Genetic Analysis of Daily Maximum Milking Speed by a Random Walk Model in Dairy Cows

B. Karacaören[1], L. L. G. Janss[2], H. N. Kadarmideen[3]

[1]Akdeniz University, Turkey, [2] Aarhus University, Denmark,[3]University of Copenhagen, Denmark

ABSTRACT: Data were obtained from dairy cows stationed at research farm ETH Zurich for maximum milking speed. The main aims of this paper are a) to evaluate if the Wood curve is suitable to model mean lactation curve b) to predict longitudinal breeding values by random regression and random walk models of maximum milking speed. Wood curve did not provide a good fit to the data set. Quadratic random regressions gave better predictions compared with the random walk model. However random walk model does not need to be evaluated for different orders of regression coefficients. In addition with the Kalman filter applications: random walk model could give online prediction of breeding values. Hence without waiting for whole lactation records, genetic evaluation could be made when the daily or monthly data is available.
Keywords: Kalman Filter; Gibbs sampling; Random Walk model; Breeding value estimation

Introduction

Kalman Filter was used in navigation software of the Apollo that put the Neil Armstrong on the moon. Kalman Filter could be useful to track the signal from error contaminated serial observations. Under optimal conditions Kalman Filter leads to obtain online prediction of random effects without storing and inverting data matrices. Karacaören (2006) and Karacaören et al. (2012) defined the Kalman Filter in state space form for predicting longitudinal breeding values of body condition scores over days in milk of dairy cows by an animal model. Forni et al. (2009, 2009) used a dynamic linear model via Kalman Filter for genetic analyses of longitudinal traits.

Incomplete gamma curve was proposed by Wood (Wood, (1967)), probably the most common choice to model test day (TD) measurements for milk yield, although different models showed advantages over the Wood function; but they do require more parameters to be estimated (Macciotta etal. (2004)). Milking speed is an important trait in dairy cattle production and it has been shown to be influenced by strong genetic factors in Swiss dairy cattle (Ilahi and Kadarmideen (2004)). Using cross-sectional Bayesian models, they showed that heritabilities of milk flow in both Brown Swiss and Simmental to be high (0.44 to 0.48) and in Swiss Holsteins to be intermediate (0.25). In practice, milking speed is often measured by subjective scoring whereas in the present study maximum milking speed (MAXMS) was measured electronically on daily basis. It's characteristic curve has not been studied in detail due to difficulty of collection of objective measurements.

The main aims of this paper are a) to evaluate if the Wood curve is suitable to model mean lactation curve of MAXMS and b) to predict longitudinal breeding values for MAXMS by random regression (Karacaören et al. (2006)) and random walk-kalman filter models (Karacaören et al. (2012)).

Materials and Methods

Data. Data (*N*=72) were obtained from first lactation records of dairy cows stationed at the Chamau research farm of the Swiss Federal Institute of Technology, Switzerland over the period of April 1994-2004. MAXMS were recorded daily using automated units by METATRON (American Calan Inc., North The Wood, NH, USA). The animals were housed in a free-stall barn. The concentrate, roughage, minerals andvitamins were fed according to calculated needs (Karacaören et al. (2012)). Data samples were created based on different sampling frequencies and named as scheme 1 to scheme 5; daily sample (scheme 1), 10 day sample (scheme 2), 20 day sample (scheme 3), 30 day sample (scheme 4) and 60 day sample (scheme 5). Mean lactation curves based on daily samples are given in Figure 1 for MAXMS.
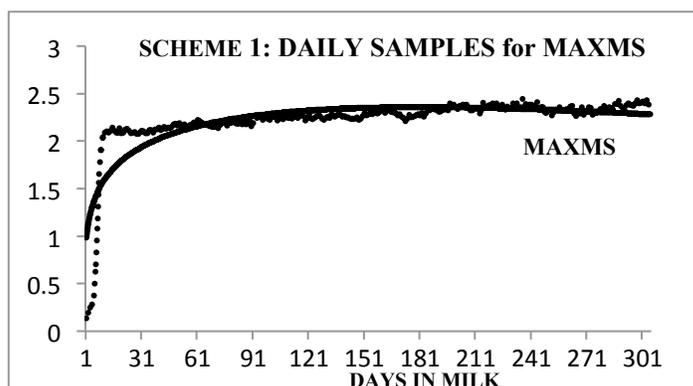

Figure 1 Actual (.) and estimated (-) average lactation curve for scheme 1(daily sampling) for maximum milking speed.

Statistical analyses. We fitted the Wood function with and without logarithmic transformation to obtain parameter estimates for MAXMS. The Wood function with

the associated logarithmic transformation could be written as (Macciotta etal. (2004))

$$\log\big(y(t)\big) = \log(a) + b\log(t) - ct$$

where $y(t)$ is the average daily production at time t, and a is an estimate of an initial milk yield, b is a measurement of the rate of milk yield increase until the peak and c is a measurement of the rate of decreasing milk yield after the peak. Using estimated parameters following properties of the lactation curves could be predicted;

the lactation persistency, p, $p=-(b+1)*\ln(c)$,

the time at which the peak production is attained t_m, $t_m=b/c$,

the peak yield $y_m = a(t_m)^b e^{-b}$ (3).

We estimated the linearized version of the wood function by PROC REG (SAS, (2013)) and non linear version by PROC(NLIN) using Marquardt iterative method. In order to check if there is any correlation among residuals over DIM Durbin-Watson statistics (D-W) were compared. Model fit assessed by visual inspection of residuals and by adjusted determination of coefficient (Adj.R-Square). In nonlinear models we investigated the 95% confidence intervals and correlations between the estimates of parameters.

Random Regression Model. For MAXMS, individual deviations from the mean lactation curve were fitted for additive genetic effects and permanent environmental effects using the random regression methodology (Hadfield, (2010)). The animal model fitted was as follows;

$$y_{tij} = f + \sum_{l=0}^{2} \mu_{tl} a_{jl} + \sum_{l=0}^{2} \mu_{tl} p_{jl} + e_{tij} \quad (1)$$

where y_{tij} is MAXMS produced by the jth cow on day t. f consist of fixed effects including season, age at calving (in months), and year-season. Terms a_{jl} and p_{jl} are random regression coefficients for the genetic effects and permanent environmental effects respectively; μ_t is the vector of the three first legendre polynomial coefficients made on day t and e_{tij} is the random residual variance (assumed fixed over lactation).

Random Walk Model: For genetic analyses of traits following mixed model is normally used in animal breeding;

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z_a a} + \mathbf{Z_p p} + \mathbf{e} \quad (2)$$

where **y** is the vector of observations, $\boldsymbol{\beta}$ is the vector of fixed effects, **a** is the vector of breeding values, **p** is the vector of random permanent environmental effects, $\mathbf{X}$, $\mathbf{Z_a}$, $\mathbf{Z_p}$ are design matrices and **e** is the vector of random residual effects.

For the random effects it was assumed that

$$Var\begin{pmatrix}\mathbf{a}_0\\ \mathbf{p}_0\\ \mathbf{e}_0\end{pmatrix} \sim N\left[\mathbf{0}; \begin{pmatrix}\mathbf{A}\sigma_a^2 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\sigma_p^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_e^2\end{pmatrix}\right],$$

$$Var\begin{pmatrix}\mathbf{a}_t\\ \mathbf{p}_t\\ \mathbf{e}_t\end{pmatrix} \sim N\left[\begin{pmatrix}\mathbf{a}_{t-1}\\ \mathbf{p}_{t-1}\\ \mathbf{e}_{t-1}\end{pmatrix}; \begin{pmatrix}\mathbf{A}\sigma_a^2 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\sigma_p^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_e^2\end{pmatrix}\right]$$

where σ_a^2, σ_p^2, and σ_e^2 are genetic, permanent environment and error variances. $\mathbf{A}$ is the additive genetic relationship matrix for the animals; $\mathbf{I}$ is an identity matrix. Details of the random walk-kalman filter model could be found in (Karacaören (2006); Karacaören et al. (2012)).

Results and Discussion

Wood Curve for Maximum Milking Speed. Mean of MAXMS was 2.10 kg/min with s.e. of 0.12. Estimate of persistency, p, ranged from 7.02 to 9.07 (table 1). Correlation of estimates of parameters for Wood function found to be generally high and ranged from -0.97 (a and b) to 0.81 (a and c) and it was -0.91 among b and c for scheme 1. Although residual variance was smaller (0.03) compared to total variance (508.5) in scheme 1, auto correlations were still found among residuals (figure is not shown) and this was observed for other schemes as well. In scheme 1 (daily samples) peak production was estimated about day 182 with correspondent peak yield as 1.94(kg/min), and observations showed peak at day 236 and 2.39 (kg/min) of peak yield. Due to nonlinearity we investigated the approximate 95% confidence intervals of the estimates instead of adjusted determination of coefficients, and they were found reasonable as estimated for a 0.98(0.89-1.08), for b 0.21(0.18-0.23) and for c -0.0011(-0.0012-(-0.0013)). In scheme 2 (10 daily samples) peak production was estimated about day 181 with correspondent peak yield as 1.94(kg/min), observations showed peak at day 109 with 2.40(kg/min). The approximate 95% confidence intervals of the estimates of parameters were found reasonable as estimated for a 0.89(0.54-1.24), for b 0.24 (0.12-0.34) and for c -0.001 (-0.002-(-0.0002)). In scheme 3 (20 daily samples) peak production was estimated by the model as about day 170 with correspondent peak yield as 2.13(kg/min), and

observations showed peak at day 109 with 2.40(kg/min). The approximate 95% confidence intervals of the estimates of parameters estimated for a was0.5(0.16-0.95), for b was 0.36 (0.16-0.55) and for $c$was-0.0021(-0.0038-(-0.00018)). In scheme 4 (monthly samples) peak production was estimated by the model as about day 172 with correspondent peak yield as 2.23(kg/min), and observations showed peak at day 109 with 2.30(kg/min). Approximate 95% confidence intervals for the function of the Wood parameters estimated as a was 0.41(0.005-0.82), for b was 0.43(0.15-0.70) and for c was-0.0024(-0.005-(-0.00013)) and found to be larger compared with more frequent sampling schemes. In scheme 5 (60 daily samples) peak production was estimated by the model as about day 185 with correspondent peak yield as 2.47(kg/min), observations showed peak at day 305 with 2.40(kg/min). The approximate 95% confidence intervals of the estimatedparameters were as follows: for a was 0.23(-0.25-0.71), for b was 0.56(0.2-1.10) and for c -was 0.003(-0.007-0.001).

Scheme	P	t_m	y_m
Scheme 1	7.02	182	1.94
Scheme 2	7.17	181	1.94
Scheme 3	7.88	170	2.13
Scheme 4	8.31	172	2.23
Scheme 5	9.07	185	2.47

Table 1 Various statistics (lactation persistency, $p_,,$the time at which the peak production is attained,t_m, the peak yield, y_m,) estimated by the wood function for maximum milking speed by different sampling schemes.

Estimates of MAXMS had high correlation; ranged from -0.97 (a and$b$) to 0.75 (a and c) and it was -0.88 among b and c for scheme 1. Confidence intervals were also found to be narrowed by the increment of sampling size from scheme 1 to scheme 5. However its characteristic curve did not show linear decrease and some fluctuations were also observed over DIM after the peak. Due to variation over DIM after the peak its characteristics was not predicted correctly by none of the schemes using the Wood function.

Random regression analyses. For random regression model; the markov chain was run for 13.000 iterations with a burn-in period of 3.000 and a thinning interval of 10 iterations. Default uninformative priors were employed. We performed model comparisons based on deviance information criterion (DIC) using intercept, first degree and quadratic legendre random regressions for genetic and permanent environmental effects. Due to estimability problems (for quadratic random regressions) we did not include any fixed effects apart from intercept to the all models. DIC value was lower for quadratic random regressions (DIC= 55113.88) compared with intercept (60794.9) and first degree (58834.76) random regressions

models. We also estimated correlations between observations and predictions. Again correlations were found to be higher for quadratic random regressions (0.76) compared with first degree (0.73) and intercept (0.72) random regression models. Based on quadratic random regression model: estimates of heritabilities varied from 0.50 to 0.64 ($\bar{h}^2$ =0.60). We found the heritability to be the highest around middle of the lactation.

Random walk analyses. For the random walk model the markov chain was run for 1.000 iterations with a burn-in period of 100. Estimates of heritability were found to be around 0.10. Correlations between observations and predictions were found to be 0.50. Compared to Ilahi and Kadarmideen (2004) who reported heritability in the range of 0.45, Berry et al. (2013) found the heritability for maximum milk flow rate as 0.17 and Edwards et al. (2013) as 0.31. Kramer et al. (2013) compiled the estimates of heritabilities from various studies and concluded that ranged from moderate to high when it is measured electronically. However due to estimability problems we did not include and fixed effects into the random regression model apart from the intercept. In addition; for comparison purposes (with the random walk model) we assumed the residuals were constant over lactation. The differences concerning the experimental design, number of animals and statistical methodology may explain the differences between results of current and the other studies.

Conclusion

Wood curve did not provide a good fit to the mean lactation curve for MAXMS data set. Quadratic random regressions gave better predictions compared with the random walk model. However random walk model does not need to be evaluated for different orders of regression coefficients. In addition with the Kalman Filter applications: random walk model could give online prediction of breeding values. Hence without waiting for whole lactation records genetic evaluation could be made when the daily (Figure 2) or monthly data are available.
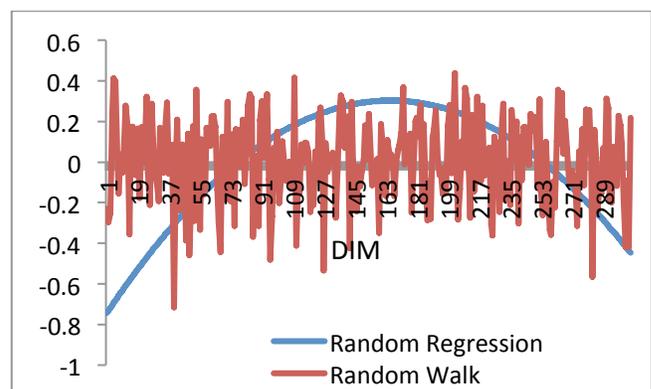
Figure 2: Comparison of longitudinal breeding values obtained by random walk (x10) and quadratic random regression models for an animal 830.

Acknowledgements

Literature Cited

Berry, D. P., Coyne, J., Coughlan, B., et al. (2013). Animal. 7: 1750-1758.

Edwards, J. P., Jago, J. G., Lopez-Villalobos, N. (2014). J. Dairy Sci. 97: 259-269.

Forni, S., Gianola, D. Rosa, G. J. M. (2009). J. Anim. Sci. 87: 3854-3864.

Forni, S., Gianola, D., Rosa, G. J. M., et al. (2009). J. Anim. Sci. 87: 3845-3853.

Hadfield, J D. (2010). J Stat S. 33: 1-22.

Karacaören, B., Jaffrézic F., Kadarmideen H.N. (2006). J. Dairy Sci. 89: 791-798.

Karacaören, B., Janss, L.L.G., Kadarmideen H. N., (2012). Kafkas Uni. Vet. Fak. Derg 18: 627-632.

Karacaören, B.: (2006). Ph D Dissertation. Swiss Federal Institute of Technology Zurich, Zurich, Switzerland.

Kramer, M., Erbe, M., Bapst, B. et al. (2013). J. Dairy Sci. 96: 5954-5964.

Ilahi, H., Kadarmideen, H.N. (2004). Genet Sel Evol. 36: 563-576

Macciotta, P.P.N., Borlino-Cappio, A., Pulina, G. (2004). 97-146.

SAS (2013). SAS User's Guide.SAS Institute Inc., Cary, NC, USA.

Wood, P.D.P. (1967). Nature. (1967). 216:164-165.