BURLEIGH DODDS SERIES IN AGRICULTURAL SCIENCE

Advances in precision dairy and beef farming technologies

Edited by Professor Daniel Berckmans and Professor Tomás Norton, Katholieke University of Leuven, Belgium

E-CHAPTER FROM THIS BOOK

burleigh dodds SCIENCE PUBLISHING

Modeling the lactation curves of dairy ruminants

Meike B. van Leerdam, Cornell University, USA; Lucia Trapanese, University of Naples Federico II, Italy; Chen Yongyan, Guangzhou Nansha Modern Agriculture Industry Group Co., China; Kristof Hermans and Matthieu Salamone, Ghent University, Belgium; and Enhong Liu and Miel Hostens, Cornell University, USA

1 Introduction
2 The different use cases of lactation curves
3 Modeling the lactation curve
4 The use of machine learning for lactation curve modeling
5 Collecting the data for the lactation curve
6 Conclusion
7 Declaration of funding and/or conflicts of interest
8 Where to look for further information
9 References

1 Introduction

A lactation curve is a visualization of milk production over time, usually describing a single lactation. The x-axis typically represents days in milk, while the y-axis often displays milk yield per unit of time. However, this structure can also be applied to other traits that change over the course of a single lactation, such as fat yield, protein yield, or somatic cell count. A typical lactation curve forms an asymmetrical shape with a sharp rise, a rounded peak, and a gradual decline (Fig. 1). This characteristic shape arises from the combined effect of several simultaneous biological processes (Sorensen et al., 2006). These include the normal growth of udder capacity during late pregnancy and early lactation, followed by the gradual senescence of secretory tissue as lactation progresses, as well as additional influences on production caused by, among others, environmental conditions, health events, and nutritional management. These underlying processes are largely occult, in the sense that they cannot be directly observed or separated in routine data; instead, only their net effect on

http://dx.doi.org/10.19103/AS.2025.0167.19
© The Authors 2026. This is an open access chapter distributed under a Creative Commons Attribution 4.0 License (CC BY).
Chapter taken from: Berckmans, D. and Norton, T. (eds.), *Advances in precision dairy and beef farming technologies*, pp. 513–558, Burleigh Dodds Science Publishing, Cambridge, UK, 2026, (ISBN: 978 1 83545 194 6; www.bdspublishing.com)

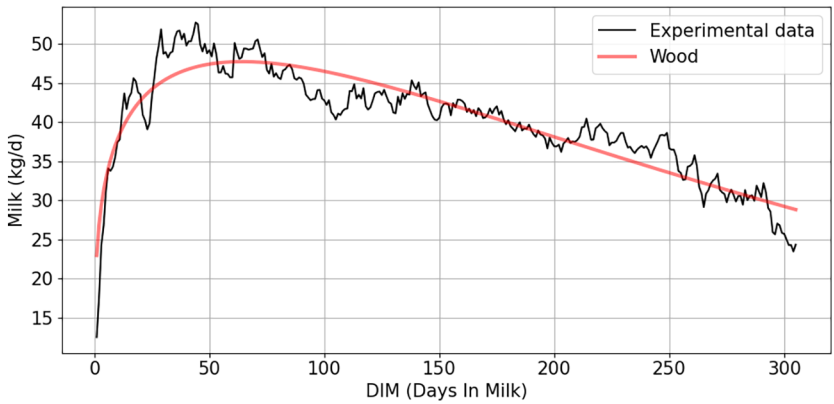

Figure 1 Example of the typical shape of the lactation curve. The Wood model (Wood, 1967) was linearly fitted to the measured daily milk yield of a single lactation.

milk production can be measured. Lactation curve models, therefore, attempt to summarize the combined outcome of these interacting processes using a mathematical representation.

Lactation curves are not a novel development; over the course of the past century, many different approaches to model lactation curves have been suggested, each with its own benefits and limitations, tailored to the specific needs of the application. For basic decision-making, such as ranking cows by cumulative production to identify which cows are most profitable to breed, keep, or cull, a lactation curve with lower accuracy may be sufficient for farmers. However, more advanced applications, such as using lactation curve perturbations for disease detection (Adriaens et al., 2021; Salamone et al., 2024), as a resilience indicator (Abdelkrim et al., 2021; Poppe et al., 2020), or predicting optimal voluntary waiting periods for individual cows (Chen et al., 2023a), require higher precision to achieve meaningful results. Despite their potential, many of these techniques have failed to be widely implemented in the dairy industry, and one possible reason for this is the limited accuracy of commonly used lactation curves (Rojas de Oliveira et al., 2024).

Lactation curves have been extensively covered in various review papers (Angeles-Hernandez et al., 2021; Beever et al., 1991; Bouallegue and M'Hamdi, 2019; Gipson and Grossman, 1990; Macciotta et al., 2011; Murphy et al., 2018; Rojas de Oliveira et al., 2024; Solodneva et al., 2022). These prior works have thoroughly covered the technical details behind lactation curve methods and applications, providing valuable insights for scientists working in this area. However, no one has offered a comprehensive overview of the considerations involved in applying or interpreting lactation curves in real-world scenarios.

In this chapter, we want to emphasize the importance of the complete methodology of the lactation curve, starting with the formulation of a use

case, followed by data collection, model formation, fitting, and validation. By advocating for more informed considerations, we hope to elevate the accuracy of the lactation curves in practice, thereby improving and expanding their use for decision-making. Furthermore, we want to stress the need for transparency and documentation of the methodologies applied, as this will facilitate the reproducibility of the lactation curve applications. Finally, this chapter highlights emerging, underexplored advancements in the field, such as the application of machine learning techniques, the integration of multi-source datasets for lactation curve modeling, and data quality management of the input features used in the models. These novel approaches have significant potential to improve the accuracy, adaptability, and applicability of lactation curves in modern dairy science.

2 The different use cases of lactation curves

Since their formulation over a century ago, lactation curves have had a widespread implementation, both in research and in practice. Their original intention, proposed by Brody et al. in 1923, remains relevant today: lactation curves can serve as a tool to compare cows at different stages of lactation, either with other cows or with themselves, based on limited milk yield measurements (Brody et al., 1923). In addition to individual lactation curves, aggregate curves facilitate the comparison of groups of cows with each other or the individual with the average.

The most accurate lactation curve is achieved by meticulously measuring every single milking session during the lactation of a cow and plotting this result over a timeline (Sitkowska et al., 2019). In the past, this approach was not feasible due to constraints in time, money, and equipment (Hare et al., 2004). Models were therefore deployed to connect different milk yield measurements and subsequently use this graph to estimate the cumulative milk yield and the shape of the lactation curve. Nowadays, with increased use of continuous milk measurements through milk meters, the course of the lactation can be assessed in near real-time. This development does, however, not make lactation curve models obsolete, but their focus has shifted to predicting the course of the lactation rather than evaluating the past lactation, and as a tool to remove measurement noise from yield data. The application of lactation curves can be divided into three categories: genetic selection, herd management, and research.

2.1 Genetic selection

One of the most widespread applications of lactation curves is in genetic selection. Traditionally, lactation curves were used to obtain an estimation of the

cumulative 305-day milk production based on limited test-day milk recordings. The result was used as a single trait to calculate breeding values and identify genomic markers (Cole et al., 2023). This methodology is known as the lactation yield model (Wu et al., 2023a). One of the simplest ways to do this is by applying the Test Interval Method, which is not based upon a lactation curve model per se, but instead takes the average milk production between two milk recordings and adds these up to generate a full lactation yield (Sargent et al., 1968). This method is still used by many milk recording organisatios and is the standard in buffalo breeding value estimation (Gómez-Carpio et al., 2023). However, in order to produce timely genetic predictions, lactation records can be projected to a 305-day basis after only two or three monthly milk recordings, thanks to lactation curve models. This has been applied since 1975 as part of the USDA animal model, increasing genetic progress in dairy cows by up to 10% (Weigel et al., 2017).

Lactation curves for breeding have seemingly been less important since the 1990s, when test day models (TDM) gained popularity. Instead of calculating breeding values based on the cumulative 305-day production, the timing, frequency, and magnitude of the individual milk yield measurements, known as test days, influenced the resulting breeding values (Jensen, 2001). In its simplest form, the separate trait TDM, no lactation curves were applied (Wiggans and Goddard, 1997). This model assumed that each test day record is independent and is therefore treated as a unique trait, leading to overparameterization and excessive computations. From a biological perspective, different test day yields are more likely to be correlated with each other (Ali and Schaeffer, 1987), even though the influence of specific genes may vary throughout lactation due to time-dependent gene expression (Strucken et al., 2015, 2011). Lactation curves have therefore been brought back to model changing environmental and genetic influences on daily milk yield measurements over the course of the lactation (Weigel et al., 2017). This approach also facilitates taking into account differences in the shape of the lactation curve, which have a heritable component (De Roos et al., 2001; Jamrozik et al., 1997; Rojas de Oliveira et al., 2024; Schaeffer et al., 2000). The test day yields can be calculated in multiple ways, including the repeatability TDM, the random regression TDM, and the block multiple-trait TDM (Misztal et al., 2000). The structure of the model greatly influenced the number of parameters that need to be estimated to deploy a TDM for breeding value calculation and thereby influence computational demand (Swalve, 2000). Complex models might be better at accurately estimating the daily milk yield, but a compromise needs to be found between model complexity and computational capacity. To simplify the fitting process and reduce computational demands, lactation curves were therefore incorporated either as a fixed effect, representing an average production pattern, and/or

as a random effect, allowing for individual deviations in lactation trajectories among cows (Jensen, 2001; Rojas de Oliveira et al., 2024; Swalve, 2000). The use of test day models has resulted in more accurate breeding values compared to the breeding values calculated with the cumulative milk yield, as it can take into account changing genetic and environmental effects over the course of the lactation (Oliveira et al., 2019; Sawalha et al., 2005).

The use of lactation curves in breeding applies not only to traits related to total milk yield, but also to milk constituents such as fat, protein, and lactose, which can be modeled similarly (Costa et al., 2019; Silvestre et al., 2009). This is not only valuable from a production point of view, but there is some evidence that changes in lactose yield across lactations are an indicator of longevity and udder health (Costa et al., 2025, 2020). Additionally, lactation curve shape can be translated to breeding values, such as breeding values for persistency, lactation length, peak production magnitude, and productive life (Atashi et al., 2020; Miglior et al., 2017; VanRaden et al., 2006). Studies have shown that persistent cows can have positive effects in a herd, such as improved milk yield, improved conception rates, extended productive lifetimes, and reduced culling rates (Dekkers et al., 1998; Hadley et al., 2006; Togashi et al., 2016). Therefore, interest has grown to identify and selectively breed cows with high persistency (Cole & VanRaden, 2006; Dekkers et al., 1996; Togashi & Lin, 2009). Beyond traditional traits, lactation curves have facilitated the development of resilience breeding values, which quantify deviations from expected milk production in response to challenges during lactation (Chen et al., 2023b; Poppe et al., 2021; Wang et al., 2022). This approach could potentially reduce animal disease, making livestock management easier and improving animal welfare (Berghof et al., 2019).

Specific lactational phenotypes can also be identified (Ehrlich, 2013). For example, cows tolerating long milking intervals while maintaining high daily milk yield are suggested to be more efficient while using automated milking systems (Masía et al., 2020). Or cows who distribute their total milk production more equally over their lactation might experience a less severe negative energy balance, improving health and fertility (Strucken et al., 2015; Yamazaki et al., 2011). As an illustration, two cows with the same 305-day milk yield (M305) may follow very different lactation trajectories (Fig. 2). Cow A exhibits a lower peak production and a slower rate of decline after peak lactation, whereas Cow B has a higher peak followed by a steeper decline. Cow A may be considered superior for several reasons. During early lactation, she produces less milk, potentially reducing the severity of negative energy balance (Butler, 2005; Martens, 2023; Wathes et al., 2007). Additionally, her ability to maintain a steady milk production rate makes her a strong candidate for extended lactation strategies, potentially improving overall efficiency (Kok et al., 2019; Sehested et al., 2019; Sorensen et al., 2008). However, traditional evaluation

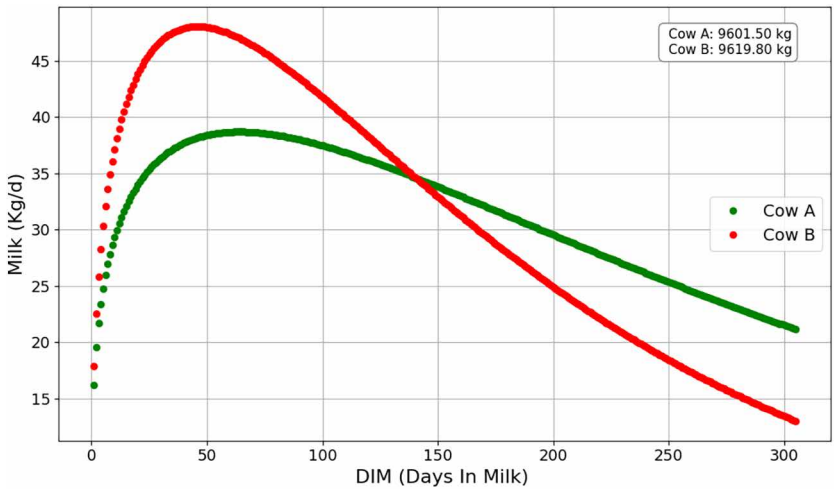

Figure 2 The lactation curves of two cows with approximately the same cumulative 305 days' milk production, but with different shapes.

metrics such as summarized milk yields may fail to capture these differences, highlighting the need for alternative assessment methods, such as metrics derived from the shape of the lactation curve.

2.2 Herd management

Although dairy farmers rarely analyze lactation curves directly, dairy herd improvement (DHI) performance parameters derived from these curves are widely used in management decisions. One of the most prominent applications is around the culling decision. Milk sales are often the primary source of income for a farm and, therefore, a key determinant of a cow's economic value. To make optimal culling decisions, farmers need to estimate both current and future milk production to be able to adequately compare a cow to a replacement heifer (Lehenbauer and Oltjen, 1998). Examples of metrics deploying lactation curves for this purpose are the Dairy Comp 305 module called **Cow Value**, commonly used in the US (Sorge et al., 2007), and **lactation value (lactatiewaarde)** by CRV used in the Netherlands (CRV, 2015).

Lactation curve-derived metrics can also be used at the herd level to track production trends over time and compare groups of cows based on factors such as parity or lactation stage. These metrics are valuable tools for identifying management inefficiencies and areas needing improvement. The simplest application is to use the curve to derive the cumulative production at a certain cut-off point, mostly at 305 days for dairy cows, after fitting an individual lactation curve based upon milk recordings during the lactation. This

cut-off point is species dependent; for instance, buffalo lactations are often capped at 270 DIM, while for sheep and goats there is no standard lactation length available as it differs too much between management systems (ICAR Wiki contributors, 2025a). Other examples include the Standardized 150-day Milk, commonly used in the US, which represents the average daily milk yield of a herd adjusted to simulate uniform conditions, accounting for parity, calving season, breed, and lactation stage (Oleggini et al., 2001). And the Net Profit ("Netto opbrengst"), a metric employed by CRV in the Netherlands. This measure quantifies the revenue of a lactation while accounting for variables such as calving season, cow age at calving, and the expected calving interval, allowing for standardized comparisons (CRV, 2020, 2015). But also, metrics describing the shape of the average lactation curve of a herd can be insightful, like the magnitude and timing of peak production.

Herd management systems provide these metrics, but there is a lack of standardized definitions, leading to wide variation, low transparency, and limited comparability across systems and countries. While some of the underlying calculations are not publicly available, they are assumed to be based on lactation curves. As a result, performance parameters for cows with atypical lactation curves may be over- or underestimated, potentially leading to misinformed management decisions.

In addition to direct applications in cow and herd management, lactation curves are used in whole-farm simulation models, such as the Ruminant Farm Systems (RuFaS) (Hansen et al., 2021; Kebreab et al., 2019), the bioeconomic model of De Vries (de Vries, 2004), and the DairyPrint model of da Silva et al. (da Silva and Cabrera, 2024). The goal of these simulations is to estimate the impact of management decisions on farm production and the environment, valuable for farmers, researchers, and policymakers.

2.3 Research

Lactation curves can provide insights into the physiological mechanisms involved in milk production. The comparison of fitted models facilitates the discovery of temporal, geographic, and management effects on lactation curve parameters, helping to explain milk yield variability (Li et al., 2022). Lactation curves allow for the analysis of disease impacts on milk production in both the short and long term, by looking at perturbations of the expected lactation curve or by comparing the lactation curve shape of diseased versus healthy cows. These approaches reveal effects that may not be apparent when examining cumulative milk yield (Adriaens et al., 2021; Andersen et al., 2011a; Hostens et al., 2012; Masia et al., 2022; Wilson et al., 2004; Yamazaki et al., 2009). Lactation curves also facilitate the calculation of economic disease impact, as lactation curves help quantify milk loss due to disease (Bruijnis et al., 2010;

Steeneveld et al., 2011, 2007). For example, Bruijnis et al. showed that hoof disorders caused a reduction in farm revenue of $75 dollars per cow per year in the Netherlands, whereby milk production loss accounted for almost half of the economic consequences, which is often underestimated by farmers (Bruijnis et al., 2010).

Similarly, explorations have been conducted into the potential associations between the shape of the lactation curve and other cow-related features such as age at first calving, the voluntary waiting period, and dry period length to help determine their optimum value and maximize milk yield (Andersen et al., 2011b; Atashi et al., 2021, 2013; Burgers et al., 2021; Calanni Macchio et al., 2025; Chen et al., 2016, 2024; Elahi Torshizi, 2016; Inchaisri et al., 2011; Lehmann et al., 2019).

Lactation curves can also be used to study the effects of herd management, such as feed composition and feeding systems (Chen et al., 2016; van Hoeij et al., 2017; Różańska-Zawieja et al., 2021). As shown by Różańska-Zawieja et al., feeding management affected not only the total milk yield but also the lactation curve dynamics, with total mixed ration-fed cows reaching peak production later and exhibiting a gentler post-peak decline than cows fed partial mixed rations (Różańska-Zawieja et al., 2021).

Furthermore, they are applied to studies on the response of vaccinations and treatments (Fluck et al., 2024; Wilson et al., 2008). As an example, it was shown that J5 vaccination is associated with reduced milk production loss after clinical mastitis. However, this could only be proven to be significant by using a lactation curve model to adjust for significant covariate factors, including a herd effect (Wilson et al., 2008).

3 Modeling the lactation curve

Since no universally optimal model exists (Murphy et al., 2018), the choice of the most appropriate approach depends largely on the intended use case and available data. Key considerations include whether the model is intended to visualize aggregated curves or individual lactation curves, whether its application is retrospective (e.g. for calculating cumulative production) or predictive (e.g. for forecasting future yields), the frequency and quality of the available data, and the level of accuracy required for the task at hand (Macciotta et al., 2011). Once the intended use case is established for the lactation curve, a model can be chosen on which the data can be fitted. This model can be a predefined function, but a methodology whereby the model function is derived from the data is also possible. After this, the data must be fitted to the model, the results need to be validated, and the applied methodology needs to be documented.

Published by Burleigh Dodds Science Publishing Limited, 2026.

3.1 The creation of a model

While more complex models may offer enhanced accuracy, they often come with higher computational demands, an increased risk of overfitting (capturing noise rather than true trends), and a false sense of precision when applied to low-quality milk measurement data (Robinson, 2023). Therefore, when selecting a model, a balance must be made between complexity and accuracy to ensure both practicality and reliability.

Creating a lactation curve can be as simple as connecting the samples, as if connecting the dots, using methods such as the Test Interval Method (Sargent et al., 1968). More sophisticated dot connecting techniques also exist, such as piecewise linear regression, which fits data using connected straight-line segments with breakpoints to capture changes in trend (Lee and Park, 2022). Or smoothing with cubic splines, fitting a piecewise cubic function to test-day measurements, but with a roughness penalty to balance closeness to the data and overall smoothness (Verbyla and Verbyla, 2009). These methods are highly flexible and can often provide a more accurate fit than parametric or orthogonal polynomial models (White et al., 1999). However, their flexibility can lead to overfitting, especially in the presence of noise, which may reduce their predictive power (Macciotta et al., 2010). Additional drawbacks include increased computational demands due to the need to estimate multiple piecewise segments instead of a single global function and the difficulty of interpreting spline coefficients in a biological context (White et al., 1999). Another limitation of smoothing methods is the poor extrapolation beyond the sampling range. Natural cubic splines partially address this by enforcing linearity at the boundaries (Verbyla and Verbyla, 2009), but predictions outside the observed range remain unreliable, making these methodologies unusable for the prediction of future milk yields (Zielesny et al., 2011).

When wanting to use an extrapolatable lactation curve model, e.g., for future prediction, or when a model is desired with fewer parameters compared to cubic splines, but higher accuracy than the Test Interval Method, an alternative approach is to fit the lactation curve to a predefined function. There are many different predefined functions to choose from, each with its own benefits and pitfalls. The first lactation curve function was made in the 1920s by Brody et al., aiming to quantify differences in the decline of milk production across breeds and strains of cows (Brody et al., 1923). Initially, their model was formulated as an exponential function to depict the declining phase of the lactation curve. One year later, this model was improved by incorporating the modeling of the ascending phase (Brody et al., 1924). Although this model had its limitations, including the tendency to underestimate milk yield in mid-lactation and overestimate yield toward the end of lactation (Cobby and Du,

1978), the initial shape of the lactation curve has remained largely unchanged in the past century.

Historically, the model functions behind the lactation curve have been characterized by a high variability in approaches, including empirical vs. mechanistic, parametric vs. non-parametric, and linear vs. non-linear methods. Overall, the preponderant approach to lactation curve modeling has been empirical, parametric, and non-linear. The subsequent segment will clarify the definitions of these terms in the context of lactation curve modeling, followed by a brief chronological overview of different model developments.

3.2 Different types of model functions

Essentially, empirical models rely exclusively on the descriptive aspects of lactation. This is in contrast to mechanistic models, which are based on underlying theoretical assumptions (Macciotta et al., 2011). As an example of this binary distinction, a mechanistic model could base its milk yield prediction on underlying energy compartmentation theories, whereas an empirical model would only look at milk yield measurements over time to predict the shape of the curve, without accounting for underlying biological processes. This distinction is often described as choosing between practical value (empirical) or generating new knowledge (mechanistic) (Nestorov et al., 1999). However, nuance is necessary when models are differentiated within this binary view. Assuming theoretical assumptions are only reserved for mechanistic models would overshadow the complexity of lactation curve modeling (Macciotta et al., 2011). Concretely, even the simplest empirical model contains theoretical assumptions, such as the expectation of the occurrence of a peak yield, ascending phase, and descending phase.

A second distinction that can be made is between parametric and non-parametric models. Parametric models rely on specific assumptions about the underlying distribution of the population being studied. In the case of lactation curves, this would mean that the distribution of milk yields for a certain day is normally distributed. Parametric models generally have more statistical power, require smaller sample sizes, and their results are relatively easy to interpret compared to non-parametric models. But when the assumptions about the data are not met, they can lead to misleading results. Examples of parametric models include linear and logistic regression. Non-parametric models, on the other hand, make fewer assumptions about the distribution of the population from which samples were drawn. This makes them more flexible, robust to outliers, and less prone to errors, but they require more samples to reach the same statistical power. Neural networks, like the lactation curve model proposed by Liseune et al. (2021a), are an example of these types of models.

Published by Burleigh Dodds Science Publishing Limited, 2026.

Lastly, model functions can be divided into being linear or non-linear in their parameters. A linear model in this sense is not the same as a model describing a linear relationship, but it can be written as a linear combination of subfunctions for each parameter. Examples of linear models include straight lines and polynomials. If a parameter is inside a fraction, exponent, logarithm, or product with another parameter, the function is nonlinear in parameters. This is important to consider when looking at fitting options, as will be described below (Zielesny et al., 2011).

3.3 Predefined lactation curve functions from a historical perspective

Table 1 enumerates some of the principal lactation curve functions and their mathematical properties. After the Brody model in 1924, the Sikka model (Sikka, 1950) was built, which provided a better fit for primiparous cows, but performed poorly for multiparous cows (Gahlot et al., 1988). Next, the Fischer model (Fischer, 1958) was introduced, using a linear decline instead of an exponential decline. Fischer was one of the first to experiment with using corrections for environmental effects to improve the fit of the model, although this did not lead to model improvement.

The early lactation curve models had to be relatively simple as they were fitted manually. This started to change around 1960 as computers became more accessible and the first specialized statistical software appeared. One of the pioneers in statistical computing, Nelder, proposed an inverse function to improve the lactation curve, which limited the curve to positive values and fitted more easily to a non-symmetrical shape around the peak of lactation (Nelder, 1966). Wood also started using computers and introduced one of the most popular models still used today. Wood proposed encompassing the entire lactation curve through an incomplete gamma-type function, recognizing the limitations of the previous exponential models in accurately capturing the ascending phase of the lactation (Wood, 1967). The model presented a good fit and had a computationally less demanding linearized fitting option. This classic Wood model emerged as the most widely recognized lactation curve model (Dijkstra et al., 2010) and served as an inspiration for subsequent improvements and innovations (Cobby and Du, 1978; Congleton and Everett, 1980; Dematawewa et al., 2007; Dhanoa, 1981). While the latter models enhanced the Wood model in various ways, their practicality and difficult interpretability seemed to hamper further implementations (Dijkstra et al., 2010).

The addition of a mechanistic layer to the lactation curve models was studied at the end of the 1980s. Starting with the publication of Wilmink (1987a), the correction of predictions for environmental effects was successfully implemented. While these corrections were computationally challenging,

Table 1 A non-inexhaustible list of lactation curve models developed with their formula and main features.

Model	Formula ($Y(t) =$)	Remarks	Type of model
Brody et al. (1923)	ae^{-k1t}	Only depicts the declining phase of milk production.	Empirical
Brody et al. (1924)	$ae^{-k1t} - be^{-k2t}$	Incorporates the modeling of the ascending phase.	Empirical
Sikka (1950)	ae^{bt-ct^2}	The parabolic exponential function is effective for primiparous cows but less suitable for multiparous cows.	Empirical
Fischer (1958)	$a - bt - ae^{-ct}$	Improves the Brody models by combining the exponential decline with a linear decline.	Empirical
Nelder (1966)	$\frac{t}{a + bt + ct^2}$	Prevents unrealistic fit as the inverse function is non-negative, bounded, and has no built-in symmetry.	Empirical
Wood (1967)	$at^b e^{-ct}$	Incomplete gamma function; most popular due to its simplicity, its stability in the presence of missing data, and its computational ease.	Empirical
Dhanoa (1980)	$at^{bc}e^{-ct}$	Less correlations between characteristics; characteristic b can directly estimate the peak lactation date.	Empirical
Ali & Schaeffer, (1987)	$a + bt + c(t/305)^2 + d\log\left(\frac{305}{t}\right) + k\left[\log\left(\frac{305}{t}\right)\right]^2$	Linear regression model of a quadratic polynomial in standardized time ($t/305$), and a quadratic polynomial in the log-transformed inverse time ($\log(305/t)$). The addition of the log enables more flexibility in modelling the ascending phase and peak of the lactation.	Empirical
Wilmink (1987)	$a + be^{-kt} - ct$	A parametric model that combines an exponentially decaying component with a linear decline. The parameter a defines the baseline level of milk yield, the exponential term predominantly describes early lactation dynamics and peak formation, whereas the linear term captures the approximately linear decline in milk yield during later lactation. The decay parameter k is often fixed in advance to make the model easier to fit.	Empirical
Rook (1993)	$a\left(1 - b_0 e^{-b_1 t}\right)\frac{1}{1+ct}$	Consists of a scaler, a growth curve with asymptote 1 and a death curve with asymptote 0. Because of this clear division, the death function can be interpreted as model for persistency.	Empirical

Dijkstra (1997)	$ae^{\left[\frac{b\left(1-e^{-ct}\right)}{c}-dt\right]}$	Describes the proliferation and death of mammary gland cells during lactation to model resulting milk yield.	Mechanistic
Pollott (2000)	$\left\{\frac{N}{1+\frac{1-P_0}{P_0}e^{-Gt}}-\frac{L}{1+\frac{1-Q_0}{Q_0}e^{-Dt}}\right\}\left\{S_a\left[1-\exp\left(-S_b t\right)\right]\right\}$	Also includes cellular secretion rate of mammary tissue. N = total number of parenchyma cells which become active during early lactation, P_0 = proportion of N present at the start of lactation, G = the relative growth rate in active cell numbers (0<G<1), t = day of lactation, L = total number of parenchyma cells dying off, Q_0 = proportion of L dead at start of lactation, D = the relative death rate of cells (0<D<1). S_a = the maximum secretion rate (in kg/cell per day) and S_b (0<S_b) = the relative rate of change in secretion rate.	Mechanistic
Martin and Sauvant (2010)	Model too big to fit in table	Complex mechanical model consisting of multiple sub-models based on the compartmentation and prioritization of energy towards different physiological statuses (growth, maintenance, gestation and lactation).	Mechanistic
MilkBot (Ehrlich, 2010)	$a\left(1-\frac{e^{\frac{c-t}{b}}}{2}\right)e^{-dt}$	Derived from a theoretical-mechanistic hypothesis, characteristics can be interpreted from two perspectives: their impact on the curve's shape and their association with mechanistic hypotheses.	Mix

a viable model was presented. This was possible using pre-fitting data engineering, such as dimension reduction by splitting covariables into classes (e.g. ten age classes) and separating the effect estimation of environmental factors (linear mixed model) from the fitting to the lactation curve (exponential curve). However, it's important to note that the performance of the Wilmink model was compromised when the first test day occurred after 60 days in milk (Silvestre et al., 2006). Consequently, the effectiveness of the Wilmink model in making accurate predictions is significantly influenced by the interval between calving and the first test day. During the development of the first test day models, Stanton et al. (1992) similarly integrated corrections for environmental factors, though these corrections were not applied to a lactation curve but to group/herd averages. Ptak and Schaeffer hypothesized that corrected daily predictions would allow more accurate breeding values, but at the cost of a drastic increase in computational power (Ptak and Schaeffer, 1993). Nevertheless, by the late 1990s, multiple national breeding organizations had adopted a test day model for evaluating milk production traits, many of which included a lactation curve; for example, the Wilmink function was integrated

into the Canadian and the New Zealand Test-Day Models (Schaeffer et al., 2000; Swalve, 2000; Weller et al., 2017). To this day, test-day models remain widely used by breeding organizations from all over the world (Interbull Centre, 2026)

Fully mechanistic models developed by Dijkstra et al. (Dijkstra et al., 1997) and Pollot (Pollott, 2000) approached the modeling of milk yield by exploiting the close relation between milk production and the number of cells in the mammary gland. These models describe the proliferation and death of mammary gland cells during lactation to model milk yield. An alternate approach within a mechanistic framework was formulated by Martin and Sauvant (Martin and Sauvant, 2010), where the compartmentation and prioritization of energy towards different physiological statuses (growth, maintenance, gestation, and lactation), using an array of sub-models, dictated the shape of the lactation curve. Although these mechanistic models showed similar fits to the standard empirical models (Dijkstra et al., 2010; Martin and Sauvant, 2010), the direct link with biological processes would increase the interpretability of the model. However, Pollott (2000) indicated the limitations formed by this kind of model's overparameterization and the difficulties fitting it in a monthly-based recording system.

The point that a binary distinction between empirical and mechanical models cannot be made is proved by Ehrlich (Ehrlich, 2010), who proposed an empirical model derived from a theoretical-mechanistic hypothesis, concerning the development and decay of udder capacity. This led to four parameters, each corresponding to a specific aspect of the lactation curve shape or magnitude, which could be interpreted both in terms of the effect that they have on the curve and in terms of the mechanistic hypothesis. Consequently, the model offered greater flexibility in accounting for the influence of diseases and management practices, facilitating the analysis of their effect (Ehrlich, 2013; Hostens et al., 2012).

Figure 3 illustrates the fitting of different lactation models to the same daily milk yield data of a single lactation. Notably, the most significant differences between the models occur near the peak of lactation. The Fischer model predicts an earlier peak compared to the other models, which exhibit more accurate timing of the peak but tend to underestimate its magnitude. These mathematical models are designed to capture the overall lactation trend; consequently, they often fail to account for short-term disturbances in milk yield, as seen around 150 DIM in the figure.

This variability in model fit is further demonstrated in Fig. 4, where the Wood and MilkBot functions are applied to the same lactation data again using an identical linear fitting method. Clear differences in timing of peak, peak yield, and overall trajectory can be observed. Solely varying the predefined function gave a difference of nearly 200 kg of total lactation milk yield. This demonstrates the sensitivity of lactation curve modeling to the selected predefined function, even when the fitting technique remains consistent.

Published by Burleigh Dodds Science Publishing Limited, 2026.

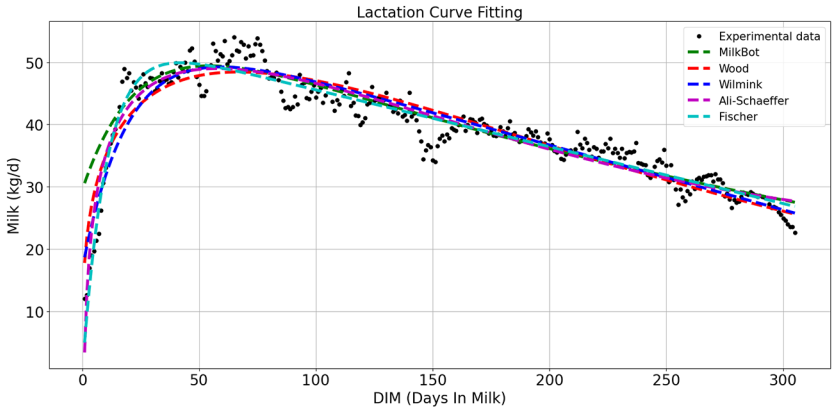

Figure 3 Comparison of five different lactation curve models for the same lactation using the same linear fitting method but different predefined functions (MilkBot, Wood, Wilmink, Ali-Schaffer, and Fischer).

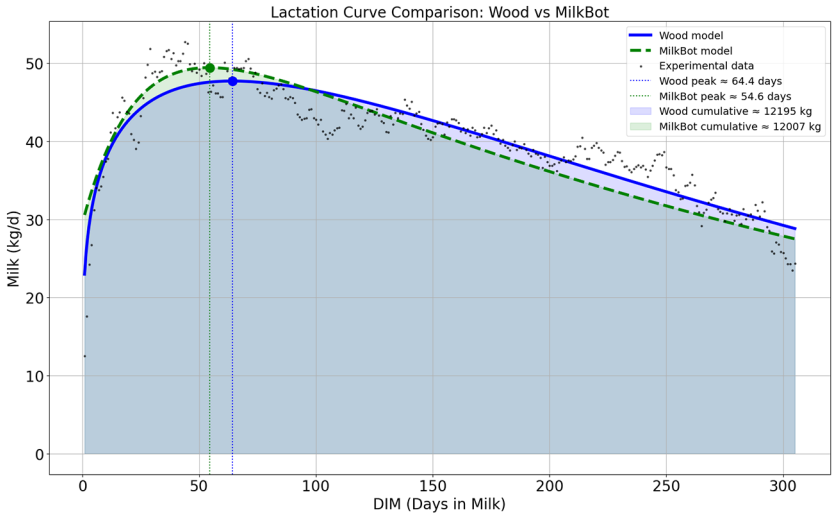

Figure 4 Comparison of two different lactation curves of the same lactation using the same linear fitting method but different predefined functions (Wood versus Milkbot).

3.4 Fitting data to a lactation curve function of a parametric model

After selecting the model function, the next step involves choosing an appropriate method to estimate the values of the corresponding parameters. This process is partially dictated by the chosen model architecture, but

often offers several methodological options, each with inherent strengths and weaknesses (Zielesny et al., 2011). For mechanistic models, this entails estimating the underlying parameters; for instance, Dijkstra et al. (Dijkstra et al., 1997) initially measured the total DNA contents of the mammary gland during pregnancy and lactation to use as a parameter in their model. In contrast, empirical models allow data to be fitted to the curve directly. The specific fitting method available depends on the mathematical model function, but the goal remains consistent: minimizing an evaluation metric, such as the sum of squared residuals. Functions that are linear in their parameters can be optimized algebraically using derivatives, as they possess a single global minimum. However, numeric optimization approaches are necessary when dealing with functions that are non-linear in their parameters, have correlated parameters, involve unevenly spaced test day data, or use very large datasets (Hastie et al., 2009).

Some non-linear model functions can be transformed into a linear variant to make them easier to solve, for example, by taking the natural logarithm of the parameters of the Wood curve (Wood, 1967). However, this also changes the problem being solved; from optimizing the squared difference in actual milk yield to the squared difference in log-yields, which disproportionately weights smaller values and thus biases the fit towards lower-yielding periods. Additionally, the transformation alters the underlying statistical assumptions: it changes the error structure from additive to multiplicative, meaning the variance of the residuals is no longer constant on the original scale. This induces heteroscedasticity, where prediction errors increase with yield, which may violate assumptions of constant variance and affect the reliability of parameter estimates or predictions.

Parameter fixation is another approach to create an algebraically solvable model, a method commonly applied to the Wilmink function, whereby the exponent is estimated in a preliminary analysis as the best-fitting value for the data (Schaeffer et al., 2000; Silvestre et al., 2006; Wilmink, 1987b). However, it must be noted that fixing parameters could lead to reduced model fit as even the educated guess of the fixed parameter is likely to be suboptimal (Olori et al., 1999).

As an alternative to converting a model to a linear variant, iterative numerical methods such as Marquardt, Newton, Gauss-Newton, and Does Not Use Derivatives (DUD) can be used for the fitting of lactation curves (Bouallegue and M'Hamdi, 2019). These methods work by progressively refining initial parameter estimates to minimize the difference between observed and predicted values, continuing until a satisfactory fit is achieved. The solution that is found using numerical methods is therefore an approximation, in contrast to linear methods, where there is only one result possible. A benefit of these techniques for model fitting is, next to their broad applicability, that

Published by Burleigh Dodds Science Publishing Limited, 2026.

they allow the incorporation of prior knowledge through starting values or constraints, which can guide the fitting process toward biologically plausible solutions, especially when data is sparse or incomplete (Radjabalizadeh et al., 2022). This advantage is especially strong in Bayesian frameworks, where prior distributions are formally specified. Methods such as Gibbs sampling and maximum a posteriori estimation can estimate model parameters even in the absence of data; in such cases, the posterior distribution reflects the prior, resulting in the return of population-mean parameter values (Ehrlich, 2013; Rekaya et al., 2000). Numerical methods also have limitations. They can be more error-prone as the predefined maximum number of iterations can be reached before the optimum is found or, in the case of non-linear functions, the iteration can get trapped in a local minimum instead of converging to the global minimum (Zielesny et al., 2011). A careful estimation of the start values of the model parameters is important to mitigate this risk. Models with biologically interpretable parameters, such as peak milk yield or time to peak yield, have an advantage here, as their values typically fall within defined intervals that can be derived from descriptive statistics of the dataset. The fact that fitting strategy influences model fit is illustrated in Fig. 3, whereby a lactation is fitted on the MilkBot model using both linear and Bayesian fitting. This fitting resulted in different model trajectories, especially around the peak (Fig. 5)

3.5 Lactation curve modelling in other ruminants

Although most research about lactation curve models has focused on cows, these models can, in principle, be applied to all lactating animals, although the scale differs per species. For example, lactation curves of goats, sheep, and water buffalo generally resemble those of cows, characterized by a rapid rise to peak production followed by a gradual decline (Dijkstra et al., 2010; Elvira et al., 2013; Omar et al., 2024). However, marked species-specific differences also exist. In goats, extended lactation strategies are commonly applied, leading to a prolonged plateau after the post-peak decline, and some lactation curves show a mid-lactation bend linked to photoperiod length (Arnal et al., 2018; Salama et al., 2005).

Many of the models described above are therefore transferable across species, but their application requires careful evaluation of underlying assumptions and appropriate adaptation to the target species. Such adaptations may include specifying species-appropriate priors in Bayesian frameworks, recalibrating parameters in mechanistic models, or scaling with an appropriate lactation length in the Ali and Schaeffer model. Consequently, it is recommended that multiple candidate models be tested and their goodness of fit assessed before selecting a model for application in a new species.

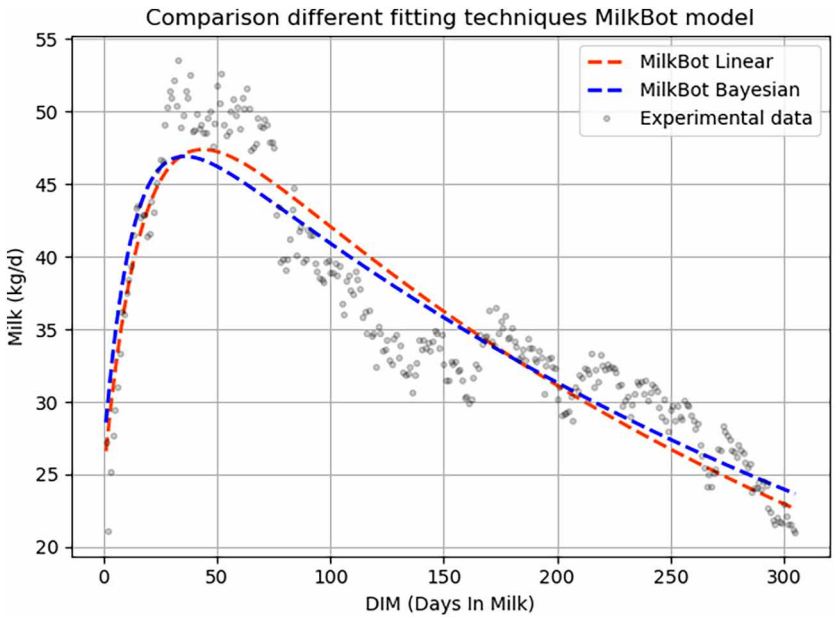

Figure 5 Comparison of two different fitting strategies for the same lactation and the same model function (MilkBot model).

3.6 Optimizing and evaluating the lactation curve

There are many metrics that can be used to guide model fitting or to judge whether a fit is successful. The coefficient of determination (R^2), the Pearson correlation coefficient (PCC), the mean absolute error (MAE), the mean absolute percentage error (MAPE), and the root mean square error (RMSE) are commonly applied to assess lactation curves (Liseune et al., 2020). These metrics can be applied to assess the prediction of an individual test day or the complete lactation through the cumulative 305 milk yield, whereby the MAE serves as a good metric to evaluate accuracy, while RMSE can be used as a measure for precision (Cole et al., 2012). While comparing lactation curve models with multiple metrics, Liseune et al. (2020) showed that models with the exact same R^2 and MAPE can have different RMSE, emphasizing that the selection of the best performing model can be influenced by the metric chosen. Furthermore, Ehrlich (2013) argued that while strict minimization of MSE may achieve lower error values, Bayesian methods often better capture real-world phenomena, regardless of a higher MSE, especially in cases of sparse data. This highlights the importance of considering multiple metrics and statistical frameworks, as different approaches may prioritize aspects of fit that are more biologically or practically meaningful.

Optimizing based upon the widely used RMSE is the best choice when the differences between the milk yield predicted by the model and the measured test day milk yield are independent, evenly and normally distributed, and have a constant variance. If these assumptions are violated, for instance, because the lactation curve function was transformed, absolute error metrics like MAE or scale-independent metrics like MAPE may provide more informative evaluations of fit. Another valuable approach to assess model fit is to examine the absolute residuals, as a well-fitting function will produce residuals that are comparable in magnitude to the measurement errors. When comparing multiple models, evaluation should extend beyond absolute goodness-of-fit to account for differences in model complexity. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) can be used to penalize the addition of extra parameters, thereby helping to distinguish genuine performance improvements from overfitting.

3.7 Extracting the lactation curve characteristics

Once a lactation curve is fitted, key lactation curve characteristics (LCCs) can be extracted that quantify the visual shape of the lactation curve, including time to peak, peak yield, cumulative milk yield (e.g. at 305 DIM), and persistence (Atashi et al., 2020). Calculating these characteristics helps translate the mathematical representation into tangible and interpretable metrics that can be used for hypothesis testing and management applications. For example, the effects of a dietary intervention are more readily understood in terms of changes in persistence or peak yield than through changes in model parameters.

Figure 6 graphically illustrates the Lactation Curve Characteristics (LCCs) on a typical lactation curve.

These characteristics can be derived either analytically or numerically from the fitted model's parameter estimates. Analytical derivations are especially useful when a simple formula can be obtained, as is the case for the Wood model, where, for instance, time to peak is computed as the ratio of parameters b and c. However, for more complex models, such as the (Prasad, 2003) model, the formulas for LCCs are lengthy and impractical. In such cases, numerical methods to approximate the LCCs are typically simpler and more effective. Table 2 presents the LCCs analytically derived from the most applied lactation models.

Time to peak represents the number of days from calving to maximum milk production. It is identified as the days in milk where the derivative of the lactation curve function equals zero. Peak yield corresponds to the y-value of the lactation curve at this point, representing the highest milk yield during the lactation period. Cumulative yield refers to the total milk production over a

specific period, typically 305 days, and can be calculated by integrating the lactation function over the desired range of days in milk.

Persistency reflects the ability to maintain the level of production throughout the lactation and has several mathematical interpretations in the literature (Gengler, 1996; Madsen, 1975). These can broadly be classified into four categories: measures derived from lactation curve models, measures based on variation in test-day yields, measures obtained through the extraction of latent variables, and ratio-based measures comparing yields at different stages of lactation (Arnal et al., 2018; Gengler, 1996). This diversity of definitions reflects both different conceptual interpretations of persistency and different practical objectives, but it also complicates comparisons across studies and management systems. For this reason, persistence has been excluded from Table 2.

A common intuitive interpretation of persistency is the steepness of the decline in milk yield after peak production. The simplest way to quantify this slope is as the post-peak average rate of decline; however, this implicitly assumes a linear decay. Because most lactation curve models do not exhibit a linear decline, such estimates depend on the assumed lactation length or on the specific interval in days in milk over which the slope is calculated. One way to reduce this dependency is to evaluate the slope of the model curve at a fixed time point during the declining phase (e.g. 100 days in milk) by calculating the derivative of the function at that point. However, this measure remains confounded with the absolute production level and, to a lesser extent, with the shape of the early lactation curve. As a result, selection for increased persistency may inadvertently favor lower-producing animals, even when the objective is to maintain or extend production at a similar yield level. In contrast,

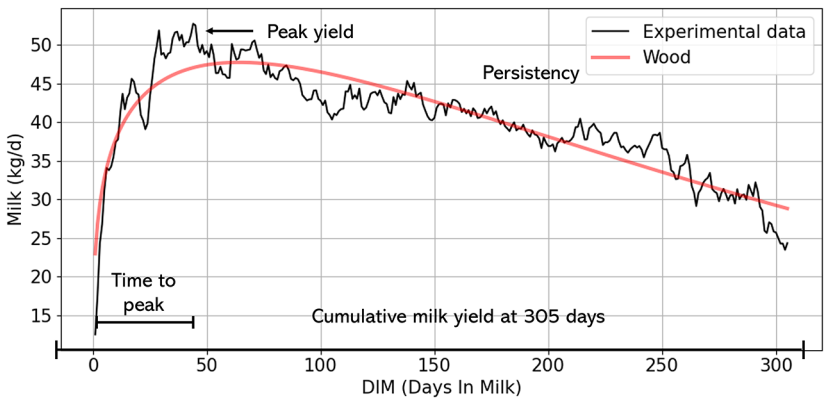

Figure 6 Graphical representation of lactation curve characteristics on a typical lactation curve.

Table 2 The formulas used to extract lactation curve characteristics from mathematical models. Lengthy expressions are available through the associated lactation curve package (van Leerdam et al., 2026)

Model	Formula ($Y(t) =$)	**Time to peak (days)**	**Peak yield (kg/d)**	**Cumulative milk yield (kg/305d)**
Brody et al. (1923)	ae^{-kt}	Not applicable	Not applicable	$\frac{a}{k} * \left(e^{-k} - e^{-305k}\right)$
Sikka (1950)	ae^{bt-ct^2}	$\frac{b}{2c}$	$ae^{\left(b^2/4c\right)}$	package
Fischer (1958)	$a - bt - ae^{-ct}$	$\frac{1}{c}\ln\left(\frac{ac}{b}\right)$	$a + b * \frac{ln\left(\frac{b}{ac}\right)}{c} - \frac{b}{c}$	$305a - \frac{a}{c} + \frac{ae^{-305c}}{c} - \frac{93025b}{2}$
Nelder (1966)	$\frac{t}{a + bt + ct^2}$	$\sqrt{\frac{a}{c}}$	$\frac{\sqrt{a}}{\sqrt{ab} + 2a\sqrt{c}}$	package
Wood (1967)	$at^b e^{-ct}$	$\frac{b}{c}$	$a\left(\frac{b}{c}\right)^b e^{-b}$	package
Dhanoa (1980)	$at^{bc}e^{-ct}$	b	$a\left(be^{-1}\right)^{-bc}$	package
Emmans and Fisher (1986)	$ae^{\left(-e^{d-bt}\right)}e^{-ct}$	$\frac{d - \ln\left(\frac{c}{b}\right)}{b}$	$a * e^{\left(-\frac{c}{b}\left(1+d-\ln\left(\frac{c}{b}\right)\right)\right)}$	package
Ali and Schaeffer (1987)	$t1 = \frac{t}{305}$	package $a + bt1 + ct1^2 + dw1 + kw1^2$	package	package
Wilmink (1987)	$a + be^{-kt} - ct$	$\ln((\frac{ck}{b})^{\frac{1}{k}})$ $w1 = log\left(\frac{305}{t}\right)$	$a + b\frac{\ln(\frac{ck}{b})}{k} + \frac{b}{k}$	$305a + \frac{93025b}{2} + \frac{c}{k} - \frac{c * e^{-305k}}{k}$
Hayashi (1993)	$b\left(e^{-t/c} - e^{-t/ac}\right)$	$\frac{ac\ln\left(a\right)}{a-1}$	$b\left(e^{-\frac{aln(a)}{(a-1)}} - e^{-\frac{aln(a)}{(a-1)}}\right)$	package
Rook (1993)	$a\left(\frac{1}{1+\frac{b}{c+t}}\right)e^{-dt}$	$\frac{\sqrt{b} * \sqrt{bd+4}}{2\sqrt{d}} - \frac{b}{2} - c$	package	package
Dijkstra (1997)	$ae^{\left[\frac{b\left(1-e^{-ct}\right)}{c} - dt\right]}$	$c^{-1}\ln\left(\frac{b}{d}\right)$	$a\left(\frac{b}{d}\right)^{d/c} e^{\frac{(b-d)}{c}}$	package
Prasad (2003)	$a + bt + ct^2 + \frac{d}{t}$	package	package	package
MilkBot (2011)	$a\left(1 - \frac{e^{\frac{c-t}{b}}}{2}\right)e^{-dt}$	$c - b\ln\left(\frac{2bd}{1+bd}\right)$	$a\left(2bd\right)^{bd}\left(bd+1\right)^{-bd-1}$ e^{-cd}	package

scale-independent measures allow persistency to be interpreted as a distinct trait, facilitating more meaningful genetic evaluation and selection decisions.

To address these limitations, alternative definitions of persistency have been proposed that aim to reduce or eliminate dependence on overall milk yield. For example, persistency can be expressed as a ratio of milk yield at peak relative to yield at the end of lactation (Gengler, 1996). Another, more statistical, approach defines persistence as a weighted sum of test-day deviations across lactation, where the weights depend on days in milk and are constructed to be orthogonal to total yield (VanRaden, 1998). Similarly, persistence can be extracted as a latent variable from principal component analysis or multivariate analysis of the correlation matrix of milk yield data (Macciotta et al., 2006). These approaches explicitly recognize persistency as a characteristic distinct from scale, but differ substantially in interpretation and implementation.

Building on this idea, persistence can be defined more generally as a relative, rather than an absolute, rate of decline. This relative rate can be obtained by taking the derivative of the logarithm of the lactation-curve function, which yields a proportional (percentage) rate of change over time. Because this measure expresses decline relative to the magnitude of the function itself, it is inherently scale-independent and therefore less confounded with overall production level than absolute slope-based measures, while at the same time being more intuitive as it can be easily related back to the shape of the lactation curve. However, as lactation curve models are non-linear, this relative rate is dependent on the DIM at which it is evaluated.

A more direct interpretation emerges in lactation curve models that explicitly separate growth and decline components. In the MilkBot model, for example, the multiplicative growth term approaches unity as lactation progresses, such that late-lactation is dominated by the exponential decay component. Under these conditions, persistency can be approximated directly, independent of DIM, by the decay parameter $-d$, or equivalently expressed in terms of the half-life of the milk production (Ehrlich, 2013). Likewise, in the Rook model, milk yield is represented as the product of a growth function and a decay function, with the latter providing a direct and interpretable measure of persistency (Rook et al., 1993).

Regardless of the chosen methodology, one limitation of lactation curve modeling remains: the inevitable lag between predicted and actual milk yields. This model inaccuracy, influenced by the data quality and quantity of the input variables, model architecture, and fitting methods, should be carefully considered when interpreting model outputs and their applications.

Published by Burleigh Dodds Science Publishing Limited, 2026.

4 The use of machine learning for lactation curve modeling

The mathematical models described above are widely used and accepted by the technical and scientific community. However, the continuous search to improve model accuracy and the pursuit of greater flexibility, particularly in addressing individual lactations deviating from the average pattern, has driven more recent studies away from traditional parametric models (Macciotta et al., 2011). Machine learning (ML) offers a fundamentally different approach to modeling, characterized by its ability to automatically improve through iterative experience and learning, rather than relying on predefined mathematical functions (Jordan and Mitchell, 2015). ML models are typically nonparametric and nonlinear in their parameters, making them distinct from traditional approaches. ML is an umbrella term for a multitude of models. To create lactation curves, supervised models are mostly used, where the model is created with a labeled data set.

4.1 Benefits of machine learning approaches

Using ML to model a lactation curve has some clear benefits. First, as non-parametric models, ML eliminates the need to verify assumptions such as normality and autocorrelation in the input data. Additionally, ML models can integrate multi-source input data (e.g. milk yield, milk ejection dynamics, health status, and environmental factors), allowing for the construction of a unified model that accommodates diverse scenarios for different cow subgroups (Grzesiak et al., 2006; Liseune et al., 2020). Moreover, as the amount of data collected on farms increases, including individual cow monitoring (Cabrera and Fadul-Pacheco, 2021; Monteiro et al., 2021), opportunities are opened up to use this data to improve the accuracy of milk yield predictions, especially at the individual level. With increasing interest in individual dairy cow management (Maltz, 2020), this is a progressively important aspect of lactation curve modeling.

Predefined traditional functions have an inherent shape, a convex curve increasing at the beginning, followed by a longer and gradual decline. However, not all lactations from dairy ruminants follow this classic pattern and instead may have an alternative shape (Macciotta et al., 2005; Rekik and Gara, 2004). For example, cows calving in autumn in a pasture-based system can have a bimodal lactation curve shape due to a boost in milk production when the quality of the grass increases during spring (García and Holmes, 2001). Because ML is a data-driven approach, these models can overcome this limitation by flexibly adapting to different lactation curve shapes. Their fitting is not constrained by predefined theoretical assumptions and can be continuously updated as new

data becomes available. Furthermore, Slob et al., in their systematic literature review, have shown that ML models often outperform traditional parametric methods in terms of model performance in the dairy science domain (Slob et al., 2021).

4.2 Applied machine learning techniques to lactation curve modelling

Although machine learning encompasses a wide range of model classes, only a subset has been explored for lactation curve modeling to date (Slob et al., 2021). Existing applications largely focus on neural network–based approaches. These range from early models with only a handful of parameters to more specialized architectures designed to capture temporal structure and latent representations in milk production data. This section reviews these applied ML techniques, highlighting how increasing computational capacity and data availability have enabled more complex model formulations, while also shaping the current scope and limitations of ML-based lactation curve modeling.

The Artificial Neural Network (ANN) is among the most widely studied ML approaches applied in lactation curve modelling (Murphy et al., 2018). ANN is a class of supervised ML models that learn a functional relationship between a set of input variables and one or more output variables directly from data. Conceptually, ANNs are inspired by biological neural systems, but in practice, they are purely mathematical models composed of layers of interconnected computational units ("neurons"). Each neuron applies a weighted linear combination to its inputs, followed by a nonlinear activation function, allowing the network to represent complex, nonlinear relationships. During training, the network parameters (weights and biases) are iteratively adjusted to minimize the difference between predicted and observed outputs using optimization algorithms such as backpropagation. These input features can represent all kinds of data, such as previous milk yield, along with animal metadata (e.g. lactation number, disease records, or a genic profile).

Early work by Lacroix et al. (1995) demonstrated the potential of multilayer perceptrons (MLPs), a simple type of ANN, in forecasting total 305-day milk yield with similar accuracy in respect to traditional mathematical models. Their study laid the foundation for ANN use in lactation curve modeling, showcasing how multi-source input variables, such as management and physiological factors, improved predictive performance, particularly during early lactation. Building on this early work, later research has concentrated on methodological refinements, including data preprocessing techniques, the optimization of ANN architectures, and training procedures, to increase predictive accuracy. (Lacroix et al., 1997). For example, Sharma et al. (2007) systematically evaluated

variations in the backpropagation algorithm, activation function, hidden layer structure, and learning rate in MLPs to predict first lactation total 305-day milk yield. Among the tested configurations, the Levenberg-Marquardt algorithm with Bayesian regularization showed the highest accuracy (92%), but only marginally exceeded the performance of a conventional multiple linear regression model (91%).

Several other studies have applied MLPs to predict cumulative milk production traits with comparable objectives. Salehi et al. (1998) proposed a two-stage approach in which cows were first classified into high and low producing groups, followed by specialized neural networks to predict 305-day milk, fat, and protein yields. Njubi et al. (2010) and Dongre et al. (2012) specifically focused on predicting first-lactation cumulative milk yield, while Gorgulu (2012) targeted only Brown Swiss cattle. Beyond dairy cattle, MLPs have also been applied successfully to model cumulative lactation performance in small ruminants, including sheep, as demonstrated by Ince & Sofu (2013) and Kominakis et al. (2002). And lastly, in direct comparisons with traditional parametric models, Khazaei et al. (2008) found that ANNs achieved superior predictive accuracy relative to the summation of the Wood lactation curve. However, not all attempts were successful. For instance, Grzesiak et al. (2003) reported no statistical difference between the predictions of ANN and a multiple regression model, likely due to a relatively small training set compared to the complexity of their model. And while trying 11 different machine learning algorithms, Fieguth et al. reported issues with model overfitting, making the models unusable for practice (Knapp et al., 2025).

Predicting the 305-day milk yield, however, is not equivalent to modeling the full lactation curve. As emphasized earlier, a lactation curve contains substantially more information than cumulative production alone, including features such as peak yield, time to peak, and the pattern of decline.

One approach to bridge this gap is to use artificial neural networks (ANNs) to predict the parameters of a traditional lactation curve function, such as the Wood model. This strategy, applied by Torres et al. (2005), allows the lactation curve shape to be derived indirectly from a small number of estimated parameters. Compared with predicting an entire daily trajectory, this approach is computationally simpler, as only three parameters need to be estimated rather than a full set of daily yields. Alternatively, ANNs can be trained to predict specific lactation curve characteristics directly, such as peak yield or time to peak, rather than cumulative yield alone (Guevara et al., 2023). A more computationally intensive strategy is to predict daily milk yield and subsequently construct the lactation curve by connecting the predicted values over time. As this approach is not constrained by a predefined model function, it can accommodate a wider range of lactation curve shapes and is better suited to capturing short-term deviations in milk production, such as perturbations associated with mastitis.

One of the first studies to predict daily milk yield was conducted by Grzesiak et al. (2006), followed by numerous studies employing different model architectures, training strategies, and input features. For example, Vergani et al. (2024) developed a relatively simple multilayer perceptron with a single hidden layer and demonstrated that incorporating genomic breeding values improved forecasting accuracy of individual cow daily milk production. Similarly, Nguyen et al. (2020) predicted subsequent daily milk yields by integrating previously observed milk production and dietary parameters using three machine learning models, a multilayer perceptron, a support vector machine, and a random forest, while also considering computational efficiency in model selection.

MLPs treat observations independently, making them less suited for sequential or time-series data. To address this limitation, dynamic architectures like Nonlinear Auto-Regressive networks with Exogenous Inputs (NARX) have been proposed, as demonstrated by Murphy et al. (2014). NARX networks integrate past inputs and outputs, enabling them to better capture temporal dependencies and therefore should achieve superior prediction performance compared to both static ANNs and regression-based methods. Despite these advantages, Grzesiak et al. (2021) found no statistical difference in performance when comparing an NARX model to the Wood model and a Seasonal Auto-Regressive Integrated Moving Average (SARIMA) model for predicting average milk yield at successive lactation stages. Building on the limitations of previous time-dependent models, W. Zhang et al. (2020) explored Long Short-Term Memory (LSTM) networks, another more advanced architecture designed to analyze long-term dependencies in sequential data. The researchers optimized their LSTM model using a genetic algorithm and demonstrated improvements in predictive accuracy. However, their study compared model performance only between the optimized LSTM and a standard LSTM without genetic algorithm optimization, leaving its superiority over other models, such as NARX or regression-based approaches, unclear.

Autoencoders (AE) are an alternative ANN architecture that accounts for temporal dependencies in sequential data. Unlike traditional MLPs or NARX models, autoencoders compress lactation data into low-dimensional representations and reconstruct a lactation curve from this compressed space. This capability makes autoencoders effective at imputing missing data at variable time intervals, as demonstrated by Liseune et al. (2020), who showed that sequential autoencoders outperformed MLPs across various patterns of missing data imputation and handled complex, incomplete datasets with greater accuracy.

An example of unsupervised machine learning for lactation curve modeling is provided by Zhang et al. (2022), who used clustering to group milk yield records into ten different groups based on their similarity. A random forest model classified these clusters for new data based on within-herd relationships.

Published by Burleigh Dodds Science Publishing Limited, 2026.

A support vector machine model translated the cluster assignments back into milk yields to construct a lactation curve before a single milk yield was measured. This method outperformed traditional approaches for pre-lactation yield estimation, based on the average milk yield of paternal siblings.

There is an important distinction between ML models that predict lactation curves without relying on milk recordings as input features and those that use milk recordings to predict missing or future yields. ML models that do not utilize milk recordings as input features are often compared with traditional statistical models, such as mixed linear models. Models that do use milk yield measurements should ideally be benchmarked against traditional parametric functions, such as the widely used Wood model (1967), as they are still the standard in the lactation curve domain. However, direct comparisons are sometimes overlooked in existing research. For example, Liseune et al. (2020) trained their ML models using extensive milk yield datasets but did not explicitly compare their model performance to parametric approaches. As a result, while such studies demonstrate the flexibility and potential accuracy of ML methods, they cannot definitively establish ML approaches as superior to traditional models for lactation curve analysis.

4.3 Limitations of machine learning approaches

Although, as illustrated above, many ML models have been proposed, adoption in the field has been limited. This fits in a broader trend in the dairy industry where most published models end up in the so-called "technology graveyard" (Sharma et al., 2026). Several factors may contribute to this, and understanding these pitfalls may help mitigate them in future applications.

Firstly, ML models typically require extensive computational resources and large datasets to ensure reliable parameter estimation. For instance, the original Wood model (1967) was fitted for every single lactation using one test day milk sample per week and was validated on data from 524 heifers, while the ML model of Liseune et al. (2020) was trained and tested using 12,790,342 milk yield recordings from 59,122 lactations and took much longer to fit. Also, more complex traditional methodologies, like the Wilmink model (1987a), used almost 60 times fewer milk yield recordings to be developed and tested. This highlights the potential barrier of data availability and ease of model use for ML adoption.

Secondly, data quality remains a critical limiting factor. Despite advances in automated on-farm data collection, lactation data are still affected by measurement noise, redundancy, and missing values. As a result, when the signal-to-noise ratio is low, even well-designed ML models may have limited predictive performance, reflecting the continued relevance of the "garbage in, garbage out" principle (Carbone, 2022).

Thirdly, most ML approaches are purely data-driven and nonmechanistic, which can compromise their ability to extrapolate beyond the range of the training data. Consequently, ML-based predictions are generally reliable only for conditions that closely resemble those observed during model training, including lactation length and production level, because they are not based upon biological processes that are expected to hold across animals and environments (Carbone, 2022). Similarly, unlike traditional parametric models, parameters estimated by most ML models typically lack a direct biological interpretation, limiting their usefulness for understanding underlying physiological mechanisms.

Lastly, some of the ML models, like neural networks, are generally perceived as "black boxes," making them challenging to interpret and explain (Hassija et al., 2024). This lack of transparency is a notable drawback, as clear reasoning behind model outputs is critical for acceptance and trust among both farmers and researchers (Pietersma et al., 2002). In contrast, other ML approaches, such as decision trees, are generally considered more intuitive and easier to interpret (Pietersma et al., 2003), and although the decision-making processes of the 'black box' models themselves are considered hard to explain, there are techniques to decipher this, such as feature importance and model extraction (Hassija et al., 2024; Rudin et al., 2022).

Although ML has shown promising results in a multitude of fields (Jordan and Mitchell, 2015), evidence from the general livestock domain is more mixed. A review by Nayeri et al. (2019) found that traditional approaches, such as linear regression, can sometimes outperform ML methods. This variability highlights that ML performance is highly problem-dependent and influenced by factors such as problem formulation and data quality.

5 Collecting the data for the lactation curve

In practice, data used to fit models for lactation curve estimation are collected through two main approaches. Traditionally, milk recording organizations coordinate monthly milk sampling (test-day recording) by either sending a trained technician to the farm to record milk volumes, collect milk samples, and deliver them to a laboratory for analysis, or by allowing farmers to record milk production themselves and submit samples for testing. In this system, milk recording organizations play a key role in ensuring data quality and consistency.

More recently, the widespread adoption of milk meters in both milking parlors and automatic milking systems, along with advancements in real-time milk component analysis, has enabled the collection of daily individual cow-level lactation data without reliance on milk recording organizations (Kamphuis et al., 2015). By making the milk recording process as automated as possible, there is less room for human error. Automation does not, however, inherently

guarantee data quality, as demonstrated by Carlström et al. (2013), who attempted to use data from automatic milking systems for genetic analysis. Their experience revealed significant limitations: nearly half of the data and approximately a quarter of the cows had to be excluded due to incomplete or inconsistent information. Moreover, increased reliance on automated systems can make it more difficult to intuitively assess or detect potential biases, as the process becomes increasingly hands-off.

Furthermore, although large volumes of milk production data are collected, they are not always readily accessible for further analysis. Questions regarding the ownership and control of data generated by milk meters and automatic milking systems remain unresolved, which can hinder data extraction for applications such as fitting lactation curve models. In Europe, Regulation (EU) 2023/2854 establishes harmonized rules on access to and use of data, strengthening the rights of users, including farmers, to access and share data generated by connected devices (European Parliament and Council of the European Union, 2023). Similar initiatives have been proposed globally to improve data governance in favor of enhanced data usage (Wolfert et al., 2017). Nevertheless, in practice, limited data accessibility continues to represent a significant barrier to technology adoption (Sharma et al., 2026).

Regardless of the collection approach, the performance of lactation curve models is highly dependent on the quality and quantity of the underlying data, like any other statistical method. Insufficient and/or poor-quality data can compromise model fit, leading to biased parameter estimates and insufficient predictive performance, and therefore lead to suboptimal lactation curve use for decision-making. The significance of data quality issues varies depending on the modeling approach taken. When an individual lactation curve is fitted to a predefined function, such as the Wood model, a faulty measurement will affect only the specific curve for that animal. However, systemic measurement errors have more impact, as they can distort every curve-fitting attempt across the dataset. The implications of inaccurate data are even more pronounced in machine learning models. Training an ML model on incorrect data introduces systemic bias, affecting not only the model's predictability but also its future calculations with new data.

5.1 Causes of data quality loss

Data quantity, influenced by the frequency and distribution of sampling, is widely recognized as an important factor affecting model fit (Berry et al., 2005; Silvestre et al., 2006). However, variability in lactation curve modeling may also arise from differences in the accuracy and precision of milk yield measurements. Data quality can be affected by several factors, including the performance of measuring devices, the calibration of equipment, and the experience of the

operator conducting the measurement (Andreen et al., 2020; Mein et al., 2001). Additional sources of error include incorrect animal identification and sample mislabeling, which may lead to incorrect assignment of milk yield records. Missing values may also occur due to human error, missing animals during recording, or equipment malfunction.

Data manipulation represents another potential source of variability, particularly when 24-hour milk yield is estimated from a subset of daily milkings (Čandek-Potokar et al., 2006; Wu et al., 2023a). Several methods exist for estimating daily milk yield, but their accuracy depends on factors such as milking frequency, irregular milking intervals, milk secretion rates, parity, and lactation stage (Gantner et al., 2009; Liu et al., 2000; Migose et al., 2020; Wu et al., 2022; Wu et al., 2023b). The estimation of milk fat percentage is particularly sensitive to these factors (Duplessis et al., 2019; Gilbert et al., 1973). In a conventional milking parlor, permanent milk meters can improve the accuracy of 24-hour yield estimates by averaging production over multiple days (Miglior et al., 2007), although such improvements are less evident for milk composition estimates. In an automatic milking system, irregular milking intervals are common because cows determine their own milking frequency. To account for these irregular intervals, daily milk yield is often estimated by calculating milk yield per unit of time (e.g. per hour) over a preceding interval and extrapolating this value to 24 h (Hand et al., 2006; Hanuš et al., 2014; Hogeveen et al., 2001; Lazenby et al., 2003). However, the length of the interval used for extrapolation differs among milk recording organizations and may influence the accuracy of yield estimation (ICAR Wiki contributors, 2025b). As a result, challenges remain in obtaining accurate daily milk yield estimates and milk component measurements, particularly in systems characterized by irregular milking patterns. Because these sources of data quality loss can introduce systematic or random errors into the recorded milk yield data, explicit assessment of data reliability is a critical step prior to lactation curve analysis.

5.2 The assessment of data quality

There are limited options to validate the reliability of the milk yield and composition. Validation strategies fall into three broad categories: clerical, rule-based, and probabilistic/statistical approaches. The clerical method relies on user expertise to assess whether datasets meet predefined criteria for their intended use (Winkler, 2004). The assessment is typically expressed as an overall percentage or classified into categories like poor, good, or excellent data quality. The clerical review method is not a reliable approach for the routine evaluation and assessment of data quality due to its inherent subjectivity. The reliance on human judgment introduces variability and inconsistency, leading

to potential errors in identifying inaccuracies and anomalies. Moreover, the outcome of a clerical review is an estimation rather than a precise measurement of data quality, and the manual evaluation of large datasets, particularly those used for research purposes, is neither feasible nor scalable.

Rule-based approaches are more objective, applying deterministic checks against predefined rules (Liu et al., 2019; Wang et al., 2020). The outcome of rule-based testing is binary; the data either passes or fails according to these rules. Examples of basic conditional checks include detecting physically impossible values, such as negative milk yield or milk yield recorded for a male animal. The number and complexity of applicable validation rules depend on the availability of contextual information. Enhancing data collection by integrating secondary data sources, such as dairy herd management software, can provide this extra context (Caraviello et al., 2006; Liseune et al., 2021b; Shahinfar et al., 2014). Similarly, contextual information provided by milk meters, such as milking interval, milking speed, and refusals, could be applied to rule-based testing. Nonetheless, leveraging secondary data introduces its own challenges. These sources were not originally designed for lactation curve development, and there is limited control over farm-level data collection practices and associated protocols, complicating oversight and standardization in milk recording.

While measurements like milk yield are often expected to fall within a minimum and maximum range, defining strict upper and lower thresholds is not straightforward. In the case of milk yield, several biological and contextual factors make it difficult to establish universally valid cut-off values. For example, the recording of an amount of 100 L for an individual cow given current genetics falls into a gray area. The largest part of the cow population will produce far less milk than 100 L per day. However, there are reports of milk recordings that exceed 100 L, leaving it very difficult to determine a maximum cut-off value to create an unequivocal test result. In addition, milk yield is not constant throughout lactation but follows a dynamic curve that changes with each day in milk. Consequently, even if it were possible to define a biologically plausible maximum or minimum yield, such thresholds would need to be adjusted relative to the stage of lactation. This is further complicated by differences in milk production levels across parities.

The third method is a probabilistic or statistical approach (Afriliana et al., 2021; Batini et al., 2009). This approach validates data relative to the overall dataset, with outliers detected based on deviations from expected patterns. The number of outliers is an indicator of low data quality, which can be removed to improve the data. Outliers can be detected using, for instance, the mean ± three times the standard deviation when the recording is normally distributed (Zhang et al., 2021) or by calculating the Mahalanobis distance against a chi-square threshold (Moawed et al., 2024). The granularity of the overall dataset

hereby matters as expected yield is known to vary between parities, days in milk, and herds.

Another statistical approach involves creating machine learning models to identify faulty measurements. The Local Outlier Factor algorithm is an example of such a technique. It calculates the local density deviation of a data point relative to its neighbors, with points exhibiting substantially lower density considered outliers as they do not adhere to the characteristics of the rest of the data (Satoła and Satoła, 2024). These methods are particularly useful for detecting anomalies and improving the identification of outliers. Furthermore, models can be applied to use contextual information and adjacent milk yields for missing value imputation (Liseune et al., 2021a; Macciotta et al., 2002; Vasconcelos et al., 2004)

The error of the input data set the highest achievable accuracy of a lactation curve model. It is therefore good practice to determine and report the error of the samples used to fit a model, as it helps determine model performance and aids in detecting overfitting (Zielesny et al., 2011).

5.3 Management of data quality

The outcome of the data quality assessment informs how data quality issues should be managed, either through preventive measures during data collection or through corrective actions applied before model fitting. The preferred method is the prevention of data quality issues. This involves establishing clear protocols, regularly calibrating measurement equipment, and implementing robust data governance practices (Andreen et al., 2020; Gengler, 2019). If preventive measures are insufficient, data cleaning can be applied as a secondary solution, which may include omitting outliers and inputting missing values. However, minimizing data manipulation is crucial, as each additional manipulation step introduces variability, which can accumulate and significantly affect the results. Furthermore, the indiscriminate removal of outliers may inadvertently exclude true perturbations, thereby reducing the ability of lactation curve models to accurately capture individual variation. However, it is impossible to validate the correctness of imputed variables. As an alternative for data manipulation, Liseune et al. (2020) implemented a score to identify the average recording quality of the lactation cycle and successfully used this as input in their lactation curve model, causing the model to take into account data quality issues.

In the end, the goal of data quality management is to get more reliable input variables for the lactation curve models, enabling the development of more accurate models, especially when this is combined with well-informed model selection, appropriate fitting techniques, and targeted application.

Published by Burleigh Dodds Science Publishing Limited, 2026.

6 Conclusion

Lactation curve modeling remains a cornerstone of dairy science, providing valuable insights into milk production dynamics and supporting a wide range of research and management applications. As demonstrated throughout this chapter, the development of a useful lactation curve model extends far beyond selecting a mathematical equation. Decisions regarding data collection, preprocessing, model selection, parameter estimation, validation, and interpretation all influence the quality and usefulness of the final model.

No single modeling approach is universally optimal. The most appropriate method depends on the research objective, available data, and intended application. Understanding the assumptions, strengths, and limitations of different approaches is therefore essential for making informed methodological choices and for correctly interpreting model outputs.

Continued progress in measuring equipment, data validation procedures, and analytical methodologies will further expand the possibilities of lactation curve modeling. At the same time, transparent reporting and thorough documentation remain critical to ensure reproducibility and enable future researchers to build upon existing knowledge. Although more than a century of research has substantially advanced the field, many opportunities for innovation remain. By combining high-quality data with appropriate modeling strategies, lactation curve models will continue to play an increasingly important role in advancing dairy research and supporting evidence-based decision making in the dairy industry.

7 Declaration of funding and/or conflicts of interest

The authors of this chapter declare they have not received any funding related to the chapter and have no conflicts to declare.

8 Where to look for further information

Lactation curves have been extensively covered in various review papers including Angeles-Hernandez et al., 2021; Beever et al., 1991; Bouallegue and M'Hamdi, 2019; Gipson and Grossman, 1990; Macciotta et al., 2011; Murphy et al., 2018; Rojas de Oliveira et al., 2024; Solodneva et al., 2022. Another useful resource is the ICAR wiki that includes guidelines for the use of lactation curve models (https://wiki.icar.org/index.php/Guidelines) and the recently developed lactation curve Python package and fitting API (van Leerdam et al., 2026).

This open-source Python package called *lactation curve* provides implementations of 14 lactation curve models, including traditional parametric

models, Bayesian fitting approaches, and an AI-based model. In addition to curve fitting, the package includes functionality for deriving biologically relevant lactation characteristics such as time to peak, peak yield, cumulative yield, and persistency. It also serves as a reference implementation for 305-day yield calculations.

The package can be integrated into analytical workflows to fit and compare lactation curves, estimate 305-day yields, and derive lactation characteristics. To facilitate broader accessibility, these functionalities are complemented by an interactive web platform (https://tools.bovi-analytics.org/) for visualizing lactation curve models and exploring their behavior, as well as an openly available GitHub repository containing the source code and documentation (https://github.com/Bovi-analytics/bovi). The platform additionally includes a validation tool that allows users to compare self-estimated 305-day yields against both reference calculations and observed daily milk yields using standardized lactation datasets.

9 References

Abdelkrim, A.B., et al. 2021. Lactation curve model with explicit representation of perturbations as a phenotyping tool for dairy livestock precision farming. *Animal* 15. https://doi.org/10.1016/j.animal.2020.100074.

Adriaens, I., et al. 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.

Afriliana, N., et al. 2021. Computational intelligence techniques for assessing data quality: towards knowledge-driven processing, In: Paszynski, M., et al. (Eds.), *Computational science – ICCS 2021*. Springer International Publishing, pp. 392–405. https://doi.org/10.1007/978-3-030-77967-2_33.

Ali, T.E. and Schaeffer, L.R. 1987. Accounting for covariances among test day milk yields in dairy cows. *Can J Anim Sci* 67, 637–644. https://doi.org/10.4141/cjas87-067.

Andersen, F., et al. 2011a. Mastitis and the shape of the lactation curve in Norwegian dairy cows. *J Dairy Res* 78, 23–31. https://doi.org/10.1017/S0022029910000749.

Andersen, F., et al. 2011b. Associations between the time of conception and the shape of the lactation curve in early lactation in Norwegian dairy cattle. *Acta Vet Scand* 53, 5. https://doi.org/10.1186/1751-0147-53-5.

Andreen, D.M., et al. 2020. *Technical note:* method for improving precision of in-parlor milk meters and adjusting milk weights for stall effects. *J Dairy Sci* 103, 5162–5169. https://doi.org/10.3168/jds.2019-17479.

Angeles-Hernandez, J.C., et al. 2021. Physiology of milk production and modelling of the lactation curve. *CABI Rev* 2021. https://doi.org/10.1079/PAVSNNR202116056.

Arnal, M., et al. 2018. Diversity of dairy goat lactation curves in France. *J Dairy Sci* 101, 11040–11051. https://doi.org/10.3168/jds.2018-14980.

Atashi, H., et al. 2020. Genome-wide association for milk production and lactation curve parameters in Holstein dairy cows. *J Anim Breed Genet* 137, 292–304. https://doi.org/10.1111/jbg.12442.

Atashi, H., et al. 2013. Association between dry period length and lactation performance, lactation curve, calf birth weight, and dystocia in holstein dairy cows in iran. *J Dairy Sci* 96, 3632–3638. https://doi.org/10.3168/jds.2012-5943.

Atashi, H., et al. 2021. Association between age at first calving and lactation performance, lactation curve, calving interval, calf birth weight, and dystocia in Holstein dairy cows. *PLoS One* 16, 1–13. https://doi.org/10.1371/journal.pone.0244825.

Batini, C., et al. 2009. Methodologies for data quality assessment and improvement. *ACM Comput Surv* 41, 16:1-16:52. https://doi.org/10.1145/1541880.1541883.

Beever, D.E., et al. 1991. A review of empirical and mechanistic models of lactational performance by the dairy cow. *Livest. Prod. Sci* 29, 115–130. https://doi.org/10.1016/0301-6226(91)90061-T.

Berghof, T.V.L., et al. 2019. Opportunities to improve resilience in animal breeding programs. *Front Genet* 9. https://doi.org/10.3389/fgene.2018.00692.

Berry, D.P., et al. 2005. Accuracy of predicting milk yield from alternative milk recording schemes. *Anim Sci* 80, 53–60. https://doi.org/10.1079/asc34880053.

Bouallegue, M. and M'Hamdi, N. 2019. Mathematical modeling of lactation curves: a review of parametric models, In: *Lactation in Farm Animals - Biology, Physiological Basis, Nutritional Requirements, and Modelization*. IntechOpen. https://doi.org/10.5772/intechopen.90253.

Brody, S., et al. 1923. The rate of decline of milk secretion with the advance of the period of lactation. *J Gen Physiol* 5, 441–444. https://doi.org/10.1085/jgp.5.4.441.

Brody, S., et al. 1924. The relation between the initial rise and the subsequent decline of milk secretion following parturition. *J Gen Physiol* 6, 541–545. https://doi.org/10.1085/jgp.6.5.541.

Bruijnis, M.R.N., et al. 2010. Assessing economic consequences of foot disorders in dairy cattle using a dynamic stochastic simulation model. *J Dairy Sci* 93, 2419–2432. https://doi.org/10.3168/jds.2009-2721.

Burgers, E.E.A., et al. 2021. Fertility and milk production on commercial dairy farms with customized lactation lengths. *J Dairy Sci* 104, 443–458. https://doi.org/10.3168/jds.2019-17947.

Butler, W. 2005. Nutrition, negative energy balance and fertility in the postpartum dairy cow. *Cattle Pract* 13, 13–18.

Cabrera, V.E. and Fadul-Pacheco, L. 2021. Future of dairy farming from the dairy brain perspective: data integration, analytics, and applications. *Int Dairy J* 121, 105069. https://doi.org/10.1016/j.idairyj.2021.105069.

Calanni Macchio, A., et al. 2025. The role of age at first calving in shaping production and reproductive outcomes in Italian buffaloes. *J Dairy Sci* 108, 7235–7247. https://doi.org/10.3168/jds.2025-26369.

Čandek-Potokar, M., et al. 2006. The uncertainty of results when estimating daily milk records. *Anim Res* 55, 521–532. https://doi.org/10.1051/animres:2006037.

Caraviello, D.Z., et al. 2006. Analysis of reproductive performance of lactating cows on large dairy farms using machine learning algorithms. *J Dairy Sci* 89, 4703–4722. https://doi.org/10.3168/jds.S0022-0302(06)72521-8.

Carbone, M.R. 2022. When not to use machine learning: a perspective on potential and limitations. *MRS Bull* 47, 968–974. https://doi.org/10.1557/s43577-022-00417-z.

Carlström, C., et al. 2013. Feasibility of using automatic milking system data from commercial herds for genetic analysis of milkability. *J Dairy Sci* 96, 5324–5332. https://doi.org/10.3168/jds.2012-6221.

Chen, J., et al. 2016. Effects of dry period length and dietary energy source on lactation curve characteristics over 2 subsequent lactations. *J Dairy Sci* 99, 9287–9299. https://doi.org/10.3168/jds.2016-11253.

Chen, S.-Y., et al. 2023a. Genomic-based genetic parameters for resilience across lactations in North American Holstein cattle based on variability in daily milk yield records. *J Dairy Sci* 106, 4133–4146. https://doi.org/10.3168/jds.2022-22754.

Chen, Y., et al. 2023b. Prediction of persistency for day 305 of lactation at the moment of the insemination decision. *Front Vet Sci* 10. https://doi.org/10.3389/fvets.2023.1264048.

Chen, Y., et al. 2024. Association between days post-conception and lactation persistency in dairy cattle. *J Dairy Sci* 107, 5794–5804. https://doi.org/10.3168/jds.2023-24282.

Cobby, J.M. and Du, Y.L.P.L. 1978. On fitting curves to lactation data. *Anim Sci* 26, 127–133. https://doi.org/10.1017/S0003356100039532.

Cole, J.B., et al. 2012. *Short communication:* projecting milk yield using best prediction and the MilkBot lactation model. *J Dairy Sci* 95, 4041–4044. https://doi.org/10.3168/jds.2011-4905.

Cole, J.B., et al. 2023. The effects of breeding and selection on lactation in dairy cattle. *Anim Front* 13, 62–70. https://doi.org/10.1093/af/vfad044.

Cole, J.B. and VanRaden, P.M. 2006. Genetic evaluation and best prediction of lactation persistency. *J Dairy Sci* 89, 2722–2728. https://doi.org/10.3168/jds.S0022-0302(06)72348-7.

Congleton, W.R. and Everett, R.W. 1980. Error and bias in using the incomplete gamma function to describe lactation curves. *J Dairy Sci* 63, 101–108. https://doi.org/10.3168/jds.S0022-0302(80)82894-3.

Costa, A., et al. 2019. Heritability and repeatability of milk lactose and its relationships with traditional milk traits, somatic cell score and freezing point in Holstein cows. *Animal* 13, 909–916. https://doi.org/10.1017/S1751731118002094.

Costa, A., et al. 2020. Changes in milk lactose content as indicators for longevity and udder health in Holstein cows. *J Dairy Sci* 103, 11574–11584. https://doi.org/10.3168/jds.2020-18615.

Costa, A., et al. 2025. Mastitis has a cumulative and lasting effect on milk yield and lactose content in dairy cows. *J Dairy Sci* 108, 635–650. https://doi.org/10.3168/jds.2024-25467.

CRV. 2015. NO Kengetallen E- chapter 3 Netto Opbrengst en Lactatiewaarde.

CRV. 2020. CRV actualiseert berekening kerngetallen NO, ISK, LM en BSK - CRV [WWW Document]. https://crv4all.nl/nl/news/crv-actualiseert-berekening-kerngetallen (accessed 4.4.25).

da Silva, T.E. and Cabrera, V.E. 2024. The DairyPrint model: a decision support model to help dairy farmers and other stakeholders toward improved sustainability. *J Dairy Sci* 107, 10998–11015. https://doi.org/10.3168/jds.2024-24946.

De Roos, A., et al. 2001. Random regression test-day model in the Netherlands. *Interbull Bull* 155–155.

de Vries, A. 2004. Economics of delayed replacement when cow performance is seasonal. *J Dairy Sci* 87, 2947–2958. https://doi.org/10.3168/jds.S0022-0302(04)73426-8.

Dekkers, J.C.M., et al. 1996. Genetic and economic evaluation of persistency in dairy cattle. In: *Proceedings of the International Workshop on Genetic Improvement of Functional Traits in Cattle*, Interbull Bulletin No. 12, 111–22. Uppsala, Sweden: Interbull.

Dekkers, J.C.M., et al. 1998. Economic aspects of persistency of lactation in dairy cattle. *Livest. Prod. Sci* 53, 237–252. https://doi.org/10.1016/S0301-6226(97)00124-3.

Dematawewa, C.M.B., et al. 2007. Modeling extended lactations of Holsteins. *J Dairy Sci* 90, 3924–36. https://doi.org/10.3168/jds.2006-790.

Dhanoa, M.S. 1981. A note on an alternative form of the lactation model of wood. *Anim. Prod* 32, 349–351. https://doi.org/10.1017/S0003356100027276.

Dijkstra, J., et al. 1997. A model to describe growth patterns of the mammary gland during pregnancy and lactation. *J Dairy Sci* 80, 2340–2354. https://doi.org/10.3168/jds.S0022-0302(97)76185-X.

Dijkstra, J., et al. 2010. Evaluation of a mechanistic lactation model using cow, goat and sheep data. *J Agric Sci* 148, 249–262. https://doi.org/10.1017/S0021859609990578.

Dongre, V.B., et al. 2012. Comparative efficiency of artificial neural networks and multiple linear regression analysis for prediction of first lactation 305-day milk yield in Sahiwal cattle. *Livest Sci* 147, 192–197. https://doi.org/10.1016/j.livsci.2012.04.002.

Duplessis, M., et al. 2019. Assessment of the canadian model predicting daily milk yield and milk fat percentage using single-milking dairy herd improvement samples. *Can J Anim Sci* 99, 521–531. https://doi.org/10.1139/cjas-2018-0215.

Ehrlich, J.L. 2010. Quantifying shape of lactation curves, and benchmark curves for common dairy breeds and parities. *Bov Pract*, 88–95. https://doi.org/10.21423/bovine-vol45no1p88-95.

Ehrlich, J.L. 2013. Quantifying inter-group variability in lactation curve shape and magnitude with the MilkBot® lactation model. *PeerJ* 1, e54. https://doi.org/10.7717/peerj.54.

Elahi Torshizi, M. 2016. Effects of season and age at first calving on genetic and phenotypic characteristics of lactation curve parameters in Holstein cows. *J Anim Sci Technol* 58, 8. https://doi.org/10.1186/s40781-016-0089-1.

Elvira, L., et al. 2013. Factors affecting the lactation curves of intensively managed sheep based on a clustering approach. *J Dairy Res* 80, 439–447. https://doi.org/10.1017/S0022029913000381.

European Parliament and Council of the European Union. 2023. Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data (Data Act), Official Journal of the European Union.

Fischer, A. 1958. Research with Württemberg spotted mountain cows on the shape of the lactation curve and how it may be influenced by non-genetic factors. *Ziichtungskunde* 30, 296–304.

Fluck, A.C., et al. 2024. Lactation performance, feed efficiency, and blood metabolites of dairy cows treated with recombinant bovine somatotropin: a systematic review and meta-analysis. *Res Vet Sci* 173, 105274. https://doi.org/10.1016/j.rvsc.2024.105274.

Gahlot, G.C., et al. 1988. Pattern of lactation curve in Rathi cattle. *Indian J Anim Sci* 58, 1112–1114.

Gantner, V., et al. 2009. Methods for estimation of daily and lactation milk yields from alternative milk recording scheme in Holstein and Simmental cattle breeds. *Ital J Anim Sci* 8, 519–530. https://doi.org/10.4081/ijas.2009.519.

García, S.C. and Holmes, C.W. 2001. Lactation curves of autumn- and spring-calved cows in pasture-based dairy systems. *Livest Prod Sci* 68, 189–203. https://doi.org/10.1016/S0301-6226(00)00237-2.

Gengler, N. 1996. Persistency of lactation yields: a review. *Interbull Bull* 12, 87–96.

Gengler, N. 2019. *Symposium review:* challenges and opportunities for evaluating and using the genetic potential of dairy cattle in the new era of sensor data from automation. *J Dairy Sci* 102, 5756–5763. https://doi.org/10.3168/jds.2018-15711.

Gilbert, G.R., et al. 1973. Diurnal variations in milk yield, fat yield, milk fat percentage, and milk protein percentage of Holstein-Friesian cows. *J Dairy Sci* 56, 409–410. https://doi.org/10.3168/jds.S0022-0302(73)85187-2.

Gipson, T.A. and Grossman, M. 1990. Lactation curves in dairy goats: a review. *Small Rumin Res* 3, 383–396. https://doi.org/10.1016/0921-4488(90)90019-3.

Gómez-Carpio, M., et al. 2023. Genetic parameters for reproductive traits in the Italian Mediterranean buffalo using milk yield as a correlated trait. *J Dairy Sci* 106, 9016–9025. https://doi.org/10.3168/jds.2023-23257.

Gorgulu, O. 2012. Prediction of 305-day milk yield in brown Swiss cattle using artificial neural networks. *South Afr J Anim Sci* 42, 280–287. https://doi.org/10.4314/sajas.v42i3.10.

Grzesiak, W., et al. 2003. A comparison of neural network and multiple regression predictions for 305-day lactation yield using partial lactation records. *Can J Anim Sci* 83, 307–310. https://doi.org/10.4141/A02-002.

Grzesiak, W., et al. 2006. Methods of predicting milk yield in dairy cows—Predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs). *Comput Electron Agric* 54, 69–83. https://doi.org/10.1016/j.compag.2006.08.004.

Grzesiak, W., et al. 2021. Lactation milk yield prediction in primiparous cows on a farm using the seasonal auto-regressive integrated moving average model, nonlinear autoregressive exogenous artificial neural networks and Wood's model. *Anim Biosci* 34, 770–782. https://doi.org/10.5713/ajas.19.0939.

Guevara, L., et al. 2023. Application of machine learning algorithms to describe the characteristics of dairy sheep lactation curves. *Animals* 13, 2772. https://doi.org/10.3390/ani13172772.

Hadley, G.L., et al. 2006. Dairy cattle culling patterns, explanations, and implications. *J Dairy Sci* 89, 2286–2296. https://doi.org/10.3168/jds.S0022-0302(06)72300-1.

Hand, K.J., et al. 2006. *Short Communication:* comparison of protocols to estimate twenty-four-hour fat and protein percentages for herds with a robotic milking system. *J Dairy Sci* 89, 1723–1726. https://doi.org/10.3168/jds.S0022-0302(06)72240-8.

Hansen, T.L., et al. 2021. The ruminant farm systems animal module: a biophysical description of animal management. *Animals* 11, 1373. https://doi.org/10.3390/ani11051373.

Hanuš, O., et al. 2014. Reliability of milk recording data under conditions of automatic milking system. *Acta Univ Agric Silvic Mendel Brun* 62, 911–917. https://doi.org/10.11118/actaun201462050911.

Hare, E., et al. 2004. Duration of herd participation in dairy herd improvement milk recording in the United States. *J Dairy Sci* 87, 2743–2747. https://doi.org/10.3168/jds.S0022-0302(04)73401-3.

Hassija, V., et al. 2024. Interpreting black-box models: a review on explainable artificial intelligence. *Cogn Comput* 16, 45–74. https://doi.org/10.1007/s12559-023-10179-8.

Hastie, T., et al. 2009. *The elements of statistical learning*, Springer Series in Statistics. Springer. https://doi.org/10.1007/978-0-387-84858-7.

Hogeveen, H., et al. 2001. Milking interval, milk production and milk flow-rate in an automatic milking system. *Livest Prod Sci* 72, 157–167. https://doi.org/10.1016/S0301-6226(01)00276-7.

Hostens, M., et al. 2012. On-farm evaluation of the effect of metabolic diseases on the shape of the lactation curve in dairy cows through the MilkBot lactation model. *J Dairy Sci* 95, 2988–3007. https://doi.org/10.3168/jds.2011-4791.

ICAR Wiki contributors. 2025a. ICAR Wiki [WWW Document]. Guidelines. http://wiki.icar.org/index.php?title=Guidelines&oldid=4632 (accessed 3.23.26).

ICAR Wiki contributors. 2025b. *Section 02 – Cattle milk recording*. ICAR Wiki.

Ince, D. and Sofu, A. 2013. Estimation of lactation milk yield of Awassi sheep with artificial neural network modeling. *Small Rumin Res* 113, 15–19. https://doi.org/10.1016/j.smallrumres.2013.01.013.

Inchaisri, C., et al. 2011. Analysis of the economically optimal voluntary waiting period for first insemination. *J Dairy Sci* 94, 3811–3823. https://doi.org/10.3168/jds.2010-3790.

Interbull Centre. 2026. PREPdb - Performance Recording, Evaluation and Publication database [Online].

Jamrozik, J., et al. 1997. Multiple trait random regression test day model for production traits. *Interbull Bull* 16, 43–43.

Jensen, J. 2001. Genetic evaluation of dairy cattle using test-day models1. *J Dairy Sci* 84, 2803–2812. https://doi.org/10.3168/jds.S0022-0302(01)74736-4.

Jordan, M.I. and Mitchell, T.M. 2015. Machine learning: trends, perspectives, and prospects. *Science* 349, 255–260. https://doi.org/10.1126/science.aaa8415.

Kamphuis, C., et al. 2015. Devices used by automated milking systems are similarly accurate in estimating milk yield and in collecting a representative milk sample compared with devices used by farms with conventional milk recording. *J Dairy Sci* 98, 3541–3557. https://doi.org/10.3168/jds.2014-8714.

Kebreab, E., et al. 2019. A new modeling environment for integrated dairy system management. *Anim Front* 9, 25–32. https://doi.org/10.1093/af/vfz004.

Khazaei, J., et al. 2008. Approximating milk yield and milk fat and protein concentration of cows through the use of mathematical and artificial neural networks models, In: World Conference on Agricultural Information and IT, pp. 91–105.

Knapp, J.R., et al. 2025. Proceedings of the 2024 meeting of the animal science modelling group. *Can J Anim Sci* 105, S1–S6. https://doi.org/10.1139/cjas-2024-0152.

Kok, A., et al. 2019. Production, partial cash flows and greenhouse gas emissions of simulated dairy herds with extended lactations. *Animal* 13, 1074–1083. https://doi.org/10.1017/S1751731118002562.

Kominakis, A.P., et al. 2002. A preliminary study of the application of artificial neural networks to prediction of milk yield in dairy sheep. *Comput Electron Agric* 35, 35–48. https://doi.org/10.1016/S0168-1699(02)00051-0.

Lacroix, R., et al. 1995. Prediction of cow performance with a connectionist model. *Trans ASAE* 38, 1573–1579. https://doi.org/10.13031/2013.27984.

Lacroix, R., et al. 1997. Effect of data preprocessing on the performance of artificial neural networks for dairy yield prediction and cow culling classification. *Trans ASAE* 40(3), 839–846. https://doi.org/10.13031/2013.21294.

Lazenby, D., et al. 2003. *Methods to estimate 24-hour yields for milk, fat and protein in robotic milking herds*. Brill. https://doi.org/10.3920/9789086865116_010.

Lee, S. and Park, J. 2022. A vector representation of lactation curves for dairy cows. *Agriculture* 12, 395. https://doi.org/10.3390/agriculture12030395.

Lehenbauer, T.W. and Oltjen, J.W. 1998. Dairy cow culling strategies: making economical culling decisions. *J Dairy Sci* 81, 264–271. https://doi.org/10.3168/jds.S0022-0302(98)75575-4.

Lehmann, J.O., et al. 2019. Extended lactations in dairy production: economic, productivity and climatic impact at herd, farm and sector level. *Livest Sci* 220, 100–110. https://doi.org/10.1016/j.livsci.2018.12.014.

Li, M., et al. 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.

Liseune, A., et al. 2020. Leveraging latent representations for milk yield prediction and interpolation using deep learning. *Comput Electron Agric* 175, 105600. https://doi.org/10.1016/j.compag.2020.105600.

Liseune, A., et al. 2021a. Leveraging sequential information from multivariate behavioral sensor data to predict the moment of calving in dairy cattle using deep learning. *Comput Electron Agric* 191, 106566. https://doi.org/10.1016/j.compag.2021.106566.

Liseune, A., et al. 2021b. Predicting the milk yield curve of dairy cows in the subsequent lactation period using deep learning. *Comput Electron Agric* 180, 105904. https://doi.org/10.1016/j.compag.2020.105904.

Liu, H., et al. 2019. A rule based data quality assessment architecture and application for electrical data, In: Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing, AIIPCC '19. Association for Computing Machinery, New York, NY, USA, pp. 1–6. https://doi.org/10.1145/3371425.3371435.

Liu, Z., et al. 2000. Approaches to estimating daily yield from single milk testing schemes and use of a.m.-p.m. Records in test-day model genetic evaluation in dairy cattle. *J Dairy Sci* 83, 2672–2682. https://doi.org/10.3168/jds.S0022-0302(00)75161-7.

Macciotta, N.P.P., et al. 2002. Test day and lactation yield predictions in Italian Simmental cows by ARMA methods. *J Dairy Sci* 85, 3107–3114. https://doi.org/10.3168/jds.S0022-0302(02)74398-1.

Macciotta, N.P.P., et al. 2005. Detection of different shapes of lactation curve for milk yield in dairy cattle by empirical mathematical models. *J Dairy Sci* 88, 1178–1191. https://doi.org/10.3168/jds.S0022-0302(05)72784-3.

Macciotta, N.P.P., et al. 2006. Use of multivariate analysis to extract latent variables related to level of production and lactation persistency in dairy cattle. *J Dairy Sci* 89, 3188–3194. https://doi.org/10.3168/jds.S0022-0302(06)72593-0.

Macciotta, N.P.P., et al. 2010. Comparison of parametric, orthogonal, and spline functions to model individual lactation curves for milk yield in Canadian Holsteins. *Ital J Anim Sci* 9, e87. https://doi.org/10.4081/ijas.2010.e87.

Macciotta, N.P.P., et al. 2011. The mathematical description of lactation curves in dairy cattle. *Ital J Anim Sci* 10, e51. https://doi.org/10.4081/ijas.2011.e51.

Madsen, O. 1975. A comparison of some suggested measures of persistency of milk yield in dairy cows. *Anim Sci* 20, 191–197. https://doi.org/10.1017/S0003356100035182.

Published by Burleigh Dodds Science Publishing Limited, 2026.

Maltz, E. 2020. Individual dairy cow management: achievements, obstacles and prospects. *J Dairy Res* 87, 145–157. https://doi.org/10.1017/S0022029920000382.

Martens, H. 2023. Invited review: increasing milk yield and negative energy balance: a Gordian knot for dairy cows? *Animals* 13, 3097. https://doi.org/10.3390/ani13193097.

Martin, O. and Sauvant, D. 2010. A teleonomic model describing performance (body, milk and intake) during growth and over repeated reproductive cycles throughout the lifespan of dairy cattle. 1. Trajectories of life function priorities and genetic scaling. *Anim Consort* 2030–2047. https://doi.org/10.1017/S1751731110001357.

Masia, F., et al. 2022. Quantifying the negative impact of clinical diseases on productive and reproductive performance of dairy cows in central Argentina. *Livest Sci* 259, 104894. https://doi.org/10.1016/j.livsci.2022.104894.

Masía, F.M., et al. 2020. Modeling variability of the lactation curves of cows in automated milking systems. *J Dairy Sci* 103, 8189–8196. https://doi.org/10.3168/jds.2019-17962.

Mein, G.A., et al. 2001. *Limits of error for permanently-installed milk meters used for herd recording or for daily herd management purposes*. Brill. https://doi.org/10.1163/9789004683679_025.

Miglior, F., et al. 2007. *Report of the ICAR Working Group on Lactation Calculation Methods. xA survey on milk recording strategies in dairy cattle*. Brill. https://doi.org/10.3920/9789086865932_033.

Miglior, F., et al. 2017. *A 100-Year Review:* identification and genetic selection of economically important traits in dairy cattle. *J Dairy Sci* 100, 10251–10271. https://doi.org/10.3168/jds.2017-12968.

Migose, S.A., et al. 2020. Accuracy of estimates of milk production per lactation from limited test-day and recall data collected at smallholder dairy farms. *Livest Sci* 232. https://doi.org/10.1016/j.livsci.2019.103911.

Misztal, I., et al. 2000. Strategies for estimating the parameters needed for different test-day models. *J Dairy Sci* 83, 1125–1134. https://doi.org/10.3168/jds.S0022-0302(00)74978-2.

Moawed, S.A., et al. 2024. Milk yield prediction in Friesian cows using linear and flexible discriminant analysis under assumptions violations. *BMC Vet Res* 20, 392. https://doi.org/10.1186/s12917-024-04234-1.

Monteiro, A., et al. 2021. Precision agriculture for crop and livestock farming—brief review. *Animals* 11, 2345. https://doi.org/10.3390/ani11082345.

Murphy, M.D., et al. 2014. Comparison of modelling techniques for milk-production forecasting. *J Dairy Sci* 97, 3352–3363. https://doi.org/10.3168/jds.2013-7451.

Murphy, M., et al. 2018. A review of milk production forecasting models. In: *Dairy farming, operations management, animal welfare and milk production*. Nova Science Publishers, pp. 14–61.

Nayeri, S., et al. 2019. A review of traditional and machine learning methods applied to animal breeding. *Anim Health Res Rev* 20, 31–46. https://doi.org/10.1017/S1466252319000148.

Nelder, J.A. 1966. Inverse polynomials, a useful group of multi-factor response functions. *Biometrics* 22, 128–141. https://doi.org/10.2307/2528220.

Nestorov, I., et al. 1999. Empirical versus mechanistic modelling: comparison of an artificial neural network to a mechanistically based model for quantitative structure

pharmacokinetic relationships of a homologous series of barbiturates. *AAPS PharmSci* 1, 5–13. https://doi.org/10.1208/PS010417/METRICS.

Nguyen, Q.T., et al. 2020. Comparison of forecast models of production of dairy cows combining animal and diet parameters. *Comput Electron Agric* 170, 105258. https://doi.org/10.1016/j.compag.2020.105258.

Njubi, D.M., et al. 2010. Use of test-day records to predict first lactation 305-day milk yield using artificial neural network in Kenyan Holstein–Friesian dairy cows. *Trop Anim Health Prod* 42, 639–644. https://doi.org/10.1007/s11250-009-9468-7.

Oleggini, G.H., et al. 2001. Effect of region and herd size on dairy herd performance parameters. *J Dairy Sci* 84, 1044–1050. https://doi.org/10.3168/jds.S0022-0302(01)74564-X.

Oliveira, H.R., et al. 2019. Application of single-step genomic evaluation using multiple-trait random regression test-day models in dairy cattle. *J Dairy Sci* 102, 2365–2377. https://doi.org/10.3168/jds.2018-15466.

Olori, V.E., et al. 1999. Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. *Livest Prod Sci* 58, 55–63. https://doi.org/10.1016/S0301-6226(98)00194-8.

Omar, A.I., et al. 2024. Factors affecting the milk production traits and lactation curve of the indigenous river buffalo populations in Bangladesh. *Animals* 14. https://doi.org/10.3390/ani14081248.

Pietersma, D., et al. 2002. Machine-learning assisted knowledge acquisition to filter lactation curve data. *Trans ASAE* 1637–1650. https://doi.org/10.13031/2013.11049.

Pietersma, D., et al. 2003. Induction and evaluation of decision trees for lactation curve analysis. *Comput Electron Agric* 38, 19–32. https://doi.org/10.1016/S0168-1699(02)00105-9.

Pollott, G.E. 2000. A biological approach to lactation curve analysis for milk yield. *J Dairy Sci* 83, 2448–2458. https://doi.org/10.3168/jds.S0022-0302(00)75136-8.

Poppe, M., et al. 2020. Exploration of variance, autocorrelation, and skewness of deviations from lactation curves as resilience indicators for breeding. *J Dairy Sci* 103, 1667–1684. https://doi.org/10.3168/jds.2019-17290.

Poppe, M., et al. 2021. Genetic analysis of resilience indicators based on milk yield records in different lactations and at different lactation stages. *J Dairy Sci* 104, 1967–1981. https://doi.org/10.3168/jds.2020-19245.

Prasad, S. 2003. Modeling of lactation curves of dairy animals. *J Appl Anim Res* 24, 79–84. https://doi.org/10.1080/09712119.2003.9706437.

Ptak, E. and Schaeffer, L.R. 1993. Use of test day yields for genetic evaluation of dairy sires and cows. *Livest. Prod. Sci* 34, 23–34. https://doi.org/10.1016/0301-6226(93)90033-E.

Radjabalizadeh, K., et al. 2022. Estimation of genetic parameters of Wood's lactation curve parameters using Bayesian and REML methods for milk production trait of Holstein dairy cattle. *J Appl Anim Res* 50, 363–368. https://doi.org/10.1080/09712119.2022.2080211.

Rekaya, R., et al. 2000. Bayesian analysis of lactation curves of Holstein-Friesian Cattle using a nonlinear model. *J Dairy Sci* 83, 2691–2701. https://doi.org/10.3168/jds.S0022-0302(00)75163-0.

Published by Burleigh Dodds Science Publishing Limited, 2026.

Rekik, B. and Gara, A.B. 2004. Factors affecting the occurrence of atypical lactations for Holstein–Friesian cows. *Livest Prod Sci* 87, 245–250. https://doi.org/10.1016/j.livprodsci.2003.09.023.

Robinson, S. 2023. Exploring the relationship between simulation model accuracy and complexity. *J Oper Res Soc* 74, 1992–2011. https://doi.org/10.1080/01605682.2022.2122740.

Rojas de Oliveira, H., et al. 2024. Phenotypic and genomic modeling of lactation curves: a longitudinal perspective. *JDS Commun* 5, 241–246. https://doi.org/10.3168/jdsc.2023-0460.

Rook, A.J., et al. 1993. On the mathematical description of lactation curves. *J Agric Sci* 121, 97–102.

Różańska-Zawieja, J., et al. 2021. The effect of feeding management and culling of cows on the lactation curves and milk production of primiparous dairy cows. *Animals* 11, 1959. https://doi.org/10.3390/ani11071959.

Rudin, C., et al. 2022. Interpretable machine learning: fundamental principles and 10 grand challenges. *Stat Surv* 16, 1–85. https://doi.org/10.1214/21-SS133.

Salama, A.A.K., et al. 2005. Effect of pregnancy and extended lactation on milk production in dairy goats milked once daily. *J Dairy Sci* 88, 3894–3904. https://doi.org/10.3168/jds.S0022-0302(05)73075-7.

Salamone, M., et al. 2024. Milk yield residuals and their link with the metabolic status of dairy cows in the transition period. *J Dairy Sci* 107, 317–330. https://doi.org/10.3168/jds.2023-23641.

Salehi, F., et al. 1998. Improving dairy yield predictions through combined record classifiers and specialized artificial neural networks. *Comput Electron Agric* 20, 199–213. https://doi.org/10.1016/S0168-1699(98)00018-0.

Sargent, F.D., et al. 1968. Test interval method of calculating dairy herd improvement association records. *J Dairy Sci* 51, 170–179. https://doi.org/10.3168/jds.S0022-0302(68)86943-7.

Satoła, A. and Satoła, K. 2024. Performance comparison of machine learning models used for predicting subclinical mastitis in dairy cows: bagging, boosting, stacking, and super-learner ensembles versus single machine learning models. *J Dairy Sci* 107, 3959–3972. https://doi.org/10.3168/jds.2023-24243.

Sawalha, R.M., et al. 2005. Genetic evaluation of dairy cattle with test-day models with autoregressive covariance structures and with a 305-d model. *J Dairy Sci* 88, 3346–3353. https://doi.org/10.3168/jds.S0022-0302(05)73018-6.

Schaeffer, L.R., et al. 2000. Experience with a test-day model. *J Dairy Sci* 83, 1135–1144. https://doi.org/10.3168/jds.S0022-0302(00)74979-4.

Sehested, J., et al. 2019. Review: extended lactation in dairy cattle. *Animal* 13, s65–s74. https://doi.org/10.1017/S1751731119000806.

Shahinfar, S., et al. 2014. Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algorithms. *J Dairy Sci* 97, 731–742. https://doi.org/10.3168/jds.2013-6693.

Sharma, A.K., et al. 2007. Prediction of first lactation 305-day milk yield in Karan Fries dairy cattle using ANN modeling. *Appl Soft Comput* 7, 1112–1120. https://doi.org/10.1016/j.asoc.2006.07.002.

Sharma, S., et al. 2026. *Invited review:* milking the data for value-driven dairy farming. *J Dairy Sci*. https://doi.org/10.3168/jds.2025-27011.

Sikka, L.C. 1950. A study of lactation as affected by heredity and environment. *J Dairy Res* 17, 231–252. https://doi.org/10.1017/S0022029900005811.

Silvestre, A.M., et al. 2006. The accuracy of seven mathematical functions in modeling dairy cattle lactation curves based on test-day records from varying sample schemes. *J Dairy Sci* 89, 1813–1821. https://doi.org/10.3168/jds.S0022-0302(06)72250-0.

Silvestre, A.M., et al. 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., et al. 2019. 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.

Slob, N., et al. 2021. Application of machine learning to improve dairy farm management: a systematic literature review. *Prev Vet Med* 187, 105237. https://doi.org/10.1016/j.prevetmed.2020.105237.

Solodneva, E., et al. 2022. Lactation curves as a tool for monitoring the health and performance of dairy cows-a mini-review. *J Agric Biol* 57, 257–271. https://doi.org/10.15389/agrobiology.2022.2.257eng.

Sorensen, A., et al. 2008. Extended lactation in dairy cows: effects of milking frequency, calving season and nutrition on lactation persistency and milk quality. *J Dairy Res* 75, 90–97. https://doi.org/10.1017/S0022029907002944.

Sorensen, M.T., et al. 2006. Cell turnover and activity in mammary tissue during lactation and the dry period in dairy cows. *J Dairy Sci* 89, 4632–4639. https://doi.org/10.3168/jds.S0022-0302(06)72513-9.

Sorge, U.S., et al. 2007. Evaluation of the dairy comp 305 module "cow value" in two Ontario dairy herds. *J Dairy Sci* 90, 5784–5797. https://doi.org/10.3168/jds.2006-0813.

Stanton, T.L., et al. 1992. Estimating milk, fat, and protein lactation curves with a test day model. *J Dairy Sci* 75, 1691–1700. https://doi.org/10.3168/jds.S0022-0302(92)77926-0.

Steeneveld, W., et al. 2007. Stochastic modelling to assess economic effects of treatment of chronic subclinical mastitis caused by Streptococcus uberis. *J Dairy Res* 74, 459–467. https://doi.org/10.1017/s0022029907002828.

Steeneveld, W., et al. 2011. Cow-specific treatment of clinical mastitis: an economic approach. *J Dairy Sci* 94, 174–188. https://doi.org/10.3168/jds.2010-3367.

Strucken, E.M., et al. 2011. Lactation curve models for estimating gene effects over a timeline. *J Dairy Sci* 94, 442–449. https://doi.org/10.3168/jds.2009-2932.

Strucken, E.M., et al. 2015. Go with the flow—biology and genetics of the lactation cycle. *Front Genet* 6. https://doi.org/10.3389/fgene.2015.00118.

Swalve, H.H. 2000. Theoretical basis and computational methods for different test-day genetic evaluation methods. *J Dairy Sci* 83, 1115–1124. https://doi.org/10.3168/jds.S0022-0302(00)74977-0.

Togashi, K. and Lin, C.Y. 2009. Economic weights for genetic improvement of lactation persistency and milk yield. *J Dairy Sci* 92, 2915–2921. https://doi.org/10.3168/jds.2008-1647.

Togashi, K., et al. 2016. Genotype-by-environment interaction on genetic relationships between lactation persistency and conception measures in Japanese Holstein cows. *Livest Sci* 183, 40–47. https://doi.org/10.1016/j.livsci.2015.11.008.

Torres, M., et al. 2005. Approximating the sheep milk production curve through the use of artificial neural networks and genetic algorithms. *Comput Oper Res* 32, 2653–2670. https://doi.org/10.1016/j.cor.2004.06.025.

Van Hoeij, R.J., et al. 2017. Consequences of dietary energy source and energy level on energy balance, lactogenic hormones, and lactation curve characteristics of cows after a short or omitted dry period. *J Dairy Sci* 100,10, 8544-8564. https://doi.org/10.3168/jds.2017-12855.

van Leerdam, M.B., et al. 2026. Bovi-analytics/lactation_curve_core: v.1.1.1. https://doi.org/10.5281/zenodo.18715144.

VanRaden, P.M. 1998. Best prediction of lactation yield and persistency, In: Proc. 6th World Congr. Genet. Appl. Livest. Prod. (Armidale, Australia), pp. 347–350.

VanRaden, P.M., et al. 2006. Productive life including all lactations and longer lactations with diminishing credits. *J Dairy Sci* 89, 3213–3220. https://doi.org/10.3168/jds.S0022-0302(06)72596-6.

Vasconcelos, J., et al. 2004. Prediction of daily and lactation yields of milk, fat, and protein using an autoregressive repeatability test day model. *J Dairy Sci* 87, 2591–2598. https://doi.org/10.3168/jds.S0022-0302(04)73384-6.

Verbyla, K.L. and Verbyla, A.P. 2009. Estimated breeding values and association mapping for persistency and total milk yield using natural cubic smoothing splines. *Genet Sel Evol* 41, 48. https://doi.org/10.1186/1297-9686-41-48.

Vergani, A.M., et al. 2024. Predicting bovine daily milk yield by leveraging genomic breeding values. *Comput Electron Agric* 219, 108777. https://doi.org/10.1016/j.compag.2024.108777.

Wang, A., et al. 2022. Exploring milk loss and variability during environmental perturbations across lactation stages as resilience indicators in Holstein cattle. *Front Genet* 13. https://doi.org/10.3389/fgene.2022.1031557.

Wang, Z., et al. 2020. A rule-based data quality assessment system for electronic health record data. *Appl Clin Inform* 11, 622–634. https://doi.org/10.1055/s-0040-1715567.

Wathes, D.C., et al. 2007. Influence of negative energy balance on cyclicity and fertility in the high producing dairy cow. *Theriogenology, Proceedings of the International Conference on Farm Animal Reproduction* 68, S232–S241. https://doi.org/10.1016/j.theriogenology.2007.04.006.

Weigel, K.A., et al. 2017. *A 100-Year Review:* methods and impact of genetic selection in dairy cattle—From daughter–dam comparisons to deep learning algorithms. *J Dairy Sci* 100, 10234–10250. https://doi.org/10.3168/jds.2017-12954.

Weller, J.I., et al. 2017. Invited review: a perspective on the future of genomic selection in dairy cattle. *J Dairy Sci* 100, 8633–8644. https://doi.org/10.3168/JDS.2017-12879.

White, I.M.S., et al. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. *J Dairy Sci* 82, 632–638. https://doi.org/10.3168/jds.S0022-0302(99)75277-X.

Wiggans, G.R. and Goddard, M.E. 1997. A computationally feasible test day model for genetic evaluation of yield traits in the United States. *J Dairy Sci* 80, 1795–1800. https://doi.org/10.3168/jds.S0022-0302(97)76113-7.

Wilmink, J.B.M. 1987a. Comparison of different methods of predicting 305-day milk yield using means calculated from within-herd lactation curves. *Livest Prod Sci* 17, 1–17. https://doi.org/10.1016/0301-6226(87)90049-2.

Wilmink, J.B.M. 1987b. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livest Prod Sci* 16, 335–348. https://doi.org/10.1016/0301-6226(87)90003-0.

Wilson, D.J., et al. 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.

Wilson, D.J., et al. 2008. Milk production change following clinical mastitis and reproductive performance compared among j5 vaccinated and control dairy cattle. *J Dairy Sci* 91, 3869–3879. https://doi.org/10.3168/jds.2008-1405.

Winkler, W.E. 2004. Methods for evaluating and creating data quality. *Inf Syst* 29, 531–550. https://doi.org/10.1016/j.is.2003.12.003.

Wolfert, S., et al. 2017. Big data in smart farming: a review. *Agric Syst* 153, 69–80. https://doi.org/10.1016/j.agsy.2017.01.023.

Wood, P.D.P. 1967. Algebraic model of the lactation curve in cattle. *Nature* 216, 164–165. https://doi.org/10.1038/216164a0.

Wu, X.-L., et al. 2022. Statistical methods revisited for estimating daily milk yields: how well do they work? *Front Genet* 13. https://doi.org/10.3389/fgene.2022.943705.

Wu, X.-L., et al. 2023a. Updating test-day milk yield factors for use in genetic evaluations and dairy production systems: a comprehensive review. *Front Genet* 14. https://doi.org/10.3389/fgene.2023.1298114.

Wu, X.-L., et al. 2023b. Daily milk yield correction factors: what are they? *JDS Commun* 4, 40–45. https://doi.org/10.3168/jdsc.2022-0230.

Yamazaki, T., et al. 2009. Relationship between the lactation curve and udder disease incidence in different lactation stages in first-lactation Holstein cows. *Anim Sci J* 80, 636–643. https://doi.org/10.1111/j.1740-0929.2009.00695.x.

Yamazaki, T., et al. 2011. Phenotypic relationship between lactation persistency and change in body condition score in first-lactation Holstein cows. *Asian-Australas J Anim Sci* 24, 610–615. https://doi.org/10.5713/ajas.2011.10271.

Zhang, F., et al. 2022. Predicting daily milk yield for primiparous cows using data of within-herd relatives to capture genotype-by-environment interactions. *J Dairy Sci* 105, 6739–6748. https://doi.org/10.3168/jds.2021-21559.

Zhang, L., et al. 2021. Appropriate data quality checks improve the reliability of values predicted from milk mid-infrared spectra. *Anim Open Access J* 11, 533. https://doi.org/10.3390/ani11020533.

Zhang, W., et al. 2020. Daily milk yield prediction of dairy cows based on the GA-LSTM algorithm, In: 2020 15th IEEE International Conference on Signal Processing (ICSP). Presented at the 2020 15th IEEE International Conference on Signal Processing (ICSP), pp. 664–668. https://doi.org/10.1109/ICSP48669.2020.9320926.

Zielesny, A., et al. 2011. *From curve fitting to machine learning*. Springer.