

39. Predicting nitrogen use efficiency of individual dairy cows by mid-infrared spectra

R. Shi^{1,2}, W. Lou^{1,2}, B. Ducro², A. van der Linden², H.A. Mulder², S.J. Oosting², S. Li¹ and Y. Wang^{1*}

¹China Agricultural University, Beijing 100193, China; ²Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; wangyachun@cau.edu.cn

Abstract

Dairy farming may have negative impacts on the environment, such as ammonia emissions to air and nitrate leaching to water. By selecting more efficient cows, the amount of nitrogen excreted per cow and consequently the adverse environmental impacts will be reduced. The aim of this study was to predict nitrogen use efficiency (NUE) of individual dairy cows using mid-infrared (MIR) spectra of milk. A total of 600 feeding and MIR records of 56 Holstein cows were collected from a farm in Beijing, China. NUE was calculated as the ratio of nitrogen in milk to nitrogen intake. The coefficient of determination of the best model was 0.69, 0.62 and 0.70 for NUE, nitrogen loss and dry matter intake, respectively. The MIR wavenumbers around 981.00 and 1506.93 cm^{-1} and daily milk yield were the most important variables for prediction. These results show potential for large-scale genetic evaluations of nitrogen efficiency.

Introduction

As widely emphasised, nitrogen surpluses at dairy farms have negative impacts on the environment, such as ammonia emission and nitrate leaching to ground and surface water. Meanwhile, the increasing demand for animal products will result in higher production and thus results in more intensive dairy farming with higher total emissions of nitrogen (N) if N losses are not controlled. Therefore, a more efficient, and environmentally friendly production system is needed, in which NUE of dairy cows plays a key role. If efficient cows are selected, the fraction of intake that is not ending up in faeces or urine will be lower, which will contribute to lower N losses to the environment. Generally, this kind of resource allocation trait is difficult to measure individually (Lopez-Villalobos *et al.*, 2018). For instance, it requires assessment of daily feed intake, which is costly. To genetically improve NUE, routinely recording and cost-effective proxies are essential to initiate genetic evaluations. Recently, Grelet *et al.* (2020) obtained reasonable proxies for N related traits such as NUE and nitrogen loss (NL) by including MIR spectra of dairy cows in their prediction models. A maximum coefficient of determination (R^2) of 0.82 was observed in their validations, which indicated the proxies were reasonable for further genetic analysis.

However, these prediction models were only based on records from Holstein cows in early lactation. Individual NUE and NL in other lactation stages have not been predicted yet with MIR data. Literature also indicated that non-informative signals may exist in original MIR data, which will decrease the relation between phenotypes and MIR spectra (Bresolin and Dórea, 2020). Pre-treated MIR data may be beneficial for constructing robust prediction models. Therefore, different pre-processing methods were assessed in the current study. Additionally, it is of interest to quantify the contribution of different input variables. The objectives of this study were (1) to develop the best prediction model of NUE, NL and dry matter intake (DMI) for individual dairy cows in China, (2) to compare different prediction models of these traits, and (3) to investigate the importance of input variables in the prediction.

Materials & methods

A feeding trial was conducted in an experimental farm in Beijing, China between March 2017 and August 2017. A total of 56 Chinese Holstein cows in mid and late lactation stage (154-452 days in milk (DIM)) were offered the same diet during the experimental period in a free-stall design barn. Daily feed intake of

individual cows was recorded by an automatic system (Roughage Intake Control System, Insentec B.V., Marknesse, the Netherlands). Daily DMI was calculated for each cow based on feed intake and dry matter content of the diet. Each cow was evaluated monthly for body condition scores (BCS) by two technicians. Additionally, the daily milk yield for each cow was recorded by the milking system.

Individual MIR spectra were obtained from Fourier transform spectrometers (Bentley, Chaska, USA) at the Beijing Dairy Cattle Centre. Fat, lactose, milk urea nitrogen (MUN) and total protein of milk samples were estimated from the MIR analysis.

NUE was defined as ratio of total N output in milk to total N intake from feed and NL was defined as total N intake from feed minus total N output in milk (Calsamiglia *et al.*, 2010). Two pre-processing methods were applied to the original MIR data to reduce the influence of noise in the MIR spectra, namely multiplicative scatter correction (MSC) and standard normal variate (SNV). Subsequently, wavenumbers induced by water were omitted, resulting in 215 data points for each record (Grelet *et al.*, 2016). The MIR spectra were then combined with feeding trial data, providing 600 records of 56 cows.

The selected MIR spectra, parity (1 and 2 or more), corresponding daily milk yield, BCS, DIM_g (actual DIM grouped by 5 days) and additional parameters generated by MIR analysis (protein, fat, lactose and MUN) were included for model development. MIR spectra were included in prediction either as original spectra or as result from pre-processing (MSC-spectra or SNV-spectra).

Furthermore, three different algorithms were applied for prediction: partial least squares (PLS), ridge regression (RR), and support vector machine (SVM) regression. Consequently, a total of 48 models were applied for predicting each trait (Table 1).

To validate the models, the dataset was randomly split in test and training sets in a ratio of 1 to 3, and a cow can only be in the test set or in the training set. Prediction models were constructed using the training sets and validated using the test sets. The performance metrics included R², root mean square error (RMSE), relative error (RE, calculated as RMSE/mean of the global data), and ratio of standard deviation of the global data to RMSE from the validation (RPD). Besides, the Spearman correlation coefficient (SpearR) between true values and predictions was included to test re-ranking of the best model. Importance of each input variable for the PLS and RR methods was defined as the absolute value of the corresponding regression coefficient.

Table 1. Prediction models for nitrogen use efficiency, nitrogen loss and dry matter intake.

Models	Predictors ¹	Number of input variables	Spectra	Algorithms	Count ²
Model 1	MIR	215	Original, MSC, SNV	PLS, RR, SVM	9
Model 2	+ Milk yield, parity	217			9
Model 3	+ BCS	218			9
Model 4	+ DIM_g	219			9
Model 5	+ MUN	220			9
Model 6	+ Protein, fat, lactose	8	No MIR		3

¹ For model 2 to 5, the additional predictor of next model is based on previous model, and model 6 includes all additional predictors, except for MIR spectra.

² Number of prediction models developed using this set of predictors: 3 algorithms times 3 types of MIR spectra for models 1-5 = 9 models. See main text for abbreviations.

Results

Individual daily NUE was on average 26.4%, with a standard deviation of 8.6%, NL was on average 0.51 kg/day, with a standard deviation of 0.15 kg/day, and DMI was on average 25.15 kg/day, with a standard deviation of 5.95 kg/day.

Model 6 was the least accurate for NL and DMI regardless of the performance metrics (Figure 1). Model 2 generated the best prediction for NUE and DMI, with a relatively high R^2 and RPD, as well as a small RE. Model 1 had a better application to NL, from which the highest R^2 and RPD, as well as lowest RE were obtained. For most models with MIR, the pre-processing step resulted on average in a 12.3% higher R^2 and 8.6% higher RPD, in combination with a 11.2% smaller RE on average.

Model 2 using the SVM method, and model 1 and model 2 using the PLS method were the best models for predicting NUE, NL and DMI, respectively (Table 2). SNV-spectra were included in these models, and the best model for NL only included MIR spectra.

The most important variables were MIR wavenumbers 973.54 to 988.46 cm^{-1} , 1,480.80 to 1,574.07 cm^{-1} , and daily milk yield based on feature score.

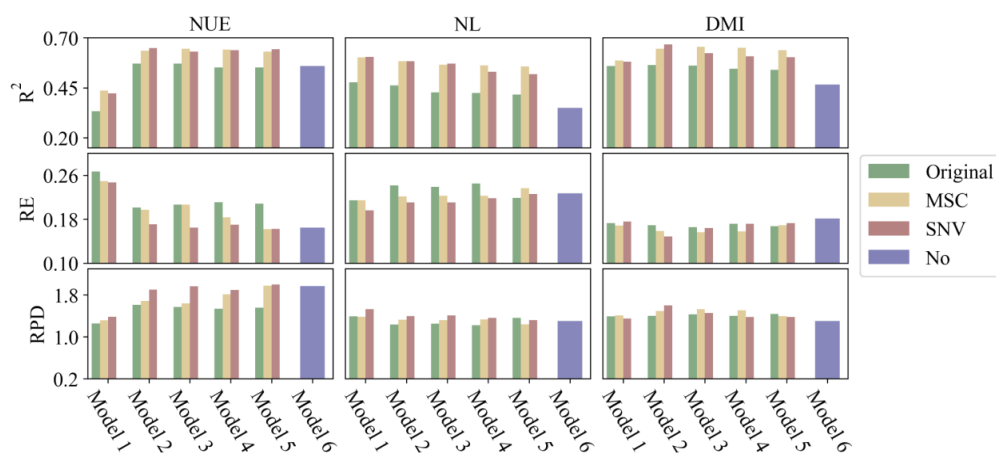


Figure 1. Statistics of different prediction models for nitrogen use efficiency, nitrogen loss and dry matter intake.

Table 2. Performance of best model for each trait¹.

Trait	Model information	R^2	RE	RPD	SpearR
NUE	Model 2, SNV-spectra, SVM	0.69	0.15	2.11	0.84
NL	Model 1, SNV-spectra, PLS	0.62	0.20	1.46	0.79
DMI	Model 2, SNV-spectra, PLS	0.70	0.14	1.70	0.81

¹ See main text for abbreviations.

Discussion

The current study developed accurate prediction models for daily NUE, NL and DMI of individual cows by comparing different prediction algorithms and precorrection methods for MIR data. Furthermore, the contribution of input variables was evaluated. The performance metrics of the best models for these three traits were comparable with those in the study of Grelet *et al.* (2020), which reported a range of R^2 from 0.59 to 0.68, RE from 0.14 to 0.23, and RPD from 1.57 to 2.07. It should be noted that the equations in Grelet *et al.* (2020) were based on three farms in three countries and therefore they might be more robust than prediction equations based on one farm. As long as feeding regimes are very similar on other farms, the prediction equations may facilitate genetic evaluation on a larger data set of MIR spectra. For instance, Chen *et al.* (2021) applied the prediction models of Grelet *et al.* (2020) to a large dataset for genetic evaluation of predicted NUE and NL in early lactation.

In this study, adding MIR spectra in the best models increased the R^2 and RPD by 10 to 30%, as well as reduced the RE by 0.03 to 0.10 compared to using model 6 (Figure 1). Meanwhile, this improvement was more obvious when MIR spectra were pre-treated. These results indicate that MIR includes additional information for better prediction of NUE, NL and DMI. Pre-treated MIR spectra were better than raw MIR spectra for developing accurate prediction models. Milk yield was an important variable in all models because it is highly correlated with N output (phenotypic correlation=0.96). The specific wave range of MIR that was important in prediction comprised C-H, C-N and N-N stretching (Grelet *et al.*, 2015; Xia *et al.*, 2018), which is related to proteins.

The results of current study indicate that the methodology is suitable to exploit MIR data for prediction of NUE, NL and DMI as reflected by a relatively high R^2 . The resulting predictions can be applied to improve nutrition and management of dairy herds in China and to perform further genetic evaluations.

Acknowledgements

The China Scholarship Council (No. 201913043) is acknowledged for funding the PhD projects of Rui Shi and Wenqi Lou, and the Beijing Dairy Cattle Centre is acknowledged for providing data. This research is supported by China Agriculture Research System of MOF and MARA.

References

- Bresolin T., and Dórea J.R.R. (2020) *Front. Genet.* 11:923. <https://doi.org/10.3389/fgene.2020.00923>
- Calsamiglia S., Ferret A., Reynolds C.K., Kristensen N.B., and van Vuuren A.M. (2010) *Animal* 4:1184–1196. <https://doi.org/10.1017/S1751731110000911>
- Chen Y., Vanderick S., Mota R.R., Grelet C., and Gengler N. (2021) *J. Dairy Sci.* 104:4413–4423. <https://doi.org/10.3168/jds.2020-18849>
- Grelet C., Fernández Pierna J.A., Dardenne P., Baeten V., and Dehareng F. (2015) *J. Dairy Sci.* 98:2150–2160. <https://doi.org/10.3168/jds.2014-8764>
- Grelet C., Bastin C., Gelé M., Davière J.-B., and Johan M. *et al.* (2016) *J. Dairy Sci.* 99:4816–4825. <https://doi.org/10.3168/jds.2015-10477>
- Grelet C., Froidmont E., Foldager L., Salavati M., and Hostens M. *et al.* (2020) *J. Dairy Sci.* 103:4435–4445. <https://doi.org/10.3168/jds.2019-17910>
- Lopez-Villalobos N., Correa-Luna M., Burke J.L., Sneddon N.W., and Schutz M.M. *et al.* (2018) *Proc. of NZSAP 2018*, Lincoln, New Zealand.
- Xia Y., Ugarte C.M., Guan K., Pentrak M., and Wander M.M. (2018) *Soil Sci. Soc. Am. J.* 82:1415–1427. <https://doi.org/10.2136/sssaj2018.05.0175>