Computers and Electronics in Agriculture 176 (2020) 105657

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Use of discriminant statistical procedures for an early detection of persistent lactations in dairy cows

Elisabetta Manca, Alberto Cesarani, Nicolò P.P. Macciotta, Alberto S. Atzori, Giuseppe Pulina, Corrado Dimauro*

Department of Agraria, University of Sassari, Viale Italia 39, 07100 Sassari, Italy

ARTICLE INFO

Keywords:
Extended lactation
Persistency
Discriminant analysis

ABSTRACT

With the development of precision dairy farming, data on individual daily milk production are easily available in many herds. The aim of the present research was to develop an algorithm able to early detect dairy cows with a potential persistent lactation using daily production data. In this study, 2295 lactations belonging to primiparous (1015) and multiparous (1280) Holstein cows from two different farms equipped with the Afimilk system were used. Based on daily milk yield at 305 days in milk (DIM), animals were grouped into three production classes: low (LC) with milk yield < 20 kg, middle (MC) with milk yield between 20 kg and 32 kg, and high (HC) with milk yield > 32 kg, respectively. Lactations of MC or HC were considered as suitable for becoming long lactations. Four different models (Wood, Ali & Schaeffer, Legendre polynomials and 4th degree polynomials) were fitted to individual lactations by using the first 90, 120 and 150 DIM. Estimated model parameters were considered as variables in two multivariate discriminant techniques. The canonical discriminant analysis was used to test for possible differences between the extreme classes LC and HC. The discriminant analysis was performed to assign animals to the two production classes. The canonical discriminant analysis significantly separated LC from HC both for primiparous and multiparous cows. Among the different lactation models, the 4th degree polynomial was the most precise when the discriminant analysis was used to assign animals to the two production classes. In particular, by using the data of the first 150 DIM, the percentage of LC lactations incorrectly assigned to HC was 5% for primiparous and 7% for multiparous. Errors slightly increased when data of 120 (6% and 8% for primiparous and multiparous) and 90 (7% and 12% for primiparous and multiparous) DIM were used.

The entire procedure could be automated by implementing, for example, the Afifarm's report with a statistical computer software and it could be applied at farm level or using data from different associated farms. In practice, a historical database with previous complete lactations should be firstly created. As a new lactation proceeds, the recorded milk production data are fitted by using the 4th degree polynomial model and the estimated parameters submitted to the discriminant analysis. The lactation will be assigned to LC or HC.

1. Introduction

In the traditional yearly cycle of calving in dairy cattle farming, the lactation length is fixed at 305 days in milk (DIM). Lasting the average cow pregnancy 283 days (nine months and ten days), a cow is normally inseminated approximately around 60–90 DIM and it is dried off at 305 DIM. Under this condition, the minimum dry period of 60 days for recovering the mammary gland and leaving undisturbed the final period of gestation (Fig. 1a) is ensured. However, when high yielding cows are inseminated in the first 60–90 DIM, troubles related to fertility (Pryce et al., 2004; Walsh et al., 2011), sanitary risks as mastitis (Bates and Dohoo, 2016) or high milk yielded at dry off (sometimes higher than 30 kg/day) are often reported. The consequence is that several cows have a lactation longer than the traditional end. VanRaden et al. (2006) related that more than 55% of US Holsteins had lactations over 305 days whereas, in a recent study, Capper and Cady (2020) found that the average lactation length of US Holsteins was 357 DIM. In Italy, 42% of Italian Holsteins had lactations longer than the traditional length (Steri et al., 2009). In particular, in 2018, in the 120 Italian higher producing farms, the lactation length was 378 DIM (ANAFI, 2020). However, if only the 100 top yielding Italian Holstein cows are considered, the lactation length is, on average, 410 DIM (ANAFI, 2020).

* Corresponding author.
E-mail address: dimauro@uniss.it (C. Dimauro).

https://doi.org/10.1016/j.compag.2020.105657
Received 23 March 2020; Received in revised form 20 July 2020; Accepted 20 July 2020

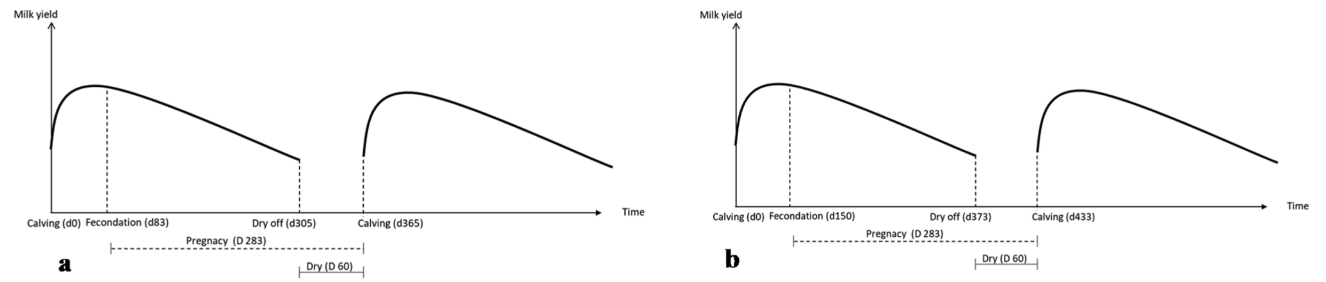

Fig. 1. a. The yearly cycle of dairy cow. Starting from the day of calving (d0), to maintain the inter-calving at 365 days with a dry period of 60 days (D60), being the pregnancy length of 283 days (D283), the cow would be fecundated at 83 days in milking (d83) and would be dry off at 305 days in milking (d305). **b.** The extended lactation resulting from a voluntary waiting period (interval calving – fecundation) of 150 days. Starting from the day of calving (d0), to maintain the dry period at 60 days (D60), being the pregnancy length 283 days (D283) and being cows fecundated at 150 days in milking (d150), the dry off results at 373 days in milking (d373) and the inter-calving at 433 days (d433).

A cow will have an extended lactation if, for different reasons, it becomes pregnant after 90 DIM. However, if the milk yield is not sufficiently high after the traditional lactation end, profit losses for the farm could occur. In high-yielding Holstein primiparous cows, Mellado et al. (2016) found an average daily milk yield of around 32 kg during the first 305 DIM and of 30 kg in the subsequent 253 DIM. Similar proportions of milk yield were reported for multiparous cows (around 35 and 32 kg for the first and second lactation period, respectively). In all parities, however, a standard deviation around 10 kg was observed. This result indicates that there are some animals with a low milk yield after 305 DIM.

An extended lactation could be obtained by delaying the voluntary waiting period, i.e. the interval during the postpartum period in which producers do not decide to breed cows even if estrus occurs (Fig. 1b). This practice could bring to several benefits for the farm. First, a cow could have longer time for restarting the normal ovarian cycle with a consequent reduction of hormonal treatments to control anestrus. In a study on reproductive failure in Holsteins, Butler et al. (2010) reported an average number of services per cow of 2.8 and 1.8, for calving intervals of 12 and 24 months, respectively. Inchaisri et al. (2010) found that the probability of insemination successes tends to increase with the progress of lactation. They also found that positive inseminations before peak yield were 6% lower than those after peak. No difference between primiparous and multiparous cows was found.

A voluntary lengthening of the voluntary waiting period should be, however, economically sustainable for farmers. Apart from management and sanitary savings, animals should have a suitable milk production over the standard lactation length (305 DIM). Thus, only cows with a possible future high persistent lactation could be suitable for the voluntary delay of mating in order to optimize the farm performances. With the development of precision dairy farming, data about individual daily milk production are easily available in many herds (Cabrera et al., 2020). Several mathematical models are currently exploited to describe and analyze the overall lactation and its different parts as peak production, time at peak and persistency. In particular, the last feature depends on the slope of the lactation curve after peak yield and, by using ordinary statistical techniques (e.g. regression models), it can be evaluated only in late lactation (Cole et al., 2009).

Lehmann et al. (2017) found that the information on milk daily production available for a cow at 90 DIM well correlates with milk yielded after 305 DIM and, therefore, that information could be used to select most suitable cows for extended lactation. An attempt to estimate the total milk production of a cow by using lactation data from the first 10 weeks was also carried out by Goodall and Sprevak (1985). In their recursive procedure, prior lactation information obtained from previous year's data were required. The method estimated the total milk yield with an error of about 6%.

In the present study, a new statistical approach to early estimate milk yielded by a dairy cow after 305 DIM is proposed. The algorithm exploits only daily lactation data recorded till 150 DIM obtained using automated milking devices. The method combines the features of continuous modeling in depicting lactation curve main traits with the ability of multivariate statistics to highlight differences between groups (lactations with low or high persistency in this paper).

2. Materials and methods

2.1. The data

Data consisted of individual daily milk production supplied by one farm located in Italy (Arborea, Oristano) and one in Hungary (Tiszaalpar, Bács-kiskun). The first farm had 1394 mature cows and contributed with 1984 lactations recorded from 2001 to 2008. The second farm had 244 mature cows with 391 lactations recorded from 2008 to 2016. In both farms, Holstein Friesian cows of high genetic merit were intensively managed indoor on cubicle housing system. The individual production per each milking was measured with the Afimilk equipment supervised by the software Afifarm (Afimilk, Kibbutz Afikim Israel) which allows the user to gather individual milk yield data directly from the milking machine. For each production cycle of cows, a lactation was considered for further analyses if it had records ranging from, at least, 10 to the 305 DIM. The total number of lactations that matched these requirements (2295) was split between primiparous cows with 1015 lactations (first parity group, FPG) and multiparous cows with 1280 lactations (multiple parity group, MPG).

2.2. Lactation curve models

Four lactation curve models, among those available in literature, were used to fit individual lactation curves. The Wood's equation was chosen for its ability in describing the shape of the lactation curve and because its parameters, and their combinations, have a clear biological meaning (Macciotta et al., 2005). The remaining three models were polynomial equations. The four models were:

i) the incomplete gamma function of Wood (1967) (Wood):

$$Y_t = at^b e^{ct}$$

where Y_t is the milk yield at time t and a, b, c are parameters that define the shape of the curve and the height on the ordinate axis;

ii) the five-parameter polynomial regression of Ali and Schaeffer (1987) (A&S):

$$Y_t = a + b\left(\frac{t}{305}\right) + c\left(\frac{t}{305}\right)^2 + d\ln\left(\frac{305}{t}\right) + e\left[\ln\left(\frac{305}{t}\right)\right]^2$$

where a is a parameter associated with the peak yield, b and c are parameters associated with slope in the decreasing phase, d and e are parameters associated with increasing slope in the phase until peak and t is the time, in days, from 10 to 305 days after parturition (Silvestre et al., 2006);

iii) the fourth-order Legendre orthogonal polynomial (Legendre) (Macciotta et al., 2005):

$$Y_t = \alpha_0 P_0 + \alpha_1 P_1 + \alpha_2 P_2 + \alpha_3 P_3 + \alpha_4 P_4$$

where αi are parameters to determine and Pj are calculated with values published by Schaeffer (2004); iv) the 4th degree polynomial (4th Pol):

$$Y_t = a + bt + ct^2 + dt^3 + et^4$$

where a, b, c, d and e are parameters to determine and t is the time in days from 10 to 305 days after parturition.

At first the Wood model was fitted to each individual lactation and the milk yield at 305, 400, 500 and 600 DIM was calculated. The correlation between the predicted milk yield at 305 DIM and at 400, 500 and 600 DIM was estimated to ascertain if the production at 305 DIM can be considered a good indicator for milk yielded in the subsequent stages of lactation.

Based on milk predicted at 305 DIM, lactations were grouped into three classes of production: low (LC) with milk yield lower than 20 kg, middle (MC) with milk yield between 20 kg and 32 kg, and high (HC) with milk yield greater than 32 kg. Both MC and HC contained lactation whose yield at 305 DIM was greater than 20 kg. Therefore, only lactations belonging to LC were considered not suitable to become an extended lactation. These boundaries were fixed considering that a milk yield lower than 20 kg is considered not economically profitable (Bach et al., 2020). However, a farmer, according to his breed and his objectives, could use other class limits.

The four curve models (Wood, A&S, Legendre and 4th Pol) were repeatedly fitted to data using, at each run, only the first 90, 120 and 150 DIM. For every model, the estimated regression parameters were retained and arranged in a multivariate manner. The three (90, 120 and 150 DIM) per four (Wood, A&S, Legendre and 4th Pol) obtained datasets, 12 for FPG and 12 for MPG, had lactations as rows and the regression parameters as columns. One categorical variable containing the production classes was also added to each dataset. The 24 datasets were very different from each other. Every dataset contained its specific regression parameters according to the lactation model was used (Wood, A&S, Legendre and 4th Pol) and to the DIM was chosen (90, 120 and 150 DIM).

To enhance the discriminant power of the method, only LC and HC were considered in further analyses. In the multivariate procedure, a lactation with only early records available will be assigned to one of the two production classes. However, a wrong class assignment could occur. According to the aims of this research, it is crucial to avoid that a lactation belonging to LC is assigned to HC. In this case, the first insemination will be delayed and a cow with a low persistency will have an extended lactation. This event could determine a loss of profit for the farm. Only this erroneous assignment was considered an error. No matter if a HC lactation is assigned to LC or a MC because, in any case, the farm does not suffer losses.

2.3. Discriminant analysis

Two multivariate statistical techniques were used to assign lactations to LC or HC: the canonical discriminant analysis (CDA) and the discriminant analysis (DA). CDA is a multivariate technique which allows researchers to ascertain, using a specific set of variables, if two or more groups of objects belong to different populations or not. Unlike the cluster analysis, in the CDA the group an individual belongs to is known. If k is the number of groups, the CDA derives $k - 1$ linear equations, called canonical functions (CAN) that are linear combinations of the original variables and are used to assign objects to groups. The statistical significance in group separation can be evaluated by means of the Mahalanobis' distance and the corresponding Hotelling's T-square test (De Maesschalck et al., 2000). The DA is a multivariate technique capable to classify objects into one of the involved groups. In this case, an individual is assigned to one group if its discriminant score produced by the CANs is lower than the cutoff value obtained by calculating the weighted mean distance among group centroids (Mardia et al., 1979).

First, CDA was applied to the 24 previously obtained datasets to check if the considered DIM were able to discriminate LC from HC, both for FPG and MPG, using the CANDISC procedure of SAS software (ver. 9.4, SAS Institute Inc., USA, Cary, NC). The 24 datasets were different each other.

Second, each dataset was randomly divided into training (80%) and validation (20%). Only the training datasets contained the variable indicating the production classes each lactation belonged to. This partition of data was iterated 5,000 times by using a bootstrap resampling procedure (Efron, 1979). The DA was applied to the training dataset by using the DISCRIM procedure of SAS. The obtained discriminant criterion was then used in the validation dataset to predict the corresponding production classes. At each run, errors in assignment were recorded. We only count as error when a LC was classified as HC. The final average percentage error, after 5000 iterations, was calculated as the ratio between the total number of errors and the total number of LC in the 5000 validation datasets.

3. Results

The mean lactation curve for FPG had a peak (around 32 kg/day) lower than MPG (around 40 kg/day) with a greater persistency (7.3 vs 6.7). FPG reached the peak at the 90th day after calving, whereas the MPG showed the peak after the 50th day. The number of lactations between production classes differ in the two parity groups (Table 1). In both cases, most of lactations were in MC. For primiparous cows, the remaining 11% and 18% of lactations were in HC and LC, respectively. For multiparous, only 8% of lactations belonged to HC, whereas 37% of them were in LC. The average milk production at 305 DIM was nearly equal between the two parities with a slight preponderance of FPG for LC and MC and of MPG for HC. Anyhow, the variation in milk production at 305 DIM was greater in MPG than FPG.

Pearson's correlations between the predicted daily milk yielded at 305 DIM and at 400, 500 and 600 DIM were almost all above 90% (Table 2), except for that between 305 and 600 DIM for FPG that was 85%.

Except for Wood at 90 DIM in FPG (p-value = 0.46), the CDA significantly separated the two production classes both for FPG and MPG (p-value < 0.01). In FPG, the greatest Mahalanobis' distance was obtained with the 4th Pol at 150 DIM (89) followed by Legendre (80) in the same situation, whereas the lowest Mahalanobis' distances were obtained with Wood (lower than 12 for all DIM). In MPG the greatest distance between HC and LC was achieved by Wood at 150 DIM (82), followed by the 4th Pol at 150 DIM (64). The lowest distance in MPG was obtained with Legendre at 90 DIM (27). Fig. 2 displays the plot of the CAN, obtained with the 4th Pol model, for FPG and MPG at 150

Table 1
Mean (and standard deviation) milk production (kg) at 305 days in milk in the three production classes both for first (FPG) and multiple (MPG) parity groups.

Class[1]	FPG		MPG	
	n	Mean (st.dev)	n	Mean (st.dev)
LC	182	17.51 (1.96)	468	16.64 (2.46)
MC	718	25.47 (3.11)	711	24.78 (3.09)
HC	115	35.41 (3.35)	101	37.13 (4.64)

[1]LC low class with milk yield lower than 20 kg.
MC middle class with milk yield between 20 kg and 32 kg.
HC high class with milk yield greater than 32 kg.

Table 2
Pearson's correlations between milk production predicted by the Wood's model at 305 and those yielded at 400, 500 and 600 days in milk (DIM).

	First Parity Group		
	400 DIM	500 DIM	600 DIM
305 DIM	0.97	0.91	0.85
	Multiple Parity Group		
305 DIM	0.98	0.94	0.90

DIM. The ability of the CAN in distinguishing LC from HC was greater in FPG than in MPG. Due to the low performances of CDA in assigning lactations when the Wood's parameters were used, only the three polynomial models were exploited in the subsequent DA. Table 3 displays the average percentage errors in assigning lactations to HC after the bootstrap resampling procedure. The 4th Pol was the model that, in the DA, gave better results. With this model, the DA correctly assigned 95% of lactations by using the first 150 DIM in the FPG. With only 120 and 90 DIM, the correct assignations were 94% and 93%, respectively. In the MPG, with the same lactation model (the 4th Pol), the DA correctly assigned 93% of lactations at 150 DIM, 92% at 120 DIM and 88% at 90 DIM, respectively.

4. Discussion

In the present research, an algorithm able to early assess whether a lactation will have a persistency suitable for becoming a long lactation and, therefore, continuing to produce over the standard 305 DIM threshold was developed. Data used was individual daily milk production recorded in two different farms and in different years. The distribution of relative frequency of lactations across the three production classes was nearly equal in the two farms, even if the Hungarian one contributed only with 391 lactations. However, the addition of this data enhanced the variability of lactations used to develop the discriminant procedure, thus assuring the ability of the proposed method in correctly allocate lactations obtained also in different conditions.

Results from the Wood model substantially confirmed what is reported in literature about peak yield, total milk production and persistence for Holstein cows (Olori et al., 1999). Daily milk production at 305 DIM was estimated using Wood and three different classes of production (LC, MC and HC) were considered (Table 1). The percentage of lactations allocated in the three classes was rather different. In both cases the largest class was MC. In FPG around 18% of lactations belonged to LC whereas for MPG this value was 37%. These results indicate that the probability to have a long lactation is higher in FPG than in MPG because primiparous cows have, on average, a greater persistency than multiparous (Tekerli et al., 2000). Actually, this figure is maintained also as the lactation proceeds over 305 DIM. In FPG, at 305 DIM, the number of lactations in HC was 115 (Table 1). However, after 400, 500 and 600 DIM this number slowly falls to 63, 50 and 45 respectively. On the other hand, in MPG, lactations in HC decreased from 101 at 305 DIM to 43, 29 and 23 at 400, 500 and 600 DIM respectively. The average milk production at 305 DIM was slightly lower in MPG than in FPG for the first two production classes (Table 1). However, this value was greater in MPG than in FPG for HC. Moreover, in MPG, correlations between milk production at 305 DIM and that produced at 400, 500 and 600 DIM (Table 2) were greater than FPG. These results indicate that a multiple parity cow could have a more productive extended lactation than a primiparous one.

Table 3
Mean percentage of low class lactations incorrectly assigned to high class at 90,120 and 150 days in milk (DIM) for first and multiple parity groups Holstein cows.

First parity group	Legendre	A&S	4th Pol
90 DIM	9	8	7
120 DIM	7	8	6
150 DIM	6	8	5
Multiple parity group			
90 DIM	12	12	12
120 DIM	11	11	8
150 DIM	9	10	7

Legendre = the fourth-order Legendre orthogonal polynomial (Macciotta et al., 2005).
A&S = the five-parameter polynomial regression of Ali and Schaeffer (Ali and Schaeffer, 1987).
4th Pol = the 4th degree polynomial.

As in the present research, a multivariate approach was previously proposed by Atzori et al. (2013) to improve managerial strategies of dairy herds. With the same aim, in this study we exploited two multivariate techniques to test differences between production classes (the CDA) and to early assign new lactations to one of them (the DA). In this process two kind of errors can be made. One of them, the incorrect assignment of a HC lactation to LC, was not considered an effective error because, in this case, the cow will be normally managed with no economic consequence for the farm. Therefore, this kind of error was flagged as false error and it was not considered in the results. On the contrary, a wrong assignation of a LC lactation to HC could cause problems to the farm management, because the cow will be inseminated later to respect its contemporaries. The consequence is that the ordinary yearly calving pattern is broken, and the cow will have a

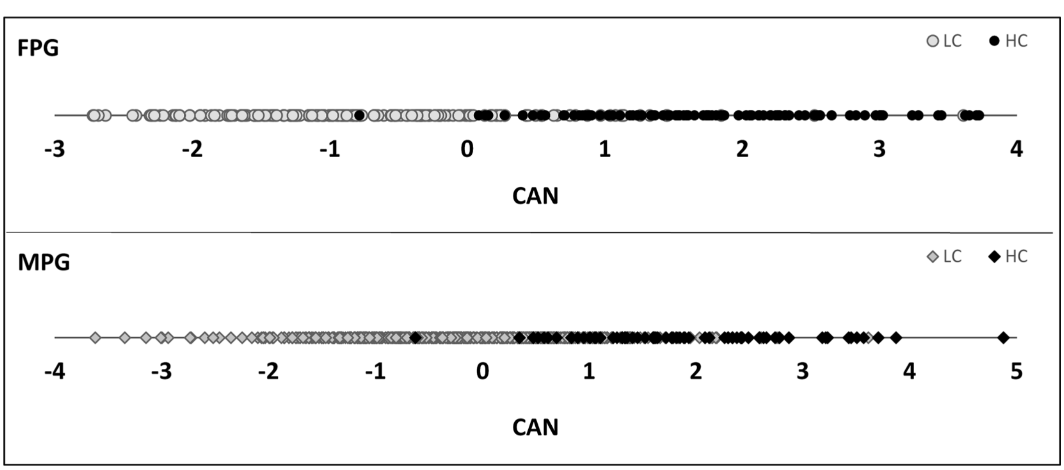

Fig. 2. Score plot of the canonical function (CAN), obtained by using the 4th degree polylomial, for first (FPG) and multiple (MPG) parity groups at 150 DIM between the low (LC) and high (HC) classes of production.

long lactation despite the low persistency. This kind of error was, therefore, flagged as true error (Table 3). No matter if a MC lactation is assigned to LC or HC. The cow will be ordinary managed with no consequence for the farm management.

The high statistical significance of most of Mahalanobis' distances highlighted a clear separation between the two classes of production. Only for FPG at 90 DIM, the Wood model was not able to significantly separate LC from HC. This probably occurred because Wood is a "rigid" model (Macciotta et al., 2005) and it requires data both in the ascending and in the descending phase of the curve to catch the correct shape of the lactation. Polynomial models are, on the contrary, flexible and they adapt to data better than Wood in the fitting procedure. The lowest error obtained in assigning LC lactations incorrectly to HC (5%) was obtained in FPG (Table 3) by using the 4th Pol at 150 DIM. Errors slightly increased in scenarios involving 120 and 90 DIM, with an error of 6% and 7%, respectively. This figure occurred because, as obvious, the greater the number of exploited daily milk production data the better the curve models fit the lactation shape. In MPG, errors in assignment increased compared with FPG. This probably happens because MPG lactations have a greater variation than FPG in determining the three production classes (Table1). However, also in this case, the 4th Pol was the model that better contributed to obtain a good classification. The error in assignment was, however, acceptable in the 150 DIM and 120 DIM (7% and 8%, respectively) scenarios. A total error of 12% was observed at 90 DIM.

4.1. Practical implications

If a farmer wants to select in his herd cows suitable for having an extended lactation by using the method here proposed, he need first to have a database with some complete lactations belonging both to LC and HC. To develop the discriminant procedures (CDA and DA) a full rank data matrix is required (Dimauro et al., 2011). In our research the data matrix is made up by a column with the production classes and other five columns with the five regression parameters of 4th Pol. So, its rank is ≥ 5 and, in theory, six lactations, distributed in the two classes of production, would be enough to develop the method. However, to obtain reliable results, it is important that most of the possible lactation shapes are in the training dataset. So, in our opinion, lactations recorded in different years and belonging to different farms should be joined to obtain a suitable set of data. Then the 4th Pol should be fitted to lactations and three data-matrices, each with regression parameters obtained considering milk yielded till 90, 120 and 150 DIM, would be created. These datasets, three for primiparous and three for multiparous cows, are the training datasets. As, in the farm, lactations proceed, the 4th Pol model is fitted to milk production data of each lactation thus obtaining the validation dataset for, as example, 90 DIM or for 120 DIM or for 150 DIM. Then the DA is applied to allocate lactations belonging to the validation dataset.

5. Conclusions

The algorithm developed in the present study could help farmers to select a quota of their herd to be destined to a long lactation. These cows will be managed in a different manner to respect the others, by delaying the voluntary waiting period that could bring several benefits for the farm. After an adequate training dataset is obtained, the method is able to early assign lactations to HC with a low error, from 5 to 12% depending on the chosen stage of the lactation. The entire procedure could be automated by implementing, for example, the Afifarm's report with a statistical computer software as SAS or a free software as R. However, the method can be efficiently implemented only if the farm has a high degree of automatization, with a robotic milking system supervised by a software able to produce daily reports on milk production.

CRediT authorship contribution statement

Elisabetta Manca: Methodology, Investigation, Writing - original draft. **Alberto Cesarani:** Data curation, Writing - review & editing. **Nicolò P.P. Macciotta:** Supervision, Writing - review & editing. **Alberto S. Atzori:** Data curation, Writing - review & editing. **Giuseppe Pulina:** Writing - review & editing. **Corrado Dimauro:** Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The study was supported by the Fondo di Ateneo per la Ricerca, Dimauro 2019, University of Sassari, Italy. The authors are thankful to dr. Enrico Costa and dr. Sergio Beltrame for providing data used in the research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2020.105657.

References

Ali, T.E., 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.

ANAFI (Association of Italian Holstein Friesan Breeders) http://www.anafi.it/it/pubblicazioni-statistiche/dati-aia-primi-allevamenti-frisona-italiani (accessed 13rd March 2020).

Atzori, A.S., Tedeschi, L.O., Cannas, A., 2013. A multivariate and stochastic approach to identify key variables to rank dairy farms on profitability. J. Dairy Sci. 96, 3378–3387. https://doi.org/10.3168/jds.2012-6256.

Bach, A., Terré, M., Vidal, M., 2020. Decomposing efficiency of milk production and maximizing profit. J. Dairy Sci. 103 [in press] https://doi.org/10.3168/jds.2019-17304.

Bates, A.J., Dohoo, I., 2016. Risk factors for peri-parturient farmer diagnosed mastitis in New Zeland dairy herd: findings from a retrospective cohort study. Prev. Vet. Med. 127, 70–76. https://doi.org/10.1016/j.prevetmed.2016.03.009.

Butler, T., Shalloo, L., Murphy, J.J., 2010. Extended lactations in a seasonal-calving pastoral system of production to modulate the effects of reproductive failure. J. Dairy Sci. 93, 1283–1295. https://doi.org/10.3168/jds.2009-2407.

Capper, J.K., Cady, R.A., 2020. The effects of improved performance in the U.S. dairy cattle industry on environmental impacts between 2007 and 2017. J. Anim. Sci. 98, 1–14. https://doi.org/10.1093/jas/skz291.

Cabrera, V.E., Barrientos-Blanco, A.J., Delgado, H., Fadul-Pacheco, L., 2020. Real-time continuous decision making using big data on dairy farms. J. Dairy Sci. 103, 3856–3866. https://doi.org/10.3168/jds.2019-17145.

Cole, J.B., Null, D.J., VanRaden, P.M., 2009. Best prediction of yields for long lactations. J. Dairy Sci. 92, 1796–1810. https://doi.org/10.3168/jds.2007-0976.

De Maesschalck, R., Jouan-Rimbaud, D., Massart, D.L., 2000. The Mahalanobis distance. Chemometr. Intell. Lab. Syst. 50, 1–18. https://doi.org/10.1016/S0169-7439(99)00047-7.

Dimauro, C., Cellesi, M., Pintus, M.A., Macciotta, N.P.P., 2011. The impact of the rank of marker variance–covariance matrix in principal component evaluation for genomic selection applications. J. Anim. Breed. Genet. 28 (6), 440–445. https://doi.org/10.1111/j.1439-0388.2011.00957.x.

Efron, B., 1979. Bootstrap Methods: Another look at the jackknife. The Ann. Stat. 7, 1–26.

Goodall, E.A., Sprevak, D., 1985. A Bayesian estimation of the lactation curve of dairy cow. Anim. Prod. 40, 189–193. https://doi.org/10.1017/S0003356100025290.

Inchaisri, C., Jorritsma, R., Vos, P.L.A.M., Van der Weijden, G.C., Hogeveen, H., 2010. Economic consequences of reproductive performance in dairy cattle. Theriogenology 74, 835–846. https://doi.org/10.1016/j.theriogenology.2010.04.008.

Lehmann, J.O., Mogensen, L., Kristensen, T., 2017. Early lactation production, health, and welfare characteristics of cows selected for extended lactation. J. Dairy Sci. 100, 1487–1501. https://doi.org/10.3168/jds.2016-11162.

Macciotta, N.P.P., Vicario, D., Cappio Borlino, A., 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.

Mardia, K.V., Kent, J.T., Bibby, J.M., 1979. Multivariate Analysis. Academic Press, London.

Mellado, T.L., Flores, J.M., de Santiago, A., Veliz, F.G., Macías-Cruz, U., Avendaño-Reyes, L., García, J.E., 2016. Extended lactation in high-yielding Holstein cows: characterization of milk yield and risk factors for lactations > 450 days. Livest. Sci. 159, 50–55. https://doi.org/10.1016/j.livsci.2016.05.004.

Olori, V.E., Brotherstone, S., Hill, W.G., McGuirk, B.J., 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.

Pryce, J.E., Royal, M.D., Garnsworthy, P.C., Mao, I.L., 2004. Fertility in the high-producing dairy cow. Livest. Prod. Sci. 86, 125–135. https://doi.org/10.1016/S0301-6226(03)00145-3.

Schaeffer, L.R., 2004. Application of random regression models in animal breeding. Livest. Prod. Sci. 86, 35–45. https://doi.org/10.1016/S0301-6226(03)00151-9.

Silvestre, A.M., Petim-Batista, F., Colaço, J., 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.

Steri, R., Cappio-Borlino, A., Macciotta, N.P.P., 2009. Modelling extended lactation curves for milk production traits in Italian Holsteins. Ital. J. Anim. Sci. 8 (2), 165–167. https://doi.org/10.4081/ijas.2009.s2.165.

Tekerli, M., Akinci, Z., Dogan, I., Akcan, A., 2000. Factors affecting the shape of lactation curves of Holstein cows from the Balikesir Province of Turkey. J. Dairy Sci. 83, 1381–1386. https://doi.org/10.3168/jds.S0022-0302(00)75006-5.

VanRaden, P.M., Dematawewa, C.M.B., Pearson, R.E., Tooker, M.E., 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.

Walsh, S.W., Williams, E.J., Evans, A.C.O., 2011. A review of the causes of poor fertility in high milk producing dairy cows. Anim. Reprod. Sci. 123, 127–138. https://doi.org/10.1016/j.anireprosci.2010.12.001.

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