Indian Journal of Animal Sciences **87** (9): 1163–1167, September 2017/Short communication

Prediction of lifetime performance traits by principal component analysis in Jersey crossbred cattle at an organized farm of eastern India

POONAM RATWAN[1], AJOY MANDAL[2], MANOJ KUMAR[3] AND ATISH KUMAR CHAKRAVARTY[4]

ICAR-National Dairy Research Institute, Karnal, Haryana 132 001 India
and
Eastern Regional Station, ICAR-National Dairy Research Institute, Kalyani, Nadia, West Bengal 741 235 India

Received: 10 February 2017; Accepted: 27 March 2017

Key words: Early selection, Jersey crossbred, Lifetime performance, PCA

As per All India Livestock Census (2012). India has 190.90 million (37.28%) cattle and 108.7 million (21.23%) buffaloes. Out of total cattle population, about 39.73 million are crossbred cattle in our country. The decade wise trend in livestock population (1997 to 2012) shows a distinct shift in composition of dairy animal stock in favour of buffaloes and crossbred cattle, as their numbers increased by 3.19 and 20.18% respectively, while that of indigenous cattle declined by 8.94% (19th Livestock Census 2012). In India, there are 40 registered breeds of cattle (NBAGR 2016). In spite of the presence of large and diverse cattle genetic resources, the productivity of cows remains low in the country, for various reasons, such as inadequate nutrition, poor genetic potential, inadequate animal health services, harsh climatic conditions and other management related problems. In India, average milk productivity of crossbred cows, indigenous cows and buffaloes is about 7.02, 2.36 and 4.8 kg/day, respectively as per Ministry of Agriculture, GoI (2012–13). Although, crossbred cattle are less in number but are contributing more to the total milk yield of the country as compared to the indigenous cattle.

The profitability of any dairy enterprise is function of overall lifetime production of the cows maintained at a farm. Prediction of lifetime milk production on the basis of early lactation traits with maximum accuracy is one of the criteria of selection for lifetime profitability as it would not be economically viable to wait for an animal to complete its lifespan in the herd (Gopal and Bhatnagar 1972, Gandhi and Gurnani 1988, Pundir and Raheja 1995). There is need to develop efficient planning and management tools in order to increase milk availability. Principal component analysis (PCA) is a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components (Pearson 1901, Lukibisi *et al.* 2008, Hotelling 1933, Rao 1964). 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 predict the performance traits. Thus, particular principal component (s) may be of greater use in predicting the performance traits. Puri and Sharma (1965) studied first lactation yield and age at first calving for prediction of lifetime production and determined their relative importance for selection purposes in Red Sindhi and crossbred cows. Malhotra and Singh (1980) predicted lifetime production (total milk yield in the first 3 lactations) on the basis of early life traits for Red Sindhi cows. The advantages and disadvantages of principal components regression (PCR) and partial least squares (PLS) for livestock management research were investigated by Rugoor *et al*. (2000). Several workers (Pundir *et al.* 2011, Vohra *et al.* 2015 and Khargharia *et al.* 2015) applied principal component analysis technique for analysis of biometric traits to explain the body conformation of animals. Very few literatures are available on use of principal component analysis for prediction of lifetime performance traits. In the present study, first lactation traits, relationships among them and with lifetime milk yield up to fourth lactation were studied along with development of synthetic components (latent) to predict lifetime performance traits for early selection in Jersey crossbred cattle using principal component analysis.

The present study was based on the records of first lactation traits of 159 Jersey crossbred animals maintained at the Eastern Regional Station, ICAR-National Dairy Research Institute, Kalyani, West Bengal spread over a period of 29 years (1986–2014). The normal reproduction and production records were considered and records showing abortion, dystocia and other reproductive disorders were not included in the study. Initially, nine traits viz., dam's weight at calving, age at first calving, first lactation length, peak yield, days to attain peak yield, first calving interval, milk yield per day of lactation length, total milk

Present address: [1,3]Ph.D. Scholar (punam.ratwan@gmail.com, drmanojneemwal@gmail.com), [4]Principal Scientist and Head (ak_chakravarty@yahoo.co.in), AGB Division. [2]Principal Scientist (ajoymandal@gmail.com), AGB Division, ERS, ICAR-NDRI, Kalyani, Nadia, West Bengal.

yield per day of calving interval and 305 days milk yield were considered for the analysis to predict lifetime milk yield up to fourth lactation. Out of these, only 4 traits i.e. peak yield (PY), 305 days milk yield (305MY), milk yield per day of lactation length (MY/LL) and total milk yield per day of calving interval (TMY/CI) were further utilized for prediction purpose due to their medium to high correlation with lifetime time milk yield up to fourth lactation. Each considered trait has its significance like peak yield is the point where the cow reaches the highest milk production level during the entire lactation. Heifers peak at 70 to 75% of mature cows and second lactation cows peak at 90% of mature cows. Normally the peak is reached four to ten weeks after calving. Higher producing animals tend to peak later than low producing ones. A high peak yield normally means a higher total yield. First lactation 305 days milk yield is used as a general criterion for selection of animals and has a significant role in predicting lifetime milk yield. Other two traits considered namely, MY/LL and TMY/CI are also known as milk production efficiency traits. MY/LL was derived by dividing first lactation 305 days milk yield with standard 305 days or less first lactation length whereas TMY/CI was obtained by dividing first lactation total milk yield with first calving interval. Milk production efficiency traits reflect both production and reproduction status of the animals.

To adjust the significant effect of periods and seasons, the entire period was classified into six groups (5 year each) and each year was further subdivided into three seasons i.e. summer (March to June), rainy (July to October) and winter (November to February). The least squares maximum likelihood program of Harvey (1990) was used to estimate and study the effect of genetic and non-genetic factors on PY, 305MY, MY/LL and TMY/CI in Jersey crossbred cattle.

$$\text{Model: } Y_{ijkl} = \mu + S_i + (Sea)_j + P_k + e_{ijkl}$$

where, Y_{ijkl}, l^{th} observation of cattle which is progeny of i^{th} sire calved in j^{th} season and k^{th} period; μ, overall population mean; S_i, random effect of i^{th} sire; $(Sea)_j$, fixed effect of j^{th} season; P_k, fixed effect of k^{th} period and e_{ijkl} is the random residual error associated with each observation which is normally and independently distributed with mean zero and unit variance. The least squares constants obtained for significant effects for a particular trait were used for adjusting that trait. Principal component analysis was carried out using SPSS (2001) statistical package for social sciences.

PCA is a multivariate statistical technique dealing with the reduction of a set of observable variables. The objective of PCA is to account for the maximum portion of variance present in original set of variables with a minimum number of composite variables. Varimax rotation was used for rotation of principal components through transformation of components to approximate a simple structure. Kaiser-Meyer-Olkin (KMO) test of sampling adequacy and Bartlett's test of Sphericity were computed to establish the validity of data set at 1% level of significance. Bartlett test (Bartlett 1950) was performed to check whether the data set of 159 animals with 4 traits could be factored or not. Maxwell (1959) suggested that the test should be used prior to the application of factor analysis. The following formula was used to compute Bartlett's test of Sphericity:

$$x^2 = [(n–1)1/6(2p + 5) \log_e |R|]$$

where, n, sample size; p, number of variables; |R|, determinant of correlation matrix. It follows χ^2 distribution with [p (p–1)/2] degree of freedom.

Means of first lactation traits studied in Jersey crossbred were 11.59±0.92, 2320.92±51.74, 7.75±0.17 and 5.98±0.15 for PY, 305DMY, MY/LL and TMY/CI, respectively. The random effect of sire was highly significant on all the traits. Similar significant effect of sire on lactation traits was reported by several workers (Singh and Gurnani 2004, Verma and Thakur 2013, Ratwan *et al.* 2015, Ratwan *et al.* 2016) in crossbred cattle. Season of calving had significant influence on PY, 305MY and MY/LL but it did not influence TMY/CI. These findings were in accordance with the observations of Mandal *et al.* (2013), Kumar *et al.* (2014) and Ratwan *et al.* (2016 a,b). The present study revealed significant effect of period of calving on all the considered lactation traits which may be due to different management and environmental conditions. Mandal *et al*. (2013), Wondifraw *et al.* (2013), Ratwan *et al.* (2015) and Kumar *et al.* (2016) also reported significant effect of period of calving on different lactation traits. A total of 6 correlations were estimated and all are positively correlated with each other. Correlations among first lactation traits ranged from 0.744 (PY and TMY/CI) to 0.981 (305MY and MY/LL) as evident from Table 1. The KMO observed in the present investigation was 0.825, however, Vohra *et al*. (2015)

Table 1. Phenotypic correlation among four different first lactation traits in 159 adult female Jersey crossbred.

Trait	PY	305MY	MY/LL	TMY/CI
PY		0.828**	0.852**	0.744**
305MY			0.981**	0.913**
MY/LL				0.910**
TMY/CI				

PY, Peak yield; 305MY, 305 days milk yield; MY/LL, milk yield per day of lactation length; TMY/CI, total milk yield per day of calving interval.** P<0.01.

Table 2. Communalities and unique factor of first lactation traits in adult female Jersey crossbred

First lactation trait	Communalities	Unique factor
PY	0.992	0.008
305MY	0.969	0.031
MY/LL	0.973	0.027
TMY/CI	0.955	0.045

where, PY, Peak yield; 305MY, 305 days milk yield; MY/LL, milk yield per day of lactation length; TMY/CI, total milk yield per day of calving interval.

Table 3. Total variance explained by different components in adult female Jersey crossbred (n = 159)

Component	Initial Eigen values			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.620	90.489	90.489	3.620	90.489	90.489	2.332	58.296	58.296
2	0.270	6.755	97.244	0.270	6.755	97.244	1.558	38.948	97.244
3	0.093	2.317	99.562						
4	0.018	0.438	100.000						

reported comparatively low (0.74) KMO in Gojri buffaloes. The KMO revealed the proportion of variance in different first lactation traits caused by underlying factors. The overall significance of correlation tested with Bartlett's test of Sphericity for first lactation traits was significant and provided enough support for the validity of factor analysis of data. The communality estimate ranged from 0.955 (TMY/CI) to 0.992 (PY) and unique factor varied from 0.008 (PY) to 0.045 (TMY/CI) (Table 2). The estimated factor loadings extracted by factor analysis, eigen values and variations explained by each factor are presented in Table 3. Only one component was extracted using Kaiser Rule criterion (Johnson and Wichern 1982) to determine the number of components i.e. retaining only component that have Eigen value >1 (Table 3). Another criterion for determination of number of component is scree plot that could be used to decide the actual number of components to be retained for analysis. Scree plot can depict various components and the components having eigen value up to the bent of elbow are usually considered (Fig. 1). The following principal component was retained:

PC = 0.370 (TMY/CI) + 0.537 (305MY) + 0.575 (MY/LL) + 0.896 (PY)

Extracted one principal component accounted for 90.49% of the variance in data. The extracted component was represented by significant positive loading for TMY/CI, 305 MY and MY/LL (Table 4). The principal component regression analysis for lifetime milk yield (LTMY4) evolved equation, LTMY4 = $2400.56^{**} + 0.698^{**}$PC could explain

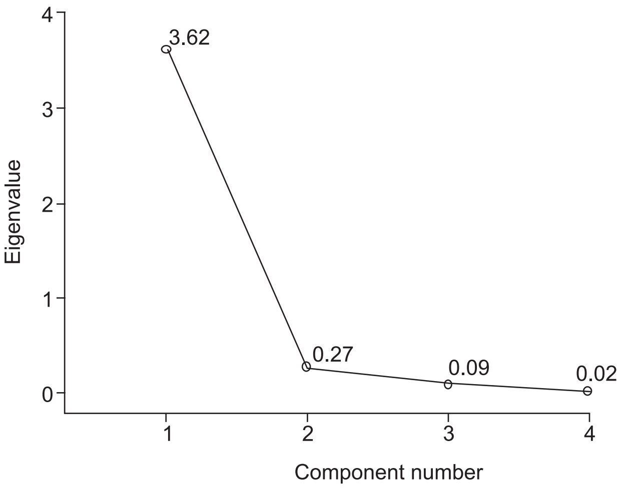

Fig. 1. Scree plot showing component number with Eigen values.

Table 4. Varimax rotated component matrix of different factors for first lactation traits in adult female Jersey crossbred

Trait	Principal component
TMY/CI	0.904
305MY	0.825
MY/LL	0.802
PY	0.436

PY, Peak yield; 305MY, 305 days milk yield; MY/LL, milk yield per day of lactation length; TMY/CI, total milk yield per day of calving interval.

48.8% variation in the estimated values with adjusted R^2 = 48.5%. Khan *et al*. (2013) reported almost similar results in Vrindavani cattle. They evolved an equation for prediction of LTMY up to 4^{th} lactation using part lactation records up to two lactations and their data set explained 54.46% variation in estimated values with adjusted R^2 = 53.20%. Dhara and Chakravarty (1995) compared multiple regression analysis and principal component analysis to predict breeding value of 305 days milk yield and found results of principal component analysis better than multiple regression analysis for prediction. However, Khan *et al*. (2012) formulated the lifetime prediction model using regression analysis based on birth weight, age at first calving (AFC), first service period (FSP), first dry period (FDP), first calving interval (FCI) and first lactation length (FLL) and part lactation records of 100, 170 and 240 days of first lactation, second lactation and their respective total milk yields and explained 40.32% variation in estimated lifetime yields (total of first 4 lactations) in Vrindavani cattle. Bhatacharya and Gandhi (2005) have compared multiple regression analysis and principal component analysis to predict lifetime milk production and found that total variance was lower from the model having PCs as compared to original variables in the regression model. The evolved prediction equation may be helpful in early selection of Jersey crossbred cattle based on first lactation records. Further, the factor extracted can be used in breeding programme with sufficient reduction in the number of production traits to be recorded to explain the lifetime performance of Jersey crossbred cattle.

SUMMARY

Data on first lactation traits viz. peak yield, 305 days milk yield, milk yield per day of lactation length and total milk yield per day of calving interval were used for

prediction of lifetime milk yield (LTMY) upto 4th lactation in Jersey crossbred animals maintained at the organized herd of the Eastern Regional Station, ICAR-National Dairy Research Institute, Kalyani, Nadia, West Bengal using principal component analysis (PCA). The evolved equation, LTMY4 = $2400.56^{**} + 0.698^{**}$PC could explain 48.8% variation in the estimated values with adjusted R^2 = 48.5%. Early selection of Jersey crossbred cattle based on first lactation records can be done with the help of predicted equation. The study also revealed that factor extracted could be used in breeding programs with sufficient reduction in the number of first lactation traits to be recorded for explanation of maximum variability for prediction of lifetime performance traits in Jersey crossbred cattle.

ACKNOWLEDGEMENTS

The authors are thankful to the Director cum Vice Chancellor ICAR-NDRI & Head ERS of ICAR-NDRI, Kalyani for providing the necessary facilities and ICAR for providing financial support. We wish to acknowledge the hard work and sincerity of the staffs of the cattle farm.

REFERENCES

19th Livestock census. 2012. All India Report. *dahd.nic.in/dahd/WriteReadData/Livestock.pdf.*

Bartlett M S. 1950. Tests of significance in factor analysis. *British Journal of Psychology* **3**: 77–85.

Bhatacharya T K and Gandhi R S. 2005. Principal components versus multiple regression analysis to predict lifetime production of Karan Fries cattle. *Indian Journal of Animal Sciences* **75**: 1317–20.

Dhara S K and Chakravarty A K. 1995. Comparison of PC and MR Models for prediction of breeding value of milk production in Murrah buffaloes. *Tropical Agricultural Research* **7**: 168–76.

Gandhi R S and Gurnani M. 1988. Association amongst different productive and reproductive traits in Sahiwal cattle. *Asian Journal of Dairy Research* **7**: 171–74.

Gopal D and Bhatnagar D S. 1972. The effect of age at first calving and first lactation yield on lifetime production in Sahiwal cattle. *Indian Journal of Dairy Science* **25**: 129–34.

Harvey W R. 1990. *Guide for LSMLMW, PC-1 Version, mixed model least squares and maximum likelihood computer programme.* Mimeograph Ohio State University., USA.

Hotelling H. 1933. Analysis of the complex of statistical variables into principal components. *Journal of Educational Psychology* **24**: 417–41.

Johnson R A and Wichern D W. 1982. *Applied Multivariate Statistical Analysis*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, USA.

Khan T A, Tomar A K S and Dutt T. 2012. Prediction of lifetime milk production in synthetic crossbred cattle strain Vrindavani of North India. *Indian Journal of Animal Sciences* **82**: 367–71.

Khan T A, Tomar A K S, Dutt T and Bhushan B. 2013. Principal component regression analysis in lifetime milk yield prediction of crossbred cattle strain Vrindavani of North India. *Indian Journal of Animal Sciences* **83**(12): 1288–91.

Khargharia G, Kadirvel G, Kumar S, Doley S, Bharti P K and Das M. 2015. Principal component analysis of morphological traits of assam hill goat in Eastern Himalayan India. *Journal of Animal and Plant Sciences* **25**(5): 1251–58.

Kumar M, Vohra V, Ratwan P, Valsalan J, Patil C S and Chakravarty A K. 2016. Estimates of genetic parameters for fat yield in Murrah buffaloes. *Veterinary World* **9**(3): 295–98.

Kumar N, Eshetie A, Tesfaye A and Yizengaw H A. 2014. Productive performance of indigenous and HF crossbred in Gondar, Ethiopia. *Veterinary World* **7**(3): 177–81.

Lukibisi F B, Muhuyi W B, Muia J M K, Ole Sinkeet S N and Wekesa W F. 2008. Statistical use and interpretation of principal component analysis in applied research. *Egerton University's 3rd Annual Research Week and International Conference.*

Malhotra P K and Singh P. 1980. Estimation of life-time production in Red Sindhi cattle using ridge-trace criterion. *Indian Journal of Animal Sciences* **50**: 215–18.

Mandal A, Roy P K, Ghosh M K, Chatterjee A and Das S K. 2013. Genetic and environmental effects on first lactation traits of Jersey crossbred cattle in an organized herd of Eastern India. *Indian Journal of Dairy Science* **66**(2): 130–33.

Maxwell A F. 1959. Statistical methods in factor analysis. *Psychological Bulletin* **56**(1): 228–35.

Ministry of Agriculture, GOI. 2012–2013. http: //www.indiastat.com/table/agriculture/2 milk and dairyproducts /167/788613/data.aspx.

NBAGR. 2016. http: //www.nbagr.res.in/registeredbreed.html.

Pearson K. 1901. On lines and planes of closet fit to a system of points in space. *Philosophical Magazine* **2**: 557–72.

Pundir R K and Raheja K L. 1995. Prediction of lifetime milk yield by early traits in Sahiwal and Hariana cows. *Indian Journal of Dairy Science* **48**: 146–19.

Pundir R K, Singh P K, Singh K and Dangi P S. 2011. Factor analysis of biometric traits of Kankrej cows to explain body conformation. *Asian Australasian Journal of Animal Sciences* **24**(4): 449–56.

Puri T R and Sharma K N S. 1965. Prediction of lifetime production on basis of first lactation yield and age at first calving for selection of dairy cattle. *Journal of Dairy Science* **48**: 462– 67.

Rao C R. 1964. The use and interpretation of principal component analysis in applied research. *Sankhya A* **26**: 329–58.

Ratwan P, Mandal A, Kumar M and Chakravarty A K. 2015. Genetic analysis of milk production efficiency traits in Jersey crossbred cattle. *Indian Journal of Animal Research* **51**(4): 644–47.

Ratwan P, Mandal A, Kumar M, Kumar A and Chakravarty A K. 2016a. Genetic analysis of lactation traits in Jersey crossbred cattle. *Indian Journal of Dairy Science* **69**(2): 1–4.

Ratwan P, Kumar M and Mandal A .2016b. Influence of genetic and non-genetic factors on lactation traits in dairy cattle: A Review. *Research & Reviews: Journal of Dairy Science and Technology* **5**(3): 7–22.

Rougoor C W, Sundaram R and van Arendonk J A M. 2000. The relation between breeding management and 305–day milk production determined via principal components regression and partial least squares. *Livestock Production Science* **66**: 71– 83.

Singh M K and Gurnani M. 2004. Performance evaluation of Karan Fries and Karan Swiss cattle under closed breeding System. *Asian Australasian Journal of Animal Sciences* **17**(1): 1–6.

SPSS. 2001. Statistical Package for Social Sciences. SPSS Inc.,

444 Michigan Avenue, Chicago, IL, USA.
Verma N and Thakur Y P. 2013. Effect of genetic and non-genetic factors on production efficiency traits of Red Sindhi × Jersey crossbred cows maintained under sub-temperate Indian conditions. *Livestock Research International* **1**: 58–60.
Vohra V, Niranjan S K, Mishra A K, Jamuna V, Chopra A, Sharma N and Jong D K. 2015. Phenotypic characterization and multivariate analysis to explain body confirmation in lesser known buffalo (*Bubalus bubalis*) from North India. *Asian Australasian Journal of Animal Sciences* **28**: 311–17.
Wondifraw Z, Thombre B M and Bainwad D V. 2013. Effect of non-genetic factors on milk production of Holstein Friesian × Deoni crossbred cows. *International Journal of Livestock Production* **4**(7): 106–12.