



Fourier transform infrared spectroscopy and multivariate analysis for the detection and quantification of different milk species

N. Nicolaou, Y. Xu, and R. Goodacre¹

School of Chemistry and Manchester Interdisciplinary Biocentre, The University of Manchester, 131 Princess Street, Manchester, M1 7DN United Kingdom

ABSTRACT

The authenticity of milk and milk products is important and has extended health, cultural, and financial implications. Current analytical methods for the detection of milk adulteration are slow, laborious, and therefore impractical for use in routine milk screening by the dairy industry. Fourier transform infrared (FT-IR) spectroscopy is a rapid biochemical fingerprinting technique that could be used to reduce this sample analysis period significantly. To test this hypothesis we investigated 3 types of milk: cow, goat, and sheep milk. From these, 4 mixtures were prepared. The first 3 were binary mixtures of sheep and cow milk, goat and cow milk, or sheep and goat milk; in all mixtures the mixtures contained between 0 and 100% of each milk in increments of 5%. The fourth combination was a tertiary mixture containing sheep, cow, and goat milk also in increments of 5%. Analysis by FT-IR spectroscopy in combination with multivariate statistical methods, including partial least squares (PLS) regression and nonlinear kernel partial least squares (KPLS) regression, were used for multivariate calibration to quantify the different levels of adulterated milk. The FT-IR spectra showed a reasonably good predictive value for the binary mixtures, with an error level of 6.5 to 8% when analyzed using PLS. The results improved and excellent predictions were achieved (only 4–6% error) when KPLS was employed. Excellent predictions were achieved by both PLS and KPLS with errors of 3.4 to 4.9% and 3.9 to 6.4%, respectively, when the tertiary mixtures were analyzed. We believe that these results show that FT-IR spectroscopy has excellent potential for use in the dairy industry as a rapid method of detection and quantification in milk adulteration.

Key words: Fourier transform infrared spectroscopy, milk adulteration, authenticity of milk products, chemometrics

INTRODUCTION

Milk quality is important in the production of all types of cheese, especially in regard to cheese quality and characteristics. Goat and sheep milks are of higher value than cow milk and are used for the production of a variety of specialty cheeses. This presents the potential for financial gain by unscrupulous producers adulterating either goat or sheep milk with cow milk, thus resulting in nonauthentic milk products (Maudet and Taberlet, 2001). In addition to the ethical, religious, and cultural implications (Shatenstein and Ghadirian, 1998), consumers need to be protected from this kind of practice because of potential intolerance and allergic reactions to the cow milk component of these adulterated products (Bischoff, 2006; Venter, 2009; Chafen et al., 2010). Indeed, the European Union has legislation in place for the correct display of the constituents of dairy products protecting their authenticity (European Union, 2001), and various legislation on food labeling and authenticity exists among other member countries (Dennis, 1998).

Until now, several methods have been investigated, both within academic institutions and in industry, for their accuracy and practicality in detecting dairy product adulteration. These include several analytical approaches based on immunological, electrophoretic, and chromatographic techniques, as well as DNA-based processes such as species-specific PCR. Antibody-based assays used for the quantification and detection of species-specific milk proteins (antigens) form the basis of immunological methods. These have targeted proteins such as cow caseins (whole, γ_3 -, β - and α_{S1} -caseins), cow β -lactoglobulin, and cow IgG for the detection of cow milk adulteration, whereas goat whey proteins and goat α_{S2} -casein have been used for the detection of goat milk adulteration (Hurley et al., 2004). Enzyme-linked immunosorbent assays have been routinely used for this purpose and are performed using a variety of processes including direct, indirect competitive, and sandwich techniques (Levieux and Venien, 1994; Anguita et al., 1996; Beer et al., 1996; Anguita et al., 1997; Hurley et al., 2006). Even though ELISA techniques require rela-

Received July 13, 2010.

Accepted September 2, 2010.

¹Corresponding author: roy.goodacre@manchester.ac.uk

tively less sample preparation than other techniques, they are costly, because they rely on the use of expensive antibodies that cannot be reused and have a limited shelf-life. In addition, the reliance of these methods on specific protein identification and quantification is a potential drawback in the analysis of processed milk because proteolysis and heat denaturation can cause the loss of antibody-specific epitopes (Mayer, 2005; Hurley et al., 2006).

Polyacrylamide gel electrophoresis and isoelectric focusing are the main nonimmunological methods employed for the detection of casein or whey proteins in milk but they are slow and laborious for routine use in the dairy industry (Amigo et al., 1992; Levieux and Venien, 1994; Malin et al., 1994; Mayer, 2005; Addeo et al., 2009). Importantly though, isoelectric focusing has been adopted by the European Commission as the reference method for detecting cow κ -casein, with a detection limit of 1% (vol/vol) cow milk (European Union, 2001) in other types of milk.

Separation techniques have also been used for the detection of milk adulteration and both HPLC and GC have been used, often with MS, and are based on the detection of characteristic fatty acids and proteins in dairy products (Romero et al., 1996; Chen et al., 2004; Hurley et al., 2004; Gutierrez et al., 2009). The main disadvantage of these techniques is that they are time consuming (chromatography typically takes 30–60 min) and laborious, and the increased requirements for sample handling during preparation can adversely affect the quality of the analysis (Karoui and Baerdemaeker, 2007).

Finally, PCR techniques have been developed over the last decade, aiming to exploit the presence of somatic cells in milk by detecting genomic DNA from different species. These DNA-based techniques have been used to identify milk adulteration rapidly and with relatively high sensitivity. However, they are not very practical for routine industrial use, and quantification aspects may be hindered by environmental factors such as mastitis that lead to increases in the numbers of somatic cells in milk or by milk processing factors such as milk heat treatment (Bania et al., 2001; Lopez-Calleja et al., 2005; Cheng et al., 2006; Maskova and Paulickova, 2006).

In general, there exists a useful set of analytical approaches for the detection of milk adulteration; however, all the above techniques have the main disadvantage that they are slow and laborious and this delay in milk analysis makes these tools of little value for routine screening of milk in the dairy industry. By contrast, Fourier transform infrared (**FT-IR**) spectroscopy is a very rapid biochemical fingerprinting technique (typically 30 s or faster per sample) that can potentially

resolve many of these problems and produce milk analysis results in under a minute after minimal sample preparation (Nicolaou and Goodacre, 2008). Compared with other techniques it is simple to use, with high sensitivity and low operational costs. When combined with appropriate multivariate statistical methods such as partial least squares (**PLS**) regression or, as we report here, kernel PLS, FT-IR spectroscopy may be an ideal solution for the detection and quantification of the adulteration of milk.

Fourier transform infrared spectroscopy, in combination with PLS, has been used in the past to classify different types of oils (Dahlberg et al., 1997; Ozen and Mauer, 2002) and honey (Hennessy et al., 2008) and detect adulteration of extra virgin oil with palm oil (Rohman and Man, 2010) and hazelnut oil (Ozen and Mauer, 2002) with very good predictive values. In addition, it has been successfully applied in the detection of spoilage in meat such as beef (Ellis et al., 2004) and chicken (Ellis et al., 2002), as well as in cow milk (Nicolaou and Goodacre, 2008). Our aim in this study was therefore to investigate whether FT-IR spectroscopy is an accurate and valid technique for the detection and quantification of the adulteration of goat or sheep milk with cow milk in both binary and tertiary mixtures.

MATERIALS AND METHODS

Sample Preparation

Three types of full-fat, fresh pasteurized milk were used in this study: cow, goat, and sheep milk. The milk samples used were purchased from national retail outlets and analyzed immediately. From these, 4 milk type combinations were prepared: (1) sheep milk adulterated with cow milk; (2) goat milk adulterated with cow milk; (3) sheep milk adulterated with goat milk; and (4) a tertiary mixture containing sheep, goat, and cow milk.

For each of these combinations, various samples were created, with the primary milk type adulterated with a different type of milk from 0 to 100% in successive increasing steps of 5%. The concentration levels for the tertiary mixture (4) are shown in Supplementary Table S1 (<http://www.journalofdairyscience.org/>).

The different milk combinations were then poured into sterile flasks and placed in a rotational incubator for 15 min to ensure a homogeneous mixed sample. One-milliliter milk samples were then obtained and subsequently used for FT-IR analysis.

FT-IR High-Throughput Spectroscopy

A Bruker Equinox 55 infrared spectrometer equipped with a deuterated triglycine sulfate (DTGS) detector

(Bruker Ltd., Coventry, UK) employing a motorized microplate module HTS-XT was used for the FT-IR analysis (Nicolaou and Goodacre, 2008). After a brief mixing of the sample using a rotational mixer for 1 min, 3 μL of each sample was pipetted onto ZnSe infrared plates (sample carriers), and these were then dried in the oven at 50°C for 30 min. Three replicates from each sample were analyzed in randomized order on the sample carrier. The wavelength range in the mid-IR of 4000 to 600 cm^{-1} was employed and 64 scans were co-added and averaged to improve the signal-to-noise ratio; all spectra were collected with a resolution of 4 cm^{-1} . In total, 378 spectra were collected from the different milk combinations and the collection time for each spectrum was approximately 30s.

Data Analysis

Preprocessing. The ASCII data were exported from the Opus software used to control the FT-IR instrument and imported into Matlab ver. 7 (The Mathworks Inc., Natick, MA). The data were preprocessed using standard normal variate (SNV; Barnes et al., 1989, 1993; Dhanoa et al., 1994). To perform this correction, each FT-IR spectrum was first mean centered and then divided by its standard deviation, a process also called autoscaling (Goodacre et al., 2007).

To investigate the relationship between the FT-IR spectra and the concentration of milk, multivariate statistical methods were used including cluster analysis and supervised regression-based techniques.

Unsupervised Analysis. Unsupervised analysis was performed in 2 steps as described fully elsewhere (Goodacre et al., 1998). First, principal component analysis (PCA) was used; PCA is a well-known technique for reducing the dimensionality of multivariate data while preserving most of the variance and results in the creation of uncorrelated variables named principal components (PC). The second step involved the use of discriminant function analysis (DFA); DFA can predict dependent variables from independent variables by discriminating data based on the PC with the prior knowledge of which spectra are machine replicates, a process that does not bias the analysis in any way.

Supervised Analysis. If the targets are known for some of the data, then supervised learning methods can be used where the knowledge of both input (FT-IR spectra) and outputs (% adulteration target values) can be used in the calibration phase. The aim of supervised learning is to construct a robust valid model that correctly associates inputs with outputs. In this study we used PLS and kernel PLS as linear and nonlinear regression techniques, respectively.

Supervised methods are very powerful and it is therefore important to validate both PLS and kernel PLS. To do this for each of the 4 mixtures, a training set and independent hold-out set were generated. For the binary mixtures the training set was 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100% of one of the milks, and the test set was 5, 15, 25, 35, 45, 55, 65, 75, 85, and 95%. The samples used in the training and test sets for tertiary mixtures are shown in the supporting information (Supplementary Table S1; <http://www.journalofdairyscience.org/>).

PLS Regression. Partial least squares is a commonly used multivariate regression method (Martens and Naes, 1989), especially in the field of spectroscopic study. This is because PLS is able to handle effectively the problem of multicollinearity, which is always the case in spectroscopic data, whereas standard multivariate regression will fail because of the rank deficiency problem. Partial least squares can predict either a single predictive variable using a PLS1 model or predict several predictors simultaneously using a PLS2 model. We used PLS1 on the binary milk mixtures (cow–sheep, cow–goat, and sheep–goat milk mixtures), whereas we used PLS2 on the tertiary milk mixture. The number of PLS components was optimized using a k -fold cross-validation on the training set only.

Kernel PLS. We also employed kernel PLS (Shawe-Taylor and Christianini, 2004) as a nonlinear regression method, considering that the response of IR spectra might not always be linear when measuring different compositions of different milks. Kernel PLS is a nonlinear extension of PLS model that made use of the recent development of the concept of kernel learning. The idea of kernel learning is that by projecting the data into an appropriate higher dimensional feature space, many nonlinear problems (e.g., regression, classification) can be solved by using a linear modeling method. The projection is achieved by employing a so-called kernel function. An introduction to various kernel methods can be found in Shawe-Taylor and Christianini (2004). In this study, we employed a radial basis function as the kernel function. The optimal combination of the kernel parameter and the number of PLS factors were optimized using a grid search approach coupled with a k -fold cross-validation on the training set; k is the number of concentration levels we kept in the training set.

RESULTS AND DISCUSSION

Representative FT-IR spectra collected from pure cow, goat, and sheep milks are shown in Figure 1. Overall, at the qualitative level, the spectra look very similar and

this is particularly the case for cow and goat milks. In contrast, some quantitative differences were observed in sheep milk. The first difference involved the CH_2 absorption band at approximately 2927 cm^{-1} , which is related to the acyl chain on fatty acids. As expected, the degree of absorption for this band correlates with the fat quantity in each type of milk, with a higher fat content resulting in higher IR absorption. As reported in the nutritional information that accompanied these milks, full-fat sheep milk has a 4.5% fat composition, whereas cow and goat milk have 3.1 and 3.7% fat, respectively, and this was reflected in the CH_2 stretch.

The other obvious visible differences between milk types appeared on the absorption bands related to the remaining milk components, protein (at 1654 cm^{-1} and 1544 cm^{-1} for amide I and II, respectively) and lactose (at 1159 cm^{-1} and 1076 cm^{-1}). These appeared to be present in higher quantities in sheep milk than in cow and goat milk, with the former displaying higher IR absorption for these bands. Integration of the bands derived from the spectra for each milk type confirm the above differences in the composition of sheep milk with higher absorption values for the CH_2 and C–O absorption bands (Table 1).

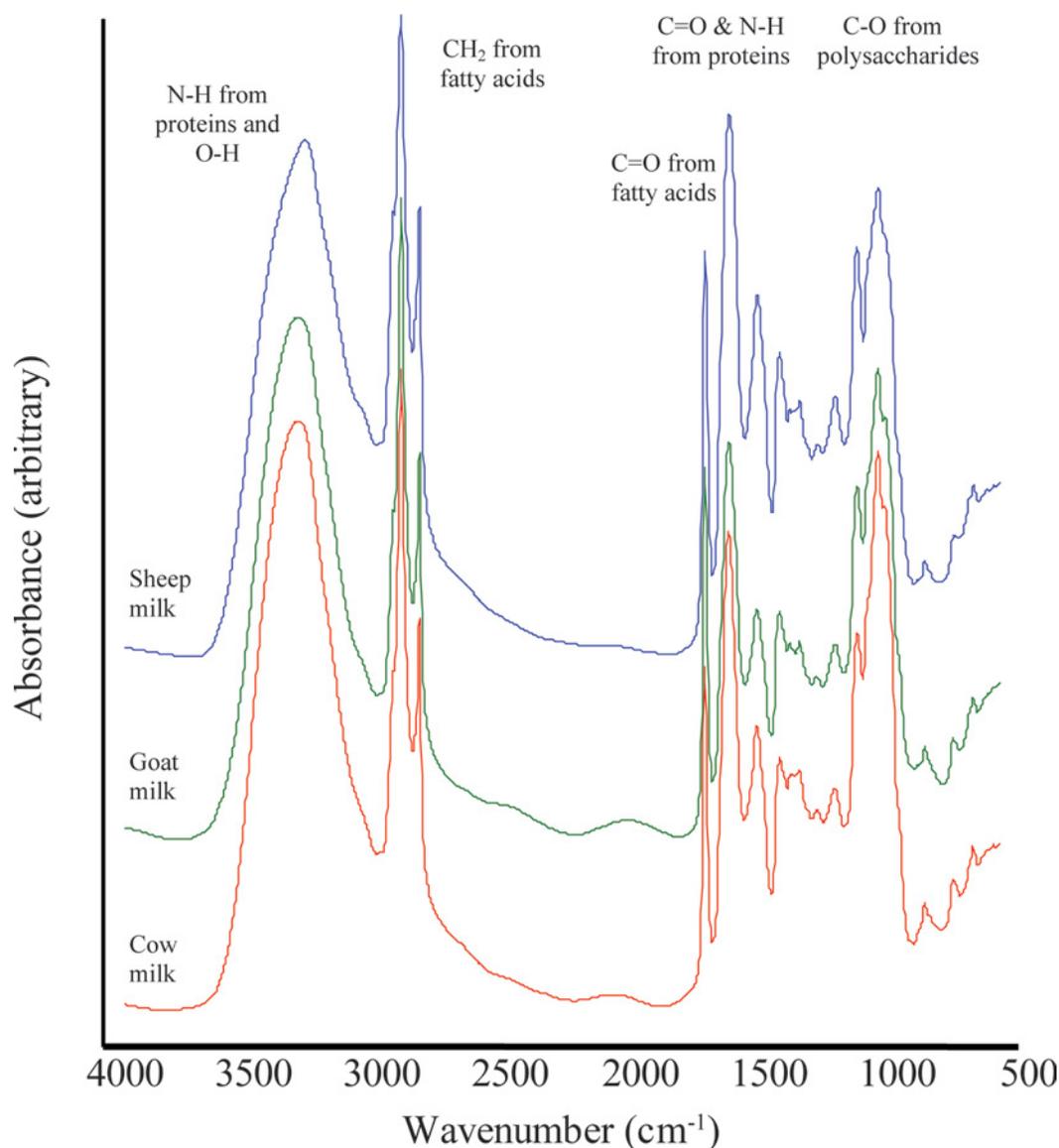


Figure 1. Fourier transform infrared spectra for pure cow, goat, and sheep milk. These spectra are offset to allow visualization of any difference. Color version available in the online PDF.

Table 1. Integrated absorption bands from the 3 measured spectra for each type of milk; numbers are mean \pm standard deviation

Source	CH ₂ from fatty acids, 2976–2884 cm ⁻¹	C–O from polysaccharides, 1134–1018 cm ⁻¹	CH ₂ /C–O
Cow	98.5 \pm 0.59	539.6 \pm 7.15	0.182 \pm 0.0034
Sheep	121.7 \pm 9.32	614.4 \pm 4.67	0.198 \pm 0.0141
Goat	111.2 \pm 0.48	524.5 \pm 2.39	0.212 \pm 0.0005

Analysis of Binary Mixtures of Milk

Because of the subtle differences highlighted above it was not possible to use simple visual inspection to quantify the level of adulteration of these milk. Thus, following spectral collection from the different milk combinations, the relationship between the spectra was investigated using PC-DFA.

Using the combination of goat and sheep milks as an example (Figure 2A), visual inspection of the results identifies a clear trend in PC-DFA space with respect to mixture levels, although this does not follow a linear trajectory, rather a parabolic one. The samples with high concentrations of goat milk are recovered in the same region on the right of the panel; a decrease in the concentration of goat milk, with a concurrent increase in the concentration of sheep milk, appears to cause a spread of the spectra first upward to the middle of the pane and then latterly toward the bottom left of the pane. Inspection of the PC-DFA loadings matrix (Figure 2B) indicates that for the separation of goat and sheep milk, larger differences were observed in the fatty acid vibrations from CH_x stretches in the region 2800 to 3000 cm⁻¹ (and see Figure 1 for annotation) than for either the protein or polysaccharide regions. For the other 2 mixtures (cow–sheep and cow–goat), a parabolic trend in PC-DFA space was also observed (data not shown). It was clear, therefore, that a nonlinear trend existed with respect to milk concentration and we thus sought to explore the use of linear and nonlinear multivariate regression techniques for quantification.

Quantification of Binary Milk Mixtures

As detailed above, 3 binary milk combinations were produced. For brevity, we shall use goat–sheep milk as an example. Samples containing 0 to 100% goat milk (in 5% increments) in sheep milk were prepared and the 21 mixtures were analyzed in triplicate using FT-IR spectroscopy. As detailed above, the data were preprocessed using SNV and were then split into a training set (0, 10, 20, ..., 90, and 100% goat milk) and a test set (5, 15, 25, ..., 85, and 95%) and analyzed by PLS and KPLS.

During calibration of the PLS model, the training data were used and these were subsampled using leave-one-out to generate a cross validation set to choose the optimum number of PLS factors (latent variables) for calibration. Following this, the independent test set was used to challenge the PLS model. The PLS regression results for the goat and sheep milk combinations are shown in Figure 3. In this plot the estimated goat milk concentration versus the known goat milk concentration values follow the expected $y = x$ and gave relatively accurate results. As detailed in Table 2, the root mean squared (RMS) error for the training data (root mean square error for calibration; **RMSEC**) was 3.73%, the cross-validation set selected 3 PLS factors for this model, the RMS error for the cross validation set was 5.57%, and the RMS error in the independent test was 8.03% (root mean square error for predictions in the test set; **RMSEP**). The correlation coefficient for the test set (**Q**²) for this model was 0.92.

Root mean square error for predictions and **Q**² are 2 unbiased metrics to assess the predictive ability of a model in regression analysis. Generally, RMSEP gives an unbiased estimation of the error of the prediction when the model is used to predict the concentration of unknown samples from a data set not used to construct the model (the test set). The smaller the RMSEP value, the better the model. The **Q**² value is an independent metric scale used to quantify the predictive ability of the model with values similar to the commonly used squared correlation coefficient **R**² in univariate regression analysis. The closer the value of **Q**² to 1, the better the model. Thus, the predictions from the goat–sheep milk mixture were relatively encouraging at 8.03% and 0.92 for the RMSEP and **Q**², respectively.

Kernel PLS (**KPLS**) is a nonlinear extension of PLS and we decided to investigate this algorithm because the PC-DFA had shown a nonlinear trajectory. The results for KPLS are also shown in Figure 3 and the associated statistics are given in Table 2. It can be seen that the greatest improvement of using KPLS over PLS was in the model constructed from the goat–sheep milk mixture, with a RMSEP of 3.95% and a **Q**² of 0.98. This improvement of KPLS over PLS was also observed for the cow–sheep and cow–goat mixtures (Figure 3 and

Table 2. Comparison of the partial least squares (PLS) regression, and the nonlinear kernel PLS of the Fourier transform infrared spectra for determining the percentage volume of cow, goat, and sheep milk mixtures¹

Milk mixture	PLS	Kernel PLS
Goat–sheep milk		
Factors	3	3
RMSECV	5.57	3.96
RMSEC	3.73	3.09
RMSEP	8.03	3.95
R ²	0.97	0.99
Q ²	0.92	0.98
Cow–sheep milk		
Factors	3	5
RMSECV	6.87	6.24
RMSEC	6.24	2.30
RMSEP	6.55	5.84
R ²	0.95	0.99
Q ²	0.94	0.95
Cow–goat milk		
Factors	5	5
RMSECV	5.90	5.75
RMSEC	4.07	3.29
RMSEP	7.42	5.62
R ²	0.89	0.97
Q ²	0.91	0.96

¹RMSECV = root mean square error for the cross-validation; RMSEC = root mean square error for the calibration; RMSEP = root mean square error for the predictions produced; R² = correlation coefficient for the training set; Q² = correlation coefficient for the test set.

Table 2), again highlighting the usefulness of employing a nonlinear regression algorithm.

Mixtures of the 3 Types of Milk

Following a similar strategy to that used for the binary mixtures, the first stage in the analysis was to look at the natural relationship between the FT-IR spectra collected from the 63 milk samples containing various concentrations of sheep, goat, and cow milk. The results of the PC-DFA are shown in Figure 4 as a pseudo 3-dimensional plot of the first 3 PC discriminant functions (DF), and it is clear that these spectra are spread in 3 different dimensions. Three color scales are used for each of the milk species: high concentrations of cow milk samples are presented in red; sheep milk is presented in green; and a higher proportion of goat milk is presented in blue.

It can be seen from Figure 4 that when the contribution of one milk type starts to dominate the mixture, a “tentacle” extends from the center of the milk mixtures and the tips of the tentacle contain that pure milk. Domination of these tentacles appears to develop when the concentration of one of the contributing milk types increases above the 55% level. In addition, a fourth tentacle (yellow color) develops when the concentration of cow and sheep milk in the tertiary sample both increase above the 40% contribution level for these milk types.

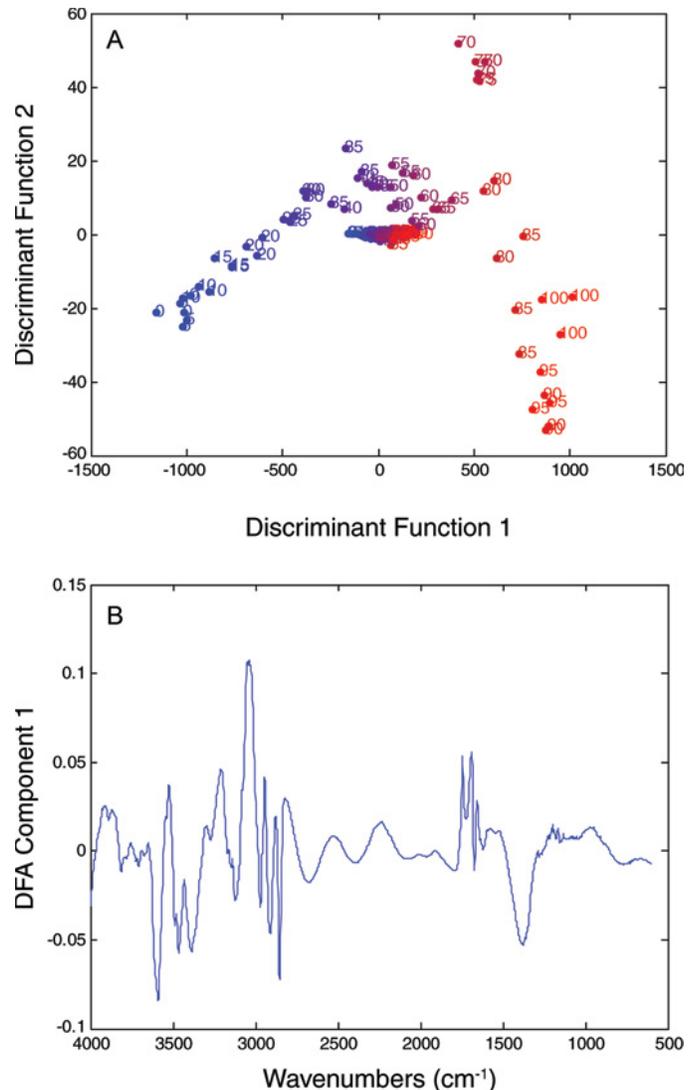


Figure 2. (A) Principal components (PC)-discriminant function analysis (DFA) plot on high-throughput Fourier transform infrared spectra for the 3 repeat experiments of goat milk when added to sheep milk. Principal components 1 to 20 (accounting for 99.45% of the total variance) were used by the DFA algorithm with a priori knowledge of machine replicates. The different numbers show the concentration of goat milk in the mixture and the color (gray) scale the level of adulteration. The blue (black) color indicates a low concentration of goat milk in the mixture, the red (light gray) color indicates a high concentration of goat milk, and the purple (dark gray) color indicates the concentrations where goat and sheep milk have very close concentrations. (B) Principal component-DFA loadings plot from the first vector (discriminant function 1) showing which infrared regions are important: the positive part of DF 1 reflects areas that are increased in goat milk and the negative half of the plot those regions that are higher in sheep milk. Color version available in the online PDF.

In the middle 3 smaller tentacles can be observed, which are colored dark green, turquoise, and purple. These contain samples in which the 3 milk types are in similar concentrations (dark green) and in which 2 milk types are in high concentration; the turquoise represents milk

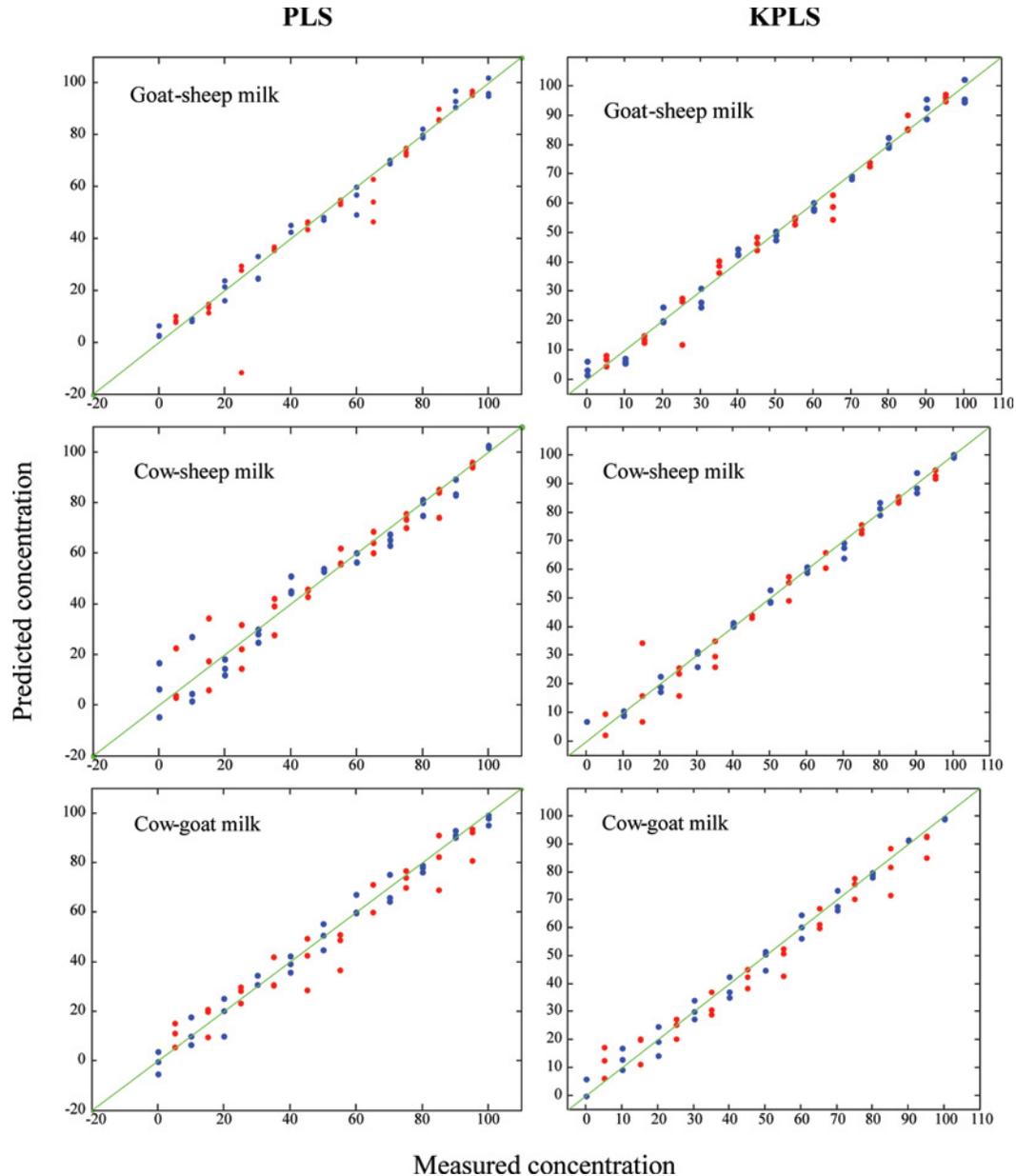


Figure 3. Plot showing the predicted concentrations of goat milk for the first set of mixtures and cow milk predicted concentrations for the second and third mixtures versus the actual concentrations for the 3 different mixtures of milk using partial least squares (PLS) and kernel PLS (KPLS). The blue (black) circles represent the training data set and the red (gray) circles the test set. Color version available in the online PDF.

with a high concentration of sheep–goat milk and the purple represents milk with a high concentration of cow–goat milk (a similar plot giving the concentration levels of each milk is shown in Supplementary Figure 1; <http://www.journalofdairyscience.org/>). In conclusion, the cluster analysis shows that clear trends exist in these data related to the various combinations of milk but that the ability to quantify the level of milks in these tertiary mixtures is unlikely from the PC-DFA, given the complexity of this space.

Quantification of Tertiary Milk Mixtures

The data were split into training and test set data and then analyzed using linear and nonlinear supervised learning techniques; as 3 milk concentrations were to be predicted we employed 3 output Y-variables (one for each type of milk) in PLS2 and KPLS2. The results from the PLS2 model for the training and test data are shown in Figure 5 and Figure 6, respectively, and overall showed a good prediction for the 3 different types of

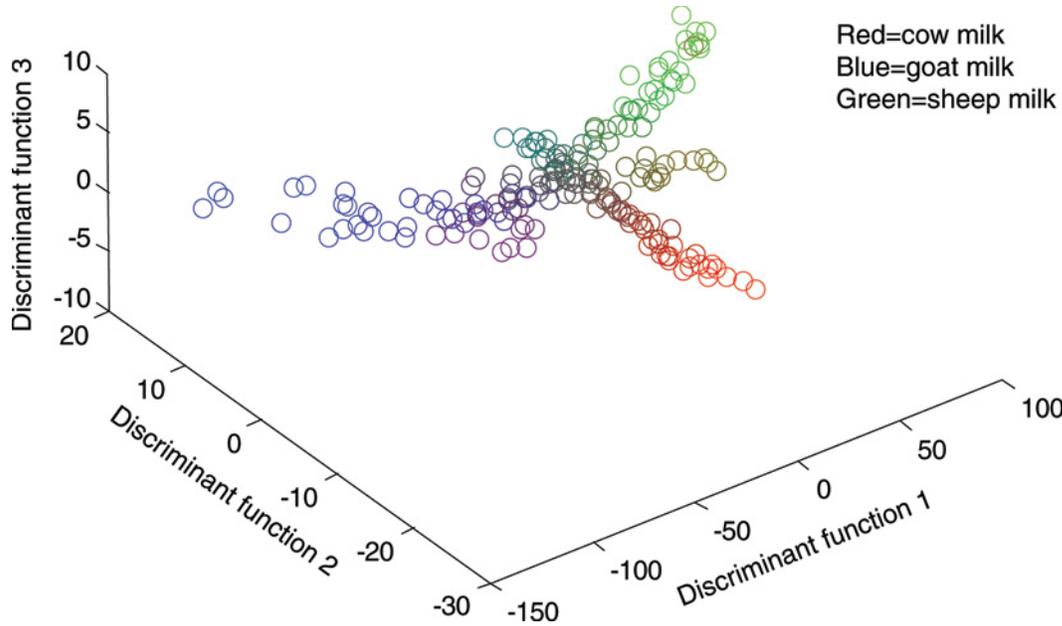


Figure 4. Principal components (PC)-discriminant function analysis (DFA) plot on high-throughput Fourier transform infrared spectra for the 3 repeat experiments of the mixtures from the 3 types of milk. Principal components 1 to 20 (accounting for 99.45% of the total variance) were used by the DFA algorithm with a priori knowledge of machine replicates (i.e., 1 class per time point, giving 63 classes in total). Samples with high concentrations of cow milk are presented in red, those with high concentrations of sheep milk are green, and those with high concentrations of goat milk are blue. In the middle appear to be samples that contain 2 out of the 3 milk types in high concentrations and these samples are a different color from the combination of all 3 types of milk.

milk. Table 3 shows the summary statistics for PLS2 and KLP2 using the same training and test set splits and, in this case, the PLS2 algorithm outperformed KLP2.

In general, all milks were predicted with a similar level of accuracy and the RMSEP values were 3.4 to 4.9% for PLS2 with Q^2 correlations of 0.94 to 0.97.

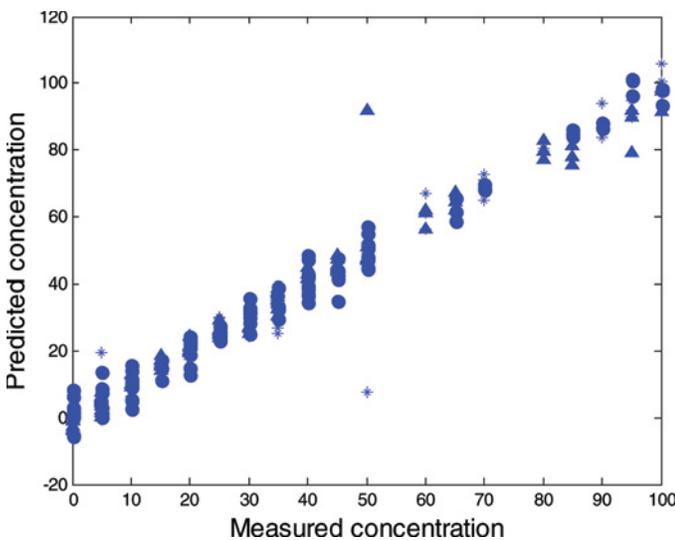


Figure 5. Plot showing the predicted concentrations of cow, goat, and sheep milk versus the actual concentrations for the training set in the mixtures of the 3 types of milk by using partial least squares 2. The circles represent the goat milk samples, triangles the sheep milk samples, and the crosses cow milk samples. Color version available in the online PDF.

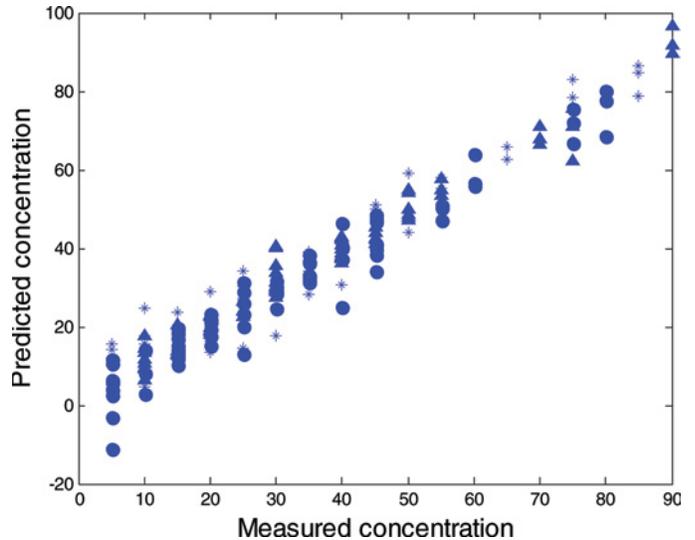


Figure 6. Plot showing the predicted concentrations of cow, goat and sheep milk versus the actual concentrations for the test set in the mixtures of the 3 types of milk by using partial least squares 2. The circles represent the goat milk samples, the triangles the sheep milk samples, and the crosses cow milk samples. Color version available in the online PDF.

Table 3. Comparison of the partial least squares 2 (PLS2) regression, and the nonlinear kernel PLS 2 (KPLS2) of the Fourier transform infrared spectra for determining the percentage volume of cow, goat, and sheep milks from mixtures containing the 3 types of milk together¹

Item	Cow milk	Sheep milk	Goat milk
PLS2			
Factors		10	
RMSECV	5.63	3.10	4.96
RMSEC	5.53	4.94	3.67
RMSEP	4.89	3.36	4.83
R ²	0.95	0.96	0.98
Q ²	0.94	0.97	0.94
KPLS2			
Factors		16	
RMSECV	6.05	3.97	4.91
RMSEC	4.20	3.60	2.84
RMSEP	6.40	5.61	3.89
R ²	0.97	0.98	0.98
Q ²	0.92	0.93	0.96

¹RMSECV = root mean square error for the cross-validation; RMSEC = root mean square error for the calibration; RMSEP = root mean square error for the predictions produced; R² = correlation coefficient for the training set; Q² = correlation coefficient for the test set.

CONCLUSIONS

In this study, we demonstrated that in binary and tertiary mixtures of milk, FT-IR spectroscopy in combination with multivariate analysis, such as linear PLS and nonlinear Kernel PLS, provides an accurate, simple, and rapid technique for the quantitative assessment of the adulteration of sheep, goat, and cow milks. The typical errors that were found were in the region of 5 to 8% for all milk species; an unscrupulous producer would unlikely adulterate at such a level because it would not be financially viable. For this reason and the speed of analysis (30 s per sample in batches of 96 or 384), we believe that FT-IR spectroscopy has excellent potential for use in the food industry to replace less efficient and more time-consuming techniques for the detection of milk adulteration. For this approach to be employed within the industry, future studies should consider geographical as well as seasonal variations in milk production.

ACKNOWLEDGMENTS

YX and RG acknowledge the Symbiosis-EU (www.symbiosis-eu.net) project (no 211638) financed by the European Commission under the 7th Framework programme for RTD. The information in this document reflects only the authors' views and the European Community is not liable for any use that may be made of the information contained therein.

REFERENCES

- Addeo, F., R. Pizzano, M. A. Nicolai, S. Cairra, and L. Chianese. 2009. Fast isoelectric focusing and antipeptide antibodies for detecting bovine casein in adulterated water buffalo milk and derived mozzarella cheese. *J. Agric. Food Chem.* 57:10063–10066.
- Amigo, L., M. Ramos, L. Calhau, and M. Barbosa. 1992. Comparison of electrophoresis, isoelectric-focusing, and immunodiffusion in determinations of cows and goats milk in serra-da-estrela cheeses. *Lait* 72:95–101.
- Anguita, G., R. Martin, T. Garcia, P. Morales, A. Haza, I. Gonzalez, B. Sanz, and P. Hernandez. 1996. Immunostick ELISA for detection of cow's milk in ewe's milk and cheese using a monoclonal antibody against bovine β -casein. *J. Food Prot.* 59:436–437.
- Anguita, G., R. Martin, T. Garcia, P. Morales, A. I. Haza, I. Gonzalez, B. Sanz, and P. E. Hernandez. 1997. Detection of bovine casein in ovine cheese using digoxigenated monoclonal antibodies and a sandwich ELISA. *Milchwissenschaft* 52:511–513.
- Bania, J., M. Ugorski, A. Polanowski, and E. Adamczyk. 2001. Application of polymerase chain reaction for detection of goats' milk adulteration by milk of cow. *J. Dairy Res.* 68:333–336.
- Barnes, R. J., M. S. Dhanoa, and S. Lister. 1993. Correction to the description of standard normal variate (SNV) and de-trend (DT) transformation in practical spectroscopy with applications in food and beverage analysis, 2nd ed. *J. Near Infrared Spectrosc.* 1:185–186.
- Barnes, R. J., M. S. Dhanoa, and S. J. Lister. 1989. Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* 43:772–777.
- Beer, M., I. Krause, M. Stapf, C. Schwarzer, and H. Klostermeyer. 1996. Indirect competitive enzyme-linked immunosorbent assay for the detection of native heat-denatured bovine β -lactoglobulin in ewes' and goats' milk cheese. *Z. Lebensm. Unters. Forsch.* 203:21–26.
- Bischoff, S. C. 2006. Food allergies. *Curr. Gastroenterol. Rep.* 8:374–382.
- Chafen, J. J., S. Newberry, M. Riedl, D. Bravata, M. Maglione, M. Suttorp, V. Sundaram, N. Paige, A. Towfigh, B. Hulley, and P. Shekelle. 2010. Diagnosing and managing common food allergies: A systematic review. *JAMA* 303:1848–1856.
- Chen, R.-K., L.-W. Chang, Y.-Y. Chung, M.-H. Lee, and Y.-C. Ling. 2004. Quantification of cow milk adulteration in goat milk using high-performance liquid chromatography with electrospray ionisation mass spectrometry. *Rapid Commun. Mass Spectrom.* 18:1167–1171.
- Cheng, Y.-H., S.-D. Chen, and C.-F. Weng. 2006. Investigation of goats' milk adulteration with cows' milk by PCR. *Asian-australas. J. Anim. Sci.* 19:1503–1507.
- Dahlberg, D. B., S. M. Lee, S. J. Wenger, and J. A. Vargo. 1997. Classification of vegetable oils by FT-IR. *Appl. Spectrosc.* 51:1118–1124.
- Dennis, M. 1998. Recent developments in food authentication. *Analyst (Lond.)* 123:151R–156R.
- Dhanoa, M., S. Lister, R. Sanderson, and R. J. Barnes. 1994. The link between multiplicative scatter correction (MSC) and standard normal variate (SNV) transformations of NIR spectra. *J. Near Infrared Spectrosc.* 2:43–47.
- European Union. 2001. EC 213/2001: Methods for the analysis and quality evaluation of milk and milk products. *Off. J. L* 44:37/1–L37/99.
- Ellis, D. I., D. Broadhurst, B. K. Douglas, J. J. Rowland, and R. Goodacre. 2002. Rapid and quantitative detection of the microbial spoilage of meat by Fourier transform infrared spectroscopy and machine learning. *Appl. Environ. Microbiol.* 68:2822–2828.
- Ellis, D. I., D. Broadhurst, and R. Goodacre. 2004. Rapid and quantitative detection of the microbial spoilage of beef by Fourier transform spectroscopy and machine learning. *Anal. Chim. Acta* 514:193–201.

- Goodacre, R., D. Broadhurst, A. Smilde, B. S. Kristal, J. D. Baker, R. Beger, C. Bessant, S. Connor, G. Capuani, A. Craig, T. Ebbels, D. B. Kell, C. Manetti, J. Newton, G. Patemostro, R. Somorjai, M. Sjöström, J. Trygg, and F. Wulfer. 2007. Propose minimum reporting standards for data analysis in metabolomics. *Metabolomics* 3:231–241.
- Goodacre, R., E. M. Timmins, R. Burton, N. Kaderbhai, A. M. Woodward, D. B. Kell, and P. J. Rooney. 1998. Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks. *Microbiology* 144:1157–1170.
- Gutierrez, R., S. Vega, G. Diaz, J. Sanchez, M. Coronado, A. Ramirez, J. Perez, M. Gonzalez, and B. Schettino. 2009. Detection of non-milk fat in milk by gas chromatography and linear discriminant analysis. *J. Dairy Sci.* 92:1846–1855.
- Hennessy, S., G. Drowney, and C. O'donnell. 2008. Multivariate analysis of attenuated total reflection-Fourier transform infrared spectroscopy data to confirm the origin of honeys. *Appl. Spectrosc.* 62:1115–1123.
- Hurley, I. P., R. C. Coleman, H. E. Ireland, and J. H. H. Williams. 2006. Use of sandwich IgG ELISA for the detection and quantification of adulteration of milk and soft cheese. *Int. Dairy J.* 16:805–812.
- Hurley, I. P., H. E. Ireland, R. C. Coleman, and H. H. H. Williams. 2004. Application of immunological methods for the detection of species adulteration in dairy products. *Int. J. Food Sci. Technol.* 39:873–878.
- Karoui, R., and J. D. Baerdemaeker. 2007. A review of the analytical methods coupled with chemometric tools for the determination of the quality and identity of dairy products. *Food Chem.* 102:621–640.
- Levieux, D., and A. Venien. 1994. Rapid, sensitive 2-site ELISA for detection of cows' milk in goats' or ewes' milk using monoclonal antibodies. *J. Dairy Res.* 61:91–99.
- Lopez-Calleja, I., I. Gonzalez, V. Fajardo, I. Martin, P. E. Hernandez, T. Garcia, and R. Martin. 2005. Application of polymerase chain reaction to detect adulteration of sheep's milk with goats' milk. *J. Dairy Sci.* 88:3115–3120.
- Malin, E., J. Basch, J. Shieh, B. Sullivan, and V. Holsinger. 1994. Detection of adulteration of buttermilk powder by gel-electrophoresis. *J. Dairy Sci.* 77:2199–2206.
- Martens, H., and T. Naes. 1989. *Multivariate Calibration*. 1st ed. John Wiley & Sons, Chichester, UK.
- Maskova, E., and I. Paulickova. 2006. PCR-based detection of cow's milk in goat and sheep cheeses marketed in the Czech Republic. *Czech J. Food Sci.* 24:127–132.
- Maudet, C., and P. Taberlet. 2001. Detection of cows' milk in goat's cheeses inferred from mitochondrial DNA polymorphism. *J. Dairy Res.* 68:229–235.
- Mayer, H. K. 2005. Milk species identification in cheese varieties using electrophoretic, chromatographic and PCR techniques. *Int. Dairy J.* 15:595–604.
- Nicolaou, N., and R. Goodacre. 2008. Rapid and quantitative detection of the microbial spoilage in milk using Fourier transform infrared spectroscopy and chemometrics. *Analyst* 133:1424–1431.
- Ozen, B. F., and L. J. Mauer. 2002. Detection of hazelnut adulteration using FT-IR spectroscopy. *J. Agric. Food Chem.* 50:3898–3901.
- Rohman, A., and Y. B. C. Man. 2010. Fourier transform infrared (FTIR) spectroscopy for analysis of extra virgin olive oil adulterated with palm oil. *Food Res. Int.* 43:886–892.
- Romero, C., O. Perez-Andujar, A. Olmedo, and S. Jimenez. 1996. Detection of cow's milk in ewe's or goat's milk by HPLC. *Chromatographia* 42:181–184.
- Shatenstein, B., and P. Ghadirian. 1998. Influence on diet, health behaviours and their outcome in select ethnocultural and religious groups. *Nutrition* 14:223–230.
- Shawe-Taylor, J., and N. Christianini. 2004. *Kernel Methods for Pattern Analysis*. Cambridge University Press, Cambridge, UK.
- Venter, C. 2009. Cow's milk protein allergy and other hypersensitivities in infants. *J. Fam. Health Care* 19:128–134.