



Comparison of Bayesian regression models and partial least squares regression for the development of infrared prediction equations

V. Bonfatti,^{*1} F. Tiezzi,[†] F. Miglior,^{‡§} and P. Carnier^{*}

^{*}Department of Comparative Biomedicine and Food Science, University of Padova, 35020, Legnaro, Italy

[†]Department of Animal Science, North Carolina State University, Raleigh 27695

[‡]Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, N1G 2W1, Ontario, Canada

[§]Canadian Dairy Network, Guelph, N1K 1E5, Ontario, Canada

ABSTRACT

The objective of this study was to compare the prediction accuracy of 92 infrared prediction equations obtained by different statistical approaches. The predicted traits included fatty acid composition ($n = 1,040$); detailed protein composition ($n = 1,137$); lactoferrin ($n = 558$); pH and coagulation properties ($n = 1,296$); curd yield and composition obtained by a micro-cheese making procedure ($n = 1,177$); and Ca, P, Mg, and K contents ($n = 689$). The statistical methods used to develop the prediction equations were partial least squares regression (PLSR), Bayesian ridge regression, Bayes A, Bayes B, Bayes C, and Bayesian least absolute shrinkage and selection operator. Model performances were assessed, for each trait and model, in training and validation sets over 10 replicates. In validation sets, Bayesian regression models performed significantly better than PLSR for the prediction of 33 out of 92 traits, especially fatty acids, whereas they yielded a significantly lower prediction accuracy than PLSR in the prediction of 8 traits: the percentage of C18:1n-7 *trans*-9 in fat; the content of unglycosylated κ -casein and its percentage in protein; the content of α -lactalbumin; the percentage of α_{S2} -casein in protein; and the contents of Ca, P, and Mg. Even though Bayesian methods produced a significant enhancement of model accuracy in many traits compared with PLSR, most variations in the coefficient of determination in validation sets were smaller than 1 percentage point. Over traits, the highest predictive ability was obtained by Bayes C even though most of the significant differences in accuracy between Bayesian regression models were negligible.

Key words: infrared spectra, fatty acid, protein fraction, Bayesian regression

INTRODUCTION

Mid-infrared spectroscopy is a recognized tool for predicting novel milk traits for payment systems, management of dairy cows, and selective breeding purposes (Gengler et al., 2016). The utility of prediction (regression) equations for practical applications depends mostly on their accuracy. According to Soyeurt et al. (2011), equations with a coefficient of determination (R^2) greater than 0.95 in cross-validation are useful in milk payment systems. For management purposes, the usefulness of equations depends on the correlation between the predicted traits and the management indicators (e.g., prevalence of metabolic disorders). For selective breeding, the usefulness of infrared predictions relies mostly on their heritability and on their genetic correlation with the breeding goal traits. Even though equations with medium to low accuracy might be successfully used for breeding purposes, more accurate prediction equations would lead to a faster genetic progress because a positive relationship exists between the R^2 in cross-validation and the estimated genetic correlation between measured and predicted traits (Rutten et al., 2010; Bonfatti et al., 2017). Thus, there is interest in finding chemometric methods that can increase the accuracy of prediction models.

To date, prediction equations have been developed mostly using partial least squares regression (PLSR). Recently, Bayesian models adopted for regression on high-dimensional genotypes have been reported to dramatically increase the prediction accuracy of infrared prediction equations for the prediction of 8 traits related to fatty acid (FA) composition and technological properties of milk (Ferragina et al., 2015). However, results might depend on the traits as well as on the conditions under which models are developed. For example, the presence of noise regions in the development of prediction models might affect the accuracy of PLSR (De Marchi et al., 2009; Bonfatti et al., 2011; Eskildsen et al., 2014) without influencing the accuracy of Bayesian regressions, as those methods allow shrinkage and perform variable selection, down-weighting or exclud-

Received October 25, 2016.

Accepted May 4, 2017.

¹Corresponding author: valentina.bonfatti@unipd.it

ing the uninformative variables (de los Campos et al., 2013).

The objective of this study was to compare the accuracy of prediction equations obtained by PLSR and by Bayesian regression methods for predicting a large number of traits related to milk fine composition and technological properties. Comparison among models was performed considering different numbers of spectral variables and PLSR terms for prediction equation development.

MATERIALS AND METHODS

Reference Data

A total of 1,330 individual milk samples of Simmental cows were collected for reference analyses. Samples were collected during the morning milking in 21 herds located in northern Italy. Herd size ranged from 30 to 125 cows. Cows were between 5 and 484 DIM and ranged from 1 to 9 parities. All samples, with the exception of those that were lost during the analysis due to nonmatching sample numbers or poor preservation, were analyzed for pH, milk coagulation properties, micro-cheese yield, micro-cheese composition, and protein profile. Measures of pH and milk coagulation properties were available for 1,296 samples. Contents of α_{S1} -CN, α_{S2} -CN, β -CN, γ -CN, glycosylated and unglycosylated κ -CN, β -LG, and α -LA of individual milk samples were measured in 1,137 samples. Curd yield and composition were obtained for 1,177 samples using a micro-cheese making procedure and measuring DM content, protein content, and fat content in the micro-curds.

Samples from cows with known sire and dam (1,040 cows in 20 herds; daughters of 378 sires) were also analyzed for FA composition. Part of these samples, depending on budget constraints, was also analyzed for mineral contents and lactoferrin (LF). Contents of Ca, P, Mg, and K were analyzed in 689 samples from 15 herds, whereas LF was measured in 558 samples from 11 herds. All samples with measures of mineral contents, LF, and FA also had measures of pH, milk coagulation properties, protein profile, and cheese yield. In total, 92 traits were investigated. Details on the methods used to obtain the reference traits, as well as descriptive statistics of all the investigated traits, can be found in Bonfatti et al. (2016).

Milk Infrared Spectra

Infrared absorption spectra (1,060 variables) were collected on all samples by the Friuli Venezia Giulia Milk Recording Agency laboratory (Codroipo, Italy) using a MilkoScan FT6000 (Foss Electric A/S, Hillerød,

Denmark). Spectra exhibiting a Mahalanobis distance from the population centroid greater than 3 were considered to be outliers and were discarded. Records with a trait value above 4 or below -4 SD from the mean were also excluded. For each trait, the number of outliers ranged from 5 to 11 samples. Spectra variables were standardized to a null mean and a unit variance before the analysis.

Due to the interference of water absorption, the O–H bending and stretching regions of the spectra (between 1,628 and 1,658 cm^{-1} and between 3,105 and 3,444 cm^{-1} , respectively) are assumed to contain no useful chemical information and to have very high coefficients of variation and a very low heritability (Soyeurt et al., 2010); for these reasons, they are often discarded before the chemometric analysis (De Marchi et al., 2009; Bonfatti et al., 2011; Eskildsen et al., 2014). Analyses were performed on the totality of the spectra variables and on the spectra variables cleared of the 2 water absorption regions of the spectra. In the latter case, prediction equations were developed using 872 spectra variables. The derivative of the spectral data is often used because it enhances resolution by sharpening the absorption bands and it removes baseline offset (Burns and Ciurczak, 2001). Models were fitted to both raw spectra and spectra transformed with a first derivative mathematical treatment. The results obtained from treated and untreated spectra were very similar, and only those obtained for the raw spectra are shown.

Prediction Equations

Prediction equations can be considered as a series of partial regression coefficients, in number equal to the number of wavelengths, providing a prediction of a measurable trait in a sample. In the present study, these regression coefficients were estimated using 6 different methods. The PLSR implemented in the R (R Development Core Team, 2013) package PLS (Mevik and Wehrens, 2007) was used as the reference method. The PLSR models were fitted by either (1) choosing the number of PLSR components that minimized the prediction error in the training set (**PLSR-M**) or (2) setting the maximum number of components to 16 (**PLSR-16**). The PLSR-16 was performed only on the totality of the spectra variables and was carried out to compare the results with previous literature estimates obtained using WinISI II software (InfraSoft International, State College, PA), for which the maximum number of terms is 16 by default. In addition, PLSR and modified PLSR (**MPLSR**; Shenk and Westerhaus, 1991) were performed on 1,060 spectra variables using WinISI II software to enable the comparison between results yielded by different software products.

The regression coefficients that are part of the prediction equations can be obtained by multiple linear regression implemented in a Bayesian framework. The algorithms used in this study were Bayesian ridge regression (**BRR**), Bayes A (**BA**), Bayes B (**BB**), Bayes C (**BC**), and Bayesian least absolute shrinkage and selection operator (**LASSO**; **BL**) and were tested as alternative methods to PLSR. These are penalization methods (i.e., they are used to reduce the estimator variance to guarantee the stability of the estimates). The models differ in the form of the prior density assigned to the effects. In BRR, effects are assigned Gaussian priors. This shrinks the estimate toward zero, and the extent of shrinkage is homogeneous across effects. In BA, effects are assigned a scaled t density (Meuwissen et al., 2001). The estimated effects of variables that are weakly correlated with the trait are strongly shrunk toward zero, whereas those of variables with strong association with the trait are shrunk to a lesser extent. The prior for BB and BC is a 2-component mixture with a point of mass at zero and with a slab that can be either a scaled t (in BB; Habier et al., 2011) or a Gaussian (in BC; Habier et al., 2011) density. These 2 models perform variable selection. For BL, the prior for the effects is a double-exponential density (Park and Casella, 2008).

All the Bayesian regression models listed above were implemented in the BGLR package (de los Campos and Pérez-Rodríguez, 2014) of the R software. In all the Bayesian models, the number of iterations was 100,000, with a burn-in of 20,000. The high-order hyperparameters of the models were specified using built-in BGLR rules that select default values for these unknowns to yield proper but relatively uninformative priors. The priors are chosen in a way that, a priori, they obey a variance partition of the y variable into components attributable to the error terms and to each of the elements of the linear predictors. This partition (R^2 of the model) is set equal to 0.5. Pérez and de los Campos (2014) provide a detailed description of the models and algorithms implemented in BGLR. Briefly, in BRR the normal distribution assigned to the effects has mean zero and variance σ_β^2 . The variance parameter is assigned a scaled-inverse chi-squared density, with parameters df_β and S_β . The BGLR package sets $df_\beta = 5$ and solves for the scale parameter S_β to match the R^2 of the model. In BA, the marginal distribution of the effects is a scaled t density, with parameters df_β and S_β . The BGLR package sets $df_\beta = 5$, whereas $S_\beta \sim \text{gamma}(r, s)$. It also sets $s = 1.1$ and solves for the rate r so that the total contribution of the linear predictor matches the R^2 of the model. Models BB and BC extend BA and BRR, respectively, by introducing an additional parameter π representing the prior proportion

of nonzero effects. In BGLR, $\pi \sim \text{beta}(p_0, \pi_0)$, with $p_0 = 10$ and $\pi_0 = 0.5$. In BL, effects are assigned independent normal densities with the null mean and effect-specific variance parameter $\tau_{jk}^2 \times \sigma_\varepsilon^2$. The residual variance is assigned a scaled-inverse chi-squared density, and the effect-specific scale parameters, τ_{jk}^2 , are assigned independent and identically distributed exponential densities with rate parameter $\lambda^2/2$, where $\lambda^2 \sim \text{gamma}(r, s)$, $s = 1.1$, and r is solved to match the expected R^2 of the model.

Model Comparison

In practice, prediction equations are developed using data from a restricted number of samples and applied over a larger population. Prediction models should cover the variability expected in the population over which the models are to be applied. The effect of the herd represents the main source of spectra variability. Hence, testing the predictive ability of models using validation sets formed by herds that are not included in the training set would be the most conservative approach. In our study, this approach was not possible due to the limited number of herds. Therefore, individual records were randomly assigned to training and validation sets. As a result, records from the same herd are likely to appear in both training and validation sets, leading to a possible overestimation of model robustness. Data available for each trait, after outlier elimination, were randomly assigned to 10 nonoverlapping subsets. Prediction equations were developed using 9 subsets at a time and validated on the remaining subset. For each trait, models were tested on the same subsets of data.

Prediction equations were therefore obtained for each trait in each training set by either PLSR or Bayesian multiple regression models. The regression coefficients were then multiplied by the respective wavelengths of the samples included in the validation set to obtain a predicted value for that trait. Prediction performance was assessed by computing the R^2 between the predicted and measured values in training (R^2_t) and validation (R^2_v) and the root-mean-squared error of prediction in training ($RMSEP_t$) and validation ($RMSEP_v$). The difference in R^2_t , R^2_v , $RMSEP_t$, and $RMSEP_v$ between models was assessed for each trait by paired t -tests performed using the R software.

RESULTS AND DISCUSSION

Accuracy of PLSR Models Performed Under Different Scenarios

The mean of the estimated model parameters and the distribution of the number of terms obtained with

PLSR-16 and PLSR-M is reported in Table 1. With PLSR-M, only 4 traits had a maximum number of terms equal to or lower than 16: the time from rennet addition to curd firming, LF, fat curd yield, and protein content in the curd. For 88 traits, a higher number of latent variables (on average 55) was required to reach the minimum prediction error.

With PLSR-16, values of R^2_t and R^2_v across traits were on average 13 and 9 percentage points lower than those yielded by PLSR-M, respectively. The difference in R^2_v ranged from -31 (for the percentage of C18:1n-7 *trans*-9 in fat) to -2 (for the content of SFA) percentage points. In all traits, prediction accuracy in training was significantly affected by the number of PLSR terms chosen, whereas the accuracy in validation was significantly different for 67 traits. Overall, in these 67 traits, the decrease in R^2_v obtained by PLSR-16 compared with PLSR-M was approximately 12 percentage points. Hence, limiting the number of PLSR terms to 16 reduced the prediction accuracy of the models in the large majority of the traits. One may argue that increasing the number of PLSR terms might lead to overfitting. However, it is worth noting that all PLSR models were performed under a 10-fold cross-validation, thus ensuring the identification of the optimal number of PLSR terms while avoiding overfitting.

A maximum number of PLSR terms was used to enable the comparison of our results with those that would be expected using the WinISI II software, for which the maximum number of terms is 16. Under this condition, the PLSR performed using the R software led to very similar results as the MPLSR procedure implemented in the WinISI II. Across traits, the correlation between the 2 systems for R^2_t was 0.97, with a bias of -0.001 (results not reported in tables). This indicates that R and WinISI II are expected to yield consistent results, and a higher number of terms might effectively enhance their predictive performances.

When PLSR-M was used on 872 spectral variables instead of 1,060, the average number of PLSR terms necessary to reach the minimum prediction error across traits decreased from 55 to 29, with 65 models having fewer than 35 components (data not reported in tables). Even though the accuracy of prediction equations built after spectral variable selection was very similar to that yielded by models developed on the entire spectra, the average number of terms necessary to reach that accuracy was almost halved. This suggests that PLSR models perform better when water absorption regions of the spectra are excluded from the analysis, and it indicates that models developed on 872 variables are expected to be more robust.

However, also in this scenario, the number of PLSR terms necessary to minimize the prediction error was less than or equal to 16 for few traits: DM curd yield, fat curd yield, protein content in the curd, and percentage of β -LG in protein. Figure 1 shows the relationship between the average R^2_v obtained for each trait by PLSR-16 and PLSR-M using either the whole spectra or the selected spectra variables. Results suggest that 16 PLSR terms are not sufficient to exploit all the potential accuracy of models, even when noise spectral regions are omitted.

Results of the comparison between parameters yielded by models applied to the whole spectra (1,060 variables) and to the spectra cleared of the water absorption regions (872 variables) are reported in Table 2. Across traits, the average R^2_v obtained using 872 spectral variables was 3.2 percentage points higher than that obtained using the whole spectrum. For 18 traits, the difference in R^2_v under the 2 approaches was significant and was on average 5.3 percentage points, with a maximum difference of 8.9 percentage points for the content of unglycosylated κ -CN.

The overall difference in R^2_v between PLSR-16 performed on 1,060 spectra variables and PLSR-M performed on 872 variables was around 12 percentage points, with 69 traits (i.e., 75% of the investigated traits) being significantly affected by the different scenario in which PLSR models were fitted. For those traits, the average difference in R^2_v for the 2 methods was 15 percentage points (results not reported in tables). These results indicate that PLSR models can have a very different predictive ability depending on the conditions under which models are fitted to the data.

Table 1. Mean \pm SD of predictive performance parameters obtained by a partial least squares regression (PLSR) in which the maximum number of terms was set to 16 (PLSR-16) and a PLSR in which the number of terms was chosen to minimize the prediction error (PLSR-M)¹

Parameter ²	PLSR-16	PLSR-M
R^2_t	59.5 \pm 17.8	72.1 \pm 17.9
R^2_v	51.6 \pm 20.7	60.4 \pm 21.4
RMSEPt	1.63 \pm 7.09	1.51 \pm 7.09
RMSEpv	1.75 \pm 7.41	1.68 \pm 7.41
No. of PLSR terms		
Mean	15.6	54.6
Minimum	5	5
Maximum	16	92
No. of traits with no. of PLSR terms \leq 16	92	4

¹Comparison between PLSR-16 and PLSR-M was performed using the whole spectra (1,060 variables).

² R^2_t = coefficient of determination in training; R^2_v = coefficient of determination in validation; RMSEPt = root-mean-squared error in training; RMSEpv = root-mean-squared error in validation.

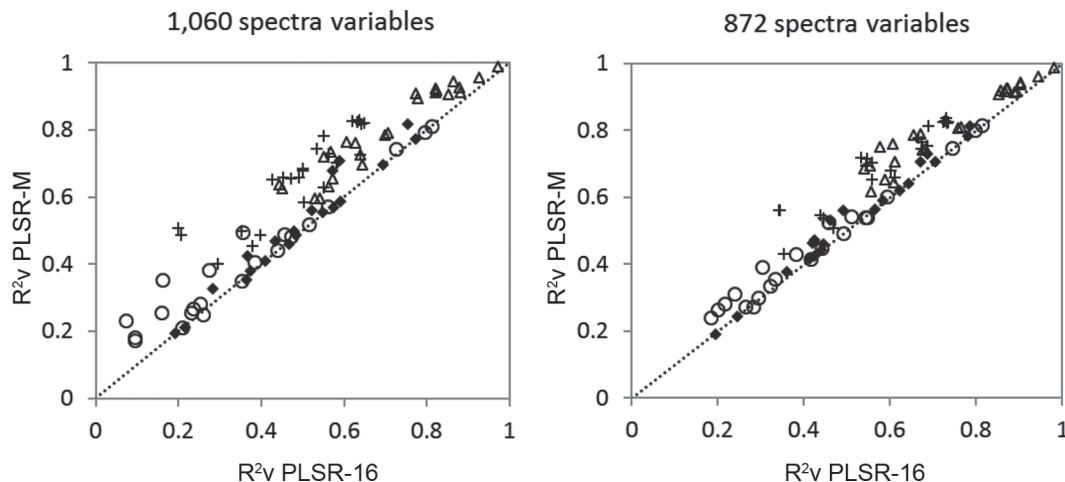


Figure 1. Relationship between the coefficient of determination of validation obtained by a partial least squares regression (PLSR) in which the maximum number of terms was set to 16 (PLSR-16) and a PLSR in which the number of terms was chosen to minimize the prediction error (PLSR-M). Predicted traits were fatty acid contents (Δ); fatty acid percentage in fat (+); protein fraction contents and percentage in total protein (\circ); and technological traits, lactoferrin, and minerals (\blacklozenge).

Accuracy of Bayesian Models Developed Using the Whole Spectra or Selected Spectra Regions

Results of the comparison between model parameters obtained with Bayesian regression models applied on 1,060 and 872 spectral variables are reported in Table 2. Generally, results obtained for Bayesian regression models were consistent with those observed for PLSR-M, with an increased prediction accuracy after the exclusion of the water absorption regions. For BRR, BA, and BL, the average difference in R^2v when using 1,060 or 872 spectral variables ranged from 4 to 7 percentage points. The models less affected by the inclusion of the noise regions were BB and BC, for which the difference in R^2v was significant in only 3 and 5 traits, respectively. This result was expected because those models perform variable selection excluding the uninformative variables (de los Campos et al., 2013).

For BB, the traits favorably affected by the use of the selected spectra regions were the content of unglycosylated κ -CN and the percentages of unglycosylated κ -CN and total κ -CN in protein. For BC, in addition to such traits, the percentage of Σ C18:1 *trans* in fat and the content of C18:1n-7 *trans*-9 had a significantly higher R^2v when prediction equations were developed using 872 spectral variables.

Theoretically, informative wavelengths might also exist in the spectra regions corresponding to water absorbance, and the a priori exclusion of those regions could lead to a loss of potential useful information for the prediction of new phenotypes. Prediction equations as obtained with BB or BC would benefit from the use of the whole spectra because those 2 Bayesian methods

might select informative wavelengths potentially hidden in the noise regions.

The BB and BC models performed on 1,060 variables, however, yielded the same prediction accuracy of models carried out on 872 variables, and the prediction for few traits was negatively affected (-15.5 percentage points in R^2v , on average) by the presence of the noise regions. This indicates that BB and BC were not able to extract any additional information from the variables included in the water absorption regions of the spectra.

Accuracy of PLSR-M Models Across Traits

Values of R^2v obtained by PLSR for all the investigated traits are reported in the Appendix tables. Results obtained by PLSR analysis of the data used in this study have already been published (Bonfatti et al., 2016), albeit no measures of variability in the performance parameters were reported. Although results were in the range of variation of those presented in this study, the R^2 values reported by Bonfatti et al. (2016) were slightly lower. This might be attributed to the different approach used to determine the optimal number of PLSR terms. In Bonfatti et al. (2016), no additional PLSR components were retained if the increase in the R^2 of the model was smaller than 0.002. In this study, the number of PLSR terms was the one corresponding to the minimum prediction error, leading to a slightly higher number of terms. A remarkable difference between the 2 studies was detected for the content of C18:3n-3, for which the R^2 reported previously (29%) was much lower than the one estimated in this study (71%). Comparison of the accuracy of the PLSR mod-

Table 2. Mean \pm SD of the prediction performance parameters obtained across traits and across models using the whole spectra (1,060 variables) and mean \pm SD of the difference between parameters obtained using the spectra with the exclusion of the water absorbance regions (872 variables) and the whole spectra

Parameter ¹	No. of spectra variables	No. of traits ²	PLSR-M ³	Bayes RR ⁴	Bayes A	Bayes B	Bayes C	LASSO ⁵
R ² _t	1,060	All	71.1 \pm 17.9	63.2 \pm 23.4	65.0 \pm 22.3	66.7 \pm 20.8	67.2 \pm 20.8	64.6 \pm 22.9
	872	All	-1.2 \pm 2.5	4.6 \pm 8.5	2.6 \pm 5.7	1.3 \pm 3.7	1.2 \pm 4.3	3.7 \pm 6.5
	872	Selected	-1.5 \pm 2.7 (69)	5.3 \pm 8.9 (81)	3.6 \pm 6.4 (67)	1.8 \pm 4.3 (66)	1.8 \pm 5.2 (60)	4.6 \pm 6.9 (75)
R ² _v	1,060	All	60.4 \pm 21.4	56.5 \pm 24.5	58.8 \pm 23.7	61.2 \pm 22.4	62.4 \pm 22.4	58.0 \pm 24.0
	872	All	3.2 \pm 2.5	6.6 \pm 7.5	4.4 \pm 4.9	1.5 \pm 3.3	1.3 \pm 3.5	5.5 \pm 5.6
	872	Selected	5.3 \pm 1.8 (18)	10.9 \pm 8.1 (47)	9.0 \pm 5.3 (30)	15.5 \pm 5.5 (3)	12.0 \pm 8.3 (5)	9.4 \pm 5.7 (43)
RMSEPt	1,060	All	1.51 \pm 7.09	1.55 \pm 6.85	1.55 \pm 6.82	1.53 \pm 6.79	1.53 \pm 6.84	1.55 \pm 6.86
	872	All	-0.09 \pm 0.89	-0.02 \pm 0.22	-0.03 \pm 0.10	-0.02 \pm 0.11	-0.00 \pm 0.20	-0.03 \pm 0.15
	872	Selected	-0.14 \pm 1.12 (60)	-0.03 \pm 0.25 (71)	-0.04 \pm 0.11 (79)	-0.02 \pm 0.12 (80)	-0.00 \pm 0.24 (66)	-0.03 \pm 0.17 (69)
RMSEPV	1,060	All	1.68 \pm 7.41	1.68 \pm 7.33	1.66 \pm 7.30	1.62 \pm 7.24	1.63 \pm 7.30	1.67 \pm 7.31
	872	All	-0.09 \pm 0.30	-0.08 \pm 0.13	-0.07 \pm 0.17	-0.04 \pm 0.10	-0.03 \pm 0.09	-0.07 \pm 0.12
	872	Selected	-0.04 \pm 0.09 (30)	-0.07 \pm 0.09 (42)	-0.07 \pm 0.09 (28)	-0.09 \pm 0.09 (4)	-0.09 \pm 0.10 (4)	-0.06 \pm 0.09 (38)

¹R²_t = coefficient of determination in training; R²_v = coefficient of determination in validation; RMSEPt = root-mean-squared error in training; RMSEPV = root-mean-squared error in validation.

²Selected traits correspond to the traits for which the difference between estimates of a parameter obtained using 1,060 or 872 spectra variables was significant ($P < 0.05$). The number of selected traits is reported in parentheses.

³PLSR-M = partial least squares regression in which the number of terms was chosen to minimize the prediction error.

⁴RR = ridge regression.

⁵LASSO = least absolute shrinkage and selection operator.

els for all the investigated traits with other results in the literature can be found in Bonfatti et al. (2016).

Accuracy of Bayesian Models Compared with PLSR for the Prediction of FA Contents

The comparison between performance parameters of PLSR-M and Bayesian regression models was based on the analysis of spectra cleared of the water absorption regions (i.e., in the scenario in which all models gave the best prediction accuracy). Overall, R^2_t yielded by Bayesian regressions was 2 to 8% lower than that provided by PLSR-M, depending on the Bayesian method. The lower R^2 was also accompanied by a 4 to 5% higher RMSEP_t (results not reported in tables).

Comparisons among the R^2_v of the 6 different models for the prediction of FA contents are reported in Figure 2. Despite a lower prediction accuracy in training sets, Bayesian regression models yielded a significantly higher R^2_v than PLSR-M, indicating that they are more robust than PLSR-M. In particular, BRR and BA performed significantly better than PLSR-M for the prediction of 3 traits, whereas BB, BC, and BL models performed significantly better than PLSR-M in 9, 11, and 6 traits, respectively. Hence, the model providing the best results across traits was BC, and short- and medium-chain individual FA were the traits benefiting more from the use of this approach. However, the maximum difference in R^2_v due to the use of Bayesian methods compared with PLSR-M was 1.5 percentage points, obtained for $\Sigma C_{16:1}$ with BC. For all the other traits, the difference in R^2_v ranged from 0.1 to 0.4 percentage points. Hence, although being statistically significant, most variations in R^2_v produced by Bayesian models were smaller than 1 percentage point. The RMSEP_v yielded by Bayesian models across traits was on average 0.5 to 1.8% lower than that obtained by PLSR-M. For the traits for which the model used significantly affected prediction accuracy, BC yielded on average a 3% lower RMSEP_v than PLSR-M and the improvement in RMSEP_v was never greater than 4% (results not reported in tables).

Accuracy of Bayesian Models Compared with PLSR for the Prediction of FA Percentages

Results of the comparison between models obtained for percentages of FA were very similar to those observed for FA contents. Prediction equations developed using Bayesian regression models generally led to lower values of R^2_t and slightly greater (+5%) RMSEP_t. The differences between the R^2_v of Bayesian models and PLSR-M for the prediction of FA percentages are

reported in Figure 2. Bayesian models yielded similar or better results than those produced by PLSR-M, with the only exception of BRR for the prediction of C18:1n-7 *trans*-9, which led to a slightly worse accuracy (−2 percentage points of R^2_v) than PLSR-M. Like for FA content, the best performances were obtained by BC. The improvements in R^2_v yielded by the different models were on average 1 percentage point, with a maximum value of 2.3 percentage points. Hence, the benefits deriving from the use of Bayesian models instead of PLSR for FA percentage were slightly greater than those observed for FA contents. On average, for the 13 traits for which models yielded a significantly different accuracy, the RMSEP_v of Bayesian models was 3% lower than that obtained by PLSR-M.

This was in contrast to the findings of Ferragina et al. (2015), who reported that BRR, BA, and BB showed significantly higher prediction accuracy than PLSR-16 or MPLSR. Comparison between values of R^2_v across studies might not always be appropriate due to the different experimental conditions under which validation has been performed (herds vs. individuals assigned to training and validation sets). For this reason, results obtained by PLSR-16 and PLSR-M in the current or other literature studies were compared with the R^2 values estimated by Ferragina et al. (2015) on their entire data set, when individuals were randomly assigned to the training or validation sets.

Results by Ferragina et al. (2015) showed that the improvements in accuracy provided by the Bayesian models compared with PLSR-16 were around 30 percentage points for the R^2 calculated on the entire data set. Differences in accuracy compared with MPLSR were lower but still remarkable (on average, 19 percentage points in R^2). Using PLSR-16, values of R^2 estimated by Ferragina et al. (2015) on the entire data set for C10:0, C16:0, and C18:0 were 0.48, 0.44, and 0.27, respectively. However, the performance in R^2 in training when using a nonrestricted PLSR seems to be more promising. For example, R^2 for the same traits was 75, 71, and 51% in Rutten et al. (2009). In addition, in agreement with results from the current study, R^2 values for these traits were 0.93, 0.88, and 0.75 in Ferrand-Calmels et al. (2014), who compared penalization methods, such as LASSO and elastic nets, with PLSR-M for the prediction of FA contents and percentages, concluding that the best results were obtained with PLSR-M.

A possible explanation for the inconsistent results might be the use of the PLSR-M instead of PLSR and MPLSR with a maximum number of 16 terms, as used in the previous study. Some of the traits investigated by Ferragina et al. (2015), namely C10:0, C16:0, and

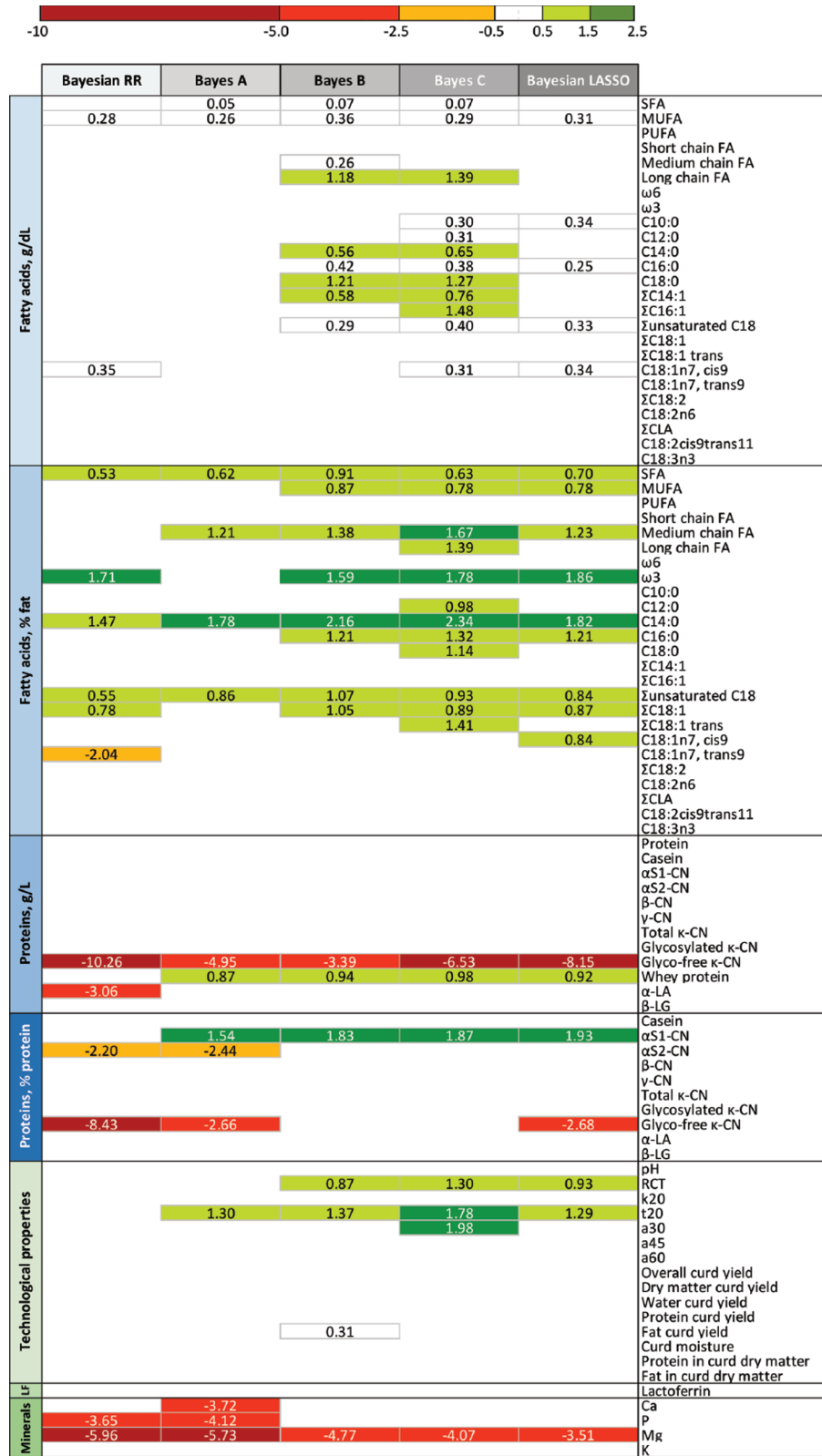


Figure 2. Difference (percentage points) between the coefficient of determination of validation obtained by Bayesian models and by a partial least squares regression in which the number of terms was chosen to minimize the prediction error for the prediction of all the investigated traits. RR = ridge regression; LASSO = least absolute shrinkage and selection operator. Only significant differences ($P < 0.05$) are shown. Color version available online.

C18:0 percentage in total fat as well as rennet coagulation time, were also among our traits, and values of R^2_v obtained for such traits with PLSR-16 were remarkably lower than those estimated using PLSR-M. The R^2_v for C10:0, C16:0, and C18:0 percentage in total fat was 25, 28, and 16 percentage points lower, respectively, than that obtained by PLSR-M and 14 percentage points lower than the R^2_v yielded by PLSR-M for rennet coagulation time (results not reported in tables). Thus, it can be hypothesized that the settings under which the PLSR and MPLSR were performed in Ferragina et al. (2015) might have penalized the accuracy of those methods in comparison with the Bayesian models.

Accuracy of Bayesian Models Compared with PLSR for the Prediction of Protein Composition

Consistent with results for FA, Bayesian models led, on average, to 7% lower R^2_t and 4% greater RMSEP_t than PLSR-M in the prediction of protein composition. Comparisons between R^2_v values obtained with the 6 different models for the prediction of protein fraction contents and percentages are reported in Figure 2. In validation, Bayesian models offered very limited or no benefits when compared with PLSR-M. For some traits, the R^2_v yielded by Bayesian methods was significantly lower than that yielded by PLSR-M. This occurred for the content of unglycosylated κ -CN (from -3.4 percentage points with BB to -10.3 percentage points with BRR) and for its percentage in total protein (from -2.7 percentage points with BA and BL to -8.4 percentage points with BRR). In addition, BRR showed a lower accuracy than PLSR-M when predicting the content of α -LA and α_{S2} -CN percentage (-3.1 and -2.2 percentage points, respectively). Conversely, BA, BB, BC, and BL yielded a significantly higher R^2_v for the content of total whey protein and α_{S1} -CN percentage, which were the only protein fractions predicted more accurately by Bayesian methods than by PLSR-M. For the content and percentage of unglycosylated κ -CN, RMSEP_v was significantly higher than that of PLSR-M when using BRR (+4 and +5%, respectively). When using BB and BC, the same parameter was 1% lower than that of PLSR-M in the prediction of whey protein. For α_{S1} -CN percentage, the decrease in RMSEP_v due to the use of Bayesian models was not significant. These results indicate that Bayesian models offer a small advantage over PLSR-M for the prediction of protein composition, particularly when considering that the improvements in the prediction accuracy for the content of total whey protein and α_{S1} -CN percentage were lower than 2 percentage points of R^2_v .

Accuracy of Bayesian Models Compared with PLSR for the Prediction of Technological Properties, LF, and Minerals

For the prediction of milk coagulation properties, curd yield and composition, LF, and minerals, R^2_t yielded by Bayesian models was 3 to 4 percentage points lower than that obtained with PLSR-M, with a corresponding 4% higher RMSEP_t. Differences between the R^2_v obtained with the 6 models for the prediction of such traits are reported in Figure 2. For coagulation properties, BB, BC, and BL gave slightly better results when predicting rennet coagulation time, time from rennet addition to curd firming, and curd firmness after 30 min from rennet addition. Using BC, the RMSEP_v decreased from 2 to 3% in comparison with PLSR-M. However, the maximum difference in R^2_v between the PLSR-M and Bayesian methods was lower than 2 percentage points. No significant difference was detected in the prediction accuracy for the traits related to curd yield and composition when using either the PLSR-M or Bayesian models, with the exception of a slightly higher R^2_v (0.3 percentage points) and a slightly smaller RMSEP_v (-0.63%) provided by BB for fat curd yield.

In contrast with our results, Ferragina et al. (2015) reported a remarkable difference (from 6 to 14 percentage points) in the R^2 yielded by BB and MPLSR for rennet coagulation time, cheese yield, protein recovery, and fat recovery. However, like for FA, values were generally lower than other literature estimates obtained for the same traits and breed using PLSR-M (Cecchinato et al., 2009; De Marchi et al., 2009). As hypothesized for FA percentage, the conditions under which PLSR and MPLSR were fitted, together with the use of the whole and raw spectra, might have reduced the accuracy of those models and enhanced the differences between those methods and Bayesian models.

The prediction accuracy of LF was not influenced by the model used, whereas it decreased for Ca, P, and Mg when Bayesian models were used. In particular, for Mg, the R^2_v was 3.5 to 6 percentage points lower and the RMSEP_v was 4% higher than that yielded by PLSR-M. Hence, PLSR should be preferred over Bayesian methods in the development of prediction equations for mineral contents.

Differences in Prediction Ability Among Bayesian Models

Significant differences in the predictive ability of Bayesian models were detected. However, they were limited and, for R^2_v , generally lower than 0.5 percentage point (Appendix Tables A1, A2, A3, and A4). Our

results indicate that the effect of the noise spectral regions on the accuracy of prediction equations depends on the statistical method used, with BRR, BA, and BL being the most sensitive (Table 2). As the major noisy spectral regions were excluded from the analysis, all Bayesian models provided very consistent results.

Ferragina et al. (2015) performed model comparison on the whole spectra, including the noise regions, and reported statistically significant and quantitatively remarkable differences among Bayesian models. According to their results, BB exhibited the best performances, as it was highly selective in identifying the most informative wavelengths in the 1,060 spectral variables. Conversely, BRR assigned small effects to almost all wavelengths, even within the regions affected by water absorbance, and PLSR estimated greater effects for those regions. This explained the better performance provided by BB compared with the other methods as well as the intermediate results obtained by BRR.

It is worth noting that water absorption regions are not the only noisy or noninformative spectral regions that might potentially affect model predictive ability. In preliminary analyses, we excluded the entire region ranging from 3,000 to 5,000 cm^{-1} (Eskildsen et al., 2014; Grelet et al., 2015), but this did not affect the performance of PLSR and MPLSR models. However, the presence of other noninformative wavelengths might affect the predictive accuracy of PLSR and MPLSR while having less or no effect on the accuracy of BB and BC models. Spectra mathematical treatments (e.g., derivatives, smoothing) can also have a potential effect on model accuracy because they influence the amount of noise. For this reason, the use of raw or pretreated spectra might influence the accuracy of prediction models, depending on the method used to build the equations. The application of a first derivative mathematic treatment to spectra further decreased differences in accuracy between PLSR-M and Bayesian models: across traits, BB and BC obtained the highest prediction accuracy within Bayesian models, but produced significantly better results than PLSR for only 7 traits over 92, with a maximum difference in R^2_v of 1.32% (results not presented in tables).

CONCLUSIONS

For the first time, BC and LASSO were tested as chemometric tools for the development of infrared prediction equations and BRR, BA, and BB were tested for the prediction of contents of FA, protein profile, LF, and minerals in milk. Overall, prediction accuracies yielded by BC were slightly higher than those provided by the other methods, in particular for the prediction of FA profile. Despite being significant, most of the

differences in prediction accuracy between models were small or negligible; the only exceptions were the prediction of Mg and unglycosylated κ -CN contents, for which Bayesian methods had worse predictive abilities compared with PLSR. Differences between regression methods were ascribed mainly to their ability of coping with spectral noise. The presence of noninformative wavelengths and the use of raw spectra affect predictive accuracy of PLSR while having less or no effect on the accuracy of BB and BC models. Hence, spectra variable selection and mathematical pretreatments might influence the accuracy of prediction models depending on the method used to build the equations. Differences in prediction accuracy between PLSR and Bayesian regression models were null or negligible after reducing noise by spectral mathematical treatments.

ACKNOWLEDGMENTS

The staff of the Friuli Venezia Giulia Milk Recording Agency (AAFVG, Codroipo, Italy) is gratefully acknowledged for milk sample collection and spectra acquisition. We warmly acknowledge Walter Manzon (AAFVG) for his help with spectra collection and database creation. Financial support for this study was provided by the University of Padova (Padova, Italy; Progetto di Ateneo 2012, CPDA122982).

REFERENCES

- Bonfatti, V., L. Degano, A. Menegoz, and P. Carnier. 2016. Short communication: Mid-infrared spectroscopy prediction of fine milk composition and technological properties in Italian Simmental. *J. Dairy Sci.* 99:8216–8221.
- Bonfatti, V., G. Di Martino, and P. Carnier. 2011. Effectiveness of mid-infrared spectroscopy for the prediction of detailed protein composition and contents of protein genetic variants of individual milk of Simmental cows. *J. Dairy Sci.* 94:5776–5785.
- Bonfatti, V., D. Vicario, A. Lugo, and P. Carnier. 2017. Genetic parameters of measures and population-wide infrared predictions of 92 traits describing the fine composition and technological properties of milk in Italian Simmental cattle. *J. Dairy Sci.* 100:5526–5540. <https://doi.org/10.3168/jds.2016-11667>.
- Burns, D. A., and E. W. Ciurczak. 2001. *Handbook of Near-Infrared Analysis*. 2nd ed. Marcel Dekker, New York, NY.
- Cecchinato, A., M. De Marchi, L. Gallo, G. Bittante, and P. Carnier. 2009. Mid-infrared spectroscopy predictions as indicator traits in breeding programs for enhanced coagulation properties of milk. *J. Dairy Sci.* 92:5304–5313.
- de los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler, and M. P. L. Calus. 2013. Whole genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193:327–345.
- de los Campos, G., and P. Pérez-Rodríguez. 2014. BGLR: Bayesian Generalized Linear Regression. Version 1.0.3. Accessed Jun. 9, 2017. <https://cran.r-project.org/web/packages/BGLR/index.html>.
- De Marchi, M., C. C. Fagan, C. P. O'Donnell, A. Cecchinato, R. Dal Zotto, M. Cassandro, M. Penasa, and G. Bittante. 2009. Prediction of coagulation properties, titratable acidity, and pH of bovine milk using mid-infrared spectroscopy. *J. Dairy Sci.* 92:423–432.
- Eskildsen, C. E., M. A. Rasmussen, S. B. Engelsen, L. B. Larsen, N. A. Poulsen, and T. Skov. 2014. Quantification of individual fatty

- acids in bovine milk by infrared spectroscopy and chemometrics: Understanding predictions of highly collinear reference variables. *J. Dairy Sci.* 97:7940–7951.
- Ferragina, A., G. de los Campos, A. I. Vazquez, A. Cecchinato, and G. Bittante. 2015. Bayesian regression models outperform partial least squares methods for predicting milk components and technological properties using infrared spectral data. *J. Dairy Sci.* 98:8133–8151.
- Ferrand-Calmels, M., I. Palière, M. Brochard, O. Leray, J. M. Astruc, M. R. Aurel, S. Barbey, F. Bouvier, P. Brunschwig, H. Cailat, M. Douguet, F. Faucon-Lahalle, M. Gelé, G. Thomas, J. M. Trommenschlager, and H. Larroque. 2014. Prediction of fatty acid profiles in cow, ewe, and goat milk by mid-infrared spectrometry. *J. Dairy Sci.* 97:17–35.
- Gengler, N., H. Soyeurt, F. Dehareng, C. Bastin, F. Colinet, H. Hammami, M.-L. Vanrobays, A. Lainé, S. Vanderick, C. Grelet, A. Vanlierde, E. Froidmont, and P. Dardenne. 2016. Capitalizing on fine milk composition for breeding and management of dairy cows. *J. Dairy Sci.* 99:4071–4079.
- Grelet, C., J. A. Fernández Pierna, P. Dardenne, V. Baeten, and F. Deharen. 2015. Standardization of milk mid-infrared spectra from a European dairy network. *J. Dairy Sci.* 98:2150–2160.
- Habier, D., R. Fernando, K. Kizilkaya, and D. Garrick. 2011. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186–197.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Mevik, B. H., and R. Wehrens. 2007. The PLS package: Principal component and partial least squares regression in R. *J. Stat. Softw.* 18:1–24.
- Park, T., and G. Casella. 2008. The Bayesian Lasso. *J. Am. Stat. Assoc.* 103:681–686.
- Pérez, P., and G. de los Campos. 2014. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198:483–495.
- R Development Core Team. 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rutten, M. J. M., H. Bovenhuis, K. A. Hetingga, H. J. F. van Valenberg, and J. A. M. van Arendonk. 2009. Predicting bovine milk fat composition using infrared spectroscopy based on milk samples collected in winter and summer. *J. Dairy Sci.* 92:6202–6209. <https://doi.org/10.3168/jds.2009-2456>.
- Rutten, M. J. M., H. Bovenhuis, and J. A. M. van Arendonk. 2010. The effect of the number of observations used for Fourier transform infrared model calibration for bovine milk fat composition on the estimated genetic parameters of the predicted data. *J. Dairy Sci.* 93:4872–4882.
- Shenk, J. S., and M. O. Westerhaus. 1991. Population definition, sample selection and calibration procedures for near infrared reflectance spectroscopy. *Crop Sci.* 31:469–474.
- Soyeurt, H., F. Dehareng, N. Gengler, S. McParland, E. Wall, D. P. Berry, M. Coffey, and P. Dardenne. 2011. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. *J. Dairy Sci.* 94:1657–1667.
- Soyeurt, H., I. Misztal, and N. Gengler. 2010. Genetic variability of milk components based on mid-infrared spectral data. *J. Dairy Sci.* 93:1722–1728.

APPENDIX

Table A1. Mean \pm SD of the coefficient of determination in validation (R^2_v), obtained by partial least squares regression (PLSR) in which the number of terms was chosen to minimize the prediction error (PLSR-M), and mean \pm SD of the difference between the R^2_v yielded by Bayesian models and by PLSR-M for prediction equations developed for fatty acid (FA) contents (g/dL of milk)

Trait ¹	Difference between Bayesian models and PLSR-M ²						S^3
	R^2_v PLSR-M	Bayes RR	Bayes A	Bayes B	Bayes C	LASSO	
SFA	98.9 \pm 0.3 ^a	0.0 \pm 0.1 ^{ab}	0.1 \pm 0.1 ^b	0.1 \pm 0.1 ^c	0.1 \pm 0.1 ^c	0.0 \pm 0.1 ^a	+
MUFA	94.3 \pm 1.5 ^a	0.3 \pm 0.3 ^b	0.3 \pm 0.3 ^b	0.4 \pm 0.3 ^b	0.3 \pm 0.4 ^b	0.3 \pm 0.2 ^b	+
PUFA	78.8 \pm 3.7	-0.2 \pm 1.4	-0.4 \pm 1.8	-0.3 \pm 1.6	0.0 \pm 1.1	-0.1 \pm 1.2	
Short-chain FA	91.7 \pm 3.5 ^{ab}	0.0 \pm 0.5 ^a	0.0 \pm 0.6 ^{ab}	0.1 \pm 0.5 ^{ab}	0.1 \pm 0.4 ^{ab}	0.1 \pm 0.5 ^b	
Medium-chain FA	96.3 \pm 0.8 ^{ab}	0.0 \pm 0.3 ^a	0.2 \pm 0.3 ^b	0.3 \pm 0.3 ^c	0.2 \pm 0.3 ^{abc}	0.1 \pm 0.2 ^b	+
Long-chain FA	81.1 \pm 3.3 ^{ab}	0.3 \pm 1.6 ^a	0.6 \pm 1.5 ^b	1.2 \pm 1.6 ^c	1.4 \pm 1.5 ^c	0.7 \pm 1.4 ^b	+
n-6	74.3 \pm 5.2 ^{ab}	-0.3 \pm 1.7 ^b	-0.6 \pm 2.1 ^a	-0.4 \pm 2.0 ^{ab}	-0.2 \pm 1.8 ^{ab}	-0.2 \pm 1.7 ^{ab}	
n-3	75.9 \pm 3.9	0.0 \pm 2.1	-0.7 \pm 2.7	-0.5 \pm 2.4	-0.6 \pm 1.9	-0.1 \pm 2.1	
C10:0	90.8 \pm 2.1 ^a	0.2 \pm 0.4 ^{ab}	0.2 \pm 0.5 ^{ab}	0.3 \pm 0.4 ^{ab}	0.3 \pm 0.3 ^{bc}	0.3 \pm 0.4 ^c	+
C12:0	92.1 \pm 1.5 ^a	0.2 \pm 0.3 ^{ab}	0.2 \pm 0.6 ^{ab}	0.3 \pm 0.5 ^{ab}	0.3 \pm 0.3 ^b	0.2 \pm 0.3 ^{ab}	+
C14:0	91.4 \pm 3.2 ^a	0.3 \pm 0.6 ^a	0.4 \pm 0.6 ^{ab}	0.6 \pm 0.6 ^c	0.6 \pm 0.8 ^c	0.4 \pm 0.5 ^{bc}	+
C16:0	93.7 \pm 1.4 ^a	0.1 \pm 0.3 ^a	0.2 \pm 0.4 ^{ab}	0.4 \pm 0.4 ^c	0.4 \pm 0.5 ^{bc}	0.3 \pm 0.3 ^{bc}	+
C18:0	80.8 \pm 2.9 ^{ab}	0.2 \pm 1.7 ^a	0.7 \pm 1.5 ^b	1.2 \pm 1.6 ^c	1.3 \pm 1.4 ^c	0.7 \pm 1.6 ^b	+
Σ C14:1	64.4 \pm 5.1 ^{ab}	-0.2 \pm 0.8 ^a	0.2 \pm 0.6 ^b	0.6 \pm 0.7 ^c	0.8 \pm 0.6 ^c	0.3 \pm 0.6 ^b	+
Σ C16:1	74.3 \pm 5.2 ^{ab}	0.4 \pm 2.2 ^a	0.6 \pm 2.3 ^{ab}	0.8 \pm 2.3 ^b	1.5 \pm 2.0 ^c	0.7 \pm 2.2 ^b	+
Σ unsaturated C18	92.6 \pm 2.3 ^a	0.2 \pm 0.2 ^b	0.2 \pm 0.3 ^b	0.3 \pm 0.4 ^c	0.4 \pm 0.3 ^c	0.3 \pm 0.3 ^{bc}	+
Σ C18:1	92.6 \pm 1.7	0.1 \pm 0.5	0.1 \pm 0.5	0.1 \pm 0.5	0.2 \pm 0.4	0.1 \pm 0.5	
Σ C18:1 <i>trans</i>	68.6 \pm 5.3 ^{abc}	-0.5 \pm 2.8 ^a	-0.4 \pm 3.4 ^{ab}	0.4 \pm 2.7 ^{bc}	0.8 \pm 2.1 ^{bc}	0.5 \pm 2.7 ^c	
C18:1n-7, <i>cis</i> -9	82.5 \pm 4.1 ^a	0.4 \pm 0.3 ^b	0.1 \pm 0.6 ^a	0.2 \pm 0.6 ^a	0.3 \pm 0.4 ^b	0.3 \pm 0.4 ^b	+
C18:1n-7 <i>trans</i> -9	69.5 \pm 4.2 ^{abc}	-0.7 \pm 1.7 ^{ab}	-0.8 \pm 2.5 ^a	-0.1 \pm 2.6 ^{bc}	0.8 \pm 1.3 ^c	0.4 \pm 1.7 ^c	
Σ C18:2	78.5 \pm 3.1 ^{ab}	-0.2 \pm 1.3 ^{ab}	-0.6 \pm 1.3 ^a	-0.2 \pm 1.0 ^b	-0.2 \pm 0.8 ^{ab}	-0.2 \pm 0.9 ^{ab}	
C18:2n-6	75.2 \pm 3.2 ^{abc}	-0.6 \pm 1.5 ^{ab}	-1.2 \pm 2.1 ^a	-0.7 \pm 1.9 ^{bc}	-0.4 \pm 1.3 ^c	-0.4 \pm 1.7 ^c	
Σ CLA	61.7 \pm 4.4 ^{ab}	-1.6 \pm 4.9 ^a	-1.4 \pm 5.2 ^a	-1.0 \pm 5.2 ^{ab}	-0.5 \pm 4.1 ^b	-0.6 \pm 4.3 ^b	
C18:2 <i>cis</i> -9, <i>trans</i> -11	65.3 \pm 6.5 ^{ab}	-1.3 \pm 2.3 ^a	-1.5 \pm 2.3 ^a	-0.5 \pm 2.0 ^b	-0.1 \pm 1.9 ^b	-0.5 \pm 2.0 ^b	
C18:3n-3	70.7 \pm 4.1 ^{ab}	-0.4 \pm 1.7 ^{ab}	-0.8 \pm 2.1 ^a	-0.3 \pm 2.2 ^b	-0.2 \pm 1.9 ^b	-0.3 \pm 1.7 ^{ab}	

^{a-c}Different superscripts indicate significantly different ($P < 0.05$) parameter estimates.

¹Short-chain FA = FA from C4 to C10; medium-chain FA = FA from C12 to C16; long-chain FA = FA from C18 to C24; Σ = sum of FA.

²RR = ridge regression; LASSO = least absolute shrinkage and selection operator.

³ S = presence of at least 1 significant difference ($P < 0.05$) between Bayesian models and PLSR.

Table A2. Mean \pm SD of the coefficient of determination in validation (R^2_v), obtained by partial least squares regression (PLSR) in which the number of terms was chosen to minimize the prediction error (PLSR-M), and mean \pm SD of the difference between the R^2_v yielded by Bayesian models and by PLSR-M for prediction equations developed for the prediction of fatty acid (FA) contents (g/100 g of fat)

Trait ¹	R^2_v PLSR-M	Difference between Bayesian models and PLSR-M ²					S^3
		Bayes RR	Bayes A	Bayes B	Bayes C	LASSO	
SFA	83.6 \pm 4.0 ^a	0.5 \pm 0.6 ^{bc}	0.6 \pm 0.7 ^b	0.9 \pm 0.8 ^c	0.6 \pm 0.7 ^c	0.7 \pm 0.7 ^{bc}	+
MUFA	82.6 \pm 3.3 ^a	0.4 \pm 0.7 ^{ab}	0.6 \pm 0.9 ^{ab}	0.9 \pm 0.9 ^c	0.8 \pm 1.0 ^c	0.8 \pm 0.7 ^{bc}	+
PUFA	69.4 \pm 6.4 ^{ab}	-0.7 \pm 1.6 ^{ab}	-0.9 \pm 1.6 ^a	-0.6 \pm 1.3 ^b	0.0 \pm 2.0 ^b	-0.5 \pm 1.5 ^{ab}	
Short-chain FA	75.4 \pm 5.4	-0.1 \pm 1.6	-0.2 \pm 2.0	0.1 \pm 1.5	0.4 \pm 1.9	-0.1 \pm 1.5	
Medium-chain FA	76.6 \pm 4.2 ^a	0.6 \pm 1.0 ^{ab}	1.2 \pm 1.4 ^{bc}	1.4 \pm 1.3 ^c	1.7 \pm 1.2 ^c	1.2 \pm 1.1 ^c	+
Long-chain FA	74.4 \pm 4.5 ^{abc}	0.0 \pm 1.5 ^a	0.3 \pm 1.5 ^b	0.7 \pm 1.5 ^c	1.4 \pm 1.2 ^d	0.1 \pm 1.5 ^{abc}	+
n-6	71.6 \pm 4.5 ^{ab}	-0.1 \pm 2.6 ^b	-0.9 \pm 2.5 ^a	-0.3 \pm 2.7 ^b	0.1 \pm 2.4 ^b	0.0 \pm 2.3 ^b	
n-3	65.8 \pm 8.2 ^a	1.7 \pm 2.0 ^b	1.3 \pm 2.2 ^a	1.6 \pm 2.1 ^b	1.8 \pm 1.9 ^b	1.9 \pm 1.9 ^b	+
C10:0	77.8 \pm 6.0 ^{ab}	-0.1 \pm 0.9 ^a	0.3 \pm 1.2 ^{ab}	0.3 \pm 1.3 ^{ab}	0.5 \pm 1.1 ^{ab}	0.4 \pm 1.1 ^b	
C12:0	81.3 \pm 4.4 ^a	0.5 \pm 0.9 ^{ab}	0.4 \pm 1.3 ^{ab}	0.7 \pm 1.3 ^{ab}	1.0 \pm 1.0 ^b	0.7 \pm 1.0 ^{ab}	+
C14:0	68.0 \pm 9.9 ^a	1.5 \pm 2.0 ^b	1.8 \pm 2.3 ^{bc}	2.2 \pm 2.4 ^c	2.3 \pm 2.1 ^c	1.8 \pm 1.8 ^c	+
C16:0	70.3 \pm 4.3 ^a	0.5 \pm 0.9 ^{ab}	0.9 \pm 1.8 ^{abc}	1.2 \pm 1.7 ^{bc}	1.3 \pm 1.5 ^c	1.2 \pm 1.6 ^{bc}	+
C18:0	74.5 \pm 4.6 ^{abc}	0.1 \pm 1.5 ^a	0.4 \pm 1.6 ^{ab}	1.0 \pm 1.5 ^c	1.1 \pm 1.4 ^d	0.5 \pm 1.5 ^b	+
Σ C14:1	50.6 \pm 6.9 ^{ab}	0.1 \pm 2.1 ^a	0.4 \pm 1.8 ^a	0.9 \pm 1.8 ^b	1.2 \pm 1.7 ^b	0.5 \pm 1.8 ^a	
Σ C16:1	54.5 \pm 5.8 ^{ab}	0.7 \pm 2.6 ^a	0.7 \pm 3.2 ^{ab}	0.8 \pm 2.9 ^{ab}	1.4 \pm 2.6 ^{ab}	1.2 \pm 2.4 ^b	
Σ unsaturated C18	82.7 \pm 4.9 ^a	0.6 \pm 0.6 ^b	0.9 \pm 0.9 ^{bc}	1.1 \pm 0.9 ^d	0.9 \pm 0.8 ^{bc}	0.8 \pm 0.7 ^{cd}	+
Σ C18:1	82.1 \pm 5.9 ^a	0.8 \pm 0.9 ^b	0.7 \pm 1.0 ^{ab}	1.1 \pm 1.3 ^b	0.9 \pm 1.0 ^b	0.9 \pm 0.9 ^b	+
Σ C18:1 <i>trans</i>	56.2 \pm 5.6 ^{ab}	-1.6 \pm 2.5 ^a	-1.6 \pm 2.9 ^a	-0.9 \pm 2.5 ^a	1.4 \pm 1.9 ^c	0.2 \pm 2.0 ^{bc}	+
C18:1n-7, <i>cis</i> -9	91.8 \pm 2.6 ^a	0.7 \pm 1.1 ^{ab}	0.4 \pm 1.2 ^{ab}	0.6 \pm 1.2 ^{ab}	0.7 \pm 1.0 ^{ab}	0.8 \pm 1.0 ^b	+
C18:1n-7, <i>trans</i> -9	56.1 \pm 6.0 ^{bc}	-2.0 \pm 2.7 ^a	-1.9 \pm 3.3 ^{ab}	-1.1 \pm 3.3 ^{abc}	0.1 \pm 2.0 ^c	-0.5 \pm 2.2 ^c	+
Σ C18:2	69.3 \pm 3.8	0.1 \pm 2.8	0.3 \pm 3.5	0.5 \pm 3.1	0.4 \pm 2.6	0.5 \pm 2.7	
C18:2n-6	72.0 \pm 3.9 ^{abc}	-0.3 \pm 2.2 ^{bc}	-1.3 \pm 2.4 ^a	-0.7 \pm 2.2 ^b	0.2 \pm 2.0 ^c	-0.4 \pm 2.2 ^{bc}	
Σ CLA	43.2 \pm 7.4 ^{abcd}	-2.0 \pm 5.6 ^b	-2.5 \pm 5.9 ^a	-1.4 \pm 5.4 ^{bc}	-0.7 \pm 5.0 ^d	-1.5 \pm 5.3 ^c	
C18:2 <i>cis</i> -9, <i>trans</i> -11	53.8 \pm 5.3 ^{abc}	-1.8 \pm 4.6 ^b	-2.6 \pm 5.2 ^a	-1.6 \pm 4.8 ^b	-0.8 \pm 4.4 ^c	-0.9 \pm 4.2 ^c	
C18:3n-3	65.4 \pm 5.9 ^{ab}	0.0 \pm 2.4 ^b	-0.5 \pm 2.6 ^a	-0.1 \pm 2.5 ^{ab}	0.3 \pm 2.4 ^b	0.2 \pm 2.2 ^b	

^{a-d}Different superscripts indicate significantly different ($P < 0.05$) parameter estimates.

¹Short-chain FA = FA from C4 to C10; medium-chain FA = FA from C12 to C16; long-chain FA = FA from C18 to C24; Σ = sum of FA.

²RR = ridge regression; LASSO = least absolute shrinkage and selection operator.

³ S = presence of at least 1 significant difference ($P < 0.05$) between Bayesian models and PLSR.

Table A3. Mean \pm SD of the coefficient of determination in validation (R^2_v), obtained by partial least squares regression (PLSR) in which the number of terms was chosen to minimize the prediction error (PLSR-M), and mean \pm SD of the difference between the R^2_v yielded by Bayesian models and by PLSR-M for prediction equations developed for the prediction of protein fractions (g/L of milk and g/100 g of protein)

Trait ¹	R^2_v PLSR-M	Difference between Bayesian models and PLSR-M ²					S^3
		Bayes RR	Bayes A	Bayes B	Bayes C	LASSO	
Protein	80.0 \pm 8.9 ^{abc}	-0.0 \pm 1.2 ^a	0.3 \pm 1.0 ^{bc}	0.3 \pm 1.1 ^{bc}	0.2 \pm 1.1 ^b	0.2 \pm 1.1 ^c	
Protein fractions, g/L							
Casein	81.7 \pm 6.7 ^{abc}	-0.0 \pm 1.3 ^a	0.2 \pm 1.2 ^b	0.3 \pm 1.2 ^c	0.1 \pm 1.2 ^b	0.1 \pm 1.2 ^b	
α_{S1} -CN	74.8 \pm 5.3 ^{ab}	0.3 \pm 1.1 ^a	0.5 \pm 1.0 ^b	0.6 \pm 1.1 ^b	0.5 \pm 1.0 ^b	0.4 \pm 1.1 ^{ab}	
α_{S2} -CN	54.2 \pm 6.7 ^{abc}	-0.9 \pm 3.8 ^a	-0.7 \pm 3.6 ^a	-0.2 \pm 3.4 ^{bc}	0.0 \pm 3.4 ^c	-0.3 \pm 3.5 ^b	
β -CN	60.1 \pm 6.8 ^{abc}	0.3 \pm 2.5 ^a	0.7 \pm 2.3 ^b	1.2 \pm 2.1 ^c	1.0 \pm 2.4 ^c	0.8 \pm 2.2 ^{ab}	
γ -CN	35.8 \pm 5.8 ^{abc}	-2.6 \pm 4.6 ^a	-1.5 \pm 3.3 ^b	-0.9 \pm 3.3 ^c	-0.7 \pm 3.6 ^c	-1.6 \pm 3.7 ^{ab}	
Total κ -CN	43.1 \pm 5.2 ^{abc}	-0.9 \pm 1.8 ^{ab}	-1.0 \pm 2.1 ^a	-0.5 \pm 1.9 ^b	0.2 \pm 1.6 ^c	-0.6 \pm 2.0 ^{ab}	
Glycosylated κ -CN	52.5 \pm 6.3 ^{abc}	-0.1 \pm 3.0 ^b	-0.7 \pm 3.4 ^a	-0.4 \pm 3.1 ^{ab}	0.8 \pm 3.0 ^c	-0.5 \pm 4.7 ^{ab}	
Glyco-free κ -CN	26.4 \pm 7.7 ^a	-10.3 \pm 7.6 ^b	-5.0 \pm 7.7 ^c	-3.4 \pm 6.6 ^d	-6.5 \pm 8.5 ^{bc}	-8.2 \pm 9.6 ^{bc}	+
Whey protein	54.2 \pm 7.7 ^a	0.7 \pm 1.4 ^{abc}	0.9 \pm 1.2 ^b	0.9 \pm 1.1 ^{bc}	1.0 \pm 1.2 ^c	0.9 \pm 1.2 ^{bc}	+
α -LA	27.4 \pm 5.4 ^{bed}	-3.1 \pm 4.2 ^a	-1.1 \pm 3.4 ^d	-0.9 \pm 3.1 ^d	-3.1 \pm 4.4 ^{ab}	-1.8 \pm 3.8 ^c	+
β -LG	49.2 \pm 8.6 ^{ab}	0.2 \pm 1.2 ^a	0.4 \pm 1.0 ^{ab}	0.5 \pm 1.0 ^{ab}	0.3 \pm 1.3 ^b	0.4 \pm 1.1 ^b	
Protein fractions, %							
Casein	54.2 \pm 6.2 ^{ab}	0.1 \pm 1.1 ^a	0.2 \pm 1.0 ^{ab}	0.2 \pm 1.1 ^{ab}	0.3 \pm 1.1 ^b	0.2 \pm 1.1 ^b	
α_{S1} -CN	27.4 \pm 9.9 ^a	1.4 \pm 2.1 ^{abc}	1.5 \pm 2.0 ^b	1.8 \pm 2.1 ^c	1.9 \pm 2.2 ^c	1.9 \pm 2.2 ^{bc}	+
α_{S2} -CN	31.1 \pm 6.4 ^{bc}	-2.2 \pm 2.8 ^a	-2.4 \pm 2.7 ^a	-1.4 \pm 2.5 ^b	-0.0 \pm 2.3 ^c	-1.3 \pm 2.4 ^b	+
β -CN	44.8 \pm 5.5 ^{abc}	-0.3 \pm 1.7 ^a	0.2 \pm 1.4 ^b	0.6 \pm 1.5 ^c	0.7 \pm 1.3 ^c	0.3 \pm 1.5 ^b	
γ -CN	33.7 \pm 5.5 ^{ab}	-1.3 \pm 3.1 ^a	-0.9 \pm 2.8 ^a	-0.4 \pm 2.9 ^b	-0.2 \pm 2.8 ^b	-0.9 \pm 2.7 ^{ab}	
Total κ -CN	28.3 \pm 4.5 ^{abc}	-0.3 \pm 2.6 ^{ab}	-1.0 \pm 3.4 ^a	-0.2 \pm 3.0 ^b	0.8 \pm 2.4 ^c	-0.5 \pm 2.6 ^{ab}	
Glycosylated κ -CN	39.2 \pm 9.0 ^{abc}	-0.3 \pm 2.5 ^{bc}	-1.3 \pm 2.9 ^a	-0.8 \pm 2.7 ^b	0.4 \pm 2.1 ^c	-1.4 \pm 5.5 ^{abc}	
Glyco-free κ -CN	24.1 \pm 6.6 ^c	-8.4 \pm 7.2 ^a	-2.7 \pm 3.4 ^b	-2.1 \pm 3.6 ^c	-2.2 \pm 6.4 ^{bc}	-2.7 \pm 3.1 ^b	+
α -LA	30.1 \pm 8.2 ^{ab}	-1.4 \pm 3.2 ^a	-0.6 \pm 2.6 ^b	-0.5 \pm 2.5 ^b	-1.3 \pm 3.1 ^a	-0.7 \pm 2.8 ^b	
β -LG	41.5 \pm 7.1 ^{ab}	0.1 \pm 1.5 ^{ab}	0.1 \pm 1.4 ^{ab}	0.2 \pm 1.4 ^b	0.1 \pm 1.5 ^a	0.2 \pm 1.2 ^{ab}	

^{a-d}Different superscripts indicate significantly different ($P < 0.05$) parameter estimates.

¹Protein = casein + whey protein; casein = α_{S1} -CN + α_{S2} -CN + β -CN + γ -CN + κ -CN; κ -CN = glycosylated κ -CN + unglycosylated κ -CN; whey protein = α -LA + β -LG.

²RR = ridge regression; LASSO = least absolute shrinkage and selection operator.

³ S = presence of at least 1 significant difference ($P < 0.05$) between Bayesian models and PLSR.

Table A4. Mean \pm SD of the coefficient of determination in validation (R^2v), obtained by partial least squares regression (PLSR) in which the number of terms was chosen to minimize the prediction error (PLSR-M), and mean \pm SD of the difference between the R^2v yielded by Bayesian models and by PLSR-M for prediction equations developed for the prediction of technological traits, lactoferrin, and mineral contents

Trait ¹	R^2v PLSR-M	Difference between Bayesian models and PLSR-M ²					S^3
		Bayes RR	Bayes A	Bayes B	Bayes C	LASSO	
pH	81.4 \pm 5.4 ^{ab}	0.3 \pm 1.1 ^{ab}	0.1 \pm 1.2 ^a	0.2 \pm 1.2 ^a	0.3 \pm 1.1 ^{ab}	0.3 \pm 1.2 ^b	
Coagulation properties							
RCT, min	73.0 \pm 4.2 ^a	0.7 \pm 1.2 ^a	0.8 \pm 1.2 ^{ab}	0.9 \pm 1.1 ^{bc}	1.3 \pm 1.0 ^c	0.9 \pm 1.1 ^c	+
k20, min	46.1 \pm 6.1 ^{ab}	0.7 \pm 2.1 ^a	0.5 \pm 2.3 ^a	0.8 \pm 2.4 ^{ab}	1.2 \pm 2.0 ^b	0.6 \pm 2.2 ^{ab}	
t20, min	70.5 \pm 3.5 ^a	1.1 \pm 1.3 ^b	1.3 \pm 1.4 ^{bc}	1.4 \pm 1.5 ^{bc}	1.8 \pm 1.4 ^c	1.3 \pm 1.2 ^{bc}	+
a30, mm	37.8 \pm 4.2 ^a	1.1 \pm 2.2 ^a	0.9 \pm 2.6 ^a	1.2 \pm 2.4 ^a	2.0 \pm 2.1 ^b	1.2 \pm 2.4 ^{ab}	+
a45, mm	24.4 \pm 9.2 ^{abc}	-1.1 \pm 2.9 ^a	-0.4 \pm 2.6 ^{bc}	-0.2 \pm 2.5 ^c	-0.8 \pm 2.8 ^a	-0.7 \pm 2.9 ^{ab}	
a60, mm	19.1 \pm 5.3 ^{abc}	-0.7 \pm 1.3 ^a	0.4 \pm 1.2 ^c	0.4 \pm 1.1 ^c	-0.3 \pm 1.1 ^b	-0.4 \pm 2.1 ^{abc}	
Curd yield, g/100 g of milk							
Overall	64.0 \pm 6.9	0.4 \pm 1.1	0.4 \pm 1.1	0.5 \pm 1.0	0.4 \pm 1.2	0.5 \pm 1.0	
DM	78.2 \pm 8.6	0.1 \pm 0.9	0.2 \pm 0.8	0.2 \pm 0.9	0.2 \pm 0.9	0.2 \pm 0.9	
Water	62.2 \pm 6.4	0.3 \pm 1.6	0.3 \pm 1.4	0.2 \pm 1.5	0.3 \pm 1.4	0.1 \pm 1.6	
Protein	56.5 \pm 9.9 ^{abc}	-0.5 \pm 1.8 ^a	-0.0 \pm 1.5 ^c	-0.1 \pm 1.6 ^c	-0.5 \pm 1.8 ^a	-0.3 \pm 1.7 ^b	
Fat	70.6 \pm 6.4 ^a	0.2 \pm 0.4 ^a	0.3 \pm 0.4 ^a	0.3 \pm 0.4 ^b	0.2 \pm 0.4 ^a	0.2 \pm 0.4 ^a	+
Curd composition, %							
Moisture	59.1 \pm 10.2 ^{ab}	-0.3 \pm 1.6 ^a	0.0 \pm 1.4 ^b	-0.2 \pm 1.5 ^{ab}	-0.1 \pm 1.4 ^{ab}	-0.2 \pm 1.3 ^{ab}	
Protein in DM	42.6 \pm 10.9 ^{ab}	-0.8 \pm 3.2 ^a	-0.1 \pm 2.5 ^b	-0.0 \pm 2.6 ^b	-0.7 \pm 3.1 ^a	-0.9 \pm 3.2 ^a	
Fat in DM	41.5 \pm 10.0 ^{abc}	-0.2 \pm 1.8 ^a	0.3 \pm 1.7 ^{bc}	0.4 \pm 1.7 ^c	0.1 \pm 1.8 ^{bc}	0.1 \pm 1.7 ^b	
Lactoferrin, μ g/mL	44.4 \pm 11.8 ^{ab}	-3.4 \pm 4.9 ^a	-2.0 \pm 4.5 ^b	-1.9 \pm 4.5 ^b	-3.3 \pm 4.6 ^a	-3.1 \pm 4.6 ^a	
Minerals, mg/L							
Ca	56.2 \pm 8.8 ^{bc}	-2.7 \pm 4.2 ^b	-3.7 \pm 4.2 ^a	-3.0 \pm 4.3 ^b	-1.7 \pm 3.9 ^c	-1.8 \pm 4.0 ^c	+
P	47.3 \pm 8.6 ^c	-3.7 \pm 4.7 ^b	-4.1 \pm 5.1 ^a	-3.1 \pm 5.0 ^{bc}	-2.9 \pm 5.3 ^c	-2.6 \pm 5.4 ^c	+
Mg	53.0 \pm 4.3 ^d	-6.0 \pm 3.0 ^a	-5.7 \pm 3.6 ^{ab}	-4.8 \pm 3.8 ^b	-4.1 \pm 3.7 ^c	-3.5 \pm 3.2 ^c	+
K	46.3 \pm 5.1 ^{ab}	-3.4 \pm 5.5 ^{ab}	-3.7 \pm 5.8 ^a	-3.4 \pm 5.7 ^a	-3.4 \pm 6.1 ^{ab}	-2.9 \pm 5.7 ^b	

^{a-d}Different superscripts indicate significantly different ($P < 0.05$) parameter estimates.

¹RCT = rennet coagulation time; k20 = curd firming time; t20 = time from rennet addition to k20; a30 = curd firmness at 30 min from rennet addition; a45 = curd firmness at 45 min from rennet addition; a60 = curd firmness at 60 min from rennet addition.

²RR = ridge regression; LASSO = least absolute shrinkage and selection operator.

³ S = presence of at least 1 significant difference ($P < 0.05$) between Bayesian models and PLSR.