



## Recent prediction tools for estimation of first lactation 305 days milk yield using monthly test days for genetic evaluation in crossbred cattle

T S ARUNA<sup>1</sup>✉, S M DEB<sup>2</sup>, A P RUHIL<sup>3</sup>, RAVINDER MALHOTRA<sup>3</sup> and SALEEM YOUSUF<sup>3</sup>

ICAR- National Dairy Research Institute, Karnal, Haryana 132 001 India

Received: 30 September 2021; Accepted: 2 December 2021

### ABSTRACT

First lactation 305 days milk yield (FL305DMY) of progenies is an important criteria for selection of sires. In field progeny testing programme, prediction of FL305DMY based on test day records will reduce the cost of recording. The present study was aimed at evaluating the efficiency of different methods of prediction for FL305DMY based on monthly test day milk yield records in Karan Fries cattle and evaluation of sires based on the performance of its progenies. Methods of prediction were Wood's incomplete gamma function and Artificial neural network (ANN). Efficiency of prediction was compared using error of prediction. Root mean square error was least for ANN method (5.72%). The attempt was made to evaluate Karan Fries sires based on actual and predicted FL305DMY so obtained utilising BLUP- AM method of sire evaluation. The breeding values of 48 Karan Fries sires having 3 or more daughters were estimated. Comparison of ranking of sires based on different methods of prediction was done using Spearman's Rank correlation and RMSE values. The sire evaluation using ANN method based predicted values showed marginally higher rank correlation (0.91) and least RMSE (2.39%) with actual yield based on monthly test day records. Thus it was concluded that recoding at monthly intervals and use of ANN and conventional fitting Wood's function can be used for evaluation of sires. Though both the methods are equally efficient, ANN showed marginal superiority.

**Keywords:** Artificial neural network, BLUP-AM, Monthly test days, Wood's incomplete gamma function

Accurate selection of superior sire is an important task in breeding programme. Since selection of bulls can bring about 61% of genetic gain in dairy cattle (Robertson and Rendel 1954). In India, sires are evaluated mainly on the basis of first lactation 305-day milk yield at organized herds. Sires are selected on the basis of his progenies' performance, so this parameter is an important criterion for evaluation of sire. But, due to smaller herd size and limited number of daughters per sire in organized farm, the accuracy of evaluating sire is low. To overcome this situation, field progeny testing is being carried out by incorporating more number of breedable females from the field. Recording the daily milk yield of the progenies in the field is a difficult and capital intensive task. Here comes the importance of test day milk recording. Prediction of 305 day milk production can be done using few monthly test day milk yields.

Over the years, different conventional methods are used for prediction of FL305DMY based on test day records.

Present address: <sup>1</sup>Animal Genetics Division, Indian Veterinary Research Institute, Izatnagar, Uttar Pradesh. <sup>2</sup>Animal Genetics and Breeding Division, ICAR-National Dairy Research Institute, Eastern Regional Station, Kalyani, West Bengal. <sup>3</sup>ICAR-National Dairy Research Institute, Karnal, Haryana. ✉Corresponding author email: tsarunathoppil2010@gmail.com

Incomplete gamma function (Wood 1967) is a nonlinear lactation curve model, which has been still used for modeling of lactation curve and prediction of FL305DMY (Banu *et al.* 2012 and Kong *et al.* 2018). Recent prediction tools like Artificial neural network (ANN) has been used for 305 day milk yield prediction based on test day models in different breeds of cattle (Dongre *et al.* 2012 and Singh *et al.* 2020) and buffalo (Rana *et al.* 2021). ANN is a computational mirror of its biological counterpart which is composed of multilayer of interconnected neurons. These models are machine leaning systems based on artificial intelligence which can learn from previous experience. Since ANN is a nonlinear technique which can take any kind of input, it can be used for prediction of complex traits.

Limited information is available about the prediction of 305 days milk yield using recent prediction tools which may be useful under field conditions. As the actual 305 days milk yield of Karan Fries cattle is available at the institute, the accuracy of prediction of the same using recent prediction tools can be ascertained.

### MATERIALS AND METHODS

*Data collection:* The data of Karan Fries cattle on FL305DMY, MTDY were collected from history cum pedigree sheets and daily milk yield registers maintained

at ICAR-NDRI, extended over a period of 20 years from 1998 and consist of records on 1,027 animals. Only those Karan Fries cattle were considered for study which had produced a minimum of 500 kg of milk for at least 100 days. After standardization and normalization of raw data, 722 animals having all ten test days were selected for prediction of FL305DMY. For prediction based on MTDY all ten test days were utilized. The data was classified into subclasses according to the season (4), period of calving (4) and age at first calving (11). The whole data set was divided into training and test data set. Training data set (60%) was used for training in ANN. Prediction was done for test data set (40%).

*Prediction of FL305DMY:* Prediction of FL305DMY from MTDY were done by fitting Wood’s incomplete gamma function and ANN.

*Wood’s incomplete Gamma function*

Prediction of FL305DMY based on MTDY were done by fitting Wood’s incomplete gamma function (Wood 1967).

$$Y_t = at^b e^{-ct}$$

where,  $Y_t$ , Average daily yield in the  $t^{th}$  day;  $a$ , Approximate initial milk yield just after calving;  $b$ , Inclining slope parameter up to peak yield;  $c$ , Declining slope parameter after peak yield. 305-days milk yield of each animal were calculated.

*Artificial Neural Network*

ANN is an intelligent computing method mainly used for prediction based on nonlinear technique of machine learning. It consists of network of interconnected artificial neurons. ANN has three layers (input, output and hidden layers). The ANN models can learn predictive ability based on input variables and output data from the training data set. Input variables are test day milk yields (MTDY) and output variable is FL305DMY. Back propagation of errors technique is the commonly used method for training the network till it converges to a predictive value (Fausett 2006). In the present study, a multilayer feed forward neural network with back propagation of error learning mechanism was developed using Neural Network Toolbox (NNT) of MATLAB 7.8.0 (MATLAB Users’ Guide, R2009a) for prediction of FL305DMY. Network was trained using Bayesian regularization (BR) algorithm up to 6000 epochs or until the algorithms are truly converged. Each input is weighed with an appropriate weight and weighed input and bias form the input to the transfer function. Log sigmoid (log sig) and tan sigmoid (tan sig) are the transfer functions used to convert input signals to output signals. Parameters like learning rate (0.01), momentum (0.05), and error goal (0) were set as default network parameters. Accuracy of ANN model was compared based on coefficient of determination ( $R^2$  value) and root mean square error (RMSE) values of the test data using the equations:

$$R^2 = 1 - \frac{SS_E}{SS_T}$$

where  $SS_T$  is total sum of squares and  $SS_E$  is error sum of squares.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{n}}$$

where  $\hat{Y}_i$ , estimated 305-day milk yield of the  $i^{th}$  animal;  $Y_i$ , actual 305-day milk yield of the  $i^{th}$  animal and  $n$ , number of animals in the test group.

Network training was done using two hidden layers and different number of neurons in each hidden layer till the best result was obtained. The most suitable method of prediction among the above mentioned methods was identified on the basis of least RMSE and Average error values.

*Estimation of breeding value:* The breeding values of sires were estimated by BLUP- AM (Henderson 1975), using Wombat (Meyer 2007) software. Following general model was used.

$$Y = Xb + Zu + e$$

where,  $Y$ , vector of observations for trait;  $b$ , the vector of observations of fixed effects (season, period and age first calving);  $u$ , vector of observations of random effects (sire);  $X$  and  $Z$ , Incidence matrix of fixed and random effects respectively;  $e$ , Random error.

Comparison of ranking of sires using estimated breeding values, based on actual and predicted FL305DMY were done by estimating rank correlation coefficient (Spearman 1904) and RMSE values.

RESULTS AND DISCUSSION

Wood’s incomplete gamma function was fitted to each animal for predicting FL305DMY.  $R^2$  ranged from 3.3–99.9% with an average of 71.6%. Similar degree of accuracy was reported by Gupta *et al.* (2016) in Sahiwal cattle ( $R^2 = 74.51\%$ ) and Banu *et al.* (2012) in Karan Fries cattle ( $R^2 = 79.05\%$ ) based on MTDY.

For prediction of FL305DMY using ANN, comparison of  $R^2$  and RMSE values in different number of hidden layer neurons combinations and activations functions are given in Table 1. In most of the times, algorithm was truly converged that means that the error goal was achieved. Highest  $R^2$  (92.68%) with least RMSE (201.88 kg) was given by two hidden layer combinations with four and three neurons each with tangent sigmoid activation function. Dongre *et al.* (2012) used Bayesian regularization (BR) back propagation algorithm in ANN based on initial 6, 7, 9, 10, 11 and 18 weekly test day milk yield and reported  $R^2$  value of 81.29% in Sahiwal cattle for predicting first lactation 305 day milk yield. Rana (2017) reported RMSE value of 90.47 kg (4.28%) based all monthly test days using ANN in Murrah buffalo.

Comparison of actual and predicted FL305DMY using the above methods was given in Table 2. Table 2 shows that both the methods were efficient in predicting FL305DMY using monthly test day milk yield records.

Table 1. R<sup>2</sup> and RMSE values of Bayesian regularization (BR) algorithm with test data

Number of neurons in hidden layer		Training function for first and second hidden layers			
First layer	Second layer	log sig		tan sig	
		RMSE (kg)	R <sup>2</sup> (%)	RMSE (kg)	R <sup>2</sup> (%)
4	3	204.57	92.49	201.88	92.68
4	2	204.11	92.52	201.91	92.52
3	2	238.60	89.78	205.47	92.42
6	3	212.51	91.89	210.39	92.05
6	2	210.59	92.03	219.03	91.39
6	5	214.00	91.78	210.93	92.01
6	4	209.29	92.13	218.05	91.46
5	4	207.60	92.26	202.44	92.31
5	3	217.87	91.48	210.45	92.04
5	2	208.00	92.23	211.19	91.99

Table 2. Comparison of actual and predicted FL305DMY using different methods

Method of Prediction	Average error (kg)	Average error (%)	RMSE (kg)	RMSE (%)
ANN	140.71	3.99	201.88	5.72
Wood's	152.15	4.31	212.13	6.01

Based on ANN method, deviation of predicted FL305DMY from actual FL305DMY was found to be 140.71 kg as average error (3.99%), 201.88 kg as RMSE (5.72%). The ANN method was found to be marginally better with least error.

In order to maintain the uniformity of data set used for genetic studies, only sires from test data set were taken. Breeding values of 48 Karan Fries sires having at least three daughters were estimated based on their daughters' actual and predicted FL305DMY based on the above two methods using BLUP – AM. Table 3 shows the average breeding value estimates of sires for actual and predicted FL305DMY.

Comparison of relative accuracy of sire ranking based on predicted FL305DMY with actual FL305DMY was done

Table 3. Descriptive statistics of estimated breeding values

	Actual FL305DMY	Predicted FL305DMY	
		ANN	Wood's function
Average EBV (kg)	3580.83	3563.72	3608.00
Number of sires above average EBVs	25 (52.08%)	26 (54.17%)	27 (56.25%)
Number of sires below average EBVs	23 (47.92%)	22 (45.83%)	21 (43.75%)
Min. EBV (kg)	2973.41	3124.13	3139.99

Table 4. Rank correlations of sires ranking using actual and predicted FL305DMY

	Sire ranks based on predicted FL305DMY using ANN	Sire ranks based on predicted FL305DMY using Wood's function
Sire ranks based on actual FL305DMY	0.914	0.905

in terms of rank correlation coefficient (Table 4) and RMSE value. It was observed that rank correlations among estimated breeding values were near to 0.91 in both methods of prediction, among these ANN shows marginal superiority. The estimates of rank correlation were highly significant ( $P < 0.01$ ). RMSE value calculated as the difference between estimated breeding values based on actual and predicted first lactation 305 day milk yield showed that ANN method was having least RMSE, 85.40 kg (2.39%) while value using Wood's function was 91.97 (2.57%). Rana (2017) used BLUP- AM for estimation of breeding values for actual and predicted FL305DMY in Murrah buffalo and showed RMSE of 0.39% based on ANN method of prediction. Similarly Nosrat *et al.* (2021) observed RMSE of 0.93% in Holstein cows.

Comparison of estimated breeding values calculated by BLUP-AM showed that estimated breeding values of sires using ANN method of prediction were having high rank correlation and least RMSE compared with breeding values based on actual first lactation 305 day milk yield.

For reducing the cost of recording without compromising accuracy of prediction and evaluation of sires, recoding at monthly intervals and use of artificial neural network and conventional Wood's function can be used. Though all the methods are equally efficient, artificial neural network showed marginal superiority with least error for prediction of first lactation 305 days milk yield. The sire evaluation using artificial neural network method based predicted value showed marginally higher rank correlation and least root mean square error with actual yield.

## REFERENCES

- Banu R, Singh A, Malhotra R, Gowane G, Kumar V, Jaggi S, Verghese E, Gandhi R S, Chakravarty A K and Raja T V. 2012. Comparison of different lactation curve models in Karan Fries cattle of India. *Indian Journal of Animal Sciences* 82(11): 1377–80.
- Dongre V, Gandhi R S and Singh A. 2012. Comparison of different lactation models in Sahiwal cows. *Turkish Journal of Veterinary and Animal Sciences* 36(6): 723–26.
- Fausett L V. 2006. *Fundamentals of Neural Networks: Architectures, Algorithms and Applications*. Pearson Education, India.
- Gupta A, Gandhi R S, Singh M, Singh A, Prakash V, Dash S K and Dash S. 2016. Comparison of different lactation curve models in Sahiwal cattle up to fourth parity using monthly test day milk yields. *Indian Journal of Dairy Science* 69(4): 460–66.
- Henderson C R. 1975. Use of relationship among sires to increase

- accuracy of sire evaluation. *Journal of Dairy Science* **58**(11):1731–37.
- Kong L, Li J, Li R, Zhao X, Ma Y, Sun S, Huang J, Ju Z, Hou M and Zhong J. 2018. Estimation of 305-day milk yield from test-day records of Chinese Holstein cattle. *Journal of Applied Animal Science* **46**(1): 791–97.
- Meyer K. 2007. WOMBAT version 1.0 User notes. Uni New England, Armidale, NSW, Australia.
- Nosrati M, Hafezian S H and Gholizadeh M. 2021. Estimating heritabilities and breeding values for real and predicted milk production in holstein dairy cows with artificial neural network and multiple linear regression models. *Iranian Journal of Applied Animal Science* **11**(1): 67–78.
- Rana E. 2017. 'Genetic evaluation of Murrah buffaloes based on test day milk yield and peak yield records.' M V Sc. Thesis, National Dairy Research Institute, Deemed Univesity, Karnal, India.
- Rana E, Gupta A K, Singh A, Ruhil A P, Malhotra R, Yousuf S and Ete G. 2021. Prediction of first lactation 305-day milk yield based on bimonthly test day milk yield records in Murrah buffaloes. *Indian Journal of Animal Research* **55**(4): 486–90.
- Robertson A and Randel J M. 1954. The performance of heifer got by artificial insemination. *Journal of Animal Science* **44**: 184–92.
- Singh N, Usman S, Maurya V, Dutt T, Bhatt N and Kumar A. 2020. Comparative analysis of artificial neural network algorithms for prediction of FL305DMY in Murrah buffalo. *International Journal of Livestock Research* **10**(9): 205–09.
- Spearman. 1904. The proof and Measurement of association between two things. *The American Journal of Psychology* **15**(1): 22–26.
- Wood P D P. 1967. Algebraic model of lactation curve in cattle. *Nature, London* **216**: 164.