Classification of lactation curves on French dairy goats

Mathieu Arnal[1, 2], Christèle Robert-Granié[1], Hélène Larroque[1]

[1]GenPhySE, Université de Toulouse, INRA, INPT, ENVT, 31326 Castanet-Tolosan, France
[2]Institut de l'Elevage, 31321 Castanet-Tolosan, France

Abstract. The objective was to achieve a typology of lactation curves and to analyze the influence of environmental factors on these curves. The data used consists of 2,245,628 monthly test-day records of milk on 213,534 Saanen and Alpine French goats. A principal component analysis (PCA) was performed on milk test-day records with R package fdapace. Three principal components were found: one related to the level of milk production, another to the persistency and a third to the shape in middle lactation. The first principal component was mainly explained by herd effect, milk yield estimated breeding value (EBV), parity and region. The second one was mainly explained by month of kidding, breed, age at kidding, dry period length, gestation stage, somatic cell score (SCS) EBV, and parity. The third one was mainly driven by the month of lactation. A classification was used on PCA lactation scores and 5 clusters were found.

Keywords. Lactation curve, lactation persistency, functional clustering, dairy goat.

Titre. Classification des courbes de lactation de chèvres laitières françaises

Résumé. *L'objectif de cette étude était d'obtenir une typologie des courbes de lactation et d'analyser l'influence de facteurs environnementaux sur ces courbes. Les données utilisées consistent en 2 245 628 enregistrements mensuels de la production laitière sur 213 534 chèvres Saanen et Alpines françaises. Une analyse en composantes principales (ACP) a été réalisée sur les productions journalières avec un package R (fdapace). Trois composantes principales ont été trouvées: l'une relative au niveau de la production laitière, l'autre à la persistance et la dernière à la production au milieu de la lactation. La première composante principale a été expliquée principalement par l'effet du troupeau et la valeur génétique pour la production laitière. La seconde a été principalement expliquée par le mois de mise bas, la race et l'âge à la mise bas. La troisième a été principalement influencée par le mois de mise bas. Une classification a été utilisée sur les scores de l'ACP et 5 groupes de courbes ont été identifiés.*

Mots-clés. *Courbe de lactation, persistance laitière, classification de courbes, chèvre laitière.*

I – Introduction

France is one of the world's leading countries in terms of goat milk production thanks to its herds with high genetic potential, composed mainly of two internationally recognized breeds (Alpine and Saanen). This production is carried out by 850,000 goats distributed into: 3,000 farms of 225 goats on average delivering their milk directly to dairies and producing 80% of the French volume, and 2900 farms of 79 goats on average processing part or all of their production (Maigret, 2016). Every 4 to 5 weeks, a technician measures the dairy production of goats called test-day in farms that adhere to the official performance control. All these data are stored in the national database of the Genetic Information Treatment Center (CTIG) of the INRA.

The shape of the lactation curve is of interest for several reasons. First, for the same total production during a lactation, a goat with a "flat" curve is said to be persistent and will have a lower production level at the lactation peak. This is particularly interesting because a high level of production at the lactation peak can lead to an energy deficit of the goat due to its inability to ingest enough food to compensate for its large production. The goat then draws on its body reserves, which can lead to metabolic and reproductive problems (Gipson et Grossman, 1990). A persistent goat is able to be fed with basic forages and few concentrate (Sölkner et Fuchs, 1987), which is an advantage, in a context of new environmental policy measures. It is known that environmental factors influence the shape of the lactation curve (Gipson et Grossman, 1990). The objective of this study was to achieve a typology of lactation curve forms in the two main French dairy goat breeds (Alpine and Saanen) from a large dataset of milk production records from the French national database and to analyze the influence of different environmental factors on these curves.

II – Material and methods

1. Data

After data selection, measurements consist of 2,245,628 monthly test-day from 319 975 lactations of 213 534 French Saanen and Alpine goats collected between September 2008 and June 2012 in 910 French herds by the French official record program. The weights of morning and evening milk produced were added to calculate the daily milk yield of a goat. Lactations should last between 180 and 350 days to be conserved. Test-day records before the 7th and beyond 300th day of lactation were not considered. Within a lactation, a goat was required to have at least 4 test-day records per lactation with the first one occurring before the 81st day of lactation. Only goats from herds having at least 30 conserved lactations per year were kept. On average, a goat had 7 test-days per lactation and Alpine goats represented 60% of the data.

2. Analysis

All statistical analyses were completed using R software. A Principal Component Analysis (PCA) was performed on milk yield test-days, for each lactation, using the R package fdapace (Dai et al., 2016). This package was appropriated to our dataset: sparse data, irregular measurement stage, irregular number of measurements per lactation. Based on the principal component (PC) scores of the lactations, a cluster analysis was performed using the R package Rmixmod (Langrognet et al., 2016). Rmixmod contains a set of functions to fit a mixture model of multivariate Gaussian or multinomial components to a dataset. The optimal number of clusters was researched between 2 and 5 clusters and the Bayesian Information Criterion (BIC) was used to determine it.

The effects of environmental factors on each principal component of PCA were tested with a linear Gaussian model. The fixed effects considered were: the breed, the French herd region, the parity, the age at kidding for the three first parturitions, the kidding month, the dry-period length before their last kidding, the gestation stage at 300 days after the last parturition, the genetic milk estimated breeding value (EBV) based on total lactation milk yield (kg), the somatic cell score (SCS) EBV (high SCS index are correlated with small SCS), with respectively 2, 3, 8, 5, 12, 5, 7, 4 and 5 levels. For each model, a factor was successively removed in order to study its relative impact on each PC using a Fisher test and the determination coefficient (R^2). Least-square means (LS-means) were calculated for each fixed effect to interpret the correlation between the environmental factor and the PC. The influence of the herd on the PC scores was studied from the residuals of each linear Gaussian model.

III – Results and discussion

1. Principal Component Analysis and Classification

In average, goats produced 31.4 hg of daily milk during 283.4 days for a total milk yield of 964 kg on a lactation. The first 3 components explained 99.5% of the original variance: 81%, 15%, and 3.5% for PC1, PC2, and PC3 respectively. Only these three PC were studied due to the high percentage of variance explained. PC1 represented an indicator of the level of production, PC2 a persistency indicator, and PC3 an indicator of linearity deviation in the middle of lactation. Macciotta et al. (2006) performed a PCA on dairy cow correlation test-days milk yield matrix. They did not used all the lactation stage but they summarized the information by the test-day rang in the lactation. They obtained two PC: the first one representing the total lactation production and the second one the persistency. These results were in accordance with ours.

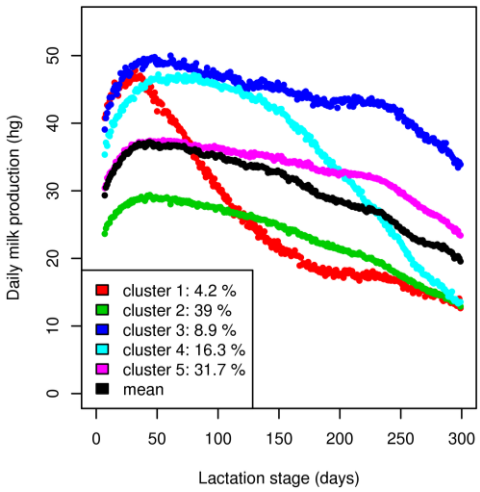

Fig. 1: The average lactation curves for each cluster and the general mean curve (in black) on French dairy goat data. Each point corresponded to the average production of animals in the cluster, and for each cluster, the percentage of lactations was noted.

Five clusters were obtained according to the BIC score. A greater number of clusters was not kept in order to maintain sufficient cluster sizes. Procedures were run several times on the whole dataset and the same clusters were found each time. The averages of daily milk production were plotted according to each cluster number (Figure 1). The cluster 2 was characterized by the lowest total level of milk production and a curve shape similar to the mean curve. The cluster 5 was characterized by a "flat" curve. The mean curves of clusters 1, 3 and 4 (nearly 30% of the data) were very different to the mean curve. Indeed, the cluster 3 had the higher level of milk production and the higher persistency. The cluster 1 had a very marked lactation peak. The cluster 4 was characterized by a "bombed" curve (between the 70th and 200th lactation stage). The mean curves showed a great coherence between two consecutive stages with smooth profiles. The PC3 contributed significantly to the diversity of the shape of curves highlighted and had led to build clusters 1 and 3.

2. Relationship between environmental factors and Principal Components

The milk EBV and the region were the most correlated effects with PC1 with a correlation equal to +0.51.The goats bred in South of France produced less than the goats in North-East which produced less than those in North-West. The breed, the gestation stage, the SCS EBV, the age at kidding, the kidding month and the length of dry period were correlated with PC2. The goats were more persistent when their gestations were less advanced at 300 days after their previous parturition, when they were dried during a short period before their last kidding, when they had a high SCS EBV. Saanen goats were more productive and more persistent than the Alpine goats. The goats which kidded younger had a lower level of milk production but a higher persistency. From the first to the fourth parity, the goats produced in average more milk but their lactations were less persistent at each parturition. After the fourth parity, they produced less and less milk and they were a little bit less persistent than in the fourth lactation. Lactations were less and less persistent when the goats kidded from January to June and were more persistent in the second part of the year. The kidding month was the most linked to the PC3 and presented a great variability with a maximal value in February and a minimum value in August. The effect of the kidding month reflects the effect the evolution of the photoperiod length 120 days later when the bulge of the curve is observed. The influence of herd on residuals of the linear Gaussian

models showed important R^2, equal to 0.397, 0.178 and 0.162 for the 1st PC, 2nd PC and 3rd PC respectively. Bouloc (1991) obtained similar results on French goats for the effects of herd, parity, breed, region and age at kidding. The effect of kidding month on persistency had already been highlighted by León et al. (2012) in Murciano-Granadina dairy goats. This effect could be linked to the duration of the day. The effect of the duration of the day on the lactation curve was also found by Delouis et Mirman (1984): an important exposure to artificial light affects the milk yield for all the lactation stage, until +33% in France. Knight and Wilde (1988) showed that the gestation induce a decrease of milk yield comparing to non-gestating goats. Caja et al. (2006) studied the impact of the dry-period length and they found that goats with no dry-period had a smaller peak and less milk during the lactation than goat dried. The relationship between high SCS EBV and the persistent goats did not seem to be known. A hypothesis could be that goats with low SCS genetic level are more able to protect themselves against diseases allowing to maintain a high milk production. EBVs for type traits were studied and no effect on PCs was found. It was not possible to test here the effect of the number of kids whereas Leon et al (2012) showed that this factor affected the amount of milk and the shape of the lactation curve. Analysis with more data on the animal health, alimentation system, weather,… would be interesting to better understand the large effect of the herd.

IV – Conclusions

This study, based on a large number of goats, allowed to characterize different shape of lactation curves and to analyze the influence of environmental factors on each cluster. Three principal components were found: one relating to the level of milk production, another to the persistency and a third to the shape in middle lactation. Five clusters were found with very different lactation curves. This work is a first step to model the lactation curves according to the environmental factors and to propose a genetic evaluation model to study the persistency.

References

Bouloc N., 1991. Analyse de la forme de la courbe de lactation. Thèse de doctorat INA-PG, Paris

Caja G., Salama A. A. K., Such X., 2006. Omitting the dry-off period negatively affects colostrum and milk yield in dairy goats. In: *J. Dairy Sci.,* (89), 11, p. 4220–4228.

Dai X., Hadjipantelis P. Z., Ji H., Mueller H.-G., Wang J.-L., 2016. fdapace: Functional Data Analysis and Empirical Dynamics. R package version 0.2.5. < https://cran.r-project.org/web/packages/fdapace/index.html > (consulted 18 January 2017)

Delouis C., Mirman B., 1984. Influence de la durée quotidienne d'éclairement sur la production laitière de la chèvre. In:*Compte-Rendu 9e Journ. Rech. Ovine Caprine*. p. 352–360.

Gipson T. A., Grossman M., 1990. Lactation curves in dairy goats: a review. In:*Small Rumin. Res.,* (3), 4, p. 383–396.

Maigret C., 2016. Chiffres clés 2016–Productions caprines. <http://www.idele.fr> (consulted 13 June 2017)

Knight C. H., Wilde C. J., 1988. Milk production in concurrently pregnant and lactating goats mated out of season. In:*J. Dairy Res., (*55), 4, p. 487–493.

Langrognet F., Lebret R., Poli C., Iovleff S., Auder B., Biernacki C., Celeux G., Govaert G., 2016. Rmixmod: Supervised, Unsupervised, Semi-Supervised Classification with MIXture MODelling. R Package Version 2.1.1. < https://cran.r-project.org/web/packages/Rmixmod/index.html > (consulted 18 January 2017)

León J. M., Macciotta N. P. P., Gama L. T., Barba C., Delgado J. V., 2012. Characterization of the lactation curve in Murciano-Granadina dairy goats. In:*Small Rumin. Res.,* (107), 2–3, p. 76‑ 84.

Macciotta N. P. P., Vicario D., Cappio-Borlino A., 2006. Use of multivariate analysis to extract latent variables related to level of production and lactation persistency in dairy cattle. In: *J. Dairy Sci.,* (89), 8, p. 3188–3194.

Sölkner J., Fuchs W., 1987. A comparison of different measures of persistency with special respect to variation of test-day milk yields. In:*Livest. Prod. Sci.,* (16), 4, p. 305–319.