Application of a Hybrid Variable Selection Method for Determination of Carbohydrate Content in Soy Milk Powder Using Visible and Near Infrared Spectroscopy

XIAOJING CHEN*,[†,‡] AND XINXIANG LEI[§]

College of Physics and Electronic Information and Department of Chemistry, Wenzhou University, Wenzhou 325027, China, and Department of Physics, Xiamen University, Xiamen, 361005, China

Visible and near-infrared (Vis-NIR) spectroscopy was investigated to fast determine the carbohydrate content in soy milk powder. A hybrid variable selection method was proposed. In this method, a simulate annealing (SA) algorithm was first operated to search the optimal band (OB) in the wavelet packet transform (WPT) tree. The OB with 47 variables was further selected by SA (WTP-OB-SA). Finally, the number of variables was reduced from 47 to 20. The best partial least-squares prediction with a high residual predictive deviation (RPD) value of 12.2242 was obtained using these 20 variables with the correlation coefficient (r) and root-mean-square error of prediction (RMSEP) being 0.9967 and 0.1669, respectively. The results indicated that Vis-NIR spectroscopy could efficiently determine the carbohydrate content in soy milk powder. The WPT-OB-SA selection method eliminated redundant variables and improved the prediction ability.

KEYWORDS: Visible and near-infrared spectroscopy; variable selection; soy milk powder; wavelet packet transform; simulated annealing

INTRODUCTION

Soy milk (also called soya milk or soybean milk) is a beverage made primarily from soy beans and milk. Soy milk is produced by soaking dry soybeans and subsequently grinding them with milk. Soy milk powder is a manufactured dairy product made by evaporating soy milk to dryness. Carbohydrates are one of the three main components of food that provide energy and other factors to the body. Sugars, including fructose, glucose, and lactose, are the main forms of carbohydrates. Carbohydrates should be part of a healthy diet, but they can also affect the plasma sugar level. Because carbohydrate consumption is directly related to diabetes and other illnesses such as obesity, the issue that will strongly dominate the health food market is blood sugar management and healthy foods with optimal carbohydrate contents (*1*). It has been estimated that there were more than 30 million people in China with diabetes in 2007. Thus, the rapid measurement of carbohydrate content in soy milk powder is important for people's health.

There are few reports that have evaluated the quality of soy milk powder. Some research has applied chemical analysis or sensory analysis techniques for evaluating the quality of milk or milk powder. These techniques include Rose–Gottlieb,

Soxhlet extraction, Babcock, and Gerber methods (*2*), the Kjeldahl method (*3*), high-performance liquid chromatography–mass spectrometry (*4*), atomic absorption spectrometry (*5*), and NMR spectroscopy (*6*). However, these methods are expensive and labor intensive, result in sample destruction, and are not suitable for the rapid carbohydrate content determination and real-time monitoring during the production of soy milk powder. Alternative fast procedures that preserve sample integrity and reduce both financial and personnel costs are desirable, especially if the accuracy and precision of the techniques are comparable to the traditional chemical methods.

Visible and near-infrared (Vis-NIR) spectroscopy has wide applicability due to key features such as high speed, low cost, and reliable detection for quantitative and qualitative analysis (*7*). The NIR technique has been applied for evaluating the quality and discrimination of milk powder (*8*–*13*). However, few reports have reported on the quality evaluation of soy milk powder using Vis-NIR spectroscopy. Vis-NIR spectra have typical characteristics of wide broad, nonspecific, and overlapping bands (*14*). The analysis of Vis-NIR spectra requires optimization because of the wide wavelengths as input variables and the large number of samples. The selection of variables would influence the quality of multivariate calibration models (*15*). Some irrelevant variables may affect multivariate calibration, and the elimination of such variables can predigest calibration modeling and improve prediction results in terms of accuracy and robustness.

* To whom correspondence should be addressed. Tel: +86-577-86593111. Fax: +86-577-86593111. E-mail: chenxj10@yahoo.cn.
† College of Physics and Electronic Information, Wenzhou University.
‡ Department of Physics, Xiamen University.
§ Department of Chemistry, Wenzhou University.

Table 1. Carbohydrate Content Statistic Values of 300 Soy Milk Powder Samples from Five Brands

data set	no. of samples	maximum	minimum	mean $\pm$ SD[a]
brand 1	60	5.822	4.882	5.432 $\pm$ 0.171
brand 2	60	10.489	10.300	10.397 $\pm$ 0.051
brand 3	60	10.292	9.813	10.023 $\pm$ 0.148
brand 4	60	11.398	10.905	11.127 $\pm$ 0.151
brand 5	60	8.618	7.901	8.313 $\pm$ 0.167
calibration set	200	11.372	4.882	9.061 $\pm$ 2.039
prediction set	100	11.398	5.121	9.052 $\pm$ 2.059

[a] Standard deviation.

We propose a hybrid variable selection method in combination of wavelet packet transform (WPT) with simulated annealing (SA). The objective of this study was to investigate the feasibility of using Vis-NIR spectroscopy to predict the carbohydrate content in soy milk powder. A partial least-squares (PLS) model was established based on selected variables from the optimal band (OB) in the WPT tree using an SA algorithm. The performance of the proposed variable selection method was compared with other PLS models using the following three inputs: all spectral variables, selected variables from the whole spectra using SA, or variables from the OB.

MATERIALS AND METHODS

Sample Preparation. Five popular brands of soy milk powder found in China with different production times were purchased in several local markets, including Beingmate (brand 1), Xiyangyang (brand 2), Sanlu (brand 3), Weiwei (brand 4), and Sandun (brand 5). All of these soy milk powder brands were frequently purchased by Chinese people. The first three brands were standard soy milk powders found in markets, whereas the remaining two brands were considered to be a health-conscious soy milk powder for diabetics because the carbohydrate content is lower.

Before the experiment, the soy milk powder samples were stored in the laboratory at a constant temperature of 25 $\pm$ 1 °C for more than 48 h to ensure room temperature equalization. A total of 300 samples (60 samples for each variety) were prepared for further treatments. Forty samples were selected randomly from each variety, and a total of 200 soy milk powder samples were used in the calibration set for calibration and validation, whereas the remaining 100 samples (20 for each variety) were for the prediction set. No single sample was used at the same time in the calibration and prediction sets. To compare the performance of the different calibration models, the samples in the calibration and prediction sets remained unchanged for all calibration models, and this was set as a basic condition in this paper.

Spectral Collection and Reference Methods for Carbohydrate Content. The Vis-NIR diffuse reflectance (375−1024 nm) spectra were determined using a U-4100 spectrophotometer (Hitachi High Technologies, Inc., Tokyo, Japan). Before the calibration stage, the spectral data should be preprocessed to ensure optimal performance. Standard normal variate (SNV) (*16*) was applied for light scatter correction and reducing the changes in the light path length.

The reference value of the carbohydrate content was measured by an ABBebenchtop refractometer (model: WAY-2S, Shanghai Precision & Scientific Instrument Co. Ltd., Shanghai, China). The refractive index accuracy was $\pm$0.0002, and the Brix (%) range was 0−95% with temperature correction. A 15 w/v % of samples was prepared in distilled water, which required the dissolution of 15 g of soy milk powder in 100 mL of distilled water.

The statistic values of carbohydrate content in soy milk powder are shown in **Table 1**. The difference between the maximum and the minimum values was insignificant for the carbohydrate content in the soy milk powder obtained from the same brand. The reason for this observation was that soy milk powder was a mixed powder, and the carbohydrate content was similar for samples obtained from the same brand. Moreover, the samples in the calibration and prediction sets were selected randomly. Therefore, the range and mean values of the
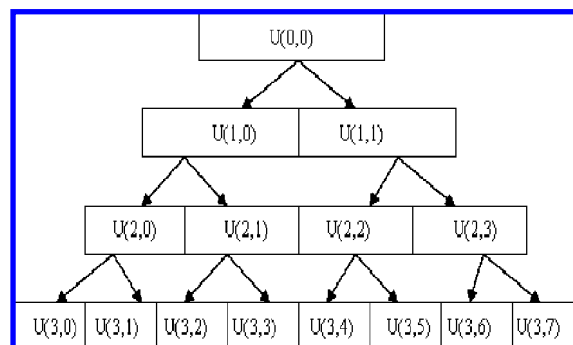
Figure 1. Full WPT binary tree; each node is identified by the couple of indices U (j,k), where j is the level of decomposition and k is the position of the node at that level of decomposition.

carbohydrate contents were similar in the calibration and prediction sets. However, carbohydrate contents in the calibration and prediction sets covered an adequate range, and this observation was valuable in developing the models.

Wavelet Packet Transform (WTP). As an extension of wavelet transform (WT) (*17*), WTP is a powerful signal processing technique. It transforms the raw spectral data into different frequency bands, and the frequency component in different bands has a different contribution to the multivariate model (*18*). As such, finding the most useful band that represents the most contribution to the model is an important issue.

In fast WT (*19*), a partial multiresolution analysis is performed. Only the approximation coefficients (low-pass node) are employed to deduce both scale and wavelet coefficients at the next resolution level. However, WPT allows a full multiresolution analysis; both the approximation and the detail coefficients (high-pass node) are involved to simultaneously decompose at the next resolution level (*20*). As a result, a library of sub-band including low and high frequencies is obtained. A schematic diagram for the WPT decomposition method is shown in **Figure 1**.

SA Algorithm. A SA algorithm, a simulation of an annealing process used for metals, was proposed by Kerkpatrick et al. in 1983 (*21*). Arguably, it offers the simplest and the most elegant solution with the "best" record for solving combinatorial optimization problems. Unlike other algorithms, the SA algorithm allows various types of transitions in which some of them may oppose the goal (*22*). Hence, the SA algorithm has been widely applied to many optimization problems, such as multiobjective optimization of a constrained problem (*23*), the maximum clique problem (*24*), and multiparameter of water optical properties from above-water remote-sensing reflectance (*25*).

In the SA algorithm, a problem starts with an initial solution, and this solution can be easily changed. Yet, as the temperature T, which is the control parameter in the analogy with temperature in the physical annealing processing, is decreased, changing configuration is increasingly difficult. Finally, if T is lowered sufficiently, no further changes in the solution space are possible. To avoid being frozen at a local optimum, the SA algorithm moves slowly through the solution space. This controlled improvement of the objective value is accomplished by accepting nonimproving moves with a certain probability that decreases as the algorithm progresses (*23−25*).

WPT-OB-SA Method. This method is a combination method of WTP and SA and has two steps. In the first, the raw spectra were decomposed by WTP, and all of the bands constituting the WPT tree were obtained. The SA algorithm was first employed to search for the OB in the WPT tree, which has the most contribution to calibration model. Because of the effective searching ability of SA, all of the bands were efficiently searched. In the second step, variables of the OB were further selected by SA to obtain a more parsimonious and efficient model. On the aspect of informative variables selection, the key advantages of the WPT-OB-SA method are the number of variables to the build PLS model is reduced without significantly compromising the prediction ability of the model.

Determination of SA Algorithm Parameters. In this work, the initial temperature T_I, of the SA algorithm was 100 and the termination temperature T_S was 0. Student's t distribution was employed to generate a new solution in the SA algorithm. The random disturbance can be

regarded as jumping the optimal model. The metropolis criterion uses the change in function values between the current point and the new point to determine whether the new point was acceptable or not. The annealing schedule was an exponential annealing schedule, which updated the current temperature based on the initial temperature and the current annealing parameter k (the number of evaluations of the objective function):

$$T_{n+1} = 0.95^k T_n \tag{1}$$

In general, there are two stopping rules; the first one was that the number of temperature transitions satisfies the temperature termination rules, and the second rule was that the neighbor solution was not improved after a certain period (*26*). In our strategy, the algorithm stopped when the average change on the value of the fitness function at the current point was less than 1×10^{-6} after 500 iterations.

Evaluation of the Fitness Function. The performance of the SA was evaluated through a fitness function, also known as an objective function. The function value was the criterion for guiding SA to the global optimum. The prediction ability of the calibration model was evaluated with the parameters of correlation coefficient (r) and root-mean-square error of cross-validation (RMSECV). The ideal model should have a high r value and low RMSECV values (*27*). So, the fitness function was defined as follows:

$$\max f(X) = \frac{R}{1 + \text{RMSECV}} = \frac{\sum_{i=1}^{n}(c_i^p - \bar{c}^p)(c_i^o - \bar{c}^o)}{\left\{1 + \left[\frac{1}{n}\sum_{i=1}^{n}(c_i^o - c_i^p)^2\right]^{\frac{1}{2}}\right\} \cdot \left[\sum_{i=1}^{n}(c_i^p - \bar{c}^p)^2 \cdot \sum_{i=1}^{n}(c_i^o - \bar{c}^o)^2\right]^{\frac{1}{2}}} \tag{2}$$

where c_i^o is the actual concentration of spectrum i in the calibration set, $\bar{c}^o$ is the mean of c_i^o, c_i^p is the concentration predicted by the model, $\bar{c}^p$ is the mean of c_i^p, and n is the number of calibration sets. The fitness function was measured in the form of leave-one-out cross-validation using PLS. Different numbers of latent variables (LVs) were applied to build the calibration models. The optimal number of LVs of PLS was determined according to RMSECV. Thus, the model was built using $n - 1$ training samples, and the one left out was used for prediction.

RESULTS AND DISCUSSION

Absorbance Spectra of Soy Milk Powder and Statistical Analysis of the Carbohydrate Content. Typical spectra of soy milk powder samples are shown in **Figure 2a**. After SNV pretreatment, most effects caused by the environment have been eliminated. In **Figure 2b**, spectra from all five soy milk varieties have similar gross absorbance patterns. There are some cross-overs and overlaps among these samples. There is a strong absorption in the violet region between 375 and 400 nm. Absorbance values decrease as the wavelength increases, and the spectral curves flatten. Although the spectra trends are similar, some latent differences and features are present due to the chemical components and color variance, which cannot be distinguished visibly. Some obvious differences in the region of 450−500 nm are observed following close inspection of the spectra. These differences may have resulted from the color variance caused by the different production stages and chemical reactions (*28*). The production stages and chemical reactions are related to the concentration of carbohydrate content. Therefore, different varieties of soy milk powder with different internal qualities such as sugars and organic acids are reflected in the Vis-NIR spectra. However, it was not easy to visibly

observe differences. Chemometrics has been used for the further analysis. Latent features of the spectra need to be mined for the determination of the carbohydrate content in soy milk powders.

PLS Models with the Whole Spectra Data. PLS models were developed using the preprocessed spectra data by SNV. The whole Vis-NIR spectra were used to establish the PLS model. The predictive capability was evaluated by the following indices: r, RMSECV, and root-mean-square error of prediction (RMSEP). A good model should have a high r, low RMSECV, and RMSEP. These standards should be taken into consideration for distinguishing systematic errors and studying the correlation between the carbohydrate content and the Vis-NIR models. No outliers were detected in the calibration set during the development of the PLS models. In the development on all PLS calibration models, leave-one-out full cross-validation was used to validate their quality. Different numbers of LVs were applied to build the calibration models. The optimal number of LVs for PLS was determined as nine when the REMSECV reached a minimal value. Results of the calibration and prediction sets are shown in **Table 2**. The r and RMSECV for the calibration set were 0.9933 and 0.2617, and the r and RMSEP for prediction set were 0.9940 and 0.2596.

PLS Models with Selected Spectra Data from Whole Spectra. As aforementioned, whole Vis-NIR spectra as input set of PLS model contained useless or irrelevant information for the calibration model, for example, noise and background signals. As an excellent search method, SA was employed to seek the optimal wavelengths in the Vis-NIR region. In the SA process, each wavelength variable had two logical values, 1 and 0; 1 represented the selection of the corresponding wavelength, whereas 0 meant that the wavelength was not selected.

The selected wavelengths and the best function value are shown in **Figure 3**. After SA selection, 309 wavelengths of total 650 wavelengths were selected to obtain the best fitness function value of 0.8243. Spectra of selected wavelengths were applied to build the PLS model. The optimal number of LVs for PLS was determined as 10. Results of the calibration and prediction sets are shown in **Table 2**. The r and RMSECV for calibration set were 0.9957 and 0.2030, and the r and RMSEP for the prediction set were 0.9951 and 0.2113. However, as the SA was executed using all of the spectral variables in the Vis-NIR region, the use of this large (i.e., hundreds) set of variables resulted in the SA process being complex and time consuming. Thus, we used WPT to first find the OB that represents the most featured information. Here, it should be noted that the number of variables as inputs of PLS model was changing during the SA search running, but the optimal LVs remained unchanged and was determined by the minimal RMSECV value of the original variables. In the final calibration model, optimal LVs were determined by minimal RMSECV value of selected variables.

OB Searched by SA (WPT-SA). After all frequency bands were obtained by WPT decomposition, SA was employed to seek the OB in the WTP tree. Before SA was employed to search for the OB, the search range of the SA approach should be defined. The ranges include lower and upper bound constraints of the decomposition level and the max node number in each level. Lower and upper bound constraints of the WPT decomposition level should be initially determined, upper bound, namely, for a signal with a length of N, the theoretically maximum decomposition level J, which was calculated by the Matlab command "wmaxlev"; the lower bound of the decomposition level was defined as zero, namely, the signal was not
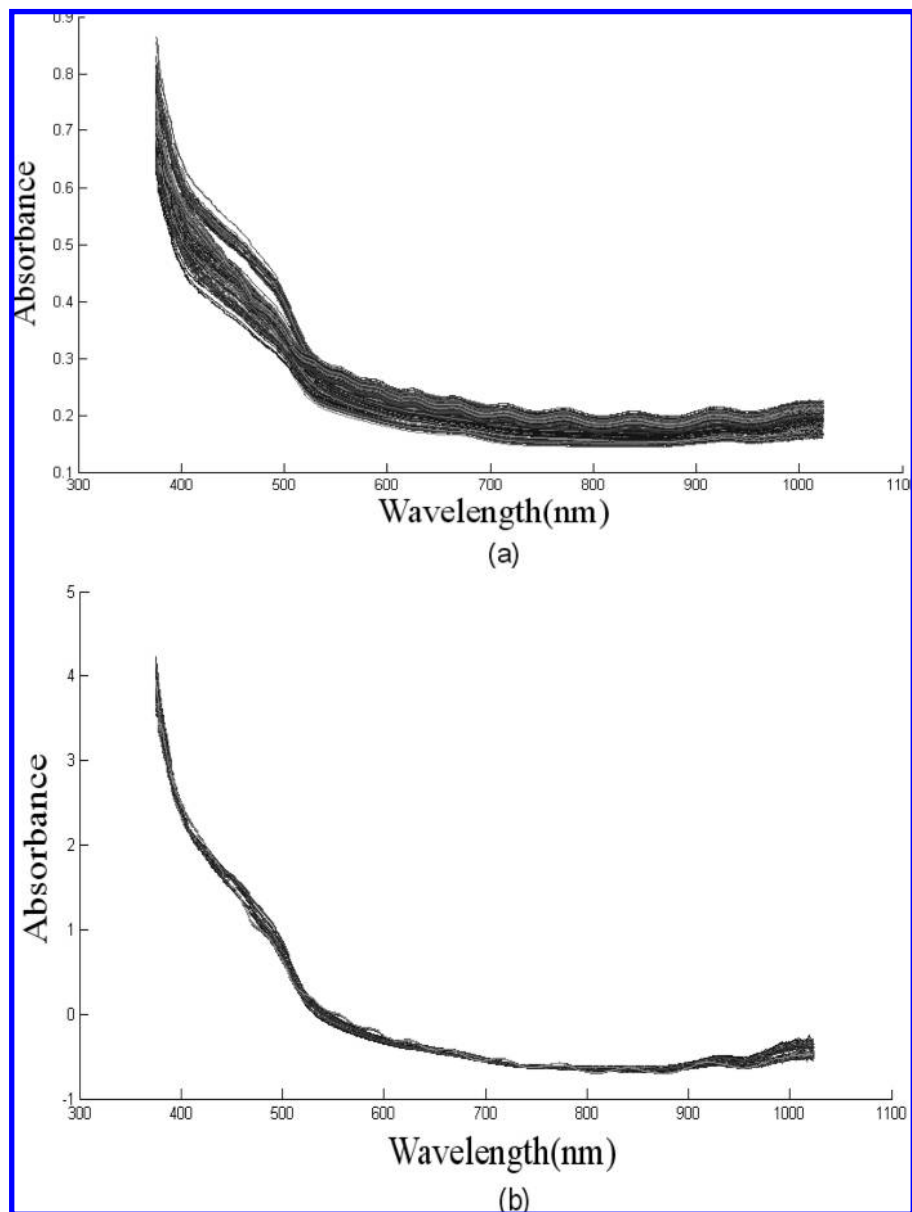
Figure 2. Original absorbance spectra (**a**) and SNV pretreated absorbance spectra (**b**) of a total of 300 soy milk powder samples of five varieties at Vis-NIR regions (375−1024 nm).

Table 2. Prediction Results of Carbohydrate Content of 100 Soy Milk Powder Samples in Prediction Set Using Vis-NIR Spectra Based on Different Variable Selection Methods

| variable selection method	no. of selected variables	no. of LVs of PLS	calibration		prediction	
			r^c	RMSECV	r	RMSEP
none	650	9	0.9933	0.2617	0.9940	0.2596
WPT-SA[a]	47	11	0.9967	0.1929	0.9959	0.1978
SA[b]	309	10	0.9957	0.2030	0.9951	0.2113
WPT-OB-SA	20	11	0.9973	0.1668	0.9967	0.1669

[a] Optimal band of WPT. [b] Simulated annealing. [c] Correlation coefficient.

decomposed. Subsequently, the constraint range of node number was determined according to the decomposition level, because WTP decomposed the signal as the full binary tree (**Figure 1**). The upper bound of the node number was set as 2^l, where l was the decomposition level. The low bound number of the node was set to zero. In this work, the Db4 wavelet function

was employed, and all calculations were performed using MATLAB 7.6 (the Mathwork, Natick, MA).

The optimal decomposition level, the node number, and the best function value are shown in **Figure 4**. The optimal decomposition level was determined as four, and the optimal node index was (4, 0). The best fitness function value of 0.8337 was obtained by SA. The OB, which had 47 variables, was set as the inputs of the PLS model. The optimal number of LVs for PLS was determined as 11. Results of the calibration and prediction sets are shown in **Table 2**. The r and RMSECV for calibration set were 0.9967 and 0.1929, and the r and RMSEP for the prediction set were 0.9959 and 0.1978. The prediction results were found to be better than using optimal wavelengths proposed by SA in the Vis-NIR region.

Further Variable Searching on OB Using SA (WPT-OB-SA). In the OB, there were 47 variables; therefore, redundancy was still present. SA was again employed to seek the optimal variables in the OB. In the SA process, the 47 variables had two logical values, 1 and 0, which represented selected or not selected, respectively. The selected wave-
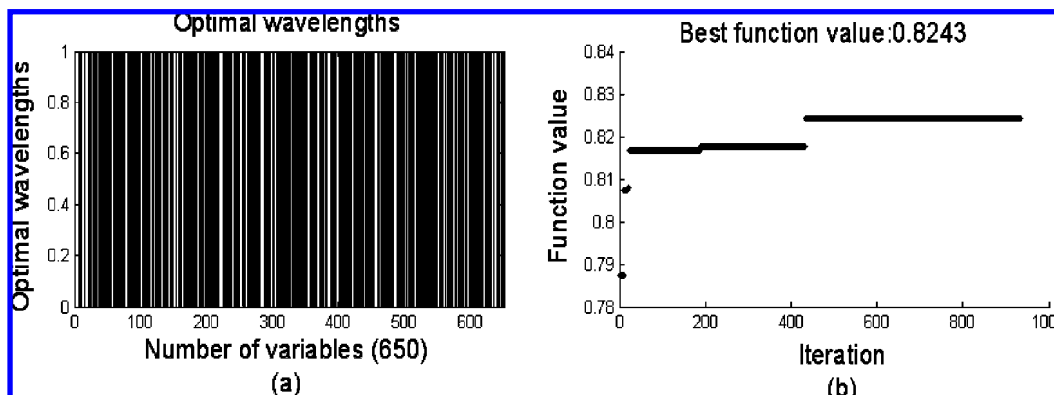
Figure 3. Selection results of optimal wavelengths in the full Vis-NIR region proposed by SA. (**a**) Optimal wavelengths and (**b**) best fitness function value. The white column represents the unselected wavelengths.
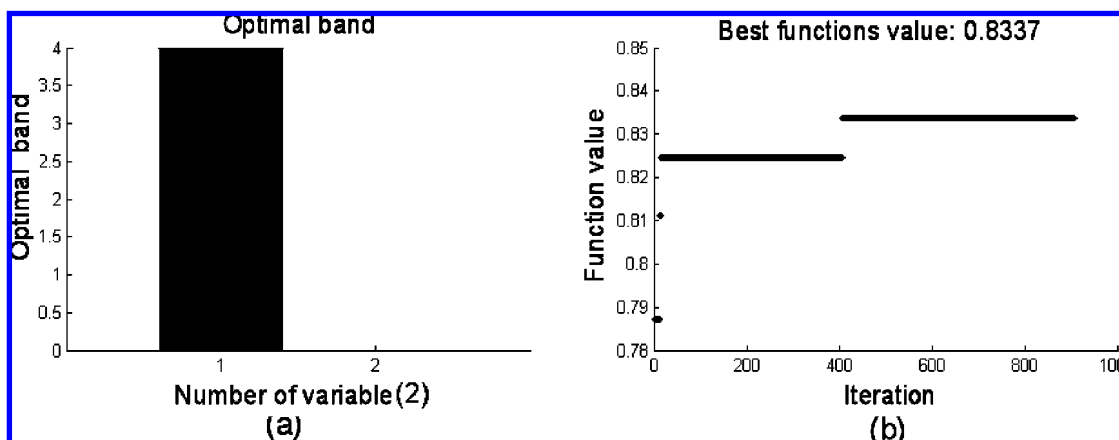
Figure 4. Result of optimal node by SA. (**a**) Optimal node index and (**b**) best fitness function value.
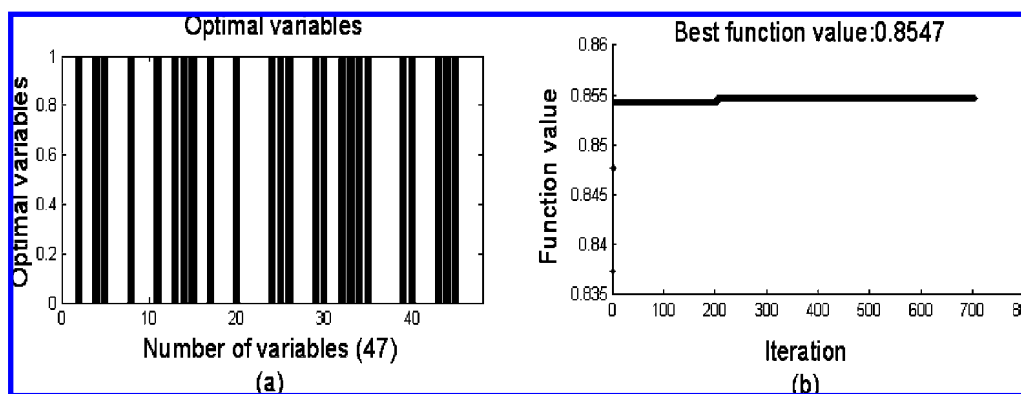

Figure 5. Selection results of optimal variables in the OB of WTP by SA. (**a**) Optimal variables and (**b**) best fitness function value. The white column represents the unselected wavelengths.

lengths and the best function value are shown in **Figure 5**. After SA selection, 20 variables out of the 47 variables in the OB were selected to obtain the best fitness function value of 0.8547. The retained variables were applied to build the calibration PLS model (WPT-OB-SA-PLS). The optimal number of LVs for PLS was determined as 11. The prediction results of the calibration and prediction sets are shown in **Table 2**. The r and RMSECV for the calibration set were 0.9973 and 0.1668, and the r and RMSEP for prediction set were 0.9967 and 0.1669. Moreover, a high residual predictive deviation value (RPD) of 12.2242 was obtained. The RPD is the standard deviation of reference data for the validation samples divided by the RMSECV (*29*). In general, RPD values > 8.0 are considered adequate for any application (*29*). Thus, the ability of the model to predict carbohydrate content

in soy milk powder was significantly high. **Figure 6** shows the predicted vs reference charts for carbohydrate content of samples in prediction set. The solid line represents the regression line.

The results demonstrated that the best performance was achieved with these 20 variables selected by WPT-OB-SA. The r values for the calibration and prediction set were larger, and the RMSEP and RESECV values were smaller than any of the other models that were established based on the whole 650 spectral variables. These mined 20 variables obtained the best predictive information on the carbohydrate content prediction in soy milk powder. Furthermore, r for the calibration and prediction sets and the RMSEP and RESECV were very close. It indicated that the overfitting problem was avoided, and a good stability and generalization were

Determination of Carbohydrate Content in Soy Milk Powder

J. Agric. Food Chem., Vol. 57, No. 2, 2009 **339**
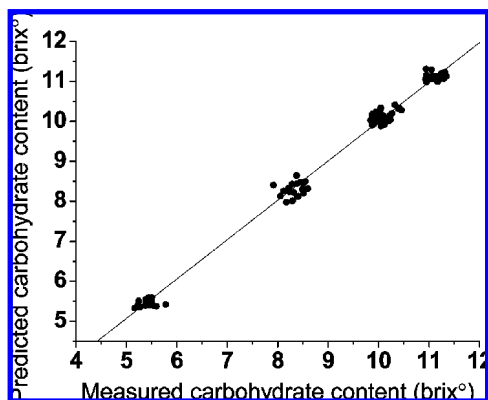
Figure 6. Predicted vs measured values of the carbohydrate content in soy milk powder samples in the prediction set by the WPT-OB-SA-PLS model. The solid line is the regression line.

achieved by this new model. According to the results of the calibration and prediction set, the 20 variables as feature information could represent main diagnostic information, and these variables could be applied instead of the whole original spectra for predicting the carbohydrate content in soy milk powder.

In conclusion, the determination of the carbohydrate content in soy milk powder could be successfully performed through Vis-NIR spectroscopy. A hybrid variable selection method, WPT-OB-SA, was proposed. The overall results indicated that Vis-NIR spectroscopy was an effective spectroscopic technique to determine the carbohydrate content in soy milk powder based on PLS models. Taking into account that the original 650 input variables were reduced to 20 and that the accuracy and precision results were better than the whole 650 spectral variables, the proposed hybrid variable selection method, WPT-OB-SA, is a powerful tool to compress the spectral data sets and select the optimal predictive information. This hybrid selection model not only reduced the number of analysis variables but also improved the prediction ability. However, as the 20 variables still represent a rather set for this application, further variable selection on these 20 variables will be performed to find more optimal variables.

ABBREVIATIONS USED

Vis-NIR, visible and near-infrared; SA, simulated annealing; OB, optimal band; WPT, wavelet packet transform; SNV, standard normal variate; PLS, partial least-squares; LVs, latent variables; NIR, near-infrared; RMSECV, root-mean-square error of cross-validation; RMSEP, root-mean-square error of prediction.

LITERATURE CITED

(1) Sloan, A. E. Healthy vending and other emerging trends. *Food Technol.* **2005**, *59*, 26–35.

(2) Rosenthal, I.; Rosen, B. 100 years of measuring the fat content of milk. *J. Chem. Educ.* **1993**, *70*, 480–482.

(3) Kamizake, N.; Gonçalves, M.; Zaia, C.; Zaia, D. Determination of total proteins in cow milk powder samples: a comparative study between the Kjeldahl method and spectrophotometric methods. *J. Food Compos. Anal.* **2003**, *16*, 507–516.

(4) Luykx, D.; Cordewener, J.; Ferranti, P.; Frankhuizen, R.; Bremer, M.; Hooijerink, H.; America, A. Identification of plant proteins in adulterated skimmed milk powder by high-performance liquid chromatography-mass spectrometry. *J. Chromatogr. A* **2007**, *1164*, 189–197.

(5) Wagley, D.; Schmiedel, G.; Mainka, E.; Ache, H. J. Direct determination of some essential and toxic elements in milk and milk power by graphite furnace atomic absorption spectrometry. *At. Spectrosc.* **1989**, *10*, 106–111.

(6) Hu, F.; Furihata, K.; Kato, Y.; Tanokura, M. Nondestructive quantification of organic compounds in whole milk without pretreatment by two-dimensional NMR spectroscopy. *J. Agric. Food Chem.* **2007**, *55*, 4307–4311.

(7) Yan, Y. L.; Zhao, L. L.; Han, D. H.; Yang, S. M. *The Foundation and Application of Near-Infrared Spectroscopy Analysis*, 1st ed.; China Light Industry Press: Beijing, China, 2005; p 32.

(8) Borin, A.; Ferrao, M. F.; Mello, C.; Maretto, D. A.; Poppi, R. J. Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk. *Anal. Chim. Acta* **2006**, *579*, 25–32.

(9) Maraboli, A.; Cattaneo, T. M. P.; Giangiacomo, R. Detection of vegetable proteins from soy, pea and wheat isolates in milk powder by near infrared spectroscopy. *J. Near Infrared Spectrosc.* **2002**, *10*, 63–69.

(10) Wu, D.; Feng, S.; He, Y. Infrared spectroscopy technique for the nondestructive measurement of fat content in milk powder. *J. Dairy Sci.* **2007**, *90*, 3613–3619.

(11) Wu, D.; He, Y.; Feng, S.; Sun, D. W. Study on infrared spectroscopy technique for fast measurement of protein content in milk powder based on LS-SVM. *J. Food Eng.* **2008**, *84*, 124–131.

(12) Wu, D.; Feng, S.; He, Y. Short-wave near-infrared spectroscopy of milk powder for brand identification and component analysis. *J. Dairy Sci.* **2008**, *91*, 939–949.

(13) Wu, D.; He, Y.; Feng, S. Short-wave near-infrared spectroscopy analysis of major compounds in milk powder and wavelength assignment. *Anal. Chim. Acta* **2008**, *610*, 232–242.

(14) Shao, X. G.; Wang, F.; Chen, D.; Su, Q. D. A method for near-infrared spectral calibration of complex plant samples with wavelet transform and elimination of uninformative variables. *Anal. Bioanal. Chem.* **2004**, *378*, 1382–1387.

(15) Ying, Y. B.; Liu, Y. D. Nondestructive measurement of internal quality in pear using genetic algorithms and FT-NIR spectroscopy. *J. Food Eng.* **2008**, *84*, 206–213.

(16) Barnes, R. J.; Dhanoa, M. S.; Lister, J. S. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *J. Appl. Spectrosc.* **1989**, *43*, 772–777.

(17) Cocchi, M.; Corbellini, M.; Foca, G.; Lucisano, M.; Pagani, A.; Tassi, L.; Ulrici, A. Classification of bread wheat flours in different quality categories by a wavelet-based feature selection/classification algorithm on NIR spectra. *Anal. Chim. Acta* **2005**, *544*, 100–107.

(18) Chen, D.; Shao, X. G.; Hu, B.; Su, Q. D. A Background and noise elimination method for quantitative calibration of near infrared spectra. *Anal. Chim. Acta* **2004**, *511*, 37–45.

(19) Mallat, S. G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693.

(20) Cocchi, M.; Seeber, R.; Ulrici, A. WPTER: Wavelet packet transform for efficient pattern recognition of signals. *Chemom. Intell. Lab. Syst.* **2001**, *57*, 97–119.

(21) Kirkpatrick, S.; Vecchi, M. P. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680.

(22) Hijrchner, U.; Kalivas, J. H. Further investigation on a comparative study of simulated annealing and genetic algorithm for wavelength selection. *Anal. Chim. Acta* **1995**, *311*, 1–13.

(23) Suman, B. Study of simulated annealing based algorithms for multiobjective optimization of a constrained problem. *Comput. Chem. Eng.* **2004**, *28*, 1849–1871.

(24) Geng, X.; Xu, J.; Xiao, J.; Pan, L. A simple simulated annealing algorithm for the maximum clique problem. *Inf. Sci.* **2007**, *177*, 5064–5071.

(25) Salinas, S. V.; Chang, C. W.; Liew, S. C. Multiparameter retrieval of water optical properties from above-water remote-sensing reflectance using the simulated annealing algorithm. *Appl. Opt.* **2007**, *46*, 2727–2742.

340 *J. Agric. Food Chem.,* Vol. 57, No. 2, 2009

Chen and Lei

(26) Ghazanfari, M.; Alizadeh, S.; Fathian, M.; Koulouriotis, D. E. Comparing simulated annealing and genetic algorithm in learning FCM. *Appl. Math. Comput.* **2007**, *192*, 56–68.

(27) Liu, F.; He, Y. Classification of brands of instant noodles using Vis/NIR spectroscopy and chemometrics. *Food Res. Int.* **2008**, *41*, 562–567.

(28) Wang, F. Y. *Modern Techniques for Food Fermentation*; China Light Industry Press: Beijing, China, 2004.

(29) Williams, P. C. Implementation of near-infrared technology. In *Near-Infrared Technology in the Agricultural and Food Industries*, 2nd ed.; Williams, P., Norris, K., Eds.; American Association of Cereal Chemists: St. Paul, MN, 2001; pp 145–169.