Exploiting opportunities in dairy cattle breeding using mid-infrared spectral data associated to novel traits in the Walloon Region of Belgium

N. Gengler[1], H. Soyeurt[1], C. Bastin[2], C. Bertozzi[2], F.G. Colinet[1], E. Froidmont[3], A. Gillon[2], C. Grelet[3], H. Hammami[1], J. Leblois[2], P. Mayeres[2], E. Piraux[4], R. Reis Mota[1], S. Vanderick[1], A. Vanlierde[3], M.-L. Vanrobays[1], D. Veselko[4] and F. Dehareng[3]

[1] University of Liège – Gembloux Agro-Bio Tech, 5030 Gembloux, Belgium
[2] awé groupe, 5190 Ciney, Belgium
[3] Walloon Agricultural Research Center, 5030 Gembloux, Belgium
[4] Milk Committee, 5030 Battice, Belgium

Abstract

The Walloon Region of Belgium has started very early (since 2005) to collect mid-infrared (MIR) milk spectral data produced during routine milk performance recording by Dairy Herd Improvement (DHI) organizations and to research the possibilities offered by this technology. Already in 2008, a Walloon Research and Development (R&D) consortium was created. This partnership developed a framework to collect, to store, to research and to use milk MIR spectral data first from DHI, later from milk payment. This strategy was internationalized and related innovative "open" calibration schemes were developed. New international partners can benefit from this expertise. In order to get appropriate scientific and industry impacts, several major advances were achieved, the three most important being: 1) common strategies and specifications to access milk MIR data on many spectrometers from different laboratories and countries, to store and use this data; 2) standardization of MIR data across different spectrometers overtime to generate harmonized MIR data (i.e., organized through an international network); 3) making different, often heterogeneous, reference data useful for the development of novel calibration equations. In this "open" calibration strategy, partners get access to the latest version of the prediction equations and updates when new partners join, still retaining full control and confidentiality of their data, only the calibration equation building organizations have access to all data restricted to its use for equation building. We will present opportunities and challenges for two groups of MIR based traits, fatty acids and methane emission proxy traits in dairy cattle breeding. Similarly, R&D are ongoing for the use of MIR data for many other traits as milk and milk product (i.e., cheese making) quality, animal efficiency and resilience, health and welfare traits (e.g., resistance to heat stress) to be used in future genomic evaluations. We have experienced that there are countless opportunities in MIR based breeding, only restricted by the limits in financial and human resources available in the Walloon Region.

Key words: experience, spectral data, fatty acids, methane, spectrometry

Introduction

Research started in 2005 in the Walloon Region of Belgium on the topic of using milk mid-infrared (MIR) spectral data produced during routine milk performance recording by milk recording (MR) organizations providing dairy herd improvement (DHI) services. Based on an MSc Thesis leading to a PhD (Soyeurt, 2008), since 2005, the local milk recording organization, the regional milk analysis lab, science and extension joined forces to collect MIR spectra during routine milk performance recording and to perform associated research and development (R&D). Within 3 years, collection of milk MIR spectra was organized from first very few herds to all herds in Wallonia. This paper will report experiences learned from novel traits and phenotypes collection based on MIR data, but also challenges and opportunities in its use for animal breeding.

Walloon Context

In this context, being small was an advantage with a very simple and coherent structure of already highly interconnected groups. In 2008, the "Futurospectre Consortium" was created associating:
- Science and extension: Walloon Agricultural Research Center (**CRA-W**) and University of Liège – Gembloux Agro-Bio Tech (**ULiège-GxABT**);
- MR and DHI services providing organization: Walloon Breeding Association, now awé groupe – elevéo, (hereafter called **awé**);
- Milk analysis laboratory: Milk Committee (Comité du Lait - **CdL**).

The same strategy was later extended to the collection and, anonymous, storage of MIR spectra from the analysis of milk for milk payment.

What is Infrared Spectrometry?

Infrared spectroscopy is a class of spectroscopy that exploits the infrared region of the electromagnetic spectrum. A wide range of techniques, exploiting the interaction of infrared light with matter, are potentially used. In the context of the analysis of raw milk for routine MR and milk payment, the most common techniques is absorption spectroscopy using the mid-infrared region. In contrast to other uses of spectroscopy, our context is about quantification, and therefore, the term spectrometry was coined. Absorbances of infrared light measured at different wavelengths, often called wavenumbers defined as the inverse of the wavelength, are measured in instruments called spectrometers. Before 2005, the spectrometers were used as black boxes generating standard MR traits like protein, fat, lactose, and urea contents used in DHI by combining observed absorbances towards predictions of traits of interest. (De Marchi et al., 2014; Gengler et al., 2016) using manufacturer's equations.

Standardization, Calibration and Internationalization

An essential element in the development of management and breeding tools based on MIR data is a R&D step called "calibration". Predicting novel traits from MIR spectral data needs the generation, the validation and the organization of the use of prediction or calibration equations for traits of interest. By using appropriate statistical methods based on advanced multivariate analysis (chemometrics), appropriate calibration coefficients are estimated linking reference values to spectral data for reference data points. Calibration needs the largest possible (and expected) variability in reference phenotypes covering the whole range. As an example, if values between 1 and 10 are possible, reference data from 1 to 10 are needed for calibration, potentially representing 1/10 of the data set. Similarly, MIR spectral data necessary for calibration should also cover the largest possible (and expected) variability.
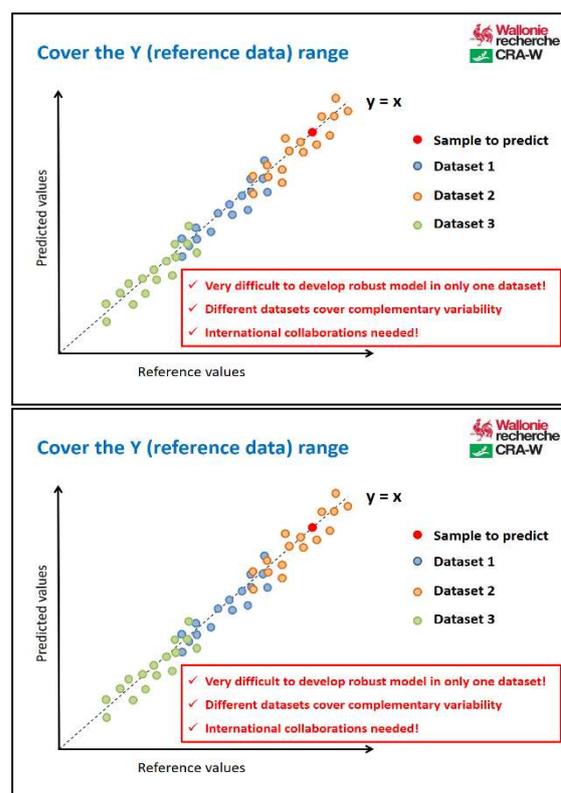
Figure 1. Illustration of the need to combine different datasets to cover enough range of phenotypic (reference data) variation.

As illustrated in Figures 1 and 2, this need for sufficient variability require the assembly of different and variable data sets.
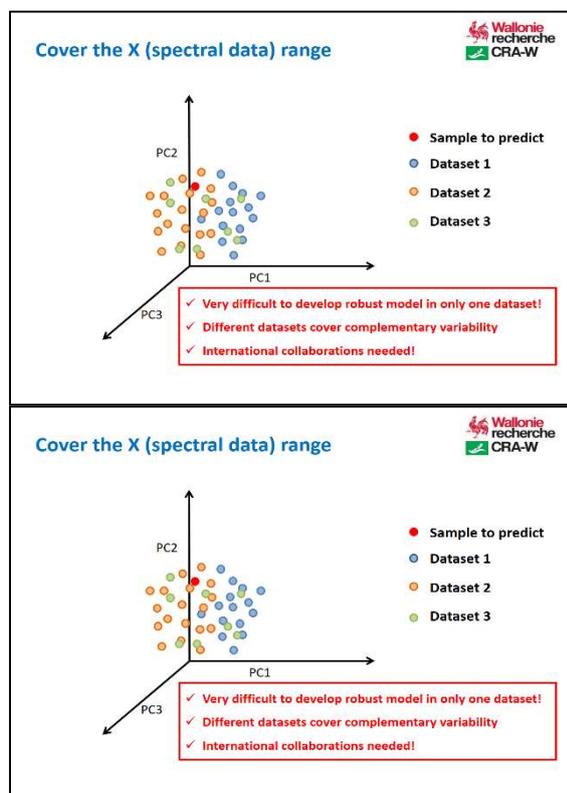
Figure 2. Illustration of the need to combine different datasets to cover enough range of MIR spectral variation, here represented by the three largest principal components (PC).

Due to the small size of Wallonia, limiting funding and animal variability, the need for international innovative calibrations appeared very early. This internationalization was also achieved through collaborations in several successful European projects such as inside the FP7 framework: RobustMilk, Greenhousemilk, and GplusE; and inside the INTERREG NWE context: OptiMIR and HappyMoo. Especially the INTERREG projects extended our philosophy of close academia-industry collaboration to a European dimension.

Even if the importance of international collaborations is obvious, these collaborations must be organized. We organized calibration as an "open" process similar to the strategy commonly used in near-infrared (e.g., NIRSC in the USA - http://www.nirsconsortium.org). Open means here that new international partners (industry and science) can join by simply adding relevant reference data to the calibration data pool. Everybody can join and get the same rights, i.e., access to the prediction equations and updates when new partners arrive or methodology has been improved. There are several other advantages:

- Calibration equation developing entities are established, who manage the partnership and organize the research load across all the scientific partners;
- All partners keep full control over their own data;
- Only calibration equation developing entities (here CRA-W and ULiège-GxABT) have access to all the data explicitly only for equation building;
- All partners are treated alike as they contribute, proportionally, the same and receive the same results.

These schemes have been very successful for novel traits such as milk fatty acids and methane emissions, but they were also developed for milk and milk product quality, health, welfare, robustness and efficiency traits.

Another issue is that the assembly of different datasets obtained across different instruments needs to be done considering spectral differences. These differences can be revealed by analyzing the same samples of these instruments. Since 2012, Grelet et al. (2015) developed a strategy based on piecewise direct standardization (**PDS**), that, as illustrated in Figure 3, stabilizes the spectral data towards reference machines, here called "master". Grelet et al. (2015) extended the PDS approach to allow standardization of MIR spectra across different brands and types of instruments. A commercial service for MR is organized by European Milk Recording (**EMR**) (http://www.milkrecording.eu), an EEIG network, and the scientific support for research projects is provided by CRA-W.

The standardization of MIR spectra has two major applications. First, as mentioned before, it is needed to assemble correctly calibration data. This allows to get prediction equations

that can be instrument, but also brand, independent. Second, for the correct application of these equations, this standardization can be considered as an alternative to routine post-prediction secondary calibration defined as the adjustment of secondary slope and intercept using linear regression (Kaylegian et al., 2006) as currently used on raw MIR fat, protein, lactose predictions. Indeed, the secondary calibration need reference samples across a large range of values and this for every considered trait. Moreover, for some traits (e.g., methane emission) it is not possible to generate such reference samples.
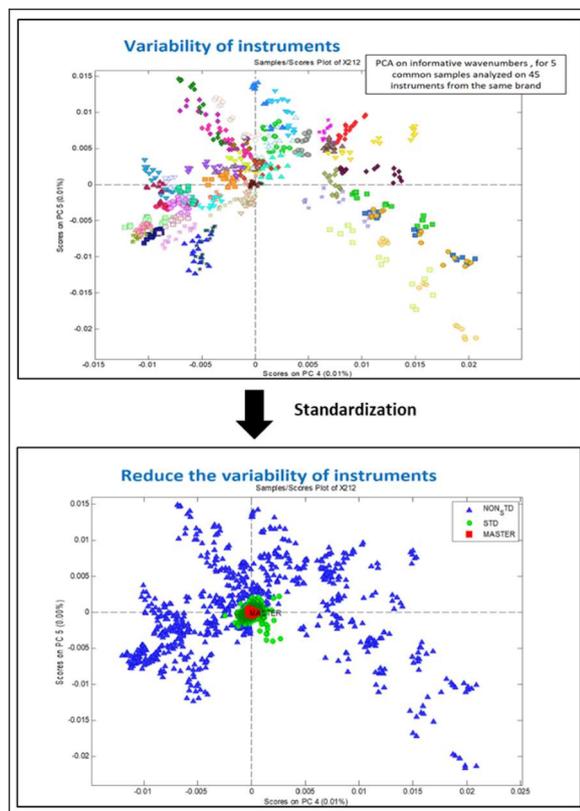
Figure 3. Illustration of the standardization process as developed by Grelet et al. (2015) and used in Grelet et al. (2017).

Opportunities (and Challenges) in Dairy Cattle Breeding

We can present two successful examples of MIR consortia leading to opportunities in dairy cattle breeding, the development of fatty acid (**FA**) equations and of MIR based methane (**CH$_4$**) proxy.

The development of the FA consortium was initiated with several projects. The study of Soyeurt et al. (2011) showed the moment when the calibration equations became stable. Several scientific studies have started towards the development of a genetic evaluation system for FA (e.g., Gengler et al., 2010; Bastin et al., 2011). However, this very early research showed also the challenges and limitation as the quantity and deepness of data, and the lack of international evaluations. Research into appropriate modeling of these traits, with adapted test-day models for massive multi-variate, multi-lactation, longitudinal data, is still far from conclusive. The research done currently in the Walloon Region for advanced genomic evaluation systems (Colinet et al., 2018) integrating external information, here external predictor traits (e.g., milk, fat%, protein%) provided by INTERBULL multiple across country evaluations (**MACE**), opens new opportunities but more research is required. Also, as in many countries, efforts are ongoing to strengthen the directly phenotyped cow contribution to the genomic reference population. Unfortunately, interest in this topic has strongly diminished as there is still a large uncertainty how to define "economics" for FA. Further progress is here pending but, outside of the human health context, milk FA are more and more recognized as excellent biomarkers for cow health and cow herd management (Gengler et al., 2016).

The development of MIR based on CH$_4$ proxies is another ongoing effort. Reference data currently used include respiration chamber and SF$_6$ methane data, but extensions to the use of Greenfeed and other novel methane sensor data are considered in new projects (e.g., Smartcow). A total of seven countries / research groups are currently involved, and this figure is expanding. Latest status of equations is reported regularly in research papers (e.g., Vanlierde et al., 2018). In general, modeling CH$_4$ for breeding purposes needs to address very many different traits recorded on different time scales, on different related animals, in different environments, potentially mixing direct measurement ("breeding goal" traits) and proxies ("index"

traits). As for FA, "economics" in a very wide sense are important and the crucial question is "why" should we select for a novel trait? An important aspect for CH_4 is its link to efficiency, a question that has not yet been definitively answered (e.g., Flay et al., 2019). Recently the concept of "Residual CH_4" (e.g., Manzanilla-Pech et al., 2016) representing emitted CH_4 ("Observed CH_4") minus the "Expected CH_4" for a given production (e.g., energy corrected milk) and maintenance (e.g., weight) level. This concept has first the advantage that it does not rely on assuming that less CH_4 is desirable, but that for a given expected level of CH_4, emitting less CH_4 is desirable. Moreover, the definition of "Residual CH_4" allows the setting up of appropriate bi-variate models were a local MIR based CH_4 proxy is considered the "Observed CH_4" and the "Expected CH_4" is based on MACE EBV for milk, fat, protein and other traits. As for FA using this setup would allow to blend this external information into the Walloon genomic evaluation system (Colinet et al., 2018). Importance of international collaboration in the context of genomic evaluations, also through the combination of cow reference populations, has to be stressed. Finally, research is required to get a clear message about genetic correlations between CH_4 and its MIR proxy, with other traits and about the "economics" for CH_4.

Conclusions

The Walloon Region of Belgium was first to get access to milk MIR data on large scale allowing research into its use. This opened very early opportunities, but we faced also challenges. The need for international collaborations was established and some initiated, but there is still room to do more. Early focus for MIR based novel phenotypes was on FA and CH_4. Retrospectively, there may be doubts these were most suitable choices because of the described issues as the complexity of the data and required associated modeling, their far from perfect availability and the lack of known genetic correlations with traits of interest and of precise economic values. Future research may still focus on CH_4 strengthening collaborations and

supporting female reference data through funding. Also, even if economic values of FA are uncertain, there are many opportunities for animal health, wellbeing and other important contributions to breeding goals linked to these traits.

Moreover, R&D allowed to achieve several major advances, the three most important being: 1) common strategies and specifications to access MIR data on many spectrometers from different labs and countries, to store and to use this data; 2) standardization of MIR data across these different spectrometers overtime to generate harmonized MIR data (i.e., organized through an international network); 3) making different, often heterogeneous, reference data useful for the development of novel calibration equations. Similarly, and based on these advances, R&D is ongoing for the use of MIR data for many other traits as quality of milk and milk product (i.e., cheese making), animal efficiency and resilience, health and welfare traits (e.g., resistance to heat stress) to be used in future genomic evaluations. These traits were researched but results were not (yet) widely publicized, illustrating the fact that there are countless opportunities in MIR based breeding.

Acknowledgements

References

Bastin, C., Gengler, N., Soyeurt, H., 2011. Phenotypic and genetic variability of production traits and milk fatty acid contents across days in milk for Walloon Holstein first-parity cows. *J. Dairy Sci. 94,* 4152–4163.https://doi.org/10.3168/jds.2010-4108

Colinet, F.G., Vandenplas, J., Vanderick, S., Hammami, H., Mota, R.R., Gillon, A., Hubin, X., Bertozzi, C., Gengler, N., 2018. Bayesian single-step genomic evaluations combining local and foreign information in Walloon Holsteins. *Animal 12,* 898–905. https://doi.org/10.1017/S1751731117002324

De Marchi, M., Toffanin, V., Cassandro, M., Penasa, M., 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

Flay, H.E., Kuhn-Sherlock, B., Macdonald, K.A., Camara, M., Lopez-Villalobos, N., Donaghy, D.J., Roche, J.R., 2019. Hot topic: Selecting cattle for low residual feed intake did not affect daily methane production but increased methane yield. *J. Dairy Sci. 102,* 2708–2713.
https://doi.org/10.3168/JDS.2018-15234

Gengler, N., Soyeurt, H., Dehareng, F., Bastin, C., Colinet, F., Hammami, H., Vanrobays, M.-L., Lainé, A., Vanderick, S., Grelet, C., Vanlierde, A., Froidmont, E., Dardenne, P., 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., Vanderick, S., Arnould, V., Soyeurt, H., 2010. Genetic evaluation for milk fat composition in the Walloon Region of Belgium. *Interbull Bull. 42,* 81–84.

Grelet, C., Fernández Pierna, J.A., Dardenne, P., Baeten, V., Dehareng, F., 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., Pierna, J.A.F., Dardenne, P., Soyeurt, H., Vanlierde, A., Colinet, F., Bastin, C., Gengler, N., Baeten, V., Dehareng, F., 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

Kaylegian, K.E., Houghton, G.E., Lynch, J.M., Fleming, J.R., Barbano, D.M., 2006. Calibration of infrared milk analyzers: Modified milk versus producer milk.
J. Dairy Sci. 89, 2817–2832. https://doi.org/10.3168/JDS.S0022-0302(06)72555-3

Manzanilla-Pech, C.I. V., De Haas, Y., Hayes, B.J., Veerkamp, R.F., Khansefid, M., Donoghue, K.A., Arthur, P.F., Pryce, J.E., 2016. Genomewide association study of methane emissions in Angus beef cattle with validation in dairy cattle. *J. Anim. Sci. 94,* 4151–4166.
https://doi.org/10.2527/jas.2016-0431

Soyeurt, H., 2008. Study of genetic variability of fatty acid profile in bovine milk and fat using mid-infrared spectrometry. PhD Thesis. Gembloux Agricultural University, Gembloux, Belgium.

Soyeurt, H., Dehareng, F., Gengler, N., McParland, S., Wall, E., Berry, D.P., Coffey, M., Dardenne, P., 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

Vanlierde, A., Soyeurt, H., Gengler, N., Colinet, F.G., Froidmont, E., Kreuzer, M., Grandl, F., Bell, M., Lund, P., Olijhoek, D.W., Eugène, M., Martin, C., Kuhla, B., Dehareng, F., 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