



2950 Niles Road, St. Joseph, MI 49085-9659, USA  
269.429.0300 fax 269.429.3852 hq@asabe.org www.asabe.org

An ASABE Meeting Presentation

Paper Number: 097428

## Optimal Wavelengths Selection Using Hierarchical Evolutionary Algorithm for Prediction of Firmness and Soluble Solids Content in Apples

**Min Huang, Ph.D., Associate Professor**

1. School of Communication and Control Engineering, Jiangnan University, China, 214122
2. USDA ARS, 207 Farrall Hall, MSU, East Lansing, MI 48824. e-mail: huangmzqb@163.com

**Renfu Lu, Ph.D., Supervisory Agricultural Engineer**

USDA ARS Sugarbeet & Bean Research Unit, 224 Farrall Hall, MSU, East Lansing, MI 48824

Written for presentation at the  
**2009 ASABE Annual International Meeting**  
Sponsored by ASABE  
**Grand Sierra Resort and Casino**  
Reno, Nevada  
June 21 – June 24, 2009

**Abstract.** *Hyperspectral scattering is a promising technique for noninvasive measurement of quality attributes of apple fruit, and extraction of the most useful information from hyperspectral scattering data is critical to prediction of fruit firmness and soluble solids content (SSC). A hierarchical evolutionary algorithm (HEA) approach coupled with subspace decomposition and partial least squares (PLS) regression was proposed to select the optimal wavelengths from the hyperspectral scattering profiles of 'Golden Delicious' apples for predicting fruit firmness and SSC. Seventeen optimal wavelengths were selected for firmness, which nearly spanned the entire spectral range of 500 - 1,000 nm, and 16 optimal wavelengths, none of them below 600 nm, were selected in the SSC prediction model. The model using 17 optimal wavelengths for predicting firmness yielded better results ( $r_p = 0.857$ , root mean square error of prediction or RMSEP = 6.2 N) than the full spectrum model ( $r_p = 0.848$ , RMSEP = 6.4 N). For predicting SSC, the model using 16 optimal wavelengths model also yielded better results ( $r_p = 0.822$ , RMSEP = 0.78%) than the full spectrum model ( $r_p = 0.802$ , RMSEP = 0.83%). The proposed HEA approach provided an effective means for optimal wavelengths selection and improved the prediction of firmness and SSC in apples compared with the approach of using full spectrum.*

**Keywords.** Hyperspectral scattering, Hierarchical evolutionary algorithm, Partial least squares, Subspace decomposition, Fruit, Apple, Firmness, Soluble solids content, Nondestructive sensing.

---

The authors are solely responsible for the content of this technical presentation. The technical presentation does not necessarily reflect the official position of the American Society of Agricultural and Biological Engineers (ASABE), and its printing and distribution does not constitute an endorsement of views which may be expressed. Technical presentations are not subject to the formal peer review process by ASABE editorial committees; therefore, they are not to be presented as refereed publications. Citation of this work should state that it is from an ASABE meeting paper. EXAMPLE: Author's Last Name, Initials. 2009. Title of Presentation. ASABE Paper No. 09----. St. Joseph, Mich.: ASABE. For information about securing permission to reprint or reproduce a technical presentation, please contact ASABE at rutter@asabe.org or 269-429-0300 (2950 Niles Road, St. Joseph, MI 49085-9659 USA).

---

## Introduction

Firmness and soluble solids content (SSC) are two key parameters in evaluating and grading postharvest quality of apples. Current standard methods for firmness and SSC measurement would destroy fruit samples in the testing process (Park et al. 2003). Hence much research has been reported over the past decades on the development of nondestructive technologies for measuring firmness and SSC. Among them is near-infrared (NIR) technology that provides an effective means for measuring SSC (Lammertyn et al., 1998; Lu, 2001; Lu and Arinana, 2002). However, NIR measurement does not correlate well with the standard Magness-Taylor (MT) firmness measurement (McGlone and Kawano, 1998; Lu et al., 2000; Lu and Ariana, 2002).

In recognizing the shortcomings of NIR technology in quantification of light absorption and scattering – the two basic phenomena when light interacts with plant materials, Lu (2003, 2004) proposed a spectral scattering technique to measure diffuse reflectance profiles at the surface of fruit resulting from the incidence of a small light beam, for assessing apple firmness and SSC. The technique may be implemented in multispectral imaging mode (Lu, 2004) for selected wavelengths or in hyperspectral imaging mode (Lu, 2003) for a broad spectral region. Hyperspectral scattering images provide a large amount of spatial and spectral information about a sample. It is thus critical that an appropriate method be used in analyzing hyperspectral scattering features to predict apple firmness and SSC. Qin and Lu (2008) determined absorption and reduced scattering coefficients from the spatial scattering profiles of apples using a fundamental diffusion theory model, and these coefficients were then correlated to fruit firmness and SSC. Lu (2007) calculated mean reflectance from the hyperspectral scattering images for a specific scattering distance and then developed neural network models to predict fruit firmness and SSC. Peng and Lu (2008) proposed a modified Lorentzian distribution function with four parameters to characterize spatial scattering profiles for the wavelengths of 450 -1,000 nm.

A critical problem encountered in analyzing hyperspectral scattering images is extraction of meaningful information from such large quantity of data. One approach is to use conventional linear methods such as principal component analysis or partial least squares (PLS) to relate linear combinations of spectral wavelengths to independent physical variables of interest. These approaches are powerful, but have limitations in that they only apply linear relationships between spectral variables and do not identify specific wavelengths which are most important for assessing quality of apple such as firmness and SSC (Steward, 2005). Another approach is to select the optimal wavelengths using multi-linear regression (Lu and Peng, 2006). This wavelengths selection method is time-consuming and may not be optimal.

Evolutionary computation is an intelligent optimization technology that has been widely used in various fields. It is an adaptive search method that is especially powerful for dealing with difficult search problems without being caught at local extremes in the search space (Goldberg, 1989). As a sub-class of the method, genetic algorithm (GA) applies the “survival of the fittest” approach to modeling data. The GA method has been used in NIR spectral analysis for assessing SSC of fruit coupled with PLS (Leardi and Lupiáñez González, 1998; Leardi, 2000; Kleyen et al., 2003; Zou et al., 2007). In these reported studies, there was only one single parameter for each wavelength, i.e., one spectrum for each sample. It would be difficult to encode in GA when multi-parameters respond to one wavelength, as in the case of the modified Lorentzian distribution function with four parameters (Peng and Lu, 2008) or in the diffusion theory model with two parameters for each wavelength (Qin and Lu, 2008). As a variant of GA, hierarchical evolutionary algorithm (HEA) has an intrinsic property of encoding the parameters of considered problem in a hierarchical manner (Tang et al, 1998). This particular property makes HEA potentially useful for automatic spectral wavelengths selection.

Hence, the objectives of this research were to:

- Use HEA to search optimal wavelengths in the multi-parameter scattering spectra that would be most closely related to firmness and SSC, and
- Develop calibration models using partial least squares method to predict the firmness and SSC of apples.

## Materials and Methods

### *Apples Samples and Data Acquisition*

Six hundred 'Golden Delicious' apples, harvested from the orchards of Michigan State University Horticultural Teaching and Research Center in Holt, MI and Clarksville Horticultural Experiment Station in Clarksville, MI during the 2006 harvest season, were used in the experiment. The apples were kept in controlled atmosphere storage (2% O<sub>2</sub> and 3% CO<sub>2</sub> at 0 °C) for about five months prior to the experiment. The equatorial diameter of each sample was measured using a digital caliper and this information was later used for correcting the spectral scattering profiles.

Two-dimensional hyperspectral scattering images were obtained from the equator of each fruit using a hyperspectral imaging system (Qin and Lu, 2008) developed at the U.S. Department of Agriculture Agricultural Research Service (USDA/ARS) postharvest engineering laboratory in East Lansing, Michigan. A detailed description of the hyperspectral imaging system and experimental procedure is given in Qin and Lu (2008).

After imaging, standard firmness and SSC measurements were made from the same imaging area of each apple using a Texture Analyzer (model TA.XT2i, Stable Micro Systems, Inc., Surrey, U.K) with a Magness-Taylor (MT) firmness test probe of 11 mm diameter and a digital refractometer (model PR-101, Atago Co., Tokyo, Japan).

### *Analysis of Hyperspectral Scattering Images*

A typical raw hyperspectral scattering image for the test apples is shown in figure1 (left), where the horizontal axis represents a spatial dimension and the vertical axis shows spectral. The spatial resolution was 0.20 mm per pixel along the spatial axis and the spectral resolution was 4.54 nm per pixel along the spectral axis. Hence, each scattering image in effect consisted of more than one hundred of spatial scattering profiles, each representing a specific wavelength. The interested region covering 500 - 1,000 nm and a total spatial distance of 20 mm were selected from each image.

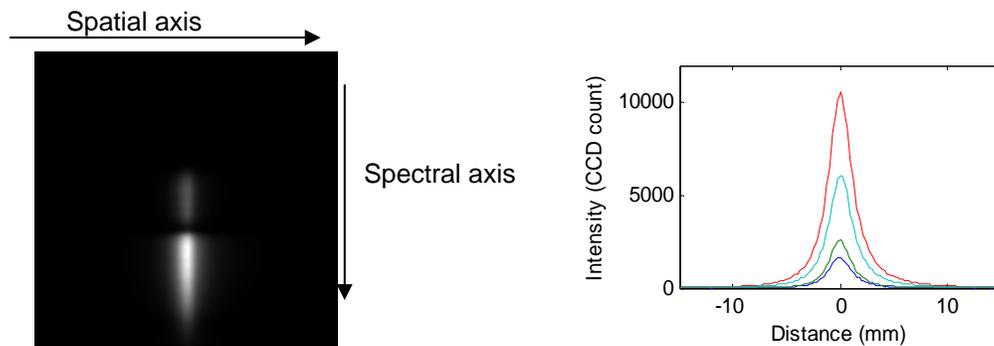


Figure 1. Hyperspectral scattering image of an apple (left), and raw spatial scattering profiles at four wavelengths (right).

In comparing different empirical mathematical models, Peng and Lu (2006) reported that a modified Lorentzian function with four parameters provided better fit to the scattering profiles and Lorentzian parameters correlated well with fruit firmness. Hence, the modified Lorentzian function with four parameters was used to fit the spatial scattering profiles at different wavelengths:

$$R(x) = a + \frac{b}{1 + \left(\frac{x}{c}\right)^d} \quad (1)$$

Where  $R$  is the reflectance;  $x$  is the scattering distance;  $a$  is the asymptotic value;  $b$  is the peak value of estimated light intensity at the distance closest to the light incident point;  $c$  is the full scattering width at half maximal value (FWHM);  $d$  is the slope around the FWHM area. Since the spatial scattering profiles were symmetrical about the incident point (or the zero point position in figure 1(right)), they were first averaged over the symmetrical point. The averaged profiles were then fitted by equation (1) for a scattering distance of 10 mm, from which estimated values for the four parameters ( $a$ ,  $b$ ,  $c$ , and  $d$ ) were obtained. The parameter estimation procedure was implemented using a nonlinear least-square inverse algorithm in Matlab (Peng and Lu, 2006). These parameters uniquely characterized the scattering features of each apple at a specific wavelength. Prior to the curve fitting, each scattering profile was corrected for nonuniform instrument response and fruit size by following the procedures described in Qin and Lu (2008).

After Lorentzian parameters for the apple samples were obtained, the same procedure was applied to extract parameters from the scattering profiles for a reference standard (a white Teflon disk). Each Lorentzian parameter for the apple samples was divided by the corresponding Lorentzian parameter for the reference and was subsequently multiplied by a constant (i.e., first value of the parameter for every 10 samples) to avoid changing the physical meanings of the parameters (Peng and Lu, 2007).

### ***Optimal Wavelengths Selection***

The 600 samples were arranged in ascending order for fruit firmness (or SSC). The first apple was selected for validation, and the second and third apples were selected for calibration for every three apples. This procedure resulted in 400 apples (66.7%) for the calibration set and 200 (33.3%) for the validation set. The calibration set was used to select the optimal wavelengths.

Hyperspectral scattering images are represented by high dimensional spatial and spectral data, which contain redundant information between wavelengths. Hence there exists strong correlation between adjacent wavelengths. Figure 2 shows correlation curves of 675 nm, 720 nm, and 970 nm with other wavelengths for the spectral region of 500 - 1,000 nm. High correlations between the adjacent wavelengths suggested the need of selecting the optimal wavelengths to reduce processing and computing time for the hyperspectral image data.

The optimal wavelengths selection was based on these two criteria: 1) the selected wavelengths had the most information, and 2) the selected wavelengths had smallest relevance with other wavelengths.

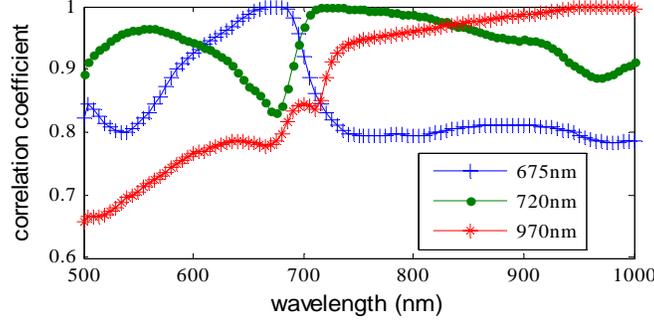


Figure 2. Correlations of 675 nm, 720 nm, and 970 nm with other wavelengths.

### Subspace Decomposition

To reduce the code length of hierarchical chromosome and simplify the complexity of hierarchical evolutionary algorithm, subspace decomposition and adaptive wavelength selection in each subspace were performed.

Subspace decomposition mainly depends on the correlation matrix  $F$  between different wavelengths. Elements of the correlation matrix  $F$  are defined as:

$$f_{i,j} = \frac{\sum_{k=1}^n (x_{k,i} - \mu_i)(x_{k,j} - \mu_j)}{\sqrt{\sum_{k=1}^n (x_{k,i} - \mu_i)^2} \sqrt{\sum_{k=1}^n (x_{k,j} - \mu_j)^2}} \quad (2)$$

for  $i = 1, 2, \dots, 101$  and  $j = 1, 2, \dots, 101$  for wavelength,  $n$  is the number of samples for calibration,  $x_{k,i}$  is a Lorentzian function parameter ( $a$ ,  $b$ ,  $c$  or  $d$ ) for the  $k^{th}$  sample at  $i^{th}$  wavelength, and  $\mu_i$  and  $\mu_j$  are the mean value of all samples at  $i^{th}$  and  $j^{th}$  wavelength. Note that  $f_{i,j} = f_{j,i}$  for all  $i = j$ . Four correlation matrices for the four Lorentzian parameters were obtained by equation (2) and the final correlation matrix  $F$  used in this study consisted of maximum correlation coefficients of four matrices for any two wavelengths. If the correlation between two wavelengths was greater than  $T_r$ , a given threshold value, the wavelengths between these two wavelengths would constitute one subspace. All wavelengths in this subspace would have similar correlations.

Based on the correlation matrix  $F$  and the threshold value, the full spectrum space with 101 wavelengths was adaptively decomposed into different subspaces with different numbers of wavelengths (figure 3.). By adjusting the threshold value, the number of wavelengths for each subspace and the number of subspaces were adaptively changed. The number of selected wavelengths in each subspace was reduced as the threshold value increased, while the number of subspaces increased. In this study,  $T_r$  was chosen to be 0.93 for subspace decomposition.

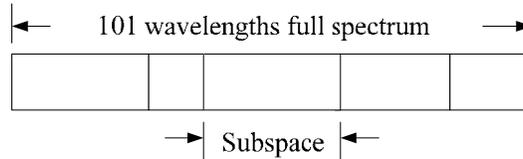


Figure 3. Subspace decomposition.

## Wavelengths Selection in Subspace

To satisfy the principle of optimal wavelengths selection, wavelength index was used as the criterion to select appropriate wavelengths in each subspace. Wavelength index was first calculated in each subspace by equation (3). All wavelengths in the subspace were then sorted by the index on descending order to prepare for wavelengths selection.

$$Index_i = \frac{\sigma_i}{(f_{i-1,i} + f_{i,i+1})/2} \quad (3)$$

where  $Index_i$  is the index value of  $i^{th}$  wavelength and  $\sigma_i$  is the calibration correlation coefficient based on a single wavelength for calibration modeling. The greater the coefficient, the more information the wavelength provides.

In every sorted subspace, wavelengths with higher index would be selected for constituting a new entire space. Several methods were researched for selecting wavelength combinations with higher index of each subspace. One is the same number (M) of wavelengths selected from each subspace. Another is based on the given index threshold  $T_w$ ; a wavelength is selected if its index is greater than  $T_w$ . Since the number of wavelengths and wavelength index in each subspace are non-uniform, these two methods, depending on M and  $T_w$ , cannot guarantee selection of wavelengths from every subspace. To ensure a reasonable quantity of wavelength selections from every subspace and sufficient search space for HEA, the same selection ratio  $R_s$  ( $=0.7$ ) for each subspace was used for firmness and SSC. Hence the number of wavelengths selected for the subsequent implementation of HEA was reduced from 101 to 71.

## Optimal Wavelengths Selection Using Hierarchical Evolutionary Algorithm

After the representative wavelengths with more information and lower correlation between them were selected, the next step was to select the optimal wavelengths from the representative wavelengths in order to develop models predicting the firmness and SSC of apples. To overcome the disadvantages of conventional GA in encoding and mating for multi-parameters responding to one wavelength, a hierarchical evolutionary algorithm was proposed.

The hierarchical chromosome consists of a control layer chromosome and a parameter chromosome. The selected wavelengths after subspace decomposition were encoded using the hierarchical chromosome shown in figure 4. The control layer was encoded for wavelengths as a binary string and the parameter layer was encoded for Lorentzian function parameters as a real number string. The bits of '1' in the control layer indicate that these wavelengths were selected in the evolutionary process and the corresponding parameter chromosomes were active, and the bits of '0' in the control layer indicate that these wavelengths were excluded and the corresponding parameter chromosomes were inhibitive. Generally, the control layer and the parameter layer chromosome are mating at the same evolutionary process (He et al., 2002; Lai and Chang, 2009) to solve different problems. In this study, the control layer was used to select wavelengths and the parameter layer only represented scattering features (i.e., Lorentzian parameters). Thus, the genes in the parameter layer weren't mating in wavelengths selection process.

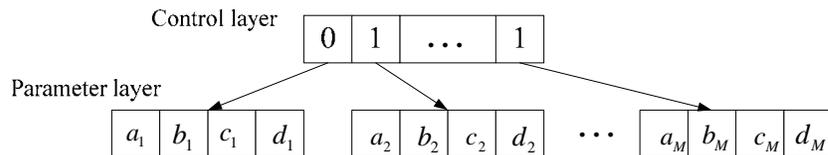


Figure 4. Hierarchical chromosome for wavelengths selection.

Similar to the process of implementation with GA, evaluation, crossover, mutation and selection operators were used based on fitness in HEA. PLS regression model and cross validation were adopted for evaluation. The performance of the PLS model was evaluated in terms of root mean square error of cross validation (RMSECV). For RMSECV, a modified contiguous blocks (calibration set separated by five groups) cross validation was used by HEA. Although the samples were still selected and left out in contiguous blocks, the starting position of the first block shifted randomly through the data. This helped reduce the likelihood of over-fitting. The HEA was completed after a finite number of iterations or after some percentage of individuals in the population were using identical variable subsets.

The control parameters of HEA were population size, crossover probability and mutation probability and maximum iterative generation or identical variable number. After considering the complexity and running time of HEA, we chose a population size of 64, crossover probability of 0.5, mutation probability of 0.005 and the maximum iterative generation of 100 or identical variable percentage of 50% in this study.

After the termination of iterations, wavelengths were sorted by selected frequency on descent order and added to the PLS regression model one by one. Optimal wavelengths for modeling were determined based on the smallest RMSECV. Because HEA is a random search algorithm, the selection of initial population and the implementation process of genetic operators (selection, crossover and mutation) would be highly random. Therefore, the algorithm was run five times to reduce the randomness effect in the process.

### ***Development and Validation of Calibration Models***

Partial least squares (PLS) method was used to build calibration models using optimal wavelengths. PLS normally requires single spectra for all samples, while the Lorentzian function generated four parameter spectra for each sample, which differed greatly in their scale of values. An autoscaling method, a normalization procedure, was used to overcome the problem of large disparity in values among the four parameters.

$$Y_{N,i} = \frac{Y_i - \bar{Y}}{SD} \quad (4)$$

where  $Y_{N,i}$  is the rescaled parameter,  $Y_i$  is the original parameter of  $a$ ,  $b$ ,  $c$  or  $d$  for sample  $i$  at a given wavelength  $\lambda$ , and  $\bar{Y}$  and  $SD$  denote the mean value of the calibration samples and the corresponding standard deviation, respectively. These rescaled parameters have zero means and unit variance, and each parameter would preserve its essential features but have a similar scale of values.

After autoscaling, a new single spectrum was created by cascading four parameters spectra one after another. This procedure was applied to both calibration and validation samples to create new spectra. After the new spectra had been created, PLS coupled with contiguous blocks cross validation (calibration set separated by five groups) was then applied to develop calibration models with optimal wavelengths for firmness and SSC. The models were validated with a separate set of validation samples based on the same wavelengths.

## **Results and Discussion**

The 600 'Golden Delicious' apples ranged between 60 mm and 83 mm in equatorial diameter, between 31.2 N and 88.5 N for MT firmness, and between 8.0% and 15.7% for SSC. The mean values of MT firmness and SSC for the samples were 56.7 N and 10.7%, respectively, whereas the standard deviations were 12.1 N for firmness and 1.4% for SSC.

Spectra of the four Lorentzian parameters for five ‘Golden Delicious’ apples are shown in figure 5. The patterns of the four parameters spectra were distinctively different since each had a different physical interpretation. Parameter *a* represented asymptotic values ranging between -200 and 200 among the 600 samples. In principle, the asymptotic value should be positive. However, we permitted negative values for this parameter during the curve-fitting process in view of the empirical nature of the function. Parameter *b*, representing the peak value, changed greatly over the wavelengths of 500 - 1,000 nm. Compared with parameters *a* and *b*, parameters *c* and *d* were relatively consistent over the spectral region and had much lower values. A downward or upward peak was observed for parameters *a*, *b*, *c*, and *d* around 675 nm, which corresponded to an absorption waveband for chlorophylls.

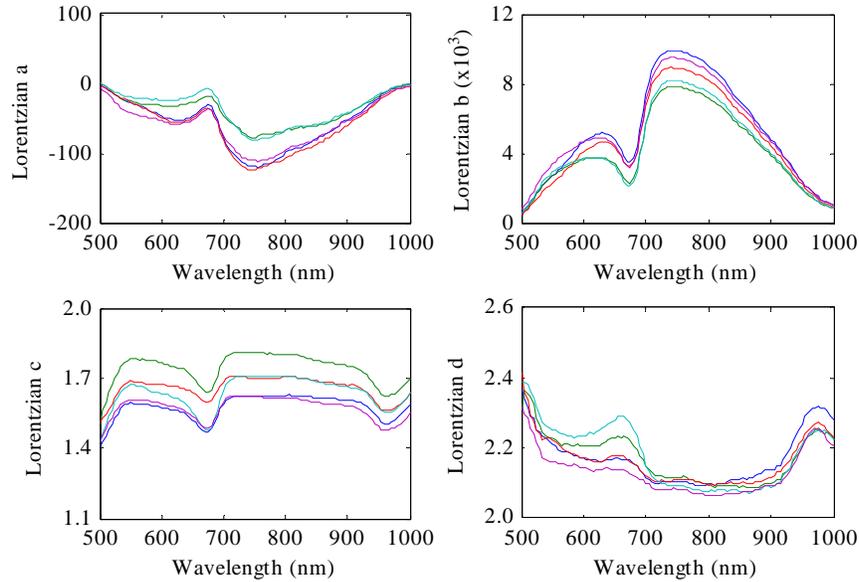


Figure 5. Spectra of Lorentzian parameters for selected ‘Golden Delicious’ apples.

Table 1 shows values of the correlation matrix *F* of firmness for a few selected wavelengths, calculated by equation (2) from the calibration set. After subspace decomposition, the full

Table 1. Correlation Matrix *F* for selected wavelengths.

wavelength	500 nm	520 nm	540 nm	560 nm	580 nm	600 nm	620 nm	640 nm
500 nm	1							
520 nm	0.967	1						
540 nm	0.940	0.991	1					
560 nm	0.926	0.980	0.997	1				
580 nm	0.903	0.955	0.981	0.992	1			
600 nm	0.880	0.936	0.966	0.980	0.995	1		
620 nm	0.857	0.921	0.947	0.965	0.987	0.997	1	
640 nm	0.840	0.898	0.919	0.940	0.969	0.985	0.995	1

spectrum space was decomposed into five subspaces for firmness. The range of each subspace was 500 – 550 nm, 555 – 640 nm, 645 – 695 nm, 700 – 855 nm and 860 -1,000 nm, and the corresponding number of wavelengths in each subspace was 11,18, 11, 32 and 29, respectively. For SSC, the full spectrum space was decomposed into five spaces. The wavelength range of each subspace was 500 - 570 nm, 575 – 650 nm, 655 - 695nm, 700 – 885 nm and 890 – 1,000 nm, and these subspaces had 15, 16, 9, 38 and 23 wavelengths, respectively.

Table 2 summarizes the results of five HEA-PLS models and the PLS model using full spectrum for predicting firmness and SSC of apples. Values for RMSEC (root mean square error of calibration) of the HEA-PLS models except the second and third HEA-PLS models for SSC prediction were generally higher than that for the full-spectrum model, while RMSECV and RMSEP (root mean square error of prediction) of the HEA-PLS models were lower than those of the full-spectrum model. These results indicated that although HEA did not result in better PLS models for calibration samples, their prediction ability was improved, as shown in Table 2. In terms of prediction ability for the cross validation set and validation set, the second HEA-PLS model with 17 wavelengths and 15 factors presented the best result for firmness ( $r_{cv} = 0.868$  and RMSECV = 6.0 N,  $r_p = 0.857$  and RMSEP = 6.2 N.). The third HEA-PLS model using 16 wavelengths and 21 factors had the best result for SSC ( $r_{cv} = 0.817$  and RMSECV = 0.80%,  $r_p = 0.822$  and RMSEP = 0.78%). Compared with the full-spectrum PLS model for firmness, RMSECV and RMSEP values for the best model were reduced by 3.2% and 3.1% respectively,

Table 2. Calibration and prediction results for firmness and soluble solids content (SSC) calculated from five HEA-PLS models and the full-spectrum PLS model.

Output	Model Type	$N_w^*$	$N_f^*$	$r_c^{**}$	RMSEC	$r_{cv}^{**}$	RMSECV	$r_p^{**}$	RMSEP
Firmness	PLS	101	14	0.890	5.5	0.858	6.2	0.848	6.4
	1	18	15	0.888	5.6	0.867	6.0	0.853	6.3
	2	17	15	0.887	5.6	0.868	6.0	0.857	6.2
	3	27	15	0.889	5.5	0.867	6.0	0.856	6.2
	4	21	14	0.887	5.6	0.868	6.0	0.853	6.3
	5	19	15	0.888	5.6	0.867	6.0	0.855	6.2
	Ave.			0.888	5.6	0.867	6.0	0.855	6.3
SSC	PLS	101	19	0.855	0.72	0.758	0.92	0.802	0.83
	1	13	20	0.849	0.73	0.810	0.81	0.810	0.81
	2	18	21	0.859	0.71	0.813	0.81	0.812	0.80
	3	16	21	0.862	0.70	0.817	0.80	0.822	0.78
	4	11	21	0.850	0.73	0.815	0.81	0.818	0.80
	5	12	21	0.851	0.73	0.818	0.80	0.810	0.81
	Ave.			0.854	0.72	0.815	0.81	0.815	0.80

\*  $N_w$  = number of wavelengths;  $N_f$  = number of factors.

\*\*  $r_c$  = correlation coefficient for calibration;  $r_{cv}$  = correlation coefficient for cross validation;  $r_p$  = correlation coefficient for validation.

while the number of wavelengths used in the model reduced to 17 from 101. The RMSECV and RMSEP values of the best HEA-PLS model for SSC decreased by 13.0% and 6.0% and the number of wavelengths used in the model reduced to 16 from 101. The best HEA-PLS model for SSC also yielded better result for calibration samples than the PLS model; the RMSEC value was reduced by 2.8%. Compared with the PLS model, the average values of RMSECV and RMSEP for the five HEA-PLS models were reduced by 3.2% and 1.6% for firmness and by 12.0% and 3.6% for SSC (Table 2). The improvement of SSC prediction ability was better than that for firmness.

The 17 optimal wavelengths for firmness prediction and 16 optimal wavelengths for SSC prediction are shown in Table 3. Wavelengths for firmness prediction nearly spanned the entire spectral range of 500 -1,000 nm, while none of the wavelengths below 600 nm was selected in the SSC prediction model. The optimal wavelengths selected in this study are overall in

agreement with the reported works using NIR spectroscopy that the spectral range between 600 nm and 1,000 nm is more appropriate for SSC prediction (McGlone et al., 2002).

Table 3. Optimal wavelengths selected for prediction models for fruit firmness and SSC.

Output	Number of wavelengths	Optimal wavelengths (nm)
Firmness	17	505, 525, 540, 580, 585, 640, 645, 655, 670, 700, 720, 735, 800, 875, 915, 985, 990
SSC	16	600, 620, 645, 760, 790, 820, 825, 830, 835, 845, 850, 855, 900, 935, 960, 990

Figures 6 and 7 show firmness and SSC predictions for the validation set of 200 apples using PLS models with full spectrum and the PLS models with 17 and 16 optimal wavelengths. The PLS models with the optimal wavelengths improved the correlation between the predicted values and actual values of the samples.

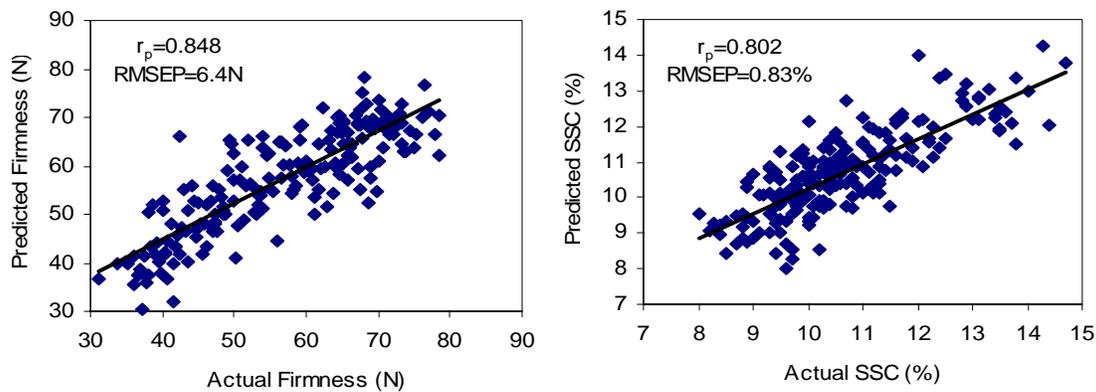


Figure 6. Prediction of the firmness (left) and soluble solids content (SSC, right) of 'Golden Delicious' apples by the PLS model with full spectrum.

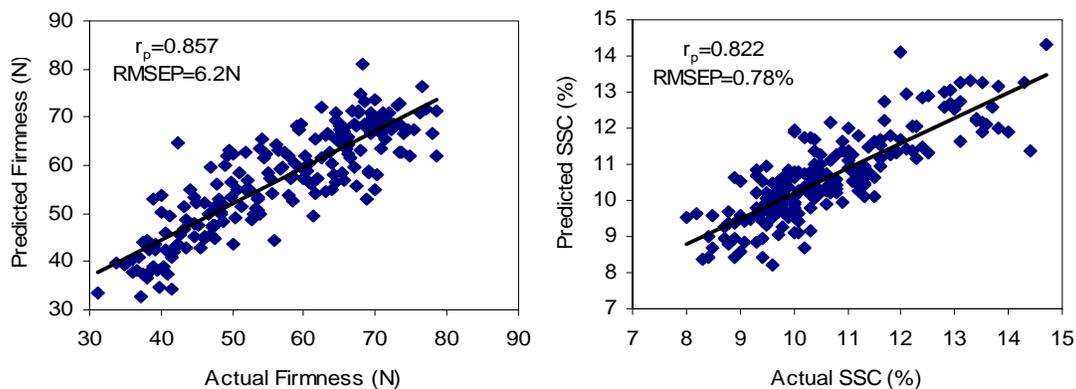


Figure 7. Prediction of the firmness (left) and soluble solids content (SSC, right) of 'Golden Delicious' apples by the PLS model with 17 and 16 optimal wavelengths respectively.

## Conclusion

Hierarchical evolutionary algorithm (HEA) was suitable for selecting optimal wavelengths from the hyperspectral scattering data and subspace decomposition reduced the code complexity of hierarchical chromosome. Seventeen and 16 optimal wavelengths were selected to achieve the

best prediction of apple firmness and SSC, respectively. The models using optimal wavelengths gave better predictions of fruit firmness with  $r_p = 0.857$  and RMSEP = 6.2 N than the full spectrum model ( $r_p = 0.848$ , RMSEP = 6.4 N), and of SSC with  $r_p = 0.822$  and RMSEP = 0.78% than the full spectrum model ( $r_p = 0.802$ , RMSEP = 0.83%). The HEA approach provided an effective means for selecting optimal wavelengths and improved prediction of firmness and SSC in apples compared with the approach of using full spectrum.

### **Acknowledgements**

The authors would like to thank Dr. Diwan Ariana, Ms. Haiyan Cen, Ms. Chenglu Wen and Mr. Benjamin Bailey for their technical support and advice for this research. Dr. Min Huang gratefully acknowledges financial support from the National Natural Science Foundation of China (Grant No. 60805014).

### **References**

- Goldberg, D. E. 1989. Sizing populations for serial and parallel genetic algorithms. *In Proc. 3rd international Conference on Genetic Algorithms*, p.70-79, San Mateo, Calif: Morgan Kaufman.
- He, Y., F. Chu, and B. Zhong. 2009. Hierarchical evolutionary algorithm for constructing and training wavelet networks. *Neural Comput. & Applic.* 10(4): 357-366.
- Kleynen, O., V. Leemans, and M. -F. Destain. 2003. Selection of the most efficient wavelength bands for 'Jonagold' apple sorting. *Postharv. Bio. Techn.* 30(3): 221-232.
- Lai, C. and C. Chang. 2009. A hierarchical evolutionary algorithm for automatic medical image segmentation. *Expert Systems with Applica.* 36(1): 248-259.
- Lammertyn, J., B. Nicolai, K. Ooms, V. De Smedt, and J. De Baerdemaeker. 1998. Non-destructive measurement of acidity, soluble solids, and firmness of Jonagold apples using NIR-spectroscopy. *Trans. ASAE* 41(4): 1089-1094.
- Leardi, R. 2000. Application of genetic algorithm-PLS for feature selection in spectral data sets. *J. chemometrics* 14(6): 643-655.
- Leardi, R., and A. Lupiáñez González. 1998. Genetic algorithms applied to feature selection in PLS regression: how and when to use them. *Chemometr. Intell. Lab. Syst.* 41(2): 195-207.
- Lu, R. 2001. Predicting firmness and sugar content of sweet cherries using near-infrared diffuse reflectance spectroscopy. *Trans. ASAE* 44(5): 1265-1271.
- Lu, R. 2003. Imaging spectroscopy for assessing internal quality of apple fruit. *ASAE Paper No. 036012*. St. Joseph, Mich.: ASABE.
- Lu, R. 2004. Multispectral imaging for predicting firmness and soluble solids content of apple fruit. *Postharv. Bio. Techn.* 31(2): 147-157.
- Lu, R. 2007. Nondestructive measurement of firmness and soluble solids content for apple fruit using hyperspectral scattering images. *Sens. & Instrumen. Food Qual.* 1(1): 19-27.
- Lu, R. and D. P. Ariana. 2002. A near-infrared sensing technique for measuring internal quality of apple fruit. *Appl. Eng. and Agric* 18(5): 585-590.
- Lu, R., D. E. Guyer, and R. M. Beaudry. 2000. Determination of firmness and sugar content of apples using near-infrared diffuse reflectance. *J. Texture Studies* 31(6): 615-630.
- Lu, R. and Y. Peng. 2006. Hyperspectral scattering for assessing peach fruit firmness. *Biosystems Eng.* 93(2): 161-171.

- McGlone, V. A., R. B. Jordan, and P.J. Martinsen. 2002. Vis/NIR estimation at harvest of pre- and post- storage quality indices for 'Royal Gala' apple. *Postharv. Bio. Techn.* 25(2): 135-144.
- McGlone, V. A. and S. Kawano. 1998. Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. *Postharv. Bio. Techn.* 13(2): 131-141.
- Park, B., J. A. Abbott, L. Lee, C. Choi, and K. Choi. 2003. Near-infrared diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of Delicious and Gala apples. *Trans. ASAE* 46(6): 1721-1731.
- Peng, Y. and R. Lu. 2006. An LCTF-based multispectral imaging system for estimation of apple fruit firmness: Part I. Acquisition and characterization of scattering images. *Trans. ASABE* 49(1): 259-267.
- Peng, Y. and R. Lu. 2007. Prediction of apple fruit firmness and soluble solids content using characteristics of multispectral scattering images. *J. Food Engineering* 82(2): 142-152.
- Peng, Y. and R. Lu. 2008. Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit firmness and soluble solids content. *Postharv. Bio. Techn.* 48(1): 52-62.
- Qin, J. and R. Lu. 2008. Measurement of the optical properties of fruits and vegetables using spatially resolved hyperspectral diffuse reflectance imaging technique. *Postharv. Bio. Techn.* 49(3): 355-365.
- Steward, B. L., A. L. Kaleita, R. P. Ewing, and D. A. Ashlock. 2005. Genetic algorithms for hyperspectral range and operator selection. *ASAE Paper No. 053063*. St. Joseph, Mich.: ASABE.
- Tang, K., K. Man, S. Kwong, and Z. Liu. 1998. Design and optimization of IIR filter structure using hierarchical genetic algorithms. *IEEE Transactions on industrial electronics* 45(3): 481-487.
- Zou, X., J. Zhao, X. Huang, and Y. Li. 2007. Use of FT-PLS spectrometry in non-invasive measurements of soluble solid contents (SSC) of 'Fuji' apple based on different PLS models. *Chemometr. Intell. Lab Syst.* 87(1): 43-51.