

WILHELM GRZESIAK, JERZY WOJCIK and BARBARA BINEROWSKA

## **Prediction of 305-day first lactation milk yield in cows with selected regression models**

### **Abstract**

In the present study, the prognostic values of multiple and spline regression models were tested for 305-day lactation milk yield of cows. The predictors were: HF genes proportion in cow's genotype; average milk yield for 305-day lactation from 4 first milkings of all cows in a barn, in which a cow was used in a given year; month of calving; and average daily milk yield from first four test-day milkings. Models were developed basing on 628 first lactations of BW cows with average 71% HF genes proportion. Subsequently, the predictive values of the models examined were verified on the grounds of next 105 first lactations. Prognostic differences of the models examined were determined, finding the prognosis obtained with the spline regression more accurate (smaller prediction error, higher coefficient of correlation for prognosis and real values in the model containing information from first three test-day milkings). These models are easy to construct and may be useful in practical estimation of cow lactation yields. They may be used for predicting the actual lactation yield in order to minimise production costs and achieve better production results.

Key Words: HF-crossbreeds, milk yield, prediction, multiple regression, spline regression

### **Zusammenfassung**

Titel der Arbeit: **Vorhersage für die 305-Tage Erstlaktationsleistung von Kühen unter Verwendung ausgewählter Regressionsmodelle**

In den vorliegenden Untersuchungen wurden prognostische Eigenschaften von Modellen der mehrfachen und gestückelte Regression für die 305-Tage Erstlaktationsleistung geprüft. Als prognostische (prädikative) Variablen dienten: Prozentualer Anteil an Holstein-Friesian (HF)-Genen im Genotyp der Kuh, Anzahl der Laktationstage der Kuh, durchschnittliche Milchleistung für 305 Laktationstage der ersten vier Prüfgemelke aller Kühe im Stall, in dem die jeweilige Kuh im jeweiligen Jahr gehalten wurde, Abkalbemonat sowie mittlere Tagesmilchleistung der ersten vier Prüfgemelke. Die Modelle wurden anhand von 628 ersten Laktationen der schwarzbunten Kühe mit einem durchschnittlichen HF-Anteil von 78% aufgebaut. Anschliessend wurden Vorhersageeigenschaften der untersuchten Modelle anhand von 105 nacheinanderfolgenden ersten Laktationen von Kühen überprüft. Prognostische Differenzen der untersuchten Modelle wurden ermittelt, wobei die Prognosen mit Hilfe der gestückelten Regression sich als genauer erwiesen (kleinerer Prognosefehler, höherer Korrelationskoeffizient für die Prognose und für tatsächliche Werte bereits beim Modell mit Informationen von drei ersten Prüfgemelken). Die Modelle sind einfach zu erstellen und können zur praktischen Schätzung der Laktationsleistung von Kühen verwendet werden. Sie können zur laufenden Prognose der Laktationsleistung genutzt werden, um Produktionskosten zu minimieren und bessere Produktionsergebnisse zu erzielen.

Schlüsselwörter: Holstein-Friesian Kreuzungen, Milchleistung, Prognose, gestückelte Regression, teilweise Regression

### **Introduction**

In many countries the analysis of milk yield for 305-day lactation is a basis for evaluating the breeding value of milk cattle. Every month, or every second month, test-day milkings are being carried out, which are the base for calculating 305-day lactation milk yield. Commercial value of a cow is compared with the yield of cows

being in the same lactation, calved in the same year and season and coming from the same herd. Many factors are included into the model, as a herd-year-season effect (HYS). The estimated and the actual yield of a cow may differ considerably, however, in particular when the interval between test-day milkings are longer or also when the last test-day milking is collected before the end of 305-day lactation.

In recent years studies have been performed on the use of lactation curves for calculating the milk yield. The application of specific lactation equation may enable more accurate estimation of the performance of cows for 305 days. Most frequently, the regression models of WOOD (1967), WILMINK (1987) or ALI and SCHAEFFER (1987) are used, though the proposed lactation curves do not always allow to obtain satisfactory results in relation to indigenous cattle population (STRABEL, 2002). When estimating the breeding value, the yield on the test day is taken into account (PTAK and SCHAEFFER, 1993) as well as random regression models (JAMROZIK and SCHAEFFER, 1997). Some possibilities for milk performance evaluation may arise from application of automatic milking systems, where data are based on real milk yields, however this approach still needs verification (BOHLSSEN et al., 2003).

Knowledge of the lactation curve may be used for supplementing incomplete cow lactations, determining food rations, estimating more accurately production costs and possible profits (HOLMAN et al., 1984; GERNAND et al., 1999) as well as for diagnosing *mastitis* (LESCOURRET et al., 1995), *ketosis* (DETILLEUX et al., 1994) or relationship between mastitis and milk composition throughout lactation (AMIN, 2001). Milk yield prediction in for same cows, and also in the whole herd, may facilitate the producers in taking adequate breeding decisions, being instrumental this way in reducing the costs and increasing the profits.

Taking into account the foregoing, an attempt was undertaken in the present study to test the values of selected multiple and spline regression models based on the yields of first four test-day milkings. The received milk yield predictions for 305-day lactation were compared with the real yields obtained by a cow in that period.

### Material and methods

In the study were used the data on 733 Black-and-White (BW) cows with various proportion of Holstein-Friesian (HF) genes referring to their first lactations lasting at least 305 days. The lactations were split into 2 groups:

- 1 group - 628 lactations, for constructing regression models. The lowest proportion of HF genes for cows of that group was 25%, the largest almost 98.6% (67% on the average). These cows were used in 7 barns in the northern part of Poland in 1994-1999. Their milk performance was calculated basing on monthly test-day milkings (STOLZMAN, 1982).
- 2 group - 105 lactations, picked out from the same 7 barns. These lactations were terminated in 2000. The average proportion of HF genes in that group of cows was 71%.

The 305-day lactation milk yield in examined cows in particular barns was at the level of 4528 kg of milk, with the average daily yield being 19.8 kg. Butterfat content ranged 3.54-4.54%, and protein - 2.95-3.35%.

The highest average milk yield was recorded in cows usually during the first month of lactation. Later on, the milk yield decreased gradually (by 1 kg per month on the average).

Each cow was characterised by a group of 8 variables - the predictors (**X**):

**X**<sub>1</sub> - percentage of HF genes in cows genotype;

**X**<sub>2</sub> - number of days in cows lactation;

**X**<sub>3</sub> - average 305-day lactation milk yield from 4 first milkings of all cows of a barn, in which a cow was used this year;

**X**<sub>4</sub> - month of calving;

**X**<sub>5</sub> - average daily milk yield from test-day milking falling on the average on the 18th day of lactation (between the 5th and the 28th day);

**X**<sub>6</sub> - average daily milk yield from test-day milking falling on the average on the 48th day of lactation (between the 36 and the 59th day);

**X**<sub>7</sub> - average daily milk yield from test-day milking falling on the average on the 76th day of lactation (between the 66th and the 89th day);

**X**<sub>8</sub> - average daily milk yield from test-day milking falling on the average on the 104th day of lactation (between the 95th and the 118th day).

The explanatory variable (*y*) was total milk yield of a cow for the lactation determined as the real yield estimated according to obligatory methods and recorded in breeding documentation, as opposed to the yields predicted with the regression models ( $\hat{y}$ ).

In the analyses were used four multiple regression (MLR) and four spline regression models with break-points in arithmetic mean (PRA).

MLR may be written down as the following equation:

$$\hat{y} = \mathbf{X}\beta + e \quad \text{where:}$$

$\hat{y}$  - vector of values calculated on the basis of estimated model;  $\beta$  -  $(k+1) \times 1$  vector of coefficients,  $\beta = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ ;  $\mathbf{y}$  -  $n \times 1$  vector of variable Y observations;  $\mathbf{X}$  -  $n \times (k+1)$  matrix of observations on explanatory variables,  $e$  - vector of random errors.

In PRA a situation is considered, when dependent variable (*y*) changes together with the values of independent variables, and the linear correlation may remain within the wide range of different levels of production; however, above a certain value a discontinuity may appear in the correlation between variables. The moment of this discontinuity, called a break-point, divides the regression equation in two estimations, above and below a break-point. Thus, the model of spline regression consists of two independent equations, applied according to the value of dependent variable (PEDHAZUR, 1973). The break-point in the median allows to predict more accurately in the situation when distribution of traits is characterised by larger or smaller asymmetry:

$$\hat{y} = \mathbf{X}_1 \beta_1 \text{ (for } p \leq \bar{y}) + \mathbf{X}_2 \beta_2 \text{ (for } p > \bar{y}); \text{ where:}$$

**X**<sub>1</sub> - matrix of observations for explanatory variables for *p* (break-point) being less than or equal to the arithmetic mean of observations; **X**<sub>2</sub> - matrix of observations for explanatory variables for *p* (break-point) being more than the arithmetic mean;  $\beta_1$ ,  $\beta_2$  - vectors of coefficients;  $\bar{y}$  - arithmetic mean.

All analysed models had identical first four variables: **X1**, **X2**, **X3**, **X4**. Differences in the models resulted from using additional predictors: average milk yields from test-day

milkings in successive months of the test. The analysed models of multiple regression are:

**MLR1 - X5**

**MLR2 - X5, X6**

**MLR3 - X5, X6, X7**

**MLR4 - X5, X6, X7, X8**

and the models of spline regression are:

**PRA1 - X5**

**PRA2 - X5, X6**

**PRA3 - X5, X6, X7**

**PRA4 - X5, X6, X7, X8**

The analysed models were augmented originally with three additional variables: age of cows, inter-calving period and study year. The input of these variables in explaining the variability in milk yield with lactation was non-significant, just as analogous regression coefficients. Considering the excess of information, the setting-up of an adequate model would be difficult, therefore these variables were not taken into account in analysed models.

For estimating regression functions, Statistica 5.1 v.PL software package was used (Statistica PL, 1997). PRA models were approximated with the quasi-Newton's estimation algorithm, which estimates the function at each successive estimation step in different points, in order to determine derivatives of the 1st and the 2nd order. The derivatives indicate the direction and the gradient of changes in the function slope. The algorithm follows a certain path to the minimum of loss function, which estimates the volume of deviations between the predicted and the observed values (GILL and MURRAY, 1972). Each deviation represents a certain loss in the accuracy of prediction. Minimizing the loss function is a procedure in estimating the coefficients of regression equation. The loss function ( $F_L$ ) was:

$$F_L = \sum_{i=1}^n (y - \hat{y})^2 = \min; \text{ where:}$$

$y$  - observed values;  $\hat{y}$  - predicted values.

The quality of the obtained models was determined with the coefficient of determination, which indicates the extent, in per cent, in which the variability of the explained trait ( $Y$ ) has been explained by explanatory variables ( $X_1, X_2, \dots$ ) occurring in a given model, with the coefficient of determination being given at the same time corrected according to the formula below, which will enable the comparison of coefficients from models differing in the number of explanatory variables.

$$R_A^2 = 1 - \frac{n-1}{n-k} (1 - R^2); \text{ where:}$$

$n$  - number of observations;  $k$  - number of predictors;

$R^2$  - coefficient of determination;  $R_A^2$  - adjusted  $R^2$ .

Moreover, using the Shapiro-Wilk's test the normality of distribution of the differences between the values predicted by the model and the real values was verified. Also an assumption was analysed on the homoscedasticity, i.e. on the constancy of variance of the residual component basing on the following statistics:

$$F = \frac{(n_1 - 1)(n_2 - k - 1) \cdot \sigma_1^2}{(n_2 - 1)(n_1 - k - 1) \cdot \sigma_2^2}; \text{ where:}$$

$n_1$  and  $n_2$  - observations split into subsets according to the arithmetic mean;  $\sigma_1^2$  and  $\sigma_2^2$  - variances in both subsets;  $k$  - number of explanatory variables.

In the analysed models the possibility of explanatory variables being correlated or not was verified as well.

The obtained models were used for generating the prediction for group 2 of previously selected 105 cows. The yield predictions derived by respective models were compared with the real 305-day lactation yields, which were attained by these cows (according to documentation). The test for the difference between means (paired two group t-test or dependent sample t-test, or repeated measures t-test) was employed for related variables to verify the yields predicted with regression and the real yields. Also the coefficient of correlation ( $r$ ) was determined between the real values and the values predicted by respective models. The mean relative error of prediction (PE) was calculated as well (according to DITTMANN, 1999, under assumption that time momentum  $T = 0$ ):

$$PE = \frac{1}{n} \sum_{i=1}^n \frac{|y - \hat{y}|}{y};$$

For each model, numerical ranges and absolute percentage differences for the predictions and the real values were composed in 200-kg intervals in order to see the number of predictions that fell within particular limits of the arithmetic mean for the examined cows of group 2.

## Results

Tables 1-2 present the regression coefficients for the eight models analysed. All the coefficients were highly significant ( $P < 0.01$ ). The tables also show the coefficients of determination ( $R^2$ ), which ranged 0.74 to 0.91, thus being very high (LUSZNIEWICZ and SLABY, 2001).

Table 1

Multiple regression models with a variable number of predictors (Multiple Regressionsmodelle mit variabler Prädikatorenzahl)

	Multiple regression			
	MLR 1	MLR 2	MLR 3	MLR 4
Intercept	-4168.30	-4211.17	-4102.59	-4239.16
X <sub>1</sub>	7.94	5.64	4.71	4.26
X <sub>2</sub>	16.67	16.21	16.02	15.69
X <sub>3</sub>	0.39	0.28	0.21	0.18
X <sub>4</sub>	-25.02	-9.18	4.59	19.96
X <sub>5</sub>	105.67	55.81	39.00	33.81
X <sub>6</sub>	-	79.52	38.36	26.34
X <sub>7</sub>	-	-	74.10	39.02
X <sub>8</sub>	-	-	-	74.49
R <sup>2</sup>	0.74	0.79	0.83	0.88
F	**	ns	*	ns

(R<sup>2</sup> - coefficient of determination of the model; coefficient from F test for homoscedasticity of the model - ns - non-significant, \* - significant at  $P < 0.05$ , \*\* - significant at  $P < 0.01$ )

Table 3 presents prediction effects of the examined models for 105 cows selected. Differences between the average real milk yield for 305-day lactation and the predicted yields were negative and ranged -278.8 to -33.6 kg (MLR1 and PRA1). This difference was a result of the comparisons of particular observations, deviating either in plus or in minus from the real values. With the use of the dependent sample t-test, the predictions from respective models were compared with the real milk yields of the analysed set of 105 cows. In most cases, these differences were non-significant (except for MLR1 and MLR2). The lowest prediction error for the multiple regression amounted to 0.08% (MLR3 and MLR4), and to 0.06% (PRA4) for the spline one. Only PRA1 had a larger prediction error than the best models of multiple regression with 3 and 4 first test-day milkings.

Table 2

Spline regression models at the break-point equal to the arithmetic mean with a variable number of predictors (Modelle der gestückelten Regression im Knoten, der dem arithmetischen Mittelwert mit variabler Prädikatorenzahl gleicht)

	Spline regression (breakpoint at mean)							
	PRA1		PRA2		PRA3		PRA4	
	≤ 5164.51	>5164.51	≤ 5164.51	>5164.51	≤ 5164.51	>5164.51	≤ 5164.51	>5164.51
Intercept	-1003.92	-1713.28	-1425.33	-2160.29	-1685.07	-2373.92	-2129.19	-2839.19
X <sub>1</sub>	4.08	10.08	3.48	7.34	2.77	7.19	3.00	6.02
X <sub>2</sub>	9.50	12.62	10.19	13.60	10.76	13.80	11.10	14.12
X <sub>3</sub>	0.20	0.32	0.16	0.24	0.12	0.19	0.11	0.17
X <sub>4</sub>	-21.59	-28.85	-12.01	-16.52	-5.30	2.18	3.63	21.58
X <sub>5</sub>	76.33	72.56	50.17	37.51	41.37	22.46	38.21	19.02
X <sub>6</sub>	-	-	49.84	63.35	27.19	27.57	21.28	18.43
X <sub>7</sub>	-	-	-	-	50.35	66.36	31.32	35.02
X <sub>8</sub>	-	-	-	-	-	-	54.06	67.14
R <sup>2</sup>	0.84		0.87		0.89		0.91	
F	ns		*		ns		ns	

ns - non-significant, \* - significant at P<0.05, \*\* - significant at P<0.01)

Table 3

Milk yield parameters for 305-day lactation, prediction errors and the coefficient of correlation for predictive and real values (r) of the analysed models (n=105) (Milchleistungsparameter für 305-Tage-Laktation, Prädikationsfehler sowie Korrelationskoeffizienten für prognostizierte und tatsächliche Werte (r) von ausgewerteten Modellen)

	Mean milk yield	SD	Mean error	Error SD	Prediction error (%)	r
			Observed values			
	5255.08	1013.37	-	-	-	-
			Predicted values			
			multiple regression			
MLR1*	5470.48	865.14	-278.80	739.14	0.13	0.70
MLR2*	5366.39	807.34	-111.31	626.84	0.10	0.79
MLR3	5371.97	812.44	-116.89	509.44	0.08	0.87
MLR 4	5317.85	939.62	-126.17	458.85	0.08	0.87
			spline regression ( $\bar{y} = 5165.51$ kg)			
PRA1	5225.24	972.65	-33.56	596.15	0.09	0.82
PRA2	5250.14	992.45	-58.46	536.61	0.08	0.86
PRA3	5268.82	1010.09	-77.14	460.74	0.07	0.90
PRA4	5286.95	981.58	-95.26	404.79	0.06	0.92

\* - means for observed milk yields and for a given model differ significantly at P<0.05

The coefficients of correlation between the predictions and the real values ranged from 0.70 for MLR1 to 0.92 for PRA4. The values of those coefficients may be referred to linear models, which are based on the day of lactation.

The value of the average absolute differences for the real and the predicted values is shown in Table 4. The average absolute differences ranged from 632.66 to 310.91 kg milk, which made for 12.04 to 5.91% of mean real milk yield of the analysed population. The variability of absolute differences was high, and the maximum discrepancy between the predicted and the real values ranged in some cases from 1305 kg milk for PRA4 to 2045 kg milk for PRA1.

Table 4

Mean absolute differences for real and predicted values of respective regression models (n = 105) (Absolute mittlere Differenzen zwischen den tatsächlichen und prognostizierten Werten für einzelne Regressionsmodelle)

Models	Mean difference	Difference SD	Min.	Max.	95% confidence interval
Multiple regression					
MLR1	632.66	469.79	7.05	1994.42	541.75 – 723.58
MLR2	523.49	358.86	8.87	1523.99	454.03 – 592.93
MLR3	423.65	303.53	0.69	1449.95	364.91 – 482.38
MLR 4	379.99	284.32	2.13	1321.32	324.96 – 435.01
Spline regression ( $\bar{y} = 5165.51$ kg)					
PRA1	460.71	377.15	2.64	2045.12	387.72 – 533.70
PRA2	413.04	345.19	2.28	1814.63	346.24 – 479.84
PRA3	349.19	308.52	0.17	1427.30	289.49 – 408.90
PRA4	310.91	274.63	2.44	1305.56	257.76 – 364.06

Table 5 presents numerical and proportional values of predictions divided into intervals, which specify the value of absolute difference between the predicted and the real values. Most of matching predictions (almost 47%), which deviate at the utmost by 200 kg from the real values occurred in PRA4, whereas the least number of them was in MLR1 (16%). Model PRA4 had proportionally the highest number of matched predictions in each interval.

Table 5

Numerical and proportional prediction data in difference intervals between real and predicted milk yields for respective models (Zahlen- und Prozentzusammenstellung von Prognosen in den Differenzbereichen zwischen den tatsächlichen und prognostizierten Milchleistungen für einzelne Modelle)

Models	Difference interval											
	0 – 200 kg		0 – 400 kg		0 – 600 kg		0 – 800 kg		0 – 1000 kg		>1000 kg	
Multiple regression												
	n	%	n	%	n	%	n	%	n	%	n	%
MLR1	17	16.19	23	38.09	19	56.19	13	68.57	12	80.00	21	20.00
MLR2	22	20.95	18	38.09	29	65.71	15	80.00	9	88.57	12	11.43
MLR3	22	20.95	36	55.23	21	75.24	11	85.71	11	96.19	4	3.81
MLR 4	30	28.57	35	61.90	22	82.86	8	90.48	5	95.24	6	4.76
Spline regression ( $\bar{y} = 5165.51$ kg)												
PRA1	30	28.57	26	53.33	23	75.24	8	82.86	7	89.52	11	10.48
PRA2	30	28.57	34	60.95	16	76.19	11	86.67	7	93.33	7	6.67
PRA3	42	40.00	27	65.71	16	80.95	11	91.43	3	94.29	6	5.71
PRA4	49	46.67	25	70.48	15	84.76	9	93.33	5	98.09	2	1.91

## Discussion

The daily milk yield of cows with a similar proportion of HF genes was very similar according to test-day milkings in successive lactation months, i.e. during the first 30

days after the calving the cows attained their maximum daily milk yields, while in next months a smaller or a greater decrease in milk yield was observed. Such a course of yield during lactation is consistent with the reports of other authors (OLORI et al., 1997; SWALVE and GUO, 1999) and might have been considered to be normal if the high milk yield had been sustained for a longer time and the decrease had not been over-intensive. GARCIA and HOLMES (2001) stress that development of milk yield in cows during lactation is influenced also by the level of milk yield within the herd and the calving season in cows. CHMIELNIK et al. (1998) confirmed the increase of milk yield in cows with a higher proportion of HF genes; however, the course of lactation, in particular in cows with more than 85% content of HF genes, was characterised by two distinct peaks.

Great part of milk yield variability has been explained with the proposed set of predictive variables. In that respect, the best model was PRA4 ( $R^2 = 0.91$ ). In the studies of other authors (SHIVE-KUMAR, 1995), the proposed model for predicting the milk yield in cows in the 1st lactation with regard to body weight and some zoometric parameters, was characterised by low coefficient  $R^2 = 0.36$ . On the other hand, PAN et al. (1997) applied the model of multiple regression ( $Y = 200.548 + 59.122 \cdot X_1 + 1.447 \cdot X_2$ ), in which the explanatory variables were: peak milk yield during lactation ( $X_1$ ) and initial milk yield ( $X_2$ ). The coefficient of determination for that model amounted to 0.64 and was lower than that in the presented study. OLORI et al. (1999), when analysing the standard models of the lactation curves (incomplete gamma, inverse quadratic polynomial, polynomial regression, exponential, mixed log) for predicting some milk yield parameters in cows from the 1st lactation in a small herd, obtained very high  $R^2$  (0.944 - 0.996) for peak milk yield. For the whole lactation yield, however, these coefficients were much lower and were at the level of 0.66. These authors stressed that coefficient  $R^2 > 0.70$  points to well-matched model, whereas the model is disqualified when  $R^2 < 0.40$ .

Some authors (PERZ and SOBEK, 1999) suggest however that it is better to take into account the lowest variance of estimation error as the criterion of model quality instead of the value of  $R^2$ ; though one should stress that in the present paper are quoted the corrected coefficients  $R^2$ , which makes a comparative criterion for various models. One should remember, however, that every additional variable in a model increases the value of that coefficient (RENCHEK and CEAYONG, 1980). The analysed residuals were of normal distribution, except for MLR1 and MLR2. The homoscedasticity test showed that the equality of variances occurred for MLR2, MLR4, PRA1, PRA3 and PRA4, the obtained estimators being biased in the remaining models; moreover some variables in those models were inter-correlated, which hinders or even rules out the use of those models.

Logarithmic and polynomial equations presented by BADNER and ANDERSON (1985) were characterised by the coefficients of correlation from 0.77 to 0.96. The model of JENKINS et al. (1984) was characterised by the coefficient of correlation equal to 0.72. GUO and SWALVE (1995) received the coefficients of correlation above 0.97 for the analysed polynomial curves for the real and the predicted milk yields. The models analysed by OLORI et al. (1999) were characterised in the majority by randomness, with a differentiated value of the coefficient of correlation for the predicted and the real milk yields for particular cows (0.0 to 0.99).



Improvement of model predictability resulted in general from taking into account in models the average milk yield of test-day milking of the successive month, which in general increased the percentage of predictions in particular intervals and reduced the number of observations deviating by almost 1000 kg from the real milk yields. This is also evident for the parameters of particular models (Table 3 and 4), at the same time one may compare respective models for the full set of multiple regression data and incomplete spline regression models, which generally were more precise. PRA models were of similar values. The differences between predicted and observed values for spline regression (Table 3) to some degree resemble those obtained by SWALVE and GUO (1999) for herds of similar yield; using their own model, the authors recorded positive differences (7 to 70 kg of milk) for the first lactation cows, depending on the herd.

PERZ and SOBEK (1999) analysed the usefulness of 22 various models for predicting the milk yield basing on different test-day data, using many mathematical models: from the simplest - a simple regression, to more complex models with trigonometric functions and their reciprocals, concluding that a more complex model is not always better in predicting than the more simple one.

Models PRA 3 and PRA 4 predicted the milk yield for 305-day lactation with a difference below 350 kg milk, surpassing all MLR models. Unfortunately, the predictive values of certain models are ruled out by heteroscedasticity (MLR1, MLR3, PRA2), and despite their rather good other quality parameters, their predictions cannot be reliable.

Model PRA3 with a lower number of prognostic variables matched in the quality parameters MLR4 model with the full set of explanatory variables. When using that model, a prediction is possible just after 3 first test-day milkings (approximately 90 days of lactation), with the accuracy matching the predictions based on 4 test-day milkings for multiple regression. PERZ and SOBEK (1999) stress that more simple models are less sensitive to fluctuations appearing in the data, though they do not render truly the real course of lactation curve.

Recently, test-day milkings are being used more and more frequently, which may be directly included into evaluation of breeding value (*Test-Day Model*). Thanks to that, it is possible to see, for instance, on effect of health state, weather conditions, feeding method, maintenance conditions in a barn and other factors that eliminate the influence of environment. The accuracy of evaluation of the breeding value of cows increases when compared with 305-day milk yield by approximately 4-8% (JENSEN, 2001; STRABEL and SZWACZKOWSKI, 2000; SCHAEFFER et al., 2000). The lactation curve equation allows taking into account the changes in the quantity of milk obtained from a cow depending on the lactation stage. Also the models with random regression are applied (*Random Regression Model*), where the effect of individual is replaced by curvilinear regression, which takes into consideration the shape of lactation curve in estimating the genetic value of cow (SCHAEFFER and DEKKERS, 1994; JAMROZIK et al., 1997). The genetic value refers in that case to daily milk yield of cows.

### Recapitulation

In the models presented here a genetic agent has not been taken into account, though the estimated coefficients of regression may be used in the model of random

regression. According to the presented models, in particular those, the predictive values of which were the best (coefficient of correlation between the real value and the prediction -  $r > 0.90$ ), it is reasonable to predict the milk yield of cows basing on first four test-day milkings (approximately 100 days of lactation).

An important advantage of the presented models consists in the ease of design and in the clarity of interpreting the model parameters. Their comparison, as is showed by the present study, does not require such large data sets as in case of models with random regression. This speaks well for using these models in practical application, in particular for producers for actual predictions with the aim to minimize production costs and to achieve the best production results possible. One should remember, however, that environmental conditions and animal productive abilities change with time, hence it is necessary to modify regression coefficients of the model.

### References

- AMIN, A.A.:  
Lactation and Sample Test\_Day Multi-trait animal model for genetic evaluation of somatic cell scores in Hungarian Holstein-Friesian crossbreeds. *Arch. Tierz., Dummerstorf* **44** (2001) 3, 263-275
- ALI, T.E.; SCHAEFFER, L.R.:  
Accounting for covariances among test day milk yields dairy cows. *Can. J. Anim. Sci.* **67** (1987), 637-644
- BADNER, G.B.; ANDERSON, C.R.:  
Evaluation of five lactation curve models fitted from daily milk weights. *J. Dairy Sci.* **68** (1985) (Suppl. 1): 226 (abstr.).
- BOHLSSEN, E.; WABMUTH, R.; ORDOLFF, D.:  
Aussagegenauigkeit der Milchleistungsprüfung unter Bedingungen automatischer Melkverfahren – Vergleich deutscher und kanadischer Modellansätze. *Arch. Tierz., Dummerstorf* **46** (2003) 1, 3–15
- CHMIELNIK, H.; MAKULSKA, J.; SAWA, A.:  
Course of three consecutive lactations of BW cows improved with HF breed in Kujawsko-Pomorskie Region. [Przebieg trzech kolejnych laktacji krów czarno-białych doskonalonych rasą holsztyńsko - fryzyską w rejonie Pomorsko-Kujawskim]. *Zesz. Nauk. AR Wrocław, nr 331, Konferencje XVII. "25 lat doskonalenia krajowego bydła czarno-białego i czerwono-białego rasą holsztyńsko-fryzyską. Wyniki i perspektywy.* (1998), 109-116. [In Polish].
- DETILLEUX, J.C.; GROHN, Y.T.; QUAAS, R.L.:  
Effects of clinical ketosis on test day milk yields in Finnish Ayrshire cattle. *J. Dairy Sci.* **77** (1994), 3316-3323.
- DITTMANN, P.:  
Methods of sales prediction in the company. [Metody prognozowania sprzedaży w przedsiębiorstwie]. Wyd. AE we Wrocławiu (1999). [In Polish].
- GARCIA, S.C.; HOLMES, C.W.:  
Lactation curves of autumn- and spring-calved cows in pasture-based dairy systems. *Lives. Prod. Sci.* **68** (2001) 2-3, 189-203
- GERNAND, R.; RECKNAGEL, R.; HEERBACH, B.:  
Möglichkeiten zur kostensparenden Milchleistungsprüfung. 2. Mitt.: Untersuchungen zur Zuverlässigkeit von aufwandsreduzierten Verfahren zur Feststellung der Laktationsleistung. *Arch. Tierz., Dummerstorf* **42** (1999) 2, 117-134
- GILL, P.E.; MURRAY, W.:  
Quasi-Newton methods for unconstrained optimization. *Journal of the Institute of Mathematics and its Applications*, (1972) 9, 91-108
- GUO, Z.; SWALVE, H.H.:  
Modelling of the lactation curve as a sub-model in the evaluation of test day records. *INTERBULL Mtg., Prague, Sept. 7-8 INTERBULL Bull.* (1995), No. 11. *Int. Bull. Eval. Service*, Uppsala, Sweden
- HOLMAN, F.J.; SHUMWAY, C.R.; BLAKE, R.W.; SCHWARTZ, R.B.; SUDWEEKS, E.M.:  
Economic value of days open for Holstein cows of alternative milk yields with varying calving intervals. *J. Dairy Sci.* **67** (1984), 636-643
- JAMROZIK, J.; KISTEMAKER, G.J.; DEKKERS, J.C.M.; SCHAEFFER, L.R.:  
Comparison of Possible Covariates for Use in a Random Regression Model for Analyses of Test Day Yields. *J. Dairy Sci.* **80** (1997), 2550-2556

- JAMROZIK, J.; SCHAEFFER, L.R.:  
Estimates of Genetic Parameters for a Test Day Model with Random Regression for Yield Traits of First Lactation Holsteins. *J. Dairy Sci.* **80** (1997), 62-770
- JENKINS, T.G.; FERREL, C.L.:  
A note on lactation curves of crossbred cows. *Anim. Prod.* **39** (1984), 479
- JENSEN, J.  
Genetics evaluation of dairy cattle using test-day models. *J. Dairy Sci.* **84** (2001), 2803-2812
- LESCOURRET, F.; COULON, J.B.; FAYE, B. :  
Predictive model of mastitis occurrence in dairy cow. *J. Dairy Sci.* **78** (1995), 2167-2177
- LUSZNIEWICZ, A.; SŁABY, T.:  
Statistics with computer package STATISTICA™ PL; Theory and applications. [Statystyka z pakietem komputerowym STATISTICA™ PL, Teoria i zastosowania]. Wydawnictwo C.H. BECK, Warszawa.(2001) [In Polish].
- OLORI, V.E.; BROTHERSTONE, S.; HILL, W.G.; MCGUIRK, B.J.:  
Effect of gestation stage on milk yield and composition in Holstein Friesian Dairy Cattle. *Livest. Prod. Sci.* **52** (1997), 167-176
- OLORI, V.E.; BROTHERSTONE, S.; HILL, W.G.; MCGUIRK, B.J.:  
Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. *Livest. Prod. Sci.* **58** (1999), 55-63
- PAN, S.; KOLEY, N.; DUTTAGUPTA, R.; ROY, A.M.  
Prediction of milk on the basis of peak yield and initial yield in Holstein-Friesian X Hariana cows (F1). *Environment-and-Ecology* **4** (1997), 860-861
- PEDHAZUR, E.J.:  
Multiple regression in behavioral research. Holt, Rinehart & Winston, (1973), New York.
- PERZ, W.P.; SOBEK, Z.:  
Evaluation of selected mathematical models of lactation curves. *Biometrical Letters* **36** (1999) 2, 169-178
- PTAK, E.; SCHAEFFER, L.R.:  
Use of test day yields for genetic evaluation of dairy sires and cows. *Livest. Prod. Sci.* **34** (1993), 23-34
- RENCHER, A.C.; PUN, FU CEAYONG:  
Inflation of  $R^2$  in best subset regression. *Technometrics* **22** (1980) 1, 49-53
- SCHAEFFER, L.R.; DEKKERS, J.C.M.:  
Random regression in animal models for test-day production in dairy cattle. In: Proc. 5<sup>th</sup> World Congr. Gen. Appl. Livest. Prod., Guelph, ON, Canada, (1994)Vol. XVIII, 443.
- SCHAEFFER, L.R.; JAMROZIK, J.:  
Multiple – trait prediction of lactation yields for dairy cows. *J. Dairy Sci.* **79** (1996), 2044-2055
- SCHAEFFER, L.R.; JAMROZIK, J.; KISTEMAKER, G.J.; DOORMAAL, B.J.:  
Experience with a Test-Day Model. *J. Dairy Sci.* **83** (2000), 1135-1144
- SHIVE-KUMAR; PRASAD, R.B.; ARUN-KUMAR; DEVENDRA-SINGH; KUMAR, S.; KUMAR, A.; SINGH, D.:  
Prediction of first lactation milk yield from body measurements in crossbred cows. *Journal-of-Dairying,-Foods-and-Home-Sciences* **14** (1995) 1-2, 41-46
- STATISTICA PL:  
General concepts and statistics. [Ogólne konwencje i statystyki]. (1997), Vol I-III, StatSoft. Kraków [In Polish]
- STOLZMAN, M.:  
A hand-book for cattle evaluating animal husbandry specialist. Collective work under editor supervision of Dr M. Stolzman. [Poradnik zootechnika oceny bydła. Opracowanie zbiorowe pod kierunkiem dr M. Stolzman]. PWRiL W-wa, (1982), 26-30. [In Polish].
- STRABEL, T.; SZWACZKOWSKI, T.:  
Partial lactation in genetic evaluation based on test day model. *J. Anim. and Feed Sci.* **9** (2000), 47-53
- STRABEL, T.:  
Adaptation of model with random regressions to evaluation of the breeding value of milk cattle on the grounds of test-day milking. [ Dostosowanie modelu z losowymi regresjami do oceny wartości hodowlanej bydła mlecznego na podstawie próbných udójów]. *Rocz. Nauk. Zoot., Supl.*, (2002) 15, 13-18. [In Polish].
- SWALVE, H.H.; GUO, Z.:  
An illustration of lactation curves stratified by lactation yields within herd. *Arch. Tierz., Dummerstorf* **42** (1999) 6, 515-525
- WILMINK, J.B.M.:  
Comparison of different methods of predicting 305-day milk yield using means calculated from within-herd lactation curves. *Livest. Prod. Sci.* **17** (1987), 1-17

WOOD, P.D.P.:

Algebraic model of the lactation curve in cattle. *Nature* **216** (1967), 164-165

Received: 2003-01-20

Accepted: 2003-05-05

Author's address

WILHELM GRZESIAK, PhD; JERZY WOJCIK, DSc; BARBARA BINEROWSKA, MSc  
Department of Ruminants Science,  
Agricultural University of Szczecin,  
ul. Dr Judyma 10  
71-466 Szczecin  
Poland

E-Mail: [w.grzesiak@biot.ar.szczecin.pl](mailto:w.grzesiak@biot.ar.szczecin.pl)