



Modeling and forecasting of lifetime milk production in Haryana cattle

MAHANTESH SHETKAR¹, VIJAY KUMAR^{1✉}, S P SINGH¹, YAJUVENDRA SINGH¹ and MUNEEENDRA KUMAR¹

Pt. Deen Dayal Upadhyaya Veterinary University, Mathura, Uttar Pradesh 281 001 India

Received: 6 June 2022; Accepted: 16 March 2023

ABSTRACT

The present study was conducted on performance records of 648 Haryana cattle with respect to first lactation and lifetime traits spread over a period of 57 years (1962-2018) from DUVASU farm, Mathura and 23 years (1995–2017) from State livestock farm Babugarh, Hapur. Lifetime production traits, viz. Lifetime 2 Milk Yield (LT2MY), Lifetime 3 Milk Yield (LT3MY), Lifetime 4 Milk Yield (LT4MY) and Lifetime 5 Milk Yield (LT5MY) were predicted by multiple linear regression analysis. Step-wise regression procedure was used to predict the lifetime milk yield from first lactation traits, viz. age at first calving (AFC), first calving interval (FCI) and first lactation total milk yield (FLTMY). The optimum models were selected with the help of Akaike information criterion (AIC), Bayesian information criterion (BIC), Mallows's conceptual predictive value and variance inflation factor. The models with lower AIC, BIC and RMSE values and with higher co-efficient of determination were selected as optimum. The optimal models for lifetime production traits, viz. LT2MY, LT3MY, LT4MY and LT5MY with co-efficient of determination were 75.46%, 66.63%, 59.4% and 55.03%, respectively.

Keywords: Coefficient of determination, FCI, FDP, Haryana cattle, Regression

Livestock sector including dairying, which is the component of the agriculture sector in India, has shown remarkable development in the recent past (Valsalan *et al.* 2014). Lifetime performance of the dairy cattle determines the profitability of dairy enterprise because the overall productivity of dairy animals depends on their lifetime performance rather than single lactation performance (Shetkar *et al.* 2021a). Lifetime milk production (LMP) is important because the costs of rearing replacement cows represent a substantial part of the expenses in the dairy cattle production system. An early prediction of the phenotypic value of lifetime traits from lactation traits would help in culling process and reduce the expenditure in maintaining animals having less lifetime production. Lifetime production is one of the important parameters to obtain the most economic return from a particular animal (Shetkar *et al.* 2021b). Thus, dairyman would have a considerable interest in the relationship of these traits in lifetime milk production and other production traits. In any dairy cattle production enterprise, the length of life of a dairy cow has a substantial impact on economic performance. Arthur *et al.* (1992) reported that longer life-span in dairy cows allows producers to be more selective in choosing replacement heifers because only a few have to be chosen each year. Higher longevity also reduces the

cost of herd replacements, increases the number of animals available for marketing, and increases the proportion of the high producing, mature animals in the breeding herd (Arthur *et al.* 1992).

Although lactation records are widely used in assessing the genetic merit of cattle but selection of dairy sires is invariably based on the first one or two lactation records in most of the breeding programmes (Kuralkar and Raheja 1997). But still, there is further need to study the relationship between first, later lactations and lifetime performance traits for overall better evaluation of the genetic worth of the individual's performance. Several procedures have been often proposed for prediction of lifetime milk production from early lactation traits but out of them, Multiple Regression Analysis technique has been widely used. Patil *et al.* (2013) showed the importance of predicting fertility (DPR) of Murrah bulls using service period, which subsequently helps in sire evaluation under the Progeny Testing Programme in India. Considering the above facts, an investigation was done in Haryana cattle with the objectives to develop predictive model for lifetime performance traits.

MATERIALS AND METHODS

In the present study, the breeding information were collected from the history sheet registers and herd inventory registers of Haryana cattle maintained at DUVASU farm, Mathura and Babugarh Farm, Hapur. Total of 648 animal records were used. The data on productive animals with date of birth from 1962 to 2018 were collected from DUVASU

Present address: ¹Department of Animal Genetics and Breeding, College of Veterinary Science and Animal Husbandry, Pt. Deen Dayal Upadhyaya Veterinary University, Mathura, Uttar Pradesh. ✉Corresponding author email: jaisvet@gmail.com

farm, Mathura. Whereas, the data of productive animals with date of birth from 1995 to 2018 were collected from Babugarh farm, Hapur for the study.

Standardization of records: To ensure the normal distribution, the outliers were removed and data within the range of $\text{mean} \pm 3\text{SD}$ were considered for the study. The records of animals with equal or more than 400 L of milk production and equal or more than 100 days of lactation length in any of its lifetime lactation were considered for the study.

Variance inflation factor (VIF): When correlation exists among predictors, the value of coefficient of determination is biased. In that case standard error of predictors coefficients will increase and consequently the variance of predictors coefficients are inflated (Kumar *et al.* 2019). VIF is a tool used to measure and quantify how much the variance is inflated. To interpret the value of VIF the following rule is used (Table 1).

Table 1. VIF rule followed in the study

VIF Value	Conclusion
VIF = 1	Not correlated
$1 < \text{VIF} \leq 3$	Moderately correlated
VIF > 3	Highly correlated

Multiple regression analysis: The multiple regression analysis was done as suggested by Draper and Smith (1987) with the help of following model:

$$Y_i = a + b_1X_1 + b_2X_2 + \dots + b_nX_n + e_i$$

Where, Y_i , Variable to be predicted; $a, b_1, b_2, \dots, b_n$, Unknown parameters to be estimated; $X_1, X_2, \dots, X_n$, Traits whose values are known; e_i , Random error; NID $(0, \sigma^2)$.

Estimation of coefficient of determination (R^2): The coefficient of determination (R^2) indicates that out of hundred per cent of variability of the prediction how much variation was contributed by a set of independent variables and is expressed in terms of percentage. The R^2 estimated in different models from analysis of variance.

Development and selection of optimum model for prediction of lifetime traits in Haryana cattle: Research workers often use models to approximate unknown relationship between a set of predictor variables and the response variable. They try different types of models, which explain the variability in the data in a better way. The main objective of the model building is to predict response variable using the predictor variables (Kumar *et al.*, 2019). Optimum model(s) were developed for prediction of variables in Haryana cattle using combination of AIC value, BIC value, MCp value, R^2 and RMSE (root mean squares error) for each model. The models with lower AIC, BIC and RMSE values with higher co-efficient of determination (R^2) were selected as optimum.

RESULTS AND DISCUSSION

Variance inflation factor (VIF) was used to know the correlations existing among the independent traits in the

study. After use of various combinations (one or more than one trait combination) of AFC, FDP, FCI and FLTMY as independent variables to predict the lifetime traits LT2MY (total milk yield up-to second lactation), LT3MY (total milk yield up-to third lactation), LT4MY (total milk yield up-to fourth lactation) and LT5MY (total milk yield up-to fifth lactation) we observed the high correlations between FDP and FCI during the study. For estimating unbiased co-efficient of determination, the FDP was excluded from VIF analysis. On excluding FDP, the lower VIF values for AFC, FCI and FLTMY were observed which means that these traits were not correlated.

Multiple regression analysis was carried out for the prediction of lifetime traits, viz. LT2MY, LT3MY, LT4MY and LT5MY. First lactation production and reproduction traits were taken as independent variables viz. AFC, FCI and FLTMY. Prediction equations for prediction of lifetime traits from first lactation traits are presented in the table 2. The best model for the lifetime traits were selected based on their higher co-efficient of determination (R^2) value and lower RMSE, AIC and BIC values. MCp was also estimated for interpretation of result.

Lifetime 2 milk yield (LT2MY): To predict LT2MY the production and reproduction traits like AFC, FCI and FLTMY were used as independent variables for multiple linear regression method. Step wise regression procedure was used to predict the lifetime 2 milk yield from first lactation traits, viz. AFC, FCI and FLTMY under the study. On analysis the co-efficient of determinations of 0.007, 1.49, 75.46, 1.51, 75.46, 75.47 and 75.47% for LT2MY were observed when AFC, FCI, FLTMY, AFC-FCI, AFC-FLTMY, FCI-FLTMY and AFC-FCI-FLTMY were used as independent variables, respectively. The models for prediction of LT2MY using production and reproduction traits developed are presented in the table 2. The developed models were in different combination of AFC, CI and FLTMY traits. Out of 07 models developed, model number 03 having one independent variable i.e., first lactation total milk yield (FLTMY) has fulfilled many of the criterions such as higher R^2 (co-efficient of determination) and lower RMSE, AIC and BIC values than the rest of the models.

Hence, model number 03 was adjudged as the best optimum model for the prediction of LT2MY. The linear regression equation developed is given below.

$$\text{LT2MY} = 330.24 + 1.627 \text{ FLTMY}$$

The estimated co-efficient of determination from the model was 75.46%. The R^2 value of prediction was relatively high suggesting that the relationship between the predictors and the response variable is linear.

Lifetime 3 milk yield (LT3MY): To predict LT3MY the production and reproduction traits like AFC, FCI, and FLTMY were used as independent variables for multiple linear regression method. The analysis for predicting lifetime 3 milk yield was done with independent traits, viz. AFC, FCI, FLTMY, AFC-FCI, AFC-FLTMY, FCI-FLTMY and AFC-FCI-FLTMY and co-efficient of determinations

of 1.5×10^{-6} , 1.39, 66.63, 1.39, 66.68, 66.64 and 66.69% were observed, respectively. In this study, the FLTMY had a more influential effect on animals LT3MY. The models for prediction of LT3MY using first lactation production and reproduction traits developed are presented in the table 2. The developed models were in different combination of AFC, FCI and FLTMY traits. Out of 07 models developed model number 10 having FLTMY as independent variable had fulfilled the most criteria such as highest R^2 (co-efficient of determination) and lower RMSE, AIC and BIC values. Hence, the model number 10 was adjudged as the best optimum model for the prediction of LT3MY. The linear regression equation developed is given below.

$$LT3MY = 760.39 + 2.182 X_1$$

The estimated co-efficient of determination from the model was 66.63%. The R^2 value of prediction was high suggesting that the relationship between the predictors and the response variable is linear. Dalal *et al.* (2004) and Shinde *et al.* (2010) reported lower estimates of co-efficient of determination (R^2 value) for predicting LT3MY. Satija (1989) had observed lower R^2 value for predicting LT3MY in Kankrej cattle by using AFC, FL305DMY, FLL, FSP and FDP as independent traits. Kumar (2003) had observed similar R^2 in Sahiwal cattle by using FLL, F305DMY and

FL305DFY as independent traits for predicting LT3MY.

Lifetime 4 milk yield (LT4MY): To predict LT4MY the first lactation production and reproduction traits like AFC, FCI, and FLTMY were used as independent variables for multiple linear regression method. The models for prediction of LT4MY were developed using first lactation production and reproduction traits and are presented in the Table 2. The developed models are in different combinations of AFC, FCI and FLTMY traits. In total 07 models were developed for prediction of LT4MY. This was observed that the model number 17 having first lactation total milk yield as independent variable had fulfilled the most criteria such as relatively highest R^2 (co-efficient of determination) and lower RMSE, AIC and BIC values. Hence, model number 17 was adjudged as the best optimum model for the prediction of LT4MY. The linear regression equation developed is given below.

$$LT4MY = 1295.20 + 2.675 FLTMY$$

The estimated co-efficient of determination from the model was 59.40%. The R^2 value of prediction was relatively high suggesting that the relationship between the predictors and the response variable is linear.

Lifetime 5 milk yield (LT5MY): To predict LT5MY the first lactation production and reproduction traits like AFC,

Table 2. Prediction equations for predicting lifetime traits from first lactation traits

Model no.	Dependent variable	Independent variable	R^2 (%)	Linear regression equation
1	LT2	AFC	0.007	$LT2 = 1958.52 - 0.017AFC$
2	LT2	FCI	1.49	$LT2 = 1610.93 + 0.535FCI$
3	LT2	FLTMY	75.46	$LT2 = 330.24 + 1.627FLTMY$
4	LT2	AFC, FCI	1.51	$LT2 = 1656.73 - 0.027AFC + 0.537FCI$
5	LT2	AFC, FLTMY	75.46	$LT2 = 318.81 + 0.006AFC + 1.628FLTMY$
6	LT2	FCI, FLTMY	75.47	$LT2 = 311.99 + 0.033FCI + 1.626FLTMY$
7	LT2	AFC, FCI, FLTMY	75.47	$LT2 = 301.96 + 0.006AFC + 0.033FCI + 1.626FLTMY$
8	LT3	AFC	1.5×10^{-6}	$LT3 = 2957.58 - 0.004AFC$
9	LT3	FCI	1.39	$LT3 = 2510.84 + 0.753FCI$
10	LT3	FLTMY	66.63	$LT3 = 760.39 + 2.182FLTMY$
11	LT3	AFC, FCI	1.39	$LT3 = 2539.28 - 0.017AFC + 0.755FCI$
12	LT3	AFC, FLTMY	66.68	$LT3 = 647 + 0.066AFC + 2.183FLTMY$
13	LT3	FCI, FLTMY	66.64	$LT3 = 778.73 - 0.035FCI + 2.184FLTMY$
14	LT3	AFC, FCI, FLTMY	66.69	$LT3 = 667.07 + 0.067AFC - 0.041FCI + 2.186FLTMY$
15	LT4	AFC	2.4×10^{-5}	$LT4 = 4055.31 + 0.020AFC$
16	LT4	FCI	2.09	$LT4 = 3330.88 + 1.324FCI$
17	LT4	FLTMY	59.40	$LT4 = 1295.20 + 2.675FLTMY$
18	LT4	AFC, FCI	2.09	$LT4 = 3320.06 + 0.006AFC + 1.324FCI$
19	LT4	AFC, FLTMY	59.40	$LT4 = 1296.56 - 0.0008AFC + 2.675FLTMY$
20	LT4	FCI, FLTMY	59.43	$LT4 = 1210.4 + 0.167FCI + 2.665FLTMY$
21	LT4	AFC, FCI, FLTMY	59.43	$LT4 = 1214.54 - 0.002AFC + 0.167FCI + 2.665FLTMY$
22	LT5	AFC	6.2×10^{-5}	$LT5 = 5269.17 - 0.041AFC$
23	LT5	FCI	2.98	$LT5 = 3976.05 + 2.183FCI$
24	LT5	FLTMY	55.03	$LT5 = 1838.89 + 3.169FLTMY$
25	LT5	AFC, FCI	2.98	$LT5 = 4034.86 - 0.036AFC + 2.183FCI$
26	LT5	AFC, FLTMY	55.04	$LT5 = 1769.74 + 0.042AFC + 3.170FLTMY$
27	LT5	FCI, FLTMY	55.43	$LT5 = 1425.63 + 0.813FCI + 3.128FLTMY$
28	LT5	AFC, FCI, FLTMY	55.44	$LT5 = 1354.76 + 0.043AFC + 0.813FCI + 3.129FLTMY$

FCI, and FLTMY were used as independent variables for multiple linear regression method. Lifetime 5 milk yield (LT5MY): The analysis for predicting lifetime 5 milk yield was done with independent traits, viz. AFC, FCI, FLTMY, AFC-FCI, AFC-FLTMY, FCI-FLTMY and AFC-FCI-FLTMY and co-efficient of determinations of 6.21×10^{-5} , 2.98, 55.03, 2.98, 55.04, 55.43 and 55.44% were observed, respectively. In the present study the FLTMY had a more influential effect on animals LT5MY. The models for prediction of LT5MY were developed using first lactation production and reproduction traits and are presented in the Table 2. The developed models were in different combinations of

AFC, FCI and FLTMY traits. Out of 07 models developed model number 24 having first lactation total milk yield has fulfilled the most criteria such as highest R^2 (co-efficient of determination) and lower RMSE, AIC and BIC values. Hence, model number 24 was adjudged as the best optimum model for the prediction of LT5MY. The linear regression equation developed is given below.

$$\text{LT5MY} = 1838.89 + 3.169 \text{ FLTMY}$$

The estimated co-efficient of determination from the model was 55.03%. The R^2 value of prediction was quite high suggesting that the relationship between the predictors and the response variable is linear. Dalal *et al.* (2004) observed higher R^2 value by using first lactation length, first lactation milk yield and second lactation milk yield as independent traits for predicting LT5MY in Haryana cattle. Chander (1977) had observed lower co-efficient of determination (R^2) in Tharparkar cattle by using AFC, FL305DMY and BE as independent traits for predicting LT5MY. Kumar (2003) had observed lower R^2 in Sahiwal cattle by using AFC, F305DMY and FL305DFY as independent traits for predicting LT5MY.

In the present study, lifetime milk yield predictions modeling was done with the help of multiple linear regression methods. It was found that early expressed lifetime traits had more accuracy in prediction. In conclusion, these models can be improved with incorporation of other production and reproduction traits such as growth data, age at maturity and other traits for more accurate prediction of lifetime milk yield.

REFERENCES

- Akaike H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**: 716–23.
- Arthur P F, Makarechian M, Beng R J and Weingardt R. 1992. Longevity and lifetime productivity of cows in a purebred Hereford and two multibred systematic groups under range conditions. *Journal of Dairy Science* **71**: 1142–47.
- Chander S. 1977. 'Genetic evaluation of lifetime production and reproduction in Tharparkar cattle.' MSc thesis submitted to Kurukshetra University, Kurukshetra (Haryana), India.
- Dalal D S, Malik Z S, Chhikara B S and Chander R. 2004. Prediction of lifetime milk production from early lactation traits in Haryana cattle. *Indian Journal of Animal Sciences* **74**(11): 1145–49.
- Draper N R and Smith H. 1987. *Applied Regression Analysis*. John Wiley and Sons Inc., New York.
- Kumar D. 2003. Genetic studies on breeding efficiency in crossbred cows. *Indian Journal of Animal Sciences* **73**(10): 1180–81.
- Kumar V, Chakravarty A K, Magotra A, Patil C S and Shivahare P R. 2019. Comparative study of ANN and conventional methods in forecasting first lactation milk yield in Murrah buffalo. *Indian Journal of Animal Sciences* **89**(11): 1262–68.
- Kuralkar S V and Raheja K L. 1997. Relationship among early performance, lifetime production and reproduction traits in Murrah buffaloes. *Indian Journal of Animal Sciences* **67**(9): 798–801.
- Mallows C L. 1973. Some Comments on Cp. *Technometrics* **15**: 661–75.
- Patil C S, Chakravarty A K, Singh A, Kumar V, Valsalan J and Vohra V. 2013. Development of a predictive model for daughter pregnancy rate and standardization of voluntary waiting period in Murrah buffalo. *Tropical Animal Health and Production* **46**(1): 279–84.
- Satija C S. 1989. 'Construction of selection indices in Kankrej cattle.' PhD thesis submitted to National Dairy Research Institute, Karnal (Haryana), India.
- Schwarz G E. 1978. Estimating the dimension of a model. *Annals of Statistics* **6**: 461–64.
- Shetkar M, Kumar V, Singh S P, Singh Y Kumar M and Nath S. 2021b. Age at first calving and lifetime performance of Haryana cattle at organized farms. *Indian Journal of Animal Sciences* **91**(12): 1106–08.
- Shetkar M, Kumar V, Singh S P, Singh Y Kumar M and Singh K. 2021a. Genetic analysis of first dry period and lifetime performance in Haryana cattle. *Indian Journal of Animal Sciences* **91**(12): 1103–05.
- Shinde N V, Mote M G, Khutal B B and Jagtap. 2010. Prediction of lifetime milk production on the basis of lactation traits in Phule Triveni crossbred cattle. *Indian Journal of Animal Sciences* **80**(10): 986–88.
- Valsalan J, Chakravarty A K, Patil C S, Dash S, Mahajan A, Kumar V and Vohra V. 2014. Enhancing milk and fertility performances using selection index developed for Indian Murrah buffaloes. *Tropical Animal Health and Production* **46**(6): 967–74.