

Efficacy of near infrared spectroscopy to segregate raw milk from individual cows between herds for product innovation and traceability

K. K. Ejeahalaka¹, L. Cheng², D. Kulasiri¹, G. R. Edwards³ and S. L. W. On^{1*}

¹Department of Wine, Food and Molecular Biosciences, Lincoln University, P.O. Box 85084, Canterbury, New Zealand;

²Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Dookie Campus, Victoria 3647, Australia;

³Department of Agricultural Sciences, Lincoln University, P.O. Box 85084, Canterbury, New Zealand

Corresponding Author: S. L. W. On, stephen.on@lincoln.ac.nz

Received: 3 December 2019 / Accepted: 5 April 2020 / Published: 4 July 2020.

© 2020 Codon Publications

OPEN ACCESS 

RESEARCH ARTICLE

Abstract

Cows with specialised characteristics and requirements can be aggregated into different herds for targeted nutritional management and to facilitate on-farm segregation of raw milk for the production of high-value niche dairy products, offering improved economic returns. Rapid methods for independent verification of product quality and origin are desirable to support validation and traceability of such products. This study examined the use of near infrared spectroscopy (NIRS) to segregate raw milk from individual cows of multiple breeds from different herds fed on the same or differing feeding regimes, and to correlate and evaluate the efficacy of the predictions for crude protein and the milk fatty acid (FA) phenotypes for each of the herds. Reference values and near infrared spectra were obtained from representative freeze-dried raw milk samples ($n = 220$) collected from 847 lactating cows of 3 breeds from the Lincoln University dairy farm in New Zealand. The feed sources (i.e. pasture or pasture with lucerne silage) significantly influenced the protein and the FA values, and these differences were reflected in NIRS analyses. The partial least square regression models for crude protein determination showed excellent results, whereas for the most dominant FA, they were not appreciable. Maximum separation was obtained between the herds on the same feeding regime (mean specificity = 95.2%) using the partial least square discriminant analysis, and its overall performance in differentiating the objects was better than that of the soft independent modelling of class analogy. The multiclass analyses conducted in this study offer improvements to current approaches for evaluating and validating raw milk for the manufacture of specific dairy products, and for enhancing product traceability.

Keywords: milk segregation, near infrared spectroscopy, partial least square, soft independent modelling

1. Introduction

The composition of raw milk is influenced by breed, nutrition, health status and time of lactation of individual cows in the herds (Katz *et al.*, 2016). Consequently, the composition of bulk milk (i.e. the sum of the milk contributed by all the animals that milked into the storage vats) can also vary. The recent volatility in the global dairy trade has prompted producers to reduce dependency on low value bulk commodities, to focus on premium segregated

milk products that meet the consumer preferences. There is significant potential in segregating raw milk between herds to assist the manufacture of specific/differentiated/high value dairy products (Dooley *et al.*, 2005), as well as providing quality assurance and traceability benefits. However, on-farm 'made-to-order' segregation of raw milk by herds is not yet a common practice by farmers in New Zealand, despite these commercial and logistic advantages. Nevertheless, the key nutrients of fat and protein in cow milk vary tremendously between herds,

and they can be influenced through nutrition and feeding management (Heinrichs *et al.*, 1997). In recent years, consumers have come to consider animal diet as a major criterion to estimate the quality of dairy products (Andueza *et al.*, 2013), with those based on grass enjoying positive image and demand (Martin *et al.*, 2004). Thus, to meet the growing demand of consumers who are prepared to pay premium prices for tailored products, grass-fed milk may be differentiated between herds as such segregation imbues value. In addition, tracing the herds of origin of raw milk may be valuable in the instances of fraud and contamination. There remains a need for analytical methods to independently verify milk quality prior to sale or distribution.

Several researchers have successfully characterised the nutritional quality of raw milk from different feeding systems based on their dominant FA profiles (Collomb *et al.*, 2002; Hurtaud *et al.*, 2014). However, those characterisation involved cumbersome and expensive analytical reference methods such as gas chromatography (GC) which are generally less suitable for extensive routine authentication of dairy products (Coppa *et al.*, 2012).

Multi-parametric instruments based on near infrared spectroscopy (NIRS) have been developed to provide precise, rapid, low-cost and non-destructive analytical solutions, and they have been shown to have potential when combined with chemometric tools for characterising cow milk according to feeding systems and origin (Coppa *et al.*, 2012). However, those studies used bulk milk samples collected from different regions without any reference to the variances between individual cows and their herds of origin. Consequently, there are no guarantees that the baseline quality of such samples was preserved right from milking until the commencement of the laboratory analysis. It is well known that the genetic background of any given herd could also affect the milk traits (i.e. within breed variation) (Marchitelli *et al.*, 2013). So far, to the best of our knowledge, no studies have been undertaken to assess the potential of NIRS to differentiate the nutritional quality of cow milk from different herds on an ongoing basis at individual animal level. In our study, we applied NIRS to freeze-dried raw milk samples from individual dairy cows from mixed breeds, which were aggregated into different herds under differing feeding regimes, in a carefully controlled rearing environment in New Zealand and examined the classification and correlation with FA and crude protein contents, to assess the utility of the approach in differentiating milk quality between herds. Thus, in this article, different chemometric methods such as principal component analysis (PCA), soft independent modelling of class analogy (SIMCA), partial least square discriminant analysis (PLS-DA) and partial least square regression (PLSR) were adopted to

analyse and to make key decisions from the near infrared spectral data sets.

2. Materials and methods

Sample collection, handling and preparation

A total of 50 mL sample of fresh raw milk was collected during the morning (0600 h) and afternoon (1,430 h) milking processes from each of the 220 lactating cows from four herds (i.e. 56 in herd 1, 64 in herd 2, 49 in herd 3 and 51 in herd 4) on differing feeding regimes using inline milk meters (DeLaval International AB, Tumba, Sweden) as part of a 2-year biomonitoring study which commenced in September 2013 at Lincoln University Research Dairy Farm. The dairy cows (Friesian, Jersey and Friesian x Jersey) in herds 1 and 2 calved in late winter (August), whereas those in herds 3 and 4 calved in mid-spring (October). The cows were of mixed age (between 3 and 10 years old), in good health and in the middle of their lactation period. The mean liveweights for the cows in herds 1, 2, 3, and 4 were 507 ± 48.0 kg, 477 ± 55.0 kg, 499 ± 45.0 kg and 474 ± 50.0 kg, while their days in milk (DIM) were 273 ± 25.0 days, 259 ± 30.0 days, 265 ± 23.0 days and 263 ± 27.0 days, respectively. The treatment grouping was equitably done to ensure that there was a balance within the herds for calving date, milk yield, DIM and liveweight while maximising the spread between herds. Herds 1 and 2 were fed mainly pasture with the addition of lucerne (*Medicago sativa*, alfalfa) silage during the 2013–2014 milking season, whereas herds 3 and 4 were fed only pasture during the 2014–2015 season. The pasture was composed mainly of standard mixtures of perennial ryegrass (*Lolium perenne* L.) and white clover (*Trifolium repens*) base. All cows in the same herd were fed with the same diet throughout the milking season. The main difference between the herds was the style of their nutritional management which in principle was designed to meet the herd-specific requirements in order to optimise efficiency and milk quality. The raw milk samples, collected from individual cows, were tagged with their identities and thoroughly mixed to ensure uniformity. Then, they were frozen at -21 °C without any preservative before freeze-drying (Cuddon Ltd New Zealand, Model E.D.5.3) at 4 °C in preparation for spectra measurements using the near infrared spectrophotometer. The freeze-dried milk samples were stored immediately at 4 °C until analysis.

Compositional analysis by primary methods

Milk crude protein determination

Crude protein content was measured in this study using the combustion (Dumas) method implemented using the Rapid Max N Exceed Elemental Analyzer (Elementar

GmbH, Germany). In brief, about 200 mg of the freeze-dried milk samples was weighed in duplicates into the stainless-steel crucibles and combusted at 900 °C in an oxygen atmosphere. The combustion process converted any elemental nitrogen into N₂ and NO_x; NO_x moieties were then reduced to N₂, while the other volatile combustion products were trapped and eliminated. Thereafter, the N₂ gases were passed through a thermal conductivity cell in order to measure the total nitrogen. The crude protein content of the sample was then calculated by applying the nitrogen-to-protein conversion factor of 6.38, and it was expressed in percent (or g per 100 g) of freeze-dried milk powder.

Milk fatty acid determination

The method employed for the determination of FA concentrations was based on two major phases, namely one-step methylation and GC analysis as previously described (Ejeahalaka and On, 2019a). In brief, 100 µL of internal standard (C21:0 ester) was added into each of the two Kimax tubes containing 0.15 g of the freeze-dried milk samples. Then, 900 µL of heptane and 4.0 mL of 0.5 M NaOH/dried methanol were added into the tubes followed by the addition of 2.0 mL of heptane and 2.0 mL of distilled water. The tubes were then capped, vortexed and centrifuged to separate the heptane layer of the FA esters.

The GC analysis was carried out by injecting 1.0 µL of the extracted FA methyl esters (FAME) onto a Varian CP7420, tailor-made fused silica capillary column using the AOC-20i auto-sampler fitted to a Shimadzu GC-2010 gas chromatograph. The individual fatty acids in the sample were quantified in milligrams using the GC peak areas of the internal (C21:0) and external standards together with the mass of internal standard in the sample and the response factor of the individual fatty acids. Results were expressed in absolute terms in grams of fatty acid per 100 g of freeze-dried milk samples by dividing the individual FA in milligrams by the mass of the sample in grams.

Spectral measurements

Near infrared spectroscopic measurements were carried out on the freeze-dried milk samples using a FOSS NIRSystem (model DS 2500F) scanning spectrometer as previously described (Ejeahalaka and On, 2019a). In brief, 5 g of each of the 220 samples was placed in a 10 mL DS 2500 ring cup (cup type: 2004) and scanned at 0.5 nm intervals in the wavelength range of 850–2,500 nm in triplicates. A total of 3,300 absorbance values were captured per NIRS spectrum averaging 32 scans. The triplicate spectra acquired for each sample were averaged and then subjected to multivariate chemometric data analysis after undergoing the necessary pretreatments.

Spectral preprocessing and chemometric analysis

Open source software, R Project for Statistical Computing, version 3.5.1 (R Core Team, 2018), was used to process the spectral data and to develop models for exploring the resolution of the freeze-dried raw milk samples to herd level. It was also used to analyse the significant differences between the mean values of the reference data (i.e. for crude protein and the FA). The NIRS data were first subjected to PCA after mean-centring to evaluate potential relationships or groupings of readings obtained from the milk samples derived from the different animals and herds examined. PCA has been shown to be useful in identifying how one sample is different from another and which variables contribute most to this difference, and whether those variables contribute in the same way (i.e. are correlated) or independently (i.e. uncorrelated) from each other (Lavine and Workman, 2004; Wishart, 2007). Consequently, the PCA loadings were plotted to identify the most sensitive wavelengths in the spectral data. To build the classification rules and to predict the reference (i.e. crude protein and FA) concentration values, the raw spectra data were preprocessed using Savitzky–Golay (SG) algorithm (Savitzky and Golay, 1964) (window size = 55 points, second-order polynomial fit) followed by the Extended Multiplicative Signal Correction (EMSC) method (Martens and Stark, 1991) to preserve higher moments, remove baseline effects and suppress wavelength-dependent light-scattering variations. Zimmermann and Kohler (2013) demonstrated that the use of SG with EMSC for preprocessing results in simpler and often better models, since the SG differentiation effectively suppresses the broad underlying baselines, while the EMSC principally has the feature of removing the multiplicative effect. Zimmermann and Kohler (2013) further showed that EMSC generally performed better on SG differentiated spectroscopic data than the multiplicative signal correction (MSC) because of the lack of higher terms in the MSC model as opposed to the linear and quadratic terms in the EMSC model. To proceed with the analyses, the preprocessed spectra data were split randomly in the ratio of 4:1 into training sets (i.e. 42 in herd 1, 48 in herd 2, 37 in herd 3 and 38 in herd 4) to build the calibration models and test sets (i.e. 14 in herd 1, 16 in herd 2, 12 in herd 3 and 13 in herd 4) to measure the predictive ability of the models on unseen samples. The leave-one-out cross-validation method was used for internal validation of the training data in order to choose the models that will perform best on the test sets.

Quantitative prediction of raw milk phenotypes by herds

Partial least square regression is used for relating and modelling the structure of two data matrices, X and Y, with many noisy, collinear and even incomplete variables (Wold *et al.*, 1984). This supervised multivariate

technique mathematically correlates the spectral intensities at different wavelengths (called the absorbance values) as the independent X variables to the reference sample composition (e.g. crude protein or fatty acid contents) as the dependent Y variable. The power of PLSR lies in its ability to derive latent variables (LV) or factors (i.e. linear combinations of the X variables) that correlate to as great an extent as possible with the Y variable as well as having maximal predictive power on Y (Meilgaard *et al.*, 1999; Wold, 1966). Two approaches can be followed in carrying out a PLSR analysis, namely PLS1 (used when only one Y variable is predicted) and PLS2 (used when several Y variables are predicted). In this study, PLS1 approach was used for modelling the crude protein contents whereas PLS2 was used for modelling the selected five of the most dominant FA, that is, palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1 c9), rumenic acid (C18:2 c9,t11), and α -linolenic acid (C18:3 c9,c12,c15) and the FA groups namely: saturated (SFA) and the unsaturated fatty acids (UFA) in the freeze-dried raw milk samples. The models were built using the full spectra range (850–2,500 nm) and the windows of spectral wavelengths that contained the most relevant information as determined by interval-PLS (Nørgaard *et al.*, 2000). The predictive performances of the regression models were evaluated using the root mean squared error-observations standard deviation ratio (RSR) as previously described (Ejeahalaka and On, 2019b). Although lower values of the root mean squared error of prediction (RMSEP) indicate higher model predictive accuracy, RSR provides a delimiter of what is considered a low RMSEP based on the standard deviation of the measured values (Moriassi *et al.*, 2007). RSR standardises RMSEP using the observations standard deviation (Moriassi *et al.*, 2007), and as such, it is the reciprocal of the residual prediction deviation (RPD) of the model. RSR varies from the optimal value of 0, which indicates zero RMSEP and therefore perfect model prediction, to a large positive value (Moriassi *et al.*, 2007). However, as RPD > 2 is desired for a good calibration (Karoui *et al.*, 2006), so also is RSR < 0.5.

Raw milk near infrared segregation by herds

PLS-DA and SIMCA were employed to segregate the raw milk samples into their herds of origin. PLS-DA is a supervised discriminant classification approach requiring at least two classes to be defined (Barker and Rayens, 2003), and it is essentially based on the PLS2 algorithm that searches for LV with a maximum covariance with Y variables (Wold, 1966). However, the Y-block in PLS-DA is a qualitative variable as it describes which objects are in the classes of interest (Wold, 1966). On the other hand, SIMCA is specifically designed to study and describe one single class at a time, and it checks for compatibility of unknown samples with the class being modelled (Oliveri and Downey, 2012;

Wold and Sjöström, 1977). SIMCA is a supervised classification (or pattern recognition) method that helps to demarcate the boundaries of predefined classes and to assign future objects to those classes. In this work, the PLS-DA and the SIMCA models were built using both the full spectral range (850–2,500 nm) and the subset of the original data set covering the sensitive bands selected from the PCA loadings. The classification performances of the models were evaluated using the sensitivity, specificity, correct classification rate, precision, and the Matthews correlation coefficient as previously defined in the literature (Ballabio and Todeschini, 2009; Oliveri, 2017).

3. Results and discussion

Characteristics of NIRS profiles

Figure 1 shows the mean NIRS values of the freeze-dried raw milk samples from individual cows from each of the four herds in the near infrared region from 850 to 2,500 nm. Spectral profiles of milk from each of the herds displayed similar absorption characteristics, with some features typical for NIRS analyses of oven-dried goat milk (Núñez-Sánchez *et al.*, 2016) and full cream milk powder (Frankhuizen, 2001). However, differences between milk from herd types receiving differing nutrition were seen, with generally higher absorbance values observed in milk from cows fed on pasture alone. This finding correlates with those of Mouazen *et al.* (2009) who reported higher absorbance values for pasture feeding as compared to soya-bean feeding in their study of the properties of ewe milk. These authors described the milk as having higher density and lesser water content.

Based on previous research (Frankhuizen, 2001), six bands that are most useful for analysing the major constituents of milk powders are distinguishable in Figure 1 as indicated at about 1,649, 1,722, 1,758, 1,934, 2,218 and 2,308 nm in each of the spectra. The PCA loadings plot (Figure 3) also identified the range from 1,580 to 2,305 nm as the optimal wavelengths and therefore was included in our models. It is likely that the constituents contributing to absorption at these bands correspond to water, protein and lactose (1,650 nm); protein and fat (1,734 nm); fat and protein (1,759 nm); water (1,940 nm); fat, protein and lactose (2,230 nm); and fat and protein (2,310 nm) as previously reported (Frankhuizen, 2001). Also, the three typical water absorption bands that characterise the milk spectrum are distinguishable at about 950 nm, 1,445 nm and 1,934 nm.

NIRS analysis with PCA

The PCA performed on the 3,300 independent spectral variables (Figure 2) of the freeze-dried milk powder

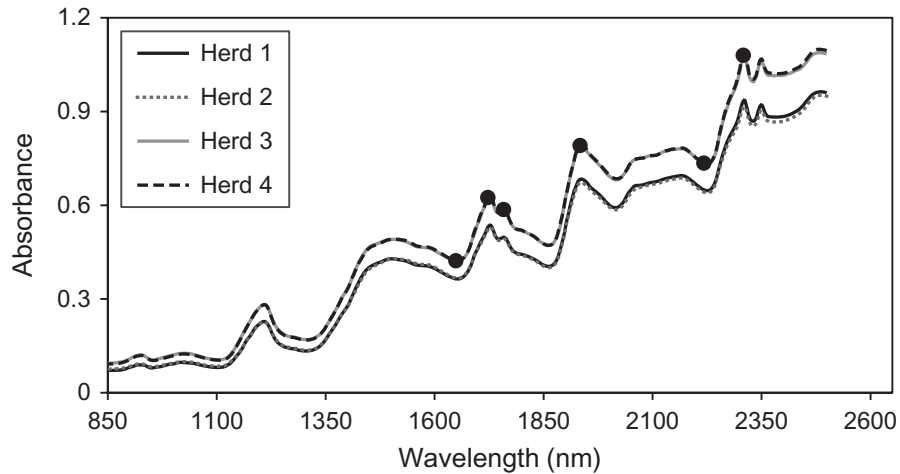


Figure 1. Average near infrared spectra of the freeze-dried raw milk for each of the herds (• = most useful selected bands for milk analytics).

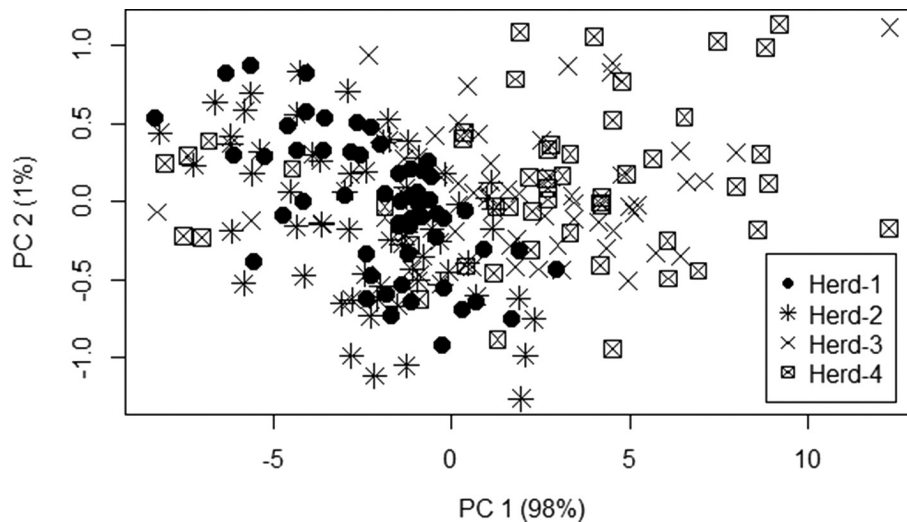


Figure 2. PCA Score plot of the near infrared spectra for the 4 herds of the dried raw milk samples.

supported the mean profile analysis in displaying substantive separation of samples taken from herds fed with different sources. The score plots showed an overlap between the objects of herds 1 and 2 and also between those of herds 3 and 4, thus requiring the use of multivariate classification methods (i.e. SIMCA and PLS-DA) for their possible differentiation.

The loadings variable plot showing the wavelengths that had the largest effect on PC 1 and PC 2 is presented in Figure 3. The identified wavelengths include 921, 1,210, 1,390, 1,580, 1,722, 1,945, 2,100, 2,305 and 2,345 nm. However, only the wavelength range 1,580–2,305 nm was considered optimal for building the chemometric models as it yielded the most stable predictions.

Reference data analysis and the PLSR modelling of milk phenotypes

It can be seen from Table 1 that the total UFA for herds 2 and 3 were statistically higher than those for herds 1 and 2, whereas the total SFA remained the same for all the herds.

The inclusion of lucerne in the diet resulted in a statistically significant decrease in oleic acid (C18:1c9), rumenic acid (C18:2c9t11) and α -linolenic acid (C18:3c9c12c15). This result concurred with the findings of Dierking *et al.* (2010) and Rugoho *et al.* (2014). Dierking *et al.* (2010) found from their study that grasses had higher amounts of α -linolenic acid compared with lucerne. Similarly, Rugoho

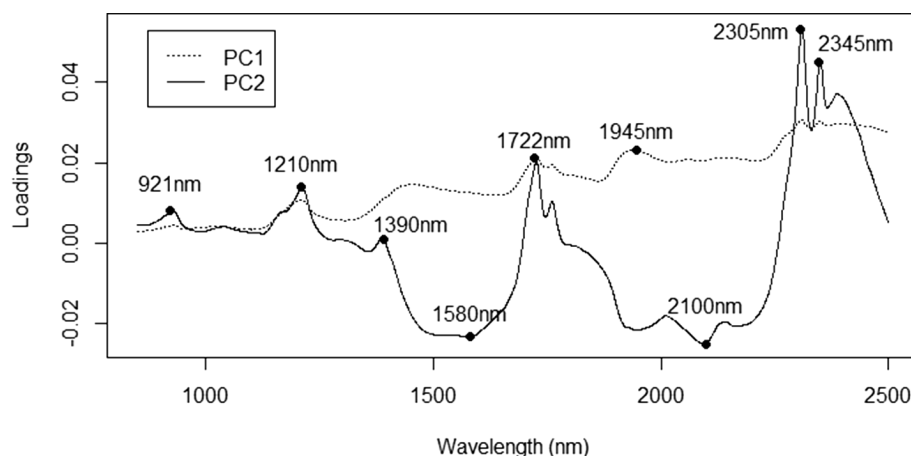


Figure 3. Loadings variable plot of the first two principal components, PC 1 and PC 2, showing the optimal wavelengths

et al. (2014) reported that the milk produced from cows consuming high-quality pasture contained high levels of α -linolenic acid and rumenic acid. Overall, the coefficients of variation of the FA concentrations were consistent with that of Fleming *et al.* (2017) and the values found for herds 3 and 4 were higher than those for herds 1 and 2.

The crude protein contents for the four herds of the freeze-dried raw milk samples are shown in Table 2. It can be seen from the table that the average values were higher for herds 1 and 2 than for herds 3 and 4. The recorded differences in mean values of the crude protein contents can be attributed to the effects of the inclusion of lucerne (a legume high in nitrogen) in the feeding regimes of herds 1 and 2. Woodward *et al.* (2010) showed in their study that a change from ryegrass to lucerne reduced milkfat and increased milk protein concentrations. The PLSR models for crude protein determination (Table 2) using the iPLS variable selection showed excellent results ($RSR \leq 0.41$).

For FA determination (Table 1), the RSR values ranged from 0.40 to 1.21. In general, only 8 objects had RSR values ≤ 0.50 that was desired for good predictions. To provide more details, we observed that the PLSR models developed with the full spectral range (850–2,500 nm) yielded four good predictions, whereas those built with iPLS variable selection gave four optimal predictions. Thus, the selection of sensitive wavelength intervals for building the calibration models did not improve the overall prediction outcome. Among the individual FA, 16:0 had the lowest RSR across all the herds, with herd 4 having the best prediction ($RSR = 0.44$). The minimum RSR recorded for the five C18 fatty acids predicted was 0.61 which was above the cut-off of 0.50 desired for good calibrations. Thus, the prediction accuracy of the PLSR models for the selected FA was inadequate. Nevertheless, there was a slight improvement in the global prediction performance across the herds for the FA groups. The RSR for UFA

ranged from 0.40 to 0.98, whereas that for SFA ranged from 0.41 to 0.79 across the herds. Several researchers obtained similar loss of prediction accuracy for FA and they attributed it to the effects of the differences in breeds of individual cows on the sample matrices. Tsenkova *et al.* (2000) obtained poorer results for near infrared determination of fat contents when the samples (in liquid form) from individual cows of multi-breeds were combined and predicted collectively than when they were simulated individually. These authors attributed the discrepancies in accuracy between the predictions to the differences in chemical composition and physical structure of the milk from each cow.

Multiclass chemometric differentiation of raw milk by herds

Tables 3 and 4 show the respective model performance metrics for PLS-DA and SIMCA for multiple classification of the NIRS spectra of the freeze-dried raw milk samples into herds. Three different treatments of the calibration sets are presented. In the first treatment, no preprocessing was applied, and the full spectra range (850–2,500 nm, 3,300 wavelengths) was used for the calibrations. In the second and third treatments, the same preprocessing (SG + EMSC) was applied but using the full spectra range (850–2,500 nm) and the selected sensitive wavelength range (1,580–2,305 nm, 1,451 wavelengths) for analysing the major constituents of milk powder, respectively.

As seen in the tables, the main diagonal elements represent fractions of true positives, whereas the off-diagonal elements indicate fractions of false positives for each of the herds. For the first treatment, the mean values recorded across the herds for SENS, SPEC, CCR, P and MCC for the PLS-DA model were 71.6%, 93.3%, 88.2%, 0.78 and 0.67, respectively, whereas those

Table 1. Fitting statistics of the NIRS prediction models for the dominant fatty acids of herd-classified raw milk using development data sets expressed on per milk basis as g/100 g of milk.

Prediction performance metrics ²						
Fatty acids	Descriptive stats ¹		SG + EMSC Full		SG + EMSC IPLS	
	Mean	CV	R ² _p	RSR	R ² _p	RSR
C16:0						
Herd 1	12.47 ^a	16.17	0.64	0.58	0.65	0.57
Herd 2	13.32 ^a	17.64	0.68	0.73	0.70	0.68
Herd 3	12.96 ^a	22.90	0.67	0.58	0.71	0.55
Herd 4	12.71 ^a	25.13	0.80	0.45	0.86	0.44
C18:0						
Herd 1	3.20 ^{bc}	20.44	0.03	1.00	0.11	0.91
Herd 2	2.93 ^c	19.68	0.16	1.03	0.18	1.12
Herd 3	3.52 ^{ab}	23.35	0.40	0.88	0.52	0.72
Herd 4	3.69 ^a	28.99	0.55	0.89	0.62	0.80
C18:1c9						
Herd 1	4.72 ^b	15.46	0.32	0.85	0.45	0.72
Herd 2	4.44 ^b	14.96	0.28	1.21	0.32	1.19
Herd 3	6.99 ^a	15.83	0.72	0.61	0.72	0.67
Herd 4	7.03 ^a	23.45	0.32	0.81	0.63	0.62
C18:2c9t11						
Herd 1	0.43 ^b	29.12	0.05	0.96	0.19	0.88
Herd 2	0.29 ^c	40.02	0.35	1.00	0.53	0.76
Herd 3	0.59 ^a	41.78	0.16	0.92	0.39	0.76
Herd 4	0.59 ^a	42.60	0.10	0.94	0.28	0.82
C18:3c9c12c15						
Herd 1	0.27 ^b	17.35	0.08	1.00	0.13	0.99
Herd 2	0.28 ^b	16.72	0.34	0.81	0.36	0.80
Herd 3	0.32 ^a	28.23	0.28	0.92	0.47	0.82
Herd 4	0.33 ^a	27.08	0.22	0.86	0.37	0.82
SFA						
Herd 1	24.93 ^a	13.58	0.51	0.76	0.54	0.71
Herd 2	25.37 ^a	15.17	0.73	0.77	0.74	0.79
Herd 3	25.95 ^a	19.94	0.88	0.41	0.89	0.41
Herd 4	25.94 ^a	22.95	0.88	0.50	0.91	0.44
UFA						
Herd 1	9.13 ^b	11.92	0.22	1.00	0.50	0.70
Herd 2	8.82 ^b	50.07	0.03	0.99	0.07	0.98
Herd 3	12.44 ^a	10.99	0.84	0.50	0.87	0.40
Herd 4	12.52 ^a	17.75	0.13	1.04	0.54	0.67

¹Mean and coefficient of variation estimated for Herd 1 ($n = 56$), Herd 2 ($n = 64$), Herd 3 ($n = 49$) and Herd 4 ($n = 51$) with a total of 220 samples.

²SG + EMSC = Savitzky–Golay plus extended multiplicative signal correction preprocessing methods performed on full spectral range, 850–2,500 nm, of the calibration sets ($n = 166$) and using interval-PLS. R²_p = coefficient of determination for prediction ($n = 54$). RSR = RMSE-observation standard deviation ratio calculated as 1/RPD (ratio of prediction to deviation). Optimal RSR ≤ 0.5.

^{a–c}Same superscript letters within a row are not significantly different at $P \leq 0.05$.

SG = Savitzky–Golay; EMSC = Extended Multiplicative Signal Correction; PLS = partial least square; RSR = RMSE-observation standard deviation ratio; SFA = saturated fatty acids; UFA = unsaturated fatty acids; CV = cross-validation.

computed for the SIMCA model were 79.7%, 78.8%, 79.1%, 0.57 and 0.54, respectively. For the second treatment, the mean values recorded across the herds for SENS, SPEC, CCR, P and MCC for the PLS-DA model were 73.3%, 95.2%, 90.0%, 0.83 and 0.72, respectively,

whereas those computed for the SIMCA model were 89.5%, 69.2%, 74.1%, 0.49 and 0.51, respectively. Finally, for the third treatment, the mean values recorded across the herds for SENS, SPEC, CCR, P and MCC for the PLS-DA model were 69.1%, 92.8%, 87.3%, 0.75 and 0.64,

Table 2. Fitting statistics of the NIRS prediction for crude protein contents (%) of herd-classified raw milk.

Prediction performance metrics ²						
Crude protein	Descriptive statistics ¹		SG + EMSC Full		SG + EMSC iPLS	
	Mean	Range	R ² p	RSR	R ² p	RSR
Herd 1	26.16 ^a ± 1.15	24.09–28.01	0.85	0.53	0.90	0.41
Herd 2	25.59 ^a ± 2.10	23.15–30.15	0.94	0.25	0.98	0.21
Herd 3	23.55 ^a ± 3.33	20.05–30.56	0.92	0.25	0.93	0.23
Herd 4	23.70 ^a ± 3.21	19.11–31.73	1.00	0.11	1.00	0.08

^aSame superscript letters within a column are not significantly different at $P \leq 0.05$.

¹Mean ± standard deviation estimated for Herd 1 ($n = 56$), Herd 2 ($n = 64$), Herd 3 ($n = 49$) and Herd 4 ($n = 51$) using Dumas Method.

²SG + EMSC = Savitzky–Golay plus extended multiplicative signal correction preprocessing methods performed on full spectral range, 850–2,500 nm, of the calibration sets ($n = 166$) and using interval-PLS. R²p = coefficient of determination for prediction ($n = 54$). RSR = RMSE-observation standard deviation ratio calculated as 1/RPD (ratio of prediction to deviation). Optimal RSR ≤ 0.5 .

SG = Savitzky–Golay; EMSC = Extended Multiplicative Signal Correction; PLS = partial least square; RSR = RMSE-observation standard deviation ratio.

Table 3. Multi-class PLS-DA classification of freeze-dried raw milk for four herds using different spectra preprocessing techniques.

Model with no pretreatment ¹										
True Class	Test sets assigned into classes				Model performance metrics					
	Herd 1	Herd 2	Herd 3	Herd 4	LV	SENS	SPEC	CCR	P	MCC
Herd 1	10/14	15/16	12/12	12/13	11	71.4	95.1	89.1	0.83	0.70
Herd 2	11/14	14/16	12/12	13/13	11	87.5	92.3	90.9	0.82	0.78
Herd 3	14/14	16/16	7/12	10/13	15	58.3	93.0	85.5	0.70	0.55
Herd 4	14/14	16/16	9/12	9/13	11	69.2	92.9	87.3	0.75	0.64
Model with pretreatment ²										
Herd 1	12/14	14/16	12/12	13/13	12	85.7	95.1	92.7	0.86	0.81
Herd 2	12/14	14/16	12/12	13/13	8	87.5	94.9	92.7	0.88	0.82
Herd 3	14/14	16/16	7/12	10/13	12	58.3	93.0	85.5	0.70	0.55
Herd 4	14/14	16/16	11/12	8/13	13	61.5	97.6	89.1	0.89	0.68
Model with variable selection ³										
Herd 1	11/14	14/16	12/12	13/13	9	78.6	95.1	90.9	0.85	0.76
Herd 2	12/14	15/16	11/12	13/13	5	93.8	92.3	92.7	0.83	0.83
Herd 3	14/14	16/16	6/12	8/13	14	50.0	88.4	80.0	0.55	0.40
Herd 4	14/14	16/16	10/12	7/13	5	53.8	95.2	85.5	0.78	0.56

¹Full spectrum (850–2,500 nm) models calibrated ($n = 165$) with no preprocessing; test sets $n = 55$.

²Full spectrum models preprocessed with Savitzky–Golay algorithm in addition to the extended multiplicative signal correction method.

³Variable selection models (1,580–2,305 nm) preprocessed with Savitzky–Golay algorithm in addition to the extended multiplicative signal correction method.

LV = latent variables; SENS = sensitivity (%); SPEC = specificity (%); CCR = correct classification rate; P = precision; and MCC = Matthews correlation coefficient for evaluating the model efficiency.

respectively, whereas those computed for the SIMCA model were 92.7%, 73.9%, 78.7%, 0.54 and 0.59, respectively. Thus, the PLS-DA model had higher specificity and Matthews Correlation Coefficient, correct classification rate and precision (i.e. implying smaller standard deviation and coefficient of variation). The SIMCA model only had higher mean sensitivity than the PLS-DA model and as such performed better in assigning objects to their native classes/herds. However, the SIMCA model, with lower specificity, presented higher false positives while

differentiating the objects from the herds that received the same diet in the same milking year. The implication of higher specificity for the PLS-DA model is that it provided better differentiation between the herds on the same feeding regime which were shown to be overlapped in Figures 1 and 2 (i.e. herds 1 and 2, and herds 3 and 4) with fewer false positives/misclassifications. This result has further echoed the findings of Dooley *et al.* (2005) that raw milk varies between herds, making it possible to segregate it on-farm for the manufacture of specific dairy products.

Table 4. Multi-class SIMCA classification of freeze-dried raw milk for four herds using different spectra preprocessing techniques.

True Class	Model with no pretreatment ¹									
	Test sets assigned into classes				Model performance metrics					
	Herd 1	Herd 2	Herd 3	Herd 4	PC	SENS	SPEC	CCR	P	MCC
Herd 1	12/14	13/16	12/12	13/13	10	85.7	92.7	90.9	0.80	0.77
Herd 2	3/14	13/16	12/12	13/13	13	81.2	71.8	74.5	0.54	0.49
Herd 3	13/14	14/16	9/12	5/13	8	75.0	74.4	74.5	0.45	0.42
Herd 4	14/14	16/16	2/12	10/13	13	76.9	76.2	76.4	0.50	0.47
Model with pretreatment ²										
Herd 1	13/14	6/16	11/12	13/13	5	92.9	73.2	78.2	0.54	0.58
Herd 2	4/14	13/16	12/12	12/13	8	81.2	71.8	74.5	0.54	0.49
Herd 3	14/14	15/16	11/12	1/13	5	91.7	69.8	74.5	0.46	0.51
Herd 4	10/14	15/16	1/12	12/13	5	92.3	61.9	69.1	0.43	0.46
Model with variable selection ³										
Herd 1	13/14	8/16	12/12	12/13	4	92.9	78.0	81.8	0.59	0.63
Herd 2	3/14	15/16	12/12	13/13	11	93.8	71.8	78.2	0.58	0.60
Herd 3	14/14	16/16	11/12	2/13	4	91.7	74.4	78.2	0.50	0.56
Herd 4	14/14	16/16	0/12	12/13	8	92.3	71.4	76.4	0.50	0.55

¹Full spectrum (850–2,500 nm) models calibrated ($n = 165$) with no preprocessing; test sets $n = 55$.

²Full spectrum models preprocessed with Savitzky–Golay algorithm in addition to the extended multiplicative signal correction method.

³Variable selection models (1,580–2,305 nm) preprocessed with Savitzky–Golay algorithm in addition to the extended multiplicative signal correction method.

LV = latent variables; SENS = sensitivity (%); SPEC = specificity (%); CCR = correct classification rate; P = precision; and MCC = Matthews correlation coefficient for evaluating the model efficiency.

However, freeze-drying the samples for several days for *in situ* measurements may pose a problem of using NIRS as a quick method for on-farm segregation of raw milk. Our methodology was undertaken under small-scale laboratory conditions involving freeze-drying using available resources and it is feasible that raw liquid milk may alternatively be used in a compatible NIRS liquid analyser to achieve similar on-farm segregation efficiency. Indeed, Andueza *et al.* (2013) showed that there were no significant differences when fresh and freeze-dried cheeses were used to compare the ability of two NIRS methods to distinguish between pasture and preserved-forage cheeses. It has also been established ((De la Roza-Delgado *et al.*, 2017) that handheld NIRS instruments can successfully be used *in situ* to estimate the changes in individual raw liquid milk composition. Nonetheless, while we believe the principles of NIRS-based raw milk classification are well founded in our study, further research is needed to better establish the practicalities. Furthermore, as the Matthews correlation is a contingency matrix method for calculating the Pearson product-moment correlation coefficient (Powers, 2011), it will invariably take the same range of values as follows as adopted from Chan (2003): at least 0.8 for very strong, 0.6 up to 0.8 for moderately strong, 0.3–0.5 for fair and less than 0.3 for poor linear relationship. Thus, the performance of the PLS-DA model can be described as reasonably good while that of SIMCA was only fair. Overall, the preprocessed PLS-DA

calibration using the full spectra range (treatment 2) provided the best model for this study. The other preprocessing techniques, for example, treatment 3, could not better the outcome of treatment 2 using PLS-DA since they yielded lower CCR, P and MCC values. In addition, the calibration performed using the raw spectra without preprocessing (i.e. treatment 3) did not improve the results.

4. Conclusions

The average near infrared spectra and the PCA established that there was a difference in the freeze-dried milk powder samples between herds on different feeding regimes, but it was not clear whether such differences exist between those on the same diet. This study has demonstrated that although SIMCA models achieved high sensitivity, they could not reliably differentiate the near infrared spectra of the milk powder samples from different herds on the same feeding regimes where the PCA model had failed to describe the classes because of their superimposed objects. That implies that SIMCA models excelled only where PCA showed clear separation between the objects. However, the PLS-DA models were reasonably effective in differentiating between the herds on the same diets and their overall performance was better than that of SIMCA, thus highlighting the possibility of segregating raw milk on-farm between herds for the manufacture of specific dairy products and

for enhancing product traceability. These results support value for NIRS for characterising milk and efficacy in validation of feeding regimes, which are useful for independent quality assurance. It has also been shown that the inclusion of lucerne in the diets significantly impacted on the quality of the milk powder. Furthermore, PLSR models were developed and they provided excellent predictions for crude protein contents but not for the selected fatty acid profiles. The poor accuracy of the fatty acid determinations resulted from the collective prediction of milk powder samples, with differing matrices, from individual cows of multiple breeds. However, a seldom used but meaningful performance metric called 'RMSE-observations standard deviation ratio' was successfully used in this study for the evaluation of the PLSR models. Thus, while further studies are needed to examine the practicalities of on-farm analysis of raw milk to achieve the results we describe here, our results indicate that useful, independent and cost-effective, rapid quality analytics of milk powder to discriminate at herd level can be achieved, which may be beneficial to manufacturers of high-value dairy products.

Acknowledgements

The authors thank Andrea Hogan, Jenny Zhao, Rosy Tung and Shuang Jiang in the Lincoln University analytical team in the John Burton Building for their technical and logistical assistance.

Conflicts of interest

The authors have declared no conflict of interest in this article.

Compliance with ethical standards

This study does not involve any human testing.

Funding

Financial support from the Ministry of Science and Innovation (DRCX 0802; Dairy Systems for Environmental Protection; NZ) is acknowledged for provision of milk samples.

References

- Andueza, D., Agabriel, C., Constant, I., Lucas, A. and Martin, B., 2013. Using visible or near infrared spectroscopy (NIRS) on cheese to authenticate cow feeding regimes. *Food Chemistry* 141: 209–214. <https://doi.org/10.1016/j.foodchem.2013.02.086>
- Ballabio, D. and Todeschini, R., 2009. Multivariate classification for qualitative analysis. *Infrared spectroscopy for food quality analysis and control* 83: e102. <https://doi.org/10.1016/B978-0-12-374136-3.00004-3>
- Barker, M. and Rayens, W., 2003. Partial least squares for discrimination. *Journal of Chemometrics: A Journal of the Chemometrics Society* 17: 166–173. <https://doi.org/10.1002/cem.785>
- Chan, Y., 2003. *Biostatistics 104: correlational analysis*. Singapore Medical Journal 44: 614–619.
- Collomb, M., Bütikofer, U., Sieber, R., Jeangros, B. and Bosset, J.-O., 2002. Composition of fatty acids in cow's milk fat produced in the lowlands, mountains and highlands of Switzerland using high-resolution gas chromatography. *International Dairy Journal* 12: 649–659. [https://doi.org/10.1016/S0958-6946\(02\)00061-4](https://doi.org/10.1016/S0958-6946(02)00061-4)
- Coppa, M., Martin, B., Agabriel, C., Chassaing, C., Sibra, C., Constant, I., Graulet, B. and Andueza, D., 2012. Authentication of cow feeding and geographic origin on milk using visible and near-infrared spectroscopy. *Journal of Dairy Science* 95: 5544–5551. <https://doi.org/10.3168/jds.2011-5272>
- De la Roza-Delgado, B., Garrido-Varo, A., Soldado, A., Arrojo, A.G., Valdés, M.C., Maroto, F. and Pérez-Marín, D., 2017. Matching portable NIRS instruments for in situ monitoring indicators of milk composition. *Food Control* 76: 74–81. <https://doi.org/10.1016/j.foodcont.2017.01.004>
- Dierking, R., Kallenbach, R. and Roberts, C., 2010. Fatty acid profiles of orchardgrass, tall fescue, perennial ryegrass, and alfalfa. *Crop Science* 50: 391–402. <https://doi.org/10.2135/cropsci2008.12.0741>
- Dooley, A., Parker, W., Blair, H. and Hurley, E., 2005. Implications of on-farm segregation for valuable milk characteristics. *Agricultural Systems* 85: 82–97. <https://doi.org/10.1016/j.jagsy.2004.07.012>
- Ejeahalaka, K.K. and On, S.L., 2019a. Chemometric studies of the effects of milk fat replacement with different proportions of vegetable oils in the formulation of fat-filled milk powders: implications for quality assurance. *Food Chemistry* 295: 198–205. <https://doi.org/10.1016/j.foodchem.2019.05.120>
- Ejeahalaka, K.K. and On, S.L., 2019b. Effective detection and quantification of chemical adulterants in model fat-filled milk powders using NIRS and hierarchical modelling strategies. *Food Chemistry*: 125785. <https://doi.org/10.1016/j.foodchem.2019.125785>
- Fleming, A., Schenkel, F., Chen, J., Malchiodi, F., Bonfatti, V., Ali, R., Mallard, B., Corredig, M. and Miglior, F., 2017. Prediction of milk fatty acid content with mid-infrared spectroscopy in Canadian dairy cattle using differently distributed model development sets. *Journal of Dairy Science* 100: 5073–5081. <https://doi.org/10.3168/jds.2016-12102>
- Frankhuizen, R., 2001. NIR analysis of dairy products. In: Ciurczak, E.W. and Burns D.A. (eds.), *Handbook of near-infrared analysis*. CRC Press, New York, NY, USA, pp. 499–535.
- Heinrichs, J., Jones, C. and Bailey, K., 1997. Milk components: understanding the causes and importance of milk fat and protein variation in your dairy herd. *Dairy Animal Science* 5: 1e–8e.
- Hurtaud, C., Dutreuil, M., Coppa, M., Agabriel, C. and Martin, B., 2014. Characterization of milk from feeding systems based on herbage or corn silage with or without flaxseed and authentication through fatty acid profile. *Dairy Science & Technology* 94: 103–123. <https://doi.org/10.1007/s13594-013-0147-0>
- Karoui, R., Mouazen, A.M., Dufour, É., Pillonel, L., Picque, D., Bosset, J.-O. and De Baerdemaeker, J., 2006. Mid-infrared spectrometry: a tool for the determination of chemical parameters

- in Emmental cheeses produced during winter. *Le Lait* 86: 83–97. <https://doi.org/10.1051/lait:2005040>
- Katz, G., Merin, U., Bezman, D., Lavie, S., Lemberskiy-Kuzin, L. and Leitner, G., 2016. Real-time evaluation of individual cow milk for higher cheese-milk quality with increased cheese yield. *Journal of Dairy Science* 99: 4178–4187. <https://doi.org/10.3168/jds.2015-10599>
- Lavine, B. and Workman, J.J., 2004. Chemometrics. *Analytical Chemistry* 76: 3365–3371. <https://doi.org/10.1021/ac040053p>
- Marchitelli, C., Contarini, G., De Matteis, G., Crisà, A., Pariset, L., Scatà, M.C., Catillo, G., Napolitano, F. and Moioli, B., 2013. Milk fatty acid variability: effect of some candidate genes involved in lipid synthesis. *Journal of Dairy Research* 80: 165–173. <https://doi.org/10.1017/S002202991300006X>
- Martens, H. and Stark, E., 1991. Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis* 9: 625–635. [https://doi.org/10.1016/0731-7085\(91\)80188-F](https://doi.org/10.1016/0731-7085(91)80188-F)
- Martin, B., Fedele, V., Ferlay, A., Grolier, P., Rock, E., Gruffat, D. and Chilliard, Y., 2004. Effects of grass-based diets on the content of micronutrients and fatty acids in bovine and caprine dairy products, Land use systems in grassland dominated regions. Proceedings of the 20th General Meeting of the European Grassland Federation, Luzern, Switzerland, 21–24 June 2004. vdf Hochschulverlag AG an der ETH Zurich, pp. 876–886.
- Meilgaard, M.C., Carr, B.T. and Civille, G.V., 1999. Sensory evaluation techniques, 3rd edition. CRC Press; Taylor and Francis, New York, USA.
- Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D. and Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE* 50: 885–900. <https://doi.org/10.13031/2013.23153>
- Mouazen, A., Dridi, S., Rouissi, H., De Baerdemaeker, J. and Ramon, H., 2009. Prediction of selected ewe's milk properties and differentiating between pasture and box feeding using visible and near infrared spectroscopy. *Biosystems Engineering* 104: 353–361. <https://doi.org/10.1016/j.biosystemseng.2009.08.001>
- Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J.P., Munck, L. and Engelsen, S.B., 2000. Interval Partial Least-Squares Regression (i PLS): a comparative chemometric study with an example from near-infrared spectroscopy. *Applied Spectroscopy* 54: 413–419. <https://doi.org/10.1366/0003702001949500>
- Núñez-Sánchez, N., Martínez-Marín, A., Polvillo, O., Fernández-Cabanás, V., Carrizosa, J., Urrutia, B. and Serradilla, J., 2016. Near infrared spectroscopy (NIRS) for the determination of the milk fat fatty acid profile of goats. *Food Chemistry* 190: 244–252. <https://doi.org/10.1016/j.foodchem.2015.05.083>
- Oliveri, P., 2017. Class-modelling in food analytical chemistry: development, sampling, optimisation and validation issues—a tutorial. *Analytica Chimica Acta* 982: 9–19. <https://doi.org/10.1016/j.aca.2017.05.013>
- Oliveri, P. and Downey, G., 2012. Multivariate class modeling for the verification of food-authenticity claims. *Trends in Analytical Chemistry* 35: 74–86. <https://doi.org/10.1016/j.trac.2012.02.005>
- Powers, D.M., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies* 2: 37–63.
- R Core Team, 2018. R: a language and environment for statistical computing. Available at: <http://www.R-project.org>.
- Rugoho, I., Liu, Y. and Dewhurst, R., 2014. Analysis of major fatty acids in milk produced from high-quality grazed pasture. *New Zealand Journal of Agricultural Research* 57: 165–179. <https://doi.org/10.1080/00288233.2014.899505>
- Savitzky, A. and Golay, M.J., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* 36: 1627–1639. <https://doi.org/10.1021/ac60214a047>
- Tsenkova, R., Atanassova, S., Itoh, K., Ozaki, Y. and Toyoda, K., 2000. Near infrared spectroscopy for biomonitoring: cow milk composition measurement in a spectral region from 1,100 to 2,400 nanometers. *Journal of Animal Science* 78: 515–522. <https://doi.org/10.2527/2000.783515x>
- Wishart, D.S., 2007. Current progress in computational metabolomics. Briefings in bioinformatics 8. Academic Press, Inc., New York, pp. 279–293. <https://doi.org/10.1093/bib/bbm030>
- Wold, H., 1966. Estimation of principal components and related models by iterative least squares. *Multivariate Analysis*. Academic Press, Inc., New York, pp. 391–420.
- Wold, S., Ruhe, A., Wold, H. and Dunn, I., WJ, 1984. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing* 5: 735–743. <https://doi.org/10.1137/0905052>
- Wold, S. and Sjostrom, M., 1977. SIMCA: a method for analyzing chemical data in terms of similarity and analogy. *Chemometrics: Theory and Application* 52: 243–282. <https://doi.org/10.1021/bk-1977-0052.ch012>
- Woodward, S., Waghorn, G., Attwood, G. and Li, D., 2010. Ryegrass to lucerne-effects of dietary change on intake, milk yield and rumen microflora bacteria of dairy cows, Proceedings of the New Zealand Society of Animal Production. New Zealand Society of Animal Production, pp. 57–61. 23 June – 25 June 2010, Palmerston North, New Zealand.
- Zimmermann, B. and Kohler, A., 2013. Optimizing Savitzky-Golay parameters for improving spectral resolution and quantification in infrared spectroscopy. *Applied Spectroscopy* 67: 892–902. <https://doi.org/10.1366/12-06723>