Indian Journal of Animal Sciences **85** (5): 477–479, May 2015/Article

Prediction of FL 305 DMY from monthly part lactation milk yield records using artificial intelligence in Sahiwal cattle

U T MUNDHE[1], R S GANDHI[2], D N DAS[3] V B DONGRE[4] and ATUL GUPTA[5]

National Dairy Research Institute, Karnal, Haryana 132 001 India

Received: 11 June 2014; Accepted: 26 November 2014

ABSTRACT

The present investigation was carried out on 4,760 first lactation monthly part lactations milk yield records of 476 Sahiwal cows sired by 34 bulls maintained at National Dairy Research Institute (NDRI), Karnal over a period of 51 years (1961–2011) to predict first lactation 305-day milk yield (FL305DMY). The comparison was made between multiple linear regression analysis and artificial neural network method for prediction of first lactation 305-day milk yield in Sahiwal cows. In multiple regression method, an optimum equation having 3 variables namely PL2, PL5 and PL8 gave an accuracy of prediction of 88.80%, whereas under artificial neural network, this equation gave an accuracy of prediction of 89.29%. It was concluded that artificial neural network could be used as an alternative to multiple regression method for prediction of FL305DMY in Sahiwal cows.

Key words: Artificial neural network, Milk yield, Monthly lactation records, Multiple linear regression, Sahiwal cows

Generally sires are evaluated on the basis of first lactation 305-day or less milk yield of their daughters which leads to enhanced generation interval, lower genetic gain per unit of time and fewer numbers of daughters per sire due to smaller herd size. Further, under field conditions it is cumbersome task to record daily milk yields and therefore, part lactation milk records can be used as an alternative. The prediction of first lactation milk yield based on monthly part lactation records is useful in early culling of genetically inferior cows, early evaluation of dairy sires on the basis of part lactations of their daughters or in cases where cow records are needed to be extended (Gandhi *et al*. 2009, Raja 2010, Dongre *et al*. 2012). Hence, the present investigation was undertaken to predict first lactation 305-day milk yield on the basis of monthly part lactation records using multiple regression analysis and artificial neural network in Sahiwal cattle.

MATERIALS AND METHODS

Data on 4,760 first lactation monthly part lactations milk yield records of 476 Sahiwal cows maintained at National Dairy Research Institute, Karnal over a period of 51 years (1961–2011) were analyzed. The individual monthly part lactation records were calculated by adding daily milk yield at an interval of 30 days. The colostrum yield for the first 5 days after calving were not added to calculate the monthly milk yields and individual monthly part lactation records were generated by adding daily milk yield from sixth day to 305-day of lactation. The records of animals with less than 300 kg of milk production or less than 100 days of lactation length were discarded. Further, records of abortion and still birth were also discarded. Multiple linear regression and artificial neural network were used and compared for the prediction of FL-305 DMY from monthly part lactation milk yields. The model used for prediction of first lactation 305 DMY by multiple linear regression analysis was:

$$\hat{Y}_i = a + b_i \Sigma X_i$$

where, $\hat{Y}_i$, estimated first lactation 305 day or less milk yield of the i^{th} animal; X_i, monthly part lactation record of i^{th} animal; a, intercept; b_i, regression coefficient of first lactation 305 day or less milk yield on monthly part lactation records.

The per cent coefficient of determination (R^2) was calculated as:

R^2 = (sum of squares due to regression /total sum of square) ×100

Second method of prediction was artificial neural network which is a multilayer forward neural network with back propagation of error learning mechanism was developed using neural network toolbox (NNT) of MATLAB 7.0 to predict the first lactation 305-day or less milk yield (FL305 DMY). The network was trained and

Present address: [1]PhD Scholar (utmundhe12@gmail.com), [3]Principal Scientist, Department of Animal Genetics and Breeding, NDRI, Southern Regional Station, Bengaluru. [2]Assistant Director General (Animal Production and Breeding), Indian Council of Agricultural Research, New Delhi, India. [4]Assistant Professor, Department of Animal Genetics and Breeding, College of Veterinary Science and Animal Husbandry, Junagadh Agricultural University, Junagadh. [5]Assistant Professor, Department of Animal Genetics & Breeding, SVPUA&T, Meerut.

simulated using Bayesian regularization (BR) up to 4,000 epochs or till the algorithms truly converged. The input and target data was pre-processed so that the mean is 0 and the standard deviation is 1 using the feature prestd available in NNT as per the requirement of algorithms. Network parameters such as learning rate (0.01), momentum (0.05), and error goal (0) were used as the default setting of the algorithms. The whole data were divided into training (75%) and test data sets (25%).

For analysing the data, the network was tested with 1 hidden layer (3, 5, 7 and 10 neurons) and 2 hidden layers (3:3, 3:5, 5:5 and 7:10 neurons). Initial weights and bias matrix were randomly initialized between –1 to 1. A non-linear transformation function tangent sigmoid was used to compute the output from summation of weighted inputs of neurons in each hidden layer. A pure linear transformation function was used at output layer for getting network response. The designed network was trained in supervisory mode with several variant of back propagation of error learning algorithms.

The performance efficiency of the network model was evaluated using the value of coefficient of determination (R^2-value).

RESULTS AND DISCUSSION

In multiple regression method monthly part lactation milk yields were considered to predict first lactation 305-day milk yield (FL305DMY) using backward elimination method. When all monthly part lactation records (PL1 to PL8) were incorporated in an equation, the R^2-value was 89.37%. However, in second equation, when PL1, PL2, PL3, PL5, PL7 and PL8 were involved in an equation, it gave 89.30% R^2-value. In the third equation, when PL3 was dropped from the equation, it resulted in slight decrease in accuracy of prediction (89.16%). Further, when PL3, PL4, and PL7 was dropped from the equation, the R^2-value decreased to 88.99%. Finally, with PL2, PL5 and PL8 gave an optimum equation with R^2 –value of 88.80% (Table 1).

Three part lactations records incorporated in above optimum equation were used as input variables for artificial neural network (Table 2).

Table 1. Prediction equations for first lactation 305-day or less milk yield on the basis of monthly part lactation milk yields by backward elimination method

Sl.No.	Prediction equations	R^2-value (%)
1	v = 93.9156 + 0.8140PL1 + 1.5106PL2 + 0.6579PL3 +0.5573PL4 + 1.5979PL5 + 0.5756PL6 + 0.9596PL7 + 2.6294PL8	89.37
2	v = 94.2410 + 0.8047PL1 + 1.4877PL2 + 0.9378PL3 + 2.1447PL5 + 1.3185PL7 + 2.6145PL8	89.30
3	v = 107.7379 + 0.6718PL1 + 2.1970PL2 + 2.4970PL5 + 1.2682PL7 + 2.6063PL8	89.16
4	v = 111.5145 + 0.6428PL1 + 2.2006PL2 + 2.8562PL5 + 3.5155PL8	88.99
5	v = 120.6400 + 2.7273PL2 + 2.7808PL5 + 3.5146PL8	88.80

Table 2. Description of input and output variables

Input variables	Day of lactation	Output variable
PL2	36st to 65th day of lactation	First lactation 305 day milk yield (FL305 -DMY)
PL5	126th to 155th day of lactation	
PL8	216th to 245th day of lactation	

Table 3. R^2 values of *Bayesian regularization (BR)* algorithm with training (75%) and test (25%) data sets

Input sets	Layers	Neurons	R^2 (%)
1	1	3	70.01
1	1	5	71.45
1	1	7	70.34
1	1	10	69.34
1	2	3:3	69.36
1	2	3:5	73.64
1	2	5:5	70.48
1	2	7:10	71.85
2	1	3	80.72
2	1	5	80.21
2	1	7	79.22
2	1	10	81.25
2	2	3:3	82.42
2	2	3:5	87.16
2	2	5:5	85.21
2	2	7:10	83.23
3	1	3	88.21
3	1	5	86.76
3	1	7	84.20
3	1	10	89.29
3	2	3:3	86.71
3	2	3:5	89.24
3	2	5:5	88.21
3	2	7:10	88.12

Further, 3 input sub-sets were prepared with a total of 3 part lactations milk yield records, which were used as input variables. The first input set included one part lactation milk yield record (PL2), second input set included 2 part lactation milk yield records (PL2 and PL5) and third set included 3 part lactations (PL2, PL5 and PL8).

The artificial neural network (ANN) was trained on the training data set having 3 monthly part lactations records. The network was tested with 1 hidden layer (3, 5, 7 and 10 neurons) and 2 hidden layers (3:3, 3:5, 5:5 and 7:10 neurons) (Table 3).

The prediction performance was tested using a new data set. It was inferred that the accuracy of prediction (R^2-value) showed increasing trend while decreasing the percentage of test data set. The significant difference observed in training data-test data: 75–25% between multiple linear regression (88.80%) and ANN (89.29%). In accordance with

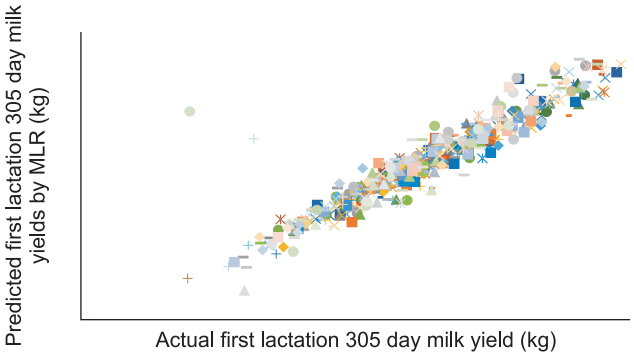

Fig 1. Actual versus the predicted first lactation 305-day milk yield (kg) from best MLR model (kg).

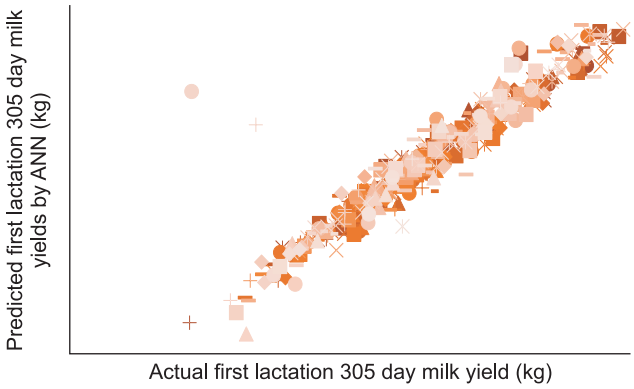

Fig 2.Actual versus the predicted first lactation 305-day milk yield (kg) from best ANN model (kg).

the present finding, Dongre *et al.* (2012) reported that artificial neural network explained higher value of coefficient of determination for predicting FL305 DMY using fortnightly test days records.

Comparison of ANN in different algorithms: It was observed that the accuracy of prediction of the algorithm increased with the addition of part lactations milk yields as input variables and network with single hidden layer with 10 neurons showed the best result. Accuracy of the models was tested on the basis of coefficient of determination value of the test data. It was further observed that the accuracy of prediction of the models increased with the addition of part lactation records in the set. The highest coefficient of determination (R^2-value = 89.29%) was explained by Bayesian regularization (BR) algorithm in set-3. The FL305DMY predictions obtained by the best ANN model and the MLR model are graphically shown in Figs 1 and 2.

It was concluded that first lactation 305 day milk yield can be predicted using monthly part lactation milk yields during early (ascending phase), middle (persistent phase) and later stage (descending phase) of lactation curve in Sahiwal cattle. The accuracy of prediction obtained from artificial neural network was slightly higher than multiple linear regression for prediction of FL305DMY using monthly part lactation milk yield records in Sahiwal cows. Hence, it is recommended that artificial neural network can also be used for prediction of FL305DMY in Sahiwal cows.

REFERENCES

Ahunu B K and Danbaro G. 1992. Prediction of lactation yield in crossbred cattle. *Indian Journal of Dairy Science* **45** (5): 241–45.

Dongre V B, Gandhi R S, Avtar Singh and Ruhil A P. 2012. Comparative efficiency of artificial neural networks and multiple linear regression analysis for prediction of first lactation 305-day milk yield in Sahiwal cattle. *Livestock Science* **147:** 192–97.

Gandhi R S, and Gurnani M. 1988. Association amongst different productive and reproductive traits in Sahiwal cattle. *Asian Journal of Dairy Research* **7** (3): 171–74.

Gandhi R S, Raja T V, Ruhil A P and Kumar A. 2009. Prediction of lifetime milk production using artificial neural network in Sahiwal cattle. *Indian Journal of Animal Sciences* **79**: 1038–40.

Grzesiak W, Blaszczyk P and Lacroix R. 2006. Methods of predicting milk yield in dairy cows–predictive capabilities of Wood's lactation curve and Artificial Neural Networks (ANN). *Computers and Electronics in Agriculture* **54**: 69–83.

Joshi B K, Tantia M S, Vij P K, Kumar P and Gupta N. 1996. Performance of Hariana cows under farmer's herd condition. *Indian Journal of Animal Sciences* **66:** 383–97.

Kannan D S. 2002. 'Lifetime performance evaluation of Sahiwal cattle.' M.Sc.Thesis, NDRI, Deemed University, Karnal, India.

Raja T V. 2010. 'Part lactation records for Sahiwal sire evaluation.' unpublished Ph.D thesis, National Dairy Research Institute, Karnal, India.