LIÈGE université
Gembloux
Agro-Bio Tech

Developing decision-support tools with economic and environmental interests for the dairy sector using milk mid-infrared spectrometry

Pauline Delhez

COMMUNAUTÉ FRANÇAISE DE BELGIQUE
UNIVERSITÉ DE LIÈGE – GEMBLOUX AGRO-BIO TECH

Developing decision-support tools with economic and environmental interests for the dairy sector using milk mid-infrared spectrometry

Pauline DELHEZ

Dissertation originale présentée en vue de l'obtention du grade de docteur en sciences agronomiques et ingénierie biologique

Promoteurs : Prof. Hélène Soyeurt & Prof. Nicolas Gengler
Année civile : 2020

Abstract

Maintaining satisfying economic outcomes and limiting environmental impacts are key challenges in dairy farming today and this requires good decision-making regarding actions to make on farms. The analysis of milk by Fourier-transform mid-infrared (MIR) spectrometry provides valuable information on milk composition. This technique has already demonstrated qualities to support decision-making, for example, through the well-established predictions of milk fat and protein contents or the latest development of prediction models for novel traits. However, its full potential remains partly uninvestigated. Hence, the objective of this thesis was to contribute to the development of decision-support tools with economic and environmental interests for the dairy sector using milk MIR spectrometry. The research conducted in the framework of this thesis covered three different approaches of using MIR for decision support: (1) the development of a MIR calibration equation to predict a trait of interest, (2) the development of a test-day model to predict milk MIR spectra for management purposes, and (3) the combination of MIR-predicted data with other data streams as a means of providing additional information for decision-making.

First, we explored different strategies to predict the pregnancy status of dairy cows (pregnant vs. open) in Australia using milk MIR spectra and partial least squares discriminant analysis. Correctly identifying the pregnancy status of cows is imperative for a profitable dairy farm. Early pregnancy could not be detected satisfactorily, but promising results were obtained using MIR spectra recorded 151 days or more after insemination (i.e., mid- and late gestation), with the area under the receiver operating characteristic curve of 0.76 on the testing set. A potential application that needs to be explored further is the development of a screening tool to detect mid- to late-term fetal abortion.

Secondly, we studied the ability of a test-day mixed model to predict milk MIR spectra from first parity Holstein cows for management purposes (e.g., for the detection of problems, simulations, predictions of future data). The spectral data used for modeling originated from the Walloon milk recording database. The average correlation between observed and predicted values of each spectral wavenumber was 0.85 for the modeling set and ranged from 0.36 to 0.62 for different scenarios that corresponded to situations with more or less information known about the cows. Correlations between milk fat, protein and lactose contents predicted from the observed spectra and from the modeled spectra ranged from 0.83 to 0.89 for the modeling set and from 0.32 to 0.73 for the scenarios. These results demonstrated a moderate but promising ability to predict milk MIR spectra using a test-day model. Different improvements of the model are possible before potential practical applications that could have economic or environmental implications for dairy farming, depending on the MIR traits subsequently predicted from the modeled spectra.

Thirdly, we investigated the univariate relationships (correlations) between dairy cow enteric methane (CH_4) production (g/day) predicted from milk MIR spectra and 42 technico-economic variables from 206 Walloon dairy herds over a period of 8 years. Enteric CH_4 is an important part of the carbon footprint of milk production. Significant correlations ranged between |0.06| and |0.38|. Low MIR CH_4 production tended to be associated with more extensive or suboptimal management practices, which could lead to lower profitability. The observed weak correlations suggest intricate interactions between MIR CH_4 and technico-economic variables due to the use of real farm data with large variability in management practices. This implies the need for further research to unravel these complex relationships for a better understanding of factors associated with CH_4 production on dairy farms in order to better target mitigation strategies.

Lastly, we discussed, in the light of the research carried out in this thesis, strengths as well as issues and considerations regarding the development of decision-support tools using milk MIR. In particular, key strengths of MIR are the low cost and rapidity of the technology as well as the standard procedures for milk sample collection and analysis, allowing the acquisition of MIR data on a large scale for the development of various customized tools to assist decision-making on dairy farms. Issues and considerations covered the prediction of indirect MIR traits, the quality and variability of spectral and reference data, the choice and validation of models, the utilization of MIR indicators, the study of MIR traits in the population, the timing of milk sampling, and the uptake of MIR tools by farmers.

In conclusion, this thesis contributed (1) to establish the first steps of the development of new MIR tools and studies to support decision-making in dairy farming with potential economic and environmental benefits; and (2) to gain insight into the benefits and considerations of using milk MIR for the development of decision-support tools.

Résumé

Maintenir des résultats économiques satisfaisants et limiter les impacts environnementaux sont des enjeux essentiels en élevage laitier de nos jours et cela demande des prises de décisions adéquates par rapport aux actions à réaliser dans les exploitations laitières. L'analyse du lait par la spectrométrie moyen infrarouge à transformée de Fourier (MIR) fournit des informations intéressantes sur la composition du lait. Cette technique a déjà démontré des qualités pour accompagner les prises de décisions, par exemple via les prédictions des teneurs en matière grasse et protéine du lait déjà bien établies ou les récents développements de modèles de prédiction pour des caractères nouveaux. Cependant, son plein potentiel reste encore partiellement inexploré. De ce fait, l'objectif de cette thèse est de contribuer au développement d'outils d'aide à la décision avec des intérêts économiques et environnementaux pour le secteur laitier en utilisant la spectrométrie MIR. Les recherches réalisées dans le cadre de cette thèse couvrent trois approches différentes de l'utilisation de la spectrométrie MIR pour l'aide à la décision : (1) le développement d'une équation de calibration MIR pour prédire un caractère d'intérêt, (2) le développement d'un modèle jour de test pour prédire le spectre MIR du lait dans un but de management, et (3) la combinaison de données prédites à partir de la spectrométrie MIR avec d'autres sources de données afin de fournir des informations supplémentaires pour les prises de décisions.

Premièrement, différentes stratégies ont été explorées pour prédire le statut de gestation des vaches laitières (gestantes vs. non gestantes) sur base de spectres MIR du lait collectés en Australie et à l'aide de la méthode d'analyse discriminante par les moindres carrés partiels (PLS-DA). Identifier de façon correcte le statut de gestation des vaches est essentiel pour la rentabilité des exploitations laitières. La gestation à un stade précoce n'a pas pu être détectée de manière satisfaisante, mais des résultats prometteurs ont été obtenus en utilisant des spectres MIR collectés 151 jours ou plus après insémination (c.-à-d. en milieu et fin de gestation), avec une aire sous la courbe ROC de 0.76 pour le jeu de valdiation. Une application potentielle à explorer davantage serait le développement d'un un outil de screening pour alerter quant à de potentiels avortements à des stades moyens à avancés.

Deuxièmement, nous avons étudié la capacité d'un modèle jour de test mixte pour prédire le spectre MIR du lait de vaches Holstein en première lactation dans un but de management (p.ex. détection de problèmes, simulations, prédictions de données futures). Les données spectrales utilisées pour la modélisation proviennent du contrôle laitier en Wallonie. La corrélation moyenne entre les valeurs observées et prédites pour chaque nombre d'onde du spectre est de 0.85 pour le jeu de modélisation et varie entre 0.36 et 0.62 pour différents scénarios correspondant à des situations où on connait plus ou moins d'informations à propos des vaches. Les corrélations entre les teneurs en matière grasse, protéine et lactose prédites sur base du spectre observé et sur base du spectre modélisé varient de 0.83 à 0.89 pour le jeu de modélisation et de 0.32 à 0.73 pour les scénarios. Ces résultats montrent une capacité moyenne mais

prometteuse de prédire le spectre MIR du lait à partir d'un modèle jour de test. Différentes améliorations du modèle sont possibles avant de potentielles applications pratiques qui pourraient avoir des implications économiques ou environnementales selon les caractères prédits à partir du spectre modélisé.

Troisièmement, nous avons analysé les relations univariées (corrélations) entre les émissions de méthane (CH_4 ; g/jour) entérique des vaches laitières prédites à partir du spectre MIR du lait et 42 variables technico-économiques pour 206 troupeaux laitiers Wallon sur une période de 8 ans. Le CH_4 entérique constitue une partie importante de l'empreinte carbone de la production de lait. Les corrélations significatives varient entre |0.06| et |0.38|. Des émissions plus faibles de MIR CH_4 ont tendance à être associées avec des pratiques de management plus extensives ou sous-optimales, qui peuvent conduire à des résultats économiques plus faibles. Les faibles corrélations observées suggèrent des relations complexes entre le MIR CH_4 et les variables technico-économiques du fait de l'utilisation de données réelles issues d'exploitations avec une large variabilité de pratiques de management. Cela implique la nécessité de réaliser des études plus approfondies afin de décrypter ces relations complexes pour une meilleure compréhension des facteurs associés aux émissions de CH_4 afin de mieux cibler les stratégies de réduction des gaz à effet de serre dans les fermes laitières.

En dernier lieu, nous avons discuté, à la lumière des recherches réalisées dans cette thèse, de certains points positifs et points d'attention concernant le développement d'outils d'aide à la décision en utilisant la spectrométrie MIR du lait. Plus particulièrement, les avantages principaux de la spectrométrie MIR sont le coût peu élevé et la rapidité des analyses ainsi que les procédures standardisées pour la collecte et l'analyse des échantillons, permettant l'obtention de données MIR à large échelle afin de développer différents outils personnalisés pour accompagner les prises de décision dans les fermes laitières. Les points d'attention discutés concernent les prédictions de caractères MIR indirects, la qualité et la variabilité des données de référence et des données spectrales, le choix et la validation des modèles utilisés, l'utilisation d'indicateurs MIR, l'étude des caractères MIR dans une population, la fréquence de collecte des échantillons de lait, et l'adoption des outils MIR par les éleveurs.

En conclusion, cette thèse (1) contribue à établir les premières étapes du développement de nouveaux outils et études MIR pour accompagner les prises de décisions en élevage laitier, avec de potentiels bénéfices économiques et environnementaux ; et (2) contribue à une meilleure compréhension des bénéfices et des éléments à considérer lors de l'utilisation de la spectrométrie MIR pour développer des outils d'aide à la décision.

Acknowledgments

Firstly, I would like to express my sincere gratitude to the National Fund for Scientific Research (FNRS). This thesis would not have been possible without their financial support through the "Research Fellow" fellowship, and for the international conferences and the research stay in Australia.

I am deeply indebted to my supervisors, Hélène Soyeurt and Nicolas Gengler, for their continuous guidance and support, as well as for sharing their knowledge in their respective fields. This has been crucial to achieve this thesis.

I am very grateful to Jennie Pryce and Phuong Ho from Agriculture Victoria Research (AgriBio, Melbourne) for welcoming me to their research department for 6 months as part of my research stay in Australia, as well as for their precious guidance.

I would also like to thank the members of my thesis committee and the members of the jury: Frédéric Dehareng, Didier Stilmant, Jérôme Bindelle, Yves Brostaux, and Phuong Ho, for their insightful comments and suggestions.

Many thanks to Frédéric Colinet for helping me with data on the servers, to Astrid Loriers and Florence Van Stappen from the CRA-W for exchanging information and ideas about DECiDE and environmental aspects of dairy production, and to Eric Froidmont and his team for accepting my presence during experiments of the Grassmilk project at the farm of the CRA-W. I wish to thank Amélie Vanlierde for her explanations about enteric methane emissions and the equations she developed. I am also grateful to Edouard Reding for providing data from the awé accounting database. Moreover, I gratefully acknowledge the co-authors of my publications for their help.

Finally, I would like to offer my special thanks to my colleagues from the SIMa for the very pleasant work atmosphere and for their everyday support.

Table of contents

List of figures

List of tables

List of abbreviations

- ANN: artificial neural network
- AUC: area under the receiver operating characteristic curve
- BHB: beta-hydroxybutyrate
- BLUP: best linear unbiased prediction
- Ca: calcium
- calf_meat: value of meat production from calves, expressed per cow
- calf_meat.L: value of meat production from calves, expressed per litre fat and protein corrected milk
- calves_price: price of sold calves
- CH_4: methane
- CI: calving interval
- cm^{-1}: 1/centimetre, wavenumber unit
- CO_2: carbon dioxide
- concentrate_feed: kilograms of concentrate equivalents fed to dairy cows per year
- concentrate_feed.L: kilograms of concentrate equivalents fed to dairy cows per litre fat and protein corrected milk
- conserved_forages: percentage of conserved forages in the forage area
- cow_meat: value of meat production from cows, expressed per cow
- cow_meat.L: value of meat production from cows, expressed per litre fat and protein corrected milk
- culled_cows_price: price of culled cows
- culling_age: age of culled cows
- culling_percent: culling rate
- d: day
- DIM: days in milk
- DMI: dry matter intake
- dL: decilitre
- EU: European Union
- FA: fatty acid
- fat_percent: milk fat percentage
- feed_costs: purchased feed costs per cow
- feed_costs.L: purchased feed costs per litre fat and protein corrected milk
- first_calving: age at first calving
- fixed_costs: fixed costs per cow
- fixed_costs.L: fixed costs per litre fat and protein corrected milk
- forage_area: size (in hectares) of the forage area
- forage_area_costs: variables costs related to the forage area, expressed per cow
- forage_area_costs.L: variables costs related to the forage area, expressed per litre fat and protein corrected milk
- forage_milk: litres of milk per cow produced from forages
- FPCM: fat and protein corrected milk

- g: gram
- Gen: additive genetic effect
- GH: standardized Mahalanobis distance
- GHG: greenhouse gas
- grass_milk: litres of milk per cow produced from grass
- gross_margin: gross margin per cow
- gross_margin.L: gross margin per litre fat and protein corrected milk
- H_2: hydrogen
- ha: hectare
- herd_costs: herd costs per cow
- herd_costs.L: herd costs per litre fat and protein corrected milk
- herd_year: associated with each herd and year
- HTDr: herd-test-day effect
- HTM: herd-test-month effect
- HTY: herd-test-year effect
- HYMIR-CH_4: methane production for a specific herd and year predicted from milk Fourier-transform mid-infrared spectra
- IR: infrared
- K: potassium
- kg: kilogram
- KNN: k-nearest neighbors
- l: litre
- LU: livestock unit
- LU.ha: number of livestock units per hectare of forage area
- meat.ha: meat production per hectare of forage area
- milk.ha: milk per hectare of forage area
- milk_production: value of milk production per cow
- milk_production.L: value of milk production per litre fat and protein corrected milk
- MIR: Fourier-transform mid-infrared
- MIR-CH_4: methane production predicted from milk Fourier-transform mid-infrared spectra
- mo: month
- n_calvings: average number of calvings per cow per year
- n_cows: number of dairy cows
- n_cows.labour: number of dairy cows per labour supply
- nitrogen_grass: kilograms of nitrogen applied on grasslands per year
- no.: number
- Nobs: number of observations
- PC: principal component
- PCA: principal component analysis
- PE: permanent environmental effect
- PLS-DA: partial least squares discriminant analysis
- profit: profit per cow

- profit.L: profit per litre fat and protein corrected milk
- prot_percent: milk protein percentage
- r: correlation
- R^2: coefficient of determination
- R^2cv: cross-validation coefficient of determination
- REML: restricted maximum likelihood
- RMSE: root-mean-square error
- ROC: receiver operating characteristic
- SCC: somatic cell count
- SD: standard deviation
- SECV: standard error of cross-validation
- SF_6: sulfur hexafluoride
- SVM: support vector machine
- TDM: test-day model
- TMR: total mixed ration
- VFA: volatile fatty acid

1

General introduction

1. Context: current challenges of dairy farming

For centuries, milk and dairy products have been an important source of nutrients for the global population. In the European Union (EU), milk production is the second largest agricultural sector, representing more than 13% of total agricultural outputs in 2018 (Eurostat, 2019). Total milk production was estimated at about 170 million tons per year and there were around 21 million cows in the EU in 2018 (European Commission, 2019a; Eurostat, 2019). However, the European and global dairy sector has been facing many changes and challenges in recent years.

First, there is reduced governmental involvement in regulating agricultural commodity prices worldwide, such as the abolishment of milk quotas in the EU in 2015 and the openness to the world market, or the recent international trade agreements (e.g., EU-Mercosur), leading to fluctuating (typically declining) milk prices to align with global prices (Barkema et al., 2015; DG AGRI, 2017; European commission, 2019b). The fluctuating and low milk price is a major factor in reduced profitability. On top of that, some countries like Belgium are noticing an increase in agricultural land price, mainly driven by an increasing demand for land for non-agricultural uses (La Spina, 2014).

Besides, the trend toward fewer but larger dairy farms continues in the EU and across the globe (Barkema et al., 2015). Since the abolishment of milk quotas, farmers in the EU are facing growing pressures to focus on the economy of scale by increasing the size of their herds (Norton and Berckmans, 2017). At the same time, there is lower involvement of family labour on the farm and more farms depend on non-family labour, often expensive and sometimes less skilled (La Spina, 2014). More animals and less workforce lead to more complex and time-consuming monitoring and management of cows, potentially affecting farm profitability (Norton and Berckmans, 2017).

Additionally, there are increasing regulations related to consumer protection, food safety and quality, sanitary measures (i.e., preventing animal and zoonotic diseases, and reduction of medical treatment), and the environment as well as increasing administrative work (Barkema et al., 2015; Turlot, 2019). Also, the consumers, influencing processors and retailers, are being increasingly concerned with animal welfare, food safety, health, and environmental protection (De Graaf et al., 2016). This requires updated infrastructures, adapted management practices, better monitoring of animals, and tools to control food quality, environmental parameters as well as animal welfare. In this context, farmers can face costs for compliance with regulations, which can impact profitability (DG AGRI, 2017).

Moreover, climate change and environmental disruptions will likely become increasingly important in the future. Global temperature has increased steadily in the last decades and is expected to continue (Allen et al., 2018). In many areas of the world, scientists forecast warmer temperatures year-round, greater variation in precipitations (i.e., periods of drought and excess rainfall) and more severe weather incidents (Seneviratne et al., 2012). Changing climatic conditions lead to more difficult crop and fodder production for feed as well as animal health and welfare issues (e.g., heat stress), also exerting weight on farm profitability (Raitens and Combs, 2019). Also, assessing the environmental impacts of dairy farms (e.g., soil degradation, ecotoxicity, greenhouse gas emissions) is necessary to better target actions to mitigate environmental footprint.

Lastly, there is a need to improve on-farm working conditions and the quality of life of farmers. Farmers seek less physical work, to save time, more and more flexibility in organizing their work (i.e., to adapt to their family life) and to reduce the mental workload by anticipating events such as inseminations or health problems (Hostiou et al., 2017).

Summing up, dairy farming in the EU and across the globe is facing many challenges related to the three pillars of sustainability, i.e., economic, environmental and social aspects (Wezel et al., 2014).

2. The need for new decision-support tools in dairy farming

In this context, there is an increasing importance of good decision-making regarding actions to make on farms in order to maintain sustainable dairy farming, and there is a need for the development of new management tools to cope with these different challenges. The farmer has a key role in farm decision-making, but he is not the only one to take decisions regarding dairy farming. Several stakeholders (agricultural advisors, feed or fertilizer salesmen, etc) may be involved in the analysis of problems and the selection of appropriate solutions. Also, the authorities can encourage certain practices by specific support, or restrict others as a means of regulation (Meynard et al., 2002).

Traditionally, farmers and stakeholders made their decisions based on intuitive methods, experiential learning, experts' knowledge, or statistics from historical records (Jago et al., 2013). The two last approaches can lead to approximate decisions that can be too generic (i.e., not adapted to each farm and not specific enough for individual cows) and insensitive to the unstable changes in the market and environment. Dairy farms are complex systems with many interacting inputs and outputs. Each type of feed crop or animal has its own set of interacting components upon which observations need to be made, decisions need to be taken, and operations need to be controlled (Figure 1-1, Heinemann, 2009). To deal with decision-making in such complex systems and to customize what actions to make for each specific farm, decision-support tools have been developed to guide the farmers and stakeholders towards the best alternative solutions from an economic, social or environmental point of view. Decision-support tools assist the users in their decision process and support, rather than replace, human judgment (Moureaux, 2016). There are different types of decision-support tools, with many of them being computer-based. For instance, these can be dynamic tools, whose recommendations vary according to the users' inputs, and they lead the users through clear steps and suggest optimal decisions (Rose et al., 2016). Other decision-support tools may not be dynamic, but rather descriptive or predictive and act as information sources to improve the evidence base for decisions. These can use field observations and measurements or associated predictions to adjust interventions to the farm (Meynard et al., 2002; Rose et al., 2016).

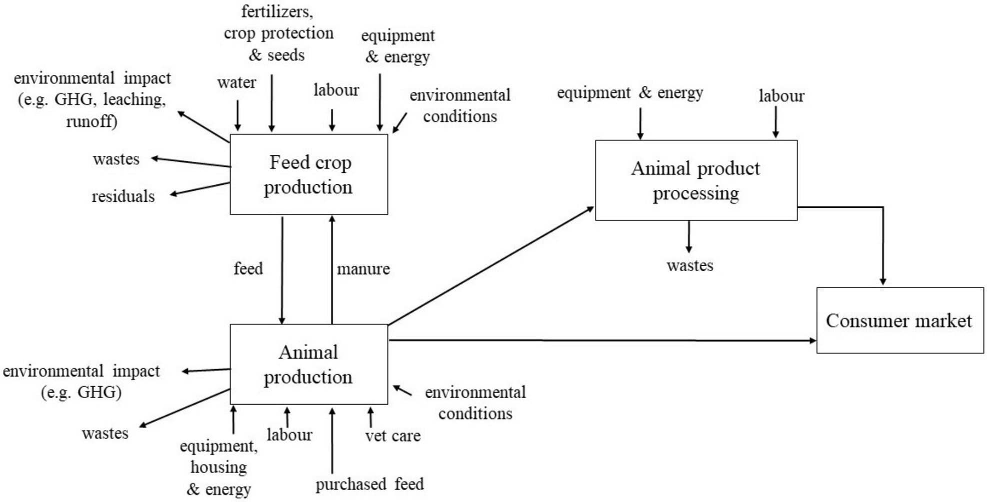

Figure 1-1. Diagram of a specialized dairy farming system (based on Heinemann, 2009).

Many management decision-support tools have been developed so far to assist dairy farmers and stakeholders. They cover different components of the dairy farming system (Figure 1-1). Examples of operational decision-support tools include tools for fertilizers and manure recommendations (e.g., Valor, CRA-W, valor.cra.wallonie.be; Date N'Prairie, Arvalis, www.datenprairie.arvalis-infos.fr), the assessment of environmental impact (e.g., DeCIDE, CRA-W, decide.cra.wallonie.be), heat detection (e.g., Heatime, SCR, www.fr.scrdairy.com/cow-intelligence/heatime-hr-systeme.html), feed management (e.g., Eva'lait, INRA, evalait.bretagne.chambagri.fr), grassland management (e.g., GrassMan, www.grassman.fr) or herd health management (e.g., Parasit'sim, idele.fr/services/outils/parasitsim-outil-de-simulation-du-risque-parasitaire.html).

The availability of high-performance computing resources, new technologies (i.e., precision farming), and mass data in dairy farming are prompting the development of new and more customized decision-support tools (Eastwood et al., 2012; Jago et al., 2013). Among the potential technologies, Fourier-transform mid-infrared (MIR) spectrometry is already routinely used in the dairy industry in many areas of the world to analyze major milk components (e.g., fat, protein, lactose contents) for milk payment and individual milk recording (ICAR, 2017). Some advantages of this technology are the reasonable cost of analysis and rapid testing, allowing for large-scale use. MIR has already demonstrated its ability for the development of some management decision-support tools, which will be discussed later, but its full utility is still partly unexplored.

3. MIR analysis of milk

In this section, we will give a more detailed explanation of the general concept of spectroscopy and spectrometry, Fourier-transform mid-infrared (MIR) spectrometry, and the use of MIR for the analysis of milk.

Spectroscopy is the study of the interaction between matter and electromagnetic radiation. Electromagnetic radiation comprises different regions depending on the wavelengths. The infrared region is lying between the visible and microwave regions (from ~ 800 nm to ~1 mm) and is used to identify and study chemical substances, either solid, liquid, or gaseous. When matter is crossed by the infrared electromagnetic radiation, the bonds between the atoms of a molecule vibrate at a precise frequency, interact with infrared rays having the same frequency and absorb the energy of the rays (Subramanian and Rodriguez-Saona, 2009; Khan et al., 2018). On the basis of supplied radiation energy and the amount passing through or absorbed by the sample, it is possible to determine an absorption spectrum representing the chemical composition of different types of samples (De Marchi et al., 2014). Because of the quantitative aspects of infrared spectroscopy (i.e., measuring the electromagnetic radiation as a means of obtaining information), the word spectrometry is also commonly used (Gengler et al., 2016). The resulting absorption spectrum is commonly expressed in % transmittance or absorbance (i.e., representing the amount of radiation absorbed by the sample) versus wavenumber (i.e., wavenumber (cm^{-1}) is the inversed function of wavelength, Khan et al., 2018). The transmittance (T) and absorbance (A) are calculated as follows:

$$T = I/I_0 \qquad A = \log(I_0/I) = -\log(T)$$

where I_0 is the intensity of light emitted by the source and I is the intensity of light after passing through the sample (Khan et al., 2018). Expressed in wavenumbers, the infrared region of the electromagnetic radiation ranges from approximately 40 cm^{-1} to 14,000 cm^{-1} and is divided into three main portions: near-infrared (4000 to 14000 cm^{-1}), mid-infrared (400 to 4000 cm^{-1}), and far-infrared (40 to 400 cm^{-1}; De Marchi et al., 2014).

Milk is usually analyzed in the mid-infrared region, and more rarely in the near-infrared region, because the mid-infrared region provides the clearest signal (i.e., best signal-to-noise ratio) and more relevant information (Grelet, 2019). Spectrometers currently used for milk analysis are Fourier-transform mid-infrared spectrometers, i.e., the most common type of mid-infrared spectrometer. Unlike dispersive instruments such as monochromators, Fourier-transform mid-infrared spectrometers collect all wavelengths simultaneously. They use different beams containing a combination of different wavelengths of electromagnetic radiation at once and measure the total beam intensity. Then, data are processed posteriorly using a technique called Fourier-transform to obtain the desired spectrum (Gengler et al., 2016; Khan et al., 2018). The main manufacturers of Fourier-transform mid-infrared spectrometers for milk analysis are Foss (Hillerød, Denmark), Perten Instruments (formerly Delta Instruments, Drachten, the Netherlands) and Bentley (Chaska, MN, USA). These three brands provide MIR spectra with different wavenumber ranges, from 925.66 to 5,010.15 cm^{-1} for Foss spectrometers, from 649.03 to 3,998.59 cm^{-1} for Bentley spectrometers and from 397.31 to 4,000 cm^{-1} for Perten spectrometers (Grelet, 2019).

The use of milk MIR spectrometry is common in dairy cattle for the analysis of milk of individual cows for milk recording, and also for the analysis of bulk tank milk for milk payment. Figure 1-2 illustrates the process of individual MIR samples analysis as part of milk recording, including a representation of the typical shape of the resulting MIR spectra. Milk samples are collected routinely on farms, usually on a monthly basis, for herds enrolled in milk recording (e.g., every 4 or 6 weeks in the Walloon region of Belgium). Samples are either from the morning milking, the evening milking (usually alternately between test days), or more commonly a mixture of both. At the same time, other useful information is collected for each cow, e.g., test date, animal identification, milk yield, breed, birth date, calving date, pedigree. The samples are then analyzed by a Fourier-transform spectrometer in a certified milk laboratory (ICAR, 2017; awé, 2019). The absorbance peaks of the resulting MIR spectra represent wavenumbers at which significant amounts of mid-infrared radiation were absorbed by the samples. The height or area of a peak is proportional to the concentration of the chemical bond they represent in milk. Hence, after spectral data processing, the milk MIR spectra are used both for a qualitative and quantitative purpose as they indicate the presence of a substance in milk and also the quantity of this substance in milk. Absorbance values at different wavenumbers can be linked to reference values in order to develop calibration equations allowing the prediction of milk traits of interest (Gengler et al., 2016). Spectra and predicted traits are stored in databases and can subsequently be reported to farmers (ICAR, 2017; awé, 2019).

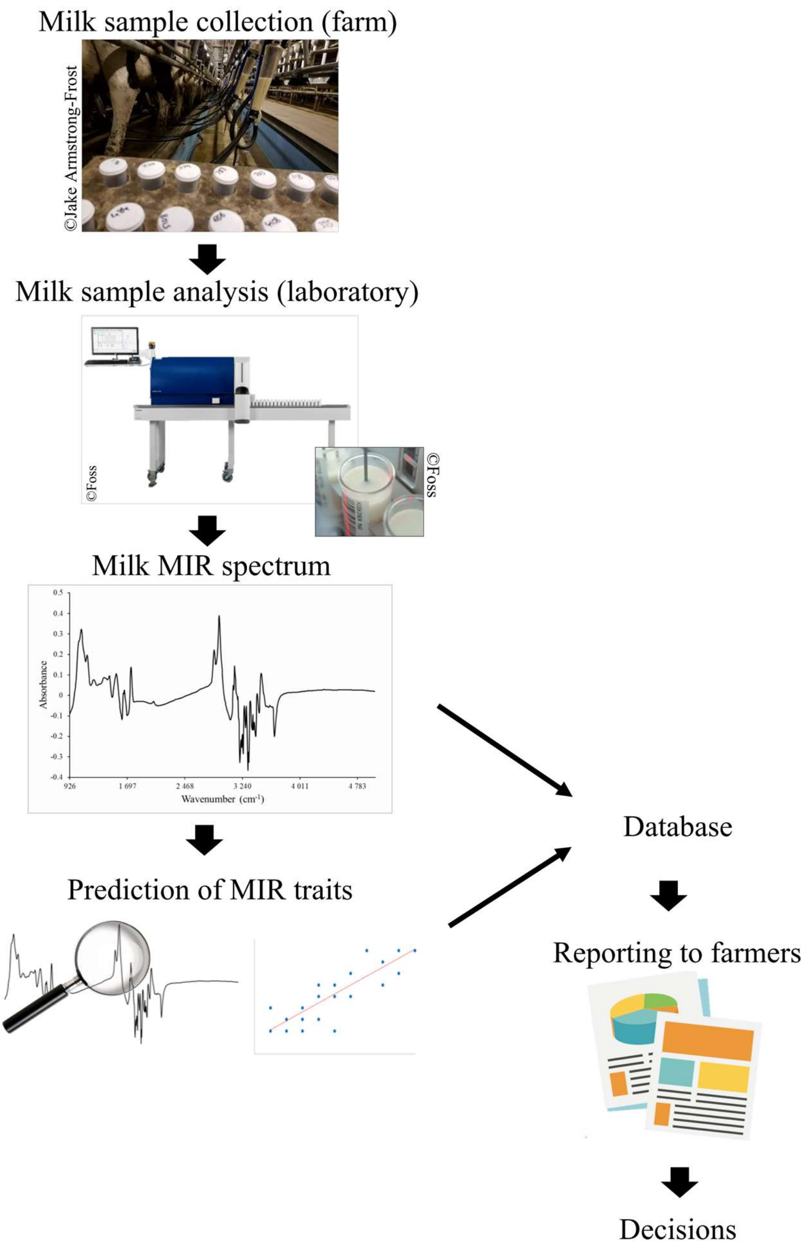

Figure 1-2. Pictorial summary of milk MIR analysis as part of milk recording.

4. Milk MIR for decision support

Some major advantages of milk MIR for the development of decision-support tools for dairy farming are the easy collection of milk samples on farms on a large scale, the low cost and rapidity of MIR analyses in the laboratory, the comparability of MIR data within and between farms, and the ability to develop customized tools at the herd and at the animal levels. The advantages of milk MIR will be explained and discussed in more detail in section 1 of Chapter 5, including examples from the research conducted in this thesis.

MIR-based tools directly or indirectly target different aspects of the dairy farming system described in Figure 1-1, such as feed, health and vet care, product processing, or the environmental impact.

Prediction equations are the most commonly developed tools with MIR and can support decision-making at the herd level through bulk tank milk analysis and most commonly at the individual level through milk recording. Figure 1-3 explains the usual steps for the development of MIR prediction equations. Some of these stages are part of the traditional process to develop a decision-support tool. The first usual step is the selection, manipulation, and cleaning of the data. The next stage is the development of the model itself (e.g., calibration of the equation), followed by validation (e.g., application of the equation to an independent dataset) to test the performances of the model on real-world data before the implementation. At the implementation stage, it is important to inform the users about the objectives, benefits, conditions of utilization, and limits of the tool.

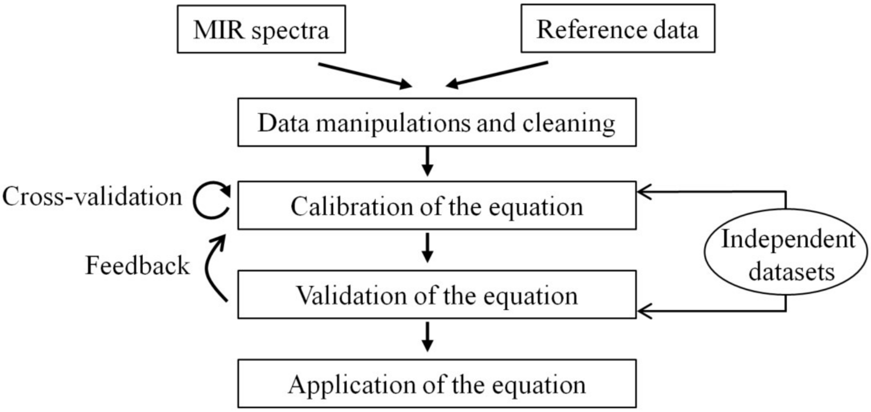

Figure 1-3. Usual steps in the development of a MIR prediction equation.

Traditional milk components predicted using MIR calibration equations (e.g., milk fat, protein, lactose, and urea contents) are the most common traits communicated to farmers and used for decision-making. For example, the evolution of fat and protein contents over the lactation and the year is used to detect abnormalities or to make decisions regarding feeding, culling, or breeding. The ratio of fat to protein content in milk is used to detect unbalanced diet: a ratio > 1.5 may indicate that the diet is deficient in energy (i.e., risk of ketosis) and a ratio < 1.1 may indicate that the diet is too rich in energy and not fibrous enough (i.e., risk of acidosis; Koeck et al., 2014; awé, 2020). Monitoring urea content in milk is useful for feeding management and to estimate feed nitrogen efficiency, which is also related to the environmental impact of excessive nitrogen excretion (Bastin et al., 2009).

Besides the conventional milk components such as fat, protein, lactose and urea contents, milk composition is much more complex. Following advances in informatics and exploration of MIR spectra (e.g., using multivariate statistics), the number of studies exploring the use of MIR to predict other informative traits has strongly increased since the mid-2000s (De Marchi et al., 2014). Models have been recently built to attempt to predict fine milk composition, such as fatty acid composition (Soyeurt et al., 2006; De Marchi et al., 2011; Maurice-Van Eijndhoven et al., 2013), protein composition (Bonfatti et al., 2011; Rutten et al., 2011), minerals (Soyeurt et al., 2009) or lactoferrin (Soyeurt et al., 2007). Other research explored MIR to predict milk technological properties, such as milk coagulation and cheese yield (Ferragina et al., 2013; Visentin et al., 2015) or milk acidity (De Marchi et al., 2009). Recent studies also considered MIR to predict novel traits mainly related to the cow physiology, for instance methane (CH_4) emissions (Vanlierde et al., 2018), nitrogen use efficiency (Grelet et al., 2020), body energy status (McParland et al., 2011), body weight (Soyeurt et al., 2019), fertility (Ho et al., 2019), energy intake and efficiency (McParland et al., 2014) or milk components related to physiological pathologies like ketosis (Van der Drift et al., 2012; Grelet et al., 2016). In addition, MIR has been investigated to develop other novel models, for example to have information about the geographical origin of milk (Scampicchio et al., 2016; Caredda et al., 2017) or to detect milk adulteration (Balabin and Smirnov, 2011). Most of these MIR prediction equations have been developed in the framework of research projects. Only few of them have been implemented in practice in some countries and reported to farmers, or are in the process of being deployed, to support management decision-making. For instance, in the Walloon region of Belgium, detection of ketosis using several MIR traits (e.g., β-hydroxybutyrate (BHB), fatty acids, acetone) and milk technological properties (e.g., for farms wishing to diversify production and process milk on the farm) are under study to be included in the milk recording results reported to farmers (C. Bastin, 2020, personal communication). To take another example, in Canada, the Ketolab service provides BHB predictions for ketosis detection, and fatty acids predictions on bulk tank milk have recently been available for farmers for feed management (Valacta, 2011; Santschi et al., 2019). Many other novel traits developed by scientific institutions are promising tools to guide various decisions and to optimize herd management, but they need to be elaborated further to be used as decision-support tools in practice.

In addition to management decision support based on predicted MIR phenotypes, conventional and novel MIR traits are of great interest for genetic and genomic evaluations because of the large amounts of phenotypes available through the use of

MIR (Gengler et al., 2016). Genetic evaluations are important for breeding decisions and some other management decisions (e.g., culling). Also, several studies showed the interest of modeling MIR traits predicted from spectra collected routinely using test-day models (TDM), for instance to predict future values, detect problems by comparing observed and predicted (i.e., expected) values or by evaluating the deviation of herd effects (e.g., Koivula et al., 2007; Bastin et al., 2009; Gillon et al., 2010). As a practical example, in Finland, monthly herd-management effect solutions from TDM of milk, fat and protein yields, and SCC are provided for authorized users such as farmers and dairy advisors to recognize management problems (Koivula et al., 2007).

5. Research objectives and outline

This introductory chapter has shown that dairy farming is facing multiple challenges compromising the sustainability of the sector. Dairy farmers consider the economic viability as the most important pillar of sustainability, while society is increasingly concerned by the environmental aspects (Vandreck, 2018). In this context, milk MIR would be a promising method to guide decisions leading to potential improvement of economic and environmental performances of dairy herds. However, the full utility of MIR is still partly unexplored. Therefore, the aim of this thesis was to contribute to the development of novel decision-support tools with economic and environmental interests for the dairy sector using milk MIR spectrometry. This manuscript consists of a compilation of three published scientific papers, covering different approaches of using MIR for decision support and addressing gaps in the current diversity of MIR tools. An overview of the next chapters is given in Figure 1-4.

The first paper (i.e., Chapter 2) focuses on the use of MIR to diagnose the pregnancy status of dairy cows. It illustrates the development of a calibration equation using milk MIR spectra and considers an economically important trait that has not been successfully predicted with MIR yet. This tool would be dedicated to dairy farmers directly or to dairy advisors who could transfer the outputs to farmers and advise them regarding, for example, mating decisions.

The second paper (i.e., Chapter 3) investigates the use of a test-day mixed model to predict milk MIR spectra with the perspective of herd management. As mentioned earlier in the introduction, there is an interest in modeling MIR traits using test-day models. Depending on the trait, this could contribute to optimizing economic and/or environmental performances of dairy herds. However, given the large number of MIR trait prediction equations, developing separate models for each trait would be time- and resource-consuming. Therefore, modeling directly the whole MIR spectrum and subsequently applying existing MIR calibration equations to the modeled spectra could be a better option. This tool would rather be dedicated to dairy advisors who could interpret the outputs and subsequently give advice to farmers, but also to researchers.

The third paper (i.e., Chapter 4) illustrates the implementation of a MIR tool and the interest of combining MIR-predicted data with other sources of data to provide more comprehensive information for decision-making. More specifically, we studied the relationships between milk MIR-predicted enteric CH_4 production and the technical and financial performances of commercial dairy herds. MIR equations to predict enteric CH_4 production have recently been developed (e.g., Vanlierde et al., 2016). A significant part of greenhouse gases emitted on a dairy farm comes from enteric CH_4 (e.g., 45% in Belgium), meaning it is an important factor in the carbon footprint of dairy products and a key trait to improve the environmental performances of dairy farms (FPS Public Health, 2018). This study was originally dedicated for research purposes to better understand management practices associated with CH_4 production, with potential implications for decisions at the political level (e.g., regarding climate policies) and subsequently at the farm level.

Chapter 5 discusses the main strengths as well as some issues and considerations regarding the use of milk MIR for the development of decision-support tools, mainly based on the learning acquired through the three research studies developed in this thesis. Lastly, a general conclusion and future prospects are addressed.

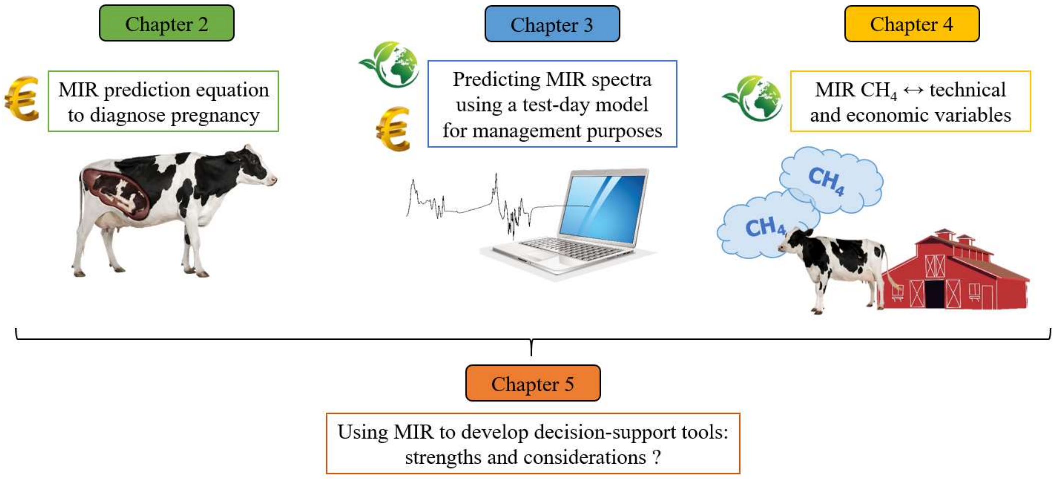

Figure 1-4. Overview of Chapters 2-5 of the thesis.

6. References

Allen, M.R., O.P. Dube, W. Solecki, F. Aragón-Durand, W. Cramer, S. Humphreys, M. Kainuma, J. Kala, N. Mahowald, Y. Mulugetta, R. Perez, M. Wairiu, and K.Z. 2018. Framing and Context. In: Global warming of 1.5°C. An IPCC special report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty.

awé. 2019. Règlement général du contrôle laitier. www.awenet.be/awe/UserFiles/file/asbl/document/Reglement-controle-laitier-2019.pdf

awé. 2020. MyAwenet. www.awenet.be/awe/commun/asbl/gestion/myawenet.php.

Balabin, R.M., and S. V. Smirnov. 2011. Melamine detection by mid- and near-infrared (MIR/NIR) spectroscopy: a quick and sensitive method for dairy products analysis including liquid milk, infant formula, and milk powder. Talanta 85:562–568. https://doi.org/10.1016/j.talanta.2011.04.026.

Barkema, H.W., M.A.G. von Keyserlingk, J.P. Kastelic, T.J.G.M. Lam, C. Luby, J.P. Roy, S.J. LeBlanc, G.P. Keefe, and D.F. Kelton. 2015. Invited review: changes in the dairy industry affecting dairy cattle health and welfare. J. Dairy Sci. 98:7426–7445. https://doi.org/10.3168/jds.2015-9377.

Bastin, C., L. Laloux, A. Gillon, F. Miglior, H. Soyeurt, H. Hammami, C. Bertozzi, and N. Gengler. 2009. Modeling milk urea of Walloon dairy cows in management perspectives. J. Dairy Sci. 92:3529–3540. https://doi.org/10.3168/jds.2008-1904.

Bonfatti, V., G. Di Martino, and P. Carnier. 2011. Effectiveness of mid-infrared spectroscopy for the prediction of detailed protein composition and contents of protein genetic variants of individual milk of Simmental cows. J. Dairy Sci. 94:5776–5785. https://doi.org/10.3168/jds.2011-4401.

Caredda, M., M. Addis, I. Ibba, R. Leardi, M.F. Scintu, G. Piredda, and G. Sanna. 2017. Building of prediction models by using mid-infrared spectroscopy and fatty acid profile to discriminate the geographical origin of sheep milk. LWT - Food Sci. Technol. 75:131–136. https://doi.org/10.1016/j.lwt.2016.08.053.

De Graaf, S., E.J. Van Loo, J. Bijttebier, F. Vanhonacker, L. Lauwers, F.A.M. Tuyttens, and W. Verbeke. 2016. Determinants of consumer intention to purchase animal-friendly milk. J. Dairy Sci. 99:8304–8313. https://doi.org/10.3168/jds.2016-10886.

De Marchi, M., C.C. Fagan, C.P. O'Donnell, A. Cecchinato, R. Dal Zotto, M. Cassandro, M. Penasa, and G. Bittante. 2009. Prediction of coagulation properties, titratable acidity, and pH of bovine milk using mid-infrared spectroscopy. J. Dairy Sci. 92:423–432. https://doi.org/10.3168/jds.2008-1163.

De Marchi, M., M. Penasa, A. Cecchinato, M. Mele, P. Secchiari, and G. Bittante. 2011. Effectiveness of mid-infrared spectroscopy to predict fatty acid composition of Brown Swiss bovine milk. Animal 5:1653–1658. https://doi/org.10.1017/S1751731111000747.

De Marchi, M., V. Toffanin, M. Cassandro, and M. Penasa. 2014. Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits. J. Dairy Sci. 97:1171–1186. https://doi.org/10.3168/jds.2013-6799.

Directorate-General for Agriculture and Rural Development (DG AGRI). 2017. Modernising and simplifying the CAP. Background document. Socio-economic challenges facing EU agriculture.

Eastwood, C.R., D.F. Chapman, and M.S. Paine. 2012. Networks of practice for co-construction of agricultural decision support systems: case studies of precision dairy farms in Australia. Agric. Syst. 108:10–18. https://doi.org/10.1016/j.agsy.2011.12.005.

European Commission. 2019a. Milk and dairy products. Accessed Jan. 17, 2020. https://ec.europa.eu/info/food-farming-fisheries/animals-and-animal-products/animal-products/milk-and-milk-products_en.

European Commission. 2019b. Le nouvel accord UE-Mercosur. L'accord de principe (28 juin 2019).

Eurostat. 2019. Agriculture, forestry and fishery statistics. Statistical books.

Ferragina, A., C. Cipolat-Gotet, A. Cecchinato, and G. Bittante. 2013. The use of Fourier-transform infrared spectroscopy to predict cheese yield and nutrient recovery or whey loss traits from unprocessed bovine milk samples. J. Dairy Sci. 96:7980–7990. https://doi.org/10.3168/jds.2013-7036.

Federal Public Service (FPS) Public Health. 2018. Belgian National Debate on Carbon Pricing: Final Report.

Gengler, N., H. Soyeurt, F. Dehareng, C. Bastin, F. Colinet, H. Hammami, M.L. Vanrobays, A. Lainé, S. Vanderick, C. Grelet, A. Vanlierde, E. Froidmont, and P. Dardenne. 2016. Capitalizing on fine milk composition for breeding and management of dairy cows. J. Dairy Sci. 99:4071–4079. https://doi.org/10.3168/jds.2015-10140.

Gillon, A., S. Abras, P. Mayeres, C. Bertozzi, and N. Gengler. 2010. Adding value to test-day data by using modified best prediction method. Proc. ICAR 37th Annu. Meet. 1–8.

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

Grelet, C. 2019. Promoting international prediction models through standardization of milk mid-infrared spectra. Doctoral thesis. Gembloux Agro-Bio Tech, Université de Liège, Belgium.

Grelet, C., E. Froidmont, L. Foldager, M. Salavati, M. Hostens, C.P. Ferris, M.A. Crowe, M.T. Sorensen, J.A.F. Pierna, A. Vanlierde, N. Gengler, G. Consortium, F. Dehareng, M. Salavati, Z. Cheng, A. Fouladi, G. Pollott, M. Bell, M. Vaneetvelde, K. Hermans, and M. Hostens. 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.

Heinemann, P.H. 2009. Decision support systems for food and agriculture. Pages 122-135 in Systems analysis and modeling in food and agriculture. K.C. Ting, D.H. Fleisher, L.F. Rodriguez, ed. EOLSS Publishers, Oxford, UK.

Ho, P.N., V. Bonfatti, T.D.W. Luke, and J.E. Pryce. 2019. Classifying the fertility of dairy cows using milk mid-infrared spectroscopy. J. Dairy Sci. 102:10460–10470. https://doi.org/10.3168/jds.2019-16412.

Hostiou, N., J. Fagon, S. Chauvat, A. Turlot, F. Kling-Eveillard, X. Boivin, and C. Allain. 2017. Impact of precision livestock farming on work and human-animal interactions on dairy farms. A review. Biotechnol. Agron. Soc. Environ. 21:268–275. https://doi.org/ 10.25518/1780-4507.13706.

International Committee for Animal Recording (ICAR). 2017. Section 2 - Guidelines for dairy cattle milk recording. Int. Comm. Anim. Record., Rome, Italy.

Jago, J., C. Eastwood, K. Kerrisk, and I. Yule. 2013. Precision dairy farming in Australasia: adoption, risks and opportunities. Anim. Prod. Sci. 53:907–916. https://doi.org/10.1071/AN12330.

Khan, S.A., S.B. Khan, L.U. Khan, A. Farooq, K. Akhtar, and A.M. Asiri. 2018. Fourier transform infrared spectroscopy: fundamentals and application in functional groups and nanomaterials characterization. Pages 317-344 in Handbook of materials characterization. S.K. Sharma, ed. Springer International, Switzerland.

Koeck, A., J. Jamrozik, F.S. Schenkel, R.K. Moore, D.M. Lefebvre, D.F. Kelton, and F. Miglior. 2014. Genetic analysis of milk β-hydroxybutyrate and its association with fat-to-protein ratio, body condition score, clinical ketosis, and displaced abomasum in early first lactation of Canadian Holsteins. J. Dairy Sci. 97:7286–7292. https://doi.org/10.3168/jds.2014-8405.

Koivula, M., J.I. Nousiainen, J. Nousiainen, and E.A. Mäntysaari. 2007. Use of herd solutions from a random regression test-day model for diagnostic dairy herd management. J. Dairy Sci. 90:2563–2568. https://doi.org/10.3168/jds.2006-517.

La Spina, S. Quel etat de santé pour notre agriculture wallonne ? Echangeons sur notre agriculture ! Accessed Jan. 17, 2020. https://agriculture-natpro.be/2014/04/01/quel-etat-de-sante-de-notre-agricult.

Maurice-Van Eijndhoven, M.H.T., H. Soyeurt, F. Dehareng, and M.P.L. Calus. 2013. Validation of fatty acid predictions in milk using mid-infrared spectrometry across cattle breeds. Animal 7:348–354. https://doi.org/10.1017/S1751731112001218.

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

McParland, S., E. Lewis, E. Kennedy, S.G. Moore, B. McCarthy, M. O'Donovan, S.T. Butler, J.E. Pryce, and D.P. Berry. 2014. Mid-infrared spectrometry of milk as a predictor of energy intake and efficiency in lactating dairy cows. J. Dairy Sci. 97:5863–5871. https://doi.org/10.3168/jds.2014-8214.

Meynard, J.-M., M. Cerf, L. Guichard, M.-H. Jeuffroy, and D. Makowski. 2002. Which decision support tools for the environmental management of nitrogen? Agron. EDP Sci. 23:407–418. https://doi.org/10.1051/agro:2002064.

Moureaux, B. Les OAD vont-ils révolutionner le métier d'agriculteurs ? Accessed Jan. 17, 2020. https://www.arvalis-infos.fr/les-outils-d-aide-a-la-decision-modifient-ils-le-metier-d-agriculteur--@/view-23316-arvarticle.html.

Norton, T., and D. Berckmans. 2017. Developing precision livestock farming tools for precision dairy farming. Anim. Front. 7:18–23. https://doi.org/10.2527/af.2017.0104.

Raiten, D.J., and G.F. Combs. 2019. Nutritional Ecology: understanding the intersection of climate/environmental change, food systems and health. Pages 68-80 in: Agriculture for Improved Nutrition: Seizing the Momentum. S. Fan, S. Yosef and R. Pandya-Lorch, ed. CABI International, Boston, USA.

Rose, D.C., W.J. Sutherland, C. Parker, M. Lobley, M. Winter, C. Morris, S. Twining, C. Ffoulkes, T. Amano, and L. V. Dicks. 2016. Decision support tools for agriculture: Towards effective design and delivery. Agric. Syst. 149:165–174. https://doi.org/10.1016/j.agsy.2016.09.009.

Rutten, M.J.M., H. Bovenhuis, J.M.L. Heck, and J.A.M. van Arendonk. 2011. Predicting bovine milk protein composition based on Fourier transform infrared spectra. J. Dairy Sci. 94:5683–5690. https://doi.org/10.3168/jds.2011-4520.

Santschi, D.E., D. Warner, and D.M. Lefebvre. 2019. Implementation of a routine FT - IR procedure for fatty acid analysis in bulk tank milk Milk fatty acids : wy do we want them ? ICAR annual conference, Pragues, Czech Republic.

Scampicchio, M., D. Eisenstecken, L. De Benedictis, C. Capici, D. Ballabio, T. Mimmo, P. Robatscher, L. Kerschbaumer, M. Oberhuber, A. Kaser, C.W. Huck, and S. Cesco. 2016. Multi-method approach to trace the geographical origin of alpine milk: a case study of tyrol region. Food Anal. Methods 9:1262–1273. https://doi.org/10.1007/s12161-015-0308-2.

Seneviratne, S., N. Nicholls, D. Easterling, C. Goodess, S. Kanae, J. Kossin, Y. Luo, J. Marengo, K. McInnes, M. Rahimi, M. Reichstein, A. Sorteberg, C. Vera, and X. Zhang. 2012. Changes in climate extremes and their impacts on the natural physical environment. Pages 109-230 Managing the risk of extreme events and disasters to advance climate change adaptation. A special report of working groups I and II of the Intergovernmental Panel on Climate Change (IPCC). ed. Cambridge University Press, Cambridge, UK, and New York, NY, USA.

Subramanian, A., and L. Rodriguez-Saona. 2009. Fourier transform infrared (FTIR) spectroscopy. Pages 145-178 in Infrared spectroscopy for food quality analysis and control. D.W Sun, ed. Academic Press, New York, NY, USA.

Soyeurt, H., P. Dardenne, F. Dehareng, G. Lognay, D. Veselko, M. Marlier, C. Bertozzi, P. Mayeres, and N. Gengler. 2006. Estimating fatty acid content in cow milk using mid-infrared spectrometry. J. Dairy Sci. 89:3690–3695. https://doi.org/10.3168/jds.s0022-0302(06)72409-2.

Soyeurt, H., F.G. Colinet, V.M.R. Arnould, P. Dardenne, C. Bertozzi, R. Renaville, D. Portetelle, and N. Gengler. 2007. Genetic variability of lactoferrin content estimated by mid-infrared spectrometry in bovine milk. J. Dairy Sci. 90:4443–4450. https://doi.org/10.3168/jds.2006-827.

Soyeurt, H., D. Bruwier, J.-M. Romnee, N. Gengler, C. Bertozzi, D. Veselko, and P. Dardenne. 2009. Potential estimation of major mineral contents in cow milk using mid-infrared spectrometry. J. Dairy Sci. 92:2444–2454. https://doi.org/10.3168/jds.2008-1734

Soyeurt, H., E. Froidmont, I. Dufrasne, D. Hailemariam, Z. Wang, C. Bertozzi, F.G. Colinet, F. Dehareng, and N. Gengler. 2019. Contribution of milk mid-infrared spectrum to improve the accuracy of test-day body weight predicted from stage, lactation number, month of test and milk yield. Livest. Sci. 227:82–89. https://doi.org/10.1016/j.livsci.2019.07.007.

Turlot, A. Organisation du travail dans des exploitations wallonnes. Accessed Jan.17, 2020. http://www.cra.wallonie.be/fr/organisation-du-travail-dans-des-exploitations-wallonnes.

Valacta, 2011. Cétolab. https://www.valacta.com/_media/document/858/v2k-22f1-rpt50600f-cetolab-doc-explic-uv-qc-2011-05-6p-pays-port-3.pdf.

Van der Drift, S.G.A., R. Jorritsma, J.T. Schonewille, H.M. Knijn, and J.A. Stegeman. 2012. Routine detection of hyperketonemia in dairy cows using Fourier transform infrared spectroscopy analysis of β-hydroxybutyrate and acetone in milk in combination with test-day information. J. Dairy Sci. 95:4886–4898. https://doi.org/10.3168/jds.2011-4417.

Vandreck, S. 2018. Si les exploitations ne misent pas sur la durabilité, elles sont vouées à disparaître. Accessed Jan. 17, 2020. https://www.rtbf.be/info/regions/detail_si-les-exploitations-ne-misent-pas-sur-la-durabilite-elles-sont-vouees-a-disparaitre?id=9979305.

Vanlierde, A., H. Soyeurt, N. Gengler, F.G. Colinet, E. Froidmont, M. Kreuzer, F. Grandl, M. Bell, P. Lund, D.W. Olijhoek, M. Eugène, C. Martin, B. Kuhla, and F. Dehareng. 2018. Short communication: Development of an equation for estimating methane emissions of dairy cows from milk Fourier transform mid-infrared spectra by using reference data obtained exclusively from respiration chambers. J. Dairy Sci. 101:7618–7624. https://doi.org/10.3168/jds.2018-14472.

Vanlierde, A., M.L. Vanrobays, N. Gengler, P. Dardenne, E. Froidmont, H. Soyeurt, S. McParland, E. Lewis, M.H. Deighton, M. Mathot, and F. Dehareng. 2016. Milk mid-infrared spectra enable prediction of lactation-stage-dependent methane emissions of dairy cattle within routine population-scale milk recording schemes. Anim. Prod. Sci. 56:258–264. https://doi.org/10.1071/AN15590.

Visentin, G., A. McDermott, S. McParland, D.P. Berry, O.A. Kenny, A. Brodkorb, M.A. Fenelon, and M. De Marchi. 2015. Prediction of bovine milk technological traits from mid-infrared spectroscopy analysis in dairy cows. J. Dairy Sci. 98:6620–6629. https://doi.org/10.3168/jds.2015-9323.

Wezel, A., M. Casagrande, F. Celette, J.F. Vian, A. Ferrer, and J. Peigné. 2014. Agroecological practices for sustainable agriculture. A review. Agron. Sustain. Dev. 34:1–20. https://doi.org/10.1007/s13593-013-0180-7.

2

Diagnosing the pregnancy status of dairy cows: how useful is milk mid-infrared spectrometry?

Adapted from: Delhez, P., P.N. Ho, N. Gengler, H. Soyeurt, and J.E. Pryce. 2020. Diagnosing the pregnancy status of dairy cows: How useful is milk mid-infrared spectroscopy? J. Dairy Sci. 103: 3264-3274.

https://doi.org/10.3168/jds.2019-17473

Outline

Pregnancy is a meaningful trait for dairy farms' sustainability because of its economic importance. Being able to detect early pregnancy is particularly useful to know if a cow should be re-bred or to forecast calving dates. Previous research achieved limited results in predicting the pregnancy status of dairy cows using MIR and suggested that further research would be needed on this topic. The interest in predicting pregnancy using MIR originates from the fact that pregnancy establishment would result in changing milk composition. Therefore, this chapter investigated innovative strategies to attempt to find a successful approach to diagnose pregnancy using a MIR calibration equation. This study was conducted using milk MIR spectra and records of insemination from Australian dairy cows. The algorithm used to calibrate the equations developed in the article mentioned in this chapter was partial least squares discriminant analysis (PLS-DA). However, alternative machine learning algorithms were also tested and are presented in section 2.3.1 of Chapter 5.

Abstract

Pregnancy diagnosis is an essential part of successful breeding programs on dairy farms. Milk composition alters with pregnancy and this is well documented. Fourier-transform mid-infrared (MIR) spectrometry is a rapid and cost-effective method for providing milk spectra that reflect the detailed composition of milk samples. Therefore, the aim of this study was to assess the ability of MIR spectrometry to predict the pregnancy status of dairy cows. The MIR spectra and insemination records were available from 8,064 Holstein cows of 19 commercial dairy farms in Australia. Three strategies were studied to classify cows as open or pregnant using partial least squares discriminant analysis models with random cow-independent 10-fold cross-validation and external validation on a cow-independent test set. The first strategy considered 6,754 MIR spectra after insemination used as independent variables in the model. The results showed little ability to detect the pregnancy status as the area under the receiver operating characteristic curve was 0.63 and 0.65 for cross-validation and testing, respectively. The second strategy, involving 1,664 records, aimed to reduce noise in the MIR spectra used as predictors by subtracting a spectrum before insemination (i.e., open spectrum) from the spectrum after insemination. The accuracy was comparable with the first approach, showing no superiority of the method. Given the limited results for these models when using combined data from all stages after insemination, the third strategy explored separate models at seven stages after insemination comprising 348 to 1,566 records each (i.e., progressively greater gestation) with single MIR spectra after insemination as predictors. The models developed using data recorded after 150 d of pregnancy showed promising prediction accuracy with the average value of area under the receiver operating characteristic curve of 0.78 and 0.76 obtained through cross-validation and testing, respectively. If this can be confirmed on a larger dataset and extended to somewhat earlier stages after insemination, the model could be used as a complementary tool to detect fetal abortion.

Key words: gestation, prediction accuracy, milk composition, discriminant analysis

1. Introduction

An accurate and timely detection of pregnancy is fundamental to successful breeding programs in dairy cows because it enables non-pregnant cows to be re-bred as soon as possible (Velek et al., 2012). Pregnancy diagnosis methods that are currently available include the observation of non-return to estrus, transrectal palpation, transrectal or transcutaneous ultrasonography, and analysis of progesterone and pregnancy-associated glycoproteins in milk or blood (Fricke et al., 2016). However, these methods have a certain cost and efficacy and some require animal handling, which might limit their practical implementation.

Fourier-transform mid-infrared (MIR) spectrometry is already routinely used in the dairy industry worldwide to analyze major milk components (e.g., fat, protein, lactose contents, and urea) for milk payment, herd management, quality control, or genetic evaluation programs (ICAR, 2017a). Additionally, MIR can be used to predict other phenotypes associated with milk composition in dairy cows with reasonable accuracy, such as fatty acids (Soyeurt et al., 2006), ketone bodies (Grelet et al., 2016), methane emissions (Vanlierde et al., 2018), or energy intake and feed efficiency (McParland and Berry, 2016). Because the establishment of pregnancy affects milk composition, through altering nutrient partitioning between physiological functions, it might be hypothesized that MIR would be used to detect the pregnancy of a dairy cow. Indeed, several authors have indicated, for instance, increasing milk fat and protein contents of pregnant compared with non-pregnant cows, especially in the last months of gestation (Olori et al., 1997; Penasa et al., 2016; Lainé et al., 2017).

Toledo-Alvarado et al. (2018) investigated the possibility to predict the pregnancy status of dairy cows from multiple breeds using the whole raw milk MIR spectrum alone or in combination with other effects including DIM; parity; and concentration of fat, protein, and lactose in milk. They found promising, but limited, prediction accuracy for classification of open and pregnant cows [i.e., around 0.60 for the area under the receiver operating characteristic curve (AUC)]. The imperfect prediction accuracy was attributed to the complicated nature of pregnancy status and its indirect association with milk composition. Lainé et al. (2014) used residual MIR spectra which were obtained after correcting for several effects associated with open cows (i.e., fixed effects of parity, breed, month of test-day, DIM and random effects of cows) to discriminate between pregnant and open cows. They initially reported very good prediction accuracy for classification of open and pregnant cows (i.e., specificity and sensitivity higher than 90%), which they achieved through random cross-validation. Unfortunately, the prediction accuracy did not hold in external validation, with a drop up to 50% reported in a later study (A. Lainé, 2019, University of Liège, Gembloux, Belgium, personal communication). The explanation was that the way they removed fixed effects to obtain the residual MIR spectrum somehow artificially created dependencies to the dataset used, but this cannot be done with future data where pregnancy status is not known a priori. These inconsistent results facilitate a consideration of new approaches to investigate the association between MIR and pregnancy, as also indicated by Toledo-Alvarado et al. (2018).

This study aimed to investigate the potential of milk MIR to predict the pregnancy status of dairy cows while considering some new aspects that are unexplored in the existing literature on this subject. The first innovation of this work lies in the use of novel residual MIR spectrum, which is the difference between the spectrum after an insemination and a spectrum before this insemination at a specific stage during the same lactation. By doing this, we expected to simplify the MIR signal after insemination by removing effects specific to each cow when being open while preserving the potential pregnancy signal. The second novelty is the exploration of predictions at different stages after insemination, because the stage of gestation has been reported to influence milk composition. The third novelty is the use of cow-independent validation because the commonly applied method of random cross-validation has recently been reported to produce overoptimistic results (Wang and Bovenguis, 2019).

2. Materials and methods

2.1. *Data*

The data used in this study, including milk MIR spectra and records of insemination from 8,064 Holstein cows, were obtained from 19 commercial dairy herds located in Victoria, Tasmania, and New South Wales of Australia in 2016, 2017 and 2018. Milk samples were collected 1 to 8 times per cow and sent to TasHerd Pty. Ltd. (Hadspen, Tasmania, Australia) for analysis by an infrared spectrometer (Bentley Instruments NexGen Series FTS Combi machine, Chaska, MN) to obtain the MIR spectrum from which fat, protein, and lactose contents were estimated using manufacturer prediction equations. The obtained MIR spectrum was expressed in absorbance, with 899 wavenumbers covering the absorption of light in the infrared region located from 649 to 3,999 cm^{-1}.

The pregnancy status, defined as pregnant or open, was determined for each test date using records of insemination and the corresponding calving date. Where an insemination did not result in a pregnancy, MIR records occurring before and after that insemination were set to open status. Where an insemination resulted in a pregnancy, MIR records occurring before and after that insemination were set to open and pregnant statuses, respectively. Pregnancy was confirmed by the following actual calving record (Figure 2-1a). Cows with no calving records were also retained and considered as open cows.

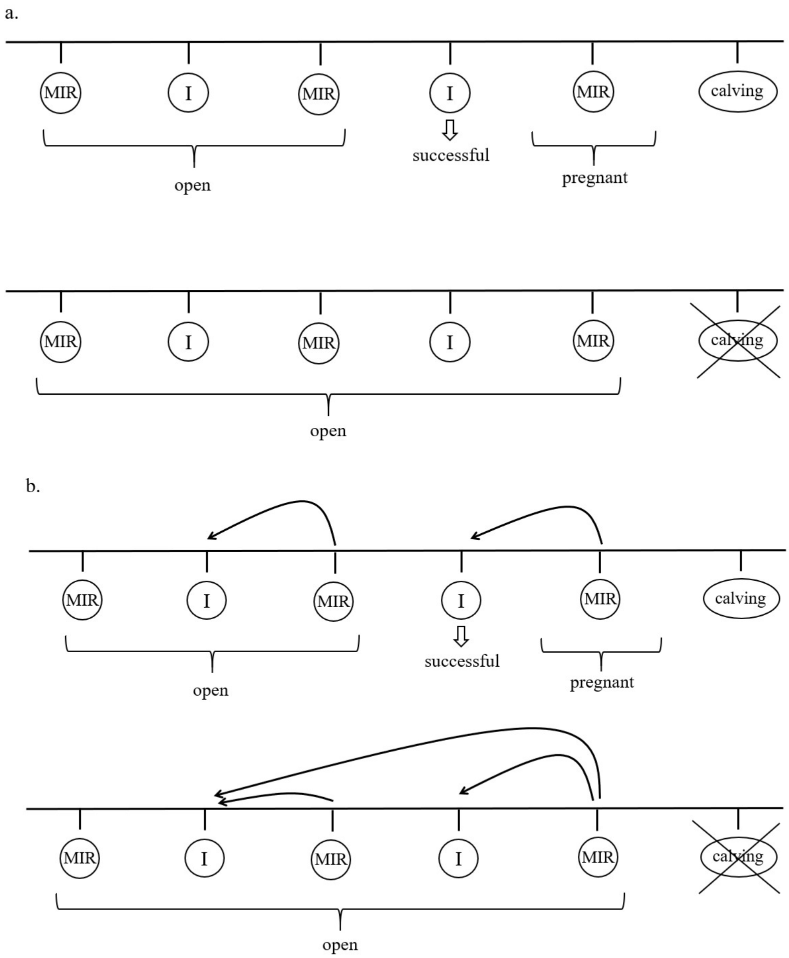

Figure 2-1. (a) Timeline diagram explaining how the pregnancy status (i.e., open or pregnant) was defined for milk mid-infrared spectrum (MIR) records. (b) Timeline diagram explaining the selection of spectra after insemination. The MIR records with open status can be associated with several inseminations. The MIR records with pregnant status are only associated with the successful insemination. I = insemination.

2.2. Data preprocessing

The first derivative of raw MIR spectra was calculated for each wavenumber X as the difference between the wavenumber X-2 and the wavenumber X+2 to remove the baseline variation (Soyeurt et al., 2011). On the reduced spectra, noisy regions with low signal-to-noise ratio induced by water absorption were first removed, leaving 538 spectral points out of 899 for the study (Bonfatti et al., 2011; Grelet et al., 2016). These 538 wavenumbers were in the regions from 928 to 1,596 cm^{-1} and from 1,693 to 3,025 cm^{-1}. Subsequently, potential spectral outliers were excluded by calculating the standardized Mahalanobis distance or global distance, and records with global H distance > 3 were eliminated (De Maesschalck et al., 2000; Grelet et al., 2017).

The data were also checked for unusual milk fat and protein contents (i.e., 1.5 g/dL of milk < fat < 9 g/dL of milk, 1 g/dL of milk < protein < 7 g/dL of milk) and SCC (0.01% upper values were deleted) following the recommendation of the International Committee for Animal Recording (ICAR, 2017b). Finally, gestation length was restricted to be from 250 to 300 d, because the average for Australian Holsteins is 281 d (Haile-Mariam and Pryce, 2019).

2.3. Modeling strategies

2.3.1. Single spectra after insemination

In this strategy, only spectral records after insemination were considered, which is similar to the study by Toledo-Alvarado et al. (2018). However, Toledo-Alvarado et al. (2018) did not consider cows with no calving records, in contrast to the present study. The selection of spectral records after insemination is described in Figure 2-1b. MIR records with open status, as previously described, could be associated with several inseminations, while those with pregnant status were only associated with the successful (i.e., last) insemination.

2.3.2. Spectral differences

The concept behind this strategy is to look at the change in a spectrum arising from pregnancy, which was done by subtracting a spectrum collected before insemination (i.e., open spectrum) from the one collected after insemination (i.e., open spectrum if the animal was not pregnant and pregnant spectrum if the animal was pregnant). We hypothesized that, by taking the difference and given the animal was pregnant, the resulting spectrum would be less noisy and retain among others the information potentially related to pregnancy establishment. This approach is similar to Lainé et al. (2014) except that we subtracted observed spectra and not modeled spectra. For each spectrum after insemination, we selected an open spectrum before insemination within a restricted window of 5 to 30 DIM, to have control spectra at a similar lactation stage and before the first insemination for all cows.

2.3.3. Stages after insemination

We hypothesized that the pregnancy signal in milk might vary by gestation stage, with the signal being stronger towards the end of gestation. Therefore, grouping records based on the period after insemination and modeling each group separately may improve the accuracy of the models by reducing the spectral and the pregnancy signal variabilities. Consequently, only single spectra after insemination were selected and divided into 7 classes based on the number of days after insemination (Table 2-1).

Table 2-1. Number of open and pregnant records and cows, and DIM (mean, with SD in parentheses) for the calibration and test sets for the 3 strategies

			Calibration set			Test set		
Strategy	Class[1]	Status	Records, no.	Cows, no.	DIM	Records, no.	Cows, no.	DIM
(1) Single spectra after insemination		open	3,377	1,342	208 (127)	863	495	229 (114)
		pregnant	3,377	1,698	239 (113)	863	658	246 (126)
(2) Spectral differences		open	832	456	161 (103)	176	114	169 (114)
		pregnant	832	485	166 (105)	176	127	174 (113)
(3) Stages after insemination	1 (1-30 d)	open	783	611	121 (70)	166	124	144 (97)
		pregnant	783	612	152 (102)	166	128	150 (97)
	2 (31-60 d)	open	688	570	164 (80)	161	116	174 (85)
		pregnant	688	638	201 (118)	161	147	190 (108)
	3 (61-90 d)	open	743	553	201 (92)	180	130	219 (90)
		pregnant	743	678	232 (113)	180	161	233 (111)
	4 (91-120 d)	open	481	314	263 (105)	139	80	257 (99)
		pregnant	481	400	298 (102)	139	118	292 (105)
	5 (121-150 d)	open	282	205	291 (94)	87	57	325 (102)
		pregnant	282	240	329 (100)	87	73	332 (95)
	6 (151-180 d)	open	174	112	314 (87)	62	34	295 (80)
		pregnant	174	155	341 (98)	62	51	352 (126)
	7 (>= 181 d)	open	226	122	361 (81)	68	37	355 (81)
		pregnant	226	157	360 (95)	68	48	349 (89)

[1] The number of days after insemination for each class of strategy 3 is indicated in parentheses.

2.4. Model development and evaluation of performance

For each strategy, approximately 80% of the data was randomly selected for calibration and the remaining 20% was selected for testing (the specific number of records can be seen in Table 2-1). All test sets were cow-independent, meaning that cows from each test set were different from those in the corresponding calibration set. All calibration and test sets were forced to have a balanced ratio of open and pregnant records by randomly sampling the majority class to be the same size as the minority class.

Partial least squares discriminant analysis was used to classify open and pregnant records (1 = pregnant, 0 = open), using the Caret package in R (version 3.6.0; Kuhn, 2019). Spectra after insemination were used as independent variables for the first and third strategies (i.e., single spectra after insemination and stages after insemination), whereas in the second strategy (i.e., spectral subtraction) spectral differences were used as independent variables. Partial least squares discriminant analysis was used because it performs dimensionality reduction while simultaneously carrying out classification (Rozenstein et al., 2015), which is usually preferred to conventional regression techniques, for example, logistic regression when predictor variables are highly correlated. As such, this method is appropriate due to the correlations that exist between wavenumbers.

The model performance was evaluated in two ways: random cow-independent 10-fold cross-validation and external validation. Random cow-independent 10-fold cross-validation means that 10% of the cows were randomly removed from the calibration set, and a model built with the remaining cows was used to classify the excluded records. The procedure was repeated 10 times to obtain predictions for all records. Hence, for each of the 10 folds, cows in the training set were different from those in the validation set. In addition to assessing the predictive performance of the model, cross-validation was also used to fine-tune the model (i.e., determine the most appropriate number of components of the partial least squares discriminant analysis models). However, the maximum number of components was fixed to 20 to avoid overfitting. External validation was carried out using the cow-independent testing data consisting of cows that were not used to build up and tune the model.

Model performance for calibration, cross-validation, and testing were assessed using the sensitivity, specificity, AUC, and median predicted class probability for correctly classified records. This median predicted probability is explained as follows: only records correctly classified were retained and, for each of them, the predicted probability amongst the 2 classes (i.e., open vs. pregnant) with a value greater than the 0.5 threshold was retained and the median of these predicted probabilities was calculated. Values that are close to 1 indicate high confidence in prediction results for correctly classified records, whereas values around 0.5 indicate low confidence. Sensitivity is defined as the proportion of records belonging to pregnant cows that were correctly identified as pregnant, and specificity is defined as the proportion of records belonging to open cows that were correctly identified as open. Sensitivity and specificity are calculated at a decision threshold probability of 0.5, and when this threshold is modified, the values of both measures are also modified. Consequently, the projection of these measures at different threshold probabilities defines a curve referred to as the receiver operating characteristic curve. The area under the receiver operating characteristic curve indicates how well the model can distinguish between the 2 categories. The AUC values are between 0 and 1, with 0.5 being a random guess (Šimundić, 2009).

For strategy 2, to have a fair comparison of model performance obtained using the spectral differences and single spectra after insemination, we ran the 2 models on the same dataset. Finally, to explore if the patterns of effects of pregnancy on MIR would change as gestation progresses, the coefficients of the 7 models in the third strategy (i.e., stages after insemination) were extracted and analyzed. Accordingly, the correlations between the models' coefficients are presented as a heatmap. If the correlation between the coefficients of 2 models is strong, similar wavenumbers might have been used for the prediction of the pregnancy status.

3. Results and discussion

A timely and cost-effective pregnancy diagnosis is desirable in the management of dairy farms, enabling decisions such as feed budgets, planning which cows to cull or rebreed, and when to dry off. MIR spectrometry allows low-cost, high throughput, and large-scale milk analysis and is already routinely used in many countries worldwide in milk recording programs. Hence, using MIR data for pregnancy diagnosis would be inexpensive and simple to conduct on farms. The expectation that MIR could be used to predict pregnancy status originates from the fact that a pregnancy establishment would result in changing milk composition (Olori et al., 1997; Roche, 2003; Penasa et al., 2016), which is commercially analyzed using MIR. This study shows that generally MIR can predict pregnancy status with promising accuracy in the late but not early stage of gestation.

3.1. Data description

Table 2-1 describes the datasets used for the 3 modeling strategies. Due to the repeated records per cow, some cows could have records that appear in both calibration and test sets, or in both training and validation sets for cross-validation, which is likely to lead to overoptimistic results (Shetty et al., 2017; Wang and Bovenhuis, 2019). Consequently, we decided to use random cow-independent cross-validation and cow-independent test sets (external validation; i.e., cows from each validation or test set are different from those in the corresponding training or calibration set). Such validation strategies were not implemented in the previous studies on diagnosing the pregnancy status of dairy cows (Lainé et al., 2014; Toledo-Alvarado et al., 2018).

Regarding the third strategy (i.e., our hypothesis that the pregnancy signal in milk might vary by gestation stage), the number of records decreased from the first to the last class (Table 2-1). The small number of spectral records available for pregnant cows in late gestation stages was caused by the dry-off period, which is usually at least six weeks before calving (Dairy Australia, 2017). Records from cows that were still open later in lactation and a long interval from insemination could be from cows that were not re-inseminated because of fertility or health issues but remained in the herd until the end of the lactation. We were unable to verify this assumption with certainty because culling records were sparse.

As expected, an increasing pattern existed for the average DIM from class 1 to class 7, but in each class, the DIM for open and pregnant records were in similar ranges. This was important to make sure that the models did not discriminate between open and pregnant cows solely based on changes in milk composition associated with different lactation stages (Mayeres et al., 2004), instead of the true effects of pregnancy.

The distribution of open and pregnant records and DIM at different periods after insemination for strategies 1 and 2 was similar to that of strategy 3. In the study by Toledo-Alvarado et al. (2018), the proportion of open cows decreased by week after insemination, meaning that open and pregnant records were not equally represented at different periods after insemination, in contrast to the present study.

3.2. Diagnosis of pregnancy status using MIR with different approaches

In this study, we explored 3 approaches to use MIR data to predict the pregnancy status of dairy cows. In the first approach, which is similar to the study by Toledo-Alvarado (2018), only single spectral records taken after insemination were retained. In terms of prediction accuracy, the value of AUC obtained through cross-validation in our study was comparable to that reported by Toledo-Alvarado (2018) when using a Bayesian model with Holstein cows and the whole MIR spectra as independent variables (i.e., 0.63 and 0.61, respectively; Table 2-2). Sensitivity and specificity values reported in this study were comparable and close to 0.60. For the test set, sensitivity was higher than specificity (i.e., 0.65 and 0.56, respectively), indicating that the model might have a slightly better ability to predict pregnant records correctly. The median probability values for correctly classified records were 0.56 for both cross-validation and testing. These low values (i.e., close to the 0.5 threshold) indicate poor confidence in the predictions. However, the negligible difference in prediction accuracy between calibration and cross-validation was a good sign because this implies that the model was unlikely to be overfitted and presented a certain robustness. One of the reasons for poor performance of the models might be that the spectra after insemination were too noisy, because they contain not only the effect of pregnancy but also other factors such as the lactation stage, herd management, or cow genetics (Collier et al., 2017). Consequently, finding a signal indicating pregnancy was challenging.

Table 2-2. Strategy 1 (single spectra after insemination): results of partial least squares discriminant analysis for calibration, random cow-independent 10-fold cross-validation (mean, with SD in parentheses), and cow-independent test set

Item[1]	Calibration	Cross-validation	Test
AUC	0.69	0.63 (0.05)	0.65
Sensitivity	0.63	0.60 (0.03)	0.65
Specificity	0.65	0.59 (0.07)	0.56
Probability	0.56	0.56 (0.00)	0.56

[1] AUC = area under the receiver operating characteristic curve; sensitivity = proportion of records belonging to pregnant cows that were correctly classified as pregnant; specificity = proportion of records belonging to open cows that were correctly classified as open; and probability = median predicted probability for correctly classified records (i.e., for each correctly classified record, the greatest predicted probability amongst the 2 classes open *vs.* pregnant was selected and the median was calculated).

To overcome the problem caused by noisy effects of multiple factors on milk composition, in the second approach, a spectrum collected before insemination at a specific lactation stage was subtracted from that collected after insemination (Table 2-3). By doing this, it was expected that the noisy effects of, for example, cow genetics, herd management, and parity might be eliminated, and the resulting spectrum would be less noisy and contain mainly the potential pregnancy signal (if there is one) and thus improve prediction accuracy. In fact, the idea of taking difference to remove noise and magnify signal has been studied in other research areas. For instance, noise reduction from speech signals by subtracting a signal during a nonspeech period is a well-known technique in audio signal processing (Boll, 1979). Similarly, in the field of earthquake engineering, Coelho et al. (2011) removed noise during seismic dynamic measurements using a spectral subtraction. Lainé et al. (2014) also used a spectral subtraction, but they derived the spectral difference by performing a mixed model on all open-cow data simultaneously (considering the fixed effects of parity, breed, month of test-day, DIM, and random effect of cows) to calculate an expected open spectrum, which was removed from the observed spectrum after insemination. In the present study, we performed the subtraction of spectra on observed data from the same cow and not using a mixed-model approach. Our strategy would, therefore, allow the removal of more specific effects to each cow and minor effects that were not accounted for in the study by Lainé et al. (2014), and without the need to know the population a priori.

Table 2-3. Strategy 2 (spectral differences): comparison of the results of partial least squares discriminant analysis with spectral differences vs. single spectra after insemination as independent variables for calibration, random cow-independent 10-fold cross-validation (mean, with SD in parentheses), and cow-independent test set

	Spectral differences			Single spectra after insemination		
Item[1]	Calibration	Cross-validation	Test	Calibration	Cross-validation	Test
AUC	0.77	0.59 (0.06)	0.58	0.78	0.60 (0.07)	0.62
Sensitivity	0.71	0.58 (0.07)	0.59	0.73	0.62 (0.06)	0.61
Specificity	0.72	0.54 (0.09)	0.52	0.72	0.56 (0.09)	0.47
Probability	0.60	0.59 (0.01)	0.58	0.59	0.58 (0.01)	0.60

[1]AUC = area under the receiver operating characteristic curve; sensitivity = proportion of records belonging to pregnant cows that were correctly classified as pregnant; specificity = proportion of records belonging to open cows that were correctly classified as open; and probability = median predicted probability for correctly classified records (i.e., for each correctly classified record, the greatest predicted probability amongst the 2 classes open vs. pregnant was selected and the median was calculated).

Unexpectedly, the prediction accuracy for calibration, cross-validation and testing using spectral differences were not noticeably different from those obtained using single spectra after insemination on the same data, with values of AUC, sensitivity specificity and median probability for testing of 0.58 versus 0.62, 0.59 versus 0.61, 0.52 versus 0.47 and 0.58 versus 0.60, respectively. These results were obtained on the same dataset to make sure the comparison is fair. Indeed, the restricted DIM window for the selection of spectra before insemination limited the number of data available for modeling. Different DIM windows were tested, but no improvements could be observed (results not shown). Further, the considerable drops in prediction accuracy between calibration and cross-validation (10-20%) implied a lack of robustness. This may be partly because of the small size of the dataset (Hawkins, 2004). Indeed, the dataset for the single spectra after insemination approach (Table 2-1) was larger, which has been shown to result in small differences in prediction accuracy between calibration and cross-validation. Given these results, it is not possible to conclude that the spectral difference approach we tested is superior to single spectra after insemination in diagnosing pregnancy status. However, the change in chemical composition of pasture for seasonal feeding systems, as in this study, could confuse the spectral signal before and after insemination, because the spectra were collected at different periods. In TMR systems, where the feed is more consistent, this method has a better chance of working. Also, it is still unclear whether a pregnancy signal in MIR exists of sufficient size to be detected, even after all the noisy effects have been removed, which could be elucidated using control studies such as the one by Lainé et al. (2017).

We have shown that using spectral data from various pregnancy stages was not enough to accurately diagnose the pregnancy status of cows. Several studies have reported a varying effect of pregnancy on milk composition throughout gestation. For example, Olori et al. (1997), Roche (2003), and Penasa et al. (2016) reported that milk fat and protein contents, as well as lactose content to a smaller extent, increased as pregnancy advances. More recently, Lainé et al. (2017) reported that the effect of pregnancy on the whole MIR spectrum was variable according to the spectral region and the gestation stage, and tended to be stronger at the end of the gestation. In the previous studies on the prediction of the pregnancy status of dairy cows (Lainé et al., 2014; Toledo-Alvarado et al., 2018), the number of days after insemination was limited to 120 and 91 respectively, and the pregnancy status was studied without a consideration of different gestation periods. In this study, we had data for later stages after insemination and we hypothesized that separating data by stage after insemination might reduce spectral variability as well as pregnancy signal variability and thus improve prediction accuracy. This approach is comparable to the first one (i.e., using single spectra after insemination as independent variables in the model), but the data were grouped into 7 classes based on the days after insemination.

Table 2-4 shows that, in general, for each class after insemination, the prediction accuracies for cross-validation and testing were relatively similar, whereas they were higher for calibration, indicating potential lack of robustness in all models. Similar to the results of the spectral difference approach, the lack of robustness might just be a consequence of having a small dataset in each class.

Interestingly, the prediction accuracy for class 7 (i.e., records taken 181 or more days after insemination) was 20% higher than that of classes 1 to 5. We also observed higher cross-validation and test AUC, sensitivity, and median probability for class 6 (i.e., records between 151 and 180 days after insemination) compared with classes 1 to 5. Globally, the median probabilities for correctly classified records were relatively close to the 0.5 threshold, showing that the model was unable to separate well the data, though these probabilities were slightly higher in the last 2 groups. These results imply that although MIR may not be sufficient to predict the pregnancy status of dairy cows in early and mid-stages after insemination (from 1 to 150 d), promising results were obtained for records taken 151 d or more after insemination.

Table 2-4. Strategy 3 (stages after insemination): results of partial least squares discriminant analysis for calibration, random cow-independent 10-fold cross-validation (mean, with SD in parentheses), and cow-independent test set for the 7 classes of records based on the number of days after insemination[1]

Class	Data set	AUC	Sensitivity	Specificity	Probability
1 (1-30 d)	Calibration	0.74	0.70	0.68	0.58
	Cross-validation	0.63 (0.05)	0.60 (0.05)	0.58 (0.09)	0.58 (0.00)
	Test	0.62	0.57	0.58	0.58
2 (31-60 d)	Calibration	0.67	0.62	0.67	0.55
	Cross-validation	0.64 (0.03)	0.58 (0.05)	0.63 (0.06)	0.52 (0.01)
	Test	0.61	0.59	0.57	0.54
3 (61-90 d)	Calibration	0.74	0.66	0.69	0.58
	Cross-validation	0.64 (0.03)	0.60 (0.08)	0.61 (0.10)	0.58 (0.01)
	Test	0.59	0.61	0.52	0.58
4 (91-120 d)	Calibration	0.75	0.66	0.69	0.59
	Cross-validation	0.62 (0.06)	0.57 (0.07)	0.60 (0.13)	0.58 (0.01)
	Test	0.65	0.60	0.66	0.58
5 (121-150 d)	Calibration	0.78	0.70	0.73	0.59
	Cross-validation	0.63 (0.10)	0.64 (0.10)	0.59 (0.16)	0.59 (0.02)
	Test	0.61	0.62	0.59	0.58
6 (151-180 d)	Calibration	0.88	0.78	0.75	0.63
	Cross-validation	0.71 (0.06)	0.64 (0.12)	0.63 (0.13)	0.63 (0.03)
	Test	0.70	0.71	0.55	0.61
7 (>= 181 d)	Calibration	0.91	0.81	0.85	0.65
	Cross-validation	0.86 (0.05)	0.76 (0.10)	0.79 (0.16)	0.65 (0.02)
	Test	0.82	0.75	0.74	0.63

[1] AUC = area under the receiver operating characteristic curve; sensitivity = proportion of records belonging to pregnant cows that were correctly classified as pregnant; specificity = proportion of records belonging to open cows that were correctly classified as open; and probability = median predicted probability for correctly classified records (i.e., for each correctly classified record, the greatest predicted probability amongst the 2 classes open vs. pregnant was selected and the median was calculated).

In late gestation, cows are subject to more significant endocrine changes, reduced feed intake, and greater nutritional demand of the growing calf compared to the early gestation period (Ingvartsen and Andersen, 2000; Penasa et al., 2016). Indeed, almost no increase in estrogen level exists during the first 3 mo of pregnancy, but it gradually rises from the third to the seventh month, before suddenly increasing after 7 mo of pregnancy (Parkhie et al., 1966; Penasa et al., 2016). The increase in estrogen levels in maternal blood during the last weeks of pregnancy affects nutrient partitioning, which has a deleterious effect on milk yield and, consequently, milk composition (Parkhie et al., 1966; Olori et al., 1997). Olori et al. (1997) also mentioned that the varying effect of pregnancy stages on milk composition suggests that the mechanism for the secretion of the solid and nonsolid components of milk might be unequally affected by pregnancy. This may offer an explanation of why good accuracy was obtained for the models using records after 150 d after insemination but not in early gestation.

A potential application of MIR for pregnancy diagnosis at advanced stages after insemination is for detection of late fetal abortion. Mid- to late-term fetal losses are often detected using careful observations because of vaginal discharge or expulsion of placenta. However, not all abortions are detected, especially those that occur before 180 d of pregnancy (Bronner et al., 2015), and therefore, MIR could be used as a complementary tool, provided that it can be used for slightly earlier diagnosis than 151 d. This would need to be confirmed using a larger dataset. Adding data collected from other analyses such as metabolomics could further improve the prediction accuracy. Also, using a much larger dataset, it is worth exploring whether deep learning techniques can help provide a better prediction.

The results obtained for strategy 3 are consistent with Figure 2-2, which is a heatmap of correlations between coefficients of the 7 models associated with the 7 classes. Correlations were generally low, with a maximum of 0.41. However, an interesting finding was that the closer the groups, the higher the correlations (e.g., group 7 had high correlations with groups 5 and 6, but lower correlations with group 1 to 4). These results potentially imply that pregnancy was detected using different spectral wavelengths or with different weights at different stages after insemination, with an evolving pattern from the beginning towards the end of the pregnancy. This confirms the poor prediction accuracy of the global model for pregnancy detection in the present study and the previous studies by Lainé et al. (2014) and Toledo-Alvarado et al. (2018).

The low frequency of milk recoding, which is monthly in most countries, could be considered as an obstacle in applying MIR to predict not only pregnancy status but also other traits, such as indicators of early lactation metabolic diseases (e.g. β-hydroxybutyrate, fatty acids). Only when scientists have been able to demonstrate the costs and benefits of having MIR collected more frequently, will farmers be willing to alter their milk-testing practice so that these MIR tools can be applied more effectively.

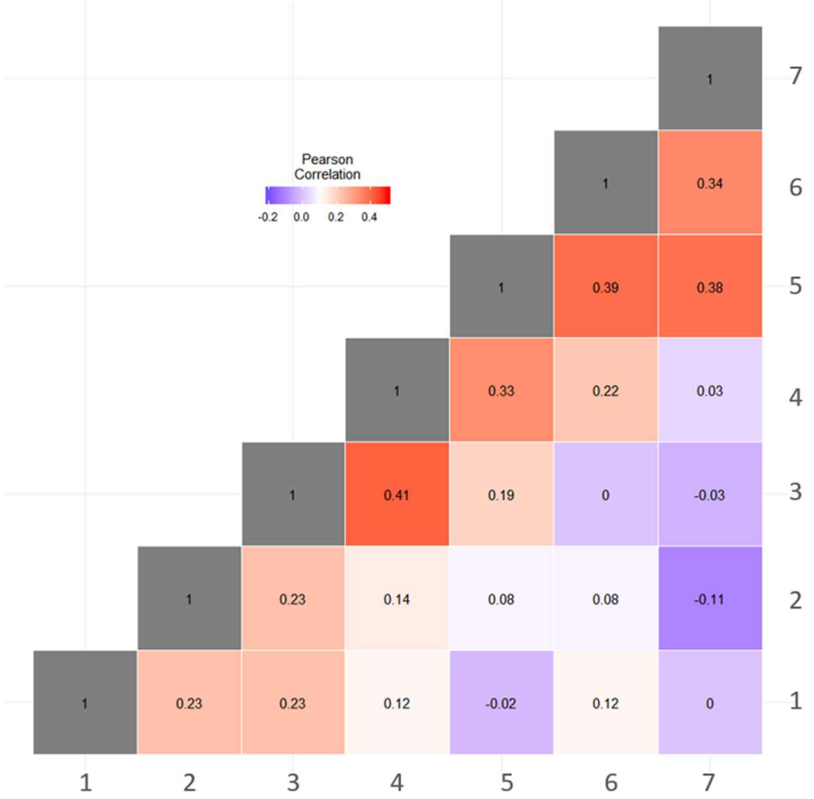

Figure 2-2. Heatmap of correlations between coefficients of the 7 models associated with the classes of records based on the number of days after insemination (class 1 = 1 to 30 d, class 2 = 31 to 60 d, class 3 = 61 to 90 d, class 4 = 91 to 120 d, class 5 = 121 to 150 d, class 6 = 151 to 180 d, class 7 = ≥181 d).

4. Conclusions

In this study, we have shown that milk MIR spectral data collected at different stages after insemination, when used directly or taking a spectral difference, were not sufficient to detect the pregnancy status of dairy cows. However, the models developed using data recorded after 150 d of pregnancy showed promising prediction accuracy, with the value of AUC around 78% obtained through random cow-independent cross-validation. If this can be confirmed using a larger dataset and can be done a little earlier, the models could be used as a complementary tool to detect fetal abortion.

5. Acknowledgments

The milk mid-infrared spectral data were obtained as part of the MIRforProfit project “Integrating very large genomic and milk mid-infrared data to improve profitability of dairy cows”, funded by the Australian Government Department of Agriculture (Canberra, Australia) as part of the Rural R&D for Profit program. The authors thank DairyBio project, funded by Dairy Australia (Melbourne, Australia), The Gardiner Foundation (Melbourne, Australia); and Agriculture Victoria (Melbourne, Australia) for supporting this research. The staff of DataGene (Melbourne, Australia) are gratefully acknowledged for providing the data used in this study. Special thanks are given to Peter Nish from TasHerd Pty. Ltd. (Hadspen, Tasmania, Australia) for providing MIR spectral data, and Hico Pty. Ltd. (Maffra, Victoria, Australia) for collecting milk samples, and the farmers whose data were used. The first author, Pauline Delhez, who is Research Fellow of the National Fund for Scientific Research (FNRS, Brussels, Belgium), completed this research while on a sabbatical at Agriculture Victoria, supported by the FNRS and the University of Liège through funding for research stays abroad. The authors have not stated any conflicts of interest.

6. References

Boll, S.F. 1979. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust. 27:113–120. https://doi.org/10.1109/TASSP.1979.1163209.

Bonfatti, V., G. Di Martino, and P. Carnier. 2011. Effectiveness of mid-infrared spectroscopy for the prediction of detailed protein composition and contents of protein genetic variants of individual milk of Simmental cows. J. Dairy Sci. 94:5776–5785. https://doi.org/10.3168/jds.2011-4401.

Bronner, A., E. Morignat, V. Hénaux, A. Madouasse, E. Gay, and D. Calavas. 2015. Devising an indicator to detect mid-term abortions in dairy cattle: a first step towards syndromic surveillance of abortive diseases. PLoS One 10:1–16. https://doi.org/10.1371/journal.pone.0119012.

Coelho, B., P. Hölscher, and F. Barends. 2011. Enhancement of double integration procedure through spectral subtraction. Soil Dyn. Earthq. Eng. 31:716–722. https://doi.org/10.1016/j.soildyn.2010.12.013.

Collier, R.J., Y. Xiao, and D.E. Bauman. 2017. Regulation of factors affecting milk yield. Pages 3-15 in Nutrients in Dairy and Their Implications for Health and Disease. 1st ed. R.R. Watson, V.R. Preedy, and R.J. Collier, ed. Acad. Press, Elsevier, Amsterdam, the Netherlands. http://doi.org/10.1016/B978-0-12-809762-5.00001-2.

Dairy Australia. 2017. Decide dry cow strategy drying off dates, antibiotic. Dry cow and internal teat sealant treatment. Pages 1:9 in Late Lactation, Technote.

Accessed June 7, 2019. https://www.dairyaustralia.com.au/-/media/dairyaustralia/documents/farm/animal-care/mastitis/research-reports-and-papers/new/technote--14-dry-cow-strategy.ashx?la=en&hash=7DE338D5C50BB5CC2C55403EFAF15EEB6FD4A77B.

De Maesschalck, R., D. Jouan-Rimbaud, and D.L. Massart. 2000. The Mahalanobis distance. Chemom. Intell. Lab. Syst. 50:1–18.

Fricke, P.M., A. Ricci, J.O. Giordano, and P.D. Carvalho. 2016. Methods for and implementation of pregnancy diagnosis in dairy cows. Vet. Clin. North Am. Food Anim. Pract. 32:165–180. https://doi.org/10.1016/j.cvfa.2015.09.006.

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

Grelet, C., 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.

Haile-Mariam, M., and J.E. Pryce. 2019. Genetic evaluation of gestation length and its use in managing calving patterns. J. Dairy Sci. 102:476–487. https://doi.org/10.3168/jds.2018-14981.

Hawkins, D.M. 2004. The problem of overfitting. J. Chem. Inf. Comput. Sci. 44:1–12.

ICAR. 2017a. Section 2 - Guidelines for Dairy Cattle Milk Recording. Int. Comm. Anim. Record., Rome, Italy.

ICAR. 2017b. Section 13 - Guidelines for On-Line Milk Analysis. Int. Comm. Anim. Record., Rome, Italy.

Ingvartsen, K.L., and J.B. Andersen. 2000. Integration of metabolism and intake regulation: a review focusing on periparturient animals. J. Dairy Sci. 83:1573–1597. https://doi.org/10.3168/jds.s0022-0302(00)75029-6.

Kuhn, 2019. Package ‘caret’. R package version 6.0-84. R Foundation for Statistical Computing, Vienna, Austria.

Lainé, A., H. Bel Mabrouk, L. Dale, C. Bastin, and N. Gengler. 2014. How to use mid-infrared spectral information from milk recording system to detect the pregnancy status of dairy cows. Comm. Appl. Biol. Sci 79:33–38.

Lainé, A., C. Bastin, C. Grelet, H. Hammami, F.G. Colinet, L.M. Dale, A. Gillon, J. Vandenplas, F. Dehareng, and N. Gengler. 2017. Assessing the effect of pregnancy stage on milk composition of dairy cows using mid-infrared spectra. J. Dairy Sci. 100:2863–2876. https://doi.org/10.3168/jds.2016-11736.

Mayeres, P., J. Stoll, J. Bormann, R. Reents, and N. Gengler. 2004. Prediction of daily milk, fat, and protein production by a random regression test-day model. J. Dairy Sci. 87:1925–1933. https://doi.org/10.3168/jds.s0022-0302(04)73351-2.

McParland, S., and D.P. Berry. 2016. The potential of Fourier transform infrared spectroscopy of milk samples to predict energy intake and efficiency in dairy cows. J. Dairy Sci. 99:4056–4070. https://doi.org/10.3168/jds.2015-10051.

Olori, V.E., S. Brotherstone, W.G. Hill, and B.J. McGuirk. 1997. Effect of gestation stage on milk yield and composition in Holstein Friesian dairy cattle. Livest. Prod. Sci. 52:167–176. https://doi.org/10.1016/S0301-6226(97)00126-7.

Parkhie, M.R., L.O. Gilmore, and N.S. Fechheimer. 1966. Effect of successive lactations, gestation, and season of calving on constituents of cows' milk. J. Dairy Sci. 49:1410–1415. https://doi.org/10.3168/jds.s0022-0302(66)88103-1.

Penasa, M., M. De Marchi, and M. Cassandro. 2016. Short communication: effects of pregnancy on milk yield, composition traits, and coagulation properties of Holstein cows. J. Dairy Sci. 99:4864–4869. https://doi.org/10.3168/jds.2015-10168.

Roche, J.R. 2003. Effect of pregnancy on milk production and bodyweight from identical twin study. J. Dairy Sci. 86:777–783. https://doi.org/10.3168/jds.s0022-0302(03)73659-5.

Rozenstein, O., T. Paz-Kagan, C. Salbach, and A. Karnieli. 2015. Comparing the effect of preprocessing transformations on methods of land-use classification derived from spectral soil measurements. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8:2393–2404. https://doi.org/10.1109/JSTARS.2014.2371920.

Shetty, N., G. Difford, J. Lassen, P. Løvendahl, and A.J. Buitenhuis. 2017. Predicting methane emissions of lactating Danish Holstein cows using Fourier transform mid-infrared spectroscopy of milk. J. Dairy Sci. 100:9052–9060. https://doi.org/10.3168/jds.2017-13014.

Šimundić, A.-M. 2009. Measures of diagnostic accuracy: basic Definitions. Measures of Diagnostic Accuracy: Basic Definitions. EJIFCC. 19:203-211.

Soyeurt, H., P. Dardenne, F. Dehareng, G. Lognay, D. Veselko, M. Marlier, C. Bertozzi, P. Mayeres, and N. Gengler. 2006. Estimating fatty acid content in cow milk using mid-infrared spectrometry. J. Dairy Sci. 89:3690–3695. https://doi.org/10.3168/jds.s0022-0302(06)72409-2.

Soyeurt, H., F. Dehareng, N. Gengler, S. McParland, E. Wall, D.P. Berry, M. Coffey, and P. Dardenne. 2011. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. J. Dairy Sci. 94:1657–1667. https://doi.org/10.3168/jds.2010-3408.

Toledo-Alvarado, H., A.I. Vazquez, G. de los Campos, R.J. Tempelman, G. Bittante, and A. Cecchinato. 2018. Diagnosing pregnancy status using infrared spectra and milk composition in dairy cows. J. Dairy Sci. 101:2496–2505. https://doi.org/10.3168/jds.2017-13647.

Vanlierde, A., H. Soyeurt, N. Gengler, F.G. Colinet, E. Froidmont, M. Kreuzer, F. Grandl, M. Bell, P. Lund, D.W. Olijhoek, M. Eugène, C. Martin, B. Kuhla, and F. Dehareng. 2018. Short communication: development of an equation for estimating methane emissions of dairy cows from milk Fourier transform mid-infrared spectra by using reference data obtained exclusively from respiration chambers. J. Dairy Sci. 101:7618–7624. https://doi.org/10.3168/jds.2018-14472.

Velek, K., J. Lawrence, R. Linscott, E. Martel, H. Pearse, R. Hildreth, and C. Egli. 2012. IDEXX Milk Pregnancy Test — a new tool for pregnancy diagnosis in cattle. IDEXX Laboratories Inc., Westbrook, ME.

Wang, Q., and H. Bovenhuis. 2019. Validation strategy can result in an overoptimistic view of the ability of milk infrared spectra to predict methane emission of dairy cattle. J. Dairy Sci. 102:6288-6295. https://doi.org/10.3168/jds.2018-15684.

3

Predicting milk mid-infrared spectra from first parity Holstein cows using a test-day mixed model with the perspective of herd management

Adapted from: Delhez, P., F. Colinet, S. Vanderick, C. Bertozzi, N. Gengler, and H. Soyeurt. Predicting milk mid-infrared spectra from first parity Holstein cows using a test-day mixed model with the perspective of herd management. J. Dairy Sci. In press.

Outline

Unlike lactation models, test-day models (TDM) model individual repeated measurements over time that are specific to a particular testing day. This chapter aimed to model milk MIR spectra using a TDM for management purposes. The research was carried out using data from first parity Holstein cows extracted from the Walloon milk recording database. Existing or future MIR prediction equations can be applied to the modeled spectra to predict traits of interest in different situations, without the need to develop one TDM separately for each trait. Interests for management are for instance the comparison between expected (i.e., modeled) records and observed records or the analysis of animal or herd effect solutions of the model to detect individual or herd abnormalities (e.g., feed or health problems), the prediction of missing or future data, or simulations by changing effects in the model. Depending on the MIR traits predicted from the modeled spectra, this could contribute to optimizing the economic and/or environmental performances of dairy herds.

Abstract

The use of test-day models to model milk mid-infrared (MIR) spectra for genetic purposes has already been explored; however, little attention has been given to their use to predict milk MIR spectra for management purposes. The aim of this paper was to study the ability of a test-day mixed model to predict milk MIR spectra for management purposes. A dataset containing 467,496 test-day observations from 53,781 Holstein dairy cows in first lactation was used for model building. Principal component analysis was implemented on the selected 311 MIR spectral wavenumbers to reduce the number of traits for modeling; 12 principal components (PC) were retained, explaining approximately 96% of the total spectral variation. Each of the retained PC was modeled using a single trait test-day mixed model. The model solutions were used to compute the predicted scores of each PC, followed by a back-transformation to obtain the 311 predicted MIR spectral wavenumbers. Four new datasets, containing altogether 122,032 records, were used to test the ability of the model to predict milk MIR spectra in four distinct scenarios with different levels of information about the cows. The average correlation between observed and predicted values of each spectral wavenumber was 0.85 for the modeling dataset and ranged from 0.36 to 0.62 for the scenarios. Correlations between milk fat, protein and lactose contents predicted from the observed spectra and from the modeled spectra ranged from 0.83 to 0.89 for the modeling set and from 0.32 to 0.73 for the scenarios. Our results demonstrated a moderate but promising ability to predict milk MIR spectra using a test-day mixed model. Current and future MIR traits prediction equations could be applied to the modeled spectra to predict all MIR traits in different situations instead of developing one test-day model separately for each trait. Modeling MIR spectra would benefit farmers for cow and herd management, for instance through prediction of future records or comparison between observed and expected wavenumbers or MIR traits for the detection of health and management problems. Potential resulting tools could be incorporated into milk recording systems.

Key words: mid-infrared spectrometry, mixed model, milk composition, management

1. Introduction

Fourier-transform mid-infrared (MIR) spectrometry is a valuable technique to describe the molecular structure of food materials. It involves the absorption of electromagnetic radiation by a sample at frequencies (400 to 4,000 cm^{-1}) that are characteristic of specific chemical bonds of a molecule (Van de Voort, 1992). The MIR spectrometry provides analyses with high throughput, at low cost, and on a large scale. Therefore, it is used in the dairy industry worldwide to predict major milk components (e.g., lactose, fat, protein contents, or urea) for milk quality control, milk payment, management of herds, or genetic studies (Gengler et al. 2016; ICAR, 2017). More recently, studies have focused on milk MIR spectrometry to predict other traits such as fine milk composition (Soyeurt et al., 2009; Bonfatti et al. 2011), milk technological properties (Ferragina et al., 2013; Visentin et al., 2015), body energy status (McParland et al., 2011), enteric methane emissions (Vanlierde et al., 2018), body weight (Soyeurt et al., 2019) or geographical origin of milk (Scampicchio et al., 2016; Caredda et al., 2017).

Test-day models (TDM) model individual test-day records (i.e., repeated measurements over time that are specific to a particular testing day) like milk yield or MIR traits records in dairy cattle. One of several benefits of TDM is the ability to account for individual animal effects and for environmental factors occurring on the day of milk recording (Wiggans and Goddard, 1997). Their use in genetic evaluations has been widely explored (e.g., De Roos et al., 2004; Hammami et al., 2009; Leclercq et al., 2013). However, relatively few published studies considered the use of TDM for prediction and management purposes. For example, Mayeres et al. (2004), Caccamo et al. (2008) and Gillon et al. (2010) investigated the use of TDM to predict future daily milk, fat and protein yields for management. Koivula et al. (2007) studied herd-management effect solutions from milk yield, fat and protein contents, and SCC TDM, and Bastin et al. (2009) explored the solutions and predictions of a TDM for milk urea.

Modeling MIR traits with TDM would have numerous benefits for herd management, such as the prediction of future records or missing lactation records, the evaluation of cow responses to herd management changes by adjusting herd factors in the model, or simulation of new records by modifying effects in the models. Decision-support tools incorporating such models could help farmers to detect problems affecting individual cows or the whole herd through direct comparison between the actual and predicted traits, predict the production potential of heifers, analyze the results of diet or environmental changes, and so on. With the increasing number of MIR traits (Gengler, 2016), implementing one TDM for each trait would be time- and resource-consuming from a workload and computational point of view. Therefore, predicting the whole milk MIR spectrum using TDM would be beneficial because of the reduction of the number of models to implement. All current and future MIR prediction equations could be applied to the modeled spectra to predict all MIR traits, without the need to do one TDM separately for each trait. Also, in some instances, using the whole spectra instead of specific MIR traits could bring more comprehensive information.

Several authors already modeled milk MIR spectral wavenumbers using TDM for genetic purposes (e.g., Soyeurt et al., 2010; Wang et al., 2016; Rovere et al., 2019), but few tackled the possible use for predictive and management purposes (e.g, Dagnachew et al., 2013b; Lainé et al., 2017). Therefore, the objective of this paper was to study the ability of a test-day mixed model to predict milk MIR spectra from first parity Holstein cows with the perspective of herd management. To do so, we compared observed and predicted spectral wavenumbers and we examined predicted MIR milk components obtained using the observed vs. predicted spectra. We tested different scenarios to assess the accuracy of the model when knowing more or less information about the cows.

2. Materials and methods

2.1. Modeling data

The data used to build the model were collected from January 2012 to July 2017 by the Walloon Breeding Association (Ciney, Belgium) during the Walloon routine milk recording. A total of 467,496 test-day records from 53,781 Holstein dairy cows in first lactation within 541 herds were selected. Each record included the identification number of the cow and herd, the lactation stage (i.e., DIM), the test date, milk composition information (fat, protein and lactose contents, SCC), the milk MIR spectrum and pedigree data. Pedigree data contained 139,385 animals extracted from the database used for the official Walloon genetic evaluation and were limited to animals born after 1985. Milk MIR spectra were obtained by the analysis of individual milk samples on MilkoScan FT6000 spectrometers (Foss, Hillerod, Denmark) at the 'Comité du Lait' laboratory (Battice, Belgium). Milk MIR spectra included 1,060 spectral wavenumbers expressed in absorbance and covering the absorption of light in the infrared region located from 900 to 5,000 cm^{-1}. All studied cows had at least 5 test-day records per lactation and belonged to herds with more than 10 recorded cows in first lactation on average over the studied period. Records with fat and protein contents as well as milk yield out of the limits set by the International Committee for Animal Recording (ICAR, 2017) were discarded (i.e., 3 L < milk yield < 99.9 L, 1.5 g/dL of milk < fat < 9 g/dL of milk, 1 g/dL of milk < protein < 7 g/dL of milk). Records within the 0.1% upper values and 0.1% lower values for lactose content as well as within the upper 0.1% values for milk somatic cells were removed. Records with DIM values higher than 563 (1% upper values) were also discarded.

2.2. Preprocessing of modeling MIR spectra

To remove baseline variation, the first derivative at wavenumber X was calculated on the raw spectra as the difference between the spectral wavenumber X-2 and the spectral wavenumber X+2. A total of 311 spectral wavenumbers out of the 1,060 were retained for this study, covering three MIR spectral regions: 933 to 1,589 cm^{-1}, 1,704 to 1,809 cm^{-1} and 2,553 to 2,981 cm^{-1}. Some spectral regions were excluded based on the experience of the research team; these are regions that cannot be used effectively as they have low signal-to-noise ratio or little relevant chemical information (e.g., Iñón et al, 2004; Soyeurt et al., 2010; Capuano et al., 2014). Previous studies involving MIR spectra also retained similar spectral regions (e.g., Grelet et al., 2016; Vanlierde et al., 2018).

Spectral wavenumbers were standardized (i.e., centered and scaled) by subtracting the corresponding mean and dividing by the corresponding standard deviation. Then, principal components analysis (PCA) was carried out on the 311 standardized spectral wavenumbers. We performed standardization and PCA using the “stats” package in R (version 3.3.3; R Core Team, 2017). The PCA method extracts the information from a multivariate dataset and expresses it as a set of new variables called principal components (PC). These PC are a linear combination of the original variables keeping the largest amount of information contained in the original data (i.e., the PC explain most of the variance of the original data, Palm, 1998). The use of PCA has multiple advantages for this study: decreasing the number of traits to reduce computational operations and the independence of PC allowing separate modeling without losing information about the interactions between spectral wavenumbers (Soyeurt et al., 2010). Let $\mathbf{X}_{(\mathrm{NxK})}$ be the matrix for the 311 spectral wavenumbers for the 467,496 test-day records where N is the number of test-day records and K is the number of spectral wavenumbers. The contributions of the spectral wavenumbers to each PC are given by the eigenvectors. The amount of variance retained by each PC is measured by its eigenvalue. PCA for **X** is expressed as:

$$\mathbf{X}_{(\mathrm{N}\times\mathrm{K})} = \mathbf{Z}_{(\mathrm{N}\times\mathrm{M})} \times \mathbf{V}^{\mathrm{T}}_{(\mathrm{M}\times\mathrm{K})} + \mathbf{E}_{(\mathrm{N}\times\mathrm{K})} \quad [1]$$

where M is the number of PC selected (M <= K), $\mathbf{Z}_{(\mathrm{N}\times\mathrm{M})}$ is the matrix of the PC scores, $\mathbf{V}^{\mathrm{T}}_{(\mathrm{M}\times\mathrm{K})}$ is the transpose of the eigenvector matrix and $\mathbf{E}_{(\mathrm{N}\times\mathrm{K})}$ is the residual matrix after the M PC are extracted from $\mathbf{X}_{(\mathrm{N}\times\mathrm{K})}$. The number of PC selected was based on the proportion of variance of the spectral wavenumbers explained and on the accuracy of the model.

2.3. Model

A total of M single-trait test-day mixed models was applied on the PC scores for the M selected PC (i.e., one model was run separately for each PC). Using single-trait models was preferred over a multi-trait model, because PC are phenotypically de-correlated and because of the computational advantage for large datasets. Running several single-trait models in parallel (i.e., data parallelism) is faster and less computationally demanding than running one multi-trait model, and therefore easier to implement in practice (Shallue et al., 2019). The single-trait test-day mixed model used was defined as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e} \quad [2]$$

where **y** is the vector of observations (PC scores), **b** is the vector of fixed effects [including herd-test-year (HTY) effect, herd-test-month (HTM) effect and DIM], **u** is the vector of random effects [including herd-test-day (HTDr) effect, additive genetic effect, and permanent environmental effect] and e is the vector of random residual effects. **X** and **Z** are the corresponding incidence matrices. The distributional assumption about the random terms of the model was

$$\text{HTDr} \sim N(\mathbf{0}, \mathbf{I}\sigma^2_{HTDr}),$$

$$\text{additive genetic} \sim N(\mathbf{0}, \mathbf{A}\sigma^2_{g}),$$

$$\text{permanent environnement} \sim N(\mathbf{0}, \mathbf{I}\sigma^2_{PE}),$$

$$\text{e} \sim N(\mathbf{0}, \mathbf{I}\sigma^2_{e})$$

where **I** was an identity matrix, **A** was the pedigree relationship matrix, σ^2_{HTDr} was the variance of the herd test-day effect, σ^2_{g} was the additive genetic variance, σ^2_{PE} was the variance of the permanent environmental effect, and σ^2_{e} was the error variance. We divided DIM into 38 classes of 15 d. Mayeres et al. (2002, 2004) suggested a remodeling of the HTD fixed effect usually used in genetic models by replacing it with three herd-test-related effects to allow prediction of future test-day measurements for predictive purposes. These three effects are a fixed herd-test month-period effect, a fixed herd-test-year effect, and a random herd-test-day effect. The herd-test month-period effect and the herd-test-year effect represent the herd level and its seasonal trend and allow the prediction of future records, while the herd-test-day effect takes into account the effect of the herd at a specific date and is not assigned to the two other herd effects. In our study, we considered a simple herd-test-month effect instead of a herd-test month-period effect because the number of years considered in our study was higher. Variance components for random effects were estimated using expectation-maximization REML as described by Misztal et al. (2018).

2.4. Spectral wavenumber predictions for the modeling dataset

The solutions of the model were obtained using the Best Linear Unbiased Prediction (BLUP) method solved using the Preconditioned Conjugate Gradient algorithm (Tsuruta et al., 2001; Misztal et al., 2018). Estimates of the fixed and random effects were used to compute the predicted scores of each PC. Then the predicted 311 MIR spectral wavenumbers (i.e., $\widehat{\boldsymbol{X}}$ matrix) were obtained using the equation derived from [1]:

$$\widehat{\boldsymbol{X}} = \widehat{\boldsymbol{Z}} \times \mathbf{V}^{\mathrm{T}} \text{ [3]}$$

where $\widehat{\boldsymbol{Z}}$ is the matrix of the predicted PC scores and $\mathbf{V}^{\mathrm{T}}$ is the transpose of the eigenvector matrix of the PCA; and by adding the mean and multiplying by the standard deviation as data were standardized. For each spectral wavenumber, correlations were computed between observed and predicted values. Estimated variance components were also back-transformed to spectral basis using the eigenvector matrix.

2.5. Spectral wavenumber predictions for different scenarios

Four new datasets, containing altogether 122,032 records from first parity Holstein cows, were used to test the ability of the model to predict accurately milk MIR spectrum in four distinct scenarios with different levels of information about the cows. These new datasets had different levels of relatedness with the modeling dataset. Information about each scenario and considered effect estimates for predictions are presented in Table 3-1.

The 2 first scenarios corresponded to common practical situations. The first scenario represented predictions of future test-day spectra for cows in production. It included cows for which all effects were known except the HTDr. Data were collected from August to early December 2017 (i.e., over a four-month period after modeling data). As expected, the median DIM was high (i.e., 310 d), but was within the DIM range of the modeling set. A similar scenario with the same known and unknown effects could be applicable to the imputation of missing spectral records for a cow with previous recorded spectral data. The second scenario represented predictions of records for a new cow in a known herd (e.g., the evaluation of the potential of a heifer before it produces milk). Data were collected from August to early December 2017 and included cows with no known test date (i.e., unknown HTDr) and no previous animal data (i.e., the permanent environmental effect was unknown).

The third and fourth scenarios were created based on records discarded when selecting data for the modeling dataset. The third scenario included cows with permanent environment as the only unknown effect. This could represent a situation when we want to evaluate the potential of a new cow like a heifer and compare it to existing test-day records of the other cows of the herd. The fourth scenario included cows with minimal information, i.e., only the DIM effect and the genetic effect were known. Data were unrelated to the modeling set regarding cows and herds, but were collected over the same period. This scenario, rather theoretical, would represent predictions of records for farms that have no spectral data, but only pedigree information. The interest, more academic, is to see how the accuracy would evolve when very little information is known.

For each scenario, the predicted PC scores were obtained using the solutions from solving equations associated with the mixed model [2]. New datasets used had different levels of relatedness with the previous modeling dataset adding scenario records. The trait values (i.e., PC scores) of the scenario records were considered unknown during solving. The pedigree was updated compared to the modeling dataset in order to add animals related to the scenario cows (i.e., 191,685 animals in total). This strategy permitted estimation of predicted PC scores for scenario records by summing the solutions (i.e., fixed and random effect estimates) equivalent to those obtained using the modeling data. Missing effects, depending on the scenario, were set to zero. By extending the pedigree, solutions for the genetic effect were automatically computed for new animals with scenario records. These genetic solutions can be considered being the estimated breeding value for a cow that had previous production records (i.e., scenario 1), or parent averages derived through the pedigree in all other situations (i.e., scenarios 2, 3 and 4). The permanent environmental effect solutions were known for cows with previous production records, but were set to an expected value of zero for other cows. The predicted PC scores were back-transformed to the 311 spectral wavenumbers using the PCA eigenvectors of the modeling set (Equation [3]) followed by de-standardization. For each spectral wavenumber, correlations were computed between observed values and predicted values.

Table 3-1. Description of the 4 scenarios (number of records, cows and herds; effect estimates; and meaning)

				Effect estimate[1]						
Scenario	Records	Cows	Herds	HTY	HTM	DIM	HTDr	Gen	PE	Example of meaning in practice
1	10,057	4,246	422	X	X	X		X	X	Future or missing records for cows with previous records
2	17,025	7,035	430	X	X	X		X*		Potential of heifers or new cows
3	32,315	12,126	538	X	X	X	X	X*		Potential of heifers or new cows for a specific test date
4	62,625	7,507	414			X		X*		Cows with minimal information

[1] Crosses in the effect estimate columns indicate known estimates used to calculate the predictions of spectral data. Unknown effect estimates were set to zero. HTY = herd-test-year fixed effect; HTM = herd-test-month fixed effect; DIM = fixed effect of days in milk; HTDr = random herd-test-day effect; Gen = random additive genetic effect; PE = random permanent environmental effect

*An asterisk means that the genetic solutions are based on parent averages. The absence of an asterisk means that the genetic solutions are based on the estimated breeding value of the cow

2.6. MIR trait predictions

To assess the usefulness of predicted MIR spectra for practical applications, fat, protein, and lactose contents in milk were predicted from observed MIR spectra and predicted MIR spectra for the modeling dataset and scenarios. The prediction equations for fat, protein, and lactose contents had a cross-validation R^2 of 0.99, 0.99, and 0.91, respectively and a root mean square error of prediction of 0.06, 0.04, and 0.06 g/dL of milk, respectively. Reference values to build these equations came from the predicted phenotypes obtained using the Milkoscan FT6000 (i.e., these phenotypes were based on the MIR spectra) because no phenotypes were available from chemical analysis. This explains the high R^2 values obtained for these prediction equations. Traits predicted from observed and modeled MIR spectra were compared using correlations, descriptive statistics and the root-mean-square error (RMSE). As a comparison, we also modeled directly milk fat, protein, and lactose contents using the TDM in equation [2]. We compared correlations between traits predicted from observed spectra vs. traits predicted from spectra modeled from a TDM with correlations between traits predicted from observed spectra vs. traits modeled directly from a TDM. This way, we can compare the accuracy of using prediction equations on MIR spectra predicted from a TDM with the accuracy of modeling MIR traits directly using a TDM.

3. Results and discussion

3.1. Preprocessing of modeling MIR spectra

Several preprocessing methods exist for MIR data (Rinnan et al., 2009). We chose first derivative because, in our study, data were collected with different spectrometers (but originating from the same laboratory and same brand) over several years and derivative is a useful technique to remove baseline variation that may occur because of instabilities between instruments of the same brand and over time (Owen, 1995). First-derivative preprocessing made data in our study more conform to normality as skewness and excess kurtosis were globally closer to zero after derivative (Figure 3-1; Kim, 2013). Further analyses of our data and comparison of models using derived and raw spectral data suggested that first derivative was a useful pretreatment for our study and improved results accuracy (results not shown). Several authors who modeled MIR spectra using TDM previously did not precorrect spectral data (e.g., Bittante and Cecchinato, 2013; Wang et al., 2016; Zaalberg et al., 2019) while some precorrected MIR spectra using methods such as derivatives (e.g., Belay et al., 2017; Lainé et al., 2017) or extended multiplicative signal correction (e.g., Dagnachew et al., 2013a; Belay et al., 2017). Among these authors, Belay et al. (2017) indicated that spectral preprocessing improved prediction accuracy. Some authors stated that it is important to test different pretreatment methods to make the most advised choice and that pretreatment might only improve accuracy for some traits (De Marchi et al., 2011; Soyeurt et al., 2011; Mineur et al., 2017). This should be explored further.

One advantage of PCA was the reduction of spectral variables for modeling to decrease computer operations. Use of PCA to reduce spectral dimensions was also implemented in other studies on MIR spectra (e.g., Soyeurt et al., 2010; Dagnachew et al., 2013a,b; Bonfatti et al., 2017). We selected a total of 12 PC, representing 96% of the information (i.e., total variance) contained in spectral wavenumbers. Other authors (e.g., Dagnachew et al., 2013a; Bonfatti et al., 2017) indicated that even less than 1% loss of total variation could lead to loss of relevant information. However, in the present study, we considered 12 PC as an optimum between dimension reduction and model accuracy. Indeed, although 4% of the spectral variance was lost, the average correlation between observed and predicted spectral wavenumbers did not substantially increase when adding extra PC [i.e., the correlation increased by only 0.01 and 0.004 when increasing the number of PC from 12 to 23 (i.e., 99% of total variance) for modeling and scenarios, respectively; Figure 3-2]. Bonfatti et al. (2017) did not mention using spectral pretreatment, in contrast to the present study. When using raw spectra, a larger part of the variability may be noise and the remaining percentages may be interesting spectral variability. This might explain why even a 1% loss of spectral variation might lead to loss of relevant information.

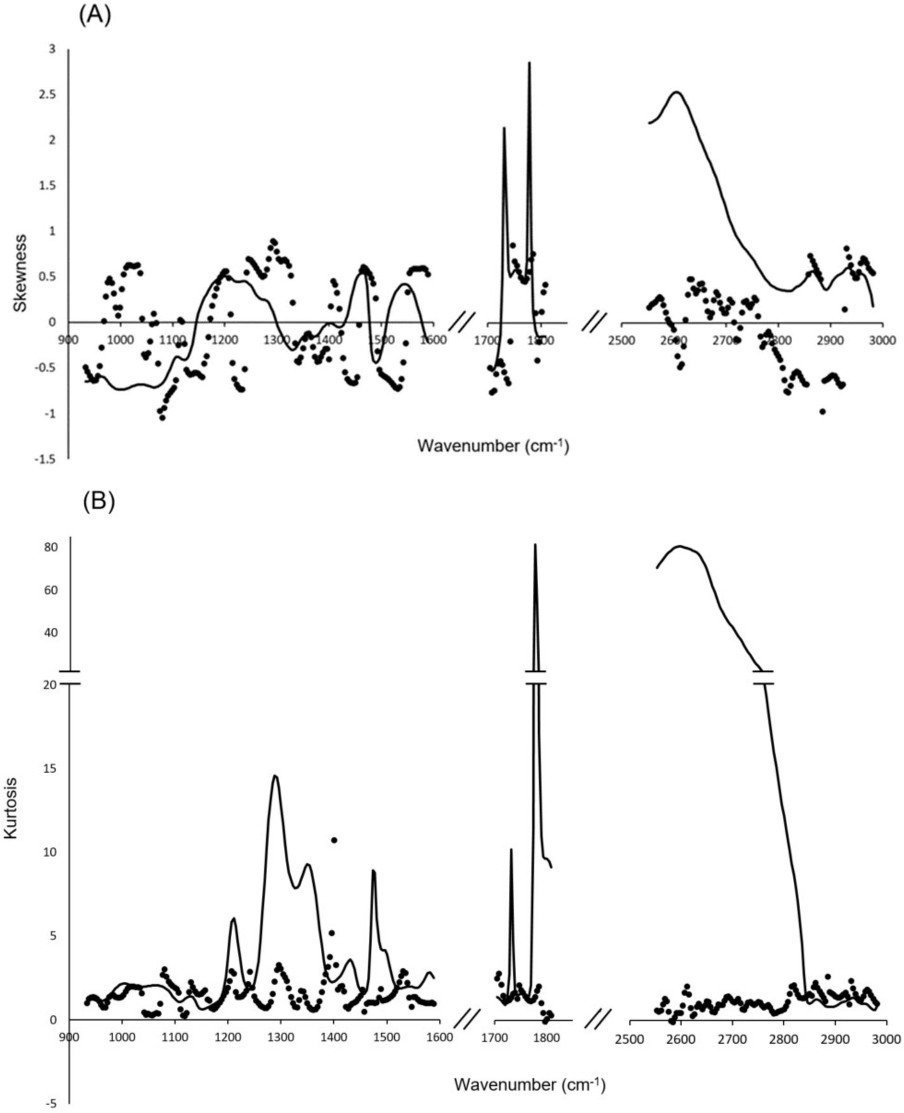

Figure 3-1. (A) Skewness for each of the 311 selected spectral wavenumbers for the modeling dataset. The continuous line represents the raw spectra; circles represent the derived spectra. (B) Excess kurtosis for each of the 311 selected spectral wavenumbers for the modeling dataset. The continuous line represents the raw spectra; circles represent the derived spectra.

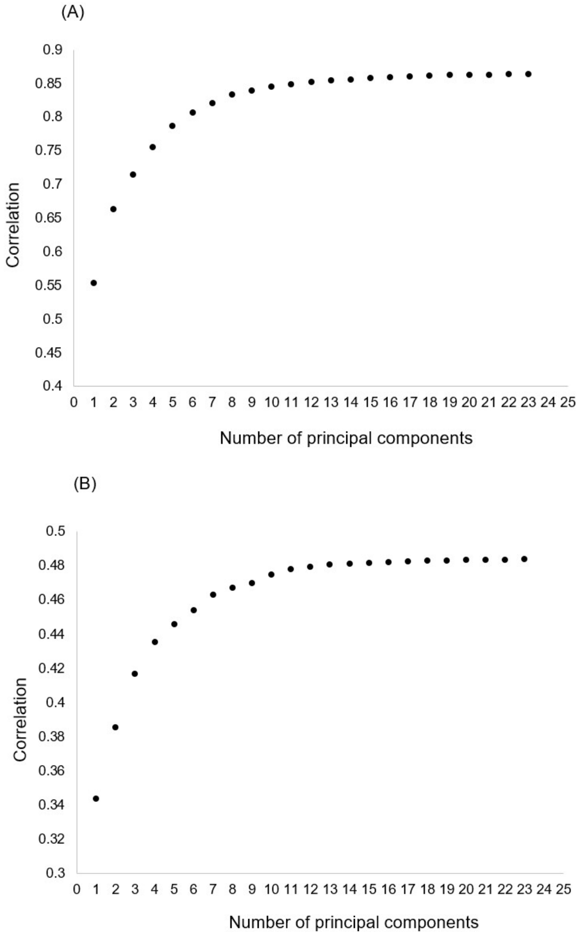

Figure 3-2. Average correlation between observed and predicted values for the 311 spectral wavenumbers according to the number of principal components selected (A) for the modeling dataset and (B) for the 4 scenarios datasets together.

3.2. Variance components

Figure 3-3 represents the percentage of total spectral variation retained by the model explained by the genetic, permanent environment, HTDr, and residual effects. Genetic and residual effects explained on average respectively 37% and 41% of the total variation for 79% of the wavenumbers (i.e., from 933 to 1,589 cm^{-1}, 1,704 to 1,786 cm^{-1} and 2,777 to 2,981 cm^{-1}) while permanent environment and HTDr explained on average 10% and 12% for the same regions, respectively. This general pattern was observed in other studies (Wang et al., 2016; Lainé et al., 2017).

In the present study, the HTDr effect was the most important effect in the spectral regions from 1,790 to 1,809 cm^{-1} and 2,553 to 2,773 cm^{-1}. Similarly, in the study of

Dagnachew et al. (2013a) on dairy goat milk spectra, HTDr also had a high variance ratio between 2,400 and 2,800 cm^{-1} and around 1771 cm^{-1}. Wang et al. (2016) observed a lower variance ratio explained by the genetic effect around 2,400 cm^{-1} to 2,800 cm^{-1}, similarly to the present study. Wang et al. (2016) and Lainé et al. (2017) found a lower proportion of variance explained by the genetic effect around 1,300 to 1,500 cm^{-1}, but this was not observed in our study.

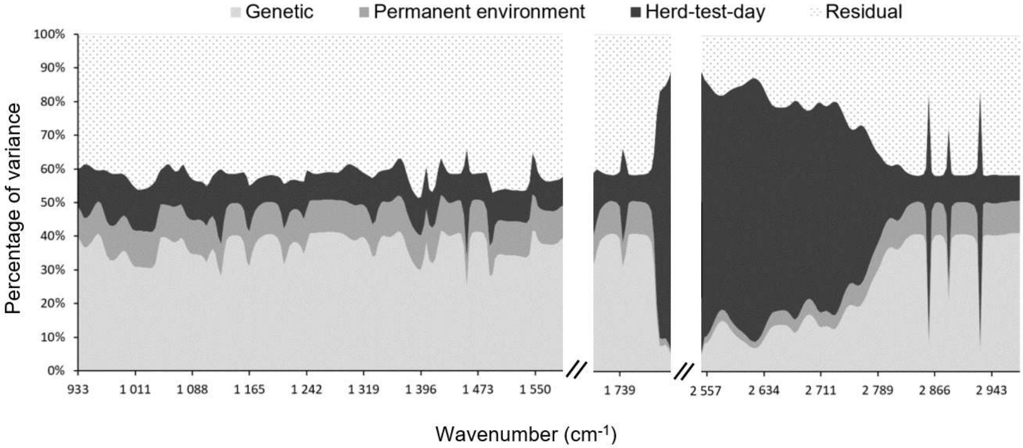

Figure 3-3. Percentage of total variance of spectral wavenumbers retained by the model explained by genetic, permanent environmental, herd-test-day and residual effects.

3.3. Predictions of spectral wavenumbers

The average correlation between observed and modeled values of the spectral wavenumbers was 0.85 for the modeling dataset, the minimum correlation value was 0.67 and the maximum was 0.95 (Figure 3-4). Regarding the scenarios, on average correlations were the highest for scenario 3 and the lowest for scenario 4 (0.62 and 0.36, respectively), with scenarios 1 and 2 in between (0.56 and 0.37, respectively, Table 3-2). We observed low correlations for spectral regions from 1,790 to 1,809 cm^{-1} and from 2,553 to 2,773 cm^{-1} for scenarios 1, 2 and 4 (Figure 3-5). When removing these regions, the 1st scenario (i.e., prediction of future records) had the highest average correlation (r = 0.63). Comparing this with Figure 3-3, these results seemed logical because the variance was mainly explained by the HTDr effect in these regions. The high proportion of variance explained by HTDr suggests that the wavenumbers are influenced by test-day factors, such as daily changes of feed, climatic conditions, or laboratory environment. Therefore, when the HTDr effect is not known (i.e., scenarios 1, 2, and 4), wavenumbers predictions for this region have low accuracy. According to Socrates (2001), these regions are not very associated with main chemical information like fat, protein, or lactose content, but these could be associated with minor components.

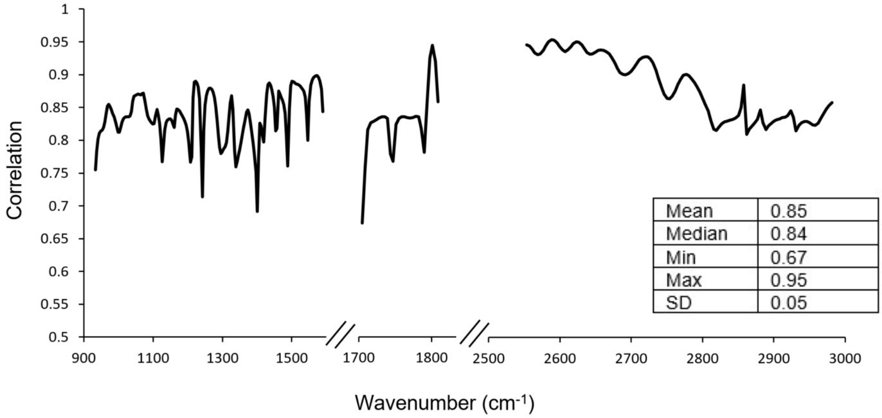

Figure 3-4. Correlations between observed and predicted values for the 311 spectral wavenumbers for the modeling dataset. Min = minimum; Max = maximum.

Table 3-2. Descriptive statistics of correlations between observed and predicted values for the 311 spectral wavenumbers for the 4 scenarios

Scenario	Mean	Median	Minimum	Maximum	SD
1	0.56	0.62	0.06	0.72	0.15
2	0.37	0.37	0.00	0.61	0.13
3	0.62	0.59	0.30	0.89	0.12
4	0.36	0.38	0.05	0.63	0.14

To our knowledge, no other authors studied the predictability of spectral wavenumbers, so comparison with the existing literature is difficult. However, Dagnachew et al. (2013a) showed that for goat milk spectra PC scores for future records could be predicted with reasonable accuracy (correlations between observed and predicted PC scores ranged between 0.48 and 0.75 for the first 7 PC). In the present study, correlations between observed and predicted PC scores for the 12 retained PC varied from 0.83 to 0.98 for the modeling dataset. Correlations between observed and predicted PC scores could not be calculated for the scenarios, because observed PC scores were not available as PCA was applied on modeling spectra only. The interpretation of individual spectral wavenumbers correlations and variance ratio variability is complicated because milk MIR spectra represent a combination of many different molecules in milk (Soyeurt et al., 2010). Not all wavenumbers have the same contribution in the prediction of specific traits and variation in spectral wavenumbers predictions might affect differently MIR traits predictions. For instance, chemical bonds that include nitrogen molecules (e.g., N-H, C-N) are specific to protein but less interesting to predict milk fat content (Socrates, 2001). Hence, comparing MIR traits predicted on modeled MIR spectra would be relevant.

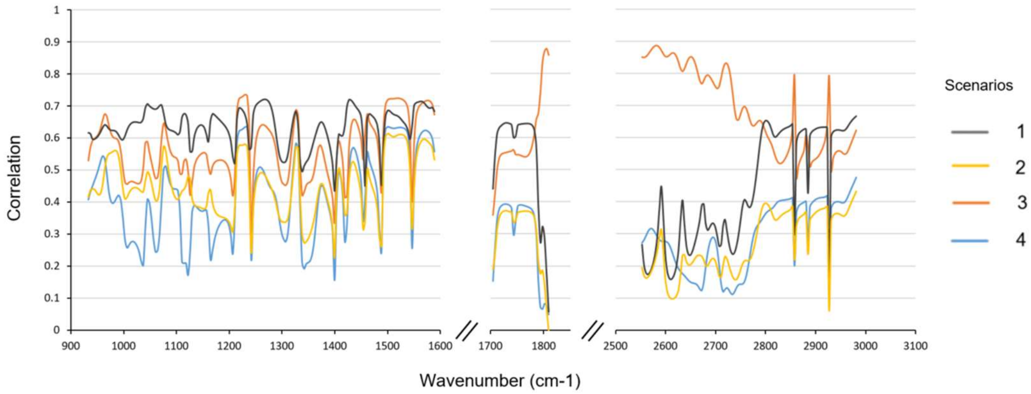

Figure 3-5. Correlations between observed and predicted values for the 311 spectral wavenumbers for the 4 scenarios.

3.4. Predictions of MIR traits

One interest of the prediction of MIR traits on the modeled spectra was to see if the accuracy did not decrease compared to the prediction of spectral wavenumbers and if predicted phenotype values were in an acceptable range. The left side of Table 3-3 presents correlations between predictions of milk fat, protein, and lactose contents from observed or modeled spectra. Correlations were relatively high for the modeling dataset as they were above 0.80. Regarding the scenarios, the first one had the highest correlation values for fat and lactose content (0.63 for both traits), and scenario 3 had the highest values for protein content (0.73). Scenarios 2 and 4 had the lowest values for the three milk components (i.e., 0.36, 0.62, 0.46 and 0.40, 0.64, and 0.32 for fat, protein, and lactose contents, respectively). Compared to fat and lactose contents, correlations for protein content showed a lower overall variation for the four scenarios. Globally, correlations for fat, protein and lactose contents varied in accordance with the evolution of correlations for spectral wavenumbers in the regions 933 to 1,589 cm^{-1},1,704 to 1,786 cm^{-1} and 2,777 to 2,981 cm^{-1} (Figure 3-5). These spectral regions are associated with molecular functional groups belonging to major milk components like lipids, proteins or carbohydrates (Socrates, 2001; Iñón et al., 2004; Dagnachew et al., 2013a). Milk fat is mainly associated with two spectral regions where the carbon-hydrogen groups (C–H) and the carbonyl groups (C=O) of milk fat absorb, that is to say at 2,873 and 1,747 cm^{-1}, respectively (Socrates, 2001; Iñón et al., 2004). The region around 1,100 cm^{-1} is associated with lactose content (Picque et al., 1993). MIR regions located between 1,200 to 1,450 cm^{-1} and 1,500 to 1,600 cm^{-1} correspond to protein content (Sivakesava and Irudayaraj, 2002). Superior accuracy for wavenumbers and MIR traits predictions for scenario 1 and 3 were justified by the higher number of known effects in the model compared to scenarios 2 and 4. We expected that scenario 4 would produce the least accurate results given the little number of known effects included in the model. However, we noticed that scenario 4 marginally outperformed scenario 2 for some wavenumbers and for the prediction of fat and protein contents, even though HTY and HTM effects were unknown. This might partly be the consequence of the different herds and cows and the different number of observations between these two scenarios. Indeed, scenario 4 was composed of different cows and herds from those in the modeling set. In contrast, scenarios 1, 2, and 3 were partially connected to the modeling set and between them because they were composed of similar herds (i.e., known HTY and HTM), though they were of different sizes. Unfortunately, we are unable to verify this hypothesis with certainty due to the structure of the datasets.

Table 3-3. Left side: correlations between fat, protein, and lactose contents (g/dL of milk) predicted from observed spectra vs. from spectra modeled from a test-day model (TDM); right side: correlations between fat, protein, and lactose contents (g/dL of milk) predicted from observed spectra vs. modeled directly using a TDM[1]

	Traits predicted from observed spectra vs. from modeled spectra from TDM			Traits predicted from observed spectra vs. modeled directly from TDM		
Item	Fat	Protein	Lactose	Fat	Protein	Lactose
Modeling set	0.83	0.89	0.83	0.83	0.90	0.86
Scenario 1	0.63	0.68	0.63	0.63	0.68	0.59
Scenario 2	0.36	0.62	0.46	0.37	0.60	0.37
Scenario 3	0.53	0.73	0.46	0.53	0.72	0.46
Scenario 4	0.40	0.64	0.32	0.40	0.64	0.31

[1] Results are presented for the modeling dataset and the four scenarios.

For the 3 studied milk components, comparing the left and right parts of Table 3-3, correlations between traits predicted from observed spectra vs. traits predicted from spectra modeled from a TDM were very similar to correlations between traits predicted from observed spectra vs. traits predicted directly from a TDM. This implies that, for these traits, using prediction equations on MIR spectra predicted from a TDM did as good as modeling traits directly using a TDM. For future research, it might also be useful to consider other traits such as fine milk components that are less correlated with the major sources of variation of the spectra (Bonfatti et al., 2017).

Table 3-4 displays mean values and standard deviations for fat, protein, and lactose contents predicted from observed or modeled spectra and RMSE between these traits predicted from observed vs. modeled spectra. For the modeling dataset and the four scenarios, mean values predicted from modeled spectra were very similar to mean values predicted from observed spectra, but standard deviations were smaller. It shows the ability of the model to predict values in a similar range as traits predicted from observed spectra on average (i.e., close to reality), but with lower variability. In all situations, the RMSE was the highest for milk fat content, followed by protein content and then lactose content. Values for RMSE were in similar ranges compared to the standard deviation for the traits predicted from the observed spectra for the modeling set. This indicates that the error is relatively large compared to the expected variation of the observed traits in the population. RMSE was slightly lower for the first scenario (i.e., prediction of future records) compared with the 3 other cases, in line with globally higher correlations (Table 3-3). Even when knowing very little information about the cow, like for scenario 4 (i.e., only genetic and DIM effects were known), correlations for the 3 studied milk traits were still higher than zero and the RMSE did not increase sharply (Tables 3-3 and 3-4).

Table 3-4. Mean (SD in parentheses) for fat, protein, and lactose contents (g/dL of milk) predicted from observed and modeled spectra and root mean square error (RMSE) between fat, protein, and lactose contents predicted from observed vs. modeled spectra for the modeling dataset and the 4 scenarios

	Fat			Protein			Lactose		
	Mean obs. spectra[1]	Mean pred. spectra[2]	RMSE	Mean obs. spectra	Mean pred. spectra	RMSE	Mean obs. spectra	Mean pred. spectra	RMSE
Model.	4.01 (0.68)	4.01 (0.53)	0.38	3.41 (0.38)	3.42 (0.32)	0.17	4.83 (0.17)	4.82 (0.14)	0.10
Scen. 1	4.30 (0.70)	4.22 (0.46)	0.55	3.68 (0.36)	3.67 (0.26)	0.27	4.72 (0.17)	4.75 (0.12)	0.14
Scen. 2	3.94 (0.64)	3.89 (0.43)	0.63	3.36 (0.33)	3.34 (0.28)	0.27	4.77 (0.17)	4.80 (0.11)	0.16
Scen. 3	4.00 (0.74)	3.96 (0.50)	0.64	3.35 (0.42)	3.34 (0.33)	0.29	4.84 (0.19)	4.85 (0.11)	0.17
Scen. 4	4.01 (0.71)	4.04 (0.31)	0.65	3.39 (0.40)	3.44 (0.25)	0.31	4.79 (0.18)	4.79 (0.06)	0.17

[1] Mean of the trait predicted from the observed spectra
[2] Mean of the trait predicted from the modeled spectra

3.5. Practical use for herd management

The objective of this study was to test the ability of a test-day mixed model to predict milk MIR spectra with the perspective of herd management. Today numerous traits are predicted from milk MIR spectra (Gengler, 2016). One advantage of modeling the MIR spectrum compared with modeling MIR traits directly is that a limited number of models are required (i.e., one for each PC) instead of developing one model separately for each of the existing MIR traits. Then the existing calibration equations for MIR traits can be applied to the modeled spectra to predict the different phenotypes without additional calculations. Hence, it reduces the workload associated with the development of several models for the numerous MIR traits, as well as computation time and resources [e.g., there is no need to estimate variance components (REML) for each trait, which is computationally demanding]. Also, the information contained in the MIR spectrum is richer than the information in a few predicted traits and could be used as such. The information resulting from MIR spectra modeling could benefit farmers with their cow and herd management in many ways. Any unexpected variation from a usual pattern may indicate a problem. Consequently, the difference between observed and modeled (i.e., expected) spectra and resulting predicted MIR traits could be a way to detect problems like metabolic disorders, feeding problems, or mastitis at the cow or herd level. It would help saving medical treatment costs and avoid loss of money caused by decreasing productivity or degradation of health. For example, changes in milk fat and protein ratio compared to an expected pattern can be used to detect the risk of metabolic disorders in lactating cows such as ketosis (Duffield et al., 1997); or unexpected changes in milk lactose, protein and minerals content can be used to detect mastitis (Hamann and Krömker, 1997). In addition, MIR spectra modeling could be used for simulations through adjustments of factors in the model (e.g., to predict the results of diet or genetic changes), for prediction of missing or future records or prediction of heifers productive potential. Also, as suggested by Mayeres et al. (2004), Koivula et al. (2007) and Bastin et al. (2009), studying any deviations in the solutions of the HTDr effect when running the model after each milk recording would enable the identification of herd-specific phenomena and management problems such as feeding problem, or seasonal difficulties. Following further research on the topic, also to reach sufficient accuracy, MIR spectra modeling and resulting tools could be implemented in practice into official milk recording systems.

The results presented in this study showed that the model had a moderate accuracy. Consequently, its use in the current state would be limited for instance for imputation of missing data or rough estimations for simulations when precise data is not necessary. However, several improvements of the studied model are possible, and needed, to use it for more precise applications such as the detection of cow or herd problems. Adding extra effects in the model such as the age of the cow or the gestation stage could be necessary, but such data are not always easily available for all cows and herds. Besides, the studied model focused only on first-lactation cows, but, in a multi-lactation model, the herd effects and permanent environmental effects would be more precise as a consequence of having more individual data. Extending the model to a random regression model (e.g., using Legendre polynomials) might also improve predictions. There is also a need to explore further if capturing more spectral variation or more diverse variation would improve the accuracy, for instance using an alternative selection of PC or investigating other methods for the reduction of spectral variables. Moreover, utilization of co-variance between PC in REML and BLUP though multi-trait analysis would possibly improve the accuracy. Even though PC are phenotypically orthogonal, they have genetic, permanent environmental, HTDr and residual co-variance structures (Dagnachew et al., 2013a,b; Bonfatti et al., 2017; Belay et al., 2017). But such multi-trait models are currently very computationally demanding on large datasets and could impede the implementation, which is a reason why we preferred single trait models in this study. All these possible improvements need to be elaborated further, also to avoid limitations of the practical use of the model.

4. Conclusions

This study demonstrated the moderate ability to predict milk MIR spectra using a test-day mixed model. The prediction accuracy varied for the different spectral wavenumbers and depended on the effects known in different situations. This influenced the prediction accuracy of related MIR traits. More research is required to improve the accuracy of predictions for potential promising applications for dairy herd management.

5. Acknowledgments

The first author acknowledges the support of the National Fund for Scientific Research (FRS-FNRS, Brussels, Belgium) through a grant scholarship. The authors acknowledge the milk laboratory Comité du Lait (Battice, Belgium) and the Walloon Breeding Association for providing spectral data. They also acknowledge for computational resources: the Plateau de Calcul et Modélisation Informatique (CAMI), partly supported by the FRS-FNRS under Grant n° T.0095.19 (PDR "DEEPSELECT"), of the University of Liège, Gembloux Agro-Bio Tech, and the Consortium des Équipements de Calcul Intensif (CÉCI), also supported by the FRS.-FNRS under Grant n° 2.5020.11. Additional support for the work was provided by the FRS-FNRS under Grant n° J.0174.18 (CDR "PREDICT 2"). The authors have not stated any conflicts of interest.

6. References

Bastin, C., L. Laloux, A. Gillon, F. Miglior, H. Soyeurt, H. Hammami, C. Bertozzi, and N. Gengler. 2009. Modeling milk urea of Walloon dairy cows in management perspectives. J. Dairy Sci. 92:3529–3540. https://doi:10.3168/jds.2008-1904.

Belay, T.K., B.S. Dagnachew, Z.M. Kowalski, and T. Ådnøy. 2017. An attempt at predicting blood β-hydroxybutyrate from Fourier-transform mid-infrared spectra of milk using multivariate mixed models in Polish dairy cattle. J. Dairy Sci. 100:6312–6326. https://doi:10.3168/jds.2016-12252.

Bittante, G., and A. Cecchinato. 2013. Genetic analysis of the Fourier-transform infrared spectra of bovine milk with emphasis on individual wavelengths related to specific chemical bonds. J. Dairy Sci. 96:5991–6006. https://doi.org/10.3168/jds.2013-6583.

Bonfatti, V., G. Di Martino, and P. Carnier. 2011. Effectiveness of mid-infrared spectroscopy for the prediction of detailed protein composition and contents of protein genetic variants of individual milk of Simmental cows. J. Dairy Sci. 94:5776–5785. https://doi.org/ 10.3168/jds.2011-4401.

Bonfatti, V., D. Vicario, L. Degano, A. Lugo, and P. Carnier. 2017. Comparison between direct and indirect methods for exploiting Fourier transform spectral information in estimation of breeding values for fine composition and technological properties of milk. J. Dairy Sci. 100:2057–2067. https://doi.org/10.3168/jds.2016-11951.

Caccamo, M., R.F. Veerkamp, G. De Jong, M.H. Pool, R. Petriglieri, and G. Licitra. 2008. Variance components for test-day milk, fat, and protein yield, and somatic cell score for analyzing management information. J. Dairy Sci. 91:3268–3276. https://doi.org/10.3168/jds.2007-0805.

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.

Caredda, M., M. Addis, I. Ibba, R. Leardi, M.F. Scintu, G. Piredda, and G. Sanna. 2017. Building of prediction models by using Mid-Infrared spectroscopy and fatty acid profile to discriminate the geographical origin of sheep milk. LWT - Food Sci. Technol. 75:131–136. https://doi.org/10.1016/j.lwt.2016.08.053.

Dagnachew, B.S., A. Kohler, and T. Ådnøy. 2013a. Genetic and environmental information in goat milk Fourier transform infrared spectra. J. Dairy Sci. 96:3973–3985. https://doi.org/10.3168/jds.2012-5972.

Dagnachew, B.S., T.H.E. Meuwissen, and T. Ådnøy. 2013b. Genetic components of milk Fourier-transform infrared spectra used to predict breeding values for milk composition and quality traits in dairy goats. J. Dairy Sci. 96:5933–5942. https://doi.org/10.3168/jds.2012-6068.

De Marchi, M., M. Penasa, A. Cecchinato, M. Mele, P. Secchiari, and G. Bittante. 2011. Effectiveness of mid-infrared spectroscopy to predict fatty acid composition of Brown Swiss bovine milk. Animal 5:1653–1658. https://doi:10.1017/S1751731111000747.

De Roos, A.P.W., A.G.F. Harbers, and G. De Jong. 2004. Random herd curves in a test-day model for milk, fat, and protein production of dairy cattle in the Netherlands. J. Dairy Sci. 87:2693–2701. https://doi:10.3168/jds.S0022-0302(04)73396-2.

Duffield, T.F., D.F. Kelton, K.E. Leslie, K.D. Lissemore, and J.H. Lumsden. 1997. Use of test day milk fat and milk protein to detect subclinical ketosis in dairy cattle in Ontario. Can. Vet. J. 38:713–718.

Ferragina, A., C. Cipolat-Gotet, A. Cecchinato, and G. Bittante. 2013. The use of Fourier-transform infrared spectroscopy to predict cheese yield and nutrient recovery or whey loss traits from unprocessed bovine milk samples. J. Dairy Sci. 96:7980–7990. https://doi.org/10.3168/jds.2013-7036.

Gengler, N., H. Soyeurt, F. Dehareng, C. Bastin, F. Colinet, H. Hammami, M.-L. Vanrobays, A. Lainé, S. Vanderick, C. Grelet, A. Vanlierde, E. Froidmont, and P. Dardenne. 2016. Capitalizing on fine milk composition for breeding and management of dairy cows. J. Dairy Sci. 99:4071–4079. https://doi.org/10.3168/jds.2015-10140.

Gillon, A., S. Abras, P. Mayeres, C. Bertozzi, and N. Gengler. 2010. Adding value to test-day data by using modified best prediction method. Pages 171-178 in Proc. ICAR 37th Annu. Meet., Riga, Latvia, ICAR, Rome, Italy.

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

Koivula, M., J.I. Nousiainen, J. Nousiainen, and E.A. Mäntysaari. 2007. Use of herd solutions from a random regression test-day model for diagnostic dairy herd management. J. Dairy Sci. 90:2563–2568. https://doi.org/10.3168/jds.2006-517.

Hamann, J., and V. Krömker. 1997. Potential of specific milk composition variables for cow health management. Livest. Prod. Sci. 48:201–208.

Hammami, H., B. Rekik, H. Soyeurt, C. Bastin, E. Bay, J. Stoll, and N. Gengler. 2009. Accessing genotype by environment interaction using within- and across-country test-day random regression sire models. J. Anim. Breed. Genet. 126:366–377. https://doi.org/10.1111/j.1439-0388.2008.00794.x.

ICAR. 2017. Section 2 - Guidelines for Dairy Cattle Milk Recording. Int. Comm. Anim. Record., Rome, Italy.

Iñón, F.A., S. Garrigues, and M. De La Guardia. 2004. Nutritional parameters of commercially available milk samples by FTIR and chemometric techniques. Anal. Chim. Acta 513:401–412. https://doi.org/10.1016/j.aca.2004.03.014.

Kim, H.-Y. 2013. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Restor. Dent. Endod. 38:52-54. https://doi.org/10.5395/rde.2013.38.1.52.

Lainé, A., C. Bastin, C. Grelet, H. Hammami, F.G. Colinet, L.M. Dale, A. Gillon, J. Vandenplas, F. Dehareng, and N. Gengler. 2017. Assessing the effect of pregnancy stage on milk composition of dairy cows using mid-infrared spectra. J. Dairy Sci. 100:2863–2876. https://doi.org/10.3168/jds.2016-11736.

Leclercq, G., N. Gengler, H. Soyeurt, and C. Bastin. 2013. Genetic variability of the mid-infrared prediction of lactoferrin content in milk for Walloon Holstein first-parity cows. Livest. Sci. 151:158–162. https://doi.org/10.1016/j.livsci.2012.11.002.

Mayeres, P., J. Stoll, R. Reents, and N. Gengler. 2002. Alternative modeling of fixed effects in test-day models to increase their usefulness for management decisions. Interbull Bull. 29:128–132.

Mayeres, P., J. Stoll, J. Bormann, R. Reents, and N. Gengler. 2004. Prediction of daily milk, fat, and protein production by a random regression test-day model. J. Dairy Sci. 87:1925–1933. https://doi.org/10.3168/jds.s0022-0302(04)73351-2.

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

Mineur, A., A. Köck, C. Grelet, N. Gengler, C. Egger-Danner, and J. Sölkner. 2017. First results in the use of milk mid-infrared spectra in the detection of lameness in Austrian dairy cows. Agric. Conspec. Sci. 82:163–166.

Misztal, I. 2018. BLUPF90 Family of Programs. Accessed Jan. 2018. http://nce.ads.uga.edu/wiki/doku.php/.

Owen, A.J. 1995. Uses of derivative spectroscopy, UV-visible spectroscopy, application note, Agilent Technologies, Santa Clara, CA.

Palm, R. 1998. L'analyse en composantes principales: principes et applications. Faculté universitaire des Sciences agronomiques de Gembloux.

Picque, D., D. Lefier, R. Grappin, and G. Corrieu. 1993. Monitoring of fermentation by infrared spectrometry. Alcoholic and lactic fermentations. Anal. Chim. Acta 279:67–72. https://doi.org/10.1016/0003-2670(93)85067-T.

R Core Team. 2017. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. Accessed Dec. 2017. https://www.R-project.org/.

Rinnan, Å., L. Nørgaard, and F. Van Den Berg. 2009. Data Pre-processing. Pages 29-50 in Infrared spectroscopy for food quality analysis and control. D.W. Sun, ed Academic Press, New York, NY.

Rovere, G., G. de los Campos, R.J. Tempelman, A.I. Vazquez, F. Miglior, F. Schenkel, A. Cecchinato, G. Bittante, H. Toledo-Alvarado, and A. Fleming. 2019. A landscape of the heritability of Fourier-transform infrared spectral wavelengths of milk samples by parity and lactation stage in Holstein cows. J. Dairy Sci. 102:1354–1363. https://doi.org/10.3168/jds.2018-15109.

Scampicchio, M., D. Eisenstecken, L. De Benedictis, C. Capici, D. Ballabio, T. Mimmo, P. Robatscher, L. Kerschbaumer, M. Oberhuber, A. Kaser, C.W. Huck, and S. Cesco. 2016. Multi-method approach to trace the geographical origin of alpine milk: a case study of Tyrol region. Food Anal. Methods 9:1262–1273. https://doi.org/10.1007/s12161-015-0308-2.

Shallue, C.J., J. Lee, J. Antognini, J. Sohl-Dickstein, R. Frostig, and G.E. Dahl. 2019. Measuring the effects of data parallelism on neural network training. J. Mach. Learn. Res. 20:1–49.

Sivakesava, S., and J. Irudayaraj. 2002. Rapid determination of tetracycline in milk by FT-MIR and FT-NIR spectroscopy. J. Dairy Sci. 85:487–493. https://doi.org/10.3168/jds.s0022-0302(02)74099-x.

Socrates, G. 2001. Infrared and Raman characteristic group frequencies contents: Tables and charts. 3rd ed. Wiley, Chichester, NY.

Soyeurt, H., D. Bruwier, J.-M. Romnee, N. Gengler, C. Bertozzi, D. Veselko, and P. Dardenne. 2009. Potential estimation of major mineral contents in cow milk using mid-infrared spectrometry. J. Dairy Sci. 92:2444–2454. https://doi.org/10.3168/jds.2008-1734.

Soyeurt, H., I. Misztal, and N. Gengler. 2010. Genetic variability of milk components based on mid-infrared spectral data. J. Dairy Sci. 93:1722–1728. https://doi.org/10.3168/jds.2009-2614.

Soyeurt, H., F. Dehareng, N. Gengler, S. McParland, E. Wall, D.P. Berry, M. Coffey, and P. Dardenne. 2011. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. J. Dairy Sci. 94:1657–1667. https://doi.org/10.3168/jds.2010-3408.

Soyeurt, H., E. Froidmont, I. Dufrasne, D. Hailemariam, Z. Wang, C. Bertozzi, F.G. Colinet, F. Dehareng, and N. Gengler. 2019. Contribution of milk mid-infrared spectrum to improve the accuracy of test-day body weight predicted from stage, lactation number, month of test and milk yield. Livestock Science 227:82-89.

Tsuruta, S., I. Misztal, and I. Strandén. 2001. Use of the preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications. J. Anim. Sci. 79:1166–1172. https://doi.org/10.2527/2001.7951166x.

Van de Voort, FR. 1992. How does material resource adequacy affect innovation project performance? A meta-analysis. J. Prod. Innov. Manag. 34:842–863. https://doi.org/10.1111/jpim.12368.

Vanlierde, A., H. Soyeurt, N. Gengler, F.G. Colinet, E. Froidmont, M. Kreuzer, F. Grandl, M. Bell, P. Lund, D.W. Olijhoek, M. Eugène, C. Martin, B. Kuhla, and F. Dehareng. 2018. Short communication: development of an equation for estimating methane emissions of dairy cows from milk Fourier transform mid-infrared spectra by using reference data obtained exclusively from respiration chambers. J. Dairy Sci. 101:7618–7624. https://doi.org/10.3168/jds.2018-14472.

Visentin, G., A. McDermott, S. McParland, D.P. Berry, O.A. Kenny, A. Brodkorb, M.A. Fenelon, and M. De Marchi. 2015. Prediction of bovine milk technological traits from mid-infrared spectroscopy analysis in dairy cows. J. Dairy Sci. 98:6620–6629. https://doi.org/10.3168/jds.2015-9323.

Wang, Q., A. Hulzebosch, and H. Bovenhuis. 2016. Genetic and environmental variation in bovine milk infrared spectra. J. Dairy Sci. 99:6793–6803. https://doi.org/10.3168/jds.2015-10488.

Wiggans, G.R., and M.E. Goddard. 1997. A computationally feasible test day model for genetic evaluation of yield traits in the United States. J. Dairy Sci. 80:1795–1800. https://doi.org/10.3168/jds.S0022-0302(97)76113-7.

Zaalberg, R.M., N. Shetty, L. Janss, and A.J. Buitenhuis. 2019. Genetic analysis of Fourier transform infrared milk spectra in Danish Holstein and Danish Jersey. J. Dairy Sci. 102:503–510. https://doi.org/10.3168/jds.2018-14464.

4

Relationships between milk mid-infrared predicted gastro-enteric methane production and the technical and financial performance of commercial dairy herds

Adapted from: Delhez, P., B. Wyzen, A. Dalcq, F.G. Colinet, E. Reding, A. Vanlierde, F. Dehareng, N. Gengler, and H. Soyeurt. 2018. Relationships between milk mid-IR predicted gastro-enteric methane production and the technical and financial performance of commercial dairy herds. Animal 12:1981–1989.

https://doi.org/10.1017/S1751731117003378

Outline

This chapter illustrates the implementation of a MIR tool and the interest of combining MIR-predicted data with other data streams to provide more comprehensive information for decision-making. The research uncovers the relationships between MIR enteric methane (CH_4) production and technical and economic variables of commercial dairy herds in the Walloon region of Belgium. Enteric CH_4 is becoming an increasing concern because of its global warming potential and its importance in the environmental footprint of dairy production. The obtained results contribute to identifying practices associated with reduced lactating dairy cow CH_4 production in practice, and to assess the economic viability of herds emitting little CH_4 in a context of sustainable dairy farming. This could influence both farmers' approach to achieving CH_4 reduction and the ways in which policymakers could provide effective support, for example in the context of climate policies targeting agricultural emissions.

Abstract

Considering economic and environmental issues is important in ensuring the sustainability of dairy farms. The objective of this study was to investigate univariate relationships between lactating dairy cow gastro-enteric methane (CH_4) production predicted from milk mid-infrared spectra and technico-economic variables by the use of large scale and on-farm data. A total of 525,697 individual CH_4 predictions from milk mid-infrared spectra [MIR-CH_4 (g/day)] of milk samples collected on 206 farms during the Walloon milk recording scheme were used to create a MIR-CH_4 prediction for each herd and year (HYMIR-CH_4). These predictions were merged with dairy herd accounting data. This allowed a simultaneous study of HYMIR-CH_4 and 42 technical and economic variables for 1,024 herd and year records from 2007 to 2014. Pearson correlation coefficients (r) were used to assess significant relationships ($P < 0.05$). Low HYMIR-CH_4 was significantly associated with, amongst others, lower fat and protein corrected milk (FPCM) yield ($r = 0.18$), lower milk fat and protein content ($r = 0.38$ and 0.33, respectively), lower quantity of milk produced from forages ($r = 0.12$) and suboptimal reproduction and health performance (e.g. longer calving interval ($r = -0.21$) and higher culling rate ($r = -0.15$)). Concerning economic results, low HYMIR-CH_4 was significantly associated with lower gross margin per cow ($r = 0.19$) and per litre FPCM ($r = 0.09$). To conclude, this study suggested that low lactating dairy cow gastro-enteric CH_4 production tended to be associated with more extensive or suboptimal management practices, which could lead to lower profitability. The observed low correlations suggest complex interactions between variables due to the use of on-farm data with large variability in technical and management practices.

Keywords: dairy cow, methane production, economy, herd management, mid-infrared

1. Introduction

According to the Intergovernmental Panel on Climate Change (IPCC), agriculture and forestry are responsible for 24% of the greenhouse gases (GHG) emitted worldwide (Intergovernmental Panel on Climate Change, 2014). Cattle production represents the largest part of this, generating carbon dioxide, CH_4 and nitrous oxide. Gastro-enteric CH_4 from anaerobic fermentation of feed in the rumen and hindgut is the largest source of cattle GHG, representing 46.5% of total cattle emissions (FAO, 2013). In addition to environmental pollution, 2 to 12% of gross energy is lost through CH_4 emissions, which reduces the consumed energy available for the cow (Johnson and Johnson, 1995).

Understanding factors influencing gastro-enteric CH_4 production from dairy cows is an important step in reducing its impact. These factors have been widely studied and include: the level of dry matter intake (DMI); feed composition (e.g., starch, fiber, and lipid content); feed processing; herd management (e.g., breed, available grasslands); and genetics (e.g., Gerber et al., 2013; Hristov et al., 2013; Pickering et al., 2015). However, such studies were usually conducted using a small number of animals and farms. In addition, the effects of a limited number of factors at a time were usually studied under controlled conditions in experimental settings, with other factors being fixed (e.g., Brask et al., 2013; Hart et al., 2015; Hatew et al., 2016).

Beyond this, results on the relationship between gastro-enteric CH_4 and the economic performance of dairy herds are scarce in the research literature. The only economic aspects studied are the estimation of the cost of specific enteric CH_4 mitigation measures, such as the adoption of specific feeding strategies (e.g., Van Middelaar et al., 2014; Moraes et al., 2015). Studies analyzing the link between all GHG emissions at the farm level and economic results have also been carried out in several countries, but these did not focus on gastro-enteric CH_4 production only and assumed equal CH_4 emissions for all cows in the herd, often based on literature reference values or mechanistic models (e.g., Beukes et al., 2009; Thomassen et al., 2009; O'Brien et al., 2015). Given the economic difficulties encountered by dairy farmers (i.e., volatile milk price, increasing input prices), it would be useful to study such relationships in order to avoid profitability loss when global management strategies to mitigate CH_4 production are adopted.

A major issue for any study of gastro-enteric CH_4 at a herd level in a commercial environment is the difficulty in providing reasonable estimates of CH_4 production on a large scale. Recent advances have been reported on the use of milk mid-infrared (MIR) spectra and lactation stage-based predictions (Vanlierde et al., 2015), which have already been used successfully in a large-scale study on lactating dairy cows (Vanlierde et al., 2016). In this context, the objective of the present study was to investigate the univariate relationships between lactating dairy cow gastro-enteric CH_4 production (g/day) predicted from milk MIR spectra and technical and economic variables in commercial dairy herds. To achieve this objective, individual cow milk MIR CH_4 predictions were pooled for each herd and each year and associated with technical and economic variables from the corresponding herds and years. Correlations were used to study univariate relationships between milk MIR CH_4 and technico-economic variables.

2. Materials and methods

Operations of data editing, processing, and statistical analyses were implemented using SAS® 9.4 software (Statistical Analysis Software, SAS Institute Inc. Cary, NC, USA) and R 3.4.1.

2.1. Data

Two different datasets were merged to achieve the objectives of this study: the Walloon milk recording database and the dairy herd accounting database (containing economic and technical variables) both managed by the Walloon Breeding Association (awé groupe, Ciney, Belgium). The milk recording dataset includes milk production and milk composition data as well as milk MIR spectra expressed on individual cow and test-day level. Milk MIR spectra are obtained by the analysis of individual milk samples on MilkoScan FT6000 spectrometers (Foss, Hillerod, Denmark) at the 'Comité du Lait' laboratory (Battice, Belgium). Milk recording data were collected every 4 or 6 weeks. This study used 525,697 individual records from 40,314 dairy cows in 206 herds registered for the Walloon milk recording from January 2007 to April 2014. The accounting dataset contains technical and economic data expressed at herd and year level. All studied herds had Holstein as the main breed, at least 5 months of milk recording per accounting year, more than 10 cows per recording day, and more than 60% of cows registered for milk recording. Data were combined to obtain 1,024 records including technical and economic variables as well as a milk MIR CH_4 prediction calculated on a herd and year basis.

2.1.1. Technical variables

The list and descriptive statistics of technical variables, divided into four categories, are presented in Table 4-1. The category 'dairy herd characteristics' includes variables relative to the number of cows, calvings, and culling. 'Meat production' includes meat production per hectare of forage area. The category 'milk production' includes milk yield variables and milk composition variables. The category 'feed' relates to forage areas management and concentrate feed. All technical variables used in this study concern dairy cows only and do not take young stock into account, except the following variables: meat production per ha forage area, size of the forage area, mineral nitrogen on grasslands, number of livestock units per ha of forage area and percentage of conserved forages in the forage area. These apply to young animals from the dairy herd such as heifers and calves, as well as to the dairy cows.

Table 4-1. Description of the technical variables for the 1,024 dairy cow herd_year records

Variables	Mean	SD
Dairy herd characteristics		
No. dairy cows (n_cows)	84	36
No. dairy cows per unit labour supply (n_cows.labour)	53	18
Culling rate (%) (culling_percent)	25.7	7.7
Age at first calving (month) (first_calving)	29	2.8
Age of culled cows (month) (culling_age)	74.6	10.1
No. calvings per cow (n_calvings)	0.92	0.12
Calving interval (days) (CI)	436	28
Meat production		
Meat production per ha of forage area (kg/ha) (meat.ha)	328.6	96.4
Milk production		
Milk yield (l/cow per year) (milk)	7,253.4	1,007.4
Fat and protein corrected milk (FPCM) yield (l/cow per year) (FPCM)	7,534.7	1,023
Milk produced per ha of forage area (l/ha) (milk.ha)	9,978.7	2,389.2
Milk fat (%) (fat_percent)	4.26	0.16
Milk protein (%) (prot_percent)	3.47	0.09
Milk solids (kg/cow per year) (milk_solids)	560.5	76.7
Feed		
Size of the forage area (ha) (forage_area)	61.6	23.7
Mineral nitrogen on grasslands (kg/ha) (nitrogen_grass)	116	57
No. livestock units (LU) per ha forage area (LU/ha) (LU.ha)	2.89	0.59
Concentrate equivalents fed to dairy cows (kg/cow per year) (concentrate_feed)	1,863.9	584.8
Concentrate equivalents fed to dairy cows per litre FPCM (kg/l) (concentrate_feed.L)	0.245	0.064
Milk produced from forages (l/cow per year) (forage_milk)	3,801.8	1,027.4
Milk produced from fresh grass (l/cow per year) (grass_milk)	861.3	1,255.6
Percentage of conserved forages in the forage area (%) (conserved_forages)	53.5	12.3

2.1.2. Economic variables

The list and descriptive statistics of economic variables, divided in six categories, are presented in Table 4-2. All economic variables are part of two economic indicators: the gross margin and profit allocated to the dairy cows, expressed per cow

(€/cow per year), or per litre of fat and protein corrected milk (FPCM) (€/litre FPCM). They are calculated in the accounting database as:

$$\text{Gross margin} = \begin{aligned}[t] &\textit{(value of milk production}\\ &+ \textit{value of meat production from cows}\\ &+ \textit{value of meat production from calves)}\\ &- \textit{(purchased feed costs}\\ &+ \textit{proportional costs related to the forage area}\\ &+ \textit{herd costs)} \qquad (1)\end{aligned}$$

$$\text{Profit} = \textit{gross margin} - \textit{fixed costs} \qquad (2)$$

All variables included in equations (1) and (2) and in Table 4-2 concern dairy cows only, without taking young stock into account. The value of meat production from calves, which only takes into account the value of new-born calves, is also included in the dairy cow gross margin and profit as it is a production from the dairy cows. Herd costs include veterinary costs, medicines, litter, and miscellaneous costs. The gross margin and profit (per cow or litre FPCM) were corrected to remove the year effect caused by large fluctuations in milk and purchased feed prices over the years. This step is necessary in order to analyze data from different years together as one dataset. The two variables 'value of milk production' and 'purchased feed costs' (per cow or litre FPCM) were therefore corrected as follows:

$$\text{Corrected value of milk production} = \frac{Value\ of\ milk\ production}{Average\ annual\ price\ for\ 100\ l\ of\ milk} \times \textit{Average price (for all years) for 100 l of milk} \qquad (3)$$

$$\text{Corrected purchased feed costs} = \frac{Purchased\ feed\ costs}{Average\ annual\ price\ of\ the\ concentrate\ equivalent} \times \textit{Average price (for all years) of the concentrate equivalent} \qquad (4)$$

Corrected gross margin and profit were calculated by summing their respective components (equations (1) and (2)), including the two corrected variables (equations (3) and (4)).

Table 4-2. Description of the economic variables for the 1,024 dairy cow herd_year records

Variables	Mean	SD
Meat production		
Value of meat production from cows (€/cow per year) (cow_meat)	-143.8	73.1
Value of meat production from cows per litre fat and protein corrected milk (FPCM) (€/l) (cow_meat.L)	-0.019	0.010
Price of culled cows (€) (culled_cows_price)	632.1	141.5
Value of meat production from calves (€/cow per year) (calf_meat)	124.5	36.6
Value of meat production from calves per litre FPCM (€/l) (calf_meat.L)	0.017	0.005
Price of sold calves (€) (calves_price)	133.9	45.1
Milk production		
Value of milk production (€/cow per year) (milk_production)	2,450.1	363.6
Value of milk production per litre FPCM (€/l) (milk_production.L)	0.325	0.014
Feed costs		
Purchased feed costs (€/cow per year) (feed_costs)	519	163.7
Purchased feed costs per litre FPCM (€/l (feed_costs.L)	0.068	0.017
Other variable costs		
Variable costs related to the forage area (€/cow per year) (forage_area_costs)	208	74.1
Variable costs related to the forage area per litre FPCM (€/l) (forage_area_costs.L)	0.028	0.009
Herd costs (€/cow per year) (herd_costs)	161.2	56.1
Herd costs per litre FPCM (€/l) (herd_costs.L)	0.021	0.007
Fixed costs		
Fixed costs (€/cow per year) (fixed_costs)	900.7	219.9
Fixed costs per litre FPCM (€/l) (fixed_costs.L)	0.121	0.031
Economic indicators		
Gross margin (€/cow per year) (gross_margin)	1,542.6	287.9
Gross margin per litre FPCM (€/l) (gross_margin.L)	0.205	0.028
Profit (€/cow per year) (profit)	641.9	329.5
Profit per litre FPCM (€/l) (profit.L)	0.084	0.041

2.1.3. Herd- and year-based milk MIR methane predictions

Milk MIR CH_4 predictions from individual cows were pooled to obtain one prediction for each herd and year. First, CH_4 production (g/day) (MIR-CH_4) from individual lactating dairy cows was predicted from the recorded milk MIR spectra using the dependent-lactation-stage equation developed by Vanlierde et al. (2016). The equation has cross-validation R^2 and standard error of 0.70 and 70 g/day, respectively. We consider the equation to be appropriate for the milk MIR spectra used in this study as reference values in the calibration set came from similar Walloon herds (Vanlierde et al., 2016). MIR-CH_4 records were predicted only for cows with days in milk between 5 and 365. To discard potential outlier records, MIR-CH_4 predictions with a standardized Mahalanobis distance from the calibration set higher than 10 were discarded (Whitfield et al., 1987) as well as the 0.5% upper and 0.5% lower data. Second, one MIR-CH_4 prediction was calculated for each herd and each year (i.e., herd_year based MIR-CH_4; HYMIR-CH_4) in order to match technico-economic variables expressed on a herd and year basis. In this way, for each herd and year, the median of individual MIR-CH_4 records was calculated for each month of milk recording. The median was used instead of the mean because it is less affected by potential extreme values that could occur within a test day of milk recording. The median values were corrected for the year effect as follows:

Corrected MIR-CH_4 monthly median = *(MIR-CH_4 monthly median – annual MIR-CH_4 mean) + MIR-CH_4 mean for all years*

Then, for each year, the corrected herd_year median values were modeled using the GLM with herd and month as fixed effects. The least squares mean of the herd effect obtained was finally used as HYMIR-CH_4 for all further analyses. CH_4 production (g/day) was used as methane trait preferentially to CH_4 intensity (g/kg milk) in this study because it appeared that CH_4 intensity was very highly correlated with milk yield ($r = -0.94$ vs. $r = 0.11$ for HYMIR-CH_4), meaning that most studied relationships would have been dependent on milk yield.

2.2. *Relationships study*

The normality of the distribution for HYMIR-CH_4 and technico-economic variables was assessed by measuring skewness and kurtosis. Non-normal variables were transformed using a base 10 logarithm function. The relationships between each technico-economic variable and HYMIR-CH_4 were studied individually by using Pearson correlation coefficients (r). The correlation values were deemed significantly different from zero when P values were lower than 0.05. Relationships between HYMIR-CH_4 and technico-economic variables were tested for quadratic effect by using the generalized linear model with a quadratic term to ensure they were appropriate for correlation analysis requiring approximate linear relation.

3. Results

The mean for HYMIR-CH_4 was 476 g/day, with a SD of 15.8 g/day. The minimum and maximum values were 423 g/day and 526 g/day, respectively.

3.1. Technical variables

Correlations between HYMIR-CH_4 and technical variables are presented in Figure 4-1. Significant r ($P < 0.05$) ranged between |0.06| and |0.38|. Regarding herd characteristics, low HYMIR-CH_4 herds had a higher culling rate (r = -0.15), higher age at first calving (r = -0.09), lower age of culled cows (r = 0.06), longer calving interval (r = -0.21) and a lower number of calvings per cow (r = 0.10). Low HYMIR-CH_4 was also associated with a higher number of dairy cows ($\log_{10}$) (r = -0.13) and a higher number of dairy cows per unit labour supply (r = -0.10). Looking at meat production, the number of kilograms of meat produced per hectare of forage area did not change depending on the level of HYMIR-CH_4 ($P > 0.05$). Concerning milk production variables, milk fat and protein percentages were lower in low HYMIR-CH_4 herds; the correlation with HYMIR-CH_4 was the highest for these two variables (r = 0.38 and 0.33, respectively). Low HYMIR-CH_4 herds produced less milk (r = 0.11), less FPCM (r = 0.18), less milk solids (r = 0.20) and less milk per hectare of forage area (r = 0.08) compared to high HYMIR-CH_4 emitters. As for feed, the size of the forage area ($\log_{10}$) was larger (r = -0.14) and the number of livestock units per hectare was lower (r = 0.07) for herds producing less HYMIR-CH_4. Lower HYMIR-CH_4 production was associated with a lower quantity of mineral nitrogen applied on grasslands (r = 0.15) as well as a lower quantity of milk produced from forages (r = 0.12) and from fresh grass (r = 0.09). The quantity of concentrate equivalents fed per litre FPCM and the proportion of conserved forages in the forage area did not show any significant trend for low or high HYMIR-CH_4 herds ($P > 0.05$), though there was a trend ($0.05 < P < 0.10$) for higher concentrate equivalents per cow for high HYMIR-CH_4 herds.

3.2. Economic variables

Correlations between HYMIR-CH_4 and economic variables are presented in Figure 4-1. Significant r ($P < 0.05$) ranged between |0.07| and |0.19|. HYMIR-CH_4 was positively linked to gross margin per cow and per litre FPCM (r = 0.19 and 0.09, respectively). The profit per cow and per litre FPCM followed the same positive trend as the gross margin (r = 0.18 and 0.16, respectively). Regarding meat production, low HYMIR-CH_4 was associated with a lower value of meat production from cows expressed either per cow or per litre FPCM (r = 0.15 and 0.19, respectively). The value of meat production from calves per cow ($\log_{10}$), the price of culled cows, and the price of calves ($\log_{10}$) were also lower for herds producing less HYMIR-CH_4 (r = 0.07, 0.08 and 0.08, respectively). The value of milk production per cow was lower for herds producing less HYMIR-CH_4 (r = 0.16), but there was no significant relationship for the value of milk production per litre FPCM ($\log_{10}$) ($P > 0.05$). Feed costs per cow and per litre FPCM were not significantly different for herds with low or high HYMIR-CH_4 ($P > 0.05$). Regarding other costs, low HYMIR-CH_4 herds had lower variable costs related to the forage area per cow ($r = 0.13$). There was a trend ($0.05 < P < 0.10$) for lower variable costs related to the forage area per litre FPCM for low HYMIR-CH_4 herds. No relation was observed for the herd costs per cow ($P > 0.05$), but low HYMIR-CH_4 was associated with higher herd costs per litre FPCM ($\log_{10}$) ($r = -0.07$). Fixed costs per litre FPCM ($\log_{10}$) were higher for herds emitting less HYMIR-CH_4 ($r = -0.14$), but there was no significant relationship for fixed costs per cow ($P > 0.05$).

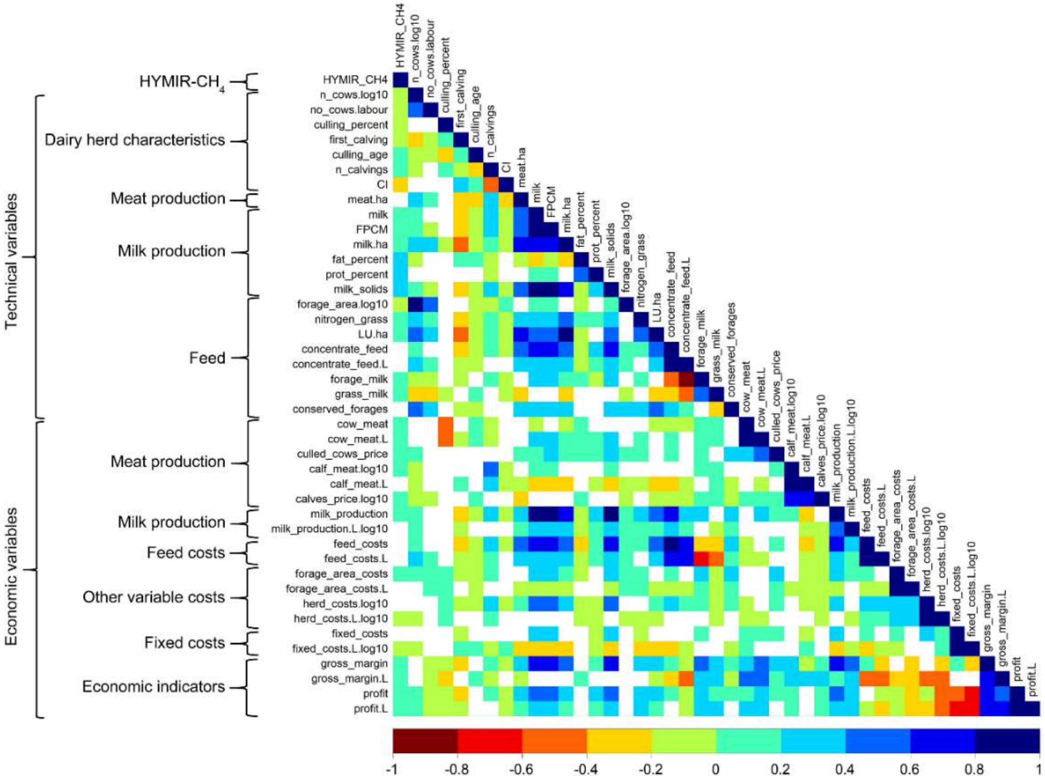

Figure 4-1. Heatmap of correlations between lactating dairy cow herd_year based milk MIR CH_4 production (HYMIR-CH4) and herd_year technico-economic variables. Abbreviations for technico-economic variables are provided in Table 4-1 and Table 4-2. Variable with names ending with '.log10' were log-transformed. White colour means non-significant relationship ($P > 0.05$).

4. Discussion

4.1. *Herd- and year-based milk MIR methane predictions*

HYMIR-CH_4 values were within the range of CH_4 measurements reported in the literature, varying from 300 to 600 g/day, though the SD was inferior to those reported in previous papers for individual cows, varying from 50 to 150 g/day (e.g., Hart et al., 2015; Hatew et al., 2016; Hristov et al., 2016). The reason is that HYMIR-CH_4 is expressed at the herd level, which reduces the individual cow variability existing within herds.

An important consideration about the obtained results was that the univariate relationships between HYMIR-CH_4 and technico-economic variables were weak as r were always lower than |0.38|. Despite this, significant relationships were highlighted and general assumptions could be made. The observed relationships could be direct or indirect and do not imply causality.

4.2. *Relationship with milk composition*

It was observed that, amongst technical variables, milk fat and protein content were the variables that best correlated with HYMIR-CH_4 ($r = 0.38$ and 0.33 respectively; Figure 4-1). An increase in milk fat and protein content was associated with higher HYMIR-CH_4 (Figure 4-1). These relationships were not just an artifact of the method used to predict CH_4 in this study. Indeed, MIR-CH_4 predictions obtained by the equation were not just the result of specific milk components already predicted by MIR but the result of the new recombination of global spectra information using other spectral regions (Dehareng et al., 2012; Vanlierde et al., 2015). The relationship between HYMIR-CH_4 and fat content is partially explained by the fact that milk fat content and composition and CH_4 production are both dependent on ruminal fermentation (Dehareng et al., 2012). Carbohydrates are the primary source of energy for the cow. The end products of carbohydrate degradation are, amongst others, volatile fatty acids (VFA; the three principle VFA are acetate, butyrate, and propionate), CO_2, and H_2. During the fermentation process, acetate and butyrate promote CH_4 production. The same VFA also increase the fat content in milk, especially short-chain saturated fatty acids (Moss et al., 2000; National Research Council, 2001). In practice, on the one hand, Soyeurt et al. (2006) showed a correlation of 0.95 between milk fat content and the proportion of milk saturated fatty acids (FA) (constituting 70% of total milk FA). On the other hand, Chilliard et al. (2009) showed a positive relationship between daily CH_4 production and milk saturated FA. This could explain why milk fat content was positively linked to HYMIR-CH_4 in the present study.

Regarding protein content, there is little evidence in the literature to suggest that milk protein content is related to CH_4 production. A possible explanation might be that higher DMI might be needed in order to provide more energy for higher milk protein production. This would also lead to more CH_4 production as DMI is a driver for CH_4 production (National Research Council, 2001; Muetzel, 2011). Additionally, there could be an indirect relationship between HYMIR-CH_4 and milk protein content due to the positive correlation existing between milk fat content and milk protein content (Figure 4-1).

4.3. Relationship with milk yield

In the current study, increasing milk or FPCM production as well as milk solids yield was associated with increasing HYMIR-CH_4 (Figure 4-1). Garnsworthy et al. (2012a) showed that CH_4 production was positively related to milk yield ($R^2 = 0.50$). This also agrees with the findings of Gerber et al. (2011). The reason mentioned by Garnsworthy et al. (2012a) was that milk yield is positively linked to DMI, which is a primary driver for CH_4 production. In contrast, Bell et al. (2014) showed that the link between milk yield and CH_4 production (mg/l of eructed air) varied amongst commercial farms; CH_4 production increased with increasing milk production on most farms, but decreased on some farms. The author suggested that this disparity may reflect differences in cow diet composition, energy efficiency, and management. Kandel et al. (2017) also cited similar arguments to explain the weak relationship observed between MIR CH_4 production and milk yield ($r = 0.33$) in their study. For example, Holstein herds selected for higher feed efficiency can produce more milk at a given level of DMI compared to Holstein herds with a lower feed efficiency, which means higher milk yield without higher production of CH_4 (as CH_4 production is correlated with DMI) (Muetzel, 2011; Kristensen et al., 2015). The role of different diets fed to low- or high-yielding dairy cows (e.g., forages vs. concentrates) can also have an impact on the relationship between milk yield and CH_4 production (cf. paragraph below). This might explain the weak relationships observed in the current study involving herds with different genetic performances and management practices.

4.4. Relationship with feed

The effect of diet type on CH_4 has largely been studied in controlled experiments. For example, concentrate feed is known to decrease CH_4 emissions per kg DMI through higher propionate production in the rumen (hydrogen sink) compared to acetate (hydrogen source used by methanogens to generate CH_4) and through accelerated passage rate in the digestive tract. However, it could increase CH_4 production per cow and day in some cases because of increased feed intake, though different effects were observed depending on the context of the experiment (Moss et al., 2000; Knapp et al., 2014). In the current study, there was no relationship observed between HYMIR-CH_4 and the quantity of concentrate equivalents given per cow or per litre FPCM. There was, however, a slight trend ($0.05 < P < 0.01$) towards lower HYMIR-CH_4 for herds fed less concentrate equivalents per cow, in line with some literature results (Figure 4-1, Knapp et al., 2014). It also appeared that low HYMIR-CH_4 herds had a lower quantity of milk produced from forages and fresh grass (Figure 4-1). This could be due to lower intake of forages and fresh grass. This lower intake could be the result of lower genetic DMI capacity, which is known to reduce CH_4

production (Muetzel, 2011), or a lower proportion of grass in the diet compared to other feed for some herds. Indeed, a lower proportion of forages and grass may lead to less CH_4 per unit DMI because of lower acetate production in the rumen compared to propionate (Moss et al., 2000). Lower DMI could also be due to lower forage quality (more fibrous and less digestible – reflected by less nitrogen on grasslands in this study; Figure 4-1) (Warner et al., 2015). Similarly to our study, Bell et al. (2014) investigated differences in CH_4 production between different diets on commercial dairy farms in the same country (partial mixed ration v. partial mixed ration with grazed pasture), but found no significant difference. The reason suggested by the author was that diets could be all of relatively equal quality despite different components, or that more detailed information on diet composition was needed. Besides, in the current study low HYMIR-CH_4 herds seemed to show some characteristics of more extensive management of forage areas – that is a lower number of livestock units per ha, lower milk production per ha, and lower quantity of mineral nitrogen applied on grasslands (Figure 4-1).

4.5. Relationship with herd characteristics

In addition, the obtained results revealed that herds producing less HYMIR-CH_4 seemed to have lower reproduction and health performances (i.e., herd characteristics, Figure 4-1). Information about similar relationships is scarce in the literature, but indirect relationships could be assumed. According to Lucy (2001), a longer calving interval or lower number of calvings per cow (observed in this study for low HYMIR-CH_4 emitters; Figure 4-1) might reflect reproduction problems. According to Drogoul et al. (2004), higher age at first calving, as observed for low HYMIR-CH_4 emitters in this study (Figure 4-1), can be associated with suboptimal nutrition or reproduction problems in heifers. A higher culling rate and a lower age of culled cows, also observed for low HYMIR-CH_4 emitters (Figure 4-1), could be associated with herds that have more health (e.g. lameness, mastitis), fertility or production problems (Adamczyk et al., 2017). According to Lucy (2001) and Windig et al. (2005), it is possible that suboptimal reproductive management could be linked to less global care of the herd by the farmer, suboptimal feeding and health conditions, all three of which are also often linked to lower milk production performances (associated with low HYMIR-CH_4 in this study). A higher number of cows per labour supply for low HYMIR-CH_4 emitters (Figure 4-1) might also mean that the farmer has less time to care for all cows well. These assumptions could explain the observed indirect relationships with HYMIR-CH_4.

4.6. Relationship with economic results

Lower gross margin and profit per cow and litre FPCM were associated with low HYMIR-CH_4, though large variability between herds was observed, meaning that it is possible to find lactating dairy herds with both reduced CH_4 production and good economic results. Due to the lack of information regarding dairy cow gastro-enteric CH_4 production and economy, the interpretation of the relationships between HYMIR-

CH_4 and economic variables was based on interactions with the technical variables discussed previously.

The lower gross margin per cow for low HYMIR-CH_4 herds was the result of the lower value of milk production and the lower value of meat production from cows and from calves expressed per cow ($r_{\text{value of milk production per cow|gross margin per cow}} = 0.75$, $r_{\text{value of meat production from cows per cow|gross margin per cow}} = 0.28$ and $r_{\text{value of meat production from calves per cow|gross margin per cow}} = 0.26$, Figure 4-1). In this study, the value of meat production from cows was calculated on the basis of the value of the kg exported out of the dairy herd (i.e., sold animals such as culled cows, dead animals), plus kg related to inventory change, minus kg imported into the dairy herd (i.e., heifers becoming cows and purchased cows). The observed lower value of meat production from cows expressed per cow could thus partly be due to higher a culling rate observed in the technical part ($r_{\text{culling rate|value of meat production from cows per cow}} = -0.47$, Figure 4-1). Indeed, a higher culling rate would mean more heifers are needed, which usually have higher value than culled cows. Another reason could be the slightly lower value for culled cows ($r_{\text{price of culled cows|value of meat production from cows per cow}} = 0.36$, Figure 4-1).

The value of meat production from calves took into account the value of new-born calves sold or raised on the farm. The observed lower value of meat production from calves expressed per cow for low HYMIR-CH_4 emitters could be partly explained by a lower number of calvings per cow ($r_{\text{number calvings per cow|value of meat production from calves per cow}} = 0.40$, Figure 4-1). The observed lower value of milk production from cows associated with low HYMIR-CH_4 emitters was the result of lower milk production ($r_{\text{milk yield|value of milk production per cow}} = 0.95$, Figure 4-1). The lower variable costs related to the forage area per cow for herds producing less HYMIR-CH_4 might be partly explained by decreased chemical inputs for low HYMIR-CH_4 emitters ($r_{\text{mineral nitrogen on grasslands|variable costs related to the forage area per cow}} = 0.20$, Figure 4-1) or by lower contracting costs.

Regarding the gross margin per litre FPCM, the observed lower value for low HYMIR-CH_4 emitters was probably mainly the result of lower value of meat production from cows per litre FPCM and higher herd costs per litre FPCM ($r_{\text{value of meat production from cows per litre FPCM|gross margin per litre FPCM}} = 0.41$ and $r_{\text{herd costs per litre FPCM|gross margin per litre FPCM}} = -0.51$, Figure 4-1). Lower profit per litre FPCM for low HYMIR-CH_4 herds was also due to increased fixed costs per litre FPCM ($r_{\text{fixed costs per litre FPCM|profit per litre FPCM}} = -0.74$, Figure 4-1). The lower value for meat production from cows per litre FPCM could partly be due to a higher culling rate ($r_{\text{culling rate|value of meat production from cows per litre FPCM}} = -0.46$, Figure 4-1) or a slightly lower value for culled cows ($r_{\text{price of culled cows|value of meat production from cows per litre FPCM}} = 0.40$, Figure 4-1). The observed reduction in herd costs and fixed costs per litre FPCM with increasing HYMIR-CH_4 was most probably due to the dilution effect of slightly higher FPCM yield, because herd costs and fixed costs per cow did not vary with HYMIR-CH_4.

4.7. Intensity of the relationships

The weak relationships between HYMIR-CH_4 and technico-economic variables in this study could be due to other factors in addition to those already explained above for specific variables. First, this study used a CH_4 indicator (cross-validation $R^2 = 0.70$; Vanlierde et al., 2016) and no direct measurements on animals, which could have led to some imprecisions. In contrast, studies involving respiration chambers for CH_4 measurement are expected to reduce CH_4 variation compared with on-farm CH_4 measurements (Garnsworthy et al., 2012b; Bell et al., 2014). Also, the present study comprised technico-economic variables and CH_4 predictions expressed at the herd level. A similar study with individual cow data could make it possible to confirm these relationships or to find new ones, but some individual data are not easily available on a large scale. Besides, as mentioned in the introduction, CH_4 production is influenced by several factors, with some important drivers reported in scientific literature being DMI and feed composition, which are themselves influenced by many other factors. Therefore, other variables that were not available in the databases used in this study might be more strongly associated with HYMIR-CH_4 (e.g., herd genetics, concentrate and forage type and composition, farm environment, Hristov et al., 2013, Knapp et al., 2014). Moreover, it is important to highlight that there was a large diversity of animals and management practices between herds in the studied dataset. This means that several management factors varied simultaneously and in different ways for all commercial farms considered, with possible antagonistic effects and interactions. This differed from many studies reported in the literature in which animals often had a similar genetic background and were housed under the same conditions. In those studies, the effect of only one factor on CH_4 production was usually assessed with other factors being fixed (e.g., Garnsworthy et al., 2012a; Warner et al., 2015). These reasons can explain the low correlations obtained in the present study, which arguably better represent the reality on commercial farms.

5. Conclusion

In conclusion, this study highlighted the fact that low lactating dairy cow gastro-enteric CH_4 production was associated, on a herd and year level, with lower milk fat and protein percentages and lower FPCM yield, as well as with other characteristics such as more extensive management of forage areas or suboptimal reproduction and health performances. Results suggested that some of these characteristics lead to lower economic performances for herds producing less CH_4. Due to the use of on-farm data, the relationships observed in the present study were weaker than in research conducted in controlled conditions, and large variability in technical and management practices was observed for herds with similar CH_4 production. This implies that the numerous relationships between HYMIR-CH_4 and technico-economic variables should be further studied in more detail, including the consideration of interactions and co-evolution aspects between variables. Conclusions drawn in this study regarding CH_4 production could contribute to advanced studies taking into account other GHG and environmental impacts such as those assessed by life cycle analyses.

6. Acknowledgments

The authors acknowledge the milk laboratory Comité du Lait (Battice, Belgium) for providing spectral data and the Plateau de Calcul et Modélisation Informatique (CAMI) of the University of Liège, Gembloux Agro-Bio Tech for computing resources.

7. References

Adamczyk, K., J. Makulska, W. Jagusiak ,and A. Węglarz. 2017. Associations between strain, herd size, age at first calving, culling reason and lifetime performance characteristics in Holstein-Friesian cows. Animal 11:327-334.

Bell, M.J., SL. Potterton, J. Craigon, N. Saunders, R.H. Wilcox, M. Hunter, J.R. Goodman, and P.C. Garnsworthy. 2014. Variation in enteric methane emissions among cows on commercial dairy farms. Animal 8:1540–1546.

Beukes, P.C., P. Gregorini, A.J. Romera, G. Levy, and G.C. Waghorn. 2009. Modelling the efficacy and profitability of mitigation strategies for greenhouse gas emissions on pastoral dairy farms in New Zealand. Proceeding in: 18th World IMACS/MODSIM Congress, p.635-64, Cairns, Australia.

Brask, M., P. Lund, M.R. Weisbjerg, A.L.F. Hellwing, M. Poulsen, M.K. Larsen, and T. Hvelplund. 2013. Methane production and digestion of different physical forms of rapeseed as fat supplements in dairy cows. J. Dairy Sci. 96:2356–2365.

Chilliard, Y., C. Martin, J. Rouel, and M. Doreau, 2009. Milk fatty acids in dairy cows fed whole crude linseed, extruded linseed, or linseed oil, and their relationship with methane output. J. Dairy Sci. 92:5199-5211.

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

Drogoul, C., R. Gadoud, M.M. Joseph, R. Jussiau, M.J. Lisberney, B. Mangeol, L. Montmeas, and T. Tarrit. 2004. Nutrition et alimentation des animaux d'élevage, Tome 2. Educagri éditions, Dijon, France.

Food and Agriculture Organization of the United Nations (FAO). 2013. Greenhouse gas emissions from ruminant supply chains, a global life cycle assessment. FAO, Rome, Italy.

Garnsworthy, P.C., J. Craigon, J.H. Hernandez-Medrano, and N. Saunders. 2012a. On-farm methane measurements during milking correlate with total methane production by individual dairy cows. J. Dairy Sci. 9:3166–3180.

Garnsworthy, P.C., J. Craigon, J.H. Hernandez-Medrano and N. Saunders. 2012b. Variation among individual dairy cows in methane measurements made on farm during milking. J. Dairy Sci. 95:3181–3189.

Gerber, P., T. Vellinga, C. Opio and H. Steinfeld. 2011. Productivity gains and greenhouse gas emissions intensity in dairy systems. Livest. Sci. 139, 100–108.

Gerber, P.J., B. Henderson, and H.P.S. Makkar. 2013. Animal production and health. Mitigation of greenhouse gas emissions in livestock production. A review of technical options for non-CO2 emissions. FAO, Rome, Italy.

Hart, K.J., J.A. Huntington, R.G. Wilkinson, C.G. Bartram, and L.A. Sinclair. 2015. The influence of grass silage-to-maize silage ratio and concentrate composition on methane emissions, performance and milk composition of dairy cows. Animal 9:983–991.

Hatew, B., A. Bannink, H. Van Laar, L.H. De Jonge, and J. Dijkstra. 2016. Increasing harvest maturity of whole-plant corn silage reduces methane emission of lactating dairy cows. J. Dairy Sci. 99: 354-368.

Hristov, A.N., J. Oh, C. Lee, R. Meinen, F. Montes, T. Ott, J. Firkins, A. Rotz, C. Dell, A. Adesogan, W. Yang, J. Tricarico, E. Kebreab, G. Waghorn, J. Dijkstra, and S. Oosting. 2013. Mitigation of greenhouse gas emissions in livestock production – A review of technical options for non-CO2 emissions.FAO, Rome, Italy.

Hristov, A.N., J. Oh, F. Giallongo, T. Frederick, M.T. Harper, H. Weeks, A.F. Branco, W.J. Price, P.J. Moate, M.H. Deighton, S.R.O. Williams, M. Kindermann, and S. Duval. 2016. Short communication: comparison of the GreenFeed system with the sulfur hexafluoride tracer technique for measuring enteric methane emissions from dairy cows. J. Dairy Sci. 99:5461-5465.

Intergovernmental Panel on Climate Change (IPCC). 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland.

Johnson, K.A., and D.E. Johnson. 1995. Methane emissions from cattle. J. Anim. Sci. 73:2483-2492.

Kandel, P.B., M.L. Vanrobays, A. Vanlierde, F. Dehareng, E. Froidmont, N. Gengler, and H. Soyeurt. 2017. Genetic parameters of mid-infrared methane predictions and their relationships with milk production traits in Holstein cattle. J. Dairy Sci. 100:5578-5591.

Knapp, J.R., G.L. Laur, P.A.Vadas, W.P. Weiss, and J.M. Tricarico. 2014. Invited review: Enteric methane in dairy cattle production: Quantifying the opportunities and impact of reducing emissions. J. Dairy Sci. 97:3231–3261.

Kristensen, T., C. Jensen, S. Østergaard, M.R. Weisbjerg, O. Aaes, and N.I. Nielsen. 2015. Feeding, production, and efficiency of Holstein-Friesian, Jersey, and mixed-breed lactating dairy cows in commercial Danish herds. J. Dairy Sci. 98:263-274.

Lucy, M.C. 2001. Reproductive loss in high-producing dairy cattle: where will it end? J. Dairy Sci. 8:277-1293.

Moraes, L.E., J.G. Fadel, A.R. Castillo, D.P. Casper, J.M. Tricarico, and E. Kebreab. 2015. Modeling the trade-off between diet costs and methane emissions: A goal programming approach. J. Dairy Sci. 98:5557–5571.

Moss, A., J.P. Jouany, and J. Newbold. 2000. Methane production by ruminants: its contribution to global warming. Annales de zootechnie, INRA/EDP Sciences 49:231-253.

Muetzel, S. 2011. Effect of level of feed intake on methane production per kg of dry matter intake. Technical Paper. Ministry of Agriculture and Forestry, Wellington, New-Zealand.

National Research Council. 2001. Nutrient Requirements of Dairy Cattle. 7th rev. ed. Natl. Acad. Press, Washington, DC.

O'Brien, D., T. Hennessy, B. Moran, and L. Shalloo. 2015. Relating the carbon footprint of milk from Irish dairy farms to economic performance. J. Dairy Sci. 98:7394–7407.

Pickering, N.K., V.H. Oddy, J. Basarab, K. Cammack, B. Hayes, R.S. Hegarty, J. Lassen, J.C. McEwan, S. Miller, C.S. Pinares-Patiño, and Y. de Haas. 2015. Animal board invited review: genetic possibilities to reduce enteric methane emissions from ruminants. Animal 9:1431–1440.

Soyeurt, H., P. Dardenne, A. Gillon, C. Croquet, S. Vanderick, P. Mayeres, C. Bertozzi, and N. Gengler. 2006. Variation in fatty acid contents of milk and milk fat within and across breeds. J. Dairy Sci. 89:4858–4865.

Thomassen, M.A., M.A. Dolman, K.J. Van Calker, I.J.M. de Boer. 2009. Relating life cycle assessment indicators to gross value added for Dutch dairy farms. Ecol. Econ. 68:2278–2284.

Van Middelaar, C.E., J. Dijkstra, P.B.M. Berentsen, and I.J.M. De Boer. 2014. Cost-effectiveness of feeding strategies to reduce greenhouse gas emissions from dairy farming. J. Dairy Sci. 97:2427–2439.

Vanlierde, A., M.L. Vanrobays, F. Dehareng, E. Froidmont, H. Soyeurt, S. McParland, E. Lewis, M.H. Deighto, F. Grandl, M. Kreuzer, B. Gredler, P. Dardenne, and N. Gengler. 2015. Hot topic: Innovative lactation-stage-dependent prediction of methane emissions from milk mid-infrared spectra. J. Dairy Sci. 98:5740-5747.

Vanlierde, A., M.L. Vanrobays, N. Gengler, P. Dardenne, E. Froidmont, H. Soyeurt, S. McParland, E. Lewis, M.H. Deighton, M. Mathot, and F. Dehareng. 2016. Milk mid-infrared spectra enable prediction of lactation-stage- dependent methane emissions of dairy cattle within routine population-scale milk recording schemes. Anim. Prod. Sci. 56:258–264.

Warner, D., S.C. Podesta, B. Hatew, G. Klop, H. Van Laar, A. Bannink, and J. Dijkstra. 2015. Effect of nitrogen fertilization rate and regrowth interval of grass herbage on methane emission of zero-grazing lactating dairy cows. J. Dairy Sci. 98:3383–3393.

Whitfield, R.G., M.E. Gerger, and R.L. Sharp. 1987. Near-infrared spectrum qualification via mahalanobis distance determination. Appl. Spectrosc. 41:1204-1213.

Windig, J.J., M.P.L. Calus, and R.F. Veerkamp. 2005. Influence of herd environment on health and fertility and their relationship with milk production. J. Dairy Sci. 88:335–347.

5

General discussion

Outline

The objective of this chapter is to explore, in the light of the work carried out in this thesis, some strengths as well as some issues and considerations regarding the development of management decision-support tools in dairy farming using milk MIR. In the end, a general conclusion and future prospects will be drawn.

1. Strengths of MIR for the development of decision-support tools

1.1. *Easy, cheap and large-scale acquisition of MIR data*

Significant strengths of milk MIR for the development and utilization of decision-support tools are the easy availability of milk MIR spectra and MIR traits at low cost and on a large scale via milk samples collected routinely in official milk recording schemes and milk payment systems. First, MIR data are easy to obtain because the collection of milk samples is simple to implement on the farms and does not require any supplemental manipulation of the animals beyond daily milking (i.e., non-invasive method), and because subsequent analyses of milk samples are fast and inexpensive (i.e., a spectrometer can offer a high throughput of around 500 samples per hour – the main cost coming from the sample collection; Gengler and Soyeurt, 2020). For instance, in the Walloon region of Belgium, the price for milk recording, including MIR analysis for conventional traits and somatic cell count, is approximately 250€ per test day for a herd of 90 cows and with samples collected by a technician (awé groupe, 2020). In addition, many existing and future MIR traits could be predicted at a low price for the farmers, because all traits are predicted on the same spectrum that is already collected (i.e., no supplementary sample collection or analysis in the laboratory would be required, apart from the costs associated with the development of the prediction equations). MIR data are collected regularly on a large scale, i.e., bulk tank milk samples are collected at each milk collection on the farm (± every 2 days) for all herds, and individual milk samples are collected usually on a monthly basis for herds enrolled in milk recording services. In practice, the proportion of dairy farmers participating in a milk recording scheme is still limited in some countries (e.g., approximately 35% of dairy cows are recorded in the Walloon region of Belgium and in Poland) while it is higher in other countries (e.g., around 90% of German dairy cows and 70% of Canadian dairy cows are recorded; GGI-Spermex, n.d; Bucek et al., 2014; awé groupe, 2018; Canadian Dairy Information Centre, 2018). Different assumptions can be made to explain the disparity in participation rates in official milk recording services, such as the education level of farmers, incentives or obligations from governments, or the understanding of benefits of MIR (e.g., farmers could be more likely to enroll in a milk recording program if scientists demonstrate its ability to translate into increased profits).

1.2. *Association and comparability of MIR data*

The use of MIR spectrometry through milk recording and milk payment systems enables to collect spectral data on individual farms and then associate data from different farms. Comparison of MIR spectral data within and between farms is quite straightforward because (1) milk samples are collected and analyzed following general standard procedures and the spectrometers regularly undergo quality controls to ensure the acquisition of accurate spectral data (ICAR, 2017, 2019), and (2) collected data are gathered and can be saved in regional or national central databases for an extended period of time. For instance, in the Walloon region of Belgium, MIR spectra are available in a central database since 2007. Different spectrometers have specific instrumental responses, especially different models or brands. To associate spectra from different instruments (e.g., from different countries), adjustments of predicted data to reference standards can be made (i.e., slope and bias correction, Lynch et al., 2006). This method is only applicable when reference samples are available (e.g., for fat, protein and lactose contents), which is too expensive or even impossible for many new traits like CH_4 production or body energy status (McParland et al., 2011; Vanlierde et al., 2018; Grelet et al., 2019). Hence, standardization procedures have been developed to harmonize directly the spectral responses of instruments and allow the merging of data from different instruments, as described by Grelet et al. (2015, 2017a), Bonfatti et al. (2017) and Tiplady et al. (2019). Standardization will be discussed further in connection with the research conducted in this thesis in section 2.2.2 of the present chapter. In comparison, associating and comparing information from different farms is more difficult with other tools such as on-farm precision farming tools, which are mostly separated from standard recording systems. Information collected by on-farm technologies often stays on the farm (Gengler, 2019). Comparing and sharing information is more complicated with these technologies, among others because standardization among apparatus is more difficult and because computer tools are required to exchange data between systems and to export data to central databases independent from the manufacturer (e.g., Ori-Automate, FIDOCL Conseil Elevage, n.d.). Consistency and comparability of collected MIR spectral data are important for research studies and the development of MIR tools as well as for the application of the developed tools (Grelet et al., 2015). As a practical example, regarding Walloon milk recording services, statistics on milk composition predicted from MIR are reported monthly to the farmers (both for each animal and at the herd level) and the herd values for milk fat and protein contents are compared with the mean values of all enrolled Walloon farms so that the farmer has a concrete benchmark for comparison. The same procedure could be imagined for other predicted MIR traits.

1.3. A versatile method for the development of customized tools at the herd and animal levels

MIR has the advantage of being a versatile method, with multiple potential applications to support herd and individual cow management decisions. To make improvements in the dairy sector and enhance its resilience, more specific information to each herd and animal should be evaluated in the future (Marchewka et al., 2018). As such, MIR enables the development of customized tools at the herd or at the animal level, i.e., tools that provide personalized, and not generic, information for decision-making about individual animals or a specific herd. Measuring indicators on individual animals to improve management strategies and performances is in the scope of precision dairy farming (Bewley, 2010). As already mentioned, MIR is commonly used in the field to predict major milk components (e.g., fat and protein contents) of individual animals and the whole herd (ICAR, 2017). Besides, many research institutions have recently explored MIR to predict other direct or indirect traits of interest, although perspectives of applications on farms were more contrasted. These traits were related for instance to fine milk composition (e.g., Soyeurt et al., 2006a; Bonfatti et al., 2011), milk technological properties (e.g., Visentin et al., 2015) or cow physiological status (e.g., Soyeurt et al., 2019; Ho et al., 2019), as also approached in Chapters 2 and 4 with the pregnancy status and CH_4 production. In addition to the development of prediction equations, the large amount of available MIR spectra collected over long periods of time and stored in central databases allows to carry out research studies and develop tools requiring a great number of phenotypes and that can provide valuable information for decision-making, such as genetic evaluations or population studies. Test-day models using MIR traits are widely used in genetic studies (e.g., to derive breeding values in genetic evaluations) but, as shown in Chapter 3, they can also be used for individual cow and herd management purposes (e.g., for the detection of problems or prediction of future data). These models require a large number of records collected on a regular basis because a sufficient amount of data by class of each effect is needed to have a reliable estimation of solutions (Bastin et al., 2009). In the context of this thesis, 467,496 individual spectra from 53,781 Holstein dairy cows in first lactation were used in Chapter 3 to model milk spectra using test-day models. Population studies (i.e., study of a MIR trait in a population) providing information to support decision-making, similar to the study in Chapter 4, are also possible thanks to the large-scale availability and comparability of MIR data. In Chapter 4, a high number of herds was required to conduct a large-scale study, and a high number of records was required for each herd, with an even distribution throughout the year, in order to have a representative estimate for CH_4 production at the herd and year level.

1.4. From day-to-day decisions to long-term impacts and strategic decisions

Tools developed with MIR and applied to large databases, for instance as part of milk recording or milk payment, can guide decisions with short, medium or long-term impacts for dairy farming, and can also guide decisions at the political or research level. Thus, the scope of MIR to support decisions is quite large.

Short-term impacts of MIR decision-support tools refer to day-to-day management decisions taken on dairy farms. In the field, milk recording agencies or dairy advisors would be able to interpret outputs of MIR tools and transfer them to farmers to advise or alert them (e.g., when threshold values for traits of interest are reached). This could help with feeding decisions (e.g., adjusting the diet based on herd or individual performances), culling decisions (e.g., identifying cows performing below the average), pasture management, mating decisions, heifers replacement, grouping decision (e.g., separating cows in different feeding or management groups), etc. (Bewley, 2010; ICAR, 2017). This better day-to-day monitoring of dairy farmers' management systems would improve their dairy production on numerous aspects including economic aspects (e.g., reducing costs or increasing revenues), reduction of adverse environmental impacts, improved animal welfare, or improved product quality (Bewley, 2010). In Chapter 2, we concluded that MIR, at the current stage, is probably not sufficient to diagnose pregnancy at early stages. However, a possible application that could arise from the model is the detection of mid and late-term fetal abortion because the model accuracy was promising for advanced pregnancy. That would help farmers decide, for example, which cows need to be re-inseminated after pregnancy loss (i.e., mating decisions) and avoid loss of money due to non-calving or long delay of calving (De Vries, 2006). This is an example of short-term decisions based on MIR and with an economic interest. In addition, in Chapter 3, the aim of potential future applications related to MIR spectra modeling (e.g., comparison of observed vs. expected MIR spectra or traits, or studying deviations in the solutions of the herd effects) is also to provide short-term decisions regarding individual cows and herd management (e.g., detecting problems by identifying meaningful deviations from the expected pattern).

Medium and long-term impacts of MIR tools concern decisions about herd development (e.g., herd replacement) as well as breeding programs and selection decisions. MIR traits can help make decisions for breeding programs (e.g., cows with low milk solid content or poor fertility can be inseminated with appropriate bulls to improve these traits over time in the herd). MIR traits can also be used in genetic and genomic evaluations that support selection decisions, for example by deriving breeding values or selection indices using MIR traits (Gengler et al., 2016). From an environmental point of view, environmental-related traits such as MIR-predicted CH_4 could be included in selection indices to breed for CH_4-efficient cows, providing it does not impede the selection of other important traits (Vanrobays, 2019). Also, decisions related to breeder association activities and selling breeding animals require recording data in many cases (ICAR, 2017).

Beyond decisions taken on farms, MIR tools are interesting for research and strategic decisions, that can indirectly influence farm management practices. For example, MIR tool outputs can be used in research programs aiming to broaden general knowledge and understanding in dairy farming (e.g., Soyeurt et al., 2006b; Visentin et al., 2018), as shown in Chapter 4 of this thesis (i.e., the aim was, among others, to understand management practices associated with CH_4 emissions). Outputs of MIR tools and subsequent research could also guide political decisions (or milk buyers) about possible incentives in dairy farming (e.g., bonus/subsidies or penalties/taxes). In Chapter 4, findings about the relationships between MIR-predicted CH_4 production and technico-economic results could be useful to guide future research on the carbon footprint of milk or for policymakers. We found out that low MIR enteric CH_4 production (g/day) tended to be associated with more extensive or suboptimal management practices, which could lead to lower profitability. However, an important point is that observed relationships (i.e., correlations) were weak, which made it difficult to associate specific practices with CH_4 emissions with certainty. This indicates that the links between CH_4 production and farm practices are complex, with possible antagonistic effects and interactions with various management factors. These results suggest that relationships between CH_4 and farming practices should be studied in more detail in future research (i.e., with real farm data, and not experimental data in a controlled environment) to know more about determinants of CH_4 emissions. This is also a sign that policymakers should not take decisions on policies aiming to reduce enteric CH_4 emissions on farms before it is studied and understood further.

2. Issues and considerations regarding the development of MIR decision-support tools

Following on from the above discussion, we cannot deny that MIR has many strengths for the development of management decision-support tools. Recently, several research projects (e.g., OptiMIR – developing prediction tools based on MIR) have been optimistic regarding the ability of MIR to predict many traits of interest for practical use. However, as also stated in the introduction, finally only a limited number of MIR management tools are currently used on farms. Several models developed in a research context proved to be too little accurate or robust, either at the first research stages or during the validation process (e.g., Soyeurt et al., 2009; Eskildsen et al., 2014; Visentin et al., 2016; Bonfatti et al. 2020). In addition, challenges also appear at the implementation stage, thereby limiting the number of tools effectively used in practice. In this context, the research conducted in the framework of this thesis raised several issues and considerations regarding the development of MIR tools to support decision-making. For more clarity, we separated our different remarks in four sub-categories: (1) the prediction of indirect MIR traits, (2) data selection (quality and variability of data), (3) choice and validation of models, and (4) practical implementation of the developed tools.

2.1. On the prediction of physiological status-related traits and other indirect MIR traits

MIR has proven to be an accurate method for the quantification of the main milk components that have a direct footprint in the milk spectrum and that are present in a high concentration such as fat, protein, or lactose contents. In this case, the final purpose of MIR is to replace the gold standard/reference values even if slope and bias corrections must be performed regularly to ensure this high reliability (Barbano and Clark, 1989; Lynch et al., 2006). Subsequently, MIR has been explored for the quantification of minor milk components present in lower concentration in milk (e.g., mineral contents, fatty acids, protein fractions), with somewhat more limited accuracy (e.g., Soyeurt et al., 2006a; Soyeurt et al., 2009; Bonfatti et al., 2011). More recently, MIR has been used to investigate the prediction of several physiological status-related traits and other indirect traits with the aim of improving economic or environmental performances of dairy production or animal welfare. Examples are the body energy status (McParland et al., 2011), likelihood of conception (Ho et al., 2019), methane emissions (Vanlierde et al., 2018), body weight (Soyeurt et al., 2019) or pathologies like ketosis (Grelet et al., 2016, 2019). Indeed, milk composition is considered as a biomarker and accessible source of information on the physiological status of dairy cows because of the interaction between blood circulation and milk composition (Gengler et al., 2016). However, many of these traits have complex relationships with milk composition and an indirect and sometimes little detectable footprint in the milk spectrum. MIR is not expected to detect very small differences in milk composition or differences in components present in very low concentration in milk, and changes in milk composition must be relevant to be captured by chemometric methods (Eskildsen et al., 2014). Besides, the relationship between some indirect traits (e.g., lameness, early pregnancy) and milk composition might be quite erratic or not obvious (Bonfatti et al., 2020). Therefore, the prediction accuracy for such traits is often limited (i.e., poor to moderate) and sometimes MIR fails altogether to predict the targeted trait. For example, Bonfatti et al. (2020) concluded that results of the prediction of the lameness score using MIR were considered to be too poor to envisage a practical application as on-farm tools or to be used as large-scale phenotype for animal breeding purposes. Grelet et al. (2017b) found out that MIR was able to predict the energy status of dairy cows (i.e., energy balance, residual feed intake, dry matter intake, and blood metabolites) with only limited accuracy as R^2 of cross-validation ranged between 0.33 and 0.68 depending on the trait.

Similarly, low ability of MIR to predict the pregnancy status of dairy cows was observed in Chapter 2 of this thesis. Early stages of gestation were the most interesting stages to diagnose, but they could not be predicted successfully from MIR information. Only the model developed using data recorded at 151 days or more after insemination (i.e., late pregnancy) showed promising prediction accuracy with the average value of AUC of 0.78 and 0.76 obtained through cross-validation and testing, respectively. This low global ability to predict pregnancy with MIR could be partly attributed to the indirect and complicated relationship between pregnancy and milk composition and to the uncertain existing footprint in milk spectra in early gestation stages. Indeed, different factors such as endocrine changes, reduced feed intake and greater nutritional demand for the growing fetus during pregnancy potentially contribute to altering milk composition, and several research studies reported changes in milk composition only after a few months of gestation (e.g., Olori et al., 1997; Roche, 2003; Penasa et al., 2016). Beyond what has been discussed in Chapter 2, we noticed that testing AUC were slightly higher than 0.5 (i.e., between 0.59 and 0.65) for the models at early stages after insemination (before 151 days after insemination). We hypothesized that these values could partly reflect the ability of the cow to get pregnant (i.e., likelihood of conception), as studied by Hempstalk et al. (2015) and Ho et al. (2019), and not only the true pregnancy status. One must be careful that a trait might actually be related to another and the model might not predict only the true effect of the targeted trait, but also catch the signal associated with a closely related trait, as also suggested by Eskildsen et al. (2014).

Likewise, the MIR CH_4 equation used in Chapter 4 of this thesis had a moderate prediction accuracy (i.e., cross-validation R^2 and standard error of 0.70 and 70 g/day, respectively; Vanlierde et al., 2016). Enteric CH_4 production is a complex trait and has an indirect relationship with milk composition (especially fatty acids) because both are dependent on fermentations in the rumen (Dijkstra et al., 2011; Vanrobays et al., 2016). In ruminants, milk fatty acids come from dietary lipids transported in the blood (long-chain fatty acids) and de novo synthesis in the mammary gland based on the volatile fatty acids acetate and butyrate coming from dietary carbohydrates (short- and medium-chain fatty acids). CH_4 emissions are also associated with higher acetate and butyrate concentration in the rumen (Moss et al., 2000; Dehareng et al., 2012), which explains the relationship between milk fatty acids and CH_4 emissions. CH_4 is also a good illustration of the influence of other factors on the relationships between a trait and milk components. Indeed, milk fatty acid profiles are influenced by lipomobilization during the lactation, leading to a non-constant link between CH_4 and fatty acids and therefore non-constant prediction coefficients throughout the lactation (Vanlierde et al., 2015; Vanrobays et al., 2016).

In short, physiological status-related traits and other indirect traits usually have complex relationships with milk composition; their prediction is not always straightforward and requires good knowledge about how these traits interact with milk composition. As many of these traits are difficult to predict with high accuracy, they can be used as indicators or proxies that have a low cost but are reliable enough, rather than replace reference methods (De Marchi, 2014; Gengler et al., 2016). The utilization of MIR indicators will be discussed further in section 2.4.1. Nevertheless, we have to acknowledge that MIR is not flawless and can fail to predict certain traits with sufficient accuracy for practical use, even as indicators (e.g., early pregnancy), especially if the fingerprint/signal in the milk MIR spectrum is not strong enough to be detectable.

2.2. Data selection

This section addresses the importance of the quality and variability of reference and spectral data as a prerequisite for the development of accurate and robust MIR tools and for appropriate implementation. The accuracy of a model is the degree to which its results conform to the correct value or standard (Lexico, n.d.). The robustness of a model is its ability to take uncertainties into account and to perform well even if it is applied in various situations or if its assumptions are somewhat violated (i.e., flexibility; Mathematical Modelling Company, n.d.). Robustness is particularly important for practical implementation purposes.

2.2.1. Quality of reference data

MIR spectrometry is an indirect technique and needs precise reference analyses (i.e., good quality phenotypes) to develop accurate prediction equations. However, this is difficult for some traits like lameness, CH_4 emissions, mastitis, or pathologies like ketosis (e.g., Bonfatti, 2020; Grelet et al., 2019; Vanlierde et al., 2018). Regarding the

MIR CH_4 prediction equation used in Chapter 4, reference values were measurements conducted with the sulfur hexafluoride (SF_6) tracer gas technique (Vanlierde et al., 2016). The basic idea behind the method is that CH_4 production is measured based on the CH_4/SF_6 ratio in a representative gas sample from which the emission of SF_6 (the tracer gas in a bolus placed in the rumen) is known (Storm et al., 2012). As explained by Vanlierde (2019), adaptations of the classical protocol were made to ensure reliable SF_6 measurements of CH_4 emissions. However, the open-circuit respiration chamber method, measuring all gas flows and composition, is still regarded as the gold standard for CH_4 measurement (Grainger et al., 2007). Reasons for the use of SF_6 over chambers is the lower cost (i.e., no need for special expensive infrastructures), it does not require confinement of the animals (e.g., measurements can be done for grazing animals) and it is less labour-intensive (Grainger et al., 2007; Storm et al., 2012). Recently, Vanlierde (2019) added chamber reference values to the SF_6 values to refine the prediction of MIR CH_4, also adding more variability to the dataset (i.e., increased number of countries, cows, breeds, lactation stages, parities and herd management practices). Repeating the study in Chapter 4 using the latest MIR prediction equation would be useful to see if results would be modified.

As for reference values in Chapter 2, the pregnancy status was defined using records of insemination and the corresponding calving date. These reference data can be considered very accurate. The only minor concern is that early embryonic losses (during the first days after insemination) might not have been detected, i.e., the cow was pregnant for a short period and then returned to heat, but in the data this short pregnancy period was not mentioned if it was not noticed by the farmer. Thus, a small proportion of spectra could have been erroneously associated with open status even if their status was actually pregnant. From a hypothetical point of view, this might affect the model. However, as this happens at very early stages after insemination, it is not expected to affect milk composition yet and, as such, it should not affect the model accuracy. Data related to later abortions were excluded from the beginning already.

2.2.2. Quality of spectral data

The quality of spectral data is important for both the development and the implementation of MIR tools such as prediction equations and test-day models. This quality is ensured, among others, by the milk sampling procedure, laboratory analyses, cleaning and preprocessing of spectra, and potential standardization of spectra if required (Rinnan et al., 2009b; Grelet et al., 2015; ICAR, 2019).

Quality checks should be performed on spectral data to identify inconsistent or inaccurate data (i.e., outliers) that should be removed in order to avoid adverse effects on subsequent analyses or models (Zhang et al., 2019). For example, in this thesis, unusual spectral records were removed based on inconsistent values for some common MIR-predicted traits such as fat or protein contents (cf. Chapter 3) or based on the standardized Mahalanobis distance (i.e., global H distance, GH; cf. Chapter 2). Other methods also exist, for instance regarding quantitative prediction models, samples with residuals higher than 2.5 times the standard deviation of all residuals could be considered as outliers (Rousseeuw et al., 2006; Grelet et al., 2019).

In addition, spectral preprocessing is a suitable step to remove unwanted variation (e.g., instrumental artifacts) to focus more on the variation of interest (Engel et al, 2013). Spectral preprocessing compensates for deviations from linear relationships between absorbance and concentration and consequently improves the linear relationships between the spectral signal and analyte concentrations (i.e., Lambert-Beer's law; Rinnan et al., 2009b). The most common types of preprocessing techniques are those which do not need reference values for the preprocessing operations; they are referred to as reference-independent preprocessing (Rinnan et al., 2009b). Reference-independent preprocessing techniques can be divided into two subgroups: derivative methods (e.g., first or second derivatives, Savitzky–Golay) and scatter correction methods (e.g., multiplicative signal correction, standard normal variate; Rinnan et al., 2009b; Engel et al., 2013). Based on previous experience from our research team, the first derivative was chosen as preprocessing method in Chapters 2 and 3 to remove baseline variation that may occur because of instabilities between instruments of a given brand and over time (Owen, 1995). In Chapter 4, a first derivative was also calculated on the raw spectra, as required to use the MIR CH_4 equation developed by Vanlierde et al. (2016). Besides, in Chapters 2 and 3, spectra were centered and scaled as a common preprocessing technique in PLS-DA models and PCA, and this also corrects for scattering effects (Rinnan et al., 2009b). We only chose the first derivative as main pre-correction to avoid applying too severe preprocessing that could remove the valuable information, as suggested by Rinnan et al. (2009a). In many cases, preprocessing strategies are selected based on past experience or on the identification of artifacts in the spectra, but this does not always guarantee appropriate selection. Indeed, as stated by Engel et al. (2013), the choice of one or several optimal preprocessing method(s) may strongly influence the results of the spectral analyses or models, but it is not straightforward, and different techniques need to be tested depending on the characteristics of our specific dataset and the goal of our analyses. This is a point we need to address more conscientiously for future research.

Before developing a model using MIR spectra, it is common to select the spectral wavenumbers located in the most informative regions and to remove noisy regions with low signal-to-noise ratio induced by water absorption (e.g., Iñón et al, 2004; Capuano et al., 2014; Grelet et al., 2016). Water absorption regions are situated between ~1,600 and 1,700 cm^{-1} (O-H bending region) and above ~3,000 cm^{-1} (O-H stretching region; Tiplady et al., 2019). In Chapter 2, we selected 538 spectral wavenumbers out of 899 (i.e., Bentley spectrometer) for the study, in the regions from 928 to 1,596 cm^{-1} and from 1,693 to 3,025 cm^{-1}. In Chapter 3, a total of 311 wavenumbers out of the 1,060 (i.e., Foss spectrometer) were retained, covering the regions from 933 to 1,589 cm^{-1}, from 1,704 to 1,809 cm^{-1} and from 2,553 to 2,981 cm^{-1}. Similar regions were conserved in both studies, even though more restrictive regions were used in the second one. Figure 3-3 in Chapter 3 revealed that the variance in the regions from 1,790 to 1,809 cm^{-1} and from 2,553 to 2,773 cm^{-1} was mainly explained by the herd-test-day effect (i.e., the wavenumbers are influenced by test-day factors, such as daily changes of feed, climatic conditions or laboratory environment). Some authors reported that the spectral wavenumbers between ~1,800 and ~2,800 cm^{-1} (encompassing the two abovementioned regions with high herd-test-day variance) have no specific bands or useful chemical information (e.g., Iñón et al., 2004; Dagnachew et al., 2013). Hence, this region was removed in some studies in the literature (e.g., Grelet et al. (2016), Grelet et al. (2019), and Lainé et al. (2017) retained in total 212 spectral wavenumbers). Nevertheless, other authors stated that, although these wavenumbers are not associated with main chemical information, they could be associated with minor components (e.g., Socrates, 2001). Regions with high herd-test-day variance may be useful to study test-day effects for the purpose of detecting daily herd or analysis problems. Further, some authors even stated that water absorption regions should still be considered for investigation (after reduction of interference from water absorption) because they include absorbance peaks for chemical bonds related to non-water milk components (e.g., Bittante and Cecchinato, 2013; Wang and Bovenhuis, 2018). These noisy water absorption regions could also be used to identify potential dependence structures in the data that could affect model accuracy (Wang and Bovenhuis, 2019). These considerations are worth exploring in the development of future MIR tools.

Standardization was described in section 1.2 of this chapter in the context of the combination of spectra from different spectrometers that have different instrumental responses (Grelet et al., 2017a; Bonfatti et al., 2017; Tiplady et al., 2019). In addition, when the spectra are collected over different years, there might be some shifts in the machine response and spectra across time, even when using the same instrument (Tiplady et al., 2019). These variations can be the consequence of, for example, electric drift or detector instability (Bonfatti et al., 2017). Standardization of spectra can be used to correct for such shifts over time. However, we used non-standardized spectra in the three studies carried out in this thesis. The reasons were that we used spectra from the same spectrometer brand and the same laboratory in each study, and we applied a first derivative on the spectral data, which partly corrects for potential spectral drifts over time (Owen, 1995). Regarding spectra modeling (Chapter 3), we tried to build the same model with standardized spectra and the results were very similar, showing that standardization was not needed. However, when we want to use data coming from different laboratories and/or spectrometer brands, the standardization of MIR spectra is useful. In future studies, we can make sure that there are not any potential spectral drifts over time (even when using standardization because shifts might also appear if standardization coefficients are not calculated frequently enough) using exploratory analysis such as plotting the scores of principal component analysis (PCA) performed on the spectra over time (Bonfatti et al., 2017).

2.2.3. Variability of reference and spectral data

A sufficient number of reference data and spectral data with the largest possible variability (e.g., various cows, breeds, lactation stages, parities, diets, management practices) is needed to develop robust MIR models for practical use (Berry et al., 2013; Vanlierde et al., 2018). The dataset used to build a model must include a maximum of variability that could be encountered in the field. Applying a model to a sample not covered by the calibration dataset used to build the model is likely to lead to flawed predictions. Consequently, it is also important to ensure that the model is suitable for each new milk spectrum, for instance by assessing if the new data are in the variability covered by calibration data or by calculating the GH between the spectrum to be predicted and the calibration spectra used to build the model (Berry et al., 2013; Grelet, 2019). Collecting a large number of data with high variability is often challenging (especially for reference data), mainly because of the cost and time required. Many MIR prediction equations reported in the literature were developed on a limited dataset, sometimes because they were first developed for research purposes (e.g., Bonfatti et al., 2011; Grelet et al., 2019; Visentin et al., 2015). When publishing results of a prediction equation and other models, it is important to mention clearly the limits of validity and, when appropriate, to advertise that the models developed on limited data need to be refined with a larger amount of data and more variable data to increase their robustness for practical implementation.

In Chapter 2 of this thesis, we used a moderate number of available data to develop the MIR prediction equation for the pregnancy status, especially for the second and third strategy (i.e., 6,754 calibration records the first approach, 1,664 calibration records for the second approach and between 348 and 1,566 calibration records for the third approach depending on the class). Data were from a limited number of herds (19) with mainly similar management practices (pasture-based). Adding extra data with higher variability, for example following collaboration with other countries, would be beneficial to confirm the obtained results, including the potential application to detect fetal abortion. In this case, standardization of spectral data would be needed for harmonization of spectral data among different spectrometers (Bonfatti et al., 2017; Grelet et al., 2017a; Tiplady et al., 2019).

In Chapter 4, the MIR CH_4 prediction equation used was created using, among others, reference values from Walloon dairy cows (Vanlierde et al., 2016). Therefore, we considered the equation to be appropriate for the MIR spectra used in the study, originating from the Walloon milk recording scheme. Additionally, the GH was used to discard spectra that were too distant from spectra in the calibration set used to build the equation. The MIR CH_4 equation validity domain has been extended recently, among others by including data from more countries with different management practices (Vanlierde, 2019).

Regarding Chapter 3, MIR spectra used in the test-day models were limited to herds enrolled in the Walloon milk recording system and we selected only first lactation Holstein cows. Consequently, one has to be aware that the accuracy (cf. Figure 3-4 and Table 3-2) and utilization of the model is only valid within these limits. When the model is applied to a different population or when there are changes over time (i.e., when there is a potential (co)variance change), variance components need to be re-estimated (van der Werf, n.d). There is a need to add more data, including other lactations (at least until the third lactation then we can assume that the variance components estimated for this third lactation could be equal to the ones estimated for later lactations) and breeds (e.g., multiple-lactation model and one separate model per breed) to expand the use of the model to more diverse situations.

2.3. Model development

The choice of an appropriate model according to the data and the objective, and the subsequent validation of the model are important to have accurate and robust MIR tools that are easy to implement. Some aspects of model selection and validation will be discussed based on Chapters 2 and 3.

2.3.1. Choice of the model

In Chapter 2, we used the classical PLS-DA method to classify cows as open or pregnant based on MIR spectra. PLS-DA is the most widely used supervised discrimination method for spectral data because it performs dimensionality reduction while carrying out classification at the same time, which is appropriate when there are many highly correlated explanatory variables (Rozenstein et al., 2015). Accordingly, this method is relevant due to the high number of spectral wavenumbers, of which some, especially the ones close to each other, are highly correlated (i.e., collinearity). However, other machine learning algorithms have recently been used in the development of MIR prediction models (e.g., Shahinfar et al., 2014; Hempstalk et al., 2015; Pradle et al., 2018; Tremblay et al., 2019). Machine learning is an application of artificial intelligence that provides computers the ability of automatically building statistical models that can generate predictions based on what they have learned and without being explicitly programmed (Hempstalk et al., 2015). It is legitimate to ask if alternative machine learning algorithms could outperform PLS-DA for the prediction of the pregnancy status, so we did extra research to find out more. Following on from the results in Chapter 2, we focused on the third strategy (the most promising strategy studied), i.e., the consideration of separate models for 7 classes of records based on the number of days after insemination (progressively greater gestation). We tested several alternative machine learning algorithms to classify cows as open or pregnant (using the Caret package in R, version 3.6.0; Kuhn, 2019), i.e. random forest, support vector machine (SVM), artificial neural network (ANN), logistic regression, naive Bayes classifier and k-nearest neighbors (KNN). The models used are classifiers. More information about how these algorithms work can be found in Kotsiantis et al. (2006) and Hempstalk et al. (2015). These particular models were selected because they are commonly used for classification, they represent different machine learning approaches to classify data and they have different levels of complexity. For example, the three last machine learning methods (naive Bayes classifier, logistic regression, and KNN) are simple methods that can be more appropriate for small datasets than more abstract methods like ANN (Forman and Cohen, 2004; Kotsiantis et al., 2006). A limitation in the use of other methods than PLS-DA lies in the collinearity and the high dimensionality of spectral data, especially compared to the little number of data in each of the 7 classes after insemination (between 348 and 1,566 records for calibration). A large number of predictive variables could among others cause the models to overfit (i.e., the model fits too closely to a particular dataset and learns from features that represent noise) and perform poorly in external validation (Hempstalk et al., 2015). Therefore, dimension reduction while limiting collinearity is recommended before implementing the machine learning models. Many methods for variable selection and dimension reduction exist (e.g., Andersen and Bro, 2010; Anzanello and Fogliatto, 2014), including the popular PCA method and the partial least squares (PLS; or PLS-DA for classification) factorization method. PCA is commonly used to reduce the dimensions of datasets with correlated variables. This methodology takes into account only the variability of the independent variables (i.e., spectral data) and not the variability of the trait to predict (i.e., the pregnancy status), which could lead to loss of relevant information. Therefore, we decided to use PLS-DA factorization, for which the latent variables are calculated in such a way to explain the greatest variance between the predictors and the response (Nguyen and Rocke, 2002). PLS-DA latent variables were extracted from the PLS-DA model used in Chapter 2. The optimal number of PLS-DA latent variables was chosen by the PLS-DA model (i.e., number maximizing cross-validation performances) and could be different for each of the 7 classes. Validation of the models and performance evaluation were carried out as presented in Chapter 2 (same protocol as for strategy 3, but other machine learning algorithms with prior dimension reduction were tested instead of PLS-DA). The results for the test set are presented in Table 5-1 for the 2 last classes of records after insemination only (i.e., 151 days or more after insemination), because these are the groups for which the results were the most promising. The area under the receiver operating characteristic curve (AUC), sensitivity, and specificity of the alternative machine learning algorithms were very similar to PLS-DA. The class probabilities for correctly classified records (i.e., “score of certainty” for classification) are not comparable between different algorithms because they are calculated differently. In short, given these results, we cannot conclude that any other algorithm tested was superior to PLS-DA for the prediction of the pregnancy status in dairy cows. Conclusions for the 5 first classes of records after insemination were similar (results not shown). Other methods of dimension reduction or variable selection to reduce the number of predictors are worth exploring for future research, such as recursive feature elimination (Granitto et al., 2006; Spetale et al., 2016; Bahl et al., 2019).

Beyond the accuracy of the model, it is interesting to consider the computational efficiency and ease of implementation when the goal is to deploy MIR models through a production environment, which is not often taken into account and discussed in scientific papers. For example, some powerful algorithms such as neural networks usually take a long time to run the first time but, when new data are incorporated in the calibration set, it is possible to update the coefficients in some models considering only the new data instead of having to train the entire model again (i.e., incremental learning, Castro et al., 2018). This is appreciated when new data are regularly available. Figuring out the best algorithm to use in terms of accuracy and efficiency is relative to each specific problem and application needed. This is a point we need to consider as a perspective for future research.

Table 5-1. Strategy 3 (stages after insemination): results of different machine learning algorithms for the cow independent test sets for the 2 last classes of records based on the number of days after insemination[1]

	Class 6 (151–180 d)			Class 7 (≥181 d)		
	AUC	Sens.	Spec.	AUC	Sens.	Spec.
PLS-DA	0.71	0.71	0.55	0.82	0.75	0.74
Random forest	0.71	0.71	0.56	0.81	0.78	0.72
Linear SVM	0.71	0.74	0.55	0.82	0.77	0.74
ANN	0.70	0.73	0.59	0.80	0.78	0.68
Logistic regression	0.70	0.74	0.53	0.82	0.78	0.71
Naïve bayes	0.68	0.61	0.61	0.81	0.79	0.69
KNN	0.71	0.77	0.53	0.82	0.74	0.74

[1] AUC = area under the receiver operating characteristic curve; sens. (sensitivity) = proportion of records belonging to pregnant cows that were correctly classified as pregnant; spec. (specificity) = proportion of records belonging to open cows that were correctly classified as open

In Chapter 3, there were different possibilities of test-day models for the modeling of MIR spectra for management purposes. We decided to start with a simple model (i.e., single trait mixed test-day model with a limited number of effects) as this research was an exploratory study on the prediction of MIR spectra. Nonetheless, we are aware that further research should be carried out to refine the model, as also mentioned in Chapter 3. In our mixed test-day model, random effects variations in the course of the lactation were assumed to be constant. However, this assumption that random effects variances are homogenous throughout the lactation is suboptimal (Ptak & Schaeffer, 1993; Vanderick, 2017). Therefore, a first step to improve the model would be using random regression coefficients for each random effect (e.g., using Legendre polynomials). In this case, the shape of the modeled trait along the lactation for an individual cow could be viewed as two sets of regressions on DIM. Fixed regressions for cows belonging to the same class of fixed effects describe the general shape of the modeled trait throughout the lactation, and the random regressions for a given cow describe the deviations from this general shape, which allows each cow to have its own shape (Jamrozik and Schaeffer, 1997). Other potential improvements explained in Chapter 3 could be the addition of extra effects in the model, such as the age of the cow or the calving date and the development of a multi-lactation model. However, it should be noticed that, even with the best model possible, many predictions will still differ from observations. This is one of the benefits of such models for management purposes, for example to identify suspect observations by comparison with their expected (i.e., modeled) value to detect problems. Practically speaking, test-day models are known to be computationally demanding (Swalve, 2000), which can sometimes hinder easy development and practical use even if recent advances in high-performance computing have the potential to make them more viable. A challenge for the practical implementation of test-day models is the computing time and memory required to estimate variance components and, to a lesser extent, to calculate the solutions of the model. The re-estimation of variance components is done every now and then, e.g., when there is a (co)variance change over time due to environmental and/or genetic change (van der Werf, n.d.). In contrast, the calculation of model solutions (e.g., BLUP) is carried out each time we want to have new predictions, but requires less computing power. That is why (1) we assumed that it would be easier to model the MIR spectrum instead of the various existing and future MIR traits, and to subsequently apply developed MIR prediction equations to the modeled spectrum (this reduces the computational requirements and workload of developing several test-day models adapted to each trait); and (2) we reduced spectral information using PCA and we modeled each selected principal component (PC) separately using a single-trait model. Parallel computing of several single-trait models was preferred over a multi-trait model including several PC for computational reasons and because PC are phenotypically de-correlated, meaning that genetic correlations are also assumed to be small (Soyeurt et al., 2008; Shallue et al., 2019). However, excessive reduction in the complexity of the model for computational reasons should not be done at the expense of model performances. This research study was preliminary and there is still food for thought regarding the most suitable test-day model (in terms of accuracy, robustness, and computational requirements) for the prediction of MIR spectra.

As we have seen, the choice of an appropriate model to develop a MIR tool is not trivial. It is wise to start with simple models that are easy to deploy as a baseline to understand our problem better and inform us on the best way to approach it. Once we have built a baseline model, we are in the best position to decide which steps to take next to attempt to improve it. Regarding the calibration of MIR prediction equations, classical methods (e.g., PLS) are most often used, but it is worth exploring other algorithms. Besides, when developing a model at the research stage, it is important to think about the potential practical applications and build a model that is easy to implement (e.g., not too computationally demanding).

2.3.2. Model validation

Validation is the process by which model outputs are compared to independent real-world observations to judge the correspondence with reality and confirm that the model actually achieves its intended purpose (Jager, 2016). This is, therefore, an important part of model development.

Many MIR prediction equations performances presented in scientific papers are based on the commonly-applied random cross-validation (i.e., a random subset, or subsets, of data is excluded from model training and used for evaluation), especially at the first stages of the development of the model and when the dataset is too small to carry out external validation (e.g., Dehareng et al., 2012; Bonfatti et al., 2016; Grelet et al., 2019). However, this technique is sub-optimal and might lead to dependencies between the training and validation sets, for instance if there are multiple records per animal as it is often the case in models involving MIR data or if there are unaccounted experimental treatment structures within the data (e.g., Shetty et al., 2017; Wang and Bovenhuis, 2019). If different records of the same cow appear both in the training and validation sets, this could lead to an overoptimistic evaluation of model accuracy because the training sets learn from data that have similarities with the validation sets (i.e., validation data are not entirely new). In this situation, block cross-validation, where data are split strategically rather than randomly, would be preferred. Therefore, in Chapter 2, we decided to carry out random cow-independent 10-fold cross-validation (i.e., for each fold, cows in the training set were different from cows in the validation set) because of repeated records per cow in the dataset. Comparison of the results for PLS-DA with classic random 10-fold cross-validation and random cow-independent 10-fold cross-validation is presented in Table 5-2 for scenario 2 (i.e., spectral differences, cf. Chapter 2). This scenario was selected because of more distinct differences between both cross-validation methods. The results clearly illustrate the overoptimistic performances of classical random 10-fold cross-validation.

Table 5-2. Strategy 2 (spectral differences, cf. Chapter 2): comparison of the results of partial least squares discriminant analysis with classical random 10-fold cross-validation and random cow-independent 10-fold cross-validation (i.e., block cross-validation)[1]

	Random cow-independent 10-fold cross-validation	Random 10-fold cross-validation
AUC	0.59	0.69
Sensitivity	0.58	0.64
Specificity	0.54	0.65

[1] AUC = area under the receiver operating characteristic curve; sensitivity = proportion of records belonging to pregnant cows that were correctly classified as pregnant; specificity = proportion of records belonging to open cows that were correctly classified as open

In our study, we also performed external validation on a cow-independent test set consisting of cows that were not used to build up and tune the model. Ideally, implementing for instance a herd-independent external validation or herd- and year-independent external validation would have been better as cows from the same herd have common information that could both be found in the calibration sets and in the test sets (e.g., same diet, environmental or management practices), leading to dependencies. This is also a way to test the robustness of the model. Such external independent validation was for example implemented by Ho et al. (2019) and Wang and Bovenhuis (2019). However, this was difficult to implement properly in our study because of the little number of herds and the highly different number of records per herd. An issue was to be able to keep sufficient information in the calibration set to cover the variability encountered in the test set. This could be resolved using a larger and more diverse dataset, for example combining data from different countries.

In Chapter 3, the test-day mixed model for the prediction of MIR spectra was assessed in different situations (i.e., scenarios) with more or less information known about the cows. This can be considered as a kind of validation of the model in new real-world based situations (e.g., prediction of the production potential of heifers, prediction of future records). In addition to the assessment of the prediction quality of spectral wavenumbers, we applied MIR prediction equations for milk fat, protein, and lactose contents to the observed spectra and to the modeled spectra and we compared the predictions for the modeling dataset and the scenarios. This lets us know if MIR spectral wavenumbers were predicted with sufficient accuracy to enable subsequent satisfying predictions of MIR traits. However, milk fat, protein, and lactose contents are very accurately predicted milk components (i.e., R^2 of cross-validation (R^2cv) of 0.99, 0.99, and 0.91 for the equations used in Chapter 3, respectively). It would also be interesting to assess traits that are predicted with lower accuracy (e.g., novel MIR traits) to see if the modeled MIR spectra are also able to predict such traits satisfactorily. Therefore, to complement our results presented in Chapter 3 (Tables 3-3- and 3-4) and validate the model further, we examined the correlations between novel MIR traits predicted from observed spectra and from modeled spectra (Table 5-3). We selected the following traits: minor milk components (calcium (Ca), potassium (K)), indicators of metabolic diseases (citrate, beta-hydroxybutyrate (BHB)), an indicator of udder health (lactoferrin) and fatty acids (omega-3 fatty acids and omega-6 fatty acids). Calibration equations of these novel traits originated from different research projects. They were developed using 212 spectral wavenumbers and their R^2cv ranged from 0.55 to 0.89. Complementary information about the equations is available in Table A-1 in the Appendix.

Table 5-3. Correlations between novel MIR traits predicted from the observed spectra and from the modeled spectra for the modeling dataset and the 4 scenarios. For comparison, the results for fat, protein, and lactose contents (extracted from Table 3-3), were added in the first 3 columns

Trait[1]	Fat	Protein	Lactose	Ca	K	Citrate	BHB	Lactoferrin	Omega-3	Omega-6
R^2cv[2]	0.99	0.99	0.91	0.82	0.55	0.89	0.71	0.72	0.66	0.72
Model. set	0.83	0.89	0.83	0.77	0.77	0.83	0.64	0.67	0.74	0.72
Scen. 1	0.63	0.68	0.63	0.52	0.54	0.62	0.40	0.35	0.50	0.51
Scen. 2	0.36	0.62	0.46	0.45	0.30	0.34	0.26	0.28	0.35	0.31
Scen. 3	0.53	0.73	0.46	0.58	0.58	0.48	0.39	0.57	0.60	0.54
Scen. 4	0.40	0.64	0.32	0.51	0.51	0.28	0.21	0.42	0.40	0.34

[1] Ca=calcium, K=potassium, BHB= beta-hydroxybutyrate

[2] R^2cv = R^2 of cross-validation of the prediction equation for each trait (cf. Table A-1 in Appendix for additional information)

Table 5-4. Mean (SD into parentheses) for calcium (Ca), potassium (K), citrate, beta-hydroxybutyrate (BHB), lactoferrin, omega-3 fatty acids, omega-6 fatty acids predicted from observed and from modeled spectra for the modeling dataset and the 4 scenarios

	Ca (mg/kg)		K (mg/kg)		Citrate (mmol/L)		BHB (µmol/L)		Lactoferrin (mg/L)		Omega-3 (g/dl)		Omega-6 (g/dl)	
	Obs. spectra	Pred. spectra	Obs. spectra	Pred. spectra	Obs. spectra	Pred. spectra	Obs. spectra	Pred. spectra	Obs. spectra	Pred. spectra	Obs. spectra	Pred. spectra	Obs. spectra	Pred. spectra
Modeling	1178	1174	1512	1514	9.12	9.10	2.25	2.24	142.29	143.32	0.026	0.026	0.102	0.102
set	(114.25)	(85.86)	(97.75)	(72.76)	(1.41)	(1.10)	(0.11)	(0.07)	(99.97)	(72.96)	(0.006)	(0.004)	(0.017)	(0.012)
Scen. 1	1248	1237	1488	1468	8.93	8.66	2.29	2.27	163.49	192.78	0.028	0.026	0.106	0.104
	(123.08)	(75.03)	(101.68)	(59.22)	(1.50)	(1.00)	(0.11)	(0.06)	(93.85)	(53.27)	(0.006)	(0.003)	(0.017)	(0.010)
Scen. 2	1169	1166	1567	1549	8.87	8.78	2.25	2.25	108.37	110.61	0.024	0.023	0.100	0.100
	(95.38)	(66.94)	(83.43)	(48.47)	(1.41)	(0.91)	(0.11)	(0.05)	(86.40)	(59.13)	(0.005)	(0.003)	(0.017)	(0.010)
Scen. 3	1175	1169	1514	1531	9.06	9.14	2.25	2.24	125.06	110.87	0.026	0.025	0.106	0.103
	(119.37)	(83.68)	(104.29)	(72.85)	(1.47)	(0.87)	(0.12)	(0.05)	(113.22)	(79.76)	(0.007)	(0.005)	(0.020)	(0.012)
Scen. 4	1172	1183	1510	1504	9.15	9.08	2.26	2.25	147.10	157.18	0.026	0.026	0.102	0.102
	(120.23)	(62.22)	(101.90)	(53.29)	(1.48)	(0.40)	(0.12)	(0.03)	(101.71)	(50.93)	(0.006)	(0.003)	(0.018)	(0.062)

Regarding the modeling set, we see that correlations between traits predicted from observed and modeled spectra were somewhat lower for the novel traits compared to fat, protein, and lactose contents, except for citrate (Table 5-3). Correlations for the scenarios tended to be lower too, but it was not always the case. For example, K has the lowest R^2cv but, regarding some scenarios, correlations between traits predicted from observed or modeled spectra were higher than some other traits. One reason is that the different traits are predicted using different spectral regions which can be more or less well modeled. In addition, we examined the mean values and standard deviation for each trait predicted from observed spectra or modeled spectra (Table 5-4). Mean values were close, although the standard deviation tended to be smaller for traits predicted from modeled spectra, as also observed for fat, protein, and lactose contents in Chapter 3 (Table 3-4). These mean and standard deviation values were also in the same range as reference values used to build the prediction equations (i.e., they were within normal ranges; cf. Table A-1 in Appendix).

As mentioned several times in Chapter 3 and the discussion, a valuable perspective of test-day models for management is to detect problems by comparisons between observations and expectations (i.e., modeled values). Validation of such a forecasting tool will require direct comparison with the happenings on farms. Reliable field records on individual animal health, feed, herd management, and environmental conditions will be essential to link the deviations between observed and modeled records to modeling inaccuracies or real cow or herd problems.

In brief, this section emphasizes the importance of proper validation to avoid overoptimistic results and to confirm that the model achieves its intended purpose for practical use.

2.4. Implementation of MIR tools

This section addresses some considerations regarding the implementation of developed MIR tools, more specifically the utilization of MIR indicators, the studies of MIR traits in relation with other data streams, the timing of milk sampling, and the uptake of MIR tools by dairy farmers.

2.4.1. On the utilization of MIR indicators

As previously mentioned, most MIR traits related to cow physiological status or indirect traits (e.g., pregnancy in Chapter 2 and CH_4 in Chapter 4), but also some minor milk components, are not accurate enough to be proposed for analytic purposes, i.e., they cannot replace the reference methods. However, some of these traits can be used for instance as indicators for screening in routine, for acquiring phenotypes for research purposes or as phenotypic information to genetically improve the trait in question on a large scale (Visentin et al., 2016). Using single indicators for management purposes is acceptable when used at the herd level (i.e., average of a group) because, under the hypothesis that prediction errors are random, the accuracy of predictions for the group will increase (Gengler et al., 2016). However, we should be careful at the individual level. It is for example wiser to use several indicators and not just one to avoid imprecisions and faulty conclusions or diagnosis, or use relative scales or thresholds instead of a quantitative value that could be misleading. For example, in Chapter 4 of the present thesis, the moderate accuracy of the MIR CH_4 prediction equation means that MIR CH_4 is considered as an indicator of CH_4 production. The use of this indicator was appropriate in Chapter 4 because individual MIR CH_4 predictions were pooled at the herd and year level before comparison with technico-economic data expressed at the same scale. If used at the individual level, an idea could be to assign a class (i.e., relative scale) to MIR CH_4 predictions, e.g., low emitters – moderate emitters – high emitters, instead of a quantitative value. In Chapter 2, if further studies confirm the possibility to develop a tool to detect pregnancy loss in mid- to late pregnancy, this could be used as an indicator in routine alerting the farmer when there is a risk that a cow aborted (i.e., screening tool in routine). This way, the farmer could verify the assumption of the MIR prediction by an accurate method such as rectal palpation. Regarding genetic evaluations, MIR indicators are particularly important in the case of traits that are expensive and difficult to measure directly such as CH_4 production or health traits (Gengler et al., 2016; Gengler and Soyeurt, 2020). In this context, the genetic correlation with the trait of interest is important, rather than the phenotypic correlation. It has been demonstrated that low prediction accuracy (e.g., $R^2 < 0.50$) in a chemometrical and management context can be considered sufficient to extract useful spectral variation in an animal breeding context (Gengler and Soyeurt, 2020).

2.4.2. The interest to study MIR traits in the population

Studying MIR traits evolution and behavior in a population, as well as relationships with other variables, is a good way to demonstrate their usefulness and learn more about their sources of variation or associated management practices (cf. Chapter 4). This could provide information for decision-making, for example regarding the adjustment of management practices to influence a MIR trait of interest. Although such large-scale population studies can be considered as strengths of MIR due to the availability and quantity of spectral data (cf. section 1.3 of this Chapter), an issue is that large-scale studies combining novel MIR traits with external data to milk recording or milk payment databases such as economic or management data are not often considered in the literature. Examples of phenotypic population-level studies with novel MIR traits, using only milk recording or milk payment data or in combination with management data, are described by Santschi et al. (2016) for BHB, Visentin et al. (2018) for mineral components, and Woolpert et al. (2016) for fatty acids. Vanlierde et al. (2016) previously demonstrated the usefulness of a developed MIR CH_4 prediction by applying it on spectra from the Walloon milk recording and revealed, among other, interesting individual, herd, seasonal and regional patterns. They suggested that further explanations were needed to explain differences in CH_4 emissions between Walloon herds. This was a reason for the research presented in Chapter 4. In this study, relationships between MIR CH_4 and technico-economic data were analyzed globally and results showed general trends, but studying these relationships between and within different groups with similar characteristics (typology) might strengthen or highlight some interesting tendencies (Dalcq et al., 2018). This would ideally require a larger amount of data. Among others, the access to accounting data from other accounting associations could be useful to increase the number of data and be more representative of the Walloon Region of Belgium, but standardization of accounting data from different sources would be needed. Databases from feed companies could also bring useful and more detailed information about feeding and diet composition for large-scale studies in combination with MIR data, but an obstacle is to make them available for research institutions.

2.4.3. Timing of milk sampling

A disadvantage of the current usage of milk recording to obtain individual MIR spectra is that samples are collected on a monthly basis and not in real time, in contrast to some on-farm precision tools providing daily monitoring. This is a limit when rapid and frequent monitoring is needed in the context of individual management on farms, for example to quickly detect troubles and find prompt solutions. In the case of health or status-traits prediction, metabolic changes have to occur at the right moment to be detectable in milk if milk tests are distant. For example, timely detection of early-lactation diseases like ketosis requires MIR data early in the lactation (Luke et al., 2019); pregnancy or abortion detection (cf. Chapter 2) would require more frequent recordings to take prompt decisions. Increasing the milk sampling frequency in the framework of classical milk recording or even altering milk recording procedures to focus on some critical periods of the lactation would be ideal (e.g., analyses from postpartum cows may provide more valuable information than analyses from cows in late lactation). However, it would be more costly, and farmers will be willing to alter their milk-testing practices only if the benefits outweigh the costs. More frequent testing of postpartum cows might also be technically rather impractical because, in many herds, cows calve at different periods and are at all times at different lactation stages, making the organization of milk testing difficult.

On-farm daily milk analysis would provide high-throughput valuable information, but until now, MIR has not been used on farms for real-time monitoring for economic and technical reasons. MIR technology is still expensive and, practically speaking, the MIR spectrometers need a very thin homogeneous layer of milk, which would be a challenge in on-line systems, and regular reference measurements would be needed to ensure accurate analyses (Waaben, 2016). The possibility that these technical limits will be resolved in the future and enable MIR analysis on the farm cannot be excluded. Nevertheless, because of these current limits, other technologies better adapted to daily on-farm measurements, such as near-infrared (NIR) spectrometry, are currently under investigation (e.g., Melfsen et al., 2012; Kaniyamattam and De Vries, 2014; Hanus et al., 2016). Advantages of NIR on-farm systems compared to MIR are among others their lower cost, robustness to tougher conditions, and easier on-line sampling of milk (Waaben, 2016). Drawbacks of NIR are that (1) the signal is not as strong as MIR, (2) the accuracy and repeatability are not as good as MIR, and (3) as mentioned in section 1.2 of this Chapter, it is more difficult to associate and compare data between farms and to standardize process among apparatus (Kaniyamattam and De

Vries, 2014; Waaben, 2016). Even if on-farm milk testing systems are implemented successfully in the future, official MIR analyses in a laboratory will most probably still be needed to benchmark and correct on-farm systems for deviations. Hence, on-farm and off-farm systems will most likely be complementary.

In contrast to milk recording, bulk tank milk samples are collected and analyzed more often (i.e., at every milk collection on the farm) via milk payment systems. Despite the loss of individual animal monitoring, this opens a door for the development of herd-level decision-support tools fed with higher throughput information, providing that bulk tank data are available for research and development institutions. For example, observing values for specific traits predicted from bulk tank spectra or modeling periodic trends in comparison with other herds (for benchmarking) is a way to identify poor or superior performances and detect problems. Using bulk tank milk data allows to have access to information about all dairy herds in a research context and would provide management tools for all herds, in contrast to milk recording data that are limited to herds enrolled in official milk recording systems (e.g., approximately 30% in the Walloon region of Belgium; awé groupe, 2018). However, a disadvantage of bulk tank milk is that the quantity and quality of milk in the tank are influenced by high-yielding cows (i.e., the milk composition is the weighted average by milk production of individual cow milk compositions). For example, a cow producing 30 litres of milk on a specific day will contribute twice more to bulk tank milk than a cow producing 15 litres of milk. Consequently, we should be aware that MIR traits predictions from bulk tank milk would not represent the average herd, but would be drawn towards cows producing more milk, and conclusions based on bulk tank milk may be biased for some traits or more difficult to interpret correctly. Also, some MIR equations were developed for individual spectra and are less adapted to bulk samples. This is the case with the equation for enteric CH_4 production used in Chapter 4, which takes into account the DIM of the cow (i.e., the evolution of the trait along the lactation) and is therefore designed for individual milk spectra (Vanlierde et al., 2016). It could be interesting to investigate the possibility to adapt such equations to bulk tank milk to be able to predict a global value for the herd with a single analysis of tank milk. Moreover, novel traits such a pregnancy (cf. Chapter 2), body weight (Soyeurt et al., 2019), or likelihood of conception (Ho et al., 2019) are only interesting at the individual level.

In short, MIR tools used in the framework of classical milk recording may be limited by the discontinuous periodic milk sampling. On-farm milk analysis would be a promising solution, but it is still challenging at the moment. Bulk tank milk analysis provides high-throughput information, but it is limited to herd management under certain conditions.

2.4.4. Uptake and acceptability of MIR tools by the farmers

Although research institutions have generated many MIR tools, of which some are suitable for implementation in the field, on-farm integration and acceptability by the farmers sometimes remain difficult and limit the number of tools effectively used in practice. Farmers will be more likely to adopt and pay for a specific tool if direct or indirect economic interests have been demonstrated. To illustrate this in the framework of this thesis, it is unlikely that farmers will pay for the prediction of CH_4 without economic incentives for the reduction of greenhouse gas emissions or the tangible proof that reducing CH_4 production will improve economic returns, even though CH_4 can be considered as a direct loss of gross energy (2-12%; Johnson and Johnson, 1995). Likewise, farmers selling their milk to a dairy plant will most probably show little interest in MIR predictions for milk technological properties without incentives, in contrast to farmers processing milk on the farm. However, beyond deliberate choice, farmers could be forced by governments or dairy industries to adopt some MIR tools, for example in the context of milk quality or environmental controls or regulations. For instance, assessing CH_4 production using MIR predictions could be an obligation in the future to monitor the carbon footprint of dairy products. Such MIR traits related to quality insurance or environment were often first developed following industry or governmental interests, in contrast to traits related for example to cow health or fertility that are rather a request from farmers.

Moreover, it is useful to inform farmers about the benefits of some novel MIR traits and tools that they may be too little aware of. An example is the fatty acid profiles that could help with better feeding strategies decisions, but that is used in a minority of countries (Valacta, 2018). The wish to predict fatty acids originated from an analytics company providing MIR spectrometers at first and not directly from dairy advisors, which is also a reason why some MIR tools are less used in practice. Sub-optimal transfer of information and knowledge about potential MIR decision-support tools, as well as their little visibility, constitutes an obstacle to their implementation in practice. In the future, proposing a clear document or online platform to farmers listing tools available in the framework of their national or regional milk recording and milk payment systems, with the associated price, reliability, explanation of benefits, etc. would give them the opportunity to choose tools that would fit best to their individual needs. Regarding MIR trait predictions, these could be grouped in packages corresponding to different objectives (e.g., milk processing, milk quality, health indicators, feed monitoring, etc). For instance, in the Walloon region of Belgium, a package with several MIR predictions related to milk processing (i.e., milk technological properties and quality) is under study to be included in the milk recording reports transferred to farmers (C. Bastin, 2020, personal communication).

When transferring a tool to farmers for implementation, we need to be careful about how it would shape actions that farmers should make. We should not only transfer the tool as it is, but we should also give some advice regarding the actions to make or how to adapt management practices depending on the results obtained from the tool.

3. Conclusion and perspectives

Good decision-making regarding actions to make on farms is required to maintain satisfying economic performances and limit environmental impacts in dairy farming. Analysis of milk by MIR spectrometry provides a large amount of information on milk composition. This technique has demonstrated qualities to support decision-making, for example through the well-established predictions of milk fat and protein contents or the recent development of prediction models for novel traits. However, its full potential remains partly uninvestigated. Hence, the objective of this thesis was to contribute to the development of decision-support tools with economic and environmental interests for the dairy sector using milk MIR spectrometry.

In practical terms, the different approaches for pregnancy diagnosis using a MIR prediction equation (Chapter 2) unfortunately failed to detect satisfactorily early pregnancy stages that are of great importance in dairy farming. However, a first tool arising from the findings of this research could be a screening tool for the detection of mid- to late-term abortion in individual cows, because we obtained promising results for the prediction of advanced pregnancy. Such a screening tool could be implemented in routine in official milk recording programs to alert the farmers when there is a risk that a cow aborted several months after an insemination, so that they can identify problem cows more easily and find prompter solutions. The next step would be to validate the model on a larger and more variable dataset to confirm the obtained results and explore further the prediction of this trait. More generally, further improvement to models aiming to predict novel traits (e.g., health and wellness traits) using MIR could be made by incorporating zootechnical parameters routinely recorded on farms such as milk yield, parity or DIM, or other easily available data streams such as information from rumination collards and accelerometers. However, the inclusion of several data sources increases the technical difficulty for data consolidation and quality control (especially for data from on-farm tools). In the recent years, there has also been interest in combining phenotypic data like MIR traits and genetic information to help improve prediction accuracy (Ho et al., 2019; van der Heide et al., 2020; Wang and Bovenhuis, 2020). Besides, using omics approaches (e.g., metabolomics or proteomics) may elucidate some relationships between novel traits and milk composition or, if appropriate, they may be combined with MIR to improve prediction models (e.g., Chapinal et al., 2012; Melzer et al., 2013; Goldansaz et al., 2017). If large datasets are available for calibration, it is worth exploring deep learning algorithms that can discover new features in complex and high dimensional datasets (LeCun et al., 2015). In addition, when combining different types of data, the relationships between predictor and response variables may be more complicated than linear. This puts more emphasis on the need to explore non-linear machine and deep learning algorithms.

A second tool with potential future application in practice would be the test-day model to predict MIR spectra (Chapter 3). It would rather be aimed at advisors (who can interpret results and communicate them to farmers) and researchers. Examples of utilization of this tool are the comparison between traits predicted from the expected (i.e., modeled) and observed spectra for the detection of anomalies at the herd or cow level (this could for example result in economic savings) or simulations of modifications of effects in the model (e.g., to assess the evolution of environmental-related traits such as MIR CH_4 or urea). Different improvements of the model are needed before potential applications. The first enhancements to make would be the use of random regression coefficients, adding second and third lactations in the model, and testing the possibility to use a multi-trait model including several principal components without being computationally too excessive. If used as a forecasting tool to detect problems, the model should be validated using reliable field recordings to identify whether deviations are due to modeling problems or to herd or cow problems.

The last research study (Chapter 4) used an existing MIR tool (i.e., MIR CH_4 prediction equation applied to a large database) in combination with technico-economic data to provide additional information for decision-making. This is not a new tool per se, but rather a study including an existing tool and with an objective of decision support. Results of this study are primarily dedicated for research purposes to better understand management practices associated with CH_4 production, but these could contribute to guiding potential decisions at the political level (e.g., which practices should be supported on farms to reduce CH_4 production?) and subsequently affect actions to make at the farm level. Given the obtained results, revealing complex relationships and low correlations between MIR-predicted CH_4 production and technico-economic variables, it is difficult to make any clear choice regarding management practices that should be adopted on farms to reduce CH_4 production. This implies that further studies, with real farm data, are needed to better understand the factors associated with CH_4 emissions and its impact on profitability in order to better target mitigation strategies (e.g., studies with more precise data about diets and feed composition or segmenting the data into different typology groups). As additional perspective, MIR CH_4 production could also be associated with other environmental variables or indicators to carry out a more comprehensive study on relationships between dairy farm environmental footprint and technico-economic factors. Indeed, MIR CH_4 covers only partly the environmental impacts of milk production because CH_4 emissions from young stock, other greenhouse gas emissions, carbon sinks, and other environmental aspects (e.g., eutrophication, biodiversity conservation) are not considered. In our study, we considered MIR CH_4 expressed in g/day, but CH_4 emissions can be expressed in different metrics, which have different interpretations (e.g., g/day, g/kg dry matter intake, or g/kg milk). The unit of CH_4 measurement to consider depends mainly on the specific research question and application, but there is no general concencus and research efforts have been concerned with all of these metrics.

In addition to the three research studies using MIR for decision support, this thesis highlighted strengths of MIR spectrometry for the development of decision-support tools for the dairy sector, but also raised some issues and consideration in the light of the work conducted in the framework of this thesis. MIR spectrometry is a non-invasive, fast, and cost-effective technology, allowing the collection of individual and herd MIR data on a large scale. The standard procedures for milk sample collection and analysis, and the storage of MIR data in central databases facilitate the comparison of data within and between herds. These advantages enable the development of various customized tools and studies requiring large amounts of comparable data, which can support decision-making in dairy farming with impacts from the short to the long term. Still, MIR has some limitations, and considerations need to be accounted for at the development and implementation stages of decision-support tools. Among others, issues and considerations were raised about the prediction of indirect MIR traits, the quality and variability of spectral and reference data, the choice and validation of models, the utilization of MIR indicators, the study of MIR traits in the population, the timing of milk sampling, and the uptake of MIR tools by farmers. As we have seen, the development of MIR tools is not straightforward and is a lengthy process that requires reflection at every step.

To conclude, this thesis contributed to the first steps of the development of MIR tools and studies to support decision-making with potential economic and environmental benefits for the dairy sector; and contributed to stimulate thinking and gain insight into the benefits and points of attention regarding the development of decision-support tools using milk MIR spectrometry. This thesis is not an end to itself: additional studies are required to refine the research developed in this manuscript for potential practical applications, and it is worth continuing to explore deeper the potential of milk MIR spectrometry to provide innovative and valuable tools for progressing towards more sustainable dairy farming.

4. References

Andersen, C.M., and R. Bro. 2010. Variable selection in regression – a tutorial. J. Chemometrics 24: 728–737. https://doi.org/10.1002/cem.1360.

Anzanello, M.J., and F.S. Fogliatto. 2014. A review of recent variable selection methods in industrial and chemometrics applications. European J. of Industrial Engineering 8:619-645. https://doi.org/10.1504/EJIE.2014.065731.

awé groupe, 2018. Rapport d’activités 2018. http://www.awenet.be/awe/commun/asbl/qui/rapport.php.

awé groupe. 2020. Simulateur de Prix Par Contrôle. Accessed Jan. 24, 2020. http://www.awenet.be/awe/commun/asbl/lait/services.php?language=nl.

Bahl, A., B. Hellack, M. Balas, A. Dinischiotu, M. Wiemann, J. Brinkmann, A. Luch, B.Y. Renard, A. Haase, G. Federal, R. Assessment, and P. Safety. 2019. Nanoimpact recursive feature elimination in random forest classification supports nanomaterial grouping. NanoImpact 15:100179. https://doi.org/10.1016/j.impact.2019.100179.

Barbano, D.M., and J.L. Clark. 1989. Symposium: instrumental method for measuring components of milk. Infrared milk analysis – challenges for the future. J. Dairy Sci. 72:1627–1636. https://doi.org/10.3168/jds.S0022-0302(89)79275-4.

Bastin, C., L. Laloux, A. Gillon, F. Miglior, H. Soyeurt, H. Hammami, C. Bertozzi, and N. Gengler. 2009. Modeling milk urea of Walloon dairy cows in management perspectives. J. Dairy Sci. 92:3529–3540. https://doi.org/10.3168/jds.2008-1904.

Berry, D.P., S. Mcparland, C. Bastin, E. Wall, N. Gengler, and H. Soyeurt. 2013. Phenotyping of robustness and milk quality. Animal 4:600–605. https://doi.org/10.1017/S2040470013000150.

Bewley, J. 2010. Precision dairy farming: advanced analysis solutions for future profitability. The First North American Conference on Precision Dairy Management 2010, Toronto, Ontario, USA.

Bittante, G., and A. Cecchinato. 2013. Genetic analysis of the Fourier-transform infrared spectra of bovine milk with emphasis on individual wavelengths related to specific chemical bonds. J. Dairy Sci. 96:5991–6006. https://doi.org/10.3168/jds.2013-6583.

Bonfatti, V., G. Di Martino, and P. Carnier. 2011. Effectiveness of mid-infrared spectroscopy for the prediction of detailed protein composition and contents of protein genetic variants of individual milk of Simmental cows. J. Dairy Sci. 94:5776–5785. https://doi.org/10.3168/jds.2011-4401.

Bonfatti, V., L. Degano, A. Menegoz, and P. Carnier. 2016. Short communication: Mid-infrared spectroscopy prediction of fine milk composition and technological properties in Italian Simmental. J. Dairy Sci. 99:8216–8221. https://doi.org/10.3168/jds.2016-10953.

Bonfatti, V., A. Fleming, A. Koeck, and F. Miglior. 2017. Standardization of milk infrared spectra for the retroactive application of calibration models. J. Dairy Sci. 100:2032–2041. https://doi.org/10.3168/jds.2016-11837.

Bonfatti, V., P.N. Ho, and J.E. Pryce. 2020. Usefulness of milk mid-infrared spectroscopy for predicting lameness score in dairy cows. J. Dairy Sci. 103:2534-2544. https://doi.org/10.3168/jds.2019-17551.

Bucek, P., K. Zottl, F. Onken, D. Radzio, Z. Barac, Q. Testing, D. Farmers, C.A. Agency, and S. Republic. 2014. Practical aspects in milk recording in Central and Eastern Europe and its effects on the guidelines. 39th ICAR Session, Berlin, Germany.

Canadian Dairy Information Centre. 2018. Enrollments on milk recording. Accessed Feb. 10, 2020. https://dairyinfo.gc.ca/index_e.php?s1=dff-fcil&s2=mrr-pcle&s3=emr-ipcl&menupos=01.01.12.

Castro, F.M., M.J. Mar, and C. Schmid. 2018. End-to-end incremental learning. European Conference on Computer Vision 2018, Munich, Germany.

Chapinal, N., M.E. Carson, S.J. Leblanc, K.E. Leslie, S. Godden, M. Capel, and J.E.P. Santos. 2012. The association of serum metabolites in the transition period with milk production and early-lactation reproductive performance. J. Dairy Sci. 95:1301–1309. https://doi.org10.3168/jds.2011-4724.

De Marchi, M., V. Toffanin, M. Cassandro, and M. Penasa. 2014. Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits. J. Dairy Sci. 97:1171–1186. https://doi.org/10.3168/jds.2013-6799.

De Vries, A. 2006. Economic value of pregnancy in dairy cattle J. Dairy Sci. 89:3876–3885. https://doi.org/10.3168/jds.S0022-0302(06)72430-4.

Dagnachew, B.S., A. Kohler, and T. Ådnøy. 2013. Genetic and environmental information in goat milk Fourier transform infrared spectra. J. Dairy Sci. 96:3973–3985. https://doi.org/10.3168/jds.2012-5972.

Dalcq, A., Y. Beckers, P. Mayeres, E. Reding, B. Wyzen, F. Colinet, P. Delhez, and H. Soyeurt. 2017. The feeding system impacts relationships between calving interval and economic results of dairy farms. Animal 12: 1662-1671. https://doi.org/10.1017/S1751731117003020.

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

Dijkstra, J., S.M. Van Zijderveld, J.A. Apajalahti, A. Bannink, and W.J.J. Gerrits. 2011. Relationships between methane production and milk fatty acid profiles in dairy cattle. Anim. Feed Sci. Technol. 166–167:590–595. https://doi.org/10.1016/j.anifeedsci.2011.04.042.

Engel, J., J. Gerretzen, E. Szymanska, J.J. Jansen, G. Downey, L. Blanchet, and L.M.. Buydens. 2013. Breaking with trends in pre-processing? Trends Anal. Chem. 50:96–106. https://doi.org/10.1016/j.trac.2013.04.015.

Eskildsen, C.E., M.A. Rasmussen, S.B. Engelsen, L.B. Larsen, N.A. Poulsen, and T. Skov. 2014. Quantification of individual fatty acids in bovine milk by infrared spectroscopy and chemometrics: understanding predictions of highly collinear reference variables. J. Dairy Sci. 97:7940–7951. https://doi.org/10.3168/jds.2014-8337.

FIDOCL Conseil Elevage. Ori-automate. Accessed Feb. 10, 2020. http://www.fidocl.fr/content/ori-automate.

Forman, G., and I. Cohen. 2004. Learning from little: comparison of classifiers given little training. 8th European conference on principles of data mining and knowledge discovery, Pisa, Italy.

Gengler, N., H. Soyeurt, F. Dehareng, C. Bastin, F. Colinet, H. Hammami, M.L. Vanrobays, A. Lainé, S. Vanderick, C. Grelet, A. Vanlierde, E. Froidmont, and P. Dardenne. 2016. Capitalizing on fine milk composition for breeding and management of dairy cows. J. Dairy Sci. 99:4071–4079. https://doi.org.10.3168/jds.2015-10140.

Gengler, N. 2019. Symposium review: challenges and opportunities for evaluating and using the genetic potential of dairy cattle in the new era of sensor data from automation. J. Dairy Sci. 102:5756–5763. https://doi.org/10.3168/jds.2018-15711.

Gengler, N., and H. Soyeurt. 2020. The use of mid-infrared spectral data to predict traits for genetic selection in dairy cattle. Pages 289-312 in Advances in breeding of dairy cattle. J. van der Werf and J. Pryce, ed. Burleigh Dodds Science Publishing Limited, Cambridge, UK.

GGI-Spermex. Independent Milk Recording. Accessed Jan. 24, 2020. http://ggi.de/en/holstein-breeding/organization/.

Goldansaz, S.A., A.C. Guo, T. Sajed, M.A. Steele, G.S. Plastow, and D.S. Wishart. 2017. Livestock metabolomics and the livestock metabolome : A systematic review. PLoS One. 12:1-26. https://doi.org/10.1371/journal.pone.0177675.

Grainger, C., T. Clarke, S.M. McGinn, M.J. Auldist, K.A. Beauchemin, M.C. Hannah, G.C. Waghorn, H. Clark, and R.J. Eckard. 2007. Methane emissions from dairy cows measured using the sulfur hexafluoride (SF6) tracer and chamber techniques. J. Dairy Sci. 90:2755–2766. https://doi.org/10.3168/jds.2006-697.

Granitto, P.M., C. Furlanello, F. Biasioli, and F. Gasperi. 2006. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. Chemom. Intell. Lab. Syst. 83:83–90. https://doi.org/10.1016/j.chemolab.2006.01.007.

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

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

Grelet, C., J.A.F. Pierna, P. Dardenne, H. Soyeurt, A. Vanlierde, F. Colinet, C. Bastin, N. Gengler, V. Baeten, and F. Dehareng. 2017a. 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., A. Vanlierde, E. Froidmont, GPlusE Consortium, and F. Dehareng. 2017b. Prediction of energy status of dairy cows using MIR milk spectra. Proceedings in: EAAP annual conference, 28 Aug. 2017, p 403, Tallin, Estonia.

Grelet, C. 2019. Promoting international prediction models through standardization of milk mid-infrared spectra. Doctoral thesis. Gembloux Agro-Bio Tech, Université de Liège, Belgium.

Grelet, C., A. Vanlierde, M. Hostens, L. Foldager, M. Salavati, K.L. Ingvartsen, M. Crowe, M.T. Sorensen, E. Froidmont, C.P. Ferris, C. Marchitelli, F. Becker, T. Larsen, F. Carter, G. Consortium, and F. Dehareng. 2019. Potential of milk mid-IR spectra to predict metabolic status of cows through blood components and an innovative clustering approach. Animal. 13:649–658. https://doi.org/10.1017/S1751731118001751.

Hanuš, O., L. Stádník, M. Klimešová, M. Tomáška, L. Hasoňová, D. Falta, J. Kučera, J. Kopecký, R. Jedelská, F. Daniel, K. Josef, K. Jaroslav, and J. Radoslava. 2016. The evaluation of real time milk analyse result reliability in the Czech Republic. Acta Universitatis Agriculturae et silviculturae mendelianae brunensis. 64:1155–1166.

Hempstalk, K., S. Mcparland, and D.P. Berry. 2015. Machine learning algorithms for the prediction of conception success to a given insemination in lactating dairy cows. J. Dairy Sci. 98:5262–5273. https://doi.org/10.3168/jds.2014-8984.

Ho, P.N., V. Bonfatti, T.D.W. Luke, and J.E. Pryce. 2019. Classifying the fertility of dairy cows using milk mid-infrared spectroscopy. J. Dairy Sci. 102:10460–10470. https://doi.org/10.3168/jds.2019-16412.

Iñón, F.A., S. Garrigues, and M. De La Guardia. 2004. Nutritional parameters of commercially available milk samples by FTIR and chemometric techniques. Anal. Chim. Acta 513:401–412. https://doi.org/10.1016/j.aca.2004.03.014.

ICAR. 2017. Section 2 - Guidelines for Dairy Cattle Milk Recording. Int. Comm. Anim. Record., Rome, Italy.

ICAR. 2019. Section 12 - Guidelines for Milk Analysis. Int. Comm. Anim. Record., Rome, Italy.

Jager, T. 2016. Dynamic modeling for uptake and effects of chemicals. Pages 71-98 in Marine ecotoxicology. J. Blasco, Chapman, P.M., Campana O., and M. Hampel, ed. Academic Press, London, UK.

Jamrozik, J., and L.R. Schaeffer. Estimates of genetic parameters for a test day model with tandom regressions for yield traits. J. Dairy Sci. 80:762–770. https:/doi.org/10.3168/jds.S0022-0302(97)75996-4.

Johnson, K.A., D.E. Johnson, and K.A. Johnson. 1995. Methane emissions from cattle. J. Anim. Sci. 73:2483-2492.

Kaniyamattam, K., and A. De Vries. 2014. Agreement between milk fat, protein , and lactose observations collected from the Dairy Herd Improvement Association (DHIA) and a real-time milk analyzer. J. Dairy Sci. 97:2896–2908. https://doi.org/10.3168/jds.2013-7690.

Kotsiantis, S.B., I.D. Zaharakis, and P.E. Pintelas. 2006. Machine learning: a review of classification and combining techniques. Artif Intell Rev 26:159–190. https://doi.org/10.1007/s10462-007-9052-3.

Kuhn, M. 2019. Package ‘caret’. R package version 6.0–84. R Foundation for Statistical Computing, Vienna, Austria.

Lecun, Y., Y. Bengio, and G. Hinton. 2015. Deep learning. Nature 251: 436-444. https://doi.org/10.1038/nature14539.

Lexico. n.d. Definition of accuracy. Accessed Feb. 12, 2020. https://www.lexico.com/en/definition/accuracy.

Luke, T.D., S. Rochfort, W.J. Wales, V. Bonfatti, L. Marett, and J.E. Pryce. 2019. Metabolic profiling of early-lactation dairy cows using milk mid-infrared spectra. J. Dairy Sci. 102:1747–1760. https://doi.org/10.3168/jds.2018-15103.

Lynch, J.M., D.M. Barbano, M. Schweisthal, and J.R. Fleming. 2006. Precalibration evaluation procedures for mid-infrared milk aanalyzers. J. Dairy Sci. 89:2761–2774. https://doi.org/10.3168/jds.S0022-0302(06)72353-0.

Marchewka, J., J. Lassen, L. Lukacs, G. Hoffmann, and R. Bodas. EIP-AGRI Focus Group. Robust and resilient dairy production systems mini-paper - The implication of precision livestock farming (PLF) on dairy farms robustness.

Mathematical Modelling Company. n.d. Robust mathematical modeling. Accessed Feb. 12, 2020. http://www.scmsa.com/RMM/robust1.htm.

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

Melfsen, A., E. Hartung, and A. Haeussermann. 2012. Accuracy of in-line milk composition analysis with diffuse reflectance near-infrared spectroscopy. J. Dairy Sci. 95:6465–6476. https://doi.org/10.3168/jds.2012-5388.

Melzer, N., D. Wittenburg, S. Hartwig, S. Jakubowski, U. Kesting, L. Willmitzer, and J. Lisec. 2013. Investigating associations between milk metabolite profiles and milk traits of Holstein cows. J. Dairy Sci. 96:1521–1534. https://doi.org/10.3168/jds.2012-5743.

Moss, A., J. Jouany, J. Newbold, A. Moss, J. Jouany, and J. Newbold. 2000. Methane production by ruminants: its contribution to global warming. Ann. Zootech. 2000:231-253.

Nguyen, D. V, and D.M. Rocke. 2002. Tumor classification by partial least squares using microarray gene expression data. Bioinformatics 18:39-50.

Olori, V.E., S. Brotherstone, W.G. Hill, and B.J. McGuirk. 1997. Effect of gestation stage on milk yield and composition in Holstein Friesian dairy cattle. Livest. Prod. Sci. 52:167–176. https://doi.org/10.1016/S0301-6226(97)00126-7.

Owen, A.J. 1995. Uses of derivative spectroscopy, UV-visible spectroscopy, application note. Agilent Technologies, Waldbronn, Germany.

Penasa, M., M. De Marchi, and M. Cassandro. 2016. Short communication: Effects of pregnancy on milk yield, composition traits, and coagulation properties of Holstein cows. J. Dairy Sci. 99:4864–4869. https://doi.org/10.3168/jds.2015-10168.

Pralle, R.S., K.W. Weigel, and H.M. White. 2018. Predicting blood β-hydroxybutyrate using milk Fourier transform infrared spectrum, milk composition , and producer-reported variables with multiple linear regression , partial least squares regression , and artificial neural network. J. Dairy Sci. 101:4378–4387. https://doi.org/10.3168/jds.2017-14076.

Ptak, E., and L. R. Schaeffer. 1993. Use of test day yields for genetic evaluation of dairy sires and cows. Livest. Prod. Sci. 34:23–34.

Rinnan, A., F. van den Berg, and S.B. Engelsen. 2009a. Review of the most common pre-processing techniques for near-infrared spectra. Trends Anal. Chem. 28:1201–1222. https://doi.org/10.1016/j.trac.2009.07.007.

Rinnan, Å., L. Nørgaard, F. Van Den Berg, J. Thygesen, R. Bro, and S.B. Engelse. 2009b. Data Pre-processing. Pages 29-50 in Infrared spectroscopy for food quality and control. D.W. Sun, ed. Academic Press, New York, USA.

Roche, J.R. 2010. Effect of pregnancy on milk production and bodyweight from identical twin Study. J. Dairy Sci. 86:777–783. https://doi.org/10.3168/jds.s0022-0302(03)73659–5.

Rousseeuw, P.J., M. Debruyne, S. Engelen, and M. Hubert. 2006. Robustness and outlier detection in chemometrics 242:221–242.

Rozenstein, O., T. Paz-Kagan, C. Salbach, and A. Karnieli. 2015. Comparing the effect of preprocessing transformations on methods of land-use classification derived from spectral soil measurements. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8:2393–2404. https://doi.org/10.1109/JSTARS.2014.2371920.

Santschi, D.E., R. Lacroix, J. Durocher, M. Duplessis, R.K. Moore, and D.M. Lefebvre. 2016. Prevalence of elevated milk β-hydroxybutyrate concentrations in Holstein cows measured by Fourier-transform infrared analysis in Dairy Herd Improvement milk samples and association with milk yield and components. J. Dairy Sci. 99:9263–9270. https://doi.org/10.3168/jds.2016-11128.

Shahinfar, S., D. Page, J. Guenther, V. Cabrera, P. Fricke, and K. Weigel. 2014. Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algorithms. J. Dairy Sci. 97:731–742. https://doi.org/10.3168/jds.2013-6693.

Shallue, C.J., J. Lee, J. Antognini, J. Sohl-Dickstein, R. Frostig, and G.E. Dahl. 2019. Measuring the effects of data parallelism on neural network training. J. Mach. Learn. Res. 20:1-49.

Shetty, N., G. Difford, J. Lassen, P. Løvendahl, and A.J. Buitenhuis. 2017. Predicting methane emissions of lactating Danish Holstein cows using Fourier transform mid-infrared spectroscopy of milk. J. Dairy Sci. 100:9052–9060. https://doi.org/10.3168/jds.2017-13014.

Socrates, G. 2001. Infrared and Raman characteristic group frequencies contents: Tables and charts. 3rd ed. Wiley, Chichester, New York, USA.

Soyeurt, H., P. Dardenne, F. Dehareng, G. Lognay, D. Veselko, M. Marlier, C. Bertozzi, P. Mayeres, and N. Gengler. 2006a. Estimating fatty acid content in cow milk using mid-infrared spectrometry. J. Dairy Sci. 89:3690–3695. https://doi.org/10.3168/jds.s0022-0302(06)72409-2.

Soyeurt, H., P. Dardenne, A. Gillon, C. Croquet, S. Vanderick, P. Mayeres, C. Bertozzi, and N. Gengler. 2006b. Variation in fatty acid contents of milk and milk fat within and across breeds. J. Dairy Sci. 89:4858–4865. https://doi.org/10.3168/jds.S0022-0302(06)72534-6.

Soyeurt, H., V. Arnould, D. Bruwier, P. Dardenne, J.M. Romnee, and N. Gengler, 2008. 2008 Joint ADSA-ASAS Annual Meeting. J. Anim. Sci. Vol. 86, E-Suppl. 2/J. Dairy Sci. Vol. 91, E-Suppl. 1. https://orbi.uliege.be/handle/2268/24971.

Soyeurt, H., D. Bruwier, J. Romnee, N. Gengler, C. Bertozzi, D. Veselko, and P. Dardenne. 2009. Potential estimation of major mineral contents in cow milk using mid-infrared spectrometry. J. Dairy Sci. 92:2444–2454. https://doi.org/10.3168/jds.2008-1734.

Soyeurt, H., E. Froidmont, I. Dufrasne, D. Hailemariam, Z. Wang, C. Bertozzi, F.G. Colinet, F. Dehareng, and N. Gengler. 2019. Contribution of milk mid-infrared spectrum to improve the accuracy of test-day body weight predicted from stage, lactation number, month of test and milk yield. Livest. Sci. 227:82–89. https://doi.org/10.1016/j.livsci.2019.07.007.

Storm, I.M.L.D., A.L. Hellwing, N.I. Nielsen, and J. Madsen. 2012. Methods for measuring and estimating methane emission from ruminants. Animals 2:160–183. https://doi.org/10.3390/ani2020160.

Swalve, H.H. 2000. Theoretical basis and computational methods for different test-day genetic evaluation methods. J. Dairy Sci. 83:1115–1124. https://doi.org/10.3168/jds.S0022-0302(00)74977-0.

Tiplady, K.M., R.G. Sherlock, M.D. Littlejohn, J.E. Pryce, S.R. Davis, D.J. Garrick, R.J. Spelman, and B.L. Harris. 2019. Strategies for noise reduction and standardization of milk mid-infrared spectra from dairy cattle. J. Dairy Sci. 102:6357–6372. https://doi.org/10.3168/jds.2018-16144.

Tremblay, M., M. Kammer, H. Lange, S. Plattner, C. Baumgartner, and J.A. Stegeman. 2019. Prediction model optimization using full model selection with regression trees demonstrated with FTIR data from bovine milk. Prev. Vet. Med. 163:14–23. https://doi.org/10.1016/j.prevetmed.2018.12.012.

van der Heide, E.M.M., R.F. Veerkamp, M.L. van Pelt, C. Kamphuis, and B.J. Ducro. 2020. Predicting survival in dairy cattle by combining genomic breeding values and phenotypic information. J. Dairy Sci. 103:556–571. https://doi.org/10.3168/jds.2019-16626.

van der Werf, n.d. Estimation of genetic parameters. https://jvanderw.une.edu.au/Estimation%20of%20Variance%20Components.pdf.

Valacta, 2018. La passion de faire plus grâce au profil des acides gras du lait. Pages 30-32 in Le producteur de lait Québécois, April 2018.

Vanderick, 2017. Improvement of genetic evaludation systems for maternally influenced traits and multi-breed livestock populations. Doctoral thesis. Gembloux Agro-Bio Tech, Université de Liège, Belgium.

Vanlierde, A., M. Vanrobays, F. Dehareng, E. Froidmont, H. Soyeurt, S. Mcparland, and E. Lewis. 2015. Hot topic: Innovative lactation-stage-dependent prediction of methane emissions from milk mid-infrared spectra. J. Dairy Sci. 98:5740–5747. https://doi.org/10.3168/jds.2014-8436.

Vanlierde, A., M.L. Vanrobays, N. Gengler, P. Dardenne, E. Froidmont, H. Soyeurt, S. McParland, E. Lewis, M.H. Deighton, M. Mathot, and F. Dehareng. 2016. Milk mid-infrared spectra enable prediction of lactation-stage-dependent methane emissions of dairy cattle within routine population-scale milk recording schemes. Anim. Prod. Sci. 56:258–264. https://doi.org10.1071/AN15590.

Vanlierde, A., H. Soyeurt, N. Gengler, F.G. Colinet, E. Froidmont, M. Kreuzer, F. Grandl, M. Bell, P. Lund, D.W. Olijhoek, M. Eugène, C. Martin, B. Kuhla, and F. Dehareng. 2018. Short communication: Development of an equation for estimating methane emissions of dairy cows from milk Fourier transform mid-infrared spectra by using reference data obtained exclusively from respiration chambers. J. Dairy Sci. 101:7618–7624. https://doi.org/10.3168/jds.2018-14472.

Vanlierde, 2019. Development of equations to predict methane eructed by lactating cows from milk mid-infrared spectra. Doctoral thesis. Gembloux Agro-Bio Tech, Université de Liège, Belgium.

Vanrobays, M., C. Bastin, J. Vandenplas, H. Hammami, H. Soyeurt, A. Vanlierde, and F. Dehareng. 2016. Changes throughout lactation in phenotypic and genetic correlations between methane emissions and milk fatty acid contents predicted from milk mid-infrared spectra. J. Dairy Sci. 99:7247–7260. https://doi.org/10.3168/jds.2015-10646.

Vanrobays, M.-L. 2019. Large-scale predictions of methane emission and body weight as tools contributing potentially to the genetic improvement of environmental sustainability of milk production. Doctoral thesis. Gembloux Agro-Bio Tech, Université de Liège, Belgium.

Visentin, G., A. McDermott, S. McParland, D.P. Berry, O.A. Kenny, A. Brodkorb, M.A. Fenelon, and M. De Marchi. 2015. Prediction of bovine milk technological traits from mid-infrared spectroscopy analysis in dairy cows. J. Dairy Sci. 98:6620–6629. https://doi.org/10.3168/jds.2015-9323.

Visentin, G., M. Penasa, P. Gottardo, M. Cassandro, and M. De Marchi. 2016. Predictive ability of mid-infrared spectroscopy for major mineral composition and coagulation traits of bovine milk by using the uninformative variable selection algorithm. J. Dairy Sci. 99:8137–8145. https://doi.org/10.3168/jds.2016-11053.

Visentin, G., M. Penasa, G. Niero, M. Cassandro, G. Visentin, M. Penasa, G. Niero, M. Cassandro, G. Visentin, G. Niero, M. Cassandro, and M. De Marchi. 2018. Phenotypic characterisation of major mineral composition predicted by mid-infrared spectroscopy in cow milk. Ital. J. Anim. Sci. 17:549–556. https://doi.org/10.1080/1828051X.2017.1398055.

Waaben, 2016. FTIR or NIR for dairy analysis? Accessed Feb. 3, 2020. https://www.fossanalytics.com/it-it/news-articles/videos/dairy/ftir-or-nir-for-dairy-analysis.

Wang, Q., and H. Bovenhuis. 2019. Validation strategy can result in an overoptimistic view of the ability of milk infrared spectra to predict methane emission of dairy cattle. J. Dairy Sci. 102: 6288-6295. https://doi.org/10.3168/jds.2018-15684.

Wang, Q., and H. Bovenhuis. 2020. Combined use of milk infrared spectra and genotypes can improve prediction of milk fat composition. J. Dairy Sci. 103:2514–2522. https://doi.org/10.3168/jds.2019-16784.

Woolpert, M.E., H.M. Dann, K.W. Cotanch, C. Melilli, L.E. Chase, R.J. Grant, and D.M. Barbano. 2016. Management, nutrition, and lactation performance are related to bulk tank milk de novo fatty acid concentration on northeastern US dairy farms. J. Dairy Sci. 99:8486–8497. https://doi.org/10.3168/jds.2016-10998.

Zhang, L., Y. Ma, F. Dehareng, C. Grelet, F. Colinet, N. Gengler, and H. Soyeurt. 2019. Can the calculation of a global H distance ensure the quality of international base MIR predictions? ICAR conference, Prague, Czech Republic.

Appendix

Table A-1. Information about novel traits prediction equations: unit of measurement; project that supported the development of the equation; number of observations (Nobs), mean and SD for the calibration dataset; R^2 of cross-validation (R^2cv); and standard error of cross-validation (SECV)

Trait	Unit	Project	Nobs calibration set	Mean calibration set	SD calibration set	R^2cv	SECV
Calcium (Ca)	mg/kg milk	NovaUdderHealth & OptiMIR	1094	1149.95	124.74	0.82	53.38
Potassium (K)	mg/kg milk	NovaUdderHealth & OptiMIR	1090	1526.71	130.89	0.55	88.14
Citrate	mmol/L milk	CompoMILK	498	9.054	2.189	0.89	0.720
β- hydroxybutyrate (BHB)	µmol/L milk	CompoMILK	419	2.256	0.257	0.71	0.136
Lactoferrin	mg/L milk	European Milk Recording (EMR)	924	292	244	0.72	131
Omega-3	g/dl milk	RobustMILK & OptiMIR	1779	0.026	0.0097	0.66	0.0056
Omega-6	g/dl milk	RobustMILK & OptiMIR	1784	0.1031	0.0281	0.72	0.0149