



Diagnosing the pregnancy status of dairy cows: How useful is milk mid-infrared spectroscopy?

P. Delhez,^{1,2}  P. N. Ho,^{3*}  N. Gengler,²  H. Soyeurt,²  and J. E. Pryce^{3,4} 

¹National Fund for Scientific Research (F.R.S.-FNRS), Egmont 5, Brussels 1000, Belgium

²Terra Teaching and Research Centre, Gembloux Agro-Bio Tech, University of Liège, Gembloux 5030, Belgium

³Centre for AgriBioscience, AgriBio, Agriculture Victoria, Bundoora, Victoria 3083, Australia

⁴School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia

ABSTRACT

Pregnancy diagnosis is an essential part of successful breeding programs on dairy farms. Milk composition alters with pregnancy, and this is well documented. Fourier-transform mid-infrared (MIR) spectroscopy is a rapid and cost-effective method for providing milk spectra that reflect the detailed composition of milk samples. Therefore, the aim of this study was to assess the ability of MIR spectroscopy to predict the pregnancy status of dairy cows. The MIR spectra and insemination records were available from 8,064 Holstein cows of 19 commercial dairy farms in Australia. Three strategies were studied to classify cows as open or pregnant using partial least squares discriminant analysis models with random cow-independent 10-fold cross-validation and external validation on a cow-independent test set. The first strategy considered 6,754 MIR spectra after insemination used as independent variables in the model. The results showed little ability to detect the pregnancy status as the area under the receiver operating characteristic curve was 0.63 and 0.65 for cross-validation and testing, respectively. The second strategy, involving 1,664 records, aimed to reduce noise in the MIR spectra used as predictors by subtracting a spectrum before insemination (i.e., open spectrum) from the spectrum after insemination. The accuracy was comparable with the first approach, showing no superiority of the method. Given the limited results for these models when using combined data from all stages after insemination, the third strategy explored separate models at 7 stages after insemination comprising 348 to 1,566 records each (i.e., progressively greater gestation) with single MIR spectra after insemination as

predictors. The models developed using data recorded after 150 d of pregnancy showed promising prediction accuracy with the average value of area under the receiver operating characteristic curve of 0.78 and 0.76 obtained through cross-validation and testing, respectively. If this can be confirmed on a larger data set and extended to somewhat earlier stages after insemination, the model could be used as a complementary tool to detect fetal abortion.

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

INTRODUCTION

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

Fourier-transform mid-infrared (MIR) spectroscopy is already routinely used in the dairy industry worldwide to analyze major milk components (e.g., fat, protein, lactose contents, and urea) for milk payment, herd management, quality control, or genetic evaluation programs (ICAR, 2017a). Additionally, MIR can be used to predict other phenotypes associated with milk composition in dairy cows with reasonable accuracy, such as fatty acids (Soyeurt et al., 2006), ketone bodies (Grelet et al., 2016), methane emissions (Vanlierde et al., 2018), or energy intake and feed efficiency (McParland and Berry, 2016). Because the establishment of pregnancy affects milk composition, through altering nutrient partitioning between physiological functions,

Received August 20, 2019.

Accepted December 10, 2019.

*Corresponding author: phuong.ho@agriculture.vic.gov.au

it might be hypothesized that MIR would be used to detect the pregnancy of a dairy cow. Indeed, several authors have indicated, for instance, increasing milk fat and protein contents of pregnant compared with nonpregnant cows, especially in the last months of gestation (Olori et al., 1997; Penasa et al., 2016; Lainé et al., 2017).

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

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

influence milk composition. The third novelty is the use of cow-independent validation because the commonly applied method of random cross-validation has recently been reported to produce overoptimistic results (Wang and Bovenhuis, 2019).

MATERIALS AND METHODS

Data

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

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

Data Preprocessing

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

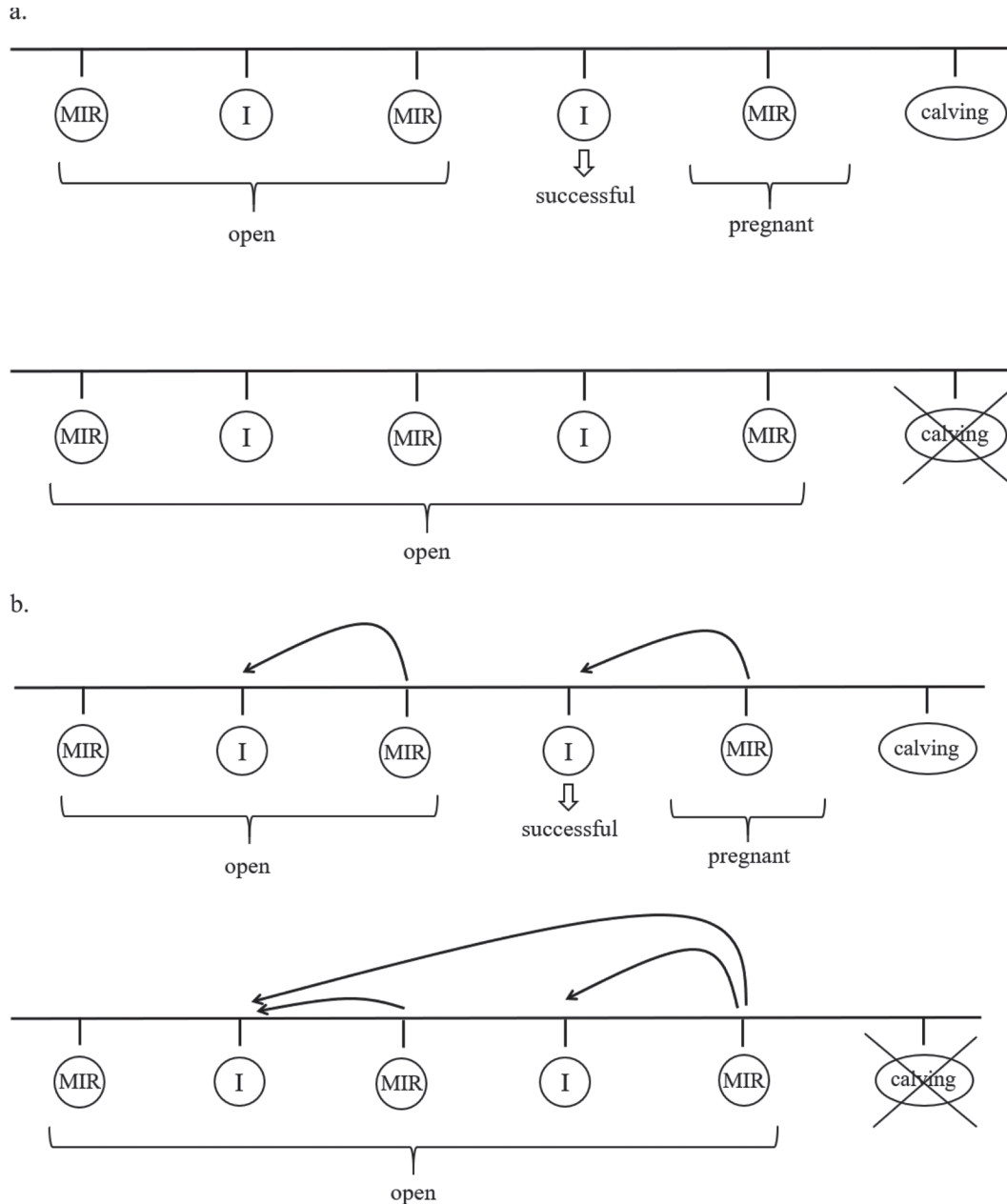


Figure 1. (a) Timeline diagram explaining how the pregnancy status (i.e., open or pregnant) was defined for milk mid-infrared spectrum (MIR) records. (b) Timeline diagram explaining the selection of spectra after insemination. The MIR records with open status can be associated with several inseminations.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

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

Modeling Strategies

(1) Single Spectra After Insemination. In this strategy, only spectral records after insemination were considered, which is similar to the study by Toledo-Alvarado et al. (2018). However, Toledo-Alvarado et al. (2018) did not consider cows with no calving records, in contrast to the present study. The selection of spectral records after insemination is described in Figure

1b. The MIR records with open status, as previously described, could be associated with several inseminations, whereas those with pregnant status were only associated with the successful (i.e., last) insemination.

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

(3) Stages After Insemination. We hypothesized that the pregnancy signal in milk might vary by gestation stage, with the signal being stronger toward the end of gestation. Therefore, grouping records based on the period after insemination and modeling each group separately may improve the accuracy of the models by reducing the spectral and the pregnancy signal variabilities. Consequently, only single spectra after insemination

were selected and divided into 7 classes based on the number of days after insemination (Table 1).

Model Development and Evaluation of Performance

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

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

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

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

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

appropriate due to the correlations that exist between wavenumbers.

The model performance was evaluated in 2 ways: random cow-independent 10-fold cross-validation and external validation. Random cow-independent 10-fold cross-validation means that 10% of the cows were randomly removed from the calibration set, and a model built with the remaining cows was used to classify the excluded records. The procedure was repeated 10 times to obtain predictions for all records. Hence, for each of the 10-folds, cows in the training set were different from those in the validation set. In addition to assessing the predictive performance of the model, cross-validation was also used to fine tune the model (i.e., determine the most appropriate number of components of the partial least squares discriminant analysis models). However, the maximum number of components was fixed to 20 to avoid overfitting. External validation was carried out using the cow-independent testing data consisting of cows that were not used to build up and tune the model.

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

For strategy 2, to have a fair comparison of model performance obtained using the spectral differences and single spectra after insemination, we ran the 2 models on the same data set. Finally, to explore whether the patterns of effects of pregnancy on MIR would change

as gestation progresses, the coefficients of the 7 models in the third strategy (i.e., stages after insemination) were extracted and analyzed. Accordingly, the correlations between the model's coefficients are presented as a heatmap. If the correlation between the coefficients of 2 models is strong, similar wavenumbers might have been used for the prediction of the pregnancy status.

RESULTS AND DISCUSSION

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

Data Description

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

Regarding the third strategy (i.e., our hypothesis that the pregnancy signal in milk might vary by gestation stage), the number of records decreased from the first to the last class (Table 1). The small number of spectral records available for pregnant cows in late gestation stages was caused by the dry-off period, which is usually at least 6 wk before calving (Dairy Australia, 2017). Records from cows that were still open later in

lactation and a long interval from insemination could be from cows that were not reinseminated because of fertility or health issues but remained in the herd until the end of the lactation. We were unable to verify this assumption with certainty because culling records were sparse.

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

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

Diagnosis of Pregnancy Status Using MIR with Different Approaches

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

ics (Collier et al., 2017). Consequently, finding a signal indicating pregnancy was challenging.

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

Unexpectedly, the prediction accuracy for calibration, cross-validation, and testing using spectral differ-

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

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

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

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

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

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

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

We have shown that using spectral data from various pregnancy stages was not enough to accurately diagnose the pregnancy status of cows. Several studies have reported a varying effect of pregnancy on milk

composition throughout gestation. For example, Olori et al. (1997), Roche (2003), and Penasa et al. (2016) reported that milk fat and protein contents, as well as lactose content to a smaller extent, increased as pregnancy advances. More recently, Lainé et al. (2017) reported that the effect of pregnancy on the whole MIR spectrum was variable according to the spectral region and the gestation stage, and tended to be stronger at the end of the gestation. In the previous studies on the prediction of the pregnancy status of dairy cows (Lainé et al., 2014; Toledo-Alvarado et al., 2018), the number of days after insemination was limited to 120 and 91, respectively, and the pregnancy status was studied without a consideration of different gestation periods. In this study, we had data for later stages after insemination and we hypothesized that separating data by stage after insemination might reduce spectral variability as well as pregnancy signal variability and thus improve prediction accuracy. This approach is comparable to the first one (i.e., using single spectra after insemination as independent variables in the model), but the data were grouped into 7 classes based on the days after insemination.

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

Interestingly, the prediction accuracy for class 7 (i.e., records taken 181 or more days after insemination) was 20% higher than that of classes 1 to 5. We also observed higher cross-validation and test AUC, sensitivity, and probabilities for class 6 (i.e., records between 151 and 180 d after insemination) compared with classes 1 to 5.

Globally, the median probabilities for correctly classified records were relatively close to the 0.5 threshold, showing that the model was unable to separate well the data, though these probabilities were slightly higher in the last 2 groups. These results imply that although MIR may not be sufficient to predict the pregnancy status of dairy cows in early and mid stages after insemination (from 1 to 150 d), promising results were obtained for records taken 151 d or more after insemination.

In late gestation, cows are subject to more significant endocrine changes, reduced feed intake, and greater nutritional demand of the growing calf compared with the early gestation period (Ingvarsen and Andersen, 2000; Penasa et al., 2016). Indeed, almost no increase in estrogen level exists during the first 3 mo of pregnancy, but it gradually rises from the third to the seventh month, before suddenly increasing after 7 mo of pregnancy (Parkhie et al., 1966; Penasa et al., 2016). The increase in estrogen levels in maternal blood during the last weeks of pregnancy affects nutrient partitioning, which has a deleterious effect on milk yield and, consequently, milk composition (Parkhie et al., 1966; Olori et al., 1997). Olori et al. (1997) also mentioned that the varying effect of pregnancy stages on milk composition suggests that the mechanism for the secretion of the solid and nonsolid components of milk might

be unequally affected by pregnancy. This may offer an explanation of why good accuracy was obtained for the models using records after 150 d after insemination but not in early gestation.

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

The results obtained for strategy 3 are consistent with Figure 2, which is a heatmap of correlations between coefficients of the 7 models associated with the 7 classes. Correlations were generally low, with a maximum of 0.41. However, an interesting finding was that the closer the groups, the higher the correlations (e.g., group 7 had high correlations with groups 5 and 6 but lower correlations with groups 1 to 4). These

Table 4. Strategy 3 (stages after insemination): results of partial least squares discriminant analysis for calibration, random cow-independent 10-fold cross-validation (mean, with SD in parentheses), and cow-independent test set for the 7 classes of records based on the number of days after insemination¹

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

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

results potentially imply that pregnancy was detected using different spectral wavelengths or with different weights at different stages after insemination, with an evolving pattern from the beginning toward the end of the pregnancy. This confirms the poor prediction accuracy of the global model for pregnancy detection in the present and the previous studies by Lainé et al. (2014) and Toledo-Alvarado et al. (2018).

The low frequency of milk recoding, which is monthly in most countries, could be considered as an obstacle in applying MIR to predict not only pregnancy status but also other traits, such as indicators of early lactation metabolic diseases (e.g., β -hydroxybutyrate, fatty acids). Only when scientists have been able to demonstrate the costs and benefits of having MIR collected

more frequently will farmers be willing to alter their milk-testing practice so that these MIR tools can be applied more effectively.

CONCLUSIONS

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

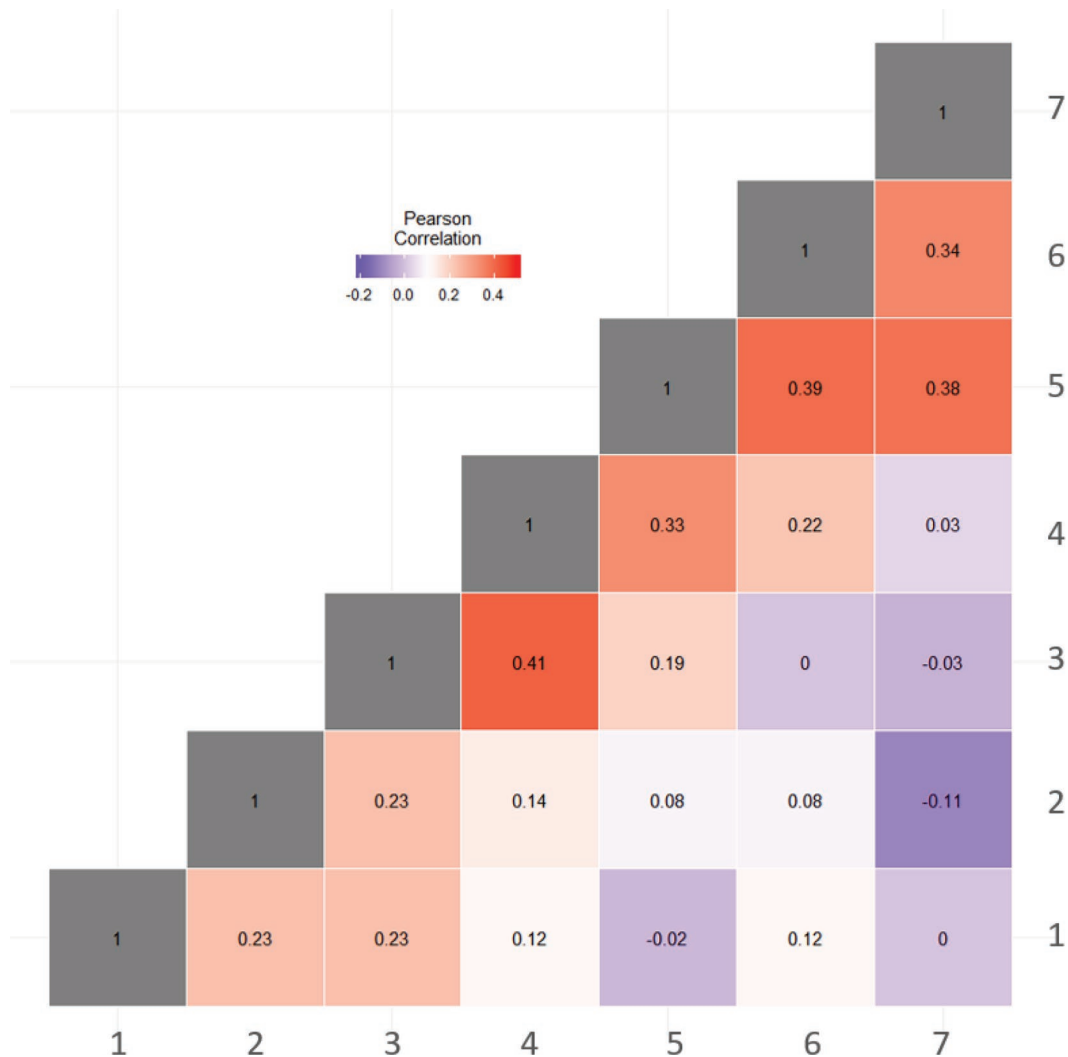


Figure 2. Heatmap of correlations between coefficients of the 7 models associated with the classes of records based on the number of days after insemination (class 1 = 1 to 30 d, class 2 = 31 to 60 d, class 3 = 61 to 90 d, class 4 = 91 to 120 d, class 5 = 121 to 150 d, class 6 = 151 to 180 d, class 7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

set and can be done a little earlier, the models could be used as a complementary tool to detect fetal abortion.

ACKNOWLEDGMENTS

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

REFERENCES

- Boll, S. F. 1979. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust.* 27:113–120. <https://doi.org/10.1109/TASSP.1979.1163209>.
- Bonfatti, V., G. Di Martino, and P. Carnier. 2011. Effectiveness of mid-infrared spectroscopy for the prediction of detailed protein composition and contents of protein genetic variants of individual milk of Simmental cows. *J. Dairy Sci.* 94:5776–5785. <https://doi.org/10.3168/jds.2011-4401>.
- Bronner, A., E. Morignat, V. Hénau, A. Madouasse, E. Gay, and D. Calavas. 2015. Devising an indicator to detect mid-term abortions in dairy cattle: A first step towards syndromic surveillance of abortive diseases. *PLoS One* 10:e0119012. <https://doi.org/10.1371/journal.pone.0119012>.
- Coelho, B., P. Hölscher, and F. Barends. 2011. Enhancement of double integration procedure through spectral subtraction. *Soil. Dyn. Earthquake Eng.* 31:716–722. <https://doi.org/10.1016/j.soildyn.2010.12.013>.
- Collier, R. J., Y. Xiao, and D. E. Bauman. 2017. Regulation of factors affecting milk yield. Pages 3–15 in *Nutrients in Dairy and Their Implications for Health and Disease*. 1st ed. R. R. Watson, V. R. Preedy, and R. J. Collier, ed. Acad. Press, Elsevier, Amsterdam, the Netherlands. <https://doi.org/10.1016/B978-0-12-809762-5.00001-2>.
- Dairy Australia. 2017. Decide dry cow strategy drying off dates, antibiotic. Dry cow and internal teat sealant treatment. Pages 1–9 in *Late Lactation, Technote*. Accessed Jan. 19, 2020. <https://www.dairyaustralia.com.au/-/media/dairyaustralia/documents/farm/animal-care/mastitis/research-reports-and-papers/new/technote-14-decide-dry-cow-strategy.ashx?la=en&hash=C9929BD8EA6FC6CBB5BB125A22F17A87B3D473C7>.
- De Maesschalck, R., D. Jouan-Rimbaud, and D. L. Massart. 2000. The Mahalanobis distance. *Chemom. Intell. Lab. Syst.* 50:1–18. [https://doi.org/10.1016/S0169-7439\(99\)00047-7](https://doi.org/10.1016/S0169-7439(99)00047-7).
- Fricke, P. M., A. Ricci, J. O. Giordano, and P. D. Carvalho. 2016. Methods for and implementation of pregnancy diagnosis in dairy cows. *Vet. Clin. North Am. Food Anim. Pract.* 32:165–180. <https://doi.org/10.1016/j.cvfa.2015.09.006>.
- Grelet, C., C. Bastin, M. Gelé, J.-B. Davière, M. Johan, A. Werner, R. Reding, J. A. Fernandez Pierna, F. G. Colinet, P. Dardenne, N. Gengler, H. Soyeurt, and F. Dehareng. 2016. Development of Fourier transform mid-infrared calibrations to predict acetone, β -hydroxybutyrate, and citrate contents in bovine milk through a European dairy network. *J. Dairy Sci.* 99:4816–4825. <https://doi.org/10.3168/jds.2015-10477>.
- Grelet, C., J. A. F. Pierna, P. Dardenne, H. Soyeurt, A. Vanlierde, F. Colinet, C. Bastin, N. Gengler, V. Baeten, and F. Dehareng. 2017. Standardization of milk mid-infrared spectrometers for the transfer and use of multiple models. *J. Dairy Sci.* 100:7910–7921. <https://doi.org/10.3168/jds.2017-12720>.
- Haile-Mariam, M., and J. E. Pryce. 2019. Genetic evaluation of gestation length and its use in managing calving patterns. *J. Dairy Sci.* 102:476–487. <https://doi.org/10.3168/jds.2018-14981>.
- Hawkins, D. M. 2004. The problem of overfitting. *J. Chem. Inf. Comput. Sci.* 44:1–12. <https://doi.org/10.1021/ci0342472>.
- Ingvarsen, K. L., and J. B. Andersen. 2000. Integration of metabolism and intake regulation: A review focusing on periparturient animals. *J. Dairy Sci.* 83:1573–1597. [https://doi.org/10.3168/jds.S0022-0302\(00\)75029-6](https://doi.org/10.3168/jds.S0022-0302(00)75029-6).
- ICAR. 2017a. Section 2—Guidelines for Dairy Cattle Milk Recording. Int. Comm. Anim. Record., Rome, Italy.
- ICAR. 2017b. Section 13—Guidelines for On-Line Milk Analysis. Int. Comm. Anim. Record., Rome, Italy.
- Kuhn, M. 2019. Package ‘caret’. R package version 6.0–84. R Foundation for Statistical Computing, Vienna, Austria.
- Lainé, A., C. Bastin, C. Grelet, H. Hammami, F. G. Colinet, L. M. Dale, A. Gillon, J. Vandenplas, F. Dehareng, and N. Gengler. 2017. Assessing the effect of pregnancy stage on milk composition of dairy cows using mid-infrared spectra. *J. Dairy Sci.* 100:2863–2876. <https://doi.org/10.3168/jds.2016-11736>.
- Lainé, A., H. Bel Mabrouk, L. Dale, C. Bastin, and N. Gengler. 2014. How to use mid-infrared spectral information from milk recording system to detect the pregnancy status of dairy cows. *Comm. Agric. Appl. Biol. Sci.* 79:33–38.
- Mayeres, P., J. Stoll, J. Bormann, R. Reents, and N. Gengler. 2004. Prediction of daily milk, fat, and protein production by a random regression test-day model. *J. Dairy Sci.* 87:1925–1933. [https://doi.org/10.3168/jds.S0022-0302\(04\)73351-2](https://doi.org/10.3168/jds.S0022-0302(04)73351-2).
- McParland, S., and D. P. Berry. 2016. The potential of Fourier transform infrared spectroscopy of milk samples to predict energy intake and efficiency in dairy cows. *J. Dairy Sci.* 99:4056–4070. <https://doi.org/10.3168/jds.2015-10051>.
- Olori, V. E., S. Brotherstone, W. G. Hill, and B. J. McGuirk. 1997. Effect of gestation stage on milk yield and composition in Holstein Friesian dairy cattle. *Livest. Prod. Sci.* 52:167–176. [https://doi.org/10.1016/S0301-6226\(97\)00126-7](https://doi.org/10.1016/S0301-6226(97)00126-7).
- Parkhie, M. R., L. O. Gilmore, and N. S. Fechner. 1966. Effect of successive lactations, gestation, and season of calving on constituents of cows’ milk. *J. Dairy Sci.* 49:1410–1415. [https://doi.org/10.3168/jds.S0022-0302\(66\)88103-1](https://doi.org/10.3168/jds.S0022-0302(66)88103-1).
- Penasa, M., M. De Marchi, and M. Cassandro. 2016. Short communication: Effects of pregnancy on milk yield, composition traits, and coagulation properties of Holstein cows. *J. Dairy Sci.* 99:4864–4869. <https://doi.org/10.3168/jds.2015-10168>.
- Roche, J. R. 2003. Effect of pregnancy on milk production and body-weight from identical twin study. *J. Dairy Sci.* 86:777–783. [https://doi.org/10.3168/jds.S0022-0302\(03\)73659-5](https://doi.org/10.3168/jds.S0022-0302(03)73659-5).
- Rozenstein, O., T. Paz-Kagan, C. Salbach, and A. Karnieli. 2015. Comparing the effect of preprocessing transformations on methods of land-use classification derived from spectral soil measurements.

- IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8:2393–2404. <https://doi.org/10.1109/JSTARS.2014.2371920>.
- Shetty, N., G. Difford, J. Lassen, P. Løvendahl, and A. J. Buitenhuis. 2017. Predicting methane emissions of lactating Danish Holstein cows using Fourier transform mid-infrared spectroscopy of milk. *J. Dairy Sci.* 100:9052–9060. <https://doi.org/10.3168/jds.2017-13014>.
- Šimundić, A.-M. 2009. Measures of diagnostic accuracy: Basic definitions. *EJIFCC* 19:203–211.
- Soyeurt, H., P. Dardenne, F. Dehareng, G. Lognay, D. Veselko, M. Marlier, C. Bertozzi, P. Mayeres, and N. Gengler. 2006. Estimating fatty acid content in cow milk using mid-infrared spectrometry. *J. Dairy Sci.* 89:3690–3695. [https://doi.org/10.3168/jds.S0022-0302\(06\)72409-2](https://doi.org/10.3168/jds.S0022-0302(06)72409-2).
- Soyeurt, H., F. Dehareng, N. Gengler, S. McParland, E. Wall, D. P. Berry, M. Coffey, and P. Dardenne. 2011. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. *J. Dairy Sci.* 94:1657–1667. <https://doi.org/10.3168/jds.2010-3408>.
- Toledo-Alvarado, H., A. I. Vazquez, G. de los Campos, R. J. Tempelman, G. Bittante, and A. Cecchinato. 2018. Diagnosing pregnancy status using infrared spectra and milk composition in dairy cows. *J. Dairy Sci.* 101:2496–2505. <https://doi.org/10.3168/jds.2017-13647>.
- Vanlierde, A., H. Soyeurt, N. Gengler, F. G. Colinet, E. Froidmont, M. Kreuzer, F. Grandl, M. Bell, P. Lund, D. W. Olijhoek, M. Eugène, C. Martin, B. Kuhla, and F. Dehareng. 2018. Short communication: Development of an equation for estimating methane emissions of dairy cows from milk Fourier transform mid-infrared spectra by using reference data obtained exclusively from respiration chambers. *J. Dairy Sci.* 101:7618–7624. <https://doi.org/10.3168/jds.2018-14472>.
- Velek, K., J. Lawrence, R. Linscott, E. Martel, H. Pearce, R. Hildreth, and C. Egli. 2012. IDEXX Milk Pregnancy Test—A New Tool for Pregnancy Diagnosis in Cattle. IDEXX Laboratories Inc., Westbrook, ME.
- Wang, Q., and H. Bovenhuis. 2019. Validation strategy can result in an overoptimistic view of the ability of milk infrared spectra to predict methane emission of dairy cattle. *J. Dairy Sci.* 102:6288–6295. <https://doi.org/10.3168/jds.2018-15684>.

ORCIDS

- P. Delhez  <https://orcid.org/0000-0002-2699-1135>
P. N. Ho  <https://orcid.org/0000-0001-9481-9612>
N. Gengler  <https://orcid.org/0000-0002-5981-5509>
H. Soyeurt  <https://orcid.org/0000-0001-9883-9047>
J. E. Pryce  <https://orcid.org/0000-0002-1397-1282>