Bayesian Hierarchical Modeling Applied to Test Day Milk Yield Data

E. Karaman[1], E. Kaya Başar[1], M.Z. Firat[1]

[1] Department of Animal Science, Akdeniz University, TURKEY.
(E-mail: emrekaraman@akdeniz.edu.tr, ebrukaya@akdeniz.edu.tr, mzfirat@akdeniz.edu.tr)

ABSTRACT

Fitting non-linear models to the milk yield records of dairy cows is of particular importance for both breeding and management purposes as they provide a summary of the pattern of milk production with a few parameters that are biologically interpretable in animal science. In this study, hierarchical modeling of the test day milk yield data collected on monthly interval was considered and the non-linear three parameter Wood's incomplete gamma function was used to obtain the lactation curves of 200 Holstein cows. Bayesian approach was utilized in estimation of model parameters. Some characteristics of lactation such as persistency, peak time and peak yield were also derived as well as total milk yield for each cow.

Keywords: Lactation, Wood's function, Bayesian approach.

1. INTRODUCTION

The genetic evaluation in dairy breeding schemes for milk production has depended for many decades on the analysis of 305-day lactation milk production which is based on a set of individual test day (TD) yields, the so called test day milk yields (TDMY), usually taken on a monthly interval during lactation period [1]. The graph of TDMY against time is the lactation curve, and many mathematical functions were provided to summarize the whole process in terms of a few parameters that have biological meaning. The non-linear incomplete gamma function proposed by Wood [2] has been considered by various researchers and fitted directly in its non-linear form [3] or after a log transformation [4].

Individual deviations from the average curve, or the deviations of the parameter vector of each individual from the average of the population should be taken into account not only in modeling the lactation curves, but also in the growth curves of livestock species. As such, both the breeding and management decisions can be made more efficiently. Recent progresses in theory and computing technology allow for the consideration of this approach through the use of non-linear mixed effect models. There are some reports on the use of non-linear mixed modeling in broilers [5], in quail [6,7,8], in pigs [9], and in cattle [10,11].

Bayesian modeling arises with its own hierarchical structure where the prior distribution for the unknown parameters $\boldsymbol{\theta}$ is defined with its prior parameters $\boldsymbol{\alpha}$ indicating a one level of hierarchy. In order to capture the complex structure of the data set, a prior can be constructed

to yield a series of conditional distributions called hierarchical stages of the assumed prior distribution [12]. That is, in this type of modeling, the observations are given distributions conditional on parameters, and the parameters in turn have distributions conditional on additional parameters called hyperparameters [13]. In this respect, Bayesian hierarchical modeling is a class of modeling methodology where a prior distribution is also attributed on the prior parameters $\boldsymbol{a}$.

In this study, we are concerned with the mixed effects modeling of lactation curves in dairy cows through the use of Bayesian hierarchical modeling methodology. Correlations among model parameters were also considered and some characteristics of lactation such as persistency, peak time and peak yield were derived.

2. MATERIAL AND METHODS

2.1. Data

The data of this study is a subset of the data used previously in [1,14,15]. Ten TD records belonging to 200 cows, whose first measurement was taken at 15th day after calving, were used in modeling the lactation curves by Wood's function. Some descriptive statistics of the data is given in Table 1. The pattern of changes in milk yields over 10 TDs depicts the main feature of lactation curves of dairy cows with a rapid increase in production reached a peak at TD2 followed by a decrease up to TD10.

Table 1. Some descriptive statistics of the data set.

	N	Mean	SD	Min	Max
TD1	200	20.85	2.926	12.0	28.0
TD2	200	23.96	2.871	16.4	31.4
TD3	200	22.82	3.007	14.8	30.4
TD4	200	21.33	2.993	12.8	28.8
TD5	200	19.91	2.803	11.6	27.0
TD6	200	18.50	3.045	9.6	25.2
TD7	200	17.53	3.143	8.8	26.6
TD8	200	16.52	3.445	8.2	24.8
TD9	200	15.19	3.535	6.6	23.4
TD10	200	13.25	3.413	4.0	22.2

2.2. The Wood's (Incomplete Gamma) Function

The Wood's incomplete gamma function [2] has been used widely for the lactation curve of dairy cows as it has the advantage of having a limited number of parameters with assessed direct reference to main features of lactation curve [10]:

$$y_j = \beta_0 t_j^{\beta_1} e^{\beta_2 t_j}$$

where, y_j is the milk yield (kg) at TD t_i, e is the Neper number (2,718), β_0 is the initial milk yield after calving, β_1 is the inclining slope parameter up to peak yield (PY), and β_2 is the declining slope parameter after PY. The model estimates a PY, the highest recorded test day milk yield, of $\beta_0(\beta_1/\beta_2)^{\beta_1}e^{-\beta_2}$, which occurs β_1/β_2 months after calving (peak time-PT)

[16]. As proposed by Wood [2] a measure of the distribution of yield over the lactation, persistency (P), is obtained as $-(\beta_1 + 1)ln(\beta_2)$. The area under the curve of *i*'th cow represents the estimated total milk yield [4]:

$$TY_i = \beta_0 \int_0^j t^{\beta_1} e^{\beta_2 t}\, dt$$

2.3. Bayesian Method of Estimation

Any non-linear model for the *i*'th individual with n_i records can be given as:

$$y_{ij} = f(\boldsymbol{\theta}_i, \boldsymbol{x}_{ij}) + \varepsilon_{ij}$$

y_{ij} is the *j*'th response for the *i*'th individual, $\boldsymbol{x}_{ij}$ is the predictor vector for the *j*'th response of *i*'th individual, f is a non-linear function of the predictor vector and a parameter vector $\boldsymbol{\theta}_i$ of length *r*, and ε_{ij} is normally distributed noise term [17]. Replacing the Wood's function in the above formula, we obtain:

$$y_{ij} = \beta_{0i} t_{ij}^{\beta_{1i}} e^{\beta_{2i} t_{ij}} + \varepsilon_{ij}$$

where ε_{ij} is the random error, t_{ij} is the TD, and y_{ij} is the milk yield at t_{ij} for $j = 1, \ldots n_i$. Errors are assumed to be uncorrelated and normally distributed with mean zero and a constant variance, $\varepsilon_{ij} \sim N(0, \sigma_e^2)$. Parameter vector $\boldsymbol{\theta}$ for the *i*'th cow can be written as:

$$\boldsymbol{\theta}_i = (\beta_0 \;\; \beta_1 \;\; \beta_2)'$$

The likelihood of the data can now be written as follows:

$$l(\boldsymbol{y}_i | \boldsymbol{\theta}_i, \sigma_e^2, \boldsymbol{x}_i) = \frac{1}{(2\pi)^{\frac{1}{2}n_i} \sigma_e^{n_i}} exp\left\{ -\frac{1}{2\sigma_e^2} \sum_{j=1}^{n_i} \left[y_{ij} - f(\boldsymbol{\theta}_i, \boldsymbol{x}_{ij}) \right]^2 \right\}$$

The lactation curve of any cow can be seen a realization of a random variable $\boldsymbol{\Theta}$ having a posterior density function $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ on the population of cows. We can further assume that the distribution of $\boldsymbol{\theta}_i = (\beta_0 \;\; \beta_1 \;\; \beta_2)'$ is multivariate normal yielding [18],

$$h(\boldsymbol{\theta} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{3}{2}} (|\boldsymbol{\Sigma}|)^{\frac{1}{2}}} exp\left\{ -\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}) \right\}$$

Where $\boldsymbol{\mu}$ is a column vector of length *r* and $\boldsymbol{\Sigma}$ is a symmetric and positive definite matrix with dimensions $r \times r$. For conjugacy, $\boldsymbol{\mu}$ was assumed to follow a Multivariate Normal distribution while $\boldsymbol{\Sigma}$ was assumed to follow an Inverted-Wishart distribution with parameters were set as follows:

$$h(\boldsymbol{\mu}) \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1.0 \times 10^6 & 0 & 0 \\ 0 & 1.0 \times 10^6 & 0 \\ 0 & 0 & 1.0 \times 10^6 \end{bmatrix} \right)$$

$$h(\mathbf{\Sigma}) \sim IW\left(\begin{bmatrix} 1.0\times10^{-3} & 0 & 0 \\ 0 & 1.0\times10^{-3} & 0 \\ 0 & 0 & 1.0\times10^{-3} \end{bmatrix}, 3\right)$$

We further assumed an inverse Gamma prior for σ_e^2 with $v = s = 0.001$.
Analyses were performed using WinBUGS software [19] through the R2WinBUGS package of R program [20]. Chain length of 500,000 cycles was considered with a burn-in period of 50,000 cycles. Thinning interval was set to 250, and in total 1800 samples were used to estimate the features of marginal distributions.

3. RESULTS

A model was fitted taking into account the individual data, however, only the summary statistics for the posterior distributions of parameters and derived quantities (P, PY, and PT) in the population level are presented in Table 2. The posterior mean of the parameter β_0, which is related with initial milk yield, was found as 18.278. Denoting the inclining slope of the lactation curve, posterior mean of parameter β_1 was found to be 0.094 while the posterior mean of declining slope parameter β_2 was found to be 0.003. Posterior means for P, PY and PT were found to be 6.458, 23.213 and 34.371, respectively. In addition, posterior mean of the error sum of squares, σ_e^2, was estimated to be 3.816.

In Bayesian approach, frequentist confidence intervals are replaced by Bayesian credible intervals (BCI). The interpretation of a 95% BCI is that the probability that the parameter falls into the given interval is 0.05. Also, the highest probability density interval (HPDI), the shortest possible interval enclosing 95% of the posterior mass, is the choice if the posterior distribution is an asymptotic one [21]. Both BCIs and HPDs in Table 2 almost indicate the same interval for the parameters of interest.

Table 2. Posterior summary statistics of the parameters, P, PY and PT.

			BCI		HPDI	
	Mean	SD	2.5%	97.5%	2.5%	97.5%
β_0	18.278	0.868	16.950	19.650	16.940	19.630
β_1	0.094	0.014	0.072	0.117	0.072	0.118
β_2	0.003	0.000	0.002	0.003	0.002	0.003
P	6.458	0.073	6.314	6.603	6.311	6.597
PY	23.213	0.219	22.790	23.650	22.790	23.640
PT	34.371	4.114	26.919	41.981	26.400	41.260
σ_e^2	3.816	0.199	3.463	4.215	3.459	4.206

Estimated P, PY and PT of the top %20 of cows with the highest TYs are given in Table 3. A close examination of Tables 2 and 3 reveals that the cows with a higher milk production over the herd also have a peak yield occurring later than the average time of peak (34.371) of the herd. The P values for the cows in Table 3 have higher persistence than the average of the herd (6.458), except the “Cow 87”. For management purposes, cows with persistent lactations are more profitable than cows having lactations with low persistency when the TYs are equal.

Table 3. Estimated P, PY and PT of the top %20 of cows with the highest TYs.

Cow	P	PY	PT	TY	Cow	P	PY	PT	TY
6	6.90	27.6	58.7	7401.2	102	6.74	26.0	49.2	6793.5
83	6.93	27.2	59.5	7335.0	119	6.80	25.5	50.5	6751.5
23	7.03	26.6	63.4	7310.1	3	6.68	26.1	47.1	6726.9
88	7.10	26.3	66.7	7293.0	85	6.83	25.1	51.2	6725.7
137	6.94	26.9	59.7	7279.0	170	6.63	26.5	45.8	6722.3
184	6.75	27.7	52.4	7209.0	49	6.63	26.4	45.8	6713.3
90	6.72	28.0	51.1	7206.4	48	6.55	27.0	43.7	6694.8
136	7.00	25.8	60.2	7070.8	11	6.50	27.5	42.5	6679.4
36	6.84	26.3	54.0	7022.8	40	6.92	24.6	53.5	6675.4
93	7.34	24.6	76.2	7004.4	100	6.49	27.6	42.1	6673.6
190	7.12	24.9	64.2	6951.9	78	6.61	26.3	44.9	6662.1
84	6.84	26.0	53.2	6929.8	33	6.60	26.3	44.3	6624.7
107	6.89	25.6	54.9	6911.2	96	6.64	25.8	45.2	6602.3
165	7.30	24.3	76.4	6909.2	7	6.58	26.2	43.6	6584.5
18	6.86	25.6	53.7	6880.3	86	6.57	26.2	43.2	6563.8
13	7.12	24.6	63.0	6871.2	87	6.41	27.9	39.9	6551.8
161	6.81	25.8	51.6	6852.7	47	6.73	24.9	46.5	6525.3
172	7.17	24.3	65.5	6843.3	37	6.70	24.9	45.7	6501.4
71	6.74	26.0	49.3	6801.1	120	6.93	23.8	51.4	6482.6
97	6.89	25.2	53.7	6795.8	59	6.89	23.9	50.0	6475.0

CONCLUSIONS

In conclusion, the Bayesian modeling of lactation curves, both in herd and individual level simultaneously, allowed us to evaluate several characteristics of the lactation as well as to investigate the usefulness of the methodology in making decisions governing the breeding and management. Model and methodology used here can be extended to include not only the phenotypic relationships, but also the genotypic relationships between parameters.

REFERENCES

[1] M.Z. Firat, C.M. Theobald, R. Thompson, 1997. Univariate analysis of test day milk yields of British Holstein Friesian heifers using Gibbs sampling. *Acta Agriculturae Scandinavica, Section A–Animal Science*. 47: 213-220.

[2] P.D.P. Wood, 1967. Algebraic model of the lactation curve in cattle. *Nature*. 216: 164-169.

[3] A.M. Silvestre, A.M. Martins, V.A. Santos, M.M. Ginja, J.A. Colaço, 2009. Lactation curves for milk, fat and protein in dairy cows: A full approach. *Livestock Science*. 122: 308–313

[4] Tekerli, M., Akinci, Z., Dogan, I., Ackan, A., 2000. Factors affecting the shape of lactation curves of Holstein cows fromthe Balikesir province of Turkey. *Journal of Dairy Science*. 83: 1381–1386.

[5] Z. Wang, M.J. Zuidhof, 2004. Estimation of growth parameters using a nonlinear mixed Gompertz model. *Poultry Science*. 83: 847–852.

[6] K. Kizilkaya, M.S. Balcioglu, H.I. Yolcu, K. Karabag, and I.H. Genc, 2006. Growth curve analysis using nonlinear mixed model in divergently selected Japanese quails. *Archiv fur Geflugelkunde*. 70: 181-186.

[7] S.E. Aggrey, 2009. Logistic nonlinear mixed effects model for estimating growth parameters. *Poultry Science*. 88: 276-280.

[8] E. Karaman, D. Narinc, M.Z. Firat, T. Aksoy, 2013. Nonlinear mixed effects modeling of growth in Japanese quail. *Poultry Science*. 92: 1942-1948.

[9] A.B. Strathe, A. Danfær, H. Sørensen and E. Kebreab, 2010. A multilevel nonlinear mixed-effects approach to model growth in pigs. *Journal of Animal Science*. 88: 638-649.

[10] R. Rekaya, M.J. Carabano, M.A. Toro, 2000. Bayesian Analysis of Lactation Curves of Holstein-Friesian Cattle Using a Nonlinear Model. *Journal of Dairy Science*. 83: 2691–2701.

[11] S. Forni, M. Piles, A. Blasco, L. Varona, H.N. Oliveira, R.B. Lôbo, L.G. Albuquerque, 2009. Comparison of different nonlinear functions to describe Nelore cattle growth. *Journal of Animal Science*. 87(2): 496-506.

[12] I. Ntzoufras. Bayesian Modeling Using WinBUGS: An introduction. Wiley Series in Computational Statistics. NJ, USA, 2009.

[13] J. Albert. Bayesian Computation with R. Springer. NY, USA, 2009.

[14] M.Z. Firat, C.M. Theobald, R. Thompson, 1997. Multivariate analysis of test day milk yields of British Holstein Friesian heifers using Gibbs sampling. *Acta Agriculturae Scandinavica, Section A–Animal Science*. 47: 221-229.

[15] B. Simsek, M.Z. Firat, 2011. Application of multilevel analysis in animal sciences. *Applied Mathematics and Computation*. 218(3): 1067–1071.

[16] L. Sherchand, R.W. McNew, D.W. Kellog, Z.B. Johnson, 1995. Selection of Mathematical Model to Generate Lactation Curves Using Daily Milk Yields of Holstein Cows. *Journal of Dairy Science*. 78: 2507-2513.

[17] M.J. Lindstrom, D.M. Bates, 1990. Nonlinear Mixed Effects Models for Repeated Measures Data. *Biometrics*. 46: 673-687.

[18] C.S. Berkey, 1982. Bayesian approach for a nonlinear growth model. *Biometrics*. 38(4) :953-61.

[19] D. Lunn, A. Thomas, N. Best, D.J. Spiegelhalter, 2000. WinBUGS-a Bayesian modeling framework: concepts, structure, and extensibility. *Statistics and Computing*. 10: 325-337.

[20] R Development Core Team. 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

[21] P. Waldmann, T. Ericsson, 2006. Comparison of REML and Gibbs sampling estimates of multi-trait genetic parameters in Scots pine. *Theoretical and Applied Genetics*. 112:1441-1451.