Shortwave infrared hyperspectral imaging for detection of pH value in Fuji apple

Guo Zhiming[1,2], Huang Wenqian[1,3], Chen Liping[1,3*], Peng Yankun[2], Wang Xiu[1,3]

(1. *Beijing Research Center of Intelligent Equipment for Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China*;

2. *College of Engineering, China Agricultural University, Beijing 100097, China*;

3. *National Engineering Research Center of Intelligent Equipment for Agriculture, Beijing 100097, China*)

Abstract: pH value is regarded as one of the most important attributes that affect sensory characteristics and edible quality of apple.   The objective of the research was to explore the feasibility of applying shortwave infrared hyperspectral imaging system to detect the pH value of apple.   A shortwave infrared hyperspectral imaging system was developed over the wavelength region of 1 000-2 500 nm and used to acquire hyperspectral images of apple samples.   After reflectance calibration, mean reflectance spectral was calculated by averaging the intensity of all pixels within the roundness region of interest (ROI). Synergy interval partial least squares (siPLS) algorithms as an effective multivariable method was conducted on the calibration of regression model to estimate the pH value in Fuji apple.   The performance of the final model was back-evaluated according to root mean square error of calibration (*RMSEC*) and correlation coefficient (R_c) in calibration set, and tested in prediction set. The optimal prediction siPLS model was obtained with correlation coefficient (R_p) of 0.8474 and mean square error of prediction (*RMSEP*) of 0.0398.   The results indicated that shortwave infrared hyperspectral imaging combined with siPLS chemometrics could be an accurate and fast method for nondestructive prediction of pH value in Fuji apple.

Keywords: shortwave infrared hyperspectral imaging, synergy interval partial least squares, pH value, apple

DOI: 10.3965/j.ijabe.20140702.016

1   Introduction

As a leader of apple production, China produced 36 million tons of apples according to FAOSTAT, which accounted for half of the world's annual production.

'Fuji' is one of the mostly cultivated apple varieties in China.   It is planted in both East and West China, mainly in Shandong Province, Shaanxi Province, Xinjiang Uygur Autonomous Region, etc.   Apple is considered an important choice of a healthy diet and is one of the most frequently consumed fruits[1,2].   Apple taste and appearance are two important criteria for consumers' purchase decision making.   Apple quality characteristics include external factors such as appearance (size, shape, color, gloss, and consistency), texture, flavor, and internal factors (chemical, physical, and microbial). Among these quality parameters, pH value is regarded as one of the most important attributes that affects the edible quality of apple[3].   pH value is positively correlated with taste, and has a substantial influence on overall consumer satisfaction.   Despite its importance, pH value is one of the attributes that are most difficult to evaluate

before purchase, because it is invisible from appearance and highly variable[4,5]. Traditionally, it can be precisely measured by instrumental methods such as pH meter measurement. These methods are reliable but have the disadvantages of being slow and destructive for industrial applications. It is impossible to use these methods as routine analysis in the apple processing industry. Therefore, it is desirable to develop a fast, non-destructive, accurate, and on-line technique to predict pH value of apple.

Conventional spectral analysis techniques have proved some success as an objective method in fruit acidity prediction for titratable acidity (TA), malic acid, and citric acid. Optical techniques, particularly near infrared spectroscopy (NIRS), have become popular in evaluating fruit quality and physiology because they are generally fast, easy and nondestructive in implementation. Jamshidi, et al.[6] investigated the feasibility of reflectance Vis/NIR spectroscopy for taste characterization of Valencia oranges based on taste attributes including soluble solids content (SSC) and TA, as well as taste indices including SSC to TA ratio (SSC/TA) and BrimA. Xie, et al.[7] conducted an experiment to simultaneously measure TA, malic acid, and citric acid of bayberry fruit in a nondestructive manner using near-infrared (NIR) transmittance spectroscopy and chemometrics. Shao, et al.[8] evaluated the application of Vis/NIRS in measuring the quality characteristics of tomato 'Heatwave' (Lycopersicumesculentum), including fruit firmness (indicated by compression force and puncture force), SSC and acidity (pH value). Shao, et al.[9] also analyzed spectral with two types of preprocessing to build a best model for predicting the SSC and acidity (pH value) of bayberry juice. Unfortunately, the distribution information of the chemical composition of a sample cannot be obtained via this technique because conventional NIR spectroscopic instruments are considered as point-based scanning instruments that provide one spectrum of the target sample without giving any spatial information. In reality, there are some cases where spatial distribution of quality parameters is needed. Hyperspectral imaging (HSI) can be utilized as the

basis for developing such systems due to its high spectral and spatial resolution, non-invasive nature and capability for large spatial sampling areas[10,11]. With the development of optical sensors, hyperspectral imaging integrates spectroscopy and imaging techniques to provide spectral information as well as the spatial information of measured samples. Hyperspectral imaging is a technique whereby hundreds of reflectance images are captured over a broad wavelength range at contiguous and narrow intervals, forming a three-dimensional structure of multivariate data (hypercube). When the hyperspectral data are appropriately processed, it is possible to automatically identify the location of features that display specific spectral signatures and to map the gradient and spatial distribution of specific attributes[12]. HSI, as a powerful analytical tool, has recently emerged for nondestructive evaluation of fruit quality and safety, such as prediction of physico-chemical and sensory characteristics of table grapes[13], determination of quality attributes for strawberry[14], prediction of firmness and soluble solids content of blueberries, detection of apple mealiness[15], inspection of banana fruit quality and maturity stages[16], detection of early bruises in apples[17,18], and assessment of apple fruit firmness and soluble solids content[19]. However, hyperspectral imaging technology has not yet been directly executed for predicting pH value of apple for nondestructive quality detection.

Hyperspectral imaging technology in both visible and NIR regions have been widely used in predicting food quality and safety. The two spectral ranges were named visible and short-wave near infrared (VNIR) spectral range (400-1 000 nm) and long-wave near infrared spectral range (LW-NIR, 900-1 700 nm or 1 000-2 500 nm), respectively. Although the hyperspectral imaging system with the spectral range of 400-1 000 nm includes the visible region, and can be used for the color measurement, the hyperspectral imaging system with LW-NIR range has better ability of measuring internal quality attributes, because the molecular overtone and combination vibrations in VNIR are more serious than those in LW-NIR[20]. Therefore, it is of our aim to implement LW-NIR hyperspectral imaging for predicting

pH value in apple.

The specific objectives of this study were (1) to develop a hyperspectral imaging system in the spectral region of 1 000-2 500 nm for fast acquisition of LW-NIR hyperspectral images from apple sample, (2) to apply a feature selection technique to search for an optimized spectral interval and an optimized combination of spectral regions selected from informative regions in near-infrared spectroscopy for pH value determination, (3) to establish multivariate calibration models to quantitatively correlate spectral information to the corresponding pH values based on the whole range of spectra or spectra at the selected region.

2   Materials and methods

2.1   Sample preparation

To make the experimental results accurate and reliable and reduce the system error in the experiment, all apples free from any abnormal features such as defects, bruises, diseases and contamination were selected. Sample was collected at Feicheng Fuji Apple Research Station and at China National Research Center for Apple Engineering and Technology in Tai'an city, Shandong Province.   This orchard, located at 116°50′33″E and 36°14′6″N, has an elevation of 110 m above sea level, an annual temperature ranging between 12.5°C and 13.0°C, and an annual rainfall of 659 mm.   A total number of 160 apples were selected in this experiment.   Apples were stored at 4°C to keep the quality prior to testing. Fruits were removed from the storage and left in room temperature (20±1)°C before the shortwave infrared hyperspectral image acquisition.

2.2   Shortwave infrared hyperspectral imaging system

A laboratory shortwave infrared hyperspectral imaging system (Figure 1) was developed to acquire hyperspectral reflectance images of apple in the spectral region of 1 000-2 500 nm.   The system consisted of hyperspectral camera, illumination unit, frame grabber, sample transportation plate, and computer with data acquisition software.   The hyperspectral camera included an imaging spectrograph (ImSpector V25E, Spectral Imaging Ltd., Oulu, Finland) and a shortwave infrared charged couple device (CCD) camera (Xeva-2.5-320,

Xenics Infrared Solutions, Belgium).   The digital camera operates HgCdTe detector array (up to 2.5 μm) with 320×256 pixel resolution.   It outputs 14 bit data and comes in a 60 Hz or 100 Hz version.   The camera interfaces to a PC via standard USB 2.0.   The digital camera combines a four stage thermo-electrically cooled detector head.

The position and orientation of the hyperspectral camera were adjusted by trial and error in order to minimize shape distortions caused by misaligned line scanning and also to match up with the detector orientation.   Two 150 Watt tungsten halogen lamps provided by regulated power supply were used to achieve a uniform NIR illumination for the sample in the field of view of the optics.   An industrial computer system and Labview software (version 8.3, National Instruments Corporation, Austin, USA) are used to capture all the line-scan images and save them for further analysis. The whole setup is placed in a dark room to avoid any stray light from the surrounding environment.
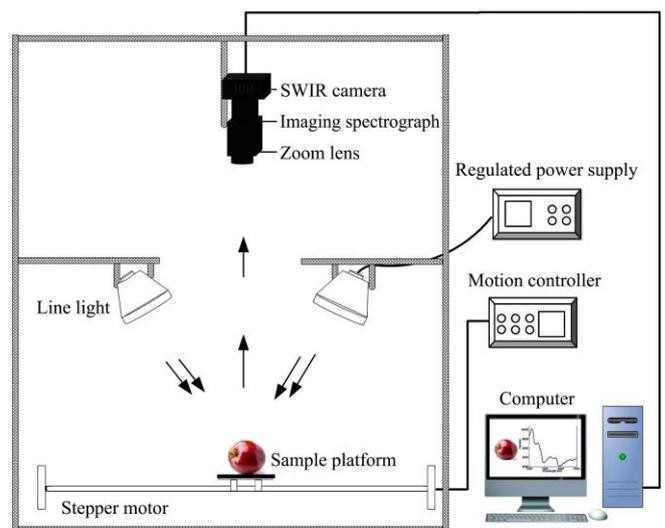
Figure 1   Schematic of the shortwave infrared hyperspectral imaging system for apple fruit

2.3   Hyperspectral image acquisition and correction

Each apple sample was placed on the translation platform and then moved at a speed of 45 mm/s under the field of view of the camera to be scanned line by line. Each image was recorded in the NIR region of 930-2 548 nm with 6.34 nm intervals between contiguous bands, producing a hyperspectral image with 256 wavelengths channels.   The spectral data for further processing were limited to 238 different wavelengths

ranging from 1 000 nm to 2 500 nm, as beyond this range the noise level of the camera was high, and the signal efficiency of the light source was low.

To eliminate the effects of dark current noise from camera and the non-uniformity of the lighting, the hypercube needed to be corrected with a white and a dark reference.   The dark image was collected by closed shutter, turning off all light sources and covering the lens with a black cap.   A PTFE white reference spectral with 99% reflection efficiency was used to obtain a white reference image.   The calibrated image (R) was then defined by the following Equation (1):

$$R_{\lambda,n} = \frac{S_{\lambda,n} - D_{\lambda,n}}{W_{\lambda,n} - D_{\lambda,n}} \times 16384 \tag{1}$$

where, λ, n means pixel n and at wavelength λ; R is the intensity of a raw image; W is the intensity of the white reference spectral on plate; and D is the intensity of the dark image.   Since the shortwave infrared CCD camera had 14-bit digital output, a coefficient of 16 384 was used to scale the pixel values of the images back to the original intensity range of the system.

2.4   Extraction of spectral data

To conduct spectral data extraction from each sample in the hyperspectral image, the region of interest (ROI) function of HIS Analyzer software was used to extract the sample spectral from the center region.   A square ROI with a size of 50×50 pixels around the center of each image was selected.   The reflectance spectrum in Figure 3 was obtained by averaging ROI for a hyperspectral image.   The spectral profiles of apple samples from 1 000 nm to 2 500 nm are shown in Figure 3.   It shows the spectral profile of calibration set sample from each origin. This process was executed manually one by one.

2.5   Spectral preprocessing

Spectral pretreatment is usually performed to mathematically treat the extracted spectral data to correct undesired effects such as light scattering and random noise resulting from variable physical sample properties or instrumental effects.   In the present work, it was only applied to the extracted spectral data, where scattering corrections are frequently applied.   Multiplicative scattering correction (MSC) is used to correct for light scattering variations in reflectance spectroscopy.   The

standard normal variate (SNV) transformation is used to remove interferences due to light scattering and path length variations.   Spectral derivatives may also be employed to improve resolution and to highlight the selectivity towards a particular analyte when strong multicollinearity is presented.

2.6   Multivariate calibration methods

In general, partial least square (PLS) regression is the most commonly used multivariate method for constructing calibration models to predict the constituent of interest[21].   It is proved to be very efficient for multivariate calibration when the measured variables such as spectral data are highly correlated[22,23].   However, the selection of informative spectral regions should be considered to quantify highly complicated samples such as biological samples.   A drawback of the technique of feature selection when it is applied to spectral data is that the selected features are usually scattered throughout the spectrum.

An approach dividing the full spectrum into regions of defined size, synergy interval partial least squares (siPLS) regression was proposed by Norgaard and co-workers[24,25].   The basic principle of siPLS is given as follows: First, it splits the spectra into some smaller equidistant regions.   Next, PLS regression models are developed in all possible combinations of two, three or four intervals.   Thereafter, root mean square error of calibration (RMSEC) is calculated for every combination of intervals.   The combination of intervals with the lowest RMSEC is chosen.

The performance of the calibration model was evaluated in terms of RMSEC.   Root mean square error of prediction (RMSEP) was used to evaluate the performance of the prediction set in the prediction process.   Correlation coefficient (R) between the predicted and the measured value were calculated for both the calibration and the prediction sets.   Generally, a good model should have lower RMSEC, RMSEP and higher R but small differences between RMSEP and RMSEC.

3   Results and discussion

3.1   pH reference values

The reference descriptive values of apple pH were

measured using the pH meter (model FE20 with Inlab micro) manufactured by Mettler Toledo Scale Company Shanghai corp., Switzerland.   The mean, standard deviation and the range for pH value by reference measurement was summarized in Table 1.   Before making calibration, samples with a standard deviation on the measured result, three times higher than the mean standard deviation of all samples are deleted. No outlying replicate is detected.   Then, all 173 samples were randomly divided into two subsets.   The first subset was called the calibration set with 115 samples to be used for model building, while the other one was called the prediction set with 58 samples to be used for testing the robustness of the model.   Both datasets have comparable ranges, means, and standard deviations for all measured parameters, which are important for calibration models to be used later in the prediction.   The results of the T-test revealed that there was no significant difference between the calibration and validation sets for pH value.   The pH value distribution of the calibration set samples shown in Figure 2 is substantially normal, with good representation.
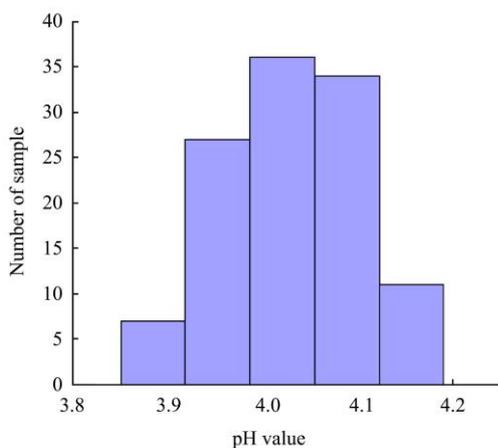
Figure 2   pH value distribution of apple samples of the calibration set

Table 1   Descriptive statistics of pH value measured by the standard reference method

Subset	Samples No.	Range	Mean	Standard deviation
Calibration set	115	3.85-4.19	4.0627	0.0781
Prediction set	58	3.86-4.19	4.0060	0.0737

3.2   Spectral profiles

LW-NIR reflectance spectral profiles of calibration set apple samples extracted from the ROI of hyperspectral image was shown in Figure 3.   Scrutinized from the

spectra of the original data (Figure 3), it is readily apparent that no outliers can be observed by naked eyes and all the spectra were quite homogeneous.   As obviously noticed in the spectra, there were two pronounced valleys observed at 1 450 and 1 940 nm were related to the O–H first stretching overtone and bending combination of water, respectively.
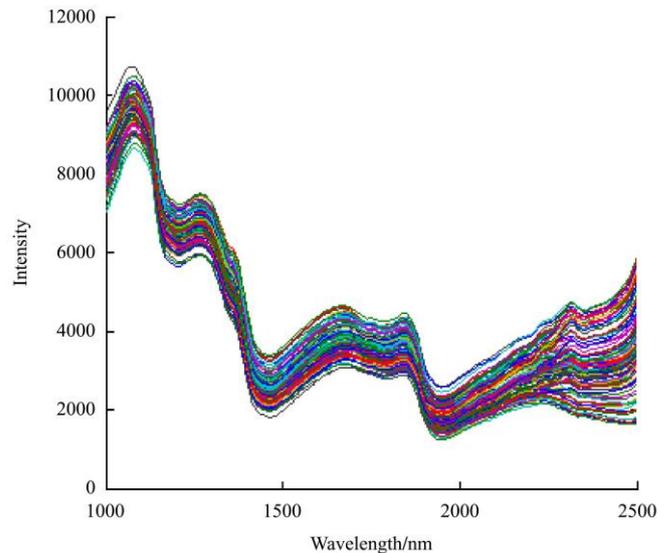
Figure 3   LW-NIR reflectance spectral profiles of apple samples of the calibration set

3.3   Results of PLS model

In the application of PLS algorithm, it is generally known that the number of PLS factor is a critical parameter in model calibrating.   PLS regression model was developed under full cross validation using the full spectral range (238 wavelengths), and the optimum number of PLS factors was identified at the lowest value of RMSEC.   Table 2 shows the performance of PLS models on the raw spectra and those developed with various pre-treatment routines (SNV, MSC, 1st derivative and 2nd derivative) for the prediction of pH value.   The results indicated that 1st derivative spectral pretreatment methods offered an improvement in the model accuracy. All others even had deteriorative influence on the model predictability compared with raw spectra.   Results presented in Table 2 showed that the PLS models developed using the 1st derivative spectra used in predicting pH value are reasonably good with R_c at 0.8494, RMSEC at 0.0411, R_p at 0.8407 and RMSEP at 0.0412 with 8 PLS factors.   However, the prediction models could also be further enhanced by including more

samples or building calibration models based on non-linear approaches.

Subset	PLS factor	R_c	RMSEC	R_p	RMSEP
Raw	10	0.8431	0.0419	0.8204	0.0435
SNV	9	0.8335	0.0430	0.8102	0.0441
MSC	9	0.8038	0.0481	0.7779	0.0487
1st derivative	8	0.8494	0.0411	0.8407	0.0412
2nd derivative	8	0.8344	0.0429	0.8143	0.0437

3.4   Results of siPLS model

The siPLS algorithm develops PLS regression models for all possible combinations of two, three or four intervals. The number of intervals was also optimized according to the RMSEC in siPLS model calibration. The combination of intervals with the lowest RMSEC is chosen. Considering the upper results, we selected the 1st derivative pretreatment spectral for the further analysis. When four intervals were considered, the best siPLS model was achieved. The number of intervals was 19, and the 3, 5, 8 and 10 intervals were chosen for pH value siPLS model. The optimal combination of siPLS model was obtained when combined four subintervals. The best number of intervals was 19, and the 3, 5, 8 and 10 intervals were chosen for pH value siPLS model. The spectral regions labeled with wavelength were shown in Figure 4. According to the investigation of the spectrum, we selected spectral regions to build a PLS model, but the absorption band around 1 940 nm, corresponding to O-H stretching and deformation was excluded in the analysis. Figure 5 is the scatter plot showing a correlation between reference measured and spectral predicted in the calibration set by siPLS model. Here, the value of RMSEC is 0.0406, and correlation coefficient R_c is 0.8529 in the calibration set. When the performance of siPLS model was evaluated by the samples in the prediction set, the RMSEP is 0.0398 and R_p is 0.8474 in the prediction set.

For full-spectrum PLS model, all variables were used to calibrate global model. In some NIR spectral regions, there may contain some useless or irrelevant information, which can worsen the predictability of the model. The elimination of irrelevant variables can predigest

calibration modeling and improve the results in terms of accuracy and robustness. In addition, there were some collinear variables in NIR spectral region, and they inevitably weaken the performance of PLS model. Synergy interval PLS actually gives overview spectral data to select the interesting spectral region and remove some noisy regions. SiPLS combine several informative spectral regions in building calibration models. Therefore, siPLS selected fewer spectral variables than full-spectrum PLS models, and got equal performance. This result was slightly better than the previous models developed with the full spectral range, indicating that the method for feature region selection using siPLS was efficient.
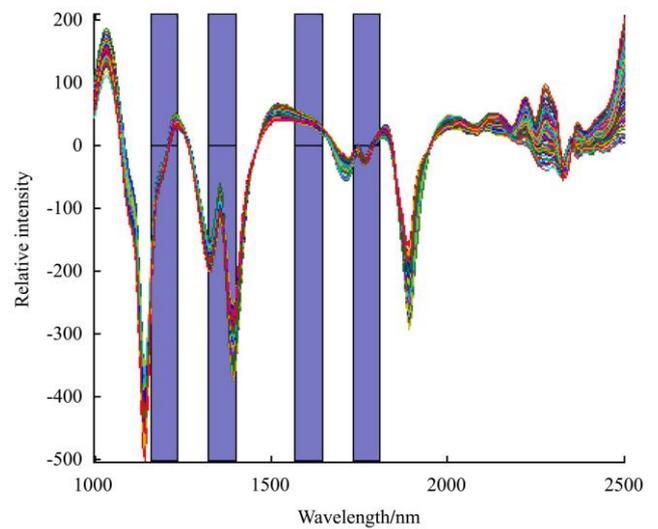
Figure 4   Spectral intervals selected by siPLS method
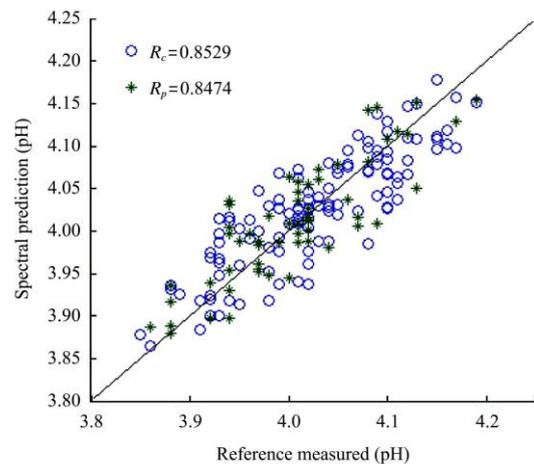
Figure 5   Predicted values of the calibration (○) and prediction (∗) sets against reference values by siPLS

Comparing with the investigation of Jamshidi, et al.[6] and Shao, et al.[8], Vis/NIR spectroscopy was performed

to measure TA in oranges and bayberry fruit, respectively. The best models obtained with PLS method had predicted correlation coefficients at 0.86 and 0.83, respectively. The result was similar with ours for apple. Therefore, more research should be conducted to improve prediction accuracy of fruit pH value using shortwave infrared hyperspectral imaging technique.

4   Conclusions

To predict pH value in apple, a systematic approach was developed using LW-NIR hyperspectral imaging in conjunction with PLS regression as well as siPLS. This study has shown that LW-NIR hyperspectral imaging offers an alternative to analytical methods for rapid and non-destructive prediction of pH value in apple. The partial least squares regression analysis showed a strong performance in predicting pH value with 1st derivative pretreated spectral with the full range (R_c at 0.8494, RMSEC at 0.0411, R_p at 0.8407 and RMSEP at 0.0412). Meanwhile, siPLS algorithm revealed its superiority to enhance the model for pH value prediction by selected the optimal combination of intervals. It can be concluded that LW-NIR hyperspectral imaging with siPLS algorithm can significantly improve the efficiency of quality control and assurance of agro-food production. More investigation is needed to improve the accuracy and robustness of the prediction models for on-line application.

Acknowledgements

[References]

[1]   Vedwan N. Culture, climate and the environment: Local knowledge and perception of climate change among apple growers in Northwestern India. Journal of Ecological Anthropology, 2006; 10: 4-18.

[2]   Wang Y, Shao M, Zhu Y, Liu Z. Impacts of land use and plant characteristics on dried soil layers in different climatic regions on the Loess Plateau of China. Agricultural and Forest Meteorology, 2011; 151(4): 437-448.

[3]   Ben fez E I, Genovese D B, Lozano J E. Effect of pH and ionic strength on apple juice turbidity: Application of the extended DLVO theory. Food Hydrocolloids, 2007; 21(1): 100-109.

[4]   Harker F R, Kupferman E M, Marin A B, Gunson F A, Triggs C M. Eating quality standards for apples based on consumer preferences. Postharvest Biology and Technology, 2008; 50(1): 70-78.

[5]   Roble C, Auty M A, Brunton N, Gormley R T. Evaluation of fresh-cut apple slices enriched with probiotic bacteria. Innovative Food Science & Emerging Technologies, Volume 11, 2010; 11(1): 203-209.

[6]   Jamshidi B, Minaei S, Mohajerani E, Ghassemian H. Reflectance Vis/NIR spectroscopy for nondestructive taste characterization of Valencia oranges. Computers and Electronics in Agriculture, 2012; 85: 64-69.

[7]   Xie L, Ye X, Liu D, Ying Y. Prediction of titratable acidity, malic acid, and citric acid in bayberry fruit by near-infrared spectroscopy. Food Research International, 2011; 44(7): 2198-2204.

[8]   Shao Y, He Y, Gómez A H, Pereir A G, Qiu Z, Zhang Y. Visible/near infrared spectrometric technique for nondestructive assessment of tomato 'Heatwave' (Lycopersicumesculentum) quality characteristics. Journal of Food Engineering, 2007; 81(4): 672-678.

[9]   Shao Y, He Y. Nondestructive measurement of the internal quality of bayberry juice using Vis/NIR spectroscopy. Journal of Food Engineering, 2007; 79(3): 1015-1019.

[10]   Kamruzzaman M, ElMasry G, Sun D, Allen. Non-destructive assessment of instrumental and sensory tenderness of lamb meat using NIR hyperspectral imaging. Food Chemistry, 2013; 141: 389-396.

[11]   Baiano A, Terracone C, Peri G, Romaniello R. Application of hyperspectral imaging for prediction of physico-chemical and sensory characteristics of table grapes. Computers and Electronics in Agriculture, 2012; 87: 142-151.

[12]   Kamruzzaman M, ElMasry G, Sun D, Allen. Non-destructive prediction and visualization of chemical composition in lamb meat using NIR hyperspectral imaging and multivariate regression. Innovative Food Science and Emerging Technologies, 2012; 16: 218-226.

[13]   ElMasry G, Wang N, ElSayed A, Ngadi M. Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. Journal of Food Engineering, 2007; 81: 98-107.

[14]   Leiva-Valenzuela G A, Lu R, Aguilera J M. Prediction of firmness and soluble solids content of blueberries using

hyperspectral reflectance imaging. Journal of Food Engineering, 2013; 115: 91-98.

[15] Huang M, Lu R. Apple mealiness detection using hyperspectral scattering technique. Postharvest Biology and Technology, 2010; 58: 168-175.

[16] Rajkumar P, Wang N, EImasry G, Raghavan G S V, Gariepy R Y. Studies on banana fruit quality and maturity stages using hyperspectral imaging. Journal of Food Engineering, 2012; 108: 194-200.

[17] ElMasry G, Wang N, Vigneault C, Qiao J, ElSayed A. Early detection of apple bruises on different background colors using hyperspectral imaging. LWT-Food Science and Technology, 2008; 41(2): 337-345.

[18] Baranowski P, Mazurek W, Wozniak J, Majewska U. Detection of early bruises in apples using hyperspectral data and thermal imaging. Journal of Food Engineering 110 2012; 110: 345-355.

[19] Peng Y, Lu R. Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit firmness and soluble solids content. Postharvest Biology and Technology, 2008; 48: 52-62.

[20] Wu D, Sun D, He Y. Application of long-wave near infrared hyperspectral imaging for measurement of color distribution in salmon fillet. Innovative Food Science and Emerging Technologies, 2012; 16: 361-372.

[21] Guo Z, Chen Q, Chen L, Huang W, Zhang C, Zhao C J. Optimization of informative spectral variables for the quantification of EGCG in green tea using Fourier transform near-infrared (FT-NIR) spectroscopy and multivariate calibration. Applied Spectroscopy, 2011; 65(9):1062-1067.

[22] Taghizadeh M, Gowen, A, O'Donnel, C P. Prediction of white button mushroom (Agaricusbisporus) moisture content using hyperspectral imaging. Sensing & Instrumentation Food Quality, 2009; 3(4): 219-226.

[23] Zou X, Zhao J, Malcolm J W, Mel H, Mao H. Variables selection methods in near-infrared spectroscopy. Analytica Chimica Acta, 2010; 667(1): 14-32.

[24] Leardi R, Nørgaard L. Sequential application of backward interval PLS and genetic algorithms for the selection of relevant spectral regions. Journal of Chemometrics, 2004; 18(11): 486-497.

[25] Norgaard L, Saudland A, Wagner J, Nielsen J P, Munck L, Engelsen S B. Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. Applied Spectroscopy, 2000; 54(3): 413-419.