. Development of pistachio sorting system using principal component analysis (PCA) assisted artificial neural network (ANN) of impact acoustics. Expert Systems with Applications, 37, 7205-7212.

Park B., Abbott J.A., Lee K.J., Choi C.H., and Choi K.H., 2003. Near-infrared diffused reflectance for quantitative and qualitative measurement of soluble solids and firmness of Delicious and Gala apples. Transactions of the ASAE, 46(6), 1721-1731.

Sahari M.A., Barzegar M., and Radfar R., 2007. Effect of varieties on the composition of dates (*Phoenix dactylifera* L.). J. Food Sci. Technol. Int., 13(4), 269-275.

Schmilovitch Z., Hoffman A., Egozi H., Ben-Zvi R., Bernstein Z., and Alchanatis V., 1999. Maturity determination of fresh dates by near infrared spectrometry. J. Sci. Food Agric., 79, 86-90.

Williams P.C. and Norris K., 2001. Near-Infrared technology in the Agricultural and Food industry. AACC Press, St. Paul, MN, USA.