The use of milk Fourier-transform mid-infrared spectroscopy to diagnose pregnancy and determine spectral regional associations with pregnancy in US dairy cows

Piush Khanal and Robert J. Tempelman*

Department of Animal Science, Michigan State University, East Lansing 48824-1225

ABSTRACT

Accurate early diagnosis of pregnancy is important for timely reproductive management of dairy farms. Fourier-transform mid-infrared (FT-MIR) milk spectral data are routinely used for determining milk components such as fat and protein, whereas milk composition is known to change with advancing stages of pregnancy. The objectives of this study were to compare partial least squares discriminant analysis (PLS-DA) and a Bayesian variable selection regression model (BayesC) for the diagnosis of pregnancy status (PS) from milk FT-MIR data and to infer any spectral regions that might be highly associated with PS at various stages of pregnancy. Conception dates on confirmed pregnant cows were obtained from Holstein cows within 123 herds in Michigan, Ohio, and Indiana during 2018 and 2019. Milk samples from these pregnant cows at 7 different stages of pregnancy were case-control matched to open contemporary herd mates to be within the same stage (± 10 d for days in milk) of lactation for the same milk sample test date. The FT-MIR data were obtained for all of these milk samples. Ten-fold herd-independent cross-validation was used to compare PLS-DA versus BayesC using the area under the receiver operating characteristic curve (AUC). The BayesC model demonstrated higher mean AUC compared with PLS-DA at all stages exceeding 60 d of pregnancy. The mean BayesC AUC at stage 1 (1–30 d) was 0.58 ± 0.02 , which was superior to a random guess ($\text{AUC} = 0.50$) yet too low to be of practical use. The mean BayesC AUC at stage 7 (≥ 180 d) was 0.13 greater compared with that of stage 1 (1–30 d) and 0.07 to 0.10 greater compared with stages 2, 3, 4, 5, and 6 (31–180 d in 30-d increments). The mean AUC of stages 2 to 6 were 0.03 to 0.06 greater compared with stage 1 yet again too low to be of practical use. Because of high multicollinearity between many adjacent wavenumbers,

a spatially constrained clustering algorithm was used to adaptively partition wavenumbers into 68 windows before inferring associations of spectral regions with pregnancy. Pregnancy status was highly associated with wavenumber windows 1,063 to 1,134 cm^{-1} , 1,201 to 1,257 cm^{-1} , and 1,260 to 1,432 cm^{-1} based on an estimated BayesC posterior probability of association (PPA) approaching 100% for each of these windows at all pregnancy stages. Other windows ranging from 1,730 to 1,764 cm^{-1} , 1,775 to 1,992 cm^{-1} , 1,995 to 2,163 cm^{-1} , and 2,167 to 2,316 cm^{-1} had varying medium to high PPA (30% to 100%) across stages. The estimated PPA in wavenumber regions from 1,477 to 1,507 cm^{-1} , and 1,510 to 1,574 cm^{-1} was weaker in stages 1 and 2 compared with later stages, whereas for the regions 2,984 to 3,077 cm^{-1} and 3,081 to 3,133 cm^{-1} the effect of pregnancy was greater for stage 1 compared with other stages. Despite our conclusion that milk FT-MIR data poorly diagnose PS, our study provides new insights into spectral regions that are strongly associated with PS and warrant greater attention.

Key words: Bayesian variable selection, mid-infrared spectrum, pregnancy

INTRODUCTION

Fourier-transform (FT) mid-infrared (MIR) spectroscopy is a relatively inexpensive, fast, and accurate method used to analyze milk samples for components such as fat, protein, and lactose (ICAR, 2017). The FT-MIR absorbance readings at individual wavenumbers within the MIR region are due to stretching and bending of specific chemical bonds present in milk, thereby helping characterize its chemical composition. There has been substantial research on MIR spectra for predicting important phenotypes including methane emissions (Vanlierde et al., 2018), ketone bodies (Grellet et al., 2016), fatty acids (Soyeurt et al., 2006; De Marchi et al., 2014), total protein, casein, β -LG, whey protein, and glycosylated κ -CN (Bonfatti et al., 2011).

An early and accurate diagnosis of pregnancy status (PS) is important for timely reproductive management

Received July 28, 2021.

Accepted December 21, 2021.

*Corresponding author: tempelma@msu.edu

in livestock (Balhara et al., 2013) and is considered to be an ethical issue regarding the culling of pregnant livestock (More et al., 2017). Pregnancy diagnosis is routinely conducted using rectal palpation, ultrasonography, and analysis of pregnancy-associated glycoproteins in milk or blood between 26 and 40 d after insemination (Reese et al., 2018). During pregnancy, milk yield typically declines and milk composition changes with advancing stages of gestation due to altered nutrient partitioning between various physiological functions (Penasa et al., 2016). As this should also be reflected in changes in milk spectral absorbance at various wavenumbers, some researchers have recently focused their attention toward diagnosis of PS using FT-MIR data on milk since this does not require any extra phenotyping and animal handling costs (Toledo-Alvarado et al., 2018; Delhez et al., 2020; Brand et al., 2021). Toledo-Alvarado et al. (2018) investigated the diagnosis accuracy of PS from raw milk spectral data in Italian dairy cattle and determined, via a random 10-fold cross-validation, a mean area under the receiver operating characteristic curve (AUC) of 60%, albeit across a broad expanse of pregnancy stages. Delhez et al. (2020) highlighted the use of FT-MIR data to diagnose PS after 150 d of gestation in Australian Holsteins with an average AUC of 78% obtained through random cow independent cross-validation; nevertheless, diagnosing pregnancies this late might be far too late to be useful for management decisions.

Wang and Bovenhuis (2019) demonstrated with a dramatic illustration how naive random cross-validation strategies could result in overly optimistic predictions from MIR spectra to predict methane emissions. They recommended block cross-validation strategies whereby all animals within a contemporary group or herd are kept together within the same training or testing data sets and which we refer to as herd-independent cross-validation. Naive random cross-validation strategies can be suboptimal because they often fail to recognize the dependence structures (i.e., within herds) that persist as dependencies in the residuals between cows in training sets with their contemporary herd mates in testing sets. A good general discussion of this problem within other contexts is also provided in Roberts et al. (2017). Herd-independent cross-validation seems far more appropriate for assessing the potential of a new diagnostic for dairy industry applications because it should more closely reflect out-of-sample predictive ability of milk spectral data for PS on cows within herds that are not part of the research study in question.

Relative to, say, the tens or hundreds of thousands of SNP markers used currently in genomic evaluations (Wiggans et al., 2017), milk infrared spectra are low dimensional (~1,000 wavenumbers). Furthermore, ab-

sorbances on neighboring wavenumbers are often highly correlated such that the effective number of wavenumbers is substantially less (Rovere et al., 2021). The vast majority of FT-MIR studies predict phenotypes using partial least squares (PLS) for continuous responses and PLS discriminant analysis (PLS-DA) for binary outcomes (Bresolin and Dórea, 2020). Bayesian regression models originally intended for high dimensional regression of traits on genotypes (Meuwissen et al., 2001) have also been used to predict responses on various traits using FT-MIR spectra (Toledo-Alvarado et al., 2018). In a recent review, Bresolin and Dórea (2020) reported that, compared with PLS, Bayesian variable selection models such as BayesB (Meuwissen et al., 2001) or BayesC (Habier et al., 2011) can result in improved prediction accuracies when using milk spectral data to predict complex phenotypes. Ferragina et al. (2015) reported that BayesB outperformed PLS methods for predicting various fatty acids, whereas El Jabri et al. (2019) demonstrated that PLS generally outperformed Bayesian regression models for cheese-making properties. Furthermore, Bonfatti et al. (2017) reported that the differences in prediction accuracies between PLS and Bayesian methods depended upon the trait. However, formal tests of difference in performance between these 2 classes of methods are lacking in most of these studies. There may be considerable merit to using Bayesian over PLS methods as many Bayesian methods involve both shrinkage estimation and variable selection, whereas PLS is a dimension reduction procedure that often requires judicious deletion of wavenumbers to avoid suboptimal predictions on the trait of interest (Bresolin and Dórea, 2020). The comparison of PLS-DA and Bayesian methods for the diagnosis of PS using MIR spectra has not been extensively studied.

Joint analyses of the effects of all wavenumbers together can improve precision and account for population structure in a manner similar to that advocated for GWAS (Kang et al., 2008). That is, associations between individual wavenumbers and phenotypes may be biased by background management strategies and dietary rations if associations between wavenumber absorbances and the trait of interest are estimated one wavenumber at a time as is often done (e.g., Toledo-Alvarado et al., 2018). On the other hand, it has also been demonstrated in GWAS that SNP-based associations that simultaneously fit all other SNP markers can be badly maligned by a high degree of local multicollinearity involving adjacent SNP markers (Fernando et al., 2017). That is, in a joint association study involving FT-MIR wavenumbers, any single wavenumber may not display a strong association with the trait after adjusting for the effects of highly correlated neighboring wavenumbers. Hence, it seems then not surprising that various

researchers (Lainé et al., 2017; Toledo-Alvarado et al., 2018) have reported single wavenumber associations with PS based on one wavenumber at a time analyses. Lainé et al. (2017) determined that the standardized solutions of coefficients are greater for days pregnant compared with milk production traits (milk yield, fat and protein contents) within some spectral regions when using spectral absorbance measures of individual wavenumbers as the response variables.

One strategy to mitigate this multicollinearity problem in variable selection models and adaptively reduce the number of hypotheses to be tested is to aggregate or cluster wavenumbers into windows based on correlation. Ghosh and Ghattas (2015) demonstrated the utility of determining joint inclusion probabilities for windows of correlated covariates, rather than the marginal inclusion probabilities of individual covariates, in a Bayesian model application that also involved infrared spectral data. Windows-based strategies have been similarly advocated for GWAS (Chen et al., 2017) and, most recently, for associating MIR regions with fatty acid composition in Michigan Holsteins (Rovere et al., 2021).

The objectives of this study were (1) to compare PLS-DA versus Bayesian methods, specifically BayesC, for diagnosing pregnancy using milk MIR spectral data at different various of gestation in large Midwest US dairy herds using herd-independent cross-validation, and (2) to identify specific wavenumber regions that are potentially important for the diagnosis of pregnancy at various stages after conception.

MATERIALS AND METHODS

Field Data Retrieval and Editing

Milk spectral data and records of insemination were obtained from the Dairy Records Management Systems (www.drms.org) on Holstein cows from herds located in Michigan, Ohio, and Indiana. Data were collected in 2018 and 2019 from 123 herds milking at least 100 cows on a DHI test-day. Milk samples were sent to Central Star Cooperative Inc. (Grand Ledge, MI) for analysis using 2 FT-MIR spectrometers (Bentley Instruments NexGen Series FTS Combi machine). The milk FT-MIR absorbances were collected on 899 spectral wavenumbers ranging from 649 to 3,999 cm^{-1} .

The absorbance data for each wavenumber was scaled to a mean of zero and variance of one. A separate principal component analysis was performed on spectral absorbance data within each herd. The Mahalanobis distance was calculated from the first 15 principal components as these scores generally explained more than 95% of the spectral absorbance variance.

Outlier determination was based on comparing these Mahalanobis distances to a chi-squared distribution with degrees of freedom 15 using a type I error rate of 5% and Bonferroni-adjusted for the number of samples within each herd. Potential spectral outliers (11,936 out of 1,600,760 records) were removed because of this edit.

All cows that became pregnant during the study period were first retrieved from the database. Pregnancy was confirmed by requiring a subsequent calving record to be within 3 standard deviations (265–295 d) of the typical Holstein gestation length (Vieira-Neto et al., 2017) after the matching insemination date, thereby leading to confirmed pregnancies on 28,242 cows. We subsequently matched 94,728 milk samples from conception up until 275 d of pregnancy on these 28,242 cows. Each of these 94,728 milk samples were subsequently matched with milk samples from an open cow within the same contemporary group (herd-test-day) and such that the open cow's corresponding DIM was specified to be within a range of ± 10 d of the milk sample from the pregnant cow. This case-control matching was used to avoid potentially confounding biases due to the effect of stage of lactation, i.e., energy balance, when inferring spectral data associations with pregnancy. The PS, defined as pregnant (PS = 1) or open (PS = 0), was determined for each test date by matching insemination records with subsequent calving dates.

As milk components change with advancing pregnancy (Roche, 2003; Penasa et al., 2016), a stronger FT-MIR based signal should be expected with later stages of pregnancy (Lainé et al., 2017; Delhez et al., 2020). Hence, the data set was further partitioned into 7 different stages of pregnancy (i.e., days after conception) in a manner similar to that considered by Delhez et al. (2020). These 7 different stages of pregnancy were 1–30 d, 31–60 d, 61–90 d, 91–120 d, 121–150 d, 151–180 d, and ≥ 181 d as further summarized in Table 1. Separate analyses were conducted within each stage, again recognizing that the pregnancy signals, and hence associations with milk spectral data, could vary over gestation.

Spectral Data Preprocessing Strategies

We considered preprocessing of spectral data using several different strategies advocated previously in the literature. One such strategy is piecewise direct standardization (PDS) based on the use of principal component regression as illustrated by Grelet et al. (2015). We conducted principal component regression of wavenumber absorbances from one spectrophotometer ("slave") onto the other ("master") using 42 different sets of calibration samples from Eastern Laboratory Services (Medina, OH) over the course of the study

period, each set based on 24 raw milk samples. We were able to identify these calibration samples and their FT-MIR spectra from the spectrophotometers at various intervals ranging from one week to 3 mo. Using the raw absorbance data as provided, we adapted the approach of Grelet et al. (2015) based on sliding windows of 5 adjacent wavenumbers to do a 4-component principal component regression with the exception of the first 2 (649 and 652 cm^{-1}) and the last 2 wavenumbers (3,995 and 3,999 cm^{-1}) for which a simple linear regression was conducted. Spectral data from commercial milk samples were then calibrated using the principal component regression-based prediction equations to the respective calibration samples to which they were closest in time. This standardization is intended to reduce systematic bias caused by the use of different machines and potential machine drift over time (Grelet et al., 2015).

Derivatives have also been used to remove baseline variability (Rinnan et al., 2009) as an alternative to the use of PDS in dairy science research (Delhez et al., 2020). Hence, we also considered the use of first-derivative and second-derivative processing of raw spectra based on the smoothing method of Savitzky and Golay (1964). We specifically used the prospectr package (Stevens and Ramirez-Lopez, 2020) in R (2020; <https://www.r-project.org/>) based on a window size of 7 and polynomial degree of 3. Our final choice of data preprocessing (i.e., leaving spectral absorbances as is, PDS, or the use of first or second derivatives) was based on a 10-fold herd-independent cross-validation strategy separately within each stage using the computationally expedient PLS-DA model. In all cases, spectral data for each wavenumber was rescaled to a mean 0 and variance 1 before PLS-DA.

To formally compare these different preprocessing strategies, a separate ANOVA for each stage was fitted to AUC (defined later) with data preprocessing strategy as a fixed effects factor and fold as a random effects factor. The best fitting data preprocessing strategy was based on a formal post-ANOVA mean AUC comparison using Tukey's test.

PLS Analyses

The PLS-DA method within the caret package (Kuhn, 2008) in R was used to classify PS with milk spectral absorbance records after insemination being the independent variables. Herd-independent cross-validation was performed, which is described later in detail. The AUC was used to fine tune the model with respect to the number of components with the maximum number of PLS-DA components set to 30 to prevent overfitting.

BayesC

Our analysis was based on a Bayesian generalized linear model with a probit link function. The probability $\Pr(y_i = 1 | X_i)$ of a confirmed pregnancy ($y_i = 1$) on cow i was defined as

$$\Pr(y_i = 1 | X_i) = \Phi\left(\beta_0 + \sum_{j=1}^{899} X_{ij}\beta_j\right), i = 1, 2, \dots, n,$$

where $\Phi(\cdot)$ represents the cumulative distribution function of the standard normal distribution. Also, β_0 is an intercept, X_{ij} are the scaled (i.e., to a mean of 0 and variance 1 by wavenumber $j = 1, 2, \dots, 899$) absorbances for the milk sample from cow i , β_j is the effect of wavenumber j such that $\beta_j \sim N(\beta_j | 0, \sigma_\beta^2)$ with prob-

Table 1. Number of pregnant and open records and cows, and DIM (mean, with SD in parentheses) for different stages of pregnancy

Stages after insemination	Status	No. of records	No. of cows	DIM
1 (1–30 d)	Pregnant	5,119	4,922	120 (40)
	Open	5,119	4,520	115 (39)
2 (31–60 d)	Pregnant	5,805	5,505	147 (40)
	Open	5,805	5,329	140 (40)
3 (61–90 d)	Pregnant	6,334	5,882	178 (41)
	Open	6,334	5,861	171 (41)
4 (91–120 d)	Pregnant	6,429	6,061	209 (41)
	Open	6,429	6,109	201 (42)
5 (121–150 d)	Pregnant	6,545	6,166	241 (43)
	Open	6,545	6,165	234 (43)
6 (151–180 d)	Pregnant	6,729	6,266	272 (43)
	Open	6,729	6,048	265 (44)
7 (≥ 181 d)	Pregnant	10,403	8,289	309 (42)
	Open	10,403	7,287	303 (43)

ability π and $\beta_j = 0$ with probability $1 - \pi$ with π being the prior probability of a nonzero wavenumber effect. Furthermore, σ_β^2 is the variance for nonzero wavenumber effects.

This model can be implemented as the binary BayesC model using the BGLR package (Pérez and de los Campos (2014) in R. We chose BayesC as opposed to the more popular BayesB method since BayesC is not likely to be sensitively different from BayesB for binary data yet is nearly half as computationally costly for each Markov chain Monte Carlo (MCMC) cycle. This additional cost for BayesB is because of the additional data augmentation required to sample wavenumber-specific variances at each cycle (Meuwissen et al., 2001) leading to Student t distributed, rather than normally distributed, nonzero wavenumber effects.

In BayesC, the prior proportion of nonzero effect (π) is assigned a Beta prior such that $\pi \sim \text{Beta}(p_0, \pi_0)$ with $p_0 = 10$ representing the prior count and $\pi_0 = 0.2$ representing the prior mean using the parameterization of Pérez and de los Campos (2014). The prior density of the variance component σ_β^2 is specified to be a scaled inverse chi-squared density with hyperparameters df_0 and s_0 . We specified df_0 as 15.6 and s_0 as 0.33 based on prior experiences with the data to facilitate reasonable mixing with the MCMC algorithm invoked with the BGLR R package yet allow a fairly diffuse prior. A total of 240,000 MCMC iterations were conducted with the first 60,000 iterations discarded as burn-in and a subsequent thinning interval of 10 iterations such that 18,000 MCMC samples were used to estimate posterior means. Convergence of the models was checked using visual inspection of trace plots and by calculating the effective sample size of the slowest mixing parameters, π and σ_β^2 , post burn-in using the CODA package (Plummer et al., 2006) in R. The number of MCMC iterations were chosen to ensure that effective sample size of these parameters was greater than 100.

Cross-Validation

To evaluate the predictive ability of the models, a stratified 10-fold herd-independent cross-validation scheme was used to split the data into training (~90% of herds) and testing sets (~10% of herds) separately within each pregnancy stage such that the distribution of herd sizes was similar between all 10 testing sets. The cross-validation assessment of model performance was evaluated using the AUC. A predicted probability for pregnancy (PPP) of 0.5 is typically used as the threshold for classification (i.e., those animals with PPP ≥ 0.5 are classified as pregnant whereas those animals with PPP < 0.5 are classified as open). The ROC curve plots

the true positive rate against the false positive rate as the PPP threshold changes from 0 to 1, whereas the AUC represents the area underneath this ROC curve. An AUC of 0.5 implies a random guess [i.e., the method is not effective (Zou et al., 2007)]. The AUC was estimated using training data-based predictions to predict testing set phenotypes using the pROC package (Robin et al., 2011) in R.

Paired t -tests of PLS-DA-based assessments of AUC between noncalibrated, PDS-calibrated, first derivative and second derivative processed spectral data, blocking on cross-validation folds, was performed separately for each of the 7 pregnancy stages. The best performing calibration was then used to compare PLS-DA and Bayesian regression model predictions based on a paired t -test of AUC for each of the 7 pregnancy stages.

Posterior Probability of Association

Marginal Associations. The BayesC model is a Bayesian variable selection model such that one can readily estimate the probability of association (PPA) of any wavenumber j with PS by simply determining the proportion of all MCMC samples that are nonzero for that particular wavenumber effect β_j . In other words, the PPA for each wavenumber j (i.e., PPA_j) was

estimated as $PPA_j = \frac{\sum_{l=1}^N \left[I(\beta_{j(l)} \neq 0) \right]}{N}$, where N denotes the number of MCMC cycles saved for posterior inference, and $\beta_{j(l)}$ is a draw from the full conditional distribution of β_j at MCMC cycle l . Also, $I(\cdot)$ is the indicator function which is equal to 1 if the condition within the parentheses is true; otherwise it is equal to 0. These PPA were determined based on separate analyses of the data sets for each stage of gestation.

Joint Associations. Recent statistical research (Ghosh and Ghattas, 2015) has indicated that estimated PPA for individual covariates (i.e., wavenumbers) can be badly misleading when these covariates are highly correlated with each other, as is generally true for MIR data. They advocated that any Bayesian variable selection involving associations of responses with highly correlated covariates be based on windows of joint (i.e., multiple covariate) rather than marginal (i.e., single covariate) inferences.

Windows of highly correlated wavenumbers were created using a spatially constrained hierarchical agglomerative clustering (HAC) technique whereby highly correlated and adjacent wavenumbers are adaptively clustered. This HAC was implemented using the adjclust package in R with details provided in Ambrose et al. (2019). The number of windows were determined using the slope heuristic method described in Ambrose

et al. (2019) and implemented using the capushe function in the adjclust package.

The joint PPA of all n_k wavenumbers within a window k was computed using
$$\text{PPA}_k = \frac{\sum_{l=1}^N I\left(\sum_{j=1}^{n_k} |\beta_{kj(l)}| \neq 0\right)}{N}$$
 as described in further detail but within a genomic context by Chen et al. (2017). Here $\beta_{kj(l)}$ is the random draw from the full conditional distribution of β_j for wavenumber j located within window k drawn during MCMC cycle l . Note that $I\left(\sum_{j=1}^{n_k} |\beta_{kj(l)}| \neq 0\right)$ is equal to 1 when any of the draws of $\beta_{kj(l)}$ within window k are sampled to be nonzero.

For each pregnancy stage, we also determined the posterior mean of the proportion (s^2) of the total variance on the underlying latent scale due to spectral data associations in a manner not different from that determined for heritabilities in genomic studies except that spectral data absorbances rather than SNP genotypes are used as markers (Lehermeier et al., 2017). Denote the spectral absorbance merit of animal or record i sampled at MCMC cycle l as $s_{i(l)} = \sum_{j=1}^{899} X_{ij} \beta_{j(l)}$. Then the sampled spectral variance at MCMC cycle is
$$\sigma_{s(l)}^2 = n^{-1} \sum_{i=1}^n \left(s_{i(l)} - \sum_{j=1}^{899} \bar{X}_j \beta_{j(l)} \right)^2$$
 where $\bar{X}_j$ denotes the average spectral absorbance for wavenumber j in the corresponding data set. This further simplifies to
$$\sigma_{s(l)}^2 = n^{-1} \sum_{i=1}^n \left(s_{i(l)} \right)^2$$
 because absorbances were standardized to a mean of 0. Noting that the residual variance is equal to 1 on the underlying latent variable scale for a probit link model, we define the “spectral repeatability” or s^2 sampled in MCMC cycle l to be $s_{(l)}^2 = \sigma_{s(l)}^2 / (\sigma_{s(l)}^2 + 1)$ such the mean and standard deviation of $s_{(l)}^2$ across the N MCMC cycles are the corresponding posterior mean and standard deviation estimates for s^2 .

RESULTS AND DISCUSSION

Data Characterization and Summary Statistics

Descriptive statistics for pregnant and open records are presented in Table 1. Here, milk samples at various test dates are obtained from multiparous cows at various stages of lactation (35–420 d). Mean DIM at the milk test naturally increased from early (i.e., stage 1) to later (i.e., stage 7) stages of pregnancy. Nevertheless, as indicated previously, mean DIM of pregnant and matching open cows within each pregnancy stage were in the range of 10 DIM because of the case-control

matching strategy described previously, although our sorting and matching algorithm did lead to a slightly higher mean DIM for pregnant animals. This DIM pairing within each pregnancy stage was used to avoid the partial confounding of lactation stage with pregnancy stage as milk components change across lactation stages (Mayeres et al., 2004; Delhez et al., 2020). Roughly 70% of pregnant cows had repeated records across the pregnancy stages, whereas a small percentage (5–7%) of cows had multiple records within pregnancy stages.

Assessment of Spectral Data Processing Strategy

Before formally comparing PLS-DA and Bayesian methods for their prediction of PS, we set out to determine the most appropriate preprocessed form for the spectral data for use in that comparison using just PLS-DA for computational reasons.

We determined that the mean AUC using raw spectral data did not typically differ ($P > 0.05$) from mean AUC using either first or second derivative processed spectra but was often greater ($P < 0.05$) than the mean AUC of PDS-calibrated spectral data at several different stages (2, 3, 4, 6, and 7) of pregnancy (Supplemental Web File S1, <https://doi.org/10.6084/m9.figshare.15017046>; Khanal and Tempelman, 2021). Hence, to simplify and for ease of interpretation, all subsequent results reported in this paper are based on raw unprocessed spectral data, albeit, again, standardized to a mean 0 and variance 1. In extensive literature reviews, De Marchi et al. (2014) and Bresolin and Dórea (2020) reported mixed results in assessing the prediction performance of various forms of spectral data preprocessing from a large number of different studies, suggesting no clear evidence of spectral data preprocessing affecting prediction performance.

Comparison of Models for Diagnosing Pregnancy

We compared the cross-validation performance of PLS-DA and BayesC based on whole unprocessed spectra after rescaling absorbances for each wavenumber to a mean 0 and variance 1. The cross-validation prediction accuracies (as measured by AUC) at different stages for PLS-DA and BayesC are presented in Table 2. The BayesC method demonstrated higher mean AUC ($P < 0.05$) compared with PLS-DA at all stages except for nonsignificant differences at the first 2 stages. Estimated AUC ($\sim 0.58 \pm 0.02$) for both methods at stage 1 (1–30 d) were particularly low. Nevertheless, there was evidence that the mean AUC were greater than 0.50 (i.e., a random guess) for every stage and for both methods ($P < 0.0001$). In a recent review paper, Bresolin and Dórea (2020) reported that Bayesian models

often have better prediction properties compared with PLS-DA methods for some complex phenotypes. Some previous comparisons (Ferragina et al., 2015; Bonfatti et al., 2017; El Jabri et al., 2019; Grelet et al., 2020) have determined that Bayesian methods that incorporate variable selection outperform PLS-DA if whole spectra are used, whereas comparable prediction properties have been noted if various regions known to be highly sensitive to water absorption are removed. Our results demonstrated better predictive performance of BayesC for a categorical trait (PS). Bayesian variable selection methods, such as BayesB and BayesC, discount covariates that are clearly unimportant (de los Campos et al., 2013), whereas the dimension reduction strategy used in PLS-DA can be suboptimal when all spectral data are included (Visentin et al., 2015; Bonfatti et al., 2017). However, as recently noted by Rovere et al. (2021) and Tiplady et al. (2019), strong signals of association are often determined in nominal water regions.

For comparisons between different stages of pregnancy for predictive performance and for wavelength associations, we focused on estimates from BayesC because it had superior performance compared with that of PLS-DA as previously noted. Mean AUC were unsurprisingly larger for later stages of gestation. From Table 2, estimated AUC (0.71 ± 0.02) at stage 7 (≥ 180 d) was 0.13 greater ($P < 0.0001$) compared with that of stage 1 (1–30 d) and 0.07 to 0.10 greater ($P < 0.0001$) compared with stages 2, 3, 4, 5, and 6 (31–180 d). The mean AUC of stages 2 to 6 did not vary substantially from each other but were 0.03 to 0.06 greater ($P < 0.001$) compared with stage 1.

The increasing trend of mean AUC across advanced stages of pregnancy in our study agreed with Delhez et al. (2020), although the mean AUC for all stages were lower in our study. This difference between the 2 studies may be due to greater number of commercial herds (123) varying widely in their management conditions as well as the herd-independent cross-validation strategy used in our study, whereas a random cow independent cross-validation was conducted using a much smaller number (19) of herds by Delhez et al. (2020). Wang and Bovenhuis (2019) reported large differences in prediction accuracies between random versus herd-independent cross-validation and suggested that random cross-validation underestimates the error rate of the prediction equation when cows from the same batch (i.e., herd) are included in both training and validation sets and there are systematic differences between batches. Herd-independent cross-validation recognizes that each herd has its own independent management and feeding systems which influence milk components (Chilliard et al., 2007) and potentially then the association between milk spectral data and pregnancy. This

Table 2. Estimated receiver operating characteristic areas under the curve (AUC) for partial least squares discriminant analysis (PLS-DA) and BayesC based on 10-fold cross-validation at various stages of pregnancy

Stages after insemination	PLS-DA ¹	BayesC ¹
1 (1–30 d)	0.57 (0.02) ^{a,x}	0.58 (0.02) ^{a,x}
2 (31–60 d)	0.61 (0.02) ^{b,x}	0.61 (0.02) ^{b,x}
3 (61–90 d)	0.61 (0.02) ^{b,x}	0.62 (0.01) ^{b,y}
4 (91–120 d)	0.61 (0.01) ^{b,x}	0.62 (0.01) ^{b,y}
5 (121–150 d)	0.60 (0.01) ^{b,x}	0.61 (0.01) ^{b,y}
6 (151–180 d)	0.62 (0.01) ^{b,x}	0.64 (0.01) ^{b,y}
7 (≥ 181 d)	0.69 (0.01) ^{c,x}	0.71 (0.02) ^{c,y}

^{a–c}Within a method (i.e., column), stages not sharing common superscript differ ($P < 0.05$) from each other.

^{x,y}Within a stage (i.e., row), methods not sharing same letter differ ($P < 0.05$) from each other.

¹Estimate (SE).

cross-validation strategy would be most pertinent for inferring the effectiveness of using FT-MIR data to diagnose pregnancy within herds not included in this study. Herd-independent cross-validation strategies have not been seemingly considered in previous studies involving the use of spectral data to diagnose pregnancies in dairy cattle (Toledo-Alvarado et al., 2018; Delhez et al., 2020; Brand et al., 2021).

Brand et al. (2021) recently reported impressive prediction sensitivities and specificities exceeding 85% for pregnancy diagnosis using deep learning methods; however, they also reported sensitivities and specificities that were almost as impressive (82%) using PLS-DA. Conversely, we observed sensitivities ranging from 54% to 65% and specificities ranging from 58 to 64% from stages 1 to 7 using PLS-DA (results not reported), whereas we determined sensitivities and specificities for BayesC to be only marginally better. The difference in PLS-DA performance between the Brand et al. (2021) study and our study could be attributed to several factors. First, as with many other studies, the Brand et al. cross-validation scheme was based on purely random partitioning of training and validation data sets, whereas we conducted block or herd-independent cross-validation. Second, spectral data collected for nonpregnant or open records in their study were all based on milk samples collected before the first insemination, thereby partially confounding lactation stage with PS. Third, milk records for pregnant cows from a broad expanse of gestational stages in their study were combined together instead of partitioning them into different stages of pregnancy as we did such that it was not possible to parse out the stronger late gestation stage signals from the weaker early stage lactation signals.

Milk production decreases, whereas fat, lactose, and protein content all increase as pregnancy advances

(Olori et al., 1997; Penasa et al., 2016). In later stages of gestation, the growth and nutrient requirements of conceptus increases compared with earlier stages (Loker et al., 2009). Furthermore, estrogen levels in maternal blood gradually increase from 3 mo with no meaningful differences in levels between 3 to 5 mo and 5 to 7 mo of pregnancy, followed by a sudden increase thereafter (Parkhie et al., 1966), thereby potentially affecting nutrient partitioning (Olori et al., 1997). This could be a reason for observing a higher mean AUC at later stages compared with earlier stages.

It is generally recognized that diagnostic procedures should have AUC >0.70 before they are considered to be useful (Mandrekar, 2010). Hence our study implies that milk FT-MIR data has limited use for early diagnosis of PS in dairy cows even though the AUC values generally increased over the stages of gestation. Detection of PS at early ages with higher accuracies is important to replace more conventional methods of pregnancy diagnosis in order for FT-MIR-based predictions to be useful.

We do recognize that some cows classified as open in our study may have actually conceived and lost their embryo after their corresponding milk sample collection. This could be a contributing source of attenuating bias, particularly in stage 1 where such losses are known to approach 35% of conceptions (Santos et al., 2004). Even within the first half of stage 2 (30–45 d), embryo losses can occur in as high as 13% of pregnancies (Santos et al., 2004). Hence, this may be a contributing factor to the increasing signals associated with later stages of pregnancy in our analyses.

Wavenumber Associations With Pregnancy

We inferred the marginal associations of individual wavenumbers with PS as well as the joint associations of adaptively clustered windows of correlated wavenumbers with PS. In our study, we inferred 68 different windows based on the HAC algorithm with the number of wavenumbers per window ranging from 1 to 79 with a median of 7. These window ranges, along with the number of wavenumbers per window, are presented in Table S2 in Supplemental Web File S2 (<https://doi.org/10.6084/m9.figshare.15017046>; Khanal and Tempelman, 2021). The marginal PPA of individual wavenumbers for each pregnancy state are presented in the left panels of Figure 1, whereas the joint PPA of windows of clustered wavenumbers for each pregnancy stage are presented in the right panels of Figure 1. These estimated PPA are also tabulated in detail in Supplemental Web File S2 (Tables S3 and S4; <https://doi.org/10.6084/m9.figshare.15017046>; Khanal and Tempelman, 2021). For all stages, the marginal PPA

of individual wavenumbers were generally lower compared with the corresponding joint PPA of windows-based associations within the same regions. These differences may be attributed to the multicollinearity problem characterized by Ghosh and Ghattas (2015) and further illustrated by Fernando et al. (2017) in the context of GWAS. Furthermore, there seemed to be far greater consistency with inferences on peaks (i.e., high PPA) associated with windows as opposed to individual wavenumbers across the various stages of pregnancy as can be observed from comparing the left versus right panels of Figure 1.

We determined that pregnancy was consistently highly (PPA = 1) associated with wavenumber regions 1,063 to 1,134 cm^{-1} , 1,201 to 1,257 cm^{-1} , and 1,260 to 1,432 cm^{-1} for all pregnancy stages (Supplemental Table S4). As an additional check (results not reported), we noted that the joint PPA for each training set within each cross-validation fold for each stage ranged from 0.85 to 1 for both regions 1,063 to 1,134 cm^{-1} and 1,260 to 1,432 cm^{-1} . The corresponding joint PPA of wavenumbers within the region 1,201 to 1,257 cm^{-1} ranged from 0.3 to 1.0 across folds and stages; nevertheless, the median joint PPA of this region was 0.75 across the different stages and folds. Our results agreed with those reported by Toledo-Alvarado et al. (2018) who noted that wavenumbers 1,245 and 1,399 cm^{-1} were associated with pregnancy. Bittante and Cecchinato (2013) reported that the range of wavenumbers from 930 to 1,582 cm^{-1} correspond to peaks of absorbance pertaining to the bonds C–H, CH_2 , aromatic C=C, C=O, and N–O all contained within the “fingerprint region” (Karoui et al., 2010), which has been used to distinguish between a large number of molecular compounds based on absorbance band patterns. Absorbances at wavenumbers near 1,375 cm^{-1} have been previously associated with pregnancy in Holstein cows (Lainé et al., 2017) and also with the stretching mode of N–O and NO_2 (Socrates, 2004), which are chemical bonds believed to constitute pregnancy-associated proteins (Lainé et al., 2017).

We determined that the signals for pregnancy in range of wavenumbers from 1,477 to 1,507 cm^{-1} and from 1,510 to 1,574 cm^{-1} were lower for stages 1 and 2 compared with later stages. These regions are known to be associated with protein and amines where chemical bonds that contain nitrogen molecules (C–N, N–N) are specific to the region from 1,492 to 1,546 cm^{-1} (Kaylegian et al., 2006; Lynch et al., 2006). Zaalberg et al. (2020) also reported that the bonds (CO–N and N–H) associated with protein molecules are associated with wavenumbers that range from 1,500 to 1,550 cm^{-1} and the region around 1,600 cm^{-1} . The wavenumber 1,485 cm^{-1} is associated with $(\text{CH}_3)_3\text{N}^+$ asymmetric bending

Figure 1. Posterior probability of association (PPA) for individual wavenumbers (column A) and adaptively clustered windows wavenumbers (column B) at different stages of pregnancy (stage 1: 1–30 d, stage 2: 31–60 d, stage 3: 61–90 d, stage 4: 91–120 d, stage 5: 121–150 d, stage 6: 151–180 d, and stage 7: ≥ 181 d). Vertical reference lines provided at known associations for fat A ($1,747\text{ cm}^{-1}$) and fat B ($2,873\text{ cm}^{-1}$). Alternating colors denote ranges of 68 different windows (see also Table S2 in Supplemental Web File S2; <https://doi.org/10.6084/m9.figshare.15017046>; Khanal and Tempelman, 2021) in column B.

(Stuart, 2004). Lainé et al. (2017) and Toledo-Alvarado et al. (2018) also found important associations with pregnancy at around $1,550\text{ cm}^{-1}$, although, again, their definitions of pregnancy were much broader than the 7 different stages in our study. The higher PPA in later stages compared with earlier stages for these regions could be due to the higher protein content of milk in later stages of gestation (Roche, 2003) compared with earlier stages. The wavenumber in this region is known to be associated with the *ANKH* gene (Zaalberg et al., 2020), which is strongly associated with α -LA and β -LG (Sanchez et al., 2017). Bleck et al. (2009) reported an increase in α -LA and β -1,4-galactosyltransferase in milk of cows as gestation advances and both enzymes are associated with lactose concentration in milk. Lower lactose levels in earlier stages of gestation (Olori et al., 1997) may explain the lower PPA for stage 1 to 30 d in the wavenumber regions of $1,477$ to $1,507\text{ cm}^{-1}$ and $1,510$ to $1,574\text{ cm}^{-1}$.

Our results showed that windows from $1,730$ to $1,764\text{ cm}^{-1}$, $1,775$ to $1,992\text{ cm}^{-1}$, $1,995$ to $2,163\text{ cm}^{-1}$, and $2,167$ to $2,316\text{ cm}^{-1}$ had widely varying PPA (from 0.20 to 1.00) with PS with no obvious trends across different stages of pregnancy. The wavenumber $1,747\text{ cm}^{-1}$ within the first window is known to be associated with fat A ($\text{C}=\text{O}$ stretch; Stuart, 2004; Kaylegian et al., 2006).

The windows of range from $2,693$ to $2,839\text{ cm}^{-1}$ and $2,842$ to $2,962\text{ cm}^{-1}$ had medium to high PPA (0.47 to 1.00) across different stages. We did not find any obvious trend in PPA across different stages but a higher PPA was determined for the latter 2 stages (stage 6 and stage 7). The vibrations of ester linkage and $\text{C}-\text{H}$ stretch groups related to fatty acids are located between $2,800$ to $3,000\text{ cm}^{-1}$ (Sivakesava and Irudayaraj, 2002). Bittante and Cecchinato (2013) and Grelet et al. (2016) reported the $\text{C}-\text{H}_3$ (methyl) group is associated with $2,862\text{ cm}^{-1}$. An important bandwidth for PS was previously inferred between $2,872$ and $2,973\text{ cm}^{-1}$ (Toledo-Alvarado et al., 2018) but again for a much broader definition of pregnancy. In this region, fat B (absorbance by carbon-hydrogen) is filtered (Stuart, 2004; Kaylegian et al., 2006). Between $2,862\text{ cm}^{-1}$ (CH_3) and $2,927\text{ cm}^{-1}$ (CH_2), a significant association of pregnancy was determined by Lainé et al. (2017). The 2 bands at $2,850$ and $2,920\text{ cm}^{-1}$ are attributable to asymmetrical and symmetrical methylene (CH_2) stretching vibration respectively (Sinclair et al., 1952).

Our results showed that wavenumber regions ranging from 906 to 973 cm^{-1} , $1,577$ to $1,626\text{ cm}^{-1}$, $1,667$ to $1,693\text{ cm}^{-1}$, $2,984$ to $3,077\text{ cm}^{-1}$, $3,081$ to $3,133\text{ cm}^{-1}$, $3,662$ to $3,703\text{ cm}^{-1}$, and $3,707$ to $3,998\text{ cm}^{-1}$ had highly varying PPA (0.06 to 0.95) across various stages. The wavenumber region $1,577$ to $1,627\text{ cm}^{-1}$ had greater

PPA at stages 5, 6, and 7 compared with stage 1 and 2, whereas the wavenumber regions $2,984$ to $3,077\text{ cm}^{-1}$ and $3,081$ to $3,133\text{ cm}^{-1}$ had greater PPA at stage 1 compared with later stages. All of these regions overlap substantially with “water” regions as reported by Tip-lady et al. (2019), Toledo-Alvarado et al. (2018), Bittante and Cecchinato (2013), and Grelet et al. (2015) and are believed to have low signal-to-noise ratios. We observed that the region 649 to 900 cm^{-1} was universally associated with near-zero PPA across all stages, whether for marginal or joint associations (Figure 1). These results were somewhat anticipated as this region has been associated with an area of the spectrum where the calcium fluoride cuvette of the spectrophotometer should render it opaque to infrared light (Rovere et al., 2021). Toledo-Alvarado et al. (2018) determined a peak at $3,683\text{ cm}^{-1}$ for pregnancy which is close to those known for characterizing the bonds: $\text{C}=\text{CH}_2$ and $\text{O}-\text{H}$ (alcohols, phenols, and carboxyl) group. Wang et al. (2016) and Wang and Bovenhuis (2018) identified wavenumbers ($1,619$ to $1,674\text{ cm}^{-1}$; $3,073$ to $3,667\text{ cm}^{-1}$) associated with fat, protein, or lactose percentage and associated with polymorphisms in the *DGAT1* gene. Zaalberg et al. (2020) also reported that $1,604\text{ cm}^{-1}$ is associated with *DGAT1* gene and $1,696\text{ cm}^{-1}$ is associated with the *PAEP* gene. Contrary to what has been often practiced for FT-MIR predictions of key traits in other studies (Bonfatti et al., 2011; Delhez et al., 2020), we, like Bittante and Cecchinato (2013), recommend using wavenumbers in nominal water regions, at least for those $>900\text{ cm}^{-1}$, for FT-MIR predictions. That is, these regions include absorbance peaks that are associated with nonwater components of milk as we have previously observed for fatty acids (Rovere et al., 2021) and as we seem to observe for PS in our study.

Proportion of Variance Explained by Milk Infrared Spectra

The proportion of variance explained by milk infrared spectrum, which we denote as spectral repeatability (s^2), was calculated in analogy to a genomic heritability but using spectral data instead of genomic markers. Spectral repeatability estimates are presented on Table 3. As anticipated, spectral repeatability estimates generally increased with increasing stages of pregnancy, ranging from 0.05 ± 0.02 (posterior mean $\pm$ posterior standard deviation) for stage 1 to 0.38 ± 0.12 for stage 7. Hence milk infrared spectral data at later stages seem more effective for diagnosing pregnancy relative to earlier stages as the spectral repeatability estimates were higher at later stages, thereby confirming our earlier reported cross-validation assessments. However, as also previously noted, the low estimates of s^2 at earlier

Table 3. Spectral repeatability (s^2) or estimated proportion of variance explained by the milk infrared spectrum at different stages of pregnancy using BayesC

Stages after insemination	Spectral repeatability ¹
1 (1–30 d)	0.05 (0.02)
2 (31–60 d)	0.09 (0.02)
3 (61–90 d)	0.08 (0.01)
4 (91–120 d)	0.11 (0.04)
5 (121–150 d)	0.09 (0.01)
6 (151–180 d)	0.13 (0.04)
7 (≥ 181 d)	0.38 (0.12)

¹Posterior mean (posterior SD).

stages limits the practical use of FT-MIR data for early detection of pregnancy.

We believe that the low early stage estimates of s^2 are somewhat counterbalanced by the strong associations (i.e., PPA approaching 100%) inferred between some spectral regions with PS in stage 1 as previously reported in this paper. By clear analogy, promising GWAS analyses have led to major QTL regions being discovered for pregnancy rates in spite of heritability estimates being lower than 10% (Ma et al., 2019).

CONCLUSIONS

Our study suggests that the potential for MIR spectroscopy to diagnose early the PS of dairy cows is limited. We compared PLS-DA regression and a Bayesian variable selection model BayesC and determined that BayesC had better out-of-sample prediction accuracies in diagnosing PS compared with PLS-DA. Cross-validation prediction accuracies were higher at later stages of gestation (greater than 150 d); however, those stages are too late to be of practical use for management decisions. Nevertheless, this study demonstrated some associations between pregnancy at all stages with various wavenumber regions in the mid-infrared spectrum that include plausible candidates for chemical bonds in molecules believed to be associated with pregnancy. The nature of these inferences is not much unlike what is typically reported for genomic studies or GWAS involving fertility in dairy cattle (i.e., low heritability estimates yet some genomic regions known to be strongly associated with various measures of fertility). Further research is needed to improve prediction accuracies in earlier pregnancy stages by combining the use of FT-MIR spectra with other omics data sources such as metabolomics, and genomic and molecular data. Window-based associations require greater consideration as associations of individual wavenumbers with PS can be muted by multicollinearity with individual wavenumbers, whereas one wavenumber at a

time regression may not fully account for population structures induced by management clusters.

ACKNOWLEDGMENTS

Financial support from the USDA-National Institute of Food and Agriculture (Washington, DC; grant number 2017-67007-25947) and from the Michigan Animal Agriculture Alliance is gratefully acknowledged. The authors also greatly appreciate the efforts of Central Star Co-operative (Grand Ledge, MI) and Bentley Instruments Inc. (Chaska, MN) in providing the spectral data on the corresponding milk samples. The authors also acknowledge the co-operation of John Clay from Dairy Records Management Systems (Raleigh, NC) for providing breeding and calving records on those same cows. The authors have not stated any conflicts of interest.

REFERENCES

- Ambroise, C., A. Dehman, P. Neuvial, G. Rigai, and N. Vialaneix. 2019. Adjacency-constrained hierarchical clustering of a band similarity matrix with application to genomics. *Algorithms Mol. Biol.* 14:22. <https://doi.org/10.1186/s13015-019-0157-4>.
- Balhara, A. K., M. Gupta, S. Singh, A. K. Mohanty, and I. Singh. 2013. Early pregnancy diagnosis in bovines: Current status and future directions. *ScientificWorldJournal* 2013:1–10. <https://doi.org/10.1155/2013/958540>.
- Bittante, G., and A. Cecchinato. 2013. Genetic analysis of the Fourier-transform infrared spectra of bovine milk with emphasis on individual wavelengths related to specific chemical bonds. *J. Dairy Sci.* 96:5991–6006. <https://doi.org/10.3168/jds.2013-6583>.
- Bleck, G., M. Wheeler, L. Hansen, H. Chester-Jones, and D. Miller. 2009. Lactose synthase components in milk: Concentrations of α -lactalbumin and β 1,4-galactosyltransferase in milk of cows from several breeds at various stages of lactation. *Reprod. Domest. Anim.* 44:241–247. <https://doi.org/10.1111/j.1439-0531.2007.01047.x>.
- 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>.
- Bonfatti, V., F. Tiezzi, F. Miglior, and P. Carnier. 2017. Comparison of Bayesian regression models and partial least squares regression for the development of infrared prediction equations. *J. Dairy Sci.* 100:7306–7319. <https://doi.org/10.3168/jds.2016-12203>.
- Brand, W., A. T. Wells, S. L. Smith, S. J. Denholm, E. Wall, and M. P. Coffey. 2021. Predicting pregnancy status from mid-infrared spectroscopy in dairy cow milk using deep learning. *J. Dairy Sci.* 104:4980–4990. <https://doi.org/10.3168/jds.2020-18367>.
- Bresolin, T., and J. R. R. Dórea. 2020. Infrared spectrometry as a high-throughput phenotyping technology to predict complex traits in livestock systems. *Front. Genet.* 11:923. <https://doi.org/10.3389/fgene.2020.00923>.
- Chen, C., J. P. Steibel, and R. J. Tempelman. 2017. Genome-wide association analyses based on broadly different specifications for prior distributions, genomic windows, and estimation methods. *Genetics* 206:1791–1806. <https://doi.org/10.1534/genetics.117.202259>.
- Chilliard, Y., F. Glasser, A. Ferlay, L. Bernard, J. Rouel, and M. Doreau. 2007. Diet, rumen biohydrogenation and nutritional qual-

- ity of cow and goat milk fat. *Eur. J. Lipid Sci. Technol.* 109:828–855. <https://doi.org/10.1002/ejlt.200700080>.
- de los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler, and M. P. L. Calus. 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193:327–345. <https://doi.org/10.1534/genetics.112.143313>.
- 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>.
- Delhez, P., P. N. Ho, N. Gengler, H. Soyeurt, and J. E. Pryce. 2020. Diagnosing the pregnancy status of dairy cows: How useful is milk mid-infrared spectroscopy? *J. Dairy Sci.* 103:3264–3274. <https://doi.org/10.3168/jds.2019-17473>.
- El Jabri, M., M. P. Sanchez, P. Trossat, C. Laithier, V. Wolf, P. Groperrin, E. Beuvier, O. Rolet-Répécaud, S. Gavoye, Y. Gaüzère, O. Belysheva, E. Notz, D. Boichard, and A. Delacroix-Buchet. 2019. Comparison of Bayesian and partial least squares regression methods for mid-infrared prediction of cheese-making properties in Montbéliarde cows. *J. Dairy Sci.* 102:6943–6958. <https://doi.org/10.3168/jds.2019-16320>.
- Fernando, R., A. Toosi, A. Wolc, D. Garrick, and J. Dekkers. 2017. Application of whole-genome prediction methods for genome-wide association studies: A Bayesian approach. *J. Agric. Biol. Environ. Stat.* 22:172–193. <https://doi.org/10.1007/s13253-017-0277-6>.
- Ferragina, A., G. de los Campos, A. I. Vazquez, A. Cecchinato, and G. Bittante. 2015. Bayesian regression models outperform partial least squares methods for predicting milk components and technological properties using infrared spectral data. *J. Dairy Sci.* 98:8133–8151. <https://doi.org/10.3168/jds.2014-9143>.
- Ghosh, J., and A. E. Ghattas. 2015. Bayesian variable selection under collinearity. *Am. Stat.* 69:165–173. <https://doi.org/10.1080/00031305.2015.1031827>.
- 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., P. Dardenne, H. Soyeurt, J. A. Fernandez, A. Vanlierde, F. Steevens, N. Gengler, and F. Dehareng. 2020. Large-scale phenotyping in dairy sector using milk MIR spectra: Key factors affecting the quality of predictions. *Methods*. <https://doi.org/10.1016/j.ymeth.2020.07.012>.
- Grelet, C., J. A. Fernández Pierna, P. Dardenne, V. Baeten, and F. Dehareng. 2015. Standardization of milk mid-infrared spectra from a European dairy network. *J. Dairy Sci.* 98:2150–2160. <https://doi.org/10.3168/jds.2014-8764>.
- Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick. 2011. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186. <https://doi.org/10.1186/1471-2105-12-186>.
- ICAR. 2017. Section 2—Guidelines for Dairy Cattle Milk Recording. *Int. Comm. Anim. Record. ICAR*.
- Kang, H. M., N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman, M. J. Daly, and E. Eskin. 2008. Efficient control of population structure in model organism association mapping. *Genetics* 178:1709–1723. <https://doi.org/10.1534/genetics.107.080101>.
- Karoui, R., G. Downey, and C. Blecker. 2010. Mid-infrared spectroscopy coupled with chemometrics: A tool for the analysis of intact food systems and the exploration of their molecular structure-quality relationships—A review. *Chem. Rev.* 110:6144–6168. <https://doi.org/10.1021/cr100090k>.
- Kaylegian, K. E., G. E. Houghton, J. M. Lynch, J. R. Fleming, and D. M. Barbano. 2006. Calibration of infrared milk analyzers: Modified milk versus producer milk. *J. Dairy Sci.* 89:2817–2832. [https://doi.org/10.3168/jds.S0022-0302\(06\)72555-3](https://doi.org/10.3168/jds.S0022-0302(06)72555-3).
- Khanal, P., and R. Tempelman. 2021. Supplemental Web Files for “The use of milk Fourier transform mid-infrared spectroscopy to diagnose pregnancy and determine spectral regional associations with pregnancy in US dairy cows” by Khanal and Tempelman (2021). figshare. <https://doi.org/10.6084/m9.figshare.15017046.v4>.
- Kuhn, M. 2008. Building predictive models in R using the caret package. *J. Stat. Softw.* 28:1–26. <https://doi.org/10.18637/jss.v028.i05>.
- 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>.
- Lehermeier, C., G. de los Campos, V. Wimmer, and C. C. Schön. 2017. Genomic variance estimates: With or without disequilibrium covariances? *J. Anim. Breed. Genet.* 134:232–241. <https://doi.org/10.1111/jbg.12268>.
- Loker, S., F. Miglior, J. Bohmanova, L. R. Schaeffer, J. Jamrozik, and G. Kistemaker. 2009. Short communication: Effect of preadjusting test-day yields for stage of pregnancy on variance component estimation in Canadian Ayrshires. *J. Dairy Sci.* 92:2270–2275. <https://doi.org/10.3168/jds.2008-1806>.
- Lynch, J. M., D. M. Barbano, M. Schweisthal, and J. R. Fleming. 2006. Precalibration evaluation procedures for mid-infrared milk analyzers. *J. Dairy Sci.* 89:2761–2774. [https://doi.org/10.3168/jds.S0022-0302\(06\)72353-0](https://doi.org/10.3168/jds.S0022-0302(06)72353-0).
- Ma, L., J. B. Cole, Y. Da, and P. M. VanRaden. 2019. Symposium review: Genetics, genome-wide association study, and genetic improvement of dairy fertility traits. *J. Dairy Sci.* 102:3735–3743. <https://doi.org/10.3168/jds.2018-15269>.
- Mandrekar, J. N. 2010. Receiver operating characteristic curve in diagnostic test assessment. *J. Thorac. Oncol.* 5:1315–1316. <https://doi.org/10.1097/JTO.0b013e3181ec173d>.
- 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).
- Meuwissen, T. H., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829. <https://doi.org/10.1093/genetics/157.4.1819>.
- More, S., D. Bicout, A. Botner, A. Butterworth, P. Calistri, K. Depner, S. Edwards, B. Garin-Bastuji, M. Good, C. Gortazar Schmidt, V. Michel, M. A. Miranda, S. Saxmose Nielsen, A. Velarde, H. Thulke, L. Sihvonen, H. Spoolder, J. A. Stegeman, M. Raj, P. Willeberg, D. Candiani, and C. Winckler. 2017. Animal welfare aspects in respect of the slaughter or killing of pregnant livestock animals (cattle, pigs, sheep, goats, horses). *EFSA J.* 15. <https://doi.org/10.2903/j.efsa.2017.4782>.
- 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. Fecheimer. 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>.
- Pérez, P., and G. de los Campos. 2014. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198:483–495. <https://doi.org/10.1534/genetics.114.164442>.
- Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: Convergence diagnosis and output analysis for MCMC. *R News* 6:7–11.
- Reese, S. T., M. H. C. Pereira, J. L. Edwards, J. L. M. Vasconcelos, and K. G. Pohler. 2018. Pregnancy diagnosis in cattle using pregnancy associated glycoprotein concentration in circulation at day 24 of gestation. *Theriogenology* 106:178–185. <https://doi.org/10.1016/j.theriogenology.2017.10.020>.
- Rinnan, Å., F. van den Berg, and S. B. Engelsen. 2009. Review of the most common pre-processing techniques for near-infrared spectra. *Trends Analyt. Chem.* 28:1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007>.

- Roberts, D. R., V. Bahn, S. Ciuti, M. S. Boyce, J. Elith, G. Guille-ra-Arroita, S. Hauenstein, J. J. Lahoz-Monfort, B. Schröder, W. Thuiller, D. I. Warton, B. A. Wintle, F. Hartig, and C. F. Dormann. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40:913–929. <https://doi.org/10.1111/ecog.02881>.
- Robin, X., N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J. C. Sanchez, and M. Müller. 2011. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77. <https://doi.org/10.1186/1471-2105-12-77>.
- 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).
- Rovere, G., G. de los Campos, A. L. Lock, L. Worden, A. I. Vazquez, K. Lee, and R. J. Tempelman. 2021. Prediction of fatty acid composition using milk spectral data and its associations with various mid-infrared spectral regions in Michigan Holsteins. *J. Dairy Sci.* 104:11242–11258. <https://doi.org/10.3168/jds.2021-20267>.
- Sanchez, M. P., A. Govignon-Gion, P. Croiseau, S. Fritz, C. Hozé, G. Miranda, P. Martin, A. Barbat-Leterrier, R. Letaïef, D. Rocha, M. Brochard, M. Boussaha, and D. Boichard. 2017. Within-breed and multi-breed GWAS on imputed whole-genome sequence variants reveal candidate mutations affecting milk protein composition in dairy cattle. *Genet. Sel. Evol.* 49:68. <https://doi.org/10.1186/s12711-017-0344-z>.
- Santos, J. E. P., W. W. Thatcher, R. C. Chebel, R. L. A. Cerri, and K. N. Galvão. 2004. The effect of embryonic death rates in cattle on the efficacy of estrus synchronization programs. *Anim. Reprod. Sci.* 82–83:513–535. <https://doi.org/10.1016/j.anireprosci.2004.04.015>.
- Savitzky, A., and M. J. E. Golay. 1964. Smoothing and differentiation of data by simplified least-squares procedures. *Anal. Chem.* 36:1627–1639. <https://doi.org/10.1021/ac60214a047>.
- Sinclair, R. G., A. F. McKay, and R. N. Jones. 1952. The infrared absorption spectra of saturated fatty acids and esters. *J. Am. Chem. Soc.* 74:2570–2575. <https://doi.org/10.1021/ja01130a033>.
- Sivakesava, S., and J. Irudayaraj. 2002. Rapid determination of tetracycline in milk by FT-MIR and FT-NIR spectroscopy. *J. Dairy Sci.* 85:487–493. [https://doi.org/10.3168/jds.S0022-0302\(02\)74099-X](https://doi.org/10.3168/jds.S0022-0302(02)74099-X).
- Socrates, G. 2004. Infrared and Raman Characteristic Group Frequencies: Tables and Charts. John Wiley & Sons Ltd.
- 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).
- Stevens, A., and L. Ramirez-Lopez. 2020. Package “prospectr” Miscellaneous Functions for Processing and Sample Selection of Spectroscopic Data.
- Stuart, B. H. 2004. Infrared Spectroscopy: Fundamentals and Applications. Analytical Techniques in the Sciences. John Wiley & Sons Ltd.
- Tiplady, K. M., R. G. Sherlock, M. D. Littlejohn, J. E. Pryce, S. R. Davis, D. J. Garrick, R. J. Spelman, and B. L. Harris. 2019. Strategies for noise reduction and standardization of milk mid-infrared spectra from dairy cattle. *J. Dairy Sci.* 102:6357–6372. <https://doi.org/10.3168/jds.2018-16144>.
- Toledo-Alvarado, H., A. I. A. I. Vazquez, G. de los Campos, R. J. 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>.
- Vieira-Neto, A., K. N. Galvão, W. W. Thatcher, and J. E. P. Santos. 2017. Association among gestation length and health, production, and reproduction in Holstein cows and implications for their offspring. *J. Dairy Sci.* 100:3166–3181. <https://doi.org/10.3168/jds.2016-11867>.
- Visentin, G., A. McDermott, S. McParland, D. P. Berry, O. A. Kenny, A. Brodkorb, M. A. Fenelon, and M. De Marchi. 2015. Prediction of bovine milk technological traits from mid-infrared spectroscopy analysis in dairy cows. *J. Dairy Sci.* 98:6620–6629. <https://doi.org/10.3168/jds.2015-9323>.
- Wang, Q., and H. Bovenhuis. 2018. Genome-wide association study for milk infrared wavenumbers. *J. Dairy Sci.* 101:2260–2272. <https://doi.org/10.3168/jds.2017-13457>.
- 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>.
- Wang, Q., A. Hulzebosch, and H. Bovenhuis. 2016. Genetic and environmental variation in bovine milk infrared spectra. *J. Dairy Sci.* 99:6793–6803. <https://doi.org/10.3168/jds.2015-10488>.
- Wiggans, G. R., J. B. Cole, S. M. Hubbard, and T. S. Sonstegard. 2017. Genomic selection in dairy cattle: The USDA experience. *Annu. Rev. Anim. Biosci.* 5:309–327. <https://doi.org/10.1146/annurev-animal-021815-111422>.
- Zaalberg, R. M., L. Janss, and A. J. Buitenhuis. 2020. Genome-wide association study on Fourier transform infrared milk spectra for two Danish dairy cattle breeds. *BMC Genet.* 21:9. <https://doi.org/10.1186/s12863-020-0810-4>.
- Zou, K. H., A. J. O'Malley, and L. Mauri. 2007. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* 115:654–657. <https://doi.org/10.1161/CIRCULATIONAHA.105.594929>.

ORCIDS

Piush Khanal <https://orcid.org/0000-0003-1079-5102>
 Robert J. Tempelman <https://orcid.org/0000-0002-7833-6730>