Random forest regression for estimating dry matter intake of grazing dairy cows

L. Leso[1], J. Werner[2], D. McSweeney[3], E. Kennedy[3], A. Geoghegan[3] and L. Shalloo[3]

[1]*Department of Agriculture, Food, Environment and Forestry, University of Florence, 50145 Firenze, Italy;* [2]*Animal Nutrition and Rangeland Management in the Tropics and Subtropics, University of Hohenheim, 70599 Stuttgart, Germany;* [3]*Teagasc, Animal & Grassland Research and Innovation Centre, Moorepark, Fermoy Co. Cork, P61 C997, Ireland*

Corresponding author: lorenzo.leso@unifi.it

Abstract

The understanding of individual feed intake of grazing cows is essential for monitoring the nutrient intake of the animal, to calculate feed efficiency and increase animal and pasture productivity as well as tailored herd and pasture management. However, the measurement of individual herbage dry matter intake (hDMI) of individual cows is laborious and may be challenging in commercial grass-based dairy systems. The aim of this study was to assess the application potential of machine learning algorithms (random forest) to estimate the individual hDMI of grazing dairy cows in an intensive grazing management system using animal behaviour characteristics. This study was performed at a research farm on n = 41 animals with the established reference value for measuring feed intake based on the n-alkane technique. The database for model development included individual cow information, on-field grass measurements, grass quality as well as detailed individual cow behavioural characteristics based on the RumiWatchSystem. Random forest regression was used to predict hDMI. Recursive feature elimination (RFE) was used to select the best subset of predictors to be included in the model. To overcome issues associated with the relatively small sample size of n = 68 weekly values in total a nested cross-validation procedure was implemented. Results showed that RF has good potential for the prediction of hDMI in grazing dairy cattle. However, further studies are required to fully assess performance of this method and identify new potential predictors for hDMI.

Keywords: grazing, dairy cattle, dry matter intake, random forest, cross validation

Introduction

Pasture growth and utilisation have been reported as the main factors affecting profitability of pasture-based dairy farms (Hanrahan *et al.*, 2018). A better understanding of individual herbage dry matter intake (hDMI) at pasture has the potential to improve herd and grassland management. However, conventional methods for hDMI estimation, such as the n-alkane technique, are laborious and require specific equipment (Hellwing *et al.*, 2015). In recent years, PLF technologies such as behaviour sensors have become largely available to dairy producers and can provide fast and easy access to a wide range of information about both pasture and cows (Shalloo *et al.*, 2018). The aim of this study was to assess the application of a machine learning algorithm (random forest) for the estimation of individual hDMI of grazing dairy cows based on readily available data and information collected by PLF sensors.

Material and methods

The study was performed during the spring and summer of 2016 (27 March 2016 – 31 July 2016) at the Teagasc's research farm in Moorepark, Fermoy, Ireland. Forty-one spring calving dairy cows were included in the study, 24 of them were Holstein-Friesian, nine were Jersey and eight were crossbreeds. Ten primiparous cows and 31 multiparous cows were included, with an average parity of 2.57±1.41. All cows calved between 26 January

2016–28 February 2016. The body weight (BW) of cows included in the study was 470±65 kg while milk yield was 20.8±2.3 kg/cow/day.

During the experiment, cows grazed a perennial ryegrass-based pasture with no supplementary feeding. Cows were milked twice daily (at 7:00 and 14:30) and fresh pasture was offered after each milking. As the cows were involved in another on-going experiment regarding restricted pasture allocation (Werner *et al.*, 2019), a wide range of daily herbage allowances (DHA) was available. On average DHA was 13.6 kgDM/cow/day, ranging from 9.0–17.0 kgDM/cow/day (3.5 cm above the ground). Pre- and post-grazing compressed grass heights (PREGH and POSTGH) were measured daily using a rising plate meter (Jenquip, Fielding, New Zealand). The PREGH and POSTGH were 8.8±1.9 cm and 3.8±1.0 cm, respectively. Representative grass samples from each paddock were collected weekly and sent to an external lab for analysis. Results showed a crude protein (CP) content of 21.4±1.7%, 23.3±3.1% acid detergent fibre (ADF), 40.1±2.4% neutral detergent fibre (NDF), 26.3±2.7% non-forage carbohydrates (NFC) and 11.84±0.17 MJ/kgDM metabolizable energy (ME).

The experiment consisted of two observation periods, which lasted five weeks during spring and one week during summer. During observations, cows were fitted with the RumiWatchSystem (Itin+Hoch GmbH, Liestal, Switzerland), which consisted of a noseband sensor and a pedometer. This system was capable of monitoring cows' grazing behaviour, rumination and physical activity in a detailed and accurate manner (Werner *et al.*, 2018). Behaviour parameters measured by the RumiWatchSystem and retained in the final dataset are reported and described in Table 1.

Table 1. Behaviour parameters measured by the RumiWatchSystem and retained in final model for analysis

Parameter (unit of measure)	RumiWatch Output	Mean±SD
Ruminating chews frequency (n/min)	CHEWSPERMINUTE	49.5±5.8
Ruminating bouts started (n/day)	RUMIBOUTSTART	13.6±2.0
Ruminating time (min/day)	RUMINATETIME	456±65
Grazing bites (n/day)[1]	EATBITE	33,413±4,818
Feeding bite frequency (n/min)[2]	EAT1CHEW/EAT1TIME	74.8±3.6
Grazing bite frequency (n/min)[1]	EATBITE/EATTIME	54.6±5.3
Grazing bouts (n/day)	GRAZINGSTART	8.0±1.6
Total feeding time (min/day)	EATTIME	611±51
Feeding time with head down (min/day)	EAT1TIME	512±56
Feeding time with head up (min/day)	EAT2TIME	99±33
Other activities time (min/day)	OTHERACTIVITYTIME	368±83
Other chews (n/day)	OTHERCHEW	1,186±401
Walking time (min/day)	WALKTIME	95±20
Laying bouts (n/day)	LAYDOWN	7.7±1.7
Laying time (min/day)	LAYTIME	517±83

[1]Grazing bites are defined as prehension bites or jaw movements for ripping the grass; [2]Feeding bites are defined as all the feeding jaw movements taken with head down

Individual hDMI was estimated with the n-alkane technique (Dillon & Stakelum, 1989). Cows were dosed with a C32 n-alkane marker. Faecal samples were collected twice daily in the paddocks, before both am and pm milking and analysed in a lab. Estimation of hDMI was based on the amount of marker found in faeces and yielded weekly average values for each individual. As the n-alkane technique has demonstrated reliable estimates of individual intake (Wright *et al.*, 2019), values obtained with this method were used as the reference hDMI.

Data preparation and analysis

A dataset containing all possible predictors and reference hDMI was built. As the reference hDMI was only available as weekly means, all other data originally recorded in different time frames were converted to obtain weekly values. Initially, 103 variables were acquired which were then filtered to improve model prediction performance. Twenty-one descriptors were excluded for having the same value for more than 95% of the dataset. Fifty-two redundant variables were filtered off until no remaining pair-wise correlations exceeded $r_p = 0.9$ (Pearson correlation). As RF cannot deal with missing values, incomplete observations had to be removed as well. The final dataset consisted of 68 cow-week observations (rows) and 30 variables (columns) including cow characteristics (cow ID, breed, BW, parity, DIM), grass field measurements (DHA, PREGH and POSTGH), grass lab analysis (NDF, ADF, NFC, CP, ME), cows' behaviour data (CHEWSPERMINUTE, RUMIBOUTSTART, RUMINATETIME, EATBITE, EAT1CHEW/EAT1TIME, EATBITE/EATTIME, GRAZINGSTART, EATTIME, EAT1TIME, EAT2TIME, OTHERACTIVITYTIME, OTHERCHEW, WALKTIME, LAYDOWN, LAYTIME; Table 1) and hDMI.

Data analysis was performed using the open source R statistical software (R Core Team, 2018). Random forest (RF) regression was used to predict hDMI. The RF model is a machine learning method for classification and regression analysis that uses an ensemble of randomised decision trees to define its output (Breiman, 2001). To overcome issues associated with the relatively small sample size (n. 68), avoid overfitting and obtain a robust evaluation of model performance, a nested cross-validation (CV) procedure was implemented (Figure 1). This method consisted of an inner and outer repeated CV loop analysis, each with three folds and ten repetitions. The inner CV was used to tune model parameters, subset predictors and select the best model. Then, generalisation error was estimated by averaging the test set scores over several dataset splits in the outer CV loop. In the inner CV, recursive feature elimination was used to select the best subsets of predictors to be included in the models. Variable importance was computed for any variable included in the models selected in the inner CV and described the deterioration in prediction accuracy of the model when excluding that particular descriptor. Root-mean-square error (RMSE) and explained variance were used to evaluate model performance in both the inner and outer CV loops.

Results and discussion

The hDMI of cows included in the experiment, as estimated with the n-alkane method, was 16.35±4.4 kgDM/cow/day. The RF models produced were capable of explaining 64.1±8.5% of the variance in hDMI, with an average root-mean-square error (RMSE) of 2.66±0.37 kgDM/cow/day (Outer CV; Table 2).

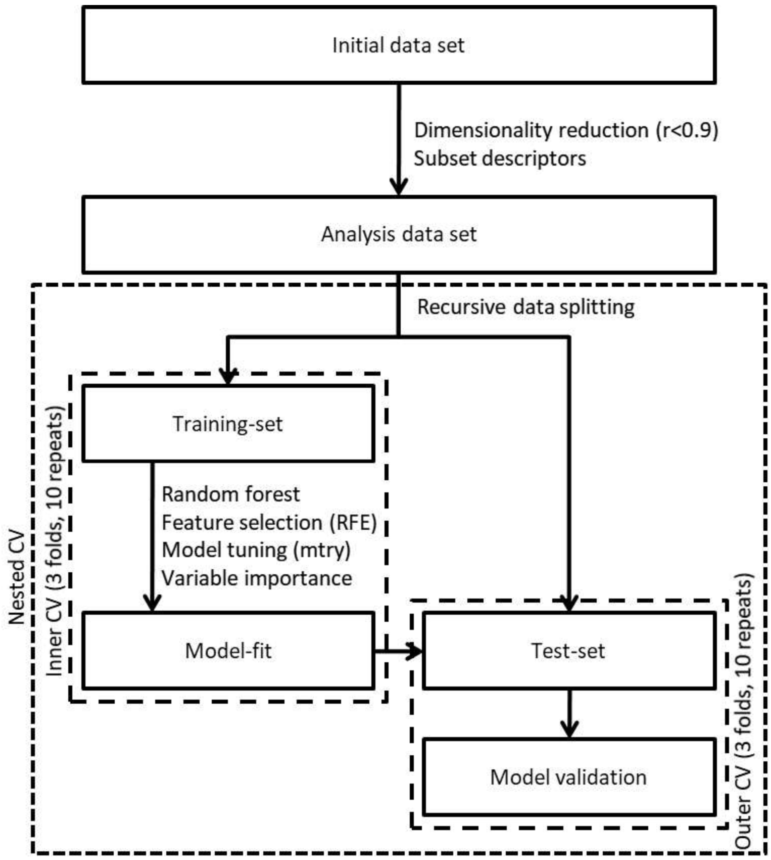

Figure 1. Workflow of the data preparation and modelling process based on a nested CV procedure

Table 2. Performance estimates of the random forest models in the inner (training set) and outer (test set) CV loops

	Inner CV	Outer CV
mtry[1]	3.30±1.26	
n. of predictors	10.57±3.86	
Explained variance (%)	67.3±13.4	64.1±8.5
RMSE[2]	2.63±0.55	2.66±0.37

[1]number of variables randomly sampled as candidates at each split; [2]root-mean-square error

By recursively splitting the dataset in train and test sets, the nested CV procedure ensured data used to train the model was not used to test it providing a robust evaluation of model performance (Cawley & Talbot, 2010). The relatively large variation in model performance metrics for both the inner and outer CV loops shows that RF may produce instable results when dealing with small sample sizes. The method employed showed an effective tool to overcome this issue as variation in model performance is captured and can therefore be evaluated. Also, consistency between performance measured in the inner and outer CV loops indicates the method used was capable of preventing overfitting.

The number of predictors retained in the models by RFE (inner CV) was 10.57±3.86 (Table 2). Figure 2 shows the improvement in model performance (RMSE) with increasing number of variables included during the model training process. Variable importance analysis (Figure 3) showed that grass quality parameters (NDF, ADF, NFC) are the most important predictors of hDMI. In the current study, grass quality was more important than DHA and PREGH and POSTGH in predicting hDMI. These outcomes confirm that maintaining high grass quality is key to maximise hDMI at pasture (Dillon, 2006). Cows' bodyweight, DIM and parity were also highly associated with hDMI. This indicates that a high variation in hDMI may exist among individuals. Overall, behavioural measures were shown to affect hDMI to a rather limited extent with just two variables (EAT1CHEW/EAT1TIME and LAYTIME) ranked among the ten most important predictors. Surprisingly, although a number of grazing bites is thought to be strictly related to hDMI (Umemura *et al.*, 2009; Oudshoorn *et al.*, 2013), EATBITE was among the least important variables being excluded from one-third of the models trained in the inner CV. Obviously this suggests that predicting hDMI based on behavioural measures alone may pose some challenges.

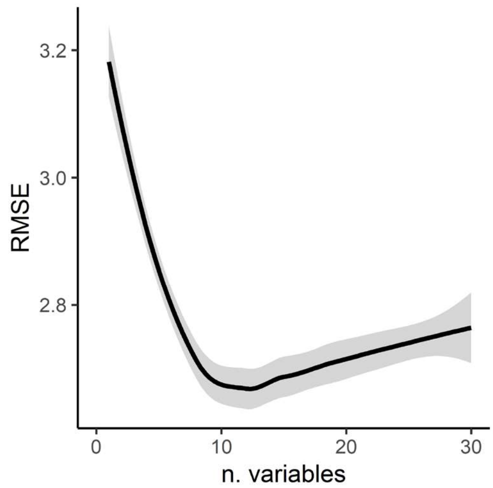

Figure 2. Root mean square error (RMSE) vs number of variables included in the models during model training process (inner CV)

In a recent study Rombach *et al.* (2019) also explored the possibility of modelling hDMI of grazing dairy cows based on behavioural data measured with the RumiWatchSystem and other accessible measures including cow characteristics, milk production and grass-related variables. Their results also highlighted that behavioural characteristics alone may not allow a sufficiently accurate estimation of individual hDMI.

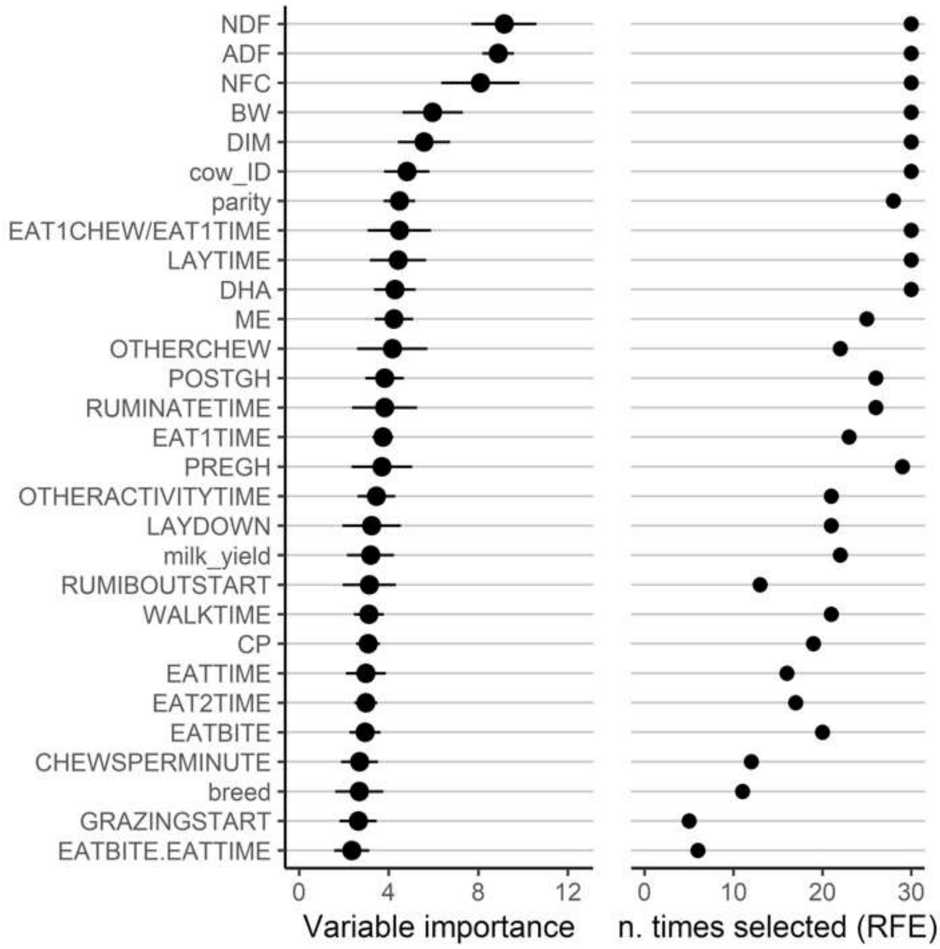

Figure 3. Random forest variable importance plot (left) and n. of times variables were selected by RFE (right) within the inner CV (train set)

Conclusions

Results showed that maintaining high grass quality is paramount to achieve high hDMI in pasture-based dairy systems. Random forest has been shown to have good potential for the prediction of hDMI in grazing dairy cattle. The nested CV procedure used in the current study allowed a robust estimation of model performance, even with a small sample size. Behavioural measurements showed limited importance in the RF model developed. Further research is required to identify new potential predictors and fully assess performance of machine learning algorithms for modelling of individual hDMI of grazing dairy cattle.

Acknowledgements

This project was funded by Science Foundation Ireland through the PRECISIONDAIRY project. The authors wish to sincerely thank all members of the Moorepark farm staff for their care of the experimental animals and assistance with measurements taken during the study.

References

Breiman, L. (2001). Random forests. *Machine Learning,* 45(1): 5–32.

Cawley, G.C. and Tablot, N.L. (2010). On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *Journal of Machine Learning Research,* 11: 2079–2107.

Dillon, P. (2006). Achieving high dry-matter intake from pasture with grazing dairy cows. In: Elgersma A., Dijkstra J. and Tamminga S. (eds.) *Fresh Herbage for Dairy Cattle*, 1–26. Springer. Printed in the Netherlands. pp. 1–26.

Dillon, P. and Stakelum, G. (1989). Herbage and dosed alkanes as a grass measurement technique for dairy cows. *Irish Journal of Agricultural Research* 28,104 (Abstr.).

Hanrahan, L., McHugh N., Hennessy, T., Moran, B., Kearney, R., Wallace, M. and Shalloo, L. (2018). Factors associated with profitability in pasture-based systems of milk production. *Journal of Dairy Science,* 101(6): 5474–5485.

Hellwing, A.L.F., Lund, P., Weisbjerg, M.R., Oudshoorn, F.W., Munksgaard, L. and Kristensenc, T. (2015). Comparison of methods for estimating herbage intake in grazing dairy cows. *Livestock Science,* 176: 61–74.

Oudshoorn, F.W., Cornou, C., Hellwing, A.L.F., Hansen, H.H., Munksgaard, L., Lund, P. and Kristensen, T. (2013). Estimation of grass intake on pasture for dairy cows using tightly and loosely mounted di- and tri-axial accelerometers combined with bite count. *Computers and Electronics in Agriculture* 99: 227–235.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rombach, M., Südekum, K.-H., Münger, A. and Schori, F. (2019). Herbage dry matter intake estimation of grazing dairy cows based on animal, behavioral, environmental, and feed variables. *Journal of Dairy Science,* 102: 1–15.

Shalloo, L., O' Donovan, M., Leso, L., Werner, J., Ruelle, E., Geoghegan, A., Delaby, L. and O'Leary, N. (2018). Review: Grass-based dairy systems, data and precision technologies. *Animal*, 12(s2): s262-s271.

Umemura, K., Wanaka, T. and Ueno, T. (2009). Technical note: Estimation of feed intake while grazing using a wireless system requiring no halter. *Journal of Dairy Science,* 92: 996–1000.

Werner, J., Leso, L., Umstatter, C., Niederhauser, J., Kennedy, E., Geoghegan, A., Shalloo, L., Schick, M. and O'Brien, B. (2018). Evaluation of the RumiWatchSystem for measuring grazing behaviour of cows. *Journal of Neuroscience Methods,* 300: 138–146.

Werner, J., Umstatter, C., Kennedy, E., Grant, J., Leso, L., Geoghegan, A., Shalloo, L., Schick, M. and O'Brien, B. (2019). Identification of possible cow grazing behaviour indicators for restricted grass availability in a pasture-based spring calving dairy system. *Livestock Science,* 220: 74–82.

Wright, M.M., Lewis, E., Garry, B., Galvin, N., Dunshea, F.R., Hannah, M.C., Auldist, M.J., Wales, W.J., Dillon, P. and Kennedy, E. (2019). Evaluation of the n-alkane technique for estimating herbage dry matter intake of dairy cows offered herbage harvested at two different stages of growth in summer and autumn. *Animal Feed Science and Technology,* 247: 199–209.