HOW TO USE MID-INFRARED SPECTRAL INFORMATION FROM MILK RECORDING SYSTEM TO DETECT THE PREGNANCY STATUS OF DAIRY COWS

A. LAINÉ*, H. BEL MABROUK*, L-M. DALE*, C. BASTIN* AND N. GENGLER*
* University of Liege, Gembloux Agro-Bio Tech, Animal Science Unit, Gembloux, Belgium

INTRODUCTION

Ensuring that a cow is effectively pregnant after an insemination is a key element in the management of dairy farms. Currently, the most common pregnancy diagnosis methods are echography and transrectal palpation which have to be done by a veterinarian or another qualified person between 30 and 90 days after the insemination date. Those methods present a certain cost, might be risky and have a given efficiency (Purohit, 2010).

The mid-infrared spectroscopic (MIR) analysis of milk is the method of choice used internationally for the quantification of milk composition, i.e., for major milk components such as fat, protein and lactose but recently also for fine milk components such as fatty acids. Mid-infrared spectroscopy is used both for samples taken as a part of large-scale milk recording and for milk payment systems. It is a rapid and non-expensive method. A milk sample analysed by MIR provides a spectrum which is a unique fingerprint of the whole milk composition. A MIR milk spectrum represents the absorption of infrared light through the milk sample at wavelengths between 900 cm-1 and 5,000 cm-1 (Coates, 2000).

The aim of this study was to investigate the potential of MIR analysis of milk to identify changes in the pregnancy status of dairy cows. The long-term aim of the research was to develop a strategy transferable in routine in the context of the milk recording system that can identified open cows from pregnant cows after an insemination and within the first 50 days after this insemination.

MATERIAL AND METHODS

Overall approach

The general approach of this study was to compare an observed spectrum from a given cow at one given day in milk to the expected spectrum if this cow would have been open at this given day in milk. To obtain the expected spectrum, relevant effects were estimated based on spectral data from open cows only. This strategy was followed by Sloth et al. (2003) in order to assess the udder health status by adjusting some milk components on a subset of healthy observations. Expected spectra were calculated and then removed from the observed spectra to obtain residual spectra. These residual spectra were then used to predict if the spectrum belonged to a pregnant cow or not.

Data were obtained from the milk recording system of the Walloon Region of Belgium (managed by the Walloon Breeding Association, Ciney, Belgium). Test-day observations were selected from 1st November 2011 until 15 October 2013. For each test-day, available data were the complete production and animal information obtained through the milk recording system (e.g., dates of calving, dates of inseminations, number of lactations, production traits such as milk yields or fat percentage, days in milk) and also the spectral data obtained from the MIR-analysis of milk samples.

Dataset editing

Based on historical information of known reproductive events coupled with test-day and spectral information, the data set was constructed. The pregnancy status was assigned to each available test-day spectrum using the following algorithm. For each cow, according to the theoretical gestation length for the breed (e.g., 282 days for Holstein cows), the theoretical date of the successful insemination was calculated as the re-calving date minus the gestation length. By comparison with observed dates of insemination recorded in the database, the real date of successful insemination was identified as the nearest observed date within an interval of 20 days around the theoretical date. All test days that occur before the real date of successful insemination were coded as "open" and all test days that occur after this date were coded as "pregnant" until the end of the lactation. To ensure the quality of the pregnancy status code, if doubts existed or if an irregularity in the recorded dates was observed, test days were be coded as "unknown". Only spectra with a known pregnancy status ("open" or "pregnant") were kept for the study.

Dataset pre-processing

First derivative was calculated on the raw spectra as the difference between a spectral value at data point X and the spectral value at data point X+5. The aim of this pre-treatment was to set all spectra at a common baseline (McParland et al., 2011). In order to avoid introducing noises, only informative area from the spectra were kept. These ranges of informative wavelengths are the same than those used routinely to predict fatty acid contents in cow milk (Soyeurt et al., 2006).

Outliers were deleted following two steps. First, milk yield, fat content and protein content were required to be within the range of ICAR norms (ICAR, 2012). Second, spectra with a value of standardized Mahalanobis distance greater than 3 compared to the mean of our dataset were considered as outliers and removed from the dataset (Shenk and Westerhaus, 1990). Also, only observations from 5 to 365 days in milk (DIM) were kept.

The final edited dataset included a total of 411,406 spectra (114,338 spectra from pregnant cows and 297,018 spectra from open cows) from 68,998 cows in 1,045 herds.

Modelling the expected open spectra

A mixed model was applied to the subset of spectra recorded on open cows only. Lactations with at least one observation considered as open before all observations considered as pregnant in the lactation were kept for modelling (256,238 spectra). Expected open spectra were then estimated for all the spectra of the database (pregnant or open) as follow:

$$\hat{y}_{ijkl} = parity_i + breed_j + monthTD_k + (cDIM * DIM) + (cDIM2 * DIM^2) + anlact_l + (cDIManlact_l * DIM) + (cDIM2anlact_l * DIM^2)$$

where

$\hat{y}_{ijkl}$ = expected value of spectra for the l^{th} lactation at the i^{th} parity for the j^{th} breed and at the k^{th} month of test day

$parity_i$ = fixed effect of parity i

$breed_j$ = fixed effect of breed j

$monthTD_k$ = fixed effect for the k^{th} month of test day

$cDIM$ = regression coefficient for the days in milk (DIM)

$cDIM2$ = regression coefficient for the squared DIM

$anlact_l$ = random effect of cow within lactation l

$cDIManlact_l$ = random regression coefficient for the DIM across the l^{th} lactation

$cDIM2anlact_l$ = random regression coefficient for the squared DIM across the l^{th} lactation.

The Table 1 show the distribution of observations by level for each factor used.

Table 1. Distribution of observations by level (n=256,238) to the systematic factors used in the model.

Systematic factors	Level	N	%
Parity	1	59,457	23.2%
	2+	196,781	76.8%
Breed	HOL	238,26	93.0%
	BBL	8,141	3.2%
	MON	3,847	1.5%
	Other	5,991	2.3%
Milking moment	AM/PM	225,323	88.0%
	AM	15,359	6.0%
	PM	15,356	6.0%
Month of test-day	January	23,89	9.3%
	February	21,321	8.3%
	March	23,296	9.1%
	April	21,601	8.4%
	May	22,336	8.7%
	June	20,813	8.1%
	July	9,388	3.7%
	August	21,526	8.4%
	September	25,052	9.8%
	October	17,679	6.9%
	November	25,022	9.8%
	December	24,314	9.5%
Animal x lactations	92,133*	2.8**	

HOL: Holstein and Red-Holstein; BBL: dual purpose Belgian Blue; MON: Montbeliard; *: number of level; **: average number of data by level

Residual spectra

Expected open spectra, obtained above, were removed from observed spectra in order to obtain residual spectra. Residual spectra are the results of factors that were not taking account for the calculation of the expected spectra (errors, pregnancy status, and unaccounted factors).

Predictive discriminant analysis

In order to discriminate residual spectra of open cows from residual spectra of pregnant cows, a predictive discriminant analysis was performed on a training dataset that included residual spectra recorded after 20 days in

milk and recorded from 20 days to 120 days after an insemination. Moreover, the dataset was equilibrated to obtain the same proportion of residual spectra from pregnant and open cows. For the construction of the discriminant function the final dataset included a total of 2,491 residual spectra. The discriminant function was constructed using the PROC DISCRIM procedure from SAS (SAS Institute, 2009).

The discriminant function obtained on the training dataset was then applied on a validation dataset obtained from observations of lactations which were not present in the calibration dataset. This validation dataset included only observations occurring from 20 to 120 days after an insemination and contained 14,883 residual spectra. The error rates of classification was calculated as the number of data misclassified divided by the number of observations in the validation dataset.

Results were expressed in terms of specificity and sensibility. Specificity is defined as the ability of the equation to predict correctly the non-event (open cows) among all observations which are not pregnant. Sensibility is defined as the ability of the equation to predict correctly the event (pregnant cows) among all observations belonging to pregnant cows.

In order to show the interest of using residual spectra, and not the raw spectra, the discriminant function was constructed for the same observations but based on the raw spectra (i.e. the spectra not adjusted for the systematic factors) and then applied on the raw spectra of the validation dataset.

RESULTS AND DISCUSSION

When the predictive discriminant equation obtained on residual spectra was applied on the whole validation dataset, the classification error rate was 0.7%. Specificity was 86.2% and sensibility was 99.7%. Meaning that 86.2% of observations belonging to open cows were correctly classified by the equation as open cow and 99.7% of observations belonging to pregnant cows were correctly classified by the equation as pregnant cow.

When the discriminant function was constructed and applied on the raw spectra, the error rate of classification was 55.5%. In the case of a two group classification, the base probability to find the correct response is already 50%. Using the raw spectra was therefore not useful to make this prediction. However, the use of the residual spectra allowed distinguishing much better spectral observation from an open or a pregnant cow.

Table 2 shows the result of classification in terms of error rates by classes of 10 days after an insemination event. Table 3 shows the same results expressed as specificity and sensibility.

Results 21 days after an insemination are over 97% of correct classification which is a very good result regarding to classical pregnancy detection (e.g. echography). For instance, the echography can be made only 30 days after an insemination and the efficiency is less than 95% of correct detection (Purohit, 2010). A trend can be observed in both tables, classification is getting better when the number of days after an insemination increases. This is an expected result since more the gestation is advanced more the embryo influences the cow and therefore her milk composition. Because of the editing on the dataset, there were no observations 48 days after insemination belonging to open cows. Therefore, this study could not yet validate the method beyond this limit. One can however expect that given the results, the discriminant function used allows reasonable classifications for several weeks after 48 days post insemination.

Table 2. Number of data and error rates of classification on residual spectra by classes of 10 days after an insemination date (IDATE).

Days after IDATE	N Open	N Pregnant	Error rates
21 - 30	202 (14.7%)	1,168 (85.3%)	2.3%
31 - 40	122 (9.5%)	1,162 (90.5%)	2.0%
41 - 50	38 (3.0%)	1,222 (97.0%)	0.8%

Table 3. Results of classification on residual spectra by classes of 10 days after an insemination date (IDATE) expressed in terms of specificity and sensibility.

Days after IDATE	Specificity*	Sensibility**
21 - 30	89.1%	99.2%
31 - 40	83.6%	99.5%
41 - 50	84.2%	99.7%

*Specificity: Proportion of data belonging to open cows properly classified as open, it concerns data from cows that have been inseminated but it was not successful; **Sensibility: Proportion of data belonging to pregnant cows properly classified as pregnant, it concerns data from cows after the successful insemination.

CONCLUSION

To conclude, results showed that the changes in the MIR milk spectrum can provide indications of the change in the pregnancy status of the cow. This can be derived in a milk testing setting. Therefore, advisory tools based on MIR analysis of milk could be developed to help dairy farmers in the management of their herd fertility. Such tools could be used to give an early warning to the farmer when a change in the reproductive status of an animal having been inseminated appears in the milk that is analysed in the milk testing lab. Different uses of such tools could be made, the most evident being the detection of inseminated animals considered pregnant as suspicious not to be. In such case, the breeder could be advised to make a check on the problematic animal.

Further studies will investigate the prediction of the pregnancy status of the cows beyond 50 days after insemination. Indeed, the current calibration and validation datasets did not included open cows beyond 48 days after an insemination event.

The proposed approach can be adapted inside various milk recording systems as done currently through the OptiMIR project. It can thus provide added value to the information provided to dairy farmers by these milk recording organisations.

Another interesting point is that the basic idea to compare observed spectra from expected spectra has high potential in the detection of metabolic and health (e.g., udder health) disorders of dairy cows based on the MIR analysis of their milk.

ACKNOWLEDGEMENTS

Support for the Interreg IVb project OptiMIR (www.optimir.eu) by the European Commission (ERDF) and the Service Public of Wallonia (SPW) is acknowledged. The authors thank Service Public of Wallonia (SPW-DG03)

also for the support for the NovaUdderHealth project (D31-1273). The Walloon Breeding Association (AWE), the Milk Committee of Battice (Belgium) and the Agricultural Walloon Research Center (CRA-W) are acknowledged as partner of the OptiMIR project.

REFERENCES

Coates J. (2000). Interpretation of infrared spectra, a practical approach. Pages 1085-1087 in Encyclopedia of Analytical Chemistry. R. A. Mayeres, ed. John Wiley & Sons, New York, NY.

ICAR Guidelines (2012). International agreement of recording practices. Guidelines approved by the general assembly held in Cork, Ireland on June 2012.

McParland S., Banos G., Wall E., Coffey M.P., Soyeurt H., Veerkamp R.F., and Berry D.P. (2011). The use of mid-infrared spectrometry to predict body energy status of holstein cows. J. Dairy Sci. 94: 3651-3661.

Purohit G. (2010). Methods of Pregnancy Diagnosis in Domestic Animals: The Current Status. WebmedCentral REPRODUCTION 2010;1(12):WMC001305.

SAS Institute (2009). SAS/STAT User's guide: Version 9.2 SAS institute Inc., Cary, NC.

Shenk J.S., and Westerhaux M.O. (1990). Population structuring of near infrared spectra and modified partial least squares regression. Crop Sci. 31: 1548-1555.

Sloth K.H.M.N., Friggens N.C., Lovendhal P., Andersen P.H., Jensen J., and Ingvartsen K.L. (2003). Potential for improving description of bovine udder health status by combined analysis of milk parameters. J. Dairy Sci. 86: 1221-1232.

Soyeurt H., Dardenne P., Dehareng F., Lognay G., Veselko D., Marlier M., Bertozzi C., Mayeres P., and Gengler N. (2006). Estimating fatty acid content in cow milk using mid-infrared spectrometry. J. Dairy Sci. 89: 3690-3695.