Genetic evaluation of milk yield in Alpine goats for the first four lactations using random regression models

F.G. Silva[1], R.A. Torres[2], L.P. Silva[2], H.T. Ventura[3], F.F. Silva[2], A.P.S. Carneiro[4], M. Nascimento[4] and M.T. Rodrigues[2]

[1]Departamento de Zootecnia e Extensão Rural,
Universidade Federal de Mato Grosso, Cuiabá, MT, Brasil
[2]Departamento de Zootecnia, Universidade Federal de Viçosa, Viçosa,
MG, Brasil
[3]Associação Brasileira dos Criadores de Zebuínos, Uberaba,
MG, Brasil
[4]Departamento de Estatística, Universidade Federal de Viçosa,
Viçosa, MG, Brasil

Corresponding author: F.G. Silva
E-mail: felipe.melhoramento@gmail.com

ABSTRACT. Random regression models have been used in evaluating test-day milk yield, providing accurate estimates of genetic values in animals. However, herd evaluation with only information from the first lactation may not be the best option from an economic perspective. Other factors should be taken into account, particularly other lactations. Our objective in this study was to analyze the genetic divergence between the first four lactations of Alpine goats. The RENPED software was used to perform descriptive statistics, check for errors in pedigree, recode the data, and for Pearson's and Spearman's correlations. The WOMBAT software was used to estimate the variance components and predict the breeding values. The CALC software was adopted to calculate the

percentage of coincidence between the ranking of the animals and the animals kept in common at each lactation evaluation. The results show that selection using only the first lactation in small herds with a low degree of technology can be employed as a palliative measure, in view of the difficulty in evaluating all lactations. However, the selection of breeding goats and the production of catalogues should not be based only on the first lactation, because the results demonstrate inversions in the classification of the best breeders when other lactations are analyzed.

Key words: Animals selected in common; Legendre; Genetic value

INTRODUCTION

The methods of selecting genetically superior individuals in a population are under constant change. Artificial selection has achieved important results for humanity, particularly in terms of the domestication of species.

The genetic evaluation of dairy goats has undergone many advances, such as empirical evaluation, the evaluation of visual traits correlated with milk yield, the evaluation of cumulative milk yield (CMY) using the sire model, the animal model, the simultaneous evaluation of more than one lactation using repeatability models, and multi-trait modes. More recently, random regression models (RRMs) have been used in regard to test-day milk yield (TDMY) instead of CMY. RRMs have the advantage of not only being able to deal with a greater volume of data, but also result in a greater detailing of the phenotype, which allows for more accurate estimates of the genetic values.

Brito et al. (2011) and Silva et al. (2013a) assessed fixed effects that affect CMY and TDMY, respectively. The former demonstrated that fixed effects should be included in single-trait, multi-trait, and repeatability models, whereas the latter demonstrated that fixed effects should be included in RRMs. Although many studies testing RRMs in dairy cattle choose Wilmink's parametric functions (WPF) rather than Legendre orthogonal polynomials (LOPs), Menezes et al. (2010) found that LOP was the better model, which contained the order of four for genetic additive variance and six for permanent individual effects.

Evaluating 980 RRMs using LOPs, Silva (2011) demonstrated that the best RRMs might not be those that have been previously tested. The most parsimonious model, with results identical to the full model using the likelihood ratio test (LRT), was the model which contained the orders of four, two, and seven for the fixed, genetic additive, and permanent individual regressions, respectively, in addition to five classes of heterogeneity of variance.

Brito (2012) evaluated RRMs using equidistant B-spline functions, which improve the predictions when heavily parameterized models are required when using longitudinal data. This type of model reduces the problem of multicollinearity in variance-component matrices. However, Brito (2012) did not find any model that outperformed the best LOP found by Silva (2011), according to the AIC and BIC criteria for the same population. The efficiency of non-equidistant B-spline functions may be better than the models tested by Brito (2012).

Great improvements have been made using RRMs; however, they have two shortcomings: how does one use the results of the curve parameters to select the animals? In addition, they only evaluate one lactation at a time, which is not adequate, since the genes affecting each lactation may not be the same. For the first problem, Silva (2011) suggested integrating the

area below the genetic curve of each animal using the RENPED software. Silva (2011) also provides the user manual and the source code for the software. Lidauer et al. (2000) described a matrix solution to this problem, involving a self-function matrix. The second problem could be solved using an index that would be the sum of the cumulative genetic values found for each of the lactations. Therefore, the animals with high values would be those having the best genes for milk production during their entire productive lives.

The objective of this study was to evaluate genetic divergence between the first four lactations of Alpine goats, using single-trait RRMs.

MATERIAL AND METHODS

TDMY data from goats at the Universidade Federal de Viçosa (Viçosa, MG, Brazil) were used in this study. The animals were reared in a free-stall system, and were fed diets based on corn silage and hay as roughage and a concentrate mix, supplied according to the nutritional requirements of the animals. The milk controls were carried out weekly by mechanical milking performed twice daily. The morning collection began at 6 a.m. and the afternoon collection started at 2 p.m. The TDMY corresponds to the sum of the milk collected in these two periods.

After a critical analysis of the data, there were 50,398 remaining records of TDMY, from 2946 lactations of 1093 Alpine goats. Four new files were generated, each containing information on one of the first four lactations. The descriptive statistics of the TDMY sample values were obtained using the RENPED software (Silva, 2011). The same program was used to correct pedigree errors and recode the alphanumeric data, which are procedures that improve and enable data analysis, respectively.

The WOMBAT software [Meyer, 2007; updated version (September 2013)], was used in all the genetic analyses, providing the estimates of the parameters and predictions of the genetic values for each evaluated animal.

The models adopted in the TDMY evaluations contained combinations of adjustment for the polynomials of the fixed, genetic random, and permanent environmental curves of the orders of four, two, and seven, in addition to the number of classes of residual variances equal to five, following the best model found by Silva et al. (2013b) for the same herd. The single-trait random regression animal model used in each of the four lactations is generally described as follows:

$$y_{ij} = FE_i + \sum_{m=0}^{4-1} b_m \emptyset_m(t_{ij}) + \sum_{m=0}^{2-1} \alpha_{im} \emptyset_m(t_{ij}) + \sum_{m=0}^{7-1} \gamma_{im} \emptyset_m(t_{ij}) + \varepsilon_{ij}$$

where y_{ij} corresponds to the milk yield by goat i in test j, FE_i is the set of fixed effects in which the classifying part is composed of the mean of the genetic group (Alpine > 90% = 1; 80% < Alpine ≤ 90% = 2; 70% < Alpine ≤ 80% = 3; 60% < Alpine ≤ 70% = 4; 50% < Alpine ≤ 60% = 5), the year-season, and the type of parturition. Goat age at calving was included as a covariate with linear and quadratic effects, b_m is the regression coefficient m of the TDMY on a Legendre polynomial to model the mean curve of the population, α_{im} and γ_{im} are the m genetic

additive and permanent environmental regression coefficients, respectively, for goat i, 4-1, 2-1, and 7-1 are the degrees of the Legendre orthogonal polynomials, t_{ij} is the control variable, lactation week of goat i standardized for the interval of -1 to 1, as described by Kirkpatrick et al. (1990), $Ø_m(t_{ij})$ is the Legendre polynomial function for parameter m evaluated for age t_{ij}, and e_{ij} consists of the specific effects at each observation not explained by the correction factors, or by regressions of the model.

We used the same fixed effects as Silva et al. (2013a) to evaluate TDMY, because they were data from the same herd. For distinct herds or traits, the authors recommended carrying out new tests. To study the genetic divergences between the lactations, all the analyses were conducted using the same model, or there may have been confusion between structure and modeling effects.

Model comparison criteria, such as the logarithm of the restricted likelihood function (LogL) and Akaike's information criterion (AIC; Akaike, 1973) cannot be used to compare different databases, because they were designed to compare different models using the same dataset. In addition, their use would confuse the sampling effect and the volume of data, and would not indicate divergences or similarities in the genetic structures that coordinate milk yield in each lactation. However, for illustrative rather than for comparative purposes, only LogL was used.

The condition number (CN; further details in Montgomery and Peck, 1981) is also presented, but for information only, not for model comparison. The LRT (Rao, 1973) could not be used in this study, because all the analyses had the same parameter number.

The results of the genetic values for the parameters of the additive model were integrated using RENPED, in order to obtain the accumulated genetic value of each animal in each lactation. The same software was also used to calculate Pearson's and Spearman's correlations between the genetic values for each lactation. The PROCV function of the CALC software (LibreOffice) was used to calculate the percentage of coincidence of the ranking of the animals, and the percentage of animals selected in common, when 20, 30, and 40% of the animals were discarded. These tests allowed us to check whether the individuals selected in the first lactation were also the best in the other lactations. The same software was also used to calculate the genetic correlation over the different days in lactation (5 to 290) for each lactation, using matrix processes.

RESULTS AND DISCUSSION

Figure 1 shows the fluctuation of the variance (phenotypic, additive, permanent, and residual) components of the model studied according to the number of days in milk, for the first four lactations. The results for the genetic additive variances, permanent, and residual effects were not significantly different when the four lactations were compared. However, the third lactation was slightly different to the others, by exhibiting a lower increase in genetic additive variance at the end of the curve. Some divergence was expected regarding the genes associated with its persistence in relation to the other lactations. This may also be related to alterations in management, feeding, or requirements that may require distinct genes of adaptability in this lactation in relation to the others, or even to some problem with the estimates obtained. For lactations one, two, and four, however, the same genes of production, adaptability, and disease resistance should be expressed in a similar manner.
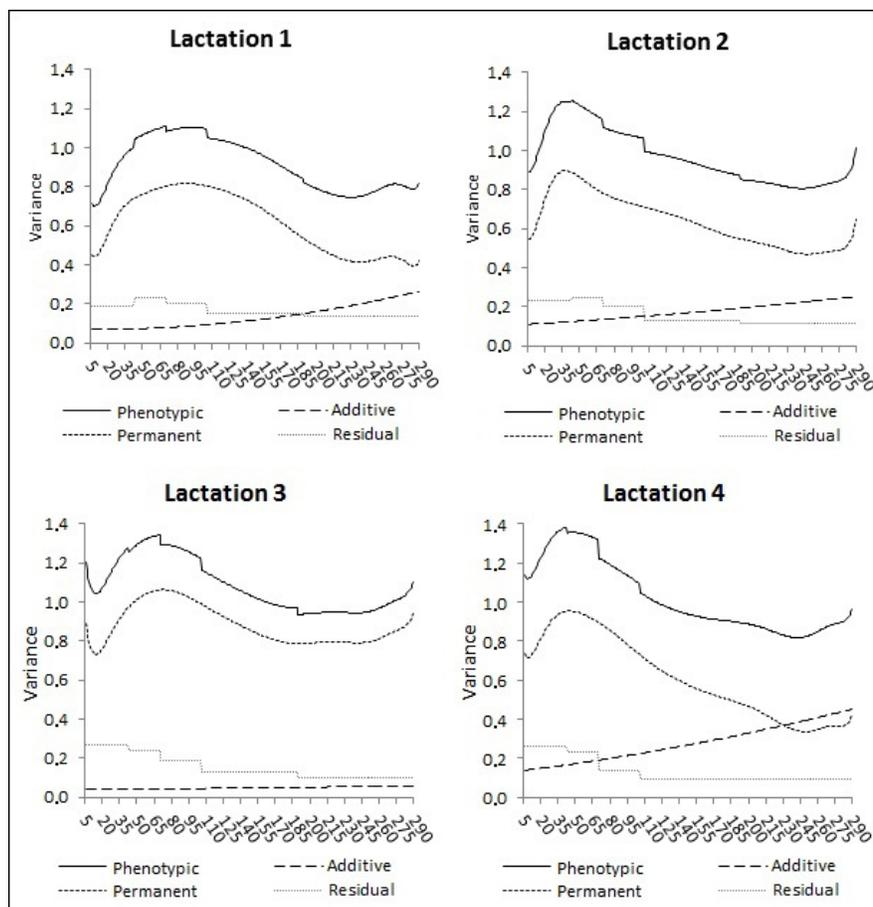
Figure 1. Components of phenotypic, genetic additive, permanent individual, and residual variances obtained as a function of the number of days in milk for the first four lactations in Alpine goats.

The variation in a dataset is termed sample variance; when fixed effects are removed, the variance related to random effects is termed phenotypic variance. In RRMs, part of the phenotypic variance is adjusted in the model by equations related to genetic additive regression and permanent effects. The main difference between the two equations is that the former uses parentage information, whereas the permanent effect equations utilize only information from the same animal from a covariance structure. Permanent regression is sometimes termed permanent environment, and may be incorrect occasionally, because this equation also seeks to adjust non-transmissible genetic effects, like the most part of dominance and epistatic effects.

Figure 2 presents the heritability estimates throughout the days in milk for the first four lactations, identified as L1, L2, L3, and L4. The percentage of phenotypic variation was explained by the regressions of the model. The heritability curves of the four lactations diverged significantly, with the curves of the first and second lactations being a little closer. These divergences may indicate alterations in the genetic structure related to milk yield in

each lactation; not necessarily the genes bound directly to production physiology, but genes of persistence and disease resistance, which had a greater impact on the animals from the third lactation. The changes in the estimates can also be related to other factors, such as the low quality of the estimates as lactation number increases, which can occur due to the ever-decreasing number of observations.
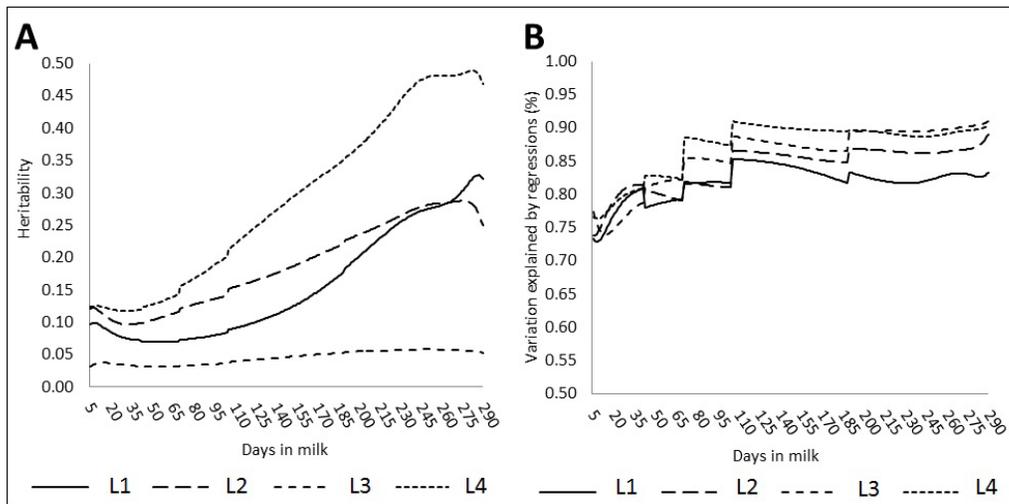
Figure 2. Heritability (**A**) and percentage of the phenotypic variation explained by the regressions utilized in the model (**B**) as a function of the number of days in milk for the first four lactations, identified as L1 to L4.

The correlations between the different days in milk (from 5 to 290) were all high, indicating little variation in the genetic structure that affects milk production throughout the lactation, although the average structure of the population had its characteristic shape. It is unnecessary to represent the results graphically, because the lowest correlation values found between lactation days for each one of the first four calving orders were 95.8, 95.3, 99.7, and 98.1, respectively.

Figure 2 shows that the first lactation exhibited the lowest percentage of variation explained by the regression model over time, even though the model was chosen from the 980 models tested by Silva et al. (2013b); this is probably because in the first lactation there were goats with different ages at first calving, and because this covariate with linear and quadratic effects was not sufficient to correct the variations caused when the female calved too early, which in this case could be a result of many productive- and reproductive-precocity genes being involved. For the age at the second, third, and fourth calvings, however, we expected to find a lower divergence of milk yield in the beginning of lactation due to the age at calving, because all of the animals were sufficiently developed to undergo their reproductive and productive phases. It is also possible that the difficulty in modeling the data in the beginning of the lactations was due to a scale effect, in which higher values will generate greater dispersions. In addition, it is apparent that the stress at the beginning of lactation could result in greater alterations, not only in quantity, but also in proportion, whereas at the end of lactation changes were of a lower magnitude.

Table 1 presents the descriptive statistics for the sample values of TDMY in the four lactations, and the LogL, AIC, CNA (Additive CN), and CNP (Permanent CN) values, which were used as criteria for model comparisons for the different lactations. Too high or undetermined values for CNA and CNP may indicate that the model presents redundant parameters for genetic additive and permanent individual regressions, respectively, for the lactation in question. As shown in Table 1, this occurred for CNA in the third lactation and CNP in the fourth lactation. The LogL values decreased sharply from one lactation to another, which might have been an effect of a reduction in the volume of data. The reduction in the data occurred not only in the number of animals analyzed, but also in the lower proportion of records per animal, which directly compromises the quality of adjustment of regression models. However, in this case this will not serve to contrast the lactations.

Table 1. Descriptive statistics for the sample values of test-day milk yield, containing the number of observations (NO and N%), the lowest (Min), the mean (Mean), the highest (Max) value obtained, the standard deviation (SD), and the criteria for model comparisons in the different lactations (LogL, AIC, and the condition numbers for the genetic additive and permanent individual regressions, termed CNA and CNP, respectively).

Lactation	NO	N%	Min	Mean	Max	SD	LogL	NCA	NCP
1	21.610	42.89	0.1	2.20	6.9	0.9888	4351.303	13.43	182.44
2	14.738	29.24	0.1	2.05	7.5	1.0736	2841.662	199.00	134.78
3	8.776	17.41	0.1	1.96	6.5	1.1063	1626.664	Ind.	362.40
4	5.274	10.46	0.1	1.97	7.2	1.1002	1428.146	231.56	Ind.

The term "Ind." indicates indetermination, which occurred when the lowest estimated eigenvalue was too close to zero, because the lowest eigenvalue found was used in the denominator of the calculation of the condition number.

Table 2 presents the percentages of coincidence, Pearson's and Spearman's correlations, and the percentages of animals kept in common, after the discard of 20, 30, and 40% of the animals with the lowest genetic values. Higher discards can be practiced in highly technological systems, which will increase the selection intensity, which in turn may promote an increase in genetic gain. In herds with a lower technological level, lower discard levels may be performed. The results shown in Table 2 demonstrate that at certain discard rates the percentage of animals kept in common for the four lactations studied was high, which suggests that if the producer conducts a milk control and genetic evaluation, at least for the first lactation, this may result in the selection of animals with superior genes for all lactations.

Table 2. Percentage of coincidence between the ranking of genetic values (GV), as well as Pearson's and Spearman's correlations for the GV, in addition to the percentages of animals kept in common (60, 70, and 80%).

Calving order		Percentage of coincidence	Spearman's correlation (%)	Pearson's correlation (%)	Percentage of individuals kept in common when discarded		
					20%	30%	40%
1	2	0.37	49.10	51.56	88.26	81.53	75.00
1	3	0.51	35.59	36.23	85.22	78.68	69.43
1	4	0.34	41.16	41.77	84.95	77.42	72.38
2	3	0.50	46.68	53.08	87.58	79.17	71.43
2	4	0.25	57.93	52.76	86.33	82.08	78.66
3	4	1.00	31.52	30.18	84.33	76.70	68.20

The previous measure can help improve the herds of small producers, and convince them to begin the milk control activity at least for the first lactation, so they will be able to also have gains in the other lactations. However, in more technologically advanced systems, it is recommended to invest in resources to determine all possible lactations, obtaining responses of greater economic return. The results obtained by Silva et al. (2013b), claiming that less technological systems can obtain good selection responses if milk controls are practiced with maximum intervals of 21 days, may also help convince producers, given that weekly controls, as was the case in the herds under study, can be viewed as a great barrier by producers, due to the global trend in increases in skilled labor.

According to the Pearson's and Spearman's correlation values, and the percentage of coincidence between the ranking of individuals by the genetic values obtained from the evaluation of each lactation, under no circumstances it would be appropriate to recommend the production of catalogues of parents using information of the first lactation only. This is because the dissemination of the best ranked animals in catalogues influences their use. When only information of the first lactation is used, the best parents will not be those that on average will produce the daughters that will provide the best economic return in all lactations.

Our results indicated differences between the additive regressions for each lactation; further studies should be conducted to determine the best model for each lactation, when aiming at new genetic evaluations.

The high degree of importance and applicability of the subject addressed in the present study allows us to suggest future studies on multi-trait RRMs (MRRMs), which have the advantage of considering the existing covariance structure between lactations, better utilizing the data, and reducing problems stemming from a lack of information in the more advanced lactations.

However, great care should be taken in these studies, given that MRRMs also have major disadvantages, such as the large increase in the number of parameters to be estimated by the model, and the difficulty in converging. Therefore, these models and routines are too recent and have not been tested enough, and may lead to biased estimates and poor decisions in the selection process.

CONCLUSIONS

Differences between the genetic structures of each lactation should be studied, so as to properly plan selection strategies in herds of dairy breeds. The herd size and the level of financial and technological resources should be combined with the results of genetic studies, to provide more effective socioeconomic results. Future analyses testing different models for the second, third, and four parturitions, and assessing the use of MRRMs that include all lactations would extend our knowledge of this subject significantly.

ACKNOWLEDGMENTS

REFERENCES

Akaike H (1973). Information Theory and an Extension of the Maximum Likelihood Principle. Proceedings of 2nd International Symposium on Information Theory, Budapest, 267-281.

Brito LF (2012). Modelagem da Produção de Leite de Cabras das Raças Alpina e Saanen Utilizando Regressão Aleatória. UFV, 2012. Master's thesis, Programa de Pós-Graduação em Genética e Melhoramento, Universidade Federal de Viçosa, Viçosa.

Brito LF, Silva FG, Melo AL, Caetano GC, et al. (2011). Genetic and environmental factors that influence production and quality of milk of Alpine and Saanen goats. *Genet. Mol. Res.* 10: 3794-3802.

Kirkpatrick M, Lofsvold D and Bulmer M (1990). Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* 124: 979-993.

Lidauer M, Mantysaari EA, Stradén I and Poso J (2000). Multiple-trait random regression test-day model for all lactations. *Interbull Bull.* 25-81.

Menezes GRO, Caetano GC, Brito LF and Silva FG (2010). Estudo da Produção de Leite em Cabras da Raça Saanen Utilizando Modelos de Regressão Aleatória. In: Anais do VIII Simpósio Brasileiro de Melhoramento Animal Maringá.

Meyer K (2007). WOMBAT: a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *J. Zhejiang Univ. Sci. B* 8: 815-821.

Montgomery DC and Peck EA (1981). Introduction to Linear Regression Analysis. John Wiley and Sons, New York, 504.

Rao CR (1973). Linear Statistical Inference and its Applications. 2nd ed. John Wiley and Sons, New York, 417-420.

Silva FG (2011). Estudo da Produção de Leite de Caprinos Utilizando Modelos de Regressão Aleatória. Master's thesis, Programa de Pós-Graduação em Genética e Melhoramento, Universidade Federal de Viçosa, Viçosa.

Silva FG, Brito LF, Torres RA, Ribeiro Junior JI, et al. (2013a). Factors that influence the test day milk yield and composition. *Genet. Mol. Res.* 12: 1522-1532.

Silva FG, Torres RA, Brito LF, Euclydes RF, et al. (2013b). Random regression models using Legendre orthogonal polynomials to evaluate the milk production of Alpine goats. *Genet. Mol. Res.* 12: 6502-6511.