Research Article

Authenticating Raw from Reconstituted Milk Using Fourier Transform Infrared Spectroscopy and Chemometrics

Lijuan Du,1,2 Weiying Lu,2 Boyan Gao,2 Jing Wang,1 and Liangli (Lucy) Yu3

1Beijing Advanced Innovation Center for Food Nutrition and Human Health, Beijing Technology & Business University (BTBU), Beijing 100048, China
2Institute of Food and Nutraceutical Science, Department of Food Science and Technology, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai 200240, China
3Department of Nutrition and Food Science, University of Maryland, College Park, MD 20742, USA

Correspondence should be addressed to Weiying Lu; weiying.lu@sjtu.edu.cn and Boyan Gao; gaoboyan@sjtu.edu.cn

Received 1 November 2018; Revised 12 January 2019; Accepted 11 February 2019; Published 24 March 2019

Guest Editor: Xavier Cetó

Copyright © 2019 Lijuan Du et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fourier transform infrared (FTIR) spectroscopy combined with chemometrics was used to authenticate raw milk from their reconstituted counterparts. First, the explanatory principal component analysis (PCA) was employed to visualize the relationship between raw and reconstituted milk samples. However, the degree of separation between two sample classes was not significant according to direct observation of the scores plot, indicating FTIR spectra may contain complicated chemical information. Second, partial least-squares-discriminant analysis (PLS-DA) that incorporate additional class membership information as modelling input was further calculated. The PLS-DA scores yielded clear separation between two classes of samples. Additionally, possible components from the model loading were studied, and the PLS-DA model was validated internally under the model population analysis framework, as well as externally using an independent test set. This study gave insights into the authentication of milk using FTIR spectroscopy with chemometrics techniques.

1. Introduction

Milk is one of the most consumed food items, which has significant nutritional and economical importance. The reconstitution of milk is an act that adulterates skimmed or whole milk powder in part into raw milk or completely substitutes raw milk [1, 2]. Such fraud can achieve marginal economic gain since the shelf life of milk powder is longer than their raw counterparts. Adulteration of powdered milk in their raw counterparts may alter the original nutritional and functional value of raw milk, and thus, it may provoke a crisis of confidence to consumers for milk industry. Therefore, a rapid, simple, and automated method for milk adulteration detection is required.

Adulteration of commercial milk powder is even more challenging to detect than many other common milk adulterants such as melamine or plant protein, due to the extremely similar chemical composition. Therefore, measurements with both high sensitivity and resolution are preferred. For instance, two-dimensional gel electrophoresis combined with matrix-assisted laser desorption/ionization-mass spectrometry was reported to detect powdered milk in raw cow’s milk based on the modified peptide including oxidation, lactosylation, and deamination protein products [1]. The detection of furosine [3] and lysinoalanine [4] by liquid and gas chromatography was introduced. Rather than seeking specific marker components, other kinds of methods applied empirical models or fingerprints to detect adulteration. Differentiation of raw from reconstituted milk by the stable isotope ratios of oxygen and hydrogen was also reported [5]. However, the above methods often require either time and cost-consuming mass spectrometric detection or labor-intensive sample pretreatment or analysis procedures, which render these methods inapplicable to large-scale assay.

Rapid analytical techniques such as spectroscopy or electronic noses, with the combination of empirical modelling, provide a convenient approach to characterize complex food matrices. For example, the adulteration of
whole milk with milk powder was detected by spectro-photometry. The ultraviolet and visible spectroscopy has been applied to the detection and quantification of raw milk with reconstituted full-fat milk powder [2]. The transmittance of raw milk adulterated with full-fat dry milk powder reconstituted milk was observed and possibly explained the phenomenon by turbidity variation induced from low degree of homogenization [6]. In addition, the fluorescence of advanced Maillard products and soluble tryptophan (FAST) index had been devised for distinguishing milk heat treatments [7]. However, these researches were based on empirical observations without clear metrics or limits, and thus limited information is extracted from the spectra. Fingerprints combined with chemometrics methods were suitable for processing complex analytical data in an automated and objective decision-making manner. For instance, the adulteration of reconstituted milk or water with electronic noses constructed with ten different metal oxide sensors was monitored with chemometrics modelling [8].

Fourier transform infrared spectroscopy (FTIR) has been widely used for food quality monitoring including authenticity and traceability, due to its fast speed and nondestructive capabilities [9, 10]. FTIR spectroscopy has been successfully demonstrated in milk authentication such as to detect soymilk adulterated in cow or buffalo milk [11]. It is therefore interesting to test whether FTIR spectroscopy could further identify any reconstitution in raw milk.

Adulteration in food ingredients such as milk or olive oil suggested that chemometrics modelling is becoming an essential part in the fingerprinting analyses [12–14]. Specifically, infrared and Raman spectroscopy studies on detection of food adulterations had resulted in a wide range of successful applications. Raman spectroscopy could detect melamine adulterant in milk powder at the detection limit of 0.13% (w/w) by two vibration modes at 673 and 982 cm$^{-1}$ [12]. Additionally, machine-learning methods provide possibilities to a wide range of application of infrared spectroscopy in food authentication and quality control. For instance, near-infrared reflectance spectral were used to examine the authentication of skim and nonfat dry milk powder using analysis of variance- (ANOVA-) principal component analysis (PCA), pooled-ANOVA, and partial least-squares-regression (PLSR) [13]. The potential of near-infrared (NIR) spectroscopy combined with chemometrics for nontargeted detection of adulterants in skim and nonfat dry milk powder was also studied [14]. Therefore, it is interesting to test whether infrared spectroscopy combined with chemometric modelling techniques can be applied in detecting milk powder in raw milk.

In this study, FTIR combined with chemometrics was developed for the detection of milk adulteration. Specifically, infrared spectral fingerprints combined with chemometrics were tested in detecting reconstituted milk powder in raw milk. The workflow is demonstrated in Figure 1. This study aimed at detecting milk powder adulterated in raw milk using FTIR spectroscopy combined with chemometrics. This work may serve as a reference for quality assurance of raw milk and its related dairy products.

2. Materials and Methods

2.1. Sample Collection. Twenty raw milk samples were provided by local milk farms located in Qingdao, Shandong province, China. These farms were certified suppliers of the Nestle Corporation (Vevey, Switzerland). Each raw milk sample was stored in a separate 100 mL polyethylene bottle. All samples were immediately frozen after collection. The bottles were placed in a portable Styrofoam box with ice packs and dry ice to maintain optimum low temperature and stored at $-20^\circ$C once transferred to the laboratory. Four anonymously branded commercial milk powders with unrevealed processing techniques were purchased from local groceries in Shanghai, China.

2.2. Sample Pretreatment. Raw milk samples were directly lyophilized using a Labconco freeze dryer (Kansas City, MO, USA). The freeze-drying process removes any moisture that may interfere the FTIR measurement. It was served as a pretreatment step that maintains the original chemical compositions of raw milk as much as possible and made storage and testing of a large batch of samples possible.

For the preparation of adulterated milk with reconstituted milk powder, first, raw milk was randomly selected as the standard sample. Then, each commercial milk powder was added to the authentic liquid milk in 0.5, 1, 3, 5, and 10% (w/v), resulting in five partially reconstituted samples, respectively. After that, the mixtures were sonicated for 20 min. Finally, the mixtures were lyophilized. The lyophilizates were subjected to FTIR analysis.

2.3. FTIR Analysis. All fingerprints were collected using a Nicolet 6700 FTIR spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) equipped with a Smart iTR single bounce germanium crystal attenuated total reflectance (ATR) sampling accessory (Thermo Fisher Scientific). The spectra were collected in the transmittance mode by an average of 60 scans ranging between 650 and 4000 cm$^{-1}$ with a 0.48 cm$^{-1}$ interval. Before each measurement, an independent background scan was performed and subtracted immediately to minimize atmospheric interference and instrument fluctuation.

All samples were prepared and tested in triplicates, including 60 raw milk samples (20 raw milk samples × 3 replicates) and 60 reconstituted milk samples (4 milk powders × 5 adulteration levels × 3 replicates), resulting in totally 120 spectra.

2.4. Chemometrics Modelling. All raw data were imported to MATLAB (version R2018a, The MathWorks, Natick, MA, USA). Different preprocessing strategies such as wave-number region selection, autoscaling, standard normal variate (SNV), and derivative were applied. All chemometric analyses including preprocessing, PCA, and partial least-squares-discriminant analysis (PLS-DA) were performed using in-house MATLAB routines running on a personal
computer under Windows 7 operating system (Microsoft Corporation, Redmond, WA, USA).

For internal validation, statistically relevant comparisons were achieved by the model population analysis (MPA) framework [15]. The MPA is essentially based on cross validation of a series of submodels obtained from the original dataset through random sampling. In this work, MPA extract statistical information from models to achieve a statistically unbiased estimation of performance. The internal validation process was evaluated repeatedly for 100 bootstraps. For external validation, the Latin partition approach was employed to split the whole data set into training and test sets prior to classification. To evaluate the result, prediction accuracy of the data set is used. Prediction accuracy is an estimated percentage of correct identifications when the model is applied for unknown samples, which is widely applied to assess the overall performance of a specific classification model.

3. Results and Discussion

3.1. FTIR Spectral Characteristics. The FTIR spectral fingerprints contained representative information for different components in milks. The mean spectra of raw and reconstituted milks are shown in Figure 2. The absorption bands observed at 1630 to 1680 cm\(^{-1}\) and 1510 to 1570 cm\(^{-1}\) may be induced by C=O stretching vibrations of absorption of amide I and N-H and C-H bending vibration absorption of amide II from milk protein, respectively [16, 17]. The bands around 2920, 2850, and 1743 cm\(^{-1}\) may be anti-symmetric and symmetric CH\(_2\) stretching and carbonyl group C=O double bond stretching from milk fat, respectively [18]. The absorption bands located at 3200 to 3800 cm\(^{-1}\), 1030 to 1200 cm\(^{-1}\), 900 to 930 cm\(^{-1}\), and 755 to 785 cm\(^{-1}\) may be associated with carbohydrate [19, 20]. These peaks also resemble the largest differed variable ranges in fingerprints. However, noting that, the mean spectra occurred in high overlap, suggesting a strong compositional similarity. Additionally, no evident peaks were determined as marker peaks since any single component is unlikely to be a critical differentiation factor. Consequently, it is hard to detect milk adulteration with mere visual inspection. Therefore, applying multivariate methods to address the overall spatial distribution of the data is necessary.

3.2. PCA Explanatory Study. PCA was performed to preliminary visualize the multivariate distribution of all fingerprints. Figure 3 shows the PCA scores plot, with autoscaling preprocessing applied. The PCA scores plot suggested that there were no obvious discriminations between raw and reconstituted samples using raw fingerprints. Specifically, no separation tendencies between raw and reconstituted samples were observed along the axes of both principal components PC1 and PC2, the two largest principal components. The combined variances explained by PC1 and PC2 were 88% of the total variances, indicating that the most dominant variances of the fingerprints do not closely relate to the reconstituted milk. The PCA result was also consistent with the result from visual inspection. Different preprocessing methods, including SNV alone and SNV combined with first- and second-order derivatives, were also studied by observing the PCA scores plot (data not shown). Regardless of preprocessing methods or the combinations used, there were no obvious discriminations between raw and reconstituted samples. By selecting the wavenumber region of 800–1800 cm\(^{-1}\), the degree of separation between raw and reconstituted milk cannot be improved either (data not shown). Consequently, it is not confirmed by PCA that there can be characteristic bands in the fingerprint region, nor there exhibit characteristics between two kinds of spectra. However, supervised multivariate classification may be capable of extracting information from complex data because the class memberships of samples were also included as the model input. Therefore, PLS-DA was applied to analyze the fingerprints further.
3.3. PLS-DA Model Evaluation. PLS-DA is perhaps one of the most well-known supervised classification methods in chemometrics. This method is based on partial least-squares-regression of continuous predictor variables, which seek for optimal latent variables with maximum covariance. Similar to PCA, PLS-DA was firstly applied as an explanatory approach to study the overall distribution. Different preprocessing methods were applied to the data, including wavenumber selection, autoscaling, first derivatives, and different combinations. It was indicated that PLS-DA achieved generally good separation of classes by PLS-DA scores. The best separation is shown in Figure 4, which is the X-scores (scores of the spectral data block) plot of the PLS-DA model by first selecting the wavenumber at 800–1800 cm\(^{-1}\), where the spectral differences were larger than other regions, with autoscaling and first derivative preprocessing. The two largest latent variables are displayed in Figure 4. Although with small portion of overlap, the distribution of tested samples clearly showed two clusters, indicating an intrinsically different fingerprint patterns among two classes of samples. A trend related to the adulteration level was also observed. Specifically, samples adulterated at 0.5%, the lowest adulteration level in this study, is located at the partial overlap with the raw milk sample cluster. Contrarily, samples adulterated at 10% are more significantly apart from raw milk, compared with those at lower adulteration levels.

In explanatory studies, both PCA and PLS-DA scores plots limit their indicative abilities in only two dimensions, namely, the first and second principal components or latent variables. Such analysis approach relies heavily on the final judgement of the researcher for the analysis of visual patterns instead of objective performance metrics. In comparison, the PLS-DA model is able to overcome this shortcoming by the automatic model-building process with a reasonable number of variables. By selecting 90% of original data as the training set, with 11 latent variables though internal validation procedure described in the next section, a final PLS-DA model was built and validated. Figure 5 shows the regression coefficients of the PLS-DA model. Positive and negative coefficients represent the relationships of the peaks to pure and reconstituted samples, respectively. The absolute magnitude of coefficients indicated the relative importance of peaks. Some interesting peaks arise in the PLS-DA coefficients. Peaks at 904 to 1288 cm\(^{-1}\) were generally associated with C-H bending, C-O-H in-plane bending, and C-O stretching vibrations of lipids, organic acids, and carbohydrate derivatives. Compared with the raw spectra shown in Figure 2, some peaks (904–1288 cm\(^{-1}\)) may be associated with carbohydrates. This might be attributed to a series of the Maillard reaction occurred in milk powder, which result in the reduction of lysine-rich proteins and lactose [21]. Peaks at 1583 cm\(^{-1}\) corresponded to unspecified compounds. The result is relevant with the PCA study that characteristic peaks may arise in the fingerprint region when authenticating raw milk. However, it did not agree with our previous findings that PCA and PLS-DA performed consistently in classifying pure milk and their counterparts adulterated with other powdered proteins [22], probably due to the complexity of spectra. It was indicated that, for the complex FTIR spectral fingerprints, the application of supervised classification methods is important because exploratory methods such as PCA did not yield a complete clear characterization.

3.4. PLS-DA Model Validation. Although PLS-DA model finds the possible characteristics between raw and reconstituted milk samples, evaluating the validity of the PLS-DA model is necessary, since PLS-DA may be prone to overfitting. Specifically, the quantitative metrics of PLS-DA prediction power were tested by both internal and external approaches to indicate the suitability and generalizability of the model. Firstly, the complete data set was split into training and external test sets. Secondly, the internal validation was performed solely on the training set by splitting the training set into internal training and calibration set. In internal validation, statistically relevant validation of PLS-DA modelling was achieved by MPA. To achieve a statistically unbiased estimation of performance, a series of PLS-DA models were built and evaluated repeatedly for 100 bootstraps. The average classification accuracy was 98% when 11 latent variables were applied, suggesting a reliable performance.

In external validation, the Latin partition approach was employed to split the whole data set into training (90%) and test (10%) sets prior to the PLS-DA classification. Unlike the previous PCA scores plot that used only two principal components to find possible separation between pure and reconstituted samples, 11 latent variables were applied for the final building of the PLS-DA model after bootstrapped Latin partition evaluation, indicating that there were many independent components presented in the sample to establish an effective model. Figure 6 shows the final prediction output of the PLS-DA model for external validation. All samples in the test set were correctly classified by PLS-DA.

It is also interesting to study the differences between different adulteration levels since Figure 4 presented differences as previously discussed. Therefore, PLSR was
applied to model the adulteration level. All other parameters remained the same as PLS-DA. The final external validation yielded a root-mean-squared error of 3.0, indicating an effective quantification of the adulteration level.

Figure 3: Principal component analysis scores plots of FTIR fingerprints.

Other than a 9:1 (training set/test set) split ratio, further evaluations by different split ratios of 2:1 and 1:1 were performed to prevent model overfitting. Except that, all other calculations remain unchanged. The result was consistent with that from the previous condition. Specifically, only one test sample was misclassified when the split ratio was 1:1, and all other predictions were correct. It can be concluded that the MPA modelling approach is robust and still reliable even when half of the data were removed.
4. Conclusion

FTIR spectroscopy combined with chemometrics has been successfully demonstrated to detect possible presence of reconstituted milk in raw milk. This work indicates FTIR spectroscopy has great potentials in quality control of milk and their related products because the PLS-DA model yielded satisfactory separation of the two spectral fingerprints. Noting that, due to the limited sample size and variability, careful selection of liquid and powdered milk in a larger data set may be necessary before practice to assure the universality of the final model. Additionally, simpler methods such as sampling without lyophilization and quantitating the level of adulteration need to be investigated in the future.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The authors would like to acknowledge the financial support by the National Natural Science Foundation of China (Grant no. 31501553); Beijing Advanced Innovation Center for Food Nutrition and Human Health, Beijing Technology and Business University (BTBU); and research fund from Nestec Ltd.

References


