Tropical Agricultural Research Vol. 7 1995, 168-176

Comparison of PC and MR Models for Prediction of Breeding Value of Milk Production in Murrah Buffaloes

S.K. Dhara and A.K. Chakravarty

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

ABSTRACT. *Records of growth, reproduction and production of 687 Murrah buffaloes maintained at NDRI from 1973 to 1993 were used for developing four multiple regression (MR) models for the prediction of breeding value (BV) of 305 days milk yield. Each model consisted of 10 characters at a time. Forty principal components were generated by using the same structure of data. The number of principal components exceeding 95% cumulative variance were 7, 6, 4 and 4 in principal component (PC) models I, II, III and IV, respectively. The efficiency of MR and PC models were compared and Model IV of PC models was found best over MR models. Only four variables/components were required under best principal component model compared to 10 variables of multiple regression model to obtain a better prediction of breeding value of 305 days milk yield in Murrah buffaloes. The coefficient of determination (r^2) of the best model was 18.63% under principal component analysis.*

INTRODUCTION

Many economic traits are important for the prediction of breeding value (BV) of milk production. Although the multiple regression (MR) model can be a major resort in the hand of the breeder for such a prediction, the problems of inter associations among the traits may misinterpret the result (Baker *et al.*, 1988). Principal component (PC) model has the potentiality to make the traits independent of each other by developing principal components each of which consists of linear combinations of all traits with their relative contribution to the breeding value of milk production. Thus, particular principal component(s) may be of greater use in predicting the breeding value of milk production. The present study was, therefore, undertaken to compare the models developed through multiple regression and principal component analyses for the prediction of the breeding value of milk production in Murrah buffaloes.

MATERIALS AND METHODS

The study was based on records of growth, reproduction and production traits of six hundred and eighty seven Murrah buffaloes maintained at the National Dairy Research Institute herd, Karnal. The data were spread over a period of 21 years (1973 to 1993). To adjust the significant effect of periods and seasons, the entire period was classified into four groups depending on the sires used. A group of sires was used for about 4 to 5 years in the farm. Each year was further subdivided into five seasons depending on the meteorological information provided by Central Soil Salinity Research Institute (CSSRI), Karnal (Singh, 1983). The least squares analysis (Harvey, 1975) of fitting constants was used by taking period and season as fixed effects in the model, and significance of each effect on all the traits was judged through least-squares analysis of variance. The least-squares constants obtained for significant effect for a particular trait were used for adjusting that character.

By using paternal half-sibs method, the heritability of 305 days milk yield was found as 0.26±0.08, after necessary adjustment of significant non-genetic factor. The repeatability was computed as 0.28±0.02 for 305 days' milk yield. The breeding values (BV) of 305 days' milk yield were estimated by using the herd average and average deviation of individual buffalo performance from their contemporaries with the help of below given equation.

$$BV = \mu + [nh^2/1+(n-1)r]\ (\bar{X}_i-\overline{XC}_i)$$

where,

μ	=	herd average estimated as 1994.75 kg
n	=	number of lactation records available for each buffalo
h^2	=	heritability
r	=	repeatability of 305 days' milk yield
$(\bar{X}_i - \overline{XC}_i)$	=	average deviation of individual buffalo performance from their contemporaries

Forty principal components were generated in four models. Each model consisted of the combination of ten different traits (which are indicated in the next paragraphs) of Murrah buffalo, since STATGRAPH programme (Anonymous, 1986) available at the Computer Centre, NDRI, Karnal for principal component analysis was able to accommodate a maximum of only ten variables.

In view of the significant genetic associations of BV with growth and production traits, Model I consisted of characters like birth weight (BWT), weight at 12 months (WT12M), weight at 24 months (WT24M), weight at first calving (WFC), first lactation 120 days' milk yield (FL120MY), first lactation overall average (FOA), second lactation 120 days' milk yield (SL120MY), 305 days' milk yield (S305DY) and milk yield per day of first dry period (SYPDFDP). However, since predicting the BV of milk production at the earliest possible is desirable characters such as weight at 12 months, weight at first calving, first service period, first lactation milk yield of 30 days, 60 days, 90 days, 120 days and 305 days, first lactation length and overall average were taken into Model II by considering their genetic association with the breeding value of milk production. Model III considered only the production traits other than part lactation records of the first two parities. Finally, in Model IV only the traits which had the highest genetic association with the breeding value of milk production in Murrah buffaloes were considered.

The eigen values (τ_i) were computed by using correlation matrix ($\underset{\sim}{R}$) and the significance of eigen values were judged by sphericity test of Bartlett (Kendall *et al.*, 1983). Each principal component (PC_i) was generated by normalizing the correlation matrix subjected to the restriction of $\underset{\sim}{U}'_i\underset{\sim}{U}_i = 1$. Thus,

$$PC_i = \underset{\sim}{U}'_i\underset{\sim}{R}\underset{\sim}{U}_i$$

where,

$\underset{\sim}{U}_i$ = normalized vector of i^{th} column.

The contribution of each component in explaining the variance out of the original set of data was determined by the variance accounted by

$$PC_i(\%) = \tau_{i/p} \times 100$$

where,

PC_i = i^{th} principal component

p = number of independent traits.

The number of PC to be extracted was decided by Bartlett's X^2 criterion (Kendall *et al.*, 1983), for the selected number of principal components accounted for more than 95% of variability.

Four PC models were developed by using

$$Y_i = Q_o + \Sigma^n_{i=1} \Sigma^{p-r}_{j=1} Q_j P_{ji} + e_i$$

where,

Y_i	=	breeding value of 305 days' milk production of i^{th} buffalo,
Q_o	=	constant,
Q_j	=	partial principal component regression coefficients for j^{th} PC,
P_{ji}	=	i^{th} observation of j^{th} PC and
e_i	=	random error assumed to be normally independently distributed $(0, \sigma_e^2)$.

Similarly, four MR models were developed by taking ten original traits, with the same combination in each model, with the help of the stepwise regression method; and the efficiency of both PC and MR models were compared as per Draper and Smith (1981).

For comparing the efficiency of MR and PC models developed separately, the use of number of traits/components, the estimated coefficients of determination (r^2%), mean sum of square due to error (MS_e) and predicted mean BV of milk production under each model were taken into consideration.

The intercept values alongwith partial regression coefficients had been used to predict the BV of 305 days' milk production in buffaloes through MR and PC models. The significance of partial regression coefficients of MRA (b_i's) and of PCA (Q_j's) were evaluated by t-test, using the following formula:

t(n-p) = estimated regression coeffients/standard error of regression coefficients,

where,

n = total number of observations and
p = total number of parameters in each model

The standard error of partial regression coefficients obtained by underroot of $E(MS_e)$ for each model multiplied with the corresponding diagnol element of $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$ where $\underset{\sim}{X}$ was the design matrix or matrix of inputs for each model.

The partial regression coefficient indicated the change of BV of 305 days' milk production per unit change of any independent variable in the model by keeping other independent variables at the same level. In the case of MRA, the partial regression coefficients were calculated when the traits were associated; whereas we estimated the partial regression coefficients by making the traits independent of each other i.e. by removing the multi colinearity of traits in PCA.

RESULTS AND DISCUSSION

The estimated intercept values (constant) and partial regression coefficients (b_i's) to the four MR models are presented in Table 1. The total number of principal components exceeded 95% cumulative variance obtained through eigen values, and the percentages of variance were 7, 6, 4 and 4 in Model I, II, III and IV, respectively. The estimated intercept values (constant) and partial principal component regression coefficients (Q_j 's) of four PC models are also presented in Table 1. It has been observed that in almost all principal component models, the intercept as well as the partial principal component regression coefficients were significantly higher than those estimated through MR models (Table 1).

Critical appraisal of Table 2 indicated that the r^2 (%) value of Model IV which consisted of high genetically correlated ten traits was the highest, and therefore, is the best of the models under multiple regression analysis (MRA). With the principal component analysis (PCA) the r^2 (%) value was highest in Model III, which consisted of ten production traits other than part lactation records. The higher r^2 (%) value under Model III in PCA was due to less number of non-identical repeat runs of the observations for some traits in comparison to Model IV of PCA. Draper and Smith (1981) have also imposed a similar opinion. Although there was approximately 1% greater r^2 (%) value observed in Model III in PCA, Model IV could be acceptable as it fulfilled the other criteria of judging the best model. On the other hand, by using the same criterion, Model II under MRA (r^2 = 10.55%) and PCA (r^2 = 9.35%) was found to be the most inferior.

On judging the mean sum of squares due to error (MS_e) of the four models under MRA and PCA, it was observed that in both analyses, Model IV had the minimum MS_e. Therefore, on the basis of MS_e, Model IV was found to be the best of all the models in both analyses. Again, the predicted mean breeding value was also observed to be more accurate through least squares analyses. Moreover, the predicted mean breeding value was more accurate in PCA than MRA, since the expected value of errors was found minimum.

Table 1. **Estimated constants and partial regression coefficients in MRA and PCA* models.**

Model	Constant value b_0 (Q_0)	b_1 (Q_1)	b_2 (Q_2)	b_3 (Q_3)	b_4 (Q_4)	b_5 (Q_5)	b_6 (Q_6)	b_7 (Q_7)	b_8 (Q_8)	b_9 (Q_9)	b_{10} (Q_{10})
I	1062.5049	0.2187	-44.8608	0.6185	0.2075	-0.1044	0.2501	1.5276	0.7782	-0.7733	02237
	(2006.5096)	(69.1299)	(-1.9608)	(-85.2495)	(-3.9689)	(-9.5216)	(-44.2436)	(-65.2153)	-	-	-
II	1195.4216	0.4774	-0.3062	-0.3277	-50.3557	0.3125	-0.6210	-0.7663	0.6062	0.7516	05310
	(1974.7965)	(41.4259)	(8.2121)	(28.3266)	(31.2763)	(-61.2545)	(43.3564)	-	-	-	
III	1289.9441	14.1867	0.0878	-0.3693	-4.3270	40.8250	-2.5843	0.0860	-0.0510	-0.4241	724066
	(1962.5044)	(65.7087)	(19.0145)	(-1.0778)	(-60.5879)	-	-		- -		--
IV	1048.0277	12.8355	-0.0473	-3.8640	0.0984	-0.1011	17.0504	-0.0145	0.0831	-1.0882	11437
	(1994.8372)	(60.4142)	(20.5223)	(-6.1926)	(-16.8586)	-	-	-	-	-	-

* values in parantheses represent the constants and partial regression coefficients of PCA models.

Table 2. Comparison of different models developed through MRA and PCA.

Model	Number of traits	Number of principal components	r^2 (%)		MS_e		Predicted mean BV (kg)	
	MRA	PCA	MRA	PCA	MRA	PCA	MRA	PCA
I	10	7	17.77	17.12	145743.19	143913.44	2006.17	2006.30
II	10	6	10.55	9.35	132320.23	131650.09	1974.59	1974.79
III	10	4	21.18	19.61	169896.80	111204.18	1962.22	1962.50
IV	10	4	25.15	18.63	103330.48	110093.41	1990.01	1994.83

Thus in both MRA and PCA, Model IV was judged the best model. However, as far as the adoption of a model under MRA and PCA were concerned, it could be concluded that Model IV under PCA should be considered for the prediction of breeding value in Murrah buffaloes, because the required number of variables/components in Model IV was only 4, in comparison to 10 variables required in Model IV of MRA. The advantage of selecting Model IV of PCA was that the inclusion of a minimum number of variables/components might help in handling data structure of large size more efficiently, without loosing much information. This is in agreement with the view of Baker *et al.*, (1988). Information is not available on comparison of two kinds of models for the prediction of breeding value of milk production by using multiple regression and principal component analysis.

CONCLUSIONS

Four MR models were developed for the prediction of BV of 305 days' milk yield. Each model consisted of ten characters at a time. Four PC models were also developed by using 7, 6, 4 and 4 principal component variables. It could be concluded that PC Model IV is the best model, compared to MR Model IV, for obtaining more accurate prediction of mean BV of 305 days' milk production in Murrah buffaloes, as the expected value of errors was found minimum under PC Model IV.

ACKNOWLEDGEMENTS

The first author is thankful to N.D.R.I. (Deemed University) for providing financial assistance in the form of a Junior Research Fellowship.

REFERENCES

Anonymous (1986). Statgraphics by Statistical Graphics Corporation, STSC, Inc.

Baker, J.F., Steward, T.S., Long, C.R. and Cartwright, T.C., (1988). Multiple regression and principal component analysis of puberty and growth in cattle. J. Anim. Sci. 66:2147-2158.

Draper, N.R. and Smith, H. (1981). Applied regression analysis. 2nd ed., John Wiley and Sons, Inc. New York.

Harvey, W.R. (1975). Least-squares analysis of data with unequal sub-class numbers, AR-H-4, USDA. Washington, D.C.

Kendall, M.G., Stuart, A. and Ord, J.R. (1983). The advanced theory of statistics, Vol.3. Charles Griffin and Co., HD.

Singh, O.P. (1983). Climate of Karnal. Bull. 8. Central Soil Salinity Research Institute (ICAR). Karnal, India.