J. Dairy Sci. TBC
https://doi.org/10.3168/jds.2023-23843

Prediction of key milk biomarkers in dairy cows through milk MIR spectra and international collaborations

C. Grelet,[1] T. Larsen,[2] M. A. Crowe,[3] D. C. Wathes,[4] C. P. Ferris,[5] K. L. Ingvartsen,[2] C. Marchitelli,[6] F. Becker,[7] A. Vanlierde,[1] J. Leblois,[8] U. Schuler,[9] F. J. Auer,[10] A. Köck,[11] L. Dale,[12] J. Sölkner,[13] O. Christophe,[1] J. Hummel,[14] A. Mensching,[14]* J. A. Fernández Pierna,[1] H. Soyeurt,[15] M. Calmels,[16] R. Reding,[17] M. Gelé,[18] Y. Chen,[15] N. Gengler,[15] GplusE consortium,† and F. Dehareng[1]

[1] Walloon Agricultural Research Center (CRA-W), Gembloux, Belgium
[2] Dept Animal and Veterinary Sciences, Aarhus University, Tjele, Denmark
[3] University College Dublin (UCD), Dublin, Ireland
[4] Royal Veterinary College (RVC), London, United Kingdom
[5] Agri-Food and Biosciences Institute (AFBI), Belfast, Northern Ireland
[6] Research Center for Animal Production and Aquaculture (CREA), Roma, Italy
[7] Leibniz Institute for Farm Animal Biology (FBN), Dummerstorf, Germany
[8] EEIG European Milk Recording (EMR), Ciney, Belgium
[9] Qualitas, Zug, Switzerland
[10] LKV-Austria, Vienna, Austria,
[11] ZuchtData, Vienna, Austria
[12] LKV Baden Württemberg, Stuttgart, Germany
[13] University of Natural Resources and Life Sciences, Vienna, Austria,
[14] University of Göttingen, Göttingen, Germany
[15] University of Liège, Gembloux Agro-Bio Tech (Ulg-GxABT), Gembloux, Belgium
[16] Seenovia, Saint Berthevin, France
[17] Convis, Ettelbruck, Luxembourg
[18] Idele, Paris, France,

ABSTRACT

At the individual cow level, sub-optimum fertility, mastitis, negative energy balance and ketosis are major issues in dairy farming. These problems are widespread on dairy farms and have an important economic impact. The objectives of this study were: 1) to assess the potential of milk Mid Infrared (MIR) spectra to predict key biomarkers of energy deficit (citrate, isocitrate, glucose-6P, free glucose), ketosis (BHB and acetone), mastitis (NAGase and LDH), and fertility (progesterone); 2) to test alternative methodologies to partial least square regression (PLS) to better account for the specific asymmetric distribution of the biomarkers; and 3) to create robust models by merging large data sets from 5 international or national projects. Benefiting from this international collaboration, the data set comprised a total of 9,143 milk samples from 3,758 cows located in 589 herds across 10 countries and represented 7 breeds. The samples were analyzed by reference chemistry for biomarker contents while the MIR analyses were performed on 30 instruments from different models and brands, with spectra harmonized into a common format. Four quantitative methodologies were evaluated to address the strongly skewed distribution of some biomarkers. PLS was used as the reference basis, and compared with a random modification of distribution associated with PLS (Random-downsampling-PLS), an optimized modification of distribution associated with PLS (KennardStone-downsampling-PLS) and Support Vector Machine (SVM). When the ability of MIR to predict biomarkers was too low for quantification, different qualitative methodologies were tested to discriminate low vs high values of biomarkers. For each

Received June 6, 2023.
Accepted September 23, 2023.
* Present address: Vereinigte Informationssysteme Tierhaltung (VIT), Verden, Germany
† List of authors within the GplusU consortium: Niamh McLoughlin, Alan Fahey, Elizabeth Matthews, Andreia Santoro, Colin Byrne, Pauline Rudd, Roisin O'Flaherty, Sinead Hallinan, Claire Wathes, Zhangrui Cheng, Ali Fouladi, Mazdak Salavati, Geoff Pollott, Dirk Werling, Beatriz Sanz Bernardo, Alistair Wylie, Matt Bell, Miel Hostens, Mieke Vaneetvelde, Kristof Hermans, Geert Opsomer, Sander Moerman, Jenne Dekoster, Hannes Bogaert, Jan Vandepitte, Leila Vandevelde, Bonny Vanranst, Leslie Foldager, Martin T. Sorensen, Johanna Hoglund, Susanne Dahl, Soren Ostergaard, Janne Rothmann, Mogens Krogh, Else Meyer, Charlotte Gaillard, Jehan Ettema, Tine Rousing, Federica Signorelli, Francesco Napolitano, Bianca Moioli, Alessandra Crisà, Luca Buttazzoni, Jennifer McClure, Daragh Matthews, Francis Kearney, Andrew Cromie, Matt McClure, Shujun Zhang, Xing Chen, Huanchun Chen, Junlong Zhao, Liguo Yang, Guohua Hua, Chen Tan, Guiqiang Wang, Michel Bonneau, Andrea Pompozzi, Armin Pearn, Arnold Evertson, Linda Kosten, Anders Fogh, Thomas Andersen, Matthew Lucey, Chris Elsik, Gavin Conant, Jerry Taylor, Eric Froidmont, Nicolas Gengler, Michel Georges, Frédéric Colinet, Marilou Ramos Pamplona, Hedi Hammami, Catherine Bastin, Haruko Takeda, Aurelie Laine, Anne-Sophie Van Laere, Martin Schulze, Sergio Palma Vera.

biomarker, 20% of the herds were randomly removed within all countries to be used as the validation data set. The remaining 80% of herds were used as the calibration data set. In calibration, the 3 alternative methodologies outperform the PLS performances for the majority of biomarkers. However, in the external herd validation, PLS provided the best results for isocitrate, glucose-6P, free glucose and LDH (R^2v = 0.48, 0.58, 0.28, and 0.24). For other molecules, PLS-Random-downsampling and PLS-KennardStone-downsampling outperformed PLS in the majority of cases, but the best results were provided by SVM for citrate, BHB, acetone, NAGase and progesterone (R^2v = 0.94, 0.58, 0.76, 0.68, and 0.15). Hence, PLS and SVM based on the entire data set provided the best results for normal and skewed distributions, respectively. Complementary to the quantitative methods, the qualitative discriminant models enabled the discrimination of high and low values for BHB, acetone, and NAGase with a global accuracy around 90%, and glucose-6P with an accuracy of 83%. In conclusion, MIR spectra of milk can enable quantitative screening of citrate as a biomarker of energy deficit and discrimination of low and high values of BHB, acetone, and NAGase, as biomarkers of ketosis and mastitis. Finally, progesterone could not be predicted with sufficient accuracy from milk MIR spectra to be further considered. Consequently, MIR spectrometry can bring valuable information regarding the occurrence of energy deficit, ketosis and mastitis in dairy cows, which in turn have major influences on their fertility and survival.

Keywords: Fourier transform mid-infrared spectrometry, ketosis, negative energy balance, mastitis, fertility

INTRODUCTION

At the individual cow level, sub-optimum fertility, mastitis, negative energy balance, and associated metabolic diseases are major issues in dairy farming. These problems have widespread incidence and have an important economic impact. Energy deficit can be considered as a central issue as almost all dairy cows experience a period of negative energy balance after calving. This is associated with immunosuppression and depending of the intensity of this deficit, between 30% and 50% of dairy cows suffer from associated metabolic and infectious diseases (Leblanc, 2010; Wathes et al., 2021). Among the associated diseases, ketosis is particularly damaging. McArt et al. (2012) reported an average incidence of subclinical and clinical ketosis of 43%, while economic losses per cow may range from 188 to 347€ (McArt et al., 2015; Gohary et al., 2016). Mastitis is also a major challenge in dairy herds, with Riekerink et al. (2008) estimating the incidence of clinical mastitis at 23%, while the associated cost per cow per year are estimated between 356 and 716€ (Puerto et al., 2021). Finally, fertility problems such as delayed resumption of cyclicity, prolonged luteal phase, fertilization failure or failure to sustain pregnancy are also important issues affecting particularly high yielding dairy cow (Leroy et al., 2008) and is leading to increase of inter-calving interval, ratio of insemination to conception while decreasing rate to first insemination (Roche et al., 2011).

Having relevant and frequent 'indicators' of these issues at the individual cow level, through associated milk biomarkers, could support herd management by proposing an early detection of subclinical or clinical issues. Estimation of biomarkers in milk by Fourier transform mid-infrared spectrometry (**MIR**) could meet this role. Indeed, milk is available daily without any invasive process and MIR analysis is now common place and accessible in many dairy regions.

Regarding energy deficit, citrate in milk has been highlighted by Bjerre-Harpøth et al. (2012), among various biomarkers, as having the greatest response during a period of negative energy balance. Citrate plays a central role in cellular energy metabolism, being an intermediate in the citric acid cycle and a central component in the de novo synthesis of fatty acids in the mammary glands (Garnsworthy et al., 2006; Akkerman et al., 2019). More recently, Billa et al. (2020) and Pires et al. (2022) also highlighted isocitrate, free glucose, and glucose-6 phosphate (**glucose-6P**) as relevant biomarkers of energy balance. Krogh et al. (2020) observed that variation of these molecules was largely due to between-cow variation, which supports their use as biomarkers at the individual cow level. Energy deficit and excessive body fat mobilization may cause subclinical or clinical ketosis when the supply of nonesterified fatty acids overloads the liver and their degradation products are diverted to ketone bodies (Ingvartsen, 2006; Esposito et al., 2014). Ketosis is already routinely monitored through well-known milk biomarkers that are the major ketone bodies, i.e., β-hydroxybutyrate (**BHB**) and acetone (Enjalbert et al., 2001). For mastitis, somatic cell count (**SCC**) is the main milk biomarker. However, this measurement relies on flow cytometry, which is a complex and expensive methodology. Therefore, some authors have suggested alternative indicators to improve detection and diagnosis of mastitis. For example, Chagunda et al. (2006) and Hovinen et al. (2016) suggested N-acetyl-β-d-glucosaminidase activity (**NAGase**) as a relevant indicator because it reflects damage to epithelial cells rather than the counting of somatic cells, and may help in discriminating between minor and major pathogen infections. Alternatively, lactate dehydrogenase (**LDH**) is an enzyme that is

part of the glycolytic pathway and is known to increase with mastitis, and shows high sensitivity to mastitis detection (Chagunda et al., 2006). Krogh et al. (2020) observed that NAGase and LDH were not heavily influenced by herd factors and concluded that they could be considered as useful biomarkers for mastitis at cow level for precision livestock farming. Finally, measuring progesterone in milk would provide important information on the reproductive status of dairy cows, particularly for the detection of resumption of ovulation, pregnancy or atypical ovarian patterns due to delayed ovulation or presence of a persistent corpus luteum (Bulman and Lamming, 1978; Friggens and Chagunda, 2005; Crowe, 2008)

However, the development of MIR models to predict these biomarkers raises several challenges or questions. First, although MIR models predicting citrate, acetone, and BHB already exist (De Roos et al., 2007; Grelet et al., 2016), to our knowledge this is the first study attempting to predict milk isocitrate, glucose-6P, free glucose, NAGase, LDH, and progesterone with MIR while predicting in routine these biomarkers could be of great interest. Moreover, concentrations of some of these molecules are far below the considered detection threshold of 100 ppm with MIR (Dardenne et al., 2015). Previous works have shown the possibility to get information on totally indirect phenotypes, e.g., predicting body energy status of cows, or their methane emissions (McParland et al., 2011; Dehareng et al., 2012). These models rely on the global changes in milk composition and their associations with the phenotypes of interest, but the possibility to extract information from milk MIR spectra linked with isocitrate, glucose-6P, free glucose, NAGase, LDH, and progesterone is unknow. In addition, some molecules show a particular distribution, with most healthy cows having low content, while sick cows show an exponential increase of the biomarker of interest. The resulting positively skewed distribution is particularly difficult to model, as it goes beyond the linear relationship of milk molecules and spectral absorbance values. Other authors emphasis the lack of compatibility between the main modeling method with MIR, i.e., partial least square regression (PLS), with these asymmetric distributions of molecules, as well as the need for alternative modeling methodologies (Soyeurt et al., 2020; Kostensalo et al., 2023). Finally, to be used for large-scale phenotyping, MIR models must be robust (i.e., provide reliable predictions under all conditions) covering as much variability as possible to avoid extrapolation (Grelet et al., 2021).

Therefore, the objectives of this study were: 1) to evaluate the potential of milk MIR spectra to predict key biomarkers of energy deficit (citrate, isocitrate, glucose-6P, free glucose), ketosis (BHB and acetone), mastitis (NAGase and LDH), and fertility (progesterone); 2) to test alternative methodologies to PLS, Random-downsampling-PLS and KennardStone-downsampling-PLS, specifically designed to better account for the specific asymmetric distribution of biomarkers, and SVM, which is known for its capacity to handle non-linear relationships; and 3) to create robust models by merging large data sets from 5 international or national projects, reaching a total number of 9143 samples.

MATERIALS AND METHODS

Projects and data

Data used in this study were collected in 3 international and 2 national projects, allowing the merger of a total of 9,143 samples from 3,758 cows across 589 herds in 10 countries, collected from 2013 to 2020. Samples were collected in experimental and commercial herds within the frame of OptiMIR project (Interreg IVB NEW) and GplusE project (Genotype Plus Environment, FP7-Project, http://www.gpluse.eu) and in commercial herds within IndiKuh project (IndiKuh, funding code: 2817905815), D4Dairy project (FFG comet with support of the Austrian government, project 872039, https://d4dairy.com/), and a Swiss national project (data provided by Qualitas). The main characteristics of each sampling (i.e., countries, breeds, number of samples…) are reported in Table 1. The merging of these data sets provided wide variability in terms of breeds, lactation stages, parities, diets, seasons, management practices, and geographical areas.

Milk analysis for biomarkers

All samples were collected following the guidelines edited by the International Committee for Animal Recording (ICAR Dairy Cattle Milk Recording Working Group, 2017) and with ICAR approved milk samplers. Morning and evening samples were collected for MIR analysis and only morning samples were analyzed by reference analysis. Morning samples were therefore split in 2 aliquots, for both reference and MIR analysis. Aliquoting was realized directly in the milking room when milk was still at udder temperature to prevent from fat-aqueous phasing and unperfect aliquoting. Aliquots for biomarkers analysis were store at 4°C right after sampling, without preservative and were stored within 2 h at −18°C until shipping. Samples were sent frozen, with refrigerated delivery with dry ice or ice blocks to the respective labs. Samples from the OptiMIR, IndiKuh and D4Dairy projects were analyzed at CRA-W (Belgium) for BHB, acetone and citrate. Analysis were performed with a continuous flow analyzer (Scan ++, Skalar, Bre-

Table 1. Main sampling characteristics per projects

Project	Year of sampling	Countries	Breeds	Lactation stage	Herds	Cows	Samples
OptiMIR	2013–2014	France, Germany, Luxembourg	Abondance, Montbéliarde, Holstein, and Normande	Entire lactation	429	755	1,170
GplusE	2015–2017	Belgium, Denmark, England, Germany, Ireland, Italy, and Northern Ireland	Holstein	Calving to DIM 50	50	2,670	5,947
Qualitas	2016	Swiss	Holstein, Brown Swiss, and Swiss Fleckvieh	Calving to DIM 125	6	210	1,732
IndiKuh	2017–2018	Germany	Holstein	Calving to DIM 41	5	44	195
D4Dairy	2020	Austria	Simmental, Holstein, Fleckvieh, and crossedbred cows	Entire lactation	99	99	99
Total					589	3,758	9,143

da, The Netherlands) following the procedure described by (De Roos et al., 2007; Grelet et al., 2016). All samples were analyzed twice, and samples with variation higher than 5% were re-analyzed. Samples from the GplusE project were analyzed at the Department of Animal and Veterinary Sciences, Aarhus University, Denmark, for isocitrate, glucose-6P, free glucose, BHB, NAGase, LDH, and progesterone. Fluorometric end-point analyses were used to determine milk glucose and glucose-6P (Larsen, 2015), isocitrate (Larsen, 2014a) and BHB (Larsen and Nielsen, 2005). The indigenous enzymes LDH (EC. 1.1.1.27) and NAGase (EC 3.2.1.30) were analyzed by fluorometric assays according to Larsen (2005) and Larsen et al. (2010) to provide results in µmol product (4-MeU)/min*ml, but are later expressed in the document in Unit/L. For these molecules, intra plate assay repeatability was below 4.5%CV and inter plate assay reproducibility was below 6.0%CV. Milk progesterone was determined using a commercial ELISA assay (Ovucheck, Biovet, Canada), based on the competitive binding of unlabeled progesterone present in the standard or whole milk sample, and a fixed quantity of progesterone labeled with the enzyme alkaline phosphatase (AP), to binding sites on a limited amount of specific progesterone antibodies. After incubation, all components other than those bound to the plate wells were washed away. The amount of bound AP-labeled progesterone remaining on the wells was inversely proportional to the concentration of the unlabeled progesterone present in the sample. The bound labeled progesterone was then measured by reacting the AP with its substrate during a second incubation. The color produced was measured spectrophotometrically and the concentration of progesterone in the milk was determined from a standard curve. The recommendations given by the manufacturer were followed. Intra plate assay repeatability was 5.2%CV and inter plate assay reproducibility was 6.8%CV. Samples from the Swiss national project were analyzed for acetone with an AutoAnalyzer 3 (BRAN +LUEBBE). Not all 9 biomarkers were analyzed in all projects or sub-projects. Of the 9,143 samples, the number of samples analyzed for each biomarker ranged from 600 for citrate to 7,166 for BHB. Table 2 shows the number of analysis for each biomarker according to countries.

Milk samples analysis for MIR spectra

Aliquots for MIR analysis were stored at 4°C with 0.02–0.03% bronopol until analysis. Analyses were performed locally on a wide range of instruments with a total of 30 spectrometers used to analyze the samples: 24 Foss instruments of models FT2, FT6000, FT+, and FT7 (Foss, Hillerød, Denmark), 5 Bentley FTS (Bent-

Table 2. Number of samples used in models per biomarker and per country

	Isocitrate (mmol/l)	Citrate (mmol/l)	Glucose6P (mmol/l)	Free glucose (mmol/l)	BHB (μmol/l)	Acetone (mmol/l)	NAGase (Unit/l)	LDH (Unit/l)	Progesterone (ng/ml)
Austria		98			99	86			
Belgium	594		594	594	617		594	594	291
Denmark	332		338	338	341		338	338	338
England	2,014		2,015	2,015	2,043		2,004	2,004	
France		316			681	149			
Germany	105	82	105	105	500	44	105	105	105
Ireland	1,145		1,139	1,112	1,164		1,143	1,142	156
Italy	439		439	439	650		439	439	439
Luxembourg		104			197	49			
Northern Ireland	874		872	873	874		874	874	651
Switzerland						1671			
Total	5,503	600	5,502	5,476	7,166	1,999	5,497	5,496	1,980

NAGase = N-acetyl-β-d-glucosaminidase; LDH = lactate dehydrogenase.

ley, Chaska, United States) and 1 Standard Lactoscope FT-MIR automatic (Delta Instruments, Drachten, The Netherlands). All the spectra from the different instruments were standardized to be merged into a common data set following the procedure described in Grelet et al. (2015). Morning and evening samples were analyzed separately but merged into a daily spectrum, to be in line with milk recording 24H milk samples, by a weighted average considering the AM and PM milk yields.

Data editing and MIR models development

To avoid erroneous association between spectra and sample, as well as analytical issues during analysis, a local fat model was applied to the spectra and the generated predictions were compared with the predictions provided by the laboratories. Records with a difference above 0.3 g/100mL between local and laboratory predictions were discarded (n = 218 records deleted) to prevent of wrong association between spectra and samples, or other analytical errors. This threshold was highlighted in a study of Zhang et al. (2021). Spectra with a standardized Mahalanobis distance (GH) greater than 10 were eliminated (n = 54 records deleted). A high GH value was intentionally used as threshold as previous work shown that keeping large spectral variability in the data set was benefiting to the robustness of the developed model (Grelet et al., 2021). Additionally, only records with DIM between 5 and 365 were retained (n = 88 records deleted) and reference values being under quantification thresholds were discarded. All these aforementioned edits eliminated 9% of the samples, and resulted in a data set of 8,783 records. In descriptive statistics, skewness of each variable was calculated as:

$$Skewness = \frac{N}{(N-1)(N-2)} \sum \frac{x_i - mean^3}{SD},$$

with N being the number of samples and x_i the ith observation within each variable.

The MIR spectra were pretreated by a first derivative with a gap of 5 wavenumbers. The selected spectral area consisted of 212 wavenumbers from 968.1 to 1,577.5 cm^{-1}, 1,731.8 to 1,762.6 cm^{-1}, 1,781.9 to 1,808.9 cm^{-1}, and 2,831.0 to 2,966.0 cm^{-1} to exclude spectral areas not reproducible between instruments (Grelet et al., 2021) and the absorbance values were mean-centered.

For each biomarker, a data set was constituted by keeping only the records with reference values and MIR spectra. Each data set was then split to create an external herd validation data set, randomly removing 20% of the herds (i.e., 120 herds out of 589) across all countries to be used as the validation data set. The remaining 80% were used as the calibration data set. The modifications to the distribution described below were only performed in the calibration data set and the original distribution for each biomarker in the validation data set was kept unchanged, to be as close as possible to the real field application conditions. In both the calibration and validation data sets for each biomarker, all samples were retained without removing outliers.

As a basis for comparison, the first models were developed using partial least squares regression (**PLS**), this being the most commonly used method. The number of latent variables was set according to the breakpoint of the Root Means Square Error (RMSE) slope during the cross-validation step. However, as mentioned by Soyeurt et al. (2020), PLS is not adapted to handle asymmetric distributions of molecules due to its linear structure. Therefore, to better take into account the strongly right skewed distributions of some molecules, 2 methodologies of modification of the distribution were

specifically developed and tested to move toward a normal distribution. PLS regression was later applied on the modified data sets.

The first methodology was the one used in Grelet et al. (2016), randomly removing a portion of the low values to reduce the over-representation of low values in the data set. The thresholds for discriminating low and high values were obtained from the literature or were optimized during the cross-validation step regarding the model R^2cv, and the proportion of low values to be removed was calculated to balance equally the proportion of low and high values. The random elimination was performed by cyclic iterations. After down-sampling, the proportion of high values, and especially extreme high values, were still under-represented. Therefore, a logarithmic (10) transformation was applied to the reference values to be closer to a normal distribution. PLS was then applied to the reduced data set. This random modification of the distribution associated with PLS is later referred to as "**Random-downsampling-PLS**."

However, the random elimination of samples, by discarding samples of interest, inevitably leads to loss of variability and robustness of the models. Therefore, a second methodology was also tested for an optimized modification of the distribution to keep only the most informative samples in the data set while harmonizing the distribution of samples over the range of reference values. To do this, a 3-step methodology was applied which is schematically represented in Figure 1. The data set for each biomarker was first divided into 20 sub-sets of equal Y-interval across the Y range. In a second step, within each subset, a fixed number of samples (e.g., n = 100) was selected using the Kennard-Stone algorithm, which iteratively selects the 2 most spectrally different samples until the number of samples to be selected was reached. The number of samples to be selected was manually optimized between 50 and 200. When the number of samples in the subset was lower than the number of samples to be selected (i.e., at the right end of the distribution containing samples with high content of biomarkers), all the samples were retained. In a third step, the selected samples among the 20 subsets were merged to compose a data set with a harmonized distribution along the Y range, but keeping the most informative samples from the initial data set. Finally, a logarithmic (10) transformation was applied to the reference values and a PLS regression was applied to the data as the extreme values were still under-represented. This optimized modification of the distribution associated with PLS is later referred to as "**KennardStone-downsampling-PLS**."

Support vector machines regression (**SVM-R**) was tested as an alternative quantitative methodology to overcome the specific distribution. SVM-R is a linear method adapted to nonlinear relationships due to its capacity to find a linear link in a space of higher dimension. In regression the support vectors are the ones including a maximum of samples in the regression within an acceptable margin (Brereton and Lloyd, 2010). SVM-R was used after a PLS compression to reduce the dimension of the data set to 14 latent variables. The LIBSVM algorithm was used with the epsilon version and a radial basis function kernel. The gamma, cost and epsilon hyperparameters were optimized with a grid-search to minimize the RMSEcv. Due to its capacity to handle nonlinear data, no Log-transformations were applied to biomarker contents.

Finally, when the ability of MIR to predict biomarkers was too low to enable rough quantitative screening (R^2cv < 0.74; Grelet et al., 2021), a qualitative methodology was tested to assess the possibility of discriminating low vs high values of biomarkers. Discriminant models were developed with partial least squares discriminant analysis (**PLS-DA**) using the full data set, the randomly balanced data set (**Random-downsampling-PLS-DA**), or the optimized balanced data set (**KennardStone-downsampling-PLS-DA)**. Thresholds to discriminate low and high values came from the literature, 200 μmol/L for BHB (Denis-Robichaud et al., 2014), 150 mmol/L for acetone (De Roos et al., 2007), 4.3 Unit/L for LDH (Chagunda et al., 2006), 5 ng/ml for progesterone (Roelofs et al., 2006), or from personal communication for NAGase (8 Unit/L; Torben Larsen, Aarhus University, Denmark, personal communication). No threshold could be found for 3 biomarkers of energy deficit, isocitrate, free glucose, and glucose-6P, therefore discriminant values for feed restrictions were visually estimated from published data; 0.15 mmol/L for isocitrate (Pires et al., 2022), 0.25 mmol/L for free glucose (Pires et al., 2022), and 0.3 mmol/L for glucose-6P (Billa et al., 2020).

The models were evaluated using the external herd validation set. When a logarithmic transformation was used, predictions were back-transformed ($10^{\text{prediction}}$) to further evaluate model performance in a usual scale and for graphical representation. Model statistics, both in the calibration and validation steps, were expressed in terms of R^2 (coefficient of determination), and RMSE. Ratio of RMSE/standard deviation of global data set (RPD) was not calculated as being not relevant for models with asymmetric distribution, for which SD does not describe correctly the spread of the population (Bellon-Maurel et al., 2010). As a last step, final models were developed using all the data, calibration and validation, and internal cross-validation R^2cv and RMSEcv (10 subsets constituted by cyclic iteration) are shown. Discriminant model statistics were expressed

in terms of sensitivity (percentage of good classification in the high content group), specificity (percentage of good classification in the low content group) and global accuracy (global percentage of correct classification). Computations and models were carried out with programs developed in Matlab 2022 (The Mathworks,

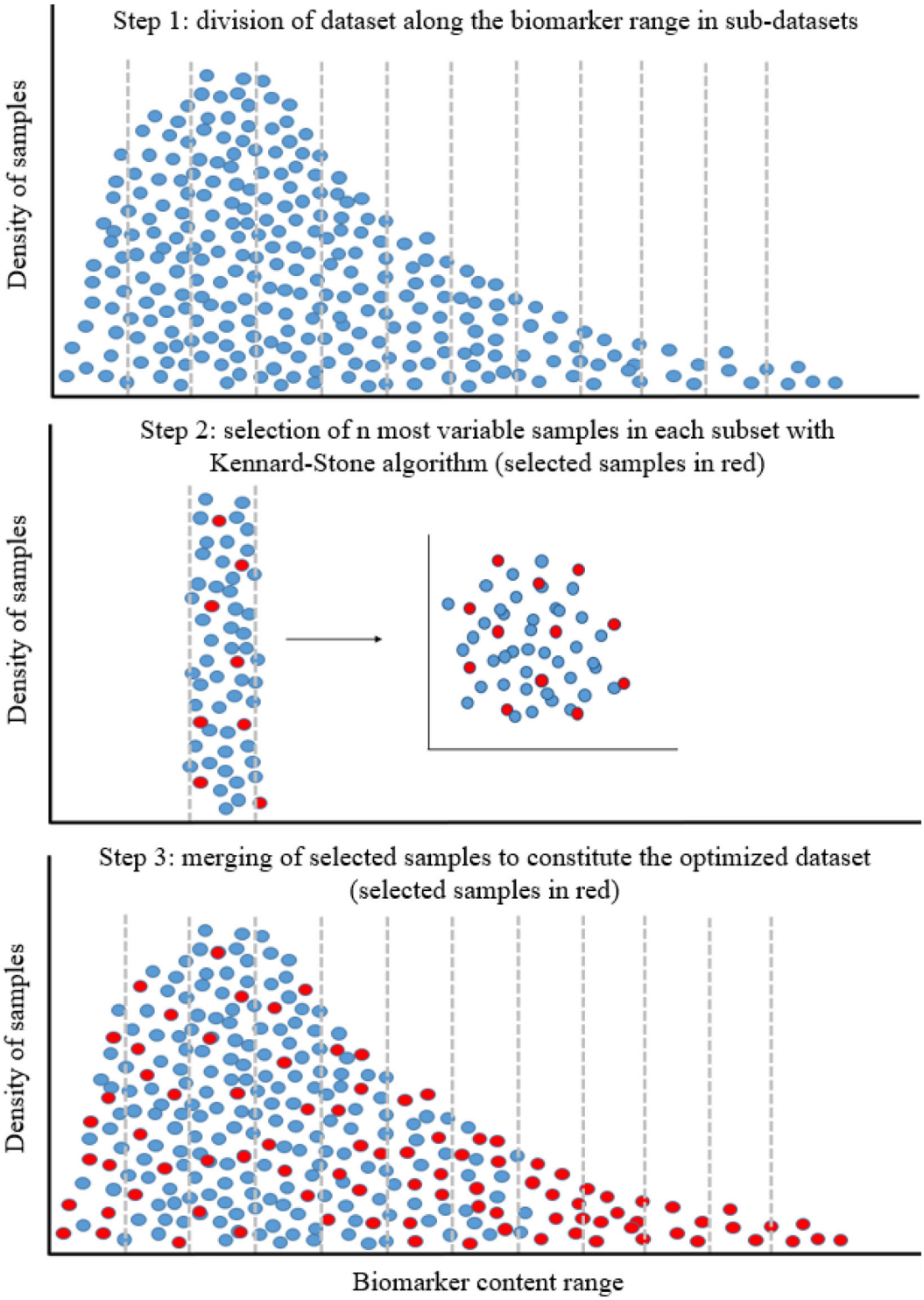

Figure 1. Schematic representation of the modification of biomarkers distribution using Kennard-Stone algorithm to sub-data sets. Samples in red are selected to constitute the calibration data set.

Table 3. Descriptive statistics of the chemical analysis results for the 9 biomarkers

	Minimum	Maximum	Mean	Standard deviation	Skewness
Isocitrate (mmol/l)	0.016	0.447	0.156	0.053	0.86
Citrate (mmol/l)	3.88	16.12	8.90	2.19	0.33
Glucose-6P (mmol/l)	0.001	0.808	0.121	0.075	1.52
Free Glucose (mmol/l)	0.001	0.998	0.284	0.120	0.62
BHB (µmol/l)	20	1989	90	102	5.88
Acetone (mmol/l)	0.005	3.355	0.070	0.148	11.49
NAGase (Unit/l)	0.00	25.10	1.91	1.61	4.52
LDH (Unit/l)	0.00	45.96	2.53	2.70	5.99
Progesterone (ng/ml)	0.50	22.44	5.28	2.76	1.21

NAGase = N-acetyl-β-d-glucosaminidase; LDH = lactate dehydrogenase.

Inc., Natick, MA, USA) and the PLS toolbox v. 8.5.1 (Eigenvector Research, Inc.,Wenatchee, WA, USA).

RESULTS AND DISCUSSION

Descriptive statistics of reference values

Descriptive statistics are shown in Table 3. The mean value for each biomarker was in the same order of magnitude as in previous studies, e.g., means of 0.179 mmol isocitrate (Larsen, 2014a), 9.04 mmol/L citrate (Grelet et al., 2016), 0.081mmol/L glucose-6P and 0.331 mmol/L free glucose (Larsen and Moyes, 2015), 146 µmol/L BHB (De Roos et al., 2007) and 0.100 mmol/L acetone (Denis-Robichaud et al., 2014), 2.7 Unit/L NAGase and 2.4 Unit/L LDH (Åkerstedt et al., 2011), and progesterone between 0.8 and 22.8 ng/ml (Ginther et al., 1976). However, it should be noted that glucose-6P is higher than in the above referenced study (+33%) while BHB, acetone and NAGase are considerably lower, −62%, −47%, and −41% respectively. This could be explained by the over-representation of healthy cows in our data set, as most of the sampling did not focus on comparing healthy and sick cows, or by minor differences in the analytical process. Table 3 also shows that both ketosis (BHB and acetone) and mastitis biomarkers (NAGase and LDH) have asymmetric distributions with strong positive skewness. Figure 2 shows the distribution of each molecule according to sampling countries. It also particularly highlights the positive skewness of distribution of ketosis and mastitis biomarkers and their exponential increase in case of disorder.

The Pearson correlation table between biomarkers is shown in Table 4. As different projects analyzed different biomarkers, not all the correlations could be calculated. Isocitrate is derived from citrate during the Kreb's cycle and both molecules were found to be positively correlated in Larsen (2014b). Citrate decreases with the increase of de novo fatty acids synthesis (Garnsworthy et al., 2006) and is mentioned as an early indicator of physiological imbalance of the animal (Bjerre-Harpøth et al., 2012b). In the current data set the correlation coefficients (r) between isocitrate and citrate could not be calculated, but both are weakly to moderately positively correlated with BHB (r = 0.24 and r = 0.45 for isocitrate and citrate respectively). This moderate correlation with BHB may reflect that they are also associated with physiological imbalance but at a different degree of severity. A decrease of citrate has also been observed in case of mastitis (Hyvönen et al., 2010), however, in the present study a weak positive correlation is observed between isocitrate and NAGase and LDH, both of which increasing in cases of mastitis. Glucose is not synthetized in the mammary epithelial cells and is therefore directly dependent of blood glucose absorbed in the mammary gland (Larsen and Moyes, 2015). Free glucose was negatively correlated with both Glucose-6P and isocitrate as observed by (Pires et al., 2022). Free glucose is positively correlated with plasma glucose and milk lactose, and decreases in case of energy deficit (Larsen and Moyes, 2015; Pires et al., 2022). On the contrary, glucose-6P is negatively correlated with milk lactose and increases in case of energy deficit, probably due its role in the pentose phosphate cycle and as an intermediate in glycolysis (Larsen and Moyes, 2015; Pires et al., 2022). The 2 molecules were not correlated with BHB, reflecting their complementarity. Low glucose level associated with high BHB indicates an imbalanced cow, whereas high BHB associated with normal glucose may be associated to a false ketosis diagnistic. Both glucose-6P and free glucose were moderately correlated with NAGase and LDH, which is probably explained by the damaging effect of mastitis on mammary epithelial cells. The 2 biomarkers of ketosis (BHB and acetone) were strongly correlated, with an r = 0.69. A similar relationship (r = 0.71) was observed between the mastitis biomarkers, NAGase and LDH.

Quantitative MIR models

Table 5 shows the performance of the 4 different modeling methodologies with the 9 biomarkers during the calibration step with 80% of herds. The results are shown in terms of calibration RMSE (RMSEc) and coefficient of determination (R^2c). The 3 alternative methodologies to PLS seemed to outperform PLS for most of biomarkers. PLS-Random-downsampling provided the best results for acetone, PLS-KennardStone-downsampling the best results for isocitrate and progesterone, while SVM-R the best results for citrate, glucose-6P, Free Glucose, BHB, NAGase, and LDH, suggesting a better ability to take into account the specificities of the distributions. However, as the distribution is modified in PLS-Random-downsampling and PLS-KennardStone-downsampling, the RMSEc is artificially inflated due to the elimination of low values which are better predicted than high values. Alternatively, the removing of low samples inflates the R^2 which is directly

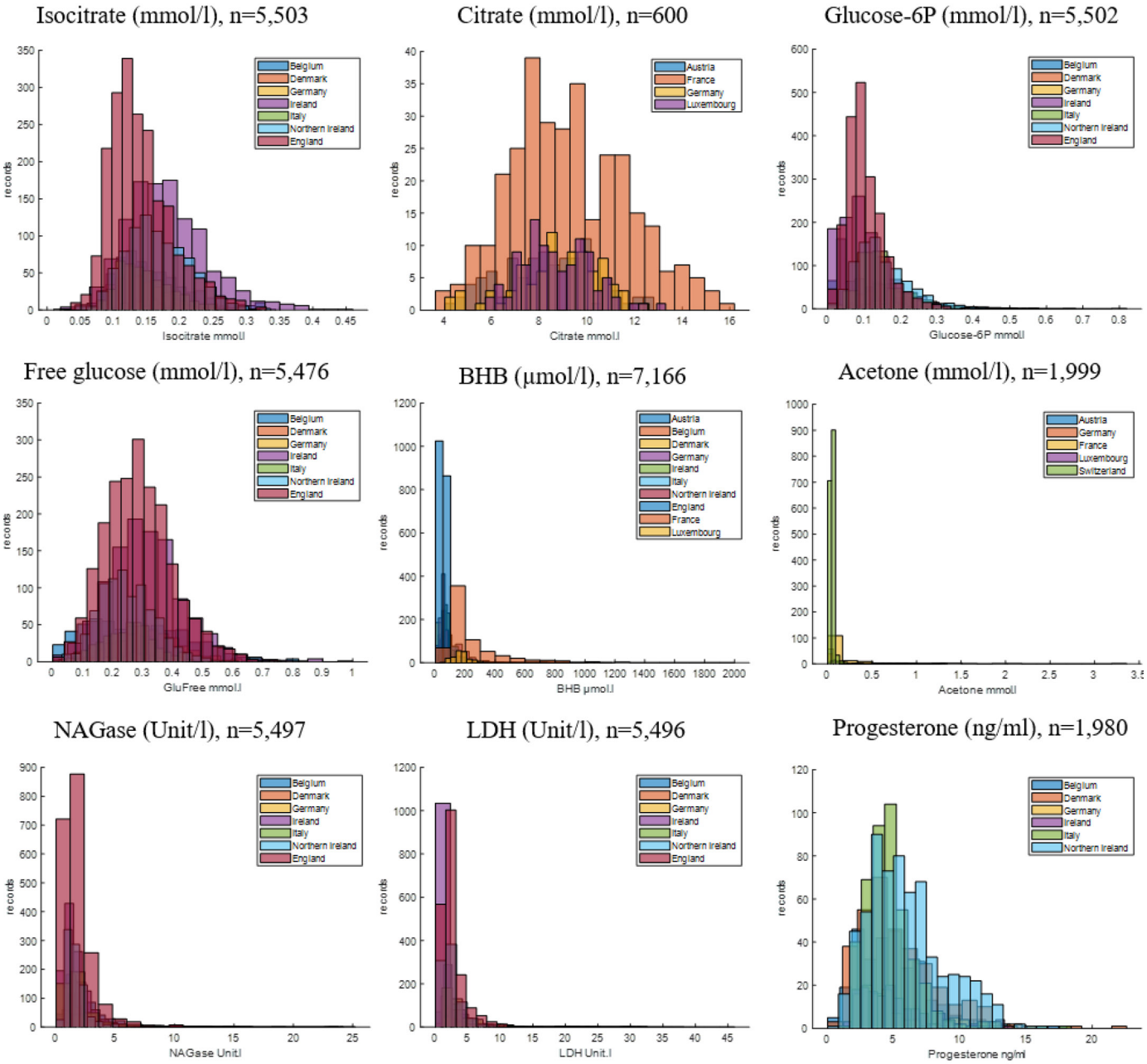

Figure 2. Distribution of each biomarker according to the sampling country.

Table 4. Correlation between the reference measurement of biomarkers

	Isocitrate (mmol/l)	Citrate (mmol/l)	Glucose-6P (mmol/l)	Free Glucose (mmol/l)	BHB (µmol/l)	Acetone (mmol/l)	NAGase (Unit/l)	LDH (Unit/l)
Isocitrate (mmol/l)								
Citrate (mmol/l)	0.41***							
Glucose-6P (mmol/l)	−0.31***		−0.44***					
Free Glucose (mmol/l)	0.24***	0.45***	−0.01^{NS}	−0.15***				
BHB (µmol/l)		0.39***			0.69***			
Acetone (mmol/l)	0.22***		0.40***	−0.30***	0.09***			
NAGase (Unit/l)	0.18***		0.39***	−0.31***	0.03*		0.71***	
LDH (Unit/l)	−0.26***		−0.22***	0.19***	−0.05*		−0.16***	−0.20***
Progesterone (ng/ml)								

NS for $P \geq 0.05$; * for $P \leq 0.05$; ** for $P \leq 0.01$; *** for $P \leq 0.001$; NAGase = N-acetyl-β-d-glucosaminidase; LDH = lactate dehydrogenase.

affected by a better distribution over the range (Davies and Fearn, 2006). Finally, in the calibration step some methodologies may overfit the hyperparameters to this particular data set. Therefore, the calibration statistics does not allow to compare the methods on a common basis and the results should be considered with caution and in an informative way. Consequently, for the selection of best model, the external validation statistics are to be considered only.

Table 6 shows the performance of models in external herd validation, by applying the models to the 25% excluded herds across countries. Looking at the R^2v and RMSEv, the best models in calibration were not necessarily the best in the validation step. Indeed, PLS provided the best results for isocitrate, glucose-6P, free glucose and LDH. SVM-R provided the best results for citrate, BHB, acetone, NAGase, and rogesterone. PLS-Random-downsampling and PLS-KennardStone-downsampling provided better results than PLS for some biomarkers (glucose-6P, BHB, acetone, LDH, and progesterone for PLS-Random-downsampling and BHB, acetone, and progesterone for PLS-Kennard-Stone-downsampling). These molecules were the ones with highly skewed distributions. However, for these molecules, the best models were obtained using SVM-R. Therefore, although these 2 methodologies might provide some improvement compared with PLS, SVM-R seems to be the best strategy to take into account non-linearity, i.e., the exponential increase of minor molecules not associated with exponential changes in the main components. Therefore, with the current data set, it is worth using a method adapted to non-linearity rather than modifying the distribution to approach normality. Additionally, SVM-R provides better results with the entire data set than when associated with modification of distribution (results not shown), which enables keeping all the variability of the calibration data set and provides a higher robustness to the models, leading to better results in the external validation. Therefore, classical PLS and SVM-R based on the full data set provided the best results for the normal and skewed distributions, respectively. The results also indicate that retaining the full variability of the calibration data set increases the robustness of the model and the quality of predictions when applied to external data. This demonstrates the benefits of expanding data sets through collaborations.

As a final step, the calibration and validation data sets were merged to develop final models across the entire data set, covering a maximum of variability. These final models were developed with the best methodologies highlighted for each molecule during the herd validation step, i.e., PLS for isocitrate, glucose-6P, Free Glucose, and LDH an SVM-R for citrate, BHB, ac-

Table 5. Calibration performance of MIR models. Calibration step is performed with 80% of herds among countries (no outliers were removed). Distributions of biomarkers were modified in PLS-Random-downsampling and PLS-KennardStone-downsampling, which affect both RMSEc and R^2c

			R^2c				RMSEc			
Calibration 80% herds	n	#LV	PLS	PLS Random-downsampling	PLS KennardStone-downsampling	SVM-R	PLS	PLS Random-downsampling	PLS KennardStone-downsampling	SVM-R
Isocitrate (mmol/l)	4,211	13	0.49	0.53	0.53	0.52	0.037	0.045	0.060	0.036
Citrate (mmol/l)	520	8	0.87	0.86	0.87	0.91	0.777	0.874	0.950	0.661
Glucose-6P (mmol/l)	4,200	11	0.59	0.64	0.51	0.73	0.047	0.064	0.084	0.044
Free Glucose (mmol/l)	4,324	12	0.42	0.48	0.53	0.56	0.091	0.108	0.110	0.080
BHB (µmol/l)	5,794	14	0.40	0.60	0.50	0.72	77.0	134.2	173.0	55.6
Acetone (mmol/l)	1,671	13	0.37	0.82	0.68	0.61	0.120	0.198	0.284	0.097
NAGase (Unit/l)	4,204	13	0.38	0.43	0.35	0.56	1.31	2.42	2.34	1.08
LDH (Unit/l)	4,203	13	0.33	0.28	0.27	0.43	2.37	5.25	4.71	2.23
Progesterone (ng/ml)	1,640	8	0.14	0.25	0.28	0.18	2.50	2.95	3.12	2.44

R^2c = coefficient of determination in calibration; RMSEc = root mean square error of calibration; n = number of samples in calibration; #LV = number of latent variables in PLS models; PLS = partial least square regression; SVM-R = support vector machine regression; NAGase = N-acetyl-β-d-glucosaminidase; LDH = lactate dehydrogenase.

Table 6. Validation performance of MIR models. External validation is performed by removing 20% herds among countries to be used as validation set

		R^2v				RMSEv			
External 25% herds validation	n	PLS	PLS Random-downsampling	PLS KennardStone-downsampling	SVM-R	PLS	PLS Random-downsampling	PLS KennardStone-downsampling	SVM-R
Isocitrate (mmol/l)	1,292	**0.48**	0.42	0.42	0.44	**0.04**	0.04	0.04	0.04
Citrate (mmol/l)	80	0.91	0.88	0.88	**0.94**	0.70	0.84	0.92	**0.59**
Glucose-6P (mmol/l)	1,288	**0.58**	0.59	0.49	0.49	**0.05**	0.07	0.07	0.06
Free Glucose (mmol/l)	1,152	**0.28**	0.28	0.25	0.26	**0.11**	0.10	0.11	0.11
BHB (µmol/l)	1,372	0.37	0.48	0.42	**0.58**	90.83	85.03	100.92	**76.50**
Acetone (mmol/l)	325	0.49	0.66	0.61	**0.76**	0.10	0.12	0.08	**0.07**
NAGase (Unit/l)	1,292	0.38	0.35	0.31	**0.68**	1.25	2.03	1.92	**0.81**
LDH (Unit/l)	1,291	**0.24**	0.25	0.20	0.20	**1.93**	2.98	2.60	1.93
Progesterone (ng/ml)	337	0.13	0.15	0.15	**0.15**	2.81	3.12	2.97	**2.81**

R^2v = coefficient of determination in external herd validation; RMSEv = root mean square error of external herd validation; n = number of samples in validation; PLS = partial least square regression; SVM-R = support vector machine regression. Bold R^2v and RMSEv highlight the best modeling method; NAGase = N-acetyl-β-d-glucosaminidase; LDH = lactate dehydrogenase.

etone, NAGase, and progesterone. These models could only be validated in cross-validation, with 10 subsets constituted by cyclic iteration. Figure 3 shows these final models, with measured versus predicted values. Energy balance biomarkers were predicted with R^2cv of 0.50, 0.88, 0.59, and 0.40 and RMSEcv of 0.037, 0.76, 0.048, and 0.093 mmol/L for isocitrate, citrate, glucose-6P and free glucose respectively. Biomarkers of ketosis were predicted with R^2cv of 0.61 and 0.60, and RMSEcv of 64.6 μmol/L and 0.094 mmol/L for BHB and acetone respectively, which is comparable to results obtained by De Roos et al. (2007), with RMSE of 65μmol/L and 0.070mmol/L for BHB and acetone and shows a lower error than obtained by Heuer et al. (2001), with error of 0.210 mmol for acetone. Mastitis biomarkers were predicted with R^2cv of 0.42 and 0.29, and RMSEcv of 1.23 and 2.28 unit/L for NAGase and LDH respectively. Finally, progesterone as an indicator of the reproductive status of dairy cows, was predicted with a R^2cv of 0.13 and a RMSEcv of 2.57 ng/ml, which is not surprising knowing than mean progesterone was approximately 5ppb in milk, and than other changes associated to fertility status in milk composition are probably insignificant compared with the detection threshold of 100ppm with MIR in milk.

Regarding the possibility of using MIR to predict biomarkers of energy balance, citrate was very well predicted, with a R^2v of 0.94 and a relative error lower than 10%, which means that accurate quantitative screening can be considered with this model. The glucose-6P model shows a R^2v of 0.58 but a very high relative error (RMSE/mean) of 41%, indicating that the model is very inaccurate and should only be considered to detect extreme values (Grelet et al., 2021). Therefore, it is more reliable to focus on citrate predictions to obtain information on cow status. It is recommended not to use isocitrate and free glucose prediction models, due to low R^2v of 0.48 and 0.28, respectively.

Regarding biomarkers of ketosis, BHB models have similar performances in cross-validation and validation, while the final acetone model shows a R^2cv of 0.6, which is lower than the R^2 in external validation (0.76). Therefore, the R^2v is probably too optimistic, due to the random exclusion of herds that were likely well predicted in the validation data set, and it seems more relevant to consider the R^2cv (0.6) as a quality indicator of the acetone model. Therefore, BHB and acetone models show similar performances, with both a R^2v or R^2cv of approximately 0.6 and a very large relative RMSE. This performance is similar to previous findings (Grelet et al., 2016), and confirms that these models should be used with caution due to their high inaccuracy, and can only be used to detect extreme values. This is nevertheless potentially sufficient to detect cows suffering from sub-clinical and clinical ketosis. In the case where BHB and acetone information would be redundant in practical use, it is recommended to focus on BHB as it is based on 7,166 records compared with 1,999 for acetone. This large number, which is associated with a better coverage of countries and local conditions, should bring a better robustness to the model by limiting extrapolation to unknown conditions when using with real field data as shown in Grelet et al. (2021). Furthermore, the inconsistency between validation results and cross-validation based on the whole data set shows that the acetone model is dependent on a few extreme high points, indicating that the model is not yet sufficiently robust.

Regarding mastitis biomarker models, as for Acetone, the NAGase R^2v is probably too optimistic, due to the random exclusion of herds that were likely well predicted in the validation data set. The NAGase model shows a more realistic R^2 in cross validation than in validation, and R^2cv should be considered rather than R^2v. NAGase is predicted with a R^2cv of 0.42 and a high relative error (42%), which is not precise enough to be used for individual cow monitoring. Performance of the LDH model, with R^2v of 0.24 and a high relative error (76%) is also too low to be considered for use. Finally, the progesterone prediction model, with an extremely low R^2v (0.15) and a high relative error (53%), cannot be used for cow management.

Qualitative MIR models

When the ability of MIR to predict biomarkers was too low to enable approximate quantitative screening (R^2cv < 0.74; (Grelet et al., 2021)), a qualitative methodology was tested to assess the possibility of discriminating low vs high values for all biomarkers except citrate. Discriminant models were developed using different methodologies associating PLS-DA and under-sampling or not. The best results obtained during the external herd validation are shown in Table 7. PLS-DA on the full data set provided the best classification results for isocitrate, free glucose, acetone, and progesterone, while PLS-DA associated with random under-sampling provided the best results for BHB and LDH, while PLS-DA associated with KennardStone under-sampling provided the best results for glucose-6P and NAGase. The lack of consistency in the best methodology for both normally distributed and skewed molecules shows that no general conclusions can be drawn and that different methodologies should be tested for each biomarker. The results also show that under-sampling methods can be useful to balance the data set before discrimination. Three biomarkers were predicted with a very high percentage of correct clas-

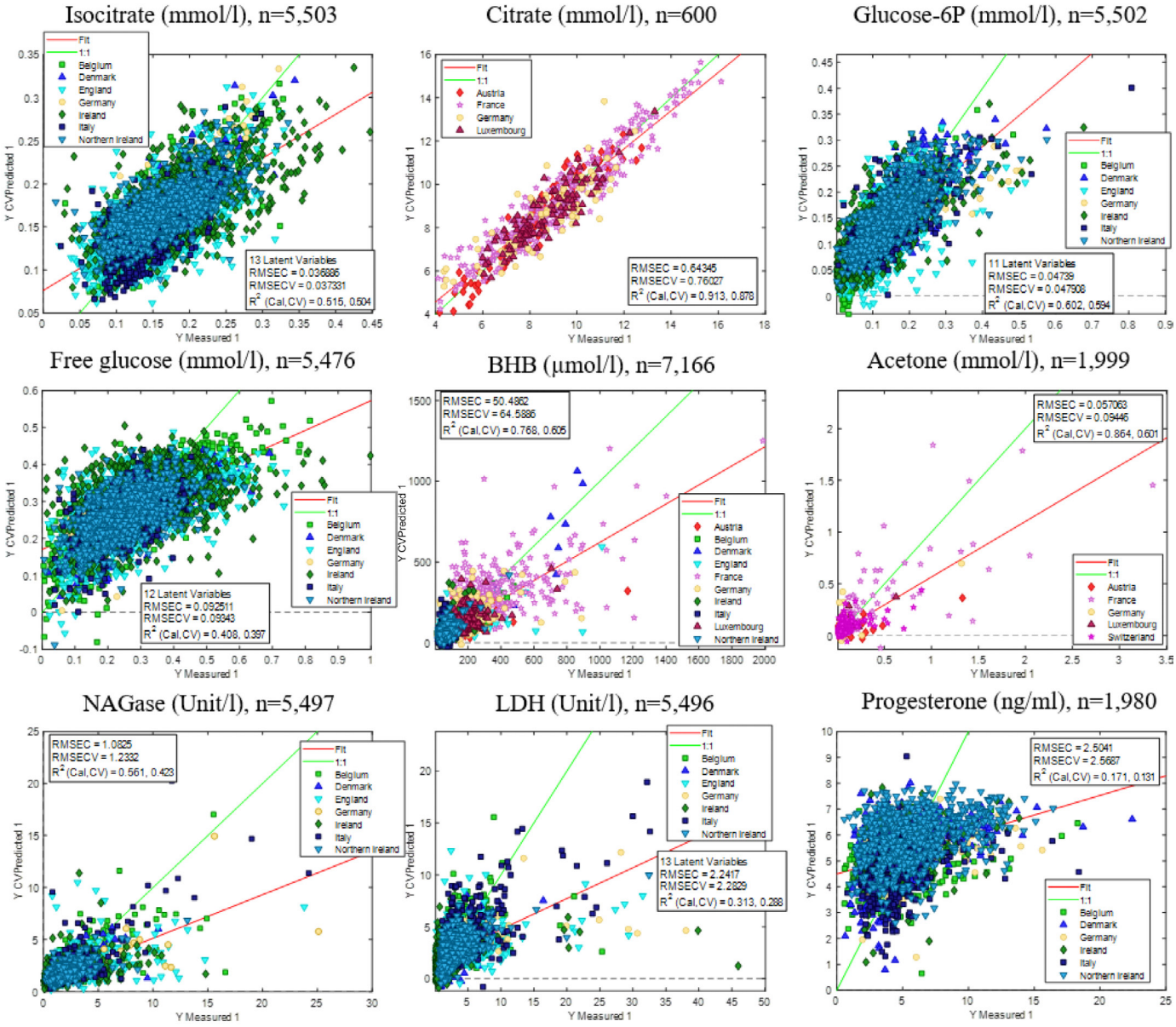

Figure 3. Cross-validation performances of the final models developed on the full data sets. Plots showing measured biomarkers versus predicted biomarkers in the 10-fold cross-validation. The final models were developed with the best methodologies highlighted for each molecule during the herd validation step, i.e., PLS for isocitrate, glucose-6P, free glucose and LDH an SVM-R for citrate, BHB, acetone, NAGase and progesterone. RMSEc = root mean square error of calibration; RMSEcv = root mean square error of 10-folds cross-validation; R^2cal = coefficient of determination of the calibration; R^2cv = determination coefficient of 10-folds cross-validation; NAGase = N-acetyl-β-d-glucosaminidase; LDH = lactate dehydrogenase

sification, with global accuracy around 90% for BHB, acetone, and NAGase. This validates that MIR spectrometry can discriminate high and low values of BHB and acetone. It also indicates that, unlike quantitative models, qualitative discriminant models can discriminate low and high NAGase contents with good accuracy, which could help in the management and detection of mastitis. glucose-6P is predicted with a global accuracy of 83%. The added value of such accuracy needs to be tested in the field to evaluate a potential use of this model. Furthermore, (Billa et al., 2020) mentioned a breed effect on glucose-6P that should be taken into account in the use of raw predictions. Accuracy of other models was less than 75%, therefore it seems difficult to consider their use for herd management.

Table 7. External herd validation results of the qualitative discrimination between low and high values for 8 biomarkers. Only the best methodology between PLS-DA, Random-downsampling-PLS-DA, and KennardStone-downsampling-PLS-DA is shown

Biomarker	Best method	Threshold	Sensitivity	Specificity	Accuracy
Isocitrate	PLS-DA	0.150 mmol/L	84%	60%	75%
Glucose-6P	KennardStone-downsampling-PLS-DA	0.3 mmol/L	83%	83%	83%
Free Glucose	PLS-DA	0.250 mmol/L	81%	57%	67%
BHB	Random-downsampling-PLS-DA	200 µmol/L	88%	92%	92%
Acetone	PLS-DA	0.15 mmol/L	81%	91%	91%
NAGase	KennardStone-downsampling-PLS-DA	8 Unit/L	85%	88%	88%
LDH	Random-downsampling-PLS-DA	4.3 Unit/L	75%	69%	70%
Progesterone	PLS-DA	5 ng/ml	49%	65%	56%

PLS-DA = partial least square discriminant analysis; NAGase = N-acetyl-β-d-glucosaminidase; LDH = lactate dehydrogenase.

Perspectives and limitations

The objectives of this study were: 1) to evaluate the potential of MIR milk spectra to predict key biomarkers of energy deficit (citrate, isocitrate, glucose-6P, free glucose), ketosis (BHB and acetone), mastitis (NAGase and LDH), and fertility (progesterone), 2) to test alternative methodologies to PLS to better account for the specific asymmetric distribution of biomarkers, and 3) to create robust models by merging large data sets from international or national projects.

Milk citrate is predicted with sufficient accuracy to allow accurate quantitative screening of energy deficit of cows. Quantitative and qualitative models predicting BHB and acetone are able to discriminate high from low values, which seems sufficient to detect cows suffering from sub-clinical ketosis. The qualitative discriminant model predicting NAGase, showing a global accuracy of 88% of correct classification, also demonstrates an ability to discriminate high from low values. It could play a role in the detection of mastitis, especially if flow cytometry is not available to provide SCC information, e.g., if in future MIR instruments are miniaturized to be placed in farms. Models predicting progesterone were not good enough to contribute to cow fertility monitoring. However, mastitis, energy deficit and metabolic disease are all major causes of subfertility in dairy cows, so their detection in early lactation will help to predict future fertility issues (Wathes, 2012; Lou et al., 2022). Consequently, MIR spectrometry can provide valuable information in relation to 4 main factors leading to involuntary culling of dairy cows (De Vries and Marcondes, 2020; Dallago et al., 2021). Previous versions of the citrate, BHB and acetone models (Grelet et al., 2016) were already routinely used in the framework of milk recording in several European countries, but contained only a few hundreds of samples. The 3 updated quantitative models should increase the quality of predictions due to the increased variability in the calibration data sets, which has been facilitated by the international collaboration.

However, the predicted raw values should not be considered as an end in themselves. Biomarker concentration may be highly dependent on days in milk, parity, breed and potentially other variables. Therefore, their use and the way information is disseminated at farm level must take these important aspects into account. For example, assessing the dynamics of biomarkers throughout the lactation stage, or comparing with animals of similar characteristics may be more appropriate than using the predicted raw value. These predicted biomarkers also come in addition to the wide range of already existing parameters, and their complementarity, or marginal effect, should be considered. In particular, the possibility to predict NAGase content on such large scale is rather new, and its complementarity with the SCC, as well as its added values, should be investigated. For ketosis, energy deficit biomarkers reflecting plasma glucose may be highly complementary to ketosis biomarkers. Indeed, their combination may enable to screen cow with low glucose and high ketone bodies, which are the critical cases of ketosis to detect (Moyes et al., 2013; Foldager et al., 2020).

In terms of modeling methodologies, the results show that it is worth using methods adapted to non-linearity rather than trying to modify the distribution to approach normality. In the present work, SVM-R has particularly shown good performances for skewed distributions. However, this method based on a higher dimension space and numerous hyperparameters to tune is extremely time consuming to calculate. Working with an elevated number of samples (e.g., several thousand samples) especially increases calculation time as the matrix dimension reduction is performed with XX' (dimension n samples * n samples) instead of X'X (dimension p absorbances*p absorbances). Therefore, not all the potential algorithms and optimizations could be tested and further investigations, requiring higher computation power, may enable accuracy of current results to be improved. Additionally, not all the external herd validation combinations could be tested, which would be necessary to have a stable estimation of

external herd validation performances. SVM-R is only one of the many methods adapted to non-linearity and future research should investigate the potential of other algorithms such as neural networks, kernel methods or weighted regressions.

In addition to herd management, predicted biomarkers can be valuable for genetic evaluations as proxies of dairy cow challenges. McParland et al. (2015) and Bonfatti et al. (2017) mentioned that models predicting phenotypes with low accuracy may successfully contribute in making genetic progress if genetic correlations exist with other traits of interest (e.g., direct health traits).

CONCLUSIONS

Thanks to international collaborations, the working data set consisted of a total of 9,143 records. PLS and SVM-R based on the full data set provided the best results for normal and skewed distributions, respectively. Regarding the ability of MIR to predict biomarkers of energy balance, citrate was very well predicted, which allowed for quantitative screening. Qualitative models indicate that MIR spectrometry can discriminate low and high values of BHB and acetone, as biomarkers of ketosis, with accuracy around 90%. A qualitative discriminant model can discriminate low and high NAGase contents with good accuracy (88% good classification), which could help in mastitis management and detection, especially if SCC through flow cytometry is not available, e.g., in farm measurements. Finally, progesterone could not be predicted with sufficient accuracy to be further considered. Further investigations are needed, especially to evaluate the performances of models when used in routine with real field data. Consequently, MIR spectrometry can provide valuable information on energy deficit, ketosis and mastitis in dairy cows, all of which influence fertility.

ACKNOWLEDGEMENT AND DISCLAIMER

The authors gratefully thank Petimat Kitaeva, Claire Darimont, Mohamed El Morabit, Matthieu Dubuisson, Quentin Vigneron, Lora Renault, Olivier Genard, Maxence Didelez, François Rouelle and Sophie Mathieux for their precious work.

OptiMIR project has received funding from INTERREG IVB NEW and the Walloon Region. GplusE project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n° 613689. The views expressed in this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission. IndiKuh project 'Evaluation of Animal Welfare in Dairy Farming—Indicators for the Metabolism and Feeding' (IndiKuh, funding code: 2817905815) was supported by funds of the German Federal Ministry of Food and Agriculture (BMEL). COMET-Project D4Dairy (Digitalisation, Data Integration, Detection and Decision Support in Dairying, project 872039) was supported by BMK, BMDW and the provinces of Lower Austria and Vienna in the framework of the COMET-Competence Centers for Excellent Technologies. The COMET program is handled by the FFG.

REFERENCES

Åkerstedt, M., L. Forsbck, T. Larsen, and K. Svennersten-Sjaunja. 2011. Natural variation in biomarkers indicating mastitis in healthy cows. J. Dairy Res. 78:88–96. https://doi.org/10.1017/S0022029910000786.

Akkerman, M., L. B. Larsen, J. Sørensen, and N. A. Poulsen. 2019. Natural variations of citrate and calcium in milk and their effects on milk processing properties. J. Dairy Sci. 102:6830–6841. https://doi.org/10.3168/jds.2018-16195.

Bellon-Maurel, V., E. Fernandez-Ahumada, B. Palagos, J. M. Roger, and A. McBratney. 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. Trends Analyt. Chem. 29:1073–1081. https://doi.org/10.1016/j.trac.2010.05.006.

Billa, P. A., Y. Faulconnier, T. Larsen, C. Leroux, and J. A. A. Pires. 2020. Milk metabolites as noninvasive indicators of nutritional status of mid-lactation Holstein and Montbéliarde cows. J. Dairy Sci. 103:3133–3146. https://doi.org/10.3168/jds.2019-17466.

Bjerre-Harpøth, V., N. C. Friggens, V. M. Thorup, T. Larsen, B. M. Damgaard, K. L. Ingvartsen, and K. M. Moyes. 2012a. Metabolic and production profiles of dairy cows in response to decreased nutrient density to increase physiological imbalance at different stages of lactation. J. Dairy Sci. 95:2362–2380. https://doi.org/10.3168/jds.2011-4419.

Bjerre-Harpøth, V., N. C. Friggens, V. M. Thorup, T. Larsen, B. M. Damgaard, K. L. Ingvartsen, and K. M. Moyes. 2012b. Metabolic and production profiles of dairy cows in response to decreased nutrient density to increase physiological imbalance at different stages of lactation. J. Dairy Sci. 95:2362–2380. https://doi.org/10.3168/jds.2011-4419.

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.

Brereton, R. G., and G. R. Lloyd. 2010. Support Vector Machines for classification and regression. Analyst (Lond.) 135:230–267. https://doi.org/10.1039/B918972F.

Bulman, D. C., and G. E. Lamming. 1978. Milk progesterone levels in relation to conception, repeat breeding and factors influencing acyclicity in dairy cows. Reproduction 54:447–458. https://doi.org/10.1530/jrf.0.0540447.

Chagunda, M. G. G., T. Larsen, M. Bjerring, and K. L. Ingvartsen. 2006. L-lactate dehydrogenase and N-acetyl-β-D-glucosaminidase activities in bovine milk as indicators of non-specific mastitis. J. Dairy Res. 73:431–440. https://doi.org/10.1017/S0022029906001956.

Crowe, M. A. 2008. Resumption of Ovarian Cyclicity in Post-partum Beef and Dairy Cows. Reprod. Domest. Anim. 43(s5):20–28. https://doi.org/10.1111/j.1439-0531.2008.01210.x.

Dallago, G.M., K.M. Wade, R.I. Cue, J.T. McClure, R. Lacroix, D. Pellerin, and E. Vasseur. 2021. Keeping Dairy Cows for Longer: A Critical Literature Review on Dairy Cow Longevity in High Milk-Producing Countries. Animals 2021, Vol. 11, Page 808 11:808. https://doi.org/10.3390/ani11030808.

Dardenne, P., D. Vincke, V. Baeten, and J. A. Fernandez Pierna. 2015. Untargeted multivariate characterization of contaminants in powdered milk. Pages 97–124 in Biotechnol. Agron. Soc. Environ. Biotechnol. Agron. Soc. Environ., Gembloux.

Davies, A., and T. Fearn. 2006. Back to basics: calibration statistics. Spectroscopy Europe 18:31–32.

De Roos, A. P. W., H. J. C. M. Van Den Bijgaart, J. Hørlyk, and G. De Jong. 2007. Screening for subclinical ketosis in dairy cattle by fourier transform infrared spectrometry. J. Dairy Sci. 90:1761–1766. https://doi.org/10.3168/jds.2006-203.

De Vries, A., and M. I. Marcondes. 2020. Review: Overview of factors affecting productive lifespan of dairy cows. Animal 14:s155–s164. https://doi.org/10.1017/S1751731119003264.

Dehareng, F., C. Delfosse, E. Froidmont, H. Soyeurt, C. Martin, N. Gengler, A. Vanlierde, and P. Dardenne. 2012. Potential use of milk mid-infrared spectra to predict individual methane emission of dairy cows. animal 6:1694–1701. https://doi.org/10.1017/S1751731112000456.

Denis-Robichaud, J., J. Dubuc, D. Lefebvre, and L. DesCôteaux. 2014. Accuracy of milk ketone bodies from flow-injection analysis for the diagnosis of hyperketonemia in dairy cows. J. Dairy Sci. 97:3364–3370. https://doi.org/10.3168/jds.2013-6744.

Enjalbert, F., M. C. Nicot, C. Bayourthe, and R. Moncoulon. 2001. Ketone bodies in milk and blood of dairy cows: Relationship between concentrations and utilization for detection of subclinical ketosis. J. Dairy Sci. 84:583–589. https://doi.org/10.3168/jds.S0022-0302(01)74511-0.

Esposito, G., P. C. Irons, E. C. Webb, and A. Chapwanya. 2014. Interactions between negative energy balance, metabolic diseases, uterine health and immune response in transition dairy cows. Anim. Reprod. Sci. 144:60–71. https://doi.org/10.1016/j.anireprosci.2013.11.007.

Foldager, L., C. Gaillard, M. T. Sorensen, T. Larsen, E. Matthews, R. O'Flaherty, F. Carter, M. A. Crowe, C. Grelet, M. Salavati, M. Hostens, K. L. Ingvartsen, M. A. Krogh, N. McLoughlin, A. Fahey, A. Santoro, C. Byrne, P. Rudd, S. Hallinan, C. Wathes, Z. Cheng, A. Fouladi, G. Pollott, D. Werling, B. Sanz Bernardo, C. Ferris, A. Wylie, M. Bell, M. Vaneetvelde, K. Hermans, G. Opsomer, S. Moerman, J. De Koster, H. Bogaert, J. Vandepitte, L. Vandevelde, B. Vanranst, J. Hoglund, S. Dahl, S. Ostergaard, J. Rothmann, E. Meyer, J. Ettema, T. Rousing, V. H. Silva De Oliveira, C. Marchitelli, F. Signorelli, F. Napolitano, B. M. Moioli, A. Crisà, L. Buttazzoni, J. McClure, D. Matthews, F. Kearney, A. Cromie, M. McClure, S. Zhang, X. Chen, H. Chen, J. Zhao, L. Yang, G. Hua, C. Tan, G. Wang, M. Bonneau, M. Sciarretta, A. Pearn, A. Evertson, L. Kosten, A. Fogh, T. Andersen, M. Lucy, C. Elsik, G. Conant, J. Taylor, D. Triant, N. Gengler, M. Georges, F. Colinet, M. Ramos Pamplona, H. Hammami, C. Bastin, H. Takeda, A. Laine, A. S. Van Laere, R. Mota, S. Naderi Darbagshahi, F. Dehareng, A. Vanlierde, E. Froidmont, F. Becker, M. Schulze, and S. Palma Vera. 2020. Predicting physiological imbalance in Holstein dairy cows by three different sets of milk biomarkers. Prev. Vet. Med. 179:105006. https://doi.org/10.1016/j.prevetmed.2020.105006.

Friggens, N. C., and M. G. G. Chagunda. 2005. Prediction of the reproductive status of cattle on the basis of milk progesterone measures: model description. Theriogenology 64:155–190. https://doi.org/10.1016/j.theriogenology.2004.11.014.

Garnsworthy, P. C., L. L. Masson, A. L. Lock, and T. T. Mottram. 2006. Variation of milk citrate with stage of lactation and de novo fatty acid synthesis in dairy cows. J. Dairy Sci. 89:1604–1612. https://doi.org/10.3168/jds.S0022-0302(06)72227-5.

Ginther, O. J., L. C. Nuti, M. C. Garcia, B. C. Wentworth, and W. J. Tyler. 1976. Factors Affecting Progesterone Concentration in Cow's Milk and Dairy Products. J. Anim. Sci. 42:155–159. https://doi.org/10.2527/jas1976.421155x.

Gohary, K., M. W. Overton, M. von Massow, S. J. LeBlanc, K. D. Lissemore, and T. F. Duffield. 2016. The cost of a case of subclinical ketosis in Canadian dairy herds. Can. Vet. J. 57:728.

Grelet, C., C. Bastin, M. Gelé, J. B. Davière, M. Johan, A. Werner, R. Reding, J. A. Fernandez Pierna, F. G. Colinet, P. Dardenne, N. Gengler, H. Soyeurt, and F. Dehareng. 2016. Development of Fourier transform mid-infrared calibrations to predict acetone, β-hydroxybutyrate, and citrate contents in bovine milk through a European dairy network. J. Dairy Sci. 99:4816–4825. https://doi.org/10.3168/jds.2015-10477.

Grelet, C., P. Dardenne, H. Soyeurt, J. A. Fernandez, A. Vanlierde, F. Stevens, N. Gengler, and F. Dehareng. 2021. Large-scale phenotyping in dairy sector using milk MIR spectra: Key factors affecting the quality of predictions. Methods 186:97–111. https://doi.org/10.1016/j.ymeth.2020.07.012.

Grelet, C., J. A. Fernández Pierna, P. Dardenne, V. Baeten, and F. Dehareng. 2015. Standardization of milk mid-infrared spectra from a European dairy network. J. Dairy Sci. 98:2150–2160. https://doi.org/10.3168/jds.2014-8764.

Heuer, C., H. J. Luinge, E. T. G. Lutz, Y. H. Schukken, J. H. Van Der Maas, H. Wilmink, and J. P. T. M. Noordhuizen. 2001. Determination of Acetone in Cow Milk by Fourier Transform Infrared Spectroscopy for the Detection of Subclinical Ketosis. J. Dairy Sci. 84:575–582. https://doi.org/10.3168/jds.S0022-0302(01)74510-9.

Hovinen, M., H. Simojoki, R. Pösö, J. Suolaniemi, P. Kalmus, L. Suojala, and S. Pyörälä. 2016. N-acetyl -β-D-glucosaminidase activity in cow milk as an indicator of mastitis. J. Dairy Res. 83:219–227. https://doi.org/10.1017/S0022029916000224.

Hyvönen, P., T. Haarahiltunen, T. Lehtolainen, J. Heikkinen, R. Isomäki, and S. Pyörälä. 2010. Concentrations of bovine lactoferrin and citrate in milk during experimental endotoxin mastitis in early- versus late-lactating dairy cows. J. Dairy Res. 77:474–480. https://doi.org/10.1017/S0022029910000579.

ICAR Dairy Cattle Milk Recording Working Group. 2017. Section 2 -Guidelines for Dairy Cattle Milk Recording (https://www.icar.org/Guidelines/02-Overview-Cattle-Milk-Recording.pdf).

Ingvartsen, K. L. 2006. Feeding- and management-related diseases in the transition cow: Physiological adaptations around calving and strategies to reduce feeding-related diseases. Anim. Feed Sci. Technol. 126:175–213. https://doi.org/10.1016/j.anifeedsci.2005.08.003.

Kostensalo, J., M. Lidauer, B. Aernouts, P. Mäntysaari, T. Kokkonen, P. Lidauer, and T. Mehtiö. 2023. Short communication: Predicting blood plasma non-esterified fatty acid and beta-hydroxybutyrate concentrations from cow milk—addressing systematic issues in modelling. animal 100912. https://doi.org/10.1016/j.animal.2023.100912.

Krogh, M.A., M. Hostens, M. Salavati, C. Grelet, M.T. Sorensen, D.C. Wathes, C.P. Ferris, C. Marchitelli, F. Signorelli, F. Napolitano, F. Becker, T. Larsen, E. Matthews, F. Carter, A. Vanlierde, G. Opsomer, N. Gengler, F. Dehareng, M.A. Crowe, K.L. Ingvartsen, and L. Foldager. 2020. Between- and within-herd variation in blood and milk biomarkers in Holstein cows in early lactation. animal 14:1067–1075. https://doi.org/10.1017/S1751731119002659.

Larsen, T. 2005. Determination of lactate dehydrogenase (LDH) activity in milk by a fluorometric assay. J. Dairy Res. 72:209–216. https://doi.org/10.1017/S0022029905000865.

Larsen, T. 2014a. Fluorometric determination of free and total isocitrate in bovine milk. J. Dairy Sci. 97:7498–7504. https://doi.org/10.3168/jds.2014-8018.

Larsen, T. 2014b. Fluorometric determination of free and total isocitrate in bovine milk. J. Dairy Sci. 97:7498–7504. https://doi.org/10.3168/jds.2014-8018.

Larsen, T. 2015. Fluorometric determination of free glucose and glucose 6-phosphate in cows' milk and other opaque matrices. Food Chem. 166:283–286. https://doi.org/10.1016/j.foodchem.2014.06.017.

Larsen, T., and K.M. Moyes. 2015. Are free glucose and glucose-6-phosphate in milk indicators of specific physiological states in the cow?. animal 9:86–93. https://doi.org/10.1017/S1751731114002043.

Larsen, T., and N. I. Nielsen. 2005. Fluorometric Determination of β-Hydroxybutyrate in Milk and Blood Plasma. J. Dairy Sci. 88:2004–2009. https://doi.org/10.3168/jds.S0022-0302(05)72876-9.

Larsen, T., C.M. Rntved, K.L. Ingvartsen, L. Vels, and M. Bjerring. 2010. Enzyme activity and acute phase proteins in milk utilized

as indicators of acute clinical E. coli LPS-induced mastitis. animal 4:1672–1679. https://doi.org/10.1017/S1751731110000947.

Leblanc, S. 2010. Monitoring Metabolic Health of Dairy Cattle in the Transition Period. J. Reprod. Dev. 56(S):S29–S35. https://doi.org/10.1262/jrd.1056S29.

Leroy, J. L. M. R., G. Opsomer, A. van Soom, I. G. F. Goovaerts, and P. E. J. Bols. 2008. Reduced Fertility in High-yielding Dairy Cows: Are the Oocyte and Embryo in Danger? Part I </br>The Importance of Negative Energy Balance and Altered Corpus Luteum Function to the Reduction of Oocyte and Embryo Quality in High-yielding Dairy Cows*. Reprod. Domest. Anim. 43:612–622. https://doi.org/10.1111/j.1439-0531.2007.00960.x.

Lou, W., H. Zhang, H. Luo, Z. Chen, R. Shi, X. Guo, Y. Zou, L. Liu, L. F. Brito, G. Guo, and Y. Wang. 2022. Genetic analyses of blood β-hydroxybutyrate predicted from milk infrared spectra and its association with longevity and female reproductive traits in Holstein cattle. J. Dairy Sci. 105:3269–3281. https://doi.org/10.3168/jds.2021-20389.

McArt, J. A. A., D. Nydam, and G. R. Oetzel. 2012. Epidemiology of subclinical ketosis in early lactation dairy cattle. J. Dairy Sci. 95:5056–5066. https://doi.org/10.3168/jds.2012-5443.

McArt, J. A. A., D. Nydam, and M. W. Overton. 2015. Hyperketonemia in early lactation dairy cattle: A deterministic estimate of component and total cost per case. J. Dairy Sci. 98:2043–2054. https://doi.org/10.3168/jds.2014-8740.

McParland, S., G. Banos, E. Wall, M. P. Coffey, H. Soyeurt, R. F. Veerkamp, and D. P. Berry. 2011. The use of mid-infrared spectrometry to predict body energy status of Holstein cows. J. Dairy Sci. 94:3651–3661. https://doi.org/10.3168/jds.2010-3965.

McParland, S., E. Kennedy, E. Lewis, S. G. Moore, B. McCarthy, M. O'Donovan, and D. P. Berry. 2015. Genetic parameters of dairy cow energy intake and body energy status predicted using mid-infrared spectrometry of milk. J. Dairy Sci. 98:1310–1320. https://doi.org/10.3168/jds.2014-8892.

Moyes, K. M., T. Larsen, and K. L. Ingvartsen. 2013. Generation of an index for physiological imbalance and its use as a predictor of primary disease in dairy cows during early lactation. J. Dairy Sci. 96:2161–2170. https://doi.org/10.3168/jds.2012-5646.

Pires, J. A. A., T. Larsen, and C. Leroux. 2022. Milk metabolites and fatty acids as noninvasive biomarkers of metabolic status and energy balance in early-lactation cows. J. Dairy Sci. 105:201–220. https://doi.org/10.3168/jds.2021-20465.

Puerto, M. A., E. Shepley, R. I. Cue, D. Warner, J. Dubuc, and E. Vasseur. 2021. The hidden cost of disease: I. Impact of the first incidence of mastitis on production and economic indicators of primiparous dairy cows. J. Dairy Sci. 104:7932–7943. https://doi.org/10.3168/jds.2020-19584.

Riekerink, R. G. M. O., H. W. Barkema, D. F. Kelton, and D. T. Scholl. 2008. Incidence Rate of Clinical Mastitis on Canadian Dairy Farms. J. Dairy Sci. 91:1366–1377. https://doi.org/10.3168/jds.2007-0757.

Roche, J. R., C. R. Burke, S. Meier, C. G. Walker, J. R. Roche, C. R. Burke, S. Meier, and C. G. Walker. 2011. Nutrition × reproduction interaction in pasture-based systems: is nutrition a factor in reproductive failure? Anim. Prod. Sci. 51:1045–1066. https://doi.org/10.1071/AN10162.

Roelofs, J. B., F. J. C. M. van Eerdenburg, W. Hazeleger, N. M. Soede, and B. Kemp. 2006. Relationship between progesterone concentrations in milk and blood and time of ovulation in dairy cattle. Anim. Reprod. Sci. 91:337–343. https://doi.org/10.1016/j.anireprosci.2005.04.015.

Soyeurt, H., C. Grelet, S. McParland, M. Calmels, M. Coffey, A. Tedde, P. Delhez, F. Dehareng, and N. Gengler. 2020. A comparison of 4 different machine learning algorithms to predict lactoferrin content in bovine milk from mid-infrared spectra. J. Dairy Sci. 103:11585–11596. https://doi.org/10.3168/jds.2020-18870.

Wathes, D. C. 2012. Mechanisms Linking Metabolic Status and Disease with Reproductive Outcome in the Dairy Cow. Reprod. Domest. Anim. 47(s4):304–312. https://doi.org/10.1111/j.1439-0531.2012.02090.x.

Zhang, L., C. Li, F. Dehareng, C. Grelet, F. Colinet, N. Gengler, Y. Brostaux, and H. Soyeurt. 2021. Appropriate Data Quality Checks Improve the Reliability of Values Predicted from Milk Mid-Infrared Spectra. Animals 2021, Vol. 11, Page 533 11:533. https://doi.org/10.3390/ani11020533.