Estimation of lactation curves of Gyr cattle and some associated production parameters in the Colombian low tropic

Estimación de Curvas de lactancia de bovinos Gyr y algunos parámetros asociados en el trópico bajo Colombiano

Estimação de curvas de lactação de bovinos Gyr e alguns parâmetros associados no trópico baixo da Colômbia

Daniela Ferro[1] iD; Jhon Gil[1] iD; Ariel Jiménez[2] iD; Carlos Manrique[3] iD; Carlos A Martínez[4]* iD.

[1]*Escuela de Matemáticas y Estadística, Universidad Pedagógica y Tecnológica de Colombia, Facultad Seccional Duitama, Colombia.*

[2]*Asociación Colombiana de Criadores de Ganado Cebú ASOCEBU, Bogotá, Colombia.*

[3]*Departamento de Producción Animal, Universidad Nacional de Colombia, Sede Bogotá, Colombia.*

[4]*Dirección de Investigación y Desarrollo, Corporación Colombiana de Investigación Agropecuaria - AGROSAVIA, Sede Central, Mosquera, Colombia.*

Abstract

Background: The Gyr breed is widely used in Colombian low tropic dairy production systems. During the last 10 years, the Asociación Colombiana de Criadores de Ganado Cebú† - ASOCEBU, has been leading a dairy milk control program which led to the creation of a dataset that permits to carry out the first analysis of milk yield in Gyr cattle in the country using records from several herds. **Objectives:** To study milk production dynamics of Gyr cattle in the Colombian low tropic through the estimation of lactation curves and four derived production parameters: total milk yield between 5 and 305 days (TMY305), peak milk yield (PMY), days at peak (DP) and persistency (P). **Methods:** 13,798 daily milk yield records from 1,510 cows performing in 103 herds were used; the total number of lactations was 2,480. Four models were considered: Wood, Wiltmink, Papajcsik & Bordero, and a second-degree polynomial. Mean square error, mean absolute error, mean square error of prediction, Akaike and Bayesian information criteria were used to select the model better describing each lactation using the majority rule, that is, the model selected by most criteria was the chosen one. The shape of each fitted lactation curve was checked using basic results from calculus which permitted the classification of the estimated curves into two groups: typical and atypical; only typical functions were used to compute the four aforementioned production parameters. **Results:** The second-order polynomial was the model most frequently selected, while the Papajcsik & Bordero model had the lowest frequency. Average TMY305, PMY, DP and P were 3,489.86 kg, 17.28 kg, 57.17 days, and 0.83, respectively, with coefficients of variation: 0.27, 0.21, 0.41, and 0.16. **Conclusions:** This study permitted to identify individuals with outstanding phenotypic performance. To the best of our knowledge, this is the first study of this kind involving thousands of lactations from Gyr cows performing in several regions of Colombian low tropic.

Resumen

Antecedentes: La raza Gyr es ampliamente utilizada en las lecherías de trópico bajo en Colombia. Durante los últimos 10 años, la Asociación Colombiana de Criadores de Ganado Cebú – ASOCEBU, ha liderado un programa de control lechero que generó una base de datos que permite llevar a cabo el primer análisis de producción de leche de la raza Gyr en el país, considerando individuos de varias fincas. **Objetivos:** Caracterizar la dinámica de producción de leche de ganado Gyr en el trópico bajo colombiano mediante la estimación de curvas de lactancia y cuatro parámetros de producción derivados: producción total entre 5 y 305 días (TMY305), pico de lactancia (PMY), días al pico (DP) y persistencia (P). **Métodos:** Se utilizaron 13.798 registros de producción diaria de leche de 1.510 vacas provenientes de 103 fincas, el total de lactancias fue 2.480. Se consideraron cuatro modelos: Wood, Wiltmink, Papajcsik & Bordero, y un polinomio de segundo grado. Los criterios usados para elegir el modelo que mejor describió cada lactancia fueron: error cuadrático medio, error absoluto medio, error cuadrático medio de predicción, criterio de información de Akaike y criterio de información Bayesiano. Se utilizó el criterio de mayoría, esto es, el modelo seleccionado fue aquel elegido por más criterios. La forma de cada una de las curvas de lactancia estimadas fue chequeada utilizando resultados básicos del cálculo, esto permitió clasificar las curvas estimadas en dos grupos: típicas y atípicas; solamente las curvas típicas fueron empleadas para calcular los cuatro parámetros antes mencionados. **Resultados:** El polinomio de segundo grado fue el modelo que se seleccionó con mayor frecuencia, mientras que el modelo Papajcsik & Bordero tuvo la menor frecuencia. Los promedios para TMY305, PMY, DP y P fueron 3.489,86 kg, 17,28 kg, 57,17 días, y 0,83, respectivamente, con coeficientes de variación 0,27, 0,21, 0,41 y 0,16. **Conclusiones:** Este estudio permitió identificar individuos con desempeño fenotípico sobresaliente. De acuerdo al estado del arte, este es el primer estudio de este tipo que considera miles de lactancias de vacas Gyr provenientes de varias regiones del trópico bajo colombiano.

Palabras clave: <u>*Bos*</u> <u>*indicus*</u>; *bovinos; comparación de modelos; forma de la función de lactancia; ganado; lactancia; lecherías tropicales; modelamiento estadístico; modelos lineales, modelos no lineales; parámetros de lactancia; producción de leche; rendimiento lechero; vacas.*

Resumo

Antecedentes: Gir é uma raça *Bos indicus* amplamente utilizada em sistemas de produção leiteira no trópico baixo Colombiano. A Asociación Colombiana de Criadores de Ganado Cebú- ASOCEBU, lidera um programa de controle de leite nos últimos 10 anos, o que permitiu à coleta de um conjunto de dados para realizar a primeira análise de produção de leite em bovinos Gir no país com informações de vários rebanhos. **Objetivo:** Estudar a dinâmica da produção de leite por meio da estimativa das curvas de lactação e quatro parâmetros de produção derivados: produção total de leite entre 5 e 305 dias (TMY305), produção de leite no pico de lactação (PMY), dias em produção de leite no pico de lactação (DP) e persistência (P). **Métodos:** Foram utilizados 13.798 registros de produção diária de leite de 1.510 vacas de 103 fazendas , totalizando 2.480 lactações. Foram considerados quatro modelos: Wood, Wiltmink, Papajcsik & Bordero e um polinômio de segundo grau. Para selecionar o modelo que melhor descreve cada lactação foram utilizados os seguintes critérios: erro quadrado médio, erro absoluto médio, e erro quadrado médio de predição. O modelo selecionado pela maioria dos parâmetros, de acordo com os critérios de informação de Akaike e Bayesiano, foi o escolhido. A forma de cada curva de lactação ajustada foi verificada utilizando os resultados básicos do cálculo, isso permitiu classificar as curvas estimadas em dois grupos: típico e atípico, e apenas funções típicas foram utilizadas para calcular os quatro parâmetros de produção acima mencionados. **Resultados:** O polinômio de segunda ordem foi o modelo mais frequentemente selecionado, enquanto o modelo Papajcsik & Bordero apresentou a menor frequência. A média de TMY305, PMY, DP e P foram 3.489,86 kg, 17,28 kg, 57,17 d e 0,83, respectivamente, com coeficientes de variação de 0,27, 0,21, 0,41 e 0,16. **Conclusões:** Este estudo permitiu identificar indivíduos com excelente desempenho fenotípico. De acordo com a literatura atual, este é o primeiro estudo envolvendo milhares de lactações de vacas Gir em várias regiões do trópico baixo colombiano.

Palavras-chave: <u>*Bos*</u> <u>*indicus*</u>; *bovinos; comparação de modelos; forma da função de lactação; gado; lactação; laticínios tropicais; modelagem estatística; modelos lineares; parâmetros de lactação; produção de leite; modelos não lineares; vacas.*

Introduction

Gyr is a *Bos indicus* breed widely used in the Colombian low tropic dairy production systems. Originally from India, it can be found in Brazil and other Latin American countries (Quiroz *et al.*, 2015). This breed performs well in regions with high temperatures and features resistance to tropical diseases. Although in certain regions it has gained popularity as a dual-purpose breed, Colombian breeders have focused on milk production (Mellado *et al.*, 2011). Thus, due to its ability to produce milk under the harsh conditions of low tropic ecosystems, Gyr germplasm can be found in several regions of Colombia and, consequently, it has a significant impact on Colombian low tropic dairy operations. Therefore, it is relevant to study the milk yield dynamics of this breed under Colombian conditions. In addition to the characterization of milk yield trajectory, estimating lactation curves allows to predict total yield and conduct analyses to identify individuals with outstanding phenotypic performance (Cañas *et al.*, 2012). Moreover, lactation curves allow farmers to make management decisions. For example, designing nutritional plans and defining when cows should be dried off. It is worth mentioning that some of the most important production parameters derived from lactation curves are: total milk yield between two given time points -typically from the beginning of lactation to 305 days in milk (DIM)-, peak milk yield (PMY), DIM at PMY (DP), and persistence. These parameters allow breeders to make better management decisions (López *et al.*, 2015).

During the last 10 years, the Asociación Colombiana de Criadores de Ganado Cebú (ASOCEBU) has been leading a milk control program which led to the creation of a dataset that permits to carry out the first analysis of milk yield dynamics of Gyr cattle in the country using records from several herds (ASOCEBU, n.d.).

Therefore, the objective of this study was to characterize milk production dynamics of Gyr cattle in the Colombian low tropic by fitting linear and nonlinear statistical models.

Materials and methods

Records

Data were collected by trained staff from ASOCEBU between 2008 and 2018 in 130 herds located in low tropic regions of Colombia. These herds were distributed in 65 municipalities from 22 Departments. In herds with two milkings per day, the two partial yields were recorded. After discarding records corresponding to more than 400 DIM and lactations with less than four records, the resulting dataset contained 13,798 records corresponding to 2,480 lactations from 1,510 cows. Parity ranged from 1 to 13. The distribution of records according to parity is shown in Figure 1.
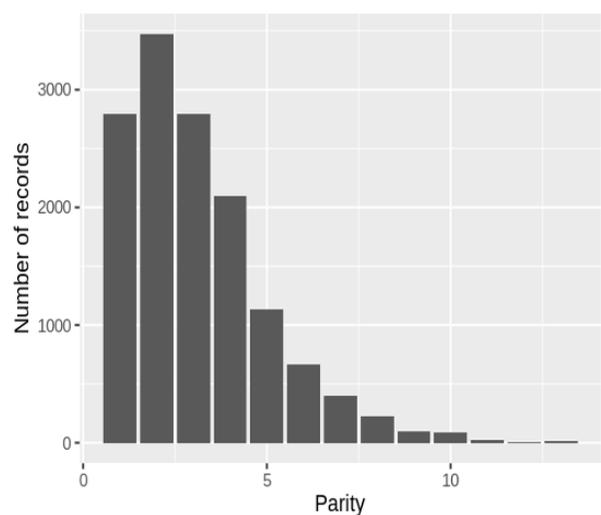
Figure 1. Distribution of records according to parity.

Data analysis

Estimation of lactation Curves and associated parameters. Four statistical models (two linear and two nonlinear) were fitted to each individual lactation. The linear models were a second-order polynomial (Rawlings *et al.*, 1998), and Wiltmink (Wiltmink, 1987), while the nonlinear models were Wood (Wood, 1967), and the Papajcsik and Bordero (PB) model (Papajcsik and Bordero, 1988). Model parameters were estimated using ordinary least squares for linear models, and non-linear least squares for non-linear models (Rawlings *et al.*, 1998). Normal equations corresponding to non-linear models were minimized using the Gauss-Newton algorithm.

For each lactation, these models were compared using five criteria: mean square error (MSE) as defined in Tedeschi (2004), mean absolute error (MAE), mean square prediction error (MSEP), Bayesian information criterion (BIC; Schwarz, 1987), and Akaike's information criterion (AIC; Akaike, 1973). Analla (1998) proposed the MSE as a criterion to compare statistical models based on the residuals of the linear regression of y_i on $\hat{y}_i$, denoted as ε_i^*, and it is computed as follows:

$$MSE = \frac{\sum_{i=1}^{n} \varepsilon_i^{*2}}{n-2} \qquad (1)$$

On the other hand, the MAE measures the average absolute difference between observed and predicted values (Tedeschi, 2004), that is,

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} \qquad (2)$$

The MSEP is computed as the average squared difference between observed and predicted values (Tedeschi, 2004).

$$MSEP = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n} \qquad (3)$$

Finally, the BIC and AIC are based on two components, one indicates goodness of fit and the other one can be thought of as a penalty for the number of parameters (AIC; Akaike, 1974) or the number of parameters and sample size (BIC):

$$AIC = -2 \log (\hat{L}) + 2k \qquad (4)$$
$$BIC = -2 \log (\hat{L}) + \log(n)\, k \qquad (5)$$

where $\hat{L}$ is the likelihood evaluated at the maximum likelihood estimator, i.e., the maximum value of the likelihood function, k is the number free parameters in the model and n is sample size. When using these criteria, models with smaller values are preferred.

The four models were fitted and the five criteria to compare these models were computed for each lactation. Subsequently, the majority criterion was used to select a model for each lactation; that is, the model selected by most criteria was the chosen one. For example, assuming that for a given lactation, MSE, MAE and MSEP preferred the Wood model, BIC suggests the Wiltmink model and AIC the quadratic model, then the Wood model would be selected.

Once the model that best fit the data from each lactation was selected, curves yielding negative predicted values at some point within the interval [5, 305] were discarded. The next step was checking the shape of the estimated lactation curve. This step was necessary to detect curves that did not have a typical shape; that is, curves that were convex (or concave upward), strictly increasing or changed from convex to concave in the set of positive reals; for example, those having a minimum and an inflection point.

Consequently, using estimated parameters from the selected model for each lactation, the shape of the corresponding function was evaluated utilizing results from calculus. Basically, these results led to find conditions on the signs of estimated parameters which permitted defining the shape of the function (Steward, 2008). The shape checking step can be summarized as follows. Shape of a differentiable function can be determined using basic results from calculus. For example, a concave function has a strictly decreasing first derivative, i.e., its second derivative is always negative. For Wood's model, if β_1 and β_2 have the same sign, its systematic component corresponds to a function that has a stationary point in the set of positive real numbers; therefore, the limit when the independent variable tends to zero from the right permits to determine whether the function has a typical shape or not.

This analysis reveals that if β_1 is negative, the function is atypical; this result also applies when β_2 is positive. Moreover, when β_1 is positive and β_2 is negative, the function is strictly increasing; that is, it shows an atypical shape. On the other hand, for Papajcsik & Bordero's model, if β_1 is positive the function has a saddle point at $2/\beta_1$ and a global maximum is located at $1/\beta_1$; consequently, it shows a typical shape. If β_1 is negative the function has a global minimum in the set of negative real numbers and is strictly increasing in the set of positive real numbers; therefore, in this case Papajcsik & Bordero's model is not useful for modelling milk production. Moreover, the systematic components of the two linear models correspond to functions

without saddle points; hence, they are concave or convex in their entire domain; if β_2 is negative, the functions are concave. For Wiltmink's model, it is easy to show that if $\beta_1 > 0$ and $\beta_2 < 0$, the function is strictly increasing; thus, it is not useful for modelling milk yield and corresponds to an atypical curve.

In the following step the four production parameters of interest were computed for typical (concave, having a shape change from concave to convex in the set of positive real numbers or strictly decreasing) estimated lactation curves. The PMY corresponds to the maximum of the lactation function, while DP is the corresponding maximum argument or argmax, the point in the function's domain where it is maximized. On the other hand, total milk yield between five and 305 DIM (TMY305) amounts to the definite integral of the lactation curve with these two time points as lower and upper bounds, respectively; that is,

$$TMY305 = \int_5^{305} f(t)\, dt \qquad (6)$$

Notice that when computing TMY305 the first four days are discarded due to colostrum production. Persistency may be computed in several ways but, in general, it is a term referring to the behavior of the curve after PMY; thus, a persistent lactation is one in which the rate of decline after PMY is small (Yanar and Güler, 2009). Here, persistency was computed as defined by Johansson and Hansson (1940) who proposed to use the ratio between total milk production during the second 100 days and the first 100 days as a measure of persistency. Consequently, this parameter was computed as the ratio of two definite integrals. Table 1 shows the functional forms of each of the four models considered in this study and the corresponding associated production functions.

It is important to notice that the domain of the lactation curves is a subset of the set of positive real numbers, hence, PMY and DP are defined in this set; as a consequence, when the global maximum of the estimated curve is attained at a negative value of DIM (according to the expressions shown in Table 1), the estimated DP is zero and the estimated PMY is equal to the corresponding value of the function evaluated at 0.

The integrals required to compute TMY305 and persistency in the Wood's model were numerically evaluated using the Newton method because they do not have a closed form solution.

Results

We found that 1,382 lactations had a typical shape, which corresponded to 55.73% of the total number of lactations (2,480). Moreover, among curves with atypical shapes,

Table 1. Functional forms of the fitted models and the corresponding associated production parameters.

Model	Functional form	TMY305	PMY	DP	Persistency
WOOD	$E[y_i] = \beta_0 t_i^{\beta_1} e^{(-\beta_2 t_i)}$	No closed form	$\beta_0 \left(\dfrac{\beta_1}{\beta_2}\right)^{\beta_1} e^{-\beta_1}$	$\dfrac{\beta_1}{\beta_2}$	No closed form $\dfrac{\int_{105}^{205} \beta_0 t_i^{\beta_1} e^{(-\beta_2 t_i)}\, dt}{\int_{205}^{305} \beta_0 t_i^{\beta_1} e^{(-\beta_2 t_i)}\, dt}$
PB	$E[y_i] = \beta_0 t_i e^{(-\beta_1 t_i)}$	$\dfrac{\beta_0}{\beta_1^{2}}[e^{-5\beta_1}(5\beta_1 + 1) - e^{-305\beta_1}(305\beta_1 + 1)]$	$\dfrac{\beta_0}{\beta_1} e^{-1}$	$\dfrac{1}{\beta_1}$	$\dfrac{e^{-105\beta_1}(5\beta_1 + 1) - e^{-205\beta_1}(305\beta_1 + 1)}{e^{-205\beta_1}(5\beta_1 + 1) - e^{-305\beta_1}(305\beta_1 + 1)}$
WILTM	$E[y_i] = \beta_0 + \beta_1 t_i + \beta_2 e^{\left(-\frac{t_i}{20}\right)}$	$300[\beta_0 + 155\beta_1] + 20\beta_2 \left[e^{-\frac{1}{4}} - e^{-\frac{61}{4}}\right]$	$\beta_0 + 20\beta_1 \left(1 - \ln\left[\dfrac{20\beta_1}{\beta_2}\right]\right)$	$-20 \ln\left[\dfrac{20\beta_1}{\beta_2}\right]$	$\dfrac{15[\beta_0 + 155\beta_1] + \beta_2\left[e^{-\frac{21}{4}} - e^{-\frac{41}{4}}\right]}{15[\beta_0 + 255\beta_1] + \beta_2\left[e^{-\frac{41}{4}} - e^{-\frac{61}{4}}\right]}$
POL2	$E[y_i] = \beta_0 + \beta_1 t_i + \beta_2 t_i^2$	$300[\beta_0 + 155\beta_1 + 31525\beta_2]$	$\beta_0 - \dfrac{\beta_1^2}{4\beta_2}$	$-\dfrac{\beta_1}{2\beta_2}$	$\dfrac{\beta_0 + 155\frac{\beta_1}{2} + 74575\frac{\beta_2}{3}}{\beta_0 + 255\frac{\beta_1}{2} + 197575\frac{\beta_2}{3}}$

TMY305: Total milk yield between 5 and 305 days; PMY: Peak milk yield; DP: Days at peak; PB: Papajcsik and Bordero model; POL2: Second-order polynomial model; WILTM: Wiltmink model.

the POL2 model was the most frequent, while the Wood model had the smallest frequency; however, the frequency of estimated atypical lactation curves from the Wiltmink's model was nearly identical to the one from the POL2 model (Table 2).

Table 2. Percentages (relative to 2,480 lactations) of typical (PTFL), atypical (PAFL), and total lactations fitted (PFL) by each model.

Model	PTFL	PAFL	PFL
PB	3.47%	0%	3.47%
POL2	21.61%	19.96%	41.57%
WILM	15.40%	19.96%	35.36%
WOOD	15.24%	4.35%	19.60%
Overall	**55.73%**	**44.27%**	**100%**

PTFL: Percentages of typical curves; PAFL: Percentages of atypical curves; PFL: Total lactations fitted; PB: Papajcsik and Bordero model; POL2: Second-order polynomial model; WILTM: Wiltmink model.

Table 3 shows the frequencies of curves with and without negative estimated values relative to the 1382 curves without shape problems for each one of the fitted models.

Table 3. Frequencies (relative to 1,382 lactations) of curves with and without negative estimated values.

Model	Negative	Positive
PB	0%	6.22%
POL2	2.03%	36.76%
WILTM	3.40%	24.24%
WOOD	0%	27.35%
Overall	**5.43%**	**94.57%**

PB: Papajcsik and Bordero model; POL2: Second-order polynomial model; WILTM: Wiltmink model.

Thus, 5.43% of the typical lactations yielded negative estimates, while 94.57% of them were strictly positive (1307 lactations). As to the selected models, among typical lactations with positive estimated milk yield values the POL 2 model exhibited the best fit for most of the lactations, followed by Wood's model, Wiltmink's model, and the PB model.

Table 4 presents the frequency of selection of each model by parity group computed from the 1307 lactations with typical shapes and positive estimated values. The second-order polynomial was the most frequently selected model across all lactation groups.

Table 4. Frequencies of selected models by parity group.

Parity group	PB %	POL2 %	WILTM %	WOOD %
1	6.7	37.3	34	22
2	5.5	43.9	28.5	22.1
3 – 5	5.6	37.7	25.2	31.5
6-13	9	34.3	24.7	31.9

PB: Papajcsik and Bordero model; POL2: Second-order polynomial model; WILtM: Wiltmink model.

Subsequently, the 1,307 estimated lactation curves were used to estimate the four production parameters of interest. The overall averages of the four estimated parameters were: PMY: 17.28 kg, DP: 57.17 days, TMY305: 3489.86 kg, persistence: 0.83 (Table 5).

Table 5. Descriptive statistics for the 1,307 estimated lactation curves of Gyr cattle performing in Colombian low tropic.

Descriptive Statistics	TMY305	DP	PMY	Persistency
Minimum	1530.39	0.30	12.01	0.46
Maximum	6657.63	109.44	30.44	1.22
Average	3489.86	57.17	17.28	0.83
Standard deviation	958.21	23.41	3.68	0.13
Coefficient of variation	0.27	0.41	0.21	0.16

TMY305: Total milk yield between 5 and 305 days; PMY: Peak milk yield; DP: Days at peak.

Table 6 shows some descriptive statistics of the four parameters of interest for each of the following groups: first lactation, second lactation, third to fifth lactation and six or more lactations.

According to these results, the highest averages for TMY305, DP and PMY were found in the third to fifth lactation group, and the highest average for P was found in the first lactation group.

Table 6. Descriptive statistics of estimated lactation curves for the first lactation, second lactation, third to fifth lactation and six or more lactation groups of Gyr cattle performing in Colombian low tropic.

Parity group	Statistic	TMY305	DP	PMY	Persistency
1	Minimum	1,530.39	11.13	12.01	0.52
	Maximum	5,362.45	109.44	28.13	1.22
	Average	3,187.98	58.92	15.80	0.85
	Standard deviation	803.48	19.92	3.58	0.13
	Coefficient of variation	0.25	0.34	0.23	0.16
2	Minimum	1,930.03	6.34	12.10	0.49
	Maximum	6,359.38	104.84	26.48	1.20
	Average	3,520.89	56.49	16.98	0.83
	Standard deviation	973.58	22.89	3.57	0.14
	Coefficient of variation	0.28	0.41	0.21	0.16
3-5	Minimum	1531.59	0.30	12.16	0.52
	Maximum	6,657.63	109.33	30.44	1.22
	Average	3,667.67	57.15	18.04	0.84
	Standard deviation	983.66	25.08	3.66	0.13
	Coefficient of variation	0.27	0.44	0.20	0.16
6-13	Minimum	1,614.24	11.51	12.21	0.46
	Maximum	5,160.98	108.29	29.78	1.05
	Average	3,089.92	55.87	16.99	0.80
	Standard deviation	820.88	22.68	3.43	0.13
	Coefficient of variation	0.27	0.41	0.20	0.17

TMY305: Total milk yield between 5 and 305 days; PMY: Peak milk yield; DP: Days at peak.

Discussion

In this study, milk yield records from a Gyr cattle population performing in the Colombian low tropic were used to study milk yield dynamics through the estimation of lactation curves and four production parameters derived from it. To the best of our knowledge, this is the first study involving thousands of lactations from individuals performing in several regions of Colombian low tropic in this particular breed. Four models were fitted to each lactation and five criteria were used to choose the model with the best fit.

As pointed out before, lactation curves of Gyr cattle have not been studied in Colombia. On the other hand, several studies have been conducted in Brazil. For example, Pereira *et al.* (2012) studied lactation curves in first parity Gyr cows in Brazil. Those researchers focused on comparing several measures of persistency in order to find one well-suited for genetic evaluation. However, their study aimed to predict breeding values for these measures of persistency and their modelling approach was based on random regression models.

The average persistency suggests that, in general, cows from this population are good at keeping the productivity level after PMY; the overall mean of persistency was close to the unit (Table 5). This is an indication of a good productive performance of animals from this population which can be a consequence of genetic, management and environmental factors.

Milk yield is expected to increase after parturition until reaching the peak milk yield and then decrease until the end of lactation; this is a typical behavior. Thus, milk yield dynamics is expected to be described by functions showing this trajectory. Hence, a concave function or a function changing from concave to convex may describe such shape as far as they meet the condition of having a maximum within a certain subset of the set of positive real numbers. This subset corresponds to the range where peak milk yield is usually attained. For example, a function having a maximum at 250 days would not be useful because it does not describe the usual behavior. However, curves not following this trajectory (atypical curves) are frequently found. In the Colombian low tropic, it has been found that certain individuals do not show a peak, milk yield is described by a strictly

decreasing function (Ossa *et al.*, 1997); thus, this type of curve was considered typical in this study. As pointed out by Macciota *et al.* (2003) when analyzing atypical curves, it is important to determine if they are the result of some biological phenomena or data-specific features such as missing data. Hence, studying these curves permits to find errors in data as well as individuals with extreme or erratic performance. Further, these individuals could be used to detect management conditions that induce these lactation trajectories.

Regarding data problems that cause atypical curves, a common one is the lack of records during the first weeks of lactation. So far, this problem has not been studied in depth; hence, there is not a threshold establishing an upper limit for the time at which the first record must be collected (e.g., at most 40 DIM). Days at first record is not always used as a criterion to edit datasets to be used for lactation curve estimation; when doing so, the threshold is defined ad hoc, and it is set around 40 or 45 days. For instance, Pereira *et al.* (2012) discarded lactations not having the first record within 45 days postpartum. Here, it was found that 74% of atypical curves did not meet the condition of having a record within 45 DIM.

As to the factors inducing atypical lactation curves, it has been found that these can be the result of environmental and management conditions of tropical dairy systems with low individual production levels using Bos taurus-Bos indicus crossbred cattle (Vinay-Vadillo *et al.*, 2012). Notwithstanding, this kind of curves is not exclusive of these systems because they have also been found in dairy operations located in temperate regions using specialized breeds (Macciota *et al.*, 2003). Tekerli *et al.* (2000) studied factors affecting the shape of lactation curves in Holstein cattle in Turkey. They found that farm, calving year and season, parity, and interval between calving and conception were associated with curve shape. Therefore, results from this research may be used for further studies concerning lactation dynamics in Colombian Gyr cattle.

Problems with data may also cause negative predictions. Such predictions do not occur in the observed DIM interval, but when extrapolating too far from its limits. Thus, negative predictions for this lactation stage may be found because of lack of records in early lactation. A similar situation can be found with unfinished lactations when projecting total yield to 305 DIM. However, in these cases, useful inferences can be performed within the range of observed records. For instance, if the DIM range [85, 170] was available, then PMY and DP could be estimated. In addition, this occurs with linear models because the functional form of their systematic components may take negative values. Notice that for $\beta_0 > 0$, the range of the functions describing the systematic component of the two non-linear models considered in this study is the set of positive real numbers.

Records from lactations with an atypical shape were sent to ASOCEBU in order to detect possible errors in the dataset. Some errors that may lead to estimated lactation curves exhibiting an atypical shape are: typos corresponding to very high yields towards the end of lactation which may have a high leverage (this may be due to incorrect recording of yield or days in milk) and errors when recording parity (what seems to be a single lactation may correspond to records from two consecutive lactations). After discarding errors in the dataset, further analysis should be performed in order to identify why these lactations showed such trajectories.

In general, the POL2 model had the best fit in most lactations, while the PB model was the one selected with the lowest frequency (Table 2). On the other hand, Wood's model, one of the most frequently used models to fit lactation curves (Macciotta *et al.*, 2001; Quintero *et al.*, 2007), was the third most frequently selected. However, when checking the frequencies of typical curves without negative estimated milk yield, the Wood model was second in the ranking of most fitted models; in this group, the POL2 model was still the one that fitted most of the lactations (Table 2). Our results are in agreement with those reported by Ossa *et al.*

(1997) who found that, with the exception of first-parity cows, a second-order polynomial fit milk yield data better than the Wood's model in dual purpose-hybrid cows from the Caribbean region of Colombia. It is worth mentioning that, as can be seen in Table 2, the PB model did not yield atypical functions because β_1 was always positive. Using simple results from calculus, it can be shown that under this setting this function cannot have such a shape. In addition, this function either has a global maximum or is strictly increasing in the set of positive real numbers. Consequently, the high frequency of lactations in which the observed data showed an almost flat or u-shaped trend is one of the reasons explaining the low overall frequency of this model being selected as the one with the best fit (last column of Table 2). However, among lactations with typical trajectories, the PB model was still the one with the smallest selection frequency. Hence, our results suggest a poor fit of the PB model to milk yield records from Gyr cattle in Colombian low tropic dairy systems located in the regions considered in this study.

It is very important for Gyr cattle breeders to assess the productive performance of their herds because it is a relevant input for decision making. Some herds have the infrastructure to record milk yield on a daily basis; hence, they know their actual productive performance. However, this is not the case in many dairy operations, especially in low tropic systems where economic resources are limited. In these cases, statistical modelling would permit the computation of dairy production parameters such as total milk yield during a given period, peak milk yield, days at peak, and persistency.

Our results set the basis for the characterization of milk yield dynamics and assessment of productive performance of purebred Gyr cattle under the environmental and management conditions of low tropic areas in Colombia. To the best of our knowledge, this is the first study modelling Gyr cattle lactation curves in Colombia, considering records from thousands of individuals performing in several herds. Hopefully, these results along with those derived from other analyses carried out with the same dataset (e.g., cross-validation of total milk yield computed using partial milkings) will contribute to data-driven decisions in herds from the studied region.

Declarations

Conflicts of interest

The authors declare they have no conflicts of interest with regard to the work presented in this report.

Author contributions

Paper writing: D. Ferro, C.A. Martínez, J. Gil, A. Jiménez, C. Manrique. Data edition: D. Ferro, J. Gil, A. Jiménez. R programming: C.A. Martínez, J. Gil. Statistical analysis: D. Ferro, C. Manrique, C.A. Martínez. Mathematical derivations: C.A. Martínez and J. Gil.

References

Akaike H. A New Look at the Statistical Model Identification. Ieee Transactions On Automatic Control decem 1974; ac–19, ko. 6 DOI: https://doi.org/10.1109/TAC.1974.1100705

Analla M. Model Validation Through The Linear Regression Fit To Actual Versus Predicted Values. Agric Syst 1998; 57: 115–119. URL: https://doi.org/10.1016/S0308-521X(97)00073-5

ASOCEBU. La raza Gyr especializada en leche. n.d. [accessed date: May 2019] URL: http://www.asocebu.com/index.php/el-cebu/razas/gyr

Brody S, Turner CW, Ragsdale AC. The Relation Between The Initial Rise And The Subsequent Decline Of Milk Secretion Following Parturition. The Journal of General Physiology 1924; 6: 541–545. DOI: https://doi.org/10.1085/jgp.6.5.541

Butler Farms. Gyr Cattle. n.d. [accessed date: may 2018] URL: https://butlerfarms.us/gyr-cattle/

Cañas A, Cerón-Muñoz M, Corrales J. Modelación y parámetros genéticos de curvas de lactancia en bovinos Holstein en Colombia. Revista MVZ Córdoba 2012; 17(2): 2998–3003. URL: https://doi.org/10.21897/rmvz.234

Cuellar MA, Pereira RJ. Evaluación lechera en razas mestizas y gir lechero puro en condiciones sub-tropicales (período 1984 – 2001). Santa Cruz (Bolivia): Departamento de Santa Cruz Facultad de Ciencias Veterinarias, UAGRM; 2002. URL: https://www.academia.edu/8330574/EVALUACI%C3%93N_LECHERA_EN_RAZAS_MESTIZAS_Y_GIR_LECHERO_PURO_EN_CONDICIONES_SUB-TROPICALES_Departamento_de_Santa_Cruz_per%C3%ADodo_1984_2001_1

De Souza R, Alcalde CR, Lopes CA, De Labio BS, Fonseca FD, Couto L, Hygino B, Silva AP. Lactation curves and economic results of Saanen goats fed increasing dietary energy levels obtained by the addition of calcium salts of fatty acids. Revista Brasileira de Zootecnia 2014; 43 (2): 73–79. URL: http://dx.doi.org/10.1590/S1516-35982014000200004

Faraway J. Linear Models with R. Boca Raton (FL): CRC Press, 2015.

Ferreira AG, Henrique S, Vieira RA, Maeda EM, Valotto AA. Fitting mathematical models to lactation curves from holstein cows in the southwestern region of the state of Parana, Brazil. Anais da Academia Brasileira de Ciências 2015; 1 (87): 503–517. DOI https://doi.org/10.1590/0001-3765201520130514

Johansson I, Hansson A. Causes of Variation in Milk and Butterfat Yields of Dairy Cows. Kungliga Lantbruksakademiens Handlingar 1940; 79 (62): 127. URL: https://www.cabdirect.org/cabdirect/abstract/19420100118

López S, France J, Odongo NE, McBride RA, Kebreab E, AlZahal O, Dijkstra J. On the analysis of Canadian Holstein dairy cow lactation curves using standard growth functions. Journal of Dairy Science 2015; 98 (4): 2701–2712. URL: https://doi.org/10.3168/jds.2014-8132

Nicolò PP, Corrado D, Salvatore PG, Roberto S, Giuseppe P. The Mathematical Description Of Lactation Curves In Dairy Cattle. Italian Journal of Animal Science 2011; 10 (51): 213–223. URL: https://doi.org/10.4081/ijas.2011.e51

Mellado M, Coronel F, Estrada A, Ríos F. Lactation Performance Of Holstein And Holstein X Gyr Cattle Under Intensive Condition In A Subtropical Environment. Tropical and Subtropical Agroecosystems 2011; 14 (3): 927–931. URL: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1870-04622011000300031&lng=es&nrm=iso

Ossa G, Torregroza L, Alvarado L. Determinación de la curva de lactancia en vacas mestizas de un hato de doble propósito en la Región Caribe de Colombia. Revista Corpoica july 1997; 2 (1): 54 – 57. URL: https://doi.org/10.21930/rcta.vol2_num1_art:163

Wood PDP. Algebraic Model of the Lactation Curve in Cattle. Nature octubre 1967; 216(5111), 164–165. DOI: https://doi.org/10.1038/216164a0

Pereira RJ, Verneque RS, Lopes PS, Santana Jr. ML, Lagrotta MR, Torres RA, Vercesi AE, Machado MA. Milk yield persistency in Brazilian Gyr cattle based on a random regression model. Genetics and Molecular Research 2012; 11 (2): 1599-1609. URL: http://dx.doi.org/10.4238/2012.June.15.9

Papajcsik IA, Bodero J. Modelling Lactation Curves Of Friesian Cows In A Subtropical Climate. Animal Production 1988; 47(2): 201-207. URL: https://doi.org/10.1017/S0003356100003275

Quiroz J, Granados L, Barrón M. Producción De Leche De Ganado Gyr Y F1 Y (Holstein X Gyr). Actas Iberoamericanas de Conservación Animal 2015; 6: 294-299. URL: https://docplayer.es/47865539-Produccion-de-leche-de-ganado-gyr-y-f1-y-holstein-x-gyr.html

Rawlings JO, Pantula SG, Dickey DA. Applied regression analysis: a research tool. New York (NY): Springer-Verlag; 1998.

Rivera J. Producción bovina para leche. Pasto (Colombia): Universidad de Nariño; 1997.

Tedeschi L. Assessment of the Adequacy of Mathematical Models. Workshop on Mathematical Model Analysis and Evaluation. Sassari (Italy), 2004. URL: https://www.nutritionmodels.com/papers/Tedeschi2004.pdf

Western Canadian Dairy Herd Improvement Services. DHI Lactation Curves. Western Diary Digest 2001, n.d. [accessed date: may 2019] URL: http://www.dairyweb.ca/Resources/WDD24/WDD2424.pdf

Wiltmink JBM. Comparison of Different Methods of Predicting 305-Day Milk Yield Using Means Calculated From Within-Herd Lactation Curves. Livestock Production Science 1987; 17: 1–17. URL: http://anzoo.org/publicaciones/index.php/anzoo/article/download/52/40

Yanar M, Güler O. Factors Influencing the Shape of Lactation Curve and Persistency of Holstein Friesian Cows in High Altitude of Eastern Turkey. Journal of Applied Animal Research 2009; 35(1): 39–44. URL: https://doi.org/10.1080/09712119.2009.9706981