145. Classifying the likelihood of conception in dairy cow with milk mid-infrared spectra before the first insemination

W. Lou[1,2], R. Shi[1,2], B. Ducro[2], A. van der Linden[2], H.A. Mulder[2], S.J. Oosting[2], L. Liu[3] and Y. Wang[1*]

[1]China Agricultural University, Beijing, 100193, China; [2]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [3]Beijing Dairy Cattle Center, Beijing, 100192, China; wangyachun@cau.edu.cn

Abstract

Accurate and early identification of the likelihood of conception (LC) in cows is imperative for a profitable dairy farm. This study aims to use the milk mid-infrared (MIR) spectra in different intervals before the first insemination and partial least squares discriminant analysis (PLS-DA) to predict LC. The results show that the MIR data within 30 to 50 d after calving and close to insemination had a better prediction in LC (accuracy = 73.9 and 72.3%) than 0 to 30 d. And specificity (74.4 to 84.9%) was higher than sensitivity (67.8 to 73.1%). Once the expected date of insemination is given, the model can predict LC before the actual insemination, and intervene in advance for cows predicted poor LC such as delayed insemination and treatment can be initiated. The predicted LC also provide a novel and convenient way to accumulate extra reproductive phenotypes for genetic evaluation.

Introduction

As one of the most important factors in reproduction performance of dairy cow, the likelihood of conception (LC) has impact on the management and subsequent milk production of the herd because cows with more insemination or infertility are often culled and it does not fit the farms calving pattern or is not profitable, e.g. due to a prolonged dry period of the cow or lower milk revenues compared to a replacement heifer. Prior knowledge of the LC to a given insemination would enable farmers to optimize breeding decisions. The producer may decide not to use the cow where a poor LC is predicted but instead to wait for a subsequent estrus event or treatment. Sexed semen or expensive semen could be used where a greater LC is predicted. Thus, early diagnosis of LC in dairy cow needs to be investigated for defining more effective reproductive strategies. Dramatic changes in milk composition (e.g. sexual hormone) before insemination underlie the prediction of LC for the cows (Toledo-Alvarado *et al.*, 2018). And milk compositions are routinely quantified by mid-infrared (MIR) spectroscopy in many milk recording organizations (De Marchi *et al.*, 2014). In recent years, a few studies have attempted to calibrate milk spectra to predict LC in dairy cows before an insemination, reporting AUC of prediction results were 0.487 to 0.675 (Hempstalk *et al.*, 2015) and 0.66 to 0.75 (Ho *et al.*, 2019). The MIR predicted LC has greater heritability than true LC phenotypes (van den Berg *et al.*, 2021). These results highlighted the potential of milk MIR spectra as a predictor of LC, and MIR-predicted reproductive indicators can be used as a new trait for traditional reproductive phenotypes that are used in genetic evaluation. Nevertheless, there are no studies that determine which optimal time of milk testing after calving and before insemination to obtain the highest prediction accuracy. Therefore, this study aimed to: (1) using MIR spectra to classifying good and poor LC of the cow at the first insemination after calving and find the optimal test period; (2) identify the most relevant wavenumbers from MIR spectra that contribute to prediction accuracy of LC.

Materials & methods

Data. Original records of MIR spectra and reproduction from Chinese Holstein cows were obtained from December 2016 to March 2019 from 12 dairy farms located in Beijing, China. The MIR spectra were tested monthly using the machine FTS (Bentley, Chaska, USA) at the Beijing Dairy Cattle Centre. Spectra located in the infrared range from 649.03 to 3,999.59 cm^{-1} (i.e. containing 899 wavenumbers) were included. Only

R.F. Veerkamp and Y. de Haas (eds)

DOI: 10.3920/978-90-8686-940-4_145, © W. Lou *et al.* 2022

cows within the range of 1 to 375 days in milk and up to 4 parities were selected. To ensure the cow had normal fertility performance, the records of gestation length ranging from 268 to 296 d, interval from calving to first insemination ranging from 20 to 230 d were included in the analysis. In this study, three-class of LC at first insemination were grouped: good (cow that conceived at first insemination), average (cows that conceived at 2 or more inseminations), and poor (cows with no conception event recorded and that had only 1 insemination). From an economic and breeding perspective, it would be very valuable to be able to predict cows with poor LC, thus the cows with good and poor LC were considered in this study.

Methods. Savitzky-Golay second-order derivative algorithm was used to eliminate the chaos and perturbations from water and others in milk (Grelet *et al.*, 2020). Based on previous studies and known water related regions, 538 spectral points (928 to 1,596 cm^{-1}, 1,693 to 3,025 cm^{-1}) out of the 899 wavenumbers were kept for further analyses. A further detection of outliers was performed using the standardized Mahalanobis distance combined with principal component analysis (Grelet *et al.*, 2020). Finally, 95% of data was kept according to the standardized Mahalanobis distance. Here, the dates of calving and first insemination were set on day 0 and the last day. MIR data in different intervals after calving include 0 to 30 d (early stage), 30 to 50 d (middle stage), 50 d to last day (late stage), 0 to 50 d, and 0 to last day were considered. The two time points 30 d and 50 d were chosen to ensure data balance. Based on previous results, MIR spectra, days in milk, parity, somatic cell count, milk yield, and interval from calving to first insemination were selected as the independent variables in the model.

In this study, PLS-DA was used to develop the prediction model. When one herd was used for external validation, then the remaining data was used for training model and internal validation (random, 10-fold cross-validation). Cross-validation was used to evaluate model performance and enable comparison between the different models. The indicators accuracy, area under the curve (AUC), specificity, and sensitivity were used to define which model performed best.

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \quad (1)$$

$$\text{Sensitivity} = TP/(TP+FN) \quad (2)$$

$$\text{Specificity} = TN/(TN+FP) \quad (3)$$

Where *TP*, *TN*, *FP*, and *FN* represent total numbers of true positives, true negatives, false positives, and false negatives as predicted by the model, respectively. Here, positive is the cow with good LC and negative is the cow with poor LC.

Results

The prediction results of external validation at different intervals before first insemination are shown in Table 1. The accuracies and AUCs ranged from 70.2 to 77.3% and 0.698 to 0.775, respectively. Sensitivity and specificity respectively represent the probability of correct prediction of individuals with good LC and poor LC in the real data. Specificity (74.4 to 84.9%) was higher than sensitivity (67.8 to 73.1%). In addition, the MIR data within 30 to 50 d after calving had best prediction of poor LC (accuracy = 73.9%) cows compared to that within 0 to 30 d and 50 to last d (Table 1).

To find out the most relevant wavenumbers that contribute to prediction in LC of the cows in this study, the regression coefficient in PLS-DA after normalization was considered for the different intervals (0 to 30 d, 30 to 50 d, 50 to last d). The wavenumbers that were the most relevant to predict the LC were found in five small regions, names from 936 to 1,189, 1,395 to 1,596, 1,693 to 1,790, 2,346 to 2,364, and 2,857 to

Table 1. The prediction results of herd-by-herd external validation based on mid-infrared data at different intervals before first insemination.

Interval	Records	No. cow	Accuracy[1]	AUC	Sensitivity[2]	Specificity[3]
0[4]-last[5] d	2,766	2,140	77.3%	0.775	73.1%	84.9%
0-50 d	1,886	1,422	74.3%	0.746	70.8%	78.5%
0-30 d	967	724	70.2%	0.698	67.8%	74.4%
30-50 d	956	727	73.9%	0.741	71.2%	76.9%
50-last d	955	725	72.3%	0.731	69.5%	75.3%

[1] Accuracy = (TP+TN)/(TP+TN+FP+FN).

[2] Sensitivity = TP/(TP+FN).

[3] Specificity = TN/(TN+FP). TP, TN, FP, and FN represent total numbers of true positives, true negatives, false positives, and false negatives as predicted by the model, respectively.

[4] The 0 day = the date of calving. [5]the last day = the date of first insemination. See main text for other abbreviations.

3,017 cm^{-1} (Figure 1). In addition, the regression coefficients after normalization of the three intervals had moderate to low Spearman correlations, ranged from 0.604 (0 to 30 d and 30 to 50 d) to 0.284 (0 to 30 d and 50 to last d).

Discussion

The innovation of this study is to defining the optimal test period to better classify good and poor LC based on MIR spectra. In addition, the MIR region associated with this reproductive trait was identified. The ability to accurately predict the conception outcome for a future mating would be of considerable benefit for producers in deciding what mating plan to implement for a given cow. The results of this study can be used in the development of a novel tool for monitoring the fertility and health in dairy cows before insemination. And accurate prediction implemented in current study also provide a new trait to accumulate reproductive phenotypes for genetic evaluation, especially before routine collection of reproductive information is established.

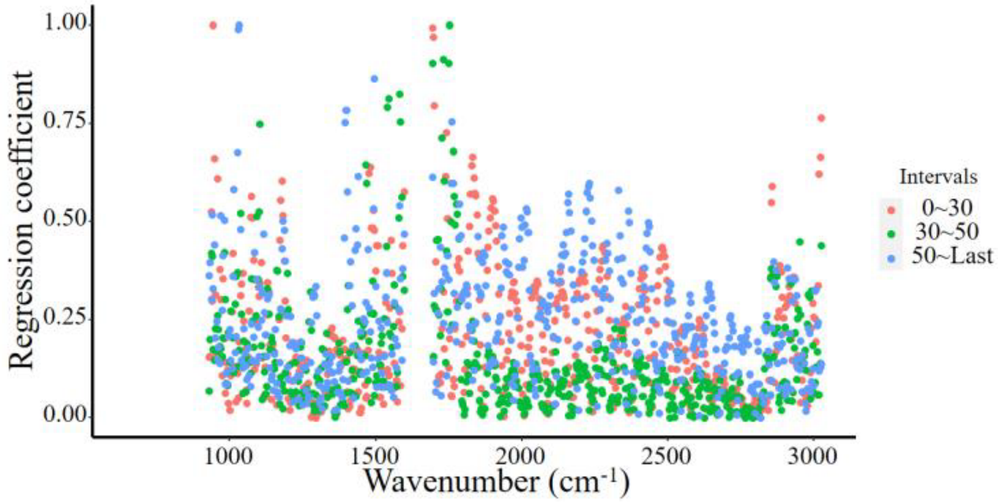

Figure 1. The regression coefficient after normalization for different wavenumbers in intervals of 0 to 30 (early stage), 30 to 50 (middle stage), and 50 to last (late stage) days.

Compared to the published studies, the accuracy (70.2 to 77.3%) and AUC (0.698 to 0.775) of external validation in this research are within these reported results (AUC: 0.487 to 0.750) (Hempstalk *et al.*, 2015; Ho *et al.*, 2019). Such differences among prediction is mainly due to differences in MIR data selection, spectral processing, and regression method employed. Specificity (74.4 to 84.9%) was higher than sensitivity (67.8 to 73.1%) in this study, which means the model has stronger ability to identify a cow with poor LC than a cow with good LC and this tendency is more important for reproductive management. The MIR data within 30 to 50 d after calving and close to insemination had the better prediction in LC (Accuracy = 73.9 and 72.3%) than 0 to 30 d. This indicated that collecting MIR spectra during these periods may obtain optimal prediction accuracy.

The most important wavenumbers of MIR may help in understanding why MIR spectra before insemination can predict LC. The regions from 936 to 1,189, 1,395 to 1,596, 1,693 to 1,790, 2,346 to 2,364, and 2,857 to 3,017 cm^{-1} are related to the chemical bonds including C-H, C-C, C=O, C-O-C, nitrogen-oxy compounds, aldehydes, and esters (Duffy, 1972). Many sex hormones (e.g. progesterone and follicle-stimulating hormone) contain the chemical bonds like esters, aldehyde, and nitrogen-oxy compounds. A previous study indicated that sex hormone exist in milk and that the level changes significantly before insemination (Toledo-Alvarado *et al.*, 2018), which means that MIR data before insemination contains information on sex hormone changes, especially within 30-50 d after calving and close to insemination. In addition, some studies found high concentration of blood ketone body before insemination will increase the risk of non-pregnancy (Walsh *et al.*, 2007), and MIR spectroscopy is a mature tool to qualify the concentration of ketone body in blood and milk (Benedet *et al.*, 2019). This indicated that the five spectral regions maybe reflect energy imbalance information, because the ketone body contain the chemical bonds like C-H, C-C, and C=O. The variation in other milk compositions could also be related to fertility and health of dairy cows. Given the most important wavenumbers in MIR data to predict LC, prediction based on MIR data before inseminations seems to be (partly) based on reflecting abnormal levels in sex hormones and energy imbalance. Meanwhile, the intervals of 30-50 d after calving and close to insemination might contain more changes in physiological state than the other periods, but it needs to be interpreted with caution.

Acknowledgements

The China Scholarship Council (No.201913043) is acknowledged for funding the PhD projects of Wenqi Lou and Rui Shi. This study is supported by China Agriculture Research System of MOF and MARA.

References

Benedet A., Manuelian C.L., Zidi A., Penasa M., and De Marchi M. (2019). Animal 13(8):1676-1689. https://doi.org/10.1017/S175173111900034X

De Marchi M., Toffanin V., Cassandro M., and Penasa M. (2014) J. Dairy Sci 97(3):1171-1186. https://doi.org/10.3168/jds.2013-6799

Duffy N.V. (1972) J. Chem Educ 49(9):30-45. https://doi.org/10.1021/ed049p652.1

Grelet C., Dardenne P., Soyeurt H., Fernandez J.A., Vanlierde A., and *et al.* (2020) Methods 186(3):97-111. https://doi.org/10.1016/j.ymeth.2020.07.012

Hempstalk K., Mcparland S., and Berry D.P. (2015) J. Dairy Sci 98(8):5262-5273. https://doi.org/10.3168/jds.2014-8984

Ho P.N., Bonfatti V., Luke T.D.W., and Pryce J.E. (2019) J Dairy Sci 102(11):10460-10470. https://doi.org/10.3168/jds.2019-16412

Toledo-Alvarado H., Vazquez A.I., Gustavo De L.C., Robert J.T., and Gianfranco G. (2018) J. Dairy Sci 101(10):9135-9153. https://doi.org/10.3168/jds.2018-14480

Van den Berg I., Ho P.N., Haile-Mariam M., and Pryce J.E. (2021) J. Dairy Sci 2(6):361-365. https://doi.org/10.3168/jdsc.2021-0141

Walsh R.B., Walton J.S., Kelton D.F., LeBlanc S.J., Leslie K.E, and *et al.* (2007) J. Dairy Sci 90(6):2788-2796. https://doi.org/10.3168/jds.2006-560