Predicting body condition score change in early lactating Irish dairy cows using milk mid infrared spectra

M. Frizzarin[1,2], I.C. Gormley,[2] D.P. Berry,[1] and S. McParland[1]

[1]*Teagasc, AGRIC, Moorepark, Fermoy P61 P302, Co. Cork, Ireland*
Corresponding Author: donagh.berry@teagasc.ie
[2]*School of Mathematics and Statistics, UCD, Ireland*

Abstract

Body condition score (BCS) is a subjective metric used worldwide to reflect the fat stored in mammals. BCS, as well as its change in early lactation, has been associated with productive, reproductive, and health traits. The objective of the present study was to predict, using different statistical techniques, BCS change (ΔBCS) in early lactation dairy cows using milk mid infrared (MIR) spectra. A total of 73,193 BCS records from 6,572 cows were collected from five research farms. To generate daily ΔBCS, splines with 6 knot points across days in milk (DIM) were fitted through individual test-day records of BCS. Daily BCS was interpolated from the splines and used to calculate daily ΔBCS. ΔBCS observations were merged with MIR spectra recorded on the same week. Data in the first 120 DIM were retained. Three statistical methods were used to predict ΔBCS from the spectra; partial least squares regression (PLSR), generalized additive mixed model (GAMM), and neural networks (NN). Spectra and DIM were used as predictors in NN and PLSR, while the first 20 principal components of the spectra and a spline fitted through DIM were used as predictors in GAMM. Tuning parameters of PLSR were determined using 10 fold cross-validation. The NN model had two hidden layers and a Bayesian regularization applied to the input layer. To compare predictive ability across the approaches, the dataset was divided in 4 sub-datasets, and iteratively 3 sub-datasets were used to train the methods, while the remaining sub-dataset (the test dataset) was used to test the methods. Prediction accuracy was evaluated according to the root mean square error of the test dataset (RMSEV; here multiplied by 1000) and the correlation (r) between the actual and the predicted ΔBCS. The RMSEV and r obtained from the four test datasets were averaged. Body condition score change was predicted with an average RMSEV of 1.02 BCS units (SD=0.010) and r of 0.87 (SD=0.004) from NN; the SD of actual ΔBCS was $2.05*10^{-3}$ BCS units. Partial least squares regression performed better than GAMM with an average RMSEV of 1.06 (SD=0.010) and 1.10 (SD=0.010) BCS units, respectively, and an r of 0.86 (SD=0.004) and 0.84 (SD=0.004) for PLSR and GAMM, respectively. Results from the present study demonstrate the potential to use milk MIR spectra to predict ΔBCS, which can be used to support farm decisions and can be incorporated in dairy cow breeding programs.

Keywords: Mid-infrared spectroscopy, body condition score change, early lactation, machine learning.

Introduction

Body condition score (BCS) is a metric used worldwide to reflect fat stored in mammals (Roche *et al.*, 2009). In dairy cows, BCS tends to be lost after parturition and it is generally restored before a new calving (Bauman and Carrie, 1980). Body condition score change is used on farms as an indicator of cow energy balance, and its trajectory in early lactation is generally used to support fertility decisions. In fact, BCS, as well as its change in early lactation, have been associated with fertility and health traits (Berry *et al.*, 2007; Buckley *et al.*, 2003; Roche *et al.*, 2007). Buckley *et al.* (2003), for example, reported that the mean BCS at 60 and 100 days of lactation was positively associated with both submission for breeding in the first 21 days of the breeding season and the likelihood of pregnancy after 42 days of breeding. Butler and Smith (1989) documented the effect of the change of body condition score in early lactation on the rate of conception at first serving, reporting a conception rate of just 17% to first service in cows that lost more than 1 BCS unit (5-point scale) after calving, while cows that lost <1 BCS unit had a conception rate of 53% to first service. Body condition score, as well as its change (ΔBCS), are therefore useful metrics for producers. Body condition score change can be calculated from multiple BCS records albeit these may not always be available.

Mid infrared spectroscopy of milk is a fast, cheap, and non-disruptive technique to generate spectra from milk samples (De March *et al.*, 2013). The spectra are produced from the interaction between the light emitted from the spectrometer and the molecules within the milk and, utilizing suitable statistical techniques, it is possible to predict milk and animal traits from the spectra (Soyeurt *et al.*, 2006; McParland *et al.*, 2014). Milk MIR spectra are used to predict milk fatty acids with good accuracy (Soyeurt *et al.*, 2006) and, as reported by Nogalski *et al.* (2012), cows that lost >1 BCS unit in early lactation had a different milk fatty acid profile compared to cows that lost < 1 BCS units. Thus there is a strong biological hypothesis as to why ΔBCS could potentially be predicted from milk MIR spectra. McParland *et al.* (2014) pioneered the research on predicting DBCS from milk spectra in dairy cows, reporting a correlation between the actual and the predicted ΔBCS ranging from 0.57 to 0.75 using different spectra combination and partial least squares regression (PLSR) as the prediction method.

The objective of the present study was to predict DBCS using morning milk spectra in early lactating dairy cows and to compare predictions from PLSR with those from generalized additive mixed model (GAMM) and neural networks (NN). Predicting ΔBCS from milk MIR spectra can routinely provide producers with DBCS estimates, which can be used in making reproductive decisions, as well as providing phenotypes for use in dairy cow breeding programs.

Materials and methods

Data and data editing

A total of 73,193 BCS records from 6,572 cows were recorded in 5 Teagasc research farms between 2015 and 2019. Body condition score was assessed by trained scorers using a 5-point scale with increment of 0.25 (Edmonson *et al.*, 1989). BCS was recorded every 18 days, with a repeatability of BCS within a 7-day period being 0.69. Cows were from different parity orders (from parity 1 to 12) and different breeds (Holstein-Friesian, Jersey, Norwegian Red, as well as their crosses). Daily BCS for each lactation was calculated after fitting a cubic spline with 6 knot points at 20, 70, 120, 170, 220, and 270 days in milk (DIM) through individual test-day records of BCS with a covariance structure fitted among knot points (McParland *et al.*, 2014). Daily ΔBCS was calculated from the fitted splines as the BCS on a given day minus the BCS of the previous day. To avoid potential problems with interpolation, only DBCS calculated on the day of an actual BCS records were retained. Lactations greater than 10 were deleted, and only DBCS records between DIM 5 and 120 were retained. Parities were grouped as first, second, third, fourth, and fifth or greater, and DBCS outliers were deleted for each

parity group as DBCS records greater than 3 standard deviations (SD) from the mean ∆BCS of the respective parity group.

From the same cows and across the same time period, 423,532 individual milk spectra were collected from morning milking. Spectra outliers were deleted following the same procedure as described by Frizzarin *et al.* (2021a), and were standardized across time using the standardization coefficients provided as part of the European Milk Recording project ring testing program (Grelet *et al.*, 2015; Grelet *et al.*, 2017). Body condition score change records were merged with the closest spectra recorded within one week. The final dataset comprised 13,492 ∆BCS records across 2,489 lactations from 1,250 cows.

Prediction methods

All the analyses were conducted using the statistical software R (R core team). Three different prediction methods were used to predict ∆BCS: PLSR, GAMM, and NN. For the PLSR analyses, the a.m. spectra as well as a fourth order polynomial of DIM were used as predictor variables. The R package pls (Mevik *et al.*, 2019) was used to develop the prediction equation. The number of PLSR factors were defined using 10 folds cross-validation (CV). For the GAMM analyses, the first 20 principal components of the a.m. spectra were used as linear predictor variables with DIM fitted as a spline; cow-lactation was accounted as repeated measures in the model. The R package gamm4 (Wood and Scheipl, 2020) was used to develop the prediction equation. Lastly, a Bayesian regularized NN was developed using the a.m. spectra and the DIM as predictor variables. The R package brnn (Perez Rodrigez and Gianola, 2020) was used to develop the prediction equation, and the default tuning parameters were chosen, which included two hidden layers and a Bayesian regularization to the input layer to improve generalizability.

Measures of prediction performance

To test the predictive performance of the methods on unseen data, the original dataset was divided into four sub-datasets, three of these were used to train the prediction equation (i.e., training dataset), and the fourth sub-dataset was used to test the equation (i.e., test dataset). This process was repeated until all the four sub-datasets were considered as test dataset once.

Different metrics were used to evaluate prediction performance, such as the root mean square error on the test dataset (RMSEV), the correlation between the predicted ∆BCS values in the test dataset and the actual ∆BCS values (r), the bias of the prediction, the slope, and the ratio of performance to interquartile distance (RPIQ). The bias corresponds to the average of the residuals, the slope corresponds to the slope of the line where the real ∆BCS values are plotted against the predicted ∆BCS values, and the RPIQ is the ratio between the interquartile range of the observed trait values and the RMSEV. All these metrics were calculated for each of the test datasets, and were subsequently averaged. The SD of the performance metrics across the four test datasets was also calculated and was considered as a reflection of robustness of the prediction method. The F-test was used to compare the RMSEV across prediction methods.

Results

The prediction performance of the different prediction methods on the test datasets used to predict ΔBCS are summarized in Table 1. Neural networks had the lowest RMSEV (P < 0.05) as well as the greatest r and RPIQ; thus, for this dataset, NN was the prediction method which predicted ΔBCS with greatest accuracy. Partial least squares regression had a slightly greater (P < 0.05) RMSEV compared to NN. Generalized additive mixed model was the prediction method with the poorest prediction performance (P < 0.05). For all prediction methods, the SD of the RMSEV and r was low, indicating that the prediction performance were robust across all the test datasets.

In Figure 1 is presented the average actual ΔBCS across DIM, as well as the average predicted values from PLSR, GAMM, and NN. Both GAMM and NN predictions perfectly followed the actual ΔBCS trend, while PLSR slightly over predicted the actual ΔBCS between week 3 and week 8 of lactation, and slightly under predicted actual ΔBCS between week 8 and week 12 of lactation.
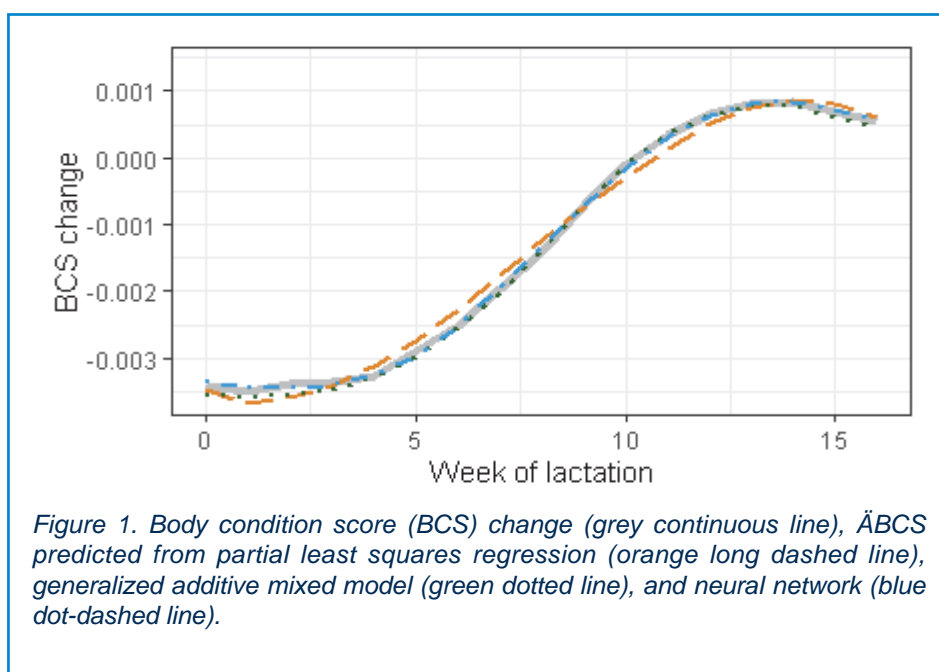
Table 1. Prediction performance[1] on the test dataset of partial least squares regression (PLSR), generalized additive mixed model (GAMM), and neural network (NN) to predict body condition score change across 120 days in milk (DIM).

Method	RMSEV[2,3] (SD[2,3])	Bias[2] (SD[2])	r (SD)	Slope (SE)	RPIQ (SD)
PLSR	1.06 (0.010)	0.00 (0.042)	0.86 (0.004)	1.00 (0.010)	3.16 (0.11)
GAMM	1.10 (0.010)	0.04 (0.057)	0.84 (0.004)	0.99 (0.010)	3.04 (0.11)
NN	1.02 (0.010)	0.00 (0.040)	0.87 (0.004)	0.99 (0.010)	3.27 (0.10)

[1] RMSEV = root mean square error in validation data set; r = Pearson correlation between the actual observed value of the trait and predicted value of the trait; SD = standard deviation; SE = standard error.
[2] Values presented are values *1,000
[3] BCS units

Figure 1. Body condition score (BCS) change (grey continuous line), ÄBCS predicted from partial least squares regression (orange long dashed line), generalized additive mixed model (green dotted line), and neural network (blue dot-dashed line).

Body condition score, as well as ∆BCS, have been related to both fertility and health traits in lactating dairy cows (Berry *et al.*, 2007; Buckley *et al.*, 2003; Roche *et al.*, 2007). Body condition score change is used by producers as an indicator of cow energy status, and, as demonstrated in the present study, it can be predicted from the MIR milk spectra.

One of the main challenge in predicting ∆BCS is accounting for scorer subjectivity. In fact, BCS evaluation, even if executed by trained scorers, is a subjective measure. Roche *et al.* (2004) studied the variability of international scoring systems, and reported a strong correlation between the scoring systems across countries. Moreover, in the same study, Roche *et al.* (2004) reported that the accuracy of the BCS assessed depended also on whether the assessment was just visual or a combination of visual plus tactile appraisal of the cow. This suggests that BCS assessment is somewhat scorer-dependent and may vary according to the scale and the method (visual or tactile appraisal) used by the scorer. In the present study, BCS was assessed by trained scorers using both visual and tactile appraisal. Using ∆BCS as opposed to BCS in the development of the prediction should negate the impact of mean score differences among scorers. As a further strategy to minimize the impact of the BCS uncertainty on the development of prediction methods, only ∆BCS predicted from the splines on the same day of an actual BCS record were retained. Nonetheless, considering the uncertainty of the trait being predicted, perfect or almost perfect prediction performance should not be expected.

McParland *et al.* (2014) previously predicted ∆BCS across 305 day of lactation in 1,018 lactating dairy cows from milk MIR using PLSR; they reported a correlation between the actual and the predicted ∆BCS of 0.75. Machine learning approaches have sometimes been shown to slightly improve the accuracy of predicting traits from milk MIR when compared to predictions developed using PLSR (Frizzarin *et al.*, 2021b; Soyeurt *et al.*, 2020). Partial least squares regression has the potential to be an interpretable statistical method, but assumes linearity between the trait and the latent variables of the spectra (Wold *et al.*, 2001). Generalized additive mixed model had low variability in the ∆BCS predictions, and NN is less generalizable than simpler models, requires more tuning parameters identification, and requires large data sets. Therefore, it is important to consider the advantages and disadvantages of the different statistical methods when deciding which method to use for the final spectra predictions.

While the present study related individual cow milk MIR to ∆BCS, the prediction models could also possibly extend to the routinely taken herd bulk tank samples to assess mean herd energy status. Nevertheless, herd averages could hide individual cows mobilizing considerable body condition. Whereas, the present study was based on milk MIR samples taken weekly, this is usually not always practical in commercial farms. Nonetheless, technologies are being developed to routinely assess milk samples using different regions of the spectrum. These wavelengths are simply overtones of the mid-infrared region; hence, it could therefore possible to re-derive prediction equations for ∆BCS using data from different regions of the spectrum. Body condition score change can be used by the farmers to take specific animal decisions, like diet requirements, or whether inseminate a cow or not. Lastly, the prediction of ∆BCS after milk recording permits to generate a large amount of phenotypes which can be included in breeding programs (e.g., as correlated traits in a multi-trait genetic evaluation).

Discussion

The results from the present study demonstrated the potential of using MIR spectra to predict ∆BCS in early lactating dairy cows. Body condition score change was predicted with a correlation between the actual and the predicted ∆BCS of 0.87 using NN and spectra obtained during morning milking. The prediction of ∆BCS from MIR spectra

Conclusion

can be useful to obtain information on individual cow energy status, as well as routinely generate a phenotype for using in genetic evaluations.

Acknowledgments

References

Bauman, D. E., and W. B. Currie. 1980. Partitioning of nutrients during pregnancy and lactation: A review of mechanisms involving homeostasis and homeorhesis. J. Dairy Sci. 63:1514–1529.

Berry, D. P., J. M. Lee, K. A. Macdonald, K. Stafford, L. Matthews, and J. R. Roche. 2007. Associations between body condition score, body weight, somatic cell count, and clinical mastitis in seasonally calving dairy cattle. J. Dairy Sci. 90:637–648.

Buckley, F., K. O'sullivan, J.F. Mee, R.D. Evans, and P. Dillon. 2003. Relationships among milk yield, body condition, cow weight, and reproduction in spring-calved Holstein-Friesians. J. Dairy Sci. 86(7), 2308-2319.

Butler, W. R., and R. D. Smith. 1989. Interrelationships between energy balance and postpartum reproductive function in dairy cattle. J. Dairy Sci. 72:767-783.

De Marchi M.D., Toffanin V., Cassandro M. and Penasa M. 2014. Invited review: mid-infrared spectroscopy as phenotyping tool for milk traits. J. Dairy Sci. 97, 1–16.

Edmonson, A. J., I. J. Lean, L. D. Weaver, T. Farver, and G. Webster. 1989. A body condition scoring chart for Holstein dairy cows. J. Dairy Sci. 72:68–78.

Frizzarin, M., O'Callaghan, T.F., Murphy, T.B., Hennessy, D. and Casa, A., 2021a. Application of machine-learning methods to milk mid-infrared spectra for discrimination of cow milk from pasture or total mixed ration diets. J. Dairy Sci., 104(12), pp.12394-12402.

Frizzarin M., I.C. Gormley, D.P. Berry, T.B. Murphy, A. Casa, A. Lynch, and S. McParland, 2021b. Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods. J. Dairy Sci. 2021, 104, 7438–7447.

Grelet, C., J.F. Pierna, P. Dardenne, V. Baeten, and F. Dehareng, 2015. Standardization of milk mid-infrared spectra from a European dairy network. J. Dairy Sci, 98(4), pp.2150-2160.

Grelet, C., J.F. Pierna, P. Dardenne, H. Soyeurt, A. Vanlierde, F. Colinet, C. Bastin, N. Gengler, V. Baeten, and F. Dehareng, 2017. Standardization of milk mid-infrared spectrometers for the transfer and use of multiple models. J. Dairy Sci., 100(10), pp.7910-7921.

McParland, S., E. Lewis, E. Kennedy, S. G. Moore, B. McCarthy, M. O'Donovan, S. T. Butler, J. Pryce, and D.P. Berry. 2014. Mid-infrared spectrometry of milk as a predictor of energy intake and efficiency in lactating dairy cows. J. Dairy Sci 97(9):5863-5871.

Mevik B.H, R. Wehrens and K. H. Liland. 2019. pls: Partial Least Squares and Principal Component Regression. R package version 2.7-2.

Nogalski, Z., Wrotski, M., Sobczuk-Szul, M., Mochol, M., and Pogorzelska, P. 2012. The Effect of Body Energy Reserve Mobilization on the Fatty Acid Profile of Milk in High-yielding Cows. Asian-Australasian Journal of Animal Sciences. Asian Australasian Association of Animal Production Societies.

Perez Rodriguez P. and D. Gianola. 2020. brnn: Bayesian Regularization for Feed-Forward Neural Networks. R package version 0.8.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Roche, J. R., P. G. Dillon, C. R. Stockdale, L. H. Baumgard, and M.J. VanBaale. 2004. Relationships among international body condition scoring systems. J. Dairy Sci. 87:3076–3079

Roche, J. R., K. A. Macdonald, C. R. Burke, J. M. Lee, and D. P. Berry. 2007. Associations among body condition score, body weight, and reproductive performance in seasonal-calving dairy cattle. J. Dairy Sci. 90:376–391.

Roche, J. R., N. C. Friggens, J. K. Kay, M. W. Fisher, K. J. Stafford, and D. P. Berry. 2009. Invited review: body condition score and its association with dairy cow productivity, health, and welfare. J. Dairy Sci. 92:5769–5801.

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(9):3690-3695.

Soyeurt H., C. Grelet, S. McParland, M. Calmels, M. Coffey, A. Tedde, P. Delhez, F. Dehareng, and N. Gengler. A comparison of 4 different machine learning algorithms to predict lactoferrin content in bovine milk from mid-infrared spectra. J. Dairy Sci 103.12 (2020): 11585-11596.

Wold, S., M. Sjostrom, and L. Eriksson. 2001. PLS-Regression: A basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58:109–130.

Wood S. and F. Scheipl. 2020. gamm4: Generalized Additive Mixed Models using 'mgcv' and 'lme4'. R package version 0.2-6.