Advance Journal of Food Science and Technology 11(4): 344-349, 2016
DOI:10.19026/ajfst.11.2421
ISSN: 2042-4868; e-ISSN: 2042-4876
© 2016 Maxwell Scientific Publication Corp.
Submitted: August 21, 2015 Accepted: September 11, 2015 Published: June 05, 2016

Research Article

Sugar Content Detection of Red Globe Grape Based on QGA-PLSR Method and Near-infrared Spectroscopy

Qiaohua Wang, Yihua Tang and Yufei Duan
College of Engineering, Huazhong Agricultural University, Wuhan, 430070, China

Abstract: Nowadays, the sugar content detection of red globe grape is destructive, inefficient and cumbersome. In this study, in order to find a rapid non-destructive detection method for the sugar content of red globe grape, the experiment was conducted to study the relationship between sugar content of red globe grape and the near-infrared spectra. The near-infrared spectra of 160 red globe grapes were acquired with a wavelength of from 4000 to 10000 cm^{-1}. The model established in all band was analyzed by using different spectral pretreatments combined with three quantitative analysis models which were Multiple Linear Regression (MLR), Partial Least Squares Regression (PLSR) and Principal Component Regression (PCR). The results illustrated that the reliability of PLSR was the best and PCR followed by. In order to get a better prediction model, the number of wavebands was reduced from 1557 to 650 by using quantum genetic algorithm and partial least squares regression (QGA-PLSR). At the same time, the correlation coefficient (RC) of prediction model and its Root-Mean-Square Error of Prediction (RMSEP) were improved obviously. RC was increased from 0.975 to 0.995 and RMSEP was decreased from 0.8 to 0.495. With the QGA-PLSR method, the number of wavebands was reduced greatly which made full use of the wavebands information. And the sugar content prediction model of red globe grape was established. This provided technical support for the quality classification of red globe grape.

Keywords: Near-infrared spectroscopy, partial least squares regression, quantum genetic algorithm, red globe grape, sugar content

INTRODUCTION

The red globe grape is a good variety of grape, which is full of high nutrition and tastes delicious (Min, 2013). It contains many vitamins, amino acids, essential trace elements. Therefore, it is widely spread and planted. At present, red globe grape is mainly used as delicious fruit. With the increase of consumption group and the ascension of consumption concept, consumers are seeking for food with high quality. In China, the standards for red globe grape contain size, sugar content, color and coloring rate and so on. The sugar content has a direct impact on the quality of red globe grape. Nowadays, the detecting method of sugar content for red globe grape is mainly manual measuring. The method is cutting out a few grapes from each string of red globe grape, squeezing juice, measuring the juice with Brix meter and calculating the average of sugar content as the sugar content of the string. The process of this method is tedious and inefficient.

Near-infrared spectrum is a rapid, convenient and non-destructive analysis method, which analyzes the sample quantitatively or qualitatively by the acquisition of multi-band spectral information. As the near-infrared has low cost, fast detection, nondestructive advantages, it is wildly used in food, chemical, pharmaceutical and other fields (Hua *et al*., 2010; Zhu *et al*., 2011; Rong *et al*., 2010; Jun *et al*., 2012; Fang *et al*., 2008). Domestic scholars did a lot of researches to detect fruit sugar content with near-infrared spectrum. Fei *et al*. (2014) found that using the spectral information in the umbilical part of watermelon to establish prediction model was slightly better than that of the stem part. Dai Fen detected the total sugar content of orange with near-infrared spectrum and found that extracting 14 optimum wavelengths to establish model by SPA spectral compression was better than that of full spectrum (Fei *et al*., 2011). Wu *et al*. (2008) measured the fat content in milk with LS-SVM and the results showed that the fat content in milk powder could be detected successfully. The determination coefficient was 0.9796 and prediction error was 0.8367. Foreign scholars also had done a lot of researches on fruits (Slaughter *et al*., 2003; Wedding *et al*., 2011; Cayuela, 2008).

According to the above, detecting the sugar content of soluble solid is feasible by near-infrared spectrum. However, most of the research is aimed at large fruit, such as apple, citrus, watermelon. This study wants to find an effect model to detect the sugar content of red globe grape.

Corresponding Author: Qiaohua Wang, College of Engineering, Huazhong Agricultural University, Wuhan, 430070, China
This work is licensed under a Creative Commons Attribution 4.0 International License (URL: http://creativecommons.org/licenses/by/4.0/).

MATERIALS AND METHODS

The red globe grape samples: The object of this research was the red globe grape, which was bought from fruit market. This experimental samples were 160 red globe grapes, which were randomly selected from 40 bunches. Among them, 120 grapes were used as train samples and the other 40 grapes were used as prediction samples.

Main equipments and instruments: The instrument to collect spectrum is FT-NIR ANALYZER, whose effective detecting wavelength range is 1000~2500 nm. And the corresponding wave number range is 10000~4000 cm^{-1}. The instrument to measure the sugar content of red globe grape is Brix meter, type VBS2T/ATC and its accuracy is 0.2% Brix.

Experimental procedure:

- **Preheating:** Turn on the FT-NIR ANALYZER and keep it on for 1 h
- Open the RESULT Integration software and set the collection number to 32 times and the wave number resolution to 4 cm^{-1}
- Scan the background to avoid the background influence
- **Scanning the sample:** Place the sample to erect. Set the maximum longitudinal diameter as the scanning direction. The original acquisition spectrum is shown in Fig. 1
- Break the red globe grape and measure the sugar content of 3 different parts on one grape. Take the average value as the actual sugar content
- The actual sugar contents of training samples and prediction samples are showed in Table 1. It shows that the sugar contents of training samples is close to the prediction samples, which satisfies the modeling requirement.

Data pretreatment and modeling method: In the process of spectra collecting, there exist the factors, such as noise, un-uniform size of red globe grape and so on. Therefore, pretreatment should be done before processing spectra. In this research, the employed pretreatment method mainly are Savitzky-Golay filter, Norris derivative filter, first-order differential and second-order differential.

After preprocessing, modeling analysis is done on the spectrum data. The model of Multiple Linear Regression (MLR), Principal Component Regression (PCR) and partial Least Squares Regression (PLSR) are established respectively combined with preprocessing data. The optimum prediction treatment and modeling method are picked out by comparing predictive effect.

MLR refers to a regression analysis when there is a linear relation between the multiple independent variables and dependent variable. After utilizing independent variables to analyze principal component and reduce dimensions, PCR is using the obtained principal component to regress the dependent variables. It can eliminate multi-collinearity. And this is a regularly used linear regression method. PLSR is deriving a few variables from the independent variable system. Therefore, they have a strong ability to explain independent variable system as they generalize the

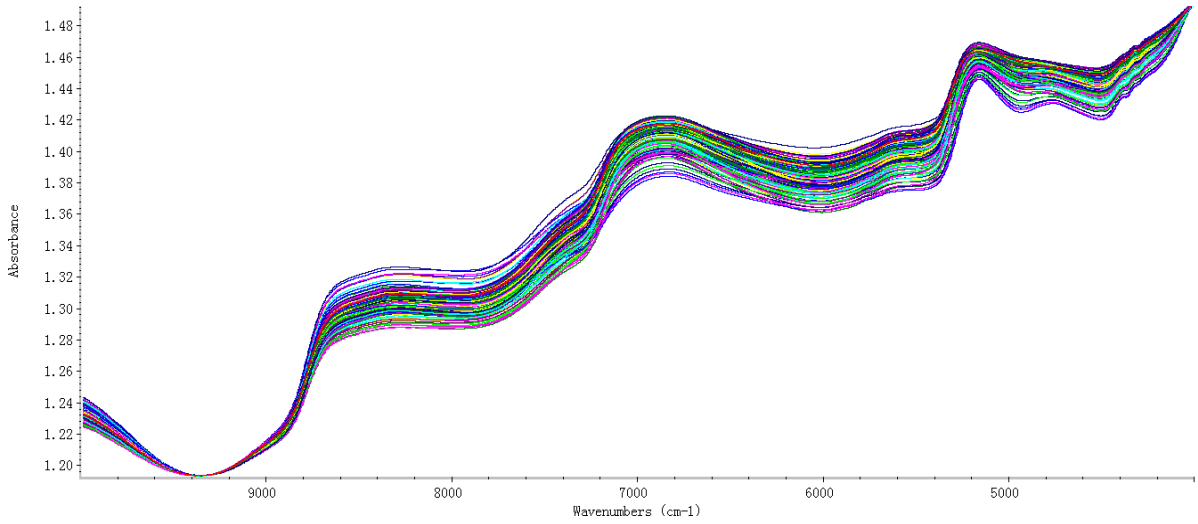

Fig. 1: Original spectra

Table 1: Sugar contents of red globe grape samples

Samples	Sample number	Minimum of sugar content	Maximum of sugar content	Average of sugar content	S.D.
Training samples	120	11.7	19.53	16.22	1.618
Prediction samples	40	12.83	18.83	15.55	1.678

S.D.: Standard Deviation

Table 2: Prediction results of different pretreatments and modeling methods in all bands

Pretreatment methods	MLR			PCR			PLSR		
	RC	RMSEC	RMSEP	RC	RMSEC	RMSEP	RC	RMSEC	RMSEP
SG	0.672	1.22	1.4	0.822	0.931	1.09	0.814	0.949	1.07
FD	0.803	0.974	1.48	0.834	0.904	0.996	0.975	0.364	0.8
FD+ND	0.865	0.828	1.04	0.837	0.896	0.954	0.947	0.524	0.827
SD+ND	0.826	0.911	1.01	0.853	0.853	0.972	0.969	0.402	0.912

SG: Short for Savitzky-Golay filter, ND: Short for Norris derivative filter, FD: Short for first derivative, SD: Short for second derivative

original variables. Then, these variables are used to regress. Generally, the above 3 methods are used to establish models and the models' RC of validation set and RMSEP of prediction set are looked as evaluation criteria. The bigger the RC is and the smaller the RMSEP is, the better the prediction effect is.

QGA is the product of quantum computation and genetic algorithm. It is a newly developed probability evolutionary algorithm. Genetic algorithm is simulating the relationship between evolutionary biology evolution rules and chromosome. Then, the optimal individual is searched by selecting, crossing, variation and so on. The objective function of genetic algorithm is the global adapting search by the guidance of probability. It has high robustness and universality. If the manner of selecting, crossing and variation is wrong, genetic algorithm may converge slowly and fall into local extremum. However, QGA leads the state vector expression of quantum into genetic code. It uses quantum logic gate to realize the chromosome evolution. It has a better effect than general genetic algorithm.

RESULTS AND DISCUSSION

Spectrum pretreatment in all band and modeling analysis: Model is established on the all band spectra data and the analyzing result is showed in Table 2. In Table 2, RC, RMSEC and RMSEP represents related coefficient, root mean standard error of calibration and root mean square error of prediction respectively. From Table 2, it shows that the prediction effect of the combination of first-order differential and PLSR is the best. The RC is 0.975 and the RMSEP is 0.8. Among the 3 modeling methods, the overall prediction effect of PLSR is preferable. The all band prediction effect is not ideal. The possible reason is that there exists interfering information in all band. To improve the prediction ability of model, QGA is used to do further treatment on the spectra data.

Modeling method of QGA combined with PLSR: To obtain better prediction model, this research uses QGA to choose effective wavebands. Then, the model is established by PLSR method. QGA is a product of the quantum computing combined with genetic algorithm, which is a newly developed probability evolutionary algorithm (Xiang *et al*., 2004; Liang and Wang, 2011; Zhang *et al*., 2014).

The method of QGA-PLSR is made up of 4 sections:

- **Initialization:** Segment treatment is done on the spectra data. There are 1557 wavebands in the sample spectra. All of the wavebands is divided into 31 sections. The former 30 sections contains 50 wavebands and the 31_{st} wavebands contains 57 wavebands. Every section is looked as a gene. Then, the gene number of chromosome is m = 31. Every chromosome is changed to quantum state. The calculation formula is as follows:

$$|x_i\rangle = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \cos(\frac{2\pi(X_i - a_i)}{b_i - a_i}) \\ \sin(\frac{2\pi(X_i - a_i)}{b_i - a_i}) \end{bmatrix}$$

  Among them, a_i and b_i are the upper and lower bounds of x_i. The scale of determined population is 40. The population q_i is initialized and the probability amplitude of all the chromosomes are set as 1/(20.5p(t0)). Measure each of the population $Q(t_0)$ and the corresponding determined solution can be got.
- Establishing fitness function model and calculating the fitness value of every chromosome. Fitness value represents the individual is good or bad in the model. Individual with bigger fitness is retained so that it can be duplicated to the next generation with more ease. The model used in this research is PLSR model. Leave-one-out method is employed to do cross validation. The evaluation criteria is the RC of validation set and RMSEP of prediction set. When RC is bigger and RMSEP is smaller, the prediction model is better. Therefore, the fitness function is:

$$\text{fitness} = \frac{RC}{1 + \sqrt{2 * \text{RMSEP}}}$$

  p(t0) gained in the step1is calculated and the fitness value of every individual can be got. Then, judge whether the calculating process can be

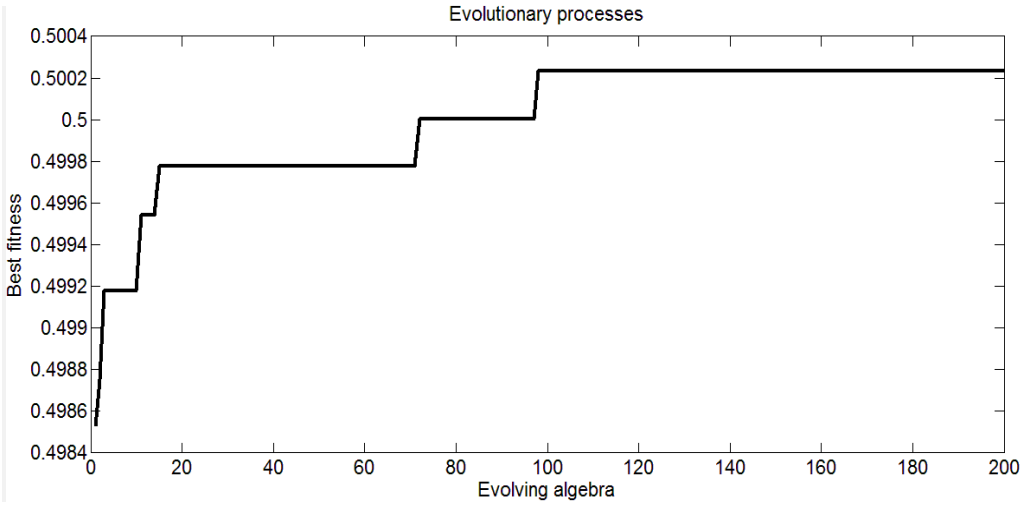

Fig. 2: Results of QGA-PLSR iterative

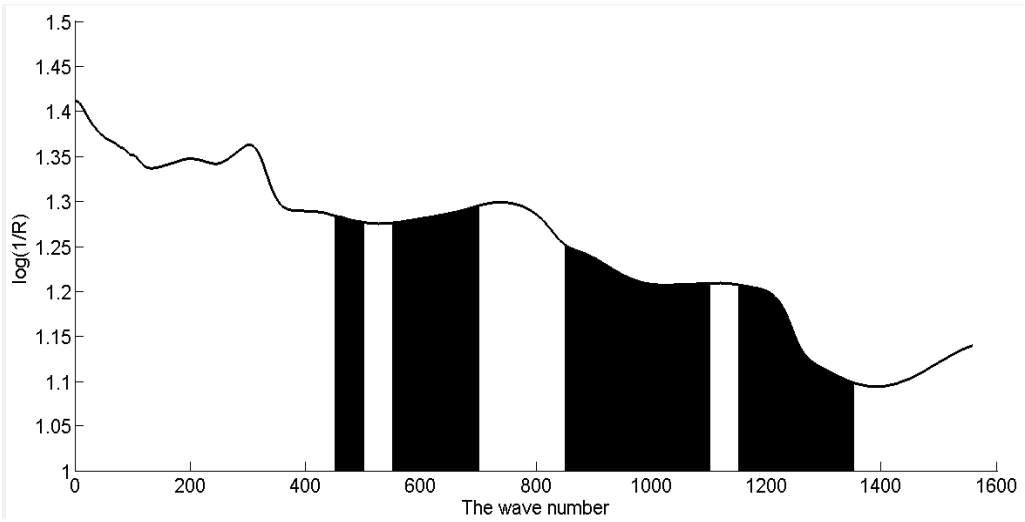

Fig. 3: Effective bands chosen by GQA method

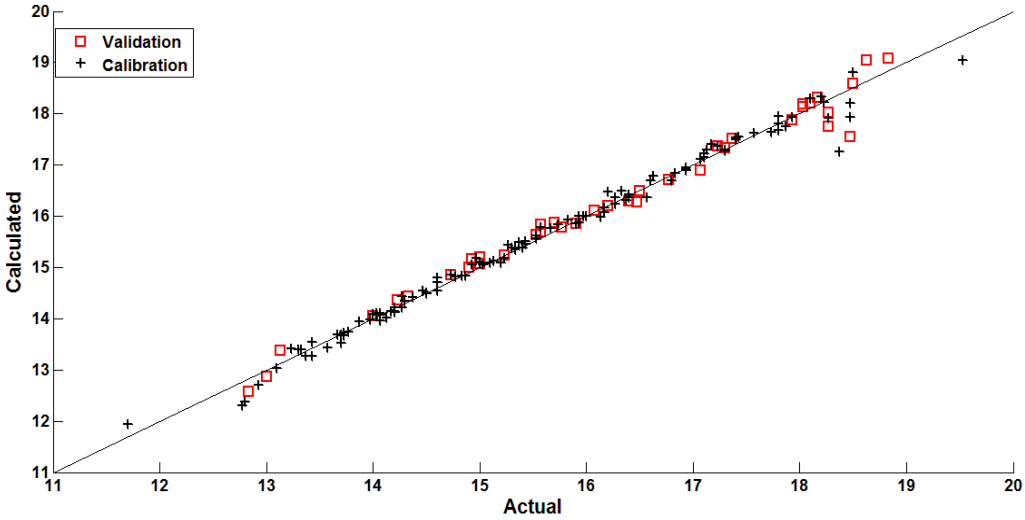

Fig. 4: Prediction results by QGA+PLSR modeling method

Table 3: Prediction results by QGA-PLSR and QGA+PCR modeling method

	QGA+PLSR			QGA+PCR		
Pretreatment methods	RC	RMSEC	RMSEP	RC	RMSEC	RMSEP
SG	0.909	0.681	0.643	0.751	1.08	1.18
FD	0.995	0.177	0.489	0.774	1.04	1.09
FD+ND	0.985	0.28	0.395	0.755	1.07	1.12
SD+ND	0.983	0.296	0.515	0.733	1.11	1.3

finished or not. If the conditions are met, quit calculating. If not, continue calculating.

- Measure each of the population Q(t) and the corresponding determined solution can be got. Then, the fitness value of the solution is evaluated. A new population Q(t+1) can be gained after adjusting individual by quantum rotating gate U(t). The quantum rotating gate is as follows:

$$\mathrm{U}(\theta_i)=\begin{bmatrix}\cos(\theta_i) & -\sin(\theta_i)\\ \sin(\theta_i) & \cos(\theta_i)\end{bmatrix}$$

Among them, θ_i is rotation angle. The rotation angle is selected according to the following formula:

$$\theta_i=-\mathrm{sgn}(B_i)\theta_0\exp\left(\frac{\left|Z_i^j-\min(Z_i^j)\right|}{\max(Z_i^j)-\min(Z_i^j)}\right)$$

Among them:

$$\mathrm{B}_i=\left\|\begin{vmatrix}\alpha_i\\ \beta_i\end{vmatrix}-\begin{vmatrix}\alpha_i'\\ \beta_i'\end{vmatrix}\right\|,\quad \mathrm{Z}_i^j=\begin{vmatrix}\alpha_i\\ \beta_i\end{vmatrix}-\begin{vmatrix}\alpha_i'\\ \beta_i'\end{vmatrix}$$

θ_0 is the initial value of rotation angle. $\begin{vmatrix}\alpha_i\\ \beta_i\end{vmatrix}$ and $\begin{vmatrix}\alpha_i'\\ \beta_i'\end{vmatrix}$ are parent and child respectively.
The renewal process is as follows:

$$\begin{bmatrix}\alpha_i'\\ \beta_i'\end{bmatrix}=\mathrm{U}(\theta_i)\begin{bmatrix}\alpha_i\\ \beta_i\end{bmatrix}=\begin{bmatrix}\cos(\theta_i) & -\sin(\theta_i)\\ \sin(\theta_i) & \cos(\theta_i)\end{bmatrix}\begin{bmatrix}\alpha_i\\ \beta_i\end{bmatrix}$$

- Loop computation is conducted on step 3. Then go back to step 2 and make a decision. The optimal solution and iterations is got until the overflow condition is met or the iteration times reach to the setting times.

In this study, the setting times is 200 and the result of iterative evolution is showed in Fig. 2. From Fig. 2, it shows that the fitness value increases with the increasing of iteration time. A stable region is got when the iteration time reaches 20. When the iteration time is near 100, the fitness value reaches the maximum. At this time, the obtained 2-dimensional code is 0000 0000 0101 1100 0111 1101 1110 000, among which 0 represents removing and 1 represents selecting. Then, the adjacent wavebands are combined into 4 wavebands. The total number of wavebands is 650. And the selected effective wavebands is showed in Fig. 3. From Fig. 3, it can be known that the selected wavebands is significantly less than the number of all band spectrum 1557.

After choosing effective wavebands by QGA, the prediction result of the model established by these effective wavebands is showed in Table 3. Compared with the prediction result modeled in all band in Table 2, it can be found out that the chosen effective wavebands not only can improve the prediction model by PLSR, but also the prediction model by PCR. Among them, the model established by FD+QGA+PLSR shows the best predictive effect. The related coefficient of calibration set reaches 0.995 and the RMSEP reduces to 0.489. This illustrated that QGA can choose effective spectral information from all the band and reduce interfering information which can improve the correlation and veracity of model obviously (Fig. 4).

CONCLUSION

This study used near infrared spectrum to detect the sugar content of red globe grape. By using 4 different spectra data pretreatment methods combined with 3 quantitative analysis models, modeling prediction analysis was done in the spectral bands. In order to improve the prediction ability of model and reduce interfering information, QGA-PLSR was used to establish prediction model. The results showed that QGA could extract effective wavebands which decreased the number of wavebands from 1557 to 650. At the same time, RC and RMSEP both were improved a lot. Model established by QGA-PLSR could predict the sugar content of red globe grape very well. This realized the rapid non-destructive detection of sugar content of red globe grape which provided technical support for the quality classification of red globe grape in later period.

ACKNOWLEDGMENT

This study was supported byNatural Science Fund of Hubei Province (2012FKB02910) and Research and Development Projects in Hubei Province (2011BHB01 6).

REFERENCES

Cayuela, J.A., 2008. Vis/NIR soluble solids prediction in intact oranges (*Citrus sinensis* L.) cv. Valencia Late by reflectance. Postharvest Biol. Tec., 47(1): 75-80.

Fang, W.G., L.X. Huang and Y. He, 2008. Study on grape berry sugar degree of visible or near infrared spectrum detection. Spectrosc. Spect. Anal., 28(9): 2090-2093.

Fei, D., T.S. Hong and X.J. Xue, 2011. The total soluble sugar sugar orange visible and near infrared nondestructive testing [J]. J. Agric. Mach., 42(4): 133-138.

Fei, J.D., M. Chen, L.J. Xie *et al*., 2014. To improve the NIR Prediction model suitable precision detection parts of watermelon. J. Agr. Eng., 30(9): 229-234.

Hua, W.J., Z. Cheng, Z.R. Li *et al*., 2010. A portable visible or near infrared diffuse transmission detection technology of pear hardness. J. Agric. Mach., 11: 129-133.

Jun, L.L., N. Chen, X.S. Liu *et al*., 2012. Rapid detection and purification process of grape seed polyphenols with macroporous resin by near infrared spectroscopy. Chem. Anal., 4: 26.

Liang, L.J. and C.Q. Wang, 2011. Application of quantum-inspired evolutionary algorithm in the analysis of near infrared diffuse transmission spectroscopy of apples. J. Zhejiang Univ., 37(4): 453-459.

Min, L., 2013. Grape grain size non-destruction detection based on machine vision. M.Sc. Thesis, Huazhong Agricultural University, Wuhan, China.

Rong, H.F., T. Pan, G.L. Zhang *et al*., 2010. Rapid determination of zinc content in soil using near infrared reflectance spectroscopy. Opt. Precis. Eng., 18(3): 587-590.

Slaughter, D.C., J.F. Thompson and E.S. Tan, 2003. Nondestructive determination of total and soluble solids in fresh prune using near infrared spectroscopy [J]. Postharvest Biol. Tec., 28(3): 437-444.

Wedding, B.B., R.D. White, S. Grauf, C. Wright, B. Tilse, P. Hofman and P.A. Gadek, 2011. Non-destructive prediction of 'Hass' avocado dry matter via FT-NIR spectroscopy [J]. J. Sci. Food Agr., 91(2): 233-238.

Wu, D., Y. He, S.J. Feng and Y.D. Bao, 2008. Application of infrared spectra technique based on LS-Support vector machines to the non-destructive measurement of fat content in milk power. J. Infrared Millim. W., 27(3): 181-184.

Xiang, Z.G., N. Li, W.D. Jin and L.Z. Hu, 2004. A novel quantum genetic algorithm and its application. Electron. Sinica, 32(3): 476-479.

Zhang, X.C., H.L. Gao, H.F. Huang, L. Guo and S. Xiao, 2014. Optimization design of mathematical morphology filter based on quantum genetic algorithm. J. Southwest Jiao Tong Univ., 49(3): 462-469.

Zhu, Z., X.Y. Li, P.W. Li *et al*., 2011. Detection of moldy chestnut GA-LSSVM and near infrared based on Fourier transforms. J. Agr. Eng., 27(3): 331-335.