Productive lifespan and resilience rank can be predicted from on-farm first parity sensor data but not using a common equation across farms

I. Adriaens[1,2,3*], N.C. Friggens[4], W. Ouweltjes[5], H. Scott[3], B. Aernouts[1], J. Statham[3]

[1] KU Leuven, Department of Biosystems, Biosystems Technology Cluster, Campus Geel, Kleinhoefstraat 4, 2440 Geel, Belgium

[2] KU Leuven, Department of Biosystems, Division Mechatronics, Biostatistics and Sensors, Kasteelpark Arenberg 30, 3001 Leuven, Belgium

[3] RAFT Solutions Ltd., Mill Farm, Studley Road, Ripon HG4 2QR, United Kingdom

[4] INRA, Modélisation Systémique Appliquée aux Ruminants, 16 Rue Claude Bernard, 75005, Paris, France

[5] Wageningen Research, Breeding & Genomics, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands

* Corresponding author: Ines Adriaens, ines.adriaens@kuleuven.be, Kasteelpark Arenberg 30 box 2456, 3001 Leuven, Belgium +32 (0) 16 32 88 73

ABSTRACT

A dairy cow's resilience and her ability to re-calve gains importance on modern dairy farms as it affects all aspects of the sustainability of the sector. Many modern farms today have milk meters and activity sensors that accurately measure yield and activity at a high frequency for monitoring purposes. We hypothesized that these same sensors can be used for precision phenotyping of complex traits such as resilience or productive lifespan. The objective of this study was to investigate if resilience and productive lifespan of dairy cows can be predicted using sensor-derived proxies of first parity sensor data. We used a data set from 27 Belgian and British dairy farms with an automated milking system containing at least 5 years of successive measurements. All of these

farms had milk meter data available, and 13 of these farms were also equipped with activity sensors. This subset was used to investigate the added value of activity meters to improve prediction model accuracy. To rank cows for resilience a lifetime score was attributed to each cow based on her number of re-calvings, her 305-day milk yield, her age at first calving, her calving intervals and the days in milk or culling. For analysis, cows were classified as either first (top 33%), medium (middle 33%) or last (bottom 33%). In total 45 biologically-sound sensor features were defined from the time-series data, including measures of variability, lactation curve shape, milk yield perturbations, activity spikes indicating oestrous events and activity dynamics representing health events. These features, calculated on first lactation data, were used to predict lifetime resilience rank. A common equation across farms to predict this rank could not be found. However, using a specific linear regression model progressively including stepwise selected features (cut-off *p*-value of 0.2) at farm level, classification performances were between 35.9% and 70.0% (46.7 ± 8.0, mean ± standard deviation) for milk yield features only and between 46.7% and 84.0% (55.5 ± 12.1, mean ± standard deviation) for lactation and activity features together. Using these individual farm models, only 3.5% and 2.3% of the cows were classified high while being low and vice versa. This analysis shows (1) the need to consider local (and evidence based) culling management rules when developing such decision support tools for dairy farms; and (2) the potential of precision phenotyping of complex traits using readily available sensor data from which biologically meaningful features can be derived. We conclude that first lactation milk and activity sensor data have the potential to predict cows' lifetime resilience but that consistency over farms is lacking.

Key Words: resilience; precision phenotyping; prediction model; longevity; precision livestock farming.

INTERPRETIVE SUMMARY. First lactation sensor data predicts resilience and productive lifespan but not with a common equation across farms. Adriaens. Increased longevity largely affects the sustainability of the dairy sector. Prediction of resilience as early as the first lactation allows for selection of cows to breed replacement heifers that cope well with the local management conditions. Using sensor features derived from daily milk yield and activity data of the first lactation on farms with an automatic milking system allowed to predict a resilience ranking, but the variability over different farms was too high to find a common equation across the farms.

PREDICTION OF RESILIENCE USING SENSOR DATA

INTRODUCTION

Recent technological developments led to increasing numbers of farmers implementing sensor systems and automation to improve their herd management and reduce labour requirements (Steeneveld and Hogeveen, 2015). While many of these sensor systems are currently used for the detection of health problems or fertility events, their full potential for management support on farm is not yet exploited. Today we see that the development of generally valid tools is hampered by the challenges imposed by data accessibility and variability, by technical sensor issues and by the differences in management style and farmers' preferences (Friggens and Thorup, 2015).

The most common sensor systems on dairy farms include milk meters and activity sensors, producing high-frequent measurements of respectively milk yield and cow behaviour. Milk yield dynamics can very informative to determine a cow's physiology and how she copes with challenges: as a modern dairy cow's system is highly predicated to milk production, each change in feed intake or energy allocation e.g. for producing an immune response may result in altered milk yield dynamics (Ben Abdelkrim et al., 2019). These perturbations, their recovery rate and associated milk losses cannot be derived from lower-frequent time series such as these collected in

national milk recording herd health programs (ICAR International committee for animal recording, 2014). Similarly, activity meters inform farmers on potential oestrous periods, but might also provide more general health and fertility characteristics that describe their ability to cope and still express behaviour under stress when the dynamics in these time series are properly extracted (Rutten et al., 2013).

Additional benefits of sensor technology will come from the calculation of precision phenotypes and their use for the characterization of overall and relative performance of the animals within the farm context and compared to herd mates. To this end, meaningful sensor features (**SF**) should be derived from the high-frequency time series as proxies of the targeted traits. This requires biologically sound data processing techniques to provide valid interpretation, and contextualization of the derived features. Precision phenotypes based on sensor data not only allow for more detailed estimation of the classical traits such as milk production and fertility (Royal et al., 2000; Tenghe et al., 2015; Sorg et al., 2017), but also have the potential to characterize complex traits such as lifetime performance, resilience or feed efficiency. Accordingly, when sensor data can be used to this purpose, selection of animals on these more complex traits becomes possible, which when combined with the genetic merit of each animal can boost future breeding efforts at farm and population level.

To increase the sustainability of dairy farms, resilience and the ability to reproduce are extremely important (e.g. for the fulfilment of societal demands and to reduce the impact of the sector in the modern socio-ecological context). Moreover, a dairy heifer only starts to make profit for the farmer during the second lactation and she only reaches her full production potential during her third parity (Cabrera, 2018). Breeding animals that have a high probability of completing several lactations within a specific farm context (e.g. management conditions, infection pressure, feed quality) will therefore gain importance in the modern competitive and high-demanding dairy sector. Accordingly, the driving hypothesis of this study was that we can predict a complex trait

such as resilience from first parity sensor data in order to help a farmer in his advanced breeding (e.g. sexed semen, embryo transfer, beef semen) and culling decisions. If so, this would allow the farmer to identify valuable animals to breed the replacement heifers from as early as after the first lactation, that are specifically expressing their genetic potential phenotypically in terms of these more complex traits in the herd environmental context. As such, there is still time to take breeding decisions that directly contribute to the farm's efficiency by selecting animals which perform well on that particular farm.

The objective of this study was to develop a model that uses first parity SF as proxies for performance, health and fertility to predict their lifetime resilience and recalving ability on farm. To this end, we analysed a dataset of 27 British and Belgian dairy farms and developed mathematical rules to derive meaningful SF from the high-frequency time-series data. These SF reflect the overall cow performance, her health and fertility contrasted to her herd peers.

MATERIALS AND METHODS

Data Collection and Selection

From a database of 34 Belgian and 42 British farms, 27 farms were selected based on (1) the accessibility and reliability of at least 5 years of contiguous data, and (2) the availability of daily milk yield or activity records. All these 27 farms had automated milking systems (**AMS**) of Lely (Lely Industries N.V., Maasluis, the Netherlands; No. = 16) or Delaval (DeLaval International, Tumba, Sweden; No. = 11). On average 2.4 lactations were recorded per cow's life. All farms had intensive production systems with cows kept indoors and fed with both forage and concentrates. Other management practices differed among herds but were not further documented.

All data tables were extracted from the restored back-up files of the management AMS software using SQL Server Management Studio (Microsoft, Redmond, WA, United States). The further data mining, pre-processing and merging of these data tables and the rest of the analysis

described below was done in Matlab R2017a (The MathWorks Inc., Natick, MA, United States). In this study, we used both the full dataset of 27 farms all having daily milk records (data set 1, **DS1**) and a subset of 13 farms also having daily activity data available (data set 2, **DS2**, all milked by a Lely AMS). An overview of the characteristics of both data sets is given in **Table 1**.

After extraction of data tables, cows were selected based on the availability of data for their entire lifetime production. Because exact culling dates were not always available, we elected to apply a criterion to discriminate between cows that had been dried off towards the end of the available data set for that farm (not to be included in the analysis) or removed from the herd (to be included in the analysis). For each farm, the 95% confidence interval on the average dry period length was calculated. If the last milk record was before the end of the data set minus the upper 95% CI boundary, that cow had 97.5% chance that she was removed from the herd and she was included. Accordingly, only cows were selected that met this criterion and for which the date of her first calving was within the timespan of the available data for that farm. An overview of the characteristics of this selection is provided in **Table 1**. Data set 1 consisted of 3754 unique cows and 9395 unique lactations, while DS2 included 2075 cows with 5286 lactations. Per farm, respectively 24 to 308 cows (139 ± 82, mean ± SD) with in total 44 to 799 lactations (348 ± 229) and 57 to 308 cows (160 ± 84) with 113 to 799 lactations (407 ± 264) were selected for DS1 and DS2.

The milk yield sensor data were recorded by the AMS using ICAR-approved milk meters as integrated in the Lely and Delaval robots. The available activity sensor data were recorded by Lely (Lely Industries N.V., Maasluis, the Netherlands) and SCR (SCR Engineers Ltd., Netanya, Israel) activity sensors and consisted of raw 2-hourly measures that were aggregated in daily activity records, but no further details of the individual sensor systems were available.

Calculation of Resilience Ranking

For this study, the lifetime resilience of a cow was considered primarily as the cumulative result of her ability to recalve (and thus, to extend her productive lifespan) supplemented with secondary corrections for age at first calving, calving intervals, 305-day milk yield, health events and number of inseminations (Friggens and De Haas, 2019). With a high weight given to each newly started lactation (i.e. parity number), the additional secondary corrections mainly allow discrimination between all cows reaching a certain parity. For example, for two given cows that first calve at 24 months of age, who both reach the second parity and are inseminated twice to get pregnant and who both have a 305-day milk yield of 8500 kg and no health events on the same farm, the one with a calving interval of 400 days will rank higher than the one with a calving interval of 410 days. This definition was agreed upon within the H2020 GenTORE consortium consisting of researchers, animal experts, veterinarians, technology suppliers and geneticists and is further detailed in (Friggens and De Haas, 2019). Because the number of inseminations and health events were not consistently available for all herds over the entire time period, the final equation used in this study for calculating resilience scores (**RS**) was (Eq. 1):

$$
\begin{aligned}
RS_i = \overline{CI} & \\
&+300 * L_i \\
&+(730 - AFC_i) \\
&+\sum_{j=1}^{L_i-1}\left(\overline{CI}_j - CI_{i,j}\right) \\
&+\sum_{j=1}^{L_i}\left(\frac{\sum_{k=1}^{DIM_{i,j}}\left[MY_{i,j,k}(k \leq 305)\right]}{\sum_{k=1}^{DIM_{i,j}}\left[\overline{MY_{l,J,k}(k \leq 305)}\right]} - 1\right) * 100 \\
&+\left[\left(100 - DIM_{i,L_i}\right)\left(DIM_{i,L_i} < 100\right)\right]
\end{aligned}
\qquad \textbf{(Eq. 1)}
$$

With:

RS_i	Resilience score for cow *i*
$\overline{CI}$	Average calving interval of the herd

L_i	Lactation number in which cow *i* exited the herd (last lactation number of a cow)
AFC_i	Age at first calving of cow *i* in days
$CI_{i,j}$	Calving interval of cow *i* between the start of lactation *j* and (*j* + 1)
$\overline{CI}_j$	Average calving interval between lactation *j* and (*j+1*) in the herd
$MY_{i,j,k}$	Milk production (in kg) of cow *i* at day *k* of lactation *j*
$\overline{MY_{i,j,k}}$	Average milk production (in kg) of cow *i* at day *k* of lactation *j*
$DIM_{i,j}$	Days in milk (**DIM**) of cow *i* at the end of lactation *Li*

This way, each RS consists of a (1) baseline equal for all cows in the herd to avoid negative resilience scores (this does not contribute to the ranking); (2) a bonus given for each recalving (newly started lactation) equal to 300 points; (3) a penalty score given to cows older than 24 months at their first calving equal to 1 point per day longer than 730 days (i.e. 24 months); (4) a penalty or bonus score equal to the number of days the calving interval is shorter or longer respectively than the average calving interval of the same parity in the herd; (5) a penalty or bonus score equal to the percentage the 305-day milk production is lower or higher respectively compared to the average 305-day production of the corresponding parity for all lactations in the herd; and (6) a penalty score for cows exiting the herd before day 100 in lactation equal to 100-DIM_{exit}, assuming that these cows are involuntarily removed from the herd.

Based on this score, which assumes that these factors reflect the accumulated effects of the cows' resilience, cows were ranked within farms. For example, if the average CI of a herd is 400 days, the average first parity CI is 380 days and the average 305-day milk production in first lactation is 8000 kg, then a cow that calved twice (the first time aged 24 months and 45 days), had a calving interval of 420 days between first and second lactation, produced 5% more milk than average in the first 305 days of the first lactation (8400 kg) and 20% less than her herd peers in the

first 80 days of the second lactation, and that was culled day 80 in the second lactation will receive a resilience score of RS = 400 + 300 * 2 - 45 + (380 – 420) + 5 – 20 = 900 points. Because the weights in the RS are based on expert knowledge and the main interest is to distinct the least from the most resilience animals, the RS was converted into an on-farm resilience rank (**RR**) to rank the cows for their lifetime resilience performance. In these RR, high ranked cows ('highly resilient animals') represent animals recalving many times, having the (theoretically) optimal age at first calving and short calving intervals and thus good reproductive performance compared to their herd mates of the same parity, and producing a good amount of milk compared to their herd mates. The number of lactations, affects this ranking the most because of the high weight for each new lactation started. Before entering the RR in the models, the following scaling at farm level was applied to ensure the ranking for each farm varied between 0 (i.e. the first ranked cow) and 1 (i.e. the last ranked cow), and thus, no scale effects would influence the prediction models: $RR_{scaled} = (RR-RR_{min})/(RR_{max} - RR_{min})$ with RR_{min} equals 1 and RR_{max} equals the maximum rank (equivalent to the number of cows included in the ranking for that farm).

Sensor Features

Milk Yield Sensor Features. The time-series data of two sensors were included in this study: (1) milk meter sensors from which daily milk yields were calculated and (2) activity sensors from which the two-hourly raw data were aggregated into daily activity records. Sensor features (**SF**) were calculated for each of the cows for which the first lactation was longer than 200 days (because 200 days is enough to grasp a good image of the time-series dynamics). Only the data of the first 305 days in lactation were included for the calculations.

In total 30 milk yield SF were calculated for each first lactation in the following categories: (1) lactation shape characteristics including peak yield, consistency, days in milk of peak, etc. ; (2) goodness-of-fit and variability measures including the characteristics of lactation model residuals;

and (3) perturbation features. For these calculations, we started from a theoretical lactation curve shape by iteratively fitting a Wood curve on the daily milk yield time series using $\boldsymbol{TMY} = A*e^{-B*DIM}*DIM^{C}$ with *A*, *B* and C the parameters of the Wood model, *TMY* is the total daily milk yield in kg and *DIM* the days in milk expressed in days (Wood, 1967). During each iteration, first the Wood model was fitted and the residuals calculated by subtracting the model from the milk yield data. Next, all the residuals smaller than 85% of the theoretical curve (i.e. Wood's model) were removed and the model was refitted in a next iteration. This procedure was repeated until the difference between two iterations of the average root mean squared error (**RMSE**) was smaller than 0.10 kg or for at most 20 iterations. An example of the daily milk yield data, the iterated Wood model and the corresponding residuals is shown in **Figure 1**.

Next, the final parameters of Wood's model (*A*, *B* and *C*), the residuals of all daily milk yield records and the periods identified as perturbations from these residuals were used to calculate the lactation SF. For the latter, major events (i.e. periods of at least 10 days of successively negative residuals with at least one day of milk production lower than 80% of the theoretical production) were discriminated from minor events (i.e. periods of at least 5 days of successively negative residuals with at least one day of milk production between 90% and 80% of the expected production). It was assumed that large perturbations (major events) may represent severe health problems which might influence culling and rebreeding decisions, while smaller (minor) perturbations are probably linked to chronical or subclinical infections, with a different effect on culling or longevity. A detailed description of the milk yield SF and how they were calculated can be found in **Appendix A**.

Activity Sensor Features. In addition to the lactation SF, for DS2 also activity SF were calculated. Fifteen different SF were defined in the following categories: (1) features related to the absolute (within-herd) levels, i.e. variability and autocorrelation; (2) fertility-related characteristics based on short spikes representing oestrous behaviour; and (3) overall activity-related

characteristics based on changes in average activity during longer periods of time which possibly relates to e.g. health events. To identify the short spikes of the second category, a median smoother using a window of 4 days was used and subtracted from the raw daily activity data to obtain residual activity levels, while for the identification of the longer-term patterns in the data, a 20 day-window median smoother was applied. The details for the activity SF calculations are given in **Appendix B**.

Standardization of the Sensor Features. Before entering the SF in the models, each SF was standardized within herd using mean centring and by dividing them by the within-herd standard deviation to correct for differences in their order of magnitude and for interpretability issues. The objective was to develop a tool to evaluate and forecast the (phenotypic) performance of an animal in the herd early in her productive life to still have time to take breeding decisions that would directly contribute to the farms' performance, and so that 'high risk' animals might be monitored closer. Therefore, in this study only the first parity SF were taken into account as proxies for performance, health and fertility to predict their lifetime resilience and re-calving ability on farm. Sensor features (both milk yield and activity) with values outside of the average plus or minus three times the within-farm standard deviation were considered as outliers and were replaced by the average values (i.e. zero) to avoid missing and unbalanced data.

Exploratory Analysis

In this study, a model was sought to predict the RR of all the animals on a specific farm. Ideally, a common model structure that is valid for all farms would be obtained, as this would allow the calculation of a limited and universal number of SF indicative for the animals' resilience and longevity. As a first step, the Pearson linear correlation coefficient between each SF and the RR at individual farm level was calculated. Highly positive and negative correlations would indicate a

strong effect of that SF on the RR, and thus a potential candidate for inclusion in further prediction models.

In a second step, mutual correlations between the SF were explored for all farms together. This initial data exploration using data of all farms together pointed out some significant (but small) linear correlations between the SF. However at individual farm level, these correlations were often inconsistent. To investigate whether a underlying latent structure existed in the SF and avoid future multi-collinearity in the prediction models, a principal component analysis (PCA) was carried out on both DS1 and DS2. These PCA showed that respectively 8 and 24 principal components with eigenvalues higher than 1 (Kaiser criterion) explained only 71% and 74% of the variance, suggesting that a latent structure for data reduction over all farms did not exist.

Model Development

Several multivariate modelling techniques including partial least squares and general linear mixed models were tested, but all had poor prediction performance or showed significant overfitting of the data. Ultimately, a separate multivariate linear regression model relating the SF to the RR within farm was constructed as follows (Eq. 2):

$$RR_{scaled} = \beta X + \varepsilon \qquad \textbf{(Eq. 2)}$$

With RR_{scaled} the scaled RR between 0 and 1 as defined above. The β vector contains the regression coefficients for the standardized SF in the design matrix X and ε are the residual errors. A backward stepwise regression procedure was used to identify redundant SF in X applying a p-value of 0.2 as the in- and exclusion threshold. The chosen threshold might seem uncommonly high but given the high variability in the SF both between and within farms, we deemed relevant to include any feature having a tendency towards significance.

Ten-fold cross-validation (**CV**) was performed to evaluate the prediction performance of the obtained models and identify overfitting. To this end, ten times all the cows of each farm were

assigned to a calibration (90% of the animals) or validation (10% of the animals) set using random sampling from a uniform distribution, but using the additional criterion that both the calibration and the validation set contained at least one animal ranked in the first, one in the middle and one in the last 33%. In each CV-cycle, the cows in the calibration set were used to estimate the regression coefficients β and the obtained model was used to predict the RR_{scaled} of the cows in the validation set. The average prediction results over all ten CV-cycles were considered to represent the final model performance.

Model Evaluation

The initial model fit at farm level was evaluated using the RMSE and the $R^2_{adjusted}$, calculated as (Eq. 3):

$$R^2_{adjusted} = 1 - \frac{\left(1-\left(1-\frac{SSE}{SSTO}\right)\right)(n-1)}{1-(k-1)} \qquad \textbf{(Eq. 3)}$$

With *SSE* = residual sum of squares of the regression, SSTO the total sum of squares (i.e. the mean value of the outcome RR_{scaled}), *n* the number of data points of each farm and *k* the number of SF retained in the final model for that farm. For the herds in DS2, separate models both including and excluding the activity SF were built.

To evaluate classification performance and discriminate between high and low ranked cows (which is of practical relevance), the cows of each farm were divided into three different categories based on their ranking: first (**F**; best 33%), medium (**M**; middle 33%) and last (**L**, worst 33%) resilient animals. When the predicted RR_{scaled} did not cover the full range of 0 to 1, and to be able to calculate these high, medium and low ranked categories for each farm, a farm-individual correction factor was applied on the predicted RR_{scaled} scores as follows (**Eq. 4**):

$$\widehat{RR}_{scaled,corr,i} = \frac{\widehat{RR}_{scaled,i} - \mathrm{A}}{\mathrm{B} - \mathrm{A}} \qquad \textbf{(Eq. 4)}$$

With $\widehat{RR}_{scaled,i}$ equal to the predicted RR_{scaled} of the ith cow and A and B are farm-specific coefficients representing respectively the minimum and maximum of all the predicted RR_{scaled} for that farm in the calibration set of each CV-cycle. The RR_{scaled} of the cows in the validation set of each CV-cycle were predicted using each individual farm model (**Eq. 2**) and their category (*F*, *M*, *L*) was determined after applying the correction using the farm-specific coefficients (**Eq. 4**). Both the root mean squared error of cross-validation (**RMSECV, Eq. 5**) and the classification accuracy were evaluated in CV to assess the models' prediction performance.

$$RMSECV = \sqrt{\frac{1}{N}\sum_{1}^{N}\left(RR_{scaled,i} - \widehat{RR_{scaled,i}}\right)^2} \qquad \textbf{(Eq. 5)}$$

To evaluate whether a common model structure across farms could be identified or whether specific features are highly correlated to resilience in all farms, we evaluated the overlap in retained features for each farm, both in terms of their inclusion or exclusion in each model and the sign of their regression coefficients.

Prediction performance improvement of the models including and excluding activity features was assessed using a one-sided paired *t*-test on the percentage correctly classified using the null hypothesis "*activity features do not improve (i.e. increase) the percentage correctly classified animals*" and on the proportion oppositely classified using testing the null hypothesis "*activity features do not improve (i.e. decrease) the percentage oppositely classified animals"*.

RESULTS

Data Overview and Summary Sensor Features

Table 2 gives a summary of the most important SF over all farms for DS1 and DS2. As can be seen from the minimum and maximum value of each SF, there are some extremes present in the data sets which might result from the way each SF is calculated rather than from real deviating curves. As indicated in the materials and methods section, these outliers were excluded from the analysis.

Figure 2 shows the lifetime RR plotted against the last lactation of each cow for one farm as an example (first ranked cows have higher last lactation numbers). This figure confirms that in general, the number of lactations has the biggest impact on the final resilience ranking. The cows ranked lower than their herd mates with higher last lactation number are seen as spikes in **Figure 2**. These are mainly cows that exit their herds very early in a new lactation, thus for which DIM_{i,L_i} in **Eq. 1** plays a significant role.

Predicting Resilience Ranking – DS1

The Pearson linear correlation coefficients between the lactation SF and RR are shown in **Figure 3**. Average correlations per SF over all farms varied between ρ = -0.134 and ρ = +0.1492, although some of the individual features showed correlations of more than +/- 0.4 for some of the farms. Visual exploration of the correlation scatter plots (results not shown) did not show clear non-linear relationships either. What mainly stands out is the marked variability in correlations between farms and the inconsistency in sign of the correlations (negative vs. positive). The highest and most consistent correlations are seen for SF representing model fit and size of the residuals (RMSE of the Wood model, number of residuals below 85% of the predicted value, average size of the 3 largest negative residuals).

The stepwise multilinear regression showed that the main SF associated with the RR were the goodness-of-fit of the estimated Wood curves (SF21, 22, 25, 27, 28, 29), the size and number of perturbations (SF13, 14, 15), and their associated milk losses (SF11, 12, 18). The lactation curve characteristics (peak height and DIM, slopes, rate of the increasing phase of the lactation and persistency of the lactation after the peak) were included in only 12 out of 27 farms, indicating that for example having a high milk production in the first lactation compared to herd mates barely

affects the RR, i.e. the ability to recalve in later lactations and longevity. Another feature often retained in the models is the loss during the minor perturbations. These can be interpreted as for example subclinical or chronic infections, which also can be a reason to cull the cows.

The $R^2_{adjusted}$ of the individual stepwise multivariate linear regression models varied between 0.03 and 0.61 (0.22 ± 0.16, mean ± SD) and the RMSE was between 0.17 and 0.27 (0.23 ± 0.03, mean ± SD). These models included 2 to 12 milk yield features (**Table 3**). All SF were included at least once in one of the models of the individual farms. The variability between farms is demonstrated by the fact that for only 2 out of the 30 SF the regression coefficients were consistently above or below zero, and so had a consistently positive or negative effect on the RR. Within-farm CV showed similar performance (RMSECV 0.24 ± 0.03) to the initial models with all animals included. On average 46.7 ± 8.0% of the animals were classified in the correct F, M, L category, and on average 4.4 ± 3.5% of the cows were classified 'high' where they should have been 'low' and vice versa. Also here the large differences in model performance between the different farms stands out (range correctly classified: [35.8% ; 70.0%], range oppositely classified [0.0% ; 16.5%]). When looking deeper into which cows are predicted in the right category, it was found that the models do not solely predict animals exiting the herd after the first lactation correctly but also cows exiting the herd in a later lactation.

Predicting Resilience Ranking – DS2

Data set 2 consisted of 13 farms for which, besides milk yield features (No. = 30 features), activity features (No. = 15 features) could also be calculated. These features included both general characteristics of the daily activity (skewness, variability, absolute daily level) and specific features associated with short term and longer period activity changes. Farm-individual Pearson correlation coefficients (**Figure 4**) between the activity features and the RR varied between ρ = -0.41 and ρ = +0.44 and only the number of activity peaks in L1 (SF34) was consistently associated with a higher

ranking (lower number of peaks is lower RR, ρ = 0.29 ± 0.10, [0.16 ; 0.44]). Several other activity features also had correlations almost consistently above or below zero, but these correlations stayed relatively low on average.

The stepwise linear regression models had R^2_{adj} between 0.2 and 0.76 with 6 to 24 SF included (both activity and milk yield features), which agreed with RMSE between 0.128 and 0.24. The number of activity features retained in the final models was between 2 and 10, so the activity sensors seemed to be of added value for all of the farms in predicting their RR. Including activity features gave a higher R^2_{adj} and a lower RMSE in the calibration, while the number of features retained was sometimes higher and sometimes smaller. There was very little consistency over the different farms in which features were included in the final models. None of the SF was kept in the model of all farms. The number of activity peaks and days in milk of the first peak were retained most often (respectively 8 and 11 out of 13 times) and with a consistently positive regression coefficient. Three of the SF (6.6%) were never retained in any of the individual farm models.

The cross validation, using the same training and validation sets for these farms as in DS1, showed reasonable performance, with an RMSECV of 0.22 ± 0.03 [0.15 ; 0.26] for the validation sets. On average 55.5 ± 12.1% [43.5% ; 84.0%] of the cows were predicted in the correct category (*F*, *M*, *L*) and 2.3 ± 2.1% [0.0% ; 6.7%] of them were predicted in ‘first’ where they were actually ‘last’ or vice versa. Over all the farms, including activity features improved the correct classification with 9.3 ± 7.9% (p-value < 0.001). The classification worsened in only two farms compared to when only milk yield features were included. The proportion classified in the opposite category decreased with on average 3.5 ± 4.5% from 5.9% to 2.3% for these farms (significant difference, p-value 0.0086).

DISCUSSION

Study Opportunities and Limitations

In this study, the possibility to predict ‘resilience’ from both lactation and activity features of the first lactation was explored. To assess resilience, all cows on farm were given a score at culling taking the number of lactations, their 305-day milk yield, their age at first calving and their calving intervals into account. This was done according to the definitions put forward by the H2020 GenTORE consortium and described in (Friggens and De Haas, 2019). The idea is that resilient cows are animals that have long productive lifespans, are highly fertile and conceive easily and produce well with high adaptability to for challenges. Ideally penalties or bonus points for health status and the number of inseminations needed would also be included, but the health and insemination records available were currently not sufficiently complete over the whole time period for all farms. Consistent and correct registration, collection, mining and storage of data remains an impactful challenge on many (dairy) farms (Hudson et al., 2018).

The lactation and activity data of respectively 27 and 13 commercial dairy operations with an AMS was studied. This is a unique dataset; the large amount of high-frequency sensor data allowed for the inclusion of milk yield perturbations and daily activity dynamics, while the availability of at least 5 years of successive data uniquely permitted to study the accumulated effects of health and fertility traits over the lifetime of many animals. Accordingly, the present study represents only a start of what can be investigated using real farm high-frequency (e.g. AMS) data. Besides providing a unique opportunity, the commercial origin (in contrast to a test-farm origin) of these datasets also poses considerable challenges. For example, the RR currently studied is highly influenced by the farmer-applied management of culling, reproduction and health on farm. Although in the short term this management might be considered fixed, in the long term the motivations and preferences of the farm staff, the economic context, the animals’ genotypes and phenotypes, the farm facilities, the feed etc. are probably highly dynamic. As the resilience ranking is affected by longevity, reproduction and health performance, non-constant and unknown management factors inevitably complicate the analysis and cannot be compensated for by including additive herd-year effects in

the linear models. These management factors differ both within and between farms. Both aspects are reflected in the presented results: the linear correlations between the SF and the individual rankings strongly differed between farms, and the prediction accuracy across farms was rather low. Moreover, a common structure for the prediction models in which the same features were included for each farm could not be found, while the initial hypothesis was that similar SF would have similar effects on the RR for each farm.

Predicting Productive Lifespan and Resilience from Sensor Data

Reliable prediction of resilience ranking within a farm would allow for a more evidence-based approach to the management actions concerning advanced breeding (e.g. sexed semen, embryo transfer) or culling decisions after the first lactation (Vandeweerd et al., 2012). In this way, breeding decisions for cows in the second parity and higher could be taken using both the genetic/genomic (available once the animal is born) and phenotypic information (once an animal completed her first lactation). As such, sustainable productivity from the available animals on farm is optimized, while in the meantime also the 'overall' phenotypic information on complex traits for future breeding goals is obtained at herd level. This will allow future identification of animals and sires that perform well in many different environments. Moreover, targeting low ranked animals for increased monitoring purposes also becomes possible. In practice, it would be enough to discriminate between 'first' and 'last' ranked cows. The exact ranking is of less importance because (1) the scoring system is artificial and defined using expert knowledge and (2) a farmer's decision would not generally be different for e.g. the 5th vs. the 10th ranked cow in the herd. For example, highly resilient cows would be selected for advanced breeding techniques (e.g. ovum pick-up and in vitro production, multiple ovulation embryo transfer, sexed semen), while the lowest ranked cows

would be inseminated with beef semen and not used to breed replacement heifers (Mapletoft and Hasler, 2002; Boichard et al., 2015; Cabrera, 2018).

Today, cows exit the herd for many different reasons, for which the most common are (1) poor reproduction performance, (2) udder health problems, (3) metabolic disorders in early lactation and (4) claw health and locomotion disorders (Ahlman et al., 2011; Santos et al., 2016). In this study, it was assumed that these problems are at least partly reflected in the high-frequency sensor data, for example through perturbations in lactation curves or in the absence of activity peaks. The included SF were therefore initially defined from expert knowledge on how these disorders could affect the sensor time series. The SF retained most often in the models suggest indeed that most farmers do take severe health problems, reflected in the milk yield perturbations into account when taking culling or re-insemination decisions. However, different culling motivations might be reflected differently in the sensor data and in some cases only the combined effect of different features or only the extreme values might be linked with productive lifespan and RR. For example, a cow with severe clinical mastitis is likely to show a large and sudden drop in milk yield (Rajala-Schultz et al., 1999; Gröhn et al., 2004; Andersen et al., 2011) while a cow with subclinical mastitis may have a relatively normal lactation curve which can be well modelled with a lactation model; both can have impact on reproduction performance and the probability of culling (Lavon et al., 2010; Wathes, 2012; Wolfenson et al., 2015), but mathematically capturing the differences without more complete information about the cows is extremely difficult. So, that the correlations found in this study are rather weak is logical, because one might expect that for the cows culled in the first lactation, health reasons can be a major factor (Pinedo et al., 2010). However, for cows culled at a later stage, it might not be expected that first lactation features are predictive for the RR, unless these characteristics are highly repeatable over time or represent chronic or repeating conditions that fail to cure. The presence of outliers in the SF data also demonstrates that mathematically defining features is not always error-free; the numbers should be interpreted with care. Unfortunately, no

simple solution exists to solve this issue or take these complex interactions into account. In smaller and thoroughly selected datasets, future research might zoom in to the relation between curve characteristics and registered health events.

Despite the variability between and within farms and the fact that we could not find SF that were generally informative to predict RR over all farms, the prediction and classification performance of the individual farm models was in many cases significantly higher than the product of a random classification (i.e. 33%). Further, including activity features demonstrated a significant added value compared with using the daily milk yield features alone (p-value < 0.001). A correct classification of up to 84% of the animals suggests that at least part of the variability contained in the RR is correctly captured by the SF. The poor prediction performance on some other farms can be explained by (1) the fact that only features of the first lactation were included; (2) the fact that resilience is based on the limited data available in the commercial situation; (3) the large differences in the number of animals included per farm in the analysis; (4) the large time span over which animals were included; (5) the differing management practices over this timespan, as discussed previously. This observation suggests mainly that the need for tools to support more consistent and evidence based breeding and culling decisions should also recognise the need to take farm-specific characteristics (e.g. young stock, but also robot or parlour capacity, etc.) and farmers preferences (large focus on fertility, invisibility in the herd, udder health, robustness, etc.) into account. Furthermore, although first lactation features may not consistently predict lifetime resilience rank on all farms, SF may offer decision support for management changes that are measurable in the subsequent lactation.

Model use and practical implications

The results of this study underpin the need for more consistent and evidence-based decision making on one hand, while on the other hand show the potential of high-frequency sensor data to

provide vital information on the phenotypical performance of the cows on a specific farm, which can be combined into complex traits such as resilience. They both need dedicated data processing in which the biology of the cows within their farm contexts is taken into account in order to identify relevant traits. Before decision support and rationalization is possible, broader farm context measures should be included in the model and monitoring systems, such as key indicators of management for reproduction performance, health and treatment records and herd-level aspects. The high-frequency sensor data allows for monitoring of the animals' response to challenges through the characterisation of perturbations. As a result, ranking the cows using our process and biological knowledge could already make a huge difference for many farms, because the rationalization of decisions may bring consistent and more economically-sound and sustainable management actions.

CONCLUSIONS

In this study, we demonstrate that resilience ranking and productive lifespan of modern dairy cows on AMS farms in Belgium and the UK could be predicted using farm-individual models based on first lactation sensor data. With the milk yield and activity SF selected at farm level, we reached classification performances (poor, moderate, high resilience) of up to 84%, while only 2.3 ± 2.1% (mean ± SD) of the cows were predicted in the opposite category. This shows the potential of high-frequency milk yield and activity sensor data to rationalize evidence-based breeding and culling decisions. However, a common model structure across all farms could not be found, which shows the variability between farms and which highlights the need for biologically sound and context-dependent data processing tools. Once a resilience-predicting tool is established, the farmer and the livestock sector could not only benefit from them at management and decision support level, but also at genetic level in the context of new precision phenotyping proxies for complex traits.

ACKNOWLEDGEMENTS

Ines Adriaens received funding from the Research Foundation Flanders through grants No. 11ZG916N and V423719N and from a KU Leuven postdoctoral mandate grant No. PDM/19/132. Part of the data were collected in the context of the VLAIO LA-trajectory 'MastiMan', grant No. HBC.2016.0774 by DVM Igor Van den Brulle and dr. DVM Sofie Piepers from the University of Ghent, Flanders (Department of Reproduction, Obstetrics and Herd Health, M-Team and Mastitis and Milk Quality Research Unit). Also Katherine Lumb (RAFT Solutions Ltd., Ripon, United Kingdom) contributed to the data collection. This work is part of the GenTORE project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 727213. We thank dr. Carmen Adriaens (Bernstein Laboratory, Harvard, Boston, United States) for her critical reading of the manuscript.

REFERENCES

Ben Abdelkrim, A., L. Puillet, P. Gomes, and O. Martin. 2019. Lactation curve model with explicit representation of perturbations as a phenotyping tool for dairy livestock precision farming 661249. doi:http://dx.doi.org/10.1101/661249.

Ahlman, T., B. Berglund, L. Rydhmer, and E. Strandberg. 2011. Culling reasons in organic and conventional dairy herds and genotype by environment interaction for longevity. J. Dairy Sci. 94:1568–1575. doi:10.3168/jds.2010-3483.

Andersen, F., O. Østerås, O. Reksen, and Y.T. Gröhn. 2011. Mastitis and the shape of the lactation curve in Norwegian dairy cows.. J. Dairy Res. 78:23–31. doi:10.1017/S0022029910000749.

Boichard, D., V. Ducrocq, and S. Fritz. 2015. Sustainable dairy cattle selection in the genomic era. J. Anim. Breed. Genet. 132:135–143. doi:10.1111/jbg.12150.

Cabrera, V.E. 2018. Invited review: Helping dairy farmers to improve economic performance utilizing data-driving decision support tools. Animal 12:134–144. doi:10.1017/S1751731117001665.

Friggens, N., and Y. De Haas. 2019. Resilience: reference measures based on longer-term consequences are needed to unlock the potential of precision livestock farming technologies for quantifying this trait. Available on www.gentore.eu.

Friggens, N.C., and V.M. Thorup. 2015. From Monitoring to Precision Phenotyping: Towards a Systemic Use of Precision Livestock Measures in Dairy Herds. Pages 145–148 in New Zealand Society of Animal Production.

Gröhn, Y.T., D.J. Wilson, R.N. González, J. a Hertl, H. Schulte, G. Bennett, and Y.H. Schukken. 2004. Effect of pathogen-specific clinical mastitis on milk yield in dairy cows.. J. Dairy Sci. 87:3358–74. doi:10.3168/jds.S0022-0302(04)73472-4.

Hudson, C., J. Kaler, and P. Down. 2018. Using big data in cattle practice. In Pract. 40:396–408. doi:10.1136/inp.k4328.

ICAR International committee for animal recording. 2014. ICAR Recording Guidelines. Int. Agreem. Rec. Pract. 619.

Lavon, Y., G. Leitner, H. Voet, and D. Wolfenson. 2010. Naturally occurring mastitis effects on timing of ovulation, steroid and gonadotrophic hormone concentrations, and follicular and luteal growth in cows. J. Dairy Sci. 93:911–921. doi:10.3168/jds.2009-2112.

Mapletoft, R., and J. Hasler. 2002. Assisted reproductive technologies in cattle: a review. Sci. Tech. Rev. Off. Int. des Epizoot. 24:393–403. doi:10.1057/palgrave.development.1100303.

Pinedo, P.J., A. De Vries, and D.W. Webb. 2010. Dynamics of culling risk with disposal codes reported by Dairy Herd Improvement dairy herds. J. Dairy Sci. 93:2250–2261. doi:10.3168/jds.2009-2572.

Rajala-Schultz, P.J., Y.T. Gröhn, C.E. McCulloch, and C.L. Guard. 1999. Effects of clinical mastitis on milk yield in dairy cows.. J. Dairy Sci. 82:1213–20. doi:10.3168/jds.S0022-0302(99)75344-0.

Royal, M.D., A.O.R. Darwash, A.P.F.F. Flint, R. Webb, J.A. Woolliams, and G.E. Lamming. 2000.

Declining fertility in dairy cattle: Changes in traditional and endocrine parameters of fertility. Anim. Sci. 70:487–501. doi:10.1017/S1357729800051845.

Rutten, C.J., a G.J. Velthuis, W. Steeneveld, and H. Hogeveen. 2013. Invited review: sensors to support health management on dairy farms. J. Dairy Sci. 96:1928–1952. doi:http://dx.doi.org/10.3168/jds.2012-6107.

Santos, J.E.P., R.S. Bisinotto, and E.S. Ribeiro. 2016. Mechanisms underlying reduced fertility in anovular dairy cows. Theriogenology 86:254–262. doi:10.1016/j.theriogenology.2016.04.038.

Sorg, D., M. Wensch-Dorendorf, K. Schöpke, G. Martin, R. Schafberg, N. Reinhold, S. Pache, and H. Swalve. 2017. Genetic analysis of new progesterone-based fertility traits in dairy cows measured on-farm. J. Dairy Sci. 100:8205–8219. doi:10.3168/jds.2016-11865.

Steeneveld, W., and H. Hogeveen. 2015. Characterization of Dutch dairy farms using sensor systems for cow management. J. Dairy Sci. 98:709–717. doi:10.3168/jds.2014-8595.

Tenghe, A.M.M., A.C. Bouwman, B. Berglund, E. Strandberg, J.Y. Blom, and R.F. Veerkamp. 2015. Estimating genetic parameters for fertility in dairy cows from in-line milk progesterone profiles.. J. Dairy Sci. 98:5763–5773. doi:10.3168/jds.2014-8732.

Vandeweerd, J.M., N. Kirschvink, P. Clegg, S. Vandenput, P. Gustin, and C. Saegerman. 2012. Is evidence-based medicine so evident in veterinary research and practice? History, obstacles and perspectives. Vet. J. 191:28–34. doi:10.1016/j.tvjl.2011.04.013.

Wathes, D.C. 2012. Mechanisms linking metabolic status and disease with reproductive outcome in the dairy cow. Reprod. Domest. Anim. 47:304–312. doi:10.1111/j.1439-0531.2012.02090.x.

Wolfenson, D., G. Leitner, and Y. Lavon. 2015. The disruptive effects of mastitis on reproduction and fertility in dairy cows. Ital. J. Anim. Sci. 14:650–654. doi:10.4081/ijas.2015.4125.

Wood, P. 1967. Algebraic model of the lactation curve in cattle. Nature 216:164–165.

APPENDICES

Appendix A – Definition and Calculation of Milk Yield Features

In Table A1, the details of the different sensor features (**SF**) included in the prediction models are given. '**ITW'** stands for Iterated Theoretical Wood model, which is the result of the iterative fitting and refitting procedure excluding perturbations to estimate the shape of the theoretical lactation curve. All SF are standardized before entering them in the models, by = $\mathrm{SF}_{\mathrm{standardized}} = \frac{SF - \overline{SF}}{SD\,(SF)}$ with SF each sensor feature, $\overline{SF}$ the average of each SF for a herd and parity (within herd), and *SD(SF)* the standard deviation for that SF for a herd and parity (within herd).

Table A1. Lactation sensor features (**SF**) and their calculation included in the prediction models for resilience ranking

SF No.	Name	No. of times (%) retained Lactation SF / activity & lactation SF	How calculated?
SF1	305-day yield	**6 (22.2%)** **4 (30.8%)**	Total milk yield in the first 305 days of lactation
SF2	RMSE ITW[1]	**5 (18.5%)** **2 (15.4%)**	Root mean squared error of the iterated theoretical Wood model
SF3	Peak yield ITW[1]	**5 (18.5%)** **3 (23.1%)**	Maximal value of the iterated theoretical Wood model
SF4	DIM[4] peak yield ITW[1]	**4 (14.8%)** **0 (0.0%)**	Days in milk of the peak milk yield (SF3)
SF5	Rate increase ITW[1]	**5 (18.5%)** **4 (30.8%)**	Peak yield (SF3) / days in milk to peak (SF4)
SF6	Persistency ITW[1]	**10 (37.0%)** **5 (38.5%)**	Persistency, slope of the linear decreasing phase of the iterated theoretical Wood model
SF7	Sum of strong negative residuals of ITW[1]	**2 (7.4%)** **4 (30.8%)**	Sum of the residuals smaller than 0.85 * the iterated theoretical Wood model
SF8	Ratio of positive vs. negative residuals of ITW[1]	**5 (18.5%)** **4 (30.8%)**	Ratio of the number of positive vs. negative residuals calculated from the iterated theoretical Wood model
SF9	Number of major[2] perturbations	**5 (18.5%)** **2 (15.4%)**	Number of perturbations with milk production at least one day below 80% of the iterated theoretical Wood model and lasting at least 10 days
SF10	Average number of days needed for recovery from major[2] perturbations	**5 (18.5%)** **3 (23.1%)**	Number of days needed for recovery from a major perturbation (milk production at least one day below 80% of the iterated theoretical Wood model and lasting at least 10 days)
SF11	Average milk loss during recovery phase of major[2] perturbations	**7 (25.9%)** **2 (15.4%)**	Average milk losses during recovery phase of a major perturbation (milk production at least one day below 80% of the iterated theoretical Wood model and lasting at least 10 days)
SF12	Average milk loss during development phase of major[2] perturbations	**6 (22.2%)** **3 (23.1%)**	Average milk losses during development phase of a major perturbation (milk production at least one day below 80% of the iterated theoretical Wood model and lasting at least 10 days)
SF13	Average number of days	**6 (22.2%)**	Number of days needed to reach the minimum of a major

	needed for development of minor[3] perturbations	**5 (38.5%)**	perturbation (milk production at least one day below 80% of the iterated theoretical Wood model and lasting at least 10 days)
SF14	Average minimum milk yield of the major[2] perturbations	**7 (25.9%)** **3 (23.1%)**	Average minimal milk yield of the major perturbations (milk production at least one day below 80% of the iterated theoretical Wood model and lasting at least 10 days)
SF15	Number of minor[3] perturbations	**6 (22.2%)** **1 (7.7%)**	Number of perturbations with milk production at least one day between 90 and 80% of the iterated theoretical Wood model and lasting at least 5 days
SF16	Average number of days needed for recovery from minor[3] perturbations	**5 (18.5%)** **1 (7.7%)**	Number of days needed for recovery from a minor perturbation (milk production at least one day between 90 and 80% of the iterated theoretical Wood model and lasting at least 5 days)
SF17	Average milk loss during recovery phase of minor[3] perturbations	**4 (14.8%)** **0 (0.0%)**	Average milk losses during recovery phase of a minor perturbation (milk production at least one day below 80% of the iterated theoretical Wood model and lasting at least 10 days)
SF18	Average milk loss during recovery phase of minor[3] perturbations	**6 (22.2%)** **2 (15.4%)**	Average milk losses during development phase of a minor perturbation (milk production at least one day below 80% of the iterated theoretical Wood model and lasting at least 10 days)
SF19	Average milk loss during development phase of minor[3] perturbations	**3 (11.1%)** **4 (30.8%)**	Number of days needed to reach the minimum of a major perturbation (milk production at least one day between 90 and 80% of the iterated theoretical Wood model and lasting at least 5 days)
SF20	Average minimum milk yield of the minor[3] perturbations	**5 (18.5%)** **5 (38.5%)**	Average minimal milk yield of the major perturbations (milk production at least one day between 90 and 80% of the iterated theoretical Wood model and lasting at least 5 days)
SF21	Number of periods in which milk yield drops below 85% of the expected yield (ITW) and that last > 5 days	**6 (22.2%)** **5 (38.5%)**	Number of periods in which milk yield drops below 85% of the expected yield (iterated theoretical Wood model) and that last > 5 days
SF22	Number of periods in which milk yield drops below 85% of the expected yield and that last > 10 days	**6 (22.2%)** **3 (23.1%)**	Number of periods in which milk yield drops below 85% of the expected yield (iterated theoretical Wood model) and that last > 5 days
SF23	Percentage of days in which milk yield drops below 85% of the expected yield (ITW)	**5 (18.5%)** **2 (15.4%)**	Number of periods in which milk yield drops below 85% of the expected yield (iterated theoretical Wood model)
SF24	Average milk loss of the 3 largest perturbations	**4 (14.8%)** **1 (7.7%)**	Average milk loss of the 3 largest perturbations
SF25	Largest negative residual	**6 (22.2%)** **1 (7.7%)**	Largest negative residual
SF26	Number of sign changes of the residuals throughout the lactation	**6 (22.2%)** **2 (15.4%)**	Number of sign changes of the residuals of the iterated theoretical Wood model throughout the lactation
SF27	Average residual	**9 (33.3%)** **7 (53.8%)**	Average of the residuals
SF28	Average absolute value of the residuals	**7 (25.9%)** **5 (38.5%)**	Average of the absolute values of the residuals
SF29	Variance of the residuals of the initial	**6 (22.2%)** **5 (38.5%)**	Variance of the residuals
SF30	Ratio of the initial Wood curve fitted on all data and the ITW[1]	**3 (11.1%)** **4 (30.8%)**	Ratio of the initial Wood curve fitted on all the data and the iterated theoretical Wood curve

[1] ITW: Iterated Wood Model

[2] Major perturbations: periods of at least 10 days of successively negative residuals with at least one day of milk production lower than 80% of the theoretical production

[3] Minor perturbations: periods of at least 5 days of successively negative residuals with at least one day of milk production between 90% and 80% of the expected production

[4] DIM: days in milk

Appendix B. Definition and calculation of activity features

First, the two-hourly activity data were aggregated in daily sums. Next, a moving median using a window of 4 days was calculated on these daily data time series to identify short spikes associated with oestrous behaviour (level 1; all spikes > 0.4* the maximal residual of the time series minus the moving median). A moving median of 20 days was calculated to identify periods with generally lower or higher activity (possibly associated with health events). A threshold of 20% of the minimal or maximal activity values was set to identify these altered activity period. The below explained calculated sensor features are based on the deviations from these median windows.

Table B1. Activity sensor features (**SF**) and their calculation included in the prediction models for resilience ranking

SF No.	Activity feature	No. of times (%) retained	How calculated?
SF31	Daily mean	**6 (46.2%)**	Daily mean activity value
SF32	Daily overall skewness	**1 (7.7%)**	Skewness of the daily activity time series
SF33	Daily overall variance	**9 (69.2%)**	Overall variance of the daily activity time series
SF34	Number of first level peaks	**11 (84.6%)**	Number of sharp peaks, larger than 0.4 times the maximal peak height after correction for 4-day median activity level
SF35	Average peak height of first level peaks	**5 (38.5%)**	Average peak height of peaks larger than 0.4 times the maximal peak height after correction for 4-day median activity level
SF36	Maximal peak height of first level peaks	**5 (38.5%)**	Maximal peak height of peaks larger than 0.4 times the maximal peak height after correction for 4-day median activity level
SF37	Days in milk of first peak	**8 (61.5%)**	Days in milk of first peak larger than 0.4 times the maximal peak height after correction for 4-day median activity level
SF38	Number of periods with higher activity compared to long-term median activity	**3 (23.1%)**	Number of periods with activity level higher than 0.2 times the maximal level compared to long-term 20-day median activity level
SF39	Average activity level of periods with higher activity compared to long-term median activity	**3 (23.1%)**	Average activity level of periods with activity level higher than 0.2 times the maximal level compared to long-term 20-day median activity level
SF40	Maximal activity level in periods with higher activity compared to long-term median activity	**2 (15.4%)**	Maximal activity level of periods with activity level higher than 0.2 times the maximal level compared to long-term 20-day median activity level
SF41	Duration of periods with higher activity compared to long-term median activity	**4 (30.8%)**	Duration of periods with activity level higher than 0.2 times the maximal level compared to long-term 20-day median activity level
SF42	Number of periods with lower activity compared to long-term median activity	**2 (15.4%)**	Number of periods with activity level lower than 1.2 times the minimal level compared to long-term 20-day median activity level
SF43	Average activity level of periods with lower activity compared to long-term median activity	**3 (23.1%)**	Average activity level of periods with activity level lower than 1.2 times the minimal level compared to long-term 20-day median activity level
SF44	Minimal activity level in	**0 (0.0%)**	Minimal activity level of periods with activity level lower than 1.2

	periods with lower activity compared to long-term median activity		times the minimal level compared to long-term 20-day median activity level
SF45	Duration of periods with lower activity compared to long-term median activity	**2 (15.4%)**	Duration of periods with activity level lower than 1.2 times minimal level compared to long-term 20-day median activity level

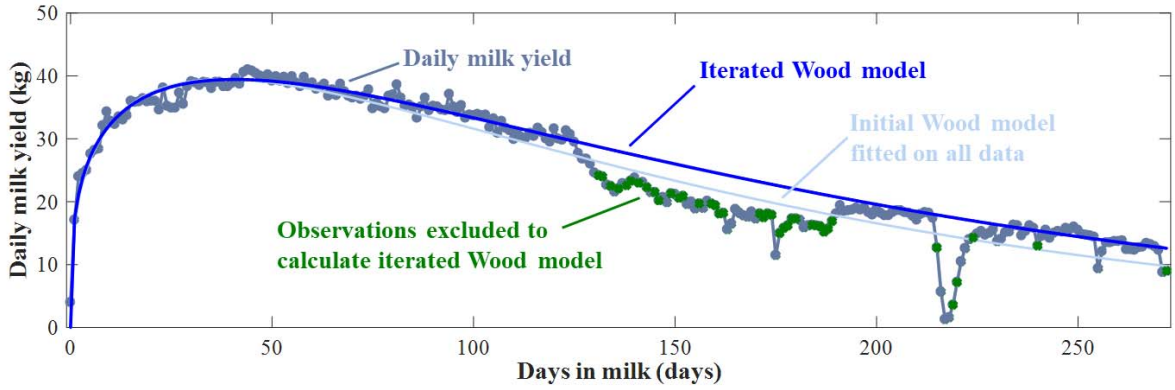
Figure 1. Example of a lactation curve with daily milk yields, the corresponding initial Wood model fitted on all data and the final Wood model fitted iteratively by excluding daily milk yields lower than 85% of the estimated curve.

Figure 2. Example of the relation between the resilience rank (**RR**) of a farm (No. of cows = 110) against the parity number in which each animal exits the herd.

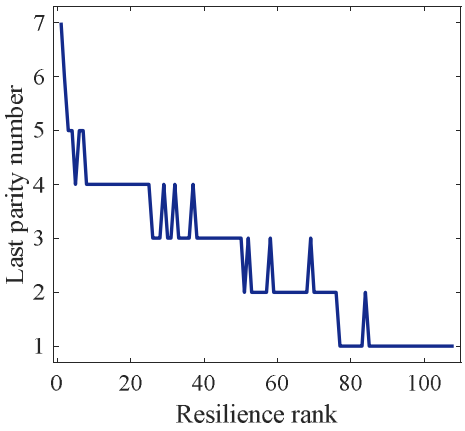
Figure 3. Pearson linear correlation coefficients between the resilience rank at farm level and the 30 lactation sensor features (**SF**) calculated on first parity data. Each individual thin line represents the correlations for a particular farm (total No. = 27). The shading represents the 95% confidence interval on the correlation coefficients over all farms. The details of the SF are described in Appendix A.

Figure 4. Pearson linear correlation coefficients between the resilience rank at farm level and the 15 activity sensor features (**SF**) calculated on first parity data. Each individual thin line represents the correlations of a particular farm (total No. = 13). The shading represents the 95% confidence interval on the correlation coefficients over all farms. The details of the activity SF are described in Appendix B.

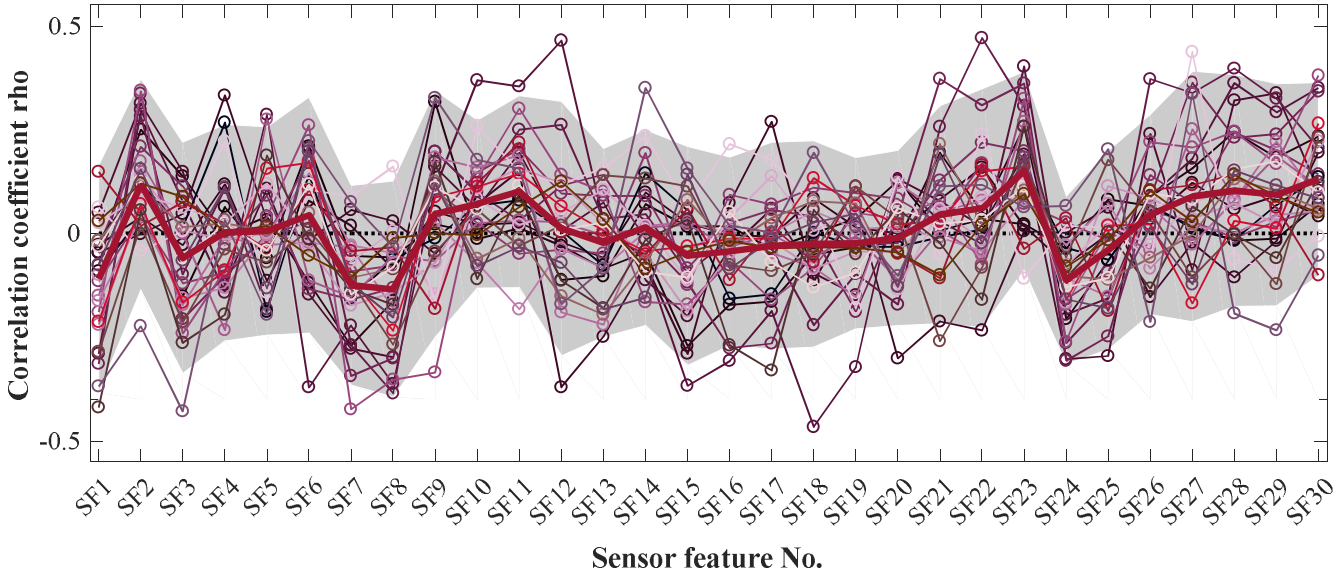
Adriaens, Figure 1



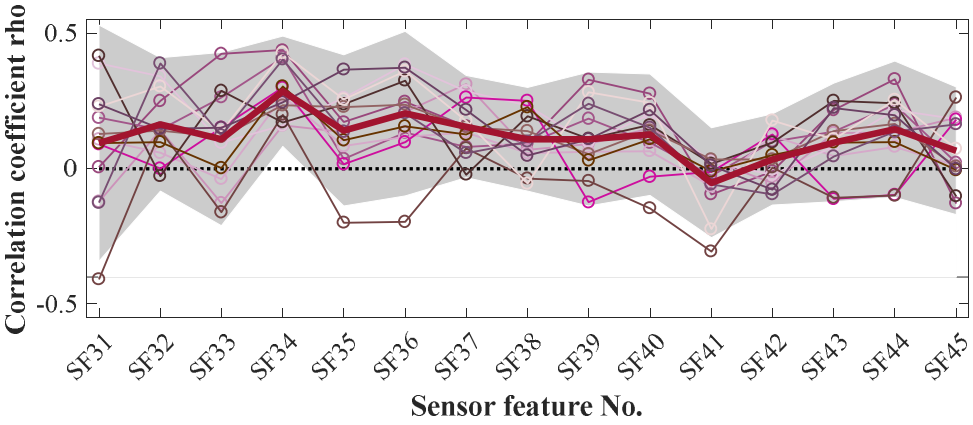
Adriaens, Figure 2



Adriaens, Figure 3



Adriaens, Figure 4



Adriaens, Table 1. Overview of the available data sets (DS). Data set 2 is a subset of DS1 for which also activity data were available besides milk meter data.

Descriptor	Data set 1 (DS1) Average ± SD [range]	Data set 2 (DS2) Average ± SD [range]
No. of days data per farm	3001 ± 828 [1830 ; 4977]	3020 ± 815 [1837 ; 4101]
No. of cows in total	3754	2075
No. of cows per farm	139 ± 82 [24 ; 308]	160 ± 94 [57 ; 308]
No. of lactations in total	9395	5286
No. of lactations per farm	348 ± 229 [44 ; 799]	407 ± 264 [113 ; 799]
Age at first calving per farm (years)	2.2 ± 0.1 [2.0 ; 2.4]	2.2 ± 0.1 [2.0 ; 2.3]
Average 305-day milk yield per farm (kg)	9557 ± 1007 [7931 ; 11739]	9691 ± 1061 [8244 ; 11739]
Average calving interval per farm (days)	407 ± 16 [377 ; 433]	403 ± 15 [377 ; 432]
Average dry period length per farm (days)	56 ± 13 [35 ; 91]	59 ± 16 [35 ; 91]

Adriaens, Table 2. Overview of a selection of lactation and activity sensor features calculated on first parity data for data set 1 (DS1) and data set 2 (DS2). A detailed overview for all the included features is given in **Appendices A** and **B**.

	Data set 1 (DS1) **Average ± SD [range]**	**Data set 2 (DS2)** **Average ± SD [range]**
No. of cows included	**3754**	**2022**
Theoretical 305-day yield	8261 ± 1772 [2688 ; 13755]	8317 ± 1874 [2689 ; 13755]
RMSE[1] [kg.day^{-1}]	3.4 ± 1.4 [1 ; 28]	3.8 ± 1.5 [1.2 ; 28]
Peak yield [kg]	34 ± 6 [15 ; 61]	34 ± 6 [14 ; 61]
Days in milk of peak lactation [day]	59 ± 25 [1 ; 248]	63 ± 27 [1 ; 207]
Rate increasing part of the lactation curve [kg.day^{-1}]	0.64 ± 0.28 [0.08 ; 6.7]	0.62 ± 0.31 [0.08 ; 6.7]
Persistency of the lactation curve [kg.day^{-1}]	-0.05 ± 0.03 [-0.15 ; 0.12]	-0.04 ± 0.03 [-0.15 ; 0.12]
No. of perturbations per lactation	6.0 ± 3.1 [0 ; 17]	6.3 ± 3.2 [0 ; 17]
Losses in perturbations per lactation [kg]	148 ± 124 [0 ; 1137]	154 ± 132 [0 ; 1129]
Activity skewness		-0.05 ± 1.64 [-5.19 ; 8.77]
No. of sharp and short activity peaks		3.9 ± 2.7 [0 ; 18]
Days in milk of first activity peak [days]		56 ± 64 [1 ; 305]
No. of general activity changes		24 ± 16 [2 ; 94]

1 RMSE: Root mean squared error

Adriaens, Table 3. Prediction results of the individual stepwise regression models in calibration and 10-times cross-validation (**CV**)

	Calibration, milk yield features only					Calibration, milk yield & activity features								CV[8], milk yield features only				CV, milk yield & activity features			
FARM ID	**R^2adj**	**N**[1]	**N_{LAC}**[2]	**N_{GOF}**[3]	**N_{PERT}**[4]	**R^2_{adj}**	**N**	**N_{LAC}**	**N_{GOF}**	**N_{PERT}**	**N_{GEN}**[5]	**N_{L1}**[6]	**N_{L2}**[7]	**$RMSE_{TR}$**[9]	**$RMSE_{CV}$**[10]	**P_C**[11]	**P_O**[12]	**$RMSE_{TR}$**	**$RMSE_{CV}$**	**P_C**	**P_O**
FARM1	0.10	8	2	4	2									0.26	0.27	45.9	3.7				
FARM2	0.16	5	0	2	3									0.24	0.27	44.3	1.1				
FARM3	0.05	4	1	1	2									0.25	0.27	41.5	3.6				
FARM4	0.27	3	0	1	2									0.21	0.21	50.0	0.0				
FARM5	0.29	8	0	2	6									0.22	0.25	45.8	6.8				
FARM6	0.53	5	1	1	3									0.17	0.18	70.0	0.0				
FARM7	0.31	4	0	2	2									0.23	0.24	59.0	2.6				
FARM8	0.42	11	2	5	4									0.19	0.23	45.6	6.3				
FARM9	0.08	2	0	1	1									0.25	0.25	44.6	2.2				
FARM10	0.16	8	2	2	4									0.24	0.27	48.9	5.6				
FARM11	0.14	7	2	0	5									0.24	0.26	37.1	3.4				
FARM12	0.40	9	1	3	5	0.51	8	1	2	2	1	2	0	0.20	0.24	55.9	1.7	0.18	0.20	52.5	1.7
FARM13	0.21	8	2	3	3	0.32	7	1	3	0	1	2	0	0.23	0.25	45.0	11.2	0.21	0.22	52.6	1.0
FARM14	0.28	6	1	3	2									0.21	0.23	45.6	1.3				
FARM15	0.06	3	2	0	1									0.19	0.19	41.8	1.8				
FARM16	0.10	9	3	2	4	0.30	15	3	2	2	1	4	3	0.23	0.24	44.8	7.4	0.20	0.21	51.5	1.8
FARM17	0.08	6	1	2	3	0.27	14	1	3	5	2	2	1	0.26	0.27	35.8	6.1	0.23	0.26	45.3	4.0
FARM18	0.53	7	2	2	3	0.76	13	0	3	6	2	1	1	0.16	0.19	55.0	5.0	0.10	0.15	77.5	0.0
FARM19	0.18	6	3	1	2	0.34	11	2	4	2	0	2	1	0.23	0.25	55.0	8.3	0.19	0.24	52.3	1.5
FARM20	0.14	5	0	2	3									0.24	0.25	40.7	5.0				
FARM21	0.21	4	1	1	2	0.33	5	0	1	1	1	0	2	0.22	0.22	35.9	5.3	0.20	0.22	43.5	0.0
FARM22	0.30	6	1	0	5	0.67	23	3	5	9	2	2	2	0.18	0.20	45.6	1.4	0.11	0.17	66.2	0.0
FARM23	0.11	5	0	1	4	0.24	14	0	1	4	1	4	4	0.26	0.26	39.3	2.8	0.23	0.25	44.4	6.7
FARM24	0.08	5	1	1	3	0.36	13	1	3	1	2	4	2	0.26	0.26	43.4	5.9	0.21	0.22	53.2	0.5
FARM25	0.03	3	0	2	1	0.20	12	1	4	1	2	2	2	0.26	0.26	41.9	16.5	0.23	0.24	49.4	5.2
FARM26	0.61	12	1	2	9	0.62	8	1	1	4	1	1	0	0.15	0.21	62.0	2.0	0.16	0.18	84.0	4.2
FARM27	0.15	6	1	3	2	0.32	13	2	2	5	0	3	1	0.26	0.26	40.4	2.5	0.22	0.24	48.5	3.7

[1] **N**: Total number of sensor features retained in the stepwise linear regression;

[2] **N_{LAC}**: Number of features linked with the milk yield shape characteristics retained in the final regression models;

[3] **N_{GOF}**: Number of features linked with the goodness-of-fit of the iterated Wood curve fitted on the lactation data retained in the final models;

[4] **N_{PERT}**: Number of features linked with the identified perturbations retained in the final models;

[5] **N_{GEN}**: Number of features linked with the general activity time series retained in the final models;

[6] **N_{L1}**: Number of features linked with the activity peaks assumed to relate to oestrus behaviour retained in the final models;

[7] **N_{L2}**: Number of features linked with periods of high or low average activity assumed to relate to general deviation in behaviour or activity level in the final models;

[8] **CV**: Results of the 10-fold cross-validation

[9] **$RMSE_{TR}$**: Average root mean squared error of the fitted scaled resilience ranking of the cows in the training set;

[10] **$RMSE_{CV}$**: Average root mean squared error of the predicted scaled resilience ranking of the cows in the validation set;

[11] **P_C**: Proportion of the animals on each farm predicted in the correct category (best, B, middle, M or worst, W);

[12] **P_O**: Proportion of the animals on each farm predicted in the opposite category: cows predicted as being in the lowest 33% while being in the highest 33% and vice versa

362