Original scientific paper - Izvorni znanstveni rad UDK: 637.115

Methods for early prediction of lactation flow in Holstein heifers

Vesna Gantner[1], Klemen Potočnik[2], Krešimir Kuterovac[3], Ranko Gantner[1], Boris Antunović[1]*

[1]Poljoprivredni fakultet, Sveučilište J.J. Strossmayera u Osijeku, Trg Svetog Trojstva 3, Osijek, Hrvatska
[2]Univerza v Ljubljani, Biotehniška Fakulteta, Oddelek za zootehniko, Groblje 3, 1230 Domžale, Slovenija
[3]Agrokor d.d., Trg D. Petrovića 3, Zagreb, Hrvatska

Received - Prispjelo: 23.04.2010.
Accepted - Prihvaćeno: 22.11.2010.

Summary

The aim of this research was to define methods for early prediction (based on I. milk control record) of lactation flow in Holstein heifers as well as to choose optimal one in terms of prediction fit and application simplicity. Total of 304,569 daily yield records automatically recorded on a 1,136 first lactation Holstein cows, from March 2003 till August 2008., were included in analysis. According to the test date, calving date, the age at first calving, lactation stage when I. milk control occurred and to the average milk yield in first 25th, T_1 (and 25th-45th, T_2) lactation days, measuring month-calving month-age-production-time-period subgroups were formed. The parameters of analysed nonlinear and linear methods were estimated for each defined subgroup. As models evaluation measures, adjusted coefficient of determination, and average and standard deviation of error were used. Considering obtained results, in terms of total variance explanation (R^2_{adj}), the nonlinear Wood's method showed superiority above the linear ones (Wilmink's, Ali-Schaeffer's and Guo-Swalve's method) in both time-period subgroups (T_1 - 97.5 % of explained variability; T_2 - 98.1 % of explained variability). Regarding the evaluation measures based on prediction error amount ($e_{avg} \pm e_{SD}$), the lowest average error of daily milk yield prediction (less than 0.005 kg/day), as well as of lactation milk yield prediction (less than 50 kg/lactation (T_1 time-period subgroup) and less than 30 kg/lactation (T_2 time-period subgroup)); were determined when Wood's nonlinear prediction method were applied. Obtained results indicate that estimated Wood's regression parameters could be used in routine work for early prediction of Holstein heifer's lactation flow.

Key words: Holstein heifers, daily milk yield, early prediction, regression method, lactation flow

Introduction

The efficient dairy production depends greatly on proper decisions made at different production stages. Numerous genetic and environmental factors that influence animals performance could complicate decisions making, therefore, some toolkit for facilitation of breeding and management decisions on dairy farm is required. Lactation curves, that are mathematical formulas that describe flow of milk yield during lactation, could be used for production monitoring. The shape of the curve provides the breeders with some valuable information. For instance, cows with a very high peak yield are unable to take the sufficient amount of nutritional substances during the first stage of lactation, which could probably induce negative energy balance, reduced reproduction rates, as well as increased susceptibility to diseases (Jakobsen et al., 2002; Swalve, 2000). Conversely, cows with a flat lactation curve are more resistant to

*Corresponding author/Dopisni autor: E-mail: vgantner@net.hr

metabolic stress during the first lactation phase and have their energy demands more balanced, which in turn reduces feeding costs (Dekkers et al., 1998). Knowledge of the lactation curve allows prediction of total milk yield from a single or several test days early in lactation. With such knowledge, a dairy producer can make management decisions early based on individual production. Predicted milk yield may also facilitate breeding decisions, e.g., selection, culling. There are various mathematical models that could be used in describing lactation curves in dairy cows, for instance Wood's (1967) incomplete gamma function; parametric curves such as the Wilmink curve (Wilmink, 1987), the Ali-Schaeffer curve (Ali and Schaeffer, 1987), and the Guo-Swalve curve (Guo and Swalve, 1995); orthogonal polynomials (Olori et al., 1999); as well as natural cubic splines (White et al., 1999). The aim of this research was to define methods for early prediction (based on I. milk control record) of lactation flow in Holstein heifers as well as to choose optimal one in terms of prediction fit and application simplicity.

Material and methods

Datasets

The investigation was carried out on a Holstein dairy farm from March 2003 till August 2008. The Strangko milking equipment installed in the barn and linked to a computer system was used for automatic data recording. The cows were milked twice a day, in the morning and in the evening, meaning that the daily records were equal to the sum of partial ones. Total of 304,569 daily yield records of 1,136 first lactation cows were included in the analysis. Yields lower than 3 kg and higher than 50 kg of milk per day were treated as extreme values and deleted from the dataset. The yields recorded after the 500th day in milk were also deleted. According to the test date, records were divided in measuring month subgroups (M_1 - January, M_2 - February, ..., M_{12} - December), while according to the calving date, cows were divided in calving month subgroups (C_1 - January, C_2 - February, ..., C_{12} - December). In regard to lactation stage when first milk control occurred two time-period subgroups were created (T_1 - I. milk control occurred till 25th day in lactation; T_2 - I. milk control occurred from 25th till 45th day in lactation). In every time-period subgroup, cows were divided in production subgroups (P_1, P_2, P_3 and P_4) regarding the average daily yield in the first 25 lactation days (P_1, P_2, P_3 and P_4 include animals that produce ≤18 kg/day; 18-22 kg/day; 22-26 kg/day and >26 kg/day) as well as regarding the average daily yield in period from 25th till 45th lactation day (P_1, P_2, P_3 and P_4 included animals that produce ≤22 kg/day; 22-24 kg/day; 24-26 kg/day and >26 kg/day). Regarding the age at first calving cows were divided in two age subgroups (A_1 and A_2 that included animals first calved at age ≤27 and >27 months). In order to detect outliers, analysed prediction methods were applied to the data. Residuals over three standard deviations were taken as outliers and deleted from the dataset. Finally, dataset appropriate for analysis contained 266,746 daily yield records from 969 first lactation cows.

Lactation curves

The equations used to describe the lactation curves are presented in Table 1. In the incomplete gamma function proposed by Wood (1967), y represents the average daily milk yield during given time t, t the time expressed in days, a the parameter related to the peak lactation, b the parameter related to the ascending part of the curve between calving and peak lactation, c is the parameter related to the descending part of the curve following the peak lactation. Respecting the fact that a nonlinear regression does not guarantee convergence, natural logarithms were taken of both sides of the equation. Since the parameters of Wood's curves were

Table 1. Equations used to describe the lactation curve
Tablica 1. Jednadžbe korištene za opis laktacijske krivulje

Equation	Functional form
Wood	$y_i = at^b e^{-ct}$
Wilmink	$y_i = a + bt + ce^{-0.05t}$
Ali-Schaeffer	$y_i = a + b(t / 305) + c(t / 305)^2 + d \ln(305 / t) + f(\ln(305 / t))^2$
Guo-Swalve	$y_i = a + b(t / 305) + c \sin(t / 100)t^2 + d \sin(t / 100)t^3 + fe^{-0.055t}$

highly correlated, the Levenberg-Marquardt method was chosen for estimation using the SAS package NLIN procedure (SAS/STAT, 2002-2003). For each combination of above defined subgroup; measuring month, calving month, age, production, time-period; set of parameters were estimated.

According to Wilmink (1987), the parameters *a*, *b*, and *c* are associated with the level of production, the increase of production before the peak, and with the subsequent decrease, respectively. In the polynomial regressions by Ali and Schaeffer (1987) as well as by Guo and Swalve (1995), *a* is a parameter associated with the peak yield, *d* and *f* are parameters associated with increasing slope, while *b* and *c* are parameters associated with decreasing slope. In all linear regressions, y_i is the daily milk yield in kg, while t is stage of lactation in days. The parameters of linear regressions that are Wilmink's, Ali-Schaeffer's and Guo-Swalve's methods were estimated for every defined subgroup (measuring month, calving month, age, production, time-period) using the SAS package GLM procedure (SAS/STAT, 2002-2003).

Evaluation measures

For evaluation of prediction fit of the applied prediction methods following measures were used:

- *Adjusted coefficient of determination*, R^2_{adj}, which impart the percentage of the variance of daily yields explained by the method:

$$R^2_{adj} = 1 - \frac{MS_E}{MS_T}$$

where:

- MS_E - the model error variance,
- MS_T - the total model variance.

- *Average and standard deviation of prediction error*, $e_{avg} \pm e_{SD}$, which measure the prediction error in absolute terms (Guo and Swalve, 1995).

Results

Prediction methods were applied separately for each time-period subgroup (T_1 - I. milk control occurred till 25th day in lactation; T_2 - I. milk control occurred from 25th till 45th day in lactation).

The evaluation measures (R^2_{adj}, $e_{avg} \pm e_{SD}$) of applied prediction methods when I. milk control occurred till 25th day of lactation are presented in Table 2. The determined values of adjusted coefficients of determination (R^2_{adj}) for linear methods were considerably lower (Wilmink's: 0.416; Ali-Schaeffer's: 0.426; Guo-Swalve's: 0.427) in comparison to the determined R^2_{adj} value of nonlinear method (Wood's: 0.975). Applied linear prediction methods explained less than 45 % of total variability indicating poor prediction fit. Slightly higher amount of explained variability was provided by the application of the Ali-Schaeffer's as well as by the Guo-Swalve's method in comparison to the Wilmink's method. Similarly, Jamrozik et al. (1997) observed that goodness of models fit slightly increases with the number of model parameters.

Olori et al. (1999) emphasize that R^2 higher than 0.70 indicates a goodness of fitness, whereas values lower than 0.4 disqualify the model. Wood (1967) obtained R^2_{adj} in amount of 0.79 for models based on monthly yields. Freeze and Richards (1992) determined $R^2_{adj} = 0.51$ in cow's first lactation, while Olori et al. (1999) in their research obtained R^2_{adj} in amount of 0.94. The R^2_{adj} values in interval from 0.316 till 0.602 regarding the genotype-age-calving season subgroups for first lactating cows were determined in study of predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs) by Grzesiak et al. (2006). The R^2_{adj} for ANNs obtained in the same study in amount of 0.77 indicate its enhanced applicability in regard to Wood's model. Obtained results in this study indicate, in terms of total variance explanation, prominence of nonlinear Wood's method above linear ones.

Regarding the value of prediction error (e_{avg}, e_{stderr}) of daily milk yield in different lactation stages (≤100 days; 100-200 days; 200-300 days; >300 days) predicted values overestimated and underestimated actual ones for less than 0.5 kg/day when Wilmink's method was applied. Application of Ali-Schaeffer's and Guo-Swalve's method resulted in more accurate prediction that is in average prediction error less than 0.05 kg/day. The lowest value of prediction error (less than 0.005 kg/day) was observed when Wood's method was applied. Lower values of models prediction error ($e_{avg} \pm e_{SD}$) indicate more accurate predictions. Regarding the prediction

Table 2. Adjusted coefficients of determination (R^2_{adj}) and prediction error (e_{avg}, e_{stderr}) of daily and lactation milk yield in accordance to applied prediction method (I. control occurred till 25th day)
Tablica 2. Korigirani koeficijent determinacije (R^2_{adj}) i pogreška procjene (e_{avg}, e_{stderr}) dnevne te laktacijske količine mlijeka u ovisnosti o metodi procjene (I. kontrola do 25. dana laktacije)

Lactation stage	Method; R^2_{adj}							
	Wilmink; 0.416		Ali-Schaeffer; 0.426		Guo-Swalve; 0.427		Wood; 0.975	
	e_{avg}	e_{stderr}	e_{avg}	e_{stderr}	e_{avg}	e_{stderr}	e_{avg}	e_{stderr}
Daily milk yield								
≤100	0.293	0.014	0.031	0.014	0.036	0.014	-0.001	0.013
100, 200	-0.131	0.015	-0.045	0.015	-0.067	0.015	0.005	0.014
200, 300	-0.457	0.017	-0.008	0.017	0.022	0.017	0.006	0.015
>300	0.473	0.024	0.045	0.024	0.027	0.024	0.004	0.018
0, 500	$2.8*10^{-13}$	0.008	$1.4*10^{-13}$	0.008	$5.1*10^{-13}$	0.008	0.003	0.007
Lactation milk yield								
	69.628	1.685	63.455	1.641	63.423	1.643	48.999	1.304

Table 3. Adjusted coefficients of determination (R^2_{adj}) and prediction error (e_{avg}, e_{stderr}) of daily and lactation milk yield in accordance to applied prediction method (I. control from 25th till 45th day)
Tablica 3. Korigirani koeficijent determinacije (R^2_{adj}) i pogreška procjene (e_{avg}, e_{stderr}) dnevne te laktacijske količine mlijeka u ovisnosti o metodi procjene (I. kontrola od 25.-45. dana laktacije)

Lactation stage	Method; R^2_{adj}							
	Wilmink; 0.492		Ali-Schaeffer; 0.497		Guo-Swalve; 0.498		Wood; 0.981	
	e_{avg}	e_{stderr}	e_{avg}	e_{stderr}	e_{avg}	e_{stderr}	e_{avg}	e_{stderr}
Daily milk yield								
≤ 100	0.237	0.013	0.037	0.013	0.027	0.013	0.001	0.012
100, 200	-0.124	0.014	-0.058	0.014	-0.057	0.014	0.004	0.014
200, 300	-0.351	0.016	-0.005	0.016	0.022	0.016	0.005	0.015
> 300	0.387	0.023	0.054	0.229	0.023	0.023	-0.002	0.018
0, 500	$-2.6*10^{-12}$	0.008	$-2.7*10^{-12}$	0.008	$-2.5*10^{-12}$	0.008	0.003	0.007
Lactation milk yield								
	37.027	1.362	31.212	1.340	31.182	1.340	27.088	1.210

error of lactation, milk yield lowest value in amount of 48.99 kg per lactation was determined in application of Wood's method for early prediction of lactation flow.

When I. milk control occurred from 25th till 45th day of lactation, the highest amount of explained variability (98.1%) was provided by the application of the Wood's method (Table 3).

Regarding the values of prediction error of daily and lactation milk yield, lowest values in approximate amount of 0.005 kg/day and 27 kg/lactation were determined when Wood's method were applied. Regarding the time-period subgroups (T_1; T_2), obtained results (Table 2 and 3) indicate more accurate prediction when I. milk control occurred from 25th till 45th day in lactation. Grossman and Koops, 1988; and Sherchand et al., 1995 noticed that application of Wood's curve results in an acceptable fit to milk yield data, with tendency of over-prediction during early and late lactation and under-prediction during mid-lactation. Silvestre et al. (2006) in research of mathematical function accuracy in modelling lactation curves using daily data determined average prediction error in amount of 0.0 for all compared models (Wood, Wilmink, Ali-Schaeffer, cubic splines and 3 Legendre polynomials). Use of 4-weekly data by lactation instead of daily data resulted in average prediction errors in amount of 0.5 kg/day indicating that model's prediction accuracy is highly affected by the reduction of the sample dimension. Macciotta et al. (2005) concluded that the availability of test day records before the peak yield is crucial for correct prediction of lactation curve shape. Tekerli et al. (2000) emphases that lactation curve traits are also affected by environmental variables such days in milk at first test day, calving age, calving year, calving season, parity, and pregnancy status.

Currently in Croatia there is a need for reduction of costs correlated with milk recording. On the other side, some facilitation toolkit, which will enable accurate and early (based on I. milk control record) lactation curve prediction, for breeding and management purpose on dairy farm are necessary. This situation implicate necessity of prediction method election optimal from the point of prediction fit and simplicity of application in routine solving of various problems that dairy breeders encounter in their day-to-day activity.

Conclusions

Considering obtained results, in terms of total variance explanation (*Adjusted coefficient of determination, R^2_{adj}*), the nonlinear Wood's method showed superiority above the linear ones (Wilmink's, Ali-Schaeffer's and Guo-Swalve's method) in both time-period subgroups (T_1 - 97.5 % of explained variability; T_2 - 98.1 % of explained variability). Regarding the evaluation measures based on prediction error amount (*Average and standard deviation of error, $e_{avg} \pm e_{SD}$*), the lowest average error of daily milk yield prediction (less than 0.005 kg/day), as well as of lactation milk yield prediction (less than 50 kg/lactation (T_1 time-period subgroup) and less than 30 kg/lactation (T_2 time-period subgroup)); were determined when Wood's nonlinear prediction method were applied. Obtained results indicate that estimated Wood's regression parameters could be used in routine work for early prediction of Holstein heifer's lactation flow.

Metode za ranu procjenu laktacijskog tijeka u holstein prvotelki

Sažetak

Ciljevi su provedenog istraživanja bili definirati metode za ranu procjenu (temeljem podatka utvrđenog pri I. kontroli mliječnosti) laktacijskog tijeka u prvotelki holstein pasmine te izabrati optimalnu s aspekta predikcijske točnosti te aplikacijske jednostavnosti. Analizom je obuhvaćeno ukupno 304.569 zapisa dnevne količine mlijeka izmjerenih na 1136 prvotelki holstein pasmine u razdoblju od ožujka 2003. do kolovoza 2008. godine. U ovisnosti o datumu mjerenja, datumu teljenja, dobi pri prvom teljenju, stadiju laktacije pri I. provedenoj kontroli mliječnosti te prosječnoj dnevnoj proizvodnji u prvih 25, T_1 (te 25-45, T_2) dana laktacije, kreirane su podgrupe mjesec mjerenja - mjesec teljenja - dob - proizvodnost - vremenski period. Parametri analiziranih nelinearnih i linearnih metoda procijenjeni su za svaku definiranu podgrupu. Kao mjere evaluacije korišteni su korigirani koeficijent determinacije te prosjek i standardna devijacija pogreške. Utvrđeni rezultati, s aspekta udjela pojašnjene varijabilnosti (R^2_{adj}), ukazuju na superiornost nelinearne Woodove metode u odnosu na linearne (Wilminkova, Ali-Schaefferova, te Guo-Swalvejeva metoda) u obje podgrupe vremenskog perioda (T_1 - 97,5% pojašnjene varijabilnosti; T_2 - 98,1% pojašnjene varijabilnosti). Obzirom na iznos predikcijske pogreške ($e_{avg} \pm e_{SD}$), najmanja pogreška procjene dnevne (manje od 0,005 kg/dan), te laktacijske količine mlijeka (manje od 50 kg/laktaciji (T_1 vremenski - period podgrupa), te manje od 30 kg/laktaciji (T_2 vremenski - period podgrupa)); utvrđena je pri aplikaciji Woodove nelinearne metode procjene. Rezultati utvrđeni provedenim istraživanjem ukazuju na mogućnost uporabe procijenjenih Woodovih regresijskih parametara za rutinsku ranu procjenu laktacijskog tijeka u prvotelki holstein pasmine.

Ključne riječi: holstein prvotelke, dnevna količina mlijeka, rana procjena, regresijske metode, laktacijski tijek

References

1. 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.
2. Dekkers, J.C.M., Ten Hag, J.H., Weersink, A. (1998): Economic aspect of persistency of lactation in dairy cattle. *Livest. Prod. Sci. 57*, 237-252.
3. Farhangfar, H., Rowlinson, P.,Willis, M.B. (2000): Estimation of lactation curve parameters for Iranian Holstein dairy cows using non-linear models. http://www.bsas.org.uk /meetings /annlproc/PDF2000/109.pdf.
4. Freeze, B.S., Richards, T.J. (1992): Lactation curve estimation for use in economic optimization models in the dairy industry. *J. Dairy Sci. 75*, 2984-2989.
5. Grzesiak,W., Błaszczyk, P., Lacroix, R. (2006): Methods of predicting milk yield in dairy cows - Predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs). *Computers and Eletronics in Agriculture. 54*, 69-83.
6. Grossman, M., Koops, W. J. (1988): Multiphasic analysis of lactation curves in dairy cattle. *J. Dairy Sci. 71*, 1598-1608.
7. Guo, Z., Swalve, H.H. (1995): Modelling of the lactation curve as a sub-model in the evaluation of test day records. *Interbull. Bull. 11*, 22-25.
8. Icar - International Committee for Animal Recording (2003): Guidelines approved by the General Assembly held in Interlaken, Switzerland, on 30 May 2002, Roma, 19-39.
9. Jakobsen, J., Madsen, P., Jensen, J., Peersen, J., Cristiensen, L.G., Sorensen, D.A. (2002): Genetic parameters for milk production and persistency for Danish Holsteins estimated in random regression models using REML. *J. Dairy Sci. 85*, 1607-1616.
10. Jamrozik, J., Kistemaker, G.J., Dekkers, J.C.M., Schaeffer, L.R. (1997): Comparison of possible covariates for use in a random regression model for analyses of test day yields. *J. Dairy Sci. 80*, 2550–2556.
11. 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.
12. 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.
13. Sherchand, L., Mcnew, R. W., Kellogg, D.W., Johnson, Z.B. (1995): Selection of a mathematical model to generate lactation curves using daily milk yields of Holstein cows. *J. Dairy Sci. 78*, 2507-2513.
14. SAS/STAT User's Guide. (2002-2003): Version 9.1.3. SAS Institute Inc., Cary, NC, USA.
15. Silvestre, A. M., Petim-Batista, F., Colaco, 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.
16. 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.
17. White, I.M.S., Thompson, R., Brotherstone, S. (1999): Genetic and environmental smoothing of lactation curves with cubic splines. *J. Dairy Sci. 82*, 632-638.
18. Wilmink, J.B.M. (1987): Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livest. Prod. Sci. 16*, 335-348.
19. Wood, P.D.P. (1967): Algebraic model of the lactation curve in cattle. *Nature 216*, 164-165.