*Sequential minimal optimization regression for the description of dairy sheep lactation with different shape curves*

L. Guevara1, F. Castro-Espinoza2, A. Fernandes1, J. Reyna-Hernández3, M. F. Vital-Hernández3, J. C. Angeles-Hernandez3

*1Centro de Ciências e Tecnologias Agropecuárias, Universidade Estadual do Norte Fluminense, Brasil*

*2Instituto de Ciencias Básicas e Ingeniería, Universidad Autónoma del Estado de Hidalgo, México*

*3Instituto de Ciencias Agropecuarias, Universidad Autónoma del Estado de Hidalgo, México.*

Corresponding author: angeles@uaeh.edu.mx

***Key words*** Artificial intelligence, atypical curve, ewe, machine learning, milk, wood model.

***Application***

To understand the characteristics of ovine lactation is a valuable tool in research and at the herd level. It is useful for estimating total production in lactations with incomplete records, making management decisions, implementing and assessing genetic improvement programs, health monitoring, calculating feeding requirements, and economic-administrative aspects***.***

***Introduction***

In the analysis of lactation characteristics in sheep, different production patterns have been identified: typical and atypical lactations (Macciotta et al., 2005). Although techniques such as mathematical modelling and the implementation of machine learning algorithms allow the determination of lactation characteristics, the variability of lactation patterns can affect the quality of fit of these techniques. For this reason, several studies have suggested that shape of lactations should be analyzed separately. The aim of the study was to evaluate the goodness of fit of the Sequential Minimal Optimization Regression algorithm (SMOreg) and a traditional empirical model in estimating the parameters of typical and atypical lactation curves.

***Material and methods***

*Data base*

A total of 1186 monthly records were used, corresponding to 156 sheep lactations from a commercial farm located in Queretaro, Mexico. Total milk yield (TMY), peak yield (PY) and time to peak yield (TPY) were calculated using the centre-day (Sargent, 1968). Descriptive analysis and outlier screening were performed, resulting in a final database of 104 lactations.

*Mathematical modelling*

Lactation curves were fitted to the incomplete gamma empirical model (Wood, 1967), with model parameters estimated using the 'nlsLM' function from the 'minpack.lm' package in the Rstudio. Estimated parameters were then used to obtain predictions for the lactation traits TMY, PY and TPY. The shape of the lactation curve was identified based on the parameters; when parameters b and c correspond to positive values (b<0 and c<0), a typical curve is found. The other combinations of b and c represent an atypical curve.

*Formulation of the machine learning model*

We chose to implement the Sequential Minimal Optimisation Regression (SMOreg) algorithm. SMOreg is an algorithm for efficiently solving optimization problems that arise when training a support vector machine, and is proposed to deal with regression problems.

*Goodness of fit and statistical analysis*

Model fit was assessed using correlation coefficient (r), mean absolute error (MAE), root mean square error (RMSE), relative absolute error (RAE) and relative root mean square error (RRSE). Two-way ANOVA is used to evaluate the effect on curve shape and prediction technique. The Tukey test used to compare means at the 95% level of significance*.*

***Results***

It was found that 51% of the modelled curves were considered atypical. Equations 1 and 2 below show the estimated mean parameters of the Wood model for typical and atypical curves., respectively.

Y = (0.3001)t(0.5451)*exp*−(0.0127)t  (1)

Y = (7.5698)t(-0.3962)*exp*−(0.0001)t  (2)

According to the goodness of fit criteria, the SMOreg algorithm shows a better performance in the estimation of sheep lactation characteristics, with more significant differences in the fit for atypical curves (Table 1). The estimated values for the milk characteristics are shown in Table 2. An effect of the model used to predict TMY was observed, with the Wood model tending to overestimate this trait. For TPY, the results show no significant differences in the estimation for typical curves. However, for atypical curves, the Wood model tends to underestimate this characteristic. In this study, values with p < 0.1 were considered as trends. Therefore, a trend towards overestimation of atypical curves by the Wood model can be observed in the PY estimates.

**Table 1.** Goodness of fit of the Wood model and SMOreg algorithm for the characteristics of lactation TMY, PY, and TPY in typical and atypical shape curves of dairy sheep.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Total Milk Yield (TMY) | | | | Time to Peak Yield (TPY) | | | | Peak Yield (PY) | | | |
|  | Typical | | Atypical | | Typical | | Atypical | | Typical | | Atypical | |
|  | Wood | SMOreg | Wood | SMOreg | Wood | SMOreg | Wood | SMOreg | Wood | SMOreg | Wood | SMOreg |
| r | 0.96 | 0.96 | 0.62 | 0.95 | 0.36 | 0.55 | 0.24 | 0.68 | 0.65 | 0.61 | -0.30 | 0.88 |
| MAE | 10.52 | 8.88 | 29.64 | 10.46 | 19.84 | 15.63 | 19.98 | 8.68 | 0.24 | 0.17 | 6.74 | 0.12 |
| RMSE (l) | 13.21 | 11.66 | 62.40 | 13.97 | 31.00 | 20.11 | 27.37 | 14.21 | 0.33 | 0.25 | 30.02 | 0.15 |
| RAE (%) | 34.11 | 28.29 | 79.86 | 27.72 | 94.11 | 72.60 | 127.83 | 53.33 | 88.24 | 62.28 | 2716.19 | 48.50 |
| RSSE (%) | 35.62 | 31.06 | 143.67 | 31.66 | 129.29 | 81.81 | 142.18 | 71.43 | 101.99 | 78.33 | 9438.84 | 47.53 |
| \*SMOreg: sequential minimal optimization regression, r: coefficient of correlation, MAE: mean absolute error, RMSE: root mean squared error, RAE: relative absolute error, and RRSE: relative root mean squared error. | | | | | | | | | | | | |

**Table 2.** Statistical differences of lactation characteristics for typical and atypical shape curves between the actual lactation and estimated ones by the techniques of prediction (Wood model and SMOreg algorithm).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Typical | | | Atypical | | | p value | | | SE |
|  | Actual | SMOreg | Wood | Actual | SMOreg | Wood | Technique | Shape | Tx\*Shape |
| Total milk yield (l) | 108.2**b** | 108.5**b** | 115.1**a** | 95.4**b** | 95.8**b** | 123.1**a** | 0.007 | 0.259 | 0.169 | 0.362 |
| Time to peak yield (day) | 39**a** | 40**a** | 45**a** | 22**b** | 21**b** | 2**c** | 0.005 | <0.001 | <0.001 | 0.149 |
| Peak yield (l) | 0.991 | 0.999 | 0.811 | 0.93 | 0.931 | 7.644 | 0.075 | 0.105 | 0.063 | 0.096 |
| **a,b, c** Within a row, means followed by a common superscript do not differ significantly in tukey test (*p* < 0.05) | | | | | | | | | | |

***Conclusion and implications***

SMOreg is a powerful algorithm that improves predictions for atypical curves; however, it is necessary to establish techniques to identify this type of pattern within the sheep flock. Therefore, this machine learning algorithm is proposed to be used for better prediction of milk production in dairy sheep systems.

***Acknowledgements***

Guevara L.P. thanks the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 for supporting her research stay at the Intituto de Ciencias Agropecuarias (ICAP) under the direction of Dr. Juan Carlos Angeles Hernandez and the Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil (CNPq) for supporting her doctoral studies.

***References***

Macciotta, N. P. P., Vicario, D., & Cappio-Borlino, A. (2005). Detection of Different Shapes of Lactation Curve for Milk Yield in Dairy Cattle by Empirical Mathematical Models. *J of Dairy Science*, *88*(3), 1178–1191.

Sargent, F. D. (1968). Test interval method for calculating Dairy Herd Improvement Association records. *J of Dairy Science*, *51*, 170–179.

Wood, P. D. P. (1967). Algebraic Model of the Lactation Curve in Cattle. *Nature*, *216*(5111), Article 5111.