

Improving robustness and accuracy of predicted daily methane emissions of dairy cows using milk mid-infrared spectra

Short title: Improving milk spectra predictions of methane emissions

Amélie Vanlierde ^a, Frédéric Dehareng ^{a*}, Nicolas Gengler ^b, Eric Froidmont ^c, Sinead McParland ^d, Michael Kreuzer ^e, Matthew Bell ^{f1}, Peter Lund ^g, Cécile Martin ^h, Björn Kuhla ⁱ and Hélène Soyeurt ^b

^a Walloon Agricultural Research Centre, Valorization of Agricultural Products, 5030 Gembloux, Belgium

^b Gembloux Agro-Bio Tech - University of Liège, AGROBIOCHEM Department and Research and Teaching Centre (TERRA), 5030 Gembloux, Belgium

^c Walloon Agricultural Research Centre, Production and Sectors Department, 5030 Gembloux, Belgium

^d Teagasc, Animal and Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork, Ireland

^e ETH Zürich, Institute of Agricultural Sciences, 8092 Zürich, Switzerland

^f Agri-Food and Biosciences Institute (AFBI), Large Park, Hillsborough, BT26 6DR, United Kingdom

^g Aarhus University, Department of Animal Science, AU Foulum, 8830 Tjele, Denmark

^h Université Clermont Auvergne, INRA, VetAgro Sup, UMR Herbivores, F-63122 Saint-Genès-Champanelle, France

ⁱ Leibniz Institute for Farm Animal Biology (FBN), Institute of Nutritional Physiology, 18196 Dummerstorf, Germany

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1002/jsfa.10969](https://doi.org/10.1002/jsfa.10969)

Accepted Article

*Corresponding Author

ABSTRACT

BACKGROUND

A robust proxy for estimating methane (CH₄) emissions of individual dairy cows would be valuable especially for selective breeding. This study aimed to improve the robustness and accuracy of prediction models that estimate daily CH₄ emissions from milk Fourier-transform mid-infrared (FT-MIR) spectra by 1) increasing the reference dataset and 2) adjusting for routinely recorded phenotypic information. Prediction equations for CH₄ were developed using a combined dataset including daily CH₄ measurements (n = 1,089; g/day) collected using the SF₆ tracer technique (n = 513) and measurements using the respiration chambers (RC, n = 576). Furthermore, in addition to the milk FT-MIR spectra, the variables of milk yield (MY) on the test day, parity (P) and breed (B) of cows were included in the regression analysis as explanatory variables.

RESULTS

Models developed based on a combined RC and SF₆ dataset predicted the expected pattern in CH₄ values (in g/day) during a lactation cycle, namely an increase during the first weeks after calving followed by a gradual decrease until the end of lactation. The model including MY, P and B information provided the best prediction results (cross-validation statistics: R² = 0.68 and standard error = 57 g CH₄/day).

CONCLUSIONS

The models developed accounted for more of the observed variability in CH₄ emissions than previously developed models and thus were considered more robust. This approach is suitable for large-scale studies (*e.g.* animal genetic evaluation) where robustness is paramount for accurate predictions across a range of animal conditions.

Key Words: methane, milk, MIR spectra, dairy, phenotype, reference method

INTRODUCTION

Estimation of individual methane (**CH₄**) emissions from a large number of animals and development of strategies to reduce such emissions represent major challenges for the dairy sector. Indeed, CH₄ has a global warming potential 28 times that of carbon dioxide¹. Livestock sector represents at global level 14.5% of human-induced greenhouse gases (**GHG**) where emissions of enteric CH₄ contribute to 39%².

Measuring individual daily CH₄ emissions (g/day) using the classical reference techniques (i.e., respiration chamber, SF₆ tracer gas technique) is labor intensive and expensive. These techniques are typically applied in research scenarios and are not feasible for large-scale studies, such as for the purpose of genetic selection where thousands of data records are needed³. So, a robust proxy to estimate CH₄ would be valuable.

Previous studies highlighted the use of milk components such as fatty acids to estimate enteric CH₄ emissions^{4,5,6}, suggesting an indirect link between CH₄ emissions and milk composition. Furthermore, previous investigations^{7, 8, 9} have shown that the milk Fourier-transform mid-infrared (**FT-MIR**) spectra can be a convenient proxy for daily enteric CH₄ emissions, as the analysis of milk samples can be performed quickly and at reasonable cost. Indeed, individual milk FT-MIR spectra are already collected routinely during the milk recording process and thus present a potential solution for estimating individual CH₄ emissions during a day on a large number of cows maintained under commercial conditions^{10, 11, 12}. This prospect is particularly interesting for large-scale studies such as those needed for genetic selection^{3, 13, 14, 15} or for benchmarking farms to detect herd management practices which can be related to low CH₄ emissions. Vanlierde *et al.*⁷ demonstrated the potential of using milk FT-MIR spectra to predict CH₄ emissions from individual cows, and especially after inclusion of explanatory variables that take into account the stage of lactation for each cow. Vanlierde *et al.*⁸ developed a prediction model based on CH₄ measurements from cows

Accepted Article

obtained using the SF₆ tracer technique in Belgium and in Ireland. Vanlierde et al.⁹ also developed a prediction model based on CH₄ measurements from cows in respiration chambers (RC) during diet comparison trials in France, Northern Ireland, Switzerland, Denmark and Germany. However, when these two prediction models were applied to 558,510 milk spectra from Holstein cows collected through the Walloon milk recording system between 1 January 2016 and 31 December 2017 (after removal of spectra with a Mahalanobis distance (GH) value higher than 5 to avoid extrapolation during the prediction step), the CH₄ emission profiles during lactation were different (**Figure 1**). An increase in CH₄ emissions is expected in the first weeks post-calving until a peak in milk yield, after which a decline in emissions until the end of the lactation occurs¹⁶. The model developed from SF₆ observations does not predict the expected increase in emissions and instead predicts a relatively constant emission of CH₄ at the beginning of lactation and then a decrease after 180 DIM. In contrast, the model based on RC observations predicts the increase in emissions in early lactation, but then the predictions increase again after 220 DIM, which is unexpected. Therefore, these two models need refinement.

Improved accuracy and robustness of these previously mentioned prediction models may be achieved by adding new reference data and/or by including additional explanatory variables. Both potential improvements will be investigated in the present study. Indeed, increasing the number of reference values in the calibration dataset and potential variability among animals (i.e. breeds, diet, milk yield, etc.) is a recognized way to improve the predictive model since it is known that development of a calibration equation requires a sufficiently varied set of data¹⁷. This is why the performance of a model to estimate CH₄ emissions of dairy cows from milk FT-MIR spectra based on the combination of both available datasets (SF₆ and RC representing 1,116 records) was investigated. The incorporation of phenotypic information as explanatory variables in the prediction model may

Accepted Article

be beneficial. Shetty et al.¹⁸ included milk yield (**MY**), herd, parity (**P**), lactation stage and season as additional explanatory variables to predict CH₄ emissions from milk MIR spectra. Their CH₄ observations were obtained using the “sniffer” measurement approach, whereas the current study was based on CH₄ measurements from SF₆ and RC studies.

Within this context, this paper has two objectives: (1) to assess the effect of combining SF₆ and RC reference CH₄ data to construct a predictive model that explains more of the variability in CH₄ emissions among cows using FT-MIR spectra and (2) to assess the effect of including additional phenotypic explanatory variables for cows such as MY, P and breed (**B**) to increase the accuracy of the model.

MATERIALS AND METHODS

Existing datasets

Two reference datasets were used to carry out the present study. These datasets were described by Vanlierde et al.⁸ for the data collected using the SF₆ tracer gas technique (n = 532), and by Vanlierde et al.⁹ for data collected from RC studies (n = 584). Each dataset consists of individual daily CH₄ measurements from dairy cows with the corresponding milk FT-MIR spectra calculated as the average of the two milk spectra obtained from the morning and evening milkings weighted proportionally depending on the milk yield for each milking. This produces a daily FT-MIR spectra in a similar format to daily milk recording data and thus, ensures the practical application of the models. With the exception of spectra which predates the development of the standardization technique (N = 557), milk FT-MIR spectra were standardized according to Grelet et al.¹⁹ to correct for the instrument interference and to ensure that the milk FT-MIR spectra are comparable regardless of the spectrometer used at the date of analysis. Also, a first derivative was applied to the milk FT-MIR spectra to correct the baseline drift as recommended by Soyeurt et al.²⁰. The spectral regions studied were from wavenumbers 968 to 1,577 cm⁻¹, 1,720 to 1,809 cm⁻¹, and 2,561 to 2,966 cm⁻¹ (i.e., 289

spectral data points). As explained in detail by Vanlierde et al.⁷, constant (P0), linear (P1), and quadratic (P2) modified Legendre polynomials were computed from days in milk (**DIM**) on the day of CH₄ measurement. The resulting Legendre polynomials were applied to each wavenumber of spectra to take into account the metabolic status of each cow during lactation. Each final modified spectrum was therefore based on a total of 867 data points (i.e. 3 × 289 data points).

Cross-validation of CH₄ values obtained using existing predictive models based on SF₆ and RC observations

To highlight the interest to merge the existing SF₆ and RC reference datasets, the CH₄ prediction models by Vanlierde et al.⁸ using SF₆ measurements and by Vanlierde et al.⁹ using RC measurements were tested by cross-validation. The existing model based on SF₆ observations⁸ was used to predict CH₄ emissions from RC measurements, and the existing model based on RC observations⁹ was used to predict CH₄ emissions from SF₆ observations. To achieve this specific objective, some editing was applied on the datasets. At a spectral level, 106 observations from the RC-based dataset (18.4% of the RC observations) had a GH > 3, meaning that those spectra were too far from the values used in the calibration of the SF₆-based prediction model. Consequently, to prevent spectral extrapolation, the SF₆-based model was applied on only the remaining 478 RC observations. Within the SF₆-based dataset 27 observations (5.3% of the SF₆ observations) had a GH > 3. Thus, the RC-based prediction model was applied on the remaining 504 observations. The root mean square error of predictions (**RMSEP**) was calculated as well as the relative predictive error (**RPE**) defined as the ratio between each RMSEP and the mean of the corresponding observations.

Combined SF₆ and RC calibration dataset

A new calibration dataset was developed from merging SF₆ and RC-based datasets (n = 1,116) to capture more variability among animals with the aim to create a more robust prediction model. The GH value was calculated for each spectrum based on their coordinates

on principal components. Principal components were used since some of the spectral variables were highly correlated (11% of spectral variables had an absolute Pearson correlation higher than 95%). Spectra with a GH value greater than 3 were considered as different from the total reference dataset and removed^{21, 22}. Indeed, during the calibration process, the presence of outlying spectra in the reference dataset means we have gaps in the spectral data. This could negatively influence the quality of the prediction model. During this calibration process, 27 data values were considered outliers and discarded (2.4% of the initial dataset). This is a little higher than reported by Withfield et al.²³ who discarded 1% of the dataset based on a GH threshold > 3. The final dataset for model development comprised 1,089 CH₄ records of 413 ± 102 g/d (mean ± SD; 513 records from SF₆ and 576 records from RC studies including 299 different cows between 7 and 322 DIM) and corresponding milk FT-MIR spectra (**Table 1**). Reference data were collected in seven countries: Belgium (BE) and Ireland (IE) for SF₆ data, and United-Kingdom (UK), France (FR), Denmark (DK), Switzerland (CH) and Germany (D) for the RC data. Seven breeds were included in this dataset: Holstein (HO, 42 cows from BE, 98 from IE, 9 from CH, 5 from UK, 9 from FR, 9 from DK and 50 from D), Jersey (JER, 10 cows from DK), Brown Swiss (BS, 39 cows from CH), Red Holstein (RH, 8 cows from CH), Swedish Red Crossbred (SRX, 7 cows from UK), Norwegian Red (NR, 6 cows from IE) and Holstein Crossbred (HOX 6 cows from IE and 1 cow from CH). Within the dataset 74% of cows were HO breed. A summary for each breed within the reference dataset is provided in **Table 2**. The DIM of each cow is included in the fitted model at the spectral level through the modified Legendre Polynomials. **Table 3** gives details on the number of cow records within the reference dataset with regard to stage of lactation, the country of origin and the method used to measure enteric CH₄ emissions. Of the 1,089 records, 291, 401 and 397 were from cows in their first, second or third and greater parity, respectively. The lack of data from cows in first lactation and measured in RC (Table

1) is partly because heifers are not ideal for dietary trials as they have specific dietary needs due to growth requirements. Additionally, they are more difficult to manage in a RC.

Attention was paid to avoid increasing the error of the model that could be attributed to the merging of two datasets which used different methods to measure CH₄ emissions. This step is detailed in Appendix 1, and found that, in this case, the SF₆ and RC reference datasets can be combined without adjusting the CH₄ values for either of the measurement methods used.

Calibration process

To avoid collinearity problems in the regression due to the high correlation existing between some spectral datapoints, a modified partial least square (**PLS**) regression was used for each calibration model as implemented in the WinISI software (version 4.6; Foss, Hillerød, Denmark) including an auto-scale process. This regression is based on PLS latent variables, which are a linear combination of explanatory variables built by taking into account the variability of X (i.e., matrix of explanatory variables) and y (i.e., vector of CH₄ observations). More details about the PLS methodology can be found in Despaigne et al.²⁴. The number of latent variables was estimated using a 5-group internal cross-validation procedure. This means that the reference data set was divided randomly into five groups; five different calibration models were then developed on four groups by removing each time one of the groups. To assess the robustness of developed equations, R² of calibration (**R²c**) and of cross-validation (**R²cv**) as well as the standard errors of calibration (**SEC**) and cross-validation (**SECV**) obtained in the PLS procedure of WinISI 4 software (Foss, Hillerød, Denmark) were compared. The data were not pre-corrected for the animal effect using a random model. Indeed, including this correction in the model would involve having a reference CH₄ measurement for all cows for which we want to have a prediction. This last point would not be possible for all cows on the field for practical and financial issues.

Usefulness of including additional explanatory variables

Previous studies by our group^{7,8,9} that described the performance of equations to estimate CH₄ emissions were based only on milk FT-MIR spectra regressed by the modified Legendre polynomials to take into account the expected metabolic status of cows associated with their lactation stage. The inclusion of additional explanatory variables in the model could improve the accuracy of CH₄ prediction. Explanatory variables based on information routinely recorded on commercial farms are preferable to ensure the applicability of the models developed to commercial farms. Therefore diet information was not added as an explanatory variable as dietary traits are not routinely collected in commercial farms. In the present study, milk yield, parity and breed information were included either individually or together as explanatory variables in model development, as well as the milk FT-MIR spectra, to predict CH₄ emissions. This led to eight different calibration equations based on different combinations of explanatory variables. Abbreviations were given to these models in order to improve the clarity of the presentation of results (**Table 4**). Milk yield was considered as a quantitative variable in the model. Parity (1, 2 and > 2) was also considered as a quantitative variable since a trend exists between CH₄ emissions as the milk yield and the dry matter intake increase when the cow gets older²⁵. Different breeds of cows were categorized as dummy variables: HO, JER, BS, RH, SRX and “others” for the remaining breeds (0: absence; 1: presence).

Similarly to the methodology detailed in **Appendix 1** about the consideration of an adjustment based on the CH₄ measurement method used, an analysis of variance was performed on residuals to test the significance of effects related to DIM (indirectly included in the spectral modification), parity, breed and country. Those variables were considered separately as fixed effects in a one-way ANOVA using Proc GLM procedure in SAS software (version 9.4., SAS Inst. Inc., Cary, NC, USA) for each of the eight models developed.

Accepted Article

In addition to the cross-validation procedure previously mentioned, a cow and country dependent external validation (**CCDEV**) was conducted using R version 3.5.2 (R Core Team, 2017, Vienna, Austria) and using packages “pls” version 2.7-1²⁶ and “cvTools” version 0.3.2²⁷ to evaluate the robustness of the models as done in a previous study⁹. As there are often several measurements on the same cow in the calibration dataset, this procedure prevents data from the same cow being present both in the calibration and in the validation set. All the data from 20% of the cows per country were removed simultaneously, resulting in the removal of all data from 8 cows from BE, 22 cows from IE, 10 cows from D, 12 from CH, 4 from DK, 1 from FR and 2 from UK. The calibration model was developed on the remaining 80% of cows per country, and the model was subsequently tested on the removed data to avoid having the same animals in the calibration and in the validation set (unlike the aforementioned cross-validation procedure). The CCDEV process was repeated 500 times for each developed model to test a large number of combinations of randomly removed cows, and the variations of the R² of CCDEV and of the RMSEP of CCDEV between each calibration model were used to compare them. This allowed evaluation of the models ability to estimate CH₄ emissions for independent cows under the same conditions as cows included in the calibration data. In this study, these various conditions are dependent on the specific trials from each country.

Breed-specific models were tested to determine the accuracy of predictions within a specific breed. A PLS regression was performed as described before, with a 5-group cross validation process and a cow and breed dependent external validation (**CBDEV**). Regarding this CBDEV step, 20% of cows from one breed were removed randomly and simultaneously, with the calibration models developed on the remaining cows of this same breed, and subsequently tested using the removed data. The CBDEV process was repeated 200 times on each breed considered to test a maximum number of combinations of randomly selected cows. The R² and the RMSEP for CBDEV were determined for each calibration model.

In addition, the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the concordance correlation coefficient (CCC) were obtained to determine the optimal prediction model (i.e. lowest values for AIC and BIC and the highest value for CCC).

RESULTS AND DISCUSSION

External validation of existing predictive models based on SF₆ and RC observations

The calibration model based on SF₆ observations assessed using 478 RC observations leading to a RMSEP of 105 g of CH₄/day (RPE of 26.2%). The RMSEP for the model based on RC observation assessed using 504 SF₆ observations was 140 g of CH₄/day (RPE of 32.6%). These resulting RMSEPs values were expected due to differences between both datasets in terms of technological (reference methods used to measure CH₄), zootechnical (breeds, management, genetics, study conditions, etc.) and phenotypic (range in CH₄ values, milk yields, etc.) attributes. **Figure 2** shows predicted CH₄ emissions using either the SF₆- and RC-based models against observed CH₄ emissions for the cross-validation test. The aforementioned lower RMSEP value using the model based on SF₆ observations on the RC data can be related to the tighter distribution of the validation dataset in Figure 2 a) than in Figure 2 b). The better predictive performance of the equation based on SF₆ observations was also expected due to the greater range of reference CH₄ values included in the SF₆ reference dataset in comparison with the RC-based reference dataset.

The equation based on SF₆ observations applied to RC data, specifying the country of origin, resulted in RMSEP values of 152 g CH₄/d for UK, 79 g CH₄/d for FR, 121 g CH₄/d for DK, 65 g CH₄/d for CH and 87 g CH₄/d for D. The CH-dataset had the lowest RMSEP. It is interesting to note that if only data from two of the three experiments included in the CH-dataset were considered (111 reference data from 46 cows), the CH-statistics were significantly improved with a RMSEP of 60 g CH₄/d. Differences can be observed within and

between countries depending on the dataset in terms of the accuracy of predictions. This should be linked with the differences in data from each country regarding observed values and zootechnical attributes (Tables 1, 2 and 3).

More than the zootechnical characteristics of the animals, **Figure 3** graphically represents the milk FT-MIR spectrum as a function of the reference dataset (SF_6 vs. RC) and for the first three principal components (explaining 82% of the total spectral variability). It can be noted that the information and the variability they contain respectively are complementary. This also infers that the models were not robust enough to perform external validation tests from one country or one reference method to another in the current state. Therefore, there is a real interest to combine these CH_4 reference data to develop a new model.

This section highlights the importance of ensuring that the characteristics of animals or their milk FT-MIR spectra for which predictions are required are represented in the calibration dataset used to develop the model.

Calibration equations combining RC and SF_6 reference datasets and inclusion of phenotypic information as explanatory variables

Calibration and cross-validation statistical results for the eight developed calibration equations built from raw CH_4 data and including different combinations of milk yield, parity and breed in addition to milk spectra as explanatory variables are detailed in **Table 5**. With the addition of zootechnical information as explanatory variables, R^2c and R^2cv slightly increased compared to the MS model. Consequently, a decrease in SEC and SECV was expected with the inclusion of additional phenotypic information (i.e. milk yield, parity and breed). The highest R^2 values and minimum errors were observed for the MSMYPB model with an R^2c of 0.73, a R^2cv of 0.68, a SEC of 53 g CH_4/d and a SECV of 57 g CH_4/d . Regarding the CCDEV procedure, MSMYPB presented the highest R^2 (0.60) and lowest

RMSEP (65 g CH₄/d). Moreover, the MSMYPB model presented the highest CCC value (0.81 meaning accurate predictive ability²⁸) and the lowest AIC and BIC values (-186 and -181 respectively from the AIC and BIC of the MS model).

In comparison with the statistics obtained with the models based on one kind of measurement method, R²c and R²cv shown in Table 5 (between 0.68 and 0.73, and between 0.64 and 0.68 respectively) are lower than for the equation based only on SF₆ observations⁸ (0.74 and 0.70, respectively) and higher than for the equation based only on RC observations⁹ (0.65 and 0.57, respectively). Regarding the errors, the SEC and SECV obtained with the models developed here and shown in Table 5 (between 58 and 53, and between 61 and 57 CH₄/d, respectively) are all lower than the ones obtained for the existing model based on SF₆ observations⁸ (66 and 70 g CH₄/d, respectively) and greater than the ones obtained for the existing model based on RC observations⁹ (43 and 47 g CH₄/d, respectively). The statistical results for the MSMYPB version were similar to the R²c and R²cv for the model based on SF₆ observations⁸ with lower errors, which suggests an improvement of accuracy. Moreover, by considering a higher number of observations the MSMYPB model described more of the variability in CH₄ emissions than previous models^{8,9} based on SF₆ and RC observations, leading to an improvement of the robustness of the model. **Figure 4** illustrates the CH₄ observations compared to the predictions obtained with the MSMYPB model.

Based on the reference datasets available for this study, RPEcv and CCC parameters reach respectively 14.1% and 0.81 for the MSMYPB model (Table 5). These statistics are slightly better than the best models based on ingestion, dietary components and/or animal attributes (but no milk MIR spectra) summarized in Niu et al.²⁹ (respectively 14.7% and 0.76 in the best conditions). The advantage of predictive models based on milk MIR spectra is that spectra are routinely collected and permit to have easily a value at individual animal level at large scale. MSMYPB model developed in this study also has better statistics than Shetty et

al.¹⁸ who observed a R^2 of prediction and a RMSEP of 0.13 and 111 L of CH_4/d respectively for the model based only on the milk FT-MIR spectra while these parameters were about 0.39 and 94 L/d respectively when they also considered MY, herd, P, DIM and season as variables in the model.

Figure 5 details the trends of the predictions obtained when these new models are applied on the same spectral dataset described in the introduction ($n = 558,510$ milk FT-MIR spectra from Holstein cows collected through the Walloon milk recording system) as a function of the DIM of the animal. The CH_4 curves of the eight models are very similar and are now much more in accordance with the literature¹⁶, with an increase during the first weeks of lactation followed by a moderate decrease until the end of lactation.

The effect on the residuals of adding DIM, parity, country and breed into each of the eight developed models are specified in **Table 6**. It can be noted that DIM had no significant effect on the residuals. Since the lactation stage is accounted for by using the modified Legendre polynomials to regress spectral data, this non-significant influence of DIM on the residuals was expected. The daily milk yield has a significant impact on the residuals for the models where milk yield is not an explanatory value of the regression. So, considering milk yield in the model is relevant. The inclusion of parity in the regression model has a low influence compared to other variables (Table 5). When parity is not included in the model, the resulting residuals are significantly dependent on the parity of the cow (Table 6). Adding parity information in the model seems therefore to be appropriate. The country factor has a highly significant effect on residuals in models where breed was excluded. This is because some breeds are represented in only one country (e.g. JER, BS, NR). Finally, breed had a significant effect on the residuals in all models including those with breed as a factor. Some breeds had a very low representation in the calibration set and were also linked with particular experimental protocols and diet, which can explain this effect. However, the breed effect was

less significant when breed was included in the model as a parameter in combination with parity or/and milk yield (Table 6). Accordingly, these phenotypic combinations permitted a better description of the animal and its performance than breed alone. Models with MSPB and MSMYPB were found to best describe variability in CH₄ among this dataset.

Even for MSPB and MSMYPB models, the breed effect was not totally accounted for by the model as the effect on the residual was still significant. Therefore, the prediction accuracy of cross-validation can be impacted as the CCDEV procedure was conducted randomly regarding cows selected for calibration or validation datasets.

The results obtained during the CBDEV procedure are given in **Table 7**. Depending on the version of the equation, the average R² values ranged from 0.57 to 0.60 for the HO breed. At the statistical level, the MSMYP model presents the most interesting combination of highest coefficients of determination and lowest errors for both calibration and validations steps.

The statistical performance of the CBDEV models were not greater than the predictive models based on the total dataset with all breeds included and were similar to the results obtained during the CCDEV procedure. Data from other breeds may bring added value to the models. This would allow predictions from milk spectra collected on additional breeds than HO as long as there is not sufficient data to consider the development of breed-specific models.

The combined information given by including milk yield, parity and breed improved the prediction of daily CH₄ emissions of dairy cows by milk FT-MIR spectra. As the phenotypic information included is routinely available, the MSMYPB model can be applied to commercial farms. The prediction model can be chosen depending on available phenotypic information. The more explanatory variables included that describe the animal, the better the

accuracy of the prediction model. Therefore, inclusion of other routinely available traits presents the possibility of further improvement. There is still potential for this model to be improved and to evolve by the inclusion of new explanatory variables and as new data becomes available.

A next step will be to determine the performance of the models developed in the present study under practical conditions by measuring the methane emissions of animals estimated to differ in this trait. Once the equations prove to be robust in this respect, the use of milk FT-MIR spectra could be introduced as a tool for estimating CH₄ emissions for a large number of dairy cows. In addition, the present observations confirm the necessity to collaborate between research teams from different countries to test the models in specific local conditions and to include local variability (at spectral and phenotypical levels) in the reference dataset if it is not yet present. This is why a further improvement of the robustness of the models could occur in the future with the availability of such new relevant reference databases.

CONCLUSIONS

The models developed in the current study combining CH₄ observations obtained with SF₆-tracer and RC measurement methods accounted for more variability in spectral and zootechnical data. They explained more variation in CH₄ emissions among cows and consequently are more robust than previous versions based on those datasets considered independently. Additionally, the inclusion of milk yield, parity and breed as explanatory variables in the prediction model improved the description of variability. Moreover, the trends of the predictions obtained by these new models applied on a large spectral database adjusted for the lactation stage of animals are now more in accordance with the literature. By taking

care of removing outlying spectral data and cow phenotypic information not included in the model, an improvement in FT-MIR derived CH₄ predictions under practical conditions is expected.

ACKNOWLEDGEMENTS

The authors are grateful to their fellow researchers with regard to collecting methane and milk samples. Part of the work and facilitation of data exchanges have been feasible through projects funded by European Union's Seventh Framework Programme: GreenHouseMilk (<https://cordis.europa.eu/project/rcn/92792/>) which particularly permitted exchange of data with Teagasc Moorepark in Ireland (grant agreement n° 238562); and GplusE (<https://cordis.europa.eu/project/rcn/111473/>), especially for research conducted at the Agri-Food and Biosciences Institute (AFBI) in Northern Ireland (grant agreement n° 613689). The views expressed in this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission. The OptiMIR (INTERREG IVB North-West Europe funded; <https://www.keep.eu/project/7026/>) and the COST Methagene (COST-Horizon 2020 funded; <https://www.cost.eu/actions/FA1302/>) European projects are gratefully acknowledged for facilitating networking between research teams and data exchange. The authors thank the Walloon Breeding Association (Ciney, Belgium) and Comité du lait (Battice, Belgium) for providing access to data related to the Walloon milk recording. Qualitas AG is sincerely thanked for initiating exchanges with Switzerland on this topic. The Association of Swiss Cattle Breeders (Zollikofen, Switzerland) is acknowledged for providing the milk Fourier transform mid-infrared spectral data of the Swiss samples. Research conducted at Leibniz Institute for Farm Animal Biology (FBN) in Germany was partially funded by the optiKuh project supported by the German Federal Ministry of Food and Agriculture (BMBL) through the Federal Office for Agriculture and Food (BLE). Research conducted in France at the French National Institute for Agricultural Research (INRA) was

funded by the French National Research Agency (ANR) through the Joint Programming Initiative on Agriculture, Food Security and Climate Change (FACCE-JPI) Global Network program (ANR-13-JFAC-0003-01). Research conducted in Denmark at Aarhus University was funded by the Danish Milk Levy Fund and Aarhus University.

REFERENCES

1. Core Writing Team, Pachauri RK, Meyer LA (eds). 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. Geneva: IPCC; 2014.
2. Gerber PJ, Steinfeld H, Henderson B, Mottet A, Opio C, Dijkman J, Falcucci A, Tempio G. (eds). Tackling Climate Change through Livestock: A Global Assessment of Emissions and Mitigation Opportunities. Rome: Food and Agriculture Organization of the United Nations (FAO); 2013.
3. Negussie E, de Haas Y, Dehareng F, Dewhurst RJ, Dijkstra J, Gengler N, Morgavi DP, Soyeurt H, van Gastelen S, Yan T and Biscarini F, Invited review: Large-scale indirect measurements for enteric methane emissions in dairy cattle: A review of proxies and their potential for use in management and breeding decisions. *J Dairy Sci* **100**(4):2433-2453 (2017). <https://doi.org/10.3168/jds.2016-12030>
4. Chilliard Y, Martin C, Rouel J and Doreau M, 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 (2009). <https://doi.org/10.3168/jds.2009-2375>
5. van Gastelen S and Dijkstra J, Prediction of methane emission from lactating dairy cows using milk fatty acids and mid-infrared spectroscopy. *J Sci Food Agric* **96**(12):3963-3968 (2016). <https://doi.org/10.1002/jsfa.7718>
6. Van Lingen HJ, Crompton LA, Hendriks WH, Reynolds CK and Dijkstra J, Meta-analysis of relationships between enteric methane yield and milk fatty acid profile in dairy cattle. *J Dairy Sci* **97**(11):7115-7132 (2014). <https://doi.org/10.3168/jds.2014-8268>
7. Vanlierde A, Vanrobays M-L, Dehareng F, Froidmont E, Soyeurt H, McParland S, Lewis E, Deighton MH, Grandl F, Kreuzer M, Gredler B, Dardenne P, Gengler N, Hot Topic: Innovative lactation stage dependent prediction of methane emissions from milk mid-infrared spectra. *J Dairy Sci* **98**(8):5740-5747 (2015). <https://doi.org/10.3168/jds.2014-8436>
8. Vanlierde A, Vanrobays M-L, Gengler N, Dardenne P, Froidmont E, Soyeurt H, McParland S, Lewis E, Deighton MH, Mathot M and Dehareng F, 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 (2016). <https://doi.org/10.1071/AN15590>

9. Vanlierde A, Soyeurt H, Gengler N, Colinet FG, Froidmont E, Kreuzer M, Grandl F, Bell M, Lund P, Olijhoek DW, Eugène M, Martin C, Kuhla B and Dehareng F, 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**(8):7618-7624 (2018). <https://doi.org/10.3168/jds.2018-14472>
10. Dehareng F, Delfosse C, Froidmont E, Soyeurt H, Martin C, Gengler N, Vanlierde A and Dardenne P, Potential use of milk mid-infrared spectra to predict individual methane emission of dairy cows. *Animal* **6**(10):1694–1701 (2012). <https://doi.org/10.1017/S1751731112000456>
11. van Gastelen S, Mollenhorst H, Antunes-Fernandes EC, Hettinga KA, van Burgsteden GG, Dijkstra J and Rademaker JLW, Predicting enteric methane emission of dairy cows with milk Fourier-transform infrared spectra and gas chromatography-based milk fatty acid profiles. *J Dairy Sci* **101**(6):5582-5598 (2018). <https://doi.org/10.3168/jds.2017-13052>
12. Bittante G and Cipolat-Gotet C, Direct and indirect predictions of enteric methane daily production, yield, and intensity per unit of milk and cheese, from fatty acids and milk Fourier-transform infrared spectra. *J Dairy Sci* **101**(8):7219-7235 (2018). <https://doi.org/10.3168/jds.2017-14289>
13. de Haas Y, Pszczola M, Soyeurt H, Wall E, and Lassen J, Invited review: Phenotypes to genetically reduce greenhouse gas emissions in dairying. *J Dairy Sci* **100**(2):855-870 (2017). <https://doi.org/10.3168/jds.2016-11246>
14. Kandel PB, Vanrobays ML, Vanlierde A, Dehareng F, Froidmont E, Gengler N and Soyeurt H, Genetic parameters of mid-infrared methane predictions and their relationships with milk production traits in Holstein cattle. *J Dairy Sci* **100**(7):5578-5591 (2017). <https://doi.org/10.3168/jds.2016-11954>
15. Vanrobays M-L, Bastin C, Vandenplas J, Hammami H, Soyeurt H, Vanlierde A, Dehareng F, Froidmont E and Gengler N, 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**(9):7247-7260 (2016). <https://doi.org/10.3168/jds.2015-10646>
16. Garnsworthy PC, Craigon J, Hernandez-Medrano JH, Saunders N, Variation among individual dairy cows in methane measurements made on farm during milking. *J Dairy Sci* **95**:3181-3189 (2012). <http://dx.doi.org/10.3168/jds.2011-4606>

17. Davies AMC and Fearn T. Back to basics: calibration statistics. *Spectroscopy Europe* **18**:31-32 (2006).
18. Shetty N, Difford G, Lassen J, Løvendahl P and Buitenhuis AJ, Predicting methane emissions of lactating Danish Holstein cows using Fourier transform mid-infrared spectroscopy of milk. *J Dairy Sci* **100**: 9052-9060 (2017).
<http://dx.doi.org/10.3168/jds.2017-13014>
19. Grelet C, Fernandez Pierna JA, Dardenne P, Soyeurt H, Vanlierde A, Colinet F, Bastin C, Gengler N, Baeten V and Dehareng F, Standardization of milk mid-infrared spectrometers for the transfer and use of multiple models. *J Dairy Sci* **100**:7910–7921 (2017). <https://doi.org/10.3168/jds.2017-12720>
20. Soyeurt H, Dehareng F, Gengler N, McParland S, Wall EPBD, Berry DP, Coffey M and Dardenne P, Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. *J Dairy Sci* **94**(4):1657-1667 (2011).
<https://doi.org/10.3168/jds.2010-3408>
21. Shenk JS and Westerhaus MO, Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. *Crop Sci* **31**(2):469–474 (1990).
doi:10.2135/cropsci1991.0011183X003100020049x
22. Soyeurt H, Bastin C, Colinet FG, Arnould VM-R, Berry DP, Wall E, Dehareng F, Nguyen HN, Dardenne P, Schefers J, Vandenplas J, Weigel K, Coffey M, Théron L, Dettelleux J, Reding E, Gengler N, McParland S, Mid infrared prediction of lactoferrin content in bovine milk: potential indicator of mastitis. *Animal* **6**:1830–1838 (2012).
<https://doi.org/10.1017/S1751731112000791>
23. Whitfield RG, Gerger ME and Sharp RL, Near-infrared spectrum qualification via mahalanobis distance determination. *Appl Spectrosc* **41**(7):1204-1213 (1987).
24. Despagne F, Massart DL and Chabot P, Development of a robust calibration model for nonlinear in-line process data. *Analytical chemistry* **72**(7):1657-1665 (2000).
25. Bittante G, Cecchinato A and Schiavon S, Dairy system, parity, and lactation stage affect enteric methane production, yield, and intensity per kilogram of milk and cheese predicted from gas chromatography fatty acids. *Journal of dairy science*, **101**(2):1752-1766 (2018).
<https://doi.org/10.3168/jds.2017-13472>
26. Mevik BH, Wehrens R and Hovde Liland K, (2019). pls: Partial Least Squares and Principal Component Regression. R package version 2.7-1. <https://CRAN.R-project.org/package=pls>
27. Alfons A, (2012). cvTools: Cross-validation tools for regression models. R package version 0.3.2. <https://CRAN.R-project.org/package=cvTools>

28. Altman DG, (1997). Practical statistics for medical research. Chapman Hall, London, UK.
29. Niu M, Kebreab E, Hristov AN, Oh J, Arndt C, Bannink A, Bayat AR, Brito AF, Boland T, Casper D, Crompton LA, Dijkstra J, Eugène MA, Garnsworthy PC, Haque MN, Hellwing ALF, Huhtanen P, Kreuzer M, Kuhla B, Lund P, Madsen J, Martin C, McClelland SC, McGee M, Moate PJ, Muetzel S, Muñoz C, O'Kiely P, Peiren N, Reynolds CK, Schwarm A, Shingfield KJ, Storlien TM, Weisbjerg MR, Yáñez-Ruiz DR and Yu Z, Prediction of enteric methane production, yield and intensity in dairy cattle using an intercontinental database. *Global Change Biology* **24**(8), pp. 3369-3389 (2018). <https://doi.org/10.1111/gcb.14094>

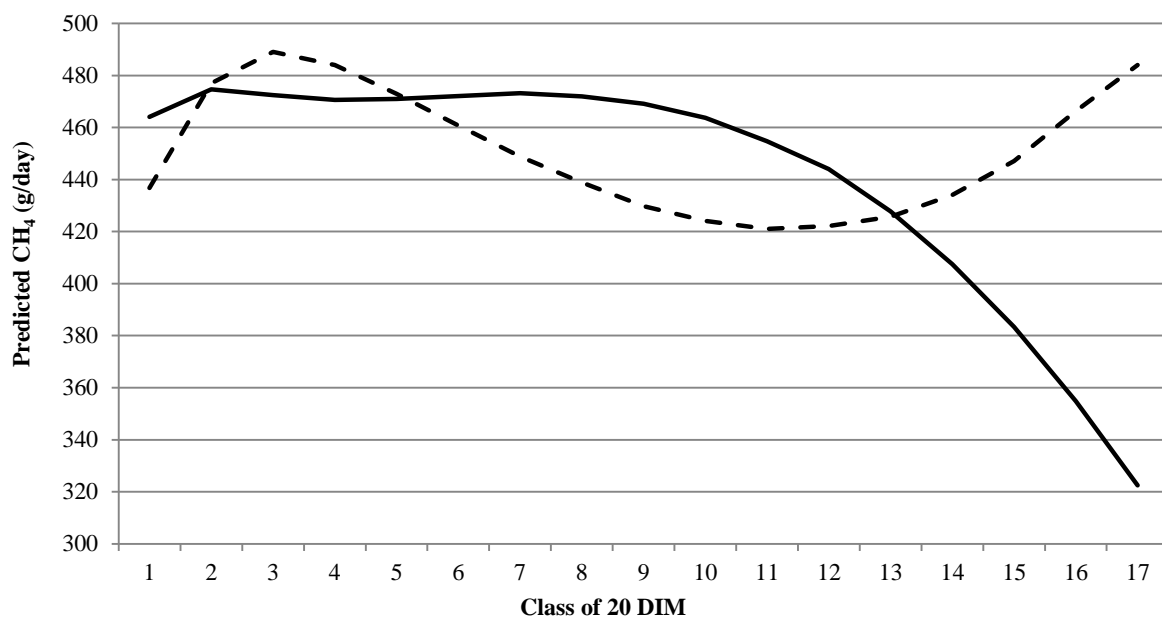
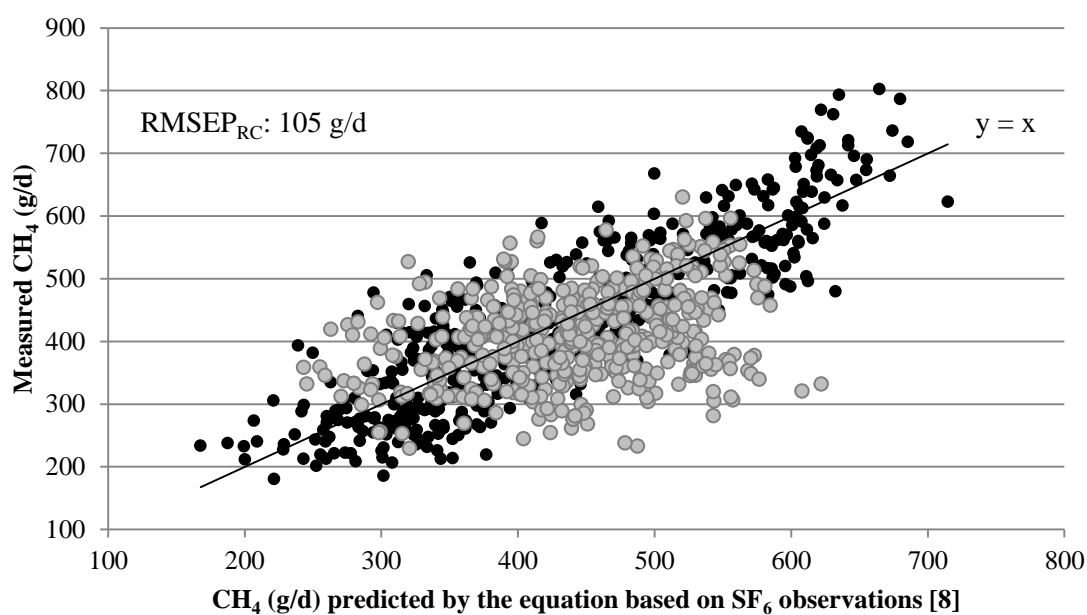


Figure 1: Average predicted CH₄ (g/day) for each class of 20 days in milk (DIM) during lactation using equation of Vanlierde et al.⁸ (— SF₆) or equation of Vanlierde et al.⁹ (--- Respiration chamber) based on spectral data collected from Holstein cows through the Walloon milk recording system between 1 January 2016 and 31 December 2017 (n = 538,510).

a)



b)

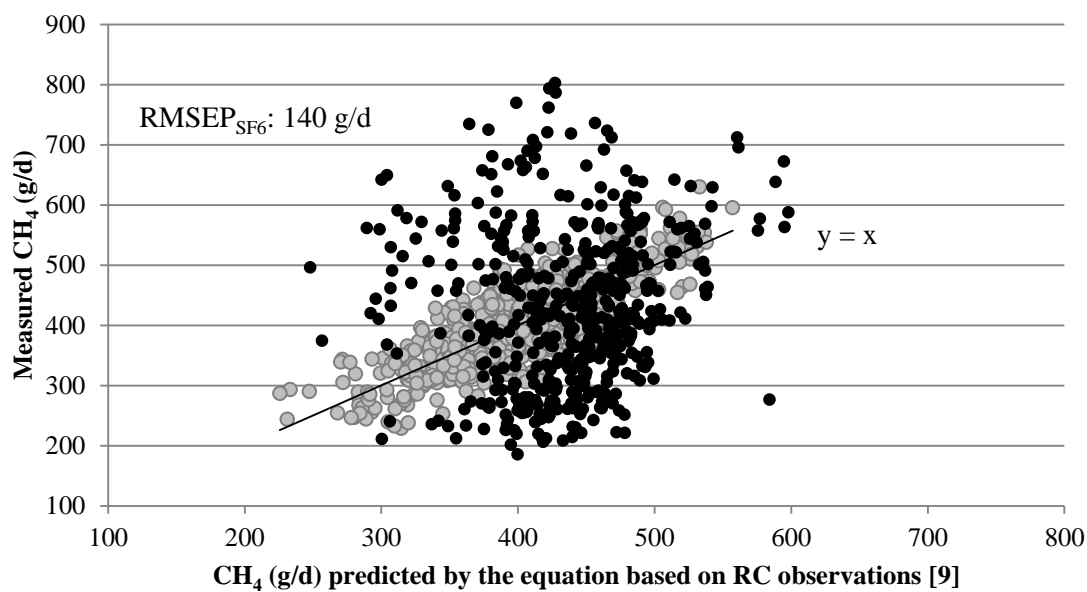


Figure 2: Measured methane emissions (g CH₄ per day) using SF₆ tracer technique (●) or respiration chamber (●) methods against predicted methane (a) using equation calibrated only on data collected with SF₆ tracer technique⁸ (validation set is the RC dataset) or (b) using equation calibrated only on data collected with respiration chamber technique⁹ (validation set is the SF₆ dataset).

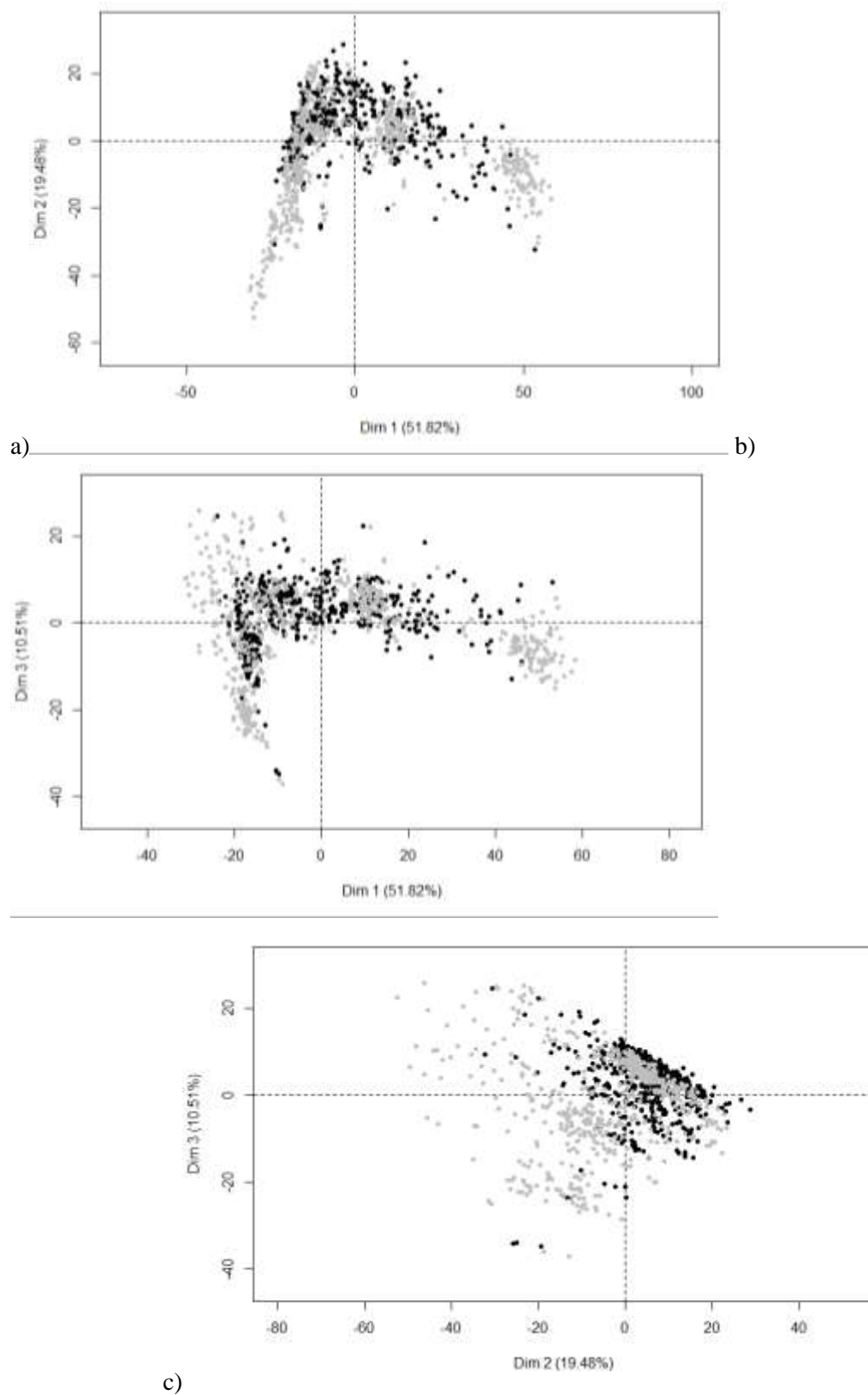


Figure 3: Graphical representation of reference milk FT-MIR spectra as a function of the first three principal components (PC) and the original dataset (● = SF₆ dataset; ● = RC dataset); a) PC 2 vs. PC 1, b) PC 3 vs. PC 1, c) PC 3 vs. PC 2

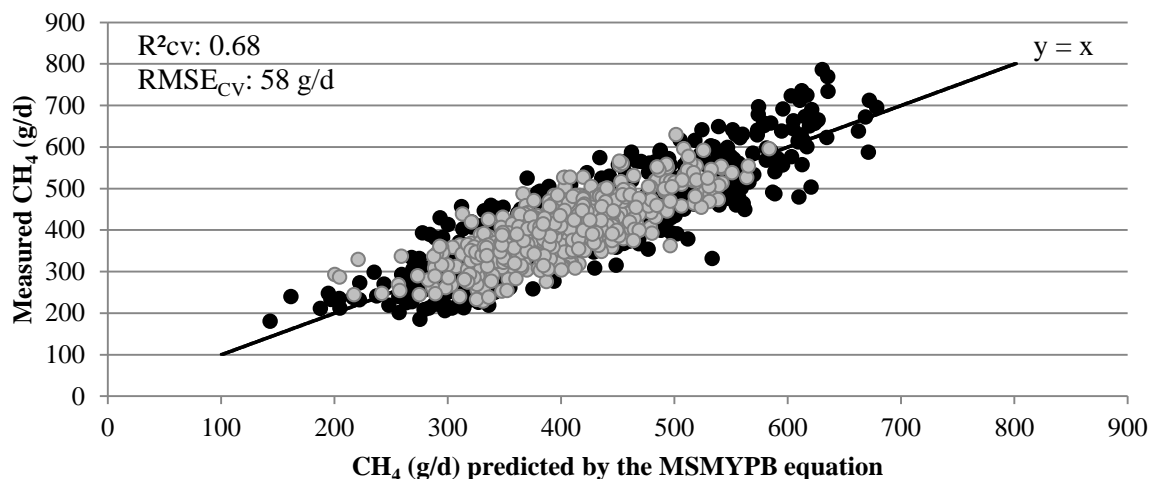


Figure 4: Measured methane (g CH₄ per day) from SF₆ and respiration chamber datasets (●: data collected with the SF₆ tracer technique⁸, n = 531; ●: data collected with the respiration chamber method⁹, n = 576) compared to predicted values using the MSMYPB model.

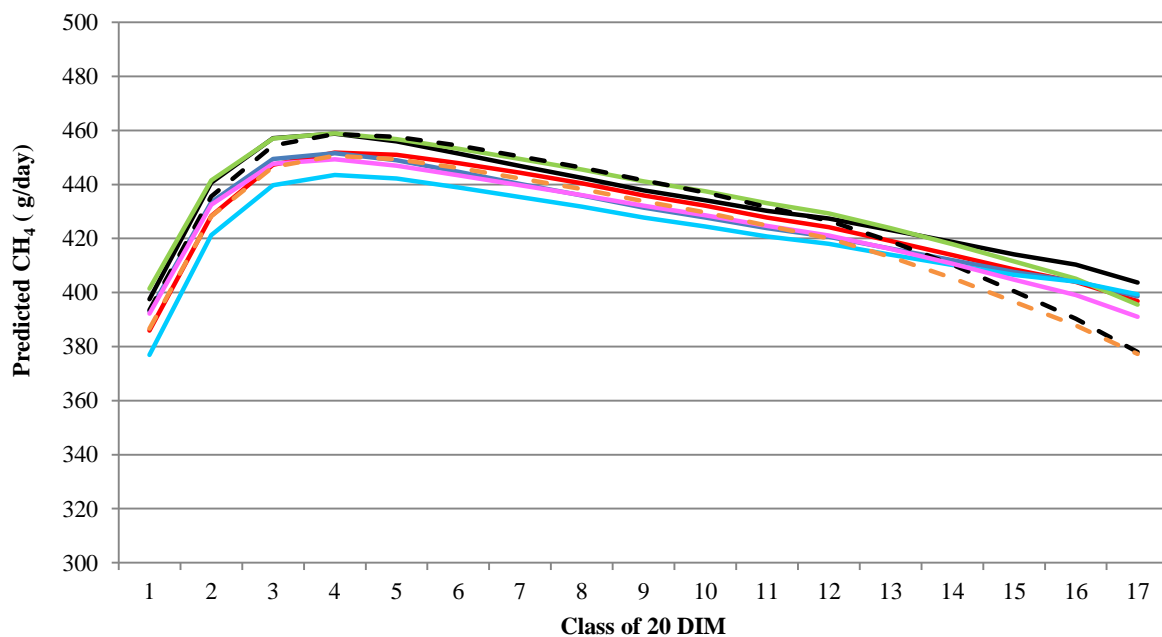


Figure 5: Average CH₄ predictions (g/day) from spectral data collected from Holstein cows through the Walloon milk recording system between 1 January 2016 and 31 December 2017 (n = 538,510) adjusted for days in milk (DIM) and explanatory variables included in the prediction model (— MS; — MSMY, — MSP, — MSB, — MSMYP, - - - MSMYB, — MSPB, - - - MSMYPB).

Table 1: The number and the range of methane (CH₄) records, number of cows and in each parity for data obtained from SF₆ and respiration chamber (RC) studies across countries included in the present study.

Reference method	Country	n data	n cows	Parity (n cows)			CH ₄ (g/d) mean ± SD	
				1	2	> 2		
SF ₆	BE	252	42	16	11	15	510 ± 105	427 ± 127
	IE	261	110	45	29	36	347 ± 89	
RC	UK	24	12	/	3	9	365 ± 44	400 ± 72
	FR	81	9	/	7	2	366 ± 61	
	DK	129	19	8	4	7	367 ± 64	
	CH	135	57	8	16	33	451 ± 75	
	D	207	50	/	34	16	405 ± 60	
TOTAL		1,089	299	77	104	118	413 ± 102	

SD: standard deviation; Belgium (BE), Ireland (IE), United-Kingdom (UK), France (FR), Denmark (DK), Switzerland (CH) and Germany (D)

Table 2: Number of methane (CH₄) records, number of cows and their parity, and mean methane emissions for each breed included in the present study.

Breed	n data	% of data	n cows	% of cows	Parity (n cows)			CH ₄ (g/d) mean ± SD
					1	2	> 2	
HO	891	82	222	74	63	84	75	415 ± 107
JER	67	6	10	3	4	2	4	342 ± 42
BS	78	7	39	13	7	7	25	458 ± 69
RH	21	2	8	3	/	5	3	427 ± 74
SRX	14	1	7	2.5	/	2	5	348 ± 37
NR	9	1	6	2	/	2	4	423 ± 79
HOX	9	1	7	2.5	3	2	2	427 ± 54

HO: Holstein; JER: Jersey; BS: Brown Swiss; RH: Red Holstein; SRX: Swedish Red crossed; NR: Norwegian red; HOX: Holstein crossed.

Table 3: Number of methane (CH₄) records included in the reference dataset with regard to stage of lactation (DIM), country of origin and the method used to measure enteric CH₄ emissions

Class of DIM	BE	IE	UK	FR	DK	CH	D	Total SF ₆	Total RC	TOTAL
0-20	-	-	-	-	-	6	-	-	6	6
21-40	4	-	-	-	-	24	63	4	87	91
41-60	8	6	-	-	-	4	7	14	11	25
61-80	15	17	-	-	-	4	-	32	4	36
81-100	15	30	-	-	-	31	9	45	40	85
101-120	20	29	-	-	-	30	56	49	86	135
121-140	33	32	-	3	-	10	-	65	13	78
141-160	12	33	-	11	3	4	2	45	20	65
161-180	25	40	6	15	12	2	8	65	43	108
181-200	14	25	16	13	16	4	2	39	51	90
201-220	12	26	-	18	12	-	2	38	32	70
221-240	38	18	2	16	17	2	2	56	39	95
241-260	26	5	-	5	32	6	2	31	45	76

261-280	19	-	-	-	21	2	2	19	25	44
281-300	5	-	-	-	14	4	31	5	49	54
301-320	3	-	-	-	2	-	21	3	23	26
321-340	3	-	-	-	-	2	-	3	2	5
TOTAL	252	261	24	81	129	135	207	513	576	1,089

DIM: Days In Milk, RC: Respiration Chamber, Belgium (BE), Ireland (IE), United-Kingdom (UK), France (FR), Denmark (DK), Switzerland (CH) and Germany (D).

Table 4: Versions of calibration equations developed based on milk FT-MIR spectra modified with Legendre Polynomials, with or without added phenotypic factors as explanatory variables.

Model	Number of explanatory variables	Name
Modified spectra	867	MS
Modified spectra + milk yield	868	MSMY
Modified spectra + parity	868	MSP
Modified spectra + breed	873	MSB
Modified spectra + milk yield + parity	869	MSMYP
Modified spectra + milk yield + breed	874	MSMYB
Modified spectra + parity + breed	874	MSPB
Modified spectra + milk yield + parity + breed	875	MSMYPB

Modified spectra: milk FT-MIR spectra regressed by the modified Legendre Polynomials to take into account the lactation stage of the cow⁷.

Table 5: Calibration statistics for each developed calibration equation (n = 1,089).

Model	n factors	R ² _c	R ² _{cv}	SEC (g/day)	SEC _v (g/day)	RP _{Ec} (%)	RPE _{cv} (%)	R ² for CCD EV Mean ± SD	RMS EP for CCD EV Mean (g/day) ± SD	AIC	BIC	CC _C
MS	14	0.6	0.6	58	61	13.	14.9	0.55	70 ±	12,0	12,1	0.7

		8	4			9		\pm 0.07	4.5	55	35	9
MSMY	14	0.6 9	0.6 5	57	60	13. 6	14.7	0.57 \pm 0.07	70 \pm 4.5	11,9 97	12,0 77	0.7 9
MSP	14	0.7 0	0.6 5	56	60	13. 6	14.8	0.57 \pm 0.07	68 \pm 4.2	11,9 90	12,0 70	0.7 9
MSB	14	0.7 0	0.6 6	56	60	13. 5	14.5	0.57 \pm 0.06	68 \pm 4.4	11,9 89	12,0 69	0.8 0
MSMY P	14	0.7 0	0.6 6	56	60	13. 4	14.5	0.58 \pm 0.07	67 \pm 4.1	11,9 62	12,0 42	0.8 0
MSMY B	15	0.7 2	0.6 8	54	58	13. 0	14.3	0.60 \pm 0.06	66 \pm 4.2	11,8 97	11,9 82	0.8 1
MSPB	14	0.7 1	0.6 7	55	59	13. 2	14.5	0.59 \pm 0.06	66 \pm 4.1	11,9 27	12,0 07	0.8 1
MSMY PB	15	0.7 3	0.6 8	53	57	12. 8	14.1	0.60 \pm 0.06	65 \pm 4.1	11,8 69	11,9 54	0.8 1

MS: Modified spectra, MSMY: Modified spectra + milk yield, MSP: Modified spectra + parity, MSB: Modified spectra + breed, MSMYB: Modified spectra + milk yield + breed, MSPB: Modified spectra + parity + breed, MSMYPB: Modified spectra + milk yield + parity + breed, SEC: Standard Error of Calibration, SECV: Standard Error of Cross Validation, RPEc: relative predictive error of calibration; RPEcv: relative predictive error of cross validation; R² for CCDEV: cow and country dependent external validation coefficient of determination, RMSEP for CCDEV: Root Mean Square Error of prediction for cow and country dependent external validation, SD: standard deviation, AIC: Akaike information criterion, BIC: Bayesian information criterion; CCC: concordance correlation coefficient.

Table 6: Effect of days in milk (DIM), daily milk yield, parity, country and breed on the residuals of each calibration model developed

Model	DIM	Milk Yield	Parity	Country	Breed
MS	NS	***	***	***	***
MSMY	NS	*	***	***	***
MSP	NS	***	NS	**	***
MSB	NS	***	***	NS	***
MSMYP	NS	NS	NS	***	***
MSMYB	NS	NS	***	NS	**
MSPB	NS	***	NS	NS	*
MSMYPB	NS	NS	NS	NS	*

DIM: Days in milk, MS: Modified spectra, MSMY: Modified spectra + milk yield, MSP: Modified spectra + parity, MSB: Modified spectra + breed, MSMYB: Modified spectra + milk yield + breed, MSPB: Modified spectra + parity + breed, MSMYPB: Modified spectra + milk yield + parity + breed, DIM: Days In Milk, NS: Non Significant, * P<0.05, ** P<0.01, *** P<0.001

Table 7: Cow and breed dependent external validation statistics for developed calibration models for Holstein cows

Model	R ² for CBDEV		RMSEP for CBDEV	
	Mean \pm SD	Min - Max	Mean \pm SD	Min - Max
MS	0.57 \pm 0.08	0.38 – 0.77	71 \pm 5.0	57 – 87
MSMY	0.59 \pm 0.08	0.36 – 0.77	71 \pm 5.0	57 – 87
MSP	0.59 \pm 0.08	0.36 – 0.79	69 \pm 4.9	56 – 82
MSMYP	0.60 \pm 0.08	0.36 – 0.78	68 \pm 4.7	54 – 82

MS: Modified spectra, MSMY: Modified spectra + milk yield, MSP: Modified spectra + parity, MSMYP: Modified spectra + milk yield + parity, R² for CBDEV: cow and breed dependent external validation coefficient of determination, RMSEP for CBDEV: Root Mean Square Error of prediction for cow and breed dependent external validation