Exploring Relationships between Dairy Herd Improvement Metrics in Minas Gerais – Brazil Dairy Herds

Gabriel Machado Dallago[1], Darcilene Maria de Figueiredo[2], Roseli Aparecida Santos[2], Paulo César de Resende Andrade[3], Diego Charles de Almeida Santos[4]

[1]Master Student, Animal Science Department, Federal University of Jequitinhonha and Mucuri Valleys, Diamantina, Brazil. [2]Animal Science Department, Federal University of Jequitinhonha and Mucuri Valleys, Diamantina, Brazil. [3]Science and Technology Institute, Federal University of Jequitinhonha and Mucuri Valleys, Diamantina, Brazil. [4]Executive Director - Holstein Livestock Breeders Association of Minas Gerais, Brazil.

A paper from the Proceedings of the

14th International Conference on Precision Agriculture

June 24 – June 27, 2018

Montreal, Quebec, Canada

Abstract. *The objective of the present study was to apply principal component analysis (PCA) on Brazilian Dairy Herd Improvement (DHI) data to discover the subset of most meaningful variables to describe complete lactations. The Holstein Livestock Breeders Association of Minas Gerais provided data collected between 2005 and 2016 from 122 dairy farms located in the State of Minas Gerais – Brazil. Twelve numerical variables were selected from the original dataset and four additional variables were created. The final dataset contained 28379 observations of 16 numerical variables. They were entered into a Pearson correlation matrix and highly correlated variables (r > 0.94) were evaluated for exclusion based on biological relevance. The PCA was performed on selected variables (n = 12) after they were standardized to mean = 0 and standard deviation = 1. Five variables were PCA-selected as meaningful to describe the variation of complete lactations. They were age at calving, lactation number, milk yield on first test day, energy-corrected milk, and total solid yield on 305 days of lactation. These variables could be used to evaluate complete lactations and future work using Brazilian DHI metrics could focus on modeling the relative importance of each of the selected variables.*

Keywords. *Dairy farms, dairy herd improvement data, multivariate statistics, precision dairy production, principal component analysis.*

Introduction

The overall success of a dairy enterprise rely on the integration of multiple factors. Dairy producers decides on a daily basis about adoption of different technologies and usage of different products while trying to maintain an equilibrium between all factors involved on milk production (McBride and Johnson 2006). However, it is hard to conduct an overall assessment of the activity without biasing towards any single variable, which in turn may or may not effectively evaluate the activity given its multifactorial characteristics.

Using Dairy Herd Improvement (DHI) metrics could be an option to conduct and overall evaluation of dairy production. Dairy breeders association routinely collect information about milk production from associated farms in Brazil similarly as North-American DHI Associations. The advantage of using this information rely on its consistency and availability in addition to describe different aspects related to animal performance (Brotzman et al. 2015). However, such data sets are usually over parameterized making it hard to extract meaningful information from it. Principal component analysis (PCA) is a multivariate statistical technique indicated to analyze quantitative variables from over-parameterized data sets as a variable reduction method, selecting variables that are the most meaningful in describing the variation of data (Borcard et al. 2011).

Therefore, the objective of the present study was to apply PCA on Brazilian DHI data to select the most meaningful subset of variables to describe complete lactations.

Materials and Methods

The Holstein Livestock Breeders Association of Minas Gerais provided the data used in this study from a pre-existing dataset. Therefore, no approval was necessary from the Ethics Committee on the Use of Animals from the Federal University of the Jequitinhonha and Mucuri Valleys in order to carry on this analysis. Data processing and modelling were performed in the statistical software R (R Core Team 2017).

Creating the Working Data Set

Twenty-two variables were collected between 2005 and 2016 from 136 dairy farms located in Minas Gerais State – Brazil resulting in 87193 observations of 45166 animals. Twelve variables were initially selected from the original data set, describing milk production and udder health. Four additional variables were created based on existing information. They were fat to protein ratio (FPR) on complete lactation, energy corrected milk (ECM), age at calving, and calving interval. ECM was calculated using the following equation proposed by Tyrrell and Reid (1965):

$$\text{ECM (kg)} = 12.55 \times \text{fat (kg)} + 7.39 \times \text{protein (kg)} + 0.2595 \times \text{milk yield (kg)}$$

Variables (n = 16) were then entered into a Pearson correlation matrix to check for linear correlations between them. Variables with a greater than 0.94 correlation coefficient were evaluated for exclusion based on biological relevance. Twenty numerical variables were kept for further PCA analysis.

Animals without observations of at least two following lactations were excluded as well as observations with more than 10% of missing values. In addition, outliers were removed based on methodology proposed by Leys et al. (2013). The final data set contained 28379 observations on 17846 dairy cows from 122 dairy farms collected between 2005 and 2016.

PCA

Sixteen variables were entered into a Pearson correlation matrix to check for linear correlations between them. Variables with a greater than 0.94 correlation coefficient were evaluated for exclusion based on biological relevance. Twelve numerical variables were kept for further PCA analysis.

The PCA was performed using the function *rda* from the package *vegan* (Oksanen et al. 2017) on variables scaled to a uniform matrix of mean = 0 and standard deviation = 1. Eigenvalues were calculated to find out the proportion of variation explained by each principal component. Significant eigenvalues were determined by the Kaiser-Guttman (Borcard et al. 2011).

Results

The first 5 eigenvalue dimensions were significant based on Kaiser-Guttman criterion [eigenvalue dimension higher than the average of all eigenvalues dimensions (Borcard et al. 2011)] and are depicted on Figure 1. The first principal component (PC1) with an eigenvalue of 2.84 explained 23.6% of the total variation and the second principal component (PC2) with an eigenvalue of 1.92 explained 16.0% of the variation. Altogether, PC1 and PC2 explained 39.6% of the total variation (Figure 1). Variable contrast was evaluated on all significant eigenvalue dimensions, but many redundancies were found after the PC2. Therefore, PC1 and PC2 were enough for the purpose of this study.

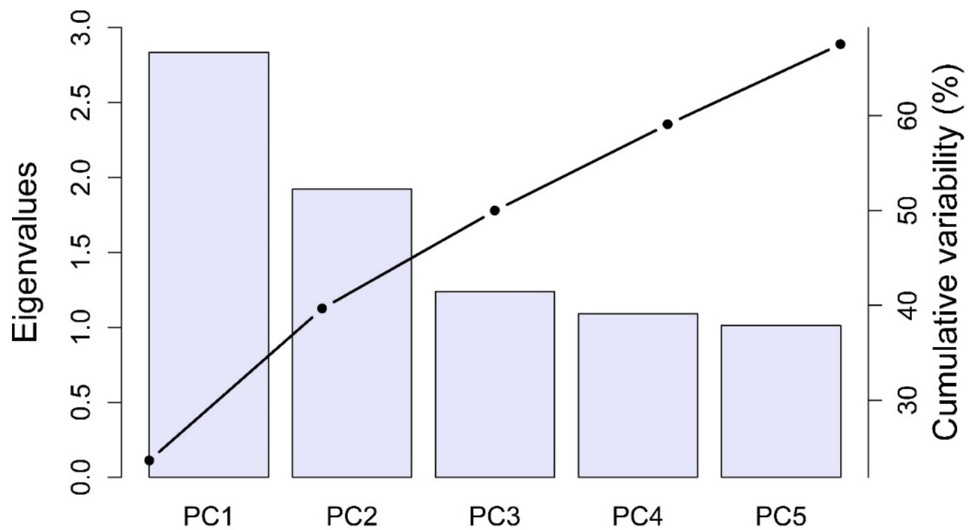

Figure 1. Cumulative variance plot and five significant eigenvalues according to kaiser-Guttman criterion (Borcard et al. 2011) extracted from principal components (PC) generated using principal component analysis (PCA).

The PCA vector ordination plot of PC1 and PC2 are depicted on Figure 2. Total solids yield of 305-day lactation, energy-corrected milk of 305-day lactation, milk yield on first test day, lactation number, and age at calving explained more than average of the total variation. Therefore, they were considered the most meaningful set of variable to describe the variation of complete lactations. Table 1 shows a Pearson correlation matrix of the five PCA-selected variables.

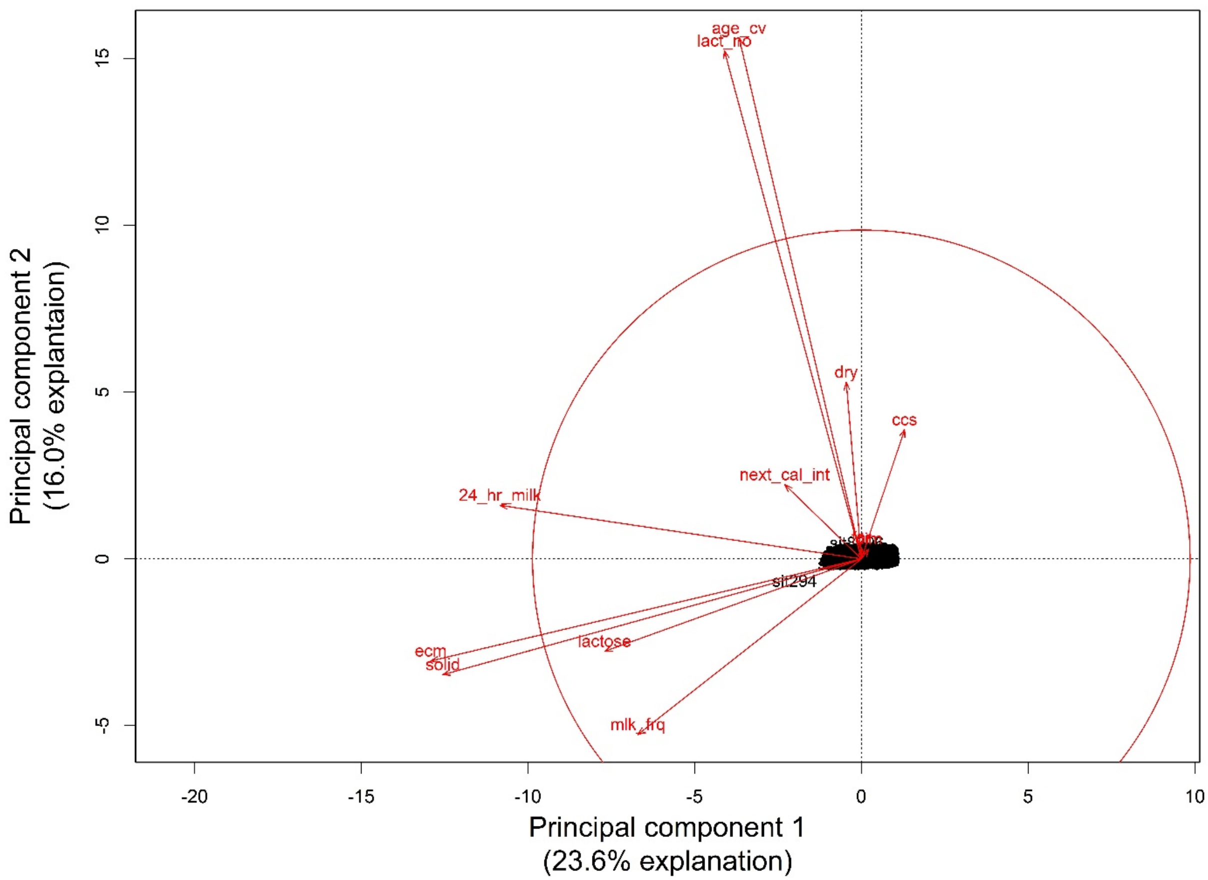

Figure 2. Biplot in the principal component 1 and 2 plane, depicting the directionality of variables and the amount of variation (arrow length) explained by each of them standardized to mean = 0 and standard deviation = 1 versus the mean of eigenvalues (○) for all standardized variables. Each dot in the center represents one observation. ccs = mean somatic cell count of 305-day lactation, standardized; dry = length of dry period between lactations, standardized; age_cv = age at calving, standardized; lact_no = lactation number, standardized; next_cal_int = calving interval, standardized; hr_24_milk = milk yield (kg) on first test day, standardized; ecm = energy-corrected milk (kg) of 305-day lactation, standardized; solid = total solids yield in 305-day lactation, standardized; lactose = total lactose yield in305-day lactation, standardized; mlk_frq = milking frequency, , standardized.

Table 1. Pearson correlation between five principal component analysis-selected variables from Brazilian Dairy Herd Improvement data set.

Item1	age_cv	lact_no	24_hr_milk	ecm	solid
age_cv	1.00				
lact_no	0.83	1.00			
24_hr_milk	0.24	0.24	1.00		
ecm	0.06	0.09	0.57	1.00	
solid	0.04	0.07	0.52	0.89	1.00

[1]age_cv = age at calving; lact_no = lactation number; hr_24_milk = milk yield (kg) on first test day; ecm = energy-corrected milk (kg) of 305-day lactation; solid = total yield of solids in 305-day lactation.

Discussion

We have applied PCA to Brazilian DHI dataset and successfully identified a subset of variables that best describe the variation of complete lactations without biasing the selection towards any single variable. The calculated variable ECM was PCA-selected as more useful to describe the variation than the individual parameters used in its calculation (Figure 2).

Regarding overall herd productive performance, ECM and total milk solid yield in 305-day lactation has been PCA-selected in our study. A close correlation between these two variables can be inferred from Figure 2 and Table 1. Milk production has been previously show to have a negative correlation with reproductive performance of dairy cows (Butler and Smith 1989). However, the magnitude of the negative energy balance from which dairy cows suffer with the onset of a new lactation seems to be closely related its return to reproductive cycle (Nebel and McGilliard 1993; Butler 2000). Milk lactose and protein content has been shown to be good indicators of reproductive performance while milk yield was not (Buckley et al. 2003). Even though lactose and protein have not been PCA-selected in our study, selected variables (i.e. ECM and total solids yield) integrates these variables. In addition, high average ECM was related to low mortality (Alvåsen et al. 2012). Altogether, it indicates the overall usefulness of using such PCA-selected variables to evaluate complete lactations.

Milk yield in the first test day of a new lactation reflects the overall success of the transition period. The transition period is defined as 3-week before and after calving (Drackley 1999; Grummer 1995). Age at calving as well as disorders that occurs during the postpartum transition period impair early postpartum reproductive performance of dairy cows (Fonseca et al. 1983) as well as milk yield (Chapinal et al. 2012; Drackley 1999; Gantner et al. 2016). For instance, Heuer et al. (1999) have found that milk yield on the first test day is a better predictor to common transition period metabolic disorders than body condition score or change of score. Therefore, total milk yield of the first test in addition to age at calving could be used to assess fresh cow management while evaluating complete lactations.

Lactation number is closely related to total milk yield. It is well stablished that multiparous dairy cows produce more 305-day milk than primiparous (Ray et al. 1992). Multiparous cows are also more likely to suffer from metabolic disorders such as subclinical ketoses (McArt et al. 2012) and subclinical hypocalcemia (Reinhardt et al. 2011) than cows on their first lactation (Gröhn et al. 1995). Consequently, it leads to the cascade of events that will result on poor animal performance. Therefore, lactation number is a variable of importance in evaluating complete lactations.

We have demonstrated the effective usefulness of multivariate statistical technique PCA to select meaningful variables to describe complete lactations without biasing toward any single variable. The results here presented could help on evaluating the overall success of dairy enterprises, potentially drawing attention to variation patterns across different variables resulting in a more thorough evaluation.

Conclusion

Five variables were PCA-selected as meaningful to describe the variation of complete lactations. They were age at calving, lactation number, milk yield on first test day, energy-corrected milk, and total solid yield on 305 days of lactation. These variables could be used to evaluate complete lactations and future work using Brazilian DHI metrics could focus on modeling the relative importance of each of the selected variables.

Acknowledgements

Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM) and Programa de Apoio à Participação em Eventos Técnicos-Científicos (PROAPP) for financial support, Associação dos Criadores de Gado Holandês de Minas Gerais – ACGHMG for data supply, and CNPq/CAPES for Masters Scholarship granted to Dallago, G.M.

References

Alvåsen, K., Jansson Mörk, M., Hallén Sandgren, C., Thomsen, P. T., & Emanuelson, U. (2012). Herd-level risk factors associated with cow mortality in Swedish dairy herds. *Journal of Dairy Science, 95*(8), 4352-4362, doi:https://doi.org/10.3168/jds.2011-5085.

Borcard, D., Gillet, F., & Legendre, P. (2011). *Numerical Ecology with R* (1 ed. ed., Use R!). New York: Springer.

Brotzman, R. L., Cook, N. B., Nordlund, K., Bennett, T. B., Gomez Rivas, A., & Döpfer, D. (2015). Cluster analysis of

Dairy Herd Improvement data to discover trends in performance characteristics in large Upper Midwest dairy herds. *Journal of Dairy Science, 98*(5), 3059-3070, doi:http://dx.doi.org/10.3168/jds.2014-8369.

Buckley, F., O'Sullivan, K., Mee, J. F., Evans, R. D., & Dillon, P. (2003). Relationships Among Milk Yield, Body Condition, Cow Weight, and Reproduction in Spring-Calved Holstein-Friesians. *Journal of Dairy Science, 86*(7), 2308-2319, doi:https://doi.org/10.3168/jds.S0022-0302(03)73823-5.

Butler, W. R. (2000). Nutritional interactions with reproductive performance in dairy cattle. *Animal Reproduction Science, 60-61*, 449-457, doi:https://doi.org/10.1016/S0378-4320(00)00076-2.

Butler, W. R., & Smith, R. D. (1989). Interrelationships Between Energy Balance and Postpartum Reproductive Function in Dairy Cattle. *Journal of Dairy Science, 72*(3), 767-783, doi:https://doi.org/10.3168/jds.S0022-0302(89)79169-4.

Chapinal, N., Carson, M. E., LeBlanc, S. J., Leslie, K. E., Godden, S., Capel, M., et al. (2012). The association of serum metabolites in the transition period with milk production and early-lactation reproductive performance. *Journal of Dairy Science, 95*(3), 1301-1309, doi:https://doi.org/10.3168/jds.2011-4724.

Drackley, J. K. (1999). Biology of Dairy Cows During the Transition Period: the Final Frontier? *Journal of Dairy Science, 82*(11), 2259-2273, doi:http://dx.doi.org/10.3168/jds.S0022-0302(99)75474-3.

Fonseca, F. A., Britt, J. H., McDaniel, B. T., Wilk, J. C., & Rakes, A. H. (1983). Reproductive Traits of Holsteins and Jerseys. Effects of Age, Milk Yield, and Clinical Abnormalities on Involution of Cervix and Uterus, Ovulation, Estrous Cycles, Detection of Estrus, Conception Rate, and Days Open. *Journal of Dairy Science, 66*(5), 1128-1147, doi:https://doi.org/10.3168/jds.S0022-0302(83)81910-9.

Gantner, V., Bobić, T., & Potočnik, K. (2016). Prevalence of metabolic disorders and effect on subsequent daily milk quantity and quality in Holstein cows. *Arch. Anim. Breed., 59*(3), 381-386, doi:10.5194/aab-59-381-2016.

Gröhn, Y. T., Eicker, S. W., & Hertl, J. A. (1995). The Association Between Previous 305-day Milk Yield and Disease in New York State Dairy Cows. *Journal of Dairy Science, 78*(8), 1693-1702, doi:https://doi.org/10.3168/jds.S0022-0302(95)76794-7.

Grummer, R. R. (1995). Impact of changes in organic nutrient metabolism on feeding the transition dairy cow. *Journal of Animal Science, 73*(9), 2820-2833, doi:10.2527/1995.7392820x.

Heuer, C., Schukken, Y. H., & Dobbelaar, P. (1999). Postpartum Body Condition Score and Results from the First Test Day Milk as Predictors of Disease, Fertility, Yield, and Culling in Commercial Dairy Herds. *Journal of Dairy Science, 82*(2), 295-304, doi:https://doi.org/10.3168/jds.S0022-0302(99)75236-7.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764-766, doi:http://dx.doi.org/10.1016/j.jesp.2013.03.013.

McArt, J. A. A., Nydam, D. V., & Oetzel, G. R. (2012). Epidemiology of subclinical ketosis in early lactation dairy cattle. *Journal of Dairy Science, 95*(9), 5056-5066, doi:https://doi.org/10.3168/jds.2012-5443.

McBride, W., & Johnson, J. D. (2006). Defining and Characterizing Approaches to Farm Management. *Journal of Agricultural and Applied Economics, 38*(01).

Nebel, R. L., & McGilliard, M. L. (1993). Interactions of High Milk Yield and Reproductive Performance in Dairy Cows. *Journal of Dairy Science, 76*(10), 3257-3268, doi:https://doi.org/10.3168/jds.S0022-0302(93)77662-6.

Oksanen, J., Blanchet, F. G., Friendly, M., Roeland Kindt, Pierre Legendre, Dan McGlinn, et al. (2017). vegan: Community Ecology Package. (Vol. https://CRAN.R-project.org/package=vegan,). R package version 2.4-4.

R Core Team (2017). R: A language and environment for statistical computing. (Version 3.4.3 "Kite-Eating Tree" ed.). Vienna, Austria: R Foundation for Statistical Computing.

Ray, D. E., Halbach, T. J., & Armstrong, D. V. (1992). Season and Lactation Number Effects on Milk Production and Reproduction of Dairy Cattle in Arizona1. *Journal of Dairy Science, 75*(11), 2976-2983, doi:https://doi.org/10.3168/jds.S0022-0302(92)78061-8.

Reinhardt, T. A., Lippolis, J. D., McCluskey, B. J., Goff, J. P., & Horst, R. L. (2011). Prevalence of subclinical hypocalcemia in dairy herds. *The Veterinary Journal, 188*(1), 122-124, doi:https://doi.org/10.1016/j.tvjl.2010.03.025.

Tyrrell, H. F., & Reid, J. T. (1965). Prediction of the Energy Value of Cow's Milk. *Journal of Dairy Science, 48*(9), 1215-1223, doi:doi.org/10.3168/jds.S0022-0302(65)88430-2.