



# Predicting blood $\beta$ -hydroxybutyrate using milk Fourier transform infrared spectrum, milk composition, and producer-reported variables with multiple linear regression, partial least squares regression, and artificial neural network

R. S. Pralle, K. W. Weigel, and H. M. White<sup>1</sup>

Department of Dairy Science, University of Wisconsin, Madison 53706

## ABSTRACT

Prediction of postpartum hyperketonemia (HYK) using Fourier transform infrared (FTIR) spectrometry analysis could be a practical diagnostic option for farms because these data are now available from routine milk analysis during Dairy Herd Improvement testing. The objectives of this study were to (1) develop and evaluate blood  $\beta$ -hydroxybutyrate (BHB) prediction models using multivariate linear regression (MLR), partial least squares regression (PLS), and artificial neural network (ANN) methods and (2) evaluate whether milk FTIR spectrum (mFTIR)-based models are improved with the inclusion of test-day variables (mTest; milk composition and producer-reported data). Paired blood and milk samples were collected from multiparous cows 5 to 18 d postpartum at 3 Wisconsin farms (3,629 observations from 1,013 cows). Blood BHB concentration was determined by a Precision Xtra meter (Abbot Diabetes Care, Alameda, CA), and milk samples were analyzed by a privately owned laboratory (AgSource, Menomonie, WI) for components and FTIR spectrum absorbance. Producer-recorded variables were extracted from farm management software. A blood BHB  $\geq 1.2$  mmol/L was considered HYK. The data set was divided into a training set ( $n = 3,020$ ) and an external testing set ( $n = 609$ ). Model fitting was implemented with JMP 12 (SAS Institute, Cary, NC). A 5-fold cross-validation was performed on the training data set for the MLR, PLS, and ANN prediction methods, with square root of blood BHB as the dependent variable. Each method was fitted using 3 combinations of variables: mFTIR, mTest, or mTest + mFTIR variables. Models were evaluated based on coefficient of determination, root mean squared error, and area under the

receiver operating characteristic curve. Four models (PLS-mTest + mFTIR, ANN-mFTIR, ANN-mTest, and ANN-mTest + mFTIR) were chosen for further evaluation in the testing set after fitting to the full training set. In the cross-validation analysis, model fit was greatest for ANN, followed by PLS and MLR. Diagnostic strength after cross-validation was poorest for MLR and was similar for ANN and PLS. Models that used mTest + mFTIR variables performed marginally better than models that used only mFTIR or mTest variables. These results suggest that blood BHB prediction models that use mFTIR + mTest variables may be useful additions to existing HYK diagnostic and management programs.

**Key words:** ketosis, transition cow, diagnostic tools, neural network

## INTRODUCTION

Hyperketonemia (HYK) is a metabolic disorder that impairs milk production, reproduction, and health outcomes in lactating dairy cows (Duffield, 2000; McArt et al., 2012). The postpartum prevalence of HYK ranges worldwide from 15 to 22%, although it can vary greatly between farms (Suthar et al., 2013; Santschi et al., 2016; Chandler et al., 2018). Measurement of blood BHB concentration early postpartum is the predominant diagnostic method, and HYK is commonly defined as blood BHB  $\geq 1.2$  mmol/L (Iwersen et al., 2009; McArt et al., 2012; Gordon et al., 2017). Milk ketone body concentrations have been considered for HYK diagnosis because they are correlated with blood concentrations (Marstorp et al., 1983; Andersson, 1984; Enjalbert et al., 2001). The most accurate methods for detecting milk ketone bodies are flow-injection analysis, GLC, and enzymatic assays; however, these techniques are expensive, time consuming, and difficult to automate. Fourier transform infrared (FTIR) spectrometry could provide a practical alternative for predicting concentrations of milk ketone bodies because it is already

Received November 1, 2017.

Accepted January 8, 2018.

<sup>1</sup>Corresponding author: heather.white@wisc.edu

used to evaluate milk composition in DHI milk testing programs (Rutten et al., 2009, 2011).

Studies in which milk FTIR data have been used to predict milk ketone body concentrations have reported moderate correlations between predicted and assayed milk concentrations (Hansen, 1999; Heuer et al., 2001; de Roos et al., 2007). However, correlation of FTIR-based milk ketone body concentration predictions with blood BHB concentration is low (Chandler et al., 2018), and use of FTIR-predicted milk ketone bodies for HYK diagnosis has produced tests with insufficient sensitivity to diagnose individual cows (van der Drift et al., 2012; Chandler et al., 2018). Marginal improvement of FTIR-predicted milk ketone bodies used as an HYK diagnostic was made by including FTIR-predicted milk ketone bodies with cow test-day information in logistic (van der Drift et al., 2012) and multiple linear regression (MLR; Chandler et al., 2018) models predicting blood BHB concentration. At this time, these predictions of blood BHB concentration are recommended as a tool for monitoring HYK prevalence at the herd level, but sensitivity and specificity are insufficient for use as an individual cow diagnostic tool (van der Drift et al., 2012; Chandler et al., 2018).

Inclusion of milk FTIR spectrum data, combined with FTIR-predicted milk ketone body concentrations and milk composition in models, could generate more robust models with enhanced prediction of blood BHB and HYK diagnosis by capturing residual spectrum predictive abilities. Furthermore, advancements in computing technology have made flexible and computationally demanding methods, such as machine learning, increasingly available for such applications. Artificial neural networks (ANN) are a type of machine-learning prediction method with the ability to self-learn relationships from labeled experimental data and generalize to unlabeled situations. This advanced modeling technique may allow optimization of blood BHB concentration prediction from milk FTIR spectrum, milk component, and producer-reported variables.

We hypothesized that the use of more advanced modeling techniques, such as partial least squares regression (PLS) and ANN, would result in improved predictions of HYK status from producer-reported and milk composition variables compared with MLR. Furthermore, we hypothesized that predictions of blood BHB could be optimized by inclusion of milk FTIR spectrum, milk component, and producer-reported variables. Based on these hypotheses, our objectives were to (1) develop and evaluate blood BHB prediction models using MLR, PLS, and ANN methods and (2) evaluate whether milk FTIR spectrum-based models are improved by the inclusion of milk composition and producer-reported variables.

## MATERIALS AND METHODS

Cows from 2 privately owned Holstein dairy farms in southern Wisconsin and the University of Wisconsin–Madison Emmons Blaine Dairy Cattle Research Center were enrolled in the study. Dairy farms were selected based on the following criteria: early-lactation cows grouped in a separate pen, availability of headlocks for blood sampling, capability of proportional milk sampling, use of management software, and willingness to participate. All animal use and handling protocols were approved by the University of Wisconsin–Madison College of Agricultural and Life Sciences Animal Care and Use Committee. Farms were visited twice a week for regular sampling of paired blood and milk samples from multiparous Holstein cows ( $n = 1,013$ ) between 5 and 18 DIM. This provided an opportunity to collect up to 4 paired blood and milk samples per cow within the 2-wk sampling period. Routine HYK testing within the first 10 DIM at the University of Wisconsin–Madison farm allowed some cows to contribute 5 samples as described by Rathbun et al. (2017).

### Sample Collection and Analysis

Morning milk samples were collected by an International Committee on Animal Recordings–approved sampling system using a proportional sampler that had been calibrated within the previous 12 mo. Animal identification numbers were recorded by automatic radio-frequency identification collection and verified by visual identification of animal identification tags to prevent inaccurate identification recording. Samples were preserved with 2-bromo-2-nitropropane-1,3-diol (Advanced Instruments Inc., Norwood, MA) and transported to AgSource Cooperative Services (Menomonie, WI) for analysis according to standard test-day procedures. In brief, all milk samples were preheated to 40°C and mixed before analysis of milk fat and milk protein by FTIR using the Foss MilkoScan FT+ (Foss Analytical, Hillerød, Denmark) in accordance with the instrument manufacturer's instructions and ISO 9622/IDF 141:2013 (AOAC official method 972.16; AOAC International, 2016). Analysis of SCC was performed using Fossomatic FC (Foss Analytical). Milk BHB and milk acetone concentrations were predicted by FTIR using Foss Ketolab (Foss Analytical) based on the calibrations of de Roos et al. (2007). Additionally, the Foss MilkoScan FT+ analysis of milk samples provided predictions for the proportions of SFA, UFA, *trans* fatty acids (FA), and short-, medium-, and long-chain FA in milk based on Foss FTIR FA prediction models (Foss Analytical). Per the DHI's standard operating procedures, milk samples were analyzed on equipment that

is calibrated weekly with 12 standards, and standards are rechecked daily and hourly with a subset of 6 of the 12 standards. Intra-assay coefficients of variation for all variables were maintained at <7%. Interassay coefficients of variation are not available for all variables; however, interassay coefficients of variation for fat and protein are maintained at <2 and <1.5%, respectively.

Immediately following the morning milk sampling, each cow was confined to a headlock along the feed alley before or during the morning feeding. At this time, a blood sample was collected from the coccygeal vessel into an evacuated Vacutainer serum collection tube containing clot activator (BD Diagnostics, Franklin Lakes, NJ). Whole-blood BHB concentration was quantified cowside with the hand-held Precision Xtra blood glucose and ketone meter (Abbott Diabetes Care, Alameda, CA). This meter has been validated for use in dairy cattle (Iwersen et al., 2009; Bach et al., 2016; Sailer et al., 2018). Cows were diagnosed with HYK for a given sampling if blood BHB was  $\geq 1.2$  mmol/L. Incidences of HYK were reported to farm staff for treatment per the respective farm's standard operating procedures.

### Data Collection

Milk sample composition was obtained from Ag-Source Cooperative Services (Menomonie, WI) and included variables routinely predicted from milk FTIR spectrum: protein percentage, fat percentage, fat:protein ratio, SNF percentage, SCC, MUN, FA categories (detailed above), acetone, and BHB. Prediction accuracy and precision of the preceding milk components were not evaluated in the present study; rather, variables were used as provided to represent data as available during routine analysis. The FTIR spectrum absorbance (1,060 wavenumbers per sample) was exported separate from the milk composition variables and provided by AgSource Cooperative Services. Producer-recorded variables regarding the previous and current lactation, which are routinely exported during milk test, were exported from DairyComp305 (Valley Ag Services, Tulare, CA) for all cows in the study. These variables included calving date, parity, DIM, age at first calving, previous lactation length, previous lactation 305-d mature equivalent milk, gestation length, and dry period length. A single observation comprised a blood BHB measurement and a paired milk sample composition and FTIR spectrum from a single sample collection day. For each observation, producer-reported variables listed above for each cow were aligned with the observational data. A total of 4,147 observations were collected, but some were excluded due to missing

blood BHB ( $n = 26$ ), failure to capture FTIR spectrum ( $n = 462$ ), missing producer-reported data ( $n = 25$ ), and misreported data ( $n = 5$ ), leaving 3,629 observations for further analysis.

### Model Fitting and Validation

All model fitting was performed with JMP Pro 12 (SAS Institute Inc., Cary, NC). Blood BHB concentration did not have a normal distribution, which is an assumption of MLR; therefore, the square root transformation of blood BHB concentration was the dependent variable for all models to better satisfy MLR assumptions and provide consistency. Modeling methods used in this study included MLR, PLS, and ANN. Each method was fitted with 3 different groups of potential explanatory variables: producer-reported and milk composition variables (**mTest**), FTIR spectrum absorbance variables (**mFTIR**), and the union of mTest and mFTIR variables (**mTest + mFTIR**). This resulted in 9 method-variable (e.g., MLR-mFTIR) combinations.

Observations within cow were considered independent and randomly assigned to a training set ( $n = 3,020$ ) and a testing set ( $n = 609$ ) before analysis. Observations in the training set were used to perform a 5-fold cross-validation for all 9 method-group combinations as described below. Data within the testing set were maintained separately and not used until external model validation.

### MLR

For the MLR method, an initial model was fitted to predict blood BHB from potential explanatory variables in mTest, mFTIR, and mTest + mFTIR using the stepwise fit function and minimum corrected Akaike information criterion stopping rule. When explanatory variables with  $P > 0.15$  were included in the model, the variable with the greatest  $P$ -value was removed and the model was refitted. This was repeated until all variables were significant at  $P \leq 0.15$ . Then, collinearity was evaluated for each variable using a variable inflation factor (**VIF**), and variables with  $VIF \geq 10$  were considered collinear. When collinearity was present, the variable with the greatest VIF was removed and the model was refitted. This process was repeated until all variables had a  $VIF < 10$ . Removing collinear variables can drastically alter the  $P$ -values of retained variables, and previously significant variables can become nonsignificant. In those instances, nonsignificant variables were removed successively, and once only significant variables remained the collinearity check was repeated.

### PLS Regression

An initial PLS model was fitted with potential explanatory variables in mTest, mFTIR, and mTest + mFTIR, and the model was fitted with 15 factors (latent variables). After the initial fitting, the model was refitted using variables with variable importance projection values  $>0.8$ , again with 15 factors.

### ANN

For ANN models, the neural predictive modeling function was implemented. A single hidden layer with 15 nodes and a tangent activation identity provided the framework for the ANN with the following specifications: least absolute values of residuals, weight decay penalty method, and 750 tours (epochs). When too many variables of low predictive ability are fitted with ANN, computing time increases, model robustness decreases, and the risk of overfitting increases. Therefore, only explanatory variables with variable importance projection values  $>0.8$  in the PLS analyses were included in the respective ANN models using mTest, mFTIR, and mTest + mFTIR explanatory variables.

### Cross-Validation and External Validation

In this study, cross-validation was performed by randomly dividing observations from the training data set into 5 subsets or folds (A through E), each with 604 observations. Each subset was used to develop MLR, PLS, and ANN models using mTest, mFTIR, and mTest + mFTIR variables independently. The resulting models were fitted to the other 4 subsets, in turn, and the following statistics were recorded: coefficient of determination ( $R^2$ ), root mean squared error (RMSE), and area under the receiver operating characteristic curve (AUC). Of the 9 models evaluated by cross-validation, 4 models were chosen for further evaluation based on  $R^2$ , RMSE, and AUC. These models were fitted to the full training data set and validated externally using the aforementioned testing set. Statistics used to evaluate model fit in the external validation included the concordance correlation coefficient (CCC) and mean squared error (MSE) of predicted blood BHB concentrations (Tedeschi, 2006). In addition, accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of predicted HYK diagnoses were evaluated for each model using the optimal diagnostic thresholds that were determined as the predicted BHB value that maximized the difference between sensitivity and  $1 - \text{specificity}$ .

## RESULTS AND DISCUSSION

In the full data set, a total of 24, 97, 182, 685, and 25 individual cows contributed 1, 2, 3, 4, and 5 samples, respectively. The proportion of samples diagnosed with HYK was 12.3, 12.1, and 13.6% for the full data set ( $n = 3,629$ ), training set ( $n = 3,020$ ), and testing set ( $n = 609$ ), respectively. Mean blood BHB concentrations were 0.78, 0.77, and 0.81 for the full data set, training set, and testing set, respectively. Within specific folds of the training set, the proportion of samples diagnosed with HYK ranged from 10.8 to 13.4%, and mean blood BHB concentration ranged from 0.75 to 0.80 mmol/L (Table 1). The means and standard errors of producer-reported and milk composition variables are reported in Tables 2 and 3, respectively. It is likely that the proportion of HYK samples and BHB concentration was reduced by the inclusion of multiple samples from individual healthy cows as well as the inclusion of samples from HYK cows collected after treatment.

### Model Development and Cross-Validation in the Training Set

**Prediction Methods.** Relative fit of models was the greatest for ANN, with mean  $R^2$  of  $0.44 \pm 0.01$ ,  $0.36 \pm 0.01$ , and  $0.27 \pm 0.02$  for ANN, PLS, and MLR methods, respectively, when averaged across the mTest, mFTIR, and mTest + mFTIR groups of explanatory variables (Table 4). Similarly, relative performance based on average RMSE was  $0.16 \pm 0.003$ ,  $0.18 \pm 0.003$ , and  $0.27 \pm 0.06$  for ANN, PLS, and MLR, respectively. To our knowledge, no previous studies have investigated the use of advanced statistical modeling, such as PLS and ANN, to develop blood BHB prediction models. In the current study, diagnostic ability of the ANN and PLS methods was similar, with an average AUC of  $0.85 \pm 0.007$ , whereas MLR had poorer diagnostic ability, with an average AUC of  $0.81 \pm 0.012$ .

**Explanatory Variables.** When averaged across methods, models that used mTest + mFTIR potential explanatory variables provided the greatest  $R^2$  ( $0.39 \pm 0.02$ ) as shown in Table 4. However, average RMSE for models with mTest + mFTIR explanatory variables was similar to that of models with mTest explanatory variables, and average AUC values were similar for models with mTest, mFTIR, or mTest + mFTIR explanatory variables (Table 4). Models that used only mFTIR variables had the least favorable fit statistics, with a mean  $R^2$  of  $0.33 \pm 0.02$  and RMSE of  $0.27 \pm 0.061$ . Improvement of  $R^2$  by using mTest + mFTIR explanatory variables is expected because the inclusion

**Table 1.** Blood BHB concentration, proportion of samples diagnosed with hyperketonemia (HYK), and proportion of samples from each farm for the training set and cross-validation subsets used for model development and the testing set used for external validation

Variable	Cross-validation subset <sup>1</sup>					Training set	Testing set
	A	B	C	D	E		
Sample size, no.	604	604	604	604	604	3,020	609
Blood BHB, mmol/L							
Mean <sup>2</sup>	0.8	0.8	0.8	0.8	0.8	0.77	0.81
Minimum	0.1	0.1	0.1	0.2	0.2	0.1	0.1
Maximum	4.7	3.9	4.5	3.6	3.8	4.7	5.4
HYK, %	11	13	12	13	12	12.12	13.63
Farm, <sup>3</sup> %							
X	51.7	48.0	48.7	46.2	48.5	48.6	47.5
Y	41.4	43.7	44.2	46.7	45.4	44.3	45.8
Z	6.9	8.3	7.1	7.1	6.1	7.1	6.7

<sup>1</sup>Training set observations were randomly assigned to 1 of 5 data subsets (A through E), which were used to perform 5-fold cross-validation during model development.

<sup>2</sup>BHB standard error  $\pm 0.02$  mmol/L.

<sup>3</sup>Proportion of samples contributed from each farm, anonymized as X, Y, and Z.

of more variables generally improves relative fit. The poorer fit of mFTIR variables compared with mTest + mFTIR and mTest variables should be interpreted with caution. Model fit statistics are noticeably poorer for the MLR models with mFTIR variables compared with all other models (Table 4). This decreased the mean performance of models based on mFTIR variables. When comparing PLS and ANN models fit with mFTIR variables ( $0.38 \pm 0.02$ ,  $0.17 \pm 0.004$ , and  $0.86 \pm 0.006$  for  $R^2$ , RMSE, and AUC, respectively) with those fit with mTest variables ( $0.38 \pm 0.02$ ,  $0.17 \pm 0.004$ , and  $0.84 \pm 0.008$  for  $R^2$ , RMSE, and AUC, respectively) or mTest + mFTIR variables ( $0.43 \pm 0.02$ ,  $0.16 \pm 0.003$ , and  $0.86 \pm 0.008$  for  $R^2$ , RMSE, and AUC, respectively), mean performance of mFTIR variables was similar to mTest variables. Both mFTIR and mTest variables were marginally poorer than mTest + mFTIR variables (Table 4). This study supports and

expands upon the work of Chandler et al. (2018), who reported improvements in correlation, AUC, and accuracy of MLR models by incorporating additional milk composition and producer-reported variables into models based on individual FTIR milk ketone body tests. Previously, van der Drift et al. (2012) also proposed utilizing producer-reported and milk composition variables to augment milk FTIR-based predictions of milk ketone bodies when predicting blood BHB and diagnosing HYK. Neither of these previous studies explored use of mFTIR data in prediction models. Results from this study support mFTIR variables being able to explain a marginal amount of additional variation in models already containing mTest data.

**Evaluation of Method–Variable Group Combination.** A summary of the  $R^2$ , RMSE, and AUC of all 9 prediction method–variable group combinations is provided in Table 4. Overall, the model with the best

**Table 2.** Sample means and SE of producer-reported variables extracted from DairyComp305 (Valley Agricultural Software, Tulare, CA) for the training set and cross-validation subsets used for model development and the testing set used for external validation

Variable	Cross-validation subset <sup>1</sup>					SE	Training set	SE	Testing set	SE
	A	B	C	D	E					
DIM	11.5	11.1	11.8	11.3	11.5	0.2	11.5	0.1	11.7	0.2
Parity	2.9	2.9	3.0	3.0	2.9	0.1	3.0	0.0	3.0	0.1
AFC, <sup>2</sup> mo	23.0	23.2	23.2	23.2	23.1	0.1	23.1	0.0	23.4	0.2
PLL, <sup>3</sup> d	338.7	337.6	339.2	337.5	333.5	2.1	337.3	1.0	336.3	2.3
PME305, <sup>4</sup> kg	15,778	15,774	15,739	15,836	15,657	94	15,757	43	15,608	96
Gestation, d	276.8	277.0	277.2	277.0	276.8	0.7	277.0	0.3	277.7	0.5
Dry period, d	56.3	55.3	56.0	56.4	56.2	0.5	56.0	0.2	55.9	0.5

<sup>1</sup>Training set observations were randomly assigned to 1 of 5 data subsets (A through E), which were used to perform a 5-fold cross-validation during model development.

<sup>2</sup>Age at first calving.

<sup>3</sup>Previous lactation length.

<sup>4</sup>Previous 305-d mature equivalent milk.

**Table 3.** Sample means and SE of milk composition variables provided by AgSource Cooperative Services (Menomonie, WI) for data sets and cross-validation subsets used for model fitting and evaluation

Variable <sup>1</sup>	Cross-validation subset <sup>2</sup>					SE	Training set	SE	Testing set	SE
	A	B	C	D	E					
Yield, kg	41.16	39.93	40.32	39.58	39.21	0.54	40.05	0.23	40.10	0.54
Fat, %	4.31	4.43	4.31	4.35	4.34	0.04	4.35	0.02	4.33	0.04
Protein, %	3.43	3.45	3.41	3.46	3.42	0.02	3.43	0.01	3.40	0.02
Fat:protein	1.27	1.30	1.28	1.27	1.28	0.01	1.28	0.01	1.29	0.01
Lactose, %	4.76	4.73	4.74	4.74	4.74	0.01	4.74	0.004	4.75	0.01
SNF, %	9.25	9.25	9.22	9.26	9.24	0.02	9.24	0.01	9.21	0.02
SCC, ×10 <sup>3</sup> cells/mL	217.13	208.47	198.02	223.16	218.75	26.73	213.11	13.85	154.43	15.92
MUN, mg/dL	11.05	11.51	11.45	11.23	11.37	0.11	11.32	0.05	11.51	0.10
Acetone	0.09	0.09	0.09	0.09	0.09	0.004	0.09	0.002	0.10	0.01
BHB	0.08	0.08	0.08	0.08	0.08	0.003	0.08	0.001	0.08	0.003
Fatty acids, <sup>3</sup> g/100 g of milk										
MUFA	1.35	1.41	1.38	1.38	1.37	0.02	1.38	0.01	1.37	0.02
PUFA	0.30	0.30	0.30	0.30	0.30	0.002	0.30	0.001	0.30	0.002
SFA	2.58	2.64	2.55	2.60	2.59	0.03	2.59	0.01	2.59	0.03
SCFA	0.40	0.41	0.39	0.40	0.40	0.01	0.40	0.002	0.40	0.01
MCFA	1.86	1.92	1.85	1.88	1.86	0.03	1.87	0.01	1.85	0.02
LCFA	1.67	1.74	1.71	1.70	1.69	0.02	1.70	0.01	1.69	0.02

<sup>1</sup>Milk composition variables were measured in composite milk samples preserved with 2-bromo-2-nitropropane-1,3-diol (Advanced Instruments Inc., Norwood, MA) collected during the a.m. milking. Milk fat percentage, protein percentage, lactose percentage, SNF percentage, and MUN were determined by the Foss MilkoScan FT+ (Foss Analytical, Hillerød, Denmark). Somatic cell count was determined by the Fossomatic FC (Foss Analytical). Acetone and BHB concentrations were determined using Foss Ketolab (Foss Analytical). Milk fatty acid groups were determined by Foss MilkoScan FT+ (Foss Analytical).

<sup>2</sup>Training set observations were randomly assigned to 1 of 5 data subsets (A through E), which were used to perform a 5-fold cross-validation during model development.

<sup>3</sup>SCFA = short-chain fatty acids; MCFA = medium-chain fatty acids; LCFA = long-chain fatty acids.

fit was ANN–mTest + mFTIR, with R<sup>2</sup>, RMSE, and AUC of 0.46 ± 0.01, 0.16 ± 0.002, and 0.86 ± 0.005, respectively. However, RMSE and AUC of ANN–mTest + mFTIR were improved only marginally relative to ANN–mTest (0.17 ± 0.001 and 0.85 ± 0.004, respectively), ANN–mFTIR (0.17 ± 0.002 and 0.86 ± 0.004, respectively), and PLS–mTest + mFTIR (0.17 ± 0.002

and 0.86 ± 0.005, respectively). Based on the differences in predictive ability between methods and groups of explanatory variables in the cross-validation analysis, 4 models (ANN–mTest + mFTIR, ANN–mTest, ANN–mFTIR, and PLS–mTest + mFTIR) were chosen for external validation in the testing set. Variables included in these 4 externally validated models are

**Table 4.** Fit statistic means and SE for 5-fold cross-validation of blood BHB prediction models in the training set during model development

Method <sup>1</sup>	Variable <sup>2</sup>	R <sup>2</sup>	SE	RMSE <sup>3</sup>	SE	AUC <sup>4</sup>	SE
MLR	Averaged across variables	0.27	0.02	0.27	0.060	0.81	0.012
PLS	Averaged across variables	0.36	0.01	0.18	0.003	0.85	0.007
ANN	Averaged across variables	0.44	0.01	0.16	0.003	0.85	0.005
Averaged across methods	mFTIR	0.33	0.02	0.27	0.061	0.84	0.012
Averaged across methods	mTest	0.35	0.02	0.18	0.003	0.83	0.007
Averaged across methods	mTest + mFTIR	0.39	0.02	0.18	0.003	0.83	0.010
MLR	mFTIR	0.23	0.02	0.45	0.072	0.81	0.014
MLR	mTest	0.29	0.01	0.19	0.001	0.82	0.006
MLR	mTest + mFTIR	0.31	0.01	0.18	0.002	0.82	0.008
PLS	mFTIR	0.34	0.01	0.18	0.002	0.86	0.004
PLS	mTest	0.33	0.01	0.18	0.002	0.84	0.006
PLS	mTest + mFTIR	0.40	0.01	0.17	0.002	0.86	0.005
ANN	mFTIR	0.42	0.01	0.17	0.002	0.86	0.004
ANN	mTest	0.43	0.01	0.17	0.001	0.85	0.004
ANN	mTest + mFTIR	0.46	0.01	0.16	0.002	0.86	0.005

<sup>1</sup>Model prediction methodology: artificial neural network (ANN), partial least squares (PLS), and multiple linear regression (MLR).

<sup>2</sup>Potential explanatory variable group: producer-reported and milk composition variables (mTest), milk Fourier transform infrared spectrum absorbance variables (mFTIR), and all collected variables (mTest + mFTIR).

<sup>3</sup>Root mean squared error.

<sup>4</sup>Area under the curve of the receiver operating characteristic curve.

provided in Supplemental Table S1 (<https://doi.org/10.3168/jds.2017-14076>).

**External Validation of the Testing Set.** Overall, the ANN-mTest + mFTIR model provided greater fit in the independent testing set, with  $R^2$ , RMSE, and CCC of 0.56, 0.16, and 0.69, respectively. The other model-explanatory variable combinations, namely PLS-mTest + mFTIR, ANN-mFTIR, and ANN-mTest, performed similarly in the testing set based on  $R^2$ , RMSE, and CCC (Table 5; Figure 1). Decomposition of MSE revealed minimal mean and slope bias for all models, with 99.0, 98.9, 98.3, and 99.0% of MSE attributed to error for PLS-mTest + mFTIR, ANN-mFTIR, ANN-mTest, and ANN-mTest + mFTIR, respectively (Table 5). Model AUC values were 0.88, 0.83, 0.87, and 0.88 for PLS-mTest + mFTIR, ANN-mFTIR, ANN-mTest, and ANN-mTest + mFTIR, respectively (Table 6). The PLS-mTest + mFTIR method-variable combination had the lowest accuracy (0.73) of HYK diagnosis due to low specificity (0.71), but it provided the greatest sensitivity (0.89). Meanwhile, ANN-mTest + mFTIR and ANN-mTest had equal accuracy (0.80), although ANN-mTest + mFTIR had marginally greater sensitivity, specificity, and NPV (Table 6). Despite the lowest AUC (0.83), ANN-mFTIR had intermediate accuracy, sensitivity, specificity, PPV, and NPV values of 0.78, 0.76, 0.78, 0.35, and 0.95, respectively, compared with other model-variable combinations.

Comparison of the PLS-mTest + mFTIR and ANN-mTest + mFTIR to determine whether the added complexity of the ANN method was advantageous indicated that ANN marginally improved fit statistics for blood BHB prediction. In addition, ANN improved accuracy, specificity, and PPV of HYK diagnosis compared with PLS-mTest + mFTIR (Table 6). Based on the models

developed herein, it appears that the added complexity of the ANN method may improve HYK diagnosis compared with PLS-mTest + mFTIR. However, the improvement in goodness of fit statistics for ANN compared with PLS may be marginal. Inclusion of mTest + mFTIR explanatory variables in ANN models provided marginal improvements in fit and diagnostic statistics, although ANN-mTest + mFTIR was most preferable and ANN-mFTIR was least preferable (Tables 5 and 6). This indicates that ANN with milk composition and producer-reported variables are somewhat better than ANN with milk FTIR spectrum for prediction of blood BHB, and fitting ANN with all explanatory variables seems to confer marginal improvement. This could be due to residual effects within the milk FTIR data not captured by mTest variables, which include FTIR-predicted milk ketone body concentrations. However, it is important to note that some of the between-samples variation in milk FTIR spectrum may be captured in the milk composition variables represented in the mTest data. For example, milk-based predictions of the concentrations of acetone, SFA, or UFA may explain some of the variation present in the milk FTIR data, and this could diminish the difference in prediction accuracy between models that were developed from mTest explanatory variables and those that were developed with mTest + mFTIR explanatory variables. The ANN-mTest model performance was only marginally reduced compared with ANN-mTest + mFTIR and was slightly improved compared with ANN-mFTIR, which may indicate that the addition of FTIR spectrum data is not essential. Use of mTest data may also be a more practical solution from a data management perspective because mTest models would not require the additional capture of FTIR spectrum

**Table 5.** Model fit statistics for blood BHB prediction models in external validation using the testing set

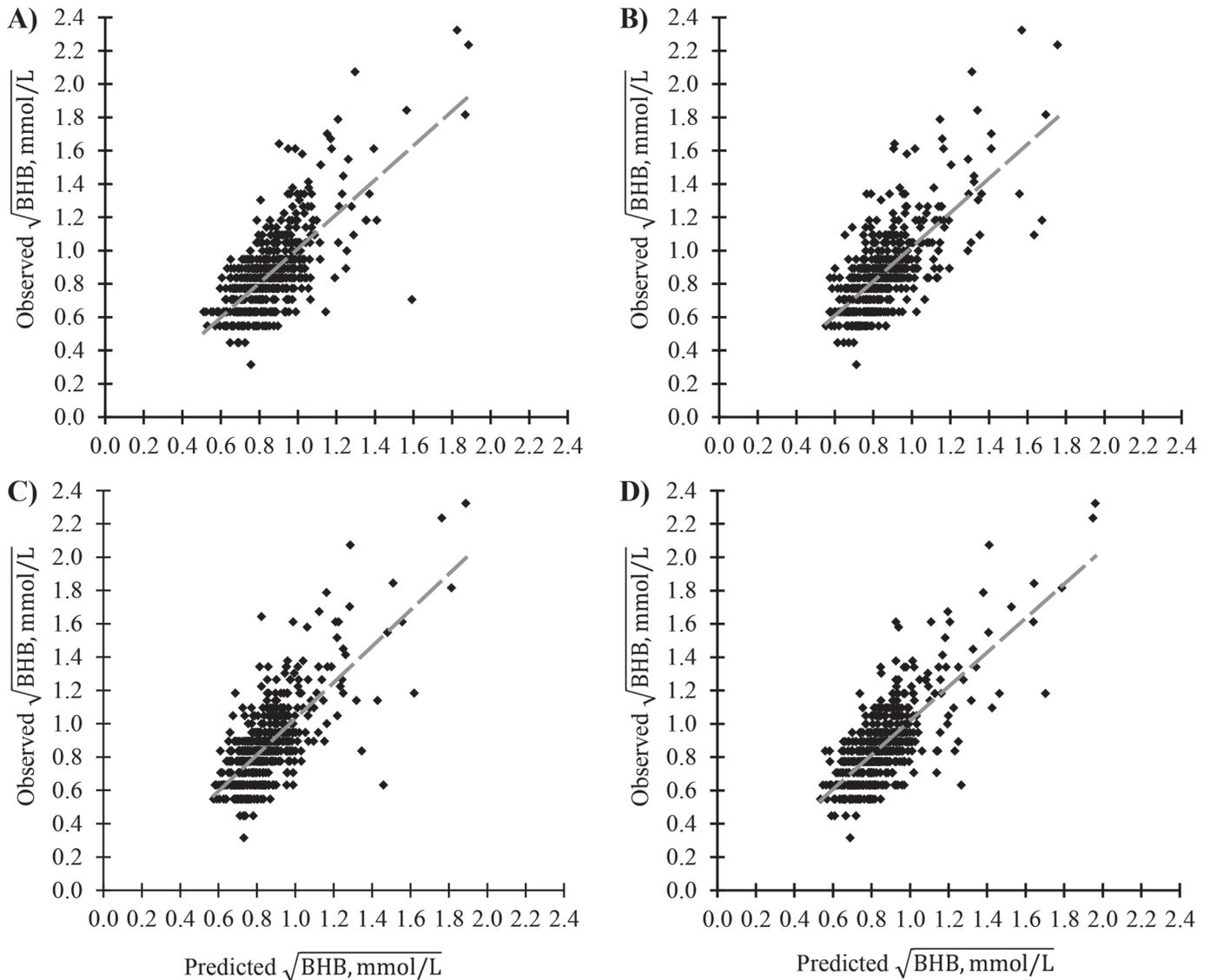
Fit statistic	Model <sup>1</sup>			
	PLS-mTest + mFTIR	ANN-mFTIR	ANN-mTest	ANN-mTest + mFTIR
$R^2$	0.50	0.50	0.51	0.56
RMSE <sup>2</sup>	0.17	0.17	0.17	0.16
CCC <sup>3</sup>	0.66	0.66	0.67	0.69
Mean bias	0.00004	0.00027	0.00029	0.00001
MSE decomp <sup>4</sup>				
Mean	0.0	1.0	1.0	0.0
Slope	1.0	0.1	0.7	1.0
Error	99.0	98.9	98.3	99.0

<sup>1</sup>Partial least squares regression with milk test and milk Fourier transform infrared spectrum absorbance variables (PLS-mTest + mFTIR), artificial neural network with Fourier transform infrared spectrum absorbance variables (ANN-mFTIR), artificial neural network with milk test variables (ANN-mTest), and artificial neural network with milk test and milk Fourier transform infrared spectrum absorbance variables (ANN-mTest + mFTIR).

<sup>2</sup>Root mean squared error.

<sup>3</sup>Concordance correlation coefficient.

<sup>4</sup>Mean squared error decomposition, %.



**Figure 1.** Observed versus predicted plots of the square root of BHB (mmol/L; best fit line = gray and dashed) for (A) partial least squares regression with milk test and milk Fourier transform infrared spectrum absorbance variables, (B) artificial neural network with Fourier transform infrared spectrum absorbance variables, (C) artificial neural network with milk test variables, and (D) artificial neural network with milk test and milk Fourier transform infrared spectrum absorbance variables from external validation in the testing set.

data and merging of the spectrum data with mTest data by cow. However, models fitted only with milk FTIR spectrum may be more reliable and ubiquitous for application in the field than models including producer-reported variables. Producer-recorded variables rely heavily on input from farm staff: they must use herd management software regularly, record potential explanatory variables accurately, and allow the milk testing service provider to access these data freely. Precision and accuracy of these data are also likely to be variable across dairy farms. Conversely, models based solely on milk FTIR spectrum do not require farms to record or provide management data and can be read-

ily implemented by analysis of proportionally collected milk samples, but are dependent on the type of equipment and quality control of sample analysis at the milk testing laboratory.

### **Prediction Models as an HYK Diagnosis Strategy**

Diagnosis of HYK on privately owned dairy operations can be hampered by the expense and labor associated with cowside and enzymatic tests. Predictions of blood BHB using data retrieved from weekly or monthly milk samples and recorded herd management data could provide a high-throughput, efficient, and cost-effective

**Table 6.** Accuracy, sensitivity, specificity, and predictive values of blood BHB prediction models for hyperketonemia<sup>1</sup> diagnosis in external validation using the testing set

Diagnostic statistic <sup>2</sup>	Model <sup>3</sup>			
	PLS-mTest + mFTIR	ANN-mFTIR	ANN-mTest	ANN-mTest + mFTIR
AUC <sup>4</sup>	0.88	0.83	0.87	0.88
Accuracy <sup>5</sup>	0.73	0.78	0.80	0.80
Sensitivity	0.89	0.76	0.80	0.83
Specificity	0.71	0.78	0.80	0.81
Positive predictive value	0.32	0.35	0.39	0.38
Negative predictive value	0.98	0.95	0.96	0.97

<sup>1</sup>Hyperketonemia diagnosed as blood BHB concentration  $\geq 1.2$  mmol/L as quantified cowside by the Precision Xtra meter (Abbott Diabetes Care, Alameda, CA).

<sup>2</sup>Predicted blood BHB threshold for hyperketonemia diagnosis was determined as the value that maximized the formula sensitivity – (1 – specificity).

<sup>3</sup>Partial least squares regression with milk test and milk Fourier transform infrared spectrum absorbance variables (PLS-mTest + mFTIR), artificial neural network with Fourier transform infrared spectrum absorbance variables (ANN-mFTIR), artificial neural network with milk test variables (ANN-mTest), and artificial neural network with milk test and milk Fourier transform infrared spectrum absorbance variables (ANN-mTest + mFTIR).

<sup>4</sup>Area under the curve of the receiver operating characteristic curve.

<sup>5</sup>Calculated as the proportion of correctly identified observations among all observations tested.

means of diagnosis and monitoring HYK. A potential limitation of implementing blood BHB prediction models using milk FTIR variables may be the frequency of routine DHI milk testing, which is typically every 4 to 5 wk. At this frequency, half of the cows would not have a single milk sample collected during the high HYK risk period (5–18 DIM; McArt et al., 2012), and the remaining cows would have only 1 milk sample during the high risk period from which to predict blood BHB and HYK risk. It may be feasible for some large dairy farms to implement weekly milk sampling of cows in early lactation (5–18 DIM), especially if these cows are grouped separately. Frequent sampling of these cows would allow prediction of individual cow HYK based on FTIR data, with or without producer-reported variables, as an effective diagnostic and monitoring tool. This tool could be particularly valuable if other diagnostics (e.g., SCC) were provided for these early-lactation cows.

All models validated externally in the testing set (PLS-mTest + mFTIR, ANN-mFTIR, ANN-mTest, and ANN-mTest + mFTIR) had “very good” diagnostic capacity (AUC  $\geq 0.83$ ; Šimundić, 2009) for HYK; however, no model achieved the “excellent” diagnostic strength that is typical with ideal reference tests (AUC  $\geq 0.90$ ; Šimundić, 2009). Furthermore, none achieved the sensitivity and specificity of currently available cowside blood tests (88 and 96%, respectively; Iwersen et al., 2009). The relatively low PPV values (ranging from 0.32 to 0.39; Table 6) imply the potential for misdiagnosis of some healthy cows as HYK, which could result in unnecessary treatments and increased costs. Alternatively, the cost-effectiveness and labor efficiency of milk-based prediction models for HYK diagnosis may justify the reduced diagnostic strength compared with

enzymatic and cowside assays, especially when the cost of an untreated HYK case is much greater than the cost of HYK treatment (McArt et al., 2014). Milk-based prediction models could also be used for HYK screening programs, with predicted positives undergoing more precise cowside or enzymatic tests. In this instance, models with lower specificity and PPV may be of less concern because there would be further testing, and the primary goal of screening would be to identify all cows that are potentially HYK positive.

## CONCLUSIONS

Blood BHB prediction models developed from producer-reported variables, milk composition, and milk FTIR spectrum may be adequate for HYK diagnosis and monitoring due to reasonable diagnostic strength and efficient implementation. Overall, ANN provided the most favorable fit statistics and diagnostic performance, although improvement over PLS was marginal. Inclusion of mTest + mFTIR variables in prediction models improved HYK diagnostic strength slightly compared with mTest or mFTIR variables separately, but there may be technical advantages to predicting HYK status solely from mFTIR or mTest variables without merging these data sets by cow.

## ACKNOWLEDGMENTS

This research is based on work that was supported by the National Institute of Food and Agriculture, USDA Agriculture and Food Research Initiative Critical Agricultural Research and Extension (CARE; 2015-67028-23572) and Early Concept Grants for Exploratory Re-

search (EAGER; 2017-67007-25947). Additional funding was from the USDA Hatch grant number WIS01878 from the Wisconsin Agricultural Experiment Station (Madison). The authors recognize and appreciate the support of the owners and herdsman at the privately owned dairies as well as the research staff at the University of Wisconsin–Madison Emmons Blaine Dairy Research Center. Furthermore, this work would not have been possible without the assistance of the following individuals: F. M. Rathbun, R. C. Oliveira, C. A. Getschel, L. C. Resende, C. H. Nova, K. J. Sailer, and T. Mack (all individuals from University of Wisconsin–Madison).

## REFERENCES

- Andersson, L. 1984. Concentrations of blood and milk ketone bodies, blood isopropanol and plasma glucose in dairy cows in relation to the degree of hyperketonaemia and clinical signs. *Zentralbl. Veterinarmed.* A 31:683–693. <https://doi.org/10.1111/j.1439-0442.1984.tb01327.x>.
- AOAC International. 2016. *Official Methods of Analysis*. 20th ed. AOAC International, Gaithersburg, MD.
- Bach, K. D., W. Heuwieser, and J. A. A. McArt. 2016. Technical note: Comparison of 4 electronic handheld meters for diagnosing hyperketonemia in dairy cows. *J. Dairy Sci.* 99:9136–9142. <https://doi.org/10.3168/jds.2016-11077>.
- Chandler, T. L., R. S. Pralle, J. R. R. Dórea, S. E. Poock, G. R. Oetzel, R. H. Fourdraine, and H. M. White. 2018. Predicting hyperketonemia by logistic and linear regression using test-day milk and performance variables in early-lactation Holstein and Jersey cows. *J. Dairy Sci.* 101:2476–2491. <https://doi.org/10.3168/jds.2017-13209>.
- de Roos, A. P., H. Van Den Bijgaart, J. Hørlyk, and G. de Jong. 2007. Screening for subclinical ketosis in dairy cattle by Fourier transform infrared spectrometry. *J. Dairy Sci.* 90:1761–1766. <https://doi.org/10.3168/jds.2006-203>.
- Duffield, T. 2000. Subclinical ketosis in lactating dairy cattle. *Vet. Clin. North Am. Food Anim. Pract.* 16:231–253.
- Enjalbert, F., M. C. Nicot, C. Bayourthe, and R. Moncoulon. 2001. Ketone bodies in milk and blood of dairy cows: Relationship between concentrations and utilization for detection of subclinical ketosis. *J. Dairy Sci.* 84:583–589. [https://doi.org/10.3168/jds.S0022-0302\(01\)74511-0](https://doi.org/10.3168/jds.S0022-0302(01)74511-0).
- Gordon, J. L., T. F. Duffield, T. H. Herdt, D. F. Kelton, L. Neuder, and S. J. LeBlanc. 2017. Effects of a combination butaphosphan and cyanocobalamin product and insulin on ketosis resolution and milk production. *J. Dairy Sci.* 100:2954–2966. <https://doi.org/10.3168/jds.2016-11925>.
- Hansen, P. W. 1999. Screening of dairy cows for ketosis by use of infrared spectroscopy and multivariate calibration. *J. Dairy Sci.* 82:2005–2010. [https://doi.org/10.3168/jds.S0022-0302\(99\)75437-8](https://doi.org/10.3168/jds.S0022-0302(99)75437-8).
- Heuer, C., H. J. Luinge, E. Lutz, and Y. H. Schukken. 2001. Determination of acetone in cow milk by Fourier transform infrared spectroscopy for the detection of subclinical ketosis. *J. Dairy Sci.* 84:575–582. [https://doi.org/10.3168/jds.S0022-0302\(01\)74510-9](https://doi.org/10.3168/jds.S0022-0302(01)74510-9).
- Iwersen, M., U. Falkenberg, R. Voigtsberger, D. Forderung, and W. Heuwieser. 2009. Evaluation of an electronic cowside test to detect subclinical ketosis in dairy cows. *J. Dairy Sci.* 92:2618–2624. <https://doi.org/10.3168/jds.2008-1795>.
- Marstorp, P., T. Anfält, and L. Andersson. 1983. Determination of oxidized ketone bodies in milk by flow injection analysis. *J. Dairy Sci.* 149:281–289. [https://doi.org/10.1016/S0003-2670\(00\)83184-0](https://doi.org/10.1016/S0003-2670(00)83184-0).
- McArt, J. A. A., D. V. Nydam, and G. R. Oetzel. 2012. Epidemiology of subclinical ketosis in early lactation dairy cattle. *J. Dairy Sci.* 95:5056–5066. <https://doi.org/10.3168/jds.2012-5443>.
- McArt, J. A. A., D. V. Nydam, G. R. Oetzel, and C. L. Guard. 2014. An economic analysis of hyperketonemia testing and propylene glycol treatment strategies in early lactation dairy cattle. *Prev. Vet. Med.* 117:170–179. <https://doi.org/10.1016/j.prevetmed.2014.06.017>.
- Rathbun, F. M., R. S. Pralle, S. J. Bertics, L. E. Armentano, K. Cho, C. Do, K. A. Weigel, and H. M. White. 2017. Relationships between body condition score change, prior mid-lactation phenotypic residual feed intake, and hyperketonemia onset in transition dairy cows. *J. Dairy Sci.* 100:3685–3696.
- Rutten, M. J. M., H. Bovenhuis, J. M. L. Heck, and J. A. M. van Arendonk. 2011. Predicting bovine milk protein composition based on Fourier transform infrared spectra. *J. Dairy Sci.* 94:5683–5690. <https://doi.org/10.3168/jds.2011-4520>.
- Rutten, M. J. M., H. Bovenhuis, K. A. Hetinga, H. J. F. van Valenberg, and J. A. M. van Arendonk. 2009. Predicting bovine milk fat composition using infrared spectroscopy based on milk samples collected in winter and summer. *J. Dairy Sci.* 92:6202–6209. <https://doi.org/10.3168/jds.2009-2456>.
- Sailer, K. J., R. S. Pralle, R. C. Oliveira, G. R. Oetzel, and H. M. White. 2018. Technical note: Validation of BHBCheck blood  $\beta$ -hydroxybutyrate meter as a diagnostic tool for hyperketonemia. *J. Dairy Sci.* 101:1524–1529. <https://doi.org/10.3168/jds.2017-13583>.
- Santschi, D. E., R. Lacroix, J. Durocher, M. Duplessis, R. K. Moore, and D. M. Lefebvre. 2016. Prevalence of elevated milk  $\beta$ -hydroxybutyrate concentrations in Holstein cows measured by Fourier-transform infrared analysis in Dairy Herd Improvement milk samples and association with milk yield and components. *J. Dairy Sci.* 99:9263–9270. <https://doi.org/10.3168/jds.2016-11128>.
- Šimundić, A. M. 2009. Measures of diagnostic accuracy: Basic definitions. *EJIFCC* 19:203–211.
- Suthar, V. S., J. Canelas-Raposo, A. Deniz, and W. Heuwieser. 2013. Prevalence of subclinical ketosis and relationships with postpartum diseases in European dairy cows. *J. Dairy Sci.* 96:2925–2938. <https://doi.org/10.3168/jds.2012-6035>.
- Tedeschi, L. O. 2006. Assessment of the adequacy of mathematical models. *Agric. Syst.* 89:225–247. <https://doi.org/10.1016/j.agry.2005.11.004>.
- van der Drift, S. G. A., R. Jorritsma, J. T. Schonewille, H. M. Knijn, and J. A. Stegeman. 2012. Routine detection of hyperketonemia in dairy cows using Fourier transform infrared spectroscopy analysis of  $\beta$ -hydroxybutyrate and acetone in milk in combination with test-day information. *J. Dairy Sci.* 95:4886–4898. <https://doi.org/10.3168/jds.2011-4417>.