Peer-reviewed paper: 8th International Conference on Goats

Prediction of weekly goat milk yield using autoregressive models

C. Fernández[1#], J. Gómez[2], P. Sánchez-Seiquer[1], L. Gómez-Chova[2], E. Soria-Olivas[2], L. Mocé[1] and C. Garcés[1]

[1]Dept. Producción Animal y Ciencia de los Alimentos. Universidad Cardenal Herrera CEU. 46113 Moncada, Spain
[2]Dept. Ingeniería Electrónica. Universidad de Valencia. 46100 Burjassot (Valencia), Spain

Abstract

This paper proposes the use of autoregressive models to predict weekly milk yield in a goat farm. Twenty-eight goats were used to build the model and eight goats were used to validate it. The best models obtained were those in which the prediction was directly related to the present milk yield and previous milk yield (both observed and predicted by the model). This emphasises the strong correlation in terms of time series which exists between consecutive values (weekly in our case) of milk production. The best model provided the best results in terms of accuracy (root mean square error, RMSE = 0.4225 kg/d) and bias (mean error, ME = 0.0044 kg/d).

Keywords: Goat milk, milk yield, time series prediction, autoregressive models
#Corresponding author. E-mail: cjfernandez@uch.ceu.es

Introduction

Goat production in Spain is located principally in the central and southern regions of the country (Falagán et al., 1995), and approximately 20% of the national goat herd consists of Murciano-Granadina dairy goats. Murciano-Granadina goats are well adapted to the Mediterranean climate, characterized by semi-arid conditions with low rainfall and high temperatures. Breeding programs are based primarily on milk yield and milk composition (Analla *et al*., 1996). Most Murciano-Granadina goats are machine milked once a day and official milk records for the Herd Book are obtained by monthly testing during lactation. Farmers' income is generated from the milk produced, thus composition and the accurate measurement or prediction of milk yield (MY) are essential to their economy. A knowledge of the milk production of an ewe over the entire lactation period is a key element in determining optimum strategies of replacement, nutrition and artificial insemination. Very few studies have modelled lactation curves of Murciano-Granadina goats (Pedauye, 1989; Fernández *et al*., 2002). The goal of this study was to test the feasibility of autoregressive models for prediction of the next week's daily MY based on the current week's MY, using individual goat management data.

Materials and Methods

The experiment was conducted at a commercial farm that is a member of ACRIMUR (Murciano-Granadina Goat Breeder Association). ACRIMUR is responsible for the Official Herd Book. Thirty-six Murciano-Granadina dairy goats were used to form a homogeneous group. The 36 goats were chosen according to the number of lactation periods, the number of births, milk production in the previous lactation period and body weight. The experimental trial had a lactation period of approximately five months, and during this time MY and body weight (Grupanor-Cercampo electronic scale) were recorded once a week (4 wks x 5.25 months = 21 observations per goat). In the present experiment we followed the typical routine for the Murciano-Granadina, consisting of a once daily milking (8:30) and kids removed from the dam 24-72 h after birth, ensuring a sufficient colostrum intake and reared by artificial feeding. A portable milking machine was used for milking and the goats were allocated to pens.

The goats were fed two commercial total mixed rations (TMR) with the same chemical composition (92% DM; 160 g CP/kg; 320 g NDF/kg 18 MJ GE/kg DM) and ingredients. The compounds differed only in the source of protein, *viz.* either soyabean meal or sunflower meal (18 goats were fed with the TMR based on soyabean meal containing 460 g protein/kg, and the other 18 goats the TMR containing sunflower meal with 300 g protein/kg). The amount offered per day was 3 kg/d split into two equal portions; at 9:00 (after milking) and at 15:00. Water was freely available at all times. The TMR was obtained from NANTA S.A., and the ingredients were lucerne hay, barley, maize, dehydrated beet pulp, beet molasses, cotton seed and the sources of protein. The diet was balanced based on the recommended values of INRA (1988) and AFRC (1993) for energy, protein, fibre, calcium, phosphorus, sodium and chloride. The diet was supplemented with a vitamin-mineral premix (5 g/kg) provided by Trouw Nutrition S.A. All the goats were housed in a building in which the environment was partially controlled (the temperature varied between 16 and 20 ºC). Throughout the trial, the goats were handled in accordance with the guidelines for the care of animals in experimentation, published by NRC (1998).

The objective of the work is to use autoregressive models to predict milk yield. We have worked with the time series of milk production. A time series is a set of values of a variable that is observed every certain period of time. This period of time is fixed in this time sequence. To define one of these series we will use the following notation: $x_1, x_2, ..., x_n$ where the subscripts are the time index and x designates the variable (Makridakis, 1998). In prediction tasks, we try to answer the following question: "Given the values of a variable until an instant k, what is its value at the instant $k+1$?". A suitable model that could be applied would be the following one:

$$\tilde{x}_{k+1} = a_0 + a_1 \cdot x_k + a_2 \cdot x_{k-1} + ... + a_s \cdot x_{k-s+1} + e_k \quad (1)$$

where e_k denotes the error carried by the model at moment k, $\tilde{x}_k$ - x_k. This model is known as an autoregressive model of order s, AR(s). In (1) x_{k+1} (in the first member) does not appear since what is obtained with the model is an estimation of this value, $\tilde{x}_{k+1}$. Thus, the objective in the prediction of the interest variable is to obtain the optimal parameters a_i. Another possible way to obtain the prediction of the variable would be from the errors of the model in each instance, that is:

$$\tilde{x}_{k+1} = b_0 + b_1 \cdot e_{k-1} + b_2 \cdot e_{k-2} + ... + b_s \cdot e_{k-s} + e_k \quad (2)$$

This model is known as Moving Average of order s, MA(s) model. As above, the problem is to find the optimal factors, b_i. A combination of the two previously commented approaches also exists, which is known as ARMA (q,p) (Makridakis, 1998) and can be written as follows:

$$\tilde{x}_{k+1} = b_0 + b_1 \cdot e_{k-1} + ... + b_p \cdot e_{k-p} + e_k + a_1 \cdot x_k + a_2 \cdot x_{k-1} + ... + a_q \cdot x_{k-q+1} \quad (3)$$

When constructing a model it is necessary to keep a balance between the complexity and accuracy of the model. These two factors are opposite in most of the cases and if the expressions are not carefully handled they can produce incorrect predictions. With complex models we can obtain small errors of adjustment and great errors when trying to extrapolate the model. This is known as overfitting. Therefore, it is necessary to control the complexity of the model in order to obtain a good predicting system. There are multitudes of criteria that control this complexity. One of the most used criteria, and the one used in this research, is the well-known criterion of Akaike (Ljung, 1998).

Results and Discussion

In this work, a methodology different from the standard in prediction problems has been followed. The data set has been divided into two sets (28 goats for training purposes, and eight for the validation of models); with the first set - training set - the model has been developed and with the second one - validation set - the accuracy of the model was tested to be extrapolated to the rest of the cases. Thus, we had a way to validate the developed model.

In order to develop the different autoregressive models we used the system identification toolbox of MATLAB. This software environment is a world-wide standard in numerical calculation being specially optimized to work with vectors and matrices. In addition to the basic package, MATLAB has an extensive set of specialized toolboxes that include the one of identification of systems.

The data pre-processing before using this toolbox was the common one in prediction problems. First we reviewed the subgroups of data associated with each animal to look for lost patterns. If the number of incorrect patterns in a subgroup was low, we completed these patterns by means of interpolation techniques. In opposite cases this subgroup was rejected for the study.

After the data pre-processing, different models were generated. In the first approaches it was verified that the autoregressive (AR) models were those that worked better. The model was selected using the Akaike Information Criterion (AIC) of different AR models (1st, 2nd and 3rd order) and obtaining the Final Prediction

Error (FPE). Evidently this model could be complemented with a MA part obtaining an ARMA model. However, it was preferred not to complicate the model too much.

The excellent adjustment that was obtained in the best model, AR(1) is remarkable. At general level (milk production of the whole population); we observe these graphs for individual goat. Figure 1 shows the accuracy in the prediction. The graphical accuracy of the prediction model is illustrated graphically in Figure 1 and the values of the mean error (ME), the root mean square error (RMSE) and the mean absolute error (MAE) are summarised in Table 1.

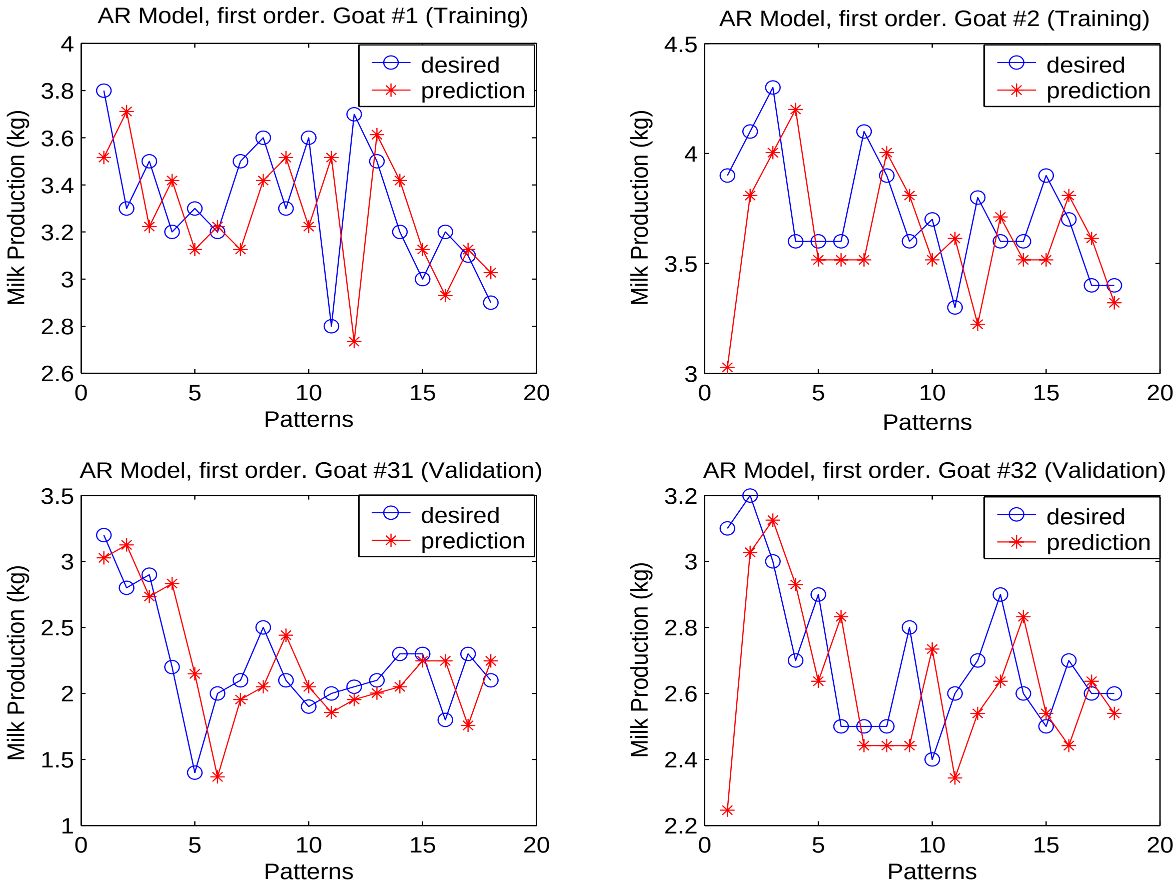

Figure 1 Prediction and desired signal for two goats of the training set and two goats of the validation set

Table 1 Values of the mean error (ME), the root mean square error (RMSE) and the mean absolute error (MAE) obtained for the training and validation sets

Data Set	ME (kg/d)	RMSE (kg/d)	MAE (kg/d)
Training	0.0243	0.3997	0.3074
Validation	0.0044	0.4225	0.3217
Mean milk yield for the training data set:		2.21 ± 0.82 kg/d	
Mean milk yield for the test data set:		2.10 ± 0.65 kg/d	

This accuracy of the models is strongly related with the precise methodology followed to collect and obtain the data used to develop the models. Good data provide reliable and fit models.

Peer-reviewed paper: 8th International Conference on Goats

Conclusions

A set of autoregressive models has been developed for the weekly prediction of goat milk of a given flock. The development of these models has been carried out using a MATLAB computer science package that is a world-wide standard. A strong dependency between consecutive values of the milk production has been shown. Therefore, the best models are the simplest order 1 and 2 AR models. Moreover, we have tested the accuracy of the model and its capability of generalization using a different set from the one used to construct the autoregressive models.

Acknowledgments

This research was supported by the Universidad Cardenal Herrera-CEU Project (ref. PRUCH-03/38).

References

Analla, M., Jiménez-Gamero, I., Muñoz-Serrano, A., Serradilla, J.M. & Falagán, A., 1996. Estimation of genetic parameters for milk yield and fat and protein contents of milk from Murciano-Granadina goats. J. Dairy Sci. 79, 1895-1998.

Falagán, A., Guerrero, J.E. & Serrano, A., 1995. Systèmes d'elevage caprin dans le sud de l'Espagne. Pages 38-50 in Proc. Goat production systems in the Mediterranean. EAAP Publication No 71. Wageningen Pers, Wageningen, The Netherlands.

Fernández, C., Sánchez, A. & Garcés, C., 2002. Modelling the lactation curve for test-day milk yield in Murciano-Granadina goats. Small Rumin. Res. 46, 29-41.

Makridakis, S., Wheelwright, S.C. & Hyndman, R.J., 1998. Forecasting: Methods & Applications. Wiley, NY, USA.

Ljung, L., 1998. "System identification: Theory for the user". Prentice Hall, London.

Pedauye, J., 1989. Lactation curve and milk composition in Murciano-Granadina goat breed. Anales de Veterinaria 5, 3-11.