



Development of a subacute ruminal acidosis risk score and its prediction using milk mid-infrared spectra in early-lactation cows

A. Mensching,^{1,2*} M. Zschiesche,³ J. Hummel,³ C. Grelet,⁴ N. Gengler,⁵ S. Dänicke,⁶ and A. R. Sharifi^{1,2}

¹Animal Breeding and Genetics Group, Department of Animal Sciences, University of Goettingen, 37075 Goettingen, Germany

²Center for Integrated Breeding Research, University of Goettingen, 37075 Goettingen, Germany

³Ruminant Nutrition Group, Department of Animal Sciences, University of Goettingen, 37077 Goettingen, Germany

⁴Walloon Agricultural Research Center, Knowledge and Valorization of Agricultural Products Department, 5030 Gembloux, Belgium

⁵TERRA Research and Training Center, Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium

⁶Institute of Animal Nutrition, Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, 38116 Brunswick, Germany

ABSTRACT

A routine monitoring for subacute ruminal acidosis (SARA) on the individual level could support the minimization of economic losses and the ensuring of animal welfare in dairy cows. The objectives of this study were (1) to develop a SARA risk score (SRS) by combining information from different data acquisition systems to generate an integrative indicator trait, (2) the investigation of associations of the SRS with feed analysis data, blood characteristics, performance data, and milk composition, including the fatty acid (FA) profile, (3) the development of a milk mid-infrared (MIR) spectra-based prediction equation for this novel reference trait SRS, and (4) its application to an external data set consisting of MIR data of test day records to investigate the association between the MIR-based predictions of the SRS and the milk FA profile. The primary data set, which was used for the objectives (1) to (3), consisted of data collected from 10 commercial farms with a total of 100 Holstein cows in early lactation. The data comprised barn climate parameters, pH and temperature logging from intrarecticular measurement boluses, as well as jaw movement and locomotion behavior recordings of noseband-sensor halters and pedometers. Further sampling and data collection included feed samples, blood samples, milk performance, and milk samples, whereof the latter were used to get the milk MIR spectra and to estimate the main milk components, the milk FA profile, and the lactoferrin content. Because all measurements were characterized by different temporal resolutions, the data preparation consisted of an aggregation into values on a daily basis and merging it into one data set. For the development of the SRS, a total of 7 traits were selected, which were derived from measurements of pH and temperature in

the reticulum, chewing behavior, and milk yield. After adjustment for fixed effects and standardization, these 7 traits were combined into the SRS using a linear combination and directional weights based on current knowledge derived from literature studies. The secondary data set was used for objective (4) and consisted of test day records of the entire herds, including performance data, milk MIR spectra and MIR-predicted FA. At farm level, it could be shown that diets with higher proportions of concentrated feed resulted in both lower daily mean pH and higher SRS values. On the individual level, an increased SRS could be associated with a modified FA profile (e.g., lower levels of short- and medium-chain FA, higher levels of C17:0, odd- and branched-chain FA). Furthermore, a milk MIR-based partial least squares regression model with a moderate predictability was established for the SRS. This work provides the basis for the development of routine SARA monitoring and demonstrates the high potential of milk composition-based assessment of the health status of lactating cows.

Key words: SARA, milk mid-infrared spectra, partial least squares regression, indicator trait

INTRODUCTION

Technological progress has facilitated collection, management, and use of “big data,” providing support for the development of innovative and complex functional traits, opening new possibilities in animal research and commercial farming. Particularly an improved monitoring for early detection of subclinical disorders such as SARA could facilitate timely interventions and thus prevent clinical diseases in dairy farming. In this regard, sensor-based data acquisition provides the basis for monitoring systems that can be superior to detection by human sensory perception. Because signs of a disease are generally reflected in several measurable traits, the combination of different data acquisition systems, in particular, can offer the basis for innova-

Received August 24, 2020.

Accepted November 10, 2020.

*Corresponding author: andre.mensching@uni-goettingen.de

tive concepts. For sensor-based lameness detection, for example, Van Hertem et al. (2013) used the neck and rumination activity recorded with a heat detection system and combined it with information on the milk performance and achieved a cross-validated sensitivity of 0.89 and specificity of 0.85. However, the situation is different with regard to SARA, as, by definition of the subacute condition, no clear clinical signs on the individual level are present, which makes diagnosis very difficult (Enemark, 2008). In this respect, the integrative combination of different information sources from several sensor systems could therefore be useful and theoretically should provide a better basis for the assessment of the health status.

Dairy herds receiving diets high in concentrate and low in physically effective fiber tend to show various clinical signs of related diseases (Kleen et al., 2003). These include reduced feed intake, lower milk production efficiency, altered fecal consistency, increased frequency of claw diseases, and an overall higher culling rate. Some of these signs are per construction herd-based parameters, and their reliability is based essentially on statistical aggregation of data from the entire herd. However, a routine monitoring system at individual level would be preferable, as this would both improve management and contribute to ensuring animal welfare, and, in addition, the obtained data would be of great value for animal breeding. Current approaches already rely on sensor-based continuous intrarecticular pH measurements (e.g., Denwood et al., 2018). Besides SARA, Stangaferro et al. (2016), for example, showed that rumination behavior provides a high potential for individual animal monitoring for ketosis and displaced abomasum.

However, for the monitoring of SARA, one major challenge is that a high individual animal variance concerning susceptibility to SARA has been observed (Humer et al., 2015). Consequently, a need exists to identify innovative indicators such as biomarkers, which provide the highest possible degree of association with SARA on the individual level and which are at best applicable on a large scale in commercial farming.

In the last decade, several studies demonstrated that milk mid-infrared (MIR) analysis, which is routinely used to determine standard milk components relevant for milk payment (e.g., fat content), can provide further relevant information for both breeding and management purposes (Gengler et al., 2016). Using MIR spectroscopy, it is possible to differentiate the milk components even further, such as breaking down the fat fraction into its fatty acid (FA) composition (Soyeurt et al., 2011). Recent approaches include the MIR-based prediction of more complex traits of animal associated characteristics such as the energetic or metabolic

status (De Koster et al., 2019) or methane emission (Vanlierde et al., 2018). Because the milk FA profile is associated with ruminal fermentation (Vlaeminck et al., 2006; Jing et al., 2018) and energy metabolism (Gross et al., 2011), it can be suggested that MIR-based models have a high potential for predicting and optimizing the monitoring of functional traits, which is relevant in terms of animal welfare and economy. Once an adequate MIR-based prediction model for health-related traits is developed, the main advantage will be that it can be applied to test day data collected during the routine recording and subsequently could be used in commercial farming. A drawback when working with MIR spectroscopy is that intra- and interinstrument variability occurs, and that predictions are unstable over time (Gengler et al., 2016). During routine application, post-prediction corrections are therefore carried out for the standard milk components (fat, protein) using standardized reference samples. However, this is not possible for all other traits that are not covered by the standardized reference samples for post-prediction correction. To solve this problem, Grelet et al. (2015) developed a method to standardize the spectra across different devices and manufacturers, while at the same time eliminating the temporal instability.

Based on data collected under field conditions, the objectives of this study were (1) to develop an integrative indicator trait for detecting SARA by combining information from different data collection systems into one SARA risk score (SRS). To evaluate this novel trait, its association with feed composition, blood metabolites, performance data, and fine milk composition was examined (2). Furthermore, a milk MIR spectra-based prediction equation of this novel reference trait SRS was developed (3) and applied to an external data set consisting of test day records to study the association between MIR-predicted SRS and the FA profile of the milk (4).

MATERIALS AND METHODS

The experiment was conducted in accordance with the German legislation on animal protection (Animal Welfare Act) and was approved by the Lower Saxony State Office for Consumer Protection and Food Safety (LAVES, Oldenburg, Germany; AZ: 33.9-42502-05-17A106).

Data Recording

The data of this study consisted of a primary and a secondary data set. The primary data set was used for the development of the SRS and the milk MIR-based partial least squares (PLS) regression model. The sec-

ondary data set served to investigate the associations between milk MIR-predicted SRS and the milk composition under field conditions.

Primary Data Set

Following a standardized sampling scheme (see Figure 1), data recording was carried out sequentially in 10 commercial farms on a total of 100 Holstein cows in the northwest of Lower Saxony, Germany, between April 2017 and March 2018. Compared with an average 305-d milk yield of 8,500 kg and 63 cows per farm in Germany in 2016 (Lindena et al., 2018), the 10 farms can be classified as above average in terms of performance and size, with 305-d milk yields between 9,200 and 11,100 kg and herd sizes of approx. 200 to 600 cows. All farms had a loose-housing outdoor climate barn with resting pens, fed a partial mixed ration (PMR) or TMR, and milked the cows twice a day. Per farm, 10 cows were selected so that the theoretical calving date was as close as possible and at most 2 wk before the actual data collection period. Additionally, we attempted to ensure that all parities of the early-lactation cows (1, 2, 3, and ≥ 4) were covered equally. On each farm, the 2-wk data collection during early lactation (15 ± 6 DIM on d 8 in each farm; see Figure 1) was preceded by a 1-wk preparation and adaptation phase, allowing the cows to get used to the measuring equipment.

In each farm, feed samples were collected during the data collection period on d 8, 10, 15, and 17 (see Figure 1). Samples of equal size were taken immediately after the distribution of PMR or TMR at 3 different positions in the trough (beginning, middle, end) and then were pooled and mixed. At first, measurements with a Penn State Particle Separator (Nasco, Fort Atkinson,

WI) were performed on subsamples of the pooled and mixed feed samples to determine the physically effective NDF > 8 mm (**peNDF** > 8) of the PMR and TMR, defined as the proportion of material retained on an 8-mm sieve multiplied by the NDF content of the diet. The corresponding procedure was performed by the same person throughout the experiment. Then, additional subsamples of the pooled and mixed feed samples were vacuum-packed and stored at -20°C until further analysis. Concentrated feed samples were also taken at the farms where PMR was fed. The analyses were carried out according to the methods of the Association of German Agricultural Analytic and Research Institute (VDLUFA, 1993), and the fiber analysis was performed using the filter bag system (Ankom Technology, Macedon, NY). The analysis was conducted in the laboratory of the Ruminant Nutrition Group at the University of Goettingen (Goettingen, Germany). Further information on the specific analytical methods of each chemical component can be found in Supplemental Table S1 (<http://dx.doi.org/10.6084/m9.figshare.13669966>). An overview of the composition of the diets is given in Supplemental Table S2. The weight of the distributed feed as well as the leftovers was also recorded to determine the average DMI of the respective feeding group. Furthermore, the quantities of distributed concentrate via transponder stations were recorded in farms with PMR feeding.

On each farm, 3 to 6 Tinytag climate data loggers (Tinytag Plus 2 TGP-4500, Gemini Data Loggers, Chichester, UK) were used to record climate data in the barn. Thereby, both the temperature and the relative air humidity were recorded at a height of about 2 m in the feeding, lying, and drinking areas, with measurements taken every 15 min.

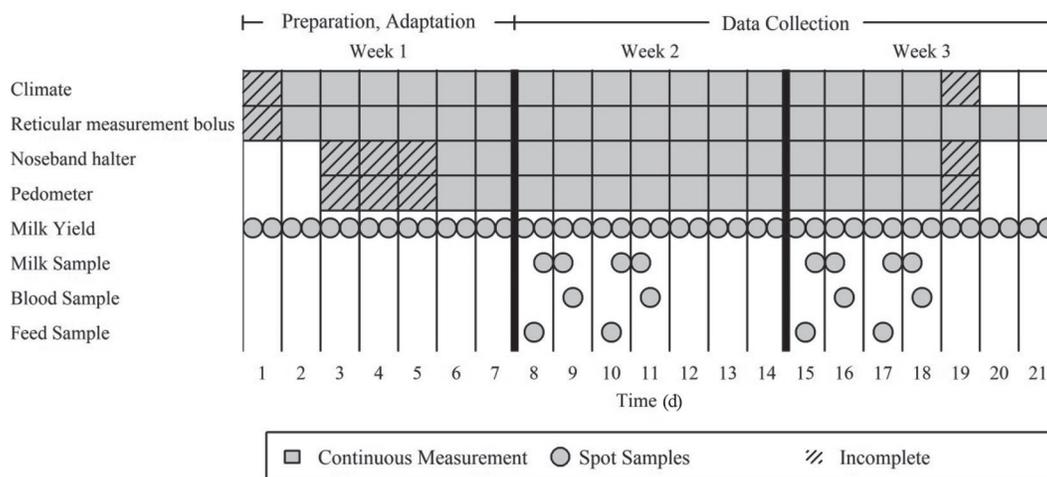


Figure 1. Data collection scheme, used sequentially in each of the 10 farms (adapted from Mensching et al., 2020c).

All 100 cows were equipped with a measurement bolus (eCow Ltd., Exeter, Devon, UK) for the continuous intrareticular measurement of pH and temperature. The boluses were inserted in each of the 10 farms on the first day of the preparation and adaptation phase, so that there were a maximum of 21 d between the insertion and the last day of data collection. Before insertion, all boluses were activated in a water bath at 38.4°C and calibrated in buffer solutions with pH 4 and pH 7. The raw data were provided as mean values of 15-min time intervals for both the reticular temperature and pH.

Jaw movement and locomotion behaviors were recorded using noseband-sensor halters and pedometers (RumiWatch, ITIN + HOCH GmbH, Liestal, Switzerland). The latter were attached to a rear leg. The raw data were converted using the RumiWatch Converter v. 0.7.4.13 (RumiWatch, ITIN + HOCH GmbH) into a 1-h resolution.

Milk yield was measured via on-farm milk quantity recording systems every day at each milking. During the data collection period, milk samples were taken on the evenings of d 8, 10, 15, and 17 and each subsequent morning (see Figure 1). The samples were taken separately and proportionally to the milk yield according to ICAR guidelines (ICAR, 2017) and were preserved with Bronopol (Georg Hansen e.K., Wrestedt, Germany). Morning and evening samples were analyzed separately by the Landeskontrollverband Weser-Ems e.V. in Leer, Germany, on the same spectrometer (MilkoScan FT+, Foss, Hillerød, Denmark). The standard components (fat, protein, lactose, urea) were estimated using the MIR-based prediction models provided by the manufacturer of the spectrometer. Somatic cell count was measured using a flow cytometer (Fossomatic FC, Foss) combined with the spectrometer. In addition to the standard milk components the MIR spectra were provided (1,060 wavenumbers per sample). The MIR spectra consisting of transmittance values (T) were converted into absorbance values (A) using a logarithm function with $A = -\log_{10}(T)$, as is usually done in spectrometry (Gengler et al., 2016). In the remainder of this paper, only absorbance spectra are considered. The raw spectral data were standardized according to Grellet et al. (2015). Afterward, 27 further traits containing information on the fine milk composition were predicted using models based on the studies of Soyeurt et al. (2007, 2011, 2012) as well as on further unpublished work from the Gembloux Agro-Bio Tech, University of Liège (GxABT, Gembloux, Belgium) and the Walloon Agricultural Research Center (CRA-W, Gembloux, Belgium). Included were predictions of short-chain FA (chain length of 4 to 10 carbons), medium-chain FA (chain length of 12 to 16 carbons), long-chain FA

(LCFA, chain length of 17 to 18 carbons), and milk lactoferrin content (LF). Only the predictions on FA with 15 carbon atoms within the medium-chain FA group are from the unpublished work, whereas all other predictions are based on published results. The applied MIR-based PLS regression models for prediction of FA related traits are characterized by a coefficient of determination of cross-validation of $0.46 \leq R_{cv}^2 \leq 0.99$, a ratio between standard deviation (SD) of the trait to the root mean squared error of cross-validation ($RMSE_{cv}$) of $1.4 \leq RPD_{cv} \leq 10.0$, and a coefficient of determination of validation of $0.39 \leq R_{val}^2 \leq 0.98$. For LF, the model statistics are $R_{cv}^2 = 0.72$, $RPD_{cv} = 1.86$, and $R_{val}^2 = 0.60$.

During the data collection period, blood samples were collected from the vena caudalis mediana on d 9, 11, 16, and 18, up to a maximum of 3 h after morning milking (see Figure 1). Separate tubes with serum clot activator were used, one for BHB and nonesterified FA (NEFA) in serum and another one for glucose determination. The analysis was conducted in the laboratory of the Institute of Veterinary Medicine at the University of Goettingen (Goettingen, Germany). See Supplemental Table S1 (<http://dx.doi.org/10.6084/m9.figshare.13669966>) for further details.

Secondary Data Set

The data set for model application consists of the monthly milk test day recordings of the entire herds of the same farms and was gathered for 1 yr from the beginning of the primary data collection described previously. Only data from farms that perform the A4 milk testing scheme according to the ICAR guidelines (ICAR, 2017) with proportional milk test day sampling consisting of evening and morning milk samples were considered. This was the case for 9 out of 10 farms. The pooled samples were analyzed on the same spectrometer as for the primary data set. The resulting secondary data set comprised 21,024 spectra from a total of 3,381 animals up to 305 DIM. In addition to the determination of the standard milk components, the spectra were standardized and the same predictions were made as for the primary data set.

The temperature in the barn was not recorded during collection of the secondary data. However, because it is needed in the further course of this work for correction purposes of fixed environmental effects, it was estimated using a linear model [coefficient of determination (R^2) = 0.99, residual SD = 0.67°C] that was developed by Mensching et al. (2020c) and established on the primary data set. In this model, external climate

data from a nearby weather station of the German Weather Service (DWD Climate Data Center, 2018), and estimated farm effects were used to predict the temperature in the barn.

Data Preparation

Primary Data. To investigate the association between the feed analysis data and the other traits at farm level, the feed analysis data were calculated using the median per trait and farm. For farms with supplementary feeding of concentrates via transponder stations, the composition of the total diet was estimated proportionally to the herd-based average DMI and the average amount of concentrates. In farms with PMR feeding, it was assumed that the peNDF > 8 of the entire diet is traced back only to the PMR and not to the supplemented concentrate feed. Considering the average herd DMI of PMR and concentrate feed, the peNDF > 8 of the entire diet was estimated by multiplying the share of PMR of the entire DMI with the calculated peNDF > 8 content of PMR. See Supplemental Table S3 (<http://dx.doi.org/10.6084/m9.figshare.13669966>) for the resulting estimated chemical compositions of the entire diets.

The other data recordings were based on a variety of different data collection systems and methods. They differed particularly in their temporal resolution. Therefore, a data aggregation and a subsequent merging were required. Because the final models are intended to be applicable to proportional sampling test day records, all data were aggregated on a daily basis.

The continuously recorded temperatures of the barn were aggregated by calculating daily arithmetic mean values ($\bar{s.T}$). The daily reticular mean pH ($\bar{r.pH}$) and the daily reticular pH range ($r.\Delta pH$) were derived from the continuously recorded reticular pH data. Because the reticular temperature is significantly influenced by water intake (Liang et al., 2013), the median ($r.T_{med}$) was used as a robust statistic for the aggregation. An arithmetic mean, however, would systematically underestimate the daily average reticular temperature by about 0.25°C. The jaw movement data in hourly resolution with information on eating and rumination duration per hour were aggregated by calculating the sum of the daily rumination and eating duration ($\Sigma.Rt$ and $\Sigma.Et$). In addition, the daily mean rumination chewing frequency (\overline{RCf}) per minute was determined by dividing the sum of daily rumination chews by the daily rumination duration. Based on the hourly pedometer data, the daily lying duration ($\Sigma.Lt$) was calculated by summation. The milk yield was calculated

from the sum of evening and morning milking according to the regular test day sampling procedure. Further, the MIR-predicted milk traits and MIR spectra were pooled proportionally to the morning and evening milk yield. The average milk yields, pooled milk traits, and MIR spectra were assigned to the day of the evening milking sample. The blood values were allocated to the previous day. This approach is also supported by the results of Loor et al. (2007), who observed that NEFA, BHB, and glucose levels responded with a 1-d delay to a feed restriction. Thus, the aim was to compensate for the time delay caused by the ruminal fermentation as well as by the actual digestion, to ensure a realistic time assignment of the individually measured variables.

To normalize the distributions of the BHB and NEFA values of the blood as well as the SCC in the milk, the former were log10 transformed (e.g., Grelet et al., 2019), and the latter were transformed by calculation of SCS according to Wiggans and Shook (1987). The FA traits predicted by standardized MIR spectra were considered as proportion of the total milk fat.

Data acquisitions with automated recording systems and sensitive measuring instruments in difficult environments such as agricultural practice are prone to failure and thus have to be investigated with caution. Therefore, a multivariate plausibility assessment according to Mensching et al. (2020c) was applied to the data set, which was aggregated on a daily basis. In this procedure, the observations of all traits were classified as “physiologically normal,” “physiologically extreme,” or “implausible,” considering various simultaneously recorded data. Observations classified as implausible were set to missing in the data set. This concerned sensor, performance, blood, and milk MIR spectral data as well as milk MIR-based predictions. However, if it was noticed that extreme spectra were accompanied by extreme observations in other traits, they were not removed from the data set. Observations of 3 animals were removed even before the plausibility of the data was assessed, as they showed unspecific clinical signs and thus were treated by the herd manager with medications, which might affect rumen fermentation. See Mensching et al. (2020c) for more details on this procedure.

The MIR-based prediction equations for the FA traits and lactoferrin were developed in a setting that covered large variability across multiple breeds, production systems, and countries. However, as outliers can always be expected due to variability that is not yet covered but is present in new MIR data, precautions were taken in an additional step using a statistical approach. Not only potential preposterous predictions of the MIR-predicted FA traits and the lactoferrin (values <0) but also extreme values (0.5% of lowest and highest pre-

dictions) were filtered out by setting them to missing before associations with the SRS were investigated.

Secondary Data. In the secondary data set, the MIR-predicted FA traits and the lactoferrin were filtered in the same way as in the case of the primary data set. Because the model development based on the primary data set only covers the early-lactation stage ($5 \leq \text{DIM} \leq 37$), the secondary data set was reduced to the same lactation period for PLS model application purposes, with 2,444 remaining observations of 2,254 cows.

SRS: Underlying Concept

For the development of the SRS, different sources of information, or, rather, traits, were combined to generate an integrative SARA-associated phenotype that best reflects the risk of suffering from SARA. It is expected that this will create a trait that is more robust and more accurate for displaying SARA than individual traits alone. A total of 7 different traits were selected, which directly or indirectly can be attributed to SARA. The selection of the traits was based on current knowledge of science derived from the literature as well as on new findings from subsequent data analysis, as follows:

- By definition, a low ruminal pH value and the corresponding acidotic conditions in the rumen are indicators of SARA (e.g., Nocek, 1997). Therefore, a low $\overline{\text{r.pH}}$ reflecting the ruminal pH should increase the SRS.
- As demonstrated by Mensching et al. (2020a), the ingestion behavior in the course of the day is reflected by the pH development. It could be deduced that a uniform ingestion behavior is associated with a limited pH range. However, irregular feed intake is considered a sign of SARA (e.g., Enemark, 2008). This is in line with the fact that the daily average pH and the pH fluctuation range are negatively correlated (Mensching et al., 2020b). For these reasons, high $\text{r.}\Delta\text{pH}$ should increase the SRS.
- The ruminal temperature is negatively correlated with pH (AlZahal et al., 2008) and is also increased by physiological stress (AlZahal et al., 2011). For this reason, elevated $\text{r.}T_{\text{med}}$ is supposed to increase the SRS.
- The daily rumination duration is associated with the amount of secreted saliva and thus the ability to buffer the ruminal pH (Beauchemin, 2018). For this reason, a reduced daily rumination duration ($\Sigma.\text{Rt}$) should increase the SRS.

- At herd level, a reduced DMI is a sign of SARA (Enemark, 2008). Reduced DMI is basically linked to a reduced eating duration and a reduced amount of secreted saliva during eating (Jiang et al., 2017). Therefore, a reduced daily eating duration ($\Sigma.\text{Et}$) is supposed to increase the SRS.
- In a preliminary analysis it was determined that from several chewing behavior traits, such as total chews, mean number of ruminant chews per bolus, or mean rumination chewing frequency, the $\overline{\text{RCf}}$ in particular had a significant positive association with the daily mean pH value. Oba and Allen (2000), for example, also observed that higher NDF contents are associated with a higher rumination chewing frequency. Derived from this, low $\overline{\text{RCf}}$ should increase the SRS.
- Khafipour et al. (2009a), for example, observed a decreased milk yield in a SARA induction experiment. Therefore, low milk yields should also be associated with the SRS, where a high SRS should be accompanied by low milk yields.

The directed associations with regard to SARA are visualized in Supplemental Figure S1 (<http://dx.doi.org/10.6084/m9.figshare.13669966>).

Statistical Implementation

The challenge was that the data are subject to a complex hierarchical structure and that factors such as lactation stage, parity, and barn temperature have significant effects on the 7 traits. In a first step, adjusted values were determined for each trait by correcting the raw observations for the influence of fixed effects. For each trait, the following linear mixed model (LMM) was used:

$$y_{ijklm} = \beta_0 + \beta_{1s} \cdot \overline{T}_{ijklm} + P_i + \beta_{2,i} \text{DIM}_{ijklm} \times P_i + F_j + F_j \times \text{TD}_k + A_1 + e_{ijklm}, \quad [1]$$

where y_{ijklm} is observation m of cow l in farm j on test day k for trait y , \overline{T}_{ijklm} is the daily mean temperature in the barn, P_i is the fixed effect of parity (1, 2, 3, or ≥ 4), and $\text{DIM}_{ijklm} \times P_i$ is the interaction of lactation stage DIM_{ijklm} and parity P_i ; β_0 is the intercept, and β_1 and $\beta_{2,i}$ are regression coefficients. The farm F_j , the interaction of farm and the test day $F_j \times \text{TD}_k$, and the animal A_1 are considered as normally distributed random effects with $F_j \sim N(0, \sigma_F)$, $F_j \times \text{TD}_k \sim N(0, \sigma_{F \times \text{TD}})$, and $A_1 \sim N(0, \sigma_A)$. Further, e_{ijklm} is a random error with $e_{ijklm} \sim N(0, \sigma_e)$. The lme4 package (Bates et al., 2015) was used for model fitting in R (R Core Team, 2019).

After estimating the fixed effects β_0 , β_1 , $\beta_{2,1}$, and P_i , the adjusted values y_{ijklm}^* were determined as follows:

$$\begin{aligned} y_{ijklm}^* &= y_{ijklm} - (\beta_0 + \beta_1 s \cdot \bar{T}_{ijklm} + P_i + \beta_{2,i} \text{DIM}_{ijklm} \times P_i) \\ &= F_j + F_j \times \text{TD}_k + A_l + e_{ijklm}. \end{aligned} \quad [2]$$

Consequently, the adjusted values contain only variance that is attributed to the random effects of the farm (e.g., feeding and management), test day, and animal, as well as the residuals. To standardize the scaling, a z-transformation was performed as follows:

$$z(y_{ijklm}^*) = \frac{y_{ijklm}^* - \mu}{\sigma}, \quad [3]$$

with μ being the arithmetic mean and σ the SD of y_{ijklm}^* .

Due to loss of data during the data collection and through the plausibility checks, $n = 357$ of theoretically 400 observations (100 cows with 4 repeated measurements) with MIR spectra remained. Thereof, only 280 showed complete cases for the 7 traits of interest. Because this additional data loss of about 22% of the observations is caused by only 7.6% of missing values, a multiple imputing procedure based on a principal component analysis (PCA) model was applied to impute missing data according to Josse and Husson (2016), to permit analysis of at least all 357 cases with available MIR data. This was done using the *MIPCA* function from the *missMDA* (Husson et al., 2019) package in R. In this procedure, 100 imputed data sets were created using the bootstrap method from which the final estimates were determined. The optimal number of principal components (PC) was determined via cross-validation using the *estim_ncPCA* function from the same package. The decision was made based on graphical examination of the mean squared error of prediction depending on the number of used PC.

In a final step, the SRS was calculated as a linear combination of the z-transformed and adjusted values $z(y_{ijklm}^*)$ and a directed weight w_π , as follows:

$$\text{SRS} = \sum_{\pi=1}^7 z(y_{ijklm}^*)_{\pi} \times w_{\pi}, \text{ with } w_{\pi} \in \{-1, 1\}. \quad [4]$$

Traits for which low values are supposed to cause a high SRS were weighted with $w_{\pi} = -1$ and the others with $w_{\pi} = 1$ (see Supplemental Figure S1, <http://dx.doi.org/10.6084/m9.figshare.13669966>). After calcula-

tion of the SRS, an LMM according to Equation [1] was set up to investigate the variance components of the random effects.

Association with Other Traits and Feed Data

To visualize the association between SRS and other reference traits, performance data, blood characteristics, and MIR-predicted traits, the observations were divided into 5 groups of equal size using the SRS, with the quintiles of the SRS as thresholds. Then group affiliation was tested for all traits with an LMM, according to Equation [1], with additional consideration of the SRS class as a fixed effect. The estimated group means (LSM) were compared with Tukey's post hoc test. For illustration purposes, the LSM of milk main components and MIR-predicted FA were plotted in a spider map, which was created with the *radarchart* function of the *fmsb* (Nakazawa, 2019) package in R.

The feed analysis data averaged per farm were examined in a PCA. For the SRS and the other 7 traits, LSM estimates corrected for DIM, parity, and temperature in the barn were determined for the 10 farms using an LMM according to Equation [1], considering the farm as fixed instead of random effect. Afterward, these data were projected into the space spanned by the first 2 PC of the feed data.

Development of MIR-Based Prediction Equations

Milk MIR spectra are characterized by very high correlations between the values of the individual number of wavenumbers. Therefore, PLS regression was used to develop the prediction equations for both the SRS and the available traits derived from the sensor data, most of which were also used for the construction of the SRS. The latter was done to assess how much the individual traits are associated with the milk MIR data. Regarding SARA, prediction equations for the 2 pH parameters r.pH and r. Δ pH are also of interest. An accurate MIR-based prediction of the traits would offer a major advantage over time-consuming and costly measurements, such as those using the eCow bolus or the RumiWatch halter, as it could be implemented cost-effectively and on a large scale in the existing milk recording system.

The supervised learning method PLS regression corresponds to a multiple regression in which the regression coefficients are estimated in an iterative procedure that is analogous to the PCA regression technique. What matters is that the design matrix \mathbf{X} (MIR spectra) is decomposed in a reduced amount of uncorrelated latent variables (LV), consisting of scores and loadings, in such a way that results in maximum covariance to

the response variable y (e.g., SRS). In a further step, a regression is performed on those reduced variables to predict y . The number of LV used to derive the calibration model affects model performances in terms of root mean squared error ($\mathbf{RMSE}_{\text{cal}}$) and coefficient of determination ($\mathbf{R}_{\text{cal}}^2$), and is a sensitive parameter to tune, as it may induce under- or overfitting of the model.

The number of LV were determined for each setting with an interleaved stratified 10-fold cross-validation, whereby the folds were divided in such a way that the repeated measurements of an animal were all situated into a common fold, to avoid a single cow having measurements in different folds. Then, for each trait, the model with the smallest $\mathbf{RMSE}_{\text{cv}}$ was selected, but the number of LV was limited to 15. In this work, the SIMPLS (de Jong, 1993) algorithm was used, applying the PLS (Mevik and Wehrens, 2007) package in R. The resulting model, which can be used for predictions based on new milk MIR data, can be described as

$$y_i = \beta_0 + \sum_{j=1}^n \beta_j x_{ij} + e_i, \quad [5]$$

where y_i is the response variable, β_0 is the estimated intercept, and β_1, \dots, β_n are the regression coefficients for the $1, \dots, n$ values x_{ij} of the respective wavenumber of the optionally pretreated and reduced spectrum. Further, e_i is a random error with $e_i \sim N(0, \sigma_e)$.

Numerous options for PLS modeling exist, of which the following settings have been systematically evaluated, starting with different spectral pretreatments:

- None
- 0/2/5, 1/2/5, 2/2/5 = Savitzky-Golay smoothing and first and second derivative with a polynomial of second degree and a window length of 5 wavenumbers (e.g., Wallén et al., 2018)
- Gap/1/5, gap/2/5 = first and second linear derivative with a gap of 5 wavenumbers (e.g., Soyeurt et al., 2011)
- LP(DIM,1) = Legendre polynomial transformation of the spectrum according to Vanlierde et al. (2015) up to the first degree to account linearly for the DIM

In addition, the part of the spectrum used in the PLS modeling was varied:

- Full spectrum with 1,060 wavenumbers
- Reduced spectrum with 516 wavenumbers, according to Grelet et al. (2015)

- Reduced spectrum with 212 wavenumbers, according to Grelet et al. (2016)

In each case, the pretreatment was carried out first and then the selection of wavenumbers was performed.

To get an objective view on the model performances, the primary data set was divided into a calibration and a validation data set. This was done individually for each trait, using data from 8 farms for model development and data from 2 farms for external validation. To select the farms, the mean values of the respective traits per farm were calculated and ordered. Observations from the second- and ninth-ordered farm were then selected for validation. This was done to take into account the hierarchical data structure and to ensure that both the calibration and validation data set cover, as far as possible, the range of the respective traits. To evaluate the models, common statistics were calculated. These include $\mathbf{RMSE}_{\text{cal}}$, $\mathbf{RMSE}_{\text{cv}}$, $\mathbf{RMSE}_{\text{val}}$, $\mathbf{R}_{\text{cal}}^2$, \mathbf{R}_{cv}^2 , $\mathbf{R}_{\text{val}}^2$, and \mathbf{RPD}_{cv} . For more information on these statistics, see Soyeurt et al. (2011) and Williams and Sobering (1993).

Application on the Secondary Data

To further validate the SRS prediction equation, the model was applied to the secondary data set. The milk MIR-predicted SRS values were assigned to an SRS class using the same quintile-based group boundaries as for the primary data set. Then the association between the milk MIR-predicted SRS class and the other MIR-predicted traits was examined, using the same LMM as for the primary data set. Again, a spider map was created to illustrate estimated LSM for the milk main components and the MIR-predicted FA, depending on the milk MIR-predicted SRS class.

All data preparation, creation of figures, and statistical analyses were performed with programs developed in the software environment R.

RESULTS AND DISCUSSION

Descriptive statistics of sensor data-derived traits, blood characteristics, performance, and milk composition-related traits can be found in the Supplemental Tables S4 and S5 (<http://dx.doi.org/10.6084/m9.figshare.13669966>).

SRS Development

By combining several traits into one integrative indicator trait, we accounted for the fact that signs of a disease are often reflected not only in one but in several

measurable traits. As the health state with regard to SARA can vary in a continuous range between physiologically healthy and clinically diseased (Enemark et al., 2002), the design of the SRS underlies a continuous scale. The traits used for its development were derived from 3 different independent data acquisition systems (intrareticular measurement bolus, noseband-sensor halter, and milk performance). Therefore, the SRS should reflect the health status of the animal more accurately than one trait alone. Alternative approaches are also possible, such as those of Grelet et al. (2019) and De Koster et al. (2019), who divided the metabolic status of cows by k-means clustering of 4 blood metabolites into 3 classes as new trait or phenotype, respectively, which was modeled using MIR data.

The PCA-based imputation procedure for the 7 adjusted and z-transformed traits solved the problem of missing values and maximized the number of observations for the downstream analysis. The optimum number of PC determined for imputing using the *estim_ncPCA* (Husson et al., 2019) function was 3. After imputation, a variable factor map (Figure 2) was generated, which shows the quality of the imputation as well as visualizing the relationships of the 7 traits. This variable factor map displays the loadings of the variables in the space of the first 2 PC, which together explain 48.81% of the variance. The points correspond to the loadings that were determined during the 100 imputations with the bootstrap method, whereas the thin arrows show the loadings of the final averaged data set. Because the point clouds around the arrowheads are only slightly scattered, the imputing quality can be classified as good enough to use those imputed values. As expected, the signs of the loadings of all traits of the first PC are in line with the directed weights, which were used for the development of the SRS. Additionally, the SRS was projected into the same space as a supplementary variable (bold arrow). The SRS lies almost completely along the first PC, so that it covers a large part of the observed variance of the 7 traits.

The design of the SRS presented here is only a first approach and can be further optimized and extended. So far, only the direction (-1 or 1) of the selected weights was defined for the SRS calculation. It is conceivable to individually adjust the weights while maximizing the association to a gold standard of the true SARA status—for instance, based on veterinary diagnostics or ventral ruminal pH measurements. However, because such data are not available, the possibilities in the present study are limited.

Analyzing the SRS with an LMM, the following variance components were estimated: $\sigma_F = 1.20$, $\sigma_{F \times TD} = 0.49$, $\sigma_A = 2.81$, and $\sigma_e = 2.02$. Considering the fact that all 10 farms are above average regarding perfor-

mance and the assumption of high interanimal variance (Humer et al., 2015; Jing et al., 2018), the variance components estimated here seem plausible and meet the expectations with regard to SARA.

Associations of the SRS with Other Traits and Feed Data

The relationship between the developed SRS and the other variables used for its development is illustrated in Figure 3. Furthermore, information on 8 additional variables is reported, including lying duration, 3 performance parameters (fat yield and protein yield), SCS, and milk LF content, as well as 3 blood characteristics (BHB, NEFA, and glucose). Shown are the LSM of each variable depending on the quintile-based SRS class, corrected for the fixed effects of the temperature in the barn, the parity, and the interaction of DIM and parity. The figure shows that the weighting of each variable used to construct the SRS is reflected in the estimated LSM with significant ($P < 0.05$) and directed differences along the means of quintile-based SRS classes, except for r.pH. Nevertheless, a trend can also be seen for r.pH, which would have a significant effect if analyzed with SRS as regression instead of the SRS class in the otherwise identical model. With regard to the lying duration, significant differences can also be observed.

Particularly remarkable is the drop in performance by approximately 20% with increasing SRS classes, which affects not only milk performance but also daily fat and protein yield. The SCS and LF levels are significantly higher in the highest SRS class than in the lowest. Both are well-known indicators of mastitis (Soyeurt et al., 2012). This indicates that the SRS may be linked to other diseases. This was to be expected, particularly as the decrease in performance, eating duration and rumination activity are unspecific signs that can be attributed to clinical conditions of various disorders, such as acidosis, mastitis, or claw disorders (Fleischer et al., 2001; Enemark, 2008). To a certain degree, this is inevitable and generally not a disadvantage, as the superior goal to differentiate between diseased and healthy animals is pursued. In addition, we assume that the value added by taking into account the additional traits compensates for the lower specificity. However, this should be verified with an external data set including veterinary diagnostics or ventral ruminal pH measurements.

The LF content of milk is known to be influenced by the lactation stage and is also negatively correlated with milk yield (Cheng et al., 2008). Consequently, higher LF values are to be expected as a result of lower milk

yield. Nevertheless, when considering the daily amount of LF excreted in milk instead of the concentration, a positive trend ($P = 0.058$) was still observed with SRS as regression term instead of the SRS as a class variable in the model. Because LF plays an important role in the immune system of mammals, including iron-binding abilities and bactericidal effects (Baker and Baker, 2004), the increased LF could indicate increased immune responses.

Blood contents of BHB and NEFA, which are often used as indicators for metabolic disorders (Oetzel, 2004), do not show a directional relationship to the SRS class. However, an indirect relationship via SARA-related lower feed intake, the resulting reduced energy balance, and the higher BHB and NEFA levels associ-

ated with body fat mobilization is conceivable (Baird, 1982; Enemark, 2008). An increase in the blood glucose level is observed if SRS is high. This can be consistent with the results of Khafipour et al. (2009a,b) and Li et al. (2012), who reported increased blood glucose concentration as consequence of a SARA challenge. Considering that propionate is known for its glucogenic role as precursor of gluconeogenesis (Miettinen and Huhtanen, 1996), a reason for this could be a higher concentration of propionate in ruminal VFA caused by diets rich in concentrate (Lechartier and Peyraud, 2011). This process could also be in line with the fact that SARA-promoting diets can lead to an enrichment of bacterial endotoxins in the form of lipopolysaccharides in both the rumen and blood plasma, which in turn can lead

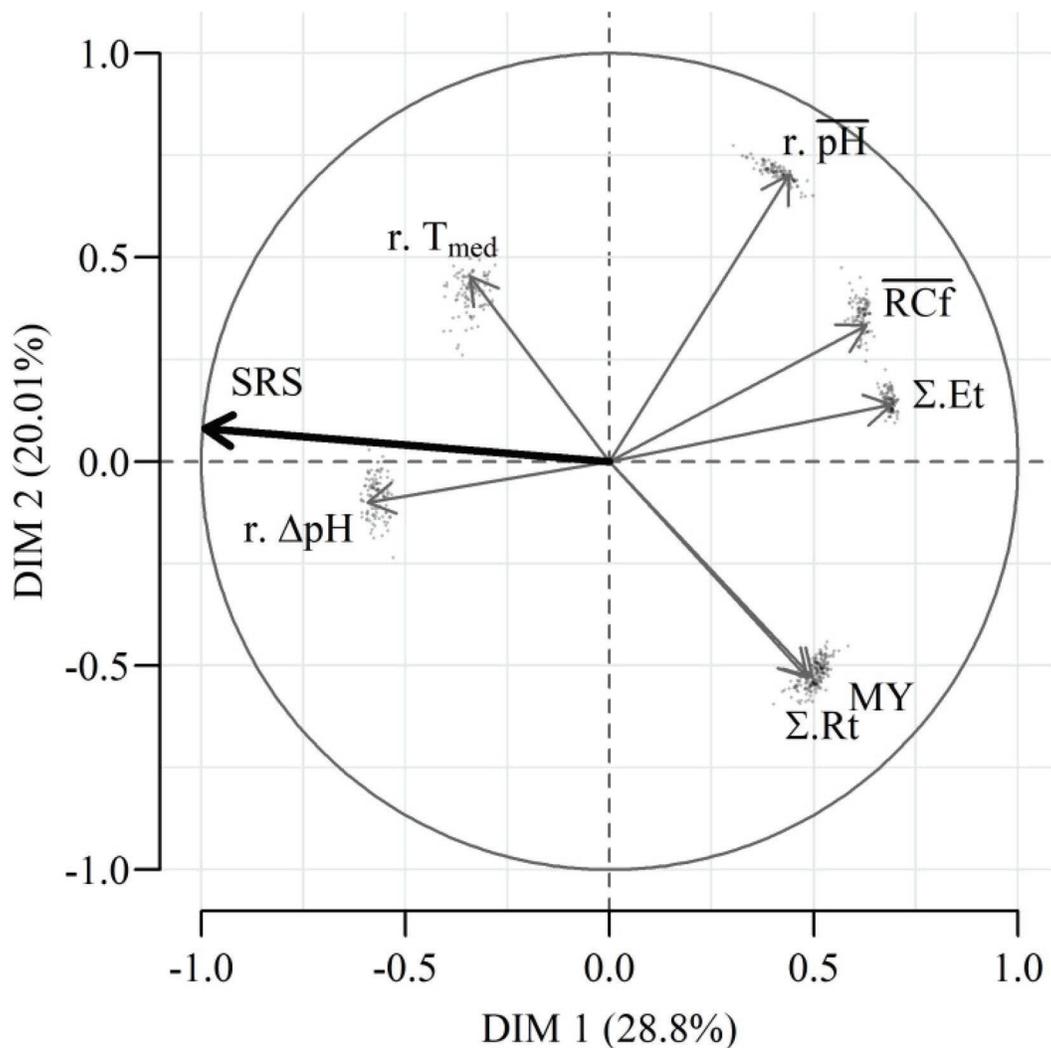


Figure 2. Variable factor map using the first 2 principal components based on the averaged multiple imputed data set of the 7 adjusted and z-transformed traits. In addition to the loadings from the variables of the averaged data set (gray arrows), the loadings from the 100 imputed data sets are projected in the same space (points). Furthermore, the SARA risk score was projected into the same space (bold black arrow). $\overline{\text{RCf}}$ = daily mean rumination chewing frequency; $r.\overline{\text{pH}}$ = reticular daily mean pH; $r.T_{\text{med}}$ = median of the daily reticular temperature; SRS = SARA risk score; MY = milk yield; $r.\Delta\text{pH}$ = daily reticular pH range; $\Sigma.\text{Et}$, $\Sigma.\text{Rt}$ = daily durations of eating and rumination, respectively.

to inflammatory reactions (Khafipour et al., 2009b). However, an activated immune system is characterized by increased body temperatures and an increased glucose requirement (Kvidera et al., 2017). Taking into account the fact that glucose is an important precursor of lactose (Kleiber et al., 1955), the significant lower milk yield and thus reduced release of lactose via milk could be linked to the higher SRS.

The PCA results of dietary characteristics as well as the LSM of SRS and the other 7 traits per farm are shown in Figure 4. The feed data-based variable factor map shows the space spanned by the first 2 PC explaining together 67.23% of the variance. The tips of the

gray arrows correspond with the loadings of the feed-related variables. In addition, the SRS, as well as the other 7 traits were projected into the same space as supplementary variables, as shown using black bold arrows. It can be seen from this figure that the diets of the 10 farms differ mainly with regard to the ratio of properties associated with fiber (NDF after amylase treatment and ashing, crude fiber, peNDF > 8) to energy (starch, nitrogen-free extract, NE_L), as they position opposite along the first PC. As expected, the daily average pH also tends to point in the direction of the fiber-associated variables, whereas it is opposite to the variables that are typical for diets rich in concentrate.

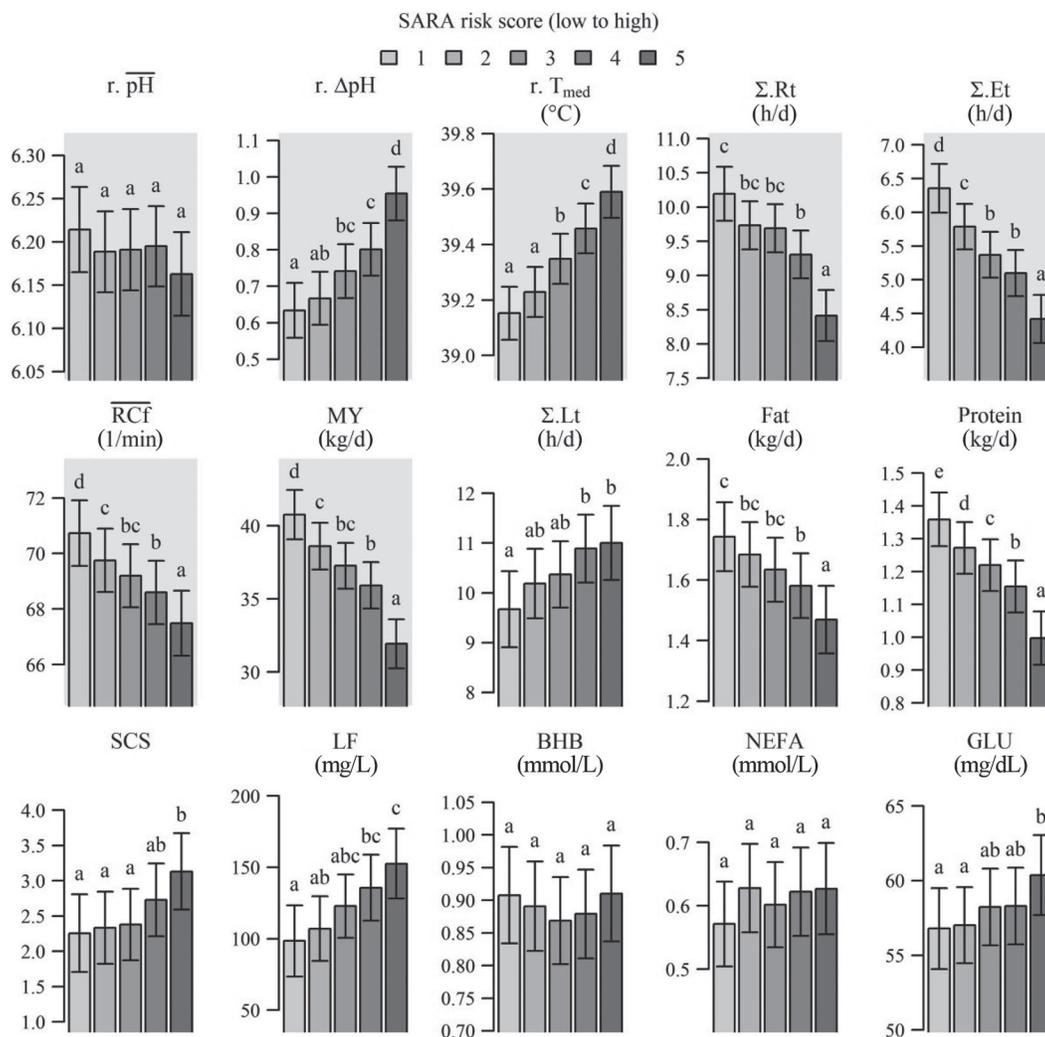


Figure 3. LSM (± 1.96 SE) for the 7 reference traits used for the construction of the SARA risk score (SRS; light gray background), further performance traits, and blood characteristics ($307 \leq n \leq 354$), depending on the quintile-based SRS classes. Post hoc comparison of LSM with Tukey's test at $P = 0.05$. Same superscripts indicate nonsignificant differences. The \log_{10} BHB and NEFA estimates were back transformed to the original scale. BHB = blood β -hydroxybutyrate level; GLU = blood glucose; RCf = daily mean rumination chewing frequency; r.pH = reticular daily mean pH; r.T_{med} = median of the daily reticular temperature; SCS = somatic cell score according to Wiggins and Shook (1987); LF = milk lactoferrin; MY = milk yield; NEFA = blood nonesterified fatty acid level; r.ΔpH = daily reticular pH range; Σ.Et, Σ.Rt, and Σ.Lt = daily duration of eating, rumination, and lying, respectively.

Thus, a large consensus with the results of Mensching et al. (2020b) can be seen, where feed properties and pH parameters of meta-analytical data were also presented jointly in a variable factor map. Consequently, Figure 4 confirms the basic relationships between diet properties and pH development with regard to SARA (Kleen et al., 2003; Enemark, 2008). Interestingly, the SRS is positioned in the direction of starch and nitrogen-free extract, as well as opposite to the fiber-associated properties and the $r.p\bar{H}$. This generally confirms that the SRS has a high potential to indicate SARA.

Presumably, the observed relationships will not only be attributable to the composition of the diet but can also be related to the individual feed intake behavior of

the animals. In this respect, Coon et al. (2019) demonstrated that part of the observed interanimal variance regarding susceptibility to SARA is attributable to behavioral differences in sorting against long particles.

The relationship between the milk main components, FA traits, and SRS are shown in Figure 5 in a spider map. In this figure, one can see that most of the milk traits display a directed relationship with the SRS class. With regard to the milk main components, higher SRS are accompanied by increased fat and decreased protein and lactose. It is known that lower lactose levels are associated with udder diseases, as is the case with the previously discussed increased SCS (Berning and Shook, 1992). In terms of SARA, a milk fat depression

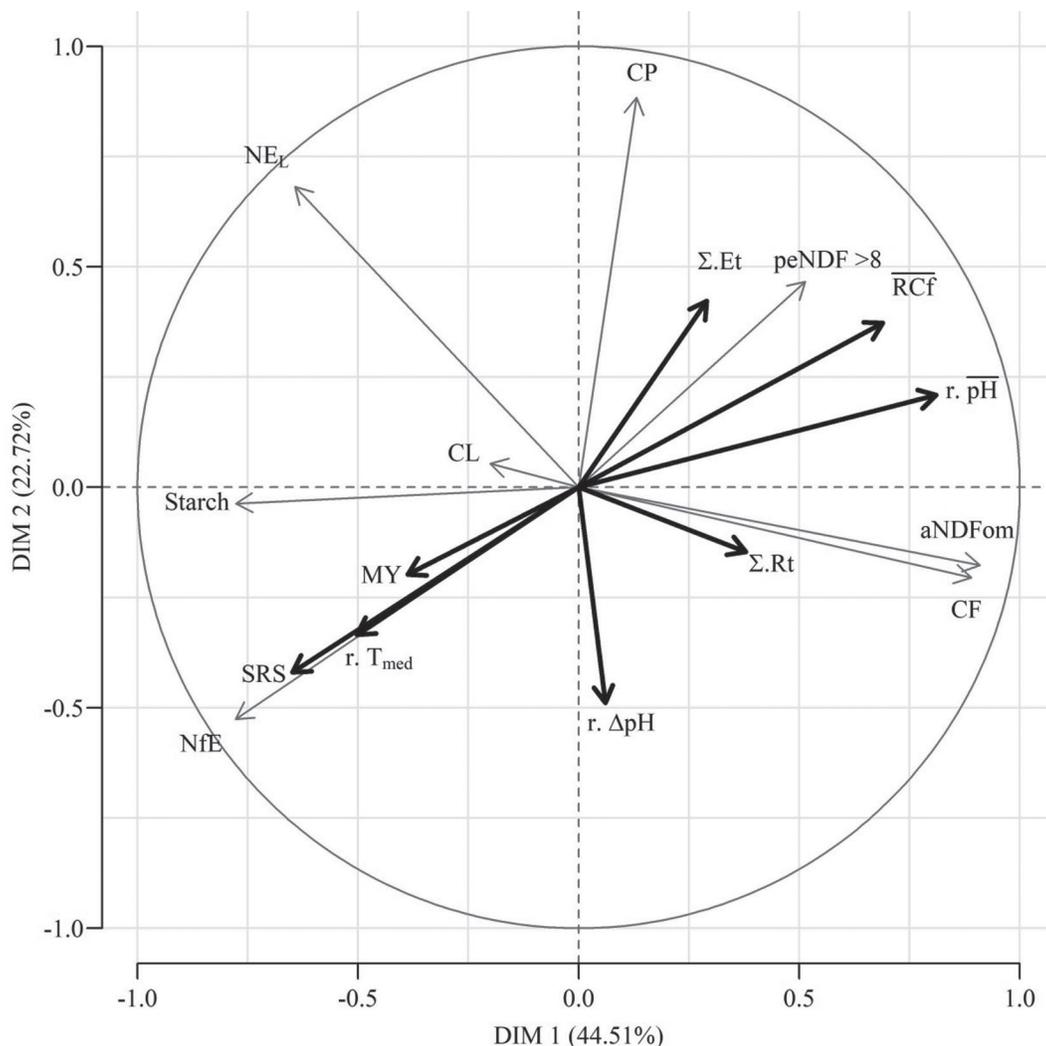


Figure 4. Variable factor map using the first 2 principal components of the feed analysis data. In addition to the loadings from the feed-related variables (gray arrows), the SARA risk score (SRS), and the other 7 traits used for its development were projected into the same space (bold black arrows). aNDFom = NDF after amylase treatment and ashing; CF = crude fiber; CL = crude fat; LF = milk lactoferrin; MY = milk yield; NfE = nitrogen-free extract; peNDF > 8 = physically effective NDF retained on an 8-mm sieve; \overline{RCf} = daily mean rumination chewing frequency; $r.p\bar{H}$ = reticular daily mean pH; $r.T_{med}$ = median of the daily reticular temperature; $r.\Delta pH$ = daily reticular pH range; $\Sigma.Et$, $\Sigma.Rt$, and $\Sigma.Lt$ = daily duration of eating, rumination, and lying, respectively.

would be expected (Enemark, 2008). However, with regard to SRS, the opposite was observed in this study, namely higher fat levels accompanied by significantly lower daily fat yields, the latter being mainly related to the lower milk yield (see Figure 3). This could be attributed to the fact that all cows were in early lactation with DIM ≤ 37 . Enemark et al. (2004) have described a numerically negative but not significant association between pH and milk fat content in early-lactation cows, from which they concluded that the individual fat content is not suitable as an indicator for SARA at this stage of lactation. The increase in milk fat might be caused by a reversed dilution effect as a consequence of reduced milk performance.

Further, lower levels of short- and medium-chain FA are observed with a higher SRS, whereas both significant and nonsignificant associated FA are found in LCFA. No significant association between SRS class and *trans* LCFA can be observed, whereas significantly higher levels of the *cis* LCFA C18:2 *cis-9,cis-12*, C18:1 *cis*, and C18:1 *cis-9* can be found with an increasing SRS class. Some patterns in the milk FA profile are similar to those already observed by Gross et al. (2011) as a consequence of a negative energy balance. These include elevated C18:1 *cis-9* as well as lower levels of the FA from the de novo synthesis (≤ 16 carbon atoms) as a result of a higher proportion of fat resulting from the mobilization of adipose tissue. However, a negative

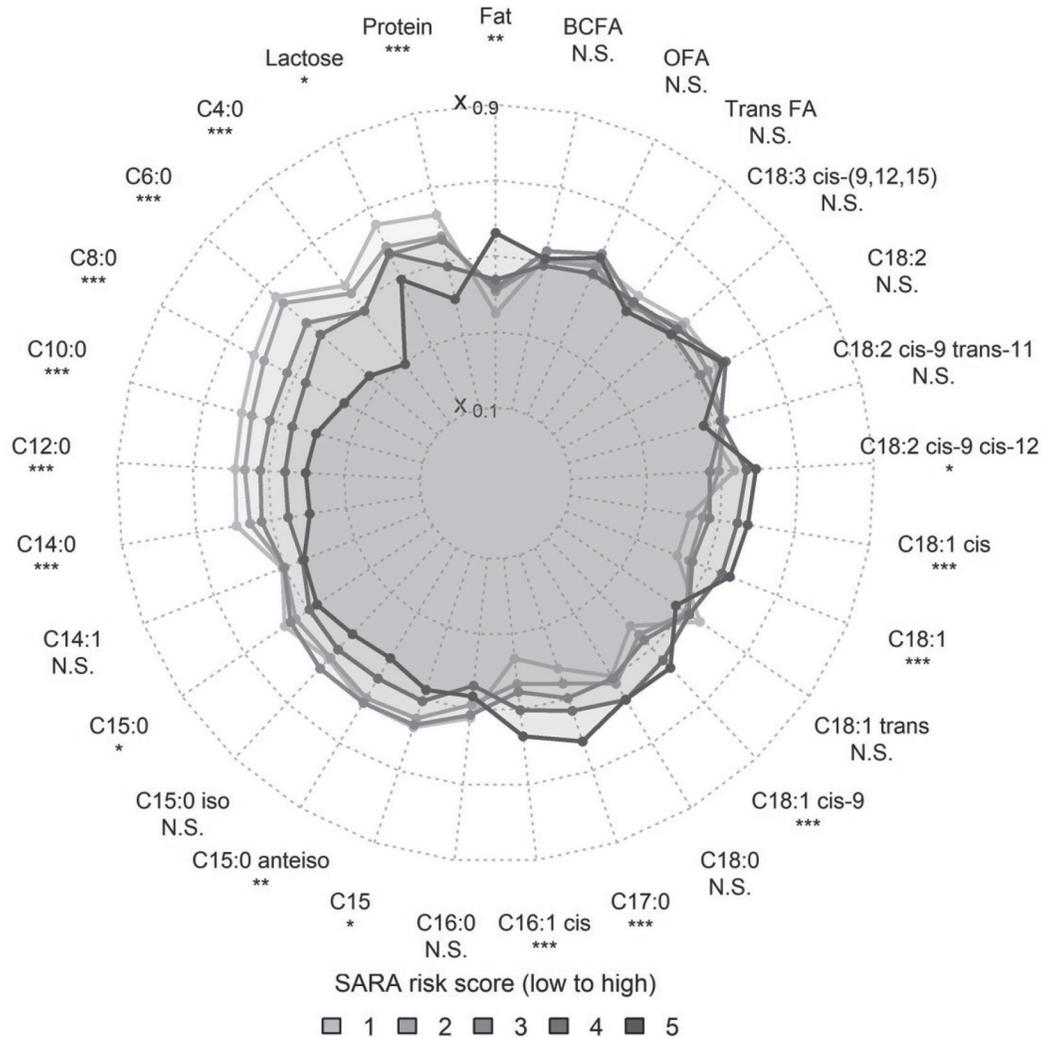


Figure 5. Spider plot (Nakazawa, 2019) to visualize standardized LSM of milk main components (%) and mid-infrared-predicted fatty acid traits (% of fat) depending on a quintile-based classification of the SARA risk score for the primary data set ($351 \leq n \leq 357$). The axes of the individual variables range from the 10 ($x_{0.1}$) to the 90% ($x_{0.9}$) quantile. The Cx:y-coding includes the number of C-atoms (x) as well as the number of double bonds (y) of the respective fatty acid. P-values for the effect of class affiliation on the content of the respective fatty acid from the ANOVA table: *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; N.S. = $P \geq 0.05$. FA = fatty acids; BCFA = branched-chain FA; OFA = odd-chain FA.

energy balance is plausible, as poor body conditions are also considered to be an indication for SARA (Nocek, 1997).

Additionally, the milk odd- and branched-chain FA, which are essentially derived from bacteria that have been built and left the rumen, are considered to be indicators of ruminal fermentation (Vlaeminck et al., 2006). In this regard, significantly higher levels of C17:0 and lower levels of all C15 FA could be associated with increased SRS in the primary data set. In comparison, Vlaeminck et al. (2006) determined a positive association for C15:0 with the ruminal propionate concentration, whereas they found an opposite association for C15:0 iso. Also, Van Nespen et al. (2005) observed a changed FA profile of the milk with increased iso C17:0, C15:0, C18:2 *cis*-7,*trans*-11, and C18:1 *trans*-10 + C18:1 *trans*-11, and lower contents of iso C14:0 as consequence of increased dietary starch content.

MIR-Based PLS Regression Models

Descriptive statistics for all traits for which milk MIR-based PLS models were developed are summarized in Table 1 separately for the data sets used for calibration and for external validation. All traits are nearly symmetrical and approximately normally distributed. However, it is noticeable that the $r.\Delta\text{pH}$ tends to show extreme values in the upper range, and $r.\overline{\text{pH}}$, daily rumination, and eating duration in the lower range, which can be indicative for SARA or even acute ruminal acidosis-related conditions. Considering the observed SARA prevalence of about 20% in German herds (Kleen et al., 2013), this is plausible, as it can be assumed that the majority of the animals were in a physiologically healthy condition and only a small number were diseased.

Table 1 also contains model characteristics, model statistics, and the results of the external validation on individual PLS calibration models. Furthermore, Figure 6 shows for all traits a plot observed against the predicted values from cross-validation and external validation. The use of the entire spectrum led to the best results only for RCF; otherwise, the reduced spectrum with 212 or 516 wavenumbers always led to the best model performances. This was expected, considering that large parts of the spectrum contain noisy regions due to the water content of milk (Grelet et al., 2015). Furthermore, it was observed that the choice of spectral pretreatment is dependent on the respective trait. The fit statistics of the best models only indicate moderate model performances. Looking at R_{val}^2 of the 8 traits, only the models for SRS and $r.\Delta\text{pH}$ show a

noteworthy potential, with $R_{\text{val}}^2 = 0.37$ and $R_{\text{val}}^2 = 0.49$. Consequently, these models explain only a part of the observed variation and are not able to make quantitative accurate predictions. All other models are characterized by $R_{\text{val}}^2 < 0.10$ and thus show inferior performance.

Especially for the SRS and for $r.\text{Tmed}$, models with a relatively high number of LV emerged. As a consequence, a lack of robustness of the model is to be expected, which is also reflected in a large discrepancy between the results of calibration and cross-validation. But this is also true for all other models. Another reason could be the small size of the data set, and associated lack of variability (Davies and Fearn, 2006).

In their current forms, these models cannot be used for individual cow management. However, some models seem to be able to distinguish between extreme values. In this context, Gengler et al. (2016) argued that, under the assumption of random prediction errors, the application of such models to a group of animals could increase the accuracy of the predictions. Considering the fact that these traits are very complex and do not correspond to direct components of the milk, these associations are appreciable and indicate a high potential to obtain additional information that goes beyond the standard components relevant for milk payment. Because the present PLS models are developed with a relatively small amount of data, with few suspected cases of SARA, it is likely that the reduced data volume causes an underestimation of the potential R^2 (Davies and Fearn, 2006). Therefore, it would be of great interest to implement a data set with a higher proportion of SARA-affected cows, to evaluate potential increase of the model performances.

It is interesting that the R_{cv}^2 and R_{val}^2 for $r.\overline{\text{pH}}$ are very low, whereas they are at least moderate for $r.\Delta\text{pH}$. This could have several reasons. In terms of $r.\overline{\text{pH}}$, it could be that it is generally not well predictable. However, the low accuracy could also be attributable to the reticular pH measurement. Although the reticular and (ventral) ruminal pH are highly associated, it is not possible to infer precisely from reticular pH to ruminal pH, with the latter being the target pH with regard to SARA (Falk et al., 2016). Another reason could be the accuracy of the measurement technique. It is well known that drift is a problem of indwelling boluses (e.g., Villot et al., 2018), although it is more likely to occur several months after insertion of the bolus in the reticulum. However, the drift would primarily affect the accuracy of $r.\overline{\text{pH}}$ and not of $r.\Delta\text{pH}$.

Table 1. Descriptive statistics of the calibration and validation data, model characteristics and statistics, as well as validation results of the milk mid-infrared spectra-based partial least squares calibration models¹

Unit	Trait							
	SRS	r.pH	r.ΔpH	r.T _{med} , °C	Σ.Rt, h/d	Σ.Et, h/d	RCf, 1/min	Σ.Lt, h/d
Descriptive statistics of calibration data								
n	290	253	233	259	277	280	268	268
n _A	79	70	70	73	78	78	79	79
Mean	-0.12	6.19	0.76	39.36	9.44	5.32	68.15	10.57
SD	3.48	0.16	0.22	0.33	1.27	1.26	5.13	2.50
Min	-6.51	5.54	0.38	38.30	3.39	1.65	55.48	3.37
Median	-0.56	6.23	0.73	39.30	9.57	5.29	67.94	10.73
Max	16.45	6.52	1.67	41.30	11.94	9.04	79.87	18.38
Descriptive statistics of validation data								
n	67	54	74	62	59	60	75	66
n _A	20	19	19	20	17	18	18	19
Mean	0.32	6.17	0.78	39.41	9.42	5.40	68.74	10.57
SD	4.12	0.17	0.21	0.40	1.33	1.28	5.69	3.61
Min	-7.16	5.77	0.40	38.90	5.00	1.46	59.57	1.38
Median	0.20	6.21	0.74	39.30	9.57	5.45	68.48	10.75
Max	14.07	6.49	1.56	40.70	11.54	7.87	83.23	17.25
Model characteristics								
Pretreat	gap/1/5	gap/1/5	2/2/5	gap/1/5	—	—	0/2/5	2/2/5
LP	—	—	—	—	LP(DIM,1)	—	—	—
WN	516	212	516	212	212	212	1,060	212
LV	12	3	3	13	8	8	3	3
Model statistics and validation results								
RMSE _{cal}	2.36	0.15	0.18	0.24	1.02	1.06	4.34	2.30
RMSE _{cv}	2.91	0.16	0.19	0.30	1.10	1.17	4.85	2.40
RMSE _{val}	3.35	0.18	0.15	0.41	1.45	1.29	5.76	3.42
R ² _{cal}	0.56	0.16	0.33	0.52	0.37	0.32	0.29	0.16
R ² _{cv}	0.32	0.06	0.23	0.26	0.26	0.17	0.13	0.09
R ² _{val}	0.37	0.05	0.49	0.09	0.00	0.03	0.04	0.10
RPD _{cv}	1.19	1.02	1.14	1.12	1.16	1.08	1.06	1.04

¹LP, LP(DIM,1) = Legendre polynomial transformation of the spectrum; see Material and Methods. n = sample size; n_A = number of animals; min, max = minimum and maximum. Pretreat = spectral pretreatments; see Material and Methods. R²_{cal}, R²_{cv}, R²_{val} = coefficient of determination of calibration model, of cross-validation, and of external validation. RMSE_{cal}, RMSE_{cv}, and RMSE_{val} = root mean squared error of the calibration model, cross-validation, and external validation, respectively. RPD_{cv} = ratio of SD of reference to the SE of cross-validation. RCf = daily mean rumination chewing frequency; r.pH = reticular daily mean pH; r.T_{med} = median of the daily reticular temperature; r.ΔpH = daily reticular pH range; SD = empirically estimated SD; SRS = SARA risk score; WN = number of wavenumbers; LV = latent variables. Σ.Rt, Σ.Et, and Σ.Lt = daily duration of rumination, eating, and lying, respectively.

Comparing the achieved model statistics for the SRS with the results of others who have worked on MIR-based prediction of complex indirect traits, higher performance can be found. McParland et al. (2014), for example, were able to achieve a cross-validated correlation between observed and predicted values when modeling the energy balance with correlation coefficient (r) = 0.67 and R²_{cv} = 0.45, respectively. In the MIR-based prediction of methane emissions measured with a respiration chamber Vanlierde et al. (2018) even achieved an R²_{cal} of 0.65 and R²_{cv} of 0.57 using a cross-country data set. However, those studies are based on a large vari-

ability, with more than 1,500 records and data from 5 countries. This encourages increase of the current data set with complementary records, which, at best, also cover the entire lactation period.

Because both the FA traits and the SRS can be estimated at least moderately with the milk MIR spectra and they show moderate associations, the predictability of the SRS is likely to be based on altered milk composition. Therefore, the following causality chain can be hypothesized: diet (composition), environment (e.g., temperature), individual animal (e.g., microbiome, genetics, behavior, lactation stage, parity), as well as interactions between them, influence the ruminal fer-

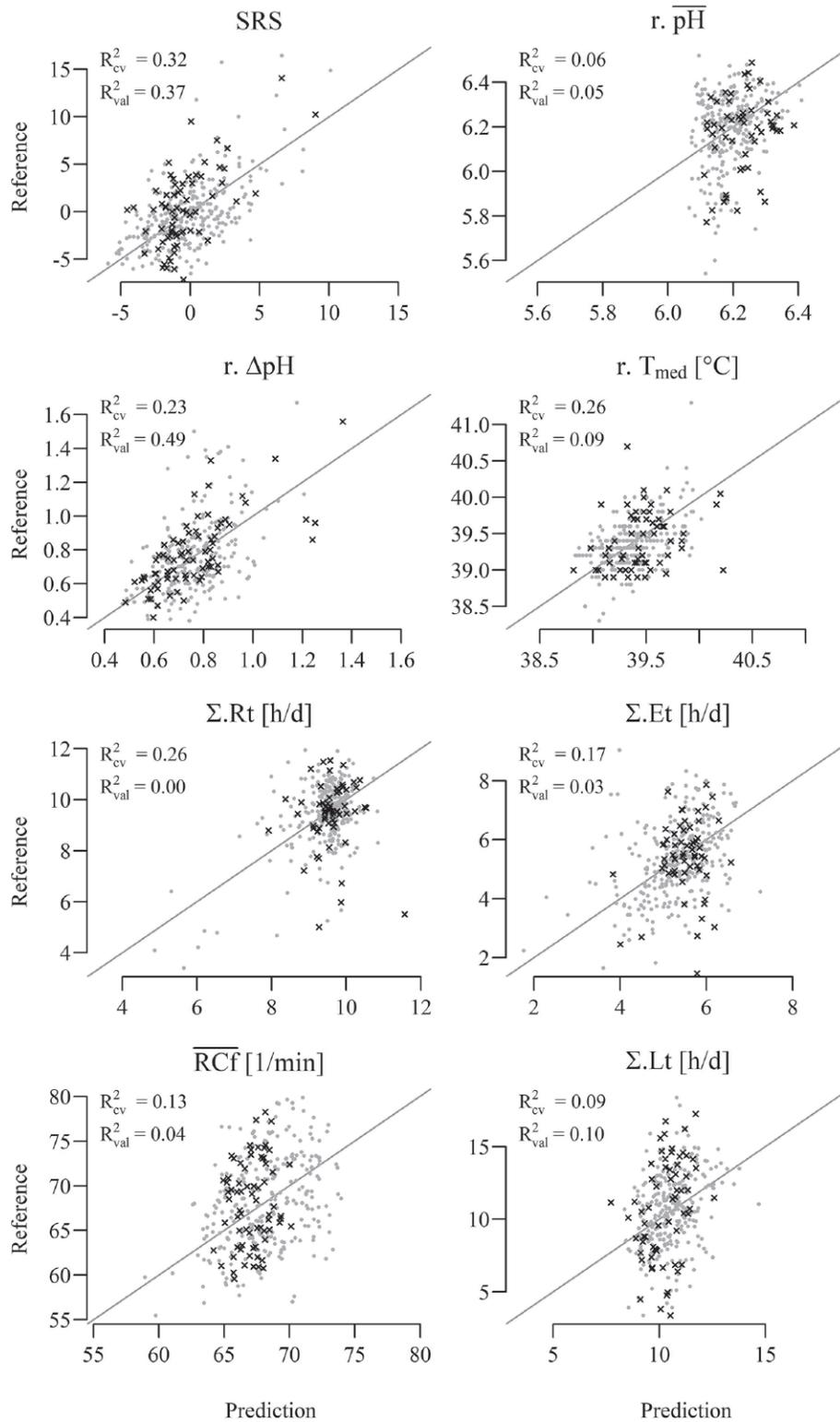


Figure 6. Scatterplot for reference values against predicted values from cross-validation (light gray dots) and external validation (crosses) based on mid-infrared-based partial least squares regression models. R_{cv}^2 , R_{val}^2 = coefficient of determination of cross-validation and of validation, respectively; \overline{RCf} = daily mean rumination chewing frequency; $r.pH$ = reticular daily mean pH; $r.T_{med}$ = median of the daily reticular temperature; SRS = SARA risk score; $r.\Delta pH$ = daily reticular pH range; $\Sigma.Et$, $\Sigma.Rt$, and $\Sigma.Lt$ = daily duration of eating, rumination and lying, respectively ($306 \leq n \leq 356$).

mentation process. This is reflected in the ruminal FA pattern and subsequently also to some extent in the FA pattern of the milk. In turn, this is related to a modified MIR spectrum.

Because milk fat is considered the most variable component of milk, in terms of both concentration and composition (Palmquist, 2006), the major part of spectral variability is attributable to altered milk fat composition. Therefore, it can be assumed that the determined associations of the other 7 traits are also essentially attributable to an altered milk fat composition, whereby the model quality achieved for each trait provides information about the strength of this relationship.

Application on the Secondary Data Set

The developed prediction equation for the SRS was applied to the secondary data for observations in the same lactation stage ($5 \leq \text{DIM} \leq 37$). As no reference values of the SRS are available, this is not an external validation in the narrow sense. Nevertheless, the SRS predictions were associated with the main milk components and the FA profile. The results of this investigation are shown in Figure 7 in a spider map (Nakazawa, 2019). Comparing this spider map with the one in Figure 5, a very high level of agreement is noticeable. It seems that for some of the traits, the directional relationships are even more evident than in

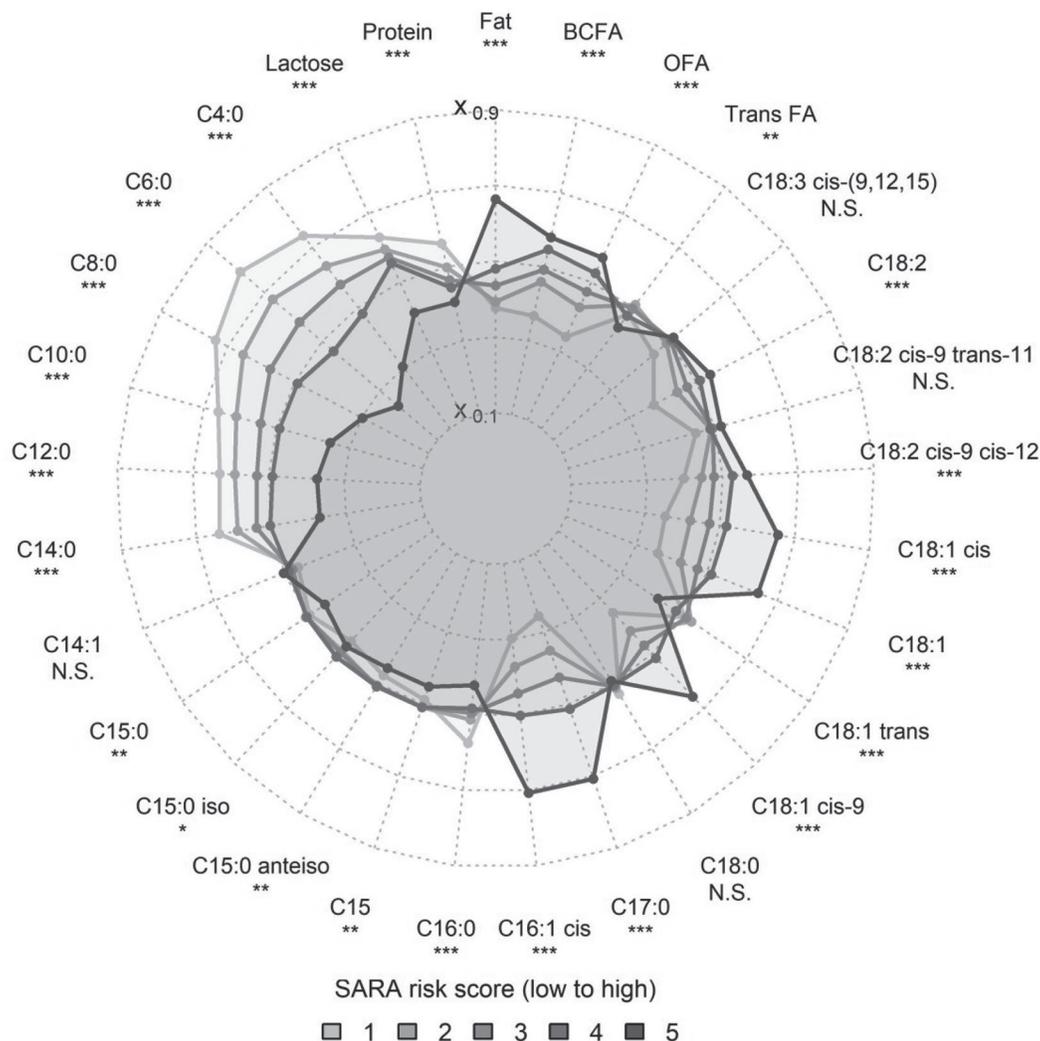


Figure 7. Spider plot (Nakazawa, 2019) to visualize standardized LSM of milk main components (%) and mid-infrared (MIR)-predicted fatty acid traits (% of fat) depending on a quintile-based classification of the MIR-predicted SARA risk score for the secondary data set ($n = 2,444$; $5 \leq \text{DIM} \leq 37$). The axes of the individual variables range from the 10% ($x_{0.1}$) to the 90% ($x_{0.9}$) quantile. The Cx:y-coding includes the number of C-atoms (x) as well as the number of double bonds (y) of the respective fatty acid. P-values for the effect of class affiliation on the content of the respective fatty acid from the ANOVA table: *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; N.S. = $P \geq 0.05$. FA = fatty acids; BCFA = branched-chain FA; OFA = odd-chain FA.

the primary data set. Due to the small number of observations in the primary data set, it is likely that more SARA cases are present in the secondary data set, with an increased discrimination between extreme groups. Complementary, more FA variation is obviously present in the secondary data set.

The fact that higher milk MIR-predicted SRS are accompanied with higher levels of odd- and branched-chain FA, as well as C17:0, validates the SRS approach, as such directed relationships between the milk FA profile and ruminal fermentation properties have already been discussed in terms of SARA as a consequence of diets rich in concentrate (Fievez et al., 2012). However, in view of the results generated here, it is crucial that the observed changes in the FA profile of Figure 7 are attributed to different values of the predicted SRS, as the farm and thus implicitly also the feeding was considered as a random effect in the used LMM. Nevertheless, farm- and therefore potential feed-specific effects are reflected in the frequency distribution of the predicted SRS classes within individual farms (see Supplemental Figure S2, <http://dx.doi.org/10.6084/m9.figshare.13669966>). But because all SRS classes, and implicitly the varying FA profiles, occur in all farms, a large part of the observed variation is presumably related to high interanimal variance.

CONCLUSIONS

This study developed an innovative approach in which information from different independent data acquisition systems (intrareticular measurement bolus, noseband-sensor halter, and milk performance) were combined to design an integrative indicator trait for SARA called SRS. At the farm level, higher SRS were related to lower daily average pH values, higher levels of easily fermentable carbohydrates, and lower levels of physical structure in the diet. On the individual level, an increased SRS was also found to be linked to a modified FA profile, in which lower levels of saturated and short-chain FA and higher levels of C16:1 *cis*, C17:0, and odd- and branched-chain FA could be observed. Furthermore, using a PLS regression model, a milk MIR spectra-based prediction equation with a moderate predictability was established, demonstrating the high potential of milk composition-based characterization of the health status of lactating cows. Based on our research, a first approach for a SARA monitoring could be implemented on a large scale in routine milk performance testing using the MIR-based prediction of the SRS.

ACKNOWLEDGMENTS

This work was done within the project “Evaluation of Animal Welfare in Dairy Farming – Indicators for the Metabolism and Feeding” (IndiKuh, Goettingen, Germany; funding code: 2817905815). The project was supported by funds of the Federal Ministry of Food and Agriculture (BMEL, Bonn, Germany) based on a decision of the parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE, Bonn, Germany) in the course of the program for promotion of innovations. The authors especially thank Henrike Maria Jansen and Dirk Albers from the Chamber of Agriculture of Lower Saxony (Oldenburg, Germany) for their extensive contribution to the data collection. In addition, the authors thank Ernst Bohlsen, representative of the Landeskontrollverband Weser-Ems e.V. (LKV Weser-Ems, Leer, Germany) for carrying out the milk analyses, and the Vereinigte Informationssysteme Tierhaltung w.V. (vit, Verden, Germany) for its support in data management. The authors thank all further contributors involved in data collection and laboratory analysis. In particular, we thank the owners of the farms who made the whole project possible by participating. Clément Grelet and Nicolas Gengler acknowledge the support of the Walloon Government (Service Public de Wallonie – Direction Générale Opérationnelle Agriculture, Ressources Naturelles et Environnement; SPW-DGARNE, Namur, Belgium) and the access to the computation resources of the University of Liège – Gembloux (Belgium) Agro-Bio Tech provided by the technical platform Calcul et Modélisation Informatique (CAMI) of the TERRA Teaching and Research Centre (Gembloux, Belgium), partly supported by the Fonds de la Recherche Scientifique – FNRS under grants no. T.0095.19 (PDR “DEEP-SELECT”) and J.0174.18 (CDR “PREDICT-2”). The authors have not stated any conflicts of interest.

REFERENCES

- AlZahal, O., H. AlZahal, M. A. Steele, M. Van Schaik, I. Kyriazakis, T. F. Duffield, and B. W. McBride. 2011. The use of a radio-telemetric ruminal bolus to detect body temperature changes in lactating dairy cattle. *J. Dairy Sci.* 94:3568–3574. <https://doi.org/10.3168/jds.2010-3944>.
- AlZahal, O., E. Kebreab, J. France, M. Froetschel, and B. W. McBride. 2008. Ruminal temperature may aid in the detection of subacute ruminal acidosis. *J. Dairy Sci.* 91:202–207. <https://doi.org/10.3168/jds.2007-0535>.
- Baird, G. D. 1982. Primary ketosis in the high-producing dairy cow: Clinical and subclinical disorders, treatment, prevention, and outlook. *J. Dairy Sci.* 65:1–10. [https://doi.org/10.3168/jds.S0022-0302\(82\)82146-2](https://doi.org/10.3168/jds.S0022-0302(82)82146-2).
- Baker, H. M., and E. N. Baker. 2004. Lactoferrin and iron: Structural and dynamic aspects of binding and release. *Biometals* 17:209–216. <https://doi.org/10.1023/B:BIOM.0000027694.40260.70>.

- Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67:1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Beauchemin, K. A. 2018. Invited review: Current perspectives on eating and rumination activity in dairy cows. *J. Dairy Sci.* 101:4762–4784. <https://doi.org/10.3168/jds.2017-13706>.
- Berning, L. M., and G. E. Shook. 1992. Prediction of mastitis using milk somatic cell count, N-acetyl- β -D-glucosaminidase, and lactose. *J. Dairy Sci.* 75:1840–1848. [https://doi.org/10.3168/jds.S0022-0302\(92\)77943-0](https://doi.org/10.3168/jds.S0022-0302(92)77943-0).
- Cheng, J. B., J. Q. Wang, D. P. Bu, G. L. Liu, C. G. Zhang, H. Y. Wei, L. Y. Zhou, and J. Z. Wang. 2008. Factors affecting the lactoferrin concentration in bovine milk. *J. Dairy Sci.* 91:970–976. <https://doi.org/10.3168/jds.2007-0689>.
- Coon, R. E., T. F. Duffield, and T. J. DeVries. 2019. Short communication: Risk of subacute ruminal acidosis affects the feed sorting behavior and milk production of early lactation cows. *J. Dairy Sci.* 102:652–659. <https://doi.org/10.3168/jds.2018-15064>.
- Davies, A. M. C., and T. Fearn. 2006. Back to basics: Calibration statistics. *Spectrosc. Eur.* 18:31–32.
- de Jong, S. 1993. SIMPLS: An alternative approach to partial least squares regression. *Chemom. Intell. Lab. Syst.* 18:251–263. [https://doi.org/10.1016/0169-7439\(93\)85002-X](https://doi.org/10.1016/0169-7439(93)85002-X).
- De Koster, J., M. Salavati, C. Grelet, M. A. Crowe, E. Matthews, R. O’Flaherty, G. Opsomer, L. Foldager, and M. Hostens. 2019. Prediction of metabolic clusters in early-lactation dairy cows using models based on milk biomarkers. *J. Dairy Sci.* 102:2631–2644. <https://doi.org/10.3168/jds.2018-15533>.
- Denwood, M. J., J. L. Kleen, D. B. Jensen, and N. N. Jonsson. 2018. Describing temporal variation in reticuloruminal pH using continuous monitoring data. *J. Dairy Sci.* 101:233–245. <https://doi.org/10.3168/jds.2017-12828>.
- DWD Climate Data Center (CDC). 2018. Historical hourly station observations of 2 m air temperature and humidity for Germany, version V006. Accessed Oct. 25, 2019. https://opendata.dwd.de/test/CDC/observations_germany/climate/hourly/air_temperature/historical/DESCRIPTION_obsgermany_climate_hourly_tu_historical_en.pdf.
- Enemark, J. M. D. 2008. The monitoring, prevention and treatment of sub-acute ruminal acidosis (SARA): A review. *Vet. J.* 176:32–43. <https://doi.org/10.1016/j.tvjl.2007.12.021>.
- Enemark, J. M. D., R. J. Jørgensen, and N. B. Kristensen. 2004. An evaluation of parameters for the detection of subclinical rumen acidosis in dairy herds. *Vet. Res. Commun.* 28:687–709. <https://doi.org/10.1023/B:VERC.0000045949.31499.20>.
- Enemark, J. M. D., R. J. Jørgensen, and P. St. Enemark. 2002. Rumen acidosis with special emphasis on diagnostic aspects of subclinical rumen acidosis: A review. *Vet. IR Zootech.* 20:16–29.
- Falk, M., A. Münger, and F. Dohme-Meier. 2016. Technical note: A comparison of reticular and ruminal pH monitored continuously with 2 measurement systems at different weeks of early lactation. *J. Dairy Sci.* 99:1951–1955. <https://doi.org/10.3168/jds.2015-9725>.
- Fievez, V., E. Colman, J. M. Castro-Montoya, I. Stefanov, and B. Vlaeminck. 2012. Milk odd- and branched-chain fatty acids as biomarkers of rumen function—An update. *Anim. Feed Sci. Technol.* 172:51–65. <https://doi.org/10.1016/j.anifeedsci.2011.12.008>.
- Fleischer, P., M. Metzner, M. Beyerbach, M. Hoedemaker, and W. Klee. 2001. The relationship between milk yield and the incidence of some diseases in dairy cows. *J. Dairy Sci.* 84:2025–2035. [https://doi.org/10.3168/jds.S0022-0302\(01\)74646-2](https://doi.org/10.3168/jds.S0022-0302(01)74646-2).
- 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. <https://doi.org/10.3168/jds.2015-10140>.
- Grelet, C., C. Bastin, M. Gelé, J.-B. B. Davière, M. Johan, A. Werner, R. Reding, J. A. A. Fernandez Pierna, F. G. 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., 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>.
- Grelet, C., A. Vanlierde, M. Hostens, L. Foldager, M. Salavati, K. L. Ingvarsen, M. Crowe, M. T. Sorensen, E. Froidmont, C. P. Ferris, C. Marchitelli, F. Becker, T. Larsen, F. Carter, and F. Dehareng. 2019. Potential of milk mid-IR spectra to predict metabolic status of cows through blood components and an innovative clustering approach. *Animal* 13:649–658. <https://doi.org/10.1017/S1751731118001751>.
- Gross, J., H. A. Van Dorland, R. M. Bruckmaier, and F. J. Schwarz. 2011. Milk fatty acid profile related to energy balance in dairy cows. *J. Dairy Res.* 78:479–488. <https://doi.org/10.1017/S0022029911000550>.
- Humer, E., K. Ghareeb, H. Harder, E. Mickdam, A. Khol-Parisini, and Q. Zebeli. 2015. Peripartur changes in reticuloruminal pH and temperature in dairy cows differing in the susceptibility to sub-acute rumen acidosis. *J. Dairy Sci.* 98:8788–8799. <https://doi.org/10.3168/jds.2015-9893>.
- Husson, A.F., J. Josse, and M.F. Husson. 2019. missMDA: Handling missing values with multivariate data analysis, version 1.14: 1–39. <https://husson.github.io/>.
- ICAR. 2017. Section 2—Guidelines for Dairy Cattle Milk Recording. Accessed Mar. 14, 2020. <https://www.icar.org/Guidelines/02-Overview-Cattle-Milk-Recording.pdf>.
- Jiang, F. G., X. Y. Lin, Z. G. Yan, Z. Y. Hu, G. M. Liu, Y. D. Sun, X. W. Liu, and Z. H. Wang. 2017. Effect of dietary roughage level on chewing activity, ruminal pH, and saliva secretion in lactating Holstein cows. *J. Dairy Sci.* 100:2660–2671. <https://doi.org/10.3168/jds.2016-11559>.
- Jing, L., L. Dewanckele, B. Vlaeminck, W. M. Van Straalen, A. Koopmans, and V. Fievez. 2018. Susceptibility of dairy cows to sub-acute ruminal acidosis is reflected in milk fatty acid proportions, with C18:1 *trans*-10 as primary and C15:0 and C18:1 *trans*-11 as secondary indicators. *J. Dairy Sci.* 101:9827–9840. <https://doi.org/10.3168/jds.2018-14903>.
- Josse, J., and F. Husson. 2016. missMDA: A package for handling missing values in multivariate data analysis. *J. Stat. Softw.* 70. <https://doi.org/10.18637/jss.v070.i01>.
- Khafipour, E., D. O. Krause, and J. C. Plaizier. 2009a. Alfalfa pellet-induced subacute ruminal acidosis in dairy cows increases bacterial endotoxin in the rumen without causing inflammation. *J. Dairy Sci.* 92:1712–1724. <https://doi.org/10.3168/jds.2008-1656>.
- Khafipour, E., D. O. Krause, and J. C. Plaizier. 2009b. A grain-based subacute ruminal acidosis challenge causes translocation of lipopolysaccharide and triggers inflammation. *J. Dairy Sci.* 92:1060–1070. <https://doi.org/10.3168/jds.2008-1389>.
- Kleen, J. L., G. A. Hooijer, J. Rehage, and J. P. T. M. Noordhuizen. 2003. Subacute ruminal acidosis (SARA): A review. *J. Vet. Med. A Physiol. Pathol. Clin. Med.* 50:406–414. <https://doi.org/10.1046/j.1439-0442.2003.00569.x>.
- Kleen, J. L., L. Uppgang, and J. Rehage. 2013. Prevalence and consequences of subacute ruminal acidosis in German dairy herds. *Acta Vet. Scand.* 55:48. <https://doi.org/10.1186/1751-0147-55-48>.
- Kleiber, M., A. L. Black, M. A. Brown, C. F. Baxter, J. R. Luick, and F. H. Stadtman. 1955. Glucose as precursor of milk constituents in the intact dairy cow. *Biochim. Biophys. Acta* 17:252–260. [https://doi.org/10.1016/0006-3002\(55\)90357-7](https://doi.org/10.1016/0006-3002(55)90357-7).
- Kvidera, S. K., E. A. Horst, M. Abujamieh, E. J. Mayorga, M. V. S. Fernandez, and L. H. Baumgard. 2017. Glucose requirements of an activated immune system in lactating Holstein cows. *J. Dairy Sci.* 100:2360–2374. <https://doi.org/10.3168/jds.2016-12001>.
- Lechartier, C., and J.-L. Peyraud. 2011. The effects of starch and rapidly degradable dry matter from concentrate on ruminal digestion in dairy cows fed corn silage-based diets with fixed forage proportion. *J. Dairy Sci.* 94:2440–2454. <https://doi.org/10.3168/jds.2010-3285>.

- Li, S., G. N. Gozho, N. Gakhar, E. Khafipour, D. O. Krause, and J. C. Plaizier. 2012. Evaluation of diagnostic measures for subacute ruminal acidosis in dairy cows. *Can. J. Anim. Sci.* 92:353–364. <https://doi.org/10.4141/cjas2012-004>.
- Liang, D., C. L. Wood, K. J. McQuerry, D. L. Ray, J. D. Clark, and J. M. Bewley. 2013. Influence of breed, milk production, season, and ambient temperature on dairy cow reticulorumen temperature. *J. Dairy Sci.* 96:5072–5081. <https://doi.org/10.3168/jds.2012-6537>.
- Lindena, T., H. Tergast, R. Ellßel, and H. Hansen. 2018. Steckbriefe zur Tierhaltung in Deutschland: Milchkühe [Fact sheets on animal husbandry in Germany: Dairy cows]. Accessed Sep. 30, 2020. https://www.milchtrends.de/fileadmin/milchtrends/Literatur_Milchproduktion/Steckbrief_Milchkuehe2019.pdf.
- Loor, J. J., R. E. Everts, M. Bionaz, H. M. Dann, D. E. Morin, R. Oliveira, S. L. Rodriguez-Zas, J. K. Drackley, and H. A. Lewin. 2007. Nutrition-induced ketosis alters metabolic and signaling gene networks in liver of periparturient dairy cows. *Physiol. Genomics* 32:105–116. <https://doi.org/10.1152/physiolgenomics.00188.2007>.
- McParland, S., E. Lewis, E. Kennedy, S. G. Moore, B. McCarthy, M. O'Donovan, S. T. Butler, J. E. Pryce, and D. P. Berry. 2014. Mid-infrared spectrometry of milk as a predictor of energy intake and efficiency in lactating dairy cows. *J. Dairy Sci.* 97:5863–5871. <https://doi.org/10.3168/jds.2014-8214>.
- Mensinging, A., K. Bünemann, U. Meyer, D. von Soosten, J. Hummel, A. O. Schmitt, A. R. Sharifi, and S. Dänicke. 2020a. Modeling of reticular and ventral ruminal pH of lactating dairy cows using ingestion and rumination behavior. *J. Dairy Sci.* 103:7260–7275. <https://doi.org/10.3168/jds.2020-18195>.
- Mensinging, A., J. Hummel, and A. R. Sharifi. 2020b. Statistical modeling of ruminal pH parameters from dairy cows based on a meta-analysis. *J. Dairy Sci.* 103:750–767. <https://doi.org/10.3168/jds.2019-16802>.
- Mensinging, A., M. Zschiesche, J. Hummel, A. O. Schmitt, C. Grelet, and A. R. Sharifi. 2020c. An innovative concept for a multivariate plausibility assessment of synchronously recorded data. *Animals (Basel)* 10:1412. <https://doi.org/10.3390/ani10081412>.
- Mevik, B.-H., and R. Wehrens. 2007. The pls package: Principal component and partial least squares regression in R. *J. Stat. Softw.* 18:1–23. <https://doi.org/10.18637/jss.v018.i02>.
- Miettinen, H., and P. Huhtanen. 1996. Effects of the ratio of ruminal propionate to butyrate on milk yield and blood metabolites in dairy cows. *J. Dairy Sci.* 79:851–861. [https://doi.org/10.3168/jds.S0022-0302\(96\)76434-2](https://doi.org/10.3168/jds.S0022-0302(96)76434-2).
- Nakazawa, M. 2019. fmsb: Functions for medical statistics book with some demographic data, version 0.7.0: 1–61. <http://minato.sip21c.org/msb/>.
- Nocek, J. E. 1997. Bovine acidosis: Implications on laminitis. *J. Dairy Sci.* 80:1005–1028. [https://doi.org/10.3168/jds.S0022-0302\(97\)76026-0](https://doi.org/10.3168/jds.S0022-0302(97)76026-0).
- Oba, M., and M. S. Allen. 2000. Effects of brown midrib 3 mutation in corn silage on productivity of dairy cows fed two concentrations of dietary neutral detergent fiber: 2. Chewing activities. *J. Dairy Sci.* 83:1342–1349. [https://doi.org/10.3168/jds.S0022-0302\(00\)75001-6](https://doi.org/10.3168/jds.S0022-0302(00)75001-6).
- Oetzel, G. R. 2004. Monitoring and testing dairy herds for metabolic disease. *Vet. Clin. North Am. Food Anim. Pract.* 20:651–674. <https://doi.org/10.1016/j.cvfa.2004.06.006>.
- Palmquist, D. L. 2006. Milk fat: Origin of fatty acids and influence of nutritional factors thereon. In: *Advanced Dairy Chemistry: Lipids*. 3rd ed. P. F. Fox and P. L. H. McSweeney, ed. Springer Science+Business Media, New York, NY.
- R Core Team. 2019. R: A language and environment for statistical computing, Version 3.6.2. R Foundation for Statistical Computing, Vienna, Austria.
- Soyeurt, H., C. Bastin, F. G. Colinet, V. M.-R. Arnould, D. P. Berry, E. Wall, F. Dehareng, H. N. Nguyen, P. Dardenne, J. Schefers, J. Vandenplas, K. Weigel, M. Coffey, L. Théron, J. Detilleux, E. Reding, N. Gengler, and S. McParland. 2012. Mid-infrared prediction of lactoferrin content in bovine milk: Potential indicator of mastitis. *Animal* 6:1830–1838. <https://doi.org/10.1017/S1751731112000791>.
- Soyeurt, H., F. G. Colinet, V. M.-R. Arnould, P. Dardenne, C. Bertozzi, R. Renaville, D. Portetelle, and N. Gengler. 2007. Genetic variability of lactoferrin content estimated by mid-infrared spectrometry in bovine milk. *J. Dairy Sci.* 90:4443–4450. <https://doi.org/10.3168/jds.2006-827>.
- 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. <https://doi.org/10.3168/jds.2010-3408>.
- Stangaferro, M. L., R. Wijma, L. S. Caixeta, M. A. Al-Abri, and J. O. Giordano. 2016. Use of rumination and activity monitoring for the identification of dairy cows with health disorders: Part I. Metabolic and digestive disorders. *J. Dairy Sci.* 99:7395–7410. <https://doi.org/10.3168/jds.2016-10907>.
- Van Hertem, T., E. Maltz, A. Antler, C. E. B. Romanini, S. Viazzi, C. Bahr, A. Schlageter-Tello, C. Lokhorst, D. Berckmans, and I. Halachmi. 2013. Lameness detection based on multivariate continuous sensing of milk yield, rumination, and neck activity. *J. Dairy Sci.* 96:4286–4298. <https://doi.org/10.3168/jds.2012-6188>.
- Van Nespen, T., B. Vlaeminck, W. Wanzele, W. Van Straalen, and V. Fievez. 2005. Use of specific milk fatty acids as diagnostic tool for rumen acidosis in dairy cows. *Commun. Agric. Appl. Biol. Sci.* 70:277–280.
- Vanlierde, A., H. Soyeurt, N. Gengler, F. G. Colinet, E. Froidmont, M. Kreuzer, F. Grandl, M. Bell, P. Lund, D. W. Olijhoek, M. Eugène, C. Martin, B. Kuhla, and F. Dehareng. 2018. Short communication: Development of an equation for estimating methane emissions of dairy cows from milk Fourier transform mid-infrared spectra by using reference data obtained exclusively from respiration chambers. *J. Dairy Sci.* 101:7618–7624. <https://doi.org/10.3168/jds.2018-14472>.
- Vanlierde, A., M. L. Vanrobays, F. Dehareng, E. Froidmont, H. Soyeurt, S. McParland, E. Lewis, M. H. Deighton, F. Grandl, M. Kreuzer, B. Gredler, P. Dardenne, and N. Gengler. 2015. Hot topic: Innovative lactation-stage-dependent prediction of methane emissions from milk mid-infrared spectra. *J. Dairy Sci.* 98:5740–5747. <https://doi.org/10.3168/jds.2014-8436>.
- VDLUFA. 1993. Handbook for Methodology of Agricultural Experiments and Investigations (Book of Methods), Volume III: The Chemical Analysis of Feedstuffs. Association of German Agricultural Analytic and Research Institutes (VDLUFA), Speyer, Germany.
- Villot, C., B. Meunier, J. Bodin, C. Martin, and M. Silberberg. 2018. Relative reticulo-rumen pH indicators for subacute ruminal acidosis detection in dairy cows. *Animal* 12:481–490. <https://doi.org/10.1017/S1751731117001677>.
- Vlaeminck, B., V. Fievez, R. J. Tamminga, R. J. Dewhurst, A. van Vuuren, D. De Brabander, and D. Demeyer. 2006. Milk odd- and branched-chain fatty acids in relation to the rumen fermentation pattern. *J. Dairy Sci.* 89:3954–3964. [https://doi.org/10.3168/jds.S0022-0302\(06\)72437-7](https://doi.org/10.3168/jds.S0022-0302(06)72437-7).
- Wallén, S. E., E. Prestløkken, T. H. E. Meuwissen, S. McParland, and D. P. Berry. 2018. Milk mid-infrared spectral data as a tool to predict feed intake in lactating Norwegian Red dairy cows. *J. Dairy Sci.* 101:6232–6243. <https://doi.org/10.3168/jds.2017-13874>.
- Wiggans, G. R., and G. E. Shook. 1987. A lactation measure of somatic cell count. *J. Dairy Sci.* 70:2666–2672. [https://doi.org/10.3168/jds.S0022-0302\(87\)80337-5](https://doi.org/10.3168/jds.S0022-0302(87)80337-5).
- Williams, P. C., and D. C. Sobering. 1993. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. *J. Near Infrared Spectrosc.* 1:25–32. <https://doi.org/10.1255/jnirs.3>.

ORCID

- A. Mensing  <https://orcid.org/0000-0002-7302-9010>
 J. Hummel  <https://orcid.org/0000-0002-8876-7745>
 C. Grelet  <https://orcid.org/0000-0003-3313-485X>
 N. Gengler  <https://orcid.org/0000-0002-5981-5509>