Indian Journal of Animal Sciences **79** (10): 1038–1040, October 2009

Prediction of lifetime milk production using artificial neural network in Sahiwal cattle

R S GANDHI[1], T V RAJA[2], A P RUHIL[3] and AMIT KUMAR[4]

National Dairy Research Institute, Karnal, Haryana 132 001

Received: 10 December 2008; Accepted: 25 May 2009

ABSTRACT

First lactation records (1493) of Sahiwal cows spread over a period of 40 years (1966–2005) were used to predict lifetime milk production. Data were divided into two sets namely training set (1120 records) and test set (373 records) to compare the accuracy of prediction for lifetime milk production from artificial neural network and multiple regression analysis. The traits considered were age at first calving (AFC), first lactation 305-day or less milk yield (FL305DMY), first lactation length (FLL), first service period (FSP) and first dry period (FDP). The accuracy of prediction of lifetime production from multiple regression analysis was 25.92% from the training set when all the 5 traits were incorporated in the equation, while it was 28.09% for the test data set. An equation incorporating AFC, FL305DMY and FLL was considered to be optimum with an accuracy of prediction of 25.62% for training data set and 27.00% for test data set. The R^2– values for prediction the lifetime production by artificial neural networks from training and test set data were 29.81% and 28.88%, respectively; while for optimum equation the corresponding values were 30.04% and 27.78%, respectively. Further, the root mean square errors of prediction were also lower from artificial neural networks in comparison to multiple regression analysis for all prediction equations developed. Higher estimates of accuracy of prediction of lifetime milk yield from artificial neural networks in comparison to multiple regression analysis from overall and optimum equations in both the data sets revealed that this methodology can be used as an alternate approach to predict lifetime milk production in Sahiwal cattle.

Key words: Accuracy of prediction, Artificial neural network, First lactation traits, Lifetime milk yield, Multiple regression analysis, Sahiwal cattle

The profitability from a dairy herd depends on the overall lifetime production of the cows maintained in a herd. Prediction of lifetime milk production on the basis of early lactation traits with maximum accuracy is one of the criteria of selection for lifetime profitability as it would not be desirable to wait for an animal to complete its lifespan in the herd (Gandhi and Gurnani 1988, Pundir and Raheja 1995, Kannan and Gandhi 2006). Mostly, the prediction of lifetime production is done on the basis of prediction equations constructed by multiple regression analysis, which does not consider dependency among explanatory variables and may lead to biased results.

On the other hand, the connectionist models also known as artificial neural networks (ANNs) are the algorithmic and mathematical models, which are imitating the learning process of human brain and can be applied to non-linear and complex data, even if the data are imprecise and noisy. Further, ANN approach is completely different from conventional statistical methods, which need a specified algorithm to be transformed by a computer program (Grzesiak *et al*. 2003, Sharma *et al*. 2006, Hosseinia, 2007). Therefore, the present investigation was undertaken to predict lifetime milk production on the basis of first lactation traits by multiple regression analysis and artificial neural networks approach and to compare their effectiveness for prediction of lifetime milk production in Sahiwal cattle.

MATERIALS AND METHODS

The data on lifetime performance records of 1493 Sahiwal cows spread over a period of 40 years (1966–2005) were collected from three herds namely National Dairy Research Institute (NDRI), Karnal, Government Livestock Farm (GLF), Chak Ganjaria, Lucknow and Government Cattle Breeding Farm (GCBF), Anjora, Durg. The traits considered were age at first calving (AFC), first lactation 305-day or less milk yield (FL305DMY), first lactation length (FLL), first service period (FSP) and first dry period (FDP). Lifetime

Present address: [1]Principal Scientist and Head (ravindersinghgandhi@yahoo.co.in), [2]Ph.D. Scholar, [3]Senior Scientist, Computer Centre.

[4]Scientist, Animal Genetics Division, IVRI, Izatnagar 243 122.

milk production (LTMY) was the total milk produced by a cow from initiation of first lactation till the day of last completed lactation.

This data were divided into 2 sets namely training set (1120 records) and test set (373 records). Two sets of data were made as required for ANN analysis. The first set (training set) was used to estimate the regression parameters and the second set (test set) was used to validate the estimated regression parameters in terms of accuracy of prediction from the training set. The test set comprising 25% of the records was extracted randomly from the whole set of data. All five first lactation traits were used to predict lifetime milk production from the training and test data sets by multiple regression analysis (MRA) as described by Draper and Smith (1987). Both these data sets were also used to predict lifetime production by the ANNs and their effectiveness was compared from both the methods using R^2– value, residual error variance and root mean square errors.

In order to train ANN, 5 variables pertaining to first parity and one lifetime milk production variable of Sahiwal cows were introduced as input and output variables, respectively, to MATLAB software version 6.1.0 (2001). Then the minimum and maximum values of each variable were mapped to the mean and standard deviation of 0 and 1, respectively. The constructed network was a feed forward back propagation artificial neural network with three layers of input, hidden and output. The layers had 5, 3 and 1 neurons, respectively. The hyperbolic tangent sigmoid transformation function was used for input and hidden layer and pureline linear transfer function was used for output layer.

The ANN program was trained using a maximum number of 2000 epochs with an interval of 50 epochs. An epoch represented one cycle of the iteration procedure to have an efficient convergent solution. The training parameter goal was fixed as 0.01. The Bayesian regulation back propagation training function was used for training the network to have optimum solution.

RESULTS AND DISCUSSION

The means for AFC, FL305DMY, FLL, FSP, FDP and LTMY are given in Table 1. The phenotypic correlations of first lactation traits with lifetime milk yield revealed that AFC and FDP were having negative but significant association with lifetime milk production, while FL305DMY and FLL had positive and highly significant (P<0.01) correlation with LTMY (Table 2). On the other hand, the correlation of FSP with LTMY was not significantly different from zero.

Table 1. Overall population means for whole data set

Traits	Min	Max	Mean	SD	CV%
AFC (Days)	772	2506	1255.14	246.22	19.62
FL305DMY (kg)	57	4362	1594.18	563.65	35.36
FLL (Days)	20	709	323.79	85.89	26.52
FSP (Days	33	517	211.51	114.58	54.17
FDP (Days)	25	720	169.66	102.91	60.66
LTMY (kg)	700.20	34318.40	8543.51	5682.26	66.51

Table 2. Phenotypic correlation matrix for whole data set

Traits	AFC	FL305DMY	FLL	FSP	FDP	LTMY
AFC	1.0000	–0.0698	0.0581	0.1465	0.1305	–0.1207
FL305 DMY		1.0000	0.3615	–0.0427	–0.3001	0.4871
FLL			1.0000	0.4279	–0.2811	0.2848
FSP				1.0000	0.6793	–0.0372
FDP					1.0000	–0.2480
LTMY						1.0000

The accuracy of prediction of lifetime milk production from all the five first lactation traits by MRA was 25.92% for the training data set and 28.09% for the test data set (Table 3). These estimates were lower than the estimates reported by Raheja and Gaur (1995), Gandhi and Bhattacharjia (2001) and Kannan and Gandhi (2006). The lower estimates of R^2– values for prediction of lifetime production under the present investigation might be attributed to wide range of variability in the lifetime milk production of Sahiwal cattle. The residual errors from the training data set were high as compared to the test data set revealing that fitting the above multiple regression equation to the test data set was more effective. Similarly, root mean square errors were also lower from the test data set than that from the training data set.

An equation having AFC, FL305DMY and FLL was considered to be optimum giving R^2– value (25.62%) very

Table 3. Criteria of judging the effectiveness of MRA and ANN analyses

	Training data set All 5 variables		Test data set	
Number of records	1120		373	
	MRA	ANN	MRA	ANN
All 5 variables				
MSE	23449681	22218587	25291309.3	25015031
RMS	4829.50	4701.02	4988.43	4961.11
R-square value	0.2592	0.2981	0.2809	0.2888
Pearson's correlation coefficient	0.5091	0.5460	0.5329	0.5388
3 variables (Optimum equation)				
MSE	23501220	22104860	25535948	25263216
RMS	4839.14	4693.18	5026.14	4999.23
R-square value	0.2562	0.3004	0.2700	0.2778
Pearson's correlation coefficient	0.5062	0.5482	0.5217	0.5289

Fig 1. Regression of ANN predicted lifetime milk yield on first lactation traits in training data set (all 5 variables)

Fig 2. Regression of ANN predicted lifetime milk yield estimated from optimum equation on first lactation traits in training data set

near to the equation having all the five variables for the training data set. Similar results were obtained in the test data set (Table 3). The multiple regression equations fitted on the training data set having all the 5 variables and 3 variables (optimum equation) are given below:

Y= 1823.96 – 1.6861 AFC + 4.1739 FL305DMY + 8.4822 FLL – 2.3758 FDP – 0.8928 FSP

Y = 1221.64– 1.8462 AFC + 4.3186 FL305DMY + 8.4204 FLL

The R^2– value of prediction of LTMY from the test data set was slightly lower (28.88%) when the same network was trained. This may be due to lower number of observations in the latter set of data. When ANN was trained on the training data set having 3 variables as that was incorporated in optimum equation from regression analysis, an accuracy of prediction of 30.04% was obtained, which was also higher than that estimated from regression analysis for same number of variables (25.62%). The estimate of R^2– value from the test data set was 27.78%, which was lower than that from training data set (Table 3).

The prediction equations derived from ANN analysis fitted on the training data set having all (5 variables) and 3 variables (optimum equator) are given below:

$$Y = 6153.77 + 0.2721\ T$$

$$Y = 6007.90 + 0.2943\ T$$

where, Y was the predicted value for lifetime milk yield on the basis of first lactation traits and T was the actual targeted value for LTMY. The actual (targeted) and predicted lifetime milk yields estimated from ANNs were plotted (Fig. 1 and 2).

The mean error sum of squares (MSE) and root mean square errors (RMS) estimated from ANN method were lower than that estimated from MRA in both the data sets for ordinary and optimum equations. However, the Pearson's correlation coefficients of predicted lifetime milk yields from ANN with the actual lifetime milk yields of cows were higher than that from the predicted lifetime milk yields by MRA for all the equations (Table 3).

All the criteria of judging the effectiveness of various equations revealed that the prediction of lifetime milk yield from first lactation traits was more accurate from artificial neural networks analysis in comparison to multiple regression analysis, though the difference in accuracy of prediction was small. Hence, it was suggested that artificial neural networks could be used as supplement to multiple regression analysis for prediction of lifetime milk yield on the basis of early lactation traits in Sahiwal cattle.

REFERENCES

Draper N R and Smith H. 1987. Applied Regression Analysis. John Wiley & Sons Inc. New York.

Gandhi R S and Bhattacharjia T K. 2001. Prediction of lifetime production on the basis of first lactation traits in Karan Fries cattle. *Indian Journal of Animal Science* **71**: 474–75.

Gandhi R S and Gurnani M. 1988. Optimum combinations of early traits for prediction of 8-year lifetime production in Sahiwal cattle. *Asian Journal of Dairy Research* **7**: 7–12.

Grzesiak W, Lacroixr R, Wojcik J and Blaszczyk P. 2003. A comparison of neural network and multiple regression predictions for 305-day lactation yield using partial lactation records. *Canadian Journal of Animal Sciences* **83**: 307–10.

Hosseinia P, Edrisi M, Edriss M A and Nilforooshan M A. 2007. Prediction of second parity milk yield and fat percentage of dairy cows based on first parity information using neural network system. *Journal of Applied Sciences* **7**: 3274–79.

Kannan D and Gandhi R S. 2006. Prediction of lifetime production in Sahiwal cattle. *Indian Journal of Animal Sciences* **76**: 768–69.

MATLAB 2001. Software Version 6.1.0.

Pundir R K and Raheja K L. 1995. Prediction of lifetime milk yield by early traits in Sahiwal and Hariana cows. *Indian Journal of Dairy Science* **48**: 146–49.

Raheja K L and Gaur G K. 1995. Prediction of lifetime yield from first lactation traits in crossbred cattle. *Indian Journal of Dairy Science* **48**: 388–89.

Sharma A K, Sharma R K and Kasana H S. 2006. Empirical comparisons of feed forward connectionist and conventional regression models for prediction of first lactation 305-day milk yield in Karan Fries dairy cows. *Neural Computing and Applications* **15**: 359–65.

Tadeusiewicz R. 1993. Neural networks. AOW, Warsaw (In Polish).