COMMUNICATION

Determination of Calcium Content in Powdered Milk Using Near and Mid-Infrared Spectroscopy with Variable Selection and Chemometrics

Di Wu · Pengcheng Nie · Yong He · Yidan Bao

Abstract Near infrared (NIR) and mid-infrared (MIR) spectroscopy techniques were evaluated to determine calcium content in powdered milk. A hybrid spectral variable selection algorithm combined with uninformation variable elimination (UVE) and successive projections algorithm (SPA) selected 11 NIR and 15 MIR variables from full 2,756 NIR and 3,727 MIR variables, respectively. Predicted results of least-squares support vector machine models for the samples in the prediction set show that the 15 MIR variables obtained much better results (0.930 for coefficient of determination (r^2), 3.703 for residual predictive deviation (RPD), 30.162 for root mean square error of prediction set (RMSEP) and 5.22% for relative errors of prediction (RSEP)) than 11 NIR variables did (0.636 for r^2, 1.587 for RPD, 78.815 for RMSEP, and 13.40% for RSEP). The overall results indicate that MIR spectroscopy could be applied as a precision and rapid method to determine calcium content in powdered milk. The good performance shows a potential application using UVE-SPA to select NIR and MIR effective variables.

D. Wu · P. Nie · Y. He (✉) · Y. Bao
College of Biosystems Engineering and Food Science,
Zhejiang University,
388 Yuhangtang Road,
Hangzhou, Zhejiang 310058, People's Republic of China
e-mail: yhe@zju.edu.cn

P. Nie
College of Electronic and information Engineering,
Nanchang Hangkong University,
Nanchang 330069, China

Introduction

Powdered milk is a manufactured dairy product made by evaporating milk to dryness. Powdered milk is high in minerals like calcium and can be used as ingredients to fortify other manufactured food products that are poor in calcium. The role of Ca^{2+} as an intracellular messenger has been recognized since the 1970s (Berridge 1975). Calcium plays a critical role in maintaining bone health and preventing osteoporosis. Calcium is also crucial for nerve conduction, muscle contraction, heartbeat, blood coagulation, the production of energy, and maintenance of immune function. So, a fast, non-destructive and accurate quality control of calcium content in powdered milk is of great importance, not only for establishing their nutritional values but also for consumer information. Some chemistry methods were applied for calcium content determination in powdered milk, such as reaction cell inductively coupled plasma mass spectrometry (Chen and Jiang 2002), energy-dispersive X-ray fluorescence spectrometry (Alvarez et al. 2005), and ion-selective electrode (Hitti and Thomas 1985). However, these methods have some disadvantages such as being destructive, time-consuming, laborious, costly, and they require complex and professional laboratory operations.

Nowadays, near infrared (NIR) and mid-infrared (MIR) spectroscopy technologies are employed as alternatives to wet chemistry procedures for quantitative analysis in agriculture and food quality evaluation (Wilson and Tapp 1999; Nicolai et al. 2007; Woodcock et al. 2008; Alexandrakis et al. 2009; Hu et al. 2010). For quality evaluation of powdered milk, we did analyses on the content prediction of protein, fat, and carbohydrate using short-wave NIR, NIR, and MIR spectroscopy (Wu et al. 2007, 2008a; Wu et al. 2008b, c). Borin et al. (2006)

studied the common adulterants in powdered milk by near infrared spectroscopy. There are some works on calcium study in organic and food using infrared spectroscopy. Byler and Farrell (1989) found out about infrared spectroscopic evidence for calcium ion interaction with carboxylate groups of casein. In the absence of Ca^{2+}, the O–C–O stretching vibrations of these carboxylates give two characteristic infrared absorption bands near 1,400 and 1,575 cm^{-1}. When Ca^{2+} ions are present, this pair of bands was observed to shift about 10 cm^{-1} to approximately 1,410 and 1,565 cm^{-1}, respectively. Upreti and Metzger (2006) found that the spectral region ~980 cm^{-1} had the highest correlation with the measured concentrations of bound Ca in cheese. Mesubi (1982) did infrared study of zinc, cadmium, and lead salts of some fatty acids. From the O–C–O stretching frequencies Mesubi deduced that the coordination of the carboxylate group to the metal ions is unsymmetric chelating bidentate. Nara and Tanokura (2008) found out the assignments of the O–C–O anti-symmetric stretch have been ensured on the basis of the spectra of calcium-binding peptide analogs. The downshift of the O–C–O anti-symmetric stretching mode from 1,565 cm^{-1} to 1,555–1,540 cm^{-1} upon Ca^{2+} binding is a commonly observed feature of MIR spectra for EF-hand proteins. However, to the best of our knowledge, there were few reports on the quantification determination of calcium content in powdered milk by using NIR or MIR spectroscopy.

Due to the characteristics of overtone and combination, the molar absorbance of NIRS is low (low sensitivity). However, NIRS can be operated with little sample preparation which makes the whole operation more convenient and online possible. The NIR instruments are cheaper than MIR. Thus, two techniques have their advantages and disadvantages and need to be considered in different applications. In this study, NIR and MIR spectroscopy techniques were compared to evaluate which one was better to determine calcium content in powdered milk.

As it is broad and weak due to combinations and overtones of functional groups, NIR is largely dependent on chemometric methods for the quantitative analysis of multicomponent systems or mixtures (Bokobza 1998). Chemometrics is also useful to rapid extract relevant information from complex MIR data sets (Downey 1998). However, with hundreds or thousands of wavenumbers as input variables and hundreds of samples, both NIR and MIR spectral data are too complicated to be trained directly in the chemometric models. The training procedure is time-consuming and not convenient. Some wavenumbers or wavenumber bands may contain useless or irrelevant information for calibration model-like noise and background than relevant information to models. Moreover, these wavenumbers or wavenumber bands may contribute

more collinearity and redundancies (Ye et al. 2008). The elimination of uninformation variables can predigest calibration modeling and improve prediction results in terms of accuracy and robustness. Moreover, the selected effective wavenumber variables can be used to develop simple instruments or online detection and control of powdered milk during the produce stage.

However, selecting variables from the full spectrum to obtain the maximum accuracy is still a challenging task, especially when spectra display strong overlapping, include mass and complex data, and have imperceptible distinctive features, as is the case with NIR and MIR spectroscopy. The commonly used variable selection methods include generalized simulated annealing (Kalivas et al. 1989), genetic algorithm (GA) (Jouanrimbaud et al. 1995), x-loading weights (Wu et al. 2008b), and regression coefficients (Wu et al. 2008b). However, some selection methods (SA and GA for instance) have stochastic nature, results are realization dependent and variable selections may not be reproducible. Other methods introduced artificial judgments which are time-consuming and not precise. In this work, a hybrid spectral variable selection algorithm, uninformation variable elimination (UVE) by PLS with successive projections algorithms (SPA), was applied for both NIR and MIR spectral variable selection. Both UVE and SPA are deterministic algorithms, which do not employ stochastic operations. There are some papers on using UVE–SPA to do NIR spectral variable selection (Ye et al. 2008; Wu et al. 2009b), but there are a few studies focused on the MIR variable selection using UVE–SPA and comparing UVE–SPA on NIR and MIR variable selection.

The specific aims of this study were: (1) to investigate the feasibility of using NIR and MIR spectroscopy to predict the calcium content in powdered milk; (2) to compare the prediction ability of NIR and MIR spectra based on least-square support vector machine (LS-SVM); (3) to reduce the input variables of LS-SVM models and obtain effective wavenumber variables by variable selection method of UVE-SPA; and (4) to compare the prediction abilities of LS-SVM models based on different inputs, namely, variables selected by UVE-SPA and only selected by UVE, and the full spectral variables.

Materials and Methods

System Setup and Data Measurement

In this study, the JASCO model FT/IR-4000 Fourier transform infrared spectrometer (Japan) was used, with a valid range of 7,800~350 cm^{-1} and auto scan speed of 2.0 mm/s, to obtain transmittance (percent transmission)

spectrum. In order to make the powder have different calcium contents, seven brands of powdered milk were provided by the Inner Mongolia Yili Industrial Group Co., Ltd., a well-known corporation in the Chinese powdered milk market. They were 0–6-month infant powdered milk, 6–12-month infant powdered milk, 1–3-year toddler powdered milk, 3–6-year child powdered milk, student powdered milk with zincum, woman nutrition powdered milk and whole powdered milk. The temperature was kept at about 25°C during the whole experiment. Powdered milk sample was mixed with KBr at the ratio of 1:49. Each brand included 52 to 60 samples, respectively (Table 1). Each sample was scanned 40 times and the data were averaged to one as the spectra of this sample. Finally the spectra of 409 samples were obtained. Due to the potential system imperfection, obvious scattering noises were observed at the beginning and end of the spectral data, and therefore should be eliminated. Finally the NIRS analysis was done based on 4,000 to 6,666 cm^{-1} while MIRS analysis was done between 400 and 4,000 cm^{-1}. The smoothing algorithm of Savitzky–Golay with three segments and zero polynomials combined with standard normal variate were applied as the spectral pretreatment. The pretreatment processes were implemented by "The Unscrambler V 9.7" (CAMO Process AS, OSLO, Norway). The reference values of calcium were determined by atomic absorption spectrometry using PerkinElmer AAnalyst™ 800 high-performance atomic absorption spectrometer. The absorbed atomic resonance line was 422.7 nm according to Official Method 965.17 (AOAC 2000).

Successive Projections Algorithms

SPA is a forward variable selection algorithm for multivariate calibration to select wavenumbers which their information content is minimally redundant, in order to solve the collinearity problems (Araujo et al. 2001). SPA performs simple projection operations in a vector space to obtain subsets of useful variables with minimum of collinearity

(Araujo et al. 2001). The principle of variable selection by SPA is that the new variable selected is the one among all the remaining variables, which has the maximum projection value on the orthogonal subspace of the previous selected variable. The detailed description of SPA can be seen in Araujo et al. (2001). However, variables selected by SPA may be with low signal/noise ratio (S/N) or useless for multivariate calibration, which can affect the model's precision (Ye et al. 2008). Thus, in this study, UVE was firstly employed to select informative variables, and SPA was followed to select variables. The process of SPA was executed in MATLAB 7.6 (The Math Works, Natick, USA).

Uninformative Variable Elimination

In full spectral variables, there are some of the variables that can be noisy, have high variances but small covariance with y, and/or do not contain information relevant for modeling Y. Eliminating these variables from the explanatory part of data can improve the model. Employing the variables selected by UVE for partial least squares (PLS) modeling can avoid a model overfitting and usually improve its predictive ability. UVE is user independent and does not present any configuration problems, compared with other variable selection methods (Centner et al. 1996). UVE evaluates the reliability of each variable in the model through the stability of each variable, and removes the variables of X that have no relation with the property of interest Y (Wu et al. 2009a).

This elimination is based on the analysis of the stability of the regression coefficient b:

$$C_i = \frac{\mathrm{mean}\,(b_i)}{S(b_i)} \tag{1}$$

where c_j is the stability of the regression coefficients. It is the ratio between the mean of b_j and the standard deviation of b_j, obtained by leave-one-out cross-validation for each

Table 1 Statistical values of calcium (milligrams per 100 g) in powdered milk in calibration and prediction sets

Data set	Sample no.	Range	Mean	Standard deviation
Brand 1	52	243.1–302.7	272.4	17.8
Brand 2	59	382.2–436.4	408.0	15.7
Brand 3	60	411.7–452.0	429.3	12.5
Brand 4	60	627.2–722.7	673.8	27.6
Brand 5	60	427.1–486.8	457.0	17.8
Brand 6	59	404.8–553.4	479.6	45.3
Brand 7	59	445.3–555.7	502.0	32.9
Calibration	209	243.1–722.8	462.7	114.8
prediction	200	243.3–722.7	458.9	114.1
All samples	409	243.1–722.8	460.0	114.3

variable j. To determine the uninformative variables, an artificial random variable matrix, with a range of approximately 10^{-10} (noise level), is added to the dataset, and their C values are computed. If $\left|c_j^{\text{dataset}}\right| < \left|\max\left(c_j^{\text{random}}\right)\right|$, the jth experimental variable is considered to be uninformative. This variable is considered to give more information than the random variables, and is eliminated from the dataset. There are some variants of UVE, like α-UVE, where the cutoff value is the value of the alpha percent of the $\left|C_j\right|$ of the artificial random variables. In this paper the cutoff value used was 99% of the $\left|c_j^{\text{random}}\right|$. The process of UVE was executed in MATLAB 7.6 (The Math Works, Natick, USA).

Chemometric Calibration of LS-SVM

Support vector machines is a powerful methodology for solving problems in nonlinear classification, function estimation, and density estimation which has also led to many other recent developments in kernel-based methods in general. LS-SVM are reformulations to the standard SVMs which lead to solving linear KKT systems. LS-SVMs are closely related to regularization networks and Gaussian processes but additionally emphasize and exploit primal–dual interpretations. As giving a good performance under general smoothness assumptions on handling the nonlinear relationships between the spectra and target attributes, RBF kernel was used in this study. Grid search technique was applied to find out the optimal parameter values which include regularization parameter gam (γ) and the RBF kernel function parameter sig^2 (σ^2). In this study, these parameters were optimized with values of γ in the range of 2^{-1}–2^{10} and σ^2 in the range of 2–2^{15} with adequate increments. These ranges were chosen from previous studies where the magnitude of parameters to be optimized was established. For each combination of γ and σ^2 parameters, the root mean square error of cross-validation (RMSECV) was calculated and the optimum parameters were selected when produced smaller RMSECV. The details of LS-SVM description could be found in the literature (Wu et al. 2008d).

Model Evaluation Standard

The evaluation indices of predictive capability for all developed models were coefficient of determination (r^2), residual predictive deviation (RPD) and root mean square error of prediction set (RMSEP). RPD is the standard deviation of reference data for the validation samples divided by the standard error of prediction (SEP) and provides a standardization of the SEP (Williams 2001; Wu et al. 2010). Generally, a good model should have higher r^2 and RPD value, lower RMSEP values. The process of

LS-SVM was executed in MATLAB 7.6 (The Math Works, Natick, USA)

Results and Discussion

NIR and MIR Spectral Investigation

The original spectra transmission values of typical powdered milk samples from seven different brands at MIRS and NIRS regions are shown in Fig. 1. It could be observed that the trend of different curves was similar in the NIR region. They were mainly parallel and there were three absorption peaks. Peak at 4,260 cm^{-1} was assigned to CH$_2$ symmetric stretching vibration +CH$_2$ deformation vibration of HC=CHCH$_2$. Peak at 4,330 cm^{-1} was assigned to C–H stretching +C–H deformation vibration of CH$_2$. H$_2$O absorption appeared at 5,155 cm^{-1}. Absorption bands observed in the MIR region were associated with fundamental valence vibrations of functional groups of the molecule (Karoui et al. 2005). The broad and strong peak from 3,200 to 3,600 cm^{-1} was due to O–H stretching vibrations. CH$_2$ symmetric stretching vibration showed peak at 2,854 cm^{-1} and CH$_2$ asymmetric stretching vibration shows peak at 2,925 cm^{-1}. Broad peak between 2,000 and 2,250 cm^{-1} as R′–C≡C–R. C=O stretching vibration showed peak at 1,745 cm^{-1}. C=C stretching vibration shows peak at 1,650 cm^{-1}. Peaks at 875 and 898 cm^{-1} were assigned to C–H deformation vibration. However, we did not know which peaks were related to calcium iron absorption and were useful for the calcium content determination. Moreover, when there were many samples, their spectral curves would be overlapped, and their absorbed peaks were hard to identify and analyze.
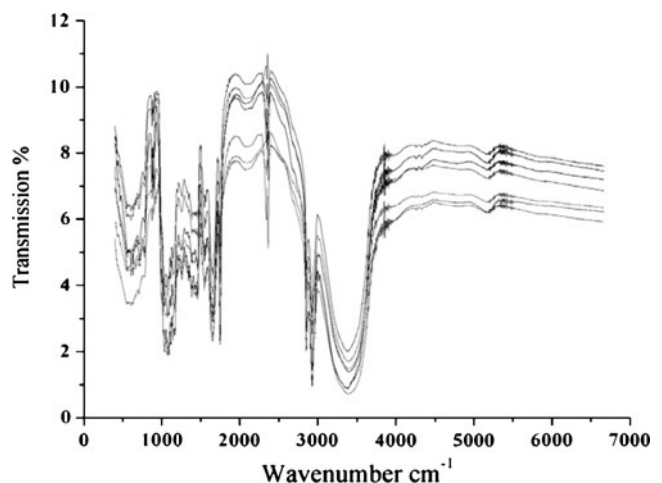

Fig. 1 NIR and MIR transmission spectra of powdered milk samples from seven different brands

Therefore, it was important to perform variable selection in multivariable analysis.

Calibration and Prediction Sets

The descriptive statistics of calcium in powdered milk determined by standard laboratory methods are presented in Table 1. All the samples were divided into calibration (209 samples) and prediction (200 samples) sets according to the Kennard–Stone (KS) algorithm (Kennard and Stone 1969). KS is a classic method to extract a representative set of objects from a given data set. It selects a representative data set consisting of samples separated by large Euclidean distances. The calibration and validation samples were used in the LS-SVM procedures. A broad range of concentration variation was observed in the calibration and prediction sets (Table 1). This situation would be helpful to build a stable and robust calibration model for the calcium content determination of powdered milk.

Full-Spectrum LS-SVM Model

LS-SVM models were developed to determine calcium content in powdered milk. The full spectra of NIR and MIR were inputted into LS-SVM, respectively. MIR spectra obtained much better results (with r^2=0.916, RPD=3.107 and RMSEP=35.974) than NIR (with r^2=0.571, RPD= 1.323, and RMSEP=84.762). RPD values of MIR up to 3 shows a reliable ability for calcium content determination and was considered suitable for screening (Williams 2001). In order to evaluate the performance of using LS-SVM for the model calibration, partial least square regression (PLS), a classical spectral calibration algorithm, was used also for the NIR and MIR spectral calibration model establishment. The PLS analysis was operated on "The Unscrambler V 9.7" (CAMO Process AS, OSLO, Norway). Both four latent variables were selected for NIR and MIR model. The r^2 decreased to 0.365 compared to 0.571 of LS-SVM for NIR analysis, and decreased 0.785 compared to 0.916 of LS-SVM. LS-SVM showed better prediction ability than PLS. Therefore, based on LS-SVM, a good determination performance could be achieved by MIRS, and NIR did not

show good prediction ability. The reason might be that NIRS is low sensitive compared to MIR. The superposition of many different overtones and combination bands in the NIR region causes a very low structural selectivity. In MIRS region, many fundamentals can usually be observed in isolated positions (Karoui and De Baerdemaeker 2007) which might contain metal ion information. MIRS can monitor low levels of key analytes and functional groups (Roychoudhury et al. 2006). From the plot of original spectra transmission in NIRS region (Fig. 1), it could be seen that the transmission spectral curves are flat. There were lots of high frequency noises in the NIRS region. The results have much noise and other unrelated information arises from overtones and combinations of such vibrations, rendering them more difficult to interpret (Reid et al. 2005). Although it was hard to distinguish each curve with different metal ion contents through the naked eye, the LS-SVM could extract useful information from the MIR spectra to establish the calibration model and did a precise prediction of calcium content in powdered milk.

However, as aforementioned, there were 2,756 and 3,727 variables in the full NIR or MIR spectra, respectively. Thousands of variables caused the LS-SVM process to be complex and time-consuming. Moreover, variables selected by SPA may be with low signal/noise ratio (S/N) or useless for multivariate calibration, which can affect model precision of prediction (Ye et al. 2008). Thus, we proposed UVE to firstly eliminate uninformation variables.

Variable Elimination Using UVE

In the process of UVE, the useful variables in the full spectrum of NIR and MIR were selected, and then inputted into LS-SVM. Different latent variables (LVs) numbers of PLS model were used to establish UVE models. The numbers of eliminated variables were different based on different LVs used in UVE. Numbers of LVs were calculated from 1 to 30 and the best LV numbers was determined based on RMSEP. When LV number was 8 and 19, best NIR and MIR results were obtained, respectively (Table 2). Although the number of selected MIR variables was much larger than NIR, the prediction results based on

Table 2 Prediction results of calcium content in powdered milk using selected NIR and MIR wavenumbers based on five different PC numbers which obtained the five small RMSEP in UVE

NIR					MIR				
LVs	Variables	r^2	RMSEP	RPD	LVs	Variables	r^2	RMSEP	RPD
5	905	0.683	73.554	1.523	18	1,261	0.937	29.011	3.872
6	562	0.685	77.512	1.438	19	1,241	0.941	27.770	4.074
7	160	0.670	72.057	1.550	20	1,242	0.938	28.695	3.893
8	159	0.685	71.678	1.558	21	1,241	0.934	29.069	3.847
9	106	0.658	72.775	1.539	22	1,225	0.931	29.679	3.751

selected variables of MIR were much better than NIR. Moreover, the prediction results were improved using selected spectral variables of both NIR and MIR, compared with results predicted based on full spectrum.

Figure 2a and b show the stability of each variable in the NIR and MIR for calcium content prediction by UVE of 8 and 19 LVs, respectively. Wavenumber variables were at left of the vertical line, while random variables were at the right side. Two horizontal lines show the lower and upper cutoff. The wavenumbers whose stability is within the cutoff lines should be treated as uninformative and eliminated. Finally, 159 and 1,241 variables were selected by UVE-PLS for NIR and MIR, respectively. The UVE selected spectral variables were only 5.8% for NIR spectra and 33.3% for MIR spectra. The RPD of UVE selected variables were improved compared to the full spectra by 17.8% for NIR spectra and 31.1% for MIR spectra. The RMSEP of UVE selected variables were reduced compared to full spectra by 15.4% for NIR spectra and 22.8% for MIR spectra. Figure 3 is the plots of 159 (a) and 1,241 (b) selected variables by UVE for NIR and MIR, respectively. Columns represent selected variables. The curve shows the original spectrum of sample one of the calibration data set for comparison. Some retained variable regions might attribute to some specific chemical structure related to calcium ion, while others which might not relate to calcium ion might attribute to calcium prediction model.

Further Variable Selection Using SPA

After UVE analysis, the numbers of input variables were reduced from 2,756 to 129 for NIR and from 3,727 to 1,241 MIR. SPA was carried out on the UVE selected NIR and MIR variables, respectively. The optimal variable number was selected according to the smallest root mean square error (RMSE) value. The selection process was operated by comparing the RMSE values of different variable numbers from 1 to 30. Finally, 11 variables (RMSE=77.23) and 15 (RMSE=41.78) variables were selected as the best ones for NIR and MIR, respectively.

Figure 3 shows the plots of 11 (c) and 15 (d) variables selected by SPA on NIR and MIR, respectively. Columns represent selected wavenumber variables. Curves show the

original spectrum of sample one of the calibration data set for comparison. The 11 selected NIR wavenumber variables by UVE-SPA were 4,089; 4,331; 4,486; 4,802; 4,984; 5,276; 5,316; 5,346; 5,728; 6,058; and 6,236 cm^{-1}. Some retained wavenumber regions might attribute to some specific chemical structure related to calcium ion. Associated to an absorbance peak in Fig. 1, 4,331 cm^{-1}, was assigned to C–H stretching+C–H deformation vibration of CH_2. The value, 4,485 cm^{-1}, was assigned to N–H stretching+NH_3 deformation vibration of some amino acids. The value, 4,802 cm^{-1}, was assigned to O–H stretching+O–H deformation vibration of R–OH. The value, 4,984 cm^{-1}, was assigned to N–H symmetric stretching+amide II of $CONH_2.CONHR$. The value, 6,058 cm^{-1}, was assigned to 2×C–H stretching of R–CH–CH. Other selected wavenumber variables might not relate to calcium ion but might attribute to calcium prediction model. As there were many different overtone and combination bands in the NIR region, some obtained effective wavenumber variables were not obvious in the NIR spectra. In addition, H_2O absorption appeared at 5,155 cm^{-1}, which has less information about calcium, were not selected.

The 15 selected MIR wavenumber variables by UVE-SPA were 671; 865; 1,401; 1,427; 1,475; 1,541; 1,602; 1,660; 1,731; 1,799; 1,876; 2,526; 2,902; 3,588; and 3,698 cm^{-1}. The values, 3,588 and 3,698 cm^{-1}, were assigned to O–H stretching vibrations. Wavenumbers between 2,500 and 1,900 cm^{-1} assigned to C≡C were not selected. It could be deduced that C≡C has less information about calcium ion. As many effective wavenumbers were between 1,900 and 1,400 cm^{-1} which were assigned to double bond stretching vibration, C=O and C=C vibrations might attribute to calcium ion. The result was similar to other previous studies (Mesubi 1982; Byler and Farrell 1989; Nara and Tanokura 2008).

The selected 11 NIR and 15 MIR variables were set as the input variables of LS-SVM models. The 15 MIR variables obtained much better results (with r^2=0.930, RPD=3.703, RMSEP=30.162, and relative errors of prediction (RSEP)=5.22%) than the 11 NIR variables (with r^2=0.636, RPD=1.417, RMSEP=78.815 and RSEP= 13.40%). A good RPD of 3.703 was obtained by MIR

Fig 2 Stability in the NIR (a) and MIR (b) for calcium content prediction by UVE-PLS of 8 and 19 LVs, respectively. The *two blue horizontal dotted lines* indicated the lower and upper cutoff
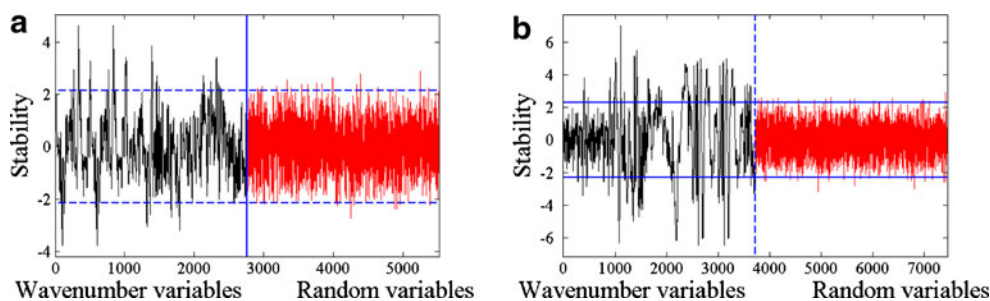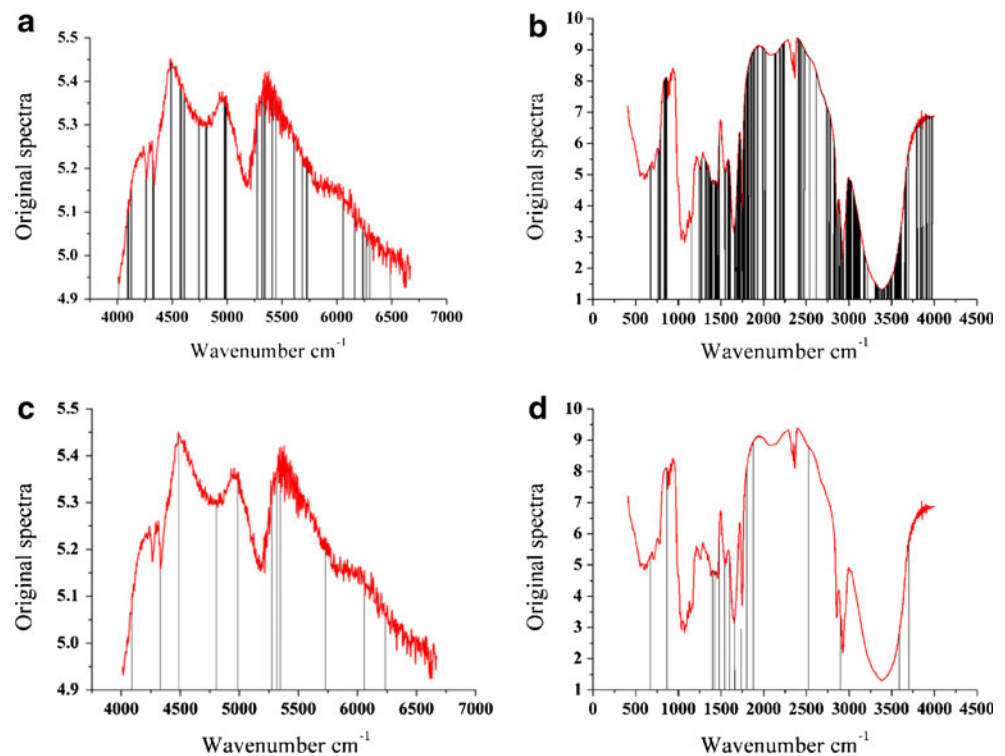
Fig. 3 Plots of selected wave-number variables. One hundred fifty-nine (**a**) and 1,241 (**b**) variables were selected by UVE for NIR (**a**) and MIR (**b**), and 11 (**c**) and 15 (**d**) variables were selected by UVE-SPA for NIR (**c**) and MIR (**d**), respectively. *Columns* represent selected wavenumber variables. *Curves* show the original spectrum of sample one of the calibration data set for comparison

spectra. It indicates the model was suitable for screening. The predicted vs. reference values of calcium content in powdered milk based on 11 NIR (**a**) and 15 MIR (**b**) variables are shown in Fig. 4. MIR showed its excellent calcium prediction ability than NIR. MIR spectroscopy had a good prediction ability of calcium content in powdered milk. The samples in Fig. 4b were distributed more closely to the regression line, compared with Fig. 4a.
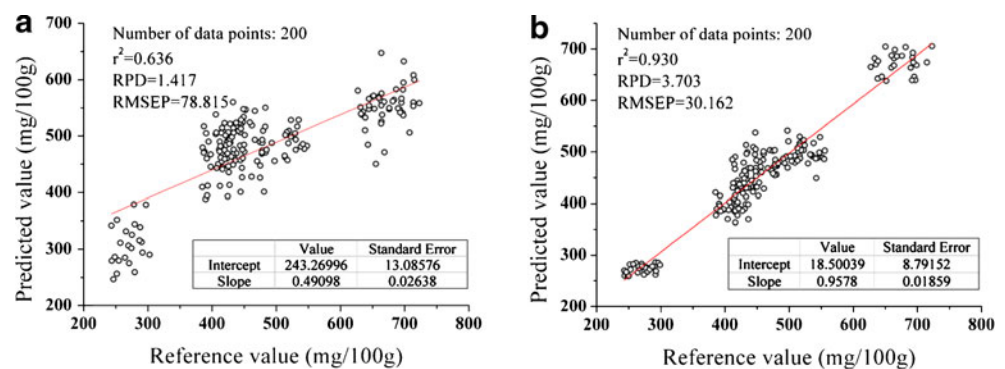
The results of both 11 NIR and 15 MIR variables selected by UVE-SPA were little worse than those of 159 NIR and 1241 MIR variables selected by UVE. The RPD of UVE-SPA selected variables were decreased compared to UVE selected variables by 9.1% for NIR spectra and 9.1% for MIR spectra. The RMSEP of UVE-SPA selected variables were increased compared to UVE selected variables by 10.0% for NIR spectra and 8.6% for MIR spectra. The reason might be that

few variables were used. Only 6.9% of UVE selected NIR spectral variables and 1.2% of UVE selected MIR spectral variables were selected. However, the training time using LS-SVM increased with the square of the number of training samples and linearly with the number of variables (Chauchard et al. 2004). Therefore, the performances of UVE-SPA were still acceptable. Hence, the application of UVE-SPA would be helpful for the spectral variable selection. UVE-SPA is more parsimonious and simple to interpret the spectral wavenumbers.

Conclusion

NIR and MIR spectroscopy techniques were evaluated to determine calcium content in powdered milk. After UVE

Fig. 4 Predicted vs reference values of calcium content in powdered milk by LS-SVM models based on 11 NIR (**a**) and 15 MIR (**b**) wavenumbers selected by UVE-SPA

analysis, the variable numbers were reduced to 159 and 1,241 for NIR and MIR, respectively. Finally, 11 NIR and 15 MIR variables were selected by SPA, respectively. A good RPD of 3.703 obtained shows the ability of the UVE-SPA-LS-SVM model based on MIR spectra. The 15 MIR variables obtained much better results than the 11 NIR variables (0.930 for r^2, 3.703 for RPD, 30.162 for RMSEP, and 5.22% for RSEP vs. 0.636 for r^2, 1.587 for RPD, 78.815 for RMSEP, and 13.40% for RSEP). The results of both 11 NIR and 15 MIR variables selected by UVE-SPA were a little worse than those of 159 NIR and 1241 MIR variables selected by UVE, but were improved, compared to those of full NIR and MIR spectral variables, respectively. The performances of UVE-SPA were still acceptable as the variable numbers were much smaller than those selected only by UVE. The good performance showed a potential application using UVE-SPA to select NIR and MIR effective variables. The overall results indicated that MIR spectroscopy could be applied as a precision and rapid method for the determination of calcium content in powdered milk. The calcium content prediction using NIR spectroscopy needs to be further investigated.

References

Alexandrakis, D., Downey, G., & Scannell, A. G. M. (2009). Rapid non-destructive detection of spoilage of intact chicken breast muscle using near-infrared and Fourier transform mid-infrared spectroscopy and multivariate statistics. *Food and Bioprocess Technology*. doi:10.1007/s11947-009-0298-4. in press.

Alvarez, M., Mazo-Gray, V., Alvarez, M., & Mazo-Gray, V. (2005). Determination of potassium and calcium in powdered milk by energy-dispersive X-ray fluorescence spectrometry. *X-Ray Spectrometry, 19*(6), 285–287.

AOAC (2000). *Official methods of analysis 17th Ed. Method 965.17*. Maryland: The Association of Official Analytical Chemists

Araujo, M. C. U., Saldanha, T. C. B., Galvao, R. K. H., Yoneyama, T., Chame, H. C., & Visani, V. (2001). The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometrics and Intelligent Laboratory Systems, 57*(2), 65–73.

Berridge, M. J. (1975). The interaction of cyclic nucleotides and calcium in the control of cellular activity. *Advances in Cyclic Nucleotide Research, 6*, 1–98.

Bokobza, L. (1998). Near infrared spectroscopy. *Journal of Near Infrared Spectroscopy, 6*(1998), 3–17781.

Borin, A., Ferrao, M. F., Mello, C., Maretto, D. A., & Poppi, R. J. (2006). Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk. *Analytica Chimica Acta, 579*(1), 25–32.

Byler, D. M., & Farrell, H. M. (1989). Infrared spectroscopic evidence for calcium-ion interaction with carboxylate groups of casein. *Journal of Dairy Science, 72*(7), 1719–1723.

Centner, V., Massart, D. L., deNoord, O. E., deJong, S., Vandeginste, B. M., & Sterna, C. (1996). Elimination of uninformative variables for multivariate calibration. *Analytical Chemistry, 68*(21), 3851–3858.

Chauchard, F., Cogdill, R., Roussel, S., Roger, J. M., & Bellon-Maurel, V. (2004). Application of LS-SVM to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes. *Chemometrics and Intelligent Laboratory Systems, 71*(2), 141–150.

Chen, K. L., & Jiang, S. J. (2002). Determination of calcium, iron and zinc in milk powder by reaction cell inductively coupled plasma mass spectrometry. *Analytica Chimica Acta, 470*(2), 223–228.

Downey, G. (1998). Food and food ingredient authentication by mid-infrared spectroscopy and chemometrics. *TrAC, Trends in Analytical Chemistry, 17*(7), 418–424.

Hitti IKA & Thomas JDR. (1985). Ion-selective electrode determination of calcium and potassium in bovine skim-milk powder. *Analytical Letters Part a-Chemical Analysis, 18*(8), 975–983.

Hu, X., Toyoda, K., Yamanoue, M., Ihara, I., & Nakai, K. (2010). Evaluation of fatty acid profile of wagyu beef by ATR-FTIR spectroscopy. *Food and Bioprocess Technology, 3*(6), 883–891.

Jouanrimbaud, D., Massart, D. L., Leardi, R., & Denoord, O. E. (1995). Genetic algorithms as a tool for wavelength selection in multivariate calibration. *Analytical Chemistry, 67*(23), 4295–4301.

Kalivas, J. H., Roberts, N., & Sutter, J. M. (1989). Global optimization by simulated annealing with wavelength selection for ultraviolet visible spectrophotometry. *Analytical Chemistry, 61*(18), 2024–2030.

Karoui, R., & De Baerdemaeker, J. (2007). A review of the analytical methods coupled with chemometric tools for the determination of the quality and identity of dairy products. *Food Chemistry, 102*(3), 621–640.

Karoui, R., Dufour, E., Pillonel, L., Schaller, E., Picque, D., Cattenoz, T., et al. (2005). The potential of combined infrared and fluorescence spectroscopies as a method of determination of the geographic origin of emmental cheeses. *International Dairy Journal, 15*(3), 287–298.

Kennard, R. W., & Stone, L. A. (1969). Computer aided design of experiments. *Technometrics, 11*(1), 137.

Mesubi, M. A. (1982). An infrared study of zinc, cadmium, and lead salts of some fatty-acids. *Journal of Molecular Structure, 81*(1–2), 61–71.

Nara, M., & Tanokura, M. (2008). Infrared spectroscopic study of the metal-coordination structures of calcium-binding proteins. *Biochemical and Biophysical Research Communications, 369*(1), 225–239.

Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., et al. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biology and Technology, 46*(2), 99–118.

Reid, L. M., Woodcock, T., O'Donnell, C. P., Kelly, J. D., & Downey, G. (2005). Differentiation of apple juice samples on the basis of heat treatment and variety using chemometric analysis of MIR and NIR data. *Food Research International, 38*(10), 1109–1115.

Roychoudhury, P., Harvey, L. M., & McNeil, B. (2006). At-line monitoring of ammonium, glucose, methyl oleate and biomass in a complex antibiotic fermentation process using attenuated total

reflectance-mid-infrared (ATR-MIR) spectroscopy. *Analytica Chimica Acta, 561*(1–2), 218–224.

Upreti, P., & Metzger, L. E. (2006). Utilization of Fourier transform infrared spectroscopy for measurement of organic phosphorus and bound calcium in cheddar cheese. *Journal of Dairy Science, 89*(6), 1926–1937.

Williams, P. C. (2001). *Near-infrared technology in the agricultural and food industries*. Saint Paul: American Association of Cereal Chemists.

Wilson, R. H., & Tapp, H. S. (1999). Mid-infrared spectroscopy for food analysis: recent new applications and relevant developments in sample presentation methods. *TrAC, Trends in Analytical Chemistry, 18*(2), 85–93.

Woodcock, T., Fagan, C. C., O'Donnell, C. P., & Downey, G. (2008). Application of near and mid-infrared spectroscopy to determine cheese quality and authenticity. *Food and Bioprocess Technology, 1*(2), 117–129.

Wu, D., Feng, S., & He, Y. (2007). Infrared spectroscopy technique for the nondestructive measurement of fat content in milk powder. *Journal of Dairy Science, 90*(8), 3613–3619.

Wu, D., Feng, S., & He, Y. (2008a). Short-wave near-infrared spectroscopy of milk powder for brand identification and component analysis. *Journal of Dairy Science, 91*(3), 939–949.

Wu, D., He, Y., & Feng, S. (2008b). Short-wave near-infrared spectroscopy analysis of major compounds in milk powder and wavelength assignment. *Analytica Chimica Acta, 610*(2), 232–242.

Wu, D., He, Y., Feng, S. J., & Sun, D. W. (2008c). Study on infrared spectroscopy technique for fast measurement of protein content in milk powder based on LS-SVM. *Journal of Food Engineering, 84*(1), 124–131.

Wu, D., Yang, H. Q., Chen, X. J., He, Y., & Li, X. L. (2008d). Application of image texture for the sorting of tea categories using multi-spectral imaging technique and support vector machine. *Journal of Food Engineering, 88*(4), 474–483.

Wu, D., Chen, X. J., Shi, P. Y., Wang, S. H., Feng, F. Q., & He, Y. (2009a). Determination of alpha-linolenic acid and linoleic acid in edible oils using near-infrared spectroscopy improved by wavelet transform and uninformative variable elimination. *Analytica Chimica Acta, 634*(2), 166–171.

Wu, D., Wu, H. X., Cai, J. B., Huang, Z. H., & He, Y. (2009b). Classifying the species of exopalaemon by using visible and near infrared spectra with uninformative variable elimination and successive projections algorithm. *Journal of Infrared and Millimeter Waves, 28*(6), 423–427.

Wu, D., He, Y., Nie, P. C., Cao, F., & Bao, Y. D. (2010). Hybrid variable selection in visible and near-infrared spectral analysis for non-invasive quality determination of grape juice. *Analytica Chimica Acta, 659*(1–2), 229–237.

Ye, S. F., Wang, D., & Min, S. G. (2008). Successive projections algorithm combined with uninformative variable elimination for spectral variable selection. *Chemometrics and Intelligent Laboratory Systems, 91*(2), 194–199.