Journal of Dairy Research

cambridge.org/dar

Predicting dairy herd resilience on farms with conventional milking systems

Roxann S. C. Rikkers, Bart J. Ducro, Rianne van Binsbergen and Claudia Kamphuis

Wageningen University & Research, Animal Breeding & Genomics, Wageningen, The Netherlands

Research Article

Cite this article: Rikkers RSC, Ducro BJ, van Binsbergen R and Kamphuis C. Predicting dairy herd resilience on farms with conventional milking systems. *Journal of Dairy Research* https://doi.org/10.1017/S0022029923000432

Received: 19 December 2022
Revised: 16 May 2023
Accepted: 30 May 2023

Keywords:
Herd resilience; management; milk production data; prediction

Corresponding author:
Claudia Kamphuis;
Email: claudia.kamphuis@wur.nl

Abstract

This research paper addresses the problem that, thus far, there is no method available to predict herd resilience for farms that do not use automated milking systems (AMS). Recently, a methodology was developed to estimate both individual cow as well as herd resilience using daily milk yield observations at individual cow level from farms with AMS. This AMS-based method, however, is not suitable on farms that use conventional milking systems (CMS) where such individual cow milk yield observations are lacking. Therefore, this research aimed at predicting herd resilience using herd performance data that is commonly available on CMS farms. To do so, data consisting of 585 Dutch AMS farms where herd resilience estimates using the AMS-based method were available was examined. To predict herd resilience with herd performance data, only those data that are also commonly available on CMS farms were used in a 5-fold cross validation Random Forest model. These herd resilience estimates were subsequently compared with the AMS-based herd resilience estimates. Results showed that it is possible to predict with a 69.9% probability whether a herd performs with above or below average herd resilience using only variables available on CMS farms. Especially, the proportion of cows with an indication of rumen acidosis, proportion of cows with an elevated somatic cell count and the fluctuation in herd size over the years are good predictors of herd resilience. Since herd management decisions appear to affect herd resilience, a lower predicted herd resilience could be taken as a general indication that tactical or strategic management changes could be taken to improve the herd resilience.

Resilient cows are those that are minimally affected by environmental disturbances, such as pathogens or extreme weather, and that quickly recover if they are affected by these disturbances (Colditz and Hine, 2016; Berghof *et al*., 2019). Farms with good herd management could support less resilient cows, whereas limitations in herd management could potentially be compensated by the resilience of cows. Herd resilience is not simply the average of the individual cow resilience values in a herd, since it depends on the adaptive capacity of the animals together with management decisions that affect the performance of these animals and their environment (Blanc *et al*., 2013). In the context of dairy herds, this means that resilient herds show less milk yield deviations at herd level and thus the herd as a whole is assumed to be less affected by disturbances. There are differences in herd resilience and these differences could partly be explained by herd management: Poppe *et al*. (2021) have shown that resilient herds tended to have a lower somatic cell score (SCS), a lower proportion of cows with elevated somatic cell count (SCC), a higher survival to second lactation, a shorter calving interval, a lower proportion of cows with either a ketosis indication or a rumen acidosis indication and, finally, a lower age at first calving when compared to non-resilient herds.

The general idea of proxies for individual cow resilience (Elgersma *et al*., 2018; Poppe *et al*., 2020) and dairy herd resilience (Poppe *et al*., 2021) is based on fluctuations of the actual measured daily milk yield around the expected daily milk yield curves. These fluctuations can be used to compute variances or auto-correlations that express different aspects of resilience (Poppe *et al*., 2020). Daily milk yield recordings are, therefore, necessary to be able to compute proxies for resilience, and thus, resilience can only be estimated on farms that use automated milking systems (AMS), sometimes known as milking robots. In contrast to AMS, many milking parlours and other conventional milking systems (CMS) do not record daily milk yield. In the Netherlands the percentage of farms with an AMS has increased from approximately 10% in 2010 to 31% (4,574 farms) in 2022 (Stichting KOM, 2022). Worldwide, the estimated number of farms with an AMS increased from 8,000 in 2009 (de Koning, 2011) to approximately 25 000 farms in 2015 (Barkema *et al*., 2015). Even though the number of farms with an AMS is increasing, a large portion of farms still use CMS. These farms have less daily milk yield information and individual or herd resilience estimation becomes problematic to estimate. Furthermore, farms that install an AMS need to wait one full lactation before it is possible to estimate individual or herd resilience of their cows, since this estimation is based on one full lactation (Poppe *et al.*, 2020). Even though individual cow resilience cannot be estimated for these CMS farms, it could be helpful to know their herd resilience. Farms with low herd resilience could try to improve the herd resilience *via* management decisions that affect the performance of their animals and their environment.

The aim of this study is to assess whether herd resilience can be predicted using herd performance data commonly available on CMS farms. To do so, we used data from AMS farms and used the methodology of Poppe *et al.* (2021) to estimate herd resilience, which served as a reference value. We, subsequently, applied a Random Forest model to predict herd resilience using only variables that are also available on CMS farms.

Material and methods

To address the aim of this study the following information was needed: (1) a herd resilience indicator for each herd and (2) herd performance variables that could predict this indicator. Thus far, herd resilience was only possible to estimate using daily milk yield observation from the AMS and, therefore, this study used data from AMS farms. However, to predict herd resilience only data commonly available on CMS farms, i.e. herd performance variables, were used. First, the variable which is expected to represent general herd resilience is described. Second, a description is given of the variables that were used to predict this general herd resilience indicator and last, the analyses performed in this study are detailed.

General herd resilience indicator

The general herd resilience indicators used in this study were the herd-year (HY) effects, originating from Poppe *et al.* (2021)'s model to estimate individual cow resilience. In that particular study, individual cow resilience was estimated by fitting an individual lactation curve using a quantile polynomial regression method. This lactation curve was used as the expected daily milk yield of each individual cow. Deviations between this expected curve and the actual observed milk yield were obtained and used for calculating the natural log-transformed variance (LnVar) from these deviations. Low values of this LnVar indicate that a cow would have a smaller deviation from this expected milk yield curve and thus is considered to be resilient. The mixed animal model, as used by Poppe *et al.* (2021) to estimate individual cow resilience for primiparous cows, was as follows:

$$\boldsymbol{y_{ijk} = HY_i + YS_j + \ a_k \ + \ e_{ijk}}$$

where y_{ijk} was the ln-transformed variance from an individual expected lactation curve, i.e. the individual cow resilience of cow k in HY class i and year-season class j; HY_i was the fixed effect of HY of calving i; YS_j was the fixed effect of year-season of calving j; a_k was the random genetic effect of animal k; and e_{ijk} was the random error term. A detailed description of this model is provided by Poppe *et al.* (2021). The HY effects represent the effect on individual cow resilience in a given herd and a given year. These HY effects could be interpreted as the average resilience in a certain herd for all primiparous cows that calved in a certain year, corrected for their breeding values and general year-season effects. Low HY estimates would indicate less variation of deviations at herd level and thus the herd as a whole is assumed to be less affected by disturbances, indicating good resilience. Therefore, the HY effects originating from this mixed animal model will be used to represent herd resilience hereafter, where low values indicate good herd resilience.

Transformation of predictive variables

For this study, an existing dataset from 2,644 Dutch AMS farms between 2011 and 2017 was available (data previously described in Poppe *et al.*, 2021). The data consisted of herd performance variables, information obtained by AMS and the herd resilience indicator at HY level, i.e. one record per herd per year. To predict herd resilience, seventeen herd performance variables commonly available on farms that use milk recording were used: mean daily milk yield per herd (average of all cows), mean fat content, mean protein content, mean lactose content, mean urea content, mean somatic cell score (SCS), mean calving interval, mean age of first calving, proportion cows with elevated somatic cell count (PropSCC), proportion cows with rumen acidosis indication (PropACID), proportion cows with ketosis indication (PropKET), proportion cows that survived till second lactation (PropSURV), mean parity, mean age, number of cows in a herd, proportion of cows not 100% Holstein Friesian and proportion cows herd-book registered (Table 1). Mean SCS was derived from the SCC using (CRV, 2018):

$$\text{SCS} = 1,000 + 100 \times \left[\log_2\left(\frac{\text{SCC}}{1,000}\right)\right]$$

PropSCC was the proportion of cows with ≥1 cases of elevated SCC during lactation, where SCC was considered elevated if SCC >100 000 cells/ml. PropACID was the proportion of cows with ≥1 indications of rumen acidosis during lactation. A rumen acidosis indication was based on the fat and protein content; when fat content was lower than the protein content and below 4% the cow received a rumen acidosis indication. PropKET was the proportion of cows with a ketosis indication during lactation, where a ketosis indication was based on the fat–protein ratio and Fourier transformed infrared measurements of milk acetone and milk β-hydroxybutyric acid (Poppe *et al.*, 2020). Incomplete observations, i.e. HY combinations with missing values on any of the aforementioned variables, were removed, resulting in a subset of 585 herds between 2012 and 2016 (2,925 HY combinations). To include possible fluctuations in herd size and health indicators such as the mean SCS, acidosis indication and ketosis indication, the variance of these HY combinations was computed (Table 1). Thus, in total 17 different 5-year averaged and 17 different 5-year variance predictive variables commonly available on dairy farms were used to predict the 5-year averaged herd resilience indicator. Predicting herd resilience for AMS farms using herd performance data has not been done before. Moreover, AMS farms have six additional herd performance variables specific for these types of farms (and thus unavailable for CMS farms). For reference purposes, we also performed the analysis including these additional six herd performance variables. These six additional variables are: both the 5-year average and the 5-year variance for (1) the daily herd milk yield, (2) the number of days with missing milk yield records, and (3) the number of days with records (Table 1). These additional six variables will be added to the 34 herd performance variables to study whether they improve the prediction of herd resilience for AMS farms specifically.

Table 1. Herd performance variables, type of cows on which each herd performance variable was based, 5-year average (mean, min and max) and 5-year variance (mean, min and max) of 585 herds

		5-year average		5-year variance	
Variables[a]	Cows on which variable is based[b]	Mean	Min – Max	Mean	Min – Max
Milk (kg × 10)	Primi	256.9	186.9–356.4	1.68	0.01–15.27
Fat (% × 100)	Primi	442.7	391.8–504.1	118.9	1.8–787.4
Protein (% × 100)	Primi	359.0	333.6–394.1	38.7	1.4–177.1
Lactose (% × 100)	Primi	464.5	450.6–475.3	9.3	0.3–69.0
Urea	Primi	22.7	12.5–29.7	2.5	0.1–16.4
SCS	Primi	1565	1469–1653	1114	19–6242
CI (days)	Primi	402	361–526	338	4–2852
AFC (months)	Primi	25.2	22.8–31.8	0.4	0.0–5.8
PropSCC	Primi	0.67	0.30–0.96	0.01	0.00–0.01
PropACID	Primi	0.19	0.02–0.55	0.01	0.00–0.05
PropKET	Primi	0.08	0.00–0.43	0.004	0.00–0.04
PropSURV	Primi	0.88	0.73–0.99	0.004	0.00–0.02
Parity	All	2.6	2.0–3.6	0.02	0.00–0.12
Age (years)	All	4.0	3.4–5.3	0.03	0.00–0.20
Herd size	All	117	51–362	276	2–4816
PropNonHF	All	0.19	0.00–0.78	0.003	0.00–0.03
PropReg	All	0.96	0.72–1.00	0.000	0.00–0.02
Daily milk (kg)[c]	Primi	26.7	19.2–36.4	1.8	0.0–21.0
Missing records (days)[c]	Primi	1.4	0.1–6.5	2.7	0.0–28.9
Records (days)[c]	Primi	298.5	213.7–328.8	139.9	1.3–2722.4
Herd resilience[d]	Primi	1.30	0.70–1.86	0.03	0.00–0.25

[a]Milk = mean daily kilogram milk per farm (×10); fat = mean fat content; protein = mean protein content; lactose = mean lactose content; ureum = mean ureum content; SCS = mean somatic cell score; CI = mean calving interval from first to second lactation; AFC = mean age at first calving; PropSCC = proportion of cows with at least 1 elevated somatic cell count; PropACID = proportion of cows with at least 1 rumen acidosis indication; PropKET = proportion of cows with at least 1 ketosis indication; PropSURV = proportion of cows that survived to second lactation; Parity = mean parity; age = mean age; herd size = number of cows calved; PropNonHF = proportion of cows that are not 100% Holstein Friesian; PropREG = proportion of cows that are herd-book registered; daily milk = mean daily milk yield records from automated milking system; missing records = number of days with a missing milk yield record; records = number of days with a milk yield record.
[b]Primi = primiparous cows with a resilience indicator in the herd-year class; All = all cows in the herd-year class.
[c]Variables only available on farms with an automated milking system.
[d]Herd resilience indicator provided by Poppe *et al.* (2021).

Analyses

Van der Heide *et al.* (2019) and Ouweltjes *et al.* (2021) have shown previously that a Random Forest algorithm can produce reliable predictions with, respectively, the prediction of dairy cow survival till second lactation and the prediction of a lifetime resilience score in dairy cows. Lifetime resilience predicted by Ouweltjes *et al.* (2021) was defined as the cumulative results of a cow's ability to recalve and thus the ability to extend her productive lifespan (Adriaens *et al.*, 2020). The current study also used the Random Forest model based on the algorithm of Breiman (2001) using R package RandomForest (Liaw and Wiener, 2002). Tuning of the model parameters was done using an extensive grid; number of generated trees were 500 (default), 1000, 1500, 2000 and 5000, minimum number of herds per branch were 5 (default), 10 and 15, and the random number of candidate variables at each split 1 till 34 (in total 34 predictive variables were available), where the default setting was the total number of variables divided by three. This extensive grid search resulted in 510 Random Forest models with different parameter settings. Each of these 510 Random Forest models 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, i.e. each herd was used four times in the training data and once in the validation set. This grid was used on the dataset containing 34 predictive variables commonly available on CMS farms and the dataset containing 40 predictive variables (CMS variables extended with six AMS variables). To determine the best performing Random Forest model, the five validation datasets were combined. The best model was the model with the highest Pearson correlation between herd resilience and predicted herd resilience.

The importance of the predictive variables was assessed, and because a 5-fold cross validation was used, the order of importance variables could be different for each k-fold. For example, the top 10 predictive variables of k1 could be different from the top 10 of k3, because a different portion of the dataset was used as training and validation sets. To summarize this into one top 10 of most predictive variables, a ranking was made, where a value of 10 was assigned to the 1st important variable and a value of 1 to the 10th variable. For example, if the predictive variable '5-year average herd size' was 2nd most important in k1, 2nd most important in k2, 3rd most important in k3, 2nd most important in k4 and 1st most important in k5, in total 5-year average herd size would be quantified with a value of 45 (9 + 9 + 8 + 9 + 10) in the total ranking of importance of predictive variables. Analyses were done using RStudio version 3.40.6 (R Studio Team, 2016). Data handling and visualization was done with, respectively R package dplyr version 0.8.5 (Wickham *et al*., 2020) and ggplot2 version 3.3.1 (Wickham, 2016).

Results

Prediction accuracy

Using the dataset containing 34 CMS features, the best performing Random Forest model was the one that generated 500 random trees, with a minimum of five herds per branch, and that randomly selected four predictive variables per split. The Pearson correlation between herd resilience and predicted herd resilience was 0.55 ± 0.06 and the Spearman rank correlation was 0.56 ± 0.06. The average for both the observed and the predicted herd resilience was 1.30. However, the range differed since the range of herd resilience was 0.70–1.86, whereas the range of the predicted herd resilience was 1.10–1.50. Nevertheless, herds with below average herd resilience tended to have below average predictions (197 herds) and herds with above average herd resilience tended to have above average predictions (212 herds; Figure 1). Thus with a 69.9% probability it was possible to predict whether a herd performs above average or below average.

Using the dataset containing 34 CMS features and 6 additional AMS features, the best performing Random Forest model had the same parameter settings as the one using the dataset containing only the 34 CMS features (500 generated trees, minimum of five herds per branch and randomly four selected variables per split). Adding these six AMS variables resulted in a slightly higher prediction performance, with a Pearson correlation of 0.58 ± 0.05 and a Spearman rank correlation of 0.58 ± 0.05. Consequentially, a few more herds with below average herd resilience also had below average predicted herd resilience (203 herds) and above average herds had above average predicted herd resilience (214 herds; Figure 2), increasing the probability to predict below or above average herd resilience to 71.3%. Furthermore, the Pearson and Spearman rank correlations of herd resilience predicted from either the dataset with 34 CMS or the dataset with 34 CMS and 6 AMS features were both 0.96. In other words, farms with high prediction in the first analysis also received high predictions in the second analysis.

Importance of predictive variables

The importance of the predictive variables using the best performing model with 34 variables is quantified in Table 2. In each of the 5 folds the most important predictive variable was the 5-year average proportion of cows with at least 1 rumen acidosis indication (importance = 50). This was followed by the 5-year average proportion of cows with at least 1 elevated somatic cell count (SCC) and the 5-year variance of herd size. The Pearson correlations showed that farms with poor herd resilience tended to have a higher proportion of cows with a rumen acidosis indication (0.46), a higher proportion of cows with elevated SCC (0.29) and a more variable herd size over the years (0.17).

The quantified top 10 of the best performing model using the 34 variables and additionally six AMS variables are presented in Table 3. Besides the proportion of cows with a rumen acidosis indication and a variable herd size over the years, the average number of missing records (rank 2) and variance of missing

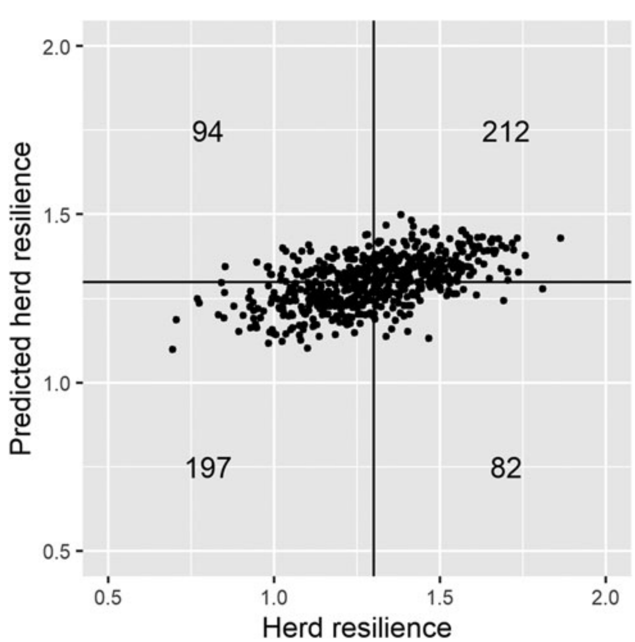

Figure 1. Scatterplot of herd resilience *vs.* predicted herd resilience using 34 predictive variables commonly available on farms with conventional miking systems. The horizontal and vertical line represent the mean herd resilience (1.30). The number in each quartile represents the number of herds in that corresponding quartile.

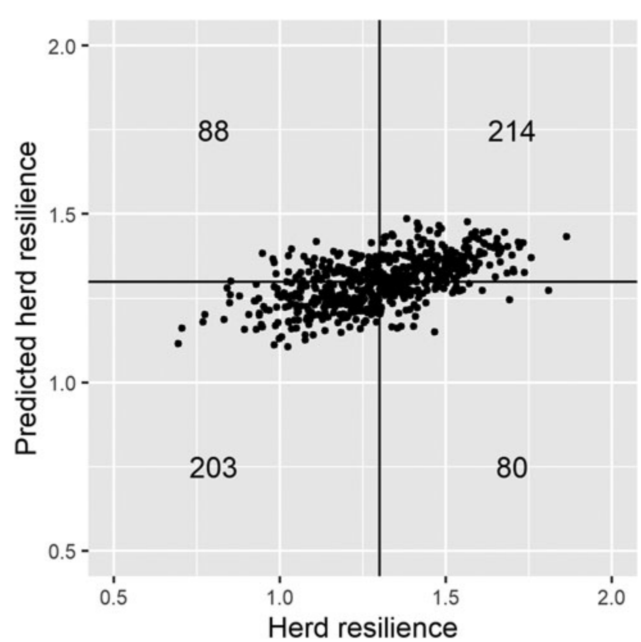

Figure 2. Scatterplot of herd resilience *vs.* predicted herd resilience using 40 predictive variables commonly available on farms with conventional miking systems and available on farms that use an automated milking system. The horizontal and vertical line represent the mean herd resilience (1.30). The number in each quartile represents the number of herds in that corresponding quartile.

https://doi.org/10.1017/S0022029923000432 Published online by Cambridge University Press

Table 2. Top 10 predictive variables using 34 herd performance variables commonly available on farms with conventional milking systems

Ranking	Predictive variables[b]	5-year average or 5-year variance	Importance	Pearson correlation[a]
1	PropACID	Average	50	0.46
2	PropSCC	Average	42	0.29
3	Herd size	Variance	36	0.17
4	SCS	Average	36	0.29
5	Milk (kg × 10)	Average	31	0.23
6	Herd size	Average	28	0.22
7	Fat (%)	Average	22	−0.26
8	PropReg	Average	11	−0.12
9	CI (d)	Average	10	0.19
10	PropACID	Variance	3	0.14

[a]Pearson correlation with the herd resilience indicator.
[b]PropACID = proportion of cows with at least 1 rumen acidosis indication; PropSCC = proportion of cows with at least 1 elevated somatic cell count; Herd size = number of cows calved; SCS = mean somatic cell score; Milk = Mean kilogram milk (×10) from national database; Fat = mean fat content; PropREG = proportion of cows that are herd-book registered; CI = mean calving interval from first to second lactation.

records over the years (rank 10) from the AMS tended to be important to predict herd resilience. Similar as the results from the dataset using 34 CMS variables, the Pearson correlations showed that farms with poor herd resilience tended to have a higher proportion of cows with a rumen acidosis indication (0.46), a more variable herd size over the years (0.17), and more missing daily milk yield records (0.15).

Discussion

This study investigated the possibility of predicting herd resilience using only herd performance variables commonly available on CMS farms. Low herd resilience values indicate less variation of milk yield deviations at herd level and thus the herd as a whole is assumed to be less affected by disturbances, indicating good resilience. This study showed that it is possible to predict whether a herd performs above average or below average with a 69.9% probability. We also showed that adding six additional variables that are only available on AMS farms increased this probability to 71.3%. Furthermore, the Pearson correlation between the analysis containing 34 CMS variables and adding six AMS variables was 0.96, indicating farms with a high prediction in the first analysis (34 variables) also received high predictions in the second analysis (40 variables). Thus, no difference was observed between the predictions using CMS or those using CMS + six additional AMS variables. This study also showed that the 5-year average proportion of cows with ≥1 rumen acidosis indication, the 5-year average proportion of cows with ≥1 elevated SCC and the 5-year variance of herd size are important variables to predict herd resilience. Three of the additional six AMS also appeared in the top 10 most important predictive variables (average number of missing records, daily milk yield observations and variance of missing records over the years). These results indicate that besides health parameters (acidosis indication and elevated SCC), the size of a farm and the information from the AMS is relevant to predict herd resilience, although this did not significantly increase the accuracy of prediction.

Table 3. Top 10 predictive variables using 34 herd performance variables commonly available on farms with conventional milking systems and additionally six automated milking system variables (bold)

Ranking	Predictive variables[b]	5-year average or 5-year variance	Importance	Pearson correlation[a]
1	PropACID	Average	50	0.46
2	**Missing records (d)**	Average	42	0.15
3	Herd size	Variance	36	0.17
4	PropSCC	Average	29	0.29
5	SCS	Average	29	0.29
6	**Daily milk (kg)**	Average	27	0.25
7	Milk (kg × 10)	Average	26	0.23
8	Herd size	Average	17	0.22
9	Fat (%)	Average	10	−0.26
10	**Missing records (d)**	Variance	4	0.04

[a]Pearson correlation with the herd resilience indicator.
[b]PropACID = proportion of cows with at least 1 rumen acidosis indication; Missing records = Number of days with a missing milk yield record; Herd size = number of cows calved; PropSCC = proportion of cows with at least 1 elevated somatic cell count; SCS = mean somatic cell score; Daily milk = mean daily milk yield records from automated milking system; Milk = Mean kilogram milk (×10) from national database; Fat = mean fat content.

https://doi.org/10.1017/S0022029923000432 Published online by Cambridge University Press

As previously suggested by Poppe *et al.* (2021), herd resilience is the ability to control the number and severity of disturbances in a herd. Since we used HY effects from an animal model that is corrected for genetic effects as a herd resilience indicator, this indicator represents the ability to control the number and severity of disturbances in the herd through management decisions related to feed, health, housing and how quickly a farmer can respond to disturbances. For example, the provision of a well-balanced diet containing adequate amounts of vitamins and minerals could reduce the vulnerability to mastitis pathogens (Heinrichs *et al.*, 2009) or exposure to extreme heat or extreme cold could be reduced by proper roof insulation (Fournel *et al.*, 2017), and implementing good hygiene protocols can reduce the exposure to pathogens and reduce risk of secondary infections (Deng *et al.*, 2019).

As previously described, to estimate herd resilience, the daily milk yield observations from the AMS are needed. Thus, herd resilience of CMS farms is unknown and predicting herd resilience and simultaneously validating this prediction is not possible. Therefore, we needed to make the assumption that AMS farms also represent CMS farms. However, it is important to realize that CMS farms and AMS farms are actually different systems. For example, cows on an AMS farm can choose when they want to be milked. Hopster *et al.* (2002) has shown that social competition forces low-ranking cows to visit the AMS at different times than preferred and this potential irregularity in milking intervals could affect milk production (Ouweltjes, 1998) and SCC (Mollenhorst *et al.*, 2011). Both aspects may affect the results from this study as the resilience indicator itself is based on the variation of deviations from an expected lactation curve and the second most predictive variable was the proportion of cows with ≥1 elevated SCC event. Further research is needed to determine if these variables are indeed different between AMS and CMS farms and with what magnitude this affects the results of this study.

The dataset used in this study included 585 herds with data between 2012 and 2016 (after filtering). Herd performance variables were at herd level, meaning for each herd and each year an average of all cows was available. These herd level averages were averaged once more into five-year averages and to account for possible yearly fluctuations, the variance values of these five years were included. Results from Poppe *et al.* (2021) showed a consistency in herd resilience estimates over multiple years (high correlation between years), indicating the fluctuation of herd resilience over years is limited. We observed a low 5-year variance (0.03) of herd resilience between 2012 and 2016 supporting these reported findings. Furthermore, predicting herd resilience of 2016 using HY variables of 2012 till 2015 resulted in a similar prediction accuracy and range (results not shown). This indicates that the abolishment of the Dutch milk quota in 2015, meaning farmers could increase herd size and total milk production per year, did not affect the results in this study.

In the analysis using 34 CMS variables we observed four predictors with a relative importance of >35. The proportion of cows with a rumen acidosis indication was the most important variable (importance 50) followed by the proportion of cows with ≥1 elevated SCC (importance 42), mean SCC in the herd (importance 36) and the variance in herd size over the five years (importance 36). Pearson correlations of these predictors with herd resilience were: 0.46, 0.29, 0.29 and 0.17, respectively. Similar correlations were observed in Poppe *et al.* (2020) where a correlation between herd resilience and proportion acidosis (0.31), elevated SCC (0.20) and mean SCS (0.19) was observed. Both rumen acidosis (Krause and Oetzel, 2006; Enemark, 2008; Abdela, 2016) and mastitis (Rajala-Schultz *et al.*, 1999; Gröhn *et al.*, 2004; Halasa *et al.*, 2009) can lead to an impaired milk production and thus more deviations from an expected lactation curve, with a consequential effect on the herd resilience indicator. Furthermore, Poppe *et al.* (2021) have shown that a general resilience indicator is affected by many factors and if a strong correlation between e.g. SCC and herd resilience was observed, it would rather be a mastitis indicator than a general resilience indicator. Also, the results showed that a consistent herd size (low variance of herd size over 5 years) is related to good herd resilience. A high variation indicates a large increase or a large reduction in size and we assume that major management changes are related to this, such as a new barn or different management strategies. Thus, a consistent herd size and consistent management practice indicate good resilience. Our results support the theory that the general herd resilience indicator is affected by a combination of multiple health traits and consequently, that herd management decisions related to food, health and herd size have an effect on herd resilience.

The maximum 5-year average of the two most important predictive variables was high with 0.55 and 0.96 for the proportion of cows with a rumen acidosis indication and the proportion of cows with an elevated SCC, respectively. These maximum values indicate that there is one farm where, on average, half of the cows had a rumen acidosis indication every year during 2012 and 2026, and another farm where, on average, almost all cows had a mastitis infection every year during 2012 and 2016. Comparing ten farms with high maximum 5-year averages for PropACID (range 0.42–0.55) with the ten lowest farms (range 0.02–0.04) revealed that farms with a high average of cows with a rumen acidosis indication are significantly less resilient ($P < 0.001$; results not shown). Furthermore, comparing ten farms with high maximum 5-year averages for PropSCC (range 0.87–0.96) with the ten lowest farms (range 0.30–0.42) revealed that also farms with a high average of cows with an elevated SCC are significantly less resilient ($P < 0.001$; results not shown). This indicates that herd resilience could be improved by reducing mastitis infections and rumen acidosis on farms and underlines the importance of proper herd management on farms, such as proper hygienic management practices and feeding adequate diets, to reduce the risk of mastitis infections and rumen acidosis. We can justifiably assume that poor herd resilience relates to poor health of the herd. Poor health is also related to poor fertility (Fourichon *et al.*, 2000; Wolfenson *et al.*, 2015) and cows show behavioral changes such as increased lying behavior, less feeding or physical activity if they are sick (Dittrich *et al.*, 2019). Therefore, it is likely that adding additional data such as fertility information or activity data would improve the prediction of herd resilience.

Table 3 shows that the AMS variable number of missing records as a mean over five years as well as the variance over five years is an important predictive variable when we include six AMS specific herd performance variables in the dataset. In practice, sick cows are separated from the herd and milked separately, meaning the number of missing records in the AMS would increase. Although low, the positive correlation of 0.27 between number of missing records and proportion of cows with an elevated SCC supports these results, meaning cows that might have mastitis are separated from the herd and not milked in the AMS and thus the number of missing records increased.

To conclude, until now, it was only possible to estimate herd resilience for farms with an AMS and only after one full lactation of a cow. However, results from the current study show that also for CMS farms it is possible to predict herd resilience. Moreover, with a 69.9% probability it is possible to predict whether a herd performs above or below average using only herd performance variables. Proportion of rumen acidosis, proportion of SCC in herds, mean SCS and fluctuation in herd size over five years are the better predictors of herd resilience. Results suggest that herd management decisions affect herd resilience, therefore, a lower predicted herd resilience could be an indication that, in general, tactical or strategic management changes could be taken to improve resilience of a herd.

Acknowledgements. We acknowledge European Union's Horizon 2020 research and innovation program (GenTORE) under grant agreement No. 727213 for their financial support. We acknowledge Prof. Dr Roel Veerkamp (Wageningen University & Research) for his contribution in getting funding for the GenTORE project. In addition, we acknowledge Cooperation CRV and CRV BV for providing the data. Finally, we acknowledge Dr Marieke Poppe for preparing the data used in this study and helping the authors to understand the data. The authors have not stated any conflicts of interest.

Financial support. None.

References

Abdela N (2016) Sub-acute ruminal acidosis (SARA) and its consequence in dairy cattle: A review of past and recent research at global perspective. *Achievements in the Life Sciences* **10**, 187–196.

Adriaens I, Friggins NC, Ouweltjes W, Scott H, Aernouts B and Statham J (2020) Productive life span and resilience rank can be predicted from on-farm first-parity sensor time series but not using a common equation on farms. *Journal of Dairy Science* **103**, 7155–7171.

Barkema HW, von Keyserlingk MAG, Kastelic JP, Lam TJPM, Luby C, Roy JP, LeBlanc SJ, Keefe GP and Kelton DF (2015) Invited review: changes in the dairy industry affecting dairy cattle health and welfare. *Journal of Dairy Science* **98**, 7426–7445.

Berghof TVL, Poppe M and Mulder HA (2019) Opportunities to improve resilience in animal breeding programs. *Frontiers in Genetics* **10**, 692.

Blanc FE, Ollion L, Puillet, L, Delaby S, Ingrand M, Tichit M and Friggens NC (2013) [Quantitative evaluation of animal and herd resilience]. *Proceedings of the 20th Rencontres Autour des Recherches sur les Ruminants*, 4–5.

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

Colditz IG and Hine BC (2016) Resilience in farm animals: biology, management, breeding and implications for animal welfare. *Animal Production Science* **56**, 1961–1983.

CRV (2018) E-18 Somatic cell score with test day model. Available at https://www.cooperatie-crv.nl/wp-content/uploads/2018/10/E_18-cel_april-2017_EN.pdf (Accessed 11 August 2022).

de Koning CJAM (2011) Milking machines | robotic milking. In Fuquay JW (ed.), *Encyclopedia of Dairy Sciences*, 2nd Edn. San Diego, CA: Academic Press, pp. 952–958.

Deng Z, Koop G, Lam TJGM, van der Lans IA, Vernooij JCM and Hogeveen H (2019) Farm-level risk factors for bovine mastitis in Dutch automatic milking dairy herds. *Journal of Dairy Science* **102**, 4522–4535.

Dittrich I, Gertz M and Krieter J (2019) Alterations in sick dairy cows' daily behavioural patterns. *Heliyon* **5**, e02902.

Elgersma GG, de Jong G, van der Linde R and Mulder HA (2018) Fluctuations in milk yield are heritable and can be used as a resilience indicator to breed healthy cows. *Journal of Dairy Science* **101**, 1240–1250.

Enemark JMD (2008) The monitoring, prevention and treatment of sub-acute ruminal acidosis (SARA): a review. *Veterinary Journal* **176**, 32–43.

Fourichon C, Seegers H and Malher X (2000) Effect of disease on reproduction in the dairy cow: a meta-analysis. *Theriogenology* **53**, 1729–1759.

Fournel S, Ouellet V and Charbonneau E (2017) Practices for alleviating heat stress of dairy cows in humid continental climates: a literature review. *Animals* **7**, 37.

Gröhn YT, Wilson JD, González RN, Hertl JA, Schulte H, Bennett G and Schukken YH (2004) Effect of pathogen-specific clinical mastitis on milk yield in dairy cows. *Journal of Dairy Science* **87**, 3358–3374.

Halasa T, Nielen M, de Roos APW, van Hoorne R, de Jong G, Lam TJGM, van Werven T and Hogeveen H (2009) Production loss due to new subclinical mastitis in Dutch dairy cows estimated with a test-day model. *Journal of Dairy Science* **92**, 599–606.

Heinrichs AJ, Costello SS and Jones CM (2009) Control of heifer mastitis by nutrition. *Veterinary Microbiology* **134**, 172–176.

Hopster H, Bruckmaier RM, van der Werf JTN, Korte SM, Macuhova J, Korte-Bouws G and van Reenen CG (2002) Stress responses during milking: comparing conventional and automatic milking in primiparous dairy cows. *Journal of Dairy Science* **85**, 3206–3216.

Krause KM and Oetzel GR (2006) Understanding and preventing subacute ruminal acidosis in dairy herds: a review. *Animal Feed Science and Technology* **126**, 215–236.

Liaw A and Wiener M (2002) Classification and regression by randomForest. *R News* **2**(3), 18–22.

Mollenhorst H, Hidavat MM, van den Broek J, Neijenhuis F and Hogeveen H (2011) The relationship between milking interval and somatic cell count in automatic milking systems. *Journal of Dairy Science* **94**, 4531–4537.

Ouweltjes W (1998) The relationship between milk yield and milking interval in dairy cows. *Livestock Production Science* **56**, 193–201.

Ouweltjes W, Spoelstra M, Ducro B, de Haas Y and Kamphuis C (2021) A data-driven prediction of lifetime resilience of dairy cows using commercial sensor data collected during first lactation. *Journal of Dairy Science* **104**, 11759–11769.

Poppe M, Veerkamp RF, van Pelt ML and Mulder HA (2020) Exploration of variance, autocorrelation, and skewness of deviations from lactation curves as resilience indicators for breeding. *Journal of Dairy Science* **103**, 1667–1684.

Poppe M, Mulder HA, Kamphuis C and Veerkamp RF (2021) Between-herd variation in resilience and relations to herd performance. *Journal of Dairy Science* **104**, 616–627.

R Studio Team (2016) RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. Available at http://www.rstudio.com/ (Accessed 25 February 2022).

Rajala-Schultz PJ, Gröhn YT, McCulloch CE and Guard CL (1999b) Effects of clinical mastitis on milk yield in dairy cows. *Journal of Dairy Science* **82**, 1213–1220.

Stichting KOM (2022) Overzicht aantallen veehouders en rapporten. Available at https://www.stichtingkom.nl/index.php/stichting_kom/category/statistiek (Accessed 11 August 2022).

van der Heide EMM, Veerkamp RF, van Pelt ML, Kamphuis C, Athanasiadis I and Ducro B (2019) Comparing regression, naive Bayes, and random forest methods in the prediction of individual survival to second lactation in Holstein cattle. *Journal of Dairy Science* **102**, 9409–9421.

Wickham H (2016) *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wickham H, François R, Henry L and Müller K 2020 dplyr: A Grammar of Data Manipulation. R package version 0.8.5. Available at https://CRAN.R-project.org/package=dplyr (Accessed 10 December 2021).

Wolfenson D, Leitner G and Lavon Y (2015) The disruptive effects of mastitis on reproduction and fertility in dairy cows. *Italian Journal of Animal Science* **14**, 650–654.

https://doi.org/10.1017/S0022029923000432 Published online by Cambridge University Press