JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

J. Agric. Food Chem. **2007,** *55,* 5423−5428 **5423**

Study of Wavelet Denoising in Apple's Charge-Coupled Device Near-Infrared Spectroscopy

DAZHOU ZHU, BAOPING JI,* CHAOYING MENG, BOLIN SHI, ZHENHUA TU, AND ZHAOSHEN QING

College of Food Science and Nutritional Engineering, China Agricultural University, Tsinghua East Road 17, Haidian District, Beijing 100083, People's Republic of China

Discrete wavelet transform was used to eliminate the noise in the charge-coupled device near-infrared (CCD-NIR) spectra of apple. The influence of three parameters (wavelet function, decomposition level, and threshold) on the predictive ability of the calibration model was investigated. The result showed that the db, sym, and bior wavelet families performed well, while the coif, dmey, and haar wavelets were not able to denoise effectively. The best decomposition level was 2. The threshold selection rules of the default, Birge−Massart, and Penalty had good denoising results, while SURE, Sqtwolog, Heuristic SURE, and Minimax set all detailed coefficients to zero due to their high threshold values. The best denoising result was obtained with the combination of the bior3.3 wavelet function, two levels of decomposition, default threshold selection rule, and the soft thresholding method. The optimal model of soluble solids content was constructed. The relative standard deviation of prediction decreased from 7.79 to 5.82% after wavelet denoising.

KEYWORDS: Apple; CCD-NIR spectroscopy; wavelet transform; denoising

INTRODUCTION

The nondestructive detection of the inner quality of fruit is of great importance for grading and quality evaluation. The near-infrared (NIR) reflectance spectroscopy technique has a short measuring time with limited sample preparation and is widely used in agricultural product detection and analysis (*1*, *2*). Many parameters of apple can be measured with NIR spectroscopy, such as soluble solids content (SSC), total sugar content, titratable acidity, firmness, and the contents of glucose and fructose (*3*−*6*). However, the intensity of NIR reflectance spectroscopy is usually 1−2 orders of magnitude lower than that of midinfrared (MNR), resulting in its larger background than that of MNR. Therefore, the improvement of signal-to-noise ratio (SNR) is crucial for NIR analysis with high analytical precision. To improve the robustness of the calibration model, different preprocessing methods have been used to eliminate the background and noise (*7*). The charge-coupled device near-infrared spectroscopy (CCD-NIR) spectrometer has a rapid scan speed, so it could be used for on-line detection and real-time grading of fruit. As compared with the spectrum collected from a Fourier transform near-infrared (FT-NIR) spectrometer, the spectrum from CCD spectrometer has higher noise. Therefore, denoising seems to be very important when CCD-NIR spectra are used to construct a calibration model.

Since it was developed in 1980s, the wavelet transform (WT) (*8*) has been used as a time frequency domain analytical method in many fields due to its noise elimination. Stark et al. (*9*) first used the WT for IR spectroscopy. With the aid of WT, they separated the mineralogical information in the Fourier transform infrared (FT-IR) absorbance spectrum from noise and other signals such as absorbance from water and organics. Ming et al. (*10*) reported that the relative root-mean-square deviation of the prediction set decreased from 9.2 to 7.4% when WT was used to filter the NIR spectra of tobacco. Chen et al. (*11*) and Fu et al. (*12*) found that wavelet denoising could improve the NIR predicting precision of the oil content in instant noodles and the sugar content in vinegar. Depczynski's group (*13*) found that the boundary effect occurred when the WT was used to analyze finitely supported signals and suggested mitigating it by applying Strum−Liouville wavelets. Zhu et al. (*14*) used a new WT denoising method to eliminate the noise in the NIR spectrum of wheat based on the difference in wavelet modulus of the maximum evolution behavior between a singular signal and random noise in a multiscale space. Shao and Zhuang (*15*) applied the continuous wavelet transform (CWT) to eliminate the background of NIR spectra. They concluded that WT achieved higher precision than S−G smoothing and derivatives, and the discrete wavelet transform (DWT) was slightly better than CWT. All of these results suggest that WT efficiently eliminates the background noise from NIR spectroscopy.

Factors of the wavelet function, decomposition level, and threshold strongly influenced the denoising result. Cao et al. (*16*) discussed the influence of these three factors when DWT denoising was used to process capillary electrophoresis−eletrochemiluminescence signals. However, the noise charac-

* To whom correspondence should be addressed. Tel: +86-10-62737129. Fax: +86-10-62347334. E-mail: jbp@cau.edu.cn.

10.1021/jf063363c CCC: $37.00
Published on Web 06/19/2007

teristics vary with signals. An optimal wavelet function for a given signal is not necessarily the best for others (*17*). Therefore, it is very important to study the influence in general of these three wavelet denoising parameters on the robustness of the calibration model when wavelet denoising is used to preprocess the CCD-NIR spectra. Ying et al. (*18*, *19*) used wavelet denoising for detecting the sugar content of apple by the FT-NIR spectroscopy and discussed the performance of the db2 and db4 wavelets. Liu et al. (*20*) used DWT to eliminate the white noise that they added to the FT-NIR spectra and applied SNR to estimate the denoising effects of the different wavelet functions and decomposition levels. However, the thresholding method was not discussed in their study. Furthermore, for a real spectrum, the measurement condition differs substantially and it is difficult to calculate the value of SNR. Thus, the performance of calibration model based on using the denoised spectra is required to estimate the denoising effect.

In this study, the DWT denoising method based on thresholding was used to eliminate the noise in the CCD-NIR spectra of apple. Wavelet functions from six wavelet families that have different orders were used to investigate the influence of wavelet functions on the denoising result. The influences of decomposition level and threshold selection rules were discussed. The optimal combination of denoising parameters was obtained, and the best calibration model for SSC of apple was constructed.

MATERIALS AND METHODS

Samples. Thirty Fuji apples were purchased from the local fruit market. Apples with different colors and sizes were chosen to keep the SSC range as wide as possible. The apples came from orchards in the Shanxi province of China. Apples were stored at 20 °C for 2 days for later analysis.

Apparatus. The NIR reflectance spectra of apples were measured with a CCD-NIR spectrometer (AvaSpec-2048, Arantes, The Netherlands), which was equipped with a 2048 pixel CCD detector array. The spectrometer had a spectral range of 580−1100 nm. The integration time of CCD was set to 8 ms. For each measurement, it was scanned 64 times and the averaged spectrum was the output of the spectrometer. A $BaSO_4$ cylinder 30 mm in diameter and 5 mm in thickness was used as a reference.

Spectra Measurement. All of the reflectance spectra of apples were measured in the lab at room temperature (~20 °C). The Y-bundle fiber of the CCD spectrometer was used to measure the apples. The head of the fiber was put closely against the apple. The light was guided to the sample by the source fiber. The diffusely reflected light from the sample was detected by the detector fiber. Each apple was measured on four evenly distributed equatorial positions that were marked with a circle, avoiding the obvious disfigurement. In total, 120 spectra were obtained (30 apples times four positions per apple). Because the responses of the CCD detector in the ranges of 580−730 and 1060−1100 nm were very low, only the spectral range of 730−1060 nm was selected to construct the calibration model. Each spectrum had 1324 points.

SSC Measurement. According to the standard for fresh apple measurements GB10651-89 (*21*), a refractometer (WAY, Shanghai Precision & Scientific Instrument Co., Ltd., China) was used to measure the SSC of apple. The juice was squeezed from a core removed from the NIR sampling site, and the °Brix was recorded. Thus, 120 SSC values were obtained (30 apples times four positions per apple).

Outlier Detection. Outliers are data that have a rather large influence on the regression solution, and the occurrence of such data points can lead to considerable deviations from normality (*22*). Therefore, the outliers must be eliminated. The principal components analysis (PCA) score plot was applied to eliminate the outlier. **Figure 1** showed the score plot of the first two PCs (PC1 and PC2). Two samples marked with circles in **Figure 1** were identified as outliers and were eliminated. Similarly, other outliers were eliminated based on the score plots of PC1−PC3, PC2−PC3, and PC3−PC4. In total, seven outliers were eliminated. The numbers of the eliminated samples were 30, 37, 52, 56, 69, 104, and 113, respectively. All of these outliers were distributed in seven different apples. Therefore, the outliers might occur because of variations in the spectral measurement.

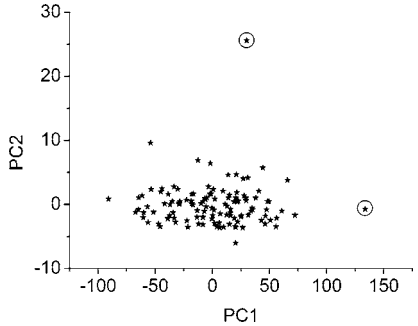

Figure 1. Score plot of the first two PCs of the NIR spectra of the apples.

Table 1. Statistic Values of Calibration and Validation Data Sets of Apple[a]

sample set	n[b]	range	mean	SD[c]
calibration	80	8.9–17.8	13.32	1.73
validation	33	10.9–14.9	12.42	1.09

[a] Unit used, °Brix. [b] n = number of samples. [c] SD, standard deviation.

The 113 retained samples were randomly divided into two data sets: The calibration set contained 80 samples, and the validation set contained 33 samples. The statistical characteristics of the SSC for two data sets are summarized in **Table 1**.

WT Theory and Algorithm. Wavelets are a series of functions that derived from the basis function (mother wavelet). A wavelet is defined as:

$$\psi_{a,b} = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right) \tag{1}$$

where a is a scaling variable and b is a translation variable. In practice, it is assumed that $S = 2^{*}m$, $b = n^{*}b_0$, (m, n $\in Z$), and the function $f(t)$ is a signal; then, the following discrete wavelet is obtained

$$\psi_{m,n(t)} = 2^{-m/2}\psi\left(\frac{t-n^{*}b_0}{2^m}\right) \tag{2}$$

The DWT can be defined as:

$$C_{m,n}(f) = \int_{-\infty}^{+\infty} \psi_{m,n}(t)\, f(t)\, \mathrm{d}_t \tag{3}$$

Mallat (*23*) proposed an efficient algorithm to perform DWT by assuming that the discrete signal $f(t)$ is $\{C(n)\}$, where n is the signal number. The Mallat algorithm is as follows.

$$C^j(n) = \sum_{k\in Z} h_{k-2n} C^{j-1}(k) \tag{4}$$

$$D^j(n) = \sum_{k\in Z} h_{k-2n} D^{j-1}(k) \tag{5}$$

where $\{C^j(n)\}$ is the approximate coefficient (low-frequency components) of the signal at the decomposition level of j and $\{D^j(n)\}$ is the detailed coefficient (high-frequency components). The original signal C^0 can be reconstructed by C^j and D^1, D^2, ..., D^j. The reconstruction formula is as follows:

$$C^{j-1}(n) = \sum_{k\in Z} h(k-2n)C^j(k) + \sum_{k\in Z} g(k-2n)D^j(k) \tag{6}$$

With the increase of decomposition level j, the more detailed characteristics of the signal can be observed.

Wavelet Denoising Steps. The characteristics of the signal and noise are considerably different under wavelet decomposition. The power spectrum of the signal is concentrated, and the absolute value of wavelet coefficient of the signal is high. However, the power spectrum of noise is dispersive comparatively and the absolute value of wavelet coefficient of noise is low. Consequently, the noise in the signal can be suppressed or eliminated by using threshold to filter the wavelet coefficient whose absolute value is below the threshold, which is the principle of wavelet denoising (*24*). The wavelet denoising on the CCD-NIR spectra is performed in three steps.

1. Wavelet Decomposition. A wavelet function and a decomposition level were selected, and DWT was applied to each spectrum. Then, the approximate coefficients and the detailed coefficients were obtained. It is noticeable from the definition formula of WT that the result of WT is closely related to the wavelet function used. Many wavelet families have been constructed. Some wavelet families contain wavelets that have different orders. Db, coif, sym, bior, and dmey wavelets have been proven to be effective in signal processing (*25*). In addition, the haar wavelet is the simplest wavelet. Thus, the denoising effects on the CCD-NIR spectroscopy of these six families of wavelets were investigated. For a signal with a length of N, the theoretically maximum decomposition level J is defined as:

$$J = \log_2^N \tag{7}$$

J was calculated by the Matlab command of "wmaxlev". An investigation was conducted in this study on the denoising effect of different decomposition levels (1, 2, ..., J).

2. Thresholding. An appropriate threshold value was given beforehand, and then, the noise component in the signal was eliminated by processing the detailed coefficients with the soft or hard thresholding strategy. The definitions of soft and hard thresholding are as follows:

hard thresholding:

$$x^* = \begin{Bmatrix} 0 & |x| \le T \\ x & x > T \end{Bmatrix} \tag{8}$$

soft thresholding:

$$x^* = \begin{Bmatrix} 0 & |x| \le T \\ \text{sign}(x)(|x| - T) & x > T \end{Bmatrix} \tag{9}$$

where T is the threshold value and x and x^* are the wavelet coefficients before and after thresholding. Normally, the signal processed by soft thresholding is smoother than that by hard thresholding. Thus, soft thresholding was selected to process the CCD-NIR spectrum.

Threshold value is a key parameter in wavelet denoising and can be determined with three kinds of approaches.

a. Estimation of Threshold Based on Original Signal. The threshold is estimated based on the SNR of original signal. There are three strategies to estimate this kind of threshold: the default, Birge−Massart, and Penalty. The default threshold is calculated as:

$$\text{thr} = \sqrt{2\log(N)^*\sigma} \tag{10}$$

where N is the signal length and σ is the standard deviation of the noise. In the Matlab Toolbox, the three thresholds are calculated by the commands of "ddencmp", "wdcbm", and "wbmpen", respectively.

b. Estimation of Threshold Based on Sample Estimator. Threshold is generated by the unbiased risk estimate according to the criterion that the square deviation between the signal denoised in the worst case and the original signal is as low as possible. It has four strategies to calculate this kind of threshold, which are SURE, Sqtwolog, Heuristic SURE, and Minimax threshold. SURE is a self-adaptive threshold selection method based on the Stein's unbiased estimate. The Sqtwolog threshold can be calculated as:

$$T = \sqrt{2\log_e[N\log_2(N)]} \tag{11}$$

Heuristic SURE is a mixture of SURE and Sqtwolog. Minimax threshold is obtained by the construction method of estimator in statistics to achieve the minimum of the maximum mean square error obtained for the worst function in a given set. These four thresholds are all calculated by the command "thselect" in Matlab, and the parameters of "thselect" are "rigrsure", "sqtwolog", "heursure" and "minimaxi", respectively.

c. Estimation of Threshold Based on Experience. After a large amount of study on a specific kind of signal and a comprehensive analysis of its characteristics, the threshold can be determined by experience. This method is very practical in engineering practice. In this study, thresholds are assumed to be very large, and all detailed coefficients are set to zero in reconstruction. In this study, the influence of these eight threshold selection methods on the denoising result was evaluated.

3. Spectrum Reconstruction. The new spectrum was reconstructed from wavelet coefficients by the WT reconstruction algorithm.

Construction and Validation of the SSC Model. After WT denoising, the partial least-square regression (PLSR) (*26*) method was used to construct the SSC model. The optimal number of factors for PLSR was determined by the four-block cross-validation method. The predictive ability of the calibration model was evaluated with parameters of the relative standard deviation of calibration (RSD_c), the relative standard deviation of prediction (RSD_p), and the correlation coefficient (R) between the measured and the predicted values for the calibration set. These parameters are defined as:

$$R = \frac{\sum_{i=1}^{m} (y_c - \bar{y}_c)(\hat{y}_c - y_m)}{\sqrt{\sum_{i=1}^{m} (y_c - \bar{y}_c)^2}\sqrt{\sum_{i=1}^{m} (\hat{y}_c - y_m)^2}} \tag{12}$$

$$\text{RSD}_c = \frac{1}{\bar{y}_c}\sqrt{\frac{1}{m-p-1}\sum_{i=1}^{m} (y_{ci} - \hat{y}_{ci})^2} \times 100\% \tag{13}$$

$$\text{RSD}_p = \frac{1}{\bar{y}_p}\sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_{pi} - \hat{y}_{pi})^2} \times 100\% \tag{14}$$

where y_c is the real SSC of calibration set; $\bar{y}_c$ is the mean of y_c; $\hat{y}_c$ is the predicted value; y_m is the mean of $\hat{y}_c$; m is the number of calibration sets; P is the number of PLS factors; y_p is the real SSC of validation set; $\hat{y}_p$ is the predicted value; $\bar{y}_p$ is the mean of y_p; and n is the number of validation set.

In this study, the complete combinations of wavelet denoising parameters were performed to investigate the influence of three main parameters. These combinations included six families of wavelets (haar, db, sym, bior, coif, and dmey) with different orders, different composition levels, and eight threshold selection rules (default, Birge−Massart, Penalty, SURE, Sqtwolog, Heuristic SURE, Minimax, and the method of setting all detailed coefficients to zero). In total, 2056 models were constructed. The optimal calibration model of SSC of apple was explored among these 2056 models.

Software. The Wavelet Toolbox 3.0 (The Math Works, Inc., Natick, MA) was used. PCA and PLS were written in Matlab language. All of the calculations were carried out on the platform of Matlab 7.0.

RESULTS AND DISCUSSION

Characteristic of Noise in the CCD-NIR Spectrum. Different signals have different kinds of noises. It is necessary to determine the noise characteristic before denoising. The sym3 wavelet was employed to decompose the 15th spectrum at a level of four. **Figure 2** shows the original spectrum, the

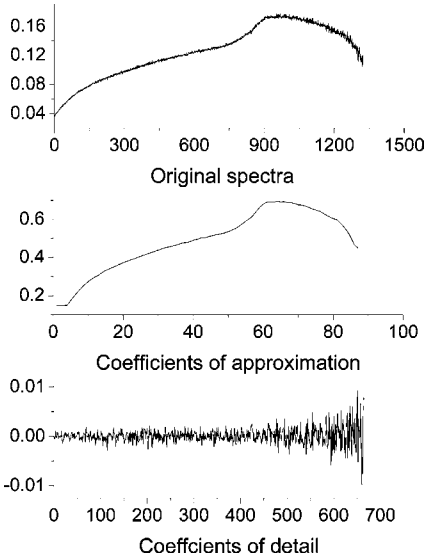

Figure 2. Noise characteristic of the CCD-NIR spectrum. Upper, original spectrum of apple; middle, coefficients of approximation; and bottom, coefficients of detail.

approximate coefficients of the fourth level, and the detailed coefficients of the first level. Generally, the signal, noise, and slow background are distributed in different frequency bands, namely, noise has a high frequency and is contained in detailed coefficients of the first level while the slow background is contained in approximate coefficients of high level. As shown in **Figure 2**, the shape of the fourth approximate coefficients and the original spectra were almost the same; the mean value of noise was nearly zero, and the intensity of noise increased with the increase of wavelength.

Influence of Wavelet Functions. Currently, there is no standard method to select a wavelet function in WT. Some criteria have been proposed to select a wavelet. One of them was that the wavelet and signal should have good similarities. The influence of the different wavelet functions on denoising effects must be conducted under the same condition. As discussed, the influence is evaluated with parameters of R, RSD_c, and RSD_p of the SSC calibration model. When a model has a high value of R and low RSD_c and RSD_p, the calibration model is robust and accurate and the denoising result is significant. **Table 2** showed the denoising results performed by different wavelet functions in a decomposition level of two and a default threshold. The calibration model of the original spectra was also shown in the table. Wavelets of the coif and dmey had poor predictive results, since the values of RSD_p were all higher than 7.79% of the original spectra. The RSD_p of the haar wavelet was almost the same as the original spectra, and its R value was even lower. Therefore, the coif, dmey, and haar wavelet families cannot effectively eliminate the noise of the CCD-NIR spectrum. For the wavelet families of the db, sym, and bior, wavelet functions with some orders had good denoising results. For the bior3.x wavelet, RSD_p values were relatively lower than other wavelets except for the bior3.1 wavelet. Especially, the bior3.3 wavelet gave the lowest RSD_p for the model and improved the predictive ability of model greatly. The db2 and sym2 and db3 and sym3 had the same results, but the db and sym wavelet families with other orders had different results; this may come from the similarity of the db and sym wavelets. Except for the dmey and haar that do not have order parameters, the calibration results for the rest four wavelet families do not show a clear tendency in varying with orders.

Table 2. Influence of Wavelet Function on CCD-NIR Spectra

wavelets	factors	R	RSD_c (%)	RSD_p (%)
db2	7	0.9546	4.06	6.45
db3	7	0.9501	4.25	7.05
db4	7	0.9546	4.06	7.87
db5	7	0.9590	3.86	6.89
db6	7	0.9531	4.12	6.89
db7	7	0.9513	4.20	8.22
db8	7	0.9547	4.05	7.64
db9	7	0.9520	4.17	6.71
db10	7	0.9003	5.89	6.94
coif1	7	0.9538	4.09	8.73
coif2	7	0.9525	4.15	8.51
coif3	7	0.9527	4.14	8.44
coif4	7	0.9528	4.13	8.37
coif5	7	0.9529	4.13	8.32
dmey	7	0.9529	4.13	8.08
haar	7	0.9577	3.92	7.74
sym2	7	0.9546	4.06	6.45
sym3	7	0.9501	4.25	7.05
sym4	6	0.9004	5.89	7.32
sym5	7	0.9524	4.17	8.06
sym6	7	0.9529	4.13	6.97
sym7	7	0.9517	4.18	6.82
sym8	7	0.9523	4.16	8.15
Bior1.1	7	0.9577	3.92	7.74
Bior1.3	6	0.8957	6.02	6.53
Bior1.5	7	0.9552	4.03	7.74
Bior2.2	7	0.9570	3.95	9.06
Bior2.4	7	0.9499	4.26	6.87
Bior2.6	7	0.9517	4.18	8.35
Bior2.8	7	0.9520	4.17	6.86
Bior3.1	6	0.9349	4.81	9.30
Bior3.3	6	0.9063	5.63	5.82
Bior3.5	6	0.9107	5.59	7.49
Bior3.7	6	0.9061	5.72	6.09
Bior3.9	6	0.9074	5.68	7.15
Bior4.4	7	0.9507	4.22	6.97
Bior5.5	7	0.9607	3.78	7.11
Bior6.8	7	0.9516	4.18	6.98
original spectra	7	0.9805	2.66	7.79

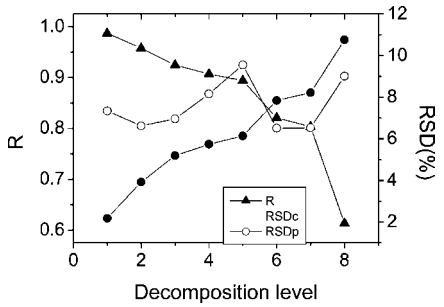

Figure 3. Influence of the decomposition level on denoising.

In addition, the factors of the PLS model did not vary with the wavelet used. After wavelet denoising, the factors of most PLS models were the same as that of the original model, and only a few models had a little change (decreased from seven to six). The wavelet families of the db, sym, and the bior had good denoising results.

Influence of Decomposition Levels. The value of the largest decomposition level J was calculated by applying the command "wmaxlev" in Matlab. The J value varies with the wavelet functions. To investigate the influence of decomposition levels, the db2 wavelet function and Brige−Massart threshold were employed to perform the WT denoising. The results are shown in **Figure 3**. With the increase of decomposition levels, the R value decreased while the RSD_c value increased and the RSD_p

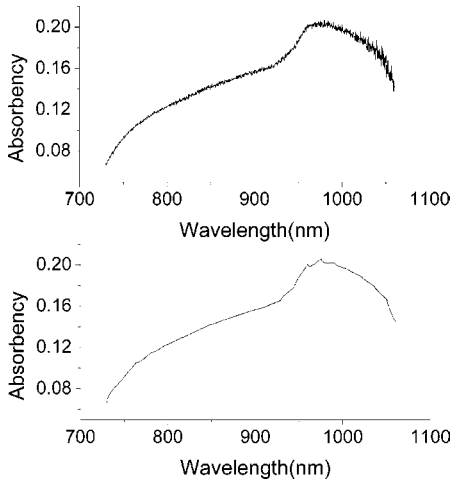

Figure 4. Spectrum before and after being denoised by db2 wavelet on a decomposition level of 5. Upper, original spectrum; bottom, spectrum after being denoised by WT.

Table 3. Effects of Different Wavelet Thresholding Methods on Denoising of CCD NIR Spectra

threshold selection roles	threshold value[a]	factors	*R*	RSD_c (%)	RSD_p (%)
default	0.0046	7	0.9546	4.06	6.45
Birge–Massart	0.0030, 0.0015	7	0.9576	3.92	6.65
Penalty	0.0036	7	0.9616	3.74	6.49
SURE	0.1361	7	0.9424	4.56	7.14
Sqtwolog	3.7917	7	0.9424	4.56	7.14
Heuristic SURE	3.7917	7	0.9424	4.56	7.14
Minimax	2.2904	7	0.9424	4.56	7.14
experienced threshold		7	0.9424	4.56	7.14

[a] The threshold value of the 25th sample that was selected by different methods.

value fluctuated. When the decomposition level reached five, RSD_c was higher than RSD_p. As shown in **Figure 4**, after the five level decomposition, the reconstructed signal was slightly distorted, suggesting that part of the useful signal was eliminated as noise with a higher decomposition level. **Figure 3** showed that the best decomposition level was two.

Influence of Threshold Selection Roles. The db2 wavelet with a decomposition level of two was employed to all of the spectra denoising procedures. Eight threshold selection roles (default, Birge−Massart, Penalty, SURE, Sqtwolog, Heuristic SURE, Minimax threshold, and the method of setting all detailed coefficients to zero) with soft thresholding methods were used. The results are shown in **Table 3**. The three original signal-based threshold selection roles (default, Birge−Massart, and Penalty) had low threshold values. Their calibration models had high *R* and low RSD_c and RSD_p, suggesting that their denoising results were good. The four sample estimator-based threshold selection roles (SURE, Sqtwolog, Heuristic SURE, and Minimax) had less effective results. Furthermore, the thresholds of all of these four methods were high, and they had the same denoising result with the method of setting detailed coefficients directly to zero. This indicated that the threshold values obtained by these four methods were so high that all detailed coefficients were changed to zero after being processed with the soft thresholding method. Therefore, they had the same result. In all of these eight threshold selection roles, the default and Penalty thresholds performed best.

Table 4. Calibration Model of SSC Obtained by Different Preprocessing Methods

processing method	*n*[a]	*R*	RSD_c (%)	RSD_p (%)
original spectra	7	0.9805	2.66	7.79
S–G smoothing	6	0.9442	4.45	6.52
DWT	6	0.9063	5.63	5.82

[a] The PLS factors.

Optimal Calibration Model of SSC. Among all the 2056 models, the combination of the bior3.3 wavelet, decomposition level of two, and default threshold selection role with the soft thresholding method gave the best denoising result. The optimal calibration model of SSC in apple is shown in **Table 4**. The results of the S−G smoothing and original spectra are also listed. The original model had low RSD_c and high RSD_p. Because the original spectra had high noise, the calibration model was overfitted and its predictive ability was bad. After WT denoising, the RSD_p decreased from 7.79 to 5.82%. Thus, the robustness and predictive ability of the SSC model were improved. As compared with the S−G smoothing, wavelet denoising gave lower RSD_p. For S−G smoothing, it is difficult to differentiate the useful signal and the noise component; thus, some useful information is often removed. WT has excellent ability for analysis of the small part of the signal, and it has been called a mathematical microscope (*27*). Therefore, we could conclude that DWT denoising was better than S−G smoothing for processing the CCD-NIR spectra.

LITERATURE CITED

(1) Weyer, L. G. Near infrared spectroscopy of organic substances. *Appl. Spectrosc. Rev.* **1985**, *21*, 1−43.
(2) Stark, E.; Luchter, K.; Margoshes, M. Near-infrared analysis (NIRA): A technology for quantitative and qualitative analysis. *Appl. Spectrosc. Rev.* **1986**, *22*, 335−339.
(3) Lammertyn, J.; Nicolai, B.; Ooms, K.; De Smedt, V.; De Baerdemaeker, J. Non-destructive measurement of acidity, soluble solids, and firmness of Jonagold apples using NIR-spectroscopy. *Trans. ASAE* **1998**, *41*, 1089−1094.
(4) Ventura, M.; Jager, A. D.; Putter, H.; Roelofs, F. P. M. Non-destructive determination of soluble solids in apple fruit by near infrared spectroscopy (NIRS). *Postharvest Biol. Technol.* **1998**, *14*, 21−27.
(5) Quilitzsch, R.; Hoberg, E. Fast determination of apple quality by spectroscopy in the near infrared. *J. Appl. Bot.−Angew. Bot.* **2003**, *77*, 172−176.
(6) Liu, Y.; Ying, Y.; Yu, H.; Fu, X. Comparison of the HPLC method and FT-NIR analysis for quantification of glucose, fructose, and sucrose in intact apple fruits. *J. Agric. Food Chem.* **2006**, *54*, 2810−2815.
(7) Delwiche, S. R.; Reeves, J. B., III. The effect of spectral pre-pretreatments on the partial least squares modeling of agricultural products. *J. Near Infrared Spectrosc.* **2004**, *12*, 177−182.
(8) Asker, A.; Cetin, A. E.; Rabitz, H. Wavelet transform for analysis of molecular dynamics. *J. Phys. Chem.* **1996**, *100*, 19165−19173.
(9) Stark, P. B.; Herron, M. M.; Matteson, A. Empirically minimax affine mineralogy estimates from fourier transform infrared spectrometry using a decimated wavelet basis. *Appl. Spectrosc.* **1993**, *47*, 1820−1829.
(10) Ming, S.; Xie, X.; Zhou, X.; Li, L.; Yan, Y. Noise filter for near infrared diffusive reflectance spectra by wavelet transform. *Chin. J. Anal. Chem.* **1998**, *26*, 34−37.
(11) Chen, B.; Fu, X.; Lu, D. Improvement of predicting precision of oil content in instant noodles by using wavelet transforms to process near-infrared spectroscopy. *J. Food Eng.* **2002**, *53*, 373−376.

(12) Fu, X.; Yan, G.; Chen, B.; Li, H. Application of wavelet transforms to improve prediction precision of near infrared spectra. *J. Food Eng.* **2002**, *69*, 461−466.
(13) Depczynski, U.; Jetter, K.; Molt, K.; Niemoller, A. The fast wavelet transform on compact intervals as a tool in chemometrics—II. Boundary effects, denoising and compression. *Chemom. Intell. Lab. Syst.* **1999**, *49*, 151−161.
(14) Zhu, S.; Wang, Y.; Zhang, X. Wavelet denoising theory and its application in wheat protein concentration with near infrared spectroscopy analysis. *J. Southwest Agric. Univ. (Nat. Sci.)* **2003**, *25*, 522−525.
(15) Shao, X.; Zhuang, Y. Determination of chlorogenic acid in plant samples by using near-infrared spectrum with wavelet transform preprocessing. *Anal. Sci.* **2004**, *20*, 451−454.
(16) Cao, W.; Chen, X.; Yang, X.; Wang, E. Discrete wavelets transform for signal denoising in capillary electrophoresis with electrochemiluminescence detection. *Electrophoresis* **2003**, *24*, 3124−3130.
(17) Pasti, L.; Walczak, B.; Massart, D. L.; Reschiglian, P. Optimization of signal denoising in discrete wavelet transform. *Chemom. Intell. Lab. Syst.* **1999**, *48*, 21−34.
(18) Ying, Y.; Liu, Y.; Fu, X. Sugar content prediction of apple using near-infrared spectroscopy processed by wavelet transform. *Spectrosc. Spectral Anal.* **2006**, *26*, 63−66.
(19) Ying, Y.; Liu, Y.; Fu, X.; Lu, H. Effect of wavelet transforms techniques upon the estimation of sugar content in apple with near-infrared spectroscopy. In *Proceedings of SPIE: Nondestructive Sensing for Food Safety, Quality, and Natural Resources*; Chen, Y.-R., Tu, S.-I., Eds.; SPIE: Philadelphia, Pennsylvania, 2004; Vol. 5587, pp 29−41.
(20) Liu, Y.; Ying, Y.; Lu, H.; Fu, X. Wavelet analysis techniques applied to removing varying spectroscopic background in calibration model for paper sugar content. In *Proceedings of SPIE: Optical Sensors and Sensing Systems for Nautral Resources and Food Safety and Quality*; Chen, Y.-R., Meyer, G. E., Tu, S.-I., Eds.; SPIE: Boston, Massachusetts, 2005; Vol. 5996, pp 413−424.
(21) The first editorial office of Standardization Administration of China. GB10651-89: Fresh apple. In *China Food Industrial Method—Fruit, Vegetable, and Their Product*, 2nd ed.; The Standardization Administration of China: Beijing, China, 2003; pp 145−155.
(22) Philips, G. R.; Eyring, E. M. Comparison of conventional and robust regression in analysis of chemical data. *Anal. Chem.* **1983**, *55*, 1134−1138.
(23) Mallat, S. G. A theory of multiresolution signal decomposition: The wavelet transform. *IEEE Trans. Pattern Anal. Machine Intell.* **1989**, *11*, 674−693.
(24) Changhong, D. Wavelet analyst in Matlab. In *The Principle and Application of Matlab Toolbox for Wavelet*; National Defense Industry Press: Beijing, China, 2005; pp 108−116.
(25) Mallat, S. G. Signularity detection and processing with wavelets. *IEEE Trans. Inform. Theory* **1992**, *38*, 617−643.
(26) Wang, H. Simple partial least-squares regression model. *The Method and Its Application of Partial Least-Squares Regression*; National Defense Industry Press: Beijing, China, 1999; pp 200−203.
(27) Daubechies, I. Orthonormal bases of compactly supported wavelets. *Commun. Pure Appl. Math.* **1988**, *41*, 909−996.

Received for review November 21, 2006. Revised manuscript received April 22, 2007. Accepted April 25, 2007. We acknowledge the financial support of the Key Technologies Program of the Ministry of Science and Technology, People's Republic of China (Project 2001BA501A16B).

JF063363C