Indian Journal of Animal Sciences **92** (3): 378–381, March 2022/Article
https://doi.org/10.56093/ijans.v92i3.122274

Predicting breeding values in Murrah buffaloes using random regression models

ANAND JAIN[1]✉, ASHISH RANJAN[2], I D GUPTA[3] and H K SINGH[4]

ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana 132 001 India

Received: 30 October 2020; Accepted: 2 December 2021

ABSTRACT

Various random regression models with different order of Legendre polynomials for permanent environmental and genetic effects were constructed to evaluate Murrah buffaloes under field condition. A total of 24,160 test-day (TD) milk yield records of 2,336 buffaloes belonging to six districts of Haryana state from 2014 to 2019 under CHRS Rohtak were considered for estimating the random covariate coefficients using REML algorithm. The R^2 and BIC improved up to seventh order of the polynomials but decreased at 8^{th} order. The lowest value of AIC was for 8^{th} order of polynomial. The models with LPs of order 5 and above gave an improved fit of the lactation curve having high R^2 value (>98%). However, the number of parameters increased from 13 for LP of order 3 to 73 for LP of order 8 indicating over parameterization of models. Best order of fit was 7^{th} order having highest R^2 (99.23%), lowest AIC and lowest BIC value. The lactation curves plotted from the ETDMYs obtained under RRMs using LP5 and LP7 were smoother as compared to the lactation curve obtained from arithmetic mean. The 301 day lactation yield under RRMs using LP of order 7 and 5 averaged 3470.42±2.76 kg and 3460.72±3.10 kg, respectively. The milk yield ranged from 3,456.38 kg in Bhiwani district to 3,480.86 kg in Sonipat district under LP7 model while it ranged from 3,426.35 kg in Jhajjar district to 3,503.95 kg in Bhiwani district under LP5 model. It was observed that less than 50% were common in top 1% of total buffaloes under the two RRMs. The RRM using LP of order 7, which provided more accurate breeding values, must be considered for selection of top ranking/ superior animals for breeding purpose.

Keywords: Breeding value, Murrah buffaloes, Prediction, Random regression model

Typical procedures used for prediction of estimated breeding values of dairy animals in India are based on least-squares, contemporary comparison methods and sometimes on BLUP models. Most methods of predicting 301/ 305 day milk yields do not consider the covariances among test-day yields, which are known to improve the accuracy of predictions. Prediction of breeding values using 301 day lactation yield might be biased due to non-consideration of individual genetic physiology and persistency (Ptak and Schaeffer 1993). On the other hand, random regression models (RRMs) directly use the test day milk yield (TDMY) to overcome the problems arising from the use of single value of 301 day milk yield.

Use of test day milk yields for genetic evaluation of dairy animals instead of 301 day milk yield has gained considerable importance in recent years (Jamrozik and Schaeffer 1997, Schaeffer *et al*. 2000, Kokate *et al.* 2013).

Present address: [1]ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana. [2]1st Class Veterinary Hospital, Wazirganj, Gaya, Animal and Fish Resources Department, Government of Bihar. [3]ICAR-National Dairy Research Institute, Karnal, Haryana. [4]Central Herd Registration Scheme, Rohtak, Haryana. ✉Corresponding author email: anand.nbagr@gmail.com

RRMs are useful to fit the TDMY data in the regression curve, treating the data as random time series (Jamrozik and Schaeffer 1997). Further, RRMs permit utilization of the covariance among test day milk yields and, thereby, improve the accuracy of 301 day milk yield predictions, help in reducing the cost of milk recording because only about 10 TDMYs need to be recorded instead of daily milk records of the complete lactation, and more number of records (TDMY) per animal become available rather than single record as in case of 301 day milk yield. Therefore, present study was conducted to estimate the breeding values of Murrah buffaloes under field conditions through analysis of test day milk yield (TDMY) records using random regression models (RRMs).

MATERIALS AND METHODS

First lactation test day milk yield (TDMY) records of Murrah buffaloes generated under Central Herd Registration Scheme, Rohtak were used for this analysis. The TDMY was recorded on superior buffaloes true to Murrah breed in the farmers' herds during 2014–18. The analyzed records pertained to six districts of the Haryana state, viz. Rohtak, Jhajjar, Jind, Hisar, Sonipat, and Bhiwani.

Buffaloes with minimum of 6 test day milk yield (TDMY) records were considered for analysis. The TDMY records used for analysis were 4660, 6354, 7303, 854, 2665, 2324 obtained from 448, 618, 708, 82, 255, 225 buffaloes from Rohtak, Jhajjar, Jind, Hisar, Sonipat, Bhiwani district respectively. Thus, a total of 24,160 TDMY records pertaining to 2,336 Murrah buffaloes were analysed.

Data were analyzed under RRMs using WOMBAT (Meyer 2007). Cubic to octic orders of Legendre polynomial (LP) functions were used for fitting fixed, animal and permanent environment effects. Fixed effect considered were the district in which the buffalo performed and year-month of recording of TDMY. The TDMY records were classified into 65 group of year-month of recordings. The residual variance structure was chosen as homogeneous. The criteria of best fit considered for the models included R^2 values, Log likelihood (Log L), Akaike information criterion (AIC), Bayesian information criterion (BIC) and number of parameters.

The model used for analysis is given below:

$$Y_{hijl} = \mu + d_h + \sum_{m=1}^{n}\beta_{ilm}Z_{ilm} + \sum_{m=1}^{n}\alpha_{jlm}Z_{jlm} + \sum_{m=1}^{n}p_{jlm}Z_{jlm} + e_{hijl}$$

where Y_{hijl}, l^{th} test day record of j^{th} buffalo performed in i^{th} year-month of recording in h^{th} district; μ, population mean; d_h is the fixed effect of h^{th} district (h=1 to 6); α_{ilm}, regression coefficient i^{th} year-month of recording l (i= 1 to 65, l=1 to 11); α_{jlm} and p_{jlm}, regression coefficients of random animal effect and permanent environment effect respectively; Z_{jlm}, covariates for fixed and random regressions; e_{hijl}, random residual effect associated with $Y_{hijl.}$

RESULTS AND DISCUSSION

The data were analyzed under RRMs using cubic to octic order of polynomials. The average test day milk yield based on arithmetic mean and the estimated test day milk yields (ETDMYs) obtained by using cubic to octic order of LPs under RRMs are given in Table 1.

The RRMs using LPs of order 3 and order 4 had lowest Log L values, lowest R^2 values, highest AIC and highest BIC. Therefore, there was no logic to consider the results obtained under these models for further discussion. The only advantage under these models was low number of parameters. On the other hand, RRM using LP of order 8 had the advantage of highest Log L and lowest AIC value but, the R^2 value was lower, BIC was higher and the number of parameters was much higher (over parameterization) under this model when compared with the RRM using LP of order 7. Perusal of the values obtained under various criteria of best fit revealed that RRM using LP of order 5 had considerably high R^2 value, high Log L, reasonably low AIC value, low BIC value and less number of parameters. The RRM using LP of order 7 had lowest BIC value, highest R^2, 2^{nd} lowest AIC value, 2^{nd} highest Log L value and reasonably less number of parameters. In view of this, the results obtained under RRMs using LPs of order 5 and 7 were considered for further discussion of the lactation curve.

The predicted breeding values of TDMYs obtained under various RRMs were adjusted for fixed effects of the district and year-month of recording of TDMY. Further, there was an increase in accuracy of ETDMYs under RRMs because the breeding value on a particular day was estimated taking into consideration its covariance with the TDMYs on other days along the lactation curve. Schaeffer *et al*. (2000) stated that a test-day model provide better accounting for herd-test date environments and confer more accurate genetic selection of bulls and cows by outweighing the short-term disadvantages. Schaeffer and Jamrozik (2008) stated that the application of RRM to test day yields in dairy cattle gave better removal of environmental effects, and therefore more accurate genetic evaluations of cows and bulls. When the genetic and permanent environmental components were both modelled by a polynomial regression, predictions became more accurate (Pool *et al*. 2000)

While estimating ETDMY of a buffalo on a particular

Table 1. Estimated TDMY (kg) at different order of LPs

TD	Avg	LP 3	LP 4	LP 5	LP 6	LP 7	LP 8
5	9.19	12.30	11.22	10.31	9.98	9.65	10.09
35	14.01	13.57	14.02	14.14	14.33	14.33	14.15
65	13.85	14.13	14.82	14.81	14.86	14.79	14.70
95	13.30	14.10	14.43	14.14	14.14	14.09	13.93
125	13.00	13.58	13.47	13.14	13.23	13.22	13.07
155	11.52	12.66	12.36	12.22	12.39	12.35	12.25
185	10.75	11.46	11.32	11.39	11.50	11.42	11.29
215	9.72	10.07	10.36	10.42	10.42	10.37	10.20
245	8.64	8.59	9.28	9.07	9.10	9.09	8.99
275	7.27	7.13	7.69	7.28	7.48	7.41	7.27
305	5.50	5.79	4.99	5.35	5.25	5.28	5.29
Parameters		13	21	31	43	57	73
R^2		89.60	95.90	98.40	98.90	99.20	98.80
Log L		–16050.50	–13209.20	–12185.40	–11831.20	–11600.70	–11528.90
1/2 AIC		16063.50	13230.20	12216.40	11874.20	11657.70	11601.90
1/2 BIC		16116.10	13315.20	12341.80	12048.20	11888.30	11897.20

day, the weightage given to the TDMY on any other day depended upon the covariance between the two points on the lactation curve.

The arithmetic mean and ETDMYs under RRMs using LP5 and LP7 were obtained from 5th through 305th day of lactation at a regular interval of 30 days. The values, so obtained were used to study the lactation curve. The lactation curve is given in Fig. 1.

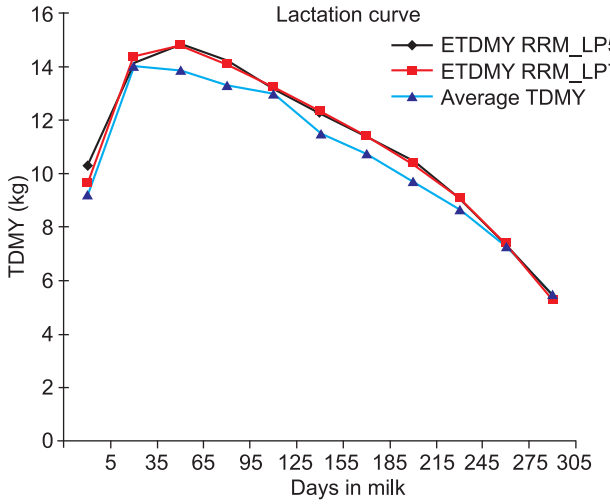

Fig. 1. Lactation curve in Murrah buffaloes.

It was observed that lactation curves plotted from the ETDMYs obtained under RRMs using LP5 and LP7 were smoother as compared to the lactation curve obtained from arithmetic mean. This explains, the accuracy achieved in the daily and lactation breeding values of the buffaloes through RRM analysis. Pool and Meuwissen (1999) stated that higher order Legendre polynomials showed a more flexible shape and the goodness of fit increased with the order of fit in Holstein Friesian cows where differences between LEG5, LEG6 and LEG7 were relatively small. Sesena *et al.* (2010) analysed the additive genetic and permanent environmental random effects of Murrah buffaloes in different herds of Brazil with quadratic to seventh order of polynomial functions and reported that DIM modelled by LP of order 5 to 7 gave best fit and were most adequate to describe the milk yield variation along the lactation.

The differences between the ETDMYs under RRMs and average TDMY were conspicuous from 5th to 245th days of lactation. Further, the RRM using LP of order 7 gave better fit in the very beginning of the lactation as compared the RRM using LP of order 5 and therefore provided more accurate estimates of the breeding values for the whole lactation. The rank correlation between the lactation breeding values obtained under the two RRMs was 0.858±0.013. The repeatability of the ETDMYs ranged from 0.784 for test day 95 to 0.942 for test day 5 under LP7 and from 0.742 for test day 125 to 0.865 for test day 5 under LP5. The repeatability were higher under LP7 when compared with LP5 for all the test days. The repeatability values were highest for test day 5 under both the polynomial.

Table 2. 301-day estimated milk yield under RRMs using LP of order 7 and 5

District	LP5	LP7
Pop. Mean	3460.72±3.10 (2336)	3470.42±2.76 (2336)
Jhajjar	3426.35±5.83 (618)	3459.53±5.32 (618)
Rohtak	3474.11±7.71 (448)	3473.44±7.29 (448)
Bhiwani	3503.95±7.86 (225)	3467.10±6.52 (225)
Hisar	3440.13±12.96 (82)	3456.38±9.40 (82)
Jind	3465.52±5.87 (708)	3476.92±5.20 (708)
Sonipat	3475.64±8.69 (255)	3480.86±7.68 (255)

District wise estimates of 301 day lactation yield were obtained by taking the averages of the breeding values of the buffaloes in each district (Table 2). The 301 day lactation yield under RRMs using LP of order 7 and 5 averaged 3470.42±2.76 kg and 3460.72±3.10 kg, respectively. The milk yield ranged from 3,456.38 kg in Bhiwani district to 3,480.86 kg in Sonipat district under LP7 model while it ranged from 3,426.35 kg in Jhajjar district to 3,503.95 kg in Bhiwani district under LP5 model. The district effect was non-significant under RRM using LP of order 7 while the effect was significant under RRM using LP of order 5.

Lactation (301-day) breeding values of top ranking 24 buffaloes, i.e. 1% of the total buffaloes considered for analysis, were tabulated and the buffaloes common under the two RRMs were studied (Table 3). It was observed that out of 24, only 11 buffaloes, i.e. less than 50% were common under the two RRMs. This implies that the RRM using LP of order 7, which provided more accurate breeding values, must be considered for selection of top ranking/ superior animals for breeding purpose. The animals from the less accurate models identified for selection are likely to be different and, therefore, may impact the genetic gains over the generations.

Under the RRM using LP of order 7, there were 10, 8, 4 and 2 animals from Rohtak, Jind, Jhajjar and Sonipat district respectively which constituted 2.23%, 1.13%, 0.65% and 0.78% of the animals under consideration from the respective district. Therefore, the maximum contribution to the top ranking animals both in absolute and percent terms was highest from the Rohtak district. Under the RRM using

Table 3. District wise number and per cent buffaloes among top ranking 24 buffaloes (1%) under RRM models using LP of order 5 and 7

District	Total number of buffaloes	LP7		LP5	
		Number	Per cent	Number	Per cent
Overall	2336	24	1.00	24	1.00
Jhajjar	618	4	0.65	3	0.49
Rohtak	448	10	2.23	6	1.34
Bhwani	225	0	0.00	0	0.00
Hisar	82	0	0.00	0	0.00
Jind	708	8	1.13	11	1.55
Sonipat	255	2	0.78	4	1.57

LP of order 5, there were 11, 6, 4 and 3 animals from Jind, Rohtak, Sonipat and Jhajjar district respectively which constituted 1.55%, 1.34%, 1.57% and 0.49% of the animals under consideration from the respective district. Under this model, the contribution to the top raking animals was maximum from district Jind in absolute terms and from district Sonipat in per cent terms.

The data generated on test day milk yields on buffaloes in the farmers' herd was used for the current study. The criteria considered to judge the goodness of fit of the model, i.e. R^2 value, log likelihood function, AIC, BIC and number of parameters helped in choosing the best fit model for the analyzed data. The RRM using legendry polynomial of order 7 provided more accurate estimates of the breeding values as compared the raw data. The methodology was helpful in providing breeding values of the milk yield on fixed days of the lactation as well for 301-day lactation yield. More comprehensive studies need to be undertaken to further establish the usefulness of the models for analysis of the data under field conditions as well as to find the loss in accuracy due to non-availability of pedigree.

ACKNOWLEDGEMENTS

The authors are thankful to the Director, ICAR-National Bureau of Animal Genetic Resources, Karnal for providing logistic support for analysis of the data. Financial assistance from Indian Council of Agricultural Research, New Delhi is thankfully acknowledged. We are also thankful to the Deputy Registrar, Central Herd Registration Scheme, Rohtak for providing the data for this study. The data used for the study were generated by a number of stockmen and in the herds of a number of farmers through their hard work. We put on records our greatest sense of gratitude to all of them.

REFERENCES

Jamrozik J and Schaeffer L R. 1997. Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins. *Journal of Dairy Science* **80**: 762–70.

Kokate L S, Singh B R, Gandhi R S, Chakarvarty A K, Gupta A K and Sachdeva G K. 2013. Genetic and non-genetic factors affecting monthly test day milk yields in Karan Fries cattle. *Indian Journal of Animal Sciences* **83**: 385–89.

Meyer K. 2007. WOMBAT—a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *Journal of Zhejiang University Science* **B8**(11): 815–21.

Pool M H, Janss L L G and Meuwissen T H E. 2000. Genetic parameters of Legendre polynomials for first parity lactation curves. *Journal of Dairy Science* **83**: 2640–49.

Pool M H and Meuwissen T H E. 1999. Prediction of daily milk yields from a limited number of test-days using test-day models. *Journal of Dairy Science* **82**: 1555–64.

Ptak E and Schaeffer L R. 1993. Use of test day yields for genetic evaluation of dairy sires and cows. *Livestock Production Science* **34**: 23–34.

Schaeffer L R, Jamrozik J, Kistemaker G J and Doormaal J V. 2000. Experience with a test-day model. *Journal of Dairy Science* **83**: 1135–44.

Schaeffer L R and Jamrozik J. 2008. Random regression models: A longitudinal perspective. *Journal of Animal Breeding and Genetics* **125**: 145–46.