Energy balance of dairy cows predicted by mid-infrared spectra data of milk using Bayesian approaches

Gabriel Rovere,1,∗ Gustavo de los Campos,2,3,4 Grum Gebreyesus,5 Rodrigo Pelicioni Savegnago,1 and Albert J. Buitenhuis5,∗

1Department of Animal Science, Michigan State University, East Lansing, MI 48824
2Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI 48824
3Institute for Quantitative Health Science and Engineering, Michigan State University, East Lansing, MI 48824
4Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824
5Center for Quantitative Genetics and Genomics, Aarhus University, DK-8000 Aarhus, Denmark

ABSTRACT

Information on dry matter intake (DMI) and energy balance (EB) at the animal and herd level is important for management and breeding decisions. However, routine recording of these traits at commercial farms can be challenging and costly. Fourier-transform mid-infrared (FT-MIR) spectroscopy is a noninvasive technique applicable to a large cohort of animals that is routinely used to analyze milk components and is convenient for predicting complex phenotypes that are typically difficult and expensive to obtain on a large scale. We aimed to develop prediction models for EB and use the predicted phenotypes for genetic analysis. First, we assessed prediction equations using 4,485 phenotypic records from 167 Holstein cows from an experimental station. The phenotypes available were body weight (BW), milk yield (MY) and milk components, weekly-averaged DMI, and FT-MIR data from all milk samples available. We implemented mixed models with Bayesian approaches and assessed them through 50 randomized replicates of a 5-fold cross-validation. Second, we used the best prediction models to obtain predicted phenotypes of EB (EBp) and DMI (DMIp) on 5 commercial farms with 2,365 phenotypic records of MY, milk components and FT-MIR data, and BW from 1,441 Holstein cows. Third, we performed a GWAS and estimated heritability and genetic correlations for energy content in milk (EnM), BW, DMIp, and EBp using the genomic information available on the cows from commercial farms. The highest correlation between the predicted and observed phenotype \( r_{yy} \) was obtained with DMI (0.88) and EB (0.86), while predicting BW was, as anticipated, more challenging (0.69). In our study, models that included FT-MIR information performed better than models without spectra information in the 3 traits analyzed, with increments in prediction correlation ranging from 5% to 10%. For the predicted phenotypes calculated by the prediction equations and data from the commercial farms the heritability ranged between 0.11 and 0.16 for EnM, DMIp and EBp, and 0.42 for BW. The genetic correlation between EnM and BW was −0.17, with DMIp was 0.40 and with EBp was −0.39. From the GWAS, we detected one significant QTL region for EnM, and 3 for BW, but none for DMIp and EBp. The results obtained in our study support previous evidence that FT-MIR information from milk samples contribute to improve the prediction equations for DMI, BW, and EB, and these predicted phenotypes may be used for herd management and contribute to the breeding strategy for improving cow performance. Key words: dry matter intake, body weight, spectrometry, high-throughput phenotyping, genome-wide association study

INTRODUCTION

Prediction of energy balance (EB) of individual cows can provide relevant information to monitor health, fertility, and general management of dairy herds. Most cows go through a period of negative EB in their first weeks of lactation. However, if either this period of negative EB is prolonged or the magnitude of the negative EB is excessive, or both, it can affect the animal’s fertility, generate metabolic disorders, and cause other health disturbances (Berglund and Danell, 1987). Although negative EB may have nutritional and genetic causes (Berglund and Danell, 1987), it is important to identify animals with negative EB at an early stage to prevent financial losses at the farm level due to reduced cow performance, increased sickness, and sometimes premature culling of animals.

Most studies defined EB as the difference between energy inputs and outputs estimated for the animals
such as milk fatty acids but auspicious enough to be observed for energy intake, DMI, and other phenotypes et al. (2011, 2012, 2014). In this sense, the prediction were reported in several contributions by McParland 2021). Predictions of EB from FT-MIR of milk samples de novo fatty acids (Bach et al., 2019; Rovere et al., 2021). Infrared spectrometry is a noninvasive technique applicable to big groups of animals, and therefore, convenient to predict complex phenotypes, which are difficult and expensive to obtain on a large scale. In this direction, several studies investigated the use of FT-MIR to predict residual feed intake and DMI (Shetty et al., 2017; Dórea et al., 2018), and technological properties of milk (Cecchinato et al., 2009, 2015), among others. Fourier-transform mid-infrared (FT-MIR) spectroscopy is used worldwide to predict traits routinely included in the dairy genetic evaluations (i.e., fat, protein, and lactose percentages), and in several studies have been used in the prediction of fine milk composition traits (Rutten et al., 2010, 2011; Fleming et al., 2017a,b, 2018), and technological properties of milk (Cecchinato et al., 2009, 2015), among others. Infrared spectrometry is a noninvasive technique applicable to big groups of animals, and therefore, convenient to predict complex phenotypes, which are difficult and expensive to obtain on a large scale. In this direction, several studies investigated the use of FT-MIR to predict residual feed intake and DMI (Shetty et al., 2017; Dórea et al., 2018). Lately, FT-MIR milk spectral data had been used to predict BW (Soyeurt et al., 2019; Zhang et al., 2021b), to predict the content of metabolites in milk related to negative EB, such as milk BHB, nonesterified fatty acid, and de novo fatty acids (Bach et al., 2019; Rovere et al., 2021). Predictions of EB from FT-MIR of milk samples were reported in several contributions by McParland et al. (2011, 2012, 2014). In this sense, the prediction accuracy for EB was lower than the prediction accuracy observed for energy intake, DMI, and other phenotypes such as milk fatty acids but auspicious enough to be used in the context of breeding programs (McParland et al., 2012; Ho et al., 2019). In particular, a positive medium-high genetic correlation between observed EB and FT-MIR-predicted EB was reported (McParland et al., 2014). We aimed to develop prediction models for EB and use the predicted phenotypes for genetic analysis. Then, in our study, we developed prediction models for EB with a specific interest in evaluating the inclusion of FT-MIR information. We considered EB as the result of energy inputs and outputs experienced by an animal at a specific moment. Traditionally, the proxy for energy inputs has been DMI, and the proxies for energy outputs have been milk yield (MY), milk components (fat, protein, and lactose), and BW. Because DMI and BW are not routinely recorded in commercial farms, we also assessed prediction models for these traits. Then, we assessed the prediction of EB using those predicted DMI and BW. We evaluated prediction models for EB considering information available at a commercial farm level combined with spectra information from labs, which are involved in the routine of milk recording. Partial least squares (PLS) regression has been the predominant method for inferring associations between FTIR absorbance and various traits (Bresolin and Dórea, 2020), however, Bayesian regression strategies have been increasingly considered for dairy science applications and variable selection-based procedures such as BayesB (Meuwissen et al., 2001; Pérez and De Los Campos, 2014) have been shown in several studies similar or better prediction accuracy than PLS (Ferragina et al., 2015; Bonfatti et al., 2017; Rovere et al., 2021). Despite the main objective of the study was to assess the inclusion of spectra data in the prediction models of complex phenotypes and not the comparison of methods, we included PLS models as a reference method as it was suggested during the review process. Finally, using genomic data, we explored the heritability and genetic associations of energy content in milk (EnM), BW, predicted DMI (DMIp) and predicted EB (EBp) using the prediction models developed in our study.

MATERIALS AND METHODS

All procedures to collect the Danish samples followed the protocols according to the National Guidelines for Animal Experimentation and the Danish Animal Experimental Ethics Committee. Infrared spectra were available from milk samples collected from routine records of dairy farms. Genotypes were available from farms, as part of the routine genotyping for genomic selection, and hence, no specific permission was required.
Data

Data were collected from the Danish Cattle Research Centre (DCRC; Foulum, Denmark) and from 5 Danish commercial farms (referred to as station and farms, respectively). Milk samples collected at DCRC and commercial herds were sent to the Eurofins-Steins laboratory (Vejen, Denmark) for FT-MIR spectral analyses using a MilkoScan FT+ (Foss, Hillerød, Denmark). The milk spectral data consisted of FT-MIR absorbance at 1,060 data points in the infrared wavenumber range from 5,008 and 925 cm⁻¹ (wavelength 2.00–10.76 μm). The following descriptions of both station and farms sets are of the final edited data used for the analyses (details on the editing can be found in Supplemental Annex S1; https://figshare.com/articles/journal_contribution/Annex_and_supplementary_Tables_Figures/24942360; Buitenhuis et al., 2024).

Station dataset consisted of 4,485 phenotypic records from 167 Holstein cows of parities 1 to 3, collected between May 2015 and December 2017 from 1 to 44 wk of lactation. Phenotypes for BW, MY, and milk components (%Fat, %Protein, and %Lactose), and DMI were measured as weekly averages. The spectra absorbances from all milk samples taken for the analysis of milk composition were available, totaling 18,238 records. When a cow had more than one record of spectra in a week, we used for the analyses the average of the absorbance of each wavenumber. Farms dataset consisted of 2,365 phenotypic records from 1,441 Holstein cows from 5 farms of parities 1 to 5, collected between April 2015 and July 2017, from 5 to 305 DIM. Test-day phenotypes were provided for BW, MY, and milk components (%Fat and %Protein). The FT-MIR spectra absorbances from respective milk sample were available for each test-day with phenotype recorded in this dataset.

The cows from farms dataset were genotyped with EuroG10K custom SNP chip, and genotypes were imputed to 50K using Minimax3 v.2.0.1 software (Das et al., 2016) as described in (Gebreyesus et al., 2021). A reference population of 6,077 bulls, including sires of the imputed cows, that were genotyped with 50K SNP chip was used for the imputation. Before the imputation, 50K genotypes in the reference population were subjected to quality control using the minor allele frequency threshold of 0.05. Imputation accuracy was evaluated for a subset of selected reference individuals as the Pearson correlation between observed and imputed genotypes (coded as 0, 1, or 2). Only SNPs with correlation higher than 0.80 were used in the analyses. Ultimately, 44,305 SNPs remained for the analyses. The SNPs used in the analysis had mean imputation accuracy of 0.979 with a standard deviation of 0.04.

EB Calculation

We calculated EB for the cows in station using the energy inputs and outputs approach. This EB calculated was considered as the observed EB to assess the prediction models. Energy balance was calculated as \( EB = Ein - Eout \); where \( Ein \) was the daily energy intake, and \( Eout \) was the energy used by the cow for milk production and maintenance. The \( Ein \) was calculated using the weekly average DMI times the energy per kilogram of the TMR provided, which was on average 6.71 MJ/kg of DM. The net energy requirement for maintenance was calculated based on the metabolic BW as \( BW^{0.75} \times 0.08 \) (NRC, 2001). The net energy for kilograms of MY was calculated as \((0.0929 \times %\text{Fat}) + (0.0588 \times %\text{Protein}) + (0.0395 \times %\text{Lactose})\) (NRC, 2001). The net energy for maintenance and content in MY was expressed in MJ using a factor conversion of 4.19 J/cal (Thompson and Taylor, 2008).

Prediction Models Assessed

First, we assessed a set of 7 simple models, which included alternatively the effect of the parity number, week of lactation, MY, or FT-MIR absorbances (Table 1).

In these equations, phenotypes DMI, BW, and EB are expressed as \( y \), \( \mu \) is an intercept, \( \text{parity}_j \) indicates the \( j \)th parity (modeled as a factor with 3 levels), \( \text{week}_k \) indicates the \( k \)th week (modeled as a factor with 44 levels), \( \text{MY}_l \) is the individual average milk yield in level \( l \).

<table>
<thead>
<tr>
<th>Model</th>
<th>Formula¹</th>
<th>FT-MIR regression coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>( y_i = \mu + \text{parity}_{j} + \varepsilon_i )</td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>( y_i = \mu + \text{week}_{k} + \varepsilon_i )</td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>( y_i = \mu + \text{MY}_{l} + \varepsilon_i )</td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>( y_i = \mu + \sum_{n=1}^{1800} W_n b_n + \varepsilon_i )</td>
<td>Bayes B</td>
</tr>
<tr>
<td>M5</td>
<td>( y_i = \mu + \sum_{n=1}^{1800} W_n b_n + \varepsilon_i )</td>
<td>Bayesian ridge regression</td>
</tr>
<tr>
<td>M6</td>
<td>( y_i = \mu + \sum_{n=1}^{1800} W_n b_n + \varepsilon_i )</td>
<td>PLS²</td>
</tr>
<tr>
<td>M7</td>
<td>( y_i = \mu + \sum_{n=1}^{351} W_n b_n + \varepsilon_i )</td>
<td>PLS</td>
</tr>
</tbody>
</table>

¹\( y \) represents the phenotypes DMI, BW, and energy balance; \( \mu \) is an intercept; \( \text{parity}_j \) indicates the \( j \)th parity (modeled as a factor with 3 levels); \( \text{week}_k \) indicates the \( k \)th week (modeled as a factor with 44 levels); \( \text{MY}_l \) is the individual average milk yield in week \( l \) (modeled as a continuous covariate); \( W_n \) and \( b_n \) are the \( n \)th wavenumber of the spectra and its effect, respectively, and \( \varepsilon_i \sim N(0, \sigma^2) \) is a normally identically and independently distributed error term associated with each observation.

²PLS = partial least squares.
week \( k \) (modeled as a continuous covariate), \( W_s \), and \( b_s \) are the \( s \)th wavenumber of the spectra and its effect, respectively, and \( \varepsilon^{uid} \sim N \left(0, \sigma^2 \right) \) is a normally identically and independently distributed error term associated to each observation.

Models M4 and M5 were Bayesian models implemented using Markov chain Monte Carlo (MCMC) methods as provided in the BGLR statistical package (Pérez and De Los Campos, 2014) R package (https://www.r-project.org/). We considered 2 different priors for regression coefficients: (1) in the case of M4, a spike-and-slab prior with a point of mass at zero and at a scaled-t-slab, in which \( b_s^{uid} \sim (1 - \pi) I_{b_s=0} + \pi I_{b_s=0} \), here \( \pi \) is the scaled-t slab with \( \pi \) degrees of freedom, known as BayesB (BB; Meuwissen et al., 2001); and (2) in the case of M5, Gaussian priors, in which case \( b_s^{uid} \sim N \left(0, \sigma^2 \right) \), known as Bayesian ridge regression (BRR; Hoerl and Kennard, 2000). In the case of model M6, PLS regressions method (Wold et al., 1983) was implemented using the PLS R package (Mevik and Wehrens, 2007). We used a 10-fold random cross-validation to determine the optimal number of components in the training set, allowing a maximum of 30 components. Subsequently, we used a PLS regression with the optimal number of components to predict the phenotypes in testing sets. The optimal number of components was defined by the model with fewest components and was yet still less than 1 standard error away from the overall best model for predictive correlation (Rovere et al., 2021). Model M7 is equal to M6 with the exception that from the spectra, 2 regions associated to water absorption, O-H stretching (between \( \sim 1,060 \text{ cm}^{-1} \) and \( \sim 1,700 \text{ cm}^{-1} \)) and O-H bending (\( >3,005 \text{ cm}^{-1} \)) were omitted, leaving 517 wavenumbers (PLS517). Although wavenumbers in these 2 regions are generally removed in studies on prediction using PLS, there is evidence that these regions contain valuable information (Wang et al., 2016; Wang and Bovenhuis, 2018) and they were used in the prediction models using Bayesian methods (Toledo-Alvarado et al., 2018; Rovere et al., 2019, 2021). However, in applications such evaluation, BGLR assigns the prior distributions \( \pi \sim \text{beta}(5,5) \), and \( \sigma^2, \sigma^2 \sim \chi^2_5 \), with a prior assumption that the random effects \( b \), explain 50% of the total phenotypic variance. We refer the reader to (Pérez and De Los Campos, 2014) for further details about the default settings of BGLR. The MCMC generated 45,000 samples, of which the first 20,000 were discarded for burn-in, and the remaining 25,000 samples were thinned at an interval of 5 samples.

Third, most of farmers have available information on individual \%Fat and \%Protein predicted by FT-MIR, we assessed the performance of the prediction models including \%Fat and \%Protein to evaluate whether FT-MIR information has an additional contribution to the prediction accuracy of the prediction models. Six models with milk components (MC; numbered MC1 to MC5) were assessed in Table 3. The references for these models are the same that were described in previous paragraphs and \%Fatm and \%Protn are \%Fat and %Protein modeled as continuous covariates.

Additionally, we assessed the prediction of DMI using Equations 1 and 2, proposed by (NRC, 2001) and by (de Souza et al., 2019), respectively, and later we used these predictions to calculate EB:

\[
\text{DMI (kg/d)} = (0.372 \times \text{FCM} + 0.0968 \times \text{BW}^{0.75}) \\
+ (1 - e^{-0.192 \times (\text{BW} + 3.67)}) \tag{1}
\]

\[
\text{DMI (kg/d)} = [(3.7 + \text{Parity} \times 5.7) + 0.305 \times \text{EnM} + 0.022 \times \text{BW} + (-0.689 + \text{Parity} \times -1.87) \times \text{BCS} \\
\times [1 - (0.212 + \text{Parity} \times 0.136) \times e^{-(0.053 \times \text{DIM})}] \\
+ \text{Parity} - e^{-0.192 \times (\text{WL} + 3.67)}] \tag{2}
\]

In these equations, FCM is 4% fat-corrected milk (kg/d), BW is body weight, WL is week of lactation, Parity is 0 for primiparous and 1 otherwise, EnM is energy in milk (National Research Council, 2001), BCS is body condition score, and DIM is days in milk. Body condition score was recorded fortnight, and it was interpolated to have weekly values; DIM was assigned as the middle day of the corresponding week of lactation; FCM was calculated following NRC (2001) as \( 0.4 \times \text{MY} + 15 \times (\%\text{Fat}/100) \times \text{MY} \).

All analyses were conducted in R programming language (R Core Team, 2020), with the Bayesian models fitted using the BGLR package (Pérez and De Los Campos, 2014). By default, BGLR assigns the prior distributions \( \pi \sim \text{beta}(5,5) \), and \( \sigma^2, \sigma^2 \sim \chi^2_5 \), with a prior assumption that the random effects \( b \), explain 50% of the total phenotypic variance. We refer the reader to (Pérez and De Los Campos, 2014) for further details about the default settings of BGLR. The MCMC generated 45,000 samples, of which the first 20,000 were discarded for burn-in, and the remaining 25,000 samples were thinned at an interval of 5 samples.

The performance of the prediction models was assessed using the Pearson correlation between the predicted and observed phenotype \( r_{\hat{y},y} \) and the root of the prediction mean square error (RPMSE), defined as
Table 2. Models that combine the effects of parity, week of lactation, and milk yield and the effects of the absorbance of each wavenumber from the milk spectra (Fourier-transform mid-infrared, FT-MIR), solved by different methods

<table>
<thead>
<tr>
<th>Model</th>
<th>Formula¹</th>
<th>FT-MIR regression coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>M8</td>
<td>( y_{ijk} = \mu + \text{parity}_j + \text{week}<em>k + \varepsilon</em>{ijk} )</td>
<td></td>
</tr>
<tr>
<td>M9</td>
<td>( y_{ijkl} = \mu + \text{parity}_j + \text{week}_k + \text{MY}<em>l + \varepsilon</em>{ijkl} )</td>
<td></td>
</tr>
<tr>
<td>M10</td>
<td>( y_{ijk} = \mu + \text{parity}<em>j + \sum</em>{s=1}^{N} W_{ijk} b_s + \varepsilon_{ijk} )</td>
<td>Bayes B</td>
</tr>
<tr>
<td>M11</td>
<td>( y_{ijk} = \mu + \text{parity}<em>j + \sum</em>{s=1}^{N} W_{ijk} b_s + \varepsilon_{ijk} )</td>
<td>Bayesian ridge regression</td>
</tr>
<tr>
<td>M12</td>
<td>( y_{ijk} = \mu + \text{parity}<em>j + \sum</em>{s=1}^{N} W_{ijk} b_s + \varepsilon_{ijk} )</td>
<td>PLS²</td>
</tr>
<tr>
<td>M13</td>
<td>( y_{ijk} = \mu + \text{parity}<em>j + \sum</em>{s=1}^{N} W_{ijk} b_s + \varepsilon_{ijk} )</td>
<td>PLS</td>
</tr>
<tr>
<td>M14</td>
<td>( y_{ijk} = \mu + \text{parity}<em>j + \text{week}<em>k + \sum</em>{s=1}^{N} W</em>{ijk} b_s + \varepsilon_{ijk} )</td>
<td>Bayes B</td>
</tr>
<tr>
<td>M15</td>
<td>( y_{ijk} = \mu + \text{parity}<em>j + \text{week}<em>k + \sum</em>{s=1}^{N} W</em>{ijk} b_s + \varepsilon_{ijk} )</td>
<td>Bayesian ridge regression</td>
</tr>
<tr>
<td>M16</td>
<td>( y_{ijk} = \mu + \text{parity}<em>j + \text{week}<em>k + \sum</em>{s=1}^{N} W</em>{ijk} b_s + \varepsilon_{ijk} )</td>
<td>PLS</td>
</tr>
<tr>
<td>M17</td>
<td>( y_{ijk} = \mu + \text{parity}<em>j + \text{week}<em>k + \sum</em>{s=1}^{N} W</em>{ijk} b_s + \varepsilon_{ijk} )</td>
<td>PLS</td>
</tr>
<tr>
<td>M18</td>
<td>( y_{ijk} = \mu + \text{parity}<em>j + \text{week}<em>k + \text{MY}<em>l + \sum</em>{s=1}^{N} W</em>{ijk} b_s + \varepsilon</em>{ijk} )</td>
<td>Bayes B</td>
</tr>
<tr>
<td>M19</td>
<td>( y_{ijk} = \mu + \text{parity}<em>j + \text{week}<em>k + \text{MY}<em>l + \sum</em>{s=1}^{N} W</em>{ijk} b_s + \varepsilon</em>{ijk} )</td>
<td>Bayesian ridge regression</td>
</tr>
<tr>
<td>M20</td>
<td>( y_{ijk} = \mu + \text{parity}<em>j + \text{week}<em>k + \text{MY}<em>l + \sum</em>{s=1}^{N} W</em>{ijk} b_s + \varepsilon</em>{ijk} )</td>
<td>PLS²</td>
</tr>
<tr>
<td>M21</td>
<td>( y_{ijk} = \mu + \text{parity}<em>j + \text{week}<em>k + \text{MY}<em>l + \sum</em>{s=1}^{N} W</em>{ijk} b_s + \varepsilon</em>{ijk} )</td>
<td>PLS</td>
</tr>
</tbody>
</table>

¹\( y \) represents the phenotypes DMI, BW, and energy balance; \( \mu \) is an intercept; \( \text{parity}_j \) indicates the \( j \)th parity (modeled as a factor with 3 levels); \( \text{week}_k \) indicates the \( k \)th week (modeled as a factor with 44 levels); \( \text{MY}_l \) is the individual average milk yield in \( \text{week}_k \) (modeled as a continuous covariate); \( W_{ijk} \) and \( b_s \) are the \( s \)th wavenumber of the spectra and its effect, respectively, and \( \varepsilon_{ijkl} \sim N(0,\sigma^2) \) is a normally identically and independently distributed error term associated with each observation.

²PLS = partial least squares.

\[ RPMSE = \sqrt{\frac{\sum_{i=1}^{N_{test}} (y_{i} - \hat{y}_{i})^2}{N_{test}}} \]

We iterated over 50 randomized replicates of a 5-fold cross-validation. In each iteration, cows and all their records were randomly assigned to one of the 5 folds to ensure that validation was cow-independent (\( CI_{cv} \)). Then, animals were assigned to training and test groups using the leave-one-fold-out method. The estimated coefficients from

Table 3. Models with milk components (MC) that combine the effects of parity, week of lactation, milk yield (MY), and milk content of fat and protein (\( \%\text{Fat} \) and \( \%\text{Prot} \), respectively) and the effects of the absorbance of each wavenumber from the milk spectra (Fourier-transform mid-infrared, FT-MIR), solved by different methods

<table>
<thead>
<tr>
<th>Model</th>
<th>Formula¹</th>
<th>FT-MIR regression coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC1</td>
<td>( y_{ijkl} = \mu + \text{week}_k + \text{MY}_l + %\text{Fat}_m + %\text{Prot}<em>n + \varepsilon</em>{ijkl} )</td>
<td></td>
</tr>
<tr>
<td>MC2</td>
<td>( y_{ijkl} = \mu + \text{parity}_j + \text{week}_k + \text{MY}_l + %\text{Fat}_m + %\text{Prot}<em>n + \varepsilon</em>{ijkl} )</td>
<td></td>
</tr>
<tr>
<td>MC3</td>
<td>( y_{ijkl} = \mu + \text{parity}_j + \text{week}_k + \text{MY}<em>l + %\text{Fat}<em>m + %\text{Prot}<em>n + \sum</em>{s=1}^{N} W</em>{ijkl} b_s + \varepsilon</em>{ijkl} )</td>
<td>Bayes B</td>
</tr>
<tr>
<td>MC4</td>
<td>( y_{ijkl} = \mu + \text{parity}_j + \text{week}_k + \text{MY}<em>l + %\text{Fat}<em>m + %\text{Prot}<em>n + \sum</em>{s=1}^{N} W</em>{ijkl} b_s + \varepsilon</em>{ijkl} )</td>
<td>Bayesian ridge regression</td>
</tr>
<tr>
<td>MC5</td>
<td>( y_{ijkl} = \mu + \text{parity}_j + \text{week}_k + \text{MY}<em>l + %\text{Fat}<em>m + %\text{Prot}<em>n + \sum</em>{s=1}^{N} W</em>{ijkl} b_s + \varepsilon</em>{ijkl} )</td>
<td>PLS²</td>
</tr>
<tr>
<td>MC6</td>
<td>( y_{ijkl} = \mu + \text{parity}_j + \text{week}_k + \text{MY}<em>l + %\text{Fat}<em>m + %\text{Prot}<em>n + \sum</em>{s=1}^{N} W</em>{ijkl} b_s + \varepsilon</em>{ijkl} )</td>
<td>PLS</td>
</tr>
</tbody>
</table>

¹\( y \) represents the phenotypes DMI, BW, and energy balance; \( \mu \) is an intercept; \( \text{parity}_j \) indicates the \( j \)th parity (modeled as a factor with 3 levels); \( \text{week}_k \) indicates the \( k \)th week (modeled as a factor with 44 levels); \( \text{MY}_l \) is the individual average milk yield in \( \text{week}_k \) (modeled as a continuous covariate); \( W_{ijkl} \) and \( b_s \) are the \( s \)th wavenumber of the spectra and its effect, respectively, and \( \varepsilon_{ijkl} \sim N(0,\sigma^2) \) is a normally identically and independently distributed error term associated with each observation.

²PLS = partial least squares.

the training dataset (4 folds) were used to predict the phenotypes responses on the testing fold. For each fold of each randomized replicated, we calculated \( r_{\hat{y}y} \) and RPMSE. The Tukey’s honestly significant difference test (\( \alpha = 0.05 \)) was used to determine the significance of the differences between models. For that, we analyze the variability of the results applying a model that considered the effect of prediction model used, the fold predicted and the interaction between these 2 effects. To ensure correctness of the test in \( r_{\hat{y}y} \), Fisher’s z-transformation was applied to \( r_{\hat{y}y} \) to perform the test. The test was implemented using the R package Agricolae (de Mendiburu, 2019). Because observed phenotypes of BW were available in both datasets, we performed a herd-independent cross-validation (HICV) for this trait. In this case, one herd was left out as the test data with data from the other 5 herds used as the training data. This was repeated 6 times (i.e., 6 folds), each time using a different herd as a testing set. The estimated coefficients from the training dataset were used to predict phenotypes responses on the remaining test data for the herd that was left out. To assess the performance of the models we calculated \( r_{\hat{y}y} \) and RPMSE as it was previously described.

**Variance Component Analysis**

Heritabilities and genetic correlations were estimated using the farms dataset. The traits EnM, BW, DMip, and EBp were analyzed using 1,347 records for MY, %Fat, and %Protein from 824 cows, and 1,273 records for BW for 805 cows (in all the cases daily records). The trait EnM was calculated using the information of components available in this dataset as \((0.0929 \times \%\text{Fat}) + (0.0588 \times \%\text{Protein}) + 0.192 \) (NRC, 2001), and expressed in MJ as it was indicated previously in this section. Because DMI and EB phenotypes were not measured in farms set, predicted phenotypes were used for these traits. Predicted phenotypes were obtained using the best prediction model assessed in the station dataset. The variance components for the 4 traits were estimated using a single trait model using a genomic relationship matrix with all genotypes. Genetic correlations between traits were estimated through bivariate analyses. The model for the univariate and bivariate analyses was defined as follows:

\[
y = Xb + Z_g g + Z_p u + e,
\]

where \( b \) is the vector of coefficients of the fixed effects and \( X \) its designed matrix. The nongenetic effects treated as fixed were the combine effect of the herd-year-season (16 levels), the effect of the number of parity (5 levels) and stage of lactation (21 fortnight periods); \( g \) is the vector of the genomic breeding values \( g \sim N(0, G \sigma_g^2) \), where \( G \) is the genomic relationship matrix built on the centered genotypes divided by total SNP variance as per method 1 in (VanRaden, 2008). \( u \) is the vector of the permanent cow effect \( u \sim N(0, I \sigma_p^2) \), \( Z_g \) and \( Z_p \) are their respective design matrices, and \( e \sim N(0, I \sigma_e^2) \) are the random residuals. For the variance estimation analyses, we used a Gibb sampler implementation of Bayesian reproducing kernel Hilbert space (de los Campos et al., 2009), which is mathematically identical to perform a genomic BLUP (VanRaden, 2008). For each analysis, we computed posterior means, posterior standard deviations and 95% credibility regions for each of the variance parameters and for the heritability \((h^2)\) defined as the proportion of total variance explained by additive genomic effects:

\[
h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_p^2 + \sigma_e^2},
\]

where \( \sigma_g^2, \sigma_p^2, \) and \( \sigma_e^2 \) are the variance terms that were previously defined.

Within the framework of the general objective of this research, we carried out a genetic association study using the farm set, which is fully presented in the Supplemental Annex S2 (https://figshare.com/articles/journal_contribution/Annex_and_supplementary_Tables_Figures/24942360; Buitenhuis et al., 2024).

**RESULTS**

**Descriptive Statistics of the Data Analyzed**

Table 4 shows the general descriptive statistics of the 2 datasets used in this study. A more detailed information on the mean and standard deviation of the traits, and number of records across the lactation are presented in Supplemental Figure S1a and S1b (https://figshare.com/articles/journal_contribution/Annex_and_supplementary_Tables_Figures/24942360; Buitenhuis et al., 2024). The records available for the analysis were well distributed along the different lactation stages in both datasets, although in the station set the number of records decreased in the last weeks of lactation. In both datasets, MY increased during the first weeks of lactation and after ~10th week decreased until the end of the lactation. As it was expected, %Fat and %Protein had their highest concentrations in the first week of lactation, afterward dropping to a mini-
Table 4. Descriptive statistics (mean with SD in parentheses) of the variables analyzed by datasets station and farms

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Parity</th>
<th>N</th>
<th>Cows</th>
<th>MY (kg)</th>
<th>%Fat</th>
<th>%Prot</th>
<th>%Lact</th>
<th>DMI (kg/d)</th>
<th>BW (kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Station</td>
<td>1</td>
<td>1,560</td>
<td>124</td>
<td>31.91 (5.83)</td>
<td>4.07 (0.69)</td>
<td>3.60 (0.37)</td>
<td>4.98 (0.38)</td>
<td>20.03 (2.78)</td>
<td>644.09 (60.13)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2,010</td>
<td>115</td>
<td>40.78 (7.85)</td>
<td>3.99 (0.70)</td>
<td>3.54 (0.37)</td>
<td>4.81 (0.38)</td>
<td>24.11 (3.13)</td>
<td>693.67 (64.14)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>915</td>
<td>59</td>
<td>43.61 (8.37)</td>
<td>4.15 (0.79)</td>
<td>3.47 (0.38)</td>
<td>4.75 (0.38)</td>
<td>24.73 (3.70)</td>
<td>718.14 (49.73)</td>
</tr>
<tr>
<td>Farms</td>
<td>1</td>
<td>1,084</td>
<td>695</td>
<td>31.90 (6.69)</td>
<td>4.09 (0.77)</td>
<td>3.51 (0.31)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>555</td>
<td>429</td>
<td>40.51 (10.58)</td>
<td>3.93 (0.86)</td>
<td>3.49 (0.37)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>390</td>
<td>291</td>
<td>41.43 (10.69)</td>
<td>4.10 (0.95)</td>
<td>3.50 (0.39)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>241</td>
<td>171</td>
<td>42.34 (10.07)</td>
<td>3.93 (0.87)</td>
<td>3.40 (0.34)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>95</td>
<td>71</td>
<td>41.32 (11.16)</td>
<td>4.05 (0.90)</td>
<td>3.50 (0.28)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

1 N = total number of observations; cows = number of cows; MY = milk yield; %Fat = fat percentage; %Prot = protein percentage; %Lact = lactose percentage.
2 Not measured.

Prediction Model Assessment

In a first set of analyses, we tested models with a unique covariate among parity, week of lactation, and MY (which are available at the farm level) to predict DMI, BW, and EB, and models with only spectra information, analyzed by different methods (Bayes methods and PLS). Among the covariates parity, week of lactation, and MY, MY predicted better DMI, parity did better for BW, and stage of lactation (week of lactation) was best for EB (Table 5, models M1–M3). Models that included only spectra information had better prediction accuracy and lower RPMSE, with a lower improvement over the previous models in the case of BW, which had low prediction accuracy in all the cases (Table 5, models M4–M7). The prediction accuracy improved for the 3 phenotypes when the 3 covariates were included in the prediction model (Table 5, M9). Afterwards, we assessed models that considered the 3 effects previously mentioned and FT-MIR information in different combinations and methods for DMI, BW, and EB in the station set. These results are shown in Table 5. For the 3 variables, the models that included the random effect of the absorbance of the wavenumbers of FT-MIR, achieved higher prediction accuracy ($r_{yy}$, and RPMSE). The highest prediction accuracy was obtained in DMI (0.88) and EB (0.80) and lowest in BW (0.69); however, the models that included FT-MIR information performed statistically better in the 3 traits with increments of 5% to 10% in $r_{yy}$ compared with models without FT-MIR information. Predicting BW was, as anticipated, more challenging; however, models achieved a better prediction performance when both datasets were joined for the analysis, being the models that included FT-MIR, the ones with highest prediction accuracy (Table 5). Between methods, there was not markedly difference between Bayesian methods and PLS; however, BB tended to have better results but without relevant differences with PLS applied to a spectrum without the water absorption regions (PLS$_{0.17}$). The RPMSE was generally lower for the prediction models with highest $r_{yy}$, although the differences between models were less clear in BW (Table 5). When we assessed the models to predict BW using a HICv, the results of $r_{yy}$ were similar to the ones obtained using a CICv using both datasets (station + farms). Even when models with spectra information tended to be signifi-
significantly better than the models without it, the differences were debatable (Table 6). Conversely, greater RPMSE was observed in all the models when HIcv was used, and no significant differences were detected among them. Beside that from the analysis of variance we observed differences in the prediction accuracy between herds, being the lowest the ones obtained for herd E and station (~0.58–60), and highest for herds A and B (~0.70–0.74), while herds C and D showed intermediate values (~0.61–0.71). Nevertheless, in any of the analysis the interaction between prediction model and fold predicted was significant, indicating that better models were better along all folds despite different level of accuracy achieved.

Table 7 shows the results for prediction models when the covariates %Fat and %Protein, from DHIA labs, were included additionally to MY, with and without the inclusion of FT-MIR information.

We included in this analysis, models M18 to M21 to assess the significance of the difference between models

Table 5. Correlations between observed and predicted phenotypes in testing set \( r_{\hat{y}, \hat{y}} \) and square root of prediction mean square error (RPMSE) for different models assessed in 50 randomizations of a 5-fold cow-independent cross-validation for the Holstein breed.1

<table>
<thead>
<tr>
<th>Model</th>
<th>Station</th>
<th>BW: Station</th>
<th>BW: Station + farms</th>
<th>EB: Station</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1: Parity</td>
<td>0.54c</td>
<td>3.16c</td>
<td>0.43b</td>
<td>60.61b</td>
</tr>
<tr>
<td>M2: Week of lactation</td>
<td>0.49d</td>
<td>3.29d</td>
<td>0.35d</td>
<td>62.87c</td>
</tr>
<tr>
<td>M3: Milk</td>
<td>0.64c</td>
<td>2.88c</td>
<td>0.09c</td>
<td>66.78c</td>
</tr>
<tr>
<td>M4: Spectra-BB</td>
<td>0.65a</td>
<td>2.80a</td>
<td>0.48a</td>
<td>59.72c</td>
</tr>
<tr>
<td>M5: Spectra-BRR</td>
<td>0.64c</td>
<td>2.89c</td>
<td>0.46c</td>
<td>59.30c</td>
</tr>
<tr>
<td>M6: Spectra-PLS1060</td>
<td>0.62b</td>
<td>2.97b</td>
<td>0.44b</td>
<td>61.00b</td>
</tr>
<tr>
<td>M7: Spectra-PLS517</td>
<td>0.64a</td>
<td>2.91ab</td>
<td>0.47b</td>
<td>59.83b</td>
</tr>
</tbody>
</table>

\[a–d\]Letters indicate groups of models with significant differences tested with Tukey’s honestly significant test with alpha = 0.05. Comparisons are done by column.

1EB = energy balance; BB = Bayes B; BRR = Bayesian ridge regression; PLS1060 = partial least squares on 1,060 wavenumbers; PLS517 = partial least squares on 517 wavenumbers.

Table 6. Correlations between observed and predicted phenotypes in testing set \( r_{\hat{y}, \hat{y}} \) and square root of prediction mean square error (RPMSE) for different model assessed in 50 randomizations of a 5-fold cow-independent cross-validation for the Holstein breed.1

<table>
<thead>
<tr>
<th>Model</th>
<th>Station</th>
<th>BW: Station</th>
<th>BW: Station + farms</th>
<th>EB: Station</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M8: Parity + WL</td>
<td>0.75d</td>
<td>2.49d</td>
<td>0.60d</td>
<td>54.36bcde</td>
</tr>
<tr>
<td>M9: Parity + WL + MY</td>
<td>0.84d</td>
<td>2.06b</td>
<td>0.59d</td>
<td>54.63bcde</td>
</tr>
<tr>
<td>M10: Parity + Spectra-BB</td>
<td>0.75d</td>
<td>2.50d</td>
<td>0.57d</td>
<td>55.86d</td>
</tr>
<tr>
<td>M11: Parity + Spectra-BRR</td>
<td>0.75d</td>
<td>2.51d</td>
<td>0.56d</td>
<td>55.99d</td>
</tr>
<tr>
<td>M12: Parity + Spectra-PLS1060</td>
<td>0.73c</td>
<td>2.57c</td>
<td>0.55d</td>
<td>56.22c</td>
</tr>
<tr>
<td>M13: Parity + Spectra-PLS517</td>
<td>0.75d</td>
<td>2.51d</td>
<td>0.57d</td>
<td>55.39bcde</td>
</tr>
<tr>
<td>M14: Parity + WL + Spectra-BB</td>
<td>0.78c</td>
<td>2.37</td>
<td>0.62c</td>
<td>53.40abcde</td>
</tr>
<tr>
<td>M15: Parity + WL + Spectra-BRR</td>
<td>0.78c</td>
<td>2.38c</td>
<td>0.61c</td>
<td>53.67b</td>
</tr>
<tr>
<td>M16: Parity + WL + Spectra-PLS1060</td>
<td>0.77c</td>
<td>2.41c</td>
<td>0.61d</td>
<td>53.59bcde</td>
</tr>
<tr>
<td>M17: Parity + WL + Spectra-PLS517</td>
<td>0.78c</td>
<td>2.37c</td>
<td>0.63b</td>
<td>52.47a</td>
</tr>
<tr>
<td>M18: Parity+WL+Milk+Spectra-BB</td>
<td>0.88a</td>
<td>1.76c</td>
<td>0.63b</td>
<td>52.75ab</td>
</tr>
<tr>
<td>M19: Parity+WL+Milk+Spectra-BRR</td>
<td>0.88a</td>
<td>1.76c</td>
<td>0.62b</td>
<td>52.99bcde</td>
</tr>
<tr>
<td>M20: Parity+WL+Milk+Spectra-PLS1060</td>
<td>0.88a</td>
<td>1.79c</td>
<td>0.62b</td>
<td>53.14abcde</td>
</tr>
<tr>
<td>M21: Parity+WL+Milk+Spectra-PLS517</td>
<td>0.88a</td>
<td>1.77c</td>
<td>0.63b</td>
<td>52.17c</td>
</tr>
</tbody>
</table>

\[a–h\]Letters indicate groups of models with significant differences tested with Tukey’s honestly significant test with alpha = 0.05. Comparisons are done by column.

1EB = energy balance; WL = week of lactation; MY = milk yield; BB = BayesB; BRR = Bayesian ridge regression; PLS1060 = partial least squares on 1,060 wavenumbers; PLS517 = partial least squares on 517 wavenumbers.
with MY and models with MY plus components. In all the cases, the inclusion of spectra information improved the prediction accuracy of the models (M18–M21 and MC3–MC6 in Table 8). In DMI, the $r_{y,y}$ was higher when FT-MIR was included to the basic model of parity and week of lactation and did not improve it by adding the FT-MIR prediction of milk components. Similar results were observed in BW using only the station set. In contrast, for EB and BW (station + farms sets) the $r_{y,y}$ was slightly higher in models that included %Fat and %Protein, and FT-MIR information (MC3 and MC4 in Table 8). These models showed lower RPMSE as well. For DMI and BW (station), the RPMSE was lowest in the models that used BB, and although no differences were observed between BB and BRR in EB, in BW (station + farms), BRR presented smaller RPMSE. Bayes models mostly overperformed PLS, however these differences were less clear in RPMSE mostly with PLS517 (PLS on 517 wavenumbers).

Table 7. Correlations between observed and predicted phenotypes of BW in testing set \( \{r_{y,y}\} \) and square root of prediction mean square error (RPMSE) for different model assessed in one-herd-out cross-validation for the Holstein breed (6 herds)

<table>
<thead>
<tr>
<th>Model(^1)</th>
<th>(r_{y,y})</th>
<th>RPMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>M8: Parity + WL</td>
<td>0.65(^{a,b})</td>
<td>66.14(^{a})</td>
</tr>
<tr>
<td>M9: Parity + WL + MY</td>
<td>0.65(^{a,b})</td>
<td>66.10(^{a})</td>
</tr>
<tr>
<td>M10: Parity + Spectra-BB</td>
<td>0.61(^c)</td>
<td>70.03(^{a})</td>
</tr>
<tr>
<td>M11: Parity + Spectra-BRR</td>
<td>0.59(^d)</td>
<td>70.79(^{a})</td>
</tr>
<tr>
<td>M12: Parity + Spectra-PLS(_{1060})</td>
<td>0.56(^d)</td>
<td>71.05(^{a})</td>
</tr>
<tr>
<td>M13: Parity + Spectra-PLS(_{517})</td>
<td>0.60(^d)</td>
<td>69.56(^{a})</td>
</tr>
<tr>
<td>M14: Parity + WL + Spectra-BB</td>
<td>0.66(^c)</td>
<td>65.77(^{a})</td>
</tr>
<tr>
<td>M15: Parity + WL + Spectra-BRR</td>
<td>0.65(^{a,b})</td>
<td>66.57(^{a})</td>
</tr>
<tr>
<td>M16: Parity + WL + Spectra-PLS(_{1060})</td>
<td>0.62(^c)</td>
<td>68.96(^{a})</td>
</tr>
<tr>
<td>M17: Parity + WL + Spectra-PLS(_{517})</td>
<td>0.65(^{a,b})</td>
<td>65.94(^{a})</td>
</tr>
<tr>
<td>M18: Parity + WL + Milk + Spectra-BB</td>
<td>0.67(^c)</td>
<td>64.53(^{a})</td>
</tr>
<tr>
<td>M19: Parity + WL + Milk + Spectra-BRR</td>
<td>0.65(^{a,b})</td>
<td>65.15(^{a})</td>
</tr>
<tr>
<td>M20: Parity + WL + Milk + Spectra-PLS(_{1060})</td>
<td>0.62(^c)</td>
<td>68.01(^{a})</td>
</tr>
<tr>
<td>M21: Parity + WL + Milk + Spectra-PLS(_{517})</td>
<td>0.66(^{a})</td>
<td>64.93(^{a})</td>
</tr>
</tbody>
</table>

\(^a\)Letters indicate groups of models with significant differences tested with Tukey’s honestly significant test with alpha = 0.05. Comparisons are done by column.

\(^b\)WL = week of lactation; MY = milk yield; effect of spectra solved using BayesB (BB), Bayesian ridge regression (BRR) or partial least squares on 1,060 (PLS\(_{1060}\)) or 517 (PLS\(_{517}\)) wavenumbers.

Table 8. Correlations between observed and predicted phenotypes in testing set \( \{r_{y,y}\} \) and square root of prediction mean square error (RPMSE) for models using the complete information on milk components and spectra assessed 50 randomizations of a 5-fold cow-independent cross-validation for the Holstein breed

<table>
<thead>
<tr>
<th>Model(^1)</th>
<th>(r_{y,y})</th>
<th>RPMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1: WL + MY + %F + %P</td>
<td>0.81(^{a})</td>
<td>2.16(^{a})</td>
</tr>
<tr>
<td>M2: Parity + WL + MY + %F + %P</td>
<td>0.84(^{a})</td>
<td>2.01(^{a})</td>
</tr>
<tr>
<td>M3: Parity + WL + Milk + Spectra-BB</td>
<td>0.88(^c)</td>
<td>1.77(^{a})</td>
</tr>
<tr>
<td>M4: Parity + WL + Milk + Spectra-BRR</td>
<td>0.88(^c)</td>
<td>1.77(^{a})</td>
</tr>
<tr>
<td>M5: Parity + WL + Milk + Spectra-PLS(_{1060})</td>
<td>0.86(^{a})</td>
<td>1.92(^{d})</td>
</tr>
<tr>
<td>M6: Parity + WL + Milk + Spectra-PLS(_{517})</td>
<td>0.88(^{a})</td>
<td>1.81(^d)</td>
</tr>
</tbody>
</table>

\(^a\)Letters indicate groups of models with significant differences tested with Tukey’s honestly significant test with alpha = 0.05. Comparisons are done by column.

\(^b\)MC = milk components; WL = week of lactation; MY = milk yield; %F = percentage of fat; %P = percentage of protein; Spectra-BB = effect of spectra solved using BayesB (BB); Spectra-BRR = effect of spectra solved using Bayesian ridge regression; Spectra-PLS\(_{1060}\) and Spectra-PLS\(_{517}\) = effect of spectra solved using partial least squares on 1,060 or 517 wavenumbers.
models using spectra information outperformed prediction models based on the predefined formulas (i.e., NRC, 2001; de Souza et al., 2019) in both parameters $r_{y,\hat{y}}$ and RPMSE (Table 9).

### Genetic Analyses

Heritabilities, phenotypic and genetic correlations estimated for EnM, BW, DMIp, and EBp are shown in Table 10. The parameters estimated for EnM and BW were based on the recorded phenotypes in the farms set, while for DMIp and EBp, the genetic parameters were estimated using predicted phenotypes obtained for the farms set applying the prediction equation trained in the station dataset (MC3). Heritabilities were medium-low for EnM, DMIp, and EBp (0.11–0.16), and medium-high for BW (0.42). The genetic associations between the traits, estimated with the bivariate analysis, showed for EnM, a low-negative genetic correlation with BW (−0.17), a medium-high-positive genetic correlation with DMIp (0.40) and a medium-high-negative genetic correlation with EBp (−0.39). Thus, the genetic correlation between DMIp and EBp was high and negative (−0.57). Beside this clear genetic associations, in the population analyzed, BW did not show genetic association with the traits analyzed, except for the already mentioned low-negative association with EnM (Table 9).

### DISCUSSION

#### Prediction Models

Concerning the prediction models, we aimed to assess the contribution of the FT-MIR data from milk samples to the prediction of complex phenotypes scarcely recorded such as DMI and EB, and secondary to BW, which is needed for EB calculation. In this regard, models that incorporated FT-MIR data outperformed the ones without FT-MIR information. In contrast, the structure of the data, mostly belonging from one experimental station, limits the extrapolation of these prediction equations to other production systems.

---

**Table 9.** Correlations between observed and predicted phenotypes in testing set ($r_{y,\hat{y}}$) and square root of prediction mean square error (RPMSE) for prediction models for DMI and EB in 50 randomizations of a 5-fold cross-validation

<table>
<thead>
<tr>
<th>Model</th>
<th>DMI</th>
<th>EB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r_{y,\hat{y}}$</td>
<td>RPMSE</td>
</tr>
<tr>
<td>M7 Parity+WL+MY+Spectra-BB</td>
<td>0.88a</td>
<td>1.76</td>
</tr>
<tr>
<td>M8 Parity+WL+MY+Spectra-BRR</td>
<td>0.88a</td>
<td>1.76</td>
</tr>
<tr>
<td>eq1</td>
<td>0.84a</td>
<td>1.97</td>
</tr>
<tr>
<td>eq2</td>
<td>0.59a</td>
<td>2.89</td>
</tr>
</tbody>
</table>

*Letters indicate groups of models with significant differences tested with Tukey's honestly significant test with alpha = 0.05. Comparisons are done by column.

EB = energy balance; M = model; WL = week of lactation; MY = milk yield; Spectra-BB = effect of spectra solved using BayesB; Spectra-BRR = effect of spectra solved using Bayesian ridge regression; eq1 = DMI predicted using the formula provided by NRC(2001); eq2 = DMI predicted using the formula provided by de Souza et al. (2019); $\hat{D}MI$ = DMI predicted with model or eq.

---

**Table 10.** Posterior mean, (95% highest posterior density region) and Monte Carlo error of heritability in the diagonal, genetic correlation (above diagonal), and phenotypic correlation (below diagonal) between energy in milk (EnM), BW, predicted DMI (DMIp), and predicted energy balance (EBp) in the farms set

<table>
<thead>
<tr>
<th>Trait</th>
<th>EnM</th>
<th>BW</th>
<th>DMIp</th>
<th>EBp</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.05, 0.19)</td>
<td>(−0.45, 0.14)</td>
<td>(0.25, 0.53)</td>
<td>(−0.53, −0.24)</td>
</tr>
<tr>
<td>M1</td>
<td>0.11</td>
<td>−0.17</td>
<td>0.40</td>
<td>−0.39</td>
</tr>
<tr>
<td>M2</td>
<td>0.03</td>
<td>0.42</td>
<td>0.05</td>
<td>0.003</td>
</tr>
<tr>
<td>M3</td>
<td>(−0.3, 0.10)</td>
<td>(0.28, 0.55)</td>
<td>(−0.12, 0.21)</td>
<td>(−0.17, 0.17)</td>
</tr>
<tr>
<td>M4</td>
<td>0.37</td>
<td>0.08</td>
<td>0.16</td>
<td>−0.57</td>
</tr>
<tr>
<td>M5</td>
<td>(0.32, 0.42)</td>
<td>(0.02, 0.15)</td>
<td>(0.13, 0.20)</td>
<td>(−0.09, −0.43)</td>
</tr>
<tr>
<td>M6</td>
<td>(−0.51, −0.42)</td>
<td>(−0.09, 0.04)</td>
<td>(−0.06, 0.05)</td>
<td>(0.10, 0.17)</td>
</tr>
</tbody>
</table>
circumstances (McParland et al., 2014). For the same reason, although CIcv was used to assess the results, it is arguable that our results may be overestimated. Despite these limitations, the results obtained are worthy to confirm results obtained in previous studies on these traits, and to confirm that the use of FT-MIR data may be useful to develop more accurate prediction models even for being used locally.

Beside the methods used in the prediction models with FT-MIR data, BayesB variable selection methods systematically had the best performance together or followed by optimized PLS, namely PLS$^{517}$ with optimum number of component for each prediction fold, as it was already showed in previous studies (Ferragina et al., 2015; Bonfatti et al., 2017; Rovere et al., 2021).

Performance of PLS was sensitive to the noise regions identified to the water absorption wavenumbers, whereas Bayes methods, especially BayesB, was more invariant to such pre-edition of wavenumbers in both $r_{y,y}$ and RPMSE (Supplemental Table S1; https://figshare.com/articles/journal_contribution/Annex_and_supplementary_Tables_Figures/24942360; Buitenhuiss et al., 2024), although some difference could be observed according with the trait analyzed.

Specifically, in PLS, the optimum number of components was lower in PLS$^{517}$ than PLS$^{1060}$ (PLS on 1,000 wavenumbers). Despite the optimal number of components frequently exceeding 20, prediction models with 20 components achieved very similar results in the traits analyzed in this study (Supplemental Table S2; https://figshare.com/articles/journal_contribution/Annex_and_supplementary_Tables_Figures/24942360; Buitenhuiss et al., 2024).

**Prediction Models for DMI**

In our study, the use of spectra information improved the accuracy of the prediction models for DMI. For the best models tested, our results ($r_{y,y} = 0.88$, $R^2 = 0.77$) rank in the upper limit of the precision reported in previous studies (0.29–0.77; Bresolin and Dórea, 2020). Among these results, the highest accuracy of prediction was reported by (Shetty et al., 2017), using PLS models. These authors compared models to predict DMI using MY, BW, and milk components (fat, protein, and lactose) and models in which milk components were replaced by FT-MIR data, using a similar dataset to what we used in our study. They did not find a benefit in using FT-MIR information. In contrast, in our results, the prediction models that included FT-MIR information outperformed models with milk components but without FT-MIR information. We focused on the Holstein population and used more records, which often improves the accuracy of trait prediction (Soyeurt et al., 2011; Tiplady et al., 2020). In addition, in our study, we used a Bayesian approach, that in several studies showed to have an equivalent or better performance in prediction accuracy than PLS (Ferragina et al., 2015; Bonfatti et al., 2017; Toledo-Alvarado et al., 2018; El Jabri et al., 2019; Rovere et al., 2021). Dórea et al. (2018) used artificial neural networks (ANN) for prediction of DMI on Holstein breed and reported a coefficient of determination of 0.70 that would correspond to a correlation of ~0.84. The authors found an improvement in the prediction accuracy of the models based on ANN when, in addition to the milk components, the raw milk spectra data were included, in agreement with our findings. The authors suggested that using raw FT-MIR spectra in the model, in addition to the FT-MIR predicted phenotypes of milk components (i.e., %Fat, %Protein), contributed with information of other compounds not represented in the information carried by %Fat and %Protein. With data from Holstein breed, Grelet et al. (2020) reported an $R^2$ of 0.66 that would correspond to values of correlation of 0.81 using a PLS prediction model approach. More recently using multicountry data and cow-independent cross-validation, Tedde et al. (2021b) reported RMSE of 3.27 ± 0.08 kg for the PLS regression and 3.25 ± 0.13 kg for ANN. On a country-independent cross-validation, the authors reported RMSE varying from 3.73 to 6.03 kg for PLS and 3.69 to 5.08 kg for ANN. In light of these results, they underlined the importance of the FT-MIR spectra in the prediction models.

**Prediction Models for BW**

Body weight at the individual level and during cow lactation is a relevant trait for management and breeding. Body weight is directly related to cow’s feed requirements and consequently, to other complex phenotypes such as EB or feed efficiency; however, BW has not been routinely recorded in commercial herds. Then, predictions using information recorded during the cow’s test day might be a useful tool. To our knowledge, 3 previous studies assessed FT-MIR predictions of BW. Soyeurt et al. (2019) reported a cross-validation precision of 0.61 that would correspond with a correlation of 0.78, higher than our best results of ~0.69. Similar results, in the range of 0.75 to 0.79 ($R^2$ range 0.55–0.63), were reported by Zhang et al. (2021a), and Tedde et al. (2021a) ($R^2$ range 0.54–0.61), in this case using a multicountry dataset.

Despite differences in the structure of population and methodology applied, we found that FT-MIR models increased by 6% to 8% prediction accuracy, in
agreement with the 7% of improvement reported by (Soyeurt et al., 2019), and slightly more than ~4% by Zhang et al., (2021b). In all the cases, parity and week of lactation (or DIM) were more positively associated with BW than MY; nevertheless, prediction models that included all the 3 variables, in addition to FT-MIR, achieved the highest accuracy of prediction. In terms of RPMSE, our results are in the range of values reported for prediction in the aforementioned studies (37–64, Soyeur et al. (2019); ~60, Zhang et al., 2021b; 52–56, Tedde et al., 2021a), and when it is expressed relative to the standard deviation of the samples varied between 1.26 and 2.80. This level of accuracy is considered limited but is still a good indicator of the BW to discriminate groups of cows useful for management and breeding. This limited prediction accuracy was similar in the results obtained from a Hcv, and worse in terms of RPMSE to the results from a 9 Hcv obtained by Tedde et al. (2021a), using data from a multiple-country dataset. We speculated that the differences in the structure of the herds analyzed would have negatively affected the prediction accuracy. In terms of the spectra data, Supplemental Figure S3 (https://figshare.com/articles/journal_contribution/Annex_and_supplementary_Tables_Figures/24942360; Buitenhuis et al., 2024) shows a PCA on the spectra analyzed from station and farms sets. Additionally, we calculated and compared the pairwise Gromov-Hausdorff (GH) distance between the herds as it was suggested by one referee and according with (Zhang et al., 2021a). The GH distance between spectra from different herds were no significantly different, disregarding the idea that the variability of the spectra in the training dataset were significantly different to the ones in the testing set. Regardless, it is important to emphasize that regularly updating and enlarging the databases used for the prediction models will ensure robust predictions; however, for BW, a trait more challenging to predict using FT-MIR data, including additional features available at the farm level, as suggested by Tedde et al. (2021a), may improve the prediction accuracy.

**Prediction Models for EB**

For this study, we defined EB as the difference between the energy acquired by the animal in its feed consumption and the energy spent in the production and maintenance. In this study, energy input or consumed had a very high correlation, and its prediction accuracy was mostly the same as for DMI. This high association can be explained by the lack of detailed information on the feed energy associated with the individual DMI recorded. Instead, we only have access to a general mean of the energy content of the feed during the period covered by our study. Therefore, we opted for reporting only DMI that was the most accurate phenotype recorded to contrast the results to previous studies. Then considering the DMI information, 2 strategies were evaluated: the first one was to assess the prediction accuracy for EB obtained by models with cow features such as parity, week of lactation, MY, and FT-MIR data; the second one sought to predict DMI, which is not routinely measured in commercial herds, and based on predicted DMI, calculate the EB. Interestingly, this second option performed significantly (P < 0.05) better than the first option, although the difference observed was small. However, when %Fat and %Protein were additionally included to MY in the prediction model, the performance between both options did not differ. Both strategies showed auspicious results, and the accuracy of the prediction models were in line with the results obtained from previous studies (McParland et al., 2011, 2014, 2015; Smith et al., 2019). The accuracy of prediction reported on these foundational studies by McParland et al. (2011, 2014, 2015) ranged between 0.67 and 0.75, and similar results were reported by Smith et al. (2019) with accuracy of prediction ~0.77 (R² = 0.60) for the external cross-validation. In line with the aforementioned studies, the assessment of the models in our study was based in the partition of the station set; consequently, training and test shared genetic and environmental similarities that play positively to the accuracy of prediction. This was pointed out by (McParland and Berry, 2016) noting the scarcity of studies that assess the accuracy in external dataset. In fact, the results from (McParland et al., 2012) showed that when equations calibrated in populations of cows that differed substantially from the one in the validation set, the accuracy of prediction decayed drastically. The authors analyzed 2 independent populations, one fed with TMR diet in confinement and one population mostly pasture based feed, which differed in MY, milk components and BW. This emphasizes, one more time, the importance of applying equations calibrated in samples with similar spectral patterns than the target population, on which the prediction wants to be applied (McParland et al., 2012).

**Genetic Analyses**

**Heritability.** The heritabilities obtained in this study were in the lower range of the values reported in previous studies, except for BW. For EnM, most of the previous studies had reported medium to medium-high heritabilities (0.22–0.46), considering estimates for the whole lactation and for different stages of lactation (Li et al., 2018; Lu et al., 2018; Manzanilla-Pech...
High values as 0.58 to 0.68 were reported by Krattenmacher et al. (2019) and Byskov et al. (2017), who considered that high values could be related to the fact that the studies were in a single well-managed research herd. In fact, Byskov et al. (2017) reported estimates ranged between 0.08 and 0.15 when the data were from commercial herds, values similar to our results (0.11). Also, similar values were reported by Spurlock et al. (2012; 0.10–0.17) and Manzanilla-Pech et al. (2016) using data belonging to 7 experimental stations in the United States (0.18). In the case of BW, the estimates obtained in this study were well in the range of results reported for Holstein breed whether they referred to primiparous or multiparous, for whole lactation or by stage of lactation (i.e., 0.25–0.48; Manzanilla-Pech et al., 2014; Lu et al., 2018; Veerkamp and Thompson, 1999). Most of the previous studies for DMI in Holstein breed obtained medium to medium-high heritabilities (0.21–0.55; Li et al., 2018; Manzanilla-Pech et al., 2014; Lu et al., 2018; Krattenmacher et al., 2019; Veerkamp and Thompson, 1999). Different from these studies, our estimates were on predicted DMI instead of observed DMI, and therefore, a reduction of the variances is expected; however, our results (0.16) were aligned with some values reported by Veerkamp and Thompson (1999; the third to fourth week of lactation, 0.23–0.13), Spurlock et al. (2019; fifth month of lactation, 0.22), and with Smith et al. (2019) and (McParland et al., 2015), which reported estimates for predicted energy intake instead (0.18 and 0.20) using prediction models with individual FT-MIR information, similar to this study. Most of the previous studies had reported low to medium heritabilities for EB (0.06–0.29; Veerkamp et al., 2000; Berry et al., 2007; Buttchereit et al., 2012); higher values (>0.4) were reported by Krattenmacher et al. (2019) in primiparous cows at 30 DIM, using random regression models. Our results (0.13) were in the average of the values reported and they are auspicious considering that are on predicted phenotypes, which are expected to have lower variability than observed phenotypes. Thus, these results also support the idea that there is some genetic variability in cows’ energy partitioning, and consequently, the possibility of selection toward the most appropriate energy profile in a more holistic genetic merit goal where traits related to reproduction and fitness are considered, as it has been established previously by Banos and Coffey (2010) and Smith et al. (2019).

Correlations. The results of this study show that cows with higher milk production and component (i.e., EnM) did not necessarily have higher weight or size, but they tended to have a higher DMIp and a more negative EBp. This suggests that the increase in DMIp was not sufficient to compensate for the energy output required for milk production. These findings are consistent with previous studies of Manzanilla-Pech (2014, 2016) and Veerkamp et al. (2000) that also reported similar estimates of low-negative to low positive correlations.

In contrast, the medium-high positive correlations between EnM and DMIp were in line with previous studies, and in concordance with which is biologically expected (Veerkamp et al., 2000; Manzanilla-Pech et al., 2014, 2016; McParland et al., 2015). The phenotypic correlations between EBp and BW or DMIp were found to be close to zero, which is in line with previous studies (McParland et al., 2015). Genetic correlations largely followed the same direction as phenotypic correlations, and EnM was found to be positively correlated with DMIp and negatively associated with EBp, consistent with previous studies (Li et al., 2018, Lu et al., 2018; Krattenmacher et al., 2019; Veerkamp et al., 2000).

However, the genetic correlation between BW and EnM was found to be low-negative, and zero with DMIp and EBp, respectively, which differs from previous studies that reported positive genetic correlations. The difference in results might be explained by the fact that the DMI and EB were based on predicted phenotypes, not observed, and that they were estimated for the whole lactation period. Further research on the genetic association along the lactation might give insights in how to optimize the antagonist relationship between EnM and EB to improve the overall efficiency of the cow.

CONCLUSIONS

Prediction models with FT-MIR increased the prediction accuracy of DMI, BW, and EB. In EB and BW, prediction equations allowed discriminating groups of cows useful still useful for informed management and breeding decisions. Genetic analyses were feasible, and genetic variability was captured in FT-MIR-predicted traits. These results support previous evidence that DMI, BW, and EB predictions using FT-MIR information from milk samples provide useful massive information applicable to herd management and breeding programs, mainly that ones with broad breeding goals that include reproduction, and fitness additionally to production traits.

ACKNOWLEDGMENTS

AJB and the generation of the data in this paper were financially supported by the projects FT-IR spektre i mælk: Genetisk variation og effekt på sundhed, frugtbarhed og energibalance of the Milk Levy Fund,

REFERENCES


ORCIDs

Gabriel Rovere https://orcid.org/0000-0001-8178-3163

Gustavo de los Campos https://orcid.org/0000-0001-5692-7129

Grum Gebreyesus https://orcid.org/0000-0003-4757-3060

Albert J. Buitenhuis https://orcid.org/0000-0002-4953-3081