J. Dairy Sci. 108:12391–12406
https://doi.org/10.3168/jds.2025-26794


# Modeling feed herbage proportion and modeling of the likelihood of feeding strategies focused on grazing and herbage consumption using milk Fourier-transform mid-infrared spectral analysis

**Killian Dichou,[1]* Charles Nickmilder,[1] Gérard Conter,[2] Romain Reding,[3] Antonino Marvuglia,[4] and Hélène Soyeurt[1]**
[1]TERRA Research and Teaching Centre, Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium
[2]Lycée Technique Agricole, 1 Kréiwénkel, L-9374, Gilsdorf, Luxembourg
[3]CONVIS, 9085 Ettelbruck, Luxembourg
[4]Luxembourg Institute of Science and Technology (LIST), 4362 Esch-sur-Alzette, Luxembourg

## ABSTRACT

Effectively evaluating and promoting pro-grazing practices necessitates the implementation of a verification system. To address this imperative, exploration of milk composition analysis as a means to assess grazing practices has garnered substantial attention. In this study, we used component predictions from milk Fourier-transform mid-infrared (FT-MIR) spectra to construct an indicator to estimate the proportion of herbage consumed by dairy cows and another indicator to validate grazing. This approach was developed and validated using 75 estimated bulk milk analyses, each associated with 3 variables related to feeding from the same day ± 3 d, totaling 526 observations. These 3 variables are based on the occupation time, harvested and conserved herbage, and other feeds from 7 farms in Luxembourg. Hierarchical clustering facilitated the effective segregation of observations into distinct groups, with one group predominantly focused on herbage and the other group on other feeds. Leveraging partial least squares discriminant analysis trained on FT-MIR predicted milk characteristics from both groups, we successfully developed an indicator—the probability of belonging to the herbage group—with a maximum accuracy of 0.93, a sensitivity of 0.94, and a specificity of 0.93 on the Luxembourg dataset. In a partial least squares regression, the cross-validation yielded results of predicting the percentage of herbage in the diet with an error of 8.77%. Notably, the indicator relied on FT-MIR predicted components expected to reflect a diet based on herbage, such as the total of C18:1 *trans* fatty acids and CLA. However, it also incorporated unexpected FT-MIR predicted parameters like milk acidity parameters, citrate content, and specific proteins such as lactoferrin. Finally, the developed indicators were tested on the 5,886,364 Walloon spectra collected between 2009 and 2023, as well as 23,718 Walloon spectra between 2023 and 2025 from 72 farms known to practice grazing. The annual trends were analyzed in the context of Walloon dairy production, helping to refine the selection of better indicators. These results could contribute to practical tools for monitoring and estimating days spent on pasture. Looking ahead, future research should aim to incorporate more comprehensive data, such as precise feed compositions, to further refine our understanding of the influence of specific herbage diets on milk composition and enhance the detection of associated changes.

**Key words:** milk FT-MIR spectra, grass diet, pasture, grazing, dairy cows

## INTRODUCTION

Grazing and herbage-based diets can be important levers in the management and feeding strategies of dairy farms. One element often emphasized is their impact on the organoleptic properties and composition of milk (Alothman et al., 2019; Birkinshaw et al., 2023). Notable advantages for human health include the enhancement of n-3, CLA, and PUFA in milk fat (Dewhurst et al., 2006; O'Brien and Hennessy, 2017). Another frequently cited aspect of grazing pertains to animal welfare, as it allows cows to express their natural behaviors (Haskell et al., 2006; Legrand et al., 2009; Olmos et al., 2009) and contributes to their overall health and mobility (White et al., 2001; De La Torre-Santos et al., 2020). Although grazing also offers benefits to farmers by reducing feeding costs (Kučević et al., 2016) and enhancing the sustainability of their farms, these practices may demand considerable time, planning, and expertise to optimize the feeding practices for each farm's specific context. Current poli-

Received April 18, 2025.
Accepted August 11, 2025.
*Corresponding author: killian.dichou@uliege.be

The list of standard abbreviations for JDS is available at adsa.org/jds-abbreviations-25. Nonstandard abbreviations are available in the Notes.

cies, such as the Common Agricultural Policy for the period 2023 to 2027 within the European Union (Dewhurst and Moloney, 2013), further endorse and encourage these practices. This endorsement is not only due to the aforementioned reasons but also because pastures confer environmental benefits, such as carbon sequestration (Dillon et al., 2005) and biodiversity protection (European Commission, 2024).

To effectively evaluate and promote grazing practices, the implementation of a verification system is critical, as it would ensure traceability and compliance with policy standards. Various possibilities have been explored to address this challenge. One option is sensor-based devices alone, such as the Aficollar, developed by Afimilk Ltd. (Kibbutz Afikim, Israel), which reported 94% accuracy for both recording grazing and rumination time (Iqbal et al., 2021). Other devices, like the Chronopâture tool, developed by Adventiel (France), include global positioning system (**GPS**) technology and are used in precision livestock farming to record the animals' position every 30 min (Axema Promotion and Services, 2023), offering a means to assess the number of days on pasture. An advanced approach that involves coupling sensor-based devices and having their data processed via methods such as artificial neural networks or random forest algorithms (Coelho Ribeiro et al., 2021) enhances the identification of grazing and nongrazing behaviors, thereby improving the reliability of such tools. Nevertheless, the widespread adoption of these tools may be hindered by substantial initial investments.

Another approach, which can be less expensive, involves using satellite data, enabling the verification process to be scaled up. For instance, high-resolution images from the Sentinel-2 (10 m), RapidEye (5 m), and Planetscope (3 m) satellite constellations have been used (Woodward et al., 2019). When combining satellite images with a few GPS collars to keep track of the whole herd's position, they estimated herbage mass available with a root mean squared error (**RMSE**) of 225 kg of DM per hectare and estimated cows' grass intake by using the pasture disappearance as a proxy. However, the reliance on optical data introduces challenges, including image resolution, orbital revisit time, and data acquisition frequency.

A third alternative explores milk composition analysis as a means to assess the proportion of herbage in the animals' diet. An indicator called GRASS1 has been proposed (Soyeurt et al., 2022) based on 48 components predicted from the Fourier-transform infrared (**FT-MIR**) spectra from milk analysis, trained on bulk milk. The most important components were found to be 6 groups of FT-MIR–predicted fatty acids (**FA**): C6:0, C10:0, C16:0, C18:2 *cis*-9, C18:2 *cis*-12, and the medium-chain FA. The GRASS1 indicator discriminates among several management practice groups. It was indirectly validated by comparison with expected seasonal patterns using a database built from 533,786 records and resulted in a 10-fold stratified cross-validation accuracy of 95% and a validation accuracy of 90%. Additionally, herbage and other feeds were differentiated using milk FT-MIR spectroscopy and partial least square discriminant analysis (**PLSDA**) with sensitivity and specificity ranging from 80% to 100%, although with challenges in identifying the specific proportion of the diet derived from grazing (Capuano et al., 2014). From a nutritional standpoint, the form in which herbage is provided to cattle affects milk composition (Alothman et al., 2019). For instance, protein content and quality in milk change depending on feed derived from grazed herbage, silage, or hay (O'Callaghan et al., 2018). Addressing this matter, different proportions of herbage types in the diet have been predicted (including hay, conserved herbage and fermented forage), reaching 85% precision for diets based on less than 50% grazing and demonstrating the potential of milk composition analysis in characterizing diet traits (Coppa et al., 2021).

Gathering data over the occupation time or grazing period, so the total time passed on pasture or the actual time spent grazing (The Forage and Grazing Terminology Committee, 2009), is challenging when addressing modeling questions regarding grazing-related traits. Instead of hypothesizing about the seasonality to train the models (Soyeurt et al., 2022), this study consists of using grazing-based traits recorded in grazing calendars completed by 7 farmers in Luxembourg. This study develops 2 main types of models: some that predict the proportion of herbage in the diet, and others that model the likelihood of feeding strategies centered on grazing and herbage consumption. Given the nearly daily availability of FT-MIR bulk milk analysis, this study suggests the potential for developing a monitoring tool to reward management practices that promote grazing and optimize farm resources.

## MATERIALS AND METHODS

To clarify this study's stages, the general workflow is outlined in Figure 1.

### *Farm Characteristics*

Gathered through the Simulating Economic and Environmental Impacts of Dairy Cattle Management Using Agent-Based Models (**SIMBA**) project, data on farm characteristics, predicted milk composition, and grazing calendars were collected in Luxembourg on 4 farms in 2019 and 2020, and 3 other farms in 2020. The sampling period concerned an extended grazing season, ranging

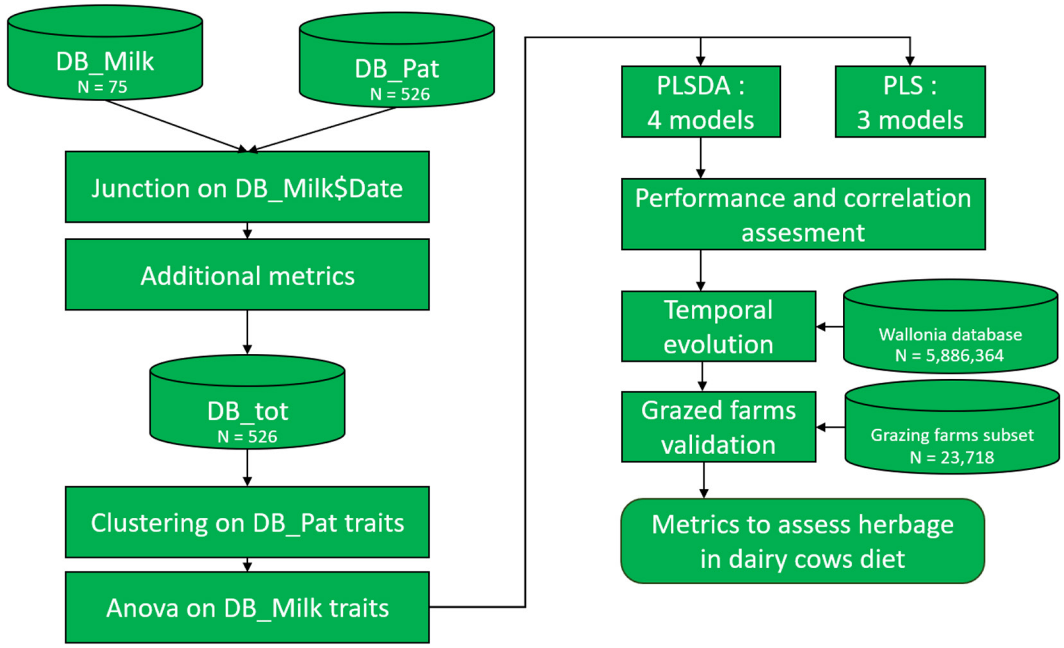


**Figure 1.** Workflow representation. DB = database; PLSDA = partial least squares discriminant analysis; PLS = partial least square analysis.

from March to November. All the farms bred Holstein cows. The database included both conventionally managed (farms 3, 4, 5, 6, and 7) and organically managed farms (farms 1 and 2).

### Predicted Milk Composition

Because it was not possible to retrieve the results coming from the FT-MIR analysis of bulk milk samples, the results obtained from the routine milk recording done on individual cows were used in this study to compute an approximate herd milk composition. In Luxembourg, the farms of the SIMBA project participated in the milk recording routine organized by CONVIS s.c., an agricultural cooperative (Ettelbruck, Luxembourg) every 4 to 6 wk. A sample of milk is collected on each producing cow following procedure 1 to compute 24-h yields approved by International Committee for Animal Recording (**ICAR**) guidelines (ICAR, 2023). Proportional milk samples from both morning and evening milkings are collected, preserved, and combined into a single composite to accurately reflect the full-day yield in accordance with ICAR standards. The sample is then analyzed using a MilkoScan spectrometer (Foss Electric A/S, Hillerød, Denmark). From this analysis, the milk fat (**%FAT**), protein (**%PROT**), lactose, and urea contents were provided. The milk FT-MIR spectrum generated during this analysis was also recorded in the database. All spectral data were standardized (Grelet et al., 2017). Additional cow-level information such as milk yield, parity, and the stage of lactation were also available. From the recorded spectral data, 54 traits were predicted using equations developed in past projects. Table 1 presents the equation characteristics. To represent the analysis of the estimated herd milk samples, a weighted average based on the milk yield was computed for all common dairy traits and FT-MIR predictions. Finally, all information was included in a database called DB_milk. Due to issues in data collection, computation, or validation of those sample data, only 75 observations were available for this study.

### Grazing Calendar

The second database (DB_Pat) contained daily observations made at the herd level by the farmer. The time spent on pasture (**Occupation_Time**) is expressed in hours per day. The quantities of harvested or conserved herbage (**compl_herbage**) and other feeds (**compl_other**) are expressed in kilograms of DM and are averaged per cow and per day. Unfortunately, compl_herbage was not detailed regarding the specific proportions of fresh herbage, herbage silage, and hay. In the same manner, for compl_other, the specific composition or characteristics of the feeds were not recorded by the farmers. Based on typical feeding practices in Luxembourg, these are assumed to include energy- and protein-rich concentrates like cereal grains such as corn or barley; soybean meal; rapeseed meal; agroindustrial byproducts such as sugar beet pulp or brewers grains; and specific nutritional supplements such as mineral and vitamin premixes.

**Table 1.** Descriptive statistics of traits used in the milk dataset, as well as the specifications of FT-MIR equations

| | | Equation specification[1] | | | | DB_milk (n = 75) | |
|---|---|---|---|---|---|---|---|
| Trait[2] | Unit | n | $R^2$ | RMSE | Reference | Mean | SD |
| Milk yield | kg/d | — | — | — | — | 25.42 | 4.008 |
| Fat | g/L milk | — | — | — | — | 4.076 | 0.321 |
| Protein | g/L milk | — | — | — | — | 3.410 | 0.156 |
| Urea | mg/L milk | — | — | — | — | 193.615 | 71.231 |
| Lactose | g/L milk | — | — | — | — | 4.755 | 0.167 |
| C4:0 | g/dL milk | 1,371 | 0.93 | 0.008 | A | 0.1 | 0.012 |
| C6:0 | g/dL milk | 1,371 | 0.91 | 0.006 | A | 0.069 | 0.006 |
| C8:0 | g/dL milk | 1,371 | 0.91 | 0.004 | A | 0.049 | 0.004 |
| C10:0 | g/dL milk | 1,371 | 0.92 | 0.010 | A | 0.123 | 0.011 |
| C12:0 | g/dL milk | 1,371 | 0.93 | 0.011 | A | 0.157 | 0.014 |
| C14:0 | g/dL milk | 1,371 | 0.94 | 0.030 | A | 0.493 | 0.037 |
| C14:1 *cis* | g/dL milk | 1,371 | 0.71 | 0.008 | A | 0.048 | 0.005 |
| C16:0 | g/dL milk | 1,371 | 0.95 | 0.091 | A | 1.205 | 0.129 |
| C16:1 *cis* | g/dL milk | 1,371 | 0.73 | 0.013 | A | 0.066 | 0.007 |
| C17:0 | g/dL milk | 1,371 | 0.81 | 0.003 | A | 0.026 | 0.002 |
| C18:0 | g/dL milk | 1,371 | 0.84 | 0.056 | A | 0.375 | 0.057 |
| Total of C18:1 | g/dL milk | 1,371 | 0.96 | 0.060 | A | 0.96 | 0.105 |
| Total of C18:1 *trans* | g/dL milk | 1,371 | 0.80 | 0.025 | A | 0.139 | 0.018 |
| Total of C18:1 *cis* | g/dL milk | 1,371 | 0.95 | 0.063 | A | 0.869 | 0.096 |
| C18:1 *cis*-9 | g/dL milk | 1,371 | 0.95 | 0.061 | A | 0.807 | 0.09 |
| Total of C18:2 | g/dL milk | 1,371 | 0.71 | 0.014 | A | 0.099 | 0.007 |
| C18:2 *cis*-9,*cis*-12 | g/dL milk | 1,371 | 0.75 | 0.011 | A | 0.056 | 0.006 |
| C18:2 *cis*-9,*trans*-11 | g/dL milk | 1,371 | 0.74 | 0.010 | A | 0.023 | 0.003 |
| C18:3 *cis*-9,*cis*-12,*cis*-15 | g/dL milk | 1,371 | 0.69 | 0.004 | A | 0.039 | 0.01 |
| SFA | g/dL milk | 1,371 | 0.99 | 0.072 | A | 2.697 | 0.233 |
| MUFA | g/dL milk | 1,371 | 0.97 | 0.059 | A | 1.12 | 0.114 |
| PUFA | g/dL milk | 1,371 | 0.79 | 0.021 | A | 0.169 | 0.014 |
| UFA | g/dL milk | 1,371 | 0.97 | 0.064 | A | 1.289 | 0.127 |
| SCFA | g/dL milk | 1,371 | 0.93 | 0.025 | A | 0.359 | 0.029 |
| MCFA | g/dL milk | 1,371 | 0.97 | 0.104 | A | 2.125 | 0.189 |
| LCFA | g/dL milk | 1,371 | 0.95 | 0.110 | A | 1.568 | 0.163 |
| Branched FA | g/dL milk | 1,371 | 0.77 | 0.013 | A | 0.109 | 0.012 |
| Total of n-3 | g/dL milk | 1,371 | 0.68 | 0.006 | A | 0.027 | 0.003 |
| Total of n-6 | g/dL milk | 1,371 | 0.74 | 0.014 | A | 0.096 | 0.008 |
| Total of odd FA | g/dL milk | 1,371 | 0.84 | 0.016 | A | 0.173 | 0.016 |
| Total of *trans* FA | g/dL milk | 1,371 | 0.82 | 0.029 | A | 0.181 | 0.022 |
| Lactoferrin | mg/L milk | 2,189 | 0.70 | 127.98 | NP | 114.892 | 66.218 |
| Casein $\alpha_{S1}$ | g/L milk | 135 | 0.81 | 0.58 | NP | 11.332 | 0.393 |
| Casein $\alpha_{S2}$ + K | g/L milk | 135 | 0.81 | 0.36 | NP | 8.191 | 0.612 |
| Casein β | g/L milk | 133 | 0.75 | 1.13 | NP | 11.245 | 0.468 |
| Total casein | g/L milk | 133 | 0.84 | 1.56 | E | 31.785 | 1.182 |
| Protein efficiency | % | 1,093 | 0.55 | 3.05 | B | 15.338 | 1.599 |
| Protein nitrogen content | g/L milk | 4,305 | 1 | 0.19 | NP | 27.111 | 0.916 |
| Sodium | mg/kg of milk | 1,019 | 0.44 | 50.98 | C | 349.9 | 24.769 |
| Calcium | mg/kg of milk | 1,094 | 0.82 | 53.38 | C | 1,191.268 | 53.001 |
| Phosphorus | mg/kg of milk | 1,083 | 0.75 | 58.71 | C | 986.346 | 37.977 |
| Potassium | mg/kg of milk | 1,090 | 0.55 | 88.14 | C | 1,496.364 | 38.501 |
| Magnesium | mg/kg of milk | 1,124 | 0.72 | 6.53 | C | 103.251 | 3.527 |
| Milk pH | — | 1,152 | 0.65 | 0.07 | NP | 6.616 | 0.039 |
| Titratable acidity | Dornic degree | 930 | 0.70 | 1.00 | NP | 14.582 | 1.324 |
| Yogurt activity | Dornic degree | 98 | 0.42 | 0.04 | NP | 0.617 | 0.013 |
| Yogurt texture | N | 51 | 0.12 | 0.02 | NP | 0.228 | 0.003 |
| Isocitrate | mmol/L of milk | 2,129 | 0.58 | 0.03 | F | 0.158 | 0.013 |
| Citrate | mmol/L of milk | 498 | 0.90 | 0.68 | D | 8.55 | 0.929 |
| Blood fructosamine | µmol/L of blood | 368 | 0.32 | 14.37 | G | 266.057 | 4.092 |
| Blood glucose | µmol/L of blood | 369 | 0.50 | 0.33 | H | 3.708 | 0.104 |
| Blood IGF-1 | ng/mL blood | 371 | 0.61 | 36.80 | H | 137.266 | 15.267 |
| Square root A30 | mm | 185 | 0.59 | 2.74 | NP | 36.405 | 1.21 |

[1]n = number of records used to build the model; RMSE = root mean squared error; NP = not published; A: Grelet et al. (2014); B: Grelet et al. (2020); C: Christophe et al. (2021); D: Grelet et al. (2016); E: Grelet et al. (2021); F: Grelet et al. (2024); G: Grelet et al. (2025); H: Grelet et al. (2019).

[2]SCFA = short-chain FA; MCFA = medium-chain FA; LCFA = long-chain FA.

Depending on farm type and feeding strategy, additional ingredients such as molasses, bakery waste, or feed fats may also be included. Because the period of occupation does not necessarily imply grazing, this study interprets grazing as a binary variable, indicating whether grazing is practiced. This consideration is crucial for assessing the potential for animals to consume fresh herbage.

To merge the 2 databases, the date of milk analysis (DB_milk) was extended to ±3 d around the date of observation in DB_Pat, resulting in 7 observations from DB_Pat being linked to a single milk analysis. This chosen extended window was based on relative time for milk composition to be affected by feeding and animal digestion, as all components should be digested after more than 48 h (Mambrini et al., 1994; Wattiaux, 1995). After merging, 2 additional metrics were computed to shed light on the feeding patterns of dairy herds and to analyze their relationship with milk composition. The first indicator, **%Filling**, helps ascertain the proportion of supplementation relative to the average ingestion capacity. In this study, the ingestion capacity, expressed in DMI, was estimated using the formula proposed by Cuvelier et al. (2021):

$$\text{DMI} = 1.4 \times (\text{BW}/100 + 2) + 0.30 \times \text{FPCM},$$

where BW is the average cow BW, set at 650 kg based on expert input from the Lycée Technique Agricole (Ettelbruck, Luxembourg), and FPCM is the quantity of fat- and protein-corrected milk, expressed in kilograms per day, estimated using the formula mentioned by De Brabander et al. (2011):

$$\text{FPCM} = 0.337 + (0.116 \times \%\text{FAT}) + (0.06 \times \%\text{PROT})] \times \text{milk yield}.$$

Although more complex formulas exist to estimate DMI (Coulon and d'Hour, 1994), they were deemed too difficult to apply to this study, which focuses on analyzing bulk milk and herd-level data. Consequently, parameters such as individual lactation periods could not be incorporated into the formulas. The proportion of supplementation (%Filling) was then approximated as follows:

$$\%\text{Filling} = 1.1 \times \text{compl\_herbage} + 1 \times \text{compl\_other}/\text{DMI} \times 100.$$

This formula considers the rumen-filling associated with 2 types of feed (i.e., herbage and other). Although the coefficients vary depending on feed type, an average value from Baumont et al. (2007) was chosen since the information was only categorized into herbage and other feeds in our database. The second parameter employed for an easier interpretation is the proportion of herbage fed to dairy cows (**%Herbage**), calculated as follows, with the proportion of compl_other fed to dairy cows regarding their estimated DMI (%other_feeds):

$$\%\text{Herbage} = 100 - \%\text{other\_feeds}.$$

This approximation is considered valid in cases where animals consistently reach their maximum ingestion capacity. Several factors support this hypothesis, such as the relatively stable milk production per cow. A general underfed state would negatively affect milk yield or composition (Leduc et al., 2021).

### Clustering

Because different feeding practices were used on the farms composing the dataset, clustering was performed to group the observations. To segregate them into distinct groups, we employed unsupervised learning, based on hierarchical clustering from Ward's method (Ward, 1963) on the Occupation_Time, %Filling, and %Herbage. This agglomerative algorithm works step by step to combine the closest records until the whole dataset is considered. The data were clustered using Ward's method after standardization by subtraction of the mean and division by the SD. The visualization resulted in a dendrogram with a final number of groups defined by the optimal inertia (i.e., the ratio between the between-class and within-class variance). The interpretation of these clusters was based on the descriptive statistics measured for the features used in the clustering, resulting in feeding strategy typology for each cluster.

### Feature Selection and PLSDA

The Occupation_Time, %Filling, and %Herbage used to make the cluster are not readily available on a routine basis. Therefore, because the consumption of herbage affects the milk composition (Elgersma et al., 2004; De La Torre-Santos et al., 2020; Musati et al., 2025), there is interest in predicting the cluster representing different feeding strategies. This can be achieved by using milk traits that are widely available and low-cost as it is the case for FT-MIR predictions. In this study, 56 FT-MIR predictions were available (Table 1). However, to provide a more robust model, it is necessary to limit the number of features (Grelet et al., 2021). Furthermore, this would also highlight the specific traits that could be related to feeding practices in milk analysis.

To select the traits that are mostly influenced by the found clusters, a first approach consisted of performing a single-trait univariate regression for each FT-MIR trait, as well as milk yield, protein, fat, urea, and lactose contained in DB_milk (i.e., 75 records). This ANOVA only included the cluster label as a fixed effect. The most relevant traits to distinguish clusters were selected based on their *P*-value on the global cluster effect. In other words, if the *P*-value was lower than 0.05, the trait was included in the subsequent model allowing the prediction of the found cluster.

The second step consisted of developing a PLSDA to predict the cluster promoting more grazing in their feeding practices, thus presenting a higher value in %Herbage and lower values in "other feeds" and %Filling. To perform this model, the records belonging to this cluster were set to 1 and the ones belonging to another cluster were set to 0. Upsampling was also performed to avoid weight problems for the clusters with the fewest observations. The number of partial least squares (**PLS**) analysis factors was estimated using the output of a 10-fold cross-validation. The number of latent variables was chosen to maximize the coefficient of determination ($R^2$). The output also provides an estimate of model accuracy, assessed from the estimation of the area under the curve. We have also estimated the Cohen's kappa coefficient, the sensitivity, and the specificity. Because the modeling was performed on 10 folds, the mean and the SD were estimated for all previously listed statistical parameters. The output of the PLSDA model not only provides a label, which is a discontinuous variable, but also provides the probability to belong to the cluster, which is often more interesting to use in practice because it is a continuous variable. To complete the approach, a PLS was conducted to directly predict the %Herbage, the %Filling, and the Occupation_Time as conducted with the same milk traits as the previous PLSDA. The number of latent variables was chosen to minimize the mean RMSE, which was calculated based on 10-fold cross-validation. The results of the PLSDA were compared with the PLS predicting the %Herbage using a linear regression that was only calibrated on the dataset, as no external validation set could be made.

All computations in this study were conducted using R v4.2.2, although the packages were compiled with R v4.2.3 (CRAN, 2023) in the RStudio development environment (POSIT, 2024). More specifically, both PLS and PLSDA were computed with the R packages caret 6.0-93 (Kuhn et al., 2023) and pls 2.8-2 (Liland et al., 2023).

### Comparison with Another Grass Model

The probability of belonging to a cluster presenting feeding practices that promote grazing can be computed for each model. As a way to interpret our models, the GRASS1 model (Soyeurt et al., 2022) was applied on DB_milk to allow a comparison with the probability developed in this study by calculating their correlation. Because GRASS1 was built without direct validation data but based on the hypothesis of seasonality, the correlation with our model will be informative for understanding how seasonality might be indirectly accounted for in modeling based on feeding practice traits.

### Application to the Walloon Database

To observe the resulting indicators on a wider dataset, we applied the obtained equations to the recorded milk database related to the FuturoSpectre agreement linking the University of Liège—Gembloux Agro-Bio Tech (Gembloux, Belgium), the Walloon Research Centre (Gembloux, Belgium), the milk laboratory Comité du Lait (Battice, Belgium), and the Walloon Breeding Association (Ciney, Belgium). This dataset was composed of 5,886,364 spectra between 2009 and 2023 from 7,268 dairy farms as part of the milk sampling conducted by dairies to determine the milk payment. Thus, the milk FT-MIR spectra were generated by the analysis of bulk milk using MilkoScan spectrometers (Foss Electric A/S, Hillerød, Denmark) at the milk laboratory Comité du Lait (Battice, Belgium) and standardized using the method developed by Grelet et al. (2017).

From this dataset, a more specific subset was verified to contain records from 72 farms that implemented grazing practices. Although the specific grazing period was not detailed, based on our comprehensive knowledge of local practices, we can infer that grazing typically occurs from April to October on average. The subset encompasses milk analysis data spanning from January 1, 2023, to March 6, 2025, yielding a total of 23,718 observations. Due to the uncertainty on the grazing periods, we only applied the most promising indicators on this dataset as an example.

## RESULTS AND DISCUSSION

### Overview of DB_Pat

Table 2 provides the descriptive statistics of the final DB_Pat dataset. The DMI falls within the expected range. The SD across the other variables reflects the variation in the dataset. The relative SE is below 5% for all variables, indicating that the mean value of each variable is a reliable estimation of the mean of the populations.

### Farm Typology

Figure 2A and 2B present, by farm, the average characteristics found in DB_Pat. Figure 2A focuses on the %Herbage, %Filling, and Occupation_Time, which are used in the clustering. Figure 2B offers a more detailed

**Table 2.** Descriptive statistics of the dataset related to grazing calendars

| Parameter | Unit | N | Mean | SD | RSE[1] |
|---|---|---|---|---|---|
| Harvested and conserved herbage | kg of DM/cow per day | 526 | 7.01 | 4.81 | 2.99 |
| Other feeds | kg of DM/cow per day | 526 | 7.72 | 3.72 | 2.10 |
| Occupation time | h/d | 526 | 9.11 | 8.39 | 4.02 |
| DMI | kg of DM/cow per day | 526 | 19.54 | 1.20 | 0.27 |
| %Filling | % | 526 | 79.23 | 32.37 | 1.78 |
| %Herbage | % | 526 | 61.12 | 17.69 | 1.26 |

[1]Relative standard error = (SE/mean) × 100.

point of view, presenting specifically the harvested or conserved herbage and the other feeds. The distinction allows an easier interpretation of the farm feeding strategies. Compared with the other farms, farms 1 and 2 exhibit the highest scores for %Herbage and the occupation period. The ratio between other feeds and harvested or conserved herbage is also clearly in favor of the latter. These observations were expected, as farm 1 and 2 are organic. Farm 7 presents the same conclusion concerning the feeding ratio, but like farms 3, 4 and 6, it also presents a high %Filling and a lower %Herbage. All farms exhibit an average occupation period that ranges between 4 to 9 h/d, with the exception of farms 1 and 2, which seem to keep the herd on pasture most of the time. On average, the highest grazed grass observations should therefore be found on farms 1 and 2, but also on farm 5, where %Herbage reaches 67% and the %Filling reaches 68%, indicating that the herd can still graze and that a good proportion of the feeds are related to herbage (grazed, harvested, or conserved). In contrast, farm 7 should include observations that focus less on grazing, with the highest %Filling and the lowest time of occupation. Because the observations vary with the weather and the time of the year and cannot definitively define a farm, this typology only draws attention to general patterns that should be observed after the clustering.

## Clustering and Interpretation

The dendrogram obtained from the hierarchical clustering performed on the Occupation_Time, %Filling, and %Herbage is shown in Figure 3. Based on the observed inertia, the clustering was divided into 4 distinct groups. Their characteristics, outlined in Table 3, highlight various feeding strategies. Indeed, cluster 2 includes 123 observations, all focused on grazing and maximizing herbage in the diet, as it presents higher Occupation_Time and %Herbage. Conversely, cluster 4 represents a contrasting feeding practice with 65 observations, featuring considerably less herbage in the ration. Additionally, all observations with an occupation time of zero were clustered within cluster 4. Cluster 3 only isolates 31 observations, notable for their herds with high levels of %Herbage and high %Filling. This suggests that access to pasture may not necessarily mean grazing, as the cows are already satiated with other forms of herbage. Finally, cluster 1 comprises 307 observations that do not align closely with the characteristics of the aforementioned clusters, displaying

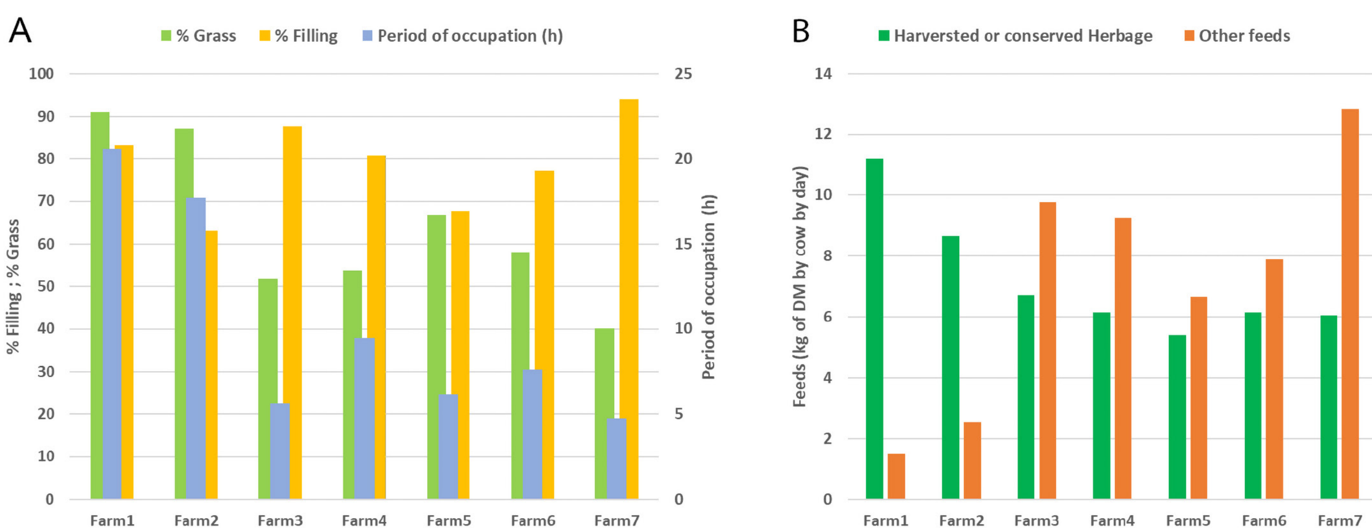


**Figure 2.** Grazing calendars information by farm.

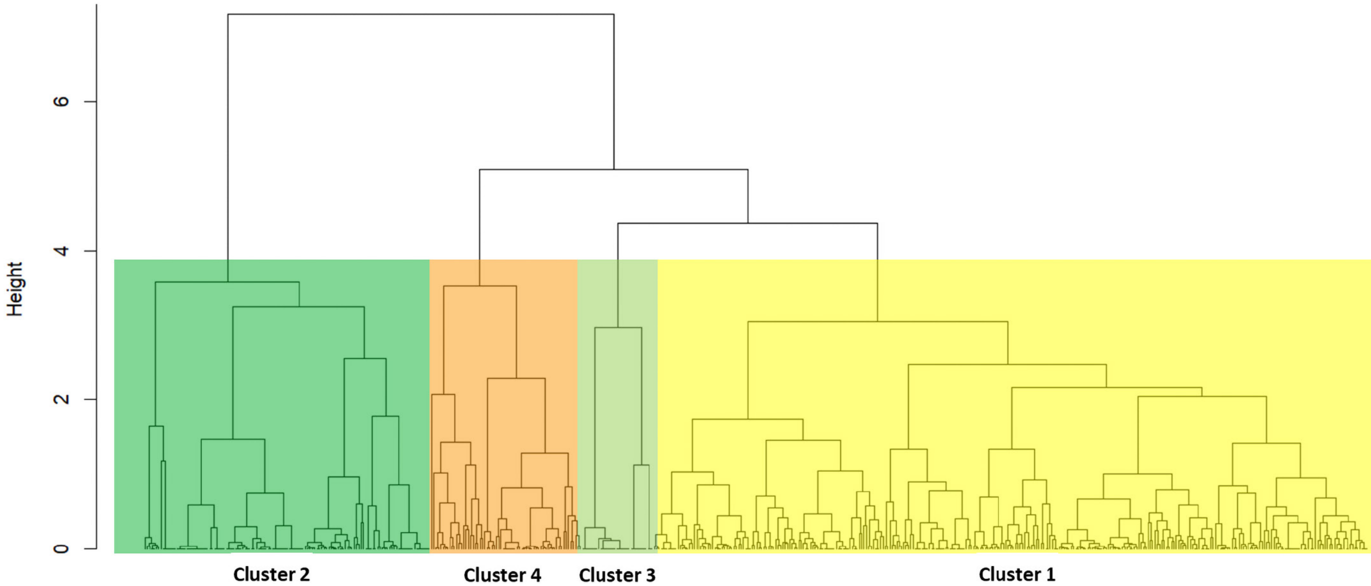


**Figure 3.** Dendrogram of the hierarchical clustering based on the Occupation_Time, %Filling, and %Herbage in the diet (n = 526).

traits that are somewhat intermediate between clusters 2 and 4. Considering the farming general context in Luxembourg, it is expected that most observations during the year align with moderate characteristics (cluster 1), optimizing when possible the efficiency of using grazed herbage, harvested or conserved herbage, and other feeds. In the seasonal context, observations in late spring or early summer are also more likely to be classified as cluster 2.

Another factor supporting the assertion of distinct management types is the evolution of cluster labels for each farm between 2019 and 2020, as depicted in Figure 4. It is evident that farms predominantly maintain their affiliation with the same cluster over time. Changes between cluster 1 and cluster 2 or cluster 1 and cluster 4 were expected, as cluster 1 exhibits intermediate traits. Any other observed changes could likely be attributed to seasonal variations and fluctuations in forage availability.

Belonging to the cluster 2 is therefore a sign that the farm feeding strategy encourages grazing. However, providing a cluster label on a routine basis is challenging because the features used in this study to perform this analysis are not routinely recorded. Therefore, there is an interest to create a model predicting the cluster based on routinely available features such as the milk FT-MIR predictions. Using milk-based predictions is relevant because it is known that milk composition is influenced by the herbage consumption on the farm (Capuano et al., 2014; Coppa et al., 2021; Manzocchi et al., 2021). Consequently, the next step is to develop a model to predict this cluster.

## ANOVA Between Clusters for the Predicted Milk FT-MIR–Based Traits

To select the most informative milk-based features and therefore achieve a more robust model (Grelet et al., 2021), an ANOVA was conducted on the 75 observations of DB_milk associated with their respective clusters in order to highlight cluster 2, assumed to reflect a diet mainly composed of herbage in various forms, from the others. Of the 54 traits examined, 25 had a $P$-value below 0.05, as detailed in Table 4. The predictions of milk yield, milk transformation traits, as well as the fat, protein, and mineral composition demonstrated discriminatory capabilities across the clusters.

The average fat content was not significantly different, and the predicted milk yield was significantly lower when

**Table 3.** Mean values and number of observations related to the found clusters

| Cluster | Occupation time (h) | Filling (%) | Total herbage feed (%) | DB_Pat (n) | DB_milk (n) |
|---|---|---|---|---|---|
| 1 | 5.81 | 81.15 | 55.74 | 307 | 45 |
| 2 | 17.89 | 42.80 | 80.82 | 123 | 18 |
| 3 | 16.26 | 139.23 | 86.41 | 31 | 4 |
| 4 | 4.71 | 110.47 | 37.20 | 65 | 8 |

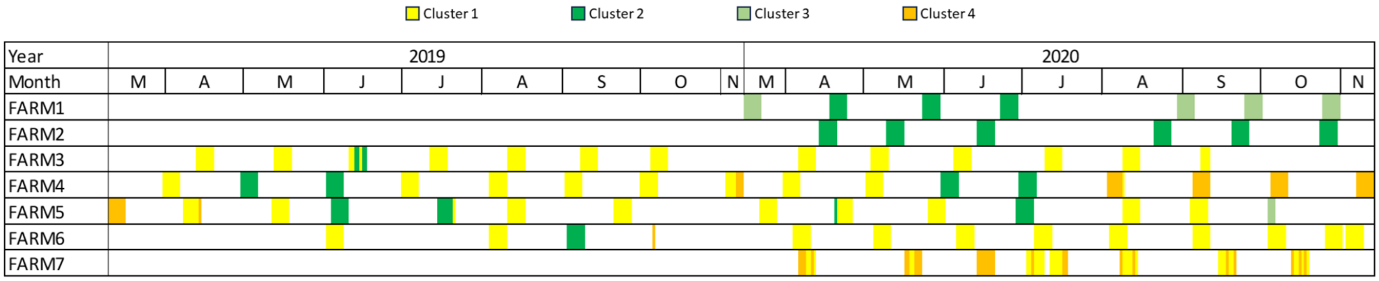


**Figure 4.** Cluster classification by farm over 2019 and 2020.

cows received herbage (Table 4). This was also observed in other studies on real measured milk yields (White et al., 2001; Gulati et al., 2018; Stergiadis et al., 2021).

Regarding the indicators about the fat composition, the results coming from other studies are more discordant, which can be explained by the influence of the season (Timlin et al., 2021; Hayes et al., 2023), the breed (Bär et al., 2020; Kostovska et al., 2024), and the exact type of feed considered. One of the most discussed FA contents when comparing grass-based diets to complementation are the CLA. They were observed to be higher in fat when a grass diet is used, as mentioned by De La Torre-Santos et al. (2020) with 2.29 versus 1.37 g/100 g fat, by White et al. (2001) with 0.72% versus 0.41% of total FA, or by Collomb et al. (2008) with 1.39 versus 1.21 g/100 g of fat. This follows the results we obtained, as CLA are part of the FT-MIR–predicted total *trans* FA, which is indeed higher with 0.20 versus 0.17 g/dL of milk or 4.92 versus 4.15 g/100 g of fat (Table 4). Following De La Torre-Santos et al. (2020) and Kučević et al. (2016), the milk fat content of C18:1 *trans* also follows the same trend as CLA, with higher proportions in the group mainly on pasture. We also observed a higher FT-MIR predicted total content of C18:1 *trans* in milk (Table 4) for cluster 2. In general, a higher total content of *trans* FA can be observed in groups related to grazing or at least herbage consumption (Chassaing et al., 2016; Cabiddu et al., 2022).

For C14:1, the value shows an opposite trend of our FT-MIR–predicted results, with seemingly higher con-

**Table 4.** The most informative FT-MIR predictions allowing to discriminate found clusters and their LSM for the 2 opposite clusters: cluster 2 for mainly herbage consumption and cluster 4 for other feeds

| FT-MIR–based traits | Unit | N | LSM Cluster 2 | LSM Cluster 4 | *P*-value |
|---|---|---|---|---|---|
| Milk yield | kg/d | 75 | 24.00 | 27.33 | 0.000*** |
| C14:1 | g/dL | 75 | 0.04 | 0.05 | 0.015* |
| C16 | g/dL | 75 | 1.15 | 1.30 | 0.043* |
| Total of C18:1 *trans* | g/dL | 75 | 0.15 | 0.13 | 0.000*** |
| MCFA | g/dL | 75 | 2.03 | 2.24 | 0.038* |
| Total of *trans* FA | g/dL | 75 | 0.20 | 0.17 | 0.000*** |
| Ca | mg/kg | 75 | 1,174.82 | 1,257.21 | 0.001** |
| P | mg/kg | 75 | 975.67 | 1,017.16 | 0.017* |
| K | mg/kg | 75 | 1,489.80 | 1,496.56 | 0.000*** |
| Mg | mg/kg | 75 | 102.77 | 107.05 | 0.000*** |
| Urea in milk | mg/L | 75 | 154.40 | 176.00 | 0.000*** |
| Lactoserum proteins | g/L | 75 | 1.24 | 1.44 | 0.000*** |
| Lactoferrin | mg/L | 75 | 116.46 | 127.56 | 0.002** |
| Protein efficiency | % | 75 | 16.22 | 14.26 | 0.049* |
| Protein N content | g/L | 75 | 26.881 | 27.70 | 0.006** |
| pH | — | 75 | 6.62 | 6.60 | 0.004** |
| Titratable acidity | Dornic degree | 75 | 14.78 | 14.35 | 0.007** |
| Yogurt activity | Dornic degree | 75 | 0.61 | 0.63 | 0.000*** |
| Yogurt texture | N | 75 | 0.23 | 0.23 | 0.005** |
| Square root A30 | mm | 75 | 36.51 | 36.44 | 0.035* |
| Isocitrate | mmol/L | 75 | 0.15 | 0.17 | 0.000*** |
| Citrate | mmol/L | 75 | 8.40 | 9.46 | 0.006** |
| Blood fructosamine | µmol/L | 75 | 265.58 | 269.55 | 0.000*** |
| Blood glucose | µmol/L | 75 | 3.69 | 3.80 | 0.001** |
| Blood IGF-1 | nmol/L | 75 | 136.29 | 140.59 | 0.004** |

$*P < 0.5$; $**P < 0.01$; $***P < 0.001$.

tent in grass-based feeding management. In White et al. (2001) in Holstein, they compare 0.80 versus 0.59 g/100 g, compared with 0.98 versus 1.22 g/100 g in our study. The FA indicator appears to be context-dependent. For example, FT-MIR–predicted C18:1 *trans* had a *P*-value higher than 0.05, indicating that the differences could not be highlighted in Ellis et al. (2006). The same happens with C14:1 in De La Torre-Santos et al. (2020) and C16 in both De La Torre-Santos et al. (2020) and White et al. (2001). Our results concerning C16 show a lower content for cluster 2, which aligns more with a smaller reported difference of −6% to −10% (Cabiddu et al., 2022) and a highly significant difference (Coppa et al., 2019).

The results yielded regarding the mineral content evolve in opposition to those from Gulati et al. (2018) for milk Ca and P contents. Indeed, those authors found the content of major minerals to be higher in the outdoor grazing group. In their study, Ca and P contents were 142.2 versus 131.8 g/100 g of milk, and 104 versus 101 g/100 g of milk, respectively. All minerals were found with a *P*-value higher than 0.05 in Stergiadis et al. (2021), who compared milk from organic and conventional farming. In contrast, our study found the FT-MIR–predicted content of CA, P, K, and Mg to also significantly differ between the 2 clusters.

Other indicators not previously reported in the literature, to our knowledge, were also found to be informative to distinguish the 2 types of milk. For example, the FT-MIR–predicted milk acidity is higher in cluster 2, whereas the FT-MIR–predicted isocitrate content is lower. As for the blood indicators, all predictions showed higher values in cluster 4.

### Cluster Predictions Through PLSDA

These 25 significant features (Table 4) were used as predictors in PLSDA models to predict the clusters. The models offer the possibility to consider a continuous variable instead of the binary attribution, thanks to the estimation of the probability to belong to a specific group. Four different models were tested and compared in this study. First, in model 1, we used the 2 extreme clusters (i.e., cluster 2 and cluster 4; Figure 3) leading us to use 26 records out of the 75. Second, because we wanted to predict cluster 2, the idea was to create a new variable where the records belonging to cluster 2 were set to 1 and the records belonging the other clusters were set to 0. This approach was realized in model 2. To predict the absence of herbage in the diet, we created another variable where the records belonging to the cluster 4 were set to 1 and the ones belonging to another cluster were set to 0. This new target was used to build model 3. Finally, for model 4, we followed the clustering dendrogram by setting the branch of clusters 2 and 3 at 1, and clusters 1 and 4 at 0. The cross-validation performances of those 4 models are exposed in Table 5, where the predictions were applied to the 75 observations. In the case of model 1, 3 options similar to the training of models 2, 3, and 4 were used to define the cross-validation set and define whether an observation was supposed to be set as 0 or 1, in order to apply the cross-validation on the 75 observations and be able to compare model 1 to the rest. Those 3 options are related to the groups to which clusters 1 and 3 were attributed. Model 2 showed the best performance, followed by model 3. Model 4 had the worst results, with a kappa value lower than 0.50, suggesting that the labeling to a specific cluster is not better than a random choice. Indeed, even if the performance on the common dataset (n = 75) is the highest for model 1, this model, being built on a lower number of observations, seems to be less robust because its cross-validation specificity is the lowest. The differences in predicted information between models can also be observed through the estimation of correlations between the probabilities to belong to a cluster (Table 6). All models are different, as no correlation is close to 1. Model 1 was more closely related to model 2, whereas models 3 and model 4 were strongly related to each other. This suggests that models 1 and

**Table 5.** Partial least squares discriminant analysis models and indicators computed with the cross-validation

| | Cluster target[1] | | | | | |
|---|---|---|---|---|---|---|
| | Model 1 | | | Model 2 | Model 3 | Model 4 |
| Item | C2 vs. C4 | | | C2 vs. C1C3C4 | C4 vs. C1C2C3 | C2C3 vs. C1C4 |
| N records | | 26 | | 75 | 75 | 75 |
| N factors | | 16 | | | 1 | 1 |
| | C2 vs. C1C3C4 | C4 vs. C1C2C3 | C2C3 vs. C1C4 | | | |
| Accuracy | 0.85 | 0.49 | 0.80 | 0.93 | 0.91 | 0.73 |
| Kappa | 0.67 | 0.14 | 0.55 | 0.82 | 0.62 | 0.40 |
| Sensitivity | 1 | 0.43 | 0.82 | 0.94 | 0.88 | 0.68 |
| Specificity | 0.81 | 1 | 0.79 | 0.93 | 0.91 | 0.75 |

[1]C1, C2, C3, C3 = cluster 1, 2, 3, and 4.

**Table 6.** Correlation between the cluster probabilities estimated from the 5 models as well as their relationships with the percentage of herbage in the diet and the Occupation_Time

| Correlation | Model1_C2[1] | Model2_C2[2] | Model3_C4[3] | Model4_C23[4] | GRASS1[5] |
|---|---|---|---|---|---|
| Model1_C2 | 1.00 | 0.82 | −0.32 | 0.26 | 0.20 |
| Model2_C2 | 0.82 | 1.00 | −0.46 | 0.39 | 0.44 |
| Model3_C4 | −0.32 | −0.46 | 1.00 | −0.88 | −0.63 |
| Model4_C23 | 0.26 | 0.39 | −0.88 | 1.00 | 0.51 |
| %Herbage | 0.35 | 0.56 | −0.56 | 0.63 | 0.43 |
| %Filling | −0.39 | −0.61 | 0.32 | −0.07 | −0.52 |
| Occupation_time | 0.35 | 0.49 | −0.45 | 0.47 | 0.31 |

[1]Probability to predict cluster 2 from model 1.

[2]Probability to predict cluster 2 from model 2.

[3]Probability to predict cluster 4 from model 3.

[4]Probability to predict the group composed of cluster 2 and 3 from model 4.

[5]Probality estimated using the model developed by Soyeurt et al. (2022).

2 use different information from models 3 and 4, given their correlation pattern.

To compare our approach based on real pasture information with the unsupervised method based on seasonal specificities made by Soyeurt et al. (2022), we predicted from our dataset the probability to belong to the GRASS1 cluster. The estimated correlations with this additional grass indicator are added in Table 6. All correlations were moderate, with the highest absolute value observed with model 3. The moderate correlation values could be related to the fact that GRASS1 might only consider the presence of herbage, whereas our clusters reflect multiple aspects associated with grazing. Indeed, in this study, even if all farms presented harvested and conserved herbage (if not grazed herbage) in their feeding practices, their typologies were different (Figure 2).

To deepen this discussion, the correlations between all those probabilities and the pasture characteristics analyzed in the current study were also estimated and added in Table 6. The highest absolute correlation with the %Herbage is observed with model 4, followed by models 2 and 3. Those 3 models have similar absolute correlations with the Occupation_Time (0.45–0.49). However, model 4 differed significantly in its relationship with %Filling, for which this model seems to be more or less independent. Model 3 is in the middle between models 2 and model 4 for this trait. The GRASS1 model presented lower correlations overall, although its tendency was closer to the one observed for model 2. In a similar way, it opposes the trend of model 3. This smaller value could also be explained by the fact that the GRASS1 model was not trained on this dataset, unlike the other models used in Table 6. For all DB_Pat traits, model 1 had the lowest absolute correlation, confirming the lack of interest that we could have for this model. Before making a choice for the best model, it is also important to observe the model behavior on a large database in order to verify if the predictions follow the expected pattern.

### Direct Prediction of the Quantitative Traits

Before applying the developed algorithm on the Walloon spectral database related to the milk payment, we completed our approach by building PLS regressions directly on the pasture variables. Their performance results are presented in Table 7. The %Herbage and the Occupation_Time were predicted with RMSE of 8.77% and 6.65 h, respectively. Because the predictor error on pasture time reached more than 6 h, representing one-fourth of the daily occupation time, this prediction appears to be unsuitable for practical applications. The prediction error for the %Filling, 18.93%, is also too high to be useful in practice. With an $R^2$ of 0.76, the prediction of the %Herbage could be considered as an interesting indicator. Its predictions were more strongly correlated with the information given by model 4. By comparison with Table 6, we can observe that the correlations were a little stronger for the prediction of %Herbage (0.63) compared with that observed using the reference %Herbage (0.72). However, the trend is still the same, potentially confirming the interest of this trait. However, some limitations of this study results should now be acknowledged. There is an uncertainty on the exact type of feed used in the

**Table 7.** Performances of equations predicting the DB_Pat traits from the 25 milk-based predictors, and correlation with the PLSDA models

| Traits | Herbage (%) | Occupation_time (h) | Filling (%) |
|---|---|---|---|
| Performance | | | |
| N records | 75 | 75 | 75 |
| N factors | 11 | 2 | 3 |
| RMSE | 8.77 | 6.65 | 18.93 |
| $R^2$ | 0.76 | 0.37 | 0.66 |
| Correlation | | | |
| GRASS1 | 0.50 | 0.67 | −0.56 |
| Model1_C2 | 0.41 | 0.43 | −0.38 |
| Model2_C2 | 0.65 | 0.69 | −0.60 |
| Model3_C4 | −0.64 | −0.83 | 0.38 |
| Model4_C2C3 | 0.72 | 0.77 | −0.06 |

category "harvested and conserved herbage." Its influence on the predicted herbage content in the diet remains difficult to estimate. Indeed, the composition of milk can vary using fresh grass, grass silage, or hay (Elgersma et al., 2004; van den Oever et al., 2021). For instance, following Manzocchi et al. (2021), the content of SFA, *trans*-11 C18:1, and PUFA were all significantly different between a fresh grass diet and the 2 groups of grass silage and hay. It would therefore be a very interesting perspective to reiterate our analysis with a broader interpretation, based on the exact type of feed, or more precisely, on feed composition.

## Cluster Prediction Upscale

The 4 developed PLSDA models were applied on the Walloon milk spectral database related to the milk payment, and their temporal evolution between 2009 and 2023 is available in Figure 5. The expected trend is an increase from spring to summer and a decrease from summer to autumn, following grazing practices in the Walloon region of Belgium. By observing all plots, there is an unexpected pattern for model 1. For most of the years, the model is not responsive and gives a probability inferior to 0.4. Model 2 has a variation that is too small within each year. Both model 1 and 2 do not match expected herbage consumption (grazed, conserved, or harvested herbage) or grazing. Models 3 and 4 both display the seasonal pattern within each year and the stability of this pattern across years. Model 4 is slightly less volatile between years and therefore seems more robust. Combined with its cross-validation performance, its higher correlation with %Herbage of 0.63, and its predicted %Herbage of 0.72, it is the model that appears to perform the best, closely followed by model 3.

Figure 6 shows the proportion of farms classified as farms encouraging grazing for each month on a specific dataset provided by the Comité du Lait, exclusively containing farms that are known to practice grazing. As an example, the probability threshold to classify was set to 0.50. In winter, the proportion of farm detected is still high, varying for the period 2023 to early 2025 ~40% to 50%. On one hand, the presence of detected grazing farms during the winter could be explained by the composition of feed, which could, for a few components, have a similar signature to herbage. For example, extruded linseed included in the diet can increase some FA in the milk composition, such as the C16, C18:1 *trans*, and n-3 content (Beauregard et al., 2023; Huang et al., 2022). On the other hand, the models were not trained to specifically differentiate the type of herbage fed to the dairy cows. The high proportion of grazing farms detected in winter could therefore be a result of

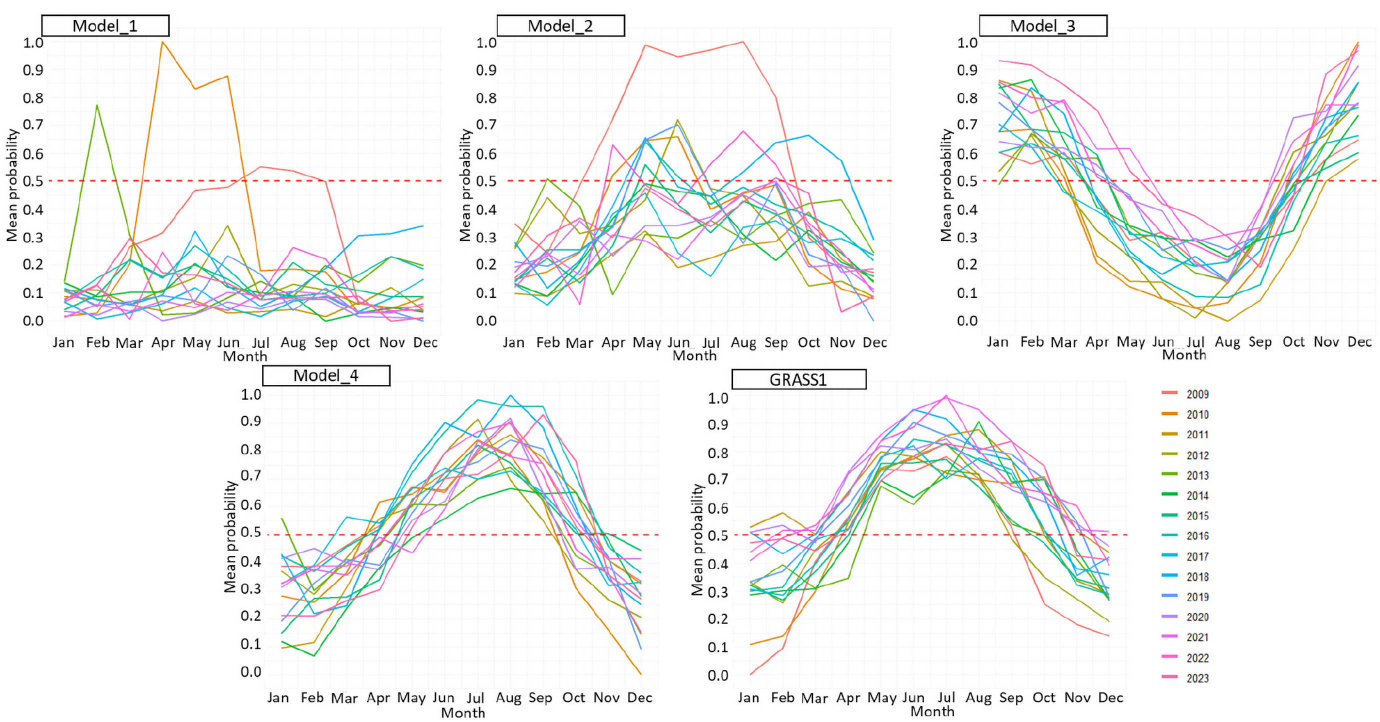


**Figure 5.** Evolution the standardized mean of Model_1, Model_2, Model_3, Model_4, and GRASS1 between 2009 and 2023 by year and by month.

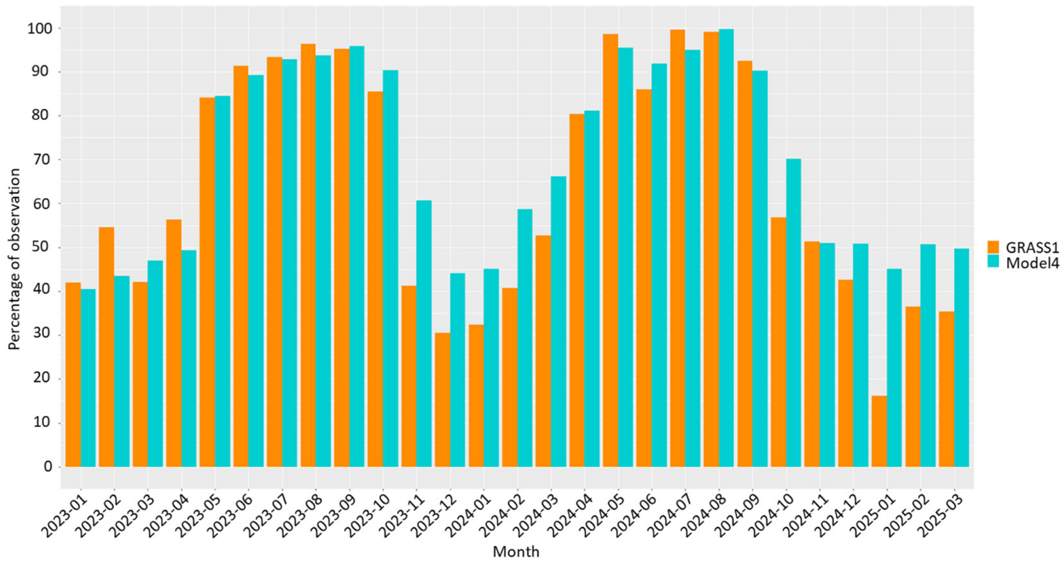


**Figure 6.** Percentage of farms practicing grazing that present values above the 0.5 threshold to detect milk based on model 4 and GRASS1.

feeding hay or silage. This hypothesis would require access to more complementary data on the TMR from several farms. Model 4 performs similarly to GRASS1, even though model 4 was not trained directly with any time variable, unlike GRASS1. This suggests that at least part of the seasonal information is contained in the effective herbage intake, the occupation time as a proxy of the grazing time, and the approximated %Filling, which were used to make the clustering. This seasonal impact could then also be transmitted through milk composition, which was used to build the models.

Consequently, relying solely on a class label would require a sensitivity analysis to select the most appropriate threshold. As it stands, the evolution of herbage probability per farm could be more informative. Indeed, for a farm practicing grazing, the probability of herbage consumption should increase during the summer.

## CONCLUSIONS

In conducting this study, several assumptions were required, inevitably introducing error into the models. Some were linked to the database, such as estimating %Filling or extrapolating herd milk composition from individual samples. Clusters were interpreted to separate feeding strategies, which added uncertainty to the conclusions. The herbage proportion in dairy cow diets was evaluated using PLS, with an RMSE of 8.77% of herbage in cross-validation. Grazing-oriented practices could also be predicted from milk analysis through PLSDA, achieving an accuracy of 0.93, sensitivity of 0.94, and specificity of 0.93. Leveraging a large database (5,886,364 spectra) further enabled discrimination of models by seasonal performance. The traits that contributed to separating herbage-based diets included both expected ones, such as FA-related variables (e.g., C18:1 *trans*, CLA), and more unexpected ones, such as acidity traits (pH, titratable acidity), protein-related traits (lactoserum proteins, protein efficiency), and citrate. Integrating herbage and grazing indicators into routine analysis could strengthen metrics, support external estimation of pasture days, and attribute grazing-based characteristics to milk.

## NOTES

This work was supported by the Service Public de Wallonie (Namur, Belgium) with the project WALLeSmart (grant number D65-1435), as well as the FNR (Esch-sur-Alzette, Luxembourg) with the grant INTERFNRS/18/12987586 and the F.R.S-FNRS (Brussels, Belgium) with the grant T.0221.19. under the bilateral Simulating Economic and Environmental Impacts of Dairy Cattle Management Using Agent-Based Models (SIMBA) project. We thank Anthony Tedde (University of Liege, Belgium) for his previous work in the SIMBA project, for his expertise, and for the time he passed on the grazing calendars. We also thank the partners of the Walloon FUTUROSPECTRE agreement, the Walloon Breeding Association (Ciney, Belgium) for providing access to data related to the Walloon milk recording, and the milk laboratory Comité du Lait (Battice, Belgium) for providing spectral data. No human or animal subjects were used, so this analysis did not require approval by an Institutional Animal Care and Use Committee or Insti-

tutional Review Board. The authors have not stated any conflicts of interest.

**Nonstandard abbreviations used:** %FAT = fat content; %Filling = proportion of supplementation relative to average ingestion capacity; %Herbage = proportion of herbage fed; %PROT = protein content; compl_herbage = quantity of harvested or conserved herbage; compl_other = quantity of feed fed to the cattle excluding herbage; DB = database; FA = fatty acid; FT-MIR = Fourier-transform mid-infrared; GPS = global positioning system; ICAR = International Committee for Animal Recording; LCFA = long-chain FA; MCFA = medium-chain FA; NP = not published; Occupation_Time = time spend on pasture; PLS = partial least square analysis; PLSDA = partial least squares discriminant analysis; RMSE = root mean squared error; RSE = relative SE; SCFA = short-chain FA; SIMBA = Simulating Economic and Environmental Impacts of Dairy Cattle Management Using Agent-Based Models.

## REFERENCES


Alothman, M., S. A. Hogan, D. Hennessy, P. Dillon, K. N. Kilcawley, M. O'Donovan, J. Tobin, M. A. Fenelon, and T. F. O'Callaghan. 2019. The "grass-fed" milk story: Understanding the impact of pasture feeding on the composition and quality of bovine milk. Foods 8:350. https://doi.org/10.3390/foods8080350.

Axema Promotion and Services. 2023. CHRONOPATURE. Pages 111–113 in AGRITECH DAY—6th edition: The International Conference of Technologies and Solutions for Efficient and Sustainable Agriculture. Paris, France. Accessed Sep. 4, 2025. https://agritechday.com/wp-content/uploads/2023/11/BookAGTD_2023_BD.pdf.

Bär, C., M. Sutter, C. Kopp, P. Neuhaus, R. Portmann, L. Egger, B. Reidy, and W. Bisig. 2020. Impact of herbage proportion, animal breed, lactation stage and season on the fatty acid and protein composition of milk. Int. Dairy J. 109:104785. https://doi.org/10.1016/j.idairyj.2020.104785.

Baumont, R., J. Dulphy, D. Sauvant, F. Meschy, J. Aufrère, and J. Peyraud. 2010. Alimentation des bovins, ovins et caprins. Besoins des animaux–Valeur des aliments. Éditions Quæ, Versailles, France. Accessed Aug. 7, 2025. https://wiki.tripleperformance.fr/images/fr/2/20/Alimentations_des_caprins%2C_ovins%2C_bovins.pdf.

Beauregard, A., M.-P. Dallaire, R. Gervais, and P. Y. Chouinard. 2023. Lactational performance of cows fed extruded flaxseed in commercial dairy herds. Animal - Open Space 2:100043. https://doi.org/10.1016/j.anopes.2023.100043.

Birkinshaw, A., M. Sutter, B. Reidy, L. Jungo, S. Mueller, M. Kreuzer, and M. Terranova. 2023. Evaluation and quantification of associations between commonly suggested milk biomarkers and the proportion of grassland-based feeds in the diets of dairy cows. PLoS One 18:e0282515. https://doi.org/10.1371/journal.pone.0282515.

Cabiddu, A., G. Peratoner, B. Valenti, V. Monteils, B. Martin, and M. Coppa. 2022. A quantitative review of on-farm feeding practices to enhance the quality of grassland-based ruminant dairy and meat products. Animal 16:100375. https://doi.org/10.1016/j.animal.2021.100375.

Capuano, E., J. Rademaker, H. van den Bijgaart, and S. M. van Ruth. 2014. Verification of fresh grass feeding, pasture grazing and organic farming by FTIR spectroscopy analysis of bovine milk. Food Res. Int. 60:59–65. https://doi.org/10.1016/j.foodres.2013.12.024.

Chassaing, C., C. Sibra, J. Verbič, O. M. Harstad, J. Golecký, B. Martin, A. Ferlay, I. Constant, C. Delavaud, C. Hurtaud, V. Ž. Pongrac, and C. Agabriel. 2016. Mineral, vitamin A and fat composition of bulk milk related to European production conditions throughout the year. Dairy Sci. Technol. 96:715–733. https://doi.org/10.1007/s13594-016-0300-7.

Christophe, O. S., C. Grelet, C. Bertozzi, D. Veselko, C. Lecomte, P. Höeckels, A. Werner, F.-J. Auer, N. Gengler, F. Dehareng, and H. Soyeurt. 2021. Multiple breeds and countries' predictions of mineral contents in milk from milk mid-infrared spectrometry. Foods 10:2235. https://doi.org/10.3390/foods10092235.

Coelho Ribeiro, L. A., T. Bresolin, G. J. Rosa, D. Rume Casagrande, M. A. C. Danes, and J. R. R. Dórea. 2021. Disentangling data dependency using cross-validation strategies to evaluate prediction quality of cattle grazing activities using machine learning algorithms and wearable sensor data. J. Anim. Sci. 99:skab206. https://doi.org/10.1093/jas/skab206.

Collomb, M., W. Bisig, U. Bütikofer, R. Sieber, M. Bregy, and L. Etter. 2008. Fatty acid composition of mountain milk from Switzerland : Comparison of organic and integrated farming systems. Int. Dairy J. 18:976–982. https://doi.org/10.1016/j.idairyj.2008.05.010.

Coppa, M., C. Chassaing, C. Sibra, A. Cornu, J. Verbič, J. Golecký, E. Engel, J. Ratel, A. Boudon, A. Ferlay, and B. Martin. 2019. Forage system is the key driver of mountain milk specificity. J. Dairy Sci. 102:10483–10499. https://doi.org/10.3168/jds.2019-16726.

Coppa, M., B. Martin, S. Hulin, J. Guillemin, J. V. Gauzentes, A. Pecou, and D. Andueza. 2021. Prediction of indicators of cow diet composition and authentication of feeding specifications of Protected Designation of Origin cheese using mid-infrared spectroscopy on milk. J. Dairy Sci. 104:112–125. https://doi.org/10.3168/jds.2020-18468.

Coulon, J., and P. d'Hour. 1994. Capacité d'ingestion comparée de génisses et de vaches de race Holstein, Montbéliarde et Tarentaise. Ann. Zootech. 43:389–392. https://doi.org/10.1051/animres:19940408.

CRAN. 2023. The Comprehensive R Archive Network. Accessed Aug. 7, 2025. https://cran.r-project.org/.

Cuvelier, C., I. Dufrasne, J.-L. Hornick, É. Froidmont, É. Knapp, and L. Istasse. 2021. L'alimentation de la vache laitière: Physiologie et besoins–Aliments, calculs de ration, indicateurs d'évaluation des déséquilibres de la ration et pathologies d'origine nutritionnelle (Livret de l'Agriculture, no 22). Service public de Wallonie, Agriculture, Ressources naturelles et Environnement. Accessed Aug. 7, 2025. https://ediwall.wallonie.be/les-livrets-de-l-agriculture-22-l-alimentation-de-la-vache-laitiere-physiologie-et-besoins-aliments-calculs-de-ration-indicateurs-d-evaluation-des-desequilibres-de-la-ration-et-pathologies-d-origine-nutritionnelle-numerique-080283.

De Brabander, D., S. De Campeneere, I. Ryckaert, and A. Anthonissen. 2011. Melkveevoeding (ILVO Mededeling no 101). Vlaamse Overheid – Departement Landbouw en Visserij. Accessed Aug. 7, 2025. https://www.rundveeloket.be/sites/default/files/inline-files/Brochure_Melkveevoeding_2011.pdf.

De La Torre-Santos, S., L. J. Royo, A. Martínez-Fernández, C. Chocarro, and F. Vicente. 2020. The mode of grass supply to dairy cows impacts on fatty acid and antioxidant profile of milk. Foods 9:1256. https://doi.org/10.3390/foods9091256.

Dewhurst, R. J., and A. P. Moloney. 2013. 8. Modification of animal diets for the enrichment of dairy and meat products with omega-3 fatty acids. Pages 257–287 in Food Enrichment with Omega-3 Fatty Acids. C. Jacobsen, N. S. Nielsen, A. F. Horn, and A.-D. M. Sørensen, ed. Woodhead Publishing. https://doi.org/10.1533/9780857098863.3.257.

Dewhurst, R. J., K. J. Shingfield, M. R. F. Lee, and N. D. Scollan. 2006. Increasing the concentrations of beneficial polyunsaturated fatty acids in milk produced by dairy cows in high-forage systems. Anim. Feed Sci. Technol. 131:168–206. https://doi.org/10.1016/j.anifeedsci.2006.04.016.

Dillon, P., J. R. Roche, L. Shalloo, and B. Horan. 2005. Optimising financial return from grazing in temperate pastures. Pages 131–147 in Utilisation of grazed grass in temperate animal systems. J. Murphy, ed. Wageningen Academic Publishers.

Elgersma, A., G. Ellen, H. van der Horst, H. Boer, P. R. Dekker, and S. Tamminga. 2004. Quick changes in milk fat composition from cows after transition from fresh grass to a silage diet. Anim. Feed Sci. Technol. 117:13–27. https://doi.org/10.1016/j.anifeedsci.2004.08.003.

Ellis, K. A., G. Innocent, D. Grove-White, P. Cripps, W. G. McLean, C. V. Howard, and M. Mihm. 2006. Comparing the fatty acid composition of organic and conventional milk. J. Dairy Sci. 89:1938–1950. https://doi.org/10.3168/jds.S0022-0302(06)72261-5.

European Commission. 2024. The common agricultural policy at a glance. European Commission. Accessed Aug. 7, 2025. https://agriculture.ec.europa.eu/common-agricultural-policy/cap-overview/cap-glance_en.

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

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

Grelet, C., E. Froidmont, L. Foldager, M. Salavati, M. Hostens, C. P. Ferris, K. L. Ingvartsen, M. A. Crowe, M. T. Sorensen, J. A. Fernandez Pierna, A. Vanlierde, N. Gengler, and E Gplus. Consortium, and F. Dehareng. 2020. Potential of milk mid-infrared spectra to predict nitrogen use efficiency of individual dairy cows in early lactation. J. Dairy Sci. 103:4435–4445. https://doi.org/10.3168/jds.2019-17910.

Grelet, C., T. Larsen, M. A. Crowe, D. C. Wathes, C. P. Ferris, K. L. Ingvartsen, C. Marchitelli, F. Becker, A. Vanlierde, J. Leblois, U. Schuler, F. J. Auer, A. Köck, L. Dale, J. Sölkner, O. Christophe, J. Hummel, A. Mensching, J. A. Fernández Pierna, H. Soyeurt, M. Calmels, R. Reding, M. Gelé, Y. Chen, N. Gengler, and E Gplus. Consortium, and F. Dehareng. 2024. Prediction of key milk biomarkers in dairy cows through milk mid-infrared spectra and international collaborations. J. Dairy Sci. 107:1669–1684. https://doi.org/10.3168/jds.2023-23843.

Grelet, C., J. A. F. Pierna, P. Dardenne, H. Soyeurt, A. Vanlierde, F. Colinet, C. Bastin, N. Gengler, V. Baeten, and F. Dehareng. 2017. Standardization of milk mid-infrared spectrometers for the transfer and use of multiple models. J. Dairy Sci. 100:7910–7921. https://doi.org/10.3168/jds.2017-12720.

Grelet, C., J. F. Pierna, H. Soyeurt, F. Dehareng, N. Gengler, and P. Dardenne. 2014. Creation of universal MIR calibrations by standardization of milk spectra: Example of fatty acids. Pages 108 in Book of Abstracts of the 65th Annual Meeting of the European Association for Animal Production. European Federation of Animal Science.

Grelet, C., H. Simon, J. Leblois, O. Christophe, M. Jattiot, N. Gaudillère, R. Reding, J. Wavreille, E. J. P. Strang, F. J. Auer, K. Goossens, E. Chevaux, N. Gengler, and F. Dehareng. 2025. Can we get information on dairy cow chronic stress biomarkers from milk mid-infrared spectra? J. Dairy Sci. 108:10186–10202. https://doi.org/10.3168/jds.2025-26502.

Grelet, C., A. Vanlierde, M. Hostens, L. Foldager, M. Salavati, K. L. Ingvartsen, M. Crowe, M. T. Sørensen, É. Froidmont, C. P. Ferris, C. Marchitelli, F. Becker, T. Larsen, F. Carter, F. Dehareng, and E Gplus. Consortium. 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.

Gulati, A., N. Galvin, E. Lewis, D. Hennessy, M. O'Donovan, J. J. McManus, M. A. Fenelon, and T. P. Guinee. 2018. Outdoor grazing of dairy cows on pasture versus indoor feeding on total mixed ration: Effects on gross composition and mineral content of milk during lactation. J. Dairy Sci. 101:2710–2723. https://doi.org/10.3168/jds.2017-13338.

Hammami, H., F. Colinet, C. Bastin, C. Grelet, A. Vanlierde, F. Dehareng, and N. Gengler. 2017. Using milk based biomarkers to monitor the physiological state of dairy cows in large populations. 68th Annual Meeting of the European Federation of Animal Science, Tallinn, Estonia. European Federation of Animal Science. https://www.cra.wallonie.be/fr/using-milk-based-biomarkers-to-monitor-the-physiological-state-of-dairy-cows-in-large-populations-2017.

Haskell, M. J., L. J. Rennie, V. A. Bowell, M. J. Bell, and A. B. Lawrence. 2006. Housing system, milk production, and zero-grazing effects on lameness and leg injury in dairy cows. J. Dairy Sci. 89:4259–4266. https://doi.org/10.3168/jds.S0022-0302(06)72472-9.

Hayes, E., D. Wallace, C. O'Donnell, D. Greene, D. Hennessy, N. O'Shea, J. T. Tobin, and M. A. Fenelon. 2023. Trend analysis and prediction of seasonal changes in milk composition from a pasture-based dairy research herd. J. Dairy Sci. 106:2326–2337. https://doi.org/10.3168/jds.2021-21483.

Huang, G., J. Wang, K. Liu, F. Wang, N. Zheng, S. Zhao, X. Qu, J. Yu, Y. Zhang, and J. Wang. 2022. Effect of flaxseed supplementation on milk and plasma fatty acid composition and plasma parameters of Holstein dairy cows. Animals (Basel) 12:1898. https://doi.org/10.3390/ani12151898.

ICAR (International Committee for Animal Recording). 2023. ICAR Guidelines. Section 2 – Procedure 1. Computing •••:24 (-Hour Yield. ICAR, Rome, Italy.).

Iqbal, M. W., I. Draganova, P. C. H. Morel, and S. T. Morris. 2021. Validation of an accelerometer sensor-based collar for monitoring grazing and rumination behaviours in grazing dairy cows. Animals (Basel) 11:2724. https://doi.org/10.3390/ani11092724.

Kostovska, R., B. Horan, G. Drouin, J. T. Tobin, T. F. O'Callaghan, A. L. Kelly, and L. G. Gómez-Mascaraque. 2024. Effects of multispecies pasture diet and cow breed on milk composition and quality in a seasonal spring-calving dairy production system. J. Dairy Sci. 107:10256–10267. https://doi.org/10.3168/jds.2024-24975.

Kučević, D., S. Trivunović, V. Bogdanović, K. Čobanović, D. Janković, and D. Stanojević. 2016. Composition of raw milk from conventional and organic dairy farming. Biotechnol. Anim. Husb. 32:133–143. https://doi.org/10.2298/BAH1602133K.

Kuhn, M., J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, R Core Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, C. Candan, and T. Hunt. 2023. caret: Classification and Regression Training (Version 6.0-94) [Logiciel]. Accessed Aug. 7, 2025. https://cran.r-project.org/web/packages/caret/index.html.

Leduc, A., S. Souchet, M. Gelé, F. Le Provost, and M. Boutinaud. 2021. Effect of feed restriction on dairy cow milk production: A review. J. Anim. Sci. 99:skab130. https://doi.org/10.1093/jas/skab130.

Legrand, A. L., M. G. von Keyserlingk, and D. M. Weary. 2009. Preference and usage of pasture versus free-stall housing by lactating dairy cattle. J. Dairy Sci. 92:3651–3658. https://doi.org/10.3168/jds.2008-1733.

Liland, K. H., B.-H. Mevik, R. Wehrens, and P. Hiemstra. 2023. pls: Partial Least Squares and Principal Component Regression (Version 2.8-3) [Logiciel]. https://cran.r-project.org/web/packages/pls/index.html.

Mambrini, M., J. Peyraud, H. Hetault, A. Stephant, and M. Texier. 1994. Mean retention time in digestive tract and digestion of fresh perennial ryegrass by lactating dairy cows: Influence of grass maturity and comparison with a maize silage diet. Reprod. Nutr. Dev. 34:9–23. https://doi.org/10.1051/rnd:19940102.

Manzocchi, E., B. Martin, C. Bord, I. Verdier-Metz, M. Bouchon, M. De Marchi, I. Constant, K. Giller, M. Kreuzer, J. Berard, M. Musci, and M. Coppa. 2021. Feeding cows with hay, silage, or fresh herbage on pasture or indoors affects sensory properties and chemical composition of milk and cheese. J. Dairy Sci. 104:5285–5302. https://doi.org/10.3168/jds.2020-19738.

Musati, M., M. Coppa, C. Delbès, I. Verdier-Metz, M. Popova, V. Niderkorn, M. Bouchon, Y. Farizon, F. Enjalbert, M. Renna, C. Lussiana, G. Mangione, B. Martin, and A. Ferlay. 2025. The ruminal and faecal microbiota, digestion processes, and milk composition of dairy cows are modified by the botanical biodiversity of pastures. Animal 19:101537. https://doi.org/10.1016/j.animal.2025.101537.

O'Brien, B., and D. Hennessy. 2017. Scientific appraisal of the Irish grass-based milk production system as a sustainable source of premium quality milk and dairy products. Ir. J. Agric. Food Res. 56:120–129. https://doi.org/10.1515/ijafr-2017-0011.

O'Callaghan, T., D. Hennessy, S. Mcauliffe, J. Sheehan, K. Kilcawley, P. Dillon, R. P. Ross, and C. Stanton. 2018. The effect of cow feeding system on the composition and quality of milk and dairy products.

Paper presented at the Grass-Fed Dairy Conference, Naas, Co. Kildare, Ireland. Accessed Aug. 7, 2025. https://www.researchgate.net/publication/328687482_The_effect_of_cow_feeding_system_on_the_composition_and_quality_of_milk_and_dairy_products.

Olmos, G., L. Boyle, A. Hanlon, J. Patton, J. J. Murphy, and J. F. Mee. 2009. Hoof disorders, locomotion ability and lying times of cubicle-housed compared to pasture-based dairy cows. Livest. Sci. 125:199–207. https://doi.org/10.1016/j.livsci.2009.04.009.

POSIT. 2024. Posit. https://www.posit.co/.

Soyeurt, H., C. Gerards, C. Nickmilder, J. Bindelle, S. Franceschini, F. Dehareng, D. Veselko, C. Bertozzi, N. Gengler, A. Marvuglia, A. Bayram, and A. Tedde. 2022. Prediction of indirect indicators of a grass-based diet by milk Fourier transform mid-infrared spectroscopy to assess the feeding typologies of dairy farms. Animals (Basel) 12:2663. https://doi.org/10.3390/ani12192663.

Stergiadis, S., N. Qin, G. Faludi, S. Beauclercq, J. Pitt, N. Desnica, A. H. Pétursdóttir, E. E. Newton, A. E. Angelidis, I. Givens, D. J. Humphries, H. Gunnlaugsdóttir, and D. T. Juniper. 2021. Mineral concentrations in bovine milk from farms with contrasting grazing management. Foods 10:2733. https://doi.org/10.3390/foods10112733.

The Forage and Grazing Terminology Committee. 2009. Grazing Terminology. Forage Information System. Accessed Aug. 7, 2025. https://forages.oregonstate.edu/fi/topics/pasturesandgrazing/grazingsystemdesign/grazingterminology.

Timlin, M., J. T. Tobin, A. Brodkorb, E. G. Murphy, P. Dillon, D. Hennessy, M. O'Donovan, K. M. Pierce, and T. F. O'Callaghan. 2021. The impact of seasonality in pasture-based production systems on milk composition and functionality. Foods 10:607. https://doi.org/10.3390/foods10030607.

van den Oever, S. P., A. Haselmann, M. Schreiner, B. Fuerst-Waltl, Q. Zebeli, H. K. Mayer, and W. Knaus. 2021. Hay versus silage: Does hay feeding positively affect milk composition? Int. Dairy J. 118:105024. https://doi.org/10.1016/j.idairyj.2021.105024.

Ward, J. H. 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58:236–244. https://doi.org/10.1080/01621459.1963.10500845.

Wattiaux, M. A. 1995. Dairy Essentials: Nutrition and Feeding, Reproduction and Genetic Selection, Lactation and Milking, Raising Dairy Heifers. Babcock Institute for International Dairy Research and Development, Madison, WI.

White, S. L., J. A. Bertrand, M. R. Wade, S. P. Washburn, J. T. Green Jr., and T. C. Jenkins. 2001. Comparison of fatty acid content of milk from Jersey and Holstein cows consuming pasture or a total mixed ration. J. Dairy Sci. 84:2295–2301. https://doi.org/10.3168/jds.S0022-0302(01)74676-0.

Woodward, S. J. R., M. B. Neal, and P. S. Cross. 2019. Preliminary investigation into the feasibility of combining satellite and GPS data to identify pasture growth and grazing. J. N. Z. Grasslands 81:47–54. https://doi.org/10.33584/jnzg.2019.81.404.

## ORCIDS

Killian Dichou, https://orcid.org/0009-0003-1633-7003
Charles Nickmilder, https://orcid.org/0000-0001-5235-2145
Gérard Conter, https://orcid.org/0009-0009-3142-9196
Romain Reding, https://orcid.org/0009-0008-5946-9459
Antonino Marvuglia, https://orcid.org/0000-0002-8360-8040
Hélène Soyeurt https://orcid.org/0000-0001-9883-9047