Multiple linear regression analysis using monthly test day milk yield predicting the first lactation production performance for sire evaluation in Murrah buffaloes

Manvendra Singh¹, Ved Prakash², Shaktikant Das³, Sonam Dixit⁴ and AK Gupta⁵

Received: 25 September 2020 / Accepted: 01 November 2020 / Published online: 28 February 2021
© Indian Dairy Association (India) 2021

Abstract: A total of 9071 first lactation monthly test-day milk yield records of Murrah buffaloes were used to predict the first lactation 305-day milk yield (FL305DMY) by using stepwise backward regression analysis. For the prediction of FL305DMY best combination of monthly test-day milk yields were selected based on adjusted R² and RMSE values. The objective of the study was to compare various methods of sire evaluation viz., least squares, simple regressed least squares, best linear unbiased prediction sire model and best linear unbiased prediction animal model in terms of accuracy and efficiency. The methods were compared on the basis of error variance, coefficient of determination, coefficient of variation and rank correlations among the methods. The accuracy of prediction of FL305DMY from monthly test-day milk yields were observed to be best for TD-2 (45th day), TD-4 (105th day) and TD-6 (165th day) combination with BLUP-AM as the most efficient method for sire evaluation. Individual, key monthly TD-6 (165th day) milk yield has high rank correlation with EBVs obtained from actual 305-day milk yield. It was concluded that the optimum combination of TD-2 (45th day), TD-4 (105th day) and TD-6 (165th day) or individual TD-6 (165th day) can be used for genetic evaluation of Murrah sires.

Keywords: FL305DMY, Multiple linear regression, Murrah buffalo, Test-day milk yield

Introduction

India is the home of the finest buffalo germplasm of the world. The country is bestowed with some of the most renowned buffalo breeds; Murrah, Nili-Ravi, Mehasana, Jaffrabadi. Presently the country is having 109.85 million heads of buffaloes (Livestock Census, 2020), constituting over 54.53 % of the total world buffalo population and holds first position in the world. Over the last few decades the contribution of buffaloes in the total milk production of the country is highest than any other livestock species. The contribution of buffaloes to the total milk production of India (187.75 million tonnes) is about 91.81 million tonnes, i.e. 49% (BAHS, 2019).

Murrah is one of the best milch breed of buffaloes with superior genetic potential for milk production and is used for grading up the local non-descript buffaloes under the genetic improvement programmes across the country. The ultimate objective of any breed improvement programme is to bring fast and accurate genetic gain in the population. In the animal selection, female (dam) has limited scope due to inadequate number of replacement stock. As sires are considered as more than half of the herd, the key step is to identify and select the superior sires on the basis of their breeding values to bring desirable rate of genetic gain. An early and accurate judgment of sire’s breeding value is crucial for long term genetic progress in the population.

In country like India, sires are generally evaluated on the basis of FL305DMY records of their daughters. This results in enhanced generation interval, lower genetic gain per unit of time, fewer numbers of daughters per sire due to smaller herd size, costly and time consuming of field recording of performance data etc. As a suitable alternative to these constraints, the test-day milk yield (TDMY) models have recently evoked considerable interest of the animal breeders for sire’s selection in buffaloes (Chakraborty et al. 2010, Singh and Taylor, 2013, Chitra et al. 2016, Sahoo et al. 2019).

The different methods of sire evaluation viz., least squares (LSM), simple regressed least squares (SRLS), best linear unbiased prediction (BLUP) and restricted maximum likelihood method (REML) have been used by different workers (Kumar et al. 2015,
Chitra et al. 2016, Sahoo et al. 2019). The main advantage of using TDMY models for sire evaluation is the possibility of selecting sires at younger age, resulting in reduced generation interval, increased intensity of selection attributed to the availability of more numbers of records on daughters. This will eventually lead to higher rates of genetic gain in milk yield. In addition, an animal can be included in the evaluation if it has at least one measurement and the larger amount of available data will increase sires evaluation accuracy and minimize possible bias due to excluding incomplete lactations (Ptak and Schaeffer, 1993, Schaeffer et al. 1977). Prediction of FL305DMY from TDMY records at various intervals has higher accuracy (> 80%) of prediction because of high genetic association between TDMY and complete milk production records (Chakraborty et al. 2010, Dongre et al. 2012, Sahoo et al. 2019).

Although test-day milk yield records offer greater advantage compared to 305-day milk yield in selection schemes but the investigation on fitting of prediction equations using monthly test-day milk yield records for sire evaluation is scanty in Murrah buffaloes. Hence, the present study was aimed to fit and compare the different prediction equations of first lactation 305-day milk yield using monthly test-day milk yields for sire evaluation in Murrah buffalo.

Materials and Methods

The present investigation was conducted on first lactation 9071 monthly test-day milk yield records of 965 Murrah buffaloes that calved during 1977 to 2012 sired by 98 bulls. The data on monthly test-day milk yields and first lactation 305 days or less milk yield were collected from history cum-pedigree sheets and daily milk yield recording registers maintained at Dairy Cattle Breeding Division, ICAR-National Dairy Research Institute, Karnal. Milk yield recorded on 15th day of calving was taken as first monthly test-day milk yield (TD1). Next 9 test-day milk yield records were taken at 30 days interval from first test-day (TD2 – TD10). The first lactation 305 days or less milk yield was also used in the present study. Only those records of Murrah buffaloes were considered that had produced milk for at least 100 days and minimum of 500 kg. To ensure the normal distribution, the outliers were removed and data within the range of mean ± 3SD were only considered for the present study.

Prediction of FL305DMY

Different monthly test-day milk yields were used for the prediction of FL305DMY by stepwise backward multiple linear regression analysis (SAS Enterprise Guide 4.3). Following model was used:

\[ Y_i = a + b_1 \times X_{1i} + b_2 \times X_{2i} + b_3 \times X_{3i} + b_4 \times X_{4i} + b_5 \times X_{5i} \]

Where, \( Y_i \) is the estimated first lactation 305-day milk yield of the \( i \)th animal; \( X_{1i} \) is the test-day record of \( i \)th animal; \( a \) is an intercept; \( b \) is the regression coefficient of first lactation 305-day milk yield on test-day records.

Sire evaluation

The breeding values of 98 Murrah sires with five or more daughters were estimated from the actual and predicted FL305DMY and key monthly test-day milk yields by regression analysis by applying four sire evaluation methods i.e., Least Squares Method (LSQ), Simple Regressed Least Squares (SRLS), Best Linear Unbiased Prediction Sire Model (BLUP-SM) and Best Linear Unbiased Prediction Animal Model (BLUP-AM).

Results and Discussion

Prediction of FL305DMY

The estimated intercept values, regression coefficients, adjusted \( R^2 \) and RMSE values for prediction of 305-day milk yield by best prediction equations using single, two and three test day combinations are presented in Table 1.

It was observed that the regression equation with 2 important variables TD-2 and TD-4 explained about 71% variability with RMSE value 218.37 kg of prediction in 305-day milk yield. Addition of one more variable TD-6 increased the predictability to about 83% i.e. an increase of 12% with RMSE value 163.18 kg. Hence regression equation with 3 variables viz. TD-2 (45th day), TD-4 (105th day) and TD-6 (165th day) was considered more appropriate (early test days with high accuracy) for prediction of FL305DMY with about 83% accuracy and 163.18 kg RMSE value. In a study on Murrah buffalo, Singh and Rana (2008) observed that the accuracy of prediction of 305-day milk yield using monthly test-day milk yields varied between 42% (TD-1) and 67% (TD-6). Sahoo et al. (2019) also observed that the regression equation with 2 important weekly test days (TD-7 i.e., 48th day and TD-22 i.e., 153rd day) explained about 86% variability with RMSE 162.41 kg and with 3 weekly test days (TD-7 i.e., 48th day, TD-22 i.e., 153rd day, and with 3 weekly test days (TD-7 i.e., 48th day, TD-22 i.e., 153rd day)

Table 1 Best prediction equations and their accuracy for estimation of FL305DMY

| Test-Days | Best Prediction Equations | Adjusted \( R^2 \) (%) | RMSE (kg) |
|-----------|---------------------------|------------------------|-----------|
| 1         | \( Y_i = 226.31 + 213.27 \times TD-4 \) | 65.95% | 247.06 |
| 1         | \( Y_i = 431.20 + 214.19 \times TD-6 \) | 69.89% | 242.10 |
| 2         | \( Y_i = 73.08 + 76.51 \times TD-2 + 155.16 \times TD-4 \) | 70.96% | 218.37 |
| 3         | \( Y_i = -35.53 + 62.73 \times TD-2 + 75.34 \times TD-4 + 125.39 \times TD-6 \) | 82.76% | 163.18 |
day and TD-34 i.e., 237th day) explained about 88% variability with RMSE 147.14 kg, respectively in Murrah buffalo. However, Dongre et al. (2012) predicted the FL305DMY with R²-value of 81.49 to 85.16% and RMSE 190.4 to 253.0 kg for different fortnightly test day milk yield combinations in Sahiwal cattle. Prediction of 305-day milk yield using individual key monthly test-day milk yields (TD-6 and TD-4) which have higher genetic and phenotypic correlations with 305-day milk yield gave prediction accuracy and RMSE value around 70%, 242.10 and 66%, 247.60 respectively.

**Sire evaluation**

The average estimated breeding values (EBVs) for Murrah sires based on the actual and predicted first lactation 305-day or less milk yield and key monthly test-day milk yields of their daughters for different sire evaluation methods are presented in Tables 2. It was observed that the BLUP-SM and BLUP-AM has the highest stability and least variability in terms of highest and lowest EBVs and found to be superior in comparison to other models on the basis of actual and predicted FL305DMY and key monthly test-day milk yields which were in agreement with the findings of Sahoo et al. (2019).

**Effectiveness of sire evaluation methods**

The sire evaluation methods were compared on the basis of results of rank correlations, error variance, coefficient of determination and coefficient of variation.

**Spearman’s Rank Correlations**

It was observed that the highest rank correlation was found between BLUP-SM and BLUP-AM (0.94) while the lowest (0.62) was observed among SRLS and BLUP-AM for actual 305-day milk yield (Table 3). Others workers (Chaudhari et al. 2014, Chitra et al. 2016, Sahoo et al. 2019) have also observed that the BLUP method was one of the most efficient sire evaluation as it has lowest error variance than other methods in Murrah buffaloes. The rank correlations of predicted 305-day milk yield using TD-2 and TD-4 and TD-2, TD-4 and TD-6 with actual 305-day milk yield were found to be highest within a particular method e.g.

| Table 2 Average expected breeding values (EBVs) of Murrah sires for actual and predicted first lactation 305-day milk yield and key monthly test-day milk yields by different sire evaluation methods |
|---|
| Sire evaluation methods | Average EBV (kg) | Number of sires above average EBV | Number of sires below average EBV | Maximum EBV (kg) | Minimum EBV (kg) |
| LSQ | 1755.98 | 41 (41.8%) | 57 (58.2%) | 2428.02 | 1274.62 |
| SRLS | 1794.82 | 49 (50.0%) | 49 (50.0%) | 2328.40 | 1302.61 |
| BLUP-SM | 1845.10 | 42 (42.8%) | 56 (57.2%) | 1950.51 | 1733.94 |
| BLUP-AM | 1805.74 | 47 (47.9%) | 51 (52.1%) | 2087.38 | 1557.51 |
| LSQ | 1750.98 | 42 (42.8%) | 56 (57.2%) | 2336.66 | 1315.46 |
| SRLS | 1800.83 | 44 (44.8%) | 54 (55.2%) | 2352.45 | 1319.06 |
| BLUP-SM | 1826.65 | 43 (43.8%) | 55 (56.2%) | 1968.61 | 1732.94 |
| BLUP-AM | 1799.73 | 42 (42.8%) | 56 (57.2%) | 2165.08 | 1572.31 |
| LSQ | 1752.00 | 40 (40.8%) | 58 (59.2%) | 2384.70 | 1323.11 |
| SRLS | 1805.17 | 49 (50.0%) | 49 (50.0%) | 2300.81 | 1381.10 |
| BLUP-SM | 1835.38 | 47 (47.9%) | 51 (52.1%) | 1924.66 | 1767.77 |
| BLUP-AM | 1804.74 | 44 (44.8%) | 54 (55.2%) | 2120.17 | 1565.59 |
| LSQ | 7.20 | 41 (41.8%) | 57 (58.2%) | 9.70 | 5.01 |
| SRLS | 7.38 | 44 (44.8%) | 54 (55.2%) | 9.60 | 5.40 |
| BLUP-SM | 7.50 | 45 (45.9%) | 53 (54.1%) | 7.98 | 7.15 |
| BLUP-AM | 7.38 | 46 (46.9%) | 52 (53.1%) | 8.84 | 6.53 |
| LSQ | 6.44 | 40 (40.8%) | 58 (59.2%) | 8.99 | 4.58 |
| SRLS | 6.56 | 48 (48.9%) | 50 (51.1%) | 4.87 | 2.29 |
| BLUP-SM | 6.54 | 50 (51.0%) | 48 (49.0%) | 6.99 | 6.21 |
| BLUP-AM | 6.55 | 49 (50.0%) | 49 (50.0%) | 7.92 | 5.39 |
BLUP-SM of predicted 305-day milk yield with BLUP-SM of actual 305-day milk yield and so on for other methods (Table 4).

The comparison of rank correlations of EBVs obtained by using TD-2, TD-4 and TD-6 vis-à-vis TD-2 and TD-4 with EBVs of actual 305-day milk yield revealed that the rank correlations were higher for TD-2, TD-4 and TD-6 by either of the methods used for sire evaluation. In the case of key monthly test-day milk yields, the higher rank correlations were obtained for TD-6 with actual 305-day milk yield by any of the four methods of sire evaluation. The rank correlations were by and large higher for TD-6 followed by TD-4 among different sire evaluation methods.

All the different methods of sire evaluation were compared for their accuracy, efficiency and stability on the basis of the values of error variance, coefficient of determination and coefficient of variation, respectively (Table 5).

**Error variance**

The error variance is one of the criteria of judging the effectiveness of different methods of sire evaluation. An efficient method of sire evaluation must have minimum error variance. The results showed that the error variance of BLUP-AM was lowest as compared to the other methods. Thus, on the basis of error variance the BLUP-AM was considered more efficient followed by LSQ in actual and predicted 305-day milk yield and also for key monthly test-day milk yields. Similar observations were reported by Sahoo et al. (2019) in Murrah buffalo.

**Coefficient of variation (CV)**

The stability of different sire evaluation methods was judge by comparing the estimates of coefficient of variation. The nearer the CV (%) of the population for sire evaluation to the CV (%) of unadjusted data, the more stable is the method. The CV (%) of unadjusted data for 305-day milk yields was 29.21 per cent.

In case of actual 305-day milk yields, the coefficient of variation was most near to unadjusted data in SRLS method (26.03) and far apart in BLUP-AM method (23.59%). The CV of LSQ and BLUP-SM was almost same for actual 305-day milk yield. The coefficient of variation for predicted 305-day milk yield by using TD-2 and
The prediction of 305-day milk yield from monthly test-day milk yields using TD-2 (45th day), TD-4 (105th day) and TD-6 (165th day) combination gave higher accuracy (82.76%) of prediction. Individual, key monthly TD-6 (165th day) milk yield has high rank correlation with EBVs obtained from actual 305-day milk yield. It was concluded that the optimum combination of TD-2 (45th day), TD-4 (105th day) and TD-6 (165th day) or individual TD-6 (165th day) can be used for genetic evaluation of Murrah sires due to their high rank correlations with EBVs of actual 305-day milk yield. BLUP-AM was found most efficient and accurate method followed by BLUP-SM for sire evaluation using actual and predicted 305-day milk yield and key monthly test-day milk yields in Murrah buffalo.

Acknowledgements

The authors are immensely grateful to Director, ICAR-NDRI, Karnal and Head DCB division for providing necessary facilities for conducting this research work.

References

BAHS (2019) Basic Animal Husbandry Statistics. Department of Animal Husbandry, Dairying & Fisheries. Ministry of Agriculture, Govt. of India

Chakraborty D, Dhaka SS, Pander BL, Yadav AS, Singh S, Malik PK (2010) Prediction of lactation milk yield from test day records in Murrah buffaloes. Indian J Anim Sci 80: 244-245

Chaudhari PN, Upadhay MR, Patel AC, Patel SB, Rank DN (2014) A comparison of different methods of sire evaluation for production and reproduction traits of Murrah graded buffaloes. Int J Res Hum Soci Sci 2: 16-24.

Chitra A, Jain A, Kumar M, Ratwan P, Gupta AK (2016) Genetic evaluation of sires for milk production and its composition traits in Murrah buffaloes. Indian J Dairy Sci 69: 721-724

Dongre VB, Gandhi RS, Singh A, Ruhil AP (2012) Comparative efficiency of artificial neural networks and multiple linear regression analysis for prediction of first lactation 305-day milk yield in Sahiwal cattle. Livest Sci 147: 192-197

Table 5 Comparison of relative efficiency and effectiveness of different sire evaluation methods for actual and predicted first lactation 305-day milk yield and key monthly test-day milk yields in Murrah buffalo

| Methods | Error Variance (kg²) | Relative Efficiency | Coefficient of determination (%) | Coefficient of variation (%) |
|---------|----------------------|---------------------|----------------------------------|-----------------------------|
| Actual first lactation 305-day milk yield | | | | |
| LSQ | 205116 | 88.50 | 20.90 | 25.79 |
| SRLS | 218346 | 83.13 | 21.46 | 26.03 |
| BLUP-SM | 207702 | 87.39 | 27.20 | 25.23 |
| BLUP-AM | 181531 | 100 | 44.97 | 23.59 |
| Predicted first lactation 305-day milk yield (TD-2+TD-4) | | | | |
| LSQ | 144584 | 84.57 | 22.35 | 21.71 |
| SRLS | 158507 | 77.14 | 29.07 | 22.10 |
| BLUP-SM | 145368 | 84.11 | 29.40 | 21.18 |
| BLUP-AM | 122278 | 100 | 50.37 | 19.42 |
| Predicted first lactation 305-day milk yield (TD-2+TD-4+TD-6) | | | | |
| LSQ | 174615 | 85.59 | 20.52 | 23.85 |
| SRLS | 187776 | 79.60 | 26.92 | 24.00 |
| BLUP-SM | 176148 | 84.85 | 25.6 | 23.25 |
| BLUP-AM | 149470 | 100 | 46.21 | 21.42 |
| Key monthly test-day milk yield (TD-4) | | | | |
| LSQ | 3.12 | 85.89 | 20.56 | 24.52 |
| SRLS | 3.35 | 80 | 26.80 | 24.78 |
| BLUP-SM | 3.13 | 85.62 | 25.60 | 23.98 |
| BLUP-AM | 2.68 | 100 | 45.90 | 22.16 |
| Key monthly test-day milk yield (TD-6) | | | | |
| LSQ | 2.83 | 80.91 | 19.18 | 26.11 |
| SRLS | 2.95 | 77.62 | 23.38 | 26.17 |
| BLUP-SM | 3.50 | 65.42 | 17.40 | 29.06 |
| BLUP-AM | 2.29 | 100 | 48.18 | 23.07 |
Kumar V, Chakravarty AK, Patil CS, Shivahre PR, Gupta AK (2015) Comparison among different conventional methods of sire evaluation in Murrah buffalo. Indian J Anim Sci 85: 670-673
Livestock Census (2020) 20th All India Livestock Census. Department of Animal Husbandry, Dairying & Fisheries. Ministry of Agriculture, Govt. of India.
Ptak E, Schaeffer LR (1993) Use of test day yields for genetic evaluation of dairy sires and cows. Livest Prod Sci 34: 23–34
Sahoo SK, Singh A, Ambhore GS, Dash SK, Dubey PP (2019) Comparative efficiency of different multiple linear regression prediction equations of first lactation 305-day milk yield for sire evaluation in Murrah buffaloes. Indian J Anim Res 53: 1287-1291
Schaeffer LR, Minder CE, McMillan I, Burnside EB (1977) Non-linear techniques for predicting 305-day lactation production of Holsteins and Jerseys. J Dairy Sci 60: 1636-1644
Singh A, Rana JS (2008) Prediction of 305-day milk yield based on test-day values in Murrah buffaloes. Indian J Anim Sci 78: 1131-1133
Singh S Tailor SP (2013) Prediction of 305 days first lactation milk yield from fortnightly test and part yields. Indian J Anim Sci 83: 166-169