TERNAK TROPIKA
Journal of Tropical Animal Production
Vol 26, No. 1 pp. 78-90, Juni 2025

DOI: 10.21776/ub.jtapro.2025.026.01.9
OPEN ACCES Freely Available Online

PREDICTION OF MILK YIELD AND LACTATION CURVE FROM EARLY STAGE MILK RECORDING DATA: A COMPARATIVE ANALYSIS OF THREE MATHEMATICAL MODELS IN TROPICAL SMALLHOLDER DAIRIES

Aswah Ridhowi[1*)], Trianti Djoharjani[1)], Sucik Maylinda[2)]

[1)] Departement of Animal Production, Faculty of Animal Science, Universitas Brawijaya, Jl. Veteran, Ketawanggede, Kec. Lowokwaru, Kota Malang, Jawa Timur, Indonesia 65145
[2)] Departement of Animal Reproduction and Breeding, Faculty of Animal Science, Universitas Brawijaya, Jl. Veteran, Ketawanggede, Kec. Lowokwaru, Kota Malang, Jawa Timur, Indonesia 65145

Corresponding author: aswah.ridhowi@ub.ac.id
Submitted 30 Mei 2025, Accepted 29 Juni 2025

ABSTRACT

Accurate milk yield prediction is essential for effective dairy herd management, particularly in smallholder dairy where daily milk recording is often limited. Predictive models that can estimate total milk yield during one lactation period using early-stage recording data offer a practical solution to support decision-making in such environments. This study aimed to evaluate the predictive performance of three mathematical models, i.e incomplete gamma function model $y_x = ax^b e^{-cx}$, orthogonal polynomial contrast model $y_x = x(a + bx + cx^2)^{-1}$, and non-linear regression model $y_x = a - bx - ae^{-cx}$ for estimating total milk production during lactation in dairy cows. Milk yield data were obtained from 164 Friesian Holstein cows across five lactation parities at a dairy cooperative in East Java, Indonesia. Milk production records over the first three months (13 weeks) of lactation were used to estimate total 305-day (44 weeks) milk yield using three predictive mathematical models, each fitted with parity-specific constants (a, b, c). Model performance was evaluated by comparing predicted and actual milk yields, using absolute and percentage errors as accuracy metrics. All models demonstrated acceptable predictive ability under weekly data conditions, with average percentage errors below 10%. The incomplete gamma function model showed the highest predictive accuracy and stability with lowest deviation (average deviation: 274.67 L; 6.17%), followed by orthogonal polynomial contrast model (324.43 L; 7.29%) and non-linear regression model (346.27 L; 7.78%). Those mathematical model exhibited stronger alignment with biological lactation patterns, and more sensitive to variation across parities. Frequent data collection enhances the accuracy of milk yield predictions. The incomplete gamma function model is recommended for initial milk yield prediction in smallholder dairy systems, offering an optimal balance between flexibility and biological plausibility. These findings support the integration of predictive modeling into routine herd management practices to improve productivity and sustainability.

Keywords: Dairy cow; milk yield prediction; mathematical model; early lactation.

How to cite : Ridhowi, A., Djoharjani, T., & Maylinda, S. (2025). Prediction of Milk Yield and Lactation Curve from Early-Stage Milk Recording Data: A Comparative Analysis of Three Mathematical Models in Tropical Smallholder Dairies. TERNAK TROPIKA Journal of Tropical Animal Production Vol 26, No 1 (78-90)

INTRODUCTION

Recording of milk production is a fundamental aspect of dairy farm management, yet it is often neglected by smallholder dairy farmers in developing countries, including Indonesia. Several factors contribute to this neglect, such as limited financial resources, time constraints, labor shortages, and a general lack of awareness regarding the importance of milk yield recording. Some studies from Pica-Ciamarra et al (2014), Migose et al., (2019) and Oliviera et al., (2024) reported that the absence of milk production records hampers the ability of farmers to track the performance of individual cows across a complete lactation period, ultimately restricting their capacity to make informed decisions for future herd development and farm profitability. To address this challenge, milk yield estimation methods have been developed that do not require complete daily milk records. Among these, the Simplified Method (SM) and Test Interval Method (TIM) (Everett and Carther, 1968; Pereira et al., 2001), or other estimation model such as iterative wood model (IW), a perturbed lactation model (PLM), and a quantile regression (QR) have been widely recognized as reliable techniques for estimating lactation milk yield using partial data collected over time (Rana et al., 2021; Ranzato et al., 2024). However, these methods still rely on milk yield data obtained at the end of the lactation period, thereby limiting their utility for early decision-making, such as identifying low-producing cows or adjusting feeding strategies during lactation.

Early prediction of total milk yield during an ongoing lactation would enable farmers to adopt timely and strategic management interventions. This includes early selection or culling of dairy cows, more accurate feed budgeting, and more precise economic forecasting. In this context, Sudarwati (1990) developed predictive models based on mathematical equations incomplete gamma function model, orthogonal polynomial contrast model, and non-linear regression model based on involving parameters a, b, and c, which have been adapted to Indonesian dairy farming conditions. These models allow for the estimation of both the total milk yield and the shape of the lactation curve at an earlier stage of lactation. Therefore, this study aims to evaluate and apply early lactation milk yield prediction models in smallholder dairy farms. By enabling earlier milk yield prediction, this approach has the potential to enhance management efficiency, reduce economic losses due to low performance, and contribute to the sustainability of dairy production systems.

MATERIALS AND METHODS

Animals and Data Source

Data comprising daily milk production records from 305 day lactation were obtained from a total of 164 lactating Friesian Holstein cows data in “Setia Kawan” Dairy Cooperative (KPSP Setia Kawan), Nongkojajar, Tutur Subdistrict, Pasuruan Regency, East Java, Indonesia. The animals were classified based on lactation number: 12 cows in first lactation, 48 in second, 48 in third, 44 in fourth, and 12 in fifth lactation.

Treatment

An observational study was conducted by comparing prediction of total milk yield and lactation curve with actual of milk yield for 305 days (44 weeks). Prediction of milk yield and lactation curve was calculated from 13 weeks milk production record at early lactation using 3 mathematical models:

1. Incomplete Gamma Function
   $y_x = ax^b e^{-cx}$
2. Orthogonal Polynomial Contrast
   $y_x = x(a + bx + cx^2)^{-1}$
3. Non Linear Regression
   $y_x = a - bx - ae^{-cx}$

   y_x denotes milk yield at time
   a, b, c are empirical constants
   e represents constat value (2.718).

The estimated total milk yield was derived by calculating the prediction bias generated by each of the three mathematical models based on average weekly milk production data during 13 weeks of lactation. The bias resulted during 13 weeks lactation record from each mathematical model was used to estimate total milk yield and lactation curve during 305 days (44 weeks). Parameters a, b, and c for each model and lactation group were adapted from Sudarwati (1990), in Tables 1.

Tabel 1. Estimated constants (a, b, and c) for the three lactation models across parity 1 to 5, derived from weekly milk yield records during the first three months of lactation.

No	Model	Parity	a	b	c
1	$y_x = ax^b e^{-cx}$	1	8,1605	0.0778	0,0150
		2	12,2105	0.0058	0,0160
		3	12,3394	0.0437	0,0151
		4	14,3651	0,0060	0,0131
		5	13,6686	0,0034	0,0113
2	$y_x = x(a + bx + cx^2)$	1	0,0339	0,0864	0,0014
		2	0,0222	0,0620	0,0017
		3	0,0223	0,0620	0,0012
		4	0,0218	0,0539	0,0012
		5	0,0215	0,0581	0,0010
3	$y_x = a - bx - ae^{-cx}$	1	9,3551	0,0736	67,6022
		2	12,3489	0,1340	27,2237
		3	13,2980	0,1180	70,1789
		4	14,3894	0,1242	20,1029
		5	13,8362	0,1064	72,0784

The analytical workflow comprised three stages:

1. Calculation of mean weekly milk production for each lactation group over a 44-week period.
2. Estimation of milk yield and modeling of lactation curves using the three mathematical models.
3. Evaluation of mathematical model performance based on the discrepancy between actual and estimated milk yields.

Statistical Analysis

Model accuracy was assessed by calculating the absolute error (in liters), relative error (percentage) and lactation curve between observed and predicted milk yield during 305 days (44 weeks). Descriptive statistics and comparative analysis were performed to determine the model with the lowest prediction error across lactation periods. All data analyses were conducted using Microsoft Excel

RESULT AND DISCUSSION

This study assessed three mathematical models in estimating total milk yield and lactation curve using the first 13 weeks milk production record at initial lactation. The models evaluated included: (1) incomplete gamma function, (2) orthogonal polynomial contrast model, and (3) non-linear regression model. The comparative analysis focused on model accuracy, consistency across parities, and sensitivity to variation in milk production patterns. Each model's strengths and limitations were examined in relation to their underlying mathematical structure and suitability for smallholder dairy systems. Detailed results for observed and predicted yields, including associated prediction errors, are presented in Table 2.

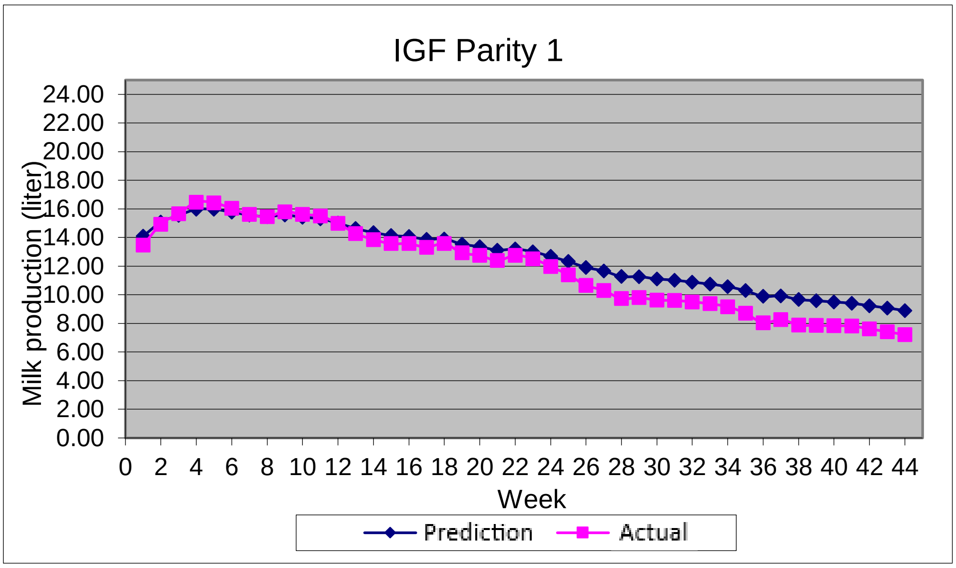
This study showed that each mathematical model in each parity has variation in milk production and lactation curve (Figure 1-3). The variation in estimated milk yield and lactation curve may result from several factors, including the values of the constants a, b, and c. The specific values assigned to these constants a, b, and c affected both the predicted milk production and the magnitude of the associated estimation bias (Val-Arreola et al., 2004; Rekik and Ben Gara, 2004; Dematawewa et al., 2007; Cole and Null, 2009).

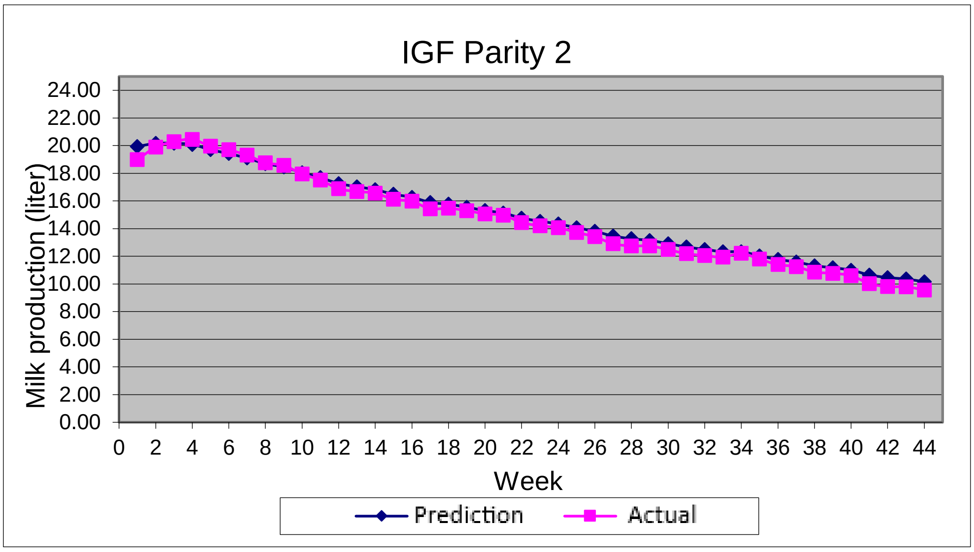
Parameter *a* reflects the overall scale of the lactation curve and generally increases with the number of parities, aligning with the animal's production level. Parameter *b* governs how quickly milk yield rises to reach its peak, while parameter *c* indicates the post-peak decline rate, with its magnitude typically increasing as parity increases. Additionally, Wood's model parameters can be employed to derive key lactation characteristics, including the time to peak yield (*tm*), the maximum yield (*ym*), and the persistency of lactation (*p*) (Nicolo et al., 2011).

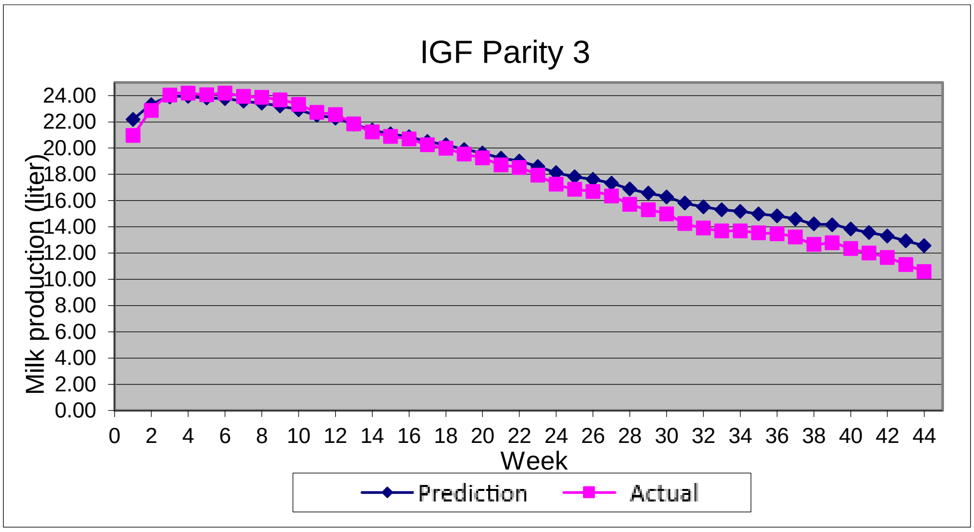
The incomplete gamma function model $y_x = ax^b e^{-cx}$, which is based on a biological representation of lactation dynamics, yielded and highest predicted accuracy with average prediction error of 274.67 liters, corresponding to 6.17% of the actual milk yield. The smallest deviation was observed in the second lactation period (89.44 L; 1.98%), while the largest occurred during the fourth period (558.20 L; 12.54%). (Table 2, Figure 1). These results indicate that the model performs relatively well when data are sufficiently dense.

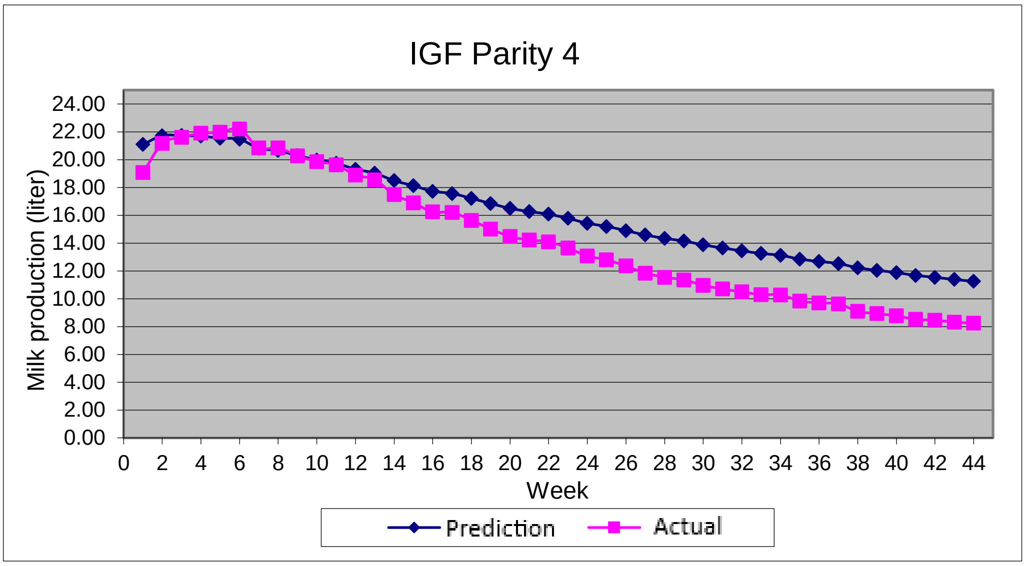
Tabel 2. Actual and estimated total milk yield for 305 day (44 weeks) with prediction errors across lactation periods using weekly data during the first 13 weeks of lactation under three mathematical models.

No	Mathematical model	Parity	n	Milk yield (lt)		Bias	
				Actual	Prediction	Liter	%
1.	$y_x = ax^b e^{-cx}$	1	8	3649.05	3898.06	249.01	6.82
		2	48	4513.70	4603.15	89.44	1.98
		3	48	5539.61	5757.19	217.58	3.93
		4	44	4451.20	5009.40	558.20	12.54
		5	12	4105.11	4364.24	259.13	6.31
Average				4451.73	4726.41	274.67	6.17
2.	$y_x = x(a + bx + cx^2)^{-1}$	1	8	3649.05	3937.83	288.78	7.91
		2	48	4513.70	4623.24	109.54	2.43
		3	48	5539.61	5796.46	256.85	4.64
		4	44	4451.20	5084.49	633.30	14.23
		5	12	4105.11	4438.80	333.70	8.13
Average				4451.73	4776.17	324.43	7.29
3.	$y_x = a - bx - ae^{-cx}$	1	8	3649.05	3909.76	260.71	7.14
		2	48	4513.70	4670.31	156.61	3.47
		3	48	5539.61	5813.66	274.05	4.95
		4	44	4451.20	5128.29	677.09	15.21
		5	12	4105.11	4467.98	362.88	8.84
Average				4451.73	4798.00	346.27	7.78


IGF Parity 1
Milk production (liter)
24.00
22.00
20.00
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44
Week
Prediction
Actual


IGF Parity 2
Milk production (liter)
24.00
22.00
20.00
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44
Week
Prediction
Actual


IGF Parity 3
Milk production (liter)
24.00
22.00
20.00
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44
Week
Prediction
Actual



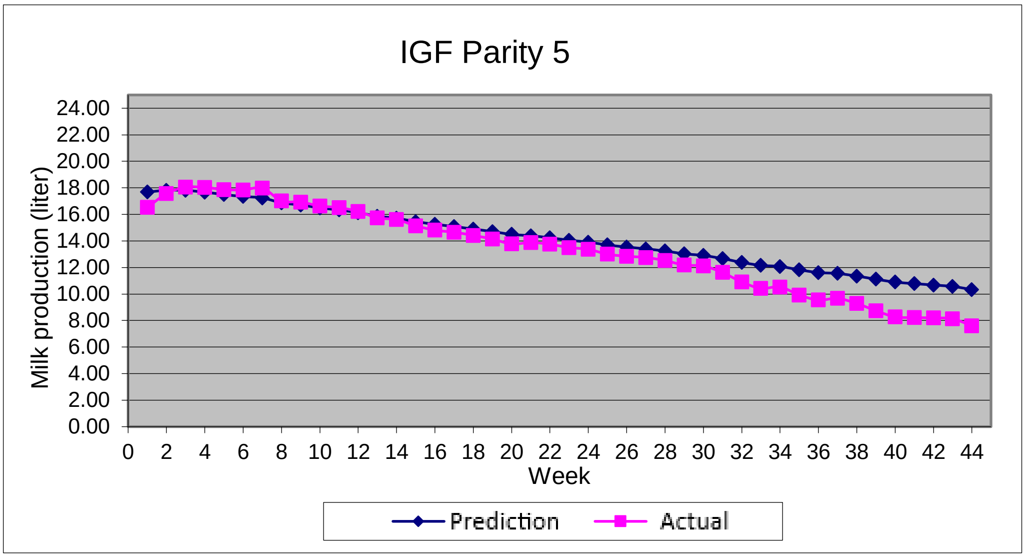
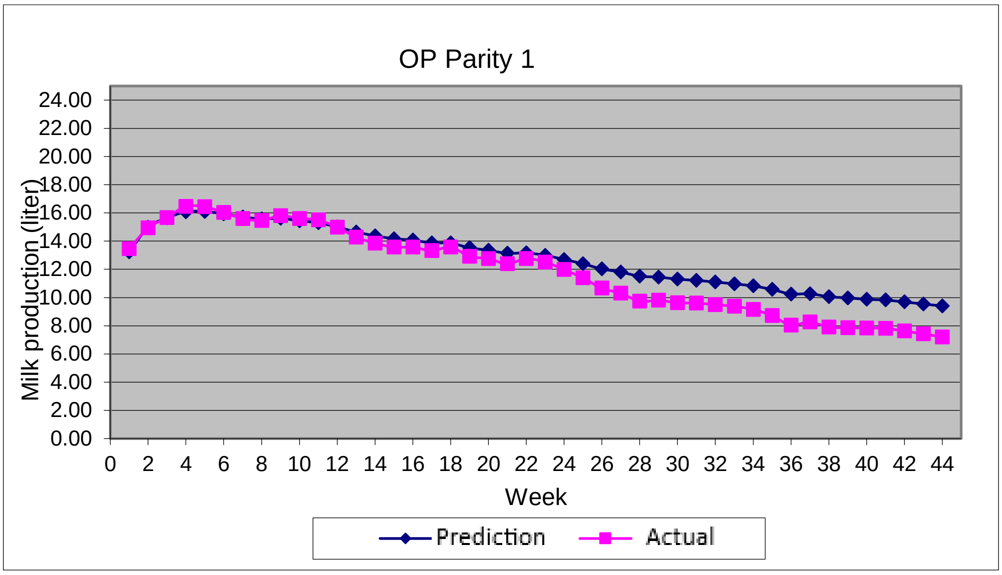
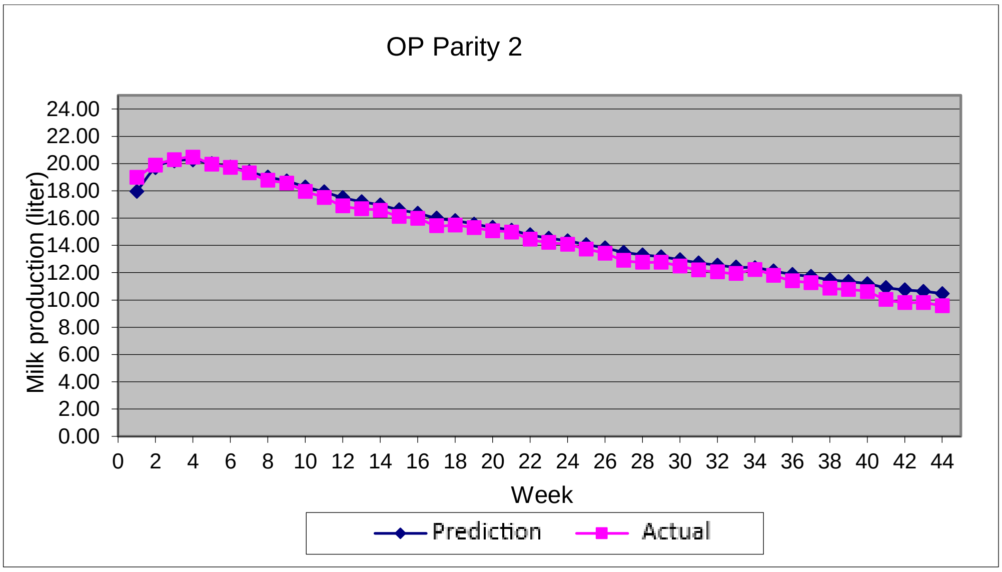


Figure 1. Prediction and actual milk yield, lactation curve based on weekly records using incomplete gamma function model $y_x = ax^b e^{-cx}$ (IGF: Incomplete Gamma Function) Parity 1, 2, 3, 4, 5

The incomplete gamma function model is widely recognized for its bell-shaped curve, which effectively mirrors the typical lactation trajectory, encompassing an initial rise, a peak, and a gradual decline (Wood, 1967). This structural alignment makes the model particularly effective when the observed data conform to the expected lactation pattern. Nevertheless, previous studies have highlighted its limitations; for instance, Wilmink (1987) and Gengler et al. (1997) reported that the model's performance deteriorates in the absence of a well-defined peak or in the presence of disruptions caused by health or nutritional stressors.

In the present study, the model appeared to underestimate milk yield in later stages of lactation, likely due to its intrinsic decay component a,b and c value, which may excessively weight the declining phase. Despite this limitation, the Gamma function model remains a useful tool for predictive purposes in structured dairy production systems where milk recording is conducted consistently. The orthogonal polynomial contrast model $y_x = x(a+bx+cx^2)^{-1}$ yielded a slightly higher average prediction error compared to the Gamma model, with a mean deviation of 324.43 liters, equivalent to 7.29% of actual yield. Although the model offered moderate overall accuracy, it exhibited greater fluctuation across lactation periods, with percentage errors ranging from 2.43% in parity 2 to 14.23% in parity 4 (Table 2, Figure 2). This variability highlights the model's sensitivity to changes in lactation dynamics.





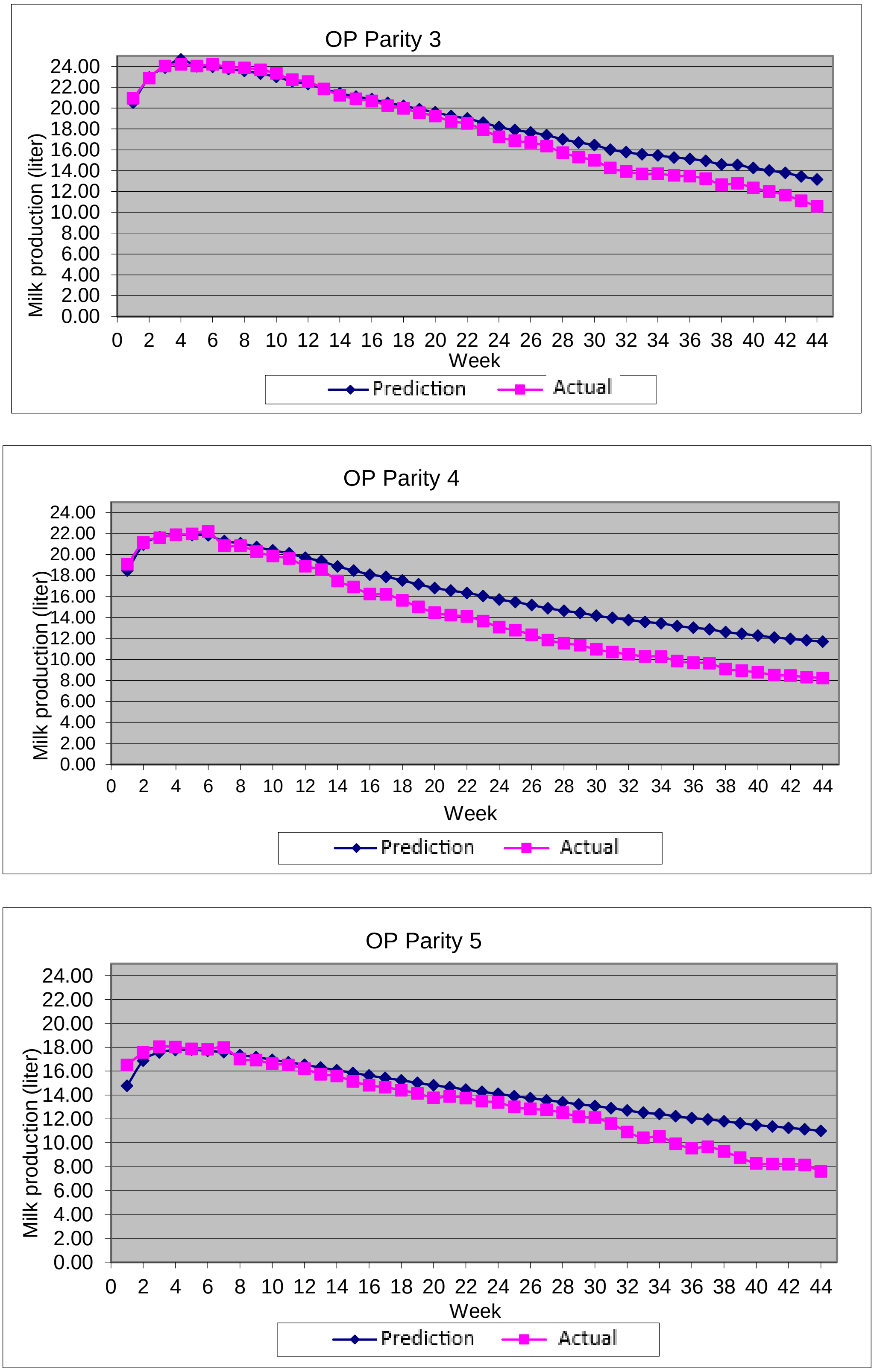
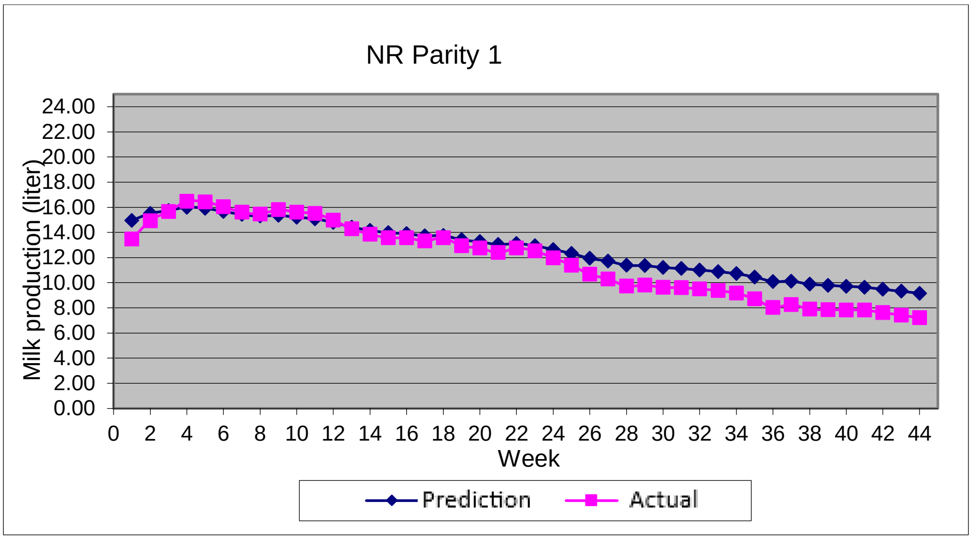


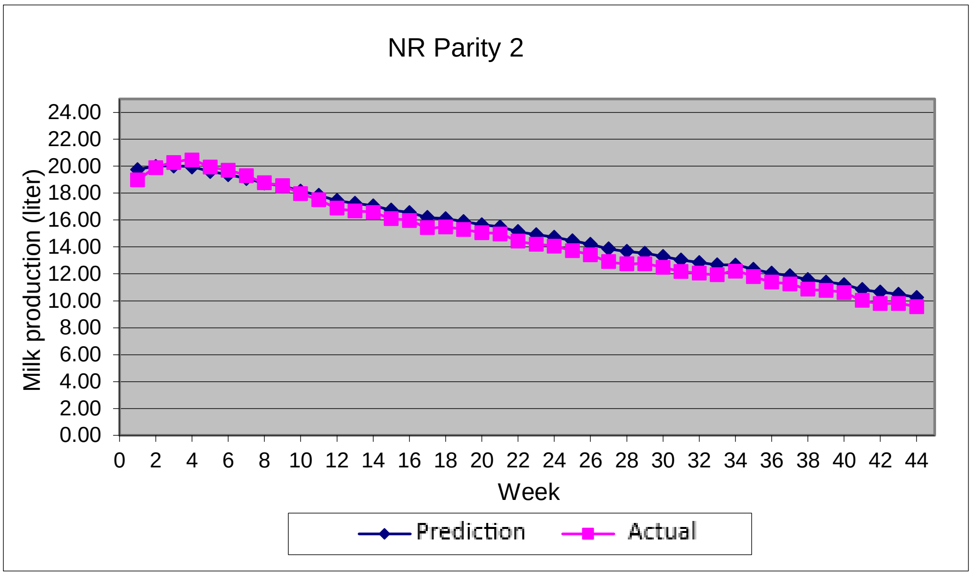
Figure 2. Prediction and actual milk yield, lactation curve based on weekly records using orthogonal polynomial contrast model $y_x = x(a + bx + cx^2)^{-1}$ (OP: Othogonal Polynomial Contrast) Parity 1, 2, 3, 4, 5

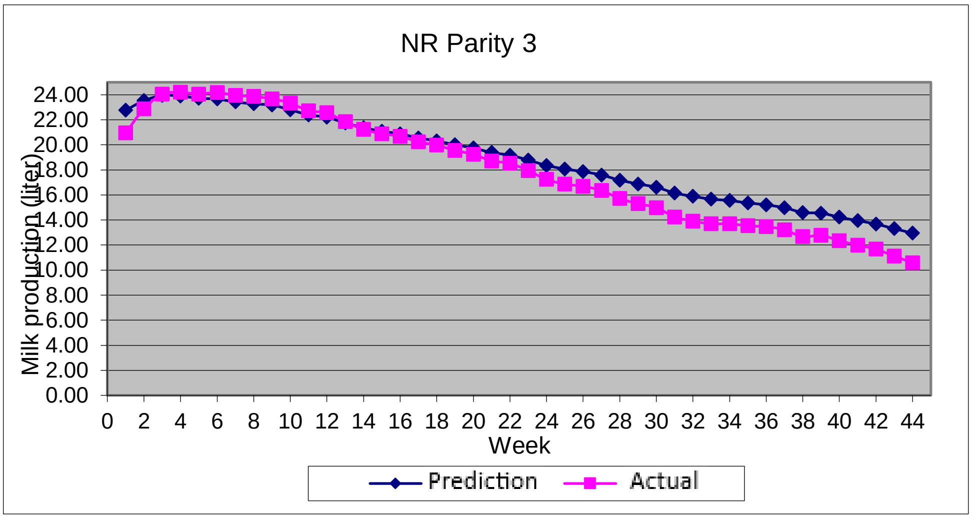
The polynomial formulation provides considerable flexibility in curve fitting; however, it may become unstable under conditions involving non-linear trends or abrupt shifts in production. A notable limitation of this model lies in its vulnerability to parameter instability, particularly when applied to high-frequency datasets where polynomial terms may capture noise rather than signal (Papajcsik & Bodero, 1988).

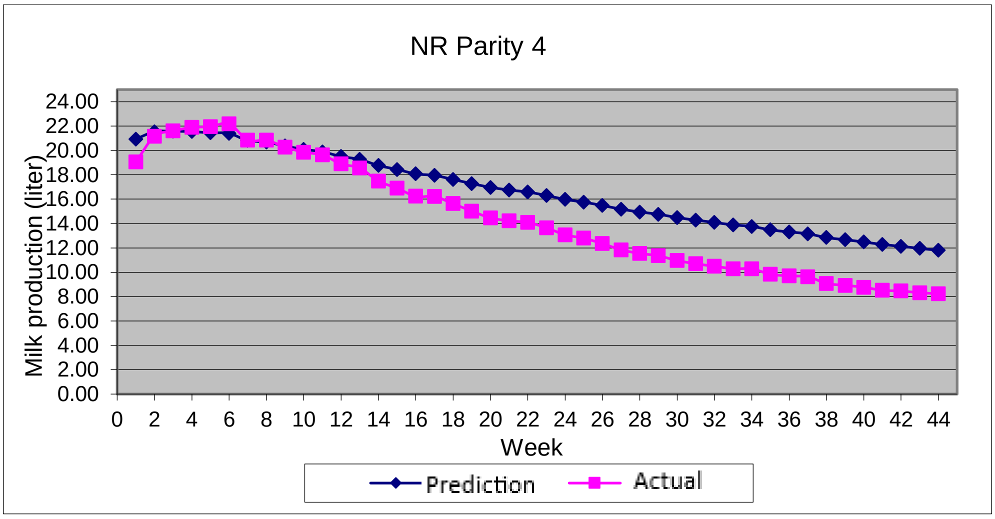
Despite this, its mathematical simplicity and ease of application make it suitable for use in field conditions characterized by moderate variability.The highest estimation error was observed in period 4, which may be attributed to transient disturbances such as environmental stressors or subclinical health issues. Compared to biologically driven models, the orthogonal polynomial contrast model appears to have a limited capacity to adapt to such anomalies, thereby constraining its predictive robustness in dynamic lactation environments.

Among the three models evaluated, the non-linear regression model $y_x = a - bx - ae^{-cx}$ demonstrated the lowest predictive accuracy, with a mean deviation of 346.27 liters or 7.78% (Table 2, Figure 3). Its accuracy decreases in periods characterized by greater yield variability. Although its average error slightly exceeded that of the incomplete gamma function model and orthogonal polynomial contrast, it exhibited greater consistency across lactation periods. This model integrates both linear and exponential decay terms, enabling it to capture gradual declines while simultaneously accommodating short-term fluctuations in milk yield.




NR Parity 2
Milk production (liter)
24.00
22.00
20.00
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44
Week
Prediction
Actual


NR Parity 3
Milk production (liter)
24.00
22.00
20.00
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44
Week
Prediction
Actual


NR Parity 4
Milk production (liter)
24.00
22.00
20.00
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44
Week
Prediction
Actual

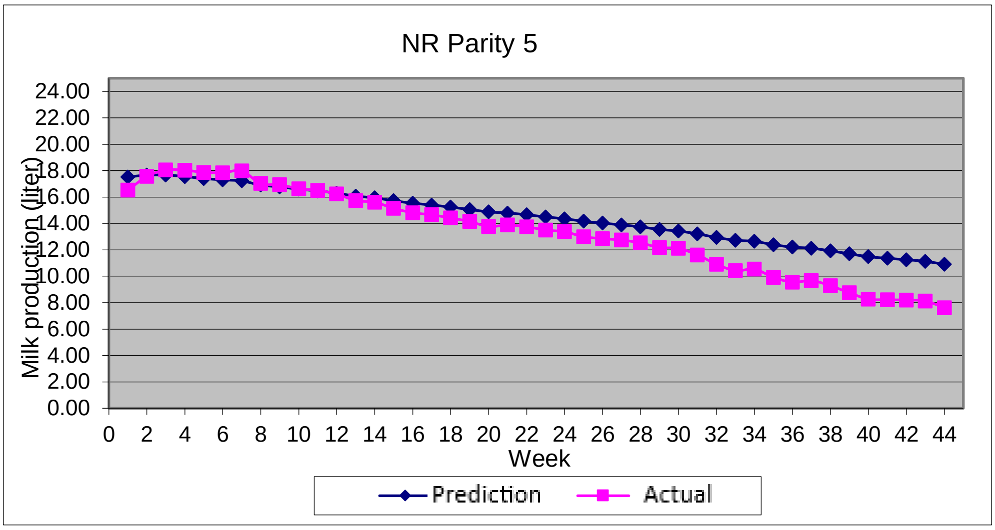


Figure 3. Prediction and actual milk yield, lactation curve based on weekly records using non-linear regression model $y_x = a - bx - ae^{-cx}$ (NR: non linear regression) Parity 1, 2, 3, 4, 5

The smallest prediction error was observed in the second lactation period (156.61 L; 3.47%), whereas the largest occurred during the fifth period (667.09 L; 15.21%). These results suggest that the non-linear regression model is not particularly well-suited for mid-lactation phases, where milk yield typically begins to transition from peak to decline particularly at 4 lactation. Even though, as noted by Guo and Swalve (1995); Montesinos, and Cham (2022), this hybrid modeling approach offers enhanced flexibility compared to single-component models.

However, the inclusion of multiple parameters also increases susceptibility to overfitting, particularly when applied to sparse or noisy datasets. In the present study, the use of weekly milk production data provided sufficient temporal resolution to ensure robust parameter estimation, thereby enhancing the model's overall predictive performance Comparing the three models, all demonstrated reasonably good predictive performance under weekly data collection conditions.

Incomplete gamma function model and orthogonal polynomial contrast models showed stronger alignment with biological lactation patterns, while the non-linear regression model offered a mathematically flexible but biologically less grounded alternative. The standard deviations of prediction errors would likely confirm that the incomplete gamma function model had lower variance, further supporting its robustness. Additionally, the percentage deviations among all models remained within 10%, affirming that weekly data collection has potency to improves model reliability.

CONCLUSION

Incomplete gamma function model $y_x = ax^b e^{-cx}$ demonstrated the highest level of predictive accuracy and stability, followed orthogonal polynomial contrast model $y_x = x(a + bx + cx^2)^{-1}$ and non-linear regression model $y_x = a - bx - ae^{-cx}$. Weekly milk recording data for 13 weeks at initial lactation can be used to estimated total milk yield and lactation curve for 305 days using Incomplete gamma function. These findings highlight the critical role of high-frequency milk yield recording and emphasize the importance of selecting appropriate lactation models to enhance the accuracy of dairy herd management and production forecasting.

REFERENCES

Cole, J. B., & Null, D. J. (2009). Genetic evaluation of lactation persistency for five breeds of dairy cattle. *Journal of Dairy Science, 92*(5), 2248–2258. https://doi.org/10.3168/jds.2008-1825

Dematawewa, C. M. B., Pearson, R. E., & VanRaden, P. M. (2007). Modeling extended lactations of Holsteins. *Journal of Dairy Science, 90*(8), 3924–3936. https://doi.org/10.3168/jds.2006-790

Everett, R. W., & Carter, H. W. (1968). Accuracy of test interval method of calculating Dairy Herd Improvement Association records. *Journal of Dairy Science, 51*(12), 1936–1941. https://doi.org/10.3168/jds.S0022-0302%2868%2964111-6

Gengler, N., Sneyers, M., & Burny, P. (1997). Test day model with repeated observations and explicit modeling of the environmental effects. *Interbull Bulletin, 15*, 60–65.

Guo, Z., & Swalve, H. H. (1995). Modeling of lactation curve as a sub-model in the evaluation of test day records. *Interbull Bulletin, 11*, 4–11.

Migose, S. A., van der Linden, A., Bebe, B. O., de Boer, I. J. M., & Oosting, S. J. (2019). Accuracy of estimates of milk production per lactation from limited test-day and recall data collected at smallholder dairy farms. *Livestock Science, 232*, 1–10.

Montesinos, L. O. A., & Cham, C. J. (2022). *Multivariate statistical machine learning methods for genomic prediction: Chapter 4: Overfitting, model tuning, and evaluation of prediction performance*. Springer. ISBN 978-3-030-89010-0.

Nicolò, P. P., Dimauro, C., Rassu, S. P. G., Steri, R., & Pulina, G. (2011). The mathematical description of lactation curves in dairy cattle. *Italian Journal of Animal Science, 10*, 213–233.

Oliveira, H. R. O., Campos, G. S., Lazaro, S. F., Jamrozik, J., Schinckel, A., & Brito, L. F. (2024). Phenotypic and genomic modeling of lactation curves: A longitudinal perspective. *JDS Communications, 5*, 241–246.

Papajcsik, I. A., & Bodero, J. (1988). Modeling lactation curves of Friesian cows in a subtropical climate. *Animal Production, 47*(2), 201–207.

Pereira, J. A. C., Suzuki, M., & Hagiya, K. (2001). Prediction of 305-day milk production from early records in dairy cattle using an empirical Bayes method. *Animal Bioscience*, 1511–1515.

Pica-Ciamarra, U., Baker, D., Morgan, N., Zezza, A., Azzarri, C., Ly, C., Nsiima, L., Nouala, S., Okello, P., & Sserugga, J. (2014). *Investing in the livestock sector: Why good numbers matter. A sourcebook for decision makers on how to improve livestock data*. World Bank. https://doi.org/10.1596/17830

Rana, E., Gupta, A. K., Singh, A., Ruhil, A. P., Yousuf, S., & Ete, G. (2021). Prediction of first lactation 305-day milk yield based on bimonthly test-day milk-yield records in Murrah buffaloes. *Indian Journal of Animal Research*, 486–490.

Ranzato, G., Aernouts, B., Lora, I., Adriaens, I., Ben Abdelkrim, A., Gote, M. J., & Coz, G. (2024). Comparison of three mathematical models to estimate lactation performance in dairy cows. *Journal of Dairy Science, 107*, 6888–6901.

Rekik, B., & Ben Gara, A. (2004). Factors affecting the occurrence of atypical lactations for Holstein-Friesian cows. *Livestock Production Science, 87*, 245–250.

Sudarwati, H. (1990). *Studi terhadap pendugaan model kurva laktasi pada sapi perah* (Tesis, Pascasarjana, Institut Pertanian Bogor).

Val-Arreola, D., Kebreab, E., Dijkstra, J., & France, J. (2004). Study of the lactation curve in dairy cattle on farms in central Mexico. *Journal of Dairy Science, 87*, 3789–3799.

Wilmink, J. B. M. (1987). Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livestock Production Science, 16*(4), 335–348.

Wood, P. D. P. (1967). Algebraic model of the lactation curve in cattle. *Nature, 216*(5111), 164–165.