ADJUSTING LACTATION CURVES WITH SEGMENTED POLYNOMIALS

Lenira El Faro[1], Lucia Galvão de Albuquerque[2,3] and Luiz Alberto Fries[2,4,5]
[1]Instituto de Zootecnia - EEZ - Ribeirão Preto - SP
[2] Departamento de Melhoramento Genético Animal - FCAV - UNESP - Jaboticabal; [3]CNPq;
[4] GenSys Consultores Associados S/C Ltda. and [5]FAPESP.

SUMMARY
A total of 7239 lactations of Caracu cows were used to adjust lactation curves using segmented polynomials with four quadratic segments (4QS). For the average lactation curve the knots were at 2, 26 and 36 weeks of lactation. The parameters were estimated using nonlinear iterative methods (Gauss-Newton). Adjusted coefficient of determination (R^2_A) and distances between actual and predicted total milk yield (%dTY), as a percentage of the first, were calculated. For the average lactation curve, R^2_A was 99.96% and the %dTY was 0.01%. For individual lactations, 73.8% of them had R^2_A higher than 70% and 84% had %dTY within $\pm$2%. For some groups of lactations this function showed good adherence.
Keywords: mathematical functions, segmented polynomial, milk yield

INTRODUCTION

Fitting lactation curves through regression is used to describe the form of the curve of an animal, its components and for estimating the total milk (TMY) yield from incomplete records. The first mathematical function describing the declining phase of a lactation curve was proposed by Brody et al. (1923) and it is known as "Gaines' function". Wood (1967) included a parameter for the phase of ascension of the curve, and the Incomplete Gamma function (IGF) has been used intensively, in its non-linear or linearized forms, or with modifications.

The adjustment of lactation curves can be done through ordinary polynomials. They are easily estimated by linear regression procedures, but suffer some statistical drawbacks, such as multicollinearity and lack of determination uniformity for the entire domain of the curve. Also, they tend to describe better the behavior of the curve in areas where there is a concentration of points (Schenkel 1989). Another problem is the difficulty in the biological interpretation of the parameters.

An alternative to these problems is the use of spline functions or segmented polynomials, which are defined by Rice (1969), as "segments of polynomials of degree P, connected at some specific points, "knots", in a way to have continuous derivatives of degree P-1." The junction points or knots indicate where a change happens in the behavior of the curve. According to Fuller (1969), these functions present the following desirable properties: are continuous; have continuous first derivatives; easily interpretable; very flexible given their multiphasic components; and are linear on the parameters. Their main disadvantage is that the knots should be known beforehand since the relationship between knots and parameters is non linear.

The objective of this work was to study the possible usefulness of spline functions to adjust lactation curves of Caracu cows.

MATERIALS AND METHODS

There were 7239 lactations of Caracu cows, weekly recorded in the period from 1978 to 1988, with a total of 299 949 records. Feeding was based on pastures and on some level of supplementation during the dry months. Milkings were twice daily, with presence of the calf.

The equation of a 4QS model to describe (average or individual) lactation curves, can be represented by:

$$Y_X = a_0 + a_1X + a_2X^2 + b_2Z_1 + c_2Z_2 + d_2Z_3 \quad \text{where, } Z_i = 0 \quad \text{if } X \leq K_i \text{ and}$$
$$Z_i = (X - K_i)^2 \quad \text{if } X > K_i \quad \text{for } i = 1,\ldots,3.$$

Y_x is the production in week X;
a_0, a_1, a_2, b_2, c_2 and d_2 are the parameters and K_1, K_2 and K_3 are the knots.

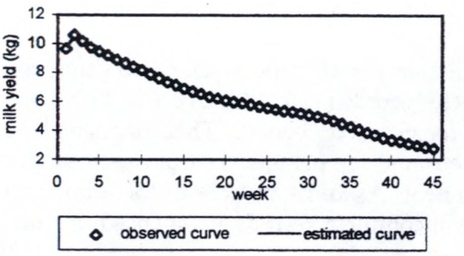
The parameters and knots were estimated using PROC NLIN, SAS, by Gauss-Newton method.

The definition of knots was done initially by inspection, through a diagram of points. Then, values for the second and third knots were fixed to obtain the estimate for the first. The smallest sum of squared residuals indicated the best estimate for the first knot. Then, values for the first and third knots were fixed and the second was estimated; and the same was done for the third point. Some statistics used to evaluate the quality of adjustment were R^2_A and %dTY.

RESULTS AND DISCUSSION

Average Lactation Curve. The 4QS has six parameters: a_0, a_1, and a_2, represent intercept, linear and quadratic regression coefficients for the first segment; and b_2, c_2 and d_2 are differences in quadratic coefficients from one segment to the another. In that way, a_2+b_2, $a_2+b_2+c_2$ and $a_2+b_2+c_2+d_2$ determine the curvatures of the second, third and fourth segments, respectively. Functional and significant changes in the curvature happened in weeks 2, 26 and 36, corresponding to knots K_1, K_2 and K_3.

Observed and estimated average lactation curves are presented in Figure 1. It shows the changes and knots and that the curve for this herd tends to be atypical, similar to what is observed for zebu breeds. That justifies the use of a model with four quadratic segments to explain the form of the curve, instead of models with two segments (such as quadratic-linear or quadratic-quadratic) that could be enough to explain the lactation of Holstein cows. The four segments provided great flexibility to the model, so that the %dTY, shown in Figure 2, were almost null and R^2_A =99.96%. The 4QS function overestimated TMY in just 0.01% and the estimated values for initial and peak production were very close to those observed.
The first quadratic segment determined peak production and time. Persistency was defined as rate of decline of milk production, determined by the linear regression coefficient (a_1).



Y=6,32+4,56X-1,23X2+1,24Z1-0,013Z2+0,015Z3

Figure 1. Lactation curves estimated by the 4QS function

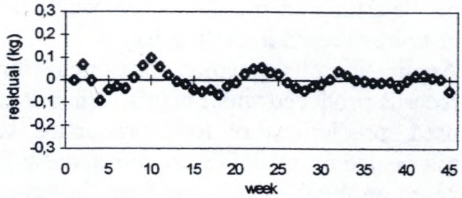


Figure 2. Residuals estimated by the 4QS function.

Individual lactations. The adjustment for the average lactation, in spite of indicating a tendency of the fit of a model, says very little with respect to the goodness of fit for individual lactations.

The 4QS provided a good approximation to the form of the average curve and to the %dTY. However for individual lactations the 4QS approach presented some problems. Estimation of knots is more difficult since each animal possesses its own curve, which changes in different points. Generalization of knots from the average curve to individual lactations produced unrealistic parameter estimates, mainly with respect to a_0, that translates to initial production.

The problem was partially solved with the estimation of knots for basically three lactation groups: low, medium and high levels of production. That allowed for an improvement in the quality of adjustment, because 74% of lactations showed R^2_A above 70%. The %dTY for complete records were within $\pm$2% for 84% of the lactations.

There are several proposals for the estimation of knots when segmented polynomials are used as approximating functions (Schenkel 1989). Gallant and Fuller (1973) propose simultaneous estimation of knots and parameters of the model, using iterative processes, given that the established relationship among them is non linear (it suffices to look at $b_2(X-K_1)^2$). The iterative process suggested by the authors is the Gauss-Newton's modified method, that demands the declaration of the mathematical model, initial values of parameters and first derivatives with respect to the parameters. According to the authors, solutions would be obtained easily for a quadratic-quadratic segmented polynomial. In the present study, with four segments, however, convergence were not obtained.

Usefulness of segmented polynomials. The main application of segmented polynomials seems to be in the description of the shape of lactation curve and its components. Of the total of 7239 studied lactations, 69% were typical, but with peaks happening between first and third weeks and with peak yields close to 12.0 kg. A large proportion of lactations without peak was expected, since the average curve of this herd presented a peak in the second week of lactation. In monthly or biweekly recording systems, the production peak probably would not be estimated and the 'curve' would be a straight line, as it often happens with zebu breeds or *taurus* breeds in the tropics.

Application of segmented polynomials for prediction of total production when using complete records produced small biases. When incomplete records (150, 180, 210 and 240 days) were used, predictions of total production were considerable biased. That happened because segmented polynomials are multiphasic functions and the parameters from each segment are based on the (few) records from the actual segment. As knots are free to 'pick up' curvature changes in each segment, it can happen that incomplete records being used are too close and/or too influential to a knot, producing gross over- or underestimates in predicted TMY.

CONCLUSIONS

Based on the conditions of the present work it can be concluded that:

1. The use of segmented polynomials can be an alternative for the adjustment and description of average lactation curves and the interpretation of its components.
2. The use of 4Q segmented polynomials may not be a plausible alternative for the adjustment of individual lactation curves, since estimating six parameters with a, say, monthly milk control system may not be effectual.
3. The alternative of estimating knots for the lactation groups solved the problem of determining them but the identification of the lactation groups is still difficult and subjective

Final remarks.

1. Other alternatives, using a smaller number of segments (e.g.: Quadratic-Quadratic or Quadratic-Linear), should be tested.
2. Individual curves could be estimated using 'known' knots, since it could be of relevant economic interest to describe the curve in some fixed periods, determined beforehand.
3. An alternative to these problems of adjusting the lactation curves, its form and knots, could be the adoption of Random Regression Models, suggested by Schaeffer (1996). The idea would be to estimate the vector of the parameters for each individual lactation, starting from the average curve parameters, weighted by functions of the vectors of variances of the estimates.

REFERENCES

Brody, S.; Turner, C.W. and Ragsdale, A.C. (1923). *J.Gen.Physiol.*, **5:**441-444.

Drapper, N.M. and Smith, H. (1981). "Applied Regression Analysis". Wiley. New York.

Fuller, W.A. (1969). *Aust. J. Agr. Statistics,* **13:**35-46.

Gallant, A.R. e Fuller, W.A. (1973). *J. Am. Statistics Association*, **68:**140-147.

Rice, J.R. (1969). "The Approximation of Functions". Addison-Wesley. Massachusetts.

Schaeffer, L.R. (1996). "Tópicos Avançados em Melhoramento Animal" UNESP.Jaboticabal

Schenkel, F.S. (1989). "*Polinômios Segmentados*" Seminário. DZ-AGRO-UFRGS.

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