Predicting dairy herd resilience on farms with conventional milking systems

R.S.C. Rikkers[1], B. Ducro[2] and C. Kamphuis[1]

*[1]Wageningen Livestock Reseacrh, Wageningen University and Research,
P.O. Box 338, 6700 AH Wageningen, The Netherlands
Corresponding Author: roxann1.rikkers@wur.nl
[2]Animal Breeding and Genomics, Wageningen University and Research,
P.O. Box 338, 6700 AH Wageningen, The Netherlands*

Abstract

Individual cow resilience depends on the capacity of cows to respond to environmental disturbances. Together with management decisions, that affect the performance of these cows, and their environment they represent herd resilience. Until now, herd resilience can be estimated with the use of daily milk yield observations from an automated milking system (AMS), leaving conventional milking farms (CMS) without information on herd resilience. Therefore, this study investigated the possibility to predict dairy herd resilience using herd performance data generally available on AMS and CMS farms. Data from 585 Dutch AMS farms including herd performance variables: herd size, kg milk, proportion acidosis, proportion ketosis were used to predict herd resilience. As prediction model, a 5- fold cross validation Random Forest model with an extensive grid to finetune model parameter settings was used. Results show that, on average, herd resilience is predictable. Both the mean herd resilience as well as the predicted mean herd resilience were 1.30. The range, however, was much wider for herd resilience (0.70-1.86) than for the predicted herd resilience (1.10-1.50). Pearson correlation between herd resilience and predicted herd resilience was 0.55 ± 0.06. Thus, using only herd performance data that are generally available on farms (AMS and CMS), it is possible to predict herd resilience but not with a high accuracy.

Keywords: Dairy herd resilience, random forest, prediction, automated milking system.

Introduction

Individual cow resilience indicates that cows are minimally affected by environmental disturbances, such as pathogens or extreme weather, and their production quickly recovers if they are affected (Colditz and Hine, 2016). The individual cow resilience together with management decisions that affect the performance of these animals and their environment represents herd resilience (Blanc *et al.*, 2013). For dairy herds, this could be defined as that resilient herds show less milk yield deviations on herd level and thereby the herd as a whole is assumed to be less affected by disturbances. Individual cow resilience (Elgersma *et al.*, 2018; Poppe *et al.*, 2020) and dairy herd resilience can only be estimated on farms that use automated milking systems (AMS). For individual cow resilience, an expected lactation curve is fitted and less deviations from this fitted lactation curve indicates good individual resilience (Poppe *et al.*, 2020). The estimated individual resilience is corrected for a fixed herd-year and year-season effect and a random genetic and error effect. This fixed herd-year effect is assumed to represent herd resilience and it has been shown to correlate with herd management decisions (Poppe *et al.*, 2021). Estimation of individual or dairy herd resilience is only possible with the use of daily milk yield observations from AMS, and only after a full lactation. Still, a large portion of farms do not have an AMS but milk conventionally (in a milking parlour). Therefore, this study investigates the possibility to predict herd resilience with the use of a Random Forest model and herd performance variables usually available on farms with and without AMS.

Material and methods

An existing dataset including 2,644 Dutch AMS farms between 2011 and 2017 was available and consisted of a herd resilience indicator (fixed herd-year effect) and variables describing herd performance: kg milk, fat percentage, protein percentage, proportion with elevated somatic cell count, proportion ketosis, proportion of survival till second lactation, parity, herd size, etc. Incomplete records were removed resulting in a subset of 585 herds between 2012 and 2016 with complete information. These five herd-year estimates were averaged per herd, resulting in a mean herd resilience indicator per herd, which was assumed to represent herd resilience of a farm between 2012 and 2016. Mean herd resilience was 1.30 ranging from 0.70 till 1.86, where low values indicates good herd resilience.

To predict herd resilience, a Random Forest model based on the algorithm of Breiman (2001) was used (R package RandomForest (Liaw and Wiener, 2002). A Random Forest model is a combination of randomly generated decision trees. Each individual tree performs quite poorly, but many trees combined into a 'forest' provides more reliable prediction results. Previously, the Random Forest algorithm has shown to produce reliable predictions with e.g. the prediction of dairy cow survival till second lactation (Heide *et al,*. 2019) and with the prediction of lifetime resilience in dairy cows (Ouweltjes *et al.*, 2021). Finetuning of the Random Forest model parameters was done using an extensive grid: number of generated trees were 500, 1000, 1500, 2000 and 5000, minimum number of herds per branch were 5, 10 and 15, and the random number of candidate variables at each split 1 till 34 (in total 34 predictive variables were available). Each of the 510 model combinations was trained on 80% of the data and validated on the remaining 20% of the data using a 5-fold cross validation stratified to herd. This means that each herd was used four times in the training data and once in the validation set.

Results and discussion

Random Forest model finetuning

The best performing Random Forest model combination, based on the Pearson correlation between herd resilience and predicted herd resilience (0.55 ± 0.06), was the model that generated 500 random trees, contained a minimum number of five herd per branch and that randomly selected four predictive variables per split. The poorest performing model combination was the model that generated 1500 trees, contained a minimum number of 15 herds per branch and randomly selected one predictive variable per split (Pearson correlation estimated vs. predicted 0.52 ± 0.06). Considering that 500 randomly generated trees and a minimum of five herds per branch are default parameters of the randomForest function (Liaw and Wiener, 2002) and the difference between the best and poorest performing model is small, the need for an extensive grid was not necessary for this dataset.

Table 1. Mean, minimum and maximum of estimated and predicted herd-year estimates.

Herd-year effect	Mean	Min – Max
Estimated	1.30	0.70 – 1.86
Predicted	1.30	1.10 – 1.50

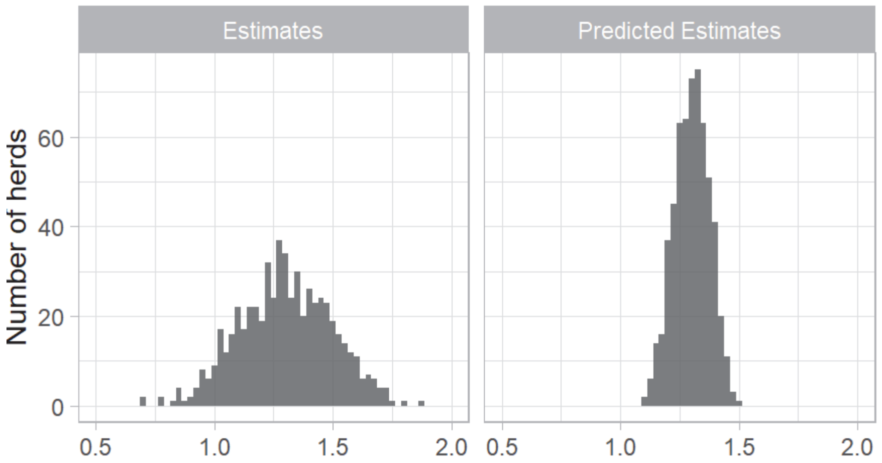

Figure 1. Histogram of herd-year estimates (range 0.70-1.86) and predicted herd-year estimates (range 1.10-1.50) using a Random Forest model.

Prediction accuracy

Pearson correlation between herd resilience and predicted herd resilience was 0.98 ± 0.00 for the 80% training datasets and 0.55 ± 0.06 for the 20% validation datasets (mean and standard deviation of the five k-folds). On average the Random Forest model performed quite well, since the average predicted herd resilience indicator of 1.30 was similar to the estimated herd resilience indicator (Table 1). However, the range of herd resilience was 0.70-1.86 while the range of predicted herd resilience was 1.10-1.50 (Table 1; Figure 1).

A method to improve prediction accuracy could be discretization, which is a pre-processing step where continuous variables are transformed to discrete variables. Either transforming the continuous predictive variables or the continuous variable herd resilience might improve the prediction accuracy, since applying discretization in complex datasets has been shown to significantly improve model performance (Lustgarten *et al.*, 2008).

Conclusion

Using only herd performance data that is generally available on farms it is possible to predict herd resilience but not with a high accuracy.

Acknowledgement

We acknowledge European Union's Horizon 2020 research and innovation program (GenTORE) under grant agreement No. 727213 for their financial support.

References

Blanc, F., E. Ollion,, L. Puillet, L. Delaby, S. Ingrand, M. Tichit, and N.C. Friggens, 2013. Evaluation quantitative de la robustesse des animaux et du troupeau: quels principes retenir. Proceedings of the 20th Rencontres Autour des Recherches sur les Ruminants, 4-5.

Breiman,, L., 2001. Random Forests. Machine Learning 45, 5–32

Colditz, I. G., and B.C. Hine, 2016. Resilience in farm animals: Biology, management, breeding and implications for Animal Welfare. Anim. Prod. Sci. 56(12), 1961.

Elgersma, G. G., G. de Jong, R. van der Linde and H.A. Mulder, 2018. Fluctuations in milk yield are heritable and can be used as a resilience indicator to breed healthy cows. J. Dairy Sci. 101(2), 1240–1250.

Heide, E.M.M. van der, R.F. Veerkamp, M.L. van Pelt, C. Kamphuis, I. Athanasiadis and B. Durco, 2019. Comparing regression, naive Bayes, and random forest methods in the prediction of individual survival to second lactation in Holstein cattle. J. Dairy Sci. 102:9409-9421.

Liaw A. and M. Wiener, 2002. Classification and Regression by randomForest. R News 2(3), 18—22.

Lustgarten, J.L., V. Gopalaksirhnan, K. Grover and S. Viswerswaran, 2008. Improving classification preformance with discretization on biomedical datasets. AMIA Annu Symp. Proc. 2008 445-449

Ouweltjes W., M. Spoelstra, B. Ducro, Y. de Haas and C. Kamphuis, 2021. A data-drive prediction of lifetime resilience of dairy cows using commercial sensor data collected during first lactation. J. Dairy Sci. 104:11759-11769.

Poppe, M., R.F. Veerkamp, M.L. van Pelt and H.A. Mulder, 2020. Exploration of variance, autocorrelation, and skewness of deviations from lactation curves as resilience indicators for breeding. J. Dairy Sci., 103(2), 1667–1684.

Poppe, M., H.A. Mulder, C. Kamphuis and R.F. Veerkamp, 2021. Between-herd variation in resilience and relations to Herd Performance. J. Dairy Sci., 104(1), 616–627.