



Leveraging unsupervised machine learning techniques for detecting outliers in the daily milk yield data of dairy cows

Shogo Higaki,^{1,2} Eduardo Noronha de Andrade Freitas,^{2,3} Ariana Negreiro,²
João R. R. Dórea,² and Victor E. Cabrera^{2*}

¹National Institute of Animal Health, National Agriculture and Food Research Organization, Tsukuba, Ibaraki 305-0856, Japan

²Department of Animal and Dairy Sciences, University of Wisconsin–Madison, Madison, WI 53706

³Department of Informatic, Federal Institute of Goiás, Goiânia, Goiás 74130-012, Brazil

ABSTRACT

The lactation curve is essential for developing effective feeding plans, optimizing breeding, and strategizing milk production for dairy farms. However, health disorders, as well as external factors such as heat stress, dietary changes, and certain management practices can cause perturbations (temporary drops in milk yield) that shift the fitted lactation curve downward, making it difficult to accurately estimate the potential lactation ability of dairy cows. This study aims to evaluate the applicability of unsupervised machine learning techniques for detecting outliers in daily milk yield data and estimating the expected lactation curve in the absence of perturbations, referred to as the unperturbed lactation curve (ULC). Using the Wood model as the baseline lactation curve, we compared ULC derived from 3 unsupervised machine learning models (UMLM), specifically one-class support vector machines, isolation forest, and local outlier factor, with those from 2 previously proposed models: the perturbed lactation model (PLM) and the iterative Wood model (IWM). We first conducted a simulation study using 1,000 simulated lactations over a 305-d period, each including 1 to 15 perturbations (mean \pm SD: 4.00 ± 1.46), to assess perturbation detection performance. Across all UMLM, sensitivities (~61%), precisions (~82%), and their harmonic means (F_1 scores, ~70%) did not differ significantly. The UMLM outperformed the baseline Wood model in sensitivity (51.5%) and F_1 score (64.2%) while maintaining comparable precision (83.8%). Their F_1 scores also exceeded those of the PLM (53.2%) and IWM (66.8%), indicating more balanced curve adjustment and improved perturbation detection. We then applied the models to observed daily milk yield data from 2,831 lactation records of 1,636 Holstein cows collected

over a 10-year period at the University of Wisconsin–Madison Agricultural Research Station. The comparison focused on the goodness-of-fit of ULC, computational efficiency, curve shape, and the validity of identified perturbations. The UMLM demonstrated relatively high computational efficiency in establishing the ULC, and these ULC showed better goodness-of-fit and shapes more consistent with the baseline Wood curve than the PLM and IWM. The upward shifts in the ULC from the UMLM were more conservative than those from the IWM and PLM, yet seemed reasonable based on previous reports on the impact of health disorders on milk yield. Additionally, these upward shifts by the UMLM may help identify potential perturbations that went undetected with the baseline Wood curve. In contrast, the PLM and IWM showed limitations in detecting potential perturbations, especially during early lactation. These findings suggest that unsupervised machine learning techniques can effectively detect potential outliers in daily milk yield data and adequately estimate the expected lactation curve in the absence of perturbations. However, the generalizability of the findings may be limited by the use of data from only Holstein cows at a single farm and the absence of health, environmental, and management records. Moreover, the current UMLM do not account for fixed effects (e.g., breed, parity, calving season) or long-term impacts of health disorders, which may hinder accurate lactation curve modeling. Future studies should consider incorporating more flexible modeling approaches and multifarm datasets with detailed background records.

Key words: cattle, lactation curve, milk production, perturbation

INTRODUCTION

The lactation curve is a reproducible pattern of milk yield that can be mathematically expressed as a function of time, characterized by an ascending phase from parturition to peak production, followed by a descending

Received October 17, 2024.

Accepted May 30, 2025.

*Corresponding author: vcabrera@wisc.edu

phase after the peak (Bouallegue and M'Hamdi, 2020). It provides valuable information for developing effective feeding plans, optimizing breeding, and strategizing milk production for dairy farms (Bouallegue and M'Hamdi, 2020). However, certain health disorders, such as mastitis, ketosis, and lameness (Fourichon et al., 1999; Rasmussen et al., 2024), are known to cause perturbations, which are temporary drops in milk yield. Additionally, external factors such as heat stress (Hou et al., 2021), dietary changes (Bach, 2023), and certain management practices (e.g., regrouping; von Keyserlingk et al., 2008) can also lead to transient milk yield depression. These perturbations shift the fitted lactation curve downward, making it difficult to accurately estimate the potential lactation ability of dairy cows (Lucey et al., 1986).

To estimate the expected lactation curve in the absence of perturbations, several methods have been proposed. For instance, Ben Abdelkrim et al. (2021) developed a perturbed lactation model (**PLM**) to explicitly represent perturbations and estimate the expected lactation curve by excluding the perturbation effects. Adriaens et al. (2021) proposed an iterative approach using the Wood model (Wood, 1967) to determine the expected lactation curve by identifying outlier data points through a set threshold and selecting clean data from daily milk yield to apply the Wood model. However, the former model is considered computationally expensive, and the latter is suggested to have issues with unstable fitting (Ranzato et al., 2024).

In recent years, significant research has focused on outlier detection using unsupervised machine learning techniques (Wang et al., 2021). Methods such as one-class support vector machines (**SVM**; Schölkopf et al., 2001), isolation forest (Liu et al., 2008), and local outlier factor (Breunig et al., 2000) have been developed. These methods are relatively computationally efficient and more robust compared with simple cutoff thresholding methods for detecting outliers (Chandola et al., 2009; Darrab et al., 2024). Additionally, they have demonstrated effectiveness in various domains, including industrial damage detection and network intrusion detection (Seliya et al., 2021). However, to our knowledge, these methods have never been applied to milk yield data.

This study aims to evaluate the applicability of unsupervised machine learning techniques for detecting outliers in daily milk yield data and estimating the expected lactation curve in the absence of perturbations, referred to as the unperturbed lactation curve (**ULC**). To assess their applicability, we compared ULC derived using 3 unsupervised machine learning techniques, specifically one-class SVM, isolation forest, and local outlier factor, with those from 2 previously proposed models by Ben Abdelkrim et al. (2021) and Adriaens et al. (2021). These models were first applied to simulated daily milk yield data with known perturbation events, allowing us

to objectively evaluate their perturbation detection performance using sensitivity, precision, and their harmonic mean (F_1 score). We then applied the models to observed daily milk yield data, focusing on the goodness-of-fit of ULC, computational efficiency, curve shape, and the validity of identified perturbations. We hypothesize that unsupervised machine learning techniques can not only accurately detect true perturbations in simulated data, but also efficiently identify potential outliers in observed data while adjusting the lactation curve to mitigate the effects of perturbations.

MATERIALS AND METHODS

Baseline Lactation Curve Model

The baseline lactation curve model used in this study was developed by Wood (1967). The formula of the Wood model is as follows:

$$Y(t) = a \times t^b \times e^{-c \times t},$$

where $Y(t)$ is the daily milk yield in kilograms at time t (DIM), and a , b , and c are parameters that define the shape of the lactation curve, and c is the base of the natural logarithm. Parameter a mainly determines the scaling of the curve, while parameters b and c determine the ascending slope before the peak and descending slope after the peak of the lactation curve, respectively. The value of the parameter a is approximately equal to the yield immediately after calving, because the parameter c is much less than 1. The peak milk yield of $a(b/c)^b e^{-b}$ is assumed to occur b/c days after calving (Wood, 1967). The 305-d milk yield was calculated by summing the daily milk yields estimated by the fitted Wood model curve. The lactation curve obtained by fitting the Wood model to all available daily milk yield data is hereafter referred to as the baseline Wood curve.

The Wood model was fitted separately to each lactation record using a consistent functional form across individuals. Because the model was applied at the individual level, no group-level fixed effects such as breed, parity, or calving season were included in the modeling process. All other models in this study adopted the same formulation. Because the Wood model forces the net impact of disease to zero, it limits the ability to capture long-term negative effects on milk yield (Wilson et al., 2004), and its use may hinder accurate modeling of lactation curves. However, we selected this model because it explains lactation curves well, has low computational cost, and is the most frequently used model, making comparisons with previous studies easier (Bouallegue and M'Hamdi, 2020). Another advantage of the Wood model is its shape stability in the presence of missing data (Wasike et al.,

2011), which was essential for the outlier filtering procedure we employed in this study.

Simulated Milk Yield Data

Simulated daily milk yield data were generated for 1,000 lactations over a 305-d lactation period using parameterized Wood model curves, implemented in Python (version 3.11.4) with the NumPy (version 1.26.4; Harris et al., 2020) and Pandas (version 2.2.3; McKinney, 2010) libraries. For each lactation, the Wood model parameters a , b , and c were sampled from beta distributions skewed toward the lower end of each parameter's range: a was drawn from 0 to 55 with a mode around 15 (beta distribution with $\alpha = 2.4$, $\beta = 4.6$), b from 0 to 0.9 with a mode around 0.22 ($\alpha = 2.2$, $\beta = 4.7$), and c from 0 to 0.01 with a mode around 0.0025 ($\alpha = 2.3$, $\beta = 4.7$). These distributions reflect the empirical patterns observed in 8.6 million lactation curves based on test-day data from Holstein cows in the United States (Li et al., 2022). To ensure a plausible range for most cows, only lactation curves that met all of the following criteria were retained: (1) a peak daily milk yield between 20 and 100 kg, (2) time to peak less than 300 DIM, and (3) a total 305-d milk yield between 5,000 and 20,000 kg.

Perturbations were then introduced by randomly inserting between 1 and 15 events per lactation, with the number of events sampled from a truncated normal distribution (mean = 4, SD = 1.5). The start day of each perturbation was randomly selected between DIM 1 and 305, allowing overlaps among multiple events within the same lactation. Each perturbation lasted for 5 to 45 d, with durations drawn from a beta distribution ($\alpha = 0.7$, $\beta = 2.1$) skewed toward shorter periods (approximate median: 12 d; interquartile range: 8 to 20 d). These numbers and durations were chosen to reflect the patterns observed in a previous study by Ranzato et al. (2024), who applied PLM and iterative Wood models to 4,441 lactation curves constructed from daily milk yield data for 2,250 dairy cows. During each perturbation, daily milk yield was reduced by a randomly sampled proportion between 10% and 20%.

To simulate day-to-day milk yield variability, Gaussian noise was added to the perturbed lactation curves. For each day, the SD of the noise was determined by multiplying that day's milk yield by a random factor drawn from a uniform distribution between 0 and 0.1, indicating that higher yields had greater variability and the SD could reach up to 10% of that day's milk yield (Moncur et al., 2021; Guinan et al., 2024). Noise values were sampled independently for each cow on each day, and the resulting noise was then added to the day's milk yield. Any negative values were truncated to zero. Each record in the simulated dataset includes an artificial cow

identifier, DIM, daily milk yield (kg), and a binary indicator of perturbation start day (1 if the day corresponds to the start of a perturbation).

Observed Milk Yield Data

All data were collected at the University of Wisconsin–Madison Agricultural Research Station (Arlington, WI) over the last 10 years (August 2013 to February 2024). Cows (Holstein, Jersey, Brown Swiss, and dairy-beef crossbreeds) were housed in 2 sand-bedded freestall barns connected by a corridor and were managed in groups according to lactation month, production, and reproductive status. The cows were provided with a balanced total mixed ration, had unlimited access to water, and were milked twice daily.

Historical daily milk yield data (kg) and calving dates with lactation numbers were collected from the farm management software BoviSync (Bovisync LLC, Eden, WI) and DairyComp305 (Valley Agricultural Software, Tulare, CA), respectively. These datasets were merged based on cow identifiers (i.e., ear tag number), resulting in data for 7,766 lactation records from 3,053 cows. To ensure consistency in statistical modeling, only Holstein cows were included in this study, as lactation curve patterns are known to vary substantially across breeds (Wood, 1980). After selecting the lactation data with at least 300 data points from 1 to 305 DIM, the final dataset consisted of 2,831 lactation records from 1,636 Holstein cows (mean lactation number: 2.80 ± 1.60 ; range: 1 to 12). In total, the dataset included 861,534 daily milk yield observations. The main reasons for the exclusion of lactation records were shortened lactation periods (e.g., due to early culling or dry-off) and missing data, which resulted from data storage issues or technical malfunctions in the milking equipment. Milk yield data during health events, including mastitis, were routinely recorded by the milking system and were therefore included in the analysis without any special handling, such as data imputation. Because this study did not involve animal procedures, it was exempt from requiring approval by an Institutional Animal Care and Use Committee.

Perturbed Lactation Model

The PLM was developed on the work by Ben Abdelkrim et al. (2021). The formula of PLM for a lactation with n perturbations is given as follows:

$$Y(t) = a \times t^b \times e^{-cxt} \times \prod_{i=1}^n \left[1 - \frac{k_{0,i} \times k_{1,i}}{k_{1,i} - k_{2,i}} \times \left(e^{-k_{2,i} \times \Delta_i(t)} - e^{-k_{1,i} \times \Delta_i(t)} \right) \right],$$

$$\Delta_i(t) = \begin{cases} 0 & \text{if } t < t_{Pi} \\ t - t_{Pi} & \text{if } t \geq t_{Pi} \end{cases},$$

where $Y(t)$ is the daily milk yield in kilograms at time t (DMI), and a , b , and c are the 3 parameters of the Wood model, which define the ULC that represents the lactation curve without the influence of perturbations. There are 4 parameters per individual perturbation i ($k_{0,i}$, $k_{1,i}$, $k_{2,i}$, and t_{Pi}). For a given perturbation i , the parameters $k_{0,i}$, $k_{1,i}$, and $k_{2,i}$ represent the perturbation intensity, collapse speed, and recovery speed parameters, respectively. $\Delta_i(t)$ is the elapsed time since the beginning of the i th perturbation and is given by t_{Pi} the start time (DIM) of the i th perturbation. So that the total number of parameters to define PLM is equal to $4 + 4 \times n$. In the code provided by Ben Abdelkrim et al. (2021), the lower bound for the t_{Pi} is set such that the model restricts detection of perturbations to those starting at least 3 d after the earliest available time point.

In practice, because the number of perturbations is unknown, a fitting strategy is implemented in 2 steps: (1) perform repeated fittings to estimate the most frequent number of detected perturbations, and (2) fix the number of perturbations to the value determined in the first step and estimate the remaining parameters of the model (Ben Abdelkrim et al., 2021). Arbitrary values need to be set for the maximum number of perturbations to be detected and the number of fittings to be performed in step 1, as well as for the maximum number of iterations required to achieve convergence of the estimation procedure in step 2. Following Ben Abdelkrim et al. (2021), we set the maximum number of perturbations to 15. However, the number of fittings and the maximum number of iterations were reduced from 100 to 50 and from 100,000 to 1,000, respectively, due to computational constraints and in accordance with a recent comparable study (Ranzato et al., 2024). Although that study noted that changes in these meta-parameters can influence the number and duration of detected perturbations, as well as the 305-d milk yield calculated from the ULC, we adopted this setting to balance computational feasibility with methodological rigor.

Iterative Wood Model

The IWM presented hereafter is based on the work by Adriaens et al. (2021). An iterative fitting procedure is applied to gradually remove potential outlier data points, following these steps: (1) in this first iteration ($i = 1$), fit the Wood model (Curve_1) on all milk yield data of the lactation; (2) calculate the residuals from this model, their SD (SD_1) and the root mean squared error (RMSE ; RMSE_1); (3) remove all milk yield data below the $\text{Curve}_1 - 1.6 \times \text{SD}_1$ to obtain the filtered milk

yield data resulting from the first iteration ($i = 1$); (4) in this next iteration (i), fit the Wood model (Curve_i) on the filtered milk yield data resulting from the previous iteration ($i - 1$); (5) calculate the residuals from this model, their SD (SD_i), and the RMSE_i ; (6) remove all milk yield data below the $\text{Curve}_i - 1.6 \times \text{SD}_i$ to obtain the filtered milk yield data resulting from the current iteration (i); (7) repeat steps 4 to 6 until the improvement in RMSE_i of the current iteration (i) compared with the previous iteration (RMSE_{i-1}) is smaller than 0.1 kg, or after 20 iterations to determine the ULC. These parameter settings followed the original implementation described by Adriaens et al. (2021), ensuring consistency with the established method to facilitate comparability across studies.

The RMSE represents the square root of the average of the squared differences between estimated and observed values (Hodson, 2022). The formula for calculating RMSE is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2},$$

where n is the number of observations, y_i represents the actual value for the i th observation, and \hat{y}_i denotes the predicted value for the i th observation. The RMSE gives a higher weight to larger errors, making it sensitive to large deviations between predicted and observed values (Hodson, 2022).

Unsupervised Machine Learning Model

For the detection of potential outlier data points, we used 3 unsupervised machine learning algorithms: one-class SVM (Schölkopf et al., 2001), isolation forest (Liu et al., 2008), and local outlier factor (Breunig et al., 2000). One-class SVM detects anomalies by learning a boundary that encompasses most data points (normal) and classifies points outside this boundary as outliers. Isolation forest builds decision trees by randomly selecting features and split values, identifying anomalies as those points isolated quickly (requiring fewer splits). Local outlier factor detects anomalies by comparing the local density of a data point to its neighbors.

We employed the scikit-learn library (version 1.2.2; Pedregosa et al., 2011) in Python to implement the 3 algorithms, specifically using the `sklearn.svm`, `sklearn.ensemble`, and `sklearn.neighbors` modules for one-class SVM, isolation forest, and local outlier factor, respectively. Because the proportion of outlier data points is unknown, we conducted an optimal search for the relevant hyperparameter values of each algorithm, as described in the following paragraph. Specifically, the parameter “nu”

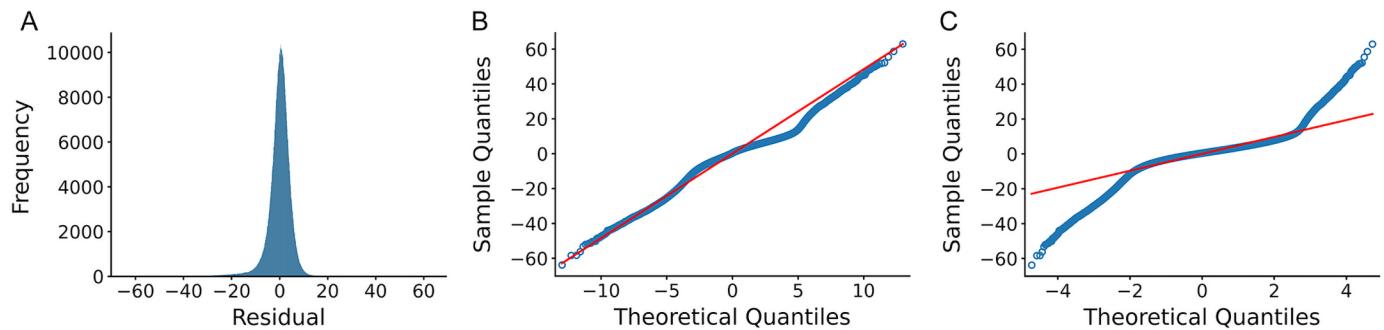


Figure 1. Frequency distribution of residuals between the Wood curve and actual daily milk yield data, and quantile-quantile (Q-Q) plots of the residuals for the Laplace and normal distributions. (A) Frequency distribution of residuals between the Wood curve and actual daily milk yield data. The kurtosis of the residuals was 9.968, and the skewness was -1.344. (B) The Q-Q plot of the milk residuals for the Laplace distribution. The result of the Kolmogorov-Smirnov test was 0.077 ($P = 0.178$). (C) The Q-Q plot of the milk residuals for the normal distribution. The result of the Kolmogorov-Smirnov test was 0.108 ($P = 0.018$). To account for the potential inflation of statistical significance due to a large sample size, we randomly selected 200 data points from the milk residuals for the Kolmogorov-Smirnov test.

for one-class SVM and the parameter “contamination” for isolation forest and local outlier factor were both set to values ranging from 0.01 to 0.5 (divided into 25 increments). This means that between 1% and 50% of the data could be classified as outliers. The value of 0.5 is the default for the parameter “nu” in one-class SVM and also represents the maximum allowable value for the “contamination” parameter in isolation forest and local outlier factor, based on the assumption that outliers constitute a minority of the data. All other hyperparameters were set to their default values.

In this model, we first optimized the hyperparameter representing the assumed proportion of outliers, and then identified and removed outliers in a single step based on this value before generating the ULC. The unsupervised machine learning model (**UMLM**) followed these steps: (1) fit the Wood model (Curve_1) on all milk yield data of the lactation; (2) calculate the residuals from this model and their mean absolute error (**MAE**; MAE_1); (3) remove all milk yield data detected as outliers within negative residuals by unsupervised machine learning with the smallest hyperparameter value; (4) fit the Wood model (Curve_2) on the filtered milk yield data; (5) calculate the residuals from this model and their MAE (MAE_2); (6) compare MAE_2 with MAE_1 ; (7) if there is an improvement, repeat steps 3 to 6 with the next smallest hyperparameter value until the improvement in MAE of the current iteration compared with the previous iteration is smaller than 0.1 kg, or the hyperparameter value reaches its upper limit (after 25 iterations) to determine the optimal hyperparameter value; (8) remove all milk yield data detected as outliers within negative residuals by unsupervised machine learning with the optimal hyperparameter value; (9) retain the milk yield data for DIM less than 5 and data where residuals are within 5% of the Curve_1 for stabilizing the model fit; (10) fit the Wood model to establish the ULC.

The MAE represents the average of the absolute differences between predicted and observed values (Hodson, 2022). The formula for calculating MAE is as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|,$$

where n is the number of observations, y_i represents the actual value for the i th observation, \hat{y}_i denotes the predicted value for the i th observation, and $|y_i - \hat{y}_i|$ is the absolute error for each observation. In our preliminary analysis, the differences between the baseline Wood curve and the actual milk yield followed a Laplacian distribution (Kolmogorov-Smirnov test yielded statistics of 0.077 with $P = 0.178$ and 0.108 with $P = 0.018$ for the Laplace and normal distributions, respectively; Figure 1). Because minimizing MAE yields the most likely model for Laplacian errors (Hodson, 2022), model improvement was evaluated based on MAE. Unlike RMSE, MAE does not give higher weight to larger errors, as it averages the magnitude of errors without considering their direction or magnitude.

Comparison of Models on Simulated Milk Yield Data

To assess each model’s performance in detecting perturbations, we compared the predicted perturbation start days with the ground truth in the simulated datasets. For all models except the PLM, a perturbation was defined as a period of at least 5 consecutive days with negative residuals between the actual milk yield and the expected yield from the baseline Wood curve or the ULC, during which the daily milk yield dropped below 80% of the

Table 1. Detection performance of perturbation start days based on simulated milk yield data (95% CI in parentheses)

Item ¹	Wood model	Perturbed lactation model (PLM)	Iterative Wood model (IWM)	Unsupervised machine learning model (UMLM)		
				One-class SVM	Isolation forest	Local outlier factor
No. of simulated perturbations per lactation				4.00 (3.91, 4.09)		
No. of predicted perturbations per lactation	2.38 ^d (2.31, 2.45)	7.90 ^a (7.82, 7.98)	3.37 ^b (3.29, 3.45)	2.89 ^c (2.82, 2.96)	2.91 ^c (2.83, 2.98)	2.85 ^c (2.77, 2.92)
Sensitivity (%)	51.5 ^c (49.9, 53.1)	81.8 ^a (80.5, 83.0)	61.2 ^b (69.6, 62.9)	61.5 ^b (59.9, 63.0)	61.2 ^b (69.7, 62.8)	61.0 ^b (59.4, 62.5)
Precision (%)	83.8 ^a (82.2, 85.3)	40.8 ^c (39.9, 41.8)	71.8 ^b (70.2, 73.5)	82.8 ^a (81.4, 84.2)	82.1 ^a (80.7, 83.6)	83.7 ^a (82.3, 85.1)
F ₁ score (%)	64.2 ^c (62.9, 65.4)	53.2 ^d (52.3, 54.2)	66.8 ^b (65.5, 68.1)	70.0 ^a (68.8, 71.3)	69.7 ^a (68.5, 70.9)	69.9 ^a (68.7, 71.2)

^{a-d}Values with different superscript letters within columns differ significantly ($P < 0.05$).

¹No. of predicted perturbations per lactation = TP + FP. Sensitivity (%) = TP/(TP + FN) × 100. Precision (%) = TP/(TP + FP) × 100. F₁ score (%) = 2 × TP/(2 × TP + FP + FN) × 100. Here, TP, FN, and FP refer to true positives, false negatives, and false positives, respectively. A predicted perturbation start day was counted as a TP if it fell within ± 3 d of a true start day. Each predicted start day could be matched to only one true start day, and vice versa. Unmatched predictions were counted as FP, and unmatched true start days were considered FN. For all models except the PLM, a perturbation was defined as a period of at least 5 consecutive days of negative residuals between the actual milk yield and the expected yield from the fitted curve, during which the daily milk yield dropped below 80% of the expected yield at least once. For the PLM, these values were determined based on the presence, number, and value of the parameter t_{pi} (the start time of the i th perturbation).

expected yield at least once (Adriaens et al., 2021). The start day of each perturbation was taken to be the first day on which the residual dropped below zero. In the case of the PLM, perturbation onset was determined using the model parameter t_{pi} , which represents the start time (DIM) of the i th perturbation.

For each lactation, true perturbation start days were extracted from the simulation records, and predicted start days were obtained from each model's output. A prediction was considered a true positive (TP) if a predicted perturbation start day was within ± 3 d of a true start day. Each predicted start day could be matched to only one true start day, and vice versa. Unmatched predictions were counted as false positives (FP), and unmatched true start days were considered false negatives (FN). Based on these counts, detection sensitivity, precision, and the F₁ score were calculated as follows:

$$\text{Sensitivity}(\%) = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100,$$

$$\text{Precision}(\%) = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100,$$

$$F_1 \text{ score}(\%) = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \times 100.$$

Comparison of Models on Observed Milk Yield Data

In this study, we did not use cow health records, as not all diseases were likely to have been detected or treated. More importantly, our aim was not to focus solely on detecting outliers or perturbations associated

with health disorders. Accordingly, the relationship between the detected outliers or perturbations and actual health events could not be directly evaluated. Instead, model performance was assessed using alternative indicators. Specifically, the models were compared across 4 aspects based on the observed milk yield data: the goodness-of-fit of the ULC, computational efficiency, the shape of the ULC, and the validity of estimated potential perturbations.

To assess the goodness-of-fit of ULC, the MAE, RMSE, and R² were computed. These values were calculated over all the tested daily milk yield data. The R² indicates how well the model explains the variability of the observed data and is calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

where n is the number of observations, y_i represents the actual value for the i th observation, \hat{y}_i denotes the predicted value for the i th observation, and \bar{y} is the mean of the observed values (Chicco et al., 2021).

To assess the computational efficiency, the total number of curve fittings performed until the ULC was established, as well as the number of data points identified as potential outliers (if applicable), were counted. The duration of the total curve-fitting processes was also measured using the “datetime.now()” function from the datetime module in Python (for the Wood model, IWM, and UMLM) and the “Sys.time()” function from the base package in R (for PLM). These measurements were performed on a conventional Microsoft Windows

Table 2. Summary statistics of the observed milk yield dataset

Item	Mean (\pm SD)
Number of unique cows	1,636
Number of lactations	2,831
Lactation 1	593
Lactation 2	886
Lactation ≥ 3	1,352
Initial milk yield ¹ (kg)	24.3 \pm 7.8
Total milk yield in 305 d ² (kg)	13,626.0 \pm 2,265.6

¹Initial milk yield = actual milk yield on d 1 of lactation.

²Total milk yield in 305 d = multiplying the average value of actual daily milk yield by 305.

laptop PC equipped with an Intel Core i7-10875H CPU and 64 GB of RAM.

To evaluate the shape of the ULC, we compared the parameters a , b , and c of the baseline Wood curve and the ULC, as well as the estimated peak milk yield, DIM to peak yield, and 305-d milk yield based on both the baseline Wood curve and the ULC. Days at peak milk yield might be within or outside the lactation period. Therefore, in this study, if the day was less than zero or greater than 305, it was assumed to be 1 or 305, respectively (Choumei et al., 2006). As a reference, the milk yield at DIM 1 was determined as the actual value of initial milk yield, and the actual 305-d milk yield was calculated by multiplying the average value of actual daily milk yield by 305.

To evaluate the validity of potential perturbations identified based on the baseline Wood curve and ULC, we compared the number of potentially perturbed lactations, the total number of potential perturbations, and the number of potential perturbations per lactation, as well as the distribution of the start day of the potential perturbations for each model. The potentially perturbed lactation refers to a lactation in which a potential perturbation has been detected. For the PLM, these values were determined based on the presence, number, and value of the parameter t_{Pi} .

Statistical Analysis

All statistical analyses were performed using Python with the statsmodels (version 0.14.0; Seabold and Perktold, 2010), SciPy (version 1.11.1; Virtanen et al., 2020), and scikit-posthocs (version 0.9.0; Terpilowski, 2019) libraries. The distribution of the milk residuals, whether following a normal distribution or a Laplace distribution, was assessed using the Kolmogorov-Smirnov test and visually examined with a quantile-quantile plot created using the “qqplot” function from the statsmodels library. To account for the potential inflation of statistical significance due to a large sample size, the residual distribution analysis was performed using a random sample

of 200 data points from the milk residuals. This subset was used exclusively for the residual distribution analysis and not in any other part of the study. Regarding the values in Tables 1, 2, 3, 4, and 5, both the Shapiro-Wilk test for normality and the Levene test for equal variances indicated violations of their respective assumptions. Therefore, the Kruskal-Wallis test was used to evaluate differences among models, followed by Dunn's test with Bonferroni correction for multiple comparisons. Values are presented in the text as mean \pm SD.

RESULTS

Comparison of Models on Simulated Milk Yield Data

Simulated daily milk yield data were generated for 1,000 lactations over a 305-d period. Each lactation included 1 to 15 perturbations (4.00 \pm 1.46 per lactation, based on unique start days), resulting in a total of 3,998 simulated perturbations.

Table 1 summarizes the perturbation detection performance of the models. Across all unsupervised machine learning algorithms, the number of predicted perturbations, sensitivity, precision, and F₁ scores based on the ULC of the UMLM were not statistically different. Compared with the baseline Wood curve, the ULC of the UMLM predicted ~0.5 more perturbations per lactation and achieved ~10% higher sensitivity (both $P < 0.05$), and no significant differences in precision were observed. The F₁ scores based on the ULC of the UMLM were significantly ($P < 0.05$) higher than that of the baseline Wood curve.

In contrast, the number of predicted perturbations based on the ULC of the UMLM was significantly lower than that of both the PLM and the IWM. The PLM achieved ~20% higher sensitivity than the UMLM, but its precision was ~40% lower (both $P < 0.05$). The sensitivity of the IWM did not differ from that of the UMLM, while its precision was ~10% lower ($P < 0.05$). The F₁ scores based on the ULC of the UMLM were significantly ($P < 0.05$) higher than those of both the PLM and the IWM, regardless of the algorithm used.

Comparison of Models on Observed Milk Yield Data

Summary statistics of the observed daily milk yield data used in this study are presented in Table 2. Figures 2 and 3 illustrate representative cases: Figure 2 shows a case where the baseline Wood curve and the ULC from the 5 models did not differ substantially, and Figure 3 shows a case where the ULC varied remarkably among models. In Figure 2, there is little difference in the ULC across the 5 models, although the number and timing of potential perturbations were different. The UMLM detected no po-

Table 3. Results of curve-fitting procedures on the observed milk yield data: goodness-of-fit and computational efficiency-related indices

Item ¹	Wood model	Perturbed lactation model (PLM)	Iterative Wood model (IWM)	Unsupervised machine learning model (UMLM)		
				One-class SVM	Isolation forest	Local outlier factor
MAE (kg)	3.24 ± 0.98 ^c	3.68 ± 1.16 ^a	3.40 ± 1.30 ^b	3.24 ± 0.96 ^c	3.24 ± 0.96 ^c	3.25 ± 0.97 ^c
RMSE (kg)	4.65 ± 1.39 ^d	5.35 ± 1.67 ^b	5.30 ± 3.65 ^a	4.81 ± 1.47 ^c	4.81 ± 1.48 ^c	4.81 ± 1.46 ^c
R ²	0.63 ± 0.18 ^a	0.52 ± 0.26 ^c	0.37 ± 4.67 ^d	0.60 ± 0.20 ^b	0.60 ± 0.20 ^b	0.60 ± 0.20 ^b
No. of curve-fitting processes per lactation ²	809.2 ± 1.4 ^a	NA	52.3 ± 1.6 ^d	21.3 ± 3.8 ^b	19.2 ± 5.4 ^c	19.4 ± 5.5 ^c
No. of potential outlier data points ³	NA	NA	51.7 ± 22.7 ^a	33.5 ± 6.3 ^d	33.7 ± 9.2 ^c	36.0 ± 9.3 ^b
Time for curve-fitting process per lactation ⁴ (s)	0.008 ± 0.007 ^c	3,358.457 ± 190.120 ^a	0.108 ± 0.045 ^d	0.527 ± 0.096 ^c	15.948 ± 0.963 ^c	0.503 ± 0.096 ^b

^{a-c}Values with different superscript letters within columns differ significantly ($P < 0.05$).¹MAE = mean absolute error; RMSE = root mean squared error. Values are presented as mean ± SD.²No. of curve-fitting processes per lactation = number of curve-fitting processes required to establish the unperturbed lactation curve (ULC).³No. of potential outlier data points = number of data points identified as potential outliers and filtered to establish the ULC. NA = not available.⁴Time for curve-fitting process per lactation = time required for the curve-fitting process to establish the ULC.**Table 4.** Results of curve-fitting procedures on the observed milk yield data: lactation curve shape-related indices¹

Item	Wood model	Perturbed lactation model (PLM)	Iterative Wood model (IWM)	Unsupervised machine learning model (UMLM)		
				One-class SVM	Isolation forest	Local outlier factor
Parameter <i>a</i>	24.4 ± 7.8 ^b	21.6 ± 6.3 ^c	31.2 ± 44.0 ^a	24.9 ± 7.4 ^b	24.8 ± 7.4 ^b	24.8 ± 7.4 ^b
Parameter <i>b</i>	0.242 ± 0.096 ^b	0.285 ± 0.068 ^a	0.200 ± 0.107 ^c	0.240 ± 0.075 ^b	0.241 ± 0.075 ^b	0.241 ± 0.076 ^b
Parameter <i>c</i>	0.0033 ± 0.0013 ^b	0.0036 ± 0.0012 ^a	0.0029 ± 0.0014 ^c	0.0033 ± 0.0013 ^b	0.0033 ± 0.0013 ^b	0.0033 ± 0.0013 ^b
Predicted peak milk yield ² (kg)	52.5 ± 10.0 ^c	55.2 ± 10.2 ^a	54.5 ± 11.7 ^b	53.7 ± 10.1 ^b	53.7 ± 10.1 ^b	53.7 ± 10.1 ^b
Predicted days in milk at peak milk yield ³ (d)	80.1 ± 36.3 ^b	85.2 ± 31.7 ^a	75.1 ± 42.2 ^c	80.4 ± 35.5 ^b	80.4 ± 35.0 ^b	80.3 ± 35.0 ^b
Predicted 305-d milk yield ⁴ (kg)	13,624.8 ± 2,268.3 ^c	14,260.2 ± 2,370.1 ^a	14,128.4 ± 2,366.7 ^a	13,954.9 ± 2,314.3 ^b	13,954.5 ± 2,316.5 ^b	13,954.0 ± 2,315.6 ^b

^{a-c}Values with different superscript letters within columns differ significantly ($P < 0.05$).¹Values are presented as mean ± SD.²Predicted peak milk yield = $a \times (b/c)^b \times e^{-b}$.³Predicted DIM at peak milk yield = b/c .⁴Predicted 305-d milk yield = sum of the predicted daily milk yield based on the baseline Wood curve or the unperturbed lactation curve.

Table 5. Results of curve-fitting procedures on the observed milk yield data: identified potential perturbation-related indices¹

Item	Wood model	Perturbed lactation model (PLM)	Iterative Wood model (IWM)	Unsupervised machine learning model (UMLM)		
				One-class SVM	Isolation forest	Local outlier factor
No. (%) of potential perturbed lactations ²	2,690 (95.0)	2,831 (100)	2,802 (99.0)	2,776 (98.1)	2,774 (98.0)	2,776 (98.1)
No. of potential perturbations ³	7,283	22,513	11,224	9,756	9,854	9,833
No. of potential perturbations per lactation ⁴	2.57 ± 1.44 ^d	7.95 ± 1.47 ^a	3.96 ± 2.02 ^b	3.45 ± 1.76 ^c	3.48 ± 1.81 ^c	3.47 ± 1.76 ^c

^{a-d}Values with different superscript letters within columns differ significantly ($P < 0.05$).

¹All values were calculated based on the definition of a perturbation as a period of at least 5 consecutive days of negative residuals between the actual milk yield and the expected yield from the fitted curve, during which the daily milk yield dropped below 80% of the expected yield at least once, except for the PLM. For the PLM, all values were calculated based on the number and value of the parameter t_{pl} (the start time of the i th perturbation).

²No. of potential perturbed lactations = number of lactations in which one or more potential perturbations were identified.

³No. of potential perturbations = number of potential perturbations identified.

⁴Values are presented as mean ± SD.

tential perturbations, whereas 5 and 1 potential perturbations were detected by the PLM and IWM, respectively. In Figure 3, there were no notable differences among the ULC from the UMLM, with 3 potential perturbations, all with similar timing, being detected. However, the ULC from the PLM showed a substantial upward deviation with 7 potential perturbations. In the IWM, the ULC appeared atypical in shape due to the extensive identification and removal of potential outlier data points during early lactation, with 2 potential perturbations.

Goodness-of-Fit of the ULC and Computational Efficiency

As shown in Table 3, across all unsupervised machine learning algorithms, we found no significant difference between the MAE of the ULC from the UMLM and that of the baseline Wood curve. Although the RMSE was larger and the R^2 was smaller for the ULC from the UMLM compared with the baseline Wood curve (both $P < 0.05$), these differences were trivial. Among the UMLM, the number of curve-fitting processes and the number of potential outliers identified by the 3 algorithms varied significantly ($P < 0.05$), though the differences were minor. Additionally, the local outlier factor algorithm required a significantly ($P < 0.05$) shorter duration for the curve-fitting process compared with the other 2 UMLM algorithms.

In contrast, when compared with the ULC from the PLM and IWM, those from the UMLM exhibited significantly ($P < 0.05$) smaller MAE and RMSE, and higher R^2 . The UMLM required fewer iterations with shorter durations for the curve-fitting process compared with the PLM, although more than the IWM (both $P < 0.05$). The number of potential outliers identified by the UMLM was significantly ($P < 0.05$) lower than that identified by the IWM.

Shape of the ULC

As shown in Table 4, regardless of the algorithm, all parameters a , b , and c , as well as the predicted DIM at peak milk yield of the ULC from the UMLM, were not significantly different from those of the baseline Wood curve. However, the estimated peak milk yield and the estimated 305-d milk yield of the ULC from the UMLM were both significantly ($P < 0.05$) higher compared with those of the baseline Wood curve.

On the other hand, the parameter a of the ULC from the UMLM was larger than that from the PLM and smaller than that from the IWM (both $P < 0.05$), which exhibited a large SD. Conversely, the parameters b and c were lower than those from the PLM and higher than those from the IWM (both $P < 0.05$). Both the estimated peak

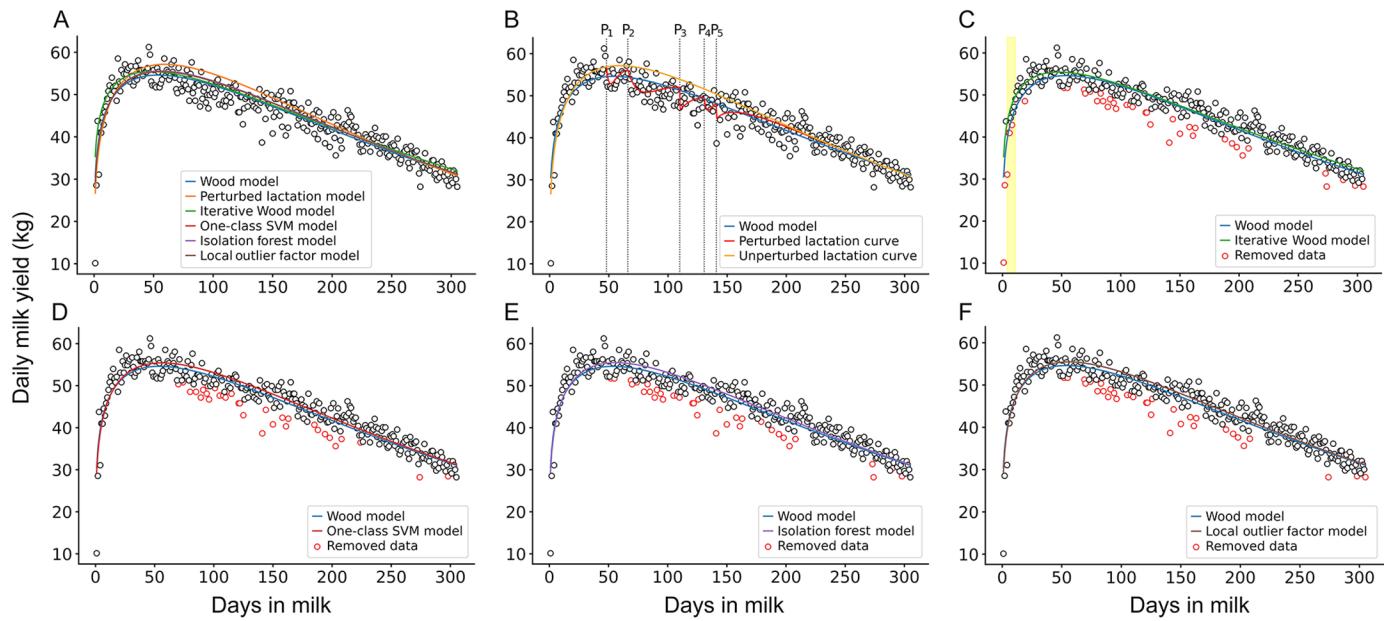


Figure 2. Baseline Wood curve and the unperturbed lactation curves (ULC) from 5 lactation curve fitting models: a representative case where the fitted curves did not show substantial differences. (A) Six fitted curves from the Wood model utilizing all daily milk yield data (blue), the perturbed lactation model (PLM; orange), the iterative Wood model (IWM; green), and unsupervised machine learning models (UMLM) utilizing one-class support vector machines (SVM; red), isolation forest (purple), and local outlier factor (brown). (B) Three fitted curves from the Wood model utilizing all daily milk yield data (blue), and the perturbated lactation curve (red) and ULC (orange), both from the PLM. (C) Two fitted curves from the Wood model utilizing all daily milk yield data (blue) and the IWM (green). (D) Two fitted curves from the Wood model utilizing all daily milk yield data (blue) and the UMLM utilizing one-class SVM (red). (E) Two fitted curves from the Wood model utilizing all daily milk yield data (blue) and the UMLM utilizing isolation forest (purple). (F) Two fitted curves from the Wood model utilizing all daily milk yield data (blue) and the UMLM utilizing local outlier factor (brown). The dotted vertical lines in panel B indicate the start day of identified potential perturbations, and the numbers above the lines (i.e., P_1 to P_5) correspond to the number of each potential perturbation. The yellow vertical bar in panel C represents the identified potential perturbation period.

milk yield and the estimated 305-d milk yield of the ULC from the UMLM were significantly ($P < 0.05$) smaller than those from the PLM. Although the estimated 305-d milk yield was also significantly ($P < 0.05$) smaller than that from the IWM, the estimated peak milk yield was not statistically different from that of the IWM. The estimated DIM at peak milk yield of the ULC from the UMLM was smaller than that from the PLM and larger than that from the IWM (both $P < 0.05$).

Perturbation Estimations Based on the ULC

As shown in Table 5, regardless of the algorithm, the number of lactations classified as potentially perturbed based on the ULC from the UMLM, as well as the number of potential perturbations and the number of potential perturbations per lactation, were higher than those from the baseline Wood curve. As shown in Figure 4, the distribution of the start days of potential perturbations identified based on the ULC from the UMLM was similar to that of the baseline Wood curve, with 3 peaks at the onset of lactation (within the first 10 d), around the peak milk yield (~60 DIM), and in late lactation

(~280 DIM). However, the second peak was more pronounced in the UMLM.

In contrast, the ULC from UMLM identified fewer lactations as potentially perturbed, fewer potential perturbations, and fewer potential perturbations per lactation compared with those from the PLM and IWM, regardless of the algorithm used (Table 5). The distribution of the start days of potential perturbations identified by the PLM showed only 2 peaks around 60 DIM and 280 DIM, with a high incidence of potential perturbations throughout the lactation period (Figure 4). On the other hand, the ULC from the IWM identified a high incidence of potential perturbations at the onset of lactation and before 60 DIM, as well as an additional peak in late lactation (Figure 4).

DISCUSSION

In this study, unsupervised machine learning techniques, including one-class SVM, isolation forest, and local outlier factor, were applied to daily milk yield data for the first time and were found to be effective in detecting potential outliers and estimating the ex-

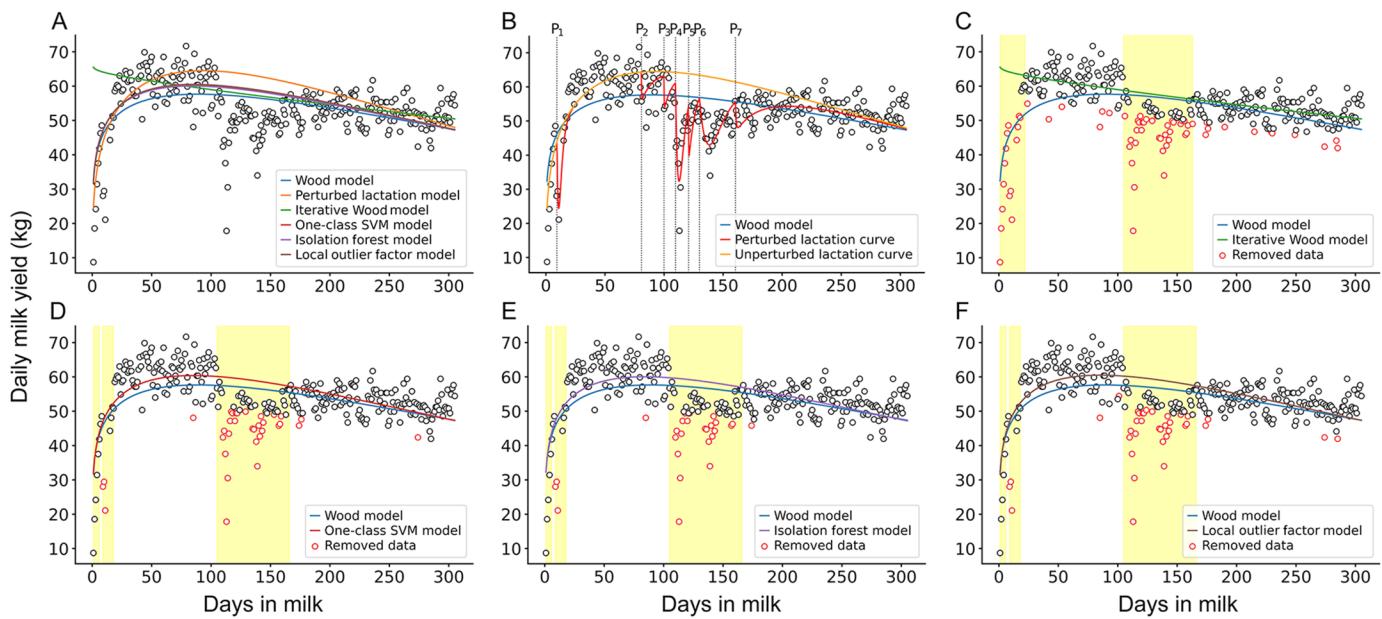


Figure 3. Baseline Wood curve and the unperturbed lactation curves (ULC) from 5 lactation curve fitting models: a representative case where the fitted curves varied remarkably. (A) Six fitted curves from the Wood model utilizing all daily milk yield data (blue), the perturbed lactation model (PLM: orange), the iterative Wood model (IWM: green), and unsupervised machine learning models (UMLM) utilizing one-class support vector machines (SVM; red), isolation forest (purple), and local outlier factor (brown). (B) Three fitted curves from the Wood model utilizing all daily milk yield data (blue), and the perturbated lactation curve (red) and ULC (orange), both from the PLM. (C) Two fitted curves from the Wood model utilizing all daily milk yield data (blue) and the IWM (green). (D) Two fitted curves from the Wood model utilizing all daily milk yield data (blue) and the UMLM utilizing one-class SVM (red). (E) Two fitted curves from the Wood model utilizing all daily milk yield data (blue) and the UMLM utilizing isolation forest (purple). (F) Two fitted curves from the Wood model utilizing all daily milk yield data (blue) and the UMLM utilizing the local outlier factor (brown). The dotted vertical lines in panel B indicate the start day of identified potential perturbations, and the numbers above the lines (i.e., P₁ to P₇) correspond to the number of each potential perturbation. The yellow vertical bars in panels C to F represent the identified potential perturbation periods.

pected lactation curve in the absence of perturbations. The simulation results supported the effectiveness of the UMLM by demonstrating that they achieved a more balanced adjustment of the lactation curve and improved overall performance in perturbation detection compared with the other models. In the analysis of the observed data, the resulting ULC from the UMLM maintained a consistent shape with the baseline Wood curve, with a reasonable upward shift, regardless of the algorithm used. Among the UMLM, the local outlier factor algorithm appeared to be the most efficient, based on its short duration for the curve-fitting process. Furthermore, the UMLM may outperform the previously proposed PLM in computational efficiency and the IWM in identifying potential outliers, as well as in estimating the lactation curve to mitigate the effects of perturbations in a more objective manner. These findings suggest that the present UMLM can be applied to daily milk yield data in real farm settings and may offer useful insights for improving farming practices and enhancing the profitability of dairy farms.

Based on the simulation study, more perturbations were identified and higher sensitivity was achieved using the

ULC from the UMLM compared with the baseline Wood curve, whereas precision did not differ among them. These results indicate that the UMLM were capable of detecting relatively small perturbations that went undetected by the baseline Wood curve. This improvement was due to the upward shift in the fitted curve, which likely reflects the effective detection and removal of outlier data points. The PLM detected even more perturbations and achieved higher sensitivity, which aligns with its formulation that allows overlapping perturbations to be detected as separate events. However, its low precision indicates that many short, minor fluctuations were misclassified as perturbations. The ULC from the IWM also led to the prediction of a large number of perturbations. Although its sensitivity was comparable to that of the UMLM, its precision was lower despite the use of the same definition of perturbation. This reflects suboptimal detection and removal of outliers, which in turn led to less effective curve adjustment. As a result, the UMLM achieved significantly higher F₁ scores than the baseline Wood curve, PLM, and IWM, indicating a more balanced adjustment of the lactation curve and improved perturbation detection performance.

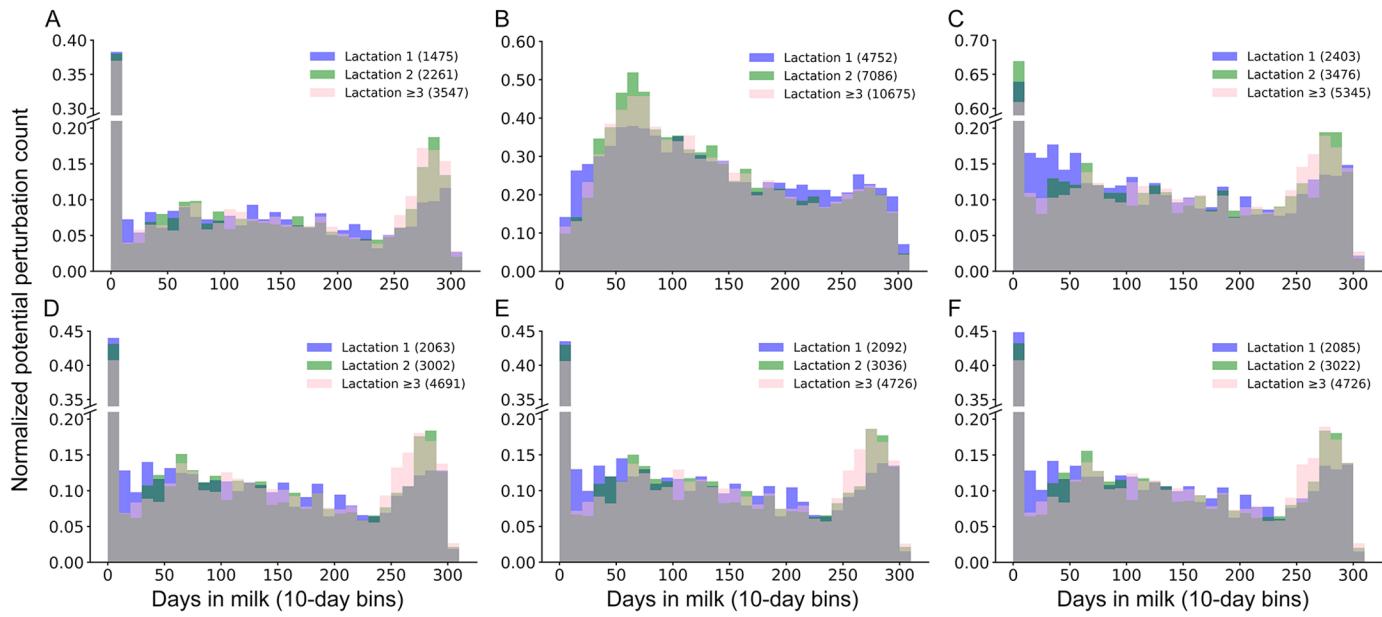


Figure 4. Normalized potential perturbation start day counts identified based on the baseline Wood curve and unperturbed lactation curves (ULC) from 5 lactation curve fitting models across lactation. (A) Normalized potential perturbation counts identified based on the lactation curve from the Wood model utilizing all daily milk yield data. (B) Normalized potential perturbation counts identified by the perturbed lactation model. (C) Normalized potential perturbation counts identified based on the ULC from the iterative Wood model. (D) Normalized potential perturbation counts identified based on the ULC from the unsupervised machine learning model (UMLM) utilizing one-class support vector machines. (E) Normalized potential perturbation counts identified based on the ULC from the UMLM utilizing isolation forest. (F) Normalized potential perturbation counts identified based on the ULC from the UMLM utilizing local outlier factor. The normalized potential perturbation count was calculated by dividing the number of potential perturbations in each 10-d bin by the number of lactations (lactation 1, 2, and ≥ 3 were 593, 886, and 1,352, respectively). Values in parentheses indicate the total number of potential perturbations identified.

According to the analysis of the observed data, the ULC from the UMLM and the baseline Wood curve exhibited similar shapes, as there were no significant differences in parameters a , b , and c , as well as in the predicted DIM of peak milk yield between them. However, both the predicted peak milk yield and the predicted 305-d milk yield were higher for the ULC from the UMLM. These findings suggest that the current UMLM shift the lactation curve upward while maintaining consistency with the shape of the baseline Wood curve. Although a direct comparison is difficult due to the lack of health records in this study, the increases of 2.3% (1.2 kg) and 2.4% (~ 330 kg) in the predicted peak milk yield and the predicted 305-d milk yield, respectively, compared with the baseline Wood curve, appear plausible. This might be supported by previous studies by Macmillan et al. (2021) and Carvalho et al. (2019), which reported that cows without health disorders during early lactation, specifically during the first 60 and 21 DIM, produced 1.7% and 4.4% more peak milk and 3.6% and 4.1% more 305-d milk, respectively, compared with cows with one or more health disorders. In these studies, 61% of cows in Macmillan et al. (2021) and 30% of cows in Carvalho et al. (2019) experienced at least one health disorder during early lactation, which would result in increases of 1.0% to 1.3% and 1.2% to

2.2% in herd averages for peak milk yield and 305-d milk yield, respectively, assuming all cows were healthy during early lactation. Given the high incidence of health disorders during early lactation, such as 66% of all cows diagnosed with at least one clinical disease by 305 DIM in the study by Carvalho et al. (2019), the observed upward shift seems reasonable.

Regarding the ULC from the PLM and IWM, the upward shifts in predicted peak milk yield (5.1% and 3.8%, corresponding to 2.7 kg and 2.0 kg, respectively) and in predicted 305-d milk yield (4.7% and 3.7%, corresponding to 635.4 kg and 503.6 kg, respectively) compared with the baseline Wood curve suggest that these 2 models may have excessively elevated the fitted curves, potentially leading to an overestimation of the cows' lactation potential. This overestimation might be supported by their larger MAE and RMSE values, along with lower R^2 values, compared with the baseline Wood curve and the ULC from UMLM. Moreover, significant differences in the parameters a , b , and c , as well as in the DIM of peak milk yield, indicate that the shapes of the ULC from these models have deviated substantially from the baseline Wood curve. These findings imply that the curve fitting performed by the PLM and IWM reduced the ability of their ULC to accurately capture the overall trends in daily

milk yield data. Additionally, the large SD of parameter a in the ULC from the IWM highlights instability in this model, particularly during early lactation. Furthermore, the PLM required a significantly higher computational cost compared with the other models, taking much longer to obtain the ULC due to the large number of curve fittings and parameters to be determined (i.e., $4 + 4 \times n$ parameters). These findings are consistent with previous studies comparing the PLM and IWM, which reported that although the PLM is computationally expensive, the IWM may suffer from unstable fitting, especially during early lactation (Ranzato et al., 2024).

The present distributions of potential perturbations, identified based on the baseline Wood curve and the ULC from the UMLM, were similar to the daily percentages of diseased cows across lactations shown in the previous study (Becker et al., 2021). The presence of 3 peaks at the onset of lactation, before peak milk yield, and in late lactation in the distribution of potential perturbations seems logical. This is because many diseases known to affect milk yield, such as dystocia, retained placenta, metritis, and mastitis, occur at calving or are most frequent in the first week of lactation (Carvalho et al., 2019). Before peak milk yield, cows experience energy deficiency (Waltner et al., 1993), making them highly susceptible to metabolic and infectious diseases (Sharma, 2010). In late lactation, several diseases, such as subclinical mastitis and lameness, occur more frequently than in other periods (Gallo et al., 2002; Hagnestam-Nielsen et al., 2009), which may be caused by increased exposure to possible infection and permanent tissue damage from previous infections (Hortet et al., 1999). However, it should be noted that the current inclusion criterion, which requires at least 300 daily records between 1 and 305 DIM, likely resulted in the selective exclusion of cows with shortened lactation periods, often due to early culling or dry-off. This may have led to an underrepresentation of perturbations associated with late-lactation health disorders in our analysis (Fourichon et al., 1999). The higher incidence of potential perturbations before peak milk yield identified by the ULC from the UMLM, compared with the baseline Wood curve, can be attributed to the upward shift in the fitting curve. This enhanced detection of potential perturbations may allow for better identification of health disorders and the implementation of effective prevention measures. Additionally, identifying cows that exhibit fewer milk yield perturbations and are less susceptible to health disorders can aid in selecting individuals and optimizing breeding strategies to improve farm productivity (Taghipoor et al., 2023).

As discussed previously and also seen in Figures 2 and 3, the significantly high number of potential perturbations identified by the PLM appears to result from its detection of short and minor milk yield fluctuations as

perturbations, as well as the individual identification of overlapping perturbations. The absence of a clear peak at the onset of lactation in the distributions of potential perturbations might be attributed to the lower bound for the perturbation onset time (t_{Pi}) in the PLM, which requires at least 3 d of available data before a perturbation can be detected (Ben Abdelkrim et al., 2021). This could be a critical drawback of this model, given that health disorders most often occur around the onset of lactation, as mentioned earlier. Unlike the other models, the IWM identified a high incidence of potential perturbations during early lactation (before 60 DIM), with a particularly high incidence at the onset of lactation. This may be attributed to the characteristics of the Wood model, which tends to overestimate milk yield during early lactation (Cobby and Le Du, 1978). Consequently, the IWM may erroneously classify many data points as outliers and filter them during early lactation, leading to an excessive upward shift in the fitting curve or an atypical curve fit. This could result in the incorrect detection of numerous potential perturbations during early lactation. In addition to these limitations, both the PLM and IWM require arbitrary value settings, such as the maximum number of perturbations to be detected in the PLM (Ben Abdelkrim et al., 2021) and a threshold for removing low milk yield data in the IWM (Adriaens et al., 2021). These values are determined based on visual evaluation of the fitted curve and the available computational resources for curve fitting (Adriaens et al., 2021; Ranzato et al., 2024). In contrast, in the UMLM, the appropriate proportion of outliers is determined automatically through unsupervised machine learning algorithms, allowing outlier detection and the identification of potential perturbations to be carried out more objectively.

The findings of this study indicate that the present UMLM offer practical advantages, including stable fitting with relatively low computational cost and an objective, reasonable upward modification of the lactation curve. However, this study has several limitations. First, the analysis was conducted using data from only Holstein cows at a single dairy farm, which may limit the generalizability of the findings to other farm settings with different herd structures and management practices. Factors such as cow breed (Adriaens et al., 2023), parity (Silvestre et al., 2009), calving season (Wood, 1969), frequency of milk yield data collection (Sitkowska et al., 2020), and lactation period (Kopec et al., 2021) are known to influence the goodness-of-fit of the Wood curve and, consequently, the detection of outlier data points by UMLM. Second, the current study lacked health records for the cows, making it challenging to directly correlate the detected potential outliers and perturbations with actual occurrences of health disorders. This is critical because the incidence rate of health disorders varies substantially

among farms (Gonçalves et al., 2022). Moreover, the frequency and impact of external factors such as heat stress, dietary changes, and management interventions also differ considerably between farms. These internal and external influences can strongly affect both the number of potential outliers and the degree of the upward shift in the fitting curve, as well as the number of potential perturbations. Therefore, further studies should include data from multiple farms and incorporate comprehensive health, management, and environmental records (e.g., heat stress indicators) to validate the applicability and robustness of UMLM across diverse dairy farm settings.

In addition to such farm-level validation, careful consideration should be given to how the lactation curve is modeled, as it directly affects the accurate estimation of lactation potential. In this study, all models were fitted at the individual animal level using the classical Wood function and did not include group-level fixed effects such as breed, parity, or calving season. Although this approach ensured a uniform modeling framework, omitting these effects may compromise the model's ability to estimate lactation potential in more heterogeneous herds, as these factors are known to influence lactation curve characteristics. Furthermore, accurately capturing a cow's lactation potential requires appropriate modeling of the entire lactation curve. However, the present Wood model-based UMLM cannot account for the long-term effects of health disorders (Wilson et al., 2004). Although short-term decreases in milk yield (perturbations) may be handled by outlier removal, this approach may be insufficient when milk yield remains at a reduced level without full recovery. To address these limitations, more flexible modeling approaches such as spline-based regression, segmented models, or nonlinear mixed-effects models should be considered (Macciotta et al., 2011). Because UMLM operate on the residuals from a fitted curve, they are compatible with any lactation curve model. This means that the Wood model used as the base model for the UMLM can be readily replaced with a more flexible alternative without altering the core outlier detection framework. Incorporating such models could improve the estimation of each cow's lactation potential by better capturing structural variability in the lactation curve and the persistent effects of health disorders.

CONCLUSIONS

This study demonstrates that unsupervised machine learning techniques can effectively detect potential outliers in daily milk yield data and provide reliable estimates of lactation curves in the absence of perturbations. The upward shifts in the ULC from the UMLM, compared with the baseline Wood curve, were consistent with previous findings on the impact of health disorders on

milk yield. These upward shifts were more conservative than those from the previously proposed IWM and PLM but remained within a reasonable range, highlighting the potential of UMLM for accurately assessing dairy cows' lactation potential. Additionally, the UMLM may outperform the PLM in computational efficiency and IWM in identifying potential outliers, as well as in estimating the lactation curve to mitigate the effects of perturbations in more objective manner. These findings suggest that the UMLM have the potential to be applied to daily milk yield data in real farm settings and may offer useful insights for improving farming practices and enhancing the profitability of dairy farms. However, the generalizability of the findings may be limited by the use of data from only Holstein cows at a single farm and the lack of records on potential confounding factors. Moreover, the current UMLM do not account for fixed effects or the long-term impacts of health disorders. To enhance robustness and applicability, future studies should consider more flexible modeling approaches and multifarm datasets with detailed background records.

NOTES

This study received no external funding. Because this study did not involve animal procedures, it was exempt from requiring approval by an Institutional Animal Care and Use Committee. The authors have not stated any conflicts of interest.

Nonstandard abbreviations used: FN = false negative; FP = false positive; IWM = iterative Wood model; MAE = mean absolute error; PLM = perturbed lactation model; Q-Q = quantile-quantile; RMSE = root mean square error; SVM = support vector machines; TP = true positive; ULC = unperturbed lactation curve; UMLM = unsupervised machine learning model.

REFERENCES

- Adriaens, I., G. Bonekamp, J. Ten Napel, C. Kamphuis, and Y. De Haas. 2023. Differences across herds with different dairy breeds in daily milk yield based proxies for resilience. *Front. Genet.* 14:1120073. <https://doi.org/10.3389/fgene.2023.1120073>.
- Adriaens, I., I. Van Den Brulle, L. D'Anvers, J. Statham, K. Geerinckx, S. De Vliegher, S. Piepers, and B. Aernouts. 2021. Milk losses and dynamics during perturbations in dairy cows differ with parity and lactation stage. *J. Dairy Sci.* 104:405–418. <https://doi.org/10.3168/jds.2020-19195>.
- Bach, A. 2023. Changes in milk production and estimated income over feed cost of group-housed dairy cows when moved between pens. *J. Dairy Sci.* 106:4108–4120. <https://doi.org/10.3168/jds.2022-22875>.
- Becker, V. A. E., E. Stamer, and G. Thaller. 2021. Liability to diseases and their relation to dry matter intake and energy balance in German Holstein and Fleckvieh dairy cows. *J. Dairy Sci.* 104:628–643. <https://doi.org/10.3168/jds.2020-18579>.
- Ben Abdelkrim, A., L. Puillet, P. Gomes, and O. Martin. 2021. Lactation curve model with explicit representation of perturbations as a pheno-

- typing tool for dairy livestock precision farming. *Animal* 15:100074. <https://doi.org/10.1016/j.animal.2020.100074>.
- Bouallegue, M., and N. M'Hamdi. 2020. Chapter 6. Mathematical modeling of lactation curves: A review of parametric models. Pages 95–114 in *Lactation in Farm Animals: Biology, Physiological Basis, Nutritional Requirements, and Modelization*. N. M'Hamdi, ed. IntechOpen Ltd.
- Breunig, M. M., H.-P. Kriegel, R. T. Ng, and J. Sander. 2000. LOF: Identifying density-based local outliers. Pages 93–104 in Proc. 2000 ACM SIGMOD Int. Conf. Manage. Data. Dallas, TX. <https://doi.org/10.1145/335191.335388>.
- Carvalho, M. R., F. Peñagaricano, J. E. P. Santos, T. R. DeVries, B. W. McBride, and E. S. Ribeiro. 2019. Long-term effects of postpartum clinical disease on milk production, reproduction, and culling of dairy cows. *J. Dairy Sci.* 102:11701–11717. <https://doi.org/10.3168/jds.2019-17025>.
- Chandola, V., A. Banerjee, and V. Kumar. 2009. Anomaly detection: A survey. *ACM Comput. Surv.* 41:1–58. <https://doi.org/10.1145/1541880.1541882>.
- Chicco, D., M. J. Warrens, and G. Jurman. 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* 7:e623. <https://doi.org/10.7717/peerj-cs.623>.
- Choumei, Y., A. Kahi, and H. Hirooka. 2006. Fit of Wood's function to weekly records of milk yield, total digestible nutrient intake and body weight changes in early lactation of multiparous Holstein cows in Japan. *Livest. Sci.* 104:156–164. <https://doi.org/10.1016/j.livsci.2006.04.015>.
- Cobby, J., and Y. Le Du. 1978. On fitting curves to lactation data. *Anim. Sci.* 26:127–133. <https://doi.org/10.1017/S0003356100039532>.
- Darrab, S., H. Allipilli, S. Ghani, H. Changaramkulath, S. Koneru, D. Brioneske, and G. Saake. 2024. Anomaly Detection Algorithms: Comparative Analysis and Explainability Perspectives. Pages 90–104 in Proc. Aust. Conf. Data Sci. and ML. Auckland, New Zealand. https://doi.org/10.1007/978-991-99-8696-5_7.
- Fourichon, C., H. Seegers, N. Bareille, and F. Beaudeau. 1999. Effects of disease on milk production in the dairy cow: A review. *Prev. Vet. Med.* 41:1–35. [https://doi.org/10.1016/S0167-5877\(99\)00035-5](https://doi.org/10.1016/S0167-5877(99)00035-5).
- Gallo, L., P. Carnier, M. Cassandro, and F. Cesarini. 2002. Health disorders and their association with production and functional traits in Holstein Friesian cows. *Ital. J. Anim. Sci.* 1:197–210. <https://doi.org/10.4081/ijas.2002.197>.
- Gonçalves, J. L., J. L. de Campos, A. J. Steinberger, N. Safdar, A. Kates, A. Sethi, J. Shutske, G. Suen, T. Goldberg, R. I. Cue, and P. L. Ruegg. 2022. Incidence and treatments of bovine mastitis and other diseases on 37 dairy farms in Wisconsin. *Pathogens* 11:1282. <https://doi.org/10.3390/pathogens1111282>.
- Guinan, F. L., R. H. Fourdraine, F. Peñagaricano, and K. A. Weigel. 2024. Genetic analysis of lactation consistency in US Holsteins using temporal variation in daily milk weights. *J. Dairy Sci.* 107:2194–2206. <https://doi.org/10.3168/jds.2023-24093>.
- Hagnestam-Nielsen, C., U. Emanuelson, B. Berglund, and E. Strandberg. 2009. Relationship between somatic cell count and milk yield in different stages of lactation. *J. Dairy Sci.* 92:3124–3133. <https://doi.org/10.3168/jds.2008-1719>.
- Harris, C. R., K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. 2020. Array programming with NumPy. *Nature* 585:357–362. <https://doi.org/10.1038/s41586-020-2649-2>.
- Hodson, T. O. 2022. Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geosci. Model Dev.* 15:5481–5487. <https://doi.org/10.5194/gmd-15-5481-2022>.
- Hortet, P., F. Beaudeau, H. Seegers, and C. Fourichon. 1999. Reduction in milk yield associated with somatic cell counts up to 600 000 cells/ml in French Holstein cows without clinical mastitis. *Livest. Prod. Sci.* 61:33–42. [https://doi.org/10.1016/S0301-6226\(99\)00051-2](https://doi.org/10.1016/S0301-6226(99)00051-2).
- Hou, Y., L. Zhang, R. Y. Dong, M. Y. Liang, Y. Lu, X. Q. Sun, and X. Zhao. 2021. Comparing responses of dairy cows to short-term and long-term heat stress in climate-controlled chambers. *J. Dairy Sci.* 104:2346–2356. <https://doi.org/10.3168/jds.2020-18946>.
- Kopeć, T., G. Chládek, D. Falta, J. Kučera, M. Večeřa, and O. Hanuš. 2021. The effect of extended lactation on parameters of Wood's model of lactation curve in dairy Simmental cows. *Anim. Biosci.* 34:949–956. <https://doi.org/10.5713/ajas.20.0347>.
- Li, M., G. J. M. Rosa, K. F. Reed, and V. E. Cabrera. 2022. Investigating the effect of temporal, geographic, and management factors on US Holstein lactation curve parameters. *J. Dairy Sci.* 105:7525–7538. <https://doi.org/10.3168/jds.2022-21882>.
- Liu, F. T., K. M. Ting, and Z.-H. Zhou. 2008. Isolation forest. Pages 413–422 in Proc. 2008 Eighth IEEE Int. Conf. Data Mining. Pisa, Italy. IEEE. <https://doi.org/10.1109/ICDM.2008.17>.
- Lucey, S., G. J. Rowlands, and A. M. Russell. 1986. Short-term associations between disease and milk yield of dairy cows. *J. Dairy Res.* 53:7–15. <https://doi.org/10.1017/S0022029900024602>.
- Macciotta, N. P. P., C. Dimauro, S. P. G. Russu, R. Steri, and G. Pulina. 2011. The mathematical description of lactation curves in dairy cattle. *Ital. J. Anim. Sci.* 10:e51. <https://doi.org/10.4081/ijas.2011.e51>.
- Macmillan, K., M. Gobikrushanth, A. Behrouzi, B. Hoff, and M. G. Colazo. 2021. Prevalence of early postpartum health disorders in Holstein cows and associations with production, reproduction, and survival outcomes on Alberta dairy farms. *Can. Vet. J.* 62:273–280.
- McKinney, W. 2010. Data structures for statistical computing in Python. Pages 51–56 in Proc. of the 9th Python in Science Conf. (SciPy 2010), Austin, TX. <https://doi.org/10.25080/Majora-92bf1922-00a>.
- Moncur, V. S., L. C. Hardie, and C. D. Dechow. 2021. Genetic analysis of daily milk yield variability in Holstein dairy cattle in an experimental herd. *Livest. Sci.* 244:104397. <https://doi.org/10.1016/j.livsci.2021.104397>.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg. 2011. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12:2825–2830.
- Ranzato, G., B. Aernouts, I. Lora, I. Adriaens, A. Ben Abdelkrim, M. Gote, and G. Cozzi. 2024. Comparison of 3 mathematical models to estimate lactation performance in dairy cows. *J. Dairy Sci.* 107:6888–6901. <https://doi.org/10.3168/jds.2023-24224>.
- Rasmussen, P., H. W. Barkema, P. P. Osei, J. Taylor, A. P. Shaw, B. Conrady, G. Chaters, V. Muñoz, D. C. Hall, O. O. Apenteng, J. Rushton, and P. R. Torgerson. 2024. Global losses due to dairy cattle diseases: A comorbidity-adjusted economic analysis. *J. Dairy Sci.* 107:6945–6970. <https://doi.org/10.3168/jds.2023-24626>.
- Schölkopf, B., J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. 2001. Estimating the support of a high-dimensional distribution. *Neural Comput.* 13:1443–1471. <https://doi.org/10.1162/089976601750264965>.
- Seabold, S., and J. Perktold. 2010. Statsmodels: Econometric and statistical modeling with python. Pages 92–96 in Proc. The 9th Python Sci. Conf. Austin, TX. <https://doi.org/10.25080/Majora-92bf1922-011>.
- Seliya, N., A. Abdollah Zadeh, and T. M. Khoshgoftaar. 2021. A literature review on one-class classification and its potential applications in big data. *J. Big Data* 8:122. <https://doi.org/10.1186/s40537-021-00514-x>.
- Sharma, N. 2010. Economically important production diseases of dairy animals. Pages 47–65 in SMVS Dairy Year Book. Sarva Manav Vikash Samiti.
- Silvestre, A., A. Martins, V. Santos, M. Ginja, and J. Colaço. 2009. Lactation curves for milk, fat and protein in dairy cows: A full approach. *Livest. Sci.* 122:308–313. <https://doi.org/10.1016/j.livsci.2008.09.017>.
- Sitkowska, B., M. Kolenda, and D. Piwczyński. 2020. Comparison of the fit of automatic milking system and test-day records with the use of lactation curves. *Asian-Australas. J. Anim. Sci.* 33:408–415. <https://doi.org/10.5713/ajas.19.0190>.
- Taghipoor, M., M. Pastell, O. Martin, H. Nguyen Ba, J. van Milgen, A. Doeschl-Wilson, C. Loncke, N. Friggens, L. Puillet, and R. Muñoz-

- Tamayo. 2023. Animal board invited review: Quantification of resilience in farm animals. *Animal* 17:100925. <https://doi.org/10.1016/j.animal.2023.100925>.
- Terpilowski, M. A. 2019. scikit-posthocs: Pairwise multiple comparison tests in Python. *J. Open Source Softw.* 4:1169. <https://doi.org/10.21105/joss.01169>.
- Virtanen, P., R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. P. Bardell, A. Rothberg, A. Hilboll, A. Klocekner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, É. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G.-L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavić, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodriguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmeler, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko, and Y. Vázquez-Baeza. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17:261–272. <https://doi.org/10.1038/s41592-019-0686-2>.
- von Keyserlingk, M. A. G., D. Olenick, and D. M. Weary. 2008. Acute behavioral effects of regrouping dairy cows. *J. Dairy Sci.* 91:1011–1016. <https://doi.org/10.3168/jds.2007-0532>.
- Waltner, S. S., J. P. McNamara, and J. K. Hillers. 1993. Relationships of body condition score to production variables in high producing Holstein dairy cattle. *J. Dairy Sci.* 76:3410–3419. [https://doi.org/10.3168/jds.S0022-0302\(93\)77679-1](https://doi.org/10.3168/jds.S0022-0302(93)77679-1).
- Wang, X., X. Wang, and M. Wilkes. 2021. Chapter 2. Developments in unsupervised outlier detection research. Pages 13–38 in *New Developments in Unsupervised Outlier Detection*. X. Wang, X. Wang, and M. Wilkes, ed. Springer, Singapore.
- Wasike, C. B., A. Kahi, and K. J. Peters. 2011. Modelling of lactation curves of dairy cows based on monthly test day milk yield records under inconsistent milk recording scenarios. *Animal* 5:1780–1790. <https://doi.org/10.1017/S1751731111000954>.
- Wilson, D. J., R. N. González, J. Hertl, H. F. Schulte, G. J. Bennett, Y. H. Schukken, and Y. T. Gröhn. 2004. Effect of clinical mastitis on the lactation curve: a mixed model estimation using daily milk weights. *J. Dairy Sci.* 87:2073–2084. [https://doi.org/10.3168/jds.S0022-0302\(04\)70025-9](https://doi.org/10.3168/jds.S0022-0302(04)70025-9).
- Wood, P. D. P. 1967. Algebraic model of the lactation curve in cattle. *Nature* 216:164–165. <https://doi.org/10.1038/216164a0>.
- Wood, P. D. P. 1969. Factors affecting the shape of the lactation curve in cattle. *Anim. Sci.* 11:307–316. <https://doi.org/10.1017/S0003356100026945>.
- Wood, P. D. P. 1980. Breed variations in the shape of the lactation curve of cattle and their implications for efficiency. *Anim. Sci.* 31:133–141. <https://doi.org/10.1017/S0003356100024363>.

ORCIDS

- Shogo Higaki,  <https://orcid.org/0000-0002-5595-6750>
 Eduardo Noronha de Andrade Freitas,  <https://orcid.org/0009-0005-4154-9707>
 Ariana Negreiro,  <https://orcid.org/0000-0002-0154-1509>
 João R. R. Dórea,  <https://orcid.org/0000-0001-9849-7358>
 Victor E. Cabrera  <https://orcid.org/0000-0003-1739-7457>