ANALYTICA CHIMICA ACTA 610 (2008) 232–242

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/aca

Short-wave near-infrared spectroscopy analysis of major compounds in milk powder and wavelength assignment

Di Wu, Yong He, Shuijuan Feng*

College of Biosystems Engineering and Food Science, Zhejiang University, 268 Kaixuan Road, Hangzhou, Zhejiang 310029, China

ARTICLE INFO

Article history:
Received 15 November 2007
Received in revised form
15 January 2008
Accepted 20 January 2008
Published on line 2 February 2008

Keywords:
Short-wave near-infrared spectroscopy (short-wave NIRS)
Milk powder
Wavelength assignment
Partial least-squares (PLS)
Least-squares support vector machine (LS-SVM)

ABSTRACT

In this study, short-wave near-infrared (NIR) spectroscopy at 800–1050 nm region was investigated for the analysis of main compounds in milk powder. Through quantitative analysis, the feasibility is further demonstrated for the simultaneous measurement of fat, proteins and carbohydrate in milk powder. Two models, partial least-squares and least-squares support vector machine, were compared and utilized for regression coefficients and loading weights. The affect of standard normal variate spectral pretreatment to model performance was evaluated. Based on the resulted coefficients and loading weights, interesting wavelength regions of nutrition in milk powder are screened and the assignment of all specific wavelengths is firstly proposed in the details associated with chemical base. Instead of the whole short-wave NIR spectral data, these assigned wavelengths which can be reliably exploited were used for the content determination. Compared with other spectroscopy technique, assigned short-wave NIR spectral wavelengths did a good work. Determination coefficients for prediction are 0.981, 0.984, and 0.982, respectively for three components. The proposed wavelength assignment in the short-wave NIR region could be used for the component contents determination of milk powder, and could be as a guidance to interpret the spectra of milk powder.

1. Introduction

Short-wave near-infrared (NIR) spectroscopy technique is promising for the fast and nondestructive analysis of biological materials. As its high transmittance ability in the region of 700–1100 nm, the short-wave NIR spectra are being applied to construct excellent detectors for the nondestructive component determination in biological materials. The short-wave NIR region allows NIR energy to penetrate more deeply into a sample with much less heating effect than the long-wave NIR region (1100–2500 nm). Single-beam data can be reliably exploited in the short-wave NIR region to reduce measurement time [1] and the effects that arise from the water vibration in long-wave NIR region can be diminished [2]. Moreover, compared with other spectroscopy technique, short-wave NIR spectra can be obtained using inexpensive fiber optics, it can be measured with inexpensive light sources (tungsten lamps) and detectors (silicon diode array) [3], and it can be designed to be an inexpensive handheld instrument. Sasic and Ozaki used short-wave NIR to quantitative analysis of fat, protein, and lactose in raw milk by partial least-squares regression and wavelength assignment [4]. He et al. studied the measurement of sugar content of yogurt using short-wave NIR [5]. Subedi et al. investigated the mango eating quality at harvest using short-wave NIR [6].

Contents of main components are one of the important factors for the quality of milk powder. Chemical composition analyses are routinely performed for the inspection of

* *Corresponding author*. Tel.: +86 571 89671143; fax: +86 571 86971143.
E-mail address: yhe@zju.edu.cn (Y. He).
0003-2670/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2008.01.056

main components in commercial milk powder, e.g. fats, proteins, and carbohydrates. However, most currently available chemical techniques for these analyses are time consuming, destructive and costly. A rapid, non-destructive, reliable and less expensive method is highly desirable in the milk powder industry. NIR spectroscopy is a feasible way for the components determination of the milk powder based on the NIR spectra [7–10]. Almost all the NIR experiments of milk powder were made for the long-wave NIR regions. Nutritional parameters in infant formulas powdered milk were also evaluated by Raman spectroscopy [11]. However, although with advantages over other spectroscopy technique, the study of quantitative and qualitative analysis of milk powder based on short-wave NIR spectra was less investigated. Moreover, there are a few investigations on the wavelength assignment in this region. The above mentioned reports were executed for the content determination based on the whole spectral data. Few reports were done for the content determination based on the assigned wavelengths. The wavelength assignment could be as a guidance to interpret the spectra of milk powder. If the performance based on wavelength assigned short-wave NIR spectra could be close to or higher than other common spectroscopy technologies, with its advantages of high transmittance ability and single-beam data reliable measurement, wavelength assigned short-wave NIR spectra can develop inexpensive on-line sensors and instruments for nondestructive determination and noninvasive diagnosed of the nutrition content in milk powder.

In this paper, the study was concentrated on the short-wave NIR spectra in the 800–1050 nm region. The feasibility of simultaneous measurement of major components in milk powder was newly investigated using this narrow wavelength, based on the quantitative analysis methods of partial least-squares (PLS) and least-squares support vector machine (LS-SVM). Finally the specific wavelength assignment was newly proposed. Instead of the whole short-wave NIR spectral data, these assigned wavelengths were used for the determination of three main components, fats, proteins and carbohydrates. The regression coefficients and loading weights for each wavelength and component were discussed to obtain the insight into the chemical base for quantitative analysis.

2. Materials and methods

2.1. Sample preparation

A total of 350 milk powder samples were used as the whole data set. Seven brands of 6–12-month infant milk powder were bought from several local super markets. Seven brands include the local brands from the common local brands, high-grade brands, and others from Chinese-foreign joint ventures corporation. Milk powder was stored in an ice filled cooler and transported to the laboratory to be kept at cold temperature ($4 \pm 1\,^{\circ}C$). The whole experiments were made at ambient temperature of 18–20 °C. Each sample was with full of the milk powder in the uniform glass containers (65 mm in diameter, 14 mm in height).

2.2. Spectra measurements

NIR reflectance spectra in the 700–1075 nm region were measured by a handheld FieldSpec Pro FR (325–1075 nm)/A110070, Trademarks of Analytical Spectral Devices, Inc. (Analytical Spectral Devices, Boulder, USA). Considering the 10° field-of-view of the spectral probe, the spectroradiometer was placed at a distance of approximately 150 mm and 45° angle away from the measurement area. As the only illumination, a light source of Lowell pro-lam 14.5 V Bulb/128690 tungsten (Ushio Lighting Inc., Japan) was applied about 300 mm away from the measurement area and 45° of horizon plane. The spectrum of each sample was the average of 30 successive scans. To avoid low signal-to-noise ratio, only the region of wavelengths (800–1050 nm) were employed for the calculations. Absorbance data were stored as log(1/R) (R = reflectance) at 1 nm intervals (250 spectra data points). Then these 30 values were averaged and stored as the absorbance value of this sample. All spectral data were stored in a computer and processed using the RS3 software for Windows (Analytical Spectral Devices, Boulder, USA) designed with a Graphical User Interface.

2.3. Content measurement of main component

Fat content was measured by Röse–Gottlieb method following GB/T 5413.3-1997 (National Standard of China). Protein content was determined by Kjeldahl method as described by GB/T 5413.1-1997 (National Standards of P.R. China) and the factor 6.38 was used to convert the nitrogen values to protein. Carbohydrate content was determined by Lane and Eynor's Method as described by GB/T 5413.5-1997 (National Standards of P.R. China). Content value is the weight per 100 g of milk powder.

2.4. Chemometric analysis

All chemometric analyses were carried out by Unscrambler ver 9.6 (CAMO PROCESS AS, OSLO, Norway) and MATLAB 7.1 (The Math Works, Natick, USA). The free LS-SVM toolbox (LS-SVM v 1.5, Suykens, Leuven, Belgium) was applied with MATLAB to derive all of the LS-SVM models. Four pretreated spectra, namely, first-derivative, second-derivative, standard normal variate (SNV) and multiplicative scatter correction (MSC) spectra led to simplified and more robust models. By calculating first and second derivation, sample-to-sample baseline variations are eliminated and also absorption peaks are enhanced. SNV [12,13] is a mathematical transformation method used to remove slope variation and to correct for scatter effects. MSC [14,15] corrects for difference in light scatter between samples before calibration. Based on total 350 samples, all the regression coefficients, loading weights of three components were automatically determined by software on two chemometric methods, namely, PLS and LS-SVM. In order to investigate the potential of short-wave NIR region for practice application, all the samples were divided into calibration and prediction sets, respectively. Hierarchical cluster analysis was established between samples. It is related to how similar the numerical properties of sample short-wave NIR spectra are. Each sample is linked to the closest sample or group of samples and a characteristic distance is used to describe this

union [11]. Samples belonged to calibration and prediction set were randomly selected based on hierarchical clusters. Stepwise linear regression analysis and LS-SVM were used to establish chemometric models with MATLAB. The predictive performance of models in this paper was evaluated by several standards, such as determination coefficient for calibration (R_c^2) and prediction (R_p^2), root mean square error of calibration (RMSEC) and prediction (RMSEP), and residual predictive deviation (RPD). Some other standards such as slope, offset and bias were taken into consideration for distinguishing systematic errors and studying the correlation between the reference and short-wave NIR models. A good model should have high determination coefficient and RPD values, low RMSEC and RMSEP.

3. Results and discussion

3.1. *Features of short-wave NIR spectra and spectral pretreatment*

Fig. 1a shows the whole 350 NIR spectra of milk powder in the 800–1050 nm region. As the water content is low in milk powder, the spectra here do not include the broad feature around 968 nm due to 2v1 + v3 (v1: symmetric stretching; v3: antisymmetric stretching) water vibration, which dominates the spectra of raw milk [4]. The changes of spectra baselines for the different milk powders can be eliminated by spectral pretreatment [16,17]. To avoid enhancing the noise, spectra were first smoothed before other pretreatments [18,19]. This smoothing was performed by using the Savitzky–Golay algorithm with smoothing points of nine. Four spectral pretreatments were executed based on the whole 350 NIR spectra of milk powder with PLS model. The results of different spectral pretreatment models are shown in Table 1. It could be seen that the best results of all three components were obtained based on SNV process. This is probably because that SNV is a scatter correction method used to remove the multiplicative interferences of scatter and particle size among the ingredients, which typically occur for milk powder. First-derivative, second-derivative

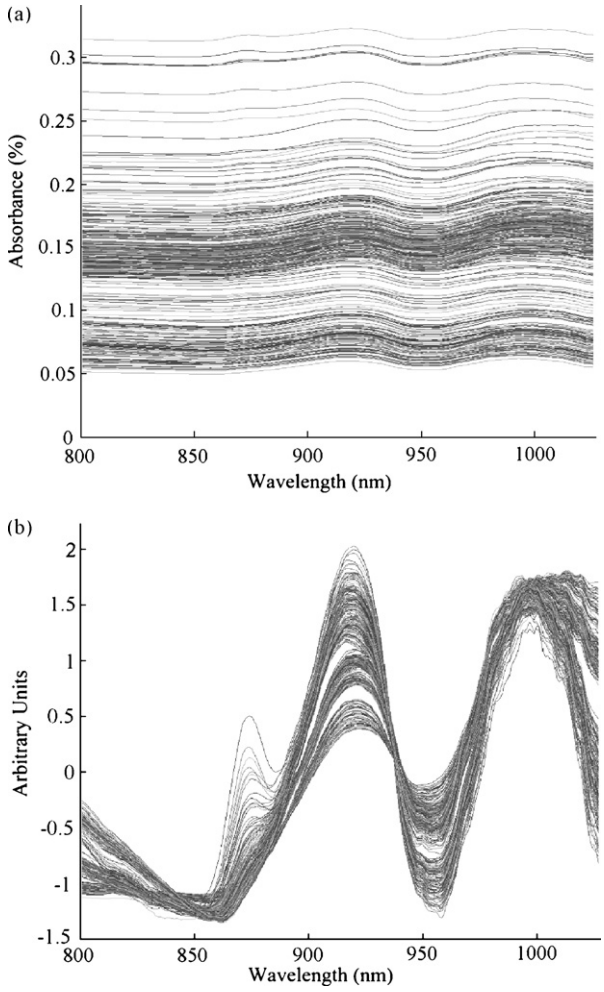

Fig. 1 – (a) Original short-wave NIR spectra of 350 milk powder samples of seven brands; (b) short-wave NIR spectra of 350 milk powder samples of seven brands after SNV pretreatment.

Table 1 – Results of PLS models with different spectral pretreatments based on short-wave NIR spectra

Component	Pretreatment	LW[a] number	Calibration		Validation				
			R_c^2	RMSEC	R_p^2	RMSEP	Bias	Slope	Offset
Fat	SNV	5	0.951	0.393	0.945	0.415	0.002	0.947	1.052
	MSC	6	0.947	0.410	0.939	0.434	0.003	0.943	1.136
	1st	5	0.933	0.460	0.925	0.481	0.002	0.928	1.434
	2nd	4	0.925	0.483	0.914	0.520	−0.001	0.910	1.787
Protein	SNV	6	0.933	0.302	0.927	0.314	−0.001	0.930	1.172
	MSC	5	0.917	0.337	0.912	0.348	0.001	0.915	1.438
	1st	5	0.925	0.318	0.925	0.320	0.001	0.922	1.256
	2nd	9	0.914	0.345	0.876	0.414	0.001	0.889	1.878
Carbohydrate	SNV	8	0.900	0.911	0.880	0.997	−0.004	0.893	5.904
	MSC	7	0.852	1.105	0.828	1.195	−0.005	0.846	8.487
	1st	10	0.885	0.931	0.870	1.057	−0.002	0.886	6.084
	2nd	4	0.861	1.071	0.839	1.156	0.005	0.846	8.507

[a] LW: loading weight.

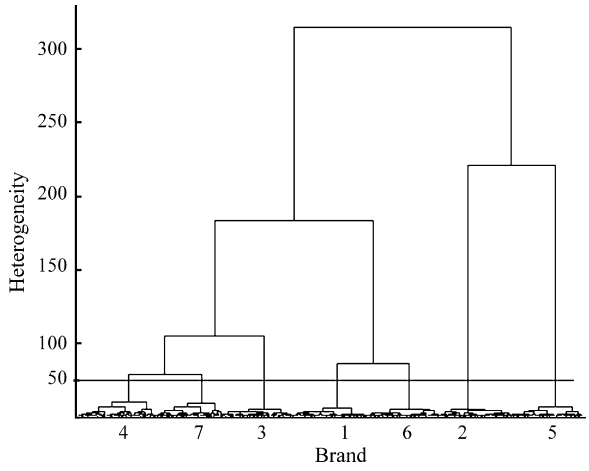

Fig. 2 – Dendrographic classification of 350 milk powder samples using the Euclidean distance with Ward linkage method.

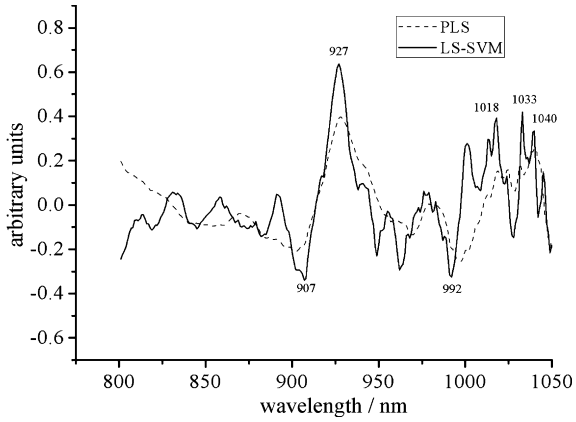

Fig. 3 – Regression coefficients of fat content in milk powder at short-wave NIR spectra.

might enhance the noise, which is an unwilling consequence of derivation. A problem using MSC is that the parameters of offset and slope might be correlated with the response. This may cause removal of information related to the response. This might make the MSC processing appear the worse result. Fig. 1b shows the spectra after the pretreatment by SNV.

3.2. *Sample clustering and selection*

In order to evaluate possible classes among samples, hierarchical clustering was carried out before chemometric analysis. Dendrogram classification using Euclidean distance with Ward linkage (inner squared distance, minimum variance algorithm) was executed to create hierarchical cluster tree upon considering the frequency range between 800 and 1050 nm on the pretreated short-wave NIR spectral data. Fig. 2 shows the dendrographic classification of samples obtained. Seven different types of samples could be identified for a cut-off value of 50 for their heterogeneity. These seven clustering types are corresponding to seven brands, respectively. Then 25 samples were randomly selected from each clustering type for calibration. Thus, 175 samples were chosen randomly as calibration set and establish chemometric models. The remaining 175 samples were used for prediction. Table 2 shows the nutrition concentration range of milk powder from the calibration and validation set.

Table 2 – Concentration range of fat, protein and carbohydrate in milk powder from the calibration and prediction set

Content	Set	Mean	Min	Max	Standard deviation
Fat	Calibration	19.97	15.93	21.80	1.77
	Prediction	19.96	15.88	21.79	1.77
Protein	Calibration	16.91	14.82	18.14	1.18
	Prediction	16.91	14.77	18.14	1.18
Carbohydrate	Calibration	55.15	50.73	60.28	2.86
	Prediction	55.13	50.66	60.27	2.89

3.3. *Fat content analysis*

Sensitive wavelengths for fat analysis in milk powder were proposed based on regression coefficients and loading weights. Two regression coefficients shown in Fig. 3 were obtained from PLS and LS-SVM models, respectively. The shapes of regression coefficients are close for two models. Regression coefficients show distinct maximums at 927 nm, which probably arises from the third overtone of a C–H stretching vibration of fat [20,21]. The peak at 968 nm due to 2v1 + v3 vibration of water is weak, which means the influence of water for fat determination is not strong. The peak at 1018 nm may be assigned to a combination mode, 2 C–H stretching and 3 C–H deformation of the CH_3 groups originating from fat [20,21]. Another peak at 1040 nm was assigned to a 2 C–H stretching + 2 C–H deformation of fat [20,21]. The trough at 992 nm was assigned to second overtone of O–H stretching of fat [16]. Besides the wavelength at 927 nm, 907 nm was supposed to come from third overtone of C–H stretching of proteins [2,20]. The one at 1033 nm was assigned to second overtone of N–H stretching [21]. These two wavelengths are demonstrated as appropriate useful for the fat calibration.

Fig. 4 shows the loading weights and explanation of *X* and *Y* variations obtained from PLS. The optimal number of loading weights was automatically determined by Unscrambler as five from the minimum value of the predicted residual error sum of squares (PRESS) by full cross validation. The percentages of spectral and concentration variances are shown in Fig. 4. About 90% of both spectral variances and concentration variances are accounted by the first two loadings. The first loading weights show distinct maximums at 1044 nm. The second loading weights develop features of fat at 926 nm and water at 967 nm. Besides the features at 928 nm, the third loading weights show another feature at 983 nm which might be second overtone of O–H stretching. The fourth and fifth loading weights develop features at 874, 908, 998 and 1049 nm. Chen et al. reported that there are very weak and unclear indications of fat wavelengths near 880 nm in the second-derivative spec-

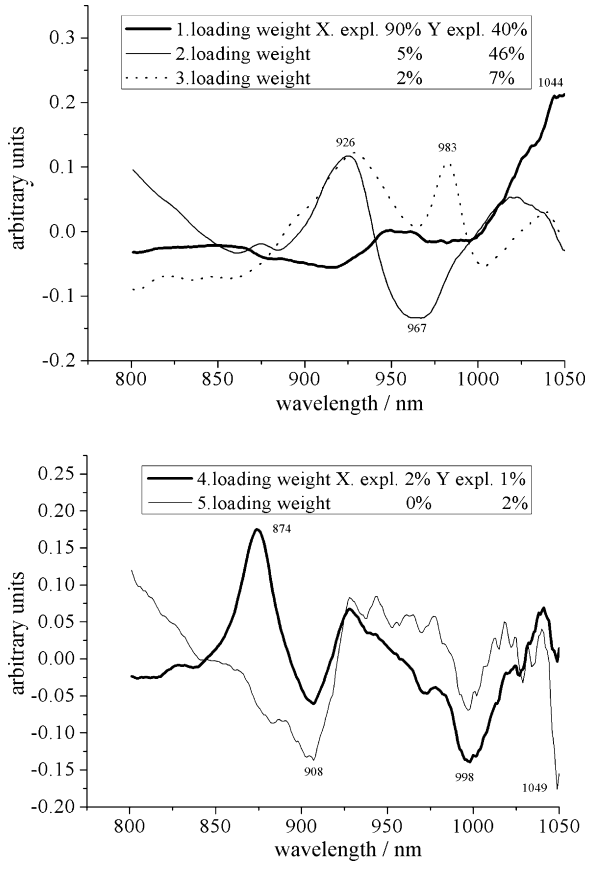

Fig. 4 – First three loading weights (a) and 4–5 loading weights (b) of PLS for fat content at short-wave NIR spectra.

tra of fat [17]. Sasic and Ozaki inferred that the wavelength at 880 nm came from the third overtone of a C–H stretching vibration of fat [4]. However, wavelengths at 874 and 983 nm are almost negligible for the spectral variances as they are concerned with not more than both 10% spectral variances and concentration variances. Thus, we propose six sensitive wavelengths for fat analysis in milk powder, 907, 927, 992, 1018, 1033 and 1040 nm.

In a short-wave NIR study of milk, Sasic et al. found that six regions have some relationship to fat, 840, 880–890, 928, 950–960, 1018, and 1042 nm, and some of these wavelengths were supposed to come from fat–water interaction [4]. Tsenkova et al. also suggested six wavelengths for inexpensive on-line fat sensor, 930, 968, 990, 1026, 1076, and 1092 nm [16]. Although, the short-wave NIR investigations of milk components are not equally comparing to the studies of milk powder components, the proposed wavelengths are similar. The different parts between their and our results are concluded to three reasons. First, milk and milk powder are close but not same materials. Second, some proposed wavelengths are out of our analyzed regions. Third, the results from Tsenkova et al. are obtained only from the peaks of positive regression coefficients but not include the negative peaks. Moreover, they did not analyze the representation of coefficient peaks. For example, 968 nm is a feature peak of water vibration.

Based on the original absorbance values of sensitive wavelengths, the fat content was predicted using two kinds of models, linear regression and LS-SVM. The results of two models are shown in Table 3. When four sensitive wavelengths were set as input of linear regression, the model achieved best performance. The performances were almost same when the number was from four to six. The performance was unacceptable when there were only two sensitive wavelengths. The RPD values of all the models were less than three.

LS-SVM models were also established based on these sensitive wavelengths without pretreatment. Table 3 shows that LS-SVM process has not much improved the mode. Linear regression model may be a better choice compared to LS-SVM for the development of a simple, low cost and efficacious instrument.

Moreover, linear regression and LS-SVM models were established based on the sensitive wavelengths after spectral pretreatment of SNV. Linear functions are shown as follow:

$$Y_{fat} = -6.05 - 4.38\lambda_{907} + 11.24\lambda_{927} + 3.80\lambda_{992} + 8.59_{1018} + 6.74_{1033} + 11.04_{1040} \quad (1)$$

$$Y_{fat} = -7.14 + 1.82\lambda_{907} + 10.36\lambda_{927} + 4.16\lambda_{992} + 8.99\lambda_{1018} \quad (2)$$

$$Y_{fat} = 1.48 + 7.72\lambda_{928} + 8.46\lambda_{1018} \quad (3)$$

where Y_{fat}, predicted content of fat; λ_i, absorbance value after SNV at i nm.

As shown in Table 3, SNV pretreatment did not much improve the performance of linear regression models except the number of sensitive wavelengths is two, but greatly improved the performance of LS-SVM model. Based on four sensitive wavelengths, the RPD value of LS-SVM model was up to seven, which is considered adequate for process control. The RPD valued of the LS-SVM model with only two sensitive wavelengths has achieved as five, which is adequate for quality control [22]. So, less sensitive wavelengths input could be achieved through LS-SVM process with spectral pretreatment. Fig. 5 shows the fat content predictive result of LS-SVM model based on four sensitive wavelengths with spectral pretreatment.

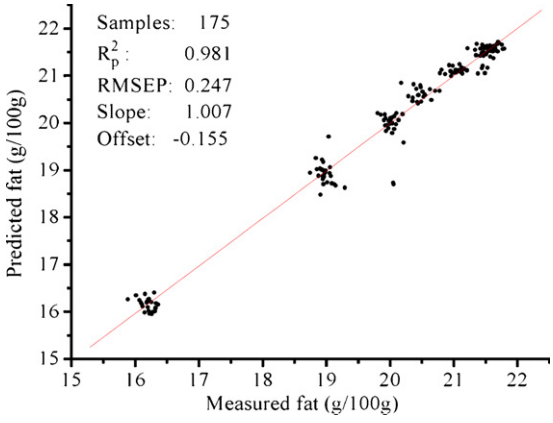

Fig. 5 – Correlation between measured and NIR predicted values for milk powder fat content of LS-SVM model based on four sensitive wavelengths with spectral pretreatment (907, 927, 992 and 1018 nm).

Table 3 – Results for fat content determination in milk powder based on sensitive wavelengths in the short-wave NIR spectra

Wavelengths (nm)	Method	R_c^2	RMSEC	R_p^2	RMSEP	Bias[c]	Slope[c]	Offset[c]	RPD
907, 927, 992, 1018, 1033 and 1040	Linear regression[a]	0.891	0.583	0.869	0.641	−0.018	0.895	2.109	2.759
	LS-SVM[a]	0.928	0.472	0.914	0.518	−0.048	0.903	1.977	3.426
907, 927, 992 and 1018	Linear regression[a]	0.888	0.591	0.870	0.637	−0.014	0.894	2.126	2.776
	LS-SVM[a]	0.926	0.429	0.918	0.506	−0.035	0.904	1.947	3.504
907 and 927	Linear regression[a]	0.678	1.002	0.730	0.925	0.058	0.683	6.257	1.916
	LS-SVM[a]	0.770	0.847	0.769	0.855	0.077	0.753	5.241	2.076
907, 927, 992, 1018, 1033 and 1040	Linear regression[b]	0.886	0.597	0.908	0.548	−0.073	0.858	2.908	3.254
	LS-SVM[b]	0.992	0.151	0.971	0.301	0.027	0.986	0.242	5.889
907, 927, 992 and 1018	Linear regression[b]	0.879	0.615	0.900	0.573	−0.091	0.852	3.038	3.127
	LS-SVM[b]	0.996	0.106	0.981	0.247	0.009	1.007	−0.155	7.162
927 and 1018	Linear regression[b]	0.858	0.664	0.870	0.644	−0.070	0.836	3.339	2.763
	LS-SVM[b]	0.983	0.225	0.965	0.335	0.050	0.990	0.134	5.342
1018	Linear regression[b]	0.477	1.277	0.473	1.284	0.040	0.464	10.660	0.424
	LS-SVM[b]	0.820	0.750	0.723	0.938	0.063	0.770	4.506	1.891

[a] With pretreatment.
[b] With out pretreatment.
[c] Prediction set.

Furthermore, when two sensitive wavelengths were selected, there are different for the models with and without pretreatment. This is because that the regression coefficients obtained here were based on the spectra data after pretreatment, and the contribution of original spectra at 1018 nm is low. The useful information of this wavelength needs to be mined out after SNV pretreatment. The spectral data is suggested to be pretreated first, and the results based on spectra pretreatment could be better, as shown in Table 3.

In order to evaluate the determination ability of short-wave NIR spectroscopy for fat content in milk powder, the results obtained here were compared with those of other spectroscopy techniques [9,11,23,24]. Wu et al. used the whole long-wave NIR spectral data and whole mid-infrared spectral data to predict fat content in milk powder [9]. In that paper, R_p^2 of LS-SVM was 0.947 and RMSEP was 2.435 using the whole long-wave NIR spectra. R_p^2 was lower than the results based on two to six sensitive wavelengths. R_p^2 was 0.956 and RMSEP was 1.667 using the whole mid-infrared spectra. R_p^2 was lower and RMSEP value is higher than ours. Moros et al. evaluated fat content in infant formulas powdered milk by Raman spectroscopy [11]. In this study, there were only 15 samples for calibration and 8 for prediction. Correlation coefficient of calibration (R_c) was 0.997, RMSEC was 0.7, RMSEP was 1.8. R_p^2 was not given. R_c is equal to result here, while both RMSEC and RMSEP are higher than those here. Barabássy used two long-wave NIR instrument to predict fat content in milk powder [23]. R_c were 0.977 and 0.986, and standard error of calibration were 0.125 and 0.128, respectively for two instruments. However, the results of prediction set were not given. The results of calibration set are close to ours. Thus, the obtained sensitive wavelengths have a high accuracy performance for the fat content determination of milk powder.

3.4. *Protein content analysis*

Sensitive wavelengths for protein analysis in milk powder were proposed based on regression coefficients and loading weights. Fig. 6 shows the regression coefficients for the PLS and LS-SVM model. The curve of LS-SVM shows distinct maximums at 800, 904, 949, 986 and 1002 nm. Eight hundred nanometers was not chosen as it is at the edge of wavelength. Nine hundred and four nanometers probably arises from the third overtone of a C–H stretching vibration of protein [20,21]. The regression coefficient at 968 nm is approximately 0, which means the influence of water for protein determination is weak. The wavelength at 1002 nm might result from second overtone of O–H stretching of water interacting with protein or second overtone of O–H stretching vibration of proteins [4]. The peak at 949 nm probably arises from the second overtone of O–H stretching of water interacting with protein [4]. For

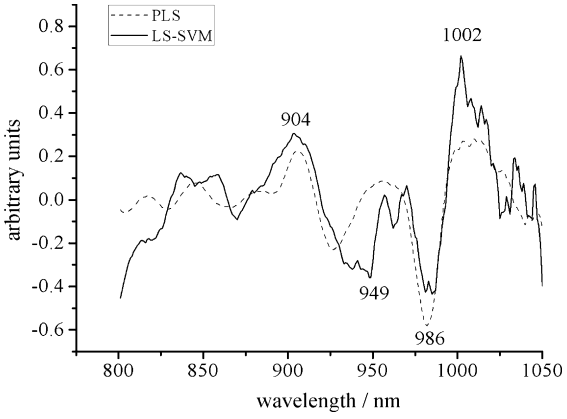

Fig. 6 – Regression coefficients of protein content in milk powder at short-wave NIR spectra.

the wavelength at 986 nm, it was identified from second overtone of O–H stretching of protein. In addition, wavelength at 1030 nm is due to a second overtone of their N–H stretching vibration, and that at 1020 nm arises from a 2 N–H stretching + 2 amide I. These two sensitive wavelengths are related to protein of milk [20,21], however, this relation is not obvious in present study. Compared to LS-SVM, the peaks of PLS are mostly at the same position except wavelength at 925 and 949 nm. The peak 949 nm is weak for PLS while the peak 925 nm has weak feature. The wavelengths below 900 nm have not significant influence on the protein prediction, and this result is consistent with the result of Sasic and Ozaki [4].

Fig. 7 shows the loading weights and explanation of *X* and *Y* variations obtained from PLS. The optimal number of loading weights was automatically determined by Unscrambler as six from the minimum value of PRESS by full cross validation. Most of the *X* variation is explained by first loading weights while the explanation of *Y* variation is not rich. On the contrast, second loading weights carried most information of *Y* variation. Thus, first loading weights and second loading weights are the most important loading weights for the protein determination of milk powder. The first loading weights show distinct maximums at 983 and 1044 nm. The maximum at 1044 nm was assigned to fat [20,21]. The second loading weights show distinct maximums at 800, 980 and 1050 nm. However, 800 and 1050 nm are at the edge of scale, and are not suggested as sensitive wavelengths of protein. The third loading weights show distinct maximums at 927, 967, 1000 and 1040 nm. The wavelengths at 927 and 967 nm have been respectively assigned to fat and water analysis and might be negligible for protein analysis. The forth to sixth loading weights develop features similar to the first third ones except the feature around 874 nm. The wavelength at 874 nm can be negligible as it is concerned with little spectral variances and not more than 10% concentration variances. Finally, we proposed four sensitive wavelengths, 904, 949, 987 and 1002 nm, for the protein content determination.

Moreover, several "sensitive" wavelengths for protein analysis were assigned to other component such as fat and water. It is likely that the concentration change of one component, which has important spectral contributions, leads the whole spectral variation. This component might affect the calibration of other components with minor spectral contributions.

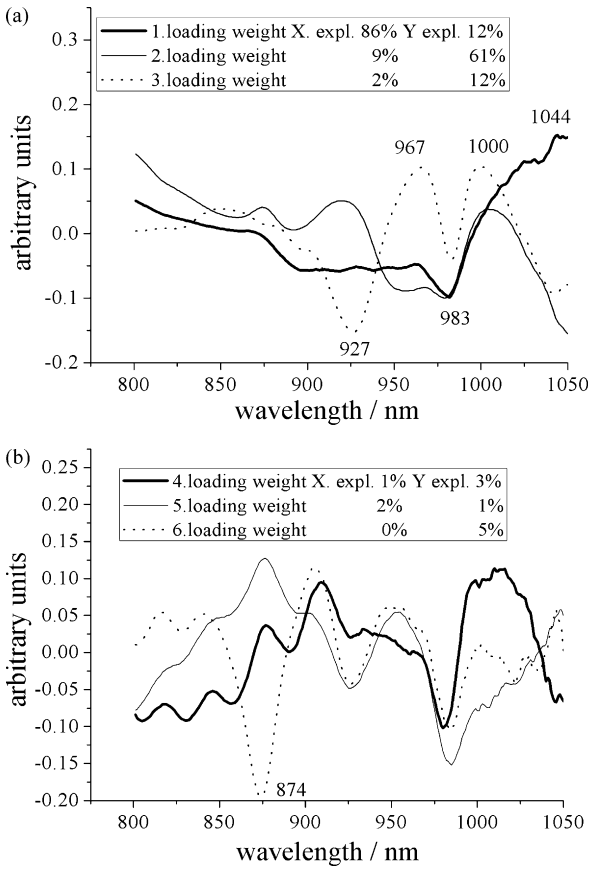

Fig. 7 – First three loading weights (a) and 4–6 loading weights (b) of PLS for protein content at short-wave NIR spectra.

The protein content was then predicted based on linear regression and LS-SVM. Models established from original absorbance values are improved through SNV pretreatment. Results of linear regression and LS-SVM models after SNV pretreatment are shown in Table 4. Linear functions are shown as follow as:

$$Y_{protein} = 15.94 - 1.02\lambda_{904} - 0.40\lambda_{949} - 7.48\lambda_{987} + 6.59\lambda_{1002} \quad (4)$$

Table 4 – Results for protein content determination in milk powder based on sensitive wavelengths in the short-wave NIR spectra with pretreatment

Wavelengths (nm)	Method	R_c^2	RMSEC	R_p^2	RMSEP	Bias[a]	Slope[a]	Offset[a]	RPD
906, 949, 987 and 1002	Linear regression	0.854	0.448	0.865	0.433	−0.013	0.848	2.074	2.720
	LS-SVM	0.994	0.091	0.984	0.148	−0.008	0.977	0.386	7.964
906, 987 and 1002	Linear regression	0.854	0.448	0.864	0.434	−0.014	0.877	2.093	2.711
	LS-SVM	0.996	0.069	0.977	0.177	−0.009	0.977	0.386	6.648
987 and 1002	Linear regression	0.812	0.509	0.822	0.497	−0.017	0.830	2.882	2.370
	LS-SVM	0.982	0.156	0.964	0.224	−0.028	0.961	0.675	5.299
987	Linear regression	0.492	0.837	0.469	0.864	−0.020	0.531	7.950	1.362
	LS-SVM	0.667	0.678	0.697	0.649	0.027	0.673	5.497	1.814

[a] Prediction set.

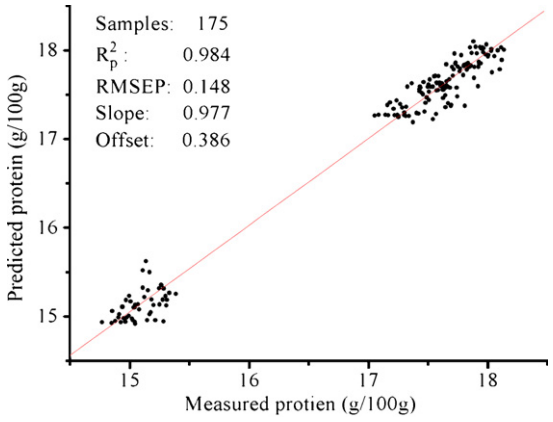

Fig. 8 – Correlation between measured and short-wave NIR predicted values for milk powder protein content of LS-SVM model based on four sensitive wavelengths with spectral pretreatment (904, 947, 987 and 1002 nm).

$$Y_{protein} = 16.13 - 0.84\lambda_{906} - 7.82\lambda_{987} + 6.81\lambda_{1002} \quad (5)$$

$$Y_{protein} = 16.86 - 8.11\lambda_{987} + 6.40\lambda_{1002} \quad (6)$$

$$Y_{protein} = 25.82 - 6.93\lambda_{987} \quad (7)$$

where $Y_{protein}$, predicted content of protein, λ_i, absorbance value after SNV at i nm.

The performance of linear regression model with SNV pretreatment has been slightly improved, while the performance of LS-SVM model has been greatly improved. Based on four sensitive wavelengths, the RPD value of LS-SVM model was close to eight, which is adequate for process control. The RPD value of the LS-SVM model with only two sensitive wavelengths was up to five, which is adequate for quality control [22]. Fig. 8 shows the protein content predictive result of LS-SVM model based on four sensitive wavelengths with spectral pretreatment.

In order to evaluate the determination ability of short-wave NIR spectroscopy for protein content in milk powder, the results obtained here were compared with those of other spectroscopy techniques [10,11,23–25]. Wu et al. used the whole long-wave NIR and whole mid-infrared spectral data respectively to predict protein content in milk powder [10]. In that paper, R_p^2 was 0.966 and RMSEP was 0.547 using the whole long-wave NIR spectra. R_p^2 is similar to the results here based on two sensitive wavelengths, while RMSEP is higher. R_p^2 was 0.990 and RMSEP was 0.294 using the whole mid-infrared spectra. The RMSEP value is also little higher than ours. Moros et al. evaluated protein content in infant formulas powdered milk by Raman spectroscopy [11]. R_c of 15 samples was 0.98, RMSEC was 1.5, RMSEP was 1.3 on 8 samples. R_p^2 was not given. Both RMSEC and RMSEP are higher than those here. Barabássy used two long-wave NIR instruments to predict protein content in milk powder [23]. R_c were 0.933 and 0.997, and standard error of calibration were 4.27 and 1.89, respectively for two instruments. The results of calibration set of later instrument are better than those of first one, and are close to ours. However, the results of prediction set were not given. Thus, the obtained sensitive wavelengths have the same predicted ability to whole long-wave NIR spectra, mid-infrared spectra and Raman spectroscopy for the protein content determination of milk powder.

3.5. *Carbohydrate content analysis*

Sensitive wavelengths for carbohydrate analysis in milk powder were proposed based on regression coefficients and loading weights. Regression coefficients of two models of PLS and LS-SVM are similar (Fig. 9). Distinct maximums appear at 800, 835, 861, 897, 919, 945, 958, 982 and 1002 nm. Very noisy regression coefficients appear at the region beyond 1015 nm. The short-wave NIR spectra up than 1015 nm are not sensitive for carbohydrate of milk powder. Eight hundred nanometers was not chosen as it is at the edge of wavelength. Eight hundred and thirty-five nanometers was supposed to come from third combination overtone of OH stretching, which might be the feature of carbohydrate [26]. The assignment for 861 and 897 nm was assigned to third overtone of C–H stretching. The wavelength at 919 nm results from third overtone of CH stretching of some carbohydrate [26]. The peak at 945 nm was probably ascribed to the third overtone of CH_2 stretching in sugars [15,27]. The wavelength at 958 nm arises from a second overtone of OH stretch of some carbohydrate [26]. Nine hundred and eighty-two nanometers was ascribed to second overtone of O–H stretching. The regression coefficients plot of lactose in milk from Sasic et al. was noisy through the whole short-wave NIR spectra at 900–1075 nm region [4]. The performance of regression coefficient curves at 800–1015 nm in our study is slightly better than Sasic and Ozaki [4].

Fig. 10 shows the loading weights and explanation of X and Y variations obtained from PLS model. The optimal number of loading weights was automatically determined by Unscrambler as eight from the minimum value of PRESS by full cross validation. By the first two loadings, more than 90% of spectral variances and 70% of concentration variances are accounted. The first loading weights show distinct maximums at 1044 nm. Second loading carried more than half information of concentration variance and its feature wavelengths focus at 925, 958 and 1007 nm. These peaks are smooth. The wavelengths

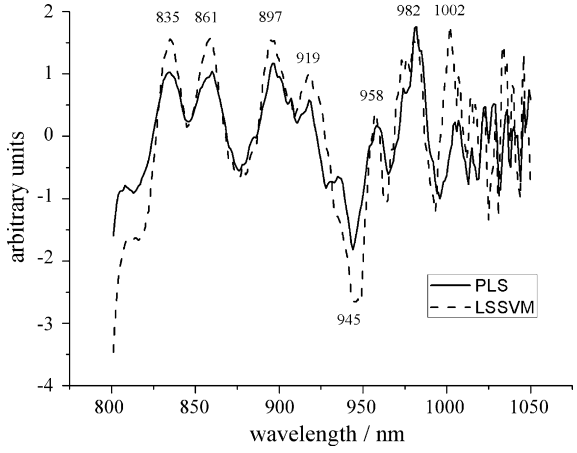

Fig. 9 – Regression coefficients of carbohydrate content in milk powder at short-wave NIR spectra.

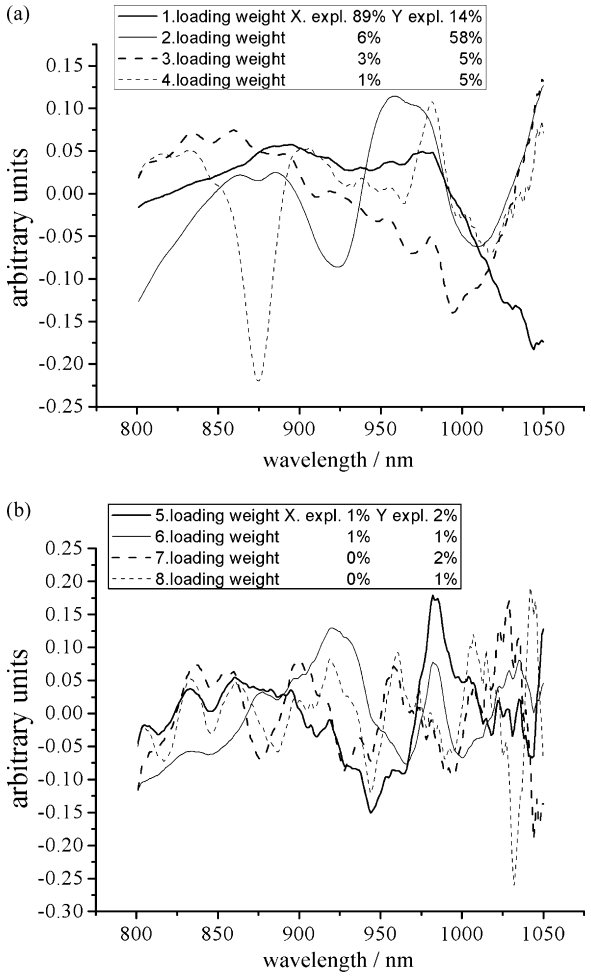

Fig. 10 – First fourth loading weights (a) and 5–8 loading weights (b) of PLS for carbohydrate content at short-wave NIR spectra.

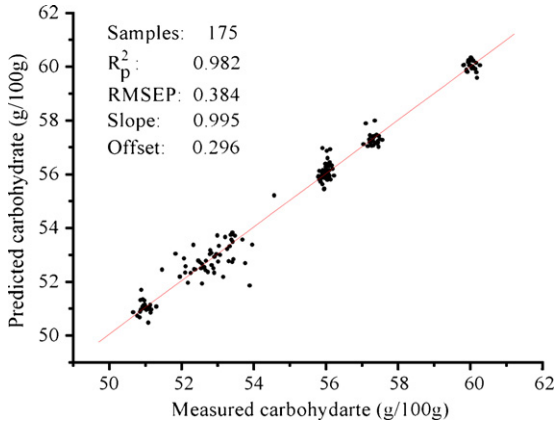

Fig. 11 – Correlation between measured and short-wave NIR predicted values for milk powder carbohydrate content of LS-SVM model based on seven sensitive wavelengths with spectral pretreatment (835, 861, 897, 919, 945, 958 and 982 nm).

at 925 and 1007 nm have been assigned to fat. Other loading weights that reveal the spectral variation in milk powder is weakly dependent on the carbohydrate content. Finally, we proposed seven sensitive wavelengths for carbohydrate content determination, 835, 861, 897, 919, 945, 958, and 982 nm.

The models were established based on sensitive wavelengths for carbohydrate content determination. Linear functions are shown as follow as:

$$Y_{carbo} = 71.17 - 1.82\lambda_{835} + 11.49\lambda_{861} + 17.88\lambda_{897} - 10.44\lambda_{919} + 9.41\lambda_{945} - 5.32\lambda_{958} + 7.18\lambda_{982} \quad (8)$$

$$Y_{carbo} = 72.66 + 11.87\lambda_{861} + 15.43\lambda_{897} - 9.25\lambda_{919} + 11.23\lambda_{945} - 5.13\lambda_{958} + 7.18\lambda_{982} \quad (9)$$

$$Y_{carbo} = 61.00 + 16.33\lambda_{861} + 13.69\lambda_{982} \quad (10)$$

where Y_{carbo}, predicted content of carbohydrate, λ_i, absorbance value after SNV at i nm.

The models were improved after SNV pretreatment compared to those established directly from original spectra. The results of linear regression and LS-SVM models after SNV pretreatment are shown in Table 5. The best performances were obtained based on six or seven wavelengths. Result of LS-SVM model can be accepted when the number of sensitive wavelengths is two. RPD values based on six or seven wavelengths are adequate for process control of carbohydrate content in milk powder. Fig. 11 shows the protein content predictive result of LS-SVM model based on four sensitive wavelengths with spectral pretreatment.

Table 5 – Results for carbohydrate content determination in milk powder based on sensitive wavelengths in the short-wave NIR spectra with pretreatment

Wavelengths (nm)	Method	R_c^2	RMSEC	R_p^2	RMSEP	Bias[a]	Slope[a]	Offset[a]	RPD
835, 861, 897, 919, 945, 958, and 982	Linear regression	0.781	1.346	0.794	1.309	0.065	0.776	12.250	2.205
	LS-SVM	0.996	0.182	0.982	0.384	0.025	0.995	0.296	7.519
861, 897, 919, 945, 958, and 982	Linear regression	0.780	0.348	0.792	1.317	0.080	0.778	12.150	2.194
	LS-SVM	0.996	0.179	0.982	0.384	−0.004	0.992	0.401	7.514
861 and 982	Linear regression	0.695	1.587	0.682	1.625	−0.034	0.676	17.880	1.775
	LS-SVM	0.947	0.662	0.910	0.863	−0.003	0.916	4.608	3.340

[a] Prediction set.

Table 6 – Proposed wavelengths assignment of fat, proteins and carbohydrate observed in loading weights and regression coefficients of PLS and LS-SVM of short-wave NIR spectra of milk powder

Wavelength position (nm)	Component	Assignment
835	Carbohydrate	Third combination overtone of O–H stretching
861 and 897	Carbohydrate	Third overtone of C–H stretching vibration
904	Protein	Third overtone of C–H stretching vibration
919	Carbohydrate	Third overtone of C–H stretching vibration
927	Fat	Third overtone of C–H stretching vibration
945	Carbohydrate	Third overtone of $C–H_2$ stretching in sugars?
949	Protein	Third overtone of $C–H_2$ stretching of carbohydrate overlapped with fat–water interaction
958	Carbohydrate	Second overtone of O–H stretch
968	Water	Second overtone of O–H stretching of water
982	Carbohydrate	Second overtone of O–H stretching
987	Protein	Second overtone of O–H stretching
992	Fat	Second overtone of O–H stretching
1002	Protein	Second overtone of O–H stretching of water interacting with protein or second overtone of O–H stretching vibration of proteins
1018	Fat	2 C–H stretching and 3 C–H deformation of the $C–H_3$ groups originating from fat
1033	Fat	Second overtone of N–H stretching
1040	Fat	2 C–H stretching + 2 C–H deformation

In order to evaluate the determination ability of short-wave NIR spectroscopy for carbohydrate content in milk powder, the results obtained here were compared with those of other spectroscopy techniques [11,23,24]. Moros et al. evaluated carbohydrate content in infant formulas powdered milk by Raman spectroscopy [11]. R_c of 15 samples was 0.993, RMSEC was 0.7, RMSEP was 2 on 8 samples. R_p^2 was not given. Both RMSEC and RMSEP are higher than those here. Barabássy used two long-wave NIR instruments to predict Lactose content in milk powder [23]. R_c were 0.929 and 0.997, and standard error of calibration were 6.79 and 1.89, respectively for two instruments. The results of calibration set of later instrument are better than those of first one. However, the results of prediction set were not given. Thus, the obtained sensitive wavelengths have the same predicted ability to whole Raman spectra and long-wave NIR spectra for the carbohydrate content determination of milk powder.

3.6. *Wavelength assignment in the short-wave NIR region*

Through quantitative analysis of fats, proteins and carbohydrates, the feasibility is further demonstrated for the simultaneous measurement of major components in milk powder using short-wave NIR spectra. The interesting wavelength regions are screened and the assignment of all interesting specific wavelengths is proposed in the details associated with chemical base. Table 6 shows the results of wavelength assignment. Some wavelengths have close relationship with some components, but have less contribution for the component determination. Some other important wavelengths need to be further assigned. Because of the increase in interest for using the short-wave NIR region for milk powder analysis, the assignment results in Table 6 may be valuable.

4. Conclusions

The present study has provided insight into the short-wave NIR region, applied newly for the emerging field of milk powder. On the basis of the regression coefficients and loading weights, we have proposed wavelength assignment in the short-wave NIR region which is newly and very useful for the research and application of milk powder. LS-SVM models of fat, proteins and carbohydrate in milk powder using the assigned wavelengths in the short-wave NIR region shows a reliable results. The predicted abilities for three main components in milk powder based on assigned wavelengths are better than the whole long-wave NIR spectroscopy, mid-infrared spectroscopy and Raman spectroscopy. It was found out that SNV was a use full pre-treatment for raw milk powder spectra. At short-wave NIR spectral region, single-beam data of proposed assigned wavelength can be reliably exploited. With its high transmittance ability, low cost, and easy to be handheld, wavelength assigned short-wave NIR spectra could be the excellent detector for the milk powder analysis and for the industrial application. The method and results are also possible to be referred for other biological material measurement using short-wave NIR spectra.

Acknowledgements

This study was supported by the Teaching and Research Award Program for Outstanding Young Teachers in Higher Education Institutions of MOE, PRC, Natural Science Foundation of China (Project No: 30671213, 60605011), Specialized Research Fund for the Doctoral Program of Higher Education (Project No: 20040335034) and Natural Science Foundation of Zhejiang Province, China (Project No: RC02067).

REFERENCES

[1] A. Bittner, R. Marbach, H.M. Heise, J. Mol. Struct. 349 (1995) 341.
[2] J.B. Reeves III, J. Near Infrared Spectrosc. 2 (1994) 199.
[3] D.M. Mayes, J.B. Callis, Appl Spectrosc. 43 (1989) 27.
[4] S. Sasic, Y. Ozaki, Anal. Chem. 73 (2001) 64.
[5] Y. He, D. Wu, S. Feng, Int. J. Food Prop. 10 (2007) 1.
[6] P.P. Subedi, K.B. Walsh, G. Owens, Postharvest Biol. Technol. 43 (2007) 326.
[7] H. Kamishikiryo-Yamashita, Y. Oritani, H. Takamura, T. Matoba, J. Food Sci. 59 (1994) 313.
[8] A. Borin, M.F. Ferrao, C. Mello, D.A. Maretto, R.J. Poppi, Anal. Chim. Acta 579 (2006) 25.
[9] D. Wu, S. Feng, Y. He, J. Dairy Sci. 90 (2007) 3613.
[10] D. Wu, Y. He, S. Feng, D.W. Sun, J. Food Eng. 84 (2008) 124.
[11] J. Moros, S. Garrigues, M. de la Guardia, Anal. Chim. Acta 593 (2007) 30.
[12] M.S. Dhanoa, S.J. Lister, R. Sanderson, R.J. Barnes, J. Near Infrared Spectrosc. 2 (1994) 43.
[13] R.J. Barnes, M.S. Dhanoa, S.J. Lister, J. Near Infrared Spectrosc. 1 (1993) 185.
[14] T. Isaksson, T. Naes, Appl. Spectrosc. 42 (1988) 1273.
[15] B.G. Osborne, T. Fearn, P.H. Hindle, Practical NIR Spectroscopy with Applications in Food and Beverage Analysis, Longman Singapore Publ. Ltd., Singapore, 1993.
[16] R. Tsenkova, S. Atanassova, K. Toyoda, Y. Ozaki, K. Itoh, T. Fearn, J. Dairy Sci. 82 (1999) 2344.
[17] Y.J. Chen, C. Iyo, S. Kawano, J. Near Infrared Spectrosc. 7 (1999) 265.
[18] P.A. Gorry, Anal. Chem. 62 (1990) 570.
[19] A. Candolfi, R. De Maesschalck, D. Jouan-Rimbaud, P.A. Hailey, D.L. Massart, J. Pharm. Biomed. Anal. 21 (1999) 115.
[20] B.G. Osborne, T. Fearn, Near Infrared Spectroscopy in Food Analysis, Longman Scientific and Technical, Essex, UK, 1986.
[21] J.J. Workman Jr., Appl. Spectrosc. Rev. 31 (1996) 251.
[22] P. Williams, K. Norris, Near-Infrared Technology in the Agricultural and Food Industries, 2nd ed., American Association of Cereal Chemists, St. Paul, MN, 2001, p. 145.
[23] S. Barabássy, Mljekarstvo 51 (2001) 263.
[24] M. Chang, P. Chu, K. Xu, Spectrosc. Spectr. Anal. 27 (2007) 43.
[25] J.L. Rodriguez-Otero, M. Hermida, J. Centeno, Agric. Food Chem. 45 (1997) 2815.
[26] M. Golic, K.B. Walsh, P. Lawson, Appl. Spectrosc. 57 (2003) 139.
[27] I. Murray, P.C. Williams, Near-Infrared Technology in the Agricultural and Food Industries, American Association of Cereal Chemists, Inc., Minnesota, 1987.