J. Dairy Sci. 100:9052–9060
https://doi.org/10.3168/jds.2017-13014

Predicting methane emissions of lactating Danish Holstein cows using Fourier transform mid-infrared spectroscopy of milk

N. Shetty,*[1] G. Difford,*† J. Lassen,‡ P. Løvendahl,* and A. J. Buitenhuis*
*Department of Molecular Biology and Genetics, Center for Quantitative Genetics and Genomics, Aarhus University, DK 8830 Tjele, Denmark
†Animal Breeding and Genomics, Wageningen University and Research, NL 6700 AH Wageningen, the Netherlands
‡Viking Genetics, DK 8960 Randers SØ, Denmark

ABSTRACT

Enteric methane (CH_4), a potent greenhouse gas, is among the main targets of mitigation practices for the dairy industry. A measurement technique that is rapid, inexpensive, easy to use, and applicable at the population level is desired to estimate CH_4 emission from dairy cows. In the present study, feasibility of milk Fourier transform mid-infrared (FT-IR) spectral profiles as a predictor for CH_4:CO_2 ratio and CH_4 production (L/d) is explained. The partial least squares regression method was used to develop the prediction models. The models were validated using different random test sets, which are independent from the training set by leaving out records of 20% cows for validation and keeping records of 80% of cows for training the model. The data set consisted of 3,623 records from 500 Danish Holstein cows from both experimental and commercial farms. For both CH_4:CO_2 ratio and CH_4 production, low prediction accuracies were found when models were obtained using FT-IR spectra. Validated coefficient of determination (R^2_{Val}) = 0.21 with validated model error root mean squared error of prediction (RMSEP) = 0.0114 L/d for CH_4:CO_2 ratio, and R^2_{Val} = 0.13 with RMSEP = 111 L/d for CH_4 production. The important spectral wavenumbers selected using the recursive partial least squares method represented major milk components fat, protein, and lactose regions of the spectra. When fat and protein predicted by FT-IR were used instead of full spectra, a low R^2_{Val} of 0.07 was obtained for both CH_4:CO_2 ratio and CH_4 production prediction. Other spectral wavenumbers related to lactose (carbohydrate) or additional wavenumbers related to fat or protein (amide II) are providing additional variation when using the full spectral profile. For CH_4:CO_2 ratio prediction, integration of FT-IR with other factors such as milk yield, herd, and lactation stage showed improvement in the prediction accuracy. However, overall prediction accuracy remained modest; R^2_{Val} increased to 0.31 with RMSEP = 0.0105. For prediction of CH_4 production, the added value of FT-IR along with the aforementioned traits was marginal. These results indicated that for CH_4 production prediction, FT-IR profiles reflect primarily information related to milk yield, herd, and lactation stage rather than individual milk fatty acids related to CH_4 emission. Thus, it is not feasible to predict CH_4 emission based on FT-IR spectra alone.
Key words: CH_4 production, CH_4:CO_2 ratio, infrared spectroscopy, prediction, validation

INTRODUCTION

Methane (CH_4) emissions are of concern because of the observed climate change effects. Methane, a potent greenhouse gas that contributes substantially to global warming, accounts for approximately 52% of the greenhouse gas emissions in both developing and developed countries (FAO, 2010). Enteric CH_4 composes 17% of global methane and is therefore the single largest source of anthropogenic CH_4 (Knapp et al., 2014). Agriculture is considered to be the major producer of anthropogenic CH_4, and most CH_4 is naturally emitted by dairy cows during the microbial fermentation of feed components (Gerber et al., 2013). In addition to its relevance on climate impact, the eructed CH_4 induces a significant loss of gross energy intake as CH_4 reduces the availability of consumed energy for the cow (Johnson and Johnson, 1995).

Thus, mitigating CH_4 emissions seems to be an obvious approach to improving sustainability and profitability of dairy production. Measurement techniques applicable on a large scale to estimate specific amounts of CH_4 emitted by a cow would therefore be valuable. Several techniques have been developed to measure enteric CH_4 emissions (Patra, 2016). However, in practice, recording CH_4 emission is expensive and difficult for a large population. Milk Fourier transform mid-infrared (**FT-IR**) profile may be useful as an indicator trait for CH_4 emission. Milk FT-IR spectral data are rou-

Received April 12, 2017.
Accepted July 26, 2017.
[1]Corresponding author: nisha.shetty@mbg.au.dk

tinely generated for all individual milk samples; if it is feasible to predict CH_4 emission using FT-IR, the prediction can be obtained at the population level at no additional cost. Globally, FT-IR spectroscopy has been used for decades at commercial milk recording agencies and dairies for routine quantification of the major milk components fat, protein, and lactose content. Several studies have investigated the potential use of milk FT-IR spectroscopy to predict detailed milk quality traits such as individual milk fatty acids (**FA**) and proteins (De Marchi et al., 2014). Recently, numerous studies have also focused on the use of FT-IR spectroscopy to predict animal-related characteristics such as energy status (intake and balance), residual feed intake (McParland et al., 2011, 2012, 2014), and CH_4 emission (Dehareng et al., 2012; Vanlierde et al., 2015, 2016).

The objective of the present study was to evaluate the feasibility of FT-IR spectroscopy of milk to predict CH_4:CO_2 ratio and CH_4 production in lactating Danish Holstein cows. We investigated the added value of milk FT-IR spectroscopy in the prediction of CH_4:CO_2 ratio and CH_4 production. The present study used a considerably larger data set to develop the prediction models compared with preceding studies (Dehareng et al., 2012; Vanlierde et al., 2015, 2016). Moreover, the robustness of generated prediction models was validated using different random test sets that are independent from the training set.

MATERIALS AND METHODS

Data Collection

Data were collected from the Danish Cattle Research Centre (**DCRC**; Foulum, Denmark) and 2 Danish commercial farms (referred to as A and B). In total, 215 Danish Holstein cows from DCRC and 285 Danish Holstein cows from commercial farms had CH_4 measurements and corresponding FT-IR spectral profile. Milk samples were collected at DCRC and commercial herds and sent to the Eurofins-Steins laboratory (Vejen, Denmark) for FT-IR spectral analyses using a MilkoScan FT+ (Foss, Hillerød, Denmark). For DCRC, CH_4 measurements were taken April 2015 to September 2016. For commercial farms, herd A consisted of 158 cows, and CH_4 measurements were performed October 2015 to January 2016; herd B consisted of 127 cows, and CH_4 measurements were taken during February and March 2016.

Total data analyzed included 3,623 CH_4:CO_2 ratio records from 500 cows and 2,202 CH_4 production (L/d) records from 490 cows. Some values for CH_4 production were missing due to the missing insemination date information, which was used in the calculation of predicted CO_2 (Madsen et al., 2010) and then used to calculate CH_4 production. For DCRC data the FT-IR spectral data records representing 2 to 6 milk samples per week per cow were averaged on a weekly basis corresponding to the weekly available CH_4 records. At commercial farms approximately 1 spectrum per month per cow (standard milk control sample) was available; therefore, the raw spectra corresponding to the week of CH_4 measurement were considered for the analysis. Effects of using weekly averaged FT-IR spectra versus raw spectral data were evaluated by building prediction models for fat and protein. Prediction results did not differ when using weekly averaged versus raw spectral profile for both fat and protein contents. However, this could be different for CH_4 production if CH_4 is more variable with day-to-day changes. Additional traits such as weekly averages for daily milk yield (**MY**; kg/d), herd, parity, lactation stage, season, weekly averaged fat, and protein content were used as predictor traits for CH_4:CO_2 ratio and CH_4 production.

Lactation stage included 1 to 44 wk and is described using a Wilmink exponential function (Wilmink, 1987). Seasonal variation is described using a Fourier series approach with sine and cosine functions of day of the year (previously used by Løvendahl and Bjerring, 2006).

Reference Data

At DCRC, CH_4 and CO_2 gas concentrations (in ppm) from breath samples were analyzed using the noninvasive "sniffer" method. At DCRC nondispersive infrared (**NDIR**) gas analyzers were installed within each automated milking station (**AMS**; Guardian NG/Gascard, Edinburgh Instruments Ltd., Livingston, UK) and at commercial farms using both NDIR and the portable Fourier transform infrared analyzer (Gasmet DX 4000, Gasmet Technologies Oy, Helsinki, Finland. Difford et al. (2016) demonstrated how measurements of CH_4 and CO_2 from both instruments can be used interchangeably. Thus, it was possible to combine the data from DCRC and commercial farms.

The CH_4 and CO_2 gas concentrations for each cow milking in the AMS were corrected for ambient barn concentrations and diurnal variation and averaged over a full week of lactation to calculate the CH_4:CO_2 ratio (Lassen et al., 2012). The barn concentrations were estimated from the ambient concentrations within the AMS during the morning cleaning cycle when no cows were present. The ECM, BW, and gestation length of each cow were used to estimate heat production (CIGR, 2002), which was converted to CO_2 production (Pedersen et al., 2008) and multiplied by CH_4:CO_2 as recommended by Madsen et al. (2010) to generate CH_4 production (L/d). The BW was calculated as the average BW from all milkings in the AMS over the week of measurement. The ECM was calculated as the average MY during the week of measurement corrected for fat, protein, and lactose content, which were estimated from the FT-IR spectra during the week of measurement (Sjaunja et al., 1990).

Prediction Model Development

Data were analyzed using MATLAB (version R2015a; The MathWorks, Natick, MA) along with PLS toolbox (Eigenvector Research Inc., Manson, WA). The FT-IR spectral data represented infrared light transmittance through the milk sample at wavenumber regions (frequencies) ranging from 925 to 5,008 cm^{-1}, which comprises 1,060 wavenumbers. The spectra were transformed from transmittance to absorbance to obey Beer's law (Swinehart, 1962). To improve the repeatability of the measurement, 241 spectral data points from 3 spectral regions (1,000–1,550 cm^{-1}, 1,705–1,820 cm^{-1}, and 2,700–2,955 cm^{-1}) were retained for the analyses, and remaining uninformative spectral regions together with high water absorption, O–H stretching (between ~1,600 and 1,700 cm^{-1}), and O–H bending (>3,005 cm^{-1}) regions were omitted (Eskildsen et al., 2014).

The partial least squares regression (**PLSR**) method (Wold et al., 1983; Martens and Naes, 1989) was used to develop the prediction models. The method has been widely used in quantitative spectroscopy to determine a relationship between spectral profile (i.e., predictors), which is collinear in nature, and related chemical or physical data (i.e., predictants). Models were developed with and without preprocessing the spectral data. The literature showed quite similar results when different preprocessing is applied to milk FT-IR spectra. One study showed better accuracies using the Savitzky–Golay derivative (Savitzky and Golay, 1964) compared with untreated spectral data, and in another study better accuracies were found using untreated data (De Marchi et al., 2014). Therefore, in the present study models were tested using both untreated data and Savitzky–Golay first derivatives with filter width 7 and polynomial order 2. Prediction models were developed using different combinations of predictor variables, namely FT-IR spectra, MY, parity, herd, lactation stage, and season. Variables with different units were equally scaled using the autoscaling method.

Optimal PLSR components were chosen based on the lowest root mean squared error of cross-validation value using 10 random iterations; this was to avoid over- or underfitting in the final PLSR model. Partial least squares regression Hotelling T^2 versus Q-residual plots were used for outlier analysis. Coefficients of determination for validation ($\mathbf{R^2_{Val}}$) and coefficients of determination for calibration along with model error root mean squared error of calibration and prediction (**RMSEP**; Esbensen, 2000) were used to evaluate the predictive ability of the obtained calibration models.

Both experimental (i.e., DCRC) and commercial farm data were combined to increase variation in data. The robustness of the obtained prediction model was evaluated using different test sets that are independent from the training set (external test sets). For CH_4:CO_2 ratio, out of the 500 cows, 20% (n = 100 cows) were randomly left out for validation and the remaining 80% (n = 400 cows) were used to train the model. For CH_4 production, out of the 490 cows, 20% (n = 98 cows) were randomly left out for validation and the remaining 80% (n = 392 cows) were used to train the model. In total, 50 random iterations were performed to compare the models for robustness.

Variable Selection

Important spectral variables were identified using the variable importance for projection (**VIP**) method and the recursive weighted partial least squares (**rPLS**) variable selection method (Rinnan et al., 2014). Variable importance for projection is a measure of how well a variable contributes to describing both predictors and response variables. A variable with a VIP score >1 indicates importance for the developed PLSR model (Andersen and Bro, 2010). In rPLS, the regression coefficients are used as weights on the original data matrix. The procedure is based on repeated PLS models, and the rPLS method iteratively uses the regression coefficients to boost important variables and thus relatively downweight less important ones. Therefore, the rPLS model helps identify most specific wavenumbers of important FT-IR regions. Large absolute weights indicate important variables, and weights close to zero indicate less important variables (Rinnan et al., 2014).

RESULTS

Data Description

Descriptive statistics of the CH_4:CO_2 ratio and CH_4 production (L/d) are shown in Table 1. Average CH_4:CO_2 ratio was 0.069 (SD = 0.013), and average CH_4 production was 437 L/d (SD = 119 L/d). Table 2 gives the Pearson correlations between different predictor traits and CH_4 measurements. Milk yield had moderate correlation with CH_4 production, whereas correlation of MY with CH_4:CO_2 ratio was lower.

Table 1. Descriptive statistics of CH_4:CO_2 ratio and CH_4 production

Item	Records (no.)	Cows (no.)	Mean	SD	Minimum	Maximum
CH_4:CO_2 ratio	3,623	500	0.069	0.013	0.023	0.124
CH_4 production (L/d)	2,202	490	437	119	53	896

Prediction and Validation

Prediction model results for CH_4:CO_2 ratio and CH_4 production using different trait combinations are shown in Table 3 and Table 4, respectively, and results shown are averages based on 50 random iterations. Spectral data preprocessing using first derivative did not show improvement compared with raw spectra models. Therefore, results shown are based on untreated spectral data.

CH_4:CO_2 Ratio and CH_4 Production Prediction

The milk FT-IR full spectral profile (241 wavenumbers) used as predictor traits in CH_4:CO_2 ratio prediction resulted in low prediction capacity when validated using the external test set for both CH_4:CO_2 ratio and CH_4 production (Tables 3 and 4). Integration of FT-IR along with traits MY and herd showed enhancement in the prediction accuracy of CH_4:CO_2; R^2_{Val} increased from 0.21 to 0.30, and RMSEP decreased from 0.0114 to 0.0107 (Table 3). On the other hand, CH_4 prediction accuracy increased when FT-IR spectra were integrated with MY; R^2_{Val} increased from 0.13 to 0.35, and RMSEP decreased from 111 to 96 L/d. However, the addition of herd showed a marginal effect (Table 4). For both CH_4:CO_2 ratio and CH_4 production, further addition of the predictor trait lactation stage showed marginal improvement, whereas adding the predictor traits parity and season did not improve accuracy (Tables 3 and 4).

Selected FT-IR Wavenumbers for CH_4:CO_2 Ratio and CH_4 Production Prediction

Figure 1a and b illustrates VIP scores and reflects the important FT-IR spectral regions in CH_4:CO_2 ratio and CH_4 production prediction, respectively. Using the rPLS variable selection method, the 13 most important spectral wavenumbers were identified for CH_4:CO_2 ratio and the 15 most relevant wavenumbers were selected for CH_4 production. Table 5 provides selected wavenumbers and corresponding chemical functional groups, milk components, feed, and nutrient information. Following the selection of these wavenumbers, they were used to develop prediction models. For CH_4:CO_2 ratio, this subset of wavenumbers explained 12% of total variation, whereas the full FT-IR spectra explained 21% of the total variation (Table 3). For CH_4 production, the selected wavenumbers explained 9% of total variation, whereas the full FT-IR spectra explained about 13% of total variation in CH_4 production (Table 4). For both CH_4:CO_2 ratio and CH_4 production, selected wavenumbers corresponded to regions of the milk spectral profile specific to the major milk components fat, protein, and lactose.

When FT-IR predicted fat is used instead of full FT-IR spectra, about 7 and 3% of the variation for CH_4:CO_2 ratio and CH_4 production, respectively, is explained. When FT-IR predicted protein is used, 3 and 7% of variation for CH_4:CO_2 ratio and CH_4 production, respectively, is explained. Combining fat and protein did not improve the prediction accuracy with respect to fat and protein alone for both CH_4:CO_2 ratio and CH_4 production (Tables 3 and 4). That indicates that other spectral wavenumbers related to lactose (carbohydrate) or additional wavenumbers related to fat and protein (amide II) are providing additional variation when using the full spectral profile (Table 3).

DISCUSSION

The basic assumption underlying multivariate prediction is that measuring phenotypes and variables carries information about the property (predictant or dependent variable) we are seeking to predict (Esbensen, 2000). In the present study, the question of interest is how the CH_4 gas concentration from breath samples is reflected in cow milk composition and thereby in milk FT-IR profile.

Prediction of CH_4 Emission Based on Milk FT-IR

In general, CH_4 emission is connected to milk FA profile, so one can assume that FT-IR spectral profile could predict CH_4 emission (van Gastelen and Dijkstra, 2016). Several studies have linked individual milk FA to CH_4 emission (Chilliard et al., 2009; van Lingen

Table 2. Pearson correlations between different predictor traits and CH_4:CO_2 ratio and CH_4 production

Trait	CH_4:CO_2 ratio	CH_4 production (L/d)
Milk yield (kg)	0.23	0.53
Fat (%)	−0.26	−0.18
Protein (%)	−0.17	−0.26
Week of lactation	0.18	0.02

Table 3. Partial least squares regression (PLSR) fit statistics for CH_4:CO_2 ratio[1,2]

Predictor trait[3]	PLSR factors	R^2_{Cal}	R^2_{Val}	RMSEC	RMSEP
MY	1	0.05	0.06	0.0123	0.0126
MY + herd	2	0.14	0.14	0.0117	0.0117
MY + herd + parity	4	0.14	0.14	0.0117	0.0117
MY + herd + parity + lactation stage	5	0.22	0.20	0.0112	0.0113
MY + herd + parity + lactation stage + season	5	0.22	0.21	0.0111	0.0115
Lactation stage	1	0.04	0.04	0.0124	0.0124
IR (full spectra)	12	0.27	0.21	0.0108	0.0114
IR (full spectra) + MY	13	0.30	0.24	0.0106	0.0111
IR (full spectra) + MY + herd	14	0.33	0.30	0.0103	0.0107
IR (full spectra) + MY + herd + parity	14	0.33	0.27	0.0104	0.0108
IR (full spectra) + MY + herd + parity + lactation stage	15	0.35	0.31	0.0098	0.0106
IR (full spectra) + MY + herd + parity + lactation stage + season	15	0.35	0.31	0.0098	0.0105
IR (full spectra) + lactation stage	13	0.27	0.20	0.0108	0.0115
IR (selected 13 wavenumbers)	10	0.14	0.12	0.0117	0.0120
Fat (%)	1	0.06	0.07	0.0123	0.0121
Protein (%)	1	0.03	0.03	0.0124	0.0126
Fat + protein	2	0.07	0.07	0.0122	0.0123
Fat + protein + MY	3	0.09	0.09	0.0121	0.0121

[1]R^2_{Cal} = coefficient of determination for calibration; R^2_{Val} = coefficient of determination for validation; RMSEC = root mean squared error of calibration; RMSEP = root mean squared error of prediction.

[2]Results are average values from 50 random iterations.

[3]MY = milk yield (kg); IR = mid-infrared.

et al., 2014; Rico et al., 2016). Eskildsen et al. (2014) demonstrated that the FT-IR predictions of milk FA are reliant on indirect correlations, which are confined to covariance structures in the data set. In their study, indirect correlations were found primarily due to the variation associated with total fat content and breed. Also, Shetty et al. (2017) found discrimination between breeds (Holstein and Jersey) in milk FT-IR profile, and it was primarily due to the fat regions of the spectra. Hoover and Miller (1991) reported that fat and fiber in feed converts to nutrient FA in the rumen, which becomes milk fat. These findings might explain why the fat regions of the milk infrared spectra were informative for both CH_4:CO_2 ratio and CH_4 production prediction (Table 5). However, in the present study, when models were obtained using FT-IR predicted total fat, rather low prediction accuracies were found for both CH_4:CO_2 ratio and CH_4 production.

Table 4. Partial least squares regression (PLSR) fit statistics for CH_4 production (L/d)[1,2]

Predictor trait[3]	PLSR factors	R^2_{Cal}	R^2_{Val}	RMSEC	RMSEP
MY	1	0.28	0.27	101	102
MY + herd	2	0.32	0.33	98	99
MY + herd + parity	4	0.33	0.31	98	98
MY + herd + parity + lactation stage	5	0.36	0.36	96	95
MY + herd + parity + lactation stage + season	5	0.37	0.36	95	95
Lactation stage	1	0.05	0.05	116	119
IR (full spectra)	12	0.21	0.13	106	111
IR (full spectra) + MY	13	0.39	0.35	93	96
IR (full spectra) + MY + herd	14	0.42	0.37	91	94
IR (full spectra) + MY + herd + parity	14	0.41	0.37	91	94
IR (full spectra) + MY + herd + parity + lactation stage	15	0.42	0.39	90	94
IR (full spectra) + MY + herd + parity + lactation stage + season	15	0.44	0.39	89	94
IR (full spectra) + lactation stage	13	0.22	0.14	105	112
IR (selected 15 wavenumbers)	10	0.12	0.09	112	115
Fat (%)	1	0.03	0.03	117	119
Protein (%)	1	0.07	0.07	115	115
Fat + protein	2	0.07	0.07	115	113
Fat + protein + MY	3	0.28	0.27	112	113

[1]R^2_{Cal} = coefficient of determination for calibration; R^2_{Val} = coefficient of determination for validation; RMSEC = root mean squared error of calibration; RMSEP = root mean squared error of prediction.

[2]Results are average values from 50 random iterations.

[3]MY = milk yield (kg); IR = mid-infrared.

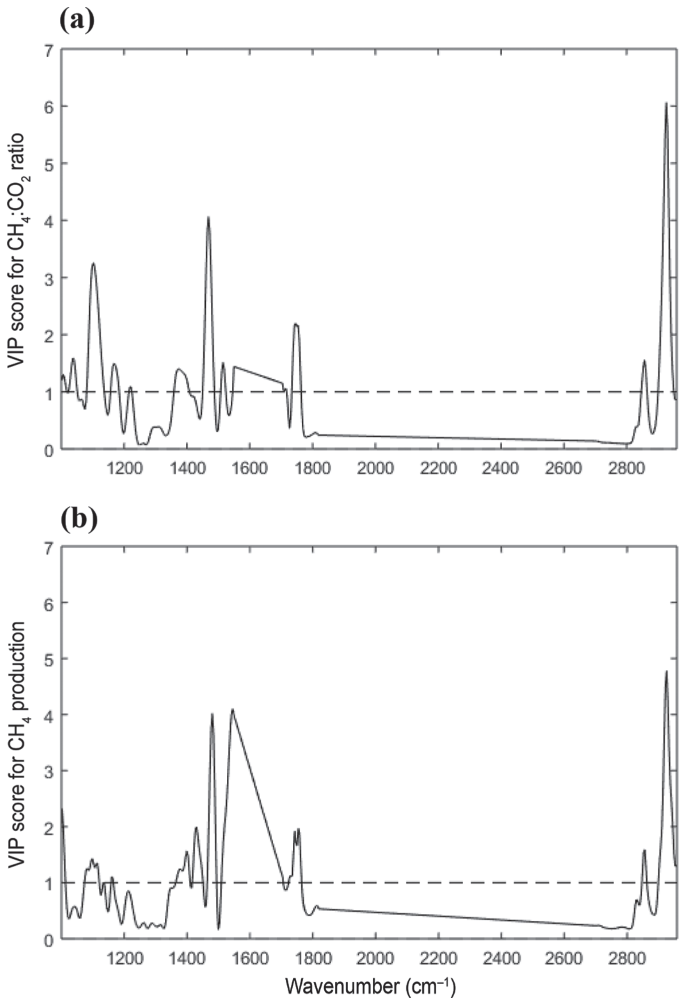

Figure 1. Variable importance for projection (VIP) scores of (a) CH_4:CO_2 ratio and (b) CH_4 production (L/d) plotted against wavenumber (cm^{-1}).

Currently, 2 studies have investigated prediction of CH_4 emission of individual cows directly using FT-IR spectral profile. Dehareng et al. (2012) used 11 lactating Holstein cows in 2 experiments and 3 dietary treatments. The prediction accuracy of models developed for CH_4 production (g/d) and CH_4 milk (g/kg of milk) on this small data set is somewhat high; the cross-validated coefficient of determination ranged from 0.68 to 0.79. In another study, lactation stage–dependent prediction models of CH_4 emission from FT-IR spectral profile were developed using 532 CH_4 measurements of 165 Holstein, Jersey, and Holstein-Jersey cows and had a cross-validated R^2 of 0.70 (Vanlierde et al., 2015, 2016). Conversely, in the present study much lower prediction accuracies were obtained when models were developed using FT-IR spectra alone as well as integration with the lactation stage information for both CH_4:CO_2 ratio and CH_4 production.

The large discrepancy in prediction accuracies between the aforementioned studies and the present study is of concern. Possible sources of the discrepancy could be differences in the methods used to measure CH_4 emission, the size and structure of the populations, prediction and validation methods, or some combination of these. In Dehareng et al. (2012) and Vanlierde et al. (2015, 2016), CH_4 emission was measured using the sulfur hexafluoride (SF_6) tracer method, whereas the present study used sniffers installed in AMS and CO_2 as a tracer gas. The 2 methods have not been compared directly; however, both have been benchmarked against the intensive respiration chamber method. McGinn et al. (2006) found an R^2_{Val} of 0.80 and a concordance correlation coefficient of 0.79 between the SF_6 method and respiration chambers for 8 beef heifers. Garnsworthy et

Table 5. Recursive weighted partial least squares regression selected important spectral wavenumbers for CH_4:CO_2 ratio and CH_4 production (L/d) along with corresponding chemical functional groups, milk components, feed, and nutrient information

Wavenumber (cm^{-1})	Milk components	Nutrients	Feed	Functional group
CH_4:CO_2 ratio				
2,957, 2,953, 2,876	Fat B[1]	Fatty acids, acetic, butyric	Fermentable fiber	Alkyl chain (C–H)
1,727	Fat A[1]	Fatty acids, acetic, butyric	Fermentable fiber	Carbonyl group (C=O)
1,376	Protein (amide II)	AA	CP	Aromatic amines (N–H)
1,249	Protein (amide III)	AA	CP	Aromatic amines (C–N)
1,172, 1,137, 1,099, 1,076, 1,068, 1,064, 1,002	Lactose (carbohydrate)	Propionic (glucose)	Sugar, starch	Hydroxyl group (O–H)
CH_4 production (L/d)				
2,957, 2,953, 2,826	Fat A[1]	Fatty acids, acetic, butyric	Fermentable fiber	Alkyl chain (C–H)
1,519	Protein (amide II)	AA	CP	Aromatic amines (N–H)
1,276	Protein (amide III)	AA	CP	Aromatic amines (C–N)
1,195, 1,130, 1,099, 1,083, 1,079, 1,072, 1,068, 1,064, 1,056, 1,029	Lactose (carbohydrate)	Propionic (glucose)	Sugar, starch	Hydroxyl group (O–H)

[1]The measurement of fat by milk Fourier transform mid-infrared spectra used the carbonyl stretch (C=O), which has been called fat A, and the symmetrical carbon hydrogen stretch (C–H), which has been called fat B, in the Fourier transform mid-infrared spectral region (Biggs et al., 1987; Barbano and Clark, 1989).

al. (2012) found measurements from the NDIR installed in the AMS to have an R^2_{Val} of 0.79 with subsequent measurements in respiration chambers for 12 Holstein cows. Likewise, Negussie et al. (2017) found CH_4 emissions using a photoacoustic infrared sniffer and CO_2 as a tracer gas to have a high concordance (concordance correlation coefficient = 0.71) and an R^2_{Val} of 0.70 in 21 Nordic Red cows. Haque et al. (2014) found CH_4 emissions from FT-IR installed in AMS and CO_2 as a tracer gas to have statistical power equivalent to SF_6. Based on these findings, one could expect some decrease in prediction accuracy using CH_4 emission from sniffers and CO_2 as a tracer gas, but not to the extent observed.

The present method of estimating CH_4 emission is based on 3 components: the measured ratio of gases, estimation of heat production units (**HPU**), and the conversion of HPU to CO_2 production. However, the specific accuracies of these components have yet to be evaluated and their relative contributions to CH_4 emission accuracy have yet to be quantified. It is important to note that the most influential predictor of HPU is ECM, which is milk production corrected for fat, protein, and lactose percentages estimated from FT-IR spectra. Based on this, one would expect some autocorrelations, resulting in higher accuracies of prediction CH_4 emission from sniffers and CO_2 tracer gases predicted from FT-IR spectra.

A further difference between the present and aforementioned studies (Dehareng et al., 2012; Vanlierde et al., 2015, 2016) is the duration of measurement and time between CH_4 measurements and milk FT-IR sampling. In the present study, weekly averaged CH_4 measurements were predicted from corresponding weekly averaged milk FT-IR spectra of an experimental farm and commercial farms from a single milk FT-IR falling within the week of CH_4 measurement, whereas Dehareng et al. (2012) and Vanlierde et al. (2015, 2016) used daily CH_4 measurements to perform the prediction models from milk FT-IR spectra averaged from the morning and afternoon milkings on each measurement day. Logically, reducing the time between milk FT-IR and CH_4 measurements should maximize the biological relationships on which the prediction equations are built. However, Moate et al. (2012) found that the DMI of the preceding day to CH_4 measurement is responsible for 30% of the variation in CH_4 on the day of measurement. Moreover, CH_4 measurements have been shown to be more stable when averaged over 3 to 5 consecutive days for SF_6 (Deighton et al., 2014; Arbre et al., 2016) and for 1 wk for the sniffer method (Difford et al., 2016). Thus, further work is needed to determine the optimization between maximizing the stability of the CH_4 measurement and minimizing the time between representative CH_4 measurements and FT-IR milk spectra and their respective effects on milk FT-IR prediction of CH_4 emission.

In Dehareng et al. (2012) and Vanlierde et al. (2015, 2016), cows were fed a variety of diets within different experimental conditions in different countries to maximize the variability of individual CH_4 emission and thereby increase the spectral variability required for a robust calibration model. However, instances where experimental treatments affect CH_4 emission as well as milk composition induce covariance structure into the data set and can inflate the prediction accuracies when not properly accounted for in the validation method. In contrast, the current study used a larger sample of Holstein cows from both experimental and commercial farms fed a TMR used in commercial practice. This sampling strategy is more representative of animals and conditions in the national Danish population and is less likely to suffer from induced covariance structure caused by extraneous factors coaffecting CH_4 emission and milk composition.

For robust FT-IR predictions, the indirect covariance structures in the calibration data set must be valid for future new samples (Eskildsen et al., 2014). In the present study, models developed were validated using an independent (data not used in the calibration model) or external test set by leaving out records of 20% cows for validation and keeping records of 80% cows for training the model with 50 random iterations. In Dehareng et al. (2012) and Vanlierde et al. (2015, 2016), models were validated using a cross-validation method. In cross-validation, data are divided into complementary subsets, training the model on one subset and validating on the other subset for a specific fold. However, if no precautions are taken, the same cow could have records in both training and validation sets for this specific fold. This may result in overfitted prediction models, more so if there is unaccounted experimental treatment structure within the data. In the present study, additional validations were performed (results not shown) to ensure that the obtained model is robust in predicting new data sets of similar types. On one hand, experimental farm data were used to build the training model and cross-validation was performed to evaluate the model performance for CH_4:CO_2 ratio and CH_4 production; results showed that similar coefficient of determination values were found for both calibration and cross-validation. On the other hand, the obtained model using experimental farm data was evaluated using commercial farm data, which is a completely independent test set; results showed negligible R^2_{val} compared with coefficients of determination for calibration for both CH_4:CO_2 ratio and CH_4 production, indicating that variation in commercial farm data is not included in the experimental data.

Prediction of CH_4 Emission Based on Factors Other than FT-IR Spectra

In their study, van Gastelen and Dijkstra (2016) suggested that integration of other factors such as nutrient composition of the feed, parity, and lactation stage along with FT-IR spectra may improve the prediction of CH_4 emission. In the present study, traits such as MY, herd, parity, lactation stage, and season along with FT-IR profiles were used as predictor traits. For the CH_4:CO_2 ratio prediction model, factors MY, herd, and lactation stage along with FT-IR spectra showed improvement; however, overall prediction accuracy still remained modest (Tables 3 and 4). For CH_4 production prediction, integration of FT-IR along with the aforementioned traits showed only a marginal improvement (Table 4). These results indicated that FT-IR profile reflected primarily information related to MY, herd, and lactation stage rather than relevant individual milk FA, which are generally assumed to be connected to CH_4 emission. In Vanlierde et al. (2015, 2016), CH_4 predictions using FT-IR spectra alone were compared with CH_4 predictions using FT-IR spectra and lactation stage information. The models were hardly different and had a cross-validated R^2 of 0.70 (vs. 0.67) and a standard error of cross-validation of 66 g/d (vs. 63). However, in contrast with the prediction using FT-IR spectra only, the lactation stage–dependent model showed biologically meaningful CH_4 values (i.e., an increase in CH_4 production between 0 and 100 DIM and a decrease thereafter). The present study also showed only marginal improvement when adding lactation stage information along with FT-IR spectra. Obtained prediction accuracies were much lower in the current study (Tables 3 and 4); therefore, it was not possible to confirm the advantage of the lactation stage–dependent model.

CONCLUSIONS

In the present study, feasibility of FT-IR spectroscopy of milk in predicting CH_4 emission is elucidated. Low prediction accuracies were obtained for CH_4:CO_2 ratio and CH_4 production when models were obtained using FT-IR spectra. When models were integrated with other factors such as MY, herd, and lactation stage, results showed improvement in the prediction accuracy; however, overall prediction accuracy still remained modest. For CH_4, production prediction added value of the FT-IR along with the aforementioned traits was marginal. This implies that for CH_4 production prediction, FT-IR profiles reflect primarily information related to MY, herd, and lactation stage rather than relevant individual milk FA, which are generally assumed to be connected to CH_4 emission. Therefore, it is not feasible to predict CH_4 emission based on FT-IR spectra alone. Moreover, the present study uses weekly averaged milk FT-IR spectra to predict weekly averaged CH_4 measurements. Improving CH_4 measurement may improve the accuracy of its prediction using FT-IR spectra. Thus, further work is needed to determine the optimization between maximizing the stability of the CH_4 measurements and minimizing the time between representative CH_4 measurements and FT-IR milk spectra.

ACKNOWLEDGMENTS

This study was financially supported by two projects: "FT-IR spektre i mælk: Genetisk variation og effekt på sundhed, frugtbarhed og energibalance" of the Milk Levy Fund, Denmark (2015–2017), and Robust and Efficient Dairy Cows, project no. 34009-14-0848 of Grønt Udviklings- og Demonstrations Program (2015–2018). The authors thank the Foundation for Registration and Milk Recording (Aarhus, Denmark) for providing data.

REFERENCES

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

Arbre, M., Y. Rochette, J. Guyader, C. Lascoux, D. P. Morgavi, G. Renand, M. Doreau, and C. Martin. 2016. Repeatability of enteric methane determinations from cattle using either the SF6 tracer technique or the GreenFeed system. Anim. Prod. Sci. 56:238–243. https://doi.org/10.1071/AN15512.

Barbano, D. M., and J. L. Clark. 1989. Infrared milk analysis: Challenges for the future. J. Dairy Sci. 72:1627–1636.

Biggs, D. A., G. Johnsson, and L. O. Sjanja. 1987. Analysis of fat, protein, lactose, and total solids by infrared absorption. Pages 21–30 in Rapid Methods for Measurement of the Major Components of Milk. Bulletin 208. Int. Dairy Fed., Brussels, Belgium.

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. https://doi.org/10.3168/jds.2009-2375.

CIGR (International Commission of Agricultural and Biosystems Engineering). 2002. 4th Report of Working Group on Climatization of Animal Houses: Heat and moisture production at animal and house levels. S. Pedersen and K. Sällvik, ed. DIAS, Horsens, Denmark.

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.

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.

Deighton, M. H., S. R. O. Williams, M. C. Hannah, R. J. Eckard, T. M. Boland, W. J. Wales, and P. J. Moate. 2014. A modified sulphur hexafluoride tracer technique enables accurate determination of enteric methane emissions from ruminants. Anim. Feed Sci. Technol. 197:47–63. https://doi.org/10.1016/j.anifeedsci.2014.08.003.

Difford, G. F., J. Lassen, and P. Løvendahl. 2016. Interchangeability between methane measurements in dairy cows assessed by comparing precision and agreement of two non-invasive infrared methods. Comput. Electron. Agric. 124:220–226. https://doi.org/10.1016/j.compag.2016.04.010.

Esbensen, K. H. 2000. Multivariate Data Analysis—In Practice. 4th ed. CAMO, Oslo, Norway.

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.

FAO (Food and Agriculture Organization of the United Nations). 2010. Greenhouse Gas Emissions from the Dairy Sector: A Life Cycle Assessment. FAO, Rome, Italy.

Garnsworthy, P. C., J. Craigon, J. H. Hernandez-Medrano, and N. Saunders. 2012. On-farm methane measurements during milking correlate with total methane production by individual dairy cows. J. Dairy Sci. 95:3166–3180. https://doi.org/10.3168/jds.2011-4605.

Gerber, P. J., H. Steinfeld, B. Henderson, A. Mottet, C. Opio, J. Dijkman, A. Falcucci, and G. Tempio. 2013. Tackling Climate Change Through Livestock: A Global Assessment of Emissions and Mitigation Opportunities. Food and Agriculture Organization of the United Nations, Rome, Italy.

Haque, M. N., C. Cornou, and J. Madsen. 2014. Estimation of methane emission using the CO_2 method from dairy cows fed concentrate with different carbohydrate compositions in automatic milking system. Livest. Sci. 164:57–66. https://doi.org/10.1016/j.livsci.2014.03.004.

Hoover, W. H., and T. K. Miller. 1991. Rumen digestive physiology and microbial ecology. Pages 311–325 in The Veterinary Clinics of North America: Food Animal Practice. Vol. 7, No. 2. C. J. Sniffen and T. H. Herdt, ed. Saunders, Philadelphia, PA.

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

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. https://doi.org/10.3168/jds.2013-7234.

Lassen, J., P. Løvendahl, and J. Madsen. 2012. Accuracy of noninvasive breath methane measurements using Fourier transform infrared methods on individual cows. J. Dairy Sci. 95:890–898. https://doi.org/10.3168/jds.2011-4544.

Løvendahl, P., and M. A. Bjerring. 2006. Detection of carryover in automated milk sampling equipment. J. Dairy Sci. 89:3645–3652. https://doi.org/10.3168/jds.S0022-0302(06)72404-3.

Madsen, J., B. S. Bjerg, T. Hvelplund, M. R. Weisbjerg, and P. Lund. 2010. Methane and carbon dioxide ratio in excreted air for quantification of the methane production from ruminants. Livest. Sci. 129:223–227. https://doi.org/10.1016/j.livsci.2010.01.001.

Martens, H., and T. Naes. 1989. Multivariate Calibration. Wiley, Hoboken, NJ.

McGinn, S. M., K. A. Beauchemin, D. Iwaasa, and T. A. McAllister. 2006. Assessment of the sulfur hexafluoride (SF) tracer technique for measuring enteric methane emissions from cattle. J. Environ. Qual. 35:1686–1691. https://doi.org/10.2134/jeq2006.0054.

McParland, S., G. Banos, B. McCarthy, E. Lewis, M. P. Coffey, B. O'Neill, M. O'Donovan, E. Wall, and D. P. Berry. 2012. Validation of mid-infrared spectrometry in milk for predicting body energy status in Holstein-Friesian cows. J. Dairy Sci. 95:7225–7235. https://doi.org/10.3168/jds.2012-5406.

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.

Moate, P. J., M. H. Deighton, M. C. Hannah, and S. R. O. Williams. 2012. Intake effects on methane emissions from dairy cows. Pages 60–61 in Proc. CCRSPI conference, Melbourne, Australia, 27–29 November.

Negussie, E., J. Lehtinen, P. Mäntysaari, A. R. Bayat, A.-E. Liinamo, E. A. Mäntysaari, and M. H. Lidauer. 2017. Non-invasive individual methane measurement in dairy cows. Animal 11:890–899. https://doi.org/10.1017/S1751731116002718.

Patra, A. K. 2016. Recent advances in measurement and dietary mitigation of enteric methane emissions in ruminants. Front. Vet. Sci. 3:39. https://doi.org/10.3389/fvets.2016.00039.

Pedersen, S., V. Blanes-Vidal, H. Joergensen, A. Chwalibog, A. Haeussermann, M. J. W. Heetkamp, and A. J. A. Aarnink. 2008. Carbon dioxide production in animal houses: A literature review. CIGR J. 10:1–19. https://doi.org/10.1017/CBO9781107415324.004.

Rico, D. E., P. Y. Chouinard, F. Hassanat, C. Benchaar, and R. Gervais. 2016. Prediction of enteric methane emissions from Holstein dairy cows fed various forage sources. Animal 10:203–211. https://doi.org/10.1017/S1751731115001949.

Rinnan, Å., M. Andersson, C. Ridder, and S. B. Engelsen. 2014. Recursive weighted partial least squares (rPLS): An efficient variable selection method using PLS. J. Chemometr. 28:439–447. https://doi.org/10.1002/cem.2582.

Savitzky, A., and M. J. E. Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36:1627–1638.

Shetty, N., P. Løvendahl, M. S. Lund, and A. J. Buitenhuis. 2017. Prediction and validation of residual feed intake and dry matter intake in Danish lactating dairy cows using mid-infrared spectroscopy of milk. J. Dairy Sci. 100:253–264. https://doi.org/10.3168/jds.2016-11609.

Sjaunja, L. O., L. Baevre, L. Junkkarinen, J. Pedersen, and J. Setala. 1990. A Nordic proposal for an energy corrected milk (ECM) formula. Pages 156–157 in Proc. European Federation of Animal Science (EAAP) Publication 50. Centre for Agricultural Publishing and Documentation, Wageningen, the Netherlands.

Swinehart, D. F. 1962. The Beer-Lambert law. J. Chem. Educ. 39:333–335. https://doi.org/10.1021/ed039p333.

van Gastelen, S., and J. Dijkstra. 2016. Prediction of methane emission from lactating dairy cows using milk fatty acids and mid-infrared spectroscopy. J. Sci. Food Agric. 96:3963–3968. https://doi.org/10.1002/jsfa.7718.

van Lingen, H. J., L. A. Crompton, W. H. Hendriks, C. K. Reynolds, and J. Dijkstra. 2014. Meta-analysis of relationships between enteric methane yield and milk fatty acid profile in dairy cattle. J. Dairy Sci. 97:7115–7132. https://doi.org/10.3168/jds.2014-8268.

Vanlierde, A., M.-L. Vanrobays, F. Dehareng, E. Froidmont, H. Soyeurt, S. McParland, E. Lewis, M. H. Deighton, 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. 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.org/10.1071/AN15590.

Wilmink, J. B. M. 1987. Adjustment of lactation yield for age at calving in relation to level of production. Livest. Prod. Sci. 16:321–334.

Wold, S., H. Martens, and H. Wold. 1983. The multivariate calibration problem in chemistry solved by the PLS method. Pages 286–293 in Proc. Conf. on Matrix Pencils, Lectures Notes in Mathematics. A. Ruhe and B. Kagstrom, ed. Springer, Heidelberg, Germany.