Food Chemistry 371 (2022) 131189

Contents lists available at ScienceDirect

Food Chemistry

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

Effectiveness of visible – Near infrared spectroscopy coupled with simulated annealing partial least squares analysis to predict immunoglobulins G, A, and M concentration in bovine colostrum

M. Franzoi, A. Costa *, A. Goi, M. Penasa, M. De Marchi

Department of Agronomy, Food, Natural resources, Animals and Environment, University of Padova, Viale dell'Università 16, 35020 Legnaro, PD, Italy

ARTICLE INFO

Keywords:
Colostrum quality
Infant formula
Infrared spectroscopy
Simulated annealing
Dairy cattle

ABSTRACT

Visible - near infrared spectroscopy coupled with variable selection using simulated annealing PLS regression was tested to predict immunoglobulin fractions (g/L) of bovine colostrum, namely IgG, IgA and IgM. Immunoglobulins were quantified in 678 samples using the gold standard radial immunodiffusion. Samples were divided in calibration (50%) and validation (50%) datasets. Maximum number of selected variables were limited to 200 and root mean squared error in cross validation ($RMSE_{CV}$) was used as loss function. Performance of the final model developed using the calibration dataset was assessed on the validation dataset. Overall, simulated annealing PLS improved validation $RMSE_{CV}$ compared to ordinary PLS regression by 3% to 17%. The present study demonstrated the effectiveness of the calibration model for accurate quantification of IgG, the most abundant immunoglobulin of bovine colostrum ($RMSE_{CV}$ = 13.28 g/L; R^2 = 0.83). These outcomes could be useful to assess colostrum quality intended for animal and human usage.

1. Introduction

Bovine colostrum is 'the early milking from dairy cows, taken up to 3 days post-partum' (McGrath et al., 2016) and the interest in its quality has increased in the scientific community and among stakeholders, as well as in the pharmaceutical industry in both veterinary and human medicine. More than 1000 original and review articles where the words "cow" + "colostrum" or "bovine" + "colostrum" appear in title or authors' keywords were published from 2010 to 2020 (Scopus, 2021).

Due to peculiar anatomical reasons, colostrum quality is important for new-born in cattle more than in other mammals. In fact, cows are characterized by a cotyledonary synepitheliochorial placenta that hampers the transfer of antibodies and other bioactive factors from dam to foetus. Since calves are deficient in immunoglobulins (Ig) at birth, passive transfer of antibodies, vitamins and growth factors is reached with colostrum administration based on the '3Q', namely quickness, quality and quantity (Costa et al., 2021b).

Among Ig, the major fractions in bovine colostrum are IgG, IgA, and IgM, representing the 80% of total protein content. The IgG is the most abundant and important to be monitored to prevent failure of passive transfer (McGrath et al., 2016). For this reason, quality of colostrum is determined based on the IgG content that should be ideally greater than 50 g/L when intended to calf feeding (Buczinski & Vandeweerd, 2016; McGrath et al., 2016). Bovine colostrum is also interesting for manufacturing. In fact, it is an emerging nutraceutical ingredient quite used in human and animal supplements, since rich in beneficial bioactive factors and antioxidants with significant positive effects on health and performance (Bagwe et al., 2015; Borad & Singh, 2018; Lee et al., 2019). Bovine colostrum is commercialized in different forms with declaration of proteins and/or IgG content (Playford et al., 2020); items containing bovine colostrum embrace supplements for athletes, infant formula, and artificial milk and supplements for new-borns of humans, livestock species, and companion animals.

Considering studies carried out in humans, it has been demonstrated that 60 g/d of colostrum supplementation is beneficial during resistance training (Duff et al., 2014) and that colostrum can counteract negative side effects of nonsteroidal anti-inflammatory drugs (Mir et al., 2009). A beneficial effect of bovine colostrum administration has been recently reported in people suffering from clinical diseases, such as ulcerative colitis, necrotizing enterocolitis, and traveller's diarrhoea, and in adults following weight loss programmes (Bagwe et al., 2015; Juhl et al., 2018; Arslan et al., 2021).

Therefore, assessing the quality of colostrum is important for farmers, veterinarians and other dairy stakeholders to make informed

* Corresponding author.
E-mail address: angela.costa@unipd.it (A. Costa).

https://doi.org/10.1016/j.foodchem.2021.131189
Received 15 June 2021; Received in revised form 13 September 2021; Accepted 16 September 2021
Available online 20 September 2021

choices, e.g., to set up the farm colostrum bank, to evaluate the potential inclusion of external high-quality colostrum, to standardize composition and Ig of colostrum at pharmaceutical or food industry level. For these reasons, colostrometers and refractometers are available for rapid and indirect evaluation of the Ig level in farms (Quigley et al., 2013). Nevertheless, the gold standard method for Ig determination in plasma, milk and colostrum is the radial immunodiffusion (RID), which provides very accurate and repeatable results (Gelsinger et al., 2015) but requires laboratory expertise, high investment (more than 5 US $ per sample) and time for incubation (24 h; Gelsinger et al., 2015; Costa et al., 2021b). For these reasons, RID kits are not suitable to day-to-day on-farm application, but can be useful as reference analysis to assess the accuracy of faster and cheaper tools that provide indirect quantifications (Abuelo et al., 2019; Bielmann et al., 2010; Rivero et al., 2012).

Infrared spectroscopy is widely adopted in livestock for phenotyping purposes and at food industry level for rapid determination of quality features in a plethora of matrices, including dairy products and milk. In this view, this indirect method could be potentially interesting for colostrum, as its compounds are generally determined at laboratory level using traditional time-consuming and expensive analyses, such as RID, HPLC, ELISA, and other chromatographic techniques. From the experience gained in milk, cheese, and meat (De Marchi et al., 2018), infrared spectroscopy is expected to be potentially useful for fast, cost-effective, and non-destructive analysis of colostrum intended for several uses.

According to the literature, the main limitation of infrared spectroscopy is the development of reliable and robust prediction models, particularly for compounds present in a low concentration. Such models need representative reference data to be collected in order to build accurate predictive equations; for this reason, the accuracy of prediction models relies on and cannot be greater than the accuracy of the reference method (De Marchi et al., 2018). Visible near-infrared spectroscopy (Vis-NIRS) exploits the region of the electromagnetic spectrum from 380 nm to 2500 nm and is based on the molecular overtone and combination of vibrations from which the Vis-NIRS spectrum originates. On this view, effective statistics algorithms are necessary to extract useful information from spectra of complex organic matrices; the most widely used algorithm is partial least squares (PLS) regression, eventually associated with other procedures like the variable selection strategies (Xiaobo et al., 2010). In addition, machine learning techniques, such as convolutional neural networks, have been successfully applied in recent years to develop prediction models from infrared spectra (Anderson, 2007; Denholm et al., 2020). The absorption in the Vis-NIRS region range is lower compared to mid-infrared spectroscopy and thus Vis-NIRS is more appropriate for the analysis of solid and semi-solid samples like colostrum, whose density ranges from 1059 to 1068 kg/m^3 (McGrath et al., 2016; Strekozov et al., 2008).

Navrátilová et al. (2006) successfully developed NIRS calibration models to determine total solids, fat, non-fatty solids, protein and lactose content of colostrum in 90 samples of 18 cows at different moments after calving. Moreover, Rivero et al. (2012) evaluated the ability of Vis-NIRS to predict bovine colostrum IgG content in 157 samples collected in 2 farms in Chile. The coefficient of determination in cross validation (R^2_{CV}) of the model proposed by Rivero et al. (2012) was very high (0.94) and predictions were highly correlated with IgG values measured through RID. Results from the literature show that Vis-NIRS models for colostrum are reliable and can be exploited on a large scale for rapid, cheap and non-destructive analysis (Buczinski & Vandeweerd, 2016; Elsohaby et al., 2016; Rivero et al., 2012), however, to the authors' knowledge no studies have evaluated Vis-NIRS for the determination of other colostrum Ig fractions, i.e., IgA and IgM.

Variable selection is an extensively used approach to improve infrared-based predictions. Basically, all variable selection algorithms point to exclude noisy spectral regions and collinear wavelengths, in order to improve model robustness and precision (Kalivas et al., 1989; Xiaobo et al., 2010). Simulated annealing (SA) for variable selection has demonstrated good performance in improving partial least squares (PLS) calibrations in several matrices (Balabin & Smirnov, 2011; Guo et al., 2020; Liu et al., 2019). The SA is a probabilistic optimization method able to accept with a certain probability non-optimizing solutions, in order to avoid getting stuck in a local minimum. Solutions improving the loss function, e.g., predicted residual sum of squares, are always accepted. Solutions worsening the loss function can be accepted with a certain probability following a Metropolis criterion. As the algorithm proceed, Metropolis criterion became more stringent, allowing convergence (Swierenga et al., 1998).

In the present study, a large number of individual colostrum samples were available to i) determine IgG, IgA, and IgM through the gold standard reference method RID and ii) collect Vis-NIRS spectra to develop prediction models using the SA-PLS.

2. Materials and methods

2.1. Protocol for samples collection

Colostrum of 678 cows was provided by 9 farms located in northern Italy from March 2019 to June 2020. Farmers did not perform vaccination for Rotavirus, Coronavirus, and E. Coli on selected cows before calving. According to the experimental protocol, only colostrum samples collected within 6 h from parturition were kept for the trial and stored in plastic sterile tubes (120 mL). Each farmer was in charge of colostrum sampling and was required to annotate the cow ID and the calving date on the tube and to freeze samples immediately after collection (−20 °C). Considering the high variability of Ig in the first hours after calving, samples collected after 6 h from calving were not considered for the present study. Periodically, the frozen samples were transferred to the University of Padova (Legnaro, Italy) for RID analysis. Official ID, date of birth, date of calving and parity of cows were retrieved from farmers' databases.

2.2. Reference analysis and spectra collection

Analyses were performed at the 'LaChi' laboratory of the University of Padua (Legnaro, Italy) as described by Costa et al. (2021b). In particular, RID kits for colostrum analyses were purchased from Triple J Farms (Bellingham, US) and handled according to the manufacturer's specific instructions. Considering the kit detection range, each colostrum sample was diluted 1:5 (v/v) for IgG and 1:3 (v/v) for IgA and IgM with ultrapure water (Arium 611UV Sartorius, Sartorius, MB, Italy). After incubation, the software ImageJ (Laboratory for Optical and Computational Instrumentation, University of Wisconsin-Madison, WI, US) was adopted to assess the diameter of each precipitated ring in duplicate (2 operators) and the average of the 2 measurements represented the final diameter considered for each sample in each Ig.

Costa et al. (2021b) derived the concentration of each Ig (g/L) through an equation developed using both reference sera (n = 3) diameter and known concentrations, i.e., 1.80, 14.72, and 28.03 g/L for IgG, 0.53, 1.94, and 3.87 g/L for IgA, and 0.62, 2.00, and 3.81 g/L for IgM. The supplier of the RID plates was Triple J Farms (Bellingham, US). A missing value was placed when a not readable or non-circular ring for one of the target components due to issues during colostrum dilution or pipetting was observed.

A preliminary test showed that using 1 well per sample was accurate enough to determine the RID target component (IgG, IgA, and IgM). Briefly, an intra-assay coefficient of variation (CV_{RID}, %) was calculated to assess RID repeatability using 4 colostrum samples tested in quintuplicate by a single operator. Separately for IgG, IgA and IgM plate, the CV_{RID} was calculated as the average of the individual CV of the 4 samples measured in quintuplicate, as:

$$CV_{RID} = \frac{\left[\left(s_1/\bar{x}_1\right) + \left(s_2/\bar{x}_2\right) + \left(s_3/\bar{x}_3\right) + \left(s_4/\bar{x}_4\right)\right]}{4} \cdot 100$$

where $\bar{x}_n$ and s_n are the mean and the standard deviation of the 5 concentrations available for the same sample. The intra-assay CV_{RID} was 7.56% for IgG, 2.46% for IgA and 3.03% for IgM. According to the guidelines of the US Department of Health and Human Services, Food and Drug Administration (2001), values < 10% are considered precise enough. Therefore, considering that each plate had 24 wells, 21 were intended to samples and 3 to reference sera.

Spectra were collected using the DS2500 (FOSS Electric A/S, Hillerød, Denmark) after homogenization of a representative amount of each colostrum sample (10 mL, 20 °C) by inversion. The device scanned from 400 nm to 2499.5 nm, with a spectral resolution of 0.5 nm, and each spectrum was an average of 32 sub-spectra recorded at 8 different points by rotating the sample cup automatically. All the spectra were recorded as absorbance, which is calculated as $\log_{10}(1/\text{reflectance})$.

2.3. Simulated annealing calibration

Samples whose spectra had Mahalanobis distance greater than 3.0 from the mean of the spectra themselves were removed from the dataset (Williams, 2007). Spectra were normalized before calibration using standard normal variate correction (Barnes et al., 1989). The initial dataset was randomly split in two subsets, each representing 50% of total samples. Calibration subset was used to train the model, while the validation subset was used to assess the final performance of the calibration model. In SA-PLS, variable selection is performed iterating several PLS changing wavenumbers included in the model and calculating root mean squared error in cross validation ($RMSE_{CV}$) as loss function. A schematic presentation of the algorithm is presented in Fig. 1. Initial variables were randomly selected and used to train the PLS model and $RMSE_{CV}$ was computed. The $RMSE_{CV}$ was calculated performing leave one out cross validation (LOOCV) on the calibration dataset at each PLS iteration. To avoid overfitting, the number of latent variables included in the PLS model were those allowing the optimization of $RMSE_{CV}$ (maximum number of latent variables is in Table 1). At this point, a candidate solution was generated randomly changing variables according to the disturbance parameter, using a random binomial variable, with probability of success equal to the parameter β. Considering such probabilistic approach, variation in the number of variables included in the model was allowed.

The new solution was evaluated for acceptance following a Metropolis criterion. Metropolis criterion probability was defined as $P(k \rightarrow j) = \exp((Fk - Fj) / t)$, where $Fk = RMSE_{CV}$ at step k, $Fj = RMSE_{CV}$ at step $j = i + 1$, and t is the control parameter. The new solution was accepted if $P(i \rightarrow j) > k$, where k is a random number between 0 and 1. Initial t was set to allow an initial transition acceptance ratio of about 0.75 (Kalivas et al., 1989). The procedure was iterated until the maximum number of new solutions or the maximum number of tested solutions were reached.

Subsequently, control parameter was modified according to the following formula: $t_j = t_k * \alpha$, and probability of variables changing was iteratively diminished at each temperature step with $\beta(k \rightarrow j) = \beta_k * \alpha + 0.01$, where β_i is the probability at step I and α is the optimization parameter. Then a new Markov chain was started. Two criteria determined the ending of the optimization algorithm: (i) the solution was not updated for two consecutive chains and (ii) the lowest t allowed was reached.

The initial parameters set for preparatory and calculation runs were the initial number of wavelengths included in the model, the maximum number of wavelengths included in the model at each step, the maximum number of latent variables, the optimization parameter α smaller than 1, the initial disturbance parameter β, the initial value for the control parameter t, the maximum number of tested solutions at each

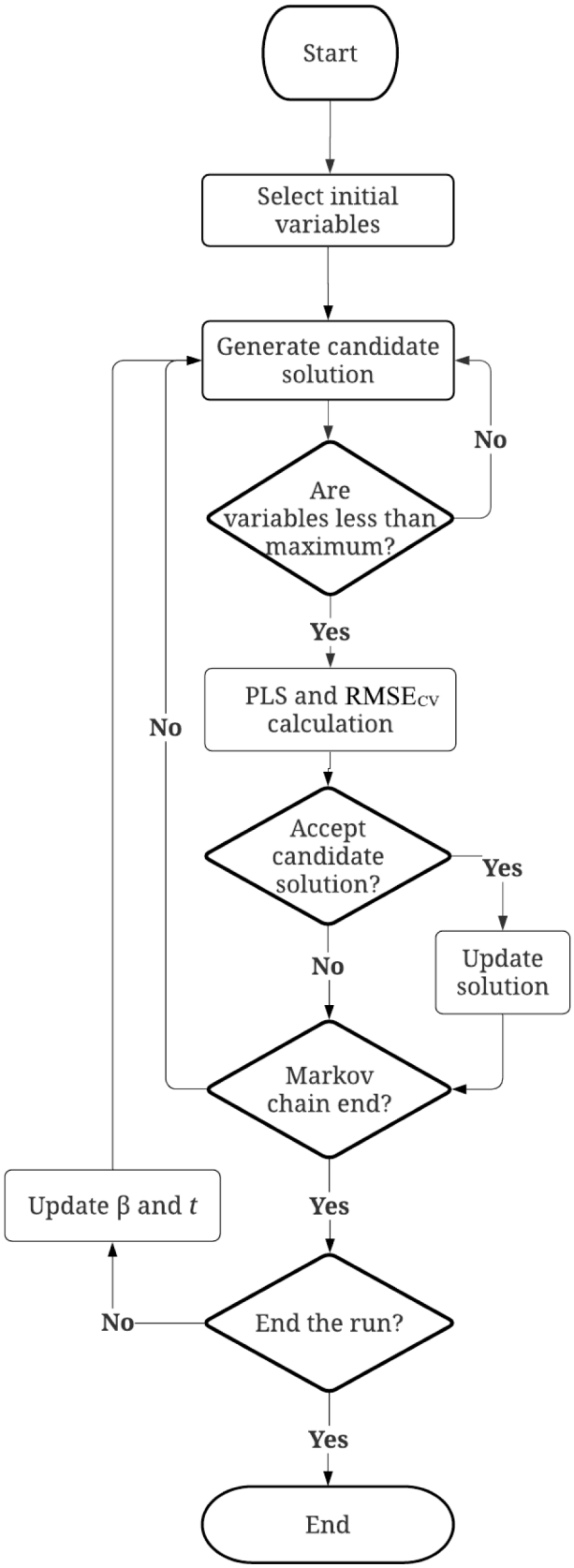

Fig. 1. Flowchart of simulated annealing PLS algorithm, where $RMSE_{CV}$ indicates the root mean squared error in cross validation.

Table 1
Simulated annealing parameters for preparatory and calculation runs.

Parameter	Preparatory runs	Calculation runs[1]		
		IgG	IgA	IgM
Number of initial wavelengths	200	100	100	100
Maximum number of wavelengths	4200	200	200	200
Maximum number of latent variables	20	20	20	20
Cross validation	Random, 20 groups	LOOCV	LOOCV	LOOCV
Optimization parameter (α)	0.950	0.975	0.975	0.975
Initial disturbance parameter (β)	0.5	0.2	0.2	0.2
Initial control parameter (*t*)	1.0	0.5	0.05	0.05
Markov chain length	250	1000	1000	1000
Maximum number accepted transitions	100	250	250	250
Lowest *t* allowed	0.0005	0.0005	0.0005	0.0005
Number of independent runs	25	5	5	5

[1] IgG = immunoglobulins G (g/L); IgA = immunoglobulins A (g/L); IgM = immunoglobulins M (g/L); LOOCV = leave-one-out cross validation.

t step, the maximum number of accepted solutions at each *t* step and the lowest value of the *t* parameter being tested. The number of independent runs for each initial set of parameters are reported in Table 1.

Performance of the final model was assessed on the validation dataset calculating coefficient of determination (R^2_V), root mean squared error in validation ($RMSE_V$) and ratio of prediction to deviation (RPD).

Simulated annealing PLS was performed using a self-built macro in SAS® 9.4 (SAS Institute Inc., Cary, NC). In the same software, the PLS, SURVEYSELECT, and other procedures were used within the macro. Moreover, a receiving operator characteristic (ROC) was carried out for the predicted IgG using LOGISTIC procedure to assess diagnostic accuracy through the area under curve (AUC) evaluation. In particular, the accuracy of predicted IgG in discriminating low- (<50 g/L of RID IgG) and high-quality colostrum (≥50 g/L of RID IgG) was evaluated. Source codes are available from the authors upon reasonable request.

3. Results and discussion

3.1. Radial immunodiffusion analyses

Means and standard deviation of IgG, IgA and IgM contents calculated for the whole dataset were 91.53 ± 32.13, 4.63 ± 2.77 and 5.07 ± 2.40 g/L, respectively, and the coefficient of variation was greater than 35% for all the Ig fractions (Table 2). Overall, concentrations ranged from 6.52 to 182.62, 0.13 to 17.51, and 0.18 to 14.01 g/L for IgG, IgA, and IgM, respectively. Results are consistent with McGrath et al. (2016) who reviewed studies on colostrum Ig and reported values of total Ig from 30 to 200 g/L. The average IgG was similar to that reported for Simmental cows by Ceniti et al. (2019), whose colostrum samples were collected within 6 h from calving, similarly to the protocol adopted in the present study. Using samples of 111 multiparous Holstein cows collected from 1 to 14.5 h after calving Cabral et al. (2016) reported IgG content in the range from 21.4 to 141.4 g/L. In colostrum of Canadian beef cattle, IgG determined through RID averaged 149.60 ± 38.70 g/L because more than three quarters of the samples were collected within 1 h from parturition (Gamsjäger et al., 2020). Sampling time is a key factor for colostrum Ig, since their concentrations quickly decline in the first hours after calving (Godden, 2008). In bovine, there is a physiological reason for the decrease in Ig concentration in the first hours post-partum. In fact, colostrum content of antibodies is maximum at birth, i.e., when calf's gut permeability is 100%, whereas it declines afterwards together with the calf's gut absorption ability (Godden, 2008; Hurley & Theil, 2011; Costa et al., 2021b). In addition, it is worth mentioning that the method to determine Ig content has to be considered when comparing studies dealing with colostrum Ig. In fact, there can be differences between RID and ELISA tests commercially available to determine bovine Ig (Gapper et al., 2007; Solórzano, 2020).

Table 2
Descriptive statistics of immunoglobulins[1] in calibration and validation sets.

Trait	n	Mean	SD	Minimum	Maximum	CV, %
Calibration set						
IgG	258	92.29	31.71	18.11	182.08	34.4
IgA	227	4.56	2.82	0.33	17.51	61.9
IgM	254	5.07	2.36	0.18	11.28	46.5
Validation set						
IgG	272	90.80	32.57	6.52	182.62	35.9
IgA	231	4.70	2.72	0.13	15.58	57.9
IgM	267	5.07	2.44	0.29	14.01	48.2

[1] IgG = immunoglobulins G (g/L); IgA = immunoglobulins A (g/L); IgM = immunoglobulins M (g/L). SD = standard deviation; CV = coefficient of variation.

Descriptive statistics of colostrum Ig in calibration and validation sets are presented in Table 2. According to unpaired *t*-test ($P < 0.05$), means calculated in the two sets within Ig were not significantly different, thus variability of each Ig in calibration and validation sets was comparable.

3.2. PLS prediction models

Results of standard PLS algorithm in predicting the concentration of different Ig in colostrum samples are summarized in Table 3. The IgG had the best prediction performances, which was expected considering they represent about 91% of total Ig content of colostrum. Calculated RPD was 2.04, i.e., slightly higher than the value of 2.00 which is considered as the minimal threshold for a prediction to be used at least for rough screening (Williams, 2014). According to intra-assay coefficient of variation for the reference method, it is reasonable the maximum achievable correlation between reference and predicted Ig to be 0.92, 0.98, and 0.97, for IgG, IgA, and IgM, respectively, considering Vis-NIRS cannot be more precise than reference data used to build the calibration model (Costa et al., 2021b; Zeaiter et al., 2004). On this view, R^2_V obtained using PLS algorithm suggest there is the possibility to improve calibration performances for all the analysed Ig. Obtained results were in agreement with previous findings of Costa et al. (2021a) who reported R^2_{CV} of 0.84 on a subset of samples used in the current study. Considering external validation, the same authors reported a lower R^2_V (0.63), lower than performance reported in the present paper (0.76). The observed difference likely arose from the different separation of samples between calibration and validation set. Results similar to the ones obtained in the present study were achieved by Elsohaby et al. (2017), Elsohaby et al. (2018), using PLS. In that case, the Pearson correlation between RID- and IR-measured IgG was 0.88, roughly corresponding to R^2 of 0.77. Considering the other analysed Ig, PLS calibrations for IgA and IgM obtained in the present study were unsatisfactory, with RPD of about 1.20 and R^2_V of 0.30 ad 0.33, respectively.

3.3. SA-PLS preparatory runs

Preparatory runs were set up to evaluate best parameters for the computational-intensive calculation runs. In preparatory runs, the maximum number of selected wavelengths was not constrained and, in order to reduce computational load, a 20-fold cross validation was performed instead of full LOOCV. Moreover, shorter Markov chains were imposed compared to calculation runs. Variable selection procedure used $RMSE_{CV}$ as loss function, as proposed by Liu et al. (2019). To check if $RMSE_{CV}$ was a reliable quality index for the final model, $RMSE_{CV}$ and $RMSE_V$ were compared for the 25 independent preparatory

Table 3
Prediction and cross-validation results for the determination of immunoglobulins[1] in colostrum using PLS and SA-PLS algorithms[2].

Algorithm	Trait	N	LV	Cross validation		External validation				
				$RMSE_{CV}$	R^2_{CV}	$RMSE_V$	R^2_V	Bias	Slope	RPD
PLS	IgG	4200	18	14.49	0.79	15.93	0.76	16.98	0.80	2.04
	IgA	4200	8	1.88	0.35	2.16	0.30	3.00	0.33	1.20
	IgM	4200	7	1.97	0.37	1.86	0.33	2.93	0.40	1.21
SA-PLS	IgG	50	19	8.77	0.92	13.28	0.83	12.42	0.86	2.45
	IgA	136	15	1.33	0.67	2.02	0.43	2.18	0.57	1.28
	IgM	121	15	1.68	0.54	1.81	0.37	2.70	0.45	1.24

[1] IgG = immunoglobulins G (g/L); IgA = immunoglobulins A (g/L); IgM = immunoglobulins M (g/L). [2]PLS = partial least squares; SA-PLS = simulated annealing partial least squares. LV = latent variables; $RMSE_{CV}$ = root mean squared error in cross validation; R^2_{CV} = coefficient of determination in cross validation; $RMSE_V$ = root mean squared error in validation; R^2_V = coefficient of determination in validation; RPD = ratio of prediction to deviation.

runs (Fig. 2). Pearson correlation between the $RMSE_{CV}$ of the selected model and the $RMSE_V$ of the external dataset was 0.95, confirming that the loss function was a good estimator for model selection. An alternative loss function was proposed by Jiang et al. (2012) combining the correlation coefficient and root mean squared error of prediction. Considering the good relationship between $RMSE_{CV}$ and $RMSE_V$, the aforementioned loss function was not investigated further.

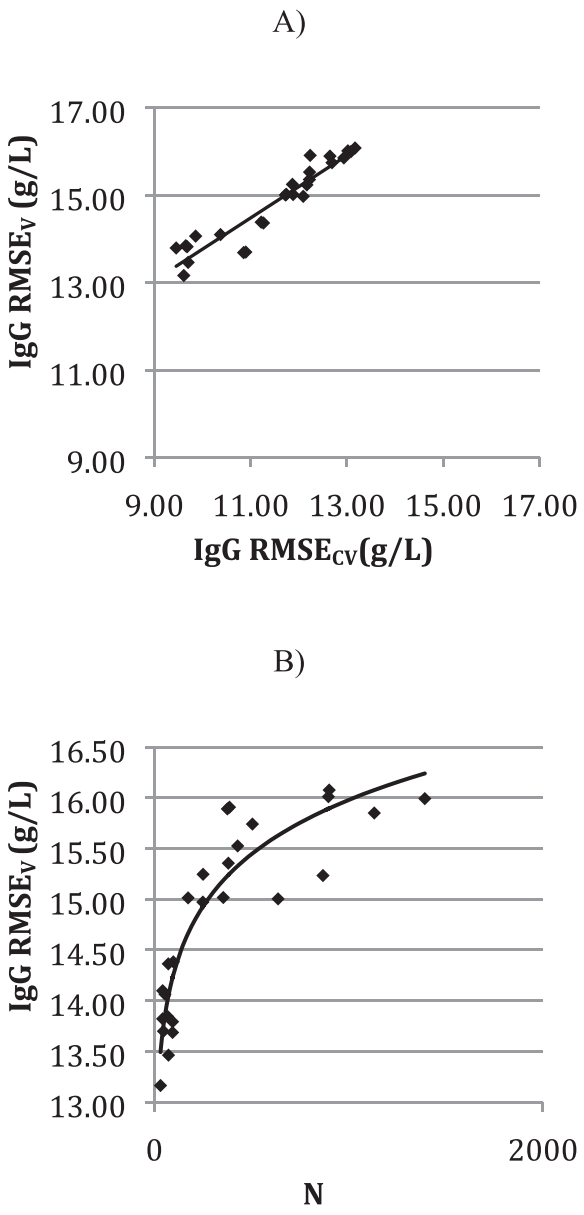

Fig. 2. Plot of A) root mean square error in cross validation ($RMSE_{CV}$) and root mean square error in validation ($RMSE_V$) in preparatory runs for immunoglobulins G (IgG, g/L) and B) number of selected variables (N) and root mean squared error in validation ($RMSE_V$) in preparatory runs.

The relationship between the number of variables in the final prediction models and the $RMSE_{CV}$ is reported in Fig. 2. It is quite clear the inverse relationship between the number of variables and the loss function. For this reason, in calculation runs the maximum number of spectral variables included in the model was set to 200, allowing lower computational load and faster calculations. Previously, Swierenga et al. (1998) tested SA-PLS performance fixing the number of selected variables at 50, 100, 150 and 200, between 1100 and 2500 nm. As a result, best average performance was achieved for 50 or 100 wavelengths. Accordingly, and considering that the number of spectral variables in the proposed algorithm was not fixed, a constrain at maximum 200 variables was believed reasonable.

3.4. Calculation runs

For calculation runs, a LOOCV was performed at each step to determine the loss function and to select the optimal number of latent variables. Starting *t* parameter allowed an initial acceptance ratio of about 0.5. For each trait, Table 3 reports the results of the best model in terms of $RMSE_{CV}$ among 5 independent calculation runs. For all the Ig, SA-PLS outperformed PLS, with a reduction of $RMSE_{CV}$ of 39%, 29% and 15% for IgG, IgA and IgM, respectively. Considering $RMSE_{CV}$ was the loss function for the selection algorithm, the result was expected. Nevertheless, even if $RMSE_{CV}$ was a good statistics for final model selection, generation of thousands of candidate solutions could lead to an overfitting of the loss function. Accordingly, robust external validation is desirable to estimate the real performance of the algorithm. Comparing $RMSE_V$ of PLS and PLS-DA, error decreased by 17%, 6% and 3% for IgG, IgA and IgM, respectively (Table 3). Even if the improvement is lower than for $RMSE_{CV}$, it is still particularly significant for IgG. Notably, the best improvement was for the IgG because good performances are obtained for this fraction using standard PLS. In a previous study, Balabin & Smirnov (2011) obtained an improvement of $RMSE_V$ between 16% and 24% on different components of biodiesel, using SA-PLS. Correlation plots for RID and predicted Ig in the validation dataset are available in the Supplementary Material (Figs. S1, S2, and S3).

Considering that for calf feeding conventionally good quality colostrum is that one presenting IgG $\geq$ 50 g/L (Godden, 2008), the capability of NIRS model to distinguish between good and poor-quality samples was tested using a ROC analysis. The ROC curve was obtained from the relationship between RID and predicted IgG in the validation set (Fig. 3) and was characterized by an AUC of 0.97. Such value can be classified as outstanding according to ROC interpretation standards (Hosmer et al., 2013).

Selected wavelengths for SA-PLS models and variable importance in projection (VIP) are displayed in Fig. 4. Notably, even if SA-PLS selected for different wavelenghts among Ig, VIP followed a similar pattern. For

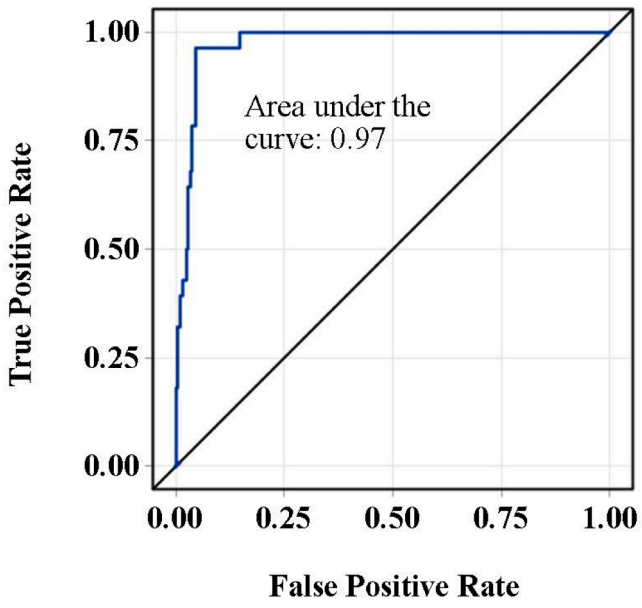

Fig. 3. Receiving operating characteristic for good and poor-quality colostrum predicted using SA-PLS infrared prediction model, calculated on the validation set. Positive samples were defined as those with RID-measured IgG < 50 g/L.

IgA and IgM, the range of visible light of the spectrum (400–600 nm) was particularly important. The Ig content of colostrum is known to be linked to total solids scattering, which determines the whiteness of samples and, consequently, the absorbtion pattern in the visible range (Cattaneo et al., 2009). Other important variables were detected between 950 and 1050 nm and between 1080 and 1120 nm. The first is usually linked to fourth overtone of C═O bond, typical for protein, and to the water content of the sample, and the second is influenced by aromatic compounds such as proteins aromatic residues (Holroyd, 2013; Workman, 2016). Another important region of spectum for IgM prediction was between 1400 and 1550 nm, which is referred to the second combination region of CH and amide II, and linked to water and protein content (López-Lorente & Mizaikoff, 2016). According to VIP, particularly important was the spectral section of the first combination region of CH (2070–2150 nm and 2240–2360 nm), linked to protein and fat content in food samples (Workman, 2016).

4. Conclusions

The present study is the first attempt to predict the concentration of the 3 major bovine colostrum Ig fractions from spectral data. Visible-near infrared spectroscopy decreases RMSEv when SA-PLS is used for calibration by 17%, 6% and 3%, for IgG, IgA and IgM, respectively, compared to the standard PLS. Nevertheless, only IgG reached satisfactory performances for routine usage. Despite this, it is worth highlighting that concentration of IgG is the key-parameter used to define colostrum quality at both farm and industry level. Therefore, visible-near infrared spectroscopy is suitable for IgG quantification in colostrum intended for different uses, since it is a fast, reliable, easy to be implemented, and cost-effective tool. Further research will focus on the prediction of other traits of interest for both farmers and dairy industry, like growth factors, vitamins, and antioxidants, in both liquid and spray-dried bovine colostrum.

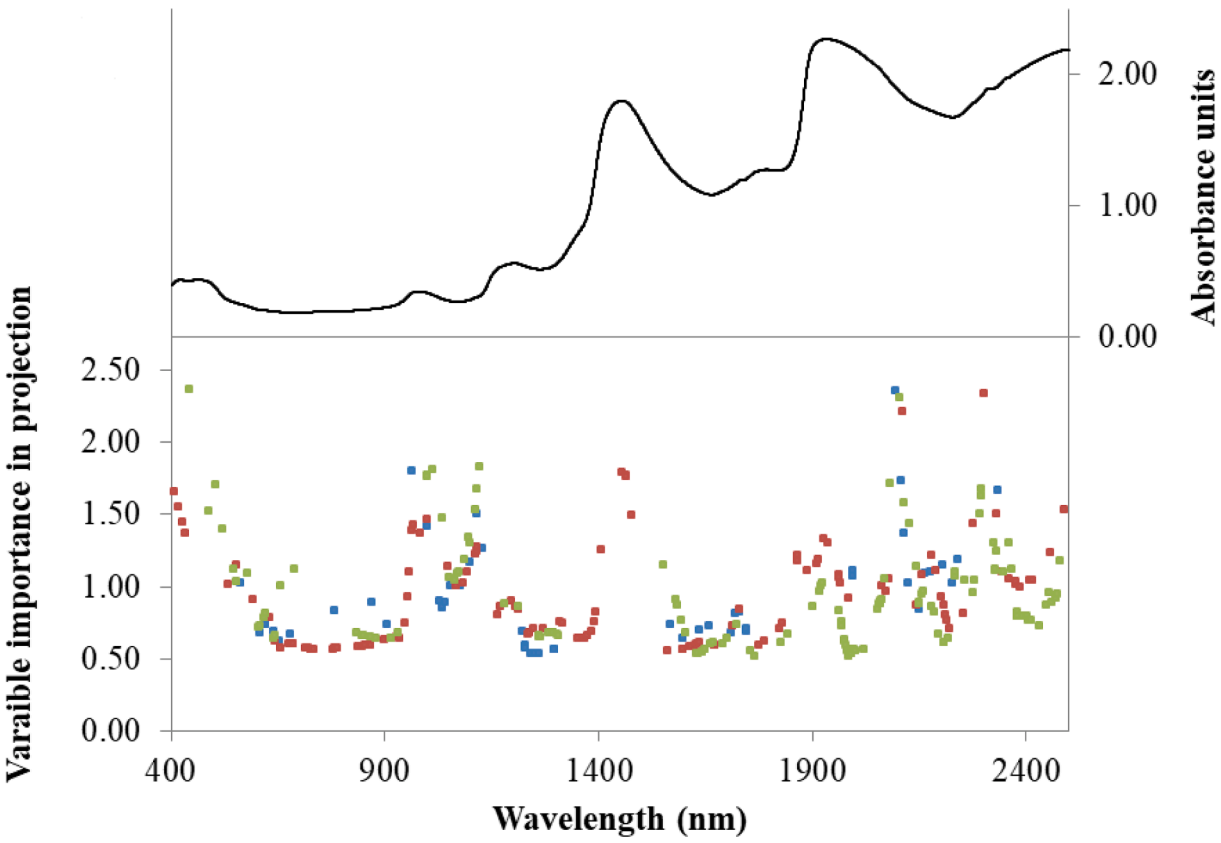

Fig. 4. Average near-infrared spectrum of analysed colostrum samples and variable importance in projection for selected variables: immunoglobulins G (g/L; blue), A (g/L; green), and M (g/L; red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

CRediT authorship contribution statement

M. Franzoi: Methodology, Software, Validation, Visualization, Formal analysis, Writing - original draft, Writing - review & editing. **A. Costa:** Data curation, Conceptualization, Methodology, Software, Validation, Visualization, Formal analysis, Writing - original draft, Writing - review & editing, Project administration. **A. Goi:** Methodology, Validation, Visualization, Formal analysis, Writing - original draft. **M. Penasa:** Validation, Visualization, Writing - original draft, Supervision. **M. De Marchi:** Validation, Visualization, Writing - original draft, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Dr Claudia De Lorenzi, Dr Giulia Nardino and Dr Laura Posenato (University of Padova) are gratefully acknowledged for their technical support. This research was supported by project “Innovamilk” (Innovations in Italian dairy industry for the enhancement of farm sustainability, milk technological traits and cheese quality) funded by AGER – Agroalimentare e Ricerca, Grant. N. 20171153.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2021.131189.

References

Abuelo, A., Havrlant, P., Wood, N., & Hernandez-Jover, M. (2019). An investigation of dairy calf management practices, colostrum quality, failure of transfer of passive immunity, and occurrence of enteropathogens among Australian dairy farms. *Journal of Dairy Science, 102*(9), 8352–8366. https://doi.org/10.3168/jds.2019-16578

Anderson, S. (2007). Determination of fat, moisture, and protein in meat and meat products by using the FOSS FoodScan Near-Infrared Spectrophotometer with FOSS Artificial Neural Network Calibration Model and Associated Database: Collaborative study. Journal of AOAC International, 90(4), 1073–1083.

Arslan, A., Kaplan, M., Duman, H., Bayraktar, A., Ertürk, M., Henrick, B. M., … Karav, S. (2021). Bovine Colostrum and Its Potential for Human Health and Nutrition. *Frontiers in Nutrition, 8*, 350. https://doi.org/10.3389/fnut.2021.651721

Bagwe, S., Tharappel, L. J. P., Kaur, G., & Buttar, H. S. (2015). Bovine colostrum: An emerging nutraceutical. *Journal of Complementary and Integrative Medicine, 12*(3), 175–185. https://doi.org/10.1515/jcim-2014-0039

Balabin, R. M., & Smirnov, S. V. (2011). Variable selection in near-infrared spectroscopy: Benchmarking of feature selection methods on biodiesel data. *Analytica Chimica Acta, 692*(1), 63–72. https://doi.org/10.1016/j.aca.2011.03.006

Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy, 43*(5), 772–777. https://doi.org/10.1366/0003702894202201

Bielmann, V., Gillan, J., Perkins, N. R., Skidmore, A. L., Godden, S., & Leslie, K. E. (2010). An evaluation of Brix refractometry instruments for measurement of colostrum quality in dairy cattle. *Journal of Dairy Science, 93*(8), 3713–3721. https://doi.org/10.3168/jds.2009-2943

Borad, S. G., & Singh, A. K. (2018). Colostrum immunoglobulins: Processing, preservation and application aspects. *International Dairy Journal, 85*, 201–210. https://doi.org/10.1016/j.idairyj.2018.05.016

Buczinski, S., & Vandeweerd, J.-M. (2016). Diagnostic accuracy of refractometry for assessing bovine colostrum quality: A systematic review and meta-analysis. *Journal of Dairy Science, 99*(9), 7381–7394. https://doi.org/10.3168/jds.2016-10955

Cabral, R. G., Chapman, C. E., Aragona, K. M., Clark, E., Lunak, M., & Erickson, P. S. (2016). Predicting colostrum quality from performance in the previous lactation and environmental changes. *Journal of Dairy Science, 99*(5), 4048–4055. https://doi.org/10.3168/jds.2015-9868

Cattaneo, T. M. P., Cabassi, G., Profaizer, M., & Giangiacomo, R. (2009). Contribution of Light Scattering to near Infrared Absorption in Milk. *Journal of Near Infrared Spectroscopy, 17*(6), 337–343. https://doi.org/10.1255/jnirs.867

Ceniti, C., Froiio, F., Britti, D., Paolino, D., & Costanzo, N. (2019). Rheological characteristics of bovine colostrum and their correlation with immunoglobulin G. *International Journal of Dairy Technology, 72*(3), 345–349. https://doi.org/10.1111/1471-0307.12593

Costa A., Franzoi M., Goi A., Visentin G., & De Marchi, M. (2021a). Genetic aspects of immunoglobulins G concentration measured with gold standard and predicted from infrared spectra in bovine colostrum. Accepted as oral presentation to the 24th Congress of the Animal Science and Production Association (ASPA), 21-24 September 2021, Padova, Italy.

Costa, A., Goi, A., Penasa, M., Nardino, G., Posenato, L., & De Marchi, M. (2021b). Variation of immunoglobulins G, A, and M and bovine serum albumin concentration in Holstein cow colostrum. *Animal, 15*(7), 100299. https://doi.org/10.1016/j.animal.2021.100299

De Marchi, M., Penasa, M., Zidi, A., & Manuelian, C. L. (2018). Invited review: Use of infrared technologies for the assessment of dairy products—Applications and perspectives. *Journal of Dairy Science, 101*(12), 10589–10604. https://doi.org/10.3168/jds.2018-15202

Denholm, S. J., Brand, W., Mitchell, A. P., Wells, A. T., Krzyzelewski, T., Smith, S. L., … Coffey, M. P. (2020). Predicting bovine tuberculosis status of dairy cows from mid-infrared spectral data of milk using deep learning. *Journal of Dairy Science, 103*(10), 9355–9367. https://doi.org/10.3168/jds.2020-18328

Duff, W. R. D., Chilibeck, P. D., Rooke, J. J., Kaviani, M., Krentz, J. R., & Haines, D. M. (2014). The effect of bovine colostrum supplementation in older adults during resistance training. *International Journal of Sport Nutrition and Exercise Metabolism, 24*(3), 276–285. https://doi.org/10.1123/ijsnem.2013-0182

Elsohaby, I., McClure, J. T., Hou, S., Riley, C. B., Shaw, R. A., & Keefe, G. P. (2016). A novel method for the quantification of bovine colostral immunoglobulin G using infrared spectroscopy. *International Dairy Journal, 52*, 35–41. https://doi.org/10.1016/j.idairyj.2015.08.004

Elsohaby, I., McClure, J. T., Cameron, M., Heider, L. C., & Keefe, G. P. (2017). Rapid assessment of bovine colostrum quality: How reliable are transmission infrared spectroscopy and digital and optical refractometers? *Journal of Dairy Science, 100*(2), 1427–1435. https://doi.org/10.3168/jds.2016-11824

Elsohaby, I., Windeyer, M. C., Haines, D. M., Homerosky, E. R., Pearson, J. M., McClure, J. T., & Keefe, G. P. (2018). Application of transmission infrared spectroscopy and partial least squares regression to predict immunoglobulin G concentration in dairy and beef cow colostrum. *Journal of Animal Science, 96*(2), 771–782. https://doi.org/10.1093/jas/sky003

Gamsjäger, L., Elsohaby, I., Pearson, J. M., Levy, M., Pajor, E. A., Haines, D. M., & Windeyer, M. C. (2020). Assessment of Brix refractometry to estimate immunoglobulin G concentration in beef cow colostrum. *Journal of Veterinary Internal Medicine, 34*(4), 1662–1673. https://doi.org/10.1111/jvim.v34.410.1111/jvim.15805

Gapper, L. W., Copestake, D. E. J., Otter, D. E., & Indyk, H. E. (2007). Analysis of bovine immunoglobulin G in milk, colostrum and dietary supplements: A review. *Analytical and Bioanalytical Chemistry, 389*(1), 93–109. https://doi.org/10.1007/s00216-007-1391-z

Gelsinger, S. L., Smith, A. M., Jones, C. M., & Heinrichs, A. J. (2015). Technical note: Comparison of radial immunodiffusion and ELISA for quantification of bovine immunoglobulin G in colostrum and plasma. *Journal of Dairy Science, 98*(6), 4084–4089. https://doi.org/10.3168/jds.2014-8491

Godden, S. (2008). Colostrum management for dairy calves. The Veterinary Clinics of North America. *Food Animal Practice, 24*(1), 19–39. https://doi.org/10.1016/j.cvfa.2007.10.005

Guo, Z., Barimah, A. O., Shujat, A., Zhang, Z., Ouyang, Q., Shi, J., … Chen, Q. (2020). Simultaneous quantification of active constituents and antioxidant capability of green tea using NIR spectroscopy coupled with swarm intelligence algorithm. *LWT, 129*, 109510. https://doi.org/10.1016/j.lwt.2020.109510

Holroyd, S. E. (2013). The use of near infrared spectroscopy on milk and milk products. *Journal of Near Infrared Spectroscopy, 21*(5), 311–322. https://doi.org/10.1255/jnirs.1055

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* ((3rd ed.).). Wiley Inc. (Chapter 5.2).

Hurley, W. L., & Theil, P. K. (2011). Perspectives on immunoglobulins in colostrum and milk. *Nutrients, 3*(4), 442–474. https://doi.org/10.3390/nu3040442

Jiang, L., Liu, F., & He, Y. (2012). A non-destructive distinctive method for discrimination of automobile lubricant variety by visible and short-wave infrared spectroscopy. *Sensors, 12*(3), 3498–3511. https://doi.org/10.3390/s120303498

Juhl, S. M., Ye, X., Zhou, P., Li, Y., Iyore, E. O., Zhang, L., Jiang, P., van Goudoever, J. B., Greisen, G., & Sangild, P. T. (2018). Bovine colostrum for preterm infants in the first days of life: a randomized controlled pilot trial. Journal of Pediatric Gastroenterology and Nutrition, 66(3), 471–478. 10.1097/MPG.0000000000001774.

Kalivas, J. H., Roberts, N., & Sutter, J. M. (1989). Global optimization by simulated annealing with wavelength selection for ultraviolet-visible spectrophotometry. *Analytical Chemistry, 61*(18), 2024–2030. https://doi.org/10.1021/ac00193a006

Lee, A., Pontin, M. C. F., Kosmerl, E., Jimenez-Flores, R., Moretti, D. B., & Ziouzenkova, O. (2019). Assessment of adipogenic, antioxidant, and anti-inflammatory properties of whole and whey bovine colostrum. *Journal of Dairy Science, 102*(10), 8614–8621. https://doi.org/10.3168/jds.2019-16509

Liu, J., Li, N., Zhen, F., Xu, Y., Li, W., & Sun, Y. (2019). Rapid detection of carbon-nitrogen ratio for anaerobic fermentation feedstocks using near-infrared spectroscopy combined with BiPLS and GSA. *Applied Optics, 58*(18), 5090–5097. https://doi.org/10.1364/AO.58.005090

López-Lorente, Á. I., & Mizaikoff, B. (2016). Mid-infrared spectroscopy for protein analysis: Potential and challenges. *Analytical and Bioanalytical Chemistry, 408*(11), 2875–2889. https://doi.org/10.1007/s00216-016-9375-5

McGrath, B. A., Fox, P. F., McSweeney, P. L. H., & Kelly, A. L. (2016). Composition and properties of bovine colostrum: A review. *Dairy Science & Technology, 96*(2), 133–158. https://doi.org/10.1007/s13594-015-0258-x

Mir, R., Singh, N., Vikram, G., Kumar, R. P., Sinha, M., Bhushan, A., … Singh, T. P. (2009). The structural basis for the prevention of nonsteroidal antiinflammatory drug-induced gastrointestinal tract damage by the c-lobe of bovine colostrum lactoferrin. *Biophysical Journal, 97*(12), 3178–3186. https://doi.org/10.1016/j.bpj.2009.09.030

Navrátilová, P., Hadra, L., Dračková, M., Janštová, B., Vorlová, L., & Pavlata, L. (2006). Use of FT-NIR spectroscopy for bovine colostrum analysis. *Acta Veterinaria Brno, 75*(1), 57–63. https://doi.org/10.2754/avb200675010057

Playford, R. J., Cattell, M., & Marchbank, T. (2020). Marked variability in bioactivity between commercially available bovine colostrum for human use; implications for clinical trials. *PLOS ONE, 15*(6), Article e0234719. https://doi.org/10.1371/journal.pone.0234719

Quigley, J. D., Lago, A., Chapman, C., Erickson, P., & Polo, J. (2013). Evaluation of the Brix refractometer to estimate immunoglobulin G concentration in bovine colostrum. *Journal of Dairy Science, 96*(2), 1148–1155. https://doi.org/10.3168/jds.2012-5823

Rivero, M. J., Valderrama, X., Haines, D., & Alomar, D. (2012). Prediction of immunoglobulin G content in bovine colostrum by near-infrared spectroscopy. *Journal of Dairy Science, 95*(3), 1410–1418. https://doi.org/10.3168/jds.2011-4532

Scopus (2021). Retrieved from www.scopus.com. Accessed June 6, 2021.

Solórzano, J. C. (2020). Genetics of colostrum, milk and serum antibodies in dairy cattle. *Acta Universitatis Agriculturae Sueciae.*

Strekozov, N. I., Motova, E. N., & Fedorov, Y. N. (2008). Evaluation of the chemical composition and immunological properties of colostrum of cows' first milk yield. *Russian Agricultural Sciences, 34*(4), 259–260. https://doi.org/10.3103/S1068367408040149

Swierenga, H., de Groot, P. J., de Weijer, A. P., Derksen, M. W. J., & Buydens, L. M. C. (1998). Improvement of PLS model transferability by robust wavelength selection. *Chemometrics and Intelligent Laboratory Systems, 41*(2), 237–248. https://doi.org/10.1016/S0169-7439(98)00055-0

US Department of Health and Human Services, & Food and Drug Administration. (2001). Guidance for industry: Bioanalytical method validation. Retrieved from http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm070107.pdf. Accessed May 30, 2021.

Williams, P. (2007). Statistical terms for evaluation of accuracy and precision. *Getting The Best Out Of Light.* 5.0 Ed. *PDK Grain, 5*, 1–17.

Williams, P. (2014). The RPD Statistic: A Tutorial Note. *NIR News, 25*(1), 22–26. https://doi.org/10.1255/nirn.1419

Workman, J. (2016). The Concise Handbook of Analytical Spectroscopy: Theory, Applications, and Reference Materials: Volume 3: Near Infrared Spectroscopy (Vol. 3). WORLD SCIENTIFIC. 10.1142/8800-vol3.

Xiaobo, Z., Jiewen, Z., Povey, M. J. W., Holmes, M., & Hanpin, M. (2010). Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta, 667*(1), 14–32. https://doi.org/10.1016/j.aca.2010.03.048

Zeaiter, M., Roger, J.-M., Bellon-Maurel, V. D., & Rutledge, N. (2004). Robustness of models developed by multivariate calibration. Part I: The assessment of robustness. *TrAC Trends in Analytical Chemistry, 23*(2), 157–170. https://doi.org/10.1016/S0165-9936(04)00307-3