Analysis in a time series of milk-yield production

Tomas Haloun[1], Jaroslav Marek[2], Radko Rajmon[3], Martina Kršíková[4]

Abstract. Monitoring of milk yield is used as a tool for the operative management of individual cows as well as management at the herd level. Nevertheless, sometimes a general knowledge of the significance of a particular problem is not sufficient for appropriate managerial decisions and a need arises to weigh up the impact of particular disorder on a particular herd. Than it is necessary to evaluate animal data from a longer period, different lactation order etc. The aim of the introduced research is to present methods for comparing the milk yield of cows in different epochs. The data from selected Holstein dairy herds in the Czech Republic from an 8-year period was used as a model. Several statistical hypotheses (zero trend, absence of seasonal variation) were being tested. The results indicated a requirement to consider the relationship between labour date and milk production. Our statistical analysis enabled us to draw up a table of milk-yield indices showing the average changes in 305-day yield values among lactations and years.

Keywords: Statistical analysis of milk production, Wood function, Nonlinear regression, Estimates of unknown parameters, time series.

JEL classification: C13
AMS classification: 62J02, 62M10

1 The introduction

Milk yield has a substantial impact on the economic profitability of dairy farms. Therefore a great deal of attention is paid to the analysis of this indicator as well as the factors which influence milk production. There exists a host of models which evaluate the effects such as the success of genetic improvement, increasing quality of care, better feed composition and seasonal effect milk production at population level, cf. [6]. At farm level the milk production is influenced by factors which impact the whole herd but at the same time there are factors which affect individual animals. In order to make managerial decisions, it can be desirable to assess the yield of individual cows in the context of the herd at a concrete farm, which means working with seriously limited data volumes and different models. The basis for yield evaluation is modelling of a 305-day yield. Using a mathematical model for the description of the lactation curve leads to the need for finding a suitable regression function for fitting of measurements of daily yield, which are performed mostly only once a month. Several methods for approximation of a standard 305-day lactation cycle have already been proposed. Many studies have focused on modelling the 305-day yield. Cf. [2], [1], [3]. The Wood function is the most preferred method for solution of this nonlinear regression problem. See [5].

To evaluate phenomena observed in the long term, which have an individual impact and occur less frequently, it is desirable to standardize the estimated 305-day yields. A timeline may be convenient to capture correctly trend, seasonal, and cyclic components. The existence of these components needs to be tested at a chosen level of significance with suitable statistical tests. Subsequently, the productivity can be stripped off substantial factors. The main goal of this paper is to propose a suitable solution to the issue of yield correction in selected animals, which will make comparisons possible in the long term.

[1]Czech University of Life Sciences Prague, Faculty of Agrobiology, Food and Natural Resources, Department of Veterinary Science, Kamycka 129, 165 21 Praha 6, Suchdol, Czech Republic, tomas.haloun@email.cz

[2]University of Pardubice, Faculty of Electrical Engineering and Informatics, Department of Mathematics and Physics, 532 10 Pardubice, Czech Republic, jaroslav.marek@upce.cz

[3]Czech University of Life Sciences Prague, Faculty of Agrobiology, Food and Natural Resources, Department of Veterinary Science, Kamycka 129, 165 21 Praha 6, Suchdol, Czech Republic, rajmon@af.czu.cz

[4]University of Pardubice, Faculty of Electrical Engineering and Informatics, Department of Mathematics and Physics, 532 10 Pardubice, Czech Republic, st32357@student.upce.cz

1.1 Measurements

The data was obtained from the company ZD Zahori register of two farms. All recordings for all 3737 animals registered between the years 2001 and 2013 were exported. The data included 76724 measurements 2803, 1928, 1232, 1531 cycles for first, second, third, four and larger lactation, respectively. The data were exported directly from the register with subsequent attributes: Cow id, cowshed stall number, date of cows birth, day of cows disposal, date of labour, order of lactation, date of the control day, daily milk yield on the control day [in kilogram], protein content [in kilogram], number of somatic cells (Tab. 1).

43539	265	1	20.11.1997	14.05.2007	07.12.2004	6	25.01.06	11.40	3.92	3.65	240.00
43539	265	1	20.11.1997	14.05.2007	04.03.2006	7	23.03.06	41.20	4.52	3.40	28.00
43539	265	1	20.11.1997	14.05.2007	04.03.2006	7	24.04.06	42.00	3.34	3.08	237.00
43539	265	1	20.11.1997	14.05.2007	04.03.2006	7	24.05.06	39.60	3.35	3.14	67.00
43539	265	1	20.11.1997	14.05.2007	04.03.2006	7	26.06.06	35.20	4.07	2.83	183.00
43539	265	1	20.11.1997	14.05.2007	04.03.2006	7	25.07.06	26.20	3.74	3.11	391.00
43539	265	1	20.11.1997	14.05.2007	04.03.2006	7	23.08.06	30.80	3.53	3.17	327.00
43539	265	1	20.11.1997	14.05.2007	04.03.2006	7	25.09.06	30.60	3.00	3.07	115.00
43539	265	1	20.11.1997	14.05.2007	04.03.2006	7	24.10.06	28.40	2.80	3.39	44.00
43539	265	1	20.11.1997	14.05.2007	04.03.2006	7	23.11.06	23.20	3.19	3.43	102.00
43539	265	1	20.11.1997	14.05.2007	04.03.2006	7	20.12.06	20.60	5.13	3.74	637.00
43539	265	1	20.11.1997	14.05.2007	04.03.2006	7	23.01.07	23.00	4.31	3.60	313.00
43539	265	1	20.11.1997	14.05.2007	04.03.2006	7	22.02.07	22.40	4.32	3.58	311.00
43539	265	1	20.11.1997	14.05.2007	04.03.2006	7	26.03.07	20.40	4.27	3.43	531.00
43539	265	1	20.11.1997	14.05.2007	04.03.2006	7	23.04.07	19.00	4.89	3.74	546.00
43542	265	1	28.11.1997	20.03.2002	26.01.2000	1	23.02.00	20.40	5.69	3.25	34.00

Table 1 Samples of the data on monthly measurements of the daily milk yield of several cows

1.2 Estimation of lactation curve by Wood function

Several mathematical models for 305-day yield are at our disposal. For detail see [2], [1], [3]. The results in [3] indicated very good properties of Wood models. Estimation of this model will be described now. Let x be the day of lactation and Y is the daily yield. Let's consider that function f models how variable x explain variable $\mathbf{Y}$. Wood model for estimating of 305-day yield is given by function

$$f(\boldsymbol{\beta}, x) = \beta_1 x^{\beta_2} e^{-\beta_3 x}. \tag{1}$$

With nonlinear regression we expect that $\mathbf{Y} - \mathbf{Y}_0$ has distribution with mean value $X\boldsymbol{\beta}$, and covariance matrix where $\mathbf{Y}$ is vector of measured daily yields, $\mathbf{Y}_0 = f(\boldsymbol{\beta}_0; x)$, $\boldsymbol{\beta}_0$ is a suitable initial vector of unknown parameters. Ordinary squares estimate of unknown vector parameter is given by formula

$$\delta\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{Y} - \mathbf{Y}_0)', \tag{2}$$

where $\mathbf{X}$ is matrix of first partial derivatives of Wood function. Dimension of the matrix is $n \times 3$. After determining the partial derivatives of Wood function our matrix $\mathbf{X}$ has a form:

$$\mathbf{X} = \begin{pmatrix} x_1^{\beta_2} e^{-\beta_3 x_1}, & \beta_1 x_1^{\beta_2} e^{-\beta_3 x_1} \log(x_1), & -\beta_1 x_1^{\beta_2+1} e^{-\beta_3 x_1} \\ & \dots & \\ x_n^{\beta_2} e^{-\beta_3 x_n}, & \beta_1 x_n^{\beta_2} e^{-\beta_3 x_n} \log(x_n), & -\beta_1 x_n^{\beta_2+1} e^{-\beta_3 x_n} \end{pmatrix}. \tag{3}$$

Estimation of unknown parameters is determined by adding up the initial solution and correction given in formula (2):

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + \delta\widehat{\boldsymbol{\beta}}. \tag{4}$$

In nonlinear regression for functions with large Bates & Watts curvature and a small area of linearization a correction of the initial solution is given by relationship

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + \alpha\delta\widehat{\boldsymbol{\beta}}. \tag{5}$$

In process of numerical calculations we can choose equidistant step eg. $s = 1/1000$. So we will get values $\alpha_i = i/1000$, $i = 1, \dots, n$. In the cycle 1000 estimates of unknow parameters is obtained. An estimate with the smallest value of criterior for estimation of regression line

$$S_e = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2 \tag{6}$$

is considered as most satisfactory.

1.3 Example of approximation

This section will give a detailed description of individual steps in calculating unknown parameters of the approximate function in one milk cow. Data are selected from seven lactation cycles of the milk cow with ID 43539. The data for this cow in the table 2 was collected and calculated on the basis of data on date of the milk yield measurement, date of the cow's birthday and subsequent labors given in Table 1. Measured values of daily yields (in kilograms) are labelled y and the serial (order) number of the lactation cycle in measured days is marked as x.

x	19	51	81	114	143	172	205	214	244	271	325
Y	41.2	42.0	39.6	35.2	26.2	30.8	30.6	28.4	23.2	20.6	23.0

Table 2 Samples of monthly measurements of the daily milk yield with order of lactation day

We need an appropriate initial vector of unknown parameters of Wood function, which is close to the true value of unknown parameters. Suitable initial solution for approximation by Wood function can be set as $(5; 0.8; 0.02)$, cf. [3]. Then vector Y_0 can be calculated by substitution of initial solution $\boldsymbol{\beta}_0$ and vector x of the order of days in lactation cycle, into Wood function. We need this vector to calculate the difference between estimated and measured milk yield. Calculated vector is $y_0 = (42.7, 39.6, 37.1, 34.6, 32.5, 30.6, 28.5, 26.8, 25.2, 23.8)'$. Numerical calculation provides the following subresults and results

$$\mathbf{X} = \begin{pmatrix} 0.94, & 125.58, & 810.36 \\ 0.87, & 155.76, & 2020.40 \\ 0.81, & 163.05, & 3005.43 \\ 0.76, & 163.74, & 3941.13 \\ 0.72, & 161.33, & 4648.61 \\ 0.67, & 157.40, & 5259.38 \\ 0.63, & 151.86, & 5848.33 \\ 0.59, & 146.45, & 6281.95 \\ 0.56, & 140.59, & 6656.25 \\ 0.52, & 135.20, & 6934.73 \end{pmatrix}, \qquad \begin{array}{c} \delta\widehat{\boldsymbol{\beta}} = (45.4125, 0.0080189, 0.0020595)', \ \alpha_{opt} = 3/1000, \\ \widehat{\boldsymbol{\beta}} = (5.1362375, 0.799976, 0.020006)', \ S_e = 79.62. \end{array}$$

Obtained estimate of unknown parameters is vector $\widehat{\boldsymbol{\beta}} = (5.1362375, 0.799976, 0.020006)'$. Estimate of lactation curve is presented in Fig. 1.

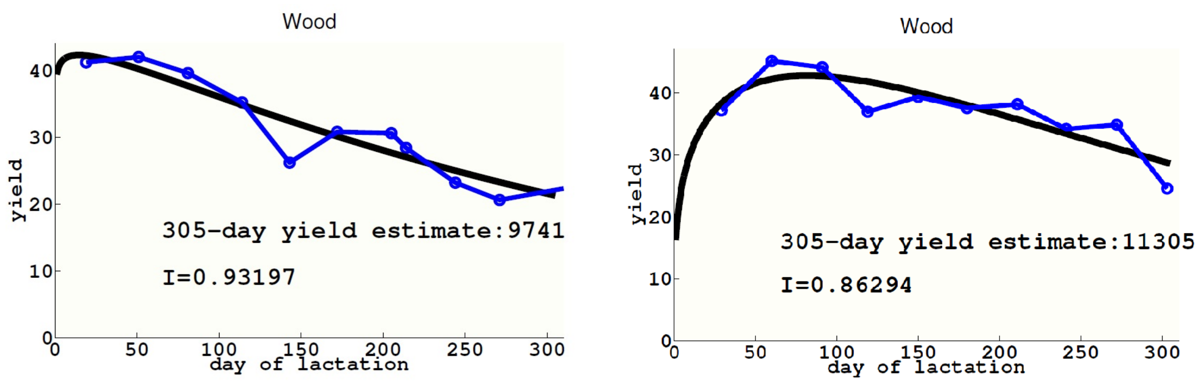

Figure 1 Estimates of 305 d yield of cow 43539 at 1th lactation and cow 54538 at 7th lactation

1.4 Applications to time series

In the process of finding a suitable methodology for data processing various test of statistical hypotheses about trend in years, quarters, months were performed. The results of these tests demonstrated a significant correlation between milk production and the years. Furthermore significant seasonal component appeared in the quarters of the year. Comparing months appeared to be problematic with regard to the different number of observations at a given interval and too many boundary observation. Coefficients (relative indices) for each quarter and each separate lactation were proposed as the optimal procedure. The final result of our study and our random sample is then a table of milk yield indices. This table facilitates comparison of milk yield of individual animals from different lactation and years.

So time series of milk yield can be described by model with trend and seasonal components

$$Y = a + bt + \alpha_1 q_1 + \alpha_2 q_2 + \alpha_3 q_3 + \alpha_4 q_4, \tag{7}$$

where q_1 is 1 only for first quarter, for another quarters of year is zero, where q_2 is 1 only for second quarter, for another quarters is zero, where q_3 is 1 only for third quarter, for another quarters is zero, where q_4 is 1 only for fourth quarter, for another quarters is zero. Analysis of milk yield time series show how trend and seasonal factor change 305-d milk yield. See Fig. 2 and Fig. 3.

Based on such time series we can directly make corrections of estimated yield for individual cows based on its classification into time interval. This leads to the possibility to relate yield to the selected period, which will be evaluated in series of milk yield indexes by value 1. Each additional period then has its own value index, which measures the relative change from baseline period in milk yield periods.

2 Results and Discussion

2.1 Trend in period 2005–2013

Obtained estimates of 305 day yield are depicted in Fig. 2. On axis x is date of calving. t-test for hypothesis of nonzero trend provide us information that trend is positive. The result obtained from ANOVA model simply provides information about existence of seasonal factors. See Table 3.

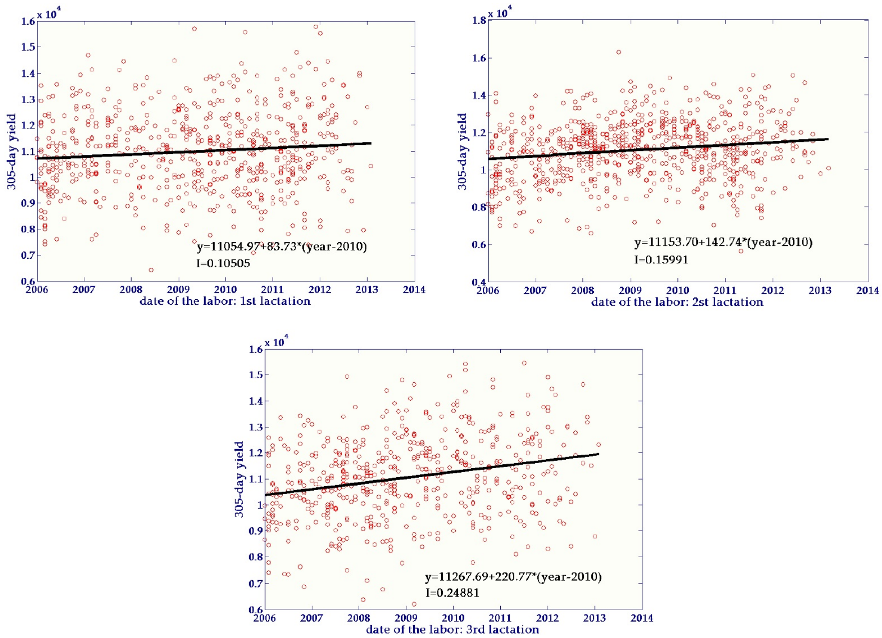

Figure 2 Time series of 305 d yield estimators between 2006 and 2013, 1st, 2nd, and 3rd lactation

A construction of graph in Fig. 3 is made on model (7). Intended milk yield indexes are presented in the Tab. 4. As an initial period with an index value 1 was chosen the first quarter of 2006.

first	second	third	fourth
10714	10005	10111	10743

Table 3 The mean values of milk-yield for quarters of year, 1st lactation

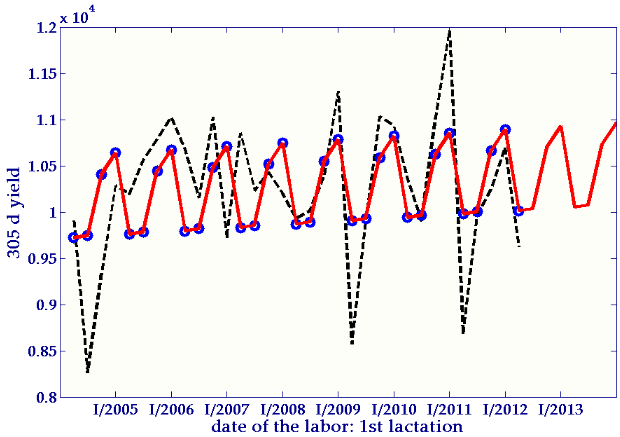

Figure 3 Estimates of trend and seasional component in 305 d yield in 2005 – 2012, 1st lactation

quarter/year	mean	n	s	index
III/2005	8258.238	180	522.13794	0.803045261
IV/2005	9340.996	204	282.69537	0.908334510
I/2006	10283.652	227	154.28539	1.000000000
II/2006	10193.088	186	156.32336	0.991193401
III/2006	10560.219	163	347.72766	0.991193401
IV/2006	10787.2	120	343.05917	1.048965873
I/2007	11027.059	138	665.71650	1.072290175
II/2007	10674.728	182	925.85370	1.038028903
III/2007	10154.052	184	662.67224	0.987397473
IV/2007	11024.403	212	389.96655	1.072031901
I/2008	9715.876	204	479.19580	0.944788486
II/2008	10853.934	184	506.42737	1.055455202
III/2008	10232.292	173	349.37335	0.995005665
IV/2008	10428.035	173	581.23500	1.014040051
I/2009	10198.219	151	825.39960	0.991692348
II/2009	9926.858	167	372.90256	0.965304738
III/2009	10016.76	169	461.18213	0.974046963
IV/2009	10400.896	182	406.96500	1.011401008
I/2010	11305.579	179	163.22684	1.099373938
II/2010	8572.864	159	1117.67100	0.833640034
III/2010	9900.303	198	164.36745	0.962722484
IV/2010	11031.356	145	315.29016	1.072708022
I/2011	10927.135	211	420.09340	1.062573393
II/2011	10353.197	184	274.22632	1.006762675
III/2011	9904.891	137	1138.0745	0.963168629
IV/2011	11021.393	176	848.06330	1.071739203
I/2012	11962.039	148	421.92883	1.163209237
II/2012	8680.581	103	981.57153	0.844114620
III/2012	9960.138	111	1048.89110	0.968540942
IV/2012	10243.146	78	250.45737	0.996061127

Table 4 The mean values of milk-yield for quarters of year, 1st lactation

3 Conlusion

When processing limited data sets or comparing individual milk cows, it is necessary to have the total 305-day yields stripped off selected factors. According to literature, a Wood function is considered as the most appropriate model for modelling a 305-day yield. Testing the statistical hypothesis indicated the existence of nonzero trend and seasonal variation. Gained knowledge allows to propose a method of 305-day yield corrections by factors of lactation and calving data. Corrections are made using the index table for years and quarters. The calculated milk yield indexes in a given cowshed can help producers when comparing cows and making decisions.

Acknowledgements

This research was supported by the institutional support of Czech University of Life Sciences and by Internal Grant Agency of University of Pardubice, the project SGS FEI_26_2016.

References

[1] Golebiewski, M., Brzozowski, P., and Golebiewski, L.: Analysis of lactation curves, milk constituents, somatic cell count and urea in milk of cows by the mathematical model of Wood. *Acta Veterinaria* Brno **1** (2011), 3-80.

[2] Leon-Velarde, C.,U., McMillan, I., Gentry, R. D., and Wilton, J. W.: Models for estimating typical lactation curves in dairy cattle. *Journal of Animal Breeding and Genetics* **112** (1995), 333-340. doi: 10.1111/j.1439-0388.1995.tb00575.x

[3] Marek, J., Rajmon, R., and Haloun, T.: Critical evaluation of seven lactation curve estimation models. In: *Intelligent Data Analysis and Applications. Proceedings of the Second Euro-China Conference on Intelligent Data Analysis and Applications, ECC 2015* (Abraham Ajith, Jiang Xin Hua, Snasel Vaclav, Pan Jeng-Shyang, eds.). Springer International Publishing, New York, 2015, 73-84. ISBN: 978-3-319-21206-7

[4] Wimmer, G., Palencar, R., Witkovsky, V.: *Spracovanie a vyhodnocovanie merania.* VEDA, Bratislava, 2002. p. 1-187. ISBN 80-224-0734-8.

[5] Wood, P.D.P.: Algebraic model of the lactation curve in References cattle. *Nature* **216** (1967), 164-165.

[6] Zavadilova, J.: Definition of subgroups for fixed regression in the test-day animal model for milk production of Holstein cattle in the Czech Republic. *Czech J. Anim. Sci.* **50**, vol. 1 (2005), 7-13.

[7] Zavadilova, J.: Genetic parameters for test-day model with random regressions for production traits of Czech Holstein cattle. *Czech J. Anim. Sci.* **50**, vol. 4 (2005), 142-150.